

Proteomic Analysis of Shrimp White Spot Syndrome Viral Proteins and Characterization of a Novel Envelope Protein VP466*

Canhua Huang‡§, Xiaobo Zhang‡§, Qingsong Lin‡, Xun Xu¶, Zhihong Hu||, and Choy-L. Hew†**

White spot syndrome virus (WSSV) is at present one of the major pathogens in shrimp culture worldwide. The complete genome of this virus has been sequenced recently. To identify the structural and functional proteins of WSSV, the purified virions were separated by SDS-PAGE. Twenty-four protein bands were excised, in-gel digested with trypsin, and subjected to matrix-assisted laser desorption ionization-time of flight mass spectrometry and electrospray ionization tandem mass spectrometry, respectively. Eighteen proteins matching the open reading frames of WSSV genome were identified. Except for three known structural proteins and collagen, the functions of the remaining 14 proteins were unknown. Temporal analysis revealed that all the genes were transcribed in the late stage of WSSV infection except for vp121. Of the newly identified proteins, VP466 (derived from band 16) was further characterized. The cDNA encoding VP466 was expressed in *Escherichia coli* as a glutathione S-transferase (GST) fusion protein. Specific antibody was generated with the purified GST-VP466 fusion protein. Western blot showed that the mouse anti-GST-VP466 antibody bound specifically to a 51-kDa protein of WSSV. Immunogold labeling revealed that VP466 protein is a component of the viral envelope. Results in this investigation thus proved the effectiveness of proteomic approaches for discovering new proteins of WSSV. *Molecular & Cellular Proteomics* 1:223–231, 2002.

White spot syndrome virus (WSSV),¹ considered to be a new virus (1), is one of the major pathogens in cultured

From the ‡Department of Biological Sciences and Tropical Marine Science Institute, National University of Singapore, Singapore 119260, ¶Key Laboratory of Marine Biotechnology, Third Institute of Oceanography, State Oceanic Administration, Xiamen 361005, People's Republic of China, and ||Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan 430071, People's Republic of China

Received, December 14, 2001, and in revised form, February 5, 2002

Published, MCP Papers in Press, February 15, 2002, DOI 10.1074/mcp.M100035-MCP200

¹ The abbreviations used are: WSSV, white spot syndrome virus; ABC, ammonium bicarbonate; ESI-MS/MS, electrospray ionization tandem mass spectrometry; GST, glutathione S-transferase; p.i., post-infection; MALDI-TOF, matrix-assisted laser desorption ionization-time-of-flight; MS, mass spectrometry; ORF, open reading

penaeid shrimp. First appearing in the 1990s in Taiwan, WSSV has spread rapidly to shrimp-farming areas around the world causing large economic losses. The virus has a broad host range, including other invertebrate aquatic organisms, such as crab and crayfish (2, 3). The WSSV particles are non-occluded and bacilliform in shape with double envelopes (4). In 1997, the WSSV genomic DNA was purified successfully from *Penaeus japonicus* shrimp in China (5), and the genomic DNA and cDNA libraries were constructed. WSSV genome contains a 305-kb double-stranded circular DNA (6), which is longer than the 293-kb genome isolated from *Penaeus monodon* shrimps in Thailand (7). Approximately 180 open reading frames (ORFs) of 50 amino acids or more were revealed by the analysis of the WSSV genomic DNA sequences (6). However, in contrast to the insect baculovirus, one of the best studied viruses, only a few genes from WSSV have been reported (8).

With the completion of the WSSV genomic DNA sequence, research has now focused on the functional analysis of the gene products. Essential to the functional analysis is to identify the proteins expressed in WSSV. To this end, a proteomic approach using mass spectrometry has been proven to be the most effective technology for the identification of proteins (9). Recently, matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry (MS) and electrospray ionization tandem mass spectrometry (ESI-MS/MS) utilizing a quadrupole time-of-flight (Q-TOF) mass spectrometer have been used as tools for the characterization of proteins because of their high sensitivity and throughput (10).

In this communication, the WSSV proteins separated by SDS-PAGE were analyzed using MALDI-TOF MS and ESI-MS/MS (Q-TOF), respectively. The resulting mass spectrometric data were searched against the theoretical ORF database of WSSV. One of the newly retrieved genes, vp466 gene, was further characterized.

EXPERIMENTAL PROCEDURES

Proliferation and Purification of Shrimp WSSV

Proliferation and Purification of WSSV Virion—The infected tissues from penaeid shrimp *P. monodon* (e.g. gill, stomach, midgut, etc.)

frame; Q-TOF, quadrupole time-of-flight; RACE, rapid amplification of cDNA ends; RT, reverse transcription; PBS, phosphate-buffered saline.

were homogenized in TN buffer (20 mM Tris-HCl and 400 mM NaCl, pH 7.4) at 0.1 g/ml. After centrifugation at $2000 \times g$ for 10 min, the supernatant was filtered (0.22- μ m filter) and injected (1:100 dilution in 0.9% NaCl) intramuscularly into healthy crayfish *Cambarus clarkii* from Hainan province, China in the lateral area of the fourth abdominal segment. Four days later, hemolymph freshly extracted from infected crayfish was layered on the top of the 10–40% (w/v) continuous sodium bromide gradient and centrifuged at $110,000 \times g$ using a SW41-Ti rotor in a Beckman ultracentrifuge (XL-90; Beckman Coulter) for 2 h at 4 °C. Virus bands were collected by side puncture, diluted 1:10 with TNE buffer (50 mM Tris-HCl, 100 mM NaCl, and 1 mM EDTA, pH 7.4), and pelleted at $119,000 \times g$ for 1 h at 4 °C. The resulting pellets were resuspended in TNE. Virus samples were examined under a transmission electron microscope (JEOL 100 CXII) for purity (11).

Purification of WSSV Nucleocapsid—Purified WSSV virion was treated with Triton X-100 for 15 min at room temperature, subjected to 20–50% continuous CsCl gradient, and centrifuged for 24–48 h at $110,000 \times g$ using a SW 41-Ti. Viral capsid band was collected by side puncture and then diluted with $1 \times$ TN buffer (1:10), subsequently sedimented at $120,000 \times g$ for 45 min. The pellet was resuspended in $1 \times$ TN buffer, pH 7.4.

Computer Analysis of the ORFs of WSSV Genome

The 305,107-bp DNA sequence of the WSSV genome (6) was analyzed with DNAMAN (Lynnon BioSoft, Vaudreuil, Canada) to identify ORFs. In total, 4443 ORFs starting with an ATG start codon and with lengths of 50 amino acids or larger were located on both strands of the WSSV genome. From these ORFs, ~181 ORFs ranging from 61 to 6077 amino acids in size were likely to encode functional proteins (6). These 181 ORFs were designated as putative genes and assigned to the WSSV ORF database. Homology searches were performed with the BLAST and BLAST2 programs. Protein motifs were analyzed by PROSITE release 16 database (12).

