Terminal Structure of Reovirus RNAs

By MALCOLM A. McCRAE

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, U.K.

(Accepted 30 March 1981)

SUMMARY

S1 nuclease analysis and 3' terminal sequencing of the reovirus genome double-stranded RNAs and in vitro produced viral mRNAs has been used to establish that the viral mRNA is a full length copy of the genome template. Sequence determination at the 3' end of the genome minus strand has by transposition allowed determination of the 5' terminal sequences of the viral mRNAs. In all but one case an AUG codon which could function in initiation of protein synthesis has been found within the determined sequence. The 3' ends of the plus strands from all genome segments were found to have a common sequence. The implications of these results on the mechanism of virus replication are discussed.

INTRODUCTION

The reovirus genome is composed of ten segments of double-stranded RNA (dsRNA), each of which give rise to a single mRNA and thus a single primary gene product in infected cells (Joklik, 1974; McCrae & Joklik, 1978). Reovirus RNA replication has been shown to occur by a fully conservative mechanism (Acs et al., 1971) with the plus-sense copy of the genome segments functioning as the template upon which new minus-sense copies are synthesized to produce progeny dsRNA molecules. Plus-sense RNA also functions by definition as the mRNA of the virus and as such is found associated with polysomes in infected cells (Stoltzfus et al., 1973). Given this dichotomy in function of the plus-sense RNA it is of interest to try and elucidate the molecular mechanism involved in determining the function(s) of a given plus-sense RNA molecule.

In influenza virus replication a similar problem is overcome by producing two types of plus-sense RNA: a full-length non-polyadenylated species to serve as template in progeny genome production and a 3' end truncated polyadenylated species which acts as mRNA (Hay et al., 1977). To investigate whether or not a similar mechanism operates in reovirus replication, we have studied the termini of both the genome dsRNA species and the mRNA transcripts of such dsRNA produced in vitro by the action of the virion-associated RNA polymerase (Skehel & Joklik, 1969). The method used was 3' terminal labelling with T4 RNA ligase (England et al., 1977) which, when applied to genome dsRNA, gives labelling of both the plus and minus strands. In vitro produced mRNA was also end-labelled and following fractionation of both lots of labelled material into individual segments the termini were studied using S1 nuclease analysis and rapid RNA sequencing procedures (Donis-Keller et al., 1977; Simoncsits et al., 1977; Peattie, 1979).

METHODS

Production of virus, virus dsRNA and virus mRNA. The growth in and purification from mouse L-cells of reovirus type 3 (Dearing strain) using low m.o.i. grown stocks was as previously described (Smith et al., 1969). Extraction of genome dsRNA and the production of...
virus mRNA in the *in vitro* transcription reaction were all carried out as described by McCrae & Joklik (1978).

**3' terminal labelling of dsRNA and mRNA.** This was carried out using T₄ RNA ligase (P & L Biochemicals) and ³²pCp basically as described by England & Uhlenbeck (1978). The reaction mixture (150 µl final vol.) contained either 20 µg mRNA or 40 µg dsRNA, 200 µCi ³²pCp (The Radiochemical Centre, Amersham), 7.5 µmol Hepes buffer pH 8.3, 0.5 µmol dithiothreitol, 1.5 µmol MgCl₂, 0.75 nmol ATP, 10% dimethyl sulphoxide (DMSO), 10% glycerol, and 20 units T₄ RNA ligase. Incubation was at 4 °C for 19 h. Following incubation the reaction was mixed with an equal volume of buffer containing 100 mM-tris–HCl pH 8, 0.2% SDS, and 1 m-NaCl. Fifty µg yeast tRNA was then added and this mixture passed through a 5 ml Sephadex G-50 column to remove unincorporated isotope. The void volume peak was pooled and precipitated with ethanol. The precipitate was washed with 80% ethanol, 100% ethanol and dried by vacuum desiccation.

