

Sequence of 1060 3'-terminal nucleotides of poliovirus RNA as determined by a modification of the dideoxynucleotide method*

(reverse transcription/RNase T1- and RNase A-resistant oligonucleotide-primed DNA synthesis/two-dimensional gel electrophoresis/sequencing gels)

NAOMI KITAMURA AND ECKARD WIMMER

Department of Microbiology, School of Basic Health Sciences, State University of New York at Stony Brook, Stony Brook, Long Island, New York 11794

Communicated by Igor Tamm, February 27, 1980

ABSTRACT The dideoxynucleotide method for sequencing DNA developed by Sanger *et al.* [Sanger, F., Nicklen, S. & Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467] was modified to allow sequence analysis of poliovirus RNA without recourse to cloning. Our method involves reverse transcription of poliovirus RNA followed by cDNA-dependent DNA synthesis in the presence of unlabeled dNTPs and 2',3'-dideoxynucleoside triphosphates, with *Escherichia coli* DNA polymerase I (Klenow) used to catalyze the reaction. DNA synthesis is primed by 5'-³²P-labeled RNase T1- or RNase A-resistant oligonucleotides generated from poliovirus RNA. The sequence of 1060 nucleotides preceding the 3'-terminal poly(A) is presented. Based on the position of termination codons we propose that viral translation terminates at nucleotide -562.

Poliovirus, a human enterovirus, is a small naked capsid virus belonging to the family Picornaviridae. This family of viruses is unique in genome structure and genome expression (1). Poliovirus contains a single-stranded RNA molecule that is covalently linked to a small protein [VPg (2-4)] at its 5' end (5) and polyadenylated at its 3' end (6). Knowledge of the total nucleotide sequence of poliovirus RNA, estimated to be 7800 nucleotides long (7), is valuable for several reasons. In combination with analyses of viral proteins it permits the determination of the initiation and termination sites of translation, the precise mapping of proteins, and the identification of cleavage sites in protein processing. Moreover, the untranslated sequences may reveal sites of functional significance such as the recognition site(s) for the RNA replicase or protein binding sites involved in maturation.

Rapid sequence determination techniques have been developed for analysis of the primary structure of both DNA and RNA (8-11). These techniques are suitable for segments up to 400 nucleotides long (12). Longer nucleic acid molecules must be cleaved into fragments, which, in the case of DNA, is facilitated by restriction enzymes (13). Because restriction enzymes for RNA are unknown, use of rapid techniques has been limited to relatively small RNAs or to 5'- or 3'-terminal regions of larger RNAs. Donis-Keller (14) recently developed a procedure that allows the site-specific fragmentation of large RNA. The method of choice for determining the primary structure of high molecular weight RNA molecules, however, has been to reverse transcribe the RNA into DNA that can be cloned and sequenced (for example, see refs. 15 and 16).

Molecular cloning of poliovirus-specific cDNA was prohibited by guidelines of the National Institutes of Health when we initiated a study of determining the primary structure of poliovirus RNA. We therefore developed a strategy that enabled us to use Sanger's dideoxynucleotide method [often referred to also as the chain-termination method (8)] without recourse

to cloned cDNA. Our procedure is useful for determining the sequence of any high molecular weight RNA. It can be applied also in studies of the primary structure of high molecular weight DNA for which a specific RNA transcript (e.g., mRNA) can be isolated.

Here we describe our method that has thus far yielded segments totaling 6500 nucleotides. Of the known sequences, we present a segment of 1060 3'-terminal nucleotides.†

MATERIALS AND METHODS

Reverse transcriptase from avian myeloblastosis virus (generously supplied by J. W. Beard, Research Resources, National Cancer Institute) was purified (17) as follows. To 1000 units of enzyme in 0.1 ml of 0.2 M potassium phosphate, pH 7.2/2 mM dithiothreitol/0.2% Triton-X-100/50% (vol/vol) glycerol was added 4 vol of column buffer (0.01 M potassium phosphate, pH 8.0/1 mM dithiothreitol/50% glycerol). The solution was passed through a column of poly(C)-agarose (P-L Biochemicals; bed volume, 0.4 ml). After a wash with 5 ml of column buffer, the enzyme was eluted with column buffer containing 0.4 M KCl; 0.3-ml fractions were assayed for activity. Nearly carrier-free [γ -³²P]ATP was prepared from ³²P_i and ADP as described (18, 19).

