Journal of General Virology (1991), **72**, 1659–1666. *Printed in Great Britain*

Sequence analysis of the turkey enteric coronavirus nucleocapsid and membrane protein genes: a close genomic relationship with bovine coronavirus

Arnold Verbeek and Peter Tijssen*

Centre de Recherche en Virologie, Institut Armand-Frappier, Université du Québec, Laval-des-Rapides, Québec H7V 1B7, Canada

The 3′ end of the turkey coronavirus (TCV) genome and the gene encoding the nucleocapsid protein (N) were cloned and sequenced. The gene encoding the membrane protein (M) was obtained by cloning a polymerase chain reaction (PCR)-amplified fragment obtained using bovine coronavirus (BCV)-specific primers. Furthermore, five TCV DNA fragments, obtained by PCR on RNA from clinical specimens and corresponding to either the N terminus of the M protein or the complete M protein were also cloned and sequenced. The sequence revealed a 3′ non-coding region of 291 bases, an open reading frame (ORF) encoding the N protein with a predicted size of 448 amino acids, or an M_r of 49K, and an ORF encoding the M protein with a predicted size of 230 amino acids and an M_r of 26K. A third ORF, encoding a hypothetical protein of 207 amino acids with an M_r of 23K was found within the N gene sequence. The amino acid sequences of both the N and M proteins were more than 99% similar to those published for BCV. Extensive similarity was also observed between the amino acid sequences of the TCV N protein and those of murine hepatitis virus (MHV) (70%) and human respiratory coronavirus strain OC43 (HCV-OC43) (98%) and between the amino acid sequences of the predicted M proteins of TCV and MHV (86%). Such striking identity suggests that BCV, TCV and HCV-OC43 must have diverged from each other only recently. A potential *N*-glycosylation site was found at the N terminus of the TCV M protein and is situated at the same location in BCV, MHV and transmissible gastroenteritis virus.

Introduction

The *Coronaviridae* family contains four antigenic groups (Pederson *et al.*, 1978; Sturman & Holmes, 1983). Viruses within each group possess partial antigenic cross-reactivities and infect a variety of mammalian and avian species (Siddell *et al.*, 1983). The viruses possess a single-stranded, polyadenylated RNA genome of about 20 to 30 kb. The genes encoding the viral structural proteins are situated on the last quarter of the 3′ end of the genome. Except for the nucleocapsid protein (N), all other structural proteins so far identified are associated with the lipid membrane. The integral membrane protein (M), which is largely embedded in the lipid bilayer targets the site of virus morphogenesis (Tooze *et al.*, 1984) and may be implicated in viral pathogenesis (Fleming *et al.*, 1989), whereas the bulbous peplomer (S) protein is responsible for virus binding (Cavanagh & Davis, 1986; Koch *et al.*, 1990) as well as virulence and tissue tropism (Wege *et al.*, 1988). An additional surface protein (HE), responsible for haemagglutination, has been found in bovine coronavirus (BCV; King *et al.*, 1985; Hogue *et al.*, 1989; Parker *et al.*, 1989), human respiratory coronavirus strain OC43 (HCV-OC43; Hogue & Brian, 1985), haemagglutinating encephalitis virus of swine (Callebaut & Pensaert, 1980), diarrhoea virus of infant mice (Sugiyama *et al.*, 1986) and turkey coronavirus (TCV; Dea *et al.*, 1986). The HE protein of BCV also exhibits an acetyl esterase receptor-destroying activity similar to the activity found in influenza C viruses (Vlasak *et al.*, 1988).

Our recent studies demonstrated a close antigenic relatedness between TCV and BCV; only a few monoclonal antibodies produced against either of the two viruses were able to differentiate between them, indicating that TCV, which is still placed in an antigenic group distinct from avian infectious bronchitis virus (IBV) and the mammalian coronaviruses, should be reclassified (Dea *et al.*, 1990). Homology between BCV and TCV was further established in hybridization assays (unpublished results). It was demonstrated that BCV-specific probes were efficient in detection and clinical diagnosis of TCV. In order to determine the extent of homology between the two viruses, we cloned and

sequenced the genes encoding the N and M proteins. One of the structural differences observed between TCV and BCV is the type of glycosylation of the M protein, which is *N*- and *O*-glycosylated in TCV and BCV, respectively (Dea *et al.*, 1989*a*; Lapps *et al.*, 1987). We therefore used the polymerase chain reaction (PCR) on nucleic acid isolated from TCV-positive clinical specimens to obtain the M gene or gene fragments corresponding to the N-terminal portion of the M protein. These fragments were cloned and sequenced to establish possible sequence differences associated with the predicted glycosylation sites and to confirm the reliability of the obtained TCV sequence.

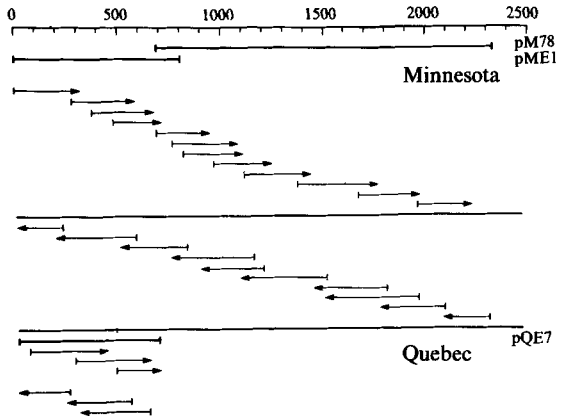

Fig. 1. Strategy used to sequence the N and M genes of TCV-Minnesota and the M gene of a TCV Quebec isolate. pM78 and pME1 represent plasmids containing cDNA inserts of 1·7 and 0·81 kbp, corresponding to the N and M genes, respectively. pQE7 contains a PCR-amplified fragment, corresponding to the M gene of TCV Quebec isolate number 6. All inserts were also subcloned in M13 mp19, analysed for their orientation and subjected to unidirectional deletion. The arrows represent sequences obtained from the deleted-insert clones.

