SEQUENCE AND ANALYSIS OF BOVINE ENTERITIC CORONAVIRUS (F15) GENOME

I. — SEQUENCE OF THE GENE CODING FOR THE NUCLEOCAPSID PROTEIN; ANALYSIS OF THE PREDICTED PROTEIN

C. Crucière [1] and **J. Laporte** [2]

*Station de Virologie et d'Immunologie, INRA,
CRJJ, Domaine de Vilvert, 78350 Jouy-en-Josas
(France)*

SUMMARY

Sequences encoding the N protein of the bovine enteritic coronavirus-F15 strain (BECV-F15) have been cloned in PBR322 plasmid using cDNA produced by priming with oligo-dT on purified viral genomic RNA. Some 265 insert-containing clones were studied. Hybridization of these inserts with poly(A)[+] RNA extracted from infected cells led to the conclusion that they were located at the 3'-end of the genome.

After subcloning in M13 phage DNA, clones were sequenced by the Sanger technique. A 1,710-nucleotide sequence corresponding to the gene coding for the viral N-protein was established. It shows 2 overlapping open reading frames (ORF). The 3'-non-coding end of the gene has an 8-nucleotide sequence in common with the homologous genome areas of MHV, TGE and IBV viruses. This sequence may represent the polymerase RNA binding site.

An upstream sequence surrounding the first AUG of the smaller ORF corresponds to a potentially functional initiation codon. The sequence of the primary translation product deduced from the DNA sequence predicts a polypeptide of 207 amino acids (22.9 Kd) with a high leucine (19.8 %) content, possessing a hydrophobic N-terminal end.

The larger ORF has a coding capacity of 448 amino acids (49.4 Kd), corresponding to the N-protein molecular weight. The deduced protein possesses 43 serine residues (9.6 % of the total amino acid content) which may be phosporylated and involved in N-protein/RNA binding. N-protein also has 5 regions with a high basic amino acid content. One of them is also serine-rich and has a strong homology site with MHV, TGE and IBV viruses. In the first part of the N-terminal, a 12-amino-acid sequence (PRWYFYYLGTGP) is highly conserved for BECV-F15, JHM, TGE and IBV viruses. BCV Mebus strain and BECV-F15 have only minor differences in their N-protein sequence.

KEY-WORDS: Coronavirus, Protein, Nucleocapside, Genome; BECV-F15 strain, N-protein sequence.

INTRODUCTION

Bovine enteritic coronavirus (BECV) belongs to the monogeneric Coronaviridae family having the avian infectious bronchitis virus as type species. They are pleiomorphic, enveloped, surrounded by a fringe of « club-shaped » spikes looking like a corona in the electron-microscope and giving the name to the family. The viral genome is a positive single-stranded RNA of approximately 18 to 20 kb, its 3'-end is polyadenylated [19, 22]. This genome codes for the viral proteins which are nucleocapsid (N), membrane (E1), spikes (E2) and several non-structural proteins. They are translated from a 3'-end co-terminal nested set of mRNA, each also having a common 5'-leader sequence [8]. Only the unique 5'-terminal sequence, not present in the next smaller RNA of the set, is translated.

It was recently established that, in fact, BECV contains 4 main structural proteins: the nucleoprotein N (50 Kd), the transmembrane E1 glycoprotein (28 Kd) and 3 peplomer glycoproteins E2, gp105 and gp95. The haemagglutinin protein E2 (125 Kd) is cleaved by reducing agents into 2 subunits having molecular weights of 65 Kd; the main neutralizing epitopes of the viral particle are located on gp105 (105 Kd) [9, 24, 6]; the structure of gp95 (95 Kd) is not clearly established.

The BECV induces very severe, often fatal, diarrhoea in young calves. It was described for the first time in the United States of America [13]; we have been able to isolate such a virus in the faeces of diarrhoeic calves in France and to experimentally reproduce the disease [4]. These 2 strains of BECV are distinguishable by using monoclonal antibodies [23].

BECV	= bovine enteritic coronavirus.	N	= nucleocapsid.
BSA	= bovine serum albumin.	ORF	= open-reading frame.
FCS	= foetal calf serum.		

Vaccines produced from cell culture of attenuated or inactivated BECV are not totally protective and they necessitate production of large volumes of viral suspension because of the low infectious titre obtained in authorized cell lines. For these reasons, we have started cloning and sequencing the French F15 strain of BECV to try and produce cheaper and more efficient vaccines by genetic engineering or by oligopeptidic synthesis.

MATERIALS AND METHODS

Cell culture and virus production.

HRT18 cells (human rectal tumour cell line) were grown in RPMI-1640 medium containing 15 % foetal calf serum (FCS) [10] except that tylosine (10 µg/ml) and lincomycine (200 µg/ml) were added to the medium instead of penicillin and streptomycin.

Bovine enteritic coronavirus F15 strain (BECV-F15) was isolated from diarrhoeic calf faeces, then directly adapted on HRT18 cells [10] and plaque-purified. It was grown as previously described [4]. Infectious titres reached 5×10^5 plaque-forming unit (PFU)/ml.

Virus purification.

After freezing and thawing of infected cells together with supernatant and then clarification, the virus was purified by 2 ultracentrifugation steps (velocity then isopycnic) [9].

Genomic RNA purification.

A 1-ml sample of purified virus suspension in distilled water was added to the same volume of 2-fold concentrated TNE buffer (20 mM pH 8 Tris-HCl, 200 mM NaCl, 2 mM EDTA) containing 400 µg of proteinase K. After incubation for 30 min at 37°C, then for 5 min at 50°C, a same volume of the same buffer containing 2 % SDS was added and incubation carried on for 30 min at 25°C.

Genomic RNA was phenol/chloroform-extracted, then precipitated in 2.5 volumes of 0.25 M sodium acetate in ethanol. After one night at -20°C, RNA suspension was centrifuged for 20 min at 10,000 g, the pellet washed with 75 % ethanol, dried and dissolved in minimal volume of distilled water. One optical density (OD) unit at 260 nm corresponded to 40 µg/ml of single-stranded RNA [12].

cDNA cloning.

The synthesis of cDNA complementary to the 3'-end of the BECV-F15 genome was carried out in a volume of 52 µl: 10 µg in 10 µl of BECV RNA, denatured at 65°C for 5 min and quickly chilled in an ice bath, were added to 42 µl of 100 mM pH 8.3 Tris-HCl at 42°C containing 100 mM KCl, 100 mM $MgCl_2$, 10 mM dithiothreitol, 4 µg actinomycin D, 500 µM each of the 4 dNTP, 75 units RNasin, 140 units reverse transcriptase (P.H. Stehelin), and as primer, 10 µg oligo-dT. Incubation was performed for 2 h at 42°C and the reaction was stopped by adding 2 µl 500 mM EDTA. Reaction products were extracted with phenol/chloroform, chloroform and ethanol precipitation. Free RNA strands non-hybridized with cDNA were digested with endonuclease T_2 [25]; these digests and free nucleotides were

removed by gel filtration on a spun column of « Sephadex-G50 » medium (Pharmacia) [12].

