Journal of General Virology (1995), **76**, 2519–2527. *Printed in Great Britain*

Complete genomic sequence of the fish rhabdovirus infectious haematopoietic necrosis virus

H. Schütze,[1]* P.-J. Enzmann,[2] R. Kuchling,[1] E. Mundt,[1] H. Niemann[2] and T. C. Mettenleiter[1]

[1] *Federal Research Centre for Virus Diseases of Animals, Institute of Molecular and Cellular Virology, Friedrich-Loeffler-Institutes, D-17498 Insel Riems and* [2] *Institute of Microbiology, Paul-Ehrlich-Strasse 28, D-72076 Tübingen, Germany*

The complete nucleotide sequence of the genome of the fish rhabdovirus infectious haematopoietic necrosis virus (IHNV) has been determined after cDNA cloning of the viral genomic RNA. Sequence analysis showed the presence of six open reading frames encoding the nucleoprotein N, the matrix proteins M1 and M2, the glycoprotein G, a so-called non-structural protein NV, and the RNA polymerase L. The genome organization is 3′ N-M1-M2-G-NV-L 5′. The extreme 5′ and 3′ ends of the genome were sequenced after RNA ligation or RACE. Prokaryotic expression products of the open reading frames predicted to encode the matrix proteins M1 and M2, the glycoprotein G and the NV protein reacted with rabbit anti-IHNV serum thereby confirming their identity. This is the first complete nucleotide sequence of a fish rhabdovirus. Knowledge of the complete sequence is an essential prerequisite for future manipulation of the genome and also serves to provide gene- and protein-specific reagents for use in further examination of the replication of the fish rhabdoviruses.

Introduction

Infectious haematopoietic necrosis (IHN) is an acute disease of salmonids characterized by rapid disease progression and high mortality. IHN virus (IHNV) is enzootic in the sockeye salmon population on the west coast of North America (McAllister, 1979). In recent years rainbow trout has gained importance as an additional major host for IHNV. In 1977 the disease was introduced into Japan and in 1987 it appeared in Europe (Bovo *et al.*, 1987; Laurencin, 1987; Sano *et al.*, 1977). It now represents a major threat to aquaculture all over Europe.

IHNV is a member of the rhabdovirus family which also includes other pathogens such as vesicular stomatitis virus (VSV) and rabies virus (RV). Ultrastructurally, rhabdoviruses display a bullet-shaped morphology with glycoprotein spikes projecting from the viral envelope. The genome is associated with the nucleoprotein N in a ribonucleoprotein core. The rhabdovirus genome consists of an unsegmented single-stranded RNA of approximately 11–12 kb with negative polarity (reviewed in Baer *et al.*, 1990; Wagner, 1990). So far, complete genomic sequences have only been determined for VSV and RV (Rose, 1980; Tordo *et al.*, 1988; Conzelmann *et al.*, 1990). The availability of complete sequence information on rhabdovirus genomes, including that for the authentic genomic termini, was a major prerequisite for the recent successful establishment of a system allowing synthesis and manipulation of infectious rhabdovirus RNA (Schnell *et al.*, 1994). This achievement opens the way for genetic manipulation of rhabdoviruses following the example set for RNA viruses containing plus-stranded genomes.

The IHNV genome has been reported to consist of approximately 11000 bases (Hill *et al.*, 1975; Kurath & Leong, 1985). Differences were observed between IHNV and other rhabdoviruses. IHNV was reported to specify six mRNAs rather than five mRNAs as is found for the other rhabdoviruses analysed to date (Kurath *et al.*, 1985; Kurath & Leong, 1987). These five mRNAs encode the viral structural proteins RNA polymerase (L), envelope glycoprotein (G), nucleocapsid protein (N) and the two matrix proteins (M1 and M2). The sixth IHNV mRNA encodes a unique so-called non-structural protein which is synthesized in infected cells but appears to be excluded from mature virions (Kurath *et al.*, 1985; Kurath & Leong, 1985, 1987). Based on the available data, the gene organization of the IHNV genome has

* Author for correspondence. Fax +49 38351 7219.

The complete nucleotide sequence of the IHNV genome (N, M1, M2, G, NV and L genes) has been deposited in GenBank and assigned the accession number X89213.

been proposed as 3′ N-M1-M2-G-NV-L 5′ (Kurath *et al.*, 1985; Kurath & Leong, 1985). However, only the G and N genes have been sequenced so far (Gilmore & Leong, 1988; Koener *et al.*, 1987).

Precise knowledge of the IHNV genome is essential for further detailed studies on the replication of fish rhabdoviruses and for possible genetic manipulation of the genome, opening up new ways for construction of vaccines and preparation of efficient diagnostic reagents. Here we report the cloning and sequencing of the complete genome of IHNV. Predicted open reading frames (ORFs) were verified by reactivity of the respective prokaryotically expressed proteins with anti-IHNV serum.

Methods

Cells and virus. The pathogenic IHNV was originally isolated from rainbow trout in Oregon in 1969, and was kindly provided by Dr P. de Kinkelin (INRA, Unité de Virologie et Immunologie Moléculares, Jouy-en-Josas, France). The virus was adapted to growth in epithelioma papulosum cyprini (EPC) cells after five end-point passages. For virus propagation the cells were infected at an m.o.i. of 0·1–1 at 15 °C.

Isolation of RNA. RNA was extracted from virions by the guanidinium thiocyanate procedure as described by Chomzynski & Sacchi (1987).

cDNA cloning. Virion RNA was used as a template for construction of cDNA libraries. Five micrograms of genomic RNA were incubated with 5 pmol phage T4 polynucleotide kinase-treated specific primer for 5 min at 65 °C and for an additional 5 min at 37 °C. After primer annealing the assay was adjusted to 50 mM-Tris–HCl at pH 8·3, 60 mM-KCl, 10 mM-$MgCl_2$, 1 mM-DTT and 20 U RNasin ribonuclease inhibitor (Promega) and incubated for 15 min at 37 °C. Thereafter 2·5 mM of each deoxynucleotide triphosphate (dNTP) and 50 U of Moloney murine leukaemia virus reverse transcriptase (Gibco BRL) were added and first-strand cDNA synthesis was allowed to proceed for 1 h at 43·5 °C. The oligonucleotide primer for the first-strand cDNA was derived from the published sequence of the IHNV N gene (5′ C GGA TCA CGA ACG ATG ACA AGC GCA C 3′) (Gilmore & Leong, 1988). Subsequent cDNA syntheses were executed by gene walking using primers deduced from our own sequence determinations. Second-strand synthesis was performed using the appropriate reaction mix from a cDNA synthesis kit (Pharmacia Biotech). The Klenow fragment generated by DNA polymerase was used to create blunt-ended cDNA before *Eco*RI adaptors (Pharmacia) were added. After selection for long cDNA fragments by agarose gel electrophoresis, they were ligated into *Eco*RI-digested λ ZAP II arms (Lambda ZAP II/*Eco*RI/Gigapack II Cloning Kit, Stratagene) or into *Eco*RI-digested and dephosphorylated pUC18 vector (Pharmacia). The resulting libraries were screened with [^{32}P]ATP-labelled oligonucleotides (phage cloning) or with fluorescein-dUTP-labelled primers (insertion into pUC18 vector) derived from our own sequence data.

