Volume 7 Number 5 1979 Nucleic Acids Research

Determination of the length distribution of poly(A) at the 3' terminus of the virion RNAs of EMC virus, poliovirus, rhinovirus, RAV-61 and CPMV and of mouse globin mRNA

Paul Ahlquist and Paul Kaesberg

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

Received 8 August 1979

ABSTRACT

Rapid, detailed, and accurate analysis of the length spectrum of 3' terminal poly(A) in an RNA population can be obtained by 3'-terminal ^{32}p-labelling of RNA with T4 RNA ligase, digestion with ribonucleases T1 and A, and use of gel sequencing methods. Length distributions of 3'-terminal poly(A) of EMC virus, poliovirus, rhinovirus, RAV-61, and CPMV virion RNAs as well as mouse globin mRNA are presented.

INTRODUCTION

A common but little understood feature of eukaryotic mRNAs is the presence of 3' terminal poly(A) which is heterogenous in length. Such poly(A) tracts can be released by digestion with RNase T1 and/or RNase A and displayed on polyacrylamide gels, but their accurate sizing has been frustrated by the anomalous electrophoretic mobility of poly(A) even under highly denaturing conditions (1). To overcome this problem, synthetic homopolynucleotides have been used directly as size markers in several studies (2-4) although such markers of appropriate and precisely known length are difficult to obtain.

We find that 3' terminal ^{32}p-labelling with T4 RNA ligase (5), which requires less than pmol quantities of polyadenylated RNA, allows convenient generation of both the intact 3' terminal poly(A) chains and poly(A) marker fragments of all possible sizes in a highly radioactive form. Displaying both sets of fragments on high resolution nucleic acid sequencing gels (6) provides base by base enumeration of 3' poly(A) tail lengths and allows their distribution to be determined directly and rapidly. We report the use of these methods to obtain length spectra of 3' terminal poly(A) of several RNAs.

MATERIALS AND METHODS

RNA was extracted (7) from EMC and poliovirus type I (Mahoney), obtained from Mark Pallansch and Roland Rueckert, and from human rhinovirus 14, obtained from John Mapoles, Joseph Icenogle, John Anderegg, and Charles McLean. These viruses had been grown and isolated as previously described (7-9). Polyadenylated RAV-61 vRNA (10), selected by binding to oligo(dT)-cellulose, was a gift of David Shealy and Roland Rueckert. Mouse globin mRNA (11) was a gift of Jeffrey Ross. CPMV (yellow strain) was grown in blackeyed cowpea from inoculum provided by Ramsey Frist and purified according to Klootwijk et al. (12). M and B virions were separated and RNA extracted according to Steele and Frist (13).

RNA ligation reactions were modelled after England and Uhlenbeck (5) and typically contained 5-10 pmol substrate RNA, 2 μM [5'-^{32}P]pCp (1000-2000 Ci/mmol, from New England Nuclear), 12 units T4 RNA ligase (PL Biochemicals), and 12 μM ATP in 20 μl 50 mM Hepes pH 8.3, 10 mM $MgCl_2$, 3.3 mM DTT, 12.5% glycerol and 10% DMSO. After the reaction was incubated 24 hours at 4-5°C in a sealed glass capillary, the RNA was precipitated twice with ethanol and washed once with ethanol.

RNase digestions and gel electrophoresis were as described for determination of RNA sequences (14) except that RNase T1 and RNase A concentrations were increased to 2 units/μg RNA and 2 x 10^{-3} μg/μg RNA respectively to give complete digestion. Urea was omitted from digests with RNase φ1. Sensitized autoradiography (15) was carried out at -80° with pre-flashed Kodak X-Omat R film and Dupont Cronex Lightning-Plus intensifying screens. Autoradiographs were densitometered on a Joyce-Loebl MkIIIc microdensitometer.

