Journal of General Virology (1992), 73, 2731–2736. *Printed in Great Britain*

Sequence analysis of the membrane protein gene of human coronavirus OC43 and evidence for *O*-glycosylation

Samir Mounir and Pierre J. Talbot*

Virology Research Center, Institut Armand-Frappier, Université du Québec, 531 boulevard des Prairies, Laval, Québec, Canada H7N 4Z3

The gene encoding the membrane (M) protein of the OC43 strain of human coronavirus (HCV-OC43) was amplified by a reverse transcription–polymerase chain reaction of viral RNA with HCV-OC43- and bovine coronavirus (BCV)-specific primers. The nucleotide sequence of the cloned 1·5 kb fragment revealed an open reading frame (ORF) of 690 nucleotides which was identified as the M protein gene from its homology to BCV. This ORF encodes a protein of 230 amino acids with an M_r of 26416. The gene is preceded by the motif UCCAAAC, analogous to the consensus coronavirus transcription initiation sequence. The M protein of HCV-OC43 shows features typical of all coronavirus M proteins studied: a hydrophilic, presumably external N terminus including about 10% of the protein, and a potential *N*-glycosylation site followed by three major hydrophobic transmembrane domains. The amino acid sequence of the M protein of HCV-OC43 has 94% identity with that of the Mebus strain of BCV, and also contains six potential *O*-glycosylation sites in the exposed N-terminal domain. Indeed, the glycosylation of the M protein was not inhibited in the presence of tunicamycin, which is indicative of *O*-glycosylation, as previously reported for BCV and murine hepatitis virus. Virions released from tunicamycin-treated cells contained the M glycoprotein but were devoid of both peplomer (S) and haemagglutinin–esterase (HE) proteins. Thus, inhibition of the *N*-glycosylation of the S and HE structural proteins prevented their incorporation into progeny virions, an indication that they are dispensable for virion morphogenesis, unlike the M protein.

Human coronaviruses (HCVs) are grouped into two major antigenic clusters, represented by the prototype strains 229E and OC43 (Siddell *et al.*, 1983). HCV-OC43 shares antigenic relationships with other coronaviruses such as mouse hepatitis virus (MHV), rat sialodacryadenitis virus, porcine haemagglutinating encephalomyelitis virus, bovine coronavirus (BCV) and rabbit coronavirus. HCV-229E is the prototype strain of another antigenic group which includes porcine transmissible gastroenteritis virus (TGEV), porcine respiratory coronavirus, canine coronavirus, feline enteric coronavirus and feline infectious peritonitis virus (FIPV).

HCVs are recognized as the causative agents of respiratory diseases, being responsible for about 15% of common colds (McIntosh, 1974). Other disease associations have been suggested but are less well documented, for example the involvement of HCV in severe diarrhoea (Resta *et al.*, 1985) or neurological disease such as multiple sclerosis (Burks *et al.*, 1980; Weiss, 1983). MHVs show diverse tropisms, and some MHV strains in rodents have been used as a model system to study chronic and acute hepatic and neurological diseases.

Coronaviruses contain a capped and polyadenylated positive-sense ssRNA molecule of 27 to 31 kb (Boursnell *et al.*, 1987; Lee *et al.*, 1991). Virus-specific mRNAs in infected cells comprise a genomic-sized mRNA plus four to eight subgenomic mRNA species. These mRNAs are arranged in a 3′-coterminal nested set structure, in which the sequence of each mRNA is contained within the sequence of the next larger mRNA. The mRNAs appear to be formed by a mechanism of leader-primed transcription, and a consensus intergenic sequence is the proposed site of fusion of the leader sequence with the mRNA coding region (Lai, 1990).

