Volume 6 Number 10 1979 Nucleic Acids Research

Sequence of 200 nucleotides at the 3'-terminus of the genome RNA of vesicular stomatitis virus

Duncan J.McGeoch and Aidan Dolan

MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR, UK

Received 30 May 1979

ABSTRACT

The sequence of 200 nucleotides at the 3'-terminus of the genome RNA of vesicular stomatitis virus, Indiana serotype, was determined by adding a poly(A) tract to the 3'-terminus of genome RNA, then using the poly(A) as a binding site for a primer to initiate reverse transcription of the RNA, and analysing the complementary DNA sequence by the dideoxynucleoside triphosphate chain termination method. Proceeding 3' to 5', the genome RNA sequence consisted of a sequence complementary to the leader RNA, followed by the sequence AAA, followed by a sequence complementary to the 5'-extremity of N protein mRNA. These results are discussed in terms of leader RNA function, mechanism of transcript processing at the junction between leader RNA and N mRNA, and N mRNA structure.

INTRODUCTION

The genome of vesicular stomatitis virus (VSV; ref. 1) consists of a single-stranded RNA molecule about 11,000 nucleotides long (2,3). In the infected cell the mRNA species coded by the virus are transcribed from this "negative" sense genome RNA (virion RNA; vRNA) by a virus-specified RNA-dependent RNA polymerase (see refs. 4 and 5 for reviews). Studies on the sensitivity to ultraviolet irradiation of synthesis of individual gene products indicate that the polymerase initiates transcription at the 3'-terminus of vRNA and sequentially synthesizes the transcript RNA species (6,7). The first transcribed RNA is a species, called the leader RNA, which is 47 or 48 nucleotides long and is complementary to the 3'-terminus of vRNA (8-10), and this is thought to be followed by the first messenger species, N protein mRNA (6,7). Unlike the mRNA species, the leader RNA is neither capped nor polyadenylated (8,9). Thus, the 3'-

© Information Retrieval Limited 1 Falconberg Court London W1V 5FG England

terminal region of the vRNA must contain a polymerase initiation site, sequences complementary to the leader RNA, sequences representing the junction between the leader RNA and N mRNA, and sequences complementary to the 5'-terminus of N mRNA. The junction between the transcripts is the site of processing events during transcription:- termination of leader RNA synthesis and the start of N mRNA synthesis, plus capping and methylation of N mRNA (5,11).

This paper describes the determination of the sequence of 200 nucleotides at the 3'-terminus of vRNA, and interprets the events of transcription in terms of the sequence. The RNA sequence was determined by adding a poly(A) tract to the 3'-terminus of vRNA, and using the poly(A) to initiate reverse transcription of the vRNA with an oligo (dT) primer; the resulting complementary DNA was sequenced by the dideoxynucleoside triphosphate (ddNTP) chain termination method (12,13). Recognition and interpretation of the sequence obtained was facilitated by published sequencing studies on the 3'-terminal 17 nucleotides of vRNA (14-16), on the leader RNA (10), on the 5'-terminal 31 nucleotides of N mRNA (17) and on the 5'-terminus of vRNA (18).

MATERIALS AND METHODS

1. Materials

dNTPs and ddNTPs, $p(dT)_8$-dA and $p(dT)_{11}$-rA were purchased from PL Biochemicals Inc. $p(dT)_8$-dA-dC was purchased from Collaborative Research Inc. Radiochemicals were from the Radiochemical Centre, Amersham.

AMV reverse transcriptase was a gift of Dr. J. W. Beard. Rat liver RNase inhibitor was a gift of G. D. Searle Company. Poly(A) polymerase was purified from *Escherichia coli* B/r by the method of Sippel (19). Peak activity fractions from the phosphocellulose column were dialysed against 700 mM NaCl, 25 mM Tris-HCl pH 7.9, 1 mM EDTA, 50% glycerol and stored at -20°.

