SEQUENCE HOMOLOGY AT THE 3′-TERMINI OF THE FOUR RNAs OF ALFALFA MOSAIC VIRUS

Madeline R. GUNN and Robert H. SYMONS
Department of Biochemistry, University of Adelaide, Adelaide, SA 5001, Australia

Received 29 October 1979

1. Introduction

Alfalfa mosaic virus (AMV) contains a coat protein-dependent tripartite genome. For infectivity, the three largest RNAs (RNAs 1–3 in order of decreasing molecular weight) plus either coat protein or the smallest RNA (RNA 4) which codes for coat protein are required [1,2]. By contrast, the coat protein-independent tripartite viruses, e.g., cucumber mosaic virus (CMV) and brome mosaic virus (BMV), required only the three largest of their four RNAs for infectivity [3,4]. A further difference is that CMV and BMV RNAs can be aminoacylated by tyrosine in the presence of plant aminoacyl tRNA synthetases whereas AMV RNAs cannot be aminoacylated by any amino acid [2,5]. This implies a fundamental difference in the 3′-terminal sequences of AMV RNAs as compared to CMV and BMV RNAs.

We have shown by hybridization analysis using labelled complementary DNA (cDNA) that, for both AMV and CMV, RNAs 1, 2 and 3 are unique molecules and that the sequence of RNA 4 is contained at the 3′-end of RNA 3 [6,7]. In addition, the hybridization results for CMV showed an homologous region of ~200 residues at the 3′-termini of the four RNAs whereas RNAs 1, 2 and 3 of AMV showed <50 residues terminal homology.

The sequence of ~270 residues from the 3′-termini of the four CMV RNAs has been determined using the dideoxynucleotide chain termination technique as applied to RNA [8] and showed regions of complete, partial or no homology between the RNAs [9]. Here, the same sequencing approach has been applied to determine the sequence of the 3′-terminal 220 residues of the four AMV RNAs. All RNAs have a common 3′-terminal sequence of 23 residues followed by a region of extensive homology up to residue 142 for RNAs 1, 3 and 4 and to residue 143 for RNA 2; thereafter there is no sequence homology. As compared with the identical sequence of RNAs 3 and 4 in the 3′-homologous region, there are 22 residues (15%) different in RNA 1 and 15 residues (10%) different in RNA 2.

Address correspondence to: Dr R. H. Symons

2. Materials and methods

The four viral RNAs of AMV (Q-strain) were purified as in [10] and polyadenylated at their 3′-termini with *Escherichia coli* ATP:RNA adenyltransferase (poly(A) polymerase) essentially as in [11].

The sequencing approach and procedure used have been described fully in [9] but brief details are as follows. Using polyadenylated RNA as template and d(pT_8G) as primer, cDNA was transcribed using reverse transcriptase in four reactions, each containing a different 2′,3′-dideoxynucleoside 5′-triphosphate (ddNTP). Reaction mixtures contained in 4 μl final vol.; 50 mM Tris–HCl (pH 8.0), 50 mM KCl, 6 mM $MgCl_2$, 10 mM DTT, ~0.1 μg polyadenylated RNA, 0.05–0.1 μg d(pT_8G), each of the four dNTPs at 15–50 μM, the appropriate ddNTP at 8–250 μM, 5 μCi each of d[α-^{32}P]CTP and d[α-^{32}P]GTP (200–350 Ci/mmol [12]), and 2.5 units of reverse transcriptase. After incubation at room temperature for 5 min and 37°C for 30 min, labelled nucleotides in each reaction mixture were separated on the basis of chain length by electrophoresis on thin 6% polyacrylamide gels [9,13]. Autoradiography of gels was carried out at −70°C using preflashed Fuji RX

medical X-ray film and Ilford fast tungstate intensifying screens [14].

3. Results and discussion

3.1. *Sequencing procedure*

The 3′-terminal residue of the AMV RNAs was shown to be C by the method for CMV in [15] so that d(pT_8G) was used as a primer for the reverse transcriptase catalysed synthesis of cDNA on polyadenylated RNA during sequencing by the dideoxynucleotide chain termination technique in [8]. By empirically altering the concentration of each ddNTP in the sequencing reaction mixtures, it was possible to obtain sequences ⩾220 residues from the 3′-terminus of the four RNAs. Routinely, thin sequencing gels nearly 80 cm long were used [9,16] rather than the standard 40 cm thin gels in [13].

