Nucleotide Sequence from the 5′ End to the First Cistron of R17 Bacteriophage Ribonucleic Acid†

Jerry M. Adams,*‡ Pierre-F. Spahr, and Suzanne Cory‡

ABSTRACT: The complete sequence of 117 nucleotides at the 5′ end of R17 bacteriophage RNA has been determined using radiochemical techniques. ^{32}P-labeled R17 RNA was partially digested with T1 ribonuclease and with ribonuclease IV (an endonuclease partially purified from *Escherichia coli* MRE 600) and the digests were fractionated by electrophoresis on large slabs of polyacrylamide gel. To identify 5′-terminal fragments, a portion of the RNA extracted from each gel band was hydrolyzed with alkali and the hydrolysates were fractionated by electrophoresis on DEAE paper in 7% formic acid. This system provides a rapid, convenient method for separating the 5′-terminal residue, pppGp, from nucleoside monophosphates. Five 5′-terminal fragments identified by the presence of pppGp were purified further by electrophoresis on cellulose acetate. The nucleotide sequences of the fragments, which proved to be 74, 79, 92, 109, and 117 nucleotides long, were studied by characterizing the products of complete digestion with T1 ribonuclease and with pancreatic ribonuclease. Partial digestion of the fragments with these two enzymes was then used to establish overlaps between the complete digestion products. The nucleotide sequence established for the first 117 residues at the 5′ end of R17 RNA is pppG-G-G-U-G-G-G-A-C-C-C-C-U-U-U-C-G-G-G-G-U-C-C-U-G-C-U-C-A-A-C-U-U-C-C-U-G-U-C-G-A-G-C-U-A-A-U-G-C-C-A-U-U-U-U-U-A-A-U-G-U-C-U-U-U-A-G-C-C-G-A-G-A-C-G-C-U-A-C-C-A-U-G-G-C-U-A-U-C-G-C-U-G-U-A-G-G-U-A-G-C-C-G-G-A-A-U-U-C-C-A-U-U-C-C-U-A-Gp.

The sequence reported here overlaps the known sequence of the ribosomal binding site at the beginning of the first cistron of R17 RNA, the A protein cistron, so that the complete sequence is now established for the first 145 residues in R17 RNA. The initiation codon for the A protein cistron is preceded by an untranslated nucleotide sequence 129 residues long. One possible secondary structure for this region includes a loop containing the initiation codon for the A protein cistron. Homology exists between the 5′-terminal sequence of R17 RNA and that of f2, MS2, and Qβ RNAs, suggesting that particular sequences in this untranslated region have some essential function.

The simplicity and small size of the RNA bacteriophages make it likely that these viruses will be the first whose properties can be understood in complete chemical detail (for a review, see Stavis and August, 1970). Phages such as R17, MS2, and f2 contain a single strand of RNA about 3400 nucleotides long which apparently codes for only three proteins: the phage coat protein, the A protein (a minor component of the phage particle), and the phage-coded component of RNA synthetase, the enzyme which copies the phage RNA (Kamen, 1970; Kondo *et al.*, 1970). In the last 2 years a number of long nucleotide sequences have been determined in the RNAs from these closely related phages (Adams *et al.*, 1969; Steitz, 1969; Adams and Cory, 1970; Gupta *et al.*, 1970; Jeppesen *et al.*, 1970a; Nichols, 1970; Cory *et al.*, 1970; Ling, 1971; Nichols and Robertson, 1971; Min Jou *et al.*, 1971) and in that from the very different phage Qβ (Billeter *et al.*, 1969; Hindley and Staples, 1969; Goodman *et al.*, 1970). One of the most significant conclusions from this work is that the nucleotide sequence near the 5′ end of these viral messengers does not code for protein. Thus, Billeter *et al.* (1969) found that no initiation codon occurs within the first 62 residues at the 5′ end of Qβ RNA. Moreover, by comparing the sequence of a 5′-terminal fragment of R17 RNA with the known nucleotide sequence (Steitz, 1969) at the beginning of the first cistron, that coding for the A protein (Jeppesen *et al.*, 1970b), we concluded that at least 91 nucleotides at the 5′ end of R17 RNA are not translated (Adams and Cory, 1970). Recently De Wachter *et al.* (1971) have reported a partially completed sequence to position 130 in MS2 RNA and have concluded that the untranslated region at the 5′ end of that molecule is 129 residues long.

We have isolated 5′-terminal fragments of R17 RNA made by partial digestion with T1 ribonuclease (Adams *et al.*, 1969) and with RNase IV (Cory *et al.*, 1970), an endonuclease partially purified from MRE 600 (Gesteland and Spahr, 1969). The sequence of these fragments has been studied by radiochemical methods (Sanger *et al.*, 1965) and we report here the determination of the complete sequence of 117 residues at the 5′ end of R17 RNA. This sequence overlaps that reported by Steitz (1969) for the ribosomal binding site at the beginning of the A protein cistron, so that the sequence is now established for the first 145 residues in R17 RNA.

Materials and Methods

T1 ribonuclease and U2 ribonuclease were obtained from Sankyo Ltd., Tokyo; pancreatic ribonuclease and snake venom phosphodiesterase, from Worthington Biochemical Corp. The venom phosphodiesterase was purified further by the method of Keller (1964). Spleen phosphodiesterase was kindly provided by Dr. A. Bernardi from this laboratory. Poly(U) and poly(C) were obtained from Miles Laboratories and *E. coli* tRNA from General Biochemicals.

Identification of 5′-Terminal Fragments of R17 RNA. The

† From the Département de Biologie Moléculaire, University of Geneva, Geneva, Switzerland. This work was supported by a Helen Hay Whitney Fellowship (J. M. A.), by a Rothmans Fellowship (S. C.), and by a research grant from the Fonds National Suisse de la Recherche Scientifique.

‡ Present address: The Walter and Eliza Hall Institute of Medical Research, P. O. Royal Melbourne Hospital, Melbourne, Victoria 3050, Australia.

* To whom correspondence should be addressed.

preparation of ^{32}P-labeled R17 RNA, its partial digestion with T1 ribonuclease and ribonuclease IV, and the fractionation of the partial digests by electrophoresis on flat slabs of polyacrylamide gel have been described previously (Gesteland and Spahr, 1970; Adams *et al.*, 1969; Cory *et al.*, 1972). The gel slabs were subjected to autoradiography, the prominent bands cut out, and the RNA recovered by phenol extraction (Cory *et al.*, 1972). The presence of pppGp in alkaline hydrolysates (Roblin, 1968a) was used to identify 5′-terminal fragments. A fraction of the RNA from each gel band was hydrolyzed with 0.25 M NaOH for 15 hr at 37° and the hydrolysates were loaded directly onto DEAE paper; the paper was wet with 7% formic acid and subjected to electrophoresis in Varsol-cooled tanks for about 18 hr at 1 kV.

Fingerprinting the Fragments. Complete digestion of 5′-terminal fragments with T1 ribonuclease or pancreatic ribonuclease required stronger conditions than those used for other fragments of R17 RNA. The RNA was digested for 30 min at 37° in 0.02 M Tris-Cl (pH 7.5)–0.002 M Na_2EDTA, with an enzyme to substrate ratio of 1:6. The digests were fractionated by electrophoresis on cellulose acetate at pH 3.5 followed by electrophoresis on DEAE paper in 7% formic acid (Brownlee *et al.*, 1968). EDTA (0.002 M) was included in the pH 3.5 buffer to reduce streaking of oligonucleotides containing pppGp (Dahlberg, 1968); its presence greatly increased their mobility but did not alter that of other oligonucleotides. To improve resolution in the second dimension, 110 cm long sheets of DEAE paper were used and electrophoresis was for 18 hr at 1.3 kV. Autoradiography of the fingerprints and subsequent elution of the products were as described by Sanger *et al.* (1965).

Analysis of Eluted Oligonucleotides. T1 OR PANCREATIC RIBONUCLEASE DIGESTION. The oligonucleotide was incubated in a capillary for 30 min at 37° in 10 μl of a solution containing 0.02 M Tris-Cl (pH 7.5), 0.002 M Na_2EDTA, and 0.1 mg/ml of enzyme and the products were fractionated at pH 3.5 on DEAE paper (Adams *et al.*, 1969).

U2 RIBONUCLEASE DIGESTION. The oligonucleotide was incubated for 2 hr at 37° in 10 μl of a solution containing 0.05 M sodium acetate (pH 4.5), 0.002 M EDTA, 1 μg each of poly(U) and poly(C), 0.1 mg/ml of bovine serum albumin, and 1 unit/ml of U2 ribonuclease. Poly(U) and poly(C) were added to reduce splits occurring after Cp and Up residues, which presumably are due to some enzymic activity present in either the U2 ribonuclease or the albumin.

