Nucleotide sequence and genome organization of foot-and-mouth disease virus

S. Forss, K. Strebel, E. Beck and H. Schaller

Microbiology, University of Heidelberg, Im Neuenheimer Feld 230, D-6900 Heidelberg, FRG

Received 21 May 1984; Revised and Accepted 24 July 1984

ABSTRACT
A continuous 7802 nucleotide sequence spanning the 94% of foot and mouth disease virus RNA between the 5'-proximal poly(C) tract and the 3'-terminal poly(A) was obtained from cloned cDNA, and the total size of the RNA genome was corrected to 8450 nucleotides. A long open reading frame was identified within this sequence starting about 1300 bases from the 5' end of the RNA genome and extending to a termination codon 92 bases from its polyadenylated 3' end. The protein sequence of 2332 amino acids deduced from this coding sequence was correlated with the 260 K FMDV polyprotein. Its processing sites and twelve mature viral proteins were inferred from protein data, available for some proteins, a predicted cleavage specificity of an FMDV encoded protease for Glu / Gly(Thr, Ser) linkages, and homologies to related proteins from poliovirus. In addition, a short unlinked reading frame of 92 codons has been identified by sequence homology to the polyprotein initiation signal and by in vitro translation studies.

INTRODUCTION
Foot and mouth disease viruses (FMDV) or aphthoviruses are picornaviruses which are the causative agent of an aggressive and economically important disease of cloven-footed farm animals. Their virion contains a single-stranded RNA genome of about 8 kb with a small protein (VPg) covalently attached to its 5' end, an internal poly(C) tract, and a poly(A) sequence at the 3' end. This RNA is of positive polarity and can act directly as a messenger RNA. Protein synthesis involves post-translational cleavage of a 260 K polyprotein which is encoded between a single major translation initiation site next to the poly(C) tract and the 3' end of the RNA. This mode of protein synthesis is common to all picornaviruses (1) and makes them interesting model systems for studying the mechanisms of translation initiation and of protein maturation by specific proteolytic cleavages (2).

To obtain more information about the FMDV genome, its control signals and its gene products we have cloned cDNA copies of the viral RNA from strain O1K and determined their nucleotide sequence. Using a set of overlapping clones we obtained a continuous sequence of 7915 nucleotides representing the 3' proximal long L segment of the FMDV RNA which is thought to contain all its coding information. In a previous publication we had already determined the initia-
Nucleic Acids Research

Ion sites for polyprotein synthesis and a long open reading frame encoding the structural proteins of FMDV (3). The present work extends this reading frame by the coding sequence for the non-structural proteins, predicting a single translation product of 2332 amino acids which could be correlated in many parts with known protein data from FMDV induced proteins. In addition, we report and discuss a sequence of 713 bases preceding the polyprotein gene to the poly(C) tract.

MATERIALS AND METHODS

Enzymes

Restriction endonucleases were isolated and purified by standard procedures or purchased from Boeringer Mannheim or Biolabs. Avian Myeloblastosis Virus reverse transcriptase was a gift from W. Keller, Heidelberg. Nuclease S1 from Aspergillus oryzae, calf intestinal phosphatase, and T4 polynucleotide kinase were from Boehringer Mannheim, and terminal deoxynucleotidyl transferase was from BRL. [α-32P]-ATP was prepared according to (4).

FMDV RNA

FMDV O K RNA, isolated after 7 and 64 passages in BHK-cells of the same field-isolate of the virus (Kaufbeuren 1966-67) was provided by K. Strohmaier and W. Keller. The virus was plaque-purified at the beginning of the passages and again after passage 16.

