The Nucleotide Sequences of some Large Ribonuclease T₁ Products from Bacteriophage R17 Ribonucleic Acid

By PETER G. N. JEPPESEN

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

(Received 25 February 1971)

A method of 'fingerprinting' high-molecular-weight ³²P-labelled RNA species, using a two-dimensional thin-layer-chromatographic separation of ribonuclease T₁ digestion products, has been applied to RNA from the *Escherichia coli* bacteriophage R17. The 'fingerprinting' technique, besides giving a unique pattern that can be used as a characterization of the RNA, has made it possible to isolate a number of the larger oligonucleotides and to determine their nucleotide sequences.

Progress in studies of sequences in RNA depends very largely on the development of methods for fractionation of the complex mixtures of oligonucleotides produced by partial degradation. Brownlee & Sanger (1969) have described a two-dimensional technique for fractionating ³²P-labelled oligonucleotides using ionophoresis on cellulose acetate in the first dimension and 'homochromatography' on t.l.c. plates in the second dimension. The method was particularly useful for separating large oligonucleotides (10-50 residues long) that could not be easily resolved by previous 'fingerprinting' techniques and was applied to partial ribonuclease digests of 6S RNA (Brownlee, 1971). The present paper describes further development of the method and its application to complete digests of the RNA from bacteriophage R17 with ribonuclease T₁. The technique, besides giving a unique pattern that can be used as a characterization of the RNA, has made it possible to isolate a number of larger oligonucleotides in a sufficiently pure form for determination of their complete nucleotide sequences.

The RNA of bacteriophage R17 contains 3300-3500 nucleotide residues and the genome comprises three cistrons which, in order along the RNA molecule, are: 5'-A-protein-coat protein-RNA replicase-3' (Jeppesen, Steitz, Gesteland & Spahr, 1970b).

MATERIALS AND METHODS

Preparation of ³²P-labelled bacteriophage R17 RNA. ³²P-labelled R17 bacteriophages were grown as described by Steitz (1969) and purified as described by Anderson, Dahlberg, Bretscher, Revel & Clark (1967). The RNA was extracted from the bacteriophage by the method of Gesteland & Boedtker (1964). The specific radioactivity of RNA thus prepared varied from batch to batch, from 1 to 2 µCi/µg.

Thin-layer homochromatography 'fingerprinting' procedure. DEAE-cellulose t.l.c. plates (20 cm × 40 cm) were prepared as described by Brownlee & Sanger (1969) from DEAE-cellulose and cellulose in the ratio 1:7.5 (w/w).

A portion (20 µg) of ³²P-labelled R17 RNA (approx. 20 µCi) was digested with 1 µg of ribonuclease T₁ (Sankyo Co. Ltd., Tokyo, Japan) in 3 µl of 0.01 M-tris-HCl buffer, pH 7.4, containing 1 mM-EDTA, for 30 min at 37°C. The digestion mixture was then applied to a strip (3 cm × 55 cm) of cellulose acetate (Schleicher & Schüll, Dassel, West Germany) as described by Sanger, Brownlee & Barrell (1965), and subjected to ionophoresis in 7 M-urea, buffered with 5% (v/v) acetic acid and pyridine to pH 3.5, until the blue and slowest pink marker dyes had separated by approx. 15 cm. The oligonucleotides, which extended from the pink spot to about 3 cm behind the blue spot (as checked by monitoring with a portable Geiger counter) were then transferred to a DEAE-cellulose t.l.c. plate by blotting and the homochromatography second dimension developed with a 3% mixture of partially hydrolysed non-radioactive RNA dissolved in 7 M-urea as described by Brownlee & Sanger (1969).

[It has been found that the commercial RNA used for homochromatography (British Drug Houses Ltd., Poole, Dorset, U.K.) varies from batch to batch in respect of the time required for the alkaline hydrolysis. To obtain separation of oligonucleotides up to 24 base residues in length for the R17 'fingerprint' shown in Plate 1 with the particular batch of RNA used required a 10 min hydrolysis time.]

After the second-dimension run, the thin layer was dried and radioautographed. Oligonucleotides were eluted from the DEAE-cellulose with 30% (v/v) triethylamine carbonate, pH 10.

Sequence analysis of the large ribonuclease T₁ products. The techniques of enzyme digestions and paper ionophoresis used were those described by Sanger *et al.* (1965).

Initial step in sequence analysis. Because of the non-radioactive RNA used as the developing mixture in homochromatography, radioactive oligonucleotides eluted from the DEAE-cellulose thin layer contain a large quantity of carrier oligonucleotides. It was assumed that an average eluted spot (from approx. 1 cm² of thin layer) contained about 200 µg of RNA. One-third portions of

each eluted ribonuclease T₁ product (which will be referred to as a 'T₁-oligonucleotide') were analysed separately as follows.

(i) Ribonuclease A digestion. The sample was incubated with 10 μ l of the enzyme (Worthington Biochemical Corp., Freehold, N.J., U.S.A.; 0.2 mg/ml in 0.01 M-tris-HCl buffer, pH 7.4, containing 1 mM-EDTA) for 1 h at 37°C. The products were fractionated by ionophoresis on DEAE-cellulose paper at pH 3.5 (Adams, Jeppesen, Sanger & Barrell, 1969), located by radioautography and determined by scintillation counting (see paragraph xiii below). Where necessary, the base compositions of the larger products were determined by alkaline hydrolysis (paragraph xii).

