Nucleotide Sequence from the Coat Protein Cistron of R17 Bacteriophage RNA

by
J. M. ADAMS*
P. G. N. JEPPESEN
F. SANGER
B. G. BARRELL
MRC Laboratory of Molecular Biology, Cambridge

The sequence of fifty-seven nucleotides in the coat protein cistron of phage R17 RNA directly confirms the genetic code, shows that the code used by the phage is degenerate and suggests that highly ordered base-paired structures exist in this RNA. Such base-paired loops may be involved in regulation of cistron expression and packing of the RNA in the phage particle.

ALTHOUGH the nature of the genetic code is well established, it has not been possible until now to determine by chemical means a sequence from a messenger RNA and to show that it is related by the code to the sequence of amino-acids in the protein that it specifies. The best characterized messenger RNAs that can be obtained in a pure form are the single-stranded RNAs containing about 3,300 nucleotide residues[1] isolated from RNA bacteriophages, such as R17, f2 and MS2. The nucleotide sequences at the ends of these molecules[2–6] have been determined and, for MS2 RNA, the sequences of the products of pancreatic ribonuclease digestion[7]. R17 RNA codes for three proteins[8], one of which is the phage coat protein of known amino-acid sequence[9]. Here we report a nucleotide sequence from the coat protein cistron of R17 RNA.

In this laboratory we have developed fractionation methods for ^{32}P-labelled oligonucleotides[10–12] which have been applied in the determination of the nucleotide sequences of *t*RNAs[13] and the 5*S* ribosomal RNA[12], which is 120 nucleotides long. The method used for separating nucleotides up to about ten residues in length is ionophoresis on a two-dimensional system using cellulose acetate in one dimension and DEAE-paper in the other[10]. Longer oligonucleotides are not well fractionated by the ionophoresis on DEAE-paper; however, those up to about thirty residues in length can be separated by a form of displacement chromatography on DEAE-paper (termed "homochromatography")[12], in which a mixture of non-radioactive nucleotides is used to develop the chromatogram and fractionate the radioactive nucleotides.

Recently we have found that even larger oligonucleotides can be resolved with homochromatography on a thin layer of DEAE-cellulose instead of on DEAE-paper; nucleotides up to fifty residues long can be fractionated very well by a new two-dimensional system—ionophoresis on cellulose acetate followed by thin layer homochromatography[14]. When R17 RNA was digested completely with T_1 ribonuclease a few large oligonucleotides were produced that could be purified by this two-dimensional system. One of these proved to be of particular interest because its nucleotide sequence corresponds to an amino-acid sequence in the coat protein of the phage.

Large Oligonucleotides from a Complete T_1 Ribonuclease Digest of R17 RNA

R17 RNA, labelled uniformly with ^{32}P, was digested with T_1 ribonuclease and the digest fractionated on a two-dimensional system using ionophoresis at *p*H 3·5 in 7 M urea on cellulose acetate as the first dimension and thin layer chromatography on DEAE-cellulose as the second dimension[14]. Fig. 1 is an autoradiograph of the fractionation. The smaller oligonucleotides are not well resolved but certain of the larger products are clearly separated from the others. These large oligonucleotides were isolated from the chromatogram and subjected to preliminary sequence analysis as follows.

First, base compositions were determined by hydrolysing a portion of each oligonucleotide with alkali; the resulting 3′-mononucleotides were fractionated by electrophoresis on paper at *p*H 3·5, located by autoradiography, and then measured by scintillation counting. A second portion of each oligonucleotide was digested with pancreatic ribonuclease and the digestion products separated by electrophoresis on DEAE-paper at *p*H 3·5 (Fig. 2). Most of the products could be identified uniquely from their position on the paper. Where confirmation was necessary, the composition of each product was analysed by alkaline hydrolysis.

Using the genetic code, one can write a set of nucleotide sequences corresponding to the known amino-acid sequence of the R17 coat protein. By comparing the information about each of the large oligonucleotides obtained as given here with the possible nucleotide sequences for the coat

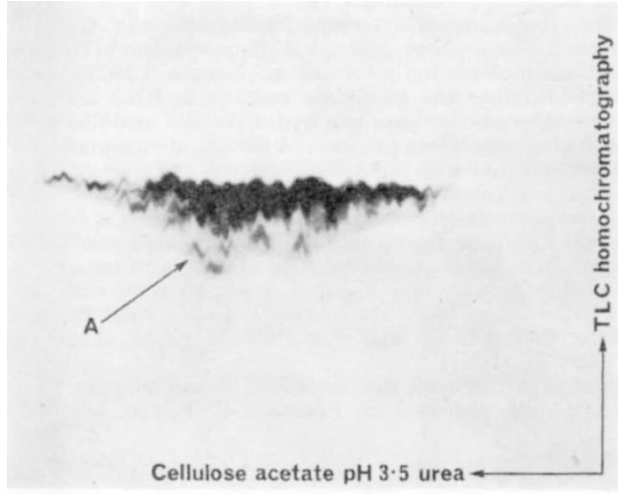

Fig. 1. A two-dimensional fractionation of a ribonuclease T_1 digest of R17 RNA. Uniformly ^{32}P-labelled bacteriophage R17 RNA was prepared as described by Dahlberg[2]. 20 μg (about 10^7 c.p.m.) was digested with 1 μg of ribonuclease T_1 dissolved in 3 μl. of 0·01 M *tris*-HCl, *p*H 7·4, containing 0·001 M EDTA, at 37° C for 30 min, fractionated on a two-dimensional system[14] using homomixture *a* for homochromatography with DEAE-cellulose and cellulose in the ratio of 1 : 7·5 on a short plate.