3.2. *Sequences adjacent to the 3′-termini of the four AMV RNAs*

Examples of sequencing gels of the 3′-terminal region of RNAs 2 and 4 are shown in fig.1; such gels of all RNAs allowed the unambiguous reading of ⩾50 residues from the 3′-terminus. However, a band corresponding to residue 2 (the first one inserted after the d(pT_8G) primer) was always missing (fig.1, RNA 4) even when d[5′-^{32}P](pT_8G) was added to the reaction mixture (fig.1, RNA 2). By contrast, it was necessary to use d[5′-^{32}P](pT_8G) for the determination of the first few residues of the CMV RNAs [9]. That residue 2 was C was shown by the appearance of all residues from no. 3 onwards in sequencing reaction mixtures containing only d[α-^{32}P]CTP as the labelled triphosphate; by comparison, residue 11 was the first to appear when d[α-^{32}P]GTP was the only labelled triphosphate present. The reason for the lack of appearance of residue 2 under the conditions used is unknown.

3.3. *Sequences up to 220 residues from the 3′-termini of AMV RNAs*

Data for each RNA were obtained from sequencing gels of several different reaction mixtures using two or more different preparations of polyadenylated RNA. This was necessary because of the variation in the quality of sequencing gels which has been found difficult to control. Regions difficult to read in one gel were resolved by reference to one or more other gels. A background banding pattern was always found in all four ddNTP tracks for these longer sequences (fig.2) but this often proved useful in lining up gel bands.

For each RNA, there were two regions of ambiguity due to compression of nucleotide bands [8,17–19] but these were resolved using sequencing reaction mixtures in which unlabelled dGTP was replaced by dITP. This approach was based on the suppression of band compression in sequencing gels of the phage Qβ RNA replicase product by substitution of GTP by

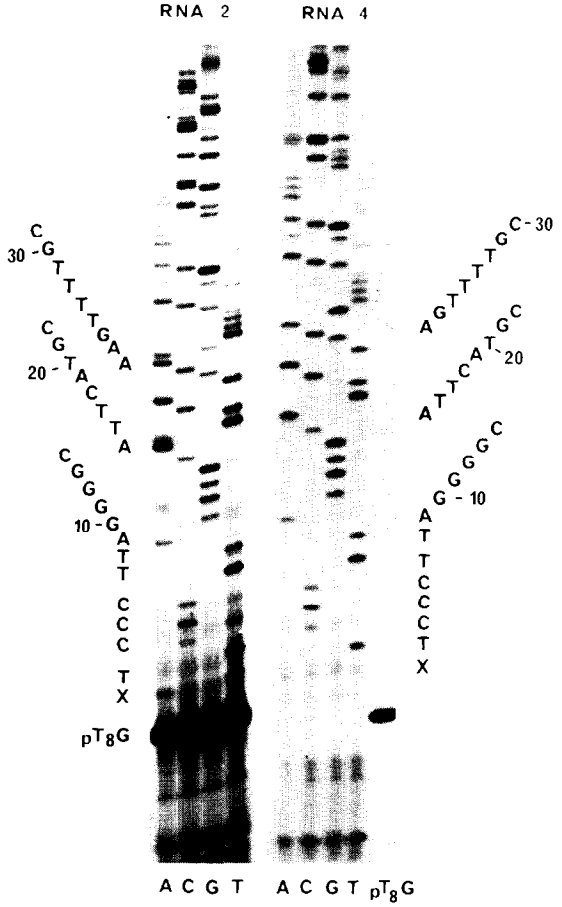

Fig.1. Sequencing gels for the 3′-ends of AMV RNAs 2 and 4. The reaction mixture for RNA 4 contained d[α-^{32}P]CTP and d[α-^{32}P]GTP while that for RNA 2 also contained d[5′-^{32}P](pT_8G). Only the bottom 30 cm of the 77 cm long 6% acrylamide gel is shown. A, C, G and T are the tracks of reaction mixtures containing ddATP, ddCTP, ddGTP and ddTTP, respectively. Residue 2 (X) does not appear as a band but was shown to be C (see text). Residues are numbered from the 3′-end of each RNA.

