SEQUENCE ANALYSIS OF THE 3' END (8740 NUCLEOTIDES) OF BECV GENOME ; COMPARISON WITH HOMOLOGOUS MHV NUCLEOTIDE SEQUENCE

Pascal Boireau[2], Nathalie Woloszyn[1], Catherine Crucière[2], Ericka Savoysky[3] and Jacques Laporte[1]

[1]Station de Virologie et d'Immunologie Moléculaires I.N.R.A.,C.R.J.J. Domaine de Vilvert, 78350 Jouy en Josas, France
[2]Laboratoire de Virologie Moléculaire, L.C.R.V. C.N.E.V.A., 22 rue Pierre Curie, BP 67, 94703 Maisons-Alfort Cedex, France
[3]Laboratoire de Microbiologie Appliquée et Industrielle, Faculté des Sciences Pharmaceutiques et Biologiques, 5 rue A. Lebrun 5400 Nancy, France

INTRODUCTION

Bovine enteritic coronavirus (BECV) is one of the major pathogen of neonatal calves. The viral genome is a unique 20 Kb large RNA molecule which is polyadenylated at its 3' end. Studies of viral mRNAs suggest that eight genes are translated (Crucière & Laporte, personnal communication) and as established for MHV[1] the possibility of functional bicistronic genes cannot be excluded. BECV is made of 4 major structural proteins[2,3,4], a 50 kD phosphorylated nucleocapsid N, a 180 kD peplomer glycoprotein S (present on the virions as 105 and 95 kD glycosylated subunits), a 125 kD haemagglutinin HE (made up of two disulfide bridge-linked glycosylated subunits of 65 kD) and a 28 kD transmembrane glycoprotein M. The complete amino-acid sequences of these proteins have recently been deduced from nucleotide sequences of the viral genome[5,6,7,8,9].

In this paper we present the complete nucleotide sequence of the first eight Kb of BECV strain F15 genome and discuss the number and the location of the genes and the possible role of the encoded proteins ; we also compare our results with the already published data concerning MHV-A59 or JHM.

RESULTS

1- BECV specific intracellular RNAs

In a one-step growth experiment, the analysis of the poly(A+) RNAs extracted from BECV infected HRT18 cells 7 hours after virus penetration showed, after electrophoresis in denaturing agarose gel, 8 different molecular mass species (Fig. 1, Table 1). Northern blot analysis of these RNAs by cDNA probes corresponding to the viral N gene established their viral origin.

Coronaviruses and Their Diseases
Edited by D. Cavanagh and T.D.K. Brown
Plenum Press, New York, 1990

Table 1. Sizes of BECV mRNAs.

BECV mRNA	MOLECULAR WEIGHT OF BECV mRNA[1]	PREDICTED SIZE OF THE mRNA (IN BASES)		PREDICTED SIZE OF THE POLYPEPTIDE ENCODED BY mRNA[4]
		EXPERIMENTAL DATA[2]	NUCLEOTIDE SEQUENCE DATA[3]	
1	5,75			
2	2,70	8120		
3	2,45	7420	8700	47 730 (HE)
4	2,10	6360	7400	150 740 (S)
5	0,87	2630	3005	12 800 (NS2)
6	0,80	2400	2735	9 585 (NS3)
7	0,71	2160	2345	26 370 (M)
8	0,56	1690	1650	49 370 (N)

(1) in kD
(2) size of BECV mRNA obtained in denaturing agarose gel .
(3) size of BECV mRNA predicted from the location of the consensus nucleotide sequence CNAAAC (N=C or T)
(4) size of polypeptide deduced from the ORF immediatly downstream the conserved sequence (in D).