Mass Spectrometric Analysis of Viral Proteins

In-gel Digestion—The proteins from purified WSSV were separated by 12% SDS-PAGE and stained with Coomassie Blue R 250. Protein bands were excised and dehydrated several times with acetonitrile. After vacuum drying, the gel bands were incubated with 10 mM dithiothreitol in 100 mM ammonium bicarbonate (ABC buffer) at 57 °C for 60 min and subsequently with 55 mM iodoacetamide (Sigma) in 100 mM ABC buffer at room temperature for 60 min. Then the gels were washed with 100 mM ABC buffer and dried. All in-gel protein digestions were performed using sequencing grade modified porcine trypsin (Promega, Madison, WI) in 50 mM ABC buffer at 37 °C for 15 h. Digests were centrifuged at $6000 \times g$. The supernatants were separated, and the gel pieces were extracted further first with 50% acetonitrile, 5% formic acid and then with acetonitrile. The extracts were combined with the original digesting supernatants, vacuum-dried, and redissolved in 0.5% trifluoroacetic acid and 50% acetonitrile (13).

MALDI-TOF MS—The matrix used was a saturated solution of α -cyano-4-hydroxycinnamic acid in 0.5% trifluoroacetic acid and 50% acetonitrile. The sample and the matrix (1:1, v/v) were loaded on the target plate. MALDI-TOF spectra of the peptides were obtained with a Voyager-DE STR Biospectrometry work station mass spectrometer (PerSeptive Biosystems, Inc., Framingham, MA). The analyses were performed in positive ion reflector mode with an accelerating voltage of 20 kV and a delayed extraction of 150 ns. Typically 180 scans were averaged. The trypsin autolysis products were used as internal calibrants. Data mining was performed using MS-Fit software against the WSSV ORF database. A mass deviation of 100 ppm was usually allowed in the database searches.

Nano-ESI-MS/MS—The in-gel digested samples were desalted using C18 ZipTip (Millipore, Bedford, MA) and dried. After dissolving in 2 μ l of 50% acetonitrile and 0.5% formic acid, the sample was loaded into a metallized glass capillary. The capillary was then mounted on the nanoflow Z-spray source of a Q-TOF2 mass spectrometer (Micromass, Manchester, United Kingdom). Flow rates usually varied from 8 to 16 nL/min. Instrument operation, data acquisition, and analysis were performed using MassLynx/BioLynx 3.5 software (Micromass). The capillary voltage and collision energy were optimized for each sample. The microchannel plate voltage was set to 2200 V. Data searches against the WSSV ORF database were performed using Global Server (Micromass).

Transcriptional Analyses of Genes

Shrimp Infection with WSSV—The tissue from *P. monodon* shrimp with pathologically confirmed infection was homogenized in TN buffer at 0.1 g/ml. After centrifugation at $2000 \times g$, the supernatant was diluted to 1:100 with 0.9% NaCl and filtered (0.22- μ m filter). 0.2 ml of the filtrate was injected intramuscularly into each healthy shrimp in the lateral area of the fourth abdominal segment. At various times post-infection (p.i.), four specimens were selected at random, and their hemolymphs were collected. The collected hemolymphs were immediately frozen and stored at -70 °C.

RT-PCR—The total RNA was isolated from WSSV-infected shrimp hemolymph according to the manufacturer's instruction (NucleoSpin RNA II; Macherey-Nagel GmbH & Co. KG, Germany). Then RT-PCR was performed with ORF-specific primers using a TITANIUM One-step RT-PCR kit (CLONTECH Laboratories, Inc.). The RT-PCR cycles were as follows: 50 °C for 1 h, 94 °C for 5 min, 30 cycles of 94 °C for 30 s, 65 °C for 30 s, 68 °C for 1 min, followed by an elongation at 68 °C for 2 min.

Characterization of WSSV vp466 Gene

Rapid Amplification of vp466 cDNA Ends (5' and 3' RACE)—The 5' and 3' cDNA of vp466 was revealed by RACE. For 5' RACE, a gene-specific primer SP1 and a nested SP2 were designed as 5'-GCTCTCCATCCGCTTAGTCACATTGGC-3' and 5'-GCCGAAGCT-GAAGGTTTTGGAGGTGC-3', respectively. For 3' RACE, the gene-specific primer SP3 was 5'-GCAGTAGCAAATCTCACCGGACCTG-TG-3'. RACE reaction was performed according to the instructions of the 5'/3' RACE kit (Roche Molecular Biochemicals).

Expression of GST-VP466 Fusion Protein in Escherichia coli—The vp466 gene was amplified using the synthesized forward primer 5'-CACGGATCCATGCTGCATCTTTAAT-3' with *Bam*HI site (*italic* and *underlined*) and the reverse primer 5'-AGACCCGGGTTATGACA-CAAACCTAT-3' with *Sma*I site (*italic* and *underlined*). The amplified DNA and plasmid vector pGEX-4T-2 were digested with *Bam*HI + *Sma*I. After purification and ligation of DNA fragments, the vp466 gene was inserted into pGEX-4T-2 vector downstream of GST and expressed in pGEX-4T-2-pLysS as a fusion protein with GST (Amersham Biosciences Corp.). The resulted recombinant plasmid was named pGEX466. The competent cells of *E. coli* BL-21 pLysS were transformed with the recombinant pGEX466, and colonies containing transformants were screened by colony PCR. The identity of pGEX466 was subsequently confirmed by both restriction enzyme digestion (*Bam*HI + *Sma*I) and DNA sequencing.

After overnight incubation at 37 °C, pGEX466-pLysS and pGEX-4T-2-pLysS (BL-21 pLysS containing pGEX466 and pGEX-4T-2, respectively) were inoculated into new media at a ratio of 1:100. When A_{600} reached 0.6, the cultures were induced with 1 mM IPTG and continued to grow for 6 h. Then the bacteria were spun down ($4000 \times g$) at 4 °C. The pellets were suspended in ice-cold phosphate-buffered saline (PBS) (containing 1% Triton X-100, 1 mM phenylmethane-

