Volume 11 Number 22 1983 Nucleic Acids Research

Structure of the FMDV translation initiation site and of the structural proteins

Ewald Beck, Sonja Forss, Klaus Strebel, Roberto Cattaneo and Gudrun Feil

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

Received 16 August 1983; Revised and Accepted 18 October 1983

ABSTRACT

A cDNA clone of Foot and Mouth Disease Virus (FMDV), strain C1, has been sequenced. The limits of the structural genes were defined by comparison with the available protein data. We identified two potential translation initiation sites for the viral polyprotein separated by 84 nucleotides. We suggest that these two initiation sites could be used to express two proteins differing only at the N-termini, P16 and P20a. This model is supported by the fact that antiserum against a bacterially synthesized polypeptide corresponding to the anterior region of the polyprotein precipitates specifically both P16 and P20a. Comparison of the C1 sequence with two other serotypes, O1K and A10 revealed variability in the major immunogenic structural protein, VP1, and also in two other capsid proteins, VP2 and VP3. P16/P20a, VP4, and the N-terminal part of the precursor of the nonstructural genes, P52, are rather conserved between the different FMDV strains.

INTRODUCTION

Foot and Mouth Disease Virus (FMDV) is immunologically divided into seven serotypes (O, A, C, SAT1, SAT2, SAT3, and AsiaI), which are further classified into about 70 partially crossreacting subtypes (1). In addition, minor changes rapidly accumulate during passage of the virus in the field (2). Therefore FMDV is an interesting model for studies of the molecular basis of antigenic drift.

The viral RNA is a single-stranded molecule of plus-strand polarity, i.e. equal to the mRNA, consisting of about 8000 nucleotides. The mechanism of protein synthesis in these viruses differs from most cellular mRNAs particularly at the initiation event in that the viral RNA lacks a cap (3) and has a large untranslated portion upstream from the initiation site, making this system an interesting model for studying translational control mechanisms. After initiation the ribosomes proceed continuously to a termination site close to the 3'-end of the genome. The initial translation product, a polyprotein with a (calculated) molecular weight of about 260 K is cleaved in statu nascendi probably by a cellular protease, giving rise to four primary precursor

© IRL Press Limited, Oxford, England.

proteins. The precursors are further processed by secondary cleavage events, mediated by a viral protease, generating the stable structural and nonstructural polypeptides. The VP4/VP2 junction is cleaved in a final separate event accompanying virus assembly (4).

To elucidate the structures involved in initiation of protein synthesis and clarify the basis of FMDV antigenic variation we cloned and sequenced the translational start site of the polyprotein and the structural genes for the viral capsid proteins of FMDV C1. The C1 sequence was compared with the sequence of serotype O1K (Forss et al., submitted) and A10 (5, this sequence starting some 450 nucleotides downstream from the translation initiation site). The sequence also revealed the complete amino acid sequence of the foremost located gene products P16 and P20a.

MATERIALS AND METHODS

(a) FMDV RNA and enzymes. FMDV C1 RNA, isolated from virus grown in BHK-cells, was kindly provided by K. Strohmaier, Tübingen. Restriction endonucleases (*Hha*I, *Hin*fI, *Hin*dIII, *Hpa*II, *Eco*RI, *Taq*I) were prepared essentially as described (6) or purchased from New England Biolabs (*Ava*I). Calf intestinal phosphatase and polynucleotide kinase were from Boehringer Mannheim GmbH. [γ-^{32}P]ATP of high specific activity was prepared according to (7).

(b) cDNA of FMDV C1 was synthesized by extension of a *Hae*III fragment from a single-stranded derivative of the O1K cDNA clone pFMDV-715 (8) covering the viral RNA from positions 3793 to 3882, rendered double-stranded by *Escherichia coli* DNA polymerase I as described (8) and ligated to synthetic *Eco*RI-linkers. As the primer fragment covered a *Hin*dIII site ∿ 240 bp downstream from the end of VP1, the cDNA was cleaved with *Eco*RI and *Hin*dIII, sized by gel electrophoresis and fragments > 1000 nucleotides were ligated into the vector pBR322 via the *Eco*RI and *Hin*dIII sites. The largest clone recovered, pFMDVC1-9, contained a 3000 bp fragment extending from the *Hin*dIII site at position 3832 to a (natural) *Eco*RI site at position 894 (see Fig. 1).

(c) Nucleotide sequence analysis. DNA sequencing was performed essentially as described (9). Sequencing gels were dried in order to enhance the sharpness of the bands and to shorten exposure times (10). Sequences were stored and processed using the computer programs of (11).