Previous studies have identified four HCV-OC43 structural proteins: a 190K peplomer (S) glycoprotein (normally present as subunits of 120K and 100K), a 130K haemagglutinin–esterase (HE) glycoprotein, a 55K nucleocapsid (N) phosphoprotein and a 26K membrane (M) glycoprotein (Hogue & Brian, 1986). Various studies have reported remarkable antigenic and genomic similarities between HCV-OC43 and BCV (Hogue *et al.*,

The nucleotide sequence data reported in this paper will appear in the EMBL and GenBank Nucleotide Sequence Databases under the accession number M93390.

```
                                                                5'- GATGTGGATGACG   13

TTTAGGTAATCCAAACATT ATG AGT AGT AAA ACT ACT CCA GCA CCA GTT TAT ATC TGG ACT GCT GAT GAA GCT   86
                    M   S   S   K   T   T   P   A   P   V   Y   I   W   T   A   D   E   A    18
                        *   *       *   *                                   *

ATT AAA TTC CTA AAG GAA TGG AAT TTT TCT TTG GGT ATT ATA CTA CTT TTT ATT ACA ATC ATA TTG CAA   155
I   K   F   L   K   E   W   N   F   S   L   G   I   I   L   L   F   I   T   I   I    L  Q    41
                        o       *

TTT GGA TAT ACA AGT CGC AGT ATG TTT GTT TAT GTT ATT AAG ATG ATT ATT TTG TGG CTT ATG TGG CCC   224
F   G   Y   T   S   R   S   M   F   V   Y   V   I   K   M   I   I   L   W   L   M   W   P    64

CTT ACT ATA ATC TTA ACT ATT TTC AAT TGC GTA TAC GCA TTG AAT AAT GTG TAT CTT GGC CTT TCT ATA   293
L   T   I   I   L   T   I   F   N   C   V   Y   A   L   N   N   V   Y   L   G   L   S   I    87

GTT TTT ACC ATA GTG GCC ATT ATT ATG TGG ATT GTG TAT TTT GTG AAT AGT ATC AGG TTG TTT ATT AGA   362
V   F   T   I   V   A   I   I   M   W   I   V   Y   F   V   N   S   I   R   L   F   I   R    110

ACT GGA AGT TTT TGG AGT TTC AAC CCA GAA ACA AAC AAC TTG ATG TGT ATA GAT ATG AAA GGA ACA ATG   431
T   G   S   F   W   S   F   N   P   E   T   N   N   L   M   C   I   D   M   K   G   T   M    133

TAT GTT AGG CCG ATA ATT GAG GAC TAT CAT ACT CTG ACG GTC ACA ATA ATA CGC GGC CAT CTT TAC ATT   500
Y   V   R   P   I   I   E   D   Y   H   T   L   T   V   T   I   I   R   G   H   L   Y   I    156

CAA GGT ATA AAA CTA GGT ACT GGC TAT TCT TGG GCA GAT TTG CCA GCT TAT ATG ACT GTT GCT AAG GTT   569
Q   G   I   K   L   G   T   G   Y   S   W   A   D   L   P   A   Y   M   T   V   A   K   V    179

ACA CAC CTG TGC ACA TAT AAG CGT GGT TTT CTT GAC AGG ATA AGC GAT ACT AGT GGT TTT GCT GTT TAT   638
T   H   L   C   T   Y   K   R   G   F   L   D   R   I   S   D   T   S   G   F   A   V   Y    202

GTT AAG TCC AAA GTC GGT AAT TAC CGA CTG CCA TCA ACC CAA AAG GGT TCT GGC ATG GAC ACC GCA TTG   707
V   K   S   K   V   G   N   Y   R   L   P   S   T   Q   K   G   S   G   M   D   T   A   L    225

TTG AGA AAT AAT ATC TAA ATTTTAAGGATG -3'                                                    737
L   R   N   N   I               M                                                           230
                                N→
```

Fig. 1. Complete nucleotide sequence of the M protein gene of HCV-OC43 and its deduced amino acid sequence. The intergenic consensus sequences are underlined. The potential *N*-glycosylation (small circle) and surface-accessible *O*-glycosylation (asterisks) sites are indicated.

1984; Lapps & Brian, 1985; Kamahora *et al.*, 1989; Zhang *et al.*, 1992).

As part of our ongoing studies on the molecular characterization of HCVs and their possible involvement in neurological diseases (Arpin & Talbot, 1990; Jouvenne *et al.*, 1990, 1992; Talbot & Jouvenne, 1992), we now report the nucleotide sequence of the gene encoding the M protein of HCV-OC43. Its predicted amino acid sequence is compared with sequences determined for other coronaviruses and shown to be closely similar to that of BCV. Moreover, it is also *O*-glycosylated.

