

Nucleotide Sequence and Expression of the Capsid Protein Gene of Feline Calicivirus

JOHN D. NEILL,^{1†*} ILENE M. REARDON,² AND ROBERT L. HEINRIKSON²

National Animal Disease Center, Agricultural Research Service, U.S. Department of Agriculture, Ames, Iowa 50010,¹ and Upjohn Co., Kalamazoo, Michigan 49001²

Received 7 June 1991/Accepted 19 July 1991

The sequence of the 3'-terminal 2,486 bases of the feline calicivirus (FCV) genome was determined. This region of the FCV genome, from which the 2.4-kb subgenomic RNA is derived, contained two open reading frames. The larger open reading frame, found in the 5' end of the subgenomic mRNA, contained 2,004 bases encoding a polypeptide of 73,467 Da. The smaller open reading frame, encoded in the 3' end of the mRNA, was composed of 318 bases, encoding a polypeptide of 12,185 Da. The AUG initiation codon of the second open reading frame overlapped the UGA termination codon of the first, with the sequence AUGA. The nucleotide sequence of the region containing this overlap resembles the -1 frameshift sequences of the retroviruses. The 5' end of the 2.4-kb subgenomic RNA was mapped by primer extension analysis. There were two apparent transcription initiation points, both of which were 5' to the AUG initiation codon of the large open reading frame. Transcription from these sites yielded RNA transcripts with 5' nontranslated leader regions of 17 and 18 bases. The total length of the 2.4-kb subgenomic RNA was 2,375 bases (from the 5'-most start site) excluding the poly(A) tail. Edman degradation of the purified capsid protein of FCV showed that the capsid protein was encoded by the large open reading frame. Western immunoblot analysis of FCV-infected cells using a feline anti-FCV antiserum demonstrated that translation of the capsid protein was detectable at 3 h postinfection and continued to accumulate until 8 h postinfection, the last time examined.

Little is known concerning the molecular biology of the caliciviruses. Members of this family include vesicular exanthema virus of swine, San Miguel sea lion virus, and feline calicivirus (FCV) (28). These are small, nonenveloped viruses that possess a single-stranded, plus-sense RNA genome and only a single capsid protein (1, 30). These viruses also produce a major subgenomic mRNA (9, 24). The genomic RNA has been shown to have a protein (VPg) of 10,000 to 15,000 Da covalently attached to the 5' end (4, 29). It is unknown whether the subgenomic mRNA has a VPg, although it has been demonstrated that there is no cap structure (10). Estimates of the size of the capsid proteins from the various family members range from 60,000 to 71,000 Da (28). Fretz and Schaffer (11) reported that the capsid protein of San Miguel sea lion virus serotype 2 is synthesized as an 86,000-Da precursor polypeptide that is processed posttranslationally to yield the mature capsid protein of 65,000 Da (cp65). Carter (6) demonstrated the capsid protein in FCV-infected cells was synthesized as a 76,000-Da precursor and is post- or cotranslationally processed to the mature 62,000-Da capsid protein. The fate of the remainder of the polypeptide is unknown.

Recent studies have shown that the virus-specific polyadenylated RNAs in FCV-infected cells are coterminal, nested transcripts with common 3' ends (24). Transcription of the viral RNAs was detected at 2 h postinfection, with steady-state levels being reached at 4 h postinfection. Studies involving sucrose gradient density centrifugation of radiolabeled RNA (2, 9) and Northern (RNA) blot analysis (24) have shown that the 2.4-kb subgenomic RNA, which is derived from the 3'-terminal 2,400 bases of the genomic

RNA, is abundant in calicivirus-infected cells. On the basis of this abundance, Ehresmann and Schaffer (9) hypothesized that the subgenomic RNA encodes the viral capsid protein. A similar inference was drawn by Black et al. (2) from an *in vitro* translation study. However, to date there has been no definitive evidence in support of this assertion.

This report is the first to establish that the 2.4-kb subgenomic RNA encodes the capsid protein of FCV. Comparison of the amino acid sequence of a portion of the capsid protein of FCV with the derived amino acid sequence from the large open reading frame (ORF) encoded in the 3'-terminal 2,400 bases of the FCV genomic RNA demonstrates that these sequences encode the capsid protein. Studies on the expression of the capsid protein gene during replication of FCV are also described.

MATERIALS AND METHODS

Materials. All restriction endonucleases, *Escherichia coli* DNA polymerase I, Klenow fragment, T4 DNA ligase, exonuclease III, mung bean nuclease, T7 and SP6 RNA polymerases, and *EcoRI* synthetic linkers were purchased from New England BioLabs, Inc. (Beverly, Mass.) and used as specified by the supplier. Reverse transcriptase was purchased from Bio-Rad (Richmond, Calif.) and used as directed by the supplier. [α -³²P]dATP and [α -³²P]UTP were from ICN Radiochemicals, Inc. (Irvine, Calif.), and [α -³⁵S]dATP was from New England Nuclear (Boston, Mass.). Feline anti-FCV antiserum was a kind gift of G. Erickson (Animal and Plant Health Inspection Service, U.S. Department of Agriculture). This antiserum was derived from a cat following an active FCV infection.

Strains and medium. FCV strain CFI/68 FIV (hereafter referred to as FCV; American Type Culture Collection) was used throughout this study. FCV was propagated in Crandell-Reese feline kidney cells by using F-15 Eagle's minimal

* Corresponding author.

† Present address: Department of Biotechnology Research, Pioneer Hi-Bred International, Inc., 7300 NW 62nd Avenue, Johnston, IA 50131.

