SEQUENCE AND ANALYSIS OF BECV F15 MATRIX PROTEIN

E. Savoysky ([1]), **P. Boireau** ([3]), **C. Finance** ([1]) and **J. Laporte** ([2])

([1]) *Laboratoire de Microbiologie Appliquée et Industrielle,*
Faculté des Sciences Pharmaceutiques et Biologiques,
5 rue A. Lebrun, 54000 Nancy (France),
([2]) *Station de Virologie et d'Immunologie Moléculaires INRA,*
Domaine de Vilvert, 78350 Jouy-en-Josas (France), and
([3]) *Laboratoire central de Recherches Vétérinaires, CNEVA,*
22 rue Pierre Curie, BP 67, 94703 Maison-Alfort Cedex (France)

SUMMARY

Clones from the bovine enteric coronavirus (F15) cDNA library were cloned in pBR322 and sequenced by the method of Sanger and Coulson. This led to the identification of a sequence of 1,300 bases which contained a single open reading frame of 690 bases yielding a protein having properties of the matrix protein (M). It was comprised of 230 amino acids with a molecular weight of 26,376 Da. It was hydrophobic and had a net charge of + 8 at neutral pH. Analysis of its secondary structure could not establish a simple transmembrane arrangement of the amino acids. Comparison of its nucleotide sequence with that of BECV Mebus strain showed only a two-base change resulting in a 100 % homology between the two amino acid sequences. Furthermore, a very conserved structure of M appeared on comparison with the Dayoff optimal alignment of MHV-A59, MHV-JHM, TGEV, IBV Beaudette and IBV 6/82M amino acid sequences. As the two strains of BECV, F15 and Mebus present some antigenic differences, this led us to reconsider the role of M in viral antigen specificity. A hypothesis is that, as it seems to possess the necessary information on its transmembrane region, it is an ideal candidate for the viral budding process.

KEY-WORDS: Coronavirus, Matrix, Protein, Sequencing; Strain BECV F15, Analysis.

INTRODUCTION

Coronaviruses are enveloped, positive-stranded RNA viruses which infect mammals and birds and are responsible for diseases of clinical and economic importance; in particular, respiratory and gastrointestinal disorders (Sturman and Holmes, 1983). A member of this *Coronaviridae* family, the bovine enteric coronavirus (BECV) was identified for the first time in the USA by several studies and was recognized as one of the main causes of severe and often fatal enteritis in neonatal calves (Mebus *et al.*, 1973). A second strain of BECV was isolated in France in the faeces of diarrhoeic calves and adapted to cell culture (Laporte *et al.*, 1980; Gouet *et al.*, 1978). These two strains are distinguishable using monoclonal antibodies against E2 (Vautherot and Laporte, 1983).

Electron microscopy analysis established that, like the other members of the family, BECV is a spherical (120 nm diameter), pleomorphic, enveloped viral particle surrounded by a characteristic fringe of "club-shaped" spikes. BECV is comprised of 4 major structural proteins, a 50-kDa phosphorylated nucleocapsid N, a 180-kDa peplomeric glycoprotein S or E2 (present on the virion as subunits of 105 and 95 kDa), a 125-kDa haemagglutinin (made up of two disulphide-linked subunits of 65 kDa) and a 28-kDa transmembrane glycoprotein M or E1 (Laporte and Bobulesco, 1981; Vautherot *et al.*, 1984; King and Brian, 1982).

The viral genome is a positive-sensed single-stranded polyadenylated RNA of approximately 18 to 20 kb which is transcribed in infected cells into a 3'-coterminal "nested set" (Siddell *et al.*, 1982) of 8 subgenomic RNA (Crucière and Laporte, personal communication). It appears that generally only the 5'-proximal region which is not present in the next smaller RNA of the set is translated (Holmes *et al.*, 1986). Furthermore, each RNA contains a common 5'-end leader sequence which is derived from the 5'-end of the genomic RNA (Lai *et al.*, 1984).

It was established in both the American and European strains that the smallest RNA encodes the N protein; only the second RNA from the American strain has been studied and has been found to encode the M protein.

In this paper, we report the cloning and sequencing of the gene coding for the M protein of the French isolate F15 in order to compare its sequence with that of American strain of BECV. This would enable us to localize some

BECV	= bovine enteric coronavirus.	ORF	= open reading frame.
HRT	= human rectal tumour	S	= spicule.
IBV	= infectious bronchitis virus.	SDS	= sodium dodecyl sulphate.
IPTG	= isopropyl thiogalactoside.	TGEV	= transmissible gastroenteritis virus.
M	= matrix.	XGal	= 5,10-bromo-4-chloro-3-indolyl-galac-
MHV	= mouse hepatitis virus.		toside.

of the antigenic differences between the two strains and to increase the knowledge of the protein structure of the virus with a view to production of a more efficient vaccine by genetic engineering or oligopeptide synthesis.

MATERIALS AND METHODS

Cell culture and virus purification.

HRT18 cell (human rectal tumour cell line) culture and BECV F15 production and purification were performed as previously described by Laporte *et al.*, 1980 and Crucière and Laporte, 1988.

Genomic RNA purification.

The viral genome was purified by two different methods.

(1) *Proteinase K/phenol extraction.* — Purified viral suspension (500 µl) was processed by the method previously described by Crucière and Laporte, 1988.

(2) *Hot acid/phenol extraction.* — Purified viral suspension, 500 µl in distilled water, was added to 1 ml of sodium acetate buffer (10 mM pH 5.2) containing 0.5 % SDS, and to the same volume of phenol saturated with this buffer (Haymerle *et al.*, 1986). After 3 min of incubation at 55°C with vigourous shaking, the solution was quickly chilled on ice while still being shaken, then centrifuged for 10 min at 7,000 rpm. The aqueous phase was reextracted by the same method and recentrifuged. The upper aqueous phase was first phenol/chloroform-extracted at room temperature with 30 min of shaking, then chloroform-extracted, ethanol-precipitated, washed with 70 % cold ethanol and dissolved in distilled water to a concentration of 1 µg/µl.

cDNA synthesis and cloning.

