An improved rapid enzymatic method of RNA sequencing using chemical modification

A.M. Mazo, T.D. Mashkova, T.A. Avdonina, N.S. Ambartsumyan, and L.L. Kisselev

Institute of Molecular Biology, Academy of Sciences of the USSR, Moscow, V-334, USSR

Received 31 July 1979

ABSTRACT

A version of rapid gel sequencing procedure based on the analysis of partial endonuclease hydrolyzates of chemically modified 5'-32P-labelled RNA is suggested. Complete and selective modification of cytidilic residues by a methoxyamine-bisulfite mixture leads to the unfolding of the RNA secondary structure and, due to this effect, to the generation of a more uniform set of fragments after partial RNAase hydrolysis. The position of cytidines in an RNA sequence can be determined by restricting the hydrolysis of phosphodiester bonds between the modified CMP residues and their 3'-neighbours with Tp and A RNAases. The method was verified with tRNA Trp (yeast) and 5S RNA (rat liver and yeast).

INTRODUCTION

Several methods have been developed recently for rapid RNA sequencing. The general principle of these methods consists in the generation of a set of labelled fragments terminated with a unique type of RNA residue followed by electrophoretic separation on polyacrylamide gel, namely: a procedure using partial specific endonuclease hydrolysis of 5'-32P-labelled RNA; a "plus-minus" method; a technique based on specific termination with triphosphate analogues in cRNA or cDNA synthesis catalyzed with Qp replicase and reverse transcriptase on an RNA template and some others. However these methods have been unable so far to give complete information about RNA sequences although they are very rapid. In the first procedure, this is due to unaccessibility of some structured regions to RNAase hydrolysis. Moreover, in the conditions of partial endonuclease hydrolysis, a pronounced difference in the rate of cleavage of phosphodiester bonds is revealed as well as insuffi-
cient specificity of some endonucleases, especially pyrimidylic ones. As a consequence of these effects, many bands are not present or weakened in the gel, preventing the unambiguous determination of an RNA sequence.

Other procedures mentioned above are also not free of shortcomings which we do not discuss here. Up to now they are not widely used for estimation of unknown RNA sequences.

In this paper, we tried to eliminate some disadvantages of the endonuclease method by using chemically modified RNA. Treatment of nucleic acids with a methoxyamine-bisulfite mixture selectively converts cytidine into $N^4$-methoxy-5,6-dihydrocytidine-6-sulfonate ($C^*$). Appearance of an $SO_3^-$ group in the modified base presumably destroys hydrogen bonds in G-C base pairs and is followed by melting of the RNA secondary structure. It is known, for example, that modification of poly(U) with bisulfite which introduces an $SO_3^-$ group in the adduct greatly decreases the melting point of the poly(U).poly(A) double-helical complex. The melting of the RNA secondary structure should yield a more uniform set of fragments after enzymatic hydrolysis, manifesting itself in the number and intensity of radioactive bands in the gel.

We have shown earlier that chemical modification of the bases in polynucleotides prevents enzymatic cleavage of the phosphodiester bonds between the modified nucleotide and its 3'-neighbour. In particular, after exhaustive selective modification of cytosines by a methoxyamine-bisulfite mixture RNAses A and T$_2$ split RNA in the conditions of complete hydrolysis only after U or A, G and U, respectively. Basing on these data, we suggest here to perform the sequencing using premodified RNA and T$_1$, U$_2$, T$_2$ and A RNAase hydrolysies as described earlier.

The proposed method of RNA sequencing has been tested with yeast tRNA$_{Trp}$ and ribosomal 5S RNAs from rat liver and yeast with known primary structures.

MATERIALS AND METHODS

Yeast tRNA$_{Trp}$ (80% purity) was provided by Dr. G. Keith,
rat liver 5S RNA by Dr. M. Saarma. Yeast 5S RNA was isolated from total low molecular weight yeast RNA by preparative electrophoresis in 10% polyacrylamide gel. T₁, T₂ and U₂ RNAs were purchased from Sankyo. Staphylococcus aureus nuclease from Worthington, alkaline phosphatase (E.coli) from Sigma. RNAase A was the gift from Dr. Yu. Lebedev, and RNAase Phy I (Physarum polycephalum) from Dr. Yu. Kozlov. Polynucleotide kinase was isolated from E.coli infected with T4 phage as described, [32P]-orthophosphate (carrier-free) was from Radiochemical Centre, Amersham, Sephadex G-50 (fine) from Pharmacia Fine Chemicals, γ-32P ATP was synthesized as described 1.

