Cloning and sequencing the messenger RNA of the N gene of viral haemorrhagic septicaemia virus

J. Bernard, 1 F. Lecocq-Xhonneux, 2 M. Rossius, 3 M. E. Thiry 2 and P. de Kinkelin 1

1 Laboratoire d’Ichtyopathologie, Centre de Recherches de Jouy en Josas, Domaine de Vilvert, 78350 Jouy en Josas, France, 2 Eurogentec S.A., Campus du Sart Tilman, 4000 – Liège and 3 Laboratoire Central de Génie Génétique, Université de Liège, Liège, Belgium

The mRNA transcribed from the N gene of viral haemorrhagic septicaemia virus (VHSV) of salmonids has been cloned in Escherichia coli and expressed. Fusion proteins were recognized by monoclonal antibody directed against the N protein from the viral particle. A 1212 bp long open reading frame (ORF) coding for 404 amino acids with a calculated Mr of 44590 was deduced from the nucleotide sequence. The ORF was preceded by a 93 bp segment including in position 42 the AACAC pentanucleotide which is presumed to be the start signal for transcription by analogy with other rhabdoviral mRNAs. The upstream 41 bp region could correspond to the covalently linked positive polarity leader RNA as also found on the N mRNA from infectious haematopoietic necrosis virus (IHNV). This may be a characteristic of fish lyssaviruses. The AAACC sequence, which is part of the leader, was not found. Amino acids 44 to 359 from IHNV and 45 to 360 from VHSV are 45.3% homologous. A strong homology which could reflect functional importance was also found for potential phosphorylation sites and hydrophobic peaks despite the fact that the two viruses evolved on different continents.

Introduction

Viral haemorrhagic septicaemia virus (VHSV) of salmonid fishes is responsible for significant losses among the trout stocks of European fish farms (de Kinkelin et al., 1979). It belongs to the Lyssavirus genus of the Rhabdoviridae family. The virus does not drastically inhibit host cell synthesis despite the fact that the yield of progeny viral particles is high (Bernard et al., 1985). The genomic, negative polarity, viral RNA is transcribed sequentially and the order of the genes from 3’ to 5’ is thought to be similar to that of other members of the family i.e. N, M1, M2, G and L (Bernard & de Kinkelin, 1985). It seems that only the N protein is phosphorylated (Lenoir & de Kinkelin, 1975; McAllister & Wagner, 1975) and its Mr, as determined by SDS–PAGE varies from 38000 to 62000 depending upon the authors (for a review, see Deuter & Enzman, 1986).

The N gene is interesting because, since it is the first protein gene to be transcribed, its mRNA should be the most abundant in infected cells and thus would be the best target for sensitive molecular probing; also, as it encodes the nucleoprotein, which is supposedly tightly bound to the viral RNA and plays an important role in the transcription and replication processes, it is thus interesting to verify whether its sequence is conserved throughout the Lyssavirus genus.

cDNA reverse-transcribed from the mRNA of VHSV-infected cells was cloned simultaneously in our two laboratories into λZAP (J. Bernard) and pUC13 (F. Lecocq-Xhonneux & M. Thiry) vectors respectively. We present here the sequence of the N gene.

Methods

Virus and cells. The 07-71 pathogenic isolate of VHSV (Le Berre et al., 1977) was grown in the epithelioma papulosum cyprini (EPC) cell line (Fijan et al., 1983) as described (de Kinkelin & Le Berre, 1978). The number of passages in cell cultures prior to utilization was limited to two (Bernard et al., 1985). The same virus seed was used by both laboratories (provided by P. de Kinkelin).

Cloning of mRNA in pUC13 and M13 vectors. Cells were infected at a multiplicity of 10 and harvested 20 h later. Total RNA was prepared as described (Chomczynski & Sacchi, 1987) and poly(A)+ RNA was purified on an oligo(dT)-cellulose column (Pharmacia). The cDNA was synthesized using a cDNA synthesis kit (Boehringer Mannheim cat. no. 1013882) and inserted in the SmaI site of the pUC13 vector (Pharmacia) before transforming Escherichia coli strain D1210 (De Boer et al., 1983). Colonies were transferred onto replica filters and screened by hybridization with labelled single-stranded cDNA prepared by random primed reverse transcription of total RNA from
VHSV-infected or uninfected cells. The clones that hybridized only with RNA from infected cells were purified and their plasmids used as probes on Northern blots of RNA from infected cells. The colonies harbouring a plasmid that hybridized with an RNA of the putative mRNA were selected. The largest insert (clone pSHV-N2) was excised by HindIII and partial EcoRI digestion and transferred into M13 mp18 and M13 mp19 sequencing vectors.

