Volume 15 Number 17 1987 Nucleic Acids Research

Potential secondary and tertiary structure in the genomic RNA of foot and mouth disease virus

B.E.Clarke, A.L.Brown, K.M.Currey[1], S.E.Newton, D.J.Rowlands and A.R.Carroll

Wellcome Biotechnology Limited, Ash Road, Pirbright, Surrey, UK and [1]Laboratory of Mathematical Biology, Division of Cancer Biology and Diagnosis, Building 10, National Cancer Institute, Bethesda, MD 20205, USA

Received June 11, 1987; Revised and Accepted August 3, 1987

ABSTRACT

The nucleotide sequence of the 5' untranslated region of foot and mouth disease virus (FMDV), serotype A_{10} has been determined. This completes the first total genomic sequence for any one serotype of FMDV. Analysis of the sequence to the 3' side of the poly (C) tract reveals the presence of a 24 nucleotide repeated motif which has homologies with a sequence located upstream of the transcriptional initiation site from several mammalian fibrinogen genes. The function of this element in FMDV is unclear. However, computer analysis of this region predicts the presence of a high degree of secondary and tertiary structure in which these repeats form an important part. The implications of these predictions are discussed.

INTRODUCTION

Foot and mouth disease virus (FMDV) is a member of the family Picornaviridae having a positive sense single stranded RNA genome of approximately 8500 bases in length (3). The genomic organization of all picornaviruses is similar, each being polyadenylated at the 3' end and having a viral coded protein, Vpg, covalently linked to the 5' end. Each RNA has a long untranslated region (741 bases for poliovirus (17)) followed by an open translational reading frame coding for the viral polyprotein terminating about 100 bases from the 3' end. This protein is then processed into specific viral proteins by viral or possibly host coded proteases (22). FMDV and the cardioviruses have a more complex genomic structure due to the presence of a tract of 100-150 cytidine residues located approximately 400 and 150 bases respectively from the 5' ends of the genomes (3, 35). Initiation of translation in FMDV occurs approximately 700 nucleotides downstream from the poly (C) tract resulting in a total untranslated region of about 1200 nucleotides. Furthermore, FMDV appears to be distinct from all other picornaviruses in that it initiates translation at two different sites producing two related leader proteins (Lab and Lb) (9, 11, 32).

© IRL Press Limited, Oxford, England.

Nucleotide sequencing of FMDV has concentrated largely on those regions coding for the viral structural proteins (28, 29, 30) or the viral polymerase (27). However for three distinct strains, A_{10}, A_{12} and O_1, more extensive sequence data has been reported (23, 27, 33). For each virus the complete nucleotide sequence representing the viral polyprotein has been reported while substantial sequence representing the untranslated region to the 3' side of the poly (C) tract has only been reported for the O_1 and A_{12} viruses. Conversely the only available data for the sequence to the 5' side of the poly (C) tract derives from the A_{10} and the O_1 BFS viruses (3). Computer analysis of this area of the genome predicted a high degree of secondary structure. Furthermore, recent studies with other members of the Picornaviridae have suggested the importance of secondary structure at the 5' end of the genome in terms of attenuation, neurovirulence, etc, (13, 14, 26). In this paper we complete the first total genomic sequence for one strain of foot and mouth disease virus (A_{10}) and examine the entire 5' untranslated region for predicted structural features.

MATERIALS AND METHODS

Preparation of Virus and Virus RNA.

Virus particles were purified and the RNA extracted as previously described (1). The viruses used were strains A_{10} 61 and A_{12} (USA).

Sequencing Procedures.

Sequencing was carried out by the primer extension dideoxy method directly from the virus RNA (2, 3), using 2μg viral RNA and synthetic deoxynucleotide primers as indicated in Figure 1. Confirmatory sequence was generated by shotgun cloning into M13 of double stranded cDNA from this region which had been synthesised using the same primers (4).

Computer Analysis

Sequence data from areas of FMDV RNA were analysed for secondary structure using a modification of the FOLD program (6) as described by Currey et al (5). Further analysis for higher order structures within the genome was based on the principle of pseudoknotting as described by Pleij et al (7, 8).