RESULTS

Figure 1a shows an autoradiogram of five separate digests and an undigested control of 3' ligase-labelled EMC virion RNA (vRNA) displayed on a sequencing gel. Near the bottom of the gel, strong bands appear only in partial digests with formamide (which gives cuts at all residues), RNase U2 (which cuts after A residues), and RNase φ1 (which cuts after all but C residues), showing that EMC vRNA terminates at its 3' end with a run of A residues. Further up the gel, bands resulting from digestion at other than A residues first appear in the complete digest with RNase T1 (which cuts after G residues). In this digest fragments from approximately 18 to over 60 residues in length occur, showing that the number

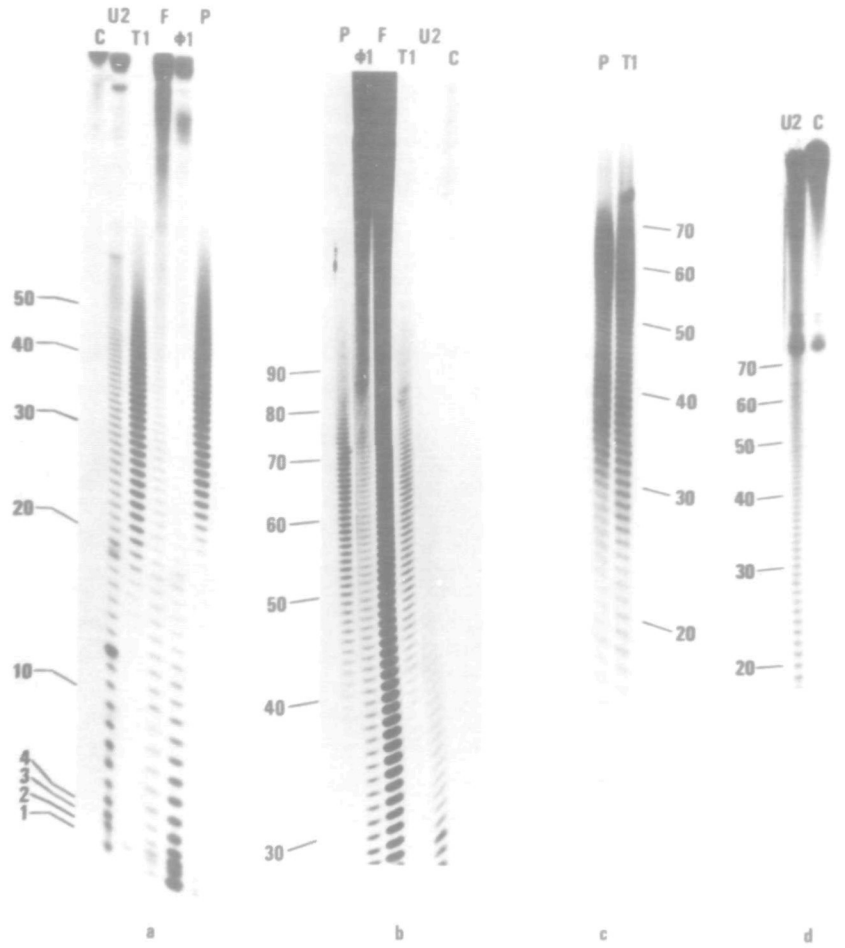

Figure 1. Autoradiograms of digests of 3' terminal ligase-labelled RNAs separated on 12% acrylamide 7 M urea thin sequencing gels. The following symbols are used to denote the various lanes: C - minus enzyme control; U2 - partial RNase U2 digest; T1 - complete RNase T1 digest; F - partial formamide digest; φ1 - partial RNase φ1 digest; P - complete RNase A digest. The numbering of the bands refers to the number of nucleotides present from the 3' end of the original RNA and does not include the terminal C added by ligation. (a) EMC vRNA digests. (b) Poliovirus vRNA digests. (c) Mouse globin mRNA digests. Peaks in the poly(A) distribution near residues 39 and 56 are clearly visible. (d) CPMV M vRNA digests.

of consecutive A residues at the 3' end of the RNA varies accordingly. This information is sufficient to define the length spectrum of 3' terminal poly(A) on EMC vRNA. However, we note that the complete digest with RNase A (which cuts after C and U residues) gives a pattern similar to that of the complete RNase T1 digest, but shifted up four nucleotides. This is in accord with the 3' terminal sequence '''U-A-G-A-G- poly(A) as reported by Merregaert et al. (16).