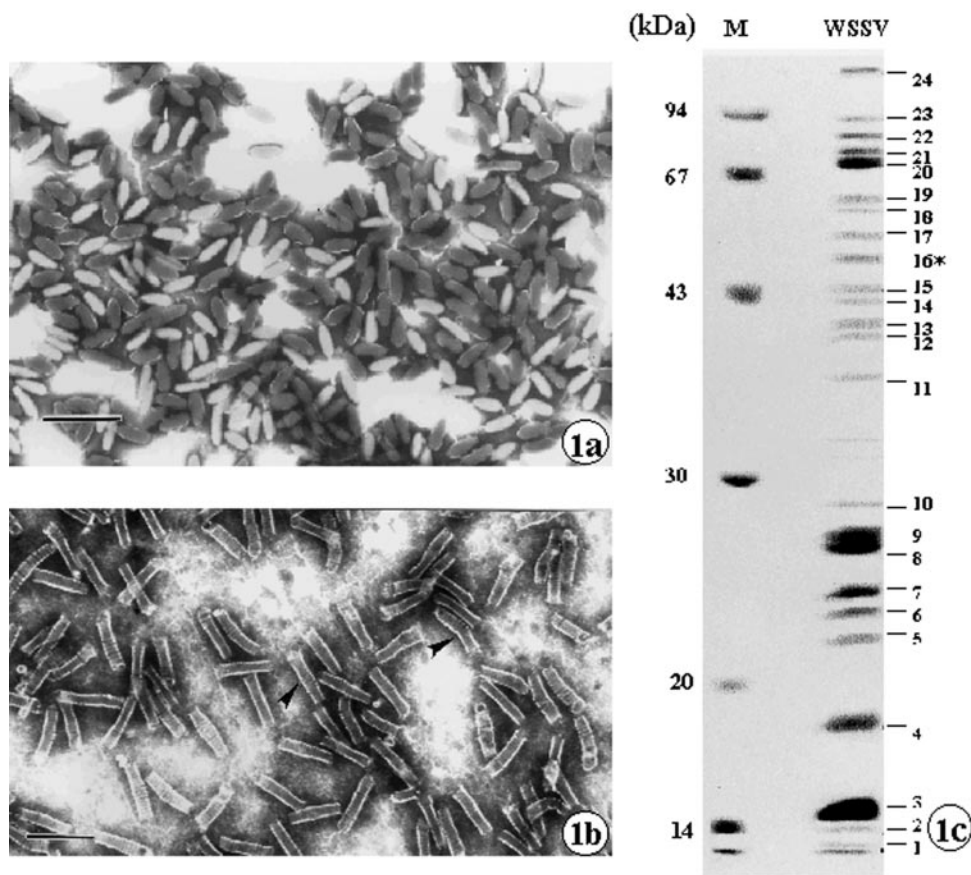


FIG. 1. **Electron micrograph of negatively stained WSSV virions and nucleocapsids; SDS-PAGE profile of purified WSSV virions.** *a*, electron micrograph of negatively stained intact WSSV virions. Scale bar = 720 nm (arrow). *b*, electron micrograph of negatively stained WSSV nucleocapsids (arrow). Scale bar = 370 nm. *c*, SDS-PAGE of purified WSSV virions. Numbers indicate the excised bands. *M*, protein molecular marker (kDa). The asterisk indicates band 16.

sulfonyl fluoride, 4 mM benzamidine, 10 μ g of leupeptin, and 10 μ g of aprotinin) and sonicated for 30 s on ice. After spinning at 60,000 $\times g$, the supernatant was mixed with 1 \times PBS-buffered glutathione-agarose beads (Sigma) and incubated at 4 $^{\circ}$ C for 2 h on a shaking device. The beads were washed three times with ice-cold 1 \times PBS and incubated in reducing buffer (50 mM Tris-HCl, 10 mM reduced glutathione, pH 8.0) at room temperature for 10 min. After centrifugation at 1000 $\times g$ for 5 min, the supernatant was collected and analyzed by SDS-PAGE.

Antibody Preparation—The purified GST-VP466 fusion protein was used to immunize mice (Swiss Albino, 3–4 weeks) once every 2 weeks by intradermal injection over an eight-week period. Titers of the antisera were 1:20,000 as determined by enzyme-linked immunosorbent assay. Protein A-SepharoseTM CL-4B was used to isolate anti-GST-VP466 IgG according to the manufacturer's instruction (Amersham Biosciences). The optimal dilution of purified IgG, after serial dilutions, was 1:1,000 as determined by enzyme-linked immunosorbent assay. Horseradish peroxidase-conjugated goat anti-mouse IgG was obtained from Sigma. For negative control, 1 \times PBS buffer was used.

Western Blot—WSSV virions were subjected to 12% SDS-PAGE, followed by transferring onto nitrocellulose membrane (Bio-Rad) in electroblotting buffer (Tris 25 mM, glycine 190 mM, methanol 20%) for 3 h. The membrane was immersed in blocking buffer (3% bovine serum albumin, 20 mM Tris, 0.9% NaCl, 0.1% Tween 20, pH 7.2) at 4 $^{\circ}$ C overnight, followed by incubation with a polyclonal mouse anti-GST-P466 IgG, pre-immune mouse serum, or mouse anti-GST IgG

for 3 h. Subsequently, horseradish peroxidase-conjugated goat anti-mouse IgG (Sigma) was used as the secondary antibody, and the detection was performed with a substrate solution containing 4-chloro-1-naphthol (Sigma).

Localization of VP466 by Immunoelectron Microscopy—WSSV virion and nucleocapsid were mounted onto the Formvar-coated, carbon-stabilized nickel grids, respectively, and the grids were then blocked with 2% AURION BSA-CTM (Electron Microscopy Sciences) for 1 h, followed by incubation with the primary antibody (purified polyclonal mouse anti-VP466 antibody IgG, 1:1000 dilution in 1% AURION BSA-CTM) for 2–3 h. After washing three to four times with 1 \times PBS, the grids were incubated with goat anti-mouse IgG conjugated with 15 nm colloid gold (Electron Microscopy Sciences) for 1 h. The grids were further washed two times with 1 \times PBS and briefly stained with 2% phosphotungstic acid (pH 7.0) for 1 min. The specimens were examined under the transmission electron microscope (JEOL 100 CXII, Japan). In the control experiments, the primary antibody was replaced with pre-immune mouse serum and mouse anti-GST antibody, respectively, and following other experimental steps.

RESULTS

Identification of WSSV Proteins by MS

WSSV was proliferated in an alternate host crayfish *C. clarkii*. The virus purification was performed according to an

