

Nucleotide sequence of the gene encoding the membrane protein of human coronavirus 229 E

Brief Report

T. Raabe and S. G. Siddell

Institute of Virology, University of Würzburg, Würzburg, Federal Republic of Germany

Accepted June 22, 1989

Summary. The sequence of the gene encoding the membrane protein of human coronavirus 229 E (HCV 229 E) has been determined. The primary translation product, deduced from the DNA sequence, is a polypeptide of 225 amino acids with a predicted molecular weight of 26,000. The polypeptide has 3 potential N-glycosylation sites. Many structural similarities with the membrane proteins of other coronaviruses can be recognized.

*

The coronaviruses are a group of positive strand RNA viruses that cause a wide spectrum of disease in mammals and birds [21]. The human coronaviruses are thought to cause 10–20% of all common colds, and about half of these are associated with the human coronavirus strain HCV 229 E [4, 5]. The HCV 229 E virion is comprised of a genomic RNA of approximately 6×10^6 molecular weight (mol.wt), a lipid envelope and three major proteins; a phosphorylated protein of 50,000 mol.wt. associated with the genome, a glycosylated 180,000 mol.wt. protein and a family of polypeptides with estimated molecular weights of 25,000, 23,000, and 21,000 [7, 16]. It is clear that these proteins represent the nucleocapsid (N), surface (S) and membrane (M) proteins characteristic of coronaviruses [reviewed in 19].

The replication of HCV 229 E appears to follow the pattern which has now been well established for several coronaviruses, notably, avian infectious bronchitis virus (IBV) and murine hepatitis virus (MHV) [reviewed in 19]. In the HCV 229 E infected cell a set of 6 3'-coterminal subgenomic mRNAs are synthesized, the smallest of which, mRNA 7, encodes the nucleocapsid protein (22, 17). As for other coronaviruses, the synthesis of the HCV 229E subgenomic mRNAs appears to involve a leader-primed mechanism of discontinuous tran-

scription in which a specific intergenic sequence (TCTAAAC for MHV and HCV) plays an important role [17]. In this paper we report the nucleotide sequence of the genomic region of HCV 229E which encodes the membrane protein. This region is adjacent to the 3' terminal nucleocapsid protein gene and corresponds to the unique region of mRNA 6.

HCV 229E virus was obtained from Dr. D. A. J. Tyrrell of the MRC Common Cold Unit, Salisbury, U.K., plaque purified and propagated in C16 cells at 33 °C as described previously [11]. The cytoplasmic RNA from 10⁸ cells which had been infected 48 h previously at an m.o.i. of 3 was extracted by standard procedures [18] and the poly A-containing fraction selected by hybridization to poly U-Sepharose. cDNA synthesis was performed according to the procedure of Gubler and Hoffman [3] using 5 µg of RNA primed with random hexanucleotides. The double-stranded cDNA was ligated to EcoRI linkers and cloned into the EcoRI site of the pBluescript vector, pKS⁺ (Stratagene, Federal Republic of Germany). After transformation, a library of recombinant clones was established and screened by colony hybridization with a synthetic oligonucleotide, 5'TTGAACATTCCAATAGCC3', which is complementary to a region 165–183 bases from the 5' end of the HCV 229E nucleocapsid gene [Myint et al., in prep.; 17]. This search identified the clone 2F7 which hybridized to all 7 virus specific mRNA species in HCV 229E infected cells (data not shown). The cDNA insert of the clone 2F7 was sequenced completely on both strands using restriction endonuclease fragments subcloned into the M13 vector, mp18 [9] and the dideoxyribonucleotide chain termination method of Sanger [15]. Universal and sequence specific primers synthesized on a Cyclone DNA synthesizer (Milli Gen, Federal Republic of Germany) were used. The sequence data were assembled and analysed using the programs of Staden [20] and the University of Wisconsin Genetics Group [2].

The sequence of the clone 2F7 corresponding to the genomic region representing the unique region of mRNA 6 is shown in Fig. 1. This region contains a single large open reading frame (ORF) of 678 nucleotides. The open reading frame is flanked on either side by the nucleotide sequence TCTAAAC (Fig. 1, nucleotide positions 33–39 and 722–728) which would be at or near the sites of fusion between the leader RNA and the mRNA 6 and mRNA 7 coding regions, respectively [17]. Fourteen nucleotides downstream from the large ORF is an AUG codon which represents the initiation codon of the nucleocapsid gene [Myint et al., in prep.] and 12 nucleotides upstream from the large ORF is a TAA codon which represents the termination codon of the 5' proximal gene [Raabe et al., in prep.].

