Use of specific endonuclease cleavage in RNA sequencing - an enzymic method for distinguishing between cytidine and uridine residues

Ramesh C. Gupta and Kurt Randerath

Department of Pharmacology, Baylor College of Medicine, Texas Medical Center, Houston, TX 77030, USA

Received 13 June 1977

ABSTRACT

The extracellular ribonuclease I of the common slime mold {Physarum polycephalum} (RNase Phy), which has recently been purified to homogeneity, has been used to distinguish between C and U residues in 3'-end-labeled oligoribonucleotides. As shown by Bargetzi and coworkers, this enzyme exhibits strong cleavage preference for U-N over C-N and N-C over N-U bonds. In the present paper, conditions are being detailed, which enable one to deduce the sequences of rather large, pyrimidine-rich, terminally labeled oligonucleotides by partial digestion with RNases U2, A, and Phy, followed by resolution of the cleavage products by size. The techniques described in this and a previous communication provide a direct means for identifying A, G, C, and U residues in end-labeled polyribonucleotides.

INTRODUCTION

As reported previously, nonradioactive RNA fragments may be sequenced by incorporation of label into terminal positions, controlled digestion with specific endonucleases, and resolution of the digestion products by size. The positions of G and A residues were identified by partial digestion with RNase T1 and RNase U2 respectively. Digestion with RNase A served to locate pyrimidine residues but additional manipulations were necessary to distinguish between C and U residues. We have previously described a chromatographic mobility shift analysis technique for this purpose.

In the present communication, we wish to report on an alternative method, which is based on preferential enzymic cleavage at uridine sites. An extracellular ribonuclease from the common slime mold {Physarum polycephalum} was discovered in 1969 by Braun and Behrens, who also demonstrated the resistance of C-N bonds to the action of this enzyme. Later work by Contreras and...
Nucleic Acids Research

Fiers confirmed the resistance of C-N linkages. In addition to this enzyme, Physarum polycephalum has been shown to produce a second extracellular RNase (RNase II) and an acid phosphatase. In the present paper, we shall use the designation RNase Phy for the extracellular RNase I of Physarum polycephalum. This enzyme was recently purified to homogeneity by Bargetzi and coworkers, who also performed detailed studies on its cleavage specificity. This work showed the cleavage rates of dinucleotides to decrease in the order UpN > ApN > GpN > CpN (thus, UpN ≈ CpN). Cleavage rates were found to depend also on the 3'-neighbor of the cleavage site, decreasing in the order NpA ≈ NpC > NpG > NpU (thus, NpC ≈ NpU). Therefore, the most susceptible compounds are UpC, UpA, ApA, and ApC, while CpU and CpC are most resistant. UpC is about 10 times more susceptible than UpU, which is however readily cleaved. On the basis of the specificities of the enzyme, we anticipated that partial digestion of end-labeled oligonucleotides with RNase Phy might be a useful tool for discriminating between C and U residues. In the present communication, the (3H)-oligonucleotide-(3') dialcohols, A-A-U-C-U-C-U-U-A-G'-T, C-A-A-C-C-A-C-C'-T, U-C-A-U-C-A-C-C-C-A-C-C-A'-T, and A-A-U-C-C-U-U-C-C-C-C-C-A-C-A'-T were used as model compounds to test this hypothesis.

MATERIALS

The sources of most materials have been indicated previously. tRNAVal of E. coli B was from Medical Research Laboratories, Irvine, CA. 92707. Yeast transfer RNA was Sigma Type I (Sigma Chem. Co., St. Louis, MO. 63178). RNase Phy, purified from the culture fluid of Physarum polycephalum, was generously donated by Professor J.-P. Bargetzi. One unit of enzymatic activity is defined as the amount of enzyme that solubilizes 35 µg RNA per min at pH 4.5 and 40° as described. A stock solution of the enzyme containing 3.4 U/ml and 40% glycerol was kept at -18°.

