

CHARACTERIZATION OF THE NONSTRUCTURAL AND SPIKE PROTEINS OF THE HUMAN RESPIRATORY CORONAVIRUS OC43: COMPARISON WITH BOVINE ENTERIC CORONAVIRUS

Samir Mounir, Patrick Labonté and Pierre J. Talbot

Centre de Recherche en Virologie
Institut Armand-Frappier, Université du Québec
Laval, Québec, Canada H7N 4Z3

ABSTRACT

The nucleotide sequence of the region between the spike (S) and the membrane (M) protein genes, and sequences of the S and ns2 genes of the OC43 strain of human coronavirus (HCV-OC43) were determined. The ns2 gene comprises an open reading frame (ORF) encoding a putative nonstructural (ns) protein of 279 amino acids with a predicted molecular mass of 32-kDa. The S gene comprises an ORF encoding a protein of 1353 amino acid residues, with a predicted molecular weight of 149,918. Sequence comparison between HCV-OC43 and the antigenically related bovine coronavirus (BCV) revealed more sequence divergence in the putative bulbous part of the S protein (S1) than in the stem region (S2). The cysteine residues near the transmembrane domain and the internal predicted protease cleavage site are conserved in the HCV-OC43 S protein. Nucleotide sequence analysis of the region between the S and M gene loci revealed the presence of an unexpected intragenomic partial leader sequence and two ORFs encoding potential proteins of 12.9 and 9.5-kDa. These two proteins were identified as nonstructural by comparison with the homologous BCV genes. *In vitro* translation analyses demonstrated that the HCV-OC43 9.5-kDa protein, like its BCV counterpart, is poorly translated when situated downstream of the 12.9-kDa ORF, but is expressed in infected cells, as shown by immunofluorescence. Interestingly, two ORFs, potentially encoding 4.9 and 4.8-kDa ns proteins in BCV are absent in HCV-OC43, indicating that they are not essential for viral replication in HRT-18 cells.

INTRODUCTION

Coronaviruses are enveloped RNA viruses, isolated from a range of animal hosts (for review, see 1). They cause respiratory, gastrointestinal and neurological disorders in human and domestic animals^{2,3}. Coronavirus genomic RNA directs the synthesis of a nested set of 6 to 8 subgenomic mRNAs by a leader-primed mechanism of transcription⁴. Besides the genes for known structural proteins, coronavirus genomes contain several open reading frames (ORFs) that could encode nonstructural (ns) proteins or novel structural proteins. The number and position of the genes encoding these proteins differ between coronavirus species.

Human coronavirus OC43 (HCV-OC43) comprises four major structural proteins: a 190-kDa peplomer (S) glycoprotein (cleaved into two subunits of 120 and 100-kDa), a

130-kDa hemagglutinin-esterase (HE) glycoprotein, a 55-kDa nucleocapsid (N) phosphoprotein, and a 26-kDa membrane (M) glycoprotein^{5,6,7}. The amino acid sequences of structural proteins HE, M and N, as well as the leader sequence of HCV-OC43 have recently been determined from cloned cDNAs^{7,8,9}. Within the genome, the structural proteins are arranged in the order 5'-HE-S-M-N-3'^{7,8}. Dispersed throughout the genome are genes that encode the RNA-dependent RNA polymerase and nonstructural proteins.

Here we report the nucleotide sequence of the region between the spike (S) and the membrane (M) protein genes and sequences of the S and ns2 genes of HCV-OC43, as well as Northern blot and *in vitro* translation analyses.

MATERIALS AND METHODS

Virus Growth, Isolation and Analysis of Nucleic Acids

The origin and cultivation of the HRT-18 human rectal tumor cell line and the OC43 strain of HCV, as well as the preparation, reverse transcription, polymerase chain reaction (PCR) amplification, cloning and sequencing of viral RNA (both mRNA and genomic RNA) were performed as described elsewhere⁸. For Northern blot analysis, poly (A)-containing RNA was selected with the PolyAtract[®] mRNA isolation system (Promega, Fisher Scientific, Montréal, Québec, Canada) according to the manufacturer's instructions, was size fractionated by electrophoresis in 1% (w/v) agarose gels containing 5.3% formaldehyde, and was transferred onto Hybond-C[™] extra (Amersham Canada Ltd., Oakville, Ontario, Canada) nitrocellulose filters¹⁰. Blots were hybridized with the random-primed ³²P-labeled (ICN Biomedicals Canada Ltd., Mississauga, Ontario, Canada) DNA probes at 42°C as described previously¹¹.

