JOURNAL OF VIROLOGY, June 1985, p. 651–660
0022-538X/85/060651-10$02.00/0

Vol. 54, No. 3

Nucleotide and Amino Acid Sequence Coding for Polypeptides of Foot-and-Mouth Disease Virus Type A12

BETTY H. ROBERTSON,[1]† MARVIN J. GRUBMAN,[1]* GREG N. WEDDELL,[2] DOUGLAS M. MOORE,[1] J. DOUGLAS WELSH,[1] THOMAS FISCHER,[1] DONALD J. DOWBENKO,[2] DANIEL G. YANSURA,[2] BARBARA SMALL,[2] AND DENNIS G. KLEID[2]

U.S. Department of Agriculture, Plum Island Animal Disease Center, Greenport, New York 11944,[1] and Genentech, Inc., South San Francisco, California 94080[2]

Received 24 September 1984/Accepted 11 February 1985

The coding region for the structural and nonstructural polypeptides of the type A12 foot-and-mouth disease virus genome has been identified by nucleotide sequencing of cloned DNA derived from the viral RNA. In addition, 704 nucleotides in the 5′ untranslated region between the polycytidylic acid tract and the probable initiation codon of the first translated gene, P16-L, have been sequenced. This region has several potential initiation codons, one of which appears to be a low-frequency alternate initiation site. The coding region encompasses 6,912 nucleotides and ends in a single termination codon, UAA, located 96 nucleotides upstream from a 3′-terminal polyadenylic acid tract. Microsequencing of radiolabeled in vivo and in vitro translation products identified the genome position of the major foot-and-mouth disease virus proteins and the cleavage sites recognized by the putative viral protease and an additional protease(s), probably of cellular origin, to generate primary and functional foot-and-mouth disease virus polypeptides.

The genome of foot-and-mouth disease virus (FMDV), an aphthovirus of the picornavirus family, is a single-stranded RNA molecule that acts directly as an mRNA. It contains a covalently bound polypeptide, VPg, at its 5′ end (14, 40) and a 100- to 150-base polycytidylic acid sequence about 300 nucleotides from the 5′ end (4, 19, 37) and terminates in a polyadenylic acid tract approximately 40 residues long (7). The RNA codes for a single polyprotein of about 250,000 daltons, which is subsequently cleaved by cellular and viral proteases to yield viral structural and nonstructural polypeptides (9, 15).

Four primary genome products result from initial protease cleavage of the 250-kilodalton polyprotein. These products are used to define the four functional genome regions (Table 1) and are similar to those delineated by Kitamura et al. (21) for the poliovirus genome and by Palmenberg et al. (32) for the encephalomyocarditis virus genome. For FMDV, these four regions are as follows: an L region, 5′ to the capsid components, that codes for a leader polypeptide (P16-L); a P1 region encoding the precursor for capsid polypeptides (P91); a P2 region coding for the precursor (P56) of polypeptides in the middle genome region; and a P3 region that encodes the precursor (P102), which includes the three VPg molecules (12), a putative viral protease (15, 23), and the viral RNA polymerase (9, 36). The polypeptides described in this manuscript are identified by their molecular weights as determined by polyacrylamide gel electrophoresis followed by map coordinates (38).

The nucleotide sequences of the P1 region and the derived amino acid sequence of the polypeptides have been published previously (5, 22, 25, 27, 47), and the cleavage sites that are involved in the processing of the mature P1 polypeptides have been identified from published amino acid sequences (1, 28, 46). Recently the nucleotide sequence of the entire open reading frame for A10 and the L and P1 regions and a portion of the P2 region of C1 FMDV have been published (3, 6). In the present communication, nucleotide-derived amino acid sequences of the entire open reading frame are presented for FMDV type A12 and compared with radiolabeled amino acid microsequencing results to map the positions of the translation-initiation codon and the nonstructural polypeptides encoded in the L, P2, and P3 regions. The peptide bonds cleaved to generate the major FMDV polypeptides have been identified, and their positions have been mapped on the genome.

MATERIALS AND METHODS

Virus growth, purification, and RNA isolation. A large-plaque ab variant of FMDV type A12 strain 119 was grown and purified as previously described (2, 8, 48). Viral RNA was extracted and purified by the method of Grubman et al. (16) and used as template in reverse transcription.

cDNA synthesis and cloning. cDNA molecules were synthesized by using reverse transcriptase and three types of primers: oligodeoxythymidylic acid, calf thymus DNA, and oligonucleotide primers synthesized to prime at specific regions of the genome. Double-stranded DNA transcripts were inserted into the *Pst*I site of the *Escherichia coli* plasmid pBR322 with oligodeoxycytidylic acid-oligodeoxyguanidylic acid tails and cloned as previously described (13, 22, 49).

Restriction endonuclease analysis. Restriction endonucleases were purchased from Bethesda Research Laboratories (Bethesda, Md.) and New England Biolabs (Waltham, Mass.). The sizes of the DNA cleavage products were determined by polyacrylamide gel electrophoresis and agarose electrophoresis with various pBR322 cleavage fragments as molecular weight markers.

DNA sequencing. Restriction fragments from the various

* Corresponding author.
† Present address: Hepatitis Branch, Centers for Disease Control, Atlanta, GA 30333.

TABLE 1. Primary and secondary cleavage products and sites which generate functional FMDV polypeptides

Genome region	Primary cleavage products	Functional polypeptides[a]	Protease cleavage[b]
L	P16-L		Gly/Gln or Gln-Ser-Gly/Asn (uncertain)[c]
P1	P91-VP4/2/3/1	VP4	Ala/Asp
		VP2	Val/Gly
		VP3	Gln/Thr
		VP1	Lys-Gln-Leu-Leu/Asn (Arg/Pro)
X?[d]			
P2	P56-2A/C	P14-2A	Lys-Gln/Leu
		P41-2C	Lys-Gln/Ile
P3	P102-3A/VPg/C/POL	P19-3A	Glu/Gly
		VPg1-3B	Glu/Gly
		VPg2-3B	Glu/Gly
		VPg3-3B	Glu/Ser-Gly
		P18-3C	Glu/Gly
		P61-3D^{POL}	(C terminus)

[a] Polypeptides of unknown function are designated by molecular weight as determined by polyacrylamide gel electrophoresis followed by the genome region in which they are located (1, 2, and 3) and then followed by the relative location within that genome region (A, B, C, and D) (38).

[b] Shills (/) indicate cleavage sites between primary and secondary cleavage products.

[c] A10 (6) and C1 (3) proposed cleavage points, actual cleavage point uncertain.