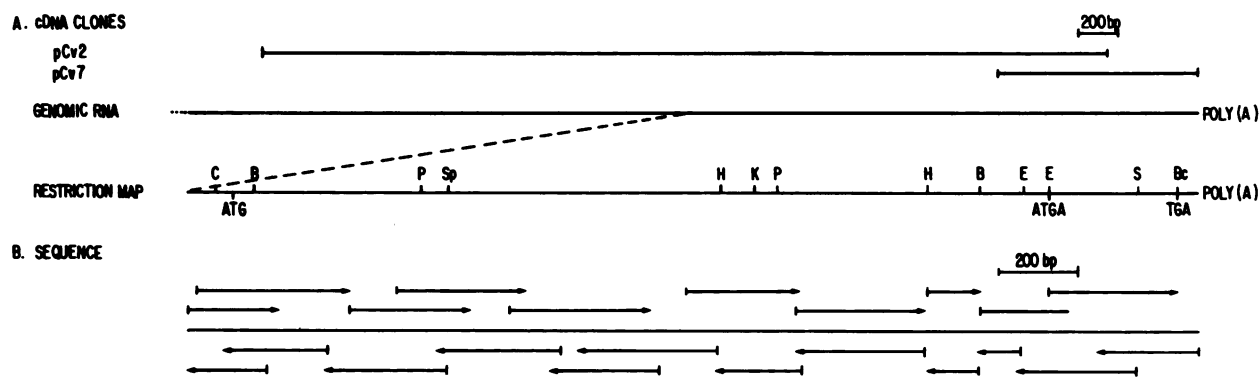


FIG. 1. Physical map and sequencing strategy of the 3'-terminal 2,486 bases of the FCV genomic RNA. (A) Positions of the cDNA clones pCv2 and pCv7 within the 3' end of the FCV genome. The entire genomic RNA is not illustrated. The restriction map of the 3'-terminal 2,486 bases is shown. Restriction sites: B, *Bam*HI; Bc, *Bcl*I; C, *Cl*aI; E, *Eco*RI; H, *Hind*III; K, *Kpn*I; P, *Pvu*II; S, *Sal*I; Sp, *Sph*I. Positions of the initiation (ATG) and termination (TGA) codons are indicated beneath the restriction map. (B) Strategy for sequencing of pCv2 and pCv7.

essential medium supplemented with 0.25% lactalbumin hydrolysate and 10% fetal calf serum. Infection of cells for RNA preparation, virus purification, and preparation of protein lysates for Western immunoblot analysis were done as previously described (24).

E. coli DH1 (12) was used for plasmid propagation. Strain JM107 was used to propagate M13mp18 and M13mp19 (34) for use in DNA sequence analysis and in vitro deletion generation. JM107 was maintained on minimal medium as previously described (34).

RNA purification and Northern blot analysis. Total cellular and poly(A)⁺ RNA from noninfected and FCV-infected cells was purified and analyzed by Northern blot analysis as previously described (24). Strand-specific FCV RNA probes were prepared by first subcloning the 1.3-kb *Eco*RI FCV cDNA clone contained in pCV3 (24) into pSP72, an RNA transcription vector containing both SP6 and T7 RNA polymerase promoters (Promega, Madison, Wis.), yielding pCV16. These probes hybridized to both the genomic and 2.4-kb subgenomic RNAs. The plus-strand RNA probe was transcribed with SP6 RNA polymerase following *Bgl*II linearization of the plasmid. The minus-strand RNA probe was generated with T7 RNA polymerase preceded by linearization of the plasmid with *Xho*I. The labeling reactions were done as described by the enzyme supplier in the presence of 50 μ Ci of [α -³²P]UTP.

DNA sequence analysis. The FCV cDNA clones pCV2 and pCV7 were sequenced by the dideoxy-chain termination procedure as described by Williams et al. (33), using [α -³⁵S]dATP. The construction of pCV2 and pCV7 has been described (24). Sequencing of the cDNA clones was facilitated by the construction of an M13-FCV cDNA library of nested deletions, using exonuclease III and mung bean nuclease as described by Shade et al. (31). The accuracy of the nucleotide sequence of the region containing the overlap of the two ORFs was confirmed by amplification of this region by the polymerase chain reaction (PCR) as described by Kawasaki (16) following first-strand cDNA synthesis using the genomic RNA as the template as previously described (24). The PCR primers used were 5'-CTGCCTCC TCCATGGGAATTC-3', which hybridized to the minus strand at nucleotides 2051 to 2072, and 5'-GCACTCCAACC CATGGACTTGGATCG-3', which hybridized to the plus strand at nucleotides 2226 to 2251.

5'-End analysis of the 2.4-kb RNA. Primer extension anal-

ysis was done to map the 5' end of the 2.4-kb subgenomic mRNA, using an oligonucleotide 20-mer, 5'-GCCAAATT CAGGTAGTAATT-3'. This primer hybridized to the FCV RNAs at nucleotides 248 to 267 of Fig. 2. The primer extension reaction was done as previously described (17). A dideoxy sequencing reaction, which was primed with this synthetic oligonucleotide primer, was electrophoresed on the same gel as a marker for transcription endpoint analysis.

FCV capsid protein sequence analysis. Since attempts to sequence the intact capsid protein by automated Edman degradation were unsuccessful, we concluded that the NH₂ terminus was blocked, and a strategy was formulated to generate fragments of the molecule for structural analysis. Accordingly, a solution containing 50 μ g (0.75 nmol) of the capsid protein in 300 μ l of phosphate-buffered saline (pH 7.2) was treated with 1.0% (wt/wt) trypsin (Sigma T-8642). After 48 h at 25°C, the solution was lyophilized. Tryptic peptides from the capsid protein were separated by reverse-phase high-performance liquid chromatography (HPLC) on a Hypersil ODS column (25 by 0.46 cm) equilibrated in 0.15% trifluoroacetic acid (TFA). The lyophilized digest was dissolved in 0.2 ml of 0.15% TFA and added to the column. Peptides were eluted over a period of 3 h with a linear gradient of increasing acetonitrile concentration from 0 to 100% in the same 0.15% TFA solvent. Separation of peptides was monitored continuously at 220 nm, and fractions corresponding to peaks of material were collected and lyophilized prior to sequence analysis in a model 470 Applied Biosystems Inc. gas-phase protein sequencer. This instrument was fitted with an on-line HPLC system for separation and quantitation of phenylthiohydantoin amino acids liberated at successive cycles of the Edman procedure.

