Proc. Natl. Acad. Sci. USA
Vol. 75, No. 9, pp. 4257–4261, September 1978
Biochemistry

3′-Terminal nucleotide sequence of encephalomyocarditis virus RNA determined by reverse transcriptase and chain-terminating inhibitors

[2′,3′-dideoxynucleoside triphosphate/picornavirus/poly(A)/poliovirus]

DAVID ZIMMERN AND PAUL KAESBERG

Biophysics Laboratory of the Graduate School and Biochemistry Department of the College of Agricultural and Life Sciences, University of Wisconsin, Madison, Wisconsin 53706

Communicated by Henry Lardy, June 29, 1978

ABSTRACT **We have adapted the chain-termination method for determining the nucleotide sequence of DNA of Sanger, Nicklen, and Coulson [(1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467] for use with reverse transcriptase (RNA-directed DNA nucleotidyltransferase) on RNA templates. With this method and using a primer (the octanucleotide pdT$_7$rC) directed at the 3′-terminal poly(A), we have determined a sequence of 166 residues in the genomic RNA of the picornavirus encephalomyocarditis virus.**

Techniques for the primary structural analysis of large RNA molecules have improved steadily (1–6), though not as rapidly as in DNA structural analysis after the introduction of restriction enzymes, recombinant DNA technology, and sequence determination on gels. Indeed, two of the recent techniques for determining the nucleotide sequence in RNA simply involve prior conversion into DNA (4–6), and all are ultimately derived from the new DNA technology. One highly successful method for determining the nucleotide sequence in DNA is the "plus and minus" technique of Sanger and Coulson (7), which has been adapted for use on RNA templates by Brownlee and Cartwright (1). Sanger *et al.* (8) have recently described a second generation rapid method for determining the nucleotide sequence of DNA, relying, like the plus and minus method, on enzymatic synthesis of a complementary strand, and improving on it in speed, accuracy, sensitivity, convenience, and range. The method involves incorporation of chain-terminating deoxynucleoside triphosphate analogues [such as the 2′,3′-dideoxy analogues (ddNTPs)] of each base in turn into the reaction mixture, and hence into the growing chain, thus generating four nested sets of fragments each terminated opposite one of the four bases in the template.

Since ddNTPs are competitive inhibitors of reverse transcriptase (RNA-directed DNA nucleotidyltransferase) (9, 10), it seemed to us that this technique could also be adapted for use in determining the nucleotide sequence of RNA. We have found that a straightforward adaptation in the spirit of Brownlee and Cartwright yields a powerful technique with which we have been able to determine the nucleotide sequence adjacent to the 3′-terminal poly(adenylic acid) tract in the picornavirus encephalomyocarditis virus (EMC virus). The only major modification necessary was to lower the ddNTP concentration considerably, since, as others have noted (9), reverse transcriptase is much more sensitive to these inhibitors than is *Escherichia coli* DNA polymerase I. The method offers many of the advantages of its parent DNA technique and promises to be especially useful in sequence analysis of large RNA molecules. A special advantage of direct methods such as this for studying pathogens of higher organisms is that they avoid whatever risks may be associated with techniques involving cloning of complementary DNA in *E. coli* (6).

We chose the poly(A)-adjacent sequences of picornavirus genomic RNAs, particularly that of EMC virus, for development of the RNA dideoxy method because primers capable of phased extension into these sequences are readily obtainable and, since the sequence of 26 nucleotides adjacent to the poly(A) tract in EMC virus had been published (11), we were able to assess the accuracy of the method. Furthermore, these regions may be relevant to the unsolved problem of the mechanism of picornaviral RNA replication.

