Glycosylation of gp116 and gp64 envelope proteins of yellow head virus of *Penaeus monodon* shrimp

Chumporn Soowannayan,¹,²,³ Jeff A. Cowley,¹ Roger D. Pearson,¹ Tristan P. Wallis,⁴† Jeffrey J. Gorman,⁴† Wojtek P. Michalski⁵ and Peter J. Walker¹,⁵

¹CSIRO Livestock Industries, Queensland Bioscience Precinct, St Lucia, QLD 4067, Australia
²Centex Shrimp, Faculty of Science, Mahidol University, Rama VI Road, Bangkok 10400, Thailand
³National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Thailand Science Park, Klong Luang, Patumthani 12120, Thailand
⁴Institute for Molecular Bioscience, Queensland Bioscience Precinct, The University of Queensland, St Lucia, QLD 4072, Australia
⁵CSIRO Livestock Industries, Australian Animal Health Laboratory, Geelong, VIC 3220, Australia

Yellow head virus (YHV) is a highly virulent pathogen of *Penaeus monodon* shrimp that is classified in the genus *Okavirus*, family *Roniviridae*, in the order *Nidovirales*. Separation of virion proteins treated with peptide-N-glycosidase-F (PNGase-F) in SDS-polyacrylamide gels and the use of glycoprotein-specific staining methods indicated that the gp116 and gp64 envelope glycoproteins possess N-linked rather than O-linked glycans. Competitive binding inhibition of lectins with various oligosaccharide specificities indicated that glycans linked to gp64 are mannose-rich, whilst glycans linked to gp116 possess terminal N-acetylgalactosamine and N-acetylglucosamine in addition to terminal mannose-type sugars. Mass spectrometry analyses of peptides generated from YHV proteins before and after deglycosylation with PNGase-F, using combinations of the endoproteinases trypsin, Asp-N and Lys-C, confirmed occupancy of six of the seven potential N-linked glycosylation sites in gp116 and three of the four potential sites in gp64.

INTRODUCTION

Yellow head virus (YHV) is an important pathogen of *Penaeus monodon* shrimp. It is an enveloped, rod-shaped virus approximately 70 × 180 nm in size (Chantanachookin et al., 1993; Flegel et al., 1995; Wongteerasupaya et al., 1995) and is classified as the type species of the genus *Okavirus*, family *Roniviridae*, in the order *Nidovirales* (Cowley et al., 1999; Cowley & Walker, 2002; Walker et al., 2005; Ziebuhr et al., 2003). The YHV genome comprises a positive-sense, 26 662 nt ssRNA containing four ORFs designated ORF1a, ORF1b, ORF2 and ORF3 (Sittidilokratna et al., 2006). ORF1a and ORF1b overlap at a ribosomal frame-shift site and encode polyproteins pp1a and pp1ab, from which replicate proteins are derived by proteolytic cleavage (Sittidilokratna et al., 2002, 2008). ORF2 encodes the 146 aa p20 nucleocapsid (N) protein (Cowley et al., 2004; Sittidilokratna et al., 2006). ORF3 encodes a 1666 aa polyprotein (pp3) containing six hydrophobic regions, which are predicted to be transmembrane (TM) domains (Jitrapakdee et al., 2003). Polyprotein pp3 undergoes post-translational cleavage to produce an N-terminal 228 aa polypeptide of unknown function, as well as envelope glycoproteins gp116 (899 aa) and gp64 (539 aa) (Jitrapakdee et al., 2003). Electron microscopy has shown that gp116 forms the prominent envelope surface projections on mature virions (Soowannayan et al., 2003). Although both glycoproteins are suspected to play crucial roles in cellular binding and entry, only antibodies to gp116, but not to gp64, inhibit virus infection (Assavalapsakul et al., 2005, 2006). Indeed, a shrimp cell-surface receptor that binds gp116 and is purported to mediate cell attachment has been identified (Assavalapsakul et al., 2006). As in other viruses, glycosylation of these envelope proteins is likely to

Correspondence
Chumporn Soowannayan
chumporn@biotec.or.th or
csoowannayan@gmail.com

Received 18 March 2010
Accepted 10 June 2010

© 2010 CSIRO

DOI 10.1099/vir.0.022111-0
be critical to their structure and function in cell-receptor attachment and membrane fusion (Quinones-Kochs et al., 2002; Reitter et al., 1998; Strauss et al., 1991).

Whilst the YHV gp116 and gp64 sequences contain seven and four putative N-linked glycosylation sites, respectively (Jitrapakdee et al., 2003), glycan occupancy and structure at these sites have not been investigated. Here, lectin-binding assays and mass spectrometry (MS) analyses of peptides generated from native and deglycosylated proteins following digestion with various combinations of proteases have been used to determine the nature of the glycans attached to N-linked sites in the mature gp116 and gp64 glycoproteins.

