Biochem. J. (1972) **128**, 993–1006
Printed in Great Britain

Nucleotide Sequences of Two Fragments from the Coat-Protein Cistron of Bacteriophage R17 Ribonucleic Acid

By P. G. N. JEPPESEN,* B. G. BARRELL, F. SANGER and A. R. COULSON
Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

(*Received* 21 *January* 1972)

Bacteriophage R17 RNA was labelled with ^{32}P and was subjected to partial digestion with ribonuclease T_1. The products were fractionated by ionophoresis on polyacrylamide gel. Two fragments were purified and their nucleotide sequences determined by methods involving complete and further partial digestion with ribonucleases A and T_1. Fragment 20 had a sequence that coded for the amino acids in positions 32–53 of the coat protein of the bacteriophage. Fragment 20X, on further purification in 7M-urea, gave rise to two smaller nucleotides whose sequences coded for the amino acids in positions 56–66 and 67–76 of the coat protein. The sequence of the two fragments was such that they could be written in the form of loops stabilized by base-pairing.

The genome of the *Escherichia coli* bacteriophage R17 is a single-stranded linear RNA molecule comprising approx. 3500 nucleotides, and it acts as a messenger coding for the synthesis of three proteins: the coat protein, which is the main protein component of the virus; the A protein or 'maturation' protein, which is also present in the mature virus; and the RNA replicase, which is responsible for copying the viral genome but is not present in the virus particle. The virion components are present in the proportions RNA:coat protein:A protein 1:180:1 (Gussin, 1966; Jeppesen *et al.*, 1970*a*). The amino acid sequence of the coat protein has been determined by Weber (1967), but little is known about the sequences of the other two proteins.

When ^{32}P-labelled bacteriophage R17 RNA was digested with ribonuclease T_1, a number of large oligonucleotides (15–24 residues long) could be separated in a pure form by a two-dimensional system involving ionophoresis on cellulose acetate and thin-layer 'homochromatography' on DEAE-cellulose (Brownlee & Sanger, 1969; Jeppesen, 1971). The sequences of these oligonucleotides were determined and some of them contained sequences that were related by the genetic code to amino acid sequences in the coat protein. Larger degradation products of the RNA could be obtained by partial digestion with ribonuclease T_1 and fractionation by ionophoresis on polyacrylamide gels (Adams *et al.*, 1969) (Plate 1). One of the bands from the fractionation (band 21) contained an RNA fragment of 57 residues whose sequence was determined, and this was shown to code for the amino acids in positions 81–100 of the coat protein.

* Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. 11724, U.S.A.

Most of the other bands have now been investigated in more detail, and two more (bands 20 and 20X) have been found to belong to the coat-protein cistron and their sequences have been determined. A preliminary account of this work has been published (Jeppesen *et al.*, 1970*b*).

Methods

Preparation of the ^{32}P-labelled bacteriophage R17 RNA, partial ribonuclease T_1 digestion, and fractionation of the products by ionophoresis on polyacrylamide gel were as described previously (Adams *et al.*, 1969).

Previously RNA had been eluted from the gel by a method involving ionophoresis on to DEAE-cellulose paper and subsequent elution from the DEAE-cellulose paper with triethylamine carbonate. Although this method was relatively efficient and easy to carry out on a large number of samples simultaneously, the products obtained were frequently found to be degraded. The reason for this is not clear; it may have been due to nuclease contamination in the large volume of buffer used in the reservoir, or in the DEAE-cellulose paper, or to slight alkaline hydrolysis during elution from the DEAE-cellulose paper. The following method, although more tedious, usually yielded intact RNA and has been generally adopted.

The gel (vol. approx. 1ml) was roughly broken up by extruding it through the orifice of a 2ml plastic disposable hypodermic syringe into 2ml of redistilled, water-saturated phenol contained in a 10ml silicone-treated glass homogenizer tube. Then 2ml of 0.1M-NaCl, buffered with 0.04M-tris–HCl, pH8.3,

containing 0.002M-EDTA, was added and the gel was finely homogenized for 1–5 min with a Teflon homogenizer rotated by an electric motor. (All of these procedures were carried out at 0–4°C.) The homogenate was kept at 0°C for about 30min with occasional mixing, after which the phases were separated by brief low-speed centrifugation and the aqueous phase was removed. The phenol/polyacrylamide slurry was then washed twice in a similar manner with 1ml portions of buffer, and the combined aqueous extracts, contained in a silicone-treated 15ml glass centrifuge tube, were precipitated by the addition of 2.5vol. of ethanol and being left at −20°C overnight. The precipitate was then collected by centrifugation at 8000rev./min for 15min at 4°C (MSE High-Speed 18 centrifuge, 8×50 rotor). The phenol, which was added primarily to remove any contaminating nuclease activity that might have been present, was also found to increase the efficiency of extraction from the gel, compared with experiments performed without the addition of phenol, and recovery of radioactivity could be as high as 80–85%. There was often a certain amount of soluble (i.e. low-molecular-weight) polyacrylamide present in gels, which was also extracted by the above procedure and which co-precipitated with the RNA on addition of ethanol. This could be an advantage in that low concentrations of RNA could be precipitated with the polyacrylamide acting as a carrier: however, subsequent operations involving solution of the RNA in small volumes (e.g. for re-running extracts on cellulose acetate strips) were made more difficult because of the gelatinous nature of the dissolved polymer. The following additional acid-precipitation step, which does not co-precipitate polyacrylamide, has been used to purify RNA extracted from gels that contained a large amount of soluble polyacrylamide (as was apparent by a heavy white ethanol precipitate). The first ethanol precipitate was redissolved in 1ml of water at 0°C, and carrier RNA was added if necessary to bring the total RNA concentration to 1.0 E_{260} unit/ml. Then 0.1ml of 10% (v/v) $HClO_4$ was added and the whole was well mixed, left for 10min at 0°C, and centrifuged at 8000rev./min for 15min at 4°C. The supernatant was poured off and the pellet was resuspended in 1.0ml of 2% (w/v) sodium acetate, pH5.0, after which the RNA was reprecipitated with ethanol as described above. Precipitation of ^{32}P-labelled RNA fragments with $HClO_4$ in this manner was estimated at better than 90% efficient and has not been shown to damage the RNA in any detectable way.

In all the two-dimensional fractionation systems used, the first dimension was by ionophoresis on cellulose acetate at pH3.5 in 7M-urea as described by Brownlee & Sanger (1969). In the t.l.c. homochromatography systems the second dimension was either in homomixture *b* (Brownlee & Sanger, 1969) or in homomixture *c*, which was prepared as in the above paper to give a final concentration of 3% (w/v) RNA.

