

Nucleotide sequence of the glycoprotein S gene of bovine enteric coronavirus and comparison with the S proteins of two mouse hepatitis virus strains

Pascal Boireau,¹ Catherine Cruciere¹ and Jaques Laporte^{2*}

¹Laboratoire de Virologie Moléculaire, L.C.R.V., C.N.E.V.A., 22 rue Pierre Curie, BP 67, 94703 Maison-Alfort Cedex and ²Station de Virologie et d'Immunologie Moléculaires, I.N.R.A., C.R.J.J., Domaine de Vilvert, 78350 Jouy-en-Josas, France

The gene encoding the spike glycoprotein (S) of bovine enteric coronavirus (BECV) was cloned and its complete sequence of 4092 nucleotides was determined. This sequence contained a single long open reading frame with a coding capacity of 1363 amino acids (M_r 150747). The predicted protein had 19 *N*-glycosylation sites. A signal sequence comprising 17 amino acids was observed starting from the first methionine residue. A potential peptidase cleavage site was located between amino acids 763 and 767. These cleavages explain the maturation of the primary product of the S gene

to S1 (M_r 104692) and S2 (M_r 84175) spike structural proteins. Two amphipathic α -helices (amino acids 1007 to 1077 and 1269 to 1294) which may constitute the 12 nm stalk of the viral spike were also observed; another α -helix (amino acids 1305 to 1335) may be involved in the anchorage of the spike in the viral membrane. Comparison of this protein sequence to the described homologous mouse hepatitis (MHV) strain A59 and MHV-JHM S protein sequences led us to suggest that MHV-A59 and MHV-JHM S genes could be derived from a deletion of the BECV S gene.

Bovine enteric coronavirus (BECV) was first identified by electron microscopy in faecal samples of calves suffering from acute enteritis (Mebus *et al.*, 1973*a*). The involvement of BECV in the aetiology of diarrhoeal diseases has been suggested in several studies (Mebus *et al.*, 1973*b*; Bridger *et al.*, 1978; Gouet *et al.*, 1978). During the acute stage of infection, virus particles are excreted in large amounts and have been identified within brush border cells of the small intestine and in differentiated colonic epithelial cells. Although propagation of BECV is difficult in conventional cell lines (Mebus *et al.*, 1973*a*), it has been grown successfully in HRT18 cells (human rectal tumour cell line; Laporte *et al.*, 1980).

As a member of the Coronaviridae family BECV is a pleiomorphic, enveloped spherical particle (120 nm in diameter) surrounded by a fringe of 20 nm long club-shaped spikes. The coronavirus genome is a positive and single-stranded capped RNA with a polyadenylated 3' end (Siddell *et al.*, 1982; Sturman & Holmes, 1983). The structural and non-structural proteins of the virus are translated from a 3'-coterminal nested set of mRNAs, each having a common 5' leader sequence (Lai *et al.*, 1984). Only the unique 5'-terminal sequence is translated; this sequence is absent from the next smaller RNA of the set.

BECV possesses five main structural proteins, i.e. a phosphorylated nucleocapsid protein (N; 50K), a trans-membrane matrix glycoprotein (M; 28K) and three peplomer glycoproteins [S1, S2 (S; spike) and haemagglutinin (HA)]. Glycoproteins S1 (105K) and S2 (95K) are the cleavage products of the S gene-encoded primary product (180K; J. F. Vautherot, unpublished results; Deregt & Babiuk, 1987). HA (M_r 125K) is split by reducing agents into two subunits of equal size with an M_r of 65K (Laporte & Bobulesco, 1981; King & Brian, 1982); the neutralizing epitopes are located on S1 and HA (Vautherot *et al.*, 1984).

The cloning and sequencing of the gene encoding the S protein of BECV, a necessary first step for the production of a genetically engineered vaccine against BECV, is reported in this paper.

BECV strain F15 (BECV-F15) was grown in HRT18 cells; virus and genomic RNA purifications were performed as described previously (Crucière & Laporte, 1988). The virus genome was used as a template for cDNA synthesis. Poly(dC)-tailed RNA-cDNA heteroduplexes were inserted into a dG-tailed *Pst*I-linearized pBR322 plasmid. Complementary DNAs were then cloned in *Escherichia coli* RR1; tetracycline-resistant, ampicillin-susceptible colonies were then transferred onto nitrocellulose filters and lysed *in situ* (Maniatis *et al.*,

