Use of specific endonuclease cleavage in RNA sequencing

Ramesh C. Gupta and Kurt Randerath

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

Received 18 March 1977

ABSTRACT

Nonradioactive RNA fragments may be sequenced by incorporation of \(^{3}H\)-label into 3'-terminal positions, controlled digestion with specific ribonucleases, and separation according to size of the digestion products on polyethyleneimine- (PEI-) cellulose thin layers. This combination of techniques allows one to measure accurately distances of specific cleavage sites from the labeled terminal positions. The cleavage specificities of RNases T\(_1\), U\(_2\), and A are utilized to identify the positions of G, A, and pyrimidine residues respectively. C and U may be distinguished by mobility differences on PEI-cellulose thin layers at pH 2.6. The procedure is simple, rapid, and highly sensitive; as little as 0.5 - 1 \(\mu\)g of a RNA of the size of tRNA will be needed to sequence all fragments in a complete RNase digest.

INTRODUCTION

The subject of this publication is a newly developed simple and highly sensitive sequencing procedure for oligonucleotides prepared by digestion of nonradioactive RNA. In this procedure, use is made of the known cleavage specificities of ribonucleases T\(_1\), U\(_2\), and A.

MATERIALS

Yeast tRNA\(_{\text{Phe}}\) and \(E.\ coli\) MRE600 tRNA\(_{\text{Tyr}}\) were from Boehringer Mannheim Corp. Yeast RNA was Sigma Type XI. RNases T\(_1\) (Sankyo), U\(_2\) (Sankyo), A (Sigma), and nuclease P\(_1\) (Yamasa Shoyu Co., Tokyo) were used without further purification. Dinucleoside monophosphates were from Sigma Chemical Co. A 2-mg sample of spleen phosphodiesterase (Boehringer Mannheim Corp.) was dialyzed at 4\(^\circ\) for a total of 24 hr against 3 changes (3 l each) of deionized water. Dialysis tubing was
pretreated as described\textsuperscript{1}. Sources of alkaline phosphatase, polynucleotide kinase, nuclease S\textsubscript{1}, (\textsuperscript{3}H)NaBH\textsubscript{4}, (\textgreek{y}-\textsuperscript{32}P)ATP, and nonradioactive chemicals were as described\textsuperscript{2}. (\textsuperscript{3}H)NaBH\textsubscript{4} (30 - 60 Ci/m mole) was dissolved in freshly prepared 0.1 N KOH at a concentration of 0.1 - 0.2 M, divided into 5-, 10- or 20-\mu l portions, lyophilized, and stored at room temperature under a hood in an evacuated implosion-proof plastic desiccator (Fisher Scientific \# 8-594-15B) over anhydrous CaSO\textsubscript{4}. Individual portions were dissolved in freshly prepared 0.01 N KOH immediately before use. The high specific activity material remained active at least 5 months when stored in this way.

PEI-cellulose thin layers were prepared as described\textsuperscript{3,4}. Commercial PEI-cellulose sheets (PEI CEL-22) were obtained from Brinkmann Instruments. PEI-cellulose layers were treated before use according to procedure 3 of ref. 4 except for sheets used for mobility shift analysis, which were pre-equilibrated as follows. A Whatman 1 wick (10 - 15 cm long), which had been stapled to the top of a commercial sheet (20 x 20 cm) was allowed to extend to the outside of the jar so as to allow the solvent to evaporate after the front had reached the upper portion of the wick. Pre-development was for about 15 hr with 0.05 M ammonium formate, pH 2.6. The layers were allowed to dry at room temperature for several hours before use. If not used the same day they were kept at -18\textdegree.

Button-type Alnico magnets (\# 372A, 1/2 x 3/8 in.) were from General Hardware Manufacturing Co., New York, N.Y.).

