Identification and analysis of gp116 and gp64 structural glycoproteins of yellow head nidovirus of Penaeus monodon shrimp

Sarawut Jitrapakdee,1,2 Sasimanas Unajak,1,2 Nusra Sittidilokratna,1,2 Richard A. J. Hodgson,3 Jeff A. Cowley,3 Peter J. Walker,3 Sakol Panyim2 and Vichai Boonsaeng1,2

Correspondence
Sarawut Jitrapakdee (at Department of Biochemistry)
scj@mahidol.ac.th

1,2CENTEX Shrimp1 and Department of Biochemistry2, Faculty of Science, Mahidol University, Bangkok 10400, Thailand
3CSIRO Livestock Industries, Long Pocket Laboratories, Indooroopilly, Queensland, Australia

Yellow head virus (YHV) is a major agent of disease in farmed penaeid shrimp. YHV virions purified from infected shrimp contain three major structural proteins of molecular mass 116 kDa (gp116), 64 kDa (gp64) and 20 kDa (p20). Two different staining methods indicated that the gp116 and gp64 proteins are glycosylated. Here we report the complete nucleotide sequence of ORF3, which encodes a polypeptide of 1666 amino acids with a calculated molecular mass of 185 713 Da (pI = 6–68). Hydropathy analysis of the deduced ORF3 protein sequence identified six potential transmembrane helices and three ectodomains containing multiple sites for potential N-linked and O-linked glycosylation. N-terminal sequence analysis of mature gp116 and gp64 proteins indicated that each was derived from ORF3 by proteolytic cleavage of the polyprotein between residues Ala 228 and Thr 229, and Ala 1127 and Leu 1128, located at the C-terminal side of transmembrane helices 3 and 5, respectively. Comparison with the deduced ORF3 protein sequence of Australian gill-associated virus (GAV) indicated 83 % amino acid identity in gp64 and 71 % identity in gp116, which featured two significant sequence deletions near the N terminus. Database searches revealed no significant homology with other proteins. Recombinant gp64 expressed in E. coli with and without the C-terminal transmembrane region was shown to react with antibody raised against native gp64 purified from virions.

INTRODUCTION

Virus diseases are a serious problem for the large and rapidly expanding shrimp culture industry in Asia and the Americas, causing mass mortalities in ponds and heavy production losses (Flegel, 1997). In the Asian region, yellow head disease has been a major concern since it first emerged in Thailand in 1990 (Limsuwan, 1991). Gross signs often associated with yellow head disease include cessation of feeding, swimming near the surface and pond edges and the development of yellow colouration of the cephalothorax and gills (Chantanachookin et al., 1993). Lymphoid organs of moribund shrimp display evidence of necrosis and contain vacuolated cells with hypertrophied nuclei and densely basophilic cytoplasmic inclusions (Chantanachookin et al., 1993). Death of infected shrimp usually occurs within 2–3 days of the first appearance of visible signs of disease.

The nucleotide sequence reported in this paper has been submitted to GenBank with accession number AF540644.
Nadala et al. (1997) originally reported that YHV particles contain four structural proteins of approximately 170, 135, 67 and 22 kDa, of which the 135 kDa protein was glycosylated. However, Wang & Chang (2000) subsequently reported only three major YHV proteins (110, 63 and 20 kDa), suggesting that the larger protein may be of cellular origin. In GAV only two genes (ORF2 and ORF3), located immediately downstream of the ORF1b gene, have been predicted to encode structural protein (Cowley et al., 2001). ORF2 encodes a 22 kDa structural protein that appears to function as the nucleoprotein (J. A. Cowley and others, unpublished), and ORF3 encodes a polypeptide with the structural characteristics of a large glycoprotein with multiple membrane-spanning domains (Cowley & Walker, 2002). In this paper, we report the nucleotide and deduced amino acid sequences of the YHV ORF3 region. We show that the region encodes the major viral structural glycoproteins (gp116 and gp64): these are synthesized as a polypeptide which undergoes post-translational proteolytic cleavage and glycosylation.

