Volume 7 Number 7 1979 Nucleic Acids Research

The nucleotide sequence at the 5' end of foot and mouth disease virus RNA

T.J.R.Harris

The Animal Virus Research Institute, Pirbright, Woking, Surrey GU24 0NF, UK

Received 23 October 1979

ABSTRACT

Foot and mouth disease virus RNA has been treated with RNase H in the presence of oligo(dG) specifically to digest the poly(C) tract which lies near the 5' end of the molecule (10). The short (S) fragment containing the 5' end of the RNA was separated from the remainder of the RNA (L fragment) by gel electrophoresis. RNA ligase mediated labelling of the 3' end of S fragment showed that the RNase H digestion gave rise to molecules that differed only in the number of cytidylic acid residues remaining at their 3' ends and did not leave the unique 3' end necessary for fast sequence analysis. As the 5' end of S fragment prepared from virus RNA is blocked by VPg, S fragment was prepared from virus specific messenger RNA which does not contain this protein. This RNA was labelled at the 5' end using polynucleotide kinase and the sequence of 70 nucleotides at the 5' end determined by partial enzyme digestion sequencing on polyacrylamide gels. Some of this sequence was confirmed from an analysis of the oligonucleotides derived by RNase T_1 digestion of S fragment. The sequence obtained indicates that there is a stable hairpin loop at the 5' terminus of the RNA before an initiation codon 33 nucleotides from the 5' end. In addition, the RNase T_1 analysis suggests that there are short repeated sequences in S fragment and that an eleven nucleotide inverted complementary repeat of a sequence near the 3' end of the RNA is present at the junction of S fragment and the poly(C) tract.

INTRODUCTION

Foot and mouth disease virus (FMDV), like poliovirus and encephalomyocarditis virus (EMCV), is a member of the family Picornaviridae. These viruses contain a single stranded 35 S RNA molecule of about 8000 nucleotides which is of positive polarity and can act directly as a messenger RNA (mRNA). Protein synthesis occurs by the post-translational cleavage of a polyprotein, the synthesis of which is initiated near the 5' end of the RNA (for a review see reference 1). The 3' ends of picornavirus RNAs are polyadenylated and the nucleotide sequence adjacent to the poly(A) has recently been determined for several picornavirus RNAs including that of FMDV (2,3). The 5' ends of picornavirus genomes are covalently linked to a small virus specific protein

termed VPg (4-10). Poliovirus mRNA, however, extracted from infected cell polyribosomes is not linked to VPg but has pUp at the 5' terminus (11,12) and it is likely that all picornavirus mRNAs have a nucleotide diphosphate at the 5' terminus rather than VPg. The presence of the protein on the virus RNA precludes direct polynucleotide kinase mediated labelling of the 5' end with ^{32}P (13) and this probably explains why there is only limited sequence data so far available at the 5' end of picornavirus RNAs (6,7).

FMDV and EMCV RNAs, in contrast to poliovirus RNA, contain a poly(C) tract in addition to the poly(A) (14,15). The poly(C) has been located near the 5' end of the RNAs (10,16,17). Ribonuclease H (RNase H) treatment of FMDV RNA in the presence of oligo(dG) specifically digests the poly(C) and produces two fragments of RNA; the small fragment (S) contains the 5' end of the RNA and the large fragment (L) contains the remainder of the RNA including the poly(A) at the 3' end - see Fig.1 (10). The S fragment of FMDV RNA has been estimated to be about 400 nucleotides long (10) whereas the S fragment of EMCV RNA, produced in the same way, is considerably shorter (c 150 nucleotides, ref.18 and T.J.R.H. unpublished results).

A comparison has been made of the polypeptides synthesised <u>in vitro</u> in a reticulocyte lysate in response to FMDV RNA and fragments L and S (19). Similar proteins are produced when L and full length RNA are translated and no specific protein seems to be made when S fragment is translated. This

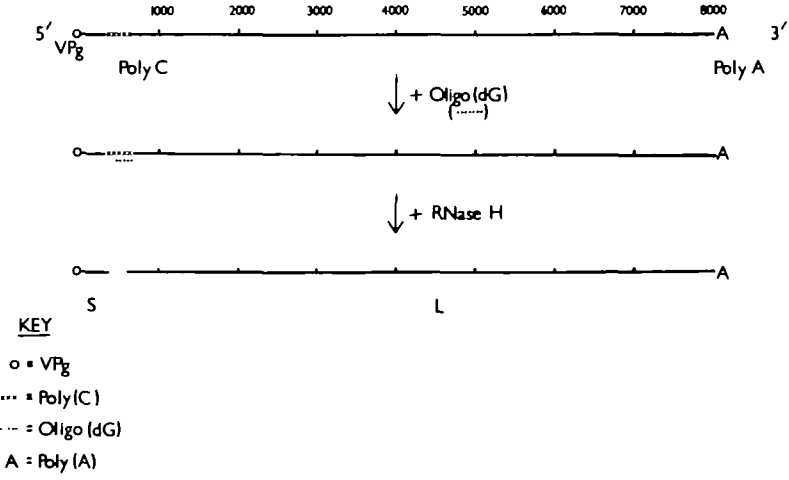

Figure 1. Diagrammatic representation of the digestion of FMDV RNA by RNase H in the presence of oligo(dG_{12-18}).

result and the small size of the S fragment from EMCV RNA suggests that the poly(C) tract and adjacent 5' sequences are not translated (19), although it is possible that this region is translated in vivo but not, under the conditions used, in vitro. This would make the 5' untranslated region of FMDV RNA much longer than that found so far for other eukaryotic and virus mRNAs (20-21), emphasising the need to obtain the nucleotide sequence of the S fragment and the poly(C) tract.

In this paper the purification and preliminary analysis of S fragment is described and the sequence of the 70 nucleotides at the 5' end presented. Some of the structural features that have become evident at this end of the RNA are discussed.

MATERIALS AND METHODS

Preparation of virus RNA.

FMDV type A_{10} (strain A_{61}) was grown in monolayers of 10^8 BHK cells in Roux bottles. Large amounts of virus containing undegraded RNA were prepared from about 4 x 10^9 cells.

The virus was purified from the medium by $(NH_4)_2SO_4$ precipitation and centrifugation through 15-45% sucrose gradients in NTE buffer (0.14M NaCl, 0.02M Tris-HCl, 0.005M EDTA pH 7.6) (10). The RNA was extracted as previously described (10) and purified further by centrifugation on 5-25% sucrose gradients in NTE buffer containing 0.1% SDS (60,000g, 20°C, 16h, Beckman SW42.1 rotor). The peak fractions of RNA sedimenting at 35S were pooled and precipitated at -20°C with 2 volumes ethanol. Using this growth and purification procedure at least 75% of the virus RNA was recovered in the 35S peak.

^{32}P labelled virus was prepared from 10^9 infected cells incubated with 20 mCi carrier free ^{32}P orthophosphate and virus RNA at a specific activity of 10^5 cpm/µg, purified as described above. Labelling of virus RNA with ^3H-uridine was as described (10).

Preparation of 35S infected cell RNA.