The synthesis of cDNA complementary to the 5'-end of the BECV F15 genome was carried out in a volume of 50 µl: 2 µl of BECV RNA (1 µg/µl) extracted according to the previously described methods were added to 5 µg of oligodT 12-18 primer (7.5 µg/µl Pharmacia) and to 5 µl of distilled water, then denatured at 65°C for 5 min and quickly chilled on ice. The solution was added to 48 µl of 100 mM Tris-HCl (pH 8.3 at 42°C), 10 mM $MgCl_2$, 10 mM dithiotreitol, 110 mM KCl, 4 U RNasin (Genofit), 1 mM of dATP, dGTP, dTTP, 0.5 mM dCTP, 10 µCi α-^{32}P-dCTP (800 Ci/mmol in aqueous solution), 15 U reverse transcriptase (BRL). The incubation was performed for 90 min at 42°C and the reaction stopped by adding 2 µl 500 mM EDTA. Reaction products were extracted with phenol/chloroform and ethanol precipitation.

The RNA-DNA hybrids were treated by RNase T2, polydC-tailed and annealed to dG-tailed *Pst*I-linearized pBR322 plasmid as previously described by Crucière and Laporte. The sizes of hybrids were estimated before cloning by alkaline agarose gel electrophoresis (NaOH 30 mM, EDTA 2 mM) and autoradiography with intensifying screen for one night at −80°C.

Competent RR1 *Escherichia coli* were transfected with this material using the $CaCl_2$ method and tetracycline resistance (Maniatis *et al.*, 1982).

Screening and mapping of the recombinants.

The tetracycline-resistant ampicillin-susceptible colonies were selected and treated by alkaline lysis (Birnboim and Doly, 1979). The sizes of inserts were estimated by digestion of 1 μg of plasmidic DNA by 2 U *Pst*I enzyme (BRL) as described by Maniatis *et al.*, 1982 and electrophoretic migration on 1 % agarose gel in TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA). Inserts longer than 1 kb were mapped by using a set of restriction enzymes having no or a single site on the pBR322 plasmid, with appropriate buffers and by agarose gel electrophoresis.

M13 cloning and sequencing.

Oriented cloning.

Viral DNA inserts 160, 174, 216 were extracted from pBR322 plasmid and treated by restriction enzymes having sites on the M13 polylinker.

The DNA fragments ranging between 0.3 and 1 kb were purified by electrophoresis and electroelution into dialysis bags (Hediger, 1986).

The M13 phage DNA was cleaved by the same enzymes and 5'-dephosphorylated by alkaline phosphatase (Boehringer-Mannheim) treatment.

Amounts of 50 to 100 ng of DNA were ligated with 20 ng of vector by addition of T4 DNA ligase (Amersham, 1984) and TG1 *E. coli* were transfected by the CaCl$_2$ method.

Random cloning.

The pBR322-plasmid-containing viral insert was sonicated and subcloned into *Sma*I site of the M13 mp10 vector (Deininger, 1983). TG1 *E. coli* were transfected by the Hanahan method (Hanahan, 1985).

The TG1 oriented and random clones were selected in IPTG- and XGal-containing medium. White plaques of random clones were then checked by hybridization with insert radioactive probes as described by Maniatis *et al.* (1982).

M13 sequencing.

M13 dideoxy sequencing was carried out according to the Sanger technique using a universal primer, α-^{35}S-dATP and buffer gradient gels (Sanger *et al.*, 1977; Amersham, 1984). Sequence data were analysed and assembled with the aid of the Queen and Korn (1984) program of the "Beckman Microgenie" program (March 1985 version, Beckman Instruments Inc.) adapted to the "IBM-PC-XT" microcomputer. The program developed by Ralph and Smith (1987) (Bisance, CITI2, Paris) was used to analyse and predict amino acid sequence secondary structure.

RESULTS

cDNA cloning.

The sizes of cDNA obtained using 2 μg of starting material and oligodT 12-18 as a primer ranged between 0.3 and 6 kb with hot-phenol-purified

genomic RNA and between 0.3 and 18 kb with phenol/proteinase-K-purified genomic RNA (fig. 1).

We obtained 178 Tet^R-Amp^S clones containing 0.3- to 2.2-kb inserts. Only 24 inserts longer than 1 kb were studied and mapped (fig. 2). Their 5'-end is supposed to code for a part of the M glycoprotein by analogy with other studies of murine and avian coronaviruses. It is known that the M protein is encoded by a gene about 800 nucleotides long which is localized at approximately 1,500 nucleotides from the 3'-end of the viral genome.

Clones having 2 or 3 restriction sites (*Sph*I, *Pvu*II, *Xba*I) which correspond to the 5'-end of N gene (Cruciere and Laporte, 1988) and a 5'-end longer than 0.8 kb from *Sph*I site were retained for cloning in M13 phage.
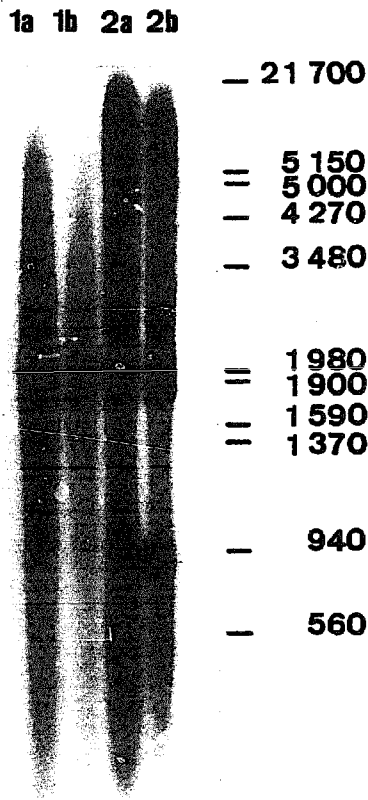

FIG. 1. — *Alkaline agarose gel electrophoresis of cDNA.*

1a) Size of the RNA-DNA hybrids obtained by action of reverse transcriptase on hot-phenol-purified genomic RNA. 1b) Hybrids treated by RNase T2. 2a) Hybrids were obtained with proteinase K/phenol-purified genomic RNA. 2b) Action of RNase T2 on the cDNA.

