Identification and sequence determination of the capsid protein gene of feline calicivirus

M. J. Carter[1], **I. D. Milton**[1], **P. C. Turner**[2], **J. Meanger**[2], **M. Bennett**[3], and **R. M. Gaskell**[4]

[1] Division of Virology, School of Pathological Sciences, New Medical School, Newcastle upon Tyne
[2] Department of Biochemistry, University of Liverpool, Liverpool
[3] Department of Veterinary Clinical Science and [4] Veterinary Pathology, University of Liverpool, Leahurst, Wirral, U.K.

Summary. We have determined 4380 bases of the sequence from a cDNA clone containing the 3' end of feline calicivirus strain F9. We find four candidate open reading frames of which three are complete and comprise 245, 317 and 2012 nucleotides. The fourth continues toward the 5' end. We have expressed the largest complete open reading frame in *E. coli*. Sera raised to this antigen react specifically with the capsid protein and its intracellular precursor molecule. N-terminal sequence analysis of purified, mature capsid protein confirms this assignment and has identified the position at which precursor is cleaved.

Introduction

The caliciviruses are a family of small, non-enveloped viruses which contain a positive-stranded RNA genome 7.5–8.3 kb in size. The 34 nm capsid is unique in structure among the animal viruses [5], and is composed of a major capsid protein 60–70 kDa in size [18]. A second, smaller capsid component has been identified by some laboratories [1, 2]. The viruses replicate in the cytoplasm producing a nested set of 8, 3' co-terminal mRNAs [4, 15]. However there is as yet no clear indication of how these RNAs are produced from the genome, and only three open reading frames (ORFs) have been identified [4, 13].

Feline caliciviruses (FCV) comprise a single serotype although strains are readily distinguished by both polyclonal [11, 16] and monoclonal antibodies [23]. The vaccine strain F9 reacts most broadly but even this strain can elicit protection against only some 50% of field isolates [12]. This suggests that there is considerable variation in capsid proteins amongst FCVs and knowledge of

the way in which such variation is achieved is required for the development of the next generation of feline vaccines.

We recently reported the cloning of 4.5 kb from the 3′ end of FCV strain F9 [4]. In order to address both the problems described above, we have determined the complete sequence of clone pFC4.2 which comprises 4380 bases, and identified a total of 4 potential ORFs. One of these has been identified as the capsid gene. This sequence overlaps that already published for strain CFI/ 68 FIV (FCV) [13]. Comparative sequence analysis indicates that the capsid protein may have both conserved and variable regions.

Materials and methods

Cells and virus

Feline calicivirus strain F9 was obtained from Prof. O. Jarrett, Glasgow Veterinary School, and grown in feline kidney cells as previously described [3]. Infected cells were processed for polyacrylamide gel electrophoresis and Western blotting as previously described [3].

DNA cloning, manipulation and sequence determination

The purification of virus particles and the derivation, identification and restriction map of clone pFC4.2 have been reported elsewhere [4]. This clone was sequenced by subcloning restriction fragments into the phagemid vector pTZ18R. Nested set deletions were prepared by the method of Henikoff [10] using a reagent kit from Pharmacia Biotechnology according to the manufacturers instructions. Sequence determination was performed on rescued single-stranded templates, or double-stranded plasmids using Sequenase V2.0 obtained from United States Biochemicals. Sequences were determined in both senses and assembled using the Assemgel program (PC-Gene, Intelligenetics) and analysed with the same software package.

Expression of cDNA

Restriction fragments prepared from clone pFC4.2 were eluted from agarose gels by filter binding [8] and ligated into expression vectors of the pEX1–3 family [21]. These vectors contain a multiple cloning site situated downstream from a lambda promoter and b-galactosidase (b-gal) gene and allow the expression of inserted cDNA fragments as fusion proteins joined to the body of b-gal protein. Use of a host cell (popc2136) containing a ts lambda repressor protein allows cloning in a non-expressed form at 28 °C. Expression can be induced by transfer to the non-permissive temperature (42 °C).

Ligated mixtures were introduced to popc2136 as described [9] and transformants were selected on ampicillin plates at 28 °C. Expression was induced, and colonies were screened for reactivity with feline antisera [20] to FCV kindly provided by Prof. O. Jarrett, University of Glasgow. Plasmids were extracted from selected bacteria and checked for reconstitution of the Pst-1 site. The reading frame expressed was confirmed using a synthetic oligonucleotide (5′-CCATCGCCATCTGCTGC-3′). This binds at position 2912 within the b-gal coding sequence to prime DNA sequencing through the MCS and into the inserted cDNA. Fusion proteins were prepared for antigen as described [7] and adjusted to a final concentration of 1.3 mg/ml.

Immunization protocols

Four mice were immunized per recombinant antigen. Each was immunized on four occasions; the first consisted of 0.1 ml iv and 3 × 50 μl inoculations of antigen in FICA subcutaneously.

Subcutaneous immunizations were repeated at 14 and 30 days after starting the procedure. A final inoculation was performed at 48 days by injection of 25 µl in FICA into the footpad. Animals were bled from the tail vein before commencing this protocol and also 14 days after the third and fourth immunizations.