\textbf{METHODS}

Isolation of oligonucleotides. Model compounds were isolated from a RNase A digest of yeast tRNA\textsubscript{Phe} \textsuperscript{5} and from RNase T\textsubscript{1} digests of \textit{E. coli} tRNA\textsubscript{Tyr} \textsuperscript{6-8} and yeast tRNA\textsubscript{UUG} \textsuperscript{9}. Solutions of oligonucleotides from the latter RNA (C-A-A-C-C-A- C-C and A-A-U-C-U-C-U-U-A-Gp) were fractions from a DEAE-cellulose column kindly donated by Dr. S. H. Chang. The column fractions were desalted by passage through a Bio-Gel P-2 column\textsuperscript{1}. For preparation of G-G-G-A-G-A-G-C, yeast tRNA\textsubscript{Phe} was subjected to combined RNase A/alkaline phosphatase digestion\textsuperscript{11}. The
digest (about 3 A260 units) was resolved by 2-D TLC on PEI-cellulose; it was applied in 10-μl portions at 2.5 cm each from the left-hand and the lower edge of a "home-made" sheet (20 x 25 cm)\(^1\). Prior to development, the origin was dried with cool air. The sheet was then soaked 5 min in 300 ml methanol/conc. ammonia (1000 : 1, by vol.) and dried with cool air. A Whatman 1 wick was attached to the sheet at a distance of 22 cm from the origin. Development was in stepwise fashion, without intermediate drying. In the first dimension, the chromatogram was developed with water to 2 cm above the origin; 0.05 M Tris-HCl, pH 8.0 to 6 cm; 0.15 M Tris-HCl, 8.5 M urea to 11 cm; 0.3 M Tris-HCl, 8.5 M urea to 15 cm; 0.5 M Tris-HCl, 8.5 M urea to 20 cm; and 0.75 M Tris-HCl, 8.5 M urea to about 2 cm on the wick. The Tris/urea solvents were prepared by diluting 1.0 M Tris-HCl, 8.5 M urea, pH 8.0 stock solution with 8.5 M urea solution. The wet sheet was soaked after development twice in 300 ml methanol 10 min each and dried with cool air. For the second dimension, a Whatman 1 wick was attached at a distance of about 16 cm from the origin. Subsequent development was with water to the origin; 7.5 M urea to 2 cm; 0.2 M Li formate, 7.5 M urea to 4 cm; 0.4 M Li formate, 7.5 M urea to 6 cm; 0.7 M Li formate, 7.5 M urea to 9 cm; 1.2 M Li formate, 7.5 M urea to 14 cm; and 1.8 M Li formate, 7.5 M urea to about 2 cm on the wick. The Li formate/urea solvents were prepared by diluting 4.0 M Li formate, 7.5 M urea, pH 3.5 stock solution with 7.5 M urea solution. The wet sheet was soaked after development twice in 200 ml methanol 10 min each and dried. Oligonucleotides were located in UV light, extracted with 4 M pyridinium formate, pH 4.2, and purified by passage through phosphocellulose\(^2\). Nucleotide sequences of the compounds of interest were analyzed by previously described\(^1,2,12\) procedures.

The nonadecamer, A-A-U-C-C-U-U-C-C-C-C-C-A-C-C-A-C-C-A, was prepared from a RNase T\(_1\) digest of E. coli tRNA\(^{Tyr}\) by a slight modification of this procedure.

**Preparation of (\(^3\)H)-labeled oligonucleotide-(3') dialcohols.** Oligonucleotides were converted to (\(^3\)H)-labeled di-alcohol derivatives\(^1,2,13\) by a procedure consisting of the following steps: (i) removal of 3'-terminal phosphate (if
required); (ii) oxidation of 3'-terminal ribose with NaIO₄; (iii) reduction of dialdehydes obtained in step (ii) with (³H)NaBH₄; and (iv) removal of excess (³H)NaBH₄ by adding acetic acid.

A standard reaction mixture (20 µl) contained about 0.05 mM oligonucleotide, 20 mM bicine, pH 8.0, and 0.1 µg/µl alkaline phosphatase. Incubation was at 38°C for 30 - 60 min. This was followed by oxidation with NaIO₄ (final concentration, 1.5 mM) and reduction with (³H)NaBH₄ (about 20 mM, 30 - 60 Ci/mmole), both reactions being performed, under the usual precautions¹⁴, 4 - 5 hr each at 23°C. Acetic acid was then added to a final concentration of about 1 M and the solution was evaporated in a stream of air. The residue was taken up in 20 µl water. The labeled oligonucleotide dialcohols were finally purified by chromatography on a "home-made" PEI-cellulose thin-layer sheet in system 1 of ref. 1, followed by fluorographic detection¹⁵, extraction with pyridinium formate, and purification of the extract with phosphocellulose².

Partial digestion of (³H)-labeled oligonucleotide-(3') dialcohols with endonucleases. This was done as follows. The standard reaction mixture for RNase T₁ or RNase A digestion contained, in 20 - 50 µl, 1 - 10 pmole (5 x 10⁴ - 5 x 10⁵ dpm) oligonucleotide dialcohol, 0.1 µg/µl yeast RNA carrier, 20 mM bicine, pH 7.8, and RNase T₁ (0.01 ng/µl)¹¹,¹⁶ or RNase A (0.1 - 0.2 ng/µl)¹⁶. For controlled digestion with RNase U₂¹⁷, the buffer was 20 mM sodium acetate, pH 4.5 containing 2 mM EDTA and the enzyme concentration was 0.1 - 0.2 milliunit/µl¹⁶. All incubations were performed for 2-40 min at about 23°C. Aliquots (5 - 10 µl) were withdrawn at suitable intervals, applied to a PEI-cellulose thin layer, and immediately dried in a stream of warm air.

For pooling of aliquots, appropriate volumes were withdrawn from the incubation mixtures at suitable intervals and applied to the same point at the origin of a PEI-cellulose layer. Each sample was dried with warm air immediately after spotting.