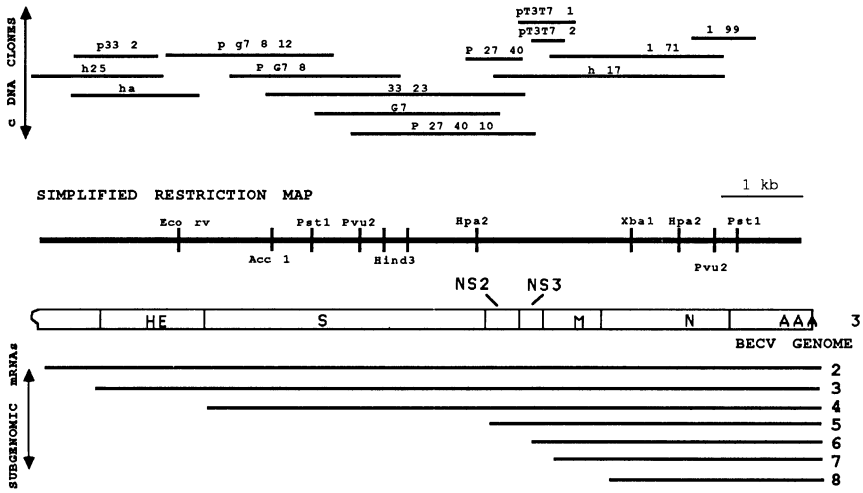

Figure 1. Organization of BECV Genome and Location of cDNA Clones Used in Sequence Determination.

Furthermore hybridization with a panel of viral probes showed that each mRNA had a common 3′ end with the next smaller mRNA.

2- cDNA cloning and sequencing

From 14 overlapping cDNA clones (Fig. 1) we have sequenced the first 8.8 Kb of the 3′ end of the viral genome. CDNA libraries were obtained by using synthetic oligonuleotide priming. CDNA inserts amplified by the polymerase chain reaction (PCR) were cloned into PT3T7 18U vector to obtain the whole length of some genes (NS2, NS3, HE). Location of cDNA clones was performed by restriction mapping, Southern and northern blot hybridizations. Nucleotide sequence of both cDNA strands was obtained from cDNA clones shown in Fig.1. From a database of nucleotide sequences containing about 60 000 nucleotides, a large unique 8740

nucleotide sequence was established. It contains six large non-overlapping open reading frames (ORF). Starting from the 3' end of the genome, the two first ORF encode, as previously reported[5,6], for the viral N and M structural proteins. Downstream two smaller ORFs NS3 and NS2 have coding capacities for 84 and 110 amino-acids respectivelly. After an intergenic sequence of 377 nucleotides, the next large ORF is 4092 nucleotides long. It belongs to the S gene encoding the major viral glycoprotein ; the predicted polypeptide (1364 amino-acids) has a relative molecular mass of about 150,5 kD[8]. The last ORF we have sequenced encodes for a polypeptide of 424 amino-acids with a relative molecular mass of 47729 D corresponding to the unglycosylated subunit of the gP 125 protein of BECV.

DISCUSSION

1- Genome organization of BECV

As determined by electrophoresis in denaturing gel the size of BECV poly(A^+) genomic RNA comprises between 18 Kb and 23 Kb. Seven subgenomic RNA species are evidenced in the infected cells and northern blot hybridization confirms that they form a 3'end coterminal nested set. These results are in good agreement with those of Keck et al.[9] concerning BECV RNA synthesis and lead us to conclude that at leat 8 non overlapping genes constitute the viral genome..

Sequencing of the first 8740 nucleotides of the 3' end of the genome demonstrate 6 different genes coding respectively for the N, M, NS3, NS2, S, and HE viral proteins. The rest of the genome should comprise the 2 missing genes (NS1 and Pol) if we compare with the MHV genome. Among the 6 sequenced genes N and HE present respectively 2 and 3 overlapping ORFs (Fig. 2). The already published sequences of IBV[10], MHV[1], or TGEV[11] demonstrate large overlapping ORFs only for genes encoding non structural proteins. Nevertheless the consensus nucleotide sequence surrounding the putative initiation codon of the secondary ORF of the BECV N gene is one of the optimal environments for initiation of mRNA translation[12]. This observation is not true for ORF2 of the HE gene which exhibits a U in +4 but the situation is the same for the initiation