Chemical modification of RNA. RNA preparations (10-50 μg) were treated in 20-100 μl with a mixture of 1.5 M CH₃ONH₂ (pH 5.0) and 1 M Na₂S₂O₃ at 37° for 16 hrs. The material was desalted on a Sephadex G-50 (fine) column and kept in 0.1 N NH₄OH at 20° for 3 hrs to demodify the partially modified uridines 16.

5'-end labelling of RNA. RNA (10-50 μg) was treated with alkaline phosphatase (1 U per mg RNA) in 0.1 ml of 0.1 M Tris·HCl (pH 8.2) at 60° for 30 min. After the incubation, RNA was precipitated with CETAVLON (to a final concentration of 1 mg/ml in 0.1 M sodium phosphate buffer, pH 5). The RNA was twice reprecipitated with EtOH from 1 M NaCl and once from 0.2 M Na·Ac, pH 5. The pellet was rinsed with 70% EtOH and dried. For the 5'-end labelling, γ-32P ATP (500-1500 Ci/mmole) and T4 kinase were used and 5'-32P-labelled RNA was purified in 10% polyacrylamide gel as described 1.

Partial RNA hydrolysis. 5'-32P-labelled RNA (1 μg, 10⁴-10⁵ cpm) was digested with RNAases T₁ and A in 0.1 M Tris·HCl (pH 7.5), 10 mM EDTA; with U₂, T₂ and Phy I in 50 mM Na·Ac (pH 4.5), 2 mM EDTA; with Staphylococcus nuclease in 30 mM Tris·HCl (pH 9), 1.4 mM CaCl₂. The reaction was run in 5 μl for 30 min at 0° except for Phy I which was run at 20°. After hydrolysis, 5 μl of formamide containing 0.05% xylene cyanol was added and the solution was kept at -40°. The hydrolysates of non-modified RNAs were heated for 1 min at 100° and quickly cooled in ice before applying on the gel. Partial non-specific hydrolysis of RNA was achieved by boiling 1 μg of RNA for 10-30
min in 0.1 ml of bidistilled water followed with evaporation of
the water till dryness. The sample was dissolved in 5 μl of
0.1 M Tris•HCl (pH 7.5), 10 mM EDTA, and 5 μl of formamide with
0.05% xylene cyanol was added.

The electrophoresis was run in 10% polyacrylamide gels con-
taining 7 M urea 1. The gels were exposed with X-ray film RT-1
at -70°.

RESULTS

Chemical modification leads to weakening of the secondary
structure and more uniform nuclease hydrolysis. Transfer RNA
molecules are known to possess rather stable secondary struc-
ture. Therefore, tRNA would be a good model to test the hypo-
thesis that, after chemical modification of C bases, the set of
enzymatically produced fragments can be random since the influ-
ence of the secondary structure on the hydrolysis rate is eli-
minated. In fact, the hypochromicity of total yeast tRNA is
about 25% at 260 nm and 0.003 M Mg$^{2+}$, 0.14 M NaCl, 0.01 M
Tris•HCl (pH 7.5) while it drops down to ~8% after modification
of C in tRNA.

We sequenced tRNA$^{\text{Trp}}$ (yeast) under standard conditions 2
before and after its chemical modification. After partial hydro-
lysis of modified RNA (C*) with T$_1$, U$_2$ and Phy I RNAases, we ob-
tained a set of fragments (Fig. 1b) which was markedly more
rich in the number of oligonucleotides than after respective
hydrolysis of non-modified RNA (Fig. 1a). A particularly great
difference was observed after T$_1$ RNAase hydrolysis: in the case
of RNA (C*), bands appeared on the film which represented G21,
23, 29, 51, 52, 63 and which were absent when non-modified RNA
was analysed. These nucleotides are indicated on the clover-
leaf model of tRNA$^{\text{Trp}}$ (Fig. 1c). It is easy to notice that all
of the fragments which appeared only with tRNA$^{\text{Trp}}$ (C*) contain-
ed terminal G's involved in the secondary structure via G-C base
pair formation. Obviously, modification of the C caused a dra-
matic weakening of the secondary structure and an appearance of
new bands derived from the regions of the molecule previously
buried from hydrolysis. Due to the melting of the secondary
structure, the $A_{20}-G_{21}$ and $A_{22}-G_{23}$ bonds became more accessible to $U_2$ RNAase hydrolysis too.