Cloning of mRNA in λZAP and pBS vectors. The cells were harvested 8 h post-infection. Total nucleic acids were extracted in guanidinium isothiocyanate buffer (Maniatis et al., 1982) and total RNA was precipitated by 2 M-LiCl. Poly(A)+ RNA was purified on Hybond 8 h post-infection. Total nucleic acids were extracted in guanidinium isothiocyanate buffer (Maniatis et al., 1982) and total RNA was precipitated by 2 M-LiCl. Poly(A)+ RNA was purified on Hybond M13 mpl9 sequencing vectors.

DNA sequencing. The chain termination method (Sanger et al., 1977) was chosen, with modifications. We used either T7 modified DNA polymerase (Sequenase, USB), T7 DNA polymerase (Pharmacia) or Klenow enzyme (Boehringer Mannheim) on either M13 mp18 and M13 mp19 single-stranded or pUC and pBS double-stranded vectors. The reaction products were separated on 40 cm polyacrylamide-urea wedge spaced gels (BRL).

Universal primers were from USB and Boehringer, KS and SK primers from Stratagene. Specific primers were synthesized in our two laboratories.

Fusion protein. Fusion protein from insert SHV-N2 was made using the pATH series of plasmids (Dieckmann & Tzagoloff, 1985). Recombinant clones were grown overnight in M9 medium supplemented with 0.5% casamino acids and 20 mg/l of tryptophan. The culture was diluted 10-fold in the same medium without tryptophan and grown 1 h at 30 °C prior to induction with 5 mg/l of indoleacrylic acid (data not shown). The homopolymer was absent from pSHV-N1.

The sequencing strategy for SHV-N2 is presented in Fig. 1. Other primers were used for SHV-N1. The sequence of nucleotides 1 to 105 was obtained from M13 mp18 and M13 mp19 vectors and thus represents both strands of pSHV-N2.

The sequence from nucleotide 105 to 1324 was repeatedly verified on both strands of both inserts (SHV-N1 and N2) with different enzymes because the increasing number of G residues rendered the interpretation of the autoradiograms difficult. pSHV-N2 and its M13 derivatives contained a homopolymeric A segment 40 to 50 nucleotides long, which impaired the progression of both Sequenase and Klenow enzymes on one strand (signals of the same intensity were found in all four lanes of the sequencing gel). The homopolymer was absent from pSHV-N1.

The ORF

The messenger sense nucleotide sequence and the deduced polypeptide are presented in Fig. 2. The first AUG, included inside a Kozak (1987a, b) box, was observed at position 92 of the insert and the in-frame stop codon at position 1305. The 1212 base long ORF codes for a protein of 404 amino acids with a calculated Mr of 44590. Of the 26 serine residues which are potential phosphorylation sites, 17 are located among the 153 C-terminal amino acids. The peptide is 35.9% hydrophobic.
Fig. 2. Nucleotide sequence (messenger sense) and deduced polypeptide sequence of the VHSV N gene. *, Stop signal; transcription signal is underlined.

