Virus Research, 22 (1992) 125–141

© 1992 Elsevier Science Publishers B.V. All rights reserved 0168-1702/92/$05.00

VIRUS 00728

Sequence analysis of human coronavirus 229E mRNAs 4 and 5: evidence for polymorphism and homology with myelin basic protein

Patricia Jouvenne [1], Samir Mounir [1], Janet N. Stewart [1], Christopher D. Richardson [2] and Pierre J. Talbot [1]

[1] *Institut Armand-Frappier, Université du Québec, Virology Research Center, Laval, Québec, Canada and* [2] *Biotechnology Research Center, National Research Council of Canada, Montréal, Québec, Canada*

(Received 1 May 1991; revision received and accepted 31 October 1991)

Summary

Human coronaviruses (HCV) are important pathogens responsible for respiratory, gastrointestinal and possibly neurological disorders. To better understand the molecular biology of the prototype HCV-229E strain, the nucleotide sequence of the 5'-unique regions of mRNAs 4 and 5 were determined from cloned cDNAs. Sequence analysis of the cDNAs synthesized from mRNA 4 revealed a major difference with previously published results. However, polymerase chain reaction amplification of this region showed that the sequenced cDNAs were produced from minor RNA species, an indication of possible genetic polymorphism in this region of the viral genome. The mutated messenger RNA 4 contains two ORFs: (1) ORF4a consisting of 132 nucleotides which potentially encodes a 44-amino acid polypeptide of 4653 Da; this coding sequence is preceded by a consensus transcriptional initiation sequence, CUAAACU, similar to the ones found upstream of the N and M genes; (2) ORF4b of 249 nucleotides potentially encoding an 83-amino acid basic and leucine-rich polypeptide of 9550 Da. On the other hand, mRNA 5 contains one single ORF of 231 nucleotides which could encode a 77-amino acid basic and leucine-rich polypeptide of 9046 Da. This putative protein presents a significant degree of amino acid homology (33%) with its counterpart found in transmissible gastroenteritis coronavirus (TGEV). The proteins in the two different viruses exhibit similar molecular weights and are extremely hydrophobic.

Correspondence to: P.J. Talbot, Centre de Recherche en Virologie, Institut Armand-Frappier, 531 boul. des Prairies, Laval, Qué., Canada H7N 4Z3.

Interestingly, a sequence homology of five amino acids was found between the protein encoded by ORF4b of HCV-229E and an immunologically important region of human myelin basic protein.

Coronavirus; Human; 229E; Myelin basic protein; Polymorphism; mRNA 4; mRNA 5; Nucleotide sequence

Introduction

The interest in the molecular biology of human coronaviruses (HCV) has increased considerably in the last few years. This observation is not surprising since they are important human pathogens responsible for up to 25% of common colds (McIntosh et al., 1974; Wege et al., 1982), some gastrointestinal infections (Resta et al., 1985), and possibly neurological disorders such as multiple sclerosis (MS) or Parkinson's disease (Fishman et al., 1985). Their potential involvement in multiple sclerosis was suggested by the observation of coronavirus-like particles in the brain of one MS patient (Tanaka et al., 1976), the isolation of coronaviruses from two MS brain tissues passaged in mice (Burks et al., 1980), the detection of intrathecal antibodies to HCV-OC43 and HCV-229E in MS patients (Salmi et al., 1982), and the preferential detection of coronavirus RNA in central nervous system (CNS) tissue from MS patients (Murray et al., 1990). Moreover, several excellent indications suggest that MS could be the consequence of a virus-induced autoimmune disease. One of the possible mechanisms involved could be a molecular mimicry resulting from a sequence homology between a viral protein and myelin basic protein (Watanabe et al., 1983; Jahnke et al., 1985; Oldstone, 1987). However, since appropriate diagnostic tools have been lacking, the association of human coronaviruses with neurological disorders has not yet been confirmed. With this objective in mind, we have initiated studies on the molecular biology of the prototype 229E strain of HCV (HCV-229E).

This virus possesses a single-stranded, positive-sense RNA genome with a molecular weight of approximately 6×10^6 (Hierholzer et al., 1981). Six subgenomic RNA species are synthesized in infected cells and appear to have lower molecular weights than their counterparts from murine hepatitis virus (MHV) (Weiss and Leibowitz, 1981). Northern blot analysis has confirmed that, like other coronaviruses, they constitute a nested set of 3′-coterminal mRNA species (Schreiber et al., 1989) of which presumably only the 5′-unique regions are translated (reviewed in Spaan et al., 1988). 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) of MHV (Sturman et al., 1985); a 47- to 53-kDa polypeptide corresponding to the nucleocapsid protein (N), and a 17- to 26-kDa membrane protein (M) which is found in both glycosylated and non-glycosylated forms (Hierholzer, 1976; Macnaughton, 1980; Schmidt and Kenny, 1982; Arpin and Talbot, 1990). Another author also reported glycopro-

teins of 31 and 65 kDa (Hierholzer, 1976). Non-structural proteins have not been described, although in vitro translation of viral mRNAs yielded potential non-structural polypeptides of 42, 28.5 and 14 kDa (Jouvenne et al., 1990), which could also be observed in infected cells (Talbot et al., unpublished results).

The nucleotide sequence of the genes encoding the N and M proteins of HCV-229E have previously been reported (Schreiber et al., 1989; Raabe and Siddell, 1989a; Jouvenne et al., 1990). Moreover, the nucleotide sequence of the unique regions of mRNAs 4 and 5 has been determined in one other laboratory (Raabe and Siddell, 1989b). According to the latter report, mRNA 4 contains one ORF of 399 nucleotides, while mRNA 5 contains two ORFs (ORF5A and ORF5B) of 264 and 231 nucleotides, respectively. However, these authors later corrected these mRNA assignments and found that mRNA 4, not 5, contained two ORFs: 4a and 4b (Raabe et al., 1990). In the present study, we report nucleotide sequence data from mRNAs 4 and 5 of HCV-229E, and observe a major difference with previously published data. The predicted amino acid sequences of the encoded polypeptides are compared with sequences known for other coronaviruses, as well as that of human myelin basic protein. Based on polymerase chain reaction analysis, we have attempted to understand the divergence of the data obtained by our two groups. We show evidence for the presence of a minor mRNA 4 species containing a large deletion and two smaller ORFs.

Materials and Methods

Cells and virus

The human embryonic lung cell line L132 (Davis and Bolin, 1960) and the HCV-229E inoculum were obtained from the American Type Culture Collection (Rockville, MD). Cells were grown at 37°C in Earle's minimal essential medium: Hank's M199 (1 : 1, v/v) supplemented with 10% (v/v) fetal bovine serum (FBS), 0.13% (w/v) sodium bicarbonate and 50 μg/ml Gentamycin (Gibco Canada, Burlington, Ont., Canada). The virus was plaque-purified twice and quantitated by plaque assay as previously described (Daniel and Talbot, 1987) except that plaques were revealed after 7 days. Three passages on L132 cells at a MOI of 0.001 were performed to yield a viral stock with a titer of 7×10^5 PFU/ml. Viral infections were performed at 33°C in medium which contained FBS reduced to a level of 2% (v/v).