PANCREATIC RIBONUCLEASE DIGESTION OF CARBODIIMIDE-BLOCKED OLIGONUCLEOTIDES. Oligonucleotides were modified with a carbodiimide and digested with pancreatic ribonuclease as described previously (Adams *et al.*, 1969). In some experiments the blocked oligonucleotides were then separated directly on paper at pH 3.5 as before. However, oligonucleotides having several U residues often streaked badly, perhaps due to incomplete modification under our conditions. Better results were obtained if the blocking groups were removed after the pancreatic ribonuclease digestion by the following adaptation of the method of Ho and Gilham (1967). *Escherichia coli* tRNA (200 μg) was added to reduce the effects of any ribonuclease activity during the subsequent manipulations. The sample was dried thoroughly, dissolved in 20 μl of diethylamine–formamide (1:4, v/v), and incubated overnight at room temperature. The sample was then dried under vacuum, dissolved quickly in 10–20 μl of ice-cold distilled water, and loaded immediately onto DEAE paper. The products were then fractionated in either 7% formic acid or pyridine acetate pH 3.5 and identified by alkaline hydrolysis.

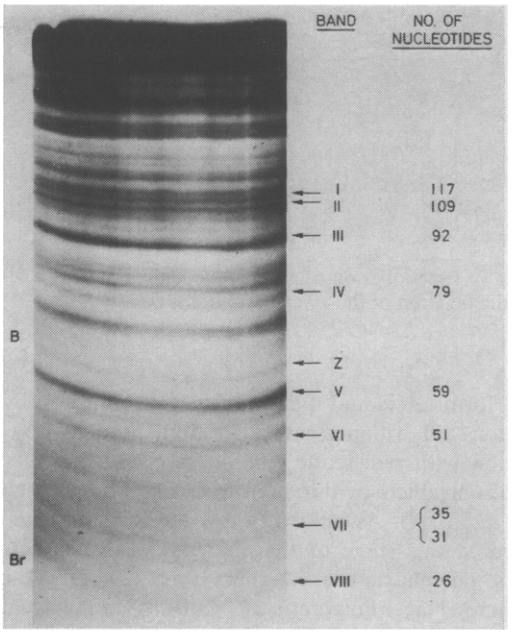


FIGURE 1: Ribonuclease IV digest of ^{32}P-labeled R17 RNA fractionated by electrophoresis on a 12.5% polyacrylamide gel slab. Some of the bands which were cut from the gel are indicated with roman numerals. The size of fragments for which sequences have been determined is also shown. B indicates the position of the xylene cyanol marker dye and Br that of bromophenol blue.

COMPLETE DIGESTION OF 3′-PHOSPHORYLATED OLIGONUCLEOTIDES WITH SNAKE VENOM PHOSPHODIESTERASE. The oligonucleotide was incubated for 2 hr at 37° in 10 μl of a solution containing 0.001 M $MgCl_2$, 0.075 M Tris-succinate (pH 6.0), and 0.15 mg/ml of venom phosphodiesterase. The products were fractionated as described previously (Cory and Marcker, 1970).

Partial Digestion of 5′-Terminal Fragments. Fragment 13a was partially digested for 30 min at 2° in 5 μl of a solution containing 0.02 M Tris-acetate (pH 7.4), 0.02 M $MgSO_4$, and either T1 ribonuclease or pancreatic ribonuclease. The enzyme to substrate ratio (w:w) was either 1:50 or 1:500 for pancreatic ribonuclease and 1:300 for T1 ribonuclease. Fragment IIa was partially digested with pancreatic ribonuclease under the same conditions except that the enzyme to substrate ratio was 1:300 and fragment Ia was partially digested with T1 ribonuclease at a ratio of 1:400.

The partial digestion products were fractionated by the two-dimensional system developed by Brownlee and Sanger (1969): electrophoresis at pH 3.5 on cellulose acetate followed by homochromatography on thin layers of DEAE-cellulose. For the second dimension we used commerical thin layers (Macherey Nagel Polygram, 20 × 40 cm), which gave very good separations. The "homomixture" used in the chromatography was 50 ml of homomixture a plus 50 ml of homomixture b (Brownlee and Sanger, 1969). The oligonucleotides were located by autoradiography, cut out from the chromatogram and eluted by a method developed by D. F. Klessig and J. E. Dahlberg (manuscript in preparation).

Each partial digestion product was subsequently analyzed by a complete digestion with T1 ribonuclease and with pancreatic ribonuclease. The digestion conditions used were 2 hr at 37° with an enzyme to substrate ratio of 1:50, where the amount of RNA was estimated assuming a recovery of 0.23 mg of RNA/cm^2 from the Macherey Nagel thin layer. The digests were fractionated by electrophoresis on DEAE paper

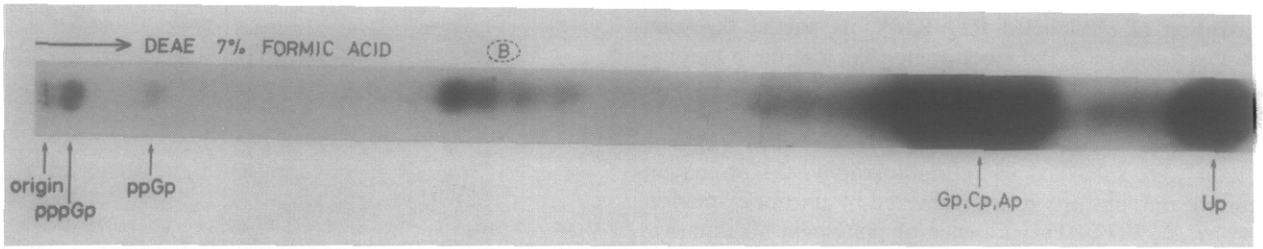


FIGURE 2: Fractionation of an alkaline hydrolysate of a 5′-terminal fragment by electrophoresis on DEAE paper in 7% formic acid. B indicates the position of the xylene cyanol marker dye.

in 7% formic acid and the products were identified by further digestion—T1 ribonuclease digestion products by further digestion with pancreatic ribonuclease and pancreatic ribonuclease products by digestion with T1 ribonuclease (Brownlee *et al.*, 1968). As only very low levels of radioactivity remained at this stage of the analysis, we fractionated these digests on plastic-backed thin layers of DEAE-cellulose (Macherey Nagel Polygram, 20 × 40 cm) in the usual pH 3.5 buffer at 1 kV for about 2 hr. (Plastic backed thin layers can be used in the Varsol-cooled tanks with the racks normally employed for DEAE paper.) This procedure increases the sensitivity of detection of nucleotides because the spots subsequently revealed by autoradiography are only about one-fifth the size of those obtained on DEAE paper. The products were identified by their relative mobilities, which are the same as on DEAE paper at pH 3.5.

Results

Isolation and Identification of 5′-Terminal Fragments. ^{32}P-Labeled R17 RNA was partially digested with either T1 ribonuclease or ribonuclease IV and the digests were fractionated by electrophoresis on large slabs of polyacrylamide gel. A typical fractionation of a ribonuclease IV digest on such a gel is shown in Figure 1. The fractionation of a T1 ribonuclease digest is shown elsewhere (Figure 3 in Sanger *et al.*, 1970).

Since the 5′ end of R17 RNA is guanosine triphosphate, 5′-terminal fragments can be identified by the presence of guanosine tetraphosphate (pppGp) in alkaline hydrolysates (Roblin, 1968a). A portion of the RNA isolated from each gel band was hydrolyzed and the products were fractionated by electrophoresis on DEAE paper in 7% formic acid. In this system (orginally suggested to us by Dr. J. Dahlberg), pppGp is well separated from the four nucleoside monophosphates, as can be seen in Figure 2. The pppGp was eluted from the DEAE paper and its identity was confirmed by its mobility at pH 3.5 (Dahlberg, 1968) and by showing that extensive digestion with snake venom phosphodiesterase (Roblin, 1968a) converted it to pGp and pyrophosphate (Figure 3).

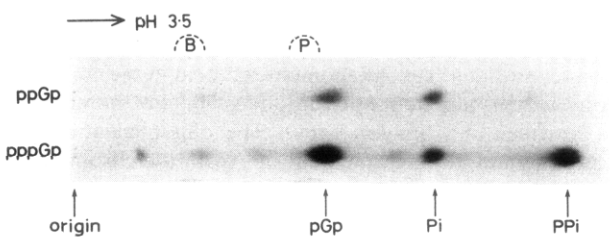


FIGURE 3: Fractionation of venom phosphodiesterase digests of pppGp and ppGp by electrophoresis on Whatman 52 paper at pH 3.5. The oligonucleotides were incubated for 3.5 hr at 37° in 10 μl of a solution containing 0.0001 M $MgCl_2$, 0.075 M Tris-succinate (pH 6.0), and 0.15 mg/ml of venom phosphodiesterase. B indicates the position of the xylene cyanol marker dye and P that of acid fuchsin.

Alkaline hydrolysates of the 5′-terminal fragments also contained significant amounts of ppGp (Figure 2). ppGp was identified by its electrophoretic mobility on the DEAE paper in 7% formic acid, its electrophoretic mobility at pH 3.5 on unmodified paper (Dahlberg, 1968), and by showing that snake venom phosphodiesterase converted it to pGp and orthophosphate (Figure 3). The following two observations suggest that ppGp is produced in large part by partial breakdown of pppGp during the alkaline hydrolysis. (1) When the hydrolysis of a 5′-terminal fragment was allowed to continue for 36 hr instead of the standard 15 hr, almost no pppGp remained and the yield of ppGp greatly increased. (2) When the pppGp was eluted from DEAE paper with 30% (v/v) triethylamine carbonate (pH 9.7) and rehydrolyzed with alkali, about one-third of it was converted to ppGp. However, we have also detected some ppGp in enzymic digests of 5′-terminal fragments. Therefore it remains possible that a fraction of the viral RNA molecules terminate in a diphosphate rather than a triphosphate group.