Products of complete ribonuclease T_1 digestion were usually fractionated with homomixture *c*. Products of partial digestion were fractionated with homomixture *b*. After elution the products of partial digestion were divided into two samples. One sample was subjected to complete digestion with ribonuclease T_1 (enzyme:RNA ratio 1:20, 30min at 37°C in 5–10μl of 0.01M-tris–HCl buffer, pH7.4, containing 0.001M-EDTA) and the products were fractionated one-dimensionally by ionophoresis on DEAE-cellulose paper in 7% (v/v) formic acid. They could be frequently identified from their mobility on this system, but this was normally confirmed by further degradation with ribonuclease A and fractionation of the products by ionophoresis on DEAE-cellulose paper at pH3.5. A second sample of the partial ribonuclease T_1 digestion products was treated with ribonuclease A (conditions were as for ribonuclease T_1)

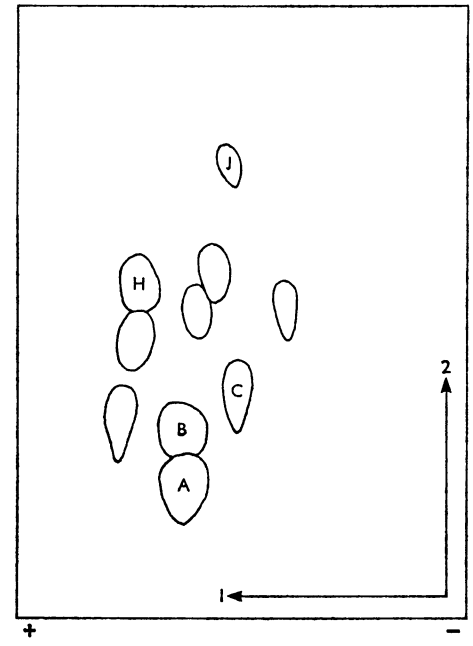


Fig. 1. *Re-fractionation of the material extracted from gel band 20X by t.l.c. homochromatography by using homomixture b*

The ethanol precipitate of the material extracted from the gel was dissolved in approx. 5μl of 7M-urea, pH3.5, and was applied to the cellulose acetate first dimension. The origin for the first dimension is to the right of the figure, as it is not transferred to the t.l.c. plate (Sanger *et al.*, 1965).

Table 1. *Sequence analysis of ribonuclease T_1 oligonucleotides from coat-protein-cistron RNA fragments contained in band 20X*

Spots C and H (Fig. 1) were eluted and digested with ribonuclease T_1 and the products were fractionated by one-dimensional ionophoresis on DEAE-cellulose paper in 7% formic acid. Sequence analyses were performed as described by Jeppesen (1971). The values in parentheses indicate molar yields of oligonucleotides to the nearest whole number, where these are not unity. 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 the molar yield respectively. -U̇- and -Ġ- represent carbodiimide-modified uridylate and guanylate residues respectively.

Oligonucleotide	Ribonuclease A digestion products	CD products (Jeppesen, 1971)	Ribonuclease U_2 digestion products	Deduced sequence
T_1-oligonucleotides from spot C				
h	← (see Table 1 of Jeppesen, 1971) →			C-A-A-A-U-A-C-A-C-C-A-U-U-A-A-A-Gp
1	A-A-Gp Cp(2) Up	U̇-A-A-Ġp Cp	—	C-C-U-A-A-Gp
2	Gp Cp Up	U̇-Cp Ġp	—	U-C-Gp
3	Gp Up	—	—	U-Gp
4	A-Gp	—	—	A-Gp
5	Gp	—	—	Gp
T_1-oligonucleotides from spot H				
6*	A-A-Cp A-Gp Cp(2) Up	U̇-Cp A-Ġp Cp A-A-Cp	(U,C_2)A-Gp (U,C_2)Ap C-A-Ap C-Ap Gp Ap	C-A-A-C-U-C-A-Gp
7*	A-Cp Gp Up	U̇-Ġp A-Cp	—	A-C-U-Gp
8	Gp Up(2)	—	—	U-U-Gp
9	A-Gp Up	—	—	U-A-Gp
10*	Gp Up	—	—	U-Gp
11*	Gp	—	—	Gp

* These oligonucleotides were also found in a ribonuclease T_1 digest of spot J (Fig. 1) from the re-fractionation of band 20X (U-Gp and Gp were in approximate molar yield in this case).

and the products were fractionated by ionophoresis on DEAE-cellulose paper at pH 1.9 (Sanger *et al.*, 1965) and identified by further digestion with ribonuclease T_1.

Other methods of sequence analysis were as described previously (Adams *et al.*, 1969; Jeppesen, 1971).

Results

Plate 1 shows a radioautograph of a preparative ionophoretic separation in polyacrylamide of a partial ribonuclease T_1 digest of bacteriophage R17 RNA. The RNA corresponding to gel bands was extracted and ribonuclease T_1 'fingerprints' were made by using the thin-layer homochromatography system

Table 2. *Analysis of partial ribonuclease T_1 digestion products of spot C from band* 20X (*Fig.* 2)

The spots numbered in Fig. 2 were eluted and analysed as described in Table 1 and by Jeppesen (1971).

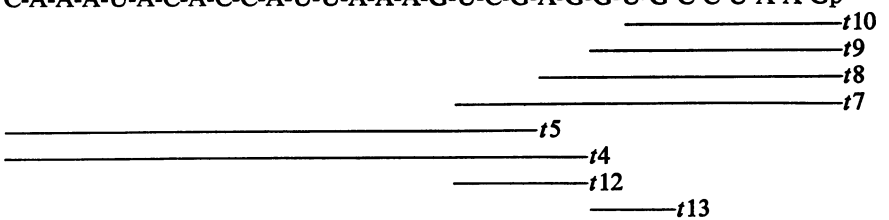
Oligonucleotide (Fig. 2)	Ribonuclease T_1 digestion products	Ribonuclease A digestion products	Deduced sequence
*t*13	U-Gp Gp	G-Up Gp	G-U-Gp
*t*12	U-C-Gp A-Gp	G-A-Gp Cp Up	U-C-G-A-Gp
*t*10	C-C-U-A-A-Gp U-Gp	A-A-Gp G-Cp Cp Up	U-G-C-C-U-A-A-Gp
*t*9	C-C-U-A-A-Gp U-Gp Gp	A-A-Gp G-Up G-Cp Cp Up	G-U-G-C-C-U-A-A-Gp
*t*8	C-C-U-A-A-Gp U-Gp A-Gp Gp	A-G-G-Up A-A-Gp G-Cp Cp Up	A-G-G-U-G-C-C-U-A-A-Gp
*t*7	C-C-U-A-A-Gp U-C-Gp U-Gp A-Gp Gp	G-A-G-G-Up A-A-Gp G-Cp Cp Up	U-C-G-A-G-G-U-G-C-C-U-A-A-Gp