RESULTS

**gp116 and gp64 contain only N-linked carbohydrates**

Purified virions of the Chonburi/1999 YHV strain were treated with peptide-N-glycosidase-F (PNGase-F) to remove N-linked carbohydrates from proteins prior to their separation by SDS-PAGE. Compared with the estimated masses of the native gp116 (approx. 116 kDa) and gp64 (approx. 64 kDa) glycoproteins, PNGase-F removal of glycans reduced their masses to approximately 100 and 58–60 kDa, respectively (Fig. 1a). Masses of the deglycosylated gp116 and gp64 proteins corresponded well with those predicted by their amino acid compositions (101.45 and 58.60 kDa, respectively), deduced from the ORF3 gene sequence determined for the same YHV strain. This indicated that the additional masses of the native glycoproteins, i.e. approximately 16 kDa for gp116 and approximately 4.6 kDa for gp64, were mostly due to N-linked glycans. In comparison, the electrophoretic mobility of the YHV p20 N protein was unaffected by treatment with PNGase-F, consistent with the absence of potential N-linked glycosylation sites. Coomassie brilliant blue (CBB) staining also revealed that the YHV protein preparation contained minor amounts of the two isoforms (approx. 75 and 73 kDa) of shrimp haemocyanin (Sellos et al., 1997) and it was noted that the relative mass of each isoform was reduced following PNGase-F treatment.

To investigate whether N-linked glycans were the only carbohydrate moieties attached to the YHV gp116 and gp64 glycoproteins, PNGase-F-treated virion proteins separated by SDS-PAGE were stained with the glycoprotein-specific Pro-Q Emerald reagent (Invitrogen). The Pro-Q Emerald staining method utilizes periodic acid to oxidize protein-linked carbohydrates, which can then conjugate to fluorescent Pro-Q Emerald 300 hydrazide dye, thus allowing glycoproteins to be detected under UV light. Staining for the native gp116 and gp64 proteins occurred, but staining of gp64 was lost and staining of gp116 was reduced markedly after N-linked glycans had been removed with PNGase-F (Fig. 1b). Similarly, three glycoprotein markers known to be N-glycosylated were not stained by Pro-Q Emerald reagent following PNGase-F digestion, confirming the effectiveness of N-glycan removal (Fig. 1b). As expected, the YHV p20 N protein was not stained by Pro-Q Emerald reagent.

**Glycan structure of gp116 and gp64**

The nature of the terminal-sugar structures on the N-linked glycans attached to gp116 and gp64 was examined by competitive binding of lectins with unique carbohydrate specificities (Table 1). As shown in Fig. 2, *Helix pomatia* agglutinin (HPA) and wheatgerm agglutinin (WGA) bound to gp116 and binding was eliminated by pre-exposure
of the lectins to N-acetyl-α-D-galactosamine (α-D-GalNAc) and N-acetylgalcosamine (GlcNAc), respectively. Canavalia ensiformis agglutinin (ConA) also bound to gp116, although binding was not inhibited completely by pre-exposure to methyl α-D-mannose. Similarly to ConA, Sambucus nigra agglutinin (SNA) bound to gp116, but binding was inhibited only marginally by pre-exposure to α-lactose. Involvement of sialic acid linked (α-2–3) to galactose in Maackia amurensis agglutinin (MAA) II binding to gp116 was not assessed, due to the unavailability of the competing sugar.

Only ConA and SNA bound to gp64 and, whilst ConA binding was inhibited markedly by methyl α-D-mannose, SNA binding was inhibited only marginally by α-lactose. MAA I and Ulex europeaus agglutinin I (UEA I), which recognize galactosyl (β-1,4)-GlcNAc and α-linked fucose, respectively, failed to bind to either gp116 or gp64 (Fig. 2).

MAA II, ConA, UEA I and SNA also bound to the p20 N protein of YHV (Fig. 2). However, glycoprotein-specific staining undertaken here (Fig. 1) and previously (Jitrapakdee et al., 2003) has shown clearly that the p20 protein is not glycosylated. Moreover, as lectin binding was not inhibited by pre-exposure to the competing sugars, these interactions presumably occurred through non-specific interactions, possibly dictated by the N protein being highly basic (pI=9.84) and containing high numbers of charged residues and prolines (Cowley et al., 2004; Sittidilokratna et al., 2006). The larger (approx. 75 kDa) isoform of shrimp haemocyanin bound specifically to ConA, indicating the presence of methyl α-D-mannose in this isoform, but not in the approximately 73 kDa isoform. The ovalbumin marker protein bound specifically to WGA, consistent with the presence of (Man)5(GlcNAc)2-Asn and (Man)6(GlcNAc)2-Asn side chains (Tai et al., 1975). For the three lectins that did not bind to either gp116 or gp64 (i.e. MAA I, MAA II and UEA I), their activity was confirmed in blots undertaken identically using human type O erythrocyte membrane proteins (see Supplementary Material (including Supplementary Fig. S1), available in JGV Online), which are known to contain several glycoproteins decorated with glycans specific for each of these lectins (Gahmberg et al., 1978; Hormia et al., 1983; Imberty et al., 2000; Kawaguchi & Osawa, 1976).