HCV-OC43 was obtained from the ATCC and propagated at 37 °C on the HRT-18 human rectal tumour cell line. Cells were grown as described previously (Jouvenne *et al.*, 1992) except that 10 units/ml TPCK-trypsin (Sigma) was added and infections (m.o.i. 0·2) were done at 37 °C.

Viral mRNA (100 ng) prepared from infected cells (Chirgwin *et al.*, 1979) was reverse transcribed using antisense primer 5′ TCGGCCCACTTGAGGATG 3′, complementary to nucleotides 147 to 165 of the HCV-OC43 N gene (Kamahora *et al.*, 1989). The cDNAs were amplified with a sense primer 5′ CTGGACACCAG-GAGTTAG 3′, located in the 3′ region of the S gene of BCV [nucleotides 290 to 308 (Abraham *et al.*, 1990)] and the antisense primer, using the polymerase chain reaction (PCR; Stewart *et al.*, 1992). Two different purified 1·5 kb PCR products were cloned into the pBluescript II SK(+) vector (Stratagene), and unidirectional deletions were created using exonuclease III and mung bean nuclease (Stratagene). Sequencing was performed on both PCR products by the dideoxynucleotide chain termination method (Sanger & Coulson, 1975) using T7 DNA polymerase (Pharmacia) and [^{35}S]dATP (Amersham). No mismatched bases, additions or deletions were found between the two clones. Sequence analyses, including hydropathy plots (Kyte & Doolittle, 1982), were performed on an Apple Macintosh computer with the MacVector 3.5 (International Biotechnologies) and GeneWorks 2.0 (IntelliGenetics) sequence analysis programs.

To study the effect of tunicamycin on the glycosylation of the viral glycoproteins, a final concentration of 5 μg/ml tunicamycin (Boehringer Mannheim) was added to infected cells and maintained throughout infection. At

```
OC43  M-------------------SSKT-TPAPVYI----------WTADEAIKFLKEWNFSLG-IILLFITIILQFGYTSRSMFVYVIKMIIL   59
BCV   .-------------------..V.-......T----------.....................-.....................   59
TCV   .-------------------..V.-......T----------.....................-.....................S.....   59
A59   .-------------------..T.QA.E...Q----------......VQ.....-.....................I..V.....   60
JHM   .-------------------..T.QA.G...Q----------......VQ.....-.....................I..V.....   60
TGEV  .K-ILLILACVIACACGERYCAMKSDTDLSCRNSTASDCESCFNGGDL.WH.AN....WSI.LIV...VL-.Y.RPQF.W...G.....M   88
FIPV  .KYILLILACIIACVYGERYCAMQ-DSGLQCINGTNSRCQTCFERGDL.WH.AN....WSV.LIV...VL-.Y.RPQF.WL..G.....M   88
229E  .------------------------SNDNC-------------TGDIVTH..N...GWNV.LTI..V.L-...HYKY.RLL.GL..LV.   51
IBV   .------------------------PNETNCT----------LDFEQSVQLF..YNLFIT-AF...L.....Y..AT..KVI.TL...V.   55

OC43  WLMWPLTIIILTIFN--CVYALNN-VYLGLSIVFTIVAIIMWIVYFVNSIRLFIRTGSFWSFNPETNNLMCIDM-KGTMYVRPIIEDYHTL   145
BCV   .............--.......-....F.....................W.....-..R.....   145
TCV   .............--.......-....F.....................W.....-..R.....   145
A59   ........V.C...--.......-..........S.VI..M.....W.....-...V.....   146
JHM   ......I.V.CM..--.......-....F.....SVV...M.....W.....-...V.....   146
TGEV  ..L..VVLA.....AYSE.QVSRY.MF.F..AGA..TFVL..M...R..Q.YR..N.W.....KAIL.VSA-L.RS..L.LEGVPTGV   177
FIPV  ..L..IVLA.....AYSE.QVSRY.MF.F.VAGAV.TFAL.MM...R.VQ.YR..K.W.....AIL.VNA-L.RS..L.LDGTPTGV   177
229E  ..L...VLA.S..DTWANWDS.W-AFVAF.LLMAVSTLV..VM..A..F...R.ART..AW...V.AITVTTV-L.QT.YQ..QQAPTGI   139
IBV   .CF...N.AVGVIS--.T.PP.T-GG.VAA.IL.VFACLSFVG.WIQ.....K.CR.W.....S.AVGS.LLTN.QQCNFA.ESVPMV.   142

OC43  TVTIIRGHLYIQGIKLGTGYSWADLPAYMTVAKVTHLCTYK-RGFLDRISDTSGFAVYVKSKVGNYRLPSTQKGSGMDTALLRNNI   230
BCV   .........M.........LS.....V.....S..L...-.....K.G.....   230
TCV   .........M.........LS.....V.....S..L...-.....K.G.....   230
A59   .A........M..V.....F.LS.....V.....S......-.A..NKVDGV.....-N.P..A......I--   228
JHM   .A......F.M..V.....F.LS.....V.....S......-.A...KVDGV.....-N.P..A..V...I--   228
TGEV  .LTLLS.N..AE.F.IAD.MNIDNL.K.VM..LPSRTIV.TLV.KKLKA.SAT.W.Y.....A.D.STEARTDNLSEQEK..HMV-   262
FIPV  .LTLLS.N..AE.F.MA..LTIEHL.K.VMI.TPSRTIV.TLV.KQLKATTAT.W.Y.....A.D.STEARTDNLSEHEK..HMV-   262
229E  ...LLS.V..VD.HR.AS.VQVHN..E.....VPSTTII.SRV.RSVNSQNST.WVF..RV.H.DFSAV.SPMSNMTENER.LHFF   225
IBV   SPI.KN.V..CE.QW.AK-CEPDH..KDIF.CTPDRRNI.R-MVQKYTGDQSGNKKRFATFVYAKQSVDT-GELESVA.GGSSLYT   225
```