Inserts 216 (1.8 kb), 174 (2.2 kb), 160 (2 kb) were used for oriented cloning and sequencing in M13 after digestion by *Pst*I-*Xba*I, *Xba*I-*Sph*I, *Sph*I-*Pst*I, *Pvu*II-*Pst*I, *Sph*I-*Pvu*II. As deduced from the sizes of N and M viral proteins, clone 174 should cover a part of the N gene and probably the whole length of the M gene. It was used for random cloning in M13 and the white plaques were screened by its *Pst*I-*Xba*I fragment.
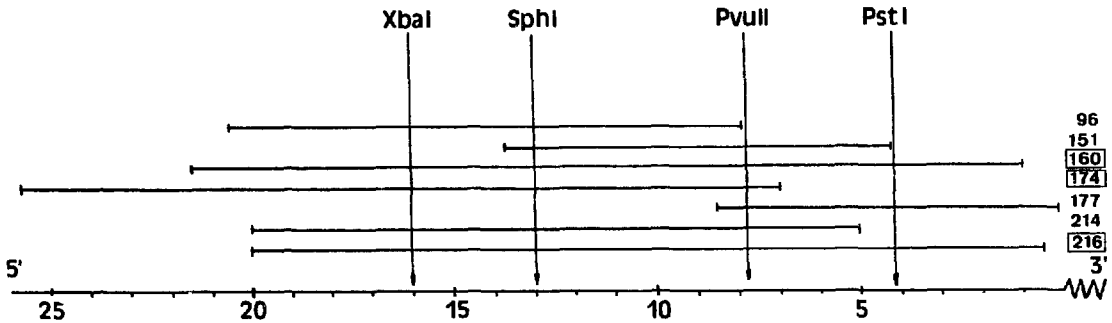
FIG. 2. — *Localization of the cDNA clones obtained.*

A simplified map of 7 of the 24 clones obtained is given. The inserts used for sequencing are in boxes.

cDNA sequencing.

After random and oriented cloning in M13 phage DNA, the nucleotide sequences of the inserted fragments were determined by multiple sequencing of both strands.

An open reading frame (ORF) of 690 bp was identified after sequence analysis by Microgenie. It stretched from the first ATG (nucleotides 131 to 133) to nucleotide 822 and had a coding capacity of 230 amino acids with a molecular weight of 26,376 Da. According to its length and position, this ORF comprised the gene of M glycoprotein (fig. 3). The 3'-end of the sequence represents the 5'-end of the nucleocapsid gene.

DISCUSSION

Nucleotide sequence analysis.

Study of the nucleotide sequence revealed the presence near each end of the ORF of two consensus sequences ATCCAAAC and ATCTAAAC which were assumed to be the start of the mRNA transcripts encoding M and N, respectively.

TATGTGTAATACCTTAGTACTGTCCCCTTCTATTTATGTGTTTAATAGAGGTAGGCAG:TTTATGAGTTTTACAATGATGTAAAACCACCAGTTCTTGATGTGGATGACGTTTAGTTAAT   120

CCAAADATTATGAGTAGTGTAACTACACCAGCACCAGTTTACACCTGGACTGCTGATGAAGCTATTAAATTCCTAAAGGAATGGAACTTTTCTTTGGGTATTATACTACTTTTTATTACA   240
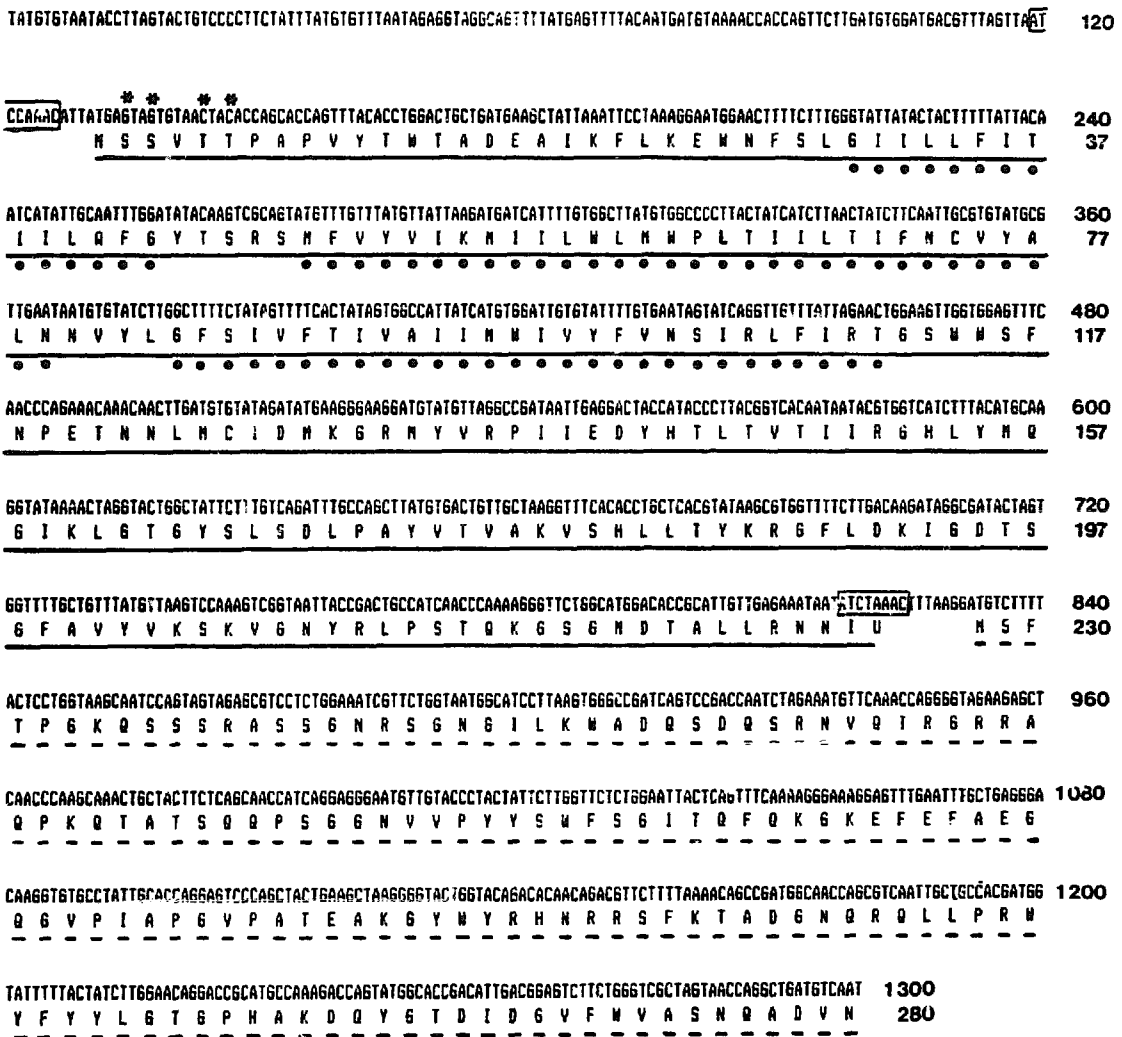 M S S V T T P A P V Y T M T A D E A I K F L K E M N F S L G I I L L F I T   37

