A Sequence of Seventy-Three Nucleotides from the Coliphage R17 Genome

By ULRICH F. E. RENSING

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

(Received 30 August 1972)

1. A sequence of 73 nucleotides of the RNA genome from coliphage R17 was determined. It can be read through in only one translational frame. The fragment is not part of the coat-protein cistron (Min Jou *et al.*, 1972), nor does it come from the untranslated sequences described previously (Steitz, 1969; Nichols, 1970; Cory *et al.*, 1970; de Wachter *et al.*, 1971; Contreras *et al.*, 1971; Cory *et al.*, 1972). It contains two sequences of 23 and 24 nucleotides, 22 of which are identical. This kind of reiteration is the first one found in bacteriophage nucleic acid.
2. Improved conditions were found and tested for blocking oligonucleotides with carbodi-imide and cleaving by ribonuclease A at cytidylate residues.
3. A synthetic medium is described which allows labelling *in vivo* with ^{32}P to give specific radioactivities higher than those obtained in the procedures used previously.

Nucleotide-sequence studies of the single-stranded RNA genome of the coliphages R17, MS2, Q β and f2 have been initiated to obtain detailed knowledge of the main functions that these RNA species fulfil. They serve as a template for replication, as a messenger in protein synthesis and are the core of an icosahedral particle with protein molecules assembled around them to constitute the infectious virus (for review see Stavits & August, 1970).

The first nucleic acid sequence that could be related by the genetic code to a known amino acid sequence (Weber, 1967) was from the coat-protein cistron of coliphage R17 RNA (Adams *et al.*, 1969). Since then the complete primary structure of the corresponding gene from coliphage MS2 RNA has been described (Min Jou *et al.*, 1972). However, the main part of the structure, some 2500 nucleotides, remains to be investigated. The completed analysis will at the same time make it possible to deduce the as yet unknown amino acid sequences of the two other proteins that the coliphage RNA codes for: the A (or maturation) protein and the coliphage-coded protein component of the 'replicase complex' (August, 1969; Kondo *et al.*, 1970; Kamen, 1970; Blumenthal *et al.*, 1972). This will also permit the description of the complete primary structure of the infectious virus. The sequence of 73 nucleotides reported here is part of one of the two unknown cistrons, or perhaps of their termination regions. The sequence may also be significant for other related RNA coliphages. Up to now only two out of 280 bases have been found to be different at the termini plus intercistronic spaces (Steitz, 1969; Nichols, 1970; Gupta *et al.*, 1970; Adams & Cory, 1970; Nichols & Robertson, 1971; Ling, 1971; de Wachter *et al.*, 1971; Contreras *et al.*, 1971; Cory *et al.*, 1972; U. F. E. Rensing & B. G. Barrell, unpublished work) and nine out of 177 bases

different in the known parts of the coat-protein cistron (Sanger, 1971; Min Jou *et al.*, 1972; and references therein).

During the course of these studies it became desirable to improve the yields of the labelled RNA *in vivo* and of the modification reaction with the water-soluble carbodi-imide reagent to allow specific cleavage at cytidylate residues. The conditions chosen for these operations are reported below.

Experimental and Results

Preparation of ^{32}P -labelled coliphage R17 RNA

Medium. This contained 0.5g of NaCl, 8.0g of KCl, 1.1g of NH_4Cl , 12.1g of Trizma base (Sigma Chemical Co.), 0.4g of $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$, 0.05g of $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$, and 0.8g of sodium pyruvate in 900ml of water. This saline solution A was adapted from that of Lindqvist & Sinsheimer (1967). It was adjusted to pH 8.5 with conc. HCl and then autoclaved. The following components otherwise used for tissue culture were then added (all from Bio-Cult Laboratories, Glasgow, U.K.) to 200ml of solution A: 10ml of 'MEM non-essential amino acids (100X, BCL 404C)', 10ml of 'Dulbecco's MEM amino acids (100X, BCL 420a)', 1.0ml of L-glutamine (200mM), 2ml of 20% (w/v) glucose and 0.15–0.3ml of 0.25M-phosphate buffer, pH 7.2.

Growth of bacteria and coliphage. To 223ml of the medium described, 25ml of *Escherichia coli* 526 in stationary phase was added. The *E. coli* used had been re-isolated from a single colony, grown on the same medium in a closed bottle without agitation, and assayed for infectibility with a bacteriophage stock of known titre, which was later used for infection.

The cells were then allowed to grow in a Roux bottle at 37°C under vigorous aeration to a cell density of about 1.5×10^8 /ml. Foaming was prevented by the addition of 0.02 ml of sterilized polypropylene glycol monolaurate (K & K Laboratories, Inc., Plainview, N.Y., U.S.A.). At this cell density 10^{12} plaque-forming units of coliphage R17, assayed on the same cells, and 2 ml of 1 M-CaCl₂ were added. After 15 min without aeration 25 mCi of carrier-free [³²P]orthophosphate (The Radiochemical Centre, Amersham, Bucks., U.K.) was injected with a syringe. Aeration was resumed and the cells were kept at 37°C for 5–8 h. Lysis was completed with 10 mg of lysozyme, 0.025 mol of EDTA and 5 drops of chloroform according to the method of Anderson *et al.* (1967). In most preparations the procedure described was carried out in duplicate, i.e. two 250 ml cultures were grown and 25 mCi of ³²P was added to each.