The RNA-cDNA heteroduplexes were then poly-dC tailed: 2 pmoles of 3' ends were dissolved in 20 μl of 25 mM Tris-HCl buffer pH 7 containing 100 mM K-cacodylate, 0.2 mM DTT, 1 mM $CoCl_2$, 0.2 mM dCTP, 50 μg bovine serum albumin (B.R.L.), 13.5 units of terminal-deoxynucleotidyl transferase (B.R.L.) and 30 μCi α-^{32}P-dCTP (3,000 Ci/mmole). The reaction was carried out at 37°C for 3 min and stopped by adding 2 μl 500 mM EDTA [16]. The product was phenol/chloroform-extracted. An average of 20 dC/3'-end of heteroduplex was obtained.

C-tailed heteroduplexes were annealed to dG-tailed *Pst*I-linearized PBR322 plasmid (1 mole for 2 moles), in a volume where the plasmid was at a concentration of 5 ng/μl at 65°C for 10 min. Competent RR1 *Escherichia coli* cells were transfected with this material [5]. The total DNA concentration was 0.25 μg/ml.

Identification of specific BECV inserts.

E. coli cells were grown overnight in a medium containing 12 μg/ml tetracycline, then treated by alkaline lysis [12]. Plasmidic DNA was extracted by phenol/chloroform treatment and ethanol-precipitated. DNA inserts were removed by *Pst*I restriction enzyme: 1.2 μl of 10-fold concentrated buffer (100 mM pH 7.5 Tris-HCl, 1 M NaCl, 100 mM $MgCl_2$, 1 mg/ml BSA) and 2 units of *Pst*I enzyme (B.R.L.) were added to 10 μl of plasmidic DNA solution. Insert size was established by electrophoretic migration in 1 % agarose gels in TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA).

Probes were prepared by nick-translation in a 20 μl volume containing 0.5 μg DNA, 2 μl of 10-fold concentrated buffer (500 mM pH 7.2 Tris-HCl, 100 mM $MgSO_4$, 1 mM DTT, 500 μg/ml BSA), 20 μM each of the 4 dNTP, 2.5 ng pancreatic DNase I (Boehringer), 40 μCi α-^{32}P-dCTP (800 Ci/mmole) and 0.8 unit DNA polymerase I. Mixture was incubated for 2 h at 16°C. Reaction was stopped by adding 3 μl 500 mM EDTA pH 8. Free nucleotides were removed by filtration through a spun column.

Northern and Southern blots were performed as described by Maniatis [12]. Probes were incubated for hybridization overnight at 42°C (Southern) or at 55°C (Northern); blots were then washed in low salt concentration solutions: three times for 15 min in 0.1 % SDS, 2 × SSC and twice for 15 min in 0.1 % SDS, × 0.1 SSC at 52°C.

DNA sequencing and sequence analysis.

M13 dideoxy sequencing was carried out according to the Sanger technique [17], using α-^{35}S-dATP (New England Nuclear). In short, the main steps were the following:

DNA replicative forms of mp18 or mp19 M13 phage were prepared [3]; they possess polylinkers with single cleavage sites for *Eco*RI, *Sac*I, *Kpn*I, *Sma*I, *Bam*HI, *Sal*I, *Pst*I, *Sph*I and *Hind*III restriction enzymes.

Viral cDNA inserts were extracted from PBR322 plasmid and treated by restriction enzymes having sites in the M13 polylinker. DNA fragments ranging between 300 and 500 bases were purified by electrophoresis in low melting point agarose (Gibco-BRL) gel. M13 phage DNA was cleaved by the same enzymes and 5' end phosphates removed by alkaline phosphatase (Boehring) treatment [12]. DNA were then phenol/chloroform-extracted and ethanol-precipitated. After ligation of the insert in the vector, performed with 50 ng of insert in a molar ratio of 3/1 TG1, *E. coli* competent cells were transfected [5].

TG1 recombinant clones were selected in a IPTG- and X-gal-containing medium. White plaques were then checked by hybrization with insert radioactive probe.

Sequencing was then performed using a primer complementary to the 3'-end of the DNA strand to be transcribed. These primers were synthesized in an automated DNA synthesizer (Biosearch 8600).

Sequence data were analysed and assembled with the aid of the program of Queen and Korn [14] of the «Beckman Microgenie» program (March 1985, version Beckman Instruments, Inc.) adapted to the «IBM PC-XT» microcomputer.

RESULTS

cDNA cloning.

Starting material for cDNA synthesis was 10 µg of purified and temperature-denatured viral RNA. When analysed by electrophoresis in alkaline agarose gels, the sizes of the cDNA obtained using oligo-dT as a primer ranged between 1.3 and 6.0 Kb. After binding of heteroduplexes to PBR322, this construction was transfected into *E. coli*-competent cells and we obtained 2×10^5 clones/µg of PBR322.

Some 265 colonies containing 0.3- to 2.0-Kb inserts were studied. Inserts of a larger size than 0.5 Kb very often showed an internal *Pst*I site (results not shown). Their viral specificity was checked, after nick-translation ^{32}P-labelling, by hybridization with purified genomic viral RNA or cellular RNA (fig. 1). Viral-specific inserts were further used for characterization of other inserts.

Insert orientation was established by hybridization with inserts having no *Pst*I site and by restriction endonuclease mapping with enzymes having no or only one cleavage site in PBR322 plasmid.

The location of the insert along the viral genome was determined by Northern blot analysis: full length or purified products of insert restriction cleavage were hybridized with poly(A)$^+$ RNA extracted from infected or non-infected cells. Before hybridization these RNA were electrophoresed in hydroxymethyl Hg-containing agarose gel. Under these experimental conditions, 8 viral-specific poly(A)$^+$ messenger RNA bands were resolved (J. Laporte and C. Cruciere; to be published). They form a specific RNA-nested set as established for other coronaviruses. All the inserts we obtained hybridized with the 8 viral RNA bands (results not shown); they were complementary to the 3' end of the viral genome.

Figure 2 presents the schematic location of the inserts we have studied. The 1.6 insert has a 2,000-nucleotide size and the 5'-end of insert 2.56 is presumably 2,400 nucleotides from the 3'-end of the viral genome. As deduced from the sizes of N and E1 viral proteins, they should cover the whole length of the N gene (1,700 nucleotides) and the beginning of the E1 5'-adjacent gene (320 nucleotides).
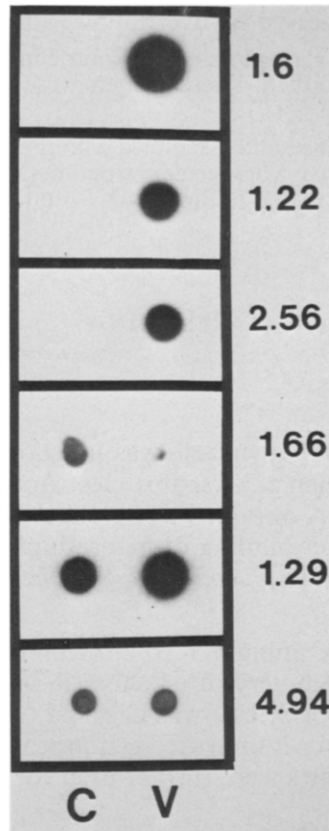
FIG. 1. — *Screening of insert virus specificity.*

Radioactive probes were prepared from insert-containing PBR322 plasmid. These probes were hybridized on nitrocellulose sheets with dots of RNA extracted from non-infected (C) or BECV-F15-infected (V) HRT18 cells. Hybridization was checked by autoradiography. In the experiment shown, inserts 1.6, 1.22 and 2.56 were clearly virus-specific.

cDNA sequencing.