**METHODS**

**Virus purification.** As source material for virus purification, 200–500 juvenile *Penaeus monodon* shrimp (average weight 20 g), were infected by intramuscular injection with a standard inoculum prepared as described previously from a 1998 Thai YHV isolate (Wongteerasupaya et al., 1995; Sittidilokratna et al., 2002). At 3 days post-infection, haemolymph was withdrawn and used as a source of virus for purification. Virions were purified by Urogaffrin (Sering) gradient ultracentrifugation as described by Wongteerasupaya et al. (1995). The stocks of purified virus were snap-frozen and stored at −80°C until required for analysis.

**SDS-PAGE and N-terminal sequencing.** Protein samples (10 μg) were analysed by SDS-PAGE (Laemmli, 1970) under reducing conditions in 7.5% or 15% discontinuous gels. The fractionated proteins were transferred to a PVDF membrane in transfer buffer (10 mM CAPS/10% methanol) using a semi-dry blotter (Hoeffer). Protein bands blotted onto the membrane were stained briefly in 0.02% Coomassie brilliant blue R250. The protein bands were then excised and analysed by N-terminal sequencing. Edman degradation was done with an Applied Biosystems Sequencer at the Center for Genetic Engineering and Biotechnology, Bangkok, Thailand.

**Polyclonal antibody production and immunoblot analysis.** Nitrocellulose-bound antigen was prepared as described by Diano et al. (1998). Briefly, 500 μg aliquots of total viral protein were blotted onto a nitrocellulose membrane (Amersham Pharmacia), excised and ground in 0.5 ml PBS (140 mM NaCl, 4 mM KCl, 2 mM KH₂PO₄, 8 mM Na₂HPO₄, pH 7.4) in liquid nitrogen. This antigen was emulsified with an equal volume of Freund’s incomplete adjuvant (Sigma) and used to immunize two BALB/c mice by intraperitoneal injection. Mice were boosted at 1 week intervals with 100 μg of viral protein in Freund’s incomplete adjuvant. The antisera were collected 1 week after the second boost and used for immunoblot analysis.

Proteins were transferred to a nitrocellulose membrane using a semi-dry electroblotting apparatus (Hoeffer). Following transfer, the membrane was blocked in 3% BSA, 0.5% Tween 20 in PBS for 3 h. The membrane was washed briefly in the same buffer without BSA and then reacted with a 1:10 000 dilution of mouse antiserum for 1 h. Goat anti-mouse polyclonal antibodies conjugated with alkaline phosphatase were then reacted for 2 h and immuno-reactive proteins were visualized by adding nitro blue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP).

**Glycoprotein staining.** Glycoprotein detection was performed using the ECL glycoprotein detection system (Amersham Pharmacia). Proteins blotted onto PVDF membranes were oxidized with 10 mM sodium metaperiodate in 100 mM acetate buffer pH 5·5 at room temperature for 20 min in the dark. The membrane was then reacted with 0·35 mM biotin hydrazone in 100 mM acetate buffer pH 5·5 for 1 h, followed by incubation with streptavidin conjugated with alkaline phosphatase. The glycoprotein bands were visualized by adding NBT and BCIP. Thymol staining was done as described by Racusen (1979).

**RT-PCR, cloning and sequencing.** YHV genomic RNA was extracted from purified virus using TRIzol Reagent (Invitrogen) and resuspended in DEPC-treated water. RT-PCR was performed in a total volume of 25 μl containing 200 ng genomic RNA, 0·4 mM of a forward primer (5’-GATCGGCGTACCTAGCTTATGCATCAGCAGCAGCTT-3’) and an oligo(dT) primer (5’-TCTAGAGGATCCTGCTGACGGTGTTTTTTTTTTTTTT-3’). The SuperScript one-step RT-PCR system (Invitrogen) was used in the presence of 1 unit of Elongase (Invitrogen) according to the instruction manual. The RT-PCR profile consisted of an initial incubation at 50°C for 30 min, 94°C for 2 min followed by 35 cycles of amplification. Each cycle consisted of denaturation at 94°C for 30 s and annealing/extension at 68°C for 8 min. The RT-PCR product was either sequenced directly or cloned in pGEM-T Easy vector (Promega). The RT-PCR product was gel-purified using the QIAquick Gel Extraction kit (Qiagen) and sequenced directly using Big Dye reagent (ABI). Nucleotide sequences obtained from initial reactions were used to design new primers to generate overlapping sequences toward the 5’- and the 3’-ends of the fragment. Sequence chromatograms were then analysed and a consensus sequence generated using SeqEd 1·0·3 (ABI).