Preparation of intracellular RNA

In order to optimize the yield of viral mRNA, we established the kinetics for synthesis of HCV-229E intracellular RNA in L132 cells and found a peak of [^{3}H]uridine incorporation at 20 h post-infection at an MOI of 0.001. Thus, intracellular RNA from uninfected cells or cells infected 20 h previously with HCV-229E was extracted according to Favaloro et al. (1980). Briefly, cells were

lysed with a Dounce homogenizer, the cytoplasmic phase was isolated and deproteinized with 200 μg/ml proteinase K and the preparations were treated with DNase I (0.1 μg/mg of nucleic acid).

cDNA synthesis and cloning

cDNA synthesis was carried out using a cDNA synthesis kit (Pharmacia, Dorval, Canada) and a synthetic oligonucleotide 5′-TGGTACAATGTCACCCGTAC-3′, complementary to nucleotides 187 to 207 at the 5′-end of the M gene (Jouvenne et al., 1990), was used as primer. *Eco*RI adapters were added to blunt-ended, double-stranded cDNAs, followed by ligation to *Eco*RI-cut, dephosphorylated pBluescript II vector (Stratagene, La Jolla, CA). *E. coli* XL-1 transformants containing HCV-229E cDNAs were identified by colony hybridization (Grunstein and Hogness, 1975) with the 5′-end-radiolabeled oligonucleotide.

cDNA sequencing and sequence analysis

Stepwise unidirectional deletions at both ends of the largest cDNA clone were created with exonuclease III, mung bean nuclease and deoxythionucleotide derivatives (Stratagene). The sequencing of both strands was performed by the plasmid sequencing technique (Hattori and Sakaki, 1986), with T7 DNA polymerase (Pharmacia). In order to confirm the nucleotide sequence, nine other clones of decreasing sizes were partially sequenced at their 5′-ends. Therefore, each nucleotide in the reported sequence is the result of three separate sequencing reactions. Sequence analyses were performed on an Apple Macintosh Plus computer with the MacGene Plus program (Applied Genetic Technology Inc., Fairview Park, OH), except for potential phosphorylation sites, which were identified with the PC/GENE program (Intelligenetics, Inc., Mountain View, CA). Potential leucine zippers and N-glycosylation sites were identified manually and confirmed with the PC/GENE program. The analysis of the RNA secondary structure was performed with the Fold prediction program (Zuker and Stiegler, 1981) contained in the Sequence Analysis Software Package of the University of Wisconsin Genetics Computer Group (Devereux et al., 1984), accessed through the CAN/SND Molecular Biology Database System (Ottawa, Canada).

Polymerase chain reaction and DNA sequencing of the amplified product

The polymerase chain reaction was performed by a modification of the original procedure (Saiki et al., 1988), which will be described elsewhere. Oligonucleotides used for amplification were synthesized on an Applied Biosystems synthesizer, with the following sequences: 5′-CTATTCCAACAGCTGGGTGTTCAC-3′ (antisense; Fig. 1, bases 396–419), 5′-AAGATCACACCGTGGCAGAGCTGC-3′ (sense; Fig. 1, bases 238–261). The amplified DNA fragment was ligated into the M13 mp18 vector and sequenced by the dideoxy chain termination method (Sanger et al., 1977), using Sequenase™ (United States Biochemical Corp., Cleveland,

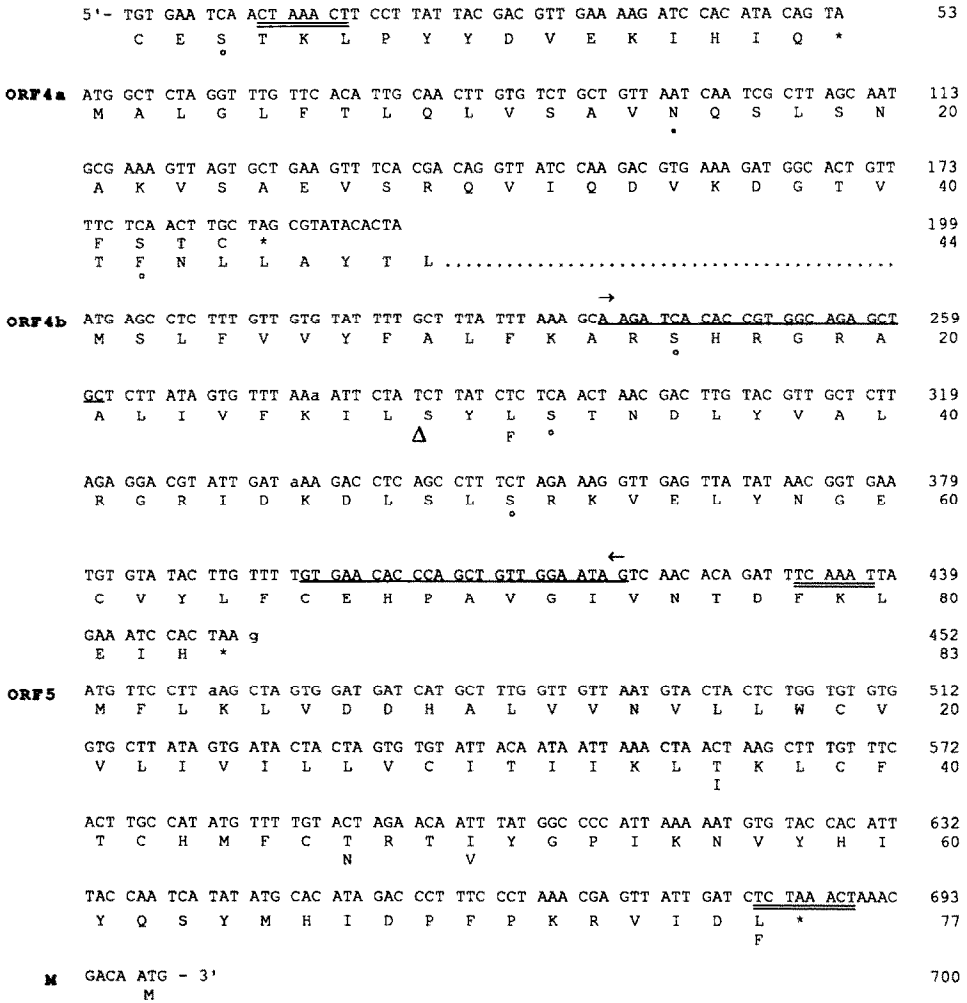


Fig. 1. Nucleotide sequence of the unique regions of mRNAs 4 and 5, as well as the predicted amino sequences of the encoded polypeptides. The intergenic sequences are doubly underlined and termination codons indicated (*). The potential N-glycosylation (•) and phosphorylation (°) sites are also indicated. The triangle shows the site of insertion of the additional 259 nucleotides reported by Raabe and Siddell (1989b) and the differences in the amino acids from mRNAs 4 and 5 published by these authors appear on the bottom row. The dots indicate that amino acid Leu-49 is followed by Met-1 in the latter sequence. The right and left-pointing arrows indicate the region selected for PCR amplification, using antisense (←) and sense (→) oligonucleotides corresponding to the underlined sequences.