As mentioned above, 400-bp fragments of the cDNA clones were subcloned in mp18 or mp19 M13 phage DNA. Their nucleotidic sequences were determined by sequencing both M13 DNA strands or by multiple sequencing of one strand. We have been able to establish a 1,710-nucleotide sequence from the 3'-end of the genome (fig. 3). This sequence has 2 overlapping open-reading frames (ORF). The main ORF stretches from nucleotide 74 to nucleotide 1,416,
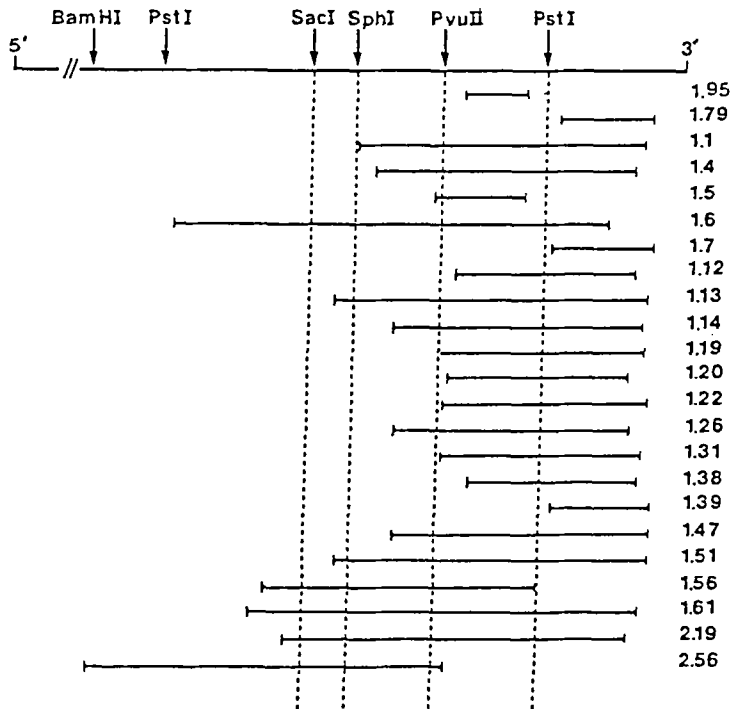
FIG. 2. — *Arrangement of some of the cDNA clones obtained using oligo-dT as primer.*

the smaller one from nucleotide 135 to nucleotide 755 (fig. 3). The first has a coding capacity for a 448-amino-acid protein, the second for a 207-amino-acid protein (fig. 4).

DISCUSSION

We have determined, by cDNA cloning of BECV-F15 genomic RNA using an oligo-dT primer, a sequence of 1,710 nucleotides.

We assume that this sequence comprises the nucleocapsid protein gene sequence.

For every coronavirus so far studied, the gene coding for the N protein is located at the 3'-end of the viral genome. The same conclusion arises from our studies on the BECV-F15 poly(A)$^+$ RNA (to be published).

The largest ORF has a 1,344-nucleotide length and encodes for a 448-amino-acid protein with a molecular weight of 49.4 Kd. Our previous results [4] had shown a 50-Kd molecular weight N protein.

Recently [11] it was described for the US Mebus strain of the related bovine corona virus (BCV), that the N protein gene was at the 3'-end of the viral genome.

Open-reading frames.

Main ORF. — The distance between the first AUG following the initiation codon and this initiation codon is 693 nucleotides. When we compared the sequence around the initiation codon to homologous sequences of different strains of MHV we found the same CTAAAC sequence upstream of the initiation AUG.

Secondary ORF. — The consensus sequence GUAAUGGC surrounding its initiation codon is one of optimal environment for starting mRNA translation [7]. Bunyaviruses and adenoviruses express 2 different proteins from only one gene by having 2 overlapping ORF [7]. So, we cannot exclude the translation of a protein from the secondary ORF. Its predicted molecular weight is 22.9 Kd for 207 amino acids. This protein has a rather high leucine content: 19.8 % compared to 5 % for the N protein. Furthermore, its N-terminal end is hydrophobic and is a potential membrane anchor region. Genes presenting 2 different ORF are also described for other coronaviruses: mRNA$_5$ of JHMvirus [20], mRNA$_D$ of IBV [2] and N protein mRNA of the Mebus BCV strain [11].

Non-coding 3'-end.

This part of the genome may play an important role during the genomic RNA trancription to the complementary minus RNA strand. Sequence homology between BECV-F15 and MHV for the last 100 nucleotides of the coding part is only 59 %, but homology increases to 75 % for the 3'-non-coding end. A 10-nucleotide sequence (GGGAAGAGCT) was found in common at the same place of this gene area for MHV and IBV viruses [2] (fig. 5). We find an identical sequence (except the last T) for BECV-F15 virus between nucleotides 1,631 and 1,640. When looking at the GETV genome se-