Results

Sequence determination

The sequence of clone pFC4.2 was analysed for potential coding regions (Fig. 1) by the method of Shepherd [19]. This method examines the distribution of purines and pyrimidines in each codon of all three potential frames. Coding regions show a bias in the position of these residues which can be detected by the program, absence of termination codons also has an effect. Predicted coding areas are displayed as a horizontal line superimposed on the position of termination codons in each frame. However the method is only a prediction, and sequence variation prevents perfect alignment between predicted coding areas (horizontal lines) and potential open reading frames (identified by absence of termination codons). For these reasons the horizontal line is not continuous and sudden changes are observed in the frame which is identified as most likely to be coding, e.g., at residue 850. These could be spurious, or alternatively suggest the existence of an overlapping gene. Such features do not imply that ribosomes would change frame at these points during translation. If the underlying ORF is continuous, it should be translated continuously. In the absence of any sequence reported from the 5′ end of the virus it is appropriate to refer to those potential ORFs identified in the sequence in order of their location from the 3′ end.
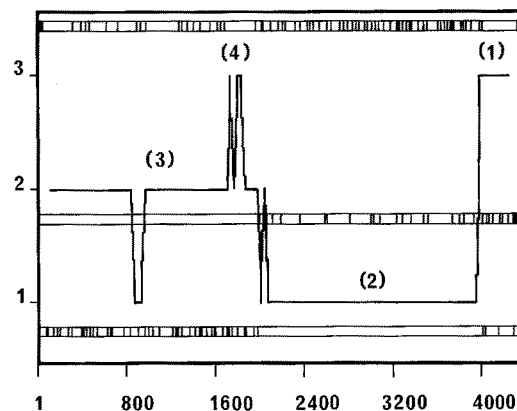
Fig. 1. Coding analysis of clone pFC42 performed by the method of Shepherd [19], using the PC-gene software package. Reading frame is indicated at the left hand side, vertical bars indicate the position of termination codons in each frame. Predicted coding regions are indicated by horizontal lines although statistical variations prevent these from being unbroken. ORFs discussed in the text are indicated in parentheses, numbered from the 3′ end: ORF 1, 4017–4334; ORF 2, 2005–4017; ORF 3, 2 (no initiation)–1999; ORF 4, 1722–1967

The small ORF described in our previous report is visible in frame 3 at the extreme 3' end of the virus (ORF 1), and the second is clearly visible extending away from this ORF in frame 1 (ORF 2). The two frames overlap by a single base. The first potential AUG initiation codon in ORF 2 occurs at position 2005 and termination should occur at position 4017. Thus ORF 2 could encode a protein of 671 residues and 73,441 in molecular weight. ORF 2 is separated by 4 bases from the 3' end of a third potential ORF in frame 2, which extends beyond the sequence presented here (ORF 3). This specifies aminoacids analogous to those determined by Neill [13] and which are thought to comprise non-structural polypeptides of the virus.

Towards the 3' terminus of ORF 3 is an area in which frames 2 and 3 appear equally likely to specify protein as determined by this method. In this region, a small potential ORF in frame 3 (245 n) overlaps the 3' end of ORF 3. We term this potential coding region, which has not been reported previously in FCV, ORF 4. This area could encode an 82 residue protein. All these potential ORFs are also present in the sequence of strain CF1/68 FIV [13, 14], and coding analysis by this method indicates that they could be functional in both viruses. The nucleotide sequence of clone pFC4.2 (strain F 9) is presented as Fig. 2, the theoretical translation products in each of the ORFs discussed above are indicated.

Expression and identification of the ORF 2 gene product

In order to identify the gene products from ORF 2, the 3' Pst-1 fragment, 2340–4380 was ligated into the expression vector pEX. Recombinant clones were identified by reaction with feline antisera to FCV. Sequence analysis confirmed that the ORF 2 construct is expressed from residue 112 in the predicted sequence ORF 2 could specify a protein of 73,441 in molecular weight. FCV infected cells contain at least three proteins of this size; two non-structural proteins, and the capsid precursor protein cpP76 [3].

We have derived monoclonal antibodies specific for the feline calicivirus capsid protein [6]. Both of these recognized the b-gal-ORF 2 fusion protein in Western blots, but had no reaction with other expressed ORFs or with the b-gal produced from a colony which contained FCV sequences inserted in the incorrect frame (data not shown). This suggested that ORF 2 could specify the capsid protein. This conclusion was confirmed by immunizing mice with the ORF 2 fusion protein or with the control b-gal antigens described above. The resulting sera were used in a Western blot. Proteins were prepared from mock-infected CRFK cells, and also from FCV-infected cultures. The infected cell


Fig. 2. Sequence determined from clone pFC42. This clone is 4380 bases long excluding synthetic homopolymer tails. Translation products in each frame are indicated, and the site of cleavage by Pst-1 is indicated

CATCCACATGGGCCACGGTGTTTACGCATCCGTAGCTCACGTGGTGAAAGGGGATTCATTTTTCTTGGGTGA 72

---I H M G H G V Y A S V A H V V K G D S F F L G E
ORF 3 (continuation)