[d] X?, 16-amino-acid peptide not present on VP1 or P56-2A/C.

cloned inserts were labeled at the 5′ ends with [^{32}P]ATP (Amersham Corp., Arlington Heights, Ill.) by T4 polynucleotidyl kinase (Boehringer Mannheim Biochemicals, Indianapolis, Ind.). Sequencing was performed by the Maxam and Gilbert procedure (29). Also, the dideoxy sequencing method was employed with single-stranded templates subcloned into M13 phage vectors (30, 31) or with restriction fragments subcloned into the *Cla*I site of pBR322 and primed with a pBR322 *Hin*dIII primer (no. 1205; New England Biolabs).

In vivo and in vitro translation of FMDV polypeptides. In vivo translation products were obtained from infected monolayers of bovine kidney (BK) cells and labeled with radioactive amino acids for 15 or 60 min as previously described (15). In vitro translation with a rabbit reticulocyte cell-free system prepared by the procedure of Schimke (44) was performed as described previously (15, 34).

Amino acid microsequencing. FMDV proteins were labeled in vitro or in vivo with [^{3}H]leucine and [^{35}S]methionine (NewEngland Nuclear Corp., Boston, Mass.) and subjected to polyacrylamide gel electrophoresis on 12.5% gels in a discontinuous Tris-glycine buffer system (26). The separated bands were located in the gels by autoradiography and sliced out, and the proteins were eluted electrophoretically (43). Microsequencing (10, 11) was performed in a Beckman 890B sequencer with 1 mg of apomyoglobin as a carrier. The myoglobin sequence was verified by thin-layer chromatography on a small portion of the sample, and the radioactivity of the remainder was determined by liquid scintillation counting (35).

RESULTS

Restriction map of the coding region of the FMDV genome. Inserts from selected cDNA clones derived from the FMDV genome after priming with calf thymus DNA, oligodeoxythymidylic acid, and synthetic oligonucleotide primers are depicted in Fig. 1. Restriction endonuclease analysis, sizing, and comparison to the plasmid insert T465, which contains the genetic information coding for VP1 (22), were used to position the plasmid inserts within the FMDV genome (Fig. 1).

Nucleotide and derived amino acid sequences. Sequencing from the 5′-labeled sites (dots on arrows in Fig. 1) of restriction fragments of cloned cDNA inserts (middle row of Fig. 1) yielded sequences for the coding region of A12 119 FMDV (Fig. 2). All sequences were confirmed by sequencing from more than one labeling or priming site.

The nucleotide sequence of the A12 FMDV contains a single open reading frame of 6,996 nucleotides (Fig. 2). The derived amino acid sequence contains recognizable sequences of structural proteins VP4, VP2, VP3, VP1, and VPg and the viral RNA polymerase (12, 22, 36). The other two reading frames contain multiple stop codons with no potential for translation into proteins larger than 10,000 daltons. In addition, 620 nucleotides were identified 5′ of the AUG codon which starts the open reading frame. There are six AUG codons in this 620-nucleotide region; however, before the AUG at nucleotides 621 through 623, termination codons are present in all three reading frames. The nucleotide sequence has not been determined in the region 5′ to the polycytidylic acid sequence or in the region between the

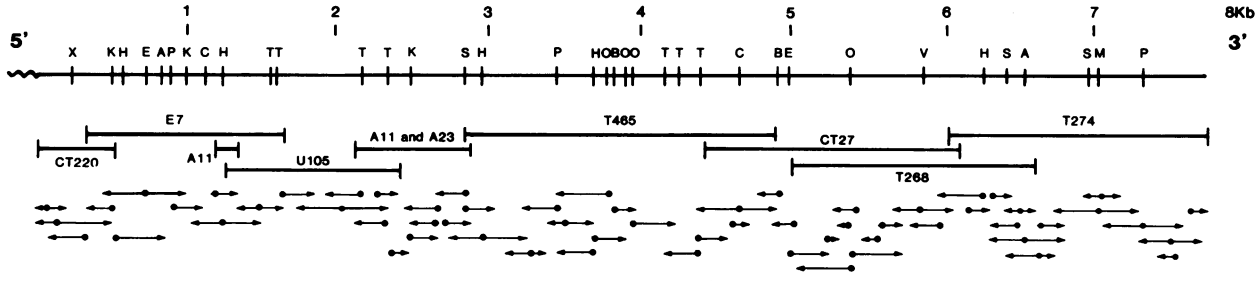


FIG. 1. Schematic representation of FMDV genome including 704 bases of the 5′ untranslated region, the coding region, and 96 bases 3′ of the translation region. Selected restriction endonuclease sites are designated by the horizontal bars. Cloned inserts from plasmids used to determine the nucleotide sequence are shown, as are the labeling or priming sites (●) and distance sequenced (arrow lines). Inserts from plasmids designated by CT and T are derived from calf thymus and oligodeoxythymidylic acid priming, respectively, whereas A, U and E prefixes for plasmid inserts indicate that synthetic oligonucleotide primers were used. Restriction endonucleases: T, *Taq*I; K, *Kpn*I; V, *Eco*RV; H, *Hin*dIII; E, *Eco*RI; P, *Pvu*II; C, *Cla*I; S, *Pst*I; A, *Sal*I; B, *Bgl*II; X, *Xba*I; O, *Xho*I; M, *Bam*HI.

polycytidylic acid tract and the initial sequences shown in Fig. 2.

Identification of the P16-L polypeptide. In a cell-free system, at 1.5 mM magnesium, P16 is the first polypeptide synthesized and the major polypeptide labeled with [^{35}S]formyl-methionyl tRNA (39) (unpublished observation). It is also found in infected cells (17). Microsequencing of this leader polypeptide (Table 1), isolated in vivo or in vitro, failed to release radioactive methionine or leucine within the first 30 residues (Fig. 3A). Since acetylation is a common amino-blocking reaction for viral proteins, in vitro translation was performed in the presence of oxaloacetic acid and citrate synthetase to inhibit acetylation (33). Analysis of P16 synthesized under the above conditions (Fig. 3B) indicated the release of methionine in sequencing cycle 1 and leucine in sequencing cycle 5. However, the derived amino acid sequence in the single open reading frame indicates that sequencing from the first methionine (Fig. 2, nucleotides 1 through 3) would yield leucine at positions 10 and 22 and methionine at position 29. The second methionine, 84 bases downstream, is followed by a leucine in position 5. Therefore, it would appear likely that translation begins at the second AUG codon in the open reading frame. Additional leucines are encoded at positions 25 and 29, but because of decreased repetitive yield these residues were obscured by a rising background of radioactivity. The methionine and leucine residues positioned the probable start of P16-L at the AUG codon more than 700 nucleotides 3′ of the polycytidylic acid tract and 6,912 nucleotides from the UAA termination codon. It should be mentioned that the above information does not allow us to unequivocally state that this AUG codon is the start site of P16-L. Also, if initiation in vitro occurs at an abnormal site that has a sequence of methionine at the first position and leucine at the fifth position, i.e., nucleotide 700, we would not be able to distinguish this from correct initiation. Nevertheless, we conclude that our suggestion is probably correct, because of the significant increase in the release of methionine after inhibition of acetylation, the release of leucine in cycle 5, and the similarity of P16-L polypeptides from in vitro and in vivo demonstrated by tryptic peptide analysis (39). Furthermore, Beck et al. (3) have also suggested that translation of P16-L begins at the comparable AUG in FMDV serotype C1, whereas the preceding AUG is recognized as the start of P20a, the major viral protein translated in vitro at high magnesium concentrations (unpublished observations). Our attempts to sequence P20a isolated from high-magnesium in vitro acetylation-inhibited translations were inconclusive.