Ten Roux of BHK cells were infected at high multiplicity with FMDV and at 1.25h after infection the cells labelled with ^3H-uridine (2 µCi/ml) in 20ml Eagle's medium. At 2.5h after infection the cell sheet was washed with isotonic buffer (0.14M NaCl, 0.01M $MgCl_2$, 0.01M Tris-HCl, pH 7.6) and the RNA extracted from the cell sheets directly by shaking with 10ml phenol and 10ml isotonic buffer. The aqueous phase from these extractions was made 0.5% SDS, extracted twice with an equal volume phenol:chloroform (2:1) and the RNA

precipitated with 2 vol of ethanol overnight at -20°C. RNA sedimenting at 35S, was obtained by sucrose gradient centrifugation and polyadenylated RNA selected by chromatography on 2ml columns of oligo(dT)-cellulose (Boehringer-Mannheim) run as described (22).

Preparation of S fragment.

Ribonuclease H from E. coli was obtained from Enzobiochem (ERN-H) at a specific activity of 2000 U/ml and also as a gift from Dr. J.L. Darlix, Département de Biologie Moléculaire, 30 Quai Ernest-Ansermet, 1211 Genève, Switzerland at a specific activity of 500 U/ml. The standard incubation mixture was developed from that used previously (10) and from the experiment shown in Fig.2. It consisted of 30 µg RNA, 1 µg oligo(dG_{12-18}) (Collaborative Research) and 2-4 units of enzyme added in that order, in 30 µl RNase H buffer (0.01M $MgCl_2$, 0.001M EDTA, 0.01M dithioethreitol, 10% glycerol, 0.05M Tris-HCl pH 7.9). Larger amounts of RNA were digested at the same ratio in proportionately larger volumes. The reaction mixtures were incubated at 30°C for 45 min and stopped by the addition of an equal volume of 0.04M EDTA containing 1% SDS and the RNA precipitated by the addition of 3 vols of ethanol at -20°C overnight. The RNA was dissolved in 25-30 µl of 0.05M Tris-HCl pH 6.8, 0.002M EDTA, 0.2% SDS, 10% glycerol containing 0.02% xylene cyanol FF and bromophenol blue, and heated at 56°C for 5 min before gel electrophoresis.

For protease treatment the RNase H digest was treated with 2.5 µl Proteinase K (Boehringer, 10 mg/ml) for 30' at 37°C before heating at 56°C. The low percentage agarose-acrylamide and the 3-20% gradient slab gels were run and processed as described (23,24). For the preparation of ^{32}P-labelled S fragment or RNA for in vitro labelling, 10% gels were run under the same buffer conditions as the gradient gels. Unlabelled RNA was located by staining with ethidium bromide (24) and eluted from the gel by crushing and soaking in 0.5M NH_4-acetate, 0.001M EDTA, 0.1% SDS. Gel pieces were removed by filtering through sterile nylon wool and RNA precipitated with 3 vols of ethanol. ^{32}P-labelled RNA was detected in gels by autoradiography and eluted in the presence of 25 µg phenol-extracted carrier E. coli tRNA (BDH).

Ribonuclease T_1 fingerprinting.

S fragment prepared from ^{32}P-labelled RNA, or labelled in vitro at the 3' end with ^{32}pCp (see below) was digested with 5 units RNase T_1 (Sankyo) in 10 µl 0.01M Tris-HCl pH 7.6, 0.001M EDTA (TE buffer) at 37°C for 1h in the presence of 25 µg carrier tRNA. 5' end ^{32}P-labelled RNase T_1 oligonucleotides were prepared by polynucleotide kinase labelling. About 1-5 µg of S

fragment in 5 µl TE buffer was digested with 5 µl RNase T_1 (300 units/ml in TE buffer) for 1h at 37°C, 12.5 µl H_2O and 2.5 µl denaturation buffer added, and kinase labelling carried out as described below. Oligonucleotides were separated by two dimensional gel electrophoresis (25) and fingerprints obtained by exposing Fuji RX film to the wet gels. For further analysis, oligonucleotides were eluted from the gels in the presence of carrier RNA and RNase A (Worthington) digestions and DEAE-paper (Whatman) electrophoresis at pH 3.5 done as described before (25). Nucleotide sequence analysis of end labelled oligonucleotides was done as described below, omitting the RNase T_1 digestions.

In vitro labelling.

S fragment was dephosphorylated by incubating with calf intestine alkaline phosphatase (Boehringer) for 30' at 37°C in 0.1M Tris-HCl pH 7.9, 0.01M $MgCl_2$ at an enzyme substrate ratio of 1:20. Ammonium sulphate was removed from the enzyme by Sephadex G100 gel filtration in 0.1M Tris-HCl pH 7.9, 0.01M $MgCl_2$ before use. The phosphatase was removed by phenol-chloroform extraction and the RNA precipitated with 3 vol ethanol at -70°C for 15'. The RNA was reprecipitated from 0.3M Na-acetate pH 5.0 by the addition of 3 vol ethanol (15 min at -70°C) before kinase labelling. For this, about 1-5 µg RNA was dissolved in 10 µl TE buffer and 2.5 µl denaturation buffer (0.05M Tris-HCl, 0.001M spermidine pH 7.9) and 12.5 µl H_2O added. After boiling for 1 min the mixture was quick-cooled and added to 100-150 µCi of dry γ^{32}P-ATP in an Eppendorf microfuge tube. Ten µl of kinase buffer (0.25M Tris-HCl pH 8.5, 0.05M $MgCl_2$, 0.025M dithioethreitol, 25% glycerol), 10 µl H_2O and 5 µl polynucleotide kinase (1000 U/ml New England Biolabs) were added and the mixture incubated for 20 min at 37°C. The reaction was stopped with 120 µl of 2.0M NH_4-acetate containing 1% SDS and 0.02M EDTA. Carrier tRNA (25 µg) was added and the solution extracted once with phenol-chloroform before precipitation of the RNA with 2 vol of ethanol at -20°C.

S fragment (1-5 µg) was labelled at the 3' end with ^{32}pCp using RNA ligase (P-L Biochemicals) by overnight incubation at 4°C (27). The enzyme was removed by phenol-chloroform extraction before the addition of 25 µg carrier tRNA and 3 vol ethanol. All radiolabelled chemicals were purchased from the Radiochemical Centre, Amersham.

Purification of end labelled S fragments was achieved by electrophoresis on 35 cm 10% or 12.5% polyacrylamide slab gels (14) and RNA eluted in the presence of 50 µg tRNA.

For ^{125}I end-labelling, FMDV RNA (100 µg) was treated with 100 µg

Proteinase K in NTE buffer containing 0.1% SDS, phenol-chloroform extracted and repurified by sucrose gradient centrifugation. This RNA was treated with Bolton and Hunter reagent as described for poliovirus RNA (24) except that 0.1% n-lauryl sarcosine was used in place of lithium dodecyl sulphate.

Nucleotide sequence analysis.