### Table 1. Biotinylated lectins, their binding specificities and competing sugars

<table>
<thead>
<tr>
<th>Source</th>
<th>Lectin (final conc.)</th>
<th>Carbohydrate-binding specificity</th>
<th>Competitor (conc. used)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maackia amurensis</td>
<td>MAA I (10 μg ml⁻¹)</td>
<td>Galactosyl (β-1,4) GlcNAc</td>
<td>α-Lactose (0.2 M)</td>
</tr>
<tr>
<td>M. amurensis</td>
<td>MAA II (5 μg ml⁻¹)</td>
<td>Sialic acid linked (2–3) to galactose</td>
<td>Human glycophorin</td>
</tr>
<tr>
<td>Sambucus nigra</td>
<td>SNA (10 μg ml⁻¹)</td>
<td>Sialic acid linked (2–6) to galactose–α-Lactose (0.2 M)</td>
<td></td>
</tr>
<tr>
<td>Ulex europeaus</td>
<td>UEA I (10 μg ml⁻¹)</td>
<td>1-Fucose</td>
<td>α-L-Fucose (0.2 M)</td>
</tr>
<tr>
<td>Canavalia ensiformis</td>
<td>ConA (10 μg ml⁻¹)</td>
<td>α-D-Mannose</td>
<td>Methyl α-D-mannose (0.2 M)</td>
</tr>
<tr>
<td>Triticum vulgare</td>
<td>WGA (10 μg ml⁻¹)</td>
<td>β-1,4-D-GlcNAc (terminal GlcNAc or chitobiose structure)</td>
<td>d-GlcNAc (0.2 M)</td>
</tr>
<tr>
<td>Helix pomatia</td>
<td>HPA (5 μg ml⁻¹)</td>
<td>Terminal α-D-GalNAc</td>
<td>d-GalNAc (0.2 M)</td>
</tr>
</tbody>
</table>

### Fig. 2. YHV proteins separated by SDS-PAGE, transferred to nitrocellulose membranes and detected using various biotinylated lectins and the same lectins pre-exposed to their specific sugar (+C). Lectins/agglutinins used were MAA I and II, which bind(s) to galactosyl (β-1,4) GlcNAc and sialic acid linked (2–3) to galactose, respectively; SNA, which binds to sialic acid linked (2–6) to galactose; UEA I, which binds to α-fucose; ConA, which binds to α-D-mannose; WGA, which binds to β-1,4-D-GlcNAc; and HPA, which binds to terminal α-D-GalNAc. The positions of YHV gp116, gp64 and p20, as well as of shrimp haemocyanin (arrow), are indicated. M1, ColorBurst prestained protein molecular mass standards; M2, full-range Rainbow prestained molecular mass markers; M3, low-range molecular mass markers.
Occupancy of potential N-linked glycosylation sites in gp116 and gp64

For accurate MS analyses, the 1666 aa ORF3 sequence of the Chonburi/1999 YHV strain (GenBank accession no. EF156405) was determined and found to possess 12 amino acid changes compared with the prototype Thai YHV strain (GenBank accession no. EU487200.1) (Jitrapakdee et al., 2003). The predicted 899 aa gp116 and 539 aa gp64 components of ORF3 respectively contained seven and four potential N-linked glycosylation sites (NXT/S, where X=all amino acids except proline), and each was located at the same position as in the prototype YHV strain (Jitrapakdee et al., 2003) (Fig. 3). To identify occupancy of each glycosylation site, MS analyses were performed using protease-digested native as well as PNGase-F-deglycosylated gp116 and gp64. The removal of N-linked glycans by PNGase-F results in the conversion of asparagine to aspartic acid at the site of cleavage and thus a 1 Da (i.e. 1 m/z) increase in mass. As this change is readily identified by MS, native and PNGase-F-treated gp116 and gp64 were digested with various combinations of trypsin, Lys-C and Asp-N and the resultant peptides were analysed by matrix-assisted laser desorption/ionization–time of flight (MALDI-TOF) MS.