ITP in the RNA replicated reaction mixtures [19]. Examples of dGTP and dITP sequencing gels for these two regions for RNA 4 are shown in fig.2; sequences between residues 53 and 60 and residues 131 and 136 were clearly resolved in the dITP gels. It has been found more difficult to obtain clearly readable gels using dITP rather than dGTP; if better conditions can be determined for the use of dITP, e.g., by variation of the dITP concentration in the sequencing reaction mixtures and the use of ddITP rather than ddGTP, then this approach could be used routinely.

The sequence data for the four AMV RNAs are given in fig.3. Since the sequence of RNA 3 was identical to that of RNA 4, these sequences are presented together. There are a number of important features of this sequence data:

1. There is substantial sequence homology between the four RNAs up to residue 142 for RNAs 1, 3 and 4, and to residue 143 for RNA 2, but there is no homology thereafter between RNAs 1, 2 and 3 or 4.
2. The sequence of RNA 3 is identical to RNA 4 as expected since hybridization data showed that the sequence of RNA 4 is at the 3′-end of RNA 3 [7].

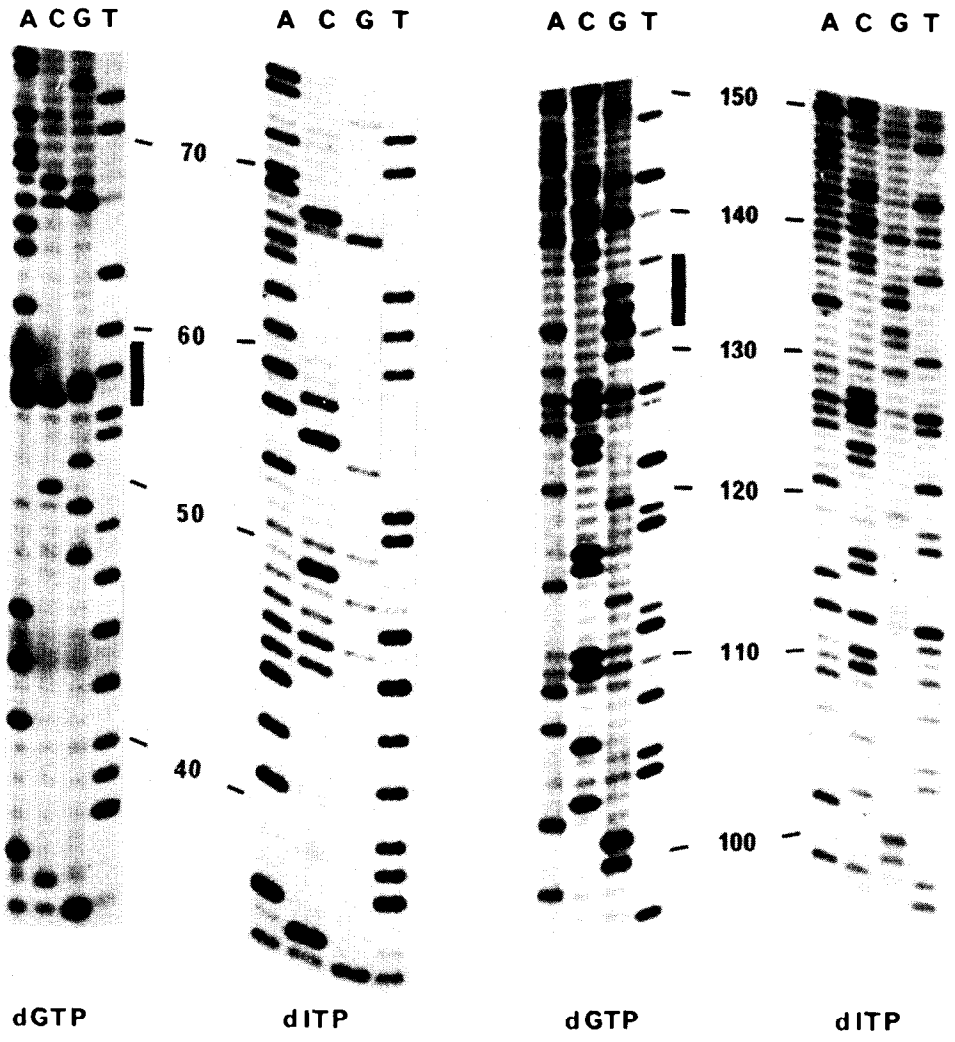

Fig.2. Comparison of sequencing gels of RNA 4 obtained using either dGTP or dITP in sequencing reaction mixtures. The region of band compression in each of the dGTP gels is shown by a vertical line. Residues are numbered from the 3′-end. A, C, G and T are as in fig.1. The two left-hand results were from different gels.