Fig. 2. Amino acid sequence comparison of the HCV-OC43 M protein with that of other coronavirus strains [BCV (Lapps *et al.*, 1987), TCV (Verbeek & Tijssen, 1991), MHV-A59 (Armstrong *et al.*, 1984), MHV-JHM (Pfleiderer *et al.*, 1986), TGEV (Laude *et al.*, 1987), FIPV (Vennema *et al.*, 1991), HCV-229E (Jouvenne *et al.*, 1990) and IBV (Boursnell *et al.*, 1984)] by alignment for maximum identity. Dots indicate residues identical to those of HCV-OC43; hyphens represent gaps introduced into the sequence. The analysis was performed with the GeneWorks 2.0 program (IntelliGenetics) using default settings.

4 h post-infection, 2·5 mCi [^{35}S]methionine/cysteine (Trans^{35}S-Label, ICN Biomedical) was added, after a 1 h methionine and cysteine deprivation period. After 2 h, unlabelled methionine and cysteine were added, as was foetal bovine serum, and infection was allowed to proceed for another 39 h. Radiolabelled virions produced in the presence or absence of tunicamycin were purified on Renografin-60 gradients and structural proteins were analysed as described previously (Arpin & Talbot, 1990).

The complete nucleotide sequence of the HCV-OC43 M gene and its predicted amino acid sequence are shown in Fig. 1, together with potential glycosylation sites and intergenic consensus sequences. The largest open reading frame (ORF) (nucleotides 33 to 725) encodes a protein of 230 amino acids with a predicted M_r of 26416, consistent with the estimated M_r of the M protein of HCV-OC43 determined by SDS–PAGE (Hogue & Brian, 1986; Schmidt & Kenny, 1982; Fig. 3).

As shown in Fig. 2, the M protein of HCV-OC43 is very similar to the corresponding protein of the Mebus strain of BCV (Lapps *et al.*, 1987), which is antigenically related. Indeed, extensive identity exists between the HCV-OC43 and both BCV and turkey coronavirus (TCV) M proteins at the amino acid level (94%). An identity of 83 to 84% is found between the M proteins of HCV-OC43 and the A59 and JHM strains of MHV, which belong to the same antigenic group. On the other hand, the M proteins of the antigenically distinct TGEV, FIPV, HCV-229E and infectious bronchitis virus (IBV) show only 37, 35, 32 and 26% identity to that of HCV-OC43, respectively. The M proteins of both HCV-OC43 and BCV are composed of the same number of residues. Moreover, they also possess identical numbers of basic and acidic amino acids, and are predicted to have similar M_rs. The M protein of HCV-OC43 contains three cysteine residues (positions 74, 126 and 183), whereas BCV lacks the C-terminal one of these. An intergenic sequence, UCCAAAC, identical to the one observed in front of the BCV M gene (Lapps *et al.*, 1987) and some other coronavirus genes [BCV mRNAs 4 and 5 (Abraham *et al.*, 1990); MHV-A59 M gene (Armstrong *et al.*, 1984); MHV-JHM M gene (Pfleiderer *et al.*, 1986)] is present 10 nucleotides upstream of the predicted initiation codon of the HCV-OC43 M protein (Fig. 1).

