J. gen. Virol. (1989), **70**, 3363–3370. *Printed in Great Britain*

Key words: *Berne virus/toroviruses/nucleocapsid protein gene*

Identification and Primary Structure of the Gene Encoding the Berne Virus Nucleocapsid Protein

By ERIC J. SNIJDER*, JOHAN A. DEN BOON, WILLY J. M. SPAAN, GEORGES M. G. M. VERJANS AND MARIAN C. HORZINEK

Institute of Virology, Veterinary Faculty, State University of Utrecht, Yalelaan 1, 3584 CL Utrecht, The Netherlands

(*Accepted 18 August 1989*)

SUMMARY

The nucleotide sequence of the nucleocapsid (N) protein gene of Berne virus (BEV; proposed family *Toroviridae*) was determined from two independent clones of a cDNA library. From the deduced amino acid sequence a basic protein of 18·3K was predicted. *In vitro* transcription and translation, followed by immunoprecipitation, were used to identify the gene. The identification was confirmed by metabolic labelling, using the knowledge that cysteine residues are absent from the amino acid sequence of the N protein. Smaller N-related polypeptides encountered in BEV-infected cell lysates were shown to be probable products of aberrant translation, due to initiation on AUG codons further downstream in the N protein gene.

INTRODUCTION

The equine Berne virus (BEV) is the prototype of a proposed new family of positive-stranded RNA viruses, the *Toroviridae* (Horzinek & Weiss, 1984). The morphologically and antigenically related bovine Breda virus (Woode *et al.*, 1985) and human torovirus (Beards *et al.*, 1984) are additional representatives of this new virus group. The BEV virion consists of a peplomer-bearing envelope, containing a tubular nucleocapsid (Weiss *et al.*, 1983) that surrounds a single, polyadenylated RNA molecule of at least 20 kb (Snijder *et al.*, 1988). The nucleocapsid can be straight or bent into a characteristic open torus. It shows periodic transverse striations (Weiss *et al.*, 1983), indicative of helical symmetry. Four proteins have been identified in BEV particles; three of them, the *N*-glycosylated peplomer protein(s) (P; 80K to 100K) and two unglycosylated proteins of 22K (E) and 37K (M), are associated with the envelope; upon detergent treatment, they can be removed from the virion. The remaining substructure is the nucleocapsid particle (Horzinek *et al.*, 1985). The main constituent of the nucleocapsid is a protein of 19K. This nucleocapsid (N) protein is also the most abundant species in the complete virion, accounting for about 80% of its protein mass (Horzinek *et al.*, 1985). The N protein is phosphorylated and possesses RNA-binding properties (Horzinek *et al.*, 1985).

In BEV-infected embryonic mule skin (EMS) cells the presence of one viral genome-sized RNA and four subgenomic mRNAs has been demonstrated (Snijder *et al.*, 1988). *In vitro* translation of the smallest RNA (RNA 5; 0·8 kb) results in the synthesis of the N protein (Snijder *et al.*, 1988). Recently a cDNA library has been prepared, using mRNAs from infected cells as a template. Northern blot hybridizations with cDNA probes from this library have shown that the five BEV RNAs form a 3′-coterminal nested set (Snijder *et al.*, 1990). This predicts the open reading frame (ORF) translated from RNA 5 to be the first ORF upstream of the poly(A) tail. In the present paper the sequence analysis of the BEV N gene and the identification of its product are described.