P1 genome region: structural polypeptides. The amino and carboxy termini of the larger viral capsid polypeptides and their precursor cleavage sites have been previously identified (1, 28, 46) (Fig. 2). Cleavage between P16-L and the smallest capsid polypeptide, VP4, was placed provisionally at positions 609 through 610 between glycine and glutamine for type A10 (6) and at positions 648 through 649 between glycine and asparagine for type C1 (3). Attempts at microsequencing P91 and VP0 (Table 1) and amino-terminal dansylation or Edman degradation of purified VP4 failed, indicating that the amino terminus of VP4 is blocked.

Polypeptides generated from the P2 genome region. Microsequencing of the primary cleavage product, P56-2AC (Table 1), of the P2 genome region (17) released leucine from position 14 and methionines from positions 21 and 25 (Fig. 2 and 4A). This placed P56-2AC 16 amino acids away from the lysine-glutamine-leucine-leucine C terminus of VP1 and identified an apparent potential cleavage sequence, arginine-proline. To verify the cleavage point, P14-2A (Table 1) was sequenced, releasing leucine from position 14 and methionines from positions 21 and 25 (data not shown), demonstrating that this polypeptide is at the 5′ end of the P2 region and confirming the arginine-proline cleavage sequence. Evidence indicates that the other polypeptide coded by the P2 region, P41-2C (17), is altered in guanidine-resistant mutants of FMDV type O (41). Microsequencing of P41-2C released leucines at positions 1, 13, 19, 22, and 24 (Fig. 2 and 4B). This positioned P41-2C at the 3′ end of P2. Thus the P2 region codes for P14-2A (154 amino acids) and P41-2C (318 amino acids), both of which have recently been reported to be associated with membranous fractions of cells (17).

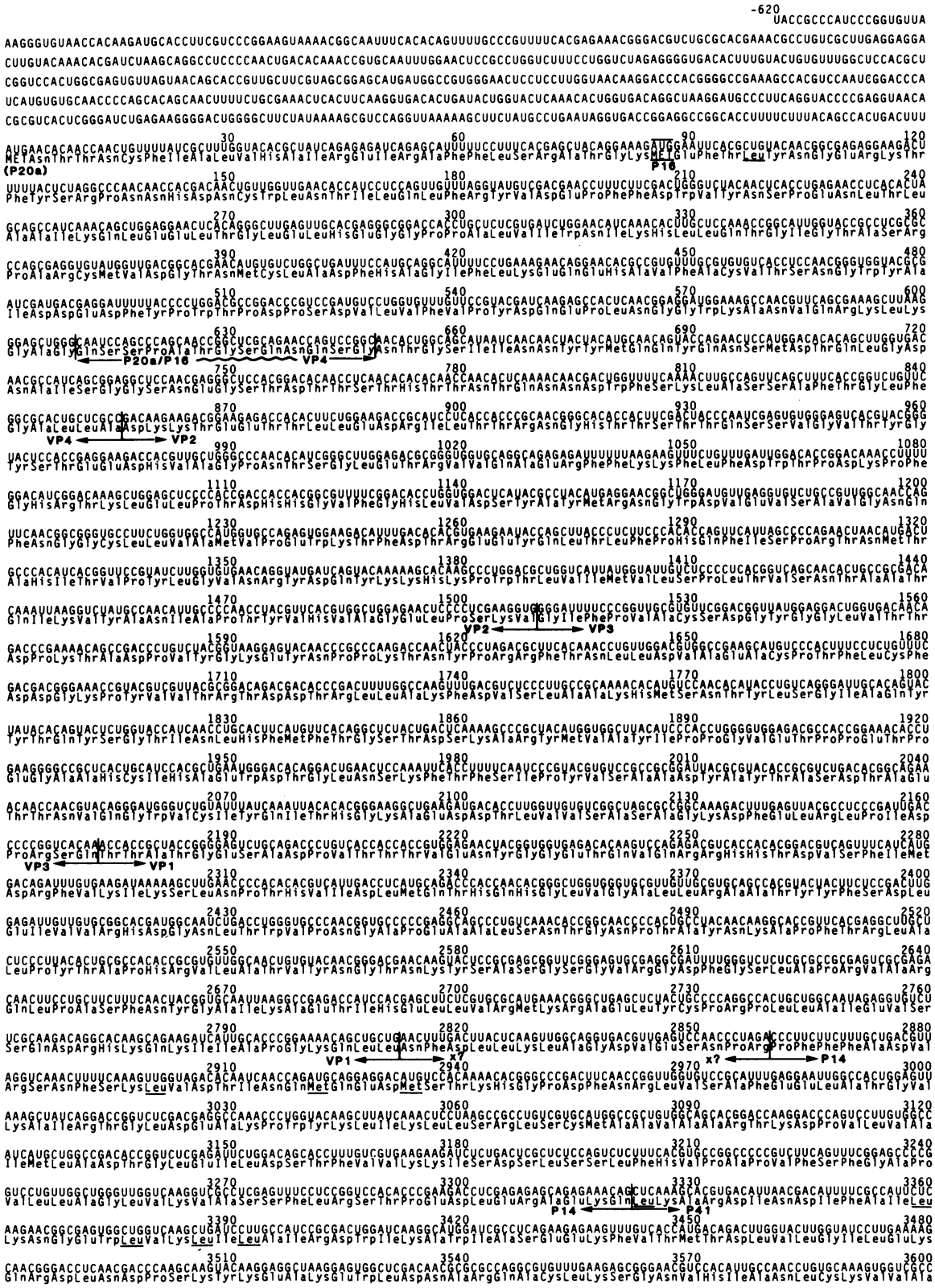
Polypeptides derived from the P3 genome region. The P3 region encodes the previously identified RNA polymerase (9, 36), a putative viral protease (15, 23), the three tandem 23- to 24-amino-acid VPg proteins (12), and an additional polypeptide of unknown function (17). The genome locations of the three VPg proteins (12) and the RNA polymerase (36) have been reported. Microsequencing of the remaining major polypeptides, P19-3A and P18-3C, from this region revealed P19-3A to be the most 5′ of these two proteins. Leucine residues were released at positions 10 and 13, and a methionine was released at position 29 (Fig. 2 and 5A). Microsequencing of P18-3C released methionine from positions 11 and 13 and leucine from positions 8, 21, and 23 (Fig. 2 and 5B). By analogy to the location of the proteases coded for by FMDV type O1k, poliovirus, and encephalomyocarditis virus, we suggest that P18-3C is the putative FMDV protease. Thus, the order of the proteins in the P3 genome region appears to be P19-3A/VPg1-VPg2-VPg3-3B/P18-3C/P61-3D^{POL}.