OH) and [^{35}S]dATP (Amersham Canada, Oakville, Ont., Canada). In some experiments, 2 μl (6.6 pmol) of [α-^{32}P]dCTP (spec. act. 3000 Ci/mmol; ICN Biomedicals Canada Ltd., Mississauga, Ont., Canada) was used in the amplification reaction, the amplified products were separated on agarose gels, which were then washed twice in water for 15 min each time and treated with 10% (w/v) trichloroacetic

acid for 15 min at room temperature and washed another two times in water before exposure to X-ray film.

Results

A cDNA library was prepared from HCV-229E intracellular RNA, using a specific oligonucleotide as primer. One clone, designated J22, contained a 1.7 kb insert that hybridized to viral mRNAs 1 through 6 in Northern blots (data not shown). The size of the RNAs, as determined with DNA markers, was similar to the ones obtained by Weiss and Leibowitz (1981). The J22 clone was selected for sequencing following creation of unidirectional deletions at both ends of the viral insert. Furthermore, the 5′-ends of nine other cDNA clones ranging in size from 0.4 to 1.5 kb were sequenced. The nucleotide sequence of the unique regions of mRNAs 4 and 5, as well as the predicted amino acid sequence of the encoded polypeptides are presented in Fig. 1. Northern blot analysis was used to confirm the proposed mRNA assignments (data not shown).

As shown in Fig. 1, messenger RNA 4 contains two ORFs: ORF4a and ORF4b. ORF4a extends from base 54 through base 185 and encodes a putative 44-amino acid polypeptide of 4653 Da. This ORF4a is preceded by a transcriptional initiation sequence, CUAAACU, which is located inside the 3′-end of the S gene. This sequence is similar to the consensus intergenic sequence found upstream of the N and M genes (Table 1). There is one potential N-glycosylation site in this predicted protein (Asn-15), as well as one potential phosphorylation site (casein

TABLE 1

Characteristics of the HCV-229E 3′-open reading frames

RNA species	Size [a]			Intergenic sequence	Distance from the initiation codon [a]	Adjacent ORF	Predicted polypeptide	
	mRNA [b]	Predicted unique region	ORF				Size [c]	Molecular weight
4	3700	700	132	CUAAACU	36	ORF4a	44 (133) [d]	4,653 (15,300) [d]
			249	absent		ORF4b	83 (88) [d]	9,550 (10,200) [d]
5	3000	400	231	UCAAAU	15	ORF5	77	9,046
6	2600	800	675 [e]	UCUAAACU [e]	8	M	225 [e]	25,822 [e]
7	1800	1800	1167 [f]	UCUAAACU [f]	10	N	389 [f]	43,366 [f]

[a] In bases.
[b] Determined from Northern blots.
[c] Number of amino acids.
[d] Data from Raabe and Siddell, 1989b; with mRNA assignment from Raabe et al., 1990.
[e] Data from Jouvenne et al., 1990.
[f] Data from Schreiber et al., 1989.

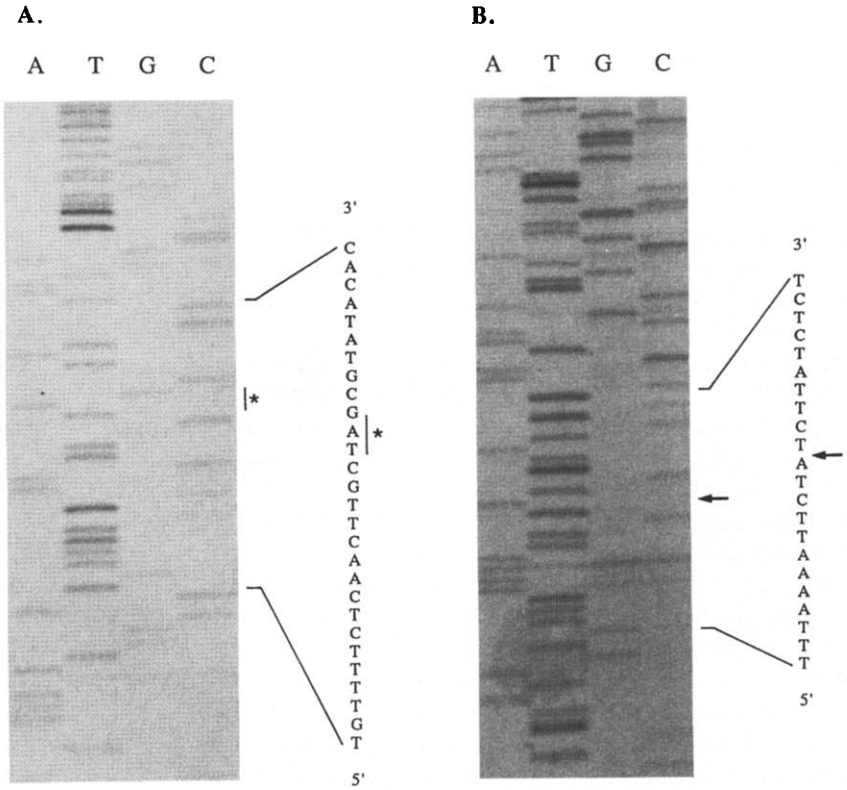


Fig. 2. Nucleotide sequence analysis from two regions of a sequencing gel corresponding to the termination codon of ORF4a (asterisk in A) and the site of an apparent deletion from the published sequence (arrows in B). The HCV-229E sequences shown correspond to nucleotides 171 to 197, located at the 3′-end of ORF4a (A), and 272 to 293, located within ORF 4b (B).

kinase II site: Ser-42). ORF4b extends from base 200 through base 448 and encodes a putative 83-amino acid polypeptide with a calculated molecular mass of 9550 Da. There are three potential phosphorylation sites in this predicted protein (protein kinase C sites: Ser-15, Ser-51; casein kinase II site: Ser-32). The ORF4b protein contains a high proportion of leucine and isoleucine (20%) as well as basic (16%) residues. No consensus intergenic sequence was found upstream of this ORF. A notable difference with a previously published sequence from this region of the genome (Raabe and Siddell, 1989b) is the absence in our sequence of 259 nucleotides, which code for an additional 56 and 33 amino acids in their ORF4a and ORF4b (Raabe et al., 1990), respectively. The site of this apparent deletion is indicated by a triangle in Fig. 1. Furthermore, a termination codon in our sequence (bases 186–188) implies that five amino acids in the published ORF4 are missing from our sequence, in the region between ORFs 4a and 4b. The actual sequencing gel data from these two regions of apparent divergence is shown in Fig. 2: panel A