MS analyses of 14 different YHV protein digests identified 97 peptides with masses that matched those predicted from in silico digests of the deduced gp116 amino acid sequence, with an overall sequence coverage of 81.2 % (730 of 899 aa; Fig. 3). From these accumulated MS data, eight peptides derived from PNGase-F-treated gp116 were identified with masses approximately 1 Da greater than predicted. These peptides encompassed six of the seven peptides containing putative N-linked glycosylation sites in gp116 (Table 2). No MS peaks were obtained in any of the gp116 digests with masses predicted for peptides containing the putative N-linked site N44ST. Thus, although N-glycosylation of all six other potential sites in gp116 was confirmed, the occupancy status of the N44ST site remains undetermined. Example total ion MS spectral fingerprints of trypsin digests of untreated and PNGase-F-treated gp116 are shown in Fig. 4(a, b). Example MS spectra expanded from these fingerprints showing the appearance of a mass peak (m/z 2559.27) following PNGase-F treatment, corresponding to a deglycosylated peptide sequence NTHPSFSAL-SYIDYDVTAGSVVR 1 Da greater in mass than the corresponding peptide predicted from the gp116 sequence, are shown in Fig. 4(c, d).

MS analyses of the various YHV protein digests identified 43 peptides with masses that matched those predicted from in silico digests of the deduced gp64 amino acid sequence, with an overall sequence coverage of 67.5 % (364 of 539 aa; Fig. 3). Analysis following PNGase-F treatment identified masses of six peptides that were 1 Da greater than predicted, encompassing three of four putative N-linked sites (Table 3; Fig. 3). A mass peak corresponding to that predicted for a peptide sequence containing the other putative N-linked site (IN93WSCLHTGSCK) was detected with both the native and PNGase-F-treated gp64 protein samples. As no masses approximately 1 Da greater than this were detected in any MS analysis, we conclude...

Fig. 3. Deduced ORF3 gene amino acid sequence of the Chonburi/1999 YHV strain, separating the N-terminal and gp116 and gp64 products cleaved from the pp3 polyprotein. Sequences covered by peptide masses identified by MALDI-TOF MS using native and PNGase-F-treated proteins digested by using various combinations of trypsin, Lys-C and Asp-N are shaded. Putative N-linked glycosylation sites (bold font) identified by MS to be occupied are boxed. The six hydrophobic TM domains are underlined and signal peptidase type 1-like cleavage positions proceeding TM domains 3 and 5 are indicated (arrows).
Table 2. MALDI-TOF MS peptide masses that matched masses predicted for various theoretical peptides in YHV gp116 that contain occupied N-linked glycosylation sites

<table>
<thead>
<tr>
<th>MS mass (Da)</th>
<th>Predicted mass (Da)</th>
<th>Peptide sequence</th>
<th>Modification*</th>
<th>No. miscleavages</th>
<th>Peptide (preparation method)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1023.54</td>
<td>1023.54</td>
<td>(R)LPVQHTEVLNID</td>
<td>-linked glycosylation</td>
<td>0 (DG, Asp-N)</td>
<td>(R)LTPQSVEHILNID (DG, Tryp, Asp-N)</td>
</tr>
<tr>
<td>1062.53</td>
<td>1061.54</td>
<td>(K)MVNVTYDORCQ</td>
<td>-linked glycosylation</td>
<td>2 (DG, Tryp)</td>
<td>(K)MVNVTYDORCQ (DG, Tryp)</td>
</tr>
<tr>
<td>1078.50</td>
<td>1077.53</td>
<td>(K)MVNVTYDORCQ</td>
<td>-linked glycosylation</td>
<td>2 (DG, Tryp)</td>
<td>(K)MVNVTYDORCQ (DG, Tryp)</td>
</tr>
<tr>
<td>1380.73</td>
<td>1379.67</td>
<td>(K)MVNVTYDORCQ</td>
<td>-linked glycosylation</td>
<td>1 (DG, Tryp, Lys-C)</td>
<td>(K)MVNVTYDORCQ (DG, Tryp, Lys-C)</td>
</tr>
<tr>
<td>2188.93</td>
<td>2187.95</td>
<td>(K)MVNVTYDORCQ</td>
<td>-linked glycosylation</td>
<td>2 (DG, Tryp)</td>
<td>(K)MVNVTYDORCQ (DG, Tryp)</td>
</tr>
<tr>
<td>2559.27</td>
<td>2558.26</td>
<td>(K)MVNVTYDORCQ</td>
<td>-linked glycosylation</td>
<td>1 (DG, Tryp)</td>
<td>(K)MVNVTYDORCQ (DG, Tryp)</td>
</tr>
<tr>
<td>3418.47</td>
<td>3417.53</td>
<td>(K)MVNVTYDORCQ</td>
<td>-linked glycosylation</td>
<td>2 (DG, Tryp)</td>
<td>(K)MVNVTYDORCQ (DG, Tryp)</td>
</tr>
<tr>
<td>3662.77</td>
<td>3661.77</td>
<td>(K)MVNVTYDORCQ</td>
<td>-linked glycosylation</td>
<td>1 (DG, Tryp)</td>
<td>(K)MVNVTYDORCQ (DG, Tryp)</td>
</tr>
</tbody>
</table>

*Met-ox, Oxidation of methionine.  
|DG, Deglycosylated (PNGase-F-treated); Tryp, trypsin-treated; Lys-C, Lys-C-treated; Asp-N, Asp-N-treated.