Cloning of the 3′ and 5′ termini of the IHNV genome. The termini of the IHNV genome were cloned by RNA ligation followed by RT–PCR. To determine the authentic 3′ and 5′ ends of the IHNV genome 5 μg of virus RNA was ligated in the presence of 100 pmol of RNA ligation primer 5′ GAG ATC ATC ATC ATC ATC GAG 3′ in a final volume of 10 μl of ligation buffer (75 mM-Tris–HCl at pH 7·5, 0·1 mM-ATP, 10 mM-$MgCl_2$, 5 mM-DTT, 10% DMSO and 50 U RNasin) with 50 U of RNase-free RNA ligase (Pharmacia). The reaction mixture was incubated at 25 °C for 6 h and the ligation product was purified by phenol–chloroform extraction and isopropanol precipitation. Prior to cloning the 3′ end of the genome the RNA ligation primer was treated with phage T4 polynucleotide kinase (Sambrook *et al.*, 1989). First-strand synthesis was primed with 3′ antisense RNA ligation primer (5′ CTC GAT GAT GAT GAT GAT CTC 3′) for cloning the 3′ end and with IHNV-specific primer complementary to nt 10880–10909 (IX3; 5′ CAT CTT CGC ATC ATT CAA ACT ACA AAG CGG 3′) for cloning the 5′ end of IHNV genome. PCR amplification of the 3′ end of the genome was performed with a specific primer complementary to nt 320–346 (INC; 5′ AGC GGG CGG TGA GTC AAG TCG GAG GAG 3′). Using 5′ antisense RNA ligation primer (5′ CTC TAG TAG TAG TAG TAG CTC 3′) and synthetic oligonucleotide IX3 the 5′ end of the IHNV genome was amplified by PCR. The RT–PCR products were cloned blunt-ended into the *Sma*I site of plasmid pSP73 and sequenced.

PCR amplification. Plasmid DNA (50–100 ng) or the complete cDNA first-strand reaction product were amplified in 30 mM-Tricine, pH 8·4, 2 mM-$MgCl_2$, 5 mM-2-mercaptoethanol, 0·01% gelatin, 0·1% Thesit, 100 pmol of selected primers, 200 μM of each dNTP and 2 U Vent DNA polymerase (New England Biolabs) (Ponce & Micol, 1992). The reaction was performed in a DNA temperature cycler Autogene II (CLF Grant) for 30 cycles. The denaturation, annealing and extension conditions were 93 °C (1 min), 55 °C (2 min) and 72 °C (30 s), respectively, followed by a final incubation for 1 min at 72 °C.

Sequence analysis. The sequence of both strands was determined by the dideoxynucleotide chain termination method according to standard protocols (Sequenase version 2.0 DNA and Sequenase 7-deaza dGTP sequencing kits, USB/Amersham). Sequences were assembled and analysed with the Genetics Computer Group package version 7.3.1-Unix (Devereux *et al.*, 1984).

Construction of expression plasmids. Specific primers containing a *Nde*I restriction site at the 3′ end of the respective gene (genomic polarity) and a *Xho*I site at the 5′ end of genes M1, G, and NV, or a *Hin*dIII site at the 5′ end of the M2 gene, were used for amplification by PCR of the distinct M1, M2, G and NV genes. As templates for PCR, the respective cDNA clones encompassing the gene of interest were used, with the exception of the glycoprotein gene, which was amplified by RT–PCR on genomic IHNV RNA. Products were cloned blunt-ended into the vector pSP73 (Promega). After digestion with *Nde*I and *Xho*I or *Hin*dIII, respectively, fragments were subcloned into *Nde*I/*Xho*I- or *Nde*I/*Hin*dIII-cleaved expression vector pET-23b (Novagen, Angewandte Gentechnologie Systeme). In the resulting recombinant plasmids the distinct genes M1, M2, G and NV were under the control of the phage T7 promoter. Translation should start at the respective authentic start codon. XL1-Blue MRF′ cells (Stratagene) that had been transformed with positive clones were screened by restriction endonuclease analysis and partially sequenced to confirm correct insertion. Plasmids were then isolated and retransformed into *Escherichia coli* BL21 pLysS cells expressing phage T7 RNA polymerase after induction by IPTG (Studier & Moffatt, 1986).

Prokaryotic expression of IHNV proteins. LB medium (2 ml) containing 50 μg/ml methicillin, 50 μg/ml ampicillin and 34 μg/ml chloramphenicol was inoculated with a freshly picked bacterial colony and incubated under constant shaking until the OD_{600} reached 0·8. Bacteria were collected by centrifugation for 5 s at 500 *g* and resuspended in 20 ml of fresh medium. The culture was further incubated at 37 °C until an OD_{600} of 0·8 was reached. Expression of the inserted gene was induced by addition of IPTG up to a final concentration of 1 mM. After an additional 1·5 h at 37 °C the bacteria were again collected by centrifugation and then washed twice in ice-cold 50 mM-Tris–HCl, pH 7·6. The final bacterial pellet was frozen at

70 °C for 10 min, cells were lysed in 2% SDS and 2% 2-mercaptoethanol, and lysates were analysed by SDS–PAGE in a Miniprotean II (Bio-Rad). Immunodetection was performed after electrotransfer of the proteins to nitrocellulose membranes. As secondary antibodies horseradish peroxidase-conjugated rabbit IgGs were used and the binding was visualized by ECL detection (USB/Amersham). Rabbit anti-IHNV serum was prepared after immunization of a rabbit with IHN virions.