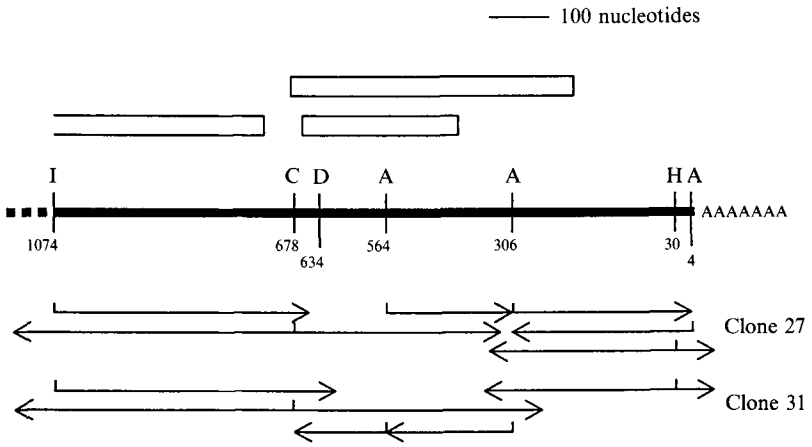

Fig. 1. Restriction map and sequencing strategy for the BEV N gene. A map of the 1074 nucleotides upstream of the poly(A) tail is shown. Numbers represent the nucleotide distance to the start of the poly(A) tail. Restriction sites are indicated by vertical bars and one-letter codes: A, *Alu*I; D, *Dde*I; C, *Eco*RI; H, *Hae*III; I, *Hin*dIII. Arrows indicate the parts of clones 27 and 31 used for sequence analysis and the direction of sequencing. Open boxes represent ORFs.

METHODS

Cells and virus. Berne virus (strain P138/72) was grown in EMS cells as described previously (Snijder *et al.*, 1988).

cDNA synthesis and cloning. Intracellular RNAs were isolated from infected EMS cells and mRNAs were purified with the aid of oligo(dT)–cellulose chromatography as described before (Snijder *et al.*, 1988). Using oligo(dT)-priming, cDNA was synthesized and cloned in plasmid pUC9 (Pharmacia). Details of cDNA synthesis and the screening of recombinant transformants will be presented elsewhere (Snijder *et al.*, 1990).

DNA sequencing. Restriction fragments were subcloned in M13 vectors (Messing, 1983) and sequenced using the dideoxynucleotide method (Sanger *et al.*, 1977). Sequence data were analysed using the computer programs of Staden (1986).

Construction of pBSN1 to pBSN4. Construct pBSN1 was generated by cutting clone 31 with *Eco*RI (position 678, see Fig. 1) and *Hin*dIII (pUC9) and cloning this fragment between the *Eco*RI and *Hin*dIII sites of transcription vector pBS (Stratagene). To reconstruct the ATG codon of the N gene, this plasmid was cut with *Eco*RI, sticky ends were filled in using the Klenow fragment of *Escherichia coli* DNA polymerase I (Maniatis *et al.*, 1982) and an *Sph*I linker (5′ CATGCATGCATG 3′; New England Biolabs) was inserted. After religation and transformation, the first ATG codon from the linker was removed by digestion with *Sph*I. The second ATG codon was eliminated by degradation of the *Sph*I sticky ends, using the 3′–5′ exonuclease activity of T4 DNA polymerase (Maniatis *et al.*, 1982). An *Xho*I linker (5′ CCTCGAGG 3′; New England Biolabs) was inserted and digestions with *Xho*I and *Hae*III (position 30, Fig. 1) were used to allow cloning of the gene between the *Xho*I and *Sma*I sites of vector pSVL (Pharmacia). Subsequently, the ORF was removed from pSVL by digestion with *Xho*I and *Bam*HI and recloned between the *Eco*RI and *Bam*HI sites of pBS; *Eco*RI and *Xho*I sticky ends were polished using the Klenow fragment of DNA polymerase I, as described above. Constructs pBSN2, pBSN3 and pBSN4 were generated by cutting pBSN1 with *Eco*RI, *Dde*I and *Alu*I, respectively. Sticky ends were filled and the pBS *Bam*HI site in pBSN1, downstream of the N gene, was used as second restriction site. Fragments were recloned between the filled *Eco*RI site and the *Bam*HI site of pBS.