(d) Immunoprecipitation of P16 and P20a. To 80 µl of a nuclease treated rabbit reticulocyte lysate (New England Nuclear) 10 µl (135 µCi) of ^{35}S-methionine, 10 µl of 0.5 M potassium acetate and 3 µl of FMDV O1K RNA (= 3 pmoles) were added and the mixture incubated at 30° C for 30 min. The reaction was

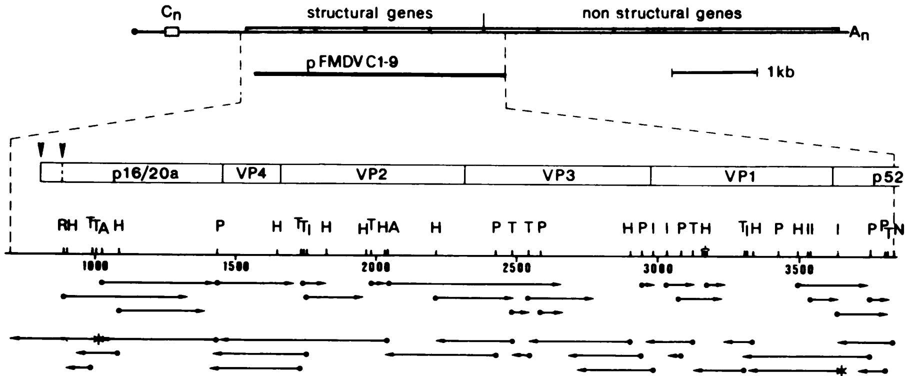

Fig. 1 Location of the cDNA clone pFMDVC1-9 in the genome of FMDV and sequencing strategy applied. C_n and A_n refer to the internal poly(C) tract and the polyadenylated 3'-end of the viral RNA. The limits of the genes are indicated. The arrowheads point to the possible start sites of the polyprotein discussed in the text. Restriction sites used in the sequence analysis are mapped and the enzymes abbreviated as follows: A = *Ava*I, H = *Hha*I, I = *Hin*fI, N = *Hin*dIII, P = *Hpa*II, R = *Eco*RI, and T = *Taq*I. Numbering of the FMDV RNA is according to (8). Direction and extent of the individual sequencing runs are indicated by horizontal arrows. Sequences obtained by primer extension with reverse transcriptase on the C1 RNA are marked by asterisks.

stopped by addition of 0.9 ml 0.15 M NaCl, 10 mM sodium phosphate buffer pH 7.5, 1 % NP40, 0.5 % sodium desoxycholate, and 0.1 % SDS. 50 µl aliquots were used for immunoprecipitation with antiserum against a bacterially synthesized polypeptide corresponding to the putative P16 encoding region of FMDV C1 (K. Strebel et al., manuscript in preparation).

