การโคลนและลักษณะสมบัติของยีนและโปรตีนที่เกี่ยวข้องกับเมแทบอลิซึมของแคลเซียม

ของกุ้งกุลาดำ *Penaeus monodon*

นาย อภิรักษ์ วัฒนสุโรโรจน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาเทคโนโลยีชีวภาพ

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2552

CLONING AND CHARACTERIZATION OF GENES AND PROTEINS
RELATED TO CALCIUM METABOLISM OF BLACK TIGER SHRIMP
Penaeus monodon

Mister Apiruck Watthanasurorot

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Biotechnology

Faculty of Science

Chulalongkorn University

Academic Year 2008

|	
Thesis Title	Cloning and Characterization of Genes and Proteins Related to Calcium Metabolism of Black Tiger Shrimp *Penaeus monodon*
By	Mister Apiruck Watthanasurorot
Field of Study	Biotechnology
Advisor	Professor Piamsak Menasveta, Ph.D.
Co-Advisor	Virak Visudtiphole, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

.... Dean of the Faculty of Science

(Professor Supot Hannongbua, Ph.D.)

THESIS COMMITTEE

.... Chairman

(Associate Professor Charoen Nitithamyong, Ph.D.)

.... Advisor

(Professor Piamsak Menasveta, Ph.D.)

.... Co-Advisor

(Virak Visudtiphole, Ph.D.)

.... Examiner

(Sirawut Klinbunga, Ph.D.)

.... External Examiner

(Sittiruk Roytrakul, Ph.D.)

นาย อภิรักษ์ วัฒนสุโรโรจน์ : การโคลนและลักษณะสมบัติของยีนและโปรตีนที่เกี่ยวข้องกับเมแทบอลิ
ซึมของแคลเซียมของกุ้งกุลาดำ *Penaeus monodon.* (Cloning and Characterization of Genes and
Proteins Related to Calcium Metabolism of Black Tiger Shrimp *Penaeus monodon*) อ.ที่ปรึกษา
วิทยานิพนธ์หลัก : อ. ที่ปรึกษา : ศ.ดร. เปี่ยมศักดิ์ เมนะเศวต, อ. ที่ปรึกษาวิทยานิพนธ์ร่วม : ดร.วิรักษ์
วิสุทธิผล 211 หน้า.

แคลเซียมไอออน (Ca^{2+}) มีบทบาทสำคัญต่อกระบวนการตอบสนองในระดับเซลล์หลายชนิดของสิ่งมีชีวิตกลุ่มยูคาริโอต
ซึ่งอวัยวะของเซลล์ (organelle) ที่ทำหน้าที่สำคัญในการควบคุมการเปลี่ยนแปลงความเข้มข้นของ Ca^{2+} ภายในเซลล์ คือ
Endoplasmic reticulum (ER) การศึกษายีน หรือ โปรตีน ที่มีบทบาทในการควบคุมความสมดุลของ Ca^{2+} เป็นอีกหนทางหนึ่งที่
นำไปสู่การเข้าใจการตอบสนองต่างๆ ที่เกี่ยวข้องกับ Ca^{2+} ในการศึกษาครั้งนี้ได้เน้นโปรตีน 3 ชนิด ที่ทำหน้าที่ร่วมกันใน
กระบวนการควบคุมสมดุล Ca^{2+} และควบคุมการพับม้วนของโปรตีนใน ER ซึ่งประกอบด้วย Calreticulin (CRT), Calnexin (CNX)
และ Endoplasmic Reticulum protein 57 (ERp57)

จากการหาลำดับนิวคลีโอไทด์ที่สมบูรณ์ของยีนที่สนใจโดยเทคนิค RACE PCR พบว่า *PmCRT* และ *PmCNX* มีนิวคลีโอ
ไทด์จำนวน 1221 และ 1778 คู่เบส ซึ่งแปลรหัสได้ 406 และ 595 อะมิโน ตามลำดับ ขณะที่ PmERp57 มี 2 isoform โดยทั้ง 2 แบบ มี
ขนาดของ ORF เท่ากัน คือ 1458 นิวคลีโอไทด์ (458 อะมิโน) แต่ส่วนที่แตกต่างกัน คือ 3'UTR ที่มีขนาด 529 และ 713 นิวคลีโอ
ไทด์ นอกจากนี้ จากการศึกษาการจัดเรียงตัวของยีนภายในจีโนมิคดีเอ็นเอของแต่ละยีน พบว่า ยีน *PmCRT* มี 4 exons และ 3
introns, และ *PmERp57* มี 10 exons กับ 9 introns ขณะที่ *PmCNX* ยังไม่เสร็จสมบูรณ์ เมื่อตรวจสอบการแสดงออกของยีนทั้ง 3 ยีน
ในเนื้อเยื่อต่างๆ คือ ก้านตา, ขาว่ายน้ำ, เหงือก, หัวใจ, รังไข่, อัณฑะ, กระเพาะอาหาร, ลำไส้, เลือด, ปมประสาทส่วนนอก, ตับอ่อน,
Lympiod organ และ ต่อมแอนเทนนา ทั้งในพ่อแม่พันธุ์กุ้งกุลาดำ และกุ้งกุลาดำขนาดอายุ 4 เดือน ผลการทดลองพบว่า ทุกยีนมีการ
แสดงออกในทุกเนื้อเยื่อที่ทำการศึกษา

นอกจากนี้ยังการตรวจสอบการแสดงออกของยีนทั้ง 3 ยีน ด้วยวิธี Semiquantitative RT-PCR และ quantitative RT-PCR
ในกลุ่มตัวอย่างที่ทำให้เกิดความเครียดโดยการเพิ่มอุณหภูมิน้ำเป็น 35°C สำหรับอวัยวะที่ถูกเก็บในการทดลองเบื้องต้นประกอบด้วย
เลือด ตับอ่อน และ เหงือก ซึ่งจากผลการทดลองพบว่า ยีนทั้งหมดที่สนใจมีการเปลี่ยนแปลงการแสดงออกที่ระดับ mRNA อย่างมี
นัยสำคัญเฉพาะในเลือดภายใต้สภาพเครียด (P<0.05) ขณะที่ในอีกสองอวัยวะที่เหลือทุกยีนยังคงมีระดับการแสดงออกเท่ากับชุด
ควบคุม (P>0.05) จากนั้นทำการยืนยันผลการแสดงออกของยีนทั้งหมดในเลือดด้วยเทคนิค real time PCR ซึ่งปรากฏว่า ผลการ
ทดลองในทั้ง 2 วิธี มีความสอดคล้องกัน คือ ภายในเลือด ยีน *PmCRT, PmCNX* และ *PmERp57* มีระดับการแสดงออกเพิ่มขึ้นเมื่อกุ้ง
ถูกทำให้เครียดด้วยการเพิ่มอุณหภูมิ (P<0.05)

โปรตีนลูกผสมของ PmCRT, PmCNX และ PmERp57 สามารถแสดงออกได้ในแบคทีเรีย โดย 2 โปรตีนแรกจะอาศัย
pGEX 4T-1 expression vector ขณะที่ ERp57 ใช้ pET 15b โดยโปรตีนลูกผสมทั้งหมดเป็นโปรตีนที่อยู่ในส่วน soluble จากนั้นนำ
โปรตีนลูกผสมดังกล่าวไปใช้ในการศึกษาหน้าที่ และ ผลิต polyclonal antibody สำหรับการศึกษาบทบาทของโปรตีน ใน *in vitro*
ได้แก่ ความสามารถในการจับกับ Ca^{2+} และการจับกับ ERp57 ของ CRT/CNX ผลการทดสอบการจับกับ Ca^{2+} พบว่ามีเฉพาะ
PMCNX เท่านั้น ที่มีความสามารถดังกล่าว ขณะที่ความสามารถในการจับกับ ERp57 ของ PmCRT และ PmCNX พบว่า โปรตีนทั้ง
สองชนิด สามารถจับกับ PmERp57 ได้โดยเกิด mobility shift band บน Native PAGE

สาขาวิชา..........เทคโนโลยีชีวภาพ..... ลายมือชื่อนิสิต ....

ปีการศึกษา.............2551.................. ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก...........................

ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม...........................

4972567323 : MAJOR BIOTECHNOLOGY
KEYWORDS : *Penaeus monodon* / GIANT TIGER SHRIMP / CALRETICULIN / CALNEXIN / ERP57

APIRUCK WATTHANASUROROT : CLONING AND CHARACTERIZATION OF GENES AND PROTEINS RELATED TO CALCIUM METABOLISM OF THE BLACK TIGER SHRIMP *Penaeus monodon* ADVISOR : PROF. PIAMSAK MENASVETA, Ph.D, CO-ADVISOR : VIRAK VISUDTIPHOLE, Ph.D, 211 pp.

Calcium ion (Ca^{2+}) plays critical role in many cellular responses of eukaryotes. The important Ca^{2+} controlling organelle (Ca^{2+} homeostasis) in cellular level is the Endoplasmic reticulum (ER). Therefore, studying genes that control Ca^{2+} homeostasis could lead to understand cellular responses related to Ca^{2+}. The focus of this study was on calreticulin (CRT), Calnexin (CNX) and Endoplasmic Reticulum protein 57 (ERp57). These genes/proteins function together to control cellular Ca^{2+} homeostasis and protein folding in the ER.

The full length cDNA of these genes were successfully characterized. A single form of *PmCRT* and *PmCNX* contained open reading frame (ORF) of 1221 and 1788 bp corresponding to deduce proteins of 406 and 595 amino acid residues, respectively. In contrast, two isoforms of *PmERp57* was found. They contained an identical ORF of 1458 bp (corresponding to 485 amino acid residues) but two different 3' UTR length of 529 and 713 bp. Furthermore, genomic organization of the three genes was characterized. The genomic *PMCRT* sequence contained 4 exons and 3 introns while that of *PmERp57* consisted of 10 exons and 9 introns. Genomic organization of *PmCNX* was partially characterized. Only 4 exons and 3 introns were completed. Tissue distribution analysis of the three gene homologues revealed their expression in eye stalk, pleopod, gill, heart, ovaries, testis, hepatopancreas, stomach, intestine, hemocytes, thoracic ganglia, lympiod organ and antennal gland of *P. monodon* juveniles and broodstocks.

Semi-quantitative RT-PCR were carried out to examine effect of thermal stress on expression levels of *CRT*, *CNX* and *ERp57* in hemocytes, hepatopancreas and gill of the juvenile black tiger shrimps. Differential expression was found only in hemocytes. The semi-quantitative result from hemocytes was confirmed by quantitative real-time PCR. Expression of the three genes in hemocytes was up-regulated within the first hour after the 30 °C heat treatment ($P < 0.05$). After that, their expression levels return to the normal levels.

Recombinant proteins of PmCRT, PmCNX and PmERp57 were produced using an *E. coli* system. All recombinant proteins were expressed in the soluble form. These recombinant proteins were purified and some protein functions were *in vitro* assayed. Binding of PMCNX to Ca^{2+} caused change in its conformation, resulting in a mobility shift on native PAGE gel while the others did not exhibit the shift. Moreover, PmCRT and PmCNX also complexed with PmERp57, causing mobility shift of the proteins bands on native PAGE gel.

Field of Study :.........Biotechnology.......   Student's Signature Apiruck Watthanasurorot

Academic Year :.........2008..............   Advisor's Signature

Co-Advisor's Signature

ACKNOWLEDGMENTS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

LIST OF ABBREVIATIONS

bp	base pair
°C	degree celcius
DEPC	diethylpyrocarbonate
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanosine triphosphate
dTTP	deoxythymidine triphosphate
DNA	deoxyribonucleic acid
HCl	hydrochloric acid
IPTG	isopropyl-thiogalactoside
Kb	kilobase
M	molar
$MgCl_2$	magnesium chloride
mg	mlligram
ml	mllilitre
mM	mllimolar
ng	nanogram
OD	optical density
PCR	polymerase chain reaction

RNA	ribonucleic acid
RNase A	ribonuclease A
rpm	revolution per minute
RT	reverse transcription
SDS	sodium dodecyl sulfate
Tris	tris (hydroxyl methyl) aminomethane
μg	microgram
μl	microlitre
μM	micromolar
UV	ultraviolet

CHAPTER I

INTRODUCTION

1.1 General introduction

Shrimps are decapod crustaceans of high economic value. In total, 343 species of economically important shrimps are reported by the FAO (Asian Shrimp Culture Council, 1996). Most wild-caught shrimps and all farm-raised shrimps belong to the Penaeidae family (Tassanakajon *et al.*, 2006). Within this family, the black or giant tiger shrimp (*Penaeus monodon)* is one of the most economically important species (Rosenberry, 2003).

.

In Thailand, farming of *P. monodon* started in 1970s using locally available broodstock captured from the sea to produce postlarvae in land-based hatcheries for pond stocking. By the early 1990s, Thailand emerged as the world's leading farmed shrimp producer and exporter based on *P. monodon* production. The country is endowed with several natural factors that facilitate the farming e.g. appropriate farming area without any serious disturb from storms, ideal soils for pond construction and small variation of seawater salinity between seasons. These factors thus caused the *P. monodon* farming to develop into a well-organized, fully-integrated industry.

The world shrimp aquaculture production, which had stabilized in the 1990s, has shown strong increase in subsequent years (Figure 1.1). Several Thai shrimp companies have well-established marketing companies in the major import markets in the U.S., Japan and Europe. Annual production of farmed *P. monodon* in Thailand alone had equaled or exceeded 200,000 metric tons since 1993 but the production was dropped owing to several relentless outbreaks of diseases and the lack of high quality wild and/or domesticated broodstock. These problems has eventually lead the farming to decline and a replacement of the monodon shrimp with Pacific white shrimp *Litopenaeus vannamei* (Figure 1.2).

Nevertheless, because of the less market competition and better market price, farming of *P. monodon* currently has been promoted to make it become intensive again. The shrimp itself is a local species, which can inhabit only in a limited region around Southeast Asia. Its farming thus is more sustainable in term of economics to the country. In addition, its unique flavor and color make the black tiger shrimp more premium and flavorable by the customers, compared with the white shrimp. As a result, to promote farming of *P. monodon* back again, researches are required to solve the farming problems ranging from the disease, reproduction and growth aspects.
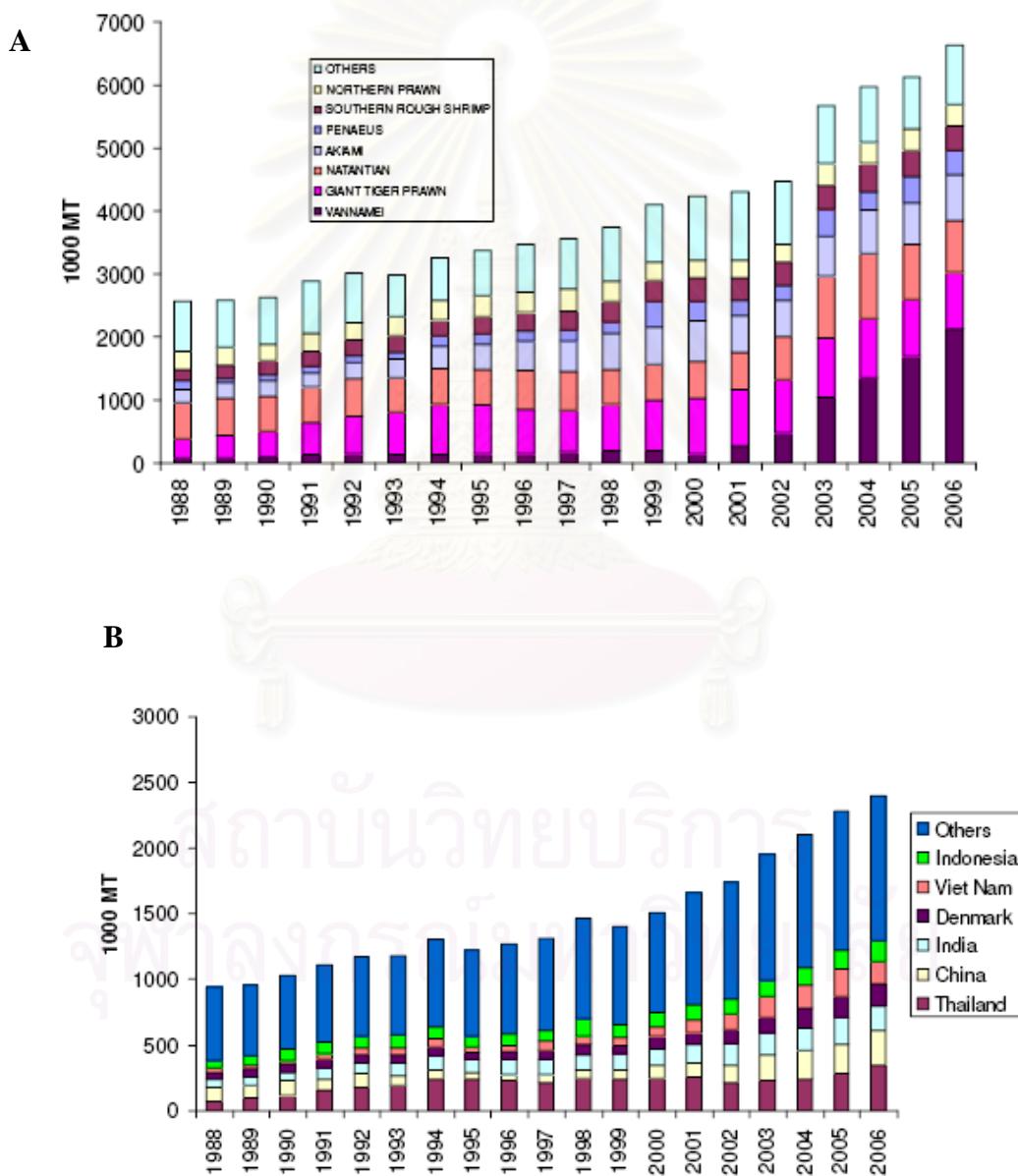
Figure 1.1 Shrimp aquaculture production from 1988-2006: classified according to types of shrimp (A) and to the exporter countries (B). (www.globefish.org).
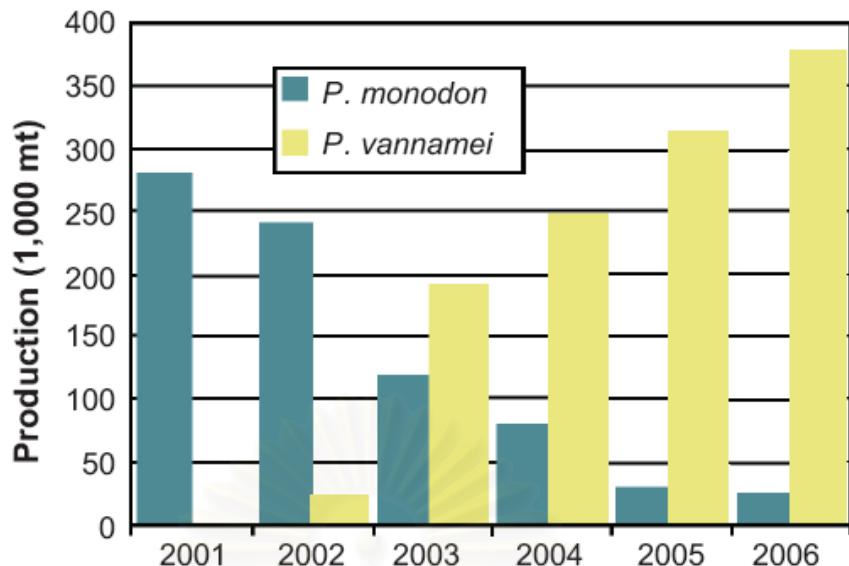
Figure 1.2 Annual shrimp farming production in Thailand. (Source: Global Aquaculture Advocate)

1.2 Taxonomy of *Penaeus monodon*

The scientific name of black tiger shrimp (giant tiger shrimp) is *Penaeus monodon* is classified as a member of Phylum Arthropoda; Subphylum Crustacea; Class Malacostraca; Subclass Eumalacostraca; Order Decapoda; Suborder Natantia; Infraorder Penaeidea; Superfamily Penaeoidea; Family Penaeidae Rafinesque, 1985; Genus *Penaeus* Fabricius, 1798 and subgenus *Penaeus* (Bailey-Brook and Moss, 1992).

1.3 Calcium homeostasis and its effect to physiology of crustaceans

Calcium ions play a critical role in many cellular biological processes. It is a cellular promoting ion that leads to activation of several cellular functions such as transcription, cell cycle, cytoplasm and mitochondrial metabolism, protein/glycoprotein synthesis and folding. Most effects are mediated by intracellular calcium oscillatory through changes of the calcium concentration, causing cells to respond. So, calcium homeostasis is necessary for biological functions of the cell (Bjorkman and Cleland, 1991; Michalak *et al.*, 2002).

1.3.1 Calcium storage and signaling: The endoplasmic reticulum

The endoplasmic reticulum (ER) is an eukaryotic cellular organelle, which stores intracellular Ca^{2+}. Ca^{2+} is required for the ER to function properly. Ca^{2+} is released from the ER to transmit cellular signals between the periphery and the central or local area of the cell (Blaustein and Golovina, 2001).

The system of endomembranes (ER in nonmuscular cells, SR in muscle cells) is potentially the most important intra cellular Ca^{2+} store because of Ca^{2+} capacity and the rapidity of exchange with the cytosol (Alvarez et al., 1999). Ca^{2+} is transported across the ER/SR endomembranes via an energy-dependent Ca^{2+} ATPase (sarcoplasmic or endoplasmic reticulum Ca^{2+} ATPase, SERCA). Crustacean SERCA has a molecular mass of approximately 100 kDa (Ushio and Watabe, 1993) and SERCA genes have been cloned and sequenced in the brine shrimp and crayfish. Rates of calcium transport by these transport systems in crustaceans have been reported to be considerably faster than the comparable SERCA proteins in vertebrates (Deamer, 1973). The ER membrane of both vertebrates and invertebrates also possess ryanodine (RyR) and inositol 1,4,5-triphosphate receptor (IP$_3$R) calcium channels which function to release stored calcium from the ER to the cytoplasm (McCormack and Cobbold, 1991).

For the ER to maintain its proper structure and function, Ca^{2+} concentration within the lumen has to be retained at optimum. Excessive depletion of Ca^{2+} in the ER results in ER fragmentation (Terasaki *et al.*, 1996) while extremely high Ca^{2+} concentration causes the ER membrane to lose its continuity. In addition, as a site for folding of newly-synthesized proteins, the ER contains several protein chaperones, some of which requires binding to Ca^{2+} for their function. Therefore, Ca^{2+} is important for the ER function in protein synthesis and folding. Also, there is an evidence showing that excessive depletion of Ca^{2+} in the ER increases protein degradation (Subramanian and Meyer, 1997; Putney and Ribeiro, 2000; Ribeiro *et al.*, 2000). Moreover, Ca^{2+} is important in glycoprotein trafficking between the ER and Golgi complex during protein modification. Ca^{2+} is involved in formation of the vesicles used to transport the pre-modified glycoproteins to the Golgi complex.

1.3.2 Importance of calcium ions in crustaceans

That most crustaceans inhabit in an overlap between the territorial and aquatic environments causes ion homeosis to become critical in maintaining the balance between their body and the environment. In addition, their behavior of the habitat migration ranging from fresh water, ocean, mangrove and territorial area forces them to continuously adapt their physiological system into the different environments (Ahearn *et al.*, 2004). This adaptation process occurring in the organism body is called homeostasis.

One of the factors which have an immediate effect to the physiological system of crustaceans is the ambient concentration of mineral ions, including Ca^{2+}. At a certain period of their life cycle, crustaceans require to shade their shell or exoskeleton to increase their size. This process is called 'molting' and associated with Ca^{2+} storage and depletion. During pre-molt, calcium stored in the shell is solubilized and transferred to another tissue for a temporary storage, before relocated to deposit in the new exoskeleton. After that, in the post-molt period Ca^{2+} is pumped in the form of $CaCO_3$ from the environment into the body to create the new shell (Ziegler *et al.*, 2002).

Synthesized in epithelioid glands (Y-organ), ecdysteroid is a hormone controlling the motling cycle in crustaceans. Synthesis of the hormone is regulated through a complex cascade of cellular signaling, involving molt-inhibiting peptide hormone (MIH) in X-organ, Ca^{2+}, cyclic guanosine monophosphate (cGMP), cAMP and protein kinase C (PKC). Ca^{2+} is engaged in expression of these proteins. In addition, Ca^{2+} was found to control the degradation of cAMP, which has an inhibition effect on the ecdysteriod production. Namely, increase of Ca^{2+} efflux inversely affects the degradation rate of cAMP (Figure 1.3).

In addition to growth, reproduction system also involves in Ca^{2+} signaling. In *M. japonicus*, Ca^{2+} and PKC are involved in the signaling pathway of vitellogenin inhibiting hormone (VIH), a neuropeptide secreted from the X organ in the eyestalks. VIH has an inhibiting effect on synthesis of vitellogenin, which is a precursor of the egg yolk protein (Okumura, 2006).
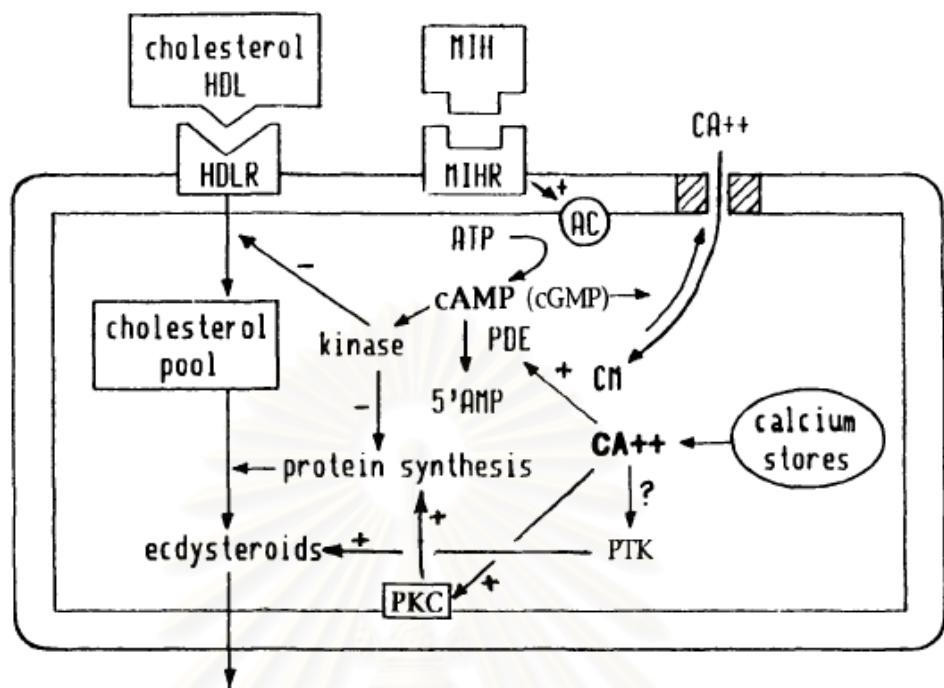
Figure 1.3 A schematic diagram displaying second messenger regulation of ecdysteriod hormone synthesis in the Y-organ. Summary for Description. Cyclic guanosine monophosphate : cGMP; protein kinase C : PKC; protein kinase C : PKC; molt-inhibiting peptide hormone : MIH; molt-inhibiting peptide hormone receptor : MIHR. (modified from Nakatsuji *et al.*, 2008)

1.4 Proteins related to calcium homeostasis

1.4.1 Calreticulin

Calreticulin (CRT) is a multi-functional protein found in the endo-/sarco-plasmic reticulum (ER/SR) of all eukaryotes, except for yeasts (Mazzarella *et al.*, 1992; Smith, 1992; Khalife *et al.*, 1993; Kwiatkowski and Bate, 1995; Jaworski *et al.*, 1996; Michalak *et al.*, 1996; Labriola *et al.*, 1999). It is considered to be the main Ca^{2+}-binding protein in ER, and among species it shares a high level of homology (Eggleton *et al.*, 1997; Nakhasi *et al.*, 1998; Michalak *et al.*, 1999).

1.4.1.1 The *CRT* gene

The *CRT* gene of many organisms studied so far contains various numbers and sizes of exons and introns (McCauliffe *et al.*, 1992; Rooke *et al.*, 1997; Xu *et al.*, 2005). The promoter of *CRT* is activated by a variety of chemical and biological stresses (Plakidou-Dymock and McGivan, 1994; Conway *et al.*, 1995; Zhu, 1996; Waser *et al.*, 1997). The proximal promoter region of the *CRT* gene contains several putative regulatory sites, including GC-rich areas, AP-1 and -2, a Sp1 site, a H4TF-1 site, a YY1, CCAAT sequence, and 2 Endroplasmic Reticulum Stress Elements (ERSE) (McCauliffe *et al.*, 1992; Waser *et al.*, 1997; Kales *et al.*, 2007). The H4TF-1 and AP-2 recognition sequences are commonly found in genes that are active during the cell proliferation while YY1 or a zinc finger transcription factor controls transcription of a large number of cellular and developmental genes (Shrivastava and Calame, 1994). ERSE (CCAATXXXXXXXXXCCACG), is necessary for induction of CRT and other luminal ER gene expression (Yoshida *et al.*, 1998; Kokame *et al.*, 2001).

1.4.1.2 The CRT protein and putative function of its domains

CRT is an evolutionarily conserved protein (Michalak *et al.*, 1992; Shago *et al.*, 1997), consisting of three major internal domains, N-, P- and C-domains aligning from N-terminus to C-terminus of protein, respectively (Figure 1.4). Half of the N-domain is globular and contains eight anti-parallel β-sheets connected by helix-turn-helix loops (Figure 1.5). The amino acid sequence of this domain is extremely conserved (Michalak *et al.*, 1999). This N-domain could interact with many biological and chemical molecules/ions e.g. DNA-binding domain of steroid receptors (Burns *et al.*, 1994), interaction α-integrin (Rojiani *et al.*, 1991), rubella virus RNA binding (Nakhasi *et al.*, 1994), protein disulphide-isomerase (PDI) and ER protein 57 (ERp57) (Baksh *et al.*, 1995a; Corbett *et al.*, 1999). In spite of inability to bind Ca^{2+}, this domain was shown to bind Zn^{2+}(Baksh *et al.*, 1995b) (Baksh and Michalak, 1991; Burns *et al.*, 1997).

The P-domain (or arm domain) is proline-rich and contains a motif of PxxIxDPDAx, three repeats of a KPEDWDE region and a GxWxPPxIxNPxYx motif,

lining up from the N- to C- termini. All of these motifs play a critical rule for the lectin-like chaperone activity of calreticulin (Vassilakos *et al.*, 1998). The domain has a high affinity of Ca^{2+} binding (Baksh and Michalak, 1991). Interestingly, amino acid sequence of the P-domain of CRT is highly homologous to those of calmergin, calnexin and CALNUC, and all of them can interact with PDIs such as ERp57 (Baksh *et al.*, 1995a; Leach *et al.*, 2002).
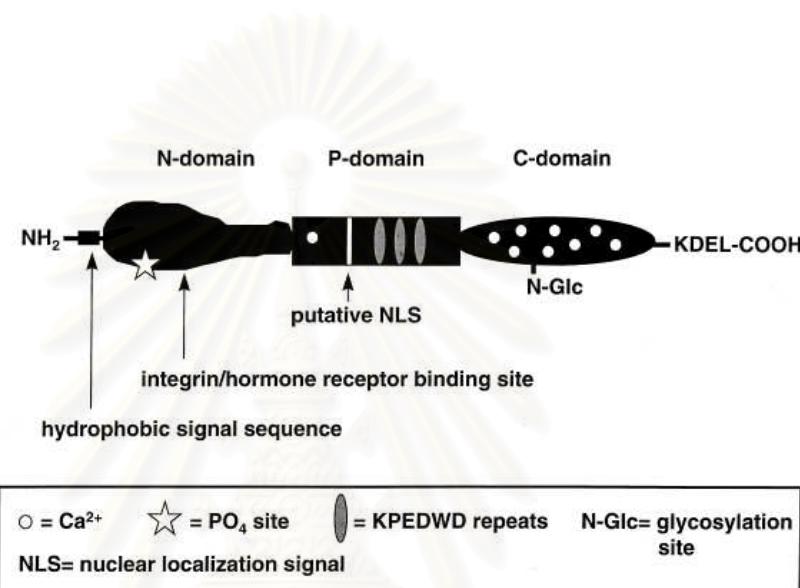
Figure 1.4 A schematic diagram displaying functional domains of calreticulin: It consists of 3 functional domains i.e. globular N-terminal domain, Proline-rich region (P-domain) and negatively–charged C-terminal domain. (Source: Coppolino and Dedhar, 1998)

The C-terminal domain is highly negatively-charged and acidic. It also contains an ER retrieval sequence (K/HDEL), a motif that helps the protein to be retained in the ER lumen. In CRT from human, rat and dog, region of the last 57 C-terminal residues is composed of 37 aspartic or glutamic acids (Fliegel *et al.*, 1989a; Smith and Koch, 1989). The C-compartment forms coiled coil structure (Waser *et al.*, 1997). This domain contains a low affinity but high capacity Ca^{2+} binding site. Binding of Ca^{2+} to this domain plays a regulatory role in its interaction with PDI (Corbett *et al.*, 1999).