FMDV cDNA clones

The clones FMDV-2735 and -2815 were constructed essentially as described by (3) using single-stranded restriction fragments (mapping at pos. 3682-3752 and 3792-3882, cf. Figure 2) from the existing cDNA clone FMDV-715 as primers for cDNA-synthesis. The double-stranded cDNA was provided with dC-tails and annealed to dG-tailed pBR322, linearized at the PstI site. Construction of clones FMDV-3214a and -3214c was carried out using a modified procedure that avoids second strand synthesis and G/C tailing (cf. Figure 1): cDNA synthesis was primed with an AvaI/HindIII restriction fragment (Pos. 692-743) from cDNA clone FMDV2735 which had been isolated from a 6% polyacrylamide strand separation gel (8). Reverse transcription was followed by RNase treatment (panc. RNase 20µg/ml, 30 min., 37°C). Oligo-dA tails of 70-80 nucleotides were added to the 3'-ends of the cDNA using Terminal Deoxynucleotidyl Transferase. This oligo-dA tailed cDNA was annealed to the DNA fragment complementary to the primer fragment. By this a partially double stranded molecule was obtained having a single stranded oligo-dA tail at its 3'-side and a double stranded portion at its 5'-side reconstituting the origi-
nal HindIII site of FMDV pos. 743. This DNA molecule was then annealed to
the vector pUC9 (6), that had been cleaved by PstI, dT-tailed, HindIII cleaved
and purified by agarose gel electrophoresis. The DNA mixture was ligated and
transformed into competent E.coli C600 cells. Clones were screened for FMDV
inserts by colony hybridization with $^{32}$P-labelled FMDV RNA as described
elsewhere (7). Clones FMDV-2735 and -2615 were derived from RNA isolated
from low passaged FMDV (7 passages), all other clones were derived from a
virus-isolate after 64 passages in BHK cells.

**Nucleotide sequence analysis**

Restriction fragments were endlabelled and chemically degraded by the base-
specific cleavage methods according to (8). 5' endlabelling was used through-
out except for the 3' end of the genome where 3' endlabelling was employed
at a HpaI site located 68 nucleotides in front of the poly(A) tail. Thin
sequencing gels (0.04 cm), either 40 or 100 cm long, were dried prior to print-
ing, according to (9) to improve resolution of the bands in the sequencing lad-
ders.

**Computer analyses**

The derived nucleotide sequences were entered into a data base, where the
information was stored and processed using the computer programs of (10).
The alignment program from Krüger and Osterburg (unpublished) was based on
the algorithm from (11). The homology comparison in Figure 4b was based on
such computer-derived alignments scoring "similar amino acids" as one third of
identical amino acids.

**RESULTS AND DISCUSSION**

**Cloning of FMDV cDNA**

A set of cDNA clones from FMDV strain O4K (FMDV-144, -512, -703, -715,
-1034, -1448) has been described which cover the two thirds of the FMDV
genome from the VP3 coding region to the 3' end (5, cf. Figure 2). To obtain
cloned cDNA copies upstream of the VP3 gene single-stranded restriction frag-
ments from the existing cDNA clones (indicated by I and II in Figure 2) were
used in two steps (7, 12) to prime cDNA synthesis close to the missing parts
of the FMDV genome (see Methods and Figure 1). As a result the cloned part
of the FMDV genome was extended into the 3'-end of the poly(C) tract.

**Nucleotide sequence analysis**

In Figure 2 exact map positions of the cloned cDNA inserts used for nucleotide
sequence analysis and the sequencing strategy are depicted. The methods of
(8) were followed for 5' $^{32}$P-endlabelling and subsequent partial chemical
degradation of restriction fragments. In order to obtain unambiguous sequence data most regions were sequenced in several independent runs, and the entire sequence was analyzed on both DNA strands (Figure 2). Using overlaps of at least hundred nucleotides between subclones, a continuous sequence was generated over the 7915 cloned nucleotides starting with 11 C residues from the poly(C) tract and ending in 102 A residues as determined in clone pFMDV-3214c and pFMDV-512, respectively (Figure 3). All cleavage sites predicted in this sequence for the restriction enzymes used during the sequence analysis were also observed experimentally, except for a Clal site at position 5026. This site overlaps two Mbol sites which are known to be targets for adenosine methylation in E. coli (13).

Minor heterologies in the sequence were observed in overlapping regions from different clones (Figure 3, 14). Such point mutations are known to occur with high probability in populations of FMDV RNA (15) and in other viral RNAs because of the intrinsically imprecise mechanism of RNA replication (16). This variation was particularly obvious when cDNA from moderately and highly passaged virus (7 and 64 passages in BHK cells, respectively) was compared. No substitutions were found in 550 bases overlapping the clones FMDV-2735 and -2815 from the same low passaged virus (14).

