Nucleotide Sequences of Similar Size from the Coliphage R17 Genome

By ULRICH F. E. RENSING and ALAN COULSON

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

and E. O. P. THOMPSON

School of Biochemistry, University of New South Wales, P.O. Box 1, Kensington, N.S.W. 2033, Australia

(Received 14 September 1972)

A sequence of 33 nucleotides from the coliphage R17 RNA genome was determined. It constitutes the main component of a mixture of fragments that migrate together on electrophoresis in a separation according to molecular weight. Fragments of comparable chain length from the 3' end of RNA from coliphage R17, from a region preceding and overlapping the coat-protein cistron ribosome binding site and from the beginning of the A-protein cistron, were also found and characterized. 'Hairpin'-like secondary structures are proposed for the longer fragments, one of which appears to have a tetranucleotide excised in the loop region.

Sequence analysis of radioactive nucleic acids depends largely on the availability of fast fractionation techniques for separating the complex mixtures of oligonucleotides obtained by partial digestion of polynucleotides. The two-dimensional-fractionation system on thin-layer plates described by Brownlee & Sanger (1969) is suitable for separating fragments of up to 40 nucleotides long. It can therefore be applied in place of the urea-containing 8% (w/v) polyacrylamide gel at pH 3.5 (Rensing, 1973; de Wachter *et al.*, 1971) for re-fractionation of the partial ribonuclease T₁ digestion products (bands '20–30'), isolated by electrophoresis at pH 8.4 in a 12.5% polyacrylamide gel as described in detail before (Rensing, 1973). We report here the results obtained with one particular gel band, '28', analysed in several different preparations of RNA from coliphage R17. The complete sequences of the five main components most consistently found in this band could be established and related to each other. Two other fragments present in lower yields were found to come from the beginning of the A-protein cistron and from the region preceding the coat-protein cistron ribosome binding site.

Experimental

All the methods used in this investigation have been described in detail before [for review of most of them see Barrell (1971) and Brownlee (1971)]. The bacteriophage was grown and its ³²P-labelled RNA prepared, partially digested, fractionated and extracted as detailed in the preceding paper (Rensing, 1973) or according to Steitz (1969), Anderson *et al.* (1967) and Adams *et al.* (1969).

Results and Discussion

Isolation and re-fractionation of gel band '28'

In the preceding paper (Rensing, 1973) the partial digestion with ribonuclease T₁ of ³²P-labelled RNA from coliphage R17 was described. It yields fragments ranging from about 250 to 25 nucleotides long, which can be isolated by electrophoresis in a 12.5% (w/v) polyacrylamide gel at pH 8.4 as previously reported by Adams *et al.* (1969). The fragment numbered '28', which migrates on electrophoresis with the Bromphenol Blue marker, was extracted and re-fractionated in the two-dimensional system shown in Plate 1. By this method more than ten different spots of varying intensity were separated. Since the second dimension (homochromatography) fractionates mainly according to size and also according to purine content (Jeppesen, 1971) it seemed possible that they had been generated by breakdown of the major component at the origin, numbered 33. However, their analysis disclosed that only two of them, components 15 and 18 (but none of the other five numbered components), were related to component 33. This is summarized in Tables 1 and 2. For the sake of simplicity we number these fragments according to chain length, and to avoid confusion with a particular mixture of fragments of similar chain length obtained by electrophoresis at pH 8.4 in the 12.5% polyacrylamide gel we will use inverted commas for these. Hence, '28' denotes a gel band in about the position 28 from the origin of electrophoresis. We thus follow the numbering system of the first gel pictures published (Adams *et al.*, 1969; Nichols, 1970).

Table 1. *Analysis of the isolated fragments from gel band '28'*

The oligonucleotides shown in Plate 1 were eluted and subjected to enzymic digestions. One part was completely digested with ribonuclease T₁; the products (Arabic numerals) were separated in one dimension on DEAE-paper in 7% formic acid and in two dimensions on DEAE-cellulose thin-layer plates, and then further analysed by digestion with ribonuclease A (first column). Some of these ribonuclease T₁ products were also analysed by blocking with carbodi-imide and then digesting with ribonuclease A (CD in second column.) Dotted symbols ($\dot{U}$ etc.) denote modified nucleotides. Some were characterized by digestion with ribonuclease U₂ (U₂ in the second column). Another part of the original material was completely digested with ribonuclease A (third column) and its products (Roman numerals) further analysed by digestion with ribonuclease T₁. This was sufficient in most cases to establish their sequences. The values of the subscripts indicate molar yields to the nearest whole number, where these are not unity (determined by liquid-scintillation counting). When the molar yield was estimated by visual inspection of the radioautograph the numbers of nucleotides in excess of 1 were underlined a corresponding number of times. If such an estimate seemed impossible, the prefixes *n* and *m* were used; *n* and *m* denote any number bigger than 3 or 4 with *n* and *m* being unequal.