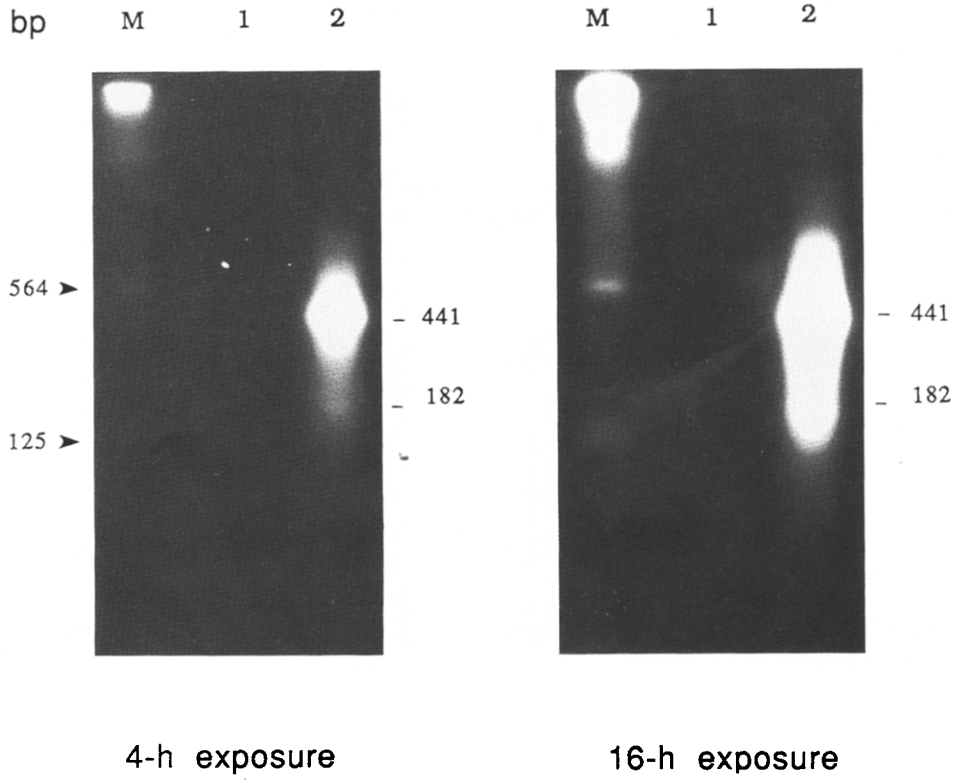


Fig. 3. Agarose gel analysis of the radiolabeled amplification products of mRNA 4. Lane 1: L132 cells used to propagate HCV-229E; lane 2: HCV-229E. *Hin*dIII digested lambda DNA was used as the molecular size marker (M). The numbers on the right indicate the lengths of the amplified fragments and the numbers on the left indicate the relevant molecular size markers. The trichloroacetic acid-treated agarose gel was exposed to X-ray film for either four (left panel) or sixteen (right panel) hours.

shows the region of the ORF4a termination codon and panel B the region of the deletion.

In order to investigate further the significance of the apparent 259-nucleotide deletion in mRNA 4, we amplified a portion of this RNA by using specific oligonucleotides flanking the target sequences (indicated in Fig. 1) and the polymerase chain reaction (PCR). As shown in Fig. 3, a major 441 bp band corresponding to the size predicted from the sequence published by Raabe and Siddell (1989b) was obtained. However, a 182 bp signal was also faintly visible when a radioactive nucleotide was included in the PCR reaction. On the basis of our cDNA sequence, this smaller fragment may correspond to an mRNA lacking the 259 nucleotides, which was used as template for the two sequenced cDNA clones. Sequencing of the major 442 bp band confirmed the published sequence of this region, with the exception of a T to G substitution at position 328 (Raabe and Siddell, 1989b), which results in the replacement of a tyrosine residue by an

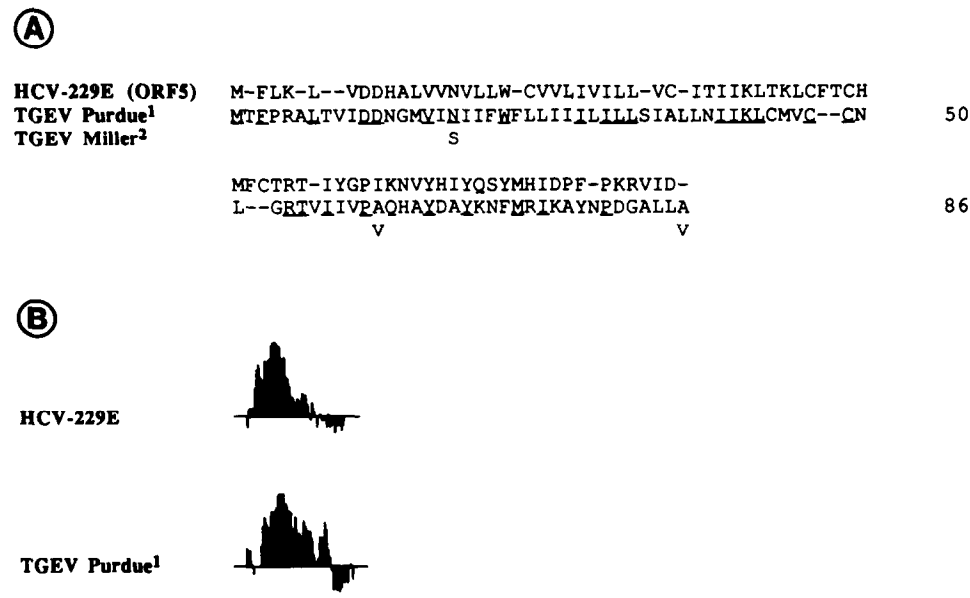
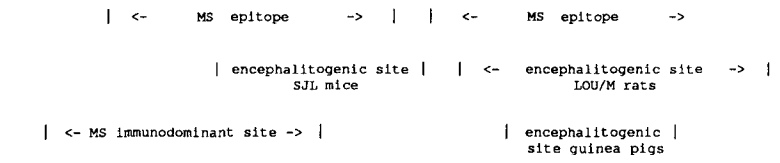


Fig. 4. (A) Comparison of the predicted amino acid sequences of mRNA 5 (ORF5) from HCV-229E (top row), TGEV Purdue ([1] Rasschaert et al., 1987; middle row) and TGEV Miller ([2] Wesley et al., 1989; bottom row) aligned for maximum homology. Regions common to the two proteins are underlined and differences with TGEV Miller are indicated on the bottom row. (B) Hydropathicity profiles of the protein encoded by mRNA 5 of HCV-229E and TGEV Purdue determined according to Kyte and Doolittle (1982). The analysis was performed on an Apple Macintosh Plus computer with the MacGene Plus program (Applied Genetic Technology Inc., Fairview Park, OH). Each point is the mean hydropathicity of a span of seven residues. Peaks extending upwards correspond to hydrophobic regions and peaks extending downwards to hydrophilic areas.

aspartic acid. Unfortunately, we were unable to recover sufficient amounts of the minor species to confirm that its sequence does correspond to the data obtained from cDNAs and presented in Fig. 1.