Results

Determination of the complete IHNV genomic sequence

cDNA libraries were constructed from genomic RNA isolated from IHNV virions. The cloning strategy is depicted in Fig. 1(*a*). The first cDNA clones were established using an oligonucleotide primer deduced from the published sequence of the IHNV N gene (nt 94–119 of EMBL/GenBank accession no. J04321; Gilmore & Leong, 1988; corresponds to nt 162–187 of the complete IHNV sequence presented here). IHNV-specific inserts were identified by hybridization with a ^{32}P-labelled oligonucleotide designated IN 783 corresponding to nt 783–800 of the published IHNV-N sequence (corresponds to nt 851–868 of the complete IHNV sequence presented here). Additional cDNA libraries were generated by priming first-strand cDNA synthesis with IHNV-specific oligonucleotides deduced from our own sequence determinations (gene walking). IHNV-specific fragments with the longest inserts were identified and sequenced. At least three independently derived clones were sequenced in both orientations to verify any particular area of the genome. Only one representative for each cDNA clone set is shown in Fig. 1(*a*).

The 3′ and 5′ ends of the IHNV genome were cloned by ligation of genomic IHNV RNA to synthetic oligonucleotides followed by PCR with an internal oligonucleotide primer deduced from our sequence. From each assay, six independent clones were sequenced to ascertain the presence of authentic 3′ and 5′ termini. Whereas the cloning of the 3′-terminal sequences proved to be highly efficient by this method the efficiency of obtaining clones containing the 5′ end was rather low.

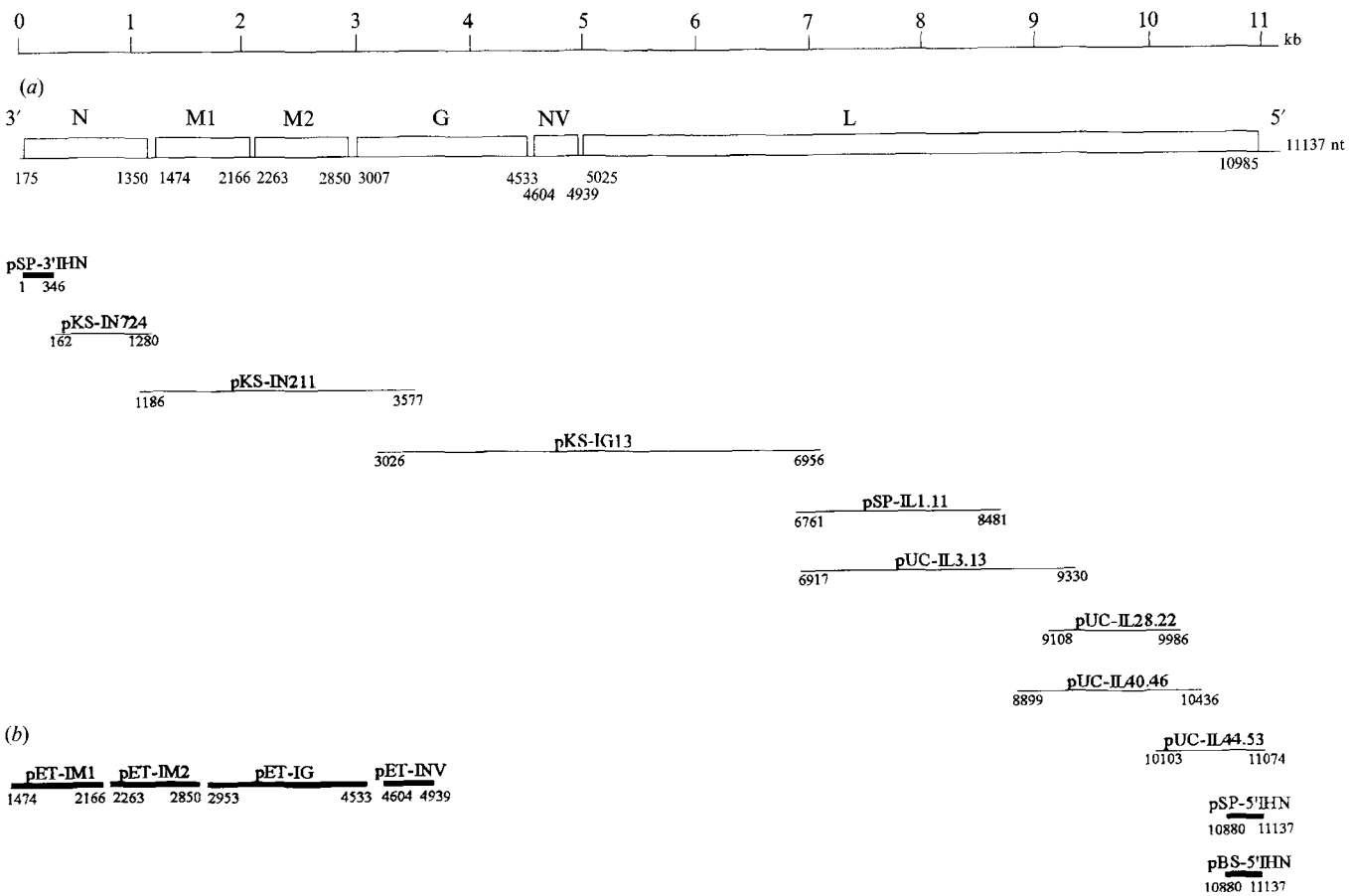

Fig. 1. Genetic map and cloning strategy of the IHNV genome. (*a*) Location of cDNA fragments derived from IHNV genomic RNA, including clones of the 3′ and 5′ termini generated by RT–PCR after RNA ligation and 5′ RACE. Only one out of each independent set of clones covering the entire genome is presented. (*b*) PCR-generated cDNA fragments used for expression of authentic viral proteins.

Table 1. *Comparison of IHNV proteins with those of other rhabdoviruses**

Protein	IHNV No. of amino acids	IHNV Mol. mass (kDa)	Identity of IHNV proteins to: VHSV	SVCV	RV	VSV
N	391	42	42	ND	19	17
M1	230	26	37	ND	17	19
M2	195	22	37	17	21	21
G	509	57	39	20	22	20
NV	111	13	ND	ND	–†	–
L	1986	225	ND	ND	23	21

* Deduced amino acid sequences of IHNV were compared with available sequences from viral haemorrhagic septicaemia virus (VHSV), spring viraemia of carp virus (SVCV), rabies virus (RV) and vesicular stomatitis virus (VSV). Percentages of identical amino acids are given. The number of amino acids and the calculated molecular mass are given for the respective IHNV proteins.

† –, Gene not found in RV or VSV.

ND, Not determined.

Therefore, 5′-RACE techniques were used as a second independent method to confirm the sequence of the authentic 5′ terminus. Using these techniques the complete nucleotide sequence of the IHNV genome has been established. It consists of 11137 nucleotides and encompasses six major predicted ORFs (Fig. 1*a*).

