Sequence and expression of the ns2 protein gene of human coronavirus OC43

Patrick Labonté, 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 complete nucleotide sequence of the ns2 gene of human coronavirus OC43 (HCV-OC43) was determined. Sequence analysis revealed an open reading frame that could encode a protein of 278 amino acids, with an estimated molecular mass of 32.2 kDa. Six potential phosphorylation sites are present but no sites of N-glycosylation were found. The amino acid sequence of the HCV-OC43 ns2 protein shows 92% identity with that of the Mebus strain of bovine coronavirus (BCV). However, a stretch of nine consecutive amino acids near the C terminus is completely different, causing it to be very hydrophilic, which contrasts with the hydrophobic nature of this region in BCV. As shown by immunofluorescence with a monospecific antiserum, the ns2 protein was expressed in the cytoplasm of HCV-OC43-infected HRT-18 cells.

Coronaviruses are enveloped viruses that contain a single-stranded positive sense RNA genome of 27 to 31 kb (Boursnell et al., 1987; Lee et al., 1991). The genomic RNA encodes six to eight capped and polyadenylated subgenomic mRNAs that are arranged in a 3'-coterminal nested set structure. Each mRNA possesses a common 5' end leader sequence derived from the 5' end of the genomic RNA. It has been suggested that the interaction of the 3' end of the leader sequence with the full-length minus strand genomic RNA, at the consensus intergenic sequences, initiates the synthesis of subgenomic mRNAs of diverse lengths (Lai, 1990).

Human coronaviruses (HCV) have so far been represented by two prototype strains, OC43 and 229E, which belong to two distinct antigenic groups. HCV-OC43 shares antigenicity with bovine coronavirus (BCV), haemagglutinating encephalomyelitis virus of swine (HEV) and murine hepatitis virus (MHV). HCV-229E is antigenically related to porcine transmissible gastroenteritis virus (TGEV) and canine coronavirus (CCV). Human coronaviruses are responsible for 15 to 35% of common colds (McIntosh, 1974; Myint, 1994; Wege et al., 1992) and have been associated with severe diarrhoea in the newborn (Resta et al., 1985). Their involvement in neurologic diseases such as multiple sclerosis has also been suggested (Murray et al., 1992; Stewart et al., 1992).

Coronaviruses are enveloped viruses that contain a single-stranded positive sense RNA genome of 27 to 31 kb (Boursnell et al., 1987; Lee et al., 1991). The genomic RNA encodes six to eight capped and polyadenylated subgenomic mRNAs that are arranged in a 3'-coterminal nested set structure. Each mRNA possesses a common 5' end leader sequence derived from the 5' end of the genomic RNA. It has been suggested that the interaction of the 3' end of the leader sequence with the full-length minus strand genomic RNA, at the consensus intergenic sequences, initiates the synthesis of subgenomic mRNAs of diverse lengths (Lai, 1990).

Human coronaviruses (HCV) have so far been represented by two prototype strains, OC43 and 229E, which belong to two distinct antigenic groups. HCV-OC43 shares antigenicity with bovine coronavirus (BCV), haemagglutinating encephalomyelitis virus of swine (HEV) and murine hepatitis virus (MHV). HCV-229E is antigenically related to porcine transmissible gastroenteritis virus (TGEV) and canine coronavirus (CCV). Human coronaviruses are responsible for 15 to 35% of common colds (McIntosh, 1974; Myint, 1994; Wege et al., 1992) and have been associated with severe diarrhoea in the newborn (Resta et al., 1985). Their involvement in neurologic diseases such as multiple sclerosis has also been suggested (Murray et al., 1992; Stewart et al., 1992).