In vitro *transcription and translation.* Plasmid DNA of constructs pBSN1 to pBSN4 was linearized downstream of the ORF with *Hin*dIII. After proteinase K treatment (200 μg/ml for 30 min at 37 °C) and phenol extraction, the DNA was ethanol-precipitated and dissolved in water. *In vitro* transcription was performed by incubating 1 μg of linearized plasmid DNA for 30 min at 37 °C in a mixture containing 40 mM-Tris–HCl pH 8·0, 50 mM-NaCl, 8 mM-$MgCl_2$, 10 mM-DTT, 2 mM-spermidine, 0·5 mM-ATP, -CTP and -UTP, 0·025 mM-GTP, 0·25 mM-^{7}mGpppG (cap analogue), 100 μg/ml bovine serum albumin, 1000 units/ml RNAsin and 750 units/ml T7 RNA polymerase. The template was removed by diluting the transcription mixture 10-fold in a buffer containing 40 mM-Tris–HCl pH 7·5, 10 mM-NaCl, 6 mM-$MgCl_2$, and 5 units/ml DNAse RQ (Promega), followed by an incubation for 15 min at 37 °C. RNA transcripts were phenol-extracted, ethanol-precipitated and dissolved in 10 μl water. For *in vitro* translation, 1 μl transcript was added to 10 μl of a rabbit reticulocyte lysate (Amersham; N-90) and 1·5 μl (25 μCi) [^{35}S]methionine. The translation mixture was incubated for 60 min at 30 °C.

Radioimmunoprecipitation and SDS–PAGE. Immunoprecipitation, using an anti-BEV rabbit antiserum, and protein gel electrophoresis were carried out as described previously (Snijder *et al.*, 1988).

Labelling of BEV proteins and virus purification. Viral proteins were labelled by incubating infected cell cultures (m.o.i. of about 10) from 9 to 15 h post-infection in medium containing 25 μCi/ml [^{35}S]methionine or [^{35}S]cysteine. ^{35}S-labelled cell lysates were prepared using the lysis method described by Horzinek *et al.* (1985). Labelled virions were isolated from the cell culture medium by ammonium sulphate precipitation and ultracentrifugation as described previously (Snijder *et al.*, 1988).

RESULTS

Sequencing and analysis of ORFs

Poly(A)-selected RNA from BEV-infected EMS cells was used to prepare an oligo(dT)-primed cDNA library as described in Methods. Recombinant transformants containing viral sequences were selected by colony hybridization with kinase-labelled fragments of alkali-digested genomic RNA, isolated from purified BEV. A hybridization with oligo(dT) was employed to identify cDNA representing the 3′ end of BEV mRNAs. For full details of this procedure see Snijder *et al.* (1990).

The oligo(dT)-positive cDNA clones 27 and 31, with insert lengths of 1·7 and 1·9 kb respectively, were selected for sequence analysis. In Fig. 1 a restriction map of the area between the poly(A) tail and a *Hin*dIII site 1074 nucleotides upstream is shown and the sequencing strategy is outlined.

The nucleotide sequences obtained from clones 27 and 31 were identical. Computer analysis revealed the presence of two overlapping ORFs at 680 to 200 and 658 to 394 nucleotides from the start of the poly(A) tail, encoding products of 160 and 88 amino acids, respectively. The 3′ terminus of an adjacent ORF was also identified (Fig. 1).

On the basis of their position in the sequence and the estimated RNA lengths (Snijder *et al.*, 1988), the two overlapping ORFs were assumed to be present on the 0·8 kb RNA 5. Northern blot hybridizations confirmed this postulate, whereas the sequence encoding the upstream ORF was absent from RNA 5 (data not shown). The nucleotide and deduced amino acid sequences of the two ORFs assigned to RNA 5 are presented in Fig. 2.

On the basis of its size and amino acid sequence, the 160 amino acid product encoded by the larger ORF was expected to be the BEV N protein. Its calculated M_r value of 18·3K is in good agreement with the 19K M_r of the BEV N protein, as it has been estimated from polyacrylamide gels (Snijder *et al.*, 1988). The product was found to be generally hydrophilic (data not shown) and to contain 22 (14%) basic and seven (4%) acidic residues, giving it a positive charge at neutral pH. The basic amino acids are clustered in two regions of the protein (Fig. 2). A group of 15 arginine residues in a 47 amino acid stretch was observed at amino acid positions 34 to 81. The remaining seven basic amino acid residues are located near the carboxy terminus of the protein, between amino acid positions 118 and 156.