MATERIALS AND METHODS

Enzyme and Nucleotides. Reverse transcriptase from avian myeloblastosis virus was generously supplied by J. W. Beard (Life Sciences, Inc.; St. Petersburg, FL) via the Office of Program Resources and Logistics, National Cancer Institute. Deoxynucleoside [α-^{32}P]triphosphates at initial specific activities of 350 Ci/mmol were obtained from Amersham/Searle. Unlabeled deoxynucleoside triphosphates were from Boehringer Mannheim. 2′,3′-Dideoxynucleoside triphosphates and oligo(dT)$_{10}$ were from P-L Biochemicals, as was the octanucleotide primer pdT$_7$rC, which was further purified by alkaline hydrolysis (12) and treatment with *E. coli* alkaline phosphatase (13) to ensure that no primer molecule contained more than one ribonucleotide (stage 1 purified primer). A further purification step by preparative electrophoresis-homochromatography (13), followed by alkaline hydrolysis to remove carrier RNA from the homomixture, was later introduced to remove an approximately 1:1 molar contamination of stage 1 pdT$_7$rC by oligo(dT)$_n$ (n = 5–8), yielding stage 2 purified primer.

Viruses and RNAs. Purified EMC virus was a gift from R. R. Rueckert. The virus was originally obtained from C. Fuerst, University of Toronto, and adapted to growth in sarcoma 180 ascites cells through 10–15 passages. Subsequent isolation of clonal lines and methods of growth and preparation of the virus have been described (14). RNA was extracted from the virus (4 mg/ml in 0.1 M NaCl/10 mM Tris·HCl, pH 7.5/1 mM EDTA) with 2 vol phenol/chloroform (1:1, vol/vol) and 1% sodium dodecyl sulfate, precipitated three times with ethanol, and stored frozen in water at −20°C. RNA concentration was measured by absorbance; $A^{260\,nm}_{0.1\%}$ was taken to be 22 (15).

Poliovirus RNA (strain Mahoney) was a gift of Mark Pal-

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.

Abbreviations: EMC virus, encephalomyocarditis virus; ddNTP, 2′, 3′-dideoxynucleoside triphosphate; cDNA, complementary DNA.

