



รายงานวิจัยฉบับสมบูรณ์

โครงการ การแยกยีนและศึกษาการแสดงออกของโปรตีนโครงสร้าง
ขนาด 65 กิโลดาลตันของเชื้อไวรัสหัวเหลือง

โดย ผศ.ดร.ศราวุฒิ จิตรภักดี

เดือน ปี ที่เสร็จโครงการ 30 มิถุนายน 2545

รายงานวิจัยฉบับสมบูรณ์

โครงการ การแยกยีนและศึกษาการแสดงออกของโปรตีนโครงสร้างขนาด 65 กิโลดาลตันของเชื้อไวรัสหัดเหลือง

คณะผู้วิจัย

ผศ.ดร.ศราวุฒิ จิตรภักดี ภาควิชาชีวเคมี คณะวิทยาศาสตร์
มหาวิทยาลัยมหิดล

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย
(ความเห็นในรายงานนี้เป็นของผู้วิจัย สกว. ไม่จำเป็นต้องเห็นด้วยเสมอไป)

กิตติกรรมประกาศ

กระผมขอขอบคุณสำนักงานกองทุนสนับสนุนการวิจัยที่สนับสนุนโครงการนี้มาตลอด 2 ปี ขอขอบคุณ ศาสตราจารย์ ดร.สกล พันธุ์ยิ้ม และศาสตราจารย์ ดร.วิชัย บุญแสง นักวิจัยที่ปรึกษาที่ให้การสนับสนุนโครงการวิจัย รวมทั้งอนุญาตให้กระผมใช้ห้องปฏิบัติการทำการวิจัย ขอขอบคุณทีมงานวิจัยกึ่งซึ่งประกอบด้วย น.ส. ศติมนัส อุณจักร์ และ น.ส. นุสรา สิทธิติลกรรัตน์ นักศึกษาปริญญาโท และนักศึกษาปริญญาเอก ภาควิชาชีวเคมี คณะวิทยาศาสตร์ มหาวิทยาลัยมหิดล ที่ช่วยเหลืองานวิจัยตลอดมา ตลอดจนผู้ร่วมงาน Dr. Peter Walker CSIRO Division of Tropical Agriculture, Queensland ประเทศ ออสเตรเลีย ขอขอบคุณ ดร. อรประไพ คชนันท์ หน่วยปฏิบัติการพันธุวิศวกรรมด้านพืช ศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพ สำหรับความช่วยเหลือในการผลิต polyclonal antibody

ขอขอบคุณคณะวิทยาศาสตร์ มหาวิทยาลัยมหิดล ที่สมทบทุนสนับสนุนนอกเพิ่มเติม เป็นจำนวน 100,000 บาท

Abstract

Project Code: PDF/52/2543

Project Title: Molecular cloning and expression of the 65 kDa Structural Protein of Yellow Head Virus

Investigator: Dr. Sarawut Jitrapakdee
Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok

Email Address: scsji@mahidol.ac.th

Project Period: July 1, 2000-June 30, 2002

Yellow head virus (YHV) is a major agent of disease in farmed penaeid shrimp. YHV virions purified from infected shrimp contain 3 major structural proteins of M_r 116 kDa (gp116), 64 kDa (gp64) and 20 kDa (p20). Two different staining methods indicated that the p116 and p65 proteins are glycosylated. Nucleotide sequence analysis of the ORF3 gene indicated that it encodes a polypeptide of 1,664 amino acids with calculated M_r of 185,713 kDa ($pI = 6.68$). Hydropathy analysis of the deduced ORF3 protein sequence identified 6 transmembrane helices and 3 ectodomains containing multiple sites for 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 Ala²²⁸ and Thr²²⁹, and Ala¹¹²⁷ and Leu¹¹²⁸ located at the C-terminal side of transmembrane helices 3 and 5, respectively. Comparison with the deduced ORF3 protein sequences of Australian gill-associated virus (GAV) indicated 83% identity in gp64 and 71% identity in gp116 which featured 2 significant sequence deletions near the N-terminus. Database searches revealed no significant homology with other proteins. Recombinant gp64 with and without C-terminal transmembrane region expressed in *E. coli* was shown to react with antibody raised against native gp64 purified from virions.

Keywords: Yellow head virus, yellow head disease, gp116, gp64, structural protein

บทคัดย่อ

รหัสโครงการ: PDF/52/2543

ชื่อโครงการ: การแยกและศึกษาการแสดงออกของโปรตีนโครงสร้าง 65 กิโลดาลตันของ เชื้อไวรัสหัดเหลือง

Investigator: ดร.ศราวุฒิ จิตรภักดี
ภาควิชาชีวเคมี คณะวิทยาศาสตร์ มหาวิทยาลัยมหิดล

Email Address: scsji@mahidol.ac.th

ระยะเวลาโครงการ: 1 กรกฎาคม 2543 - 30 มิถุนายน 2545

เชื้อไวรัสหัดเหลืองเป็นสาเหตุสำคัญของโรคหัดเหลืองซึ่งพบในฟาร์มเพาะเลี้ยงกุ้งกุลาดำ เชื้อไวรัสหัดเหลืองที่ทำการแยกให้บริสุทธิ์ประกอบด้วยโปรตีนโครงสร้างหลัก 3 ชนิด คือ gp116, gp64 และ p20 ซึ่งมีขนาด 116, 64 และ 20 กิโลดาลตันตามลำดับ โดยพบว่า gp116 และ gp64 เท่านั้นที่มีการเติมหมู่คาร์โบไฮเดรต จากการหาลำดับนิวคลีโอไทด์ของ ORF3 ทำให้ทราบว่ายีนดังกล่าวสังเคราะห์โปรตีนสายยาวซึ่งมีกรดอะมิโน 1,664 ตัว และมีน้ำหนักโมเลกุล 185 กิโลดาลตันโดยประมาณ (มีค่า pI เท่ากับ 6.68) เมื่อทำการวิเคราะห์หาคุณสมบัติที่ของกรดอะมิโนพบว่าโปรตีนดังกล่าวประกอบไปด้วยส่วนที่เป็น transmembrane 6 บริเวณ ส่วนที่เป็น ectodomain 3 บริเวณ รวมทั้งหลายบริเวณที่สามารถถูกเติมหมู่ น้ำตาลได้ จากการหาลำดับที่ปลายด้าน N ของโปรตีน gp116 และ gp64 ที่สกัดจากเชื้อไวรัสมาเทียบกับ ลำดับกรดอะมิโนที่แปลมาจากลำดับนิวคลีโอไทด์ของ ORF3 ทำให้ทราบว่าโปรตีน gp116 และ gp64 ถูกสังเคราะห์มาจาก precursor โปรตีนตัวเดียวกันซึ่งผ่านกระบวนการตัดแปลงโดยการย่อยด้วยเอนไซม์โปรตีเอสซึ่งตัดพันธะเปปไทด์ระหว่าง Ala²²⁸ กับ Thr²²⁹ และระหว่าง Ala¹¹²⁷ กับ Leu¹¹²⁸ โดยบริเวณดังกล่าวอยู่ที่ปลายด้าน C ของ transmembrane เกลียวที่ 3 และเกลียวที่ 5 ตามลำดับ เมื่อนำลำดับกรดอะมิโนของ gp116 และ gp64 ของเชื้อไวรัสหัดเหลืองไปเทียบกับ ลำดับกรดอะมิโนของ gp116 และ gp64 ของเชื้อ GAV ปรากฏว่ามีความคล้ายคลึงกัน 83% และ 71% โดยประมาณ ความแตกต่างที่เด่นชัดคือใน GAV นั้นมีการเกิด deletion ของลำดับกรดอะมิโนที่ด้าน N อยู่ 2 บริเวณ นอกจากนี้ยังพบว่าลำดับโปรตีน gp 116 และ gp64 ของเชื้อไวรัสหัดเหลืองไม่มีความคล้ายคลึงกับโปรตีนของเชื้อไวรัสตัวอื่น ๆ ในฐานข้อมูล

โปรตีน gp64 ที่ยังมี transmembrane และไม่มี transmembrane ถูกนำมาสังเคราะห์โดยวิธีทางพันธุวิศวกรรมในเชื้อแบคทีเรีย *E. coli* โดยพบว่าโปรตีนที่สร้างขึ้นมาโดยวิธีดังกล่าวสามารถให้ปฏิกิริยาผลบวกกับ antibody ซึ่งเตรียมจากไวรัสที่ทำให้บริสุทธิ์

คำหลัก: เชื้อไวรัสหัดเหลือง, โรคหัดเหลือง, gp116, gp64, structural protein

บทนำ

เชื้อไวรัสหัวเหลี่ยมมีรูปร่างเป็นแท่งคล้ายลูกกระสุนปืน (bullet shape) จึงถูกจัดจำแนกอยู่ในกลุ่มไวรัส Rhabdoviridae (Nadala et al., 1997) การวิเคราะห์สารพันธุกรรมในรายละเอียดต่อมาพบว่าไวรัสชนิดนี้มีสารพันธุกรรมเป็นชนิด RNA สายเดี่ยว (Wongteerasupaya et al., 1995) มีรายงานว่าไวรัสชนิดเดียวกันนี้มีความรุนแรงในการก่อให้เกิดโรคในกิ้งก่าต่างกันไปขึ้นอยู่กับสายพันธุ์ของเชื้อไวรัส เช่นสายพันธุ์ที่พบในประเทศไทยมีความรุนแรงมากกว่าที่พบในประเทศออสเตรเลีย (ซึ่งถูกเรียกว่า gill associated virus, GAV) ความแตกต่างของความรุนแรงที่ก่อให้เกิดโรคอาจเนื่องมาจากความผันแปรทางพันธุกรรมที่พบในเชื้อทั้งสองสายพันธุ์ (Spann et al., 1995; Spann et al., 1997; Cowley et al., 1999) สำหรับ GAV ซึ่งเป็นสายพันธุ์ที่พบในประเทศออสเตรเลียมีสารพันธุกรรมเป็น RNA เส้นเดี่ยวเช่นกันและมีการเรียงตัวของยีนต่าง ๆ คล้ายคลึงกับไวรัสในกลุ่ม coronavirus กล่าวคือ กลุ่มยีนที่เกี่ยวข้องกับการสังเคราะห์เอ็นไซม์ที่เกี่ยวข้องกับกระบวนการเพิ่มจำนวนของสารพันธุกรรมเรียงตัวอยู่ตรงกลางของยีนโนม นอกจากนี้ปลายด้าน 3' ของ ORF1a และปลายด้าน 5' ของ ORF1b ซ้อนเหลื่อมกัน 1 เบส ซึ่งเป็นผลให้เกิดการเลื่อนกรอบการสังเคราะห์โปรตีน (ribosomal frame shift) ส่วนปลายด้าน 3' ของยีนโนมเกี่ยวข้องกับการสังเคราะห์โปรตีนโครงสร้างของไวรัส ตามด้วยปลายหางของ poly(A) [Cowley et al., 2000].

จากการวิเคราะห์อนุภาคไวรัสหัวเหลี่ยมโดยนักวิทยาศาสตร์ที่ฮาวาย สหรัฐอเมริกา พบว่าประกอบไปด้วยโปรตีนโครงสร้าง 4 ชนิดซึ่งมีขนาด 170, 135, 67 และ 22 กิโลดาลตันเมื่อวิเคราะห์ด้วยวิธี SDS-PAGE โดยชิ้นโปรตีนขนาด 135 กิโลดาลตันเป็น glycoprotein (Nadala et al., 1997) อย่างไรก็ตามมีความขัดแย้งขนาดของโปรตีนโดยกลุ่มประเทศไต้หวันพบว่าโปรตีนโครงสร้างเพียง 3 ตัว คือ 110 กิโลดาลตัน 63 กิโลดาลตัน และ 22 กิโลดาลตัน (Wang and Chang, 2000) ยังไม่เป็นที่แน่ชัดว่าทำไมไวรัสสองสายพันธุ์ถึงมีโครงสร้างต่างกัน

ยังไม่มีรายงานการศึกษากลไกที่ทำให้เกิดโรคของเชื้อไวรัสในกิ้งก่าทั้งนี้เนื่องจากขาดความรู้ทางด้านอนุพันธุศาสตร์ของเชื้อไวรัสดังกล่าว โครงการนี้มุ่งเป้าไปที่การแยกยีนสังเคราะห์โปรตีนโครงสร้างของเชื้อไวรัสเนื่องจากลำดับนิวคลีโอไทด์ของยีนดังกล่าวสามารถนำไปใช้เป็นตัวตรวจจับในระดับโมเลกุลเพื่อติดตามหา route of infection รวมถึงกลไกการเพิ่มจำนวนของไวรัสภายในเซลล์ที่ติดเชื้อ นอกจากนี้ยีนที่แยกได้สามารถนำไปผลิตโปรตีนเพื่อใช้พัฒนาเป็นวัคซีนรวมถึงการผลิตแอนติบอดีสำหรับการใช้การศึกษากระบวนการเพิ่มจำนวนของเชื้อไวรัสในเซลล์ได้อีกด้วย

วิธีการทดลอง

การเตรียมเชื้อไวรัสให้บริสุทธิ์

ทำการฉีดเชื้อไวรัสเข้าสู่กุ้งกุลาดำ *Penaeus monodon* ใต้กล้ามเนื้อผ่านปล้องขาคู่ที่ 2 จากนั้นทำการเก็บ haemolymph จากกุ้งเมื่อสิ้นสุดวันที่ 3 เชื้อไวรัสถูกทำให้บริสุทธิ์ผ่านขั้นตอน gradient ultracentrifugation ตามวิธีของ Wongterasupaya et al., 1995 เชื้อไวรัสที่ได้นำไปเก็บเป็น stock โดยแช่แข็งที่ -80°C

การแยกโปรตีนด้วยวิธี SDS-PAGE และการหาลำดับกรดอะมิโนที่ปลายด้าน N

ตัวอย่างโปรตีนจากอนุภาคไวรัส (10 μg โดยประมาณ) ถูกนำมาวิเคราะห์โดยวิธี SDS-PAGE (Laemmli et al., 1970) แบบ reducing โดยแยกผ่าน 7.5% หรือ 15% โปรตีนที่แยกได้ถูกย้ายขึ้นบนแผ่น polyvinylidene difluoride (PVDF) โดยใช้เครื่อง semidry blotting ผ่านบัฟเฟอร์ซึ่งมีส่วนประกอบของ 10 mM CAPS/10% methanol เป็นเวลา 1-2 ชั่วโมง แถบโปรตีนบนแผ่น PVDF ถูกนำไปย้อมด้วย 0.3 %Coomassie blue R250 จากนั้นทำการตัดแถบโปรตีนออกจากแผ่น PVDF เพื่อนำไปหาลำดับที่ปลายด้าน N โดยวิธี Edman degradation การวิเคราะห์ดังกล่าวทำโดยหน่วยบริการชีวภาพ ศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพ

การผลิต polyclonal antibody ต่อโปรตีน gp116 และ gp64 จากหนู

ทำการแยกโปรตีน gp116 และ gp64 ปริมาณ 500 μg บน SDS-PAGE จากนั้นย้ายโปรตีนขึ้นไปบน nitrocellulose ด้วยวิธี Western blot จากนั้น ตัด band ขนาด 116 kDa (gp116) และ 64 kDa (gp64) ออกจาก membrane แล้วบดให้เป็นผงละเอียดในไนโตรเจนเหลว จากนั้นละลายในฟอสเฟตบัฟเฟอร์ (140mM NaCl, 4mM KCl, 2mM KH_2PO_4 , 8mM Na_2HPO_4 , pH 7.4) 0.5 ml จากนั้นผสมให้เข้ากับ 0.5 ml complete Freund's Adjuvant (Sigma) แล้วฉีดผ่านช่องท้อง (intraperitoneal injection) ของหนู mice BALB/C 2 ตัว ทั้งไว้ 1 อาทิตย์ จากนั้นฉีดโปรตีนดังกล่าวกระตุ้นซ้ำ 2 ครั้งห่างกัน 1 อาทิตย์ (ใช้ปริมาณ antigen 100 microgram ต่อการกระตุ้น 1 ครั้ง) ทำการเจาะเลือดหนูเพื่อเก็บ serum สำหรับทดสอบปฏิกิริยาทางภูมิคุ้มกัน (immunological reaction) กับ โปรตีนขนาด 64 kDa ที่แยกด้วยวิธี SDS-PAGE