Messenger RNA 5 contains one single ORF of 231 nucleotides which extends from base 453 through base 683 and encodes a putative 77-amino acid polypeptide of 9046 Da containing a high proportion of leucine and isoleucine residues (27%). ORF5 is preceded by the sequence UCAAAU, which resembles the consensus intergenic sequence (Shieh et al., 1989) and is located at the 3'-end of ORF4b. The predicted amino acid sequence from ORF5 is identical to that of ORF5B (later renamed ORF5) from Raabe and Siddell (1989b) except that four amino acids, one of which is part of a single potential N-glycosylation site (residue 47), are different in our sequence. Fig. 4A presents a comparison of the predicted amino acid sequences of the mRNA 5 (ORF5) between HCV-229E, TGEV Purdue (Rasschaert et al., 1987) and TGEV Miller (Wesley et al., 1989). The ORF5 of HCV-229E exhibits 32.9% homology with TGEV Purdue and 31.7% homology with TGEV Miller. This predicted protein of HCV-229E has a similar molecular



MBP 82-DENPVVHFFKNIVTPRTPPPSQGKGRG **LSLSR** FSWGAEGQKPGFGYGGR-130

229E 21-ALIVFKILSYLSTNDLYVALRGRIDKD **LSLSR** KVELYNGECVYLFCEHPA-70

Fig. 5. Homology between myelin basic protein (MBP) and ORF4b of HCV-229E at the level of a five-amino acid sequence: LSLSR. The MBP immunodominant site recognized by T cells from MS patients, two epitopes recognized by T-cell clones established from MS patients, and sites demonstrating encephalitogenic potential in experimental animals are indicated.

weight to the one from TGEV (9,241; Rasschaert et al., 1987), is relatively basic (14%) and also exhibits an extremely hydrophobic profile (Fig. 4B).

No significant homologies were found between our predicted amino acid sequences of mRNAs 4 and 5 from HCV-229E and those derived from mRNA 4 of TGEV Purdue (previously designated 3: Rasschaert et al., 1987), or TGEV Miller (Wesley et al., 1989), mRNA 3 from IBV M41 (previously designated D: Niesters et al., 1986; new nomenclature from: Cavanagh et al., 1990), mRNA 5 from IBV Beaudette (previously designated B: Boursnell and Brown, 1984), mRNAs 4 and 5 from MHV-JHM (Skinner and Siddell, 1985; Skinner et al., 1985), or mRNA 5 from MHV-A59 (Budzilowicz and Weiss, 1987).

We have also compared the HCV-229E sequences with the human myelin basic protein (MBP; Roth et al., 1987) and found a five-amino acid homology between the protein encoded by ORF4b and MBP. This sequence is the following: LSLSR (residues 48 to 52 for HCV-229E and residues 109 to 113 for human MBP; Figs. 1 and 5). This sequence belongs to exon 5 of the human MBP gene (Roth et al., 1987) and is conserved among bovine, chimpanzee, guinea pig, murine, porcine, rabbit and rat MBPs (Martenson, 1983). As shown in Fig. 5, the LSLSR sequence is situated in a biologically relevant area of human MBP. It is contained within a site shown to be encephalitogenic for LOU/M rats (residues 109–128: Hashim et al., 1991), it overlaps by two amino acids an encephalitogenic site for guinea pigs (residues 112–124: Carnegie, 1971) and it is situated two amino acids downstream from an encephalitogenic site for SJL mice (residues 94–107: Fritz and McFarlin, 1989). Moreover, it is located eight amino acids downstream from an immunodominant MBP epitope recognized by the T cells of multiple sclerosis patients (residues 82–100: Ota et al., 1990) and is contained within another site recognized by some T cell clones established from MS patients (residues 108–148: Jingwu et al., 1990), as well as four amino acids downstream from another such epitope (residues 86–105: Richert et al., 1989). It is also present in the putative protein encoded by ORF5A (later renamed ORF4b) described by Raabe and Siddell (1989b). The LSLSR homology region was searched among the genes encoding non-structural proteins (except for the putative RNA polymerase) from the other

coronaviruses. The best homology found was the sequence LSLR belonging to ORF X2b of TGEV Purdue (Rasschaert et al., 1987), ORFB of TGEV Miller (Wesley et al., 1989) and ORF2 of TGEV FS772/70 (Britton et al., 1989), which correspond to ORF4b of HCV-229E at the level of the structural organization of the genome.

Discussion

A cDNA library was constructed from viral RNA extracted at the optimum time after infection of cultured cells, using an oligonucleotide primer complementary to a 5′-region of the previously published sequence of the M gene (Jouvenne et al., 1990).

The nucleotide sequences of the genes located between the S and M genes were obtained (Fig. 1). The apparently minor species (Fig. 3) of messenger RNA 4 contains two non-overlapping ORFs and messenger RNA 5 contains only one ORF. Surprisingly, 259 nucleotides from the sequence reported by Raabe and Siddell (1989b) are absent from our sequence, which was determined from two cDNA clones. The absence of these nucleotides in our sequence and the mRNA assignments suggested by our Northern blot analysis, which are consistent with data reported by Raabe et al. (1990), have major implications for the structural organization of this region of the genome. We observe two ORFs in mRNA 4 and only one ORF in mRNA 5, an organization analogous to TGEV Purdue (Rasschaert et al., 1987). The initial report by Raabe and Siddell (1989b) of one ORF in mRNA 4 and two ORFs in mRNA 5 is similar to what has been reported for MHV (Skinner and Siddell, 1985; Skinner et al., 1985; Budzilowicz and Weiss, 1987). Our observation of multiple ORFs on mRNA 4 of HCV-229E confirms a recent report (Raabe et al., 1990), and has also been reported for IBV, where three non-overlapping ORFs were described (Boursnell et al., 1985; Niesters et al., 1986). The unique ORF reported here for HCV-229E confirms the revised assignment reported by Raabe et al. (1990), and is also seen in TGEV (Rasschaert et al., 1987; Wesley et al., 1989).

Surprisingly, the 5′-end ORF of mRNA 4 of HCV-229E (ORF4a in our sequence) is shorter than the ORF4 (later renamed ORF4a) reported by Raabe and Siddell (1989b). The latter ORF terminates in the observed additional sequence within which their ORF5A (later renamed ORF4b) initiates. The 3′-end of their ORF5A (ORF 4b) is found in our ORF4b, of which the mRNA 4 assignment was confirmed by Northern blot analysis. These missing nucleotides in our sequence led us to predict the existence of two smaller proteins: 4653 Da for ORF4a and 9550 Da for ORF4b. This observation differs from the previous report, where ORF4 (ORF4a) and ORF5A (ORF4b) specify 15300 Da and 10200 Da polypeptides, respectively (Table 1). The two non-overlapping ORFs found in the unique region of mRNA 4 of TGEV Purdue encode putative 7.7 and 18.7 kDa polypeptides (Rasschaert et al., 1987) and these two ORFs are apparently present as two distinct RNA species in TGEV Miller, with the second ORF predicting a larger

27.7 kDa protein (Wesley et al., 1989). On the other hand, the protein encoded by ORF5 is very similar to the previously reported sequence (Raabe and Siddell, 1989b), with only four amino acids changes, one of which involves the absence of a potential N-glycosylation site in our sequence (amino acid 47).