(*i*) *ORF1*

The first ORF, extending from nt 175–1350, contains 391 codons and corresponds to the previously described gene encoding the nucleoprotein N with a deduced molecular mass of 42 kDa. The deduced amino acid sequence of the N protein presented here exhibits 83% identity to the N protein sequence of isolate RB (Gilmore & Leong, 1988). Differences are mainly located in the region between aa 77–105 and in an area close to the carboxy terminus of the protein.

(*ii*) *ORF2*

The second ORF, which starts at nt 1474 and ends at nt 2166, contains 690 nucleotides with a coding capacity for a 26 kDa protein consisting of 230 amino acids. A putative transcription initiation site, 5′ AACA 3′, is located at position 1470. The deduced amino acid sequence shows 37% identity to the deduced sequence of the M1 protein of viral haemorrhagic septicemia virus (VHSV) (Benmansour *et al.*, 1994; see Table 1) which indicates that this ORF encodes the IHNV M1. No significant homology to matrix proteins of other rhabdoviruses has been detected (Rose, 1980; Tordo *et al.*, 1988; see Table 1).

(*iii*) *ORF3*

In a position collinear with that found in other rhabdoviruses, an ORF located between nt 2263 and 2850 encodes the IHNV matrix protein M2. The resulting protein consists of 195 amino acids with a calculated molecular mass of 22 kDa. A second possible start codon is found in-frame at position 2419. However, only the first start codon is in a context favourable for translation initiation (5′ RNNATG 3′) whereas the second ATG is unlikely to initiate protein synthesis due to its poor context (5′ YNNATG 3′) (Caverner, 1987). Comparison of the deduced amino acid sequences of IHNV and VHSV M2 shows 37% identity (Benmansour *et al.*, 1994; see Table 1) whereas similar to the situation for the M1 protein no significant identity to matrix proteins of other rhabdoviruses was detectable (Benmansour *et al.*, 1994; Kiuchi & Roy, 1984; Rose, 1980; Tordo *et al.*, 1988).

(*iv*) *ORF4*

The ATG at nucleotide position 3007 is in a favourable translation initiation context and most likely represents the start codon of the glycoprotein gene whereas in-

	Motif A			Motif B			Motif C			Motif D
IHNV	560	FIHVNKSLDINKFCTS	70	VFSGLKGGIEGLCQYVWTICLLLRV	12		ILAQGDNVIIT	134		ITGRIRASEMKLS
RV	610	RVTYAFHLDYEKWNNH	61	CWNGQDGGLEGLRQKGWSLVSLLMI	12		VLAQGDNQVLC	62		FRGNILVPESKRW
VSV	597	AICIANHIDYEKWNNH	59	CWQGQEGGLEGLRQKGWSILNLLVI	12		VLAQGDNQVIC	62		FRGVIRGLETKRW
SV	655	TLSCFLTTDLKKYCLN	60	FIHNPRGGIEGYCQKLWTLISISAI	12		AMVQGDNQAIA	59		YDGKILPQCLKAL
SV5	655	LAASFLTTDLKKYCLN	60	FIVSPRGGIEGLCQKAWTMISIAVI	12		SMVQGDNQAIA	59		YQGRILTQALKNA
HRSV	692	ISKCSIITDLSKFNQA	61	LYRYHMGGIEGWCQKLWTIEAISLL	12		ALINGDNQSID	58		HNGVYYPASIKKV
MarV	627	VRGASFVTDLEKYNLA	60	AYHYHLGGIEGLQQKLWTCISCAQI	12		SSVMGDNQCIT	59		LNGVQLPQSLKTM

Fig. 2. Alignment of the proposed catalytic domains of L polymerases of NNS RNA viruses. Proposed catalytic domains of deduced L polymerases of NNS RNA viruses were compared. Conserved residues are shown in bold type. Numbers correspond to the amino acid position within the respective L polymerase. Data are presented for IHNV (infectious haematopoietic necrosis virus; this report), RV (rabies virus; Tordo *et al.*, 1988), VSV (vesicular stomatitis virus; Schubert *et al.*, 1984; Yang & Lazzarini, 1983), SV (Sendai virus; Shioda *et al.*, 1986; Giesecke *et al.*, 1992), SV5 (simian virus 5; Higuchi *et al.*, 1992), HRSV (human respiratory syncytial virus; Stec *et al.*, 1991) and MarV (Marburg virus; Mühlberger *et al.*, 1992).

```
N-M1   .......... ..  .TGAAC ACAAACAGCC CCCCCCTCCT TCTTCCTCTC CCGCCCCTCG ACCCATCCAG

M1-M2  .......... .......... .......... .......... .......... .......... ..TAGACATC

M2-G   TAGGAGTCGA CCATGCCGTC TCTCACTCAC CCATCCATCG GCCGACCCAA CCCTCCTCCA TCCCCAGAGT

G-NV   .......... .......... .......... .......... .......... .......... .TAAAGGACC

NV-L   .......... .......... .......... .......... .......... .......... ..........

L-5'   .......... .......... .......... .......... .......... TAGGTACACT CCAGCAGCAC


N-M1   ATGTTTACTC CACCCTGACT CCCAGATAGA AAAAAATGGC ACTATAGTGC TTTTCAACTC AAACCACAAC

M1-M2  CTCCTCCAGG ATCCCGGCTA CCAAGACAGA AAAAAATGGC ACGCAAGTGT ATCGTTCCAA ACGAAATCCA

M2-G   CCCTCCACTC CTCACTCCGT CCAAGACAGA AAAAAATGGC ACTTTTGTGC TTTGAGACAG AACGCAACTC

G-NV   TCAATCCTCA CTTCCTCCCC ACCAGACAGA AAAAAACGGC ACATCTGTGC TGCAAAAAGA GACAATGGAC

NV-L   .......... .......... ..CAGATAGA AAAAAATGGC ACTTTTGTGC AAAAAAACTC AAGCGCGACT

L-5'   CCAGTGCCAG GGACCAAACT CCCAGATACA AAAAAATGGC TCCCGACTCC CATGGTAAGG GAACCAGTCA


N-M1   AATG..... .......... .......... .......... .......

M1-M2  AGAGTCAGTT CAAACGAGAG CATG..... .......... .......

M2-G   GCAGAGACCC ACCAAAACAA TG..... .......... .......

G-NV   CACCGCGACA TAAACACGAA CATG..... .......... .......

NV-L   CACAGAGAAC ACAACCAGCA ACGACAGCAA CCCATCACAG AAAAATG

L-5'   TTGTTTTCTG TTATTAAAAA CCTTTT.... .......... .......
```

Fig. 3. Comparison of intergenic regions within the IHNV genome. The regions between genes N and M1 (N–M1), M1 and M2 (M1–M2), M2 and G (M2–G), G and NV (G–NV), NV and L (NV–L), and the non-coding region behind ORF L (L–5′) were compared in antigenomic polarity. Stop and initiation codons are shown in bold type and are underlined. Poly(A) consensus sequences are shown in bold type.

frame upstream ATG codons at positions 2860 and 2953 are probably not used (Koener *et al.*, 1987; Kozak, 1987; Rose, 1980). The IHNV G gene extends until position 4533 and consists of 1524 nucleotides. The IHNV G gene is well characterized and our results confirm previous data (Koener *et al.*, 1987; Xu *et al.*, 1991). The identity between the G protein of the IHNV strain described here and the published sequence of IHNV isolate RB (Koener *et al.*, 1987) amounts to 98%.

