



รายงานวิจัยฉบับสมบูรณ์

โครงการ การโคลน และการแสดงออกของโปรตีนเซลล์จากหอยเชอรี่
(*Pomacea canaliculata*)

โดย

อาจารย์ ดร. จันทร์ประภา อิ่มจงใจรัก
และ รองศาสตราจารย์ ดร. ศิริพร สิทธิประณีต

ระยะเวลาสิ้นสุดโครงการ กรกฎาคม 2551

รายงานวิจัยฉบับสมบูรณ์

โครงการ การโคลน และการแสดงออกของโปรตีนเซลลูเลสจากหอยเชอรี่
(*Pomacea canaliculata*)

คณะผู้วิจัย

สังกัด

1. อาจารย์ ดร. จันทร์ประภา อิ่มจงใจรัก

ภาควิชาเทคโนโลยีทางอาหาร
คณะวิทยาศาสตร์

จุฬาลงกรณ์มหาวิทยาลัย

2. รองศาสตราจารย์ ดร. ศิริพร สิทธิประณีต

ภาควิชาชีวเคมี

คณะวิทยาศาสตร์

จุฬาลงกรณ์มหาวิทยาลัย

สนับสนุนโดยสำนักงานคณะกรรมการอุดมศึกษา และสำนักงานกองทุนสนับสนุนการวิจัย

(ความเห็นในรายงานนี้เป็นของผู้วิจัย สกอ. และ สกว. ไม่จำเป็นต้องเห็นด้วยเสมอไป)

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัยประเภททุนพัฒนาศักยภาพในการทำงานวิจัยของอาจารย์รุ่นใหม่ ปี 2549 จากสำนักงานคณะกรรมการการอุดมศึกษา (สกอ.) และสำนักงานกองทุนสนับสนุนการวิจัย (สกว.) ผู้วิจัยขอขอบพระคุณมา ณ ที่นี้ด้วย

ผู้วิจัยขอขอบคุณคณาจารย์และบุคลากรภาควิชาเทคโนโลยีทางอาหาร และภาควิชาชีวเคมี คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่มีส่วนช่วยให้โครงการนี้สำเร็จลุล่วงไปด้วยดี

ผู้วิจัยขอขอบคุณ ว่าที่ร้อยตรี ดร. ปิติ อ่ำพ่ายพ์ นักวิจัยของศูนย์พันธุวิศวกรรมและเทคโนโลยีชีวภาพแห่งชาติ ที่ได้ให้คำแนะนำที่เป็นประโยชน์ต่อการทำงานวิจัย

สารบัญ

	หน้า
1. กิตติกรรมประกาศ	3
2. บทคัดย่อภาษาไทย	5
3. บทคัดย่อภาษาอังกฤษ	6
4. เนื้อหางานวิจัย	
4.1 บทนำ	7
4.2 วิธีการทดลอง	8
4.3 ผลการทดลองและวิจารณ์	13
4.4 หนังสืออ้างอิง	31
5. Output จากโครงการวิจัยที่ได้รับทุนจาก สกว.	33
6. ภาคผนวก	34

บทคัดย่อ

รหัสโครงการ : MRG4980168
ชื่อโครงการ : การโคลน และการแสดงออกของโปรตีนเซลลูเลสจากหอยเชอรี่ (*Pomacea canaliculata*)
ชื่อนักวิจัย : อาจารย์ ดร. จันทร์ประภา อิ่มจงใจรัก และ
รองศาสตราจารย์ ดร. ศิริพร สิทธิประณีต
E-mail Address : chanprapa.i@chula.ac.th
ระยะเวลาโครงการ: 1 กรกฎาคม 2549 – 30 มิถุนายน 2551

ในงานวิจัยนี้ได้ค้นหาลำดับนิวคลีโอไทด์ที่สมบูรณ์ ศึกษาลักษณะสมบัติ วิเคราะห์การจัดเรียงตัวของยีนและศึกษาการแสดงออกของยีน glycosyl hydrolase family 10 ในหอยเชอรี่ (*Pomacea canaliculata*) จากการวิเคราะห์ลำดับนิวคลีโอไทด์พบยีน 2 ไอโซฟอร์ม (GHF10-Pc1 และ GHF10-Pc3) มีขนาด 1300 bp และ 1277 bp ตามลำดับ โดยทั้ง 2 ไอโซฟอร์มมี open reading frame (ORF) ที่สมบูรณ์มีขนาด 1188 และ 1191 bp ตามลำดับ ซึ่งสามารถถอดรหัสให้โปรตีนที่มีขนาด 395 และ 396 กรดอะมิโน เมื่อนำลำดับกรดอะมิโนของยีนมาเปรียบเทียบกับความเหมือนกับยีนที่มีรายงานไว้พบว่าลำดับของกรดอะมิโนของยีนที่ได้มีความคล้ายกับยีนเซลลูเลสในหอย *Ampullaria crosseana* 98% และ 82% ตามลำดับ และมีความคล้ายกับเซลลูเลสในกลุ่ม glycosyl hydrolase family 10 (GHF10) จากการศึกษาการจัดเรียงตัวของยีนพบว่า GHF10-Pc1 และ GHF10-Pc3 มีขนาด 4937 และ 4512 bp ประกอบด้วย 9 exon และ 8 intron ตามลำดับ จากการตรวจสอบการแสดงออกของยีนในหอยเชอรี่ระยะต่าง ๆ เช่นไข่หอย และลูกหอยอายุ 1 และ 10 วัน พบว่าไม่มีการแสดงออกของยีนทั้งสองในระยะไข่หอย และพบว่ามีแสดงออกของ GHF10-Pc1 ในลูกหอยระยะ 1 และ 10 วัน ส่วนยีน GHF10-Pc3 พบมีการแสดงออกในระยะ 1 วัน จากการศึกษาการแสดงออกของยีนเซลลูเลสใน *E. coli* โดยเหนี่ยวนำให้เกิดการแสดงออกของยีนด้วย IPTG และวิเคราะห์รีคอมบิแนนท์โปรตีนด้วย SDS-PAGE พบว่ามีการแสดงออกของรีคอมบิแนนท์โปรตีนอยู่ในส่วน inclusion bodies ซึ่งสามารถทำให้บริสุทธิ์ได้โดยใช้ Ni-NTA column chromatography เมื่อวิเคราะห์แถบโปรตีนด้วย SDS-PAGE พบว่ารีคอมบิแนนท์โปรตีนที่ได้มีขนาด 44 kDa ซึ่งเมื่อนำรีคอมบิแนนท์โปรตีนไปตรวจสอบคุณสมบัติของเอนไซม์พบว่ารีคอมบิแนนท์โปรตีนมีแอกติวิตี้ของเอนไซม์ endo- β -1,4-xylanase

คำหลัก : *Pomacea canaliculata*, เซลลูเลส, โคลน, หอยเชอรี่,

Abstract

Project Code : MRG4980168
Project Title : Molecular cloning and expression of cellulase protein from golden apple snail (*Pomacea canaliculata*)
Investigator : Dr. Chanprapa Imjongjirak and
Associate Professor Dr. Siriporn Sittipraneed
E-mail Address : chanprapa.i@chula.ac.th
Project Period : 1 July 2006 – 30 June 2008

Two cellulase cDNAs (GHF10-*Pc1* and GHF10-*Pc3*) belonging to glycoside hydrolase family 10 (GHF10) were successfully isolated and characterized from stomach tissue of golden apple snail (*Pomacea canaliculata*), a kind of herbivorous mollusca. Sequencing analysis revealed full-length cDNAs of 1300 and 1277 bp in length, respectively. The open reading frame (ORF) of cellulase cDNA was 1188 and 1191 bp, encoding 395 and 396 amino acid, respectively. Sequence alignment revealed that GHF10-*Pc1* and GHF10-*Pc3* shared high identity with glycosyl hydrolase family 10 (GHF10) and had an overall similarity of 98 and 82% to those of *Ampullaria crosseana* cellulase EGX. A neighbour-joining tree showed a clear differentiation between each species and also indicated that GHF10-*Pc1* and GHF10-*Pc3* from *P. canaliculata* and *A. crosseana* EGX are closely related phylogenetically. The genomic organization of cellulase GHF10-*Pc1* and GHF10-*Pc3* genes was also investigated. The GHF10-*Pc1* and GHF10-*Pc3* genes spanned over 4937 and 4512 bp, respectively. Both genes contained 9 exons interrupted by eight introns. The result verified the endogenous origin of the GHF10-*Pc1* and GHF10-*Pc3* genes. Analysis of RNA by RT-PCR from several ages of *P. canaliculata* revealed that neither gene was expressed in eggs. GHF10-*Pc1* was also expressed in 1- and 10-day-old juvenile snails whereas GHF10-*Pc3* was expressed only in 1-day-old juvenile snails. The result showed that two GHF10-*Pc* transcripts were developmentally expressed. The mature GHF10-*Pc1* was cloned into the expression vector pET28b with an *N*-terminal hexa-histidine tag fused in-frame, and expressed in *E. coli* as inclusion body. The recombinant (r)GHF10-*Pc1* was successfully purified by Ni-NTA chromatography. The purified rGHF10-*Pc1* showed activity of endo- β -1,4-xylanase.

Keywords : *Pomacea canaliculata*, cellulase gene, cloning, golden apple snail

4. เนื้อหางานวิจัย

4.1 บทนำ

เซลลูโลส (cellulose) เป็นโพลิเมอร์เชิงเส้นตรงของน้ำตาลกลูโคสซึ่งต่อกันด้วยพันธะไกลโคซิดิกที่ตำแหน่ง β -D-(1-4) มีคุณสมบัติไม่ละลายน้ำ ทนต่อปฏิกิริยาของเอนไซม์กรด และต่างที่เจือจาง เซลลูโลสจัดเป็นโพลิเมอร์ที่พบมากที่สุดในธรรมชาติ และเป็นส่วนประกอบหลักของผนังเซลล์พืช โดยรวมตัวอยู่กับไซแลน (xylan) และลิกนิน (lignin) ในการย่อยสลายเซลลูโลสต้องใช้เอนไซม์หลายชนิดที่ทำงานร่วมกัน ได้แก่ เซลลูเลส (cellulase) ซึ่งประกอบด้วย $\text{exo-}\beta$ -1,4-glucanase (E.C.3.2.9.11) เอนไซม์ $\text{endo-}\beta$ -1,4-glucanase (E.C.3.2.1.4) เอนไซม์ β -glucosidase (E.C. 3.2.1.21) และเอนไซม์ไซแลเนสซึ่งประกอบด้วย $\text{endo-}\beta$ -1,4-xylanase (E.C.3.2.1.8) และเอนไซม์ β -xylosidase (E.C.3.2.1.37)

ความเชื่อแต่เดิมที่ว่าสัตว์ไม่สามารถย่อยเซลลูโลสได้เนื่องจากไม่สามารถสร้างเอนไซม์ย่อยเซลลูโลสขึ้นได้เอง แต่เอนไซม์ที่ช่วยย่อยเซลลูโลสจะได้มาจากแบคทีเรียที่อยู่ในระบบทางเดินอาหาร (symbiotic gut microorganisms) (Cleveland 1924) ทำให้การวิจัยที่ผ่านมาเป็นการศึกษาเอนไซม์ในกลุ่มเซลลูเลสในสิ่งมีชีวิตที่เป็นจุลินทรีย์พวก รา และแบคทีเรียเป็นส่วนใหญ่ แต่เมื่อไม่นานมานี้ได้มีรายงานการค้นพบเอนไซม์ $\text{endo-}\beta$ -1,4-glucanase ในสัตว์ไม่มีกระดูกสันหลังหลายชนิด เช่น ปลวก หนอนตัวกลม แมลงสาบ กุ้งน้ำจืด แมลงปีกแข็ง และในหอย รายงานการวิจัยเอนไซม์ในกลุ่มเซลลูเลสในหอยเริ่มจากในปี 1977, Purchon ตรวจพบแอคติวิตีของเอนไซม์เซลลูเลสในน้ำย่อยในกระเพาะของสิ่งมีชีวิตพวกหอยสองฝา ในปี 2000 Xu และคณะ ได้แยกและศึกษาลักษณะสมบัติของเอนไซม์ $\text{endo-}\beta$ -1,4-glucanase จาก blue mussel, *Mytilus edulis* และในปี 2003 Wang และคณะได้ทำการโคลนยีนเซลลูเลสจากหอย *Ampullaria crosseana* และทำการแสดงออก (expression) เพื่อผลิตรีคอมบิแนนท์โปรตีน และศึกษาแอคติวิตีของเอนไซม์ดังกล่าว พบว่ายีนเซลลูเลสในหอยชนิดนี้มี 3 activity คือ $\text{exo-}\beta$ -1,4-glucanase, $\text{endo-}\beta$ -1,4-glucanase และ $\text{endo-}\beta$ -1,4-xylanase

หอยเชอรี่ (*Pomacea canaliculata*) หรือหอยโข่งอเมริกาใต้ หรือเป่าฮื้อน้ำจืด เป็นหอยฝาเดียวจัดอยู่ในวงศ์เดียวกับหอยโข่ง (apple snail, *Pila spp.*) หอยเชอรี่มีถิ่นกำเนิดในทวีปอเมริกาใต้ แพร่กระจายสู่ทวีปเอเชียทางประเทศญี่ปุ่น ไต้หวัน และฟิลิปปินส์ ในประเทศไทยมีคนนำมาขายตามร้านตู้ปลาเป็นหอยสวยงาม เนื้อหอยเชอรี่มีคุณค่าทางอาหารสูง โดยมีโปรตีนสูงถึง 34-53% จึงมีการนำเข้ามาขายพันธุ์และเพาะเลี้ยงเป็นการค้า แต่เมื่อไม่ได้รับความนิยม ทำให้มีการทิ้งหอยปล่อยให้แพร่ลงสู่แหล่งน้ำตามธรรมชาติจนเกิดการระบาดของหอยในนาสร้างความเสียหายให้แก่เกษตรกรเป็นอย่างมาก

หอยเชอรี่สามารถเจริญเติบโตและขยายพันธุ์ได้อย่างรวดเร็ว โดยหอยเพศเมียสามารถวางไข่ได้ตลอดปี ลูกหอยอายุเพียง 2-3 เดือน สามารถจับคู่ผสมพันธุ์ได้ และหลังจากผสมพันธุ์ตัวเมียจะคลานไปวางไข่ตามที่แห้งเหนือน้ำในเวลากลางคืน ไข่จะฟักออกเป็นตัวหอยภายใน 7-12 วัน หอยเชอรี่กินพืชน้ำได้เกือบทุกชนิด เช่น แหน แหนแดง ไข่น้ำ ผักบุ้ง ผักกระเฉด

สาหร่ายต่าง ๆ ยอดอ่อนผักตบชวา โดยสามารถกินได้อย่างรวดเร็วและกินได้ตลอดเวลา คิดเป็นน้ำหนักอาหารที่กิน 50% ของน้ำหนักตัวต่อวัน

เอนไซม์ในกลุ่มเซลล์สามารถแบ่งออกได้หลายกลุ่มโดยอาศัยความแตกต่างของโครงสร้าง 3 มิติของลำดับกรดอะมิโนในบริเวณ catalytic domain ในปัจจุบันยังไม่มีรายงานการศึกษาของกลุ่มเอนไซม์เซลล์ในหอยเชอรี่ ในขณะที่เริ่มมีรายงานการค้นพบเอนไซม์ในกลุ่มเซลล์ในหอยชนิดอื่น ๆ ดังนั้นในโครงการวิจัยนี้ จึงสนใจที่จะศึกษายีนเซลล์ในหอยเชอรี่ โดยทำการโคลน ลำดับนิวคลีโอไทด์ ศึกษาลักษณะสมบัติของยีนเซลล์จากหอยเชอรี่ในประเทศไทย และศึกษาการผลิตเอนไซม์โดยเทคนิครีคอมบิแนนท์โปรตีน

วัตถุประสงค์

1. ค้นหาลำดับนิวคลีโอไทด์ที่สมบูรณ์ (full length cDNA) ของยีนเซลล์จากหอยเชอรี่ และศึกษาลักษณะสมบัติของยีนเซลล์ในหอยเชอรี่
2. แยก และวิเคราะห์การจัดเรียงตัวของยีน (genomic organization) ในส่วนของ exon และ intron ของยีนเซลล์ในหอยเชอรี่ โดยใช้เทคนิค PCR
3. ศึกษาการแสดงออกของยีนเซลล์ใน *E. coli* สกัตรีคอมบิแนนท์โปรตีน (recombinant protein) และแยกบริสุทธิ์เพื่อนำไปศึกษาคุณสมบัติของเอนไซม์เซลล์

4.2 วิธีการดำเนินงานวิจัย

4.2.1 การเตรียมตัวอย่าง

เก็บตัวอย่างหอยเชอรี่ตัวเต็มวัย (*Pomacea canaliculata*) และนำมาตัดส่วนของกระเพาะอาหาร แชลงในไนโตรเจนเหลว (liquid nitrogen) และเก็บไว้ที่ -80°C เพื่อใช้ในการศึกษาต่อไป

4.2.2 สกัด total RNA และ mRNA จากกระเพาะอาหารของหอยเชอรี่

สกัด total RNA จากส่วนกระเพาะของหอยเชอรี่ด้วย Trizol Reagent (Gibco-BRL, USA) โดยนำเนื้อเยื่อของตัวอย่างมาบดโดยเติม liquid nitrogen และบดตัวอย่างให้ละเอียด นำเนื้อเยื่อที่บดแล้วมาใส่ในหลอด 1.5 ml และเติม Trizol Reagent ในอัตราส่วน 1 ml ของ Trizol Reagent ต่อเนื้อเยื่อประมาณ 50-100 mg ใช้แท่งแก้วบดให้ละเอียดอีกครั้ง และนำไป vortex เป็นเวลา 2-5 นาทีนำไปปั่นที่ 12000 rpm ที่ 4°C เป็นเวลา 10 นาที เพื่อกำจัดเศษเซลล์ ต่อจากนั้นดูดสารละลายส่วนบนใส่ในหลอด 1.5 ml ใหม่และเติม 0.2 ml ของ chloroform ต่อ 1 ml ของ Trizol Reagent และนำไป vortex เป็นเวลา 2-5 นาที นำไปปั่นที่ 12000 rpm ที่ 4°C เป็นเวลา 10 นาที ดูดสารละลายส่วนบนใส่ในหลอด 1.5 ml หลอดใหม่ นำไปไว้บนน้ำแข็ง จากนั้นตกตะกอน RNA โดยเติม 0.5 ml ของ isopropanol ต่อ 1 ml ของ Trizol Reagent บ่มที่ -20°C นาน 30 นาที นำไปปั่นที่ 12000 rpm ที่ 4°C เป็นเวลา 10 นาที ทิ้งสารละลายส่วนบน ล้าง RNA pellet โดยเติม 1 ml ของ 75% ethanol นำไปปั่นที่ 10000 rpm ที่ 4°C เป็นเวลา 5 นาที ทิ้งสารละลายส่วนบน นำ RNA pellet ไป air-dried โดยใช้ vacuum เป็นเวลา 3-5 นาที

ละลาย RNA pellet ด้วย RNase-free water เก็บ RNA ที่ได้ไว้ที่ -70°C วิเคราะห์ปริมาณและคุณภาพของ RNA ด้วยเทคนิค spectrophotometry จากนั้นกำจัด DNA ที่อาจปนเปื้อนมาโดยการย่อย DNA ด้วย RNase-free DNase I และสกัดแยก mRNA จาก total RNA โดยใช้ oligo (dT) cellulose column จาก QuickPrep mRNA purification kit (Amersham Biosciences, USA) และวิเคราะห์ปริมาณและคุณภาพของ mRNA ด้วยเทคนิค spectrophotometry

4.2.3 ค้นหายีนเซลล์โดยเทคนิค Reverse Transcription Polymerase Chain Reaction (RT-PCR)

4.2.3.1 ออกแบบไพรเมอร์

นำลำดับนิวคลีโอไทด์ และลำดับกรดอะมิโนของยีนเซลล์ที่มีรายงานไว้ใน GenBank มาทำ alignment โดยใช้โปรแกรม ClustalX เพื่อหาบริเวณที่สามารถใช้ออกแบบ degenerate primers ที่จะนำมาใช้ทำ RT-PCR

4.2.3.2 สังเคราะห์ first-strand cDNA

สังเคราะห์ first-strand cDNA โดยใช้ ImProm-IITM Reverse Transcription System (Promega Corporation Madison, Wisconsin, USA) ซึ่งจะประกอบด้วย 2 μg mRNA, oligo (dT)₁₂₋₁₆ primer และเอนไซม์ reverse transcriptase บ่มที่อุณหภูมิ 42°C เป็นเวลา 1 ชั่วโมง จากนั้นนำมาทำ PCR โดยใช้ไพรเมอร์ที่ออกแบบไว้ โดยทำปฏิกิริยา PCR ในปริมาตร 25 ไมโครลิตร ซึ่งประกอบด้วย 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 2 mM MgCl_2 , 100 μM ของ dATP, dTTP, dCTP และ dGTP, 0.4 μM ของแต่ละ primer และ 1 U ของ Taq DNA Polymerase ทำการเพิ่มปริมาณ DNA โดยมี PCR profiles ดังนี้คือ Predenaturation 94°C เป็นเวลา 3 นาที และตามด้วย Denaturation 94°C เป็นเวลา 1 นาที Annealing 48°C เป็นเวลา 1 นาที Extension 72°C เป็นเวลา 2 นาที จำนวน 5 รอบ และ Denaturation 94°C เป็นเวลา 1 นาที Annealing 58°C เป็นเวลา 1 นาที Extension 72°C เป็นเวลา 2 นาที จำนวน 35 รอบ ตามด้วย Final extension 72°C เป็นเวลา 10 นาที จำนวน 1 รอบ วิเคราะห์ PCR products ที่ได้ใน 1.4% agarose gel electrophoresis และย้อมด้วย ethidium bromide ตรวจสอบการเรืองแสงของ DNA โดย UV transilluminator

4.2.3.3 เชื่อมต่อชิ้นยีนเซลล์กับพลาสมิด pGEM[®]-T easy

คัดเลือกแถบ DNA และตัดแถบ DNA ของยีนเซลล์ที่ได้จากการทำ RT-PCR ออกจากอะกาโรสเจล แยกชิ้น DNA จากเจล โดยใช้ Qiaquick gel extraction kit (Qiagen, USA) จากนั้นนำมาทำการโคลนโดยวิธี TA-cloning โดยเชื่อมต่อกับพลาสมิด pGEM[®]-T easy แล้วนำเข้าสู่เซลล์ *E. coli* โดยวิธี electroporation

4.2.3.4 คัดเลือกรีคอมบิแนนท์โคลน และการหาลำดับนิวคลีโอไทด์

หาขนาดของชิ้น DNA insert โดยสุ่มคัดเลือก recombinant clones มาหาขนาดของ cDNA insert โดยใช้เทคนิค colony PCR โดยใช้ไพรเมอร์ pUC1 primer: 5'-CCGGCTCGTATGTTGTGTGGA-3' และ pUC2 primer: 5'-GTGGTGCAAGGCGA TTAAGTTGG-3' และการตัดด้วย restriction enzyme วิเคราะห์ขนาดของ DNA insert ใน

1.4% agarose gel electrophoresis จากนั้นสกัดพลาสมิดดีเอ็นเอของรีคอมบิแนนท์โคลน แล้วนำไปหาลำดับนิวคลีโอไทด์โดยใช้ automated DNA sequencer (ABI377, PE Applied Biosystems, USA) ทำการออกแบบไพรเมอร์จากข้อมูลลำดับนิวคลีโอไทด์ที่ได้ เพื่อนำไปใช้หาลำดับนิวคลีโอไทด์ full length ของยีนเซลล์ลูเลส