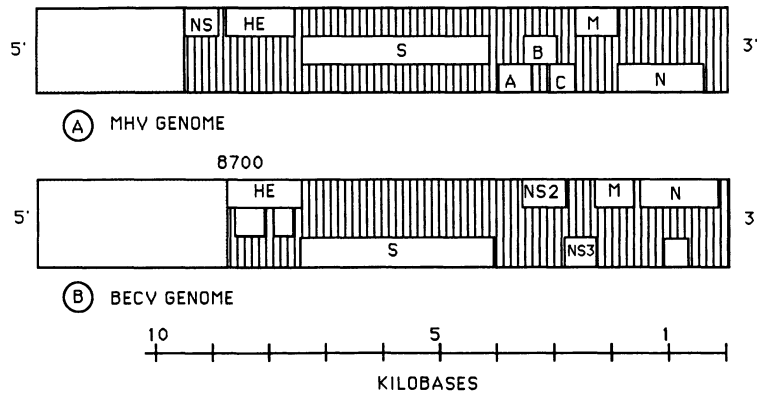

Figure 2. Location of ORFs (Larger than 200 Bases) in BECV and MHV Genomes.

codon of N, S and HE BECV genes. We cannot exclude a biological role for the secondary ORFs, but their expression in infected cells remains to be proved.

Immediatelly upstream from the initiation codon of the four structural genes we found a conserved nucleotide sequence : (A/T)C(C/T)AAAC. This sequence is very similar to the intergenic transcription initiation sites of some MHV mRNAs[13] which have the common TC(C/T)AAAC sequence. For the BECV NS2 and NS3 genes the sequence TCCAAAC is located far upstream of the putative initiation codon (80 and 130 nucleotides respectivelly). The location of the -CTAAAC- sequences that we observed for the different genes is in good agreement with the experimentally determined size of the BECV mRNAs (Table 1.) ; nevertheless because of the molecular mass markers we used, the experimental size given for the larger mRNAs appeared to be underestimated.

2- Properties of the BECV proteins predicted from the main ORFs

N and M proteins. The properties of these structural proteins and their comparison with the MHV homologous protein have been already described elsewhere[5,6].

S protein. The S gene[8] encodes the larger BECV glycoprotein (gp180) which shares some properties with other coronavirus S proteins i.e. : their amino-acid sequences revealed two main hydrophobic domains and a putative protease cleavage site. The 19 first N terminal amino-acids constitute the first hydrophobic domain consistent with a role as a signal sequence[14]. The second (amino-acids 1309 to 1335) is predicted as a transmembrane α helix and could be involved in the anchorage of the spikes in the viral membrane.

Because of the 19 putative glycosylation sites of this protein its relative molecular mass is estimated at about 185 kD. This glycoprotein has a putative protease cleavage site -Lys-Arg-Arg-Ser-Val-Arg- (amino-acids 763 to 769) colinear with the experimentally established cleavage site of MHV-A59 S protein[15]. The resulting frament S1 and S2 have a relative molecular mass (102,592 D and 85,175 D respectively) very similar to the molecular mass of the gp105 and gp 95 BECV structural proteins previously described[3,16].

HE protein. The amino-acid sequence deduced from the HE gene of BECV F15 isolated in France is very similar to the HE amino-acid sequence established for BECV Quebec strain[7] (5 amino-acids difference).

Luytjes et al.[17] have shown a fairly good similarity between the amino-acid sequence of the HA glycoprotein of Influenza C virus[18] and the MHV-A59 pseudogene product homologous to the BECV HE gene product. When comparing these 3 proteins by Dayhoff optimal alignment (Multalin program[19] Fig. 3) three highly conserved domains are evidenced : 1) from amino-acids 57 to 70, 2) from 124 to 134, 3) from 370 to 382. The esterase site of influenza C virus HA is identified as the -Gly-Phe-Gly-Asp-Ser-Arg-Thr- amino-acid sequence[20]. This sequence is the highly conserved sequence 1) except for a serine to threonine change in BECV and MHV-A59. Experiments have shown that purfied BECV particles possess such an esterase activity[21].