Methods

Virus and cells. The prototype Minnesota strain of TCV (TCV-Minnesota) (Ritchie *et al.*, 1973), kindly supplied by Dr B. S. Pomeroy (College of Veterinary Medicine, St Paul, Mn., U.S.A.), was initially serially propagated by inoculation into the amniotic cavity of 22- to 24-day-old embryonating turkey eggs and further propagated on HRT-18 cells in the presence of 1 unit/ml bovine pancreatic trypsin as described earlier (Laporte *et al.*, 1980; Dea *et al.*, 1989*b*).

Synthesis and cloning of cDNA. Purified, tissue culture-adapted TCV-Minnesota was used for the extraction of RNA (Verbeek & Tijssen, 1988), which was reverse transcribed according to standard procedures (Binns *et al.*, 1985; Gubler & Hoffman, 1983). Tailing (Roychoudhury & Wu, 1980) and cloning of cDNA molecules for the construction of a genomic library was as described for BCV (Verbeek & Tijssen, 1988).

Clone selection and DNA sequencing. Clones from the TCV cDNA library were screened in duplicate by colony hybridization assays (Grunstein & Hogness, 1975) with one probe (a ^{32}P-labelled recombinant plasmid; Rigby *et al.*, 1977) containing sequences corresponding to the 3′ end of the BCV genome as well as a part of the N gene, and another probe containing the additional upstream sequences of the BCV N gene (Verbeek *et al.*, 1990). This approach was chosen because ^{32}P-labelled BCV-specific recombinant plasmids were capable of detecting many TCV isolates and TCV present in clinical samples (unpublished results). Clones that hybridized with both probes were selected for further characterization. Two *Pst*I-generated insert fragments of recombinant plasmid pM78 were subcloned into replicative form DNA of bacteriophage M13mp19, while phage clones with opposite insert orientations, determined according to Poncz *et al.* (1982), were subjected to exonuclease III/nuclease S1 degradation (Henikoff, 1984) to create clones with a nested set of deletions. The TCV M gene was obtained by cloning a fragment amplified by PCR using BCV-specific primers [PXBAV (5′ GAA CAT TTC TAG ATT GGT CGG ACT G 3′) reverse complementary to the sequence located 1527 to 1551 nucleotides from the 3′ end and PC (5′ ATG AGT AGT GTA ACT ACA CCA GCA 3′) hybridizing to nucleotides 2314 to 2337 from the 3′ end] and TCV-Minnesota genomic RNA. The insert from recombinant plasmid pME1 was subcloned in M13mp19 and sequenced in both directions as described above for pM78. Sequencing was according to the method of Sanger *et al.* (1977). Sequences were analysed and compared with the IBI Pustell sequence programs.

Amplification by PCR using RNA isolated from TCV-positive clinical specimens. The supernatant (100 µl) of clarified intestinal contents was supplemented with 1 µg of tRNA (Sigma) before RNA extraction (Chomczynski & Sacchi, 1987). RNA was reverse-transcribed as described earlier (Verbeek & Tijssen, 1990) using a primer complementary to BCV RNA. PCR was performed on cDNA templates with BCV sequence-specific primer combinations: (i) PIORF1 (5′ GGGGGATCC TTA CAC CAG AGG TAG GGG TTC 3′, reverse complementary to the sequence located 951 to 971 nucleotides from the 3′ end) and PIORF2 (5′ GGAAGCTT ATG GCA TCC TTA AGT GGG CCG, complementary to the sequence 1554 to 1574 nucleotides from the 3′ end) to amplify the N-internal open reading frame (ORF) of 624 bp (including the translation stop codon; detecting a fragment of 640 bp containing the primer sequences), (ii) PE1E (5′ GGAAGCTT ATG AGT AGT GTA ACT ACA CCA 3′, complementary to the sequence 2317 to 2337 nucleotides from the 3′ end) and PE1F (5′ GGGGATCC TTA GAT ATT ATT TCT CAA CAA T 3′, reverse complementary to the sequence located 1645 to 1666 nucleotides from the 3′ end) to amplify the 693 bp (including the translation stop codon) TCV M gene (709 bp, including the primer sequences) and (iii) PE1E and PE1G (5′ GGGAGCTC TAA GAT GAT AGT AAG GGG CCA 3′, reverse complementary to the sequence located 2131 to 2151 nucleotides from the 3′ end) to amplify 207 bp fragments (223 bp, including the primer sequences) encoding the N terminus of the TCV M protein. Underlined primer sequences represent non-viral sequences containing restriction endonuclease sites. PCR was for 30 cycles under conditions described earlier (Verbeek & Tijssen, 1990) and in the presence of 0·5 µl of [α-^{32}P]dCTP (ICN; 3000 Ci/mmol, 3·3 µM) as a tracer for the amplified fragments.