Western immunoblot analysis. FCV capsid protein synthesis was analyzed by polyacrylamide gel electrophoresis and Western blot analysis. Protein lysates were prepared hourly from noninfected control cells and from FCV-infected cells beginning at 3 h and continuing until 8 h postinfection. Lysates were prepared as described by Carter (6) and were stored at -20°C until use. Approximately 20 μ g of total protein from each lysate was electrophoresed on a 12% polyacrylamide gel. Following electrophoresis, the proteins were blotted to nitrocellulose, using a semidry blotting apparatus (Bio-Rad) as specified by the manufacturer. The blots were probed with a feline anti-FCV antiserum, and positive reactions were detected by goat anti-feline immu-

TTCTTACAGCACAGCCTTGGAGCATTACAACAGCCAATTTAATGGTGTGGAGGCGGGAC	60
TGACCAGATCGATACGAGTGGCATGGCCGCCCTTCACTGTGATGTGTTCGAAGTTTGAGC	120
ATGTGCTCAACCTGCGCTAACGTGCTTAAATACTATGATTGGGATCCTCACATCAAATTG	180
METCysSerThrCysAlaAsnValLeuLysTyrTyrAspTrpAspProHisIleLysLeu	20
GTAATCAACCCCAACAAATTTCTACATGTTGGCTTCTGCGATAACCCCTTTAATGTGTGT	240
ValIleAsnProAsnLysPheLeuHisValGlyPheCysAspAsnProLeuMetCysCys	40
TATCCTGAATTACTACCTGAATTTGGCACCATGTGGGATTGTGATCAATCGCCACTCCAA	300
TyrProGluLeuLeuProGluPheGlyThrMetTrpAspCysAspGlnSerProLeuGln	60
GTCTACCTTGAGTCAATCTGGGTGATGATGAATGGTCTCCACTCATGAAGCAATTGAC	360
ValTyrLeuGluSerIleLeuGlyAspAspGluTrpSerSerThrHisGluAlaIleAsp	80
CCAGTTGTGCCACCAATGCATTGGGATGAAGCCGAAAAATCTTCCAACCACCCCTGGC	420
ProValValProProMetHisTrpAspGluAlaGlyLysIlePheGlnProHisProGly	100
GTCCTTATGCATCACCTCATCTGTAAAGTTGCAGAAGGATGGGACCCAAACCTGCCACTT	480
ValLeuMetHisHisLeuIleCysLysValAlaGluGlyTrpAspProAsnLeuProLeu	120
TTCCGCTTGAAGCGGACGATGGTTCCATCAGCACCTGAACAGGGAACAATGGTTGGT	540
PheArgLeuGluAlaAspAspGlySerIleThrThrProGluGlnGlyThrMetValGly	140
GGAGTCATGCTGAGCCCAACGCCAAATGTCAACCGCAGCTGACATGGCCACTGGGAAA	600
GlyValIleAlaGluProAsnAlaGlnMetSerThrAlaAlaAspMetAlaThrGlyLys	160
AGTGTGGACTCTGAGTGGGAAGCCTTCTTCTCCTTTACACTAGTGTGAAGTGGAGCACA	660
SerValAspSerGluTrpGluAlaPhePheSerPheHisThrSerValAsnTrpSerThr	180
TCTGAAACTCAGGGGAAGATACTCTTAAACAATCCTTAGGACCATTGCTCAACCCCTAC	720
SerGluThrGlnGlyLysIleLeuPheLysGlnSerLeuGlyProLeuLeuAsnProTyr	200
CCTACCCATCTTGCAAAGCTGTATGTTGCTTGGTCTGGTCTGTTGATGTTAGTTTTCT	780
LeuThrHisLeuAlaLysLeuTyrValAlaTrpSerGlySerValAspValArgPheSer	220
ATTTCTGGATCTGGTGTCTTTGGAGGGAAATAGCTGCTATTGTTGTGCCGCCAGGAAT	840
IleSerGlySerGlyValPheGlyGlyLysLeuAlaAlaIleValValProProGlyIle	240
GATCCTGTTCAAAGTACTTCAATGCTGCAATATCCTCATGCTCCTTTGATGCTCGTCAA	900
AspProValGlnSerThrSerMetLeuGlnTyrProHisValLeuPheAspAlaArgGln	260
GTTGAACCTGTTATCTTTCCATTCCCGATCTAAGAAGCACCTTATATCACCTTATGTCT	960
ValGluProValIlePheSerIleProAspLeuArgSerThrLeuTyrHisLeuMetSer	280
GACACTGATACCACATCGTTGGTAATCATGGTGTACAATGATCTTATTAACCCCTATGCT	1020
AspThrAspThrThrSerLeuValIleMetValTyrAsnAspLeuIleAsnProTyrAla	300
AATGACTCAAACCTTCGGGCTGCATTGTCACTGTGAAACTAAACCGGGCCAGATTTT	1080
AsnAspSerAsnSerSerGlyCysIleValThrValGluThrLysProGlyProAspPhe	320
AAGTTTACCTCTTAAAACCTCCTGGGTCTATGTTAACTCACGGATCTATCCCATCTGAT	1140
LysPheHisLeuLeuLysProProGlySerMetLeuThrHisGlySerIleProSerAsp	340
CTAATCCAAAATCATCTTCGCTTTGGATTGGAAATCGGTTTTGGTCTGACATAACCGAT	1200
LeuIleProLysSerSerSerLeuTrpIleGlyAsnArgPheTrpSerAspIleThrAsp	360
TTTGTAAATTCGGCCTTTTGTGTTCCAGGCAAATCGACACTTTGATTTCAACCAAGAGACA	1260
PheValIleArgProPheValPheGlnAlaAsnArgHisPheAspPheAsnGlnGluThr	380
GCAGGTTGGAGCACCCCAAGGTTTCGCCCAATTACTATCACTATCAGTGTAAAGGAGTCA	1320
AlaGlyTrpSerThrProArgPheArgProIleThrIleThrIleSerValLysGluSer	400

FIG. 2. Nucleotide and derived amino acid sequences of the 3'-terminal 2,486 bases of the FCV genomic RNA. Nucleotides in bold print represent those bases found within the large ORF at the 5' end of the FCV genome. Underlined nucleotide residues indicate the major transcription start sites of the 2.4-kb subgenomic RNA. Amino acid residues in bold type are those that were confirmed by protein sequence analysis.

noglobulin G-horseradish peroxidase-conjugated secondary antibodies.