DISCUSSION

The results in this manuscript present the nucleotide sequence and the complete nucleotide-derived amino acid sequence for the coding region of the FMDV type A12 genome. Similar results for FMDV types C1 and A10 have recently been published (3, 6). Furthermore, based on microsequencing of polypeptides isolated from infected cells and from a cell-free translation system, we demonstrate the map position of the major functional viral polypeptides and the protease cleavage sites recognized by viral and cellular proteases on the polyprotein.

Upstream of the single open reading frame, we have identified 620 nucleotides that are present at the 3′ end of the polycytidylic acid tract. Within the next 87 nucleotides there are two potential AUG codons, and our data indicate that initiation of translation of P16-L probably takes place at the second AUG codon (nucleotides 85 through 87) and that the initiating methionine is acetylated. This codon has features in common with the model proposed by Kozak (24) for efficient initiation of translation (AXX AUG G). The first AUG in the open reading frame has been postulated to be utilized for synthesis of an alternative leader component (P20a-L), which is related to P16-L by immunoprecipitation (3). Initiation at the site utilized for P20a is not normally very efficient (39) (unpublished observations), although if the magnesium concentration is raised, P20a-L is the predominant polypeptide initiated and translated in a cell-free system. Since the microsequencing of P20a-L isolated from acetylation-inhibited lysates translated under high-magnesium concentrations was inconclusive, we cannot definitively place the location of the P20a-L component.

The capsid region, as in other picornaviruses, codes for four structural polypeptides. The amino and carboxy termini



of VP1, VP2, and VP3 have been previously published (1, 28, 46). The amino terminus of VP4 could not be determined and is apparently blocked by other than an acetyl group, because neither leucine nor methionine was released from its precursor (P91-VP4/2/3/1) produced during in vitro translation when acetylation was inhibited. Consequently, the cleavage site between P16-L and VP4 was not located. Beck et al. (3) have postulated that since the cleavage between L

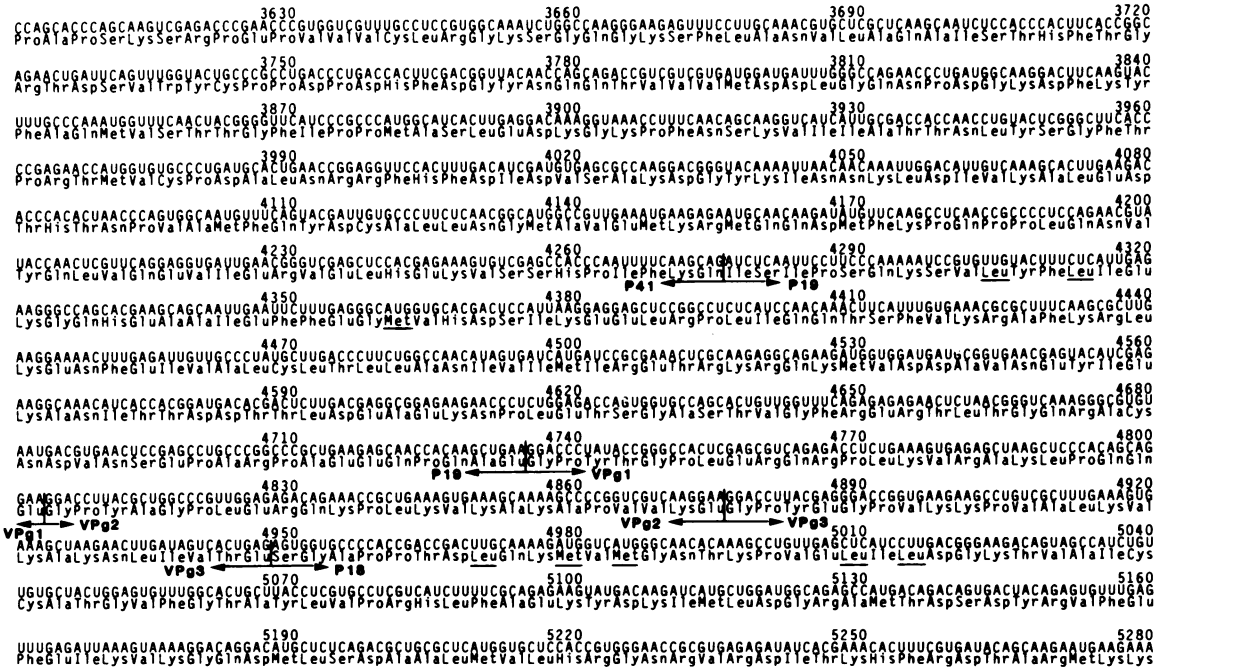


FIG. 2. Nucleotide and derived amino acid sequence for the coding region of FMDV serotype A12 large plaque, ab variant. The principal initiating methionine is depicted by lines above and below, and numbering is started at the first methionine of the 6,996-nucleotide open reading frame. Cleavage sites are indicated by the double-headed arrows. Amino acids identified by radiolabeled microsequencing are underlined. The wavy line between P16-L and VP4 indicates the uncertain cleavage point. The 16-amino-acid section between VP1 and P14-2AC which was not identified by sequencing of any precursor or product is designated by X?. Reanalysis of the nucleotide sequence has revealed an additional CUG leucine codon at the carboxy terminus of VP1 as compared with the sequence reported previously (22).

and P1 and between P1 and P2 is presumably by a host protease, some homology at these sites might be expected. A consensus sequence deduced from the P1-P2 junction of eight FMDV strains was used to suggest that the L-P1 cleavage occurred between glycine and asparagine at amino acid positions 217 and 218 of type C1 (3). Our present information demonstrates that the A12 P1-P2 cleavage occurs between leucine and asparagine and the same potential L-P1 glycine-asparagine cleavage is present at positions 216 and 217 for type A12. However, based on additional unpublished information that type A10 VP4 contains proline (6) and, by analogy to the L-P1 cleavage site in EMC virus,

