Comparison of the Nucleotide Sequence at the 5' End of RNAs from Nine Aphthoviruses, Including Representatives of the Seven Serotypes

TIMOTHY J. R. HARRIS

Department of Biochemistry, The Animal Virus Research Institute, Pirbright, Surrey, United Kingdom

The sequence of about 70 nucleotides at the 5' end of the RNAs of nine different aphthoviruses (foot-and-mouth disease viruses), including representatives of the seven serotypes of the virus, has been determined by partial enzyme digestion of ³²P-end-labeled S fragment—that part of the RNA lying to the 5' side of the poly(C) tract and including the 5' end of the molecule. The S fragments were prepared from polyadenylated virus-specific RNA extracted from infected cells by digestion with RNase H in the presence of oligo(dG)₁₂₋₁₈. The first 27 nucleotides from the 5' end were highly conserved in all the RNAs. This region was followed by a more variable region of about 15 nucleotides, showing some length and sequence heterogeneity and including potential but probably nonutilized initiation codons. In agreement with previous homology studies, the sequencing results showed that the European serotypes A, O, and C form a group distinct from the SAT serotypes and that the Asia 1 serotype is closely related to the European group. The lengths of the S fragments of two different RNAs were confirmed as containing 360 to 400 nucleotides by gel electrophoresis with reference to nucleotide markers of known size.

Foot-and-mouth disease virus (FMDV) belongs to the family Picornaviridae (subgroup, aphthovirus) and contains a single-stranded RNA genome of about 8 kilobases. Estimates of the levels of nucleotide sequence conservation between the RNAs from the seven different serotypes of the virus and between the RNAs from subtypes in some of the serotypes have been made by RNA hybridization and RNase T₁ oligonucleotide fingerprinting. By competition hybridization there is about 65% homology between the RNAs of the European serotypes A, O, and C and Asia 1 and a similar level of homology between the RNAs of the South African Territories types, SAT1, SAT2, and SAT3, but a lower level of homology (25 to 40%) between the two groups of viruses. There is a higher (ca. 80%) homology between the RNAs of subtype viruses belonging to the same serotype (10-12). The RNAs from different serotypes of FMDV and different subtypes of the same serotype, however, do not seem to share any long RNase T₁ oligonucleotides apart from the poly(C) tract near the 5' end of the RNA and the poly(A) tract at the 3' end (10, 11; T. J. R. Harris, K. J. H. Robson, and F. Brown, *J. Gen. Virol.*, in press).

Little direct nucleotide sequence information has been obtained for picornavirus RNAs. The sequence of about 50 nucleotides adjacent to the poly(A) tract at the 3' end of the RNAs has been

obtained for several picornaviruses, including five different FMDV serotypes (9), and the sequence of about 20 nucleotides at the 5' end of the RNAs of three enteroviruses, poliovirus 1, poliovirus 2, and coxsackievirus B3, has been compared recently (6). For FMDV, the sequence of 70 nucleotides at the 5' end of the RNA of a type A virus and 55 nucleotides at the 5' end of the RNA of a type O virus has been determined, and the sequences at both ends of the poly(C) tracts located about 400 nucleotides from the 5' ends of the RNAs have been compared (13, Harris et al., in press). These studies show that there is a high level of sequence conservation at the 5' and 3' ends of type A and type O FMDV RNAs. In this paper the comparison of the level of sequence conservation between FMDV RNAs has been extended by analyzing the nucleotide sequence at the 5' end of the RNAs of nine FMDVs, including representatives from the seven serotypes of the virus.

MATERIALS AND METHODS

Viruses. High-titer stocks of the following viruses were prepared by one or two passages in BHK cells: serotypes A10 (A61), O1 (O1 BFS 1848), O6 (OVI), C1 (C Noville), C2 (C997), Asia 1/1 (Pak 1/54), SAT1 (SAT1-82), SAT2 (Tan 2/75), and SAT3 (Bec 1/65). These viruses were obtained originally from the World Reference Centre for Foot and Mouth Disease at the Animal Virus Research Institute, Pirbright, Surrey, United Kingdom (see reference 8 for FMDV classifi-

cation). SAT1-82 was obtained from G. N. Mowat of the Animal Virus Research Institute.

