*Proc. Natl. Acad. Sci. USA*
Vol. 76, No. 3, pp. 1114–1117, March 1979
Biochemistry

# 3′-Terminal nucleotide sequence of alfalfa mosaic virus RNA 4

**(RNA sequence/plant viruses/replication)**

ELLEN C. KOPER-ZWARTHOFF AND JOHN F. BOL

Department of Biochemistry, State University, Wassenaarseweg 64, 2300 Ra Leiden, The Netherlands

*Communicated by Louis N. M. Duysens, December 8, 1978*

**ABSTRACT** **The sequence of the 3′-terminal 91 nucleotides of alfalfa mosaic virus RNA 4, the messenger for the viral coat protein, has been elucidated. A fragment containing the 3′ terminus of the RNA was obtained by mild digestion with RNase T1. The primary structure of the fragment was deduced by labeling it *in vitro* at its 5′ terminus and application of RNA sequencing techniques. The sequence is completely extracistronic and is believed to contain the binding sites for the viral coat protein and replicase.**

Knowledge of the primary structure of strategic sites in nucleic acids is an important aid in understanding the biological processes in which they participate. Recently, we elucidated (1) the sequence of 74 nucleotides at the 5′ terminus of alfalfa mosaic virus (AlMV) RNA 4, which is a monocistronic messenger for the viral coat protein. The sequence included the complete 5′ noncoding region of 39 nucleotides and the first part of the coat protein cistron. In this paper we report the sequence of the 3′-terminal part of AlMV RNA 4 that is supposed to contain the binding sites for the viral coat protein and replicase.

The genome of AlMV is tripartite, like that of bromo- and cucumoviruses (represented by brome mosaic virus and cucumber mosaic virus, respectively). In contrast to these viruses, however, a mixture of the AlMV genomic RNAs (RNAs 1, 2, and 3) is not infectious as such; addition of a small amount of coat protein is required to initiate the infection (2). The coat protein cistron is located in RNA 3, the smallest genome fragment (3). This cistron is not open to translation *in vitro*, but virus preparations contain a subgenomic RNA species (RNA 4) that is efficiently translated into coat protein in various cell-free systems (4). AlMV RNA 4 is approximately 800 nucleotides long (5) and has no poly(A) tail (6); the coat protein cistron contains 660 nucleotides, according to the known amino acid sequence of the coat protein (7). The four RNAs of BMV have an almost identical sequence of about 160 nucleotides at their 3′ termini, in accordance with the fact that they all can be aminoacylated with tyrosine by the appropriate enzymes (8). Evidence for a tRNA-like structure at the 3′ terminus of the genomic RNA has also been reported for several other plant viruses (9–12). However, attempts to aminoacylate the AlMV RNAs have been unsuccessful so far, suggesting that the 3′ termini of the RNAs of this particular virus have no tRNA-like character (E. M. J. Jaspars, personal communication; T. C. Hall, personal communication).

The AlMV RNAs contain a small number of sites that exhibit a high affinity for the viral coat protein, and it has been assumed that binding of coat protein to these specific sites is inherent to its function in the infection process (13). Mild digestion of RNA 4 with RNase T1 generates two fragments of approximately 90 nucleotides that contain the coat protein binding sites (14). These fragments, which were designated 29B and 29C, have a 3′-terminal $C_{OH}$-residue as has the intact RNA 4, indicating that these fragments represent the 3′-terminal part of RNA 4. Because the sequences of the subgenomic RNA 4 are probably located at the 3′-terminal region of RNA 3 (14), the specific binding of coat protein is likely to occur close to the 3′ terminus of both RNA 3 and RNA 4. It has been postulated that binding of coat protein to the 3′ terminus of the AlMV RNAs is required for a proper recognition of the RNAs by the viral replicase (14).

In the present paper we report the complete nucleotide sequence of fragments 29B and 29C. The fragments were found to contain 86 and 91 nucleotides, respectively, and the sequence of 29B is completely present in 29C. In agreement with the earlier conclusion (14), they were found to be derived from the 3′ noncoding region of AlMV RNA 4.

## MATERIALS AND METHODS

Polynucleotide kinase T4 was from Boehringer Mannheim. [$\gamma$-$^{32}$P]ATP was obtained from the Radiochemical Centre (Amersham). RNases T1 and U2 were from Sankyo (via Calbiochem). Ultrapure urea was from Schwarz/Mann. Rexyn I-300 ion exchanger was from Fisher; pure acrylamide was from Serva (Heidelberg). Nuclease P1 was obtained from Yamasa Shoyu Co. Ltd. (Tokyo).