Immunofluorescence

Immunofluorescence was performed 24 h after infection of HRT-18 cells grown on coverslips with HCV-OC43 at an MOI of 0.2. Cells were examined for fluorescence with a specific polyclonal rabbit antibody prepared against the MHV-A59 9.6-kDa protein that had been expressed in *Escherichia coli*¹², or with a polyclonal rabbit antiserum against HCV-OC43.

RESULTS AND DISCUSSION

Nucleotide Sequence Analysis

The nucleotide sequence of the region between the S and M genes of HCV-OC43 contains potential open reading frames for two proteins of 12.9-kDa and 9.5-kDa (data not shown).

The 12.9-kDa protein would contain 109 residues and show an amino acid sequence identity of 96.3% with the putative 12.7-kDa protein of BCV¹³ (Fig. 1). One potential N-linked glycosylation site is found at amino acid position 18, like in BCV¹³. The putative initiation codon for this protein is in a context not frequently used for initiation of protein synthesis as described by Kozak¹⁴. Surprisingly, a 47-nucleotide stretch that has resemblance with the 3'-half of the 82-nucleotide HCV-OC43 leader sequence⁹ is found upstream of this ORF (Fig. 2). This situation is unexpected since part of the leader sequence (47 of 82 nucleotides) is found within the genomic RNA instead of being at the 5'-end of the transcript. This may have occurred by a mechanism of recombination aimed at the conservation of the ORF encoding the 12.9-kDa protein, despite the loss of the 4.9- and 4.8-kDa putative ns proteins. This partial intragenomic leader sequence is preceded by a potential 11-residue ORF that could encode a 1.33-kDa polypeptide composed of 9 residues identical with the N-terminus of the BCV 4.9-kDa putative ns protein.

The 9.5-kDa ORF predicts an 84-amino acid protein that would have an amino acid sequence identity of 96.4% with the 9.5-kDa protein of BCV¹³ (Fig.1). As in BCV, there is a conserved methionine residue at the putative third codon of the protein. The first putative initiation codon is not in a favorable context for initiation of protein synthesis, whereas the presence of a G residue at position +4 of the second putative initiation codon would presumably ameliorate the situation¹⁴. The 9.5-kDa protein contains one large hydrophobic domain that comprises more than 50% of the molecule, which suggests a transmembrane insertion (data not shown).

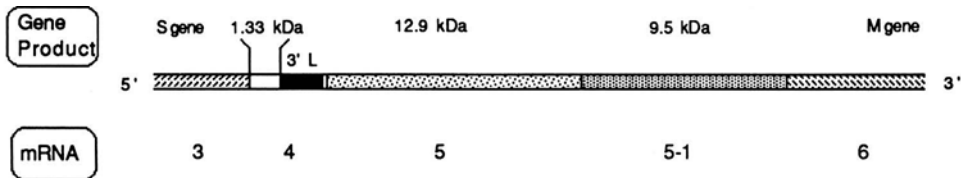


Figure 1. Organization of HCV-OC43 genome, location of ORFs and identity with BCV.
 Δ: Absence of two ORFs in HCV-OC43.

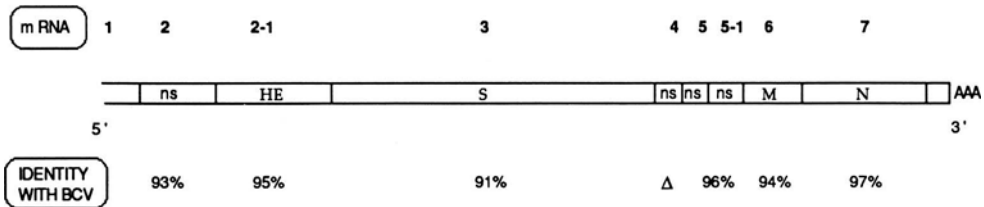


Figure 2. 3'-Leader sequence within HCV-OC43 genomic RNA.

