Volume 12 Number 6 1984 Nucleic Acids Research

# The nucleotide and deduced amino acid sequences of the encephalomyocarditis viral polyprotein coding region

Ann C.Palmenberg*, Ellen M.Kirby, Michael R.Janda, Neil L.Drake, Gregory M.Duke, Kimberly F.Potratz+ and Marc S.Collett+

Biophysics Laboratory of the Graduate School, University of Wisconsin, Madison, WI 53706, and +Molecular Genetics, Inc., 10320 Bren Road East, Minnetonka, MN 55343, USA

Received 19 December 1983; Revised and Accepted 27 February 1984

ABSTRACT

The nucleotide sequence of 7200 bases of encephalomyocarditis (EMC) viral RNA, including the complete polyprotein-coding region, was determined. The polyprotein is encoded within a unique translational reading frame, 6870 bases in length. Protein synthesis begins with the sequence Met-Ala-Thr, and ends with the sequence Leu-Phe-Trp, 126 bases from the 3' end of the RNA. Viral capsid and noncapsid proteins were aligned with the deduced amino acid sequence of the polyprotein. The proteolytic processing map follows the standard 4-3-4 picornaviral pattern except for a short leader peptide (8 kd), which precedes the capsid proteins. Identification of the proteolytic cleavage sites showed that EMC viral protease, p22, has cleavage specificity for gln-gly or gln-ser sequences with adjacent proline residues. The cleavage specificity of the host-coded protease(s) includes both tyr-pro and gln-gly sequences.

## INTRODUCTION

The family of picornaviruses contains four major subdivisions: the entero- (polio, coxsackie), cardio- (encephalomyocarditis (EMC), Mengo), rhino- (120 serotypes) and aphthoviruses (foot-and-mouth disease). Although differing greatly in pathogenicity and epidemiology, all picornaviruses share structural and morphogenic properties. Virus particles contain a single-stranded, positive-sense RNA genome ($2.7 \times 10^6$ MW) enclosed in a protein capsid shell. The capsids are composed of sixty subunits, each of which contains four non-identical polypeptide chains (1).

The 3' ends of viral RNA have poly(A), as is characteristic of most eukaryotic mRNAs (2). However, the 5' ends are not capped in the usual manner with 5'-5' triphosphate linkages. Instead, these viruses have a small, viral-coded, genome-linked protein (VPg) attached by a tyrosine-$O^4$-phosphodiester bond to the 5' pUp of the RNA (3-5). The cardio- and aphthoviral genomes are distinguished by the presence of a 5'-proximal poly(C) tract, whose length (50-150 bases) and exact location relative to the 5' end

of the RNA (150-500 bases) vary with different serotypes and isolates of virus (6-8).

Translation of picornaviral RNA begins about 700 bases from the 5' end, and produces a large precursor polypeptide which represents most of the coding capacity of the genome (9). The polyprotein is processed in a series of proteolytic cleavage steps to yield mature virion capsid proteins, as well as other viral proteins of a non-structural nature. Although protein molecular weights vary somewhat from virus to virus, the processing scheme shows remarkable similarity among the four subdivisions (1,10).

The first cleavage event occurs while the polyprotein is still nascent on a ribosome and is thought to be catalyzed by a cellular protease, whose identity and specificity remain unclear (11-13). Most subsequent maturation processing is effected by a viral-encoded protease, capable of self-cleaving and autocatalytic reactions within the polyprotein (14-16).

The nucleotide sequences of several different serotypes of polio and portions of foot-and-mouth disease virus (FMDV) have now been completed (17-22). These studies, combined with tryptic peptide mapping and protein end-group determinations, have reconfirmed the major elements of picornaviral homology between aphtho- and enteroviruses. However, the sequences have also highlighted important subdivision differences. For example, translation of FMDV polyprotein begins at two initiation sites, 5' to the capsid region, creating leader peptides attached to the capsid precursor protein (23). Polio translation begins at the start of the capsid proteins, without an attached leader sequence (24). FMDV also contains three tandemly-linked, different protein sequences for VPg, as opposed to the single VPg observed in polio (17,18,25).

There are also differences in the cleavage requirements of the viral proteases. The polio-encoded enzyme appears to have an exclusive specificity for gln-gly peptide linkages. The equivalent FMDV cleavages occur within a broader range of protein sequences, including glu-ser, glu-gly and gln-thr (21,25).

Characterization of picornaviral subdivision differences may help identify those elements responsible for variations in host range, cell specificity, pathogenicity and particle stability. Accordingly, we began to analyze the genome of EMC, a cardiovirus, as example and prototype of this group.

We now present the nucleotide and deduced amino acid sequences of the polyprotein coding region of EMC. Our data show that cardioviruses have structural and biochemical elements which appear to be intermediate between

the entero- and aphthoviruses. EMC polyprotein has a 5' leader sequence like FMDV, but the remainder of the processing map more closely resembles polio. The EMC viral protease has an apparent cleavage specificity for gln-gly and gln-ser sequences which are flanked by proline residues.

## MATERIALS AND METHODS

### Materials

EMC virus and purified viral RNA were obtained from Mark Pallansch and Roland Rueckert (26,27). E. coli K12 (strain JM101) and phage M13mp9 were supplied by J. Messing. T4 DNA ligase, DNase I, restriction endonucleases and site-specific primers for di-deoxy M13 sequencing were purchased from New England Biolabs; reverse transcriptase from Life Sciences; large fragment of E. coli DNA polymerase I (Klenow) from Promega-Biotec; $\alpha$ and $\gamma$ $^{32}$P-ATP from Amersham International; oligo(dT) primer and terminal deoxynucleotidyl transferase from P-L Biochemicals; prepacked, mini-C18 columns (Sep-Pac) were from Waters. DNA primer synthesis kits, associated reagents and protocols were obtained from New England Biolabs. E. coli strain MC1000 (28) and plasmid pKMG DNA were prepared and provided by Molecular Genetics Inc. pKMG DNA is a linear 3613 base pair derivative of vector pBR322, which has the EcoR I site at one end of the molecule, and a poly(dG)-tailed PstI site at the other.

