A new method for the size estimation of the RNA genome segments of influenza virus

M.J. Sleigh, G.W. Both and G.G. Brownlee

CSIRO Molecular and Cellular Biology Unit, P.O. Box 184, North Ryde, N.S.W., 2113 Australia

Received 9 February 1979

ABSTRACT

Previous estimates of the size of the RNA genome segments of influenza virus have been unreliable because of a lack of suitable RNA species as size markers. We have attempted to overcome this problem by utilising the ability of AMV reverse transcriptase to synthesise full length DNA copies of RNA molecules in the presence of a suitable primer. By comparing such DNA copies of the RNA segments of the influenza virus genome with sequenced restriction fragments from the E. coli plasmid pBR322, we have made more reliable estimates of the sizes of the eight genome segments from influenza virus A/NT/60/68.

INTRODUCTION

The genome of influenza A virus is segmented, and consists of eight single-stranded RNA species of negative polarity. Each segment is transcribed into a single mRNA which is subsequently polyadenylated and translated to produce a unique viral polypeptide (reviewed in reference 1). Various methods have been used to estimate the molecular weights of the genome segments, including sucrose density gradient centrifugation (2, 3) and polyacrylamide gel electrophoresis under neutral (2-5) or partially denaturing conditions (6-8). More recently, Desselberger and Palese (9), using glyoxylation to ensure RNA denaturation, obtained much lower molecular weight estimates than those reported previously, with segment sizes ranging from $8.9 \times 10^5$ to $2.1 \times 10^5$ daltons for the genome of influenza A/PR/8/34. Most of these RNA size estimations made use of 16S, 18S, 23S and 28S ribosomal RNA markers, molecules which had themselves not been accurately sized by sequence determination, although the sequence of E. coli 16S RNA is now known (10). In addition, the ribosomal RNA markers do not...
Nucleic Acids Research

adequately cover the entire size range of the influenza RNA species. A partial solution to these difficulties would be to use other RNA markers such as β-globin and ovalbumin mRNAs, which have now been fully sequenced (11, 12), although sizes for even these molecules are only approximate, since they show heterogeneity in the lengths of their poly (A) tails. In this paper, we suggest an alternative approach. Poly (A) tails are added to the 3' ends of influenza RNA segments using E. coli terminal riboadenylate transferase (13, 14). Oligo (dT)$_{10}$ or (dT)$_8$ dA molecules, hybridised to these poly (A) tails, act as primers for the synthesis by reverse transcriptase of full length DNA copies (cDNA) of the eight viral RNA segments. The radioactively labelled viral cDNA is then analysed by polyacrylamide gel electrophoresis under denaturing conditions. The molecular weights are estimated by comparing the mobilities of the cDNA species with those of appropriate completely sequenced, single stranded DNA markers derived by restriction enzyme digestion of the E. coli plasmid pBR 322 (15).

MATERIALS AND METHODS

Growth of Virus. Influenza virus strains used were A/Mem/102/72 (16), a laboratory strain constructed by recombination between a 1972 isolate of the Hong Kong (H3N2) subtype and A/PR/8/34 (HonI), with selection for haemagglutinin and neuraminidase of the Hong Kong subtype, and A/NT/60/68 (17), a field isolate antigenically identical to the 1968 Hong Kong strain. Purified virus, kindly supplied by Drs. V. Bender and B. Moss, was propagated in embryonated chicken eggs and purified by adsorption to and elution from red blood cells (18). Subsequent purification was by centrifugation through sucrose density gradients (10-40%, w/w in phosphate buffered saline (0.01M sodium phosphate, pH 7.3, 0.14M NaCl, 0.08% NaN$_3$)) in a Beckman SW27 rotor for 90 min at 20,000 rpm, followed in some cases by isopycnic banding on a 34-50% (w/w) sucrose gradient in 0.05M Tris-HCl, pH 7.4, 0.14M NaCl, 0.08% NaN$_3$ in a Beckman SW27 rotor (16 hr at 20,000 rpm).

Extraction and adenylation of viral RNA. The viral RNA was extracted and polyadenylated at its 3' terminus using E. coli terminal riboadenylate transferase (13, 14) as described previously (19). The conditions used for polyadenylation were sufficient to add 10-50 A
residues per viral RNA molecule.