Poliovirus type 1 (Mahoney) and virion RNA were isolated as described (7).

cDNA was synthesized as described (20), usually in 1 ml of a mixture containing 50 mM Tris-HCl (pH 8.3), 8 mM MgCl₂, 0.4 mM dithiothreitol, 80 mM KCl, 4 mM sodium pyrophosphate, [³H]dATP, dGTP, dCTP, and TTP, at 0.4 mM each and 6 μ g of (dT)₁₂, 50 μ g of poliovirus RNA, and 80-160 units of reverse transcriptase per ml. After incubation for 2 hr at 37°C, the reaction mixture was made 10 mM in EDTA and deproteinized with phenol, and the nucleic acids were precipitated with ethanol. The precipitate was dissolved in 0.15 M NaOH/1 M NaCl/0.1 mM EDTA and fractionated by centrifugation through 15-30% sucrose prepared in the same solution (Beckman SW 50.1 rotor, 32,000 rpm, 12 hr, at 20°C). Yield of the fastest migrating material (nearly full-length cDNA) was 10% of the input RNA.

Poliovirus [³²P]RNA (200 μ g) was digested with RNase T1 (100 units) and calf intestine alkaline phosphatase (4 units, Boehringer Mannheim) in 0.2 ml of 0.1 M Tris-HCl (pH 8.0) for 1 hr at 37°C. The larger RNA fragments together with coprecipitating smaller fragments were precipitated with ethanol, redissolved in 30 μ l of 50 mM Tris-HCl, pH 7.5/2 mM EDTA, mixed with 60 μ l of 9 M urea, 10 μ l of glycerol, and 2 μ l of 2% bromophenol blue, and subjected to two-dimensional polyacrylamide gel electrophoresis (7). Oligonucleotides were

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

* This is paper no. 5 in the series "Sequence studies of poliovirus RNA." Paper no. 4 is ref. 7.

† A preliminary account of this work has been presented (1).

eluted with H₂O (7), adsorbed to columns of DEAE-cellulose (bed volume, 0.05 ml), washed with 0.1 M triethylammonium bicarbonate (pH 7.5), eluted with 1 M triethylammonium bicarbonate (pH 7.5), and freed from salt by lyophilization. Recovery was 30–40% based on input RNA. After phosphorylation with polynucleotide kinase and [γ -³²P]ATP (10), the oligonucleotides were purified by polyacrylamide gel electrophoresis (10) and DEAE-cellulose column chromatography. RNase A-resistant oligonucleotides were prepared from 100 μ g of poliovirus [³²P]RNA by digestion with 7.5 μ g of RNase A in 100 μ l of 0.3 M NaCl/0.05 M Tris-HCl, pH 7.5/2 mM EDTA for 1 hr at 37°C followed by precipitation with ethanol and two-dimensional polyacrylamide gel electrophoresis. Individual RNase A-resistant oligonucleotides were eluted, dephosphorylated, and 5'-labeled as described above.

The 5'-³²P-labeled oligonucleotide primer (5–50 \times 10⁵ cpm) was dissolved in 6 μ l of 0.1 mM EDTA containing 1 μ g of cDNA; 1 μ l of 0.5 M NaCl/66 mM Tris-HCl, pH 7.4/66 mM MgCl₂/10 mM dithiothreitol (8) was added. The mixture was heated to 100°C for 3 min and then incubated at 67°C for 30 min in a silicone-treated capillary. DNA synthesis was as described by Sanger *et al.* (8) with modifications: 2 μ l of the ddN mixture (e.g., 0.5 mM ddGTP, 0.005 mM dGTP, and other dNTPs at 0.1 mM) were mixed with 2 μ l of the primer/cDNA solution and 1 μ l (1 unit) of DNA polymerase and incubated at 37°C for 30 min. The reaction was stopped with 1 μ l of 0.2 M EDTA and 6 μ l of 10 M urea/0.05% xylene cyanol/0.05% bromophenol blue; 3 μ l of the final mixture was heated to 100°C for 3 min and applied to a sequencing gel (12).