Like those of BCV, TCV and MHV, there is one potential *N*-glycosylation site in the predicted HCV-OC43 M protein sequence (Asn 26); it is located near the N-terminal, presumably exposed portion of the molecule. Two such exposed sites are found in IBV, TGEV and FIPV, whereas one of three sites in HCV-229E is predicted to be external. However, since the M protein

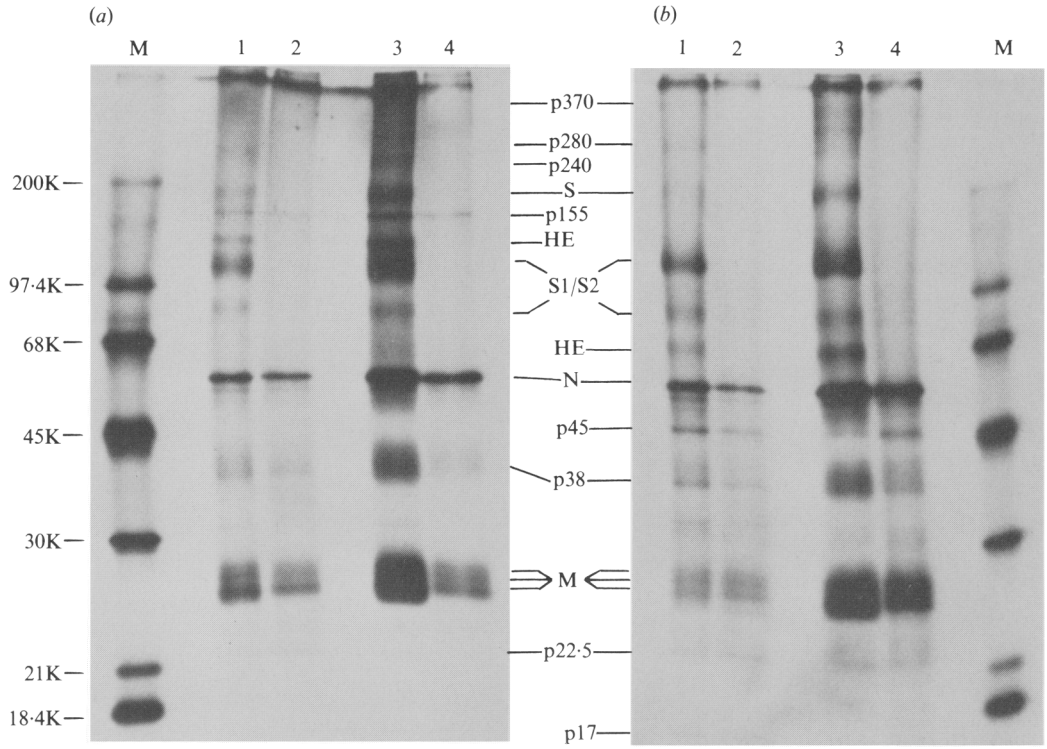

Fig. 3. Effect of tunicamycin on HCV-OC43 proteins. Radiolabelled virions released from HRT-18 cells treated (lanes 2 and 4) or not treated (lanes 1 and 3) with tunicamycin were purified using one (lanes 1 and 2) or two (lanes 3 and 4) Renografin-60 density gradients, and their proteins were identified by SDS–PAGE in the presence (*b*) or absence (*a*) of 2-mercaptoethanol, followed by fluorography (10 day exposure). M_r standards were run on the same gels (lanes M) and their sizes are indicated to the left. Viral proteins are identified either by their apparent M_r or their accepted designation in the case of the major structural proteins: S (180K), S1 and S2 subunits (105K and 80K), HE (125K in unreduced form and 65K in reduced form), N (55K) and M (26K, 27K and 28K).