Analysis of the second ORF, encoding an 88 amino acid product, revealed this polypeptide to have an M_r of 9·8K and to be generally hydrophobic (data not shown). A comparison of the sequences surrounding the AUG codons of both ORFs showed that the AUG codon of the smaller ORF is in a much less favourable context for use as a translation initiation signal (Kozak, 1987).

A sequence similarity search, using the FastP computer program (Lipman & Pearson, 1985) and the entries in the NBRF/PIR, NBRF/NEW and Swiss-Prot protein databases (releases 19.0, 37.0 and 9.0, respectively), did not result in the identification of any significant similarity between either of the BEV ORFs and other protein sequences.

Identification of the BEV N gene

The 480 nucleotide ORF was cloned downstream of the T7 promoter in transcription vector pBS as described in Methods. The resulting construct pBSN1 (Fig. 3) was transcribed *in vitro* using T7 RNA polymerase. *In vitro* translation of this transcript in a rabbit reticulocyte lysate resulted in the synthesis of a protein comigrating with the N protein from BEV-infected cells (Fig. 4*a*, lanes 2 and 4). This product could be immunoprecipitated by rabbit anti-BEV antiserum (Fig. 4*a*, lane 1).

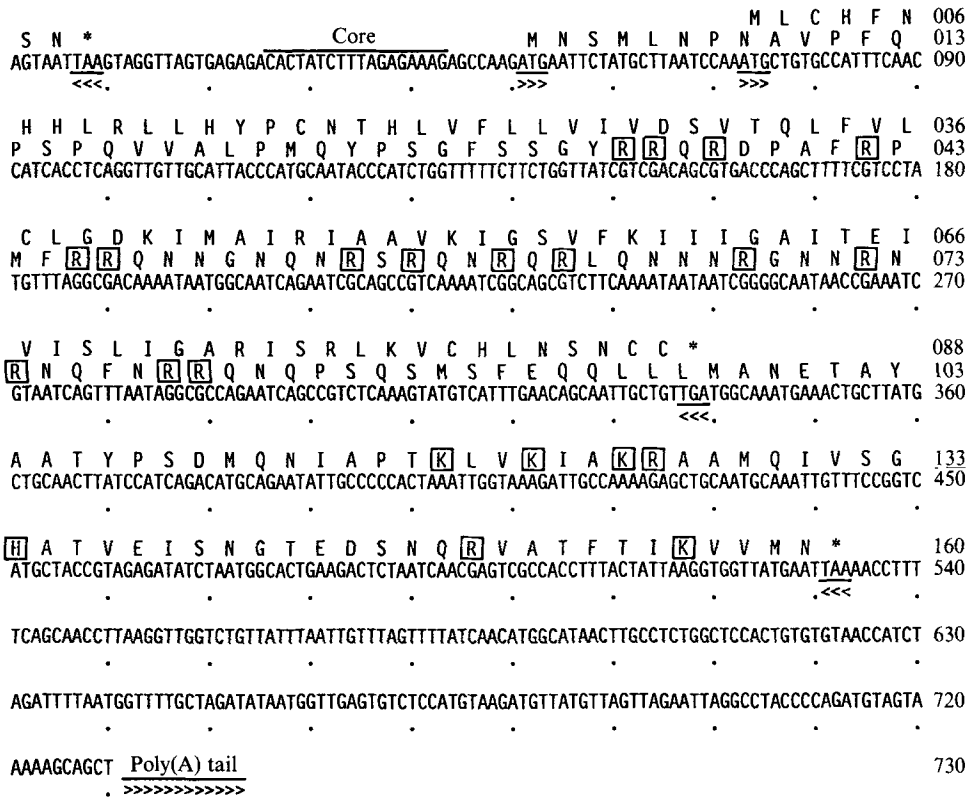

Fig. 2. Nucleotide sequence and predicted amino acid sequence of the BEV N gene. The 3′ untranslated sequences are also included in the figure. The predicted amino acid sequences of the BEV N protein and the hypothetical 9·8K product of the second ORF are presented, starting at nucleotides 51 and 73, respectively. The position of the termination codon of the preceding ORF and the postulated conserved core promoter sequence are indicated. Termination (<<<) and initiation (>>>) codons are underlined. The basic amino acid residues in the BEV N protein sequence are boxed.