FIG. 3. — *Nucleotidic sequence of the 3'-end of BECV-F15 genome.*

This 1,710-nucleotide sequence has 2 large overlapping ORF. M = potential translation initiation codons; U = translation stop codons. — = main ORF; --- = secondary ORF.

```
   1  AUGUGAGCCGAUUGCCGUGCCGUGCAUCCGCUUCACUGAUCUCUUGUUAGAUCUUUUUAUAAUCUAAACUUUAAGG
         M             U                              U

                                          U                 U              U
  74  AUGUCUUUUACUCCUCGGUAAGCCAAUCCAGUACUAGAGCCUCCUCUCGCAAAUCGUUCUGGUAAUGGCAUCCUUA
         M             U         U  U                              M
 147  AGUCGGCCCGAUCAGUCCGACCAAUCUAGAAAUGUUCAAACCCAGCGGGUAGAAGAGCCUCAACCCAAGCAAACUGC
                                  U                        U
                                       M
 220  UACUUCUCAGCAACCAUCAGGAGGGAAUGUUGUACCCUACUAUUCUUGGUUCUCUCGGAAUUACUCAGUUUCAA
                                  M
 293  AAGGGAAAGGAGUUUGAAUUUGCUGAGGGACAAGGUGUGCCCUAUUGCACCAGGAGUCCCAGCUACUGAAGCUA
                   U         U                                     U      U
 366  AGGGGUACUGGUACAGACACAACAGACGUUCUUUUAAAACACGCGAUGGCAACCAGCGUCAAUUGCUGCCACG
                                        U                      M
 439  AUGGUAUUUUUACUAUCUUGGAACAGGACCGCAUGCCAAAGACCAGUAUGGCACCGACAUUGACGGAGUCUUC
         M                                              U
                                  M              M
 512  UGGGUCGCUAGUAACCAGGCUGAUGUCAAUACCCCGGCUGACAUUCUCGAUCGGGACCCAAGUAGCGAUGAGG
              U  U        U                  U                     U     U
                         M                                     M
 585  CUAUUCCGACUAGGUUUCCGCCUGGCACGGUACUCCCUCAGGGUUACUAUAUUGAAGGCUCAGGAAGGUCUGC
              U                                          U
 658  UCCUAAUUCCAGAUCUACUUCACGCGCAUCCAGUAGAGCCUCUAGUGCAGGAUCGCGUAGUAGAGCCAAUUCU
         U                          U
 731  GGCAAUAGAACCCCUACCUCUGGUGUAACACCUGAUAUGGCUGAUCAAAUUGUCAGUCGUUGUUUUGGCAAAAC
            U                     U       U
                              U
 804  UUGGCAAGGAUGCCACUAAGCCACAGCAAGUAACUAAGCAGACUGCCAAAGAAAUCAGACAGAAAAUUUUGAA
                    U                U
              M                U                                      U
 877  UAAGCCCCGCCAGAAGAGGAGCCCCAAUAAACAAUGCACUGUUCAGCAGUGUUUUGGGAAGAGAGCCCCAAU
         U                          U  U  M
 950  CAGAAUUUUGGUGGUGGAGAAAUGUUAAAACUUGGAACUAGUGACCCACAGUUCCCCAUUCUUGCAGAACUCG
                              M         U  U
 1023 CACCCACAGCUGGUGCCGUUUUUCUUUUGGAUCAAGAUUAGAGUUGGCCCAAAGUGCAGAAUUUGUCUGGGAAUCU
                                                         U
 1096 UGAUGAGCCCCAGAAGGAUGUUUAUGAAUUACGCUACAAUGGCCGCAAUUAGAUUUGAUAGUACACUUUCAGGU
            U  U                  U                      U   U      U
              M                  M      M                M
 1169 UUUCAGAGACCAUAAUGAAGGUGUUGAAUGAGAAUUUCAAUGCAUAUCAACAACAAGAUGGUAUGAUGAAUAUGA
             M                 U                             M    M       M
               U  U                U  M              U  M                 U  U     U
 1242 GUCCAAAACCACAGCGUCAGCGUGGUCAGAAGAAUGGACAAGGAGAAAAUGAUAAUAUAAGUGUUCCAGCGCC
                                                                U  U
                                 M          M          M
 1315 CAAAAGCCGUGUGCCAGCAAAAUAAGCAGUAGAGAGUUGACUGCAGAGGACAUCAGCCUUCUUAAGAAGAUGGCAU
                   U        U                                       M         M
                                 U                                            M
 1388 GAGCCCUAUACUGAAGACACCUCAGAAAUAUAAGAGAAUGAACCUUAUGUCGGUACCUCGGUGGCAACCCCUCG
         U      U                        U
                         M          M
 1461 CAGGCAAAGUCGGGAUAAGGCCAUUCUCUAUCAGAAUGGAUGUCUUGCUGCCUAUAAUAGAUAGCACAAGCUUAUAG
                     U                  M                      U              U
                                        M                  U  U
 1534 CACACUAUAGAUUUAAUUAGUUGAAAAGUUUUCUGUGGGUAAAUGUAUAGUCGUUGGAGAAAGCUGAAAGACUUGCGGA
              U          U               U        U         U
                 U            U                M     U
 1607 AGUAAAUUGCCCACAAGUGCCCAAGCGGCAACAGCCAGCAUGUUAAGUUGCCACCCAGUAAUUAGUAAAUGGAAUG
            U                                  U              U        U     U  U  M  U
                                                                    M              U  M
 1680 AAGUUUAAUUAUGCCCAAUUCGAAGAAUCACA
             U
              M
```

quence [15], we observe the same sequence (except the first G) in the 3'-non-coding end between nucleotides 1,923 and 1,931. Our analysis strengthens Bournsell's hypothesis; this sequence, well conserved among the coronaviridae family, should have an important function during RNA replication as it is an RNA-polymerase fixation site.

N-protein-predicted amino acid sequence.

BECV-F15 N protein has very strong homology with the same protein of JHM virus (70.3 %) (fig. 6) and only 25.2 % and 24.1 %, respectively, with TGE and IBV virus N proteins. These coronavirus N proteins are phosphorylated on their serine residues [18]. Our results show 43 serine residues in BECV-F15 nucleocapsid protein (9.6 % of the total amino acids). For this virus and for JHM, TGE and IBV viruses we find 2 main areas where serine residues are clustered. For BECV-F15 and JHM viruses they are in homologous areas (nucleotides 9 to 19 and nucleotides 191 to 220) of low overall homology (58 % and 53 %). One serine cluster is common to the 4 viruses. This fact is striking because of the low sequence homology between these viruses.

It was previously established [21, 1] that N protein genomic RNA binding sites are located in the basic portions of the protein. For the complete sequence there is an excess of 19 basic residues compared to acidic residues. There are 5 basic-rich regions which are found in homologous areas of MHV, TGE and IBV viruses. Concerning BECV-F15 and MHV, 4 of these areas have 90 % homology. The fifth has only 60 % homology but is also serine-rich and possesses a sequence in common with TGE and IBV viruses (amino acids 193 to 222). It may have a more specific function in protein/RNA recognition.

We also observed a strong sequence homology, not yet described, in the first part of the N-terminal end of the N proteins of BECV-F15, BCV, MHV, TGEV and IBV viruses:

Virus	Amino acid nb	Amino acid sequence
BECV-F15	118 to 134	QLLPRWYFYYLGTGPHA
JHMV	121 to 135	QLLPRWYFYYLGTGP
GETV	89 to 101	RW FYYLGTGPHA
IBV	91 to 102	WYFYY GTGP A

This sequence has no peculiar properties: 9 hydrophilic and 8 hydrophobic residues. The biological significance of these findings is not known.