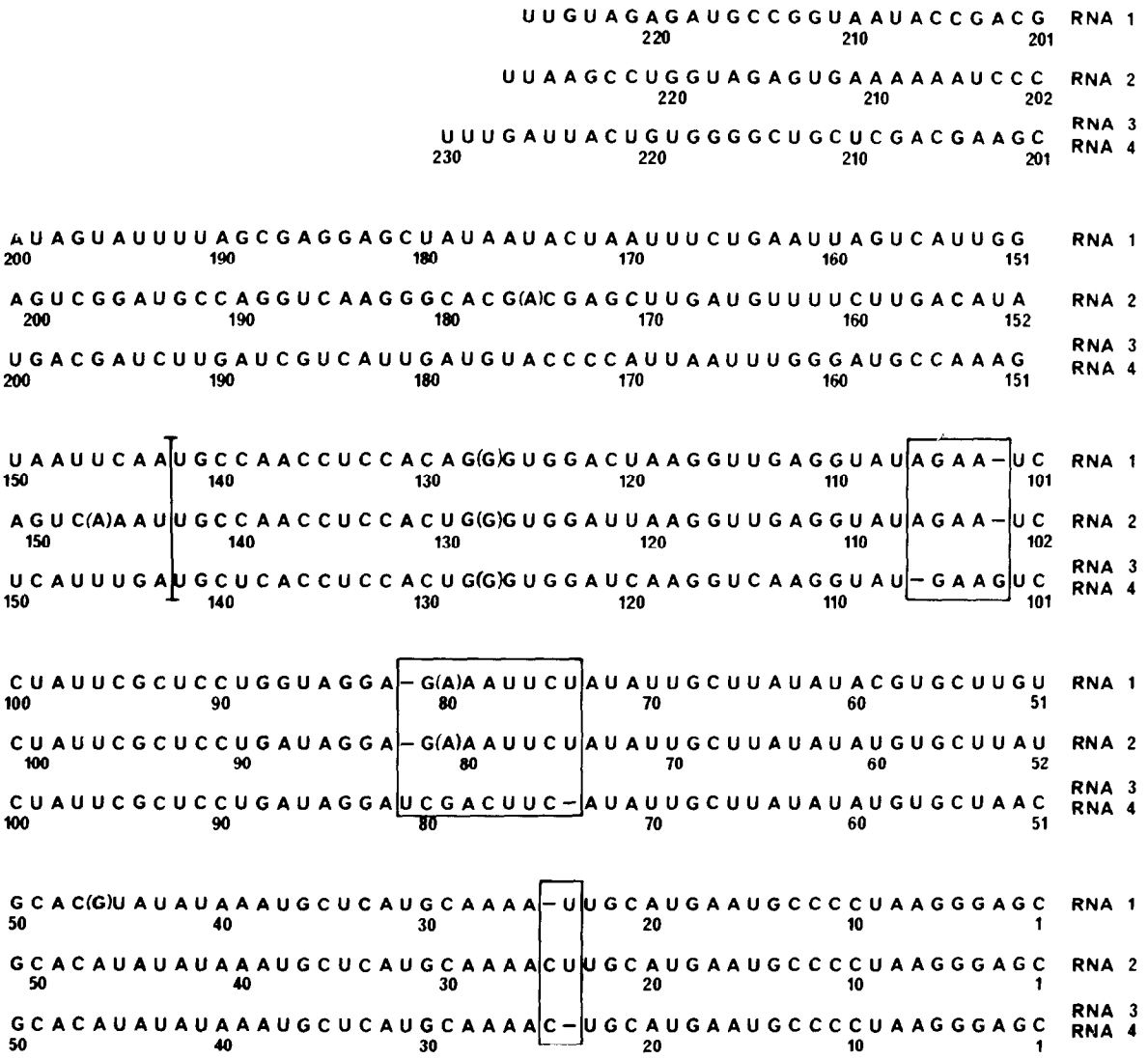

Fig.3. Sequences of the 3′-terminal regions of AMV RNAs. Sequences of RNAs 1 and 2 are arranged vertically with those of RNAs 3 and 4 to show maximum sequence homology; sequence homology terminates after residue 142 of RNAs 3 and 4 as shown by the vertical line. The absence of a residue is indicated by a dash (–). Where there is ambiguity, the most likely residue is given in parenthesis. The boxed areas are partially homologous regions with an addition or deletion of a residue.

3. As compared with RNAs 3 and 4 in the homologous region, there are 22 residues (15%) different in RNA 1 and 15 residues (10%) different in RNA 2; within the same region there are 7 differences between RNAs 1 and 2. The first 23 residues of all RNAs are identical while three regions (boxed in fig.3) contain an addition or deletion of a residue.
4. The sequence of the 3′-terminal 91 residues of RNA 4 of strain 425 of AMV have been determined in [20]. There are two differences with our data; residues 3 and 4 (AG, fig.3) are replaced by a C, and residue 91 is deleted in strain 425.
5. AMV RNA 4 contains ~825 residues [7], codes for a coat protein of ~220 amino acids [21,22] and has a 5′-untranslated region of 36 residues [23,24]. On the basis of these data, there are ~126 residues untranslated at the 3′-end of the RNA. There are six stop codons (five UGA and

one UAA) in all reading frames between residues 143 and 230 of RNA 4 so that the coat protein could terminate in this region. However, the precise termination point must await C-terminal amino acid sequence data.

3.4. *Possible base-paired regions in 3'-terminal sequences*

With the aid of the computer programme in [25], possible base-paired regions were searched for in RNA 4. The structure of fig.4 has 59% of the terminal 220 residues base-paired and has several interesting features. Of the 12 positions of RNA 4 up to residue 119 where base substitution occurs in RNAs 1 and 2, 8 are in single-strand regions and one (residue 77) is opposite a residue (residue 103) which is deleted. In two other positions of RNA 4 (residues 46 and 59), an A–U base-pair is substituted by a G–C base-pair in RNA 1. Further, in RNAs 1 and 2, the insertion of a U residue after