DISCUSSION

Reported here are analyses identifying the nature of the terminal-sugar structures present in the potential N-linked glycosylation sites occupied in the gp116 and gp64 envelope glycoproteins of YHV, a penaeid shrimp nidovirus. The nature of the attached glycans was determined by using oligosaccharide-specific staining and by assessing the binding affinity of gp116 and gp64 for various biotinylated lectins; occupancy of putative N-linked attachment sites was determined by MS.

Staining of YHV structural proteins separated by SDS-PAGE using the Pro-Q Emerald carbohydrate-specific dye confirmed that, in virions, both gp116 and gp64, but not the p20 N protein, are glycosylated (Jitrapakdee et al., 2003; Nadala et al., 1997). Following the removal of N-linked glycans from the YHV structural proteins with PNGase-F, the relative masses of both gp116 and gp64 determined by SDS-PAGE were reduced by amounts consistent with most or all glycans being attached at N-linked sites. However, as some residual Pro-Q Emerald staining of gp116 persisted following its treatment with PNGase-F, either deglycosylation at N-linked sites did not proceed to completion, possibly due to structural hindrances limiting PNGase-F access to some sites, as found in other glycoproteins (Blanchard et al., 2008; Sugrue, 2007), or low amounts of glycan(s) are attached to at least one of its N-linked glycosylation sites (Jitrapakdee et al., 2003). Whilst O-linked glycans are not common in viral glycoproteins, they do, for example, occur in envelope glycoproteins of group 1 and 3 coronaviruses (de Haan et al., 1998; Niemann et al., 1982) and the SARS (severe acute respiratory syndrome-associated) coronavirus (Oostra et al., 2006).

Removal of N-linked glycans with PNGase-F is commonly used in conjunction with MS analysis to identify N-linked sites that are occupied, as demonstrated for example in studies of the human immunodeficiency virus 1 gp120 envelope glycoprotein (Zhu et al., 2000) and the SU component of the avian sarcoma leukosis virus (ASLV) envelope glycoprotein (Kvaratskhelia et al., 2004). N-Linked site occupancy can be determined by MALDI-TOF MS, as deglycosylation of asparagine by PNGase-F results in its conversion to aspartic acid, which is accompanied by a 1 Da increase in peptide mass (Kuster et al., 1997; Kvaratskhelia et al., 2004; Mills et al., 2000; Sugrue, 2007; Tarentino & Plummer, 1994; Tarentino et al., 1985). A mass difference of this magnitude is detected readily by MALDI-TOF MS when used in reflectron mode to obtain peptide mass accuracies of approximately 150 p.p.m. (difference of approx. 0.15 Da per 1000 Da) (Henzel & Stults, 2001). To identify which of the potential N-linked sites in YHV gp116 and gp64 glycoproteins are occupied, that the putative N93WS site is not occupied in the gp64 glycoprotein of the Chonburi/1999 YHV strain.
MS data were accumulated for 14 independent preparations of either native or PNGase-F-deglycosylated YHV glycoproteins digested by using various combinations of trypsin, Asp-N and Lys-C. These data identified that six of seven N-linked sites were occupied in gp116, and three of four sites were occupied in gp64. The N93WS site in gp64 was not glycosylated, but the glycosylation status of the gp116 N44ST site remains unknown, as no masses were detected in the MS analyses that matched peptides containing this potential N-linked site.

Whilst, in any particular glycoprotein, it is common for only about two-thirds of predicted N-linked glycosylation motifs (NXT/S) to be occupied by glycans (Apweiler et al., 2001).

**Table 3.** MALDI-TOF MS peptide masses that matched masses predicted for various theoretical peptides in YHV gp64 that contain occupied N-linked glycosylation sites