RESULTS AND DISCUSSION

Nucleotide sequence. The location of the cDNA clone pFMDVC1-9 on the FMDV genome is outlined in Fig. 1. This clone covers the complete structural gene region of the virus in addition to most of the preceeding P20a/P16 coding region. The nucleotide sequence was determined by the chemical method (9) making use of different restriction endonuclease sites contained in the cDNA as shown in the lower part of the figure. The 5'-end of this clone maps at position 892 (numbering as (8)), and the sequence of the 170 nucleotides upstream from this position was derived directly by primer extension of an end-labelled DNA fragment (*Eco*RI, pos. 892 - *Ava*I, pos. 1027) on the viral RNA. Altogether, a nucleotide sequence of 3118 bases was obtained, which extends from position 719 to position 3837 on the FMDV genome (Fig. 2). The nucleo-

```
              10        20        30        40        50        60        70        80        90       100

 719 TTCTATAAAAGCGCTCGGTTTAAAAAGCTTCTATGCCTGAATAGGTAACCGGAGGCCGGCACCTTTCC-TTTACAATTAATGACCCTATGAATACAACTG C1
 719     - T    T  C A                C            G                     T    T   CC C   A AC             O1K

 818 ACTGTTTTACCGCTGTGGTAAACGCCATCAGAGAGATAAGAGCACTTTTCCTACCACGGACTACAGGAAAAATGGAATTCACGCTGCACGACGGCGAGAA C1
 818          T    T     C G  T           T A         T   T    C  C     G         C G  A   T  A    T      O1K

 918 AAAGGTCTTTTACTCCAGACCCAACAACCAAGACAACTGCTGGTTGAACACCATCCTTCAGTTGTTCAGGTACGTCGATGAACCTTTCTTCGACTGGGTC C1
 918 G   AC            G           C                  G       C                 T  A     A                O1K

1018 TACAATTCGCCCGAGAACCTCACGCTTGAAGCCATCAAGCAACTGGAGGAACTCACAGGGTTGGAGTTGCGCGAGGGTGGACCGCCCGCCCTTGTGATTT C1
1018     G      T                             GT       T        AC T  AC   AT           A  T  T  C     C  O1K
                                                                                                    P20α/P16
1118 GGAACATCAAGCACCTGCTCCACACTGGCATCGGTACCGCTTCGCGACCCAGCGAGGTGTGTATGGTGGACGGTACGGACATGTGTCTTGCTGACTTTCA C1
1118               T          C        C     C                    C        T              CT G     T  C   O1K
1146               start A10 —►T     C                          A     C                    T  C           A10

1218 TGCAGGCATTTTCATGAAAGGACAGGAACACGCTGTCTTTGCGTGTGTCACCTCCAACGGGTGGTACGCGATTGATGACGAGGACTTCTACCCATGGACG C1
1218    T         C T  G  G  A           G                                      C  T              C       O1K
1218 C                                A  G  C               AG                  C           T     G       A10

1318 CCAGACCCGTCTGATGTCCTGGTATTTGTTCCATACGATCAAGAGCCACTCAATGAAGGATGGAAGGCCAACGTTCAGCGAAAGCTCAAGGGAGCTGGGC C1
1318   G        C  C  T     G     C  G           A        C GG A       A     G     A  C        A  G     A O1K
1318   T     A  G  C   T          C  G           A          GG AC      AA AC G      AAG           T       A10

1418 AATCCAGCCCAGCGACCGGTTCGCAGAACCAATCTGGCAACACTGGCAGCATAATTAACAACTACTATATGCAGCAGTACCAAAACTCCATGGACACACA C1
1418        T           C                    T              A           C            T                    O1K
1418  G           A     C           G                                   C        A     G      T   AG      A10
                                                                                                     VP4
1518 ACTCGGCGACAACGCCATCAGTGGAGGCTCCAATGAAGGCTCCACGGACACAACCTCTACACACACAACCAACACCCAGAACAACGACTGGTTTTCCAAA C1
1518 G  T  T        A              T  C  G        C           C  C                       T        C       O1K
1518 G  T  T      TA                   C  G                 T  A                    A                 A    A10

1618 CTTGCCAGTTCAGCCTTCAGCGGTCTTTTCGGCGCCCTTCTCGCTGATAAGAAAACGGAGGAAACCACTCTCCTTGAAGACCGCATTCTCACTACCCGTA C1
1618         C  T  T                    T        C  C     G  A     G           C           C     C        O1K
1618            G  T  T C      G     T  A        C  C     G     A  G  T  G  T  G           C           C  A10

1718 ACGGGCAGACGACCTCGACAACCCAGTCGAGCGTCGGAGTCACATTCGGGTATGCAACTGCTGAAGATAGCACGTCTGGACCCAATACATCTGGTCTAGA C1
1718     C  C  C  G              A     T          A      C     A         TTTGT AGC     G  C  T  C     C   O1K
1718        C  C  T     C           T  G        G AT     CT C    AG      CA GTTG    G     C     G  CT     A10

1818 GACGCGCGTTCATCAGGCAGAGAGGTTTTTCAAAATGGCACTTTTTGATTGGGTTCCTTCACAAAATTTTGGACACATGCACAAAGTTGTTCTGCCCCAT C1
1818 A  CA A   GTG        AC            CCCAC  C  C .C     CA CAGTG CTCA  C    GTTGC  CTCC  G AA  C  GACC O1K
1818       G  GGTG           A        G A  TTT G     C   ACAA GGACA  CC       T  T  ACA   C G AG  T   ACC A10

1918 GAACCAAAAGGTGTTTACGGGGGTCTCGTCAAGTCATACGCGTACATGCGCAATGGCTGGGACGTCGAGGTGACTGCTGTTGGAAACCAGTTCAACGGCG C1
1918   C AC        C     CA C  GACTG C  G  T  A  T   A A  C        T        C  C  G     C              A  O1K
1918   T ACC C     C T    CACT G  GG C     T  A  T   A G  C        T  T     ATG   C  C  C                 A10
                                                                                                     VP2
2018 GCTGCCTCCAGGCGGCGCTCGTCCCCGAGATGGGCGATATCAGTGACA-GGGAAAAGTACCAACTAACCCTTTACCCCCACCAGTTCATCAACCCACGCA C1
2018  G     T T  TC  AA G  A  A   C TTATTC    CAAA G -     CT      G  C  A    T   T                       O1K
2018  G     T T  T   CA G  G  A   -   AA GC  TT AC   C T     A     G  T       T   A        T  T G   CA  A  A10

2117 CCAACATGACGGCACACATCACTGTGCCCTACGTGGGTGTCAACAGGTATGACCAGTACAAACAGCACAGGCCCTGGACCCTCGTGGTCATGGTTGTCGC C1
2118  G        T  G        A       TT  T  C      C C  C           GGTT    A   T        T     T        A   O1K
2117  T        T  C        G  A  G  TC T     G        C  T        GA A    AA  T     A  G  T        AC AT  A10

2217 GCCACTCACCACAAACACAGCAGGTGCCCAACAGATCAAAGTGTATGCCAACATAGCCCCAACCAACGTGCACGTGGCAGGTGAGCTCCCCTCCAAGGAG C1
2218 C  T  G   GTC     T A        CT        G              T                    C  G      T T  T          O1K
2217 A  C     GGTC G  ACA T CC    C   A     G  C  C        T         T    T        T  A     T     G  A    A10

2317 GGGATCTTCCCCGTTGCGTGTTCTGACGGTTACGGCAACATGGTGACAACTGACCCGAAAACGGCTGACCCTGCCTACGGGAAGGTTTACAACCCCCCTC C1
2318 - A  A        G  C   AGC     C  T  TGG C       C  G        G           C TT  T     A  G T         C  O1K
2317      T     A     A  CG A           AGGAC          A           A          TT     T     G  T     G  CA A10

2417 GGACTGCTCTGCCGGGGCGGTTCACAAACTACCTGGATGTTGCCGAGGCTTGTCCCACCTTCCTGATGTTCGA--G-AACGTACCTTACGTCTCAACACG C1
2417  C ACCAGT           T  T  C   CT   T     G  T     A  C  G  G  T   CGC     GG TGG      G     GA C  GAA O1K
2417 A   CAACTAC   C     C  T      CTAT     C  G     A  A           T  TCGT     CGATGGGAA   G      GTT  G   A10
                                                                                                     VP3
2514 AACTGACGGGCAAAGGCTACTGGCCAAGTTCGACGTGTCGCTGGCAGCGAAACACATGTCAAACACCTACTTGGCCGGCCTGGCCCAGTACTACACACAG C1
2517    G   TC G C   G G  T  TC    T   A    TT       A     A             T C C  A  T  T  G                O1K
2517 GG A     ACACCC T TT          T  T  C  C  T  C  A           C     A   C AT A  GA T  A                A10

2614 TACACTGGGACAATCAACCTACACTTCATGTTCACTGGGCCGACCGACGCGAAAGCTCGGTACATGGTGGCGTACGTGCCCCCTGGCAT-G-GA-CGCAC C1
2614     G   C  C        G              A  A  C  T        G  G  T        T  C    CC  A  A      - - -GC  G  O1K
2614    T    T  T        G              A  CT C  T   T A     C  C           T   A C  G T   GG G A  CGC G  A10
```