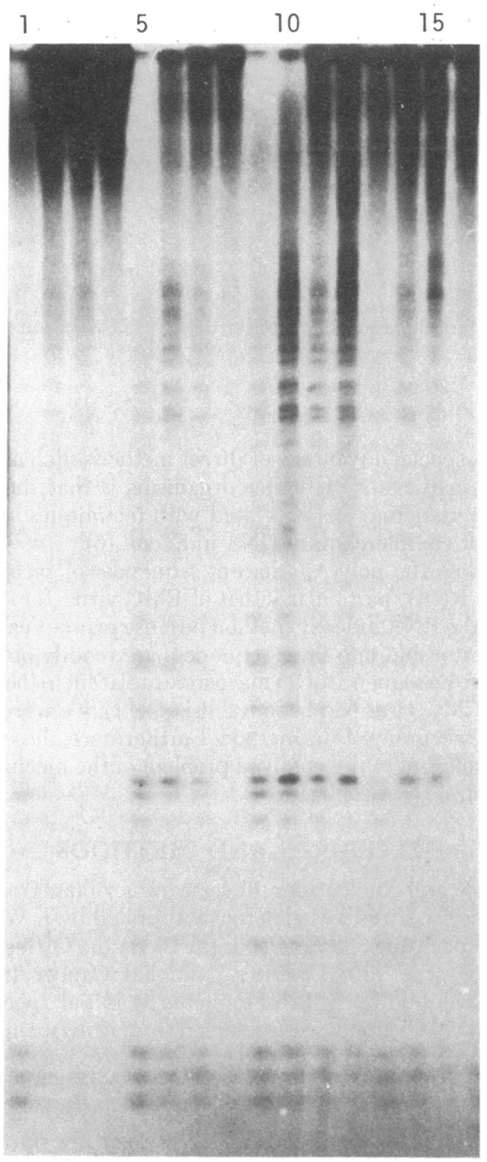

FIG. 1. Optimization of reaction conditions. The template/primer complex was polio RNA with pdT$_7$rC (stage 1). Reactions 1–4 were incubated at 25°C and reactions 5–16 at 39°C, all for 1 hr. The subsequent 20-min chase was at 39° in both cases. (Under the conditions of the chase no further termination occurs; material synthesized during the chase migrates at the top of the gel.) Reactions 1–8 were at 50 μM dCTP and dGTP; reactions 9–16 were at 10 μM. dTTP concentrations were: 5 μM in reactions 1, 5, 9, and 10; 1 μM in reactions 2, 6, 11, and 12; 0.5 μM in reactions 3, 7, 13, 14, and 15; and 0.1 μM in reactions 4, 8, and 16. ddTTP concentrations were: 10 μM in reactions 1, 5, and 9; 5 μM in reaction 10; 1 μM in reactions 2, 6, 11, and 13; 0.5 μM in reactions 3, 7, 12, and 14; and 0.1 μM in reactions 4, 8, 15, and 16. dATP was the label. Reactions were stopped by addition of 1 μl of 0.1 M EDTA; 1 μg of pancreatic RNase was added and incubation was continued for 1 more hr. The gel was run at 325 V for 17 hr.

lansch and R. R. Rueckert. The virus was obtained from M. Hatch (Center for Disease Control, Atlanta, GA).

Sequence Determinations. Template/primer mixtures were prepared by mixing either (*i*) 0.625 μg of stage 1 primer and 6.25 μg of poliovirus RNA (2.3 pmol) in 25 μl of 40 mM KCl or (*ii*) 1 μg of stage 2 primer and 17 μg of EMC RNA (about 6 pmol) in 25 μl of 80 mM KCl. Template/primer mixtures were usually preincubated for 10 min at 37°C (though this may not be strictly necessary) and were stored frozen at −20°C.

Four chain-terminating nucleotide stock mixtures ("dideoxymixes") were prepared, one for each dideoxynucleotide as follows: ddA mixture contained 100 μM dCTP, 100 μM dGTP, 100 μM dTTP, and 5 μM ddATP; ddC mixture contained 50 μM dCTP, 100 μM dGTP, 100 μM dTTP, and 10 μM ddCTP; ddG mixture contained 100 μM dCTP, 50 μM dGTP, 100 μM dTTP, and 10 μM ddGTP; and ddT mixture contained 100 μM dCTP, 100 μM dGTP, 50 μM dTTP, and 10 μM ddTTP. These mixtures are designed for use with [α-^{32}P]dATP as label, and so contain no dATP. By suitably rearranging the contents of the