Two Initiation Sites for Foot-and-Mouth Disease Virus Polyprotein in vivo

By B. E. CLARKE,* D. V. SANGAR, J. N. BURROUGHS,† S. E. NEWTON, A. R. CARROLL AND D. J. ROWLANDS

Wellcome Biotechnology Ltd, FMD Division, Ash Road, Pirbright, Woking, Surrey GU24 0NQ and Animal Virus Research Institute, Ash Road, Pirbright, Woking, Surrey GU24 0NF, U.K.

(Accepted 12 August 1985)

SUMMARY

Typically, the translation of eukaryotic mRNAs into protein is initiated at a single site. However, we have recently shown that not one but two primary products, P20a and P16, are translated from the 5' end of the coding region of the genome of foot-and-mouth disease virus (FMDV). In this paper we show by partial protease digestion of these proteins that they differ only at their N termini, thus confirming the presence of two initiation sites for translation of FMDV RNA. Sequence analysis of two subtypes of the virus (A10 and A12) confirms the presence of two initiator AUG codons in the expected position on the genome. By correlation with protein synthesis data from these subtypes it appears that the relative use of each initiation site is dependent on its surrounding nucleotide sequence. In addition, the ratio of the two proteins when synthesized in vitro differs markedly from that when they are synthesized in vivo, suggesting the presence of a control mechanism for synthesis of P20a in vivo which may be absent in vitro. We also show that the cleavage site between these two proteins and the structural protein precursor, P88, is located closer to the N terminus of the polyprotein than has previously been reported.

INTRODUCTION

Foot-and-mouth disease virus (FMDV) is a member of the family Picornaviridae, having a single-stranded RNA genome of positive sense, approximately 8500 bases in length (Sangar et al., 1980). The picornaviruses have been classified into four genera: aphthovirus (FMDV), enterovirus, cardiovirus and rhinovirus. FMDV can be further divided into seven distinct serotypes and many serologically distinct subtypes.

The mechanism of protein synthesis in picornaviruses was first studied using poliovirus (Jacobson et al., 1970) when it was shown that only one major initiation site for protein synthesis existed near the 5' end of the RNA. This initiation site is used for the synthesis of a polyprotein which is processed into the viral structural and non-structural proteins (Jacobson & Baltimore, 1968; Butterworth et al., 1971). These observations allowed the order of the genes to be determined by pactamycin mapping (for review, see Rekosh, 1977). The general strategy of protein synthesis has been shown to be similar in all picornaviruses and pactamycin mapping of FMDV RNA originally showed that three primary products were derived from the polyprotein (Sangar et al., 1977; Doel et al., 1978). Subsequently, two smaller polypeptides P20a (Lab) and P16 (Lb) [letters in parentheses refer to the nomenclature system proposed by Rueckert & Wimmer (1984)] were detected by short pulse-labelled experiments; pactamycin mapping showed them to originate from nearer the 5' end of the genome than the other primary products (Sangar et al., 1980). Furthermore, it was shown that both these proteins could be labelled with N-formyl-[35S]methionyl tRNA (N-F[35S]Met tRNA) by in vitro translation of the RNA, indicating that both these proteins are involved in the initiation of translation.

Since P20a and P16 were shown to be closely related (Sangar et al., 1980) by tryptic peptide mapping it was assumed that they must therefore exhibit a precursor-product relationship. However, pulse-chase experiments clearly showed that this was not the case. In addition, we have recently shown, using inhibitors of proteolytic cleavage, that the C termini of the two

0000-6739
proteins are the same, indicating that they may not share a common N terminus (Burroughs et al., 1984). Since both proteins contain an initiator methionine it follows that FMDV RNA is translated from either of two closely situated initiation sites, unlike eukaryotic mRNAs in general which typically possess a single initiation site. Recently, sequence analysis of three serotypes of the virus has indicated the presence of two initiator AUG codons in the expected positions on the genome in support of this model (Beck et al., 1983; Carroll et al., 1984). In this paper, we present the first direct evidence that the two proteins involved do indeed differ at their N termini and that the two sites vary in their efficiency of initiation of translation depending on the surrounding nucleotide sequence. We also present evidence that their site of cleavage from the structural protein precursor P88 (P1) is different from that which has previously been suggested (Boothroyd et al., 1982).