### Molecular Cloning of EMC

Viral RNA (20 µg) was reverse transcribed in a 200 µl reaction mixture containing 50 mM Tris-HCl, pH 8.0, 8 mM $MgCl_2$, 75 mM 2-mercaptoethanol, 28 mM KCl, 400 µM each of dATP, dGTP, dCTP and dTTP, 40 µg/ml oligo$(dT)_{12-18}$, and 40 units of reverse transcriptase. Incubation was for 90 min. at 37°C, then the reaction was terminated by the addition of EDTA to 10 mM and SDS to 0.5%. After phenol extraction and precipitation by ethanol, the material was resuspended in a small volume of 1mM EDTA, adjusted to 0.3 M NaOH, heated at 65°C for 20 min., then layered onto an alkaline 5-30% sucrose gradient containing 0.3 M NaOH, 0.5 M NaCl, and 20 mM EDTA. Centrifugation was performed at 20°C in an SW50.1 rotor at 49,000 rpm for 4 hours. Fractions corresponding to DNA of greater than 5000 nucleotides in length were pooled, neutralized and precipitated with ethanol. A portion of this cDNA (100 ng) was incubated in a 10 µl reaction containing 100 mM sodium cacodylate, pH 7.2, 2mM $CoCl_2$, 0.5 mM EDTA, 500 µM dCTP, and 5 units of terminal deoxynucleotidyl transferase for 5 min. at 37°C. The reaction was terminated by addition of EDTA to 10 mM and heating at 100°C for 1 min. This poly(dC)-tailed cDNA was

annealed to the directional cloning vector, pKMG (50 ng), in 200 mM sodium acetate, for 1-2 hours at 42°C. The annealing mixture was precipitated with ethanol, then resuspended in 20 µl of reverse transcriptase buffer (described above). After addition of reverse transcriptase (5 units) and incubation for 1 hour at 37°C, 80 µl of a solution containing 50 mM Tris-HCl, pH 7.2, 10 mM $MgCl_2$, 20 mM dithiothreitol (DTT), 1 mM ATP was added. This mixture was incubated at 20°C for 10 hours with 50-100 units of T4 DNA ligase. A portion of the resultant mixture was used to transform E. coli MC1000 cells, which were then plated on nutrient agar containing 20 µg/ml tetracycline. Plasmids from 20 tetracycline-resistant colonies were screened for the size of cloned DNA insert. One clone, pEM3, contained the largest insert DNA (6100 base pairs) and was used for subsequent analyses.

Sequence of clone pEM3

Plasmid DNA from pEM3 was digested with DNase I under limiting conditions (29) to produce a random assortment of fragments 300-1200 base pairs in length. Additional, defined pieces of pEM3 were prepared by digesting plasmid DNA with PstI plus SalI, isolating the EMC sequence-containing fragments by agarose gel electrophoresis, then redigesting the excised DNA with HpaII. DNA from the DNase I and HpaII reactions was inserted into the vector phage, M13mp9, and used to transfect E. coli JM101. White plaques developing from hybrid virus were picked and regrown. The subcloned viral DNA was sequenced by the method of Sanger (29,30). For some reactions, reverse transcriptase was substituted for Klenow fragment in the sequencing reactions (31).

Selected subclones, which gave difficult regions of band compression on sequencing gels, were also analyzed by chemical methods. Double-stranded, replicative-form viral DNA was isolated from infected JM101 cells (30). The inserted EMC fragments were excised by digestion with EcoRI and BamHI, purified by agarose gel electrophoresis, then labeled with polynucleotide kinase and $\gamma$-$^{32}$P ATP (32). DNA strands were separated by polyacrylamide gel electrophoresis, purified, then sequenced by the methods of Maxam and Gilbert (32). All sequencing gels were prepared and run as described (33).

Primer Extension with Viral RNA Template

Oligodeoxynucleotide primers (9-14 bases in length), complementary to specific, selected segments of the EMC genome, were prepared from New England Biolabs synthesis kits. The preparations (1.5 ml) were titrated to pH 7-8 with acetic acid then applied to mini-C18 columns. After washing with $H_2O$ (5 ml) and 5% methanol (5 ml) the DNA was eluted with 40% methanol (3 ml), then

concentrated (to 200 µl) by evaporation under vacuum. Some primer samples were also purified by fractionation on polyacrylamide gels.

For di-deoxy sequencing experiments, primer (0.1-10 µg/ml) and EMC RNA (50 µg/ml) were annealed in buffer (50 mM Tris-Cl, pH 7.4, 10 mM $MgCl_2$, 50 mM NaCl, 10 mM DTT) by heating for 3 minutes at 67°C and cooling (1°C/min) to 42°C. The complexes were then used in standard reactions with reverse transcriptase (31,34).

For chemical sequencing, the primers were treated with polynucleotide kinase and $\gamma$-$^{32}$P ATP (32), heated for 3 min. at 100°C, then used in reverse transcriptase reactions as described above (Molecular Cloning of EMC). The resulting cDNA was separated from excess primer by gel filtration on Sephadex G150, concentrated by lyophilization, then analyzed by the methods of Maxam and Gilbert (32).

Computer Assisted Assembly of the EMC Sequence

Sequence assembly was aided with programs made available by Rodger Staden (35). Predicted protein translations, sequence manipulation, formatting, and primer searches were carried out with programs developed by John Devereux of the University of Wisconsin Genetics Computer Group (36).

## RESULTS AND DISCUSSION

### Sequence of Clone pEM3

EMC RNA, extracted from plaque purified virus, was used as template in reverse transcriptase reactions with oligo(dT) primer. The resulting cDNA was elongated at the 5' end with dC residues, purified, then annealed with DNA of the directional cloning vector, pKMG, a derivative of pBR322. After second strand synthesis with reverse transcriptase, the hybrid DNA was treated with DNA ligase and used to transform competent *E. coli* cells. The resulting cloned DNAs were screened for length of the inserted segment, and the largest plasmid, pEM3, chosen for sequence analysis.

