Vol. 31: 9-17, 1997

Sequence analysis and expression of the M1 and M2 matrix protein genes of hirame rhabdovirus (HIRRV) *

T. Nishizawa¹, G. Kurath², J. R. Winton^{2,**}

'Faculty of Applied Biological Science, Hiroshima University, Higashi-Hiroshima, 724 Japan ²Northwest Biological Science Center, 6505 NE 65th Street, Seattle, Washington 98115, USA

ABSTRACT: We have doned and sequenced a 2318 nucleotide region of the genomic RNA of hirame rhabdovirus (HIRRV), an important viral pathogen of Japanese flounder Paralichthys olivaceus. This region comprises approximately two-thirds of the 3' end of the nucleocapsid protein (N) gene and the complete matrix protein (M1 and M2) genes with the associated intergenic regions. The partial N gene sequence was 812 nucleotides in length with an open reading frame (ORF) that encoded the carboxylterminal 250 amino acids of the N protein. The M1 and M2 genes were 771 and 700 nucleotides in length, respectively, with ORFs encoding proteins of 227 and 193 amino acids. The M1 gene sequence contained an additional small ORF that could encode a highly basic, arginine-rich protein of 25 amino acids. Comparisons of the N, M1, and M2 gene sequences of HIRRV with the corresponding sequences of the fish rhabdoviruses, infectious hematopoietic necrosis virus (IHNV) or viral hemorrhagic septicemia virus (VHSV) indicated that HIRRV was more closely related to IHNV than to VHSV, but was clearly distinct from either. The putative consensus gene termination sequence for IHNV and VHSV, $AGAYAG(A)₇$, was present in the N-M1, M1-M2, and M2-G intergenic regions of HIRRV as were the putative transcription initiation sequences YGGCAC and AACA. An Escherichia coli expression system was used to produce recombinant proteins from the M1 and M2 genes of HIRRV. These were the same size as the authentic M1 and M2 proteins and reacted with anti-HIRRV rabbit serum in western blots. These reagents can be used for further study of the fish immune response and to test novel control methods.

KEY WORDS: Hirame . Rhabdovirus . Phosphoprotein . Matrix protein

INTRODUCTION

The initial isolation of hirame rhabdovirus (HIRRV) occurred in 1984 from moribund hirame (Japanese flounder) *Paralichthys olivaceus* and ayu *Plecoglossus altivelis* in Japan (Kirnura et al. 1986). In addition, natural infections have been recorded among black seabream *Milio macrocephalus,* and mebaru *Sebastes inermis* (Sano & Fukuda 1987). The related fish rhabdoviruses, infectious hematopoietic necrosis virus (IHNV) and viral hemorrhagic septicemia virus (VHSV), are highly pathogenic for salmonids and certain species of marine fish. The 3 viruses were originally classified as 'other' members of the lyssavirus genus of the *Rhabdoviridae* (Wunner & Peters 1991), but are now regarded as unassigned vertebrate rhabdoviruses (Wunner et al. 1995). The HIRRV, IHNV and VHSV virions are composed of a negativesense, single-stranded RNA genome and 5 structural proteins including a nucleocapsid protein (N), 2 matrix proteins (M1 and M2), a glycoprotein (G), and a polymerase (L). The M1 protein of the lyssaviruses is the equivalent of the polymerase-associated protein (P) of the vesiculoviruses (Wunner et al. 1995). The entire IHNV genome has been sequenced (Koener et al. 1987, Gilmore & Leong 1988, Morzunov et al. 1995,

^{&#}x27;Sequence data in this article have been deposited with the DNA data bank of Japan (DDBJ) under Accession No. D45422

[&]quot;Addressee for correspondence. E-mail: jim-winton@nbs.gov

Schutze et al. 1995), and all but a portion of the L gene of VHSV has been determined (Bernard et al. 1990, 1992, Thiry et al. 1991, Benmansour et al. 1994, Schutze et al. 1996). These molecular studies of the IHNV and VHSV genomes have revealed a consensus gene termination sequence, $AGAYAG(A)₇$, which differs significantly from the consensus gene termination signals of the mammalian lyssaviruses, $TG(A)₇$, or of members of the vesiculovirus genus, $YATG(A)₇$ (Koener et al. 1987, Thiry et al. 1991, Benmansour et al. 1994, Morzunov et al. 1995, Bjorklund et al. 1996, Kurath et al. 1997). This unique regulatory signal and the presence of a gene between the G and L proteins that codes for a nonvirion (NV) protein (Kurath & Leong 1985, Basurco & Benmansour 1995, Kurath et al. 1997) have led to the proposal that IHNV, VHSV and HIRRV are members of a novel genus of the Rhabdoviridae (Morzunov et al. 1995, Bjorklund et al. 1996, Kurath et al. 1997).

Distinguishable from other fish rhabdoviruses by serum neutralization tests (Kimura et al. 1986), the structural proteins of HIRRV formed patterns following sodium dodecyl **sulphate(SDS)-polyacrylamide** gel electrophoresis or 2-dimensional gel electrophoresis that were similar, but not identical, to those of IHNV and VHSV (Nishizawa et al. 1991). The G, N, and M2 proteins of HIRRV were shown to share some antigenic determinants with IHNV and VHSV by western blot analysis (Nishizawa et al. 1991), and it has been suggested that HIRRV might be considered a serotype of IHNV (Winton 1992). Although both HIRRV and IHNV are widely distributed within Japan, HIRRV has not been found outside of Japan, and conversely, IHNV is believed to have been originally restricted to western North America.