METHODS

Labelling of virus-induced protein in vivo. The viruses used in this study were FMDV type A10 [Argentina, 1961 (A61)] and type A12 [119ab]. Monolayers of BHK-21 cells in 4 oz bottles were infected at 37 °C at a multiplicity of about 100. Virus was allowed to attach for 30 min after which the monolayers were washed with methionine-free Eagle’s medium (MFE). The cells were incubated for a further 90 min and then again washed with MFE medium. The cells were pulse-labelled for 5 min at 150 min post-infection with 100 μCi [35S]methionine (Amersham). In chase experiments the radioactive medium was replaced with Eagle’s medium containing 1 mM-methionine and the incubation continued for a further 30 min. The polypeptides were prepared for polyacrylamide gel electrophoresis as described previously (Doel et al., 1978) and analysed on 10% acrylamide gels (Laemmli, 1970).

Preparation of viral RNA. FMDV RNA was prepared either from purified virus as described previously (Sangar et al., 1980) or by the following modification. Monolayers of BHK-21 cells were infected with a high multiplicity of virus (approx. 20). When cytopathic effect was extensive the cells were shaken from the glass and subjected to a single cycle of freeze– thawing. Cell debris was removed by pelleting at 10000 g for 30 min. The supernatant was then centrifuged for 90 min at 100000 g to pellet virus particles. The 100000 g pellet was resuspended in 0.15 M-NaCl/Tris/EDTA (0.15 M-NaCl, 0.1 M-Tris-HCl pH 7.5, 0.01 M-EDTA) and extracted twice with phenol/chloroform/0.5% SDS/0.1% 8-hydroxyquinoline. The aqueous phase was precipitated with 2 vol ethanol at −20 °C for 16 h. The nucleic acid pellet was resuspended in 0.3 M-sodium acetate pH 5.0 containing 0.1% SDS and separated on a 5 to 25% sucrose gradient, containing the above buffer, for 4 h at 250000 g. The RNA sedimenting at about 37S was collected and precipitated with 2 vol. ethanol. The RNA pellet was washed with ethanol and stored in distilled water at −70 °C.

In vitro translation of RNA. A micrococcal nuclease-treated rabbit reticulocyte lysate system purchased from Amersham was routinely used. The lysate was supplemented with 100 mM-KCl before use. The standard reaction mixture contained 10 μl lysate, 2 μg RNA and 1 μl [35S]methionine (10 μCi/ml). Reactions were carried out at 30 °C for the times indicated. The proteins were electrophoresed for gel electrophoresis as described previously (Sangar et al., 1980).

Partial protease mapping. A10 RNA was translated in vitro for 1 h at 30 °C in the presence of [35S]methionine or N-F[35S]Met-tRNA and the proteins were separated on 15% preparative polyacrylamide gels (Laemmli, 1970). P20a and P16 were located by autoradiography, the bands cut from the dried gels and the proteins eluted into 0.1 M-Tris-HCl pH 7.9 containing 1% SDS, 2-mercaptoethanol and 1 mM-phenylmethylsulphonyl fluoride. The proteins were recovered by acetone precipitation in the presence of 25 μg bovine serum albumin, the pellets dried and treated as described by Cleveland et al. (1977) using chymotrypsin (200 μg/ml) for 30 min at 37 °C. The products were electrophoresed on a 15% polyacrylamide gel (Laemmli, 1970) which was then processed for fluorography, dried and exposed to Fuji RX film.

Cloning procedures. Recombinant clones from the A10 subtype were constructed using a synthetic oligonucleotide primer (Celltech, Slough, U.K.) which had previously been used for primer extension sequencing (Carroll et al., 1984). First-strand cDNA synthesis was performed as described by Rowlands et al. (1983). cDNA transcripts were then made double-stranded and cloned in Escherichia coli strain MC1061 (Casadaban & Cohen, 1980) using Clal-digested pAT153 and Clal linkers (Amersham) by standard procedures. Clones containing sequences representing the translational initiation sites were identified by size analysis and by screening with the same primer used for the construction (Grunstein & Hogness, 1975). DNA sequence analysis of these clones was by the method of Maxam & Gilbert (1980).

Recombinant clones from the A12 subtype were constructed in a similar manner except that the double-stranded cDNA was digested with restriction enzymes Sau3AI or MspI to generate overlapping fragments which were used
FMDV initiation sites

in libraries as shotgun clones in either BamHI-digested or AccI-digested M13mp9 respectively (Messing & Vieira, 1982). These libraries were then screened using positively identified clones from the A10 serotype to identify initiation site sequences (Benton & Davis, 1977). DNA sequencing of these clones was carried out by the dideoxy method (Sanger et al., 1977).