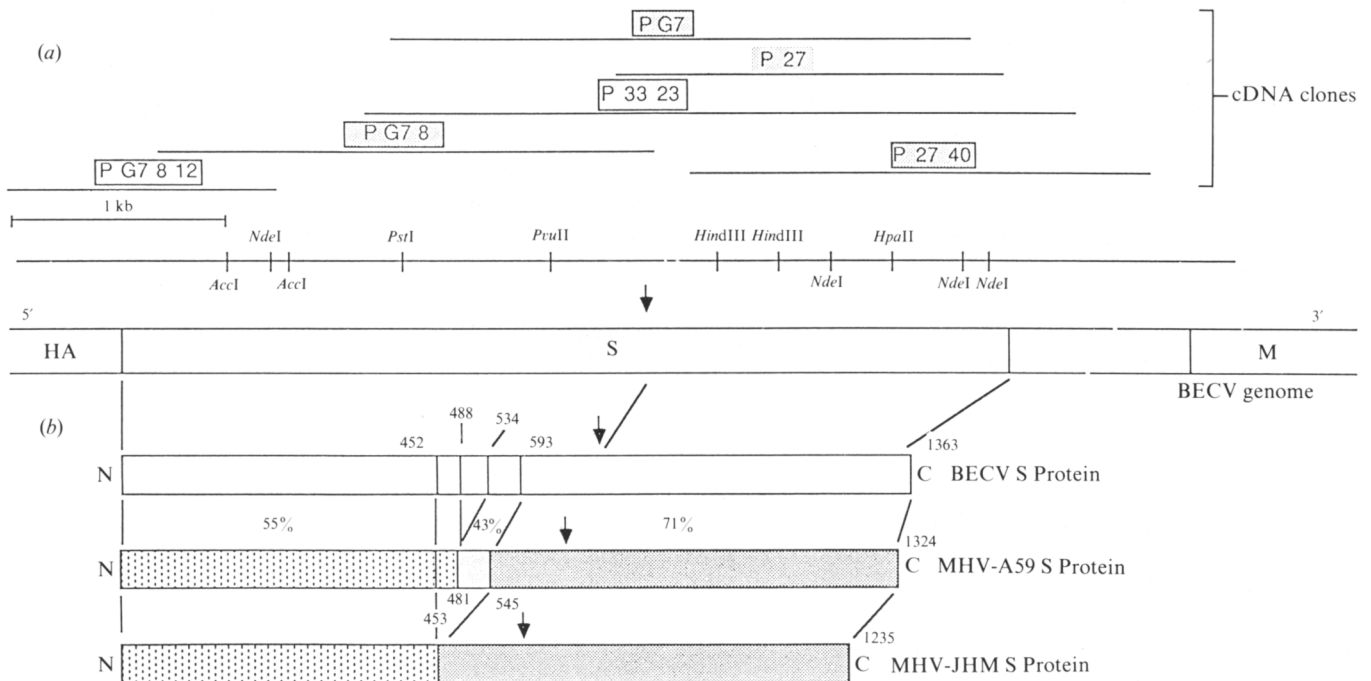


Fig. 1. (a) Schematic diagram of part of the BECV genome and location of cDNA clones. A simplified restriction map is given; the inserts used as probes to screen the cDNA library are shaded grey; the sequence of the S gene was obtained with the clones that are boxed in the figure. (b) Alignment and comparison of amino acid sequences of MHV-A59, MHV-JHM and BECV S propolypeptides. The percentage homology is given; arrows indicate putative peptidase cleavage sites.

1982). Their DNAs were hybridized overnight at 42 °C, in hybridization buffer (5 × SSC, 50% formamide, 5 × Denhardt's solution and 100 µg/ml of sonicated calf thymus DNA) containing a ³²P-labelled random primed viral insert cDNA (Feinberg & Vogelstein, 1983).

After washing the filters three times in 0.1 × SSC, 0.1% SDS at 60 °C for 20 min, they were dried and autoradiographed on X-ray films (Fuji) at -70 °C for 12 h. Mini-lysates were obtained from the positive clones (Birnboim & Doly, 1979). After digestion with *Pst*I, the selected inserts were analysed on Southern blots (Maniatis *et al.*, 1982). Northern blots were used for the localization of inserts. Total RNA from BECV-infected HRT18 cells was extracted by the guanidinium isothiocyanate method according to Vaquero *et al.* (1982) and electrophoresed in a denaturing 1% agarose gel (6% formaldehyde). Transfer, hybridization and washing were performed as described for Southern blot experiments.

The primer used to obtain viral cDNA corresponded to the *Bam*HI cleavage site (GGATCC) which was found at the beginning of our study on a cDNA clone at the 5' end of the gene encoding the N protein. (In fact this restriction site was not located at that place but was functioning randomly.) The primer yielded 3500 cDNA clones.