Some unknown compounds which had mobilities similar to that of the blue dye marker in the 7% formic acid system were also present in alkaline hydrolysates of the 5′-terminal fragments (Figure 2). These compounds presumably are not related to the 5′ terminus since they are present in hydrolysates of every fragment of R17 RNA which we have studied. They may include compounds such as pXp (X being U, C, A, or G), which Roblin (1968a,b) has identified in alkaline hydrolysates of R17 RNA.

Electrophoresis on DEAE paper in 7% formic acid proved to be a very useful method for screening for 5′-end groups, as a number of samples could be run simultaneously (up to 20 samples/sheet) and the hydrolysates could be loaded directly on the paper, which has a relatively high capacity (at least 1 mg of RNA/cm^2). Electrophoresis on unmodified paper at pH 3.5 was not as satisfactory because the pppGp streaked—the streaking can be reduced somewhat by adding EDTA to the buffer (Dahlberg, Z968)—and because the pppGp was sometimes obscured by traces of orthophosphate and residual unhydrolyzed oligonucleotides in the hydrolysates.

5′-Terminal fragments were present in bands I, II, III, and IV from the RNase IV digest (Figure 1) and in band 13 from a T1 ribonuclease digest (see Figure 3 in Sanger *et al.*, 1970). Other fragments were also present in these bands but it was possible to separate the 5′-terminal fragments from the others by electrophoresis on cellulose acetate at pH 3.5 (Adams *et al.*, 1969). All of the 5′-terminal fragments studied had a

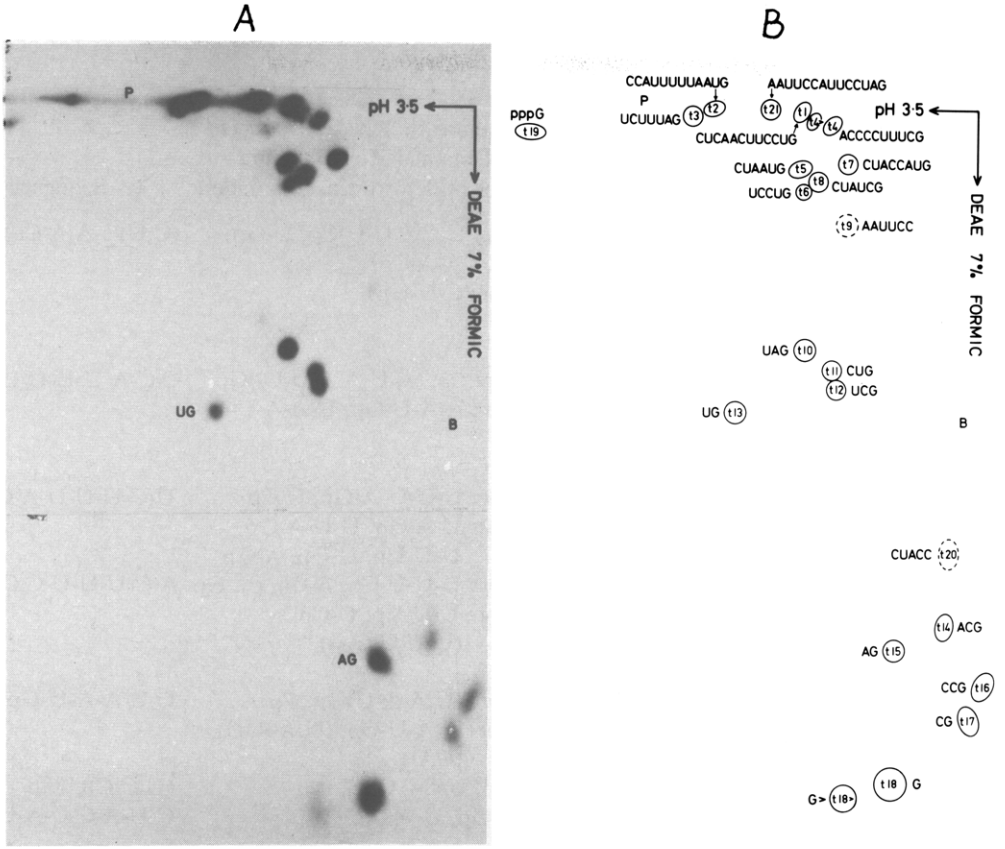


FIGURE 4: Two-dimensional fractionation of complete T1 ribonuclease digest of 5′-terminal fragment Ia (117 residues long). (A) Autoradiograph. (B) Diagram identifying the oligonucleotide which corresponds to each spot (complete circles). The 3′-terminal residue of two of the smaller 5′-terminal fragments (IIa and IVa) is not Gp and therefore the digests of these fragments each contain a product related to but smaller than one of the nucleotides in the fingerprint of the largest fragment shown here: t20 (C-U-A-C-Cp) from fragment IVa, which is related to t7 (C-U-A-C-C-A-U-Gp), and t9 (A-A-U-U-C-Cp) from fragment IIa, which is related to t21 (A-A-U-U-C-C-A-U-U-C-C-U-A-Gp). The positions of t9 and t20 are indicated by broken circles in the diagram. The letter P indicates the position of the pink marker dye (acid fuchsin) on electrophoresis in the first dimension and B the position of the blue marer dye (xylene cyanol) in the second dimension.

somewhat greater mobility in this system than other fragments from the same gel band. Most likely this is due to the presence of the 5′-terminal phosphate residues and to the fact that the 5′-terminal fragments are relatively rich in G and U, which have higher charges at pH 3.5 than A or C. The 5′-terminal fragments purified in this way are subsequently designated by an "a" after the number of the band from which they were isolated.

Sequence Analysis of the Fragments. T1 RIBONUCLEASE DIGESTION PRODUCTS. In the first stage of the sequence analysis, each 5′-terminal fragment was digested to completion with T1 ribonuclease, which splits after guanosine residues, and the products were separated on the standard two dimensional system developed by Sanger *et al.* (1965). The resulting fingerprint of fragment Ia, the largest 5′-terminal fragment that we have studied (117 residues long), is shown in Figure 4A. The accompanying diagram (Figure 4B) identifies the oligonucleotide corresponding to each spot.

Oligonucleotides Gp, C-Gp, A-Gp, C-C-Gp, U-C-Gp, C-U-Gp, and U-A-Gp were identified by their position on the fingerprint (Sanger *et al.*, 1965). The sequences of the other oligonucleotides were established by further enzymic digestion. First, the composition of an oligonucleotide was determined by quantitating the products obtained on digestion with pancreatic ribonuclease, which splits after pyrimidine residues (column 2, Table I). Oligonucleotides containing adenosine were also digested with U2 ribonuclease (Arima *et al.*, 1968), which splits after purine residues (column 3, Table I). Finally, the oligonucleotides were treated with a carbodiimide, which blocks U and G residues (Gilham, 1962), and then digested with pancreatic ribonuclease (column 4, Table I), which can split the blocked oligonucleotide only after Cp residues.

The information obtained from these three procedures allowed us to deduce the complete sequence of all of the oligonucleotides except t1, t4, t6, and t21 (Table I). The data given for oligonucleotide t21 is compatible with either of the sequences represented by A-A-U-U-C-C(A-U,U)C-C-U-A-Gp. The correct sequence is known to be A-A-U-U-C-C-A-U-U-C-C-U-A-Gp from the work of Jeppesen (1971), who has identified this oligonucleotide in a 5′-terminal fragment of R17 RNA about 130 nucleotides long. Further information on the sequences of oligonucleotides t1, t4, and t6 was obtained by overdigesting each of these oligonucleotides with either T1 ribonuclease (Brownlee and Sanger, 1967) or U2 ribonuclease. Digestion with either of these enzymes at a high enzyme to substrate ratio results in many splits in addition to those found under standard conditions. To separate the large number of products obtained, we fractionated these digests on the standard two dimensional system. The products were then identified by their position on the fingerprint together with the information provided by alkaline hydrolysis and/or pancreatic ribonuclease digestion (Table II). The products which were most important for establishing the sequence of t1, t4, and t6 are shown in italics in Table II.