ATCATATTGCAATTTGGATATACAAGTCGCAGTATGTTTGTTTATGTTATTAAGATGATCATTTTGTGGCTTATGTGGCCCCTTACTATCATCTTAACTATCTTCAATTGCGTGTATGCG   360
 I I L Q F G Y T S R S M F V Y V I K N I I L W L M M P L T I I L T I F N C V Y A   77

TTGAATAATGTGTATCTTGGCTTTTCTATAGTTTTCACTATAGTGGCCATTATCATGTGGATTGTGTATTTTGTGAATAGTATCAGGTTGTTATTAGAACTGGAAGTTGGTGGAGTTTC   480
 L N N V Y L G F S I V F T I V A I I M W I V Y F V N S I R L F I R T G S W H S F   117

AACCCAGAAACAAACAACTTGATGTGTATAGATATGAAGGGAAGGATGTATGTTAGGCCGATAATTGAGGACTACCATACCCTTACGGTCACAATAATACGTGGTCATCTTTACATGCAA   600
 N P E T N N L M C I D M K G R N Y V R P I I E D Y H T L T V T I I R G H L Y M Q   157

GGTATAAAACTAGGTACTGGCTATTCTTGTCAGATTTGCCAGCTTATGTGACTGTTGCTAAGGTTTCACACCTGCTCACGTATAAGCGTGGTTTCTTGACAAGATAGGCGATACTAGT   720
 G I K L G T G Y S L S D L P A Y V T V A K V S H L L T Y K R G F L D K I G D T S   197

GGTTTTGCTGTTTATGTTAAGTCCAAAGTCGGTAATTACCGACTGCCATCAACCCAAAAGGGTTCTGGCATGGACACCGCATTGTTGAGAAATAAATCTAAACTTTAAGGATGTCTTTT   840
 G F A V Y V K S K V G N Y R L P S T Q K G S G M D T A L L R N N I U N S F   230

ACTCCTGGTAAGCAATCCAGTAGTAGAGCGTCCTCTGGAAATCGTTCTGGTAATGGCATCCTTAAGTGGGCCGATCAGTCCGACCAATCTAGAAATGTTCAAACCAGGGGTAGAAGAGCT   960
 T P G K Q S S S R A S S G N R S G N G I L K W A D Q S D Q S R N V Q T R G R A

CAACCCAAGCAAACTGCTACTTCTCAGCAACCATCAGGAGGGAATGTTGTACCCTACTATTCTTGGTTCTCTGGAATTACTCABTTTCAAAAGGGAAAGGAGTTTGAATTTGCTGAGGGA   1080
 Q P K Q T A T S Q Q P S G G N V V P Y Y S W F S G I T Q F Q K G K E F E F A E G

CAAGGTGTGCCTATTGCACCAGGAGTCCCAGCTACTGAAGCTAAGGGGTACTGGTACAGACACAACAGACGTTCTTTTAAAACAGCCGATGGCAACCAGCGTCAATTGCTGCCACGATGG   1200
 Q G V P I A P G V P A T E A K G Y W Y R H N R R S F K T A D G N Q R Q L L P R W

TATTTTTACTATCTTGGAACAGGACCGCATGCCAAAGACCAGTATGGCACCGACATTGACGGAGTCTTCTGGGTCGCTAGTAACCAGGCTGATGTCAAT   1300
 Y F Y Y L G T G P H A K D Q Y G T D I D G V F W V A S N Q A D V N   280

FIG. 3. — *Nucleotidic sequence of BECV cDNA clones.*

The amino acid sequence deduced from the cDNA inserts is shown as positive-sense DNA from 5'- to 3'-ends. The M potential ORF is underlined. The conserved intergenic sequences are in boxes. Potential glycosylation sites are indicated by asterisks. The localization of the transmembrane regions are represented by dots. A part of the nucleocapsid gene is represented by dashes.

These sequences are very similar to the other BECV, MHV-JHM, MHV-A59 and TGEV conserved intergenic sequences (Lapps *et al.*, 1987; Crucière and Laporte, 1988; Pfleiderer *et al.*, 1986; Laude *et al.*, 1987; Armstrong *et al.*, 1984), and the most favourable initiation codon is the first ATG just downstream of the consensus sequence as described by Kozack (1983).

Furthermore, comparison of the entire nucleotide sequence with the published M-gene sequence of BECV Mebus strain (Lapps *et al.*, 1987) showed a very strong homology with only two base changes (343 and 667), which did not induce any modification in the amino acid sequence.

Amino acid sequence analysis.

Computer predictions of the amino acid sequence of BECV M gene, its secondary structure and study of its hydrophilic profile led to a model in which approximately 10 % of the N-terminal part of the molecule (28 residues) is exposed on the outer surface of the virus membrane. At this extremity, a cluster of residues SSVTT provides a potential attachment site for an *O*-glycosidically linked carbohydrate chain, as proposed by Niemann *et al.* (1984) for MHV-A59 with the analogous sequence SSTT.

Between amino acids 29 and 111, the protein was found to span the membrane three times (amino acids 29 to 43, 49 to 79 and 85 to 111) with each spanning region enriched in hydrophobic residues and separated by short hydrophilic sections at the surface of the membrane (fig. 3).

This structure is very similar to the M protein amino acid sequence of IBV as described by Boursnell *et al.*, 1984, who concluded that the molecule was buried in the viral membrane with only the hydrophilic COOH-terminal and exposed.

For the M protein of BECV, the amphiphilic C-terminal end (amino acids 112 to 185) could be associated with the inner face of the viral membrane, and the protruding very terminal end (net charge of + 8) could interact with the negatively charged RNA.

A first analysis of its amino acid sequence by "Microgenie" predicted a transmembrane structure of three β sheets, which was in contrast with Armstrong *et al.*, 1984 and Spaan *et al.*, 1988, who proposed three transmembrane α helixes. Another study by Bisance preferentially predicted a β-sheet structure; nevertheless, an α-helix structure was also possible.

Thus, it appears that a simple arrangement of the amino acids cannot be established in this transmembrane region. This may be related to very strong structural constraints in the membrane lipid bilayer, which could also explain the limited evolution of the M protein.

The coronavirus is a largely worldwide family which is divided mainly into four antigenic classes. A net antigenic specificity was found among each class species and, moreover, between two strains of a same species. However, the protein level of this specificity has not been elucidated for all coronaviruses: it is assumed to be linked to the spike and, perhaps, the matrix proteins.