We were interested in determining the relationships between HIRRV, IHNV, and VHSV at the molecular level; however, until recently (Nishizawa et

al. 1995, Bjorklund et al. 1996, Kurath et al. 1997), there has been no sequence information available for the HIRRV genome. Here, we analyze the sequence data of the partial nucleocapsid protein gene and the 2 matrix protein genes of HIRRV, and compare them with the corresponding sequences available for IHNV and VHSV including the N-M1, M1-M2, and M2-G intergenic regions. In addition, we used an Escherichia coli expression system to produce recombinant M1 and M2 proteins in order to confirm the open reading frames (ORF) of the M1 and M2 genes, to better characterize the 2 proteins, and to create reagents for further studies on the function of the proteins and which can be used to test novel control strategies.

MATERIALS AND METHODS

Virus growth and purification. Hirame rhabdovirus (strain 8401-H) was grown in the epithelioma papulosum cyprini (EPC) cell line as described by Nishizawa et al. (1995). Virus was purified from cell culture fluid by centrifugation through a discontinuous gradient composed of 50, 35, and 20% (w/w) sucrose in STE (20 mM Tris, pH 7.4, 100 mM NaC1, 1 mM EDTA) buffer followed by centrifugation through a continuous gradient of 5 to 30 % (w/w) sucrose in STE as described by Leong et al. (1981).

cDNA cloning and sequence analyses. The viral RNA was extracted from gradient-purified virions as described by Kurath & Leong (1985) and cDNA was synthesized, cloned, and sequenced as described by Nishizawa et al. (1995). Sequencing primers are listed in Table 1. Two overlapping cDNA clones, pHR9 and pHR2ls1, were used that, together, represented a 2318 nucleotide region of the HIRRV genome containing the genes of interest. The DNA sequences were assembled and analyzed with the Mac DNASIS Pro (Hitachi), Macvector (Eastman Kodak), and PCGENE (IntelliGenetics) software programs. Sequence data were taken from GenBank for the N, M1, and M2 genes of IHNV (strain K, X73872; strain WRAC, L40883), VHSV (strain FI13, X73873; strain 07-71, D00687, U02624 and U03502; and strain Makah, X59241, U02630 and U03503), and for the NS and M genes of rabies virus (strain CVS, X55727, D10499) and vesicular stomatitis virus (serotype New Jersey, K03387, M31848).

Polymerase chain reaction (PCR) amplification and subcloning. Two forward and 2 reverse primers (Table 1) were used for specific PCR amplification of the M1 and M2 genes of HIRRV using the plasmid DNA of clone pHR2lsl as a template. The forward primers, FM1 and FM2, contained NdeI sites, and the

Table 1. Primers used for sequencing and polymerase chain reaction (PCR) amplification of the M1 and M2 genes of hirame rhabdovirus

reverse primers, RM1 and RM2, contained EcoRI or HindIII sites at their 5' termini, respectively. The PCR reactions were performed using Tag polymerase and 30 cycles of PCR, at 95°C for 1 min, 50°C for 1 min, and 72°C for 1 min. The PCR products were digested with NdeI and EcoRI or HindIII and ligated into the NdeI-EcoRI site (M1 construct) or into the Nde I-Hind111 site (M2 construct) of the pET25b(+) expression vector (Novagen) which was used to transform Escherichia coli strain BL21(DH3) according to the instructions of the manufacturer.

Expression of the target gene. Clones of transformed Escherichia coli containing M1 or M2 gene expression plasmids were cultured at 37'C in LB medium (1 % Bacto tryptone, 0.5 % yeast extract, 1 % NaCl, pH 7.5) containing 50 mg ml^{-1} ampicillin. The cells were induced by resuspension in fresh LB medium containing 50 mg m l^{-1} ampicillin and 1 mM isopropyl **P-D-thiogalactopyranoside** (IPTG). Following incubation at 37°C for 2.5 h, the cells were washed and resuspended in 50 mM Tris-HC1 (pH 8.0) and 2 mM EDTA to produce the total induced fraction. After addition of 100 mg ml^{-1} lysozyme and 0.1% Triton X-100, the cell suspension was incubated at 30°C for 15 min and then centrifuged at low speed $(300 \times g)$ for 15 min. The pellets and supernatants were used as insoluble induced and soluble induced fractions, respectively. These fractions were analyzed by 12% polyacrylamide gel electrophoresis (PAGE) under the reducing conditions of Laemmli (1970). Western blots were performed as described by Towbin et al. (1979).

RESULTS AND DISCUSSION

Cloning of the genomic RNA of HIRRV strain 8401-H produced 2 cDNA clones that together represented a contiguous 2318 nucleotide region of the HIRRV genome. Sequencing both strands of the clones produced unambiguous data that showed the clones included a partial N and full-length M1 and M2 genes (Fig. 1). This sequence is available as GenBank accession number D45422. For the purposes of this paper, we refer to the HIRRV 'genes' as the entire sequence from the first nucleotide following the (A) ₇ of the previous gene through the last A of the (A) ₇ tract because the transcription start and termination signals of the fish rhabdoviruses remain experimentally unproven.