4.2.4 การหา full length cDNA ของยีนเซลล์ลูเลส โดยใช้เทคนิค Rapid Amplification of cDNA End (RACE)-PCR

4.2.4.1 3' RACE-PCR

นำ first-strand cDNA ที่ได้มาทำ 3' RACE-PCR (SMART RACE cDNA amplification kit, BD-Biosciences Clontech, USA) โดยใช้ไพรเมอร์ที่ออกแบบมาจากลำดับนิวคลีโอไทด์บริเวณ 5' UTR ของยีนเซลล์ลูเลสที่ได้และ oligo(dT)-Not I primer ทำการเพิ่มปริมาณ DNA โดยมี PCR profiles คือ Predenaturation 92 °C เป็นเวลา 3 นาที และตามด้วย Denaturation 94 °C เป็นเวลา 45 วินาที Annealing 65 °C เป็นเวลา 45 วินาที Extension 72 °C เป็นเวลา 1.30 นาที จำนวน 5 รอบ และ Denaturation 94 °C เป็นเวลา 45 วินาที Annealing 60 °C เป็นเวลา 45 วินาที Extension 72 °C เป็นเวลา 1.30 นาที จำนวน 25 รอบ ตามด้วย Final extension 72 °C เป็นเวลา 10 นาที จำนวน 1 รอบ วิเคราะห์ PCR product ใน agarose gel electrophoresis ตรวจสอบการเรียงแสงของ DNA โดย UV transilluminator

4.2.4.2 คัดเลือกรีคอมบิแนนท์โคลน และวิเคราะห์หาลำดับนิวคลีโอไทด์

ตัดแถบ DNA ของยีนเซลล์ลูเลสที่ได้จากการทำ 3' RACE-PCR ออกจากอะกาโรสเจล แยกชิ้น DNA จากเจล โดยใช้ Qiaquick gel extraction kit (Qiagen, USA) จากนั้นนำมาเชื่อมต่อกับพลาสมิด pGEM[®]-T easy แล้วนำเข้าสู่เซลล์ *E. coli* โดยวิธี electroporation หาขนาดของชิ้น DNA insert โดยวิธี colony PCR และการตัดด้วย restriction enzyme วิเคราะห์ขนาดของ DNA insert โดยวิธี agarose gel electrophoresis สกัดพลาสมิดดีเอ็นเอของรีคอมบิแนนท์โคลน แล้วนำไปหาลำดับนิวคลีโอไทด์โดยใช้ automated DNA sequencer (ABI377, PE Applied Biosystems, USA)

4.2.4.3 เพิ่มปริมาณของ full length cDNA ของยีนเซลล์ลูเลส

ออกแบบไพรเมอร์จากลำดับนิวคลีโอไทด์บริเวณ 5' UTR และบริเวณ 3' UTR ของยีนเซลล์ลูเลส และนำมาทำ PCR โดยใช้ first-strand cDNA ที่เตรียมจากกระเพาะอาหารของหอยเชอรี่เป็น template วิเคราะห์ PCR product ใน agarose gel electrophoresis แยกชิ้น DNA ของ full length ของยีนเซลล์ลูเลสออกจากอะกาโรสเจล แล้วนำมาเชื่อมต่อกับพลาสมิด pGEM[®]-T easy แล้วนำเข้าสู่เซลล์ *E. coli* ทำการคัดเลือกรีคอมบิแนนท์โคลน วิเคราะห์ขนาดของ DNA insert โดยวิธี colony PCR และ agarose gel electrophoresis สกัดรีคอมบิแนนท์พลาสมิด และวิเคราะห์หาลำดับนิวคลีโอไทด์

4.2.5 ศึกษาลักษณะสมบัติของยีนเซลล์ลูเลส

วิเคราะห์ข้อมูลของลำดับนิวคลีโอไทด์ที่ได้โดยใช้โปรแกรม GENETYX software (Software Development Inc.) และโปรแกรม SMART analysis (<http://smart.embl->

heidelberg.de/) จากนั้นนำข้อมูลที่ได้นำไปเปรียบเทียบกับข้อมูลนิวคลีโอไทด์ที่มีรายงานในธนาคารยีน (GenBank) โดยใช้โปรแกรม BLASTX และ BLASTN (<http://www.ncbi.nlm.nih.gov>) และเปรียบเทียบความเหมือนของลำดับกรดอะมิโนโดยใช้โปรแกรม Clustal W และสร้างแผนภาพความสัมพันธ์เชิงวิวัฒนาการ (phylogenetic tree)

4.2.6 การแยก และศึกษาการจัดเรียงตัวของยีนเซลล์ูลเลส

4.2.6.1 สกัดจีโนมิกดีเอ็นเอ

สกัดดีเอ็นเอจากส่วนเท้าของหอยเชอรี่โดยใช้วิธี phenol-chloroform-sodium dodecyl sulfate (SDS) วิเคราะห์ปริมาณและคุณภาพของดีเอ็นเอด้วยเทคนิค spectrophotometry และ agarose gel electrophoresis

4.2.6.2 แยกยีนเซลล์ูลเลสโดยใช้เทคนิค PCR

ออกแบบไพรเมอร์จากลำดับนิวคลีโอไทด์ของ cDNA ของยีนเซลล์ูลเลส เพื่อนำมาใช้ในการเพิ่มปริมาณของยีนเซลล์ูลเลสโดยใช้จีโนมิกดีเอ็นเอของหอยเชอรี่เป็น template ทำการวิเคราะห์ PCR product ในอะกาโรสเจล และตัดแถบ DNA ของยีนเซลล์ูลเลสออกจากอะกาโรสเจล แยกชิ้น DNA ออกจากเจล จากนั้นนำมาเชื่อมต่อกับพลาสมิด pGEM[®]-T easy แล้วนำเข้าสู่เซลล์ *E. coli* ทำการคัดเลือกกริคอมบิแนนท์โคลน วิเคราะห์ขนาดของ DNA insert โดยวิธี colony PCR และ agarose gel electrophoresis สกัดรีคอมบิแนนท์พลาสมิด และวิเคราะห์หาลำดับนิวคลีโอไทด์

4.2.6.3 ศึกษาลักษณะสมบัติของยีนเซลล์ูลเลส

ศึกษาลักษณะสมบัติของยีนเซลล์ูลเลสจากข้อมูลของลำดับนิวคลีโอไทด์ที่ได้โดยใช้โปรแกรม GENETYX software (Software Development) จากนั้นนำข้อมูลที่ได้นำไปเปรียบเทียบกับข้อมูลนิวคลีโอไทด์ที่มีรายงานในธนาคารยีน (GenBank) โดยใช้โปรแกรม BLASTX และ BLASTN (<http://www.ncbi.nlm.nih.gov>)

4.2.7 ตรวจสอบการแสดงออกของยีนด้วยเทคนิค RT-PCR

สกัด total RNA จากหอยเชอรี่ในระยะต่าง ๆ เช่น ไข่หอย และลูกหอยอายุ 1 และ 10 วัน สร้าง First-stand cDNA โดยใช้ ImProm-IITM Reverse Transcription System และนำ cDNA ที่ได้มาเพิ่มปริมาณด้วยเทคนิค PCR โดยใช้ไพรเมอร์ที่ได้ออกแบบไว้ วิเคราะห์ PCR products ใน 1.8% agarose gel electrophoresis โดยเปรียบเทียบกับยีน EF1- α

4.2.8 ศึกษาการแสดงออกของยีนเซลล์ูลเลสใน *E. coli*

4.2.8.1 โคลนยีนเซลล์ูลเลสเข้าสู่ expression vector และนำเข้าสู่เซลล์ *E. coli* Rosetta (DE3)pLysS

สังเคราะห์ full length ของยีนเซลล์ูลเลสด้วย PCR โดยใช้ข้อมูลลำดับนิวคลีโอไทด์ของยีนเซลล์ูลเลสที่ได้ โดยออกแบบไพรเมอร์ให้มีลำดับนิวคลีโอไทด์ทางด้านปลาย 3' ของ cDNA มี 6x histidine tag ที่จะแปลรหัสเป็นกรดอะมิโนฮิสติดีน 6 residues ทางด้านปลาย C-terminal ของยีนเซลล์ูลเลสที่สังเคราะห์ขึ้น เพื่อช่วยในการแยกบริสุทธิ์ของโปรตีน ทำการเพิ่มปริมาณ cDNA ของยีนเซลล์ูลเลสโดยใช้ปฏิกิริยา PCR นำ PCR product ที่ได้มาเชื่อมต่อกับเวกเตอร์

pET28b และนำเข้าสู่เซลล์ *E. coli* คัดเลือกรีคอมบิแนนท์โคลนโดยวิธี colony PCR สกัดพลาสมิดดีเอ็นเอของรีคอมบิแนนท์โคลน และนำไปหาลำดับ นิวคลีโอไทด์ นำรีคอมบิแนนท์พลาสมิดที่ได้เข้าสู่เซลล์ *E. coli* Rosetta(DE3)pLysS

4.2.8.2 การเหนี่ยวนำให้เกิดการแสดงออกของยีน

ทำการเลี้ยงเชื้อ *E. coli* ในอาหารเลี้ยงเชื้อ LB ที่อุณหภูมิ 37 °C และเหนี่ยวนำให้เกิดการแสดงออกของยีนด้วย IPTG ความเข้มข้น 1 mM บ่มเชื้อที่อุณหภูมิ 37 °C เป็นเวลา 4 ชั่วโมง ปั่นตกตะกอนเซลล์ และตรวจสอบการแสดงออกของยีนด้วย SDS-PAGE จากนั้นทำการแยกรีคอมบิแนนท์โปรตีนให้บริสุทธิ์ และวิเคราะห์โปรตีนใน SDS-PAGE ตรวจสอบแถบโปรตีนโดยการย้อมด้วย Coomassie Brilliant Blue

4.2.8.3 การแยกโปรตีนให้บริสุทธิ์

ละลายตะกอนเซลล์ด้วย binding buffer (20 mM sodium phosphate buffer, pH 7.4, 500 mM NaCl, 10 mM imidazol) และทำให้เซลล์แตกโดยใช้ sonicator แยกโปรตีนให้บริสุทธิ์โดยใช้ Ni-NTA column chromatography ทำการชะโปรตีนออกจากคอลัมน์ด้วย elution buffer (20 mM sodium phosphate buffer, pH 7.4, 500 mM NaCl, 500 mM imidazol) และวิเคราะห์โปรตีนใน SDS-PAGE นำโปรตีนที่แยกบริสุทธิ์ไปตรวจสอบแอกติวิตี เก็บโปรตีนที่ได้ไว้ที่ 4 °C

4.2.8.4 วิเคราะห์รีคอมบิแนนท์โปรตีนที่มีการแสดงออกด้วย Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

ผสมโปรตีนเข้ากับ sample buffer (50 mM Tris-HCl, pH 6.8, 10% glycerol, 1% SDS, 0.01% bromophenol blue, 50 mM DTT) ทำให้โปรตีนเสียสภาพที่อุณหภูมิ 100 °C เป็นเวลา 5 นาที วิเคราะห์ผลของโปรตีนใน 10% SDS-PAGE ตรวจสอบแถบโปรตีนด้วยการย้อมด้วย Coomassie Brilliant Blue

4.2.8.5 วิเคราะห์สมบัติของรีคอมบิแนนท์โปรตีน

ศึกษาสมบัติการเร่งปฏิกิริยาของรีคอมบิแนนท์โปรตีนบริสุทธิ์โดยอาศัยความจำเพาะต่อสับสเตรต (Kluepfel, 1988) โดยตรวจสอบแอกติวิตีของ endo- β -1,4-glucanase โดยใช้ agarose plate ที่มี 0.2 g/l carboxymethyl cellulose sodium (CMC-Na), 100 mM NaCl และ 100 mM acetate buffer, pH 5.2. บ่มที่อุณหภูมิ 30 °C เป็นเวลา 48 ชั่วโมง แล้วย้อมด้วย 1 g/l Congo red แล้วจึง destain โดยใช้ 1 M NaCl และตรวจสอบแอกติวิตีของ endo- β -1,4-xylanase โดยใช้ agarose plate ที่มี 1 g/l xylan จาก birchwood, 20 mM NaCl, 100 mM phosphate buffer, pH 5.8. บ่มที่อุณหภูมิ 30 °C เป็นเวลา 48 ชั่วโมง แล้วย้อมด้วย 1 g/l Congo red แล้วจึง destain โดยใช้ 1 M NaCl

4.3 ผลการทดลองและวิจารณ์

4.3.1 การสกัด Total RNA

เก็บตัวอย่างหอยเชอรี่ตัวเต็มวัย และตัดส่วนของกระเพาะอาหาร แช่ลงในถังไนโตรเจนเหลว และเก็บไว้ที่ -80°C ทำการสกัด total RNA จากส่วนกระเพาะของหอยเชอรี่โดยใช้ Trizol Reagent และกำจัด DNA ที่อาจปนเปื้อนมาโดยการย่อย DNA ด้วย RNase-free DNase I และสกัดแยก mRNA โดยใช้ Quick Prep mRNA purification kit วิเคราะห์ปริมาณและคุณภาพของ mRNA ด้วยเทคนิค spectrophotometry โดยเมื่อนำไปวัดค่าการดูดกลืนแสงที่ OD₂₆₀ พบว่ามีปริมาณของ mRNA ที่ได้ทั้งหมดประมาณ 3-5 μg เมื่อนำไปวัดค่าการดูดกลืนแสงที่ OD₂₆₀ / OD₂₈₀ ได้ค่าอยู่ในช่วงประมาณ 1.9 - 2.0 แสดงว่า RNA ที่ได้ค่อนข้างบริสุทธิ์

4.3.2 การค้นหายีนเซลล์ในหอยเชอรี่ (*P. canaliculata*) ด้วยเทคนิค RT-PCR

4.3.2.1 การสังเคราะห์ first strand cDNA

นำ mRNA ที่แยกได้จากกระเพาะอาหารของหอยเชอรี่มาสังเคราะห์ first-strand cDNA โดยใช้ ImProm-IITM Reverse Transcription System (Promega Corporation Medison, Wisconsin, USA) ซึ่งประกอบด้วย 2 μg mRNA, oligo(dT)-Not I primer และเอนไซม์ reverse transcriptase บ่มที่อุณหภูมิ 42°C เป็นเวลา 1 ชั่วโมง วิเคราะห์ปริมาณและคุณภาพของ first-strand cDNA ด้วยเทคนิค spectrophotometry

4.3.2.2 RT-PCR analysis

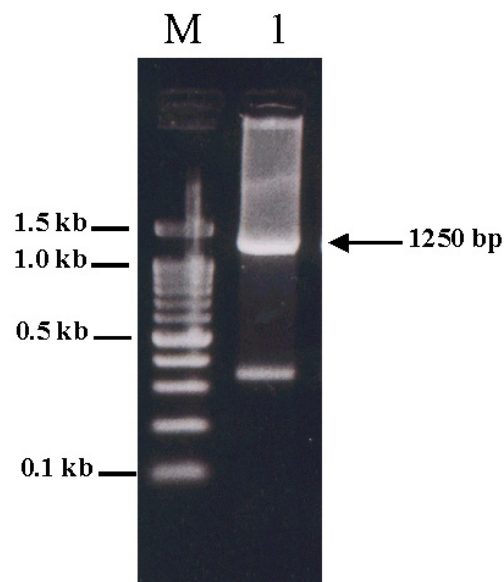
ในการค้นหา cDNA ของยีนเซลล์ด้วยเทคนิค RT-PCR โดยนำลำดับนิวคลีโอไทด์และลำดับกรดอะมิโนของยีนเซลล์จากหอย *A. crossean* (accession no. AY285999) และยีนเซลล์จากสิ่งมีชีวิตที่ใกล้เคียงกันที่มีรายงานไว้ใน GenBank มาทำ alignment โดยใช้โปรแกรม Clustal W เพื่อหาบริเวณอนุรักษ์ (conserved region) ที่สามารถนำมาใช้ในการออกแบบไพรเมอร์ ซึ่งพบว่าสามารถเลือกออกแบบ degenerate primers และ specific primer จากบริเวณ 5' UTR และบริเวณอนุรักษ์ภายในยีนเซลล์เพื่อใช้ทำ RT-PCR โดยใช้ cDNA จากกระเพาะอาหารของหอยเชอรี่เป็น template จากการทดลองจับคู่ของไพรเมอร์ต่าง ๆ และหาสภาวะที่เหมาะสมสำหรับการทำ PCR พบว่า เมื่อนำ RT-PCR product ที่ได้มาวิเคราะห์ด้วย agarose gel electrophoresis พบแถบดีเอ็นเอขนาดประมาณ 500 bp ซึ่งเมื่อแยกแถบดีเอ็นเอขนาด 500 คู่เบสที่สังเคราะห์ได้จาก RT-PCR ออกจาก agarose gel และนำมาเชื่อมต่อกับ pGEM[®]-T easy แล้วนำเข้าสู่เซลล์ *E. coli* DH5 α โดยวิธี electroporation ทำการคัดเลือก recombinant clones มาหาขนาดของ insert ของ cDNA โดยใช้ colony PCR (pUC1 primer: 5'-CCGGCTCGTATGTTG TGTGGA-3' และ pUC2 primer: 5'-GTGGTGCAAGGCGATTAAGTTGG-3') และวิเคราะห์ใน 1.0% agarose gel electrophoresis พบว่ามีแถบ DNA insert ขนาดประมาณ 530 bp จากนั้นนำ recombinant clone มาสกัด plasmid เพื่อนำไปหาลำดับนิวคลีโอไทด์ ผลจากการศึกษาเบื้องต้นโดยการทำ RT-PCR ผู้วิจัยพบว่ายีนเซลล์ในหอยเชอรี่ที่ได้มีความคล้ายคลึงกับยีนเซลล์ในหอย *A. crossean* อย่างไรก็ดีตามจากการวิเคราะห์ยีนที่แยกได้พบว่ามีลำดับนิวคลีโอไทด์เพียงบางส่วน

(partial cDNA) ยังไม่ครบสมบูรณ์ จึงทำการออกแบบไพรเมอร์จากข้อมูลลำดับนิวคลีโอไทด์ที่ได้ เพื่อนำไปใช้หาลำดับนิวคลีโอไทด์ full length ของยีนเซลล์ลูเลส

4.3.3 การค้นหา full length cDNA ของยีนเซลล์ลูเลส โดยใช้เทคนิค RACE-PCR

ได้สร้าง 3' RACE จาก mRNA ที่สกัดจากเนื้อเยื่อกระเพาะอาหารของหอยเชอรี่ เพื่อหา full-length cDNA ของยีนเซลล์ลูเลส โดยใช้ไพรเมอร์ที่ออกแบบมาจากบริเวณที่ทราบลำดับนิวคลีโอไทด์ จากการวิเคราะห์ PCR product พบแถบดีเอ็นเอขนาดประมาณ 1000 bp จึงได้ตัดแถบดีเอ็นเอมาโคลน และหาลำดับนิวคลีโอไทด์ ผลจากการวิเคราะห์ลำดับนิวคลีโอไทด์โดยนำไปเปรียบเทียบกับยีนที่มีรายงานไว้ใน GenBank พบว่ามีความเหมือนกับยีนเซลล์ลูเลสในหอย *A. crossean*

จากนั้นจึงได้ออกแบบไพรเมอร์จากลำดับนิวคลีโอไทด์บริเวณ 5' UTR (ไพรเมอร์ 5UTR-EGX) คู่กับไพรเมอร์ 3RACE-EGX ซึ่งจับได้บริเวณ 3'-UTR ของยีนเซลล์ลูเลส และนำมาทำ PCR โดยใช้ first-strand cDNA ที่เตรียมจากกระเพาะอาหารของหอยเชอรี่เป็น template เพื่อนำมาใช้ค้นหา full-length cDNA ของยีนเซลล์ลูเลสซึ่งจากการทำ PCR พบแถบดีเอ็นเอขนาดประมาณ 1250 bp (รูปที่ 1) จึงได้แยกชิ้น DNA ของ full length ของยีนเซลล์ลูเลสออกจากอะกาโรสเจล แล้วนำมาเชื่อมต่อกับพลาสมิด pGEM[®]-T easy แล้วนำเข้าสู่เซลล์ *E. coli* จากการคัดเลือกกริคอมมิแนนท์โคลนโดยวิเคราะห์ขนาดของ DNA insert โดยวิธี colony PCR แล้วสกัดกริคอมมิแนนท์พลาสมิดเพื่อนำไปวิเคราะห์หาลำดับนิวคลีโอไทด์



รูปที่ 1 Agarose gel electrophoresis ของ full-length cDNA ของยีนเซลล์ลูเลสจากหอยเชอรี่
Lane M คือ 100 bp DNA ladder ลูกศรชี้ แสดงขนาดของ RT-PCR ขนาด 1250 bp

4.3.4 การวิเคราะห์ลำดับนิวคลีโอไทด์ของ full length cDNA ของยีนเซลล์ลูเลส

จากข้อมูลของลำดับนิวคลีโอไทด์ของ cDNA ของยีนเซลล์ลูเลสจากหอยเชอรี่ (*P. canaliculata*) ที่ได้นำมาวิเคราะห์โดยใช้โปรแกรม GENETYX software (Software Development Inc.) และเปรียบเทียบกับข้อมูลนิวคลีโอไทด์ที่มีรายงานในธนาคารยีน (GenBank) โดยใช้โปรแกรม BLASTX และ BLASTN (<http://www.ncbi.nlm.nih.gov>) พบว่ามี ส่วนที่ homology กับยีนเซลล์ลูเลสของหอย *A. crossean* จากการวิเคราะห์ลำดับนิวคลีโอไทด์ ของยีนเซลล์ลูเลสของหอยเชอรี่ *P. canaliculata* พบว่าสามารถแยกยีนที่สมบูรณ์ (full-length cDNA) ได้ 2 isoform คือ GHF10-*Pc1* และ GHF10-*Pc3*