Oligonucleotide Synthesis.

Oligonucleotides were synthesized using an Applied Biosystems machine (model 381A) using cyanoethyl phosphoramidites. Crude oligonucleotides were purified on polyacrylamide gels and desalted on Sephadex G-10 columns.

RESULTS AND DISCUSSION.

Comparative Sequence Analysis of viruses A_{10}, A_{12} and O_1K.

The nucleotide sequence of the FMDV genome between the poly (C) tract and the sites for initiation of translation has previously been reported for the O_1 Kaufbeuren serotype (23). This region was found to contain 724 nucleotides and, together with the 367 nucleotides which we had shown to be present between the 5' end of the genome and the poly (C) tract of the O_1 BFS strain (3), as well as the poly (C) tract itself, this gives a total 5' untranslated region for FMDV of approximately 1240 nucleotides. The presence of such a long leader sequence in FMDV is intriguing. We have previously shown (9, 11) that for a limited portion of the untranslated sequence near the initiation sites, more homology exists across all seven serotypes of FMDV than within the 5' coding portion of the genome. We were therefore interested to compare the complete untranslated region sequence between two serotypes (A_{10} and O_1 Kaufbeuren). Partial sequence analysis was available for another virus of the A serotype (A_{12} USA) (10) which enabled us to complete this sequence for a more valid comparison between three distinct viruses.

Initially, oligonucleotide primers were synthesized which were complementary to the RNA adjacent to the translational initiation sites in FMDV. Direct dideoxy sequencing from the RNA was then carried out by the primer extension method. Successive cycles of primer synthesis and dideoxy sequencing resulted in the completion of the sequence through the untranslated region into the poly (C) tract. Sequencing of the poly (C) tract itself revealed no discontinuity. Confirmation of the direct sequences for all except 80 bases (-627 to -517) of A_{10} was obtained from cloned fragments. The sequence of this region for the three viruses is shown in Figure 1. The A_{10} sequence completes the first total genomic sequence for any one serotype of FMDV. The sequence we obtained for A_{12} USA differs in minor detail from the published data (10); these changes have been checked for our seed of this virus, obtained from Dr D M Moore, by subcloning the corresponding cDNA fragments into M13 and sequencing the resultant clones. The most obvious feature of the comparison of the three viruses was a 36 base deletion between nucleotides -588 and -552 in A_{10}. This would seem to be further evidence to indicate that this region of the genome does not code for protein.

Forss et al (23) reported the presence of a potential translational open reading frame in the untranslated region beginning at the AUG at position