Computer software. DNA sequence analysis was done by using PCS DNA analysis software (University of Iowa, Iowa City) or the Bionet data base (Intelligenetics, Mountain View, Calif.).

Nucleotide sequence accession number. The nucleotide

sequence shown in Fig. 2 has been submitted to GenBank and assigned accession number M32819.

RESULTS

Sequence analysis of the 3'-terminal 2,500 bases. The 3' 2,489 bases of the FCV genome contained in the cDNA

GCAAAGCTTGGTATGGAGTGGCCACCGACTACATTGTTCCCGGCATACCAGATGGATGG 1380
AlaLysLeuGlyIleGlyValAlaThrAspTyrIleValProGlyIleProAspGlyTrp 420

CCCGACACAACAATCCCAGGTGAGTTGGTACCTGTTGGTGACTATGCCATCACTAATGGC 1440
ProAspThrThrIleProGlyGluLeuValProValGlyAspTyrAlaIleThrAsnGly 440

ACCAACAATGATATCACCACAGCTGCGCAGTACGATGCAGCCACTGAGATTAGAAACAAC 1500
ThrAsnAsnAspIleThrThrAlaAlaGlnTyrAspAlaAlaThrGluIleArgAsnAsn 460

ACCAATTCAGAGGCATGTACATTTGTGGTCTCTTCAAAGAGCTTGGGGGATAAGAAG 1560
ThrAsnPheArgGlyMetTyrIleCysGlySerLeuGlnArgAlaTrpGlyAspLysLys 480

ATTTCAAATACTGCTTTTATCACAACCGGCACGGTTGATGGAGCCAAATTGATACCCAGT 1620
IleSerAsnThrAlaPheIleThrThrGlyThrValAspGlyAlaLysLeuIleProSer 500

AATACCATTGACCAAAACAAAATTGCCGTATTCGAAGACACATGCGAATAAGCATGTC 1680
AsnThrIleAspGlnThrLysIleAlaValPheGlnAspThrHisAlaAsnLysHisVal 520

CAGACCTCGGACGACACATGGCCCTGCTTGGTTATACTGGTATTGGTGAGGAAGCAATT 1740
GlnThrSerAspAspThrLeuAlaLeuLeuGlyTyrThrGlyIleGlyGluGluAlaIle 540

GGTGCTGACCGGATAGAGTTGTGCGAATTAGCGTCCCTCCCGAACGTGGCGCACGTGGT 1800
GlyAlaAspArgAspArgValValArgIleSerValLeuProGluArgGlyAlaArgGly 560

GGCAATCACCCAATCTCCACAAAACCTCTATCAAGCTTGGTTATGTAATTAGGTCCATT 1860
GlyAsnHisProIlePheHisLysAsnSerIleLysLeuGlyTyrValIleArgSerIle 580

GATGTGTTCAATTCTCAAATCTGCATACCTCTAGGCAACTTCCCTCAATCATTACTTA 1920
AspValPheAsnSerGlnIleLeuHisThrSerArgGlnLeuSerLeuAsnHisTyrLeu 600

TTGTGCGCTGACTCCTTGTCTATAGGATTATTGACTCTAATGGATCCTGGTTTGAC 1980
LeuSerProAspSerPheAlaValTyrArgIleIleAspSerAsnGlySerTrpPheAsp 620

ATAGGCATTGATAATGATGGATTTTCTTTTGTGGTGTATCAAGTATTGGTAAATTAGAG 2040
IleGlyIleAspAsnAspGlyPheSerPheValGlyValSerSerIleGlyLysLeuGlu 640

TTTCCTTTAACTGCCTCCTACATGGGAATTCAATTGGCAAAAATTCGACTTGCCTCTAAC 2100
PheProLeuThrAlaSerTyrMetGlyIleGlnLeuAlaLysIleArgLeuAlaSerAsn 660

ATTAGGAGTGTGATGACAAAATTATGAATTCAATTTTGGGCTTAATTGACTGTACCGA 2160
IleArgSerValMetThrLysLeuTER 668
METAsnSerIleLeuGlyLeuIleAspThrValThrA 12

ACACAATTGGCAAAGCTCAACAAATCGAATTGGATAAGGCTGCACCTGGTCAGCAACGCG 2220
snThrIleGlyLysAlaGlnGlnIleGluLeuAspLysAlaAlaLeuGlyGlnGlnArgG 32

AGCTGGCACTCCAACGTATGAACCTGGATCGCCAGGCTCTAAATAATCAAGTGGAGCAAT 2280
luLeuAlaLeuGlnArgMetAsnLeuAspArgGlnAlaLeuAsnAsnGlnValGluGlnP 52

TTAACAAACTGCTTGAGCAGAGGGTACAAGGCCCAATCCAATCTGTGCGCCTGGCACGCG 2340
heAsnLysLeuLeuGluGlnArgValGlnGlyProIleGlnSerValArgLeuAlaArgA 72

CAGCTGGTTTCAGGGTCGACCCTTACTCATACACAAATCAAACTTTTATGACGATCAAT 2400
laAlaGlyPheArgValAspProTyrSerTyrThrAsnGlnAsnPheTyrAspAspGlnL 92

TAAATGCAATCAGACTATCATATAGAAATTTGTTCAAGAATTGATCACTTAACCCCTTGG 2460
euAsnAlaIleArgLeuSerTyrArgAsnLeuPheLysAsnTER 106

GTGCCGCACTTGGCCTAACCCAGGG (POLY (A) -43) 2486

FIG. 2—Continued.