General Strategy for Poliovirus RNA Sequence Determination. Our method is summarized in Fig. 1. (i) cDNA to poliovirus RNA is synthesized with reverse transcriptase. (ii) Poliovirus RNA is digested with RNase T1 and phosphomonoesterase; oligonucleotides are separated by two-dimensional polyacrylamide gel electrophoresis (7). (iii) An oligonucleotide is selected on the basis of its position in the genomic RNA (20), eluted from the gel, and labeled at the 5' end. (iv) The 5'-³²P-labeled oligoribonucleotide is used as a primer for DNA synthesis with *E. coli* DNA polymerase I (Klenow) in the presence of unlabeled dNTP and specific chain terminators (ddNTP) (8).

In four polymerization mixtures, each containing a specific ddNTP, DNA fragments are produced that terminate opposite one of the four bases. The DNA fragments are separated by slab gel electrophoresis. Each ddNTP generates a different ladder of bands, and use of all four ladders permits a reading of the sequence complementary to the cDNA template. This ingenious method developed by Sanger and his colleagues has been modified by us in that we use labeled primers and unlabeled triphosphates instead of labeled nucleoside triphosphates. The modification proved crucial in that it reduced the background of radioactivity in sequencing gels to nearly zero. Overlaps of DNA sequences and the oligoribonucleotide sequences served as controls for the sequence generated by the dideoxy method.

RESULTS

Synthesis of cDNA to Poliovirus RNA. The products of reverse transcription were separated by zonal centrifugation through an alkaline sucrose gradient, and only cDNA in the leading fractions of the peak was used. When analyzed on alkaline agarose gels (Fig. 2), only one band of cDNA was visible; it migrated slightly faster than plasmid DNA pSM3 (estimated to be 7.7 kilobases long) and slightly slower than pSM2 DNA (estimated to be 7.0 kilobases long) (22). When analyzed in a

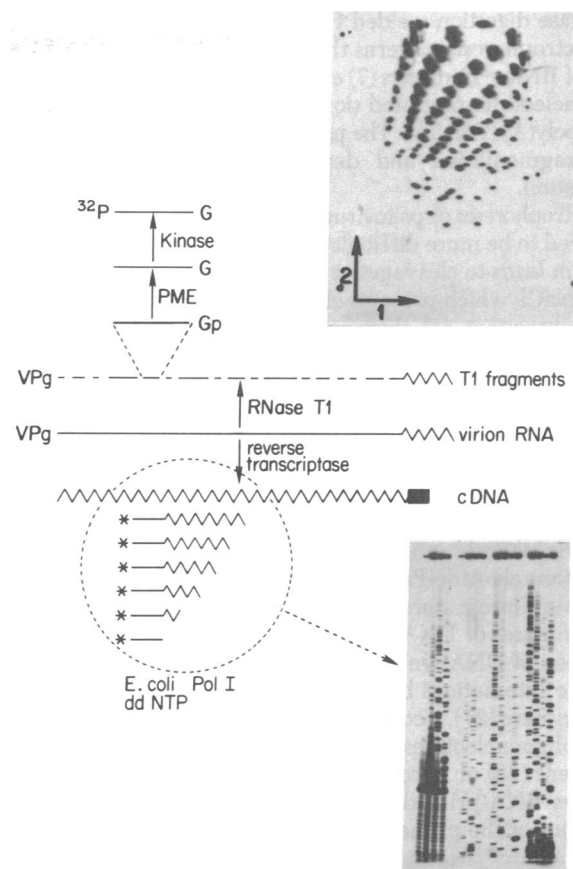


FIG. 1. Strategy for determining the primary structure of poliovirus RNA. Primers for the chain termination method are generated by RNase T1 digestion of virion RNA followed by two-dimensional gel electrophoresis (Upper Right) and enzymatic modification. (Lower Right) Sequencing gel. PME, phosphomonoesterase; *, 5'-³²P of the primers; VPg, genome-linked protein. For details, see text.

similar fashion, unfractionated reverse transcription product yielded smaller cDNA in addition to the 7.4-kilobase band; discrete DNA bands of partial transcripts (23), however, did not occur under our conditions of synthesis (data not shown). The average size of the cDNA product did not change when virus RNA that had been pretreated with proteinase K to destroy the 5'-linked protein [VPg (24)] was used as template (data not shown). Oligonucleotide-primed DNA synthesis provided evidence that the cDNA shown in Fig. 2 contained molecules with 3' ends that terminated within 100 nucleotides from the 5' end of genome RNA (unpublished results). Attempts to transfect HeLa cells (26) with the cDNA failed.