**Hybridization of 3' end-labelled mRNA species.** The diffuse nature of mRNA bands on polyacrylamide gels makes their fractionation difficult to achieve reproducibly. The genome dsRNA species give much sharper bands on polyacrylamide gels making their individual fractionation easy to achieve. Therefore, prior to their fractionation, 3' end-labelled mRNA species were hybridized to a large excess of unlabelled dsRNA so that the labelled RNA could be converted to the double-stranded form. The dried mRNA pellet was dissolved in 1 mM-EDTA and mixed with 200 µg of virus dsRNA. This mixture was made 90% for DMSO and hybridization carried out according to the method of Ito & Joklik (1972). After hybridization the mixture was ethanol-precipitated, washed with 80% ethanol, 100% ethanol and then dried in *vacuo*.

**Fractionation of dsRNA species and gel elution.** RNA pellets were resuspended in 1 mM-EDTA, mixed with an equal volume of gel buffer (1 × Loening’s E buffer, 8 M-urea, 20% glycerol) heated at 70 °C for 2 min and then electrophoresed for 44 h at 36 mA on a 7.5% polyacrylamide gel as previously described (Schuerch *et al.*, 1975). Individual RNA bands were localized by autoradiography, excised with a scalpel and then labelled RNA eluted electrophoretically as previously described (McCrae & Joklik, 1978).

**Strand separation of end-labelled dsRNAs.** Individual end-labelled dsRNA species in 1 mM-EDTA were mixed with 50 µg unlabelled viral mRNA, and hybridization to achieve displacement of the labelled plus strand was carried out using the method of Ito & Joklik (1972). Single- and double-strand RNA molecules were then fractionated on Whatman CFII cellulose columns as described by Franklin (1966). Fractions containing the separated labelled strands were pooled, carrier tRNA added and ethanol precipitated. In the case of the dsRNA fractions the dried precipitates were taken up in 90% DMSO, denatured at 45 °C for 20 min, and then immediately re-ethanol-precipitated before being used in sequence analysis.

**Sequence analysis of 3' end-labelled RNAs.** In order to reduce the error frequency in the sequencing studies two independent sequencing methods were used. The first involved the use of sequence-specific nucleases as originally described by Donis-Keller *et al.* (1977) and Simoncsits *et al.* (1977). The reactions in this method were as follows. Ladder production was achieved by heating the RNA in 50 mM-bicarbonate buffer pH 9 at 90 °C for 10 min; in the G-specific reaction RNA was digested at 37 °C for 15 min in 100 mM-tris–HCl buffer pH 7.4, 10 mM-EDTA using 5 × 10⁻⁵ units T₁RNase (Sankyo) per µg RNA; in the A specific reaction RNA was digested at 37 °C for 15 min in 100 mM-tris–HCl buffer pH 7.4, 10 mM-EDTA using 5 × 10⁻⁵ units U₂ RNase (Boehringer) per µg RNA; in the C + U reaction RNA was digested at 37 °C for 5 min in 100 mM-tris–HCl buffer pH 7.4, 10 mM-EDTA with 2 × 10⁻⁸ units pancreatic RNase (Worthington) per µg RNA; in the C reaction, RNA was digested at 37 °C for 15 min in 10 mM-EDTA buffer pH 5, 1 mM-EDTA with 1 µl 25 units/ml Physarum RNase I (P & L Biochemicals) per µg RNA. Digestions were done in
either 3 or 5 µl and following digestion samples were mixed with an equal volume of the gel buffer described by Donis-Keller et al. (1977), heated to 100 °C for 1 min and then fractionated on thin 20% polyacrylamide gels as described by Sanger & Coulson (1978). The second sequencing method used was the chemical method of Peattie (1979). Here, reactions were carried out exactly as described by Peattie (1979) with the sole addition of one extra lyophilization prior to gel analysis. Sequencing gels were pre-run at 1.4 kV for 2 h before loading, and fractionation was for 3 or 6 h at 1.4 kV. Gels were autoradiographed at -70 °C using image intensifying screens as described by Laskey & Mills (1975).

RESULTS

3' end-labelling of viral RNAs and sequencing strategy

The recently developed rapid RNA sequencing methods all depend on being able to specifically label the RNA species of interest at one of its termini. Several methods of specific terminal labelling of nucleic acids are now available (England & Uhlenbeck, 1978) and of these we chose to use the ability of T₄ RNA ligase to specifically attach a radioactive nucleotide at the 3' end of RNA molecules. Labelling of reovirus dsRNA at its 3' end in the ligase reaction followed by polyacrylamide gel fractionation gave 10 bands of equal intensity in the pattern characteristic of reovirus type 3 (Fig. 1 a). This indicated that all the dsRNA segments serve as good substrates for the enzyme, although at this stage it was not clear whether the 3' ends of both plus and minus strands were labelled.