By Northern blot analysis (data not shown), we established that the large cDNA insert G7 (2.4 kb) covered a part of the S gene (Fig. 1); sequence analysis of that clone and comparison with the sequence of the gene coding for the E2 protein of coronavirus mouse hepatitis virus strain JHM (MHV-JHM) (Schmidt *et al.*, 1987) showed that this insert mapped in the middle of the S gene. G7 cDNA was used to screen the cDNA library and to obtain clones. Clones P G7 8 and P 27 40 were used also as probes to identify cDNA clones located at the 5' and the 3' ends of the S gene, respectively (Fig. 1).

Both DNA strands were sequenced on five overlapping cDNA clones: P G7, P G7 8, P G7 8 12, P 27 40 and P 33 23 (Fig. 1a). After it had been established by restriction mapping and Southern blot analysis that these clones covered the whole length of the S gene, M13 dideoxynucleotide sequencing was carried out according to Sanger *et al.* (1977) by using sonicated cloned fragments subcloned into the *Sma*I site of the M13 mp10 vector (Deininger, 1983). Buffer gradient gels and [α -³⁵S]dATP were used according to Biggin *et al.* (1983).

Sequence data were analysed and assembled with the aid of the program of Queen & Korn (1984), the Beckman Microgenie program (March 1985 version, Beckman Instruments) adapted for the IBM PC-XT microcomputer. The nucleotide sequence obtained

(Fig. 2) contains a single long open reading frame (ORF) in the mRNA sense which extends from the first ATG codon (nucleotides 31 to 33) to nucleotide 4122. This 4092 nucleotide sequence contains 28% A, 16.2% C, 19.7% G and 36.2% T, and has a coding capacity for 1363 amino acids (Fig. 2) with an M_r of 150747 and a proposed pI of 5.6.

The length of the ORF is in the expected size range for the glycoprotein S gene sequence. Furthermore, comparison with the published S gene sequence of coronavirus MHV-A59 (Luytjes *et al.*, 1987) shows homology which increases from 61% at the 5' end of the ORF (BECV nucleotides 31 to 1513) to 73.8% at the 3' end (BECV nucleotides 2711 to 4100).

Immediately upstream from the first initiation codon there is a sequence ATCTAAACAT very similar to the conserved intergenic sequences of BECV, MHV-JHM and MHV-A59 (Lapps *et al.*, 1987; Crucière & Laporte, 1988; Luytjes *et al.*, 1987; Schmidt *et al.*, 1987); it is also closely related to the conserved sequence AACTAAAC, reported for the transmissible gastroenteritis virus (TGEV) (Rasschaert & Laude, 1987). The sequence surrounding the translation initiation codon is in a sub-optimal environment (Kozak, 1987). A similar situation has been observed for the initiation codon of the S gene of MHV (Luytjes *et al.*, 1987).

Comparison of the first 400 nucleotides of the PG7 8 12 clone (P. Boireau & J. Laporte, unpublished observations) with the recently published HA-encoding gene sequence of bovine coronavirus (Parker *et al.*, 1989) led us to the conclusion that the S gene is just downstream of the HA gene.

The deduced amino acid sequence of the BECV S protein contains 19 potential N-glycosylation sites. Assuming a mean M_r value of 2100 per carbohydrate chain (Hunter *et al.*, 1983), the M_r of the mature S glycoprotein would be approximately 190600. It appears to be hydrophobic over most of its length (35% hydrophobic amino acids).

This protein shares some properties with S proteins described for other coronaviruses. After the first methionine residue, there is a potential signal sequence with a hydrophobic core of 13 amino acids and a helix-breaking residue, glycine 17 (Watson, 1984). According to the rule established for a potential cleavage site (von Heijne, 1984), a protease could act between amino acid residues 17 and 18. The consequence would be that the first amino acid in the S protein is aspartic acid 18.

Another potential peptidase cleavage site is located in the hydrophilic peak between residues 763 and 769. This sequence, Lys-Arg-Arg-Ser-Val-Arg, is collinear with the experimentally determined cleavage site of the MHV-A59 S protein (Luytjes *et al.*, 1987), and very similar to the cleavage site of the infectious bronchitis virus (IBV)

S protein (Binns *et al.*, 1985). Furthermore, this series of basic amino acids resembles the tryptic cleavage sites of peptidic prohormones or the F0 protein of Newcastle disease virus (MacGinnes & Morrisson, 1986), which are processed in the trans-Golgi apparatus. The coronavirus S protein uses the same cellular metabolic pathway for maturation, and budding of the virus takes place in the Golgi apparatus and in the endoplasmic reticulum membrane.