Controlled digestion of oligonucleotide dialcohols with nuclease S₁ and alkaline phosphatase has been detailed pre-
viously. To produce intermediates that are sometimes obtained in low yield upon partial nuclease $S_1$ digestion, particularly those having a chain length of $n-1$, it was found advantageous to digest a portion of the parent compound with spleen phosphodiesterase as follows. About $5 \times 10^4$ dpm of oligonucleotide dialcohol was incubated, in $10 \mu l$, with $20 \text{mM}$ bicine, pH 7.8, $2 \text{mM}$ EDTA, and $0.05 \mu g/\mu l$ enzyme. After incubation for 10 - 30 min at 38°C, the digest was co-chromatographed with the nuclease $S_1$/phosphatase digest, applied to the layer as described above. In the present work, this was routinely done.

Separation of oligonucleotide dialcohol intermediates.

After application of the samples, the PEI-cellulose sheet was soaked in 300 ml methanol/conc. ammonia (1000 : 1, by vol.) 5 min and dried with cool air. The labeled products were resolved according to chain length (up to $n = 11 - 13$) by stepwise development in Tris-HCl/8.5 M urea, pH 8.0, as described (system 1 of ref. 1). ($^3H$)-labeled compounds were located by fluorography after removal of electrolytes and urea by methanol treatment. Film exposure was usually for 2 - 4 days depending on the radioactivity applied to the chromatogram.

To separate the larger fragments present in partial endonuclease digests of A-A-U-C-C-U-U-C-C-C-C-C-A-C-C-A'-T, system 1 of ref. 1 was slightly modified. A 30 cm long PEI-cellulose layer was developed with water to the origin; 0.06 M Tris-HCl, pH 8.0 to 6 cm; 0.15 M Tris-HCl, 8.5 M urea to 7.5 cm; 0.4 M Tris-HCl, 8.5 M urea to 9 cm; 0.6 M Tris-HCl, 8.5 M urea to 11 cm; 0.8 M Tris-HCl, 8.5 M urea to 16 cm; and 1.0 M Tris-HCl, 8.5 M urea to 2 cm on a Whatman 1 wick attached to the top of the sheet at a distance of 27 cm from the origin. The Tris/urea solvents were prepared from stock as described above.

**Discrimination between C and U residues.** This was done by transferring the products of controlled endonuclease digestion to a second PEI-cellulose layer and analyzing the characteristic angular mobility shifts between successive oligonucleotide spots as follows. A PEI-cellulose strip (the donor strip) containing 3'-end-labeled partial digestion products, which had been separated according to chain length,
was placed on an area about 2.5 cm from, and parallel to, one edge of a PEI-cellulose sheet (the acceptor sheet), which had been pre-equilibrated with ammonium formate as described above. Donor strip and acceptor sheet were held together firmly between two strong magnetic bars, which had been prepared by aligning a series of button-type magnets on the sticky side of a piece of 1-inch wide surgical tape so that opposite poles of adjacent magnets faced each other. Excess tape was bent upwards so as to adhere to the sides of the row of magnets, which was finally enveloped in a strip of Parafilm. The poles of each magnetic bar were appropriately color-coded for correct alignment of the two bars, ensuring optimal contact between the layers of the donor strip and the acceptor sheet during the subsequent chromatographic development. The magnetic bars could be re-used many times before a replacement of the Parafilm wrapping became necessary.

To transfer the radioactive nucleotide derivatives to the acceptor sheet, and to resolve them, development was with water to the origin; 0.2 M ammonium formate, pH 2.6 to 2 cm; 1.0 M ammonium formate, pH 2.6 to 5 cm; 2.5 M ammonium formate, pH 2.6 to 10 cm; and 4.0 M ammonium formate, pH 2.6 to 2–3 cm on a Whatman 1 wick stapled to the top of the sheet at about 17 cm from the origin. Each ammonium formate solution was prepared separately by titrating formic acid with ammonia; the molarities represent the total formic acid concentrations. After development, the acceptor sheet was dried thoroughly and soaked once in 300 ml methanol. Radioactive compounds were located by fluorography.

Procedures for the identification of positions 1 and 2. The methods described thus far enable one to deduce the sequence of 3'-end-labeled oligonucleotides containing no modified nucleosides, with the exception of positions 1 and 2. If the phosphodiester bond between these two positions is susceptible to cleavage by RNase T₁ or RNase A, treatment with either of these enzymes will liberate position 1 as a nucleoside tri-alcohol. This was found to require more vigorous conditions than those used for the partial digestions (see above). For this purpose, 2–4 μl of the partial digestes were simply
further incubated for 2 - 4 hr at 38° after adding more of the respective RNase so as to obtain an enzyme : substrate ratio of about 1 : 10, by weight. The samples were then mixed with unlabeled nucleoside trialcohols as internal markers and analyzed on silica gel or cellulose thin layers as described previously. The procedure just described is not applicable if position 2 is an A as RNase U2 was found not to cleave at 3'-terminal \( ...pApN' \). For such compounds, base analysis in combination with the results of partial endonuclease digestion, frequently sufficed to identify position 2. Since the smallest nucleotide derivatives present in partial RNase (A, T\(_1\), and U\(_2\)) digests are dinucleotide dialcohols of the structure NpN'-T, oligonucleotide dialcohols having A in position 2 were also characterized by co-chromatography of the radioactive dinucleotide dialcohols with nonradioactive ApN internal markers. For this purpose, 1 - 2 µl of aqueous solutions of the latter compounds (about 10 mM), prepared as described below, were directly applied to a particular ApN'-T spot to be analyzed. The spot was then cut from the chromatogram and transferred, using magnets as described, to a commercial PEI-cellulose acceptor sheet. Development was with the solvents given in Table 1 (see RESULTS). The dinucleotide spots were rendered visible under UV light, cut out, and counted.