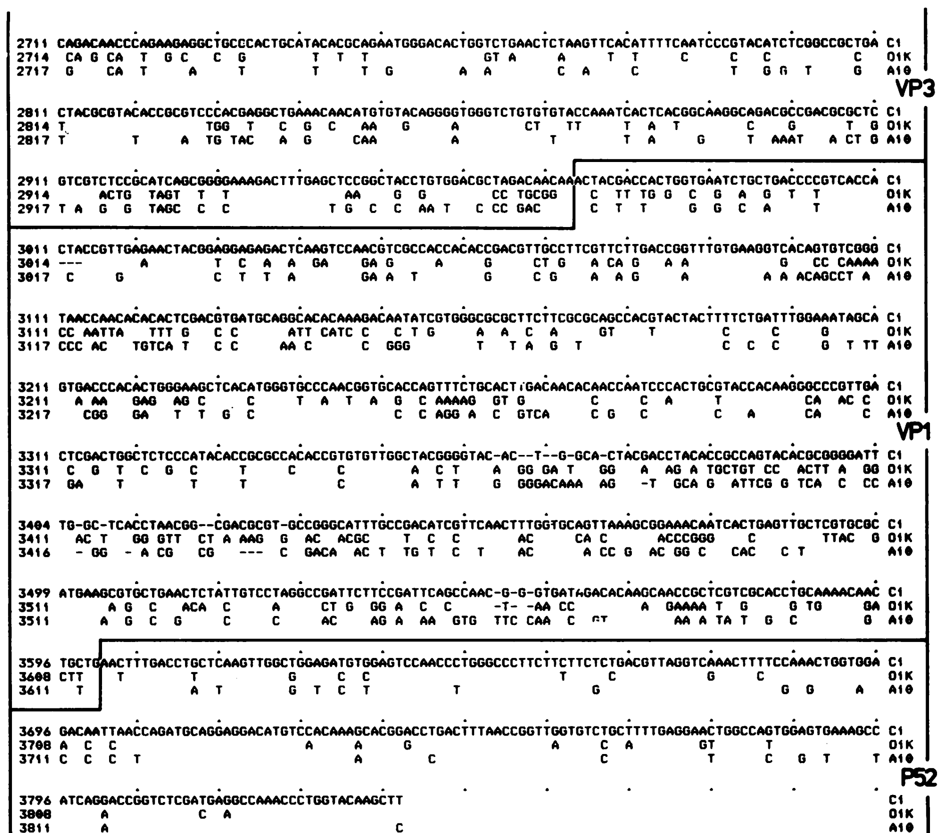

Fig. 2 Nucleotide sequence of the region in the FMDV genome encoding the start site for translation of the polyprotein and the structural proteins. The C1 sequence is shown completely. The O1K sequence (S. Forss et al., submitted and (28)) and the A10 sequence (5) have been included in the figure. Only differences to the C1 sequence are specified. Dashes represent deletions used to optimize alignment of the sequences. The limits of the proteins are indicated.

tide sequence of the serotypes O1K (Forss et al., submitted) and A10 (5), starting at position 1146, have been aligned to the C1 sequence.