Preparation of Oligonucleotide Primers. Digestion of RNA with RNase T1 and simultaneously with calf intestine alkaline phosphatase yields oligonucleotides suitable as primers. Because of nuclease contamination, bacterial alkaline phosphatase has not been used. The products of an RNase T1/phosphomono-

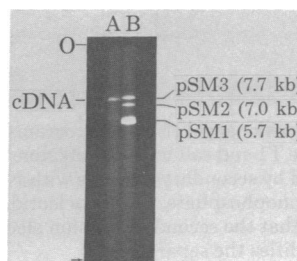


FIG. 2. Analysis of poliovirus cDNA (lane A) by alkaline agarose gel electrophoresis (25). Marker DNAs (lane B) were pSM plasmids [a gift from E. Ohtsubo (22)] cleaved once with *Eco*RI. After electrophoresis, the cylindrical gels were neutralized and stained with ethidium bromide (25). O, top of the gel; arrow, bottom of the gel; kb, kilobases.

noesterase digestion yielded two-dimensional polyacrylamide gel electrophoresis patterns that were indistinguishable from those of RNase T1 digests (7) except that the dephosphorylated oligonucleotides migrated slower in both dimensions relative to the poly(A) (Fig. 3A). The pattern shown in Fig. 3A indicates that fragmentation and dephosphorylation proceeded to completion.

Electrophoreses of poliovirus RNA after cleavage with RNase A proved to be more difficult because (i) digestion in low-salt medium leads to cleavages at A residues, and (ii) digestion in 0.3 M NaCl, which prevents cleavages at A residues, interferes with subsequent gel electrophoresis. Suitable patterns were obtained, however, when the oligonucleotides were freed from salt by precipitation with ethanol (Fig. 3B).

The use of oligonucleotides eluted from polyacrylamide gels as substrate for kinase or as primer in DNA synthesis requires purification of the oligonucleotides by DEAE-cellulose column chromatography. If this purification step is omitted, the yield of phosphorylation with kinase is low and DNA synthesis is totally inhibited.

Oligonucleotide-Primed DNA Synthesis in the Presence of Chain Terminators. In initial experiments, we used reverse transcriptase for DNA synthesis without success. Apparently, initiation of DNA synthesis with oligoribonucleotide requires specific interactions between primer, template, and reverse transcriptase (27), a condition not provided by poliovirus-specific oligonucleotides and poliovirus cDNA. The Klenow fragment (28) of *E. coli* polymerase I, on the other hand, facilitates RNA-primed DNA synthesis independently of the sequence of the primer. However, when α - ^{32}P -labeled deox-

ynucleoside triphosphates were used as label (8), background incorporation in the absence of primer was high, making it impossible to read sequences from autoradiograms of the sequencing gels. Attempts to remove putative endogenous primers, and thus reduce the background incorporation, failed. The problem was solved by using 5'- ^{32}P -labeled oligonucleotides as primers and carrying out DNA synthesis with unlabeled dNTPs.

Fig. 4B shows a set of ladders generated by T1-oligonucleotide no. 48 that has the sequence C-U-U-C-C-U-A-C-C-C-C-C-A-U-G (data not shown). The reaction mixture was loaded twice onto an 8% polyacrylamide gel under conditions such that approximately 150 nucleotides can be read. In order to read up to 250 nucleotides, the mixtures were loaded onto 6% gels and electrophoresed longer (not shown).