การวิเคราะห์ด้วยวิธี Western immunoblot

โปรตีนที่แยกผ่าน SDS-PAGE ถูกย้ายขึ้นบนแผ่น nitrocellulose ผ่านบัฟเฟอร์ 0.25 M glycine, 0.4% SDS, 25 mM Tris pH 8.3 และ 20% methanol โดยเครื่อง semidry blotter หลังจากนั้นทำการ block membrane ในสารละลาย 3% bovine serum albumin, 0.05% Tween 20 ในสารละลาย ฟอสเฟตบัฟเฟอร์ เป็นเวลา 3 ชั่วโมง จากนั้นล้างแผ่น nitrocellulose ในสารละลายฟอสเฟตบัฟเฟอร์ดังกล่าวครั้งละ 5 นาที 3 ครั้ง จากนั้นนำแผ่น nitrocellulose แช่ในแอนติบอดีที่ความเข้มข้น ต่าง ๆ (ใช้สารละลายฟอสเฟตบัฟเฟอร์เป็นตัว dilute ในปริมาณ

10 ml) เป็นเวลา 1 ชั่วโมง จากนั้นล้างแผ่น nitrocellulose ด้วยสารละลายฟอสเฟตบัฟเฟอร์ครั้งละ 5 นาที 3 ครั้ง จากนั้นแช่แผ่น nitrocellulose ในสารละลายฟอสเฟตบัฟเฟอร์ 10 ml และ 1 μ l goat anti-mouse polyclonal antibody เป็นเวลา 1 ชั่วโมง จากนั้นล้าง secondary antibody ออกด้วยสารละลายฟอสเฟตบัฟเฟอร์ 2 ครั้งก่อนนำไปแช่ในสารละลาย 35 mg/ml NBT และ 75 mg/ml BCIP ในที่มีดจนปรากฏแถบของ immunoreactive

การย้อม glycoprotein

ทำการย้อมโปรตีนโดยใช้ชุดสำเร็จรูป ECL glycoprotein detection system (Amersham Pharmacia) โปรตีนที่ย้ายขึ้นบนแผ่น PVDF ถูกนำไปทำปฏิกิริยากับสารละลาย 10 mM sodium metaperiodate ในสารละลาย acetate buffer pH 5.5 เป็นเวลา 20 นาที จากนั้นนำไปทำปฏิกิริยากับ สารละลาย 0.35 mM biotin hydrazide ในสารละลาย acetate buffer pH 5.5 เป็นเวลา 1 ชั่วโมง จากนั้นนำไปทำไปแช่ในสารละลาย streptavidin-alkaline phosphatase ก่อนทำปฏิกิริยากับ NBT และ BCIP นอกจากนี้ยังได้ย้อม glycoprotein โดยใช้ Thymol staining ตามวิธีของ Racusen et al. 1979

การแยกยีนโดยวิธี reverse transcriptase polymerase chain reaction (RT-PCR) การโคลน และหาลำดับนิวคลีโอไทด์

สกัด RNA จากไวรัสที่แยกได้โดยใช้ชุดสกัด RNA Trizol Reagent (Invitrogen) นำ RNA ที่สกัดได้มาละลายในน้ำปลอด RNAase จากนั้นทำการแยกยีนโดยการทำให้ RT-PCR โดยใช้ชุดสำเร็จรูป Superscript one-step RT-PCR system (Invitrogen) ในปริมาตร 25 μ l ซึ่งประกอบด้วย 1X reaction buffer (1.4 mM of each dNTP, 2.4 mM MgSO₄), 0.5 μ l SuperScript II RT/Taq mix, 0.3 μ l RNase Inhibitor, 2 μ l 5 mM MgSO₄, 0.75 μ l Elongase, 0.4 mM forward primer (5'-gatcggggtacctaagcttatgctatcgacct-3') ซึ่งออกแบบจากลำดับนิวคลีโอไทด์ที่ปลายด้าน 3' ของ ORF1b gene, oligo d(T) primer (5'-tctagaggatccccgggtacc(t)₂₀-3') และ RNA 1 μ g โดยปฏิกิริยาดังกล่าวประกอบด้วยการสร้างสาย cDNA ที่ 50°C 30 นาที ตามด้วย 94°C 2 นาที จากนั้นเข้าสู่ขั้นตอน PCR ด้วยการ denaturation ที่ 94°C, 30 sec; annealing ที่ 52°C, 30 sec และ extension ที่ 68°C 8 min เป็นจำนวน 35 รอบ จากนั้นนำผลผลิต PCR มาวิเคราะห์โดยวิธี agarose electrophoresis จากนั้นตัดแถบ DNA และสกัด agarose ออกโดยใช้ชุดสำเร็จรูป QIAquick Gel Extraction kit (Qiagen) DNA ที่สกัดได้ถูกนำไปหาลำดับนิวคลีโอไทด์โดยตรงโดยใช้ชุดหาลำดับ BIG Dye™ reagent (ABI, U.S.A.) ลำดับนิวคลีโอไทด์ที่อ่านได้ในแต่ละรอบของการหาลำดับจะถูกใช้ออกแบบเป็นไพรเมอร์ตัดใหม่เพื่อหาลำดับนิวคลีโอไทด์ทั้งหมด

การแสดงออกของ gp64 ในแบคทีเรีย *E. coli*

ทำการโคลนชิ้นส่วน cDNA สำหรับโปรตีน gp64 ด้วยวิธี PCR โดยออกแบบไพรเมอร์จากนิวคลีโอไทด์ที่ตรงกับ L¹¹²⁸-V¹¹³⁵ (YHV-D7

5'-gcctctagacatagctcgcctccacgacaggcagctgtt-3', introduce initiation methionine เพื่อเริ่ม translation initiation และมีลำดับ NdeI restriction site ที่ปลายด้าน 5' เพื่อสามารถโคลนเข้า pET17b expression vector) และไพรเมอร์ YHV-D8 (5'-cattgtggatcctcactagtgtgatgatgatgatgatggtggatcggttggttcgttctcatggacgt-3') ซึ่งออกแบบจาก T¹⁶⁵⁷-S¹⁶⁶⁶ ตามด้วย stop codon; จากนั้นเพิ่มจำนวนด้วยวิธี PCR ซึ่งทำขึ้นในปริมาตร 25 ml ประกอบด้วย 1x PCR buffer (10 mM Tris-HCl, pH8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.1% Triton X-100), 0.2 mM dNTP, 1 ng oligo primed cDNA, 0.25 mM each primer และ 2 U Taq polymerase ปฏิกริยา PCR ประกอบด้วย denaturation ที่ 94°C 30 วินาที , annealing/extension ที่ 68°C 2 นาที ตามด้วย final extension ที่ 72°C 10 นาที ผลผลิต PCR ที่ได้ถูกนำมาตัดด้วย NdeI และ EcoRI และต่อเข้ากับ pET17b ที่ตัดด้วยเอ็นไซม์ตัวเดียวกัน

สำหรับการโคลนชิ้นส่วน cDNA ของ gp64 ซึ่งไม่มี transmembrane สร้างขึ้นโดยวิธี PCR เช่นกัน โดยใช้ไพรเมอร์ YHV-D7 และ reverse primer จาก residues F¹⁶²⁰-D¹⁶²⁶ (YHV-D14, 5'-gaattctcactaatcccatgtcttgccgccgaa-3') ผลผลิต PCR ที่ได้ถูกโคลนเข้า pDrive vector (Qiagen) และทำการตรวจสอบลำดับนิวคลีโอไทด์ก่อนตัดด้วย NdeI และ EcoRI เพื่อโคลนเข้าสู่ pET17b expression vector

Expression construct ที่สร้างได้ถูกนำเข้าสู่แบคทีเรีย *E. coli* BL21(DE3) [Novagen] จากนั้นทำการเหนี่ยวนำให้แบคทีเรียสร้างโปรตีน gp64 โดยการเลี้ยงในอาหารเลี้ยง LB ซึ่งมี ampicillin เข้มข้น 100 µg/ml จนได้ค่า A₆₀₀ เท่ากับ 0.5 จึงทำการเติม IPTG ให้ได้ความเข้มข้นสุดท้ายเป็น 1 mM เลี้ยงแบคทีเรียต่อไปเป็นเวลา 6 ชั่วโมงจึงทำการเก็บแบคทีเรียเพื่อวิเคราะห์หาการแสดงออกของโปรตีน gp64

ผลการทดลอง

การศึกษาคุณสมบัติทางเคมีของโปรตีนโครงสร้าง gp 64

ทำการฉีดเชื้อไวรัสหัวเหลืองเข้าไปในกึ่งกลางดำและทิ้งไว้ 2-3 วันเพื่อให้กึ่งกลางแสดงอาการ การติดเชื้อ จากนั้นทำการเก็บ hemolymph และปั่นแยกไวรัสโดยวิธี ultracentrifugation โดยผ่านตัวกลาง urograffin นำไวรัสที่สกัดได้มาจากวิเคราะห์หาขนาดและจำนวนของโปรตีนโครงสร้างด้วยวิธี SDS-PAGE (12% separating gel/4% stacking gel) ดังแสดงในรูปที่ 1A พบว่าโปรตีนโครงสร้างมีทั้งหมด 3 ชนิดซึ่งมีขนาด 116, 64 และ 22 kDa โดยประมาณโดยให้ชื่อว่า gp116, gp64 และ p20 ตามลำดับ และเมื่อทำการแยกโปรตีนโครงสร้างในสภาวะ reducing ก็พบว่ายังคงพบโปรตีน 3 ขนาดเช่นเดิม แสดงให้เห็นว่าโปรตีนโครงสร้างทั้ง 3 ชนิดนี้ไม่ได้เกาะเกี่ยวกันด้วย disulfide bond ได้ทำการหาลำดับกรดอะมิโนที่ปลายด้าน N โดยวิธี Edman degradation ของโปรตีนขนาด 116 kDa และ ขนาด 64 kDa พบว่ามีลำดับคือ T-I-L-S-G-I-P-E-K-D- และ L-A-P-R-Q-A-R-V-R-G- ตามลำดับ

จากนั้นทำการทดสอบการเติมหมู่น้ำตาลที่โปรตีนดังกล่าวโดยใช้ชุดตรวจสอบ ECL glycoprotein detection system (Amersham) พบว่า gp116 และ gp64 มีการเติมหมู่น้ำตาลดังรูปที่ 1B นอกจากนี้ยังได้ยืนยันผลการทดลองโดยใช้วิธี Thymol staining ซึ่งพบว่าให้ผลเช่นเดียวกัน ซึ่งการมีหมู่คาร์โบไฮเดรตในโปรตีน gp116 และ gp64 (รูปที่ 1C) นับว่าเป็นคุณสมบัติที่ดีที่จะกระตุ้นให้มีการสร้าง antibody ใน ขั้นตอนถัดไป

การผลิต polyclonal antibody ต่อโปรตีนโครงสร้าง gp64

ทำการฉีดโปรตีนโครงสร้าง gp64 เข้าไปในหนูขาว 2 ตัว และเจาะเลือดเก็บ serum มาทำการทดสอบทางภูมิคุ้มกันกับโปรตีน gp116 และ gp64 ที่สกัดจากไวรัส

ผลการทดลองแสดงในรูปที่ 2 กล่าวคือ serum จากหนูทั้งสองตัวทำปฏิกิริยากับโปรตีนขนาด 65 kDa ในความไวที่สูงกล่าวคือ ที่ dilution 1:25,000 ของ serum ของหนู ทั้งสองตัวทำปฏิกิริยากับโปรตีน gp64 ในระดับนาโนกรัม ให้ผลชัดเจนและไม่ทำปฏิกิริยากับโปรตีน gp116 อย่างไรก็ตามพบว่า serum จากหนูตัวที่ 2 ให้ความไวสูงกว่าหนูตัวที่ 1 โดยสามารถเก็บ serum ได้ทั้งหมด 0.5 ml ต่อหนู 1 ตัว

polyclonal antibody ที่เตรียมได้จากการทดลองในขั้นนี้จะถูกใช้ในการตรวจหาการแสดงออกของยีนสังเคราะห์โปรตีน gp64 สำหรับการทำให้ Recombinant Expression ในขั้นตอนถัดไป

ลำดับนิวคลีโอไทด์ของโปรตีน gp64

เป็นไปได้อย่างสูงที่ยีนสังเคราะห์โปรตีนโครงสร้างของเชื้อไวรัสหัวเหลืองจะเรียงตัวอยู่ที่ปลายด้าน 3' ของยีนเหมือนกับของไวรัส GAV เนื่องจากมีรายงานว่า การเรียงตัวของ ORF1b ก็มีการเรียงตัวอยู่ในตำแหน่งเดียวกันเช่นกัน (ดูรูปที่ 3) ดังนั้นจึงได้ออกแบบไพรเมอร์

จากปลายด้าน 3' ของ ORF1b และปลายด้าน poly(A) tail เพื่อใช้เพิ่มจำนวนยีนโดยวิธี RT-PCR จากการทดลองมีการเพิ่มผลผลิต PCR ขนาด 6.0 kb ซึ่งขนาดใกล้เคียงกับที่คาดเดาจากการเรียงตัวของยีน

การหาลำดับนิวคลีโอไทด์พร้อมทั้งการแปลลำดับกรดอะมิโนของชิ้นส่วนดังกล่าวแสดงให้เห็นว่า ORF ที่ใหญ่ที่สุดโดยให้ชื่อว่า ORF3 ประกอบด้วยนิวคลีโอไทด์ทั้งหมด 4998 ตัวซึ่งแปลให้ลำดับกรดอะมิโนจำนวน 1,666 ตัว การวิเคราะห์ลำดับนิวคลีโอไทด์ 200 ตัวก่อนหน้า initiation codon โดยใช้ฐานข้อมูล TRSEARCH และ Sigscan ไม่พบลำดับ TATA ซึ่งเป็นลำดับเฉพาะตัวของ eukaryotic promoter อย่างไรก็ตามพบลำดับซึ่งคล้ายคลึงกับ TATA คือ TATA-like (TAAAT) ที่ตำแหน่ง -29/-24 นอกจากนี้ยังพบ inverted CAAT box ที่ตำแหน่ง -49/-37 ลำดับนิวคลีโอไทด์พร้อมทั้งกรดอะมิโนที่ถูกแปลถูกแสดงไว้ในภาพที่ 4 ORF3 แปลรหัสให้โปรตีนประกอบด้วยกรดอะมิโนจำนวน 1,666 ตัว ซึ่งมีน้ำหนักโมเลกุล 185,713 ดาลตัน และค่า pI เท่ากับ 6.68 เมื่อเทียบลำดับกรดอะมิโนที่ปลายด้าน N ของ gp116 และ gp64 ซึ่งได้โดยวิธี Edman degradation กับลำดับกรดอะมิโนที่แปลจากลำดับนิวคลีโอไทด์ทำให้ทราบว่าโปรตีน gp116 และ gp64 ถูกสร้างจาก precursor ตัวเดียวกันผ่านกระบวนการ proteolysis ซึ่งจะตัดโปรตีน precursor ระหว่าง A²²⁸ กับ T²²⁹ และระหว่าง A¹¹²⁷ กับ L¹¹²⁸ เป็นผลให้ได้โปรตีน 3 ชนิดคือ gp116 ซึ่งมีน้ำหนักโมเลกุลประมาณ 101 kDa gp64 ซึ่งมีน้ำหนักโมเลกุลเท่ากับ 58 kDa และโปรตีนอีกตัวหนึ่งซึ่งยังไม่ทราบหน้าที่มีขนาด 25.4 kDa ขนาดโปรตีนที่คำนวณได้โดยเฉพาะ gp116 และ gp64 มีขนาดใกล้เคียงกับที่พบในอนุภาคไวรัส ความแตกต่างของน้ำหนักโมเลกุลน่าจะมาจากกระบวนการเติมหมู่คาร์โบไฮเดรต (glycosylation) เป็นผลให้มีขนาดใหญ่ขึ้น

การวิเคราะห์โครงสร้างของลำดับกรดอะมิโนที่แปลจาก ORF3 ผ่านฐานข้อมูล <http://www.sosui.proteome.bio.taut.ac.jp> ทำให้ทราบว่าโปรตีนดังกล่าวเป็น transmembrane protein ประกอบด้วยบริเวณที่เป็นเกลียว α -helices 6 บริเวณโดย พบ 3 เกลียวที่บริเวณปลายด้าน N (ตำแหน่ง L²⁵-L⁴⁷, Y¹³⁸-F¹⁵⁵, F²⁰⁸-I²³⁰) 2 เกลียวตรงกลาง (G¹⁰²⁷-N¹⁰⁵⁹) และ 1 เกลียวที่ปลายด้าน C (K¹⁶³⁰-L¹⁶⁵²)

การวิเคราะห์ลำดับกรดอะมิโนโดยใช้โปรแกรม TMHMM (Krogh et al., 2001) สามารถทำนายได้ว่าปลายด้าน N ของโปรตีนยื่นเข้าไปในเซลล์ผ่านเกลียว helix อันแรก ตามด้วยส่วนที่ยื่นออกไปเป็น ectodomain ย้อนไปมาโดยปลายด้าน C ยื่นกลับเข้ามาในเซลล์ ดังรูปที่ 5