(ii) Reaction with a soluble carbodi-imide followed by ribonuclease A digestion (Gilham, 1962). The sample was incubated in 10 μ l of a 50 mg/ml solution of *N*-cyclohexyl-*N'*-[β -morpholinyl-(4)-ethyl] carbodi-imide-methyl-*p*-toluenesulphonate (CMCT) obtained from Fluka A.G., Buchs, Switzerland, in 0.05 M-tris-HCl buffer, pH 7.4, containing 5 mM-EDTA, for 16 h at 37°C, after which 2 μ l of a 1 mg/ml solution of ribonuclease A was added, and digestion carried out for 1 h at 37°C. The digestion mixture was applied as a 2 cm streak to the centre of a sheet of Whatman 3MM paper, and the modified products were fractionated by ionophoresis at pH 3.5. After location by radioautography, they were eluted with 5% (v/v) triethylamine carbonate, pH 10, and evaporated under vacuum. Water was added and they were taken to dryness twice. The blocking groups were then removed by reaction with 0.5 M-NH₃ for 16 h at room temperature, and the products identified by digestion with ribonuclease A (see paragraph vi).

Reaction with CMCT modifies uridylylate and guanylate residues, rendering the former resistant to subsequent ribonuclease A digestion. Thus the above procedure results in modified oligonucleotides (which will be referred to as 'CD-products') with -Cp or -Gp termini, the latter from the original 3'-end of the T₁-oligonucleotide (- $\dot{U}$ - and - $\dot{G}$ - will be used to signify modified uridylylate and guanylate residues respectively).

(iii) Digestion with ribonuclease U₂ (Arima, Uchida & Egami, 1968). The sample was incubated with 10 μ l of a 1 unit/ml solution of ribonuclease U₂ (Sankyo) in 0.05 M-sodium acetate buffer, pH 4.5, containing 2 mM-EDTA and 0.1 mg of bovine serum albumin/ml, for 2 h at 37°C. The products were fractionated by ionophoresis on DEAE-cellulose paper at pH 1.9. After location by radioautography and elution, the base compositions of the products were determined by alkaline hydrolysis and determination of mononucleotides (see paragraphs xii and xiii).

Ribonuclease U₂ cleaves specifically adjacent to purine residues, liberating oligonucleotides terminating in a purine residue (predominantly in the 2':3'-cyclic phosphate form). Purine-pyrimidine sequences are split more readily than purine-purine sequences, and thus it is possible to obtain a series of products differing only in the number of 3'-terminal purine residues, and this can be very useful for overlapping sequences.

(iv) Limited digestion with spleen acid ribonuclease (Bernardi & Bernardi, 1966). The sample was incubated with 10 μ l of a 2 mg/ml solution of acid ribonuclease (a gift from Dr G. Bernardi) in 0.05 M-sodium acetate buffer, pH 5.0, containing 5 mM-EDTA, for 4 h at 37°C. The

products were fractionated by ionophoresis on DEAE-cellulose paper at pH 1.9.

(v) Partial digestion with spleen phosphodiesterase. The sample was incubated for times ranging from 0 to 60 min at 37°C in 10 μ l of spleen phosphodiesterase (0.5 mg/ml) (Worthington) dissolved in 0.1 M-ammonium acetate buffer, pH 5.5, containing 2 mM-EDTA and 0.05% (v/v) Tween-80. The partial products were fractionated by ionophoresis on DEAE-cellulose paper at pH 1.9 (see Sanger *et al.* 1965).

Subsequent steps in the sequence analysis. After the first step in the analysis of an oligonucleotide eluted from a DEAE-cellulose thin layer (as detailed in paragraphs i-v above), it was assumed that the quantity of non-radioactive carrier RNA was greatly decreased, and subsequent steps were carried out under the conditions listed below. Elution of oligonucleotides from DEAE-cellulose paper was by the method of Sanger *et al.* (1965).

(vi) Ribonuclease A digestion. The sample was incubated with 10 μ l of enzyme (0.1 mg/ml) in 0.01 M-tris-HCl buffer, pH 7.4, containing 1 mM-EDTA and 1 mg of non-radioactive carrier RNA/ml for 30 min at 37°C. Products were fractionated by DEAE-cellulose paper ionophoresis at pH 3.5.

(vii) Reaction with CMCT followed by ribonuclease A digestion. Modification conditions were as described in (ii) above, except that 1 mg of carrier RNA/ml was added to the incubation mixture. Subsequent ribonuclease A digestion was by adding 1 μ l of enzyme solution (1 mg/ml), followed by incubation at 37°C for 30 min. Fractionation of products and elution were as described in (ii).

(viii) Ribonuclease U₂ digestion. The digestions were carried out with enzyme at 0.5 unit/ml in 10 μ l of buffer as described in (iii) above, containing in addition 1 mg of carrier RNA/ml, for 2 h at 37°C. Fractionation was as described in (iii) above.

(ix) Spleen acid ribonuclease digestion. Digestion conditions were as described in (iv) above, except that 1 mg of carrier RNA/ml was added to the incubation mixture. Fractionation of products was as described in (iv).