The relative yields of the T1 ribonuclease digestion prod-

TABLE I: Analysis of T1 Ribonuclease Digestion Products.[a]

Nucleotide	Pancreatic RNase Digestion Products	U2 RNase Digestion Products	Pancreatic RNase Products Obtained after Blocking Nucleotide with Carbodiimide[b]	Sequence Deduced
t1	1.0 A-A-Cp, 4.0 Up, 4.2 Cp, 1.1 Gp (average of three experiments)	Major: $(C_3,U_3)Gp$,[c] $(C_2,U)A-Ap$[c] Minor: $(C_2,U)Ap$, C-U-Cp,[d] A-Ap, Ap, (C,U)Gp	Major: U-Cp, U-U-Cp, U-Gp, A-A-Cp, C_p Minor: C-U-Gp	(C,U)C-A-A-C(C,U-U)C-U-Gp
t2	0.9 A-A-Up, 0.9 A-Up, 4.3 Up, 2.1 Cp, 1.1 Gp (average of two experiments)	U-U-U-U-U-A-Ap,[e] C-C-Ap, U-Gp	Major: *Cp*, (A-U,A-A-U,U_4)Gp Minor: A-A-U-Gp, U-A-A-U-Gp	C-C-A-U-U-U-U-U-A-A-U-Gp[f]
t3	0.9 A-Gp, 4.0 Up, 1.2 Cp (average of two experiments)		Major: U-U-U-A-Gp, U-Cp Minor- U-A-Gp, U-U-A-Gp, U-U-Up, U-Up, A-Gp	U-C-U-U-U-A-Gp[g]
t4	1.0 A-Cp, 3.0 Up, 4.0 Cp, 1.0 Gp (average of three experiments)	$(U_3,C_5)Gp$, Ap	Major: U-U-U-Cp, A-Cp, C_p, Gp Minor: U-U-Up, U-Cp, (U-U-U,C_{2-3})	A-C(U-U-U-C,C_3)Gp
t5	A-A-Up, Up, Cp, Gp	(C,U)Ap, U-Gp, Ap	Major: (U,A-A-U)Gp, Cp Minor: A-A-U-Gp, (U,A-A-U), (C,U)	C-U-A-A-U-Gp
t6	*Up*, *Cp*, Gp		U-Gp, U-Cp, Cp	(U-C,C)U-Gp
t7	1.0 A-Up, 1.0 A-Cp, 1.1 Up, 2.2 Cp, 1.1 Gp (average of two experiments)	U-Gp, C-C-Ap, C-U-Ap[d]	A-U-Gp, U-A-Cp, *Cp*	C-U-A-C-C-A-U-Gp
t8	1.0 A-Up, 1.0 Up, 2.0 Cp, 0.9 Gp (average of two experiments)	(C,U)Ap, (C,U)Gp	(A-U,U)Cp, Cp, Gp	C-U-A-U-C-Gp
t9	1.1 A-A-Up, 1.2 Up, 2.1 Cp (average of two experiments)		(A-A-U,U)Cp, Cp	A-A-U-U-C-Cp[h]
t10	A-Gp, Up			U-A-Gp
t11	Up, Cp, Gp		U-Gp, Cp	C-U-Gp
t12	Up, Cp, Gp		U-Cp, Gp	U-C-Gp
t14	A-Cp, Gp			A-C-Gp
t20	1.0 A-Cp, 1.8 Cp, 1.0 Up			C-U-A-C-Cp[i]
t21	1.0 A-A-Up, 1.1 A-Up, 3.0 Up, 3.9 Cp, 0.6 A-Gp (average of three experiments)	$(U_3,C_2)Ap$, $(U_2,C_2)Ap$, Gp, A-Ap (also some Ap)	(A-A-U,U)Cp, (A-U,U,)Cp U-A-Gp, *Cp*	A-A-U-U-C-C(A-U,U)-C-C-U-A-Gp[h]

[a] Conditions for digesting the oligonucleotides are described in Materials and Methods. The yields of the digestion products were either determined by measuring the radioactivity of the paper containing the nucleotide using a Nuclear-Chicago gas-flow counter or simply estimated by visual inspection of the autoradiograph. In the latter case, the product is in italic type if it was present in two molar yield, and in small capital letters if it was in three or more molar yield. [b] The carbodiimide reaction was not always complete in the case of oligonucleotides containing many U residues, and so products ending in Up were frequently isolated in minor yield. [c] The compositions were determined by measuring the yields of the alkaline hydrolysis products in the low-background gas-flow counter. [d] The sequence was deduced by comparing the composition of the oligonucleotide obtained by alkaline hydrolysis with that obtained by complete digestion with snake venom phosphodiesterase (see Cory and Marcker, 1970). This procedure allows one to identify both the 5′- and 3′-terminal nucleotides. [e] When electrophoresed on DEAE paper in 7% formic acid, this product barely moves off the origin, indicating that it must contain 5 Up residues. [f] Confirmation that the 5′-terminal sequence of t2 is C-Cp came from a product isolated by partial pancreatic ribonuclease digestion of fragment 13a: A-A-U-G-C-Cp. [g] Confirmation that the 5′-terminal nucleotide of t3 is Up came from a product isolated by partial pancreatic ribonuclease digestion of fragment 13a: A-A-U-G-U-C-U-U-Up. [h] A-A-Up must be 5′-terminal to overlap the pancreatic ribonuclease product p3 (G-G-A-A-Up). [i] The sequence of t20 is known from the fact that it is a 5′-terminal fragment of t7 (C-U-A-C-C-A-U-Gp).

TABLE II: Analysis of T1 Ribonuclease Digestion Products by Enzymic Overdigestion.[a]

Nucleotide	Sequence Deduced from Preceding Anal.	Products Obtained by Enzymic Overdigestion[b]	Sequence Deduced
t1	(C,U)C-A-A-C(C,U-U)C-U-Gp	With T1 RNase: *C-C-Up*,[c] *C-U-Cp*,[c] *U-C-C-Up*,[c] U-Gp, C-U-Gp, *C-C-U-Gp*, *U-C-C-U-Gp*, (U_3,C_3), (A-A-C,U_3,C_2)	C-U-C-A-A-C-U-U-C-C-U-Gp[d]
t2	C-C-A-U-U-U-U-U-A-A-U-Gp	With T1 RNase: U-Gp, A-U-Gp, A-A-U-Gp	C-C-A-U-U-U-U-U-A-A-U-Gp
t4	A-C(U-U-U-C,C_3)Gp	With T1 RNase: *C-Gp*, U-Up, U-U-Up, (U-U,C), (U-U-U,C), (U-U-U,C_3), (U-U,C_4), (U-U-U,C_4), *C-U-U-U-C-Gp*, *A-C-Cp*, *A-C-C-Cp*, (A-C-C-C,C,U), (A-C-C-C,C,U-U-U), *A-C-C-C-C-U-Up* With U2 RNase: *C-Gp*, *U-U-U-C-Gp*, *C-C-U-U-U-C-Gp*, *C-C-C-U-U-U-C-Gp*, *C-C-C-C-U-U-U-C-Gp*, (U-U-U,C), (U-U-U,C_2), (U-U-U,C_3), (U-U-U,C_4), C-C-C-C-U-U-U-Cp	A-C-C-C-C-U-U-U-C-Gp
t6	(U-C,C)U-Gp	With T1 RNase: *C-C-Up*,[e] U-Gp, *C-U-Gp*	U-C-C-U-Gp

[a] The conditions for overdigestion with T1 ribonuclease were the following: 30 min at 37° in 5 μl of a solution containing 0.02 M Tris-acetate (pH 7.5), 0.002 M Na_2EDTA, 0.5 μg of *E. coli* tRNA (General Biochemicals) and 2 μg T1 ribonuclease. The conditions for overdigestion with U2 ribonuclease were the following: 3.5 hr at 37° in 8 μl of a solution containing 0.05 M sodium acetate (pH 4.5), 0.002 M EDTA, 0.1 mg/ml of bovine serum albumin, and 1 unit/ml of U2 ribonuclease. No carrier tRNA was added. [b] The digests were fractionated by electrophoresis at pH 3.5 on cellulose acetate followed by electrophoresis on DEAE paper in 7% formic acid. The composition of each product was determined by its position on the fingerprint and by alkaline hydrolysis and/or pancreatic ribonuclease digestion. The sequence shown for each product was deduced from its composition together with the information previously deduced about the sequence of the oligonucleotide from which it was derived. [c] The sequence was deduced by comparing the composition of the oligonucleotide obtained by alkaline hydrolysis with that obtained by complete digestion with snake venom phosphodiesterase (see Cory and Marcker, 1970). This procedure allows one to identify both the 5′- and 3′-terminal nucleotides. [d] Confirmation that the 5′-terminal nucleotide of t1 is Cp came from two products isolated by partial pancreatic ribonuclease digestion of fragment 13a: C-G-G-G-U-C-C-U-G-C-U-Cp and C-G-G-G-G-U-C-C-U-G-C-U-C-A-A-Cp. [e] The sequence of C-C-Up was deduced from its mobility and from the products obtained on complete digestion with snake venom phosphodiesterase.

ucts from each of the five 5′-terminal fragments are listed in Table III. The molar yield of oligonucleotides determined experimentally agreed fairly well with that expected from the final structure, with a few exceptions. For example, the commonly occurring oligonucleotides A-Gp, C-Gp, and U-Gp were sometimes present in yield higher than expected, probably because the 5′-terminal fragments were sometimes slightly contaminated by fragments from other regions of R17 RNA.

PANCREATIC RIBONUCLEASE DIGESTION PRODUCTS. The 5′-terminal fragments were also digested to completion with pancreatic ribonuclease and the products were fractionated by electrophoresis in two dimensions. The fingerprint of fragment IIa (109 residues long) is shown in Figure 5A. The accompanying diagram (Figure 5B) identifies the oilgonucleotide corresponding to each spot. Oligonucleotides Up, Cp, A-Cp, A-A-Cp, G-Cp, A-Up, A-A-Up, and G-Up were identified by their position on the fingerprint (Sanger *et al.*, 1965). The sequences of the remaining oligonucleotides were determined by complete digestion with T1 ribonuclease, and, in two cases, by partial digestion with the 5′-exonuclease spleen phosphodiesterase (Sanger *et al.*, 1965). The results of these analyses are summarized in Table V. The relative yields of the pancreatic ribonuclease products from each of the five 5′-terminal fragments are shown in Table VI.