Human coronavirus OC43 comprises four or five major structural proteins: a peplomer (S) glycoprotein, an haemagglutinin-esterase (HE) glycoprotein, a nucleocapsid (N) phosphoprotein and a membrane (M) glycoprotein (Mounir & Talbot, 1992, 1993a; Zhang et al., 1992; Kamahora et al., 1989), as well as a predicted small membrane protein (sM) (Mounir & Talbot, 1993b), apparently similar to the one previously identified in virions of infectious bronchitis virus (IBV) and TGEV (Liu & Inglis, 1991, Godet et al., 1992). In addition to these structural proteins, HCV-OC43 possesses several open reading frames (ORF) that could encode putative nonstructural (ns) proteins (Mounir & Talbot, 1993b).

The OC43 strain of the human respiratory coronavirus and the bovine enteric coronavirus are antigenically very similar since no polyclonal serum can distinguish between them (Hogue et al., 1984). Indeed, the predicted amino acid sequences of all described structural and nonstructural proteins of these two viruses show over 91% identity (Mounir et al., 1994). The major genomic difference observed so far between HCV-OC43 and BCV is the absence on HCV-OC43 of two ORFs that could encode putative nonstructural proteins of 4.9 and 4.8 kDa in BCV (Mounir & Talbot, 1993b; Abraham et al., 1990). Further sequence analysis of the BCV genome has revealed an additional ORF located upstream of mRNA 2 (Cox et al., 1989). This gene was demonstrated to be expressed in BCV-infected cells and to encode a nonstructural phosphoprotein (ns2) of 32 kDa (Cox et al., 1991). A similar protein has been found in MHV-infected cells (Luytjes et al., 1988; Breidenbeek et al., 1990). In the present study, we have cloned, sequenced

* Author for correspondence. Fax +1 514 686 5531.

The nucleotide sequence data reported in this paper have been deposited in the EMBL and GenBank nucleotide sequence databases under accession number L37833.
and expressed in *Escherichia coli* the corresponding region of HCV-OC43 and identified a gene that could encode a protein of 32.2 kDa. With monospecific antisera, we have also detected the expression of this protein in infected cells.

The origin and cultivation of HRT-18 cells and the OC43 strain of HCV has been described previously (Mounir & Talbot, 1992). The nucleotide sequence corresponding to the HCV-OC43 ns2 gene was obtained by RT-PCR on viral RNAs. The antisense primer 5' TTAGTCTTCTTCAAGTTGAGCGAAGTTGCGTGCGTTGCATCCCGCTTCACTGATCTCTTGTTAGATCTTTTTGTAATCTAAACTTTAAAAATGGCTGTC 86