**Recombinant protein expression in *E. coli*.** The cDNA encoding full-length gp64 was generated by PCR using forward primer (YHV-D7) 5’-GCCCTCTAGACATATGCTGCTGACCACGACGGGCGTTT-3’ and reverse primer (YHV-D8) 5’-CATTGTTGACCCTCCTACAGT-GTGATGTGATGTAGGAGATGGGATCTTGGTTGGTCTCATTGCGAAGT-3’. The forward primer was designed from residues L128 to V1135 in the ORF3 polyprotein and included an initiation codon and an Ndel restriction site (underlined). The reverse primer was designed from residues T1657 to S1666 and included stop codons (bold) and a BamHI restriction site (underlined). PCR was performed in a 25 μl reaction mixture containing 1× PCR buffer (10 mM Tris/HCl pH 8·3, 50 mM KCl, 1·5 mM MgCl₂, 0·1% Triton X-100), 0·2 mM of each dNTP, 1 ng oligo-primed cDNA, 0·25 μM of each primer and 2 units of Tha DNA polymerase (Perkin Elmer). The reaction mixture was subjected to 35 cycles of denaturation at 94°C/30 s, annealing/extension at 68°C/2 min, and followed by the final extension at 72°C/10 min. The PCR products were cut with Ndel and EcoRI and ligated into the multiple cloning site of pET17b (Novagen). A transmembrane-deleted construct was generated by removal of 40 C-terminal residues (Y1627–5’GGATCTCTACTATTCCTCGTTCGTGGCCTGCA; corresponding to residues F1620–D1626 with stop codons (bold) and an EcoRI cloning site of pDrive (Novagen). The PCR product was cloned into pDrive vector (Qiagen) and sequenced. The insert was digested with Ndel and EcoRI and cloned into the multiple cloning site of pET17b. The recombinant plasmids were then transformed into *E. coli* BL21(DE3) (Novagen).
diluted 1:20 with fresh LB broth and grown at 37°C to an OD$_{600}$ of 0.5–0.6. The culture was then induced with 1 mM IPTG for 1–7 h. A 1 ml volume of the culture was pelleted, suspended in protein loading buffer, heated at 100°C for 10 min and analysed by SDS-PAGE. The recombinant protein expressed in *E. coli* was confirmed by immunoblot analysis using gp64 polyclonal antiserum.

**RESULTS**

**Analysis of structural proteins of YHV virions**

Transmission electron microscopy of YHV virions purified from shrimp haemolymph revealed typical rod-shaped,
enveloped particles as described by Wongteerasupaya et al. (1995) (data not shown). The purified virus particles were disrupted with SDS and analysed by SDS-PAGE. As shown in Fig. 1, analysis under reducing conditions revealed three relatively abundant proteins of molecular mass 116, 64 and 20 kDa. Similar results were obtained by gel electrophoresis under non-reducing conditions (data not shown), indicating that the virion proteins were not linked by intermolecular disulfide bonds. As shown in Fig. 1, haemocyanin (~68–70 kDa) is the most abundant protein in shrimp haemolymph (Figueroa-Soto et al., 1997). As there was little evidence of residual haemocyanin following purification, the virus preparation appeared to be largely free of cellular proteins.

The YHV structural proteins were examined for evidence of glycosylation using two different detection methods. The ECL glycoprotein detection assay is based on oxidation of the ketone group of sugar residues with sodium metaperiodate and conjugation of the oxidized ketone ring with biotin (Murray et al., 1989). The carbohydrate can then be visualized in situ on a membrane. The second method used the thymol reagent and is based on the hydrolysis of glycosidic bonds concomitant with the formation of a furfural derivative (Racusen, 1979). As shown in Fig. 1(B, C) each method indicated that the 116 and 64 kDa proteins were glycosylated. There was no evidence of glycosylation of the 20 kDa protein.