In conclusion, we have noticed that there are only minor changes between BECV-F15 and BCV Mebus strain N proteins. Work is in progress to sequence the other virus genes and to find out how similar in fact these two last viruses are. Because of the antigenic differences established by monoclonal antibody screening, the specificities should be found on the gene coding for the spike gp105 protein.

```
   1  AUGUGAGCGAUUGCGUGCGUGCAUCCGCUUCACUGAUCUCUUGUUAGAUCUUUUUAUAAUCUAAACUUUAAGG

                                                              M   A S   L
      M  S  F  T  P  G  K  Q  S  S  S  R  A  S  S  G  N  R  S  G N  G I   L
  74  AUGUCUUUUACUCCUGGUAAGCAAUCCAGUAGUAGAGCGUCCUCUGGAAAUCGUUCUGGUAAUGGCAUCCUUA

      S  G  P  I  S  P  T  N  L  E  M  F  K  P  G  V  E  E  L  N  P  S  K  L
      K  W  A  D  Q  S  D  Q  S  R  N  V  Q  T  R  G  R  R  A  Q  P  K  Q  T  A
 147  AGUGGGCCGAUCAGUCCGACCAAUCUAGAAAUGUUCAAACCAGGGGUAGAAGAGCUCAACCCAAGCAAACUGC

      L  L  L  S  N  H  Q  E  G  M  L  Y  P  T  I  L  G  S  L  E  L  L  S  F  K
      T  S  Q  S  P  S  G  G  N  V  V  Y  P  Y  S  W  F  S  G  I  T  Q  F  Q
 220  UACUUCUCAGCAACCAUCAGGAGGGAAUGUUGUACCCUACUAUUCUUGGUUCUCUGGAAUUACUCAGUUUCAA

      R  E  R  S  L  N  L  L  R  D  K  V  C  L  L  H  Q  E  S  Q  L  L  K  L
      K  G  K  E  F  E  F  A  E  G  Q  G  V  P  I  A  P  G  V  P  A  T  E  K  A
 293  AAGGGAAAGGAGUUUGAAUUUGCUGAGGGACAAGGUGUGCCUAUUGCACCAGGAGUCCCAGCUACUGAAGCUA

      R  G  T  G  T  D  T  T  D  V  L  L  K  H  A  M  A  T  S  V  N  C  C  H
      K  G  Y  W  Y  R  H  N  R  R  S  F  K  T  R  D  G  N  Q  R  Q  L  L  P  R
 366  AGGGGUACUGGUACAGACACAACAGACGUUCUUUUAAAACACGCGAUGGCAACCAGCGUCAAUUGCUGCCACG

      D  G  I  F  Y  T  I  L  E  G  Q  D  R  M  P  A  K  T  S  Y  M  A  P  T  L  D  T  E  V  S  S
      W  Y  F  Y  T  I  L  G  T      R  H  M  A  K  T  S  Y      G  A  P  T  D  I  D  T  E  S
 439  AUGGUAUUUUUACUAUCUUGGAACAGGACCGCAUGCCAAAGACCAGUAUGGCACCGACAUUGACGGAGUCUUC

      G  S  L  V  T  R  L  M  S  I  P  R  L  T  F  S  I  G  D  T  Q  S  V  A  M  R
      A  I  P  T  R  F  P  P  P  G  T  V  L  P  Q  Y  Y  I  E  G  S  G  R  S  A
 512  UGGGUCGCUAGUAACCAGGCUGAUGUCAAUACCCCGGCUGACAUUCUCGAUCGGGACCCAAGUAGCGAUGAGG

      L  F  R  L  G  F  R  L  A  R  Y  S  L  R  V  T  I  L  K  A  Q  E  G  L
      P  N  S  R  S  T  S  R  A  H  P  V  E  P  L  V  Q  D  R  V  V  E  P  I  L
 585  CUAUUCCGACUAGGUUUCCGCCUGCACGGUACUCCCUCAGGGUUACUAUAUUGAAGGCUCAGGAAGGUCUGC

      L  L  I  P  D  L  L  H  A  H  P  V  E  P  L  V  Q  D  R  V  V  E  P  I  L
      G  N  R  T  P  T  S  R  A  S  S  R  A  S  S  A  G  S  R  S  R  A  N  S
 658  UCCUAAUUCCAGAUCUACUUCACGCGCAUCCAGUAGAGCCUCUAGUGCAGGAUCGCGUAGUAGAGCCAAUUCU

      A  I  E  P  L  P  L  V
      G  N  R  T  P  T  S  G  V  T  P  D  M  A  D  Q  I  V  S  L  V  L  A  K
 731  GGCAAUAGAACCCCUACCUCUGGUGUAACACCUGAUAUGGCUGAUCAAAUUGUCAGUCUUGUUUUGGCAAAC

      L  G  K  D  A  T  K  P  Q  Q  V  T  K  Q  T  A  K  E  I  R  Q  K  I  L  N
 804  UUGGCAAGGAUGCCACUAAGCCACAGCAAGUAACUAAGCAGACUGCCAAAGAAAUCAGACAGAAAUUUUGAA

      K  P  R  Q  K  R  S  P  N  K  Q  C  T  V  Q  Q  C  F  G  K  R  G  P  N
 877  UAAGCCCCGCCAGAAGAGGAGCCCCAAUAAACAAUGCACUGUUCAGCAGUGUUUUGGGAAGAGAGGCCCCAAU

      Q  N  F  G  G  G  E  M  L  K  L  G  T  S  D  P  Q  F  P  I  L  A  E  L
 950  CAGAAUUUUGGUGGUGGAGAAAUGUUAAAACUUGGAACUAGUGACCCACAGUUCCCCAUUCUUGCAGAACUCG

      A  P  T  A  G  A  F  F  F  G  S  R  L  E  L  A  K  V  Q  N  L  S  G  N  L
1023  CACCCACAGCUGGUGCGUUUUUCUUUGGAUCAAGAUUAGAGUUGGCCAAAGUGCAGAAUUUGUCUGGGAAUCU

      D  E  P  Q  K  D  V  Y  E  L  R  Y  N  G  A  I  R  F  D  S  T  L  S  G
1096  UGAUGAGCCCCAGAAGGAUGUUUAUGAAUUACGCUACAAUGGCGCAAUUAGAUUUGAUAGUACACUUUCAGGU

      F  E  T  I  M  K  V  L  N  E  N  L  N  A  Y  Q  Q  Q  D  G  M  M  N  M
1169  UUUGAGACCAUAAUGAAGGUGUUGAAUGAGAAUUUGAAUGCAUAUCAACAACAAGAUGGUAUGAUGAAUAUGA

      S  P  K  P  Q  R  Q  R  G  Q  K  N  G  Q  G  E  N  D  N  I  S  V  A  A  P
1242  GUCCAAAACCACAGCGUCAGCGUGGUCAGAAGAAUGGACAAGGAGAAAAUGAUAAUAUAAGUGUUGCAGCGCC

      K  S  R  V  Q  Q  N  K  S  R  E  L  T  A  E  D  I  S  L  L  K  K  M  D
1315  CAAAAGCCGUGUGCAGCAAAAUAAGAGUAGAGAGUUGACUGCAGAGGACAUCAGCCUUCUUAAGAAGAUGGAU

      E  P  Y  T  E  D  T  S  E  I
1388  GAGCCCUAUACUGAAGACACCUCAGAAAUAUAAGAGAAUGAACCUUAUGUCGGUACCUGGUGGCAACCCCUCG

1461  CAGGAAAGUCGGGAUAAGGCAUUCUCUAUCAGAAUGGAUGUCUUGCUGCUAUAAUAGAUAGAGAAGGUUAUAG

1534  CAGACUAUAGAUUAAUUAGUUGAAAGUUUUGUGUGGUAAUGUAUAGUGUUGGAGAAAGUGAAAGACUUGCGGA

1607  AGUAAUUGCCCCACAAGUGCCCAAGGGGAAGAGCCAGCAUGUUAAGUUGCCACCCAGUAAUUAGUAAAUGAAUG

1680  AAGUUAAUUAUGGCCAAUUGGAAGAAUCACA
```

FIG. 4. — *Amino acid sequence of the proteins predicted from the main and secondary ORF.*

RÉSUMÉ

SÉQUENCE ET ANALYSE DU GÉNOME
DU CORONAVIRUS ENTÉRITIQUE BOVIN (F15)

I. — Séquence du gène codant pour la protéine nucléocapsidique ; analyse de la protéine déduite

Nous avons cloné l'ARN génomique du coronavirus entéritique bovin F15 (BECV-F15), dans le plasmide PBR322 après avoir préparé le cDNA correspondant à l'aide d'une amorce oligo-dT : 265 clones ont été étudiés. Leur hybridation avec les ARN poly(A)[+] extraits des cellules infectées nous a permis de les localiser à l'extrémité 3'-terminale du génome.

Ces clones ont été séquencés par la technique de Sanger, après sous-clonage dans l'ADN du phage M13. Nous avons déterminé une séquence de 1.710 nucléotides correspondant au gène codant pour la protéine N virale. Elle présente deux cadres ouverts de lecture (ORF) chevauchants. On observe à l'extrémité 3'-terminale non codante du génome une séquence de 8 nucléotides observée également dans la région homologue des virus MHV, GET et IBV. Cette séquence pourrait être le site de fixation de l'ARN polymérase.