To overcome the poor resolution of reovirus mRNAs on polyacrylamide gels we took mRNA that had been incubated in a ligase-labelling reaction and hybridized it to a large excess of unlabelled dsRNA before gel fractionation. The rationale of this step was to incorporate the labelled mRNA into double-stranded molecules which are well resolved on polyacrylamide gels and thereby allow the isolation of individual labelled mRNA species. The results of this procedure (Fig. 1 b) allowed six of the mRNA species to be isolated as individual species with 11, 12 and m1, m2 migrating as unresolved doublets. The reason for incomplete fractionation of mRNAs in a procedure that would be expected to allow all the mRNAs to be isolated is probably the incomplete formation of duplexes in the hybridization step giving rise to single-strand RNA ‘tails’. This ‘tailing’, which would be expected to adversely affect resolution of the dsRNA species, is normally prevented by ‘trimming’ the hybridization products with pancreatic RNase immediately before gel analysis (Ito & Joklik, 1972). However, since we wished to use the final material for sequence analysis it did not seem advisable to use this procedure. That this was probably the correct explanation of our resolution difficulties was shown by the observation that treatment of the large amount of material at the top of the gel with pancreatic RNase under the conditions used in the standard reovirus hybridization procedure converted it all into material giving the standard dsRNA gel pattern (data not shown).

We made the initial assumption that both 3' ends of dsRNA were labelled in the ligase reaction and from this the strategy for the preparation of material for sequence analysis shown in Fig. 2 was devised. The labelled dsRNA was fractionated into individual species on a 7.5% polyacrylamide gel; the individual bands excised, and the RNA eluted electrophoretically. Individual RNA species were then mixed with a large excess of unlabelled reovirus mRNA (10- to 40-fold excess) before hybridization (Ito & Joklik, 1972). The large excess of unlabelled mRNA in the hybridization was designed to displace the labelled plus strand from double-stranded to single-stranded form. After hybridization the mixtures were fractionated on CFII cellulose using the procedure devised by Franklin (1966). This gives one fraction containing the 3' end-labelled plus strand in single-stranded form and a second fraction containing the 3' end-labelled minus strand in double-stranded form. These two
Fig. 1. Gel fractionation of 3' terminally labelled RNAs. (a) Fractionation of end-labelled genome (ds)RNA. This was carried out on a 7.5% polyacrylamide gel as described in Methods. (b) Fractionation of end-labelled mRNA. Prior to gel fractionation as in (a) the labelled ss mRNA was converted to the ds form by hybridization with a large excess of unlabelled genome RNA as described in Methods.

fractions were the substrates for sequencing at the two ends of the genomic dsRNA species. This strategy required both 3' ends to be labelled in the ligase reaction, and that this was the case was shown by the approximately equal counts recovered in each Franklin column fraction for each RNA species (Table 1).

In the case of 3' end-labelled mRNA, following the fractionation of hybridization mixtures on polyacrylamide gels and the electrophoretic elution of individually excised RNA bands the material was ready for immediate sequence analysis. However, both the end-labelled mRNAs and the end-labelled minus strand from genome RNA were purified in the form of dsRNA and would thus be resistant to attack with the nuclease used in one of the sequencing methods employed. This difficulty was overcome by denaturing the dsRNA in 90% DMSO at 45 °C and precipitating with ethanol directly from the DMSO, when the RNA remains in single-stranded form only starting to re-hybridize when the dried RNA pellet is resuspended in a suitable buffer (M. A. McCrae, unpublished observation; see Table 2). Thus, by resuspending dried pellets of denatured RNA directly in the nuclease used for sequencing, the necessary digestion and gel analysis of the products could be achieved before significant re-hybridization occurred.
Table 1. Fractionation of 3' end-labelled genome RNA on CFII cellulose following hybridization with a large excess of unlabelled mRNA*