Nucleotides 1 to 812 included a partial ORF encoding 250 amino acids and a stop codon that represented the carboxyl-terminal portion of the N gene because of sequence similarities with the 3' end (mRNA sense) of the N gene of IHNV (Morzunov et al. 1995, Schutze et al. 1995) and VHSV (Bernard et al. 1990, 1992). Sequence identities between the equivalent regions of

the partial ORFs of the N genes of HIRRV, IHNV, and VHSV are shown in Table 2. The nucleotide identities between HIRRV and the 2 strains of IHNV were higher than those between HIRRV and the 3 strains of VHSV that were analyzed, indicating that HIRRV was more closely related to IHNV than to VHSV. A CLUSTAL alignment of the deduced amino acid sequences of the partial N proteins of VHSV, IHNV and HIRRV showed several small (6 to 7 amino acid) blocks of highly conserved sequence in the central portion of the gene. The partial N gene of HIRRV showed approximately the same degree of sequence identity with the IHNV and VHSV strains as did the M1 and M2 genes; however, only a portion of the entire N gene was analyzed. The ORF of the HIRRV N gene was followed by a 62 nucleotide untranslated region, that included the putative transcription termination signal (discussed below).

The M1 gene of HIRRV occurred between nucleotides 813 and 1584 (Fig. 1). The longest ORF was 684 nucleotides in length (including the stop codon) and encoded a protein of 227 amino acids having a predicted molecular mass of 25779 Da (Nishizawa et al. 1995). The isoelectric point (pI) value of the deduced amino acid sequence was 8.68, which was more basic than that estimated by isoelectric focusing (p1 7.3 to 7.4, Nishizawa et al. 1991) probably due to differences in the degree of phosphorylation of the native form of the protein as produced by fish cells. Three potential serine/threonine phosphorylation sequences (SXXD/E and/or TXXD/E) occurred at amino acids 2, 122, and 186 of the deduced amino acid sequence. The M1 proteins of IHNV and VHSV are also basic and phosphorylated (MacAllister & Wagner 1975, Benmansour et al. 1994).

Sequence identities between the ORFs of the M1 genes of HIRRV, IHNV, and VHSV are shown in Table 2. As with the partial N gene ORF, the nucleotide identities between HIRRV and the 2 strains of IHNV were higher than those between HIRRV and the 3 strains of VHSV. Comparison of the deduced amino acid sequences of the M1 proteins of IHNV and HIRRV showed a highly variable region at the amino-terminus (amino acids 11 to 65) where only 13 of 55 (24%) of the amino acids were identical. This entire region was quite hydrophilic. An analogous region of the M1 protein of VHSV also showed significant variation among strains of the virus as noted by Benmansour et al. (1994). In the central portion of the M1 proteins of HIRRV and IHNV there was a highly conserved region (amino acids 108 to 147) where 39 of 40 (97.5 %) amino acids were identical. The location of the conserved and variable blocks of the M1 protein of HIRRV appeared typical for rhabdoviruses.

Spiropoulou & Nichol (1993) described a second ORF within the P gene (equivalent to the M1 gene) of vesic-