The nucleotide sequence

The FMDV genome is divided by the poly(C) tract into two parts of different size and function. The small (S) segment to the 5' side (approx. 400 nucleotides (17)) is probably involved only in initiation of viral RNA replication, and the large (L) segment 3' of the poly(C) contains all the protein coding informa-
Fig. 2: Physical map of the FMDV genome, FMDV cDNA clones and strategy for the nucleotide sequence analysis. The top part shows the physical and the gene map of the FMDV RNA. Positions of the restriction sites of endonucleases used for 5' end-labeling are indicated on the second line. Below that, the restriction fragments used to prime cDNA synthesis (I and II) are illustrated as open arrows. Horizontal arrows show direction and extent of individual sequencing runs, grouped according to the clones they originate from. Dashed lines represent portions of clones or sequencing runs from which no information was obtained. The analysis of the clones FMDV-1034 and FMDV-144 has already been reported (26) and only the extent of the sequenced areas are indicated here. The restriction enzymes are represented by the letters: A, AvaI; B, BamHI; C, ClaI; E, HaeIII; F, HindIII; H, HhaI; I, Hpal; N, HindIII; P, HpalI; R, EcoRI; T, TaqI; U, PvuII; X, XhoI; Xb, XbaI.
The sequence of 7802 nonhomopolymeric nucleotides shown in Figure 3, represents the complete primary structure of the L segment. Assuming additional 150 bases for poly(C) and 400 for the S segment this corresponds to 94% of the FMDV genome indicating that the size of the FMDV genome is about 8500 nucleotides, i.e. 500 nucleotides longer than estimated from sizing-gels by (17).

This size correction is also supported by the sizing and the partial sequence analysis of cDNA copies that cover the as yet uncloned 5'-terminal part of the FMDV genome, i.e. the S segment and the poly(C) tract (7). So far, all attempts to clone this missing part of the viral RNA have been unsuccessful. These difficulties seem to be related to the internal poly(C) tract, which, although readily copied into cDNA (7), is probably highly unstable in E. coli, as shown for other (G:C)-homopolymers longer than 30 basepairs (19). In accordance with this notion we have only been able to clone 11 C residues from the 3' proximal part of the poly(C) tract which still contains interdispersed non-C nucleotides. The sequence obtained ...GCT(C)_{11}AAG... was confirmed by direct sequencing of the cDNA extending into the poly(C) tract (7). It is different, but shows similarities to the sequence ...

Coding regions and translation initiation sites

From the kinetics of the appearance of FMDV specific gene products it has been predicted that FMDV RNA contains a single long open translational reading frame for a 260 K polyprotein, which in turn is a precursor for all gene products. Our sequence reveals such a reading frame of 7035 nucleotides (pos. 766 to 7800) with a first possible initiation codon for the polyprotein at position 805 (see Figure 3). The coding region is preceded by 724 nucleotides of known sequence and by some additional 550 nucleotides of yet uncloned RNA comprising the poly(C) tract and the S fragment. It is followed by several stop codons in all three reading frames leaving 92 nucleotides untranslated in front of the poly(A) tail. Usually the 5' proximal AUG is used to initiate translation in an eucaryotic mRNA, which suggests that the ribosomes first recognize the capped 5' end of the RNA and then traverse downstream until an AUG is encountered (20). Clearly this model is not applicable to FMDV since translation of the major primary gene product starts at position 805 and to a lesser extent also at position 889 (3), which are the ninth and tenth AUGs downstream from the poly(C) (see Figure 3). These two start sites differ from other AUG codons in the sequence in that they are preceded at a short dis-
tance by a stretch of 11 pyrimidines, which are interrupted by no more than one purine residue and which also show significant base complementarity to the 3'-end of the 18S ribosomal RNA from eucaryotes (3). These pyrimidine runs are also present at the translational start sites of poliovirus and EMCV (21, 22). We therefore speculate that both features may be of importance for the recognition by ribosomes or initiation factors of an uncapped mRNA with the long untranslated leader sequence. Sequence complementary to the 18S rRNA terminus has also been noted for less pyrimidine rich sequences in other eucaryotic mRNAs (23).