Table 2—*continued*

*t*5	C-A-A-A-U-A-C-A-C-C-A-U-U-A-A-A-Gp U-C-Gp	A-A-A-G-Up A-A-A-Up A-Cp A-Up Gp Cp Up	C-A-A-A-U-A-C-A-C-C-A-U-U-A-A-A-G-U-C-Gp
*t*4	C-A-A-A-U-A-C-A-C-C-A-U-U-A-A-A-Gp U-C-Gp A-Gp	A-A-A-G-Up A-A-A-Up G-A-Gp A-Cp A-Up Cp Up	C-A-A-A-U-A-C-A-C-C-A-U-U-A-A-A-G-U-C-G-A-Gp

Summary of partial digestion products

Deduced sequence ... C-A-A-A-U-A-C-A-C-C-A-U-U-A-A-A-G-U-C-G-A-G-G-U-G-C-C-U-A-A-Gp



(Jeppesen, 1971). The band pattern was found to be reasonably reproducible from one preparative gel to the next, although there were often slight differences in relative mobilities of bands (as characterized by their 'fingerprints'), which were probably due to slight variations in polymerization between gels. However, after initial screening experiments in which each of the bands was 'fingerprinted', the similarities of successive gels made it possible in most cases to recognize particular bands from the gel pattern alone, although this was always confirmed by subsequent 'fingerprinting'.

The 'fingerprints' of the gel bands were considerably simpler than those for the whole bacteriophage R17 RNA, so that more smaller oligonucleotides were obtained in a pure form. Thus, whereas in the whole bacteriophage R17 RNA 'fingerprint' most oligonucleotides smaller than 12 residues were impure, in the isolated gel bands oligonucleotides larger than about eight residues were usually pure. Initially the RNA from spots was subjected to digestion with pancreatic ribonuclease and the products were separated by ionophoresis on DEAE-cellulose paper at pH3.5 (Adams *et al.*, 1969). From these results it was usually possible to determine if any of the products could be derived from the coat-protein cistron (as illustrated for band 20 in Fig. 6) or from other known parts of the RNA, such as the ribosomal binding sites (Steitz, 1969) or the ends of the RNA (Adams & Cory, 1970).

Band 21 had already been shown to be derived from the coat-protein cistron, and its sequence has been described (Adams *et al.*, 1969). On digestion with ribonuclease T_1, band 20X gave rise to oligonucleotide *h*, which had been found to have a nucleotide sequence corresponding to the amino acid sequence in position 57–61 of the coat protein (Jeppesen, 1971). Band 20, however, did not contain any of the larger oligonucleotides that had been identified in the ribonuclease T_1 digest of the whole bacteriophage R17 RNA, but preliminary sequence results on a number of the smaller oligonucleotides suggested that they could be derived from a section of the coat-protein cistron (Fig. 6). The sequences of the RNA fragments contained in these bands were therefore investigated in further detail.

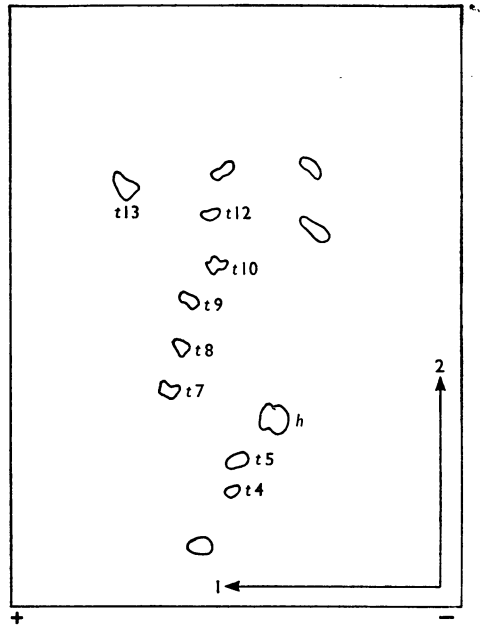


Fig. 2. *Partial ribonuclease T_1 digest of component C from gel band 20X (Fig. 1) fractionated by t.l.c. homochromatography by using homomixture b*

Digestion was with an enzyme:RNA ratio of 1:1000 in 5μl of 0.01M-tris–HCl buffer, pH7.4, containing 0.001M-EDTA, for 15min at 0°C. Analyses of the products are given in Table 2.

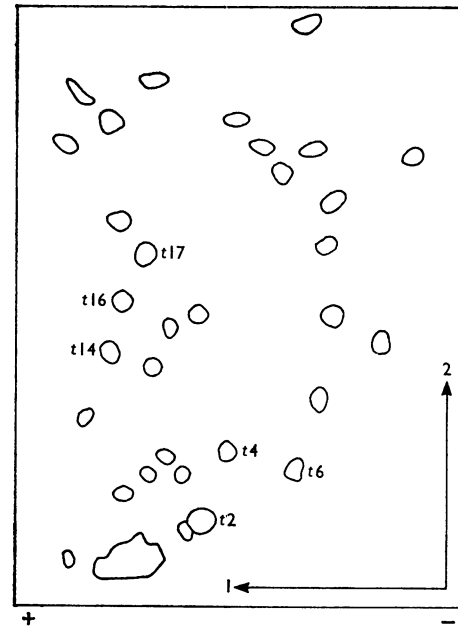
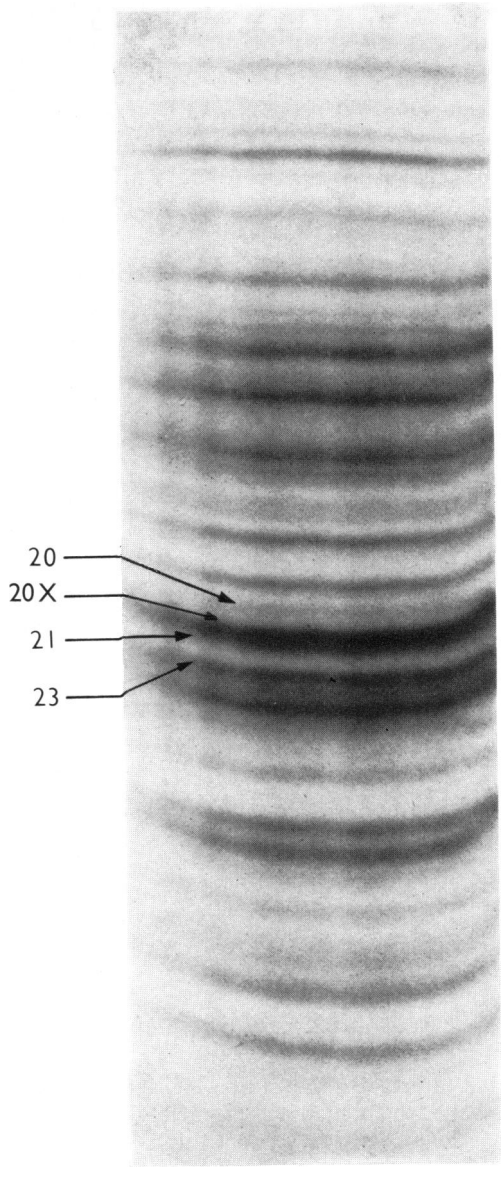