10 2 0 30 40 5 0 6 0 7 0 80 9 0 AAG ATG m **ACA TAC** W **CAG CAG MC CIC GAG 03** ClU **GCC MC ACT CAG** COI CTC **MC GAG** ClU **ACC** CIG **ATA A% ACG ACT Affi OCA** OIT CR: **R Q Q K L E R L A N S Q G L N E L T L I** 100 110 120 130 140 150 160 170 180 190
GCT CGG ATT GGG GCA GCG GTA AGA CCG GGG CAG AAG TTA ACC AAG GCC ATC TAC GGG ATC ATT CTT ATC AAC CTG ATG GAT CCC GCC ACT GCT GCC **ARIGAAVRPGQKLTKAIYGIILINLNDPATAA 200** 210 220 230 240 250 260 270 280 260 260 260 260 260 270 280 260 271 280 271 280 271 280 271 280 271 280 280 271 280 280 271 280 280 271 280 271 280 271 280 271 280 271 280 271 280 271 280 271 280 271 280 271 280 271 **CAMRLSGTGMTMVGLFNQAS** 290 300 300 310 310 320 330 340 350 360 370 370 380
GCA GAT~CTG~CTG~GAA~GAT~CTC~TGCATG^AAG^TCCATC^ATC~GAC^TCT~GCCAGGAGGATT^GTC^AAG^CTG^ATGAGGATC~GTT^GCA^GAC_GTC^GAG@AC ADLLEDLC**NKSIIDSARRIVKLMRIV 390 400 410 420 430 440 450 460 470 480 ACC OCA MC TAT** occ **ATC A% ATG AGC AGG ATG m** ccc **GAT** m **TAC m MG** TCC **TAC ax: ATC MT GAG AAC** 'TCA CGD **A= ACC** m **ATC** m **TAKYAINNSRNLGDGYFKSYGINENSRITCIL** 490 500 510 510 520 530 540 550 560 570
ATG AAC ATC AAC GAG CAG CAG TAC GAT GAG ACC AGG A CAC GA GAG GAC GO TO AGG TO TO A COCC COR A AGC CAG ACC GAG ATT GCT
MANIMUMBO ATCHERANT GAG GITT GG LAGUR GUR VSPPPRRXLAT LAT BIA 580 590 600 600 610 620 630 640 650 650 660 670
COC CITT CITC ANA ANG TAT AT CACCOGG ANC GOGA TCT TO GOGA GOGAT COT CAT CITT CITC COC CAG CAG CAG CAG CAG ANA CAG ACCOGG ANG ACCOGG ANC ACCOGGA ACCOGGA AND GOGATION GOGATION **680 690 700 710 720 730 740 750 760 770** GOC **GAG G%T GAT MG TAC GAT G40 GAG** GOC **GAG GAA GAC GGA OGG GAG GGA GAG** CC32 **GGA GAG** GAC GAA **GAG TAC TAC ?CA GACCCGGCAT CTATOFCGC GEDDKYDEEGEBDGGEGEGGEDEEYY 780 7 90 800 810 820 830 840 850 860 870** *cmxwmxc* **~MCATT ACTA~OI\TAOIVUIU MTCCC~COTAAM CMCAATCM** CAM- **AW m GAT MC** GAA GGA **GM CAG TIC** m **NsDNEGEQFF** 880 890 900 910 920 930 940 950 960 970
GAT ATT CCC AAG AAT GCT CTG GAC AGA GTT GAG GCA CGG ACG ATG TOT CCC AGG GAG GAT GGA AAG GTT OTC CGG AAA CAA GCG CCT CTA AAA GAG **KNALDRVEARTNCPREDGKVVRKQAP** 980 990 1000 1010 1020 1030 1040 1050 1050 1060 1070
GAA-CCA-AGA-CTG-GAG-GCA-GAG-CAG-AAG-GGG-TCT-CCA-AAG-AAG-CAG-GAG-AAA-CCC-CGG-GGA-ATG-CTC-CCT-CTG GAA-CHT-CTG CTT-CTG AAG-TAT-GTG-**RLEAEQKRSPKKQEKPRGMLPLEQLVLK** 1080 1090 1090 1100 1110 1110 1120 1120 1130 1140 1150 1150 1160
GTT GTG GTC GTG CTT GAT GCG CTC CHA GAG GTGA GGA GGA GTA ATT GCC CAG ATCAG CAG CAG TCT CHA GCC GAG TATG GTG C
V V C S L D A L R E F G G L I A Q I R Q S H Q A **1170 l180 1190 l200 l210 l220 1230 1240 1250 1260 GC% CPG** GCA **ACA GAG CAC** CGG **GCC AAT** CIC **CAG** OCG CTC **ACC MG** m **CAG CAC GAG CAC GAG AAA** DPC **TCC AAA GAG ATC** CIC TCG **GCA** DPC **ATC AVATERRANLQALTKSQQEHEKVSKEILSAVI** 1270 1280 1290 1300 1310 1320 1330 1340 1350
TCC ATC CGG TCC AAC CTC AAC GAG AAC TCC AGT CCC CGA CAC AAA CCC CTG GAC CTG GAC CAG GTC AAT GCG GAG AGA GCC CTC GGA TTT GGA GTC **SIRSNLNENSSPRHKPLDLDQVNAERALGFGV 1360 1370 1380 1390 1400 l410 1420 1430 1440 1450** CQi **TAC** CGG **ACC GCC TPD AAC** GlC m CC72 **AAA CTA** CGG **GGA ATC ACA CCA** GAA **GAG GCA ffiC** n*: **CM GAA** CPC **AAG MC ATG** GCh **AT AGA GAG GYRTALNVFGKLRCITPEEAGSQEVKNNAIRE 1460 1470 1480 1490 1500 1510 1520 1530 1540 1550** GCA **GAA GAA GAT GAG TAT GAG** COI **AGC AGA AGC TIC MG MG** GT2 **CrC GAT ATO** GlC **MC AAG ACC Am AGO TAGA-GC CMCAlCCCA ABEDEYEGSRSFPKKVLDNVKKTMR** 1620 1630 1560 1570 1580 1590 1600 1610 1640 1650 CTCGCCAACT CCCACTCCCC TAGATAGAAA AAAATGGCAC GTTAGTGTAC TGTTTTAACA GAAACCAAAA CCAAAGTCGA AAGAAACC ATG TCT CTC TTC AAG CGA ACC M S L F K R 1660 1670 1680 1690 1690 1700 1710 1720 1730 1730 1740 1750 1750
AAGAAGAAGAAGATTCTGATCCTCCTCCCCTCAGCCGGTTPAGGAAGACGAGCGGCGGTCGAGGTTCTAAATCAGGGGGCGGCGGTCAGAGGGGAGGGGGAGACGGAGA
KKTILIPPPPEHLLSGDBDRVTPVLNABGGBIKYL 1780 1790 1990 1800 1810 1820 1830 1840
GGG AAG CGA CCA ACC ACC CTC GAT GAG AAG ATC TAC TAC TCC ATG AGT CTG GCC GCC GCC ATT CTG GGA GGG AAT CTT CAT CCA TCT TTC CA **GKRPTTLDEKIYYSNSLAAAILGGNLHPSPQ** 1850 1860 1870 1880 1890 1900 1910 1920 1930 TOT CTG ACC TAT CTA TTT CAG CAG GAG ATG GAA TTC GGG AGC ACC ATA GAA AAG GTC AAC TTT GGG TCC CGG AAA CCG GCC CTC CTG ACA CCT $\mathbf L$ $\begin{array}{cccccccccccccc} \text{I} & & \text{E} & & \text{K} & & \text{V} \end{array}$ $\mathbf N$ \mathbf{L} L $\mathbf F$ Q Q E M E F G S T F G S R $\, {\bf P} \qquad {\bf A} \qquad {\bf P}$ 1940 1950 1960 1970 1980 1990 1990 2000 2010 2020 2020 2030
ATC AGG TG CAA AAG CCC GAG AAG TGT TCA TTC AGG CCT CTC GARC AAA AAG ATCA CCT COT CAT TAC ACT CTC AGT GTG GAG
IR WIQ KPEKC SFRPQ PLD KKIP PPQ IYT VSVEG **2040 2050 2060 2070 2080 2090 2100 2110 2120** GCG **ACC AE ACC m TCC** DCC **AGG m** m **m TCG** DCC **AGT CAT** DIC DCO %X' **GAC GAC MC** ca; **orc MC** CR: ocr aoc **cn: GAT ax m ATITFSGRPLFSASHVGCDDNRVKLAGLDGP**