Three anomalies in Fig. 1a deserve mention. First, faint, mono-, di-, and trinucleotide bands appear in the complete RNase T1 and RNase A digests. Inasmuch as these bands are quite weak and are discontinuous with the primary distribution of bands in these lanes, we believe they result from digestion of contaminating non-polyadenylated RNA fragments. A group of such contaminating RNA fragments is faintly visible in the undigested control of Fig. 1a, below full length EMC vRNA. Second, a rather broad band is found in the partial RNase U2 digest just below band 13. A similar but weaker band migrates between bands 19 and 20 in both the partial U2 and ϕ1 digests. The source of these bands is unknown although they were found in similar partial enzymatic digests of other 3' pCp-labelled polyadenylated RNAs. Such bands did not appear in partial formamide digests.

The above techniques were applied to analyze the poly(A) length spectrum of six other RNA samples. Digests of poliovirus vRNA, mouse globin mRNA, and CPMV M vRNA are shown in Fig. 1b-1d. It was convenient for us to use 5-10 pmol of polyadenylated substrate RNA for the ligation reactions. However, such reactions were typically found to yield over 100 times the incorporated label required for these experiments, indicating that substantially less substrate RNA could be used. Multiple loadings of the digests for each RNA were electrophoresed as necessary to resolve the entire distribution of bands. To maintain the count of fragment size on gels where the smallest formamide digestion fragments, and thus the origin of counting, were run off the gel, it was necessary to generate landmarks along the poly (A) ladder. Such landmarks were provided by electrophoresing digests of RNA fragments having distinctive sequences alongside digests of 3' labelled polyadenylated RNAs.

Figures 2 and 3 show the poly(A) length spectrum of the seven RNA samples tested. These spectra were determined by graphically integrating the optical density of autoradiograph bands generated by complete RNA digestion with RNase T1 or A. Adjacent bands on the autoradiograms

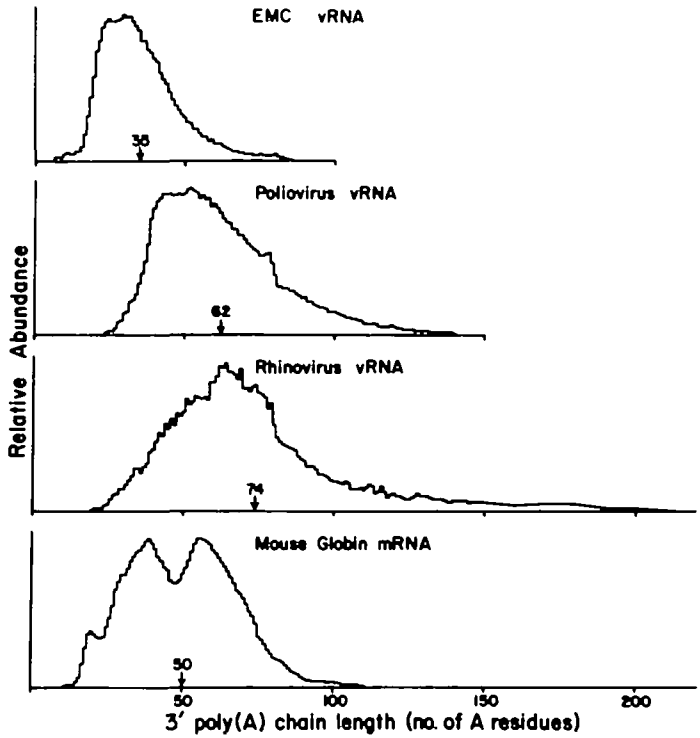

Figure 2. Length distribution of 3' terminal poly(A) chains released from picornavirus vRNAs and mouse globin mRNA by complete RNase T1 digestion. The abundance of chain lengths over 145 residues long in mouse globin mRNA was approximated as discussed in the text for every tenth residue and the resulting points were connected with straight lines. Vertical arrows mark the position of the number average 3' poly(A) chain length for each distribution. This value is indicated above the arrow.