Tryptic cleavage would explain the maturation of the S protein of BECV: the primary gene product, P150 (i.e. the S polypeptide with an M_r of 148967 without the signal peptide) is glycosylated in the endoplasmic reticulum and in the Golgi apparatus giving rise to a glycoprotein of 188867 (gp190), which is then cleaved into the S1 (104692) and S2 (84175) structural proteins. These results are in agreement with those published by Deregt & Babiuk (1987), using immunoelectrophoresis to study the biosynthesis of gp105 and gp95, and with our own observations (J. F. Vautherot *et al.*, unpublished results).

Using the approach described by de Groot *et al.* (1987) for the S proteins of feline infectious peritonitis virus, MHV and IBV and by Rasschaert & Laude (1987) for the S protein of TGEV, we also demonstrated two amphipathic α -helices for the S protein of BECV; they are located between amino acids 1007 to 1077 and between amino acids 1269 to 1294. These α -helices may constitute the stalk of viral spikes, with a length of approximately 12 nm.

Using the method described by Rao & Argos (1986) a hydrophobic transmembrane α -helix was also predicted between amino acids 1305 and 1335. Its C-terminal location, the presence of a potential myristylation site on glycine 1333 and comparison with other coronaviruses suggest that this helix is involved in the anchorage of the spike in the viral membrane. The myristylation site, surrounded by eight cysteine residues, would be located at the internal face of the viral membrane. Among coronaviruses this domain is highly conserved in structure, location and length. A stretch of eight amino acids, Lys-Trp-Pro-Trp-Tyr-Val-Trp-Leu (1305 to 1312; Fig. 2), is found in all coronavirus S protein sequences established so far; however its role is unknown.

Luytjes *et al.* (1987, 1988) have compared S amino acid sequences of the two MHV strains A59 and JHM. These sequences are very similar but the S protein of JHM was found to be shorter. As BECV belongs to the same antigenic group as these viruses, the present results enlarge upon this comparison. Dot matrix analyses (Beckman Microgenie program) of the deduced amino acid sequence of the S genes of BECV and MHV-A59 (data not shown), revealed that there is low homology (55%) between amino acids 488 to 593 of BECV and

GCTGCATGATGCTTAGACCACTAACTAAACAGTTTTGATACITTTAATTTCTTACCAATGGCTCTGCTGTATAGGAGATTAAAGTGACTACGGTTCCCAATATGATGTTGAC 120
M F L I L L I S L P M A L A V I G D L K C T T V S I N D V D 30

ACCGGTGTCCTTCTTATAGCACTGATACCTGTCGATGTTACTAATGGTTTAGGTACTTATTATGTTTTAGATCGGTGTATTTAATACIACGGCTGTGGCTTAATGGTTACTACCCCTACT 240
 T G V P S I S T D T V D V T N G L G T Y Y V L D R V Y L N T T L L L N G Y Y P T 70

TCAGGTTCTACATATCGTAATATGGCACTGAAGGGAACITTTACTATTGAGCACACTATGGTTTAAACCACCTTTTCTTCTGATTTTATTAATGGTATTTTGGTAAGGTCAAAAATACC 360
 S G S T Y R N M A L K G T L L L S T L W F K P P F L S D F I N G I F A K V K N T 110

AAGGTTAATAACATGGTGAATGTATAGTGAGTTTCTGCTATAACTATAGGTAGTACTTTTGTAAATACATCCTATAGTGTGGTAGTACACCACATACTACCAATTTAGATAATAAA 480
 K V I K H G V M Y S E F P A I T I G S T F V N T S Y S V V V Q P H T T N L D N K 150

TTACAAGTCTCTTAGAGATCTCTGTTGCCAGTATACTATGTGCGAGTACCCAAATACGATTGTGATCCTAATTTGGGTAATCGGGCGGTAGAACTATGGCATGGGATACAGGTTGT 600
 L Q G L L E I S V C Q Y T M C E Y P N T I C H P N L G N R R V E L W H W D T G V 190

GTTTCTGTTTATTAAGGCTAACTACATATGATGTGAATGCTGATTTATTGATTTCCATTTTATCAAGAAGTGGTACTTTTATGCATATTTACAGATACTGGTGTGTTACT 720
 V S C L Y K R N F T Y D V N A D Y L Y F H F Y Q E G G T F Y A Y F T D T G V V T 230

AAGTTTCTGTTAATGTTTATAGGCAAGGCTTTTCAATTATATGTCATGCTTTGACTTGTAAATAGTGTATGACTTTAGAATATGGGTTACACCTCTCACTTCTAAACAATAT 840
 K F L F N V Y L G T V L S H Y Y V M P L T C N S A M T L E Y W V T P L T S K Q Y 270

TTACTCGCTTCAATCAAGATGGTGTATTTTAACTGCTGTTGATGTAAGAGTGATTTTATGACTGAGATTAAGTGTAAAACACTATCTATAGCACCATCTACTGGTGTATGAATTA 960
 L L A F N Q D G V I F N A V D C K S D F M S E I K C K T L S I A P S T G V Y E L 310