Nucleotide	Complete ribonuclease T ₁ digestion products	Products of additional analyses	Complete ribonuclease A digestion products
30	(1) (A-A-A-U, $\dot{U}$,C)A-Gp (2) (A-A-C,C ₄₋₅ , $\dot{U}$)Gp (3) (A-A-C,C)Gp (4) A-Gp (5) Gp	(1) CD: Cp, $\dot{U}$ -Cp, ($\dot{U}$,A-A-A- $\dot{U}$)A-Gp (2) CD: Cp, $\dot{U}$ - $\dot{U}$ -Gp (3) CD: $\dot{A}$ - $\dot{U}$ -Cp, A-A-Gp	(I) A-G-A-G-Cp (II) G-G-Cp (III) A-A-A-Up (IV) A-A-Cp (V) <i>n</i> Cp (VI) $\dot{U}$ p
23	(1) ($\dot{U}$,C)Gp (2) ($\dot{A}$ -C,C, $\dot{U}$)Gp (3) (A-C,A- $\dot{U}$,C)Gp (4) C-Gp (5) Gp	(1) CD: Cp, $\dot{U}$ - $\dot{U}$ - $\dot{U}$ -A-Gp (2) CD: Cp, $\dot{U}$ - $\dot{U}$ -Gp (3) CD: $\dot{A}$ - $\dot{U}$ -Cp, A-A-Gp	(I) G-Up (II) A-Up (III) G-Gp (IV) A-Cp (V) <i>n</i> Cp (VI) <i>m</i> Up
21	(1) (C,U ₄)A-Gp (2) (C ₃ ,U ₂)Gp (3) (C,A-U)A-A-Gp (4) U-Gp	(1) CD: Cp, $\dot{U}$ - $\dot{U}$ - $\dot{U}$ -A-Gp (2) CD: Cp, $\dot{U}$ - $\dot{U}$ -Gp (3) CD: $\dot{A}$ - $\dot{U}$ -Cp, A-A-Gp	Not analysed
18	(1) (C)A-A-Gp (2) ($\dot{U}$,C)A-Gp (3) A-A-A-Gp (4) C-Gp	See Table 2 See Table 2	(I) Gp, A-A-A-Gp (II) A-A-G-Up (III) A-G-Cp (IV) Cp (V) $\dot{U}$ p
15	(1) (U,A-U,C)Gp (2) (C,U)Gp (3) ($\dot{A}$ -A-C)Gp (4) A-Gp	See Table 2 See Table 2	(I) G-A-G-A-A-Cp (II) A-Up (III) A-A-Cp (IV) Cp (V) $\dot{U}$ p
16	(1) (A-U,U ₂)Gp (2) (A-A-C,U ₂)A-Gp (3) (C,U)Gp	(1) U ₂ : Ap, U-U-U-Gp (2) CD: $\dot{U}$ -A-A-Cp, $\dot{U}$ -A-Gp (3) CD: Cp, $\dot{U}$ -Gp	(I) G-G-Up (II) A-G-Cp (III) A-A-Cp (IV) A-Up (V) Gp (VI) Cp (VII) <i>n</i> Up
14	(1) (C,U)A-Gp (2) ($\dot{A}$ -Cp,C, $\dot{U}$) no Gp (3) Gp	(1) CD: Cp, $\dot{U}$ -A-Gp (2) U ₂ : U-U-Ap, C-C-Ap, C-C-Cp(A _{OH})*	(I) A-G-Up (II) A-Cp (III) G-Cp (IV) Cp (V) $\dot{U}$ p

* The presence of adenosine was suggested by the absence of guanylate in this sequence, which is identical with the one described by Dahlberg (1968).

EXPLANATION OF PLATE I

Re-fractionation of band '28'

Band '28' from the polyacrylamide-gel electrophoresis was extracted and subjected to electrophoresis on cellulose acetate at pH 3.5 as the first dimension (1) and to homochromatography on DEAE-cellulose thin-layer plates with a 5% dialysed non-hydrolysed RNA solution as the second dimension (2), as described by Brownlee & Sanger (1969). The radioautograph shows several spots, which are labelled according to chain length. They were eluted and further analysed as detailed in the text. The dotted circle (21) indicates the position of an oligonucleotide that was found in only two preparations, not in the particular one shown here and not together with nucleotide 30. The unlabelled minor spots were also analysed. They could not be related by composition, however, to any of the others or to known sequences of RNA from coliphage R17.