AAGGATTTTTGATCTTAAGACTAATGGTGAATTTTGCTGCTTTCGCAGCACGAAAATTCTACCTAGTGCTGC 144

   R I F D L K T N G E F C C F R S T K I L P S A A

ACCTTTCTTTTCTGGGAAGCCCACTCGTGATCCGTGGGGATCCCCCGTGGCAACTGAGTGGAAGCCTAAAAT 216

   P F F S G K P T R D P W G S P V A T E W K P K M

GTACACAACAACCTCTGGAAAGATTCTGGGGTGCTTTGCAACAACTTCAACTGAAACTCACCCGGGAGACTG 288

   Y T T T S G K I L G C F A T T S T E T H P G D C

TGGCCTCCCATATATTGATGACAACGGGAGGGTGACCGGCCTCCACACTGGCTCTGGGGGACCCAAAACCCC 360

   G L P Y I D D N G R V T G L H T G S G G P K T P

AAGTGCCAAGTTGGTGGTGCCATATGTGCATATTGACATGAAGACTAAATCCGTCACTGCTCAAAAGTATGA 432

   S A K L V V P Y V H I D M K T K S V T A Q K Y D

CGTAACAAAGCCTGATATAAGTTACAAAGGCTTAATTTGTAAGCAATTGGATGAGATTAGGATTATACCAAA 504

   V T K P D I S Y K G L I C K Q L D E I R I I P K

AGGCACACGTCTCCATGTCTCCCCAGCCCACACTGAGGATTATCAAGAATGCTCACACCAACCCGCATCACT 576

   G T R L H V S P A H T E D Y Q E C S H Q P A S L

TGGAAGCGGGGATCCCCGCTGTCCAAAATCTCTCACTGCTATAGTTGTTGATTCTCTAAAACCATACTGTGA 648

   G S G D P R C P K S L T A I V V D S L K P Y C E

GAACGTTGAGGGTCCTCCACATGATGTTTTGCACAGAGTTCAAAAGATGCTTATCGACCACCTTTCAGGCTT 720

   N V E G P P H D V L H R V Q K M L I D H L S G F

TGTCCCTATGAACATTTCCTCGGAAACCTCTATGCTCTCAGCTTTCCACAAACTCAATCATGATACTTCCTG 792

   V P M N I S S E T S M L S A F H K L N H D T S C

TGGACCATACTTGGGTGGCAGAAAGAAAGATCACATGGCTAACGGTGAGCCGGACAAGCAGTTATTGGATCT 864

   G P Y L G G R K K D H M A N G E P D K Q L L D L

CCTGTCTGCAAAATGGAAATTGGCAACCCAAGGCATAGCACTACCACATGAGTACACAATTGGGCTAAAGGA 936

   L S A K W K L A T Q G I A L P H E Y T I G L K D

CGAGTTAAGGCCCGTGGAGAAGGTTAGTGAAGGGAAGAGAAGGATGATTTGGGGTTGTGATGTTGGCGTCGC 1008

   E L R P V E K V S E G K R R M I W G C D V G V A

TACTGTCTGTGCAGCTGCGTTCAAGGGTGTTAGCGATGCCATCACAGCAAACCACCAGTACGGGCCTATACA 1080

   T V C A A A F K G V S D A I T A N H Q Y G P I Q

GGTTGGTATCAACATGGATAGCCCCAGCGTCGAAGCGCTGTTCCAAAGGATCAAAAGCGCGGCCAAGGTATT 1152

   V G I N M D S P S V E A L F Q R I K S A A K V F

TGCGGTCGATTATTCCAAATGGGATTCGACCCAATCGCCTCGTGTCAGTGCAGCTTCAATTGACATCCTTCG 1224

   A V D Y S K W D S T Q S P R V S A A S I D I L R

TTACTTTTCCGATCGCTCTCCAATTGTTGACTCAGCCTCTAACACACTGAAGAGCCCTCCTGTTGCAATCTT 1296

   Y F S D R S P I V D S A S N T L K S P P V A I F

TAATGGTGTTGCTGTAAAAGTGTCCTCTGGCTTACCATCTGGAATGCCTCTTACCTCAGTAATCAATTCCCT 1368

   N G V A V K V S S G L P S G M P L T S V I N S L

```
TAATCATTGTCTGTATGTTGGGTGTGCCATTCTTCAATCCCTAGAAGCTAAGGCCATTCCCGTCACTTGGAA 1440
  N  H  C  L  Y  V  G  C  A  I  L  Q  S  L  E  A  K  A  I  P  V  T  W  N

CCTTTTCTCAACTTTTGATATCATGACTTACGGGGATGATGGTGTCTACATGTTTCCTATTATGTATGCAAG 1512
  L  F  S  T  F  D  I  M  T  Y  G  D  D  G  V  Y  M  F  P  I  M  Y  A  S

TATTAGTGACCAAATTTTTGGAAATCTTTCTTCCTATGGCCTGAAACCAACTCGGGTTGACAAGTCCGTTGG 1584
  I  S  D  Q  I  F  G  N  L  S  S  Y  G  L  K  P  T  R  V  D  K  S  V  G

AGCAATTGAGCCTATTGATCCTGACTCTGTTGTTTTCTTGAAGAGAACAATCACAAGGACACCTCAGGGGAT 1656
  A  I  E  P  I  D  P  D  S  V  V  F  L  K  R  T  I  T  R  T  P  Q  G  I

AAGGGGTTTACTTGATCGCAGCTCTATAATAAGACAATTCTATTATATTAAAGGTGAGAACTCCGATGACTG 1728
  R  G  L  L  D  R  S  S  I  I  R  Q  F  Y  Y  I  K  G  E  N  S  D  D  W
                                                     ORF4 START  M  T

GAAGAGCCCCCCAAAACATATTGACCCAACATCTCGAGGGCAACAGCTTTGGAATGCCTGTCTGTACGCTAG 1800
  K  S  P  P  K  H  I  D  P  T  S  R  G  Q  Q  L  W  N  A  C  L  Y  A  S
  G  R  A  P  Q  N  I  L  T  Q  H  L  E  G  N  S  F  G  M  P  V  C  T  L