Results

cDNA cloning and clone selection

Clone pM78, selected by colony hybridization, contained an insert of about 1·65 kbp, corresponding to the 3′ end of the TCV genome. The insert was subcloned in M13mp19

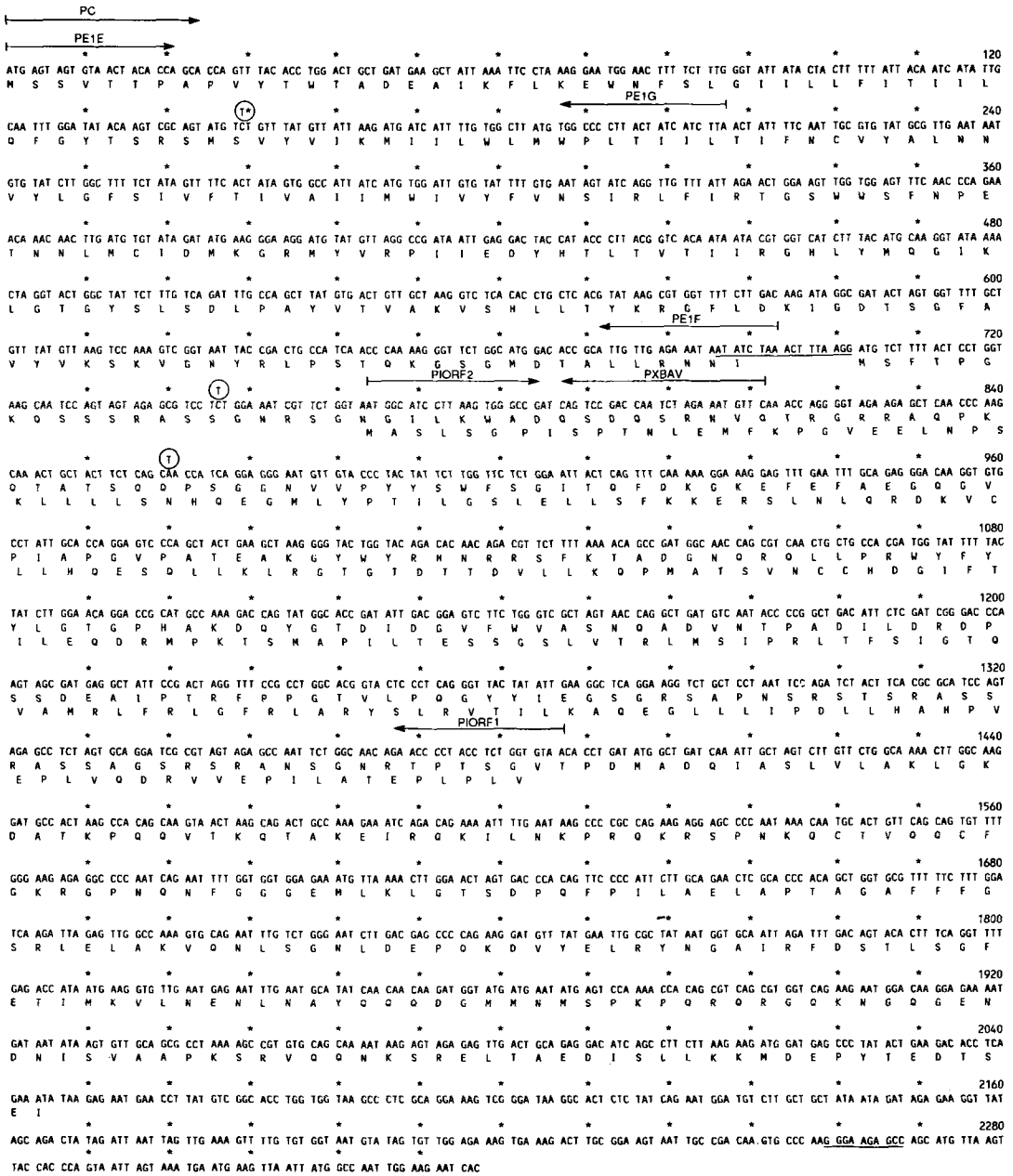

Fig. 2. cDNA sequence of the first 2337 nucleotides of the 3′ end of the TCV genome. Predicted amino acid sequences are shown for three ORFs, corresponding to the M and N genes and a reading frame inside the N gene. Nucleotide differences between the BCV and TCV sequences are indicated in circles above the sequence of TCV. The intergenic consensus region between the N and M genes and the 3′ conserved 10 base sequence are underlined. Arrows correspond to the locations of primers used in PCR amplification.

and both strands were sequenced (Fig. 1). Sequences corresponding to the ORF of the M protein were obtained by cloning a fragment amplified by PCR using BCV-specific primers (PXBAV and PC) and RNA isolated from purified TCV-Minnesota. Clone pME1 was found to contain the expected insert of 811 bp and was subcloned for sequencing (Fig. 1).

Sequence analysis of TCV-Minnesota cDNA clones

The nucleotide sequence of the 3′ end of the TCV genome, i.e. the N and M genes, and their predicted amino acid sequence are shown in Fig. 2. A non-coding region of 291 bases excluding the poly(A) tail was found at the 3′ end of the genome and contains a 10 base

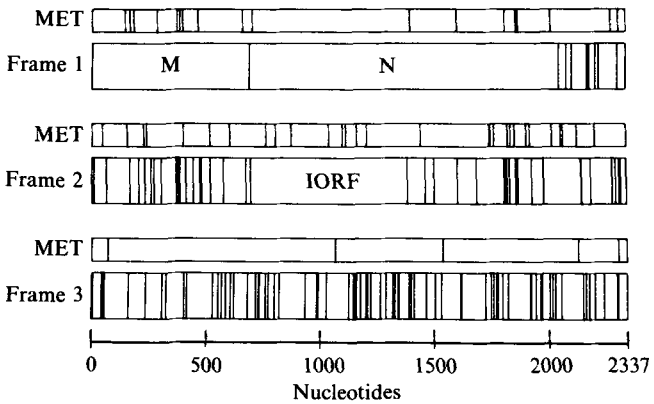

Fig. 3. Schematic design of the location of ORFs obtained when translating three frames of the 2337 nucleotides located at the 3′ end of TCV genomic RNA. Vertical lines in the translated frames represent termination codons, while lines in the 'MET' rectangles represent methionine codons that could serve as translation initiation sites.

consensus region, GGGAAGAGCC, at 70 to 79 bases from the 3′ end. The location and sequence of this region is similar to the consensus regions found in murine hepatitis virus (MHV) and IBV (Boursnell *et al.*, 1985) as well as the consensus published for porcine transmissible gastroenteritis coronavirus (TGEV) (Kapke & Brian, 1986) and two different strains of BCV (Lapps *et al.*, 1987; Crucière & Laporte, 1988).