clones pCV2 and pCV7 were subjected to DNA sequence analysis to predict the amino acid sequence of the polypeptide(s) encoded in this region of the FCV genome. The physical map and sequencing scheme are illustrated in Fig. 1. From this sequence analysis, it was determined that this region of the FCV genome contains two ORFs with the potential of being translated. The largest, consisting of 2,004 bases, encodes a polypeptide of 668 amino acids with a mass of 74,467 Da (Fig. 2). The second ORF was composed of 318 bases and encodes a polypeptide of 12,185 Da. The ORFs

overlapped at the termination codon of the large frame and the AUG initiation codon of the smaller frame (Fig. 2). The resulting overlap sequence was AUGA. Confirmation of the accuracy of the sequence of the overlap of ORFs came from the sequencing of two clones obtained from the PCR amplification of the sequences surrounding the overlap from the genomic RNA. There was no difference between the sequence of pCV7 and the PCR-derived clones. Following the UGA termination codon of the second, smaller ORF, there were 42 nontranslated bases before the poly(A) tail. Plasmid

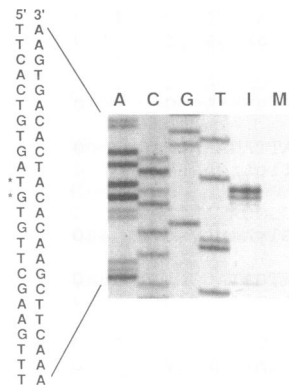


FIG. 3. Primer extension mapping of 5' end of FCV 2.4-kb subgenomic RNA transcripts. A synthetic 20-mer was used to prime reverse transcription primer extension reactions of RNA derived from FCV-infected cells (lane I) or from mock-infected control cells (lane M). The synthetic oligonucleotide was also used to prime a dideoxy sequencing region to serve as transcript endpoint markers. The A, C, G, and T reactions are marked accordingly. The sequence corresponding to the plus strand of the genomic RNA is illustrated on the left (5'), and the sequence of the minus strand (illustrated in the sequencing reaction) is that on the right (3'). Positions of the bands in the primer extension reaction in lane I are marked with asterisks.

pCV7, which contains the 3' end of the viral genome, had 43 adenine residues at the 3' end, giving this as the minimum length of the poly(A) tail (data not shown). Bases 1 through 118 in Fig. 2 are found at the end of a large ORF in the 5' end of the genomic RNA and encode the carboxy terminus of the FCV RNA-dependent RNA polymerase-like polypeptide (23). The nonstructural ORF and the ORF encoding the capsid protein are separated by two noncoding bases.

Mapping the 5' end of the 2.4-kb subgenomic RNA. The 5' end of the 2.4-kb subgenomic RNA was mapped by primer extension analysis (Fig. 3). A sequencing ladder primed with the same oligonucleotide used in the primer extension is included to provide endpoint determination. Two potential transcription start sites were identified, denoted by asterisks on the sequence. These two start points yielded 5' nontranslated leader regions of 17 and 18 bases. These start sites were located within an upstream AUG triplet which was out of frame from the large ORF. S1 protection analysis was done to confirm the results obtained with primer extension. A major band was observed, the 5' end of which mapped to the same location as that identified in the primer extension analysis (data not shown). Lane M contains the primer extension reaction of RNA derived from mock-infected control cells and contains no specific bands.

FCV capsid protein sequence analysis. One of the tryptic peptides was recovered by HPLC separation of the capsid protein digest eluted from the column at 85 min in 41%

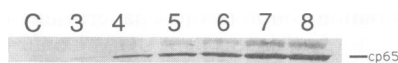


FIG. 4. Expression of the viral capsid protein during infection by FCV. Proteins from FCV-infected cells collected at hourly intervals postinfection were electrophoresed on a 12% sodium dodecyl sulfate-polyacrylamide gel, blotted to nitrocellulose, and probed with a feline anti-FCV antiserum. The capsid protein (cp65) is indicated at the right. The proteins visible above cp65 are viral proteins p73 and p75 (6).

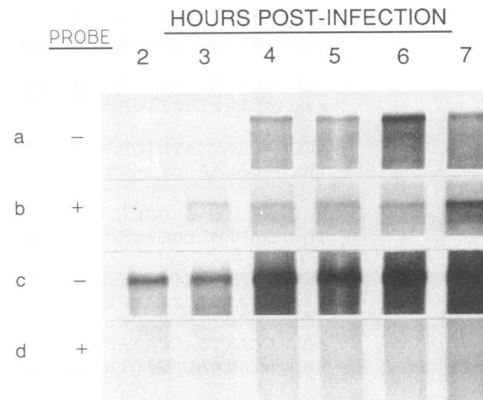


FIG. 5. Northern blot analysis of FCV RNAs using strand-specific probes. Total cellular RNA from FCV-infected cells was electrophoresed on formaldehyde-denaturing agarose gels, blotted to nitrocellulose, and probed with strand-specific, radiolabeled RNA probes. The probes used hybridize to both the genomic and 2.4-kb subgenomic RNAs (18). The + and - indicate which transcribed strand was used as the probe. (a and b) Genomic RNA probed with the minus-strand and plus-strand probes, respectively; (c and d) 2.4-kb subgenomic RNA probed with plus-strand and minus-strand probes, respectively.

acetonitrile. Automated Edman degradation revealed a single fragment with the sequence Leu-Gly-Ile-Gly-Val-Ala-Thr-Asp-Tyr-Ile-Val-Pro-Gly-Ile-Pro-Asp-Gly, corresponding to the region beginning with Leu-403 of the derived capsid protein sequence (Fig. 2). This was a major tryptic fragment, and the yield as determined from the sequence analysis was in the range of 200 pmol, or an overall yield of about 30%. Similar analysis of a fraction eluted at 70 min in 32% acetonitrile showed the sequence His-Phe-Asp-Phe-Asn-Gln-Glu, corresponding to the region beginning with His-373. Other fractions contained mixtures of peptides. For example, the peak eluted at 59 min in 27% acetonitrile gave data that could be interpreted as three sequences from the derived capsid protein sequence: (i) beginning with Gly-560, Gly-Gly-Asn-His-Pro-Ile-Phe; (ii) beginning with Ile-481, Ile-Ser-Asn-Thr-Ala-Phe-Ile-Thr-Thr; and (iii) beginning with Lys-480, Lys-Ile-Ser-Asn-Thr-Ala-Phe-Ile. The amino acid residues identified by Edman degradation are illustrated in bold type in Fig. 2.