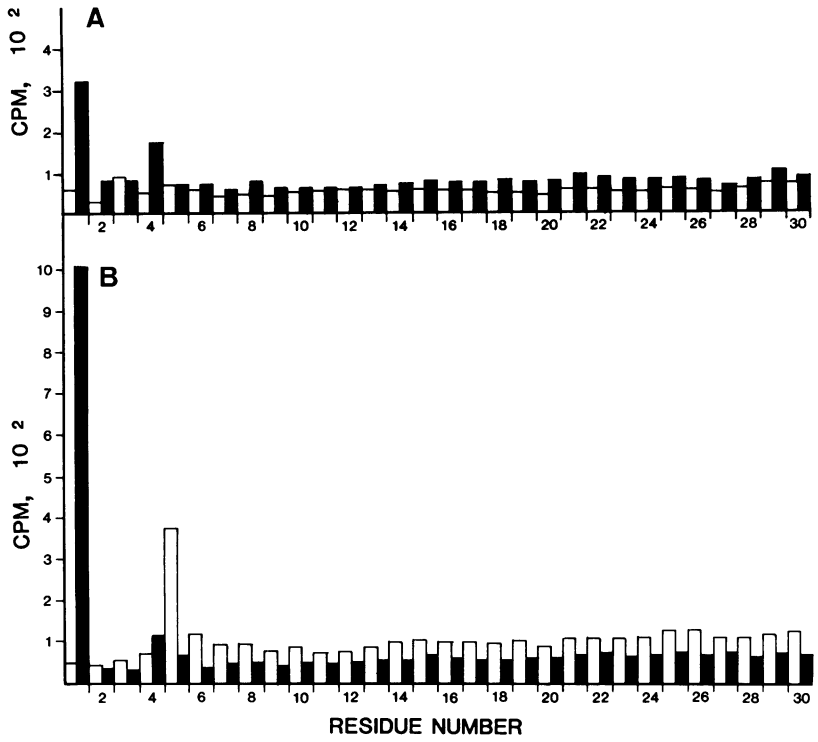


FIG. 3. Radiolabeled microsequencing of P16-L in the absence (A) and presence (B) of inhibition of amino-terminal acetylation. The sample sequenced in A was isolated from FMDV-infected cell lysates by using 3×10^6 BK cells infected with FMDV (multiplicity of infection, 100). The infected cells were labeled for 15 min at 4 h postinfection with 156 μCi of [^{3}H]leucine and 390 μCi of [^{35}S]methionine. Normal in vitro translation of FMDV RNA was performed in reticulocyte lysates (36, 37) as described previously (8); acetate salts were excluded, and the potassium concentration was adjusted to 101 mM when it was desired to inhibit amino-terminal acetylation (B). In addition, acetylation-inhibited lysates contained 0.15 μM oxaloacetic acid and 1 U of citrate synthetase (30). Translation products in the lysates were labeled with [^{3}H]leucine (50 to 500 μCi per translation) and [^{35}S]methionine (125 to 160 μCi per translation). Polypeptides were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, processed, and sequenced as described in the text. Symbols: □, [^{3}H]leucine; ■, [^{35}S]methionine.

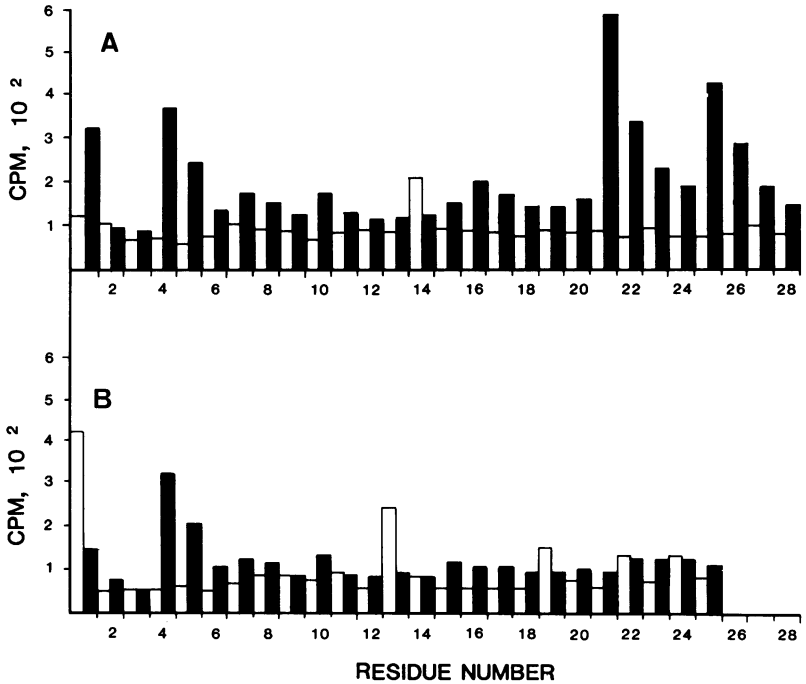


FIG. 4. Radiolabeled microsequencing of P56-2AC (A) and P41-2C (B). Polypeptides were isolated from FMDV-infected cell lysates as described for Fig. 3A. After gel separation, elution, and microsequencing, the released [^{3}H]leucine (□) and [^{35}S]methionine (■) residues were determined for each cycle. Polypeptides isolated from in vitro translations gave similar results. In the first and sometimes fourth Edman degradation cycles, there appeared to be nonspecific release of [^{35}S]methionine.

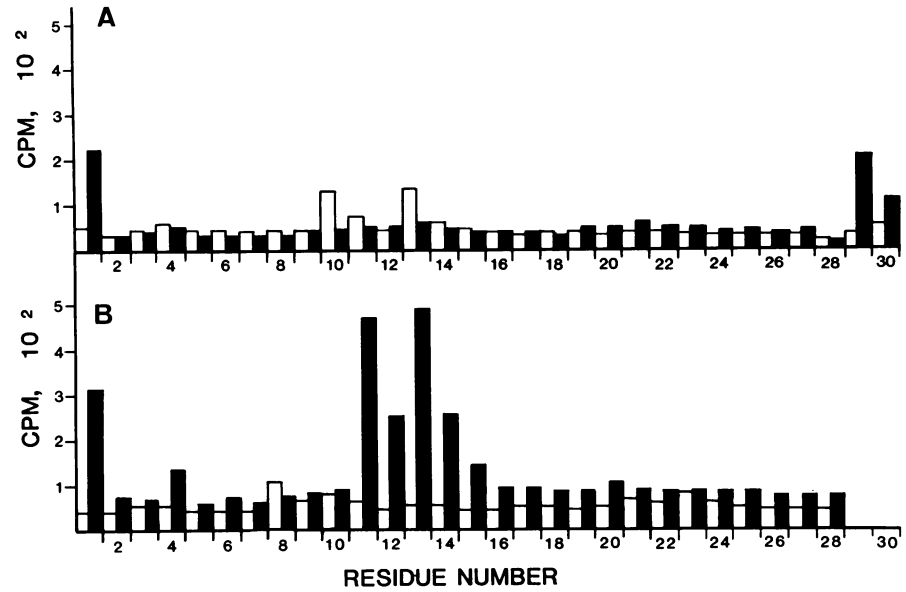


FIG. 5. Radiolabeled microsequencing of P19-3A (A) and P18-3C (B). Polypeptides were isolated from FMDV-infected cell lysates as described for Fig. 3A. After gel separation, elution, and microsequencing, the released [^{3}H]leucine (□) and [^{35}S]methionine (■) were determined for each cycle. Polypeptides isolated from in vitro translations gave similar results.