(*v*) *ORF5*

Another ORF is located between nt 4319 and 4939 overlapping in a different reading frame with the G gene. Two additional possible start codons are found at positions 4604 and 4631. The ATG at position 4604 is situated in a perfect translation initiation context as proposed by Kozak (1987) and, therefore, it is likely that it represents the start codon for a 111 amino acid protein with a molecular mass of 13 kDa. The existence of a gene

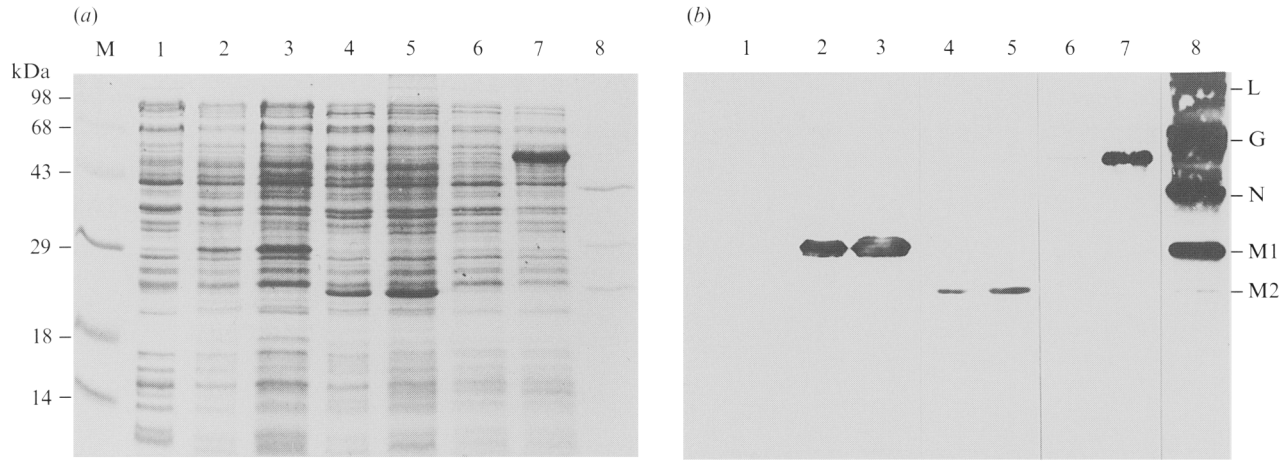

Fig. 4. Analysis of prokaryotically expressed IHNV proteins M1, M2 and G. Lysates of bacteria expressing M1 (lanes 2 and 3), M2 (lanes 4 and 5) or G (lanes 6 and 7) either before (lanes 2, 4 and 6) or after induction with IPTG (lanes 3, 5 and 7), as well as purified IHNV virions (lane 8), were fractionated in SDS–15% PAGE and either stained with Coomassie blue (*a*) or transferred to nitrocellulose and reacted with an anti-IHNV serum (*b*). M, Molecular mass marker proteins (sizes are indicated on the left). In lane 1 IPTG-induced bacteria containing the parental vector pET-23b were assayed as a negative control. IHNV virion proteins L, G, N, M1 and M2 are indicated on the right.

located between the G and L genes has been predicted by R-loop analyses (Kurath & Leong, 1985) and it has been proposed that this hypothetical gene may encode the so-called NV protein which has been detected in infected cells but not in mature virions (Kurath *et al.*, 1985; Kurath & Leong, 1985). Therefore our sequence confirms the existence of and for the first time gives actual sequence data for this rhabdoviral gene.

(*vi*) *ORF6*

The sixth ORF, which starts at nt 5025 and ends at nt 10985, is the largest gene predicted in the IHNV genome. It is capable of encoding a 1986 amino acid protein of approximately 225 kDa. The presence of highly conserved motifs A, B, C and D located between aa 558 and 832, which are typical for L proteins of viruses with a non-segmented single-stranded genomic RNA of negative polarity show that this ORF encodes the IHNV polymerase L (Tordo *et al.*, 1992). An alignment of the proposed catalytic domains of the IHNV L protein compared with other RNA-dependent RNA polymerases is shown in Fig. 2.

(*vii*) *Intergenic regions*

The intergenic region contains the conserved sequence 5′ AGAYAG/C 3′ (antigenomic polarity) which is followed by a stretch of seven adenosine residues. This sequence is present at the end of every sizeable ORF found in the IHNV genome and is similar to that found in intergenic regions of VHSV [AGATAG$(A)_7$], RV [NTG$(A)_7$] and VSV [TATG$(A)_7$]. However, no identity exists with the common eukaryotic consensus polyadenylation signal (Benmansour *et al.*, 1994; Bernard *et al.*, 1990; Thiry *et al.*, 1991; Gallione & Rose, 1983; Rose, 1980; Tordo *et al.*, 1988). A comparison of the intergenic regions of IHNV is presented in Fig. 3.

Expression of predicted IHNV ORFs in E. coli

To confirm the location of the ORFs predicted from the nucleotide sequence to encode matrix proteins M1 and M2, glycoprotein G and the NV protein, these were expressed in *E. coli* and subsequently reacted with a rabbit anti-IHNV serum.

(*i*) *Proteins M1 and M2*

The location of fragments for expression of proteins M1 and M2, as obtained by PCR, is shown in Fig. 1 (*b*). After expression, bacterial lysates were analysed by SDS–PAGE (Fig. 4). The M1 protein migrated with an apparent molecular mass of 29 kDa (Fig. 4, lanes 2 and 3), whereas M2 migrated as a 20 kDa protein (Fig. 4, lanes 4 and 5). Both sizes are comparable to those of authentic M1 and M2 proteins as found in IHNV virions (Fig. 4, lane 8). To analyse the antigenicity of the expressed proteins after transfer to nitrocellulose, they were reacted with a rabbit anti-IHNV serum. Both M1 (Fig. 4*b*, lanes 2 and 3) and M2 proteins (Fig. 4*b*, lanes 4 and 5) were recognized by the anti-IHNV serum confirming their authenticity.