EXPLANATION OF PLATE 2

Radioautograph of the two-dimensional fractionations of complete ribonuclease digests from fragment 33 (Plate 1)

(a) Ribonuclease T_1 digestion products. (b) Ribonuclease A digestion products. The fractionation was by electrophoresis at pH 3.5 on cellulose acetate (1) and at pH 1.9 on DEAE-paper (2). B denotes the position of the blue dye marker (Xylene Cyanol FF). The deduction of the sequences is described in Table 2 and in the text. (The sequences are written in a simplified version to save space. All of them carry a 5'-hydroxyl group and a 3'-phosphate; e.g., AAAG indicates A-A-A-Gp.) The 2':3'-cyclic phosphate derivative of a particular residue resulting from incomplete hydrolysis is denoted by !.

EXPLANATION OF PLATE 3

Products obtained after partial digestion with ribonuclease T_1

Fragment 33, purified as described in Plate 1, was divided into three parts, and one part was digested with 0.2 unit of ribonuclease T_1 in $7\mu\text{l}$ of 0.015M-Tris-HCl (pH 7.6)-0.007M-EDTA for 30min at 0°C and then spotted directly on the wet cellulose acetate strip. The two-dimensional fractionation (arrows 1 and 2) described for Plate 1 was used. The numbers indicate the chain length and refer to the sequences outlined underneath. Fragments 15 and 18 are those shown on Plate 1 and Table 1. In addition to the usual analyses of digesting with ribonucleases T_1 and A and subsequently with ribonucleases A and T_1 respectively (see Table 1), fragment 15 was also dephosphorylated with bacterial alkaline phosphatase, re-isolated and its 3' terminal ribonuclease T_1 product identified as A-A-C-G_{OH}. 33 is the undigested whole fragment. Fragment 18 was found in a pure form only on re-running band '28' (Plate 1). B denotes the position of the blue dye marker. The broken line indicates the presence in lower yields of A-Gp at the 3' end of product 9. (The sequence is written in the simplified version explained in Plate 2.) Products 9 and 12 are not separated in this system; the analysis of the other partial-digestion products, however, made it obvious that they could not be joined covalently.

Table 2. Analysis of the ribonuclease T_1 digestion products from fragment 33

The products obtained as described in Plate 2 were analysed by the operations indicated (CD, blocking with carbodi-imide and subsequent ribonuclease A digestion; SV, complete digestion with snake-venom phosphodiesterase). Product (3) was found only occasionally (see the text). Dotted symbols ($\dot{U}$ etc.) denote modified nucleotides.

Complete ribonuclease A digestion products of ribonuclease T_1 products (1)–(9)	Additional analysis	Molar yield of deduced structure	Deduced structure
(1) (C ₂ ,U ₄)A-Gp	(1) CD: $\dot{U}$ - $\dot{U}$ -Cp, A- $\dot{G}$ p	(1) 1.06	U-U-C-U-U-C-A-Gp
(2) (C,U,A-U)Gp	(2) CD: ($\dot{U}_2$,A) $\dot{C}$ p, $\dot{G}$ p	(2) 0.91	} U-A-U-C-Gp
(3) (C,A-U)Gp	(3) CD: (A, $\dot{U}$)Cp, $\dot{G}$ p (3) SV: pGp, pC, pU		
(4) (C,U)Gp	(4) CD: $\dot{U}$ -Cp, $\dot{G}$ p, Cp	(4) 1.05	C-U-C-Gp
(5) A-A-A-Gp	(4) SV: pGp, pC, pU	(5) 1.23	A-A-A-Gp
(6) (A-A-C)Gp		(6) 0.88	A-A-C-Gp
(7) (C)A-A-Gp		(7) 1.01	C-A-A-Gp
(8) A-Gp		(8) 1.05	A-Gp
(9) (C)Gp		(9) 1.02	C-Gp

5' -U-A-U-C-G-C-U-C-G-A-G-A-A-C-G-C-A-A-G-U-U-C-U-U-C-A-G-C-G-A-A-A-Gp-3'

. Tyr . Arg . Ser . Arg . Thr . Gln . Val . Leu . Gln . Arg . Lys .