End-labelled RNA was sequenced by the partial enzyme digestion procedure (28). RNA was digested with RNase T_1, RNase U_2 (Sankyo) and RNase Phy 1 in 10 µl 0.02M Na-citrate pH 5.0, 7M urea, 0.001M EDTA, 0.025% xylene cyanol and bromophenol blue, at 50°C for 10-15'. The pyrimidine specific B. cereus enzyme (30) digestions were done in 5 µl TE buffer for 10 min at 50°C and 5 µl 10M urea containing the dye markers added after digestion. Limited alkaline hydrolysis was achieved by incubating the RNA in 10 µl 0.1M Na_2CO_3/ $NaHCO_3$, 0.001M EDTA pH 9.0 in a sealed capillary tube at 90°C for 10-15 min (29). The digestion products were analysed on 20% 0.35 mm "sequencing" gels (31) run warm at 1500 volts. Autoradiographs were prepared at -70°C using Dupont Lightning-plus intensifying screens and preflashed Fuji RX film. Wandering spot analyses (30) were done by separating partial alkali digests on two dimensional gels (25). Nuclease P_1 digestions were done in 10 µl of 0.05M Na-acetate pH 6.0 at an enzyme:substrate ratio of 1:10 for 1h at 37°C and the 5' mononucleotides separated by 3MM paper electrophoresis at pH 3.5.

RESULTS

Digestion of RNA with RNase H.

It was shown previously that digestion of FMDV RNA with E. coli RNase H in the presence of oligo(dG) produced two fragments of RNA, S and L, which were subsequently separated by sucrose density gradient centrifugation (10). Ribonuclease T_1 oligonucleotide maps of S fragment prepared in this way showed, however, that there were a number of low molarity oligonucleotides present in addition to the major long oligonucleotides (10). This was probably due to contamination of the S fragment with other small RNAs arising by non specific degradation of the RNA. It was important, therefore, to check the enzyme digestion conditions and to use a different method to follow the production of S fragment. Consequently, ^{32}P-labelled FMDV RNA was digested with RNase H under different conditions and the digests analysed by electrophoresis on either 2% agarose-polyacrylamide or 3-20% gradient polyacrylamide gels (23,24). Fig.2 shows that as found before (10), the presence of oligo(dG) and RNase H is an absolute requirement for the production of S

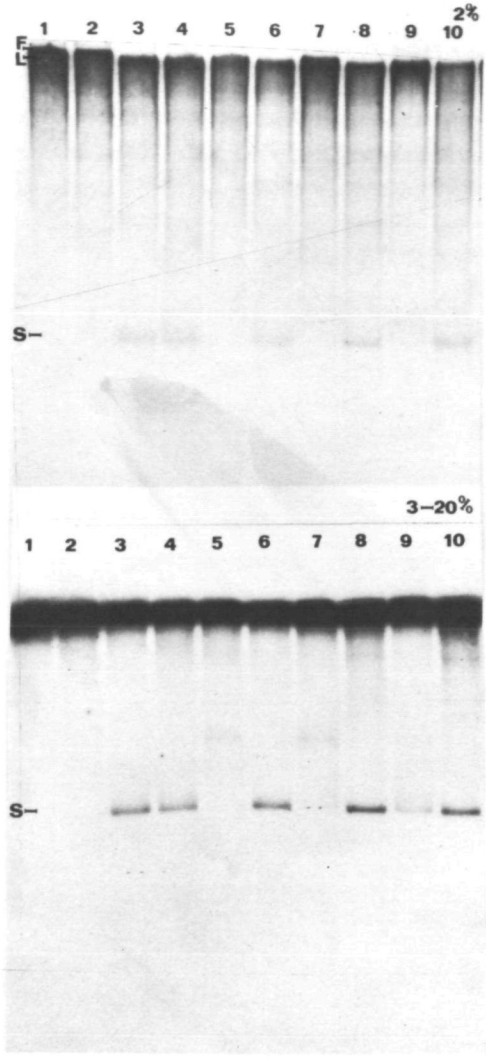

Figure 2. RNase H digestions of ^{32}P-labelled FMDV RNA analysed by 2% (top) and 3–20% (bottom) gel electrophoresis followed by autoradiography. The standard digestion mixture (lanes 3 and 4) consisted of 5 µg RNA, 0.17 µg oligo(dG_{12-18}) and 0.5 units RNase H in 30 µl buffer, incubated for 30 min at 30°C. Lanes 1 and 2 are controls, no oligo(dG) and no enzyme respectively. Other lanes are modifications to the standard digestion mixture. Lane 5, 0.03 µg oligo(dG); lane 6, 1 µg oligo(dG); lane 7, 0.1 units RNase H; lane 8, 2.5 units RNase H; lane 9, 15 min incubation; lane 10, 60 min incubation. F shows the position of full length RNA and L and S the positions of the long and short RNA fragments.

fragment. Both gel systems readily resolve the small S fragment but L fragment is only distinguishable from full length RNA on 2% agarose-acrylamide gels (Fig.2,top,lanes 2 & 3). When the concentration of either the oligo(dG) or the enzyme is reduced (Fig.2, lanes 5 & 7) no S fragment is produced. Increasing the amount of enzyme in the mixture increases the amount of S fragment formed (Fig.2, lane 8) to about the same level as that produced by extending the incubation time (lane 10). Increasing the amount of oligo(dG) (lane 6) does not affect the production of S.

Characterisation of the RNase T_1 oligonucleotides in S fragment.

S fragment was prepared by RNase H digestion of ^{32}P-labelled FMDV RNA and purified by 10% gel electrophoresis (23,24). The RNA was eluted from the gel, treated with RNase T_1 and the oligonucleotides separated by two dimensional gel electrophoresis. Fig.3(A) shows that this fingerprint is considerably clearer than the fingerprint of sucrose gradient purified S fragment (10). Several of the shorter T_1 oligonucleotides which are well resolved in

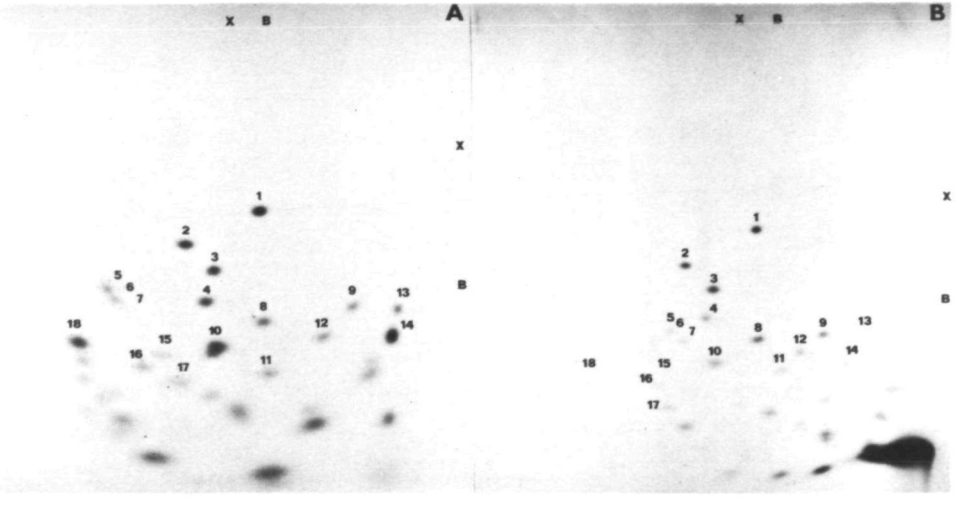

Figure 3. Two dimensional gel fingerprints of the RNase T_1 oligonucleotides of S fragment. (A), oligonucleotides from RNA labelled by growing virus in the presence of ^{32}P; (B), oligonucleotides from RNA labelled *in vitro* at the 5' end with ^{32}P using polynucleotide kinase. Oligonucleotides 1-9 from (A) were analysed further by RNase A digestion, and the same oligonucleotides from (B) by partial enzyme digestion sequencing. X and B are the positions of the xylene cyanol and bromophenol blue dye markers. The origin of the gels is top left.