Purification of AlMV RNA 4 (strain 425) was as described (15). Partial digestion of AlMV RNA 4 (4.5 mg) was done according to the procedure described by Houwing and Jaspars (14). The digested material was then centrifuged for 40 hr at 32,000 rpm (Beckman BXV Ti zonal rotor) and 4°C in a linear 10–30% sucrose gradient in 0.01 M sodium phosphate, pH 7.0/1 mM EDTA/0.05 M NaCl. Material with a sedimentation rate of about 4 S was concentrated by ethanol precipitation and subsequently electrophoresed on a 13% preparative acrylamide gel (14). Fractions containing fragments 29B and 29C were dialyzed and lyophilized. The isolation yielded approximately 11 $\mu$g of each fragment, which was sufficient for the sequence analysis. 5′-End group labeling was as described (1). The 5′-$^{32}$P-labeled fragments were purified by electrophoresis on a 20 × 40 cm polyacrylamide (20%) slab gel in 7 M urea for 24 hr at 1000 V and room temperature. Recovery of the labeled fragments was by the method of De Wachter and Fiers (16). Sequencing techniques were essentially as described (1, 17). It appeared necessary to stir the urea solutions for a few minutes with an ion exchanger (Rexyn I-300) before use to avoid streaking, especially upon gel electrophoresis at pH 3.5.

## RESULTS

**Labeling of Fragments 29B and 29C *In Vitro* and Sequence Analysis by Partial Digestion with Nuclease P1.** Fig. 1 shows the results of *in vitro* labeling of fragments 29B (track 1) and 29C (track 2) by using polynucleotide kinase T4 and [$\gamma$-$^{32}$P]ATP. The 29B preparation contained one major fragment, but the 29C preparation contained two fragments, one

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.

Abbreviation: AlMV, Alfalfa mosaic virus.

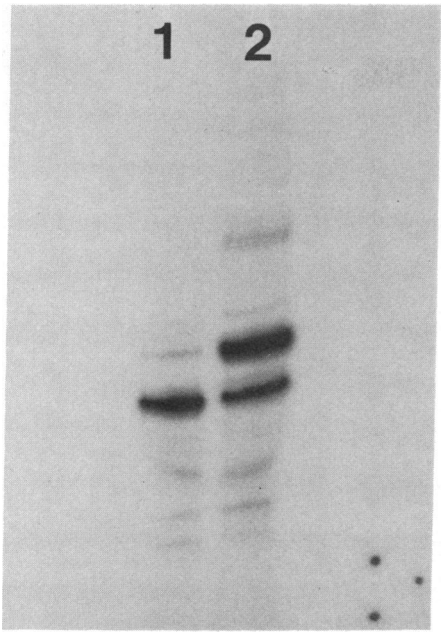


FIG. 1. Autoradiogram of 5′-$^{32}$P-labeled fragments run in a 7 M urea/20% polyacrylamide gel. Tracks: 1, 0.35 μg of fragment 29B labeled with 230 μCi of [γ-$^{32}$P]ATP; 2, 0.87 μg of fragment 29C labeled as for 29B.

of which comigrated with 29B. The major radioactive bands were excised and eluted from the gel. Partial digestion of the 5′-labeled fragments with nuclease P1 followed by electrophoresis of the digest on cellulose acetate and homochromatography yielded the sequence of approximately 20 nucleotides of each fragment. It appeared that the lower band of 29C was identical to 29B. Fig. 2 shows the autoradiograms of this experiment, together with the deduced sequence. The nature of the 5′-nucleotide of both fragments was confirmed by complete digestion with nuclease P1 and one-dimensional chromatography (results not shown). The sequences derived from Fig. 2 are:

```
          1         5         10        15        20
29B: A-U-A-G-G-A-U-C-G-A-C-U-U-C-A-U-A-U-U-G

          1         5         10        15
29C: C-U-C-U-G-A-U-A-G-G-A-U-C-G-A-C-U-U.
```

The sequence of 29C from position 6 on is the same as that of 29B from position 1 to position 13. Because the 3′ termini of both fragments (and of the intact RNA 4) are identical (14), the only difference between the two fragments is the additional sequence C-U-C-U-G at the 5′ terminus of 29C. To avoid confusion, the nucleotide numbering of 29C is used throughout this work—i.e., the first 5′-residue of 29B is designated as nucleotide 6 in subsequent experiments.