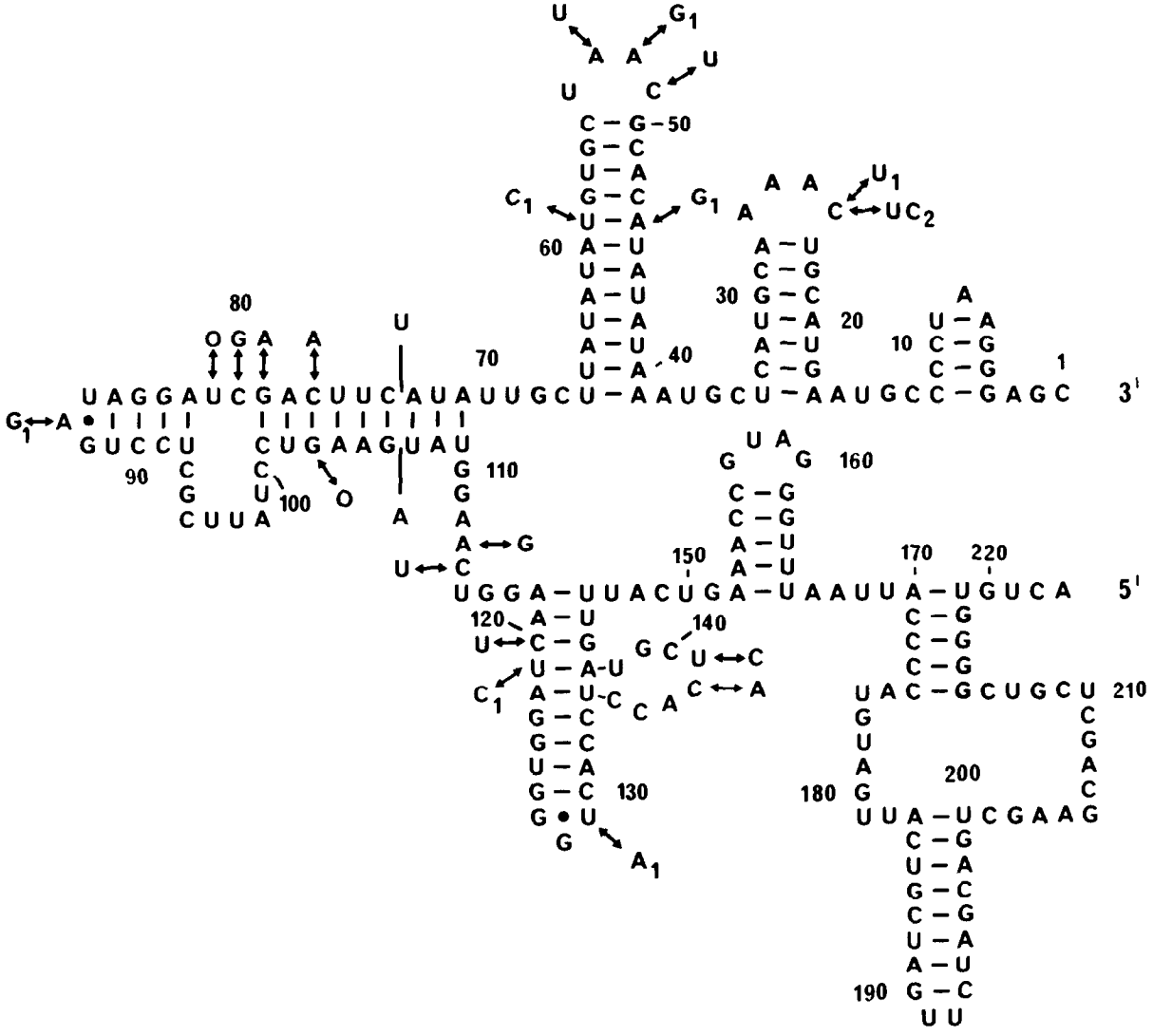

Fig.4. A possible base-paired structure for the 3'-terminal 223 residues of AMV RNA 4. Base substitutions in the 3'-terminal 142 residues of RNAs 1 and 2 shown; the subscript refers to the RNA involved while the absence of a subscript means that substitution occurs in both RNAs. The U to 0 and G to 0 substitutions at residues 81 and 103 are deletions in RNAs 1 and 2. The vertical lines after residues 73 and 106 indicate the insertion of a U and A residue in RNAs 1 and 2.

residue 73 of RNA 4 is coupled to the insertion of an A residue after residue 106. Finally, the 3′-terminal homologous region of the four RNAs terminates at the end of a loop (residue 142 of RNA 4). These considerations support the structure proposed in fig.4. However, the region of marked band compression between residues 53 and 60 (fig.2) would occur just after the loop of the hairpin structure (fig.4) and not further down the stem as would be expected on the basis of the results in [26] on the copying of MDV-1 RNA by the Qβ RNA replicase.

3.5. *Concluding comments*

The extensive sequence homology found here in the 3′-terminal sequences of the four AMV RNAs is intriguing but there is no information to indicate its possible biological significance. It would be of interest to see to what extent the actual sequence and arrangement are conserved in other strains of AMV. At least some of the terminal sequence must be involved in the recognition site of the RNA replicase responsible for the replication of the AMV RNAs [27] while the coat protein has been shown to bind to the 3′-terminal 10% (80–90 residues) of RNA 4 [28]. On the basis of the sequence data here, coat protein should also bind to the 3′-terminal region of RNAs 1, 2 and 3.