The EMC-derived segment within pEM3 was excised, fragmented by digestion with DNase I or HpaII, then subcloned into M13 phage. Di-deoxy and Maxam and Gilbert generated sequence data from 165 DNase I subclones and 38 HpaII subclones was identified, matched, oriented and assembled by computer analysis. A total of 36019 bases, representing an average of 5.9 determinations per nucleotide, were used to compile the 6069 base sequence of the EMC segment within pEM3. The entire lengths of both strands of the inserted piece were covered by these analyses. Figure 1 shows the orientation

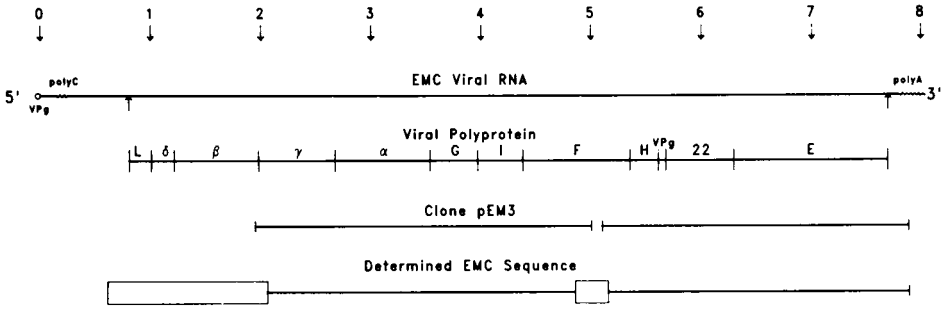



Figure 1. Schematic representation of EMC viral RNA, polyprotein, clone pEM3 and the 7200 base sequence which is depicted in Figure 2. Those bases sequenced by indirect RNA techniques are represented by boxed areas. The break within pEM3 designates the position of the 89 base deletion. The diagram is drawn to scale. Numbers at the top of the figure represent kilobase distances from the 5' end of the RNA.

of pEM3-determined sequence within the EMC genome. The cloned fragment includes the previously identified 3' terminus of the RNA (37) and extends 5' into the region coding for protein β.

The Deletion within pEM3

To encode the large polyprotein (220-260,000 MW) (10,38), EMC RNA must contain an uninterrupted translational reading frame at least 6000 bases in length. However, despite extensive, unambiguous nucleotide analysis, we were unable to find a reading frame of sufficient length within the pEM3 sequence. The longest open frame shifted phase near the middle of the cloned fragment. It was therefore apparent the pEM3 segment contained a discrepant sequence. By comparing homologies between our sequence and that of polio (17,18), we were able to localize the suspect pEM3 region to a 50-100 base sequence centering around the position of the apparent translational phase shift (see Fig 1).

The sequence of this region was then re-examined using viral RNA as template in di-deoxy reactions with reverse transcriptase. A synthetic oligodeoxynucleotide, complementary to an unequivocal EMC sequence region, was used as primer. The resulting gel patterns showed clearly that pEM3 had an 89 base deletion relative to viral RNA. Except for the deleted bases (marked with horizontal arrows on Fig. 2), the indirect RNA sequencing data exactly matched all other nucleotide assignments determined from pEM3.

While the exact molecular event responsible for the deletion within pEM3 is unknown, it is interesting to note that 9 bases at either end of the deleted

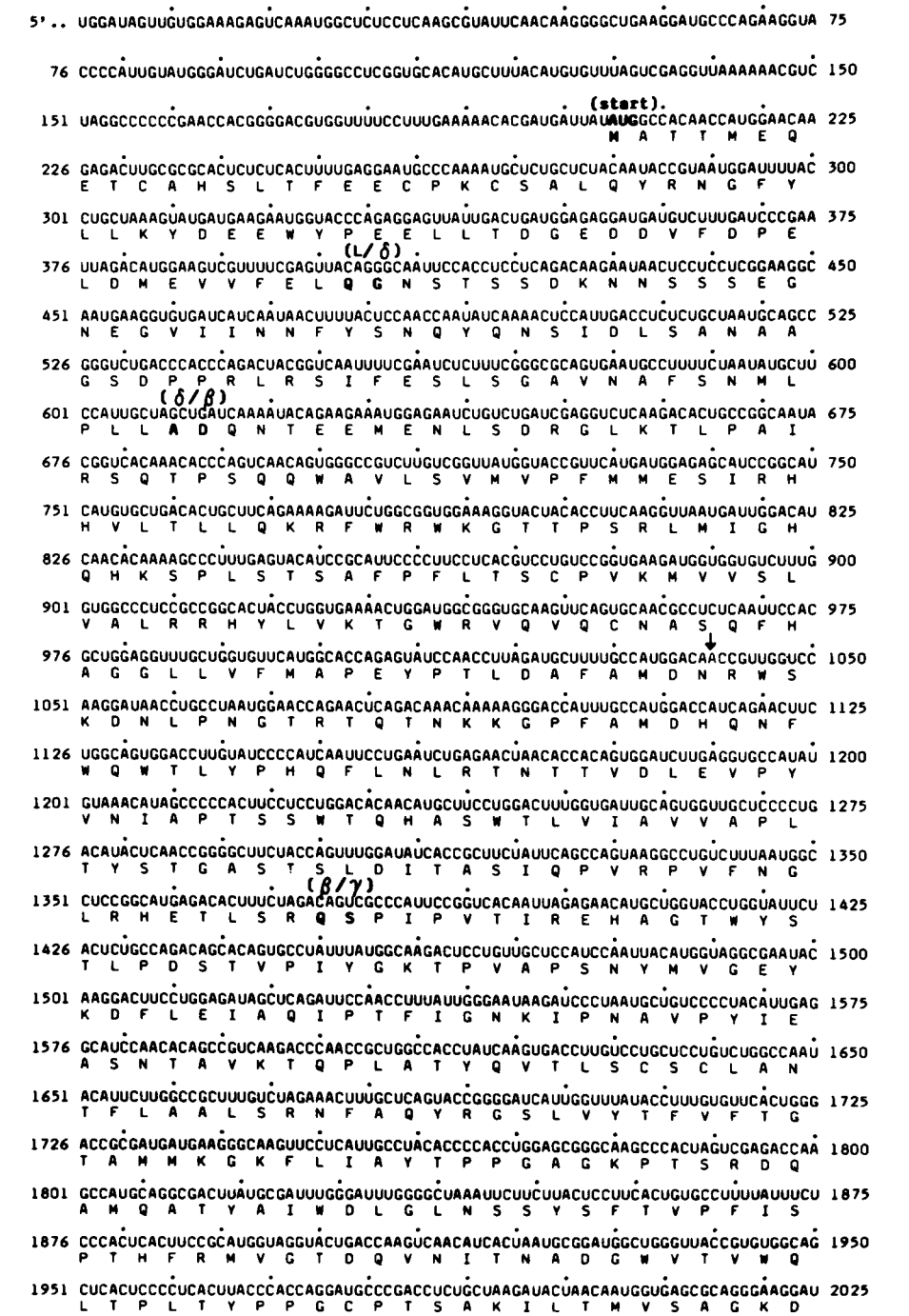
segment have almost identical sequences. The 5' sequence is CCAAUCUGC (Fig. 2, bases 4421-4429) and the 3' sequence is CCAGUCUGC (bases 4510-4518). Possibly, during viral replication or cloning, a polymerizing enzyme skipped from one repeated segment to the other, deleting the intervening bases. We consider it unlikely that such a phenomenon would occur more than once within a molecule, creating other deletions which would be undetectable unless they altered the reading frame. Therefore, we assume the sequence of pEM3, with the exception of these 89 bases, accurately reflects the sequence of the viral RNA from which it was derived.