Synthesis of a DNA Copy (cDNA) of Virion RNA. A DNA copy of total virion RNA was synthesised under conditions optimal for the production of full length single strands (20). Generally p(dT)$_{10}$ (P.-L. Biochemicals, Milwaukee, Wisc.) was used as primer, but for more accurate size estimations the primer p(dT)$_8$dA was used. The former could hybridize anywhere on the poly (A) tail, while the latter would be expected to hybridize to and prime specifically from the 5' end of the poly (A) tail, since the 3' terminal nucleotide of all the influenza RNA segments is U(3). Reaction mixtures (30 µl) contained Tris-HCl, pH 8.3 (50 mM), MgCl$_2$ (10 mM), KCl (70 mM), dithiothreitol (10 mM), dCTP, dGTP, dTTP (each 500 µM), α- dATP (100 µM), primer (40 µM), influenza RNA (20 µg) and 20 units of reverse transcriptase from avian myeloblastosis virus (Lot No. G-678, generously supplied by Dr. J. W. Beard, Life Sciences, St. Petersburg, Fla., U.S.A., under the NIH-National Cancer Institute Contract No. NO1-CP-33291). Incubation was for 60 min at 42°. The cDNA was freed of RNA by boiling the reaction mixture for 1 min, and then incubating for a further 5 min at 42° with 2 µg pancreatic RNase (Worthington). The DNA preparation was extracted with a phenol-chloroform mixture (1:1) adjusted to 0.2 M NaCl and then precipitated at -70°C after the addition of 3 volumes of ethanol.

Preparation of DNA Markers from Plasmid pBR322. The plasmid pBR322 was kindly supplied by the Plasmid Reference Center, Stanford University, and was propagated in E. coli RRL and amplified by chloramphenicol treatment (21). Plasmid DNA was prepared by the sarkosyl lysis method of Clewell (22), followed by separation on CsCl/ethidium bromide gradients. Ethidium bromide was removed from the bands of supercoiled plasmid DNA by extraction with 2-propanol, and DNA was dialysed extensively against 0.01M Tris-HCl buffer, pH 7.6, containing 1 mM EDTA.

Restriction enzymes used to digest pBR322 DNA were obtained from New England Biolabs, and digestions were carried out under the conditions recommended for each enzyme by the suppliers. In order to obtain DNA markers of suitable sizes, separate samples of pBR 322 DNA (2 µg) were digested with the enzymes AluI and Hinfl, and with the enzyme combinations EcoRI + PvuII, PvuII + BamHI and Hinfl + PstI. When digestion was complete, the DNA fragments formed were labelled at their 3' ends by adding to the reaction mixtures one unit of E. coli DNA polymerase I (Klenow
enzyme – Boehringer) and 10 μCi α-32P-dATP (Amersham). Incubation was continued at 20° for 20 min, or 40 min in the case of AluI-digested fragments (Sanger, personal communication). EcoRI-digested DNA "fills in" by the addition of two A residues, and HinfI-cut DNA is similarly labelled when one or two A residues are added. In DNA cut by PvuII or AluI, the 3' exonuclease of DNA polymerase I removes the 3' terminal G-residue, exposing an A which is labelled in a 3' exonuclease/polymerase exchange reaction. BamHI-cut termini are labelled by progressive degradation by the 3' exonuclease until an A residue is reached.

After labelling, the five restriction digests were mixed so that all marker bands had similar radioactivity. The sizes of the largest DNA fragments, to be used as markers during molecular weight estimations, were calculated from the data of Sutcliffe (15). These were 2292 and 2069 bases (the two fragments from the EcoRI + PvuII double digest), 1631 bases (the largest HinfI fragment) 1389 bases (the largest fragment from the HinfI + PstI double digest), and 910, 659 and 655 bases (the three largest fragments in the AluI digest). These sizes do not take into account staggered ends generated by restriction enzyme cleavage, or bases added or removed during the subsequent labelling reaction.

Polyacrylamide Gel Electrophoresis. RNA or DNA species were separated under denaturing conditions by electrophoresis on 2.6% acrylamide/0.14% N,N-methylene bisacrylamide thin gels containing 7M urea (23). Gel and running buffer were the Tris/borate/EDTA system of Peacock and Dingman (24). Samples (final volume 1-5 μl) were denatured before loading by heating at 100° for 1 min in 2 volumes of formamide containing 5 mM EDTA, 0.01% xylene cyanol FF and 0.01% bromophenol blue (25). Gels (0.3 x 200 x 400 mm) were run at 1000 v for 4-5 hr (gel temperature approx. 60° to favour denaturation). 32P-labelled DNA bands were identified by exposing the wet gel at -70° to Fuji RX film using Dupont Cronex Lightning plus intensifying screens (26). Unlabelled DNA or RNA was identified by staining with ethidium bromide and then photographing under UV illumination (254 nm) using a red filter.