Carroll et al. (6) proposed cleavage of L-P1 to occur at a glycine-glutamine pair (amino acids 204 and 205) in type A10. The amino acid sequence through this region (residues ~200 through 230) is identical for A12, A10, and C1. In the absence of definitive amino acid sequencing data, we feel that the L-P1 cleavage cannot be defined at this time, and the data in Fig. 2 and Tables 1, 2, and 3 reflect this uncertainty.

Within the P2 genome region of FMDV, only two products (P14-2A and P41-2C) are generated as compared with three in poliovirus and encephalomyocarditis virus (38). However, the nucleotide sequence defines a small peptide of 16 amino acids which is not present at the C terminus of VP1 (1, 33) or on the amino terminus of either the P2 genome precursor (P56-2AC) or its most 5′ product (P14-2A) (this 16-amino-acid peptide is designated in Fig. 2 by the "X?"). Since translation is contiguous, either rapid proteolytic processing removes these 16 amino acids, or an additional cleavage between an arginine-proline sequence (Table 1) has occurred. Work is currently underway with antiserum against a synthetic peptide corresponding to a portion of "X" to identify the product(s) that may contain this peptide. The identical amino acid sequence is present in A10 (6) and type C1 (3), with the exception of the residues occurring at the arginine-proline cleavage site (A12, proline-arginine-proline; C1, proline-glycine-proline; and A10, leucine-glycine-proline). Considering this difference, it will be interesting when amino acid sequencing is performed to see whether such a peptide is not found with types A10 and C1.

The P3 genome region of FMDV is known to code for the previously identified viral RNA polymerase (9, 36), three tandem VPg molecules (12), and a viral protease (15, 23). The relative position of the RNA polymerase and the three tandem VPg molecules has been confirmed, whereas the remaining two components, P19-3A and P18-3C, are the first and third polypeptides translated in this genome region. The poliovirus protease maps in the third position in the P3 genome region (18). By analogy to these systems, P18-3C may be the FMDV protease. Recently, using an *E. coli* expression vector containing FMDV cDNA of different sizes, Klump et al. (23) mapped the putative viral protease activity to a position 5′ adjacent to the viral RNA polymerase gene segment. The function of the remaining viral protein from this region, P19-3A, remains to be elucidated.

These experiments reveal that, in contrast to poliovirus and encephalomyocarditis virus proteolytic cleavages, little homology is apparent in cleavage of the FMDV polyprotein. There does appear to be recognition of glutamic acid or glutamine residues followed by glycine, serine, or threonine (Table 1). This specificity is similar to the glutamine-glycine sequence recognized by the poliovirus and encephalomyocarditis virus proteases (18, 34). In addition, a second general cleavage sequence is recognized: lysine-glutamine-leucine-leucine-/-asparagine, lysine-glutamine-/-leucine, and

TABLE 2. Charge properties and molecular weights of type A12 FMDV polypeptides calculated from the derived amino acid sequence

Polypeptide	No. of residues		Mol wt	pKI[a]
	Acidic	Basic		
P16-L[b]	24	18	21,176[c]	4.8
VP4[b]	5	2	7,592[c]	4.0
VP2	23	28	24,677	6.7
VP3	26	22	24,267	4.9
VP1	17	32	23,396	9.7
P14-2A	18	20	16,894	7.3
P41-2C	38	46	35,950	8.0
P19-3A	26	22	17,368	5.1
VPg-1-2-3B	6	16	2,644	10.5
			2,586	10.3
			2,561	10.0
P18-3C	24	33	22,993	8.7
P61-3D^{POL}	63	67	52,694	6.0

[a] Values for amino acids are based on those obtained by Shire et al. (45).
[b] Arbitrary selection of glutamine-serine at nucleotide positions 642 and 643 for calculations.
[c] Provisional due to uncertainty of the L-P1 cleavage site.

TABLE 3. Comparison of A12 FMDV polypeptide sequence homology with FMDV types A10 and C1

Polypeptide	No. of residues		Differences[a]	% Difference
P20a	A12	214[b]		
	A10	215	23 + 1 insertion	11
	C1	214	21	10
P16	A12	186[b]		
	A10	187	18 + 1 insertion	10
	C1	186	14	8
VP4	A12	71		
	A10	71	2	3
	C1	71	1	1
VP2	A12	218		
	A10	218	8	4
	C1	218	47 + 1 insertion, −1 deletion	23
VP3	A12	221		
	A10	221	11	5
	C1	219	39 − 2 deletions	19
VP1	A12	213		
	A10	212	22 − 1 deletion	11
	C1	209	62 − 4 deletions	31
X?	A12	16		
	A10	16	2	13
	C1	16	1	6
P14	A12	154		
	A10	154	4	3
P41	A12	318		
	A10	318	5	2
P19	A12	153		
	A10	153	11	7
VPg1	A12	23		
	A10	23	3	13
	C1	23	1	4
VPg2	A12	24		
	A10	24	2	8
	C1	24	2	8
VPg3	A12	24		
	A10	24	1	4
	C1	24	0	0
P18	A12	213		
	A10	214	7	3
P61	A12	470		
	A10	470	16	3

[a] All substitutions and each addition or deletion of a residue for types A10 and C1 (3, 6, 12) are counted as a difference as compared to A12.

[b] Cleavage at glutamine-serine was arbitrarily selected at nucleotide positions 642 and 643 for calculations.

lysine-glutamine-/-isoleucine. These three related sequences occur on the borders of the P2 genome region and between the two functional proteins (P14-2A and P41-2C) within this genome region. However, present information suggests that these two general cleavage sequences can be recognized by either a virus- or a host-derived protease. Although there are similarities between the specificity of the poliovirus and FMDV proteases, comparison of their amino acid sequences reveals only three limited areas of homology (unpublished observations).