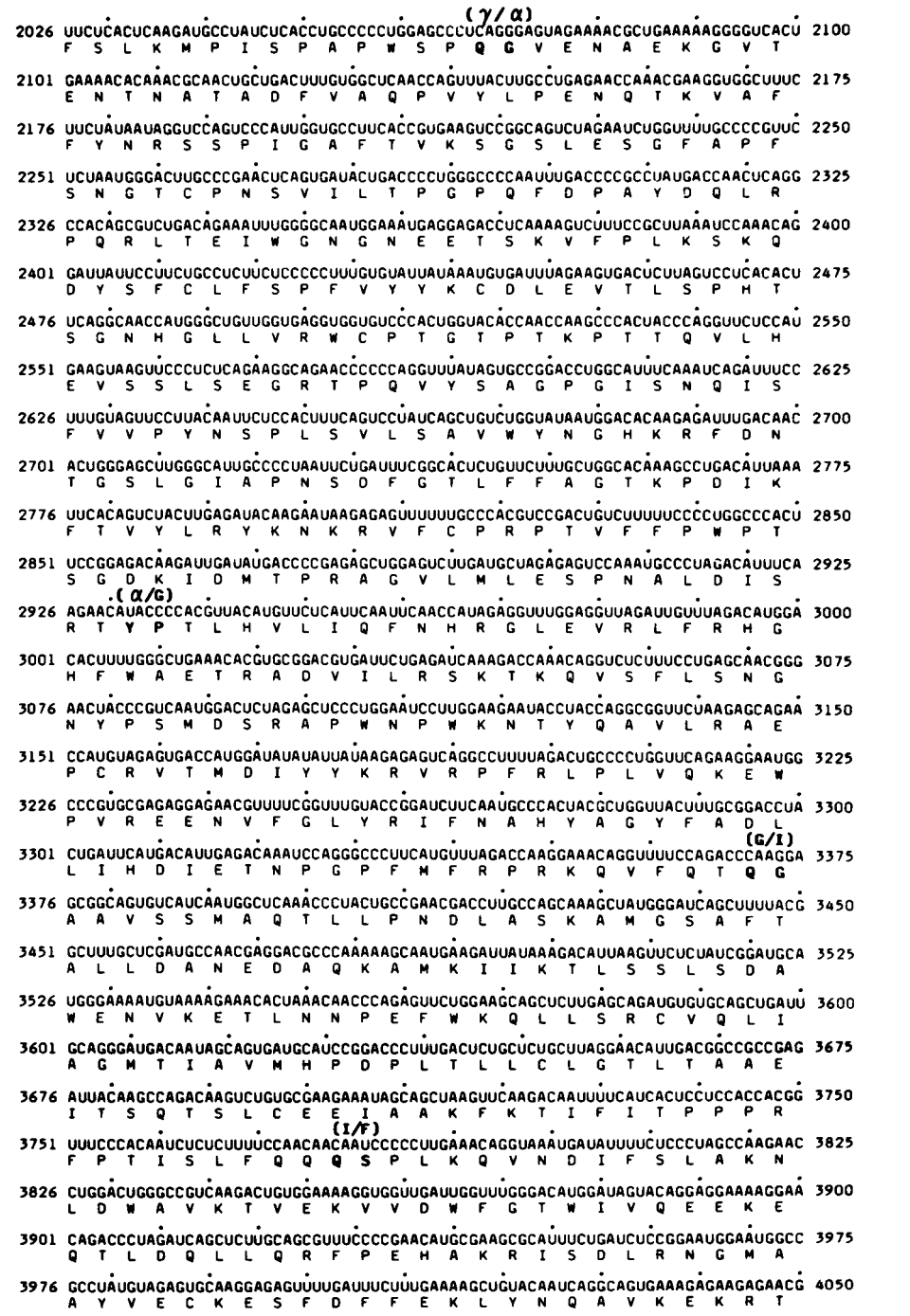
Insertion of the missing bases into the pEM3 sequence established a long open reading frame, ending at UAGUAG codons, 121 bases from the 3' end of the genome (see Fig. 3). Previous sequencing experiments identified this location as the correct termination site of the EMC polyprotein (37). The nucleotide sequence of pEM3 with the deleted bases reinserted, is shown as part of Figure 2. The 5' end of the pEM3 segment is base 1040 of this figure and the 3' end is base 7200. Our sequence corrects several previously misidentified bases near the 3' terminus (37).