Le premier AUG du plus petit ORF possède en amont une séquence nucléotidique qui en fait un site d'initiation potentiellement fonctionnel. La séquence du produit primaire de traduction que l'on en déduit est un polypeptide de 207 acides aminés (22,9 Kd) à haute teneur en leucine (19,8 %) ayant une extrémité N-terminale hydrophobe.

Le plus grand ORF a une capacité de codage de 448 acides aminés (49,4 Kd), correspondant à la masse moléculaire de la protéine N. La protéine déduite contient 43 résidus sérine (9,6 % des acides aminés), qui peuvent être phosphorylés et impliqués dans la liaison entre la protéine N et l'ARN génomique. Cette protéine présente également 5 régions fortement basiques, et l'une d'entre elles est également riche en sérine et a une forte homologie de séquence avec la région homologue des protéines N des virus MHV, GET et IBV. En outre, la première partie de l'extrémité N-terminale montre un enchaînement de 12 acides aminés (PRWYFYYLGTGP) très conservé entre ces quatre même virus.

FIG. 5. — *Nucleotide sequence homology between the 3'-end of the genomes of BECV-F15 and MHV-JHM viruses.*

```
   1     AUGUCAGCGAUU GCGUGCGU  GCAUCCGCUUCACU   GAUCUCUU   GUUAGAUCUUUUUAUAAUCU
         II  III IIII IIII III  I J II II III    I IIIII   III  I  I  I IIIIII
   1     UAUAAGAGUGAUUGGCGUCCGUACGUACCCUCUCUACUCUAAAACUCUUGUAGUUUAAAUCUAAUCUAAUCU

  63     AAACUUUAAGGAUGUCUUUUACUCCU GGUAAG CAAU CCAGUAGUAGAGCGUCCUCUCGGAAAUCGUUCUG
         IIIIIIIIIIIIIIIIIIIII  IIII II III  III II IIII III   IIIIIIIIIII II III
  73     AAACUUUAAGGAUGUCUUUUGUUCCUGGGCCAAGAAAAUGCCGGUAGCAGAAGCUCCUCUCGGAAACCGCCUG

 132     GUAAUGGCAUCCUUAAG          UGGGCCGAUCAGUCCGACCAAUCUAGAAAUGUUCAAACCACGGGUA
         IIIIIII IIIII III         IIIII II II  IIII I     IIII  IIIII  II II I
 145     GUAAUGGCAAUCCUCAAGAAGACCACUUGGCCUGACCAAACCGAGCGCGGGUUAAAUAAUCAAAAUAGAGGCA

 195     GAAGAGCUCAACCCAAGCAAACUGCUACUUCUCAGCAACCAUCAGGAGGGAAUGUUGUACCCUACUAUUCUU
         III    III IIIIIIII IIIII III II    IIIII     IIII III II III I II IIII
 217     GAAAGAAUCAGCCCAAGCAGACUGCAACUACU   CAACCCAAUUCCGGGAGUGUGGCUCCCCAAUUCACUCUU

 267     CGUUCUCUGGAAUUACUCAGUUUCAAAAGGGAAAGGAGUUUGAAUUUGCUGCAGGACAAGGCGUGCCCUAUUG
         IIII II II IIIII II II II IIIIIIII IIIIII I IIIII  I IIIIIIII IIIIIIIIII
 286     GGUUUUCGGGCAUUACCCAAUUCCAGAAGGGAAAAGAGUUUCAGUUUGCACAAGGACAAGGAGUGCCUAUUG

 339     CACCAGGCAGUCCCAGCUACUCAAGCUAAGGGGUACUGGUACAGCACACAACAGAGCGUUCUUUUAAAACACGCG
         I    III IIIIIIII I  I   IIIII II IIIIIIIIIIIIIIIII IIIIIII IIIIIIIIII  I
 358     CCAAUGGCAAUCCCAGCUUCACAGCAAAAGGGAUAUUGGUACAGACACAACCGACGUUCCUUUAAAACACCUG

 411     AUGGCAACCAGCCGUCAAUUGCUGCCACGAUGGUAUUUUUACUAUCUUGGAACAGGACCGCAUGCCAAAGACC
         IIIII I III   II I IIIII  IIIIIIIIIIIIIIIIIIIIIIIIIIII II IIII   I
 430     AUGCCCAGCAGAAGCAGCUACUCGCCCAGAUGGUAUUUUUACUAUCUUGGAACAGGCGCCCUAUGCUGCCGCAG

 483     AGUAUGGCACCGACAUUGACGGAGUCUUCUGGGUCGCUAGUAACCAGGCUGAUGUCAAUACCCCGGCUGACA
         IIIIIIII III II II IIIII   IIIIIIII II I IIIII II  I III II
 502     AGUAUGGCGACGAUAUCGAAGGCGUUGUCUGGGUCGCAAGCCAACAGCCCGAGACUAGCGACCUCUGCCCGAUA

 555     UUCUCGAUCGGGACCCAAGUAGCGAUGAGGCUAUUCCGACUAGGUUUCCGCCUGGCACGGUACUCCCUCAGG
         II I II   IIIIIIIIIIIIIII  IIIIIIIIIIII IIIIIIII IIII I IIIIII II  I
 574     UUGUUGAAAGGGACCCAAGUACCCAUGAGGCUAUUCCUACUAGGUUUGCGCCCGGUACGGUAUUGCCUCAAG

 627     GUUACUAUAUUGAAGGCUCAGGAAGGUCUGCUCCCUAAAUUCCAGAUCUACUUCACGCGCAUCCAGUAGAGCCU
         III  III IIIIIIIIIIIIIIIIIIIIIIII III I    IIII  II IIII   III I  I I
 646     GUUUUUAUGUUGAAGGCUCAGGAAGCGUCUGCCACCUGCUACUCCGAUCUGGUU  CGCG  GCCACAAUCCCGU

 699     CUAGUGCCAGGAUCCGCUAGUAGAGCCCAAUUCUCGGCCAAUAGAACCCCUACCUCUGGUGUAACACCUGAUAUGG
         I  IIIIII    IIII II III   I I     III IIIII  IIIII IIIIIIIIII
 714     GGGCCCAAAUAAUCGCG  CUAGAAGCAGUUCCAAACCAGCGCCAGCCUGCCUCUACUGUAAAGCCAAGCCUGAUAUGG

 771     CUGAUCAAAUUGUCAGUCUUGUUUUUGGCAAAACUUGGCCAAGGAUGCCACUAAGCCACAGCCAAGUAACUAACC
         I II  IIIIII    IIIIIIIIIIIII II II II II IIIIII    IIII  IIIIIIIIII IIII
 784     CCGAAGAAAUUGCUGCUCUUGUUUUUGGCCUAAGCUCCGGUAAAAGAUGCCGGCCCAGCCUAAGCAAGCAAACAAAGCC

 843     AGACUGCCAAAGAAAUCAGACAGAAAAUUUUGCAAUAAGCCCCCGCCAGAAGAGGAGCCCCAAUAAACAAUGCA
         I I IIIIIIIIII IIII IIIIIIIIIII II IIIII II II IIIIIII II II II II III
 856     AAAGUGCCAAAGAAGUCAGGCAGAAAAUUUUAAAACAAGCCUCGUCAAAAGAGGACUCCAAACAAGCAGUUCC

 915     CUCUUCAGCAGUGUUUUGGGAAGAGAGGCCCCAAUCAGAAUUUUGGGUGGUGGCAGAAAAUGUUAAAACUUGGAA
         I  I IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII II    IIIIIIIIIIIIIIIIIIIIIIIIIII
 928     CAGUGCCAGCCAGUGUUUUGGAAAGACAGGCCCCAAUCAGAAUUUUUGGAGGCCCUGAAAUGUUAAAACUUGGAA

 987     CUAGUGACCCCACAGUUCCCCAUUCUUGGCAGAAACUCGCCACCCACAGCCUGCGUUUUUCUUUGGCAUCAAGAU
         IIIIIII  IIIIIIIIIIIIIIIIIIIIIIIIIIIIII  I II II IIIIIIIIIII II IIIIIIIIIIIIII II
1000     CUAGUCAUCCACAGUUCCCCAUUCUUGCAGAGUUGGCCCCAACAGCUGGUGCCUUCUUCUUUUUGGCAUCUAAAU

1059     UAGAGUUGGGCAAAGUGCAGAAUUUUGCUCUGGCAAUCUUGAUGAGCCCCCAGAAGGAUGUUUAUGGAAUUUACCUU
         IIII IIII IIII    IIII     IIIII   I   IIIII  III   II IIIII IIIII  I I  I
1073     UAGAAUUGGUCAAA    AAGAA    CUCUGGGUGGUGCUGAUGGACCCACCAAAGAUGUGUAUGACCUGCAAU