The calculated molecular weights and pKI values of the identified structural and nonstructural proteins of type A12 FMDV are shown in Table 2. Proteins VP1 and the VPg molecules are highly basic, whereas proteins P14-2A and P41-2C both have long tracts of hydrophobic or uncharged residues as determined by the algorithm of Hopp and Woods (20), a characteristic that correlates well with their association with membranes (17). No strongly hydrophobic area was identified 5′ adjacent to the VPgs for A12 or A10 (6), as is found with poliovirus (42). Both show a strongly hydrophobic 16-amino-acid span 77 amino acids 5′ of the first VPg about midway in the P19-3A sequence. However, the potential for this region allowing membrane association of a VPg precursor seems unlikely due to preliminary evidence that VPg is processed from a P81 precursor (VPg-C-POL) (17) rather than P19 as has been shown for poliovirus (42).

A comparison of the complete coding region of FMDV A10 and A12 and the L-P1-P2 regions for A12 and C1 reveals a high degree of amino acid homology among most of the viral polypeptides (Table 3). As expected, the immunogenic VP1 polypeptide contains regions of high variability between the two subtypes A12 and A10 (about 11% different) and a significantly higher variation between serotypes A12 and C1 (31%). Although the VP4 amino acid sequence is highly conserved in all three types, as are VP2 and VP3 between A12 and A10, VP2 and VP3 show significantly more difference between the two serotypes A12 and C1 (23 and 19%, respectively, for VP2 and VP3). These differences are scattered more broadly in VP2 and especially in VP3 than for VP1 and might suggest a more indirect effect on antigenic differences shown by the two serotype groups.

Surprisingly, the nonstructural polypeptide(s) P16-L (P20a-L) shows significant differences in amino acid sequence (Table 3), even between the two more closely related A subtypes. It is difficult to speculate how the observed degree of variation would effect a presumed single function of the nonstructural leader segments. Another significant difference initially apparent in a nonstructural polypeptide is 13 consecutive amino acid changes in the viral protease (Fig. 2, nucleotides 5,179 through 5,217) between type A10 (6) and A12. However, this has been resolved through a correction in the published A10 sequence (A. R. Carroll, D. J. Rowlands, and B. E. Clarke, Nucleic Acids Res. **12**:4430, 1984). In the corrected A10 sequence, this segment is identical to the A12 amino acid sequence shown here. The amino acid sequence is highly conserved throughout the protease, consistent with the close similarity of cleavage patterns for the two strains.

The information presented in this communication can now be used to elucidate common structural features between FMDV and other picornaviruses and to characterize the effect of structure determinants on the function of individual proteins during FMDV replication and pathogenesis.

ACKNOWLEDGMENTS

We acknowledge computer assistance from the National Institutes of Health-supported Molgen project at Sumex, Stanford University Medical Center, the Cold Spring Harbor Laboratory Computer Facility, and H. L. Robertson, Southold, N.Y. We thank H. L. Bachrach for editing; Barbara Moorhouse for typing; and K. Axelson, J. Dean, and J. Mugavero for technical assistance.

LITERATURE CITED

1. **Bachrach, H. L., J. B. Swaney, and G. F. VandeWoude.** 1973. Isolation of the structural polypeptides of foot-and-mouth dis-