TABLE I
WSSV genes and their products

| Band No. | Gene | Position in the WSSV genome | | Size ^a | | Estimated size from SDS-PAGE | GenBank™ accession number | Characteristics of deduced proteins | Gene transcription p.i. | Mass spectrometry | Sequence coverage | |
|----------|--------|-----------------------------|------------------|-------------------|------|------------------------------|---------------------------|-------------------------------------|-------------------------|-------------------|-------------------|-----|
| | | start codon | stop codon | aa | mass | | | | | | MALDI | ESI |
| | | | | | | <i>kDa</i> | | | <i>h</i> | | | % |
| 1 | vp68 | 228196 | 227993 | 68 | 7 | 7 | AF411464 | Not known | 18–36 | Q-TOF | | 19 |
| 2 | vp95 | 2672 | 2388 | 95 | 11 | 12 | AF402996 | Not known | 18 | MALDI | 72 | |
| 4 | vp121 | 241637 | 241275 | 121 | 13 | 17 | AF402997 | Not known | 6–24 | MALDI, Q-TOF | 28 | 19 |
| 5 | vp184 | 173178 | 173729 | 184 | 21 | 22 | AF402998 | Not known | 30 | MALDI | 7 | |
| 6 | vp208 | 1118 | 495 | 208 | 23 | 24 | AF402999 | vp24 of WSSV (Ref. 14) | 18–30 | MALDI, Q-TOF | 61 | 33 |
| 7 | p22 | 180036 | 179425 | 204 | 22 | 25 | AF227911 | vp26 of WSSV (Ref. 14) | 18–36 | MALDI, Q-TOF | 44 | 48 |
| 8 | p204 | 244242 | 244853 | 204 | 22 | 27.5 | AF308164 | vp28 of WSSV (Ref. 14) | 18–36 | MALDI, Q-TOF | 42 | 45 |
| 9 | p204 | | | | | 28 | | | | MALDI, Q-TOF | 42 | 36 |
| 10 | p204 | | | | | 29 | | | | Q-TOF | | 19 |
| 11 | vp281 | 141696 | 142538 | 281 | 32 | 35 | AF411634 | Not known | 30 | MALDI, Q-TOF | 27 | 7 |
| 12 | p204 | | | | | 37 | | | | Q-TOF | | 12 |
| 13 | vp300 | 132994 | 133893 | 300 | 34 | 38 | AF403003 | Not known | 18 | MALDI | 24 | |
| | vp292 | 130566 | 131441 | 292 | 33 | 38 | AF411636 | Not known | 24–48 | MALDI, Q-TOF | 24 | 5 |
| 14 | p22 | | | | | 41 | | | | Q-TOF | | 4 |
| | vp357 | 58956 | 60026 | 357 | 39 | 41 | AF403004 | Not known | 36 | MALDI | 20 | |
| 16 | vp466 | 177124 | 178521 | 466 | 52 | 50 | AF395545 | Not known | 24–30 | MALDI, Q-TOF | 24 | 11 |
| | vp384 | 142545 | 143696 | 384 | 43 | 50 | AF411635 | Not known | 24 | MALDI, Q-TOF | 12 | 8 |
| 18 | vp448 | 300432 | 299089 | 448 | 50 | 55 | AY048543 | Not known | 24 | MALDI | 9 | |
| 19 | vp544 | 241775 | 243406 | 544 | 62 | 60 | AY044842 | Not known | 36 | MALDI, Q-TOF | 8 | 8 |
| 21 | vp674 | 119057 | 121078 | 674 | 76 | 76 | AY048545 | Class I cytokine receptor | 24 | MALDI | 9 | |
| 23 | vp800 | 255075 | 257474 | 800 | 89 | 90 | AY044843 | ATP/GTP binding motif | 36 | MALDI, Q-TOF | 18 | 3 |
| 24 | vp1684 | 300501 ^b | 445 ^b | 1684 | 186 | 180 | AY048547 | Collagen | 30 | MALDI | 22 | |

^a Size of ORFs in amino acids (aa) and predicted molecular mass (kDa).

^b Size = 300500 → 305107 + 1 → 445.

improved protocol (11) (Fig. 1a). After being treated with Triton X-100 and CsCl gradient ultracentrifugation, the purified viral capsid was obtained (Fig. 1b).

The proteins of the WSSV virion were separated by SDS-PAGE. More than 20 bands ranging from 5 to 190 kDa were visible with Coomassie Blue staining (Fig. 1c). Twenty-four of them were excised from the gel. Following trypsin digestion of the reduced and alkylated WSSV proteins, the peptides were first analyzed by MALDI-TOF MS. Peak lists of tryptic peptide masses of each protein were generated and were subjected to the WSSV ORF database search using the MS-Fit search engine to identify the proteins and corresponding genes. Reliable matches with the genes from the theoretical WSSV ORF database, covering 7 to 72% of amino acid sequences, were obtained for 16 of the 24 gel-excised bands (Table I). Subsequently the tryptic peptides of each band were sequenced using nano-ESI-MS/MS. To identify the proteins, the peptide sequences derived from the MS/MS data were searched against the WSSV ORF database using a global server. 14 of the 24 protein bands were identified with 3–48% coverage of amino acid sequences, respectively (Table I).

In total, 18 genes from WSSV were identified by MALDI-TOF MS and Q-TOF from SDS-PAGE. Of the 18 proteins reported in this study, only three have been described previously, VP208, P22, and P204, the products encoded by the

vp208, p22, and p204 genes, respectively (14, 15). The remaining 15 proteins were identified for the first time. The P204 protein was found in bands 8, 9, 10, and 12, and the P22 protein in bands 7 and 14, respectively. On the other hand, two genes were revealed in bands 13, 14, and 16, respectively (Table I).

Structures of Genes and Homologies with Known Proteins

A guanine residue from the beginning of the largest *Bam*HI fragment was designated as the starting point of the physical map of the WSSV genome (6). The positions of 18 genes in the WSSV genome and their accession numbers in GenBank™ are listed in Table I. A typical TATA box sequence was found in the promoter regions of all 18 genes, indicating that this sequence may be essential in WSSV for the efficient transcription of these genes. Except for the vp184, vp300, and vp674 genes, the putative polyadenylation signal sequences (AATAAAA) were present downstream of the stop codons of the remaining 15 genes. Among the 18 genes, the start codons (ATGs) were in a favorable context for efficient eukaryotic translation initiation (PuNNATGPu) (16) for 14 genes. Exceptions were the vp357, vp466, vp544, and vp800 genes. The 18 genes encoded proteins ranging from 68 to 1684 amino acids (Table I).

MSASLILDEYLKKTASAVLDVADSFEEKIGKEIQSPEEAAALSVALYGAPPKPSASAVASIIITGERTSLNDKYLSDNVLLKMSVA
 RVGQENNRKRADQAADAEIRTIMEDITGSLSGAYRQYSPLEENKVHIGIMNNKTPSIVCGYYTMDTSSISSEPLSLTDFQNPVTI
 ANVTKRMESIFSKVDSARSTRFDAFVNGVANMMDIKSSIDWANMVENVIKLPDSTPNPCSVDTIVSRDASVVKTAVNDIYASVG
 KSYCRPATQLTFMSEIEKLRKAAVVCFEALMSDTRERAFVEFLFYVSPKEDASNTNSKLFVQNKLSMSGPNRQPIKLVRRSAE
 ETLFLGLCFMFKVMPPEFMNCIINFPTIPHSTQYHGLYGTCLTPLLRYGSSFEKSWAHFEELISERANAVKKFGVNDTRIDCLD
 AVANLTGPVYVLILDVLRVTLTSAQRSCSTKFLREIKENYLLWNRFFVS