1.4.1.3 The location of mRNA and protein of CRT

The main location of calreticulin is the lumen of ER/SR. Addtionally, it could be found in cytosolic granules, in perinuclear areas, (Opas *et al.*, 1991; Krause and Michalak, 1997; Michalak *et al.*, 1999; Johnson *et al.*, 2001), sperm acrosome (Nakamura *et al.*, 1993), attached to plasma membrane of several cell lines (Gray *et al.*, 1995; White *et al.*, 1995; Arosa *et al.*, 1999; Basu and Srivastava, 1999), granules of the cytotoxic T-cell (Dupuis *et al.*, 1993; Andrin *et al.*, 1998; Fraser *et al.*, 1998) and salivary glands (Jaworski *et al.*, 1996). Calreticulin is widely expressed almost in all tissues. For example, in *Eisenia fetidaearth*, the *CRT* transcript could be found in various tissues such as sperm cells, mesenchymal lining, ventral nerves cord, the body wall muscles, cerebral ganglia, pharynx and seminal vesicle. In Chinese shrimp *Fenneropenaeus chinensis*, the transcript is expressed in all the tissues studied including hepatopancreas, gill, intestine, haemocyte and ovary. The amounts are comparable in all tissues, except for ovary whose expression is the highest (Luana *et al.*, 2007). Furthermore, its expression at the protein level has been studied in bovine. The protein is present in all the studied tissues except for erythrocytes (Khanna and Waisman, 1986). The amounts were high in pancreas and liver but moderate in kindney, spleen, parathyroid and adrenals while in cerebral cortex and muscle tissue the amounts were relatively low.

1.4.1.4 Crustacean CRT studying

The first full-length cDNA of *CRT* in crustacean was identified in Chinese shrimp *F. chinensis*. Its ORF (*FcCRT*) contains 1221 bp, encoding 406 amino acids. The *FcCRT* transcript was detected in all the studied tissues. In addition, expression of its transcript was significantly increased during the inter-molt stage, compared with pre- and post-molt ($P < 0.05$). The expression is also upregulated by infection of white spot syndrome virus (WSSV). or by heat shocking (Luana *et al.*, 2007).

1.4.1.5 CRT functions

Calreticulin has two major functions. First, it performs as a lectin-like chaperone to promote folding and also oligomeric assembly of glycoproteins. Second, it is a Ca^{2+} modulator, regulating cellular Ca^{2+} homeostasis. The Ca^{2+} modulator function is attributed to two Ca^{2+} binding sites within the molecule. One is in the P-domain having a high affinity but low capacity to bind to Ca^{2+} ($K_d = 1$ μM; $B_{max} = 1$ mol of Ca^{2+}/mol of protein), and the other is in the C-domain, having a high capacity but low affinity of Ca^{2+} binding ($K_d = 2$ μM; $B_{max} = 25$ mol of Ca^{2+}/mol of protein) (Baksh and Michalak, 1991). Ca^{2+} homeostasis activity of calreticulin is controlled by Calreticulin- Sarco(endo)plasmic reticulum Ca^{2+} ATPase 2b (SERCA2b), which contains transmembrane and luminal ER segments. Binding of SERCA2b to the P-domain of CRT results in an attenuation of the Ca^{2+} transport activity while the unbound CRT exhibits the full activity (Figure 1.5) (Verboomen *et al.*, 1994). Moreover, calreticulin is engaged with InsP$_3$-dependent Ca^{2+} release from the ER by binding to the glycosylated intraluminal loop(s) of InsP$_3$ receptor in order to modulate Ca^{2+} release (Michikawa *et al.*, 1994).
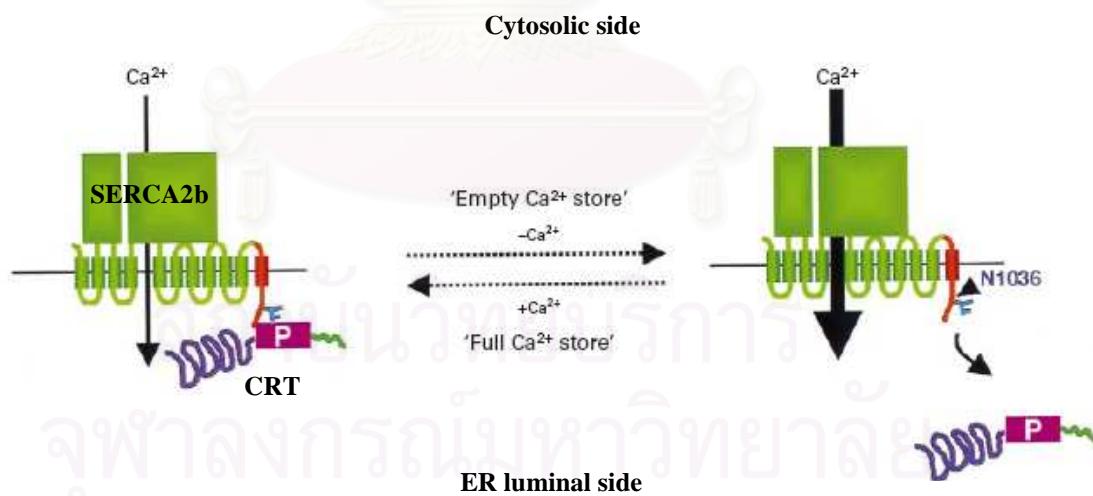
Figure 1.5 A schematic diagram representing modulation of Ca^{2+} homeostasis activity of CRT by SERCA2b. Under the conditions of full Ca^{2+} stores, CRT would bind to the luminal tail of SERCA, resulting in a decrease in SERCA2b activity. In contrast, CRT would not interact with SERCA2b for efficient refilling of the stores (Michalak *et al.*, 1999)

Moreover, CRT can bind to the KXGFF(K/R)R motif of integrin to modulate integrin's function (Coppolino and Dedhar, 1998). However, the mechanism by which calreticulin modulates function of integrin is not yet clear. Two possibilities have been hypothesized. First, calreticulin could mediate integrin's function by coordinating upregulation of expression of focal adhesion proteins i.e. vinculin and N-cadherin (Opas *et al.*, 1996). Second, it is possible that after binding to integrin, calreticulin stabilizes conformation of integrin.

In addition to the KXGFF(K/R)R motif of integrin, a similar motif of KXFFKR, commonly found in the DNA-binding domain of nuclear hormone receptors could be recognized by CRT. Its C-terminus can directly bind to the glucocorticoid and the androgen receptors to inhibit expression of steroid regulated genes (Burns *et al.*, 1994; Dedhar *et al.*, 1994).

Recently, CRT was found on the extracellular membrane surface of many cell types to serve as a mediator of adhesion and as a regulator of the immune response. In addition, it was found on the egg surface of mouse and suggested to be involved in development of both types of the germ cells and in the signal transduction during or after sperm-egg interaction (Park *et al.*, 2001; Tutuncu *et al.*, 2004).

Moreover, CRT has been found to be related to several other biological systems and disorders e.g. impairment of cardiac and brain development during embryogenesis (Rauch *et al.*, 2000; Johnson *et al.*, 2001), several inflammations and infections (Ghebrehiwet and Peerschke, 2004), rheumatic disease, lupus erythematosus, Sjorgren's syndrome, celiac disease and dermatitis herpetiformis (Sontheimer *et al.*, 1995). Recently, it also has been shown to have an effect in suppressing cancer cells (Dwarakanath *et al.*, 2008).

1.4.2 Calnexin

Calnexin (CNX also know as p88, IP90) is a type I transmembrane phosphoprotein located in the ER (Wada *et al.*, 1991). This protein is present in all eukaryotes (Boston *et al.*, 1996). CNX was first identified as a calcium binding protein of the ER membrane in microsomes of the dog pancreas and subsequently characterized as an ER chaperone in mouse and other mammals (David *et al.*, 1993; Schreiber *et al.*, 1994; Tjoelker *et al.*, 1994). It was found to be expressed at both transcriptional and translational levels in various organs of many species (Yamamoto and Nakamura, 1996) (Li *et al.*, 2001). However, unlike most ER chaperones, transcript expression of CNX is not affected by cellular stresses (Bergeron *et al.*, 1994).

1.4.2.1 The CNX protein

Similar to CRT, CNX is evolutionarily-highly conserved (Fuller *et al.*, 2004). The moleculars weight of calnexin was varied, ranging from 60-90 kDa (Wada *et al.*, 1991). Its structure could be divided into two parts, which are luminal- and transmembrane-fragments. The first part is located in luminal of ER, consisting of N- and P-domains while C-domain is in the integral membrane part (Ohsako *et al.*, 1998). Also the luminal compartment is longer than the membrane-inserted region of calnexin.

The N-domain is composed of a sandwich of two anti-parallel β sheet forming globular structure. It contains putative carbohydrate- and Ca^{2+} -binding sites. The former is important for the glycoprotein folding process. Glc_1Man_3 of oligosaccharides in a glycoprotein molecule would be docked by lining along this domain structure through hydrogen bonds (Vassilakos *et al.*, 1998). Compared with other Ca^{2+} binding sites in the CNX molecule, this domain binds to Ca^{2+} with a weaker affinity (Schrag *et al.*, 2001).

Like CRT, CNX contains a proline rich (P-) domain. In the domain are lined with four repeats of the KPEDWDE motif (1), immediately followed by four repeats of the GxWxPPxIxNPxYx (2) motif i.e. in a pattern of 11112222. This domain has a

high affinity in binding to Ca^{2+} (Hammond and Helenius, 1995). The upstream sequence of the repeated motif is associated with non-glycoprotein interaction such as ERp57 (Frickel *et al.*, 2004). This domain connects between the globular and transmembrane domains, conferring another name of 'arm domain'. (Figure 1.6)
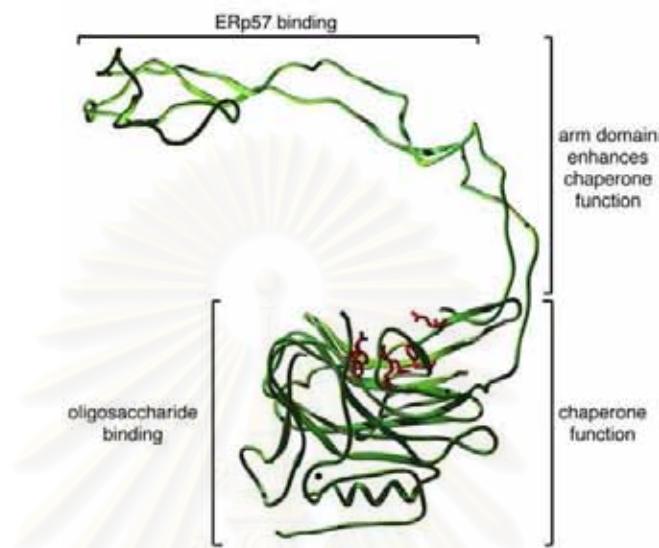
Figure 1.6 Structure and functional sites of the ER luminal domain of calnexin, containing N- or globular domain and P- or arm domain. (Source: Williams *et al.*, 2006)

The membrane-inserted C-terminal domain contains four defined regions-juxtamembrane domain, the acidic domain, the phosphorylation domain and a putative ER retrieval motif (Chevet *et al.*, 1999b). The juxtamembrane domain is lysine-rich and adjacent to the ER membrane. The phosphorylation domain contains three conserved potential serine phosphorylation sites (Chevet *et al.*, 1999a). Two of them contain a motif of SSXXD/E, which is a phosphorylation site target of protein kinase casein kinase 2 sites, while the other contains a TXXD motif, which is targeted to be phosphorylated by PKC/ proline-directed kinase (PDK) (Allende and Allende, 1995). The end of carboxyl terminus is composed of a putative ER retrieval motif (RKXRRE/D) and exposed in the ER lumen (Ho *et al.*, 1999). (Figure 1.7)

C-teminal domain of CNX

5'   **3'**

juxtamembrane   acidic domain

phosphorylation domain   Putative ER retrieval
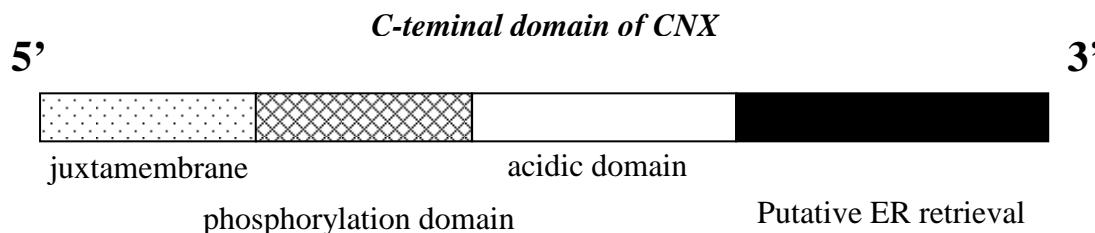
Figure 1.7 Arrangement pattern of its motifs in C-terminal domain (Source : Ho *et al.*, 1999)

1.4.2.2 Functions of CNX

CNX is a chaperone and calcium binding protein. Even so, its function mechanism seems to be poorly understood. There are two hypothetical models on its chaperone mechanism in folding of glycoproteins. The first model is "lectin-only", in which CNX binds only to the oligosaccharide moiety of glycoproteins (Hebert *et al.*, 1995). In contrast, the second model is "dual binding", where CNX interacts with both oligosaccharide and polypeptide segments of glycoproteins during the folding process (Sousa and Parodi, 1995).

For the Ca^{2+} homeostasis function, CNX cooperates with SERCA2b and the IP_3 receptor. Under resting conditions (Ca^{2+} stores are full), CNX is associated with SERCA2b and inhibits its activity. In this state, the pump is sufficiently active to maintain the ER lumen at full Ca^{2+} capacity. This environment is optimal for protein folding given the requirement of ER chaperone activity for Ca^{2+} (Corbett *et al.*, 1998). When the IP_3 signaling pathway is activated several rapid changes occur. First, IP_3-mediated Ca^{2+} release depletes Ca^{2+} from the ER with a corresponding mirror image increase in the cytosol. These cytosolic increases cause activation of the Ca^{2+} dependent phosphatase calcineurin (CN), which de-phosphorylates the PKC/PDK sites of CNX. This results in the dissociation of CNX from C-terminus of SERCA2b, removing pump inhibition. The return to maximum pumping activity rapidly refills the ER lumen and minimizes the potential risk of impaired protein folding during cytosolic Ca^{2+} signaling.

CNX was found to be involved in phagocytosis. It is a constituent of phagosomes in a mouse macrophage-like cell line. Ca^{2+} required in this process is provided from calnexin and calreticulin present in the phagocytic cup (Muller-Taubenberger *et al.*, 2001). It also acts as a modulator of the phagocytosis rate. It inihibits the gonadotropin-releasing hormone receptor (GnRHR), an effect that decreases GnRHR signaling capacity (Brothers *et al.*, 2006). Decreasing of GnRHR affects to increase gonadotropin release (Neill *et al.*, 2004). CNX is a growth-factor-regulated gene and facilitates the downstream signaling in growth-factor-treated cells (Li *et al.*, 2001). Finally, calnexin associates with several cell-surface proteins such as T-cell complex, integrin and histocompatibility complex class I antigens (Lenter and Vestweber, 1994; Ruoslahti and Reed, 1994).

1.4.3 Endoplasmic reticulum protein 57 (ERp57)

ERp57 is a protein disulfide isomerase (PDI), which catalyzes the oxidation, reduction and isomerization reactions to form disulfide bonds during the protein folding. The common structure of PDIs contains 4 thioredoxin-like domains- a, a′, b and b′ (Figure 1.8). The a and a′ thioredoxin domains are thio/disulfide oxidoreductase active sites while b and b′ are redox inactive domains (Goldberger *et al.*, 1964; Kemmink *et al.*, 1997; Ferrari and Soling, 1999; Freedman *et al.*, 2002). Different types of PDIs have different patterns of the domain arrangement, defining their sub-groups in this protein family.

ERp57 (also known as ER-60, ERp60, ERp61,GRp58, P58, HIP70, 1,25D$_3$-MARRS) is first identified from the luminal of ER. Arrangement pattern of its thioredoxin domains is abb'a', the same pattern as that of protein disulfide isomerase-like protein of the testis (PDILT) and pancreas -specific protein disulfide isomerase (PDIp). (Figure 1.9) Unlike most other proteins in the PDI family, which can form disulfide bond in the protein folding by themselves, ERp57 requires to complex with either CRT or CNX. It promotes the intra- or inter molecular disulfide bond formation of glycoproteins (Zapun *et al.*, 1998; High *et al.*, 2000; Frickel *et al.*, 2004; Russell *et al.*, 2004; Williams, 2006). ERp57 uses bb′ fragments for binding to the P-domain of CNX or CRT (Maattanen *et al.*, 2006). Especially, the b' domain is an important region for substrate or partner interaction (Koivunen *et al.*, 2005).

In addition to the disulfide formation activity, ERp57 has many other activities. It is involved in Ca^{2+} homeostasis by forming disulfide bonds with SERCA2b to inhibit the Ca^{2+} pump activity (Tutuncu *et al.*, 2004). Moreover, ERp57 acts as a nuclear receptor for 1,25-dihydroxyvitamin D_3, which plays a role in maintaining of Ca^{2+} homeostasis. It is also essential for hormone-stimulated uptake of calcium and phosphate in chick and rat intestinal cells (Nemere, 2005).

A)
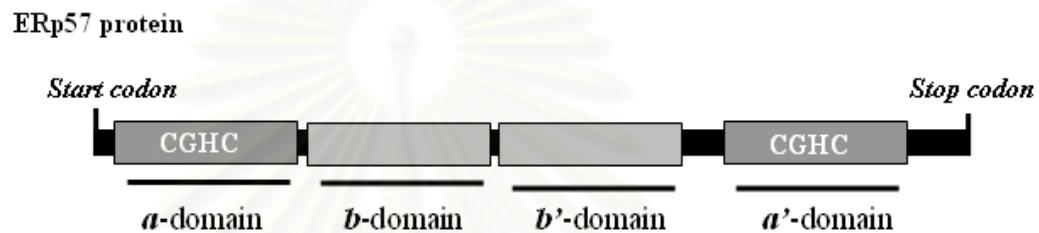
B)
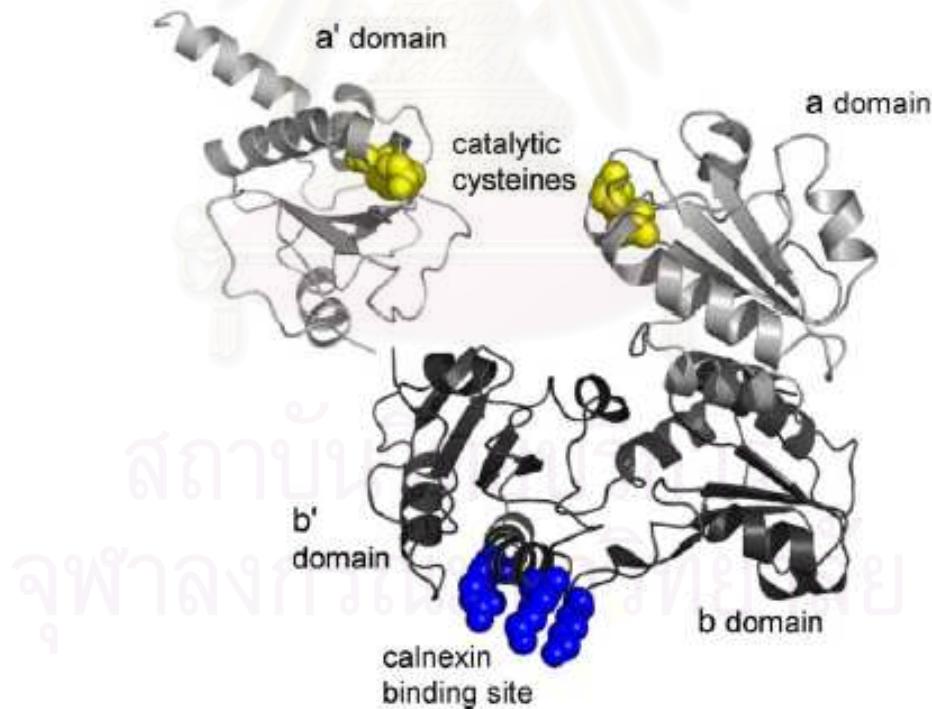
Figure 1.8 (A) Domain arrangement of ERp57 consisting of 4 thioredoxin domains-two active (a, a') and two inactive (b, b') domains. (B) X-ray crystal structure of ERp57 . (Source: Kozlov *et al.*, 2006)
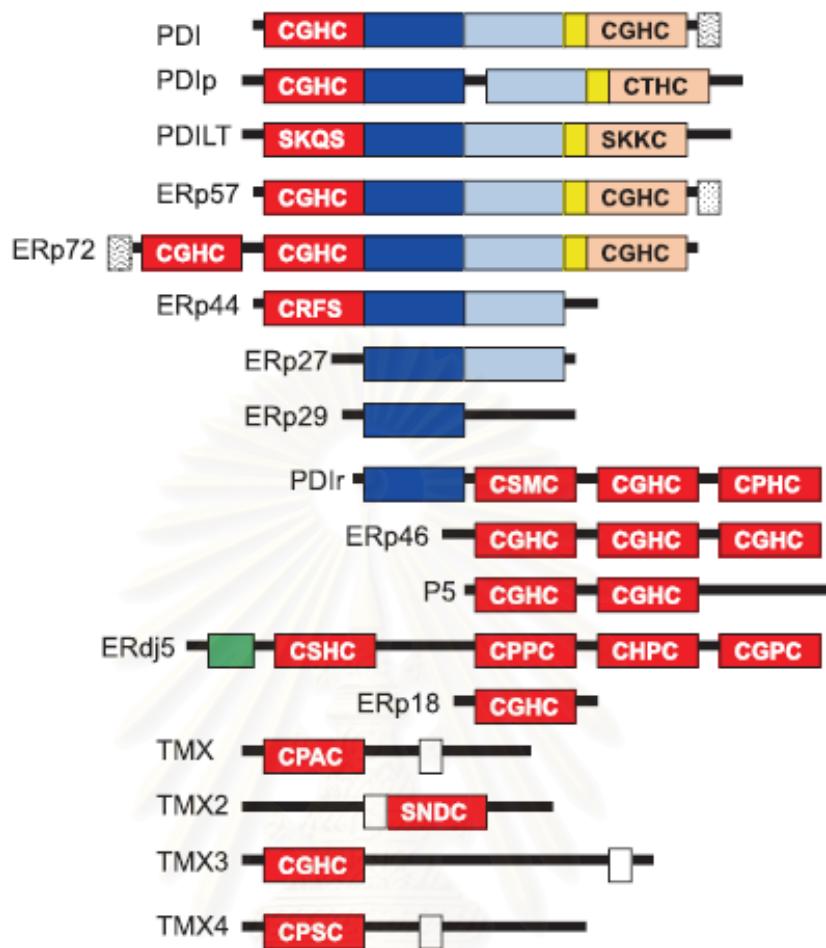
Figure 1.9 A depiction of the domain architectures of human disulfide isomerases. Thaioredoxin fold domains are represented by rectangles. Catalytic domains are represented by red (a) and light red (a'), while noncatalytic domains are represented by blue (b), and light blue (b'). (Source: Maattanen *et al.*, 2006)

ERp57 is expressed at both transcriptional and translational in various tissues of many species (Maattanen *et al.*, 2006). It also exists in acrosome of rat spermatid and on the mouse sperm membrane, playing an essential role in spermatogenesis and fertilization (Ohtani *et al.*, 1993; Ellerman *et al.*, 2006a; Ellerman *et al.*, 2006b). In humans, ERp57 is a component of sperm acrosome, which is important for the gamete fusion (Zhang *et al.*, 2007).

Furthermore, the a' domain of ERp57 has been shown to bind to certain specific DNA sequences in intron, which could be gene regulatory regions. This result suggested involvement of ERp57 in regulation of gene expression. This hypothesis is supported by an evidence showing its reductive activation of AP1 transcription factor, which is associated with stresses. ERp57 expression in fact was also found to be affected by stresses. Its expression in nuclease was upregulated by oxidative stresses (Frickel *et al.*, 2004; Grillo *et al.*, 2007). These results indicate that ERp57 may regulate other downstream stress-responded genes.

1.4.4 Correlations of CRT, CNX and ERp57

1.4.4.1 Structures and functions

Calreticulin and calnexin are luminal ER proteins, whose amino acid sequences and structures are homologous. Both are chaperone proteins, which share similar functions, including Ca^{2+} binding, lectin-like activity, and recognition of misfolded proteins. A major distinction between them is that calnexin is an integral membrane protein whereas calreticulin is a luminal protein.

Sequence homology between CRT and CNX is evident in the P-domains, containing two proline-rich motifs repeated in tandem. Motifs 1 and 2, each of which is repeated three times in CRT and four times in CNX, contain consensus sequences of KPEDWDE and GxWxPPxIxNPxYx, respectively. In addition, sequence homology can be seen in other domains (indicated as boxes A, B, as in Figure 1.10) (Leach *et al.*, 2002).

Both CRT and CNX contain N-, P- and C-domains. The N-domains of these two chaperones in human share about 39% sequence identity (Vassilakos *et al.*, 1998), and their structural conformations are globular, containing eight anti-parallel β-sheets. Even so, while both contain a glucose-binding site in this domain, only CNX accommodate a Ca^{2+} binding site in the N-domain. (Figure 1.11)
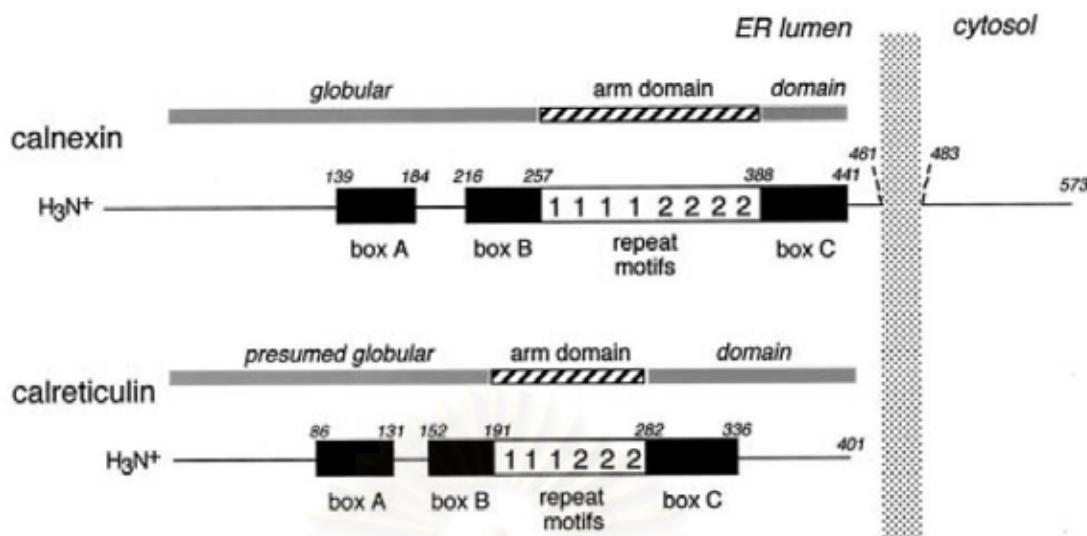
Figure 1.10 Schematic diagram demonstrating amino acid sequence homology between CRT and CNX (Source: Leach *et al.*, 2002)

The P domains of both CNX and CRT accommodate a high affinity Ca^{2+} binding site and a ERp57 binding site. Of CNX, the P-domain contains a sandwich of two β-strands folded in a hairpin configuration. Each β-strand is composed of four tandemly repeated proline-rich motif while CRT is similar to CNX but shorter, namely consisting of three rather than four tandem repeats (Ellgaard *et al.*, 2001). The N- and P-domains of both calreticulin and calnexin are relevant to protein folding.

A major difference could be found in the C-terminal regions of the two proteins. While the C-domain of CRT binds Ca^{2+} with a high capacity, that of calnexin exhibits no Ca^{2+} binding activity. Most part of the C-domain of CNX inserts itself into the ER membrane. Only a minority of its C-domain is present in the ER lumen, which, like that of CRT, could bind to SERCA2b .
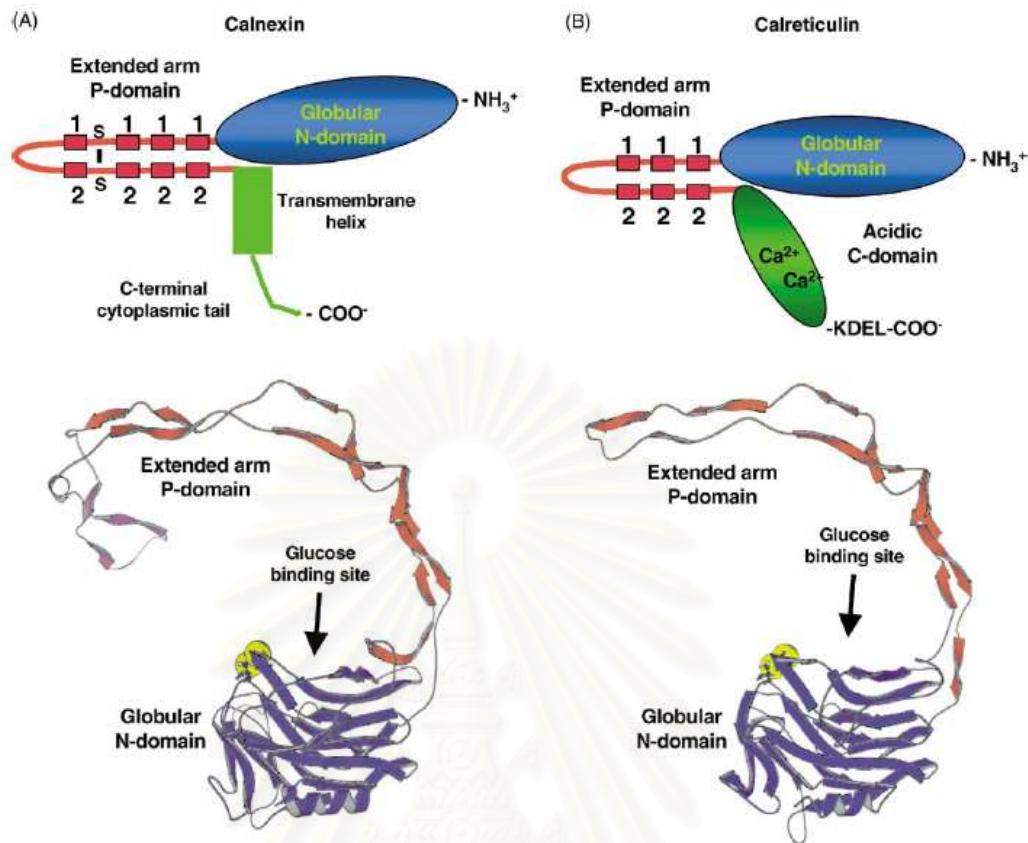
Figure 1.11 X-ray structures of "protein-folding module" of calnexin (A) and calreticulin (B). Both proteins contain a globular N-domain and a central,P-domain, which forms a characteristic loop. The locations of motif repeats 1 (KPEDWDE) and 2 (GXWXPPXIXNPXYX) in the P-domain is indicated as boxes in both proteins. (Source: Michalak *et al.*, 2002)

1.4.4.2 The CRT/CNX chaperone cycle

More than one hundred glycoproteins have been found to be folded by CNX and CRT, Either only one or both of the two chaperones bind to the nascent protein simultaneously or sequentially (Van Leeuwen and Kearse, 1996). Two possibilities of the folding process have been assumed i.e. 'lectin-only' and 'dual-binding' models (Williams, 2006).