จากการวิเคราะห์ลำดับนิวคลีโอไทด์ของ GHF10-*Pc1* (รูปที่ 2A) ที่ได้พบว่ามีขนาด 1300 bp ประกอบด้วย 5' และ 3' UTR มีความยาว 52 และ 60 bp ตามลำดับ ส่วนทางปลาย 3' พบ putative poly (A) signal consensus sequence (AATAAA) อยู่ด้านหน้าจาก poly A tail 18 bp ลำดับนิวคลีโอไทด์ในส่วนที่เป็น open reading frame (ORF) มีขนาด 1188 bp ซึ่งสามารถถอดรหัสเป็นกรดอะมิโนได้ 395 residues ซึ่งเมื่อคำนวณตาม molecular mass จะให้โปรตีนขนาดประมาณ 44438 Da และมีค่า pI 6.49 จากการวิเคราะห์ GHF10-*Pc3* (รูปที่ 2B) พบว่ามีขนาด 1277 bp ประกอบด้วยส่วนที่เป็น open reading frame (ORF) มีขนาด 1191 bp ซึ่งสามารถถอดรหัสเป็นกรดอะมิโนได้ 396 residues ซึ่งเมื่อคำนวณตาม molecular mass จะให้โปรตีนขนาดประมาณ 44820 Da และมีค่า pI 6.75 จากการวิเคราะห์โดยใช้โปรแกรม SignalP (<http://www.cbs.dtu.dk/services/SignalP>) ไม่พบกรดอะมิโนในส่วนที่เป็น signal peptide จากการหา N-glycosylation sites (NXT และ NXS) พบว่ายีนเซลล์ลูเลสที่ได้มี putative N-glycosylation sites 1 ตำแหน่ง ที่กรดอะมิโนตำแหน่ง 393 (NIS) ซึ่งแสดงว่ายีนเซลล์ลูเลสเป็น glycoprotein และเมื่อนำลำดับนิวคลีโอไทด์ของยีนเซลล์ลูเลส GHF10-*Pc1* และ GHF10-*Pc3* ของ *P. canaliculata* ไปเปรียบเทียบกับความเหมือนกับยีนที่มีรายงานไว้พบว่ามีเหมือน กับยีนเซลล์ลูเลสใน *A. crossean* 98% และ 82% ตามลำดับ

```

A AGGACCAGAGGAGGGGCATCAGTTTTGTGTGTCGACGACGCTTCAGTCAAGCGCATGCCCGC 60
                                     M P A
TGGTGCTGCTGGTGCTGCGGTGACCAGCGACATCGACAGACTGAGAAGAAGCGACATAAC 120
G A A G A G V T S D I D R L R R S D I T
GGTTCACGTGAATGTTGGTGGTAACATCAACCACGGTCAAGTGAGCATTTCGTGTGTTACA 180
V H V N V G G N I N H G Q V S I R V L Q
AAAGAAAAAGGCATTCCCGTTCCGGACATGTGTGGCCGCTGGGCTTACAACGATGGGTC 240
K K K A F P F G T C V A A W A Y N D G S
CAAAGGAGCATACCGGATTTTCATCCACCAGCACTACAACCTGGGCGGTGCCAGAAAATC 300
K G A Y R D F I H Q H Y N W A V P E N S
ACTCAAGTGGGCTAGCATCGAACCTAACAGGGGACAAAAGAATTATCAGCCTGGCCTAAA 360
L K W A S I E P N R G Q K N Y Q P G L N
CATGCTTCACGGACTGAGAAATCAACGGGATTAAGGTGAGAGGTCAACACCTGGTGTGGTC 420
M L H G L R N H G I K V R G H N L V W S
TGTCGACAATACGGTGCAGAACTGGGTCAAAGCTCTGCATGGGGATGAGCTGCCAAAGGT 480
V D N T V Q N W V K A L H G D E L R K V
TGTCATGACCATCGTGGAAACCATCAACACATTTAAGGGCCTAGTGAGGACTGGGA 540
V H D H I V E T I N T P K G L V E H W D
TGTTGAACAACGAGAACCTGCATGGCCAGTGGTACCAGCATCAACTGAATGACAATGGCTA 600
V N N E N L H G Q W Y Q H Q L N D N G Y
CAACCTGGAACTGTTCCGTATCGCACACGCGCGCGACCCCAACGTCAAACCTCTTCCTCAA 660
N L E L F R I A H A A D P N V K L F L N
CGACTACAACGTTGTGTCCAAAGTTATTCAACAAACGACTATCTTCGACAAGGTCAACA 720
D Y N V V S N S T N D Y L R Q G Q Q
GTTTAAGGCCGCTAATGTGGGTCTTTACGGTTTGGGTGCTCAATGCCACTTTGGGCGACGA 780
F K A A N V G L Y G L G A Q C H F G D E
AGCCGACCCAGAACCCGGTACTAAGCAACGTCTGGATACTTTAGCTCAAGTGGGCGTGCC 840
A D P E P G T K Q R L D T L A Q V G V P
CATCTGGGCCACTGAGTTGGATGTGGTAGCTTCGGATGAGAACAGACGAGCGGATTTCTA 900
I W A T E L D V V A S D E N R R A D F Y
CGAGCACGCGCTGACAGTCCTGTACGGCCATCATGCCGTGGAGGGCATCCTGATGTGGGG 960
E H A L T V L Y G H H A V E G I L M W G
CTTCTGGGACAAGGCCCACTGGCGTGGTGCCAGAGCTGCTCTTGTGTGCGGAGACAACCT 1020
F W D K A H W R G A R A A L V V G D N L
GCAGCTGACGCGCGCGGACGTGCGGTGCTGGAGCTCTTTGAGCACAGGTGGATGACAGA 1080
Q L T A A G R R V L E L F E H R W M T D
CGAGACGACCAACCTGGCAGCGGGCACCCAGTTTACAGTACGCGGTTTCATGGCGACTA 1140
E T H N L A A G T Q F T V R G F H G D Y
CGAGGTGCAAGTCATCGTCCAGGGTCAAGAGCACACCAACCTGAGGCAGACGTTCTCGTT 1200
E V Q V I V Q G Q E H T N L R Q T F S L
GGGCAACGGTCCCCACACCGTCAACATTAACATTAGCTAGAGCGACACTCAGAGGGCAAT 1260
G N G P H T V N I N I S *
AAA GTTGAATGATGACACACAAAAA

```

รูปที่ 2A ลำดับนิวคลีโอไทด์ (full-length cDNA) และลำดับกรดอะมิโนของโปรตีนของยีนของยีนเซลล์ GHF10-Pc1 ของหอยเชอรี่ *P. canaliculata* ตัวอักษรสีแดงเข้มแสดงตำแหน่ง polyadenylation signal (AATAAA)


```

B TTGTGACGACGCTTCAGTCAAGCGCATGCCCCGCGGTTGCTGCTGCTGCTGATGTGAGCA      60
      M P A G A A A A D V S
GCGAGATCGACAGACTGAGAAGAAGCAACATAGTGGTTACCGTGACAGCTGGTGGTAACA      120
S E I D R L R R S N I V V H V T A G G N
TCAGCCACGGTGAAGTAACATACGGGTGGTACAAAAGAAGAAGTCGTTCCCGTTCCGAA      180
I S H G E V N I R V V Q K K K S F P F G
CCGCTGTGGCGCGCTGGGCTTACAACAATGATTCCAAGACAAAATACCGGGATTTCATCC      240
T A V A A W A Y N N D S K T K Y R D F I
ACCAGCACTACAACCTGGCGGTGCCAGAAAACGAACTCAAGTGGCGCACCATCGAACCTA      300
H Q H Y N W A V P E N E L K W R T I E P
CCAGCGGACATAAGAACTATCAGCCTGCCCTGACCATGATTCACGGACTGAAAAGTCACG      360
T R G H K N Y Q P A L T M I H G L K S H
GGATTAAGGTGAGAGGTCAACAACCTGGTGTGGTCTGTCAATTCAACGGTGCAGAGCTGGG      420
G I K V R G H N L V W S V N S T V Q S W
TCAAAGCTCTGCATGGGGATGAGCTGOGAAAGTTGTCCATGACCACATTGTGGAAACCG      480
V K A L H G D E L R K V V H D H I V E T
TCAACACATTTAAGGGCTTAGTGGAGCACTGGGACGTGAACAAACGAGAACCTGCATGGCC      540
V N T F K G L V E H W D V N N E N L H G
AGTGGTACCAGCAACTGAATGACCGAACTACAACATAGAAGTTCCTCGTATCGCAC      600
Q W Y Q Q Q L N D P N Y N I E L F R I A
ACGCTGCGGACCCCAACGTCAAACCTCTCCTCAACGACTACAACGTGGTGGCTTACGGTG      660
H A A D P N V K L F L N D Y N V V A Y G
CCGCAACCAATGCCTATCTTCAGCAAGGTCAACAGTTTAAGGCGGCTAATGTGAGTCTTT      720
A A T N A Y L Q Q G Q Q F K A A N V S L
ACGGTTTGGGTGCCAGTGCCACTTTGGGATGAAGCTAACCCAAACGTGCTGGGATGA      780
Y G L G A Q C H F G D E A N P N V A G M
AGCAACATCTGGATATTTAGCTCAAGTGGGTTGCCATCTGGGCACTGAGTTGGATG      840
K Q H L D I L A Q V G L P I W A T E L D
TGTTAGCTACGAGCGAGAACAAACGAGCGGACTTCTACGAGCACGCGCTGACAGCCCTGT      900
V L A T D E N K R A D F Y E H A L T A L
ACAGCCATCATGCCGTGGAGGGCATCCTGATGTGGGGCTTCTGGGACAAGGCCCACTGGC      960
Y S H H A V E G I L M W G F W D K A H W
GGCATGAACGAGCTGCTCTTCTTGTGGAGACAACCTGCAGCTGACGGCGGCGCGGAGCTC      1020
R H E R A A L L V G D N L Q L T A A G R
GCGTGCTGGAGCTCTATGAGCACAGGTGGATGACAGACGAGACGACAACTGGCAGCGG      1080
R V L E L Y E H R W M T D E T H N L A A
GCACCCAGTTCACAGTACGCGGTTTCCATGCGCGACTACGAGGTGCACGTGATCTACCAAG      1140
G T Q F T V R G F H G D Y E V H V I Y Q
GTCAGGAGCGCACCAACCTGAAGCAGACGTTACGTTGGGCAACGCAGCCACACCGTCA      1200
G Q E R T N L K Q T F T L G N A A H T V
ACATCAACATTAGCTAGAGCGGACACTCAGAGGGCAATAAAGTTGAATGATGACACAACAA      1260
N I N I S *
AAAAAAAAAAAAAAAAAAAAA      1277

```

รูปที่ 2B ลำดับนิวคลีโอไทด์ (full-length cDNA) และลำดับกรดอะมิโนของยีนของยีนเซลล์ GHF10-Pc3 ของหอยเชอรี่ *P. canaliculata* ตัวอักษรสีแดงเข้มแสดงตำแหน่ง polyadenylation signal (AATAAA)

4.3.5 การวิเคราะห์ function domains และ conserved motifs ของยีนเซลลูเลส

เนื่องจากเซลลูเลสสามารถแบ่งออกได้เป็นหลายกลุ่ม โดยอาศัยความแตกต่างของโครงสร้าง 3 มิติของลำดับกรดอะมิโนในบริเวณ catalytic domain (Beguin and Aubert, 1994) ซึ่งในปัจจุบันสามารถแบ่งได้เป็น 106 แฟมมิลี (CAZY database; <http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html>) และจากการนำลำดับกรดอะมิโนที่ได้โดยการถอดรหัสจาก cDNA ของยีนเซลลูเลสที่ได้มาวิเคราะห์โดยโปรแกรม SMART analysis (<http://smart.embl-heidelberg.de/>) พบว่า primary structure ของโปรตีนเซลลูเลสตั้งแต่กรดอะมิโนตำแหน่งที่ 84 ถึง 336 มีความคล้ายกับเซลลูเลสในกลุ่ม glycosyl hydrolase family 10 (GHF10) ดังนั้นจึงจัดเซลลูเลสที่ได้ไว้ในกลุ่ม glycosyl hydrolase family 10 ซึ่ง glycosyl hydrolase ในกลุ่มนี้เดิมมีชื่อเรียกว่า cellulase family F ประกอบด้วยเอนไซม์ที่มีแอกติวิตีต่าง ๆ เช่น xylanase (EC: 3.2.1.8); endo-1,3-beta-xylanase (EC: 3.2.1.32) และ cellobiohydrolase (EC: 3.2.1.91) จากการวิเคราะห์ลำดับกรดอะมิโนพบว่ามีการดะมิโนที่คาดว่าจะ conserve catalytic nucleophile และ proton donor ซึ่งมีความสำคัญต่อการเร่งปฏิกิริยา คือ Glu268 ซึ่งอยู่ในบริเวณอนุรักษ WDVNNE และ Glu167 ที่อยู่ในบริเวณอนุรักษ TELD ตามลำดับ จากการวิเคราะห์ neighbor-joining tree พบว่ายีนเซลลูเลสที่ได้มีความสัมพันธ์ใกล้ชิด (close relationships) กับเซลลูเลสจาก *A. crossean*

Clostridium AKWYNSSESSQGNVSYANADLYLNWCAENGKIKVRGHCIFWEPEEWQPSWLKGLTGDALMK
 Bacillus AKWYANPEPERGKIITYEKADAMLNFAADRHLQLPVRGHALFWEVEDANPSQLRSLPNHEVYE
 Hordeum LKWYHTPEVQOQLNADADALLAFCDRLGKTVRGHCVFWSVGDGVQOQVKNLNKDQLRSA
 Triticum LKWYHTPEVQOQLNADADALLAFCDRLGKTVRGHCVFWSVGDGVQOQVKNLNKDQLRSA
 Arabidopsis LKWYWTPEEQGKLNLYQDADDMLNLGSSNNIETRGHCIFWEVQATVQOQVKNLNKDQLRNA
 Nicotiana LKWYWTPEAQGNFNKYKDADELDFCTKNNIQVRGHCIFWEVVGTVQAAVQVSLNKNLDMTA
 Oryza LKWYWTPEAQGKLNLYADADALLAFCDRLGKTVRGHCIFWEVDSAVQOQVKNLNKDQLRSA
 GHF10-Pc1 LKWASIEPNRGCKNYQPGNLMLHGLRNHGIKVRGHNLVWSVDNTVQNVKALHGDRLKRV
 Ampullaria LKWASIEPNRGCKNYQPGNLMLHGLRNHGIKVRGHNLVWSVDNTVQNVKALHGDRLKRV
 GHF10-Pc3 LKWRTIEPTRGKKNYQPALTMHGLKSHGIKVRGHNLVWSVNTVQNVKALHGDRLKRV
 ** * : * * : *** : * : * : *

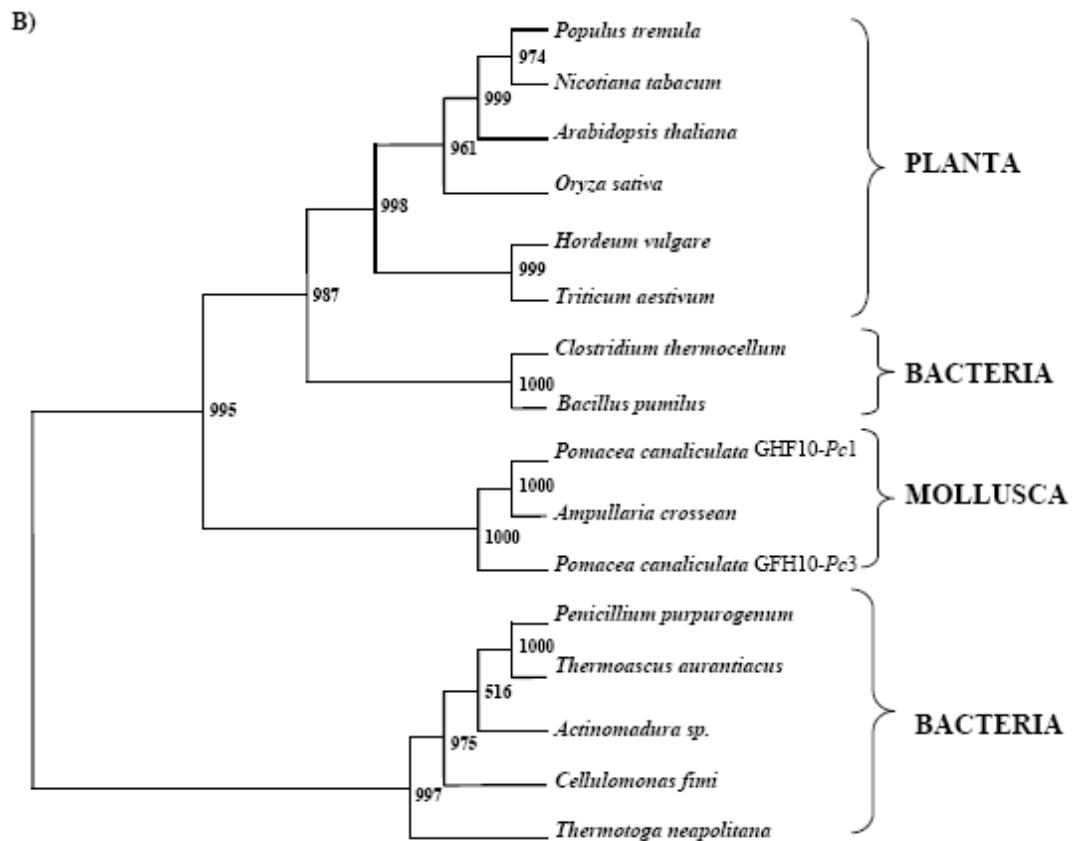
Clostridium IDARLESVVPFRGKFLHWDVNNEMLHGGFFKSRLG-ESIWPYMFKRARELDPDAKFLVN
 Bacillus MKKRLEHAGNHFKGRFRHWDVNNEMMHGSSFFKDRFG-KNIWKWMYEETKKIDPQAILFVN
 Hordeum MQSRLEGLVSRYAGRFKHYDVNNEMLHGFRFRDLRGDEDVPAYMFKEVARLDPEPVLFVN
 Triticum MQSRLEGLVSRYAGRFKHYDVNNEMLHGFRFRDLRGDEDIPAYMFKEVARLDPEPALFVN
 Arabidopsis VQNRLLDILLNRYKKGFKHYDVNNEMLHGFSFYQDKLG-KDIRVNMFKTAHQIDPSATLFVN
 Nicotiana VQNRLLTGLLTRYKKGFEHYDVNNEMMHGFSFYQDRLG-KEIRVNMFKTAHQIDPSILFVN
 Oryza VASRINGLLTRYKKGFRHYDVNNEMLHGFSFYQDKLG-AGARAAMFRAASELDPDAILFVN
 GHF10-Pc1 VHDHIVETINTFFKGLVEHWDVNNENLHGQWYQHQLNDNGYNLELFRIAHAADPNVKLFVN
 Ampullaria VHDHIVETINTFFKGLVEHWDVNNENLHGQWYQHQLNDNGYNLELFRIAHAADPNVKLFVN
 GHF10-Pc3 VHDHIVETVNTFFKGLVEHWDVNNENLHGQWYQQQLNDPNYNIELFRIAHAADPNVKLFVN
 : : : * * : ***** : * : : : : : : : * : * : *

Clostridium DYNIIITVY----EGDAYIRQTEWLLQNGAIEIDGIGVGGHFDDEDVPE--LVVKARLDNL
 Bacillus DYNVISYG----EHHAYKAHINELRQLGAPIEAIGVGGHFEERVDPE--LVVKERLDNL
 Hordeum DYNVECGN-DPNATPEKVAEQVAWLQSCGAVVRGIGLGGHVQNPVGE--VICAA-IDRLA
 Triticum DYNVERAN-DPNATPEKVAEQVAWLQRCGAVVGGIGLGGHVQNPVGE--VICAA-IDRLA
 Arabidopsis DYHIEDGC-DEKCSPEKYTEQILDLQEKGAAPVGGIGGGHIDSVPGE--IVCSA-LDKEG
 Nicotiana DYHVEDGS-DTRSSPEKYTEHILDLQEHGAPVGGIGGGHIDSVPGE--IVCSA-LDKEG
 Oryza DYNVEGACVDVRAPEAYIAQVTGLQEQAAGVGGVGLGGHVTAPEVGA--VVRAA-IDRLA
 GHF10-Pc1 DYNVVSNS----YSTNDYLROGQQFKAANVGLYGLGACGCHFGDEADP-EPGTKQRLDTLA
 Ampullaria DYNVVSNS----YSTNDYLROGQQFKAANVGLYGLGACGCHFGDEADP-EPGTKQRLDTLA
 GHF10-Pc3 DYNVVAYG----AATNAYLQOQGFKAANVSLYGLGACGCHFGDEANPNVAGMKQHLDTLA
 ** : : * * : : : : : : * : * : *

Clostridium TLGLPIWVTEYDSKTPDVNKRAENTENLYRIAFSSHFAVEGILMWGFWAGNHWRGQDAIIV
 Bacillus ELGLPIWVTEYDSVHPDPNRRADNLBALYRVAFSSHFAVGKGLMWGFWAGAHWRGEHAAIIV
 Hordeum KTGVPINATEDIMPEYDVGLRAKDLBVLVREAYAHFAVEGLVFGWFGMQGTMTWR-QNAWLIV
 Triticum KTGVPINATEDIMPEYNSVLRADLBVLVREAYAHFAVEGLVFGWFGLOGTMTWR-ENSWLIV
 Arabidopsis ILGLPIWVTELDVSSVNEHIRADDLEVMWWEAFGHPAVEGIMLWGFWEFLFMSR-DNSHLV
 Nicotiana ILGLPIWVTELDVSSGNEYIRADDLEVMLEAYAHFAVEGIMLWGFWEFLFMSR-PNAHLV
 Oryza VLGVPINATEDLDVSSANEHVRADLBAMLREAYAHFAVDGVVLWGFWEFLFMSR-DDAHLV
 GHF10-Pc1 QVGVPINATEDLDVVASDENRRADFYEHALTVLYGHHAVEGILMWGFWDKAHWRGARAALV
 Ampullaria QVGVPINATEDLDVVASDENRRADFYEHALTVLYGHHAVEGILMWGFWDKAHWRGARAALV
 GHF10-Pc3 QVGVPINATEDLVLATDENRRADFYEHALTALYSHHAVEGILMWGFWDKAHWRHERAALLV
 * : * : * : * : * : * : * : * : * : * : * : *

Clostridium DHD-WTVNEAGKRYQALL
 Bacillus NYD-WSLNEAG-----
 Hordeum DAD-GTVNEAGQMFNLNQ
 Triticum DAD-GTVNEAGQMFNLNQ
 Arabidopsis NAE-GDVNEAGKRFALVK
 Nicotiana NAE-GDINEAGKRYLALK
 Oryza DAE-GEVNEAGRLLQLK
 GHF10-Pc1 VGDNLQLTAAGRRVLELF
 Ampullaria VGDNLQLTAAGRRVLELF
 GHF10-Pc3 VGDNLQLTAAGRRVLELY
 : : * : *

รูปที่ 3A Multiple alignments ของลำดับกรดอะมิโนของยีนเซลล์ลูเลส GHF10-*Pc1* และ GHF10-*Pc3* จากหอยเชอรี่ *P. canaliculata* และเซลล์ลูเลสอื่นจาก *A. crossean* (AAP31839), *Oryza sativa* (BAD52808), *Arabidopsis thaliana* (BAB83869), *Populus tremula* x *Populus tremuloides* (AAX33301), *Nicotiana tabacum* (AAZ79232), *Triticum aestivum* (AF156977), *Hordeum vulgare* (AAB51668), *Bacillus pumilus* (AF466829), *Clostridium thermocellum* (EAM47104), *Thermotoga neapolitana* (Q60041), *Actinomadura* sp. (AAA17888), *Penicillium purpurogenum* (AAF71268), *Cellomonas fimi* (AAZ76373), and *Thermoascus aurantiacus* (AF127529) ตัวอักษรสีดำเข้มและแรเงาแสดง conserve catalytic nucleophile และ proton donor ที่ตำแหน่ง Glu268 ที่อยู่ในบริเวณนักรักษ์ (WDVNNE), และ Glu167 ที่อยู่ในบริเวณนักรักษ์ (TELD)



รูปที่ 3B Neighbor-joining tree ของยีนเซลล์ูเลส GHF10-Pc1 และ GHF10-Pc3 จากการทำ bootstrapping 1,000 ครั้ง *P. canaliculata*, *A. crossean* (AAP31839), *Oryza sativa* (BAD52808), *Arabidopsis thaliana* (BAB83869), *Populus tremula* x *Populus tremuloides* (AAX33301), *Nicotiana tabacum* (AAZ79232), *Triticum aestivum* (AF156977), *Hordeum vulgare* (AAB51668), *Bacillus pumilus* (AF466829), *Clostridium thermocellum* (EAM47104), *Thermotoga neapolitana* (Q60041), *Actinomadura sp.* (AAA17888), *Penicillium purpurogenum* (AAF71268), *Cellumonas fimi* (AAZ76373), and *Thermoascus aurantiacus* (AF127529)