Each of the three YHV structural proteins (gp116, gp64 and p20) was subjected to N-terminal sequence analysis. The N-terminal sequences for glycoproteins gp116 and gp64 were determined to be T-I-L-S-G-I-P-E-K-D- and L-A-P-R-Q-A-R-V-X-G- (X, uncertain residue), respectively. No sequence was obtained for protein p20, which appeared to be blocked at the N terminus.

**Nucleotide sequence and deduced amino acid sequence of ORF3**

A region of the YHV genome extending from the 3′-poly(A) tail to a locus at the 3′-end of the ORF1b gene was amplified by RT-PCR. Analysis of the amplified product by agarose gel electrophoresis revealed a single band of approximately 6–0 kbp. By comparison with the known complete sequence of GAV (Cowley et al., 2000; Cowley & Walker, 2002) and available partial sequence data on the YHV genome (N. Sittidilokratna and others, unpublished data), the size of the amplified product was consistent with the expected size of 3′-end of YHV genome (Fig. 2). Nucleotide sequence analysis revealed that the region contained two long open reading frames in the same sense (+) as ORF1b. The largest open reading frame (ORF3) commenced 848 nucleotides downstream of the ORF1b termination codon and comprised 4998 nucleotides. The complete nucleotide sequence and deduced amino acid sequence of YHV ORF3 is shown in Fig. 3. ORF3 encodes a polypeptide of 1666 amino acids with a predicted molecular mass of 185,713 Da and a pI of 6.68. Alignment with the N-terminal sequences of mature gp116 and gp64 identified perfect identity with residues T229–E236 and L1128–Y1135 respectively of the encoded polypeptide. The data indicated that gp116 and gp64 are derived by post-translational proteolysis of the ORF3 polypeptide. Based on the identified N-terminal sequences and presumed sites of proteolysis, the calculated molecular mass for unglycosylated gp116 was 101,734 Da and gp64 was 58,599 Da. The predicted size of these products was consistent with evidence of post-translational glycosylation of each protein (Fig. 1B, C).

A hydropathy plot of the ORF3 polypeptide sequence (http://sosui.proteome.bio.tuat.ac.jp) predicted six hydrophobic transmembrane helices – three in the N-terminal domain (residues L25–L47, V1138–E1155, E208–V230), two in the central domain (residues G1027–N1049 and F1106–L1128) and one in the C-terminal domain (residues K1630–L1652). Analysis of membrane topology using TMHMM (Krogh et al., 2001) predicted that the first transmembrane domain anchors the N terminus of the ORF3 polypeptide inside the cell. Three subsequent ectodomains would be connected by the second to the sixth transmembrane domains respectively, and the C terminus of the protein would also be anchored within the cell (Fig. 4). Proteolytic cleavage between A228 and T229, and A1127 and L1128 near the C terminus of transmembrane helices 3 and 5 respectively would generate the 25–4 kDa N-terminal fragment, gp116 and gp64. The model predicts that gp64 is a type I transmembrane glycoprotein anchored in the viral envelope at the C terminus (Fig. 4). According to the predicted topology, gp116 is a polytopic type III transmembrane glycoprotein, anchored at the C terminus but with both C terminus and N terminus external. The model also predicts the formation of an unidentified 25–4 kDa protein with three membrane-spanning domains and type III membrane topology. Prosite analysis using the NetOGlyc2.0 database (http://www.cbs.dtu.dk) (Hansen et al., 1997) predicted O-linked glycosylation sites at threonine residues 255, 414, 569, 807, 828, 1585 and at serine residues 70, 232, 413, 417, 576 and 832 of the ectodomains of the ORF3 polypeptide. As shown in Fig. 3, potential N-linked glycosylation sites were also identified in the ectodomains of gp64 (four sites), gp116 (seven sites) and the unidentified 25–4 kDa transmembrane fragment of ORF3 (two sites). The ectodomains were also rich in cysteine residues (Fig. 3).
suggesting complex folded loop structures in the mature glycoproteins.