The largest translational reading frame of 1344 nucleotides (292 to 1635 nucleotides from the 3′ end; Fig. 2 and 3) predicted a 448 amino acid protein with an M_r of 49K, which is likely to encode the N protein because of its location (Spaan *et al.*, 1988) and its predicted M_r which approaches that found for the TCV N protein (Dea & Tijssen, 1988). The consensus region, AUAUCUAAACUUUAAGG, intergenic to N and M, was the same as that for BCV and resembled closely those observed for MHV strains A59 (Armstrong *et al.*, 1983) and JHM (Skinner & Siddell, 1983), and HCV-OC43 (Kamahora *et al.*, 1989).

The second largest translational reading frame (bases 1648 to 2337 from the 3′ end; Fig. 2 and 3) was predicted to encode a protein of 230 amino acids with an M_r of about 26K, which is likely to be the M protein. The predicted protein has 113 hydrophobic residues (approximately 49% hydrophobicity) with a distribution similar to the BCV and MHV hydrophobic amino acids. The first 28 N-terminal amino acid residues contain six potential sites for *O*- and one site for *N*-glycosylation. Most basic amino acid residues (17/23) were found in the C-terminal half of the protein.

An overlapping ORF (bases 951 to 1574 from the 3′ end), predicting a protein of 207 amino acids with an M_r of 23K was found inside the coding sequence of the N protein (Fig. 2 and 3).

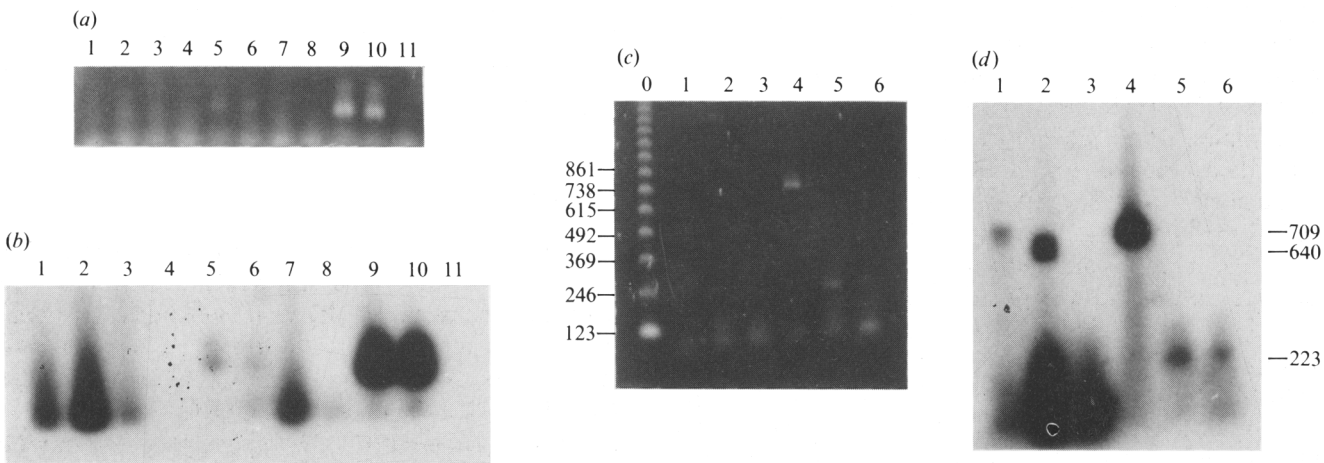

Fig. 4. Electrophoretic profiles (*a* and *c*) of PCR-amplified products and further identification of the fragments by autoradiography of the gels (*b* and *d*). (*a*) Lanes 1 to 8 refer to eight clinical samples and lanes 9 and 10 to third passage culture fluids of two other TCV isolates used to extract nucleic acid for cDNA synthesis and amplification by PCR with primers PE1E and PE1G, respectively. PCR, using the same combination of primers, was also applied to nucleic acid isolated from mock-infected HRT-18 cells (lane 11). The 223 bp amplified fragments represent gene fragments encoding the N terminus of the M protein. Samples of one-tenth of the reaction mixtures were analysed on the gel. Autoradiography of the dried gel (*b*) was for 2 h at −70 °C. Amplification by PCR was done on RNA isolated from clinical specimens 5 and 6, using a combination of primers PE1E and PE1F to amplify the translational reading frame of the M gene (709 bp) (*c*, sample 5 in lane 1; sample 6 in lane 4). Further amplification was assayed using primers PIORF1 and 2 to amplify a 640 bp fragment containing sequences of the translational reading frame inside the N gene. The IORF-amplified products from samples 5 and 6 are shown in (*c*) and (*d*), lanes 3 and 2, respectively. Lanes 5 and 6 in (*c*) and (*d*) refer to the same 223 bp amplified fragments as in (*a*). Lane 0 contains DNA markers (bp). Autoradiography of the dried gel (*d*) was for 5 h at −70 °C.

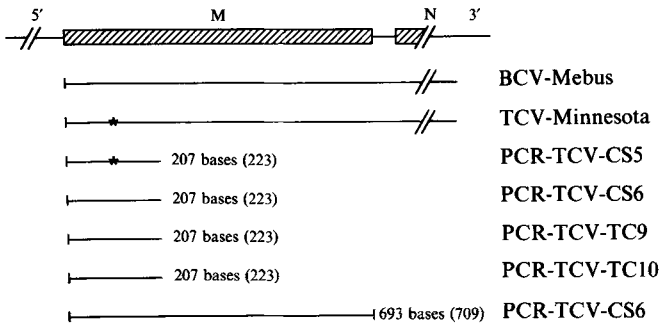

Fig. 5. Schematic representation and comparison of the sequences from cloned PCR-amplified TCV-specific fragments (see Fig. 4) with the corresponding sequence of the TCV-Minnesota and BCV-Mebus strains, respectively. Lines represent identity between the sequences of both viruses, whereas an asterisk refers to single nucleotide differences compared to the BCV sequence. CS (clinical sample) and TC (tissue culture) indicate the origin of the samples used for RNA extraction.