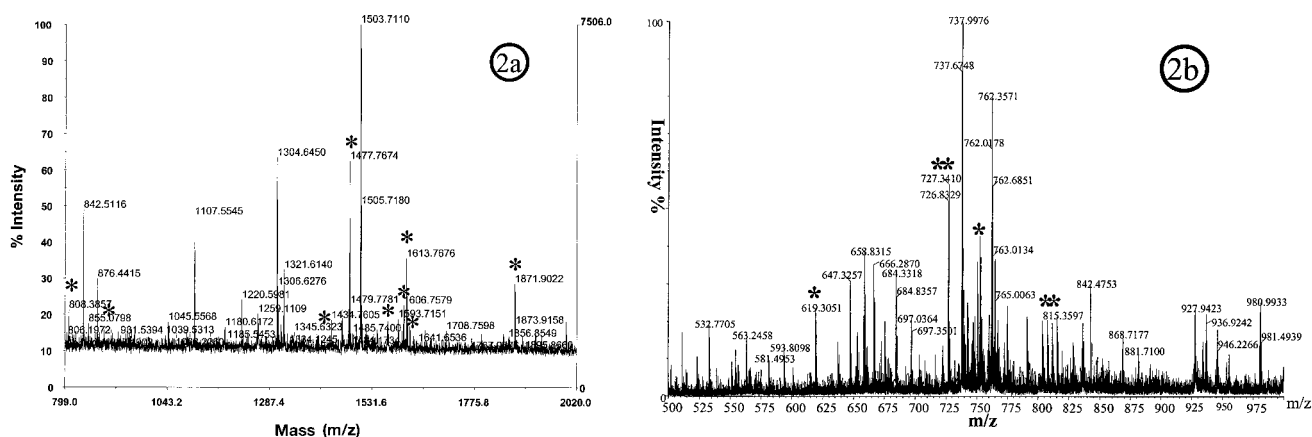


FIG. 2. **Characterization of VP466 by mass spectrometry.** a, MALDI-TOF MS spectrum of protein band 16. Peptides were produced by in-gel tryptic digestion. The peptide masses were used for the WSSV ORF database search and retrieved the vp466 gene. Tryptic peptides that map to the protein sequence encoded by the vp466 gene within a mass accuracy of 100 ppm were indicated by a dotted underline. b, nano-ESI-MS spectrum of protein band 16. Peptides were produced by in-gel tryptic digestion. The peptide solution was desalted by pipetting through C18 resin packed in a Millipore ZipTip before loading into a Micromass Q-TOF2 MS system. The band was found to be the encoding product of the vp466 gene after searching against the WSSV ORF database. Tryptic peptides that map to the protein sequence encoded by the vp466 gene within a mass accuracy of 100 ppm were indicated by a solid underline.

Based on homology searches of the 18 proteins using BLAST and BLAST2, putative functions of four of them could be assigned. Two genes contained sequence motifs based on PROSITE analysis (Table I). The remaining 12 genes showed no homology to any known proteins or sequence motifs. The vp1684 gene encoded a large protein (168 kDa) from the collagen family. This was the first time that an intact collagen gene was found in a virus. The collagen-like protein contained a typical repeat of Gly-X-Y (X was mostly proline, and Y could be any amino acid), but its function was not clear. The proteins encoded by the vp208, p22, and p204 genes showed characteristics of structural proteins (14). The p22 and p204 genes were identified further to encode envelope proteins of WSSV by immunoelectron microscopy in our earlier studies (15).

Temporal Analyses of Gene Transcriptions

RT-PCR was used to detect the ORF-specific transcripts in the total RNAs extracted from the hemolymph of adult *P. monodon* at various infection stages (0, 6, 18, 24, 30, 36, and 48 h p.i.) with WSSV. The transcripts of all 18 genes were detected at different post-infection stages (Table I). These results confirmed the coding fidelity of the ORFs. Based on

the temporal analysis, only the vp121 gene could be detected at 6 h p.i. until 24 h p.i. The remaining 17 genes were transcribed after 6 h p.i., suggesting that these genes were expressed in the late course of WSSV infection. However no putative late promoter element, ATAAG, canonical in insect baculoviruses, was found in the promoter regions of these late genes.

Characterization of WSSV vp466 Gene

Identification of WSSV vp466 Gene by Mass Spectrometry—The putative vp466 gene was associated with a 51-kDa protein, corresponding to band 16 (one of the minor bands) in the WSSV SDS-PAGE profile. Trypsin digests of reduced and alkylated protein from band 16 were first analyzed by MALDI-TOF MS (Fig. 2a). The WSSV ORF database search with the list of tryptic peptide masses identified one of the proteins as the product of the vp466 gene (termed as the VP466 protein). Four experimentally derived peptide masses were found to match the predicted peptide masses of the VP466 protein within 100 ppm, covering 21% of its amino acid sequence. The tryptic peptides of band 16 were sequenced subsequently by mass spectrometry using nano-ESI-MS/MS. The measured and calculated masses of the tryptic peptides of the nano-ESI-MS spectrum were shown in Fig. 2b. The re-

sulted amino sequences corresponded to the product of the vp466 gene after searching the WSSV ORF database.

Location of WSSV vp466 Gene—The start and stop codons of vp466 gene were present at 177,124 and 178,521 nucleotides, respectively, in the WSSV genome (6). This 1398-bp ORF presumably encoded a 466-amino acid protein (hence termed vp466 gene; GenBank™ accession number AF395545; the resulting protein termed as VP466), with a theoretical molecular mass of 51.2 kDa and with an isoelectric point of 6.51 (isoelectric point computation, EMBL). The vp466 gene was found to be located at the same 6-kb *Bam*HI fragment of WSSV genome as vp26/p22 gene but with an opposite orientation (15). A typical TATA box (TATAAAT) was present at 37 nucleotides upstream of the transcription initiation site, which was proved by 5' RACE, and 111 nucleotides upstream of the translation starting site. A putative polyadenylation (poly(A)) signal (AATAAA) is present at 161 nucleotides downstream of the translational stop codon of vp466 and 24 nucleotides upstream of poly(A) revealed by 3' RACE. The structure of vp466 gene and its deduced amino sequence are noted in Fig. 3.

Homology searches with WSSV VP466 were performed against GenBank™/EMBL, SwissProt, and PIR databases using FASTA, TFASTA, and BLAST; neither vp466 nucleotide sequence nor the deduced protein sequences shared significant homology to other proteins identified so far, which indicated that VP466 is a novel protein. Predictprotein (17) and NetOGlyc (18) showed that VP466 has three potential *N*-linked glycosylation (*N*-{P}-[ST]-{P}) and three *O*-glycosylation sites; ScanProsite predicted there are 20 possible phosphorylation motifs in VP466, which include six protein kinase C phosphorylation sites ([ST]-X-[RK]), thirteen casein kinase II phosphorylation sites ([ST]-XX-[DE]), and one tyrosine kinase phosphorylation site ([RK]-X(2)-[DE]-X(3)-Y). TopPred2 (19) predicted that there is a putative transmembrane α -helix formed by the amino acid residues 338 to 358 (Fig. 3).

Temporal Analysis of vp466 Gene Transcription—RT-PCR was used to detect the vp466 gene-specific transcript in the total RNA extracted from the hemolymph of adult *P. monodon* at various WSSV infection stages (0, 12, 18, 24, 30, 36, 48, and 60 h p.i.). The transcript was first detected at 18 h p.i. and reached the highest level at 30 h p.i. (Fig. 4), indicating that vp466 is a late gene.

Expression and Purification of GST-VP466 Fusion Protein—WSSV vp466 gene was cloned into pGEX-4T-2 vector and overexpressed as a GST fusion protein. A band corresponding to the GST-VP466 fusion protein was observed in the induced pGEX466-pLysS. No protein band was found at the same position in both of the induced and uninduced pGEX-4T-2-pLysS. The recombinant GST-VP466 was purified using affinity chromatography as shown in Fig. 5a, lane 6.