Fig. 3. *Partial ribonuclease T_1 digest of component H from gel band 20X (Fig. 1) fractionated by t.l.c. homochromatography by using homomixture c*

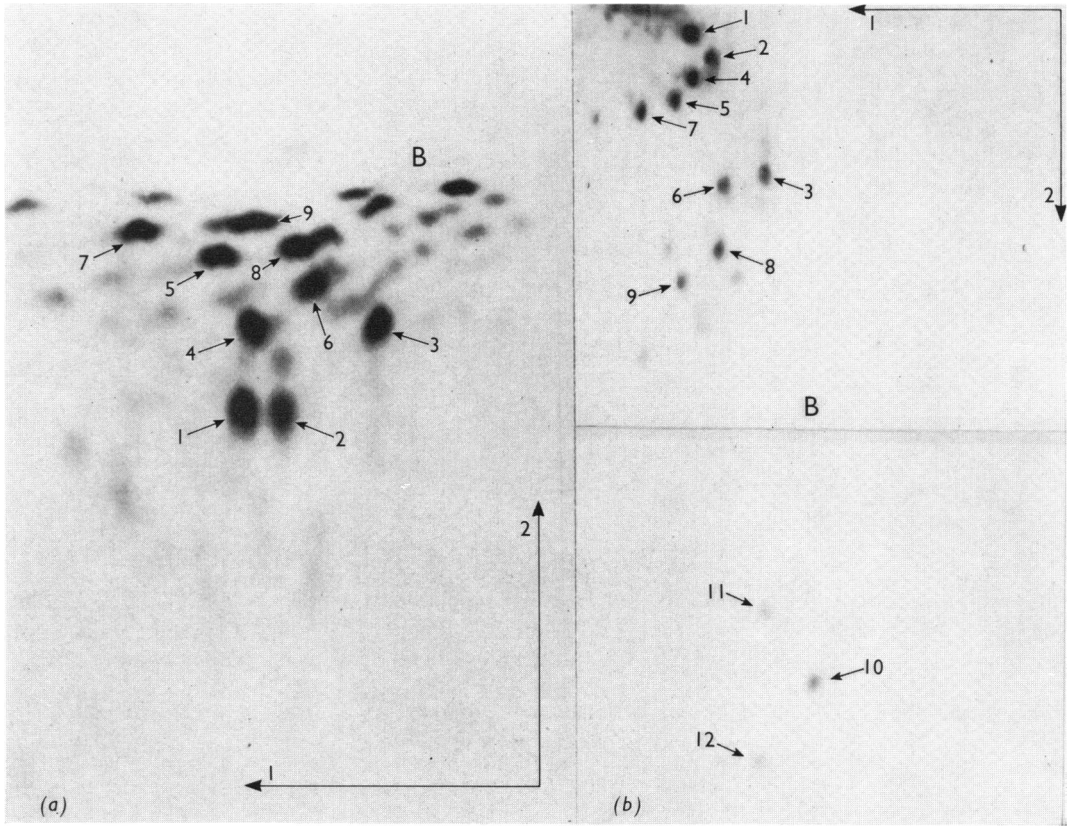
Digestion was done as in Fig. 2 with an enzyme:RNA ratio of 1:5000.



EXPLANATION OF PLATE 1

Radioautograph of a polyacrylamide-gel ionophoresis of a partial ribonuclease T_1 digest of bacteriophage R17 *RNA*

Digestion was done with an enzyme:RNA ratio of 1:500 in 0.04M-tris–acetate–0.02M-magnesium acetate buffer, pH7.4, for 16h at 0°C. Ionophoresis was in a 12.5% polyacrylamide gel as described by Adams *et al.* (1969). The positions of bands 20, 20X, 21 and 23 are indicated. Band 20X is not visible as a separate band, but represents the trailing part of the strong band indicated as 21.



EXPLANATION OF PLATE 2

Radioautographs of two-dimensional fractionations of complete ribonuclease T_1 digests of the material extracted from gel band 20 (*Plate* 1)

Digestion was done with an enzyme: RNA ratio 1:20 in 0.01 M-tris–HCl buffer, pH 7.4, containing 0.001 M-EDTA for 30 min at 37°C. (*a*) T.l.c. homochromatography system by using homomixture *c*; (*b*) ionophoresis on DEAE-cellulose paper in 7% formic acid as the second dimension. Analyses of the spots are given in Table 4. B marks the position of the blue marker (Sanger *et al.*, 1965).

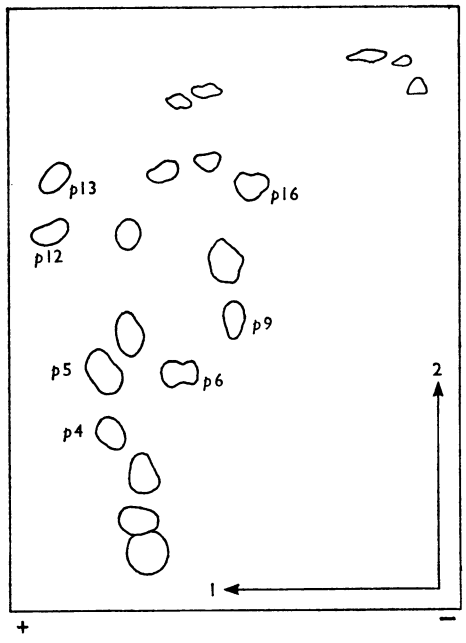


Fig. 4. *Partial ribonuclease A digest of component H from gel band 20X (Fig. 1) fractionated by t.l.c. homochromatography by using homomixture b*

Digestion was done as in Fig. 2 with an enzyme:RNA ratio of 1:1000.

