Three functionally diverged major structural proteins of white spot syndrome virus evolved by gene duplication

Mariëlle C. W. van Hulten, Rob W. Goldbach and Just M. Vlak

Laboratory of Virology, Wageningen University, Binnenhaven 11, 6709 PD Wageningen, The Netherlands

White spot syndrome virus (WSSV) is an invertebrate virus causing considerable mortality in penaeid shrimp. The oval-to-bacilliform shaped virions, isolated from infected *Penaeus monodon*, contain four major proteins: VP28, VP26, VP24 and VP19 (28, 26, 24 and 19 kDa, respectively). VP26 and VP24 are associated with the nucleocapsid and the remaining two with the envelope. Forty-one N-terminal amino acids of VP24 were determined biochemically allowing the identification of its gene (*vp24*) in the WSSV genome. Computer-assisted analysis revealed a striking similarity between WSSV VP24, VP26 and VP28 at the amino acid and nucleotide sequence level. This strongly suggests that these structural protein genes may have evolved by gene duplication and subsequently diverged into proteins with different functions in the WSSV virion, i.e. envelope and nucleocapsid. None of these three structural WSSV proteins showed homology to proteins of other viruses including baculoviruses, underscoring the distinct taxonomic position of WSSV among invertebrate viruses.

White spot syndrome virus (WSSV) is a rapidly emerging viral disease agent in shrimp in Southeast Asia and the Americas. The virus has a wide host range among crustaceans (Flegel, 1997) and induces distinctive clinical signs (white spots) under the carapace of penaeid shrimps. WSSV has a double-stranded DNA genome with a size exceeding 250 kbp (Yang *et al.*, 1997; Anon., 1999) and may be a representative of a new floating genus, *Whispovirus* (Van Hulten *et al.*, 2000*a*, *b*; Van Hulten & Vlak, 2000). Electron microscopy studies revealed that WSSV virions are enveloped rod-shaped nucleocapsids with a bacilliform to ovoid shape about 275 nm in

length and 120 nm in width. Most characteristic is the tail-like appendage at one end of the virion (Wongteerasupaya *et al.*, 1995; Durand *et al.*, 1997). WSSV nucleocapsids have a striated appearance and a size of about 300×70 nm (Wongteerasupaya *et al.*, 1995). The striations are probably the result of stacked ring-like structures consisting of rows of globular subunits about 10 nm in diameter (Durand *et al.*, 1997; Nadala *et al.*, 1998).

The characterization of the structural proteins and their genomic sequence is of major importance to determine the taxonomic position of the virus. Furthermore, the structure and interaction of the WSSV virion proteins may explain the unique morphological features of this virus. Finally, diagnostic tests could be designed based on one or more of these structural proteins. VP28 and VP26, present in the envelope and nucleocapsid, respectively, were identified previously and showed no homology with sequences available in GenBank (Van Hulten *et al.*, 2000*b*). Here we report the identification of a third major structural protein, VP24, and the surprising relatedness of this protein to the previously identified WSSV structural proteins.

Purified WSSV was used to infect shrimp, *Penaeus monodon*, by intramuscular injections in the lateral area of the fourth abdominal segment. Virions were purified from haemolymph of infected *P. monodon* as described by Van Hulten *et al.* (2000*b*). As a negative control, haemolymph was taken from uninfected *P. monodon*. The preparations were analysed by electron microscopy for the presence and purity of WSSV virions (not shown). The viral envelope was removed from the nucleocapsid by treatment with 1% NP-40 (Van Hulten *et al.*, 2000*b*). In the intact WSSV virions purified from *P. monodon* (Fig. 1*a*, lane 3), four major polypeptide species were identified with apparent molecular masses of 28 kDa (VP28), 26 kDa (VP26), 24 kDa (VP24) and 19 kDa (VP19). From the SDS–PAGE analysis (Fig. 1*a*) it can be seen that VP26 and VP24 are the major proteins present in the purified nucleocapsids (Fig. 1*a*, lane 4). VP28 and VP19 are removed by the NP-40 treatment and therefore associated with the viral envelope or tegument (Van Hulten *et al.*, 2000*b*). A schematic presentation of the WSSV virion is shown in Fig. 1(*b*).

The sizes found for the major virion proteins were similar to those described by Hameed *et al.* (1998) and Nadala *et al.*

Author for correspondence: Just Vlak.