Purification of the coliphage. The coliphage was precipitated with 75 g of (NH₄)₂SO₄/250 ml of medium and by adding 100 ml of uninfected *E. coli* suspension grown on the same medium, to stationary phase, containing 40 g of (NH₄)₂SO₄. These suspensions were kept overnight at 0°C and then centrifuged for 30 min at 23000g (MSE High Speed centrifuge, angle rotor no. 69179, r_{min} . 14.2 cm) and 4°C. The combined pellets were suspended in 40–50 ml of 0.1 M-sodium borate buffer, pH 8.8, and homogenized. They were frozen and thawed three times and made 0.05 M in CaCl₂ and 0.01 M in MgCl₂. Deoxyribonuclease II (50 µg) and 10 mg of micrococcal nuclease (crude enzymes; Worthington Biochemical Corp., Freehold, N.J., U.S.A.) were added and the mixture was incubated for 3–4 h at 37°C with occasional mixing.

After this step the standard procedure of low-speed (15 min at 5°C, 20000g, r_{min} . 10.7 cm in MSE rotor 69181) and high-speed (5°C, 3.5 h, 133000g, r_{av} . 5.9 cm) centrifugations were adopted. The coliphage was resuspended and then dissolved in 5 ml of 0.15 M-NaCl–0.015 M-sodium citrate, pH 7, containing 1 mM-MgCl₂ and incubated with 200 µg of deoxyribonuclease I (Worthington; 'ribonuclease-free, DPFF') for 90 min at 37°C. In conventional procedures a CsCl-density-gradient centrifugation is used as the last step in bacteriophage purification, but it gives good recovery only when the pellet of a first centrifugation is re-banded. A glass bead column (2 cm diam., 80 cm length) was therefore used for purification instead. This technique has been described by Gschwender *et al.* (1969) in detail. For eluent 0.15 M-NaCl–0.015 M-sodium citrate, adjusted to pH 8.4 with Tris base and containing 1 mM-EDTA, was used. Fig. 1 shows an analytical CsCl-density-gradient profile from a coliphage preparation obtained in this way, and a sucrose-gradient centrifugation of its RNA.

With the above procedure 2.6 mg of RNA with a

specific radioactivity of 1.25 mCi/mg was obtained as an average in five preparations. This compares well with 2.5 mg at 0.6 mCi/mg reported by Gesteland & Spahr (1970), who used 50 mCi of [³²P]P_i in a comparable volume, or 2.1 mg at 0.55 mCi/mg which I obtained as an average in six preparations by using the growth method described by Steitz (1969) and purification by the method of Anderson *et al.* (1967). Higher specific radioactivities may be obtained by infecting at higher cell densities. I kept the conditions for growth and early infection identical, however, to avoid a decrease in phage yield.

The medium described here has also resulted in higher incorporation of radioactivity into DNA when used for the bacteriophages fd, M13, f1 and φX174 (V. Ling, personal communication).

A medium in which the amino acids were replaced by casein hydrolysate or Bactotryptone and peptone (Difco Laboratories, Detroit, Mich., U.S.A.) did not give comparable results, although the P_i was precipitated before use, indicating a high content of organic phosphorus. The RNA preparation of high specific radioactivity (about 1.5 mCi/mg) did not show noticeable breakdown, as judged by electrophoresis of purified bands in a urea-containing acidic gel (as described in Plate 1) after 2 weeks' storage at –20°C as freeze-dried material in the presence of 100 µg of unlabelled carrier yeast tRNA in a sealed tube. Moreover, the RNA was active in the ribosome-binding system and gave the typical 'fingerprints' of ribosome binding sites (M. Yoshida, personal communication).

Methods of sequence analysis

The conditions for enzymic digestions, the 'fingerprinting' of fragments, radioautography, elution of oligonucleotides and their characterization by further enzymic digestion, were all carried out by the methods described in detail by Sanger and his colleagues (Sanger *et al.*, 1965; Brownlee & Sanger, 1967; Brownlee *et al.*, 1968; Brownlee & Sanger, 1969; Barrell, 1971; Brownlee, 1971). Conditions for limited digestion of the fragments were as follows.

Ribonuclease T₁ (Sankyo Co. Ltd., Tokyo, Japan). Digestion was for 20 min at 37°C in 5–7 µl of 0.015 M-Tris–HCl (pH 7.6)–0.007 M-EDTA with 0.08 µg of enzyme, in the presence of 50–100 µg of RNA.

Ribonuclease A (Worthington). Digestion was for 15–30 min at 0°C in 5 µl of 0.2 M-Tris–HCl (pH 7.6)–0.02 M-MgCl₂ with 0.02 µg of enzyme, in the presence of 50–100 µg of RNA.

