Volume 8 Number 18 1980 Nucleic Acids Research

Nucleotide sequences of nuclear U1A RNAs from chicken, rat and man

Christiane Branlant, Alain Krol, Jean-Pierre Ebel, Eliane Lazar+, Hélène Gallinaro+, Monique Jacob+, Joannes Sri-Widada† and Philippe Jeanteur†

Institut de Biologie Moléculaire et Cellulaire (CNRS), Université Louis Pasteur de Strasbourg, 15 rue René Descartes, 67084 Strasbourg Cédex, France

Received 8 July 1980

SUMMARY

The methods of enzymatic and chemical treatment of end-labeled RNA were applied to the determination of the nucleotide sequence of chicken and man U1A RNA and to the reexamination of that of rat U1A RNA. The chemical method allowed the easy demonstration of the cap structure. All three RNA were 165 nucleotide long. Two hitherto non described modified pyrimidines were detected close to the 5' end. Only 9 base substitutions were observed from chicken to man indicating a high degree of conservation of U1A RNA through evolution.

INTRODUCTION

Metabolically stable low molecular weight RNA were found in the nuclei of a variety of cell types (1-4). Four of these small RNA (4.5SI, U1A, U2 and U3B RNA) isolated from rat Novikoff Hepatoma cells were sequenced by Busch and coworkers (5-8). The interest for the small nuclear RNA was renewed by the finding that they may be hydrogen-bonded to premessenger RNA (9, 10) and that they were present in the ribonucleoproteins containing the premessenger RNA (11-14) which are assumed to be the site of splicing. Therefore, it was proposed that the small RNA might insure the proper alignment of premessenger RNA sequences for splicing (15-17). This idea was reinforced by the observation of at certain complementarity between a single sequence a the 5' end of U1A RNA and a consensus of the intron sequences adjacent to the splice point in premessenger RNA. As premessenger RNA from various animal species were used for the establishment of the consensus sequence and as only the published sequence of rat U1A RNA was available (6), the model implied a high conservation of the primary structure of U1A RNA through evolution (16). This prompted us to determine the complete nucleotide sequence of U1A RNA from 2 other animal species, chicken

and man (HeLa cells). As several differences from the published sequence of rat U1A RNA were observed, we also studied this latter RNA. Indeed, our results show a certain number of discrepancies with the published data.

MATERIALS AND METHODS

1. Isolation of U1A RNA. RNA from nuclear extracts of rat brain, hen liver and HeLa cells was phenol-extracted at 0-4°C, pH 7.6 (18, 19). The 4-10 S RNA were fractionated on 10 or 12 % polyacrylamide gels and U1A RNA was eluted in the presence of 0.1 % Na dodecylsulfate.

2. Preparation of 5' end-labeled U1A RNA. The three step procedure of Efstratiadis et al. (20) was used. The cap structure of U1A RNA was cleaved by tobacco acid pyrophosphatase (a generous gift of Dr M. Pinck) ; the remaining phosphate group was eliminated with alkaline phosphatase ; the resulting free 5' end was then labeled with (γ-^{32}P)ATP (Amersham) and T4 polynucleotide kinase. The RNA was phenol-extracted after each of the two first steps and the labeled RNA was purified by electrophoresis on a 15 % polyacrylamide gel.

3. Preparation of 3' end-labeled U1A RNA. Labeling was carried out with (5'-^{32}P)pCp and T4 RNA ligase according to England and Uhlenbeck (21). 25-100 µCi of pCp (Amersham) and 2-4 U of RNA ligase were used to label 1 µg of RNA.

4. Preparation of 3' end-labeled partial digestion product of U1A RNA. For a limited digestion with S1 nuclease, 1 µg of U1A RNA was first preincubated for 10 min at 37°C in 10 µl of 100 mM KCl, 10 mM Mg acetate, 50 mM Na acetate, 2 mM $ZnCl_2$ adjusted to pH 4.5 with acetic acid. It was then digested for 7 min at 37°C with 5 U of S1 nuclease. The digestion was stopped by phenolic extraction in the presence of 0.1 % Na dodecylsulfate. The resulting fragments were labeled at their 3' end as above and fractionated by electrophoresis on a 15 % polyacrylamide gel.

5. Sequence analysis of end-labeled RNA. Digestion with T1, U2, pancreatic and *Phy* I ribonucleases were previously described (22). The digestion with *Neurospora crassa* nuclease was as described by Krupp and Gross (23). The chemical digestions were performed ac-

cording to Peattie (24). The resulting products were fractionated on 0.5 x 400 x 300 mm polyacrylamide gels in Tris-borate buffer (25).