mixtures any of the other three dNTPs can be used as an alternative label. The above mixtures give 1 μM ddNTP and 5 μM of the corresponding dNTP after dilution (a "5/1" mixture). In some experiments an alternative mixture containing 20 μM of both the ddNTP and of its corresponding dNTP was used, giving 2 μM of both after dilution (a 2/2 dideoxy mixture). Concentrations of both deoxy- and dideoxynucleotide triphosphates were estimated by UV spectroscopy using the published molar extinction coefficients of the deoxynucleoside monophosphates.

Reverse transcriptase buffer (4 times concentrated stock solution) was 200 mM Tris·HCl, pH 8.0/200 mM KCl/20 mM $MgCl_2$/40 mM dithiothreitol (1). Dideoxy mixtures may be diluted and/or combined appropriately with buffer stock and frozen in aliquots at −20°C.

The reaction mixtures for sequence determinations usually contained 5 μl of distilled water, 1 μl of the relevant dideoxy mixture, 2.5 μl of reverse transcriptase buffer (4 times concentrated stock solution), 1 μl of template/primer mixture, and 0.5 μl (3.15 units) of reverse transcriptase. Exceptions are noted in the figure legends. The reaction mixtures (less enzyme) were assembled in drawnout capillary tubes while the label (usually 10 pmol = 3 μCi/reaction) was dried down in a siliconized glass tube by use of a vacuum desiccator. The dry label was taken up in the partly assembled dideoxy mixture and the reaction was started by addition of enzyme. Incubation was for 1–3.5 hr at 37°–41°C. One microliter of a 0.5 mM solution of unlabeled deoxynucleoside triphosphates was then added to chase, for an additional 0.5 hr, complementary DNA (cDNA) molecules not terminated with a ddN residue. Finally, 6 μl of 0.05% xylene cyanol FF/0.05% bromophenol blue/8 M urea was added to each incubation, and the mixture was heated to 90°–100°C for 2 min, chilled on ice, and loaded onto a 12% polyacrylamide/0.5% bisacrylamide/8 M urea gel (7) run in 87.5 mM Tris borate, pH 8.3/2.75 mM EDTA (16) in a 40 × 20 cm gel apparatus (Raven Scientific, Haverhill, U. K.). After electrophoresis the gel was fixed for 15–30 min in 10% acetic acid and autoradiographed.