(*ii*) *Glycoprotein G*

The complete G gene starting with the second possible ATG at nt 2953 was expressed in *E. coli* and analysed. The location of the fragment used for this expression is shown in Fig. 1 (*b*). The 59 kDa *E. coli* expression product

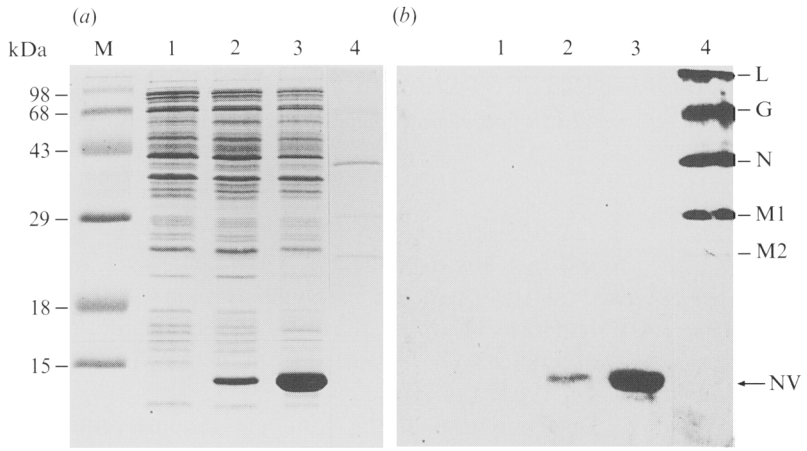

Fig. 5. Analysis of prokaryotically expressed IHNV protein NV. Lysates of bacteria expressing NV either before (lane 2) or after induction with IPTG (lane 3) as well as purified IHNV virions (lane 4) were fractionated in SDS–20% PAGE and either stained with Coomassie blue (*a*) or transferred to nitrocellulose and reacted with an anti-IHNV serum (*b*). For further details see legend to Fig. 4.

(Fig. 4*a*, lanes 6 and 7) reacted with the anti-IHNV serum after Western blotting (Fig. 4*b*, lanes 6 and 7). Faster migration of the *E. coli*-expressed G protein as compared to virion-derived G protein is presumably due to the lack of glycosylation in the prokaryotically expressed protein.

(*iii*) *Protein NV*

The NV gene was amplified from nt 4604–4939 by PCR using a cDNA clone as template. The location of the fragment is shown in Fig. 1(*b*). The predicted molecular mass of the expression product is 13 kDa. After induction of prokaryotic expression with IPTG followed by SDS–20% PAGE a prominent band of approximately 13 kDa could indeed be seen in bacterial lysates (Fig. 5*a*, lanes 2 and 3). To examine the antigenicity of the expressed NV protein, Western blot analysis with the rabbit anti-IHNV serum was performed. As shown in Fig. 5(*b*), lanes 2 and 3 the anti-IHNV serum reacted with the 13 kDa NV protein.

Discussion

We report here the first complete nucleotide sequence of the genome of a fish rhabdovirus, IHNV. The IHNV genome is 11137 nucleotides in length, which is in agreement with the previously estimated size for the virus RNA (Kurath *et al.*, 1985; Kurath & Leong, 1985, 1987). The deduced genome organization is presented in Fig. 1 and confirms previous predictions (Kurath *et al.*, 1985; Kurath & Leong, 1985).

Viruses containing a single-stranded non-segmented RNA genome of negative polarity (NNS) are composed of a limited number of proteins ranging from five (family *Rhabdoviridae*, genus *Lyssavirus*, RV; genus *Vesiculovirus*, VSV) to ten (family *Paramyxoviridae*, subfamily *Pneumovirinae*, genus *Pneumovirus*, human respiratory syncytical virus). In general, their genomes are organized into three distinct conserved gene blocks encoding (1) the nucleoprotein and polymerase cofactors, (2) matrix and envelope proteins and (3) the polymerase (Tordo *et al.*, 1992). The regions separating these conserved gene blocks may contain additional genes, such as the second nucleoprotein gene (VP30) located in the block 2–3 inter-region between the membrane protein VP24 and the L protein gene in the *Filoviridae* (Murphy *et al.*, 1990), the pseudogene ('psi') in RV located between the G and L genes (Conzelmann *et al.*, 1990; Morimoto *et al.*, 1989; Sacramento *et al.*, 1992; Tordo *et al.*, 1986), a second matrix protein (M2) in the pneumovirus RSV genome between the genes encoding the fusion protein F and polymerase L (McIntosh & Chanock, 1990) or the non-structural glycoprotein gene located between the G and L genes in the genome of bovine ephemeral fever virus (Walker *et al.*, 1992). Our sequence analysis of the IHNV genome confirmed reports of the existence of an additional gene located between the G and L genes encoding a protein designated as NV which has not been detected in virions so far but is present in IHNV-infected cells. The presence of this NV gene is one of the major differences between IHNV and other members of the *Rhabdoviridae* that have been sequenced. Computer analysis predicted five possible phosphorylation sites within the NV protein, two with specificity for casein kinase II (aa 8 and 72), and three specific for protein kinase C (aa 44, 57 and 58). Whether the IHNV NV protein is indeed phosphorylated is unknown as is its role in the viral replication cycle. Surprisingly, the prokaryotically expressed NV protein reacted with the rabbit anti-IHNV serum in a Western blot. Whether this might be indicative of a virion location for NV protein or whether the virion preparation used for immunization of rabbits contained contaminating infected cell debris is unclear. In contrast, NV protein was not detectable by Western blotting with the antiserum in our purified virion preparation. A specific anti-NV serum, which we

```
1 GTATAAAAAAAGTAACTTGACTAAGCTCAGAAGGACACAAGGACAGAAAA 50
  |||||||||||||||  |    ||||  ||||    |  | ||
11137 TATTTTTTTCATTGTCTTCCCAAGAGTTTTCCGTCAATTTTTTCCAAA 11090
```

Fig. 6. Complementarity between the 3′ and 5′ ends of the IHNV genome. Vertical lines indicate complementary nucleotides. Numbers relate to the complete IHNV genomic sequence.

are in the process of preparing, would help to solve this question.