The dark area at the bottom of part 2 of Fig. 4B represents the primer and degradation products of the primer. Limited degradation of the primer, occurring during enzymatic reactions and subsequent separation procedures, does not interfere with the analysis. We have found it impossible, however, to read the bases immediately following the primer. The reason for this phenomenon, which has also been observed by others (29), is not clear. This problem can be solved by generating overlapping sequences, a procedure required in any case for ordering sequences. In the analysis shown in Fig. 4B, an overlapping sequence was obtained by priming DNA synthesis with T1-oligonucleotide no. 50 (see Fig. 5).

Also shown in Fig. 4B is a "compression" of bands within the sequence C-G-C-T-A(G,T,C)T-C-C-T-A-G (corresponding to the nucleotide sequence -427 to -440 of the poliovirus RNA).

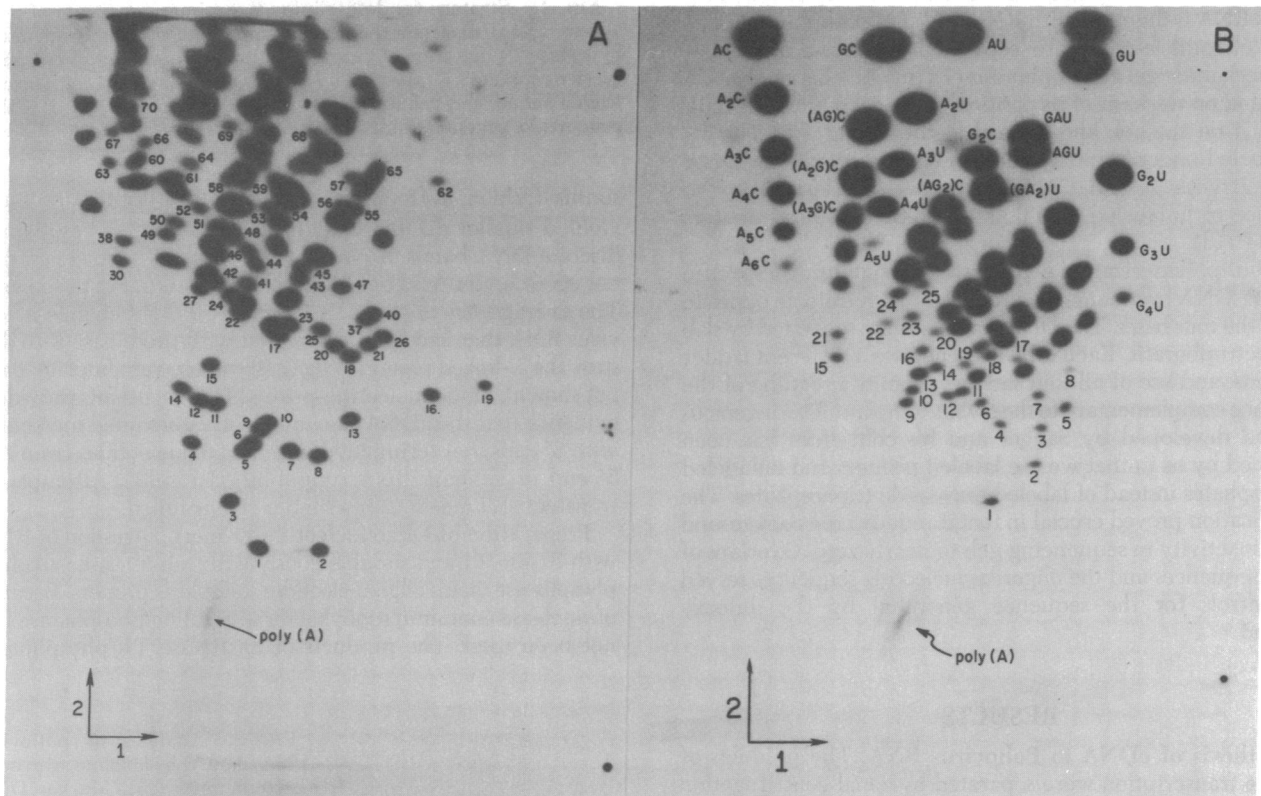


FIG. 3. Autoradiograms of two-dimensional gel electropherograms of enzymatic digests of type 1 poliovirus (Mahoney) [^{32}P]RNA. (A) RNA was digested simultaneously with RNase T1 and calf intestine alkaline phosphatase. (B) RNA was digested with RNase A. The base composition of oligonucleotides in B was determined by secondary cleavage with KOH followed by paper electrophoresis on Whatman 3MM at pH 3.5. The spot above GU is uridine 2',3'-cyclic monophosphate. Mononucleotides (Gp in A; Cp and Up in B) and some small oligonucleotides in A have been electrophoresed off the gel. Note that the second dimension electrophoresis was carried out without cooling, according to a modification of Lee *et al.* (7) that considerably simplifies the separation.