\[
\begin{align*}
    &\text{C TTT CAA UTT TGA AAT TUC AAG AEE TGU GAA AAT GAA AGT CCA AAG} \\
    &\text{G47 TCA AAT CAT AAC TCA TAA TGG TGG TAC AGA AGA ACT CAG TGC AAT} \\
    &\text{G48 TCA CAC GAC GCG CCA GCA GGA ATG AGA} \\
    &\text{G49 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G50 ATC GTC CTC ATG} \\
    &\text{G51 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G52 ATC GTC CTC ATG} \\
    &\text{G53 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G54 ATC GTC CTC ATG} \\
    &\text{G55 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G56 ATC GTC CTC ATG} \\
    &\text{G57 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G58 ATC GTC CTC ATG} \\
    &\text{G59 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G60 ATC GTC CTC ATG} \\
    &\text{G61 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G62 ATC GTC CTC ATG} \\
    &\text{G63 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G64 ATC GTC CTC ATG} \\
    &\text{G65 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G66 ATC GTC CTC ATG} \\
    &\text{G67 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G68 ATC GTC CTC ATG} \\
    &\text{G69 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G70 ATC GTC CTC ATG} \\
    &\text{G71 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G72 ATC GTC CTC ATG} \\
    &\text{G73 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G74 ATC GTC CTC ATG} \\
    &\text{G75 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G76 ATC GTC CTC ATG} \\
    &\text{G77 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G78 ATC GTC CTC ATG} \\
    &\text{G79 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G80 ATC GTC CTC ATG} \\
    &\text{G81 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G82 ATC GTC CTC ATG} \\
    &\text{G83 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G84 ATC GTC CTC ATG} \\
    &\text{G85 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G86 ATC GTC CTC ATG} \\
    &\text{G87 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G88 ATC GTC CTC ATG} \\
    &\text{G89 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G90 ATC GTC CTC ATG} \\
    &\text{G91 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G92 ATC GTC CTC ATG} \\
    &\text{G93 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G94 ATC GTC CTC ATG} \\
    &\text{G95 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G96 ATC GTC CTC ATG} \\
    &\text{G97 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G98 ATC GTC CTC ATG} \\
    &\text{G99 GAA CTG GCC CTG ACT GTC AAT GAT GAT AAC GAA AAG} \\
    &\text{G100 ATC GTC CTC ATG} \end{align*}
\]

Fig. 3. β-gal-N fusion protein. (a) The β-gal-N fusion protein and β-gal were electrophoresed in parallel and the gel was stained with Coomassie blue. M, markers: (right) 94K, 67K, 43K, 30K, 20K; (left) Rainbow (Amersham). 200K, 92.5K, 69K, 46K, 21.5K, 14K. Between 200K and 92.5K, the migration distance is not proportional to the log, of the M,. (b) The induced trpE-N, non-induced trpE-N and induced trpE bacterial extracts were electrophoresed in parallel. M, markers (Bio-Rad): 92.5K, 68K, 43K, 31K, 21.5K, 14K. (c) Gels: the β-gal–N, β-gal and purified virus proteins were electrophoresed in parallel, then transferred onto nitrocellulose membranes and treated with either F5 (N-specific) or I10 (G-specific) monoclonal antibodies at the indicated dilution. Spots: 1 to 0.13 μg (serial dilutions) of β-gal–N and β-gal proteins were spotted onto nitrocellulose and treated with monoclonal antibodies as above.

**Fusion protein**

To establish definitely that pSHV-N2 and pSHV-N1 inserts corresponded to the N gene, the EcoRI fragment from pSHV-N2 was fused in-frame with pATH1 and that of pSHV-N1 in-frame with the SmaI site of pUEX2. The overexpressed trpE-N protein (insert N2, Fig. 3b) migrated in SDS–PAGE at 70000 to 73000 Mr. As trpE accounts for 35000 Mr, the N protein is 35000 to 38000 Mr, in good agreement with the Mr of the native protein. The β-galactosidase (β-gal)–N protein (insert N1) and β-gal, which has an Mr of 120000, are presented in Fig. 3(a). Both proteins were clearly recognized by N-specific monoclonal antibody F5 but not by G-specific monoclonal I10 (Fig. 3c for N1).

**Search for homologies on the polypeptide**

Using the Microgenie (Beckman; Queen & Korn, 1984) Hopp & Woods (1981) computer programs, comparison with another fish lyssavirus, infectious haematopoietic necrosis virus (IHNV) (Gilmore & Leong, 1988) revealed a 45.3% homology between amino acids 45 to 360 of VHSV and amino acids 44 to 359 of IHNV (Fig. 3a). Of the 26 serine residues, six were in homologous positions in both IHNV and VHSV. Five were between amino acids 251 and 354 for VHSV and amino acids 254 and 353 for IHNV (Fig. 4a). Of the 145 hydrophobic residues, 50 were strictly homologous to IHNV and 37 homologous by substitution (Fig. 4a). The hydrophobicity patterns were conserved in four domains (Fig. 5).
Fig. 4. Homologies between VHSV and other rhabdoviruses. (a) Comparison of deduced polypeptide sequences of VHSV (our data) and IHNV (Gilmore & Leong, 1988). Bold: matches; underlined: homologous by substitution. (b) Comparison of the 5' transcription start signal and initiation codon of VHSV N mRNA and isofunctional sequences of other rhabdoviruses: VHSV, our data; IHNV, Gilmore & Leong (1988); rabies virus, Tordo et al. (1986a); SYNV, Zuidema et al. (1986); VSV, Rose (1980); spring viraemia of carp virus (SVCV), Roy et al. (1984).