(x) Removal of 3'-terminal phosphate group. The sample to be dephosphorylated was first treated for 1 h at 37°C with 10 μ l of 0.1 M-HCl to decyclize any 2':3'-cyclic phosphate that may be present after enzymic cleavage. The acid was then removed by freeze-drying, after which the radioactive oligonucleotide was incubated for 1 h at 37°C with 10 μ l of a 0.1 mg/ml solution of bacterial alkaline phosphatase (Worthington) in 0.05 M-tris-HCl buffer, pH 8.9, containing 0.01 M-Mg²⁺ and 1 mg of carrier RNA/ml. The dephosphorylated product was freed of enzyme and P_i by brief DEAE-cellulose paper ionophoresis at pH 1.9.

(xi) Complete hydrolysis of dephosphorylated oligonucleotides with snake venom phosphodiesterase. The sample was incubated for 2 h at 37°C with 10 μ l of snake venom phosphodiesterase (0.1 mg/ml) (Worthington) in 0.05 M-tris-HCl buffer, pH 8.9, containing 0.01 M-Mg²⁺ and 1 mg of carrier RNA/ml. The resulting 5'-mononucleotides were separated by ionophoresis on Whatman 3MM paper at pH 3.5. Yields of mononucleotides were determined as described in paragraph (xiii) below.

(xii) Base composition by alkaline hydrolysis. The sample was incubated for 16 h at 37°C with 10 μ l of 0.2 M-NaOH. The resulting mononucleotides were frac-

(b)

EXPLANATION OF PLATE 1

(a) Radioautograph of the products of ribonuclease T_1 digestion of ^{32}P -labelled R17 RNA fractionated by the two-dimensional thin-layer procedure described in the Materials and Methods section. (b) Diagram of the lower portion of (a), indicating those oligonucleotides whose sequence analysis is reported in this paper (solid lines and lettering).

tionated by ionophoresis at pH 3.5 on Whatman 3MM paper, and determined as described below.

(xiii) Determination by scintillation counting. After location of the radioactive nucleotides on paper or DEAE-cellulose paper by radioautography, they were cut out and counted for radioactivity in toluene containing 0.5% (w/v) 2,5-bis-(5-*tert.*-butylbenzoxazol-2-yl)thiophen (BBOT); [CIBA (A.R.L.) Ltd., Duxford, Cambs., U.K.] in a Unilux II counter (Nuclear-Chicago Corp.). If further digestion of the sample was required, the scintillator was washed out with toluene and the paper dried before elution.

RESULTS AND DISCUSSION

'Fingerprint' of ribonuclease T_1 digestion products of R17 RNA. Plate 1(a) shows a radioautograph of a two-dimensional fractionation of the products of complete ribonuclease T_1 digestion of ^{32}P -labelled R17 RNA, by using homochromatography on a DEAE-cellulose thin layer as the second dimension. The resolution of a particular oligonucleotide by the two-dimensional system depends both on its size and its base composition. Homochromatographic mobility depends largely on size, although the base composition also has an effect: in particular, purines appear to retard. For example, spot *g* moves slightly slower than spot *f* in the second dimension (see Plate 1) although *f* is 17 and *g* is 13 residues long (see Table 3). The high adenylate content of *g* appears to decrease its mobility on homochromatography considerably.

The very large products (say 15–24 residues), which move slowest by homochromatography, are in general well separated, whereas separation of slightly smaller oligonucleotides (10–14 residues) depends on their mobility in the first dimension: those that are particularly fast or slow (i.e. having a high or low uridylylate content respectively) being in less crowded parts of the thin layer. Oligonucleotides of more average composition tend to bunch in the centre. There is little separation of products of under 10 residues. The procedure, however, separates a number of large T_1 -oligonucleotides as pure species suitable for sequence analysis, and hence serves as a method of characterizing large RNA molecules.

Sequence analysis of large T_1 -oligonucleotides. Plate 1(b) shows a diagram of the better resolved region of the R17 T_1 'fingerprint', indicating those oligonucleotides whose sequences were investigated and are reported here. Those selected for analysis were: (1) the largest well-resolved spots, to characterize the R17 RNA molecule, and (2) certain smaller resolved oligonucleotides which preliminary sequence studies showed might come from parts of the R17 genome whose sequences were already known (e.g. the three ribosome-binding sites and the 5'- and 3'-ends). In previous work on sequences of ribonuclease T_1 products, extensive

use was made of the method of partial digestion with phosphodiesterases (e.g. Holley *et al.* 1965; Brownlee & Sanger, 1967). These methods were not in general applicable to the large oligonucleotides studied here, and a more empirical approach with various types of endonuclease digestions was found to be more profitable. Each T_1 -oligonucleotide was analysed by three basic procedures: ribonuclease A digestion; modification with CMCT followed by ribonuclease A digestion; and ribonuclease U_2 digestion (see the Materials and Methods section). The results of these three analyses are given in Table 1. In several cases, the information provided by these three procedures was sufficient to determine the sequence of an oligonucleotide unambiguously, and where this was possible, the results are shown in Table 2. As an example, the sequence of oligonucleotide *a* was deduced as follows.