which appeared by eye to be well resolved were frequently shown by densitometry to overlap slightly. Since adjacent bands were usually of comparable optical density it was assumed in such cases that spillover between bands was equal in both directions. Also, poly(A) chains over 240 residues long were not resolved on the 40 cm gels used. (They should be resolvable on longer gels.) Their abundance was estimated as follows. For poly(A) chains over 15 residues long, band mobility on the sequencing gels was found to be a linear function of ln n, and band spacing thus proportional to $\ln(\frac{n+1}{n})$, where n equals the number of

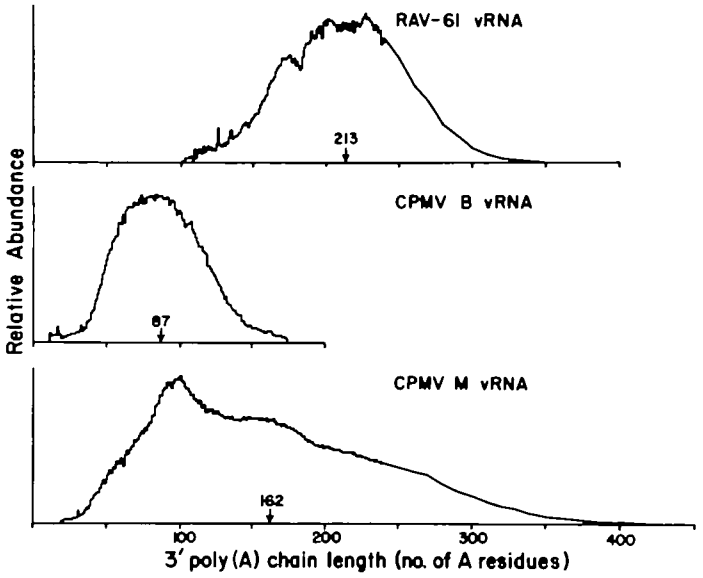

Figure 3. Length distributions of 3' terminal poly(A) chains released from RAV-61 and CPMV vRNAs by complete RNase A digestion. The abundance of chain lengths over 240 residues long in RAV-61 and CPMV vRNAs was approximated as discussed in the text for every tenth residue. The number average 3' poly(A) chain length is indicated for each distribution by a vertical arrow.

residues in the chain. The position on an autoradiogram of an unresolved poly(A) chain n residues long was determined by extrapolation according to ln n. The abundance of that length poly(A) chain was taken to be proportional to the optical density at the extrapolated point multiplied by the band spacing parameter $\ln(\frac{n+1}{n})$.

DISCUSSION

Previous determinations of picornavirus vRNA poly(A) length have failed to agree completely. The average length of the 3' terminal poly(A) chain on unfractionated EMC vRNA has been variously reported as 14 to 18 residues (17), 10 to 20 residues (18), and 59 to 67 residues (19). We find an intermediate value of 35 residues (number average). Yogo and Wimmer (20) reported the average length of poliovirus vRNA poly(A) to be 89 residues, while a later report (21) placed the value at 75 residues. Burness et al. (17) concluded that even this latter value

is an overestimate. Our determination gives the average length of poliovirus vRNA poly(A) as 62 residues. The average poly(A) length on human rhinovirus vRNA has been reported as 90 residues (22) and 150 residues (23); we find the average to be 74 residues. We further note that the poly(A) length distributions for vRNA of all three picornaviruses are asymmetric, being skewed toward the smaller lengths. Some of the conflict in the above reports may be due to actual differences in poly(A) length depending on the strain of virus and virus growth conditions used, as suggested by Gillespie et al. (24). Hruby and Roberts (19), however, found that choice of cell suspension media had little influence on the average poly(A) length in EMC vRNA.