Viral capsid protein synthesis. Synthesis of the viral capsid protein was analyzed by Western blot analysis (Fig. 4). cp65 was detected in lysates prepared at 3 h postinfection and continued to accumulate through 8 h postinfection, the last time point analyzed. The capsid protein, as determined by this analysis, was the most abundant viral protein observed. The other viral proteins also began to appear at 3 h postinfection (not shown) and were essentially those previously described by Carter (6).

Strand-specific Northern blot analysis. Strand-specific probes were used to examine the times at which the various RNAs appeared in FCV-infected cells and the relative amounts of each RNA species. Figure 5a shows the genomic RNA when probed with the minus-strand probe, and Fig. 5b shows the genomic RNA minus strand when probed with the plus-strand RNA probe. The genomic plus strand was detected at 2 h postinfection and continued to accumulate until 6 h postinfection, at which time it reached its greatest concentration. The amount of genomic plus strand decreased at 7 h postinfection, perhaps as a result of release of

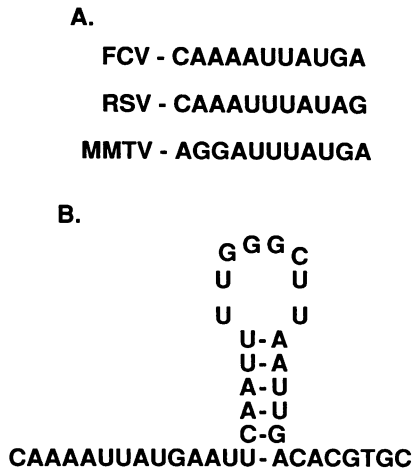


FIG. 6. Nucleotide sequence and secondary structure of the possible -1 frameshift sequences of FCV. (A) Comparison of the nucleotide sequences of the proposed ribosome-slippage sequences of FCV with those of RSV and MMTV. (B) Hairpin loop structure of FCV immediately following the proposed slippery sequences that may provide for a -1 frameshift to allow regulated translation of the second ORF at the 3' end of the 2.4-kb subgenomic mRNA. The AUGA overlap of the two reading frames of the 2.4-kb subgenomic mRNA is underlined.

progeny virus. The genomic minus strand was not detected until 3 h postinfection and appeared to be present at a slightly higher copy number than was the plus strand. After its appearance at 3 h postinfection, the minus strand continued to increase in copy number through 7 h postinfection. The concentration of the minus strand at 7 h postinfection was similar to that of the plus strand at 6 h postinfection.

Figures 5c and d show the 2.4-kb subgenomic RNA when probed with the minus-strand and the plus-strand RNA probes, respectively. The 2.4-kb plus-strand transcript was detected at 2 h postinfection, and maximum levels were observed at 4 h. The RNA was also beginning to show degradation at 7 h postinfection. However, the minus strand did not appreciably accumulate in any of the time intervals examined with the exception of a slight increase in concentration at 7 h postinfection. Both the plus- and minus-strand-probed blots were exposed on the same film, and the 2.4-kb message in Fig. 5c was overexposed to increase the signal of the minus-strand transcript in Fig. 5d. This illustrates the

difference in copy number between the 2.4-kb plus and minus strands.

DISCUSSION

Elucidation and characterization of the replication scheme of the caliciviruses has lagged behind that of other viruses. Until recently, only the number and sizes of the polypeptides and RNA species produced during FCV infection had been reported (2, 5, 7, 9, 10, 18, 25, 26, 30, 32). This work is the first to show that the capsid protein of FCV is encoded in the 3' end of the genomic RNA and thus the 2.4-kb subgenomic RNA. Sequence analysis of this region of the FCV genome revealed the presence of a large ORF that can encode a polypeptide of 73,467 Da (Fig. 1). This value is in agreement with the size of the 76,000-Da capsid precursor protein of FCV that was demonstrated by Carter (6). Confirmation that the large ORF within the 3'-terminal 2,500 bases of the FCV genome encoded the capsid protein of FCV was obtained through sequence analysis of purified capsid protein. It was determined from this analysis that the amino acid sequence derived from the nucleotide sequence of the FCV cDNA clones matches the amino acid sequence experimentally derived from purified capsid protein (Fig. 2). It is unknown at what point(s) the precursor protein is processed or how many processing steps are required to produce the mature capsid protein.

Primer extension mapping was used to map the 5' end of the 2.4-kb subgenomic RNA. This analysis revealed that the 2.4-kb subgenomic RNA contained the entire ORF encoding the FCV capsid protein. The length of the 5' leader region was 17 to 18 bases. It is not known whether both are actual start sites or one is an artifact of the primer extension reaction. The 2.4-kb subgenomic RNA is not capped (10), but it is not known whether there is a VPg covalently attached to the 5' end that may hinder the reverse transcriptase from reaching the end of the mRNA. DNA sequence analysis of this region of the FCV genomic RNA contained in pCV2 revealed that there was an upstream AUG triplet which was out of frame from the capsid ORF. The 5' ends of the transcripts began within this AUG triplet. This ensures that the AUG initiation codon of the capsid ORF will be the first encountered by the scanning 40S ribosomal subunit as proposed by Kozak (20, 21), which would allow for efficient translation of the capsid protein ORF. The AUG initiation codon of the large ORF was not a good fit with the Kozak consensus sequence, with a U at the +4 position. However, there was an A at -3, which is

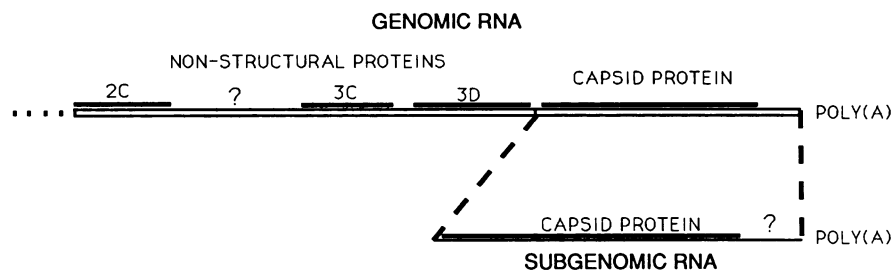


FIG. 7. Proposed genome structure of FCV. 2C, 3C, and 3D represent coding positions of 2C-, 3C cysteine protease-, and 3D RNA-dependent RNA polymerase-like polypeptides (19). Question marks represent areas of apparent coding sequences for which the function of the polypeptide product is unknown. The periods at the 5' end of the genomic RNA indicate the region of the genome for which there are currently no sequence data available.

probably more important than the nucleotide at +4 in determining initiator strength (22). Efficient translation has been observed in other viral systems with 9- and 10-base leader regions (8, 27). Consequently, the length of the 5' leaders probably provide ample space for ribosome binding and scanning to begin to allow efficient translation of the capsid reading frame.