 $\begin{array}{ccccccccc} & 2130 && 2140 && 2150 && 2160 && 2170 && 2180 && 2190 && 2200 && 2210 && 2100 &&$ **ITSPNYQRVKDYYAQETVLALSPEIPTKKGK**

2220 2230 2240 2250 2260 2270 2280 2290 2390 2300 2310
 TAG TCACAATGAC TCCCCATTTA AATCGACCCT CTCAAATCTA TCAACTCACA TC<u>AGATAGAA AAAAATGGCA C</u>ATTTGTGTG TTTTCATAAC AAACAATC

13

Fig. 1. Nucleotide (nt) sequence of a portion of the N gene and the complete M1 and M2 genes of HIRRV. The major ORFs of the N gene (nt 1 to 750), the M1 gene (nt 850 to 1533), and the M2 gene (nt 1639 to 2220) and the corresponding deduced amino acid sequences are shown. The gene junction consensus sequences and the putative transcription start sites are underlined

ular stomatitis virus (VSV), the prototype member of the vesiculovirus genus of the *Rhabdovindae.* This second ORF encoded 2 nested forms of a short (55 and 65 amino acids), highly basic, arginine-rich, protein that was detected in the cytoplasm of infected cells using rabbit antibody raised against a peptide synthesized from the deduced sequence. In the M1 gene of HIRRV, a second, very short (25 amino acid) ORF coding for a highly basic (p1 11.91), arginine-rich (5 of 25 amino acids), protein occurred in an analogous position within the second reading frame. This ORF began at nucleotides 893 to 970, 21 nucleotides downstream from a putative transcription initiation sequence (AACA) internal to the main ORF. Other short ORFs of 10 to 20 amino acids were also present in the M1 gene. One of these, at nucleotides 1067 to 1117, was also in the second reading frame and occurred 14 nucleotides downstream from an AACA sequence. Similar short, highly basic, second ORFs of the M1 (P) genes of rabies virus and IHNV were noted by Spiropoulou & Nichol (1993) and by Morzunov et al. (1995); however, like the second ORF of HIRRV, their

use remains unproven. It was interesting that the eukaryotic ribosome scanning model of Kozak (1986) predicts that the internal initiation of a second ORF is more likely when the primary start codon (AUG) is in a suboptimal context and the second AUG is in a more favorable context. This was true of the M1 gene of HIRRV where the primary ORF was only in fair context (G at position -3 , T at $+4$) while the second and third ORFs were in a better context (A at position -3, C/T at +4). In contrast, a nearly optimal Kozak's box occurs in the primary ORFs of the M1 and M2 genes of VHSV (Benmansour et al. 1994).

The M2 gene of HIRRV occurred between nucleotides 1585 and 2285 (Fig. 1). The longest ORF was 582 nucleotides in length (including the stop codon) and encoded a very basic protein of 193 amino acids having a predicted molecular mass of 21 615 Da and a p1 of 9.7 (Nishizawa et al. 1995). The primary AUG was in a good context for translation (A at position -3). While 2 short ORFs of 23 and 25 amino acids were identified near the 3' end of the M2 gene, these were in a poor Kozak's context and there was no indication that they

Table 2. Pairwise comparisons of the open reading frame nucleotide sequences (upper right) and deduced amino acid sequences (lower left) of HIRRV. 2 strains of IHNV and 3 strains of VHSV. Percent identities are shown for the partial N (top), complete M1 (middle) and complete M2 (bottom) genes. Sequence data were taken from GenBank for the N, M1 and M2 genes of IHNV (strain K, X73872; strain WRAC, L40883) and VHSV (strain FI13, X73873; strain 07-71, D00687, U02624 and U03502; and strain Makah, X59241, U02630 and U03503)

might be used. Four potential serine/threonine phosphorylation signals were evident in the M2 gene of HIRRV at amino acids 20, 43, 44, and 83. The M2 protein of VHSV was shown by Benmansour et al. (1994) to be phosphorylated, although at a lower level than the M1 protein. The equivalent protein of vesicular stomatitis virus (VSV), i.e. NS, is also phosphorylated (Wagner 1987).

Sequence identities of the M2 protein genes of HIRRV, IHNV, and VHSV are shown in Table 2. Analogous to that seen with the partial N and M1 ORFs, sequence identities between HIRRV and 2 strains of IHNV were significantly higher than the similarities between HIRRV and the 3 strains of VHSV. Comparison of the deduced amino acid virus to the cytoplasmic surface of the plasma **sample as lane A 4); lane 3: insoluble induced fraction of HIRRV M2** membrane during budding of the virus from the cell (Wagner 1987). A conserved region of basic

amino acids was present near the amino-terminus of the M2 proteins that included a triple proline sequence. This is a typical feature of M2 proteins of rhabdoviruses, and this domain is implicated in the inhibition of host transcription observed during infection (Ogden & Wagner 1986, Wagner 1987, Shipley et al. 1988).