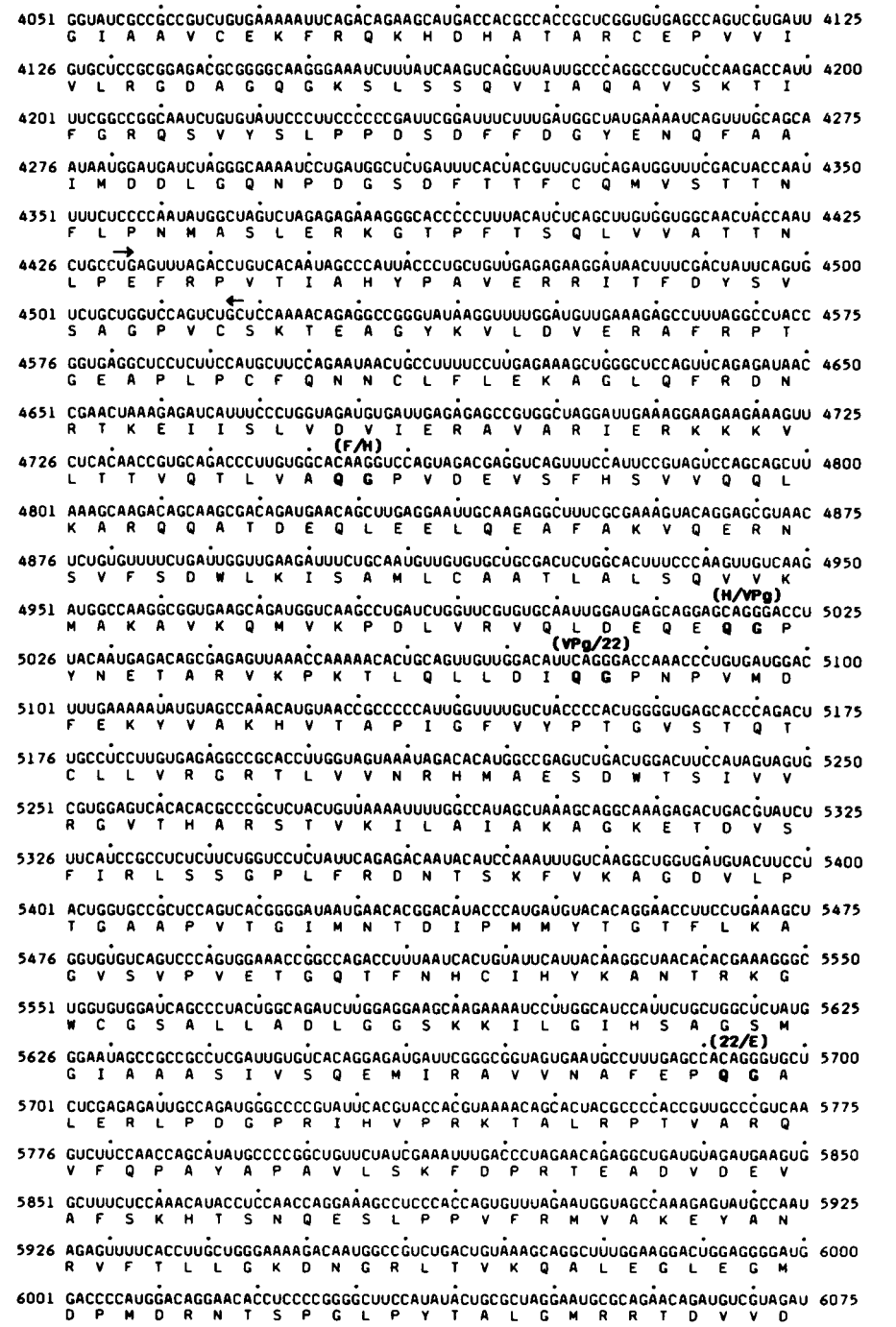
Primer Extension with Viral RNA

Although clone pEM3 was the largest clone generated in our experiments, it represents only 78% of the EMC RNA. It does not include the 5' terminal 1800 bases of the genome. Sequential primer extension reactions with viral RNA as template were used to sequence the remaining portion of the polyprotein coding region. Starting at a sequence within pEM3, a series of oligodeoxynucleotides complementary to successive segments of EMC RNA were selected and synthesized. As each primer was used in di-deoxy reactions with reverse transcriptase, the newly established bases determined the sequence of the next primer in the series. Overlapping sequences from five primers, spaced an average of 220 bases apart, allowed identification of nearly 1200 EMC nucleotides which were not present in the pEM3 clone (see Fig. 1).

The sequence of this segment was verified by chemical techniques. The synthetic oligonucleotides were radiolabeled with polynucleotide kinase and $\gamma$-$^{32}$P ATP, then used to prime reactions with reverse transcriptase and viral RNA. The resulting cDNA products were analyzed by the methods of Maxam and Gilbert. The combined indirect RNA sequencing data, with an average of 4.1 determinations per nucleotide, extended our EMC sequence to within 600 bases of the 5' end of the genome. The sequence of 7200 bases from the 3' end of EMC is shown in Figure 2.


(start).
(δ/β)
(L/δ)
(β/γ)


(γ/α)
(α/G)
(I/F)
(G/I)


(F/M)
(H/VPg)
(VPg/22)
(22/E)