Western blot analysis of FCV-infected cells using a feline FCV antiserum demonstrated that the capsid protein (65,000 Da in this strain of FCV) is detectable at 3 h postinfection. This is approximately 1 h after the 2.4-kb subgenomic RNA is first detectable (24). The largest apparent 1-h increase in capsid protein accumulation occurs between 3 and 4 h postinfection, the time at which the 2.4-kb mRNA is actively being synthesized and reaches steady-state levels (24). The capsid protein continues to accumulate to at least 8 h postinfection, the last time period analyzed in this study.

Northern blot studies of FCV-infected cells were done by using strand-specific probes to examine the appearance and accumulation of both the viral plus and minus strands and also to determine relative proportion of each. The genomic plus strand is detectable at 2 h postinfection whereas the minus strand appears at 3 h postinfection, after which they accumulate at similar rates. There was a decrease in the amount of genomic plus-strand RNA at 7 h postinfection, presumably as a result of the release of progeny virus from the cells. The genomic minus-strand RNA showed no such decrease. The minus strand appears to be present at a higher copy number at 3 h postinfection. This may represent a point at which the genomic minus strand is preferentially synthesized. A great difference was observed between the signals obtained from the 2.4-kb RNA probed with the minus strand and that probed with the plus strand. The 2.4-kb subgenomic RNA minus strand was present as only a very minor species compared with the plus strand, indicating that the minus strand probably plays no major role in the replication of FCV. It may be present as an artifact, created when the 2.4-kb subgenomic plus strand hybridizes to the genomic minus strand, with the ends being blunted by nuclease activity or perhaps by shearing forces during RNA isolation.

Sequence analysis of the FCV cDNA clones revealed the presence of a second, smaller ORF located 3' to the first ORF. This second ORF overlapped the first, with the termination codon of the large ORF making up part of the sequence of the initiation codon of the second ORF. Because the caliciviruses produce a subgenomic mRNA, it is possible that this ORF is contained within a small, undetected subgenomic mRNA. Northern blot analysis using the 3' 300 bases of the FCV genome as a probe failed to detect any small subgenomic RNA which might encode this small ORF. If this ORF is translated, it is possible that translation occurs through a frameshift mechanism that is similar to that described in the translation of retroviral polypeptides from overlapping ORFs. The nucleotide sequence of the region containing the ORF overlap in FCV shows some sequence similarity to the -1 frameshift sequences of Rous sarcoma virus (RSV) (14) and the *pro-pol* -1 frameshift sequences of murine mammary tumor virus (MMTV) (13, 15), as illustrated in Fig. 6A. The FCV sequences contain a region similar to the "slippery" sequences of RSV and MMTV and also a hairpin loop structure which may function to slow the translating ribosome at the slippery sequences (Fig. 6B). There is also the possibility of the formation of a pseudoknot structure (3) by the hybridization of nucleotides 2308 through 2318 of Fig. 2 to the nucleotides of the loop as illustrated in Fig. 6B. The formation of these secondary

structures would represent a means of controlling the expression of this ORF at the translational level.

The function of the polypeptide encoded in the second ORF is unknown. The -1 frameshift sequences of the retroviruses serve to regulate the translation of specific polypeptides encoded by these viruses which are required at low copy number. This would indicate that the polypeptide encoded by this second ORF may be required at lower copy number than is the capsid protein. Similar ORFs in several plant viruses (19) have been hypothesized to encode nucleic acid-binding proteins, based on the presence of conserved cysteine residues that may function in the formation zinc fingers. The second ORF of FCV encodes no cysteine residues but does contain a high proportion of basic and other charged amino acid residues. This feature indicates that this polypeptide may also function as a nucleic acid-binding domain and may be required attached to the capsid protein at low copy number, perhaps for the assembly of the virus particle.

Enough data are now available to propose a model for genomic organization of FCV. This model is illustrated in Fig. 7. With the evidence presented in this report, it is now apparent that the capsid protein is encoded within the 3'-terminal 2,400 bases of the FCV genomic RNA. This ORF is also transcribed into a subgenomic mRNA which probably represents the major template used to translate the capsid protein. An earlier report (23) demonstrated that a large ORF in the 5' end of the genomic RNA encoded apparent nonstructural polypeptides which are similar to picornavirus nonstructural proteins. These nonstructural polypeptides include the 2C, the 3C cysteine protease, and the 3D RNA-dependent RNA polymerase. The positions of the sequences encoding these nonstructural proteins are indicated in Fig. 7 as 2C, 3C, and 3D. The ORF encoding these nonstructural polypeptides in FCV is separated by two noncoding bases from the large ORF encoding the capsid protein (23) (Fig. 2). The question marks in Fig. 7 denote areas that apparently encode polypeptides, but the identity and function of these polypeptides remain unknown. The periods at the 5' end of the genomic RNA represent sequences for which there are no sequence data currently available. The identity and number of polypeptides encoded in this region and the nature of the 5' nontranslated sequences of the FCV genomic RNA remain to be determined.

ACKNOWLEDGMENTS

We thank Deborah Floan for technical assistance during the course of this study. Thanks are also given to R. Woods, B. Seal, and T. Casey for critical evaluation of the manuscript.