6. Determination of sequences of large oligonucleotides. U1A RNA was totally digested with T1 RNase. The released oligonucleotides were labeled with (γ-^{32}P)ATP and polynucleotide kinase. Fractionation was by the fingerprinting technique on polyacrylamide gel (26) : a first electrophoresis was performed at pH 3.5, the strip of gel containing the fractionated products was then included at the top of a 20 % polyacrylamide slab gel made up with borate buffer at pH 8.3 and a second electrophoresis was performed.

5' labeled oligonucleotides were digested for 2 hours at 100°C, in bidistilled water (22), and the resulting products were analyzed by electrophoresis on cellulose acetate in the first dimension and homochromatography in the second dimension, using the homomixture described by Brownlee et al. (27).

RESULTS

The same strategy was used to determine the nucleotide sequence of U1A RNA from hen liver, rat brain and HeLa cells. The large T_1 RNase digestion products were first sequenced by the technique of homochromatography. The new RNA sequencing methods based on polyacrylamide gel electrophoresis were then applied to 3' or 5' end-labeled RNA and finally to 3' end-labeled RNA subfragments.

1. Sequence analysis of the large T_1 RNase digestion products. The 5' end labeled oligonucleotides were prepared and fractionated by homochromatography as described under Methods. Cytidine, uridine and adenine positions were determined by mobility shifts (Fig. 1). The results are summarized in Figure 5.

2. Sequence analysis of 5' end-labeled U1A RNA. A cap structure ($m_3^{2,2,7}$GpppAm-Ump) was found at the 5' end of U1A RNA from rat tumoral cells (6, 28). Based on the following evidences, a cap structure was also present in U1A RNA from hen liver, rat brain and HeLa cells : when these RNA were 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 of products with the electrophoretic mobility of U1A RNA.

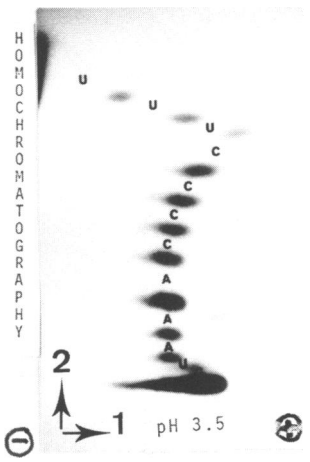

Figure 1 : Sequence analysis of 5'-end labeled T_1 oligonucleotide 1.

The decapped 5' end-labeled RNA were subjected to limited digestions with T_1, pancreatic and U2 ribonucleases and with *Neurospora crassa* nuclease. For this latter enzyme, the conditions were such that all the phosphodiester bonds were cleaved except those of the type C-Xp (23). The RNA were also treated with boiling water. The hydrolysis was random in this case. The digestion products were separated on sequencing gels (Fig. 2). The nucleotide sequence of

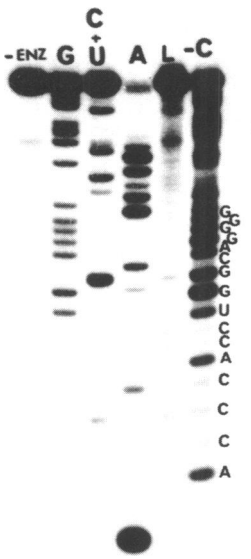

Figure 2 : 25 % sequencing gel of 5'-end labeled rat U1A RNA : from the left to the right digestions were performed with T_1, pancreatic and U2 RNases, boiling water and *Neurospora crassa* nuclease.

the quarter of the molecule located just after the cap structure was deduced from these gels (Fig. 5). The nucleotides from the cap structure were not detected since none of their phosphate linkages could be cleaved by the ribonucleases used in this system.

3. Sequence analysis of 3' end-labeled U1A RNA. U1A RNA were labeled at their 3' end and no labeling of the cap structure was observed under these conditions in spite of the presence of a free 3' OH that might have been a substrate for pCp ligation. The presence of the three methyl groups on the guanine residue might prevent ligation. The labeled RNA were partially digested with T_1, U2, pancreatic and *Phy* I ribonucleases and by chemical reagents (24) (Fig. 3). The digests were fractionated by electrophoresis on 25 %, 20 %, 15 % and 10 % polyacrylamide gels. The 28 nucleotides at the 3' end of the molecules were very resistant to ribonucleases (Fig. 3). Furthermore, in the chemical digests, the 5 guanines in positions 23 to 28 from the 3' end were