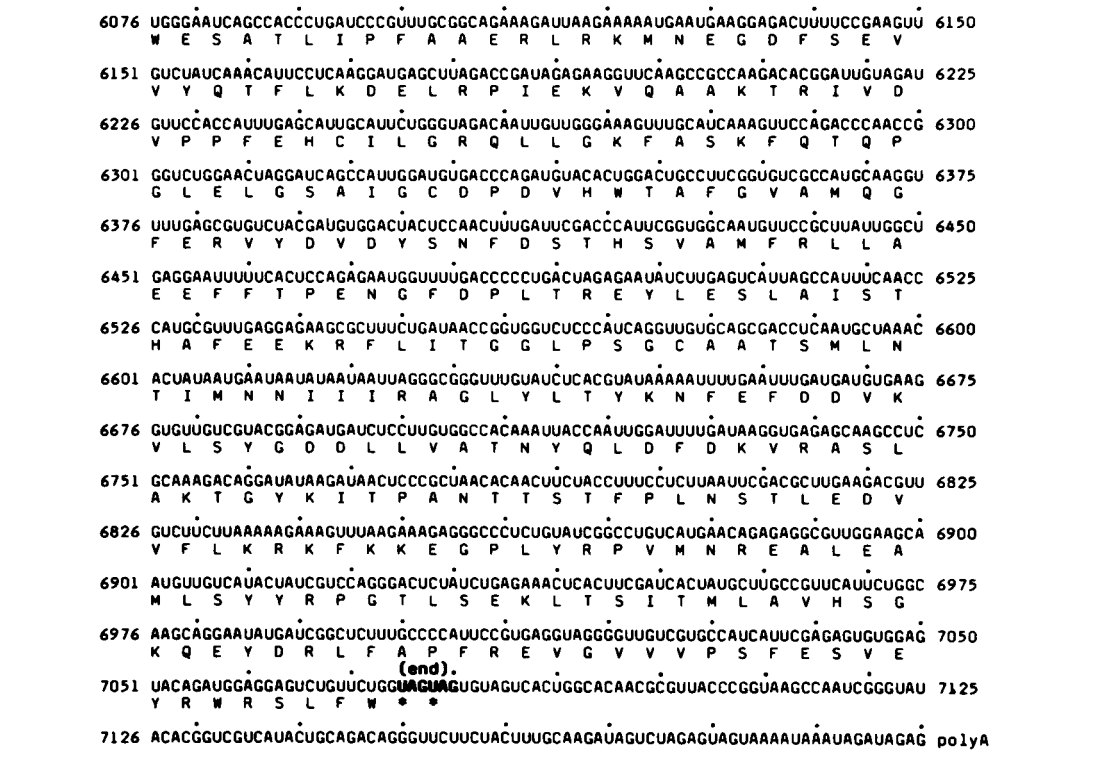



Figure 2. The nucleotide and deduced amino acid sequences of 7200 bases from the 3' end of EMC. Polyprotein translation start and stop codons and proteolytic cleavage sites are highlighted with boldface characters. The vertical arrow at position 1040 designates the 5' end of the pEM3 segment. The right and left arrows at positions 4430 and 4518 represent 5' and 3' ends respectively, of the deletion within pEM3.

## Localization of the Polyprotein

The long open reading frame first detected within pEM3 (designated as frame 1), extends 5' into the region determined by indirect RNA sequencing (see Figure 3). It is bounded on the amino end by a UAG termination codon in position 151. This figure also shows that long open reading frames are not evident in either of the two remaining positive phases. Therefore, viral translation must begin 3' to base 153 within frame 1.

The EMC polyprotein begins with the sequence Met-Ala-Thr followed by a lysine tryptic fragment containing: Glu, Cys, Met, Ser, Pro, Leu, Phe and possibly other (previously unidentified) amino acids (39). The Met-Ala-Thr combination occurs only once within our phase 1 reading frame, beginning at

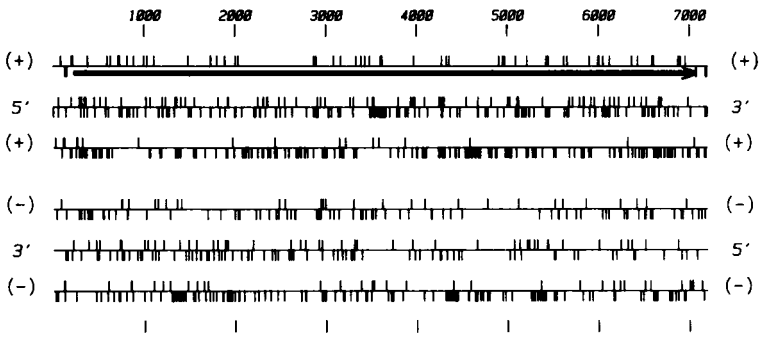



Figure 3. Start and stop codons within the EMC sequence. All positive (+) and negative (-) reading frames within the EMC sequence are represented schematically. Tic marks above the lines indicate positions of AUG codons, tic marks below the lines indicate positions of UGA, UAG or UAA codons. The open reading frame encoding the viral polyprotein is marked with a boldface arrow (top line of the figure) and is designated as reading frame 1. The values across the top of the figure orient base position according to the numbering scheme of figure 2.

base position 205. This deduced sequence is predicted to be part of a tryptic fragment ending in lysine, and containing: Glu, Cys, Met, Ser, Pro, Leu, Phe, Gln, Ala and His (see Fig. 2). Since this sequence so clearly matches the established amino end of the polyprotein, we believe EMC translation must initiate at the AUG triplet in position 205. The reading frame extends 2290 codons to the double termination triplets at position 7075-7080 and predicts a polypeptide of 255466 molecular weight.

The Proteolytic Cleavage Map

Mature virion capsid and other viral non-structural proteins arise by proteolytic cleavage of the polyprotein. Many carboxy- and amino-termini of these proteins have been identified (1,37,ACP, manuscript in preparation). This information, combined with tryptic peptide analyses (14,40), molecular weight determinations (10,12,38,41,42), amino acid compositions (1,43,44) and homology with FMDV, polio and Mengo (17,21,25,45) have allowed alignment of the viral polypeptides with the deduced amino acid sequence of the open reading frame. The proteolytic processing scheme is summarized in Figure 4.

The viral capsid proteins, δ, β, γ and α (7247, 29026, 25141 and 31703 MW, respectively) are preceded by a short leader peptide (7929 MW) and followed by two small proteins designated G (17474 MW) and I (14709 MW). Protein F (36566 MW) lies adjacent to protein H (12166 MW), a VPg-containing precursor (40). The single VPg sequence (2410 MW) is located at the carboxy terminus of H. The

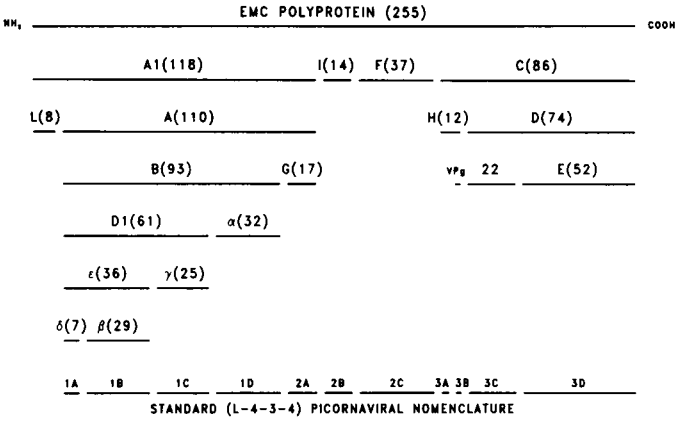



Figure 4. Proteolytic processing map of the EMC polyprotein. EMC proteins and proteolytic cleavage precursors are shown in their relative orientation to the viral polyprotein. The map is drawn to scale. Parentheses indicate the molecular weights (in kilodaltons) of each polypeptide. Standard (4-3-4) picornaviral nomenclature designations (46) are shown at the bottom of the diagram.

protease, p22 (21751 MW), and polymerase, E (51934 MW), were previously mapped (14). Precursor proteins, Al, A, B, C, D, Dl and ε, are shown above their cleavage products.