Along with the four structural proteins and the three ORFs encoding potential non-structural proteins, we found a putative gene situated immediately upstream from the HE gene, which contains an ORF that could encode a protein of 279 residues, with a predicted molecular weight of 32,277. This protein would have an amino acid sequence identity of 93% with the BCV ns2 protein¹⁵ (Fig. 1, Table 1). Nine consecutive amino acid residues of the ns2 HCV-OC43 C-terminus are different from those of BCV ns2 protein (data not shown).

Finally, the complete sequence of the S gene contains 4062 bases, encoding a protein of 1353 amino acid residues, with a predicted molecular weight of 149,918 (data not shown). This protein has an amino acid sequence identity of 91% with the BCV S protein (Fig. 1, Table 1), revealing more sequence divergence in the putative bulbous part of the S protein (N-terminus, S1) than in the S2 stalk region. The putative S ORF is preceded immediately upstream by the consensus intergenic sequence (UCUAAAC) thought to play a role in leader priming of coronavirus transcription. The deduced HCV-OC43 S protein reflects the properties of the other coronavirus spike proteins that have been characterized to date. There is an internal sequence of basic amino acid NRRSRG (KRRSRR for BCV) beginning with amino acid 753 which, on the basis of the pattern in BCV¹⁶, predicts a cleavage between amino acids 758 and 759, giving an N-terminal segment of 84,640 Da (S1) and a C-terminal segment of 65,296 Da (S2).

Northern Blot Analysis, *In Vitro* Translation and Immunofluorescence

To identify the subgenomic RNAs that encode the structural and nonstructural proteins, we performed Northern blot analysis using DNA probes. Nine HCV-OC43-specific RNAs were detected (data not shown) and have been numbered 1, 2, 2-1, 3, 4, 5, 5-1, 6, 7 in order of decreasing sizes (Table 1). A 9.5-kDa probe hybridized to RNA 1 to 5-1, but not RNAs 6 and 7. A 12.9-kDa probe hybridized to RNA 1 to 5, but not RNAs 5-1, 6 and 7. A S2 probe (which extends into the putative ns4 ORF) hybridized to RNA 1 to 4 only (Fig. 3). Thus, it appears that RNAs 5 and 5-1 encode the nonstructural proteins 12.9-kDa and 9.5-kDa, respectively, a situation similar to that for the BCV homologues¹³, and different from mouse hepatitis virus (MHV), where only one transcript (mRNA 5) is utilized for the synthesis of both the 13- and 9.6-kDa ns proteins¹⁷. Since the initiation

Table 1. HCV-OC43 mRNAs and gene product sizes

mRNA	Predicted Size of mRNAs (bases) ¹		Gene Product		
	OC43	BCV	Designation	Size (kDa) ²	
				OC43	BCV
1					
2	9232	9544	ns2	32.3	32
2-1	8385	8700	HE	47.7	47.7
3	7086	7400	S	150	150.7
4	3350	3635	ns4	1.3	4.8 + 4.9
5	3032	3005	ns5	12.9	12.7
5-1	2753	2735	ns5-1	9.5	9.5
6	2345	2345	M	26.4	26.4
7	1649	1649	N	49.3	49.4

¹Size of mRNAs predicted from the location of the intergenic nucleotide sequence.

²Size of polypeptides deduced from the ORF immediately downstream of the intergenic nucleotide sequence.

codon for the BCV 12.7-kDa ns protein is in a more favorable context for initiation of translation than the HCV-OC43 12.9-kDa ns protein¹⁴ (G instead of U in position -3), we tested the translatability of both ORFs, either independently or in tandem. In BCV, *in vitro* translation experiments using a synthetic transcript containing both the 12.7- and 9.5-kDa ORFs demonstrated that the majority of the protein synthesized was the upstream 12.7-kDa protein¹³. However, when the same experiment was performed on a MHV RNA containing both ORFs (13- and 9.6-kDa), the downstream ORF (9.5-kDa) was preferentially synthesized¹⁷. We also observed that, like in BCV, the upstream 12.9-kDa protein was preferentially synthesized when both ORFs were present but that synthetic RNA containing only one of the two ORFs directed efficient translation of both proteins.

We also showed that the HCV-OC43 9.5-kDa protein is expressed during infection, using an antibody directed against the MHV 9.6-kDa protein¹², which specifically stained HCV-OC43-infected HRT-18 cells in an immunofluorescent assay (Fig. 4).