* Present address: Institut de Biologie Moléculaire, 1211 Genève 4, Switzerland.

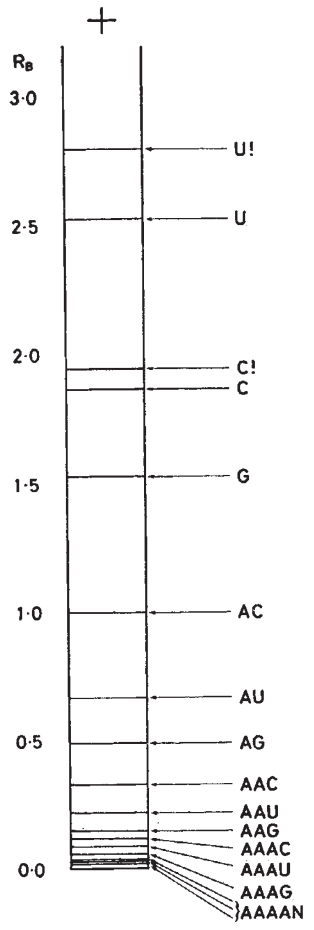

Fig. 2. Fractionation of pancreatic ribonuclease digestion products on DEAE-paper at *p*H 3·5.

protein, we found that oligonucleotide A could have come from the coat protein cistron.

Sequence of Oligonucleotide A

To establish the sequence of oligonucleotide A we used the following procedures.

(*a*) Reaction with a carbodiimide reagent: Gilham[15] showed that N-cyclohexyl-N′-(β-morpholinyl-(4)-ethyl) carbodiimide-methyl-p-toluene sulphonate (CMCT) reacts with uridylate and guanylate residues in RNA and that pancreatic ribonuclease can hydrolyse the modified RNA only after cytidylate residues. A sample of oligonucleotide A was treated with the CMCT reagent and digested with pancreatic ribonuclease; the digest was subjected to electrophoresis on paper at *p*H 3·5. Fig. 3 is a diagram of the resulting fractionation. The modified nucleotides were then eluted, incubated with ammonia to remove the blocking groups, and digested a second time with pancreatic ribonuclease. The products of the second digestion were identified by electrophoresis on paper at *p*H 3·5 (Table 1).

(*b*) Digestion with ribonuclease U_2: this enzyme, which was kindly provided by Professor F. Egami, is specific for purines when it is used at a low concentration[16]. Thus the enzyme hydrolyses oligonucleotides from a T_1 ribonuclease digest only after adenylate residues. A portion of oligonucleotide A was digested with ribonuclease U_2 and the products were fractionated by electrophoresis on DEAE-paper at *p*H 1·9. The base compositions of the products were determined by alkaline hydrolysis (Table 1). Certain products appear in pairs, differing only in the number of adenylate residues. This reflects the increased resistance of adenylate tracts to attack by ribonuclease U_2 and is useful for obtaining overlapping sequences with the pancreatic ribonuclease products.

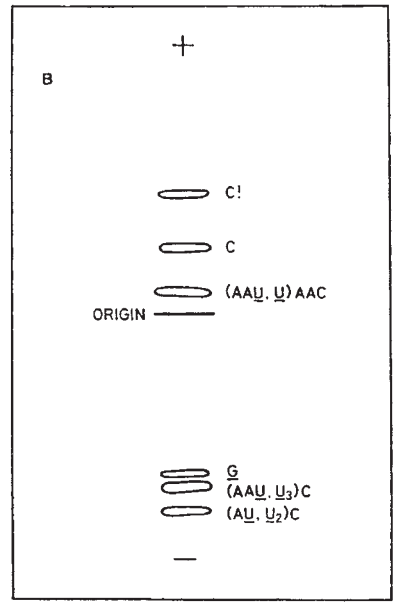

Fig. 3. Fractionation of the pancreatic ribonuclease products from carbodiimide-blocked oligonucleotide A. A sample of oligonucleotide A was incubated overnight at 37° C with 10 μl. of a 100 mg/ml. solution of CMCT dissolved in 0·01 M *tris*-HCl, *p*H 8·9, containing 0·001 M EDTA. 10 μl. of a 0·1 mg/ml. solution of pancreatic ribonuclease in 0·05 M *tris*-HCl, *p*H 7·4, containing 0·001 M EDTA, was then added and the mixture incubated for a further 30 min at 37° C. The digest was then applied as a 2 cm streak at the centre of a 57 cm strip of Whatman 3*MM* paper. Electrophoresis was then carried out in *p*H 3·5 pyridine acetate buffer for 1 h at 3 kV.