Alternatively, position 2 may be identified by a procedure for 5'-end labeling of the dinucleotide derivative with \(^{32}P\) by treatment with \((\gamma-^{32}P)\)ATP and polynucleotide kinase. This technique is applicable also if position 2 is a base-modified or ribose-modified nucleoside.

In addition to exhaustive digestion with RNase T\(_1\) or A (see above), various methods were used to identify position 1, the \((^3H)\)-labeled 3'-terminus. In situ digestion of 3'-end-labeled compounds with RNase T\(_2\) followed by co-chromatography with internal markers provides a simple, rapid, and sensitive assay for position 1. Alternatively, a scaled-down version of the method for base analysis of \((^3H)\)-labeled oligonucleotide derivatives was modified in that an aliquot containing about \(2 \times 10^4\) dpm was removed from the enzymic digestion mixture and, without further \((^3H)\)-labeling, directly co-chromato-
graphed with 2 - 4 nmole of unlabeled trialcohols of the four major nucleosides (as internal markers) on a silica gel or cellulose thin layer as described previously.

If position 2 is a ribose-methylated nucleoside, position 1 may be released by nuclease P<sub>1</sub>/alkaline phosphatase digestion of the labeled dialcohol as follows. About 10<sup>5</sup> dpm of the compound was incubated, in 5 - 10 μl of 20 mM sodium acetate, pH 5.0, with 0.3 μg/μl nuclease P<sub>1</sub> for 10 - 15 hr at 38°. An equal volume of 0.2 M bicine, pH 8.5, containing 0.2 μg/μl alkaline phosphatase was then added and the incubation continued for 30 min at 38°. The labeled trialcohol released by this treatment was identified by partition TLC as above. For identification of position 2 in such a case, see ref. 19.

Preparation of NpN' (specifically ApN') chromatographic markers. Dinucleoside monophosphate (2 mM) was oxidized with NaIO<sub>4</sub> (6 mM) for 2 hr at 23°. After 0.1 M KB<sub>4</sub> had been added to about 25 mM, the incubation was continued for another 2 hr at 23°. Acetic acid was added (final concentration, about 1 M) and the solution was dried in a stream of air. The residue was dissolved in water to give a concentration of about 10 mM.

RESULTS

Deduction of the sequence of G-G-G-A-G-A-G-C. The octanucleotide derivative, G-G-G-A-G-A-G-C', will be used as one of several model compounds to illustrate the basic features of the sequencing procedure described in the present communication. The parent oligonucleotide, G-G-G-A-G-A-G-C, which forms part of the D-arm of yeast tRNA<sub>Phe</sub>, was isolated by digestion of this RNA with RNase A and alkaline phosphatase, followed by 2-D TLC on PEI-cellulose. The (3H)-labeled dialcohol derivative of this compound, prepared as described under METHODS, was partially digested with nuclease S<sub>1</sub>/phosphatase, RNase T<sub>1</sub>, and RNase U<sub>2</sub>. TLC analysis of these digests (Fig. 1) showed the presence of four and six radioactive products respectively in the RNase T<sub>1</sub> and RNase U<sub>2</sub> digests, in addition to starting material. The radioactive products of nuclease S<sub>1</sub>/phosphatase digestion were the same as those ob-
**Fig. 1.**

Analysis by PEI-cellulose TLC of time course of controlled RNase T₁ and U₂ digestions of G-G-G-A-G-A-G-C'-T. Film detection by fluorography.

<table>
<thead>
<tr>
<th>No Enzyme</th>
<th>RNase T₁ (G)</th>
<th>RNase U₂ (A + G)</th>
<th>n</th>
<th>RNase T₁ (G)</th>
<th>(n+1) th position</th>
<th>RNase U₂ (A + G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(C')</td>
<td>(G)</td>
<td>2</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GGGAGAGC

0.5 2 5 10 | 0.5 2 5 10 Minutes
tained with RNase U2 (not shown). Tritium base composition analysis\textsuperscript{2,14} indicated the parent compound to contain 1 C, 5G, and 2 A residues. Thus, base composition analysis, as well as the fact that seven spots were visible on the fluorogram of the partial nuclease S\textsubscript{1}/phosphatase digest, provided evidence that the parent compound is an octanucleotide. As shown previously\textsuperscript{2}, partial nuclease S\textsubscript{1}/phosphatase digestion of an oligonucleotide dialcohol of chain length n yields n-2 partial digestion products that are rendered visible by fluorography after PEI-cellulose TLC of the digest, in addition to undigested starting material, i.e. a total of n-1 spots. The 3'-terminal trialcohols are lost during the methanol treatments and therefore do not show on the fluorograms.