The large open reading frame predicts a polypeptide of 225 amino acids with a molecular weight of 26,000. The predicted polypeptide has several features which are characteristic of a coronavirus membrane protein. Firstly, there are three potential N-linked glycosylation sites (Fig. 1, amino acid positions 5, 190, and 214), one of which is near the amino-terminus. Secondly, the polypeptide

1	CATAGACCC [.] TTTCCCTAAACGAGTTATTGATTTCT [.] <u>TAA</u> ACTAAACGACAATGTCAAATGAC [.]	60
		M S N D
		4
61	AATTGTACGGGTGACATTGTCACCCATTGAAGAATGGAAATTTGGTTGGAATGTTATT [.]	120
5	N C T G D I V T H L K N W N F G W N V I	24
	●	
121	CTAACCATATTCATTGTTATTCTTCAGTTGGACACTATAAATACTCCAGATTGTTTTAT [.]	180
25	L T I F I V I L Q F G H Y K Y S R L F Y	44
181	GGTTTGAAGATGCTGTACTGTGGCTTCTTTGGCCACTCGTACTTGCTTTGTCAATCTTT [.]	240
45	G L K M L V L W L L W P L V L A L S I F	64
241	GACACCTGGGCTAATTGGGATTCTAATTGGGCCTTTGTTGCATTTAGCTTTTTTATGGCC [.]	300
65	D T W A N W D S N W A F V A F S F F M A	84
301	GTATCAACACTCGTTATGTGGGTGATGTACTTCGCAAACAGTTTCAGACTTTCCGACGT [.]	360
85	V S T L V M W V M Y F A N S F R L F R R	104
361	GCTCGAACTTTTTGGGCATGGAATCCTGAGGTTAATGCAATCACGTACACAACCGTGTG [.]	420
105	A R T F W A W N P E V N A I T V T T V L	124
421	GGACAGACATACTATCAACCCATTCAACAAGCTCCAACAGGCATTACTGTGACCTTGCTG [.]	480
125	G Q T Y Y Q P I Q Q A P T G I T V T L L	144
481	AGCGGCGTGCTTTACGTTGACGGACATAGATTGGCTTCAGGTGTTTCAGGTTCCATAACCTA [.]	540
145	S G V L Y V D G H R L A S G V Q V H N L	164
541	CCTGAATACATGACAGTTGCCGTGCCGAGCACTACTATAATTTATAGTAGAGTCGGAAGG [.]	600
165	P E Y M T V A V P S T T I I Y S R V G R	184
601	TCCGTAATTCACAAAATTGCACAGGCTGGGTTTTCTACGTACGAGTAAAACACGGTGAT [.]	660
185	S V N S Q N C T G W V F Y V R V K H G D	204
	●	
661	TTTTCTGCAGTGAGCTCTCCCATGAGCAACATGACAGAAAACGAAAGATTGCTTCATTTT [.]	720
205	F S A V S S P M S N M T E N E R L L H F	224
	●	
721	<u>TTCTAAACTGAACGAAAAGAT</u> GGCTAC 747	
225	F *	

Fig. 1. Nucleotide sequence of the HCV 229 E membrane protein gene. The numbering of the nucleotide sequence is arbitrary. The predicted amino acid sequence of the membrane protein is shown in the single letter code and the position of three potential N-glycosylation sites are marked (●). The intergenic sequence TCTAAAC is overlined. The positions of the nucleocapsid gene initiation codon and the 5' upstream ORF termination signal are boxed

displays three internal hydrophobic domains (Fig. 1, amino acid positions 17–37, 48–63, and 75–95) within the amino terminal half and a relatively hydrophilic carboxy-terminus (Fig. 1, amino acid 213–221). Thirdly, the polypeptide is slightly basic with a net charge of + 4 at neutral pH.