AACGGTTACACTGTTACAGCAATTCGAGATGTTTACCGAGCTATACCTAATCTCCGATTGTAATATAGAGGCTTGGCTTAATGATAAGTCTGTGCCCTCCCAATTTGGGAAGCT 1080
 N G Y T V Q P I A D V Y R R I P N L P D C N I E A W L N D K S V P S P L N W E R 350

AAGACCTTTCAAAATGTAATTTTAAATGAGCAGCGTATGCTTTTATCCAGCAGACTCATTACTTGTAAATAATATGATGCTGCTAAGATATAGGTATGTTTCCAGCATA 1200
 K T F S N C N F N M S S L M S F I Q A D S F T C N N I D A A K I Y G M C F S S I 390

ACTATAGATAAGTTTGCATACCCAAATGGTAGGAAAGTTGACCTACAAATGGGCAATTTGGGCTATTGGCAGCTTTTAACTATAGAATGATACTACTGCTACAAGTTGTCAGTTGTAT 1320
 T I D K F A I P N G R K V D L Q L G N L G Y L Q S F N Y R I D T T A T S C Q L Y 430

TATAAATTACCTGCTGCTAATGTTTCTCCAGCAGGTTTAACTCCTTCTAATTTGGAATAGGAGATTGGGTTTACAGAACAACTGTTTTAAGCCTCAACCTGTAGGTTTACTGAT 1440
 Y N L P A A N V S V S R F N P S I W N R R F G F T E Q S V F K P Q P V G V F T D 470

CATGATGTTGTTTATGACACAACATGTTTAAAGCTCCCAAAATTTCTGCTGGTGAATTTGATGGGCTTTTGGTGTAGGTAATGGTCTGCTGATAGATGCTGGTTATAAAAATAGT 1560
 H D V V Y A Q H C F K A P T N F C P C K L D G S L C V G N G P G I D A G Y K N S 510

GGTATAGGCACTTGTCTGACAGCACTAATTTTAACTTGCATAATGCTGCCAATGTAATTTGTTGCTGACTCCAGACCCCAATACATCTAACTACAGGCTTACAAAGTCCCC 1680
 G I G T C P A G T N Y L T C H N A A Q C N C L C T P D P I T S K S T G P Y K C P 550

CAAACAAATACCTAGTTGGCAGAGTGGACACTGTTGGGCTTGTGATTTAAAGTGATTATTGTGAGGTAATCCTGTACTTGGCAACCAAGCAATTTTGGGTTGGTCTGTTGAC 1800
 Q T K Y L V G I G E H C S G L A I K S D Y C G G N P C T C Q P Q A F L G W S V D 590

TCTGTTTACAAGGGGATAGGTGAATATTTTCTAATTTTATGATGATGTTAATGAGTACTACTTGTCTACTGATTACAAAAAACAACACAGACATAATCTTGGTGT 1920
 S C L Q G D R C N I F A N F I L H D V N S G T T C S T D L Q K S N T D I I L G V 630

TGTGTTAATATGATCTTTATGTTACTGGCCAGGATTTTGTGAGGCTAATGCGACTTATTATAATAGTGGCAGAACCTTTTATATGATTCTAATGGTAACTCTATGGTTTT 2040
 C V N Y D L Y G I T G Q G I F V E A N A T Y Y N S W Q N L L Y D S N G N L Y G F 670

AGAGACTACTTAAACAACAGAACTTTTATGATTGCTAGTTGCTATAGCGGCTGTTTACAGGCGCTTTCATGCTAATCTTCCGACCCAGCACTGCTATTTCCGGATATTTAAATGCAAT 2160
 R D Y L T N R T F M I R S C Y S G R V S A A F H A N S S E P A L L F R N I K C N 710

TACGTTTTAATAACTCTTTCACGACAGCTGCAACCTATTAACATTTTATAGTACTTCTGGTGTGTTGTCATGCTGATAATAGTACTGCTAGTGGTCTCAACATGATGATCTC 2280
 Y V F N N T L S R Q L Q P I N Y F D S Y L G C V V N A D N S T A S A V Q T C D L 750

ACAGTAGGTAGTGGTTACTGTGGATTACTCTACAAAAGACGAAGCGTAAAGGCGATTACCACCTGGTATCGGTTTACTAATTTTGGAGCATTACTGTTAATTCAGTAAATGATAGT 2400
 T V G S G Y C V D Y S T K R R S V R A I T T G Y R F T N F E P F T V N S V N D S 790