After RNA(C*) hydrolysis with RNAase A, all fragments should be terminated with $U$. The absence of some bands representing $U$ from the RNAase A hydrolysate (Fig. 1b) might be explained in terms of a low rate of $U-U$ and $U-C$ cleavage and, respectively, $U-C*$ with this nuclease. The same is probably true of Phy I RNAase although this nuclease provides some additional information concerning the position of $U$ due to the fact that, after splitting with this enzyme, some new bands appeared which were not observed with RNAase A.

The appearance of a band representing $C_{13}$ in the RNAase A hydrolysate of RNA (C*) is explained by the presence of an analogous band in -E track (Fig. 1b). In this particular case, this position might be identified either by the absence of the respective band in the Phy I hydrolysate or due to a much more intensive band in the RNAase A hydrolysate of non-modified tRNA (Fig. 1a).

In our hands chemical modification of $C$ is more efficient in destroying the RNA secondary structure than elevation of temperature and use of 7 M urea as a denaturing agent.

The low rate of RNAase A hydrolysis of $U$-Pyr bonds, as well as the presence of methylated nucleotides in tRNA$^{Trp}$, does not permit to obtain the whole sequence after hydrolysis of non-modified RNA.

However, by means of hydrolysates of tRNA$^{Trp}(C*)$, all uncertainties with the determination of the $C$ positions were eliminated as well as almost all ambiguities with the $A$ positions.

Combination of the data derived from hydrolysates of modified and non-modified RNAs improves the fidelity of sequence determination. As stated above, the structural information obtained from RNA(C*) hydrolysates is much greater than that from non-modified RNA hydrolysates. However, it does not mean that nothing can be extracted from the hydrolysates of non-modified RNAs. For example, some pyrimidines ($C_{13}, \Psi_{27}$) have been precisely identified only after comparison of RNA(C) and RNA(C*) hydrolysates.

Since tRNA$^{Trp}$ has a very compact three-dimensional struc-
ture, we continued the analysis with ribosomal 5S RNAs from yeast and rat liver which were probably not so tightly folded.

Partial RNAase A hydrolyses of rat 5S RNA yielded much more fragments than those of tRNA\textsuperscript{Trp} or yeast 5S RNA (Figs 2a and 3, track A). For instance, in rat 5S RNA, RNAase A has split oligopyrimidine stretches consisting of 3 or even 4 pyrimidines which is not observed with 5S RNA (yeast) or tRNA\textsuperscript{Trp}. Probably, this behaviour is connected with a relatively weak secondary structure of rat 5S RNA.
Fig. 1. Autoradiogram of partial digests of 5'-$^{32}$P-labelled yeast tRNA$^{Trp}$ analyzed on 10% polyacrylamide gel. Non-modified tRNA$^{Trp}$, (a) and with modified C residues (b). Tracks: -E, no endonuclease; L, Ladder was prepared as described in Materials and Methods; T$_1$, incubation with RNAase T$_1$ (mixture of hydrolysates with $10^{-3}$, $10^{-4}$, $10^{-5}$ U of RNAase T$_1$; U$_2$, incubation with RNAase U$_2$ ($10^{-4}$, $10^{-5}$, $10^{-6}$ U); A, incubation with RNAase A ($10^{-4}$, $10^{-5}$, $10^{-6}$ U); Phy, incubation with RNAase Phy I as described. The samples were loaded on the polyacrylamide gel (40 x 13 x 0.06 cm) and electrophoresis proceeded at 1 kV for 6 hrs. The nucleotides are numbered from the 5' end of the molecule.

c. Clover-leef secondary structure of tRNA$^{Trp}$ (yeast). Arrows indicate the nucleotides which appear as the bands on the autoradiogram only after hydrolysis of modified tRNA$^{Trp}$.