has previously been shown to be *O*-glycosylated in both BCV (Deregt *et al.*, 1987) and MHV (Holmes *et al.*, 1981; Niemann & Klenk, 1981), we used the *N*-glycosylation inhibitor tunicamycin to verify the type of glycosylation found on HCV-OC43 glycoproteins. As shown in Fig. 3, virions produced in the presence of this drug were completely devoid of both S (180K, 105K and 80K forms) and HE (125K and 65K forms) proteins, but still contained the three forms (28K, 27K and 26K) of the M protein that have previously been shown to be glycosylated (Hogue & Brian, 1986). Therefore, it appears that the M protein of HCV-OC43 is also *O*-glycosylated. Indeed, six potential *O*-glycosylation sites are found in the external portion of the molecule (Fig. 1). These sites are conserved in both BCV and MHV (Fig. 2).

The extensive identity of the sequence of the M proteins of HCV-OC43 and BCV confirms previous reports of a close relationship between these two viruses revealed by serological analysis (Pedersen *et al.*, 1978; Gerna *et al.*, 1981), immunoprecipitation of virion structural proteins (Hogue *et al.*, 1984), oligonucleotide fingerprinting of genomic RNA (Lapps & Brian, 1985) and phylogenetic analysis of the HE gene (Zhang *et al.*, 1992). Moreover, alignment of the sequences of the N proteins of HCV-OC43 and BCV has revealed 97·5% identity (Kamahora *et al.*, 1989). We now find 94% identity between the M proteins of HCV-OC43 and BCV. Together, these findings suggest that these two viruses have diverged from each other only recently. Interestingly, these viruses share the same target cell specificity *in vitro*, but apparently show different tropisms *in vivo*. Furthermore, HCV-OC43 causes mainly respiratory illness in man, whereas BCV affects mainly the gastrointestinal system of cattle. The causes of these differences are not known. However, differences between the two viruses have been detected at the level of some non-structural protein genes (S. Mounir & P. J. Talbot, unpublished results). Sequence variations of the S glycoprotein could also be involved, as suggested previously for the murine and porcine coronaviruses (Parker *et al.*, 1989; Rasschaert *et al.*, 1990). The possibility remains that an HCV-OC43-like coronavirus could be involved in enteric infections (Gerna *et al.*, 1985), although serologically unrelated HCVs have also been reported (Resta *et al.*, 1985).

The intergenic region UCCAAAC upstream of the M

gene of HCV-OC43 has also been identified for the M genes of BCV (Lapps *et al.*, 1987) and two strains of MHV (Armstrong *et al.*, 1984; Pfleiderer *et al.*, 1986). The only other reported occurrence of this exact sequence is in mRNAs 4 and 5 of BCV (Abraham *et al.*, 1990). This conserved sequence shows only one nucleotide difference to the postulated UCUAAAC consensus leader RNA binding site found in 23 of 33 published animal coronavirus gene sequences (data not shown).

Comparison of the hydropathy profiles of all known coronavirus M proteins (data not shown) shows that the expected membrane topology (three hydrophobic domains) is likely to resemble the proposed model (Rottier *et al.*, 1986; Armstrong *et al.*, 1984). Most of the basic amino acids are present in the C-terminal half of the protein, and therefore might interact with the negatively charged RNA and the acidic residues of the N protein (Sturman *et al.*, 1980). Interestingly, only TGEV and FIPV possess an N-terminal hydrophobic sequence.

The glycosylation of the HCV-OC43 M protein was not sensitive to the inhibitory effect of tunicamycin, which is indicative of *O*-glycosylation, as has been reported previously for MHV (Holmes *et al.*, 1981; Niemann & Klenk, 1981) and BCV (Deregt *et al.*, 1987). On the other hand, both the S and the HE proteins of HCV-OC43 were sensitive to this drug and therefore are likely to be *N*-glycosylated. Interestingly, the non-glycosylated precursors of these proteins could not be detected on purified virions, which is consistent with the absence of non-glycosylated S protein from MHV virions (Holmes *et al.*, 1981). Our results suggest that the HE protein is also dispensable for the formation of the viral envelope and virus maturation and release, unlike the M protein.