RESULTS

Use of N-formyl methionyl tRNA to determine the initiation site for protein synthesis in aphthovirus RNA

All picornavirus genomes so far studied have very long leader sequences before the AUG codon initiating synthesis of the polyprotein. FMDV, like the cardioviruses, has a more complicated genomic structure due to the presence of a poly(C) tract 100 to 150 residues long located about 400 nucleotides from the 5' end of the genome (Fig. 1). In the case of FMDV, enzymic digestion of the poly(C) tract results in an RNA molecule lacking approximately 600 bases from the 5' end of the genome (Rowlands et al., 1978). In vitro translation of this RNA yields the same spectrum of proteins as the full-length RNA (Sangar et al., 1980), indicating that translation is initiated to the 3' side of the poly(C) tract. We have previously shown that for serotype A10 the use of N-[35S]Met tRNA for in vitro protein synthesis resulted in the labelling of two proteins P20a (Lab) and P16 (Lb) (Fig. 1), the latter being the more abundant (Sangar et al., 1980). We concluded from this work that FMDV RNA possessed only one site for the initiation of protein synthesis and that this site initiated both P16 and P20a; the difference between P16 and P20a was therefore assumed to be at the C terminus. A serious doubt was cast upon this conclusion, however, when it was shown that in the presence of a tripeptide protease inhibitor it was possible to identify both P20a (Lab) and P16 (Lb) linked to P88 (Fig. 1) through a common C terminus (Burroughs et al., 1984). We therefore investigated N-[35S]Met tRNA-labelled P20a (Lab) and P16 (Lb) by chymotryptic peptide analysis. In vitro translation experiments were carried out using A10 RNA in the presence of either N-[35S]Met tRNA or [35S]methionine. The gel profiles obtained for the two proteins after subsequent digestion with chymotrypsin are shown in Fig. 2. When [35S]methionine was used as label, the overall chymotryptic patterns for the two proteins were very similar (Fig. 2, lanes 5 and 7), again confirming their colinearity on the genome, However, one extra low mol. wt. band was found in the P20a (Lab) sample (Fig. 2, lane 7) which was not detectable in P16 (Lb) (lane 5). Similarly, this band was also present in the P20a (Lab) N-[35S]Met-labelled sample (lane 3) indicating that it originated from the N terminus of the protein, and since it was also absent from the N-[35S]Met-labelled P16 (Lb) sample (lane 1) this clearly showed that the two proteins differ at their N-termini. In fact, no N-[35S]Met-labelled band could be detected at all for P16 (Lb), the minor band observed probably arising from contamination of the N-formyl-Met tRNA with non-formylated Met tRNA. The absence of an N-[35S]Met-labelled P16 (Lb) band can now be explained from nucleotide sequence data (Carroll et al., 1984) which predict that the P16 (Lb) protein possesses a phenylalanine residue at +3 which would be cleaved by chymotrypsin and the resulting tripeptide would be undetected by this gel system.
Fig. 2. Fluorograph of a 15% polyacrylamide gel of polypeptides P20a (Lab) and P16 (Lb) labelled with either N-F[35S]Met tRNA (lanes 1 to 4) or [35S]methionine (lanes 5 to 8). Marker A10 polypeptides were synthesized in vivo (lane M). Lanes 1 and 5, P16 (Lb) incubated with chymotrypsin for 30 min; lanes 2 and 6, untreated P16 (Lb); lanes 3 and 7, P20a (Lab) treated as in lanes 1 and 5; lanes 4 and 8, untreated P20a (Lab).

Ratio of P20a to P16 differs between different subtypes of the A serotype

Having established that there were two initiation sites for protein synthesis for the A10 subtype of the virus we were interested to know why this phenomenon had not been observed using other serotypes and subtypes of the virus. For example, in contrast to results with A10 virus it has recently been shown for the A12 subtype that in vitro translation in the presence of N-F[35S]Met tRNA produces no detectable protein corresponding to P20a (Lab) (Grubman & Baxt, 1982; Grubman et al., 1984). We therefore decided to compare the two viruses directly. Initially, we performed a time course experiment for in vitro translation (Fig. 3). At early stages in translation the A10 virus produced significant amounts of both P20a (Lab) and P16 (Lb) which were stable throughout the translation. However, the A12 virus did not produce any protein corresponding to P20a at early stages in translation but there were significant amounts of P16. A band equivalent to P20a did appear in the A12 virus reaction after longer periods of translation (120 min) but its late appearance indicated that it was a secondary processing product.