Fig.3(A) were not clearly identified before, but the three longest oligonucleotides (numbered 1-3 in Fig.3A) have the same mobility as those numbered 7, 4 and 8 respectively in the previous analysis (10). Oligonucleotides 3 and 5, which were of low molarity in the fingerprint obtained before, are not present in Fig.3A indicating that the S fragment analysed previously was contaminated by non-specific RNA fragments. Some of the long T_1 oligonucleotides present in ^{32}P-labelled gel purified S fragment (Fig.3A) were characterised further by RNase A digestion and DEAE paper electrophoresis. The products obtained for each of the oligonucleotides are shown in Table 1.

Analysis of the sequence of the T_1 oligonucleotides.

S fragment, prepared from 50 µg of FMDV RNA, was digested with RNase T_1 and the 5' end of the oligonucleotides labelled with ^{32}P using polynucleotide kinase and the *in vitro* labelled oligonucleotides separated by two dimensional gel electrophoresis. The fingerprint obtained (Fig.3B) is very similar to that of internally ^{32}P-labelled RNA (Fig.3A). The 5' terminal oligonucleotide which is protein linked should not be apparent in Fig.3B but should be present in the fingerprint of internally ^{32}P-labelled RNA (Fig.3A). A comparison of the two fingerprints shows, however, that there is no long oligonucleotide present in Fig.3A that is not present in Fig.3B. As VPg is positively charged at pH 3.5 (A.M.Q. King, personal communication) it is possible that the 5' terminal T_1 oligonucleotide does not migrate towards the cathode in the first dimension (although this depends on the number of nucleotides attached to the protein). At present the absence of the 5' terminal T_1 oligonucleotide from Fig.3A remains unexplained. However, proteinase K treatment of the RNA after RNase T_1 digestion does not alter the fingerprint suggesting that it is not simply due to the presence of VPg.

The longer end-labelled T_1 oligonucleotides (Fig.3B) were sequenced by the partial enzyme digestion method (28). The autoradiographs of the sequencing gels of the four longest T_1 oligonucleotides are shown in Fig.4 and the sequences determined for these and the other T_1 oligonucleotides analysed are shown in Table 1. The sequences shown for oligonucleotides 1 and 2 have been confirmed by 'wandering spot' analyses of partial alkali digests. Moreover, the identity of the 5' terminal base which was difficult to read from the gels was confirmed for each end-labelled oligonucleotide by digestion with nuclease P_1 and identification of the labelled 5' nucleotide monophosphate by paper electrophoresis (not shown). It was noticed from the sequencing gels of oligonucleotides 3 and 4 (Fig.4, Table 1) that RNase U_2 did not cut after Ap when this nucleotide was next to the 3' terminal Gp.

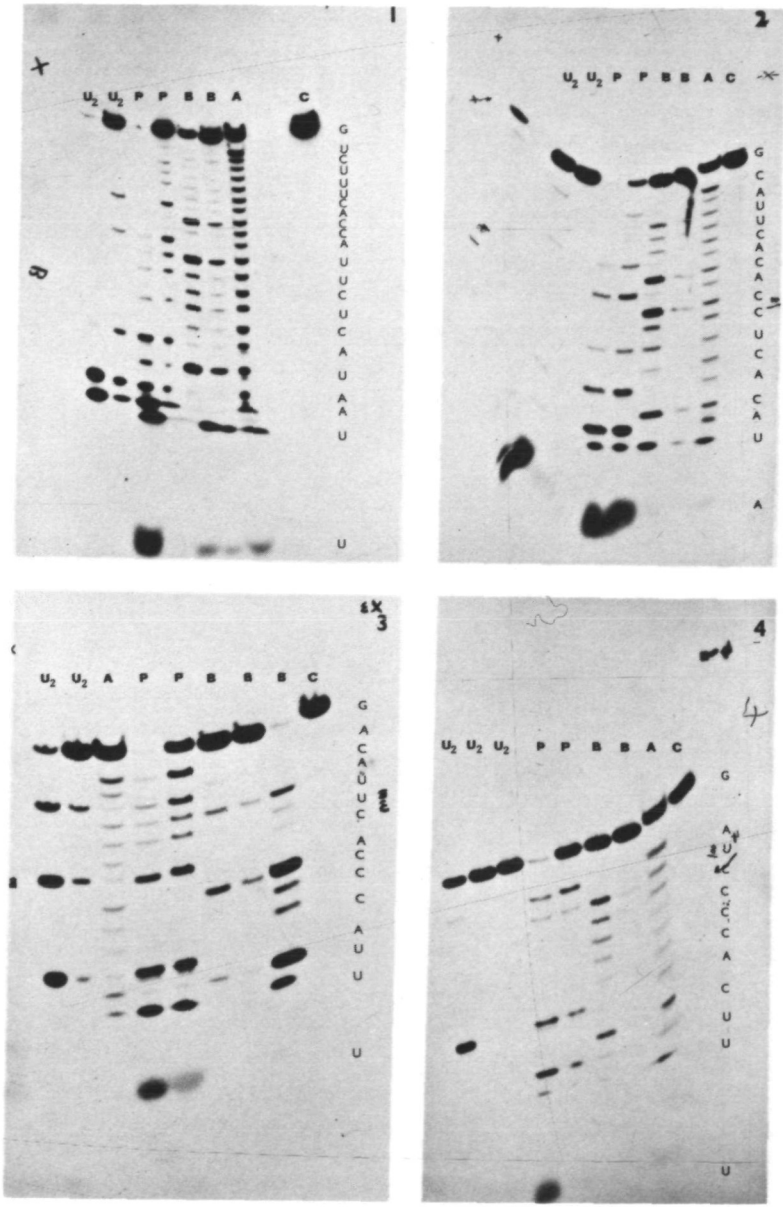

Figure 4. Autoradiographs of 20% polyacrylamide sequencing gels of 5' end-labelled RNase T_1 oligonucleotides 1-4. The partial enzyme *digestions were done as described in Methods (28) and the enzymes* used indicated above each lane; U_2, RNase U_2; P, RNase Phy 1; B, B. cereus RNase; A, alkali ladder; C, control, no enzyme.