METHODS

For the preparation of the (3H)-labeled dialcohol
derivatives, C-A-A-C-A-C-C'-T and A-A-U-C-U-C-U-U-A-G'-T, see ref. 3. U-C-A-U-C-A-C-C-A-C-A'-T and A-A-U-C-C-U-U-C-C-C-C-A-C-A-C-A'-T were prepared from E. coli tRNA\textsubscript{Val} and E. coli tRNA\textsubscript{Tyr} respectively by treatment with NaIO\textsubscript{4} and (\textsuperscript{3}H)NaBH\textsubscript{4}. The reaction mixtures contained, in a volume of 100 μl, 0.05 mM tRNA and 2.5 mM NaIO\textsubscript{4}. Incubation was on ice for 3 hr. This was followed by reduction with (\textsuperscript{3}H)NaBH\textsubscript{4}\textsuperscript{3} (about 40 mM, 40 Ci/mmole). All operations involving labeled NaBH\textsubscript{4} were performed under a well-ventilated hood. Incubation was for 30 min on ice (to destroy excess NaIO\textsubscript{4} by reaction with NaBH\textsubscript{4}) and then for 2 hr at 23°. Excess (\textsuperscript{3}H)NaBH\textsubscript{4} was decomposed by addition of acetic acid (final concentration, 1 M). The solutions were degassed by evacuation for 15 min and finally lyophilized. The 3'-end-labeled tRNA derivatives were digested completely with RNase T\textsubscript{1}\textsuperscript{8}. The 3'-terminal fragments were finally purified on PEI-cellulose thin layers by stepwise development with lithium formate (0 - 2.0 M), 7.5 M urea, pH 3.5\textsuperscript{3}, followed by contact transfer\textsuperscript{9} of the radioactive oligonucleotide spots onto another sheet and stepwise development with Tris-HCl (0 - 1.0 M), 8.5 M urea, pH 8.0\textsuperscript{3}. The radioactive compounds were rendered visible by fluorography\textsuperscript{10}. U-C-A-U-C-A-C-C-A-C-C-A'-T was extracted with 4 M pyridinium formate, pH 4.2, and purified with phosphocellulose\textsuperscript{11}. A-A-U-C-C-U-U-C-C-C-C-A-C-A'-T, which could not be extracted quantitatively with pyridinium formate, was extracted with 4 M LiCl containing 6 M urea (about 20 min at room temperature), purified with phosphocellulose and precipitated in the presence of yeast tRNA carrier (0.15 μg/μl) at 0° by addition of 0.2 vol. of 20% potassium acetate, pH 5.0, and 3 vol. of acetonitrile/ethanol (4 : 1, by vol.)\textsuperscript{12}. After 2 hr at 0°, the precipitate was collected by centrifugation. The pellet was dissolved in water, precipitated with acetonitrile/ethanol as above and the precipitate collected by centrifugation. The pellet was washed with ice-cold ethanol, dried in vacuo, and finally dissolved in water. The final solutions of the 3'-end-labeled oligonucleotide dialcohols contained about 2 x 10\textsuperscript{5} dpm/μl. Recoveries were 40 - 70% of theory.

Partial digestion of (\textsuperscript{3}H)-labeled oligonucleotide-(3') dialcohols with RNase T\textsubscript{1}, RNase U\textsubscript{2}, RNase A, and nuclease S\textsubscript{1}/alka-
line phosphatase was performed as described\textsuperscript{3,11}, except that yeast tRNA was used as the carrier. The nuclease S\textsubscript{1}/phosphatase digest was co-chromatographed with a spleen phosphodiesterase digest (see ref. 3). For controlled digestion with RNase Phy\textsubscript{1}, the reaction mixture contained, in a volume of 20 - 40 \(\mu\)l, 10\(^5\) - 10\(^6\) dpm of oligonucleotide dialcohol, 0.1 \(\mu\)g/\(\mu\)l yeast tRNA as carrier, 20 mM sodium acetate, pH 4.5, 2 mM EDTA, and 0.34 milli-unit/\(\mu\)l RNase Phy\textsubscript{1}. Incubation was performed for 10 min at about 23°. Aliquots (5 - 10 \(\mu\)l) were withdrawn at 1 - 2, 5, and 10 min, applied to the same origin spot of a PEI-cellulose thin layer and immediately dried in a stream of warm air\textsuperscript{3}. The labeled products were separated on PEI-cellulose thin layers by size\textsuperscript{3,9} and located by fluorography\textsuperscript{10}. Film exposure was for 2 - 5 days depending on the radioactivity applied to the chromatogram.