The sequence analysis of the rat liver 5S RNA fragment 10-61 gave the following information (Fig. 2): for RNA(C*) we couldn't determine the type of residues in positions 16, 25, 27, 43, 45, 51, 56 and for RNA(C) – in positions 16, 19, 26, 28-31, 33, 42, 44, 52, 58-61. In summary, the only unidentified position is A$_{16}$.
Fig. 2. Autoradiogram of partial digests of 5'-32P-labelled rat liver 5S RNA analysed on 10% polyacrylamide gel. Non-modified RNA (a) and with modified C residues (b). Tracks and electrophoresis conditions as in Fig. 1; T2, incubation with RNAase T2 (10^{-3}, 10^{-4}, 10^{-5} U). L is the mixture of T1, U2 and A digests.
Fig. 3. Autoradiogram of partial digests of $5'-^{32}P$-labelled modified at C residues yeast 5S RNA analyzed on 10% polyacrylamide gel.

Tracks as in Fig. 1; St, incubation with Staphylococcus nuclease $(2 \times 10^{-2}, 4 \times 10^{-3}$ U). The samples were loaded on the polyacrylamide gel (60 x 20 x 0.06 cm) and electrophoresis proceeded at 1.5-2 kV for 4 hrs. The nucleotides are numbered from the 5' end of the molecule.
Evidently, although the secondary structure of rat 5S RNA is relatively weak, hydrolysates of RNA(C*) are still more informative. On the other hand, since there is a certain degree of complementarity between the two sets, their comparison allows one to eliminate most of the infidelity in the RNA sequencing. In particular, in the given fragment, only one position remains uncertain, A₁₆, which is surrounded with five C's. It is known that if the 3'-neighbour of A is C or a few C's, the rate of hydrolysis of the A-C bond is very low. This fact accounts for the absence of the appropriate band from the autoradiogram (Fig. 2, track U₂). The weak band representing C₂₆ on all of the tracks is due to the cleavage of the initial RNA at that position (Fig. 2, track -E).

When analysing the sequence between the 60th and the 115th nucleotide of rat liver 5S RNA (not shown here), the difference between the modified and non-modified samples is even more remarkable.

Therefore, one may conclude that the combined use of modified and non-modified RNAs for sequence analysis diminishes the vagueness which takes place in some positions along the poly-nucleotide chain due to unequal rates of hydrolysis of phosphodiester bonds with endonuclease and, in certain cases, it eliminates it totally.

However, this is probably not a general rule. When we analysed modified 5S RNA from yeast (Fig. 3) no additional structural information was obtained after the hydrolysis of non-modified RNA (not shown). In the A₁₁-G₁₁₀ fragment, we were unable to establish the type of nucleotide in three positions (G₂₅, A₂₇ and U₃₃) after endonuclease hydrolysis of RNA(C*) with all of the aforementioned nucleases. The lack of information here is mostly due to the low rate of phosphodiester bond cleavage in G(A,U)-C(C*) and, particularly, G(A,U)-CC(C*C*). We believe that the 5'-neighbour of the nucleotide being split influences the rate of cleavage.

In 5S RNA (yeast), we have found a sequence \textsuperscript{10}CAUAUCUACCA \textsuperscript{20} in agreement with the sequence determined by Peattie \textsuperscript{12}, which differed from the sequence obtained by Hindley and Page \textsuperscript{20}. Hence RNAAse \textit{U₂} splits A-U bonds more efficiently than A-C.
bonds, the greater intensity of the $A_{13}$ band indicates that $U_{14}$ is the 3'-neighbour of $U_{13}$.

Besides, one more difference in the sequence of 5S RNA (yeast) was revealed (to be published).

It should be noted that the use of $T_2$ RNAase in modified RNA for $\text{--C}$ hydrolysis gave remarkably better results than Phy I (Figs 2 and 3, tracks $T_2$ and Phy) since $T_2$ RNAase produced a much more complete set of oligonucleotides not terminated with $\text{C}$. The presence of bands in $T_2$ and Phy I hydrolysates, and their absence in $T_1$ and $U_2$ hydrolysates, was interpreted as representing $U$ residues. The reason for that stems from the observation that, if the band representing the purine in $T_1$ and $U_2$ hydrolysates is absent, it never appears in $T_2$ and Phy I hydrolysates.