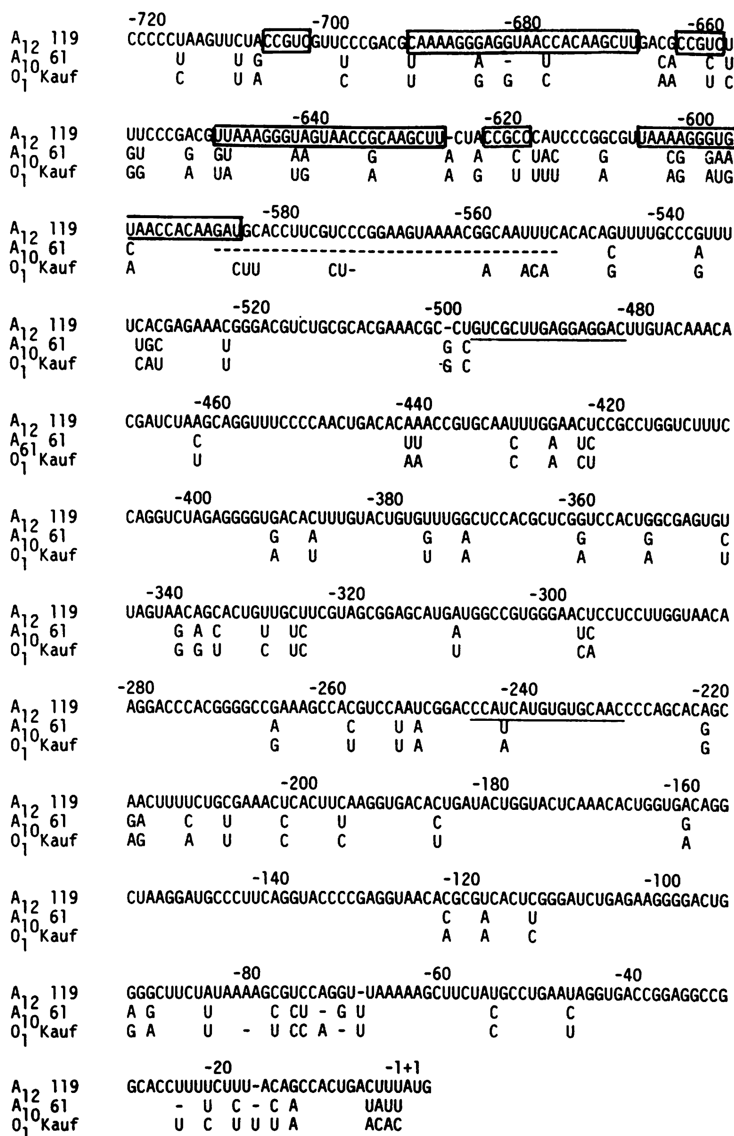

Figure 1
The complete nucleotide sequence for the untranslated region of FMDV to the 3' side of the poly (C) tract. Numbering is indicated relative to the AUG codon for the larger leader protein (Lab) as previously described (9, 11). The sequence for serotype O_1 K has been previously reported (23) and the sequence from -630 to +1 has also been reported for a similar isolate of A_{12} 119 (10; see text). Oligonucleotide primers used for primer extension sequencing are underlined. The repeated regions (see text) are indicated by the solid boxes.

-600 (Figure 1) for O_1 Kaufbeuren, which could code for a 10K polypeptide. This initiation codon is, in a sequence context, similar to the authentic initiation regions in most picornaviral RNAs being preceded at a short distance by a pyrimidine rich region. As can be seen in Figure 1 the absence of the codon in both A_{10} and A_{12} viruses suggests that the open reading frame is not of significance for viral replication. However, Forss et al (23) reported a preliminary observation that a P10 protein could be precipitated from in vitro translation products of O_1 Kaufbeuren RNA using antiserum elicited by to a synthetic peptide based on the predicted translation product of the P10 open reading frame. It is unclear whether the P10 protein is a normal product of virus replication or is simply the result of an abnormal translation initiation event in vitro.

We have previously shown that the level of nucleotide homology between the A_{10} and the A_{12} viruses in the region coding for the leader proteins is 88% (9). Analysis of the degree of homology between the three viruses in the 5' untranslated region downstream of the poly (C) tract shows a remarkable degree of conservation (A_{10} : A_{12}, 90%; A_{12} : O_1K, 90%; O_1K : A_{10}, 89%). Such a high degree of homology for a noncoding sequence infers an important structural or functional role for this region.

Detailed analysis of the sequence 3' of the poly (C) tract reveals the presence of a 24 nucleotide direct repeat interspersed with a second, smaller, five nucleotide repeat (Figure 2). This sequence, although not totally conserved, is found three times in each viral genome but the third repeat for the A_{10} genome shows more differences since it overlaps the deleted region. This interesting feature might give us an insight into the possible function(s) of such a long untranslated leader sequence in FMDV RNA. Screening the EMBL GenBank database showed that the repeat had only been found in one other situation, the genes coding for both rat and bovine β-fibrinogen (12, 24). The location of the repeated sequence in these genes is approximately 100 bases upstream of the transcriptional initiation site but has not been associated with any particular function although activity as a polymerase binding site is an interesting speculation.

Computer Predictions of Secondary and Tertiary Structure.