RESULTS

In the following, the positions of the monomer units in 3'-end-labeled oligonucleotides will be numbered in the 3' \(\rightarrow\) 5' direction, as this will simplify presentation of the data.

As shown previously\textsuperscript{3}, partial digestion of end-labeled oligonucleotides with endonucleases breaking the chain at specific sites (e.g., RNase T\textsubscript{1} at G-) enables one to measure the distances of the cleavage sites from the labeled ends, following the resolution of the digestion products by size, thus, in this example, enabling one to determine the positions of G residues in the polynucleotide chain. Similarly, RNases U\textsubscript{2} and A may be used to locate A and pyrimidine residues respectively\textsuperscript{3}. In general, specific cleavage between the \(n^\text{th}\) and \((n+1)^\text{th}\) position of a 3'-end-labeled oligonucleotide enables the identification of the \((n+1)^\text{th}\) position.

Since the previously described two-dimensional chromatographic mobility shift technique\textsuperscript{3} to distinguish between C and U residues is indirect and somewhat laborious, we have screened several nucleases for their ability to discriminate directly between C and U residues. In the following, we shall report on results obtained with one such enzyme, RNase Phy\textsubscript{1}, which we have studied thoroughly, because it afforded the most promising re-
sults in preliminary experiments.

**A-A-U-C-U-C-U-U-A-G'-T.** Sequence analysis of this compound by partial digestion with endonucleases (RNase U₂, RNase A, and nuclease S₁) and mobility shift analysis has been detailed previously³ (see also Fig. 3 of ref. 3). Controlled nuclease S₁/phosphatase digestion of 3'-end-labeled oligonucleotides breaks the chain at every position¹¹, thus providing markers of defined size³, which serve to align the specific RNase cleavage products following their separation by size. Partial digestion with the above mentioned enzymes enables one to read the sequence of this compound as A-A-(Py)₆-A-G'-T³. Fig. 1 illustrates partial cleavage of this compound by RNase Phy¹. (RNase U₂ and nuclease S₁ cleavages have also been presented for comparison.) RNase Phy¹ is seen to cleave only at the A-A, U-C, U-U, and D-A bonds; there is no visible cleavage at the A-U and the 2 C-U bonds. This is in agreement with the known specificities of RNase Phy². Since chain lengths 4 and 6 are missing in this digest, positions 5 and 7 are C. Although the dinucleotide, ApU, was found to be slightly susceptible to RNase Phy², we have observed essentially no cleavage of A-U bonds in our model compounds (see also Figs. 2 and 3). In general, N-U bonds, with the exception of U-U (Fig. 1, spot 3), appear to be extremely resistant to the action of this enzyme under partial conditions.

**C-A-A-C-C-A-C-C'-T.** In agreement with the cleavage specificities of the enzyme, the main products found in partial digests of this compound with RNase Phy¹ were the di-, penta-, and hexanucleotide derivatives (not shown), arising from A-N cleavages. As expected, these products were also present in a partial RNase U₂ digest³. No intermediates resulting from C-N cleavages were formed under the standard incubation conditions. Traces of tri- and heptanucleotide, arising from C-A cleavage, were however formed upon prolonged incubation (>30 min) with RNase Phy¹. The latter 2 compounds constituted the main cleavage products of partial RNase A digestion of this compound³. This was expected, because RNase A exhibits a strong preference for Py-Pu, in particular C-A bonds¹³.

**U-C-A-U-C-A-C-C-A-C-A'-T.** Fig. 2 illustrates the sequence analysis of this compound by controlled digestion with
Fig. 1.