CCAACATGGCTTGGAGTTTTTCAACAAGGTTTACAGGCTGGCCGAGAGGGCTGTTGAATATGAAGAGCTGCA 1872
  Q  H  G  L  E  F  F  N  K  V  Y  R  L  A  E  R  A  V  E  Y  E  E  L  H
  A  N  M  A  W  S  F  S  T  R  F  T  G  W  P  R  G  L  L  N  M  K  S  C

CTTTGAACCCCCAACATATGCTTCGGCTTTGGATCATTACAACAGCCAGTTCAATGGCGTGGAGGCGCGGTC 1944
  F  E  P  P  T  Y  A  S  A  L  D  H  Y  N  S  Q  F  N  G  V  E  A  R  S
  T  L  N  P  Q  H  M  L  R  L  W  I  I  T  T  A  S  S  M  A  W  R  R  G

TGACCAGATCGACTCGAGTGGCATGACCGCCCTACACTGTGATGTGTTCGAAGTTTGAGCATGTGCTCAACC 2016
                                           ORF2 START  M  C  S  T
  D  Q  I  D  S  S  G  M  T  A  L  H  C  D  V  F  E  V  -  ORF3 END
  L  T  R  S  T  R  V  A  -  ORF4 END

TGCGCTAACGTGCTTAAATATTATGATTGGGACCCCCATTTCAAATTGGTAATCAACCCCAACAACTTCCTC 2088
  C  A  N  V  L  K  Y  Y  D  W  D  P  H  F  K  L  V  I  N  P  N  N  F  L

TCTGTTGGCTTTTGTAGTAACCCTTTAATGTGTTGCTACCCAGAACTCCTTCCGGAATTTGGAACTGTTTGG 2160
  S  V  G  F  C  S  N  P  L  M  C  C  Y  P  E  L  L  P  E  F  G  T  V  W

GATTGCGATCGGTCACCACTTGAAATTTACCTAGAATCAATACTTGGTGATGATGAATGGGCATCCACTTTT 2232
  D  C  D  R  S  P  L  E  I  Y  L  E  S  I  L  G  D  D  E  W  A  S  T  F

GACGCTGTTGACCCAGTCGTTCCCCCAATGCACTGGGGTGCTGCTGGAAAAATTTTCCAGCCACACCCCGGT 2304
  D  A  V  D  P  V  V  P  P  M  H  W  G  A  A  G  K  I  F  Q  P  H  P  G

                    Pst-1
GTTCTCATGCACCATCTCATTGGTAAGGTTGCTGCAGGTTGGGACCCCGATCTGCCTCTAATTCGACTCGAG 2376
  V  L  M  H  H  L  I  G  K  V  A  A  G  W  D  P  D  L  P  L  I  R  L  E

GCGGATGACGGGTCAATCACAGCACCCGAGCAAGGAACAATGGTTGGCGGCGTCATCGCTGAACCCAGCGCC 2448
  A  D  D  G  S  I  T  A  P  E  Q  G  T  M  V  G  G  V  I  A  E  P  S  A
  Terminus of mature capsid

CAGATGTCAACAGCTGCTGATATGGCCACCGGGAAAAGCGTTGATTCTGAGTGGGAGGCATTCTTCTCCTTT 2520
  Q  M  S  T  A  A  D  M  A  T  G  K  S  V  D  S  E  W  E  A  F  F  S  F

CACACCAGCGTCAATTGGAGTACATCTGAAACCCAAGGAAAGATTCTCTTCAAACAATCCTTAGGCCCTTTG 2592
  H  T  S  V  N  W  S  T  S  E  T  Q  G  K  I  L  F  K  Q  S  L  G  P  L

CTCAACCCATATCTAGAACACCTTGCTAAGCTATATGTTGCGTGGTCTGGGTCGATTGAGGTTAGGTTCTCT 2664
  L  N  P  Y  L  E  H  L  A  K  L  Y  V  A  W  S  G  S  I  E  V  R  F  S

ATCTCTGGCTCTGGTGTCTTTGGTGGGAAGCTCGCAGCTATTGTTGTACCTCCTGGGGTTGATCCAGTGCAG 2736
  I  S  G  S  G  V  F  G  G  K  L  A  A  I  V  V  P  P  G  V  D  P  V  Q

AGTACTTCGATGCTACAATACCCCCATGTTTTGTTTGATGCTCGTCAGGTGGAACCAGTTATCTTCTGTCTT 2808
  S  T  S  M  L  Q  Y  P  H  V  L  F  D  A  R  Q  V  E  P  V  I  F  C  L

CCTGATCTAAGAAGCACCCTGTACCACCTTATGTCTGACACTGACACTACATCCTTGGTCATTATGGTGTAC 2880
  P  D  L  R  S  T  L  Y  H  L  M  S  D  T  D  T  T  S  L  V  I  M  V  Y

AATGATCTCATCAATCCCTATGCCAATGATGCCAACTCTTCTGGGTGTATTGTCACTGTCGAGACAAAACCT 2952
  N  D  L  I  N  P  Y  A  N  D  A  N  S  S  G  C  I  V  T  V  E  T  K  P

GGCCCTGACTTCAAGTTTCACCTCCTTAAGCCACCCGGATCTATGCTAACCCATGGCTCTATCCCTTCTGAT 3024
  G  P  D  F  K  F  H  L  L  K  P  P  G  S  M  L  T  H  G  S  I  P  S  D
```

```
TTAATTCCCAAAACATCTTCGCTCTGGATCGGTAACCGCTACTGGTCAGACATAACTGATTTTGTGATTCGG 3096
L  I  P  K  T  S  S  L  W  I  G  N  R  Y  W  S  D  I  T  D  F  V  I  R

CCGTTTGTCTTCCAAGCAAATCGTCATTTTGACTTTAATCAAGAGACCGCAGGGTGGAGCACACCACGGTTT 3168
P  F  V  F  Q  A  N  R  H  F  D  F  N  Q  E  T  A  G  W  S  T  P  R  F

CGGCCTATATCTGTTACCATTACTGAACAGAACGGAGCAAAATTGGGCATTGGGGTGGCAACAGATTACATA 3240
R  P  I  S  V  T  I  T  E  Q  N  G  A  K  L  G  I  G  V  A  T  D  Y  I

GTGCCTGGAATCCCTGATGGCTGGCCTGACACCACAATTCCTGGGGAGTTGATACCAGCTGGTGATTACGCA 3312
V  P  G  I  P  D  G  W  P  D  T  T  I  P  G  E  L  I  P  A  G  D  Y  A