<table>
<thead>
<tr>
<th>RNA species</th>
<th>Ct/min in 15% EtOH</th>
<th>Ct/min in 0% EtOH</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>1447</td>
<td>1542</td>
</tr>
<tr>
<td>L2</td>
<td>1393</td>
<td>1475</td>
</tr>
<tr>
<td>L3</td>
<td>1499</td>
<td>1420</td>
</tr>
<tr>
<td>M1</td>
<td>1715</td>
<td>1840</td>
</tr>
<tr>
<td>M2</td>
<td>1652</td>
<td>1499</td>
</tr>
<tr>
<td>M3</td>
<td>1964</td>
<td>2013</td>
</tr>
<tr>
<td>S1</td>
<td>2089</td>
<td>1910</td>
</tr>
<tr>
<td>S2</td>
<td>1479</td>
<td>1537</td>
</tr>
<tr>
<td>S3</td>
<td>1810</td>
<td>2095</td>
</tr>
<tr>
<td>S4</td>
<td>2061</td>
<td>1930</td>
</tr>
</tbody>
</table>

* CFII columns were set up and run as described in Methods. Peak fractions from the 15% and 0% EtOH washes were pooled and a small sample taken for scintillation counting.

S1 nuclease analysis

Prior to carrying out direct sequencing studies we analysed the effect of the single-strand-specific S1 nuclease on the various end-labelled RNA species to study their structure. Uniformly labelled reovirus mRNA was found to be approx. 90% sensitive to S1 nuclease (Table 2). A similar sensitivity to S1 nuclease was observed for the 3' end-labelled plus strand of genome RNA after displacement hybridization and purification on CFII cellulose. In contrast to this, when the 3' end-labelled mRNA had been hybridized to an excess of genome dsRNA it was converted to a form that was completely resistant to S1 nuclease and which could be reconverted into a nuclease-sensitive form by denaturation (Table 2). Because of the terminal location of the label in these molecules these results indicated that mRNA was not longer at its 3' end than its template, i.e. it did not contain extra nucleotides added in a
Table 2. *S1 nuclease analysis of various end-labelled RNAs* *

<table>
<thead>
<tr>
<th>Sample no.</th>
<th>Sample description</th>
<th>S1 nuclease resistance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>[3H]UTP-labelled mRNA</td>
<td>12</td>
</tr>
<tr>
<td>B</td>
<td>32P-end-labelled mRNA following hybridization to a 30-fold excess of viral genome dsRNA</td>
<td>103</td>
</tr>
<tr>
<td>C</td>
<td>As for sample B but with DMSO denaturation and ethanol precipitation before S1 nuclease treatment</td>
<td>16</td>
</tr>
<tr>
<td>D</td>
<td>15% EtOH fraction from the Franklin column (end-labelled genome + strand in ss form)</td>
<td>18</td>
</tr>
<tr>
<td>E</td>
<td>0% EtOH fraction from the Franklin column (end-labelled genome — strand in ds form)</td>
<td>97</td>
</tr>
<tr>
<td>F</td>
<td>As for sample E but with DMSO denaturation and ethanol precipitation before S1 nuclease treatment</td>
<td>15</td>
</tr>
<tr>
<td>G</td>
<td>As for sample F but with a 20 min incubation at 37 °C in nuclease buffer before addition of the S1</td>
<td>30</td>
</tr>
</tbody>
</table>

* Digestion with 1000 units/ml S1 nuclease was carried out in 50 mM-sodium acetate pH 4.5, 5 mM-zinc chloride at 37 °C for 30 min. Following digestion nuclease resistance was estimated by comparing the trichloroacetic acid-precipitable ct/min present in the sample with those present in samples incubated without addition of S1 nuclease. The amount (μg) of RNA present in each assay was identical and the total counts used ranged from 10000 to 30000 ct/min.

non-transcriptional process. They did not, however, exclude the possibility that mRNA was a truncated transcript of the genome minus strand. If the mRNA transcript was truncated at its 5' end then the 3' end-labelled minus strand purified from the CFII column should have its label at the end of a single-stranded S1-sensitive tail. However, treatment of the 3' end-labelled minus strand with S1 nuclease showed the label to be completely resistant to nuclease digestion. That this resistance to digestion was not due to either sequence peculiarity or internal 'snap-back' hybridization was demonstrated by the fact that denaturation converted the label into an S1-sensitive form which remained even after allowing a short time for snap-back to occur before adding the nuclease (Table 2).