Resolution by size on a PEI-cellulose thin layer of products obtained by controlled RNase U2, RNase Phy1, and nuclease S1 digestion of A-A-U-C-U-C-U-U-A-G'-T. For the cleavage pattern obtained by partial RNase A digestion of this compound, please refer to Fig. 3 of ref. 3. The nuclease S1 digest was dephosphorylated as described11 and co-chromatographed with a spleen phosphodiesterase digest3. Film detection by fluorography. The numbers indicate both the chain lengths of the compounds and the positions of individual residues within the sequence. The vertical column of letters refers to individual residues identified on the basis of the cleavage patterns.
Resolution by size on a PEI-cellulose thin layer of products obtained by controlled RNase U2, RNase Phy1, RNase A, and nuclease S1 digestion of U-C-A-U-C-A-C-C-A-C-A'-T. For conditions of nuclease S1 digestion and other details, please consult legend of Fig. 1. Film detection by fluorography.
RNases U₂, Phy₁, and A. A partial nuclease S₁/phosphatase digest is also shown. Partial RNase U₂ digestion indicated positions 4, 8, and 11 to be A residues (Fig. 2). Partial RNase A digestion showed positions 3, 5, 6, 7, 9, 10, 12, and 13 to be pyrimidines. The dinucleotide spot was weak on the original film and therefore does not show in Fig. 2. On the basis of these data, the sequence can be written as (Py)₂-A-(Py)₂-A-(Py)₃-A-Py-C-A'-T. (Positions 1 and 2 were determined as described previously³.) The remaining individual pyrimidines were identified by partial RNase Phy₁ digestion as follows. As shown in Fig. 2, intermediates of chain lengths 3, 7, 8, 9, 11, and 12 were present in the RNase Phy₁ digest. The fact that cleavage products of chain lengths 2, 4, 5, and 6 were present in the RNase A but not in the RNase Phy₁ digest indicated positions 3, 5, 6, and 7 to be C, in agreement with the resistance of C-N bonds to RNase Phy₁ under partial digestion conditions.

As shown by a comparison with the RNase A digest, chain lengths 9 and 12 must have arisen from Py-Py breaks. Since C-Py is resistant to RNase Phy₁, the presence of chain lengths 9 and 12 indicated positions 10 and 13 to be U. The absence of chain length 10 in the RNase Phy₁ digest (but its presence in the RNase U₂ digest) provided further evidence for U in position 10, as A-U bonds are virtually resistant to the action of RNase Phy₁, while A-C is readily cleaved (note, for example, chain lengths 3 and 7 in the RNase Phy₁ digest). Thus, the sequence of the compound is U₁₃-Py₁₂-A₁₁-U₁₀-Py₉-A₈-C₇-C₆-C₅-A₄-C₃-C₂-A₁'-T.

As to position 12, note the weakness of spot 11 relative to spot 12. In agreement with the distinct preference of the enzyme for U-C over C-A bonds², the relative intensities of spots 12 and 11 indicated that spot 12 was due to U-C cleavage and spot 11 to C-A cleavage. (In the sequence -U-U-A-, a U-U break would result in a much weaker spot than a U-A break.) The 5'-proximal sequence therefore must be U-C-A-U-. The relative intensities of spots 9 and 8 indicated similarly that these spots resulted from U-C and C-A breaks respectively. These results extended the 5'-proximal sequence to U-C-A-U-C-A-C....., thus establishing the sequence of the entire compound.
Resolution by size on a PEI-cellulose thin layer of products obtained by controlled RNase U₂, RNase Phyl, RNase A, and nuclease S₁ digestion of A-A-U-C-C-U-U-C-C-C-C-A-C-A-C-A'-T. For conditions of nuclease S₁ digestion and other details, please consult legend of Fig. 1. Film detection by fluorography. Since the film was overexposed to visualize the weak spots, spots 17 - 19 appear not well resolved, except in the nuclease S₁ digest. Separation of these compounds could be demonstrated by reducing the film exposure time from about 3 days to 1 day.
A-A-U-C-C-U-U-C-C-C-C-C-A-C-C-A-C-C-A'-T. Nuclease S₁/phosphatase and RNase U₂ digestions enabled us to write down the sequence of this compound as A-A-(Py)₁₀-A-(Py)₂-A-Py-C-A'-T. To identify the individual pyrimidines the compound was partially digested with RNase Phy₁. Separation of this digest, alongside partial RNase U₂, RNase A, and nuclease S₁/phosphatase digests, is illustrated in Fig. 3. Intermediates of chain lengths 3, 6, 12, 13, 16, and 18 were present in the RNase Phy₁ digest. Products originating from A-N breaks were identified by comparison with the RNase U₂ digest (chain lengths 3, 6, 17, and 18). Thus, chain lengths 12, 13, and 16 in the RNase Phy₁ digest must be due to U-Py breaks, since C-Py bonds are resistant to the action of this enzyme. Chain lengths 2, 4, 5, 7, 8, 9, 10, 11, 14, and 15 were present in the RNase A digest but absent in the RNase Phy₁ digest, indicating positions 3, 5, 6, 8 – 12, 15, and 16 to be occupied by C. These data established the complete sequence of the compound as A-A-U-C-C-U-U-C-C-C-C-C-A-C-C-A'-T. This example demonstrates that partial digestion with RNase Phy₁ is a powerful tool to analyze polypyrimidine sequences, as U-Py is susceptible but C-Py resistant to cleavage.