Because cysteine residues are not encoded by the 480 nucleotide ORF, BEV-infected cells were labelled with either [^{35}S]methionine or [^{35}S]cysteine. Virus particles were purified from the cell culture medium and their proteins were analysed on a polyacrylamide gel. As is shown in Fig. 5 (lane 1), the N protein could be labelled in the presence of [^{35}S]methionine. In the presence of [^{35}S]cysteine no incorporation of label into the N protein was detected, whereas the other structural BEV proteins P, M and E could be identified (Fig. 5, lane 3).

Analysis of smaller N-related proteins

In vitro translation of pBSN1 transcripts gave rise to the synthesis of three additional polypeptides (Fig. 4*a*, lane 2). Two of these translation products were also recognized by the BEV antiserum (Fig. 4*a*, lane 1) and comigrated with two N-related proteins regularly found in lysates of BEV-infected cells (Fig. 4*a*, lanes 3 and 4; Horzinek *et al.*, 1985). Limited proteolysis experiments, using *Staphylococcus aureus* V8 protease, have revealed that these proteins share oligopeptides with the N protein (Horzinek *et al.*, 1985). During *in vitro* translation of RNA 5-containing sucrose gradient fractions, the same products are synthesized, which again suggests a relationship with the N protein (Snijder *et al.*, 1988).

The presence of three additional in-frame AUG codons in the 5′ part of the N ORF (at amino acid positions 4, 23 and 44) pointed towards initiation of translation on multiple AUG codons as a mechanism for the generation of smaller N-related proteins. This possibility was studied using the pBS *in vitro* transcription system. The first three AUG codons of the N gene were sequentially removed, generating transcription vectors pBSN2 to pBSN4, and allowing the use of each AUG codon for initiation of *in vitro* translation (Fig. 3).

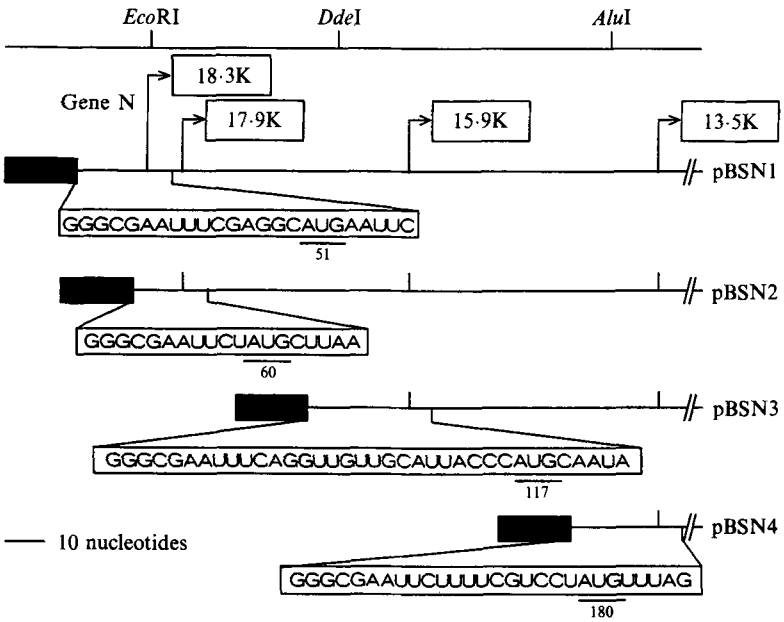

Fig. 3. Construction of transcription vectors pBSN1 to pBSN4, allowing initiation of translation at each of four in-frame AUG codons in the BEV N gene. Vertical bars indicate the positions of the AUG codons. Numbers indicate the position of the AUG codons in the sequence shown in Fig. 2. The M_r values of the predicted translation products are shown in open boxes. Dark boxes indicate the position of the T7 promoter in pBS. The sequence from the 5′ end of the transcript to the area containing the AUG codon is shown. Restriction sites used in the construction of the vectors are indicated.