**Sequence analysis**

The rapid sequencing methods used in this study do not allow the terminal nucleotide to be identified. This terminal nucleotide identification was achieved by carrying out a phosphate transfer experiment. Total alkali digestion of RNA results in the transfer of the 32P label from the added cytidine residue to the 3' terminal nucleotide of the RNA chain used in labelling. The results of this analysis (data not shown) showed that the 3' terminal nucleotide on both plus and minus strands of genome RNA and of mRNA was cytosine which is in agreement with the results of Banerjee et al. (1971).

The results of sequence analysis at the 3' ends of genome plus strands are shown in Fig. 3. The data shown are for the genome segments L3, M3 and S1 but similar sequencing data were obtained for all the other genome species. In our hands the use of the nuclease sequencing method alone would not have allowed unequivocal sequence data to be produced since it generated some anomalies, e.g. in Fig. 3 the assignment of residue 4 as a C was based on the presence of a strong band in the pancreatic RNase digest and a very weak band in the Physarum RNase track; however, there was a little digestion with T1 RNase at this position suggesting residue 4 should be a G, and only the use of the independent chemical sequencing method, which when used alone also generated anomalies, allowed the confirmation of residue 4 as a C. Thus, sequence analysis using the two sequencing methods to cross-confirm results
Fig. 3. Partial nuclease digestion sequence analysis of the 3' termini of the plus strands of species L3, M3 and S1. Nuclease digestions and gel analysis were carried out as in Methods. In this and all subsequent figures showing nuclease sequencing data the following nomenclature has been adopted. L, Partial alkali cleavage track; C, no nuclease digestion control track; T₁, digestion with nuclease T₁ (G-specific); U₂, digestion with nuclease U₂ (A-specific); Pc, digestion with pancreatic RNase (pyrimidine-specific); Ph, digestion with Physarum RNase I (G-, A- and U-specific). The labelling in the control track for each species denotes the sequence assignments for positions 2 to 11 from the 3' terminus.
showed that all species of plus strand had an identical 3' terminal sequence of seven nucleotides, -AAUCAUC, before divergence began between species. Similar sequence analysis when carried out on the isolated mRNA species and on the L1, L2 and M1, M2 pairs showed that they all had the same common seven nucleotides at their 3' terminus as the genome plus strands and that internal to this appropriate mRNA and plus strand sequences matched. Fig. 4 shows the nuclease sequence analysis for mRNA species S3 and S4. Thus, the sequence analysis shows conclusively that the in vitro transcripts of reovirus are not truncated at their 3' end but are full length copies of the template strand.
Reovirus sequencing

Fig. 5. Sequence compilation from the 3' termini of the minus strands of each genome species. The data in this figure and all other sequence information given in this paper were independently determined and cross-checked using both the nuclease and chemical sequencing methods. The notation in parentheses next to some of the species denotes the assignment of the ribosome-protected fragments of Kozak (1977) to these species. In the case of M2 the sequence beyond the UUG was not determined in this study and is taken from the results of Kozak & Shatkin (1976, 1977).

Sequencing of the minus strand of genome RNA was also undertaken using both the partial nuclease digestion and chemical sequencing methods and Fig. 5 gives a compilation of the data obtained for the minus strands of all 10 reovirus RNA species. Fig. 5 also shows by transposition the 5' end of the mRNA that would be transcribed from each of these minus strands, and the amino terminal sequence of the corresponding protein assuming that the first AUG in the chain functions as the initiator of protein synthesis.