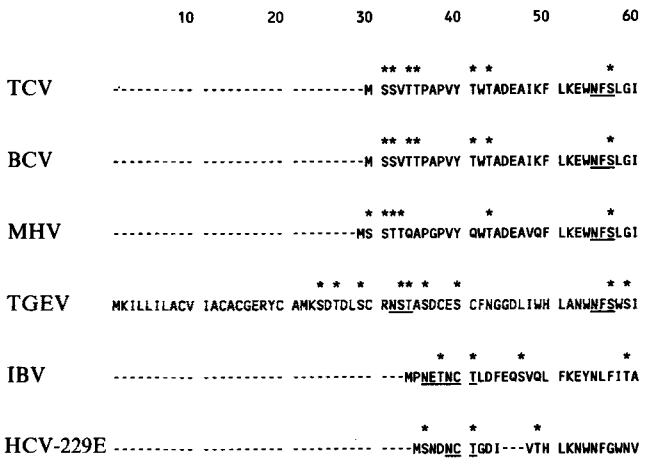

Fig. 6. Amino acid sequence comparison of the TCV 60 residue N terminus of the M protein with corresponding regions of other coronavirus strains by maximum alignment of the amino acid sequence of the complete M proteins. Potential *N*-glycosylation sites are underlined; potential sites for *O*-glycosylation are identified by an asterisk. Numbering corresponds to that of TGEV.

Amplification by PCR

The complete M gene, or gene fragments corresponding to the N terminus of the M protein, were amplified, cloned and sequenced using eight TCV-positive clinical specimens as starting material. Similarly, TCV-containing culture fluid supernatants, obtained after three passages of virus from two different clinical samples, were also subjected to PCR. Agarose gel electrophoresis of 10% of each PCR reaction mixture showed that amplification occurred in two of eight clinical specimens (Fig. 4*a*, *b*; lanes 5 and 6) and in both cultured isolates (Fig. 4*a*, *b*; lanes 9 and 10). Autoradiography revealed significant background amplification in some of the samples (Fig. 4*b*; lanes 1, 2 and 7), which is not observed in samples where actual amplification has occurred. Amplified products were absent in samples after PCR using nucleic acid isolated from mock-infected HRT-18 cells (Fig. 4*a*, lane 11). RNA from clinical samples 5 and 6 was also used for amplification with a combination of primers that would amplify the 624 bp internal ORF (IORF) (640 bp, including the primers) located inside the N gene, and the 693 bp ORF of the M protein (709 bp including the primers). Agarose gel electrophoresis (Fig. 4*c*) of 10% of the reaction products revealed that amplification could only be detected after autoradiography of the gel in two out of four reactions (Fig. 4*d*). The 709 bp amplified fragment of sample 6 (Fig. 4*c*, *d*; lane 4) and the 223 bp fragments of clinical samples 5, 6, 9 and 10 (Fig. 4*a*, *b*), were re-amplified and cloned in pUC-9 after poly(C) tailing.

Sequence analysis of cloned PCR-amplified fragments

Comparison of the sequences of cloned PCR-amplified products with those of TCV-Minnesota and the BCV Mebus strain is presented schematically in Fig. 5. The single nucleotide difference at position 149, in the M genes of TCV and BCV (Fig. 2), was also found in the sequence of the 223 bp fragment obtained from clinical specimen 5. The sequences of the other 223 bp fragments (clinical sample 6 and those obtained from the cultured TCV isolates), as well as the complete M gene (709 bp fragment; plasmid pQE7) of TCV from clinical sample 6, were identical to the sequence published for BCV (Fig. 1, 2 and 5).

Discussion

The sequence of the first 2337 nucleotides from the 3′ end of the TCV RNA genome revealed a 291 base non-coding region and two ORFs with positions corresponding to those for the N and M proteins of coronaviruses (Spaan *et al.*, 1988). The 3′ non-coding 291 base region has a 10 nucleotide sequence (GGGAAGAGCC) which is relatively conserved throughout the *Coronaviridae* family and may be involved in attachment of the polymerase to initiate negative-strand RNA synthesis (Spaan *et al.*, 1988).

The largest translational reading frame of 1344 nucleotides was predicted to encode a 448 amino acid, 49K protein which is likely to be the N protein. The predicted protein is basic and serine-rich (43/448 amino acids) and its serine residues tend to be clustered in two regions. One of these clusters is located at the N terminus of the protein, and the other cluster is situated between

amino acids 190 and 239 from the N terminus. Such clusters are also found with other mammalian and avian coronaviruses (Boursnell *et al.*, 1985; Kapke & Brian, 1986; Lapps *et al.*, 1987; Kamahora *et al.*, 1989) and possibly represent phosphorylation 'hot spots'.

The TCV N protein amino acid sequence shares extensive identity with the analogous sequences of BCV (≥ 99%; Lapps *et al.*, 1987) and HCV-OC43 (98%; Kamahora *et al.*, 1989), although it is classified in a separate antigenic group. On the other hand, little similarity was found with the corresponding sequences of IBV (approx. 30%; Boursnell *et al.*, 1985) and TGEV (approx. 30%; Kapke & Brian 1986). Furthermore, comparison of the N protein amino acid sequences of TCV with MHV, TGEV and HCV-OC43 reveals regions of up to 69 amino acids with significant sequence identity (≥ 90%) which may represent functional domains having survived evolutionary pressures. We showed previously that BCV probes specific to different regions throughout the genome were individually capable of detecting TCV isolates or TCV in clinical specimens (unpublished results). Since homology between BCV and HCV-OC43 has been confirmed by serological studies (Hogue *et al.*, 1984), and RNA fingerprinting data suggest a close resemblance between the remaining as yet unsequenced portions of the genomes (Lapps & Brian, 1985), we also expect an overall genomic relationship between TCV and HCV-OC43, although this has to be further investigated.