RESULTS AND DISCUSSION

In preliminary experiments we took as our starting point the protocol of Sanger *et al.* (8), making the relevant substitutions of enzyme, template, and buffer but retaining the original ratio of deoxynucleoside triphosphates (dNTPs) to dideoxynucleoside triphosphates (ddNTPs). Under these conditions we found that cDNA synthesis was totally inhibited. We therefore experimented with the effects of altering the ratio of dNTP to ddNTP by either lowering the ddNTP concentration or raising the concentration of the corresponding dNTP. With the dNTP concentration held at 100 μM we found that there was an abrupt transition from no synthesis (nothing on the gel) at 100 μM ddNTP to extensive synthesis (everything at the top of a 12% gel) at 10 μM, without any sign of specific termination in the size range of interest. At low dNTP concentrations (0.1–10 μM), however, we observed specific termination at dNTP:ddNTP ratios between 10:1 and 1:2 with a useful distribution

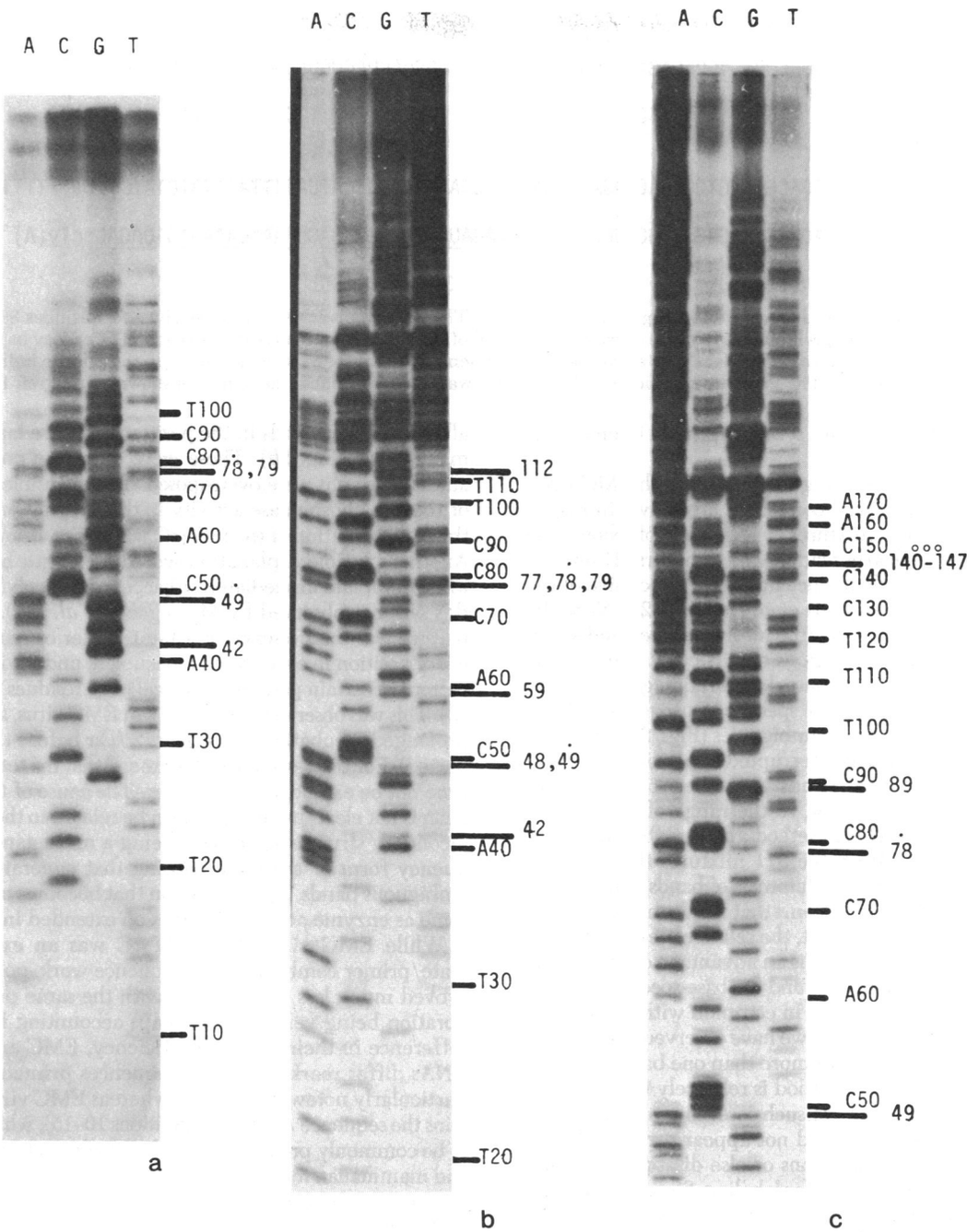

FIG. 2. Sequencing gels showing the sequence primed by pdT$_7$rC on EMC virus RNA. Reaction conditions were as described in *Materials and Methods*, with stage 2 primer and with dATP as the label, except for *c*, where the label was dTTP. (*a*) 2/2 dideoxy mixtures were used; (*b* and *c*) 5/1 dideoxy mixtures were used. The primer was not removed before electrophoresis. Gels were run at 400–500 V. Every tenth residue of the cDNA sequences read is marked. Also marked are residues that were ambiguous on a particular gel; those that were ambiguous on all the gels examined are denoted by a black dot. The open circles indicate a region of compression.

of product sizes. Exploring this range in more detail, we examined the effect of altering the concentration of driver dNTPs (those corresponding neither to the labeled nucleotide nor to the ddNTP), of delayed addition of ddNTP (20 min after the start of synthesis), and of carrying out the incubation at room temperature instead of physiological temperature (Fig. 1). None of these variables had as profound an effect as simple manipulation of the ratio and absolute values of the dNTP and ddNTP concentrations. We therefore selected two sets of conditions for further study, using dNTP:ddNTP molar ratios of 5:1 (referred to as a "5/1 mixture") and of 2:2 (a "2/2 mixture"). (Since the labeled deoxynucleoside triphosphate was usually present at 1 μM or less, a correspondingly lower ddNTP concentration was used; see *Materials and Methods*).

Since the method requires a defined starting point to ensure that consecutive bands on the gel represent elongation by consecutive bases, we used an oligo(dT) primer extended at the 3′ end by one nucleotide complementary to the base 5′-adjacent to the poly(A) in the template (1, 17). This base has been reported to be a guanine residue in both poliovirus and EMC RNAs (11, 18). We confirmed these results by determining that, for both RNAs, dCTP alone labeled oligo(dT) in the presence of enzyme (10, 19). Hence, the primer we used was pdT$_7$rC. The 3′-terminal ribonucleotide allows primer removal with