A major characteristic of rhabdovirus genomes is the presence of a conserved polyadenylation sequence present at the termini of the individual protein-coding genes. In IHNV the sequence AGAYAG/C$(A)_7$ is found at the termini of the ORFs, which is similar to respective sequences at the ends of the N, M1, M2 and G genes of VHSV (Bernard *et al.*, 1990; Benmansour *et al.*, 1994; Thiry *et al.*, 1990), VSV (Rose, 1980) and RV (Tordo *et al.*, 1988).

Another common feature in non-segmented negative-stranded RNA virus genomes is the presence of complementary nucleotide motifs at the 3′ and 5′ ends of the genome. This is also true for the IHNV genome as shown in Fig. 6. At the 3′ end this motif is part of a 174 nt non-translated leader sequence, whereas at the 5′ end 152 non-translated nucleotides are found.

To examine possible relationships between IHNV and other rhabdoviruses, deduced amino acid sequences of IHNV proteins were compared with available deduced sequences of other rhabdoviral proteins. Results are shown in Table 1. The respective proteins of IHNV and VHSV are approximately 37–42% identical, whereas the identity to RV or VSV is significantly lower at 16–23%. However, conserved motifs specific for particular proteins within the *Mononegavirales* were detected. Proteins M1 and M2 contain several potential phosphorylation sites. Similar to VHSV M1 protein, IHNV M1 protein contains a high proportion of approximately 9% of serine (Ser) and threonine (Thr) residues, resulting in five conserved Ser (SXXE/D) and one Thr (KKTT) potential phosphorylation sites in addition to further phosphorylation motifs. The first 20 amino acids of IHNV M2 exhibit an especially high identity of 75% to the corresponding region of the VHSV M2 (Benmansour *et al.*, 1994).

The L gene product exhibits structural characteristics of the family of RNA-dependent RNA polymerases. Alignment of L protein sequences of IHNV and other rhabdoviruses highlights the high degree of identity. The acidic motif DLSTLD (aa 63–68) and the basic LKR motif (aa 240–242) are also represented in the IHNV L protein. The presence of the sequence GDN in motif C is in agreement with the previously described differences between the L proteins of segmented and non-segmented RNA virus genomes. The major motifs A (aa 559–574), B (aa 645–669), C (aa 682–692) and D (aa 837–849) found in RNA-dependent RNA polymerases are localized in the predicted catalytic domain of the IHNV polymerase. However, the distance between motifs C and D differs from that reported from related viruses (Tordo *et al.*, 1992; Fig. 3).

Determination of the complete genome sequence of the fish rhabdovirus IHNV, the third complete rhabdoviral sequence determined following those of VSV and RV, forms the basis for further molecular studies on the replication of fish rhabdoviruses. Synthesis of infectious rhabdoviral particles derived from plasmid templates has recently been achieved in the rabies virus system (Schnell *et al.*, 1994). A similar approach should also be feasible for IHNV. This opens the way for further studies on the role that individual viral proteins play in the viral replication cycle and holds great promise for manipulation of the viral genome to genetically engineer a novel class of safe attenuated rhabdovirus vaccines.

The authors thank B. Klupp for help in controlling sequence data, B. Bettin and M. Schlecht for synthesis of synthetic oligonucleotides and H. Stephan for photographic assistance.

References

BAER, G. M., BELLINI, W. J. & FISHBEIN, D. B. (1990). Rhabdoviruses. In *Virology*, 2nd edn, pp. 883–930. Edited by B. N. Fields & D. M. Knipe. New York: Raven Press.

BENMANSOUR, A., PAUBERT, G., BERNARD, J. & DE KINKELIN, P. (1994). The polymerase-associated protein (M1) and the matrix protein (M2) from a virulent and an avirulent strain of viral hemorrhagic septicemia virus (VHSV), a fish rhabdovirus. *Virology* **198**, 602–612.

BERNARD, J., LECOCQ-XHONNEUX, F., ROSSIUS, M., THIRY, M. E. & DE KINKELIN, P. (1990). Cloning and sequencing the messenger RNA of the N gene of viral haemorrhagic septicaemia virus. *Journal of General Virology* **71**, 1669–1674.

BOVO, G., GIORGETTI, G., JORGENSEN, P. E. V. & OLESEN, N. J. (1987). Infectious haematopoietic necrosis: first detection in Italy. *Bulletin of the European Association of Fish Pathology* **7**, 124.

CAVERNER, D. C. (1987). Comparison of consensus sequence flanking translational start sites in *Drosophila* and vertebrates. *Nucleic Acids Research* **15**, 1353–1361.

CHOMZYNSKI, P. & SACCHI, N. (1987). Single step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Analytical Biochemistry* **162**, 156–159.

CONZELMANN, K.-K., COX, J. H., SCHNEIDER, L. G. & THIEL, H.-J. (1990). Molecular cloning and complete nucleotide sequence of the attenuated rabies virus SAD B19. *Virology* **175**, 485–499.

DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.

GALLIONE, C. J. & ROSE, J. K. (1983). Nucleotide sequence of a cDNA clone coding for the entire glycoprotein from the New Jersey serotype of vesicular stomatitis virus. *Journal of Virology* **46**, 162–169.

GIESECKE, H., NEUBERT, N., OBERMAIER, B. & DOMDEY, H. (1992). Rapid sequencing of the Sendai virus 6·8 kb large (L) gene through primer walking with an automatic sequencer. *Journal of Virological Methods* **38**, 47–60.

GILMORE, R. D. & LEONG, J. C. (1988). The nucleocapsid gene of infectious hematopoietic necrosis virus, a fish rhabdovirus. *Virology* **167**, 644–648.

HIGUCHI, Y., MIYAHARA, Y., KAWANO, M., TSURUDOME, M., MATSUMURA, H., KUSAGAWA, S., KOMADA, H., NISHIO, M. & ITO, Y. (1992). Sequence analysis of the large (L) protein of simian virus 5. *Journal of General Virology* **73**, 1005–1010.

HILL, B. J., UNDERWOOD, B. O., SMALE, C. J. & BROWN, F. (1975). Physico-chemical and serological characterization of five rhabdoviruses infecting fish. *Journal of General Virology* **27**, 369–378.

KIUCHI, A. & ROY, P. (1984). Comparison of the primary sequence of spring viremia of carp virus M protein with that of vesicular stomatitis virus. *Virology* **134**, 238–243.

KOENER, J. F., PASSAVANT, C. W., KURATH, G. & LEONG, J. C. (1987). Nucleotide sequence of a cDNA clone carrying the glycoprotein gene of infectious hematopoietic necrosis virus, a fish rhabdovirus. *Journal of Virology* **61**, 1342–1349.