Combining the results obtained with ribonuclease U_2 and pancreatic ribonuclease (both before and after blocking with CMCT), a unique sequence can be constructed for oligonucleotide A (Fig. 4). If this nucleotide sequence acts as a messenger for the synthesis of protein, it can code for three possible amino-acid sequences, depending on the phase in which the message is read. These three sequences are also shown in Fig. 4. Amino-acid sequences 1 and 2 are not found in the coat protein sequence but sequence 3 is present at positions 89 (Glu) to 95 (Phe)[9]. Because a sequence of seven amino-acids is statistically a very rare structure, it is considered highly likely that oligonucleotide A is a fragment from the coat protein cistron of R17 RNA. To extend the knowledge of the nucleotide sequence in this region of the RNA we tried to find a larger fragment of the phage RNA that contains oligonucleotide A.

Sequence of nucleotide A (G) A A U U A A C U A U U C C A A U U U U C G

Possible amino-acid sequences coded for by nucleotide A

Asn . — . Leu . Phe . Gln . Phe . Ser .
Ile . Asn . Tyr . Ser . Asn . Phe . Arg .
Glu . Leu . Thr . Ile . Pro . Ile . Phe .

Fig. 4. The nucleotide sequence of oligonucleotide A and the three possible amino-acid sequences for which it can code.

Table 1. SEQUENCE ANALYSIS OF OLIGONUCLEOTIDE A

(a) Base composition (G = 1·0)	(b) Pancreatic ribonuclease digestion products	(c) CMCT blocked pancreatic ribonuclease digestion products	(d) U_2 ribonuclease digestion products
U 8·6	2AAU*	C	A
		(AAU̇,U̇)AAC	AA
G 1·0	AAC		(C,U)A
		Ġ	
A 6·5	AU		UUA
		(AAU̇,U_3)C	
C 3·9	G		UUAA
	3C*	(AU̇,U_2)C	
	6U*		(U_2,C_2)A
			(U_2,C_2)AA
			(U_4,C)G

* These estimations were based on a visual examination of the autoradiograph together with the base composition determined in (*a*).

Oligonucleotide A was located on the thin layer of DEAE-cellulose in Fig. 1 by autoradiography, and isolated in the following manner. The DEAE-cellulose in the region of the spot was scraped off, washed with 95 per cent ethanol to remove the urea, and the nucleotide was eluted with 30 per cent triethylamine carbonate, *p*H 10·5, in a total volume of about 200 μl. The triethylamine carbonate was removed by evaporation under vacuum and the oligonucleotide subjected to preliminary sequence analysis. (*a*) The base composition was determined by hydrolysis with 5–10 μl. of 0·2 M NaOH; the products were separated by electrophoresis on paper at *p*H 3·5 (ref. 10), located by autoradiography and quantitated by scintillation counting. (*b*) The oligonucleotide was digested for 30 min at 37° C in 5–10 μl. pancreatic ribonuclease (0·1 mg/ml. enzyme in 0·01 M *tris*-HCl, *p*H 7·5, and 0·001 M EDTA). The products were separated by electrophoresis on DEAE-paper at *p*H 3·5 (Fig. 2) and identified by alkaline hydrolysis and/or mobility. (*c*) The modified nucleotide was eluted from the various bands in Fig. 3, dried, and incubated overnight with 0·2 M ammonia to remove the blocking groups. After the ammonia had been dried off, the nucleotides were digested with 10 μl. of 0·1 mg/ml. pancreatic ribonuclease for 30 min at 37° C in buffer containing 0·01 M *tris*-HCl and 0·001 M EDTA, and the products fractionated by electrophoresis on paper at *p*H 3·5. U̇ indicates the modified uridylic acid residue. (*d*) A portion of oligonucleotide A was incubated at 37° C for 2 h with 10 μl. of ribonuclease U_2 (0·1 units/ml. in 0·05 M sodium acetate, *p*H 4·5, containing 0·002 M EDTA and 0·1 mg/ml. bovine serum albumin).

Partial T_1 Ribonuclease Digest of R17 RNA

When a partial enzymic digest of ribosomal RNA is electrophoresed on a polyacrylamide gel a number of discrete bands are found[17]. We tried this approach for making specific fragments of R17 RNA. Samples of ^{32}P-labelled R17 RNA were digested with various amounts of ribonuclease T_1 at 0° C in a buffer of high ionic strength, and the partial digests were electrophoresed on a long flat slab of 12·5 per cent polyacrylamide gel by a modification of the method of Peacock and Dingman[18] which was developed in this laboratory with G. G. Brownlee. (A flat slab is particularly suitable for autoradiography and also for comparing different samples on the same gel.) Fig. 5 shows an autoradiograph of the fractionation obtained. In the undigested control sample virtually all the RNA remains at the origin because it is too large to penetrate the gel. With increasing amounts of added enzyme, however, more and more bands appear and there is a progressive increase in the amounts of the smaller, faster-moving fragments. As many as forty discrete bands can be seen in the more extensively digested samples. The RNA fragments in these bands range in size from guanosine monophosphate, in the fastest moving band, to fragments more than 300 nucleotides long near the top of the gel.

This experiment shows that there is an extremely wide range in the rate at which T_1 ribonuclease splits different guanylate residues in the molecule, presumably because of the structure of the RNA. Moreover, it shows that gel electrophoresis is capable of resolving many of the fragments that result from this very specific hydrolysis. Recently Gould, Pinder and Matthews[20] have reported that a partial digest of the RNA from the related phage μ2 also contains a number of specific fragments and we find that this is the case with f2 phage RNA as well. Gesteland and Boedtker[1] have shown by physical studies that R17 RNA has an unusually compact structure at a high ionic strength. The highly specific fragmentation found in the present experiments indicates that phage RNAs have highly organized secondary structures.