Our results therefore confirmed those of Grubman et al. (1984) in that in vitro translation indicated a single initiation site for the A12 subtype. However, when the two viruses were grown in vivo and their proteins compared a different result was obtained. Fig. 4 shows a comparison of the proteins induced by A10 and A12 RNA either in vitro or in vivo. Clearly, a protein was synthesized from A12 RNA in vivo corresponding to P20a (Lab) that was not synthesized in vitro.
Similarly, it could be seen that the amounts of P20a and P16 were similar for the A10 virus \textit{in vitro} but there was a larger amount of P20a (Lab) synthesized \textit{in vivo} relative to P16 (Lb) (see also Fig. 3). The bands equivalent to P20a and P16 on the autoradiograph of A12 proteins produced \textit{in vivo} were identified by both pactamycin mapping and partial proteolysis mapping (Cleveland \textit{et al.}, 1977) (results not shown). It therefore appears that some additional control mechanism is present in infected cells to ensure the synthesis of a P20a (Lab) protein.

\textit{Ratios of the initiating proteins are reflected by the surrounding nucleotide sequence}

Since the \textit{in vivo} profile of the A12 products indicated that a protein corresponding to P20a (Lab) was synthesized, this implied that an initiator codon for P20a (Lab) must exist on the A12 genome even though it was not used during \textit{in vitro} translation. We therefore determined the nucleotide sequences in the regions of the initiating sites in an attempt to explain the reason why the ratios of P20a to P16 (Lab to Lb) differed in the two viruses. We have recently published the nucleotide sequence of the A10 virus up to and including the P20a initiation site (Carroll \textit{et al.}, 1984). These data were derived from primer extension sequencing using a synthetic oligonucleotide primer complementary to the middle of P20a/P16 (Lab/Lb). We subsequently used this primer to generate recombinant clones covering both initiation sites from both subtypes as described in Methods. Both clones were sequenced and the resulting sequences are
Fig. 4. Comparison of the virus-specific proteins synthesized either \textit{in vitro} (lanes 1 and 2) or \textit{in vivo} (lanes 3 and 4). Lanes 1 and 3, A10 subtype; lanes 2 and 4, A12 subtype. \textit{In vitro} translation was done in rabbit reticulocyte lysates for 40 min at 30 °C using [\textsuperscript{35}S]methionine as label. Proteins were synthesized \textit{in vivo} in infected BHK-21 cells by using a 5 min pulse of [\textsuperscript{35}S]methionine at 2.5 h post-infection. The samples were prepared as described in Methods and analysed by PAGE. The symbol ⋈ indicates the positions of P20a and P16.

shown in Fig. 5. The previously reported sequence for the polyprotein of FMDV type A\textsubscript{10} is now extended into the 5' untranslated region by a further 150 nucleotides. The sequence shows the presence of translational stop codons in all three reading frames upstream from the putative P20a (Lab) AUG, supporting the conclusion that this is the correct initiation site for the polyprotein \textit{in vivo}. The sequence obtained for the A\textsubscript{12} subtype commences at nucleotide −108 and extends through 553 nucleotides of P20a/P16. As expected, the A\textsubscript{12} sequence confirms the presence of two AUG codons at positions (no. 157 and 241) corresponding to those on the A\textsubscript{10} genome. However, minor variations were observed in the surrounding sequences. Kozak (1981) has proposed that for efficient initiation of translation to occur, an AUG must exist in a favourable sequence context, the most favourable consensus sequence for eukaryotic initiation sites being PuXXAUGPu (where X is any base). Fig. 5 shows that the P16 (Lb) initiation sites for both viruses (nucleotides 241 to 243) obey this consensus as does the P20a (Lab) initiation site.
**FMDV initiation sites**