2. Addition of a poly(A) tract to VSV vRNA

vRNA of VSV Indiana was prepared as before (11). A reaction mix was set up containing, in a total volume of 250 µl, 50 µg vRNA, 50 mM Tris-HCl pH 7.9, 10 mM $MgCl_2$, 2.5 mM $MnCl_2$,

0.1 mM ATP, 200 μg/ml bovine serum albumin, 200 units of poly(A) polymerase and NaCl to a final concentration of 250 mM, including the contribution from the enzyme stock. The albumin was previously heated for 10 min at 90-100° at 10 mg/ml. The mixture was incubated for 5 min at 25°, and the RNA then recovered by extraction with phenol and ethanol precipitation. The ethanol precipitate was dissolved in 100 μl of 100 mM sodium acetate, and the RNA was then precipitated with 50 μl 1% cetyltrimethylammonium bromide (to remove ATP). The precipitate was washed twice with cold 70% ethanol containing 100 mM sodium acetate and once with cold 70% ethanol, then dried. The RNA was finally dissolved in 50 μl water and stored at -20°.

3. RNA sequence determination

Sequence determination was performed with reverse transcriptase and ddNTPs, with 1 μg of poly(A)-containing vRNA in each 5 μl reverse transcription reaction, as previously described (13), but with the following changes. For determination of the first 15 to 20 residues adjacent to the 3'-terminus, reverse transcripts were labelled using (α-^{32}P)-dCTP at 2 μM and 300 Ci/mmol, with the other dNTPs at 40 μM and ddNTPs at 10X the corresponding dNTP concentration. Incubations were terminated by addition of formamide and the products fractionated by electrophoresis in a 16% acrylamide gel (13). For determination of sequences further from the terminus, reactions contained (α-^{32}P)-dATP at 20 μM and 100 Ci/mmol, with other dNTPs at 40 μM and ddNTPs at 0.5X or 0.25X the corresponding dNTP concentration. After incubation, reaction volumes were made up to 100 μl, containing 5 μg tRNA and 5 mM EDTA, and the nucleic acids precipitated by addition of 7 μl of 15% perchloric acid. The precipitates were washed, as described above for cetyltrimethylammonium bromide precipitation, and dissolved in 10 μl formamide containing 0.025% each of xylene cyanol and bromophenol blue. This procedure conveniently removed unincorporated (α-^{32}P)-dATP, which otherwise gave a troublesome uniform background on gel electrophoresis. Aliquots were then fractionated in acrylamide slab gels as before (11,13).

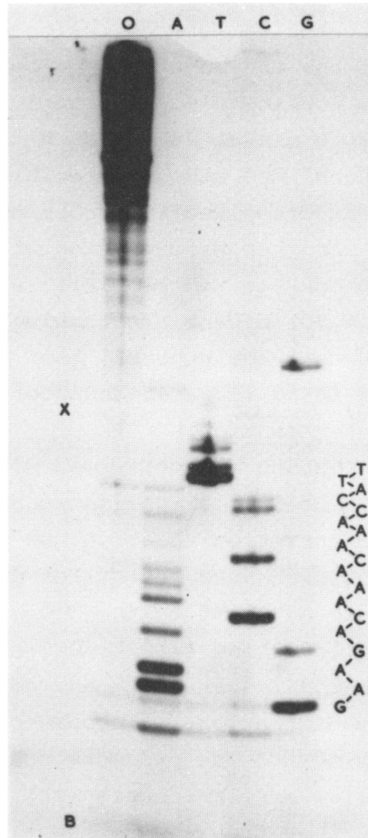

Figure 1 Sequence adjacent to 3'-terminus of vRNA.
Poly(A)-vRNA was reverse transcribed in the presence, one at a time, of high concentrations of each ddNTP, and the products fractionated on a 16% acrylamide gel, then detected by autoradiography. The ddNTP species added is shown at the head of each track. "O" indicates a reaction lacking ddNTP. X and B indicate the positions of dye markers.

