Sequence Analysis of the Membrane Protein Gene of Human Coronavirus 229E

PATRICIA JOUVENNE,* CHRISTOPHER D. RICHARDSON,† STEVEN S. SCHREIBER,‡ MICHAEL M. C. LAI,‡ AND PIERRE J. TALBOT* 1

*Institut Armand-Frappier, Université du Québec, Virology Research Center, Laval, Québec, Canada H7N 4Z3; †Biotechnology Research Center, National Research Council of Canada, Montréal, Québec, Canada H4P 2R2; and ‡University of Southern California, School of Medicine, Departments of Neurology and Microbiology, Los Angeles, California 90033

Received June 22, 1989; accepted October 18, 1989

Human coronaviruses (HCV) belong to either one of two antigenic groups, represented by the prototype strains 229E and OC43 (7). They are responsible for as much as 25% of common colds (2, 3) and have been associated with gastrointestinal disorders (4). Their possible involvement in neurological diseases was suggested by the observation of coronavirus-like particles in the brain of one multiple sclerosis (MS) patient (5), the isolation of coronaviruses from two MS brain tissues passed in mice (6), and the detection of intrathecal antibodies to HCV-OC43 and HCV-229E in MS patients (7). However, the association of human coronaviruses with neurological diseases has not yet been confirmed.

HCV-229E possesses a single-stranded, positive-sense RNA genome with a molecular weight of 5.8 x 10^6 and a poly(A) tail of about 70 nucleotides at the 3' end (8). As with other coronaviruses, six subgenomic RNAs are synthesized in infected cells (9). These appear to have lower molecular weights than viral RNAs synthesized in cells infected with murine hepatitis virus (MHV). At least four polypeptides have been found in purified HCV-229E virions: 160- to 200-kDa and 88- to 105-kDa glycoproteins which may be analogous to the spike glycoprotein S (previously designated E2) of MHV (10); a 47- to 53-kDa polypeptide corresponding to the nucleocapsid protein N and a 17- to 26-kDa M protein (previously designated E1) observed in both glycosylated and nonglycosylated forms (11–14). One author also reported glycoproteins of 31 and 65 kDa (11).

The nucleotide sequence of the genes encoding the nucleocapsid proteins as well as the mRNA leader sequences of HCV-229E and HCV-OC43 have recently been determined (15, 16). As a continuation of these studies, we report the nucleotide sequence of the gene encoding the membrane protein M of HCV-229E. Its predicted amino acid sequence is compared with sequences determined for other coronaviruses.

Clones containing the sequence of the M protein gene were obtained from a cDNA library constructed with mRNA isolated from HCV-229E-infected L132 cells, and identified using a genome-specific probe (15). One clone, designated L8, was selected for sequencing since it contained a large 3.6-kb insert overlapping by 1.2 kb the 3' end of the N protein gene. The remaining 2.4-kb fragment was excised from an internal PstI site of clone L8 and subcloned into the pBlueScript II vector (Stratagene). Unidirectional deletions of the 2.4-kb insert were created using exonuclease III, mung bean nuclelease, and deoxythionucleotide derivatives (Stratagene). The sequencing of both strands was

1 To whom requests for reprints should be addressed.
performed by the plasmid sequencing technique (17), using T7 DNA polymerase. In vitro translation of poly(A) mRNAs isolated from HCV-229E-infected L132 cells was carried out in order to determine the molecular mass of the unprocessed viral polypeptides.

The complete nucleotide sequence of the M protein gene of HCV-229E and its predicted amino acid sequence are presented in Fig. 1. The AUG codon is preceded by the consensus intergenic sequence UCUAAACU, which is identical to that upstream of the nucleocapsid protein-coding sequence (15; and Fig. 1). This sequence is conserved among coronaviruses of various species and represents the binding site of the leader RNA which mediates a discontinuous transcription of mRNAs (18). The longest open reading frame extends from base 171 through base 848 and encodes a 225-amino acid polypeptide with a calculated molecular weight of 25,822. The products of in vitro transla-
immunoprecipitation of in vitro translation products from HCV-229E mRNAs. Poly(A)' mRNAs were translated in the presence of [35S]methionine, using a rabbit reticulocyte lysate (Promega Bio- tec). The viral polypeptides were immunoprecipitated and separated by SDS–PAGE (13% acrylamide). Lane 1, molecular mass standards; lane 2, mRNAs from HCV-229E-infected cells; lane 3, mRNAs from noninfected cells; lane 4, translation without exogenous mRNA. Molecular mass standards (kDa) are indicated on the left. The calculated molecular masses of relevant viral proteins (kDa) are indicated on the right.