However, Ap is clearly the penultimate base in these two oligonucleotides as ApGp is present in their RNase A products (Table 1) and there is no cleavage at this position by the pyrimidine specific B. cereus enzyme (30) (Fig.4). This effect has been noticed before (29). Otherwise the RNase A digestion products were entirely consistent with the sequences obtained (Table 1). There is some evidence from Table 1 that there are short repeated sequences within the T_1 oligonucleotides. The sequence CACUUAC is present in oligonucleotides 2 and 3, the sequence ACCACUUU in oligonucleotides 1 and 8 and UUUCACCCC in oligonucleotides 4 and 5,6,7. The sequences obtained for oligonucleotides 5, 6 and 7 demonstrated that they were related oligonucleotides differing only in the number of cytidylic acid residues at their 3' ends. The fact that they did not terminate in Gp at the 3' end was derived from their RNase A digestion products (Table 1). The sequence obtained for these oligonucleotides is exactly the same as that found at the 5' end of the large oligonucleotide containing the poly(C) tract produced by RNase T_1 treatment of virus RNA (TJRH unpublished results). This shows that oligonucleotides 5, 6 and 7 are derived from the 3' end of S fragment and demonstrates that the S fragment preparation is heterogeneous at the 3' end, molecules differing in the number of cytidylic acid residues remaining at their 3' ends after RNase H digestion.

3' end labelling using RNA ligase.

To check this result directly, S fragment was prepared and the 3' terminus labelled with ^{32}pCp using RNA ligase (27). An apparently homogeneous band of undegraded S fragment was found when the labelled RNA was purified by electrophoresis on a 12.5% polyacrylamide, 6M urea gel (not shown). Complete digestion of this RNA with RNase T_1, however, produced a series of products rather than a single product when analysed by two dimensional gel electrophoresis (not shown). The smaller of these products migrated to the same position in the gels as oligonucleotides 5,6,7 obtained from RNase T_1 digestion of internal or 5' ^{32}P-labelled S fragment (Fig.3). This result confirmed that there was heterogeneity in the number of cytidylic acid residues remaining on the S fragment.

Attempts were made to generate molecules with a unique 3' terminus by altering the RNase H digestion conditions. However, neither increasing the concentration of RNase H nor altering the poly(C)/oligo(dG) hybrid formed (for example, by replacing oligo(dG_{12-18}) with oligo(dG_8)) led to the production of S fragment with a unique 3' end. The reason for this is not clear but is probably due to the fact that the complementary oligo(dG) is

not a complete copy of the poly(C) tract which allows differential hybrid formation and digestion by RNase H. The effect is not apparent when short hetero-oligodeoxynucleotides are used as "complementers" to form hybrid substrates for RNase H (32).

It should be possible to obtain virtually the whole sequence of an RNA molecule the size of S fragment by labelling the ends of the RNA and subjecting each labelled RNA to partial enzyme sequence analysis. As it was not possible to obtain a discrete 3' end on S fragment, attempts were made to obtain RNA uniquely labelled at the 5' end.

Preparation of labelled S fragment from mRNA.

Poliovirus mRNA isolated from infected cell polyribosomes does not have VPg at the 5' end but terminates in pUp (11,12). The nucleotide sequence at the 5' end of poliovirus mRNA, however, is identical to that at the 5' end of poliovirus RNA (38,39). This suggested the possibility of using mRNA from FMDV infected cells as a source of S fragment without a blocked 5' end. Cytoplasmic polyadenylated RNA of about 35S was isolated from infected cells, digested with RNase H in the presence of oligo(dG_{12-18}) and the products analysed by electrophoresis on 10% polyacrylamide gels in the presence of SDS. Fig.5 (lane 1) shows that three RNA fragments were produced from 35S RNA. The slower migrating fragments (A & B) were found to comigrate with S fragment prepared from virus RNA which was resolved into two bands under these conditions (Fig.5, lane 2). The other fragment from the 35S RNA (C) comigrated with S fragment from virus RNA which had been treated with Proteinase K (lane 3). As this protease removes most of the protein from the 5' end of RNA (TJRH and D.V. Sangar, unpublished results) it is probable that band C is derived from mRNA and is not protein linked. Bands A and B which are produced from both 35S RNA and virus RNA probably differ only in the protein attached to their ends as it has recently been found by isoelectric focusing that there are two forms of VPg on virus RNA (TJRH, D.V.Sangar and A.M.Q. King, in preparation). If there is no protein at the 5' end of band C then it should be possible to label S fragment preparations from infected cell RNA with ^{32}P using polynucleotide kinase. Fig.6(A) demonstrates that S fragment derived from mRNA is labelled with ^{32}P *in vitro* whereas that from virus RNA treated identically is not labelled.

Sequencing the 5' end of S fragment.

Labelled S fragment prepared from mRNA, as described above, was eluted from the gel, subjected to complete nuclease P_1 digestion and the 5' nucleotide monophosphates separated by paper electrophoresis. As can be seen from

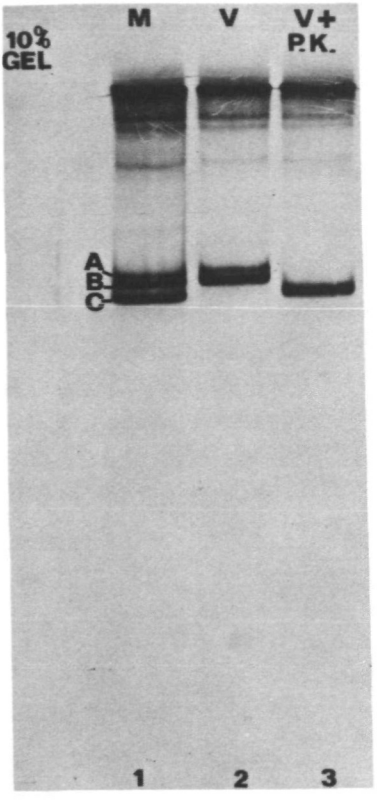

Figure 5. Fluorograph of an SDS polyacrylamide gel (24) of RNase H digests of ^{3}H-uridine labelled 35S infected cell RNA (lane 1) and virus RNA (lane 2). The effect of Proteinase K treatment on the mobility of S fragment from virus RNA is shown in lane 3.

Fig.6(B) only one labelled mononucleotide (pU) was obtained indicating that labelling had occurred at a unique 5' OH site on the RNA. The end labelled RNA was then sequenced by the partial enzyme procedure (28). Autoradiographs of two 20% gels of digests run for various times are shown in Fig.7 and the sequence derived from them in Fig.8. The sequence of about 60 nucleotides can be ascertained from these gels, and the sequence of a further 10-20 nucleotides has been obtained from gels run for a longer time (not shown). The position of the G residues found from the RNase T_1 lanes indicated that two of the long T_1 oligonucleotides that were sequenced before (Fig.4) were present at the 5' end of the S fragment, (oligonucleotides 4 and 2). This