4.3.6 การแยก และศึกษาการจัดเรียงตัวของยีนเซลล์เลส

ในการแยกยีนเซลล์เลสของหอยเชอร์รี่ด้วยเทคนิค PCR ได้ทำการออกแบบไพรเมอร์จำนวน 3 คู่ จากข้อมูลลำดับนิวคลีโอไทด์ของ cDNA ของยีนเซลล์เลส เพื่อใช้ในการเพิ่มปริมาณยีนเซลล์เลสโดยเทคนิค PCR จากการวิเคราะห์ PCR product พบแถบดีเอ็นเอขนาดประมาณ 2200, 1500 และ 1200 bp จึงได้ตัดแถบดีเอ็นเอมาโคลนและวิเคราะห์หาลำดับนิวคลีโอไทด์ จากการวิเคราะห์ลำดับนิวคลีโอไทด์ทั้ง 3 fragment พบว่าประกอบด้วยส่วนของยีนเซลล์เลสที่ overlap กัน จึงได้ออกแบบไพรเมอร์อีก 1 คู่ เพื่อใช้เพิ่มปริมาณ full length ของยีนเซลล์เลส จากการวิเคราะห์ PCR product ที่ได้ และทำการตัดแถบดีเอ็นเอขนาดประมาณ 5 kb มาโคลนและหาลำดับนิวคลีโอไทด์

จากการวิเคราะห์ลำดับนิวคลีโอไทด์พบว่ายีนเซลล์เลส GHF10-*Pc1* และ GHF10-*Pc3* ของหอยเชอร์รี่ *P. canaliculata* มีขนาด 4937 และ 4512 bp ตามลำดับ (รูปที่ 5A และ 5B) เมื่อนำลำดับนิวคลีโอไทด์ของยีนเซลล์เลสที่ได้มาเปรียบเทียบกับลำดับนิวคลีโอไทด์ของ cDNA ของยีนเซลล์เลส เพื่อวิเคราะห์หาการจัดเรียงตัวของยีนเซลล์เลส พบว่ายีนเซลล์เลสทั้งสองประกอบด้วย 9 exon และ 8 intron โดยมี start codon ATG อยู่ใน exon ที่ 1 และมี stop codon TAG อยู่ใน exon ที่ 9 เมื่อวิเคราะห์ขนาดของนิวคลีโอไทด์ในแต่ละ exon และ intron พบว่ายีนเซลล์เลสมีขนาดของ exon อยู่ระหว่าง 46 bp (exon 2) – 234 bp (exon 9) และ intron มีขนาดอยู่ระหว่าง 234 bp (intron 6) – 1071 bp (intron 8) และเมื่อวิเคราะห์บริเวณรอยต่อระหว่างแต่ละ exon และ intron พบว่าเป็นไปตาม GT/AG rule

A)

CGACGCTTCAGTCAAGCGC**ATG**CCCCGCTGGTGCTGCTGGTGCTGGGGTGACCAGCGACAT 60
M P A G A A G A G V T S D I
CGACAGACTGAGAAGAAGCGACATAACGGTTCAgtgcggtgctgctttcaatatataacaa 120
D R L R R S D I T V H
cctcacttagttctgcattgataaaaatatctcgtcagtagtgtaccaattctgtgtgtat 180
caagaaagtaacaccggttggtatataccatTTTTAAAGTAAAGCGAATTTAATCTGGAATAG 240
tacaatacttgtagtaatgcataatataaattcatgattttggtagtaataataagataaa 300
ttgtaacataaacggatagaatattaataacaatcttatgtcacgggtcatttcaagttt 360
gcacgatatacctggaatcaagaaccgaaatagtaattaaggccgaacggtgcggtgtaga 420
aaatcgtgtgcagcccgatggtggttaataactaaagaaacacaataacatccggcag 480
ccgtcgtcaagtgtgtaaaatctgacctgaaatccattcacgaagttgatgatttcttat 540
atcagCGTGAATGTTGGTGGTAACATCAACCACGGTCAAGTGAGCATTCGTgtaagtcat 600
V N V G G N I N H G Q V S I R
agatgagagcatcatccgcctgaggacagggggcaataaaaagagcaagctcatcaaagg 660
tgtcaacctcgagaatgtgccttaagtaaacttgattgagaggtacagttgaccacgtga 720
acagataggagagaagagagacaaaaaatgaaggatacaatagaagatgttaggaataata 780
ttaattttaataacttattttatatttttaattgttaattgtataaaagccaccagactaaat 840
ctttgttcttcaaggtaactgtcgttgattacagGTGTTACAAAAGAAAAAGGCATTCCC 900
V L Q K K K A F P
GTTCCGGACATGTGTGGCCGCCTGGGCCTACAACGATGGGTCCAAAGGAGCATACCGGGA 960
F G T C V A A W A Y N D G S K G A Y R D
TTTCATCCACCAGCACTACAACCTGGGCCGTGCCAGAAAACCTCACTCAAGTGGGCTAGCAT 1020
F I H Q H Y N W A V P E N S L K W A S I
CGAACCTAACAGGgtacattttttattaatgagatttataaatttaacataaattgatag 1080
E P N R
tatattaataaaaaatcttctttaacgttttacttccagtttgcacgatagtcctttaatca 1140
agaactgaaatatttttcaatactgaatgttgctgcgcataaaaatcgtttgcagcctgtc 1200
atgtgatgttactactaaagaaatacaataacatatattgaaatgatgtttaaggggtct 1260
tctgattttatatctgttccagggagagaatagtttaaaagggtactccattcacagcaagga 1320
aagacaaagttaccggttgatcatattcatagttttacttttctaaatcggtgtgtgtgtgtg 1380
tatatatattttatgtctgagcgcgcttgaaatgcttaagtcgacaacttaatgtaacagat 1440
tagttttattaagagtataactaaaaactgggttggttccctgctagccaaagaaatggggt 1500
ttgggggtattttgataatcgcacgtgtgtcttcccttggaataaaaataaacaagtaaaaa 1560
ttctggaggggaacatacttttgaggcactcgacattgagggttccggttatgatcgccatt 1620
taatgcttgctctcctccttggacatatagaaggggaaaatataatgactgttgattgtg 1680
tttttcagGGACAAAAGAATTATCAGCCTGGCCTAAACATGCTTCACGGACTGAGAAATC 1740
G Q K N Y Q P G L N M L H G L R N
ACGGgttagtaaacattaggataactgcaacccccatcacccgacattataatcatcatcgt 1800
H G
cgacttcattatcgtataataatgaaagaattggcgaattttaattattttaagaattatt 1860
aacataaaatttaattaacatttaataaatcaaagaaagtttagcgtttcagagattgctc 1920
ctcccccatgtattttactatttctatcacttcggttaaaagcaagtctcagactcagttc 1980
tgtctttttgataattccttaaaatagtcctacaaatgaaatagtaactgaaatgtctca 2040
cattgatacgaatttaagaaaaatattcacttacactaagagttaacacatgtcgtgtaca 2100
gaggtgtttgtagaaatgtttgttatagttttcttgcattccactcgagatgtttaaaaca 2160
gGATTAAGGTGAGAGGTCACAACCTGGTGTGGTCTGTGCGACAATACGGTGCAGAACTGGG 2220
I K V R G H N L V W S V D N T V Q N W
TGAAGGCTCTGCATGGGGATGAGCTGCGAAAGGTTGTCCATGACCACATCGTGGAACCA 2280
V K A L H G D E L R K V V H D H I V E T
TCAACACATTTAAGGGCCTgtaagttttatatcatttgttactttctttctaaaatgtcag 2340
I N T F K G L
tagttttattttttatttttttatcttggtattttggagaaacaaaatttactgttctacata 2400
caccaaaccagatgataccagttacaatactgaattttatgtttgtaaaaaatattgtttgaa 2460
ccttatagcaatagactgccactaattttaagaagtcacggttgcaagtgttagtttttaa 2520
ctcactcatcaacagatcaagcaagaggttttttctgttttgaccagAGTGAGCACTGG 2580
V E H W

รูปที่ 4A การจัดเรียงตัวของยีน (Genomic organization) ของยีนเซลล์ลูเลส GHF10-*Pc1* ตัวอักษรพิมพ์ใหญ่แสดงลำดับนิวคลีโอไทด์และลำดับกรดอะมิโนบริเวณ exon ตัวอักษรพิมพ์เล็กและแฉกแสดงบริเวณ intron ตัวอักษรเข้มแสดง start และ stop codon

B)

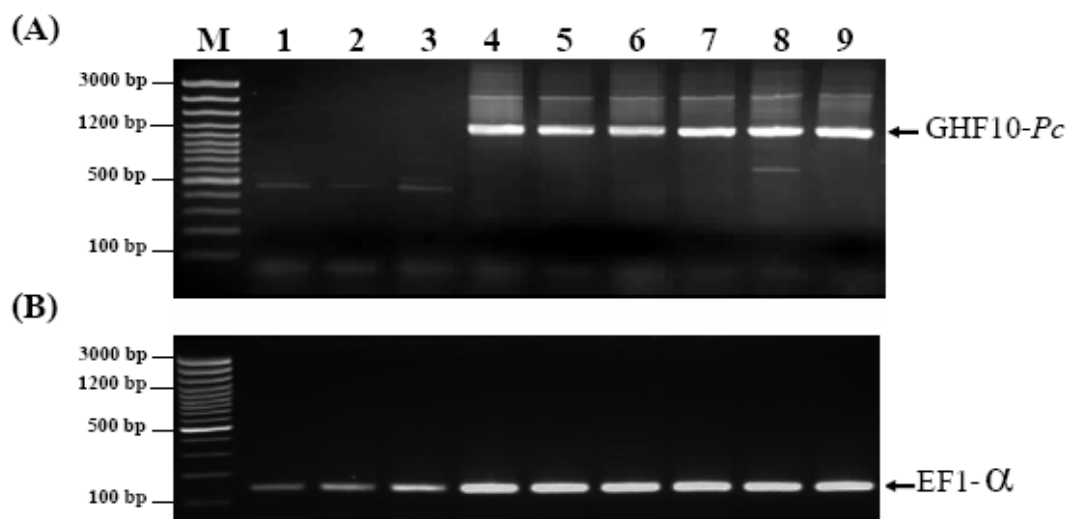
CGACGCTTCAGTCAAGCGCATGCCCCGCGGGTGCTGCTGCTGCTGATGTGAGCAGCGAGAT 60
M P A G A A A A D V S S E I
CGACAGACTGAGAAGAAGCAACATAGTGGTTCAgtgctgctgctttcaatatataacaa 120
D R L R R S N I V V H
cctcacttagttctgcattgataaaaatatctcgtcagtagtgtagtaccattctgtgtgtat 180
caataaagtaacactggtgttataaccattttttaattaaagcgaatttaattctgactatt 240
acaatattgtatgtaatcatatataattcattattttttattaataatattttataatttt 300
aacataaatggatagtagtataattaataacaatcttattttcacggcttattttcaagtttgac 360
gatgttcttttaatacaagaacttaaatatttttcaaatacagaatggttgctgctgataaaaag 420
cgtttttcagcctgtcatttggtggttactactagacaaacacaataacatcttcacaccttc 480
gttaaatgtgtataaatctgacctgaaatccatttcacgaccttgctgattttgttttcagC 540
GTGACAGCTGGTGGTAACATCAGCCACGGTGAAGTAAACATACGGgtaagtcatagttga 600
V T A G G N I S H G E V N I R
tatcatcatccgcctgaggacagagtacaacaaactaaagagctactttaccgcaagctc 660
cacctggtgtcacccccgatttgccctcggttatggcttttagtagttattttgaatattctt 720
tgtcactttgtcactagagacatctgtagcgacaagacccaaattcagcataactcttatg 780
ttgaaacataatcattttcaattctaatatcatgtgttaaactgtgttgtaggcacatttga 840
ccacgtgaacagtttagtagagaagagatacacaaaaataaaagatacagtagaagatttgg 900
ggtaataattaatttttaataacttatttttatatttttaataattaatgtataaaaaatcccc 960
agactaagtctttgttcttcaagctaactgtcggttgattacagGTGGTACAAAAGAAGAA 1020
V V Q K K K
GTCGTTCCCGTTTCGGAACCGCTGTGGCCGCTGGGCCTACAACAATGATTCCAAGACAAA 1080
S F P F G T A V A A W A Y N N D S K T K
ATACCGGGATTTCATCCACCAGCACTACAACCTGGGCGGTGCCAGAAAACGAACCTCAAGTG 1140
Y R D F I H Q H Y N W A V P E N E L K W
GCGCACCATCGAACCTACCAGGgtacactttttattaatatgattttataatttttaacata 1200
R T I E P T R
aattgatagcataatgataaacaatcttattttaacgttttacttccagtttgcatgatagt 1260
ctttaatcaagaacttaaatatttttcaatgctgaatggttgctgctgcataaaaatcggtgtg 1320
cagccgtctttgtatgttactactaaagaaatacatttaacatacattgaaatgatgttta 1380
aggggtcttgtgattttgcattttgttcaggtaatatagttttaaagggtacagattaatttt 1440
attaagactataactaaaaactgggtagtttctgctagccaaagaaatgggggttttggggg 1500
catttgataatcgacgtgtgtcttccctgtgaataaaaaataaacaagttaaaaattctgga 1560
gaaaaataacttttgatgcactcgacattgaggcttccggttatgatcgccatttaattgctt 1620
acctctcctcttgggtcacatcactccacaaatagaggaggaaaatataacagtttgattgt 1680
ggtttttcagGGACATAAGAACTATCAGCCTGCCCTGACCATGATTACGGACTGAAAAGT 1740
G H K N Y Q P A L T M I H G L K S
CACGGgttagtaaacattaagatttagcaaccaccatcaccgacattataatcatcctct 1800
H G
tcgacttcattattgtatgataatgaaagaattgggtgaatttttagttataaaaacattttat 1860
taaatgaaagagagatttagcggtttcagagattgtcctcccccatgtattttactatttt 1920
tcgtacttcgttaaaagcaagtcgcagactcagttctgactttttgataattccttaaaag 1980
tagtccatacaaatgaaatagtaactgcaaatgtctcacactgatacgaatgaagaaaat 2040
attcgcttacactaagagtcagcacatgtcgtgtacagaggtgtttgtagagatgtttgt 2100
tatagttttcttgcattcactcgagatgtttaaaatagGATTAAGGTGAGAGGTCACAAC 2160
I K V R G G H N
CTGGTGTGGTCTGTCAATTCAACGGTGCAGAGCTGGGTCAAAGCTCTGCATGGGGATGAG 2220
L V W S V N S T V Q S W V K A L H G D E
CTGCGAAAGGTTGTCCATGACCACATTGTGGAACCGTCAACACATTTAAGGGCTTgtaa 2280
L R K V V H D H I V E T V N T F K G L
gtttattttcatttattttcttttctaaaatgtcagtagttttatttgttttatgttttgattg 2340
ttgcgttatgtggagaaacaaaattttactattctacatacacccaaaccagatgataaccag 2400
ttacaatactgaatttttatgttgtaaaaatatgtttgaaccttttagcaatagactgcc 2460
ctaataagaagtcacgttgtgaagtgttagtttttaactcattcatcaacacatcaag 2520
taagagtgttttctgttttgaccagAGTGGAGCACTGGGACGTGAACAACGAGAACCTGC 2580
V E H W D V N N E N L

ATGGCCAGTGGTACCAGCAACAACCTGAATGACCCGAACCTACAACATAGAAGTGTTCGGTA	2640
H G Q W Y Q Q Q L N D P N Y N I E L F R	
TCGCACACGCTGCCGACCCCAACGTCAAACCTCTTCCTCAACGACTACAACGTGGTGGCTT	2700
I A H A A D P N V K L F L N D Y N V V A	
ACGGTGCCGCAACCAATgtcagtgatgagaaaacttgcttctatatcacctgagaacaa	2760
Y G A A T N	
atgtttccaagcatagtatgcttctatttctgcataaaaataactccttggtgtcttttagca	2820
ttatagtcttctgtgccttgctatgttcccgaaacaatgcaacttgtaactgtccatacta	2880
gtctgtagtctttgttttaaatctgtaaaaaacgaatatcatcttcacacagtctttatggt	2940
gtgtgacagGCCTATCTTCAGCAAGGTCAACAGTTTAAGGCCGCTAATGTGAGTCTTTAC	3000
A Y L Q Q G Q Q F K A A N V S L Y	
GGTTTGGGTGCCAGTGCCACTTTGGCGATGAAGCTAACCCAAACGTCGCTGGGATGAAG	3060
G L G A Q C H F G D E A N P N V A G M K	
gtaagaaaaataatatcataaacatccttcgcctttctcattttccttctcctattatta	3120
ttattctaaaaaatgaaacaaatgtacgcacacaatgaagtagttctacaatacagtaa	3180
gataaaagatttttagattttgtctgtatcattttttcttttttagttatctgcatttttta	3240
ccattttactttacatggaataacatatctttatgttgagaaatcagaaagtcttcattt	3300
tatagctgtggatttaaaaaatattcttctcacttcttctataactccataagctgtttta	3360
gtgacatctgggttttttctaattctcagCAACATCTGGATATTTTAGCTCAAGTGGGGTT	3420
Q H L D I L A Q V G L	
GCCCATCTGGGCCACTGAGTTGGATGTGTTAGCTACGGACGAGAACAACGAGCGGACTT	3480
P I W A T E L D V L A T D E N K R A D F	
CTACGAGCACGCGCTGACAGCCCTGTACAGCCATCATGCCGTGGAGGGCATCCTGATGTG	3540
Y E H A L T A L Y S H H A V E G I L M W	
GGGCTTCTGGGACAAGGCCCACTGGCGGCATGAACGAGCTGCTCTTCTTGTCGGAGACAA	3600
G F W D K A H W R H E R A A L L V G D N	
CCTGCAGgtgtgttgggttacaattaaccaaatagggatcttacaggcattatagcacag	3660
L Q	
ccagtatccagacagacggttttgtaatcactgtcatgatgagaacacattggattcttct	3720
tctattagaatttctaccagaactatccttgacttcaacattttttttctataagaga	3780
aatccgacaaatgcttactgagtgtacctatgggttagtccaaattatgcgggcatgcttt	3840
tttccattctagaattacttttagcacaaatacgaccaaataatgattaccatctaagtctca	3900
acattagtttttacattttttcatcaacgcttaaacacttttaaagacttttaaattgtaag	3960
aatcaagctttttcacagcgataaggattcttttagatgtaaccagctggtaaagtaaaac	4020
agtttaaattactatttggttagaggcaaaggagttatgatgttgtaaaggacattgtctta	4080
cataaggaatgcttcattcaaagtacttctgatagtgaatgcgaaacacatttcattaat	4140
aataaaaaaataagatatgagaattgttactaaaatatgtctgttttagtctccatgatta	4200
ataaaatgtttgcttcgttaaagttcatgaaacaccgggtgtttgtggttagacataaaat	4260
aaacgtttgattgtatagCTGACGGCGGCCGACGTCGCGTGCTGGAGCTCTATGAGCAC	4320
L T A A G R R V L E L Y E H	
AGGTGGATGACAGACGAGACGCACAACCTGGCAGCGGGCACCCAGTTCACAGTACGCGGT	4380
R W M T D E T H N L A A G T Q F T V R G	
TTCCATGGCGACTACGAGGTGCACGTGATCTACCAAGGTGAGGAGCGCACCAACCTGAAG	4440
F H G D Y E V H V I Y Q G Q E R T N L K	
CAGACGTTACGTTGGGCAACGCAGCCCACACCGTCAACATCAACATTAGCTTAGAGCGAC	4500
Q T F T L G N A A H T V N I N I S *	
ACTCAGAGGGCA	4512

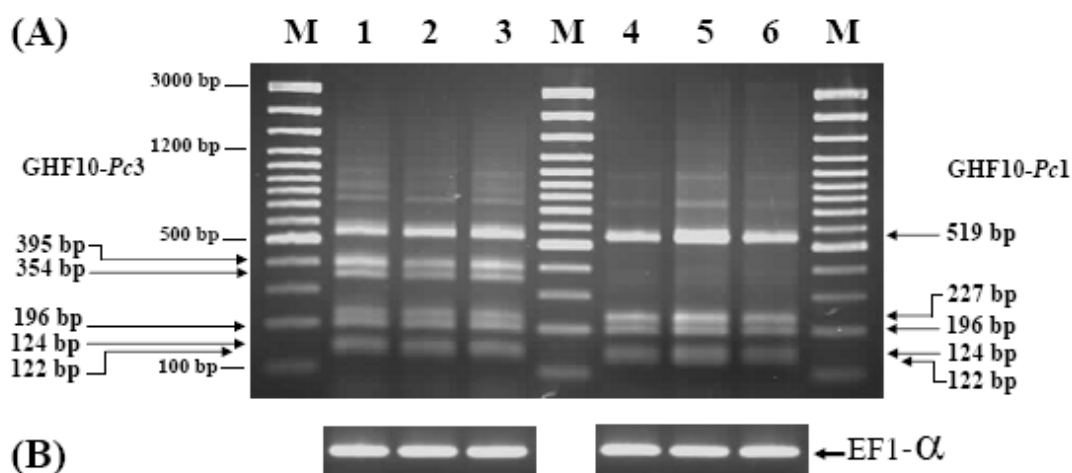
รูปที่ 4B การจัดเรียงตัวของยีน (Genomic organization) ของยีนเซลล์ลูเลส GHF10-*Pc3* ตัวอักษรพิมพ์ใหญ่แสดงลำดับนิวคลีโอไทด์และลำดับกรดอะมิโนบริเวณ exon ตัวอักษรพิมพ์เล็กและแรเงาแสดงบริเวณ intron ตัวอักษรเข้มแสดง start และ stop codon

4.3.7 การตรวจสอบการแสดงออกของยีนด้วยเทคนิค RT-PCR

จากการตรวจสอบการแสดงออกของยีน GHF10-*Pc*1 และ GHF10-*Pc*3 ในหอยเชอรี่ ระยะต่าง ๆ ได้แก่ ไข่หอยเชอรี่ ลูกหอยอายุ 1 และ 10 วัน โดยใช้เทคนิค RT-PCR และใช้ EF1 α เป็น internal control (รูปที่ 5) ผลพบว่าไม่พบมีการแสดงออกของยีนทั้งสองในระยะไข่หอย และเมื่อนำ PCR product ที่ได้มาตัดด้วยเอนไซม์ *Rsa* I ซึ่งสามารถทำให้แยกความแตกต่างระหว่าง GHF10-*Pc*1 และ GHF10-*Pc*3 ได้ ผลพบว่าการแสดงออกของ GHF10-*Pc*1 ในลูกหอยระยะ 1 และ 10 วัน ส่วนยีน GHF10-*Pc*3 พบมีการแสดงออกในระยะ 1 วัน (รูปที่ 6)