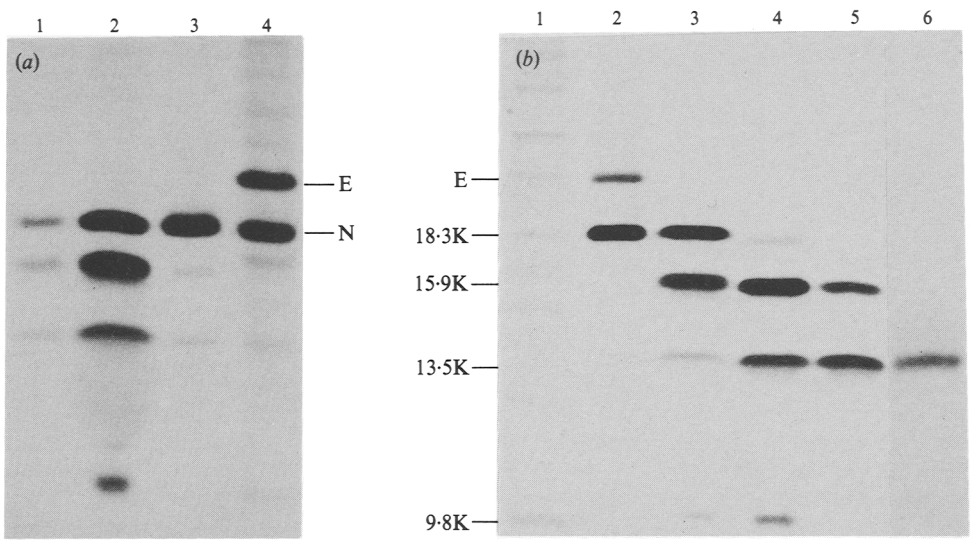

Fig. 4. *In vitro* transcription and translation of the BEV N gene and analysis of smaller, N-related polypeptides. Transcripts of constructs pBSN1 to pBSN4 were translated in the presence of [^{35}S]methionine using a rabbit reticulocyte lysate. Products were analysed by a 20% SDS–PAGE. (*a*) Lanes 2 and 4 show a direct analysis of an *in vitro* translation of pBSN1 transcripts and a [^{35}S]methionine-labelled BEV-infected EMS cell lysate, respectively. Proteins from the pBSN1-translation (lane 1) and from an infected cell lysate (lane 3) were immunoprecipitated using a rabbit immune serum. The positions of the E and N protein bands from the cell lysate are indicated. (*b*) Lanes 3 to 6 show a direct analysis of the translation products of pBSN1, pBSN2, pBSN3 and pBSN4 transcripts, respectively. In lanes 1 and 2 total lysates from uninfected (lane 1) and BEV-infected (lane 2) EMS cells were analysed.