<table>
<thead>
<tr>
<th>MS mass (Da)</th>
<th>Predicted mass (Da)</th>
<th>Δ mass (Da)</th>
<th>Modification</th>
<th>No. miscleavages (preparation method)*</th>
<th>Peptide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>859.37</td>
<td>858.39 + 1</td>
<td>+0.02</td>
<td>N-Linked glycosylation</td>
<td>0 (DG/Tryp, Asp-N)</td>
<td>(R)(N/D)YTADFK(T)</td>
</tr>
<tr>
<td>974.40</td>
<td>973.42 + 1</td>
<td>+0.02</td>
<td>N-Linked glycosylation</td>
<td>0 (DG/Lys-C, Asp-N and DG/Asp-N)</td>
<td>(C)(N/D)NFTASASEF(D)</td>
</tr>
<tr>
<td>988.47</td>
<td>987.47 + 1</td>
<td>0.00</td>
<td>N-Linked glycosylation</td>
<td>0 (DG/Lys-C, Asp-N)</td>
<td>(K)(N/D)SEYLFK(P)</td>
</tr>
<tr>
<td>1822.93</td>
<td>1821.87 + 1</td>
<td>+0.06</td>
<td>N-Linked glycosylation</td>
<td>1 (DG/Asp-N/Tryp and DG/Asp-N)</td>
<td>(K)(N/D)SEYLFKPLGQHTSN(D)</td>
</tr>
<tr>
<td>1940.88</td>
<td>1939.88 + 1</td>
<td>0.00</td>
<td>N-Linked glycosylation</td>
<td>0 (DG/Tryp)</td>
<td>(K)ADHLCNFTASASEF(DL(K)^V)</td>
</tr>
</tbody>
</table>

*DG, Deglycosylated (PNGase-F-treated); Tryp, trypsin-treated; Lys-C, Lys-C-treated; Asp-N, Asp-N-treated.
Glycosylation of YHV gp116 and gp64

1999), higher occupancy rates often occur in viral glycoproteins. For example, 10 of 11 potential sites are utilized in the ASLV SU envelope glycoprotein component (Kvaratskhelia et al., 2004). Moreover, N-linked glycans tend to occur more commonly in N-terminal, rather than C-terminal, regions of glycoproteins, possibly due to their generally more distant positioning from the lipid membrane and a higher propensity for glycan addition by oligosaccharyltransferase (OST) during protein translation (Gavel & von Heijne, 1990). Whilst threonine and serine generally occur at relatively equal frequency in N-linked glycosylation motifs, sequons containing a threonine are glycosylated at an approximately 3-fold higher frequency than those with a serine (Kaplan et al., 1988; Lennarz, 1987). In the YHV glycoproteins, NXT\NXS sequons occur at a 5:2 ratio in gp116, but equally (2:2) in gp64, and this general glycosylation trend is consistent with the N\NWS site in gp64 not being occupied. The efficiency at which OST glycosylates asparagines can vary between sites, as evidenced, for example, by the asparagine in N25GT which OST glycosylates asparagines can vary between sites, generally occur at relatively equal frequency in (Gavel & von Heijne, 1990). Whilst threonine and serine oligosaccharyltransferase (OST) during protein translation generally more distant positioning from the lipid mem-

C-terminal, regions of glycoproteins, possibly due to their tend to occur more commonly in N-terminal, rather than

http://vir.sgmjournals.org 2469

In vertebrates, N-linked glycans attached to proteins are generally less complex than those occurring in vertebrate glycoproteins and more similar to those found in plants. Moreover, plants and invertebrates possess an ability to produce fucose-linked glycans not produced in vertebrates (Wilson, 2002). However, as UEA I, which recognizes α-linked fucose, did not bind to either gp116 or gp64, these YHV glycoproteins appear not to possess fucose-linked glycan. In Caenorhabditis elegans, most N-linked glycans attached to proteins have a high mannose composition similar to that in vertebrates, but few of any very hybrid types or more complex types containing outer-chain galactose and/or sialic acid, and synthesized as in vertebrates (Schachter, 2004). For a glycoprotein expressed in lepidop-
teran Spodoptera frugiperda cells from a recombinant baculovirus, analyses of N-linked protein glycosylation using lectin-binding specificities pre- and post-treatment with endoglycosidase detected glycans rich in mannose, fucose and GlcNAc, but not galactose or sialic acid, which were detected similarly in the same glycoprotein expressed in mammalian COS cells (Jarvis & Finn, 1995). Although competing sugars were not used to confirm lectin-binding specificity, in another similar study in S. frugiperda and Trichoplusia ni, lectin binding to cellular glycoproteins suggested that many might possess some complex glycans (Davis & Wood, 1995).

This study of the crustacean ronivirus YHV provides the first insights into the nature of the carbohydrates attached to the envelope glycoproteins of a non-vertebrate nido-
virus. Based on the nature of the N-linked glycans found attached to YHV gp116 and knowledge of the enzymatic process required to obtain GalNAc and GlcNAc (Kornfeld & Kornfeld, 1985; Spiro, 2002; Wilson, 2002), it can be deduced that, after nucleocapsids have acquired envelopes by budding through membranes at the rough ER (Chantaenchaoakin et al., 1993; Flegel et al., 1995; Spann & Lester, 1997; Wang et al., 1996), glycans attached to the envelope glycoproteins are processed further as virions are transported from the ER, either via the trans side of the Golgi network or through the trans-Golgi network itself.
METHODS