The severe problem of background incorporation that we encountered in our system was solved when the newly synthesized DNA strands were labeled with 5'-³²P-labeled primers rather than with [α -³²P]dNTP. This modification also eliminated the need for a "chase step" after DNA synthesis (8). Moreover, 5'-labeling of primers with kinase and [γ -³²P]ATP is much less costly than internal labeling of the DNA chains with [α -³²P]NTP because we can synthesize the [γ -³²P]ATP in our laboratory from phosphorus-32 and ADP in a rapid and simple procedure (18, 19).

The selection of oligonucleotide primers according to their physical map position within the genome (21) permitted us to order nonoverlapping sequences during the early stages of sequencing. In general, knowledge of the map position of primers is useful, particularly when only specific segments of high molecular weight RNA are to be analyzed. This is true, for example, in studies of genetic variation of viral genomes.

In principle, our strategy for sequence analyses of RNA can be applied to any high molecular weight RNA, provided that cDNA can be prepared. For analyses of RNAs of minus-strand viruses this might be facilitated by polyadenylation of genome RNAs (16, 37).

The 3' End of Poliovirus RNA. From the 3'-terminal nucleotide sequence of poliovirus shown in Fig. 5, the following observations can be made. (i) There are 60 termination codons, 37 of which occur in all three reading frames between poly(A) and position -562. We consider it likely that the UGA at position -562 is the codon terminating translation of poliovirus RNA because of the absence of termination codons in the 500 preceding nucleotides in the same reading frame (but their presence in the other two reading frames) and the sudden high frequency of occurrence following nucleotide -562. A segment of 562 untranslated bases at the 3' end of poliovirus RNA seems surprising. It should be noted that long segments of noncoding nucleotides at the 3' end have also been observed in other eukaryotic mRNAs (34, 38).

(ii) The sequence does not contain a hexanucleotide (A-A-U-A-A-A) that characteristically occurs in the vicinity of poly(A) (within 20 bases) in all untranslated regions of mammalian cytoplasmic mRNAs known to date (see ref. 39). It has been proposed that the hexanucleotide may serve as a signal for polyadenylation of transcripts in the cell nucleus (32). The absence of A-A-U-A-A-A at the 3' terminus of poliovirus RNA is not in disagreement with this hypothesis because poly(A) in poliovirus RNA is synthesized by transcription from poly(U) and not by posttranscriptional polyadenylation (40).

We thank Akiko Kitamura for excellent technical assistance, Carol A. Carter, Nancy Reich, and Bert L. Semler for discussions and criticism of the manuscript, and Sandi Donaldson for the preparation of the manuscript. We are grateful to Hisako Ohtsubo, Eiichi Ohtsubo, and Ian Kennedy for their invaluable suggestions on DNA sequence determination techniques and the gift of pSM DNAs. This work was supported in part by Grant AI-15122 from the National Institute of Allergy and Infectious Diseases.

1. Kitamura, N., Adler, C. & Wimmer, E. (1980) *Ann. N.Y. Acad. Sci.*, in press.
2. Lee, Y. F., Nomoto, A. & Wimmer, E. (1976) *Prog. Nucleic Acid Res. Mol. Biol.* **19**, 89-95.
3. Lee, Y. F., Nomoto, A., Detjen, B. M. & Wimmer, E. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 59-63.
4. Flanagan, J. B., Pettersson, R. F., Ambros, V., Hewlett, M. J. & Baltimore, D. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 961-965.