The only significant homology between the predicted amino acid sequences encoded by mRNAs 4 and 5 of HCV-229E and other coronaviruses was found at the level of ORF5 which encodes a hydrophobic, basic and leucine-rich 9.0 kDa protein, which shows 32 to 33% homology with the 9.2 kDa protein encoded by mRNA 5 of the Miller and Purdue strains of TGEV, respectively (Wesley et al., 1989; Rasschaert et al., 1987) (Fig. 4). Despite a large proportion of leucine residues (14%) in the putative proteins encoded by mRNAs 4 and 5, no sequence motif consistent with a DNA-binding leucine zipper was identified in these regions (Abel and Maniatis, 1989). The significance of the potential phosphorylation sites identified in ORF4a (one site) and ORF4b (three sites) remains to be determined but could be involved in the functions of these potential non-structural proteins.

The extensive structural differences in the organization of HCV-229E mRNAs 4 and 5 deduced by a comparison of the sequence presented here and the one reported by Raabe and Siddell (1989b) suggest that the gene products of mRNA 4 of HCV-229E are not essential for virus replication in cell culture. Nevertheless, it remains possible that their functions are required for in vivo replication and pathogenesis. Our results extend to the human coronavirus previous observations on the non-essential nature for virus replication in transformed cells of non-structural proteins ns2, ns4 and ns5a and the HE structural protein of murine hepatitis virus (Schwarz et al., 1990; Yokomori and Lai, 1991; Yokomori et al., 1991). These observations are in contrast to the essential nature of non-structural proteins of other viruses, for example the ns1 and ns2 proteins of parvovirus H-1, which are required for viral DNA replication and efficient viral protein synthesis, respectively (Rhode, 1989; Li and Rhode, 1991). Interestingly, the putative non-structural protein encoded by ORF4b in TGEV was reported to have a predicted size of either 18.7 or 27.7 kDa, depending on the viral strain studied (Rasschaert et al., 1987; Wesley et al., 1989). Also, BCV mRNA 4 was shown to contain two open reading frames coding for proteins of 4.9 and 4.8 kDa, which appear to have arisen by a single base deletion from a single ORF encoding a 11 kDa protein, with significant homology with its counterpart in MHV (Abraham et al., 1990).

We attempted to localize putative nucleic acid binding regions within ORFs 4a, 4b and 5: no sequence compatible with either zinc fingers (Evans and Hollenberg, 1988) or leucine zippers were found, although this does not preclude their function in genome transcription and replication.

Finally, it is interesting to note that a consensus transcriptional initiation sequence is found upstream of mRNAs 4 and 5 (data summarized in Table 1). This is consistent with data accumulated so far for all coronaviruses (except for Raabe and Siddell, 1989b; although revised in Raabe et al., 1990).

Further studies are needed to ascertain the functions of the putative non-structural proteins encoded by the unique regions of mRNAs 4 and 5. The accumulation of sequence data for human coronaviruses will allow sequence-specific amplifica-

tion of the genome of these viruses from pathological specimens, and the production of specific antibodies which could be used to assess the function of the non-structural proteins.

Examination of the HCV-229E sequence did not reveal any features indicative of jumping of the RNA polymerase complex, such as a site of strong secondary structure or repeated sequences in the vicinity of the deletion site in ORF4b. Indeed, predicted RNA secondary structures were not more extensive in this region (results not shown). Therefore, our results are indicative of a possible genetic variability of a human coronavirus. Such genetic polymorphism is known to affect the pathogenesis of murine (Parker et al., 1989; Taguchi and Fleming, 1989) and porcine coronaviruses (Rasschaert et al., 1990). In the latter case, a few genomic deletions, including 672 nucleotides in the S gene, appeared to have modified viral tropism from the gastro-intestinal to the respiratory tract of infected pigs. Further studies are needed to confirm the genetic variability of human coronaviruses. Nevertheless, the findings reported in the present study, together with those reported for murine and porcine coronaviruses suggest that genomic deletions, as well as recombination, are a source of diversity in the coronavirus family.

We have compared the protein encoded by ORF4b of HCV-229E with the human myelin basic protein and found a five-amino acid homology, the sequence of which is LSLSR (Fig. 5). Since it belongs to exon 5 of the human MBP gene, this sequence is sometimes spliced among some human and murine variants (Roth et al., 1987; Takahashi et al., 1985). Assuming that the twenty amino acids are equally represented, the probability of finding a homology of five residues between two different proteins is 1 in 20^5, or 3.2×10^6. Although this homologous sequence is short, it could be sufficient to constitute a common epitope between the protein encoded by ORF4b of HCV-229E and the human MBP (Oldstone, 1987). Indeed, the minimum size of both B and T epitopes has been reported to be five amino acids (Geysen et al., 1989; Reddehase et al., 1989). On the other hand, several excellent indications suggest that multiple sclerosis could be the consequence of an autoimmune disease induced by a virus and one of the possible pathogenic mechanisms could involve molecular mimicry resulting from a sequence homology between a viral protein and myelin basic protein (Watanabe et al., 1983; Jahnke et al., 1985). If we consider the hypothesis of an autoimmune disease, the search for sequence homologies then takes its full significance.

The homologous sequence LSLSR overlaps by two amino acids an encephalitogenic site of human MBP (Carnegie, 1971). These two residues belong to a group of three amino acids which seem essential for the encephalitogenic potential (Lennon et al., 1970). Moreover, an encephalitogenic site for LOU/M rats encompasses this homology region (Hashim et al., 1991) and a similar site for SJL mice was reported slightly upstream (Fritz and McFarlin, 1989). Similar results were reported by Fujinami and Oldstone (1985), who have found a sequence homology between hepatitis B virus polymerase and the encephalitogenic site of rabbit MBP. These authors suggested that viral infection may trigger a neurologic disease resulting from a mechanism of molecular mimicry. Furthermore, the

LSLSR homologous sequence is located eight amino acids downstream from an immunodominant MBP epitope recognized by the T cells of multiple sclerosis patients (Ota et al., 1990). This immunodominant epitope may be encephalitogenic in some $DR2^+$ individuals. The biological importance of this MBP region was also suggested by Richert et al. (1989), who showed that nine out of forty MBP-specific human $CD4^+$ T cell clones recognized MBP residues 86 to 105. Also, Jingwu et al. (1990) reported that four out of 17 T cell clones established from MS patients recognized residues 108 to 148, a region which encompasses the homology region with HCV-229E. Finally, it is important to note that Allegretta et al. (1990) have isolated circulating T cells which recognize the MBP of multiple sclerosis patients, lending further support to the importance of such an autoimmune reaction in pathology.

In order to verify the hypothesis that HCV-229E may induce an autoimmune disease, it could be interesting to use a synthetic peptide corresponding to the LSLSR sequence. This peptide may, on one hand serve in proliferation tests of T cells from MS patients, and on the other hand be injected into mice in order to examine its autoimmune potential.

Acknowledgements

We thank Lucie Summerside for excellent secretarial assistance. P.J. acknowledges continuous studentship support from the Fonds de la Recherche en Santé du Québec. J.N.S is grateful to the Institut Armand-Frappier for a studentship award. This work was supported by Grant MT-9203 awarded by the Medical Research Council of Canada to P.J. Talbot, who also gratefully acknowledges salary support in the form of a University Research Scholarship from the Natural Sciences and Engineering Research Council of Canada. We thank R. Brousseau (Biotechnology Research Center, National Research Council of Canada, Montréal, Québec, Canada) for his valuable help in sequence analysis with the PC/GENE program.