The sequence of 1.3 kb preceding the polyprotein gene seems to be exceptionally long for a leader segment of a small, and otherwise compactly organized viral genome and suggests that additional short coding sequences may exist in this segment. At present the most likely candidate for such an unlinked FMDV gene is a sequence of 82 translatable codons that follows the first AUG after the poly(C) tract (pos. 209 in Fig. 3). The evidence for this hypothesis is two-fold. Firstly, a polypeptide of 10 K (tentatively named P10) has recently been identified among the products of in vitro translation of FMDV RNA by immunoprecipitation with an antiserum directed against the P10 amino acid sequence (Strebel et al, in prep.). Secondly, the translational start of the presumed P10 gene (pos. 208) is structurally very similar to other sites controlling translation initiation at internal positions of picornaviral RNAs in that it is also preceded at a short distance by the pyrimidine rich sequence noted above.

The deduced protein map

The long open reading frame encodes a polypeptide with a maximal size of 2332 amino acids and a calculated molecular weight of 258.9 K, in excellent agreement with the 260 k determined experimentally for the FMDV polyprotein (1). Its deduced amino acid sequence could be correlated in the P1 (P88) segment with all known sequence data from the structural proteins of FMDV O_K (3, 28, 27, amino acids underlined in Figure 3). Much less information
Fig. 4: Comparison of the FMDV and poliovirus genomes.

a) Schematic map of gene organization and protein processing. Open arrows indicate cleavages probably executed by cellular proteases, while filled arrows represent processing by the viral protease. The dashed arrows illustrate morphogenetic cleavages occurring during virus particle maturation. The FMDV polypeptides are termed as in Figure 3; the number of amino acids is given for each protein below the lines.

b) Sequence homology between corresponding parts of the two polyproteins. The degree of homology in different segments (see Methods) is: hatched 25-40%, crosshatched 40-65%, filled in 56%. Related gene products are connected by vertical lines and identified using the new general nomenclature for picornaviruses (according to the third Meeting of the European Study Group of Picornaviruses, Urbino, Italy Sept. 5-10, 1983).

was available for the non-structural proteins encoded in the P2 (P52) and the P3 (P100) precursors where only the VPg genes had been exactly mapped (24) and only approximate positions had been allocated for P34 and P58a (1) and for P12 and P20b (A.King, pers.comm.). A more complete protein map was established (14), using size estimations from SDS polyacrylamide gel electrophoresis and sequence homologies to poliovirus polypeptides for which the order was known (21). The exact coding limits of the individual functional proteins were predicted from the cleavage specificity for Glu (Gln)/Gly (Ser, Thr) linkages of the FMDV protease (see below) and also using very recent data from Grubman and collaborators, who determined amino acid sequences at the N-termini of the polypeptides synthesized in vitro from FMDV A12 RNA and corresponding to P52, P12, P34, P14, P20b (pers. comm.) and P56 (25). The order and limits of FMDV proteins thus obtained (Figure 4) were recently confirmed by immunoprecipitation of polypeptides of the predicted size from FMDV infected BHK cells, using antisera against bacterially synthes-
Table I: Predicted map positions and biochemical properties of FMDV polypeptides as deduced from the nucleotide sequence.