DISCUSSION

A survey of the literature on nucleases¹⁴ (see also additional citations in ref. 3) suggested to us the existence of enzymes capable of distinguishing C-N from U-N bonds in end-labeled oligonucleotides. As shown in the present work, the strong preference of RNase Phy₁ for U-N over C-N and N-C over N-U bonds² makes this enzyme a valuable tool to distinguish C from U residues in terminally labeled oligonucleotides. Using partial digestion of 3'-end-labeled oligonucleotides, in combination with resolution by size, we have been able to determine directly the location of C and U residues relative to the labeled terminal positions. This supplements the G (ref. 3), A + G (present work and ref. 3), and C + U (present work and ref. 3) cleavage patterns produced by RNase T₁, RNase U₂, and RNase A digestion respectively, thus enabling one to read directly entire sequences of polyribonucleotides containing A, G, C, and U. Noteworthy features of the enzymic RNA sequencing method include simplicity (just add
enzyme to polyribonucleotide and buffer, incubate, and separate products by size) and gentle conditions (pH range of 4.5 - 8 and incubation for a short time).

Since the C-A bond is the only C-N bond measurably cleaved by RNase Phy₁ under partial digestion conditions, its distinction from U-A deserves some comments. To decide whether a -Py-A- sequence is present in -Py-Py-A- or -Pu-Py-A-, one analyzes first the partial RNase T₁³ (if G is present), RNase U₂, and RNase A digests that have been separated alongside the RNase Phy₁ digest. This serves to display the positions of G, A, and pyrimidine residues respectively. Let us now consider the sequences -C-C-A- and -C-U-A-. While C-A appears to be resistant to cleavage by the enzyme when present in this sequence (Figs. 2 and 3), U-A is rapidly cleaved. Thus there is no ambiguity in discriminating between -C-C-A- and -C-U-A-. The sequence -U-C-A- is readily distinguished from -U-U-A-, since U-C breaks give much stronger spots than adjacent C-A breaks (Fig. 2), while the opposite is true for U-U versus U-A (U-U weak; U-A very strong). Similarly, -Pu-C-A- is easily distinguished from -Pu-U-A- because Pu-C is highly susceptible, while Pu-U is resistant to the action of RNase Phy₁. Thus, the characterization of -Py-A- sequences presents no problems, although one has to be familiar with the cleavage specificities of RNase Phy₁ so as to interpret the digestion patterns correctly.

Inspection of Fig. 2 reveals that the C-A bond next to the 5'-terminal U-C is cleaved at a greater rate than the C-A in the internal -U-C-A- sequence. This is probably a secondary effect arising from the preference of the enzyme for 5'-terminal U-C, which in turn affects the cleavage rate of the adjacent C-A bond. It is clear therefore that the intensity of the spot due to the 5'-proximal C-A break can be compared only with the intensity of the adjacent spot originating from cleavage of the 5'-terminal U-C; it cannot be compared with the intensity of a spot due to an internal U-C cleavage.