p is present among the CD-products (Materials and Methods section, paragraph ii), hence the 3'-terminal sequence is -C-Gp. The ribonuclease U_2 product terminating -Gp is (U_4 ,C)Gp, and thus its sequence must be U-U-U-U-C-Gp (1). The only CD-product which can terminate in - $\bar{U}$ - $\bar{U}$ - $\bar{U}$ - $\bar{U}$ -Cp, and hence overlap (1), is (A-A- $\bar{U}$, $\bar{U}_3$)Cp, which must then have the sequence A-A- $\bar{U}$ - $\bar{U}$ - $\bar{U}$ - $\bar{U}$ -Cp, and be preceded by a C. Thus the 3'-terminal sequence of oligonucleotide *a* is -C-A-A-U-U-U-U-C-Gp (2). There are three ribonuclease U_2 products terminating -A-Ap: A-Ap itself (which must come from the 5'-end of the oligonucleotide as there are no -A-A-A- sequences present), U-U-A-Ap and (U_2 , C_2)A-Ap. Of these, only (U_2 , C_2)A-Ap can overlap (2), and because there is no $\bar{U}$ -Cp or $\bar{U}$ - $\bar{U}$ -Cp present among the CD-products, its sequence is U-U-C-C-A-Ap. Since it is obtained by the action of ribonuclease U_2 it must be preceded by an A. The only remaining CD-product which can overlap with this is (A- $\bar{U}$, $\bar{U}_2$)Cp, whose sequence must therefore be $\bar{U}$ -A- $\bar{U}$ - $\bar{U}$ -Cp. Combining these, the sequence extending from the 3'-end is: -U-A-U-U-C-C-A-A-U-U-U-U-C-Gp (3). There is just one CD-product left, namely (A-A- $\bar{U}$, $\bar{U}$)A-A-Cp, which must have the sequence A-A- $\bar{U}$ - $\bar{U}$ -A-A-Cp to account for the three remaining ribonuclease U_2 products: A-Ap, U-U-A-Ap and (C, $\bar{U}$)Ap. The complete sequence of oligonucleotide *a* is therefore:

(Lines above the sequences in the remainder of this paper represent ribonuclease U_2 products, and lines below represent CD-products.)

The sequences of the other oligonucleotides presented in Table 2 were derived in a similar manner. For the remaining T_1 -oligonucleotides, the three initial analytical procedures were not sufficient to derive a unique nucleotide sequence, but did

Table 1. Preliminary sequence data on R17 T₁ oligonucleotides lettered in Plate 1(b)

(a) Ribonuclease A digestion, fractionation and quantitation of products were performed as described in the Materials and Methods section, paragraph (i). The values in parentheses indicate molar yields of oligonucleotides to the nearest whole number, where these are not unity. (b) Modification by CMCT followed by ribonuclease A digestion, and fractionation and characterization of products were carried out as detailed in paragraph (ii) of the Materials and Methods section. - $\dot{U}$ - and - $\dot{G}$ - represent modified uridylylate and guanylylate residues respectively. (c) Ribonuclease U ₂ digestion, and fractionation and characterization of products as described in paragraph (iii) of the Materials and Methods section. Series of products, each of which are in less than molar yield, originating from a polypurine sequence (see text) are bracketed together. (Where yields of products were estimated visually from the radioautograph, rather than determined by scintillation counting, this is indicated by underlining; for example, (—) and (==) would mean approximately twice and three times molar yield respectively.)							
Oligo-nucleotide	Ribonuclease A products	CD-products	Ribonuclease U ₂ products	Oligo-nucleotide	Ribonuclease A products	CD-products	Ribonuclease U ₂ products
<i>a</i>	A-A-A-Up(2) A-A-Cp A-Up Gp Cp(3) Up(6)	(A-A- $\dot{U}$, $\dot{U}$)A-A-Cp (A- $\dot{U}$, $\dot{U}_2$)Cp (A-A- $\dot{U}$, $\dot{U}_3$)Cp <u>Gp</u> Cp Up(6)	(U ₄ ,C)Gp {(U ₃ ,C ₂)A-Ap {(U ₂ ,C ₃)Ap U-U-A-Ap U-U-Ap (U,C)Ap A-Ap Ap	<i>j</i>	Cp(5) Up(5)	A-Cp <u>Cp</u> U- $\dot{U}$ -A-Cp U-A-A-Cp U- $\dot{U}$ -Cp U-Cp A-Gp Cp	Ap {(U ₃ ,C ₂)A-Ap {(U ₃ ,C ₂)Ap {(U ₂ ,C ₃)A-Gp {(U ₂ ,C ₃)Ap {(U ₃ ,C)Ap Gp Ap
<i>b</i>	A-A-A-A-Up A-A-Cp A-Up(3) Gp Cp(4) Up(5)	A- $\dot{U}$ -A-A-Cp (A-A-A-A- $\dot{U}$,A- $\dot{U}_2$)Cp (A- $\dot{U}$, $\dot{U}$)Cp U-Cp <u>Gp</u>	(U ₅ ,C ₂)Ap (U,C)Gp (U ₂ ,C ₃)Ap {U-A-A-A-Ap U-A-A-Ap U-A-Ap U-Ap A-Ap Ap	<i>l</i>	A-A-A-Cp A-Gp A-Cp Cp Up(4)	A-A-A-Cp (U)A-Cp <u>A-Gp</u> Cp	U-U-U-U-Ap {C-C-A-Gp {C-C-Ap {C-A-A-Ap {C-A-Ap A-Ap Gp Ap U-U-U-U-Ap {C-C-A-Gp {C-C-Ap {C-A-A-Ap A-Ap Gp Ap
<i>c</i>	A-Up(4) Gp Cp(3) Up(4)	A- $\dot{U}$ -A- $\dot{U}$ -Cp (A- $\dot{U}_2$, $\dot{U}_2$)Cp U- $\dot{U}$ -Gp Cp	(U ₃ ,C)Gp (U ₂ ,C)Ap (U,C)Ap U-Ap	<i>m*</i>	A-A-Up A-Up A-Gp Cp(4) Up(3)	(A-A- $\dot{U}$, $\dot{U}$)Cp (A- $\dot{U}$, $\dot{U}$)Cp U-A-Gp Cp <u>Cp</u>	U-U-U-U-Ap U-U-Gp {(U,C)A-Ap {(U,C)Ap U-U-Gp (A-A-A- $\dot{U}$, $\dot{U}_2$)A-Cp (A-A- $\dot{U}$, $\dot{U}$)Gp Cp A-A-Up A-Up A-Cp Gp Cp(2) Up(3)
<i>d</i>	A-A-A-Up A-A-Up(2) A-Cp Gp Cp(2) Up(3)	A-A-A- $\dot{U}$ -Cp (A-A-A- $\dot{U}$, $\dot{U}_2$)A-Cp (A-A- $\dot{U}$, $\dot{U}$)Gp Cp	{(U,C)A-Ap {(U,C)Ap U-U-Gp {A-A-Ap A-Ap C-C-A-Ap C-C-Ap Ap	<i>n*</i>	A-A-Cp A-Up A-Cp Gp Cp(2) Up(4)	(A $\dot{U}$, $\dot{U}$)Cp <u>A-A-Cp</u> A-Cp Cp Gp	(U ₄ ,C ₂)Ap (U,C)Ap C-Gp {A-Ap Ap Gp (U ₄ ,C ₂)Ap (U,C)Ap C-Gp {A-Ap Ap