The processing scheme conforms to the standard 4-3-4 picornaviral polypeptide arrangement (46), and except for the leader sequence is very similar to the primary cleavage map of polio virus (17). Since the leader is most closely analogous to that of FMDV, in some respects the overall structure of EMC can be considered intermediate between the entero- and aphthoviruses. EMC has a leader peptide like FMDV, but the rest of the polyprotein is aligned and processed like polio.

Other reports have suggested the N-terminal position of the EMC cleavage map is occupied by leader peptides (p12/p14) of 12-14000 molecular weight (47). Our protein alignments, which are consistent with the deduced amino acid sequence, indicate a much smaller (7929 MW) leader peptide. However, preliminary nucleotide sequencing data show the possible existence of another open translational reading frame, located between the genomic poly(C) tract and the polyprotein coding region. This segment is potentially long enough to code for protein(s) of at least 14000 molecular weight, and may be the source of p12 and p14. This potential reading frame does not contain a Met-Ala-Thr sequence (ACP, unpublished observations).

| VIRAL PROTEASE CATALYZED CLEAVAGE SITES | PROTEINS |
|---|---|
| Leu-Ser-Arg-GLN / SER-PRO-Ile-Pro | ε/γ |
| Trp-Ser-PRO-GLN / GLY-Val-Glu-Asn | γ/α |
| Phe-Gln-Gln-GLN / SER-PRO-Leu-Lys | I/F |
| Leu-Val-Ala-GLN / GLY-PRO-Val-Asp | F/H |
| Glu-Gln-Glu-GLN / GLY-PRO-Tyr-Asn | H/VPg |
| Leu-Asp-Ile-GLN / GLY-PRO-Asn-Pro | VPg/22 |
| Phe-Glu-PRO-GLN / GLY-Ala-Leu-Glu | 22/E |
| **HOST (?) PROTEASE CATALYZED CLEAVAGE SITES** | |
| Phe-Glu-Leu-GLN / GLY-Asn-Ser-Thr | L/δ |
| Ser-Arg-Thr-TYR / PRO-Thr-Leu-His | α/G |
| Phe-Gln-Thr-GLN / GLY-Ala-Ala-Val | G/I |
| Pro-Leu-Leu-ALA / ASP-Gln-Asn-Thr | δ/β |

CLEAVAGE ↑ SITE

Figure 5. Proteolytic cleavage sequences. The amino acid sequences surrounding all known proteolytic cleavage sites within the EMC polyprotein are tabulated. The sites are identified by the names of the final, stable processing products. See figure 4 to orient these cleavage sequences within precursor proteins Al, A, B, Dl, ε, C, D and H.

## Viral Proteolytic (p22) Cleavage Sites

The EMC viral protease, p22, cleaves between proteins ε/γ, γ/α, H/22, 22/E (14-16). By analogy with the homologous polio enzyme, p22 is also responsible for processing events between I/F, F/H and H/VPg (17). The deduced amino acid sequences surrounding known cleavage sites within the EMC polyprotein are tabulated in Figure 5. All cleavages attributed to p22 are located between gln-gly or gln-ser dipeptides. This result differs from the specificity of the polio protease, which exhibits exclusive preference for gln-gly sequences (17). The equivalent FMDV sites encompass a wider range of sequences, including glu-ser, glu-gly and gln-thr (21,25).

It is not known how the viral proteases recognise or select appropriate cleavage sequences. Both the polio and EMC polyproteins contain gln-gly pairs which do not function in proteolytic reactions (17). However, with EMC, the sequence shows that all sites known to be cleaved by p22 are also flanked by proximal proline residues (see Fig 5). There are six additional gln-gly and gln-ser pairs within the EMC polyprotein. None of these sites lies adjacent to prolines, and none are cleaved by p22. Two of these sequences can serve as processing sites (L/δ and G/I), but they are cleaved by host-derived enzymes, not the viral protease. It remains to be determined, whether the proximity of prolines at p22-catalyzed sites indicates a primary sequence recognition requirement specific to this enzyme, or simply reflects a structural preference towards helix-breaking amino acids near the cleavage points.

Host Proteolytic Cleavage Sites

Capsid precursor protein A1 is released from the nascent EMC polyprotein before the sequences encoding p22 have been translated (12-14). Proteins A and B, as cleavage products of A1, are also evident before the appearance of p22 (12). These observations suggest that release of A1, A or B is catalyzed by an enzyme(s) of cellular origin. Tabulation of amino acids surrounding these sites reveals few common sequences (Fig. 5, L/δ, α/G and G/I). Two of the cleavages occur at gln-gly pairs (not flanked by prolines), but it is not obvious how these sites are specifically distinguished from others within the polyprotein. Our sequencing data give no indication as to the nature or origin of the amino-terminal blocking group found on mature protein δ (1). It is completely unknown whether host-catalyzed proteolysis of L/δ has any effect on this reaction.

The tyr-pro protein cleavage (α/G) is at a site equivalent to the tyr-gly host-catalyzed cleavage within the polio polyprotein (17). Even though these polio and EMC sequences are both preceded by a threonine residue and have histidine in the plus 4 position, it is again not clear whether this observation implies a direct primary recognition requirement, or a selective preference towards helix-breaking amino acids.

The maturation processing of protein ε to δ plus β is the final proteolytic event within the EMC polyprotein. This cleavage is observed only during the last stages of virion morphogenesis and may play a role in stabilizing mature, RNA-containing particles (1). EMC and FMDV share considerable amino acid homology in the region surrounding this cleavage site (P-L-L-A/D-Q-N-T-E-E for EMC, and A-L-L-A/D-K-K-T-E-E for FMDV) (21,22), but neither sequence has as much in common with the equivalent polio site (P-M-L-N/S-P-N-I-E-A) (17). Although the viral or host agents responsible for these cleavages remain to be identified, the homology between EMC and FMDV suggests a similar proteolytic agent may be utilized by these two viruses, if not also by polio.

Picornaviral Comparisons

Comprehensive comparisons among EMC, polio and FMDV have been undertaken and will be published separately. Base composition, codon preference and dot matrix analyses of protein and nucleic acid sequences will be included (ACP, manuscript in preparation).