FASTA, TFASTA and BLAST searches of the deduced gp116 and gp64 sequences against GenBank/EMBL, SWISS-PROT and PIR protein databases revealed no significant similarity with other proteins including the spike glycoproteins of other nidoviruses (i.e. coronaviruses, toroviruses and arteriviruses). However, comparison with the available deduced sequence of the GAV ORF3 polyprotein (Cowley & Walker, 2002) indicated overall amino acid sequence identity of 75%. The level of amino acid sequence identity was higher in gp64 (83% identity) than in gp116 (71% identity), primarily due to two significant deletions and variability in the N terminus of GAV gp116 (see Fig. 5). Comparison of individual sites in the YHV and GAV ORF3 polyproteins indicated that all cysteine residues located in the predicted glycoprotein ectodomains were conserved. All but two (N$^{291}$–S$^{293}$ and N$^{711}$–T$^{713}$ in gp116) of the potential YHV N-glycosylation sites were preserved in GAV but four potential glycosylation sites in the GAV polyprotein were not present in YHV. All four N-glycosylation sites in gp64 were preserved and five of seven gp116 N-glycosylation sites were preserved in GAV. One common glycosylation site was present in the predicted ectodomain of the N-terminal fragment of the polyprotein and one preserved glycosylation site was located in the predicted endodomain and so would not be functional. Three of eight predicted O-glycosylation sites in YHV gp116 were not conserved and the single predicted O-glycosylation site in YHV gp64 was not conserved in GAV, suggesting that this protein may contain only N-linked glycans.

Expression of gp64 in E. coli

The gp64 coding region of YHV ORF3 was amplified by PCR. The forward primer spanned L$^{1128}$–V$^{1155}$ in ORF3 and contained an upstream in-frame initiation codon. The reverse primer spanned T$^{1657}$–S$^{1666}$ and included the endogenous translation termination codon. The PCR product was sequenced and cloned into pET17b expression vector. Recombinant plasmid pET17b-gp64 was transformed into E. coli BL21(DE3). As shown in Fig. 6(B), expression of recombinant gp64 was not evident by SDS-PAGE analysis and Coomassie blue staining of the E. coli whole cell lysate. However, by immunoblot analysis using polyclonal antibody against the native gp64 purified from virions, an immunoreactive ~58 kDa band was detected on the blot (lanes 3–6). The polyclonal antibody also reacted with native gp64 from purified virus (lane 7) but did not react with lysates prepared from E. coli transformed with pET17b alone (lane 1). Although the expression level was relatively poor, the size of the band was consistent with the molecular mass of gp64 calculated from the deduced amino acid sequence. The difference in molecular mass between the native and recombinant proteins is most likely due to the absence of post-translational glycosylation in E. coli. We also observed that the amount of the expressed ~58 kDa protein decreased 1 h after induction, indicating the recombinant gp64 produced in E. coli is not stable.
Fig. 5. Comparison of protein encoded from ORF3 of YHV and GAV. Both sequences were aligned using CLUSTAL W. (Thompson et al., 1994). The fully conserved residues are shown by asterisks (*) while the highly conserved residues are shown by (+). The residue number is also indicated.
As the transmembrane region of gp64 may cause poor expression, we subsequently generated a transmembrane-deleted construct (pET17b-gp64ΔT) [see Fig. 6A] by removing residues Y1627–S1666 located at the C terminus of gp64. This construct was transformed into E. coli and induced with IPTG. Again, there was no evidence of an overexpressed protein band corresponding in size to the gp64 lacking the transmembrane domain (i.e. ~56 kDa) on the Coomassie-stained gel. However, Western blot analysis of this gel revealed the presence of a strong immunoreactive band of approximately 56 kDa. The intensity of this band increased over the induction period (1–5 h), suggesting that recombinant gp64 without the transmembrane domain was more stable than full-length gp64.