Fig. 4. — *Comparison by "Dayoff optimal alignment" (Corpet, 1988) of six species of coronavirus.*

From Armstrong *et al.*, 1984; Binns *et al.*, 1986; Boursnell *et al.*, 1984; Laude *et al.*, 1987; and Pfleiderer *et al.*, 1986.

The main regions of homology are in boxes. Asterisks correspond to areas of exact homology, : represent a conservative change. Amino acids 1 to 21 of TGEV have no counterpart in the other species. Intramembrane segments are underlined.

```
                                    1              10             20             30
BECV F15                           MSSVT......TPAPVYT..WTADEAIKFLKEWNFSLGI
                                   ***  * ***   ******:  *****+***+**+*
MHV A59                            MSSTTQ.....APEPVYQ..WTADEAVQFLKEWNFSLGI
                                   **+***+*       ** ****  ******+****+*+****
MHV JHM                            MSSTTQ.....APGPVYQ..WTADEAVQFLKEWNFSLGI
                                   *  * *:          :        : :  *  *+**  *
TGEV 61   MKILLILACVIACACGERYCAMKSDTDLSCRNSTASDCESCFNGGDLIWHLANWNFSWSI
                                   *:  *:             *  : :  : :      *:
IBV 6/82                           MDNTIN.........CT..LGTEQAVQLFKEYNLFVTA
                                   *: *  *:         ** *  ** ******+**: **
IBV Beaudette                      MPNEIN.........CT..LDFEQSVQLFKEYNLFITA
                                                                          4

                    40          50          60          70          80
BECV F15   ILLFITIILQFGYTSRSMFVYVIKMIILWLMWFLTIILTIFN..CVYALNN.VYLGFSIV
           *********+*+*******:**:.********++*.***  ****+*** ******+*
MHV A59    ILLFITIILQFGYTSRSMFIYVVKMIILWLMWFLTIVLCIFN..CVYALNN.VYLGFSIV
           *****+*+*+*+*******+*+**+*+*+*+**+** ****:** *****+** ******+*
MHV JHM    ILLFITIILQFGYTSRSMFIYVVKMIILWLMWFLIIVLCMFN..CVYALNN.VYLGFSIV
           **:        * *:* :**+*:**: *  *:* :**: :*:      *       *
TGEV       ILIVFITVLQYGRPQFSWEVYGIKMLIMWLLWFVVLALTIFNAYSEYQVSRYVMFGFSLA
           *:       :*+***+     :*+:**::+*+  :**+  :*:            :    *
IBV 6/82   FLLFLTIILQYGYATRNKVIYIIKMIVILWCFWPLNIAVGAIS..CIYPPNT.GGLVAAIII
           ****** :*+********  ****+*+******+**+***** ** * *+**** *******+*
IBV B.     FLLFLTIILQYGYATRSKVIYTIKMIVILWCFWPLNIAVGVIS..CTYPPNT.GGLVAAIII
           _____                        4  4                  4  4

                90        100        110        120        130        140
BECV F15   FTIVAIIMWIVYFVNSIRLFIRTGSWWSFNPETNNLMCIDM.KGRMYVRPIIEDYHTLTV
           **** *::**:*+*****+**+*+*+***********+**** **:+*************
MHV A59    FTIVSIVIWIMYFVNSIRLFIRTGSWWSFNPETNNLMCIDM.KGTVYVRPIIEDYHTLTA
           *****:*:*****+*****+**+*+*+***********+**** **+************
MHV JHM    FTIVSVVMWIMYFVNSIRLFIRTGSWWSFNPETNNLMCIDM.KGTVYVRPIIEDYHTLTA
           **  *:**+*+*+*****+**:**  ********+  ::*:  **  ** *:      :*
TGEV       GAIVTFVLWIMYFVRSIQLYRRTNSWWSFNPETKAILCVSA.LGRSYVLPLEGVPTGVTL
           :   :   *+*+ : *+ *:+*+ *** **** *:   :      : :* **   :
IBV 6/82   LTVFACLSFIGYWIQSFRLFKRCRSWWAFNPESNAVGSILLTNGQQCNFAIESVPMVLSP
           *********:*+****+***+*+**+*** ********+********  ********
IBV B.     LTVFACLSFVGYWIQSIRLFKRCRSWWSFNPESNAVGSILLTNGQQCNFAIESVPMVLSP
           _____                 4

                150        160        170        180        190        200
BECV F15   TIIRGHLYMQGIKLGTGYSLSDLPAYVIVAKVSHLLTYKRG..FLD.KIGDTSGFAVYVK
           *****+*+****+:*+****:***+*+*+*+*+***** *+*  *** *:  ******+*
MHV A59    TIIRGHLYMQGVKLGTGFSLSDLPAYVIVAKVSHLCTYKRA..FLD.KVDGVSGFAVYVK
           *****+*+*+*+*:*+*+*****+*+*+*+*+** *+**  ** * ******+****
MHV JHM    TIIRGHFYMQGVKLGTGFSLSDLPAYVIVAKVSHLCTYKRA..FLD.KVDGVSGFAVYVK
           *::*+*+:*+ :*+*: * :   : :** **+** *:     *      * +*  ***
TGEV       TLLSGNLYAEGFKIADGMNIDNLPKYVMVALPSRTIVTLV..GKKLKASSATGWAYYVK
           :    *+*+**  **+  :* *:+**+*+ *+ * *  :*+ :*         *+   :  ::
IBV 6/82   IIKNGVLYCEGQWLAKC.EPDHLPKDIFVCTPDRRNIYRMVQRYTGDQSGNKREATFIY
           ********+*+*+****     ***+**+*+****+*+*:** *:******+********:*
IBV B.     IIKNGVLYCFGQWLAKC.EPDHLPKDIFVCTPDRRNIYRMVQKYTGDQSGNKREATFVY

                210        220        230
BECV F15   SKVGNYRLPSTQKGSGMDTALLRNNI
           *+******** * ** ***+**
MHV A59    SKVGNYRLPS.NKPSGADTALLRI
           *+******** ******* +***
MHV JHM    SKVGNYRLPS.NKPSGADTVLLRI
           ** *:*        :   +*+::
TGEV       SKAGDYSTEARTDNLSEQEKLLHMV
           *+       *                +*
IBV 6/82   VKHSV.DTGELESVATGGSSLYT
           *+:** *********+****
IBV B.     AKQSV.DTGELESVATGGSSLYT
```

Several studies about the M protein, especially that of MHV which was the first coronavirus matrix protein to be sequenced (Armstrong *et al.*, 1984), showed that all species share a similar structure with different degrees of homology (Kapke *et al.*, 1988).