Fax +31 317 484820. e-mail just.vlak@medew.viro.wau.nl
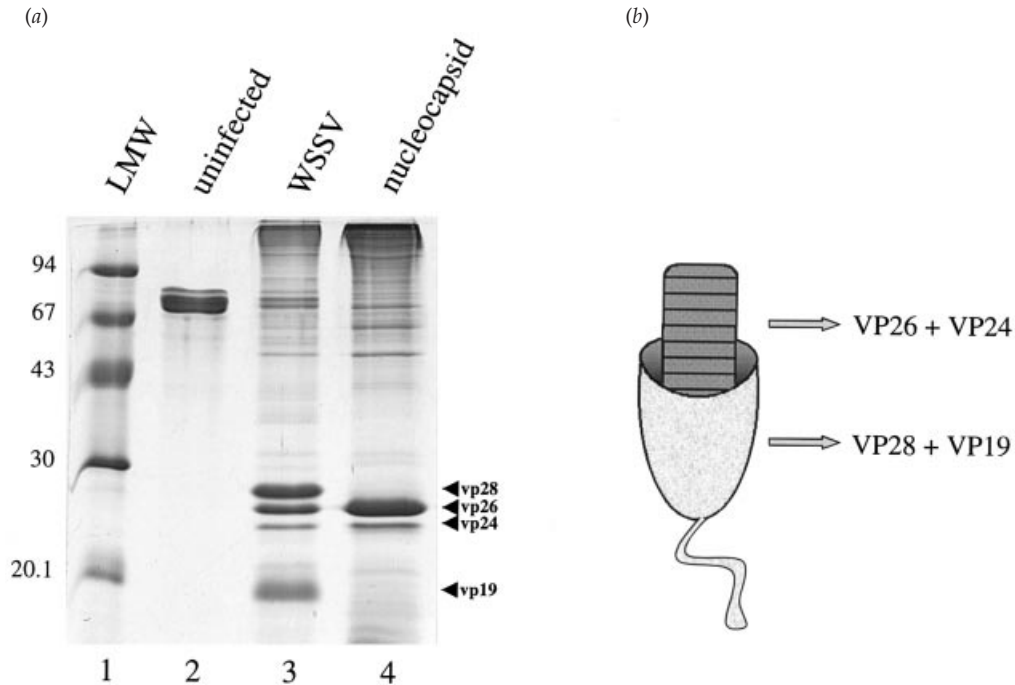
(a)

(b)

Fig. 1. (a) Coomassie brilliant blue-stained SDS–15% PAGE gel of purified WSSV. Lane 1, low molecular mass protein marker. Lane 2, mock purification. Lane 3, purified WSSV particles. Lane 4, purified WSSV nucleocapsids. (b) Schematic representation of the WSSV virion.

(1998), but somewhat different from those described by Wang *et al.* (2000). The latter authors described the presence of three major proteins in WSSV isolates from different origins with slightly different sizes of 25, 23 and 19 kDa, respectively. The N-terminal sequences of these proteins demonstrate, however, that the 25 and 23 kDa proteins correspond to our VP28 and VP26, respectively. In our WSSV isolate a protein of 24 kDa (VP24) is clearly a major component of the nucleocapsid. Here we describe the amino acid and genomic sequence of WSSV VP24 and some characteristics of this protein.

VP24 isolated from *P. monodon* was transferred from an SDS–PAGE gel onto a PVDF membrane by semi-dry blotting. The 24 kDa band, derived from two separate WSSV preparations, was excised from the membrane and each was sequenced by Edman degradation as described previously (Van Hulten *et al.*, 2000*b*). The first N-terminal sequence obtained was MHMWGVYAAILAGLTLILVVIdI, of which the aspartic acid at position 22 was uncertain. From the second VP24 band more than 40 residues were sequenced (bold font in Fig. 2) giving the sequence MHMWGVYAAILAGLTLILV VISIVV-TNIELNKKLDKKDKdA, in which a serine residue was found at position 22 and an uncertain aspartic acid at position 40.