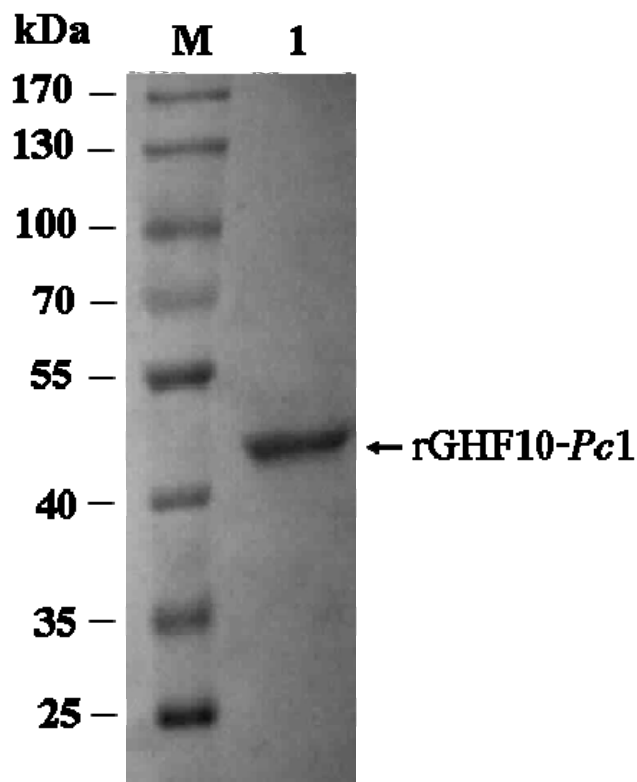
รูปที่ 5 การแสดงออกของยีน GHF10-*Pc*1 และ GHF10-*Pc*3 ของหอยเชอรี่ในระยะไข่หอย (lane 1-3), ตัวอ่อนอายุ 1 วัน (lane 4-6), ตัวอ่อนอายุ 10 วัน (lane 7-9)



รูปที่ 6 รูปแบบการตัด PCR product ด้วย *Rsa* I ของตัวอ่อนอายุ 1 วัน (lane 1-3), ตัวอ่อนอายุ 10 วัน (lane 4-6)

4.3.8 ศึกษาการแสดงออกของยีนเซลล์ใน *E. coli* และการทำบริสุทธิ์ของรีคอมบิแนนท์โปรตีน

จากการศึกษาการแสดงออกของรีคอมบิแนนท์โปรตีน GHF10-*Pc1* โดยใช้ pET28b expression vector เมื่อเหนี่ยวนำการแสดงออกของยีนโดยใช้ 1 mM IPTG และเก็บเซลล์ที่ได้ตั้งแต่ช่วงเวลา 1- 4 ชั่วโมง และวิเคราะห์รีคอมบิแนนท์โปรตีนใน 12.5% SDS-PAGE พบว่ายีน GHF10-*Pc1* สามารถเหนี่ยวนำให้มีการแสดงออกของรีคอมบิแนนท์โปรตีนมากที่สุดที่เวลา 4 ชั่วโมง และมีขนาดน้ำหนักโมเลกุลประมาณ 44 kDa ซึ่งมีขนาดตรงกับที่คาดไว้จากลำดับกรดอะมิโน จากการแยกส่วนของ soluble fraction และ insoluble fraction หลังจากการทำให้เซลล์แตก พบว่ารีคอมบิแนนท์โปรตีน GHF10-*Pc1* อยู่ในรูป inclusion bodies และได้แยกบริสุทธิ์โดยใช้ Ni-NTA column chromatography และวิเคราะห์แถบโปรตีนด้วย SDS-PAGE พบรีคอมบิแนนท์โปรตีนมีขนาดประมาณ 44 kDa (รูปที่ 7)

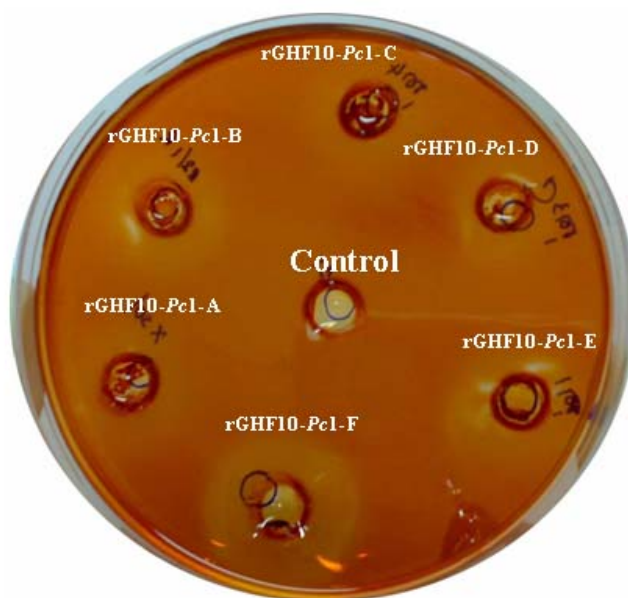


รูปที่ 7 แสดงรีคอมบิแนนท์โปรตีน GHF10-*Pc1* ที่แยกบริสุทธิ์ด้วย Ni-NTA column chromatography

4.3.9 การวิเคราะห์สมบัติของรีคอมบิแนนท์โปรตีน

จากการ refolding รีคอมบิแนนท์โปรตีนบริสุทธิ์ในสารละลาย PBS (pH 7.4) และนำรีคอมบิแนนท์โปรตีนบริสุทธิ์ที่ได้มาศึกษาสมบัติ Cellulase (endo- β -1,4-glucanase) และ Xylanase (endo- β -1,4-xylanase) ของรีคอมบิแนนท์โปรตีนบริสุทธิ์ต่อการเร่งปฏิกิริยาของสับสเตรตสองชนิด คือ carboxymethyl cellulose และ birchwood xylan ตามลำดับ โดยใช้ agar plate assay จากผลการทดลองพบว่ารีคอมบิแนนท์โปรตีนบริสุทธิ์มีแอกติวิตี้ของ endo- β -1,4-xylanase โดยสามารถย่อย birchwood xylan ซึ่งทำให้เกิด clear zone บน agar plate (รูปที่ 8) ในขณะที่การทดสอบแอกติวิตี้ endo- β -1,4-glucanase พบว่ารีคอมบิแนนท์โปรตีนบริสุทธิ์ไม่สามารถย่อย carboxymethyl cellulose ซึ่งไม่เกิด clear zone บน agar plate (รูปที่ 9) จากผลการทดลองแสดงให้เห็นว่ารีคอมบิแนนท์โปรตีนมีแอกติวิตี้ของ endo- β -1,4-xylanase

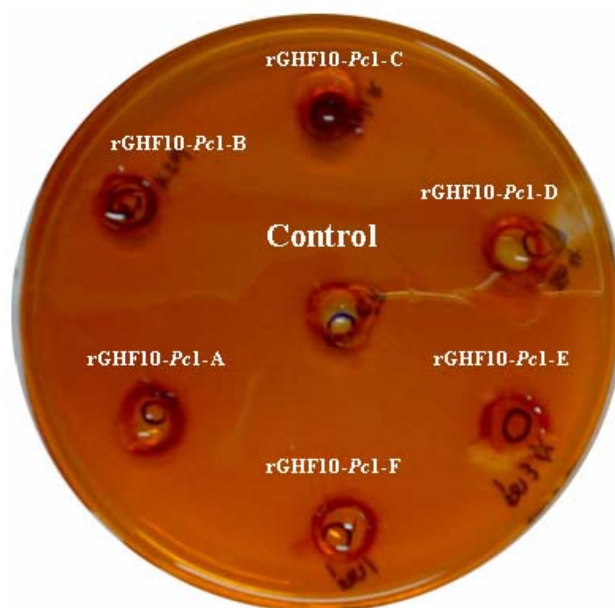
endo- β -1,4-xylanase



รูปที่ 8 แสดงแอกติวิตี้ endo- β -1,4-xylanase ของรีคอมบิแนนท์โปรตีน GHF10-Pc1 ที่แยกบริสุทธิ์

rGHF10-Pc1-A ถึง rGHF10-Pc1-F แสดงปริมาณของรีคอมบิแนนท์โปรตีนตั้งแต่ 1,2,3,4,5 และ 6 ไมโครกรัม ตามลำดับ Control แสดงสารละลาย PBS (pH 7.4)

endo- β -1,4-glucanase



รูปที่ 9 แสดงแอกติวิตี้ endo- β -1,4-glucanase ของรีคอมบิแนนท์โปรตีน GHF10-Pc1 ที่แยกบริสุทธิ์

rGHF10-Pc1-A ถึง rGHF10-Pc1-F แสดงปริมาณของรีคอมบิแนนท์โปรตีนตั้งแต่ 1,2,3,4,5 และ 6 ไมโครกรัม ตามลำดับ Control แสดงสารละลาย PBS (pH 7.4)

สรุปผลการทดลอง

จากการค้นหาและศึกษาลักษณะสมบัติของยีนเซลลูเลสในหอยเชอรี่ GHF10-*Pc1* และ GHF10-*Pc3* พบว่ามีลำดับนิวคลีโอไทด์สมบูรณ์ขนาด 1300 bp และ 1277 bp ตามลำดับ เมื่อวิเคราะห์ลำดับกรดอะมิโนของโปรตีนเซลลูเลส พบว่ามีความคล้ายกับเซลลูเลสในกลุ่ม glycosyl hydrolase family 10 (GHF10) โดยมีกรดอะมิโนที่คาดว่าจะ conserve catalytic nucleophile และ proton donor คือ Glu268 และ Glu167 ตามลำดับ ในการแยกและวิเคราะห์การจัดเรียงตัวของยีนในส่วน exon และ intron พบว่ายีนเซลลูเลส GHF10-*Pc1* และ GHF10-*Pc3* ประกอบด้วยนิวคลีโอไทด์ขนาด 4937 และ 4512 bp และเมื่อนำลำดับนิวคลีโอไทด์ที่ได้มาวิเคราะห์การจัดเรียงตัวของยีนพบว่ายีนทั้งสองประกอบด้วย 9 exon และ 8 intron โดยในบริเวณรอยต่อระหว่างแต่ละ exon และ intron เป็นไปตาม GT/AG rule เมื่อตรวจสอบการแสดงออกของยีนเซลลูเลส GHF10-*Pc1* และ GHF10-*Pc3* ในหอยเชอรี่ระยะต่าง ๆ เช่น ไข่ หอย และลูกหอยอายุ 1 และ 10 วัน ผลพบว่าไม่พบมีการแสดงออกของยีนเซลลูเลส GHF10-*Pc1* ในระยะไข่หอย และพบมีการแสดงออกในระยะ 1 และ 10 วัน ส่วนยีนเซลลูเลส GHF10-*Pc3* พบมีการแสดงออกในระยะ 1 วัน จากการศึกษาการแสดงออกของยีนเซลลูเลสใน *E. coli* โดยเหนี่ยวนำให้เกิดการแสดงออกของยีนด้วย IPTG และวิเคราะห์รีคอมบิแนนท์โปรตีนด้วย SDS-PAGE พบมีการแสดงออกของรีคอมบิแนนท์โปรตีนอยู่ในส่วน inclusion bodies ซึ่งสามารถทำให้บริสุทธิ์ได้โดยใช้ Ni-NTA column chromatography เมื่อวิเคราะห์แถบโปรตีนด้วย SDS-PAGE พบว่ารีคอมบิแนนท์โปรตีนที่ได้มีขนาด 44 kDa ซึ่งเมื่อนำรีคอมบิแนนท์โปรตีนไปตรวจสอบคุณสมบัติของเอนไซม์พบว่ารีคอมบิแนนท์โปรตีนมีแอกติวิตีของเอนไซม์ endo- β -1,4-xylanase

4.3 หนังสืออ้างอิง

- Béguin, P., and Aubert, J.P. 1994. The biological degradation of cellulose. *FEMS Microbiol Rev.* 13: 25-58
- Byrne, K.A., Lehnert, S.A., Johnson, S.E. and Moore, S.S. 1999. Isolation of a cDNA encoding a putative cellulase in the red claw crayfish *Cherax quadricarinatus*. *Gene* 239: 317-324.
- Cleveland, L.R. 1924. The physiological and symbiotic relationships between the intestinal protozoa of termites and their host, with special reference to *Reticulitermes flavipes*. *Kollar Biol Bull Mar Biol Lab.* 46: 117-227.
- Girard, C. and Jouanin, L. 1999. Molecular cloning of cDNAs encoding a range of digestive enzymes from a phytophagous beetle, *Phaedon cochleariae*. *Insect Biochem. Mol. Biol.* 29: 1129-1142.
- Kluepfel D. 1988. Screening of prokaryotes for cellulose- and hemicellulose-degrading enzymes. *Methods Enzymol.* 160: 180-186.
- Lo, N., Tokuda, G., Watanabe, H., Rose, H., Slaytor, M., Maekawa, K., Bandi, C., and Noda, H. 2000. Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. *Curr. Biol.* 10: 801-804.
- Nathan, L., Watanabe, H., and Sugimura, M. 2003. Evidence for the presence of a cellulase gene in the last common ancestor of bilaterian animals. *Proc. R. Soc. Lond. B (Suppl.)* 270: S69-S72.
- Purchon, R.D. 1997. In *The Biology of the Mollusca*, 2nd edn. pp. 24-243. Pergamon Press, Oxford.
- Rosso, M.N., Favery, B., Piotte, C. Arthaud, L., De Boer, L.M. Hussey, R.S. Bakker, J., Baum, T.J. and Abad, P. 1999. Isolation of a cDNA encoding a beta-1,4-endoglucanase in the root-knot nematode *Meloidogyne incognita* and expression analysis during plant parasitism. *Mol. Pl. Microbe Interact.* 1992: 585-591.
- Smant, G., Stokkermans, J.P.W.G., Yan, Y., Boer, J.M.D., Baum, T.J., Wang, X., Hussey, R.S. et al., 1998. Endogenous cellulases in animals: Isolation of β -1,4-glucanase genes from two species of plant-parasitic cyst nematodes. *Proc. Natl. Acad. Sci. USA.* 95: 4906-4911.
- Tokuda, G., Lo, N. Watanabe, H. Slaytor, M. Matsumoto, T., and Noda, H., 1999. Metazoan cellulase genes from termites: intron/exon structure and sites of expression. *Biochem. Biophys. Acta.* 1447: 146-159.

- Wang, J., Ding, M., Li, Y.H., Chen, Q.X., Xu, G.J., and Zhao, F.K. 2003. Isolation of a multi-functional endogenous cellulase gene from mollusc, *Ampullaria crosseana*. *Acta Biochimica et Biophysica Sinica*. 35: 941-946.
- Watanabe, H., Noda, H., Tokuda, G., and Lo, N. 1998. A cellulase gene of termite origin. *Nature*. 394: 330-331.
- Watanabe, H., and Tokuda, G. 2001. Animal cellulases. *Cell. Mol. Life Sci*. 58: 1167-1178.
- Xu, B., Ersson, B., Hellman, U. and Janson, J.C. 2000. Purification, characterization and amino-acid sequence analysis, low molecular mass endo- β -1,4-glucanase from blue mussel, *Mytilus edulis*. *Eur. J. Biochem*. 267:4970-4977.
- Xu, B., Janson, J.C. and Sellos, D. 2001. Cloning and sequencing of a molluscan endo- β -1,4-glucanase from blue mussel, *Mytilus edulis*. *Eur. J. Biochem*. 267:4970-4977.

5. Output จากโครงการวิจัยที่ได้รับทุนจาก สกว.

ผลงานตีพิมพ์ในวารสารวิชาการนานาชาติ

1. **Imjongjirak, C.**, Amparyup, A., and Sittipraneed, S. (2008) Cloning, genomic organization and expression of two glycosyl hydrolase family10 (GHF10) genes from golden apple snail (*Pomacea canaliculata*). *DNA Sequence*. 19: 224-236.

การเสนอผลงานในที่ประชุมวิชาการ

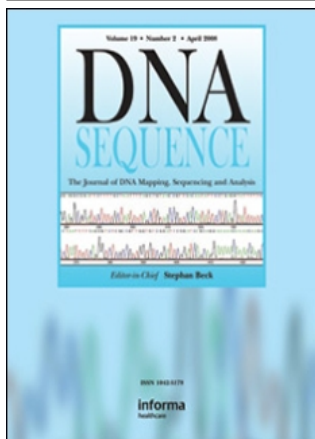
1. **Imjongjirak, C.**, Amparyup, P. and Sittipraneed, S. (2006). Cloning and nucleotide sequencing of cellulase gene from golden apple snail (*Pomacea canaliculata*) in Thailand. The 44th Kasetsart University Annual Conference, January 30 – February 2, 2006. Bangkok, Thailand, pp. 200-208.
2. **Imjongjirak, C.**, Amparyup, P. and Sittipraneed, S. (2006). Cloning, genomic organization and expression analysis of cellulase gene from golden apple snail (*pomacea canaliculata*). 32nd Congress on Science and Technology of Thailand, October 10 – 12, 2006. Bangkok, Thailand, p. 125.
3. **Imjongjirak, C.**, Amparyup, P., and Sittipraneed, S. (2006). Molecular cloning, characterization and expression analysis of two cellulase EGXs from golden apple snail (*Pomacea canaliculata*). 18th Annual Meeting of the Thai Society for Biotechnology, November, 2-3, Bangkok, Thailand, p. 125.
4. **Imjongjirak, C.**, Amparyup, P. and Sittipraneed, S. Two glycosyl hydrolase family10 (GHF10) genes from golden apple snail (*Pomacea canaliculata*): molecular cloning, genomic organization and expression analysis. การประชุม นักวิจัยรุ่นใหม่ พบเมธีวิจัยอาวุโส สกว. ครั้งที่ 7. วันที่ 11-13 ตุลาคม 2550. โรงแรม แอมบาสซาเดอร์ ซิตี้ จอมเทียน จังหวัดชลบุรี. หน้า 383.

6. ภาคผนวก

ผลงานตีพิมพ์ในวารสารวิชาการนานาชาติ

1. **Imjongjirak, C.**, Amparyup, A., and Sittipraneed, S. (2008) Cloning, genomic organization and expression of two glycosyl hydrolase family10 (GHF10) genes from golden apple snail (*Pomacea canaliculata*). *DNA Sequence*. 19: 224-236.

This article was downloaded by:[2007 - 2008 Chulalongkorn University]
On: 6 May 2008
Access Details: [subscription number 780894004]
Publisher: Informa Healthcare
Informa Ltd Registered in England and Wales Registered Number: 1072954
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



DNA Sequence

The Journal of DNA Mapping, Sequencing, and Analysis

Publication details, including instructions for authors and subscription information:
<http://www.informaworld.com/smpp/title~content=t713640135>

Cloning, genomic organization and expression of two glycosyl hydrolase family 10 (GHF10) genes from golden apple snail (***Pomacea canaliculata***)

Chanprapa Imjongjirak^a; Piti Amparyup^b; Siriporn Sittipraneed^c
^a Department of Food Technology, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

^b National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Pathumthani, Thailand

^c Department of Biochemistry, Faculty of Science, Chulalongkorn University,

Bangkok, Thailand

First Published: June 2008

To cite this Article: Imjongjirak, Chanprapa, Amparyup, Piti and Sittipraneed, Siriporn (2008) 'Cloning, genomic organization and expression of two glycosyl hydrolase family 10 (GHF10) genes from golden apple snail (***Pomacea canaliculata***)', DNA Sequence, 19:3, 224 — 236

To link to this article: DOI: 10.1080/10425170701517911

URL: <http://dx.doi.org/10.1080/10425170701517911>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

FULL LENGTH RESEARCH PAPER

Cloning, genomic organization and expression of two glycosyl hydrolase family 10 (GHF10) genes from golden apple snail (*Pomacea canaliculata*)

CHANPRAPA IMJONGJIRAK¹, PITI AMPARYUP^{2,3}, & SIRIPORN SITTIPRANEED⁴

¹Department of Food Technology, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand, ²National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), 113 Paholyothin Road, Klong 1, Klong Luang, Pathumthani, Thailand, ³Shrimp Molecular Biology and Genomics Laboratory, Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok, Thailand, and ⁴Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

(Received 16 August 2006; revised 17 May 2007; accepted 18 June 2007)

Abstract

Two cellulase cDNAs (GHF10-*Pc1* and GHF10-*Pc3*) belonging to glycoside hydrolase family 10 (GHF10) were successfully isolated and characterized from stomach tissue of golden apple snail (*Pomacea canaliculata*), a kind of herbivorous mollusca. Sequencing analysis revealed full-length cDNAs of 1300 and 1277 bp in length, respectively. The open reading frame (ORF) of cellulase cDNA was 1188 and 1191 bp, encoding 395 and 396 amino acid, respectively. Sequence alignment revealed that GHF10-*Pc1* and GHF10-*Pc3* shared high identity with glycosyl hydrolase family 10 (GHF10) and had an overall similarity of 98 and 82% to those of *Ampullaria crosseana* cellulase EGX. A neighbour-joining tree showed a clear differentiation between each species and also indicated that GHF10-*Pc1* and GHF10-*Pc3* from *P. canaliculata* and *A. crosseana* EGX are closely related phylogenetically. The genomic organization of cellulase GHF10-*Pc1* and GHF10-*Pc3* genes was also investigated. The GHF10-*Pc1* and GHF10-*Pc3* genes spanned over 4937 and 4512 bp, respectively. Both genes contained 9 exons interrupted by eight introns. The result verified the endogenous origin of the GHF10-*Pc1* and GHF10-*Pc3* genes. Analysis of RNA by RT-PCR from several ages of *P. canaliculata* revealed that neither gene was expressed in eggs. GHF10-*Pc1* was also expressed in 1- and 10-day-old juvenile snails whereas GHF10-*Pc3* was expressed only in 1-day-old juvenile snails. The result showed that two GHF10-*Pc* transcripts were developmentally expressed.

Keywords: *Pomacea canaliculata*, cellulase gene, cloning, golden apple snail

Database accession number: DQ848667, DQ848668, DQ848669, DQ848670

Introduction

Cellulose, a main component of plant cell walls, is the most abundant biopolymer in the world (Nathan et al. 2003). Cellulose is a linear biopolymer of D-glucose, linked by β -1,4-glucosyl linkages. The degradation of cellulose is considered to be achieved by the synergistic action of three types of cellulase components: exo- β -1,4-glucanase (EC 3.2.1.91), endo- β -1,4-glucanase (EC 3.2.1.4), and β -glucosidase (EC 3.2.1.21) (Beguín and Aubert 1994).

Cellulases belong to the large group of glycosyl hydrolases, of which there are several families based on amino acid sequences similarities. Over the years, the number of glycosyl hydrolase families has grown steadily and currently there are 106 known families recorded in CAZY database (<http://www.afmb.cnrs-mrs.fr/~cazy/CAZY/index.html>).

Cellulase has been shown to exist in plants (Brummell et al. 1994), fungi (Tomme et al. 1995), bacteria (Tomme et al. 1995) and protista (Moriya et al. 1998). Recently, the cellulose hydrolytic enzyme

Correspondence: C. Imjongjirak, Department of Food Technology, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand. Tel: 66 2 2185515. Fax: 66 2 2544314. E-mail: chanprapa.i@chula.ac.th

has been also reported in herbivorous invertebrates, such as arthropods (Watanabe et al. 1997; Xue et al. 1999), nematodes (Smant et al. 1998) and molluscs (Xu et al. 2000). Although, the origin of the invertebrate cellulases was initially explain as being the product of symbiotic microorganisms in the digestive tract, endogenous endo- β -1,4-glucanase genes have been cloned in some herbivorous invertebrates such as the endogenous endo- β -1, 4-glucanase *NtEG* belonging to GHF9 from termites, *Nasutitermes takasagoensis* (Tokuda et al. 1999), endoglucanase *CqEG* belonging to GHF9 from crayfish, *Cherax quadricarinatus* (Byrne et al. 1999), endoglucanase *HG-eng1*, *HG-eng2* and *GR-eng1*, *GR-eng2* belonging to GHF5 from nematodes *Heterodera glycines* and *Globodera rostochienensis* (Yan et al. 1998), endoglucanase belonging to GHF45 from blue mussel, *Mytilus edulis* (Xu et al. 2001), and endoglucanase *HdEG66* belonging to GHF9 from abalone, *Haliotis discus hannai* (Suzuki et al. 2003). It seems likely that most reported cellulases from animal belong to GHF9, GHF5 and GHF45. Recently, the GHF10, a multifunctional cellulase generally involved in microbial degradation of cellulose and xylans was isolated from *Ampullaria crosseana* (Wang et al. 2003). This gene was the first multifunctional cellulase gene that was isolated from animals. It exhibited the endo- β -1,4-glucanase, exo- β -1,4-glucanase and endo- β -1,4-xylanase activity.