ACKNOWLEDGEMENTS

This research was supported by NIH grant AI 17331. We thank Roland

Rueckert and Paul Kaesberg for helpful discussions and critical reading of the manuscript.

*To whom reprint requests should be sent

REFERENCES

1. Rueckert, R. (1976) in Comprehensive Virology, Fraenkel-Conrat, H. and Wagner, R. Eds., Vol. 6, pp 131-213, Plenum Publishing Corp., New York.
2. Ahlquist, A. and Kaesberg, P. (1979) Nucleic Acids Research 7, 1195-1204.
3. Vartapetian, A., Drygin, Yu., Chumakov, K. and Bogdanova, A. (1980) Nucleic Acids Research 8, 3729-3742.
4. Rothberg, P., Harris, T., Nomoto, A. and Wimmer, E. (1978) Proc. Nat. Acad. Sci. USA 75, 4868-4872.
5. Ambrose, V. and Baltimore, D. (1978) J. Biol. Chem. 253, 5263-5266.
6. Black, D., Stephenson, P., Rowlands, D. and Brown, F. (1979) Nucleic Acids Research 6, 2381-2389.
7. Harris, T. and Brown, F. (1976) J. Gen. Virol. 33, 493-501.
8. Chumakov, K. and Agol, V. (1976) Biochem. Biophys. Res. Com. 71, 551-557.
9. Jacobson, M. and Baltimore, D. (1968) Proc. Nat. Acad. Sci. USA 61, 77-84.
10. Butterworth, B. and Korant, B. (1974) J. Virol. 14, 282-291.
11. Shih, D., Shih, C., Zimmern, D., Rueckert, R. and Kaesberg, P. (1979) J. Virol. 30, 472-480.
12. Shih, D. and Shih, C. (1981) J. Virol. 40, 942-945.
13. Pelham, H. (1978) Eur. J. Biochem. 85, 457-462.
14. Palmenberg, A., Pallansch, M. and Rueckert, R. (1979) J. Virol. 32, 770-778.
15. Palmenberg, A. and Rueckert, R. (1982) J. Virol. 41, 244-249.
16. Gorbalenya, A., Svitkin, Yu. and Agol, V. (1981) Biochem. Biophys. Res. Comm. 98, 952-960.
17. Kitamura, N., Semler, B., Rothberg, P., Larsen, G., Adler, C., Dorner, A., Emini, E., Hanecak, R., Lee, J., van der Werf, S., Anderson, C. and Wimmer, E. (1981) Nature 291, 547-553.
18. Racaniello, V. and Baltimore, D. (1981) Proc. Nat. Acad. Sci. USA 78, 4887-4891.
19. Robertson, B., Morgan, D., Moore, D., Grubman, M., Card,J., Fischer, T., Weddell, G., Dowbenko, D. and Yansura, D. (1983) Virology 126, 614-623.
20. Harris, T. (1979) Nucleic Acids Research 7, 1765-1785.
21. Boothroyd, J., Harris, T., Rowlands, D. and Lowe, P. (1982) Gene 17, 153-161.
22. Carroll, A., Rowlands, D. and Clarke, B. (1984) Nucleic Acids Research, manuscript submitted.
23. Beck, E., Forss, S., Strebel, K., Cattaneo, R. and Feil, G. (1983) Nucleic Acids Research 11, 7873-7885.
24. Dorner, A., Dorner, L., Larsen, G., Wimmer, E. and Anderson, C. (1982) J. Virol. 42, 1017-1028.
25. Forss, S. and Schaller, H. (1982) Nucleic Acids Research 10, 6441-6450.
26. Hall, L. and Rueckert, R. (1971) Virology 43, 152-165.
27. Pallansch, M. and Rueckert, R. (1981) Methods in Enzymology 78, 315-325.
28. Casadaban, M. and Cohen, S. (1980) J. Mol. Biol. 138, 179-207.
29. Sanger,F., Nicklen, S. and Coulson, A. (1977) Proc. Nat. Acad. Sci. USA 74, 5463-5476.
30. Messing, J. (1983) Methods in Enzymology 101, 20-78.

31. Smith, A. (1980) Methods in Enzymology 65, 560-580.
32. Maxam, A. and Gilbert, W. (1980) Methods in Enzymology 65, 499-560.
33. Garroff, H. and Ansorge, W. (1981) Analyt. Biochem. 115, 454-457.
34. Zimmern, D. and Kaesberg, P. (1978) Proc. Nat. Acad. Sci. USA 75, 4257-4261.
35. Staden, R. (1980) Nucleic Acids Research 8, 3673-3694.
36. Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Research 12, 378-395.
37. Drake, N., Palmenberg, A., Ghosh, A., Omilianowski, D. and Kaesberg, P. (1982) J. Virol. 41, 726-729.
38. Butterworth, B. and Rueckert, R. (1972) Virology 50, 535-549.
39. Smith, A. (1973) Eur. J. Biochem. 33, 301-313.
40. Pallansch, M., Kew, O., Palmenberg, A., Golini, F., Wimmer, E. and Rueckert, R. (1980) J. Virol 35, 414-419.
41. Butterworth, B., Hall, L., Stoltzfus, C. and Rueckert, R. (1971) Proc. Nat. Acad. Sci. USA 68, 3083-3087.
42. Butterworth, B. and Rueckert, R. (1972) J. Virol. 9, 823-828.
43. Stoltzfus, M. and Rueckert, R. (1972) J. Virol. 10, 347-355.
44. Ziola, B. and Scraba, D. (1975) Virology 64, 228-235.
45. Ziola, B. and Scraba, D. (1976) Virology 71, 111-121.
46. Rueckert, R. and Wimmer, E. (1984) J. Virol., in press.
47. Kazachkov, Yu., Chernovskaye, T., Siyanova, E., Svitkin, Yu., Ugarova, T. and Agol, V. (1982) FEBS Letters 141, 153-156.

Note added in proof: Since submission of this manuscript, portions of the 5' end of EMC, which were originally analyzed by indirect RNA sequencing methods, have been cloned into DNA and re-examined. The clones were sequenced (both strands) by Maxam and Gilbert techniques and the newly generated nucleotide assignments agree exactly with those presented in Figure 2 (ACP, manuscript in preparation).