Based on this partial VP24 sequence a set of degenerate PCR primers was developed, with 5′ CAG*AATTC*ATGC-AYATGTGGGGNGT 3′ as forward primer, and 5′ CAG-*AATTC*YTTRTCYTTYTTRTCIARYTT 3′ as reverse primer, both containing *Eco*RI sites (italics) for cloning purposes. The

location of the primers in the final sequence is indicated in Fig. 2. PCR was performed using WSSV genomic DNA as template. A 133 bp fragment was obtained and, after purification from a 2% agarose gel, cloned into pBluescript SK(+) and sequenced. The sequence of this PCR product corresponded with the N-terminal protein sequence of WSSV VP24 and was used as probe in a colony lift assay (Sambrook *et al.*, 1989) on WSSV plasmid libraries (Van Hulten *et al.*, 2000*a*) to identify the complete ORF for VP24. An 18 kbp *Bam*HI fragment hybridizing with this fragment was selected for further analysis.

The complete *vp*24 ORF, encompassing 627 nucleotides, and the promoter region of this gene, were found on the 18 kbp *Bam*HI fragment (Fig. 2). The translational start codon was in a favourable context (AAA*ATG*C) for efficient eukaryotic translation initiation (Kozak, 1989). In the promoter region stretches of A/T-rich sequence, but no consensus TATA box, were found. A poly(A) signal overlapped the translation stop codon. The *vp*24 ORF encoded a putative protein of 208 amino acids with an amino acid sequence containing the experimentally determined N-terminal sequence of VP24. VP24 has a theoretical size of 23 kDa and an isoelectric point of 8·7. Four potential sites for *N*-linked glycosylation (N-{P}-[ST]-{P}), one site for *O*-glycosylation (Hansen *et al.*, 1998) (Fig. 2) and nine possible phosphorylation sites ([ST]-X-X-[DE] or [ST-X-[RK]) were found within VP24, but it is not known whether any of these modifications do occur. No other motifs present in the PROSITE database were

```
AAAGGCTATCCAGTACTTCAACAAAACTAGCAGAAATAATACTGCACATCATTTCAAGATGCCGGCTTCA

AAGGATCGCAACTATTCTTCTTTCGAGTATGCTGAAACGGCAGTTGCTGCTCACAATATATCTCAGTGGT

GTTTTCTACGACTAAATAAAATATAACAAACTTTAATATCTGTTTTTTTCTCTCATGACCTTTGTACAAC

TTTCTTCAAAATGCACATGTGGGGGGTTTACGCCGCTATACTGGCGGGTTTGACATTGATACTCGTGGTT
          M  H  M  W  G  V  Y  A  A  I  L  A  G  L  T  L  I  L  V  V

ATATCTATAGTTGTAACCAACATAGAACTTAACAAGAAATTGGACAAGAAGGATAAAGACGCCTACCCTG
I  S  I  V  V  T  N  I  E  L  N  K  K  L  D  K  K  D  K  D  A  Y  P  V

TTGAATCTGAAATAATAAACTTGACCATTAACGGTGTTGCTAGAGGAAACCACTTTAACTTTGTAAACGG
 E  S  E  I  I  N̲  L  T  I  N  G  V  A  R  G  N  H  F  N̲  F  V  N  G

CACATTACAAACCAGGAACTATGGAAAGGTATATGTAGCTGGCCAAGGAACGTCCGATTCTGAACTGGTA
 T  L  Q  T  R  N  Y  G  K  V  Y  V  A  G  Q  G  T  S  D  S  E  L  V

AAAAAGAAAGGAGACATAATCCTCACATCTTTACTTGGAGACGGAGACCACACACTAAATGTAAACAAAG
 K  K  K  G  D  I  I  L  T  S  L  L  G  D  G  D  H  T  L  N  V  N  K  A

CCGAATCTAAAGAATTAGAATTGTATGCAAGAGTATACAATAATACAAAGAGGGATATAACAGTGGACTC
 E  S  K  E  L  E  L  Y  A  R  V  Y  N̲  N  T  K  R  D  I  T  V  D  S

TGTTTCACTGTCTCCAGGTCTAAATGCTACAGGAAGGGAATTTTCAGCTAACAAATTTGTATTATATTTC
 V  S  L  S̲  P  G  L  N̲  A  T  G  R  E  F  S  A  N  K  F  V  L  Y  F

AAACCAACAGTTTTGAAGAAAAATAGGATCAACACACTTGTGTTTGGAGCAACGTTTGACGAAGACATCG
 K  P  T  V  L  K  K  N  R  I  N  T  L  V  F  G  A  T  F  D  E  D  I  D

ATGATACAAATAGGCATTATCTGTTAAGTATGCGATTTTCTCCTGGCAATGATCTGTTTAAGGTTGGGGA
 D  T  N  R  H  Y  L  L  S  M  R  F  S  P  G  N  D  L  F  K  V  G  E

AAAATAAACACATTTTTTAAAAATATACCTTTATTTTATTAACCAATTATTTACATTGTAGCTTGAAATT
 K  *

CATATATATACCTTTCATTGTTAGGGAGAGATTTATTGTAGTCAAACCCAG
```