References

Abel, T. and Maniatis, T. (1989) Action of leucine zippers. Nature (Lond.) 341, 24–25.

Abraham, S., Kienzle, T.E., Lapps, W.E. and 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.

Allegretta, M., Nicklas, J.A., Sriram, S. and Albertini, R.J. (1990) T-cells responsive to myelin basic protein in patients with multiple sclerosis. Science 247, 718–721.

Arpin, N. and Talbot, P.J. (1990) Molecular characterization of the 229E strain of human coronavirus. In: D. Cavanagh and T.D.K. Brown (Eds.), Advances in Experimental Medicine and Biology, Vol. 276, Coronaviruses and Their Diseases, pp. 73–80. Plenum, New York.

Boursnell, M.E.G. and Brown, T.D.K. (1984) Sequencing of coronavirus IBV genomic RNA: a 195-base open reading frame encoded by mRNA B. Gene 29, 87–92.

Boursnell, M.E.G., Binns, M.M. and Brown, T.D.K. (1985) Sequencing of coronavirus IBV genomic RNA: three open-reading frames in the 5' 'unique' region of mRNA D. J. Gen. Virol. 66, 2253–2258.

Britton, P., Lopez Otin, C., Martin Alonson, J.M. and Parra, F. (1989) Sequence of the coding regions from the 3.0 kb and 3.9 kb mRNA subgenomic species from a virulent isolate of transmissible gastroenteritis virus. Arch. Virol. 105, 165–178.

Budzilowicz, C.J. and Weiss, S.R. (1987) In vitro synthesis of two polypeptides from a non-structural gene of coronavirus mouse hepatitis virus strain A59. Virology 157, 509–515.

Burks, J.S., DeVald, B.L., Jankovsky, L.D. and Gerdes, J.C. (1980) Two coronaviruses isolated from central nervous system tissue of two multiple sclerosis patients. Science 209, 933–934.

Carnegie, P.R. (1971) Amino acid sequence of the encephalitogenic basic protein from human myelin. Biochem. J. 123, 57–67.

Cavanagh, D., Brian, D.A., Enjuanes, L., Holmes, K.V., Lai, M.M.C., Laude, H., Siddell, S.G., Spaan, W., Taguchi, F. and Talbot P.J. (1990) Recommendations of the Coronavirus Study Group for the nomenclature of the structural proteins, mRNAs, and genes of coronaviruses. Virology 176, 306–307.

Daniel, C. and Talbot, P.J. (1987) Physico-chemical properties of murine hepatitis virus, strain A59. Arch. Virol. 96, 241–248.

Davis, E.V. and Bolin, V.S. (1960) Continuous cultivation of isogenous cell lines from the human embryo. Fed. Proc. 19, 386.

Devereux, J., Haeberli, P. and Smithies, O. (1981) A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. 12, 387–395.

Evans, R.M. and Hollenberg, S.M. (1988) Zinc fingers: gilt by association. Cell 52, 1–3.

Favaloro, J., Treisman, R. and Kamen, R. (1980) Transcription maps of polyoma virus-specific RNA: analysis by two-dimensional nuclei's S1 gel mapping. In: Methods in Enzymology, Vol. 65, pp. 718–749. Academic Press, New York.

Fishman, P.S., Gass, J.S., Swoveland, P.T., Lavi, E., Highkin, M.K. and Weiss, S.R. (1985) Infection of the basal ganglia by a murine coronavirus. Science 229, 877–879.

Fritz, R.B. and McFarlin, D.E. (1989) Encephalitogenic epitopes of myelin basic protein. In: E. Sercarz (Ed.), Antigenic determinants and immune regulation, pp. 101–125, Karger, Basel.

Fujinami, R.S. and Oldstone, M.B.A. (1985) Amino acid homology between the encephalitogenic site of myelin basic protein and virus: mechanism for autoimmunity. Science 230, 1043–1045.

Geysen, H.M., Mason, T.J. and Rodda S.J. (1989) Cognitive features of continuous antigenic determinants. In: J.P. Tam and E.T. Kaiser (Eds.), Synthetic Peptides: Approaches to Biological Problems, pp. 19–30. Liss, New York.

Grunstein, M. and Hogness, D.S. (1975) Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. U.S.A. 72, 3961–3965.

Hashim, G., Vandenbark, A.A., Gold, D.P., Diamanduros, T. and Offner, H. (1991) T-cell lines specific for an immunodominant epitope of human basic protein define an encephalitogenic determinant for experimental autoimmune encephalomyelitis-resistant LOU/M rats. J. Immunol. 146, 515–520.

Hattori, M. and Sakaki, Y. (1986) Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. 152, 232–238.

Hierholzer, J.C. (1976) Purification and biophysical properties of human coronavirus 229E. Virology 75, 155–165.

Hierholzer, J.C., Kemp, M.C. and Tannock G.A. (1981) The RNA and proteins of human coronaviruses. In: V. ter Meulen, S. Siddell and H. Wege (Eds.), Advances in Experimental Medicine and Biology, Vol. 142, Biochemistry and Biology of Coronaviruses, pp. 43–69. Plenum, New York.

Jahnke, U., Fischer, E.H. and Alvord Jr., E.C. (1985) Sequence homology between certain viral proteins and proteins related to encephalomyelitis and neuritis. Science 229, 282–284.

Jingwu, Z., Chou, C.H.J., Hashim, G., Medear, R. and Raus, J.C.M. (1990) Preferential peptide specificity and HLA restriction of myelin basic protein-specific T-cell clones derived from MS patients. Cell. Immunol. 129, 189–198.

Jouvenne, P., Richardson, C.D., Schreiber, S.S., Lai, M.M.C. and Talbot, P.J. (1990) Sequence analysis of the membrane protein gene of human coronavirus 229E. Virology 174, 608–612.

Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132.

Lennon, V.A., Wilks, A.V. and Carnegie, P.R. (1970) Immunologic properties of the main encephalitogenic peptide from the basic protein of human myelin. J. Immunol. 105, 1223–1230.

Li, X. and Rhode, S.L. (1991) Nonstructural protein NS2 of parvovirus H-1 is required for efficient viral protein synthesis and virus production in rat cells in vivo and in vitro. Virology 184, 117–130.

Macnaughton, M.R. (1980) The polypeptides of human and mouse coronaviruses. Arch. Virol. 63, 75–80.

Martenson, R. (1983) Myelin basic protein speciation. Prog. Clin. Biol. Res. 146, 511–521.

McIntosh, K., Chao, R.K., Krause, H.E., Wasil, R., Mocega, H.E. and Mufson, M.A. (1974) Coronavirus infection in acute lower respiratory tract disease of infants. J. Inf. Dis. 139, 502–510.

Murray, RS., MacMillan, B., Cabirac, G. and Burks, J.S. (1990) Detection of coronavirus RNA in CNS tissue of multiple sclerosis and control patients. In: D. Cavanagh and T.D.K. Brown (Eds.), Advances in Experimental Medicine and Biology, Vol. 276, Coronaviruses and Their Diseases, pp. 505–510. Plenum, New York.