**Further Analysis of the Sequence by RNA Gel Sequencing Techniques.** The relative positions of guanine, adenine, and, by difference, pyrimidine residues were determined on one-dimensional sequencing gels as described (1, 17). Partial digestion of the 5′-labeled fragments was done with RNase T1, RNase U2, and alkali, and the digests were electrophoresed on a 7 M urea/20% polyacrylamide slab gel at pH 8.2. This separates the oligonucleotides according to size. This technique allowed us to establish the position of all guanine, adenine, and

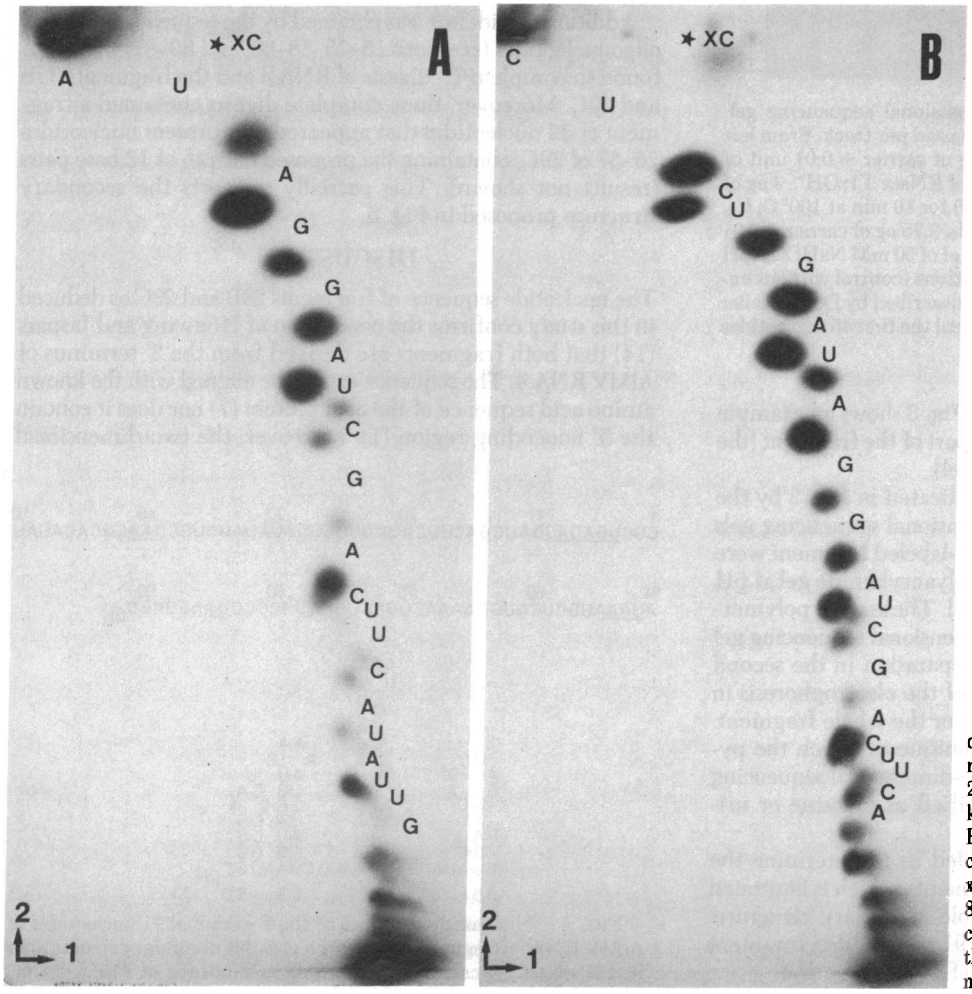


FIG. 2. Sequence analysis by two-dimensional electrophoresis/homochromatography of fragments 29B (*A*) and 29C (*B*). Approximately 250,000 Cerenkov cpm of the fragments was used. Electrophoresis in the first dimension on cellulose acetate was at 5 kV until the xylene cyanole FF dye (XC) had migrated 8 cm. The second dimension was homochromatography on DEAE-cellulose thin-layer plates with 15 mM "homomix."