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

KURATH, G. & LEONG, J. C. (1985). Characterization of infectious hematopoietic necrosis virus mRNA species demonstrates a non-virion rhabdovirus protein. *Journal of Virology* **53**, 462–468.

KURATH, G. & LEONG, J. C. (1987). Transcription *in vitro* of infectious haematopoietic necrosis virus, a fish rhabdovirus. *Journal of General Virology* **68**, 1767–1771.

KURATH, G., AHERN, K. G., PEARSON, G. D. & LEONG, J. C. (1985). Molecular cloning of the six mRNA species of infectious hematopoietic necrosis virus, a fish rhabdovirus, and gene order determination by R-loop mapping. *Journal of Virology* **53**, 469–476.

LAURENCIN, B. (1987). IHN in France. *Bulletin of the European Association of Fish Pathologists* **7**, 104.

MCALLISTER, P. E. (1979). Fish viruses and viral infections. In *Comprehensive Virology*, vol. 14, pp. 401–470. Edited by H. Fraenkel-Conrat & R. R. Wagner. New York & London: Plenum Press.

MCINTOSH, K. & CHANOCK, R. M. (1990). Respiratory syncytial virus. In *Virology*, 2nd edn, pp. 1045–1072. Edited by B. N. Fields & D. M. Knipe. New York: Raven Press.

MORIMOTO, K., OHKUBO, A. & KAWAI, A. (1989). Structure and transcription of the glycoprotein gene of attenuated HEP-Flury strain of rabies virus. *Virology* **173**, 465–477.

MÜHLBERGER, E., SANCHEZ, A., RANDOLF, A., WILL, C., KILEY, M. P., KLENK, H. D. & FELDMANN, H. (1992). The nucleotide sequence of the L gene of Marburg virus, a filovirus: homologies with paramyxoviruses and rhabdoviruses. *Virology* **187**, 534–547.

MURPHY, F. A., KILEY, M. P. & FISHER-HOCH, S. P. (1990). *Filoviridae*. In *Virology*, 2nd edn, pp. 933–942. Edited by B. N. Fields & D. M. Knipe. New York: Raven Press.

PONCE, M. R. & MICOL, J. L. (1992). PCR amplification of long DNA fragments. *Nucleic Acids Research* **20**, 623.

ROSE, J. K. (1980). Complete intergenic and flanking gene sequences from the genome of vesicular stomatitis virus. *Cell* **19**, 415–421.

SACRAMENTO, D., BADRANE, H., BOURHY, H. & TORDO, N. (1992). Molecular epidemiology of rabies in France: comparison with vaccine strains. *Journal of General Virology* **73**, 1149–1158.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular Cloning: A Laboratory Manual*. 2nd edn. New York: Cold Spring Harbor Laboratory.

SANO, T., NISHIMURA, T., OKAMOTO, N., YAMAZAKI, T., HANADA, H. & WATANABE, Y. (1977). Studies on viral diseases of Japanese fishes. VI. Infectious hematopoietic necrosis (IHN) of salmonids in the mainland of Japan, *Journal of the Tokyo University of Fisheries* **63**, 81–85.

SCHNELL, M. J., MEBATSION, T. & CONZELMANN, K.-K. (1994). Infectious rabies viruses from cloned cDNA. *EMBO Journal* **13**, 4195–4203.

SCHUBERT, M., HARMISON, G. G. & MEIER, E. (1984). Primary structure of the vesicular stomatitis virus polymerase (L) gene: evidence for a high frequency of mutations. *Journal of Virology* **51**, 505–514.

SHIODA, T., IWASAKI, K. & SHIBUTA, H. (1986). Determination of the complete nucleotide sequence of the Sendai virus genome RNA and the predicted amino acid sequences of F, HN and L proteins. *Nucleic Acids Research* **14**, 1545–1563.

STEC, D. S., HILL, M. G. & COLLINS, P. L. (1991). Sequence analysis of the polymerase L gene of human respiratory syncytial virus and predicted phylogeny of nonsegmented negative-strand viruses. *Virology* **183**, 273–287.

STUDIER, F. W. & MOFFATT, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* **189**, 113–120.

THIRY, M., LECOCQ-XHONNEUX, F., DHEUR, I., RENARD, A. & DE KINKELIN, P. (1990). Molecular cloning of the mRNA coding for the G protein of the viral haemorrhagic septicaemia (VHS) of salmonids. *Veterinary Microbiology* **23**, 221–226.

THIRY, M., LECOCQ-XHONNEUX, F., DHEUR, I., RENARD, A. & DE KINKELIN, P. (1991). Sequence of a cDNA carrying the glycoprotein gene and part of the matrix protein M2 gene of viral haemorrhagic septicaemia virus, a fish rhabdovirus. *Biochimica et Biophysica Acta* **1090**, 345–347.

TORDO, N., POCH, O., ERMINE, A., KEITH, G. & ROUGEON, F. (1986). Walking along the rabies genome: is the large G-L intergenic region a remnant gene? *Proceedings of the National Academy of Sciences, USA* **83**, 3914–3918.

TORDO, N., POCH, O., ERMINE, A., KEITH, G. & ROUGEON, F. (1988). Completion of the rabies virus genome sequence determination: highly conserved domains among the L (polymerase) proteins of unsegmented negative-strand RNA viruses. *Virology* **165**, 565–576.

TORDO, N., DEHAAN, P., GOLDBACH, R. & POCH, O. (1992). Evolution of negative-stranded RNA genomes. *Seminars in Virology* **3**, 341–357.

WAGNER, R. R. (1990). *Rhabdoviridae* and their replication. In *Virology*, 2nd edn, pp. 867–881. Edited by B. N. Fields & D. M. Knipe. New York: Raven Press.

WALKER, P. J., BYRNE, K. A., RIDING, G. A., COWLEY, J. A., WANG, Y. & MCWILLIAM, S. (1992). The genome of bovine ephemeral fever rhabdovirus contains two related glycoprotein genes. *Virology* **191**, 49–61.

XU, L., MOURICH, D. V., ENGELKING, H. M., RISTOW, S., ARNZEN, J. & LEONG, J. C. (1991). Epitope mapping and characterization of infectious hematopoietic necrosis virus glycoprotein, using fusion proteins synthesized in *Escherichia coli*. *Journal of Virology* **65**, 1611–1615.

YANG, F. & LAZZARINI, R. A. (1983). Analysis of the recombinant event generating a vesicular stomatitis virus deletion defective interfering particle. *Journal of Virology* **45**, 766–772.

(*Received 9 May 1995; Accepted 16 June 1995*)