The fragmentation of the RNA was usually sufficiently reproducible in different experiments, using different preparations of RNA or enzyme, for each band to be identified simply from the overall band pattern. To isolate enough of the fragments to characterize them, preparative digests were made with up to 5 mCi of ^{32}P-labelled R17 RNA and the digests were loaded across the width of a flat slab gel (20 cm × 40 cm × 0·3 cm). In these experiments we chose digestion conditions that would give primarily fragments of a size (up to about 200 nucleotides in length) suitable for sequence analysis. Appropriate conditions were digestion for 1 h at 0° C with an enzyme to substrate ratio of 1 : 100 (w/w) or 10 h at 0° C with a ratio of 1 : 1,000.

Because we wished to study the fragments from different gel bands, it was necessary to find a convenient method of extracting the RNA. The following method was found satisfactory. The bands were cut from the gel using the autoradiograph as a guide, broken into small pieces, and put into cylindrical tubes (10 cm × 1 cm)

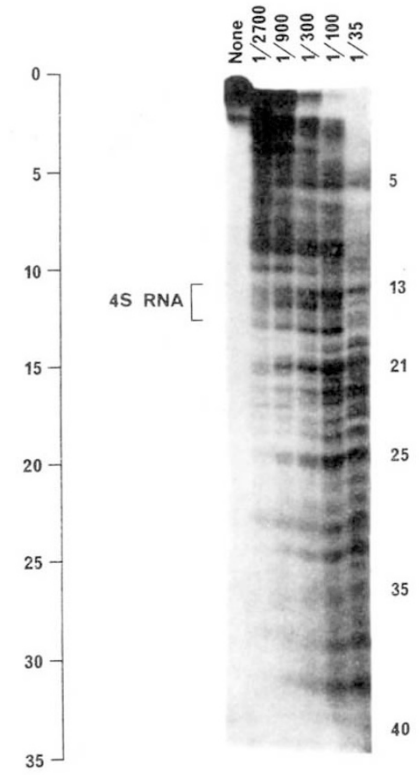

Fig. 5. Gel electrophoresis of a partial digest of R17 RNA with ribonuclease T_1. Samples (20 μg) of ^{32}P-labelled R17 RNA (specific activity 1·5 μCi/μg RNA) were digested with different amounts of ribonuclease T_1 for 1 h at 0° C in 25 μl. of a buffer consisting of 0·2 M NaCl, 0·02 M magnesium acetate, and 0·05 M *tris*-HCl (*p*H 7·5); the enzyme–substrate ratios are shown at the top of the figure. The reaction mixtures were extracted with phenol at 0° C in capillary tubes and then extracted three times with ether. The samples were made 10 per cent with respect to sucrose with a 50 per cent sucrose solution containing bromophenol blue, loaded into the sample wells of a vertical flat slab of 12·5 per cent polyacrylamide gel, and subjected to electrophoresis at 400 V and 40 mA for 16 h at 2°–4° C. At the end of the run the front glass plate was pulled away from the gel and the gel was marked with radioactive ink, covered with cellophane, and subjected to autoradiography for 30 min. The gel was prepared in the following manner. The mould consisted of two thoroughly cleaned glass plates (40 cm × 20 cm × 0·4 cm) separated at their sides by two 'Vaseline'-coated 'Perspex' spacers (40 cm × 2 cm × 0·3 cm) and sealed at the bottom with plasticine. The assembly was held together at the sides by strong spring clips. The gel was made from 12·1 per cent acrylamide and 0·4 per cent bisacrylamide (both recrystallized as described by Loening[19]) and 0·04 M *tris*-acetate, *p*H 8·3, which was also the buffer used in the reservoirs. The gel solution (250 ml.) was de-aerated, 0·6 ml. of freshly prepared 10 per cent ammonium persulphate and 0·25 ml. of N,N,N,N-tetramethylethylenediamine were added, and the solution was poured quickly into the mould. The sample wells were formed by pushing a slotted strip of 'Perspex' (15 cm × 3 cm × 0·3 cm) into the top of the gel. After gelation was complete (in about 5 min) the well-former and plasticine were removed and the bottom of the assembly was placed in a 2 l. reservoir; the top of the gel was connected to another 2 l. reservoir by a short length of Whatman 3*MM* paper. The buffer in the reservoirs was changed once during a run. On the right of the figure are shown the numbers of some of the bands that were studied. The position to which a 4*S*-RNA marker ran is also indicated.