Niesters, H.G.M., Zijderveld, A.J., Seifert, W.F., Lenstra, J.A., Bleumink-Pluym, N.M.C., Horzinek, M.C. and Van der Zeijst, B.A.M. (1986) Infectious bronchitis virus RNA D encodes three potential translation products. Nucl. Acids Res. 14, 3144.

Oldstone, M.B.A. (1987) Molecular mimicry and autoimmune disease. Cell 50, 819–820.

Ota, K., Matsui, M., Milford, E.L., Mackin, G.A., Weiner, H.L. and Hafler, D.A. (1990) T-cell recognition of an immunodominant myelin basic protein epitope in multiple sclerosis. Nature (Lond.) 346, 183–187.

Parker, S.E., Gallagher, T.M. and 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.

Raabe, T. and Siddell, S.G. (1989a) Nucleotide sequence of the gene encoding the membrane protein of human coronavirus 229E. Arch. Virol. 107, 323–328.

Raabe, T. and Siddell, S. (1989b) Nucleotide sequence of the human coronavirus HCV 229E mRNA 4 and mRNA 5 unique regions. Nucl. Acids Res. 17, 6387.

Raabe, T., Schelle-Prinz, B. and Siddell, S.G. (1990) Nucleotide sequence of the gene encoding the spike glycoprotein of human coronavirus HCV 229E. J. Gen. Virol. 71, 1065–1073.

Rasschaert, D., Gelfi, J. and Laude, H. (1987) Enteric coronavirus TGEV: partial sequence of the genomic RNA, its organization and expression. Biochimie 69, 591–600.

Rasschaert, D., Duarte, M. and Laude, H. (1990) Porcine respiratory coronavirus differs from transmissible gastroenteritis virus by a few genomic deletions. J. Gen. Virol. 71, 2599–2607.

Reddehase, M.J., Rothbard, J.B. and Koszinowski, U.H. (1989) A pentapeptide as minimal antigenic determinant for MHC class I-restricted T lymphocytes. Nature (Lond.) 337, 651–653.

Resta, S., Luby, J.P., Rosenfeld, C.R. and Siegel, J.D. (1985) Isolation and propagation of a human enteric coronavirus. Science 229, 978–981.

Rhode, S.L. (1989) Both excision and replication of cloned autonomous parvovirus DNA require the NS1 (rep) protein. J. Virol. 63, 4249-4256.

Richert, J.R., Robinson, E.D., Deibler, G.E., Martenson, R.E., Dragovic, L.J. and Kies, M.W. (1989) Human cytotoxic T-cell recognition of a synthetic peptide of myelin basic protein. Ann. Neurol. 26, 342–346.

Roth, H.J., Kronquist, K.E., Kerlero de Rosbo, N., Crandall, B.F. and Campagnoni, A.T. (1987) Evidence for the expression of four myelin basic protein variants in the developing human spinal cord through cDNA cloning. J. Neurosci. Res. 17, 321–328.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487–491.

Salmi, A., Ziola, B., Hovi, T. and Reunanen, M. (1982) Antibodies to coronaviruses OC43 and 229E in multiple sclerosis patients. Neurology 32, 292–295.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5469.

Schmidt, O.W. and Kenny, G.E. (1982) Polypeptides and functions of antigens from human coronaviruses 229E and OC43. Infect. Immun. 35, 515–522.

Schreiber, S.S., Kamahora, T. and Lai, M.M.C. (1989) Sequence analysis of the nucleocapsid protein gene of human coronavirus 229E. Virology 169, 142–151.

Schwarz, B., Routledge, E. and Siddell, S.G. (1990) Murine coronavirus nonstructural protein ns2 is not essential for virus replication in transformed cells. J. Virol. 64, 4784–4791.

Shieh, C.-K., Lee, H.-J., Yokomori, K., La Monica, N., Makino, S. and Lai, M.M.C. (1989) Identification of a new transcriptional initiation site and the corresponding functional gene 2b in the murine coronavirus RNA genome. J. Virol. 63, 3729–3736.

Skinner, M.A. and Siddell, S.G. (1985) Coding sequence of coronavirus MHV-JHM mRNA 4. J. Gen. Virol. 66, 593–596.

Skinner, M.A., Ebner, D. and Siddell, S.G. (1985) Coronavirus MHV-JHM mRNA 5 has a sequence arrangement which potentially allows translation of a second, downstream open reading frame. J. Gen. Virol. 66, 581–592.

Spaan, W., Cavanagh, D. and Horzinek, M.C. (1988) Coronaviruses: structure and genome expression. J. Gen. Virol. 69, 2939–2952.

Sturman, L.S., Ricard, C.S. and Holmes, K.V. (1985) Proteolytic cleavage of the E2 glycoprotein of murine coronavirus: activation of cell-fusing activity of virions by trypsin and separation of two different 90K cleavage fragments. J. Virol. 56, 904–911.

Taguchi, F. and Fleming, J.O. (1987) Comparison of six different murine coronavirus JHM variants by monoclonal antibodies against the E2 glycoprotein. Virology 169, 233–235.

Takahashi, N., Roach, A., Teplow, D.B., Prusiner, S.B. and Hood, L. (1985) Cloning and characterization of the myelin basic protein gene from mouse: one gene can encoded both 14 kD and 18.5 kD MBPs by alternate use of exons. Cell 42, 139–148.

Tanaka, R., Iwasaki, Y. and Koprowski, H. (1976) Intracisternal virus-like particles in brain of a multiple sclerosis patient. J. Neurol. Sci. 28, 121–126.

Watanabe, R., Wege, H. and Ter Meulen, V. (1983) Adoptive transfer of EAE-like lesions from rats with coronavirus-induced demyelinating encephalomyelitis. Nature (Lond.) 305, 150–153.

Wege, H., Siddell, S. and Ter Meulen, V. (1982) The biology and pathogenesis of coronaviruses. Curr. Top. Microbiol. Immunol. 99, 165–200.

Weiss, S.R. and Leibowitz, J.L. (1981) Comparison of the RNAs of murine and human coronaviruses. In: V. Ter Meulen, S. Siddell and H. Wege (Eds.), Advances in Experimental Medicine and Biology, Vol. 142, Biochemistry and Biology of Coronaviruses, pp. 245–259. Plenum, New York.

Wesley, R.D., Cheung, A.K., Michael, D.D. and Woods, R.D. (1989) Nucleotide sequence of coronavirus TGEV genomic RNA: evidence for 3 mRNA species between the peplomer and matrix protein genes. Virus Res. 13, 87–100.

Yokomori, K. and Lai, M.M.C. (1991) Mouse hepatitis virus S RNA sequence reveals that nonstructural proteins ns4 and ns5a are not essential for murine coronavirus replication. J. Virol. 65, 5605–5608.

Yokomori, K., Banner, L.R. and Lai, M.M.C. (1991) Heterogeneity of gene expression of the hemagglutinin-esterase (HE) protein of murine coronaviruses. Virology 183, 647–657.

Zuker, M. and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucl. Acids Res. 9, 133–148.