Our hybridization data [7] indicated that a maximum of ~50 residues of the 3′-terminal 300 residues of AMV RNA 4 were homologous with AMV RNAs 1 and 2. This sequence data illustrates the difficulty in obtaining accurate homology estimates using a hybridization approach.

Acknowledgements

The author thanks Dr R. I. B. Francki for the use of glasshouse facilities, Miss J. L. Rosey for assistance, Dr J. C. Wallace for help with the computer programme, and the Office of Program Resources and Logistics, National Cancer Institute, Bethesda, Maryland, for reverse transcriptase. This work was supported by the Australian Research Grants Committee.

References

[1] Jaspars, E. M. J. (1974) Adv. Virus Res. 19, 37–149.
[2] van Vloten-Doting, L. (1976) Ann. Microbiol. (Inst. Pasteur) 127A, 119–129.
[3] Peden, K. W. C. and Symons, R. H. (1973) Virology 53, 487–492.
[4] Lane, L. C. (1974) Adv. Virus Res. 19, 151–220.
[5] Kohl, R. J. and Hall, T. C. (1974) J. Gen. Virol. 25, 257–261.
[6] Gould, A. R. and Symons, R. H. (1977) Nucleic Acids Res. 11, 3787–3802.
[7] Gould, A. R. and Symons, R. H. (1978) Eur. J. Biochem. 91, 269–278.
[8] Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[9] Symons, R. H. (1980) Nucleic Acids Res. in press.
[10] Symons, R. H. (1978) Aust. J. Biol. Sci. 31, 25–37.
[11] Gould, A. R., Palukaitis, P., Symons, R. H. and Mossop, D. W. (1978) Virology 84, 443–455.
[12] Symons, R. H. (1977) Nucleic Acids Res. 4, 4347–4355.
[13] Sanger, F. and Coulson, A. R. (1978) FEBS Lett. 87, 107–110.
[14] Laskey, R. A. and Mills, A. D. (1977) FEBS Lett. 82, 314–316.
[15] Symons, R. H. (1975) Mol. Biol. Rep. 2, 277–285.
[16] Catterall, J. F., O'Malley, B. W., Robertson, M. A., Staden, R., Tanaka, Y. and Brownlee, G. G. (1978) Nature 275, 510–513.
[17] Zimmern, D. and Kaesberg, P. (1978) Proc. Natl. Acad. Sci. USA 75, 4257–4261.
[18] Brownlee, G. G. and Cartwright, E. M. (1977) J. Mol. Biol. 114, 93–117.
[19] Mills, D. R. and Kramer, F. R. (1979) Proc. Natl. Acad. Sci. USA 76, 2232–2235.
[20] Koper-Zwarthoff, E. C. and Bol, J. F. (1979) Proc. Natl. Acad. Sci. USA 76, 1114–1117.
[21] Collot, D., Peter, R., Das, B., Wolff, B. and Duranton, H. (1976) Virology 74, 236–238.
[22] Kraal, B., Van Beynum, G. M. A., De Graaf, J. M., Castel, A. and Bosch, L. (1976) Virology 232–235.
[23] Koper-Zwarthoff, E. C., Lockard, R. E., De Weerd, B. A., Rajbhandary, U. L. and Bol, J. F. (1977) Proc. Natl. Acad. Sci. USA 74, 5504–5508.
[24] Pinck, L., Franck, A. and Fritsch, C. (1979) Nucleic Acids Res. 7, 151–166.
[25] Staden, R. (1977) Nucleic Acids Res. 4, 4037–4051.
[26] Kramer, F. R. and Mills, D. R. (1978) Proc. Natl. Acad. Sci. USA 75, 5334–5338.
[27] Clerx, C. M. and Bol, J. F. (1978) Virology 91, 453–463.
[28] Houwing, C. J. and Jaspars, E. M. J. (1978) Biochemistry 17, 2927–2933.