In the lectin-only model, binding of CNX/CRT occurs only at the lectin site of glycoproteins. During glycoprotein synthesis, $Glc_3Man_9GlcNAc_2$ is transferred to asparagine residues of the polypeptide chain (Figure 1.12). Under the normal condition, the three terminal glucoses of $Glc_3Man_9GlcNAc_2$ are trimmed or de-glucosylated by glucosidase I and II. When the protein folding happens, both glucosidases are inhibited by their inhibitors and UDP-glucosylate:glycoprotein transferase (UGGT) re-transfers glucose to the oligosaccharide site of all the folding glycoproteins. Subsequently, calreticulin and calnexin bind to the $Glc_3Man_9GlcNAc_2$ moiety and initiate the folding process (Taylor *et al.*, 2004; Ritter *et al.*, 2005).

The dual binding model proposes that CRT and CNX recognize and bind to polypeptide-based determinants in addition to the oligosaccharide and thus behave more like classic chaperones. This model contains two steps for binding both proteins to unfolded glycoprotein. The first, both proteins bind to monoglucosylated oligosaccharide chains of unfolded glycoproteins and then bind more stably to peptide determinants, which disappear upon folding. The last step, CRT or CNX binds to glycoprotein using hydrophobic interaction (Ware *et al.*, 1995). (Figure 1.13)
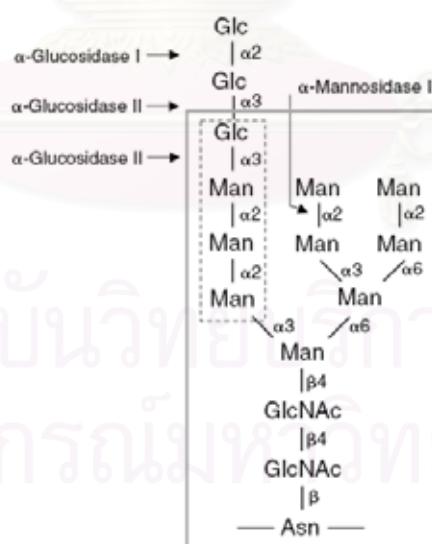
Figure 1.12 Composition of the Glc3Man9GlcNAc2 oligosaccharide initially transferred to asparagine residues of nascent glycoproteins. (Source: Williams *et al.*, 2006)
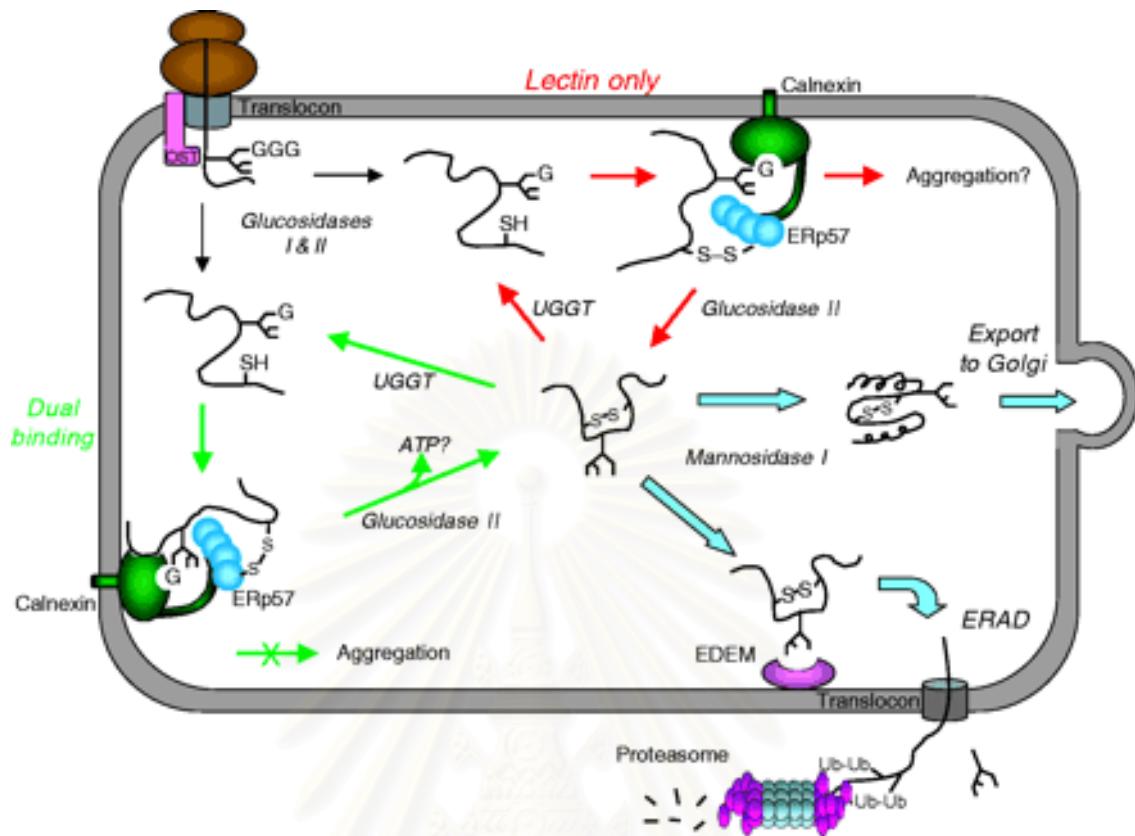
Figure 1.13 Schematic diagram displaying calreticulin and calnexin chaperone cycle in the ER lumen. In the lectin-only model (red arrows), cycles of glycoprotein release and re-binding are controlled solely by the removal and re-addition of the terminal glucose residue by glucosidase II and UDP-glucose:glycoprotein glucosyltransferase (UGGT), respectively. UGGT is the folding sensor because it only reglucosylates non-native glycoprotein conformers. Chaperone binding serves to retain non-native glycoproteins within the ER and also recruits ERp57 to promote disulfide-bond formation and isomerization. It is unclear whether binding to glycoproteins only through the lectin site is sufficient to suppress aggregation. In the dual-binding model (green arrows), Cnx and Crt recognize non-native glycoproteins through their lectin sites as well as through a polypeptide-binding site specific for non-native conformers. This allows them to prevent off-pathway aggregation reactions similarly to other molecular chaperones. Binding via the polypeptide binding site is influenced by ATP and by the free Ca^{2+} concentration, either of which may regulate the interaction (Source: Williams, 2005).

1.4.4.3 Correlation of CNX, CRT and ERp57 in Ca^{2+} homeostasis

For Ca^{2+} homeostasis, the N domain of CRT interacts with the C-terminus of SERCA2b while the C-domain of CNX binds to the C-terminus of SERCA2b. However, ERp57 can also bind to the C-terminus of SERCA2b but with a requirement of CRT. These interactions lead to decrease in the Ca^{2+} influx and eventual Ca^{2+} concentration in the ER lumen. Despite the functional cooperation of the three proteins, CRT and CNX can functionally compensate each other (Michalak *et al.*, 2002).

The Ca^{2+} homeostasis is in fact necessary for the glycoprotein folding in the ER. Concentration of Ca^{2+} in the lumen is needed to be maintained at optimum to promote an efficient folding condition. The CRT/CNX cycle exists in parallel with the action of CNX/CRT on SERCA2b. Given the homeostasis system is absent. Ca^{2+} depletion could result in protein unfolding and several subsequent responses, which can bring about cellular apoptosis. Maintaining optimal Ca^{2+} concentration in the ER is thus important for protein folding and cellular survival. Moreover, the IP_3 receptor is also known as Ca^{2+} efflux regulation protein. It can bind to CRT to induce release of Ca^{2+} from the ER (Corbett *et al.*, 1999; Roderick *et al.*, 2000).

1.5 Techniques performed in the study

1.5.1 Polymerase chain reaction (PCR)

1.5.1.1 PCR

The introduction of the polymerase chain reaction (PCR) by Mullis et al. (1987) has opened a new approach for molecular genetic studies. This method is a *in vitro* technique of DNA replication by an enzyme called DNA polymerase, using specific DNA sequences of two oligonucleotide primers, usually 18-27 nucleotides in length. Million copies of the target DNA sequence can be synthesized from the low amount of starting DNA template within a few hours.

The PCR reaction components are composed of DNA template, a pair of primers for the target sequence, dNTPs (dATP, dCTP, dGTP and dTTP), PCR buffer and heat-stable DNA polymerase (usually Taq polymerase). The amplification reaction typically consists of three steps: denaturation of double-stranded DNA at high temperature, annealing to allow primers to form hybrid molecules at the optimal temperature, and extension of the annealed primers by the polymerase. The cycle is repeated normally for 30-40 times, in each of which the amount of DNA is exponentially increased. (Figure 1.14)
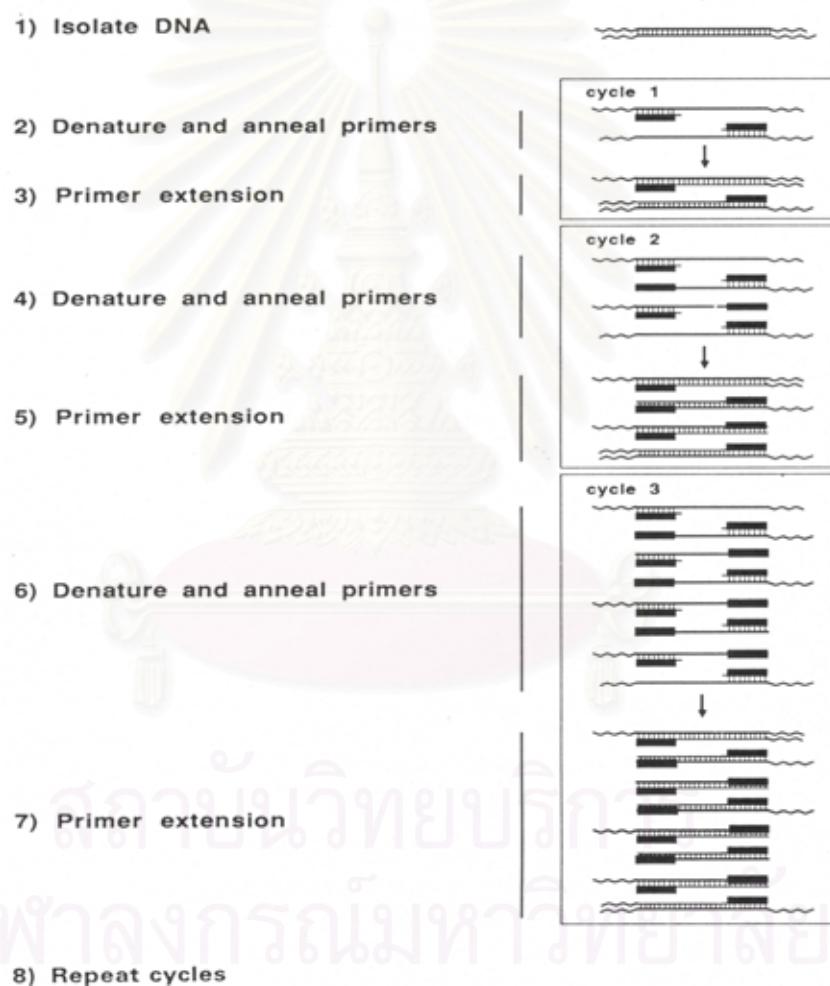
Figure 1.14 General illustration of the polymerase chain reaction (PCR) for amplification of the target DNA.

1.5.1.2 Reverse transcription-polymerase chain reaction (RT-PCR)

RT-PCR is a method used to amplify, isolate or identify a known sequence from a cellular or tissue RNA. The PCR is preceded by a reaction using reverse transcriptase and oligo(dT) or random primers to convert RNA to cDNA. The first strand cDNA, rather than genomic DNA, was then used as a template in the second-strand synthesis using a pair of gene-specific primers. It is a direct method for examination of gene expression of known sequence transcripts in the target species.

RT-PCR can also be used to identify homologues of interesting genes by using degenerate primers and/or conserved gene-specific primers from the original species and the first strand cDNA of the interesting species is used as a template. The amplified product is then further characterized by cloning and sequencing. (Figure 1.15)

1.5.2 Rapid amplification of cDNA ends-polymerase chain reaction (RACE-PCR)

RACE-PCR is a common approach used for isolation of the full length of RNA transcripts. Its procedure contains two separate reactions: 5' and 3'-end RACE PCRs. The technique is relied on addition of the SMART (**S**witching **M**echanism **AT** 5' end of **R**NA **T**ranscript) adaptor to the 5' or 3' end of the first-strand cDNA. For the 3'-RACE PCR, the adaptor is added to the 5' end of the first-strand cDNA, using an oligo-dT primer containing the SMART sequence. On the contrary for the 5'-RACE first-strand synthesis, terminal transferase activity of Powerscript Reverse Transcriptase (RT) adds 3-5 nucleotides (predominantly dC) to the 3' end of the first stand cDNA upon reaching toward the end of transcript templates. The 3' terminal dG stretch of SMART oligonucleotide then anneals to the added dC tail of the first-strand cDNA, allowing sequence of the SMART adaptor at the 5' end of the SMART oligonucleotide to be transcribed right next to the GC joining region (Figure 1.16).
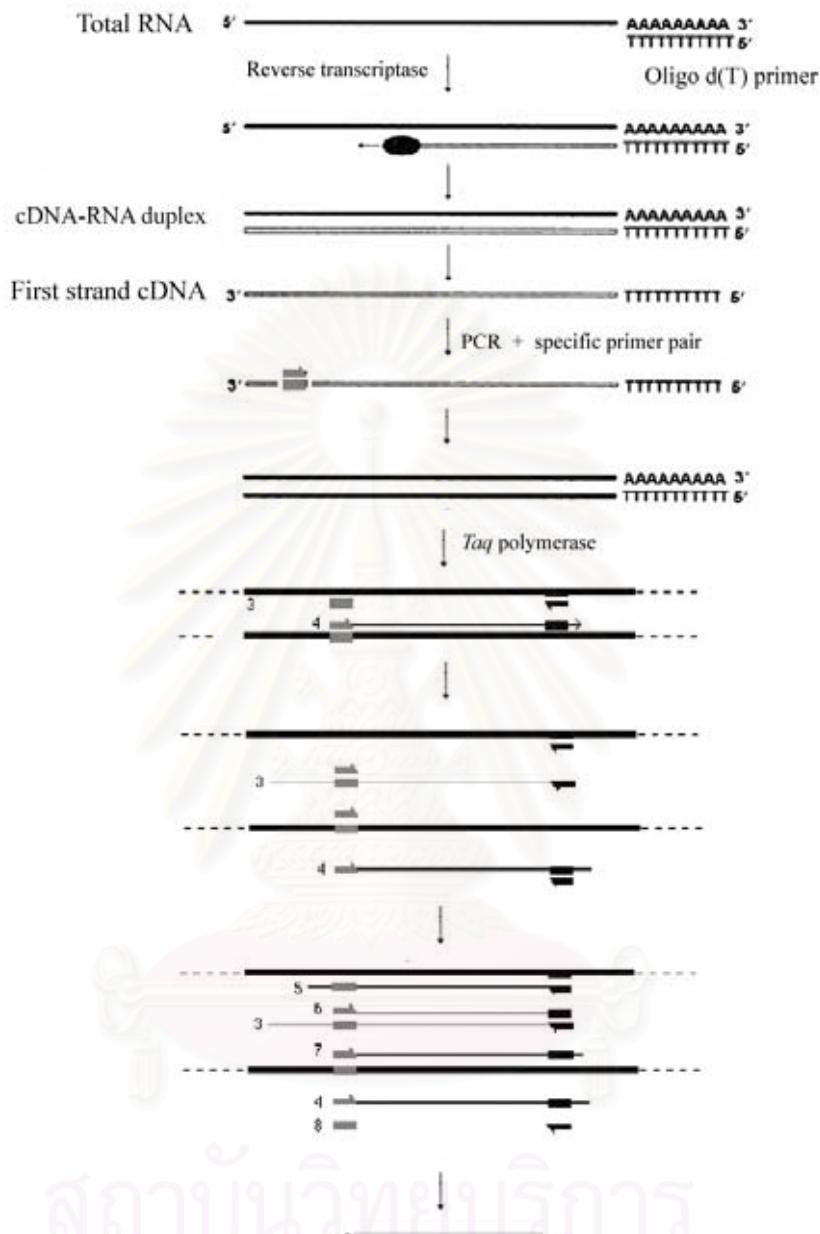
Figure 1.15 Overall concepts of RT-PCR. During the first strand cDNA synthesis, an oligo d(T) (or random primers) primer anneals and extends from sites present within mRNA. The second strand cDNA synthesis primed by the 18 – 25 base specific primer proceeds during a single round of DNA synthesis catalyzed by thermostable DNA polymerase (e.g. *Taq* polymerase).

A
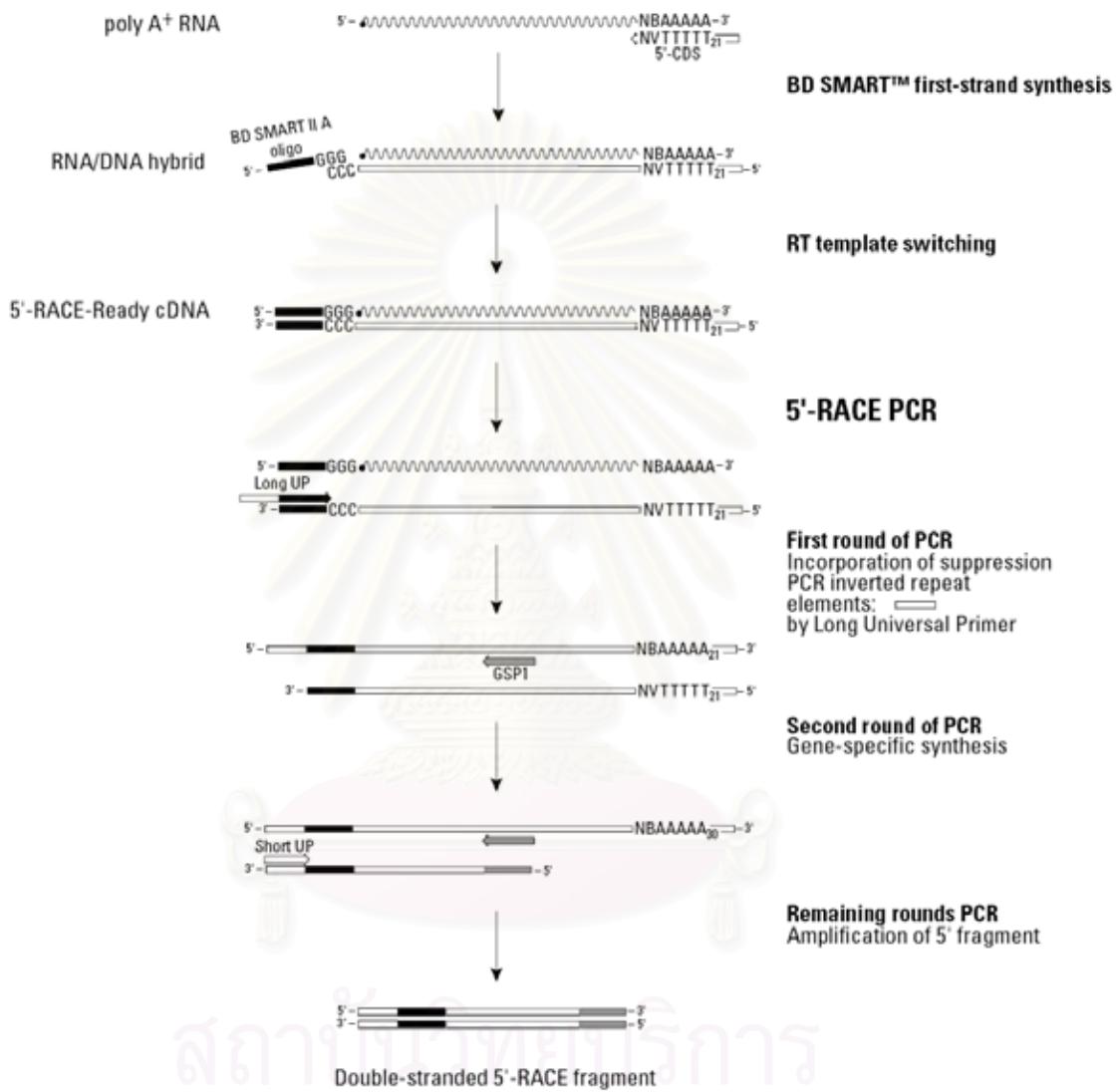
Figure 1.16 Overview of the RACE-PCR method. A) Detailed mechanism of the 5´ RACE reaction

B

poly A⁺ RNA

3'-RACE-Ready cDNA

Standard first-strand synthesis
■■■ = BD SMART II A sequence

3'-RACE PCR

GSP2

First round of PCR
Gene-specific synthesis

Long UP

Next rounds of PCR
Incorporation of suppression
PCR inverted repeat
elements: ▭
by Long Universal Primer

Short UP

Remaining rounds of PCR
Amplification of 3' fragment

Double-stranded 3'-RACE fragment
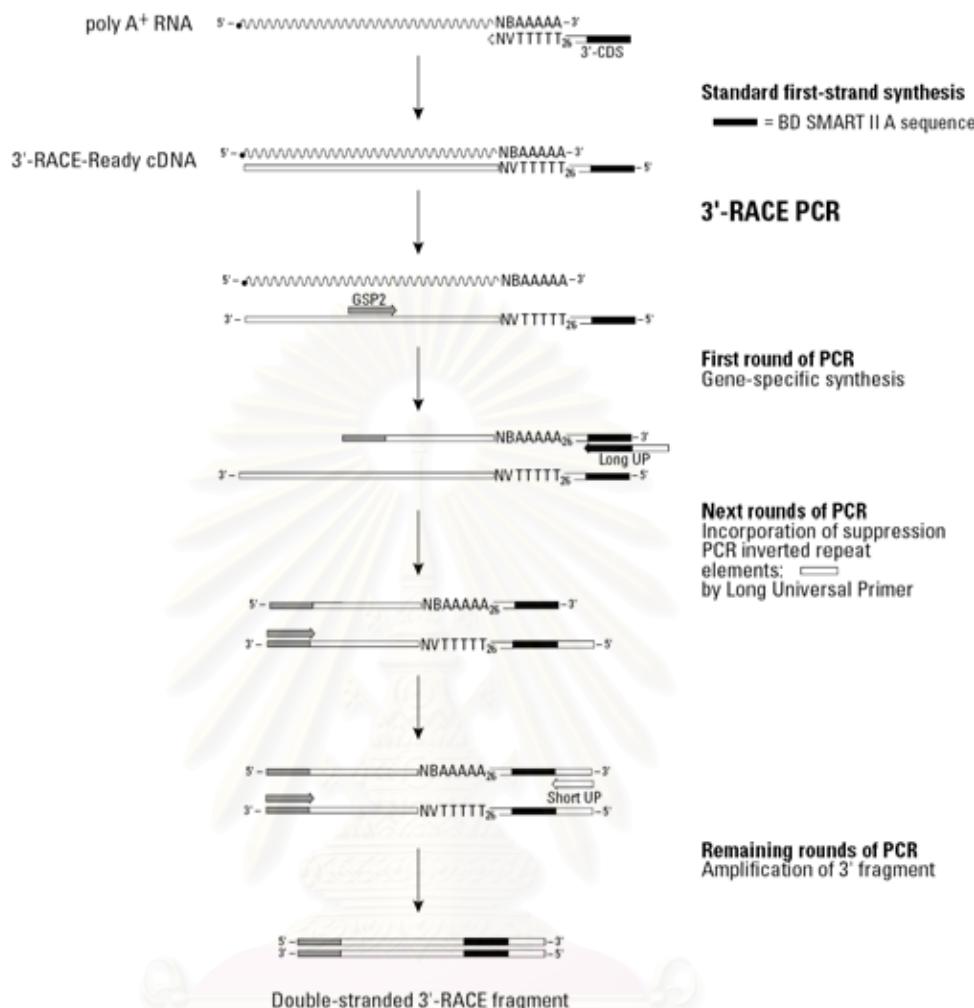
Figure 1.16 (next) Overview of 3´ RACE-PCR amplification

The modified first-strand cDNA containing the SMART sequence is then used as as the template for RACE PCR reactions in the next step. Gene specific primer (GSPs) are designed from partial sequences of the interested genes for 5'-RACE PCR (antisense primer) and 3'-RACE PCR (sense primer) and used together with the universal primer (UPM), which recognizes the SMART sequence in the modified frist-strand cDNA. RACE products are then cloned and characterized for their sequences.

1.5.3 Techniques for differential gene expression studies

Differential gene expression can be studied either semi-quantitatively or quantitatively. Both approaches are developed from the simple polymerase chain reaction (PCR). (Figure 1.17)

1.5.3.1 Semi-quantitative RT-PCR

Semi-quantitative RT-PCR is a comparatively quantitative approach, in which the target gene and the internal control (e.g. a housekeeping gene) are separately or simultaneously amplified using the same template. The internal control such as *β-actin*; *elongation factor*, *EF-1α* or G3PDH) is used under the assumption that its coding sequence is transcribed constantly and independently from the extracellular environment stimuli and that its transcript is reverse-transcribed with the same efficiency as the interested transcript.
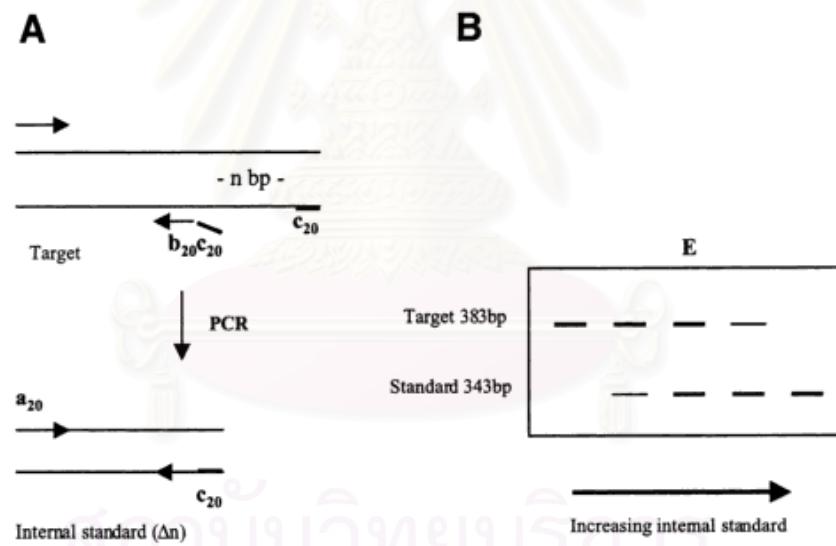
Figure 1.17 Quantitative competitive PCR. (A) Generation of an internal standard using composite primers. (B) Use of the internal standard for competitive PCR. The equivalence point (E) occurs when target and standard are present in equal starting concentration, thus permitting quantification of the target template.

1.5.3.2 Real-time PCR

Real Time PCR is a kinetic approach used to amplify and simultaneously quantify a targeted DNA molecule. It enables both detection and quantification (as the absolute number of copies or relative amount when normalized to DNA input or additional normalizing genes) of a specific sequence in the sample.

The procedure follows the general principle of PCR. Its key feature is that the amplified DNA is quantified in real time after each amplification cycle as it accumulates in the reaction. Two common methods of the product detection are the use of fluorescent dyes that intercalate with double-stranded DNA (e.g. SYBR green), and the use of modified DNA oligonucleotide probes that are fluorescent when hybridized with a complementary DNA.

In fluorescent dye-based assays, the intercalating dye binds and detects non-specifically any double-standed DNA that accumulates during the PCR like by-products and primer-dimers. With the problems of previous assays, it affects to use fluorescent probe-base assay. A labeled sequence-specific probe anneals to amplification product between the primer binding sites. Therefore, only the correct amplicon containing both primer and probe binding sites will generate a signal. However, when comparison the cost between two assays then the later assay is expensive than the further one.

The general principle of SYBR green polymerase chain reaction is composed of the first denaturation step, at which the unbound dye molecules are weakly fluorescent; the second step, primer annealing, where a few dye molecules bind to the double strand; the last step, product extension, when more dye molecules bind to the newly synthesized DNA, resulting in an increase of the fluorescence signal. Fluorescence measurement at the end of the elongation step of every PCR cycle is then performed to monitor the increasing amount of amplified DNA (Figure 1.18).

1.5.3.3 SYBR green

Real-time PCR in the laboratory can be applied to numerous applications. It is commonly used for both diagnostic and research applications. Diagnostic real-time PCR is to rapidly detect the presence of genes involved in infectious diseases, cancer and genetic abnormalities. In the research setting, real-time PCR is mainly used to provide highly sensitive quantitative measurements of gene transcription. Further, this technique may be used to follow change of gene expression over time due to physiological changes or treatments, such as response of tissues or cells to an administration of a pharmacological agent, progression of cell differentiation, or response to changes in environmental conditions.
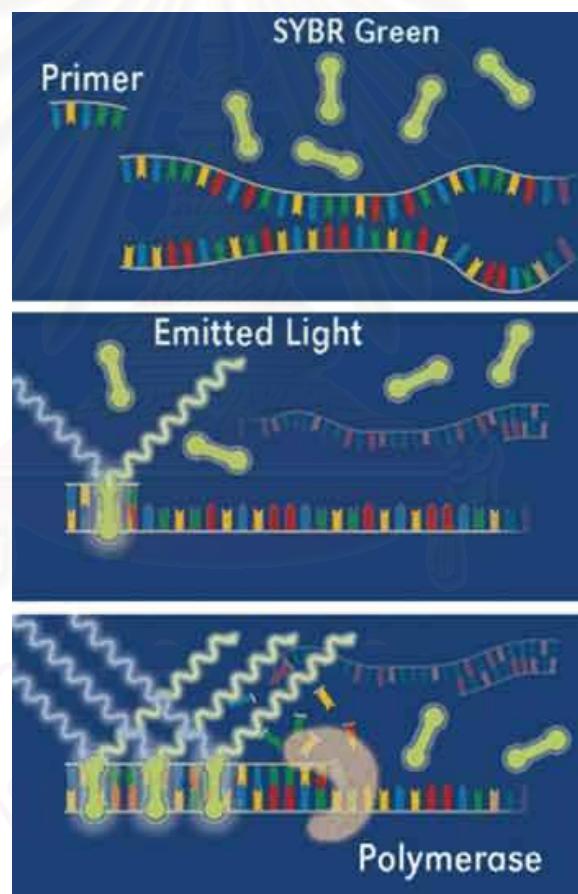
Figure 1.18 A schematic diagram showing Real-time PCR procedure using SYBR Green I Dye (Source: www.thaiscience.com/lab_vol/p23/rt_pcr/pcr5.jpg)

1.5.4 Techniques for study of genomic organization

Genome walking also known as DNA walking is a simple method to identify genomic DNA sequences of interested genes. First, a genome walking mini-library is constructed using different blunt-end generated restriction enzymes to create fragments of genomic DNA (Figure 1.19). After that, the fragments are ligated with a known-sequence adaptor using T4 DNA ligase. The adaptor-accommodated genomic DNA fragments were used as templates for PCR using Adapter primer (AP) and gene specific primer (GSP). The resulting amplified products were then cloned and characterized.