PARTIAL ENZYMIC DIGESTION. Partial enzymic digestion was used to order the sequences of the complete digestion products. Fragment 13a (74 residues long) was studied first. It was partially digested with T1 ribonuclease and with pancreatic ribonuclease and the products in each case were separated in the two-dimensional system developed by Brownlee and Sanger (1969) for the fractionation of longer oligonucleotides. Each partial digestion product was then isolated and analyzed by complete digestion with T1 ribonuclease and with pancreatic ribonuclease. The information provided by these analyses, which is given in Table IV, allowed us to deduce the sequences of 25 partial digestion products from fragment 13a. These sequences overlap as shown in Figure 6A and establish an unambiguous sequence for this fragment.

Once the sequence of the first 74 residues was known we could deduce most of the sequence up to position 117 simply by comparing the yields of the various products in successively longer 5′-terminal fragments (Tables III and V). For example, a T1 ribonuclease digest of fragment IVa (79 nucleotides long) contains C-U-A-C-Cp in addition to the oligonucleotides in the digest of fragment 13a (74 residues long). Since this oligonucleotide does not end in Gp, it must be the 3′ terminus of fragment IVa. To establish the complete sequence of the larger 5′-terminal fragments, we characterized partial digestion products from fragment IIa (109 residues long) and Ia (117 residues long). Many of these partial products were the same as those found previously but 18 came

TABLE III: T1 Ribonuclease Digestion Products from 5′-Terminal Fragments.

Nucleotide	Sequence	Molar Yield in Fragment 13a (74 Residues)		Molar Yield in Fragment IVa (79 Residues)		Molar Yield in Fragment IIIa (92 Residues)		Molar Yield in Fragment IIa (109 Residues)		Molar yield in Fragment Ia (117 Residues)	
		Experimentally Determined (2 expt)[a]	Expected from Final Structure	Experimentally Determined (1 expt)[a]	Expected from Final Structure	Experimentally Determined (3 expt[a])	Expected from Final Structure	Experimentally Determined (2 expt)[a]	Expected from Final Structure	Experimentally Determined (3 expt)[a]	Expected from Final Structure
t1	C-U-C-A-A-C-U-U-C-C-U-Gp	0.9	1	0.7	1	0.9	1	1.0	1	0.8	1
t2	C-C-A-U-U-U-U-U-A-A-U-Gp	0.8	1	0.9	1	0.9	1	0.9	1	1.0	1
t3	U-C-U-U-A-Gp	1.2	1	1.0	1	1.3	1	1.2	1	1.1	1
t4	A-C-C-C-C-U-U-U-C-Gp	0.8	1	1.3	1	0.7	1	0.8	1	0.6[e]	1
t5	C-U-A-A-U-Gp	1.2	1	0.9	1	1.0	1	1.1	1	1.1	1
t6	U-C-C-U-Gp	1.3	1	1.1	1	0.9	1	1.3	1	0.5[e]	1
t7	C-U-A-C-C-A-U-Gp	0	0	0	0	0.8	1	0.9	1	0.9	1
t8	C-U-A-U-C-Gp	0	0	0	0	1.0	1	1.1	1	0.9	1
t9	A-A-U-U-C-Cp	0	0	0	0	0	0	0.6	1	0	0
t10	U-A-Gp	0	0	0	0	0	0	2.2	2	2.2	2
t11	C-U-Gp	0	0	0.3	0	1.1	1	1.4	1	1.2	1
t12	U-C-Gp	1.1	1	1.1	1	1.1	1	1.1	1	1.3	1
t13	U-Gp	1.4	1	2.2[d]	1	1.3	1	2.3[d]	1	1.2	1
t14	A-C-Gp	0.5	1	0.9	1	1.1	1	0.8	1	0.9	1
t15	A-Gp[c]	1.9	2	2.9[d]	2	2.6[d]	2	3.0[d]	2	3.6[d]	2
t16	C-C-Gp	0	0	0	0	0	0	1.2	1	0.9	1
t17	C-Gp	1.1	1	1.7[d]	1	1.2	1	2.4[d]	1	1.3	1
t18	Gp	6.2	7	10.6[d]	7	6.3	8	11.6[d]	10	10.0	10
t19	pppGp	0.3[b]	1	Not measured	1	0.8	1	Not measured	1	0.5[d]	1
t20	C-U-A-C-Cp	0	0	0.7	1	0	0	0	0	0	0
t21	A-A-U-U-C-C-A-U-U-C-C-U-A-Gp	0	0	0	0	0	0	0	0	0.9	1

[a] Yields were determined by measuring the radioactivity of the paper containing the nucleotide using a Nuclear-Chicago low-background gas-flow counter. [b] The yield of pppGp is low because this oligonucleotide streaks extensively during the electrophoresis on cellulose acetate (see text). [c] The yield of A-Gp was usually higher than expected from the final structure (approximately 3 moles rather than 2). However, only 2 moles can come from the 5′-terminal fragments because there are only two pancreatic ribonuclease digestion products (p6 and p7) containing GAG sequences. The high yield must then result from some contamination by other fragments rich in GAG sequences. [d] The yields of these commonly occurring nucleotides are presumably high because of some contamination by RNA molecules other than the 5′-terminal fragment. The possibility that an extra mole of these nucleotides occurs in the 5′-terminal fragments is excluded by the analysis of partial digestion products given in Table IV. [e] The yields of t4 and t6 were sometimes low. Most likely this was due to incomplete digestion of the strongly hydrogen-bonded loop occurring at the 5′-terminus of R17 RNA (see text).

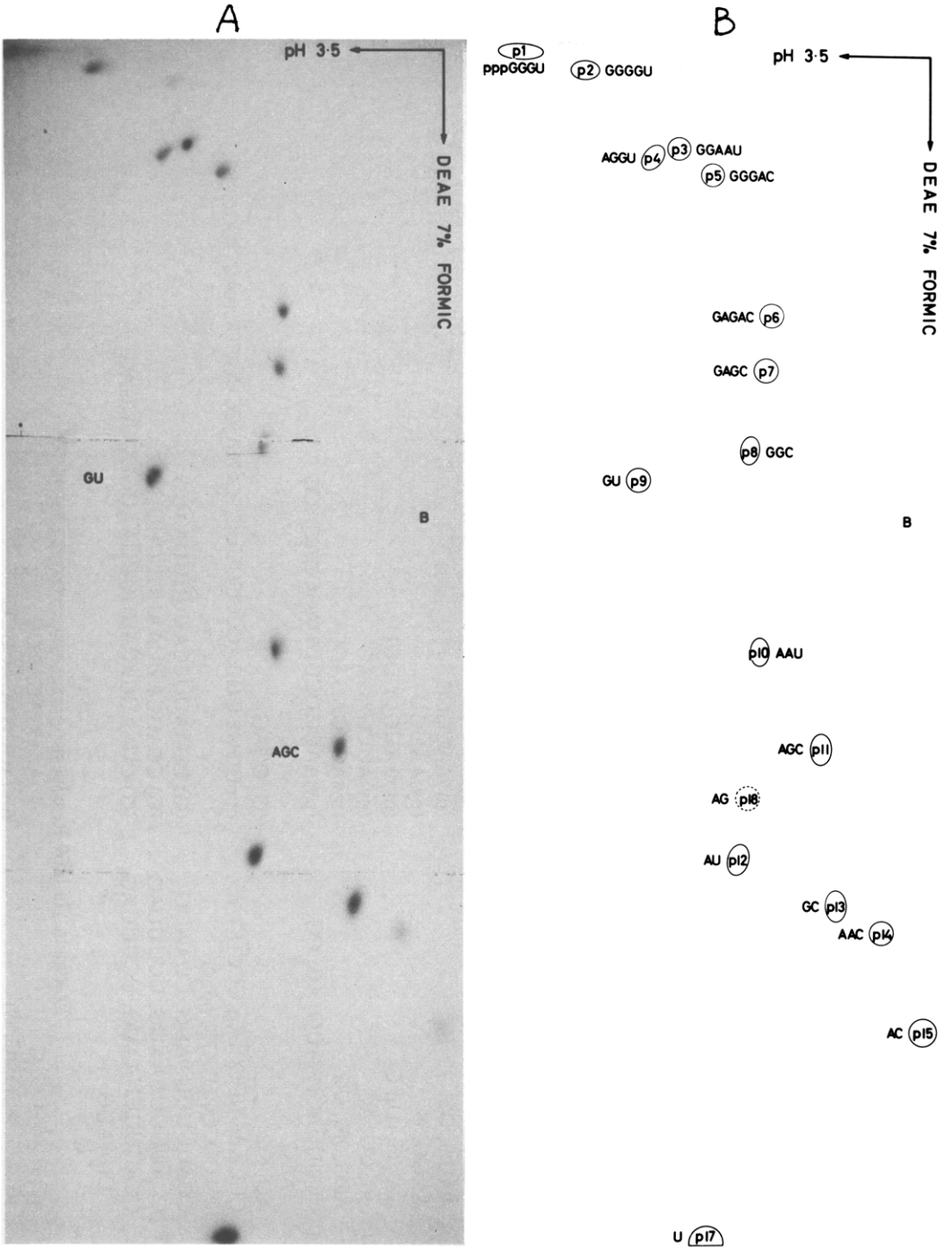


FIGURE 5: Two-dimensional fractionation of complete pancreatic ribonculease digest of 5′-terminal fragment IIa (109 residues long). (A) Autoradiograph. (B) Diagram identifying the oligonucleotide which corresponds to each spot (complete circles). In addition to these oligonucleotides, the corresponding digest of the larger 5′-terminal fragment Ia (117 residues long) contained A-Gp (p18), the position of which is indicated by a broken circle. P indicates the position of acid fuchsin on electrophoresis in the first dimension and B the position of xylene cyanol in the second dimension.

from the region between positions 65 and 117. The results of their analysis are given in Table IV. The overlaps between the sequences of these partial digestion products, which are indicated in Figure 6B, allowed us to deduce the complete sequence to position 117.