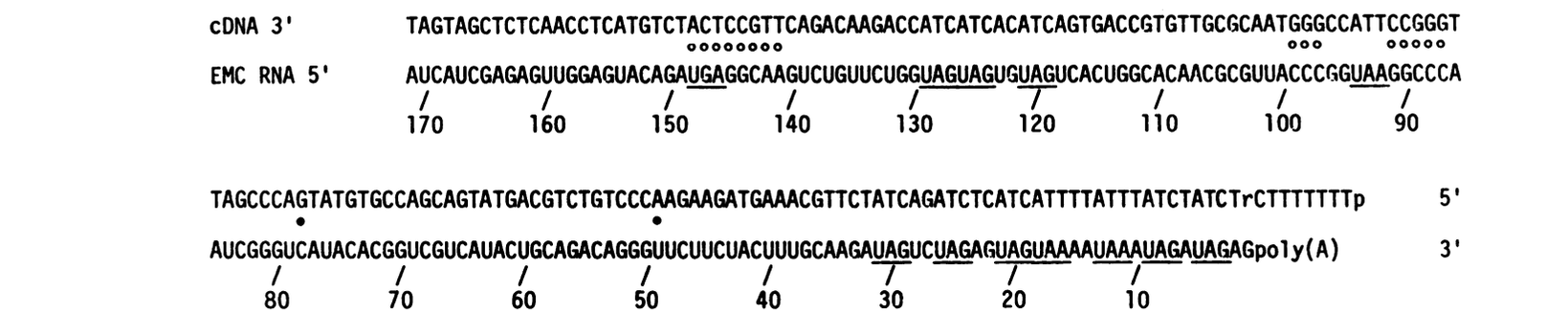

FIG. 3. Nucleotide sequences deduced from the gels shown in Fig. 2. The EMC virus RNA sequences given for residues 86–91 inclusive, and for residue 140 onwards are preliminary. Residue 49 may be C instead of U, and residue 78 may be A instead of C, or may be spurious. Ambiguities found on all gels are indicated by black dots above the RNA sequence. Regions of band compression in the gels are indicated by open circles above the RNA sequence. The sequence of residues 2–5, inclusive, was not determined; the sequence shown is from ref. 11.

pancreatic RNase or alkali after synthesis, which is sometimes an advantage.

The results we obtained using this primer with EMC RNA as a template under our trial conditions are shown in Fig. 2. With the 2/2 dideoxy mixtures a sequence of about 100 nucleotides could be clearly read in which the first 21 nucleotides corresponded exactly to nucleotides 6–26 inclusive of the sequence derived by Merregaert *et al.* (11) (Fig. 2*a*). With the 5/1 dideoxy mixtures the readable information extended out to position 171, with three regions of difficulty due to compression (an artifact of inadequate denaturation general to gel sequencing methods and not a specific consequence of the present method) and two other ambiguities (Figs. 2 *b* and *c* and 3). The readable sequence is interrupted at position 176 by an extreme compression of the bands. A further 25 residues are readable beyond this, however, so the range of the method approaches that of the DNA method and may be capable of further improvement. With the 5/1 mixtures there was no difficulty with disproportionate faintness of bands in the earlier portion of the sequence, so it seems that this formulation may be more generally useful. In fact, the relatively even intensity of all the bands in a given track is an advantage over both the RNA plus and minus method (1) and the base-specific cleavage methods that use RNases (2, 3). In common with other enzymatic copying methods (1, 7, 8), we have observed a few points of ambiguity in the gels where more than one band occurred at a given position, but the method is relatively free of them. Furthermore, in the majority of such cases the weaker band(s) of such a set of bands either did not appear reproducibly in otherwise identical incorporations or else disappeared when a different nucleotide was used for labeling. Since we have not yet observed a case where the strongest member of such a set of bands disappeared in this way, we have assumed that the strongest band gives the correct assignment. In some cases, the multiple bands arose close to runs of identical nucleotides and/or the base of possible hairpin loops (e.g., at position 78), and we assume the bands result from the enzyme pausing at these points. There are, however, other formal possibilities that could account for the existence of these minor bands, including sequence variants in the template RNA population (despite recent cloning of the virus) and minor bases in the RNA giving rise to incorporation of alternative complementary bases via wobble pairing. These possibilities remain to be investigated by other methods; however, genuine examples of such cases may be infrequent. The deduced EMC virus RNA sequence and the ambiguities we have noted are given in Fig. 3.

We noted two further imperfections in the pattern of bands obtained with EMC virus RNA as a template, although neither of them caused difficulties in reading the sequence. The first of these was the occurrence of a faint ghost band accompanying all the stronger bands in the position one base larger than the major band (Fig. 2*b*). These were obtrusive only when the autoradiographs were overexposed. Two plausible explanations of this are phosphatase activity in the enzyme preparation or the contamination of the pdT_7rC with trace amounts of pdT_8. Another possible explanation would be length heterogeneity in the template immediately adjacent to the poly(A), although this was not observed by Merregaert *et al.* (11). The second feature was the presence of a faint ladder of bands in which every position in every track is occupied, underlying the major pattern in certain portions of the gel (e.g., residues 50–57 in Fig. 2*c*). This was observed only once with EMC virus, after the most prolonged incubation attempted ($3\frac{1}{3}$ hr before the chase was started). For this reason we recommend that the total incubation time not be extended beyond 3 hr. The cause of this phenomenon is not clear, but it may again be related to the presence of secondary structures, perhaps being a more general low-frequency form of the same pausing that generates the single ambiguous bands, a phenomenon that becomes more troublesome as enzyme activity declines on extended incubation.