<i>e</i>	A-A-A-Up A-A-Cp A-Up Gp Cp(6) Up(3)	(A-Ü,Ü ₃)Gp A-A-A-Ü-Cp A-A-Cp Cp ≡	U-U-U-U-Gp (U,C ₂)Ap {C-C-C-A-A-Ap C-C-C-A-A-Ap C-C-C-Ap C-A-Ap C-Ap A-Ap Ap	A-A-Gp A-A-Cp A-Up A-Cp Up(3)	(Ü-A-Ü)A-A-Gp Ü-A-A-Cp Ü-A-Cp	{U-U-A-A-Gp U-U-A-Ap U-U-Ap (U,C)A-Ap (U,C)Ap U-Ap C-Ap Gp Ap
<i>f</i>	A-Up A-Cp A-Gp Cp(7) Up(4)	(A-Ü,Ü)Cp A-Gp Ü-Cp Ü-U-Cp A-Cp Cp ≡	{(U ₃ ,C ₄)A-Gp (U ₃ ,C ₄)Ap (U ₃ ,C ₃)Ap C-Ap Gp	A-Cp A-Up Gp Cp(8) Up(2)	(A-Ü,Ü)Cp A-Cp Ü-Cp Gp Cp ≡	(U ₃ ,C ₇)Gp (U,C)Ap C-Ap
<i>g</i>	A-A-A-A-Up A-A-Up(2) A-Gp	{U-A-A-A-Ap U-A-A-Ap U-A-Ap U-A-Gp U-Ap A-Ap Gp	{U-A-A-A-Ap U-A-A-Ap U-A-Ap U-A-Gp U-Ap A-Ap Gp	A-A-A-Cp A-Up(2) Gp Cp(3) Up	A-A-A-Cp (A-Ü,Ü)Cp A-Ü-G Cp ≡	{(U ₃ ,C)A-A-Ap (U ₃ ,C)A-Ap (U ₃ ,C)Ap C-C-Ap C-Ap U-Gp A-Ap Ap
<i>h</i>	A-A-A-Gp A-A-A-Up A-Up A-Cp(2) Cp(2) Up	(A-Ü,Ü)A-A-A-G A-A-A-Ü-A-Cp A-Cp Cp ≡	{U-U-A-A-A-Gp U-U-A-A-Ap U-U-A-Ap U-U-Ap C-A-A-Ap C-A-Ap C-C-Ap U-Ap C-Ap A-Ap Gp Ap	A-A-Up A-Up Gp Cp(2) Up(4) A-A-Cp Gp Cp(5) Up	(A-Ü, A-A-Ü,Ü)Gp Cp ≡	{U-U-U-U-A-Ap U-U-U-U-Ap C-C-Ap U-Gp Ap
<i>i</i>	A-A-Up A-Cp Gp	(A-A-Ü,Ü ₂)Cp A-Ü-Cp Ü-Gp	(U ₅ ,C ₄)Ap (U,C ₂)Gp A-Ap	A-Cp(2) Cp(3) Up(2) A-A-A-Gp A-A-Cp A-Cp	A-A-Cp Ü-Cp Gp Cp ≡	{(U,C ₄)A-Ap (U,C ₄)Ap C-C-Gp Ap
<i>j</i>	A-A-Up A-Cp Gp				A-Cp Ü-Ü-A-Cp Cp ≡	C-C-Ap C-C-C-A† U-U-Ap

* Oligonucleotides *m* and *n* are isomeric, and were not separated in ribonuclease T₁ digests of whole R17 RNA fractionated as in Plate 1. In partial ribonuclease T₁ digests of R17 RNA fractionated on polyacrylamide gels, however (see for example Adams *et al.* 1969; Nichols, 1970; Jeppesen *et al.* 1970a), the T₁-products *m* and *n* were found in different gel bands, from which their identities were established. Their sequences are included here for the sake of completeness.