Non structural proteins NS3 and NS2. The protein predicted from the ORF of the gene NS3 has a relative molecular mass of 9585 D,

Figure 3. Comparison by Dayhoff Alignment of Amino Acid Sequences Deduced from BECV HE Gene (PHAF 15), Homologous MHV A59 Pseudogene (PHA59) and Influenza C HA Gene (PHANC). Highly Conserved Esterase Site is Boxed.
(*) Exact Homology.
(:) Conservative Change.

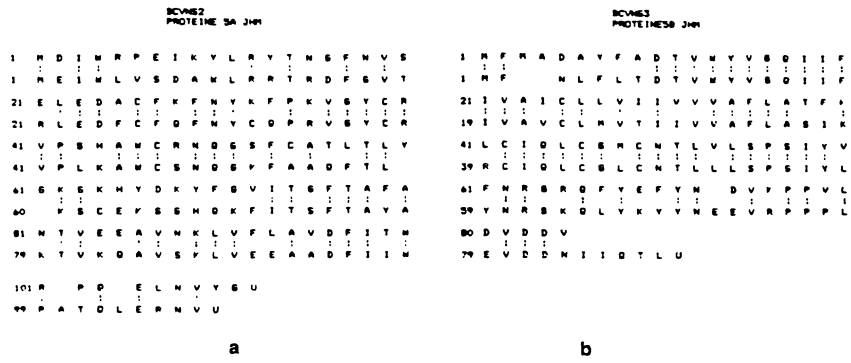

Figure 4. Alignments and Comparisons of Amino Acid Sequences of MHV JHM, 5A.5B Polypeptides and BECV NS2.NS3 Putative Polypeptides.

is highly hydrophobic and has 62% homology with the non structural MHV-JHM 10.2 kD protein[1]. There are 2 in-frame ATG at the beginning of this ORF (Fig. 4b). The second ATG seems, as a translation initiation codon, in a better environment[12] (G in +4 and T in -3). This observation is strengthened by comparison with the amino-acid sequence of the homologous MHV-JHM NS protein encoded by mRNA 4[13].

The 12.8 kD polypeptide deduced from NS2 gene (Fig 4a) has 53% homology with the 12.4 kD protein of MHV-JHM[1]. It has one potential glycosylation site and a C terminal hydrophobic α-helix structure looking like a transmembrane domain. The 12.4 kD non-structural polypeptides of IBV[10] and of MHV have a similar structure ; these 3 molecules should have a similar, but as yet unknown biological function.

This part of the BECV genome comprises only two genes (NS2 and NS3) but the homologous part of the MHV-JHM genome contains three genes A, B and C (Fig. 2). In fact between the NS and S BECV genes there is a small ORF (but no methionine as initiation codon) from which we can deduce a polypeptide having 52% homology with the C terminal end of the 15.2 kD polypeptide of MHV-JHM deduced from mRNA4[13]. A second small putative BECV polypeptide, deduced from another ORF located in the same part of the genome, has a similar homology with the N terminal end of the 15.2 kD MHV-JHM protein. Moreover, upstream of these two short hypothetical ORFs, in both cases we found a -GTAAAC- nucleotide sequence very similar to the consensus sequence -CTAAAC- putative transcription initiation site. These pseudogenes could be the memory of an ancestor gene unnecessary for virus replication at least in cell culture.