Preparation of ^{32}P -end-labeled S fragment. RNA sedimenting at about 35S was extracted from cells infected with different FMDVs at high multiplicity as described previously (5) except that the RNA was labeled with [^3H]uridine and extracted at different times after infection, depending on the virus used. For A61, OVI and SAT1-82 viruses, cells were labeled at 1.5 h after infection, and RNA was extracted 1.25 h later. For the other viruses, infected cells were labeled at 2 h after infection, and RNA was extracted 2 h later. Polyadenylated RNA was separated from rRNA by oligo(dT)-cellulose chromatography. S fragment was prepared from the bound RNA by RNase H digestion in the presence of oligo(dG)₁₂₋₁₈ and, after purification by gel electrophoresis, treated with alkaline phosphatase and labeled at the 5' end with ^{32}P , using polynucleotide kinase and [$\gamma\text{-}^{32}\text{P}$]ATP. Purification of end-labeled S fragments was achieved by polyacrylamide gel electrophoresis. These procedures have been described in more detail previously (5).

Sizing of S fragments. Restriction fragments of pBR322 DNA for use as size markers (15) were prepared by *Hinf*I digestion. DNA (2 μg) was digested with *Hinf*I (4,000 U/ml; Bethesda Research Laboratories) in 25 μl of 0.06 M Tris-hydrochloride (pH 7.5)–0.06 M MgCl_2 –0.06 M 2-mercaptoethanol–0.1 M NaCl for 2 h at 37°C, and the fragments were dephosphorylated by alkaline phosphatase treatment. After incubation for 30 min at 37°C with 10 U of enzyme, 0.125 ml of TNE buffer (0.14 M NaCl, 0.02 M Tris-hydrochloride, 0.005 M EDTA, pH 7.6) was added, the mixture was extracted once with an equal volume of phenol-chloroform (2:1), sodium acetate (pH 5) was added to 0.3 M, and the aqueous phase was precipitated with 2 volumes of ethanol for 15 min at –70°C. The DNA was 5'-end labeled by incubation with 25 μCi of [$\gamma\text{-}^{32}\text{P}$]ATP and 2 U of polynucleotide kinase in 50 μl of 0.05 M Tris-hydrochloride (pH 8.0)–0.01 M MgCl_2 –0.005 M dithiothreitol–0.1 mM spermidine–0.01 mM EDTA for 30 min at 37°C. The reaction was stopped with 120 μl of 1 M ammonium acetate–0.2% sodium dodecyl sulfate–0.02 M EDTA, and the DNA was precipitated with 2 volumes of ethanol at –20°C overnight. Before loading the DNA onto a 0.35-mm 6% polyacrylamide sequencing gel, the DNA precipitates were dissolved in 20 μl of 80% deionized formamide–0.055 M Tris-borate–0.001 M EDTA–0.02% xylene cyanole FF–0.02% bromophenol blue and heated at 90°C for 1 min. S fragments were dissolved in 20 μl of 0.01 M Tris-hydrochloride (pH 7.6)–0.001 M EDTA–5 M urea–0.02% xylene cyanole–0.02% bromophenol blue before electrophoresis.

Nucleotide sequencing. End-labeled RNAs were partially digested with specific enzymes (RNase T₁ [G]; RNase U₂ [A]; *Physarum* nuclease [U + A > G]; *Bacillus cereus* RNase [U + C]), using a procedure based on that of Donis-Keller et al. (4) as described previously (5). The fragments produced were analyzed on 20, 15, or 13.5% polyacrylamide sequencing gels (40 cm by 0.35 mm) made up in 0.1 M Tris-borate–0.002 M EDTA–7 M urea, pH 8.3, generally using two sample loadings 2 h apart. Autoradiographs of gels were

prepared at –70°C with Fuji RX film and Du Pont Hi-plus intensifying screens.