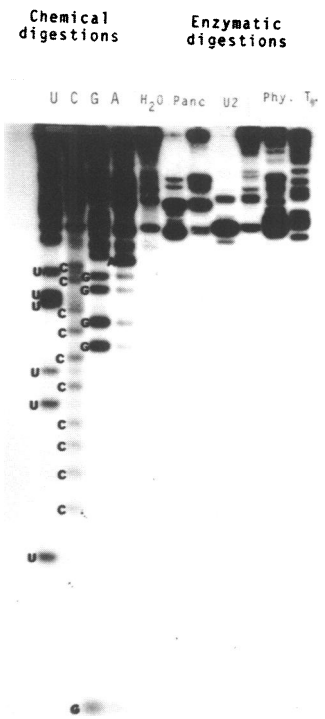

Figure 3 : 25 % sequencing gel of 3' end labeled chicken U1A RNA.

displayed only when electrophoresis was performed at temperature above 60°C. At room temperature, only 3 bands corresponding to guanines were detected. In other words, the 3' end region of U1A RNA possesses a secondary structure that is stable in the presence of 8M urea, at temperature below 60°C. As a consequence, fragments differing by one or two nucleotides had the same electrophoretic mobility and therefore comigrated on the sequencing gel. This stable structure also prevented ribonucleases action. Such band compression due to stable secondary structure was also observed for the series of pyrimidines in positions 64 to 70 from the 3' end. The order of these pyrimidines was deduced from the analysis of T_1 RNAse digestion product 1 (Fig. 1). The general problem of secondary structure of U1A RNA will be dealt with in more details in a subsequent paper.

U2 RNAse did not cleave after the adenosine at position 94 from the 3' end and no cut was observed at this level in boiling water. As these two kinds of cleavage of the phosphodiester bonds require a free 2' OH residue, we concluded to the presence of a 2'-O methyladenosine as already described in rat U1A RNA (6). Analysis of the 3' end-labeled molecules by the chemical sequencing method enabled us to determine the sequence of the 96 nucleotides at the 3' end of the three U1A RNA (Fig. 5).

4. Sequence analysis of 3' end-labeled partial digestion products.

4.1. Completion of the sequence determination. The nucleotide sequence between positions 44 and 69 from the 5' end remained to be determined. For this purpose, the three U1A RNA were partially digested with S1 nuclease and the resulting products were labeled at their 3' end. Two major labeled fragments, denoted S1 fragments 1 and 2, were obtained for each RNA species. S1 fragment 1 covered the 5' half of the molecule, S1 fragment 2 was smaller and corresponded to the 30 nucleotides at the 5' end of the molecule (Fig. 5). These two fragments were analyzed by the chemical RNA sequencing technique. In addition to the nucleotide sequences between positions 44 and 69, these experiments brought new informations on the sequences at the 5' end of the molecules. This will be discussed in the next sections.

4.2. The nucleotides at positions 6 and 7 from the 5' end. As *Neurospora crassa* nuclease hardly cleaved after the pyrimidines at positions 6 and 7 (Fig. 2), they were identified as pyrimidines.

However, the chemical analysis of fragments S2 (Fig. 4) showed that these pyrimidines were not modified by anhydrous hydrazine as should be cytidines nor by aqueous hydrazine as should be uridines. In order to clear this matter up, 5' end-labeled RNA was partially digested in boiling water and the resulting products were fractionated by two-dimensional electrophoresis on polyacrylamide gel, under conditions where cytidines and uridines can be determined by mobility shifts. Pyrimidines at positions 6 and 7 behaved like uridine. This discrepancy between the results obtained by different methods can only be explained by a post-transcriptional modification of these two pyrimidines.

Since these 2 nucleotides were recognized by pancreatic RNAse, they could not be 2'-O-methylated. They had the characteristic charge of uridylic acid, therefore they could be pseudouridylic acid, ribothymidylic acid or dihydrouridylic acid. Little is known on the behavior of these 3 modified nucleotides under the conditions of the hydrazine reaction except that pseudouridylic acid was shown not to be attacked (24). Their sensitivity to the *Neurospora crassa* nuclease is also ignored. In order to have some insight into this problem, 3' end-labeled $tRNA^{Phe}$ from yeast