Only two nucleotide differences were found between the N protein sequences of TCV and BCV. The first is located towards the N terminus of the protein and results in an amino acid change from Ser in TCV to Phe in BCV at amino acid position 15 of the protein, when compared to the BCV sequence published by Lapps *et al.* (1987). However, Crucière *et al.* (1988) reported a serine residue in the same position in another BCV strain, and this is also the case in HCV-OC43 (Kamahora *et al.*, 1989). The second nucleotide difference at amino acid position 53 of the protein is a Gln in TCV and a Leu in BCV, which again is different in HCV-OC43 and MHV strains JHM and A59.

An IORF of 624 nucleotides within the N gene is analogous to one in BCV; the corresponding region of HCV-OC43 contains two IORFs. The presence of IORFs, either inside the N gene or partially in the 3′ non-coding region, which are often preceded by an AUG codon in a favourable context for translation initiation (Kozak, 1983), is frequently observed with other coronaviruses [i.e. BCV (Lapps *et al.*, 1987), TGEV (Kapke & Brian, 1986), feline coronavirus (de Groot *et al.*, 1988), IBV (Boursnell *et al.*, 1985), HCV-OC43 (Kamahora *et al.*, 1989) and HCV strain 229E (Schreiber *et al.*, 1989)]. It is not yet known whether these IORFs are functional, but it is of interest to determine their possible translation products in virus-infected cells and to verify their translation by means of expression vectors.

Close similarity was also observed between the TCV M protein amino acid sequence and the corresponding sequences of BCV (up to 100%, with a single amino acid difference at position 50) and MHV (86%). The expected membrane topology of the M protein would therefore be likely to resemble the model proposed by Rottier and collaborators (Armstrong *et al.*, 1984; Rottier *et al.*, 1986). Most of the basic amino acids are situated in the C-terminal half of the protein and might, therefore, interact with the negatively charged RNA and the acidic residues of the N protein as suggested by Sturman *et al.* (1980).

The N and M protein and intergenic sequences of TCV-Minnesota were up to 100% the same as the sequence of BCV. Therefore, we envisaged that (i) we might have worked with a laboratory-created recombinant virus, (ii) the HRT-18 cells, which are of human origin and used for the production of all our BCV and TCV isolates, might conceal a latent infection with a closely related human coronavirus that could have been activated upon infection with another coronavirus and (iii) the inoculum may have been contaminated with BCV. The second possibility was not likely as hybridization assays with BCV-specific probes on supernatants or nucleic acid from mock-infected HRT-18 cells (Verbeek & Tijssen, 1988), did not reveal any indication of this and neither did amplification by PCR using nucleic acid from mock-infected cells (Fig. 4). In order to rule out all three possibilities, RNA from different TCV-positive clinical specimens was isolated for cDNA synthesis and amplification by PCR of fragments corresponding to the N terminus or the complete translational reading frame of the M protein (Fig. 4). The addition of tRNA to the samples, as a co-precipitator and a competitor for RNases was essential for the isolation of RNA templates suitable for PCR.

Sequence analysis of the cloned fragments from clinical samples showed that the single amino acid difference between the TCV and BCV M sequences at position 50 of the protein was found again in the N-terminal translated sequence of the amplified 223 bp fragment from clinical isolate number 5 (Fig. 5). This difference was not observed in the corresponding fragments of any other isolate (Fig. 5). Complete identity was also found between the TCV M protein sequence, which was amplified from clinical specimen 6 (Fig. 4), and the sequence of BCV (Fig. 5). The data obtained from these experiments are fully consistent with the M nucleotide sequence obtained for the TCV strain Minnesota.

The only remaining ambiguity is the type of glycosyla-

tion of the M protein which is *N*- and *O*-linked in TCV (Dea *et al.*, 1990) and BCV (Lapps *et al.*, 1987), respectively, although the nucleotide sequence of this region is the same in both viruses. In addition to the possible *O*-glycosylation sites there is one potential site for *N*-glycosylation (Asn-Phe-Ser) within the first 28 N-terminal residues of the TCV M protein and which is, by maximum alignment of the amino acid sequences of HCV-OC43, TGEV, MHV, BCV and IBV, found at the same position in MHV, BCV and TGEV, but not in IBV (Fig. 6). The latter two viruses, in which the M proteins have *N*-linked oligosaccharide side-chains, possess an additional one (TGEV) or two (IBV) potential *N*-glycosylation sites further upstream, while HCV-229E has one potential N-terminal glycosylation site at another position (Fig. 6). Whether the single *N*-glycosylation site is glycosylated in TCV remains to be seen.

The data presented here show that TCV is extremely closely related to BCV to the point that they cannot be distinguished on the basis of the nucleotide sequences so far known. There is expected to be an overall genomic homology as BCV probes to different genomic locations are efficient in detecting TCV isolates (unpublished results). We have also succeeded in amplifying the complete TCV S gene (data not shown), using primer combinations selected from a recently published BCV S gene sequence (Boireau *et al.*, 1990). It is expected that these genes may contain more differences because monoclonal antibodies against the S protein enabled differentiation of the two viruses (Dea *et al.*, 1990).