Translation initiation site. Although the nucleotide sequence does not contain the 5'-end of the FMDV genome but starts some 700 bases downstream from the poly (C) tract (K. Strebel, manuscript in prep.), it includes the start site of the polyprotein. The first 87 nucleotides of the sequence contain several stop codons in all three reading frames. The methionine codon at pos. 805, however, initiates a reading frame which is open to the end of

```
                          -30           -20           -10            1         +10
                           .             .             .             .           .
             C1:   C C G G C A C C T T T C C - T T T A C A - A T T A A T G A C C C T A T G A A T A C A A
start 805
            O1K:   C C G G C A C C T T T C C T T T T A T A - A C C A C T G A A C A C A T G A A T A C A A

             C1:   A G A G C A C T T T T C C T A C C A C G G A C T A C A G G A A A A A T G G A A T T C A
start 889
            O1K:   A A A G C A C T T T T T C T A T C A C G C A C C A C A G G G A A A A T G G A A C T G A

start polio:       C A A C A G T T A T T T C A A T C A G A C A A T T G T A T C A T A A T G G G T G C T C
```

Fig. 3 Comparison of potential start sites for translation of the viral polyprotein from FMDV, serotypes C1 and O1K and poliovirus (12). The start codons and the pyrimidine blocks mentioned in the text are underlined. In the upper lines insertions (dashes) have been introduced to optimize alignment.

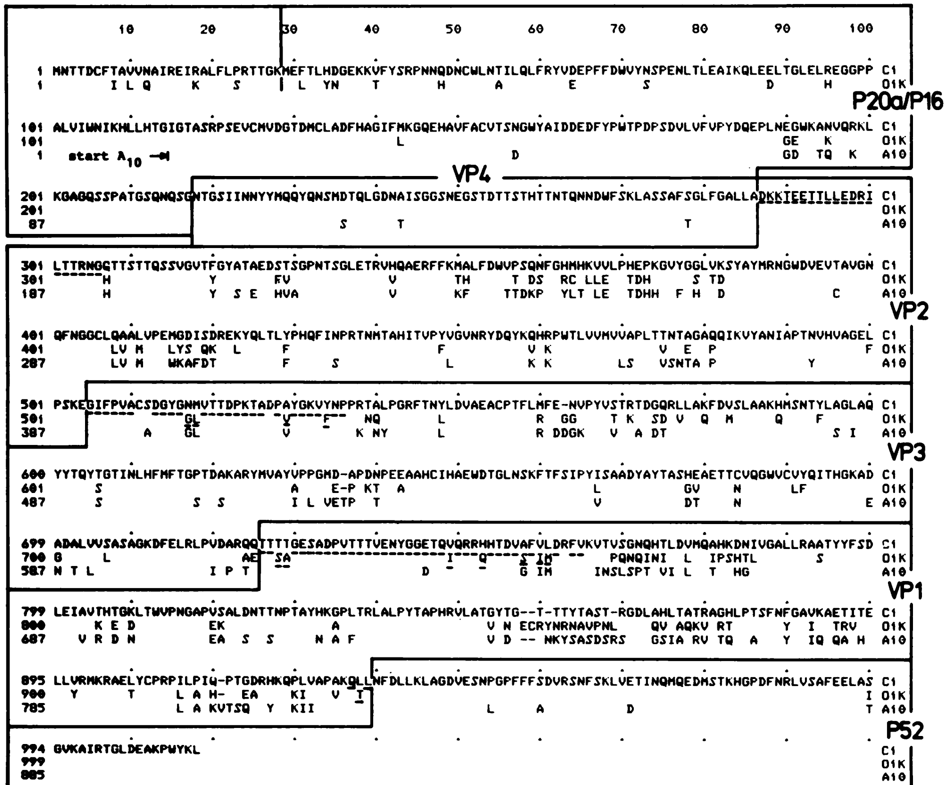

Fig. 4 Amino acid sequence of the polyprotein as deduced from the FMDV cDNA sequence in Fig. 2. The corresponding A10 and O1K sequences are also shown. Only differences to C1 are specified. Deletions (dashes) have been introduced into the sequence to optimize alignment. The limits of the proteins and the two potential start sites of translation are designated. Underlined amino acids have been experimentally determined for FMDV serotype O1K (15, 16).

VP1 ↓ P52	strain:
K Q L L N F D L	A_{24} (a)
K Q - L N F D L	A_{12} (b)
K Q L L N F D L	A_{10} (c)
R Q L L N F D L	A_5 (d)
K Q L L N F D L	C_3 (a)
K Q L L N F D L	C_1 (d)
K Q T L N F D L	O_1BFS (a)
K Q T L N F D L	O_1K (e)
+ Q X L N F D L	consensus

P20a ↓ VP4	strain:
N Q S G N T G S	C_1
N Q S G N T G S	A_{10} (c)
N Q S G N T G S	O_1K (e)
N Q S G N T G S	consensus

Fig. 5 Comparisons of the amino acid sequences at the VP1/P52 and P20a/VP4 junctions, respectively. The arrows indicate where cleavage, presumably by the same cellular protease, occurs. The sequences for the other strains are from a) (22); b) (25); c) (5); d) (21); e) Forss et al., submitted.

the determined nucleotide sequence. A second possible initiation site for translation is located 28 codons downstream from the ATG (pos. 889). The corresponding sequence of FMDV serotype O1K also contains these two possible start codons at identical positions, and strikingly both ATG codons are preceded by long stretches of pyrimidine residues (Fig. 3) which show a significant complementarity to a purine sequence close to the 3'-end of eucaryotic 18S rRNA. Interestingly the poliovirus polyprotein start site (12) has a similar structure (Fig. 3). Moreover, nucleotide variations between the C1 and the O1K sequence in the region downstream from the first start codon are mostly silent or cause conservative amino acid exchanges (Fig. 4) suggesting that this region encodes a functional gene product.