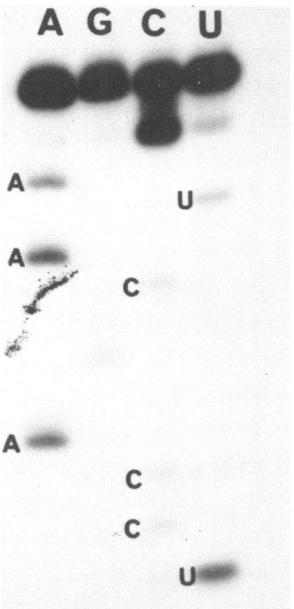

Figure 4 : Fractionation of chemical digests of 3' end-labeled S1 nuclease fragment 1 from chicken U1A RNA by electrophoresis on a 20 % polyacrylamide gel.

which contains the 3 modified nucleotides was submitted to these treatments. Pseudouridylic and ribothymidylic acids behaved like the 2 modified nucleotides of U1A RNA. They were hardly cut after hydrazine modification or by the *Neurospora crassa* nuclease. Resistance of ribothymidylic acid to hydrazine might have been predicted on a chemical basis as the substitution of an hydrogen by a methyl group renders the 5,6 double-bond less sensitive to nucleophilic attack. It is worth mentioning that hydrazine allows the cleavage of deoxythymidine residues in the DNA sequencing method and not that of ribothymidine residues under our experimental conditions. The differences of behavior may be attributed to the use of a higher temperature in the first case than in the second (20°C instead of 0°C).

Like the 2 nucleotides at positions 6 and 7 of U1A RNA, dihydrouridylic acids from $tRNA^{Phe}$ were not cleaved by *Neurospora crassa* nuclease. In contrast to these 2 nucleotides, they were attacked by both aqueous and anhydrous hydrazine under the conditions used for RNA sequencing. This was rather unexpected since the 5,6 double-bond does not exist in dihydrouridine. The mechanism of the reaction might be different from that proposed for uridine and cytidine.

According to these data, the two modified nucleotides could not be dihydrouridylic acids but pseudouridylic or ribothymidylic acids or evidently other modified nucleotides having the same characteristic behavior. Reddy et al. (29) who studied U1A RNA from Novikoff Hepatoma ascites cells found no modified nucleosides at positions 6 and 7.

4.3. The cap structure. Experiments like those of Figure 4 also brought information on the cap structure itself. Firstly, Ump and Amp residues were cleaved by chemical digestion and could be ordered. Secondly, cleavage products of the pyrophosphate bonds in the cap structure were observed in the column corresponding to cytidine. This may be explained by a nucleophilic attack of the second and mainly of the third phosphate of the cap structure by anhydrous hydrazine. Aqueous hydrazine was probably not nucleophilic enough to allow the reaction. Since this reaction and the corresponding pattern on the sequencing gel are unique, they represent an easy way to demonstrate a cap structure at the end of a molecule. We already