We have previously shown (3) that it is possible to predict a high level of secondary structure in the FMDV sequence 5' to the poly (C) tract. Similarly it has been shown that the secondary structure of the 5' untranslated region of other picornaviruses is crucially important in terms

RAT β FIBRINOGEN		AAACUGUCAAAUAUUAACUAAAGGGAGGUAACCAUUCCUGAA		
A10	(1)	CCGUC(A)	10	GUAAAAGGAAGUAAUCACAAGCUU (B)
A10	(2)	CCGCC(C)	10	GUAAAGGGAAGUAACCGCAAGCUU (D)
A10	(3)	CCGCC(E)	10	UUAACGGGGAACAACCACAAACACACUUUU (F)
A12	(1)	CCGUC	10	CAAAAGGGAGGUAACCACAAGCUU
A12	(2)	CCGUC	10	UUAAAGGGUAGUAACCGCAAGCUU
A12	(3)	CCGCC	11	UAAAAGGGU-GUAACCACAAGAU
0 Kauf	(1)	CCGUC	10	UAAAAGGGAGGUAACCACAAGCUU
0 Kauf	(2)	CCGUC	10	UAAAAGGGUGGUAACCACAAGCUU
0 Kauf	(3)	CCGUC	10	UUAAAGGGAUGAAACCACAAGACUU

Figure 2
Comparison of the repeated motifs found in three serotypes of FMDV with a homologous sequence from rat β fibrinogen (12). FMDV sequences are derived from Figure 1 and conserved nucleotides are indicated by bold type. Numbers indicate the number of nucleotides present between the repeated elements in each case. Letters in brackets correspond to those in Figure 4.

of phenotypic characteristics such as attenuation, neurovirulence and temperature sensitivity (13, 14, 26). With the completion of the genomic sequence of FMDV A_{10} we have now been able to perform more detailed secondary structural analysis for the whole 5' untranslated region of one serotype of FMDV. The predictive method used was as described by Currey et al (5) and the resultant structures are shown in Figures 3 and 4.

The two stretches of heterologous sequence seem to form two discrete regions of secondary structure separated by the poly (C) tract. This tract is not itself involved in any RNA/RNA interaction as experimentally shown by Mellor et al (15). The structure predicted for the 5' side of poly (C) by this method differs slightly from that previously published (3) in one important respect. The new model has an additional branch structure between nucleotides 13 and 32 which is "balanced" by a similar new loop between nucleotides 335 and 350. The significance of the first loop, which was predicted using partial sequence data from FMDV by Harris (19), is that it is in a region of the genome, close to the 5' end, that has been reported to form a stable hairpin structure in poliovirus (16, 5) and in many other picornaviruses (JV Maizel, personal communication). It has recently been reported that a single nucleotide deletion in the stem of this hairpin causes a poliovirus mutant to exhibit a temperature sensitive phenotype (14). It is therefore possible that this feature may be of general importance in picornavirus replication.

The secondary structure predicted for the untranslated region to the 3' side of the poly (C) tract predicts two branched stem structures which together