Search for homologies of the regulatory sequences

The pentanucleotide AACAC appeared at position 42, similar to the N mRNA transcription signal demonstrated for most other unsegmented negative strand RNA viruses (Fig. 4b) and the uncapped internal GTP-initiated in vitro transcript start signal described for vesicular stomatitis virus (VSV) (Schubert et al., 1982). This pentanucleotide was also found in position 316 but nowhere else outside the ORF. The ATATC pentanucleotide representative of the promoter sequence preceding rabies virus M1 mRNA (Tordo et al., 1986b) and each VSV mRNA (Rose, 1980) was found only once, in position 53.

Discussion

The ORF presented here is 1212 bases long and has a coding capacity for a polypeptide of $M_c = 44590$. The published values of the $M_c$ of VHSV N protein vary depending upon the authors from 38000 (Lenoir & de Kinkelin, 1975) to 62000 (Hill et al., 1975). An Np polypeptide has also been described (de Kinkelin et al., 1980) that migrates faster than the N protein, has the same tryptic pattern (J. Bernard, unpublished results) and is recognized by the same monoclonal antibody (Fig. 3c).

Between amino acids 45 to 360 of VHSV and amino acids 44 to 359 of IHNV (Fig. 4a), 144 amino acids are
strictly homologous and 36 are homologous by substitution. Such strong homology is remarkable because the two viruses evolved separately (North America for IHNV and Europe for VHSV) and suggests that this segment is of functional importance.

The basic amino acids are randomly distributed and represent only 12% of the polypeptide whereas four hydrophobic domains are conserved (Fig. 5) between VHSV and IHNV. Thus we suggest that they represent the sequence for anchorage with the genomic RNA (Ollis, 1987).

It has been demonstrated that the NS protein of VSV associates via phosphoryl residues to the N protein to form soluble complexes (Masters & Banerjee, 1988a, b) and regulate RNA synthesis. The NS protein of VSV, the N and M1 proteins of IHNV and rabies virus, but only the N protein of VHSV are phosphorylated. As six out of the 26 serine residues we found for VHSV are located at sites homologous with IHNV (Fig. 4a), they probably represent functionally important phosphorylation sites.

The AACAC transcription signal was evident at position 42 of VHSV (Fig. 2 and Fig. 4b). This is surprising since the two banks were obtained from polyadenylated RNA. Gilmore & Leong (1988) have recently described a similar interruption of monocistronic transcription between the leader sequence and the N gene for IHNV. It may be a characteristic of fish lyssaviruses. No AAACC was found upstream from AACAC as has been described both for IHNV (Gilmore & Leong, 1988) and other rhabdoviruses (Tordo et al., 1986a; Zuidema et al., 1986; Rose, 1980; Roy et al., 1984). Thus either the leader of VHSV is incomplete or it is only partially covalently linked with the N mRNA. On the other hand, the number of bases between AACAC and ATG is 2, 52 and 25 for VHSV, sonchus yellow net virus (SYNV; a plant rhabdovirus) and IHNV respectively, but only five bases for rabies virus.

ATAC, which Tordo et al. (1986a) considered to represent the promoter sequence, was found in position 53 (Fig. 2), after the transcription signal for VHSV and nowhere on the leader for IHNV (Gilmore & Leong, 1988). Thus the role of the pentanucleotide may have to be reconsidered.

We are greatly indebted to the virologists from Laboratoire de Virologie, INRA, Jouy en Josas for helpful discussion and especially to D. Rasschaert and J. Cohen who helped with the computer programs. Part of this work was supported by the AIP Genétique et Pathologie grant no. 4385.

References


(Received 30 October 1989; Accepted 9 April 1990)