1.6 The objective of this thesis

The objectives of this thesis were isolation and characterization of genes involving calcium metabolism of *P. monodon*. The full length cDNA of CRT, CNX and ERp57 genes were isolated and characterized by RACE-PCR. In addition, the expression profiles of genes in various tissues in juveniles and broodstocks *P. monodon* were examined using RT-PCR. The gene expression levels measuring of each gene under thermal stress condition were examined using semi-quantitative RT-PCR and real-time PCR. Recombinant proteins of three genes homologues were produced using a bacterial expression system and then assayed for their activities.
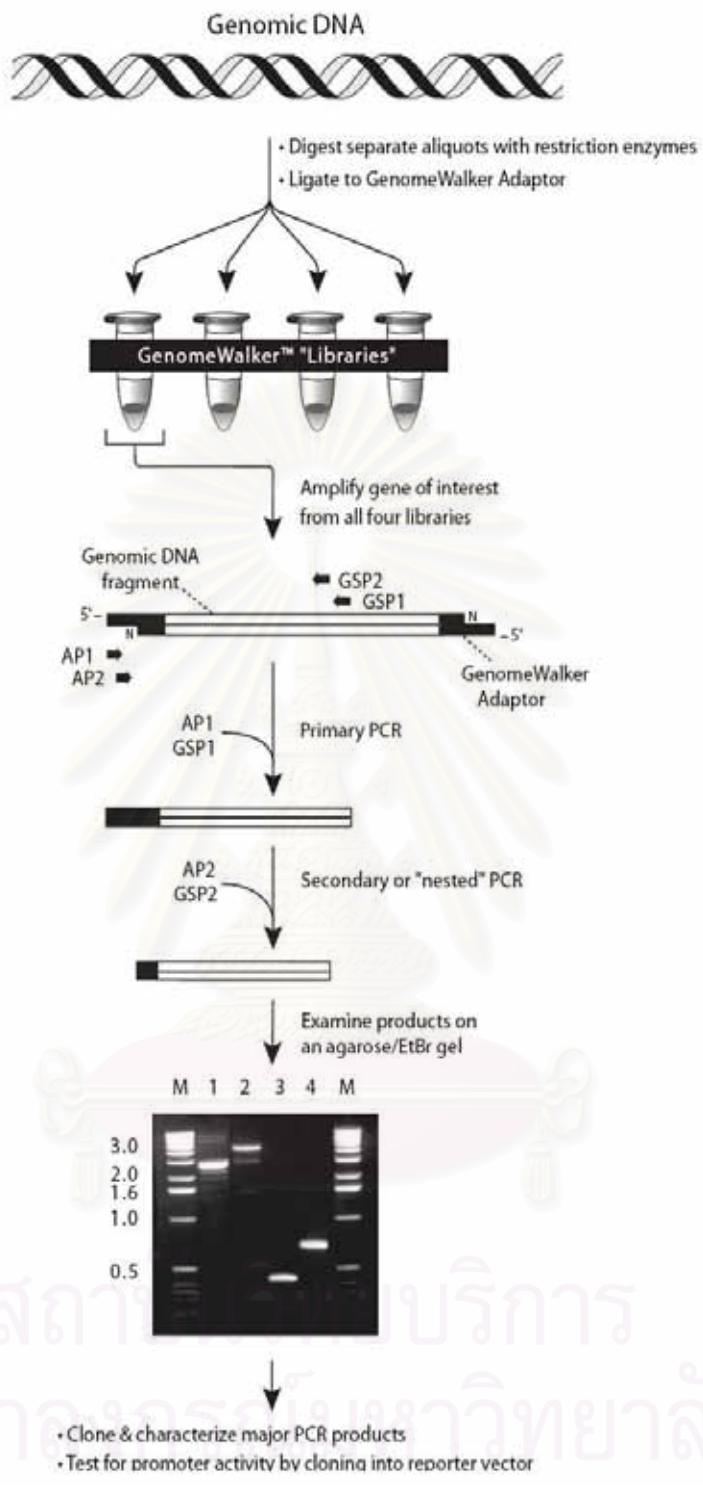
Figure 1.19 A flow chart showing the Genome walking protocol

CHAPTER II

MATERIALS AND METHODS

2.1 Experimental samples

Domesticated *P. monodon* juveniles (approximately 20g body weight, 4-month-old) were purchased from local farms in Chonburi in the eastern part of Thailand. After transported, they were acclimated under the laboratory condition for 14 days before dissected or subjected to treatments.

Male and female *P. monodon* broodstocks were wild-caught from the Andaman Sea, in the western south of Thailand. Prior to tissue dissection, the broodstock were acclimated for 3 days under the farm condition.

Tissues were dissected from live shrimps and immediately stored at -80 °C until required. Hemolymph was collected using 10% sodium citrate as an anticoagulant. The anticoagulant was removed from the sample by centrifugation at 1000 *g* for 10 minutes, and the hemocyte pellets were immediately stored at -80 °C.

2.2 Nucleic acid extraction

2.2.1 Genomic DNA extraction

Genomic DNA was extracted from a piece of pleopod tissue taken from a broodstock shrimp using a phenol-chloroform-proteinase K method (Klinbunga et al., 1999). A piece of pleopod tissue was briefly homogenized in 500 μl of a prechilled extraction buffer (100 mM Tris-HCl, 100 mM EDTA, 250 mM NaCl; pH 8.0), using a micropestle. SDS (10%) and RNase A (10 mg/ml) solutions were added to final concentrations of 1.0 % (w/v) and 100 μg/ml, respectively. The resulting mixture was then incubated at $37^{o}C$ for 1 hour. At the end of the incubation, a proteinase K solution (10 mg/ml) was added to the sample to a final concentration of 300 μg/ml and the sample was further incubated at 55 ^{o}C for 3 – 4 hours. After the incubation, the sample was gently mixed with the sample volume of buffer-equilibrated phenol: chloroform: isoamylalcohol (25:24:1) for 10 minutes. The sample was centrifuged at 13,500 g for 10 minutes at room temperature, and the upper aqueous phase was

collected. The extraction process was then repeated once with phenol: chloroform: isoamylalcohol(25:24:1) and once with chloroform:isoamylalcohol (24:1). The final aqueous phase was mixed with the sample volume of TE buffer and subsequently with one-tenth final sample volume of 3 M sodium acetate (pH 5.2). After that, DNA was precipitated by adding twice the sample volume of -20 °C-cold absolute ethanol. The mixture was incubated at -80 ^{o}C for 30 minutes, and the precipitated DNA was recovered by centrifugation at 16,300 g for 10 minutes at room temperature. The DNA pellet was then washed twice with 1 ml of -20°C-cold 70% ethanol. After air-dried, the DNA pellet was resuspended in Tricine-EDTA (TE) buffer (10 mM Tris-HCl, pH 8.0 and 0.1 mM EDTA). The DNA solution was incubated at 37 ^{o}C for 1 – 2 hours and stored at 4 ^{o}C until further needed.

2.2.2 RNA extraction

Total RNA was extracted from the dissected tissues using TRI REAGENT (Molecular Research Center). A piece of tissue was immediately placed in mortar containing liquid nitrogen and ground to fine powder under frozen condition. The tissue powder was transferred to 500 μl of TRI REAGENT and homogenized. Additional 500 μl of TRI REAGENT were later added to the homogenized sample (a final proportion of 1 ml Trizol/ 50-100 mg tissue). The homogenate was left at room temperature for 5 minutes before 0.2 ml of chloroform was added. The mixture was vortexed for at least 15 seconds, left at room temperature for 2 - 15 minutes and centrifuged at 12000 g for 15 minutes at 4 °C. The mixture was separated into the lower phenol-chloroform phase (red), the interphase, and the colorless upper aqueous phase.

The aqueous phase (containing the extracted RNA) was carefully collected. RNA was precipitated with 0.5 ml of -20 °C-cold isopropanol. The mixture was left at room temperature for 10 - 15 minutes and centrifuged at 12000 g for 10 minutes at 4 °C. The RNA pellet was then collected and washed with 1 ml of -20 °C-cold 75% ethanol prior to centrifugation at 7500 g for 10 minutes at 4°C. The washed RNA pellet was air-dried for 5-10 minutes and then dissolved in DEPC-treated H$_2$O for immediate use. Alternatively, the RNA pellet was kept in absolute ethanol at -80 °C until required. The quality of extracted total RNA was examined by electrophoresis on 1.0% agarose gel.

2.2.3 DNase treatment of the extracted RNA

Ten μg of total RNA were treated wiyh DNase I (0.5 U/1 μg of RNA, GE Healthcare) at 37ºC for 30 minutes.After the incubation, the sample was gently mixed with the a sample volume of phenol:chloroform:isoamylalcohol (25:24:1) for 10 minutes. The sample was centrifuged at 13,500 g for 10 minutes at 4 ℃, and the upper aqueous phase was collected. The extraction process was then repeated once with chloroform:isoamylalcohol (24:1) and one with chloroform. The final aqueous phase was mixed with one-tenth final sample volume of 3 M sodium acetate (pH 5.2). After that, RNA was precipitated by adding twice the sample volume of -20 ℃-cold absolute ethanol. The mixture was incubated at -80 ℃ for 30 minutes, and the precipitated RNA was recovered by centrifugation at 16,300 g for 10 minutes at room temperature. The RNA pellet was then washed twice with 1 ml of -20℃-cold 75% ethanol. Alternatively, the RNA pellet was kept in absolute ethanol at -80 ℃ until required.

2.3 Measurement of nucleic acid concentration

Concentration of the extracted DNA and RNA was estimated by spectophotometry method. Light absorbance was measured at a wavelength of 260 nm (A_{260}). An A_{260} value of 1.0 corresponds to concentration of 50 μg/ml double-strand DNA, 40 μg/ml single-strand RNA and 33 μg/ml oligonucleotide (Sambrook et al., 2001). Therefore, concentration of nucleic acids was estimated in μg/ml by using the following equation:

$$[\text{Nucleic acid}] = A_{260} \times \text{dilution factor} \times 50 \text{ or } 40 \text{ or } 33$$
(for DNA, RNA and oligonucleotides, respectively)

Purity of the nucleic acid samples can be inferred from a ratio of A_{260}/A_{280}. The ratio much lower than 1.8 indicates contamination of residual proteins or organic solvents whereas the ratio greater than this value indicates contamination of RNA in the DNA solution (Kirby, 1992).

2.4 Agarose gel electrophoresis

Appropriate amounts of agarose and 1x TBE buffer (89 mM Tris-HCl, 89 mM boric acid 2 mM EDTA, pH 8.3) were mixed together. Agarose was completely melted by heat and cooled down to approximately 60 °C before poured into a mold. The gel was left to set before submerged in 1 x TBE buffer filled in an electrophoresis chamber.

Appropriate volumes of nucleic acid samples were mixed with one-fourth sample volume of 10x loading dye (0.25% bromophenol blue and 25% Ficoll in water) and loaded into the gel wells. Electrophoresis was carried out at 100 volts until the bromophenol blue front moved to approximately one-half of the gel. The electrophorised gel was stained with an ethidium bromide solution (2.5 µg/ml) for 5 min and destained in 1 x TBE buffer. Nucleic acid products were visualized under a UV transilluminator and photographed through a red filter using a Biorad gel doc machine.

2.5 Synthesis of the first strand cDNA

One and a half micrograms of total RNA extracted from the tissues of *P. monodon* were reverse-transcribed to the first strand cDNA using an ImProm- II^{TM} Reverse Transcription System Kit (Promega). First, total RNA was combined with 0.5 µg of oligo dT$_{12-18}$ and an appropriate amount of DEPC-treated H$_2$O to make up the final volume of 5 µl. The mixture was incubated at 70 °C for 5 minutes and then immediately placed on ice for 5 minutes. Then, 5x reaction buffer, MgCl$_2$, dNTP Mix, RNasin were added to final concentrations of 1x, 2.25 mM, 0.5 mM and 20 units, respectively. Finally, 1 µl of ImProm- II^{TM} Reverse transcriptase was added and gently mixed by pipetting. The reaction mixture was incubated at 25 °C for 5 minutes and at 42°C for 90 minutes, followed by an incubation at 70 °C for 15 minutes to terminate the reverse transcriptase activity. Concentration and quality of the synthesized first strand cDNA were examined by spectophotometry and 1.2 %- agarose gel electrophoresis.

2.6 Tissue distribution analysis of gene expression

Gene expression in tissues was studied in antennular gland, eyestalk, gills, heart, hemocytes, hepatopancreas, lymphoid organs, intestine, ovaries, pleopods, stomach, testes and thoracic ganglion of both juvenile and broodstock of *P. monodon*.

The study was carried out by RT-PCR. Primers were designed, based on Expressed Sequence Tag (EST) sequences of the interested gene homologues, which had been previously identified from *P. monodon* broodstock (Preechaphol et al., 2007). The partial sequences of homologue *CRT*, *CNX* and *ERp57* transcripts brought from ovarian, lymphoid organ and hemocyte cDNA libraries of *P. monodon*, respectively (Table.2.1). Primers used to amplify *CRT, CNX* and *Erp57* are pair primer of *CRT*, *CNX* and *ERp57*, respectively (Table 2.2).

RT-PCR was performed on 25-µl reaction mixtures containing 150 ng of an appropriate first strand cDNA template derived from mRNA extracted from each dissected tissue, 10 mM Tris-HCl (pH 8.8), 50 mM KCl, 0.1 % Triton X-100, 2 mM MgCl$_2$, 100 mM each of dATP, dCTP, dGTP and dTTP, 0.2 µM of an appropriate primer pair and 1 unit of Dynazyme^{TM} DNA polymerase (FINNZYMES). The reaction thermal profile of each gene was shown in Table 2.3. Five µl of the amplification products were electrophoretically analyzed on 1.5 or 1.8 % agarose gels. Tissue distribution of the target genes was studied in reference to that of a house-keeping gene, *Elongation factor-1α* (*EF-1α*).

Table 2.1 The EST-library partial sequences of homologue interesting genes were used to design gene specific primers

Gene	EST clone no.	source of EST library
Calreticulin	OV-N-ST01-0117-W	ovary
Calnexin	LP-V-S01-0622-LF	lymphoid organ
ERp57	HC-N-N01-2411-LF	hemocyte

Table 2.2 Gene homologue, melting temperature, primer sequences and the expected sizes of the PCR product from EST of *P. monodon*

No	Name	Primer sequence	Tm(°C)	Expected size (bp)
1	*CRT*	F:5´ACATTGACTGTGGTGGAGGAT3´	58	244
		R:5´TCGTAGGTGTTGTCAGGATTG3´	58	
2	*CNX*	F:5´TGATATTTGCCTGTTGTTGTGC3´	56	493
		R:5´TATTACCCACATTCGTCACTACATTAT3´	57	
3	*ERp57*	F:5´ATGTCTCCCTTCCCAGCACT3´	62	281
		R:5´GTCTTGTTTCCCTTACACCACC3´	63	
4	*EF-1α*	F:5´ATGGTTGTCAACTTTGCCCC3´	55	500
		R:5´TTGACCTCCTTGATCACACC3´	55	

Table 2.3 Amplification condition for interesting gene expression level analysis in various tissues

Amplification condition	Gene homologue											
	CRT			*CNX*			*ERp57*			*EF-1α*		
	Temp.(°C)	Time	Number of cycles	Temp.(°C)	Time	Number of cycles	Temp.(°C)	Time	Number of cycles	Temp.(°C)	Time	Number of cycles
Initial denaturation	94	3 min	1	94	3 min	1	94	3 min	1	94	3 min	1
Denaturation	94	30s		94	30s		94	30s		94	30s	
Annealing	58	30s	30	55	45s	30	68	30s	30	53	45s	21
Extension	72	30s		72	45s		72	30s		72	45s	
Final extension	72	7min	1	72	7min	1	72	7min	1	72	7min	1

2.7 Molecular cloning of isolated PCR products

2.7.1 Elution of DNA from agarose gels

After electrophoresis, desired individual DNA bands were excised from agarose gels (200-300 mg) using a sterile scalpel. DNA was extracted from the gel pieces using HiYield™ Gel/PCR DNA Extraction Kit (RBC; Real Biotech Corporation). The extraction was carried out by following the manufacturer's instruction. In brief, the gel pieces were incubated at 55 °C with the DF buffer until they were completely dissolved. The samples were then transferred to DF columns and centrifuged at 10800 g for 30 s. The flow-through was discarded. After this step, 0.5 ml of the Wash buffer was added to the DF column and centrifuged as above. The flow-through again was discarded. The column was then centrifuged one more time to remove the trace amount of Wash solution. DNA was finally eluted from the column with 15 μl of the Elution buffer by centrifugation at 16300 g for 2 min.

2.7.2 Ligation of PCR products to the pGEM® -T easy vector

The gel-purified PCR products were cloned into the pGEM-T Easy vector (Promega) in 10-μl reactions containing 2x Rapid Ligation Buffer (60 mM Tris-HCl, pH 7.8, 20 mM MgCl$_2$, 20 mM DTT, 2 mM ATP and 10% polyethylene glycol), 3 units of T4 DNA ligase, 25 ng of pGEM®-T easy vector and 50 ng of DNA insert. The reaction mixtures were incubated overnight at 4 °C before transformed to *E. coli* JM109.

2.7.3 Transformation of ligation products into *E. coli* host cells

2.7.3.1 Preparation of competent cells

A single colony of *E. coli* JM109 was inoculated in 10 ml of LB broth (1% Bactotryptone, 0.5% Bactoyeast extract and 0.5% NaCl, pH 7.0) with vigorous shaking at 37 °C overnight. The 500 µl starting culture was then inoculated to 50 ml of LB broth and the 50 ml culture was cultured at 37 °C with vigorous shaking to an OD_{600} of 0.5 to 0.8. Cells were briefly chilled on ice for 10 min, and recovered by centrifugation at 2700g for 10 min at 4 °C. The pellets were resuspened in 30 ml of ice-cold $MgCl_2/CaCl_2$ solution (80 mM $MgCl_2$ and 20 mM $CaCl_2$) and centrifuged as above. After resuspended in 2 ml of ice-cold 0.1 M $CaCl_2$, The cell suspension was divided to 200-µl aliquots and stored at -80 °C until required.

2.7.3.2 Transformation

The prepared component cells were thawed on ice for 5 min. Two to four microlitres of the ligation products were added and gently mixed by pipetting and left on ice for 30 min. The samples were heat-shocked in a 42 °C water bath for exactly 45 s and then immediately placed on ice for 2-3 min. Next, 1 ml of prewarmed SOC medium was added (2% Bactotryptone, 0.5% Bactoyeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose), and the samples were incubated at 37 °C with shaking for 90 min. The grown cells were collected by centrifuging for 20 s at room temperature, then gently resuspended in 100 µl of SOC medium and spreaded onto a selective LB agar plates containing 50 µg/ml of ampicillin, 25 µg/ml of IPTG and 20 µg/ml of X-gal. The spreaded plates were incubated at 37 °C overnight (Sambrook and Russell, 2001).

2.7.4 Detection of recombinant clones by colony PCR

Recombinant clones were selected by a lacZ' system following standard protocols (Sambrook and Russel, 2001). Namely, only white colonies containing the inserted DNA were selected. Colony PCR was performed to identify the insert sizes of positive clones.

Colony PCRs were performed in 25-ul reactions containing 10 mM Tris-HCl (pH 8.8), 50 mM KCl, 0.1% Triton X – 100, 100 mM of each dNTP, 2 mM MgCl$_2$, 0.1 µM each of pUCl (5'-CCG GCT CGT ATG TTG TGT GGA-3') and pUC2 (5'-GTG GTG CAA GGC GAT TAA GTT GG-3'), 0.5 unit of Dynazyme™ DNA Polymerase (FINNZYMES). Selected colonies were individually picked by pipette tips and resuspended in the reaction mixtures. Condition for the thermal cycle was predenaturation at 94 °C for 3 min followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 50 °C for 1 min and extension at 72 °C for 2 min. The final extension was carried out at the same temperature for 7 min. The colony PCR products were electrophorized through 1.2% agarose gels and visualized after ethidium bromide staining.

2.7.5 Digestion of the amplified DNA insert

Colony PCR products of clones containing corresponded sizes of the inserts were digested with *Hin*dIII or *Rsa*I to verify whether the ligated PCR product possibly contained one type of sequence. The digestions were set up in a total volume of 15 µl containing an appropriate restriction enzyme buffer (buffer E; 6 mM Tris-HCl; pH 7.5, 6 mM MgCl$_2$, 100 mM NaCl and 1 mM DTT for *Hin*d III or buffer C; 10 mM Tris-HCl; pH 7.9, 10 mM MgCl$_2$, 50 mM NaCl and 1 mM DTT for *Rsa*I), 5 µl of colony PCR products and 2 units of either *Hin*dIII or *Rsa*I. The reaction mixtures were incubated at 37°C for 3 – 4 hours. Patterns of the digestion products were analyzed by agarose gel electrophoresis.

2.7.6 Extraction of recombinant plasmids

DNA plasmids were extracted from the *E. coli* host using HiYield™ Plasmid Mini Kit (Real Biotech Corporation). The extraction was carried out according to the manufacturer's protocol. In short, *E. coli* cells containing the transformed plasmids were inoculated into 3 ml of LB broth (1% tryptone, 0.5% yeast extract, 1.0 % NaCl) containing 50 µg/ml of ampicillin and incubated at 37 °C with vigorous shaking overnight. Cells were harvested from 1.5 ml of the overnight cultures by centrifuging at 19,000 g for 1 min. The cell pellets were collected and resuspended with 200 µl of the PDl buffer containing RNaseA. The resuspended cells were lysed by gently mixing with 200 µl of the PD2 buffer. The samples were left for 2 min at room temperature. After that, 300 µl of the PD3 buffer were added to neutralize the alkaline lysis step and mixed immediately by inverting the tube for 10 times. To separate the cell debris, the samples were centrifuged at 19,000 g for 15 min. The supernatants were transferred to the PD columns and centrifuged at 6,000g for 1 min. The flow-through was discarded. The columns were washed with 400 µl of the W1 buffer and centrifuged at 6,000g for 1 min. The columns were washed one more time with 600 µl of the ethanol-added Wash buffer and subjected to a final spin for a 2 min at 19,000 g to remove the residual Wash buffer. Plasmids were eluted from the columns by adding 30-50 µl of the Elution buffer. The columns were then left at room temperature for 2 min and centrifuged at 19,000 g for 2 min. Concentrations of the extracted plasmid were spectrophotometrically measured.

2.7.7 DNA sequencing and sequence identification

Plasmid clones were unidirectional sequenced using the M 13 forward and reverse primer on an automatic sequencer at Macrogen (Korea). Assembled nucleotide sequences from the sequencing results were blasted against data in GenBank using Blast*N* (nucleotide similarity) and Blast*X* (translated protein similarity). Significant similarity was considered when the probability (E) value was $< 10^{-4}$.

2.8 Isolation and characterization of full length cDNA by RACE-PCR

2.8.1 Preparation of the 5' and 3' RACE templates

Full length cDNAs of gene homologues were characterized using a SMART RACE cDNA Amplification Kit (Clontech). Template for RACE-PCR (RACE-Ready cDNA) was prepared from messenger (m) RNA purified from total RNA extracted from broodstock ovaries, using a QuickPrep *micro* mRNA Purification Kit (Amersham Phamacia Biotech). One μg of the purified mRNA was then combined with 2 μM of 5'-CDS primer (for 5'-RACE-Ready cDNA) or 3'-CDS primer A (for 3'-RACE-Ready cDNA), 1 μl of 10 μm SMART II A oligonucleotide (only for 5'-RACE-Ready cDNA), and nuclease-free H_2O in an amount that made the final reaction volume to 5 μl. The reaction was incubated at 70 °C for 2 min and immediately cooled down on ice for 2 min. After that, 2 μl of 5x First-Strand buffer, 1 μl of 20 mM DTT, 1 μl of dNTP Mix (10 mM) and 1 μl of PowerScript Reverse Transcriptase were added. The reactions were mixed by gently pipetting and centrifuged briefly before incubated at 42 °C for 1.5 hr in a thermocycler. The first strand reaction products were diluted with 250 μl of TE buffer and then heated at 72 °C for 7 min.

2.8.2 RACE-PCR

Gene-specific primers (GSPs) were designed from their corresponding homologue sequences in the *P. monodon* EST libraries. GSPs used in the 5'- and 3'-RACE PCRs used to amplify fragments of each gene are summarized in Table 2.4. Reaction mixtures for 5'- and 3'-RACE-PCR were prepared according to Tables 2.5 and 2.6, respectively. Two negative controls containing either Universal Primer A Mix (UPM) or GSP1 primer were also performed to confirm the absence of nonspecific amplified products.

Table 2.4 Gene-specific primers (GSPs) used for futher characterization of the full length cDNA of functionally important gene homologues in *P. monodon*

Gene	Name of primer	Primer sequence	Tm (ºC)
Calreticulin			
3´RACE	**CRT-F**	5´ACATTGACTGTGGTGGAGGAT3´	68
Calnexin			
5´RACE	**CNX-R**	5´TATTACCCACATTCGTCACTACATTAT3´	56
3´RACE	**CNX-F**	5´TGATATTTGCCTGTTGTTGTGC3´	57
Nested primer			
Nested 5´RACE	**5'NCNX-R**	5´CATCATCCACAGCCATTTAATCTCTCCG3´	76
Nested 3´RACE	**3'NCNX-F**	5´GCAGATGATGATGCTCCTGGAAGTGG3´	77
ERp57			
5´RACE	**ERp57-R**	5´CAGCAAGTTTTCACCTGACAAGCCTCT3´	76
3´RACE	**ERp57-F**	5´TGAACTGAATGGGTATGACCGCAAG3´	74

Table 2.5 Compositions for amplification of the 5′ end of gene homologues using 5′ RACE-PCR

Component	5′ RACE-PCR	UPM only (Control)	GSP1 only (Control)
5' RACE–Ready cDNA template	1.5 µl	1.5 µl	1.5 µl
UPM (10x)	5.0 µl	5.0 µl	-
GSP1 (10 uM)	1.0 µl	-	1.0 µl
Nuclease-free H$_2$O	-	1.0 µl	5.0 µl
Final volume	25 µl	25 µl	25 µl

Table 2.6 Compositions for amplification of the 3′ end of gene homologues using 3′ RACE-PCR

Component	3′ RACE-PCR	UPM only (Control)	GSP1 only (Control)
3' RACE–Ready cDNA template	1.5 µl	1.5 µl	1.5 µl
UPM (10x)	5.0 µl	5.0 µl	-
GSP1 (10 uM)	1.0 µl	-	1.0 µl
Nuclease-free H$_2$O	-	1.0 µl	5.0 µl
Final volume	25 µl	25 µl	25 µl

To amplify 5' RACE-PCR fragments of genes ERp57, a touchdown thermal cycle was performed i.e. 5 cycles of 94ºC for 30 s, and 72ºC for 3 min; 5 cycles of 94ºC for 30 s, 70ºC for 1 min and 72ºC for 3 min; 25 cycles of 94ºC for 30 s, 68ºC for 1 min and 72ºC for 3 min. The final extension was carried out at 72ºC for 7 min. For amplification of 3'RACE-PCR fragment of genes CRT, CNX and ERp57 and 5' RACE-PCR fragment of CNX, the thermal cycle was 25 cycles of 94ºC for 30 s, 65-68ºC for 1 min and 72ºC for 3 min. The final extension was carried out at 72ºC for 7 min.

Nested PCRs were subsequently performed for amplification of 5' and 3' RACE fragments of gene CNX when the primary RACE products contained high levels of nonspecific background. Nested Universal Primer A (NUP) and GSP2 primers, which are specific within the primary target product, were used in the nested reaction. 5'NCNX and 3'NCNX primers used for amplification of 5' and 3' RACE fragments of gene CNX, respectively (Table 2.4). Components in the nested reaction mixtures are summarized in Table 2.7. Finally, individual interested product bands were then cloned and sequenced as described above.

2.8.3 Phylogenetic analysis of CRT, CNX and ERp57

Protein sequences of interested genes from various species were retrieved from GenBank and compared with those of *P. monodon*. Multiple alignments were carried out using ClustalX (Thompson, 1994). Sequences were bootstrapped 500 times using a seqboot. The divergence between pairs of protein sequences was estimated using Prodist. A bootstrapped neighbor-joining tree was constructed to illustrate phylogenetic relationships among sequences using Neighbor and Consense. All phylogenetic programs described were routine in PHYLIP (Felsenstein, 1993).

Table 2.7 Compositions for amplification of nested PCR

Component	Nested PCR	NUP only (Control)	GSP2 only (Control)
50 time Diluted primary RACE PCR product	1.5 μl	1.5 μl	1.5 μl
NUP (10x)	5.0 μl	5.0 μl	-
GSP2 (10 μM)	1.0 μl	-	1.0 μl
Nuclease-free H_2O	-	1.0 μl	5.0 μl
Final volume	25 μl	25 μl	25 μl

2.9 Isolation and characterization of genomic sequences

Genomic organization of the interested genes was characterized by both PCR of genomic DNA and Genome walking PCR.

2.9.1 PCR amplification of genomic DNA

Genomic DNA template was prepared as mentioned previously. Primers used to amplify 2 fragments of genes CRT are ORFCRT-F with CRT-R and CRT-F with ORFCRT-R, respectively (Table 2.8). The reactions were carried out in 25-μl mixtures, containing 12.5 μl of PCR-grade water, 1x Mg^{2+}-free EXT buffer, 50 μM $MgCl_2$, 1 mM each dNTP, 2. μM each of the appropriate primers, 50 ng genomic DNA template and 0.5 μl of 50 x DyNAzyme^{TM} Ext taq polymerase (Finnzyme). The thermal cycle condition was predenaturation at 94°C for 3 min followed by 35 cycles of a 94°C denaturation step for 45 sec, a 53 °C annealing step for 1 min and a 72 °C extension step for 3 min. The final extension was carried out at 72 °C for 7 minutes. Individual interested product bands were then cloned and sequenced as described above.

2.9.2 Genome walking

2.9.2.1 Preparation of templates for Genome walking

2.9.2.1.1 Restriction digestion of genomic DNA

Two point five microgrammes of genomic DNA extracted from an individual of *P. monodon* as described above were singly digested with 40 units of a blunt-end generating restriction enzyme (*Rsa* I, *Dra* I, *Eco* RV, *Pvu* II, *Ssp* I or *Stu* I). The reaction also contained 1 x of an appropriate restriction enzyme buffer and nuclease-free H$_2$O to make the final volume to 100 µl. The digestion was carried out at 37ºC for 4 hours. Five microlitres of the digestion product was then electrophorized on 0.8% agarose gels to confirm the digestion.

2.9.2.1.2 Purification of the restriction-digested DNA

The digested DNA was purified by a phenol-chloroform-proteinase K method (Klinbunga et al., 1999). Namely, an equal sample volume of buffer-equilibrated phenol was added to the sample. The mixture was vortexed for 5-10 seconds and centrifuged for 5 minutes at room temperature to separate the aqueous and organic phases. The upper layer was collected and mixed with a sample volume of chloroform:isoamylalcohol (24:1). The sample was then centrifuged as in the previous step. The upper layer was collected and mixed with one-tenth sample volume of 3 M NaOAc (pH 4.5) sequentially with 2.5 x sample volume of -20 ºC-cold absolute ethanol. The mixture was incubated at -80 ºC for 30 min before DNA pellet was collected by centrifugation at 16,300 g for 10 minutes at room temperature. After a brief wash with ice-cols 70% ethanol, the DNA pellet was air-dried and resuspended in 10 µl of TE buffer. The purified products were confirmed by electrophoresis on 0.8 % agarose gels.

2.9.2.1.3 Ligation of the purified digested DNA to GenomeWalker adaptors

Ligation was performed in a 10 µl reaction volume containing 2.7 µl of the digested DNA, 4.75 µM of GenomeWalker Adaptor (Clontech), 1 x ligation buffer , 5 % polyethylene glycol (PEG) and 3 units of T4 DNA ligase. The reaction was carried out at 16 ºC overnight and then terminated by incubation at 70 ºC for 5 minutes. The

ligated product was ten fold diluted in TE buffer before used as template in the PCR step.

2.9.2.2 Genome walking PCR

Genome walking PCR consisted of 2 sequential steps: primary and secondary (nested) PCR. The primary PCRs were carried out in 25 µl reactions, each of which contained 10 mM Tris-HCl (pH 8.8), 50 mM KCl, 0.1% Triton X-100, 200 µM each of dATP, dCTP, dGTP and dTTP, 2 mM $MgCl_2$, 0.2 µM each of Adaptor primer1 (AP1: 5´-GTA ATA CGA CTC ACT ATA GGG C-3´) and an appropriate gene specific primer (Table 2.6), 1 µl of 10 x diluted purified digested DNA and 1.0 unit of DyNAzyme^{TM} Ext Taq DNA Polymerase (Finnzymes). The amplification reaction was carried out using a two-step cycle including 7 cycles of a denaturing step at 94 ℃ for 25 seconds and an annealing/extension step at 70 ℃ for 3 minutes followed by 35 cycles of 94 ℃ for 25 second, annealing at 65 ℃ for 3 minutes. The final extension was carried out at 67 ℃ for 7 minutes. The primary PCR products were 50 fold diluted in nuclease-free H_2O.