การวิเคราะห์ตำแหน่งที่สามารถเกิด O-linked glycosylation พบว่าสามารถเกิดได้ที่ threonine ตำแหน่งที่ 255, 414, 569, 807, 828, 1585 และที่ serine ตำแหน่งที่ 70, 232, 413, 417, 576 และ 832 นอกจากนี้ยังมี 12 บริเวณที่สามารถเกิด N-linked glycosylation (ดูรูปที่ 5) การค้นหาความคล้ายคลึงกันของลำดับกรดอะมิโนของโปรตีนที่สร้างจาก ORF3 นั้นไม่พบความคล้ายคลึงกับโปรตีนต่าง ๆ ในฐานข้อมูล GenBank/EMBL, SWISS-PROT และ PIR เมื่อค้นหาด้วยโปรแกรม FASTA, TFASTA, BLAST อย่างไรก็ตามลำดับกรดอะมิโนที่แปลจาก ORF3

ของไวรัสหัวเหลี่ยมแสดงความคล้ายคลึงกับโปรตีนตัวเดียวกันกับขอไวรัส GAV คือ gp116 ของไวรัสหัวเหลี่ยมมีความคล้ายคลึงกับของ GAV 71% ขณะที่ gp64 ของไวรัสหัวเหลี่ยมมีความคล้ายคลึงกับของ GAV 83% ความคล้ายคลึงกันของโปรตีนทั้งสองตัวจากไวรัส 2 ชนิดนี้แสดงให้เห็นในรูปที่ 6

การแสดงออกของโปรตีน gp64 ในแบคทีเรีย *E. coli*

ชิ้นส่วน cDNA สังเคราะห์โปรตีนโครงสร้าง gp64 ถูกสร้างขึ้นโดยเริ่มจาก L¹¹²⁸ ไปจนถึง V¹¹³⁵ ด้วยวิธี PCR จากนั้นโคลนเข้าสู่ pET17b ทำให้ได้ plasmid ชื่อว่า pET17b-gp64 ก่อน transform เข้าสู่ *E. coli* BL21(DE3) ทำการเหนี่ยวนำให้แบคทีเรียสร้างโปรตีนแล้ววิเคราะห์ด้วยวิธี SDS-PAGE ดังแสดงให้ดูในรูปที่ 7A เมื่อทำการย้อมโปรตีนด้วย Coomassie blue ไม่พบว่าแบคทีเรียที่มีพลาสมิด pET17b-gp64 และถูกเหนี่ยวนำด้วย IPTG มีการสร้างโปรตีนขนาด 58 kDa แต่อย่างไรถึงแม้จะเหนี่ยวนำเป็นเวลา 7 ชั่วโมง อย่างไรก็ตามเมื่อวิเคราะห์โดยวิธี Western immunoblot โดยใช้ polyclonal antibody ที่เตรียมได้จากหนูพบว่าสามารถมองเห็นแถบโปรตีนขนาด 58 kDa ที่ทำปฏิกิริยากับ antibody ดังกล่าว นอกจากนี้ยังพบว่าแถบโปรตีนขนาด 58 kDa ดังกล่าวมีปริมาณลดลงเมื่อเวลาผ่านไปนานกว่า 1 ชั่วโมง แสดงว่าโปรตีนดังกล่าวไม่เสถียร ขนาดของโปรตีนดังกล่าวมีขนาดสอดคล้องกับที่คำนวณจากการแปลลำดับนิวคลีโอไทด์แต่มีขนาดเล็กกว่าขนาดโปรตีนที่สกัดได้จากไวรัส (58 kDa vs 64 kDa) เชื่อว่าความแตกต่างนี้น่าจะมาจากขบวนการ glycosylation ซึ่งไม่พบใน *E. coli*

การที่โปรตีน gp64 ขาดความเสถียรเมื่อถูกผลิตขึ้นใน *E. coli* เชื่อว่าน่าจะมีส่วนมาจาก transmembrane ที่ปลายด้าน C เป็นที่รู้กันว่าโปรตีนที่เป็น transmembrane นั้นอาศัยระบบ endoplasmic reticulum ที่จะส่งโปรตีนไปที่เยื่อหุ้มเซลล์ ด้วยสมมุติฐานนี้จึงได้ทำการตัด transmembrane ที่ปลายด้าน C ทั้ง (Y¹⁶²⁷-S¹⁶⁶⁶) โดยการทำให้ PCR ซึ่งใช้ reverse primer ตรงกับ F1620-D1626 ทำให้สามารถกำจัดปลายด้าน C ซึ่งมีส่วนของ transmembrane ทิ้งไป จากนั้นโคลนชิ้นส่วนเข้าสู่ pET17b ทำให้ได้พลาสมิด pET17b-gp16 Δ TM ก่อนนำเข้าสู่ *E. coli* BL21(DE3) เช่นเดียวกัน

หลังจากเหนี่ยวนำให้มีการสร้างโปรตีนพบว่าแบคทีเรียที่มีพลาสมิด pET16b-gp16 Δ TM ยังคงไม่ผลิตโปรตีน gp64 ชัดเจนเมื่อวิเคราะห์ด้วย SDS-PAGE อย่างไรก็ตามเมื่อวิเคราะห์ด้วย Western immunoblot พบแถบโปรตีนขนาด 55 kDa (มีขนาดเล็กลงเนื่องจากส่วนที่เป็น transmembrane หายไป) ทำปฏิกิริยากับ polyclonal antibody อย่างชัดเจนโดยความเข้มของแถบโปรตีน 55 kDa มีความเข้มมากกว่าแถบโปรตีนขนาด 58 kDa ที่ผลิตจากพลาสมิด pET17b-gp64 ทำให้สรุปได้ว่าการกำจัดส่วน transmembrane region สามารถปรับปรุงประสิทธิภาพการแสดงออกของยีนดังกล่าวใน *E. coli*

วิจารณ์ผลการทดลอง

โปรตีนโครงสร้างของเชื้อไวรัสในกลุ่ม Nidovirus ที่ infect สัตว์เลี้ยงลูกด้วยนม มีหน้าที่เกี่ยวข้องกับการเกาะติดของไวรัสกับเซลล์ การเข้าเซลล์ ตลอดจนเกี่ยวข้องกับการปฏิกริยาทางภูมิคุ้มกัน ในไวรัสบางชนิดโปรตีนโครงสร้างมีบทบาทสำคัญกับความรุนแรงของโรค อย่างไรก็ตามมีข้อมูลน้อยมากเกี่ยวกับบทบาทของโปรตีนโครงสร้างของไวรัสในกลุ่ม Nidovirus ที่ infect สัตว์ไม่มีกระดูกสันหลังโดยเฉพาะที่ infect กุ้ง ในโครงการนี้ได้ทำการศึกษาทางอณูชีววิทยาของโปรตีนโครงสร้างของเชื้อไวรัสหัวเหลืองก่อโรคในกุ้งกุลาดำ โดยพบว่าโปรตีนโครงสร้างหลักประกอบด้วยโปรตีน 3 ตัว คือ gp116, gp64 และ p20 ซึ่งถูกสังเคราะห์จาก ORF3 gene จากนั้นถูกตัดแปลงโดยขบวนการ proteolysis

Nadala และคณะได้ทำการศึกษาโปรตีนโครงสร้างของไวรัสหัวเหลืองเป็นครั้งแรกและพบว่าโปรตีนดังกล่าวมีอยู่ 4 ตัวซึ่งมีขนาด 200 kDa, 135 kDa, 62 kDa และ 20 kDa โดยพบว่าโปรตีนขนาด 135 kDa เท่านั้นที่เป็น glycoprotein ต่อมา Wang และ Chang (2000) ได้รายงานโปรตีนโครงสร้างของเชื้อไวรัสหัวเหลืองสายพันธุ์ประเทศไต้หวันว่ามีจำนวน 3 ตัว มีขนาด 110 kDa, 63 kDa และ 20 kDa ผลการทดลองจากโครงการฉบับนี้มีความใกล้เคียงกับข้อมูลที่ได้จาก Wang และ Chang นอกจากนี้ผู้วิจัยยังได้แสดงให้เห็นว่าเฉพาะโปรตีนขนาดใหญ่เท่านั้นที่มีการ glycosylation

การนำลำดับกรดอะมิโนที่ปลายด้าน N ของโปรตีน gp116 และ gp64 ที่สกัดจากเชื้อไวรัสหัวเหลืองมาเทียบกับลำดับกรดอะมิโนที่แปลจากลำดับนิวคลีโอไทด์ของ ORF3 gene ทำให้สามารถหากรอบการอ่านโปรตีน (open reading frame) จากยีนดังกล่าวได้แม่นยำกล่าวคือ ORF3 แปลรหัสให้โปรตีนซึ่งประกอบด้วยกรดอะมิโน 1,666 ตัว ขนาดของโปรตีนซึ่งประกอบด้วยกรดอะมิโนจำนวนมากเช่นนี้พบใน vertebrate coronavirus เช่นกัน ได้แก่ avian infectious bronchitis virus (กรดอะมิโน 1,660 ตัว) [Binns et al., 1985], feline infectious peritonitis virus (กรดอะมิโน 1,452 ตัว) [de Groot et al., 1987], murine hepatitis virus (กรดอะมิโน 1,376 ตัว) [Spann et al., 1988], porcine epidemic diarrhea virus (กรดอะมิโน 1,383 ตัว) [Duarte and Laude, 1994] หรือ human respiratory coronavirus (กรดอะมิโน 1,353 ตัว) [Mounir and Talbot, 1993] อย่างไรก็ตามจากการทำนายโครงสร้างของโปรตีนโครงสร้างไวรัสหัวเหลืองนั้นมีความสลับซับซ้อนมากกว่า คือประกอบไปด้วยส่วนที่เป็น transmembrane 6 บริเวณ ส่วนที่เป็น ectodomain 3 บริเวณ และส่วนที่เป็น intra-virion 2 บริเวณ ซึ่งเมื่อเทียบกับโปรตีนโครงสร้างของ coronavirus จะพบว่าส่วนที่เป็น transmembrane นั้นมีเพียง 1 บริเวณเท่านั้นซึ่งอยู่ที่ปลายด้าน C

โปรตีนเปลือกหุ้มของ coronavirus ซึ่ง infect vertebrate นั้นประกอบไปด้วย glycoprotein ขนาดประมาณ 180 kDa (S) โดยพบว่าโปรตีนตัวตั้งต้นนี้ถูกตัดโดย protease ให้โปรตีน 2 ขนาดใกล้เคียงกันคือ S1 และ S2 ซึ่งยึดติดกันด้วย non-covalent interaction (Sturman et al., 1985; Cavanagh, 1995; Snijder and Horzinek, 1995) คล้าย ๆ กับโปรตีน

ของไวรัสหัดเหลืองคือถูกสร้างในรูปของโปรตีนตั้งต้นขนาด ~185 kDa ซึ่งถูกตัดด้วย protease จนได้โปรตีน 2 ตัวคือ gp116 และ gp64 ถึงแม้โปรตีนทั้งสองตัวจะมีกรดอะมิโน cysteine จำนวนมากแต่ก็ไม่ได้เกิด interdisulfide bond เป็นไปได้ว่า gp116 และ gp64 อาจติดกันด้วย non-covalent bond เหมือนกัน S1 และ S2 subunit ของ coronavirus โดยใน coronavirus นั้น S2 มีโครงสร้างเป็นก้านยื่นออกไปจาก membrane ขณะที่ S1 form โครงสร้างส่วนหัวของโปรตีนหนามซึ่ง interact กับ receptor ของ host cell

การที่ oligo(dT) สามารถเกาะกับปลายด้าน 3' ของอีโนม แสดงว่าปลายด้าน 3' ของ YHV มีลักษณะเป็น poly(A) ซึ่งลักษณะดังกล่าวเป็นลักษณะของ RNA ไวรัสซึ่งมีสารพันธุกรรมเป็นเส้นบวก การที่ ORF3 เรียงตัวด้าน 3' ของกลุ่มยีนที่เกี่ยวข้องกับการเพิ่มจำนวนก็เป็นลักษณะอีกประการของไวรัสในกลุ่ม nidovirus (Sithidilokratana et al., 2002)

ขนาดของ gp116 และ gp64 ที่คำนวณได้จากการแปลรหัสนิวคลีโอไทด์มีขนาดเล็กกว่ามีพบจากอนุภาคไวรัส ความแตกต่างนี้เชื่อกันว่าจะมาจากขบวนการเติมคาร์โบไฮเดรต (glycosylation) ของโปรตีนแต่ละตัว ยีนสังเคราะห์ gp64 ถูกนำมาต่อเข้ากับพลาสมิดเพื่อทดสอบการแสดงออกใน *E. coli* โดยพบว่าท่อนยีนดังกล่าวสามารถแสดงออกใน *E. coli* โดยเห็นได้จากการที่ polyclonal antibody ต่อโปรตีน gp64 สามารถทำปฏิกิริยากับโปรตีนขนาด 58 kDa ที่ผลิตจาก *E. coli* ถึงแม้ระดับการแสดงออกจะต่ำ อย่างไรก็ตามระดับการแสดงออกของ gp64 สามารถเพิ่มขึ้นได้เมื่อทำการตัดส่วนที่เป็น transmembrane ที่ปลายด้าน C ทิ้งไป

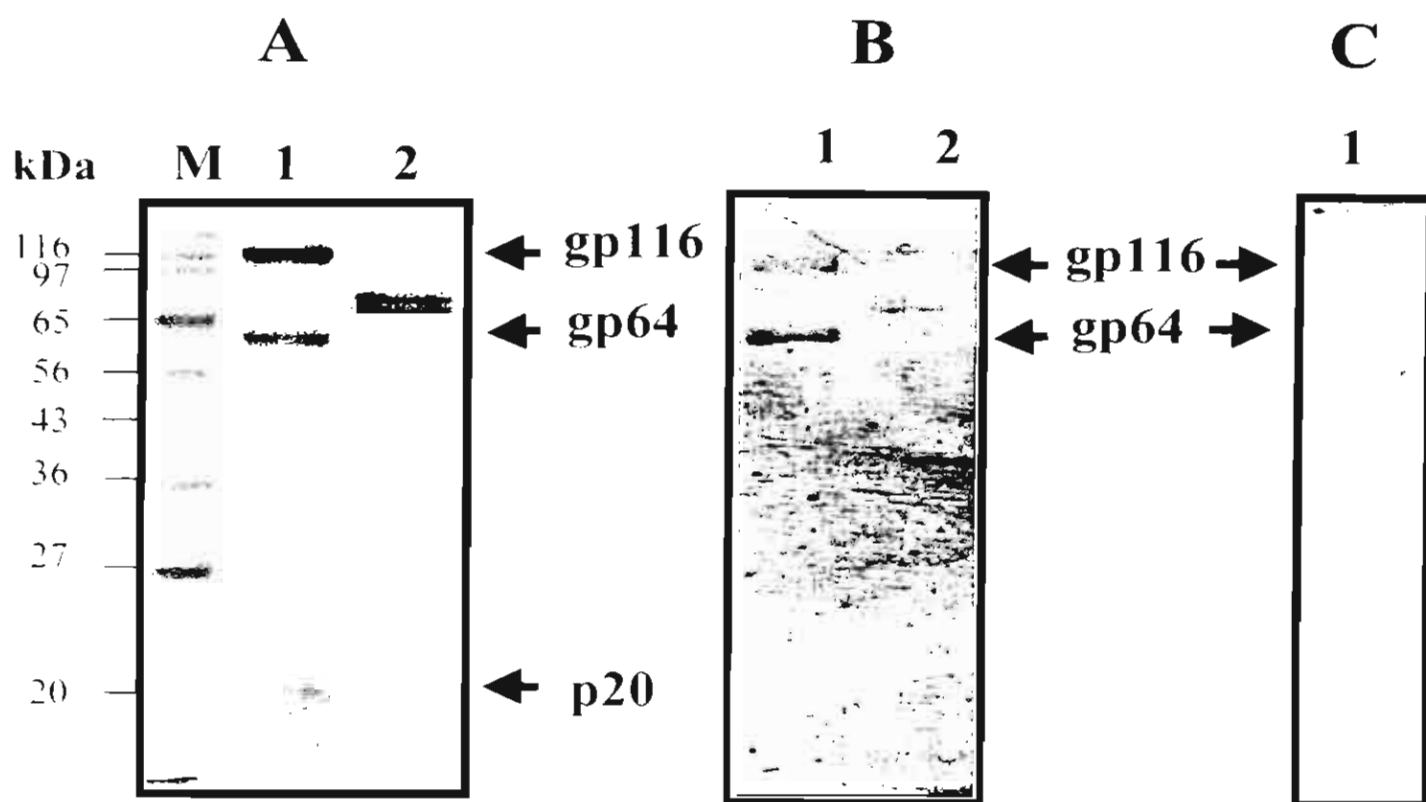
ในอนาคตระบบการแสดงออกโดยใช้ insect expression system หรือ ยีสต์ ซึ่งมีระบบ endoplasmic reticulum หรือ glycosylation น่าจะช่วยให้เราสามารถผลิตโปรตีนได้ในปริมาณเพิ่มขึ้นและมีการเติมหมู่คาร์โบไฮเดรต