A comparison of the amino acid sequences of the M proteins of HCV 229 E, TGEV, MHV, BCV, and IBV (Fig. 2) confirms that the HCV protein has a

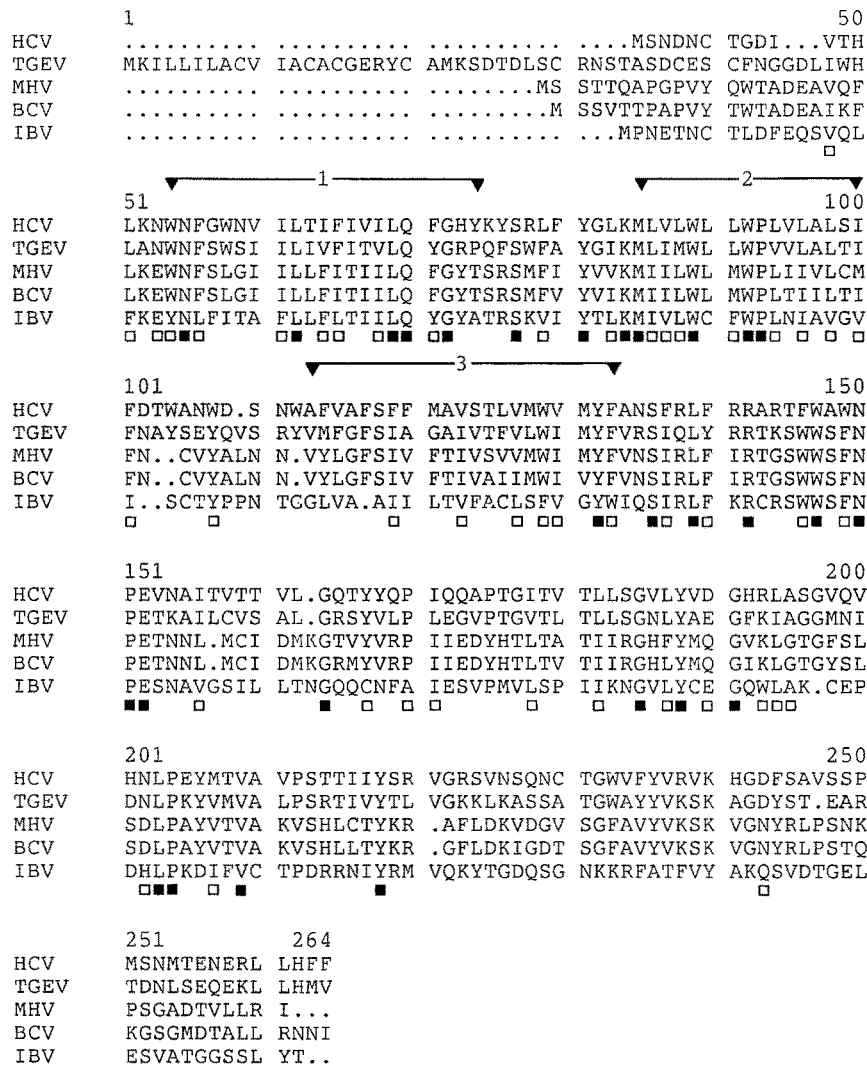


Fig. 2. Sequence similarity of the HCV 229 E, TGEV, MHV, BCV, and IBV M proteins. The sequences were aligned and percentage similarities determined using the program GAP of the UWGCG sequence analysis software. The positions of the hydrophobic transmembrane domains 1, 2, and 3 are overlined. Positions with identical amino acids are indicated (■) as well as those which are designated as similar by the UWGCG program SIMPLIFY (□). The numbering of the amino acids is arbitrary. The M protein sequences are taken from this paper and references [1, 6, 8, 10]

high sequence similarity to the other coronavirus proteins (HCV/TGEV 68%, HCV/MHV 59%, HCV/BCV 57%, and HCV/IBV 52%). Also, from this comparison it is evident that the HCV 229 E M protein, in contrast to the TGEV protein, does not possess a putative N-terminal signal sequence and, as has been noted previously [6], all coronavirus M proteins, including that of HCV 229 E, display a structurally similar central domain (Fig. 2, amino acid positions 127–152).

On the basis of structural and biochemical data Rottier and coworkers [12–14] have proposed a model for the membrane topology of the MHV M protein. In this model a short glycosylated region of the amino terminus is on the outside of the virion. The protein then enters and traverses the virion membrane three times (corresponding to the hydrophobic regions 1, 2, and 3) before emerging from the cytoplasmic face of the lipid bilayer. Basic domains in the carboxy terminal region of the protein are then proposed to interact with the nucleocapsid structure during virus maturation. The data presented here are fully consistent with this model.

Acknowledgments

This work was financed by the Deutsche Forschungsgemeinschaft (SFB 165/B 1). We would like to thank Barbara Schelle-Prinz for excellent technical assistance and Helga Kriesinger for typing the manuscript. This sequence has been deposited with the EMBL data library under the accession number X15498 Membrane Protein.