ATCACCAATGGTACTGGCAATGACATCACCACGGCTACAGGATATGACACTGCTGATATAATTAAGAACAAT 3384
I  T  N  G  T  G  N  D  I  T  T  A  T  G  Y  D  T  A  D  I  I  K  N  N

ACCAACTTTAGGGGCATGTACATATGTGGTTCGCTCCAGCGTGCCTGGGGTGATAAGAAAATTTCCAACACT 3456
T  N  F  R  G  M  Y  I  C  G  S  L  Q  R  A  W  G  D  K  K  I  S  N  T

GCCTTTATCACCACTGCCACCCTAGATGGTGACAACAACAACAAGATCAATCCCTGTAATACCATAGACCAG 3528
A  F  I  T  T  A  T  L  D  G  D  N  N  N  K  I  N  P  C  N  T  I  D  Q

TCAAAGATCGTCGTGTTTCAAGACAACCATGTTGGAAAGAAAGCGCAAACCTCAGACGATACATTGGCCCTG 3600
S  K  I  V  V  F  Q  D  N  H  V  G  K  K  A  Q  T  S  D  D  T  L  A  L

CTTGGTTACACTGGCATTGGTGAGCAGGCCATCGGGTCTGATAGGGACCGGGTTGTGCGCATCAGCACTCTC 3672
L  G  Y  T  G  I  G  E  Q  A  I  G  S  D  R  D  R  V  V  R  I  S  T  L

CCTGAAACTGGTGCTCGAGGCGGTAACCACCCAATTTTCTACAAGAACTCCATTAAATTGGGATATGTAATT 3744
P  E  T  G  A  R  G  G  N  H  P  I  F  Y  K  N  S  I  K  L  G  Y  V  I

AGGTCTATTGATGTCTTTAATTCACAAATCTTGCACACTTCCAGACAGTTATCGCTAAATCATTACCTACTC 3816
R  S  I  D  V  F  N  S  Q  I  L  H  T  S  R  Q  L  S  L  N  H  Y  L  L

CCACCTGATTCTTTTGCCGTCTATAGAATAATTGACTCAAATGGCTCGTGGTTTGATATTGGAATTGATAGT 3888
P  P  D  S  F  A  V  Y  R  I  I  D  S  N  G  S  W  F  D  I  G  I  D  S

GATGGGTTCTCTTTTGTTGGTGTTTCTGGCTTTGGTAAATTAGAATTTCCCCTTTCTGCCTCCTACATGGGA 3960
D  G  F  S  F  V  G  V  S  G  F  G  K  L  E  F  P  L  S  A  S  Y  M  G

ATACAATTGGCAAAGATCCGGCTTGCCTCTAACATTAGGAGTCCCATGACTAAGTTATGAATTCAATATTAG 4032
I  Q  L  A  K  I  R  L  A  S  N  I  R  S  P  M  T  K  L  -  ORF2 END

                                    ORF1 START  M  N  S  I  L

GCCTGATTGATACTGTTACTAACACTATTGGTAAAGCTCAGCAGATTGAATTGGACAAAGCTGCACTTGGTC 4104

G  L  I  D  T  V  T  N  T  I  G  K  A  Q  Q  I  E  L  D  K  A  A  L  G

AACAGCGTGAATTGGCTCTCCAGCGCATTGGCTTGGACCGCCAAGCCTTAAACAACCAAGTTGAGCAGTTTA 4176

Q  Q  R  E  L  A  L  Q  R  I  G  L  D  R  Q  A  L  N  N  Q  V  E  Q  F

ACAAAATTCTTGAGCAAAGGGTACAAGGCCCAATCCAATCCGTTCGCCTTGCGCGCGCGGCTGGATTTAGGG 4248

N  K  I  L  E  Q  R  V  Q  G  P  I  Q  S  V  R  L  A  R  A  A  G  F  R

TTGACCCTTACTCATACACAAACCAAAACTTTTATGACGATCAATTAAATGCAATTAGACTATCATATAAAA 4320

V  D  P  Y  S  Y  T  N  Q  N  F  Y  D  D  Q  L  N  A  I  R  L  S  Y  K

ATTTGTTTAAAATTTGATCATATATCCCTTTGGGCTGCCGCACCTGCGCCTAACCCCAGGG(poly A) 4380

N  L  F  K  I  -  ORF1 END
```
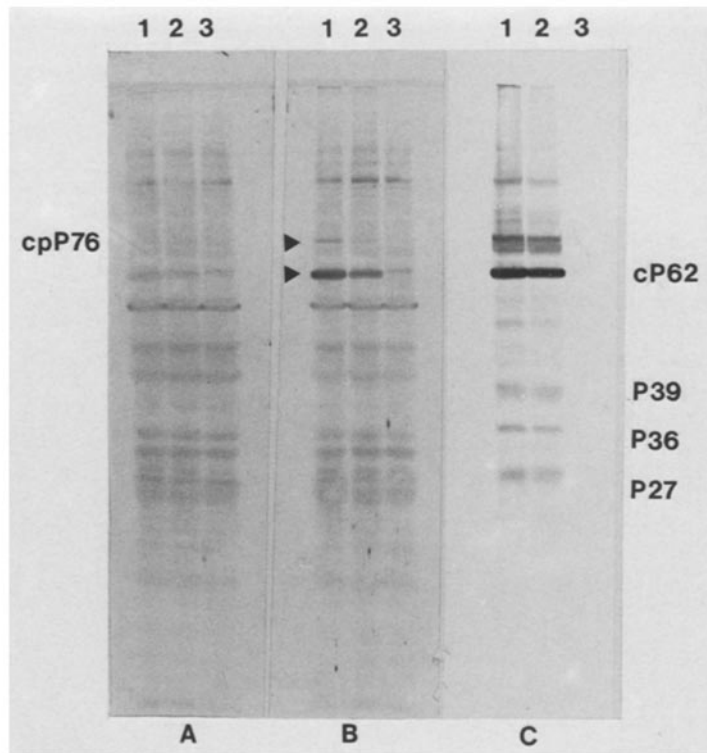
Fig. 3. Reactivity of murine sera. FCV-infected or mock-infected CRFK cell proteins were prepared for electrophoresis and analysed in triplicate by Western blotting. *1* FCV-infected CRFK cells harvested at 4h pi (37 °C 1–3 h, 42 °C 3–4 h), *2* FCV-infected CRFK cells harvested at 4 h pi (37 °C 1–4 h); *3* Mock-infected CRFK cells. Panels were developed with the following sera: **A** Murine sera from animals immunized with b-gal fused to ORF-2 in the incorrect frame (negative control); **B** Murine sera from animals immunized with b-gal fusion protein expressing ORF 2 prepared as described in the text; **C** Feline serum from a cat immunized with FCV. **A** and **B** Stained with colloidal gold-conjugated anti-murine immunoglobulin and staining was enhanced by silver precipitation (Cambio, Cambridge). **C** Developed using goat anti-feline immunoglobulins which were then detected with per-oxidase-conjugated anti-goat immunoglobulins. The capsid protein cP62, and its precursor cpP76 are marked (◄). Lower molecular weight virus-specific proteins, P39, P36 and P27 are also indicated for reference