หนังสืออ้างอิง

- Binns, M.M., Boursnell, M.E.G., Cavanagh, D., Pappin, D.J.C. and Brown
(1985). Cloning and sequencing of the gene encoding the spike protein of the coronavirus IBV. *J. Gen. Virol.* **66**, 719-726.
- Cavanagh, D. (1995). The coronavirus surface glycoprotein. *In: The Coronaviridae*. S.G. Siddell (ed.) Pp. 73-113. Plenum Press, New York.
- Cowley, J.A., Dimmock, C.M., Wongteerasupaya, C., Boonsaeng, V., Panyim, S. and Walker, P.J. (1999) Yellow head virus from Thailand and gill-associated virus from Australia are closely related but distinct prawn viruses. *Diseases of Aquatic Organisms* **36**, 153-157.
- Cowley, J.A., Dimmock, C.M., Spann, K.M. and Walker, P.J. (2000) Detection of Australian gill-associated virus (GAV) and lymphoid organ virus (LOV) of *Penaeus monodon* by RT-nested PCR. *Disease of Aquatic Organisms* **39**, 159-167.
- De Groot, R.J., , Luytjes, W., Horzinek, M.C., van der Zeijst, B.A.M., Spann, W.J.M. and Lenstra, J.A. (1987) Evidence for a coiled-coil structure in the spike proteins of coronaviruses. *J. Mol. Biol.* **196**, 963-966.
- Duarte, M. and Laude, H. (1994) Sequence of the spike protein of the porcine epidemic diarrhoea virus. *J. Gen. Virol.* **75**, 1195-1200.
- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.
- Mounir, S. and Talbot, P. (1993) Molecular characterization of the S protein gene of human coronavirus OC43. *J. Gen. Virol.* **74**, 1981-1987.
- Nadala, E.C.B., Tappy, L. M. And Loh, P. C. (1997) Yellow-head virus: a

- rhabdovirus-like pathogen of penaeid shrimp. *Diseases of Aquatic Organisms* **31**, 141-146.
- Racusen, D. (1979) Glycoprotein detection in polyacrylamide gel with thymol and sulfuric acid. *Anal. Biochem.* **99**, 474-476.
- Snijder, E.J. and Horzinek, M.C. (1995). The molecular biology of toroviruses. *In*: The Coronaviridae. S.G. Siddell (ed.) Pp. 219-238. Plenum Press, New York.
- Spaan, W., Cavanagh, D. and Horzinek, M.C. (1988) Coronavirus structure and genome expression. *J. Gen. Virol.* **69**, 2939-2952.
- Spann, K.M., Vickers, J.E. and Lester, R.J.G. (1995) Lymphoid organ virus of *Penaeus monodon* from Australia. *Diseases of Aquatic Organisms* **23**, 127-134.
- Spann, K.M., Cowley, J.A., Walker, P.J. and Lester, R.J.G. (1997) Gill-associated Virus (GAV), a yellow head-like virus from *Penaeus monodon* cultured in Australia. *Diseases of Aquatic Organisms* **31**, 169-179.
- Sittidilokratana, N., Hodgson, R.A.J., Panyim, S., Cowley, J.A., Jitrapakdee, S., Boonsaeng, V. and Walker, P.J. (2002) The complete *ORF1b*-gene sequence indicates yellow head virus is an invertebrate nidovirus. *Dis. Aquatic Organisms* **50**, 87-93.
- Sturman, L.S., Ricard, C.S. and Holmes, K.V. (1985) Proteolytic cleavage of the E2 glycoprotein of murine coronavirus: activation of cell-fusing activity of virions by trypsin and separation of two different 90K cleavage fragments. *J. Virol.* **56**, 904-911.
- Wang, Y.-C. and Chang, P.-S. (2000) Yellow head virus infection in the giant tiger prawn *Penaeus monodon* cultured in Taiwan. *Fish Pathology* **35**, 1-10.

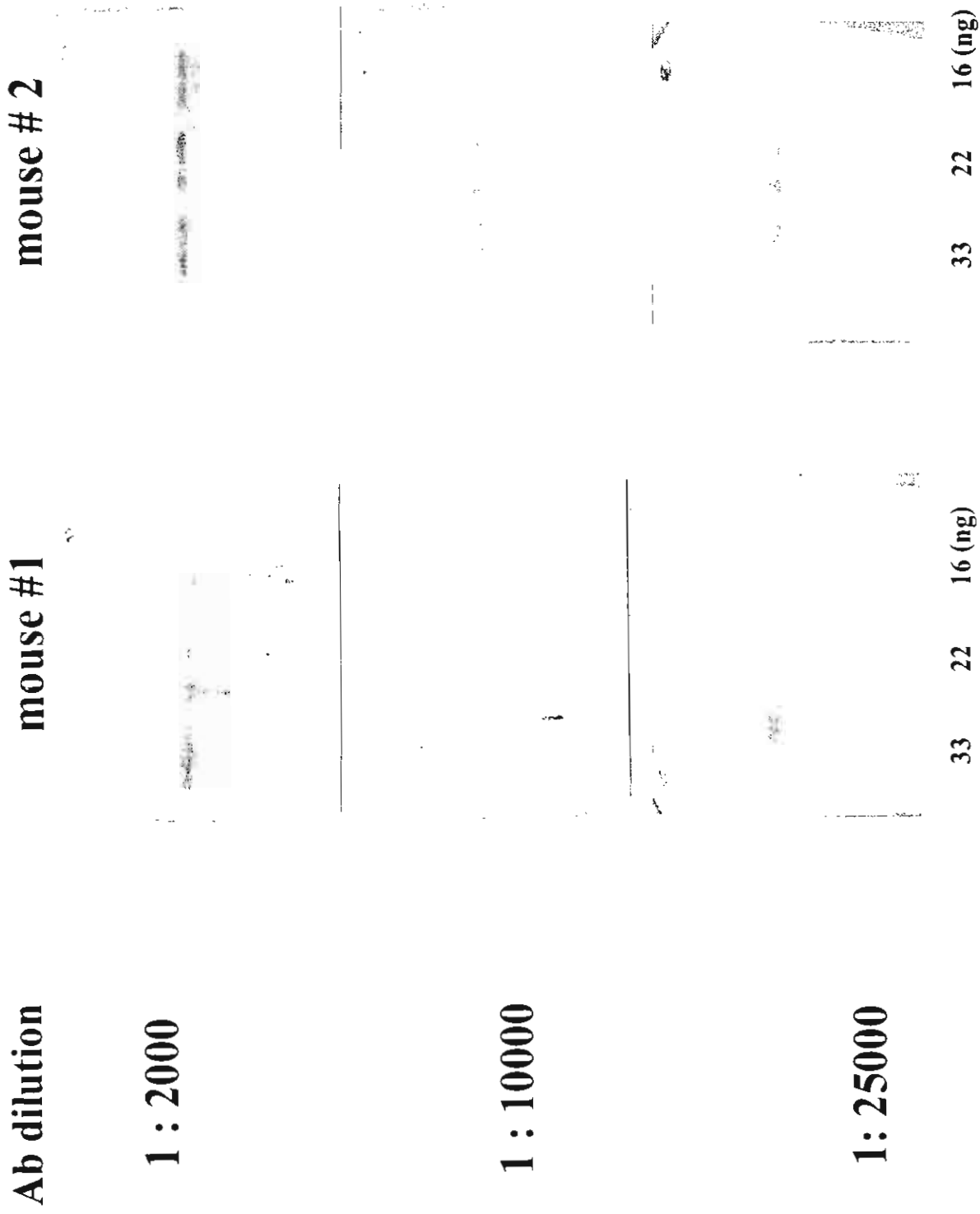
Wongteerasupaya, C., Sriurairatana, S., Vicker, J.E., Akrajamorn,S., Boonsaeng,V.,
Panyim, S., Tassanakajon, A., Withyachumnarnkul, B. And Flegel, T.W. (1995)
Yellow-head virus of *Penaeus monodon* is an RNA virus. *Diseases of Aquatic
Organisms* **22**, 45-50.



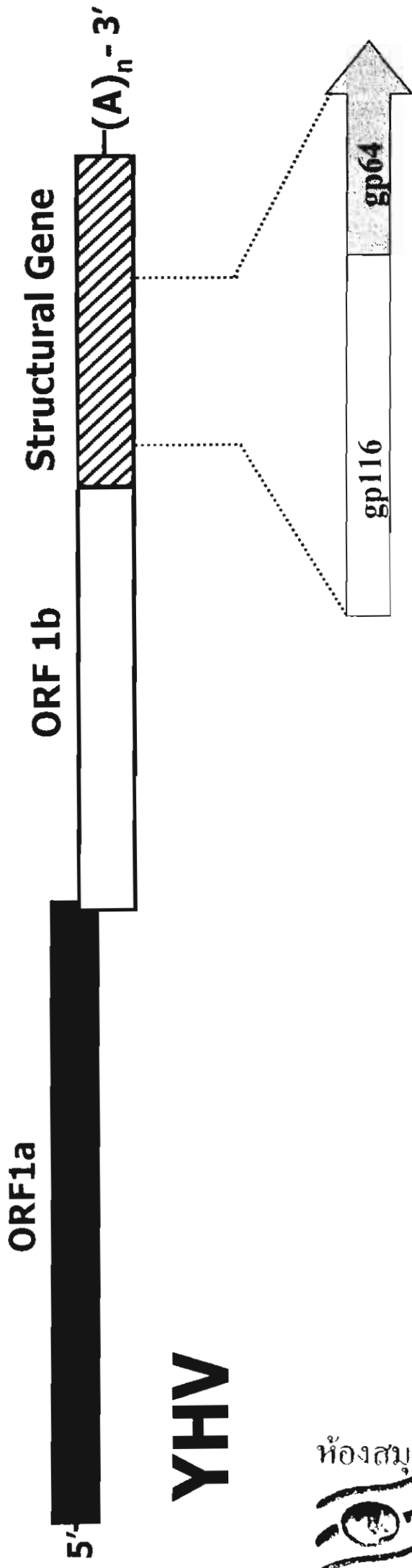
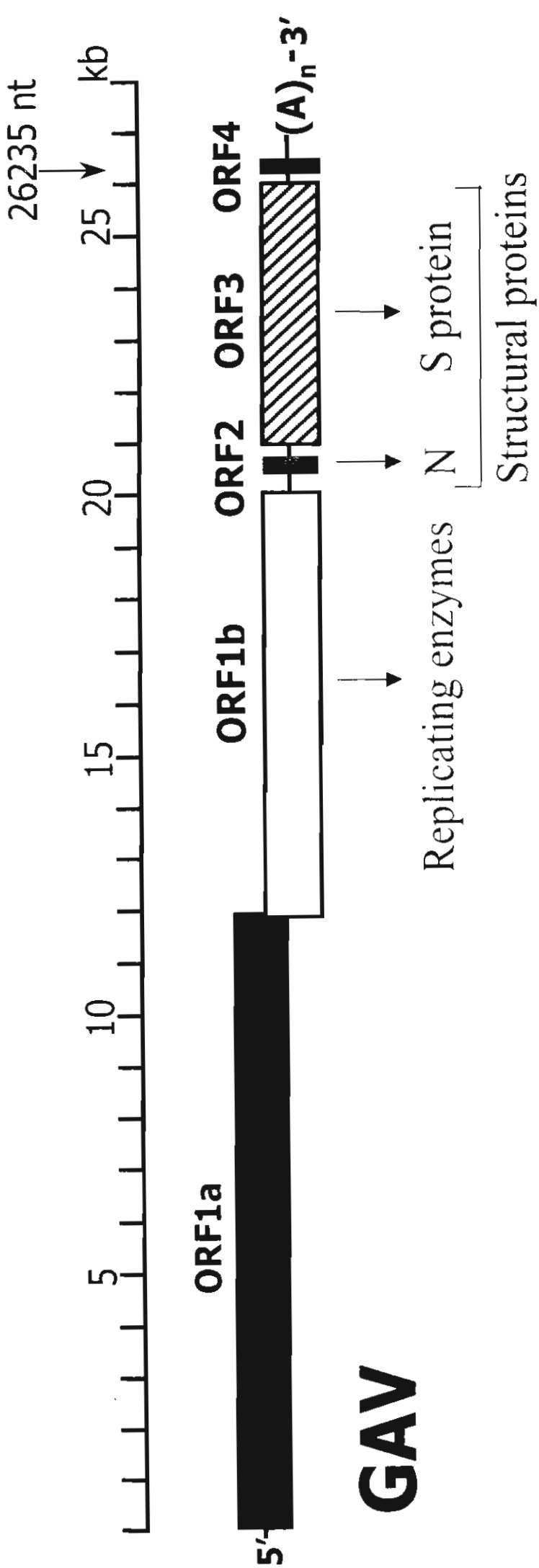
รูปที่ 1A การวิเคราะห์โปรตีนโครงสร้างของเชื้อไวรัสหัดเหลืองด้วยวิธี SDS-PAGE ตามด้วยการย้อมด้วย Coomassie blue

รูปที่ 1B การวิเคราะห์การเติมหมู่น้ำตาลของโปรตีนโครงสร้างด้วยชุดตรวจสอบ ECL glycoprotein detection system

รูปที่ 1C การวิเคราะห์การเติมหมู่น้ำตาลของโปรตีนโครงสร้างด้วยการย้อมด้วย Thymol reagent



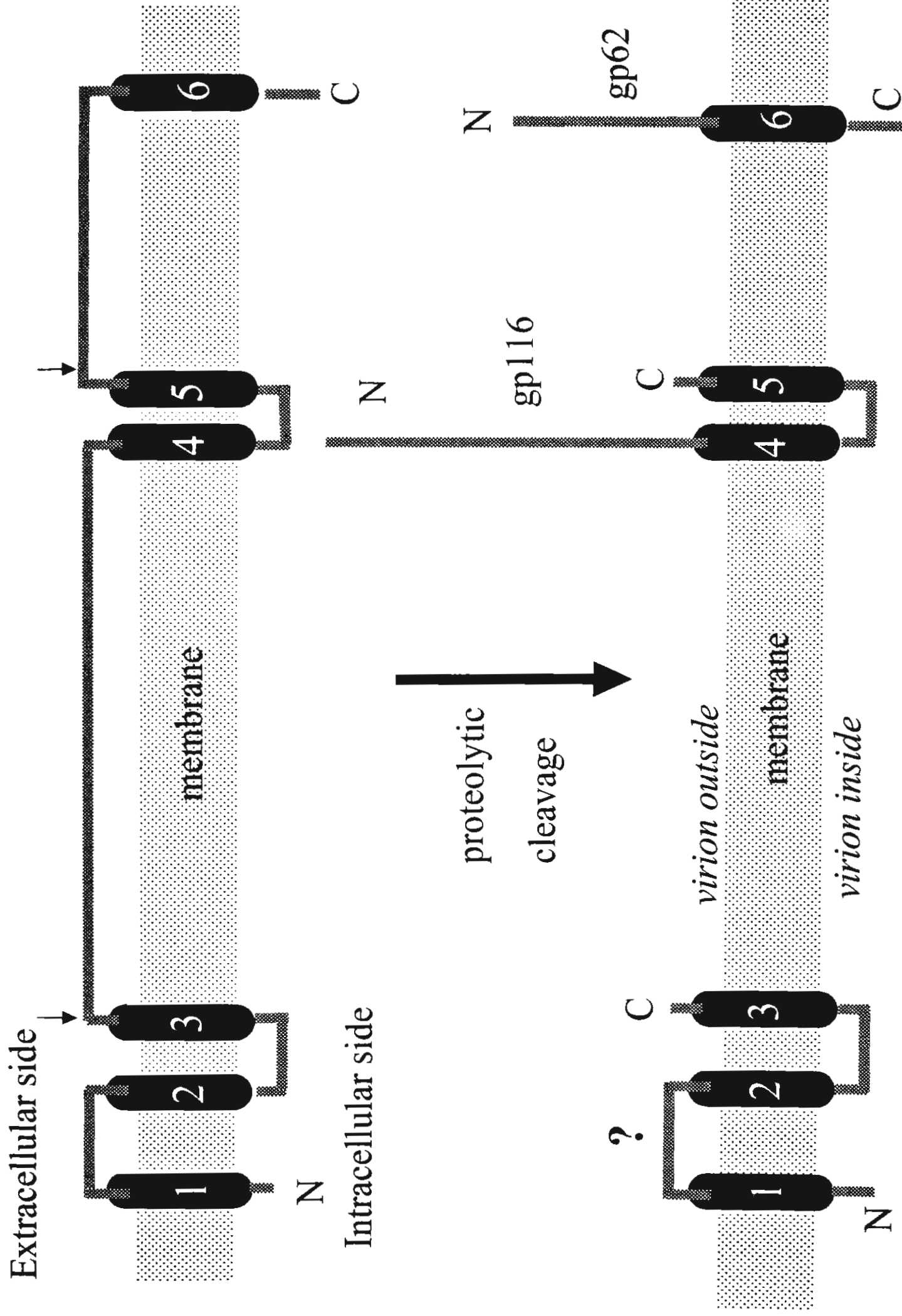
รูปที่ 2 ความไวของ polyclonal antibody ต่อโปรตีนโครงสร้าง gp64