. Ile . Ala . Arg . Glu . Arg . Lys . Phe . Phe . Ser . Glu . Ser . Arg .

. Ser . Leu . Glu . Asn . Ala . Ser . Ser . Ser . Ala . Lys .

Fig. 1. Nucleic acid sequence of spot 33 (Plate 1) and three amino acid sequences that can be derived from it by using the genetic code

None of the sequences comes from the coat protein (Weber, 1967).

Sequence analysis of a fragment 33 nucleotides long

This fragment was analysed by fractionating the ribonuclease T_1 and ribonuclease A digestion products in the two-dimensional system (Plate 2); the products appear to be very pure. The individual products were eluted and further analysed as detailed in Table 2. This information is sufficient to deduce the complete sequences of the ribonuclease T_1 digestion products. [One of these ribonuclease T_1 products, (3), was not found in the particular 'fingerprint' shown in Plate 2, but it occurred sometimes in yields concomitant with a decrease in yield of product (2). These two differ by only one uridylyte residue. It is therefore concluded that product (3) was generated from product (2) by an unspecific cleavage, possibly owing to over-digestion with ribonuclease T_1 .]

The sequences of the ribonuclease A products could be established from the results of partial digestions with spleen phosphodiesterase. Establishing the order of the complete digestion products was facilitated by the presence in the re-run of the two half-molecules 15 and 18 (compare Tables 1 and 2).

The ambiguities still remaining could be removed by performing a partial digestion with ribonuclease T_1 and analysing the products obtained as described in Plate 3. The accumulated evidence allowed us to establish the complete sequence of fragment 33.

By using the genetic code a set of three amino acid sequences from a known nucleic acid sequence can be deduced, as shown in Fig. 1. None of these could come from the coat protein of coliphage R17, whose amino acid sequence is known (Weber, 1967). The amino acid sequences of the other two proteins for which RNA from coliphage R17 has been found to code (for review see Stavits & August, 1970) are not yet known. At the moment it is impossible therefore to decide if the sequence described here comes from one of the cistrons or the probably untranslated termination regions (Cory *et al.*, 1970, 1972).

Identification and preliminary analysis of other fragments

Fragments 14 and 16: the 3' end of RNA from coliphage R17. We have described previously a

Fig. 2. Location of fragment 30

Nucleic acid sequence around the coat-protein ribosome binding site (Steitz, 1969; between the arrows) showing the location of fragment 30 from Plate 1. The numbers refer to the oligonucleotides listed in Table 1.

sequence of 51 nucleotides at the 3' end of RNA from coliphage R17 (Cory *et al.*, 1972). The analysis of fragments 14 and 16 from the re-run as listed in Table 1 was sufficient to establish their sequences and to relate them to the 3' end. Fragment 14 contains the ribonuclease T_1 product U-U-A-C-C-A-C-C-C-(A_{OH}), which is identical with the 3'-terminal nonamer described by Dahlberg (1968). (The presence of adenosine is suggested by the absence of Gp in this ribonuclease T_1 product.) Comparison of these fragments with the published (Cory *et al.*, 1972) sequence shows that a C-U-U-Gp sequence is missing between them. This C-U-U-Gp sequence helps to form a loop in a possible secondary structure (see below, under 'Possible secondary structures').

Fragment 30. The oligonucleotide of chain length 30 was not found in every gel band '28' analysed. It contains a characteristic ribonuclease T_1 digestion product whose tentative sequence (C,U-C,U-C)-(A-A-A-U,U)A-Gp can be derived from the results in Table 1. This is consistent with a tentative sequence proposed by Cory *et al.* (1970), according to Gupta *et al.* (1970), which overlaps the coat-protein cistron ribosome binding site (Steitz, 1969) and extends it towards the 5' end. As shown in Fig. 2, the other results from Table 1 are also consistent with the proposed sequence without extending it. The analysis of fragment 30, which is present in low molar yield, was therefore not carried further. It is noteworthy that a fragment of identical composition containing about four additional nucleotides was also found occasionally in the next bigger gel band, numbered '27', which migrates just behind band '28' (the one described here), on electrophoresis in a 12.5% polyacrylamide gel.