1131     ACAAUGGCGCCAAUUAGAUUUGAUAGUACACUUUCAGGGUUUGAGACCAUAUGAAGCGUUGAAUGAGAAUU
         I     II III IIIIIIIIIIIIIIIIII I    I IIIIIIIIII II  IIIII IIIIIIIIIIIIIII
1138     AUUCAGCUGCAGUUAGAUUUGAUAGUACUCUACCUGGUUUUGAGACUAUCAUGAAAGUGUUGAAUGAGAAUU

1203     UGAAUGCCAUAUCAACAACAAGAUGGU AUGAUGA AUA UGAGUCCAAAACCACAGCCGUCAGCCGUCGUCACAGA
         IIIIIII II II  I IIIIIIIII II  II  II IIII II II III II I   II I II I
1210     UGAAUGCCUACCAGAAACAAGAUGGCGUGGUGCCAGAUGUAGUGAGCCCUAAGCCUCACAGAAGAAAGAGACGGGACAA

1272     AG AAUGG  ACAACGA GAAAAUGAUAAUAUAAGCGUUGCCAGCCGCCCAAAAGCCCGUGUGCCAGCAAAAUA
         II II II  I II II III   IIIIII IIII IIIIII   IIIIIIIII III  IIIIIII IIII
1282     AGCAAAAGGCUCAGAAACAUGAAGUAGCAUAAUCGUAACGCGUUGCAAAGCCCAAAAGCUCUCUGUGCCAGCGAAAUG

1338     AGAGUAGAGAGUUGACUGCAGAGGACAUCAGCCUUCU   UAAGA   AGAUG    GAUGCAGCCCUAU
         IIIIIIIIIIII II  I IIIII   IIII   I III   I   I III I
1354     UAAGUAGACAGAGUUAACCCCUGCAGGAUCGCAGCCUUCUGGCUCAGAUCCUAGAUGAUGCCGUAGUGCCACAUC

1397     ACU  GAAGACACCUCAGAAAUAU  AAGAGAAUGAACCUUAUGUCGGCACCUGGUGCCAACCCUCCGCAGGA
         I  IIIII   III I I  I IIIIIIIIII I  IIII    IIIII  IIIII IIIIIIIIII II
1426     CGUUAGAAGAUGACUCUAAAUGUGUAAAGAGAAUGAAUCCUAUGUCGGCACUCGGUGUAACCCCUCGCGAGA

1466     AAGUCGGGAUAAGGCCACUCUCUAUCAGAAUGCGAUGUCUUGCUGCUAUAAUGAUAGAGAAGCUUAUACCAGA
         IIIIIIIIIII I  IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII  IIIIIIIIIIIIII  IIII IIIII
1498     AAGUCGGGAUAGGCACACUCUCUAUCAGAAUGGUGUCUUGCUGCAUAAACAGAUAGAGAAGGUUGUGGCCAGA

1538     CUAUAGAUUAAUUAGUUGAAAG UUUUGCGUGCGGUAAUGUAUAGUGUUGCGAGAAAGU G AAAGACUUGCG
         I  I  II IIIIIIIIIIIIIIII  III    II  I II IIII  IIIII  I I II I I II
1570     CCCUGUAUCAAUUAGUGUUGAACUAGCAAAUAUACAGAAU GUCUGUACAGAAAGUUAGCAAGUCCUACGUC

1605     GAA  GUA  AUUGCCCACAAGUCCCCAAGGGGAAGAGC   CAGCCAUGUUUAAGUUGCCCACCCAGUAAUUAG
         II   II  I  I I I I I IIII   IIIIIIIII   II II I II   I II  I I    III
1641     UAACCAUAAGAACGGCGAUAGGCCGCCCCCUGGGAAGAGCUCACAUCAGGGUACUAUUCCUGCCAAUGCCCUAG

1670     UAAAUGAAUGAAGUGUAAAUUAUGGCCCAAUUGGAAGAAUCACA
         IIIIIIIIIIIIIIIII  I  IIIIIIIIIIIIIIIIIIIIIIII
1723     UAAAUGCAAUGAACUUGAUCAUGGCCAAUUGGAAGAAUCAC
```

```
  1   M S F 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
      * * * *     * *         *   * * * * *   * * * * * *         * * * *
  1   M S F V P G Q E N A G S R S S S G N R A G N G I L K K T T W A D Q T E R

 34   R N V Q T R G R R A 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
        *     *   * * *         * * * * * *   *     *       * * *   * * * * * * *
 38   L N N Q N R G R K N Q   P K Q T A T T Q     P N S G S V V P H Y S W F S G I

 71   Q F Q K G K E F E F A E G Q G V P I A P G V P A T E A K G Y W Y R H N R
      * * * * * * *     * *     * * * * * * *   *     * *       * * * * * * * * *
 74   Q F Q K G K E F Q F A Q G Q G V P I A N G I P A S Q Q K G Y W Y R H N R

108   S F K T R D G N Q R Q L L P R W Y F Y Y L G T G P H A K D Q Y G T D I D
      * * * *     *       * * * * * * * *,* * * *       * *         * *
111   S F K T P D G Q Q K Q L L P R W Y F Y Y L G T G P Y A G A E Y G D D I E

145   V F W V A S N Q A D V N T P A D I L D R D P S S D E A I P T R F P P G T
      *   * * * *   * *       *   * * *       * * * * *   * * * * * * *   * * *
148   V V W V A S Q Q A E T R T S A D I V E R D P S S H E A I P T R F A P G T

182   L P Q G Y Y I E G S G R S A P N S R S T S R A S S R A S S A G S R S R A
      * * * *   *   * * * * * * *     * * *     * *       * *         *   *
185   L P Q G F Y V E G S G R S A P A S R S G S R P Q S R G P N N     R A R S S

219   S G N R T P T S G V T P D M A D Q I V S L V L A K L G K D A T K P Q Q V T
      *       *   *   *   *     * * * *     *       * * * * * * * * *   *     * * *
220   S N Q R Q P A S T V K P D M A E E I A A L V L A K L G K D A G Q P K Q V T

256   K Q T A K E I R Q K I L N K P R Q K R S P N K Q C T V Q Q C F G K R G P N
      * *   * * * *     * *   * * * * * * * *     * * * * *   * * * * * * * * * *
257   K Q S A K E V R Q K I L N K P R Q K R T P N K Q C P V Q Q C F G K R G P N

293   Q N F G G G E M L K L G T S D P Q F P I L A E L A P T A G A F F F G S R L
      * * * * *     * * * * * * * * * * * * * * * * * * * * * * * * * * * * *   *
294   Q N F G G P E M L K L G T S D P Q F P I L A E L A P T A G A F F F G S K L

330   E L A K V Q N L S G N L D E P Q K D V Y E L R Y N G A I R F D S T L S G F
      * *   *       *       *   *     * * * * * *   *     * *   * * * * * *   * *
331   E L V K K N S     G G A     D G P T K D V Y E L Q Y S G A V R F D S T L P G F

367   E T I M K V L N E N L N A Y Q Q Q D G M M N M   S P K P Q R Q R G Q K N
      * * * * * * * * * * * * * *     * * *           * * * * * *   * *   *
366   E T I M K V L N E N L N A Y Q N Q D G G A D V V S P K P Q R K R G T K Q K

402   G Q G E N     D N I S V A A P K S R V Q Q N K S R E L T A E D I S L L K K
      *             * *   * *   * * *       * *   * * * * * *   * *   * * *
403   A Q K D E V D N V S V A K P K S S V Q R N S V R E L T P E D R S L L A Q I

437   M D E       P Y T   E D T S E I
      *                   * * *
440   L D D G V V P D G L E D D S N V
```