RESULTS

We wished to determine nucleotide sequences by reverse transcribing the RNA from a defined starting position in the presence, one at a time, of partially inhibiting concentrations of each of the four ddNTPs. In order to generate a specific start site at the 3'-terminus of vRNA, a poly(A) tract was added

to the 3'-terminus, as described by Devos et al. for phage MS2 RNA (20). We found that lowering the incubation temperature for poly(A) addition from 37° to 25° and including bovine serum albumin to 200 µg/ml gave a greatly improved poly(A)-containing vRNA product, as judged by template activity for reverse transcriptive sequencing. No characterization of the size of the poly(A)-vRNA was performed for the material used in these experiments. Since the nucleotide sequence of the 3'-terminus of vRNA is known (14-16), sequence results should demonstrate directly whether poly(A) has been added specifically to the authentic 3'-terminus of vRNA. The 3'-terminal sequence of VSV vRNA is 5'-G-U (14,15) and so compounds of the type $p(dT)_n$-dA or $p(dT)_n$-dA-dC should constitute suitable "phasing" primers for reverse transcriptive sequencing (13,21). In early experiments $p(dT)_8$-dA, $p(dT)_8$-dA-dC and $p(dT)_{11}$-rA preparations were tested. All gave sequence specific reverse transcription with poly(A)-vRNA, but not with untreated vRNA. The cleanest results were obtained with $p(dT)_{11}$-rA, and this was used, without further purification, in the experiments described here.

The sequence determined is numbered 3' to 5' with the 3'-terminal U as number 1 (14,15). The standard protocol does not give clear results for residues 1 to around 12, due to low incorporation of label and loss of short DNA species during acid precipitation. This region was therefore determined by using high ddNTP levels, labelling with (α-^{32}P)-dCTP, and omitting acid precipitation. Difficulties and artefacts encountered in determining sequence adjacent to poly(A) were discussed previously (13). The results obtained are shown in Figure 1. The first ddNTP-specific band is in the ddGTP track at a position corresponding to $p(dT)_{11}$-rA plus two additional residues. Thus, the starting DNA sequence is 5'-ACG. This, and the further sequence, is listed in Figure 3.

The sequence from residue 12 to residue 175 was obtained with the standard protocol, with various times of gel electrophoresis in 6%, 8% and 10% acrylamide gels (11,13). Each part of the sequence was obtained from overlapping results in at least two experiments. The results obtained are illustrated in Figure 2, and the deduced complementary DNA and vRNA sequences