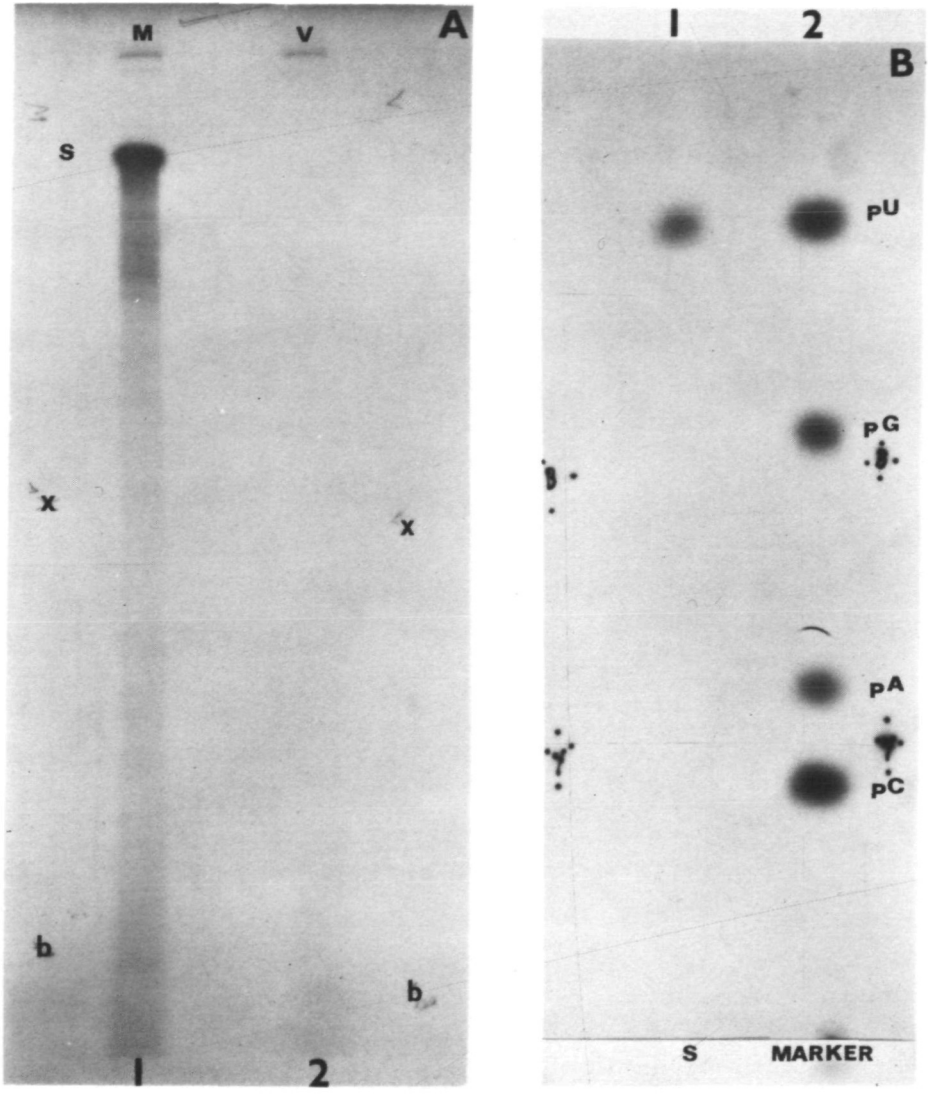

Figure 6. (A) Autoradiograph of a 35 cm 10% polyacrylamide 6M urea slab gel of 5' end labelled S fragment from infected cell RNA (lane 1) or virus RNA (lane 2).

(B) Autoradiograph of a 3MM paper electrophoresis of the 5' mononucleotides obtained by nuclease P_1 digestion of 5' ^{32}P-labelled S fragment (lane 1). The marker (lane 2) showing the position of all four 5' monophosphates is a nuclease P_1 digest of FMDV RNA labelled in vivo with ^{32}P.

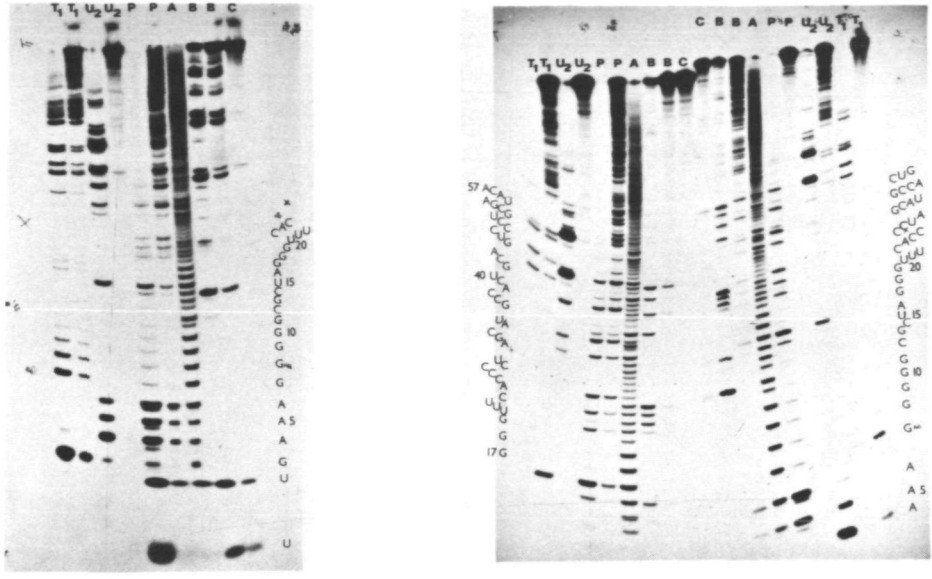

Figure 7. Autoradiographs of 20% polyacrylamide sequencing gels of partial enzyme digests of 5' labelled S fragment. The enzymes used, indicated at the top of each lane are as in the legend to Fig.4. T_1 is the partial digest with RNase T_1. The gel on the left was run for 4h and that on the right for 8h, one batch of digests being loaded 2h before the other.

provided a valuable check on the sequence at positions 19-31 (4) and 52-70 (2) and confirmed the identity of some nucleotides that were difficult to read from the gels. A good example is provided by nucleotides at positions 25-28; these were not cut by any of the enzymes but were cleaved by alkali. The nucleotides were identified as four C residues however, by the earlier unequivocal sequencing of oligonucleotide 4 (see Fig.4). The most difficult bases to distinguish were undoubtedly the pyrimidines. There must remain some doubt about whether nucleotides 36, 39 and 46-48 are C or U residues, owing to poor cutting by the B. cereus enzyme at these positions. However, as these nucleotides were cut weakly by the B. cereus enzyme and not at all by any of the other enzymes, they were designated C residues.

Correlation of mRNA and virus RNA 5' sequences.

It is possible that the 5' end sequence of S fragment derived from RNase H digestion of 35S infected cell mRNA does not correspond to the sequence at the 5' end of virus RNA. A number of observations, however,

make this possibility unlikely. From Fig.8 it is clear that the RNase T_1 oligonucleotide from virus RNA bearing the 5' protein is VPgpUpUpGp; the 5' terminal T_1 product from virus RNA after protease treatment, therefore, would have only 3 nucleotides. This probably explains the apparent absence of a 5' RNase T_1 product from the two dimensional gels (Fig.3). Also, gel isoelectric focusing studies with ribonuclease digests of FMDV RNA labelled in VPg have shown that the RNase T_1 product has only 2-3 additional negative charges compared to VPg prepared by using a mixture of ribonucleases, a result compatible with a sequence VPgpUpUpGp at the 5' end of virus RNA. More direct evidence that the 5' sequence of mRNA and virus RNA are the same is provided by the observation that two of the RNase T_1 oligonucleotides found in S fragment from virus RNA (Fig.3) are represented in the sequence derived from mRNA (Fig.8). Furthermore, preliminary sequencing of FMDV RNA labelled with ^{125}I at the 5' end in VPg using Bolton and Hunter reagent has shown the presence of a run of 4-5 G residues about 6-10 bases from the 5' end (TJRH unpublished results).

