Penaeid shrimp infectious myonecrosis virus (IMNV) was isolated at the University of Arizona Aquaculture Pathology Laboratory from farmed *Penaeus vannamei* shrimp received from growers in north-eastern Brazil (Lightner et al., 2004; Poulos et al., 2006). As the virus name implies, the disease in these shrimp was characterized by necrosis of skeletal muscle, especially in the distal abdominal segments and tail fan. This disease was reproduced in the laboratory by injection of purified IMNV virions into pathogen-free *P. vannamei* (Poulos et al., 2006).

Many basic features of IMNV were reported along with its genome sequence by Poulos et al. (2006) (GenBank accession no. AY570982). IMNV virions are icosahedral and approximately 40 nm in diameter. They contain one major capsid protein (MCP), which has a relative mass of 106 000 and an unblocked N terminus of sequence IVSMENQSEID. The genome is a single molecule of double-stranded RNA (dsRNA), 7560 bp long. It contains two extended open reading frames (ORF) in different frames of the genomic plus strand: ORF1 in frame 1 (nt 136–4953) and ORF2 in frame 3 (nt 5241–7451). ORF1 (1605 aa) encodes a 179 kDa protein that includes the N-terminal sequence of the MCP, starting at Ile705. The protein spanning aa 705–1605 of the ORF1 product has a sequence-predicted mass of 99 kDa, consistent with the relative mass of the MCP, which thus appears to be cleaved from a larger ORF1 precursor. A 60 aa region at the N terminus of the full-length ORF1 product shares sequence similarities with dsRNA-binding proteins. ORF2 (736 aa) encodes an 85 kDa protein that contains a series of motifs characteristic of an RNA-dependent RNA polymerase (RdRp). Proteins representing ORF2 and the first 704 aa (80 kDa) of ORF1 have yet to be identified, although minor protein bands reported to have relative masses of 149 000, 42 000 and 24 000 were seen in overloaded gel lanes of IMNV virions and may reflect additional virion components. Phylogenetic analysis based on the RdRp region links IMNV to members of the family Totiviridae of monosegmented dsRNA viruses with icosahedral virions and most closely to *Giardia lamblia virus* (GLV, GenBank accession no. L13218).

To date, IMNV is the only member of the Totiviridae (assignment still tentative) to infect a host other than a fungus or protozoan and therefore other distinctive features might be expected. Given the close phylogenetic relationship between the RdRp regions of IMNV and GLV, however, the apparent lack of ORF2 translation as a fusion with ORF1 in IMNV seems somewhat curious. I therefore examined the published genome sequence of IMNV for further clues about its coding strategies.

Examination of aa 1–704 in IMNV ORF1 revealed two conserved nonapeptides, GDVESNPGP and GDVEENPGP, at aa 86–94 and 370–378, respectively. A BLASTP search and review of the literature showed that these sequences are identical to the 2A or '2A-like' motifs in the polyproteins of Foot-and-mouth disease virus (GDVESNPGP), Cricket paralysis virus (GDVESNPGP), *Thosea asigna* virus (GDVEENPGP) and others (Donnelly et al., 2001;...
Palmenberg et al., 1992). The preceding list includes plus-strand RNA viruses from the families Picornaviridae, Dicistroviridae and Tetraviridae. In these other viral proteins, a cotranslational event, either autocleavage between Gly and Pro or failure of the Gly–Pro peptide bond to be formed, results in separation of the first polypeptide ending in Gly from the second polypeptide starting with Pro. The presence of two such motifs in the N-terminal third of IMNV ORF1 preceding the MCP suggests that the ORF1 product is a polyprotein that is cotranslationally ‘cleaved’ into consecutive 93, 284 and 1228 aa fragments. The 1228 aa fragment would be further cleaved either co- or post-translationally, and by a different mechanism, to yield a 327 aa fragment and the 901 aa MCP that begins with Ile705 as described above.

The functions of the 93 aa (10 kDa), 284 aa (32 kDa) and 327 aa (38 kDa) fragments of IMNV ORF1 are unknown, although the 93 aa fragment encompasses the N-terminal region previously noted to share sequence similarities with dsRNA-binding proteins (Poulos et al., 2006). Interestingly, 2A-like sequences are also found in the NS34 protein of group C rotaviruses (family Reoviridae, genus Rotavirus, of 11-segmented dsRNA viruses) and allow a C-terminal 69 aa fragment of NS34, also with similarities to dsRNA-binding proteins, to be released from the protein (Donnelly et al., 2001). Thus, at least two distinct dsRNA viruses, IMNV and group C rotaviruses, appear to use 2A-like sequences to release small N- or C-terminal dsRNA-binding fragments, which might be involved in either virus replication or modulation of the host’s innate immune response.