Fig. 2. Nucleotide and protein sequence of WSSV VP24. The sequenced N-terminal amino acids are in bold; the locations of putative *N*-glycosylation sites are underlined and an *O*-glycosylation site is double underlined. The nucleotide sequence of degenerate primer positions is in bold and italics.

found in VP24. Computer analysis of the 208 amino acids showed that a strong hydrophobic region was present at the N terminus of VP24 (Fig. 3*a*), including a putative transmembrane α-helix formed by amino acids 6 through 25. The algorithm of Garnier *et al.* (1978) predicted several other α-helices and β-sheets along the protein. It is remarkable that VP28, VP26 and VP24 roughly have the same size ($\sim$ 206 amino acids) but have distinct electrophoretic mobilities. This may be due to differences in isoelectric points, conformational differences or post-translational modifications.

Homology searches with WSSV VP24 were performed against GenBank/EMBL, SWISS-PROT and PIR databases using FASTA, TFASTA and BLAST, but no significant homology with structural proteins from other large DNA viruses could be found. Surprisingly, statistically significant similarity was found with the sequence of two other WSSV virion structural proteins, VP26 and VP28 (Van Hulten *et al.*, 2000*b*), with 41 and 46% amino acid similarity, respectively. Also, VP28 and VP26 showed a similarity of 41%. An alignment of the three WSSV proteins was made using ClustalW (Thompson *et al.*, 1994), and revealed several conserved regions (Fig. 3*b*). In the N-terminal region a well-conserved stretch of amino acids is observed at positions 15–30. A strong hydrophobic region with an α-helix is observed for all three proteins in the hydrophilicity plots (Fig. 3*a*). These residues might represent a transmembrane region, or be involved in the interaction of the structural proteins to form homo- or heteromultimers. In two other conserved regions, around positions 88–102 and positions 138–148, the algorithm of Garnier *et al.* (1978) predicted an α-helix and a β-sheet for all three proteins (Fig. 3*a*, *b*). Two-thirds of the residues conserved in the three proteins were hydrophobic and might be involved in the folding of the proteins, giving them a similar structure. The nucleocapsid and envelope proteins differed in their isoelectric points. The two nucleocapsid proteins (VP26 and VP24) both had a basic character with isoelectric points of 9·3 and 8·7, respectively, and might therefore have a close association with the viral DNA, whereas the envelope protein (VP28) was more acidic with an isoelectric point of 4·6.
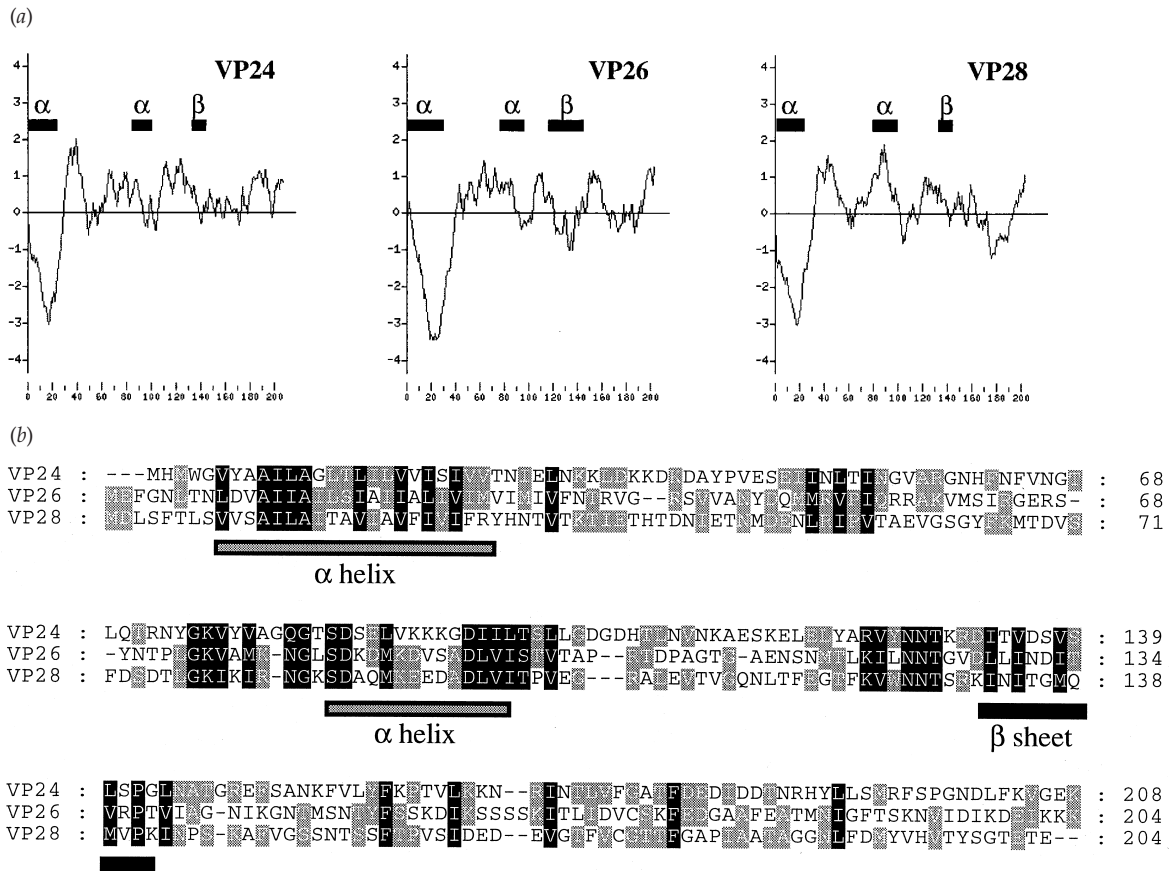
Fig. 3. (a) Hydrophilicity plots of WSSV VP24, VP26 and VP28. The amino acid number is on the abscissa, and the hydrophilicity value on the ordinate. α-Helices and β-sheets are indicated. (b) Amino acid sequence alignment of VP24, VP26 and VP28. Shading is used to indicate the occurrence (black 100%, grey 67%) of similar amino acids. Conserved α-helices and β-sheets are indicated.