Fig. 2. Immunoprecipitation of in vitro translation products from HCV-229E mRNAs. Poly(A)' mRNAs were translated in the presence of [35S]methionine, using a rabbit reticulocyte lysate (Promega Bio- tec). The viral polypeptides were immunoprecipitated and separated by SDS–PAGE (13% acrylamide). Lane 1, molecular mass standards; lane 2, mRNAs from HCV-229E-infected cells; lane 3, mRNAs from noninfected cells; lane 4, translation without exogenous mRNA. Molecular mass standards (kDa) are indicated on the left. The calculated molecular masses of relevant viral proteins (kDa) are indicated on the right.

Like TGEV (20), there are three amino acid sequences characteristic of N-glycosylation sites in the predicted M protein sequence (Asn-5; Asn-190; and Asn-214), although only one (Asn-5) is found near the N-terminus, as compared to two for TGEV. Moreover, three potential O-glycosylation sites are located in the putatively external N-terminus of the polypeptide (Ser-2; Thr-7; and Thr-12). In addition, there is only one cysteine residue (Cys-5). Other coronavirus M proteins contain two (bovine coronavirus, BCV; Ref. (21)), four (MHV-A59 and JHM; Refs. (19) and (22), respectively), eight (TGEV; Ref. (20)), or nine (infectious bronchitis virus, IBV; Ref. (23)) cysteine residues. This cysteine residue is probably important in forming interchain disulfide bridges, since M of HCV-229E has been shown to form oligomers under nonreducing conditions (14).

No significant nucleotide sequence homology exists between the M genes of HCV-229E and other coronaviruses. The highest M amino acid homology (38% or 100 of 262 residues) occurs between HCV-229E and TGEV, which was reported to be antigenically related (24). Antigenically distinct BCV, MHV, and IBV show amino acid homologies of 32, 30, and 28%, respectively. In contrast, a homology of 87% was found between the M proteins of BCV and MHV-A59 (21), which belong to another antigenic subgroup (24). On the other hand, a homology of 34% was found between the M protein of TGEV and BCV (25), which belong to two different antigenic subgroups. Figure 3 illustrates the M regions common to both HCV-229E and TGEV.

As with other coronaviruses, the M protein of HCV-229E is a highly hydrophobic membrane protein. It contains 51% hydrophobic residues, compared to 45–51% for other coronaviruses (19–23, 25). The hydrophaticity profiles of M proteins from HCV-229E, TGEV, BCV, MHV-JHM, MHV-A59, and IBV are presented in Fig. 4. The main features characterizing these M proteins include three large hydrophobic domains alternating with short hydrophilic regions. This suggests a selective pressure to maintain the potential transmembranous domains of this coronavirus protein. As with other coronaviruses (26), only about 10% (20 amino acids) of the HCV-229E M protein constitutes the hydrophilic putative external domain. On the other hand,
The large N-terminal putative signal sequence found only in the M protein of TGEV (20, 25) is not observed in HCV-229E, which is similar to the structure reported for BCV, MHV-JHM, MHV-A59, and IBV.

The coronavirus M protein is important for several reasons. This membrane protein is implicated in virus assembly and is believed to integrate the viral proteins prior to budding, most likely because of this protein's affinity for RNA (26). Moreover, some monoclonal antibodies against the M protein of MHV-JHM are protective in vivo and thus may influence the outcome of disease (27). We are currently pursuing the cloning and sequencing of other genes of HCV-229E, with emphasis on the gene coding for the spike protein S, which is potentially important in viral pathogenicity. The availability of molecular probes for human coronavirus genes opens new avenues for the verification of the potential involvement of these viruses in neurological diseases.

Fig. 3. Comparison of the predicted amino acid sequences of M proteins of HCV-229E (top row) and TGEV (bottom row) aligned for maximum homology. Regions common to both proteins are underlined. The analysis was performed on an Apple Macintosh Plus computer with the MacGene Plus program (Applied Genetic Technology Inc., Fairview Park, OH).
ACKNOWLEDGMENTS

We thank François Fossiez for helpful discussions and Lucie Summerside for typing the manuscript. This work was supported by Grant MT-9203 from the Medical Research Council of Canada to P. J. Talbot, and U.S. Public Health Services Research Grant NS18146 to M. M. C. Lai. P. J. Talbot is also the recipient of a University Research Scholarship from the Natural Sciences and Engineering Research Council of Canada. P. Jouvenne acknowledges a studentship support from the Fonds de la recherche en santé du Québec. S. S. Schreiber was supported by a postdoctoral training fellowship from the U.S. National Institutes of Health Grant NS07149.

REFERENCES