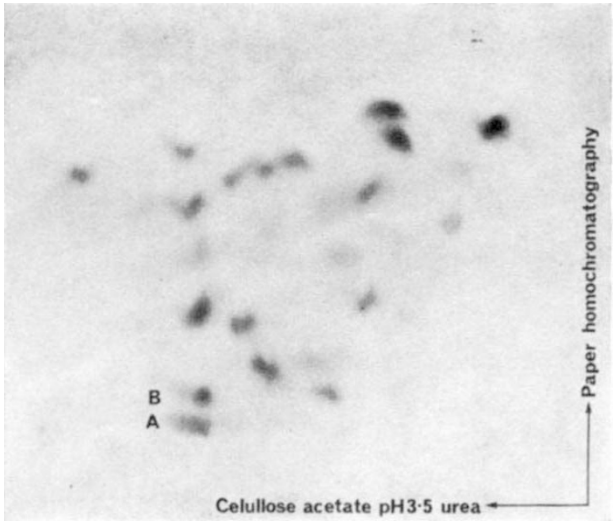

Fig. 6. A two-dimensional homochromatography fingerprint of a ribonuclease T_1 digest of gel band 19. (Indicating oligonucleotides A and B.)

having one end covered with DEAE-paper supported on cellophane. The tubes were then placed between two reservoirs containing 0·04 M *tris*-acetate, *p*H 8·3. When subjected to electrophoresis the RNA migrated out of the gel and was trapped by the DEAE-paper; it could then be eluted from the DEAE-paper with 30 per cent triethylamine carbonate (*p*H 9·7) as described earlier[10]. To get an estimate of the purity of fragments isolated from the gel, the RNA from the different bands was digested with T_1 ribonuclease and fingerprinted using ionophoresis on cellulose acetate at *p*H 3·5 as the first dimension and homochromatography on DEAE-paper as the second (homomixture *c* of ref. 14 but with a 30 min alkaline hydrolysis). The complexity of the fingerprints suggested that many of the bands in the upper part of the gel (greater than eighty nucleotides in length) contain relatively homogeneous fragments while those nearer the bottom of the gel contain several components.

Table 2. ANALYSIS OF THE NUCLEOTIDE SEQUENCE OF OLIGONUCLEOTIDE B

(*a*) Pancreatic ribonuclease digestion products	(*b*) CMCT blocked pancreatic ribonuclease digestion products	(*c*) U_2 ribonuclease digestion products	Deduced sequence
AAAU	UAC	A	UACUUAAAUAUG
AU	(U_2,AAAU,AU)G	AA	
AC		2UA	
3U		(C,U_2)A	
G		(C,U_2)AA	
		UG	

Oligonucleotide B was eluted from the homochromatography fingerprint (Fig. 6). After drying off under vacuum the following digestion procedures were used. (*a*) A sample was incubated for 2 h at 37° C with 10 μl. of a 0·2 mg/ml. solution of pancreatic ribonuclease dissolved in 0·01 M *tris*-HCl, *p*H 7·4, containing 0·001 M EDTA. (*b*) A further sample was incubated overnight at 37° C with 10 μl. of 100 mg/ml. solution of CMCT dissolved in 0·01 M *tris*-HCl, *p*H 8·9, containing 0·001 M EDTA. 10 μl. of a 0·2 mg/ml. solution of pancreatic ribonuclease dissolved in 0·05 M *tris*-HCl, *p*H 7·4, 0·001 M EDTA was then added, and the mixture incubated for a further 2 h at 37° C. (*c*) A third sample was incubated overnight at 37° C with 10 μl. of a solution of ribonuclease U_2 containing 1 unit/ml. dissolved in 0·05 M sodium acetate buffer, *p*H 4·5, containing 0·002 M EDTA and 0·1 mg/ml. bovine serum albumin. The products of these procedures were fractionated and characterized as described for oligonucleotide A in Table 1.

RNA Fragment containing Oligonucleotide A

To determine which of the fragments from the gel contained oligonucleotide A, the large T1 ribonuclease products from each fingerprint were digested with pancreatic ribonuclease; oligonucleotide A was found in the fingerprints of the material from bands 19 and 21. A fingerprint of band 19 is shown in Fig. 6. Another large nucleotide (B) was also present in these bands and by the procedures described its sequence was shown to be UACUUAAAUAUG (Table 2). This sequence can also code for three amino-acid sequences and one of these is Tyr-Leu-Asn-Met, which is found in the coat protein just before the sequence coded for by nucleotide A. Band 21, which also contained nucleotides A and B, was subjected to further analysis. Its position on the acrylamide gel suggested that it contained a fragment in the size range of 50–60 residues. The fingerprint of the material from the band was, however, considerably more complex than one would expect for a single component of this size