References

1. Bournsnel MEG, Brown TDK, Binns MM (1984) Sequence of the membrane protein gene from avian coronavirus IBV. *Virus Res* 1: 303–313
2. Devereux J, Haeblerli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 12: 387–395
3. Gubler U, Hoffman BJ (1983) A simple and very efficient method for generating cDNA libraries. *Gene* 25: 263–269
4. Hierholzer JC, Tannock GA (1988) Coronaviridae: the coronaviruses. In: Lenette EH, Halonen P, Murphy FA (eds) *Viral, rickettsial, and chlamydial diseases*. Springer, Berlin Heidelberg New York Tokyo, pp 451–483 (*Laboratory diagnosis of infectious diseases – principles and practice*, vol 2)
5. Isaacs D, Flowers D, Clarke JR, Valman B, Macnaughton MR (1983) Epidemiology of coronavirus respiratory infections. *Arch Dis Child* 38: 500–503
6. Kapke PA, Tung FYT, Hogue BG, Brian DA, Woods RG, Wesley R (1988) The aminoterminal signal peptide on the porcine transmissible gastroenteritis coronavirus matrix protein is not an absolute requirement for membrane translocation and glycosylation. *Virology* 165: 367–376
7. Kemp MC, Hierholzer JC, Harrison A, Burks JS (1984) Characterization of viral proteins synthesized in 229 E infected cells and effect(s) of inhibition of glycosylation and glycoprotein transport. In: Rottier PJM, van der Zeijst BAM, Spaan WJM, Horzinek MC (eds) *Molecular biology and pathogenesis of coronavirus*. Plenum Press, New York, pp 65–79 (*Advances in experimental biology and medicine*, vol 173)
8. Lapps W, Hogue BG, Brian D (1987) Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. *Virology* 157: 47–57
9. Messing J, Vierira J (1982) A new pair of M 13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* 19: 269–276
10. Pfeleiderer M, Skinner MA, Siddell SG (1986) Coronavirus MHV-JHM: Nucleotide sequence of the mRNA that encodes the membrane protein. *Nucleic Acids Res* 14: 6338
11. Phillpots JR (1983) Clones of MRC-C cells may be superior to the parent line for the culture of 229 E-like strains of human respiratory coronaviruses. *J Virol Methods*: 6: 267–269

12. Rottier P, Armstrong J, Meyer DI (1985) Signal recognition particle-dependent insertion of coronavirus E 1, an intracellular membrane glycoprotein. *J Biol Chem* 260: 4648–4652
13. Rottier PJM, Rose JK (1987) Coronavirus E 1 glycoprotein expressed from cloned cDNA localizes in the Golgi region. *J Virol* 61: 2042–2045
14. Rottier PJM, Welling GW, Welling-Wester S, Niesters HGM, Lenstra JA, van der Zeijst BAM (1986) Predicted membrane topology of the coronavirus protein E1. *Biochemistry* 25: 1335–1339
15. Sanger F, Nicklin S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463–5467
16. Schmidt OW, Kenny GE (1982) Polypeptides and functions of antigens from human coronaviruses 229 E and OC 43. *Infect Immun* 35: 515–522
17. Schreiber SS, Kamahora T, Lai MMC (1989) Sequence analysis of the nucleocapsid protein gene of human coronavirus 229 E. *Virology* 169: 142–151
18. Siddell S (1983) Coronavirus JHM: coding assignments of subgenomic mRNAs. *J Gen Virol* 64: 113–125
19. Spaan W, Cavanagh D, Horzinek MC (1988) Coronaviruses: structure and genome expression. *J Gen Virol* 69: 2939–2952
20. Staden R (1982) Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Res* 10: 4731–4751
21. Wege H, Siddell S, ter Meulen V (1982) The biology and pathogenesis of coronaviruses. *Curr Top Microbiol Immunol* 99: 165–200
22. Weiss SR, Leibowitz JL (1981) Comparison of the RNAs of murine and human coronaviruses. In: ter Meulen V, Siddell S, Wege H (eds) *Biochemistry and biology of coronaviruses*. Plenum Press, New York, pp 43–69 (*Advances in experimental biology and medicine*, vol 142)

Authors' address: Dr. S. G. Siddell, Institut für Virologie, Universität Würzburg, Versbacher Strasse 7, D-8700 Würzburg, Federal Republic of Germany.

Received June 15, 1989