RESULTS
We first compared the RNA from the influenza strain A/Mem/102/72 with its DNA copy analysing the products side-by-side by electrophoresis on a 2.8% acrylamide gel containing 7 M urea. The DNA and RNA bands were identified by ethidium bromide staining as shown in Fig 1a. Two further separations
FIGURE 1. Polyacrylamide gel fractionation (2.8% in 7 M urea) of influenza RNA (A/Mem/102/72) and its derivative cDNA. (a) shows RNA (1) and cDNA (2) synthesised using p(dT) fractionated side-by-side. The products were detected by ethidium bromide staining. (b) is a composite of (1) RNA detected as above and (2) a radioautograph of $^{32}$P cDNA.

of the RNA and copy DNA are shown in Fig 1b although this is a composite figure prepared from two different experiments. Fig 1b (slot 1) shows the separation of RNA band 3 from the mixture of bands 1 and 2; slot 2 shows the cDNA bands 1 to 6, detected by radioautography more clearly than is apparent in Fig 1a (slot 2). From the results, the pattern for cDNA closely parallels that for viral RNA suggesting that all eight RNA segments (labelled 1-8) were copied into DNA to produce copies of similar length to the parent RNA species. However, we have found that the relative yields of cDNA vary for different bands, with better yields from the smaller RNA segments 5-8 than from the larger segments (see Fig 1a, slot 2; and Fig 1b, slot 2). Figure 1a also shows that the cDNA bands appear somewhat larger and more diffuse than the parent RNA (at least for bands 7 and 8). This is explained by the presence of the
additional T residues due to the primer p(dT)$_{10}$ on the cDNA.

In a separate experiment, actinomycin D, which is known to inhibit the DNA-dependent DNA synthetic activity of reverse transcriptase (27), was added to the reaction at concentrations of 75-300 µg/ml. This did not change the sizes of the DNA species synthesised, suggesting that the DNA formed under our standard reaction conditions in the absence of actinomycin D is predominantly single stranded, with little, if any, double stranded DNA present. A background of material is present in both the DNA and RNA slots of Fig 1, and probably represents incomplete or degraded RNA and DNA transcripts, present in low yield.

Under the electrophoresis conditions used for this gel, we observed only two bands in the expected position for RNA bands 1-3. Even using partially denaturing conditions for electrophoresis (8) we failed to resolve three bands in this region, although there was a much clearer separation into two distinct bands. However, there is considerable evidence that for Hong Kong (H3N2) influenza subtypes there are three genome segments migrating in this region (e.g. refs. 8, 9) and on the basis of patterns previously obtained by electrophoresis of influenza RNA under denaturing conditions (9) we have identified our top band as a mixture of RNA segments 1 and 2, and the lower band as segment 3.

Figure 2a shows a radioautograph of the separation obtained by acrylamide gel electrophoresis under denaturing conditions of $^{32}$P-labelled DNA copied as already described from the RNA of influenza A/NT/60/68, a strain closely related to A/Mem/102/72. The cDNA bands are compared with $^{32}$P-labelled, denatured DNA markers derived from the E. coli plasmid pBR322 in slot 1 (see Methods). In order to avoid slight mobility differences in different channels, markers and cDNA were also mixed before electrophoresis in slots 2 and 5. The cDNA used for the separation shown in slots 2 and 3 was prepared using p(dT)$_{10}$ as primer, while for that shown in slots 4 and 5 the phased primer p(dT)$_{8}$/dA was used. Since the 3' terminal residue of the genome RNA segments is U(3), this latter primer should hybridise only to the 5' end of the poly (A) tail, leading to discrete cDNA bands rather than a mixture of closely related products as obtained with p(dT)$_{10}$ as primer.