Cleavage site at the P20a/P88 junction and structure of P16/P20a proteins. P16 and P20a, located at the N-terminus of the FMDV poylprotein, are cleaved from the polyprotein probably by a cellular protease. Since the same protease is expected to cleave the junction P88/P52, some homology between the sequences at these junctions might be expected. The "consensus" sequence Q-X-L/N-F-D deduced from the P88/P52 junction of eight different FMDV strains (listed in Fig. 5) suggests the sequence Q-S-G/N-T-G as the most likely candidate. The same cleavage site was also proposed by (5).

P16 and P20a are synthesized in a FMDV directed reticulocyte translation system as earliest gene products and contain similar peptides (14). From pulse chase experiments and our observation that the ratio of the two proteins is not changed at prolonged reaction times (data not shown) it is unlikely that P16 is a processing derivative of P20a. As we have identified two putative initiation sites for translation we propose that P20a and P16 differ only at

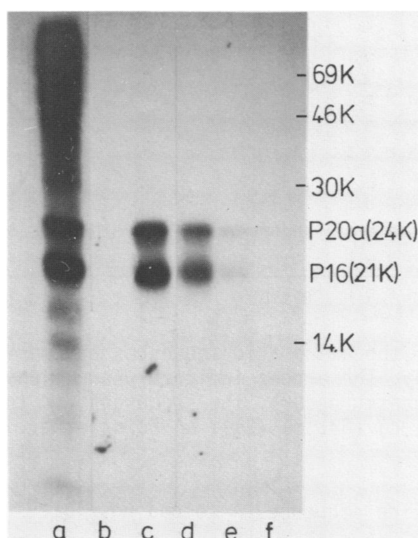

Fig. 6 Immunoprecipitation of P16 and P20a. FMDV RNA was in vitro translated in a rabbit reticulocyte lysate for 30 min at 30° C in the presence of ^{35}S-methionine. Aliquots of the translation mixture were immunoprecipitated with antiserum against a bacterially synthesized polypeptide corresponding to the putative P16 encoding region and analysed on a 12.5 % polyacrylamide gel. Lane (a) lysate untreated; lane (b) lysate immunoprecipitated with preimmunserum; lane (c) lysate immunoprecipitated with anti-P16 serum; lanes (d) - (f) lysate immunoprecipitated in the presence of 5 ng, 50 ng, and 500 ng of the bacterially expressed P16 protein. The positions of P16 and P20a and of marker proteins (bovine serum albumin 69 K, ovalbumin 46 K, carbonic anhydrase 30 K, and lysozyme 14.3 K) are indicated.

their N-termini. To substantiate this hypothesis we inserted the anterior part of pFMDVC1-9 encompassing the putative P16 coding region into a bacterial expression system and used the artificially synthesized protein to induce antiserum in rabbits (K. Strebel et al., manuscript in prep.). This antiserum precipitates specifically P16 as well as P20a (Fig. 6) demonstrating that the two proteins are related to each other and that they derive from the anterior region of the polyprotein. In our hands the apparent molecular weight of P16 and P20a on SDS gels is 20 K and 24 K, respectively (see Fig. 6), what agrees with the coding capacity of the genome between the two start codons and the suggested P20a/P88 junction at position 1455/56 (21 K and 24 K, cf. Table 1).

Structure of P88. Downstream of P20a/P16, the polyprotein encodes P88, the precursor of the structural proteins VP1 - VP4, arranged in the order NH_2 - VP4 - VP2 - VP3 - VP1 - COOH (14). Correlation of the nucleotide sequence and available amino acids sequence data results in a "P88" of 79 K (calculated mo-

Table 1

Properties of FMDV proteins as derived from the nucleotide sequence

Protein	P20a	P16	VP4	VP2	VP3	VP1	
No. of	217	189	69	218	219	209	C1
resi-	217	189	69	218	220	213	O1K
dues	-	-	69	218	221	212	A10
M_r	24443	21321	7362	24236	23973	22729	C1
	24367	21243	7362	24410	23746	23840	O1K
	-	-	7360	24649	24213	23248	A10
net	-7	-9	-3	0	-2	+14	C1
charge[a)]	-5	-7	-3	+4	-2	+14	O1K
	-	-	-3	+7	+3	+13	A10

[a)]difference between positively charged residues (Arg, His, and Lys) and negatively charged residues (Asp and Glu)

lecular weight for unmodified protein) rather than the 88 K reported in the literature.

The N-terminus of the smallest of the capsid proteins, VP4, is chemically blocked in FMDV (5) as well as in poliovirus (12). Hence, its N-terminal amino acid sequence has not been successfully analysed in any picornavirus. The coding region for VP4 was roughly located in the polyprotein sequence upstream from VP2, based on size estimations, suggesting a molecular weight of 8 K (15); if the P20a/P88 junction is situated at nucleotide 1455/56, VP4 would have a molecular weight of 7.4 K. Comparison of VP4 and the corresponding poliovirus protein shows polypeptides of almost identical size (69 and 68 amino acid residues, respectively). The C-terminal amino acid sequence of VP4 as determined from the nucleotide sequence, is highly hydrophobic, possibly preventing exposure of this part of the protein to the solvent. This could be the reason for the very late proteolytic cleavage of the VP4/VP2 junction in the virus maturation.