20 40 60 80

A$_{10}$

GGCUAAGGACUCUAGUACCCGAGUAACACCGGACACCCUUUGGGAUCUGAAGGACUGGCUUCUUAAGAG

A$_{12}$

G A

100 120 140 160

A$_{10}$

CGCUCCGGUUAAGCCUUGACCGGAGGAACCGCAGGGCACCAGUU-AUGUAACACCCCGUCCUUUCGUGACCUGAUAU

A$_{12}$

UGA

180 200 220 240

A$_{10}$

UGAAAUCACACUGCAAACGUGAAAGAGAAAGGAAAUGGCUGCCCAAUAACCACUGCUUGGACACC

A$_{12}$

G U C G U G U C C U

260 280 300 320

A$_{10}$

GUGAAUUCUCACUCUGAUAUCGGUGAAAGAGAAAGGAAAUGGCUGCCCAAUAACCACUGCUUGGACACC

A$_{12}$

G U C G U G U C C U

340 360 380 400

A$_{10}$

AUCCUCAGGUGUGAGUACGUGAAGCUCUUCUUCGAGCAUCACUCGGCAGAAUCUCACUGGCUUGAGGC

A$_{12}$

U C G A A A C A

420 440 460 480

A$_{10}$

UAUCAGGUGUGAAAGAAACUGCCCAGGCACUUGAGUGACGGGCGACCACUCCUGCCUGCUGAGAUCUGAACAAC

A$_{12}$

C A C G A C A U C

500 520 540 560

A$_{10}$

ACUUGCUCCAAACGGGUAUGCUACCUGCCUUUCAGCGGACUGUGUAGUGAUGACGGGCACGGGCAUGUGUCUUGCU

A$_{12}$

C U U G

580 600 620 640

A$_{10}$

GAAUCCCGAGGCAUUCAUGAAGCAGGACCACCGAGUGUCUGUGUACCCUCACAGCGGUGGUACGGCAU

A$_{12}$

U C C U CA

660 680 700 720

A$_{10}$

UGACCAGGACUUCUUACGCCUGGACCGCUACCCCAUGAGCCUGGCUUUGGCCAUUGGACAGAAGAACACCACUCA

A$_{12}$

C U C G G U C C U

P20a/P16(L) / VP4(P1a)

740 760 780

A$_{10}$

UGGGGAGUGAAACACUGGUUCAGAAGGCUGCUGGCUAGG / CAGUGCAGCCGACGACCCGGGUCGCGAGA

Fig. 5. Nucleotide sequence from viruses of subtypes A$_{10}$ and A$_{12}$ in the region of the translation initiation sites (indicated in bold type). The main sequence shown is that of the A$_{10}$ virus with differences indicated below. Deletions are indicated by dashed lines. The A$_{12}$ sequence extends from the Sau3A1 site at position 706. The polypyrimidine tracts precede the respective AUG codons by approximately 16 bases in each case.

for the A$_{10}$ virus. However, the A$_{12}$ virus which, as we have shown, produces a relatively small amount of P20a (Lab) protein possesses an unfavourable initiation site for the P20a (Lab) protein.

It has recently been reported for the O$_1$ serotype (Forss *et al.*, 1984) that the first AUG downstream from the poly(C) tract could potentially initiate a 10K protein. Strikingly, this particular AUG codon is preceded by a long stretch of pyrimidine residues, a feature which appears to be characteristic of functional initiation sites not only in FMDV (Fig. 5; Beck *et al.*, 1983) but also in other picornaviruses and other eukaryotic messenger RNAs (Lewin, 1980). These results suggest the potential for three initiation sites in the O$_1$ serotypes. We have been unable to detect a 10K protein supporting this in the A serotype and further nucleotide sequencing of this portion of the genome will be required to confirm whether this open reading frame is conserved.
Fig. 6. Homology between FMDV and EMC virus proteins in the region of the cleavage between the leader proteins and the structural precursor. The EMC virus sequence is derived from Palmenberg et al. (1984). Continuous lines indicate homologous amino acids and dotted lines indicate conservative changes. The proline shown to be present in FMDV VP4 is indicated by *.

Table 1. Distribution of radioactivity in specific viral proteins after labelling of virus in vivo

<table>
<thead>
<tr>
<th>Protein</th>
<th>[14C]Proline</th>
<th>[14C]Lysine</th>
<th>Proline residues</th>
<th>Lysine residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP1</td>
<td>48733</td>
<td>38978</td>
<td>40†</td>
<td>32†</td>
</tr>
<tr>
<td>VP2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP3</td>
<td>947</td>
<td>1341</td>
<td>1</td>
<td>1†</td>
</tr>
<tr>
<td>VP4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* C.p.m. in polyacrylamide gel slices containing the indicated proteins as visualized by autoradiography. † Carroll et al. (1984).