<table>
<thead>
<tr>
<th>Polypeptide</th>
<th>Map coordinate</th>
<th>Map position</th>
<th>No. of amino acids</th>
<th>Molecular weight</th>
<th>Net charge</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>P20a</td>
<td>L</td>
<td>805-1455</td>
<td>217</td>
<td>24367.4</td>
<td>-5</td>
<td>precursor</td>
</tr>
<tr>
<td>P16</td>
<td>L'</td>
<td>889-1455</td>
<td>189</td>
<td>21243.1</td>
<td>-7</td>
<td>?</td>
</tr>
<tr>
<td>P88</td>
<td>P1</td>
<td>1456-3615</td>
<td>720</td>
<td>79305.6</td>
<td>+13</td>
<td>precursor</td>
</tr>
<tr>
<td>VP4</td>
<td>la</td>
<td>1456-1662</td>
<td>69</td>
<td>7362.3</td>
<td>-3</td>
<td>capsid protein</td>
</tr>
<tr>
<td>VP2</td>
<td>lb</td>
<td>1663-2316</td>
<td>218</td>
<td>24410.4</td>
<td>+4</td>
<td>&quot;</td>
</tr>
<tr>
<td>VP3</td>
<td>lc</td>
<td>2317-2976</td>
<td>220</td>
<td>23746.4</td>
<td>-2</td>
<td>&quot;</td>
</tr>
<tr>
<td>VP1</td>
<td>ld</td>
<td>2977-3615</td>
<td>213</td>
<td>23840.5</td>
<td>+14</td>
<td>&quot;</td>
</tr>
<tr>
<td>P52</td>
<td>P2</td>
<td>3616-5079</td>
<td>488</td>
<td>54501.1</td>
<td>+8</td>
<td>precursor</td>
</tr>
<tr>
<td>(P12)</td>
<td>2b</td>
<td>3664-4125</td>
<td>154</td>
<td>16255.1</td>
<td>+2</td>
<td>?</td>
</tr>
<tr>
<td>P34</td>
<td>2c</td>
<td>4126-5079</td>
<td>318</td>
<td>35892.9</td>
<td>+8</td>
<td>?</td>
</tr>
<tr>
<td>P100</td>
<td>P3</td>
<td>5080-7800</td>
<td>907</td>
<td>100844.7</td>
<td>+21</td>
<td>precursor</td>
</tr>
<tr>
<td>(P14)</td>
<td>3a</td>
<td>5080-5538</td>
<td>153</td>
<td>17355.2</td>
<td>-5</td>
<td>?</td>
</tr>
<tr>
<td>VPg-1</td>
<td>3b-1</td>
<td>5539-5607</td>
<td>23</td>
<td>2604.5</td>
<td>+3</td>
<td>genome-linked</td>
</tr>
<tr>
<td>VPg-2</td>
<td>3b-2</td>
<td>5608-5679</td>
<td>24</td>
<td>2622.4</td>
<td>+4</td>
<td>&quot;</td>
</tr>
<tr>
<td>VPg-3</td>
<td>3b-3</td>
<td>5680-5751</td>
<td>24</td>
<td>2579.3</td>
<td>+3</td>
<td>&quot;</td>
</tr>
<tr>
<td>P20b</td>
<td>3c</td>
<td>5752-6390</td>
<td>213</td>
<td>23012.4</td>
<td>+9</td>
<td>protease</td>
</tr>
<tr>
<td>P56</td>
<td>3d</td>
<td>6391-7800</td>
<td>470</td>
<td>52760.9</td>
<td>+7</td>
<td>RNA polymerase</td>
</tr>
</tbody>
</table>

1) Nomenclature suggested for the picornaviral polypeptides at the 3rd Meeting of the European Study Group of Picornaviruses, Urbino Italy, Sept. 5-10 1983.

ized polypeptides that correspond to the respective segments of the polyprotein (Strebel et al., in prep.).

The biochemical properties of the FMDV proteins predicted from the amino acid sequences in Figure 3 are summarized in Table I and a schematic protein map is shown in Figure 4a. Our sequence-derived molecular weights often differ from earlier determinations, and correlate better with recent size estimations of FMDV proteins synthesized in infected cells (3, Strebel et al., in prep.).

The polyprotein sequence starts with two closely related "leader" proteins, L and L' ("P20a" and "P16") which differ by 28 amino acids at their N-termini and have molecular weights of 24.4 k and 21.2 k, respectively. The primary product following the leader protein L/L' in the polyprotein is P1 (MW 79.3k), formerly called "P88", the precursor of the capsid proteins which are arranged in the order VP4-VP2-VP3-VP1 (1). P2 (formerly "P52"), the precursor from
the middle part of the polyprotein has a calculated molecular weight of 54.5 K. It contains two stable proteins, 2b ("P12") and 2c ("P34"), with unknown functions. The N-terminal limit of P2 was originally set next to the carboxy-terminus of VP1 (cf. Figure 3) which had been accurately identified for FMDV O4K by (28). However, as shown for FMDV A12, the proteins P2 and P12 start 16 amino acids downstream from the C-terminus of VP1 (Grubman pers. comm., cf. Figure 3). The carboxy-terminal precursor P3 (formerly "P100", calculated molecular weight 100.8 K) comprises the sequence between amino acids 1426 and 2332. It is processed into six proteins: 3a ("P14") of 17.4 K, three VPgs (in tandem) of 2.3 K each, a candidate for a protease, 3c ("P20b"), of 23.0 K, and an RNA polymerase, 3d ("P56"), of 52.7 K.