รูปที่ 3 แสดงการเรียงตัวของยีนต่าง ๆ บนจีโนมของเชื้อไวรัสหัดเหลือง (YHV) เทียบกับ GAV

รูปที่ 4 แสดงลำดับนิวคลีโอไทด์และลำดับกรดอะมิโนที่แปลจาก ORF3 gene กรดอะมิโนที่พบจากปลายด้าน N โดยวิธี Edman degradation แสดงโดยอักษรตัวเข้ม บริเวณที่ขีดเส้นใต้คือส่วนที่เป็น transmembrane หรือเกลียว helix บริเวณที่สามารถถูกเติมหมู่คาร์โบไฮเดรต N-linked glycosylation ถูกแสดงไว้ใน box บริเวณที่ถูกตัดด้วย protease แสดงด้วยหัวลูกศร บริเวณ TATA และ CAAT ถูกแสดงโดยเส้นด้านบนและล่าง

[illegible]



รูปที่ 5 แสดงถึงขั้นตอนการเรียงตัวของ gp116 และ gp62 บน membrane รูปทรงกระบอกแทนส่วนที่เป็น transmembrane

รูปที่ 6 แสดงการเปรียบเทียบลำดับกรดอะมิโนของโปรตีนที่สร้างจาก ORF3 จากไวรัสหัวเหลือง (YHV) และ GAV ส่วนที่ล้อมกรอบสีแดงคือลำดับกรดอะมิโนที่เหมือนกันหรือมีคุณสมบัติทางเคมีใกล้เคียงกัน

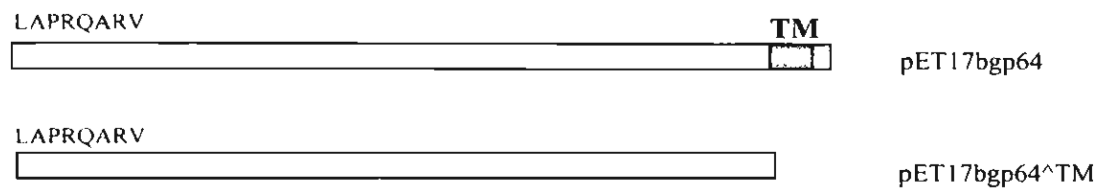
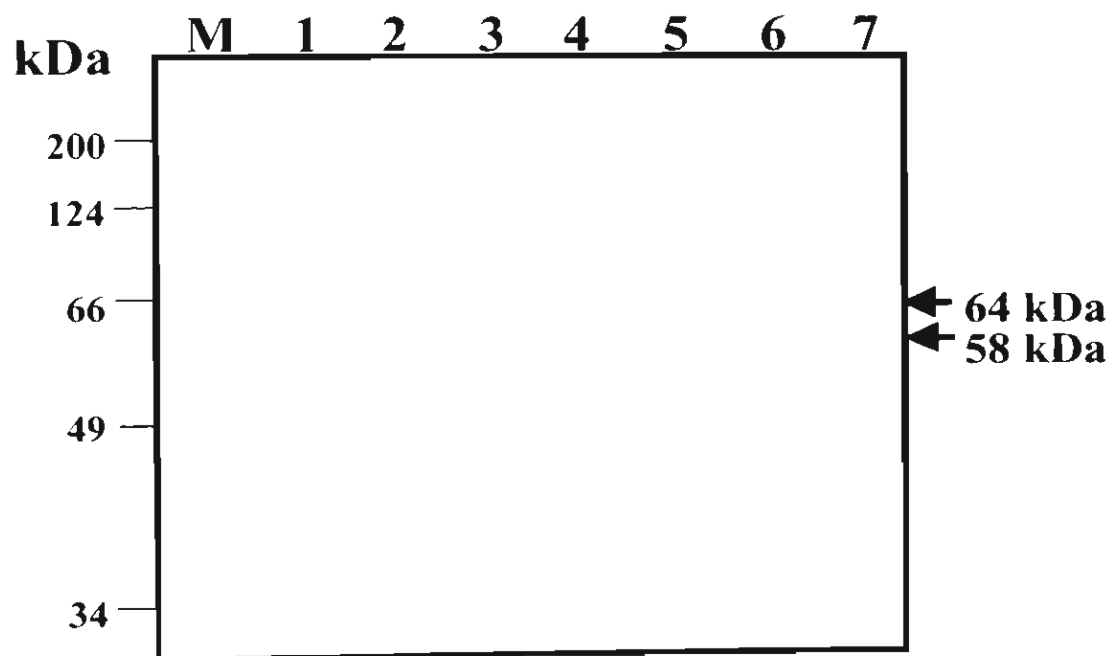
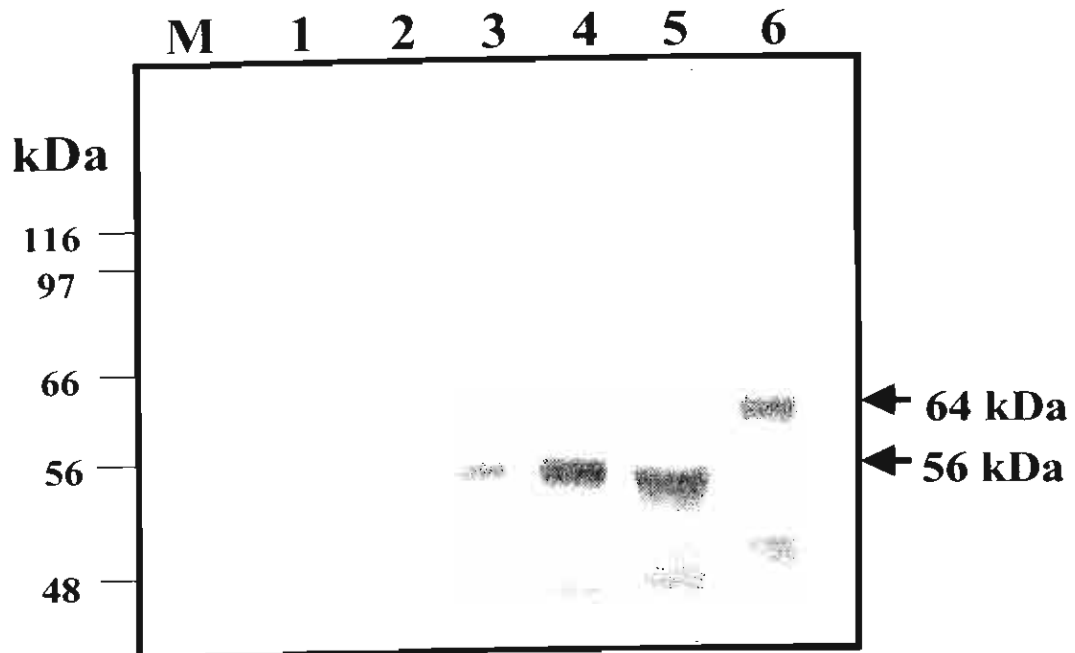
YHV	1	MQCSRFTMHFSFRSQPFSS	GLRGLFLTLFLSLTSSGSGHGF	LQFATPSSHLDAGATRTNS	ILSTCNNTITFTIKESCHLSSIM	SPVSCYDEH	HD	ISAPGSHIITYTGFAYLK	120	
GAV	1	MQCSRSTMLSSSHFPASLS	FTGLRGIFLFLLSAULFGSGHGF	LQFATPSSRSI	DAGAIRNS	ILSTCNNTITFTIKESCHLSSIM	SPVSCYDEH	HD	ISAPGSHIITYTGFAYLK	120
YHV	121	TDYECKEGYPANVTQV	FFIAILYLICSLLAYIFTKITPIAKSF	ISSCQTEMIMET	JETEKQVTVVAHRHKI	TPTSCHSIIINSTEN	ULVIGCFI	FLCFFHTEFAFATILSGIPEKDKS	240	
GAV	121	TDYECKEGYPANVTQV	FFIAILYLICSLLAYIFTKITPIAKSF	ISSCQTEMIMET	JETEKQVTVVAHRHKI	TPTSCHSIIINSTEN	ULVIGCFI	FLCFFHTEFAFATILSGIPEKDKS	229	
YHV	241	VLMAHVLEAGQTE	FFDQIHWANGDQ	FCNSTNCWSEHVQ	ILCEFCNTSSPLQLLLPHNLFRIL	FLSSSDADNPQ	VACDAGCY	SYINDYESKRTQAIKYTYT	360	
GAV	230	---ELETLESS-NHTE	FDYDMWAANGDQ	CNATNCWSEHVQ	ILCEFCNTSSPLQLLLPHNLFRIL	FLSSSDADNPQ	VACDAGCY	SYINDYESKRTQAIKYTYT	335	
YHV	361	TFEQUEMSIDEARYSD	IVSLYRINN	ITSVEGCMYNE	VSYYLHGD	SVPTCTPSTP	SGTTN	KKISQILYNH	KMVNVTVDQRC	480
GAV	336	TSQUESISIDEARYSE	IAWLYRHGNT	SAUGQVYNEAS	YYLHGD	AIPTCHPTPI	ILGTTT	FFHSIGKQILYNH	KMVNVTVDQRC	455
YHV	481	EPTLLIPFPYPRD	TKTLATHLGRV	FRNAGDYQIFILE	PGMLGRTY	DGYSYHEI	YASTH	DDCHVNM	SGDNK	600
GAV	456	KFTLLIPFPYPRD	TKTLATHLGRV	FRNAGDYQIFILE	PGMLGRTY	DGYSYHEI	YASTH	DDCHVNM	SGDNK	575
YHV	601	QYTDIEDIPAGFRDP	YDFSVDT	PTSPGPTIS	VLSEYHGD	STHETAPK	RFFIYRIM	AAARLT	PTSPQVEHLN	720
GAV	576	QYTDIDDIPAGFRDP	YDFSVDT	PTSPGPTIS	VLSEYHGD	STHETAPK	RFFIYRIM	AAARLT	PTSPQVEHLN	695
YHV	721	NEFNIONHLLAT	FGTATRTWA	MEYRHLPHFLTKRG	FYPLEPVT	GS	ADYLIVE	YNAHAS	RYSHQATYHQ	839
GAV	696	NDWNIONHLLAT	FGTATRTWA	MEYRHLPHFLTKRG	FYPLEPVT	EN	ADYLIVE	YNAHAS	RYSHQATYHQ	815
YHV	840	TSGYIFRDYDYR	KFPKFGNGLYL	GKVSAAASIG	TYSKCKQAQ	ISIPYHD	GIN	TDLGT	PPVYDSAC	959
GAV	816	TSGYIFRDYDYR	KFPKFGNGLYL	GKVSAAASIG	TYSKCKQAQ	ISIPYHD	GIN	TDLGT	PPVYDSAC	935
YHV	960	HSVTCIN	SKDNKVHVVRQ	FGYSYI	AGDPQALH	ISHNKH	MPYHS	ILKQIN	LFHSYLYQ	1079
GAV	936	HSVTCIN	SKDNKVHVVRQ	FGYSYI	AGDPQALH	ISHNKH	MPYHS	ILKQIN	LFHSYLYQ	1054
YHV	1080	GLDTKHLKI	HARHKIYI	HS	PMLGR	TFFFLWS	PIYA	ILLSI	SPASALAP	1199
GAV	1055	GLDTKHLKVHAR	HKIYI	HS	PMLGR	TFFFLWS	PIY	IMAL	FALLSPASALAP	1174
YHV	1200	SSCAYLYTSLPP	KLCDARIN	WSCLHTG	SCRN	SEYLE	FKPLG	QHTSND	YIVAN	1319
GAV	1175	SSCAYLYTSLPP	KLCDARIN	WSCLHTG	SCRN	SEYLE	FKPLG	QHTSND	YIVAN	1294
YHV	1320	TNECHSITTD	YRNYTAD	HKITIS	FNIAHP	QVAKTTFK	VGALFT	QCTTP	PKHIL	1439
GAV	1295	TNECHSITTD	YRNYTAD	HKITIS	FNIAHP	QVAKTTFK	VGALFT	QCTTP	PKHIL	1414
YHV	1440	PLGALVQC	NYDQTY	ISTIT	TARQORS	ILNGV	TYTDE	VDAL	SYMLLN	1559
GAV	1415	PLGALVQC	NYDQTY	ISTIT	TARQORS	ILNGV	TYTDE	VDAL	SYMLLN	1534
YHV	1560	THADHLCN	FTASASE	FDLKVNG	KAVMT	ITITVIT	QCDVGS	ISD	TI	1665
GAV	1535	THADHLCN	FTASASE	FDLKVNG	KAVMT	ITITVIT	QCDVGS	ISD	TI	1640

รูปที่ 7 การแสดงออกของโปรตีน gp64 จากแบคทีเรีย *E. coli*

A, แสดงพลาสมิด 2 ชนิดคือ gp64 ที่มี transmembrane domain (pET17b-gp64) และ gp64 ที่ไม่มี transmembrane (pET17b-gp64 Δ TM)

B, การวิเคราะห์ด้วยวิธี Western จากโปรตีน *E. coli* BL21(DE3) ที่ transform ด้วย pET17b-gp64 M, marker; lane 1, BL21(DE3) ที่มีพลาสมิด pET17b; lane 2-6 BL21(DE3) ที่มีพลาสมิด pET17b-gp64 ทำการเหนี่ยวนำให้สร้างโปรตีนด้วย IPTG เป็นเวลา 0, 1, 3, 5 และ 7 ชั่วโมง; lane 7, purified YHV

C, การวิเคราะห์ด้วยวิธี Western จากโปรตีน *E. coli* BL21(DE3) ที่ transform ด้วย pET17b-gp64 Δ TM M, marker; lane 1, BL21(DE3) ที่มีพลาสมิด pET17b; lane 2-5, ทำการเหนี่ยวนำให้สร้างโปรตีนด้วย IPTG เป็นเวลา 0, 1, 3, 5 ชั่วโมง; lane 6, purified YHV

A**B****C**

Output จากโครงการวิจัย

1. Manuscript entitled Identification and analysis of gp116 and gp64 structural glycoproteins of the yellow head head nidovirus of *Penaeus monodon* shrimp by Jitrapakdee, S., Unajak, S., Sidthidilokratana, N., Hodgson, R.A.J., Cowley, J.A., Walker, P.J., Panyim, S. and Boonsaeng, V. ส่งขอรับการตีพิมพ์ในวารสาร Journal of General Virology.
2. ผลงานตีพิมพ์ระดับปริญญาโทหลักสูตรชีวเคมี คือ น.ส.ศศิมนัส อุณจักษ์

ภาคผนวก

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

Sarawut Jitrapakdee¹*, Sasimanas Unajak¹, Nusra Sidthidilokratana¹, Richard A.J.

Hodgson², Jeff A. Cowley², Peter J. Walker², Sakol Panyim¹ & Vichai Boonsaeng¹

¹ *Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok, Thailand*

² *CSIRO Livestock Industries, Long Pocket Laboratories, Indooroopilly, Queensland,*

Australia

**To whom correspondence should be addressed*

Dr Sarawut Jitrapakdee

Department of Biochemistry

Faculty of Science

Mahidol University

Bangkok 10400, Thailand

Tel: 662-201-5460

Fax: 662-248-0375

Email: scsjr@mahidol.ac.th

SUMMARY

Yellow head virus (YHV) is a major agent of disease in farmed penaeid shrimp. YHV virions purified from infected shrimp contain 3 major structural proteins of M_r 116 kDa (gp116), 64 kDa (gp64) and 20 kDa (p20). Two different staining methods indicated that the p116 and p65 proteins are glycosylated. Here we report the complete nucleotide sequence of the ORF3 which encodes a polypeptide of 1,664 amino acids with calculated M_r of 185,713 kDa ($pI = 6.68$). Hydropathy analysis of the deduced ORF3 protein sequence identified 6 transmembrane helices and 3 ectodomains containing multiple sites for 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 Ala²²⁸ and Thr²²⁹, and Ala¹¹²⁷ and Leu¹¹²⁸ located at the C-terminal side of transmembrane helices 3 and 5, respectively. Comparison with the deduced ORF3 protein sequences of Australian gill-associated virus (GAV) indicated 83% identity in gp64 and 71% identity in gp116 which featured 2 significant sequence deletions near the N-terminus. Database searches revealed no significant homology with other proteins. The recombinant gp64 with and without the C-terminal transmembrane region expressed in *E. coli* was shown to react with antibody raised against native gp64 purified from virions.