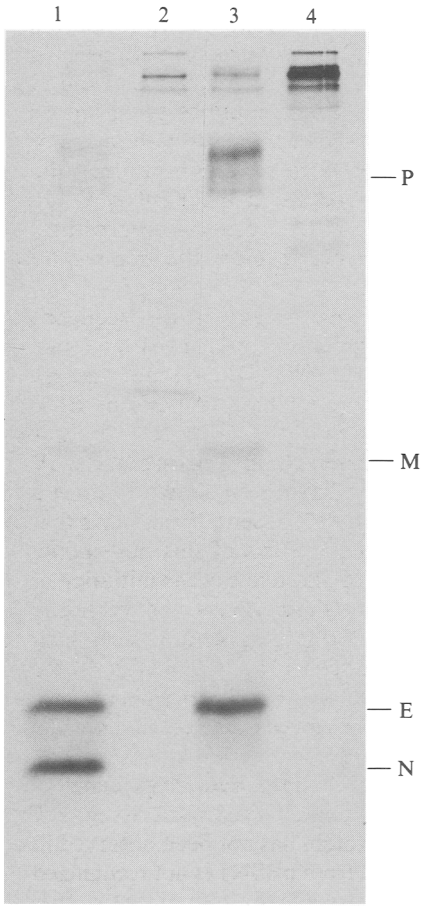

Fig. 5. Identification of the BEV N gene by metabolic labelling. Virion proteins were labelled with [^{35}S]methionine (lane 1) or [^{35}S]cysteine (lane 3) and virus preparations were analysed by 20% SDS–PAGE. The viral structural proteins are indicated. Lanes 2 and 4 show the preparations obtained from mock-infected EMS cell cultures, after labelling with [^{35}S]methionine and [^{35}S]cysteine, respectively.

The product initiated on the second AUG codon (pBSN2), which is only four amino acids shorter, virtually comigrated with the complete N protein (Fig. 4*b*, lanes 3 and 4). Initiation of translation on the second AUG codon of the N ORF was found to be inefficient compared with initiation on more downstream AUG codons of the same transcript (Fig. 4*b*, lane 4). Though the initiation context seems to be equally unfavourable for the second, third and fourth AUG codons of the N ORF (Fig. 2; Kozak, 1987), other factors, e.g. the distance from the 5′-end of a transcript to the AUG codon, may influence the initiation efficiency of translation.

The *in vitro* translation products of pBSN3 and pBSN4 transcripts did comigrate with the two smaller polypeptides found in BEV-infected cell lysates (Fig. 4*b*, lanes 5 and 6).

Translation of pBSN1 and pBSN2 transcripts also yielded a product of about 10K, which was not recognized by the anti-BEV antiserum (Fig. 4*a*, lanes 1 and 2; Fig. 4*b*, lanes 3 and 4). This polypeptide was probably translated from the second, overlapping ORF (Fig. 1), which is intact in these constructs and lacks its AUG codon in pBSN3 and pBSN4.

DISCUSSION

In this paper the sequence analysis and identification of the gene encoding the BEV N protein is reported. The N gene is situated immediately upstream of the BEV poly(A) tail. In immunoprecipitations its 19K product is recognized by rabbit anti-BEV antiserum, and metabolic labelling of virus particles has confirmed the identification of this product as the BEV N protein. Earlier experiments have demonstrated the RNA binding capacity of this 19K protein (Horzinek *et al.*, 1985). The amino acid sequence of the N protein contains two clusters of basic amino acid residues. Their presence suggests a role in the binding of nucleic acid, an essential step in virion assembly.

In BEV-infected cells the presence of varying amounts of two additional virus-specific polypeptides has been described previously (Horzinek *et al.*, 1985). V8 protease digestions have shown that they share oligopeptides with the N protein. In this study their relation to the N protein was confirmed by the fact that, like the N protein, these polypeptides could not be detected after [^{35}S]cysteine labelling of viral proteins (data not shown). *In vitro* translation of transcripts of pBSN1 to pBSN4 showed that the generation of N-related proteins is probably the result of aberrant initiation of translation on multiple AUG codons in the BEV N ORF. Although traces of one of these peptides have been encountered in nucleocapsid preparations (Horzinek *et al.*, 1985), we consider these products are unlikely to play a role in the architecture of the virion.

The AUG codon of the BEV N ORF is located just downstream of the sequence 5′ CACUAUCUUUAGAGAAAG 3′ (Fig. 2). Highly similar sequence motifs have been found upstream of the AUG codons of other BEV ORFs (Snijder *et al.*, 1990). These sequences are postulated to function as the core of promoter regions that direct the synthesis of BEV subgenomic RNAs.