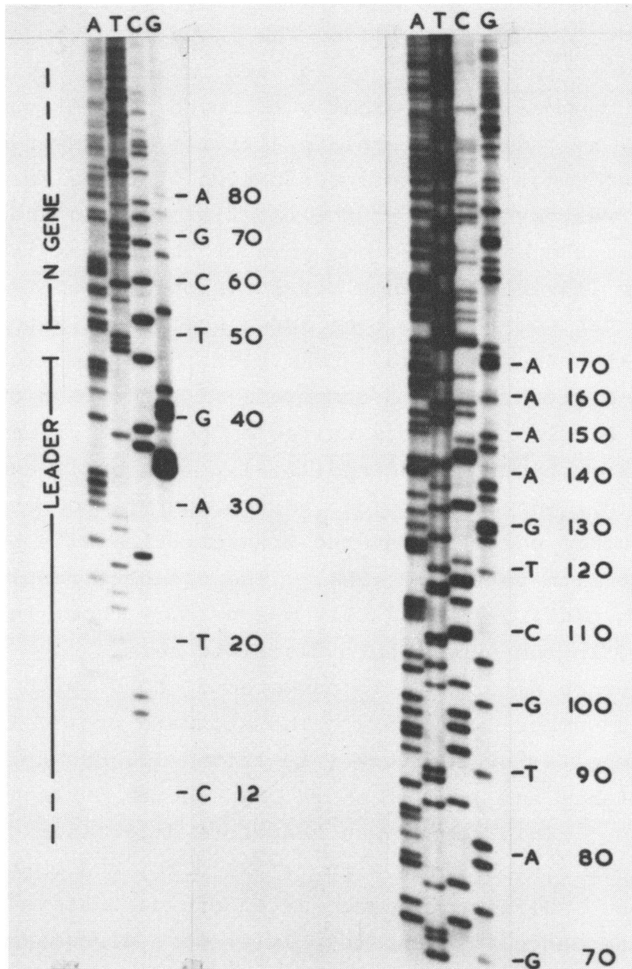

Figure 2 Autoradiographs of sequencing gels.
Reaction mixtures processed according to the standard protocol were fractionated on acrylamide gels. Left panel: portion of a 10% gel, electrophoresis time 4h at 1300-1600 V. Regions of sequence corresponding to leader RNA and N mRNA are indicated. Right panel: portion of a 8% gel, electrophoresis time 5h at 1300-1700 V. In both panels the ddNTP species is shown at the top of each track, and the identity and number (Figure 3) of every tenth residue of the DNA is indicated.

are listed in Figure 3.

Beyond residue 175 the sequencing gels exhibited an

```
              1                        25                       50
DNA  5'     ACGAAGACAAACAAACCATTATTATCATTAAAAGGCTCAGGAGAAACTTT

vRNA 3'   HOUGCUUCUGUUUGUUUGGUAAUAAUAGUAAUUUUCCGAGUCCUCUUUGAAA

              51                       75                       100
DNA         AACAGTAATCAAAATGTCTGTTACAGTCAAGAGAATCATTGACAACACAG

vRNA        UUGUCAUUAGUUUUACAGACAAUGUCAGUUCUCUUAGUAACUGUUGUGUC

              101                      125                      150
DNA         TCATAGTTCCAAAACTTCCTGCAAATGAGGATCCAGTGGAATACCCGGCA

vRNA        AGUAUCAAGGUUUUGAAGGACGUUUACUCCUAGGUCACCUUAUGGGCCGU

              151                      175                      200
DNA         GATTACTTCAGAAAATCAAAGGAGAT-CTCTTTACATCAATACTACAAAA---3'

vRNA        CUAAUGAAGUCUUUUAGUUUCCUCUA-GAGAAAUGUAGUUAUGAUGUUUU---5'
```

Figure 3 Sequence of DNA and vRNA
The DNA sequence obtained, and the deduced vRNA sequence, are listed. Beyond residue 175, there is an uncertainty, as explained in the text, and the subsequent numbering is an estimate lacking certainty.

ambiguity: a strong band was present in all tracks, and also occurred in a sample incubated without any ddNTP. Thus, an unambiguous sequence could not be determined in this region. Further clear sequence beyond this point was obtained to residue 200, but this could not be assigned a definitive continuity with the sequence of residues 1-175. Additional sequence data, with several ambiguities, were obtained to about residue 240, but are not shown here.

DISCUSSION

1. Assessment of sequencing results

Residues 1-17 of the vRNA sequence listed in Figure 3 are

identical to the 3'-terminal sequence determined directly with vRNA (16), confirming that the sequencing system is functioning as intended. Further, residues 1-47 are complementary to the leader RNA sequence of Colonno and Banerjee (10). However, they considered the 3'-terminus of the leader to be either 5'-GAAAAC or 5'-GAAAC. Our data indicate the latter and so we consider the leader as being 47 nucleotides in length. The leader complementary sequence is followed in the vRNA strand by AAA and this is followed by 31 nucleotides which are exactly complementary to the 5'-terminal sequence of N mRNA determined by Rose (17). The ambiguity near residue 175 is presumably a result of the reverse transcriptase pausing or terminating at a region of unusual template structure. This type of phenomenon constitutes a general limitation in sequencing systems which combine the operations of synthesis and of generation of base specific fragments of the nucleic acid under study (for instance, see refs. 22 and 23).