The golden apple snail, *Pomacea canaliculata*, is native to South America. These snails were spread out and introduced to Philippines, Malaysia, Thailand, Indonesia, China and the other Asian countries in the 1980s as an aquarium pet and food source. However, the commercial market demand for the snail was poor, many snail-farming projects were abandoned and in many instances the snails escaped and subsequently became a pest to crops causing serious damage to paddy rice and the other vegetables in Asia.

In Thailand, *P. canaliculata* was first discovered in the wild in 1984 (Keawjam and Upatham 1990). Because of its high rate of reproduction, the number of snails consistently increases. The genetic diversity and species-diagnostic markers in the introduced *P. canaliculata* were investigated by PCR-RFLP analysis of COI (Thaewnon-ngiw et al. 2004). Serious destructive damage of rice crops by *P. canaliculata* has been reported since 1996 and it is now regarded as being the most harmful pest to rice seedlings in Thailand.

P. canaliculata are herbivores with an ability to consume large amounts of several kinds of vegetation such as rice seedlings, algae, azolla and other succulent leafy plants. In addition, several enzymatic activities of polysaccharide-hydrolysing enzymes against laminarin, cellulose and xylan in snail viscera have previously been reported (Suzuki et al. 1986). This implies the possible potential composition

of cellulytic enzymes in golden apple snails. However, the molecular information on gene sequence and an unambiguous origin for the corresponding genes in *P. canaliculata* has not been established. Identifying and characterizing the genes involved will be a crucial step in better understanding the cellulase system in the golden apple snail, *P. canaliculata*.

Therefore, the objective of the present study was to clone and characterize the full-length cDNA of cellulase GHF10-*Pc1* and GHF10-*Pc3* genes from the golden apple snail, *P. canaliculata* in Thailand and to evaluate mRNA expression at various stages of development. The genomic organization of cellulase genes was also determined.

Materials and methods

Biological specimens

Golden apple snails (*P. canaliculata*) were collected locally (Bangkok, Thailand). The stomach and foot tissue of each snail was dissected out, immediately placed in liquid nitrogen and stored at -80°C until required. Fertilized eggs and juvenile snails were stored at -80°C until the RNA was isolated.

RNA isolation and genomic DNA extraction

The total RNA was extracted from *P. canaliculata* tissue using the TRIzol Reagent (Gibco-BRL, USA). Contaminant DNA was eliminated by incubation of the total RNA with 1 unit of DNase I for 15 min at 37°C . mRNA was isolated with a QuickPrep mRNA purification kit (Amersham Biosciences, USA). Foot muscle was homogenized followed by incubation with lysis buffer [100 mM Tris-HCl, 50 mM EDTA, and 100 mM NaCl; pH 8.0, 1.0% sodium dodecyl sulfate (SDS), 200 $\mu\text{g}/\text{ml}$ Proteinase K]. Genomic DNA was extracted with phenol/chloroform/isoamyl alcohol (25:24:1), and then re-extracted by chloroform/isoamyl alcohol (24:1). The aqueous phase was transferred to a new tube and precipitated with two volume of ice-cold absolute ethanol. DNA pellet was collected by centrifuging, washed with 70% ethanol and dissolved in 50 μl of water. RNA and DNA concentration was assessed by UV spectrophotometry and agarose gel electrophoresis.

Isolation of cellulase GHF10 cDNA by reverse-transcription PCR (RT-PCR)

Single-stranded cDNA was synthesized from two micrograms of total RNA with an oligo (dT)₁₇ primer and reversely transcribed according to the manufacturer's instructions (Improm-II Reverse Transcription System; Promega, USA). The first-stranded cDNA was subjected to PCR using oligonucleotide primer (GHF10-1F and GHF10-1R; Table I) designed from

Table I. Sequences of oligonucleotide primers used for amplification of GHF10-*Pc1* and GHF10-*Pc3* cDNAs.

Primer	Sequence (5'-3')
GHF10-1F	AGGACCAGAGGAGGGCATCAG
GHF10-1R	ATGCAGRTTCTCRITRTTNACRTC
GHF10-2F	CAAAGCTCTGCATGGGGATGAGCTGCGA
GHF10-3F	TGTCGACGACGCTTCAGTCAAGCGCATG
GHF10-3R	ACTTTATTGCCCTCTGAGTGTGCTCTA
EF1 α -F	CGAGAAGGAGGAAGCCCAGGAGGAAATG
EF1 α -R	GCATCAATGATGGTGAATGTAGAAC

5' untranslated region (UTR) of *A. crossean* EGX cellulase genes (accession number AY285999) and conserved region (DVNNENLH) of GHF10 cellulases.

The amplification reaction and PCR condition are described in Imjongjirak et al. (2005). The amplification products were cloned and electrotransformed into *Escherichia coli* XL 1-Blue. Plasmid DNA was extracted from each recombinant clone and double-stranded sequenced using an automated sequencer (ABI377, PE Applied Biosystems, USA). A core fragment of 600 bp of *P. canaliculata* cellulase cDNA was isolated by RT-PCR.

Cloning of cellulase GHF10 full-length cDNA by RACE-PCR

To obtain the full-length GHF 10 cDNA, a specific primer GHF10-2F corresponding to the cDNA sequence obtained from the internal amplification was synthesized. The 3'-rapid amplification of the cDNA end (RACE) reactions was performed with a SMART RACE cDNA Amplification kit (BD-Biosciences Clontech, USA) according to the manufacturer's instructions. RACE-PCR amplification was described in Amparyup et al. (2007). Amplification was composed of 92°C for 3 min followed by 5 cycles of 94°C for 45 s, 65°C for 45 s and 72°C for 1.30 min, followed by 25 cycles of 94°C for 45 s, 60°C for 45 s and 72°C for 1.30 min. The final extension was carried out at 72°C for 10 min. The expected DNA fragments of 3' RACE-PCR product were cloned and sequenced in both directions.

Amplification of a single fragment representing the full-length cDNA of cellulase GHF10 was performed using *Pfu* Turbo® DNA Polymerase (SRATAGENE, USA) and the specific primer primed at 5' UTR (GHF10-3F) under the following conditions: 3 min initial denaturation at 95°C; 25 cycles of a 94°C denaturation step for 1 min, a 65°C annealing step for 1 min and a 72°C extension step for 4 min. The final extension was carried out at 72°C for 10 min. The PCR product was electrophoretically analyzed through 1.6% agarose gels and visualized under a UV transilluminator after ethidium bromide staining. The resulting PCR product was cloned and sequenced in both directions.

Isolation and characterization of genomic structure of cellulase GHF10-*Pc1* and GHF10-*Pc3* genes

According to our experimentally determined GHF10-*Pc1* and GHF10-*Pc3* cDNA sequences, the GHF10-*Pc1* and GHF10-*Pc3* genes were cloned by a PCR-based strategy. The full-length genes were obtained from the amplification of genomic DNA of the *P. canaliculata* by using the primers GHF10-3F and GHF10-3R located in the 5' and 3' UTR. PCR was performed with the BD Advantage™ Genomic PCR Kit (BD-Biosciences Clontech, USA) that allow amplification of long DNA fragments from genomic DNA. The amplification reaction was described in Imjongjirak et al. (2007). The PCR condition was composed of 94°C for 3 min followed by 5 cycles of 94°C for 1 min, 55°C for 1.30 min and 68°C for 4 min, followed by 25 cycles of 94°C for 1 min, 65°C for 1.30 min and 72°C for 4 min. The final extension was carried out at 72°C for 10 min. The expected DNA fragments of PCR product were cloned and sequenced. Internal sequencing primers were subsequently designed and used for sequencing along the entire length of the insert.

DNA sequencing and data analysis

Plasmid DNA was extracted from each recombinant clones and sequenced for both directions. Nucleotide sequencing was performed by Macrogen Inc. (Korea). DNA sequences were further edited with GENETYX (Software Development Inc.) and blasted against data in the GenBank using BlastN and BlastX (<http://www.ncbi.nlm.nih.gov>). Potential functional motifs in the ORFs were searched using the PROSITE database and SMART analysis (<http://www.smart.embl-heidelberg.de>). Multiple sequence alignments of nucleotide and translated amino acids were performed using Clustal W. Aligned sequences were bootstrapped 1000 times using Seqboot. Sequence divergence between different families of cellulase was calculated based on the two parameter method using Prodist. Bootstrapped neighbour-joining trees were constructed using Neighbour and Consense. All phylogenetic reconstruction programs are routine in PHYLIP and phylogenetic trees are appropriately illustrated using TREEVIEW (<http://www.taxonomy.zoology.gla.ac.uk/rod.html>).

Expression analysis of cellulase GHF10 by reverse transcription-polymerase chain reaction

The RT-PCR was carried out to examine the expression profile of GHF10-*Pc1* and GHF10-*Pc3* genes in different developmental stages. Six individuals of each stage (e.g. 1- and 10-day-old juveniles) of *P. canaliculata* were examined. The first-stranded cDNA was synthesized from 1 µg total RNA and used

as a template for RT-PCR analysis. The GHF10 primers (GHF10-3F and GHF10-3R; Table I) were designed and used in the PCR reaction. As an internal control, elongation factor 1 α was amplified by PCR with the primer EF1 α -F and EF1 α -R (Table I). Composition and amplification conditions were identical to that described above. Ten microlitres of the amplification product was electrophoretically analysed through 1.8% agarose gel electrophoresis. To differentiate GHF10-*Pc1* from GHF10-*Pc3*, PCR product was digested with *Rsa* I and analysed in agarose gel electrophoresis. The expression level of GHF10-*Pc1* and GHF10-*Pc3* transcripts in different developmental stages of *P. canaliculata* was normalized by that of elongation factor 1 α .

Results and discussion

Isolation, cloning and characterization of cellulase GHF10-*Pc1* and GHF10-*Pc3* cDNAs

Two full-length cDNAs of *P. canaliculata* cellulase designated as GHF10-*Pc1* and GHF10-*Pc3*

were successfully isolated from stomach tissue by RACE-PCR method. The full-length cDNA of GHF10-*Pc1* was 1300 bp in length containing 52 bp of 5' UTR, and open reading frame (ORF) of 1188 nucleotides, and 60 bp of 3' UTR. A consensus sequence of polyadenylation signal (AATAAA) was located 18 bp upstream of the poly (A) tail (Figure 1A). The ORF encodes for a 395 amino acid polypeptide with a calculated molecular weight of 44,438 Da and an estimated isoelectric point (pI) of 6.49. The nucleotide sequence of GHF10-*Pc3* was 1277 bp in length and contained a 1191-bp ORF coding for a sequence of 396 amino acid residues. The calculated molecular mass was 44,820 Da with an estimated pI of 6.75 (Figure 1B). These structural characteristics indicate that the cellulase GHF10-*Pc1* and GHF10-*Pc3* cDNA are not derived from prokaryote like intestinal bacteria.

Analysis of the putative signal peptidase cleavage site by SignalP software indicated no potential signal peptide cleavage site. The predicted N-linked glycosylation site (NXS/T) was observed in deduced

```

A AGGACCAGAGGAGGGCATCAGTTTGTGTGTCGACGACGCTTCAGTCAAGCGCATGCCCCG 60
                                     M P A
TGGTGTCTGCTGGTGTCTGGGGTGACACGACATCGACAGACTGAGAAGAAGCGACATAAC 120
G A A G A G V T S D I D R L R R S D I T
GGTTACGCTGAATGTGGTGGTAACATCAACCACGGTCAAGTGAGCATTTCGTGTGTAC 180
V H V N V G G N I N H G Q V S I R V L Q
AAAGAAAAAGGCATTCCCGTTTCGGGACATGTGTGGCCGCTGGGCCTACAACGATGGGTC 240
K K K A F P F G T C V A A W A Y N D G S
CAAAGGAGCATACCGGATTTTCATCCACGACACTACAACCTGGGCGGTGCCAGAAAATC 300
K G A Y R D F I H Q H Y N W A V P E N S
ACTCAAGTGGGCTAGCATCGAACCTAACAGGGGACAAAAGAATTATCAGCCTGGCCCTAAA 360
L K W A S I E P N R G Q K N Y Q P G L N
CATGCTTCACGGACTGAGAAATCACGGGATTAAGGTGAGAGGTCACAACTGGTGTGGTC 420
M L H G L R N H G I K V R G H N L V W S
TGTCGACAATACGGTGACAGAACTGGGTCAAAGCTCTGATGGGGATGAGCTGCGAAAGGT 480
V D N T V Q N W V K A L H G D E L R K V
TGTCATGACACATCGTGGAAACCATCAACACATTTAAGGGCTAGTGGAGCACTGGGA 540
V H D H I V E T I N T F K G L V E H W D
TGTTGAACAACGAGAACCTGCATGGCCAGTGGTACCAGCATCAACTGAATGACAATGGCTA 600
V N N E N L H G Q W Y Q H Q L N D N G Y
CAACCTGGAAGTGTTCGTATCGCACACGCCCGCCGACCCCAACGTCAAACCTTCTCTCAA 660
N L E L F R I A H A A D P N V K L F L N
CGACTACAACGTTGTGTCCAACAGTTATTCAACAACGACTATCTTCGACAAGGTCAACA 720
D Y N V V S N S Y S T N D Y L R Q G Q Q
GTTTAAGGCCGCTAATGTGGGTCTTTACGGTTTGGGTGCTCAATGCCACTTTGGCGACGA 780
F K A A N V G L Y G L G A Q C H F G D E
AGCCGACCCAGAACCCGGTACTAAGCAACGCTTGGATACTTTAGCTCAAGTGGGCGTGCC 840
A D P E P G T K Q R L D T L A Q V G V P
CATCTGGGCCACTGAGTTGGATGTGGTAGCTTCGGATGAGAACAGACGAGCGGATTCTTA 900
I W A T E L D V V A S D E N R R A D F Y
CGAGCAGCGCTGACAGTCTGTACGGCCATCATGCCGTGGAGGGCATCCTGATGTGGGG 960
E H A L T V L Y G H H A V E G I L M W G
CTTCTGGGACAAGGCCACTGGCGTGGTGCCAGAGCTGCTCTGTGTGTCGGAGACACCT 1020
F W D K A H W R G A R A A L V V G D N L
GCAGCTGACGCGCGCGGACGTCGCTGCTGGAGCTCTTTGAGCACAGGTGGATGACAGA 1080
Q L T A A G R R V L E L F E H R W M T D
CGAGAGCACAACCTGGCAGCGGGCACCCAGTTACAGTACGCGGTTTCCATGGCGACTA 1140
E T H N L A A G T Q F T V R G F H G D Y
CGAGGTGCAAGTCATCGTCCAGGGTCAAGAGCACACCAACCTGAGGCAGACGTTCTCGTT 1200
E V Q V I V Q G Q E H T N L R Q T F S L
GGGCAACGTTCCCAACCGTCAACATTAACATTAGCTAGAGCGACACTCAGAGGGGCAAT 1260
G N G P H T V N I N I S *
AAA GTTGAATGATGACACAACAAAAA

```

B	TTGTCGACGACGCTTCAGTCAAGCGC ATG CCCGCGGGTGCTGCTGCTGATGTGAGCA	60
	M P A G A A A A D V S	
	GCGAGATCGACAGACTGAGAAGAAGCAACATAGTGGTTCACGTGACAGCTGGTGGTAACA	120
	S E I D R L R R S N I V V H V T A G G N	
	TCAGCCACGGTGAAGTAAACATACGGGTGGTACAAAAGAAGAGTCGTTCCCGTTCGGAA	180
	I S H G E V N I R V V Q K K K S F P F G	
	CCGCTGTGGCCGCTGGGCCTACAACATGATTCCAAGACAAAATACCGGGATTTCATCC	240
	T A V A A W A Y N N D S K T K Y R D F I	
	ACCAGCACTACAACCTGGGCGGTGCCAGAAAACGAACCTCAAGTGGCGCACCATCGAACCTA	300
	H Q H Y N W A V P E N E L K W R T I E P	
	CCAGGGGACATAAGAACTATCAGCCTGCCCTGACCATGATTCACGGACTGAAAAGTCACG	360
	T R G H K N Y Q P A L T M I H G L K S H	
	GGATTAAGGTGAGAGGTCAACCTGGTGTGGTCTGTCAATTCAACGGTGCAGAGCTGGG	420
	G I K V R G H N L V W S V N S T V Q S W	
	TCAAAGCTCTGCATGGGGATGAGCTGCGAAAGGTTGTCCATGACCACATTGTGGAAACCG	480
	V K A L H G D E L R K V V H D H I V E T	
	TCAACACATTTAAGGGCTTAGTGGAGCACTGGGACGTGAACAACGAGAACCTGCATGGCC	540
	V N T F K G L V E H W D V N N E N L H G	
	AGTGGTACCAGCAACAACCTGAATGACCCGAACCTACAACATAGAAGTGTCCGTATCGCAC	600
	Q W Y Q Q L N D P N Y N I E L F R I A	
	ACGCTGCCGACCCCAACGTCAAACCTCTTCTCAACGACTACAACGTGGTGGCTTACGGTG	660
	H A A D P N V K L F L N D Y N V V A Y G	
	CCGCAACATAGCCTATCTTCAAGGTCAACAGTTTAAGGCCGCTAATGTGAGTCTTT	720
	A A T N A Y L Q Q G Q Q F K A A N V S L	
	ACGGTTTGGGTGCCAGTGCCACTTTGGCGATGAAGCTAACCCAAACGTGCTGGGATGA	780
	Y G L G A Q C H F G D E A N P N V A G M	
	AGCAACATCTGGATATTTTAGCTCAAGTGGGGTTGCCCATCTGGGCCACTGAGTTGGATG	840
	K Q H L D I L A Q V G L P I W A T E L D	
	TGTTAGCTACGGACGAGAACAACGAGCGGACTTCTACGAGCACGCGCTGACAGCCCTGT	900
	V L A T D E N K R A D F Y E H A L T A L	
	ACAGCCATCATGCCGTGGAGGGCATCCTGATGTGGGGCTTCTGGGACAAGGCCCACTGGC	960
	Y S H H A V E G I L M W G F W D K A H W	
	GGCATGAACGAGCTGCTCTTCTGTGCGAGACAACCTGCAGCTGACGGCGGCGCGACGTC	1020
	R H E R A A L L V G D N L Q L T A A G R	
	CGGTGCTGGAGCTCTATGAGCACAGGTGGATGACAGACGAGACGACACAACCTGGCAGCGG	1080
	R V L E L Y E H R W M T D E T H N L A A	
	GCACCCAGTTACAGTACGCGGTTTCCATGGCGACTACGAGGTGCAGTGATCTACCAAG	1140
	G T Q F T V R G F H G D Y E V H V I Y Q	
	GTAGGAGCGCACCAACCTGAAGCAGACGTTACGTTGGGCAACGAGCCACACCGTCA	1200
	G Q E R T N L K Q T F T L G N A A H T V	
	ACATCAACATTAGCT TAG AGCGGACACTCAGAGGGC AATAAA TTGAATGATGACACAACAA	1260
	N I N I S *	
	AAAAAAAAAAAAAAAA	1277

Figure 1. The full-length nucleotide (above) and deduced amino acid (below) sequences of cellulase GHF10-*Pc1* (A) and GHF10-*Pc3* (B) cDNAs from *P. canaliculata*. The start codon, stop codon, putative polyadenylation signal (AATAAA) and N-linked glycosylation site (NXS/T) are in bold and underlined.

GHF10-*Pc1* and GHF10-*Pc3* protein (NIS, 393rd) and (NIS, 394th), respectively. Two cysteine residues were observed and suggested the presence of one disulfide bond in proteins. The full-length sequence was deposited in the NCBI GenBank under accession number DQ848667 and DQ848668.

A database search performed with the BLAST software shows that the *P. canaliculata* GHF10-*Pc1* and GHF10-*Pc3* deduced amino acid sequences shared an identity of 98 and 82%, respectively to those of *A. crossean* cellulase EGX (accession number AY285999). The cellulase EGX of *A. crossean* is a multifunctional cellulase with the activity of exo- β -1,4-glucanase, endo- β -1,4-glucanase and endo- β -1,4-xylanase (Wang et al. 2003).

When the predicted GHF10-*Pc1* and GHF10-*Pc3* sequence was analysed with the SMART software, it appeared that it contained a characteristic domain of family 10 of glycosyl hydrolases (GHF10).

In GHF10-*Pc1*, the domain started at position 84 aa and ended at position 336 aa. The corresponding *E*-value for the glucanase-like domain prediction was 5.3×10^{-13} . In GHF10-*Pc3*, the domain started at position 84 aa and ended at position 337 aa with the corresponding *E*-value of 3.4×10^{-14} . It is noted that both GHF10-*Pc1* and GHF10-*Pc3* appears to consist of only a single catalytic domain and lack both a cellulose-binding domain (CBD) and a spacer sequence. This is in agreement with some fungal cellulases (Yan et al. 1998), *A. crossean* EGX cellulase (Wang et al. 2003) and in endo- β -1,4-glucanase from marine blue mussels, *M. edulis* (Xu et al. 2001). The absence of a CBD and associated peptide linker has been observed in some endoglucanases and this appears to affect substrate specificity but does not prevent enzyme activity (Gilkes et al. 1991; Wang and Jones 1995a,b).