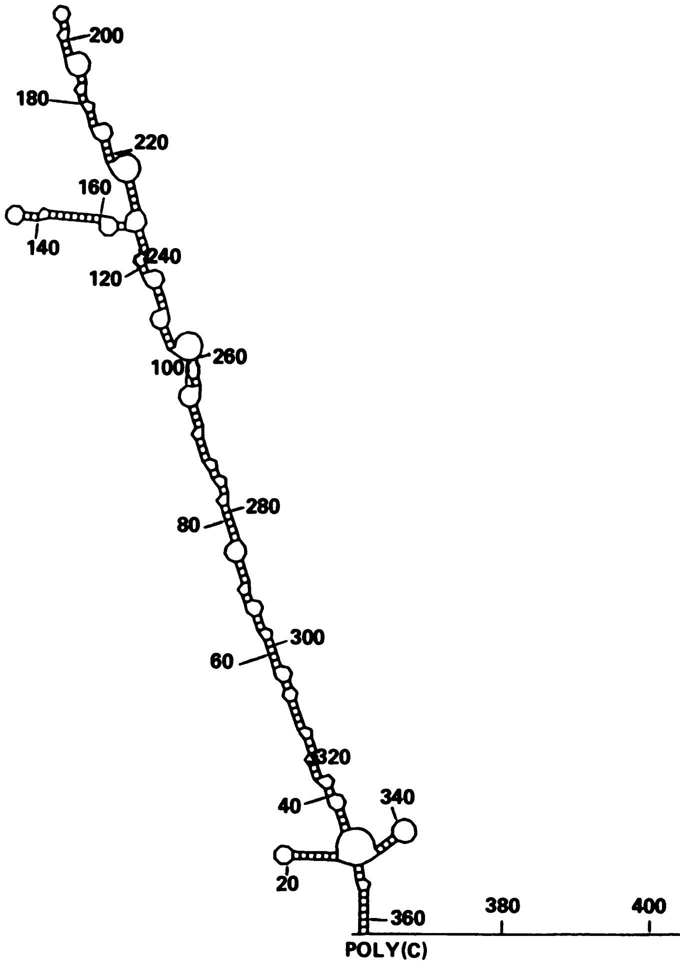

Figure 3
Predicted secondary structure of FMDV serotype A_{10} from the 5' end of the genome to the poly (C) tract. The numbering is as shown by Newton et al (3) with the poly (C) tract commencing at nucleotide 363. The new branch structure (see text) is located between nucleotides 13 and 32.

account for virtually all the primary sequence from this region of the genome. Indeed, the only region of sequence not involved in highly significant secondary structure is the poly (C) tract itself. The first of the branches is by far the larger and contains the repeated motif which was previously discussed. Although both elements of the repeat (ie, both 5 and 24 base) are involved in predicted secondary structure it is difficult to

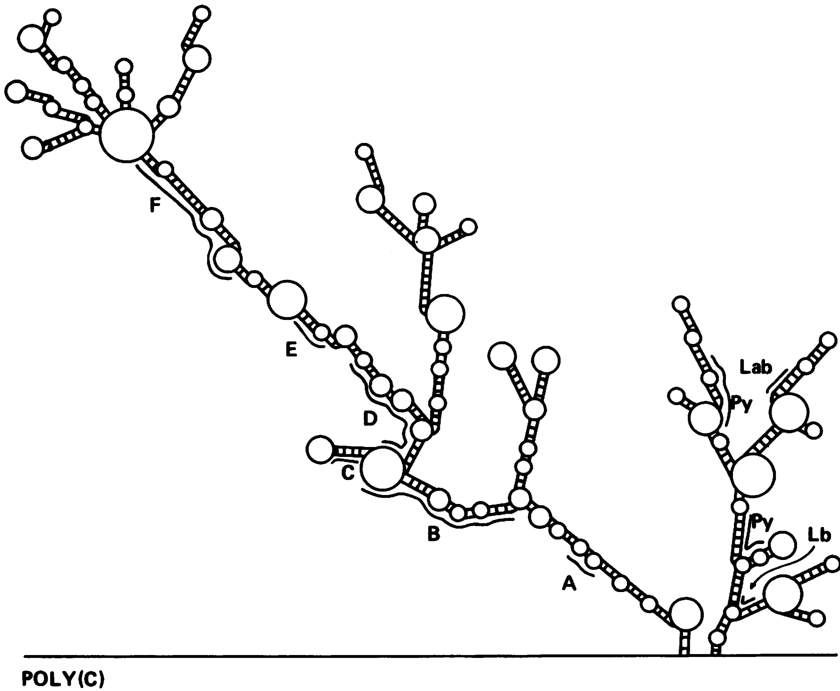

Figure 4
Predicted secondary structure of FMDV serotype A_{10} in the untranslated region to the 3' side of the poly (C) tract. The positions of the pairs of repeated motifs for A_{10} (1), A_{10} (2) and A_{10} (3) (see Figure 2) are indicated by A --- F respectively. Also indicated are the positions of the polypyrimidine tracts which may have a role in translation initiation (Py) and the two AUG codons themselves (Lab and Lb). The position of the poly (C) tract is indicated.

visualize repeated structural domains (Figure 4). The second branched structure involves both initiation sites for protein synthesis and the polypyrimidine tracts which precede them. Beck et al (18) suggested that these tracts had complementarity to 18S ribosomal RNA and might be involved in ribosomal recognition. If this model is correct then both polypyrimidine tracts would be involved in secondary structure and presumably would require some destabilization of the structure before they could function. Recent evidence has suggested that FMDV behaves unlike most other eukaryotic mRNAs in its translational strategy (9, 11, 20) in that it appears to initiate translation internally. Indeed it seems unlikely that ribosome subunits could traverse such a great length of secondary structure by binding at the 5' end of the genome and scanning for translational initiation sites (21,