In addition to the BEV N gene, the sequence of the 3′-terminal non-coding region of the BEV mRNAs was obtained (Fig. 2). Recently the five BEV mRNAs have been shown to form a 3′-coterminal nested set (Snijder *et al.*, 1989). Although formal proof is absent, this implies that the 3′ ends of mRNAs and genomic RNA are probably identical. Parts of this non-coding sequence of 200 nucleotides, excluding the poly(A) tail, are likely to play a role in the initiation of the synthesis of the negative-stranded RNA. Interestingly, a computer analysis of secondary structure of the 3′ non-coding region revealed the presence of a large potential stem–loop structure ($\Delta G^0 = -120{\cdot}4$ kJ/mol). In the region between nucleotides 574 and 703 (Fig. 2) the formation of a stem of about 60 nucleotides with a loop at positions 629–638 is predicted (data not shown). This structure could be one of the recognition signals involved in the synthesis of the negative-stranded RNA.

An overlapping ORF, encoding a 9·8K product, was found within the nucleotide sequence of the N protein gene. Although a probable product of this ORF is found in *in vitro* translations of pBSN1 and pBSN2 transcripts (Fig. 4*b*, lanes 3 and 4), a 9·8K protein has not been observed in either BEV virions or in infected cells. The 9·8K product derived from pBSN1 is not recognized by a rabbit immune serum (Fig. 4*a*, lane 1). However, we cannot exclude the possibility that this ORF is a functional gene.

In conclusion, the BEV N protein, which is at the centre of the unique morphology of this virus, has been shown to be an 18·3K protein. Though such a small N protein is not unique, it emphasizes a major difference between the toroviruses and the coronaviruses, which seem to be most related to the proposed new family. Coronaviral N proteins have a characteristic size of 45K to 55K. In addition, the absence of any significant amino acid sequence similarity between the BEV N protein and other proteins underlines the separate taxonomic positions of the toroviruses.

The authors would like to thank Peter Bredenbeek for his advice on the construction and use of the pBSN transcription vectors. This study was supported by the Division for Health Research TNO in cooperation with the Foundation for Medical Research MEDIGON (project number 900-502-081).

REFERENCES

BEARDS, G. M., HALL, C., GREEN, J., FLEWETT, T. H., LAMOULIATTE, F. & DU PASQUIER, P. (1984). An enveloped virus in stools of children and adults with gastroenteritis that resembles the Breda virus of calves. *Lancet* ii, 1050–1052.

HORZINEK, M. C. & WEISS, M. (1984). Toroviridae: a taxonomic proposal. *Zentralblatt für Veterinärmedizin* **31**, 649–659.

HORZINEK, M. C., EDERVEEN, J. & WEISS, M. (1985). The nucleocapsid of Berne virus. *Journal of General Virology* **66**, 1287–1296.

KOZAK, M. (1987). An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research* **15**, 8125–8148.

LIPMAN, D. J. & PEARSON, W. R. (1985). Rapid and sensitive protein similarity searches. *Science* **227**, 1435–1441.

MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: A Laboratory Manual.* New York: Cold Spring Harbor Laboratory.

MESSING, J. (1983). New M13 vectors for cloning. *Methods in Enzymology* **101**, 20–79.

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.

SNIJDER, E. J., EDERVEEN, J., SPAAN, W. J. M., WEISS, M. & HORZINEK, M. C. (1988). Characterization of Berne virus genomic and messenger RNAs. *Journal of General Virology* **69**, 2135–2144.

SNIJDER, E. J., HORZINEK, M. C. & SPAAN, W. J. M. (1990). A 3′-coterminal nested set of independently transcribed messenger RNAs is generated during Berne virus replication. *Journal of Virology* (in press).

STADEN, R. (1986). The current status and portability of our sequence handling software. *Nucleic Acids Research* **14**, 217–233.

WEISS, M., STECK, F. & HORZINEK, M. C. (1983). Purification and partial characterization of a new enveloped RNA virus (Berne virus). *Journal of General Virology* **64**, 1849–1858.

WOODE, G. N., SAIF, L. J., QUESADA, M., WINAND, N. J., POHLENZ, J. F. L. & GOURLEY, N. (1985). Comparative studies on three isolates of Breda virus of calves. *American Journal of Veterinary Research* **46**, 1003–1010.

(*Received 28 June 1989*)