antigens were of two types; cultures maintained throughout infection at 37 °C, and also cells infected at 37 °C, and subsequently transferred to 42 °C at 3 h pi. This procedure allows the accumulation of normally short-lived precursor molecules [3]. Replicate blots of proteins from all three sources were then immunostained with feline serum, ORF 2-specific or control murine sera described above. Bound murine antibody was detected using a colloidal-gold conjugated anti-murine immunoglobulin in combination with silver precipitation enhancement. Bound feline antibody was detected using an antibody sandwich technique, culminating in a peroxidase-conjugated detector antibody. This procedure resulted in greater background staining of panels A and B, compared to panel C (Fig. 3). The resulting blots show that the ORF 2-immunized mice

developed antibodies to the capsid protein cP62. The specificity of this reaction was confirmed since this serum also stained the capsid precursor protein cpP76, observed in the FCV-infected cultures maintained at elevated temperature. This effect is not visible in Fig. 3 C because the capsid protein precursor can be masked by other virus proteins migrating in this area of the gel [3].

Identification of the N-terminus of the mature capsid protein

The mature capsid protein is cut from a larger precursor and some 11 kDa are removed. This cleavage could take place at the N terminus, C terminus or both. We therefore purified the virus by sedimentation [4], and transferred virus cP62 to Pall Pro-blot membranes by Western blot. This was sent for polypeptide sequencing at the University of Leicester, Dept. of Biochemistry. The sequence determined was Ala Asp Asp Glu Ser Ile Thr which was found at position 125 in the sequence predicted from ORF 2 (Fig. 2). The fusion protein formed to express this ORF and which should contain residues 112–674, thus contained all the information present in the mature virus capsid protein.

Comparative sequence analysis

During the preparation of this report, the sequence of the 3′ end of FCV strain CFI/68 FIV (FCV) has been lodged in the EMBL database [14]. This sequence contains the same pattern of four potential ORF regions as that described here. The extent of nucleotide identity between the structural (ORF 2), and probable non-structural (ORF 3) genes of both strains is very similar; 80% and 78.8% respectively. Predicted protein sequences are even more closely related, 95.3% identity for ORF 1, 89.7% for ORF 2, and 91.4% in the region of overlap between ORF 3 from both strains. Modification sites and functional motifs identified by Neill [13], are common to both strains, and we find that both capsid genes encode a potential ATP/GTP binding site [25] (Fig. 4).