Dinucleotide dialcohols are also the shortest digestion products present in partial RNase T\textsubscript{1}, U\textsubscript{2}, and A digests since the first (3'-terminal) phosphodiester bond is resistant to the conditions of partial digestion.

In analogy to DEAE-cellulose column chromatography\textsuperscript{22}, oligonucleotides are resolved according to chain length on PEI-cellulose thin layers in weakly alkaline solvents containing 8 - 8.5 M urea\textsuperscript{1}. As shown in Fig. 1, the chain lengths (n) of the products of RNase T\textsubscript{1} digestion of G-G-A-G-A-G-C'-T were 3, 5, 6, and 7 respectively; di- and tetranucleotide derivatives were absent from this digest. This indicated that positions 4, 6, 7, and 8 must be occupied by G while positions 3 and 5 cannot be G. Digestion of the octanucleotide dialcohol with RNase T\textsubscript{1} under more vigorous conditions (see METHODS) yielded labeled cytidine trialcohol, establishing position 2 as G and, at the same time, position 1 as C (see base composition analysis above). Six out of the total of eight positions have thereby been accounted for. This, in combination with the results of base analysis, established the entire sequence of the compound since the two A residues can be placed only in positions 3 and 5. Direct evidence for this was provided independently by partial digestion with RNase U\textsubscript{2}, which afforded high yields of the di- and tetranucleotide derivatives (Fig. 1). RNase U\textsubscript{2} is purine-specific preferring A over G residues\textsuperscript{17}. The results, which have been summarized schematically in the right-hand panel.
of Fig. 1, show that specific cleavage between the \( n^{th} \) and \( (n+1)^{th} \) position of 3'-end-labeled oligonucleotides identifies the \( (n+1)^{th} \) position.

Other examples. For determining the sequence of C-A-A-C-C-A-C-C'-T, we have used controlled digestion with RNase A and RNase U\(_2\). (Partial digestion of this compound with nuclease S\(_1\)/phosphatase has been discussed previously\(^2\)). Base analysis had shown the presence of 5 C and 3 A residues, with C' being 3'-terminal. Digestion with RNase A (see METHODS) led to the release of labeled cytidine trialcohol indicating both positions 1 and 2 of the parent compound to be C since RNase A would not release the 3'-terminal C' if position 2 were a purine. RNase A caused the formation of tri-, tetra-, and heptanucleotide dialcohols (Fig. 2) providing evidence that positions 4, 5, and 8 are occupied by C. These results, in combination with base analysis, which indicated the presence of 3 A residues, already allows one to write down the sequence of the octanucleotide derivative. Direct evidence that the 3 A residues occupy positions 3, 6, and 7 again came from controlled RNase U\(_2\) digestion of the parent dialcohol, which yielded derivatives of chain lengths 2, 5, and 6 respectively (Fig. 2).

Sequence analysis of A-A-U-C-U-C-U-U-A-Gp, i.e. a RNase T\(_1\) digestion product containing both C and U residues, is illustrated in Fig. 3. Base analysis of the \(^3\)H-labeled dialcohol derivative of this compound showed the presence of 3 A, 4 U, and 2 C residues, in addition to the 3'-terminal G'. Fig. 3A shows a separation according to chain length of partial endonuclease digestion products while Fig. 3B gives an example for the mobility shift method to distinguish between C and U residues.

The \(^3\)H-labeled dialcohol of the parent oligonucleotide yielded eight oligonucleotide dialcohol products, in addition to starting material, upon controlled digestion with nuclease S\(_1\)/phosphatase, indicating a chain length of 10 (Fig. 3A). RNase U\(_2\) digestion gave only 2 compounds having chain lengths of 8 and 9 (Fig. 3A). Positions 9 and 10 therefore are occupied by A. RNase A digestion gave compounds of chain lengths 2, 3, 4, 5, 6, and 7 showing positions 3 through 8 to be pyrimidines. From these data, the sequence, A-A-(Py)_6-X-G'-T can be deduced. Since
Fig. 2.
Analysis by PEI-cellulose TLC of time course of controlled RNase A and U2 digestions of C-A-A-C-C-A-C'-T. Film detection by fluorography.

Base analysis had indicated the presence of 3 A residues; the data thus far obtained allow one to conclude that the remaining A residue occupies position 2. Additional evidence for this assignment was provided by co-chromatography of the labeled dinucleotide derivative with authentic ApG' as an internal marker (see METHODS). Radioactivity was found to coincide with the position of ApG' after chromatography in systems 1 and 2 (Table 1).