DISCUSSION

The S1 nuclease analysis and sequencing carried out at the 3' end of mRNA species and plus-sense strands of genome RNA showed that the mRNA made in vitro, which produces authentic viral polyproteins in in vitro translation systems (Both et al., 1975; McCrae & Joklik, 1978), is not a prematurely terminated copy of the minus-sense template strand. Li et al. (1980b) using a similar sequencing strategy reached the same conclusion for the S2 segment of reovirus type 3. S1 nuclease studies carried out on the 3' end-labelled minus strand strongly suggested that the in vitro produced mRNA was also not a truncated copy of the
template strand at its 5' end. Sequence analysis of the minus strands confirmed this suggestion allowing the assignment of the 5' terminal mRNA sequences obtained in the ribosome binding studies of Kozak & Shatkin (1976) to particular genome segments (Fig. 5). The first conclusion to be drawn from the S1 nuclease and sequencing data is that the mRNA produced in vitro is a full length copy of the minus-strand template. It therefore seems likely that in vivo reovirus will not achieve a distinction between plus-sense RNA molecules for use in RNA replication and those for use in translation by any form of terminal truncation of the mRNA molecules of type seen with influenza (Hay et al., 1977). To date, the only observations concerning structural differences between the two types of plus-sense RNA that could be functionally important are those of Skup & Millward (1980). These authors suggested that 5' terminally capped plus-sense RNA might function in viral replication and uncapped RNA would act as viral mRNA. However, given the known differences in the amounts of plus-sense RNA produced from the various genome segments (Joklik, 1974), it seems unlikely that such a simple difference will alone account for the very precise mechanism operating in genome segment selection (Zweerink, 1974).

The 3' ends of the plus and minus strands from all the genome segments were found to have different common sequences, extending for seven bases -AAUCAUC for the plus strands and only four bases -UAGC for the minus strand. In the case of the plus strand this common sequence presumably has some functional role such as acting as a replicase binding site or being involved in the correct assembly of RNA segments during virion assembly; further work is in progress to investigate this and other possibilities. The shortness of the common sequence on the minus strand means that it cannot completely delineate the transcriptase binding site. Examination of the internal sequences do not reveal any other features of primary sequence which would readily account either for transcriptase binding or the differences in transcriptional frequency of the different segments (Joklik, 1974).

Transposition of the 3' terminal sequences of the minus strand to give the 5' terminal sequences of the various mRNAs confirmed the sequencing studies of Kozak & Shatkin (1976, 1977) and extended them to include data from the L1 to L3 and S1 segments. During the course of this work Darzynkiewicz & Shatkin (1980) published results, obtained using the same sequencing strategy as ours, from the S and M species of RNA which allowed the assignment of the previously sequenced ribosome-protected fragments to particular RNA species. Our assignment of ribosome-protected fragments agrees with that of Darzynkiewicz & Shatkin (1980) and only for the S1 segment is there any major difference in sequence data. In that case our data agree with that of Darzynkiewicz & Shatkin (1980) for only 6 out of 29 bases but show a much closer agreement with the 3' terminal S1 sequence reported by Li et al. (1980a). The reason for the discrepancies is not clear, but all sequence data given in this report were cross-confirmed by two independent sequencing methods.

The derived 5' terminal sequences obtained in this work, in all cases except for L2, contain one AUG which could function as the initiator of protein synthesis. The assignment of polypeptides to genome segments is known for reovirus type 3 (McCrae & Joklik, 1978; Mustoe et al., 1978) and so using the AUG for initiation the sequence data presented in this report allow us to predict the amino terminal protein sequence for nine of the ten primary gene products (Fig. 8). Samuel & Joklik (1976) have reported initiation peptides for four reovirus primary gene products, μ1, μNS, σNS and σ3. Unfortunately none of the amino terminal peptides predicted from this and other sequencing studies agrees with these determined initiation peptides. Therefore, further work is necessary to establish whether the first AUG found on a eukaryote mRNA always represents the initiation triplet as has been proposed by Kozak (1978) in the 'minimum recognition' model, as opposed to internal AUGs operating in the initiation of reovirus proteins as suggested by the results of Samuel & Joklik (1976).

I would like to thank Mrs J. G. McCorquodale for providing excellent technical assistance throughout this work; also Dr G. G. Brownlee for providing explanation for the appearance of
Reovirus sequencing

artefactual weak bands in the nuclease sequencing method. This work was supported by a grant from the Medical Research Council.

REFERENCES


(Received 21 October 1980)