Protease cleavage sites

The present map of the FMDV genome (Figures 2 and 4) predicts that at least 12 sites in the protein sequence need to be cleaved to give rise to mature viral gene products. Seven out of these sites show similar amino acid sequences

<table>
<thead>
<tr>
<th>Site</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP2/VP3</td>
<td>Pro-Ser-Lys-Glu/Gly-Ile-Phe-Pro</td>
</tr>
<tr>
<td>VP3/VP1</td>
<td>Ala-Arg-Ala-Glu/Thr-Thr-Ser-Ala</td>
</tr>
<tr>
<td>P14/VPg-1</td>
<td>Pro-Gln-Ala-Glu/Gly-Pro-Tyr-Ala</td>
</tr>
<tr>
<td>VPg-1/VPg-2</td>
<td>Pro-Gln-Gln-Glu/Gly-Pro-Tyr-Ala</td>
</tr>
<tr>
<td>VPg-2/VPg-3</td>
<td>Val-Val-Lys-Glu/Gly-Pro-Tyr-Glu</td>
</tr>
<tr>
<td>VPg-3/P20b</td>
<td>Ile-Val-Thr-Glu/Ser-Gly-Ala-Pro</td>
</tr>
<tr>
<td>P20b/P56</td>
<td>Pro-His-His-Glu/Gly-Leu-Ile-Val</td>
</tr>
</tbody>
</table>

suggesting that they are cleaved by a single (viral) protease, recognizing the consensus sequence Glu/Gly (Ser, Thr). This specificity is similar to the one displayed by the poliovirus protease which cleaves between Gln and Gly residues at eight out of eleven processing sites in the polio polyprotein (21, 28, Fig. 4a). The sequences around the remaining five cleavage sites in FMDV

<table>
<thead>
<tr>
<th>Site</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>P20a/VP4</td>
<td>Ser-Gln/Asn-Gly-Ser-Gly/Asn-Thr-Gly-Ser</td>
</tr>
<tr>
<td>VP4/VP2</td>
<td>Ala-Leu-Leu-Ala/Asp-Lys-Asn-Thr</td>
</tr>
<tr>
<td>VP1/P52</td>
<td>Lys-Gln-Thr-Leu/Asn-Phe-Asp-Leu</td>
</tr>
<tr>
<td>P12/P34</td>
<td>Ala-Glu-Lys-Gln/Leu-Lys-Ala-Arg</td>
</tr>
<tr>
<td>P34/P100</td>
<td>Ile-Phe-Lys-Gln/Ile-Ser-Ile-Pro</td>
</tr>
</tbody>
</table>

show little sequence homology to each other and are thought to be recognized by cellular proteases. However recent experiments from this laboratory (unpublished data) suggest that the L polypeptide may also be involved in the processing of at least the L/P1 junction.
In FMDV there are 14 Glu-Gly, 3 Glu-Ser, and 9 Glu-Thr dipeptides in the polypeptide sequence which may be substrate for the FMDV protease. Of these only five, one, and one respectively are utilized, according to the protein map (Figures 3 and 4a). Therefore, not simply the primary sequence but also a certain conformation of the amino acid sequence must be recognized by the viral enzyme.

**Homology to poliovirus**

Gene organization and processing mechanisms predicted for FMDV from the nucleotide sequence in Figure 3 are summarized in Figure 4 and compared to those from poliovirus, a member of the enterovirus family. As outlined in Figure 4a, the overall organization is well conserved between the two genomes. This map is also similar to the approximate gene map established for EMCV (29), another well studied virus belonging to a third genus of picornaviridae. In an overall comparison FMDV differs from poliovirus by an increase in size of its genome by about 1000 nucleotides most of which are added in the 5’ proximal part of the genome. Within the polypeptide region the two genomes differ drastically only by the addition of the L gene and two extra VPg genes in FMDV, and by the addition of a third extended gene (2a or 2b) in the central part of the polio genome. Other corresponding genes often differ in size. Thus, the capsid proteins are larger in poliovirus, while FMDV has expanded its nonstructural proteins in the P3 segment (Figure 4).