REFERENCES

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

2. J. Laporte and P. Bobulesco, 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," pp. 181-184. Edited by V. ter Meulen, S. Siddell and H. Wege, ed., Plenum Press, New York (1981).
3. J.F. Vautherot, J. Laporte, M.F. Madelaine, P. Bobulesco and A. Roseto, Antigenic and polypeptide structure of bovine enteritic 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 and M.C. Horzinek, ed., Plenum Press, New York (1984).
4. B. King and D.A. Brian, Bovine coronavirus structural proteins. J. Virol., 42:700 (1982).
5. W. Lapps, B.G. Hogue and D.A. Brian, Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes, Virology 157:47 (1987).
6. C. Crucière and J. Laporte, Sequence and analysis of bovine enteritic coronavirus (F15) genome I.- Sequence of the gene coding for the nucleocapsid protein; analysis of the predicted protein. An. Inst. Pasteur/Virology 139: 123 (1988).
7. M.D.Parker, G.J. Cox, D.R. Deregt, D.R. Fitzpatrick and L.A. Babiuk, Cloning and *in vitro* expression of the E3 haemagglutinin glycoprotein of Bovine Coronavirus. J. Gen. Virol. 70:155 (1989).
8. P. Boireau, C. Crucière and J. Laporte, Nucleotide sequence of the glycoprotein S (E2) gene of the bovine enteric coronavirus ; comparison with mouse hepatitis virus. J. gen. Virol. in the press (1990).
9. J.G. Keck, B.G. Hogue, D.A. Brian and M.M.C. Lai, Temporal regulation of bovine coronavirus RNA synthesis. Virus Res. 9:343 (1988).
10. M.E.G. Boursnell, M.M. Binns and T.D.K. Browm, Sequencing of coronavirus IBV genomic RNA : three open reading frames in the 5' "unique" region of mRNA D. J. gen. Virol. 66:2253 (1985).
11. D. Rasschaert, J. Gelfi and H. Laude, Enteric coronavirus TGEV : partial sequence of the genomic RNA, its organization and expression. Biochimie 69:591 (1987).
12. M. Kozak, At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells, J. Mol. Biol. 196:947 (1987).
13. M.A. Skinner and S.G. Siddell, Coding sequence of coronavirus MHV-JHM mRNA 4. J. gen. Virol. 66:593 (1985).
14. M.E.E. Watson, Compilation of published signal sequences. Nucl. Acids Res. 12:5145 (1984).
15. W. Luytjes, L.S. Sturman, P.J. Bredenbeek, J. Charite, B.A.M. van der Zeijst, M.C. Horzinek and W.J.M. Spaan (1987), Primary structure of the glycoprotein E2 of coronavirus MHV-A59 and identification of the trypsin cleavage site, Virology 161:179 (1987).
16. D.Deregt and L.A. Babiuk. Monoclonal antibodies to bovine coronavirus: characteritics and topographical mapping of neutralizing epitopes on the E2 and E3 glycoproteins. Virology 161:410 (1987).
17. W. Luytjes, P.J. Bredenbeek, A.F.H. Noten, M.C. Horzinek and W.J.M. Spaan (1988), Sequence of mouse hepatitis virus A59 mRNA2 : Indication for RNA recombination between

coronaviruses and influenza C virus, Virology 166:415 (1988).
18. S. Nakada, R.S. Creager, M. Krystal, R.P. Aaronson and P. Palese, Influenza C virus hemagglutinin : comparison with influenza A and B virus hemagglutinins. J. Virol. 50:118 (1984).
19. F. Corpet, Multiple sequence aligment with hierarchical clustering. Nucl. Acids Res. 16:10881 (1988).
20. R. Vlasak, T. Muster, A.M. Lauro, J.C. Powers and P. Palese, Influenza C virus esterase : analysis of catalytic site, inhibition, and possible function. J. Virol. 63:2056 (1989).
21. R. Vlasak, W. Luytjes, J. Leider, W. Spaan and P. Palese, The E3 protein of bovine coronavirus is a receptor-destroying enzyme with acetylesterase activity. J. Virol. 62:4686 (1988).