Band 20X

The 'fingerprint' of a ribonuclease T_1 digest of band 20X showed that more than one component was present. In particular it was heavily contaminated with the material from band 21, which gave rise to oligonucleotide *a* (Jeppesen, 1971). To purify the component that gave rise to oligonucleotide *h*, the two-dimensional homochromatography re-fractionation procedure (Brownlee & Sanger, 1969) was used, and the result is shown in Fig. 1. Each of the major spots was eluted. The 'fingerprints' of the material from spots A and B, which corresponded to fragments of around 60 base residues, contained no trace of oligonucleotide *h*. However, when ribonuclease T_1 digests were made of samples of the faster-moving spots and the products were separated by one-dimensional ionophoresis on DEAE-cellulose paper in 7% formic acid, oligonucleotide *h* was found among the products of spot C, characterized by analysis of its ribonuclease A digestion products. Table 1 lists the ribonuclease T_1 products of spot C together with their sequence-analysis results. As for the component from band 21 (Adams *et al.*, 1969), these ribonuclease T_1 products could be ordered by referring to the amino acid sequence of the coat protein (see Fig. 5), but again the order was rigorously confirmed by partial ribonuclease T_1 digestion and characterization of the products (Fig. 2 and Table 2).

The paradox posed by this result was that C, a fragment of RNA only 31 nucleotides long, should run on the polyacrylamide gel more slowly than the 57-nucleotide component in band 21. However, analysis of the other major components of band 20X suggested the explanation, discussed below, that the ribonuclease T_1 digestion products of spot H (Table 1) could be arranged in such a way that if a single extra guanylate residue were inserted at the beginning they could code for the polypeptide sequence directly following that coded by component C (see Fig. 5). The order was established by partial digestions with ribonucleases T_1 and A, and analysis of the partial-digestion products as described above (Figs. 3 and 4 and Table 3). Another spot in Fig. 1, J, contained a part of sequence H (Fig. 5).

Therefore, by combining the sequences of spots C and H, and inserting the missing guanylate residue, the complete nucleotide sequence coding for amino acids 57–75 of the coat protein could be constructed (Fig. 5), although it must be emphasized that sequences C and H were never isolated covalently linked, and therefore the overlapping sequence has not been directly confirmed. Sequences C and H together constitute 58 base residues, a number consistent with the mobility of band 20X on the polyacrylamide gel. It was concluded therefore that the two migrate as a single species on the gel, held together by some interaction (e.g. hydrogen-bonded secondary structure) that is stable under the conditions of gel ionophoresis, but that is broken by the denaturing conditions of the two-dimensional re-fractionation procedure (i.e. the presence of 7M-urea.) This is readily understood if the structure is written in the form of a 'hairpin loop' (see Fig. 10 and the Discussion section).

Band 20

The ribonuclease T_1 'fingerprints' of band 20, shown in Plate 2, revealed it to be a fairly pure single major component. A preliminary analysis of the major ribonuclease T_1 products by ribonuclease A digestion indicated the possibility that they were all derived from a region of the coat-protein cistron coding for amino acids, approximately in positions 30–50 of the protein chain. This is shown by reference to Fig. 6, which illustrates the general method used for preliminary screening of ribonuclease T_1 digestion products to see if they could be derived from the coat-protein cistron. The ribonuclease A digestion products of ribonuclease T_1 oligonucleotides from band 20 are listed in column 2 of Table 4. Oligonucleotide 2 contains the sequences A-Cp, A-A-A-Gp, Cp and

Table 3. *Analysis of partial digestion products of spot H from band 20X*

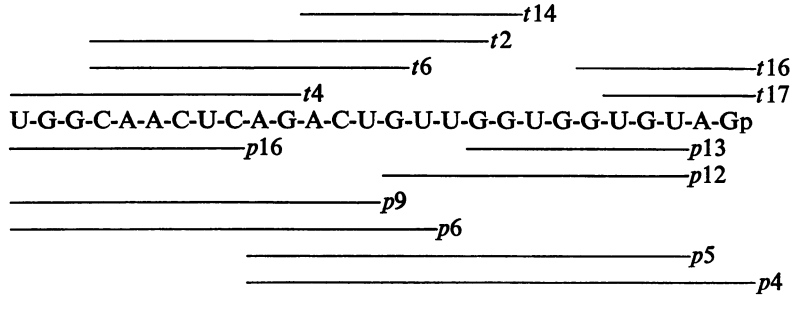
The spots numbered in Figs. 3 and 4 were eluted and analysed as described in Table 1 and by Jeppesen (1971).

Oligo-nucleotide	Ribonuclease T_1 digestion products	Ribonuclease A digestion products	Deduced sequence
Partial ribonuclease T_1 digestion products (Fig. 3):			
*t*4	C-A-A-C-U-C-A-Gp U-Gp Gp	G-G-Cp A-A-Cp A-Gp Cp Up	U-G-G-C-A-A-C-U-C-A-Gp
*t*6	C-A-A-C-U-C-A-Gp A-C-U-Gp	A-G-A-Cp A-A-Cp Gp Cp Up	C-A-A-C-U-C-A-G-A-C-U-Gp
*t*2	C-A-A-C-U-C-A-Gp A-C-U-Gp U-U-Gp	A-G-A-Cp A-A-Cp G-Up Gp Cp Up	C-A-A-C-U-C-A-G-A-C-U-G-U-U-Gp
*t*14	A-C-U-Gp U-U-Gp Gp	G-Up G-Gp A-Cp Up	A-C-U-G-U-U-G-Gp
*t*17	U-A-Gp U-Gp	G-Up A-Gp Up	U-G-U-A-Gp
*t*16	U-A-Gp U-Gp Gp	G-Up A-Gp	G-U-G-U-A-Gp
Partial ribonuclease A digestion products (Fig. 4):			
*p*13	U-Gp Gp Up	G-G-Up G-Up	G-G-U-G-G-U-G-Up
*p*12	U-U-Gp U-Gp Gp Up	G-G-Up G-Up Up	G-U-U-G-G-U-G-G-U-G-Up
*p*16	C-A-A-C-U-Cp U-Gp Gp	G-G-Cp A-A-Cp Cp Up	U-G-G-C-A-A-C-U-Cp
*p*9	C-A-A-C-U-C-A-Gp U-Gp A-C-Up Gp	G-G-Cp A-G-A-Cp A-A-Cp Cp Up	U-G-G-C-A-A-C-U-C-A-G-A-C-Up