Cleavage site of P20a/P16–P88 (Lab/Lb–P1)

The exact proteolytic cleavage site between the primary products P20a/P16 (Lab/Lb) and P88 (P1) has been difficult to ascertain because both P88 (P1) and its N-terminal product [VP4 (P1a)] are 'blocked' (Boothroyd et al., 1982). The original predictions (Boothroyd et al., 1982; Beck et al., 1983) were based on the assumption that this cleavage was carried out by the protease which cleaves the P88–P52 (P1–P2) junction. However, we have recently shown that this is not true (Burroughs et al., 1984) and we have therefore re-examined the amino acid sequence in the region where the cleavage was expected to occur. By comparison of this sequence with that of the proposed encephalomyocarditis (EMC) virus L–P1a junction (Palmenberg et al., 1984), it is apparent that there is considerable homology between the two VP4 (P1a) proteins (Fig. 6) at the N terminus and that this homology extends past the putative cleavage site (Gly–Asn) which has previously been suggested for FMDV (Boothroyd et al., 1982). If the FMDV site were correct, then the EMC virus protein would have an additional 15 amino acids at the N terminus. This would mean a 25% increase in its size compared to the FMDV protein, which is unlikely from size determinations by Sepharose chromatography in the presence of 6 M-guanidine hydrochloride (D. J. Rowlands, unpublished results). We therefore postulate that the FMDV cleavage might occur at the same relative position on the polyprotein as the EMC virus site. Examination of the FMDV sequence showed that if this were the case then the FMDV VP4 protein should contain a single proline residue that would not be present if the site originally proposed were correct. FMDV was labelled in vivo with [3H]methionine and either [14C]proline or [14C]lysine which should also be present as a single residue in VP4. The virus was then purified and the structural proteins were separated by PAGE. The distribution of radioactivity in the specific viral proteins is shown in Table 1. Clearly the amount of [14C] incorporated into VP4 was similar for both amino acids and indicates that VP4 does possess a single proline residue, suggesting that the actual cleavage site was N-distal with respect to the proline residue. We suggest that the cleavage occurs at the Gly–Gln pair since this would produce an N-terminal
**DISCUSSION**

Tryptic peptide mapping data have shown that the two proteins coded for by the 5' end of the FMDV genome are colinear (Sangar et al., 1980); more recently, immunoprecipitation of bacterially expressed protein has confirmed this (Beck et al., 1983). However, the implication that the two proteins differ at their N termini has only arisen with the publication of nucleotide sequence data from this region (Beck et al., 1983; Carroll et al., 1984) and our work using protease inhibitors (Burroughs et al., 1984). The data presented here show that when the two proteins are labelled specifically with N-F\(^{[35S]}\)Met tRNA the label is present on different chymotryptic fragments. This is therefore the first direct evidence that FMDV, unlike most eukaryotic mRNAs, utilizes multiple initiation sites for translation in vitro. Comparison of the FMDV-specific polypeptides synthesized in vitro with those found in infected cells strongly suggests that both initiation sites are also functional in vivo.

Beck et al. (1983) observed that the nucleotide sequence changes in the region between the P20a (Lab) and P16 (Lb) initiation sites were mostly silent or resulted in conservative changes, for serotypes O, and C. However, when our data for the A serotype are included in the comparison, it is apparent that this region of the protein is actually more variable in nature than is the remainder (Fig. 7). At present the functions of P20a–P88 (Lab–P1) is in the general area we have suggested.

![Fig. 7. Partial amino acid sequences for P20a (Lab) and P16 (Lb) from four distinct FMDV types. The data for C, and O, Kaufbeuren (O,K) are derived from Beck et al. (1983).](image-url)

Glutamine on P88 (P1) and VP4, and cyclization of glutamine residues is a common method of blocking N-terminal sequences in eukaryotes (Doolittle, 1970). Furthermore, the isoelectric point of VP4 (King & Newman, 1980; Boothroyd et al., 1982) indicates that it is acidic in nature; however, if our postulated N terminus were extended by a further three amino acids then a stretch of three lysine residues in four amino acids would be encountered. This presents further evidence that the cleavage site for P20a–P88 (Lab–P1) is in the general area we have suggested.
The mechanism by which eukaryotic ribosomes select the correct initiation site for translation is still not fully understood. A growing body of evidence suggests that ribosomes recognize the cap structure found at the 5' end of eukaryotic mRNAs and then scan the RNA until the first AUG is encountered (Kozak, 1980, 1981, 1983a). Consistent with this model is the observation that 95% of eukaryotic mRNAs do not contain AUG codons upstream of the translational initiation site (Kozak, 1983b). A recent modification of this model has been that initiation of translation at internal AUGs can occur provided that the upstream AUGs are closely followed by in-phase terminator codons (Liu et al., 1984; Kozak, 1984).