Fig 2b shows an experiment similar to that of Fig 2a, but in this experiment all three slots contain a mixture of influenza cDNA and the pBR322 DNA markers in varying ratios. This, together with the greater resolution obtained with increased electrophoresis time, clarified some
FIGURE 2. Polyacrylamide gel fractionation (as in Fig 1) of influenza cDNA (A/NT/60/68) and pBR322 markers. (a) contains in slot 1 pBR322 denatured markers of the stated length in bases. Slots 3 and 4 contain cDNA synthesised with the primer p(dT)\textsubscript{10} or p(dT)\textsubscript{8}dA, respectively. Slot 2 is a mixture of 1 and 3 and slot 5 a mixture of 1 and 4. (b) slots 1 to 3 contain mixtures of the pBR322 markers and cDNA synthesised using p(dT)\textsubscript{8}dA as a primer. Slot 1 has relatively more markers and slot 3 relatively more influenza cDNA. Influenza bands 1 to 8 are numbered, but band 3 is weak and only just resolves from the mixture of bands 1 and 2 (see (a)3 above).
features not evident in Fig 2a. For example, the relationship between band 8 and the 910 base marker was clearer. Measurements of migration distances in those channels containing both influenza cDNA and the pBR322 markers gave plots such as those shown in Fig 3. Migration distance was not linearly related to the log of molecular weight in any region of the gel, but because of the relatively large number of markers, we were able to compare the sizes of cDNA and markers with confidence. The results are summarised in Table 1. All the cDNA sizes were measured for samples synthesised using p(dT)$_6$A as primer. The size corresponding to the original RNA template was then obtained by subtracting eight bases (the length of the primer complementary to the poly (A) tail) from the estimated size.

**DISCUSSION**

The sizes we have calculated for the RNA segments of influenza A/NT/60/68 by measuring cDNA vary considerably from those previously obtained for related influenza strains by measuring RNA directly. For example, the estimates for the RNA segments of fowl plague virus (5) were

![Calibration curve of mobilities (from Fig 2) of pBR322 marker DNA (in cm) versus log$_{10}$ number of nucleotides (in kB). The arrows identify the mobilities of the 8 influenza cDNA species on the calibration curve.](image)
Table 1: Estimated Sizes for the RNA Genome Segments of Influenza A/NT/60/68

<table>
<thead>
<tr>
<th>RNA SEGMENT</th>
<th>ESTIMATED CDNA SIZE (bases)</th>
<th>CORRESPONDING POLYPEPTIDE</th>
<th>SIZE (DALTONS)</th>
<th>REQUIRED CODING CAPACITY (kB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2390</td>
<td>P3</td>
<td>81-85,000</td>
<td>2210-2320</td>
</tr>
<tr>
<td>2</td>
<td>2390</td>
<td>P1</td>
<td>94-96,000</td>
<td>2560-2620</td>
</tr>
<tr>
<td>3</td>
<td>2290</td>
<td>P2</td>
<td>87,000</td>
<td>2370</td>
</tr>
<tr>
<td>4</td>
<td>1760</td>
<td>Haemagglutinin</td>
<td>64,000</td>
<td>1740</td>
</tr>
<tr>
<td>5</td>
<td>1560</td>
<td>Nucleoprotein</td>
<td>50-54,000</td>
<td>1360-1470</td>
</tr>
<tr>
<td>6</td>
<td>1480</td>
<td>Neuraminidase</td>
<td>45-50,000</td>
<td>1230-1360</td>
</tr>
<tr>
<td>7</td>
<td>1060</td>
<td>Matrix</td>
<td>24-35,000</td>
<td>654-954</td>
</tr>
<tr>
<td>8</td>
<td>890</td>
<td>Non-structural (NS1)</td>
<td>23-27,000</td>
<td>627-736</td>
</tr>
</tbody>
</table>

a. Mean of the figures obtained from three separate experiments similar to those shown in Figs. 2 and 3.
b. The assignment of gene products to specific genes is reviewed in reference 1.
c. Polypeptide sizes (apoproteins only) were measured for several different influenza A virus strains. The values quoted are taken from references 7, 32, 33 and 36-44.
d. From reference 29.
e. Calculated on the basis that 3 bases code for 112 daltons of protein.

greater than our estimates for A/NT/60/68, except for bands 7 and 8, for which the sizes obtained (1080 and 870 bases, respectively) were similar to ours. Estimates for influenza A/PR/8/34 (HONI) gave molecular weights higher than our figures for bands 1-5, similar for band 6, and smaller for bands 7 and 8 (9). We believe that our figures are more likely to be accurate because of the greater degree of denaturation obtained under our electrophoresis conditions, and the use of suitably and accurately sized markers. Nevertheless, potential inaccuracies remain, as the "denaturing gel" may still not be fully denaturing for all species, and anomalous mobilities could occur. Considering errors involved in measuring of autoradiograms, constructing molecular weight versus migration plots and reading off the sizes of genome segments, we estimate that our sizes for bands 4-8 are accurate to within 2.5%. For example, we have determined the length of band 4, the haemagglutinin gene, as 1760 ± 40 nucleotides. Bands
1-3, which run in an area of the gel with lower resolution, are probably accurate to within ± 5%.