The N-M1, M1-M2, and M2-G intergenic regions of HIRRV had a conserved sequence, $AGATAG(A)$, TGGCACRTTWGTGT, that occurred at positions 800 to 826, 1572 to 1598, and 2273 to 2299 (Fig. 2). This included the sequence $AGAYAG(A)₇$, which has been described as a putative consensus sequence for transcriptional termination and/or polyadenylation in IHNV, VHSV, and HIRRV (Gilmore & Leong 1988, Benmansour et al. 1994, Morzunov et al. 1995, Bjorklund et al. 1996). This consensus termination signal is thus far unique to these fish rhabdoviruses, and differs notably from the well-conserved consensus termination signals of the mammalian lyssaviruses, $WG(A)₇$ (Bourhy et al. 1993), or the vesiculoviruses, $YATG(A)_7$ (Banerjee & Barik 1992). The length of the untranslated portions of the N, M1 and M2 genes of HIRRV was similar to that of IHNV and VHSV (on average 50 to 63 nucleotides upstream and downstream of the ORF) with the exception of the region downstream of the ORF of the M2 gene. This downstream region of the M2 gene of HIRRV was 65 nucleotides in length, considerably

Fig. 2. SDS-PAGE and western blot analysis of in *vitro* **expression products from cultures of** *Escherichia col1* **containing the M1 and M2 genes of HIRRV. (A) Polyacrylamide gel (12%) stained with** sequences of the M2 proteins of IHNV and HIRRV coomassie blue. Lanes 1 and 10: structural proteins of purified HIRRV virions; lanes 2–5: products from cells transformed with pET25b(+) showed the M2 protein was somewhat more con-
served than the M1 (Table 2). Much of the varia-
transformed with pET25b(+) containing the HIRRV M2 gene; lanes 2 served than the M1 (Table 2). Much of the varia- **transformed with pET25b(+) containing the HIRRV M2 gene. Lanes 2** and 6: uninduced fraction; lanes 3 and $\overline{7}$: total induced fraction; in the central portion (amino acids 97 to 113) **lanes 4 and 8: insoluble induced fraction; lanes 5 and 9: soluble** where only 2 of 16 amino acids were identical. **induced fraction.** (B) Proteins separated by SDS-PAGE were trans**ferred to nitrocellulose membranes, incubated with anti-HIRRV** The matrix protein (equivalent to M2) of VSV is serum and detected utilizing peroxidase conjugated anti-rabbit Ig

known to play a critical role in virus assembly by antibody. Lane 1: structural proteins of purified HIRRV antibody. Lane 1: structural proteins of purified HIRRV virions; binding the helical ribonucleoprotein core of the **lane 2: insoluble induced fraction of HIRRV M1 gene (the same**

longer that the corresponding sequence of VHSV (35 nucleotides), but somewhat shorter than the untranslated region downstream of the M2 gene of IHNV (103 nucleotides).

The sequences AACABYMCT and AACAG have been described as consensus sequences for transcription initiation and/or capping signals of lyssavirus or vesiculovirus genes (Banerjee & Barik 1992, Bourhy et al. 1993). The sequence AACA, which has been described as a potential transcription start signal for the VHSV N, M1, M2, and G genes (Bernard et al. 1990, 1992, Benmansour et al. 1994), was present in each of the 3 HIRRV N-Ml, M1-M2, and M2-G intergenic regions (nucleotides 829, 832, 839, 1607, 2308 and 2311) downstream of the termination signals and before the primary initiation codon. It also occurred ahead of the potential second and third ORFs of the M1 gene (nucleotides 869 and 1049); however, the AACA sequence is absent from IHNV gene junctions (Koener et al. 1987, Gilmore & Leong 1988, Morzunov et al. 1995), and it has not been experimentally proven that the AACA sequence functions as a transcription start signal in these rhabdoviruses of fish. In contrast, the conserved sequence, YGGCAC, was present immediately following the termination signals of nearly all of the known genes of HIRRV, IHNV, and VHSV. Only the comparable sequences in the M1-M2 junction of VHSV (strain 07-71) and the M2-G junction of VHSV

(strain Makah) are slightly altered to YGGCTC or YGGTAC, respectively (Benmansour et al. 1994). Thus, this sequence appears to be a more likely transcription initiation signal in these fish rhabdoviruses as noted by others (Morzunov et al. 1995, Bjorklund et al. 1996, Kurath et al. 199?), but this remains to be proven experimentally.

To confirm the identity of the putative ORFs and to learn more about the antigenic nature of the 2 matrix proteins, the M1 and M2 genes of HIRRV were subcloned into the pET25b(+) vector and expressed in *Escherichia coli* cells. The expression vectors encoding both the M1 and M2 protein genes were genetically stable in the host cells, and there was no significant difference in codon usage between the 2 genes. The results of SDS-PAGE analysis of the expression products from the M1 and M2 genes are shown Fig. 2. While bacterially expressed proteins are not always identical to their native counterparts in size or in the degree of phosphorylation or glycosylation due to differences in protein processing and to the presence of amino acids encoded by sequences in the expression vector, in cells transformed with pET25b(+) encoding the M1 gene, an expressed protein with an apparent size identical to the native HIRRV M1 protein was detected. The expressed M1 protein of HIRRV was present in both the insoluble induced and soluble induced fractions. In contrast, no expression product of a size similar to the HIRRV M2 protein was detected in the total induced fraction of the cells transformed with pET25b(+) containing the M2 gene construct; however, a small amount of expressed M2 protein was detectable in the insoluble induced fraction. The identities of both expression products were confirmed by reaction with anti-HIRRV rabbit serum in western blot