Inspection of Fig. 3A indicates that partial digestion of
Sequence analysis of A-A-U-C-U-C-U-U-A-G'-T by partial endonuclease digestion. Panel A, analysis of pooled time point aliquots by PEI-cellulose TLC in a Tris/urea system at pH 8.0. The nuclease S\textsubscript{1} digest was dephosphorylated as described\textsuperscript{2} and co-chromatographed with a spleen phosphodiesterase digest (see METHODS). Panel B, mobility shift analysis in an ammonium formate system at pH 2.6. Film detection by fluorography.

A-A-U-C-U-C-U-U-A-G'-T with nuclease S\textsubscript{1}/phosphatase affords a more uniform distribution of products than are obtained by partial digestion of this compound with RNase A. In the latter digest,
Table 1

Relative $R_F$ values ($R_{ApA} = 1.0$) of dinucleotide dialcohols of the structure $ApN'$ on PEI-cellulose thin layers

<table>
<thead>
<tr>
<th>Compound</th>
<th>System 1$^a$</th>
<th>System 2$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ApA$</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$ApA'$</td>
<td>1.20</td>
<td>1.27</td>
</tr>
<tr>
<td>$ApC'$</td>
<td>1.47</td>
<td>1.59</td>
</tr>
<tr>
<td>$ApU'$</td>
<td>1.41</td>
<td>0.52</td>
</tr>
<tr>
<td>$ApG'$</td>
<td>0.97</td>
<td>0.61</td>
</tr>
</tbody>
</table>

$^a$ Stepwise development in water to the origin; 0.15 M LiCl to 17 cm.

$^b$ Stepwise development in water to the origin; 0.1 N acetic acid to 7 cm; and 0.2 N formic acid to 17 cm.

A-G'-T, which is formed by U-A phosphodiester cleavage, predominates while products resulting from Py-Py phosphodiester cleavage are obtained in smaller yields. Similar observations have been made with many other oligonucleotide dialcohols (data not shown). These results are to be expected since RNase A is known to exhibit a distinct preference for Py-Pu over Py-Py phosphodiester bonds.\(^{23}\)

A ribonuclease specific for either C or U residues was not available to us. To discriminate between C and U cleavage sites we have therefore taken advantage of the observation that two adjacent oligonucleotides, $n$ and $n + \text{Py}$, differing by 1 mononucleotide unit, Py, exhibit different mobility shifts on PEI-cellulose thin layers at pH 2.6 depending on whether the terminally added monomer unit is CMP or UMP. The transfer technique described in METHODS was found convenient for this purpose. After the partial digestion products in a controlled nuclease $S_1$/phosphatase digest of A-A-U-C-U-C-U-U-A-G'-T had been resolved according to size on a PEI-cellulose thin layer and located by film detection, compounds of $n = 2 - 8$ were transferred via direct contact to a second PEI-cellulose sheet during chromatographic development in a pH 2.6 formate system. After this, the radioactive oligonucleotide derivatives were located by fluorography. The addition of Cp was found to cause a much smaller mobility shift than the addition of Up as shown in Fig. 3B. The angle between a line connecting two adjacent
spots, n and n + Py, and the vertical passing through spot n is much smaller if Py is Cp or m^5Cp (12 - 25°) than if Py is Up, m^5Up, Dp or ψp (50 - 62°), see Fig. 4. These angles were obtained by subjecting digests of several diverse oligonucleotide dialcohols to chromatographic analysis at pH 2.6 (data not shown).

Pre-equilibration of the layer at pH 2.6 (see METHODS) was found necessary to avoid the formation of secondary fronts in the acidic system.

As shown in Fig. 3B, the sequence 8 7 6 5 4 3 -U-C-U-C-U-U- may be read directly from the fluorogram obtained after chromatography of the partial nuclease S_1/phosphatase digestion products in the pH 2.6 formate system.

The mobility shift technique may be applied also to oligonucleotide dialcohols containing modified derivatives of C and/or U (see above) but additional techniques such as base analysis^2,14,24 or sequencing procedures entailing the direct identification of the modified constituents^1,2,12 are required to establish unequivocally the identity of these components. This applies also to modified purines.