Therefore, it appears that the deduced amino acid sequence of BECV F15 M protein has a 100 % homology with that of the Mebus strain. Indeed, the two strains were isolated from each side of Atlantic in very different conditions (Gouet *et al.*, 1978; Mebus *et al.*, 1973). This remarquable phenomenon led us to reconsider the role of the M protein in the viral antigenic specificity.

Comparison by "Dayoff optimal alignment" (Multalin program, Corpet, 1988) with other species of coronaviruses and between several strains of each species showed that BECV shares an 87 % amino acid homology with two strains of MHV, 43 % and 34 % homology with TGEV and two strains of IBV respectively (fig. 4).

Moreover, a strong homology was noted between the two strains of MHV (JHM and A59) which differ by only 7 residues (Pfleiderer *et al.*, 1986) and the two strains of IBV with 18 amino acid changes (Binns *et al.*, 1986). All these changes are conservative.

Unlike BECV, MHV and IBV matrix proteins, TGEV was found to possess a cleavable signal peptide at its N-terminal extremity. However, Boursnell *et al.* (1984) found a 22 in-frame codon between the consensus sequence and the initiation codon of IBV M gene which may be the remnant of an ancestral signal peptide.

Thus, the exposed part of the M protein of mammalian coronaviruses would appear to share with that of avian viruses a hypervariability with respect to size.

Whether the observed high degree of variation in the N terminus of the M protein is of biological significance is not known. These variations are readily tolerated and of little consequence with regard to the function of the protein (Cavanagh and Davis, 1988).

The hydrophobic part of the molecule contains an eight-amino-acid section (113 to 120, BECV) which is perfectly conserved among the six strains of viruses. The other main regions of homology are shown in figure 4.

The first potential membrane-spanning region has a high degree of homology between the four species, the second is well conserved within MHV-IBV pairs and the third with BECV-IBV and TGEV-MHV pairs. This may be indicative of a functional differentiation between the three segments (Laude *et al.*, 1987).

Spaan *et al.* (1988) reported that although the matrix M protein of MHV, IBV and BECV does not possess an N-terminal signal sequence, it needs the signal recognition particle for membrane insertion. So, the first and/or the third amino-terminal membrane-spanning region may function as a signal sequence (Rottier *et al.*, 1985; Boursnell *et al.*, 1984).

Comparison between the degree of homology of the matrix protein of these four coronaviruses and that of the N protein of the same species shows that there is higher conservation of the matrix protein (Crucière and Laporte, 1988; Lapps *et al.*, 1987).

M proteins possess a similar hydrophilicity profile (fig. 5) and similar membrane topology, most of the amino acid changes being conservative.

Conservation of the protein sequence in the hydrophobic N-terminal half of the polypeptide suggests a high degree of selective pressure for the maintenance of the structure of the putative transmembrane protein, resulting in limited evolution of this molecule (Boursnell *et al.*, 1984; Lapps *et al.*, 1987).

BECV, MHV, TGEV and IBV could have a common evolutionary origin, the strains diverging from a common ancestor and the M gene undergoing minimal variation.

Several studies of the MHV and BECV F15 S protein (Luytjes *et al.*, 1987; Boireau *et al.*, 1990) showed the first great divergence between these two species with a deletion of 46 residues in the MHV-A59 E2 protein and 41 residues in that of MHV-JHM.

This protein has a higher degree of evolution than the matrix protein. Thus, it appears that the antigenic specificity of BECV, and in general of coronaviruses, could be preferentially localized on the S gene; N protein being an internal molecule and having no role in immunity.

Nevertheless, if the M protein is not essential for viral antigenic specificity, it seems to possess the necessary information on its transmembrane region, probably an insertion signal as we reported before, to be an ideal candidate for use in determining the site of intracellular budding; and it is postulated that the M glycoprotein is the determinant for the budding process (Niemann *et al.*, 1984; Rottier and Rose, 1987).