TTAGAACCTGTAGGTGGTTTGTATGAAATCAAATACCTTCAAGAAATTTACTATAGGTAATATGGAGGACTTTATCAAACAAGCTCTCCTAAAGTTACTATTGATGTTCTGCTTTTGTG 2520
 L E P V G G L Y E I Q I P S E F T I G N M E E F I Q T S S P K V T I D C S A F V 830

TGTGGTATGTTGCGAGCATGTAATACACAGTTGGTGAATATGGTATGTTCTGTGACAAJATTAATGCTATACTCACAGAAGTAAATGAACACTTGTACACTACACAGTTGCAAGTAGCT 2640
 C G D C A A C K S Q L V E Y G S F C D N I N A I L T E V N E L L D T T Q L Q V A 870

AATAGTTAATGAATGGTGTACTCTTAGCCTAAGCTTAAAGATGGCGTAAATTTCAATGTAGACGACATCAATTTTCCCTGTATAGGTTGTTAGGAAGCGAATGTAATAAGTT 2760
 N S L M N G V T L S T K L K D G V N F N V D D I N F S P V L G C L G S E C N K V 910

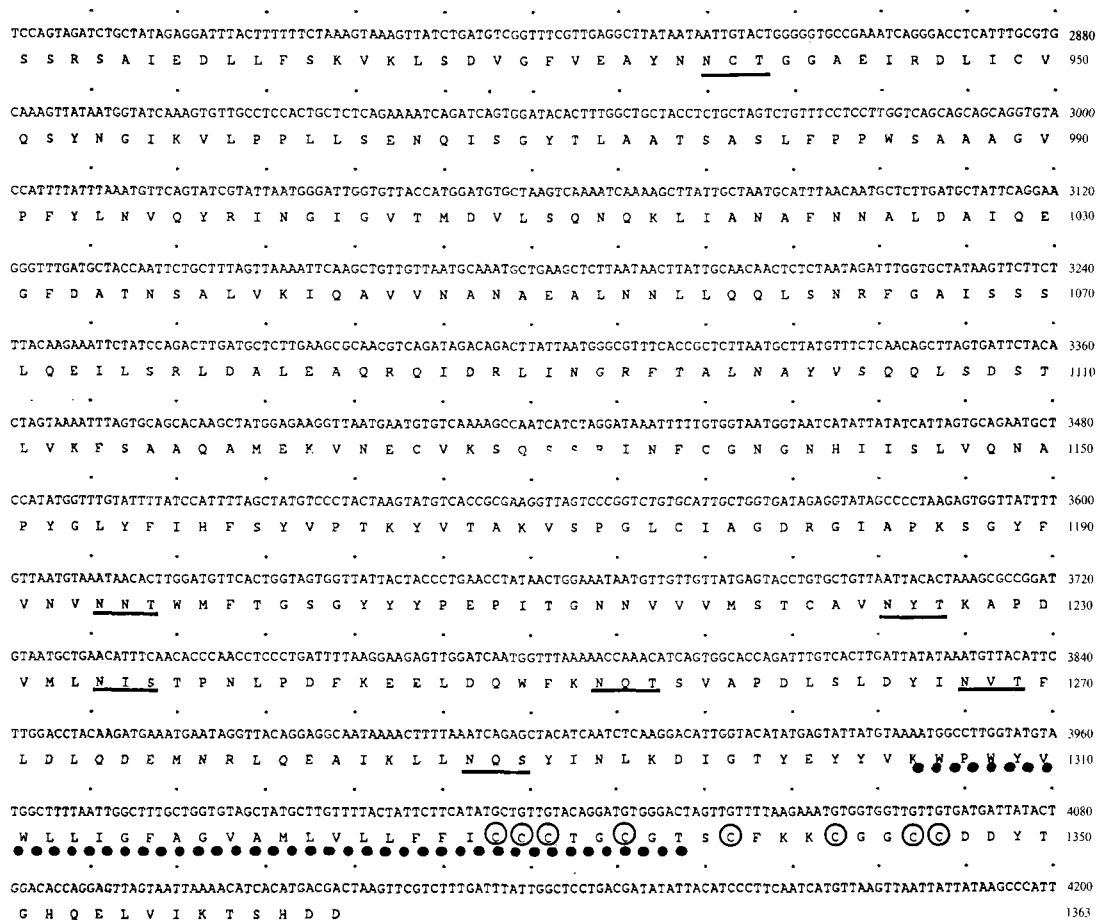


Fig. 2. Nucleotide sequence of the gene encoding the S protein and predicted amino acid sequence of the S protein. The sequence deduced from the cDNA inserts (Fig. 1a) is shown as positive sense DNA from 5' to 3' ends; the hatched bar indicates the proposed N-terminal signal sequence. An arrowhead indicates the potential tryptic cleavage site. A box surrounds the conserved intergenic sequence. Potential N-glycosylation sites are underlined (specific for the Asn-X-Ser/Thr sequence, with X different from Pro). Dots mark the proposed C-terminal membrane anchoring domain. The last eight cysteines are encircled.

amino acids 481 to 545 of MHV-A59 and apparently there is a deletion in the MHV-A59 S protein sequence. More details emerged after alignment comparison of the two sequences (Fig. 1b). Amino acids 488 to 534 of BECV have no counterpart in MHV-A59. Furthermore, amino acids 452 to 593 of the S protein of BECV have no counterpart in MHV-JHM (Fig. 1b).