Many results reported in this paper could be predicted on the basis of the work of Bargetzi and coworkers on the digestion of dinucleotides with RNase Phy₁². The effect exerted by the 5'-neighbor of C on the rate of cleavage of C-A bonds, which
could not be predicted on the basis of this work, appears to be a special case, since, for example, cleavage of A-U in the sequence A-A-U- is not facilitated by the adjacent highly susceptible A-A bond (Figs. 1 and 3). The preference of the enzyme for susceptible 5'- and 3'-proximal bonds, as shown in Fig. 1, could also not be predicted from the dinucleotide data. It may be related to the absence of terminal phosphomonoester groups in the compounds analyzed in the present work.

When applying RNase Phy₁ in sequence studies, it will be helpful to keep these important properties of the enzyme in mind:

1. C-N is resistant (C-A may be slightly digested under certain conditions, see above);
2. U-N is always cleaved;
3. N-C is cleaved (exception: C-C, see (1));
4. N-U is resistant (exception: U-U, see (2)).

While the cleavage display produced by RNase A is by itself not sufficient to distinguish C from U residues, the known cleavage preferences of RNase A¹³ (C-A > U-A ≫ C-C ≫ U-C > C-U > U-U), are clearly detectable in the examples presented in this and a previous³ communication. For example, spots originating from U-U cleavages by RNase A (Fig. 3, #13; also spot 3 of Fig. 3 in ref. 3) are the weakest, spots due to C-A breaks the strongest. Spots attributable to C-C breaks are of intermediate strength in the RNase A digests and completely missing in RNase Phy₁ digests.

It is interesting to note that in the sequence -U-C-C-C-C-C-A- the cleavage rates by RNase A of the C-C bonds gradually increase as one goes from the rather resistant U-C bond towards the very susceptible C-A bond (Fig. 3). A similar effect may be noted in Fig. 2 (sequence -A-C-C-C-A-). The presence of a highly susceptible bond such as C-A thus affects the cleavage rates at positions several nucleotides away from the C-A bond. In view of the completely different cleavage preferences of RNase A and RNase Phy₁, spot patterns generated by partial digestion with either enzyme supplement each other when compared side by side.

In conclusion, the results presented in this and a previous³ communication indicate that partial digestion of terminally labeled polyribonucleotides with suitable endonucleases (RNases T₁, U₂, A, and Phy₁) and separation of the digestion products by
size provides a simple and direct method for sequencing RNA. As previously pointed out\textsuperscript{3}, the principles of the method are not restricted to 3\(^{-}\)-(\textsuperscript{3}H)-labeled polynucleotides: other isotopes may be preferable in certain applications and the label may be attached to the 5\(^{-}\)-terminus. The resolution of the digestion products must be by size; for chain lengths of up to 20 nucleotides, anion-exchange thin-layer chromatography on PEI-cellulose in the presence of urea provides a rapid, easy and reproducible method requiring little special equipment. Polyacrylamide gel electrophoresis is better suited than chromatography to resolve larger RNA fragments by size (unpublished experiments). Work is currently underway in our laboratory on the application of RNase Phy\textsubscript{1} and other ribonucleases to read directly the terminal sequences of end-labeled RNAs.

ACKNOWLEDGMENTS

This investigation was supported by USPHS Grants CA-13591-(06) and CA-16840-(03). K.R. is a recipient of a Faculty Research Award (PRA-108) of the American Cancer Society.

We wish to express our gratitude to Dr. J.-P. Bargetzi for communicating to us results obtained in his laboratory prior to publication and for generously supplying a sample of purified RNase Phy\textsubscript{1}. We also thank Dr. S.H. Chang for a gift of 2 oligonucleotides used in this study and Dr. E. Randerath for many helpful discussions.

This is part XXIV of a series entitled "Analysis of Nucleic Acid Derivatives at the Subnanomole Level".

REFERENCES

ABBREVIATIONS