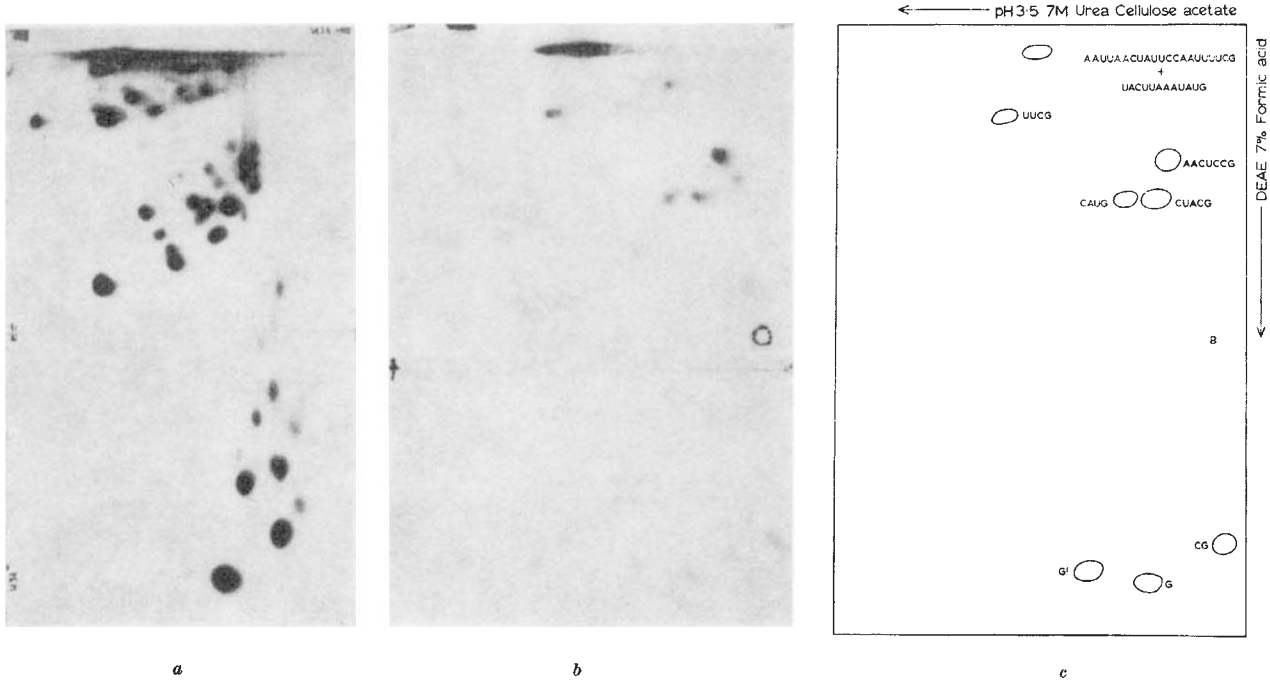

Fig. 7. *a*, A two-dimensional fractionation of a ribonuclease T_1 digest of gel band 21. *b*, Autoradiograph of ribonuclease T_1 two-dimensional fingerprint of the partially purified component from band 21 containing oligonucleotide A. *c*, Diagram of fingerprint in *b* to show the structure of the nucleotides present in the fragment from the coat protein cistron.

Table 3. ANALYSIS OF NUCLEOTIDE SEQUENCES OF REMAINING T_1 RIBONUCLEASE OLIGONUCLEOTIDES IN THE FRAGMENT OF R17 RNA (FIG. 7)

(*a*) Pancreatic ribonuclease digestion products	(*b*) CMCT blocked pancreatic ribonuclease digestion products	(*c*) U_2 ribonuclease digestion products	Deduced sequence
AAC	AAC	A	AAC(UC,C)G*
U	C	AA	
2C	UC	$(C_2,U)G$	
G	G		
AC	C	(C,U)A	CUACG
U	UAC	CG	
C	G		
G	. .		
2U	UUC	—	UUCG
C	G		
G			
AU	C . .	CA	CAUG
C	AUG	UG	
G			
	—	—	CG
	—	—	G

The oligonucleotides from the two-dimensional fingerprint shown in Fig. 7*a* were eluted[10,11]. Samples were subjected to the following procedures. (*a*) Incubation for 30 min at 37° C with 10 μl. of a 0·1 mg/ml. solution of pancreatic ribonuclease in 0·01 M *tris*-HCl, *p*H 7·4, containing 0·001 M EDTA. (*b*) Incubation overnight at 37° C with 10 μl. of a 20 mg/ml. solution of CMCT dissolved in 0·01 M *tris*-HCl, *p*H 8·9, containing 0·001 M EDTA, followed by the addition of 10 μl. of a 0·1 mg/ml. solution of pancreatic ribonuclease in 0·05 M *tris*-HCl, *p*H 7·4, 0·001 M EDTA, and incubation for a further 30 min at 37° C. (*c*) Incubation for 2 h at 37° C with 10 μl. of a 0·1 unit/ml. solution of U_2 ribonuclease in 0·05 M sodium acetate buffer, *p*H 4·5, containing 0·002 M EDTA and 0·1 mg/ml. bovine serum albumin. The products of these digestion procedures were fractionated and characterized by the usual methods.

* The sequence of this oligonucleotide could not be unambiguously assigned with the information presented in the table. In addition, a partial snake venom diesterase digestion was performed: a sample of the oligonucleotide was incubated for 30 min at 37° C with 10 μl. of a 0·1 mg/ml. solution of bacterial alkaline phosphatase dissolved in 0·05 M *tris*-HCl, *p*H 8·9, containing 0·01 M $MgCl_2$ and 1 mg/ml. carrier s-RNA. Following this, 10 μl. of a 0·05 mg/ml. solution of snake venom diesterase in the same buffer was added and digestion allowed to continue for 0 min (a control), 15 min and 30 min at room temperature. The digests were applied to DEAE-paper and the products fractionated by electrophoresis at *p*H 1·9. The following partial digestion products were identified: $(A_2C_2U)C_{OH}$, $(A_2CU)C_{OH}$, $(A_2C)U_{OH}$.