Luytjes *et al.* (1987, 1988) put forward two hypotheses to account for the difference between MHV-A59 and MHV-JHM in this domain of the S protein. The first hypothesis is that the MHV-JHM genome is deleted with respect to a nucleotide sequence corresponding to amino acids 453 to 545 of the S protein of MHV-A59; the second is that the MHV-A59 genome has acquired genomic material by non-homologous recombination. The comparisons presented above support the idea of a genetic instability in this area of the virus genome which would explain the difference in length of the S proteins of

the three virus strains. Therefore we suggest an evolutionary progression from MHV-JHM to MHV-A59 to BECV or in reverse depending upon whether non-homologous recombination events or deletions have occurred. The observation that the nucleotide sequence of the S gene encoding amino acids 470 to 480 (data not shown) is highly homologous (81%) between BECV and MHV-A59 and does not exist in MHV-JHM leads us to propose the occurrence of deletions in this genome segment. Our suggestion is supported by the identification of a functional HA gene in BECV (Parker *et al.*, 1989) whereas only a related HA pseudogene was identified in MHV-A59 (Luytjes *et al.*, 1988); this HA is most likely to occur if the BECV genome is more closely related to a possible common ancestor.

We would like to thank Jean François Vautherot, Denis Rasschaert and Michel Brémont for helpful discussions.