A sequence of 21-nucleotides at the 5' end of the A-protein cistron. As mentioned above, fragment 21 was present in low molar yields in only two out of about eight coliphage R17 preparations analysed for

Fig. 3. N-Terminus of the A-protein and fragment 21

The probable position of fragment 21 at the A-protein cistron ribosome binding site (Steitz, 1969; between the continuous arrows) is shown. The N-terminal sequence of the maturation protein has been used to determine the order of the ribonuclease T_1 products (T_1 - T_4) from fragment 21. The numbers 100, 110 and 157 refer to the position with respect to the 5' end of RNA from coliphage R17. The broken arrow indicates the cleavage by ribonuclease T_1 in a partial digest to give gel band '8', which contains the first 125 nucleotides of RNA from coliphage R17 (U. F. E. Rensing & B. G. Barrell, unpublished work). They are identical with those of RNA from coliphage MS2 (de Wachter *et al.*, 1971; Adams & Cory, 1970).

gel bands '27', '28' and '29'. It was therefore not possible yet to establish the complete sequence. The results in Table 1 are just sufficient to deduce the sequences of the ribonuclease T_1 products. Their relative order can be proposed, however, as follows: the ribonuclease T_1 products 1 and 4, C-U-U-U-U-A-Gp

and U-Gp, are also found in the A-protein cistron ribosome binding site described by Steitz (1969). These are not very likely to occur by chance twice and together in a molecule about 3500 nucleotides long. Moreover, the first nine *N*-terminal amino acids of the A-protein have been determined (Weiner *et al.*, 1972) and they can be related by the genetic code to all four ribonuclease T_1 products of fragment 21 as depicted in Fig. 3. It is thus possible to propose an extension of 12 nucleotides from the binding site into the cistron. On the basis of protein chemical information it was not clear whether the *C*-terminal amino acid was glutamine or glutamic acid. From the sequence of the oligonucleotide T_3 it becomes clear that it must be glutamine (Fig. 3). This is also an example of how protein and nucleic acid sequence determinations can complement each other in the analysis of primary structure.

Possible secondary structures

The main experimental evidence in favour of 'hairpin' loops proposed for fragments from RNA from coliphages R17 and MS2 is the frequency of cleavages in unpaired regions of the possible structures (Adams *et al.*, 1969; Nichols, 1970; Nichols & Robertson, 1971; Jeppesen *et al.*, 1972;

Cory *et al.*, 1972; de Wachter *et al.*, 1971; Min Jou *et al.*, 1972).

Fragment 33 adds another example: two, and only two, specific half-molecules (15 and 18) were found on screening all detectable minor components of gel band '28'. Fig. 4 shows that the cleavage occurred in the unpaired loop region of the proposed structure. Another cleavage right next to it was detected occasionally, owing probably to a contaminating enzymic activity.

A new finding is the observation that an excision of a ribonuclease T_1 product must have occurred in the loop region of the structure from the 3' end (fragment 31 in Fig. 4). We have previously reported a sequence of 35 nucleotides at the 3' end of RNA from coliphage R17 (Cory *et al.*, 1972), isolated from gel band '27', which migrates just behind band '28'. In band '27' two fragments were found repeatedly in good molar yields, which migrated on electrophoresis in the same regions as fragments 14 and 16. The one corresponding to fragment 14 contained the same end products of enzymic digestions and in addition the sequences C-U-U-Gp and G-G-Cp. Since neither in gel band '27' nor in band '28' could be found the complete structures of 35 or 31 nucleotides length respectively (nor any other related fragment), and because the two half-molecules migrate in the same

Fig. 4. Possible secondary structures of maximal estimated stability

'Hairpin' loops can be constructed for the longer fragments from gel band '28'. The tetramer C-U-U-Gp could not be found in band '28' (see the text for details). The continuous arrows indicate cleavage by ribonuclease T_1 ; the broken arrow indicates an unspecific cleavage observed in earlier experiments. The numbers refer to the chain length and to the particular fragment of Plate 1 and Tables 1 and 2. Fragment 30 containing a potential terminator codon in the loop region precedes the coat-protein gene (Steitz, 1969); fragment 31 is part of the 3' end of RNA from coliphage R17 (Cory *et al.*, 1972). The estimation of the stability of the structures is given in the text.

gel bands as the species 33 or 35 nucleotides long respectively in a system that is known to separate according to chain length, it is concluded that they have been held together by secondary forces during electrophoresis. Presumably the hydrogen bonds were broken in the urea-containing acidic buffer used for the subsequent fractionation and the fragments 14 and 16 were separated (Plate 1).