As shown in Fig. 5, specific endonuclease cleavage can be

Fig. 4.
Scheme illustrating the discrimination between C and U residues by mobility shift analysis at pH 2.6. For details, please consult text.
Resolution according to chain length on a PEI-cellulose thin layer of products obtained by partial RNase U$_2$ and nuclease S$_1$/phosphatase digestions of A-A-U-C-C-U-U-C-C-C-C-A-C-C-A-C-C-A'-T. For conditions of nuclease S$_1$ digestion, please consult legend to Fig. 3. Aliquots of the individual digests were pooled as indicated in METHODS. The dinucleotide derivative, which had migrated onto the wick in this particular experiment, was shown by chromatography in system 1 of ref. 1 to be present in nuclease S$_1$/phosphatase digests of the nonadecanucleotide dialcohol; it could not be detected in RNase U$_2$ digests.
used to identify the positions of A and Py in rather large (\(^{3}\)H)-labeled oligonucleotide dialcohols such as A-A-U-C-C-U-U-C-C-C-C-A-C-A-C-C-A'-T. Base analysis showed this compound to contain only A, C, and U, with A' being 3'-terminal. For separating the intermediates in controlled nuclease S\(_1\)/phosphatase and RNase U\(_2\) digests of this compound, the gradient was slightly modified as described in METHODS. From the spot distributions and data in Fig. 5 one can read the sequence A-A-(Py)\(_{10}\)-A-(Py)\(_2\)-A-Py-X\(_1\)-X\(_2\)-T. Positions 1 and 2 were identified as A'-T and Py respectively as RNase A digestion liberated (\(^{3}\)H)-labeled adenosine dialcohol. Similar spot patterns were obtained when E. coli tRNA\(_{Tyr}\) itself rather than the 3'-terminal nonadecamer was 3'-end-labeled with tritium and analyzed by controlled endonuclease digestions and TLC. The resolution of the compounds of n = 17, 18, and 19 could be improved by using 1.2 M Tris-HCl, 8.5 M urea, pH 8.0 as the final solvent and letting the solvent front run to 4 - 6 cm on the paper wick. The conditions described for the mobility shift technique are not directly applicable to pyrimidines that are located more than about 12 nucleotides from the labeled 3'-terminus. We are currently modifying the chromatographic conditions so as to accommodate intermediates of n > 12 in the partial digests. On the other hand, the positions of G residues in oligonucleotide dialcohols of this size may readily be determined by partial digestion with RNase T\(_1\) under the conditions detailed in METHODS (unpublished experiments).

DISCUSSION

For oligoribonucleotides containing the four major nucleosides, A, G, C, and U, the present method is simple and straightforward. Treatment of the oligonucleotide (or a ribonuclease/phosphatase digest of RNA) with NaIO\(_4\) and (\(^{3}\)H)-labeled NaBH\(_4\) is first used to convert the oligonucleotide(s) to the corresponding labeled dialcohol derivative(s) which are subsequently partially digested with endonucleases in the presence of carrier RNA. The digestion products are resolved according to chain length in an urea system on PEI-cellulose anion-exchange thin layers. Digestion with nuclease S\(_1\)/phosphatase yields all expected 3'-terminally
labeled digestion products\(^2\), providing a series of marker compounds of defined size. Controlled digestions with RNases A, T\(_1\), and U\(_2\) result in the formation of specific (\(^3\)H)-labeled cleavage products whose separation in a Tris/urea solvent system alongside the nuclease S\(_1\)/phosphatase digest allows one to deduce unambiguously the sequence of the parent compound. C residues are distinguished from U residues by an additional chromatographic step. As this method takes advantage of the known ribonuclease specificities, additional procedures for the identification of (\(^3\)H)-\(^1\)\(^2\),\(^12\) or (\(^32\)P)-labeled\(^2\) terminal positions of the partial digestion products are unnecessary. The procedure therefore is very rapid; a single worker can easily determine the sequences of all RNase A and RNase T\(_1\) digestion products of a RNA of \(n = 100\) in less than 2\(\frac{1}{2}\) weeks. It may be possible to reduce this time further if one uses pre-sensitized film for fluorography\(^25\).

The procedure is also highly sensitive. Standard conditions for RNase A and RNase T\(_1\) digestions\(^11,26\) can be readily scaled down to utilize about 1 \(\mu\)g RNA in 5 - 10 \(\mu\)l solution (unpublished experiments); since Tris buffer is not compatible with subsequent (\(^3\)H)-labeling\(^14\), we are using bicine buffer (20 mM, pH 7.8) in the enzymic digestions\(^12,14\). If (\(^3\)H)NaBH\(_4\) of high specific activity (50 - 60 Ci/mmole), which has recently become available commercially (see MATERIALS), is used, 1 \(\mu\)g RNA yields at least \(0.6 \times 10^6 - 10^6\) dpm of individual 3'-end-labeled oligonucleotides after elution from the PEI-cellulose layer. This is ample material to sequence each compound even if modified nucleosides are present (see below). If there are no modified nucleosides, less than 0.2 \(\mu\)g RNA per digest may be sufficient.

The present procedure is formally related to the DNA sequencing procedure recently elaborated by Maxam and Gilbert\(^27,28\), in which DNA fragments, labeled at their 5'-ends with \(^{32}\)P, are cleaved in a base-specific manner by chemical methods and the cleavage products separated according to size by polyacrylamide gel electrophoresis. The sequences of the DNA fragments may then be "read" directly from autoradiograms of the gels. In the procedure detailed in the present communication, RNA fragments, labeled at their 3'-ends with \(^3\)H, are cleaved specifically by
enzymic digestion and the cleavage products separated according to size by PEI-cellulose TLC. The sequences of the RNA fragments may then be "read" directly from fluorograms of the chromatograms. Positions 1 (the labeled 3'-end) and 2 are identified by a few simple additional manipulations, the choice of which largely depends on the nucleoside occupying position 2 as described under RESULTS.