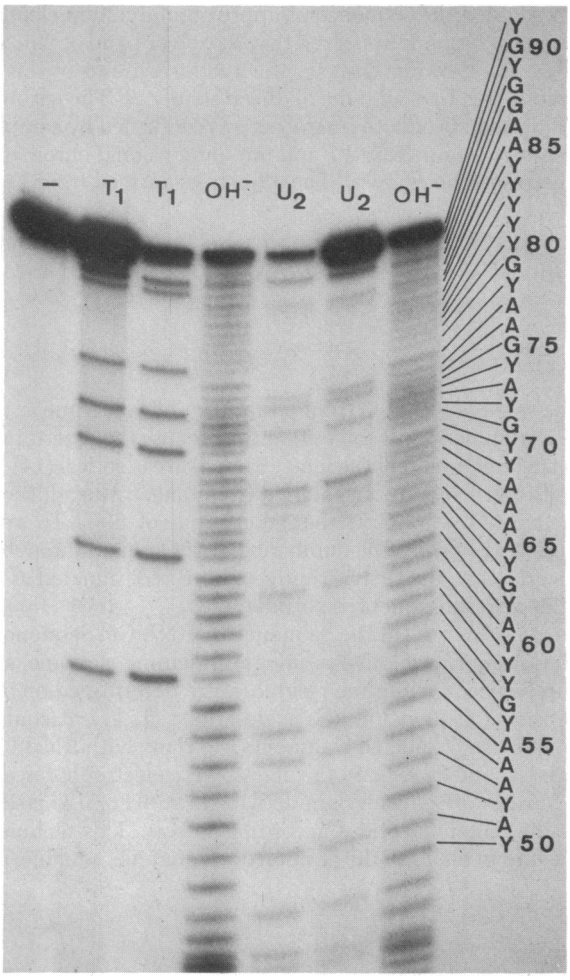


FIG. 3. Autoradiogram of a one-dimensional sequencing gel. Approximately 100,000 Cerenkov cpm was used per track. From left to right: —, no enzyme added; $T_1$, 3.75 μg of carrier + 0.01 unit of RNase T1; $T_1$, 3.75 μg of carrier + 0.1 unit of RNase T1; $OH^-$, 3 μg of carrier in 20 μl of 50 mM $NaHCO_3$ (pH 9.0) for 10 min at 100°C; $U_2$, 3.75 μg of carrier + 0.1 unit of RNase U2; $U_2$, 3.75 μg of carrier + 0.01 unit of RNase U2; $OH^-$, 3 μg of carrier in 20 μl of 50 mM $NaHCO_3$ (pH 9.0) for 7 min at 100°C. All enzyme incubations (control without enzyme included) were at 50°C for 15 min, as described by Donis-Keller *et al.* (17). Electrophoresis was at 1000 V until the first 45 nucleotides were run off the gel. Y, pyrimidines.

pyrimidine residues in the fragments. Fig. 3 shows an example of such a sequence analysis on the last part of the fragment (the first 45 nucleotides were run off the gel).

The identity of the pyrimidines, indicated in Fig. 3 by the letter Y, was established by two-dimensional sequencing gels (1, 18). Partial alkaline digests of the 5′-labeled fragment were electrophoresed in a 7 M urea/10% polyacrylamide gel at pH 3.5, and 16-cm-long strips were excised. These were polymerized into a gel, identical to the one-dimensional sequencing gel described above, that was used for separation in the second dimension. By varying the duration of the electrophoresis in both dimensions it was possible to cover the whole fragment. Fig. 4 shows an example of this technique in which the pyrimidines, already localized by the one-dimensional sequencing gels, could be unambiguously identified as cytosine or uridine.

These combined techniques enabled us to determine the whole nucleotide sequence of the fragments, which is illustrated in Fig. 5, together with the possible secondary structure. Fragment 29C was found to contain 91 nucleotides; fragment 29B is identical to nucleotides 6–91 of 29C.