Fig. 3. Dendrogram produced by CLUSTAL analysis of the deduced amino acid sequences of the M1 genes of HIRRV (HRVM1); 2 strains of IHNV, from North Amenca (IHNVM1) and Europe (IHNVKMI), and 3 strains of VHSV, from Europe (0771M1, VHSFI13Ml) and North America (MAKAHM1). The amino acid sequences of the NS protein (equivalent to phosphoprotein or M1 protein) of rabies virus strain CVS (RABCVSNS) and vesicular stomatitis virus serotype New Jersey (VSVNS) were used as representatives of the lyssavirus and vesiculovirus genera, respectively. Nearly identical dendrograms were generated by comparison of the partial N and the complete M2 amino acid sequences. Sequence data were taken from GenBank for the M1 genes of IHNV (strain K, X73872; strain WRAC, L40883) and VHSV (stram FI13, X73873, strain 07-71, U02624, and strain Makah, U02630) and for the NS genes of rabies virus strain CVS (X55727) and VSV serotype New Jersey (K03387)

analyses (Fig. 2), verifying that the ORFs identified by sequence analysis encode the viral proteins *in vivo.*

The M1 and M2 expression products will be valuable reagents for further research on the function of these proteins *in vivo* and may provide reagents for experimental control methods. Fish surviving infection with IHNV have been reported to produce anti-M1 antibody in their sera (Mourich & Leong 1992, LaPatra et al. 1993, Ristow et al. 1993) and the M1 protein of rabies virus was shown to stimulate a strong cytolytic T-cell response in mice (Larson et al. 1992). Thus, the recombinant M1 protein of HIRRV might be an attractive candidate for further research as a potential vaccine or as a supplement to recombinant vaccines based upon the G protein.

It has been shown that the expressed M2 protein of VSV is capable of inhibiting transcription *in vivo* in the absence of any other viral component (Black & Lyles 1992, Black et al. 1993, 1994). If the M2 protein of HIRRV has a similar inhibitory function, this may in part explain why the expression level of the HIRRV M2 protein in *Escherichia coli* was quite low in comparison to levels obtained for the M1 protein. It would be interesting to test the recombinant M2 protein for ability to inhibit HIRRV transcription *in vitro* and *in vivo* as a potential control strategy.

The presence of an NV gene, conserved gene junction sequences, and similarities in the nucleotide sequences of the N, M1, M2, G and NV genes have led several investigators to conclude that HIRRV belongs to the same unique group of fish rhabdoviruses that includes IHNV and VHSV and that the 3 viruses comprise a novel genus within the *Rhabdoviridae* (Morzunov et al. 1995, Nishizawa et al. 1995, Bjorklund et al. 1996, Kurath et al. 1997). This relationship

was most easily seen in dendrograms generated by CLUSTAL analysis in which the N, M1 and M2 amino acid sequences of HIRRV were separately aligned with the homologous sequences of 2 strains of IHNV (European and North American), 3 strains of VHSV (2 European and 1 North American), rabies virus (strain CVS), and vesicular stomatitis virus (serotype New Jersey). The dendrograms showed that HIRRV was more closely related to IHNV than to VHSV and that these fish viruses are quite distinct from the prototype members of the lyssavirus and vesiculovirus genera (Fig. 3) in agreement with results of Morzunov et al. (1995) and Bjorklund et al. (1996). The nucleoprotein and both matrix proteins gave very similar dendrograms suggesting the evolution rate of the 3 genes is similar, although the M2 protein showed slightly less divergence (Table 2). These data also provide evidence that HIRRV is a distinct virus, rather than a strain of IHNV, because the amino acid sequence **dif**ferences between the matrix proteins of HIRRV and their homologs in IHNV or VHSV were significantly greater than differences among even the most diverse strains of the North American and European isolates of VHSV (Basurco et al. 1995) or the North American and European isolates of IHNV (Nichol et al. 1995).