RESULTS AND DISCUSSION

Size of S fragment in FMDV RNA. The poly(C) tract in FMDV RNA is removed specifically by RNase H in the presence of oligo(dG)₁₂₋₁₈. This digestion produces the S fragment, which contains the 5' end of the RNA and is about 400 nucleotides long, and the L fragment, containing the remainder of the genome (13, 14). S fragment prepared from virus RNA has the genome-linked protein, VPg, attached to the 5' end, whereas a proportion of S fragment prepared from mRNA, isolated from infected cells, does not have the protein (D. V. Sangar, personal communication) and can be specifically labeled at the 5' end by using polynucleotide kinase (5). The presence of the protein alters the mobility of S fragment on polyacrylamide gels (5), and, as previous size estimates were made with S fragment derived from virus RNA, it was important to check the number of nucleotides in S fragments from different RNAs without VPg.

Polyadenylated 35S RNA was isolated from BHK cells infected with type A or type SAT1 FMDV, and ^{32}P -S fragment was prepared by RNase H digestion and in vitro 5'-end labeling (5). Figure 1 shows the mobilities of these S fragments on a 6% polyacrylamide–7 M urea sequencing gel compared with those of ^{32}P -end-labeled denatured *Hinf*I restriction fragments of pBR322 DNA. Both S fragments are between 360 and 400 nucleotides long, that from SAT1 RNA being slightly longer than that from A61 RNA. This agrees with the original estimate of the size of the S fragment from A61 virus RNA and confirms that the S fragment from SAT1 RNA is of a similar size to that from A61 RNA (3, 13).

Sequence at the 5' end of the RNAs from different FMDVs. Polyadenylated RNA sedimenting at 35S was prepared from cells infected with FMDVs representing all seven serotypes and two other FMDVs (subtypes) belonging to serotypes O and C. S fragments were made from each preparation by RNase H digestion and labeled at their 5' ends with ^{32}P , and the RNAs sequenced by the partial enzyme digestion procedure, using 20, 15, and 13.5% polyacrylamide gels to analyze the fragments. Figure 2 shows autoradiographs of two of the gels, one of OVI RNA, showing the sequence from nucleotide 37 to nucleotide 88, and one of SAT1 RNA, showing the sequence from nucleotide 1. The sequences obtained in this way for the nine different FMDV RNAs are compared in Fig. 3. The un-

FIG. 1. Autoradiograph of a 6% polyacrylamide gel comparing the mobilities of denatured ^{32}P -labeled *Hinf*I restriction fragments of pBR322 DNA (lanes a and d) with those of ^{32}P -end-labeled S fragments from SAT1 RNA (b) and A61 RNA (c). The sizes (in nucleotides) of the denatured restriction fragments are indicated (15). The fragments of 154 and 75 nucleotides and the xylene cyanole dye marker have been run off the end of the gel.

derlined nucleotides are not unequivocal, either owing to the difficulty of distinguishing pyrimidines or owing to compression of the nucleotide bands in the gels (see Fig. 2). The sequences of the 55 nucleotides at the 5' ends of A61 and OVI RNAs have been published previously (5; Harris et al., in press), but they are extended here and provided for comparison.

The 27 nucleotides at the extreme 5' end are highly conserved between all the RNAs. The sequences are identical apart from a purine alternative at nucleotide 13, a possible pyrimidine alternative at nucleotide 21, and either CA (in the European and Asia 1 virus RNAs) or UG or AG (in the SAT virus RNAs) at nucleotides 23

and 24. It is conceivable that the complement of this conserved sequence, at the 3' end of the negative strand, forms part or all of the replicase recognition site. The 27 nucleotides at the 5' end of the RNA, however, are more conserved than the nucleotides adjacent to the poly(A) at the 3' end of virus RNA (9), and the nucleotides are not complementary. If the replicase recognizes the sequence adjacent to the poly(A), then this result suggests that the replicase does not recognize identical sites at the 3' end of positive and negative strands. Nevertheless, the 3' end of virus RNA adjacent to the poly(A) is characterized by short conserved runs of pyrimidines (9), and the 5' end of virus RNA has conserved purine runs (e.g., AAA, nucleotides 4 through 6, and GGGG, nucleotides 7 through 11), indicating that pyrimidine tracts are present at the 3' end of negative strands. The sequence at the 5' end of poliovirus 1, poliovirus 2, and coxsackievirus B3 RNA is also conserved and contains purine tracts (1, 6).