RÉSUMÉ

SÉQUENÇAGE ET ANALYSE DE LA PROTÉINE DE LA MATRICE
DU CORONAVIRUS F15 DE L'ENTÉRITE BOVINE

Le clonage dans le plasmide pB322 et le séquençage (selon la méthode de Sanger et Coulson) de clones réalisés à partir d'une collection d'ADNc du coronavirus F15 de l'entérite bovine (BECV) nous ont permis d'identifier une séquence de 1300 bases contenant une séquence de lecture ouverte de 690 bases qui engendre une protéine dont les propriétés sont celles de la protéine M de la matrice. Cette protéine possède 230 acides aminés et son poids moléculaire est de 26376 daltons; elle est hydrophobe et sa charge nette est de + 8 à pH neutre. L'analyse de la structure secondaire ne permet pas d'établir un arrangement transmembranaire simple des acides aminés. La comparaison entre les séquences nucléotidiques afférentes pour les souches BECV F15 et BECV Mebus a montré que le seul changement de 2 bases donne une homologie de 100 % pour les séquences d'acides aminés. De plus, la structure de la protéine M apparaît très conservée quand on la compare avec les alignements optimaux de Dayoff pour les séquences d'acides nucléiques des virus MHV-A59, MHV-JHM, TGE, IBV Beaudette et IBV 6/82M. Le fait que les 2 souches, M15 et Mebus, du BECV
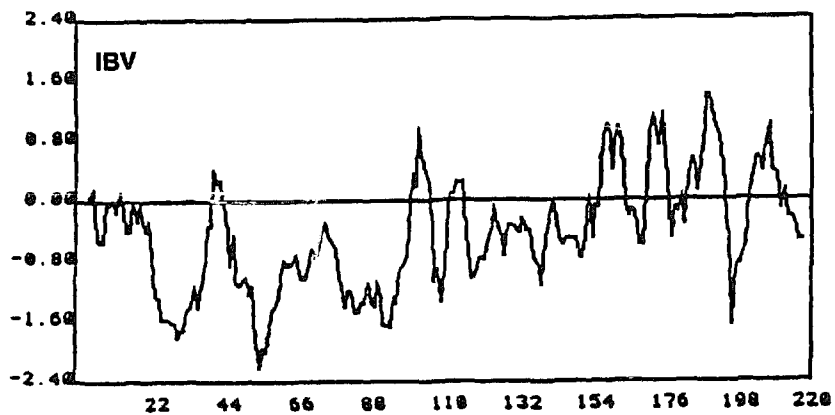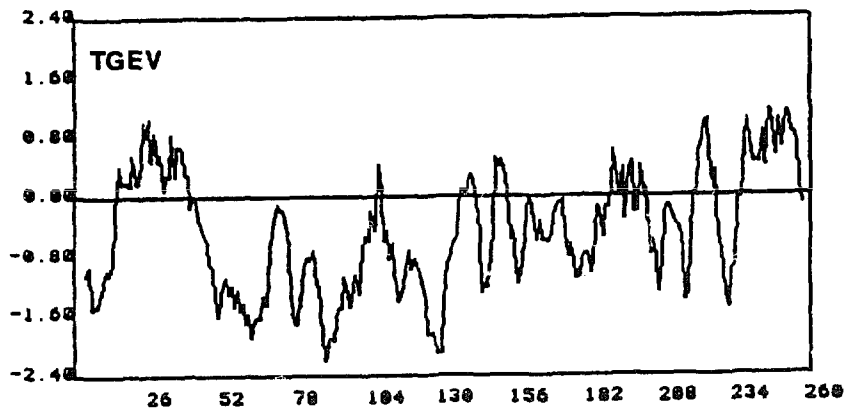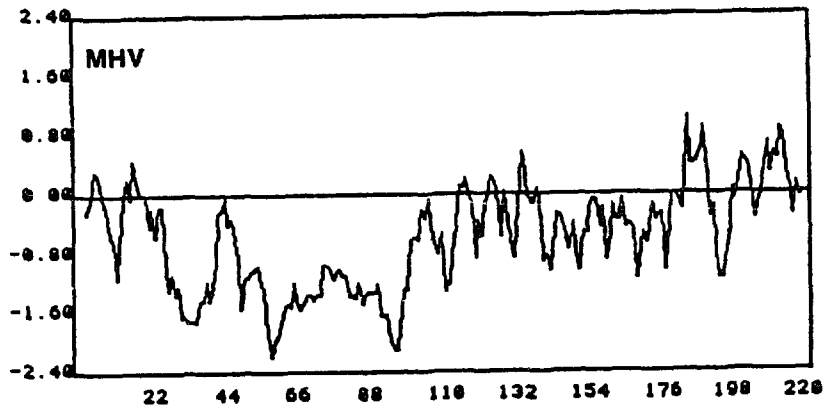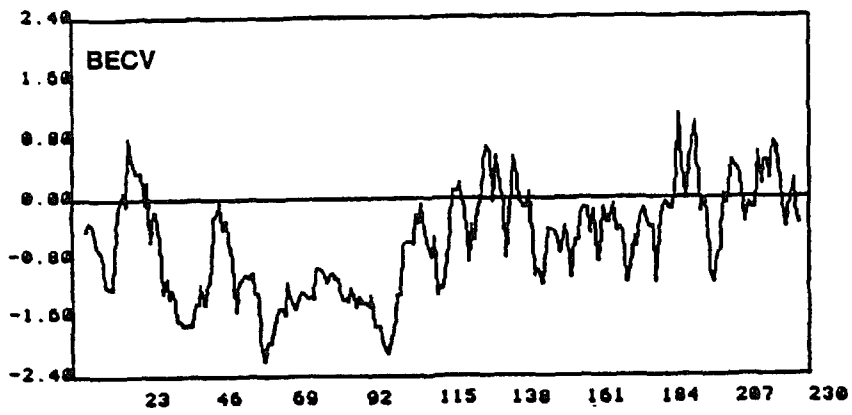
FIG. 5. — *Hydrophilicity profiles of BECV, MHV, TGEV and IBV.*

présentent quelques différences antigéniques nous conduit à reconsidérer le rôle de la protéine M dans la spécificité antigénique virale. Une hypothèse est qu'elle paraît posséder l'information nécessaire dans sa région transmembranaire pour être la candidate idéale dans le processus de bourgeonnement viral.

Mots-clés: Coronavirus, Protéine, Matrice, Séquençage; Souche BECV-F15, Analyse.

REFERENCES

AMERSHAM (1984), M13 cloning and sequencing handbook. Doc. Amersham, 54 p., Amersh. Internat. Plc.

ARMSTRONG, J., NIEMANN, H., SMEEKENS, S., ROTTIER, P. & WARREN, G. (1984), Sequence and topology of a model intracellular membrane protein, E1, glycoprotein from a coronavirus. *Nature* (Lond.), **308**, 751-753.

BINNS, M.M., BOURSNELL, M.E.G., TOMLEY, F.M. & BROWN, T.D.K. (1986), Nucleotide sequence encoding the membrane protein of the IBV strain 6/82. *Nucl. Acids Res.*, **14**, 5558.

BIRNBOIM, M.C. & DOLY, J. (1979), A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.*, **7**, 1513-1520.

BOIREAU, P., CRUCIÈRE, C. & LAPORTE, J. (1990), Sequence and analysis of bovine enteric coronavirus F15 genome. — II. Sequence of the gene coding for the glycoprotein S E2; analysis of the predicted structure of the protein. *J. gen. Virol.*, **71**, 487-492.

BOURSNELL, M.E.G., BROWN, T.D.K. & BINNS, M.M. (1984), Sequence of the membrane protein gene from avian coronavirus IBV. *Virus Res.*, **1**, 303-314.

CAVANAGH, D. & DAVIS, P.J. (1988), Evolution of avian coronavirus IBV: sequence of the matrix glycoprotein gene and intergenic region of several serotypes. *J. gen. Virol.*, **69**, 621-629.

CORPET, F. (1988), Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.*, **16**, 10881-10890.

CRUCIÈRE, C. & LAPORTE, J. (1988), Sequence and analysis of bovine enteritic coronavirus F15 genome. — I. Sequence of the gene coding for the nucleocapsid protein; analysis of the predicted protein. *Ann. Inst. Pasteur/Virol.*, **139**, 123-138.

DEININGER, P.L. (1983), Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Analyt. Biochem.*, **129**, 216-223.

GOUET, P., CONTREPOIS, H.C., DUBOURGUIER, Y., RIOU, R., SCHERRER, R., LAPORTE, J., VAUTHEROT, J.F., COHEN, J. & L'HARIDON, R. (1978), The experimental production of diarrhoea in axenic and gnotobiotic calves with enteropathogenic *E. coli*, rotavirus, coronavirus and in combinated infections of rotavirus and *E. coli*. *Ann. Rech. Vet.*, **9**, 433-440.