The proteins VP1 - VP3 are very similar in size (see Table 1) and ratio of charged amino acid residues. VP1 carries an extremely positive net charge (+ 14) which could explain its abnormal mobility on SDS gels causing some confusion in the nomenclature of VP1 and VP3.

Partial amino acid sequence data exist for three of the structural proteins of serotype O1K (15, 16). Since the N-termini of the structural proteins are

Table 2
Variation between FMDV C1, O1K, and A10

(A) Nucleotide exchanges in % (fraction silent exchanges)

Serotype	P20a[a)]	VP4	VP2	VP3	VP1	P52[b)]
O1K/C1	14 (82)	10 (100)	25 (63)	25 (69)	34 (46)	12 (86)
O1K/A10	13 (88)	13 (85)	26 (63)	27 (67)	34 (52)	10 (96)
A10/C1	13 (75)	12 (83)	29 (57)	27 (69)	32 (52)	11 (84)
O1K/C1/A10	19 (82)	17 (81)	36 (60)	36 (66)	45 (21)	16 (89)

(B) Amino acid exchanges in %

Serotype	P20a[a)]	VP4	VP2	VP3	VP1	P52[b)]
O1K/C1	6	0	17	16	30	1
O1K/A10	4	4	17	19	32	5
A10/C1	6	4	23	18	29	5
O1K/C1/A10	7	4	26	24	39	5

[a)] sequence corresponding to the C-terminal 104 amino acids of P20a
[b)] sequence corresponding to the N-terminal 79 amino acids of P52

largely homologous between the serotypes (see below) these data could be unambiguously correlated with the C1 sequence. The experimentally determined amino acid residues have been underlined in Fig. 4. The gene limits in serotype A10, which were determined also by protein sequencing (26), coincide with the gene junctions established for C1.

The junctions between the structural proteins in P88 are cleaved by a virus encoded protein, except for the capsid maturation cleavage between VP4 and VP2 which occurs late in infection and is performed possibly by still another enzyme. The remaining two junctions VP2/VP3 and VP3/VP1 in the structural gene precursor consist of Glu - Gly pairs (in C1, O1K, and A10) and of Gln - Thr (C1 and A10) and Glu - Thr (O1K), respectively. These cleavage sites look similar to the dipeptide Gln - Gly hydrolyzed by the poliovirus protease. In contrast to the polio enzyme which cleaves exclusively this dipeptide (27) the FMDV protease seems to recognize a broader sequence spectrum as a target. Other dipeptides known to be a substrate for the enzyme are mainly Glu - Gly (three out of the four cleavages liberating the three VPg's (17) and the junction between protease and RNA polymerase in the P100 precursor (18), but also a Glu - Ser (fourth site at the VPg's (17)).

The P88/P52 junction was set by the C-terminus of VP1 (16). Since no amino acid sequence data giving the exact localisation of the gene products contained in the P52 precursor are available, the last 79 amino acid residues encoded by the cDNA clone cannot be attributed to a known protein. In any case this protein is highly conserved in contrast to most of the structural proteins.

Variation in the genes. Comparison of the nucleotide sequences of the three FMDV serotypes shows conserved regions with up to 90 % sequence homology and regions with higher degrees of variation. For clarity the variation between the serotypes has been summarized in Table 2. Most of the base exchanges in P20a and VP4 and the N-terminal part of P52 are silent. Only 4 % of the amino acid residues are exchanged between these proteins. In VP2 and VP3 37 % of the nucleotides are exchanged giving 25 % alterations in the amino acid sequences but the N-terminal and the C-terminal parts of VP2 are highly conserved. Amino acid exchanges are mainly clustered in three regions, one in the center (polyprotein pos. 356 to pos. 382) and two minor regions in the second half of the protein (pos. 408 to pos. 420 and pos. 475 to pos. 481, Fig. 4). A conservative, highly positive charged region of 42 amino acid residues in VP2 (pos. 420 to pos. 461) could represent an interaction site with the viral RNA. Variations in the sequence of VP3 are more dispersed: two minor regions, in the first quarter (pos. 560 to pos. 575) and in the middle of the molecule (pos. 629 to pos. 639) are highly variable. The most variable region of P88 codes for VP1, the dominant immunogenic region of the virus (19, 20), which variability has been analysed in detail (21, 22). In this region almost 40 % of the nucleotides are altered from one serotype to another which effects differences in the primary structure of this protein of up to 30 %.