Complementary DNA, and its parent RNA from influenza strains A/Mem/102/72 and A/NT/60/68 migrated identically during electrophoresis (results not shown). Desselberger and Palese (9) found that during electrophoresis under denaturing conditions, segments 1-3, 5, 7 and 8 derived from a whole range of influenza A strains migrated identically. Only for segments 4 and 6 was there variation between strains. Therefore, our size estimates of RNA genome segments are likely to apply not only to A/Mem/102/72 and A/NT/60/68, but also, except for segments 4 and 6, to other influenza A strains.

The RNA from our most highly purified preparations (see Methods) of influenza A/NT/60/68 contained two bands in addition to the eight genome segments discussed above. These bands, which we labelled 9 and 10, appeared to be of similar size to the segments 9 and 10 observed for fowl plague virus by Skehel and Hay (28). Corresponding bands were not observed in autoradiographs of cDNA from A/NT/60/68 either because RNA bands 9 and 10 were not copied by reverse transcriptase, or because the cDNA bands were obscured by the high background in this region of the gel. Bands 9 and 10 were never observed in RNA from influenza virus A/Mem/102/72, grown and purified under the same conditions.

A potential criticism of the indirect method of RNA size estimation used in this paper is that the cDNA may not be a full length copy of the parent RNA. Partial copies may be synthesised, due either to degradation of the influenza RNA during the reaction, or to the presence of some inherent block to the progress of reverse transcriptase, for example, a minor nucleotide such as N\textsuperscript{6}-dimethyl A, or some tight "hairpin-loop" structure. Previous work (19) and control experiments (not shown) show that the RNA is essentially undegraded after poly (A) addition. We have no direct evidence that full length cDNA copies are made, other than the similarity between the electrophoretic patterns of RNA and cDNA (Fig 1a). However, previous work with reverse transcriptase demonstrates its ability to consistently synthesise full length copies under optimal conditions. Chang et al. (29) obtained high yields of full length cDNA from human β-globin mRNA, using an oligo dT primer with reverse transcriptase. Similarly, full length copies of the mRNA for rabbit β-globin (30) and chick ovalbumin (31) have been prepared in high yield. However, we have observed that the cDNA yields for the longer influenza RNA segments are
lower than for the small segments 7 and 8, presumably because a higher proportion of the large transcripts are prematurely terminated. The approach we have used to estimate the molecular weights of RNA species should be applicable to any molecules that can be polyadenylated and transcribed with reverse transcriptase.

The polypeptides encoded by mRNA from the individual influenza genome segments have been identified (for a review, see reference 1), and are listed in Table 1. Included in the table is the range of sizes reported for the eight polypeptides, and the number of bases which would be required to encode such polypeptides. The protein sizes, estimated by polyacrylamide gel electrophoresis, are only approximate. Since both nucleoprotein and NS$_1$ are phosphorylated (32) and the nucleoprotein is, in addition, arginine-rich (33), while matrix protein is extremely basic, any of these species might migrate anomalously during electrophoresis. Only for haemagglutinin is any information from amino acid sequencing available (34, and Ward and Dopheide, personal communication). However, even with these approximate figures for protein sizes, it is possible to compare our estimates of the genome segment sizes with the required coding capacities. This should allow us to determine the extent of non-coding sequences present.

Our general conclusion from Table 1 is that most of the genome segments are only slightly larger than the size required for protein coding. For the case where the protein size is known most accurately (band 4 - haemagglutinin), our estimate of genome size (1760 ± 40 bases) suggests a maximum of 60 bases of non-coding sequence. This is similar to the extent of non-coding sequences reported in preliminary investigations of some of the smaller influenza genome segments (19, 28) and suggests that in influenza virus, complex patterns of RNA splicing, such as those found for some non-segmented viruses, may not be observed.

ACKNOWLEDGEMENTS

Dr. G. G. Brownlee is grateful to Dr. G. W. Grigg and to the CSIRO for financial support and encouragement while on leave-of-absence from Cambridge, United Kingdom. We thank Mrs. Elisabeth von Krusenstierna for typing the manuscript.

*Permanent Address: MRC Molecular Biology Laboratory, Hills Road, Cambridge, CB2 2QH, UK
REFERENCES