5. Nomoto, A., Detjen, B., Pozzatti, R. & Wimmer, E. (1977) *Nature (London)* **268**, 208-213.
6. Yogo, Y. & Wimmer, E. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1877-1882.
7. Lee, Y. F., Kitamura, N., Nomoto, A. & Wimmer, E. (1979) *J. Gen. Virol.* **44**, 311-322.
8. Sanger, F., Nicklen, S. & Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
9. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560-564.
10. Donis-Keller, H., Maxam, A. M. & Gilbert, W. (1977) *Nucleic Acids Res.* **4**, 2527-2538.
11. Simonscits, A., Brownlee, G., Brown, R. S., Rubin, J. R. & Guillely, H. (1977) *Nature (London)* **269**, 833-836.
12. Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87**, 107-116.
13. Nathans, D. (1979) *Science* **206**, 903-909.
14. Donis-Keller, H. (1979) *Nucleic Acids Res.* **7**, 179-192.
15. Efstratiadis, A., Kafatos, F. C. & Maniatis, T. (1977) *Cell* **10**, 571-585.
16. Porter, A. G., Barber, C., Carey, N. H., Hallewell, R. A., Threlfall, G. & Emrtage, J. S. (1979) *Nature (London)* **282**, 471-477.
17. Marcus, S. L., Modak, M. J. & Cavalieri, L. F. (1974) *J. Virol.* **14**, 853-859.
18. Johnson, R. A. & Walseth, T. F. (1979) *Adv. Cyclic Nucleotide Res.* **10**, 135-167.
19. Larsen, G. R., Dorner, A. J., Harris, T. J. R. & Wimmer, E. (1980) *Nucleic Acids Res.* **8**, 1217-1229.
20. Kacian, D. L. & Myers, J. C. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2191-2195.
21. Nomoto, A., Jacobson, A., Lee, Y. F., Dunn, J. J. & Wimmer, E. (1979) *J. Mol. Biol.* **128**, 179-196.
22. Ohtsubo, E., Zenilman, M. & Ohtsubo, H. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 750-754.
23. Efstratiadis, A., Maniatis, T., Kafatos, F., Jeffrey, A. & Vournakis, J. (1975) *Cell* **4**, 367-378.
24. Nomoto, A., Kitamura, N., Golini, F. & Wimmer, E. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5345-5349.
25. McDonnell, M. W., Simon, M. N. & Studier, F. W. (1977) *J. Mol. Biol.* **110**, 119-146.
26. Detjen, B. M., Lucas, J. J. & Wimmer, E. (1978) *J. Virol.* **27**, 582-586.
27. Cordell, B., Swanstrom, R., Goodman, H. & Bishop, J. (1979) *J. Biol. Chem.* **254**, 1866-1874.
28. Klenow, H. & Henningsen, I. (1970) *Proc. Natl. Acad. Sci. USA* **65**, 168-175.
29. Gough, N. M., Cory, S. & Adams, J. M. (1979) *Nature (London)* **281**, 394-396.
30. Brownlee, G. G. & Cartwright, E. M. (1977) *J. Mol. Biol.* **114**, 93-117.
31. Mills, D. R. & Kramer, F. R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 2232-2235.
32. Porter, A. G., Fellner, P., Black, D. N., Rowlands, D. J., Harris, T. J. R. & Brown, F. (1978) *Nature (London)* **276**, 298-301.
33. Sanger, F. & Coulson, A. (1975) *J. Mol. Biol.* **94**, 441-448.
34. Hamlyn, P. H., Brownlee, G. G., Cheng, C.-C., Gait, M. J. & Milstein, C. (1978) *Cell* **15**, 1067-1075.
35. Zimmern, D. & Kaesberg, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4257-4261.
36. Kramer, F. R. & Mills, D. R. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5334-5338.
37. Rowlands, D. J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4793-4797.
38. Catterall, J. F., O'Malley, B. W., Robertson, M. A., Staden, R., Tanaka, Y. & Brownlee, G. G. (1978) *Nature (London)* **257**, 510-513.
39. Tucker, P. W., Marcu, K. B., Slightom, J. L. & Blattner, F. R. (1979) *Science* **206**, 1299-1303.
40. Dorsch-Häsler, K., Yogo, Y. & Wimmer, E. (1975) *J. Virol.* **16**, 1515-1527.