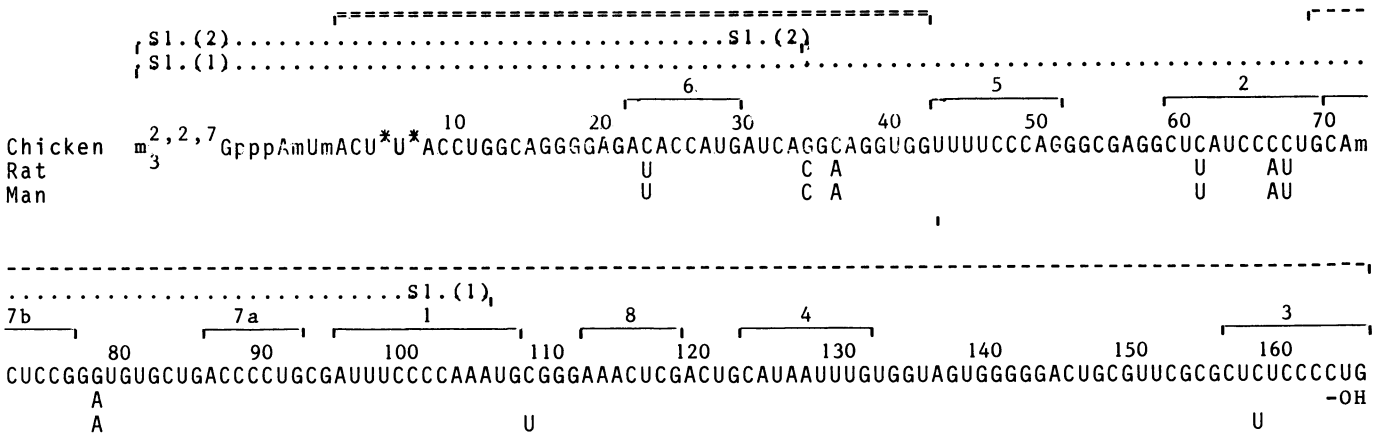

Fig. 5 : Primary structures of chicken, rat and man U1A RNA. The nucleotide sequence of chicken U1A RNA is indicated. Hyphens were omitted for reasons of space. U* indicates modified uridines as discussed in the text. Only the changes of sequences relative to chicken U1A RNA are given for rat and man U1A RNA. The lines above the sequences assign the regions sequenced by each method. (——) T_1 RNAse oligonucleotides.
(....) S1 nuclease fragments 1 and 2 ; (=====) sequence from the 5' end of the molecule ;
(----) sequence from the 3' end of the molecule.

made use of these features in a study of other nuclear RNA (unpublished results). Combined with the results obtained by the other techniques, the study of partial digestion products from U1A RNA allowed us to establish the complete nucleotide sequence of U1A RNA from chicken, rat and man (Fig. 5).

DISCUSSION

Complete primary structures of small nuclear RNA were first reported by Busch and coworkers (5-7) and these sequences supplied a number of fundamental informations. But due to difficulties of interpretation inherent to the fingerprinting technique that were formerly used, errors were made in the determination of the primary structure of U1A RNA. This RNA was 165 nucleotide long in the three animal species studied instead of 171 as previously published. The new method for RNA sequencing were far more reliable and we routinely employed both chemical end enzymatic techniques.

The expected phylogenetic relationships between chicken, rat and man were apparent in the nucleotide sequence of U1A RNA (Fig. 5). There were 7 mutations from chicken to rat. They were conserved in human U1A RNA which presented 2 additional mutations. This indicates a high degree of conservation, 96 % from chicken to rat, 99 % from rat to human U1A RNA, or 94 % between chicken and man. This is close to that observed between chicken and human 5 S RNA (95 %), but lesser than that observed between chicken and human 5.8 S RNA (99 %) (30).

The localization of the mutations was not random along the molecule. The 7 mutations between chicken and rat were localized in the 5' half of the molecules. In contrast, the 2 additional mutations in human U1A RNA were detected in the 3' half of the molecule. The 18 nucleotides at the 5' end of the molecule which are involved in splicing models (16, 17) were conserved.

Also, the nature of the mutations was not random. All of them were substitutions. In 7 cases, cytidine residues were replaced by adenine or uridine residues. In the 2 other cases, guanine residues were replaced by a cytidine and an adenine residue. In other words, these mutations tended to diminish the G + C content of the RNA (30). The overall G + C content of U1A RNA was close to that of 5 S and 5.8 S cytoplasmic RNA. However, no decrease of G + C content was observed from chicken to man for 5 S and 5.8 S RNA (30). This suggests diffe-

rent ways of evolution for these small-size RNA.

While it has been claimed that small nuclear RNA are encoded by multiple genes, of the order of 2,000 for each individual RNA (31), no sequence heterogeneity of U1A RNA was observed. Furthermore no difference was detected between the sequences of U1A RNA isolated from chicken brain and liver. This is the same situation as observed for 5 S and 5.8 S RNA.

We examined the sequence of U1A RNA for its possible coding capacity. The AUG codon at position 27 had an UGA termination codon in phase at position 84 and the GUG codon at position 80, had an UAA termination codon in phase at position 125. The encoded peptides would be short, 19 and 15 amino acids, respectively. The longest message would contain a 2'-O-methyladenosine whose function in coding is unknown. In addition, the termination codons were not repeated as is usually the case in messenger RNA. Therefore, the probability that U1A RNA codes for small peptides seems quite low.

U1A RNA contains a few modified nucleotides. In addition to the cap structure and to the 2'-O-methyladenylic acid at position 71, we detected 2 modified uridylic acids close to the 5' end of the molecule. They behave like pseudouridylic or ribothymidylic acids and other mehtods will be required for unambiguous determination. This will require much more material than used for the whole nucleotide sequence (5 µg). But this might be worthwile in as much as these nucleotides are included in the sequences of U1A RNA assumed to play a role in premessenger RNA splicing. In this respect, it should be mentioned that a guanine residue at position 19 was missing from the sequence used by the 2 groups who proposed this model (16, 17). The implications of these results on the possibility of base-pairing between U1A RNA and the consensus sequence at the extremities of introns will be discussed in a forthcoming paper dealing with the secondary structure of U1A RNA.