We have shown earlier that, in the conditions of complete hydrolysis nuclease from Staphylococcus aureus does not split $\text{C*-N}$ bonds. Therefore, this enzyme might be useful for sequence analysis. In fact, as shown in Fig. 3 (track St), hydrolysis with this nuclease helped to define more exactly the positions of some $U$'s. However, in general, we did not receive any sufficient sequence information after the application of Staphylococcus nuclease, at least in the conditions used in this work.

DISCUSSION

According to the previous experience of many investigators, the main source of errors and ambiguities during 5'-$^{32}$P-labell- ed RNA sequencing with endonucleases stems from the tightly folded regions of polynucleotide chains and from a non-equal rate of enzymatic hydrolysis of different phosphodiester bonds.

To eliminate the first of these factors, we suggested here to modify chemically C residues with a methoxyamine-bisulfite mixture to destabilize fully or partially the secondary structure of RNA. The data obtained (Fig. 1) prove unequivocally that, even in the case of a tightly folded tRNA molecule, the modification of C bases leads to a very pronounced increase in the total number of RNA fragments. Comparison of the positions
of nucleotides, which appear in the gel after modification of RNA with the clover-leaf structure of tRNA\textsuperscript{Trp}, shows that this effect is undoubtedly due to the impairment of G-C base pairs.

Therefore, the approach suggested here eliminates completely or almost completely one of the major limitations of the endonuclease technique for RNA sequencing.

Moreover, the same specific modification of C in the RNA molecule blocks the enzymatic cleavage of C*-N bonds by A, T\textsubscript{2} and staphylococcal nucleases as was shown earlier \textsuperscript{15,17,25} and confirmed here (Figs 2 and 3). This change in the enzyme specificity produced by selective chemical modification permits to discriminate between the positions of pyrimidines in a sequence and therefore to increase the total volume of structural information available.

It should be mentioned that when working with RNAs (C*) we never had any problems with the phenomenon of sequence compression which results from anomalous migration in the gel of RNA fragments containing regions of tight secondary structure (compare for example positions 24-28 in Figs 1a and 1b) and is common to all other methods of sequencing RNA on polyacrylamide gels. We suppose, that the absence of sequence compression here is due to melting of RNA (C*) secondary structure.

However, even with chemically modified RNAs, the endonucleases used in this work do not produce a complete set of fragments in the conditions of partial hydrolysis because of the great difference in the rates of hydrolysis between different nucleotide sequences. This is why some uncertainties still remain in a few positions of the sequence.

How to improve further the endonuclease technique based on chemically modified RNA? First of all, by searching nucleases that have a much lesser difference in the rates of phosphodiesters bond cleavage. Possible candidates here are pyrimidino specific RNAase from a Bacillus cereus mutant \textsuperscript{23} and commercially available Staphylococcus nuclease which, at pH 3.5 without Ca\textsuperscript{2+}, splits Pyr-N bonds more uniformly and efficiently than RNAase A does \textsuperscript{3}.

The second way to get information concerning the positions of pyrimidines is to convert C into U chemically by means of
2-3 M bisulfite at pH 5.8. We have analysed tRNA\textsubscript{Trp} after converting all of the exposed C's into U's. It turned out that it is very easy to discriminate between natural and newly made U's after A and T\textsubscript{2} RNAase hydrolyses. RNAase U\textsubscript{2} can be also useful since A residues which have become the 3'-neighbours of U produce much more intensive bands on the film (to be published). On the basis of these observations, we assume that quantitative conversion of all of the C residues into U by chemical means will solve the problem of discrimination of pyrimidines in the RNA sequence as well as purine positions which were not revealed due to 3'-C neighbours (in preparation).

Finally, the results of RNA sequencing by the improved endonuclease approach might be checked by an independent and rapid gel sequencing method based on the use of reverse transcription with terminators.

In summary, the use of chemically modified RNA as a substrate for a set of endonucleases greatly improves the pattern of fragments obtained upon hydrolysis. When combined with hydrolysis of non-modified RNA, some additional structural information may arise. If the sequence is not completely solved, one may use, in addition, an independent approach based on the use of reverse transcription of RNA in the presence of terminators.

ACKNOWLEDGEMENTS

The authors are very thankful to Drs V. Scheinker, G. Keith, M. Saarma, Yu. Kozlov and Yu. Lebedev who provided them with RNA preparations and some enzymes.

REFERENCES

Nucleic Acids Research