Table 3—*continued*

Oligo-nucleotide	Ribonuclease T_1 digestion products	Ribonuclease A digestion products	Deduced sequence
Partial ribonuclease A digestion products (Fig. 4) (cont.):			
p6	C-A-A-C-U-C-A-Gp	G-G-Cp	U-G-G-C-A-A-C-U-C-A-G-A-C-U-G-Up
	A-C-U-Gp	A-G-A-Cp	
	U-Gp	A-A-Cp	
	Up	G-Up	
	Gp	Cp	
		Up	
p5	U-U-Gp	G-G-Up	A-G-A-C-U-G-U-U-G-G-U-G-G-U-G-Up
	A-C-U-Gp	G-Up	
	U-Gp	A-G-A-Cp	
	A-Gp	Up	
	Gp		
	Up		
p4	U-U-Gp	G-G-Up	A-G-A-C-U-G-U-U-G-G-U-G-G-U-G-U-A-Gp
	A-C-U-Gp	G-Up	
	U-A-Gp	A-G-A-Cp	
	U-Gp	Up	
	Gp	A-Gp	
	A-Gp		

Summary of partial digestion products



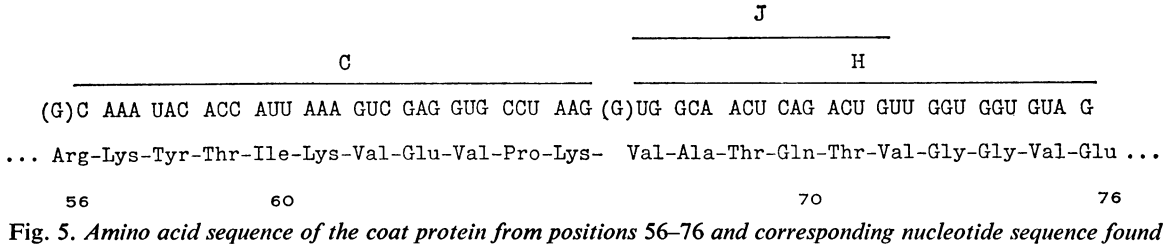


Fig. 5. *Amino acid sequence of the coat protein from positions* 56–76 *and corresponding nucleotide sequence found in the RNA fragments C, H and J from band* 20*X*

For details see Tables 1, 2 and 3, and Figs. 2, 3 and 4.

Up (2 mol). To see if it could be derived from the coat-protein cistron, a list of possible ribonuclease-digestion products was prepared, as in Fig. 6, but for the whole cistron. If we just consider the section shown in Fig. 6, there are two positions where A-A-A-Gp could occur (corresponding to positions 43–44 and 50–51 of the amino acid sequence). Both could occur in a ribonuclease T_1 digestion product having A-Cp, but the

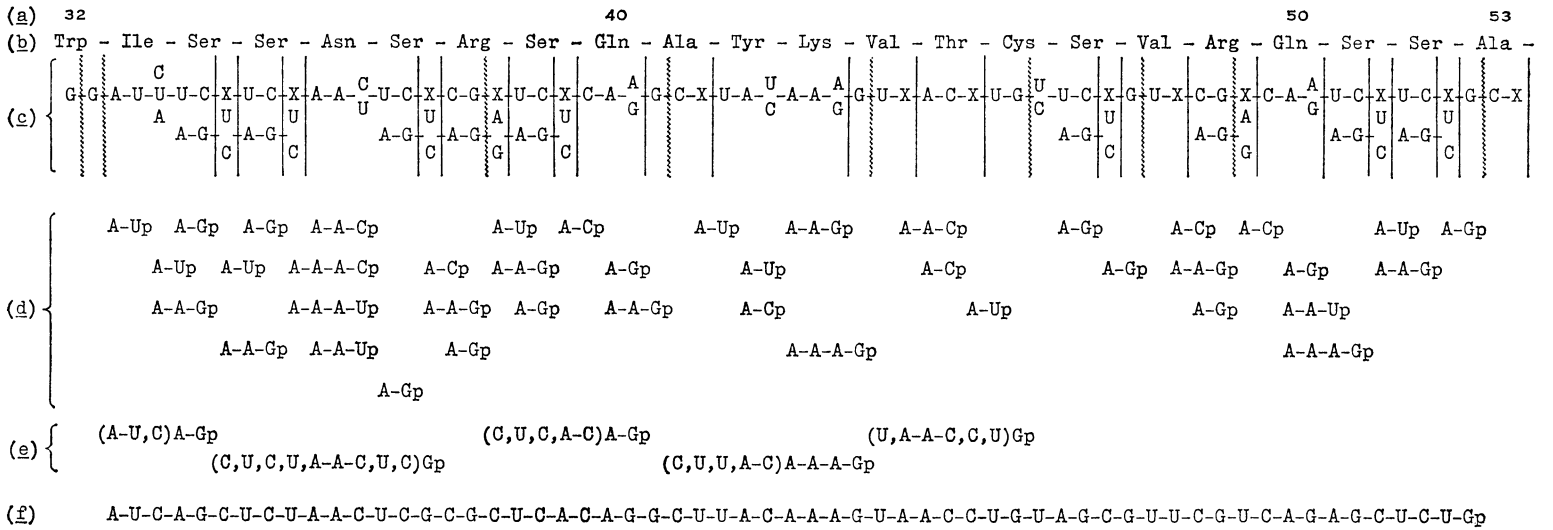


Fig. 6. *Method for screening ribonuclease T_1 digestion products to test whether they could be derived from the coat-protein cistron*

The method is applied to the oligonucleotides from band 20. (*a*) Position of amino acid residue in coat protein; (*b*) part of the amino acid sequence of coat protein; (*c*) corresponding nucleotide sequence predicted from the genetic code: } indicates an obligatory ribonuclease T_1 split, | indicates a possible ribonuclease T_1 split; (*d*) possible ribonuclease A digestion products of ribonuclease T_1 oligonucleotides from the RNA; (*e*) ribonuclease T_1 oligonucleotides from band 20 (Plate 2) characterized by their ribonuclease A digestion products (Table 4, column 2), indicating possible sequences that are compatible with (*c*) and (*d*); (*f*) nucleotide sequence of band 20 fragment as finally deduced (see Tables 4, 5, 6 and 7).

Table 4. *Sequence analysis of ribonuclease T_1 digestion products of the coat-protein-cistron RNA fragments contained in band* 20

The major spots numbered in Plate 2 were eluted and analysed as described in Table 1.