For the secondary PCRs, 1 µl of the diluted primary product was used as template. Components of the secondary PCR mixture were as the same as those of the primary PCR except for replacements of AP1 with AP2 (5´-ACT ATA GGG CAC GCG TGG T-3´) and GSP1 with GSP2 of the same concentration. GSP2 used for amplification of each product was shown in Table 2.9. PCR was carried out under the following cycle: 5 cycles of a denaturing step at 94 ℃ for 25 seconds and an annealing/extension step at 70 ℃ for 3 minutes followed by 20 cycles of 94 ℃ for 25 second and 65 ℃ for 3 minutes. The final extension at 67 ℃ was carried out for 7 minutes. Five microlitres of the secondary PCR products were electrophoretically analyzed on 1.2% agarose gels.

Finally, individual interested product bands were then cloned and sequenced as described above.

Table 2.8 Sequences and Tm of primers for PCR amplification of genomic *CRT*.

No. of pair primers	Name of sequence	Primer sequence	Tm (ºC)
1	*ORFCRT-F*	5´CAAGAGCGAAGACGGAACGATG3´	62
	CRT-R	5´TCGTAGGTGTTGTCAGGATTG3´	58
2	*CRT-F*	5´ACATTGACTGTGGTGGAGGAT3´	58
	ORFCRT-R	5´GCATCTTTAGTAGCACCCCAAGTCTC3´	60

2.10 Studies of effect of heat stress on *CRT*, *CNX* and *ERp57* expression

2.10.1 Sample preparation

P. monodon juvenile (4 month-old with the body weight estimate 20 g) were used in this experiment. They were acclimatized at the laboratory conditions (ambient temperature of 28 – 30 °C, salinity of 20 ppt) for one week. The experimental animals were fasted approximately 24 hours prior to the temperature treatment. Six shrimps were collected as a control group before the rest were treated at 35 ºC for 3 hours. After the high temperature treatment, they were transferred back to the ambient temperature. Six shrimps were collected at 0, 6, 12, 24 and 36 hours after the treatment. Haemocyte, gill and hepatopancreas were immediately dissected from the harvested shrimps, quick-frozen in liquid nitrogen and stored at –80 ºC until required. Total RNA was extracted from the collected tissues and treated with DNase as previously described. The first strand cDNA was then synthesized from the prepared RNA.

Table 2.9 Sequences and Tm of gene specific primers (GSP1) and nested gene specific primers (GSP2) for genome walking analyses

no.	Gene	Name of primer	Primer sequence	Tm (ºC)
	CRT			
	GSP1			
1		1ˢᵗ Promoter CRT GW	5´TTGTTTTGGTCCGACGTAAGAGCGGAG3´	68
	GSP2			
2		2ⁿᵈ Promoter CRT GW	5´TCCTTACACCAGCCAGGGCCGTGACCA3´	69
	CNX			
	GSP1			
3		ORFCNX-F	5´GAAGCGTGAGTCGTCCATTGACTAAA3´	62
4		ORFCNX-R	5´TTAATCTCTCCGGGACTTGCGCAGC3´	65
5		1st 5' cont GW-CNX	5´TCTGTGCTTTTATTGGTCAGCAGGGTC3´	68
6		1ˢᵗ 3'cont GW-CNX	5´TGTACTTACACGCCATGCCCCCAGTGA3´	66
	GSP2			
7		2nd 5' cont GW-CNX	5´AACCATATGTAGTGGATTCTGGGCCTTG3´	73
8		2nd 3' cont GW-CNX	5´GGCTCTACACGTACTACCTGTCTCACCTG3´	67
	ERp57			
	GSP1			
9		ORF ERp57-F	5´AATGGCTACGAGATTGTTAATACTACTCC3´	68
10		ORF ERp57-R	5´CACTCTTTACACAACACACTGCGGTGA3´	66
11		1ˢᵗ 5'cont GW-ERp57	5´GGCAGCTGTTGGGTTGTGGGCTTTCGTTTG3´	80
12		1ˢᵗ 3'cont GW-ERp57	5´CACAAATTACTGGCTTGTCACCAGGAA3´	68
13		5'1st next ERp57GW	5´TAGCAAGGTAGACGGGAGGGTCGTTG3´	68
14		3'1ˢᵗ next ERp57GW	5´CTCTACCACCTTGAAGGCCAACGACC3´	68
15		last ERp57GW R 1st	5´TCCTTAGAGGCTGGTGGTCCAACCTGTGA3´	66
	GSP2			
16		2nd 5'cont GW-ERp57	5´ACCAGACTGTCAAAGCTGTAATGTGTGAAT 3´	64
17		2nd 3'cont GW-ERp57	5´TGCTGGAAGTCGTCCTTATTGGCAACG3´	66
18		5'2nd next ERp57GW	5´CGGGAGGGTCGTTGGCCTTCAAGGTG3´	71
19		3'2nd next ERp57GW	5´TTGAAGGCCAACGACCCTCCCGTCTAC3´	69
20		last ERp57GW R 2nd	5´CTTCAGGGTAGGGTAGCCAGACACAC3´	68

2.10.2 Semi-quantitative Reverse Transcription-PCR(RT-PCR)

Expression levels of *CRT*, *CNX* and *ERp57* were semi-quatitatively examined, using primers shown in Table 2.10. Elongation factor 1 alpha (EF1-α) was used as an internal control. The amplification conditions were optimized as follow.
'

PCRs were performed in 25-µl reactions containing 0.1 µg of the first strand cDNA template, 1X PCR buffer (10mM Tris-HCl pH 8.8, 50 mM KCl and 0.1% Triton X-100), 200 µM each of dNTP and 1 unit of Dynazyme™ DNA Polymerase (FINZYMES, Finland). Concentrations of each primer and $MgCl_2$ ranged from 0.05 - 0.25 µM and 1 - 2.5 mM, respectively. For optimization of primer and $MgCl_2$, the number of cycles were showed in Table 2.11.

After RT-PCR, 5 microlitres of the PCR products were combined with 2 µl of the loading dye (5 g Ficoll type 400, 0.05 g Bromophenol blue adjust volume to 20 ml with water) before loaded to 1.8% agarose gel and electrophorized at 5-6 volt/cm. The gel was stained with 2.5 µg/ml Ethidium bromide (EtBr) for 5 min and destained in the TE buffer for 15 min. Intensity of target and control bands was quantified from photographs of the gels using the Gel Pro program.

The determined band intensity value of interested genes was normalized by that of *EF-1α*. The normalized data were then tested using one way analysis of variance (ANOVA) followed by Duncan multiple range test ($P < 0.05$).

Table 2.10 Nucleotide sequences of primers used for semi- and real-time quantitative PCR analysis of *CRT*, *CNX* and *ERp57* in *P. monodon*

No	Gene	Name of primer	Sequence	Tm(ºC)	Expected size (bp)
	CRT				
1		Real-timeCRT F	5´CATACCTCATTATGTTTGGTCCTG3´	68	
2		RealtimeCRT-Rnew	5´TTCTGGGTCCTTGATCTTCTTG3´	69	257
	CNX				
3		Real-timeCNX F	5´CACGGAAGAGCAAACAGAAGGAG3´	72	
4		Real-timeCNX R	5´TAAGAGCCAATCGTTCAGCAGGT3´	69	221
	ERp57				
5		Real-timeERp57 F	5´CTGTCACTGTTGCTGTGGGTAAG3´	71	
6		Real-timeERp57 R	5´ACATCTTCATCCTTCATCGCTTC3´	69	151
	EF-1α				
	Semi				
7		EF1-α F	5´ATGGTTGTCAACTTTGCCCC3´	58	
8		EF1-α R	5´TTGACCTCCTTGATCACACC3´	58	500
	Real-time				
9		EF1-α$_{214 F}$	5'GTCTTCCCCTTCAGGACGTC3'	58	
10		EF1-α$_{214 R}$	5'CTTTACAGACACGTTCTTCACGTTG3'	58	214

Table 2.11 Amplification condition for interesting gene expression level analysis in thermal stress shrimp using semi quantitative

Amplification condition	Gene homologue											
	CRT			CNX			ERp57			EF-1α		
	Temp.(℃)	Time	Number of cycles	Temp.(℃)	Time	Number of cycles	Temp.(℃)	Time	Number of cycles	Temp.(℃)	Time	Number of cycles
Initial denaturation	94	3 min	1	94	3 min	1	94	3 min	1	94	3 min	1
Denaturation	94	30s		94	30s		94	30s		94	30s	
Annealing	65	30s	26-35	69	45s	26-35	65	30s	26-35	53	45s	21
Extension	72	30s		72	45s		72	30s		72	45s	
Final extension	72	7min	1	72	7min	1	72	7min	1	72	7min	1

2.10.3 Quantitative real-time PCR

Expression levels of *CRT*, *CNX* and *ERp57* were examined by quantitative Real-time PCR, using primers shown in Table 2.10. *EF1-α* was used as an internal control. The primers used for amplification *EF1-α* (*EF1-α*$_{214}$) were showed in table 2.10.

2.10.3.1 Construction of standard curves

Standard curve of each gene was constructed using the according ORF-pGEMT construct prepared in section 2.13.1 as the templates. The amount of the templates in each reaction ranged from 10^3-10^8 copy numbers, which were determined from the following equation:

$$1 \text{ kb} = 6.6 \times 10^5 \text{ dalton; } 1 \text{ μg of 1 kb cDNA contains } 0.91 \times 10^{12} \text{ molecules}$$

The amplifications were performed in a reaction volume of 10 μl containing LightCycler 480 SYBR Green I Master (Roche). Concentration of each primer was 0.2 μM (except for the *CNX* amplification, in which 0.3 μM primers were used). The thermal profile was 95 °C for 5 min followed by 40 cycles of 95 °C for 10 s, 68 °C for 30 s and a final extension at 72 °C for 30 s. Subsequently, cycles for the melting curve analysis was carried out at 95 °C for 15 s, 60 °C for 1 min and at 95 °C for 15 s. Real-time RT-PCR assay was carried out in a 96 well plate and each sample was run in duplicate.

2.10.3.2 Quantitative real-time PCR

The amplifications were performed in a reaction volume of 10 μl composed of the fluorescence master mix and primers as described in 2.10.3.1 Each reaction contain 11.25 (for *EF-1α*) or 25 (for *CRT, CNX* and *ERp57*) ng of the first strand cDNA template. The thermal profile was the same as that described in the previous section. Real-time RT-PCR assay was carried out in a 96 well plate and each sample was run in duplicate. Relative expression levels of different groups of samples were statistically tested by ANOVA followed by Duncan's new multiple range test or Tukey test (*P < 0.05*).

2.11 *In vitro* expression of recombinant proteins using a bacterial expression system

2.11.1 Cloning of ORF into a cloning vector (pGEM-T easy)

ORF of the desired genes were amplified from 1st-stranded cDNA derived from ovaries. Primers for the amplifications were designed according to the derived full-length cDNA obtained from above and are shown in Table 2.12.
The amplified ORF were ligated and cloned into pGEM-T easy vector. The ligation products were transformed into *E. coli* JM109. Plasmid DNA of the positive clones was sequenced to confirm their sequences and orientation of the inserts. The ORF-pGEM-T constructs were then used as templates for amplification of the inserts to be cloned into the expression vectors in the next step.

2.11.2 Cloning of recombinant expression plasmids

Primers were designed to include an appropriate restriction site before and after the start and stop codons, respectively (Table 2.13). For mature CRT, CNX and ERp57, the first 1,173, 1359 and 1410 bp, respectively, which are the signal sequences were excluded.

The amplified products were digested with restriction enzymes that were specific to the introduced restriction sites. The digested products were then analyzed by agarose gel electrophoresis and eluted from the gel. The purified digested ERp57 product was ligated into pET15b (Invitrogen) while the others were ligated into pGEX 4T (GE Healthcare). The ligation products were transformed into *E. coli* JM109. The plasmid constructs were purified from the transformed *E. coli* JM109 cultures and sequenced. Plasmid of the positive clones was subsequently transformed into an expression host, *E. coli* BL21 (DE3) codon+ RIPL (Stratagene).

Table 2.12 Nucleotide sequences of primers used for ORF amplification of *CRT*, *CNX* and *ERp57* in *P. monodon*

No	Gene	Name of primer	Sequence	Tm(ºC)	Expected size (bp)
	CRT				
1		ORFCRT-F	5´CAAGAGCGAAGACGGAACGATG3´	62	
2		ORFCRT-R	5´GCATCTTTAGTAGCACCCCAAGTCTC3´	60	1221
	CNX				
3		ORFCNX-F	5´GAAGCGTGAGTCGTCCATTGACTAAA3´	62	
4		ORFCNX-R	5´TTAATCTCTCCGGGACTTGCGCAGC3´	65	1788
	ERp57				
5		ORF ERp57-F	5´AATGGCTACGAGATTGTTAATACTACTCC3´	68	
6		ORF ERp57-R	5´CACTCTTTACACAACACACTGCGGTGA3´	66	1458

Table 2.13 Nucleotide sequences of primers used for cloning of the expression plasmids of recombinant CRT, CNX and ERp57.

.

no.	Gene	Name of primer	Restriction site	Primer sequence
	CRT			
1		w/o CRT-GST_F	*Bam* HI	5´TTT<u>GGATCC</u>ATGAAAGTATTTTTCGAAGAAAGAT3´
2		all CRT-GST_R	*Eco* RI	5´TTT<u>GAATTC</u>TTACAGCTCGTCATGTTCAAG3´
	CNX			
3		CNX-GST_F	*Bam* HI	5´TTT<u>GGATCC</u>ATGGATGACGATGACGATGAAGAA3´
4		CNX-GST_R	*Eco* RI	5´TTT<u>GAATTC</u>TTATGGATTCTTATTGGAGTAGTTG3´
	ERp57			
5		ExpERp57-F	*Nde* I	5'TTT<u>CATATG</u>GGAGACGATGTCCTGCAATTA3'
6		ExpERp57-R	*Bam* HI	5'TTT<u>GGATCC</u>TCATCAAAGTTCAGTCTTCTTGCC3'

Note: For amplification of the two CRT forms, the same reverse primer was used

2.11.3 Expression of recombinant proteins

A single colony of recombinant *E. coli* BL21 (DE3) codon+ RIPL carrying desired recombinant plasmid was selected for the expression. Cells were grown in LB medium, containing 50 μg/ml ampicillin and 34 μg/ml chloramphenicol at 37 °C with shaking. Protein expression was induced by 1 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) at an OD_{600} of 0.4-0.6. For the time course studies, samples were time-interval taken (at 1, 2, 3, 4, 5 and 14 hr) and centrifuged at 12000 g for 1 min. The pellets were resuspended with PBS and 2 x SDS-PAGE sample loading buffer before examined by 15 % SDS-PAGE (table 2.14) (Laemmli, 1970).

2.11.4 Western blotting

Recombinant proteins were analyzed by 15% SDS-PAGE. The electrophoresed proteins were transferred to PVDF membrane (Hybind P; GE Healthcare) (Towbin, 1979). The blotted membrane was washed three times with 1X TBST (0.5 % Tween20) for 5 min, blocked with a blocking buffer (5% BSA in 1X TBST) and incubated overnight at room temperature with gentle shaking. The membrane was washed three times in 1xTBST and incubated with 1:5,000 dilution of an appropriate primary antibody (Anti-His (GE Healthcare) for ERp57 and Anti-GST (Biorad) for the rest) in the blocking buffer for 1 hr. After that, the previous wash process was repeated before the membrane was incubated with Anti-mouse-IgG-AP Conjugate (Promega; 1:10,000 dilution in the blocking buffer) for 1 hr. The alkaline phosphatase activity was detected by BCIP/NBT (Promega). The membrane was incubated in a dark place for 2-15 min.

2.11.5 Electrospray ionization mass spectrometric (ESI-LC MS/MS) analysis of recombinant proteins

The recombinant proteins were resolved by 15% SDS-PAGE and the protein bands at the expected molecular mass were excised from the gel. In-gel trypsin digestion was performed as described elsewhere (Shevchenko et al., 1996). The digestion products were injected with a flow rate of 800 nl/min into a 100×0.18 mm, 5 μm BioBasic C18 Kappa column (Thermo Electron) connected to a linear ion trap mass spectrometer (Bruku). Peptides were eluted from the column by a concentration gradient of acetonitrile from 2-65 % in 0.1% formic acid for over 40 min. The peptide mass spectra were measured by the connected mass spectrometer equipped. Using MASCOT, the results were searched against a non-redundant *P. mondon* translated protein database derived from an EST library database (http://161.200.123.190/home).

2.11.6 Protein purification

2.11.6.1 Protein sample preparation

Protein samples were obtained from 300 ml of IPTG-induced cultures prepared as previously described. Cells were harvested by centrifugation at 6,790 g at 4 °C for 15 min. The pellet was resuspended in the binding buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4). Samples were incubated on ice with 2 mg/ml lysozyme for 30 min before sonicated. After that, DNA was removed by an on-ice incubation with 0.03 mg/ml DNase supplemented with 1 mM $MgCl_2$. To separate the soluble and insoluble fractions, samples were centrifuged at 19,000 g for 30 min. Fraction containing the target expressed protein was selected to subject to the next purification process.

2.11.6.2 His-Tag/ Ni affinity system

Recombinant ERp57 tagged with 6 histidine residues was subjected to the Ni purification system using a His GraviTrap column (GE Healthcare), containing Ni ion bound to the beads. Before sample loading, the column was equilibrated with 10 x column volumes of the binding buffer. Sample prepared from the previous step was then applied to the column twice. After that, the column was washed with 10 x column volumes of the binding buffer, followed by 5-10 x column volumes of the washing buffer (20 mM sodium phosphate, 500 mM NaCl, 80 mM imidazole, pH 7.4). Target protein was finally eluted with 6 ml of the elution buffer (20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.4). The entire purification process was performed at 4 °C and the purified sample was stored at 4 °C until use.

2.11.6.3 GST-Tag/ glutathione system

Recombinant GST tagged CRT and CNX proteins were purified using a GSTrap FF column (GE Healthcare). The column was pre-equilibrated with 10 x column volumes of binding buffer (2 M NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, pH 7.3). Samples prepared from the previous step were then injected into the column. The column was washed with 5-10 column volumes of the binding buffer Target proteins were eluted with 5-10 column volumes of elution buffer (50 mM Tris-HCl, 10 mM reduced glutathione, pH 8.0). The entire purification process was performed at 4 °C and the purified sample was stored at 4 °C until needed.

2.12 *In vitro* characterization of protein activities

2.12.1 Calcium binding ability

One microgram of the purified protein was added to 1 mM of either $CaCl_2$ or EDTA and mixedThe mixtures were incubated on ice for 30 minbefore resolved through 12% SDS-PAGE gel or 10% native-PAGE gel (table 2.14). The results were detected by Coomassie blue staining.

2.12.2 Assay of binding activity of CRT and CNX to ERp57

To examine the binding activity of CRT and CNX to ERp57, 200 μg of either CRT or CNX were mixed with 200 μg of ERp57 and incubated on ice for 30 min. The mixtures were then subjected to 10% native-PAGE gel and detected by Coomassie blue staining. The shifted bands were excised and analyzed LC-MS/MS to confirm the binding.

Table 2.14 Solutions for Tris/Glycine Native and SDS-Polyacrylamide Gel Electrophoresis

Native-PAGE				SDS-PAGE					
Resolving gel		**Stacking gel**		**Resolving gel**			**Stacking gel**		
% gel	10%			% gel	12%	15%			
no. of gels	2	no. of gels	2	no. of gels	2	2			2
H$_2$O	4	H$_2$O	1.4	H$_2$O	3.89	3.22	H$_2$O		2.5
30% Acrylamide mix	3.3	30% Acrylamide mix	0.33	40% Acrylamide:Bis 37.5:1	2.7	3.376	40% Acrylamide:Bis 37.5:1		0.35
1.5 M Tris (pH 8.8)	2.5	1.0 M Tris (pH 6.8)	0.25	4xLower*	2.25	2.25	4xUpper**		1
10% APS	0.1	10% APS	0.02	10%APS	0.05	0.05	10%APS		0.05
TEMED	0.004	TEMED	0.002	10%SDS	0.09	0.09	10%SDS		0.04
				TEMED	0.02	0.02	TEMED		0.01
Total (ml)	10	Total (ml)	2	Total (ml)	9	9.006	Total (ml)		3.95

* 4xLower buffer : 1.5 M Tris, 0.4% SDS, pH 8.8

** 4xUpper buffer : 0.5 M Tris, 0.4% SDS, pH 6.8

CHAPTER III

RESULTS

3.1 Genomic DNA extraction

Genomic DNA was extracted from a piece of pleopod of *P. monodon* broodstock using a phenol-chloroform method. The quality of extracted genomic DNA was electrophoretically determined. High molecular weight DNA at the similar size as that of undigested λDNA (approximately 50 kb) along with slightly sheared DNA was obtained (Figure 3.1).

The ratio of OD_{260}/OD_{280} of extracted DNA ranged from 1.8-2.0 indicating that the acceptable quality of extracted DNA. Some DNA samples were contaminated with the residual RNA as visualized by the discrete band at the bottom of gel and exhibited a ratio of $OD_{260}/OD_{280} \geq 2.0$. However, contamination of RNA does not affect the use of the extracted genomic DNA for PCR amplification.
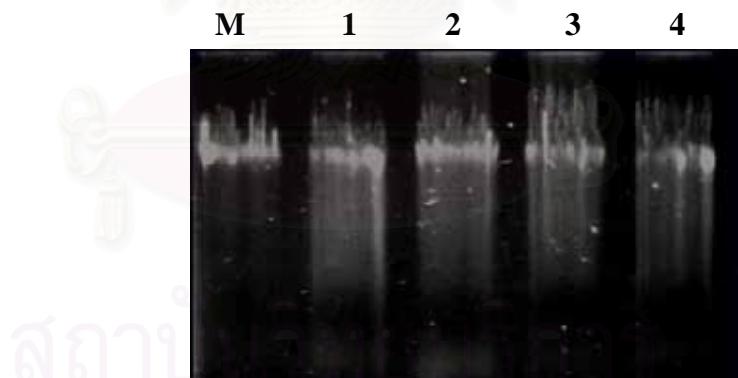
Figure 3.1 A 0.8% ethidium bromide-stained agarose gel showing the quality of genomic DNA (lanes 1 - 4) extracted from pleopods of *P. monodon* individuals. Lane M = undigested λ DNA (200 ng).

3.2 Total RNA extraction and first stand cDNA synthesis

Total RNA was extracted from various tissues of both juvenile and broodstock of *P. monodon* using TRI REAGENT. The quality of total RNA was electrophoretically determined using 1% agarose gel. Predominant discrete bands

were observed along with smeared high molecular weigth RNA (Figure 3.2). The ratios of OD_{260}/OD_{280} of purified RNA were 1.8 - 2.0. The first stand cDNA synthesized from the obtained total RNA covered the large sized products indicating the acceptable quality of the synthesized first strand cDNA (Figure 3.3)
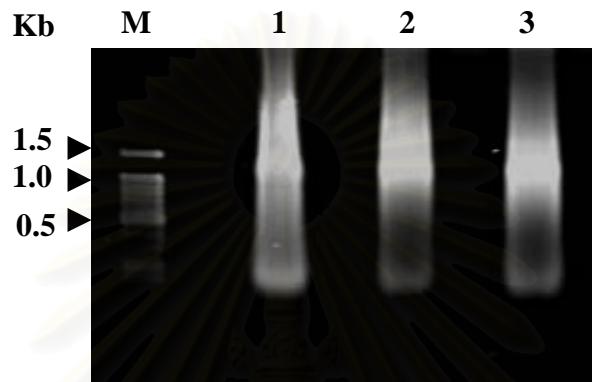
Figure 3.2 A 1.0% ethidium bromide-stained agarose gel showing the quality of total RNA extracted from various tissues of *P. monodon*. Lane M = 100 bp ladder. Lanes 1 - 3 = Total RNA from thoracic ganglia, hemocytes and lymphoid organ of *P. monodon* broodstock, respectively.
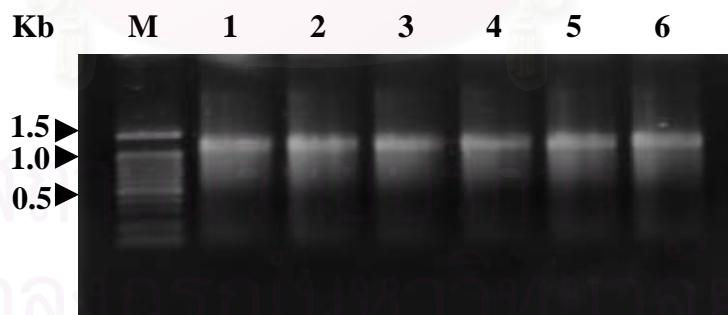
Figure 3.3 A 1.0% ethidium bromide-stained agarose gel showing the synthesized first strand cDNA from total RNA of *P. monodon* broodstock. Lane M = λ-*Hin*d III DNA marrker. Lane 1 - 6 = the first strand cDNA from intestine, ovary, testis, thoracic ganglia, heamocyte and lymphoid organ, respectively**.**

3.3 ESTs

Partial nucleotide sequences of *calreticulin* (*CRT*, clone number OV-N-ST01-0117-W, Figure 3.4A), *calnexin* (*CNX*, LP-V-S01-0622-LF; Figure 3.4B) and *endoplasmic reticulum protein 57* (*ERp57*, HC-N-N01-2411-LF; Figure 3.4C) were previously identified by the *P. monodon* EST project (http://161.200.123.190/, Tassanakajon et al., 2006; Preechaphol et al., 2007). ESTs representing *CRT*, *CNX* and *ERp57* possessed 714, 686 and 672 bp in length. CRT and Erp57 ESTs contained the putative start and stop codons of each gene homologues, respectively.

A)

```
GGGACTTACTCGAGTGACAGTCAAGAGCGAAGACGGAACGATGAAGACCTGGGTTTTTCTTGCCCTATTTGGGGTT
GCCCTAGTGGAATCTAAAGTATTTTTCGAAGAAAGATTCGACAGCCCTGATTGGGAGAAAAATTGGGTTCAGTCTG
CACACAAGGGGAAGGAGTTTGGACCCTTCAAGTTGACAGCTGGCAAATTTTATGGCGATGCTGAAAAGGATAAGGG
AATCCAGACTGGACAGGATGCCCGCTTTTATGGTCTTTCTACGAAGTTTGAGCCCTTCAGTAATAAGGATTCCCCA
CTTGTCATCCAGTTTTCTGTAAAACATGAACAGAACATTGACTGTGGTGGAGGATATCTAAAGGTCTTCGATTGCT
CTTTAGACCAGAAAGACATGCACGGAGAGTCGCCATACCTCATTATGTTTGGTCCTGATATCTGTGGCCCAGGCAC
CAAGAAGGTTCATGTAATCTTCAATTACAAGGGTGAGAACCATCTGATCAAGAAGGAAATCCGTTGCAAGGATGAC
GTATTTTCCCATCTGTATACCCTCATTGTCAATCCTGACAACACCTACGAAGTTCTTATTGACAATGAAAAAGCTC
AGTCTGGTGAACTCGAGGAGGACTGGGACTTCCTTCCACCCAAGAAGATCAAGGACCCAGAAGCCAAGAAGCCCGA
CGATTGGGATGACCGCCCCACCATTGCTGA
```

B)

```
GACAGACTGGAGTTATTCGACGAATCATCAACTACTCCAATAAGAATCCATGGCTATATGGTATCTACGTACTCCT
AGTGGCCATTCCTGTAGTGTTGATATTTGCCTGTTGTTGTGCAGAAGCAAAGGACACCAAGGAAGATAAGGATGCA
GAGAGGAAAAAGACCGATGCACCTTCTCCTGATGATCCACAATCAGAGCAAAACAAAGAAGATTCTAGTCAAACTG
CAGCAGATGATGATGCTCCTGGAAGTGGGGATGAAGCTGAAGGGGATGAAGACAAAGGTGATGAAGAGGGAGAGGA
GGAAGAAGATGAGGAAGAGGAGGAGGAAGCAGAGAAAGCTGATGCAGCTGAGGAGGTCCAGACTCGCACATCACCA
AGGCTGCGCAAGTCCCGGAGAGATTAAATGGCTGTGGATGATGTACACTACAACTTAGGAGCAAGAAGAGGTGCAA
CATGTACCTGTTGGAAGGAAGATATTTTTTGATTGGATTTTTATATGAGGCCTTAACAATATCTATTAGTACAGTA
TAAGCATTTTATGCAAGATATCTCAGTCATATAATGTAGTGACGAATGTGGGTAATATGTTCAGTATATGTTTCAT
AAATAGAATATTTTTGTCTTGTTCTAATACAGAGAAGAGTTATGAATCCTGTATTTTGCTCTGATTACGGTAAGCT
TG
```

C)

```
ATCTGAAGTCTGAGGCAGTGCCAACACAAGATGGCCCAGTCACTGTTGCTGTGGGTAAGAACTTCAATGAAGTTGT
CTCTGATGAGCGTGATGCCCTCATTGAATTCTATGCTCCTTGGTGTGGTCACTGCAAGAAATTAGCGCCCACCTAT
GATGAGCTGGGAGAAGCCGATGAAGGATGAAGATGTAGACATTGTGAAGATGGATGCCACTGCCAATGATGTTCCTC
CTCAGTACAATGTTCAAGGCTTCCCCACCATCTTCTGGAAACCCAAGGGTGGTGTTCCAAGGAATTACAATGGTGG
CCGGGAACTTGACGATTTTGTCAAGTACATTGCCCAACATTCCACAAATGAACTGAATGGGTATGACCGCAAGGGG
AAGGCAAAGAAAGGCAAGAAGACTGAACTTTGAGAATAAGAAATTTTTACGAGGAGGGTGTTGTCACCGCAGTGTG
TTGTGTAAAGAGTGCTGCAAATATGTTAGTCATGTATGTGATGATTATTCCATGCATTATAATTTTTCCATATGTAGCAAT
CAAAGGATGTCTTTTTGATGTAATATCTGTATTGTTGTATGCAATTAAACATACGCAAATTTGTTTCCTCTCCAAG
AGGCTTGTCAGGTGAAAACTTGCTGACAGCATAACAATTGGATATATTAAAATGAATTACCGGT
```

Figure 3.4 Nucleotide sequences of ESTs of *P. monodon* that were significantly matched *calreticulin* (*CRT*, clone no. OV-N-ST01-0117-W; A), *calnexin* (*CNX*, clone no. LP-V-S01-0622-LF; B) and *endoplasmic reticulum protein 57* (*ERp57*, HC-N-N01-2411-LF; C) previously identified (Tassanakajon et al., 2006). Primers were designed for further analysis. The putative start (ATG) and stop (TGA) codons are boldfaced and underlined.

3.4 Isolation and characterization of the full length cDNA of *calreticulin*, *calnexin* and *ERp57* homologues in *P. monodon* using RACE-PCR

3.4.1 *Calreticulin* (CRT)

The partial sequence of *calreticulin* from ovarian cDNA library was confirmed by re-sequencing of a 714 bp fragment generated from primers CRT-F and CRT-R. Similarity analysis using BlastX revealed that the complete 5′ end of this gene was already obtained. Therefore, only 3′RACE-PCR was carried out.

A 1344 bp fragment was generated by 3′RACE-PCR of *P. monodon calreticulin*. This fragment was cloned and sequenced. Nucleotide sequence of this fragment partially overlapped with that of the original EST for 376 bp (Figure 3.5). Nucleotide sequences of EST and 3′RACE-PCR were assembled. The full length cDNA *P. monodon calreticulin* (*PM-CRT*) was 1682 bp in length with an ORF of 1221 bp encoding a polypeptide of 406 amino acids. *PM-CRT* contained 5′ and 3′ UTR of 40 and 421 nucleotides, repectively. The poly A additional signal (AATAAA) was located between 1646-1651 of the *PM-CRT* cDNA. The closest sequence to *PM-CRT* was *calreticulin* of *Apis melliferal* (E-value = 2e-157).

The deduced PmCRT protein has the calculated molecular weight (MW) and theoretical isoelectric point (pI) of 46.76 kilodalton (kDa) and 4.3. SMART analysis shows the predicted signal peptide (MKTWVFLALFGVALVES) was located between S_{17} and K_{18}. The calreticulin domain conserved in the CRT protein family was found at positions 19-330 (2.30e-212). This conserved domain contained 3 repeats of IXDXEXXKPE(/D)DWD and a single motif of GxWxPPxIxNPxYx. The putative endoplasmic reticulum targeting tetrapepitide HDEL was found at the C-terminus of PmCRT (Figure 3.6).