For the separation of very large partial RNA digestion products according to size (n > 20 - 25), gel electrophoresis probably will afford superior resolution as compared to anion-exchange TLC in the presence of urea. As indicated in Fig. 5, it appears, however, that the present chromatographic procedure can be used to characterize up to at least 20 positions from the 3'-end of RNAs and large RNA fragments.

The analysis of large pyrimidine-rich oligonucleotides (n > 20) would be facilitated by the availability of endonucleases specific for (or at least clearly preferring) either one of the two pyrimidines. This would also obviate the chromatographic mobility shift analysis step and thus simplify the identification of pyrimidines in general. Several laboratories have reported on ribonucleases exhibiting preference for either of the pyrimidines but these enzymes have not been used thus far for partial digestions of end-labeled oligonucleotides. We are currently investigating if RNases isolated from Physarum polycephalum and human plasma might be suitable for this purpose. In preliminary experiments (unpublished), we have attempted to modify by several known chemical reactions either of the pyrimidines in end-labeled oligonucleotides so as to make them resistant to cleavage by RNase A; thus far, the mobility shift technique has consistently given more reproducible results. Whether it will be possible to work out chemical cleavage methods specific for either C or U in RNA fragments is not known at the present time.

The occurrence of the various modified nucleosides in nucleic acids presents additional analytical problems irrespective of the sequencing strategy. If the structure and chromatographic properties of a modified constituent of RNA are known, methods are now available for their qualitative and quantitative analysis.
at the subnanogram level\textsuperscript{2,14,24}. In our laboratory, various highly sensitive methods have been elaborated to determine unequivocally the position of modified nucleosides in nonradioactive RNA fragments. The procedures developed by us entail exonucleolytic\textsuperscript{1,12,19} or endonucleolytic\textsuperscript{2} degradation of the polynucleotide chain combined with terminal (\textsuperscript{3}H)-\textsuperscript{1,2,12,19} or (\textsuperscript{32}P)-labeling\textsuperscript{2,19}. Base composition analysis of each compound to be sequenced is the most important and an absolutely essential first step to demonstrate the presence of modified nucleosides. Amounts as low as 1 - 2 pmole of oligonucleotide (about 0.3 ng per mononucleotide unit; n < 20) are sufficient for accurate quantitative analysis by a scaled-down version\textsuperscript{2} of the (\textsuperscript{3}H)-method for base composition analysis. Longer polynucleotides require somewhat larger amounts of material; the method described\textsuperscript{2} is, however, suitable without modifications for the qualitative base analysis of polynucleotides of n < 200. Although pmole amounts frequently suffice to sequence oligonucleotides containing modified nucleosides, the necessity to carry out additional manipulations does affect the overall sensitivity of the present as well as any other sequencing procedure (see above).

As shown in this and a previous\textsuperscript{2} communication, the combination of end-labeling and controlled endonuclease cleavage is a powerful new tool for the sequence analysis of minute amounts of nonradioactive RNA fragments. Related procedures involving 5'-terminal (\textsuperscript{32}P)-labeling, partial nonspecific endonucleolytic digestion, and 2-D "homochromatography"\textsuperscript{37} have recently been developed independently in other laboratories\textsuperscript{24,38-40}. In the present work, the advantages derived from using specific endonuclease cleavage, in combination with end-labeling, have been demonstrated. This approach is obviously not restricted to 3'-(\textsuperscript{3}H)-labeled oligonucleotides; it should be applicable to 3'-(\textsuperscript{32}P)-\textsuperscript{24,38,41} and 5'-(\textsuperscript{32}P)-labeled\textsuperscript{24,38-40,42} RNA fragments as well. In actual sequencing work, the analysis of 5'-end-labeled oligonucleotides after controlled digestion with specific endonucleases can be expected to supplement the present procedure. For example, the sequence of the 5'-portion of a large fragment may be more readily revealed by analyzing the cleavage points in specific digests of a 5'-end-labeled derivative of the parent.
compound while the analysis of a 3'-labeled derivative may yield more information about its 3'-portion.

ABBREVIATIONS


Acknowledgements

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

We are grateful to Dr. S.H. Chang for generously providing several oligonucleotides of known sequence, and to Drs. E. Randerath, C.C. Levy, and J.-P. Bargetzi for discussions and helpful suggestions.

This is part XXIII of a series entitled "Analysis of Nucleic Acid Derivatives at the Subnanomole Level". A preliminary account of this work has been presented previously.

REFERENCES

The positions within polynucleotide chains are numbered in the 3'→5' direction as this simplifies the presentation of our data.