Snake-venom phosphodiesterase. This enzyme (Worthington) was further purified by the procedure of Sulkowski & Laskowski (1971) and Laskowski (1966). The oligonucleotides were first dephosphorylated with 10–20 µg of bacterial alkaline phosphatase in 10–20 µl of 0.1 M-Tris–HCl (pH 8.8)–0.01 M-

Fig. 1. Density-gradient centrifugation of coliphage R17(a) and its RNA (b)

(a) ^{32}P -labelled coliphage R17 was prepared according to the procedure described in the Experimental and Results section. A $5\mu\text{l}$ portion was layered on top of 5 ml of CsCl solution (0.62 g/ml) containing 0.1 M- NaCl –0.01 M-sodium citrate at pH 8. It was then centrifuged in a swinging bucket rotor (Spinco SW 50L) for 19 h at 6°C and $100\,000g$ (r_{av} , 7.3 cm). Fractions (0.2 ml) were collected from the bottom of the tube directly into scintillation vials containing 3 ml of Bray's (1960) scintillator. The radioactivity was determined in a Unilux II liquid-scintillation counter (Nuclear–Chicago). (b) A $10\mu\text{l}$ sample of

MgCl_2 , in the presence of 50–100 μg of yeast tRNA for 2 h at 37°C . They were then reisolated by one-dimensional chromatography on DEAE-cellulose or polyethyleneimine thin-layer plates. (The method of chromatography was identical with the one described in Plate 2.) Limited digestion with diesterase was in $5\mu\text{l}$ of 0.1 M- Tris-HCl (pH 8.8)–0.01 M- MgCl_2 with 5 μg of enzyme and 50 μg of unlabelled yeast sRNA for 20 and 40 min at 37°C . The products were separated by the mapping procedure of Brownlee & Sanger (1969) with a hydrolysed 3% (w/v) solution of yeast RNA (BDH Chemicals Ltd., Poole, Dorset, U.K.) in 7 M-urea, or by a standard 'fingerprint' on paper (Brownlee, 1971; Barrell, 1971). The products were identified by their position and by hydrolysis with ribonuclease A or alkali. Details of this procedure, which was adapted from techniques used by Szekeley & Sanger (1969) and Ling (1972) for depurination products, are described by Rensing & Shoenmakers (1973).

Spleen phosphodiesterase (Worthington). Digestion was for 20, 40 and 60 min at 37°C in 7–10 μl of 0.05 M-ammonium acetate–0.006 M-EDTA, pH 5.6, with an enzyme/RNA ratio of 1:5 or 1:10 (w/w). The products were separated and identified as described above for snake-venom phosphodiesterase.

Extraction of RNA from polyacrylamide gel. This was done essentially as described by Cory *et al.* (1972). The use of 4% (w/v) potassium acetate, pH 5.2, as buffer for the second and third washes of the phenol caused some shrinkage of the polyacrylamide and resulted in better yields.

Blocking with carbodi-imide reagent and subsequent digestion with ribonuclease A

Gilham (1962) and Lee *et al.* (1965) have used the water-soluble carbodi-imide reagent [*N*-cyclohexyl-*N'*-(β -morpholinyl-4-ethyl)carbodi-imide methyl-toluene-*p*-sulphonate; Fluka, Buchs, Switzerland] to modify guanosine and uridine residues and to prevent ribonuclease A digestion at uridylylate residues, allowing cleavage to occur specifically at cytidylate residues. However, neither these reaction conditions nor those described by Adams *et al.* (1969) gave a quantitative reaction and they resulted also in

^{32}P -labelled coliphage R17 RNA, prepared as described above, was layered on top of a 5–20% (w/v) sucrose gradient (5 ml) containing 0.15 M- NaCl –0.015 M-sodium citrate (pH 7.0)–0.002 M-EDTA (neutralized). It was then centrifuged in a swinging bucket rotor (Spinco SW 50L) for 4 h at 6°C and $130\,000g$ (r_{av} , 7.3 cm). Collection and counting were performed as described in (a).

Fig. 2. Ribonuclease A digestion of carbodi-imide-blocked nucleotides

The diagram shows the separation by electrophoresis in 0.5% pyridine–5% (v/v) acetic acid at pH 3.5 on Whatman 3MM paper (40V/cm for 3h) at about 15°C. Radioactive nucleotides were obtained from a complete ribonuclease T₁ digest of coliphage R17 RNA fragment separated in two dimensions by using homochromatography. Their sequences were deduced from their position and by analysing a portion with ribonuclease A before blocking with the carbodi-imide reagent. The conditions used for the blocking and subsequent digestion with ribonuclease A are described in the Experimental and Results section. The sequences of the nucleotides shown were confirmed by alkaline hydrolysis. The boxes indicated were cut out and their radioactivity was determined by liquid-scintillation counting in 0.4% (w/v) 2,5-bis-(5-t-butylbenzoxazol-2-yl)thiophen (Ciba Ltd.) in toluene as average c.p.m. over 40 min with 50 c.p.m. background count subtracted. The nucleotides represented by broken lines did not contain enough radioactivity to permit their analysis by alkaline hydrolysis: they are considered to be contaminants from the ribonuclease T₁ 'fingerprint'. They make it possible, however, to assess approximately the detection threshold, which is less than 2% of the total radioactivity, in A-Gp. A dot over a nucleotide (Ñp) denotes a modified residue. The positions of the mononucleotides are indicated at the top right. The results are presented in the form of a

cleavage at uridylyte residues, thus causing loss of radioactivity.