A)

B)

```
ACATTGACTGTGGTGGAGGATATCTGAAGGTCTTCGATTGCTCTTTAGACCAGAAAGACATGCACGGAGAGTCGCC
ATACCTCATTATGTTTGGTCCTGATATCTGTGGCCCAGGCACCAAGAAGGTTCATGTAATCTTCAATTACAAGGGT
GAGAACCATCTGATCAAGAAGGAAATCCGTTGCAAGGATGACGTATTTTCCCATCTGTATACCCTCATTGTCAATC
CTGACAACACCTACGAAGTTCTTATTGACAATGAAAAAGCTCAGTCTGGTGAACTCGAGGAGGACTGGGACTTCCT
TCCACCCAAGAAGATCAAGGACCCAGAAGCCAAGAAGCCCGACGATTGGGATGACCGCCCCACCATTGCTGATCCT
GACGATACTAAGCCTGAAGATTGGGACCAACCTGAACACATTCCTGATCCTGATGCCACCAAACCTGAGGACTGGG
ATGATGAAATGGATGGCGAGTGGGAACCACCCATGATTGACAATCCTGACTACAAGGGTGAATGGAAGCCTAAGCA
GATTGATAACCCTGATTACAAGGGTCCATGGATTCACCCTGAAATTGACAACCCAGAATACACACCTGACCCAGAG
ATCTACAAGTATGATGAGGTCTGTGCTCTTGGTTTGGATCTTTGGCAGGTAAAATCTGGTACTATCTTTGACAACT
TCCTCATCTCAAATGATCCTGAAGAAGCCCGCAAGATTGGTGAAGAGACTTGGGGTGCTACTAAAGATGCAGCTAA
GAAGATGAAGGATGAACAGGATGAAGAGGAGCGAAAGAGAGCAGAGGAAGAAGCTAAGGCAGCTGCTGATGCTGAG
AAGGATGAGGACGATGATGACGACGACGATCTTGGCGATGAAGACGAAGATGATCTTGATAATGATCTTGAACATG
ACGAGCTGTAAAGTTATATTATTTTATTAAGGAAGTTATTATTTAAATAGCCTGTTGTACTATTTAAACATCAAAG
TACAATTAACTGAACCTTTTGGTTGTACATTCTTGTAAATACCAAGGGCTTCGGTTAATTCTAGTCATGGAATCTT
TTGTTGTGGGCAAACTAAAATCCAAGCATTCCAGTAGCTGAAGCTGATTGGAGGTTCCTTGACAAGAAGCCACCTA
TTTAGGTGGTATTCATAGTACATCACTCATCATTAATCATCCCATCATCATGTATGCAGCACAGTCCTGTATGTTT
GTACGAGACAAAAGTGTGTCAGTACTGATGTCAATTTAAACATGATTCAGTATCGCTGATTGCACAATGTGTGTCT
GAGTGAGCGATTTCCAATAAACACAATCAAAAAAAAAAAAAAAAAAAAAAAAA
```

Figure 3.5 3´ RACE-PCR product of *P. monodon calreticulin* (*PmCRT*, lane 1, A) electrophoretically analyzed on a 1.5% agarose gel. A 100 bp DNA ladder (lane M, A) was included as the DNA marker. The RACE-PCR product (1344 bp) was cloned and sequenced for both directions (B). Nucleotide sequence that overlapped with that of the original EST is highlighted. The sequence of RACE-PCR primer is illustrated in boldface.

```
GGGACTTACTCGAGTGACAGTCAAGAGCGAAGACGGAACGATGAAGACCTGGGTTTTTCT    60
                                        M  K  T  W  V  F  L       7
TGCCCTATTTGGGGTTGCCCTAGTGGAATCTAAAGTATTTTTCGAAGAAAGATTCGACAG   120
A  L  F  G  V  A  L  V  E  S  K  V  F  F  E  E  R  F  D  S      27
CCCTGATTGGGAGAAAAATTGGGTTCAGTCTGCACACAAGGGGAAGGAGTTTGGACCCTT   180
 P  D  W  E  K  N  W  V  Q  S  A  H  K  G  K  E  F  G  P  F     47
CAAGTTGACAGCTGGCAAATTTTATGGCGATGCTGAAAAGGATAAGGGAATCCAGACTGG   240
 K  L  T  A  G  K  F  Y  G  D  A  E  K  D  K  G  I  Q  T  G     67
ACAGGATGCCCGCTTTTATGGTCTTTCTACGAAGTTTGAGCCCTTCAGTAATAAGGATTC   300
 Q  D  A  R  F  Y  G  L  S  T  K  F  E  P  F  S  N  K  D  S     87
CCCACTTGTCATCCAGTTTTCTGTAAAACATGAACAGAACATTGACTGTGGTGGAGGATA   360
 P  L  V  I  Q  F  S  V  K  H  E  Q  N  I  D  C  G  G  G  Y    107
TCTAAAGGTCTTCGATTGCTCTTTAGACCAGAAAGACATGCATGGAGAGTCGCCATACCT   420
 L  K  V  F  D  C  S  L  D  Q  K  D  M  H  G  E  S  P  Y  L    127
CATTATGTTTGGTCCTGATATCTGTGGCCCAGGCACAAAGAAGGTTCATGTAATCTTCAA   480
 I  M  F  G  P  D  I  C  G  P  G  T  K  K  V  H  V  I  F  N    147
TTACAAGGGTGAGAACCATCTGATCAAGAAGGAAATCCGTTGCAAGGATGACGTATTTTC   540
 Y  K  G  E  N  H  L  I  K  K  E  I  R  C  K  D  D  V  F  S    167
CCATCTGTATACCCTCATTGTCAATCCTGACAACACCTACGAAGTTCTTATTGACAATGA   600
 H  L  Y  T  L  I  V  N  P  D  N  T  Y  E  V  L  I  D  N  E    187
AAAAGCTCAGTCTGGTGAACTCGAGGAGGACTGGGACTTCCTTCCACCCAAGAAGATCAA   660
 K  A  Q  S  G  E  L  E  E  D  W  D  F  L  P  P  K  K  I  K    207
GGACCCAGAAGCCAAGAAGCCCGACGATTGGGATGACCGCCCCACCATTGCTGATCCTGA   720
 D  P  E  A  K  K  P  D  D  W  D  D  R  P  T  I  A  D  P  D    227
CGATACTAAGCCTGAAGATTGGGACCAACCTGAACACATTCCTGATCCTGATGCCACCAA   780
 D  T  K  P  E  D  W  D  Q  P  E  H  I  P  D  P  D  A  T  K    247
ACCTGAGGACTGGGATGATGAAATGGATGGCGAGTGGGAACCACCCATGATTGACAATCC   840
 P  E  D  W  D  D  E  M  D  G  E  W  E  P  P  M  I  D  N  P    267
TGACTACAAGGGTGAATGGAAGCCTAAGCAGATTGATAACCCTGATTACAAGGGTCCATG   900
 D  Y  K  G  E  W  K  P  K  Q  I  D  N  P  D  Y  K  G  P  W    287
GATTCACCCTGAAATTGACAACCCAGAATACACACCTGACCCAGAGATCTACAAGTATGA   960
 I  H  P  E  I  D  N  P  E  Y  T  P  D  P  E  I  Y  K  Y  D    307
TGAGGTCTGTGCTCTTGGTTTGGATCTTTGGCAGGTAAAATCTGGTACTATCTTTGACAA  1020
 E  V  C  A  L  G  L  D  L  W  Q  V  K  S  G  T  I  F  D  N    327
CTTCCTCATCTCAAATGATCCTGAAGAAGCCCGCAAGATTGGTGAAGAGACTTGGGGTGC  1080
 F  L  I  S  N  D  P  E  E  A  R  K  I  G  E  E  T  W  G  A    347
TACTAAAGATGCAGCTAAGAAGATGAAGGATGCACAGGATGAAGAGGAGCGAAAGAGAGC  1140
 T  K  D  A  A  K  K  M  K  D  A  Q  D  E  E  R  K  R  A      367
AGAGGAAGAAGCTAAGGCAGCTGCTGATGCTGAGAAGGATGAGGACGATGATGACGACGA  1200
 E  E  A  K  A  A  A  D  A  E  K  D  E  D  D  D  D  D          387
CGATCTTGGCGATGAAGACGAAGATGATCTTGATAATGATCTTGAACATGACGAGCTGTA  1260
 D  L  G  D  E  D  E  D  D  L  D  N  D  L  E  H  D  E  L  *    406
AAGTTATATTATTTTATTAAGGAAGTTATTATTTAAATAGCCTGTTGTACTATTTAAACA  1320
TCAAAGTACAATTAACTGAACCTTTTGGTTGTACATTCTTGTAAATACCAAGGGCTTCGG  1380
TTAATTCTAGTCATGGAATCTTTTGTTGTGGGCAAACTAAATCCAAGCATTCCAGTAGC  1440
TGAAGCTGATTGGAGGTTCCTTGACAAGAAGCCACCTATTTAGGTGGTATTCATAGTACA  1500
TCACTCATCATTAATCATCCCATCATCATGTATGCAGCACAGTCCTGTATGTTTGTACGA  1560
GACAAAAGTGTGTCAGTACTGATGTCAATTTAAACATGATTCAGTATCGCTGATTGCACA  1620
ATGTGTGTCTGAGTGAGCGATTTCCAATAAACACAATCAAAAAAAAAAAAAAAAAAAAAA  1680
AA                                                            1682
```

Figure 3.6 The full length cDNA sequences of *PmCRT*. Start and stop codons are illustrated in boldface and underlined. The predicted signal peptide is boxed. Three IXDXEXXKPE(/D)DWD repeats (position 206-218, 223-235 and 240-252) are highlighted and underlined. The consensus GxWxPPxIxNPxYx motif is double-undelined. The endoplasmic reticulum targeting sequence (HDEL) is italicized. Sequence of the 5′ RACE-PCR primer is underlined, boldfaced and italicized. The polyA additional signal is boldfaced.
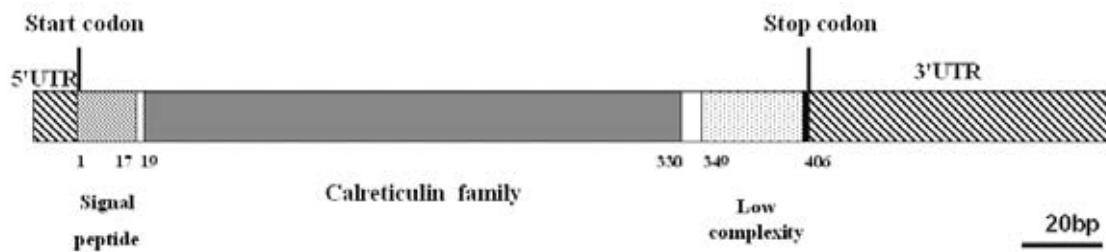
Name	Position	E-value
Signal peptide	1-17	-
Pfam : Calreticulin	19-330	2.30e-212
Low complexity	349-406	-

Figure 3.7 Diagram illustrating the full length cDNA of *PmCRT*. The signal peptide and calreticulin domain were found in positions 1-17 and 19-330, respectively. The scale bar is 200 bp in length.

3.4.2 *Calnexin (CNX)*

Several amplification fragments were obtained from 5′ and 3′ RACE-PCR of *P. monodon calnexin*. However, 1934 bp and 769 bp fragments were cloned and sequenced, respectively (Figure 3.8). Overlapping sequences between the 5′ RACE-PCR fragment and EST and the 5′ RACE-PCR fragment were 423 and 456 bp, respectively. Nucleotide sequences of EST and RACE-PCR products were assembled.

The full length cDNA of *P. monodon CNX* (called *PmCNX*) was 2509 bp in length consisting the 5′and 3′ UTRs of 130 and 591 bp and an ORF of 1788 bp encoding a polypeptide of 595 amino acids, respectively (Figure 3.9). The poly A additional signal (AATAAA) was located at positions 2460-2465 of the *PmCNX* transcript. The closest similarity of PmCNX was CNX of *Aedes aegypti* (E-value = 2e-169).

A)

B)

```
GCAGATGATGATGCTCCTGGAAGTGGGGATGAAGCTGAAGGGGATGAAGACAAAGGTGATGAAGAGGGAGAGGAGGAAGAAGATGG
AAGAGGAGGAGGAAGCAGAGAAAGCTGATGCAGCTGAGGAGGTCCAGACTCGCACATCACCAAGGCTGCGCAAGTCCCGGAGAGAT
TAAATGGCTGTGGATGATGTACACTACAACTTAGGAGCAAGAAGAGGTGCAACATGTACCTGTTGGAAGGAAGATATTTTTTGATT
GGATTTTTATATGAGGCCTTAACAATATCTATTAGTACAGTATAAGCATTTTATGCAAGATATCTCAGTCATATAATGTAGTGACG
AATGTGGGTAATATGTTCAGTTTTTGTTTCATAAATAGAATATTTTTGTCTTGTTCTAATACAGAGAAGAGTTATGAATCCTGTAT
TTTGCTCTGATTACAGTAAGCTTGTTATATTTTACATTGTAGTACCTAGTCATGTTGGGCTCAGACTATTTGAGAAGGAAAATGTA
TTGTACAATGTCAGTAGGTATTTCTGTTAGCTCTTTCAATTTTTAGGAAATGCTGTGATAGTCTGAAGCTTCTTACTTCATCTCAG
ATACATTCCTTTGGTTCTTGATCTGCGCCAGGTTATTGTATAGAAACATGAACAGAGCAGGAAGAAATGTTCTTTTCTTTTTGTTG
GATATAAAATTAATATCATAGATGTATGAATAAAATTTTGTAGTTCACTGTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

C)

```
CTAATACGACTCACTATAGGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGGGGGGGAAAGGACGACAGCTGAAGCGTGAGTCG
TCCATTGACTAAATACCTGAAATACGACCGATTTAGGAGGAGCGATGAAGTCGAGGTGGCAGAGAAAAGCAGTATTAGCACTGCTA
GTTCTTGGCCTGCTGTTACCTTTTGGTATTAAGGCAGATGACGATGACGATGAAGAAGCTGTTGTCACGGAAGAGCAAACAGAAGG
AGAGGAAGATGATGTAGAGGAGGTTGTGTATGCAACACCCAAGGCACTACCAAATGCATATCTGACTGAAACGTTTGATGACATAG
CTACTTTTGAGAAGACGTGGATCAAATCTGAAGCCAAGAAGGGCGGTGTTGATGAAAACATTGCTAAATATGATGGTGTTTGGGCA
GTAGAACCTGCTGAACGATTGGCTCTTACTGGGGATCGTGGGTTGGTGTTGAAGTCAAAGGCAAAACATGCAGCAATTGCAGCACC
ACTGAAGAAACCATTTGTATTCAGTAACAAGCCTTTTGTGGTTCAGTATGAAGTTAATTTACAGAATGGTCAAGAATGTGGTGGAG
CATATATCAAACTAATCAGTGCCCAAAAAGGACGTGTGGATCTTAAAAATTTCCATGACAAAACACCGTACACTATTATGTTTGGA
CCAGACAAATGTGGCAATGACTTCAAGTTGCATTTCATCTTCAGGCATGTTAATCCTCTTACTGAAGAAATTGAAGAAAAACATGC
TAAGAGACCACGTGACAAGATTGAGGAACCATTTAAGGACAAGAAGTCTCATTTGTACACATTAGTAATTCGACCAGACAACACCT
TTGAAATAAGCCTGGATCACGAGGTAATCAATTCAGGAAGCCTTCTGGAGGACTTCACCCCATCTGTCAACCCTCCCAAAGAAATT
GATGATCCTGAAGACTTTATGCCAGAAGACTGGGATGAAAGAGAAAAGATTCCAGATCCAGAAGCCGCAAAGCCTGATGACTGGGA
TGAGGATGCTCCCATGCAGATCCCTGATCCAGTGGCTGAGAAACCTAGTGGATGGCTGGATGATGAGCCAGAAATGGTGCCAGATC
CCACTGCTGAGAAACCTGATGACTGGGATGATGAAATGGATGGTGAATGGGAAGCTCCACTGATCACCAACCCCAAGTGTGTTGAT
GCACCGGGCTGTGGAGAGTGGAAGCCTCCCATGGTGGACAATCCTGAGTTCAAGGGCAAATGGCGTCCACCCATGATTGACAATCC
TAATTACCGTGGAAAATGGAAGCCACGAAAGATCCCCAACCCTGACTTCTTTGAAGACCTGGAACCTTTCAAGATGACTGCTATTG
ATGCTGTTGGTCTGGAATTGTGGTCAATGTCAGACAATATCCTCTTTGACAACATACTTGTCACAGACAATGTTGCTGAGGCTTAC
CAGTTTGCTCAAGAAACTTTTTGACTTGAAGGTCATGAAGATAGAGAAGGGACAGACTGGAGTCATTCGACGAATCATCAACCACTC
CAATAAGAATCCATGGCTATATGGTATCTACGTACTCCTAGTGGCCATTCCTGTAGTGTTGATATTTGCCTGTTGTTGTGCAGAAG
CAAAGGACACCAAGGAAGATAAGGATGCAGAGAGGAAAAAGACCGATGCACCTTCTCCTGATGATCCACAATCAGAGCAAAACAAA
GAAGATTCTAGTCAAACTGCAGCAGATGATGATGCTCCTGGAAGTGGGGATGAAGCTGAAGGGGATGAAGACAAAGGTGATGAAGA
GGGAGAGGAGGAAGAAGATGAGGAAGAGGAGGAGGAAGCAGAGAAAGCTGATGCAGCTGAGGAGGTCCAGACTCGCACATCACCAA
GGCTGCGCAAGTCCCGGAGAGATTAAATGGCTGTGGATGATG
```

Figure 3.8 3′ and 5′ RACE-PCR products of *P. monodon calnexin* (*PmCNX*, lanes 1 and 2, A) electrophoretically analyzed on a 1.5% agarose gel. A 100 bp DNA ladder (lane M, A) was included as the DNA marker. Arrowheads indicated bands that were cloned and sequenced. Nucleotide sequences of 3′ (769 bp, B) and 5′ (1934 bp, C) RACE-PCR products are shown. Overlapping nucleotide sequences between RACE-PCR products and EST (456 and 423 bp, respectively) are highlighted.

The calulcated p*I* and MW of the deduced PmCNX protein were 4.4 and 67.4 kDa, respectively. PmCNX contained a signal peptide of 26 amino acids and the putative peptidase cleavage site was located between A_{26} and D_{26}. A transmembrane domain was predicted at positions 478-500. PmCNX also contained the calreticulin domain located at residues 62-436 (E-value = 1.8e-233) (Figure 3.10). The calreticulin domain contained four motifs of IxDxExxKPE(/D)DWD and GxWxPPxIxNPxYx. This large domain also consisted juxtamembrane (lysine rich), acidic and phosphorylation domains. The phosphorylation domains contained a PKC (TXXD) and two CK2 (SSXXD/E) motifs, respectively. The C terminus of PmCNX possessed glutamic acid-rich region at positions 558-573. The putative ER retrieval motif (RKXRRE/D) was observed at the C-terminal end of PmCNX.

```
CTAATACGACTCACTATAGGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGGGGGGA  60
AAGGACGACAGCTGAAGCGTGAGTCGTCCATTGACTAAATACCTGAAATACGACCGATTT  120
AGGAGGAGCGATGAAGTCGAGGTGGCAGAGAAAAGCAGTATTAGCACTGCTAGTTCTTGG  180
          M   K   S   R   W   Q   R   K   A   V   L   A   L   L   V   L   G   17
CCTGCTGTTACCTTTTGGTATTAAGGCAGATGACGATGACGATGAAGAAGCTGTTGTCAC  240
  L   L   L   P   F   G   I   K   A   D   D   D   D   D   E   E   A   V   V   T   37
GGAAGAGCAAACAGAAGGAGAGGAAGATGATGTAGAGGAGGTTGTGTATGCAACACCCAA  300
  E   E   Q   T   E   G   E   E   D   D   V   E   E   V   V   Y   A   T   P   K   57
GGCACTACCAAATGCATATCTGACTGAAACGTTTGATGACATAGCTACTTTTGAGAAGAC  360
  A   L   P   N   A   Y   L   T   E   T   F   D   D   I   A   T   F   E   K   T   77
GTGGATCAAATCTGAAGCCAAGAAGGGCGGTGTTGATGAAAACATTGCTAAATATGATGG  420
  W   I   K   S   E   A   K   K   G   G   V   D   E   N   I   A   K   Y   D   G   97
TGTTTGGGCAGTAGAACCTGCTGAACGATTGGCTCTTACTGGGGATCGTGGGTTGGTGTT  480
  V   W   A   V   E   P   A   E   R   L   A   L   T   G   D   R   G   L   V   L   117
GAAGTCAAAGGCAAAACATGCAGCAATTGCAGCACCACTGAAGAAACCATTTGTATTCAG  540
  K   S   K   A   K   H   A   A   I   A   A   P   L   K   K   P   F   V   F   S   137
TAACAAGCCTTTTGTGGTTCAGTATGAAGTTAATTTACAGAATGGTCAAGAATGTGGTGG  600
  N   K   P   F   V   V   Q   Y   E   V   N   L   Q   N   G   Q   E   C   G   G   157
AGCATATATCAAACTAATCAGTGCCCAAAAAGGACGTGTGGATCTTAAAAATTTCCATGA  660
  A   Y   I   K   L   I   S   A   Q   K   G   R   V   D   L   K   N   F   H   D   177
CAAAACACCGTACACTATTATGTTTGGACCAGACAAATGTGGCAATGACTTCAAGTTGCA  720
  K   T   P   Y   T   I   M   F   G   P   D   K   C   G   N   D   F   K   L   H   197
TTTCATCTTCAGGCATGTTAATCCTCTTACTGAAGAAATTGAAGAAAAACATGCTAAGAG  780
  F   I   F   R   H   V   N   P   L   T   E   E   I   E   E   K   H   A   K   R   217
ACCACGTGACAAGATTGAGGAACCATTTAAGGACAAGAAGTCTCATTTGTACACATTAGT  840
  P   R   D   K   I   E   E   P   F   K   D   K   K   S   H   L   Y   T   L   V   237
AATTCGACCAGACAACACCTTTGAAATAAGCCTGGATCACGAGGTAATCAATTCAGGAAG  900
  I   R   P   D   N   T   F   E   I   S   L   D   H   E   V   I   N   S   G   S   257
CCTTCTGGAGGACTTCACCCCATCTGTCAACCCTCCCAAAGAAATTGATGATCCTGAAGA  960
  L   L   E   D   F   T   P   S   V   N   P   P   K   E   I   D   D   P   E   D   277
CTTTATGCCAGAAGACTGGGATGAAAGAGAAAAGATTCCAGATCCAGAAGCCGCAAAGCC  1020
  F   M   P   E   D   W   D   E   R   E   K   I   P   D   P   E   A   A   K   P   297
TGATGACTGGGATGAGGATGCTCCCATGCAGATCCCTGATCCAGTGGCTGAGAAACCTAG  1080
  D   D   W   D   E   D   A   P   M   Q   I   P   D   P   V   A   E   K   P   S   317
TGGATGGCTGGATGATGAGCCAGAAATGGTGCCAGATCCCACTGCTGAGAAACCTGATGA  1140
  G   W   L   D   D   E   P   E   M   V   P   D   P   T   A   E   K   P   D   D   337
CTGGGATGATGAAATGGATGGTGAATGGGAAGCTCCACTGATCACCAACCCCAAGTGTGT  1200
```

```
 W  D  D  E  M  D  G  E  W  E  A  P  L  I  T  N  P  K  C  V   357
TGATGCACCGGGCTGTGGAGAGTGGAAGCCTCCCATGGTGGACAATCCTGAGTTCAAGGG 1260
 D  A  P  G  C  G  E  W  K  P  P  M  V  D  N  P  E  F  K  G   377
CAAATGGCGTCCACCCATGATTGACAATCCTAATTACCGTGGAAAATGGAAGCCACGAAA 1320
 K  W  R  P  P  M  I  D  N  P  N  Y  R  G  K  W  K  P  R  K   397
GATCCCCAACCCTGACTTCTTTGAAGACCTGGAACCTTTCAAGATGACTGCTATTGATGC 1380
 I  P  N  P  D  F  F  E  D  L  E  P  F  K  M  T  A  I  D  A   417
TGTTGGTCTGGAATTGTGGTCAATGTCAGACAATATCCTCTTTGACAACATACTTGTCAC 1440
 V  G  L  E  L  W  S  M  S  D  N  I  L  F  D  N  I  L  V  T   437
AGACAATGTTGCTGAGGCTTACCAGTTTGCTCAAGAAACTTTTGACTTGAAGGTCATGAA 1500
 D  N  V  A  E  A  Y  Q  F  A  Q  E  T  F  D  L  K  V  M  K   457
GATAGAGAAGGGACAGACTGGAGTCATTCGACGAATCATCAACCACTCCAATAAGAATCC 1560
 I  E  K  G  Q  T  G  V  I  R  R  I  I  N  H  S  N  K  N  P   477
ATGGCTATATGGTATCTACGTACTCCTAGTGGCCATTCCTGTAGTGTTGATATTTGCCTG 1620
 W  L  Y  G  I  Y  V  L  L  V  A  I  P  V  V  L  I  F  A  C   497
TTGTTGTGCAGAAGCAAAGGACACCAAGGAAGATAAGGATGCAGAGAGGAAAAAGACCGA 1680
 C  C  A  E  A  K  D  T  K  E  D  K  D  A  E  R  K  K  T  D   517
TGCACCTTCTCCTGATGATCCACAATCAGAGCAAAACAAAGAAGATTCTAGTCAAACTGC 1740
 A  P  S  P  D  D  P  Q  S  E  Q  N  K  E  D  S  S  Q  T  A   537
AGCAGATGATGATGCTCCTGGAAGTGGGGATGAAGCTGAAGGGGATGAAGACAAAGGTGA 1800
 A  D  D  D  A  P  G  S  G  D  E  A  E  G  D  E  D  K  G  D   557
TGAAGAGGGAGAGGAGGAAGAAGATGAGGAAGAGGAGGAGGAAGCAGAGAAAGCTGATGC 1860
 E  E  G  E  E  E  E  D  E  E  E  E  E  A  E  K  A  D  A      577
AGCTGAGGAGGTCCAGACTCGCACATCACCAAGGCTGCGCAAGTCCCGGAGAGAT TAA AT 1920
 A  E  E  V  Q  T  R  T  S  P  R  L  R  K  S  R  R  D   *
GGCTGTGGATGATGTACACTACAACTTAGGAGCAAGAGAGGTGCAACATGTACCTGTTG 1980
GAAGGAAGATATTTTTTGATTGGATTTTTATATGAGGCCTTAACAATATCTATTAGTACA 2040
GTATAAGCATTTTATGCAAGATATCTCAGTCATATAATGTAGTGACGAATGTGGGTAATA 2100
TGTTCAGTTTTTGTTTCATAAATAGAATATTTTTGTCTTGTTCTAATACAGAGAAGAGTT 2160
ATGAATCCTGTATTTTGCTCTGATTACAGTAAGCTTGTTATATTTTACATTGTAGTACCT 2220
AGTCATGTTGGGCTCAGACTATTTGAGAAGGAAATGTATTGTACAATGTCAGTAGGTAT 2280
TTCTGTTAGCTCTTTCAATTTTTAGGAAATGCTGTGATAGTCTGAAGCTTCTTACTTCAT 2340
CTCAGATACATTCCTTTGGTTCTTGATCTGCGCCAGGTTATTGTATAGAAACATGAACAG 2400
AGCAGGAAGAAATGTTCTTTTCTTTTTGTTGGATATA AAAT TAATATCATAGATGTATG A 2460
ATAAA ATTTTGTAGTTCACTGTCAAAAAAAAAAAAAAAAAAAAAAAAAAAA         2510
```

Figure 3.9 The full length cDNA sequence of *PmCNX*. Start and stop codons are illustrated in boldfaced and underlined. The predicted signal peptide is boxed. Three IXDXEXXKPE(/D)DWD repeats (position 206-218, 223-235 and 240-252) are highlighted and underlined. Four consensus GxWxPPxIxNPxYx motif are double-underlined. A Lys-rich (juxtamembrane) domain found at residues 503-515 was boldfaced and italicized. Two putative CK2 (SXXD/E) phosphorylation motifs (residues 520-523 and 545-548) are highlighted and underlined. A PKC (TXXD) phosphorylation motif (positions 536-539 of the deduced CNX protein) is underlined. An acidic acid domain and a putative ER retrieval motif (RKXRRE/D) are double-underlined and highlighted, respectively. The polyA additional signal is boldfaced.

A)
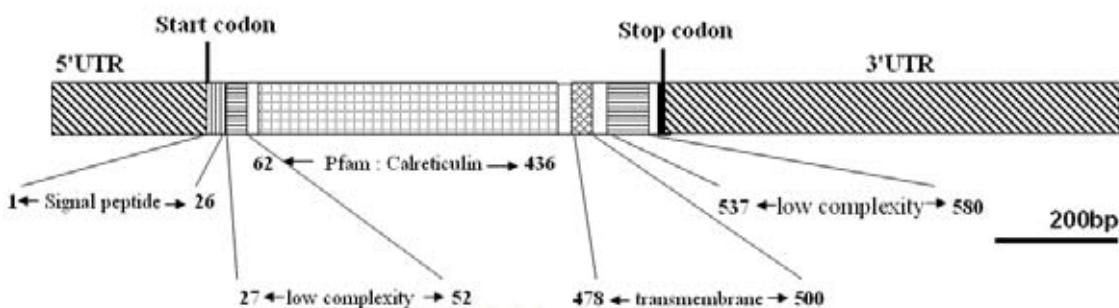
Name	Position	E-value
Signal peptide	1-26	-
Low complexity	27-52	-
Pfam : Calreticulin	62-436	1.8e-233
Transmembrane	478-500	-
Low complexity	537-580	-

B)
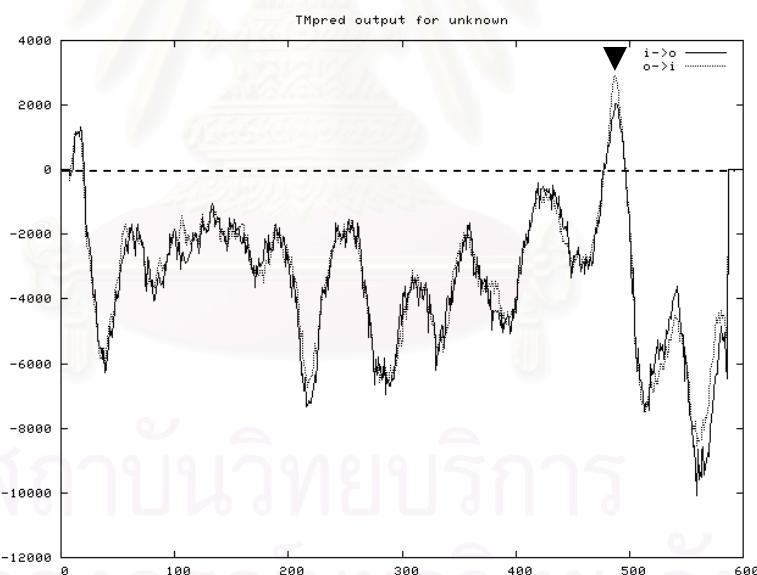
Figure 3.10 Diagram illustrating the full length cDNA of *PmCNX* (A). The signal peptide and calreticulin domain were found at residues 1-26 and 62-436 of the deduced CNX, respectively. A predicted transmembrane domain was found at positions 478-500 (arrowhead, B). The scale bar is 200 bp in length.

3.4.3 *Endoplasmic reticulum protein 57 (ERp57)*

A discrete band of 1571 bp was obtained from 5′RACE-PCR whereas both 713 and 529 bp fragments were obtained from 3′RACE-PCR (Figure 3.11). These amplified fragments were cloned and sequenced. Nucleotide sequences of these fragments were overlapped with the original EST clones for 32, 320 and 320 bp, respectively. After sequence assembly, two different isoforms of *P. monodon ERp57* with 2100 and 2284 bp in length regarded as short and long isoforms of *PmMERp57*, respectively. They shared an identical ORF of 1458 bp corresponding to a polypeptide of 485 amino acids with 5′ UTR of 113 bp and length polymorphism of the 3′ UTR for 529 and 713 bp, respectively. The putative polyadenylation signals (AATAAA) were found in both transcripts. Nucleotide sequences of both *PmCNX* isoforms were aligned and the difference between these sequences was observed due to the length of 3′ UTR (Figure 3.14). This sequence significantly matched CNX of *Bombyx mori* (E-value = 5e-154).

A)
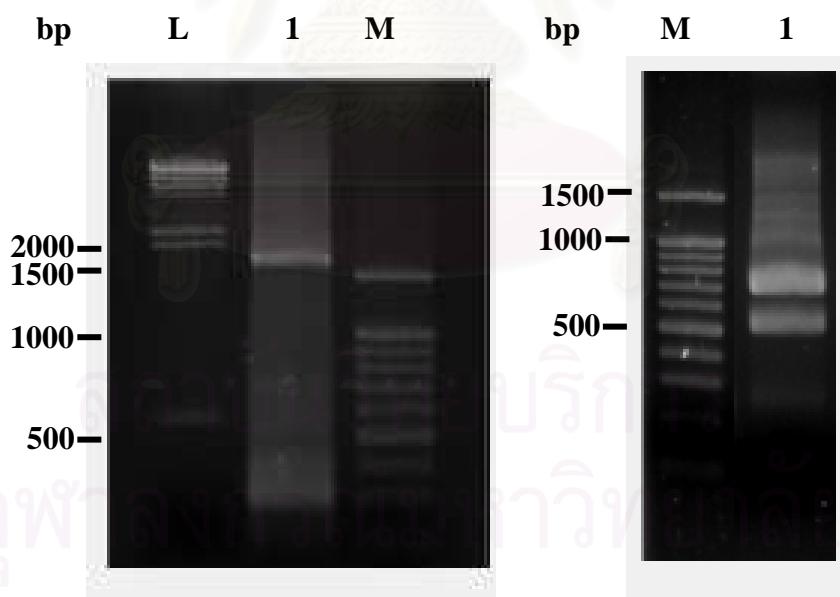
B)