Morrison et al. (25) found the average 3' poly(A) tract on mouse globin mRNA to be 50 nucleotides long. Gorski et al. (26) further showed that the length distribution of mouse globin mRNA poly(A) chains contained major peaks centered at 40 and 60 residues plus a shoulder of chains 75 to 120 residues long electrophoresing next to the peak at 60 residues. This uneven length distribution was common to both α- and β-globin mRNAs. Our results are similar. We find the average poly(A) length to be 50 residues and our length distribution contains two distinct major peaks at approximately 39 and 56 residues, respectively, with the suggestion of a shoulder on the high side of the peak at 56 residues. However, we also find a minor but distinct peak centered at 20 residues.

We find that RAV-61 vRNA bears a 3' terminal poly(A) chain averaging 213 residues in length. This is comparable to Rous sarcoma virus vRNA whose poly(A) is approximately 200 residues long (27). Since the RAV-61 vRNA used in this study was selected by binding to oligo(dT)-cellulose, we are unable to say whether any fraction of RAV-61 vRNA is unpolyadenylated, as has been found for Rous sarcoma virus vRNA (27).

In a study by El Manna and Bruening, the length distribution of 3' terminal poly(A) chains on CPMV M and B vRNAs was shown to be qualitatively similar by sucrose density gradient sedimentation (28). Recovery of RNase T1 and RNase A resistant oligonucleotides from a mixture of uniformly ^{32}P-labelled CPMV M and B vRNAs gave the average length of 3' terminal poly(A) chains as 100 residues. Steele and Frist (13) estimated the peak poly(A) lengths on both CPMV M and B vRNAs at 113 residues. Using gel electrophoresis, they found no gross difference between the overall poly(A) length distributions of the two RNAs, but this may be due to the short electrophoretic runs used which failed to move the

longest poly(A) fragments discernibly away from the origin. Using higher resolution gels and longer electrophoretic runs, we find that a substantial difference does exist in the length distribution of poly(A) chains on these two RNAs. Although the poly(A) length spectra for both CPMV M and B vRNAs peak in the range of 80-100 residues, poly(A) chains of over 300 residues are present on CPMV M vRNA while CPMV B vRNA contains essentially no poly(A) chains over 170 residues in length. It is conceivable that the extremely wide distribution in poly(A) chain length observed for CPMV M vRNA is an artifact caused by the presence of a large RNase-resistant oligonucleotide adjacent to the 3' poly(A). However, this seems unlikely in view of the highly denaturing conditions used for RNase incubation and the fact that the band pattern was not changed by increasing the RNase A concentration 25-fold. Since CPMV M vRNA (1.4 Md) and CPMV B vRNA (2.0 Md) are believed to be enclosed in identical capsids, the absence of longer poly(A) tracts on CPMV B vRNA may represent selective encapsidation. In addition to the full-length genomic RNAs whose poly(A) lengths were measured, the ligated CPMV M and B vRNA preparations contained 4 to 5 labelled bands which co-electrophoresed with polyadenylate fragments of 70 to 90 bases (Fig. 1d). The only other case of significant labelling of lower MW RNA species in these experiments occurred with the RAV-61 vRNA sample. Here the tRNA primer which is encapsidated with the genomic RNA was also labelled efficiently by RNA ligase. Inasmuch as RNA ligase will not label the 3' phosphate ends generated by common laboratory-contaminating endonucleases and the same low MW labelled bands were reproduced in preparations of both CPMV vRNAs, which are isolated independently from separate classes of virions, the possibility arises that the presence of small RNAs in the CPMV samples is of biological significance.

As reported here, end labelling with T4 RNA ligase provides several advantages for studying 3' terminal poly(A) tracts. The method is extremely rapid, requires only pmol quantities of RNA, is insensitive to internal poly(A) tracts, and, in combination with existing sequencing technology, can give very detailed results. Most importantly, poly(A) chain length is determined directly by enumeration, eliminating the accuracy problems discussed by Burness et al. (17) for previous methods. The ability to determine the complete poly(A) length spectrum of an RNA sample may make this method particularly useful in studies of poly(A) synthesis and metabolism.