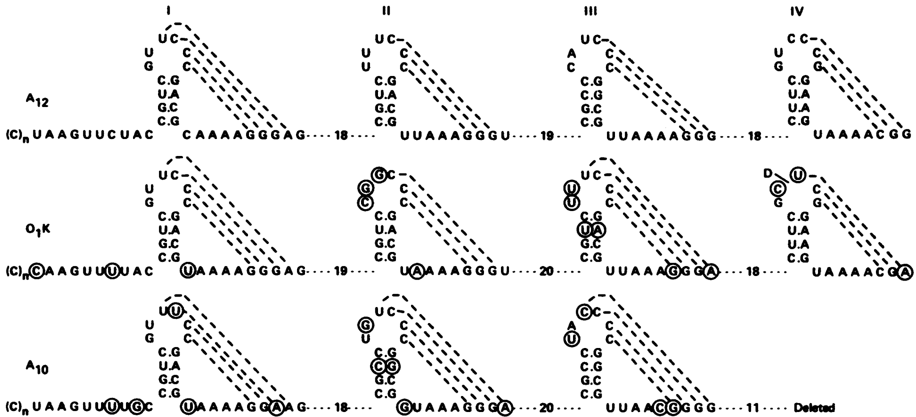

Figure 5A

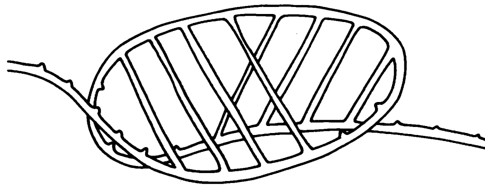

Figure 5B

Figure 5
Predicted tertiary structure in the genome of FMDV. The presence of pseudo knotted structures in the genomes of A_{12}, O_1 Kauf and A_{10}. Dotted lines indicate interactions between components of the pseudoknot. Base substitutions are highlighted by circles and the location of a deletion in the O_1 K serotype is indicated by D. The lower half of the figure shows a three dimensional diagrammatic representation of one of the pseudoknots.

36, 37). Moreover, scanning ribosomes would need to by-pass eight upstream AUG codons before reaching the authentic initiation sites in the A_{10} virus. In this respect therefore it is highly significant that the region surrounding the two AUG codons does possess a high degree of predicted secondary structure since this may be an element in direct ribosomal recognition.

It is clear that predictions for the folding of RNA chains are still in their early stages. Secondary structure predictions are only a first approximation and take no account of tertiary interactions in the formation of a three dimensional structure. An alternative method for predicting the structure of an RNA molecule has been developed by Pleij et al (7,8) which not only predicts secondary structure but allows the interaction of single

stranded elements of the secondary structure to form tertiary structures known as pseudoknots. Pseudoknots have been predicted to be present at the 3' terminus of the genomic RNAs of many plant viruses where they are involved in tRNA-like structures which are thought to play a crucial role in the RNA replication cycle (25). There is also some chemical and enzymatic evidence for the existence of these structures (7, 8). Obviously one role of the untranslated regions of FMDV will be an involvement in an RNA replication function. Consequently the 5' and 3' noncoding sequences for the three viruses were analysed by Dr C W A Pleij (University of Leiden) to ascertain specifically whether any pseudoknotted structures could be found. Analysis of the first 160 nucleotides to the 3' side of the poly (C) tract showed that four, almost identical, pseudoknots could be folded as a consequence of the three direct repeats already observed (Figure 5). It is also quite clear that the deletion in strain A_{10} corresponds to the elimination of one of these pseudoknot domains. Furthermore, compensating base changes in the stem regions of these structures (as indicated in Figure 5) as well as the location of substitutions and deletions are supportive of these predictions. This result is particularly exciting because it predicts, for the first time, that animal (polyadenylated) viral RNAs can possess pseudoknotted structures and, moreover, that they are located in a 5' noncoding region. In contrast to the plant viral RNAs no conserved pseudoknotted structures were predicted to be present in the 3' noncoding region of FMDV, nor were they consistently predicted elsewhere in the 5' noncoding region. The presence of these structures raises interesting possibilities as to their function. If they are involved in a replication function, as are some of them in plant viral structures, why are they located near the 5' end and not the 3' end of the genome and, secondly, why are they located so far from the 5' end. During the replication cycle of FMDV some mechanism exists to ensure an excess production of positive strand RNA for translational purposes and for the assembly of infectious particles. Perhaps these structures are therefore important as components of the 3' end of the negative strand RNA. A second possible function might be an involvement in the assembly of virion particles during morphogenesis. Mellor et al (31, 34) have shown that interactions between viral proteins and genomic RNA do occur within the viral particle. It would be interesting to establish whether the "pseudoknots" are involved in these interactions. We have recently shown (Sangar and Clarke, manuscript in preparation) that it is possible to assemble "empty" viral particles *in vitro* using subgenomic