Western Blot and Localization of VP466 Protein—Western blot results showed that the polyclonal mouse anti-GST-VP466 antibody reacted strongly with band 16 (51 kDa) of

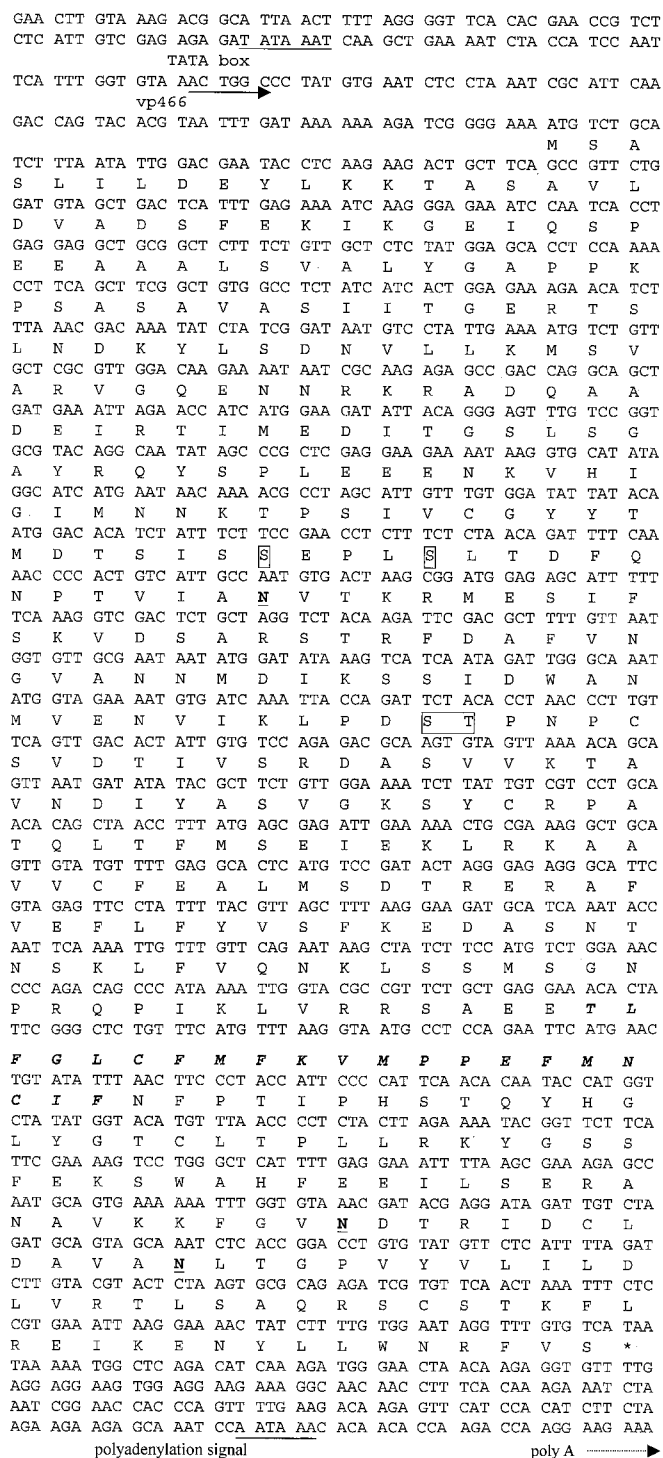


FIG. 3. WSSV-vp466 nucleotide sequence and the deduced VP466 protein sequence (one-letter code). The transcription initiation site of vp466 and position of poly(A) are indicated by solid and dashed lines with arrows, respectively. The predicted TATA box and the putative polyadenylation signal are underlined. The locations of putative *N*-linked glycosylation sites are boldfaced and underlined, and the locations of *O*-glycosylation sites are boxed. The putative transmembrane α -helix amino acids are italicized and boldfaced.

Downloaded from www.mcponline.org by on July 4, 2009

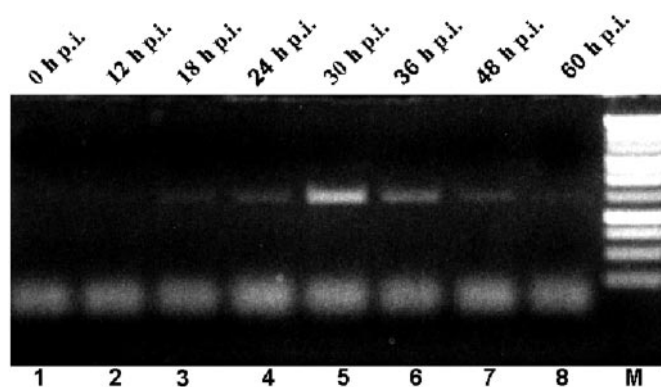


FIG. 4. Temporal analysis of vp466 transcription by RT-PCR. Lanes 1–6, RT-PCR detection of vp466 transcripts from the total mRNA collected chronologically. Lane M, DNA marker (1-kb ladder; Promega)

WSSV SDS-PAGE proteins (Fig. 5b, lane 3). No blot signals were observed if mouse anti-GST antibody was used as the primary antibody. It confirmed that band 16 was derived from WSSV. Transmission electron microscopy immunogold localization results showed that the high density gold particles were located at the viral envelope (Fig. 6a) but not the viral nucleocapsid (Fig. 6b). In the control experiments, no gold particles could be seen for either the WSSV virion or nucleocapsid (data not shown). These studies provided direct evidence that the VP466 protein should be involved functionally in the formation/assembly of WSSV envelope.

DISCUSSION

To date, about 20 shrimp viruses have been reported. Among them, WSSV is a major pathogen with a broad range of hosts, high infectivity, and high mortality. Thus there is an urgent need to study the proteins of this virus and their function to come out with a solution to prevent or cure this disease. Although previously only a few abundant structural proteins had been identified, the completion of the genomic DNA sequence of WSSV greatly facilitated the discovery of new proteins by proteomic approach. In this report, we were able to identify 18 proteins of WSSV by one-dimensional SDS-PAGE followed by mass spectrometry. Many of them might be difficult to identify by traditional methods because of their low abundance. Thus the proteomic approach was proved to be an effective and sensitive way for discovering WSSV proteins. The reliability of this approach was confirmed subsequently by ORF-specific RT-PCR of these identified genes.

MALDI-TOF MS and ESI-MS/MS are two complementary MS methods for proteomic analysis. Each of these MS methods can be used independently, but when the high throughput of MALDI-TOF MS analysis is combined with the sequence specificity of ESI-MS/MS analysis, identification of unknown proteins from database search is greatly facilitated. In this study, the proteins from WSSV SDS-PAGE gels were analyzed first by MALDI-TOF MS, and 16 WSSV genes with