HANAHAN, D. (1985), Techniques for transformation of *E. coli*, *in* "DNA cloning: a practical approach" (Glover, D.M.), vol. 1 (pp. 109-135). IRL Press, Oxford, Washington.

HAYMERLE, H., HERZ, J., BRESSAN, G.M., FRANK, R. & STANLEY, K.K. (1986), Efficient construction of cDNA librairies in plasmid expression vector using an adaptator strategy. *Nucl. Acids Res.*, **14**, 8615-8624.

HEDIGER, M.A. (1986), High resolution preparative gel electrophoresis of DNA fragments and plasmid DNA using a continuous elution apparatus. *Analyt. Biochem.*, **159**, 280-286.

HOLMES, K.V., BOYLE, J.F., WILLIAMS, R.K., STEPHENSEN, C.B., ROBBINS, S.G., BAUER, E.C., DUCHALA, C.S., FRANA, M.F., WEISMILER, D.G., COMPTON, S., McGOWAN, J.J. & STURMAN, L.S. (1986), Processing of coronavirus proteins and assembly of virion, *in* "Positive strand RNA viruses" (339-349). Alan R. Liss, Inc., New York.

KARPE, P.A., TUNG, F.Y.T., HOGUE, B.G., BRIAN, D.A., WOODS, R.D. & WESLEY, R. (1988), The amino-terminal signal peptide on the porcine transmissible gastroenteritis coronavirus matrix protein is not an absolute requirement for membrane translocation and glycosylation. *Virology*, **165**, 367-356.

KING, B. & BRIAN, D.A. (1982), Bovine coronavirus structural proteins. *J. Virol.*, **42**, 700-707.

KOZAK, M. (1983), Comparison of initiation of protein synthesis in procaryotes, eucaryotes and organelles. *Microbiol. Rev.*, **47**, 1-45.

LAI, M.M.C., BARIC, R.S., BRAYTON, P.R. & STOHLMAN, S.A. (1984), Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. *Proc. nat. Acad. Sci.* (Wash.), **81**, 3626-3630.

LAPORTE, J., BOBULESCO, P. & ROSSI, F. (1980), Une lignée particulièrement sensible à la réplication du coronavirus entéritique bovin: les cellules HTR18. *C.R. Acad. Sci.* (Paris) *(sér. D)*, **290**, 623-626.

LAPORTE, J. & BOBULESCO, P. (1981), Polypeptide structure of bovine enteritic coronavirus: comparison between a wild strain purified from faeces and a HRT18 cell adapted strain, *in* "Biochemistry and biology of coronaviruses" (Ter Meulen V. *et al.*) **14** (pp. 181-184). Plenum Press, New York.

LAPPS, W., HOGUE, B.G. & BRIAN, D.A. (1987), Sequence analysis of bovine coronavirus nucleocapsid and matrix protein genes. *Virology*, **157**, 47-57.

LAUDE, H., RASSCHAERT, D. & HUET, J.C. (1987), Sequence and N-terminal processing of the transmembrane protein E1 of the coronavirus transmissible gastroenteritis virus. *J. gen. Virol.*, **68**, 1687-1693.

LUYTJES, W., STURMAN, L.S., BREDENBEEK, P.J., CHARITE, J., VAN DER ZEIJST, B.A.M., HORZINEK, M.C. & SPAAN, W.J.M. (1987), Primary structure of the glycoprotein E2 of coronavirus MHV-A59 and identification of the trypsine cleavage site. *Virology*, **161**, 179-187.

MANIATIS, T., FRITSCH, E.F. & SAMBROOK, J. (1982), Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, New York.

MEBUS, C.A., STAIR, E.L., RHODES, M.B. & TWIEHAUS, M.J. (1973), Pathology of neonatal calf diarrhea induced by a coronavirus-like agent. *Vet. Path.*, **10**, 45-64.

NIEMANN, H., GEYER, R., KLENK, H.D., LINDER, D., STIRM, S. & WIRTH, M. (1984), The carbohydrates of mouse hepatitis virus MHV-A59: structures of the O-glycosidically linked oligosaccharides of glycoprotein E1. *EMBO J.*, **3**, 665-670.

PFLEIDERER, M., SKINNER, M.A. & SIDDELL, S.G. (1986), Coronavirus MHV-JHM: nucleotide sequence of the mRNA that encodes the membrane protein. *Nucl. Acids Res.*, **14**, 6338-6339.

QUEEN, C. & KORN, L.J. (1984), A comprehensive sequence analysis program for the IBM personal computer. *Nucl. Acids Res.*, **12**, 581-599.

RALPH, W. & SMITH, T.F. (1987), A modified Chou and Fasman protein structure algorithm. *CABIOS*, **3**, 211-216.

ROTTIER, P.J., AMSTRONG, J. & MEYER, D.J. (1985), Signal recognition particle-dependent insertion of coronavirus E1, an intracellular membrane glycoprotein. *J. biol. Chem.*, **260**, 4648-4652.

ROTTIER, P.J. & ROSE, J.K. (1987), Coronavirus E1 glycoprotein expressed from cDNA localized in the Golgi region. *J. Virol.*, **61**, 2042-2045.

SANGER, F., NICKLEN, S. & COULSON, A.R. (1977), DNA sequencing with chain termination inhibitors. *Proc. nat. Acad. Sci.* (Wash.), **74**, 5463-5467.

SIDDELL, St., WEGE, H. & TER MEULEN, V. (1982), The structure and replication of coronaviruses. *Curr. Top. Microbiol. Immunnol.*, **99**, 131-163.

SPAAN, W., CAVANAGH, D. & HORZINEK, M.C. (1988), Review article. Coronaviruses: structure and genome expression. *J. gen. Virol.*, **69**, 2939-2952.

STURMAN, L.S. & HOLMES, K.V. (1983), The molecular biology of coronaviruses, *in* "Advances in virus research" (Lauffer M.A., Maramorosch K.), **28** (pp. 35-112). Academic Press, London, New York.

VAUTHEROT, J.F. & LAPORTE, J. (1983), Utilisation of monoclonal antibodies for antigenic characterization of coronaviruses. *Ann. Rech. Vet.,* **14,** 437-444.

VAUTHEROT, J.F., LAPORTE, J., MADELAINE, M.F., BOBULESCO, P. & ROSETO, A. (1984), Antigenic and polypeptide structure of bovine enteritic coronavirus as defined by monoclonal antibodies, *in* "Molecular Biology and Pathogenesis of Coronaviruses" (Rottier P.J.M.) (pp. 117-132). Plenum Press, New York.