A 'hairpin'-like secondary structure can also be found for fragment 30 from the termination region of the A-protein cistron (Fig. 4). However, this is estimated to be less stable (-20100J/mol) at 25°C than fragment 33 (-35200J/mol) and the structure from the 3' end (-60350J/mol) by following the rules proposed by Tinoco *et al.* (1971). No such structure can be formed with fragment 21. However, since it was found only twice in low molar yields just like other minor components, whose sequences were not determined, it cannot be decided whether there was a complementary sequence present of about 12 nucleotides to make it the same molecular weight as fragment 33. Moreover, if it were proven to be single-stranded it might be expected to migrate more slowly, as indicated by findings in the preceding paper (Rensing 1971).

It is justified to question the existence of possible secondary structures which can be constructed and whose free energy can be estimated (Tinoco *et al.*, 1971) from a known nucleic acid sequence. Apart from the experimental observations mentioned above, indicating that they exist, it can be argued that it might not be expected to be possible to form as many 'hairpin' loops as have been found if the total RNA sequence of about 3500 nucleotides from coliphage R17 were completely random. Physical studies on fragments with known sequence as a function of temperature, concentration, ionic strength and components present may establish a differential melting curve according to different parts of a given structure under conditions near those *in vivo* and may thus resolve the question of secondary structure.

We thank Dr. F. Sanger, in whose laboratory this work was carried out, for his continued interest and support and for his help in the preparation of this manuscript, Dr. B. G. Barrell for discussion and for pointing out the possibility of inverting two ribonuclease T_1 products in an early tentative sequence to form the 'hairpin' loop, and also Dr. P. G. N. Jeppesen for discussion when we first looked for

the 3' end of RNA from coliphage R17. The ribonuclease U_2 used was a kind gift from Sankyo Ltd. (Tokyo, Japan). U. F. E. R. was supported by a long-term fellowship from the European Molecular Biology Organisation, and E. O. P. T. by a Royal Society and Nuffield Foundation Commonwealth Bursary.

References

- Adams, J. M. & Cory, S. (1970) *Nature (London)* **227**, 570–574
- Adams, J. M., Jeppesen, P. G. N., Sanger, F. & Barrell, B. G. (1969) *Nature (London)* **223**, 1009–1014
- Anderson, J. S., Dahlberg, J. E., Bretscher, M. S., Revel, M. & Clark, B. F. C. (1967) *Nature (London)* **216**, 1072–1076
- Barrell, B. G. (1971) *Procedures Nucleic Acid Res.* **2**, 751–779
- Brownlee, G. G. (1971) *Laboratory Techniques in Biochemistry and Molecular Biology* (Work, T. S. & Work, E., eds.), vol. 3, part 1, North-Holland Publishing Co., Amsterdam
- Brownlee, G. G. & Sanger, F. (1969) *Eur. J. Biochem.* **11**, 395–399
- Cory, S., Spahr, P. F. & Adams, J. M. (1970) *Cold Spring Harbor Symp. Quant. Biol.* **35**, 1–12
- Cory, S., Adams, J. M., Spahr, P. F. & Rensing, U. F. E. (1972) *J. Mol. Biol.* **63**, 41–56
- Dahlberg, J. E. (1968) *Nature (London)* **220**, 548–552
- de Wachter, R., Merregaert, J., Vandenberghe, A., Contreras, R. & Fiers, W. (1971) *Eur. J. Biochem.* **22**, 400–414
- Gupta, S. L., Chen, J., Schaefer, L., Lengyel, P. & Weissmann, S. M. (1970) *Biochem. Biophys. Res. Commun.* **39**, 883–888
- Jeppesen, P. G. N. (1971) *Biochem. J.* **124**, 357–366
- Jeppesen, P. G. N., Barrell, B. G., Sanger, F. & Coulson, A. (1972) *Biochem. J.* **128**, 993–1006
- Min Jou, W., Haegeman, G., Ysebaert, M. & Fiers, W. (1972) *Nature (London)* **237**, 82–88
- Nichols, J. D. (1970) *Nature (London)* **225**, 147–151
- Nichols, J. L. & Robertson, H. D. (1971) *Biochim. Biophys. Acta* **228**, 676–681
- Rensing, U. F. E. (1973) *Biochem. J.* **131**, 593–604
- Stavis, R. L. & August, J. T. (1970) *Annu. Rev. Biochem.* **39**, 527–560
- Steitz, J. A. (1969) *Nature (London)* **224**, 957–964
- Tinoco, I., Uhlenbeck, O. C. & Levine, M. D. (1971) *Nature (London)* **230**, 362–366
- Weber, K. (1967) *Biochemistry* **6**, 3144–3154
- Weiner, A. M., Platt, T. & Weber, K. (1972) *J. Biol. Chem.* **247**, 3242–3251