† The sequence C-C-C-A was determined by dividing the sample in two, and performing an alkaline hydrolysis on one half and hydrolysis with snake venom phosphodiesterase on the other.

Table 2. Sequences of $R17 T_1$ -oligonucleotides which can be deduced unambiguously from the data presented in Table 1

Lines above the sequences represent products of ribonuclease U_2 digestion, lines below represent CD-products.

Oligonucleotide	Sequence
<i>a</i>	<u>A-A-U-U-A-A-C-U-A-U-U-C-C-A-A-U-U-U-U-C-Gp</u>
<i>b</i>	<u>A-U-A-A-C-U-C-A-U-A-A-A-U-U-U-C-U-C-U-A-U-C-Gp</u>
<i>d</i>	<u>A-A-A-U-U-U-A-C-C-A-A-U-C-A-A-U-U-Gp</u>
<i>e</i>	<u>C-A-A-C-C-C-A-A-A-U-C-C-A-U-U-U-U-Gp</u>
<i>l</i>	<u>U-U-U-U-A-C-A-A-C-C-A-Gp</u>
<i>m</i>	<u>A-A-U-U-C-C-A-U-U-C-C-U-A-Gp</u>
<i>o</i>	<u>U-A-C-U-A-A-C-A-U-U-A-A-Gp</u>
<i>r</i>	<u>C-C-A-U-U-U-U-U-A-A-U-Gp</u>
<i>t</i>	<u>U-U-A-C-C-A-C-C-C-A</u>

indicate the course of subsequent analyses; the analyses of each of them are dealt with below individually.

Oligonucleotide *c*. Two alternative sequences could account for the data in Table 1:

The ambiguous U_2 products were allowed to react with CMCT and then digested with ribonuclease A, which gave:

Confirmation that alternative (ii) is correct was obtained by digesting the large CD-product with ribonuclease U_2 , after removing the modifying groups:

Oligonucleotide *f*. Limited treatment of oligonucleotide *f* with spleen acid ribonuclease (Materials and Methods section, paragraph iv) gave two main products, which were characterized by digestion with ribonuclease A:

Dephosphorylation of (i), to mark the 3'-end, followed by ribonuclease U_2 digestion gave C-A and (C₃, U_2)Ap. Combination of this additional data with that in Table 1 gives the partial sequence:

(Arrows indicate positions of cleavage by acid ribonuclease.)

Treatment of the two major ribonuclease U_2 products with acid ribonuclease (Materials and Methods section, paragraph ix) gave the following:

This suggested that the sequence is U-U-C-C-C-Ap, which was confirmed by dephosphorylation followed

by hydrolysis with snake venom phosphodiesterase, giving pU, pC(3), pA.

Its sequence is therefore U-U-C-C-C-U-C-Ap.
The complete sequence of oligonucleotide *f* is:

(For the remaining oligonucleotides, the partial sequences suggested by the data in Table 1 are written first.)

Oligonucleotide *g*.

Limited treatment of oligonucleotide *g* with acid ribonuclease gave the following products, which were characterized by digestion with ribonuclease A: A-Ap, U-A-Ap(0.4), U-A-Gp(1.0), U-A-A-Ap(1.1), and A-A-U-A-Ap(0.6). (Numbers in parentheses indicate molar yields of products, where measured.) A-A-U-A-Ap, U-A-A-Ap and U-A-Gp are therefore primary split products, A-A-U-A-Ap undergoing further breakdown to A-Ap and U-A-Ap. This can only be the case if alternative sequence (ii) above is correct:

Oligonucleotide *h*.

Limited digestion with acid ribonuclease gave only three products, which were characterized by digestion with ribonuclease A:

(ii) was further analysed by dephosphorylation followed by ribonuclease U₂ digestion, giving the following products:

(ii) therefore has the sequence U-A-C-A-C-C-Ap. Thus the sequence of *h* is:

Oligonucleotide *j*.

Treatment of the large ribonuclease U₂ product with acid ribonuclease gave the following:

Dephosphorylation of (U₂C₂)Ap, followed by hydrolysis with snake venom phosphodiesterase gave pU, pC(2) and pA. The sequence of the U₂

product is therefore U-U-U-C-C-U-U-C-Ap, and oligonucleotide *j* has the sequence:

Oligonucleotide *k*.

Limited treatment of oligonucleotide *k* with acid ribonuclease gave only three products, which were characterized by digestion with ribonuclease A: (C,U)A-Gp, (A-A-C,U₃C₂)p and (A-C,U₃)p. This requires that alternative sequence (ii) be correct:

Dephosphorylation of ribonuclease U₂ product (C,U-U-C)U-A-Ap followed by hydrolysis with snake venom phosphodiesterase gave pC, pU(3) and pA(2). The sequence of oligonucleotide *k* is therefore:

Oligonucleotide *n*.

The two ambiguous ribonuclease U₂ products were reacted with CMCT and digested with ribonuclease A:

Therefore, alternative (ii) is the correct sequence of oligonucleotide *n*.

Oligonucleotide *p*.