LITERATURE CITED

- Banerjee AK, Barik S (1992) Gene expression of vesicular stomatitis virus genome RNA. Virology 188:417-428
- Basurco B, Benmansour A (1995) Distant strains of the fish rhabdovirus VHSV maintain a sixth functional cistron which codes for a nonstructural protein of unknown function. Virology 212:741-745
- Basurco B, Vende P, Monnier AF, Winton JR, de Kinkelin P, Benmansour A (1995) Genetic diversity and phylogenetic classification of viral hemorrhagic septicemia virus (VHSV). Vet Res 26:460-463
- 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, Bremont M, Winton J (1992) Nucleocapsid gene sequence of a North-American isolate of viral hemorrhagic septicemia virus, a fish rhabdovirus. J Gen Virol 73:lOll-1014
- Bernard J, Lecoq-Xhonneux F, Rossum M, Thiry ME, De Kinkelin P (1990) Cloning and sequencing the messenger RNA of the N gene of viral hemorrhagic septicemia virus. J Gen Virol 71:1669-1674
- Bjorklund HV, Higman KH, Kurath G (1996) The glycoprotein genes and gene junctions of the fish rhabdoviruses spring viremia of carp virus and hirame rhabdovirus: analysis of relationships with other rhabdoviruses. Virus Res 42:65-80
- Black B, Brewer G, Lyles DS (1994) Effect of vesicular stomatitis virus matrix protein on host-directed translation in vivo. J Virol 68:555-560
- Black **B,** Lyles DS (1992) Vesicular stomatitis virus matrix protein inhibits host cell-directed transcription of target genes in viva. J Vir0166:4058-4064
- Black B, Rhodes RB, McKenzie M, Lyles DS (1993) The role of vesicular stomatitis virus matrix protein in inhibition of host-directed gene expression is genetically separable from its function in virus assembly. J Virol 67:4814-4821
- Bourhy H, Kissi B, Tordo N (1993) Molecular diversity of the lyssavirus genus. Virology 194:70-81
- Gilmore RD, Leong JC (1988) The nucleocapsid gene of infectious hematopoietic necrosis virus, a fish rhabdovirus. Vir010gy 167:644-648
- Kimura T, Yoshimizu M, Gorie S (1986) A new rhabdovirus isolated in Japan from cultured hirame (Japanese flounder, Paralichthys olivaceus) and ayu (Plecoglossis altivelis). Dis Aquat Org 1:209-217
- Koener JF, Passavant CW, Kurath G, Leong J (1987) Nucleotide sequence of a cDNA clone carrying the glycoprotein gene of infectious hematopoietic necrosis virus, a fish rhabdovirus. J Virol 61:1342-1349
- Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation in eukaryotic ribosomes. Cell 44:283-292
- Kurath G, Higman KH, Bjorklund **HV** (1997) Distribution and variation of **NV** genes in fish rhabdoviruses. J Gen Virol 78:113-117
- Kurath G, Leong J (1985) Characterization of infectious hematopoietic necrosis virus mRNA species reveals a nonvirion rhabdovirus protein. J Virol 53:462-468
- Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685
- LaPatra SE, Turner T, Lauda **KA,** Jones GR, Walker S (1993) Characterization of the humoral response of rainbow trout to infectious hematopoietic necrosis virus. J Aquat Anim Health 5:165-171
- Larson JK, Wunner WH, Ertl HCJ (1992) Immune response to the nominal phosphoprotein of rabies virus. Virus Res 23:73-88
- Leong JC, Hsu YL, Engelking HM, Mulcahy D (1981) Strains of infectious hematopoietic necrosis (IHNV) virus may be identified by structural protein differences. Dev Biol Stand 49:43-55
- MacAllister PE, Wagner RP (1975) Structural proteins of two salmonid rhabdoviruses. J Virol 4:733-738
- Morzunov SP, Winton JR, Nichol ST (1995) The complete genome structure and phylogenetic relationship of infectious hematopoietic necrosis virus. Virus Res 38:175-192
- Mourich DV, Leong JC (1992) Mapping the immunogenic regions of the IHNV glycoprotein in rainbow trout and mice. In: Proceedings of the Second International Symposium on Viruses of Lower Vertebrates. Oregon State University, Corvallis, p 93-100
- Nichol ST, Rowe JE, Winton JR (1995) Molecular epizootiology and evolution of the glycoprotein and nonvirion protein genes of infectious hematopoietic necrosis virus, a fish rhabdovirus. Virus Res 38:159-173
- Nishizawa **T,** Kurath G, Witon JR (1995) Nucleotide sequence of the 2 matrix protein genes (M1 and M2) of hirame rhabdovirus (HRV), a fish rhabdovirus. Vet Res 26: 408-412
- Nishizawa **T,** Yoshimizu M, Winton J, Ahne W, Kirnura T (1991) Characterization of structural proteins of hirame rhabdovirus, HRV. Dis Aquat Org 10:167-172
- Ogden JR, Wagner RR (1986) Mapping regions of the matrix protein of vesicular stomatitis virus which bind to ribonucleocapsids, liposomes, and monoclonal antibodies. J Virol58:860-868
- Ristow SS, deAvila J, LaPatra SE, Lauda K (1993) Detection and characterization of rainbow trout antibody against

infectious hematopoietic necrosis virus. Dis Aquat Org 15:109-114

- Sano T, Fukuda H (1987) Principal microbial diseases of mariculture in Japan. Aquaculture 6?:59-70
- Schutze H, Enzmann PJ, Kuchling R, Mundt E, Niemann H, Mettenleiter TC (1995) Complete genomic sequence of the fish rhabdovirus infectious haematopoietic necrosis virus. J Gen Virol 76:2519-2527
- Schutze H, Enzmann PJ, Mundt E, Mettenleiter TC (1996) Identification of the non-virion protein of fish rhabdoviruses viral haemorrhagic septicemia virus and infectious haematopoietic necrosis virus. J Gen Virol ?7:1259-1263
- Shipley JB, Pal R, Wagner RR (1988) Antigenicity, function, and conformation of synthetic oligopeptides corresponding to amino-terminal sequence of wild-type and mutant matrix proteins of vesicular stomatitis virus. J Virol 62: 2569-2577
- Spiropoulou CF, Nichol ST (1993) A small highly basic protein is encoded in overlapping frame within the P gene of vesicular stomatitis virus. J Virol 67:3103-3110
- Thiry M, Lecoq-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 hemorrhagic septicemia virus, a fish rhabdovirus. Bio-

Editorial responsibility: Jo-Ann Leong, Corvallis, Oregon, USA

chem Biophys Acta 1090:345-347

- Towbin H, Staehelin T, Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Nat Acad Sci USA 76:4350-4354
- Wagner RR (1987) The Rhabdoviruses. Plenum, New York, p 99-128
- Winton JR (1992) Evolution of fish rhabdoviruses. In: Kimura T (ed) Proceedings of the OJI International Symposium on Salmonid Diseases. Hokkaido University, Sapporo, p 88-95
- Wunner WH, Calisher CH, Dietzgen RG. Jackson AO, Kitajima EW, Lafon M, Leong JC, Nichol ST, Peters D, Smith JS, Walker PJ (1995) Rhabdoviridae. In: Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD (eds) Virus taxonomy: the classification and nomenclature of viruses. Sixth Report of the International Committee on Taxonomy of Viruses. Springer-Verlag, Wien, p 275-288
- Wunner WH, Peters D (1991) Rhabdoviridae. In: Francki RIB, Fauquet CM, Knudson DL, Brown F (eds) Classification and nomenclature of viruses. Fifth Report of the International Committee on Taxonomy of Viruses. Springer-Verlag, New York, p 250-262

Manuscript received: May **1, 1997;** Accepted: July **25, 1997** Proofs received from author(s): October 13, 1997