2. The leader RNA: N mRNA junction

As noted above, the sequences complementary to leader RNA and N mRNA are separated only by three A residues in the vRNA strand. It is not known whether transcription of this region to yield the leader RNA and the 5'-terminus of N mRNA is accomplished by terminating RNA polymerization and then restarting at the N mRNA sequence, or by transcribing through the region then generating the final transcripts by cleavage of the precursor chain. The sequence UUU, complementary to AAA (residues 48-50) in the vRNA, is not represented in the mature transcripts, so is either not synthesized or is subsequently removed. Interestingly, the leader RNA is found in two forms, one of which lacks the terminal C residue (10), so the operation of the processing mechanism may not be exact.

We have previously determined the nucleotide sequence, at the other end of the N gene, of the junction between the N and NS genes (11), and in Figure 4 this is compared with the leader/N junction, in such a way as to maximise similarities between the two sequences. The N/NS junction differs in that during transcription of this region, the virus transcriptase adds a 3'-terminal poly(A) tract to the newly synthesized N mRNA sequence,

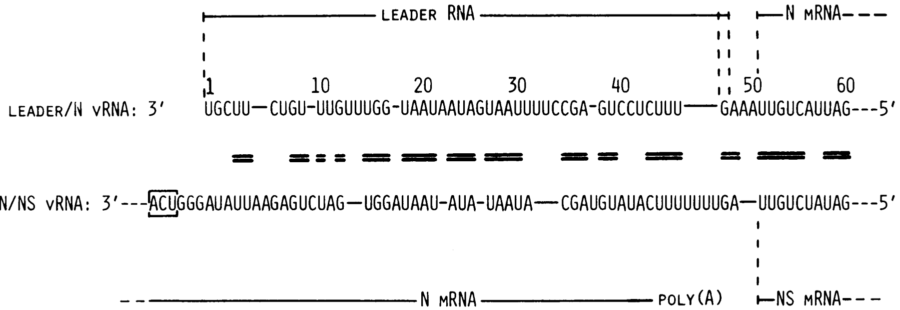

Figure 4 Comparison of the leader/N and N/NS junctions
The vRNA sequence, residues 1-60 of Figure 4, is compared with the vRNA sequence around the N/NS junction (11). The two sequences are spaced to maximise apparent homology, and homologous regions are marked by double lines between the sequences. The regions represented in the mature transcripts are indicated by lines above and below the sequences. The boxed sequence 3'-ACU-5' represents the position of translation termination in N mRNA (ref. 13; D.J. McGeoch, A. Dolan and C.R. Pringle, in preparation).

and it was postulated that synthesis of this tract was achieved by reiterative copying of a U_7 sequence in the vRNA at the junction (ref. 11, and see Figure 4). The absence of a U_7 tract at the leader/N junction is consistent with the hypothesis. In other ways the sequences are similar, most notably at vRNA residues 15-30: as written in Figure 4, 14 out of 16 residues of the leader complementary sequence are identical to residues complementary to the 3'-terminus of N mRNA. This region contains the sequence 3'-UAAUXA$\overset{*}{\text{U}}$AXUAAU-5' (in the leader/N sequence X is a purine, in the N/NS sequence X represents a gap). This structure is symmetrical around the residue marked with an asterisk; such sequences have been observed previously in the 3'-untranslated regions of eukaryote mRNAs (for instance, ref. 24). Outside this region, the two sequences contain further significant homologous regions corresponding to the 3'-terminal part of the leader RNA and the 5'-termini of N and NS mRNAs. Although the leader/N junction has no U_7 sequence, it does contain a tract of eight pyrimidine residues (numbers 39-46). As written in Figure 4 the 3'-terminus of the leader, 5'-AAAC,

```
 51                           75
GpppAACAGUAAUCAAA AUG UCU GUU ACA GUC AAG AGA AUC AUU GAC

                  Met Ser Val Thr Val Lys Arg Ile Ile Asp


        100                                 125
AAC ACA GUC AUA GUU CCA AAA CUU CCU GCA AAU GAG GAU CCA

Asn Thr Val Ile Val Pro Lys Leu Pro Ala Asn Glu Asp Pro


                    150                               175
GUG GAA UAC CCG GCA GAU UAC UUC AGA AAA UCA AAG GAG A---

Val Glu Tyr Pro Ala Asp Tyr Phe Arg Lys Ser Lys Glu ----
```