As shown in Fig. 3, other apparently structural viral proteins were observed in addition to the three envelope glycoproteins and the N protein. Of the large proteins, those with apparent M_rs of 240K, 155K and 38K have also been observed in BCV and associated with S, N and M proteins, respectively (Hogue & Brian, 1986). We also found small proteins (22·5K and 17K) and a 45K molecule, which appears to be the reduced form of a larger protein, possibly p240, which disappears upon reduction.

Six potential *O*-glycosylation sites are observed within the first 28 N-terminal residues of the HCV-OC43 M protein. Moreover, there is one potential site for *N*-glycosylation in this region. This site is conserved at the same relative position in all known sequences of coronavirus M proteins except those of IBV and HCV-229E (Fig. 2), although it is apparently not utilized in HCV-OC43, BCV and MHV.

The relative conservation of the M proteins of coronaviruses suggests that structural constraints on this protein are rigid, resulting in more limited evolution of this protein. The study of the remainder of the genome of HCV-OC43 should yield important information on the replication, tropism and pathogenesis of this important human pathogen. Such studies are in progress.

We thank Francine Lambert for excellent technical assistance. This work was supported by grant MT-92003 from the Medical Research Council of Canada. P.J.T. also gratefully acknowledges salary support in the form of a University Research Scholarship from the National Sciences and Engineering Research Council of Canada (NSERC).

References

ABRAHAM, S., KIENZLE, T. E., LAPPS, W. E. & BRIAN, D. A. (1990). Sequence and expression analysis of potential nonstructural proteins of 4·9, 4·8, 12·7 and 9·5 kDa encoded between the spike and membrane protein genes of the bovine coronavirus. *Virology* **177**, 488–495.

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

ARPIN, N. & TALBOT, P. J. (1990). Molecular characterization of the 229E strain of human coronavirus. In *Coronaviruses and Their Diseases*, pp. 73–80. Edited by D. Cavanagh & T. D. K. Brown. New York: Plenum Press.

BOURSNELL, M. E. G., BROWN, T. D. K. & BINNS, M. M. (1984). Sequence of the membrane protein gene from avian coronavirus IBV. *Virus Research* **1**, 303–313.

BOURSNELL, M. E. G., BROWN, T. D. K., FOULDS, I. J., GREEN, P. F., TOMLEY, F. M. & BINNS, M. M. (1987). Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. *Journal of General Virology* **68**, 57–77.

BURKS, J. S., DEVALD, B. L., JANKOVSKY, L. D. & GERDES, J. C. (1980). Two coronaviruses isolated from central nervous system tissue of two multiple sclerosis patients. *Science* **209**, 933–934.

CHIRGWIN, J. M., PRZYBYLA, A. E., MACDONALD, R. J. & RUTTER, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**, 5294–5299.

DEREGT, D., SABARA, M. & BABIUK, L. A. (1987). Structural proteins of bovine coronavirus and their intracellular processing. *Journal of General Virology* **68**, 2863–2877.

GERNA, G., CEREDA, P. M., GRAZIA REVELLO, M., CATTANEO, E., BATTAGLIA, M. & TORSELLINI GERNA, M. (1981). Antigenic and biological relationships between human coronavirus OC43 and neonatal calf diarrhoea coronavirus. *Journal of General Virology* **54**, 91–102.

GERNA, G., PASSARANI, N., BATTAGLIA, M. & RONDANELLI, E. G. (1985). Human enteric coronaviruses: antigenic relatedness to human coronavirus OC43 and possible etiologic role in viral gastroenteritis. *Journal of Infectious Diseases* **151**, 796–803.

HOGUE, B. G. & BRIAN, D. A. (1986). 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 virus A59. *Journal of Virology* **51**, 384–388.

HOLMES, K. V., DOLLER, E. W. & STURMAN, L. S. (1981). Tunicamycin resistant glycosylation of a coronavirus glycoprotein: demonstration of a novel type of viral glycoprotein. *Virology* **115**, 334–344.

JOUVENNE, P., RICHARDSON, C. D., SCHREIBER, S. S., LAI, M. M. C. & TALBOT, P. J. (1990). Sequence analysis of the membrane protein gene of human coronavirus 229E. *Virology* **174**, 608–612.