Immunological data suggest the major antigenic site of FMDV to be localized on VP1, supposed to cover most part of the exterior of the viral capsid. There is no evidence for participation of the two structural proteins VP2 and VP3 in the immunogenicity of the virus. However, it was not possible to mimic convincingly the full immunogenicity of the virus with isolated VP1 (24), or with bacterial VP1-fusion proteins (8, 25), or with chemically synthesized oligopeptides (19, 20). The lower immunogenicity may be explained by the failure of these proteins to adopt the proper secondary structure of the antigenic site. It cannot be excluded, however, that also other sites on other structural proteins are involved in the expression of the viral antigenicity. In this context our finding that in VP2 as well as in VP3 highly variable sites composed mainly by polar amino acid residues exist (Fig. 4) is of particular interest. These sites could possibly be exposed to the outside of the virus and

could represent parts of a more complex antigenic structure of FMDV.

ACKNOWLEDGEMENTS

We like to thank H. Schaller for discussion and comments on the manuscript and O. Steinau for patiently carrying out all the computer analysis. This work was supported by BIOGEN S.A.

REFERENCES

1. Pereira, H.G. (1977) in Developments in Biological Standardization 35, S. Karger, Ed., pp 167 - 174, Basel.
2. King, A.M.Q., Underwood, B.O., McCahon, D., Newman, J.W.I., and Brown, F. (1981) Nature 293, 479 - 480.
3. Shatkin, A.J. (1976) Cell 9, 643 - 653.
4. Jacobson, M.F., Asso, J., and Baltimore, D. (1970) J. Mol. Biol. 49, 657 - 669.
5. Boothroyd, J.C., Harris, T.J.R., Rowlands, D.J., and Lowe, P.A. (1982) Gene 17, 153 - 161.
6. Roberts, R.J., Breitmeier, J.B., Tabacnic, N.F., and Myers, P.A. (1975) J. Mol. Biol. 91, 120 - 123.
7. Johnson, R.A., and Walseth, T.F. (1979) Adv. Cyclic Nucleotide Res. 10, 135 - 137.
8. Küpper, H., Keller, W., Kurz, C., Forss, S., Schaller, H., Franze, R., Marquardt, O., Zaslavsky, W., and Hofschneider, P.H. (1981) Nature 289, 555 - 559.
9. Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymol. 65, 499 - 560.
10. Garoff, H. and Ansorge, W. (1981) Analyt. Biochem. 115, 454 - 457.
11. Osterburg, G., Glatting, K.-H., and Sommer, R. (1981) Nucleic Acids Res. 10, 207 - 216.
12. Kitamura, N., Semler, B.L., Rothberg, P.G., Larsen, G.R., Adler, C.G., Dorner, A,J., Emini, E.A., Hanecak, R., Lee, J.J., Van der Werf, S., Anderson, C.W., and Wimmer, E. (1981) Nature 291, 547 - 553.
13. Sangar, D.V., Bryant, J., Harris, T.J.R., Brown, F., and Rowlands, D. (1981) J. Virol. 39, 67 - 74.
14. Sangar, D.V., Black, D.N., Rowlands, D.J., Harris, T.J.R., and Brown, F. (1980) J. Virol. 33, 59 - 68.
15. Strohmaier, K., Wittmann-Liebold, B., and Geissler, A.W. (1978) Biochem. Biophys. Res. Comm. 85, 1640 - 1645.
16. Kurz. C., Forss, S., Küpper, H., Strohmaier, K., and Schaller, H. (1981) Nucleic Acids Res. 9. 1919 - 1931.
17. Forss, S. and Schaller, H. (1982) Nucleic Acids Res. 10, 6441 - 6450.
18. Robertson, B.H., Morgen, D.O., Moore, D.M., Grubman, M.J., Card, J., Fischer, T., Wendell, G., Dowbenko, D., and Yansura, D. (1983) Virology 126, 614 - 623.
19. Pfaff, E., Mussgay, M., Böhm, H.O., Schulz. G.E., and Schaller, H. (1982) EMBO Journal 1, 869 - 874.
20. Bittle, J.L., Houghten, R.H., Alexander, H., Shinnick, T.M., Sutcliffe, J.G., Lerner, R.A., Rowlands, D.J., and Brown, F. (1982) Nature 298, 30-33
21. Beck, E., Feil, G., and Strohmaier, K. (1983) EMBO Journal 2, 555 - 559.
22. Makoff, A.J., Payntner, C., Rowlands, D.J., and Boothroyd, J.C. (1982) Nucleic Acids Res. 10. 8285 - 8295.
23. Harris, T.J.R., Brown, F., and Sangar, D.W. (1981) Virology 112, 91 - 98.

24. Bachrach, H.L., Morgan, D.O., McKercher, P.D., and Polatnick, J. (1975) J. Immun. 115, 1636 - 1641.
25. Kleid, D.G., Yansura, D., Small, B., Dowbenko, D., Moore, D.M., Grubman, M.J., McKercher, P.D., Morgan, D.O., Robertson, B.H., and Bachrach, H.L. (1981) Science 214, 1125 - 1129.
26. Boothroyd. J.C., Highfield, P.E., Cross, G.A.M., Rowlands, D.J., Lowe, P. A., Brown, F., and Harris, T.J.R. (1981) Nature 290, 800 - 802.
27. Hanecak, R., Semler, B.L., Anderson, C.W., and Wimmer, E. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 3973 - 3977.
28. Forss, S. (1983) Thesis, University of Heidelberg.