Virus purification. The YHV strain used originated from moribund P. monodon shrimp collected in 1999 from farms in Chonburi, Thailand, experiencing outbreaks of yellow head disease (Soowannayan et al., 2003). The ORF3 gene region of this YHV strain was sequenced (GenBank accession no. EF156405). Experimental infection and virus purification methods have been described previously (Soowannayan et al., 2003). Briefly, haemolymph of infected shrimp was withdrawn into a syringe containing an equal volume of shrimp salt solution (450 mM NaCl, 10 mM KCl, 10 mM EDTA, 10 mM HEPES, pH 7.3) and pooled. Haemocytes were removed by centrifugation at 2000 × g for 10 min, the supernatant was filtered through a 0.45 μm membrane and aliquots of the crude YHV preparation were stored at −80 °C until used to infect P. monodon shrimp experimentally by tail-muscle injection. YHV particles were purified from haemolymph pooled from 50 experimentally infected shrimp by ultracentrifugation through a sucrose cushion as described previously (Soowannayan et al., 2003) and stored in aliquots at −80 °C.

PNGase-F removal of N-linked glycans from purified virions. Proteins from approximately 50 μg purified YHV virions were precipitated with methanol, dried and dissolved in 25 μl 0.1 M sodium phosphate buffer, pH 8.0, containing 0.1% SDS and 2% 2-mercaptoethanol. The protein solution was heated at 95 °C for 5 min and cooled at room temperature for 15 min. To remove N-linked glycans from proteins, the solution was incubated at 37 °C for 3 h in the presence of 2 U PNGase-F (Roche). After digestion, a 20 μl aliquot was mixed with 20 μl 2× SDS-PAGE loading buffer, heated and separated by SDS-PAGE in a 12.5% gel along the same amount (approx. 15 μg) of untreated viral protein sample. The gel was electrophoresed at 110 V for 90 min and stained with CBB R-250.

Peptide preparation to identify occupied N-linked glycosylation sites by MS. Gel bands containing either untreated or PNGase-F-treated gp116 and gp64 proteins were excised from CBB-stained gels, placed in microcentrifuge tubes and the dye was removed by incubation in two changes of 200 μl 200 mM ammonium bicarbonate, 50% acetonitrile at 37 °C for 15 min. The destained gel pieces were dried in a vacuum centrifuge (Savant DNA SpeedVac) set at medium heat (approx. 45 °C) for 1 h. Protein present in each gel band was then reduced by incubation at 65 °C for 1 h in 100 μl 20 mM DTT, 25 mM ammonium bicarbonate. The solution was discarded and reduced protein was alkylated in the dark by incubation in 100 μl 50 mM iodoacetamide in 25 mM ammonium bicarbonate at 37 °C for 40 min, with occasional vortexing. The solution was discarded and the gel piece containing reduced and alkylated protein was washed three times in 200 μl 25 mM ammonium bicarbonate at 37 °C for 15 min. The gel piece was then dried in a SpeedVac vacuum centrifuge.

In-gel proteolytic digestion of proteins. Aliquots of reduced and alkylated gp116 or gp64 glycoproteins were digested with the endopeptidases trypsin, Asp-N or Lys-C. An aliquot of peptides extracted from each digestion was analysed by MALDI-TOF MS. The remainder was digested further using a different enzyme before MALDI-TOF MS analysis. Dried gel bands containing the proteins were reswollen in 20–50 μl of 20 μg trypsin ml−1 (modified sequencing grade; Roche) in 40 mM ammonium bicarbonate, 10% acetonitrile, or 10 μg Lys-C ml−1 (Roche) in 25 mM Tris/HCl, pH 8.0, 1 mM EDTA, or 10 μg Asp-N ml−1 (Roche) in 10 mM Tris/HCl, pH 8.0, at room temperature for 1 h. An additional 50 μl of the respective digestion buffer was then added and digestion continued at 37 °C for 18 h. The supernatant was collected into a fresh tube and peptides were further extracted twice with 50 μl 0.1% trifluoroacetic acid (TFA) at 37 °C for 45 min. Supernatants containing extracted peptides were pooled and stored at −20 °C. Before use in MALDI-TOF MS analyses, the pooled supernatant from each digestion was desalted and concentrated using a C18 ZipTip pipette tip (Millipore) according to the manufacturer’s protocol. The peptides were eluted in 20 μl 60% methanol, 0.1% formic acid and stored at −20 °C.

Secondary proteolytic digestion of peptides. Aliquots (5 μl) of desalted peptides generated using trypsin, Lys-C and Asp-N were dried at 45 °C by centrifugation in a SpeedVac vacuum centrifuge. Peptides were reconstituted in 5 μl enzyme-specific digestion buffers (above) and incubated with second enzymes including 1 μl of 40 μg Asp-N ml−1 (for trypsin- and Lys-C-digested peptides), 0.5 μl of 10 μg Lys-C ml−1 (for Asp-N-digested peptides) or 0.5 μl of 0.5 mg trypsin ml−1 (for Asp-N-digested peptides) at 37 °C for 3 h. The digests were desalted and concentrated using a C18 ZipTip pipette tip and stored at −20 °C.