Interestingly, although TCV, BCV and HCV-OC43 must have only recently diverged from each other, they possess different target cell specificities *in vitro* and infect different animal species. HCV-OC43 causes mainly respiratory diseases, whereas BCV and TCV affect the gastrointestinal system. However, the pathogenicity of TCV and BCV isolates for turkey poults (unpublished results), as well as their c.p.e. *in vitro* are different (Dea *et al.*, 1990). Sequence analysis of the S protein responsible for cell attachment should reveal differences concerning regions important for host cell specificity. The predicted amino acid sequence homology between TCV and BCV, thus far analysed, supports our proposal for the reclassification of TCV, which was previously based mainly on their antigenetic relatedness (Dea *et al.*, 1990).

This research was supported by the Conseil des Recherches et Services Agricoles du Québec and the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche. A.V. acknowledges the support received from the World University Service of Canada. This report was taken in part from a dissertation submitted by A.V. to the Department of Virology, Institut Armand-Frappier, Université du Québec in partial fulfilment of the requirements for the Ph.D. degree.

References

ARMSTRONG, J., SMEEKENS, S. & ROTTIER, P. (1983). Sequence of the nucleocapsid gene from murine coronavirus MHV-A59. *Nucleic Acids Research* **11**, 833–891.

ARMSTRONG, J., NIEMANN, H., SMEEKENS, S., ROTTIER, P. & WASSER, G. (1984). Sequence and topology of a model intracellular membrane protein E1 glycoprotein, from a coronavirus. *Nature, London* **308**, 751–752.

BINNS, M. M., BOURSNELL, M. E. G., FOULDS, I. J. & BROWN, T. D. K. (1985). The use of a random priming procedure to generate cDNA libraries of infectious bronchitis virus, a large RNA virus. *Journal of Virological Methods* **11**, 265–269.

BOIREAU, P., CRUCIERE, C. & LAPORTE, J. (1990). Nucleotide sequence of the glycoprotein S gene of bovine enteric coronavirus and comparison with the S proteins of two mouse hepatitis virus strains. *Journal of General Virology* **71**, 487–492.

BOURSNELL, M. E. G., BINNS, M. M., FOULDS, I. J. & BROWN, T. D. K. (1985). Sequence of the nucleocapsid genes from two strains of avian infectious bronchitis virus. *Journal of General Virology* **66**, 573–580.

CALLEBAUT, P. E. & PENSAERT, M. B. (1980). Characterization and isolation of structural polypeptides in haemagglutinating encephalomyelitis virus. *Journal of General Virology* **48**, 193–204.

CAVANAGH, D. & DAVIS, P. J. (1986). Coronavirus IBV: removal of spike glycopeptide S1 by urea abolishes infectivity and haemagglutination but not attachment to cells. *Journal of General Virology* **67**, 1443–1448.

CHOMCZYNSKI, P. & SACCHI, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* **162**, 156–159.

CRUCIÈRE, C. & LAPORTE, J. (1988). Sequence analysis of bovine enteric coronavirus (F15) genome I. Sequence of the gene coding for the nucleocapsid protein; analysis of the predicted protein. *Annales de l'Institut Pasteur* **139**, 123–138.

DEA, S. & TIJSSEN, P. (1988). Identification of the structural glycoproteins of turkey enteric coronavirus. *Archives of Virology* **99**, 173–186.

DEA, S., MARSOLAIS, G., BEAUBIEN, J. & RUPPANNER, R. (1986). Coronavirus associated with outbreaks of transmissible enteritis (bluecomb) of turkeys in Quebec: hemagglutination properties and cell cultivation. *Avian Diseases* **30**, 319–326.

DEA, S., GARZON, S. & TIJSSEN, P. (1989*a*). Intracellular synthesis and processing of structural glycoproteins of turkey enteric coronavirus. *Archives of Virology* **106**, 239–259.

DEA, S., GARZON, S. & TIJSSEN, P. (1989*b*). Isolation and trypsin enhanced propagation of turkey enteric (bluecomb) coronavirus in a continuous human rectal tumor (HRT-18) cell line. *American Journal of Veterinary Research* **50**, 1310–1318.

DEA, S., VERBEEK, J. A. & TIJSSEN, P. (1990). Antigenic and genomic relationships among turkey and bovine enteric coronaviruses. *Journal of Virology* **64**, 3112–3118.

DE GROOT, R. J., ANDEWEG, A. C., HORZINEK, M. C. & SPAAN, W. J. M. (1988). Sequence analysis of the 3′ end of the feline coronavirus FIPV 79-1146 genome: comparison with the genome of porcine coronavirus TGEV reveals large insertions. *Virology* **167**, 370–376.

FLEMING, J. O., SHUBIN, R. A., SURSMAN, M. A., CASTEEL, N. & STOHLMAN, S. A. (1989). Monoclonal antibodies to the matrix (E1) glycoprotein of mouse hepatitis virus protect mice from encephalitis. *Virology* **168**, 162–167.

GRUNSTEIN, M. & HOGNESS, D. (1975). Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proceedings of the National Academy of Sciences, U.S.A.* **72**, 3961–3965.

GUBLER, U. & HOFFMAN, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263–269.

HENIKOFF, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351–359.

HOGUE, B. G. & BRIAN, D. A. (1985). Structural proteins of human respiratory coronavirus OC43. *Virus Research* **5**, 131–144.

HOGUE, B. G., KING, B. & BRIAN, D. A. (1984). Antigenic relationships among proteins of bovine coronavirus, human respiratory coronavirus OC43, and mouse hepatitis coronavirus A59. *Journal of Virology* **51**, 384–388.

HOGUE, B. G., KIENZLE, T. E. & BRIAN, D. A. (1989). Synthesis and processing of the bovine enteric coronavirus haemagglutinin protein. *Journal of General Virology* **70**, 345–352.

KAMAHORA, T., SOE, L. H. & LAI, M. M. C. (1989). Sequence analysis of the nucleocapsid gene and leader RNA of human coronavirus OC43. *Virus Research* **12**, 1–9.