The sequence of the large ribonuclease U_2 product (U_2, C_7)Gp was determined by treatment with acid ribonuclease, which gave the following products:

Dephosphorylation of (U, C_4)Gp followed by hydrolysis with snake venom phosphodiesterase gave pC(4) and pG. Thus the sequence of the U_2 pro-

duct is $U-C-C-C-U-C-C-C-C-Gp$, and hence oligonucleotide p is

Limited digestion with acid ribonuclease gave three main products, which were characterized by digestion with ribonuclease A:

C-C-Ap, U-Gp and ($A-A-A-C, C, U_2$)Ap, with a

minor yield of ($A-A-A-C, A-U, C_3, U$)Ap. This leads to sequence (ii) for oligonucleotide q :

Treatment of ribonuclease U_2 product ($C_3, U-C$)Ap with acid ribonuclease gave two products: C-C-Cp

and (U, C)Ap, i.e. sequence is $C-C-C-U-C-Ap$. The sequence of oligonucleotide s is therefore C-C-C-U-C-A-A-C-C-Gp.

Oligonucleotide u .

Partial digestion of oligonucleotide u with spleen phosphodiesterase gave the following series of products, which were characterized by digestion with ribonuclease A:

Table 3. Sequences of the R17 T_1 -oligonucleotides lettered in Plate 1(b)

The first nine oligonucleotides are placed in the order in which they occur in the R17 genome, and their positions in the genome indicated (see text); the positions of the remainder have not been established.

Oligonucleotide	Sequence	Position in R17 genome
<i>r</i>	C-C-A-U-U-U-U-A-A-U-Gp	49 residues from 5'-end
<i>m</i>	A-A-U-U-C-C-A-U-U-C-C-U-A-Gp	Overlaps 5'-side of A-protein initiation site
<i>s</i>	C-C-C-U-C-A-A-C-C-Gp	Coat protein ribosome binding site, to left of initiating AUG
<i>k</i>	C-U-U-C-U-A-A-C-U-U-U-A-C-U-C-A-Gp	Coat protein cistron coding for amino acids 1-6 (Fig. 1)
<i>h</i>	C-A-A-A-U-A-C-A-C-C-A-U-U-A-A-A-Gp	Coat protein cistron coding for amino acids 57-61 (Fig. 1)
<i>a</i>	A-A-U-U-A-A-C-U-A-U-U-C-C-A-A-U-U-U-U-C-Gp	Coat protein cistron coding for amino acids 89-95 (Fig. 1)
<i>q</i>	C-C-A-U-U-C-A-A-A-C-A-U-Gp	Intercistronic region between coat and replicase
<i>u</i>	A-C-A-A-C-A-A-A-Gp	Beginning of replicase cistron
<i>t</i>	U-U-A-C-C-A-C-C-C-A	3'-end (Dahlberg, 1968)
<i>b</i>	A-U-A-A-C-U-C-A-U-A-A-A-U-U-U-C-U-C-U-A-U-C-Gp	—
<i>c</i>	C-U-A-U-A-U-U-C-A-U-A-U-C-U-U-Gp	—
<i>d</i>	A-A-A-U-U-U-A-C-C-A-A-U-C-A-A-U-U-Gp	—
<i>e</i>	C-A-A-C-C-C-A-A-A-U-C-C-A-U-U-U-U-Gp	—
<i>f</i>	U-U-C-C-C-A-C-A-U-U-C-C-C-U-C-A-Gp	—
<i>g</i>	A-A-U-A-A-U-A-A-A-U-A-Gp	—
<i>j</i>	A-A-U-U-U-C-C-U-U-C-C-A-C-C-U-Gp	—
<i>l</i>	U-U-U-U-A-C-A-A-A-C-C-A-Gp	—
<i>n</i>	A-A-C-U-A-U-U-U-U-C-C-A-C-Gp	—
<i>o</i>	U-A-C-U-A-A-C-A-U-U-A-A-Gp	—
<i>p</i>	C-A-C-U-A-U-C-C-C-U-C-C-C-C-Gp	—

Therefore oligonucleotide *u* has the sequence A-C-A-A-C-A-A-Gp.

Several of the above results illustrate the usefulness of the enzyme spleen acid ribonuclease in determining sequences of T₁-oligonucleotides. This is a fairly-low-activity enzyme whose specificity is not clear: in the results obtained here, -A-U- and -C-U- bonds seem the most susceptible, with the greatest activity against the former. (There is also a tendency to remove both 5'- and 3'-terminal residues under very strong digestion conditions, but this may be due to slight contamination by exonucleases.) The specificity for -C-U- bonds was found particularly useful in determining the sequences of long pyrimidine tracts, for example, in ribonuclease U₂ products of T₁-oligonucleotides.

The sequences of all the T₁-oligonucleotides dealt with in this paper have been listed in Table 3 and, where known, their positions in the R17 genome have been indicated. Oligonucleotides *a*, *h* and *k* have sequences which can code for amino acid sequences in the R17 coat protein (Weber & Konigsberg, 1967; Weber, 1967) which are shown in Fig. 1. Larger fragments of R17 RNA from the coat-protein cistron containing T₁-oligonucleotides *a* and *h* have been reported elsewhere (Adams *et al.* 1969; Jeppesen, Nichols, Sanger & Barrell, 1970a), and sequence *k* at the beginning of the coat protein cistron has been deduced by Robinson, Frist & Kaesberg (1969) from studies with a strain of R17 having an amber mutation at position 6 of the coat protein.