M A V 3

GCT TAT GCA GAC AAG CCT AAT CAT TTT ATC AAT TTT CCA CTT ACC CAT TTT CAG GGT TTT GTG TTA AAT 155

A Y A D K P N H F I N F L H F Q G F V L N 155

GAT GAT CAT GAT GTG GAT GGG TTT CAG ATT AAA TTT CCA CTT ACC CAT TTT CAG GGT TTT GTG TTA AAT 224

Y K G L Q F Q I L E D G C K I Q T A P H I 95

AGT GCT GAC CAA TCC AAG TGG ATT GGC CAT TGC ACC ATA GCT CAA CTC ACG GAT GCA GCA CTG TCC 500

S L T M L D I Q P E D Y K S V D V A I Q E V I 118

GAT GAT ATG CAT TGG GGT GAT GGG TTT CAG ATT AAATTT GAG AAT CCT CAC ATC CTA GGA AGA TGC ATA 362

D D M H W G D G F Q I K F E N P H I L G R C I 95

GTT TTA GAT GTT AAA GGT GTA GAA GAA TTG CAT GAC GAT TTA GTT AAT TAC ATT CGT GAT AAA GGT TGT 431

V L D V K G V E E L H D D L V N Y I R D K G C 118

GTT GCT GAC CAA TCC AAG TGG ATT GGC CAT TGC ACC ATA GCT CAA CTC ACG GAT GCA GCA CTG TCC 500

V A D Q S R K W I G H C T I A Q L T D A L S 141

ATT AAG GAA AAT GAT TTT TTT AAT CCT AAT TTT AAT AAA TTT CAT AAT ACG CCA ACC TCA TCA 569

I K E N V D F I N S M Q F N Y K I T I N P S S 164

CCG GCT AGA CTT GAA ATA GAT GCT GTC GCT GAA AAG GAT GGT TTT TAT GAA ACC ATA GTT AGT 638

P A R L E I V K G A E K K D G F Y E T I V S 187

CAC TGG ATG GGA ATT CGT TTT GAA TAC ACA TCA CCC ACT GAT AAG CTA GTC ATG ATT ATG GAT GTC ATT 707

H W M I R F E Y T S P T D K L A M I G Y C 210

CGT AAG GCT TGT CAA AAT TTA GAT TGT AAT TGT TTG GGG TTT TAT GAA TCT CCA GTT GAA GAA GAC TAA 914

R K A C Q N L D C N C L G F Y E S P V E E D * 278

Fig. 1. Complete nucleotide sequence of the ns2 protein gene of HCV-OC43 and its deduced amino acid sequence. The leader sequence is underlined. The intergenic consensus sequence is doubly underlined. Potential phosphorylation sites (●) are indicated. An asterisk marks the termination codon.

As shown in Fig. 1, the region upstream of the HE gene of HCV-OC43 contains an ORF of 278 amino acids that could encode a protein with an estimated molecular mass of 32.2 kDa. The consensus intergenic sequence UCUAAC observed upstream of many other coronavirus genes (Cox et al., 1989; Shieh et al., 1987) was found 15 nucleotides upstream of the initiation codon. Six potential phosphorylation sites (Fig. 1) but no potential N-glycosylation sites were found in this predicted protein. Previous studies on the BCV ns2 protein revealed the presence of the phosphorylated amino acid residues serine and threonine (Cox et al., 1991). Interestingly, four of the six potential phosphorylation sites are shared by BCV and HCV-OC43, which is consistent with the possibility that the HCV-
OC43 ns2 protein is phosphorylated like its BCV counterpart.

As shown in Fig. 2, the putative HCV-OC43 ns2 protein shares 92% amino acid sequence identity with its BCV counterpart. Interestingly, a stretch of nine consecutive amino acids near the C terminus of the predicted ns2 protein is completely different between the two viruses. The presence of one additional nucleotide on HCV-OC43 but highly hydrophobic in BCV (data not shown). Thus, such a drastic change in this region of the ns2 protein could affect the secondary and tertiary structure of the protein and consequently modify its biological function. The identity levels between the ns2 proteins of MHV-A59, MHV-JHM and BCV are 49% and 51%, respectively (Fig. 2). Hydrophathy plots for these two ns2 proteins revealed that this stretch of nine amino acids is very hydrophilic in HCV-OC43 but highly hydrophobic in BCV (data not shown). Thus, such a drastic change in this region of the protein could affect the secondary and tertiary structure of the ns2 protein and consequently modify its biological function. The identity levels between the ns2 proteins of MHV-A59 and MHV-JHM are 49% and 51%, respectively (Fig. 2).

The ns2 proteins of MHV-A59, MHV-JHM and BCV possess three nucleotide binding domains (Cox et al., 1991; Luytjes et al., 1988). Even though the HCV-OC43 ns2 protein possesses a similar nucleotide binding domain, it is unlikely that this protein could interact with RNA since its isoelectric point is 4.8. Indeed, the negative charge of the ns2 protein and RNA would result in mutual repulsion of these molecules at physiological pH.