```
CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGAAGAAAGGTCACGTGG
AAGGAGTGGCAAGAAAGAACGTGAAGAAGCAGAAAAAGAAAAGAATGGCTACGAGATTGTTAATACTAC
TCCTCTCCCTCGTGGCCGTGGCGCTGGGAGACGATGTCCTGCAATTAAACGACGCGGATTTCGACGGGA
AAGTGGCCAGCTACGACACGGTCCTCGTCATGTTCTACGCCCCGTGGTGTGGTCACTGCAAGAGATTAA
```

AGCCTGAGTTTGAGAAGGCCTCTACCACCTTGAAGGCCAACGACCCTCCCGTCTACCTTGCTAAGGTGG
ATTGTACTGATGATGGAAAGGACAGCTGTAGCAGATTTGGTGTCTCTGGCTACCCTACCCTGAAGATCT
TCAAGGGAGGAGAGCTCTCTACGGACTACAATGGTCCACGAGATGCCAGTGGTATTGTAAAATACATGA
GGTCACAGGTTGGACCAGCCTCTAAGGAGTTGACATCCGTGGAGGCAGCAGAAGCATTCCTTGGTGCTG
CTGAAGTTGGAGTCGTTTACTTTGGAGGAGATTCCAAACTTAAAGATGCTTTCCTAAAGGCTGCTGATA
AGCTGAGGGAATCCATCCGTTTTGCACACTCCCTCGATGCCACTGTTAATGAAAGTATGGGTACAGTG
ATGTTGTTGTACTTTTCCGACCAAAACACCTGGAGAACAAATTTGAGCCTTCCTCTGTTGTATTTGAGG
GATCGGCAGACAGGGCTGAGATTGAGTCTTTCATCAAAAAGAACTTCCATGGTTTGGTAGGACACCTAA
CGCAAGACACTGCTCAGGATTTCAAACCTCCAGTTGTGATTGCTTACTACAATGTTGATTACATCAAAA
ATGTTAAGGGTACAAATTACTGGCGCAATCGTGTCCTTAAGGTGGCACAAAACTTTGCTGATGACTTCA
AGTTTGCCGTTGCCAATAAGGACGACTTCCAGCATGACCTCAATGAATATGGCCTTGATTATGTTCCTG
GTGACAAGCCAGTAATTTGTGCACGTAATGCTAAAGCCCAGAAGTTTGTCATGCAGGAAGAATTTTCAA
TGGATAACCTCCAAGCATTCCTCACCAATCTCAAGGCGGGTGAGCTTGAGCCAT**ATCTGAAGTCTGAGG**
CAGTGCCAACACAAGAT

C)

TGAACTGAATGGGTATGACCGCAAGGGGAAGGCAAAGAAAGGCAAGAAGACTGAACTTTGAG
AATAAGAAATTTTTACGAGGAGGGTGTTGTCACCGCAGTGTGTTGTGTAAAGAGTGCTGCAG
ATATGTTAGTCATGTATGATTATTCCATGCATTATAATTTTTCCATATGTAGCAATCAAAGG
ATGTCTTTTTGATGTAATATCTGTATTGTTGTATGCAATTAAACATACGCAAATTTGTTTCC
TCTCAAAGAGGCTTTCAGGTGAAAACTTGCTGACAGTATTAAGATTGGAATATATTAAAATG
AATTACCGGTGAAGCTTGTCCTGTGCCTCTGTTTGTGTGAAAGCTAAGAATGGGTGAAGTTT
TGTGCAATAGTTTTTTCCTAAATTTCTCTTGTAAAGAAAAATGACGCAATGCAGTGGACTTTT
GAAGACTAAATTGATACAATCAATCATTCCGGCCCTTGTCTGAAATACAAGAGCACAAGTTT
ATCCAGTGTAATGCATTCAGTTAACTTAATAAATTTTATAATATAAAAAAAAAAAAAAAAAAA
AAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

D)

TGAACTGAATGGGTATGACCGCAAGGGGAAGGCAAAGAAAGGCAAGAAGACTGAACTTTGAGAATAAGA
AATTTTTACGAGGAGGGTGTTGTCACCGCAGTGTGTTGTGTAAAGAGTGCTGCAGATATGTTAGTCATG
TATGATTATTCCATGCATTATAATTTTTCCATATGTAGCAATCAAAGGATGTCTTTTTGATGTAATATC
TGTATTGTTGTATGCAATTAAACATACGCAAATTTGTTTCCTCTCAAAGAGGCTTTCAGGTGAAAACTT
GCTGACAGTATTAAGATTGGAATATATTAAAATGAATTACCGGTGAAGCTTGTCCTGTGCCTCTGTTTG
TGTGAAAGCTAAGAATGGGTGAAGTTTTGTGCAATAGTTTTTCCTAAATTTCTCTTGTAAAGAAAAATG
ACGCAATGCAGTGGACTTTTGAAGACTAAATTGATACAATCAATCATTCCGGCCCTTGTCTGAAATACA
AGAGCACAAGTTTATCCAGTGTAATGCATTCAGTTAACTTAATAAATTTTATAATATATTTATTTACTT
TTTTTGAACCTTGGAAAAACTTGTGGTGGCCTGCATGGTTATTGCACAAATTTGATTACTGTAATGAGG
ACTTGAATTATGTCTGCTAAACGGATTGAATCTTGTGTGATTGTACTGTTAAAATTTCCAATGAAATAT
TTGATAGGAAATATTGACTTTTTGACATGTAAAACAAATAAAAACATTTACTCAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAA

Figure 3.11 5′ and 3′ RACE-PCR products of *P. monodon* ERp57 (*PmERp57*, lanes 1; A, left and right panels) electrophoretically analyzed in a 1.5% agarose gel. A 100 bp DNA ladder (lane M, A) and λ-*Hin*d III (Lane L, A) were included as DNA markers. Arrowheads indicated bands that were cloned and sequenced. Nucleotide sequences of 5′ (B) and different length of 3′ (C and D) RACE-PCR products are shown. Overlapping nucleotide sequences between RACE-PCR products and EST (320 bp) are highlighted.

The deduced PmERp57 has a calculated MW of 53.9 kD and a theoretical p*I* of 5.48. The predicted signal peptide of PmERp57 was 18 amino acids long. The predicted peptidase cleavage site was located between G_{18} and D_{19} residues. PmERp57 contained domains a, b, b′, and a′ typically found in the ERp57 proteins of various taxa. Two thioredoxin domains were found at residues 20-126 (E-value = 5.9e-46) and 361-463 (E-value = 1e-46) which were located in the redox-active (a and a′ domains, respectively). These thioredoxin regions contain CGHC motifs typically found in the zinc finger proteins. Domains b and b′ of ERp57 are redox-inactive and located at residues 130-232 and 235-342, respectively (Figure 3.16).

```
CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGAAGAAAG  60
GTCACGTGGAAGGAGTGGCAAGAAAGAACGTGAAGAAGCAGAAAAAGAAAAGAATGGCTA 120
                                                    M  A  T  3
CGAGATTGTTAATACTACTCCTCTCCCTCGTGGCCGTGGCGCTGGGAGACGATGTCCTGC 180
  R  L  L  I  L  L  L  S  L  V  A  V  A  L  G  D  D  V  L  Q 23
AATTAAACGACGCGGATTTCGACGGGAAAGTGGCCAGCTACGACACGGTCCTCGTCATGT 240
  L  N  D  A  D  F  D  G  K  V  A  S  Y  D  T  V  L  V  M  F 43
TCTACGCCCCGTGGTGTGGTCACTGCAAGAGATTAAAGCCTGAGTTTGAGAAGGCCTCTA 300
  Y  A  P  W  C  G  H  C  K  R  L  K  P  E  F  E  K  A  S  T 63
CCACCTTGAAGGCCAACGACCCTCCCGTCTACCTTGCTAAGGTGGATTGTACTGATGATG 360
  T  L  K  A  N  D  P  P  V  Y  L  A  K  V  D  C  T  D  D  G 83
GAAAGGACAGCTGTAGCAGATTTGGTGTCTCTGGCTACCCTACCCTGAAGATCTTCAAGG 420
  K  D  S  C  S  R  F  G  V  S  G  Y  P  T  L  K  I  F  K  G 103
GAGGAGAGCTCTCTACGGACTACAATGGTCCACGAGATGCCAGTGGTATTGTAAAATACA 480
  G  E  L  S  T  D  Y  N  G  P  R  D  A  S  G  I  V  K  Y  M 123
TGAGGTCACAGGTTGGACCAGCCTCTAAGGAGTTGACATCCGTGGAGGCAGCAGAAGCAT 540
  R  S  Q  V  G  P  A  S  K  E  L  T  S  V  E  A  A  E  A  F 143
TCCTTGGTGCTGCTGAAGTTGGAGTCGTTTACTTTGGAGGAGATTCCAAACTTAAAGATG 600
  L  G  A  A  E  V  G  V  V  Y  F  G  G  D  S  K  L  K  D  A 163
CTTTCCTAAAGGCTGCTGATAAGCTGAGGGAATCCATCCGTTTTGCACACTCCCTCGATG 660
  F  L  K  A  A  D  K  L  R  E  S  I  R  F  A  H  S  L  D  A 183
CCACTGTTAATGAAAAGTATGGGTACAGTGATGTTGTTGTACTTTTCCGACCAAAACACC 720
  T  V  N  E  K  Y  G  Y  S  D  V  V  V  L  F  R  P  K  H  L 203
TGGAGAACAAATTTGAGCCTTCCTCTGTTGTATTTGAGGGATCGGCAGACAGGGCTGAGA 780
  E  N  K  F  E  P  S  S  V  V  F  E  G  S  A  D  R  A  E  I 223
TTGAGTCTTTCATCAAAAAGAACTTCCATGGTTTGGTAGGACACCTAACGCAAGACACTG 840
  E  S  F  I  K  K  N  F  H  G  L  V  G  H  L  T  Q  D  T  A 243
CTCAGGATTTCAAACCTCCAGTTGTGATTGCTTACTACAATGTTGATTACATCAAAAATG 900
  Q  D  F  K  P  P  V  V  I  A  Y  Y  N  V  D  Y  I  K  N  V 263
TTAAGGGTACAAATTACTGGCGCAATCGTGTCCTTAAGGTGGCACAAAACTTTGCTGATG 960
  K  G  T  N  Y  W  R  N  R  V  L  K  V  A  Q  N  F  A  D  D 283
ACTTCAAGTTTGCCGTTGCCAATAAGGACGACTTCCAGCATGACCTCAATGAATATGGCC 1020
  F  K  F  A  V  A  N  K  D  D  F  Q  H  D  L  N  E  Y  G  L 303
TTGATTATGTTCCTGGTGACAAGCCAGTAATTTGTGCACGTAATGCTAAGCCCAGAAGT 1080
  D  Y  V  P  G  D  K  P  V  I  C  A  R  N  A  K  A  Q  K  F 323
TTGTCATGCAGGAAGAATTTTCAATGGATAACCTCCAAGCATTCCTCACCAATCTCAAGG 1140
  V  M  Q  E  E  F  S  M  D  N  L  Q  A  F  L  T  N  L  K  A 343
CGGGTGAGCTTGAGCCATATCTGAAGTCTGAGGCAGTGCCAACACAAGATGGCCCTGTCA 1200
  G  E  L  E  P  Y  L  K  S  E  A  V  P  T  Q  D  G  P  V  T 363
CTGTTGCTGTGGGTAAGAACTTCAATGAAGTTGTCTCTGATGAGCGTGATGCCCTCATTG 1260
  V  A  V  G  K  N  F  N  E  V  V  S  D  E  R  D  A  L  I  E 383
```

```
AATTCTATGCTCCTTGGTGTGGTCACTGCAAGAAATTAGCGCCCACCTATGATGAGCTGG 1320
  F  Y  A  P  W  C  G  H  C  K  K  L  A  P  T  Y  D  E  L  G 403
GAGAAGCGATGAAGGATGAAGATGTAGACATTGTGAAGATGGATGCCACTGCCAATGATG 1380
  E  A  M  K  D  E  D  V  D  I  V  K  M  D  A  T  A  N  D  V 423
TTCCTCCTCAGTACAATGTTCAAGGCTTCCCCGCCATCTTCTGGAAACCCAAGGGTGGTG 1440
  P  P  Q  Y  N  V  Q  G  F  P  A  I  F  W  K  P  K  G  G  V 443
TTCCAAGGAATTACAACGGTGGCCGGGAACTGGACGATTTTGTCAAGTACATTGCCCAAC 1500
  P  R  N  Y  N  G  G  R  E  L  D  D  F  V  K  Y  I  A  Q  H 463
ATTCCACAAATGAACTGAATGGGTATGACCGCAAGGGGAAGGCAAAGAAAGGCAAGAAGA 1560
  S  T  N  E  L  N  G  Y  D  R  K  G  K  A  K  K  G  K  K  T 483
CTGAACTTTGAGAATAAGAAATTTTTACGAGGAGGGTGTTGTCACCGCAGTGTGTTGTGT 1620
  E  L  *                                                      485
AAAGAGTGCTGCAGATATGTTAGTCATGTATGATTATTCCATGCATTATAATTTTTCCAT 1680
ATGTAGCAATCAAAGGATGTCTTTTTGATGTAATATCTGTATTGTTGTATGCAATTAAAC 1740
ATACGCAAATTTGTTTCCTCTCAAAGAGGCTTTCAGGTGAAAACTTGCTGACAGTATTAA 1800
GATTGGAATATATTAAAATGAATTACCGGTGAAGCTTGTCCTGTGCCTCTGTTTGTGTGA 1860
AAGCTAAGAATGGGTGAAGTTTTGTGCAATAGTTTTTCCTAAATTTCTCTTGTAAAGAAA 1920
AATGACGCAATGCAGTGGACTTTTGAAGACTAAATTGATACAATCAATCATTCCGGCCCT 1980
TGTCTGAAATACAAGAGCACAAGTTTATCCAGTGTAATGCATTCAGTTAACTTAATAAAT 2040
TTTATAATATAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2100
```

Figure 3.12 The full length cDNA and deduce protein sequences (2100 bp in length with an ORF 1458 bp corresponding to a poly peptide of 485 aa) of a short form of *P. monodon ERp57* (*PmERp57-s*). Start and stop were illustrated in boldface and underlined. The poly A additional signal is boldfaced.

```
CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGAAGAAAG 60
GTCACGTGGAAGGAGTGGCAAGAAAGAACGTGAAGAAGCAGAAAAAGAAAAGAATGGCTA 120
                                                      M  A  T 3
CGAGATTGTTAATACTACTCCTCTCCCTCGTGGCCGTGGCGCTGGGAGACGATGTCCTGC 180
  R  L  L  I  L  L  L  S  L  V  A  V  A  L  G  D  D  V  L  Q 23
AATTAAACGACGCGGATTTCGACGGGAAAGTGGCCAGCTACGACACGGTCCTCGTCATGT 240
  L  N  D  A  D  F  D  G  K  V  A  S  Y  D  T  V  L  V  M  F 43
TCTACGCCCCGTGGTGTGGTCACTGCAAGAGATTAAAGCCTGAGTTTGAGAAGGCCTCTA 300
  Y  A  P  W  C  G  H  C  K  R  L  K  P  E  F  E  K  A  S  T 63
CCACCTTGAAGGCCAACGACCCTCCCGTCTACCTTGCTAAGGTGGATTGTACTGATGATG 360
  T  L  K  A  N  D  P  P  V  Y  L  A  K  V  D  C  T  D  D  G 83
GAAAGGACAGCTGTAGCAGATTTGGTGTCTCTGGCTACCCTACCCTGAAGATCTTCAAGG 420
  K  D  S  C  S  R  F  G  V  S  G  Y  P  T  L  K  I  F  K  G 103
GAGGAGAGCTCTCTACGGACTACAATGGTCCACGAGATGCCAGTGGTATTGTAAAATACA 480
  G  E  L  S  T  D  Y  N  G  P  R  D  A  S  G  I  V  K  Y  M 123
TGAGGTCACAGGTTGGACCAGCCTCTAAGGAGTTGACATCCGTGGAGGCAGCAGAAGCAT 540
  R  S  Q  V  G  P  A  S  K  E  L  T  S  V  E  A  A  E  A  F 143
TCCTTGGTGCTGCTGAAGTTGGAGTCGTTTACTTTGGAGGAGATTCCAAACTTAAAGATG 600
  L  G  A  A  E  V  G  V  V  Y  F  G  G  D  S  K  L  K  D  A 163
CTTTCCTAAAGGCTGCTGATAAGCTGAGGGAATCCATCCGTTTTGCACACTCCCTCGATG 660
  F  L  K  A  A  D  K  L  R  E  S  I  R  F  A  H  S  L  D  A 183
CCACTGTTAATGAAAAGTATGGGTACAGTGATGTTGTTGTACTTTTCCGACCAAAACACC 720
  T  V  N  E  K  Y  G  Y  S  D  V  V  V  L  F  R  P  K  H  L 203
TGGAGAACAAATTTGAGCCTTCCTCTGTTGTATTTGAGGGATCGGCAGACAGGGCTGAGA 780
  E  N  K  F  E  P  S  S  V  V  F  E  G  S  A  D  R  A  E  I 223
TTGAGTCTTTCATCAAAAAGAACTTCCATGGTTTGGTAGGACACCTAACGCAAGCACTG 840
  E  S  F  I  K  K  N  F  H  G  L  V  G  H  L  T  Q  D  T  A 243
CTCAGGATTTCAAACCTCCAGTTGTGATTGCTTACTACAATGTTGATTACATCAAAAATG 900
  Q  D  F  K  P  P  V  V  I  A  Y  Y  N  V  D  Y  I  K  N  V 263
TTAAGGGTACAAATTACTGGCGCAATCGTGTCCTTAAGGTGGCACAAAACTTTGCTGATG 960
```

```
       K  G  T  N  Y  W  R  N  R  V  L  K  V  A  Q  N  F  A  D  D 283
ACTTCAAGTTTGCCGTTGCCAATAAGGACGACTTCCAGCATGACCTCAATGAATATGGCC 1020
       F  K  F  A  V  A  N  K  D  D  F  Q  H  D  L  N  E  Y  G  L 303
TTGATTATGTTCCTGGTGACAAGCCAGTAATTTGTGCACGTAATGCTAAAGCCCAGAAGT 1080
       D  Y  V  P  G  D  K  P  V  I  C  A  R  N  A  K  A  Q  K  F 323
TTGTCATGCAGGAAGAATTTTCAATGGATAACCTCCAAGCATTCCTCACCAATCTCAAGG 1140
       V  M  Q  E  E  F  S  M  D  N  L  Q  A  F  L  T  N  L  K  A 343
CGGGTGAGCTTGAGCCATATCTGAAGTCTGAGGCAGTGCCAACACAAGATGGCCCTGTCA 1200
       G  E  L  E  P  Y  L  K  S  E  A  V  P  T  Q  D  G  P  V  T 363
CTGTTGCTGTGGGTAAGAACTTCAATGAAGTTGTCTCTGATGAGCGTGATGCCCTCATTG 1260
       V  A  V  G  K  N  F  N  E  V  V  S  D  E  R  D  A  L  I  E 383
AATTCTATGCTCCTTGGTGTGGTCACTGCAAGAAATTAGCGCCCACCTATGATGAGCTGG 1320
       F  Y  A  P  W  C  G  H  C  K  K  L  A  P  T  Y  D  E  L  G 403
GAGAAGCGATGAAGGATGAAGATGTAGACATTGTGAAGATGGATGCCACTGCCAATGATG 1380
       E  A  M  K  D  E  D  V  D  I  V  K  M  D  A  T  A  N  D  V 423
TTCCTCCTCAGTACAATGTTCAAGGCTTCCCCGCCATCTTCTGGAAACCCAAGGGTGGTG 1440
       P  P  Q  Y  N  V  Q  G  F  P  A  I  F  W  K  P  K  G  G  V 443
TTCCAAGGAATTACAACGGTGGCCGGGAACTGGACGATTTTGTCAAGTACATTGCCCAAC 1500
       P  R  N  Y  N  G  G  R  E  L  D  D  F  V  K  Y  I  A  Q  H 463
ATTCCACAAATGAACTGAATGGGTATGACCGCAAGGGGAAGGCAAAGAAAGGCAAGAAGA 1560
       S  T  N  E  L  N  G  Y  D  R  K  G  K  A  K  K  G  K  K  T 483
CTGAACTT**TGA**GAATAAGAAATTTTTACGAGGAGGGTGTTGTCACCGCAGTGTGTTGTGT 1620
       E  L  *
AAAGAGTGCTGCAGATATGTTAGTCATGTATGATTATTCCATGCATTATAATTTTTCCAT 1680
ATGTAGCAATCAAAGGATGTCTTTTTGATGTAATATCTGTATTGTTGTATGCAATTAAAC 1740
ATACGCAAATTTGTTTCCTCTCAAAGAGGCTTTCAGGTGAAAACTTGCTGACAGTATTAA 1800
GATTGGAATATATTAAAATGAATTACCGGTGAAGCTTGTCCTGTGCCTCTGTTTGTGTGA 1860
AAGCTAAGAATGGGTGAAGTTTTGTGCAATAGTTTTTCCTAAATTTCTCTTGTAAAGAAA 1920
AATGACGCAATGCAGTGGACTTTTGAAGACTAAATTGATACAATCAATCATTCCGGCCCT 1980
TGTCTGAAATACAAGAGCACAAGTTTATCCAGTGTAATGCATTCAGTTAACTTAATAAAT 2040
TTTATAATATATTTATTTACTTTTTTTTGAACCTTGGAAAAACTTGTGGTGGCCTGCATGG 2100
TTATTGCACAAATTTGATTACTGTAATGAGGACTTGAATTATGTCTGCTAAACGGATTGA 2160
ATCTTGTGTGATTGTACTGTTAAAATTTCCAATGAAATATTTGATAGGAAATATTGACTT 2220
TTTGACATGTAAAAACA**AATAAA**AACATTTACTCAAAAAAAAAAAAAAAAAAAAAAAAAA 2280
AAAA                                                        2284
```

Figure 3.13 The full length cDNA and deduce protein sequences (2284 bp in length with an ORF 1458 bp corresponding to a poly peptide of 485 aa) of a long form of *P. monodon ERp57* (*PmERp57*-l). Start and stop were illustrated in boldface and underlined. The polyA additional signal is boldfaced.

```
PMERp57-l    CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGAAGAAAG 60
PMERp57-S    CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGAAGAAAG 60
             ************************************************************
PMERp57-l    GTCACGTGGAAGGAGTGGCAAGAAAGAACGTGAAGAAGCAGAAAAAGAAAAGAATGGCTA 120
PMERp57-S    GTCACGTGGAAGGAGTGGCAAGAAAGAACGTGAAGAAGCAGAAAAAGAAAAGAATGGCTA 120
             ************************************************************
PMERp57-l    CGAGATTGTTAATACTACTCCTCTCCCTCGTGGCCGTGGCGCTGGGAGACGATGTCCTGC 180
PMERp57-S    CGAGATTGTTAATACTACTCCTCTCCCTCGTGGCCGTGGCGCTGGGAGACGATGTCCTGC 180
             ************************************************************
PMERp57-l    AATTAAACGACGCGGATTTCGACGGGAAAGTGGCCAGCTACGACACGGTCCTCGTCATGT 240
PMERp57-S    AATTAAACGACGCGGATTTCGACGGGAAAGTGGCCAGCTACGACACGGTCCTCGTCATGT 240
             ************************************************************
PMERp57-l    TCTACGCCCCGTGGTGTGGTCACTGCAAGAGATTAAAGCCTGAGTTTGAGAAGGCCTCTA 300
PMERp57-S    TCTACGCCCCGTGGTGTGGTCACTGCAAGAGATTAAAGCCTGAGTTTGAGAAGGCCTCTA 300
             ************************************************************
PMERp57-l    CCACCTTGAAGGCCAACGACCCTCCCGTCTACCTTGCTAAGGTGGATTGTACTGATGATG 360
```

```
PMERp57-S   CCACCTTGAAGGCCAACGACCCTCCCGTCTACCTTGCTAAGGTGGATTGTACTGATGATG 360
            ************************************************************
PMERp57-l   GAAAGGACAGCTGTAGCAGATTTGGTGTCTCTGGCTACCCTACCCTGAAGATCTTCAAGG 420
PMERp57-s   GAAAGGACAGCTGTAGCAGATTTGGTGTCTCTGGCTACCCTACCCTGAAGATCTTCAAGG 420
            ************************************************************
PMERp57-l   GAGGAGAGCTCTCTACGGACTACAATGGTCCACGAGATGCCAGTGGTATTGTAAAATACA 480
PMERp57-s   GAGGAGAGCTCTCTACGGACTACAATGGTCCACGAGATGCCAGTGGTATTGTAAAATACA 480
            ************************************************************
PMERp57-l   TGAGGTCACAGGTTGGACCAGCCTCTAAGGAGTTGACATCCGTGGAGGCAGCAGAAGCAT 540
PMERp57-s   TGAGGTCACAGGTTGGACCAGCCTCTAAGGAGTTGACATCCGTGGAGGCAGCAGAAGCAT 540
            ************************************************************
PMERp57-l   TCCTTGGTGCTGCTGAAGTTGGAGTCGTTTACTTTGGAGGAGATTCCAAACTTAAAGATG 600
PMERp57-s   TCCTTGGTGCTGCTGAAGTTGGAGTCGTTTACTTTGGAGGAGATTCCAAACTTAAAGATG 600
            ************************************************************
PMERp57-l   CTTTCCTAAAGGCTGCTGATAAGCTGAGGGAATCCATCCGTTTTGCACACTCCCTCGATG 660
PMERp57-s   CTTTCCTAAAGGCTGCTGATAAGCTGAGGGAATCCATCCGTTTTGCACACTCCCTCGATG 660
            ************************************************************
PMERp57-l   CCACTGTTAATGAAAAGTATGGGTACAGTGATGTTGTTGTACTTTTCCGACCAAAACACC 720
PMERp57-s   CCACTGTTAATGAAAAGTATGGGTACAGTGATGTTGTTGTACTTTTCCGACCAAAACACC 720
            ************************************************************
PMERp57-l   TGGAGAACAAATTTGAGCCTTCCTCTGTTGTATTTGAGGGATCGGCAGACAGGGCTGAGA 780
PMERp57-s   TGGAGAACAAATTTGAGCCTTCCTCTGTTGTATTTGAGGGATCGGCAGACAGGGCTGAGA 780
            ************************************************************
PMERp57-l   TTGAGTCTTTCATCAAAAAGAACTTCCATGGTTTGGTAGGACACCTAACGCAAGACACTG 840
PMERp57-s   TTGAGTCTTTCATCAAAAAGAACTTCCATGGTTTGGTAGGACACCTAACGCAAGACACTG 840
            ************************************************************
PMERp57-l   CTCAGGATTTCAAACCTCCAGTTGTGATTGCTTACTACAATGTTGATTACATCAAAAATG 900
PMERp57-s   CTCAGGATTTCAAACCTCCAGTTGTGATTGCTTACTACAATGTTGATTACATCAAAAATG 900
            ************************************************************
PMERp57-l   TTAAGGGTACAAATTACTGGCGCAATCGTGTCCTTAAGGTGGCACAAAACTTTGCTGATG 960
PMERp57-s   TTAAGGGTACAAATTACTGGCGCAATCGTGTCCTTAAGGTGGCACAAAACTTTGCTGATG 960
            ************************************************************
PMERp57-l   ACTTCAAGTTTGCCGTTGCCAATAAGGACGACTTCCAGCATGACCTCAATGAATATGGCC 1020
PMERp57-s   ACTTCAAGTTTGCCGTTGCCAATAAGGACGACTTCCAGCATGACCTCAATGAATATGGCC 1020
            ************************************************************
PMERp57-l   TTGATTATGTTCCTGGTGACAAGCCAGTAATTTGTGCACGTAATGCTAAAGCCCAGAAGT 1080
PMERp57-s   TTGATTATGTTCCTGGTGACAAGCCAGTAATTTGTGCACGTAATGCTAAAGCCCAGAAGT 1080
            ************************************************************
PMERp57-l   TTGTCATGCAGGAAGAATTTTTCAATGGATAACCTCCAAGCATTCCTCACCAATCTCAAGG 1140
PMERp57-s   TTGTCATGCAGGAAGAATTTTTCAATGGATAACCTCCAAGCATTCCTCACCAATCTCAAGG 1140
            ************************************************************
PMERp57-l   CGGGTGAGCTTGAGCCATATCTGAAGTCTGAGGCAGTGCCAACACAAGATGGCCCTGTCA 1200
PMERp57-s   CGGGTGAGCTTGAGCCATATCTGAAGTCTGAGGCAGTGCCAACACAAGATGGCCCTGTCA 1200
            ************************************************************
PMERp57-l   CTGTTGCTGTGGGTAAGAACTTCAATGAAGTTGTCTCTGATGAGCGTGATGCCCTCATTG 1260
PMERp57-s   CTGTTGCTGTGGGTAAGAACTTCAATGAAGTTGTCTCTGATGAGCGTGATGCCCTCATTG 1260
            ************************************************************
PMERp57-l   AATTCTATGCTCCTTGGTGTGGTCACTGCAAGAAATTAGCGCCCACCTATGATGAGCTGG 1320
PMERp57-s   AATTCTATGCTCCTTGGTGTGGTCACTGCAAGAAATTAGCGCCCACCTATGATGAGCTGG 1320
            ************************************************************
PMERp57-l   GAGAAGCGATGAAGGATGAAGATGTAGACATTGTGAAGATGGATGCCACTGCCAATGATG 1380
PMERp57-s   GAGAAGCGATGAAGGATGAAGATGTAGACATTGTGAAGATGGATGCCACTGCCAATGATG 1380
            ************************************************************
PMERp57-l   TTCCTCCTCAGTACAATGTTCAAGGCTTCCCCGCCATCTTCTGGAAACCCAAGGGTGGTG 1440
PMERp57-s   TTCCTCCTCAGTACAATGTTCAAGGCTTCCCCGCCATCTTCTGGAAACCCAAGGGTGGTG 1440
            ************************************************************
PMERp57-l   TTCCAAGGAATTACAACGGTGGCCGGGAACTGGACGATTTTGTCAAGTACATTGCCCAAC 1500
PMERp57-s   TTCCAAGGAATTACAACGGTGGCCGGGAACTGGACGATTTTGTCAAGTACATTGCCCAAC 1500
            ************************************************************
PMERp57-l   ATTCCACAAATGAACTGAATGGGTATGACCGCAAGGGGAAGGCAAAGAAAGGCAAGAAGA 1560
PMERp57-s   ATTCCACAAATGAACTGAATGGGTATGACCGCAAGGGGAAGGCAAAGAAAGGCAAGAAGA 1560
            ************************************************************
PMERp57-l   CTGAACTTTGAGAATAAGAAATTTTTTACGAGGAGGGTGTTGTCACCGCAGTGTGTTGTGT 1620
PMERp57-s   CTGAACTTTGAGAATAAGAAATTTTTTACGAGGAGGGTGTTGTCACCGCAGTGTGTTGTGT 1620
            ************************************************************
PMERp57-l   AAAGAGTGCTGCAGATATGTTAGTCATGTATGATTATTCCATGCATTATAATTTTTCCAT 1680
PMERp57-s   AAAGAGTGCTGCAGATATGTTAGTCATGTATGATTATTCCATGCATTATAATTTTTCCAT 1680
            ************************************************************
PMERp57-l   ATGTAGCAATCAAAGGATGTCTTTTTGATGTAATATCTGTATTGTTGTATGCAATTAAAC 1740
PMERp57-s   ATGTAGCAATCAAAGGATGTCTTTTTGATGTAATATCTGTATTGTTGTATGCAATTAAAC 1740
            ************************************************************
PMERp57-l   ATACGCAAATTTGTTTCCTCTCAAAGAGGCTTTCAGGTGAAAACTTGCTGACAGTATTAA 1800
PMERp57-s   ATACGCAAATTTGTTTCCTCTCAAAGAGGCTTTCAGGTGAAAACTTGCTGACAGTATTAA 1800
            ************************************************************
PMERp57-l   GATTGGAATATATTAAAATGAATTACCGGTGAAGCTTGTCCTGTGCCTCTGTTTGTGTGA 1860
PMERp57-s   GATTGGAATATATTAAAATGAATTACCGGTGAAGCTTGTCCTGTGCCTCTGTTTGTGTGA 1860
```