ease virus and analysis of their C-terminal sequences. Virology **52**:520–528.
2. **Bachrach, H. L., R. Trautman, and S. S. Breese, Jr.** 1964. Chemical and physical properties of virtually pure foot-and-mouth disease virus. Am. J. Vet. Res. **25**:333–342.
3. **Beck, E., S. Forss, K. Strebel, R. Cattaneo, and G. Feil.** 1983. Structure of the FMDV translation initiation site and of the structural proteins. Nucleic Acids Res. **11**:7873–7885.
4. **Black, D. N., P. Stephenson, D. J. Rowlands, and F. Brown.** 1979. Sequence and the location of the poly C tract in aphtho- and cardiovirus RNA. Nucleic Acids Res. **6**:2381–2390.
5. **Boothroyd, J. C., T. J. R. Harris, D. J. Rowlands, and P. A. Lowe.** 1982. The nucleotide sequence of cDNA coding for the structural proteins of foot-and-mouth disease virus. Gene **17**:153–161.
6. **Carroll, A. R., D. J. Rowlands, and B. E. Clark.** 1984. The complete nucleotide sequence of the RNA coding for the primary translation product of foot-and-mouth disease virus. Nucleic Acids Res. **12**:2461–2472.
7. **Chatterjee, N. K., H. L. Bachrach, and J. Polatnick.** 1976. Foot-and-mouth disease virus RNA. Presence of 3′ terminal polyriboadenylic acid and absence of amino acid binding ability. Virology **69**:369–377.
8. **Cowan, K. M.** 1969. Immunochemical studies of foot-and-mouth disease. V. Antigenic variants of virus demonstrated by immunodiffusion analysis with 19S but not 7S antibodies. J. Exp. Med. **129**:333–350.
9. **Doel, T. R., D. V. Sangar, D. J. Rowlands, and F. Brown.** 1978. A re-appraisal of the biochemical map of foot-and-mouth disease RNA. J. Gen. Virol. **41**:395–404.
10. **Edman P.** 1950. Method for determination of the amino acid sequence in peptides. Acta Chem. Scand. **4**:283–293.
11. **Edman, P., and G. Begg.** 1976. A protein sequenator. Eur. J. Biochem. **1**:80–90.
12. **Forss, S., and H. Schaller.** 1982. A tandem repeat gene in a picornavirus. Nucleic Acids Res. **10**:6441–6450.
13. **Goeddel, D. V., H. L. Heyneker, T. Hozumi, R. Arentzen, K. Itakura, D. G. Yansura, M. J. Ross, G. Miozzari, R. Crea, and P. H. Seeburg.** 1979. Direct expression in Escherichia coli of a DNA sequence coding for human growth hormone. Nature (London) **281**:544–548.
14. **Grubman, M. J.** 1980. The 5′ end of foot-and-mouth disease virion RNA contains a protein covalently linked to the nucleotide pUp. Arch. Virol. **63**:311–315.
15. **Grubman, M. J., and B. Baxt.** 1982. Translation of foot-and-mouth disease virion RNA and processing of the primary cleavage products in a rabbit reticulocyte lysate. Virology **116**:19–30.
16. **Grubman, M. J., B. Baxt, and H. L. Bachrach.** 1979. Foot-and-mouth disease virion RNA: studies on the relation between the length of its 3′poly(A) segment and infectivity. Virology **97**:22–31.
17. **Grubman, M. J., B. H. Robertson, D. O. Morgan, D. M. Moore, and D. Dowbenko.** 1984. A biochemical map of the polypeptides specified by foot-and-mouth disease virus. J. Virol. **50**:579–586.
18. **Hanecak, R., B. L. Semler, C. W. Anderson, and E. Wimmer.** 1982. Proteolytic processing of poliovirus polypeptides: antibodies to polypeptide P3-7C inhibit cleavage at Gln-Gly amino acid pairs. Proc. Natl. Acad. Sci. U.S.A. **79**:3973–3977.
19. **Harris, T. J. R., K. J. H. Robson, and F. Brown.** 1980. A study of the level of nucleotide sequence conservation between the RNAs of two serotypes of foot-and-mouth disease virus. J. Gen. Virol. **50**:403–418.
20. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. U.S.A. **78**:3824–3828.
21. **Kitamura, N., B. L. Semler, P. G. Rothberg, G. R. Larsen, C. J. Adler, A. J. Dorner, E. A. Emini, R. Hanecak, J. J. Lee, S. vanderWerf, C. W. Anderson, and E. Wimmer.** 1981. Primary structure, gene organization and polypeptide expression of poliovirus RNA. Nature (London) **291**:547–553.
22. **Kleid, D. G., D. Yansura, B. Small, D. Dowbenko, D. M. Moore, M. J. Grubman, P. D. McKercher, D. O. Morgan, B. H. Robertson, and H. L. Bachrach.** 1981. Cloned viral protein vaccine for foot-and-mouth disease: responses in cattle and swine. Science **214**:1125–1129.
23. **Klump, W., O. Marquardt, and P. H. Hofschneider.** 1984. Biologically active protease of foot-and-mouth disease virus is expressed from cloned viral cDNA in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **81**:3351–3355.
24. **Kozak, M.** 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res. **9**:5233–5252.
25. **Kurz, C., S. Forss, H. Kupper, K. Strohmaier, and H. Schaller.** 1981. Nucleotide sequence and corresponding amino acid sequence of the gene for the major antigen of foot-and-mouth disease virus. Nucleic Acids Res. **9**:1919–1931.
26. **Laemmli, U. K.** Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. Nature (London) **227**:680–685.
27. **Makoff, A. J., C. A. Paynter, D. J. Rowlands, and J. C. Boothroyd.** 1982. Comparison of the amino acid sequence of the major immunogen from three serotypes of foot-and-mouth disease virus. Nucleic Acids Res. **10**:8285–8295.
28. **Matheka, H. D., and H. L. Bachrach.** 1975. N-terminal amino acid sequences in the major capsid proteins of foot-and-mouth disease virus types A, O, and C. J. Virol. **16**:1248–1253.
29. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
30. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
31. **Mills, D. R., and F. R. Kramer.** 1979. Structure-independent nucleotide sequence analysis. Proc. Natl. Acad. Sci. U.S.A. **76**:2232–2235.
32. **Palmenberg, A. C., E. M. Kirby, M. R. Janda, N. L. Drake, G. M. Duke, K. F. Potratz, and M. S. Collett.** 1984. The nucleotide and deduced amino acid sequence of the encephalomyocarditis viral polyprotein coding region. Nucleic Acids Res. **12**:2969–2985.
33. **Palmiter, R. D.** 1977. Prevention of NH_2-terminal acetylation of proteins synthesized in cell-free systems. J. Biol. Chem. **252**: 8781–8783.
34. **Pelham, H. R. B., and R. J. Jackson.** 1976. An efficient mRNA-dependent translation system from reticulocyte lysates. Eur. J. Biochem. **67**:147–256.
35. **Robertson, B. H., D. M. Moore, M. J. Grubman, and D. G. Kleid.** 1983. Identification of an exposed region of the immunogenic capsid polypeptide VP1 on foot-and-mouth disease virus. J. Virol. **46**:311–316.
36. **Robertson, B. H., D. O. Morgan, D. M. Moore, M. J. Grubman, J. Card, T. Fischer, G. Weddell, D. Dowbenko, and D. Yansura.** 1983. Identification of amino acid and nucleotide sequence of the foot-and-mouth disease virus RNA polymerase. Virology **126**:614–623.
37. **Rowlands, D. J., T. J. R. Harris, and F. Brown.** 1978. More precise location of the polycytidylic acid tract in foot-and-mouth disease virus RNA. J. Virol. **26**:335–343.
38. **Rueckert, R. R., and E. Wimmer.** 1984. Systematic nomenclature of picornavirus proteins. J. Virol. **50**:957–959.
39. **Sangar, D. V., D. N. Black, D. J. Rowlands, T. J. R. Harris, and F. Brown.** 1980. Location of the initiation site for protein synthesis on foot-and-mouth disease virus RNA by in vitro translation of defined fragments of the RNA. J. Virol. **33**:59–68.
40. **Sangar, D. V., D. J. Rowlands, T. J. R. Harris, and F. Brown.** 1977. Protein covalently linked to foot-and-mouth disease virus RNA. Nature (London) **268**:648–650.
41. **Saunders, K., and A. M. Q. King.** 1982. Guanidine-resistant mutants of aphthovirus induce the synthesis of an altered nonstructural polypeptide P34. J. Virol. **42**:389–394.
42. **Semler, B. L., C. W. Anderson, R. Hanecak, L. F. Dorner, and E. Wimmer.** 1982. A membrane-associated precursor to poliovirus VPg identified by immunoprecipitation with antibodies directed against a synthetic heptapeptide. Cell **28**:405–412.
43. **Semler, B. L., C. W. Anderson, N. Kitamura, P. G. Rothberg, W. L. Wishart, and E. Wimmer.** 1981. Polio replication pro-

teins: RNA sequence encoding P3-1b and sites of proteolytic processing. Proc. Natl. Acad. Sci. U.S.A. **18**:3464–3468.
44. **Schimke, R. T., R. E. Rhoads, and G. S. McKnight.** 1974. Assay of ovalbumin mRNA in reticulocyte lysates, Methods Enzymol. **30**:694–701.
45. **Shire, S. J., G. H. I. Hanania, and F. R. N. Gurd.** 1974. Electrostatic effects in myoglobin. Hydrogen ion equilibria in sperm whale ferrimyoglobin. Biochemistry **14**:1352–1358.
46. **Strohmaier, K., B. Wittmann-Liebold, and A.-W. Geissler.** 1978. The N-terminal sequence of three coat proteins of foot-and-mouth disease virus. Biochem. Biophys. Res. Commun. **85**:1640–1645.
47. **Villanueve, N., M. Davila, J. Ortin, and E. Domingo.** 1983. Molecular cloning of cDNA from foot-and-mouth disease virus C1 Santa Pau (C-S8). Sequence of protein VP1-coding segment. Gene **23**:185–194.
48. **Wagner, G. G., J. L. Card, and K. M. Cowan.** 1970. Immunochemical studies of foot-and-mouth disease. VII. Characterization of foot-and-mouth disease virus concentrated by polyethylene glycol precipitation. Arch. Ges. Virusforsch. **30**:343–352.
49. **Wickens, M. P., G. N. Buell, and R. T. Schimke.** 1978. Synthesis of double-stranded DNA complementary to lysozyme, ovomucoid, and ovalbumin mRNAs. J. Biol. Chem. **253**: 2483–2495.