The reaction conditions finally chosen were as follows. The oligonucleotide was incubated in 20 µl of a solution containing 50–100 mg of carbodi-imide reagent/ml, 0.05 M-Tris-HCl and 0.005 M-EDTA at pH 7.4 and 37°C overnight. (Incubation at 60°C for 4h also gave satisfactory results.) A portion (10 µl) of a solution of ribonuclease A (2 mg/ml of 0.005 M-Tris-HCl, pH 7.4) was then added directly and the mixture was allowed to digest at 37°C for 2 h. A separation of some oligonucleotides treated in this way is shown in Fig. 2. The reaction with uridylyte went to over 90% completion; that with guanylate varied from 70 to 90%. These quantitative results are listed in Table 1. The corresponding values obtained from a reaction at pH 8.0 under otherwise identical conditions were about 50% for both nucleotides, and those at pH 8.5 were about 30% for both nucleotides (results are not shown). The reaction conditions described in detail here have been used in this laboratory over the last 2 years by several different investigators with satisfactory results (see Brownlee, 1971).

An improved method, with a borate buffer, has also been found by L. Weith (personal communication).

Limited enzymic digestion and purification of fragments

A limited ribonuclease T₁ digestion of ribosomal and viral RNA yields specific fragments (within a lower-molecular-weight range), many of which can be resolved by electrophoresis in polyacrylamide gels (Gould *et al.*, 1969). The conditions described by Adams *et al.* (1969) have been adopted to isolate several fragments from coliphage R17 RNA, with a minor modification: the digestion was performed with an enzyme/RNA ratio of 1:1000 (w/w) for 13 h (see the legend to Plate 1). The bands obtained from this first fractionation were identified by 'fingerprinting' and subjected to a second electrophoresis in a polyacrylamide gel containing 6M-urea in 0.025M-potassium citrate, pH 3.6 (cf. Fiers *et al.*, 1971). It has previously been reported that a fractionation at an acid pH value in urea will separate RNA molecules with a hidden break that are held together by the secondary structure in a fractionation at a neutral pH value (Gould, 1967; Nichols, 1970; Jeppesen *et al.*, 1972; Cory *et al.*, 1972). To complete denaturation before application the samples were taken up in 40 µl of 6M-urea–0.01 M-potassium citrate and kept for 5 min at 100°C, and were then quickly chilled. The separation is shown in Plate 1: each band gives rise to

diagram because the photograph of the radioautograph does not show the spots of lower intensity.

EXPLANATION OF PLATE I

Isolation of fragments from coliphage R17 RNA

The radioautographs show two separations by electrophoresis in polyacrylamide gels. Arrow 1 denotes a fractionation according to the procedure of Adams *et al.* (1969) in a 12.5% polyacrylamide gel at pH 8.4. A sample (2.8 mg) of coliphage R17 RNA labelled with 7.9×10^9 d.p.m. of ^{32}P was digested with 14 units of ribonuclease T_1 for 13 h at 0°C in 0.15 ml of 0.02 M-Tris-acetate (pH 7.5)–0.04 M- MgCl_2 and loaded directly on this gel. The bands indicated were then extracted, denatured, and re-run in an 8% (w/v) polyacrylamide gel in the direction of arrow 2 for 36 h at 4°C and 10 V/cm. 0.025 M-Potassium citrate (pH 3.6), containing 6 M-urea (Aristar grade, BDH Ltd., Poole, Dorset, U.K.), was used as the buffer. (The dimensions of the gel and the ratio of cross-linker added were the same as for the first gel.) The origin of the second fractionation was on the level corresponding to the start of arrow 2. Bands 12A, 12B, 16 and the slowest heavy ones from 15 and 14 were pure (see Plates 2 and 3) and served for quantitative sequence studies.

(a)

(b)

EXPLANATION OF PLATE 2

Radioautographs of the two-dimensional fractionations of the complete digestion products from band 16

The RNA fragment from the purified gel band 16 (see Fig. 2) was divided into two parts and digested with ribonucleases T_1 (a) and ribonuclease A (b), according to established procedures (Barrell, 1971; Brownlee, 1971). The end products obtained were subjected to electrophoresis at pH 3.5 on cellulose acetate as the first dimension (1). The ribonuclease T_1 products were then separated by ascending homochromatography (2) with a hydrolysed, dialysed 3% (w/v) RNA solution on DEAE-cellulose thin-layer plates (Brownlee & Sanger, 1969). The pancreatic ribonuclease digestion products (b) were fractionated on DEAE-paper in 7% formic acid as the second dimension (2). The deduction of the sequences is described in the Experimental and Results section. B denotes the position of the blue dye marker (Xylene Cyanol FF).

EXPLANATION OF PLATE 3

Radioautograph of the two-dimensional fractionations of the complete digestion products from band 14

(a) Ribonuclease T₁ digestion products. (b) Ribonuclease A digestion products. The difference in migration on homochromatography of the ribonuclease T₁ products as compared with Plate 2 is caused by the use of a different batch of yeast RNA (BDH Ltd.) for homochromatography, by washing the thin-layer plate in an ethanol bath after transfer of the oligonucleotides from the cellulose acetate strip before chromatography and by extending the time of chromatography. A wick attached by a clip to the top of the plate, consisting of four 2cm-wide strips of Whatman 3MM paper, was used to take up the solvent front. The conditions were otherwise identical with those described in Plate 2.