Discussion

The known sequence (Steitz, 1969) of the ribosomal binding site at the beginning of the A protein cistron starts with the sequence (Py)AUUCCUAG, which corresponds to the sequence of residues 109–117 in the largest fragment that we have studied. A given sequence of this length would be expected to occur at random only once in about 130,000 residues (2×4^8). From the gene order in R17 RNA (Jeppesen *et al.*, 1970b) and the size of the three R17 cistrons (see Cory *et al.*, 1970), we know that the beginning of the A protein cistron must be located within about 400 residues of the 5′ end. Therefore it seems extremely likely that the two sequences overlap. Strong additional evidence for the overlap is that U. Rensing and B. Barrell (personal communication) have isolated a 5′-terminal fragment from a partial T1 ribo-

TABLE IV:[a] Analysis of Partial Digestion Products from 5′-Terminal Fragments.[b]

Nucleotide[c]	Complete Digestion Products[d] Pancreatic RNase	T1 RNase	Sequence Deduced[e]
		Fragment 13a (74 Residues)	
P1	pppGGGU, GGGAC, C, *U*	pppG, (U_2,C_4,A), UG, G	pppGGGUGGGACCCCUU
P2	pppGGGU, GGGAC, C, U	pppG, (U_3,C_4,A), UG, G	pppGGGUGGGACCCCUUU
P3	pppGGGU, GGGAC, C, U	pppG, (U_3,C_5,A), UG, G	pppGGGUGGGACCCCUUUC
P4	GGGGU, GC, C, *U*	UCCUG, (C_2,U), CG, G	(py)CGGGGUCCUGCUC
P5	GGGGU, GC, AAC, C, *U*	UCCUG, (C_2,U,AAC), CG, G	(py)CGGGGUCCUGCUCAAC
P6	GGGGU, GC, AAC, C, U	UCCUG, (C_2,U_2,AAC), CG, G	(py)CGGGGUCCUGCUCAACU
P7	GGGGU, GU, AAC, GC, C, U	CUCAACUUCCUG, UCCUG, CG, G, U	(py)CGGGGUCCUGCUCAACUUCCUGU
P8	GAGC, GU, AAC, C, U	(A_2, C_3, U_3)G, UCG, AG, (C, U)	(py)AACUUCCUGUCGAGCU
P9	GAGC, GU, GC, AAC, AAU, U, C	(AAC,C_3,U_3)G, CUAAUG, UCG, AG, C	(py)AACUUCCUGUCGAGCUAAUGC
P10	GAGC, C, U	AG, CG, (C, U)	(py)CGAGCU
P11	GAGC, GU, *U*, *C*	(*U*,*C*)*G*, AG, (U,C)	(py)CUGUCGAGCU
P12	GGGGU, GAGC, GU, GC, AAC, C, U	CUCAACUUCCUG, UCCUG, UCG, AG, CG, (U,G), G	(py)CGGGGUCCUGCUCAACUUCCUGUCGAGCU
P13	GGGGU, GAGC, GU, *GC*, AAC, AAU, C, U	CUCAACUUCCUG, UCCUG, CUAAUG, UCG, AG, CG, G, CC	(py)CGGGGUCCUGCUCAACUUCCUGUCGAGCUAAUGCC
P14	AAU, GC	AAUG, C	(py)AAUGC
P15	AAU, GC, C	AAUG, CC	(py)AAUGCC
P16	GU, AAU	AAUG, U	(py)AAUGU
P17	GU, AAU, U, C	AAUG, (U_4, C)	(py)AAUGUCUUU
P18	GAGAC, AGC	*AG*, CG, AC	(py)AGCGAGAC
P19	GAGAC, AGC, G	ACG, *AG*, CG	(py)AGCGAGACG
P20	GAGAC, GU, AAU, AGC, G, C, U	UCUUUAG, AAUG, AG, CG, ACG	(py)AAUGUCUUUAGCGAGACG
T1	(G, AG), C	AG, CG	(G) CGAG
T2	GGGGU, pppGGGU, GGGAC, GC, AAC, G, U, C	CUCAACUUCCUG, UCCUG, G, ACCCCUUUCG, pppG	pppGGGUGGGACCCCUUUCGGGGUCCUGCUCAACUUCCUG
T3	*AAU*, AU, GC, G, *C*, U	CCAUUUUUAAUG, CUAAUG	(G) CUAAUGCCAUUUUUAAUG
T4	GU, AAU, AG, AU, C, U	CCAUUUUUAAUG, UCUUUAG	(G) CCAUUUUUAAUGUCUUUAG
T5	GU, *AAU*, AU, AG, GC, C, U	CCAUUUUUAAUG, UCUUUAG, CUAAUG	(G) CUAAUGCCAUUUUUAAUGUCUUUAG
		Fragment IIa (109 Residues)[f]	
P1	GAGAC, AGC, GC	ACG, *AG*, CG, C	(py)AGCGAGACGC
P2	GAGAC, AGC, GC, U	*AG*, CG, ACG, (C, U)	(py)AGCGAGACGCU
P3	GGC, AU, U	AUG, G, (C, U)	(py)AUGGCU
P4	GGC, *AU*, GC, U, C	CUAUCG, AUG, G, C	(py)AUGGCUAUCGC
P5	GGAAU, AGC, C, U	(AAU, U), AG, CCG, G	(py)AGCCGGAAUU
P6	GGAAU, AGGU, AGC, C, U	(AAU, U, C_2), UAG, AG, CCG, *G*	(py)AGGUAGCCGGAAUUCC

Fragment Ia (117 Residues)[f]

T1	AU, GC, *AC*, C, G, U	CUACCAUG, ACG	(G)ACGCUACCAUG
T2	GAGAC, AU, GC, AC, G, *C*, U	CUACCAUG, ACG, AG, CG	(G) CGAGACGCUACCAUG
T3	GG, AU, AC, *C*, U	CUACCAUG, G	(G) CUACCAUGG
T4	GGC, *AU*, GC, G, *C*, *U*, *AC*	CUAUCG, CUACCAUG, ACG, G	(G) ACGCUACCAUGGCUAUCG
T5	AU, GC, G, *C*, *U*	CUAUGC, CUG	(G) CUAUCGCUG
T6	GGC, *AU*, GC, AC, G, C, U	CUAUCG, CUACCAUG, G, CUG	(G) CUACCAUGGCUAUCGCUG
T7	GAAU, AG, AU, C, U	AAUUCCAUUCCUAG, G	(G) GAAUUCCAUUCCUAG
T8	GGAAU, AG, AU, C, U	AAUUCCAUUCCUAG, CCG, G	(G) CCGGAAUUCCAUUCCUAG
T9	GU, AG	UAG, G	(G) GUAG
T10	GGAAU, GU, AGC, AG, AU, C, U	AAUUCCAUUCCUAG, UAG, CCG, *G*	(G) GUAGCCGGAAUUCCAUUCCUAG
T11	GGAAU, AGGU, AGC, AG, AU, AU, C, U	AAUUCCAUUCCUAG, *UAG*, CCG, *G*	(G) UAGGUAGCCGGAAUUCCAUUCCUAG
T12	GGAAU, AGGU, GGC, GU, AGC, AU, AG, GC, AC, C, U	AAUUCCAUUCCUAG, CUACCA-UG, CUAUCG, *UAG*, CUG, CCG, G	(G) CUACCAUGGCUAUCGCUGUAGGUAGCCGGAAUUCCAUUCCUAG

[a] To save space in this table the hyphens and 3′-terminal phosphates have been omitted from the sequences of oligonucleotides. [b] The partial enzymic digestion conditions and the methods used to fractionate and analyze the products are described in Materials and Methods section. [c] Products P1, P2, etc., are from partial pancreatic ribonuclease digests and products T1, T2, etc., from partial T1 ribonuclease digests. [d] The relative yields of the complete digestion products were estimated by visual inspection of the autoradiograph. If product is in italic type, it was present in two molar yield; if in small capital letters, in 3 or more molar yield. [e] The sequence of each oligonucleotide shown in this column was deduced using the information provided by the products of its complete digestion with T1 ribonuclease and pancreatic ribonuclease and the sequence of related but smaller partial products. [f] Only the partial digestion products derived from the 3′-terminal half of the molecule have been included in the table.

TABLE V: Analysis of Pancreatic Ribonuclease Digestion Products.