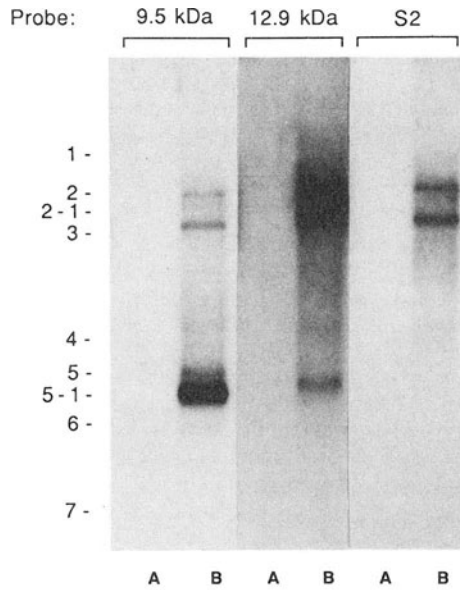


Figure 3. Northern blot analysis of RNA from uninfected cells (A) and infected cells (B). mRNAs were revealed with the probes indicated above the figure. Subgenomic mRNAs are indicated on the left.

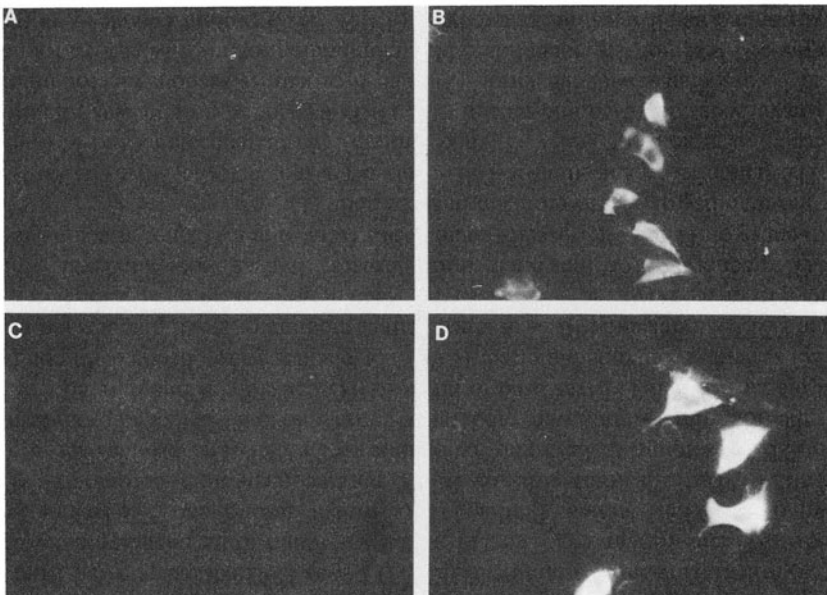


Figure 4. Immunodetection of the 9.5-kDa protein. Uninfected and HCV-OC43-infected cells revealed with an hyperimmune anti-OC43 serum (A and B, respectively), or with anti-9.5-kDa serum (C and D, respectively).

The functions of the putative 12.9-kDa and 9.5-kDa proteins in infected cells are not known. Recently, proteins analogous to the HCV-OC43 and BCV 9.5-kDa molecules were shown to be present in infectious bronchitis (IBV) and porcine transmissible gastroenteritis (TGEV) virions and were designated small membrane (SM) proteins^{18,19}. We have shown that the HCV-OC43 9.5-kDa ns protein, like its BCV and MHV counterparts, is expressed in virus-infected cells. It will also most likely be found in virions. The conservation of the 12.9-kDa protein also suggests its importance in coronavirus biology.

We found a major difference between HCV-OC43 and BCV, represented by the absence of the genes potentially encoding two nonstructural proteins of 4.9- and 4.8-kDa in BCV¹³. Their absence in HCV-OC43 indicates that they are not essential for virus replication, at least in HRT-18 cells. Similarly, the ns2, ns4 and ns5a nonstructural proteins of MHV were shown not to be essential for virus replication in tissue culture^{20,21}. It is interesting to note that the difference between mRNA 4 and 5 is only 320 bp, with one potential ORF of 11 amino acids. The presence of mRNA 4 in HCV-OC43-infected cells (Fig.3), in the absence, most likely, of an associated protein product suggests that this mutation may be a recent event. Alternatively, it is possible that this mRNA encodes non-essential gene products that have been deleted during evolution.