The nucleotides which are base paired in the putative secondary structure loop in A61 RNA (5) are largely conserved in all the RNAs. Although the U at nucleotide 29 is not conserved in the SAT RNAs, the A-to-G change at nucleotide 24 makes it possible to shift the base pairing by two nucleotides and still retain a possible stable loop (Fig. 4a). It is also possible to draw alternative stable secondary structures for all the RNAs in which the short conserved C tract is paired with the shorter of the two conserved G tracts (nucleotides 17 through 19). This results in a small, 5- to 7-nucleotide hairpin loop and a stem involving about 10 nucleotides (Fig. 4b). At present there is no direct evidence that any of these structures actually exist. Preliminary evidence from secondary structure mapping of A61 RNA by using nuclease S1 and RNases T₁ and U₂ under nondenaturing conditions (7, 17), however, shows that the nucleotides involved in the possible base pairing are not as accessible to nuclease as those in the loops, particularly for the model in Fig. 4b.

The nucleotide sequence at the extreme 5' end is followed by a run of nucleotides which varies in most of the RNAs studied. This more variable region may be considered as the sequence after nucleotide 27 and before the G(Py)₅G sequence present in all the RNAs (nucleotides 44 through 50 in A61 RNA, underlined in Fig. 3). Although the sequences in this more variable region are clearly related (e.g., the AACGA present in seven of the RNAs and the similarity of the SAT2 and SAT3 RNAs), there are changes present which are not simple nucleotide substitutions. There also appears to be some deletion or

FIG. 2. Autoradiographs of polyacrylamide sequencing gels of (a) OVI RNA from nucleotides 37 to 88, analyzed on a 13.5% gel, and (b) SAT1-82 RNA from nucleotide 1, analyzed on a 15% gel. The enzymes used (two dilutions of each enzyme) are indicated above each lane. T, RNase T₁; U, RNase U₂; P, Physarum nuclease; B, *B. cereus* RNase; A, alkali ladder; C, control, no enzyme. The brackets in (b) indicate regions of band compression, and X and Bp mark the positions of the xylene cyanole FF and bromophenol blue dye markers, respectively.

Virus	10	20	30	40	50	60	70	80	90
A 61	UUGAAAGGGGGCGCUAGGGUUUCACCCUAGCAUGCCACUGCAGUCCCUUGCGAUACACUCCACACUUACGUUCUUGUGUGCGGACCCGA								
O VI	UUGAAAGGGGGCAGCUAGGGUUUCACCCUAGCAGCCACGACAGUCCCUUGCGGUUGCAGUCCACACUUACGUUCUGGACACGCGGGA								
O ₁ BFS	UUGAAAGGGGGCAGCUAGGGUUUCACCCUAGUAAGCUAACGAUAGUCCCUUGCGGUUGCAGUCCACACUUACGUU								
C 997	UUGAAAGGGGGCGCUAGGGUUUCACCCUAGUGUACGACAGUCCCUUGCGGUUGCAGUCCACACUUACG								
C Nov	UUGAAAGGGGGCGCUAGGGUUUCACCCUAGAAAGCCAAAGCAGUCCCUUGCG								
Asia 1	UUGAAAGGGGGCGCUAGGGUUUCACCCUAGUGCCACCGGAGUCCCUUGCGUUCACUCCACACUUAGUGCGACACGUCUGG								
SAT 1	UUGAAAGGGGGCAGCUAGGGUUUGCCAGUUCGCGUAACGACCGUCCCUUGCGAUUCAUUCGGA								
SAT 2	UUGAAAGGGGGCAGCUAGGGUUUGCCUAGUGUGCGGUACGACCGUCCCUUGCGUUGAAACGCUACUUGCUCAGA								
SAT 3	UUGAAAGGGGGCAGCUAGGGUAGCCUAGUGUGCGGUACGACCGUCCCUUGCGUUGAAACGCUACUACUCAGA								

FIG. 3. Comparison of the sequence at the 5' end of nine FMDV RNAs. The box surrounds the variable region referred to in the text, and broken underlining indicates a sequence obtained from a region of band compression. Underlined (—) nucleotides are pyrimidines for which the designation as C or U is less certain than for other pyrimidines. The underlined (—) sequence of seven nucleotides is the G(Py)₅G sequence referred to in the text. The sequence of 81 nucleotides at the 5' end of A61 RNA and 55 nucleotides at the 5' end of OVI RNA has been published before (5; Harris et al., in press), but the sequences are extended here and are provided for comparison.