Nucleotide	T1 RNase Digestion Products[a]	Products of Partial Digestion with Spleen Phosphodiesterase[b]	Sequence Deduced
p1	1.0 pppGp, 2.1 Gp, 1.0 Up (average of 5 expt)		pppG-G-G-Up
p2	4.0 Gp, 1.0 Up (average of 4 expt)		G-G-G-G-Up
p3	0.9 A-A-Up, 2.3 Gp		G-G-A-A-Up
p4	A-Gp, Gp, Up		A-G-G-Up[c]
p5	A-Cp, G_p		G-G-G-A-Cp[d]
p6	A-Gp, A-Cp, Gp	G-A-Cp, A-G-A-Cp	G-A-G-A-Cp
p7	A-Gp, Gp, Cp	G-Cp, A-G-Cp	G-A-G-Cp
p8	*Gp*, Cp		G-G-Cp
p11	A-Gp, Cp		A-G-Cp

[a] The conditions for digestion with T1 ribonuclease are described in Materials and Methods. The yields of the digestion products were either determined by measuring the radioactivity of the paper containing the nucleotide using a Nuclear-Chicago gas-flow counter or simply estimated by visual inspection of the autoradiograph. In the latter case, the product is in italic type if it was present in two molar yield, and in small capital letters if it was in 3 or more. [b] The conditions for partial digestion with spleen phosphodiesterase were the following: the oligonucleotide was incubated in a capillary for 5 min at room temperature and then for 90 min at 37° in 15 μl of a solution containing 0.07 M ammonium acetate (pH 6.2), 0.002 M EDTA, and 0.2 unit/ml of spleen phosphodiesterase (unit activity as defined by Bernardi and Bernardi, 1968). Samples were taken at 0, 5, 15, and 90 min and loaded immediately on to DEAE paper prior to fractionation by electrophoresis in 7% formic acid or formic acid–acetic acid buffer (pH 1.9). The products were analysed by alkaline hydrolysis or by pancreatic ribonuclease digestion. [c] A-Gp must be 5′-terminal in p4 to overlap the T1 ribonuclease product t10 (U-A-Gp). [d] The position of p5 on the fingerprint left no doubt that it contained three Gp residues.

nuclease digest of R17 RNA which contains the oligonucleotides found in our fragment and also oligonucleotides from further into the ribosomal binding site for the A protein cistron. Therefore the sequence can now be written for the first 145 residues in R17 RNA and this is shown in Figure 7.

Apparently the first 129 residues at the 5′ end of R17 RNA and of MS2 RNA (De Wachter *et al.*, 1971) are not translated even though four potential initiation codons (three AUGs and a GUG) occur within this region before the AUG at the beginning of the A protein cistron (Figure 7). In Qβ RNA, on the other hand, the first AUG from the 5′ end, which begins at position 62 (Billeter *et al.*, 1969), is used for initiation (Staples *et al.*, 1971). It is known that ribosomes

TABLE VI: Pancreatic Ribonuclease Digestion Products from 5′-Terminal Fragments.[a]

Nucleotide	Sequence	Molar Yield in Fragment 13a (74 Residues)		Molar Yield in Fragment IVa (79 Residues)		Molar Yield in Fragment IIIa (92 Residues)		Molar Yield in Fragment IIa (109 Residues)		Molar Yield in Fragment Ia (117 Residues)	
		Experimentally Determined (3 expt)	Expected from Final Structure	Experimentally Determined (3 expt)	Expected from Final Structure	Experimentally Determined (2 expt)	Expected from Final Structure	Experimentally Determined (1 expt)	Expected from Final Structure	Experimentally Determined (1 expt)	Expected from Final Structure
p1	pppG-G-G-Up	0.7	1	0.7	1	0.9	1	Not measured	1	0.5	1
p2	G-G-G-G-Up	0.7	1	0.7	1	0.6	1	0.6	1	0.7	1
p3	G-G-A-A-Up	0	0	0	0	0	0	1.0	1	1.0	1
p4	A-G-G-Up	0	0	0	0	0	0	1.0	1	0.9	1
p5	G-G-G-A-Cp	0.8	1	0.9	1	1.0	1	0.9	1	0.8	1
p6	G-A-G-A-Cp	0.9	1	1.0	1	1.0	1	1.0	1	0.9	1
p7	G-A-G-Cp	0.9	1	1.1	1	1.0	1	1.1	1	1.0	1
p8	G-G-Cp	0	0	0.4	0	1.1	1	1.5	1	1.2	1
p9	G-Up	2.9[b]	2	3.0[b]	2	2.5	2	3.3	3	3.5	3
p10	A-A-Up	2.3	2	2.2	2	2.2	2	2.0	2	2.3	2
p11	A-G-Cp	1.0	1	1.5[b]	1	1.0	1	1.9	2	2.0	2
p12	A-Up	1.4	1	1.2	1	2.8[b]	2	2.7	3	4.2	4
p13	G-Cp	1.8	2	3.5	3	3.8	4	3.9	4	4.5	4
p14	A-A-Cp	1.1	1	1.0	1	0.8	1	0.6	1	1.0	1
p15	A-Cp	0	0	1.1	1	1.1	1	1.1	1	1.1	1
p16	Cp	Not measured	12	10.0	13	Not measured	15	Not measured	17	Not measured	19
p17	Up	Not measured	16	16.0	17	17.0	19	21.7	20	Not measured	22
p18	Gp	0.6	1	0	0	0.5	1	0	0	0	0
p19	A-Gp	0	0	0	0	0	0	0	0	Not measured	1

[a] Yields were determined by measuring the radioactivity of the paper containing the nucleotide using a Nuclear-Chicago low-background gas-flow counter. [b] The yields of these commonly occurring nucleotides are presumably high because of some contamination by RNA molecules other than the 5′-terminal fragment. The possibility that an extra mole of these nucleotides occurs in the 5′-terminal fragments is excluded by the analysis of partial digestion products given in Table IV.

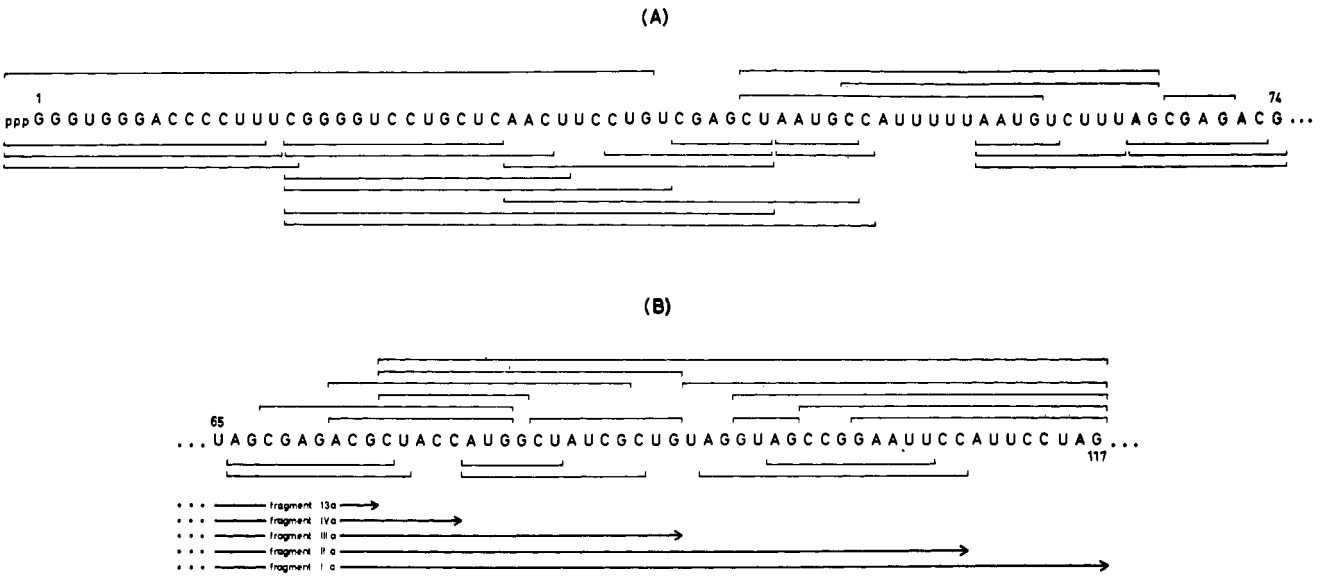
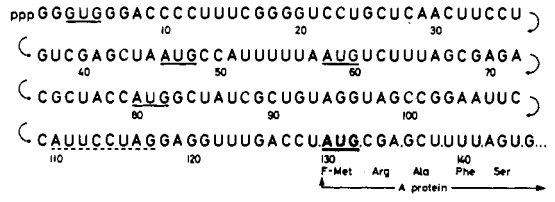


FIGURE 6: Overlaps between the sequences of partial digestion products. (A) Partial products obtained from the first 74 residues. (B) Partial products obtained from the region between residues 65 and 117. A line above either sequence indicates a partial T1 ribonuclease digestion product and a line below, a partial pancreatic ribonculease digestion product.

can attach to these phage RNAs specifically at the beginning of cistrons (Steitz, 1969; Hindley and Staples, 1969). Therefore it appears that ribosomes (Lodish, 1970a) and/or their associated initiation factors (Steitz *et al.*, 1970) recognize some special signal (s) in addition to the AUG at the start of each cistron. What these signals are, and to what extent the secondary and tertiary structure of the RNA affects their availability (Lodish, 1970b, 1971), are not well understood.