Figure 5 5'-terminal sequence of N mRNA
The 31-nucleotide 5'-terminal sequence of N mRNA obtained by Rose (17) is extended, using the vRNA data, to 125 nucleotides. Numbering is as in Figure 3. The deduced amino acid sequence of the amino-terminal region of N polypeptide is listed.

is formally homologous to part of the poly(A) tract of N mRNA. This similarity may suggest that the 3'-terminal tract of N mRNA could also be terminated with a C, at least transiently, although this has not been observed (25). In both junctions the sequences complementary to adjoining mature transcripts are closely contiguous, being separated in the leader/N case by

Table 1 Nucleotide compositions of regions of VSV (+) strand RNA

RNA	length in nucleotides	%A	%U	%G	%C	Source of data
leader	47	49	17	15	19	Sequence (this paper)
5'-region of N mRNA	125	40	22	18	20	Sequence (this paper)
3'-region of N mRNA	205	35	24	22	18	Sequence (ref. 13)
total RNA (+) strand	11,000	34	25	18	22	Digestion of vRNA (ref. 3)

3'-(G)AAA-5' and in the other by 3'-GA-5'. These common features suggest that the two junctions are the site of action of mechanistically similar processing events (with the obvious distinction that only one involves poly(A) synthesis) and that the common sequences contain sites signalling or controlling the processing mechanisms.

3. Possible function of the leader RNA

The above discussion indicates that the region of the leader RNA comprising, approximately, residues 15 through to 47 may contain sites involved in transcript processing and so may be regarded as functionally part of the leader/N junction. The remainder of the leader complementary sequence, residues 1 to 14, has only very limited homology with the corresponding region in the N gene (as written in Figure 4, any similarity is probably over-emphasized). These residues, at the extreme 3'-terminus of vRNA, presumably contain a virus polymerase binding and initiation site. Semler et al (18) have indirectly determined the sequence of 46 residues at the 5'-terminus of vRNA, and this is complementary to our 3'-terminal sequence at 15 out of the 18 terminal residues, but thereafter the sequences diverge. For purposes of replication we expect the 5'-terminus to contain the complement of a polymerase initiation site, so this comparison indicates that an initiation site could comprise up to 18 of the terminal nucleotides. Thus, the leader complementary sequence may be totally accounted for by a site for polymerase initiation (residues 1 to about 18), and sites concerned with processing at the leader/N junction (residues 15 to 47). The use of a leader RNA to start transcription can therefore be viewed as a tactic directed to synthesizing sequences around an intertranscript junction, presumably with a view to obtaining a properly modified (capped and methylated) 5'-terminus to the first messenger RNA species through operation of the transcript processing mechanisms. That is, transcript capping and methylation are seen as part of the transcript processing mechanism rather than of the transcipt initiation mechanism.

4. Sequences at the start of the N gene

The vRNA sequence, residues 51-81, is exactly complementary to the sequence determined by Rose (17) for the ribosome binding

site of N mRNA, which includes the 5'-terminus of the RNA. Thus our present data definitively place the N gene as the gene nearest the 3'-terminus of vRNA. Since gene splicing is considered not to occur with VSV (26), we can use the vRNA sequence to extend the known N mRNA sequence, and this is shown in Figure 5 for vRNA residues 51-175. Rose (17) proposed that the 5'-AUG-3' at residues 64-66 is the polypeptide initiating codon. We now see that the next candidate AUG is at residues 125-127, well outside the ribosome binding site, and so the original assignment is very probably correct. Further, the selected translation reading frame then contains no termination codons to residue 175. The deduced sequence of 37 amino acids is also shown in Figure 5.