Table 1. *Quantitative results of the carbodi-imide reaction*

The yields of the ribonuclease A digestion products shown in Fig. 2 were obtained by measuring the radioactivity of the paper containing the nucleotide, by using a Unilux II scintillation counter (Nuclear-Chicago) and 0.4% (w/v) 2,5-bis-(5-t-butylbenzoxazol-2-yl)thiophen (BBOT; Ciba Ltd., Basle, Switzerland) in toluene as scintillator. The radioactivity per nucleotide was calculated from the structures given in Fig. 2. Modified nucleotides are indicated as Gp, Up.

Nucleotide	Total radioactivity in total Gp+Up (c.p.m.)	Percentage of total radioactivity in modified nucleotides	
		Gp	Up
U-U-Gp	1015	32	63
U-Gp	1641	46	47
A-U-Gp	1060	38	50
A-Gp	1660	91	—
A-A-Gp	452	87	—
C(U,C)C-Gp	379	50.5	49.5

several sub-bands. Most of these are not related to the main component, as judged by 'fingerprinting'. This was confirmed by the analysis of partial enzymic digests on bands that had not been purified in a second gel: about half the products obtained were not part of the main component (results are not given).

In agreement with others (Richards *et al.*, 1965; Loening, 1967; Bishop *et al.*, 1967), it can be calculated that the mobility y in the 12.5% (w/v) polyacrylamide gel at pH 8.4 is usually an approximated linear function of the logarithm of the chain length x , according to the equation $y = -k_1 \times \log_{10} x + k_2$, where k_1 and k_2 are constants. The mobilities of the following bands fit this equation: '8' ($x = 128$), which overlaps the 5' end with the A-protein cistron ribosome binding site (de Wachter *et al.*, 1971; Adams & Cory, 1970; Sanger, 1971); '14' ($x = 82$); '16' ($x = 69$) (this paper); '25' ($x = 50$) (U. F. E. Rensing & J. G. G. Schoenmakers, unpublished work); '27' ($x = 35$) (Cory *et al.*, 1972); '20', '20x', '21' ($57 < x < 61$) (Adams *et al.*, 1969; Jeppesen *et al.*, 1972).

Band 23, which contains a sequence of 43 nucleotides (Nichols, 1970), does not fit the equation as well. It also migrates slower than band 25 with its sequence of 48–50 nucleotides, and as fast as a possible extension of the same sequence in band 23 ($48 < x < 53$). The reason for this is probably that for all the other sequences mentioned a 'hairpin'-like secondary structure can be constructed, in which at least 70% of the nucleotides are involved in base-pairing (see the Discussion section), whereas in band 23 less than 50% can be base-paired.

Sequence analysis of the fragment

When 'mapping' the individual bands from the polyacrylamide gel two were found whose main components were clearly related, i.e. bands 14 and 16

(Plates 2 and 3). Their sequences were established by using standard procedures with the modifications described above. The 'fingerprints' obtained by extensive digestion with ribonucleases A or T_1 are shown in Plates 3 and 4 (for details see the legends). These products were analysed further by specific endonucleolytic and exonucleolytic enzymic digestions, as summarized in Table 2. The sequences of the ribonuclease A products could be deduced from the results obtained by complete ribonuclease T_1 digestion and by limited digestion with spleen phosphodiesterase, followed by extensive digestion with ribonuclease T_1 (results are not shown).

The relative order of the complete digestion products was established by performing limited digestions of the intact fragment with ribonuclease A and T_1 . These operations are summarized in Fig. 2. Owing to a high susceptibility towards nuclease attack at the termini of fragment 14, the complete sequence of this could not be derived unambiguously. A tentative sequence is discussed below.

Discussion

Secondary structure

As can be seen in Fig. 3, the number of cleavages in the tetranucleotide G-G-A-Up (nucleotides 46–49) far exceeds that in G-G-A-G-G-Cp (28–33) in the partial ribonuclease T_1 digests, although the guanylate residues in each of these could be expected to be similarly susceptible to ribonuclease T_1 . It is tempting to relate such differences to the possible secondary structure of the molecule. Several different structures were found to be possible and two of these are presented in Fig. 4. None allowed complete base-pairing of the G-G-A-Up sequence, whereas a full complement existed for G-G-A-G-G-Cp. It may be noted that the

Table 2. Analysis of the complete T_1 ribonuclease digestion products

Enzymic digestion conditions and methods for fractionating and identifying the oligonucleotides were those described by Dr. F. Sanger and collaborators (Brownlee, 1971; Barrell, 1971). The carbodi-imide blocking and digestion of blocked nucleotides was modified (see the text). U and G represent the modified uridylylate and guanylylate residues respectively. The right-hand column gives the sequences as deduced from the results listed in columns 2, 3, 4 and 5. It includes additional results and deductions from partial digestion products obtained with spleen phosphodiesterase (represented by $\rightarrow$) and snake-venom phosphodiesterase ($\leftarrow$). These exonuclease products were identified by their position in the two-dimensional fractionations by homochromatography are those listed in column 5. The conditions for limited exonuclease digestion and their two-dimensional fractionations by homochromatography are described in the Experimental and Results section.