JOUVENNE, P., MOUNIR, S., STEWART, J. N., RICHARDSON, C. D. & TALBOT, P. J. (1992). Sequence analysis of human coronavirus 229E mRNAs 4 and 5: evidence for polymorphism and homology with myelin basic protein. *Virus Research* **22**, 125–141.

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

KYTE, J. & DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**, 105–132.

LAI, M. M. C. (1990). Coronavirus: organization, replication, and expression of genome. *Annual Review of Microbiology* **44**, 303–333.

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.

LAUDE, H., RASSCHAERT, D. & HUET, J.-C. (1987). Sequence and N-terminal processing of the transmembrane protein E1 of the coronavirus transmissible gastroenteritis virus. *Journal of General Virology* **68**, 1687–1693.

LEE, H. J., SHIEH, C.-K., GORBALENYA, A. E., KOONIN, E. V., LA MONICA, N., TULER, J., BAGDZHADZHYAN, A. & LAI, M. M. C. (1991). The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. *Virology* **180**, 567–582.

MCINTOSH, K. (1974). Coronaviruses: a comparative review. *Current Topics in Microbiology and Immunology* **63**, 85–129.

NIEMANN, H. & KLENK, H.-D. (1981). Coronavirus glycoprotein E1, a new type of viral glycoprotein. *Journal of Molecular Biology* **153**, 993–1010.

PARKER, S. E., GALLAGHER, T. M. & BUCHMEIER, M. J. (1989). Sequence analysis reveals extensive polymorphism and evidence of deletions within the E2 glycoprotein gene of several strains of murine hepatitis virus. *Virology* **173**, 664–673.

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

PFLEIDERER, M., SKINNER, M. A. & SIDDELL, S. G. (1986). Coronavirus MHV-JHM: nucleotide sequence of the mRNA that encodes the membrane protein. *Nucleic Acids Research* **14**, 6338.

RASSCHAERT, D., DUARTE, M. & LAUDE, H. (1990). Porcine respiratory coronavirus differs from transmissible gastroenteritis virus by a few genomic deletions. *Journal of General Virology* **71**, 2599–2607.

RESTA, S., LUBY, J. P., ROSENFELD, C. R. & SIEGEL, J. D. (1985). Isolation and propagation of a human enteric coronavirus. *Science* **229**, 978–981.

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

SANGER, F. & COULSON, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* **94**, 414–416.

SCHMIDT, O. W. & KENNY, G. E. (1982). Polypeptides and functions of antigens from human coronaviruses 229E and OC43. *Infection and Immunity* **35**, 515–522.

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

STEWART, J. N., MOUNIR, S. & TALBOT, P. J. (1992). Detection of coronaviruses by the polymerase chain reaction. In *Diagnosis of Human Viruses by Polymerase Chain Reaction Technology. Frontiers of Virology*, vol. 1, pp. 316–327. Edited by Y. Becker & G. Darai. Heidelberg: Springer-Verlag.

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

TALBOT, P. J. & JOUVENNE, P. (1992). Neurotropic potential of coronaviruses. *Médecine/Sciences* **8**, 119–125.

VENNEMA, H, DE GROOT, R. J., HARBOUR, D. A., HORZINEK, M. C. & SPAAN, W. J. M. (1991). Primary structure of the membrane and nucleocapsid protein genes of feline infectious peritonitis virus and immunogenicity of recombinant vaccinia viruses in kittens. *Virology* **181**, 327–335.

VERBEEK, A. & TIJSSEN, P. (1991). Sequence analysis of the turkey enteric coronavirus nucleocapsid and membrane protein genes: a close genomic relationship with bovine coronavirus. *Journal of General Virology* **72**, 1659–1666.

WEISS, S. R. (1983). Coronaviruses SD and SK share extensive nucleotide homology with murine coronavirus MHV-A59, more than that shared between human and murine coronaviruses. *Virology* **126**, 669–677.

ZHANG, X., KOUSOULAS, K. G. & STORZ, J. (1992). The hemagglutinin/esterase gene of human coronavirus strain OC43: phylogenetic relationships to bovine and murine coronaviruses and influenza C virus. *Virology* **186**, 318–323.

(Received 21 April 1992; Accepted 15 June 1992)