The alignment of the two genomes was refined using sequence homologies between functionally corresponding segments. Using a dot matrix program significant homology was detected on the amino acid level throughout most part of the coding region, and to a lower extend also on the nucleotide level indicating common structural features of general functional importance (14). As shown in Figure 4 these sequence homologies are very high in certain parts of two non-structural proteins, the polymerase (3d), and protein 2c (x/P34). Less, but still high homology was detected between the protease genes and the capsid proteins VP2 and VP3, indicating that the functional specificity of these proteins was less stringently conserved during picornaviral evolution. The evolutionary divergence is most pronounced in proteins 1d, 2a/2b and 3a. Protein 1d (VP1) is the capsid protein most exposed at the surface of the virion and, as a consequence of the pressure of the host’s immune system, its sequence is highly variable also between different aiptoviruses giving rise to seven serotypes and many more subtypes. In contrast proteins 2a/2b and 3a, although variable between FMDV and poliovirus, are highly conserved between two FMDV serotypes as exemplified by a comparison of FMDV O1K and C1O.
(unpublished results). Therefore, we conclude that these latter proteins play an important role in the FMDV life cycle but did coevolve with their targets which are most likely FMDV-specified molecules. Following the same argument, we predict that the more generally conserved picornaviral proteins (2c and 3d) depend in their functions on the interaction with host factors of conserved structure. In this context it is of note that sequences common to FMDV and poliovirus in proteins 2c and 3d are also present in Cow Pea Mosaic Virus, a plant virus which has thus been correlated with picornaviruses (30). The comparison of the two picornaviral genomes in Figure 4b also indicates that these differ in size predominantly in regions with low sequence homology. Sometimes extended blocks of nucleotides are also found inserted into well conserved genes like in gene 1b (VP2). Together with the addition of complete genes, like the L gene or the extra VPg genes in FMDV, these data indicate that evolution of picornaviral genomes involved the insertion or deletion of RNA segments of several hundred nucleotides.

CONCLUSIONS

We report the nucleotide sequence of the complete coding part of an aphthovirus genome. This sequence is useful in several ways for further studies of the viral life cycle. It has already allowed exact predictions of the genome organisation of these viruses and of the amino acid sequences of their gene products. This information facilitates the identification and characterization of these proteins, e.g. by antisera elicited against synthetic peptides. In addition, cDNA segments from specific parts of the genome can be expressed in E. coli and used as specific antigens free of any other viral protein, or be used as substrate for processing enzymes. Finally, FMDV genes and genomic signals can be transferred well defined as cDNA copies into animal cells and their functions analyzed. Such studies should also answer questions as to the function of the 1300 nucleotides long leader sequence preceding the polyprotein gene in the FMDV RNA.

During preparation of this manuscript, the complete coding sequence for the FMDV polyprotein has been reported for strain A_{10} (31). While there are extensive serotype related sequence variations in the structural proteins (3, 12) the sequence differs in the non-structural genes from the O_{1}K sequence by 331 nucleotides (7.9%) and 44 predicted amino acids (3.1%), provided that we neglect three T (U) residues (positions 5978/79, 6013/14, 6016/17), appearing in duplicate in the A_{10} sequence published.
ACKNOWLEDGEMENTS

We are indebted to Joern Wolters and Oliver Steinau for the computer analyses, and to Gudrun Feil for technical assistance. We also thank Dr. W. Keller and Dr. K. Strohmaier for gifts of FMDV RNA. We acknowledge Dr. A. King and Dr. M. Grubman for the communication of unpublished data. This work was supported by research grants from Biogen S.A. and the Deutsche Forschungsgemeinschaft (Forschergruppe Genexpression).

REFERENCES

7 Strelbel, K. (1982) Diplomarbeit, University of Heidelberg
14 Forss (1983), P.H. thesis, University of Heidelberg