FIG. 4. Examples of potential stable secondary structures that could be present at the 5' ends of FMDV RNAs. (a) Structures where the conserved Cs are paired with the Gs in the longer G tract (nucleotides 7 through 11). The structure for A61 RNA is from reference 5 and is included for comparison. (b) Structures where the conserved Cs are paired with Gs in the short G tract (nucleotides 17 through 19). Similar structures to those shown in (a) and (b) can be drawn for all RNAs. Stabilities were assessed by using the modified Tinoco rules (16) and varied from $\Delta G = -4 \text{ kcal}$ (-16.8 kJ) (C Noville RNA) to $\Delta G = -14.0 \text{ kcal}$ (-58.7 kJ) (Asia 1 RNA) for model (a) and from $\Delta G = -9.2 \text{ kcal}$ (-38.5 kJ) (C997 RNA) to $\Delta G = -26.4 \text{ kcal}$ (-110.6 kJ) (A61 RNA) for model (b).

addition of nucleotides, and this occurs even within subtype RNAs (e.g., C997 and C Noville RNAs). Some of this apparent deletion may be due to difficulties of analysis caused by band compression, which was a feature of the sequencing gels for this region of C Noville and SAT1 RNAs (Fig. 2). However, this compression could only have led to the omission of one or two nucleotides (if any) from these sequences and certainly would not account for all the length and sequence heterogeneity observed in this part of other RNAs (e.g., Asia 1 RNA).

The AUG present in A61 RNA lies in this more variable region and is not a conserved feature of the other RNAs. There is GUG present in this part of four other RNAs (C997, Asia 1, SAT2, and SAT3), but the frame of translation (and possible amino acids coded) after these potential initiation codons is not the same for each RNA, making it unlikely that they represent functional initiation sites, even though termination codons are not detected in phase sub-

sequently. The fact that the L fragment of FMDV RNA produced by RNase H digestion gives the same in vitro translation products on polyacrylamide gels as full-length untreated FMDV RNA lends support to the view that S fragment is not translated (14). If this is so, then FMDV RNA has about 550 nucleotides [S fragment plus the poly(C) tract] before the initiation site for protein synthesis.

The sequence after the more variable region in the European virus RNAs and the Asia 1 virus RNA is again highly conserved, but bears only superficial resemblance to this part of the SAT2 and SAT3 virus RNAs. The SAT1 virus RNA sequence has more in common with the European virus RNA sequence here, but it is clearly like those of SAT2 and SAT3 RNAs in other respects [e.g., the UG at nucleotides 23 and 24 and the CGUAACGACC before the G(Py)₅G sequence]. These results are in general agreement with previous homology studies in this laboratory, which showed that the Euro-

pean and SAT virus RNAs formed two groups and that Asia 1 virus RNA was more closely related to the European group (12). However, the sequence at the 5' end of the RNA of viruses belonging to the same European serotype is no more conserved than that at the 5' end of RNA from the other European serotypes, the differences being confined mostly to the variable region (compare the sequence of C Noville RNA with that of C997 RNA and the sequence of O1 BFS 1848 and that of OVI RNA in Fig. 3).

The homology observed between the European virus RNAs from this sequence work (ca. 90%) is considerably higher than the overall homology between the RNAs measured by competition hybridization (12). This probably means that the 5' end of FMDV RNA is more conserved than the remainder of the genome, although experiments comparing the sequences of other noncoding regions of the RNAs also show a higher level of homology than predicted by hybridization tests (9; Harris et al., in press). A comparison of the amino acid sequence at the NH₂ terminus of one of the structural polypeptides (VP1, the trypsin-sensitive polypeptide) of three European serotypes of FMDV shows that there is about 85% conservation in the first 26 amino acids (2). Whether this is a reflection of as high a level of nucleotide sequence conservation as apparently occurs in the noncoding regions of the RNA remains to be determined.