By multiple sequence alignment (Figure 2A), residues participating in catalytic action as proton

A

Clostridium	AKWYNSNSSGCGNVSYANADYLNNCAENGKIVRGHCIFWEVEEOWSPSLKGLTGDALMA
Bacillus	AKWYANPEERGRKITIYEKADAMLNPAHRQLPVRGHCHALFWEVEDANPSWLRLSNHEVVEA
Hordeum	LKWYHTBQQCOLNYADADALLAFCDRLGKTVRGHCVFWSVDGDVQOAWKNLKKDQLRSA
Triticum	LKWYHTBAQQCOLNYADADALLAFCDRLGKHVRGHCVFWSVDGDVQOAWKNLKKDQLRSA
Arabidopsis	LKWYWTEPEEGCKLNYQDADDMLNLCSNNNIEIRGHCIFWEVCATVQOITQNMMOTDLNNA
Nicotiana	LKWYWTBAQQGNFNFKDADELLDFCTKNNIQVRGHCIFWEVVGTVQAANVQS LNKKNDLMTA
Oryza	LKWYHTBSRQAGLYADADALLAFCDRLGKHMVCRGHCIFWEVDSAVQOAWKALPADELAA
GHF10-Pc1	LKWASIEPNRQCKNYQPLNMLHLGRNHGIKVRGHNLVMSVDNTVQNMVKALHGDRLRVV
Ampullaria	LKWASIEPNRQCKNYQPLNMLHLGRNHGIKVRGHNLVMSVDNTVQNMVKALHGDRLRVV
GHF10-Pc3	LKWRTIEPTRCHKNYQPALTMIHGLKSHGIKVRGHNLVMSVNSTVQSAWKALHGDRLRVV
	** * : * . : **** : . * : :
	↓
Clostridium	IDARLESVPVPHFRCKLFHMDVNNEMLHGDFPKSRLG-ESTWPYMFKRARELPDAKLFWN
Bacillus	MKKRLEHAGNHFKGRFRHWDVNNEMMHGSFPFKDRFG-KNIWKWMYEETTKPIDQALLFWN
Hordeum	MQSRLEGVLVSRYACRFKHYDVNNEMLHGRFFRDRLGDEDVPAYMFKEVARLDPEPVLFVN
Triticum	MQSRLEGVLVSRYACRFKHYDVNNEMLHGRFFRDRLGDEDIPAYMFKKEVARLDPEPALFVN
Arabidopsis	VQNRLTDLNLRYKCKFKHYDVNNEMLHGSEFYQDKLG-KDIRVMNPFKTAHQLDPSATLFWN
Nicotiana	VQNRLTGLLTRYKCKFKHYDVNNEMMHGSFYQDRLG-KEIRVMNPFKTAHQLDPSPIFLVN
Oryza	VASHINGLLTRYKCFRHYDVNNEMLHGSEFYQDKLG-AGARAAMFKAASELPDALLFWN
GHF10-Pc1	VHDHIVETINTFKGLVBHWDVNNENLHGQWYQHQLNDNGYNLELFRIAHAAPNPVKLFVN
Ampullaria	VHDHIVETINTFKGLVBHWDVNNENLHGQWYQHQLNDNGYNLELFRIAHAAPNPVKLFVN
GHF10-Pc3	VHDHIVETVNTFKGLVBHWDVNNENLHGQWYQQQLNDPNYNIELFRIAHAAPNPVKLFVN
	: :: : * . *.***** : * : : : . : . ** . *** *
Clostridium	DYNIITYV----EGDAMIROI EWLLQNGAEIDIGVGGCHFEDDEV--LVVKARLDNEA
Bacillus	DYNNVISYG----EHHAMKAHINELRQLGAPEIAEVGVCCHFEEVRDPE--VIVKERLDVNEA
Hordeum	DYNNVECGN-DPNATPEKMYAEQVANLQSCGAVVRGTGLGCHVQNPVGE--VICAA-IDREA
Triticum	DYNNVERAN-DPNATPEKMYAEQVANLRQCGAVVGIGLGHVQNPVGE--VICAA-IDREA
Arabidopsis	DYHIEDGC-DEKSCEPKYTEQILDLQEKGAPVGGIGLGGHIDSPVGP--IVCSA-LDKLG
Nicotiana	DYHVDEGS-DTRSSPEKYTEIHLDLQEHGAPVGGIGLGGHIDSVPGP--IVCSA-LDKLG
Oryza	DYNNVEGACVDVRAETPEANIATOVTLQEQGAAVGVGVLGGHVTAPVGA--VVRAA-IDREA
GHF10-Pc1	DYNNVVSNS----YSTNDYLROGQQGFKAANVGLYGLGACCHFGDEADP-EPGTKQRLDITLA
Ampullaria	DYNNVVSNS----YSTNDYLROGQQGFKAANVGLYGLGACCHFGDESDEP-EPGTKQRLDITLA
GHF10-Pc3	DYNNVAVG----AATNAVLOGQQGFKAANVSVLYGLGACCHFGDEANPINVAGMKQHLDTLA
	**:: : * : : : . : : * * . : * *
	↓
Clostridium	TLCLPTIWTFETEDSKTPDVNRKAENLENLYRIASFHPAVEGITIMWGFWAGNHWRGQDAAI
Bacillus	ELCLPTIWTFETDSVHPDPNRADNLEALYRVAFSHPAVKCVLMWGFWAGNHWRGEHAAIV
Hordeum	KTCVPTIWTFETDMPPEYDGLRAKDLEVLYREAYAHFAVEGIVFWGFMQGTMMWR-QNALV
Triticum	KTCVPTIWTFETDMPPEYVNSRAKDLEVLYREAYAHFAVEGIVFWGFMQGTMMWR-ENSILV
Arabidopsis	ILCLPTIWTFETDVSSSVNEHIRADDLEVMNWEAFGHFAVEGIMLGWFELFMSR-DNSHLV
Nicotiana	IILCLPTIWTFETDVSSSGNEYIRADDLEVMLREAYAHFAVEGIMLGWFELFMSR-PNAHLV
Oryza	VLCLPTIWTFETDVSASANEHVRADDEAMLREAYAHFAVDGVVWGFWELFMSR-DDAHLV
GHF10-Pc1	QVCVPTIWATELDVVASDENRRADFYEHALTVLYGHHAIVEGILMWGFDWKAHWARGAALV
Ampullaria	QVCVPTIWATELDVVASDENRRADFYEHALTVLYGHHAIVEGILMWGFDWKAHWARGAALV
GHF10-Pc3	QVCVPTIWATELDVLTADENKRADFYEHALTALYSHHAVEGILMWGFDWKAHWHERAAIL
	::* * * : * * . : * * *: : * * *
Clostridium	DHD-WTYNEAGKRYQALL
Bacillus	NYD-WSLNEAG-
Hordeum	DAD-GTVNEAGQMFLNLQ
Triticum	DAD-GTVNEAGQMFLNLQ
Arabidopsis	NAE-GDNVNEAGKRFLAVK
Nicotiana	NAE-GDINEAGKRYLALK
Oryza	DAE-GEVNEAGRRLQLK
GHF10-Pc1	VGDNLQLTAAGRRVLELF
Ampullaria	VGDNLQLTAAGRRVLELF
GHF10-Pc3	VGDNLQLTAAGRRVLELY
	. : *

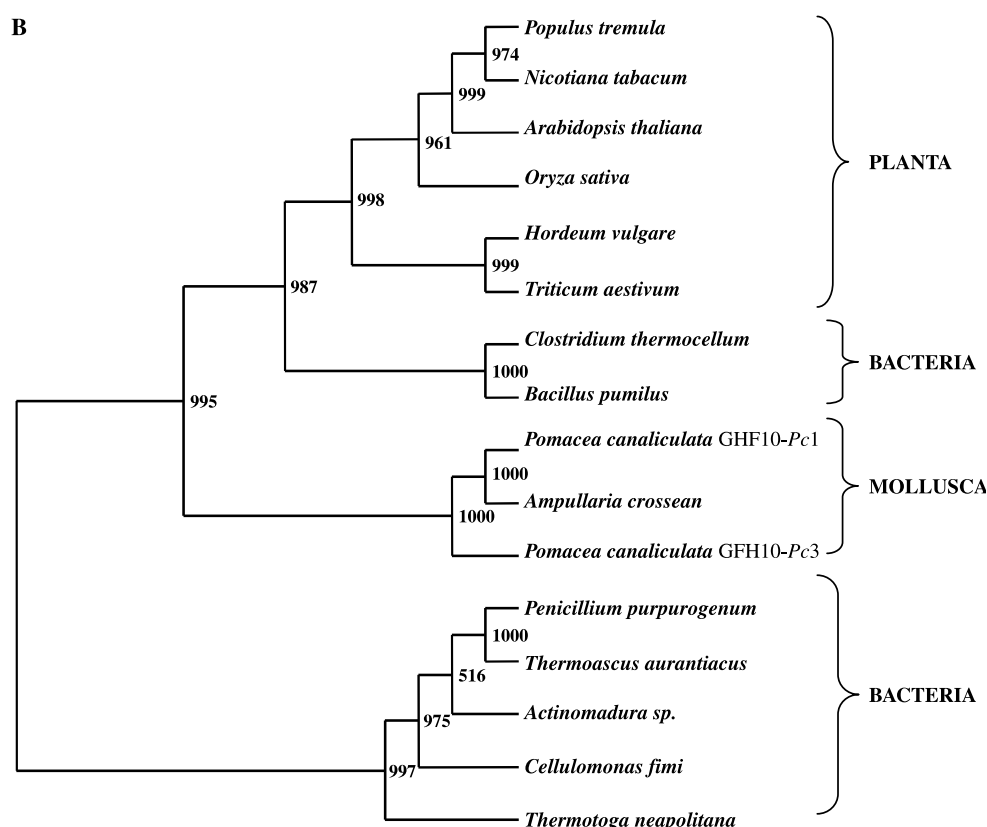


Figure 2. Multiple alignments of deduced amino acid sequence of catalytic GHF10 domain from various species (A) and a bootstrapped neighbour-joining tree (B). *P. canaliculata* cellulase gene (GHF10-Pc1 and GHF10-Pc3, this study), *A. crossean* (AAP31839), *Oryza sativa* (BAD52808), *Arabidopsis thaliana* (BAB83869), *Populus tremula* × *Populus tremuloides* (AAX33301), *Nicotiana tabacum* (AAZ79232), *Triticum aestivum* (AF156977), *Hordeum vulgare* (AAB51668), *Bacillus pumilus* (AF466829), *Clostridium thermocellum* (EAM47104), *Thermotoga neapolitana* (Q60041), *Actinomadura* sp. (AAA17888), *Penicillium purpurogenum* (AAF71268), *Cellulomonas fimi* (AAZ76373) and *Thermoascus aurantiacus* (AF 127529). Black indicates complete conservation. Grey indicates conservation in three or more species. The conserved catalytic nucleophile and proton donor of Glu268 and the Glu167 are arrowed. Values at the node of a bootstrapped neighbour-joining tree indicate the percentage of times that the particular node occurred in 1000 trees generated by bootstrapping the original deduced protein sequences.

GHF. Accordingly, *P. canaliculata* GHF10-Pc1 and GHF10-Pc3 is considered to be the second animal cellulase that belongs to family 10 GHF.

To determine the relatedness of the GHF10-Pc1 and GHF10-Pc3 with other sequences, a phylogenetic tree was constructed based on the amino acid sequence of catalytic GHF 10 domain of GHF10-Pc1 and GHF10-Pc3 and other GHF 10 members. A bootstrapped NJ tree constructed from a sequence divergence of deduced amino acids (Figure 2B) clearly indicates that GHF10-Pc1 and GHF10-Pc3 is more closely linked to mollusca (*A. crossean*) cellulase than with plant cellulases and bacterial cellulases.

Genomic organization of cellulase GHF10-Pc1 and GHF10-Pc3 genes

To identify the genomic structure of GHF10-Pc1 and GHF10-Pc3 genes and to confirm the existence of the GHF10-Pc1 and GHF10-Pc3 gene in the *P. canaliculata* chromosomal DNA, a genomic polymerase chain

reaction (PCR) approach was performed. The genomic PCR amplification generated two bands of approximately 4.6 and 5-kb fragment. PCR products were cloned and sequenced. The GHF10-Pc1 and GHF10-Pc3 gene sequences have been deposited in the GenBank nucleotide sequence database under the accession no. DQ848669 and DQ848670.

The GHF10-Pc1 and GHF10-Pc3 genomic sequences were 4937 and 4512 bp in length, respectively. The intron–exon boundaries of the GHF10-Pc1 and GHF10-Pc3 genes were determined by aligning the genomic DNA with cDNA of each corresponding gene (Figure 3A, B). Both genes were composed of 9 exons interrupted by 8 introns. The sequence of GHF10-Pc1 and GHF10-Pc3 gene was in good agreement with the cDNA sequence. The sixth intron was the shortest (234 bp in GHF10-Pc1 and 232 bp in GHF10-Pc3), the last eighth intron was the longest (1071 bp in GHF10-Pc1 and 671 bp in GHF10-Pc3) (Table II). The size of the introns between *P. canaliculata* GHF10 and *A. crossean* EGX was quite

different. The length of intron 2 in *P. canaliculata* GHF10-*Pc1* and GHF10-*Pc3* (283 and 418 bp, respectively) was approximately 3- or 2-fold smaller than that of intron 2 (1032 bp) in *A. crossean* EGX.

The sequences of all the exon-intron boundaries conformed to the typical eukaryotic splice sites, including an invariant GT at the intron 5' boundary and an invariant AG at its 3' boundary. In contrast, the exon-intron boundaries of corresponding introns in *A. crossean* EGX conformed to GT-AG rule only in intron number 1, 5, 7 and 8 (Wang et al. 2003). Therefore, in spite of the high similarity of *P. canaliculata* GHF10-*Pc1* and GHF10-*Pc3* cDNA to *A. crossean* EGX cDNA, the difference in genomic organization of these genes was observed.

Itakura et al. (2006) reported successful identification of termite species by using cDNA and genomic DNA sequences of termite origin. A pair of common primers that amplified 381-bp fragments from cDNAs encoding the endo- β -1,4-glucanases (EGases) of *Coptotermes formosanus* Shiraki and *Reticulitermes speratus* (Kolbe). The cDNAs from *C. formosanus* and *R. speratus* and genomic DNA from *R. speratus*, were amplified by using this primer pair and then cloned and sequenced. Sequences amplified from

C. formosanus cDNA displayed 97–99% identity to cDNA encoding the EGase of *C. formosanus* (*CfEG*), and 92–94% identity to cDNA encoding the EGase of *R. speratus* (*RsEG*). By contrast, cDNA from *R. speratus* displayed 99–100% identity to *RsEG* cDNA and 93–94% identity to *CfEG* cDNA. *CfEG* and *RsEG* cDNAs can therefore be used as markers for the identification of these termite species. Accordingly, the different in intron size of the cellulase genes in two snail species may be used as markers for species identification.

The GC content analysis showed a slightly greater thermal stability in exons (31–58%) than in introns (26–35%) (Table II). Intron 1, 4 and 5 interrupt ORFs after the 2nd of the codons (type 2 intron) whereas the remaining introns interrupt ORFs between two codons (type 0 intron).

All of intron sequences from GHF10-*Pc1* and GHF10-*Pc3* genes were examined to identify potential homology between introns. A strong sequence homology existed between the corresponding intron of GHF10-*Pc1* and GHF10-*Pc3*. Intron numbers 1, 4–7 of GHF10-*Pc1* and GHF10-*Pc3* shared 86–97% identical internal sequence as well as similarity in size. In contrast, the size of the last GHF10-*Pc1* intron

```

A CGACGCTTCAGTCAAGCGCATGCCCCGCTGGTCTGCTGGTGCTGGGGTGACCAGCGACAT      60
      M P A G A A G A G V T S D I
CGACAGACTGAGAAGAAGCGACATAACGGTTCAgtgctgctgctttcaatatataacaa      120
      D R L R R S D I T V H
cctcacttagttctgcattgataaaatatctcgtcagtagtgtagtaccattctgtgtgtat      180
caagaaagtaacaccggttggtataccatttttaaaagtaaaagcaatttaattctggaatag      240
tacaatacttctgtatgtaagcatatataattcatgattttgtagtaataataagataa      300
ttgtaacataaaacggatagataatataataacaattcttatgtcacgggtcatttcaagttt      360
gcacgatatcctggaatcaagaaccgaatagtaattaaggccgaacgttgccgtaga      420
aaatcgtgtgcagcccgctcatgggatgtaataactaaagaaacacaataacatccggcag      480
ccgtcgtcaagtgtgtaaaatctgacctgaaatccattcacgaagttgatgatttcttat      540
atcagCGTGAATGTTGGTGGTAACATCAACCACGGTCAAGTGAGCATTCGTgtaagtcat      600
      V N V G G N I N H G Q V S I R
agatgaagagcatcatccgcctgaggacaggggcaataaaaaagcaagctcatcaaagg      660
tgtcaacctcgagaatgtgccttaagtaaaacttgattgcgaggtacagttgaccacgtga      720
acagataggagagaagagacaaaaaatgaaggatacaatagaagatgtaggaataata      780
ttaattttaataacttattttatattttatgttaattgtataaaagccaccagactaaat      840
ctttgttcttcaaggtaactgtcgttgattacagGTGTACAAAAGAAAAGGCATTCCC      900
      V L Q K K K A F P
GTTCGGGACATGTGTGGCCGCCTGGGCCTACAACGATGGGTCCAAAGGAGCATACCGGGA      960
      F G T C V A A W A Y N D G S K G A Y R D
TTTCATCCACCAGCACTACAACCTGGGCGGTGCCAGAAAACCTCACTCAAGTGGGCTAGCAT      1020
      F I H Q H Y N W A V P E N S L K W A S I
CGAACCTAACAGGgtacattttttatattatgagatttataaatttaacataaattgatat      1080
      E P N R
tatattaataaaaaatcttctttaacgttttacttccagtttgacgatagtcctttaatca      1140
agaactgaaatattttcaataactgaatgttgctgcatataaaatcgtttgagcctgtc      1200
atttgatgttactactaaagaaatacaataacatataattggaatgatgtttaagggtct      1260
tctgatttatactgttcaggagagagaatagtttaaaagggtactccattcacagcaagga      1320
aagacaaagttaccggttgcatattcatagttttactttctaaatcgtgtgtgtgtgtg      1380
tatataatattatgtctgagcgcgcttgaaatgcttaagtcgacaacttaatgtaacagat      1440
tagttttatataagagtataactaaaaactgggttgctcctgctagccaaagaaatggggtt      1500
ttgggggtatttgataatcgcacgtgtgtcttcttctgtgaataaaataaacaagtaaaaa      1560
ttctggagggaacatacttttgaggcactcgacattgaggcttccgttatgatcgccatt      1620
taatgcttgccctctctcttgacatataagaagggaataataatgactgttgattgtg      1680
tttttcagGGACAAAAGAAATTATCAGCCTGGCCTAAACATGCTTCACGGAAGTGAATC      1740
      G Q K N Y Q P G L N M L H G L R N
ACGGgttagtaaacattagtagactgaacccccatcaccgacattataatcatcatcgt      1800
      H G

```

cgacttcattatcgtataataatgaaagaattggcgaattttaattatttaagaattatt	1860
aacataaattttaattaaacattttaataaaatcaaagaaagtttagcgtttcagagattgctc	1920
ctcccccatgtattttactattttctatcacttcgtttaaagcaagttctcagactcagttc	1980
tgtcttttttgataattccttaaaatagtcctacaaatgaaatagtaactgaaatgtctca	2040
cattgatacgaaatgaagaaatattcacttacactaaggttaacacatgtcgtgtaca	2100
gaggtgtttgttagaaatgttttagttatagttttcttgcatccactcgagatgtttaaaaca	2160
gGATTAAAGGTGAGAGGTACAACTGGTGTGGTCTGTCGACAATACGGTGCAGAACTGGG	2220
I K V R G H N L V W S V D N T V Q N W	
TGAAGGCTCTGCATGGGGATGAGCTGCGAAAGGTTGTCCATGACCACATCGTGGAAACCA	2280
V K A L H G D E L R K V V H D H I V E T	
TCAACACATTTAAGGGCCTgtaagtttatatcatttgttactttctttctaaaaatgtcag	2340
I N T F K G L	
tagttttattttttattttttatcttgtttatttggagaacaaaatttactgtttctacata	2400
caccaaacccagatgataccagttacaactactgaattttatgttgtaaaaaatattgtttgaa	2460
ccttatagcaatagactgccactaatttaagaagtcacgttgtcaagtgtagtttttaa	2520
ctcactcatcaacagatcaagcaagaggttttttctgttttgaccagAGTGGAGCACTGG	2580
V E H W	
GATGTGAACAACGAGAACCTGCATGGCCAGTGGTACCAGCATCAACTGAATGACAATGGC	2640
D V N N E N L H G Q W Y Q H Q L N D N G	
TACAACCTGGAAGTGTTCGGTATCGCACACGCCGCCGACCCCAACGTCAAACCTCTTCCCTC	2700
Y N L E L F R R I A H A A D P N V K L F L	
AACGCTACAACGTGTGTCCACAGTATTCAACAAACgtgagtgaaatgagaaaaacttg	2760
N D Y N V V S N S Y S T N	
cttctacatcacctgagaacaaatgtttgcaagcatcgtatgctttattttctgcctaaaa	2820
tatactcttgtgtcttttagcattatagtccttctgtgccttgcctatgttgccgaacaaatg	2880
caacttgtactgctccatactagtctgtagtgctctatttttttatctgtaaaaacgaat	2940
atcatcttcacaaagtctttgcggtgtgtgacagGACTATCTTCGACAAGGTCAACAGTTT	3000
D Y L R Q G Q Q F	
AAGGCCGCTAATGTGGGTCTTTACGGTTTGGGTGCTCAATGCCACTTTGGCGACGAAGCC	3060
K A A N V G L Y G L G A Q C H F G D E A	
GACCCAGAACCCGGTACTAAGgtaagaaaaataatataaacatccttcgcctttctcatt	3120
D P E P G T K	
ccccattttattattattctaaaaaataaaaacaaaggtacgcacaacaatgaattagttc	3180
tacaatgcagtaagatacaagatttcagcttctgtccgtatcctttttttctatttagtta	3240
tttgcatttttttactattatttgttaattttacgttaaatggaataacatatctttatgt	3300
tgagaaatcaaaaagtccttcatttttatagctgtggatgtaaaaaatactcatctaactgc	3360
ttctatactccaaaagctgttttagctacatgtgggtttttttctaatctcagCAACGTC	3420
Q R	
TGGATACTTTAGCTCAAGTGGGCGTGCCCATCTGGGCCACTGAGTTGGATGTGGTAGCTT	3480
L D T L A Q V G V P I W A T E L D V V A	
CGGATGAGAACAGACGAGCGGATTTCTACGAGCACGCGCTGACAGTCTGTACGGCCATC	3540
S D E N R R A D F Y E H A L T V L Y G H	
ATGCCGTGGAGGGCATCTGTATGTGGGGCTTCTGGGACAAGGCCACTGGCGTGGTGCCA	3600
H A V E G I L M W G F W D K A H W R G A	
GAGCTGCTCTTGTGTGCGGAGACAACCTGCAGgtgtgttgggttacaattaacctaattgg	3660
R A A L V V G D N L Q	
gaatcttacaggtattatagcacagccagcatccagacagacatttaggaatcactgtca	3720
tgatgagaacacattggattcttcttttattagaattttctaccagaactacccttgact	3780
tcaactttttcttttcttataagagacatctgacaaatgcttactcagtgatctacattg	3840
ttagtccagattatgcgtgtatgtttttccattttagaattacttttagcacaacacagac	3900
caaaacagattaccatctatgtgaggacaaagatagctgttctgtgtttataatgaata	3960
atattgttatgataaaactgcaaaatactcagtttgaagtagacatttttccatgaaaa	4020
ttcgcaaaaaaaagaaagaccggaatgaaccatttgaacaggtttaaaaaattgttta	4080
cagcaaaagtcctttatctctgtctataaaatgtgttgaaaggttaaatgtatatccatcag	4140
attatttgactgagaatgttttcggctgttttctcatatataatcaaaagtatcacatcag	4200
gttcattgaagaacgaggacacactggtgtgtagccaacgacattgtttgacattaatgca	4260
aataataacgtgttacacaacttaagcggtagctttgacagaagtgtaaaaattaatatt	4320
gagacttcacacgcgttaaaacattccaagactttaaaattgcaagaatcaagcttgog	4380
cgcttttttttcacatcgataaggatgtattataatgaaccagctggtaaagtgaaca	4440
ctgttttaaattactatttggttgcggcagaagagttatgacgttgttaaaggacattgtctt	4500
acttaaggattgtctcattcgaaatacttctgatagtgaaatgcgaaacgcacattccatc	4560
aataataaaagtataagatatgagaattgttactgaaatatgtctgttttagtctccatg	4620
attagccaatgattgtcttctgttaagttcatgaaacaccgggtgtttgtgtgtagacat	4680
aaaataaatgtttgattctatagCTGACGGCGGCCGACGTCGCGTGTGGAGCTCTTTG	4740
L T A A G R R V L E L F	
AGCACAGTGGATGACAGACGACGACCAACCTGGCAGCGGGCACCCAGTTCACAGTAC	4800
E H R W M T D E T H N L A A G T Q F T V	
GCGGTTTCCATGGCGACTACGAGGTGCAAGTCATCGTCCAGGGTCAAGAGCACACCAACC	4860
R G F H G D Y E V Q V I V Q G Q E H T N	
TGAGGCAGACGTTCTCGTTGGGCAACGGTCCCCACACCGTCAACATTAACATTAGCTAGA	4920
L R Q T F S L G N G P H T V N I N I S *	
GCGACACTCAGAGGGCA	4937