RNAs possessing no untranslated region at all. However, this does not preclude the possibility that pseudoknots are critical in assembly of infectious viral particles *in vivo*.

Finally, for reasons which were previously discussed regarding the translational strategy of FMDV it is possible that the pseudoknots could be recognised directly by ribosomes or other factors as a means of achieving internal initiation of translation. Several other intriguing questions regarding the presence of these pseudoknots remain to be answered. Is there any relation with the presence or function (if any) of the poly (C) tract? Are pseudoknots also predicted to be present in other picornaviral genomes. Why does the A_{10} virus possess three of these structures while both A_{12} and O_1Kaufbeureun have four? Perhaps the most enigmatic question remaining is why a portion of the untranslated region of a viral RNA genome should have homology to an untranscribed region of a mammalian DNA genome. Hopefully it will be possible to answer some or all of these questions by specific manipulation of the genome of an infectious recombinant clone of FMDV.

It is of course impossible to reconcile the two different predictions for the folding of the 5' untranslated region of FMDV RNA. The presence of the three pseudoknots (in A_{10}) would greatly change the secondary structure prediction shown in Fig 4. However there is circumstantial evidence for the existence of the pseudoknots, namely i) coincidence with a repeated sequence, ii) deletion of one domain in the A_{10} serotype and iii) compensatory base substitutions between serotypes. It should be emphasised that these structures are only predictions and it will require experimental analysis such as UV cross linking or enzymatic cleavage to ascertain which features of these predictions actually occur.

CONCLUSION

We have completed the genomic sequence of one serotype of foot and mouth disease virus. Analysis of the sequences in the 5' untranslated region of this and two other strains have revealed some extremely interesting features. The most significant of these is the presence of a 24 base repeated sequence at the 5' end of the genome which is predicted to be involved in the formation of repeated elements of tertiary structure, known as pseudoknots. This is the first example of pseudoknotted structures predicted to be found in animal viral RNAs and also at the 5' end of a genomic RNA.

ACKNOWLEDGEMENTS

We are indebted to R O Campbell for providing FMDV RNA and E Evans for typing the manuscript. We are also particularly grateful to Professor J V Maizel (JR) for help with the secondary structure predictions and Dr C W A Pleij for tertiary structure analysis.