Poulos et al. (2006) reported that IMNV ORF1 and ORF2, in frames 1 and 3, respectively, are non-overlapping. Though true that the first standard start codon (AUG, nt 5241–5243) in ORF2 begins 288 nt downstream of the ORF1 stop codon (UAA, nt 4951–4953), there are no in-frame stop codons upstream of this ORF2 start codon until the UGA at nt 4749–4751 (Fig. 1). Thus, ORF2, when defined as the region devoid of stop codons, begins 199 nt upstream of the end of ORF1, meaning that ORF1 and ORF2 are overlapping. This creates the possibility that ORF2 could be translated as a fusion with ORF1 if ribosomal −1 frameshifting was to occur within the region of overlap.

Ribosomal −1 frameshifting is often associated with a ‘shifty (or slippery) heptamer’ motif, XXXYYYYZ, where X is A, C, G or U, Y is A or U, and Z is A, C or U (Bekaert et al., 2003; Jacks et al., 1988). An RNA pseudoknot shortly downstream of the heptamer, helping to pause the translating ribosome, commonly increases the frequency of frameshifting (Brierley et al., 1989; Giedroc et al., 2000; Plant et al., 2003). For IMNV ORF2 to be translated as a fusion with ORF1, −1 frameshifting would need to occur within the region of ORF1–ORF2 overlap, nt 4752–4950. In fact, the sequence GGGUUUU is found at nt 4803–4809 (Fig. 1), which qualifies as a shifty heptamer. Immediately preceding this sequence in IMNV is the dinucleotide UC (Fig. 1), which may favour −1 frameshifting (Bekaert & Roussel, 2005). Moreover, using the web-based program Hpknotted (Huang et al., 2005) with the pknotsRG (version 1.2) kernel (Reeder & Giegerich, 2004), a class-1 H-type pseudoknot with a minimal free energy of −10.1 kcal mol$^{-1}$ (42.3 kJ mol$^{-1}$) was predicted for nt 4818–4862, shortly after the shifty heptamer (Fig. 1). Assuming that the ORF1–ORF2 fusion is cleaved after Asn704 to generate the N terminus of the MCP region at Ile705, as for the non-fused MCP (Poulos et al., 2006), the size of the fused MCP–RdRp is predicted to be 1734 aa or 196 kDa.
Phylogenetic comparisons involving the RdRp region sequences of IMNV have identified GLV as its closest relative (Crawford et al., 2006; Poulos et al., 2006). GLV is a member of the family Totiviridae and prototype of the genus Giardiaivirus. Its 6277 nt genome includes two ORFs, which are in different frames and overlap by 220 nt (Wang et al., 1993). A ribosomal −1 frameshift in the region of overlap allows ORF2 to be translated as a fusion with ORF1 (Li et al., 2001; Wang et al., 1993). The new findings for IMNV thus bring its ORF2 coding strategy more in line with that of GLV. Notably, some other members of the Totiviridae also use −1 frameshifting to translate ORF2 as a fusion with ORF1, including the well-studied Saccharomyces cerevisiae virus L-A (Dinman et al., 1991).

The newly identified coding features of IMNV are summarized in Fig. 2 and provide clear predictions for future work. The MCP–RdRp fusion would be expected to assemble in only a few copies per virion, with the MCP region occupying a position in the putatively ‘T=2’ capsid and the RdRp region projecting into the virion interior, where it could function in viral transcription as described for S. cerevisiae virus L-A (Ribas & Wickner, 1998; Naitow et al., 2002). Some of the predicted N-terminal fragments of the IMNV ORF1 polyprotein may also be assembled into virions. Although the RdRp region sequences of IMNV are most closely related to those of GLV, IMNV probably deserves to typify a new genus because of its very different host (crustacean arthropod vs flagellated protozoan) and the distinctive features of its ORF1 polyprotein.

A final finding that arose during these analyses is that 2A-like sequences are also found in the genome sequences of certain members of the family Reoviridae, genus Cypovirus, of ten-segmented dsRNA viruses that infect insects and assembly virions occluded in cytoplasmic polyhedra. Specifically, 2A-like sequences are found in the segment 5 translation products (879–881 aa) of cypovirus 1 isolates from Bombyx mori (GDIESNPGP, aa 227–235; GenBank accession no. AB035733), Dendrolimus punctatus (GDVESNP GP, aa 227–235; accession no. AY163248) and Lymnantria dispar (GDVESNP GP, aa 227–235; accession no. AF389466), as well as the cypovirus 18 isolate from Operophthera brumata (GDVESNPGP, aa 228–236; accession no. DQ192245). Thus, 2A-like ‘cleavage’ sequences appear to have been incorporated into the coding strategies of a number of different dsRNA viruses.

**Acknowledgements**

I thank Said Ghabrial for helpful discussions and Tim Baker, S. G., Don Lightner, Bonnie Poulos and Kathy Tang for helpful comments on the manuscript. I also thank the anonymous reviewer who kindly suggested that there is indeed a properly positioned, predicted RNA pseudoknot in IMNV.

**References**