REFERENCES

1. **Bachrach, H. L., and W. R. Hess.** 1973. Animal picornaviruses with a single major species of capsid protein. *Biochem. Biophys. Res. Commun.* **55**:141-149.
2. **Black, D. N., J. N. Burroughs, T. J. R. Harris, and F. Brown.** 1978. The structure and replication of calicivirus RNA. *Nature (London)* **274**:614-615.
3. **Brierley, I., P. Digard, and S. C. Inglis.** 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **57**:537-547.
4. **Burroughs, J. N., and F. Brown.** 1978. Presence of a covalently linked protein on calicivirus RNA. *J. Gen. Virol.* **41**:443-446.
5. **Burroughs, J. N., T. R. Doel, C. J. Smale, and F. Brown.** 1978. A model for vesicular exanthema virus, the prototype of the calicivirus group. *J. Gen. Virol.* **40**:161-174.
6. **Carter, M. J.** 1990. Feline calicivirus protein synthesis investigated by western blotting. *Arch. Virol.* **108**:69-79.

7. Carter, M. J., E. G. Routledge, and G. L. Toms. 1989. Monoclonal antibodies to feline calicivirus. *J. Gen. Virol.* **70**:2197–2200.
8. Dasgupta, R., D. S. Shih, C. Saris, and P. Kaesberg. 1975. Nucleotide sequence of a viral RNA fragment that binds to eukaryotic ribosomes. *Nature (London)* **256**:624–628.
9. Ehresmann, D. W., and F. L. Schaffer. 1977. RNA synthesized in calicivirus-infected cells is atypical of picornaviruses. *J. Virol.* **22**:572–576.
10. Ehresmann, D. W., and F. L. Schaffer. 1979. Calicivirus intracellular RNA: fractionation of 18-22 S RNA and lack of typical 5'-methylated caps on 36 S and 22 S San Miguel sea lion virus RNAs. *Virology* **95**:251–255.
11. Fretz, M., and F. L. Schaffer. 1978. Calicivirus proteins in infected cells: evidence for a capsid polypeptide precursor. *Virology* **89**:318–321.
12. Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**:557–580.
13. Hizi, A., L. E. Henderson, T. D. Copeland, R. C. Sowder, C. V. Hixson, and S. Oroszlan. 1987. Characterization of mouse mammary tumor virus *gag-pro* gene products and the ribosomal frameshift site by protein sequencing. *Proc. Natl. Acad. Sci. USA* **84**:7041–7045.
14. Jacks, T., H. D. Madhani, F. R. Masiarz, and H. E. Varmus. 1988. Signals for ribosomal frameshifting in Rous sarcoma virus *gap-pol* region. *Cell* **55**:447–458.
15. Jacks, T., K. Townsley, H. E. Varmus, and J. Majors. 1987. Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus *gag*-related polyproteins. *Proc. Natl. Acad. Sci. USA* **84**:4298–4302.
16. Kawasaki, E. S. 1990. Amplification of RNA, p. 21–27. In M. A. Innes, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), PCR protocols. A guide to methods and applications. Academic Press, San Diego, Calif.
17. Kingston, R. E. 1989. Primer extension, p. 4.8.1–4.8.3. In F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology. Greene Publishing Associates and Wiley-Interscience, New York.
18. Komolaffe, O. O., O. Jarrett, and J. C. Neil. 1980. Feline calicivirus induced proteins. *Microbios* **27**:185–192.
19. Koonin, E. V., V. P. Boyko, and V. V. Dolja. 1991. Small cysteine-rich proteins of different groups of plant RNA viruses are related to different families of nucleic acid-binding proteins. *Virology* **181**:395–398.
20. Kozak, M. 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res.* **9**:5233–5250.
21. Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**:8125–8148.
22. Kozak, M. 1989. The scanning model for translation: an update. *J. Cell Biol.* **108**:229–241.
23. Neill, J. D. 1990. Nucleotide sequence of a region of the feline calicivirus genome that encodes picornavirus-like RNA-dependent RNA polymerase, cysteine protease and 2C polypeptides. *Virus Res.* **17**:145–160.
24. Neill, J. D., and W. L. Mengeling. 1988. Further characterization of the viral-specific RNAs in feline calicivirus infected cells. *Virus Res.* **11**:59–72.
25. Newman, J. F. E., D. J. Rowlands, and F. Brown. 1973. A physico-chemical sub-grouping of the mammalian picornaviruses. *J. Gen. Virol.* **18**:171–180.
26. Oglesby, A. S., F. L. Schaffer, and S. H. Madin. 1971. Biochemical and biophysical properties of vesicular exanthema of swine virus. *Virology* **44**:329–341.
27. Rose, J. K. 1980. Complete intergenic and flanking gene sequences from the genome of vesicular stomatitis virus. *Cell* **19**:415–421.
28. Schaffer, F. L. 1979. Caliciviruses, p. 249–284. In H. Fraenkel-Conrat and R. Wagner (ed.), Comparative virology, vol. 14. Plenum Press, New York.
29. Schaffer, F. L., D. W. Ehresmann, M. K. Fretz, and M. E. Soergel. 1980. A protein, VPg, covalently linked to 36S calicivirus RNA. *J. Gen. Virol.* **47**:215–220.
30. Schaffer, F. L., and M. E. Soergel. 1976. Single major polypeptide of a calicivirus: characterization by polyacrylamide gel electrophoresis and stabilization of virions by cross-linking with dimethyl suberimidate. *J. Virol.* **19**:925–931.
31. Shade, R. O., M. C. Blundell, S. F. Cotmore, P. Tattersall, and C. R. Astell. 1986. Nucleotide sequence and genome organization of human parvovirus B19 isolated from the serum of a child during aplastic crisis. *J. Virol.* **58**:921–936.
32. Soergel, M. E., T. G. Akers, F. L. Schaffer, and A. T. Noma. 1976. Amino acid composition of three immunological types of a calicivirus, San Miguel sea lion virus. *Virology* **70**:527–529.
33. Williams, S. A., B. E. Slatko, L. S. Moran, and S. M. DeSimone. 1986. Sequencing in the fast lane: a rapid protocol for [α - 35 S]dATP dideoxy DNA sequencing. *Biotechniques* **4**:138–147.
34. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**:103–119.