Antigenic variation between virus strains lies predominantly in the structural proteins of the virus. The capsid precursor proteins from both strains are compared in Fig. 4. Variation between the proteins is concentrated in two areas. The first of these is the N terminus preceding the cleavage site in the precursor molecule, positions 1–121 (13% mismatch). The cleavage site itself, and the adjacent region of the protein are relatively conserved and only 3.6% of the amino acids are different between residues 122 and 395. Residues 396–525 again show increased divergence with 23% of amino acid variation and include the insertion of three asparagine residues in the sequence derived from strain F 9, which are not present in strain CFI/68 FIV (FCV). The final section of the protein is once more relatively well conserved and has only 7.5% variation. Since the N-terminal region of variation is lost during maturational cleavage, antigenic variation between virus particles may be predominantly located in the central region described above.

```
F9      -  MCSTCANVLKYYDWDPHFKLVINPNNFLSVGFCSNPLMCCYPELLPEFGT  -50
                   *            *   *       *
CFI/68  -  MCSTCANVLKYYDWDPHIKLVINPNKFLHVGFCDNPLMCCYPELLPEFGT  -50

F9      -  VWDCDRSPLEIYLESILGDDEWASTFDAVDPVVPPMHWGAAGKIFQPHPG  -100
              *    *   **           *   ** *         **
CFI/68  -  MWDCDQSPLQVYLESILGDDEWSSTHEAIDPVVPPMHWDEAGKIFQPHPG  -100
           Terminus of mature capsid
F9      -  VLMHHLIGKVAAGWDPDLPLIRLEADDGSITAPEQGTMVGGVIAEPSAQM  -150
               *   *    *    *           *               *
CFI/68  -  VLMHHLICKVAEGWDPNLPLFRLEADDGSITTPEQGTMVGGVIAEPNAQM  -150
           ATP/GTP Binding
F9      -  STAADMATGKSVDSEWEAFFSFHTSVNWSTSETQGKILFKQSLGPLLNPY  -200

CFI/68  -  STAADMATGKSVDSEWEAFFSFHTSVNWSTSETQGKILFKQSLGPLLNPY  -200

F9      -  LEHLAKLYVAWSGSIEVRFSISGSGVFGGKLAAIVVPPGVDPVQSTSMLQ  -250
              *          **                      *
CFI/68  -  LTHLAKLYVAWSGSVDVRFSISGSGVFGGKLAAIVVPPGIDPVQSTSMLQ  -250

F9      -  YPHVLFDARQVEPVIFCLPDLRSTLYHLMSDTDTTSLVIMVYNDLINPYA  -300
                          **
CFI/68  -  YPHVLFDARQVEPVIFSIPDLRSTLYHLMSDTDTTSLVIMVYNDLINPYA  -300

F9      -  NDANSSGCIVTVETKPGPDFKFHLLKPPGSMLTHGSIPSDLIPKTSSLWI  -350
              *                                     *
CFI/68  -  NDSNSSGCIVTVETKPGPDFKFHLLKPPGSMLTHGSIPSDLIPKSSSLWI  -350

F9      -  GNRYWSDITDFVIRPFVFQANRHFDFNQETAGWSTPRFRPISVTITEQNG  -400
              *                          ** *****
CFI/68  -  GNRFWSDITDFVIRPFVFQANRHFDFNQETAGWSTPRFRPITITISVKES  -400

F9      -  AKLGIGVATDYIVPGIPDGWPDTTIPGELIPAGDYAITNGTGNDITTATG  -450
                                 * *       *     **
CFI/68  -  AKLGIGVATDYIVPGIPDGWPDTTIPGELVPVGDYAITNGTNNDITTAAQ  -450

F9      -  YDTADIIKNNTNFRGMYICGSLQRAWGDKKISNTAFITTATLDGDNNNKI  -500
              * **  *                           * *  ****  *
CFI/68  -  YDAATEIRNNTNFRGMYICGSLQRAWGDKKISNTAFITTGTVDGA---KL  -497

F9      -  NPCNTIDQSKIVVFQDNHVGKKAQTSDDTLALLGYTGIGEQAIGSDRDRV  -550
              * *     *   *      * ** **                *   *
CFI/68  -  IPSNTIDQTKIAVFQDTHANKHVQTSDDTLALLGYTGIGEEAIGADRDRV  -547

F9      -  VRISTLPETGARGGNHPIFYKNSIKLGYVIRSIDVFNSQILHTSRQLSLN  -600
               *     *             *
CFI/68  -  VRISVLPERGARGGNHPIFHKNSIKLGYVIRSIDVFNSQILHTSRQLSLN  -597

F9      -  HYLLPPDSFAVYRIIDSNGSWFDIGIDSDGFSFVGVSGFGKLEFPLSASY  -650
                            *                 **        *
CFI/68  -  HYLLSPDSFAVYRIIDSNGSWFDIGIDNDGFSFVGVSSIGKLEFPLTASY  -647

F9      -  MGIQLAKIRLASNIRSPMTKL  -671
                            *
CFI/68  -  MGIQLAKIRLASNIRSVMTKL  -668
```

Fig. 4. Comparison of FCV capsid protein precursors. The proteins specified by ORF2 in strains F9 (671aa) and CFI/68 FIV (FCV) (688aa), were aligned using the PC-gene software Palign program. * Residues which differ between the two strains. The N-terminal of the mature capsid of the F9 strain, determined by sequencing; and a potential ATP/GTP binding site, identified by computer analysis, are indicated

Discussion

We have presented the sequence of the 3′ end of feline calicivirus strain F9. The distribution of potential coding regions is very similar to that observed in a second feline calicivirus whose sequence has been lodged during the preparation of this report. We have identified the major reading frame in this area as specifying the virus capsid protein. This specifies a protein of 73,441 in molecular weight which is in good agreement with the size of 76,000 determined for the capsid protein precursor [3]. Cleavage of the capsid precursor protein

at residue 125 would remove 13,961 and is thus sufficient to account for the size decrease observed during capsid protein maturation. Any trimming at the C terminus cannot therefore be extensive. The predicted sequence contains a consensus sequence for nucleotide binding [25] which is conserved in the two strains sequenced. This is usually associated with polymerase proteins but could be used in this case for RNA binding. The protein does not have extensive similarity with any non-calicivirus sequences in the database (Swiss-Prot release 13). Comparative sequence analysis suggests that the mature capsid protein has relatively conserved termini, but is more variable in the centre. This could suggest that the termini form the structural basis of the particle, whilst the centre is looped out to the surface of the virus, exposed to the immune system and subject to antigenic variation. However, sequence determination from more strains, and location of antibody binding sites will be required to confirm this point.

An mRNA synthesized to express ORF 2, and which extends to the 3' terminus would be at least 2,375 bases long. The capsid protein is the most abundant virus-specific polypeptide in the infected cell and an mRNA of 2.4–2.7 kb has been identified as the most abundant message synthesized by this virus [4, 15]. It seems likely therefore that this corresponds to the transcript of the capsid gene identified here and would also allow the presence of a short leader RNA similar to that seen in corona- and toro-virus infections [22, 24]. However the mechanism of calicivirus RNA transcription has yet to be elucidated and it is not known whether such a leader exists. Furthermore, whilst an mRNA of an appropriate size (550 n) for the expression of ORF 1 has been detected; there are no obvious coding frames of a suitable size in the sequence presented here, to account for the mRNAs of 1.5 and 1.9 kb which have also been observed [4]. mRNAs larger than the presumptive capsid protein mRNA (2.4 kb) are known to be present, one of these could direct the translation of ORF 4. Since there are thought to be at least three of these, intermediate in size between the capsid mRNA and the genome, other ORFs may exist between the end of the clone sequenced here and the 5' terminus of the virus.