DISCUSSION

From our studies on the *in vitro* translation of the L and S fragments of FMDV RNA prepared by RNase H digestion, we concluded that the major

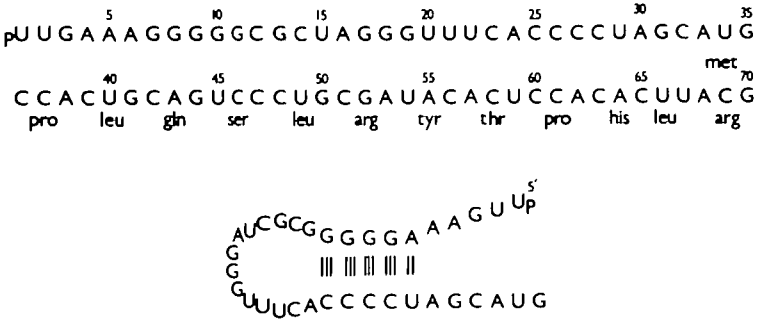

Figure 8. *Nucleotide sequence at the 5' end of FMDV RNA determined from the sequencing gels of 5' end labelled S fragment and the RNase T_1 oligonucleotide sequencing. The potential stable secondary structure that this sequence could form is also shown.

*It is possible that nucleotides 36, 39 and 46-48 are U residues, not C residues. This would make the beginning of the potential amino acid sequence met - ser - leu - gln - phe - leu -

initiation site for protein synthesis is on the 3' side of the poly(C) tract, because L fragment coded for the same products as untreated RNA and S fragment did not code for any recognisable product (19). The sequence at the 5' end of S fragment, however, shows that there is a AUG codon 33 nucleotides from the 5' end of the RNA followed by an open reading frame of at least 12 codons. There is, therefore, the possibility that S fragment does code for a protein despite failing to be translated in vitro.

It is interesting to note that two initiation sites for protein synthesis have been found on poliovirus RNA when translated in vitro in extracts from HeLa cells, one of which predominates when RNA is translated in rabbit reticulocytes (26,37). Whether the poly(C) is translated as part of this separate cistron or whether there is a stop codon before this sequence begins, remains unanswered. There are two observations, however, that have a bearing on this point. Recent studies have shown that although the 3' end of the poly(C) tract, as well as the 5' end, contains bases other than C (TJRH, unpublished results), there is still an uninterrupted run of at least 100 C residues in the tract which means that any protein product would have to contain polyproline, a possibility considered unlikely. Nevertheless, oligonucleotides 5,6,7 which come from the 5' end of the poly(C) tract show no potential stop codon.

The size of the S fragment is also of importance in this regard. It was originally estimated to be about 400 nucleotides long from gel electrophoresis under non denaturing conditions (10). The RNase T_1 fingerprints of purified S fragment (Fig.3) and the sequencing gels (Fig.7) indicate that the size is probably nearer 250–300 nucleotides. No attempt was made to estimate the exact size of S fragment, however, because the complete nucleotide sequence was expected to come from a combination of the 5' and 3' end labelling. However, as has been shown, no long 3' end sequence was obtained owing to size heterogeneity in the S fragment arising from the RNase H digestion. If the S fragment is about 250 nucleotides long it could code for a protein of about 7000 mol. wt. It is interesting to recall our initial speculation (10) that S fragment might code for VPg, a protein of about this size. Experiments are in progress to test this idea more directly by determining the amino-acid sequence at the N terminus of VPg. However, there are some other observations which do not appear to fit this hypothesis. Firstly, S fragment from EMC virus RNA is only about 150 nucleotides long (18, M.M. Chumakov, personal communication) whereas EMCV VPg is apparently larger than FMDV VPg (8). Secondly, a comparison of the RNase T_1 fingerprints of

S fragments prepared from RNA of two different serotypes of FMDV (A and O) (TJRH and KJH Robson, unpublished results) has shown that the sequence at this end of the RNA is not conserved, as might be expected for a virus specific protein with a projected role in RNA synthesis.

As picornavirus RNAs are uncapped mRNAs it is interesting to compare the sequence at the 5' end of FMDV RNA with that of other mRNAs. The most striking feature of the 5' sequence of FMDV RNA is the potential to form a stable secondary structure hairpin loop (Fig.8, ΔG^{o} = -11.2 KCal/mol; 33) ending three bases before the AUG. Although the 5' terminal base is not included in the base paired region, the structure is similar to that recently found at the 5' end of satellite tobacco necrosis virus (STNV) RNA which is also uncapped (34). The fact that C residues 25-28 were not cut by any enzyme in the partial digestions of end labelled S fragment (Fig.7) but were cut by the B. cereus enzyme when sequenced in oligonucleotide 4 (Fig.4) could be explained if they were involved in secondary base pairing. The importance of this hairpin structure as far as ribosome binding is concerned is not yet clear, although it is difficult to reconcile such a stable hairpin loop so close to the AUG with the scanning mechanism of eukaryotic ribosome binding site recognition proposed by Kozak (21). It has been suggested that the hairpin loop in STNV RNA may have an analogous role to the cap structure in other mRNAs and also serve to bring the 5' end and the AUG closer together (33). Experiments are being done to see if S fragment from FMDV RNA will bind to ribosomes in an attempt to answer this question.

Apart from the hexanucleotide UUCACC (positions 21-26) which is present at the 5' end of avian sarcoma virus RNA and ovalbumin mRNA there are no other obvious similarities in the 5' sequence presented here with those of other mRNAs (see 20,21). Although this hexanucleotide is complementary to nucleotides 19-24 at the 3' end of mammalian 18S rRNA (35) it is in a different position in each of the RNAs relative to the AUG and is of doubtful significance.

The other important feature that has become evident from this study is the presence of short repeated regions in S fragment. For example, (Table 1) the octanucleotide ACCACUUU is present in oligonucleotides 1 and 8 and the nonanucleotide UUUCACCCC is present in oligonucleotides 4 and 5,6,7. This latter sequence is present at the 5' end (nucleotides 20-28) and at the 3' end of the S fragment. Sequence reiteration has also been found at the 5' end of tobacco mosaic virus RNA (36), and it may be an element in the evolution of RNA viruses. It is possible, for example, that the poly(C)

Table 1 Sequence analysis of the long RNase T_1 oligonucleotides in S fragment

Spot No.	RNase A products					*Sequence	Length (bases)
1	1 AAU,	2-3 AC,	1G,	2-3C,	>3U	(G)UUAAUACUCUUACCACUUUCUG	(22)
2	3 AC,	1 AU,	1G,	2C,	2U	(G)AUACACUCCACACUUACG	(18)
3	2-3 AC,	1 AG,	2-3C,	high U		(G)UUUACCCACUUACAG	(15)
4	1 AC,	1 AG,	3-4C,	3-4 U		(G)UUUCACCCCUAG	(12)
5)						(G)CUUUUCACCCCCCC	(14)
6 ‡	1 AC,	+6 C,	2-3 U			(G)CUUUUCACCCCCC	(13)
7)						(G)CUUUUCACCCCC	(12)
8	2 AC,	1G,	1C,	some U		(G)ACCACUUUG	(9)
9	2 AU,	1G,	1C,	some U		(G)AUCAUUUG	(8)

‡ For RNase A digestion oligonucleotides 5,6 and 7 were taken together.