Structure deduced from columns 2-5 and terminal sequences determined independently by additional digestions with exonucleases (arrows)

Complete ribonuclease A digestion products

Oligonucleotide	Complete ribonuclease A digestion products		Partial snake-venom phosphodiesterase digestion products†¶ (apart from those outlined in column 6)	Structure deduced from columns 2-5 and terminal sequences determined independently by additional digestions with exonucleases (arrows)
	Unmodified oligonucleotides*	Carbodi-imide-blocked oligonucleotides†‡		
16a ₁	1 A-Up	(A,Ü)Gp	U-Gp	C-U-C-C-U-A-C-A-U-Gp
	1 A-Cp	Ü-Cp	C-Ap	
	3 Cp	(Ü,A)Cp	(C ₂ ,Ü-C)Ü-Ap*	
	2 Up	(Ü,C)§		
	1 Gp	Cp		
16a ₂	1 A-Cp	Ü-Gp	C-C-Ü-Gp	C-U-C-C-U-A-C-C-U-Gp
	4 Cp	C-Ü-Gp	(C ₂ ,Ü-C)Ü-Ap*	
	3 Up	Ü-Cp		
	1 Gp	(Ü,A)Cp		
		Cp		
16b	1A-A-Cp	(A,Ü)Gp	U-Gp	U-A-A-C-A-U-Gp
	1A-Up	(A,Ü)Cp	U-Ap	
	1Up	C-Ap	C-Ap	
	1 Gp	(Ü,A)Cp§	(U,C,A)Ap	
16c	A-Cp	(Ü,A)Cp§	C-Gp	C-C-U-U-A-C-Gp
	2 Cp	(Ü,A)Cp	C-C-Ü-Ap	
	2 Up	Cp	(C,U)§	
	1 Gp			
	U-Up§			

16d	4 Cp 1 Up 1 Gp 1 A-A-Cp 1 A-Gp 1 Cp 1 Up 1 A-Gp	U-Cp Gp Cp =	(C,U)Gp§	C-C-U-C-C-Gp A-A-C-A-Gp U-C-A-Gp
16e				
16f		A-Gp U-Cp	(C,U)Ap Gp	
16g	2 Cp 1 Up 1 Gp 1 A-Gp Gp 1 Up 1 A-Gp 1 A-Up 1 Gp 2 Up 1 Gp 1 A-Cp	U-Cp Gp Cp		Complete digestion* 1 pC 1 pU 1 pG
16h				
16i				
16k				
16l				
16m				
14n	3 Up 1 Gp A-Cp 1 Up 1 Gp	(U,A)Cp U-Gp A-Cp U-Gp	U-U-Ap C-U-Gp	U-A-Gp A-U-Gp U-U-Gp U-U-A-C-U-Gp
14o				A-C-U-Gp

* Molar yields were obtained by measuring the radioactivity of the paper containing the nucleotide with a Unilux II scintillation counter (Nuclear-Chicago) and the 0.4% scintillator in toluene described in Table 1. Yields are given to the nearest whole number where these are not unity.

† Unless stated otherwise the yields were estimated by visual inspection of the radioautograph; the products underlined once had an estimated molar yield of 2, those underlined twice a molar yield of 3.

‡ The products were analysed by alkaline hydrolysis.

§ The nucleotide was present in less than molar yield.

|| The product was also analysed by reaction with carbodi-imide reagent, followed by digestion with ribonuclease A, separation on Whatman 3MM paper and alkaline hydrolysis.

¶ The products were further analysed by complete ribonuclease A digestion.

Fig. 4. Possible secondary structures

Two secondary structures (*a*, *b*) of estimated similar stability are proposed for the sequence established. The arrows indicate the cleavages by ribonuclease T₁ to give the fragment from gel band 16; the sequence of the extension in gel band 14 is tentative.

second guanylate residue of G-G-A-Up can form another bulge in structures (a) and (b) (Fig. 4) to allow an additional A-U base pair to interact (with an insignificant estimated increase in stability). This guanylate residue would then appear to be exposed. Although the number of partial ribonuclease A digestion products is limited, it may also be noted that no overlap was found for the loop sequence U-U-A-C-.

Tinoco *et al.* (1971) have proposed rules for estimating a possible secondary structure for a RNA molecule of known sequence. From their rules, the free energy of formation from the single-stranded state is estimated to be $-63\,100\text{J/mol}$ for structure (a) and $-80\,400\text{J/mol}$ for structure (b) in Fig. 4. Since the fragments described here were obtained from a digestion with ribonuclease T_1 , which is specific for guanylate residues, the additional rule to maximize

for base-pairing of guanylates was followed. In structure (a) of Fig. 4, for instance, the bulges in the upper part of the stem can be moved so as to give one additional A-U base pair, and similarly in the lower stem of structure (b). If the loss in stability does not exceed three stability units according to Tinoco *et al.* (1971), the base-pairing of guanylate is preferred. Also, if two structures can be estimated to be similarly stable but differ by a guanylate residue in the loop region, the one with a base-paired guanylate residue is preferred.

Structures (a) and (b) utilize a tentative sequence for an extension found in band 14 that is consistent with the limited experimental evidence. The nucleotide sequence G-A-C-U-Gp was found in the fragment from band 14 (see Plate 3) but not in any of the pure partial enzymic digestion products. It might be located at the 3' end of structure (a) to complete a smaller 'hairpin'-like structure. A yet smaller loop can be

formed by base-pairing the -A-C-U-Gp of the 3' terminal sequence to the -C-A-G-U- (Fig. 4a).