The nucleotide sequence of the first 74 residues in R17 RNA can be drawn in the form of two hydrogen-bonded loops, the first of which appears to be very stable (Adams and Cory, 1970). A hydrogen-bonded loop having several bulges can also be constructed for residues 82–138 and this is shown in Figure 8. According to the rules for estimating the stability of secondary structures recently formulated by Tinoco *et al.* (1971), this structure would be very stable (the free energy of formation from the single-stranded state is estimated to be -16.8 kcal mole^{-1}). We cannot say how likely it is that this loop, in which the AUG at the start of the A protein cistron is buried, occurs in the intact RNA molecule; hydrogen bonding or other types of interaction with some more distal region might prevent its formation. If the loop does occur in R17 RNA, its presence might account in part for the relatively low frequency with which *E. coli* ribosomes initiate translation of the A protein cistron (see Stavis and August, 1970). W. Fiers and collaborators (personal communication) have proposed an alternative structure for the 5′-terminal region of MS2 RNA, which is somewhat like the cloverleaf model for tRNA. Unfortunately at present it is difficult to predict the secondary structure for sequences as long as those now known at the 5′ end of R17 and MS2 RNAs with any confidence because of the large number of possibilities for base pairing. The difficulty is illustrated by the case of the *E. coli* 5S RNA (120 nucleotides long), for which at least five different secondary structures have been proposed (see Tinoco *et al.*, 1971). An added difficulty with the known phage RNA sequences is that they form part of much longer sequences and we do not know to what extent the tertiary structure of the whole phage RNA molecule will influence local secondary structure.

The RNA phages R17, f2, MS2, and Qβ were independently isolated in widely separated geographical areas (Zinder, 1965). Therefore it is noteworthy that the untranslated regions at the 5′ ends of R17 and f2 RNAs are identical for at least the first 74 residues (Ling, 1971) and that MS2 and R17 RNAs are identical for at least the first 125 residues (W. Fiers, personal communication). The nucleotide sequence in the 5′-terminal region of these phage RNA molecules may have been even more closely conserved than that in other parts of the molecules. Thus the known nucleotide sequences from the R17 and MS2 coat protein cistrons differ at nine positions out of 224 residues (Min Jou *et al.*, 1971) and those in the region of the coat protein cistron of f2 and R17 at 6 out of 162 resi-



FIGHRE 7: Nucleotide sequence of 145 residues at the 5′ end of R17 RNA. The overlap between the sequence of 117 residues reported in this paper and the sequence which Steitz (1969) established for the ribosomal binding site at the beginning of the A protein cistron is indicated by a broken line. The initiation codon for the A protein cistron (residues 130–132) is indicated by bold lettering and the GUG and AUG sequences in the preceding untranslated region are underlined.

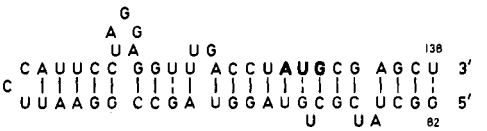


FIGURE 8: Possible secondary structure for residues 82 to 138 from the 5′-end of R17 RNA. This structure was constructed using the stability rules and the simple matrix method described by Tinoco *et al.* (1971). The initiation codon for the A protein cistron is indicated by bold lettering.

dues (Gupta *et al.*, 1970; Nichols and Robertson, 1971). Even the RNA from phage Qβ, which probably is only distantly related to the other phages, has two stretches of about 20 nucleotides within its 5′-terminal region which are very similar to the corresponding sequences in R17 RNA (Adams and Cory, 1970) though the 5′-terminal sequences of these two RNA molecules appear to be very different after these homologous regions. These findings suggest that there is selective pressure for maintaining *particular* nucleotide sequences within the untranslated region near the 5′ end of these molecules. The functions of these sequences are not yet known but we have suggested elsewhere that part of the sequence at the 5′ end of R17 RNA may function indirectly in replication since its complementary sequence at the 3′ end of the minus strand may form part of the recognition site for the R17 RNA synthetase (Adams and Cory, 1970; Cory *et al.*, 1970).

Acknowledgment

We thank Miss F. Ruinard for expert technical assistance, Dr. Ann Burgess for comments on the manuscript, and Dr. W. Fiers for communicating his results prior to publication.

Added in Proof

The complete nucleotide sequence to position 125 in MS2 RNA has now been reported by De Wachter *et al.* (1971).

References

Adams, J. M., and Cory, S. (1970), *Nature (London) 227*, 570.
Adams, J. M., Jeppesen, P. G. N., Sanger, F., and Barrell, B. G. (1969), *Nature (London) 223*, 1009.
Arima, T., Uchida, T., and Egami, F. (1968), *Biochem. J. 106*, 609.
Billeter, M. A., Dahlberg, J. E., Goodman, H. M., Hindley, J., and Weissmann, C. (1969), *Nature (London) 224*, 1083.
Brownlee, G. G., and Sanger, F. (1967), *J. Mol. Biol. 23*, 337.
Brownlee, G. G., and Sanger, F. (1969), *Eur. J. Biochem. 11*, 395.
Brownlee, G. G., Sanger, F., and Barrell, B. G. (1968), *J. Mol. Biol. 34*, 379.
Cory, S., Adams, J. M., Spahr, P. F., and Rensing, U. (1972), *J. Mol. Biol.* (in press).
Cory, S., and Marcker, K. A. (1970), *Eur. J. Biochem. 12*, 177.
Cory, S., Spahr, P. F., and Adams, J. M. (1970), *Cold Spring Harbor Symp. 35*, 1.
Dahlberg, J. E. (1968), *Nature (London) 220*, 548.
De Wachter, R., Merregaert, J., Vandenberghe, A., Contreras, R., and Fiers, W. (1941), *Eur. J. Biochem. 22*, 400.
De Wachter, R., Merregaert, J., Vandenberghe, A., Contreras, R., and Fiers, W. (1971), *Proc. Nat. Acad. Sci. U. S. 68*, 585.
Gesteland, R., and Spahr, P. F. (1969), *Cold Spring Harbor Symp. 34*, 707.
Gesteland, R., and Spahr, P. F. (1970), *Biochem. Biophys. Res. Commun. 41*, 1267.
Gilham, P. T. (1962), *J. Amer. Chem. Soc. 84*, 687.
Goodman, H. M., Billeter, M. A., Hindley, J., and Wiessmann, C. (1970), *Proc. Nat. Acad. Sci. U. S. 67*, 921.
Gupta, S. L., Chen, J., Schaefer, L., Lengyel, P., and Wiessman, S. M. (1970), *Biochem. Biophys. Res. Commun. 39*, 883.
Hindley, J., and Staples, D. (1969), *Nature (London) 224*, 964.
Ho, N. W. Y., and Gilham, P. T. (1967), *Biochemistry 6*, 3632.
Jeppesen, P. G. N. (1971), *Biochem. J. 124*, 357.
Jeppesen, P. G. N., Nichols, J., Sanger, F., and Barrell, B. (1970a), *Cold Spring Harbor Symp. 35*, 13.
Jeppesen, P. G. N., Steitz, J. A., Gesteland, R. F., and Spahr, P. F. (1970b), *Nature (London) 226*, 230.
Kamen, R. (1970), *Nature (London) 228*, 527.
Keller, E. B. (1964), *Biochem. Biophys. Res. Commun. 17*, 412.
Kondo, M., Gallerani, R., and Wiessmann, C. (1970), *Nature (London) 228*, 525.
Ling, V. (1971), *Biochem. Biophys. Res. Commun. 42*, 82.
Lodish, H. F. (1970a), *Nature (London) 226*, 705.
Lodish, H. F. (1970b), *J. Mol. Biol. 50*, 689.
Lodish, H. F. (1971), *J. Mol. Biol. 56*, 627.
Min Jou, W., Haegerman, G., and Fiers, W. (1971), *FEBS (Fed. Eur. Biochem. Soc.) Lett. 13*, 105.
Nichols, J. L. (1970), *Nature (London) 225*, 147.
Nichols, J. L., and Robertson, H. D. (1971), *Biochim. Biophys. Acta 228*, 676.
Roblin, R. (1968a), *J. Mol. Biol. 31*, 51.
Roblin, R. (1968b), *J. Mol. Biol. 36*, 125.
Sanger, F., Adams, J. M., Jeppesen, P. G. N., and Barrell, B. G. (1970), *in* Macromolecules, Biosynthesis and Function, Vol. 21, Ochoa, S., Asensio, C., Heredia, C. F., and Nachsmansohn, D., Ed., New York, N. Y., Academic Press, p 213.
Sanger, F., Brownlee, G. G., and Barrell, B. G. (1965), *J. Mol. Biol. 13*, 373.
Staples, D. H., Hindley, J., Billeter, M. A., and Weissmann, C. (1971), *Nature (London) 234*, 202.
Stavis, R. L., and August, J. T. (1970), *Annu. Rev. Biochem. 39*, 527.
Steitz, J. A. (1969), *Nature (London) 224*, 957.
Steitz, J. A., Dube, S. K., and Rudland, P. S. (1970), *Nature (London) 226*, 824.
Tinoco, I., Uhlenbeck, O. C., and Levine, M. D. (1971), *Nature (London) 230*, 362.
Zinder, N. D. (1965), *Annu. Rev. Microbiol. 19*, 455.