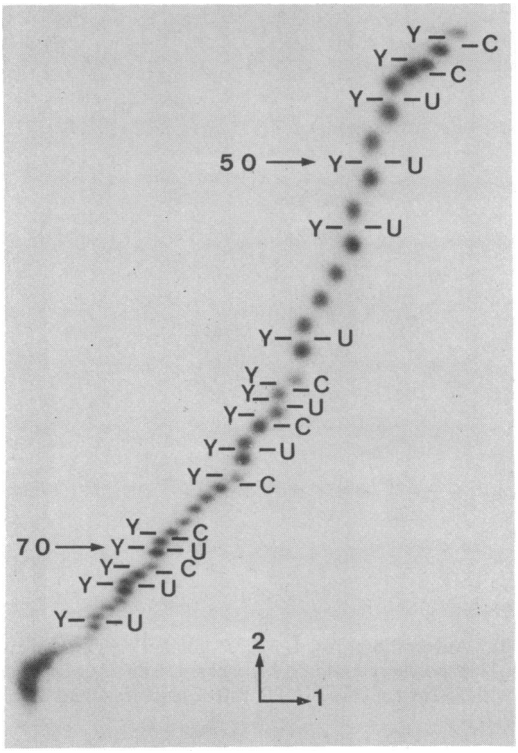


FIG. 4. Autoradiogram of a two-dimensional sequencing gel. Approximately 300,000 Cerenkov cpm of labeled fragment (3 μg) was incubated for 10 min at 100°C in 50 mM $NaHCO_3$ (pH 9.0). Electrophoresis was as described by Lockard *et al.* (18).

Additional evidence was obtained by the sequence of several oligonucleotides (positions 15–25, 58–63, and 80–87) that were found in complete T1 digests of RNA 4 and the fragments 29B and 29C. Moreover, these complete digests contained a fragment of 32 nucleotides that appeared to represent nucleotides 26–57 of 29C, containing the proposed hairpin of 12 base pairs (results not shown). This partially supports the secondary structure proposed in Fig. 5.

## DISCUSSION

The nucleotide sequence of fragments 29B and 29C as deduced in this study confirms the conclusion of Houwing and Jaspars (14) that both fragments are derived from the 3′ terminus of AlMV RNA 4. The sequence cannot be aligned with the known amino acid sequence of the coat protein (7) nor does it contain the 5′ noncoding region (1). Moreover, the two-dimensional

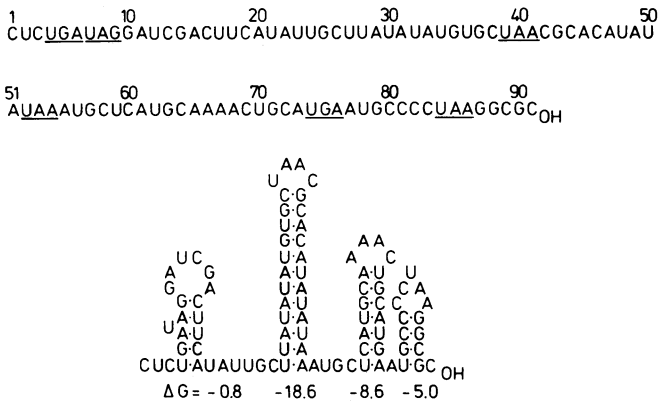


FIG. 5. Nucleotide sequence of the 3′-terminal 91 nucleotides of AlMV RNA 4 (fragment 29C) and a possible secondary structure of it. Δ*G* values are calculated according to Tinoco *et al.* (19).

sequencing gels showed that the 3′-terminal residue is not a guanine, as would be expected for an internal fragment generated by RNase T1 digestion. Although the two-dimensional gel technique enabled us to conclude that the 3′-terminal residue is not a guanine or uridine, it did not permit us to distinguish between adenosine or cytosine in this position. However, it was shown by Houwing and Jaspars (14) that both fragments terminate with a $C_{OH}$-residue as does the intact RNA 4.

From molecular weight determinations by hydrodynamic methods, AlMV RNA 4 was estimated to contain about 800 nucleotides (5). Because the 5′ noncoding region and the coat protein cistron comprise 699 nucleotides, this would suggest that the 3′-terminal nucleotide of the coat protein cistron is separated by about 10 nucleotides from the 5′-terminal nucleotide of the sequence represented by fragment 29C. However, sequence information obtained from other fragments derived from AlMV RNA 4 (unpublished results) indicate that the distance between the coat protein cistron and fragment 29C is considerably longer than 10 nucleotides. Thus, the reported number of nucleotides, 800, in AlMV-RNA 4 (5) must be an underestimate.

The presence of a 3′-terminal poly(A) segment in the genomic RNA of picornaviruses, togaviruses, and a few plant viruses is well documented (20–22). On the other hand, a number of plant viruses have a tRNA-like structure at their 3′-terminus (8–12). AlMV seems to contain neither of these two types of structures. The finding that AlMV RNA 4 does not terminate with -C-C-A may explain the lack of success of attempts to aminoacylate the AlMV RNAs.