References

- BIGGIN, M. D., GIBSON, T. J. & HONG, G. F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proceedings of the National Academy of Sciences, U.S.A.* **80**, 3963–3965.
- BINNS, M. M., BOURSNELL, M. E. G., CAVANAGH, D., PAPPIN, D. J. C. & BROWN, T. D. K. (1985). Cloning and sequencing of the gene encoding the spike protein of the coronavirus IBV. *Journal of General Virology* **66**, 719–726.
- BIRNBOIM, H. C. & DOLY, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* **7**, 1513–1523.
- BRIDGER, J. C., WOODE, G. N. & MEYLING, A. (1978). Isolation of coronaviruses from neonatal calf diarrhoea in Great Britain and Denmark. *Veterinary Microbiology* **3**, 101–103.
- CRUCIÈRE, C. & LAPORTE, J. (1988). Sequence and analysis of bovine enteric coronavirus (F15) genome I. Sequence of the gene coding for the nucleocapsid protein; analysis of the predicted protein. *Annales de l'Institut Pasteur* **139**, 123–138.
- DE GROOT, R. J., LENSTRA, J. A., LUYTJES, W., NIESTERS, H. G. M., HORZINEK, M. C., VAN DER ZEIJST, B. A. M. & SPAAN, W. J. M. (1987). Sequence and structure of the coronavirus peplomer protein. In *Biochemistry and Biology of Coronaviruses*, pp. 31–38. Edited by M. M. C. Lai & S. A. Stohman. New York: Plenum Press.
- DEININGER, P. L. (1983). Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Analytical Biochemistry* **129**, 216–223.
- DEREGT, D. & BABIUK, L. A. (1987). Monoclonal antibodies to bovine coronavirus: characteristics and topographical mapping of neutralizing epitopes on the E2 and E3 glycoproteins. *Virology* **161**, 410–420.
- FEINBERG, A. P. & VOGELSTEIN, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* **132**, 6–13.
- GOUET, PH., CONTREPOIS, M., DUBOURGUIER, H., RIOU, Y., SCHERRER, R., LAPORTE, J., VAUTHEROT, J. F., COHEN, J. & L'HARIDON, R. (1978). The experimental production of diarrhoea in colostrum-deprived axenic and gnotoxenic calves with enteropathogenic *E. coli*, rotavirus, coronavirus and in combined infection of rotavirus and *E. coli*. *Annales de Recherches Vétérinaires* **9**, 433–440.
- HUNTER, E., HILL, E., HARDWICK, M., BROWN, A., SCHWARTZ, D. E. & TIZARD, R. (1983). Complete sequence of the Rous sarcoma virus *env* gene: identification of structural and functional regions of its product. *Journal of Virology* **46**, 920–936.
- KING, B. & BRIAN, D. A. (1982). Bovine coronavirus structural proteins. *Journal of Virology* **42**, 700–707.
- KOZAK, M. (1987). At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *Journal of Molecular Biology* **196**, 947–950.
- 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. *Proceedings of the National Academy of Sciences, U.S.A.* **81**, 3626–3630.
- LAPORTE, J. & BOBULESCO, P. (1981). Polypeptide structure of bovine enteric coronavirus: comparison between a wild strain purified from feces and a HRT18 cell adapted strain. In *Biochemistry and Biology of Coronaviruses*, pp. 181–184. Edited by V. ter Meulen, S. Siddell & H. Wege. New York: Plenum Press.
- LAPORTE, J., BOBULESCO, P. & ROSSI, F. (1980). Une lignée cellulaire particulièrement sensible à la répllication du coronavirus entérique bovin: les cellules HRT18. *Comptes rendus de l'Académie des sciences* **290**, 623–626.
- LAPPS, W., HOGUE, B. G. & BRIAN, D. A. (1987). Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. *Virology* **157**, 47–57.
- 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 trypsin cleavage site. *Virology* **161**, 479–487.
- LUYTJES, W., BREDENBEEK, P. J., NOTEN, A. F. H., HORZINEK, M. C. & SPAAN, W. J. M. (1988). Sequence of mouse hepatitis virus A59 mRNA2: indications for RNA recombination between coronaviruses and influenza C virus. *Virology* **166**, 415–422.
- MACGINNES, L. W. & MORRISON, T. G. (1986). Nucleotide sequence of the gene encoding the Newcastle disease virus fusion protein and comparisons of paramyxovirus fusion protein sequences. *Virus Research* **5**, 343–356.
- MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory.
- MEBUS, C. A., STAIR, E. L., RHODES, M. B. & TWIEHAUS, M. J. (1973a). Neonatal calf diarrhoea: propagation, attenuation, and characteristics of a coronavirus-like agent. *American Journal of Veterinary Research* **34**, 145–150.
- MEBUS, C. A., STAIR, E. L., RHODES, M. B. & TWIEHAUS, M. J. (1973b). Pathology of neonatal calf diarrhoea induced by a coronavirus-like agent. *Veterinary Pathology* **10**, 45–64.
- PARKER, M. D., COX, G. J., DEREGT, D., FITZPATRICK, D. R. & BABIUK, L. A. (1989). Cloning and *in vitro* expression of the E3 haemagglutinin glycoprotein of bovine coronavirus. *Journal of General Virology* **70**, 155–164.
- QUEEN, C. & KORN, L. J. (1984). A comprehensive sequence analysis program for the IBM personal computer. *Nucleic Acids Research* **12**, 581–599.
- RAO, M. J. K. & ARGOS, P. (1986). A conformational preference parameter to predict helices in integral membrane proteins. *Biochimica et biophysica acta* **869**, 197–214.
- RASSCHAERT, D. & LAUDE, H. (1987). The predicted primary structure of the peplomer protein E2 of the porcine coronavirus transmissible gastroenteritis virus. *Journal of General Virology* **68**, 1883–1890.
- SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.
- SCHMIDT, I., SKINNER, M. & SIDDELL, S. (1987). Nucleotide sequence of the gene encoding the surface projection glycoprotein of coronavirus MHV-JHM. *Journal of General Virology* **68**, 47–56.
- SIDDELL, S., WEGE, H. & TER MEULEN, V. (1982). The structure and replication of coronaviruses. *Current Topics in Microbiology and Immunology* **99**, 131–163.
- STURMAN, L. S. & HOLMES, K. V. (1983). The molecular biology of coronaviruses. *Advances in Virus Research* **28**, 35–112.
- VAQUERO, C., SANCEAU, J., CATINOT, L., ANDREU, G., FALCOFF, E. & FALCOFF, R. (1982). Translation of mRNA from phytohemagglutinin-stimulated human lymphocytes: characterization of interferon mRNAs. *Journal of Interferon Research* **2**, 217–228.
- VAUTHEROT, J. F., LAPORTE, J., MADELAINE, M. F., BOBULESCO, P. & ROSETO, A. (1984). Antigenic and polypeptide structure of bovine enteric coronavirus as defined by monoclonal antibodies. In *Molecular Biology and Pathogenesis of Coronaviruses*, pp. 117–132. Edited by P. J. M. Rottier, B. A. M. van der Zeijst, W. J. M. Spaan & M. C. Horzinek. New York: Plenum Press.
- VON HEIJNE, G. (1984). How signal sequences maintain cleavage specificity. *Journal of Molecular Biology* **173**, 243–251.
- WATSON, M. E. E. (1984). Compilation of published signal sequences. *Nucleic Acids Research* **12**, 5145–5164.