Introduction

Viral disease is 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, 1998]. 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 coloration 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.

Yellow head virus (YHV) is an enveloped, rod-shaped particle (approximately 40 nm x 170 nm) with prominent surface projections (approximately 11 nm) and an inner helical nucleocapsid (Chantanachookin *et al.*, 1993; Wang & Chang 2000; Loh *et al.*, 1997). Primarily based on the virion morphology and the presence of a single-stranded RNA genome (Wongteerasupaya *et al.*, 1995), YHV was previously reported as a rhabdovirus (Nadala *et al.*, 1997). However, it was subsequently demonstrated that the YHV genome is positive-sense RNA [Tang & Lightner, 1999]. Sequence analysis has also revealed that, like closely related gill-associated virus (GAV) from Australia (Cowley *et al.*, 1999; Cowley *et al.*, 2000), YHV contains a large replicase gene (ORF1b) that is expressed as a polyprotein by ribosomal frame-shift at a "slippery" sequence that follows a predicted pseudoknot structure (Sittidilokratna *et al.*, 2002). Considerations of sequence identity, genome organization and gene expression have indicated that GAV and YHV are related to coronaviruses, toroviruses and arteriviruses and will be classified in new taxa (proposed family *Roniviridae*, genus *Okavirus*) within the order *Nidovirales* (Cowley *et al.*, 2000; Sittidilokratna *et al.*, 2002).

Nadala *et al.*, 1997 originally reported that YHV particles comprise 4 structural proteins of approximately 170 kDa, 135 kDa, 67 kDa and 22 kDa, of which the 135 kDa protein was glycosylated. However, Wang and Chang (2000) subsequently reported only 3 major YHV proteins (110 kDa, 63 kDa and 20 kDa), suggesting the larger protein may be of cellular origin. In GAV, structural proteins are encoded in ORF2 and ORF3 located immediately downstream of the ORF1b gene (Cowley *et al.*, 2001). ORF2 encodes a 22

kDa structural protein that functions as the nucleoprotein (Cowley *et al.*, submitted), and ORF3 encodes a polypeptide with the structural characteristics of a large glycoprotein with multiple membrane-spanning domains (Cowley and 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) that are synthesized as a polyprotein which undergoes post-translational modification including proteolytic cleavage and glycosylation.

METHODS

Viral purification

As a source of material for virus purification, 200-500 juvenile *Penaeus monodon* shrimp (average weight 20 g), were infected by intramuscular injection with a standard inoculum of a 1998 Thai YHV isolate prepared as described previously (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 Urografin (Schering AG, Germany) 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 analyzed by SDS-PAGE (Laemmli, 1970) under reducing conditions in 7.5% or 15% discontinuous gels. The fractionated proteins were transferred to a polyvinylidene difluoride (PVDF) membrane in transfer buffer (10 mM CAPS/10% methanol) using a semi-dry blotter (Hoeffer). Protein bands blotted on the membrane were stained briefly in 0.3% Coomassie brilliant blue R250. The protein bands were then excised and analyzed by N-terminal sequencing. Edman degradation was conducted using an Applied Biosystems Sequencer at the Center for Genetic Engineering and Biotechnology, Bangkok.

Polyclonal antibody production and immunoblot analysis

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

Proteins were transferred to a nitocellulose membrane (Amersham Pharmacia, Upjohn) using a semi-dry blotting apparatus (Hoeffer). Following electroblot transfer, the membrane was blocked in 3% bovine serum albumin (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 1:10,000 dilution of polyclonal antibodies for 1 h. Goat anti-mouse polyclonal antibodies conjugated with alkaline phosphatase were then reacted for 2 h. Immuno-reactive bands on the membrane were visualized by adding nitro blue tetrazolium (NBT) and 5-bromo-4-chloro-3-indoyl phosphate (BCIP).

Glycoprotein staining

Glycoprotein detection was performed using the ECL glycoprotein detection system (Amersham Pharmacia). Proteins blotted on 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 hydrazide 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. The thymol staining was essential the same as described by Racusen *et al.* (1979).

Reverse transcriptase polymerase chain reaction (RT-PCR), cloning and sequencing

YHV genomic RNA was extracted from purified virus using Trizol Reagent (Invitrogen, U.S.A.), resuspended in DEPC-treated water. The RT-PCR was performed in a total volume of 25 µl containing 200 ng of genomic RNA, 0.4 mM of forward primer (5'-gacgggggtacctaagcttatgetatgaccta-3') designed from the 3'-end of the ORF1b gene and oligo d(T) primer (5'-tctagaggatccccggtaaccttttttttttttttt-3') using SuperScript one-step RT-PCR system (Invitrogen, U.S.A.) in the presence of 1 unit of Elongase (Invitrogen, U.S.A.) 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 sec, annealing/extension at 68°C for 8 min. The RT-PCR product was either sequenced directly or cloned in pGEM-T Easy vector. The RT-PCR product was gel-purified using QIAquick Gel Extraction kit (Qiagen) and directly sequenced using BIG DyeTM reagent (ABI, U.S.A.). Nucleotide sequences obtained from an initial sequencing were used to

design new primers to sequence toward the 5'-end and the 3'-end of the fragment. Overlapping sequences were then analysed using SeqEd 1.0.3 (ABI, U.S.A.).

Recombinant protein expression in *E. coli*

The cDNA encoding full length gp64 was generated by PCR using forward primer (YHV-D7) 5'-GCCTCTAGACATATGCTCGCTCCACGACAGGCACGTGTT-3' and reverse primer (YHV-D8)

5'-CATTGTGGATCCT**CACT**AGTGATGATGATGATGATGGGATCGTTTGGCTTT-CGTTCTCATGGACGT-3'. The forward primer was designed from residues L¹¹²⁸ to V¹¹³⁵ in the ORF3 polyprotein and included initiation Met and an *NdeI* restriction site (underline) while the reverse primer was designed from residues T¹⁶⁵⁷ to S¹⁶⁶⁶ and included stop codons (bold) and a *Bam*HI restriction site (underline). PCR was performed in a 25 µl reaction mixture containing 1X 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 *Taq* DNA polymerase (Perkin Elmer, U.S.A.). The reaction mixture was subjected to 35 cycles of denaturation at 94°C 30 sec, annealing/extension at 68°C 2 min, and followed by the final extension at 72°C for 10 min. The PCR products were cut with *NdeI* and *EcoRI* and ligated into the multiple cloning site of pET17b (Novagen). The C-terminal transmembrane deleted construct was generated to remove the last 40 C-terminal residues (Y¹⁶²⁷-S¹⁶⁶⁶) by PCR using YHV-D7 and YHV-D14 [5'-GAATTC**CACTAATCCCATGTCTTGCCGCCGAA**; corresponds to residues F¹⁶²⁰-D¹⁶²⁶ with an *EcoRI* attached at the 5'-end and stop codons (bold)]. The PCR product was cloned into pDrive vector (Qiagen) and sequenced. The insert was then digested with *NdeI* and *EcoRI* and cloned into *NdeI/EcoRI* sites of pET17b. The recombinant plasmids were then transformed into *E. coli* BL21(DE3) (Novagen). Overnight culture of BL21(DE3) was 1:20 diluted with fresh LB broth and grown at 37°C to an A₆₀₀ of 0.5-0.6. The culture was then induced with 1 mM IPTG and induced for 6 h. One milliliter of culture was pelleted, suspended in protein loading buffer and heated at 100°C for 10 min before subjected to SDS-PAGE. The recombinant protein produced 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 hemolymph revealed the typical rod-shaped, enveloped particles described by Wongteerasupaya *et al.* (1995) [data not shown]. The purified particles were disrupted with SDS and analyzed by SDS-PAGE. As shown in Figure 1, analysis under reducing conditions revealed three relatively abundant proteins of M_r 116 kDa, 64 kDa, 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 Figure 1, hemocyanin (~68-70 kDa) is the most abundant protein in shrimp hemolymph (Figueroa-Soto *et al.*, 1997). As there was little evidence residual hemocyanin 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 is used the thymol reagent and is based on the hydrolysis of glycosidic bonds concomitant with the formation of fural derivative (Racusen, 1979). As shown in Figure 1B and 1C, each method indicated that the 116 kDa 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 (Sittidilokratna and associates, unpublished data), the size of the amplified product was consistent with the expected size of 3'-end of YHV genome (Figure 2). Nucleotide sequence analysis revealed the region contained 2 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. Analysis of the nucleotide 240 bp upstream of the initiation codon using Sigscan (Heinemeyer *et al.*, 1998) and TF-SEARCH (Prestridge, 1991) databases showed no typical TATA box but contained one inverted CCAAT box at position -49/-37. However there is a TATA-like sequence (TAAAT) found at positions -29/-24. The complete nucleotide sequence and deduced amino acid sequence of YHV ORF3 is shown in Fig. 3. ORF3 encodes a polypeptide of 1,666 amino acids with predicted molecular weight of 185,713 and pI of 6.68. Alignment with the N-terminal sequences of mature gp116 and gp64 identified perfect identity with residues T²²⁹ to F²³⁶ and L¹¹²⁸-V¹¹³⁵ respectively of the encoded polypeptide. The data indicated that gp116 and gp64 are derived by post-translational proteolysis of the ORF3 polyprotein. Based on the identified N-terminal sequences and presumed sites of proteolysis, the calculated molecular weight for unprocessed gp116 was 101,734 Da and for gp64 was 58,599 Da. The predicted size of these products was consistent with evidence of post-translational glycosylation of each protein (Fig. 1B and 1C).

Hydropathy plot of the ORF3 polypeptide sequence (<http://sosui.proteome.bio.tuat.ac.jp>) identified six hydrophobic transmembrane helices - three in the N-terminal domain (residues L²⁵-L⁴⁷, Y¹³⁸-I¹⁵⁵, I²⁰⁸-I²³⁰), two in the central domain (residues G¹⁰²⁷-N¹⁰⁴⁹ and I¹¹⁰⁶-L¹¹²⁸) and one in the C-terminal domain (residues K¹⁶³⁰-I¹⁶⁵²). Analysis of membrane topology using TMHMM (Krogh *et al.*, 2001) predicted that the N-terminus of the ORF3 polyprotein is anchored inside the cell via the first transmembrane domain. Three subsequent ectodomains are connected by the second to the sixth transmembrane domains respectively, and the C-terminus of the protein is also anchored within the cell (Fig. 4). Proteolytic cleavage between A²²⁸ and T²²⁹, and A¹¹²⁷ and L¹¹²⁸ near C-terminus of transmembrane helices 3 and 5 respectively would generate 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 NetOGlyc2.0 database (<http://www.cbs.dtu.dk>) [Hansen *et al.*, 1997] revealed the presence of 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 polyprotein. As shown in Fig. 3, potential N-linked glycosylation sites were also identified in the ectodomains of gp64 (5 sites), gp116 (7 sites) and the unidentified 25.4 kDa transmembrane fragment of ORF3 (1 site). The ectodomains were also rich in cysteine residues, suggesting complex folded loop structures in the mature glycoproteins (Fig. 3).

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 and Walker, 2002) indicated overall amino acid sequence similarity of 75%. The level of sequence homology was highest in gp64 (83% similarity) and lowest in gp116 (71% similarity), primarily due to 2 significant deletions at the N-terminus of the gp116 homologue in GAV (see Fig. 5). A comparison of individual sites in the YHIV and GAV ORF3 polyproteins indicated that all cysteine residues were conserved and all but two (N²⁹¹-S²⁹³ and N⁷¹¹-T⁷¹³ in gp116) of the potential YHIV N-glycosylation sites were preserved. Three of eight predicted O-glycosylation sites in YHIV gp116 were not conserved and the single O-glycosylation site in YHIV gp64 was not conserved in GAV, suggesting this protein may contain only N-glycans.

Expression of gp64 in *E. coli*

The gp64 coding region of YHIV ORF3 was amplified by PCR using a forward primer designed from L¹¹²⁸ to V¹¹³⁵ in ORF3 with an upstream initiation codon and a reverse primer designed from T¹⁶⁵⁷ to S¹⁶⁶⁶ including 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. 6B, expression of recombinant gp64 was not evident by SDS-PAGE

analysis 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 blot (lanes 3-6). The polyclonal antibody also reacted with native gp64 in purified virus (lane 7) but did not react with lysates prepared from *E. coli* transformed with pET17b alone (lane 1). Although the expression level is relatively poor, the size of the band is consistent with the molecular weight of gp64 calculated from the deduced amino acid sequence. The difference is M_r 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 level of ~58 kDa band decreased 1 h after induction, indicating the recombinant gp64 produced in *E. coli* is not stable. It is possible that the transmembrane region of gp64 may cause poor expression. We then generated the transmembrane deleted construct (pET16b-gp64 Δ TM, see Fig. 6A) by removing residues Y¹⁶²⁷-S¹⁶⁶⁶ located at the C-terminal of gp64. This construct was transformed in *E. coli* and induced with IPTG. Again there was no evidence of overexpressed protein band corresponds to the size of gp64 lacking transmembrane domain (~55 kDa) on Coomassie stained gel. However Western analysis of this gel revealed the presence of strong immunoreactive band approximately 55 kDa. This band was increased over the induction period (1-5 h) suggesting that recombinant gp64 without transmembrane domain is more stable than full length gp64.

DISCUSSION

In mammalian nidoviruses, the virion envelope glycoproteins have functions that are principally associated with 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 the determination of virulence (Dalziel *et al.*, 1987; 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 invertebrate nidovirus glycoproteins. 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 with or without transmembrane region could be expressed in *E. coli* in a form that is recognized by antibody to the native virion glycoprotein.

Nadala *et al.* 1997) have previously reported that YHV contains 4 major structural proteins (M_r = 200 kDa, 135 kDa, 62 kDa and 20 kDa) of which only the 135 kDa protein was reported to be glycosylated. Wang & Chang (2000) have reported only three structural proteins (M_r = 110 kDa, 63 kDa 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 M_r 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 1,666 amino acid ORF3 polyprotein. The size of the YHV polyprotein is similar to those of vertebrates coronaviruses including avian infectious bronchitis virus (1,160 amino acids) [Binns *et al.*, 1985], feline infectious peritonitis virus (1,452 amino acids) [de Groot *et al.*, 1987], the murine hepatitis virus (1,376 amino acids) [Spann *et al.*, 1988], porcine epidemic diarrhea virus (Duarte & Laude, 1994) [1,383 amino acids] and the human respiratory coronavirus (1,353 amino

acids) [Mounir & Talbot, 1993]. However, the predicted structure of the ORF3 polyprotein is highly complex, comprising 6 transmembrane regions, three ectodomains and two intra-virion domains. By contrast, coronaviruses contain a single transmembrane region located near the C-terminus of the protein.

The envelope spikes of vertebrate coronaviruses and toroviruses comprise a large (~180 kDa) glycoprotein (S) 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 prominent peplomers 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.*, 1987) 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, the 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, 1994; Luo *et al.*, 1999). The S2 ectodomain also contains two large amphiphatic α -helices with a heptad repeat that have been proposed to mediate coiled-coil interchain interactions (de Groot *et al.*, 1987). As reported for GAV (Cowley & Walker, 2002), examination of YHV gp64 and gp116 sequences revealed no heptad repeats or significant amphiphatic α -helices suggesting an absence of coiled-coil structures. Further work is required to determine if the processed gp116 and gp64 are associated to form the envelope spikes.

Sequence analysis of the of the YHV genome was facilitated by use of an oligo (dT) primer to amplify the 3'-terminal region. This approach was adopted in the expectation that the YHV genome would be polyadenylated - a characteristic of the viruses (+) RNA viruses in the order *Nidovirales*. The ORF3 gene, encoding the YHV structural glycoproteins, is located downstream of the ORF1b gene which encodes replication enzymes and has features characteristic of nidoviruses (Sithidilokratana *et al.*, 2002). In GAV, ORF2 (located in the genome between ORF1b and ORF3) appears to

encode the nucleoprotein (Cowley *et al.*, submitted) 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 is upstream of the structural glycoprotein gene and there is no discrete gene encoding the integral membrane (M) protein. 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 fragment also contains 3 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 towards the virion surface. Also, as p20 appears to be the nucleoprotein encoded in ORF2 (Cowley *et al.*, submitted), 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 remaining fragment of the ORF3 polyprotein.