The sequence was therefore shown to be AACUCCG.

and therefore we concluded that band 21 contained a mixture of at least two—and probably more—fragments of approximately the same size. In an attempt to separate the component containing oligonucleotide A from the others the material from band 21 was subjected to electrophoresis on cellulose acetate in 7 M urea (*p*H 3·5) at 6 kV for 3 h. After the run the RNA fragments were blotted on to DEAE-paper and an autoradiograph was made. Three main components were evident and each was eluted and digested with T_1 ribonuclease. The digests were fractionated by ionophoresis on the standard two-dimensional system: cellulose acetate in 7 M urea (*p*H 3·5) followed by DEAE-paper in 7 per cent formic acid[10]. Fig. 7*b* shows the fingerprint of the component containing nucleotides A and B (these nucleotides do not separate in this system) compared with the fingerprint of the unfractionated band 21 (Fig. 7*a*). The purification was still not complete but was sufficient to show that the spots indicated in Fig. 7*c* all come from the same fragment. The sequences of the small oligonucleotides from these spots were determined by standard procedures (Table 3).

To determine the order in which the oligonucleotides occur in the fragment, the RNA was partially digested with T_1 ribonuclease. The partial digestion products were fractionated using electrophoresis on cellulose acetate in 7 M urea (*p*H 3·5) as a first dimension, followed by either electrophoresis on DEAE-paper in 7 per cent formic acid or thin layer homochromatography as a second dimension. The former system was used to isolate the smaller products and the latter to isolate the larger ones. Each partial digestion product was then characterized by further digestion with T_1 ribonuclease and pancreatic ribonuclease.

The analysis of the partial digestion products (Table 4) permits the derivation of the complete sequence of the fragment, which is shown in Fig. 8. It can be seen that this sequence of fifty-seven nucleotides is related by the genetic code to the sequence of amino-acids in positions 81–99 of the coat protein[9].

Table 4. ANALYSIS OF PARTIAL T_1 RIBONUCLEASE DIGESTION PRODUCTS FROM THE FRAGMENT OF R17 RNA

	Ribonuclease T_1 digestion products	Pancreatic ribonuclease digestion products	Deduced sequence
$T_{II}17$	UUCG	GU	CGUUCG
	CG	G	
		2C	
		U	
T_I3	CAUG	GGC	CAUGGCGUUCG
	UUCG	GU	
	CG	AU	
	G	G	
		2C	
		U	
$T_{II}11$	CUACG	GAAC	CUACGAACUCCG
	AACUCCG	G	
		AC	
		2U	
		3C	
$T_{II}3$	AAUUAACUAUUCCAAUUUUCG	GAAC	AAUUAACUAUUCCAAUUUUCGCUACGAACUCCG
	AACUCCG	2AAU	
	CUACG	AC	
		GC	
		AAC	
		G	
		AC	
		U	
		≡	
		C	
		≡	
T_I6	AAUUAACUAUUCCAAUUUUCG	GGAAU	CAUGGCGUUCGUACUUAAAUAUGGAAUUAACUAUUCCAAUUUUCGCUACG
	UACUUAAAUAUG	AAAU	
	UUCG	GGC	
	CUACG	2GU	
	CAUG	AAU	
	CG	3AU	
	2G	GC	
		2AC	
		U	
		≡	
		C	
		≡	

A sample of the material from gel band 21 was digested with 1/500 of its weight of T_1 ribonuclease for 15 min at 0° C in about 3 μl. of 0·01 M *tris*-HCl, *p*H 7·4. The partial digestion products were fractionated (see text) and then characterized by further digestion with both T_1 ribonuclease and pancreatic ribonuclease. The structure of each nucleotide was deduced from its degradation products and from the structures listed above it in the table. Other partial degradation products were present which did not give rise to the T_1 ribonuclease digestion products listed in Table 3 and were therefore derived from the other components of gel band 21.

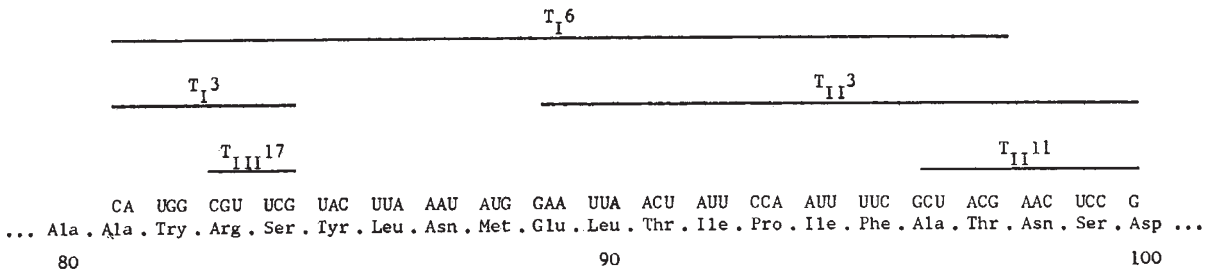

Fig. 8. Nucleotide sequence of the fragment from the coat protein cistron of R17 RNA, indicating partial T_1 digestion products and the corresponding amino-acid sequence of the coat protein.

This is the first time that a sequence from a messenger RNA has been determined by chemical means and shown to correspond to the sequence of amino-acids in the protein for which it codes; the results can be regarded as one of the most direct confirmations of the correctness of the genetic code. It is also of interest to see which codons are actually used by this bacteriophage. Table 5 shows the genetic code, in which the codons found in the above sequence are indicated by underlining the amino-acids concerned. Six amino-acids are found twice in the sequence. Two of these (Leu and Ile) are specified both times by the same codon; however, the other four (Thr, Ser, Asn, Ala) are coded for by two different codons. The data are not sufficient to make any generalizations but at least it may be concluded that the code used by the bacteriophage is degenerate.