While EMC virus RNA/pdT_7rC was an excellent template/primer combination for sequence work, poliovirus RNA proved much less forthcoming with the same primer, incorporation being very poor. Perhaps accounting for this large difference in their template efficiency, EMC and poliovirus RNAs differ markedly in the sequences primed by pdT_7rC. Particularly noteworthy is that whereas EMC virus RNA contains the sequence AAUAAA (positions 10–15), which is thought to be commonly present in the poly(A)-adjacent region of avian and mammalian mRNA (20), this hexanucleotide is absent from the sequence adjacent to the pdT_7rC primer of poliovirus RNA (data not shown). Although both we and others (18) have evidence suggesting that the nucleotide immediately adjacent to the poly(A) on polio RNA is a guanine, and hence that pdT_7rC should prime into the poly(A)-adjacent sequence at least as efficiently as at any other site, it remains possible that the sequence we observe results from priming at an internal site. This question requires further investigation.

Seven termination codons are clustered at the extreme 3′ end of the EMC virus RNA sequence (positions 31–33) in all three reading frames, but there are only five more in the next 120 nucleotides, apparently in only two of the three reading frames. It is possible on present evidence that translation continues to within 26 nucleotides of the poly(A), unusually close compared to other sequenced messengers (1, 6, 20). More definite conclusions must await further sequence determinations in RNA and protein.

We thank Prof. R. R. Rueckert for many fruitful discussions and Dr. G. G. Altman for searching the sequences for base pairing using a

computer program supplied by Prof. W. Fitch. This research was supported by Grant AI-01466 and Research Career Award AI-21942 from the National Institute of Allergy and Infectious Diseases and Grant CA-15613 from the National Cancer Institute.

1. Brownlee, G. G. & Cartwright, E. M. (1977) *J. Mol. Biol.* **114,** 93–118.
2. Donis-Keller, H., Maxam, A. M. & Gilbert, W. (1977) *Nucleic Acids Res.* **4,** 2527–2538.
3. Simoncits, A., Brownlee, G. G., Brown, R. S., Rubin, J. R. & Guilley, H. (1977) *Nature (London)* **219,** 833–836.
4. Haseltine, W. A., Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 989–993.
5. Seeburg, P. H., Shine, J., Martial, J. A., Ullrich, A., Baxter, J. D. & Goodman, H. M. (1977) *Cell* **12,** 157–165.
6. Efstratiadis, A., Kafatos, F. C. & Maniatis, T. (1977) *Cell* **10,** 571–585.
7. Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* **94,** 441–448.
8. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
9. Smoler, D., Molineux, I. & Baltimore, D. (1971) *J. Biol. Chem.* **246,** 7697–7700.
10. Taylor, J. M., Garfin, D. E., Levinson, W. E., Bishop, J. M. & Goodman, H. M. (1974) *Biochemistry* **13,** 3159–3163.
11. Merregaert, J., van Emmelo, J., Devos, R., Porter, A., Fellner, P. & Fiers, W. (1978) *Eur. J. Biochem.* **82,** 55–63.
12. Bock, R. M. (1967) *Methods Enzymol.* **12A,** pp. 218–221.
13. Barrell, B. G. (1971) in *Procedures in Nucleic Acid Research,* eds. Cantoni, C. L. & Davies, D. R. (Harper and Row, New York), Vol. 2, pp. 751–779.
14. Hall, L. & Rueckert, R. R. (1971) *Virology* **43,** 152–165.
15. Burness, A. T. H. (1970) *J. Gen. Virol.* **6,** 373–380.
16. Peacock, A. C. & Dingman, C. W. (1968) *Biochemistry* **7,** 668–674.
17. Schwartz, D. E., Zamecnik, P. C. & Weith, H. L. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 994–998.
18. Yogo, Y. & Wimmer, E. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1877–1882.
19. Proudfoot, N. J. & Brownlee, G. G. (1974) *FEBS Lett.* **38,** 179–183.
20. Proudfoot, N. J. (1976) *J. Mol. Biol.* **107,** 491–525.