*The underlining indicates those sequences which are present in more than one T_1 oligonucleotide.

tract was formed originally by sequence repetition.

The sequence GCUUUUCACCC which lies at the 3' end of S fragment and forms the junction between S fragment and the poly(C) tract is also of importance for another reason. It is an inverted complementary repeat (with one mismatched base) of a sequence near the 3' end of FMDV RNA (GAGUGAAAAGC positions 42-31, Ref.2). These sequences could be involved in secondary structure interactions or have some importance in polymerase binding during replication of positive and negative strands. It is unlikely that an inverted complementary repeat of this length at the ends of the RNA is present by chance.

The extent of these interesting structural features at the 5' end of FMDV RNA will only become clear when the complete nucleotide sequence of S fragment is known.

ACKNOWLEDGEMENTS

I would like to thank Helen Donis-Keller and John Skehel for providing RNase Phy 1, Jamie Robertson for B. cereus RNase and Amiya Banerjee for nuclease P_1. I am also grateful to David Sangar, Donald Black and John Skehel for helpful discussion and to David Rowlands and David Sangar for critically reading the manuscript. I also thank Dr. F. Brown, in whose laboratory these experiments were carried out, for overseeing and facilitating the project.

Note: Further 5' terminal sequence has been obtained by electrophoresis of the 20% gels for longer times. (Autoradiographs of these gels are available on request.) The additional sequence is as follows:

71 75 80
U/UCU/UGU/GUG/CGX
-ser-cys-val-arg

REFERENCES

1. Rekosh, D.M.K. (1977) In: The Molecular Biology of the Animal Viruses. Nayak, D.P. (ed.). Vol.1, Marcel Dekker, New York pp63-110
2. Porter, A.G., Fellner, P., Black, D.N., Rowlands, D.J., Harris, T.J.R. and Brown, F. (1978) Nature 276, 298-301
3. Zimmern, D. and Kaesberg, P. (1978) Proc. Natl Acad. Sci. USA 75, 4257-4261
4. Lee, Y.F., Nomoto, A. and Wimmer, E. (1976) Prog. Nucl. Acid Res. and Mol. Biol. 19, 89-96
5. Lee, Y.F., Nomoto, A., Detjen, B.M. and Wimmer, E. (1977) Proc. Natl Acad. Sci. USA 74, 59-63
6. Nomoto, A., Detjen, B.M., Pozatti, R. and Wimmer, E. (1977) Nature 268, 208-213
7. Flanegan, J.B., Pettersson, R.F., Ambros, V., Hewlett, M.J. and Baltimore, D. (1977) Proc. Natl Acad. Sci. USA 74, 961-965
8. Colini, F., Nomoto, A. and Wimmer, E. (1978) Virology 89, 112-118
9. Sangar, D.V., Rowlands, D.J., Harris, T.J.R. and Brown, F. (1977) Nature 268, 648-650
10. Rowlands, D.J., Harris, T.J.R. and Brown, F. (1978) J. Virol. 26, 335-343
11. Nomoto, A., Lee, Y.F. and Wimmer, E. (1976) Proc. Natl Acad. Sci. USA 73, 375-380
12. Hewlett, M.J., Rose, J.K. and Baltimore, D. (1976) Proc. Natl Acad. Sci. USA 73, 327-330
13. Wimmer, E. (1972) J. Mol. Biol. 68, 537-540
14. Porter, A., Carey, N. and Fellner, P. (1974) Nature 248, 675-678
15. Brown, F., Newman, J., Stott, J., Porter, A., Frisby, D., Newton, C., Carey, N. and Fellner, P. (1974) Nature 251, 342-344
16. Harris, T.J.R. and Brown, F. (1976) J. Gen. Virol. 33, 493-501
17. Chumakov, K.M. and Agol, V.I. (1976) Biochem. Biophys. Res. Comm. 71, 551-557
18. Black, D.N., Stephenson, P., Rowlands, D.J. and Brown, F. (1979) Nucleic Acids Research 6, 2381-2390
19. Sangar, D.V., Black, D.N., Rowlands, D.J., Harris, T.J.R. and Brown, F. (1980) J. Virol. in press
20. Barralle, F.E. and Brownlee, G.G. (1978) Nature 274, 84-87
21. Kozak, M. (1978) Cell 15, 1109-1123
22. Harris, T.J.R. and Wildy, P. (1975) J. Gen. Virol. 28, 299-312
23. Studier, F.W. (1973) J. Mol. Biol. 79, 237-248
24. Harris, T.J.R., Dunn, J.J. and Wimmer, E. (1978) Nucleic Acids Research 5, 4039-4054
25. Harris, T.J.R. and Brown, F. (1977) J. Gen. Virol. 34, 87-105
26. Knauert, F. and Ehrenfeld, E. (1979) Virology 93, 537-546
27. England, T.E. and Uhlenbeck, O.C. (1978) Nature 275, 560-561

28. Donis-Keller, H., Maxam, A.M. and Gilbert, W. (1977) Nucleic Acids Research 4, 2527-2538
29. Rommelaere, J., Donis-Keller, H. and Hopkins, N. (1979) Cell 16, 43-50
30. Lockard, R.E., Alzner-Deweerd, B., Heckman, J.E., MacGee, J., Tabor,M.W. and RajBhandary, U.L. (1978) Nucleic Acids Research 5, 37-56
31. Sanger, F. and Coulson, A.R. (1978) FEBS Letts. 87, 107-110
32. Donis-Keller, H. (1979) Nucleic Acids Research 7, 179-192
33. Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biology 246, 40-41
34. Leung, D.W., Browning, K.S., Heckman, J.E., RajBhandary, U.L. and Clark, J.M. (1979) Biochemistry 18, 1361-1366
35. Alberty, H., Raba, M. and Gross, H.J. (1978) Nucleic Acids Research 5, 425-434
36. Richards, K., Guilley, H., Jonard, G. and Hirth, L. (1978) Europ. J. Biochem. 84, 513-519
37. Brown, B.A. and Ehrenfeld, E. (1979) Virology 97, 396-405
38. Nomoto, A., Kitamura, N., Golini, F. and Wimmer, E. (1977) Proc. Natl Acad. Sci. USA 74, 5345-5349
39. Pettersson, R.F., Flanegan, J.B., Rose, J.K. and Baltimore, D. (1977) Nature 268, 270-272