The molecular weights of gp116 and gp64 calculated from the nucleotide sequence are smaller (~15 kDa and ~4 kDa respectively) than the estimated size of the mature forms as determined by SDS-PAGE. The size differences are evidently due to glycosylation as revealed by the detection of carbohydrate (Fig. 1) and the presence of several putative N- and O-linked glycosylation sites in each protein. The authenticity of cDNA encoding gp64 was also confirmed by the expression in *E. coli*. 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 weight 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 in *E. coli* independently of gp116. Our attempt to express gp64 as the fusion protein with glutathione S-transferase failed to improve the yield. The expression of gp64 in *E. coli* has been improved only after removal of transmembrane domains (Fig. 6C). Similar result was observed when gp116 with and without transmembrane domain were expressed in *E. coli* (data not shown). Expression of these proteins in insect cells which

may provide an intracellular environment similar to shrimp cells, providing 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 has a genome organization and glycoprotein expression strategy that is fundamentally different from other nidoviruses. The work further supports the classification of YHV with GAV in the new family *Roniviridae* (Cowley *et al.*, submitted; Sithidilokratana *et al.*, 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 Ph.D. Golden Jubilee program, TRF. S.P. is a TRF senior research scholar.

References

1. Binns, M.M., Bourns, M.E.G., Cavanagh, D., Pappin, D.J.C. and Brown (1985) Cloning and sequencing of the gene encoding the spike protein of the coronavirus IBV. *J. Gen. Virol.* **66**, 719-72
2. Britton, P. (1991) Coronavirus motif. *Nature* **353**, 394.
3. Cavanagh, D. (1995) The coronavirus surface glycoprotein. In: The Coronaviridae. S.G. Siddle (ed) p73-113. Plenum Press, New York
4. Chantanachookin, C., Boonyaratpalin, S., Kasornchandra, J., Sataporn, D., Ekpanithanpong, U., Supamataya, K., Sriurairatana, S. and Flegel, T.W. (1993) Histology of ultrastructure reveal a new granulosis-like virus in *Penaeus monodon* affected by yellow-head disease. *Diseases of Aquatic Organisms* **17**, 145-157.
5. Cowley, J.A., Cadogan, L.C., Spann, K.M. and Walker, P.J. (2002) The ORF2 gene encodes the nucleocapsid protein of gill-associated nidovirus of *Penaeus monodon* prawns. *J. Gen. Virol.* submitted
6. Cowley, J.A., Dimmock, C.M., Spann, K.M. and Walker, P.J. (2000) Detection of Australian gill-associated virus (GAV) and lymphoid organ virus (LOV) of *Penaeus monodon* by RT-nested PCR. *Disease of Aquatic Organisms* **39**, 159-167.
7. Cowley, J.A., Dimmock, C.M., Spann, K.M. and Walker, P.J. (2001) Gill-associated virus of *Penaeus monodon* prawns: Molecular evidence for the first invertebrate nidovirus. In: The Nidoviruses. E. Lavi, S.R Weiss and S.T. Hingley (eds.). Advances in Experimental Medicine and Biology 494, 43-48. Kluwer Academic/Plenum Press, New York.
8. Cowley, J.A. & Walker, P.J. (2002) The complete sequence of gill-associated virus of *Penaeus monodon* prawns indicates a gene organisation unique among nidoviruses. *Arch. Virol (In Press)*.
9. Dalziel, R.G., Lampert, P.W., Talbot, P.J. and Buchmeier, M.J. (1986). Site-specific alteration of murine hepatitis virus type 4 peplomer glycoprotein E2 results in reduced neurovirulence. *J. Virol.* **59**, 463-471.
10. de Groot, R.J., Maduro, J., Lenstra, J.A., Horzinek, M.C., van Der Zeijst, B.A.M. and Spaan, W.J.M. (1987) cDNA cloning and sequence analysis of the gene encoding the peplomer protein of feline infectious peritonitis virus. *J. Gen. Virol.* **68**, 2639-2646.
11. De Groot, R.J., , Luytjes, W., Horzinek, M.C., van der Zeijst, B.A.M., Spann, W.J.M. and Lenstra, J.A. (1987) Evidence for a coiled-coil structure in the spike proteins of coronaviruses. *J. Mol. Biol.* **196**, 963-966.

12. Diano, M., Le Bivic, A. and Hirn, M. (1998) Raising polyclonal antibodies using nitrocellulose-bound antigen. *Methods. Mol. Biol.* **80**, 5-13.
13. Duarte, M. and Laude, H. (1994) Sequence of the spike protein of the porcine epidemic diarrhoea virus. *J. Gen. Virol.* **75**, 1195-1200.
14. Figueroa-Soto, C.G., de la Barca, A.M.C., Vazquez-Moreno, L., Higuera-Ciapara, I. and Yepiz-Plaascencia, G. (1997) Purification of hemocyanin from white shrimp (*Penaeus vannamei* Boone) by immobilized metal affinity chromatography. *Comp. Biochem. Physiol.* **117B**, 203-208.
15. Flegel, T.W. (1997) Major viral diseases of the black tiger prawn (*Penaeus monodon*) in Thailand. *World J. Microbiol. Biotechnol.* **13**, 433-442.
18. Grosse, B. and Siddell, S.G. (1994) Single amino acid changes in the S2 subunit of the MHV surface glycoprotein confer resistance to neutralization by S1-specific monoclonal antibody. *Virology* **202**, 814-824.
19. Hanen, J.E., Lund, O., Rapacki, K. and Brunak, S. (1997) O-glycbase version 2.0- A revised database of O-glycosylated proteins. *Nucleic Acids Res.* **25**, 278-282.
20. Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A.E., Kel, O.V., Ignatieva, E.V., Ananko, E.A., Podkolodnaya, O.A., Kolpakov, F.A., Podkolodny, N.L. and Kolchanov, N.A. (1998) *Nucleic Acid Res.* **26**, 362-367.
21. Krogh, A., Laarsson, B., von Heijne, G. and Sonnhammer, E.L.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567-580.
22. Kubo, H., Yamada, Y.K. and Taguchi, F. (1994) Localization of neutralization epitopes and the receptor-binding site within the amino-terminal 330 amino acids of the murine coronavirus spike protein. *J. Virol.* **68**, 5403-5410.
23. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.
24. Limsuwan, C. (1991) Handbook for cultivation of black tiger prawns. Tansetakit Co. Ltd., Bangkok (in Thai).
25. Loh, P.C., Tapay, L.M., Lu, Y. and Nadala, E.C. Jr. (1997) *Adv. Virus Res.* **48**, 263-312.
26. Luo, Z., Matthews, A.M. and Weiss, S.R. (1999) Amino acid substitutions within the leucine zipper domain of the murine coronavirus spike protein cause defects in oligomerization and the ability to induce cell-to-cell fusion. *J. Virol.* **73**, 8152-8159.
27. Mounir, S. and Talbot, P. (1993) Molecular characterization of the S protein gene of

- human coronavirus OC43. *J. Gen. Virol.* **74**, 1981-1987.
28. Murray, M.C., Bhavanadan, V.P. and Davidson, E.E. (1989) *Carbohydr. Res.* **186**, 255-265.
 29. Nadala, E.C.B., Tappy, L. M. And Loh, P. C. (1997) Yellow-head virus: a rhabdovirus-like pathogen of penaeid shrimp. *Diseases of Aquatic Organisms* **31**, 141-146.
 30. Phillips, J.J., Chua, M.M., Lavi, E. and Weiss S.R. (1999). Pathogenesis of MHV4/MHV-A59 recombinant viruses: the murine coronavirus spike protein is a major determinant of neurovirulence. *J. Virol.* **73**, 7752-7760.
 31. Prestidge, D.S. (1991) *CABIOS* **7**, 203-206.
 32. Racusen, D. (1979) Glycoprotein detection in polyacrylamide gel with thymol and sulfuric acid. *Anal. Biochem.* **99**, 474-476.
 33. Rottier, P.J.M. (1995). The coronavirus membrane glycoprotein. . *In: The Coronaviridae*. S.G. Siddell (ed.) Pp. 115-139. Plenum Press, New York.
 34. Sittidilokratana, N., Hodgson, R.A.J., Panyim, S., Cowley, J.A., Jitrapakdee, S., Boonsaeng, V. and Walker, P.J. (2002) The complete *ORF1b*-gene sequence indicates yellow head virus is an invertebrate nidovirus. *Dis. Aquatic Organisms* **50**, 87-93.
 35. Snijder, E.J. and Horzinek, M.C. (1995). The molecular biology of toroviruses. *In: The Coronaviridae*. S.G. Siddell (ed.) Pp. 219-238. Plenum Press, New York.
 36. Spann, K.M., Vickers, J.E. and Lester, R.J.G. (1995) Lymphoid organ virus of *Penaeus monodon* from Australia. *Diseases of Aquatic Organisms* **23**, 127-134.
 37. Spann, K.M., Cowley, J.A., Walker, P.J. and Lester, R.J.G. (1997) Gill-associated virus (GAV), a yellow head-like virus from *Penaeus monodon* cultured in Australia. *Diseases of Aquatic Organisms* **31**, 169-179.
 38. Spaan, W., Cavanagh, D. and Horzinek, M.C. (1988) Coronavirus structure and genome expression. *J. Gen. Virol.* **69**, 2939-2952.
 39. Stephens, E.B. and Compans, R.W. (1988) Assembly of animal viruses at cellular membranes. *Annu. Rev. Microbiol.* **42**, 489-516.
 40. Sturman, L.S., Ricard, C.S. and Holmes, K.V. (1985) Proteolytic cleavage of the E2 glycoprotein of murine coronavirus: activation of cell-fusing activity of virions by trypsin and separation of two different 90K cleavage fragments. *J. Virol.* **56**, 904-911.
 41. Suzuki, H. and Taguchi, F. (1996) Analysis of the receptor binding site of murine coronavirus spike glycoprotein. *J. Virol.* **70**, 2632-2636.

42. Thompson, J.D. Higgins, D.G. and Gibson, T.L. (1994) *Nucleic Acid Res.* **22**, 4673-4680.
43. Tang, K. F.-J. and Lightner, D.V. (1998) A yellow head virus probe: application to in situ hybridization and determination of its nucleotide sequence. *Diseases of Aquatic Organisms* **35**, 165-173.
44. Wang, Y.-C. and Chang, P.-S. (2000) Yellow head virus infection in the giant tiger prawn *Penaeus monodon* cultured in Taiwan. *Fish Pathology* **35**, 1-10.
45. Wongteerasupaya, C., Sriurairatana, S., Vicker, J.E., Akrajamorn, S., Boonsaeng, V., Panyim, S., Tassanakajon, A., Withyachumnarnkul, B. And Flegel, T.W. (1995) Yellow-head virus of *Penaeus monodon* is an RNA virus. *Diseases of Aquatic Organisms* **22**, 45-50.

Figure legends

Figure 1. Analysis of structural proteins of YHV. Purified YHV was subjected to 15% discontinuous SDS-PAGE and stained with Coomassie brilliant blue (A) or transferred to nitrocellulose and stained using ECL glycoprotein detection system (B) or thymol (C). M, molecular weight protein markers; 1, purified YHV; 2, hemocyanin from shrimp hemolymph. Arrows indicate three distinct sizes of YHV proteins.

Figure 2. Schematic diagram of the organization of GAV and YHV genomes. Boxes represent the open reading frames (ORF) of subgenomic RNA of each virus spanning over 25 kb of genomic RNA. The discontinuous boxes between ORF1a and ORF1b represent the overlapped (-1) ribosomal frameshift site (Cowley *et al.*, 2000; Sithidilokratana *et al.*, 2002). N, nucleocapsid; S protein, structural proteins.

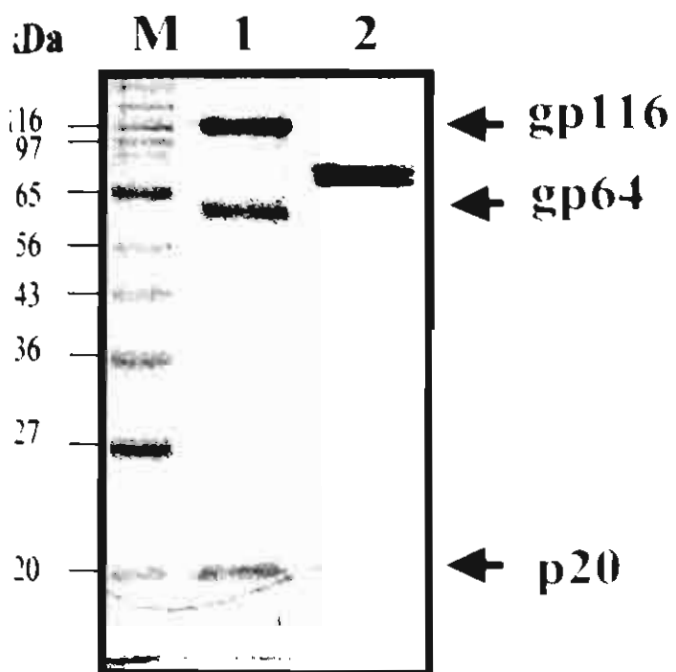
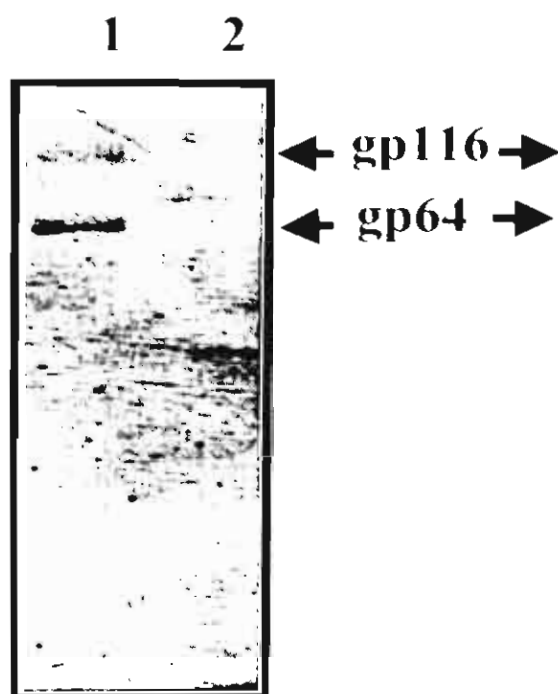
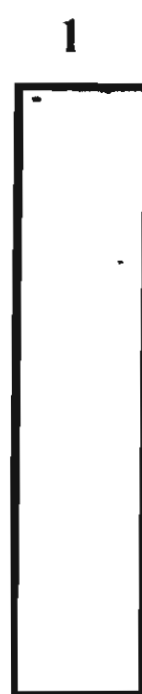
Figure 3. Nucleotide sequence and deduced amino acid sequence of YHV gp116 and gp64. Amino acid residues are numbered on the right. The amino acid residues identified by N-terminal sequencing of native gp116 and gp64 are in bold type. The predicted transmembrane helices are underlined. The potential O-linked glycosylation sites are shown in boxes. The proteolytic cleavages are shown by arrows. CCAAT box and TATA-like element are shown by double line underneath text.