FIG. 6. — *Amino acid sequence homology of the N proteins of BECV-F15 and JHM-MHV viruses.*

Les séquences des protéines N de la souche Mebus du BCV et du BECV-F15 ne présentent que des différences mineures.

Mots-clés: Coronavirus, Protéine, Nucléocapside, Génome; Souche BECV-F15, Séquence de la protéine N.

REFERENCES

[1] Armstrong, J., Smeekens, S. & Rottier, P., Sequence of the nucleocapsid gene from murine coronavirus MHV-A59. *Nucl. Acids Res.,* 1983, **11**, 883-892.

[2] Bournsell, M.E.G., Binns, M.M., Foulds, I.J. & Brown, T.D.K., Sequences of the nucleocapsid genes from two strains of avian infectious bronchitis virus. *J. gen. Virol.,* 1985, **66**, 573-580.

[3] Crouse, G.F., Frischauf, A. & Lehrach, H., An integrated and simplified approach to cloning into plasmids and single-stranded phages, *in* «Methods in Enzymology» (Grossman, L.), **101** (pp. 78-89). Academic Press, New York, London, 1983.

[4] Gouet, P., Contrepois, H.C., Dubourguier, Y., Riou, R., Scherrer, R., Laporte, J., Vautherot, J.F., Cohen, J. & L'Haridon, R., The experimental production of diarrhoea in axenic and gnotobiotic calves with enteropathogenic *E. coli,* rotavirus, coronavirus and in combined infections of Rotavirus and *E. coli. Ann. Rech. Vét.,* 1978, **9**, 433-440.

[5] Hanahan, D., Studies on transformation of *Escherichia coli* with plasmids. *J. mol. Biol.,* 1983, **166**, 557-580.

[6] King, B. & Brian, D.A., Bovine coronavirus structural proteins. *J. Virol.,* 1982, **42**, 700-707.

[7] Kozak, M., Comparison of initiation of protein synthesis in procaryotes, eucaryotes and organelles. *Microbiol. Rev.,* 1983, **47**, 1-45.

[8] Lai, M.M.C., Baric, R.S., Brayton, P.R. & Stohlman, S.A., Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. *Proc. nat. Acad. Sci.* (Wash.), 1984, **81**, 3626-3630.

[9] Laporte, J. & Bobulesco, P., Polypeptide structure of bovine enteritic coronavirus: comparison between a wild strain purified from feces and a HRT18 cell adapted strain, *in* «Biochemistry and Biology of Coronaviruses» (ter Meulen V. *et al.*) (pp. 181-184). Plenum Press, New York, 1981.

[10] Laporte, J., Bobulesco, P. & Rossi, F., Une lignée cellulaire particulièrement sensible à la réplication du coronavirus entéritique bovin: les cellules HRT18. *C.R. Acad. Sci.* (Paris) *(série D),* 1980, **290**, 623-626.

[11] Lapps, W., Hogue, B.G. & Brian, D.A., Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. *Virology,* 1987, **157**, 47-57.

[12] Maniatis, T., Fritsch, E.F. & Sambrook, J., *in* «Molecular cloning: a laboratory manual». Cold Spring Harbor Laboratory, New York, 1982.

[13] Mebus, C.A., Stair, E.L., Rhodes, M.B. & Twiehaus, M.J., Neonatal calf diarrhea: propagation, attenuation and characteristics of a corona-like agent. *Amer. J. vet. Res.,* 1973, **34**, 145-150.

[14] Queen, C. & Korn, L.J., A comprehensive sequence analysis program for the IBM personal computer. *Nucl. Acid Res.,* 1984, **12**, 581-599.

[15] Rasschaert, D., Gelfi, J. & Laude, H., Enteric coronavirus TGEV: partial sequence of the genomic RNA, its organization and expression. *Biochimie,* 1987, **69**, 591-600.

[16] ROYCHOUDHURY, R., JAY, E. & WU, R., Terminal labeling and addition of homopolymer tracts to duplex DNA fragments by terminal deoxynucleotidyl transferase. *Nucl. Acid. Res.,* 1976, **3,** 328-334.

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

[18] SIDDELL, S.G., BARTHEL, A. & ter MEULEN, V., Coronavirus JHM: a virion-associated protein kinase. *J. gen. Virol.,* 1981, **52,** 235-243.

[19] SIDDELL, St., WEGE, H. & ter MEULEN, V., The structure and replication of coronaviruses. *Curr. Top. Microbiol. Immunol.,* 1982, **99,** 131-163.

[20] SKINNER, M.A., EBNER, D. & SIDDELL, S.G., Coronavirus MHV-JHM mRNA 5 has a sequence arrangement which potentially allows translation of a second, downstream open reading frame. *J. gen. Virol.,* 1985, **66,** 581-592.

[21] SKINNER, M.A. & SIDDELL, S.G., Coding sequence of coronavirus MHV-JHM mRNA 4. *J. gen. Virol.,* 1985, **66,** 593-596.

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

[23] VAUTHEROT, J.F. & LAPORTE, J.L., Utilization of monoclonal antibodies for antigenic characterization of coronaviruses. *Ann. Rech. Vét.,* 1983, **14,** 437-444.

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

[25] VAN DER WERF, S., BREGEGERE, F., KOPECKA, H., KITAMURA, N., ROTHBERG, P.C., KOURILSKY, P., WIMMER, E. & GIRARD, M., Molecular cloning of the genome of poliovirus type 1. *Proc. nat. Acad. Sci.* (Wash.), 1981, **78,** 5983-5987.