ACKNOWLEDGMENTS

We thank Mark Pallansch, David Shealy, Charles McLean, Roland Rueckert, John Mapoles, Joseph Icenogle, John Anderegg, Jeffrey Ross, and Ramsey Frist for gifts of viruses and/or RNAs. We also thank Charles McLean for helpful discussions. This research was supported by NIH grants AI 1466 and AI 15342, NIH Career Award AI 21942 and by the Science and Education Administration of the U.S.D.A. under grant 7800002. P.A. was supported by an NSF graduate fellowship.

Abbreviations: EMC virus, encephalomyocarditis virus; RAV-61, Rous-associated virus 61; CPMV, cowpea mosaic virus.

REFERENCES

1. Burness, A., Pardoe, I. and Goldstein, N. (1975) Biochem. Biophys. Res. Commun. 67, 1408-1414.
2. Wilt, F. (1977) Cell 11, 673-681.
3. Sarkar, N., Langly, D. and Paulus, H. (1978) Biochemistry 17, 3468-3474.
4. Schultz, G., Chaconas, G. and Moore, R. (1978) J. Bacteriol. 133, 569-575.
5. England, T. and Uhlenbeck, O. (1978) Nature 275, 560-561.
6. Sanger, F. and Coulson, A. (1978) FEBS Letters 87, 107-110.
7. Shih, D., Shih, C., Kew, O., Pallansch, M., Rueckert, R. and Kaesberg, P. (1978) Proc. Nat. Acad. Sci. 75, 5807-5811.
8. Hall, L. and Rueckert, R. (1971) Virology 43, 152-165.
9. Medappa, K. and Rueckert, R. (manuscript in preparation).
10. Shealy, D. and Rueckert, R. (manuscript in preparation).
11. Ross, J., Gielen, J., Packman, S., Ikawa, Y. and Leder, P. (1974) J. Mol. Biol. 87, 697-714.
12. Klootwijk, J., Klien, I., Zabel, P. and van Kammen, A. (1977) Cell 11, 73-82.
13. Steele, K. and Frist, R. (1978) J. Virol. 26, 243-248.
14. Ahlquist, P., Dasgupta, R., Shih, D., Zimmern, D. and Kaesberg, P. Nature (in press).
15. Laskey, R. and Mills, A. (1977) FEBS Letters 82, 314-316.
16. Merregaert, J., van Emmelo, J., Devos, R., Porter, A., Fellner, P. and Fiers, W. (1978) Eur. J. Biochem. 82, 55-63.
17. Burness, A., Pardoe, I., Duffy, E., Bhalla, R. and Goldstein, N. (1977) J. Gen. Virol. 34, 331-345.
18. Giron, M., Logeat, F., Hanania, N., Fossar, N. and Huppert, J. (1976) Intervirol. 6, 367-371.
19. Hruby, D. and Roberts, W. (1976) J. Virol. 19, 325-330.
20. Yogo, Y. and Wimmer, E. (1972) Proc. Nat. Acad. Sci. 69, 1877-1882.
21. Spector, D. and Baltimore, D. (1975) J. Virol. 15, 1418-1431.
22. MacNaughton, M. and Dimmock, N. (1975) J. Virol. 16, 745-748.
23. Nair, C. and Panicali, D. (1976) J. Virol. 20, 170-176.
24. Gillespie, D., Takemoto, K., Robert, M. and Gallo, R. C. (1973) Science 179, 1328-1330.

25. Morrison, M., Merkel, C. and Lingrel, J. (1973) Mol. Biol. Rpts. 1, 55-60.
26. Gorski, J., Morrison, M., Merkel, C. and Lingrel, J. (1974) J. Mol. Biol. 86, 363-371.
27. Wang, L. and Duesberg, P. (1974) J. Virol. 14, 1515-1529.
28. El Manna, M. and Bruening, G. (1973) Virology 56, 198-206.