Picornaviruses, however, do not possess cap structures (Hewlett et al., 1976; Lee et al., 1976) and in those for which sequence data are available the initiating AUG is not the 5' proximal one. Indeed, Forss et al. (1984) have shown for the O1 serotype of FMDV that eight additional AUGs are located between the poly(C) tract and the initiation sites. Furthermore, we have recently determined the sequence of that portion of the genome to the 5' side of the poly(C) tract and three additional AUG codons have been identified (S. E. Newton, unpublished). With the single exception which we have already discussed, and which would produce a 10K protein, none of the other AUG codons would code for a polypeptide of significant length.

A further modification of the scanning model of translation initiation states that 40S ribosomal subunits may bypass AUG codons if they do not exist in an optimal sequence context (Kozak, 1984; Losson et al., 1983). The results reported here would tend to support this model since for the A10 serotype which possesses a favourable initiation site for P20a a significant amount of this protein is synthesized.

Indeed, the initiation site for P20a is so favourable that very little P16 should be observed at all. Consequently, there must also exist in vivo a mechanism for ensuring the synthesis of P16. In contrast to this the P20a initiation site in the A12 subtype is a poor one which the 40S subunits would bypass and hence favour the synthesis of P16 in this subtype, suggesting that a mechanism may exist in FMDV-infected cells ensuring that both P20a and P16 are synthesized in significant amounts.

It is interesting to note that the A10 and A12 viruses and the O1 and C1 viruses (Beck et al., 1983) have single base deletions relative to each other in the middle of the polypyrimidine tract preceding P20a (Lab). This suggests that this tract may be involved in some structural features of the genome and is probably not involved in a coding function. Apart from this deletion there is a striking conservation in the nucleotide sequence of the untranslated region. Indeed, the conservation between the A10 and A12 viruses is greater for the untranslated region (92% over 110 bases) than for the actual P20a/P16 coding portion (88% over 550 bases). Similarly, the sequence of the 5' untranslated region is markedly conserved when compared with viruses from other serotypes (Beck et al., 1983). The most likely reason for this conservation in sequence is that the 5' untranslated region exhibits a high degree of secondary structure. The involvement of secondary structure in modulating the efficiency of translation in prokaryotic systems is well known (Hall et al., 1982; Gheysen et al., 1982). It has also been shown recently that factors do exist in eukaryotic systems which interact with the 5' untranslated region of mRNAs and can either enhance or inhibit translation (Thireos et al., 1984). A model for FMDV translation could therefore be proposed in which factors which are present in vivo may be directed towards specific AUG codons on the genome by recognition signals inherent in the highly conserved secondary structure of the RNA. Furthermore, the polypyrimidine tracts preceding the functional initiation sites show significant base complementarity to the 3' end of eukaryotic 18S ribosomal RNA (Beck et al., 1983). Therefore, these initiation factors may direct the scanning ribosomal subunits towards the polypyrimidine tracts whereupon initiation of translation would begin at the next AUG. Such a function might explain the importance of having such a long untranslated region in an otherwise compact genome. In support of such a model, Dorner et al. (1984) have recently shown that internal initiation in vitro for another picornavirus (poliovirus) is dependent on the absence of additional factors present in poliovirus-infected HeLa cell extracts.

The existence of two initiation sites on certain picornavirus genomes has been suspected for some time (Celma & Ehrenfeld, 1975; Perez-Bercoff & Kaempfer, 1982; Degener et al., 1983) and our evidence supports this view. Moreover, results are also accumulating for several other
viruses, e.g. coronavirus (Skinner et al., 1985), infectious pancreatic necrosis virus (Mertens & Dobos, 1982), Sendai virus (Giorgi & Kolakofsky, 1984), which suggests that internal initiation of translation may be a more common feature than was previously assumed.

We wish to thank Andrew Syred, Alan Brown and Rosemary Clark for skilled technical assistance, and Brenda Matthew and Sarah Blackman for typing of the manuscript.

REFERENCES


(Received 12 June 1985)