The fundamental importance of the different nonstructural proteins encoded by coronaviruses remains to be explored. However, the absence of the two putative nonstructural proteins in HCV-OC43 (4.8- and 4.9-kDa), as well as amino acid differences within the S protein could be involved in the apparent preferential respiratory tropism of HCV-OC43, which contrasts with the presumed preferential enterotropism of BCV. Indeed, the S proteins of MHV and TGEV were suggested to be important in tissue tropism^{22,23}. However, replication of BCV in the respiratory tract has been reported⁷ and OC43-like human enteric coronaviruses have been isolated²⁴. Thus, it is premature to conclude on the molecular basis for a differential tropism of these two viruses.

ACKNOWLEDGMENTS

This work was supported by grant MT-9203 from the Medical Research Council of Canada to P.J.T., who also gratefully acknowledges salary support in the form of a University Research Scholarship from the National Sciences and Engineering Research Council of Canada (NSERC). We thank Dr. J. L. Leibowitz (University of Texas Health Science Center, Houston, Texas) for his generous gift of a polyclonal rabbit antiserum to the MHV 9.6-kDa protein.

REFERENCES

1. W. Spaan, D. Cavanagh, and M.C. Horzinek. *In*: "Immunochemistry of viruses, II. The basis for serodiagnosis and vaccines". M.H.V. van Regenmortel and A.R. Neurath, eds., p. 359, Elsevier, New York (1990).
2. K.V. Holmes. *In*: "Virology, 2nd edn.", B.N. Fields, D.M. Knipe *et al.*, eds., p. 841, Raven Press, New York (1990).
3. H. Wege, S. Siddell, and V. ter Meulen. *Curr. Top. Microbiol. Immunol.* 99: 165 (1982).
4. M.M.C. Lai. *Annu. Rev. Microbiol.* 44: 303 (1990).
5. B.G. Hogue, B. King, and D.A. Brian. *J. Virol.* 51: 384 (1984).
6. B.G. Hogue, and D.A. Brian. *Virus Res.* 5: 131 (1986).
7. X.M. Zhang, K.G. Kousoulas, and J. Storz. *Virology* 186: 318 (1992).
8. S. Mounir, and P.J. Talbot. *J. Gen. Virol.* 73: 2731 (1992).
9. T. Kamahora, L.H. Soe, and M.M.C. Lai. *Virus Res.* 12: 1 (1989).
10. E. Southern. *Methods Enzymol.* 69: 152 (1979).
11. M. Cogné, S. Mounir, J.L. Preud'homme, F. Nau, and P. Guglielmi. *Eur. J. Immunol.* 18: 1485 (1988).
12. J.L. Leibowitz, S. Perlman, G. Weinstock, J.R. DeVries, C. Budzilowicz, J.M. Weisseman, and S.R. Weiss. *Virology* 164: 156 (1988).

13. S. Abraham, T.E. Kienzle, W.E. Lapps, and D.A. Brian. *Virology* 177: 488 (1990).
14. M. Kozak. *J. Cell Biol.* 108: 229 (1989).
15. G.J. Cox, M.D. Parker, and L.A. Babiuk. *Nucl. Acids Res.* 17: 5847 (1989).
16. S. Abraham, T.E. Kienzle, W. Lapps, and D.A. Brian. *Virology* 176: 296 (1990).
17. C.J. Budzilowicz, and S.R. Weiss. *Virology* 157: 509 (1987).
18. D.X. Liu, and S.C. Inglis. *Virology* 185: 911 (1991).
19. M. Godet, R. L'Haridon, J.-F. Vautherot, and H. Laude. *Virology* 188: 666 (1992).
20. B. Schwarz, E. Routledge, and S.G. Siddell. *J. Virol.* 64: 4784 (1990).
21. K. Yokomori, and M.M.C. Lai. *J. Virol.* 65: 5605 (1991).
22. D. Rasschaert, M. Duarte, and H. Laude. *J. Gen. Virol.* 71: 2599 (1990).
23. T.M. Gallagher, S.E. Parker, and M.J. Buchmeier. *J. Virol.* 64: 731 (1990).
24. G. Gerna, N. Passarani, M. Battaglia, and E.G. Rondanelli. *J. Inf. Dis.* 151: 796 (1985).