Certain other sequences are contained in, or overlap, the three ribosome binding sites determined by Steitz (1969): *s* and the first eleven residues of *k* are found in the coat-protein initiation site; *u* and the last seven residues of *g* are found in the synthetase initiation site, and the whole of *q* was found by Nichols (1970) in a fragment of R17 RNA extending from the termination region of the coat protein cistron and overlapping the synthetase ribosome binding site. Oligonucleotide *r* was found in a 74-long fragment from the 5'-end of R17 RNA (Adams & Cory, 1970), and both oligonucleotides *r* and *m* occur in a larger fragment from the 5'-end of approximately 130 base residues (U. Rensing & B. G. Barrell, unpublished work). The eight-long 3'-sequence of oligonucleotide *m* is the same as the beginning of the ribosome binding site at the start of the A-protein cistron, the first of the three cistrons in order along the R17 genome (Jeppesen *et al.* 1970b), and therefore most probably overlaps it, locating the beginning of the A-protein cistron about 130 residues from the 5'-end of the molecule.

This finding that certain of the spots on the R17 RNA T₁ 'fingerprint' come from genetically defined regions of the R17 genome has made it

- (a) (G)AA UUA ACU AUU CCA AUU UUC G
... Glu - Leu - Thr - Ile - Pro - Ile - Phe - Ala ...
89 95
- (h) (G)C AAA UAC ACC AUU AAA G
... Arg - Lys - Tyr - Thr - Ile - Lys - Val ...
56 62
- (k) (G)CU UCU AAC UUU ACU CAG
f Met-Ala - Ser - Asn - Phe - Thr - Gln ...
1 6

Fig. 1. Amino acid sequences from the R17 bacteriophage coat protein coded by nucleotide sequences *a*, *h* and *k*.

possible to establish chemically the order of the three cistrons along the R17 genome by 'fingerprinting' two fragments produced from R17 RNA by the action of ribonuclease IV, which has a single primary cleavage point, and locating the 'marker' T₁-oligonucleotides in the two fragments (Jeppesen *et al.* 1970b).

The identification of large ribonuclease T₁ products forms the basis for further studies on sequences by partial digestion. Thus the 'fingerprinting' method has been used on larger products obtained by polyacrylamide gel electrophoresis (Adams *et al.* 1969) to test them for the presence of sequences from specific regions in the R17 genome, e.g. fragments which could code for amino acid sequences in the phage coat protein, and which are hence derived from the coat protein cistron (see Adams *et al.* 1969; Nichols, 1970; Jeppesen *et al.* 1970a).

The sensitivity of this 'fingerprinting' system for high-molecular-weight RNA species has been indicated by preliminary experiments with RNA from the closely related bacteriophages f₂ and MS2, which have recognizably different patterns of spots from R17 (Nichols & Robertson, 1971; J. L. Nichols & H. D. Robertson, unpublished work). These can, in certain cases, be related to simple differences in base sequences between the phages.

I thank Dr F. Sanger for his encouragement of this work and for his advice in the preparation of the manuscript, and Dr G. G. Brownlee and Mr B. G. Barrell for their helpful discussions. I also thank Dr G. Bernardi for a gift of spleen acid ribonuclease. I was supported during this work by a Scholarship from the Medical Research Council of Great Britain.

REFERENCES

- Adams, J. M. & Cory, S. (1970). *Nature, Lond.*, **227**, 570.
Adams, J. M., Jeppesen, P. G. N., Sanger, F. & Barrell, B. G. (1969). *Nature, Lond.*, **223**, 1009.
Anderson, J. S., Dahlberg, J. E., Bretscher, M. S., Revel, M. & Clark, B. F. C. (1967). *Nature, Lond.*, **216**, 1072.
Arima, T., Uchida, T. & Egami, F. (1968). *Biochem. J.* **106**, 609.

- Bernardi, A. & Bernardi, G. (1966). *Biochim. biophys. Acta*, **129**, 23.
- Brownlee, G. G. (1971). *Nature New Biol., Lond.*, **229**, 147.
- Brownlee, G. G. & Sanger, F. (1967). *J. molec. Biol.* **23**, 337.
- Brownlee, G. G. & Sanger, F. (1969). *Eur. J. Biochem.* **11**, 395.
- Brownlee, G. G., Sanger, F. & Barrell, B. G. (1968). *J. molec. Biol.* **34**, 379.
- Dahlberg, J. E. (1968). *Nature, Lond.*, **220**, 548.
- Gesteland, R. F. & Boedtker, H. (1964). *J. molec. Biol.* **8**, 496.
- Gilham, P. T. (1962). *J. Am. chem. Soc.* **84**, 687.
- Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R. & Zamir, A. (1965). *Science, N.Y.*, **147**, 1462.
- Jeppesen, P. G. N., Nichols, J. L., Sanger, F. & Barrell, B. G. (1970a). *Cold Spring Harb. Symp. quant. Biol.* **35**, 13.
- Jeppesen, P. G. N., Steitz, J. A., Gesteland, R. F. & Spahr, P. F. (1970b). *Nature, Lond.*, **226**, 230.
- Nichols, J. L. (1970). *Nature, Lond.*, **225**, 147.
- Nichols, J. L. & Robertson, H. D. (1971). *Biochim. biophys. Acta*, **228**, 676.
- Robinson, W. E., Frist, R. H. & Kaesberg, P. (1969). *Science, N.Y.*, **166**, 1291.
- Sanger, F., Brownlee, G. G. & Barrell, B. G. (1965). *J. molec. Biol.* **13**, 373.
- Steitz, J. A. (1969). *Nature, Lond.*, **224**, 957.
- Weber, K. (1967). *Biochemistry, Easton*, **6**, 3144.
- Weber, K. & Konigsberg, W. (1967). *J. biol. Chem.* **242**, 3563.