DISCUSSION

In mammalian nidoviruses, the virion envelope glycoproteins have functions that are principally associated with the recognition of target cells, the fusion of viral and cellular membranes and the release of virions from infected cells. The envelope glycoproteins also have a central role in the induction of the host immune response and are targets for protective immunity (for a review see Spaan et al., 1988). In some vertebrate nidoviruses, the surface glycoproteins may also play an important role in determining virulence (Almazan et al., 2000; Phillips et al., 1999). Little is presently known of the structure or function of the envelope glycoproteins of the newly discovered invertebrate nidoviruses infecting shrimp or of the defensive response of shrimp (or any other invertebrate) to viral infection. In this paper, we report an initial characterization of the glycoproteins of yellow head nidovirus. We show that YHV virions contain three major structural proteins including a ~20 kDa non-glycosylated protein and two envelope glycoproteins (gp116 and gp64) that are expressed as a polyprotein from ORF3 and processed by post-translational proteolysis. We also show that recombinant gp64 expressed with or without a C-terminal transmembrane region is recognized by antibodies to the native virion glycoprotein.

Nadala et al. (1997) reported previously that YHV contains four major structural proteins (molecular masses 170, 135, 67 and 22 kDa) of which only the 135 kDa protein was reported to be glycosylated. Wang & Chang (2000) reported only three structural proteins (molecular masses 110, 63 and 20 kDa). Our results are in close agreement with this more recent report and we demonstrate that each of the larger structural proteins is glycosylated. Small differences in the molecular mass of the structural proteins could well be due to differences in the conditions of electrophoresis or to variations in the pattern of glycosylation of different YHV isolates.

N-terminal sequence analysis of gp116 and gp64 allowed precise identification of sequences encoding these proteins within the 1666 amino acid ORF3 polyprotein. The size of the YHV polyprotein is similar to those of the glycoproteins of vertebrates coronaviruses including avian infectious bronchitis virus (1160 amino acids) (Binns et al., 1985), feline infectious peritonitis virus (1452 amino acids) (de Groot et al., 1987a), murine hepatitis virus (1376 amino acids) (Spann et al., 1988), porcine epidemic diarrhoea virus (Duarte & Laude, 1994) (1383 amino acids) and human
Structural glycoproteins of the yellow head virus

respiratory coronavirus (1353 amino acids) (Mounir & Talbot, 1993). However, the predicted structure of the ORF3 polyprotein is more complex, comprising six putative transmembrane regions generating three ectodomains and two intra-virion endodomains. By contrast, coronaviruses contain a single transmembrane domain located near the C terminus of the spike polyprotein.

The envelope spikes of vertebrate coronaviruses and toroviruses comprise a large (~180 kDa) glycoprotein that is often cleaved by a cellular protease to yield two similarly sized subunits (S1 and S2) which remain non-covalently associated in virions (Sturman et al., 1985; Cavanagh, 1995; Snijder & Horzinek, 1995). In YHV, gp64 and gp116 are also generated by proteolysis of a ~180 kDa polyprotein. Although each is rich in cysteine residues, SDS-PAGE under non-reducing conditions indicated that gp116 and gp64 are not covalently associated. It is possible that, like the S1 and S2 subunits of coronavirus surface glycoprotein, gp64 and gp116 are associated non-covalently to form the peplomers evident on the virion surface. In coronaviruses, the S2 subunit appears to form the membrane-bound stalk while the S1 subunit forms the globular head of the spike (de Groot et al., 1987b) which interacts with the host cell receptor (Kubo et al., 1994; Suzuki & Taguchi, 1996). The S2 subunit contains several functional domains including a membrane anchor, six strictly conserved cysteine residues and a leucine zipper motif (Britton, 1991) which has been shown to mediate the oligomerization of the subunits and induce cell fusion (Grosse & Siddell, 1991) which has been shown to mediate the oligomerization of the subunits and induce cell fusion (Grosse & Siddell, 1994; Luo et al., 1999). The S2 ectodomain also contains two large amphipathic α-helices with a heptad repeat that have been proposed to mediate coiled-coil interchain interactions (de Groot et al., 1987b). As reported for GAV (Cowley & Walker, 2002), examination of YHV gp64 and gp116 sequences revealed no heptad repeats or significant amphipathic α-helices, suggesting an absence of coiled-coil structures. Further work is required to determine whether the processed gp116 and gp64 are associated to form the envelope spikes.