Oligo-nucleotide	Ribonuclease A digestion products	CD products (Jeppesen, 1971)	Ribonuclease U_2 digestion products	Deduced sequence
1	A-A-Cp Gp Cp(3) Up(3)	U̇-A-A-Cp U̇-Cp Ġp Cp	C(U_2,C)A-Ap* C(U_2,C)Ap* (U,C_2)Gp Ap	C-U-C-U-A-A-C-U-C-Gp
2	A-A-A-Gp A-Cp Cp Up(2)	U̇-U̇-A-Cp A-A-A-Ġp Cp	C-A-A-A-Gp C-A-A-Ap C-A-Ap C-Ap (U_2,C)Ap	C-U-U-A-C-A-A-A-Gp
3	A-Gp A-Cp Cp(2) Up	U̇-Cp A-Ġp Cp A-Cp	C(U,C)Ap* C-A-Gp C-Ap Gp	C-U-C-A-C-A-Gp
4	A-A-Cp Gp Cp Up(2)	U̇-Ġp U̇-A-A-Cp Cp	(U,C_2)Gp U-A-Ap U-Ap	U-A-A-C-C-U-Gp
5	Gp Cp(2) Up(2)	U̇-Ġp U̇-Cp Cp	—	C-U-C-U-Gp*
6	A-Up A-Gp Cp	A-U̇-Cp A-Ġp	U-C-A-Gp U-C-Ap Ap	A-U-C-A-Gp
7	Gp Cp Up(2)	U̇-U̇-Cp Ġp	—	U-U-C-Gp
8	A-Gp Cp Up	U̇-Cp A-Ġp	—	U-C-A-Gp
9	A-Gp Up	—	—	U-A-Gp
10	Cp Gp	—	—	C-Gp
11	A-Gp	—	—	A-Gp
12	Gp	—	—	Gp

* The 5′-end groups in these oligonucleotides were determined by analysis of their complete snake venom phosphodiesterase-hydrolysis products after dephosphorylation (Sanger *et al.*, 1965) and comparing these with an alkaline hydrolysate.

product corresponding to positions 50–51 would not give Cp and Up with ribonuclease A, whereas the product corresponding to positions 43–44 could give Cp and Up(2), if it extended from positions 41–44. Thus oligonucleotide 2 (Table 4) could be derived from this position. This coincidence could of course be fortuitous; however, when the other oligonucleotides listed in Table 4 were examined in this way it was

found that they also could be derived from the neighbouring part of the coat-protein cistron, as shown in Fig. 6(*e*).

When the material from band 20 was re-run by the

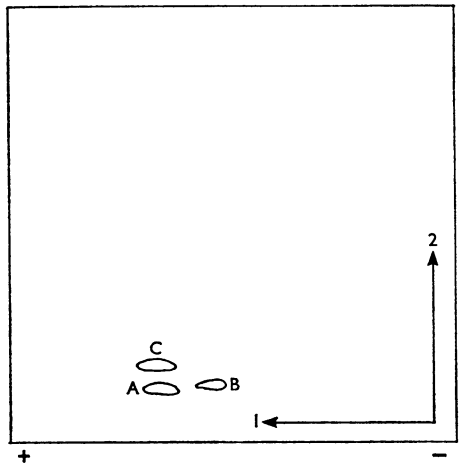


Fig. 7. *Re-fractionation of the material extracted from gel band* 20 *by t.l.c. homochromatography by using homomixture b*

For details see Fig. 1.

Table 5. *Ribonuclease A digestion products of the coat-protein-cistron RNA fragments contained in band* 20

Purified band 20 material (spot A, Fig. 7) was digested with ribonuclease A (enzyme:RNA ratio 1:20, 30min at 37°C, in 0.01M-tris–HCl buffer, pH7.4, containing 0.001 M-EDTA) and the products were fractionated in one dimension by DEAE-cellulose paper ionophoresis at pH.19. The sequences of the products were determined by analysis of their ribonuclease T_1 digestion products.

A-G-A-G-Cp	A-Cp (2mol)
A-A-A-G-Up	G-Cp (2mol)
A-G-G-Cp*	G-Up (3mol)
A-G-Cp (2mol)	Gp
A-A-Cp (2mol)	Cp (>3mol)
A-Up	Up (>3mol

* The sequence of this oligonucleotide must be as written, since there is only one A-Gp ribonuclease T_1 digestion product present in band 20 (Table 4) and this occurs in ribonuclease. A digestion product A-G-A-G-Cp.

Table 6. *Products obtained by digestion with ribonuclease T_1 and A of fragments from spots B and C (Fig.* 7) *from band* 20

Spots B and C were eluted from the t.l.c. plate shown in Fig. 7 and analysed as described by Jeppesen (1971).

	Ribonuclease T_1 digestion products	Ribonuclease A digestion products
Spot B	C-U-C-U-A-A-C-U-C-Gp	A-G-G-Cp
	C-U-U-A-C-A-A-A-Gp	A-A-A-Gp
	C-U-C-A-C-A-Gp	A-G-Cp
	A-U-C-A-Gp	A-A-Cp
	C-Gp	A-Up
	Gp	A-Cp (2mol)
		G-Cp (2mol)
		Cp (>3mol)
		Up (>3mol)
Spot C	U-A-A-C-C-U-Gp	A-G-A-G-Cp
	C-U-C-U-Gp	A-G-Cp
	U-U-C-Gp	A-A-Cp
	U-C-A-Gp	G-Up (3mol)
	U-A-Gp	Gp
	A-Gp	Cp (>3mol)
	C-Gp	Up (>3mol)

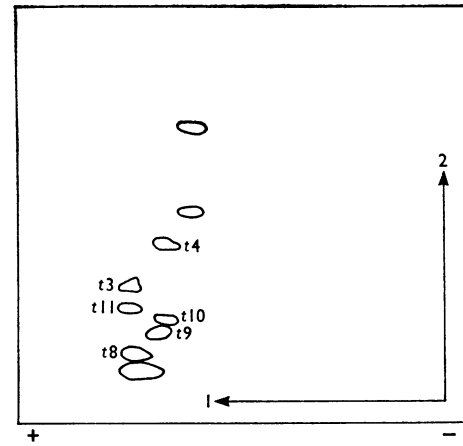


Fig. 8. *Partial ribonuclease T_1 digest of component B from gel band* 20 (*Fig.* 7) *fractionated by t.l.c. homochromatography by using homomixture b*

Digestion conditions were as in Fig. 2. The analyses of the partial digestion products are given in Table 7.

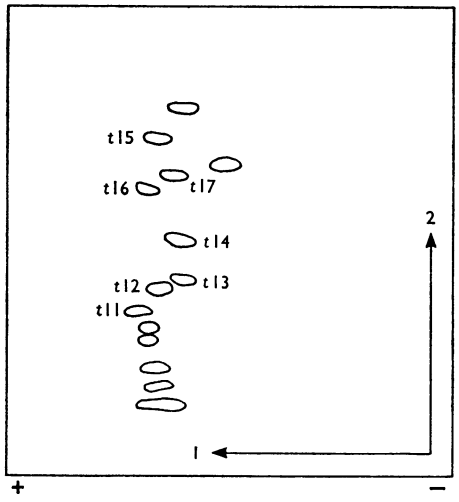


Fig. 9. *Partial ribonuclease T_1 digest of component C from gel band* 20 (*Fig.* 7) *fractionated by t.l.c. homochromatography by using homomixture b*

Digestion conditions were as in Fig. 2. The analyses of the partial digestion products are given in Table 7.

two-dimensional homochromatography system only three major components were found (Fig. 7). A ribonuclease T_1 'fingerprint' showed that spot A contained all the major-yield spots of Plate 2. The sequence of these ribonuclease T_1 products (Table 4) further supported the probability that the fragment was from the coat-protein cistron, and, together with the nucleotide sequences of the ribonuclease A-digestion products (Table 5), the complete sequence of component A could be postulated (Fig. 6).