Because it is likely that AlMV RNA 4 is homologous to the 3′-terminal part of AlMV RNA 3 (14), the sequence of fragment 29C probably also reflects the sequence of the 3′-terminal 91 nucleotides of RNA 3. This region of the genomic RNA is assumed to contain a recognition site for the viral replicase. It was postulated by Houwing and Jaspars (14) that binding of AlMV coat protein to a specific site in the sequence represented by fragment 29B is required for a proper recognition of the 3′ terminus by the viral replicase. These authors also noticed that fragment 29B had a higher affinity for coat protein than does fragment 29C. Because the primary sequence of 29B is completely present in 29C, this difference in affinity is probably the result of a difference in secondary structure. Fragment 29C may contain a weak loop structure at its 5′ terminus with $\Delta G = -0.8$ (Fig. 5). Such a loop will be absent in 29B because nucleotides 1–5 are missing in this fragment. This suggests that coat protein binds preferentially to a single-stranded region near the 5′ terminus of 29B. In this respect it should be noticed that the binding studies were done at 0°C, whereas $\Delta G$ is calculated for 25°C.

Recently it was shown by melting studies and fluorescence measurements (S. Srinivasan and E. M. J. Jaspars, personal communication) that addition of coat protein to 29B or 29C does not affect the secondary structure of the fragment. Although it is possible, for instance, that the protein binds to the most stable hairpin loop ($\Delta G = -18.6$) without changing its base pairing, such binding would not account for the differential affinity for protein of 29B and 29C. Thus, this result supports the view that coat protein binds to a single-stranded region of fragment 29B.

This work was sponsored by the Netherlands Foundation for Chemical Research (S. O. N.) with financial aid from the Netherlands Organization for the Advancement of Pure Research (Z. W. O.).

1. Koper-Zwarthoff, E. C., Lockard, R. E., Alzner-deWeerd, B., RajBhandary, U. L. & Bol, J. F. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5504–5508.
2. Bol, J. F., Van Vloten-Doting, L. & Jaspars, E. M. J. (1971) *Virology* **46,** 78–85.
3. Jaspars, E. M. J. (1974) *Adv. Virus Res.* **19,** 37–149.
4. Rutgers, A. S. (1977) Dissertation (Univ. of Leiden, Leiden, Netherlands).
5. Heijtink, R. A., Houwing, C. & Jaspars, E. M. J. (1977) *Biochemistry* **16,** 4684–4693.
6. Bol, J. F., Bakhuizen, E. G. C. & Rutgers, T. (1976) *Virology* **75,** 1–17.
7. Van Beynum, G. M. A., De Graaf, J. M., Castel, A., Kraal, B. & Bosch, L. (1977) *Eur. J. Biochem.* **72,** 63–78.
8. Dasgupta, R. & Kaesberg, P. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4900–4904.
9. Guilley, H., Jouard, G. & Hirth, L. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 864–868.
10. Briand, J. P., Jouard, G., Guilley, H., Richards, K. & Hirth, L. (1977) *Eur. J. Biochem.* **72,** 453–463.
11. Silberklang, M., Prochiantz, A., Haenni, A.-L. & RajBhandary, U. L. (1977) *Eur. J. Biochem.* **72,** 465–478.
12. Briand, J. P., Richards, K. E., Bouley, J. P., Witz, J. & Hirth, L. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 737–741.
13. Van Vloten-Doting, L. & Jaspars, E. M. J. (1972) *Virology* **48,** 699–708.
14. Houwing, C. J. & Jaspars, E. M. J. (1978) *Biochemistry* **17,** 2927–2933.
15. Bol, J. F. & Van Vloten-Doting, L. (1973) *Virology* **51,** 102–108.
16. De Wachter, R. & Fiers, W. (1972) *Anal. Biochem.* **49,** 184–197.
17. Donis-Keller, H., Maxam, A. M. & Gilbert, W. (1977) *Nucleic Acids Res.* **4,** 2527–2538.
18. Lockard, R. E., Alzner-deWeerd, B., Heckman, J., MacGee, J., Tabor, M. W. & RajBhandary, U. L. (1978) *Nucleic Acids Res.* **5,** 37–56.
19. Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature (London) New Biol.* **246,** 40–41.
20. Cooper, P. D. (1977) in *Comprehensive Virology,* eds. Fraenkel-Conrat, H. & Wagner, R. R. (Plenum, New York), Vol. 9, pp. 133–207.
21. Kennedy, S. I. T. (1976) *J. Mol. Biol.* **108,** 491–511.
22. El Manna, M. M. & Bruening, G. (1973) *Virology* **56,** 198–206.