ACKNOWLEDGEMENTS

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 à la Recherche Sientifique et Technique". We are very grateful to Dr M. Pinck and Dr J.P. Bargetzi for the generous gift of tobacco pyrophosphatase and *Phy* I RNAse. The excellent technical assistance of B. Muller and L. Kister is acknowledged.

+Laboratoire de Génétique Moléculaire des Eucaryotes (CNRS), Unité de Biologie Moléculaire et de Génie Génétique (U184, INSERM), 11 rue Humann, 67085 Strasbourg Cédex, France.

†Laboratoire de Biologie Moléculaire, Université des Sciences et Techniques du Languedoc, 34060 Montpellier Cédex, France.

REFERENCES

1. Larsen C.J., Galibert F., Lelong J.C. & Boiron M. (1967) C.R. Acad. Sci. Paris, 264, 1523-1526.
2. Prestayko A.W. & Busch H. (1968) Biochim. Biophys. Acta, 169, 327-337.
3. Weinberg R.A. & Penman S. (1968) J. Mol. Biol. 38, 289-304.
4. Hellung-Larsen P., Frederiksen S. & Plesner P. (1971) Biochim. Biophys. Acta, 254, 78-90.
5. Ro-Choi T.S., Reddy, R., Henning D., Takano T., Taylor C.W. & Busch H. (1972) J. Biol. Chem., 247, 3205-3222.
6. Reddy R., Ro-Choi T.S., Henning D. & Busch H. (1974) J. Biol. Chem., 249, 6486-6494.
7. Shibata H., Ro-Choi T.S., Reddy R., Choi Y.C., Henning D. & Busch H. (1975) J. Biol. Chem., 250, 3909-3920.
8. Reddy R., Henning D. & Busch H. (1979) J. Biol. Chem., 254, 11097-11105.
9. Flytzanis C., Alonso A., Louis C., Krieg L. & Sekeris C.E. (1978) FEBS Lett., 96, 201-206.
10. Jelinek W. & Leinwand L. (1978) Cell, 15, 205-214.
11. Deimel B., Louis C. & Sekeris C.E. (1977) FEBS Lett., 73, 80-84.
12. Northemann W., Scheurlen M., Gross V. & Heinrich P.C. (1977) Biochem. Biophys. Res. Commun., 76, 1130-1137.
13. Guimont-Ducamp C., Sri-Widada J. & Jeanteur P. (1977) Biochimie, 59, 755-758.
14. Gallinaro H. & Jacob M. (1979) FEBS Lett., 104, 176-182.
15. Murray V. & Holliday R. (1979) FEBS Lett., 106, 5-7.
16. Lerner M.R., Boyle J.A., Mount S.M., Wolin S.L. & Steitz J.A. (1980) Nature, 283, 220-224.
17. Rogers J. & Wall R. (1980) Proc. Nat. Acad. Sci. U.S.A., 77, 1877-1879.
18. Stévenin J., Gallinaro H., Gattoni R. & Jacob M. (1977) Eur. J. Biochem., 74, 589-602.
19. Gattoni R., Stévenin J. & Jacob M. (1980) Eur. J. Biochem., 108, 203-211.
20. Efstratiadis A., Vournakis J., Donis-Keller H., Chaconas G., Dougall D. & Kafatos F. (1977) Nuc. Acids Res., 4, 4165-4174.
21. England E. & Uhlenbeck O. (1978) Nature, 275, 560-561.
22. Branlant C., Krol A., Machatt M. & Ebel J.P. (1979) FEBS Lett., 107, 177-181.
23. Krupp G. & Gross H. (1979) Nuc. Acids Res., 6, 3481-3490.
24. Peattie D. (1979) Proc. Nat. Acad. Sci. U.S.A., 76, 1760-1764.
25. Sanger F. & Coulson R. (1978) FEBS Lett., 87, 107-110.
26. De Wachter R. & Fiers W. (1972) Anal. Biochem., 49, 184-197.
27. Brownlee G.G. & Sanger F. (1969) Eur. J. Biochem., 11, 395-399.
28. Ro-Choi T. & Henning D. (1977) J. Biol. Chem., 252, 3814-3820.
29. Reddy R., Ro-Choi T., Henning D., Shibata H., Choi Y. & Busch H. (1972) J. Biol. Chem., 247, 7245-7250.
30. Erdmann V. (1980) Nuc. Acids Res., 8, r31-r48.
31. Engberg J., Hellung-Larsen P. & Frederiksen S. (1974) Eur. J. Biochem., 41, 321-328.