The status of the potential ORFs numbered 1 and 4 in this analysis is less well established. The protein predicted from ORF 1 is the most well conserved sequence between the two strains compared here. This could indicate a vital role for this protein in replication. However the gene product from this region has not yet been identified.

In contrast, the predicted product from ORF 4 is truncated from 82 residues in F9, to 73 residues in strain CFI/68 FIV (FCV). Initiation in the latter strain would occur at the second methionine residue in the sequence predicted from FCV strain F9. The proteins predicted would then have 41% identity in the area of overlap which is the highest degree of variation between any of the proteins predicted from both strains.

A small ORF in this position, located at the junction of the two large genes, has been recently identified in human hepatitis E virus. This agent is a candidate calicivirus and shows generally similar gene organization in that a structural

protein gene is located at the 3' end of the genome and a large putative non-structural gene at the 5' end. The small ORF spans the two major reading frames and specifies a product which is detectable as a structural epitope on the particles [17]. It is possible therefore that the potential ORF 4 in FCV could represent a similar small structural protein for this virus. The relatively low conservation in amino acid sequence could then provide an additional mechanism for antigenic variation between viruses.

Acknowledgement

References

1. Bachrach HL, Hess WR (1973) Animal picornaviruses with a single major species of capsid protein. Biochem Biophys Res Commun 55: 141–149
2. Burroughs JN, Doel TR, Smale CJ, Brown F (1978) A model for vesicular exanthema virus, the prototype of the calicivirus group. J Gen Virol 40: 161–174
3. Carter MJ (1989) Feline calicivirus protein synthesis investigated by Western blotting. Arch Virol 108: 69–79
4. Carter MJ (1990) Transcription of feline calicivirus RNA. Arch Virol 114: 143–152
5. Carter MJ, Madeley CR (1987) Caliciviruses. In: Nermut MV, Steven AG (eds) Animal virus structure. Elsevier, Amsterdam, pp 121–128
6. Carter MJ, Routledge EG, Toms GL (1989) Monoclonal antibodies to feline calicivirus. J Gen Virol 70: 2197–2200
7. Doorbar J. Cambell D, Grand RJA, Gallimore PH (1986) Identification of the human papillomavirus-1a E4 gene products. EMBO J 5: 355–362
8. Dretzen G, Bellard M, Sassone-Corsi P, Chambon P (1981) A reliable method for the recovery of DNA fragments from agarose and acylamide gels. Anal Biochem 112: 295–298
9. Hanahan D (1985) Techniques for transformation of E. coli. In: Glover DM (ed) DNA cloning, vol 1. IRL Press, Eynsham, pp 109–135
10. Henikoff S (1984) Unidirectional deletion with exonuclease III creates target breakpoints for DNA sequencing. Gene 28, 351–359
11. Kalunda M, Lee KM, Holmes DF, Gillespie JH (1975) Serologic classification of feline caliciviruses by plaque reduction, neutralization and immunodiffusion. Am J Vet Sci 30: 197–206
12. Knowles JO, Dawson S, Gaskell RM, Gaskell CJ, Harvey CE (1990) Neutralisation patterns among recent British and North American feline calicivirus isolates from different clinical origins. Vet Rec 127: 125–127
13. Neill JD (1990) Nucleotide sequence of a region of the feline calicivirus genome which encodes picornavirus-like RNA-dependent RNA polymerase, cysteine protease and 2C polypeptides. Virus Res 17: 145–160
14. Neill JD, EMBL Database sequence number M 32819.
15. Neill JD, Mengeling WL (1988) Further characterisation of the virus-specific RNAs in feline calicivirus infected cells. Virus Res 11: 59–72
16. Povey RC (1974) Serological relationships among feline caliciviruses. Infect Immun 10: 1307–1314
17. Reyes GR, Tam AW, Kim JP, Purdy MA, Yarbough PO, Fry KE, Bradley DW (1990) Hepatitis E virus: molecular characterization of the novel agent responsible for enter-

ically transmitted non-A, non-B hepatitis (ET-NANB). In: Abstracts, VIIIth International Congress of Virology, Berlin 1990 (abstract no. W 11-007), p 38

18. Schaffer FL, Bachrach HL, Brown F, Gillespie JH, Burroughs JN, Madin SH, Madeley CR, Povey RC, Scott F, Smith AW, Studdert MJ, (1980) Caliciviridae. Intervirology 14: 1–6

19. Shepherd JCW (1981) Method to determine the reading frame of a protein from the purine/pyrimidine genome sequence and its possible evolutionary justification. Proc Natl Acad Sci USA 78: 1596–1600

20. Stanley KK (1983) Solubilization and immune detection of Beta-galactosidase hybrid proteins carrying foreign antigenic determinants. Nucleic Acids Res 11: 4077–4092

21. Stanley KK, Luzio JP (1984) Construction of a new family of high efficiency bacterial expression vectors: identification of cDNA clones coding for human liver proteins. EMBO J 3: 1429–1434

22. Spaan WJM, Delius H, Skinner M, Armstrong J, Rottier P, Smeekens S, van der Zeist BAM, Siddel SG (1983) Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. EMBO J 2: 1839–1983

23. Tohya Y, Masouka K, Takahashi E, Takeshi M (1991) Neutralizing epitopes of feline calicivirus. Arch Virol 117: 173–181.

24. deVries AAF, Chirnside ED, Bredenbeek PJ, Gravenstein LA, Horzinek MC, Spaan WJM (1990) All subgenomic mRNAs of equine arteritis virus contain a common leader sequence. Nucleic Acids Res 18: 3241–3247

25. Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the alpha and beta subunits of ATP synthetase, myosin kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J 1: 945–951

Authors' address: Dr. M. J. Carter, Division of Virology, School of Pathological Sciences, New Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH, U.K.