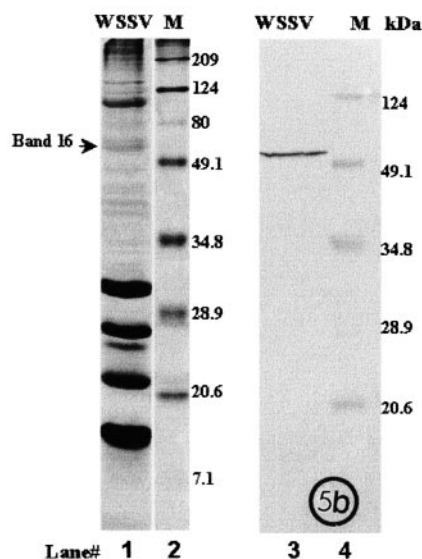
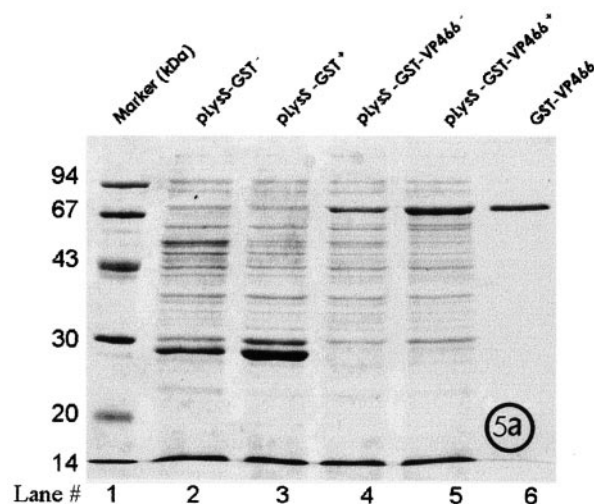


FIG. 5. Expression of VP466 in *E. coli* and Western blot analysis. a, expression of VP466 in *E. coli* BL21 pLysS. Shown are Coomassie Brilliant Blue-stained 12% SDS-PAGE gels with low molecular mass protein markers (lane 1); pGEX-4T-2 (vector only) pLysS, non-induced (lane 2); pGEX-4T-2-pLysS, induced (lane 3); pLysS-GST-VP466 (recombinant pGEX-4T-2 plasmid containing vp466 gene), non-induced (lane 4); pLysS-GST-VP466, induced (lane 5); purified GST-VP466 protein (lane 6). b, Western blot of the SDS-PAGE of WSSV virion with VP466 antibody IgG. Shown are Coomassie Brilliant Blue-stained 12% SDS-PAGE gels with purified WSSV virion (lane 1); with molecular mass protein markers (lanes 2 and 4); Western blot of the SDS-PAGE of intact WSSV with VP466 anti-mouse IgG (lane 3).

reliable matches were obtained (Table I). Despite lower sequence coverage, the peaks with stronger signals of the tryptic mass spectra of bands 5, 18, 19, and 21 were found to match the predicted peptide masses of the vp184, vp448, vp544, and vp674 genes, respectively. ESI-MS/MS analysis was used subsequently to analyze the tryptic samples be-

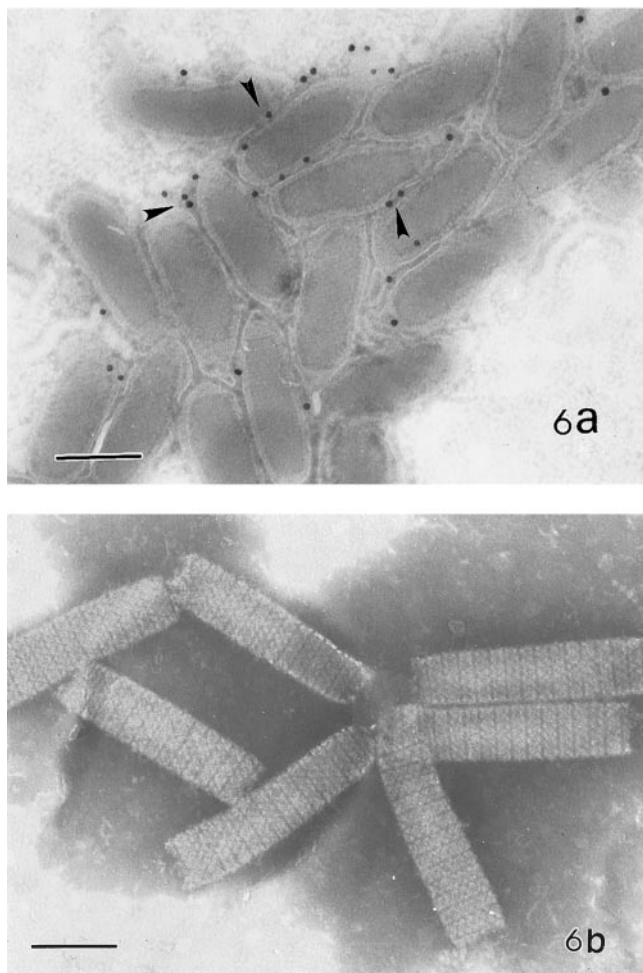


Fig. 6. **Localization of WSSV VP466 protein by immunoelectron microscopy.** a, gold particles localize in virus envelope (arrow indicated); scale bar = 180 nm. b, no labeling signal was found in the nucleocapsids; scale bar = 110 nm. The primary antibody is the purified mouse anti-GST-VP466 IgG, and the second antibody is goat anti-mouse IgG conjugated with 15 nm of gold (Electron Microscopy Sciences).

cause of its sequence specificity. Fourteen bands were revealed by Q-TOF (Table I). Moreover, some MALDI-TOF MS results were confirmed. Although the sequence coverage of bands 11, 13, 14, 16, 19, and 23 to the predicted proteins of the vp281, vp292, p22, vp384, vp544, and vp800 genes were lower, at least 10 successive amino acid sequences were matched. These sequence tags should be specific. Besides three bands containing two protein components, proteins encoded by the p204 and p22 genes were found to be present in different bands (Table I). This may be because of post-translation modifications such as glycosylation or lipidation.

In this study, vp466 gene was chosen as an example of proteomic identification and was characterized further. As this gene was derived from a minor protein band (band 16), the reliability of this assignment should theoretically be lower than those of more abundant bands. Western blot analysis showed

that the antibody against the expressed VP466 specifically recognized a 51-kDa protein of WSSV origin. Thus the authenticity of the proteomic results in this study was substantiated further. Immunogold labeling provided visualized evidence that VP466 is a WSSV envelope protein. This is the third WSSV envelope protein hitherto characterized. In the previous studies, VP28/P204, VP26/P22 have been identified as two major WSSV envelope proteins.²

Baculoviruses produce enveloped progeny viruses in the following two different ways: (i) budded virus obtains its envelope from the cell surface when the nucleocapsid buds through the plasma membrane as it exits the cell; (ii) occlusion-derived virus acquires its envelope with viral-induced membranes within the nucleoplasm (20). WSSV morphogenesis is exclusively intranuclear, where viral envelope is formed *de novo* (21, 22); this feature is closer to that of occlusion-derived virus than that of the budded virus. Although the precise mechanism of the induction of intranuclear membrane structures and protein movement to nuclear envelope is still unknown, evidence showed that some hydrophobic transmembrane domains, for example, lamin B receptor and herpes simplex virus glycoprotein B, are sufficient to direct proteins to be incorporated with nuclear membrane (20). It is thus speculated that VP466 putative transmembrane domain, formed by amino acid residues 338 to 358 (Fig. 3), may play a role in mediating the movement of this protein into the nuclear membrane and subsequently assembling as viral envelope within the nucleoplasm.