To analyse the expression of the HCV-OC43 ns2 protein in infected cells, immunofluorescence assays were performed on HRT-18 cells infected with HCV-OC43 at an m.o.i. of 0.05. At 66 h post infection, the cells were fixed with acetone and incubated first with a mouse antiserum specific for a recombinant form of the BCV ns2 protein (a kind gift of Dr Pascal Boireau, Centre national d’études vétérinaires et alimentaires, Maisons-Alfort, France), then with fluorescein-conjugated F(ab')2 fragments of goat anti-mouse antibody (Cappel Research Products, Organon Technika Inc.) and observed under UV light. As shown in Fig. 3, specific fluorescence was distributed uniformly around the nucleus, which is consistent with the expression of the ns2 protein in the cytoplasm (Fig. 3b). As expected, uninfected cells showed no fluorescence (Fig. 3a). Negative and positive controls were performed on infected cells with a preimmune serum (Fig. 3c) and with a murine monoclonal antibody directed against the HEV nucleocapsid protein (a kind gift of Dr Serge Dea, Institut Armand-Frappier, Laval, Quebec, Canada) (Fig. 3d). Similar results were obtained with a monospecific rabbit antiserum produced against the HCV-OC43 ns2 protein. To produce this antiserum, the HCV-OC43 ns2 gene was reamplified by PCR with sense and antisense primers to which hyphens represent gaps introduced into the sequence to maximize alignment.
administered with complete Freund’s adjuvant and subsequent injections (every 2 weeks) used incomplete Freund’s adjuvant. The antiserum was collected when optimal levels of specific antibodies to the MBP-ns2 fusion protein were reached; this required seven injections, an indication of the low immunogenicity of this protein, at least in rabbits.

In conclusion, we have shown that HCV-OC43 does express a protein of 32 kDa encoded by mRNA 2 (ns2), as was previously observed with BCV (Cox et al., 1991) and MHV (Bredenbeek et al., 1990; Zoltick et al., 1990). It is noteworthy that the ns2 gene has so far only been observed in coronaviruses of the same antigenic group, in which an HE gene is also present, except for MHV-A59, which does nevertheless contain a pseudogene (Luytjes et al., 1988; Shieh et al., 1989). Therefore, possible physical and/or functional interactions between the ns2 and HE proteins or genes need to be investigated.

This completes the characterization of the region downstream of the polymerase gene of HCV-OC43. The polymerase gene has been shown in other coronaviruses to encode several putative nonstructural polypeptides that have yet to be fully characterized (Boursnell et al., 1987; Denison et al., 1991; Herold et al., 1993). The genomic organization of this human coronavirus is thus as follows: 5’ pol/ns2/HE/S/ns4/sM/ns5-1/M/N 3’.

The high amino acid identities (91 to 97%) observed between HCV-OC43 and BCV structural and nonstructural proteins suggests that the two viruses have diverged only recently. However, they have developed different pathogenic properties, including an apparent lack of replication of HCV-OC43 in a bovine host (P. J. Talbot & L. A. Babiuk, unpublished observations). The previously reported genomic deletion in HCV-OC43 compared to BCV (Mounir & Talbot, 1993b; Abraham et al., 1990), remains the only major structural difference between the two viruses, although its relevance to the biology of these structurally related viruses, one a recognized respiratory pathogen (HCV-OC43) and the other a recognized enteric pathogen (BCV), remains to be established. Genetic engineering of the HCV-OC43 or BCV genome will most likely be necessary for a definitive characterization of the molecular basis of virus tropism.

We thank Dr Pascal Boireau (Centre national d’études vétérinaires et alimentaires, Maisons-Alfort, France) for his generous gift of a polyclonal mouse antiserum to the BCV ns2 protein and Dr Serge Dea (Institut Armand-Frappier, Laval, Québec, Canada) for his generous gift of a murine monoclonal antibody to the HEV N protein. We are also indebted to Francine Lambert for excellent technical assistance. This work was supported by grant MT-9203 from the Medical Research Council of Canada (MRC) to P.J.T., who also gratefully
acknowledges salary support in the form of a senior scholarship from the 'Fonds de la recherche en santé du Québec' (FRSQ). P.L. acknowledges a studentship from the 'Fonds pour la Formation de Chercheurs et l'Aide à la Recherche' du Québec (FCAR).

References


(Received 27 July 1994; Accepted 12 October 1994)