ACKNOWLEDGMENTS

I thank Helen Donis-Keller, John Skehel, and Jamie Robertson for providing some of the sequencing enzymes using in this study, John Boothroyd for a gift of pBR322 DNA and *Hinf*I restriction endonuclease, and Trevor Collen for drawing Fig. 3 and 4. I am also grateful to David Sangar, David Rowlands, Peter Lowe, and John Boothroyd for helpful discussion and to F. Brown for a review of the manuscript.

LITERATURE CITED

1. Babich, A., E. Wimmer, H. Toyoda, and A. Nomoto. 1980. The genome linked protein of picornaviruses. VI. The 5' terminal protein of poliovirus type 2 RNA is covalently linked to a nonnucleotide identical to that of poliovirus type 1 RNA. *Intervirology* 13:192-199.
2. Bachrach, H. L., D. O. Morgan, and D. M. Moore. 1979. Foot and mouth disease virus immunogenic capsid protein VPT: N terminal sequences and immunogenic peptides obtained by CNBr and tryptic cleavages. *Intervirology* 12:65-72.
3. Black, D. N., P. Stephenson, D. J. Rowlands, and F. Brown. 1979. Sequence and location of the poly(C) tract in aphtho- and cardiovirus RNA. *Nucleic Acids Res.* 6:2381-2390.
4. Donis-Keller, H., A. Maxam, and W. Gilbert. 1977. Mapping adenines, guanines and pyrimidines in RNA. *Nucleic Acids Res.* 4:2527-2538.
5. Harris, T. J. R. 1979. The nucleotide sequence at the 5' end of foot and mouth disease virus RNA. *Nucleic Acids Res.* 7:1765-1785.
6. Hewlett, M. J., and R. Z. Florkiewicz. 1980. Sequence of picornavirus RNAs containing a radioiodinated 5' linked peptide reveals a conserved 5' sequence. *Proc. Natl. Acad. Sci. U.S.A.* 77:303-307.
7. Pavlakakis, G. N., R. E. Lockard, N. Vamvakopoulos, L. Rieser, U. L. RajBhandary, and J. N. Vournakis. 1980. Secondary structure of mouse and rabbit α - and β -globin mRNAs: differential accessibility of α and β initiator AUG codons towards nucleases. *Cell* 19:91-102.
8. Pereira, H. G. 1977. Subtyping of foot-and-mouth disease virus. *Dev. Biol. Stand.* 35:167-174.
9. Porter, A. G., P. Fellner, D. N. Black, D. J. Rowlands, T. J. R. Harris, and F. Brown. 1978. 3'-Terminal nucleotide sequences in the genome RNA of picornaviruses. *Nature (London)* 276:298-301.
10. Robson, K. J. H., J. R. Crowther, A. M. Q. King, and F. Brown. 1979. Comparative biochemical and serological analysis of five isolates of a single serotype of foot and mouth disease virus. *J. Gen. Virol.* 45:579-590.
11. Robson, K. J. H., T. R. Doel, B. M. Gorman, and F. Brown. 1980. Biochemical aspects of variation in foot and mouth disease virus. *J. Gen. Virol.* 46:179-193.
12. Robson, K. J. H., T. J. R. Harris, and F. Brown. 1977. An assessment by competition hybridisation of the sequence homology between the RNAs of the seven serotypes of FMDV. *J. Gen. Virol.* 37:271-276.
13. Rowlands, D. J., T. J. R. Harris, and F. Brown. 1978. More precise location of the polycytidylic acid tract in foot and mouth disease virus RNA. *J. Virol.* 26:335-343.
14. Sangar, D. V., D. N. Black, D. J. Rowlands, T. J. R. Harris, and F. Brown. 1980. Location of the initiation site for protein synthesis on foot-and-mouth disease virus RNA by in vitro translation of defined fragments of the RNA. *J. Virol.* 33:59-68.
15. Sutcliffe, J. G. 1978. pBR322 restriction map derived from the DNA sequence: accurate DNA size markers up to 4361 nucleotide pairs long. *Nucleic Acids Res.* 5:2721-2728.
16. Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla. 1973. Improved estimation of secondary structures in ribonucleic acids. *Nature (London) New Biol.* 246:40-41.
17. Wurst, R. M., J. N. Vournakis, and A. M. Maxam. 1978. Structure mapping of 5' ³²P-labeled RNA with S₁ nuclease. *Biochemistry* 17:4493-4499.