As there is a high homology at the amino acid level among the three structural WSSV proteins, and conserved domains are present, there is reason to believe that their structures are similar. The presence of the hydrophobic domain indicates that these proteins most probably are capable of forming homo- and heteromultimers. Studies on the interaction of these proteins and their location in the virion are required to substantiate this hypothesis.

A way to explain the high degree of amino acid similarity of the three structural WSSV proteins is to assume that these genes have evolved by gene duplication and divergence. Nucleotide comparisons supported this hypothesis, as significant homology was found. Alignment of *vp24*, *vp26* and *vp28*, revealed that *vp24* has 40% nucleotide identity with *vp26* and 43% with *vp28*, whereas *vp26* has 48% nucleotide identity with *vp28*. The data presented here strongly suggest that these three WSSV structural protein genes share a common ancestor.

The most surprising observation might be that these proteins have evolved to give proteins with different functions in the WSSV virion, i.e. in the nucleocapsid and the envelope.

Such a situation is unusual in animal DNA viruses, although a parallel may exist for the virion glycoproteins of alpha-herpesviruses as their genes might have evolved by duplication and divergence (McGeoch, 1990). However, the homology of these genes is considerably lower than the homology among the WSSV virion genes. Also, the function of the alpha-herpesvirus genes has not diverged. Gene duplication and functional divergence, however, can be observed in the plant-infecting closteroviruses, where a minor regulatory protein, VP24, appears to be a diverged copy of the coat protein (Boyko *et al.*, 1992). Also, in the animal rhabdoviruses, a structural and a non-structural glycoprotein may have evolved from a common ancestral gene (Wang & Walker, 1993).