REFERENCES

1. Rowlands D J, Harris T J R & Brown F (1978) J Virol **16**, 335-343.
2. Caton A J, Brownlee G G, Yewdell J W & Gerhard W (1982) Cell **31**, 417-427.
3. Newton S E, Carroll A R, Campbell R O, Clarke B E & Rowlands D J (1985) Gene **40**, 331-336.
4. Messing J & Vieira J (1982) Gene **19**, 269-276.
5. Currey K M, Peterlin B M & Maizel J V Jr (1986) Virology **148**, 33-46.
6. Jacobson A B, Good L, Simonetti J & Zuker M (1984) Nucleic Acids Res **13**, 1717-1731.
7. Pleij C W A, Rietveld K & Bosch L (1985) Nucleic Acids Res **13**, 1717-1731.
8. Pleij C W A, Abrahams J P, van Belkum A, Rietveld K & Bosch L (1986) In "Positive Strand RNA Viruses" UCLA Symposia or Molecular and Cellular Biology, New Series, Vol **54**, Editors M A Brinton & R Rueckert; Alan Liss Inc New York, NY.
9. Clarke B E, Sangar D V, Burroughs J N, Newton S E, Carroll A R & Rowlands D J (1985) J Gen Virol **66**, 2615-2626.
10. Robertson B H, Grubman M J, Weddell G N, Moore D M, Welsh J D, Fischer T, Dowbenko D J, Yansura D G, Small B & Kleid D G (1985) J Virol **54**, 651-660.
11. Sangar D V, Newton S E, Rowlands D J & Clarke B E (1987) Nucleic Acids Res **15**, 3305-3315.
12. Fowlkes D M, Mullis N T, Couneau C M & Crabtree G R (1984) Proc Natl Avad Sci **81**, 2313, 2316.
13. Evans D M A, Dunn G, Minor P D, Schild G C, Cann A J, Stanway G, Almond J W, Currey K & Maizel J V Jr (1985) Nature **314**, 548-550.
14. Racaniello V R & Meriam C (1986) Virology **155**, 498-507.
15. Mellor E J C, Brown F & Harris T J R (1985) J Gen Virol **66**, 1919-1929.
16. Larsen G R, Semler B L & Wimmer E (1981) J Virol **57**, 328-335.
17. Racaniello V R & Baltimore D (1981) Proc Natl Acad Sci **78**, 4887-4891.
18. Beck E, Forss S, Strebel K, Cattaneo R & Feil G (1983) Nucl Acids Res **11**, 7873-7885.
19. Harris T J R (1980) J Gen Virol **36**, 659-664.
20. Sangar D V, Black D N, Rowlands D J, Harris T J R & Brown F (1980) J Virol **33**, 59-68.
21. Kozak M (1986) Proc Natl Acad Sci, USA **83**, 2850-2854.
22. Arnold E, Luo M, Vriend G, Rossmann M G, Palmenberg A C, Parks G D, Nicklin J H & Wimmer E (1987) Proc Natl Acad Sci **84**, 21-25.
23. Forss S, Strebel K, Beck E & Schaller H (1984) Nucleic Acids Res **12**, 6587-6601.
24. Chung D W, Rixon M W, MacGillivray R T A & Davie E W (1981) Proc Natl Acad Sci **78**, 1466-1470.
25. Strauss E G, Strauss J H (1983) Replication strategies of the single stranded RNA viruses of eukaryotes. In:- "Current Topics in Microbiology and Immunology", New York:- Springer-Verlag Vol **105**, p 1.
26. Cohen J I, Ticehurst J R, Purcell R H, Buckler-White A & Baroudy B M (1987) J Virol **61**, 50-59.

27. Robertson B H, Morgan D O, Moore D M, Grubman M J, Card J, Fischer T, Weddell G, Dowbenko D & Yansura D (1983) Virology **126**, 614-623.
28. Boothroyd J C, Harris T J R, Rowlands D J & Lowe P A (1982) Gene **17**, 153-161.
29. Makoff A J, Paynter C A, Rowlands D J & Boothroyd J C (1982) Nucleic Acids Res **10**, 8285-8295.
30. Kurz C, Forss S, Kupper H, Strohmaier K & Schaller H (1981) Nucleic Acids Res **9**, 1919-1931.
31. Mellor E J C (1982) PhD thesis, University of Reading.
32. Burroughs J N, Sangar D V, Clarke B E, Rowlands D J, Billiau A & Collen D (1984) J Virol **50**, 878-883.
33. Carroll A R, Rowlands D J & Clarke B E (1984) Nucleic Acids Res **12**, 2461-2472 and 4430 (Corrigenda).
34. Morrell D J, Mellor E J C, Rowlands D J & Brown F (1987) J Gen Virol **68** (in press).
35. Palmenberg A C, Kirby E M, Janda M R, Drake N L, Duke G M, Potratz K F & Collett M S (1984) Nucleic Acids Res **12**, 2969-2985.
36. Pelletier J & Sonenberg N (1985) Mol Cell Biol **5**, 3222-3230.
37. Pelletier J & Sonenberg N (1985) Cell **40**, 515-526.