The structure (a) discussed here is similar to those proposed for the fragments from the coat-protein cistron and at the termini described previously (for summary see Sanger, 1971; Cory *et al.*, 1972; Min Jou *et al.*, 1972). In addition to the evidence presented above, suggesting that the structures exist under the experimental conditions used, it may be noteworthy that the only fragment with considerably less than 70% base-pairing (band 23) does not fit the linear function between electrophoretic mobility and the logarithm of the chain length (see the Experimental and Results section).

Duplication of sequences

The two decamers 16a₁ and 16a₂ (Table 2) could be considered as reiterated sequences but for one base difference: an A to C change in the nucleotide third from the 3' end. There is no way at present to decide whether or not their similarity is a coincidence. [If they were completely identical the stability of struc-

Fig. 5. A reiterated sequence in band 14

The homologies between the 5' terminal third and the 3' terminal third of the sequence established are shown. One base change and three insertions would make the sequences identical.

ture (a) in Fig. 4 would increase by an estimated difference in free energy of -25000J/mol to a total of -90500J/mol .] However, allowing for three deletions, the homology becomes more extensive, as depicted in Fig. 5. A sequence of 22 bases would thus be duplicated in this 73 nucleotide fragment. It is the first repetitive nucleotide sequence that has been established in a bacteriophage and it represents an example of how sequences may have diverged in the course of evolution, or can be diversified in ontogenesis by using a unit oligonucleotide with just a few alterations.

Location of the fragment in the coliphage R17 genome and translational frames

By using the genetic code, a set of three peptide sequences corresponding to a given nucleic acid sequence can be written. By comparing the RNA sequence from the coliphage R17 genome established in these experiments (Fig. 6) with the known amino acid sequence of the coat protein (Weber, 1967), it is found that the nucleotide sequence is not part of the coat-protein cistron. It also does not compare with the known sequence of the coat-protein cistron of the closely related bacteriophage MS2 (Min Jou *et al.*, 1972), or with any of the known, probably untranslated, sequences at the termini or between the cistrons of RNA from coliphage R17. [The RNA of the RNA coliphages codes for three proteins (for review see Stavits & August, 1970) in the order 5'-A-protein-coat protein-polymerase-3' (Bassel, 1968; Steitz, 1969; Nichols, 1970; Konings *et al.*, 1970; Jeppesen *et al.*, 1970; Hindley *et al.*, 1970).]

Since the 5' ends of RNA from bacteriophages MS2 and R17 have been analysed up to the beginning of the A-protein cistron (Fiers *et al.*, 1971; Sanger,

Fig. 6. Potential coding properties

By using the genetic code, three protein sequences can be derived from the nucleotide sequence of gel band 16, two of which would be terminated and possibly reinitiated.

1971, and references therein), and because the sequence between the cistrons of the coat protein and the polymerase component is also known, it can be concluded that the fragments from bands 16 and 14 must come from either the cistron of the A-protein or the polymerase component, or from their termination regions. The amino acid sequences of these two proteins are not yet known and it is not therefore possible to define an 'untranslated' nucleotide sequence in these regions, although 51 bases have been described at the 3' end of RNA from bacteriophage R17 (Cory *et al.*, 1972) and 44 preceding the coat-protein cistron (Cory *et al.*, 1970), which both contain several potential terminator codons and are both identical with the corresponding sequences of RNA from bacteriophage MS2 (Fiers *et al.*, 1971).

Fig. 6 shows the three possible amino acid sequences, two of which would be terminated owing to the presence of UAA and UAG in two of the three possible translational frames. 'Read-through' could thus occur in only one translational phase and would define one sequence of 25 amino acids. Reinitiation at the AUG triplet after the UAG and UAA triplets would give an unambiguous sequence of 19 amino acids (Fig. 5) if this fragment from the bacteriophage R17 RNA genome is translated at all.

I thank Dr. F. Sanger, in whose laboratory these experiments were carried out, for supporting this work and for his help and suggestions in the preparation of the manuscript. I am grateful to Mr. B. G. Barrell, Dr. G. H. Dixon, Dr. F. Galibert, Dr. J. Sedat and Dr. E. Ziff for discussions and helpful suggestions. I am indebted to Dr. J. G. G. Schoenmakers for his help in the initial stages of the work, and to Dr. M. Yoshida for performing the sedimentation analysis. I also thank Mr. A. R. Coulson for his skilful assistance in some of the experiments. Throughout the course of this work I was supported by a long-term fellowship from the European Molecular Biology Organization. The ribonuclease U₂ used in this investigation was a kind gift from Sankyo Co. Ltd. (Tokyo, Japan).