Table 5. THE GENETIC CODE

2nd→ / 1st ↓	U	C	A	G	3rd ↓
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	—	—	A
	Leu	Ser	—	Try	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Codons found in the nucleotide sequence of the fragment from the coat protein cistron of R17 RNA have been indicated by underlining the amino-acids concerned.

Secondary Structure of the Fragment

An interesting feature of the sequence is that it can be written in the form of a simple loop showing considerable base-pairing (Fig. 9). Of the twenty-four pairs in this structure nineteen are complementary. This is very unlikely to occur by chance and therefore we believe that the sequence most probably occurs in a double helical configuration in the virus. In this structure all the guanylate residues in the sequence are involved in base pairs and would thus be expected to be resistant to T_1 ribonuclease. This would explain the presence of this fragment in the partial digest of the whole molecule. The unexpected specificity of the partial hydrolysis of R17 RNA suggests that other such highly ordered base-paired structures exist in the RNA; these may be important in the packing of the RNA into the virus particle and may also be involved in the regulation of cistron expression.

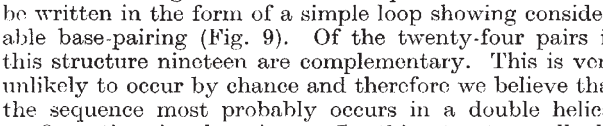

Fig. 9. Secondary structure of the fragment from the coat protein cistron of R17 RNA.

It thus appears that the sequence of a messenger RNA, at least in phage RNA, is determined not only by the need to specify an amino-acid sequence but also by its need to assume a particular secondary structure. In Fig. 9 the phasing of the codons is indicated by dots. It can be seen that the third positions do not come opposite one another. Codons that differ only in the third position often code for the same amino-acid. Thus mutations occurring in two-thirds of the base pairs could change the RNA secondary structure without altering the amino-acid sequence of the protein that is synthesized. It may be that this is one of the functions of the degeneracy of the code.

Because protein biosynthesis depends on the recognition of codons by the anticodon on *t*RNAs, it seems that the messenger RNA must be single-stranded during translation. Thus the finding of a double-stranded structure in a messenger RNA suggests that the protein-synthesizing mechanism must be capable of unfolding such a structure. Similarly the phage RNA synthetase must be able to unfold the RNA during transcription.

We thank Mr A. Coulson for his skilful technical assistance, Dr G. G. Brownlee for much helpful advice during the course of this work and Professor F. Egami for a gift of ribonuclease U_2. J. M. A. thanks the Helen Hay Whitney Foundation for a fellowship and P. G. N. J. the Medical Research Council for a scholarship.

Received July 28, 1969.

[1] Gesteland, R. F., and Boedtker, H., *J. Mol. Biol.*, **8**, 496 (1964).
[2] Dahlberg, J. E., *Nature*, **220**, 548 (1968).
[3] De Wachter, R., and Fiers, W., *J. Mol. Biol.*, **30**, 507 (1967).
[4] Weith, H. L., and Gilham, P. T., *J. Amer. Chem. Soc.*, **89**, 5473 (1967).
[5] Roblin, R., *J. Mol. Biol.*, **31**, 51 (1968).
[6] De Wachter, R., Verhassel, J-P., and Fiers, W., *FEBS Lett.*, **1**, 93 (1968).
[7] Min Jou, W., and Fiers, W., *J. Mol. Biol.*, **40**, 187 (1969).
[8] Gussin, G. N., *J. Mol. Biol.*, **21**, 435 (1966).
[9] Weber, K., *Biochemistry*, **6**, 3144 (1967).
[10] Sanger, F., Brownlee, G. G., and Barrell, B. G., *J. Mol. Biol.*, **13**, 373 (1965).
[11] Brownlee, G. G., and Sanger, F., *J. Mol. Biol.*, **23**, 337 (1967).
[12] Brownlee, G. G., Sanger, F., and Barrell, B. G., *J. Mol. Biol.*, **34**, 379 (1968).
[13] Goodman, H. M., Abelson, J., Landy, A., Brenner, S., and Smith, J. D., *Nature*, **217**, 1019 (1968). Dube, S. K., Marcker, K. A., Clark, B. F. C., and Cory, S., *Nature*, **218**, 232 (1968). Cory, S., Marcker, K. A., Dube, S. K., and Clark, B. F. C., *Nature*, **220**, 1039 (1968). Barrell, B. G., and Sanger, F., *FEBS Lett.*, **3**, 275 (1969). Yaniv, M., and Barrell, B. G., *Nature*, **222**, 278 (1969).
[14] Brownlee, G. G., and Sanger, F., *Europ. J. Biochem.* (in the press).
[15] Gilham, P. T., *J. Amer. Chem. Soc.*, **84**, 687 (1962).
[16] Arima, T., Uchida, T., and Egami, F., *Biochem. J.*, **106**, 609 (1968).
[17] Gould, H. J., *J. Mol. Biol.*, **29**, 307 (1967).
[18] Peacock, A. C., and Dingman, C. W., *Biochemistry*, **6**, 1818 (1967).
[19] Loening, U. E., *Biochem. J.*, **102**, 25 (1967).
[20] Gould, H. J., Pinder, J. C., and Matthews, H. R., *Analyt. Biochem.*, **29**, 1 (1969).