Computer analysis of VP466 predicted that there are multiple putative glycosylation, as well as phosphorylation sites. Viral glycoproteins are known to be responsible for cell tropism, spreading infection, and pathogenicity, etc., and there are accumulating evidences showing that phosphorylation of viral proteins plays important roles in initiating virus infections, e.g. it has been reported that phosphorylation of the *Plodia interpunctella* granulosis virus core protein may be required for the release of viral DNA from the nucleocapsid at the start of the infection process (23). There is little doubt that the putative post-translational modifications in VP466 should be important functionally in initiation of WSSV infection. Experiments on recombinant or deletion viral mutants will be needed to confirm the validity of these modification sites and to elucidate which of these subdomains are functionally important.

This study showed a gene-protein band pattern of WSSV based on one-dimensional PAGE. The samples used in this study were purified WSSV particles, which contain mainly structural proteins. However, many viral non-structural genes are usually transcribed at the early p.i. stages (*i.e.* 0–6 h p.i.). Because of the lack of a suitable cell line for culturing shrimp viruses, the hemolymph from various infection stages with

² VP26 was identified originally as a WSSV nucleocapsid protein (14). However, a recent study (15) confirmed that VP26/P22 is an envelope protein by immunogold labeling.

WSSV, instead of purified WSSV, should be used to reveal more genes, especially the early genes, in future studies. Two-dimensional PAGE has the capacity to separate thousands of proteins in a single analysis. The combination of two-dimensional PAGE and MS would reveal more new genes for the proteomic analysis of WSSV.

Acknowledgments—We express our thanks to Shashikant Joshi and Dr. Yunhan Hong for critical review of the manuscript, to Xianhui Wang, Xiaohua Wu, and Gek Luan Loy for technical assistance, and to professor You-an Ke and Yun Ping Lim for help in bioinformatics.

Note Added in Proof—VP19, which is identical to VP121 in our present paper, was characterized recently as a WSSV envelope protein (van Hulten, M. C., Reijns, M., Vermeesch, A. M., Zandbergen, F., and Vlaskovits, J. M. (2002) Identification of VP19 and VP15 of white spot syndrome virus (WSSV) and glycosylation status of the WSSV major structural proteins. *J. Gen. Virol.* (2002) **83**, 257–265).

* This investigation is supported by the National Science Technology Board, Singapore, under Grant “Establishment of a Laboratory of Excellence in Aquatic and Marine Biotechnology (LEAMB).” The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Contributed equally to this paper.

** To whom correspondence should be addressed: Dept. of Biological Sciences, National University of Singapore, Singapore 119260. Tel.: 65-8742692; Fax: 65-7792486; E-mail: dbshead@nus.edu.sg.

REFERENCES

- Chen, X. F., Chen, P., and Wu, D. H. (1997) Study on a new bacilliform virus in cultured shrimps. *Sci. China Ser. C* **27**, 415–420
- Lo, C. F., Ho, C. H., Peng, S. E., Chen, C. H., Hsu, H. C., Chiu, Y. L., Chang, C. F., Liu, K. F., Su, M. S., Wang, C. H., and Kou, G. H. (1996) White spot syndrome baculovirus (WSBV) detected in cultured and captured shrimp, crabs and other arthropods. *Dis. Aquat. Org.* **27**, 215–225
- Huang, C., Shi, Z., Zhang, J., Zhang, L., Chen, D., and Bonami, J. R. (1999) Establishment of a model for proliferating white spot syndrome virus *in vivo*. *Virol. Sin.* **14**, 358–364
- Chou, H. Y., Huang, C. Y., Wang, C. H., Chiang, H. C., and Lo, C. F. (1995) Pathogenicity of a baculovirus infection causing white spot syndrome in cultured penaeid shrimp in Taiwan. *Dis. Aquat. Org.* **23**, 165–173
- Yang, F., Wang, W., Chen, R. Z., and Xu, X. (1997) A simple and efficient method for purification of prawn baculovirus DNA. *J. Virol. Methods* **67**, 1–4
- Yang, F., He, J., Lin, X., Li, Q., Pan, D., Zhang, X., and Xu, X. (2001) Complete genome sequence of the shrimp white spot bacilliform virus. *J. Virol.* **75**, 11811–11820
- van Hulten, M. C., Witteveldt, J., Peters, S., Kloosterboer, N., Tarchini, R., Fiers, M., Sandbrink, H., Lankhorst, R. K., and Vlaskovits, J. M. (2001) The white spot syndrome virus DNA genome sequence. *Virology* **286**, 7–22
- Zhang, X., Xu, X., and Hew, C.-L. (2001) The structure and function of a gene encoding a basic peptide from prawn white spot syndrome virus. *Virus Res.* **79**, 137–144
- Naaby-Hansen, S., Waterfield, M. D., and Cramer, R. (2001) Proteomics-post-genomic cartography to understand gene function. *Trends Pharmacol. Sci.* **22**, 376–384
- Mann, M., and Pandey, A. (2001) Use of mass spectrometry-derived data to annotate nucleotide and protein sequence databases. *Trends Biochem. Sci.* **26**, 54–61
- Huang, C., Zhang, L., Zhang, J., Xiao, L., Wu, Q., Chen, D., and Li, J. K. (2001) Purification and characterization of white spot syndrome virus (WSSV) produced in an alternate host: crayfish, *Cambarus clarkii*. *Virus Res.* **76**, 115–125
- Hofmann, K., Bucher, P., Falquet, L., and Bairoch, A. (1999) The PROSITE database, its status in 1999. *Nucleic Acids Res.* **27**, 215–219
- Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of protein silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858
- van Hulten, M. C., Goldbach, R. W., and Vlaskovits, J. M. (2000) Three functionally diverged major structural proteins of white spot syndrome virus evolved by gene duplication. *J. Gen. Virol.* **81**, 2525–2529
- Zhang, X., Huang, C., Xu, X., and Hew, C.-L. (2002) The transcription and identification of an envelope protein gene (p22) from shrimp white spot syndrome virus (WSSV). *J. Gen. Virol.* **83**, 471–477
- Kozak, M. (1987) An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**, 8125–8148
- Bairoch, A., Bucher, P., and Hofmann, K. (1997) The PROSITE database, its status in 1997. *Nucleic Acids Res.* **25**, 217–221
- Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L., and Brunak, S. (1998) NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj. J.* **15**, 115–130
- Claros, M. G., and von Heijne, G. (1994) TopPred II: an improved software for membrane protein structure predictions. *Comput. Appl. Biosci.* **10**, 685–686
- Hong, T., Summers, M. D., and Braunagel, S. C. (1997) N-terminal sequences from *Autographa californica* nuclear polyhedrosis virus envelope proteins ODV-E66 and ODV-E25 are sufficient to direct reporter proteins to the nuclear envelope, intranuclear microvesicles and the envelope of occlusion derived virus. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 4050–4055
- Durand, S., Lightner, D. V., Redman, R. M., and Bonami, J. R. (1997) Ultrastructure and morphogenesis of white spot syndrome baculovirus (WSSV). *Dis. Aquat. Org.* **29**, 205–211
- van Hulten, M. C., Westenberg, M., Goodall, S. D., and Vlaskovits, J. M. (2000) Identification of two major protein genes of white spot syndrome virus of shrimp. *Virology* **266**, 227–236
- O’Reilly, D. R. (1997) In *The Baculoviruses* (Miller, Lois K., ed) 1st Ed., pp. 267–295, Plenum Press, New York