References

- Adams, J. M. & Cory, S. (1970) *Nature (London)* **227**, 570-574
- Adams, J. M., Jeppesen, P. G. N., Sanger, F. & Barrell, B. G. (1969) *Nature (London)* **223**, 1009-1014
- Anderson, J. S., Dahlberg, J. E., Bretscher, M. S., Revel, M. & Clark, B. F. C. (1967) *Nature (London)* **216**, 1072-1076
- August, J. T. (1969) *Nature (London)* **222**, 121-123
- Barrell, B. G. (1971) *Procedures Nucleic Acid Res.* **2**, 751
- Bassel, B. A. (1968) *Proc. Nat. Acad. Sci. U.S.* **60**, 321-328
- Bishop, D. H. L., Claybrook, J. R. & Spiegeman, S. (1967) *J. Mol. Biol.* **26**, 373-387
- Blumenthal, T., Landers, T. A. & Weber, K. (1972) *Proc. Nat. Acad. Sci. U.S.* **69**, 1313-1317
- Bray, G. A. (1960) *Anal. Biochem.* **1**, 279-285
- Brownlee, G. G. (1971) in *Laboratory Techniques in Biochemistry and Molecular Biology* (Work, T. S. & Work, E., eds.), vol. 3, part 1, p. 196, North-Holland Publishing Co., Amsterdam
- Brownlee, G. G. & Sanger, F. (1967) *J. Mol. Biol.* **23**, 337-353
- Brownlee, G. G. & Sanger, F. (1969) *Eur. J. Biochem.* **11**, 395-399
- Brownlee, G. G., Sanger, F. & Barrell, B. G. (1968) *J. Mol. Biol.* **34**, 379-412
- Contreras, R., Vandenberghe, A., Min Jou, W. & de Wachter, R. (1971) *FEBS Lett.* **18**, 141-144
- Cory, S., Spahr, P. F. & Adams, J. M. (1970) *Cold Spring Harbor Symp. Quant. Biol.* **35**, 1-12
- Cory, S., Adams, J. M., Spahr, P. F. & Rensing, U. F. E. (1972) *J. Mol. Biol.* **63**, 41-56
- de Wachter, R., Merregaert, J., Vandenberghe, A., Contreras, R. & Fiers, W. (1971) *Eur. J. Biochem.* **22**, 400-414
- Fiers, W., Contreras, R., de Wachter, R., Haegeman, G., Merregaert, J., Min Jou, W. & Vandenberghe, A. (1971) *Biochimie* **53**, 495-506
- Gesteland, R. F. & Spahr, P. F. (1970) *Biochem. Biophys. Res. Commun.* **41**, 1267-1272
- Gilham, P. T. (1962) *J. Amer. Chem. Soc.* **84**, 687-688
- Gould, H. J. (1967) *J. Mol. Biol.* **29**, 307-313
- Gould, H. J., Pinder, J. C., Matthews, H. R. & Gordon, A. H. (1969) *Anal. Biochem.* **29**, 1-21
- Gschwender, H. H., Haller, W. & Hofschneider, P. H. (1969) *Biochim. Biophys. Acta* **190**, 460-469
- Gupta, S. L., Chen, J., Schaefer, L., Lengyel, P. & Weissman, S. M. (1970) *Biochem. Biophys. Res. Commun.* **39**, 883-888
- Hindley, J., Staples, D. H., Billeter, M. A. & Weissmann, C. (1970) *Proc. Nat. Acad. Sci. U.S.* **67**, 1180-1187
- Jeppesen, P. G. N., Steitz, J. A., Gesteland, R. F. & Spahr, P. F. (1970) *Nature (London)* **226**, 230-237
- Jeppesen, P. G. N., Barrell, B. G., Sanger, F. & Coulson, A. R. (1972) *Biochem. J.* **128**, 993-1006
- Kamen, R. (1970) *Nature (London)* **228**, 527-533
- Kondo, M., Gallerani, R. & Weissmann, C. (1970) *Nature (London)* **228**, 525-527
- Konings, R. N. H., Ward, R., Francke, B. & Hofschneider, P. H. (1970) *Nature (London)* **226**, 604-607
- Laskowski, M. (1966) *Procedures Nucleic Acid Res.* **1**, 154
- Lee, J. C., Ho, N. W. Y. & Gilham, P. T. (1965) *Biochim. Biophys. Acta* **95**, 503-504
- Lindqvist, B. H. & Sinsheimer, R. L. (1967) *J. Mol. Biol.* **28**, 87-94
- Ling, V. (1971) *Biochem. Biophys. Res. Commun.* **42**, 82-88
- Ling, V. (1972) *J. Mol. Biol.* **64**, 87-102
- Loening, U. (1967) *Biochem. J.* **102**, 251-257
- Min Jou, W., Haegeman, G., Ysebaert, M. & Fiers, W. (1972) *Nature (London)* **237**, 82-88
- Nichols, J. L. (1970) *Nature (London)* **225**, 147-151
- Nichols, J. L. & Robertson, H. D. (1971) *Biochim. Biophys. Acta* **228**, 676-681
- Rensing, U. F. E. & Shoenmakers, J. G. G. (1973) *Eur. J. Biochem.* in the press
- Richards, E. G., Coll, J. A. & Gratzner, W. B. (1965) *Anal. Biochem.* **12**, 452-471

- Sanger, F. (1971) *Biochem. J.* **124**, 833–843
- Sanger, F., Brownlee, G. G. & Barrell, B. G. (1965) *J. Mol. Biol.* **13**, 373–398
- Stavis, R. L. & August, J. T. (1970) *Annu. Rev. Biochem.* **39**, 527–560
- Steitz, J. A. (1969) *Nature (London)* **224**, 957–964
- Sulkowski, E. & Laskowski, M. S. (1971) *Biochim. Biophys. Acta* **240**, 443–447
- Szekely, M. & Sanger, F. (1969) *J. Mol. Biol.* **43**, 607–617
- Tinoco, I., Uhlenbeck, O. C. & Levine, M. D. (1971) *Nature (London)* **230**, 362–366
- Weber, K. (1967) *Biochemistry* **6**, 3144–3154