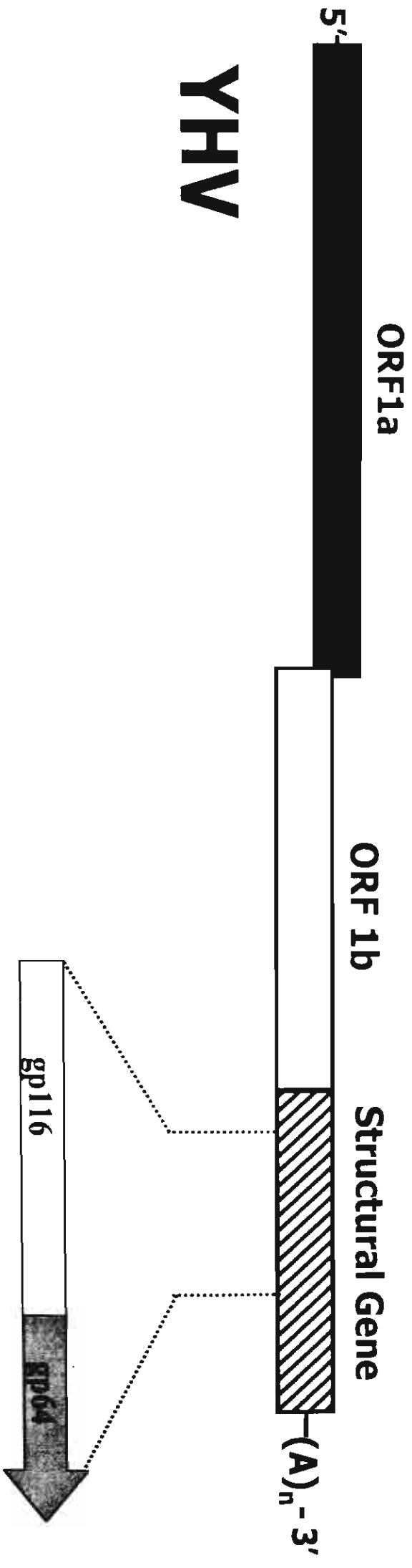
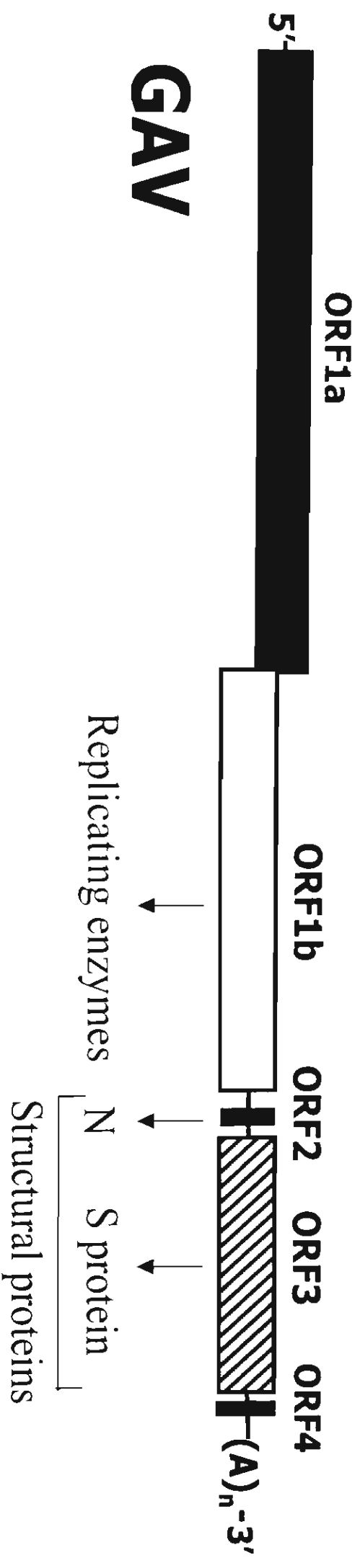
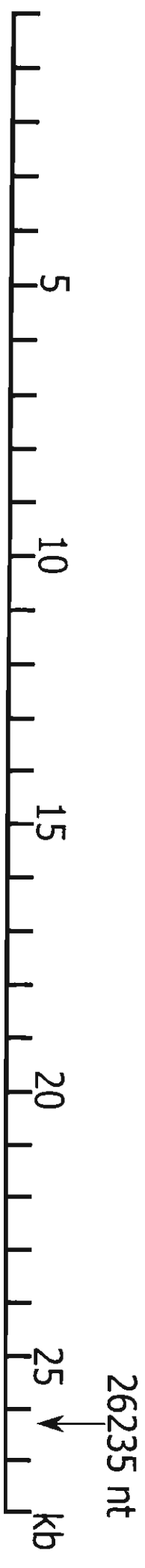
Figure 4. Predicted topology of YHV gp116 and gp64. Topology predictions were determined using TMHMM (Krogh *et al.*, 2001). Cylinders with the number represent the transmembrane domains while the solid lines represent ectodomain or intra-virion domains. Arrows represent the proteolytic cleavage sites.

Figure 5. Comparison of protein encoded from ORF3 of YHV and GAV. Both sequences were aligned using Clustal W. (Thompson *et al.*, 1994). The conserved amino acids are shown as white text against black boxes. The residue number is also indicated.

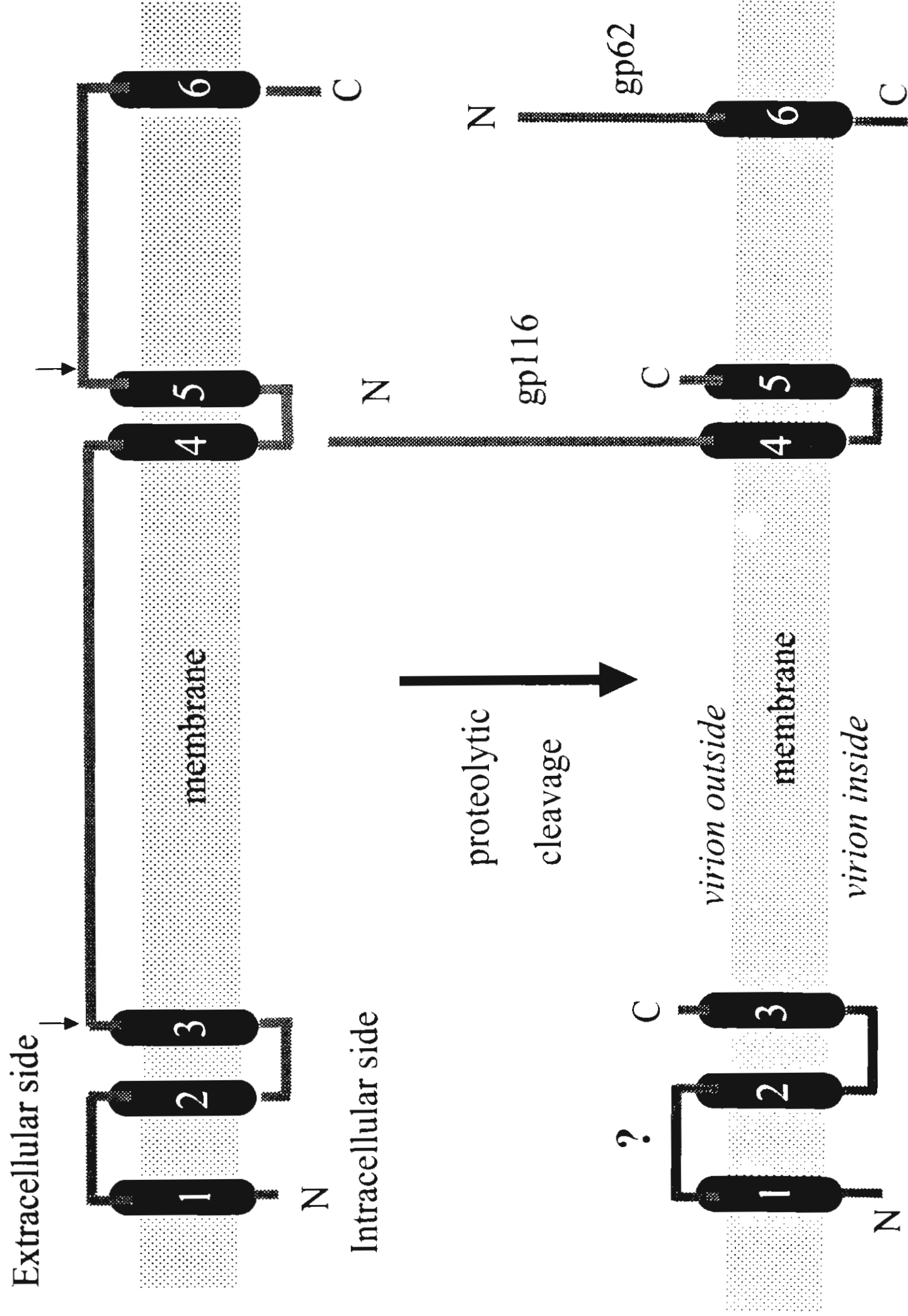
Figure 6. Expression of recombinant YHV gp64 with and without transmembrane in *E. coli*. (A) Schematic diagram of gp64 construct with or without transmembrane region. (B) Western blot of analysis of whole cell lysates of *E. coli* BL21(DE3) harbored

pET17b-gp64 induced for with IPTG. Immunoreactive proteins were detected using polyclonal antiserum raised against native gp64 in mice. M, molecular weight protein markers; 1, BL21(DE3) transformed with pET17b and induced for 3 h; 2-6, BL21(DE3) pLys transformed with pET17b-gp64, induced at 0, 1, 3, 5 and 7 h; 7, purified YHV. (C) BL21 (DE3) harbored pET17bgp64 Δ TM (transmembrane deleted). M, molecular weight protein markers; 1, BL21(DE3) transformed with pET17b and induced for 3 h.; 2-4, BL21 (DE3) transformed with pET17bgp64 Δ TM and induced for 1, 3 and 5 h. 5, purified YHV. Arrows on Western blot indicate the positions of the bands corresponding to the native gp64 and bacterially expressed gp64 with and without transmembrane region.

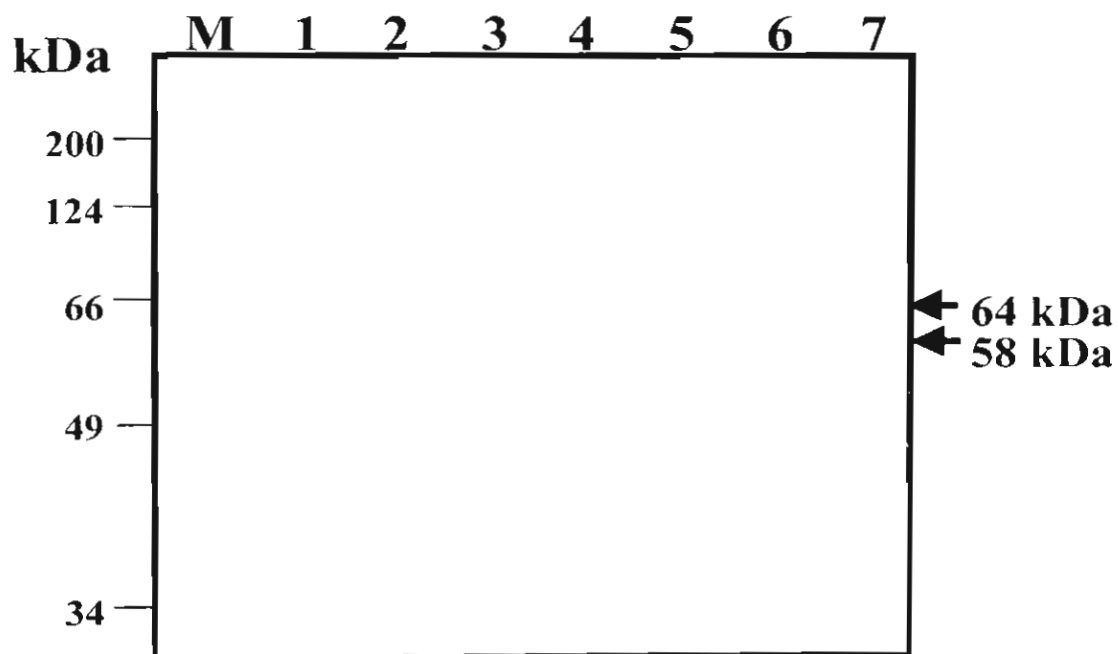
A**B****C**



[illegible]



YHV	1	MQCSRSFTMHFS	SFRSQPFSSSTGLRGLFTLLS	TILGSGHGHLOFATPSHSLDAGARTNS	IILSTCNNTITIKESCHLSSIN	SVSCYDEHHD	ISAPGSHIITYTIGFAYLK	120
GAV	1	MOCSSRSFTLSH	SSHFPASLSFTGLRGIFLIFLLS	ALGSGHGHLOFATPSRSLDAGAIRTNS	IILSTCNNTITIKESCHLSSIN	SVSCYDEHHD	ISAPGSHIITYTIGFAYLK	120
YHV	121	TDYECKEGYPPAN	TVQRFIAILYISICSLAYIFTKITPIAKSFISSCFC	TEMIMET	TEETKQGVTVVAHDRHKLTPTS	CHSIIN	STNHLVIGQFTHLQ	240
GAV	121	TDYECKEGYPPAN	TVQRFIAILYISICSLAYIFTKITPIAKSFISSCFC	TEMIMET	TEETKQGVTVVAHDRHKLTPTS	CHSIIN	STNHLVIGQFTHLQ	229
YHV	241	VLMHHLVLEAGQF	TEEDGCIHWAANGDC	HCNINCDWSEHVQMLCHQ	TCNTISSPLQLLLPHNLFRIL	ELSSSDADN	PVQAQDAGCYS	360
GAV	230	---ELETLESS-NHTEF	DMHWAANGDC	HCNINCDWSEHVQMLCHQ	TCNTISSPLQLLLPHNLFRIL	ELSSSDADN	PVQAQDAGCYS	335
YHV	361	FTEQFEMSILDEARYSD	IVSLYRINN	ITSVFGQYNE	MSYLLHGSVPVTCBTP	SESGITTYN	QJSSQILYNH	480
GAV	336	FTEQFEMSILDEARYSE	IAKLYRIQNSTAUGGVYNH	ASYYLHGDAIPITCEPTPE	IIIGTTJHHSIGKQILYNH	KMNVTVDORCKH	SENCWAYYNK	455
YHV	481	EPTLLI	TEFPYPPRDTKTLATHLGP	VRNAGDYQIFLEPGMLGR	TIDGYSYHEIYASTRH	CHYNM	SGNKGINL	600
GAV	456	KPTLLI	TEFPYPPRDTKTLATHLGP	VRNAGDYQIFLEPGMLGR	TIDGYSYHEIYASTRH	CHYNM	SGNKGINL	575
YHV	601	QYTDIEDIPAGFRDPYDF	SVTPSGPVTISVLEEYHDGDS	THE	TAPKRFYIYRIMARLTP	SOVTEHNL	ISTHAAS	720
GAV	576	QYTDIEDIPAGFRDPYDF	SVTPSGPVTISVLEEYHDGDS	THE	TAPKRFYIYRIMARLTP	SOVTEHNL	ISTHAAS	695
YHV	721	NEFNIONDILATFG	TATRTWAEYRHLPHFLTKRGFYPLEPVT	GS	ADYLIVEYNAHASRYSHQATYH	QFGHPVAKQ	IRPGVCPTPRSIRY	839
GAV	696	NDWNIONDILATFG	TATRTWAEYRHLPHFLTKRGFYPLEPVT	EN	ADYLIVEYNAHASRYSHQATYH	QFGHPVAKQ	IRPGVCPTPRSIRY	815
YHV	840	TSGYIFR	YDYIYRFPKFGNGLYLKGVSAAS	ISIGTYSKCKGAQS	ISPYHDHGINTDLGTP	PVYDSACDSAA	YTIIPVVKYNGPYSLGV	959
GAV	816	TSGYIFR	YDYIYRFPKFGNGLYLKGVSAAS	ISIGTYSKCKGAQS	ISPYHDHGINTDLGTP	PVYDSACDSAA	YTIIPVVKYNGPYSLGV	935
YHV	960	HSVTCIN	SDNKNVHVVKQ	FGMSYIYIAGDP	QALHISHNKH	PTYSILKQ	QINL	1079
GAV	936	HSVTCIN	SDNKNVHVVKQ	FGMSYIYIAGDP	QALHISHNKH	PTYSILKQ	QINL	1054
YHV	1090	GLDTRHLKHAR	HKIYIHSPLMLGRTEFLWSPIYA	IIIL	SIISPASALAH	QOARVRS	PGTFRQEP	1199
GAV	1055	GLDTRHLKHAR	HKIYIHSPLMLGRTEFLWSPIY	IM	ALLSPASALAH	QOARVRS	PGTFRQEP	1174
YHV	1200	SSCAYLYTSLPPK	LDARINWSCLHTGSCCKNSSEY	LEH	PLGQHTSNDYIVANPSN	PLQGLCH	HS	1319
GAV	1175	SSCAYLYTSLPPK	LDARINWSCLHTGSCCKNSSEY	LEH	PLGQHTSNDYIVANPSN	PLQGLCH	HS	1294
YHV	1320	TNECASITTDYRNYTAD	ERTISFNIAHPQVAKTTEFKVGAL	FTQGTTHPKH	IIYDVPKGYDAH	FGSFFSYQAE	TVPTG	1439
GAV	1295	TNECASITTDYRNYTAD	ERTISFNIAHPQVAKTTEFKVGAL	FTQGTTHPKH	IIYDVPKGYDAH	FGSFFSYQAE	TVPTG	1414
YHV	1440	PLGAIVQCNYDQ	YIYSTIIRQOORS	ILNGVTVDEVDAL	SYSLLNPHHCDFGAVNIHFA	YSTTAHVS	ILIEGDPDQ	1559
GAV	1415	PLGAIVQCNYDQ	YIYSTIIRQOORS	ILNGVTVDEVDAL	SYSLLNPHHCDFGAVNIHFA	YSTTAHVS	ILIEGDPDQ	1534
YHV	1560	TRADHLCNFTAS	SEHEDLVNKGKAM	ITTEVITQCDVGS	ISDTIVGAAGND	YGTFTTEHAF	GKATWDY	1665
GAV	1535	TRADHLCNFTAS	SEHEDLVNKGKAM	ITTEVITQCDVGA	ISDTIVGAAGND	YGTFTTEHAF	GKATWDY	1640

A**B****C**