MS. Peptide samples were diluted 1:1 with 0.1% TFA in 33% acetonitrile and then mixed with an equal volume of matrix solution (0.2 M 2,5-dihydroxybenzoic acid). An aliquot (1 μl) of each sample/matrix mixture was then spotted quickly onto a MALDI target. Spots were allowed to crystallize and dry at room temperature for about 5 min before the target was loaded into a Voyager DE STR MALDI-TOF MS instrument (Applied Biosystems; AB). The instrument was set in positive ion reflector mode with a 165 ns delayed extraction time and 20 kV accelerating voltage. Mass spectra were accumulated from 1000 laser shots and data were collected over a m/z range of 700–4000. Spectra were calibrated externally by using an ABI 4700 Proteomics Analyser Mass Standards kit (AB) containing the known-mass peptides des-Arg1-bradykinin (m/z 904.4560), angiotensin I (m/z 1296.6857), Glu1-fibrinopeptide B (m/z 1570.6758), ACTH (1–17 clip, m/z 2093.0874), ACTH (18–39 clip, m/z 2465.1980) and ACTH (7–38 clip, m/z 3657.9277). Mono-isotopic masses of peptides obtained from the reflector mode of the MALDI-TOF instrument were collected manually. Masses of peptides derived from digested YHV proteins were determined by comparing them with predicted masses calculated from amino acid sequences using the MS Digest software (http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=mdigest), which accommodated protein modifications, including the oxidation of methionine and carbamidomethylation of cysteine. Amino acid sequences of gp116 and gp64 were deduced from the YHV ORF3 gene sequence (GenBank accession no. EF156405) and peptide masses were predicted from in silico digests by trypsin, Lys-C and Asp-N or combinations thereof. MS spectra of peptides derived from untreated and PNGase-F-treated proteins digested with the proteases were also compared. Analyses took into consideration that PNGase-F cleaves asparagine (N)-linked carbohydrate moieties from the peptide backbone, converting asparagine to aspartic acid in the process (Harvey, 2003; Kuster et al., 1997; Tarentino & Plummer, 1994; Tarentino et al., 1985), and thus increases the mass of the modified peptide by 1 Da (Kvaratskhelia et al., 2004).

Glycoprotein staining. Methanol-precipitated proteins (approx. 8 μg) from YHV virions were reconstituted in 10 μl 50 mM sodium phosphate buffer (pH 8.0) and deglycosylated with 2 μl PNGase-F (1 U μl−1) at 37 °C for 3 h. An aliquot (0.5 μl, approx. 2 μg) of the CandyCane glycoprotein molecular mass standards (Invitrogen, Molecular Probes) was treated similarly by dilution with 9.5 μl phosphate buffer and incubated at 37 °C for 3 h with 0.5 μl PNGase-F (1 U μl−1). Untreated and PNGase-F-treated proteins were separated by SDS-PAGE under denaturing conditions in 12.5% gels alongside 40 μg low-range molecular mass standards (Amersham Pharmacia). Carbohydrate was detected by using the Pro-Q Emerald 300 stain (Invitrogen, Molecular Probes) following the manufacturer’s instruction protocol and visualized by using a 300 nm UV
transilluminator equipped with a digital image-documentation system (Bio-Rad).

Identification of glycoprotein glycan types using lectins. Glycans attached to the YHV gp116 and gp64 glycoproteins were detected by using lectins, as described by Casaravilla et al. (2003) with minor modifications. YHV proteins (approx. 10 μg per well) were loaded into each of seven wells of a 12.5 % gel and separated by SDS-PAGE. The masses of the proteins were estimated by comparison with proteins in the ColorBurst prestained protein molecular mass standards (Sigma), full-range Rainbow prestained markers and low-range protein molecular mass markers (both from Amersham Pharmacia). Proteins were transferred to a nitrocellulose membrane (Hybond-C; Amersham Pharmacia) by electroblotting in 25 mM Tris/HCl (pH 8.3), 192 mM glycine, 20 % methanol, 0.1 % SDS at 225 mA constant current for 1 h using a Hoefer TE77 SemiPhor semi-dry blotting system. The membrane was stained briefly with 0.1 % Ponceau red and washed in water to remove excess dye. Gel lane positions on the membrane containing the YHV proteins were cut into individual strips and blocked in TBS [10 mM Tris/HCl (pH 7.4), 150 mM NaCl] containing 0.1 % Tween 20 (TBST) overnight at 4 °C. All biotinylated lectins were purchased from Vector Laboratories Inc. except for HPA, which was purchased from Sigma. Autoradiograms of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim Biophys Acta 1473, 4–8.