KAPKE, P. A. & BRIAN, D. A. (1986). Sequence analysis of the porcine transmissible gastroenteritis coronavirus nucleocapsid protein gene. *Virology* **151**, 41–49.

KING, B., POTTS, B. & BRIAN, D. A. (1985). Bovine coronavirus structural proteins. *Virus Research* **2**, 53–59.

KOCH, G., HARTOG, L., KANT, A. & VAN ROOZELAAR, D. J. (1990). Antigenic domains on the peplomer protein of avian infectious bronchitis virus: correlation with biological functions. *Journal of General Virology* **71**, 1929–1935.

KOZAK, M. (1983). Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiological Reviews* **47**, 1–45.

LAPORTE, J., BOBULESCO, P. & ROSSI, F. (1980). Une lignée cellulaire particulièrement sensible à la replication du coronavirus enterique bovin: les cellules HRT 18. *Comptes Rendus de l'Académie des Sciences de Paris* **290**, 623–626.

LAPPS, W., & BRIAN, D. A. (1985). Oligonucleotide fingerprints of antigenically related bovine coronavirus and human coronavirus OC43. *Archives of Virology* **86**, 101–108.

LAPPS, W., HOGUE, B. G. & BRIAN, D. A. (1987). Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. *Virology* **157**, 47–57.

PARKER, M. D., COX, G. J., DEREGT, D., FITZPATRICK, D. R. & BABIUK, L. A. (1989). Cloning and *in vitro* expression of the gene for the E3 haemagglutinin glycoprotein of bovine coronavirus. *Journal of General Virology* **70**, 155–164.

PEDERSEN, N. C., WARD, I. & MENGELING, W. L. (1978). Antigenic relationships of feline infectious peritonitis virus to coronaviruses of other species. *Archives of Virology* **58**, 45–53.

PONCZ, M., SOLOWIECZYK, D., BALLANTINE, M., SCHWARTZ, E. & SURREY, S. (1982). "Non-random" DNA sequence analysis in bacteriophage M13 by dideoxy chain-termination method. *Proceedings of the National Academy of Sciences, U.S.A.* **79**, 4298–4302.

RIGBY, P. W. J., DIECKMANN, M., RHODES, C. & BERG, P. (1977). Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *Journal of Molecular Biology* **113**, 237–251.

RITCHIE, A. E., DESHMUKH, D. R., LARSEN, C. T. & POMEROY, B. S. (1973). Electron microscopy of coronavirus-like particles characteristic of turkey bluecomb disease. *Avian Diseases* **17**, 546–558.

ROTTIER, P. J., WELLING, G. W., WELLING-WESTER, S., NIESTERS, H. G., LENSTRA, J. A. & VAN DER ZEIJST, B. A. M. (1986). Predicted membrane topology of the coronavirus protein E1. *Biochemistry* **25**, 1335–1339.

ROYCHOUDHURY, R. & WU, R. (1980). Terminal transferase catalysed addition of nucleotides to the 3′ termini of DNA. *Methods in Enzymology* **65**, 43–62.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.

SCHREIBER, S. S., KAMAHORA, T. & LAI, M. M. (1989). Sequence analysis of the nucleocapsid protein gene of human coronavirus 229E. *Virology* **169**, 142–151.

SIDDELL, S., WEGE, H. & TER MEULEN, V. (1983). The biology of coronaviruses. *Journal of General Virology* **64**, 761–776.

SKINNER, M. A. & SIDDELL, S. G. (1983). Coronavirus JHM: nucleotide sequence of the mRNA that encodes nucleocapsid protein. *Nucleic Acids Research* **11**, 5045–5054.

SPAAN, W., CAVANAGH, D. & HORZINEK, M. C. (1988). Coronaviruses: structure and genome expression. *Journal of General Virology* **69**, 2939–2952.

STURMAN, L. S. & HOLMES, K. V. (1983). The molecular biology of coronaviruses. *Advances in Virus Research* **28**, 35–112.

STURMAN, L. S., HOLMES, K. V. & BEHNKE, J. (1980). Isolation of coronavirus envelope proteins and interaction with the viral nucleocapsid. *Journal of Virology* **33**, 449–462.

SUGIYAMA, K., ISHIKAWA, R. & FUKUHARA, N. (1986). Structural polypeptides of the murine coronavirus DVIM. *Archives of Virology* **89**, 245–254.

TOOZE, J., TOOZE, S. & WARREN, G. (1984). Replication of coronavirus MHV-A59 in sac-cells: determination of the first site of budding of progeny virions. *European Journal of Cell Biology* **33**, 281–293.

VERBEEK, A. & TIJSSEN, P. (1988). Biotinylated and radioactive cDNA probes in the detection by hybridization of bovine enteric coronavirus. *Molecular and Cellular Probes* **2**, 209–223.

VERBEEK, A. & TIJSSEN, P. (1990). Polymerase chain reaction for probe synthesis and for direct amplification in detection of bovine coronavirus. *Journal of Virological Methods* **29**, 243–256.

VERBEEK, A., DEA, S. & TIJSSEN, P. (1990). Detection of bovine enteric coronavirus in clinical specimens by hybridization with cDNA probes. *Molecular and Cellular Probes* **4**, 107–120.

VLASAK, R., LUYTJES, W., LEIDER, J., SPAAN, W. & PALESE, P. (1988). The E3 protein of bovine coronavirus is a receptor-destroying enzyme with acetylesterase activity. *Journal of Virology* **62**, 4686–4690.

WEGE, H., WINTER, J. & MEYERMAN, R. (1988). The peplomer protein E2 of coronavirus JHM as a determinant of neurovirulence: definition of critical epitopes by variant analysis. *Journal of General Virology* **69**, 87–98.

(Received 4 December 1990; Accepted 11 March 1991)