The two other major spots obtained by re-running band 20 (spots B and C, Fig. 7) were each found by analysis of their ribonuclease T_1 and ribonuclease A digestion products to be derived from spot A by cleavage at a single site (Table 6). Whereas all the ribonuclease T_1 digestion products of spot A are found in either spot B or C, the ribonuclease A digestion product A-A-A-G-Up is not present in spot B or C, but A-A-A-Gp is found in spot B. This establishes the site of the 'hidden break' and the relative order of the two fragments. To establish the complete sequence of the fragments from spots B and C, they were subjected to partial digestion with ribonuclease T_1 (Figs. 8 and 9) and the products were analysed by complete digestion with the same enzyme. These results are summarized in Table 7.

Table 7. *Products obtained by partial digestion of spots B and C from band* 20 (*Fig.* 7) *with ribonuclease T_1*

The numbered spots from Figs. 8 and 9 were eluted and analysed by complete digestion with ribonuclease T_1 (see the text)

Deduced sequence of spot B (Fig. 8) A-U-C-A-G-C-U-C-U-A-A-C-U-C-G-C-G-C-U-C-A-C-A-G-G-C-U-U-A-C-A-A-A-Gp

t3
t4
t11
t8
t10
t9

Deduced sequence of spot C (Fig. 9) U-A-A-C-C-U-G-U-A-G-C-G-U-U-C-G-U-C-A-G-A-G-C-U-C-U-Gp

t14
t15
t13
t16
t12
t11
t17

Discussion

The nucleotide sequences of the two fragments from bands 20X and 20 are summarized in Figs. 5

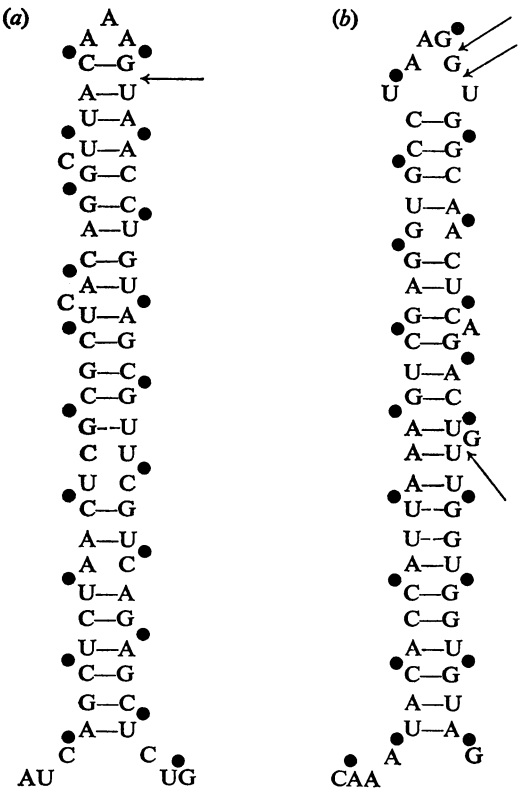


Fig. 10. *Nucleotide sequences of the fragments from bands* 20 (*a*) *and* 20*X* (*b*) *written in the form of 'hairpin loops' and showing the suggested base pairing*

Arrows indicate guanylate residues susceptible to enzyme attack. The 'reading frame' is shown by dots.

and 6, which also show the amino acid sequences in the coat protein for which they code. The significance of these and other sequence studies on the bacteriophage R17 RNA has been discussed previously (Jeppesen *et al.*, 1970*b*; Sanger, 1971) and will not be dealt with in detail here.

For the fragment from band 21 (Adams *et al.*, 1969), it was found that the nucleotide sequence could be written in the form of a hairpin loop and it was suggested that such a structure exists in the virus and may play an important part in the biological properties of the RNA. Both the fragments described in the present paper can also be written in the form of such loops (Fig. 10). In these cases it was found necessary to allow occasional 'looping out' of single residues to give a maximum number of base pairs (see Fresco *et al.*, 1960), but the structures thereby obtained correlate well with the observed susceptibility of certain guanylate residues to cleavage by ribonuclease T_1.

The methods used in the present and previous work have been shown to be applicable to the determination of relatively long sequences in the bacteriophage RNA and can clearly be extended considerably. Fiers *et al.* (1971) have been carrying out similar studies on the related bacteriophage MS2 and have reported the sequence of almost the whole of the coat-protein cistron.

We thank Dr. F. Egami for gifts of ribonuclease U_2 and Dr. G. Bernardi for spleen phosphodiesterase. P. G. N. J. thanks the Medical Research Council for a Scholarship.

References

Adams, J. M. & Cory, S. (1970) *Nature* (*London*) **227**, 570
Adams, J. M., Jeppesen, P. G. N., Sanger, F. & Barrell, B. G. (1969) *Nature* (*London*) **223**, 1009
Brownlee, G. G. & Sanger, F. (1969) *Eur. J. Biochem.* **11**, 395
Fiers, W., Contreras, R., De Wachter, R., Haegemen, G., Merragaert, J., Min Jou, W. & Vandenberghe, A. (1971) *Biochimie* **53**, 495
Fresco, J. R., Alberts, B. M. & Doty, P. (1960) *Nature* (*London*) **188**, 98
Gussin, G. N. (1966) *J. Mol. Biol.* **21**, 435
Jeppesen, P. G. N. (1971) *Biochem. J.* **124**, 357
Jeppesen, P. G. N. Steitz, J. A., Gesteland, R. F. & Spahr, P. F. (1970*a*) *Nature* (*London*) **226**, 230
Jeppesen, P. G. N., Nichols, J. L., Sanger, F. & Barrell, B. G. (1970*b*) *Cold Spring Harbor Symp. Quant. Biol.* **35**, 13
Sanger, F. (1971) *Biochem. J.* **124**, 833
Sanger, F., Brownlee, G. G. & Barrell, B. G. (1965) *J. Mol. Biol.* **13**, 373
Steitz, J. A. (1969) *Nature* (*London*) **224**, 957
Weber, K. (1967) *Biochemistry* **6**, 3144