B CGACGCTTCAGTCAAGCGC**ATG**CCCGGGTCTGCTGCTGCTGATGTGAGCAGCGAGAT 60
M P A G A A A A D V S S E I
CGACAGACTGAGAAGAAGCAACATAGTGGTTCAgtgctgctgctttcaatatataacaa 120
D R L R R S N I V V H
cctcacttagttctgcattgataaaaatctcgtcagtagtgtagtaccattctgtgtgtat 180
caataaagtaaacactgttggtataccatttttaaaattaaagcgaatttaattctgactatt 240
acaatattgtatgtaaatcatatataattcattattttttattaataattttataatttt 300
aacataaatggatagtatattaataacaatcttatttcacggcttatttcaagtttgac 360
gatgttcttttaatacaagaacttaaaatatttttcaaatacagaatgttgctgctataaaa 420
cgtttcagcctgtcattggatgttactactagacaaacacaataacatcttcatccttc 480
gttaaatgtgtaaaatctgacctgaaatccattcacgaccttgctgattttgttttcagC 540
GTGACAGCTGGTGGTAACATCAGCCACGGTGAAGTAAACATACGGgtaagtcatagttga 600
V T A G G N I S H G E V N I R
tatcatcatcgcctgaggacagagtacaacaaactaaagagctactttaccgcaagctc 660
cacctgggtgtcaccctcgatttgccttcgttatggcttagtagttatttgaatattctt 720
tgtcactttgtcactagagacatctgtagcgacaagaccaaattcagcatactcttatg 780
ttgaaacataatcatttcaattctataacatgtgtaaaactgtgtgtgaggcacatttga 840
ccacgtgaacagtttagtagagaagagatacaaaaaataaagatacagtagaagatttgg 900
gggttaataattaattttaataacttattttatatttttaataatattgtataaaaatcacc 960
agactaagtcctttgttcttcaagctaactgtcgttgattacagGTGGTACAAAGAAGAA 1020
V V Q K K K
GTCGTTCCCGTTCGGAACCGCTGTGGCCGCTGGGCCTACAACAATGATTCCAAGACAAA 1080
S F P F G T A V A A W A Y N N D S K T K
ATACCGGGATTTTCATCCACCAGCACTACAACCTGGGCGGTGCCAGAAAACGAACCTCAAGT 1140
Y R D F I H Q H Y N W A V P E N E L K W
GCGCACCATCGAACCTACCAGGgtacactttttattaatatgatttataattttaacata 1200
R T I E P T R
aattgatagcataatgataacaatcttattttaacgttttacttccagtttgcatgatagt 1260
cttttaatacaagaactttaaataatttttcaatgctgaatgttgctgcgcataaaatcgtgtg 1320
cagccgtctttgtatgttactactaaagaataacattaacatacattgaaatgatgttta 1380
aggggtcttgtgatttgcatttgttcaggtaatatagtttaaaagggtacagattaatttt 1440
attaagactatactaaaaactgggtagtttctgctagccaaagaaatgggggttttggggg 1500
catttgataatcgcacgtgtgtcttctgtgaataaaaaataaacaagtaaaaattctgga 1560
gaaaaataacttttgatgcactcgacattgaggcttccgttatgatcgccatttaagtgtt 1620
acctctcctcttgggtcacatcactccacaaatagaggaggaaaaatataacagttgattgt 1680
gggttttcagGGACATAAGAACATATCAGCCTGCCCTGACCATGATTACGGGACTGAAAAGT 1740
G H K N Y Q P A L T M I H G L K S
CACGGgttagtaaacattaagatttagcaaccaccatcaccgacattataatcatcctct 1800
H G
tcgacttcattattgtatgataatgaagaattgggtgaatttttagttataaaacattttat 1860
taaatgaaagagagtttagcgtttcagagattgtctctcccccattgtattttactatttt 1920
tcgtacttcgtttaaagcaagtcgcagactcagttctgacttttttgataattccttaaa 1980
tagtctacaaaatgaaatagtaactgcaaattgtctcacactgatacgaatgaagaaaa 2040
attcgttacactaagagtcagcacatgtcgtgtacagaggtgtttgttagagatgtttgt 2100
tatagttttcttgcattcactcagagatgttttaaaatagGATTAAGGTGAGAGTCAAC 2160
I K V R G H N
CTGGTGTGGTCTGTCAATTCAACGGTGCAGAGCTGGGTCAAAGCTCTGCATGGGGATGAG 2220
L V W S V N S T V Q S W V K A L H G D E
CTGCGAAAGGTGTCCATGACCACATTTGTGGAACCGTCAACACATTTAAGGGCTTgtaa 2280
L R K V V H D H I V E T V N T F K G L
gtttattttcatttattttcttttctaaaaatgtcagtagttttatttgtttatgttttgtattg 2340
ttgctgttatgtggagaacaaaatttactattctacatacaccaaaccagatgataccag 2400
ttacaataactgaattttatgttgtaaaaaatattgtttgaaccttttagcaatagactgcc 2460
ctaataagaagaagtcacgttgtgaagtgtagtttttaactcattcatcaacacatcaag 2520
taagagtggttttctgttttgaccagAGTGGAGCACTGGGACGTGAACAACGAGAACCTGC 2580
V E H W D V N N E N L
ATGGCCAGTGGTACCAGCAACAACCTGAATGACCCGAACATACAACATAGAAGTCTCCGTA 2640
H G Q W Y Q Q Q L N D P N Y N I E L F R
TCGCACACGCTGCCGACCCCAACGTCAAACCTCTCCTCAACGACTACAACGTGGTGGCTT 2700
I A G H A A A D P N V K L F L N D Y N V V A
ACGGTGCCGCAACCAATgtcagtgatgagaaaacttgcttctatatcacctgagaaaca 2760
Y G A A T N
atgtttccaagcatagtatgcttcatttctgcataaaaatatactcttgttgtcttttagca 2820
ttatagtccttctgtgcttctgtatgttcccgaaacatgcaacttgtactgctccatacta 2880
gtctgtagtcctttgttttaactgttaaaacgaatatcatcttcacacagtcctttatgg 2940
gtgtgacagGCCTATCTTCAGCAAGGTCAACAGTTTAAGGCCGCTAATGTGAGTCTTTAC 3000

GGTTTGGGTGCCCAGTGGCCACTTTGGCGATGAAGCTAACCCAAACGTCGCTGGGATGAAG	3060
G L G A Q C H F G D E A N P N V A G M K	
gtaagaaaaataatatacataaacatccttcgccttttctcattttccttctcctattatta	3120
ttattctaaaaaatgaacaaaatgtacgcacaacaatgaagtagttctacaatacagtaa	3180
gataaaagatttttagattttgtctgtatcattttttcttttagttatctgcatttttta	3240
ccattttactttacatggaataacatatctttatggtgagaaatcagaaagtcttcattt	3300
tatagctgtggatttaaaaaatattcttctcacttctctctatactccataagctgtttta	3360
gtgacatctgggttttttctaatactcagCAACATCTGGATATTTAGCTCAAGTGGGGTT	3420
Q H L D I L A Q V G L	
GCCCATCTGGGCCACTGAGTTGGATGTGTTAGCTACGGACGAGAACAACGAGCGGACTT	3480
P I W A T E L D V L A T D E N K R A D F	
CTACGAGCAGCGCTGACAGCCCTGTACAGCCATCATGCCGTGGAGGGCATCTGTATGTG	3540
Y E H A L T A L Y S H H A V E G I L M W	
GGGCTTCTGGGACAAGGCCACTGGCGGCATGAACGAGCTGCTTCTTGTCGAGACAA	3600
G F W D K A H W R H E R A A L L V G D N	
CCTGCAGgtgtgttgggttacaattaaccaaatagggatcttacaggcattatagcacag	3660
L Q	
ccagtatccagacagacgttttgaatcactgtcatgatgagaacacattggattcttct	3720
tctattagaatttctacccagaactatccttgacttcaacatttttttctataagaga	3780
aatccgacaaatgcttactgagtgtacctatggttagtccaaattatgcgggcatgcttt	3840
tttccatttctgaattacttttagcacaataacgaccaaataattaccatctaagtctca	3900
acattagtttttcatattttcatcaacgcttaaacactttaaagacttttaaatgttaag	3960
aatcaagcttttcacagcgataaggattcttttagatgttaaccagctggttaaagtaaaac	4020
agtttaaatctatttggtagaggcaaaaggagttagtggtttaaaggacattgtctta	4080
cataaggaaatgcttcattcacaagtacttctgtagtgatgaaacacacatttcattaat	4140
aataaaaaaataagatatgagaattgttactaaaatatgtctgttagtctccatgatta	4200
ataaaatgtttgcttctgtaaaagttcatgaaacacgggtgtttgtgtagacataaaat	4260
aaacgtttgattgtatagCTGACGGCGCGGACGTCGCGTGTGGAGCTCTATGAGCAC	4320
L T A A G R R V L E L Y E H	
AGGTGGATGACAGACGAGACGCACAACCTGGCAGCGGGCACCCAGTTCACAGTACGCGGT	4380
R W M T D E T H N L A A G T Q F T V R G	
TTCCAGTGGCGACTACGAGGTGCACGTGATCTACCAAGTTCAGGAGCGCACCAACCTGAAG	4440
F H G D Y E V H V I Y Q G Q E R T N L K	
CAGACGTTACGTTGGGCAACGCAGCCACACCGTCAACATCAACATTAGCTAGAGCGAC	4500
Q T F T L G N A A H T V N I N I S *	
ACTCAGAGGGCA	4512

Figure 3. Genomic structures of GHF10-*Pc1* (A) and GHF10-*Pc3* (B). Coding nucleotides and deduced amino acids of each exon are capitalized. Introns are shaded and illustrated with lower letters.

is 1071 bp as compared to that of GHF10-*Pc3*, which is 671 bp.

Expression analysis by reverse transcription-polymerase chain reaction

To examine the expression pattern of GHF10-*Pc1* and GHF10-*Pc3* in *P. canaliculata*, RT-PCR was carried out against the first-stranded cDNA synthesized from total RNAs of different ages of juvenile snails. Results showed that none of GHF10-*Pc1* and GHF10-*Pc3* transcripts was detected at the egg stage whereas the fragment of approximately 1200 bp RT-PCR product was amplified in 1- and 10-day-old snails (Figure 4A). Elongation factor 1 α generated band of 150 bp (Figures 4B and 5B). The undetectable of mRNA transcripts encoded by the GHF10-*Pc1* and GHF10-*Pc3* genes from the eggs suggested that the expression of both genes should start during or after the hatching period.

From sequence analysis, both GHF10-*Pc1* and GHF10-*Pc3* cDNA sequences contain *Rsa* I recognition sites at different location. Therefore, this enzyme can be used to differentiate the family

of GHF10-*Pc1* and GHF10-*Pc3* cDNAs. The *Rsa* I digested GHF10-*Pc1* showed restriction fragments of approximately 519, 227, 196, 124 and 122 bp, respectively whereas *Rsa* I digested GHF10-*Pc3* gave fragments of 395, 354, 196, 124 and 122 bp, respectively (Figure 5A). The result showed that the *Rsa* I digested PCR product of 1- day-old snails revealed restriction fragments that corresponded well to both the GHF10-*Pc1* and GHF10-*Pc3* pattern. In contrast, *Rsa* I digested PCR product of 10-day-old snails corresponded only with the GHF10-*Pc1* pattern. The restriction profile indicated that the GHF10-*Pc1* was expressed in both ages of juvenile *P. canaliculata* whereas GHF10-*Pc3* was observed in 1-day old snails but not in 10-day-old snails. Our result suggested that the GHF10-*Pc3* transcripts should have expressed in the early development of juvenile snail. However, the expression of cellulase GHF10-*Pc1* and GHF10-*Pc3* transcripts was investigated in pooled cDNA of the whole body of 1 and 10-day-old juvenile *P. canaliculata*. Therefore, it is interesting for further studies to examine the cellular type involved in cellulase GHF10-*Pc1* and GHF10-*Pc3* expression using *in situ* hybridisation.

Table II. GC content and length of exons and introns in GHF10-*Pc1* and GHF10-*Pc3* genes.

Exon	Genomic DNA (Number of nucleotides)	GC content (%)	Intron	Genomic DNA (Number of nucleotides)	GC content (%)
GHF10-<i>Pc1</i>					
1	1–93 (93 bp)	53	1	94–545 (452 bp)	34
2	546–591 (46 bp)	36	2	592–874 (283 bp)	35
3	875–1033 (159 bp)	50	3	1034–1688 (655 bp)	32
4	1689–1744 (56 bp)	35	4	1745–2161 (417 bp)	30
5	2162–2299 (138 bp)	48	5	2300–2567 (268 bp)	27
6	2568–2729 (172 bp)	48	6	2740–2973 (234 bp)	35
7	2974–3081 (108 bp)	48	7	3082–3413 (332 bp)	27
8	3414–3632 (219 bp)	56	8	3633–4703 (1071 bp)	33
9	4704–4937 (234 bp)	56			
GHF10-<i>Pc3</i>					
1	1–93 (93 bp)	48	1	94–539 (446 bp)	27
2	540–585 (46 bp)	38	2	586–1003 (418 bp)	33
3	1004–1162 (159 bp)	49	3	1163–1689 (527 bp)	31
4	1690–1745 (56 bp)	31	4	1746–2138 (393 bp)	32
5	2139–2276 (138 bp)	48	5	2277–2545 (269 bp)	28
6	2546–2717 (172 bp)	51	6	2718–2949 (232 bp)	28
7	2950–3060 (111 bp)	46	7	3061–3388 (328 bp)	26
8	3389–3607 (219 bp)	55	8	3608–4278 (671 bp)	32
9	4279–4512 (234 bp)	58			

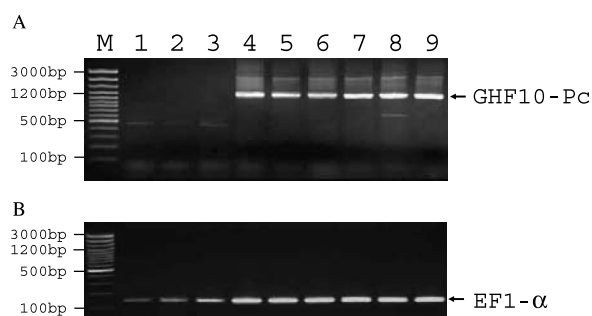


Figure 4. RT-PCR of GHF10-*Pc* (1200 bp; A) and EF1 α (150 bp; B) was examined in the egg (lane 1–3), 1- day-old (lane 4–6), and 10-day-old (lane 7–9) juveniles of *P. canaliculata*. Lanes M are 100-bp DNA marker.

Here we show that in spite of the high similarity of GHF10-*Pc1* and GHF10-*Pc3*, the two cellulase genes revealed different expression patterns. The differential expression of the two isoforms supports the idea that

these forms may play different physiological roles in *P. canaliculata*. However, the functional role of the cellulase gene and its catalytic properties remains to be elucidated.

In summary, we have successfully cloned two full-length cDNAs of cellulase GHF10-*Pc1* and GHF10-*Pc3* belonging to GHF10 from the fresh water golden apple snail, *P. canaliculata*. We also report here the gene structures of two endogenous GHF10-*Pc1* and GHF10-*Pc3* genes. Both genes contained nine exons and eight introns. The endogenous origin of these cellulase genes in *P. canaliculata* was unambiguously verified in this study. The expression analysis of GHF10-*Pc1* and GHF10-*Pc3* during several ages of *P. canaliculata* was analysed by RT-PCR. The result showed that GHF10-*Pc1* and GHF10-*Pc3* transcripts were developmentally expressed.

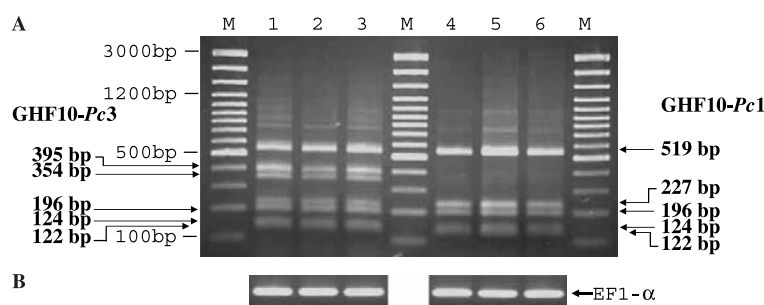


Figure 5. Restriction patterns of the *Rsa* I-digested RT-PCR products of GHF10-*Pc1* and GHF10-*Pc3* (A) and EF1 α (B). The 1-day old juveniles (lane 1–3) of *P. canaliculata* revealed restriction patterns corresponding to both the GHF10-*Pc1* (519, 227, 196, 124 and 122 bp) and GHF10-*Pc3* (395, 354, 196, 124 and 122 bp) whereas the 10-day old juveniles (lane 4–6) corresponded only with the GHF10-*Pc1* pattern. Lanes M are 100-bp DNA marker.

Acknowledgements

This work was supported by grant from the Thailand Research Fund (TRF), grant number MRG4980168. It was also supported by Ratchadaphiseksomphot Endowment Fund, Chulalongkorn University, awarded to Chanprapa Imjongjirak.

References

- Amparyup P, Jitvaropas R, Pulsook N, Tassanakajon A. 2007. Molecular cloning, characterization and expression of a masquerade-like serine proteinase homologue from black tiger shrimp *Penaeus monodon*. Fish Shellfish Immunol 22:535–546.
- Beguín P, Aubert JP. 1994. The biological degradation of cellulose. FEMS Microbiol Rev 13:25–58.
- Brummell DA, Lashbrook CC, Bennett AB. 1994. Plant endo- β -D-glucanases. ACS Symp Ser 566:100–129.
- Byrne KA, Lehnert SA, Johnson SE, Moore SS. 1999. Isolation of a cDNA encoding a putative cellulase in the red claw crayfish *Cherax quadricarinatus*. Gene 239:317–324.
- Gilkes NR, Henrissat B, Kilburn DG, Miller RC, Warren RAJ. 1991. Domains in microbial β -1,4-glycanases: Sequence conservation, function, and enzyme families. Microbiol Rev 55:303–315.
- Henrissat BA. 1991. Classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 280:309–316.
- Imjongjirak C, Klinbunga S, Sittipraneed S. 2005. Cloning, expression and genomic organization of genes encoding major royal jelly protein 1 and 2 of the Honey 20 Bee (*Apis cerana*). J Biochem Mol Biol 38:49–57.
- Imjongjirak C, Amparyup P, Tassanakajon A, Sittipraneed S. 2007. Antilipopolysaccharide factor (ALF) of mud crab *Scylla paramamosain*: Molecular cloning, genomic organization and the antimicrobial activity of its synthetic LPS binding domain. Mol Immunol 44:3195–3203.
- Itakura S, Masuta T, Tanaka H, Enoki A. 2006. Identification of two subterranean termite species (Isoptera: Rhinotermitidae) using cellulase genes. J Econ Entomol 99:123–128.
- Keawjam RS, Upatham ES. 1990. Shell morphology, reproductive anatomy and genetic patterns of three species of apple snails of the genus *Pomacea* in Thailand. J Med Appl Malacol 2:45–57.
- Moriya S, Ohkuma M, Kudo T. 1998. Phylogenetic position of symbiotic protist *Dinemympha exilis* in the hindgut of the termite *Reticulitermes speratus* inferred from the protein phylogeny of elongation factor α . Gene 210:221–227.
- Nathan L, Watanabe H, Sugimura M. 2003. Evidence for the presence of a cellulase gene in the last common ancestor of bilaterian animals. Proc R Soc Lond B (Suppl.) 270:S69–S72.
- Smant G, Stokkermans JPWG, Yan Y, De Boer JM, Baum TJ, Wang X, Hussey RS, Gommers FJ, Henrissat B, Davis EL, Helder J, Schots A, Bakker J. 1998. Endogenous cellulases in animals: Isolation of β -1,4-glucanase genes from two species of plant-parasitic cyst nematodes. Proc Natl Acad Sci USA 95:4906–4911.
- Suzuki M, Watanabe Y, Tanaka A, Ohnishi T. 1986. Bull Tokai Reg Fish Res Lab 120:53.
- Suzuki K, Ojima T, Nishita K. 2003. Purification and cDNA cloning of a cellulase from abalone *Haliotis discus hannai*. Eur J Biochem 270:771–778.
- Thaewnon-ngiw B, Klinbunga S, Phanwichien K, Sangduen N, Lauhachinda N, Menasveta P. 2004. Genetic diversity and molecular markers in introduced and Thai native apple snails (*Pomacea* and *Pila*). J Biochem Mol Biol 37:493–502.
- Tokuda G, Lo N, Watanabe H, Slaytor M, Matsumoto T, Noda H. 1999. Metazoan cellulase genes form termites: Intron/exon structure and sites of expression. Biochem Biophys Acta 1447:146–159.
- Tomme P, Warren RAJ, Gilkes NR. 1995. Cellulose hydrolysis by bacteria and fungi. Adv Microbiol Physiol 37:1–81.
- Wang H, Jones RW. 1995a. Cloning, characterization and functional expression of an endoglucanase-encoding gene from the phytopathogenic fungus *Macrophomina phaseolina*. Gene 158:125–128.
- Wang H, Jones RW. 1995b. A unique endoglucanase-encoding gene cloned from the phytopathogenic fungus *Macrophomina phaseolina*. Appl Environ Microbiol 61:2004–2006.
- Wang J, Ding M, Li YH, Chen QX, Xu GJ, Zhao FK. 2003. Isolation of a multi-functional endogenous cellulase gene from mollusc, *Ampullaria crosseana*. Acta Biochim Biophys Sin 35:941–946.
- Watanabe H, Nakamura M, Tokuda G, Yamaoka I, Scrivener AM, Noda H. 1997. Site of secretion and properties of endogenous endo- β -1,4-glucanase components from *Reticulitermes speratus* (Kolbe), a Japanese subterranean termite. Insect Biochem Mol Biol 27:305–313.
- Xue XM, Anderson AJ, Richardson NA, Anderson AJ, Xue GP, Mather PB. 1999. Characterisation of cellulase activity in the digestive system of the redclaw crayfish (*Cherax quadricarinatus*). Aquaculture 180:373–386.
- Xu B, Ersson B, Hellman U, Janson JC. 2000. Purification, characterization and amino acid sequence analysis of a thermostable, low molecular mass endo- β -1,4-glucanase from blue mussel, *Mytilus edulis*. Eur J Biochem 267:4970–4977.
- Xu B, Janson JC, Sellos D. 2001. Cloning and sequencing of a molluscan endo- β -1,4-glucanase gene from the blue mussel, *Mytilus edulis*. Eur J Biochem 268:3718–3727.
- Yan Y, Smant G, Stokkermans J, Qin L, Helder J, Baum T, Schots A, Davis E. 1998. Genomic organization of four β -1,4-endoglucanase genes in plant-parasitic cyst nematodes and its evolutionary implications. Gene 220:61–70.