One of the first observed features of the leader RNA was that it has a notably high A content (ref. 8 and Table 1). The 5'-region of N mRNA also has a high A content, lower than the leader value but clearly greater than the values for the 3'-region of N mRNA or the total (+) strand RNA. Possibly this is a consequence of a requirement for the 3'-terminal region of vRNA to have a high U content, rather than a reflection of mRNA function.

ACKNOWLEDGEMENTS

We owe thanks to Professor J. H. Subak-Sharpe for support and for a critical review of the manuscript, and to Mrs. N. Turnbull for help in early stages of this work.

REFERENCES

1. Abbreviations: VSV, vesicular stomatitis virus; vRNA, virion RNA; ddNTP, 2',3'-dideoxynucleoside 5'-triphosphate.
2. Huang, A.S. and Wagner, R.R. (1966). J. Mol. Biol. 22, 381-384.
3. Bishop, D.H.L., Aaslestad, H.G., Clark, H.F., Flamand, A., Obijeski, J.F., Repik, P. and Roy, P. (1975) in Negative Strand Viruses, Mahy, B.W.J. and Barry, R.D., Eds., Vol. 1, pp. 259-292. Academic Press, London.
4. Wagner, R.R. (1975) in Comprehensive Virology, Fraenkel-Conrat, H. and Wagner, R.R., Eds., Vol. 4, pp. 1-80.

Plenum Press, New York.
5. Banerjee, A.K., Abraham, G. and Colonno, R.J. (1977). J. Gen. Virol. 34, 1-8.
6. Ball, L.A. and White, C.N. (1976). Proc. Nat. Acad. Sci. USA. 73, 442-446.
7. Abraham, G. and Banerjee, A.K. (1976). Proc. Nat. Acad. Sci. USA. 73, 1504-1508.
8. Colonno, R.J. and Banerjee, A.K. (1976). Cell 8, 197-204.
9. Colonno, R.J. and Banerjee, A.K. (1977). Virology 77, 260-268.
10. Colonno, R.J. and Banerjee, A.K. (1978). Cell 15, 93-101.
11. McGeoch, D.J. (1979). Cell, in press.
12. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 5463-5467.
13. McGeoch, D.J. and Turnbull, N.T. (1978). Nucl. Acids Res. 5, 4007-4024.
14. Banerjee, A.K. and Rhodes, D.P. (1976). Biochem. Biophys. Res. Commun. 68, 1387-1394.
15. Keene, J.D., Rosenberg, M. and Lazzarini, R.A. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 1353-1357.
16. Keene, J.D., Schubert, M., Lazzarini, R.A. and Rosenberg, M. (1978). Proc. Nat. Acad. Sci. U.S.A. 75, 3225-3229.
17. Rose, J.K. (1978). Cell 14, 345-353.
18. Semler, B.L., Perrault, J., Abelson, J. and Holland, J.J. (1978). Proc. Nat. Acad. Sci. U.S.A. 75, 4704-4708.
19. Sippel, A.E. (1973). Eur. J. Biochem. 37, 31-40.
20. Devos, R., Gillis, E. and Fiers, W. (1976). Eur. J. Biochem. 62, 401-410.
21. Cheng, C.C., Brownlee, G.G., Carey, N.H., Doel, M.T., Gillam, S. and Smith, M. (1976). J. Mol. Biol. 107, 527-547.
22. Zimmern, D. and Kaesberg, P. (1978). Proc. Nat. Acad. Sci. U.S.A. 75, 4257-4261.
23. Maat, J. and Smith, A.J.H. (1978). Nucl. Acids Res. 5, 4537-4545.
24. Poon, R., Kan, Y.W. and Boyer, H.W. (1978). Nucl. Acids Res. 5, 4625-4630.
25. Banerjee, A.K., Moyer, S.A. and Rhodes, D.P. (1974). Virology 61, 547-558.
26. Freeman, G.J., Rao, D.D. and Huang, A.S. (1979). Gene 5, 141-157.