Structural proteins are well conserved within virus families. However, the three WSSV structural proteins identified so far have no homology to structural proteins of other viruses. The unique feature of the homologous structural virion proteins further supports the proposition that WSSV might be a representative of a new virus genus (*Whispovirus*) or perhaps a new family (*Whispoviridae*) (Van Hulten *et al.*, 2000*a*, *b*; Van Hulten & Vlak, 2000).

References

Anon. (1999). Genome project aims to combat prawn scourge. *Nature* **397**, 465.

Boyko, V. P., Karasev, A. V., Agranovsky, A. A., Koonin, E. V. & Dolja, V. V. (1992). Coat protein gene duplication in a filamentous RNA virus of plants. *Proceedings of the National Academy of Sciences, USA* **89**, 9156–9160.

Durand, S., Lightner, D. V., Redman, R. M. & Bonami, J. R. (1997). Ultrastructure and morphogenesis of white spot syndrome baculovirus (WSSV). *Diseases of Aquatic Organisms* **29**, 205–211.

Flegel, T. W. (1997). Major viral diseases of the black tiger prawn (*Penaeus monodon*) in Thailand. *World Journal of Microbiology and Biotechnology* **13**, 433–442.

Garnier, J., Osguthorpe, D. J. & Robson, B. (1978). Analysis of the accuracy and implications of simple method for predicting the secondary structure of globular proteins. *Journal of Molecular Biology* **120**, 97–120.

Hameed, A. S. S., Anilkumar, M., Raj, M. L. S. & Jayaraman, K. (1998). Studies on the pathogenicity of systemic ectodermal and mesodermal baculovirus and its detection in shrimp by immunological methods. *Aquaculture* **160**, 31–45.

Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L. & Brunak, S. (1998). NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconjugate Journal* **15**, 115–130.

Kozak, M. (1989). The scanning model for translation: an update. *Journal of Cell Biology* **108**, 229–241.

McGeoch, D. J. (1990). Evolutionary relationships of virion glycoprotein genes in the S regions of alphaherpesvirus genomes. *Journal of General Virology* **71**, 2361–2368.

Nadala, E. C. B., Tapay, L. M. & Loh, P. C. (1998). Characterization of a non-occluded baculovirus-like agent pathogenic to penaeid shrimp. *Diseases of Aquatic Organisms* **33**, 221–229.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual,* 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

Van Hulten, M. C. W. & Vlak, J. M. (2000). Genetic evidence for a unique taxonomic position of white spot syndrome virus of shrimp: genus Whispovirus. *Proceedings of the Fourth Symposium on Diseases in Asian Aquaculture*. Edited by C. Lavilla-Pitogo and others (in press).

Van Hulten, M. C. W., Tsai, M. F., Schipper, C. A., Lo, C. F., Kou, G. H. & Vlak, J. M. (2000 a). Analysis of a genomic segment of white spot syndrome virus of shrimp containing ribonucleotide reductase genes and repeat regions. *Journal of General Virology* **81**, 307–316.

Van Hulten, M. C. W., Westenberg, M., Goodall, S. D. & Vlak, J. M. (2000 b). Identification of two major virion protein genes of white spot syndrome virus of shrimp. *Virology* **266**, 227–236.

Wang, Y. & Walker, P. J. (1993). Adelaide River rhabdovirus expresses consecutive glycoprotein genes as polycistronic mRNAs: new evidence of gene duplication as an evolutionary process. *Virology* **195**, 719–731.

Wang, Q., Poulos, B. T. & Lightner, D. V. (2000). Protein analysis of geographic isolates of shrimp white spot syndrome virus. *Archives of Virology* **145**, 263–274.

Wongteerasupaya, C., Vickers, J. E., Sriurairatana, S., Nash, G. L., Akarajamorn, A., Boonsaeng, V., Panyim, S., Tassanakajon, A., Withyachumnarnkul, B. & Flegel, T. W. (1995). A non-occluded, systemic baculovirus that occurs in cells of ectodermal and mesodermal origin and causes high mortality in the black tiger prawn *Penaeus monodon. Diseases of Aquatic Organisms* **21**, 69–77.

Yang, F., Wang, W., Chen, R. Z. & Xu, X. (1997). A simple and efficient method for purification of prawn baculovirus DNA. *Journal of Virological Methods* **67**, 1–4.