Sequence analysis was facilitated by use of an oligo(dT) primer to amplify the 3′-terminal region of the YHV genome. This approach was adopted in the expectation that the YHV genome would be polyadenylated – a characteristic of the viruses in the order Nidovirales. The sequence data showed that the ORF3 gene, encoding the YHV structural glycoproteins, is located downstream of the ORF1b gene that encodes replication enzymes and has features characteristic of nidoviruses (Sittidilokratna et al., 2002). In GAV, ORF2 (located in the genome between ORF1b and ORF3) appears to encode a 22 kDa nucleoprotein (J. A. Cowley and others, unpublished) and this corresponds to the major YHV virion protein p20. Clearly, GAV and YHV are closely related in sequence and share a common gene organization which appears unique amongst nidoviruses in that the nucleoprotein gene (ORF2) is upstream of the structural glycoprotein gene (ORF3) and there is no discrete gene encoding the integral membrane (M) protein (Cowley & Walker, 2002). However, it is evident from the location of proteolytic cleavage sites in the YHV ORF3 polyprotein that the N-terminal fragment has not yet been identified in virions or infected cells. If no further cleavage occurs, this 227 amino acid (25.4 kDa) fragment will be of similar size to the integral membrane proteins of coronaviruses and toroviruses (225–262 amino acids) which occur abundantly in virions and appear to have a role in intracellular budding (Rottier, 1995). The YHV ORF3 N-terminal fragment also contains three membrane-spanning domains characteristic of the M proteins but the predicted membrane topology is in the reverse orientation, with an N-terminal cytoplasmic tail and the C terminus oriented external to the membrane. Also, as p20 appears to be the nucleoprotein encoded in ORF2 (J. A. Cowley and others, unpublished), there is no evidence that YHV has a major virion component corresponding to the coronavirus M protein. The use of antibodies against suitable peptides in the N-terminal domain should assist identification and characterization of the N-terminal fragment of the ORF3 polyprotein.

The molecular masses of gp116 and gp64 calculated from the nucleotide sequence are smaller (~15 kDa and ~4 kDa, respectively) than those estimated by SDS-PAGE for the mature forms of the proteins. The size differences can be explained in part by glycosylation as revealed by the detection of carbohydrate (Fig. 1) and the presence of putative N- and O-linked glycosylation sites in each protein. Although the expression level of recombinant gp64 was relatively poor, it was readily detected using polyclonal antibodies raised against the native form. The size of recombinant gp64 was consistent with the molecular mass calculated from deduced amino acid sequence. Poor expression of gp64 in E. coli suggests that the protein is unstable and may not be properly folded when expressed independently of gp116. An attempt to express gp64 as the fusion protein with glutathione S-transferase failed to improve the yield but expression was improved after removal of the C-terminal transmembrane domain (Fig. 6C). Similar results were obtained when forms of gp116 with and without transmembrane domain were expressed in E. coli (data not shown). Expression of these proteins in insect cells, which may present an intracellular environment similar to shrimp cells, could offer a more useful approach to investigations of the processing and assembly of the glycoproteins encoded in YHV ORF3.

Overall, these data indicate that, despite defining similarities in the ORF1b region, YHV, like GAV, has a genome organization and glycoprotein expression strategy that are fundamentally different from vertebrate nidoviruses. The unique features of the viral structural glycoproteins support the classification of YHV with GAV as new tax (family Roniviridae, genus Okavirus) within the Nidovirales (Cowley et al., 2000; Sittidilokratna et al., 2002; Cowley & Walker, 2002).
ACKNOWLEDGEMENTS

This work was supported by the National Center for Genetic Engineering and Biotechnology (BIOTEC), the Thailand Research Fund (TRF) and the Australian Centre for International Agricultural Research (ACIAR). S.J. is a recipient of the TRF postdoctoral fellowship. S.U. is supported by Graduate Studies Scholarship, Mahidol University and the TRF senior research scholarship. N.S. is supported by Graduate Studies Scholarship, Mahidol University and the PhD Golden Jubilee program, TRF. S.P. is a TRF senior research scholar. We thank Dr O. Gajanandana for her assistance with the production of antibodies and Professor John Wallace for critical reading of the manuscript.

REFERENCES