```
            ************************************************************
PMERp57-l   AAGCTAAGAATGGGTGAAGTTTTGTGCAATAGTTTTTCCTAAATTTCTCTTGTAAAGAAA 1920
PMERp57-s   AAGCTAAGAATGGGTGAAGTTTTGTGCAATAGTTTTTCCTAAATTTCTCTTGTAAAGAAA 1920
            ************************************************************
PMERp57-l   AATGACGCAATGCAGTGGACTTTTGAAGACTAAATTGATACAATCAATCATTCCGGCCCT 1980
PMERp57-s   AATGACGCAATGCAGTGGACTTTTGAAGACTAAATTGATACAATCAATCATTCCGGCCCT 1980
            ************************************************************
PMERp57-l   TGTCTGAAATACAAGAGCACAAGTTTATCCAGTGTAATGCATTCAGTTAACTTAATAAAT 2040
PMERp57-s   TGTCTGAAATACAAGAGCACAAGTTTATCCAGTGTAATGCATTCAGTTAACTTAATAAAT 2040
            ************************************************************
PMERp57-l   TTTATAATATATTTATTTACTTTTTTTTGAACCTTGGAAAAACTTGTGGTGGCCTGCATGG 2100
PMERp57-S   TTTATAATATAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2100
            ***********    *    *      **     *****                    *
PMERp57-l   TTATTGCACAAATTTGATTACTGTAATGAGGACTTGAATTATGTCTGCTAAACGGATTGA 2160
PMERp57-S   ------------------------------------------------------------
PMERp57-l   ATCTTGTGTGATTGTACTGTTAAAATTTCCAATGAAATATTTGATAGGAAATATTGACTT 2220
PMERp57-S   ------------------------------------------------------------
PMERp57-l   TTTGACATGTAAAAACAAATAAAAACATTTACTCAAAAAAAAAAAAAAAAAAAAAAAAAA 2280
PMERp57-S   ------------------------------------------------------------
PMERp57-l   AAAA 2284
PMERp57-S   ----
```

Figure 3.14 Pairwise alignment of *PmERp57-s* and *PmERp57-l*.

A)

B)
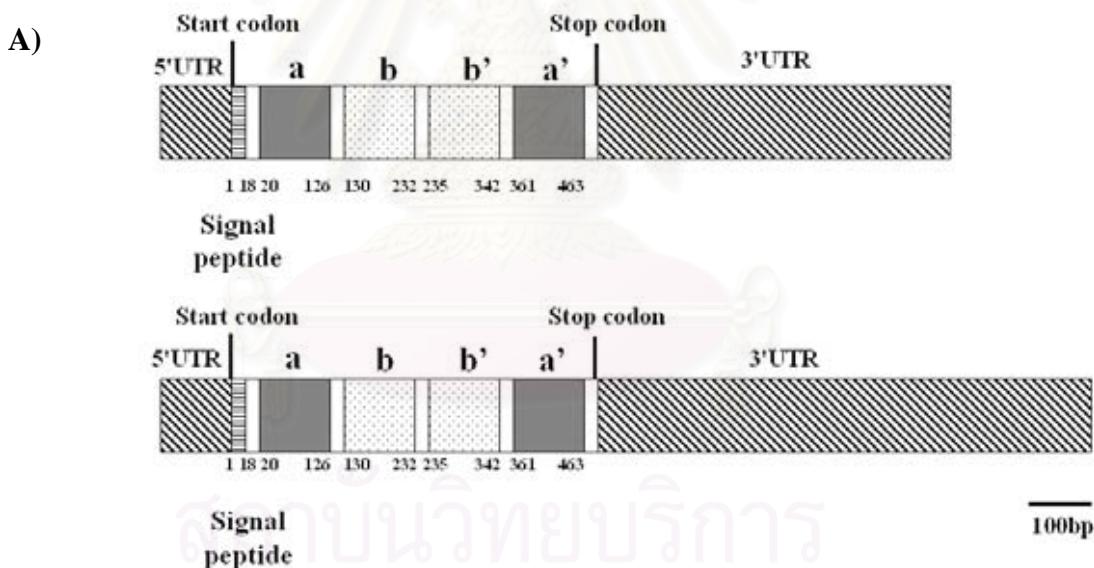
Name	Position	E-value
Signal peptide	1-18	-
Pfam: Thioredoxin	20-126	5.90e-46
Pfam: Thioredoxin	361-463	1.00e-46

Figure 3.15 Diagram illustrating the full length cDNA of the short (A) and long (B) isoforms *P. monodon ERp57*. Different isoforms of *ERp57* shared identical ORF but exhibited length polymorphism at the 3′ UTR. Two thioredoxin motif were found (known as a and a′ domains) in the deduced ERp57 protein. The scale bar is 100 bp in length.

A)

<u>**M**ATRLLILLLSLVAVALG</u>DDVLQLNDADFDGKVASYDTVLVMFYAP***WCGHCK***RLKPEFEKA
STTLKANDPPVYLAKVDCTDDGKDSCSRFGVSGYPTLKIFKGGELSTDYNGPRDASGIVKYMR
SQVGPASKELTSVEAAEAFLGAAEVGVVYFGGDSKLKDAFLKAADKLRESIRFAHSLDATVN
EKYGYSDVVVLFRPKHLENKFEPSSVVFEGSADRAEIESFIKKNFHGLVGHLTQDTAQDFKPPV
VIAYYNVDYIKNVKGTNYWRNRVLKVAQNFADDFKFAVANKDDFQHDLNEYGLDYVPGDK
PVICARNAKAQKFVMQEEFSMDNLQAFLTNLKAGELEPYLKSEAVPTQDGPVTVAVGKNFNE
VVSDERDALIEFYAP***WCGHCK***KLAPTYDELGEAMKDEDVDIVKMDATANDVPPQYNVQGFP
TIFWKPKGGVPRNYNGGRELDDFVKYIAQHSTNELNGYDRKGKAKKGKKTEL*

B)
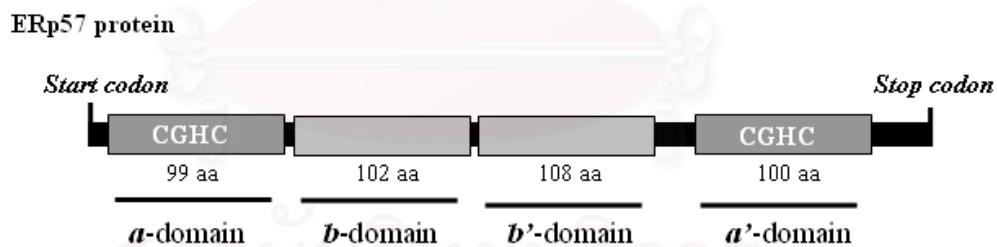
Figure 3.16 Organization of functional domains (a, b b′ and a′, respectively) in the deduce PmERp57 protein (A). The signal peptide is underlined. Catalytic sites in a and a′ domains (WCGHCK) are bold-italicized and highlighted. A schematic diagram of ERp57 also illustrated (B).

3.4.4 Phylogenetic analysis

A bootstrapped neighbor-joining tree illustrating relationships of *CRT* from different taxa was constructed (Figure 3.17). Invertebrate and vertebrate *CRT* were allocated to different groups. The tree topology of *CRT* was statistically supported as only on branch as the bootsrapping value < 50%. *PmCRT* was allocated to be a member of the invertebrate *CRT* and it clustered with that of the Chinese shrimp, *Fenneropenaeus chinensis*.

A phylogenetic tree indicated close relationships between vertebrate *CNX*. The topology of this tree was well supported by bootstrapping values and all branches showed the bootstrapping values > 50%. In addition, the bootstrapped *CNX* tree clearly supported that vertebrate *CNX* and *CMG* are closely related and shared the same evolutionary ancestor. Invertebrate *CNX* was quite diverse and *PmCNX* showed relatively large divergence from other invertebrate *CNX*. Accordingly, *PmCNX* should be regarded as a new member of this protein family (Figure 3.18).

For the construction of a phylogenetic tree for *ERp57*, protein sequences of *ERp57* from various species and those of protein disulfide isomerase (*PDI*) were retrieved. Like *CNX*, the topology of *ERp57* tree is also statistically supported. *PmERp57* clustered with *PDI* of the mud crab *Scylla paramanosain*. This strongly indicates that *ERp57* and *PDI* are members of the same protein family (Figure 3.19).
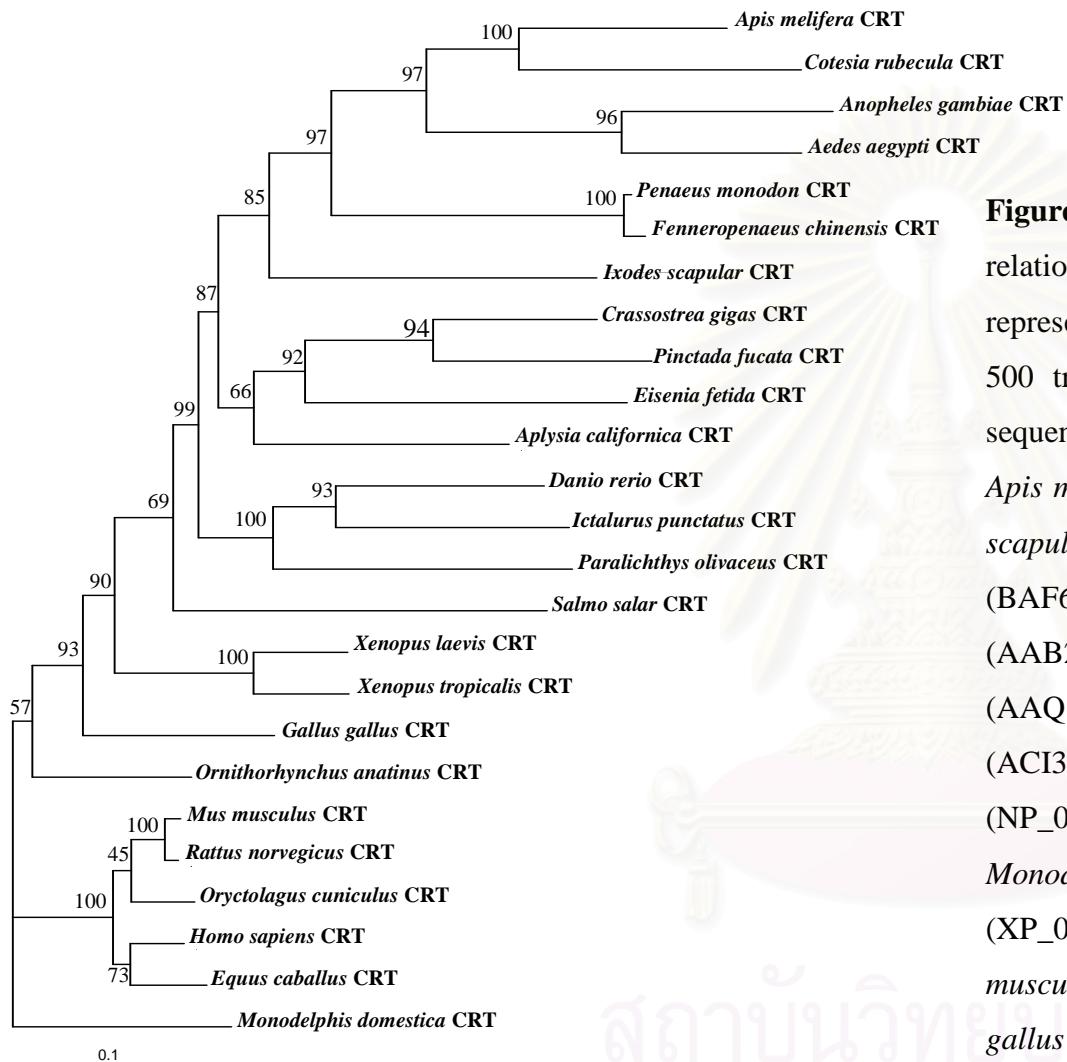
Figure 3.17 A bootstrapped neighbor-joining tree illustrating relationships between CRT of various taxa. Values at the node represent the percentage of times that the particular node occurred in 500 trees generated by bootstrapping the original data. Protein sequences of CRT from *Fenneropenaeus chinensis* (ABC50166), *Apis melifera* (XP_392689), *Cotesia rubecula* (AAN73309), *Ixodes scapular* (AAT99573, *Eisenia fetida* (ABI74618), *Crassostrea gigas* (BAF63639), *Pinctada fucata* (ABR68546), *Aplysia californica* (AAB24569), *Danio rerio* (AAH68336), *Ictalurus punctatus* (AAQ19852), *Paralichthys olivaceus* (ABG00263), *Salmo salar* (ACI33338), *Xenopus laevis* (NP_001080765), *Xenopus tropicalis* (NP_001001253), *Ornithorhynchus anatinus* (XP_001514084), *Monodelphis domestica* (XP_001377711), *Equus caballus* (XP_001504932), *Oryctolagus cuniculus* (NP_001075704), *Mus musculus* (NP_031617), *Rattus norvegicus* (NP_071794), *Gallus gallus* (AAS49610), *Anopheles gambiae* (AAL68781), *Aedes aegypti* (XP_001657748), *Homo sapiens* (NP_004334) were retrieved from the GenBank and compared with that of *P. monodon*.
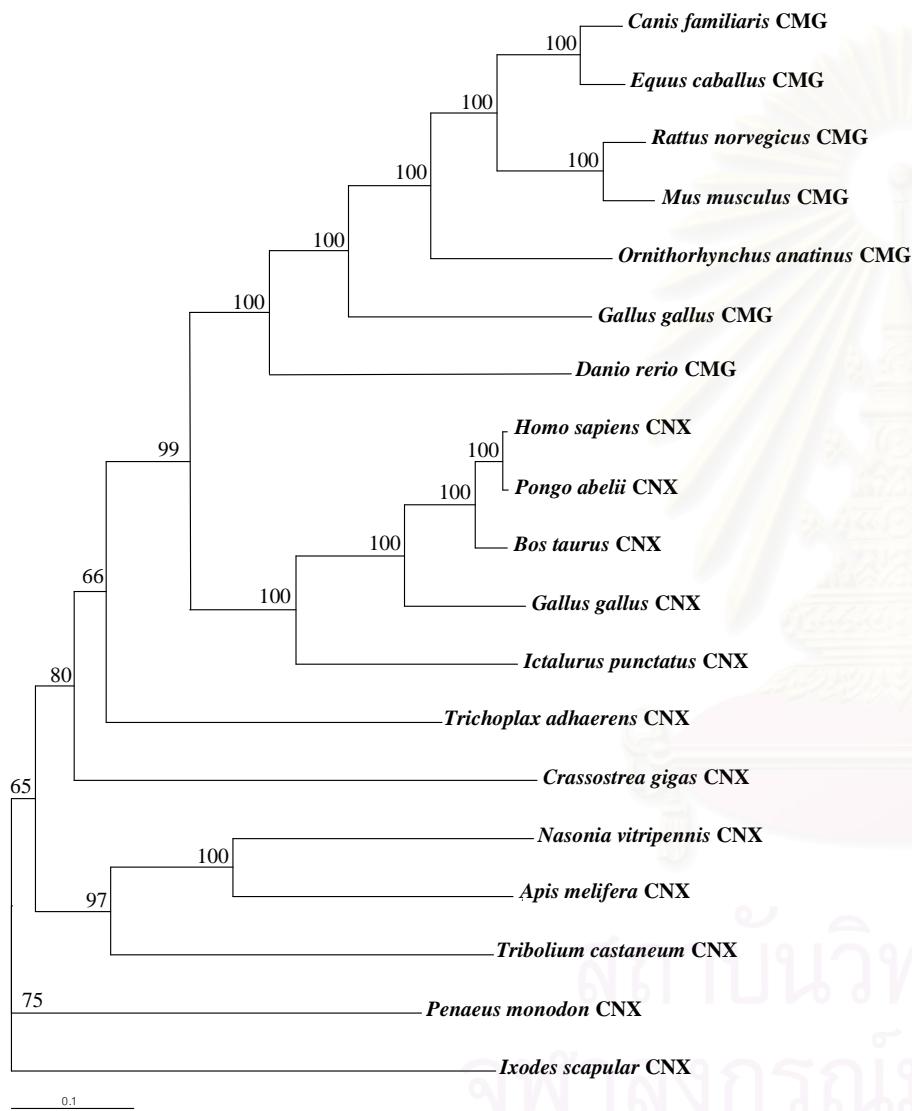
Figure 3.18 A bootstrapped neighbor-joining tree illustrating relationships between CNX of various taxa. Values at the node represent the percentage of times that the particular node occurred in 500 trees generated by bootstrapping the original data. Protein sequences of CNX from *Trichoplax adhaerens* (XP_002112490), *Apis mellifera* (XP_624907), *Nasonia vitripennis* (XP_001608138), *Tribolium castaneum* (XP_975051), *Ixodes scapularis* (EEC05774), *Gallus gallus* (CNX, NP_001025791; CMG, XP_420413), *Ornithorhynchus anatinus* (XP_001513600), *Rattus norvegicus* (NP_742005), *Mus musculus* (NP_031623), *Equus caballus* (XP_001497517), *Canis familiaris* (NP_001003232), *Danio rerio* (NP_001071009), *Ictalurus punctatus* (AAQ18011), *Bos taurus* (NP_001099082), *Homo sapiens* (NP_001737), *Pongo abelii* (CAH92563), *Crassostrea gigas* (BAF63638), were retrieved from the GenBank and compared with that of *P. monodon*.
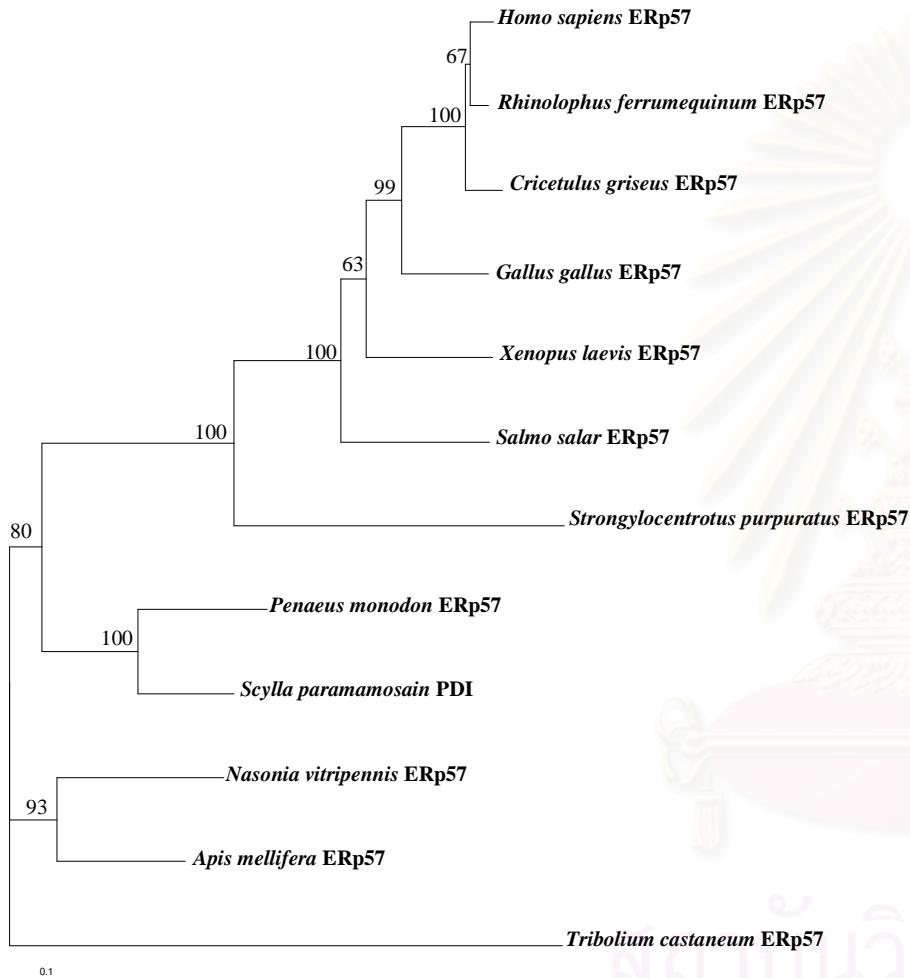
Figure 3.19 A bootstrapped neighbor-joining tree illustrating relationships between ERp57 of various taxa. Values at the node represent the percentage of times that the particular node occurred in 500 trees generated by bootstrapping the original data. Protein sequences of ERp57 from *Scylla paramamosain* (ACD44938), *Strongylocentrotus purpuratus* (XP_782981), *Cricetulus griseus* (AAL18160), *Rhinolophus ferrumequinum* (ACC69114), *Homo sapiens* (BAA11928), *Gallus gallus* (NP_989441), *Xenopus laevis* (NP_001080051), *Salmo salar* (ACI34304), *Apis mellifera* (XP_623282), *Nasonia vitripennis* (XP_001599732), *Tribolium castaneum* (XP_971685) were retrieved from the GenBank and compared with that of *P. monodon*.

3.5 Organization of *PmCRT*, *PmCNX* and *PmERp57* genes examined by genome walk analysis

3.5.1 Genomic organization of *Calreticulin*

Amplification of genomic DNA of *P. monodon* using primers ORFCRT-F & CRT-R and CRT-F and ORFCRT-R yielded the product of 1300 and 1050 bp fragments, respectively (Figure 3.20). These fragments were cloned and sequenced. The nucleotide length of these fragments was 1297 and 1036 bp, respectively (Figures 3.21 and 3.22). Nucleotide sequences of these two fragments were overlapped for 357 bp. A total of 1976 bp was obtained after sequence assembly.

In addition, 5´ and 3´ UTR of *PmCRT* were studied using genome walking analysis. A 781 bp fragment of the 5´ UTR (1st and 2nd CRT-Promoter primers with AP1 and AP2 primers) and a 1046 bp fragment of and 3´ UTR (CRT-F and AP1 primers) were obtained, respectively (Figures 3.23-3.25). Sequences of the genome walking fragments were assembled with those described above.

The complete gene sequence of *PmCRT* was 3043 bp in length (Figure 3.26). *PmCRT* was composed of 4 exons and 3 introns (Figure 3.27 and Table 3.1). The length of each exon varies from 85 bp (exon 1) - 830 bp (exon 4). The GC content reflected a slightly greater thermal stability in exons (42.8-46.7%) than in introns (28.6-34.1%) of the *CRT* (table 3.1). The exon/intron boundary sites determined by corresponding cDNA sequences were consistent with GT/AG rule. Introns 1 and 3 of the *PmCRT* gene interrupt the ORFs after the 1st or 2nd codon (type 1 intron), whereas the remain interrupted the ORFs between two codons (type 0 intron) (Figure 3.26).

The putative TATA box of this gene was found at -30 nt upstream from the transcription initiation site. The predicted CAAT regulatory box (CCAAT) was located between 95-91 nucleotide downstream from the transcription starting point. Moreover, potential *cis*-acting elements for multiple transcription factors including an ultraspiracle-TF (USP, GCTCA), a heat shock-TF (HS, GAATTTTCTC) and two endoplasmic reticulum stress elements (ERSE, CCAATXXXXXXXXXCCATG where X is any nucleotide) were found implying that *PmCRT* is regulated by both hormonal and environmental stimuli.

A)
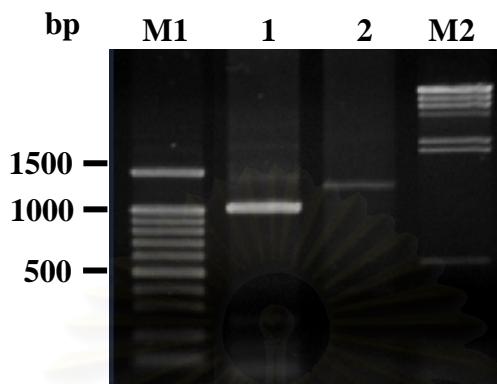
B)
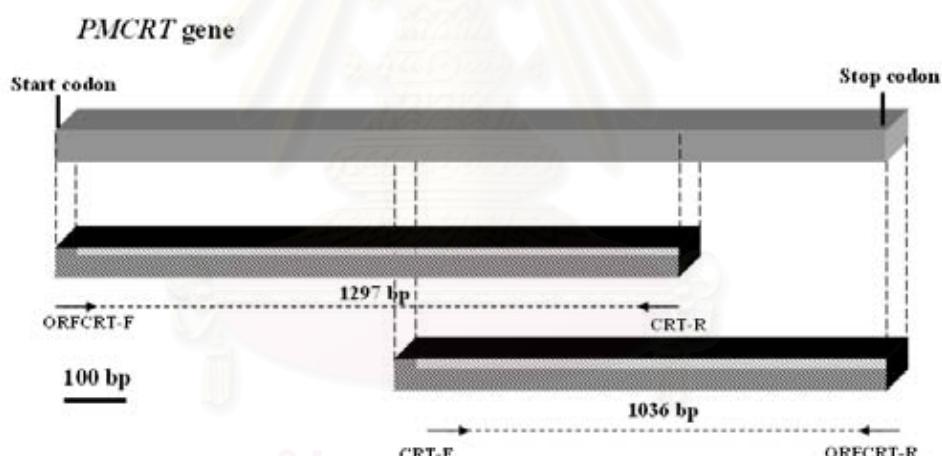
Figure 3.20 Genomic amplification of *PmCRT* using ORFCRT-R & CRT-F (lane 1) and ORFCRT-F & CRT-R (lane 2) primers. A 100 bp DNA ladder (lane M1) and λ-*Hind III* (lane M2) were used as the markers. Genomic orientation of each amplified fragment (B).

ATGAAGACCTGGGTTTTTCTTGCCCTATTTGGGGTTGCCCTAGTGGAATCTAAAGTATTTTTCGAAGAAAGA
TTCGACAGCCCTGGTAAGCTTGTATATTTTATATAAAAATCCTTCGTAATTTGTTGAATTACTTAGTTAAAT
GGTCGTCATTGACAGGCCCGGCTAATATGACCGGTGATTACAGACTTAACCCTCGGATATTAATTGTAAAAA
GGAATGCTTGTATGTTTATGTAGTGATTGTATAGTAGTTGACGACGTTTAAGGCTTTATAATGAAATTTTAA
AGTTGGAAATATTAGTTGCTTAAACTATAGCTATTTATGCCGCAAGTTGGACAAGTACAGAATTTTTTCTGT
GAAGGCGTTGCCACAAGTCATGTCTTCATCACTATTTATGTGAAAGTGGCATGTCACAACGTCATGTGATTT
GTGTAATATCTTCTTGATGATCAAGTGTGAGGAGTCCTGAAGATGCATTTTTTTTTTCTTGCAGATTGGGAG
AAAAATTGGGTTCAGTCTGCACACAAGGGGAAGGAGTTTGGACCCTTCAAGTTGACAGCTGGCAAATTTTAT
GGCGATGCTGAAAAGGATAAGGGAATCCAGACTGGACAGGTAAAATTTCTCTTGTTTAGAGACTTGCTTGTT
GTCACTCAACTTTCAGAATTTTTTTTATTATGTTGAGATATTTTGTGTATTCAACACTTCATGAATTTGCTT
AGAAATCAAGTAAAATTGAGTTGAAAGTATTTTCAAACACTGGTAAACTAATGGATAGTTCAAAACAGAAGG
TATGTGAACTTTCACATTTTGACTCCTTATTTATGTCTGCCTTTATTTTTTCAGGATGCCCGCTTTTATGGT
CTTTCTACGAAGTTTGAGCCCTTCAGTAATAAGGATTCCCCACTTGTCATCCAGTTTTCTGTAAAACATGAA
CAGA~~ACATTGACTGTGGTGGAGGATATCTAAAGGTCTTCGATTGCTCTTTAGACCAGAAAGACATGCACGGA
GAGTCGCCATACCTCATTATGTTTGGTAAGTAGCCTTCATAGCTTTAGATCTTATCCTTTAATTTTTTAAAC
TTTGACTTCATAAGATGCTTTAATAGTGTTGCTGTAGACATACTAATTTCTTTTGTCTCTTTAAAGGTCCTG
ATATCTGTGGCCCAGGCACCAAGAAGGTTCATGTAATCTTCAATTACAAGGGTGAGAACCATCTGATCAAGA
AGGAAATCCGTTGCAAGGATGACGTATTTTCCCATCTGTATACCCTCATTGT~~**CAATCCTGACAACACCTACG
A**

Figure 3.21 Nucleotide sequence of the PCR product amplified with ORFCRT-F (underlined, above) and CRT-R primers (underlined, below). Shared sequence with that of the fragment generated from ORFCRT-R & CRT-F primers is highlighted.

ACATTGACTGTGGTGGAGGATATCTAAAGGTCTTCGATTGCTCTTTAGACCAAAAAGACATGCACGGAGAGT
CGCCATACCTCATTATGTTTGGTAAGTAGCCTTCATAGCTTTAGCCATTATCCCTTAATTTTTAAAACTTTG
ACTTCATAAGATGCTTTAATAGTGTTGCTGTAGACATACTAATTTCTTTTGTCTCTTTAAAGGTCCTGATAT
CTGTGGCCCAGGCACCAAGAAGGTTCATGTAATCTTCAATTACAAGGGTGAGAACCATCTGATCAAGAAGGA
AATCCGTTGCAAGGATGACGTATTTTCCCATCTGTATACCCTCATTGTCAATCCTGACAACACCTACGAAGT
TCTTATTGACAATGAAAAAGCTCACTCTGGTGAACTCGAGGAGGACTGGGACTTCCTTCCTCCCAAGAAGAT
CAAGGACCCAAAAGCCAAGAAGCCCGACGATTGGGATGACCGCCCCACCATTGCTGATCCTGACGATACTAA
GCCTGAAGATTGGGACCAACCTGAACACATTCCTGATCCTGATGCCACCAAACCTGAGGACTGGGATGATGA
AATGGATGGCGAGTGGGAACCACCCATGATTGACAATCCTGACTACAAGGGGGAATGGAAGCCTAACCAGAT
TGATAACCCTGATTACAAGGGTCCATGGATTCCCCCTGAAATTGACAACCCAGAAAACCCACCTGACCCAAA
AATCTACAAGTATGATGAGGTCTGCGCTCTTGGTTTGGAATTTTGGCAGGGAAAATCTGGTACTATCTTTGA
CAACTTCCTCATCTCAAATGATCCTGAAGAAGCCCGCAAGATTGGTGAAGAGACTTGGGGTGCTACTAAAGA
TGCAGCTAAGAAGATGAAGGATGAACAGGATGAAGAGGAGCGAAAGAGAGCAGAGGAAGAAGCTAAGGCAGC
TGCTGATGCTGAGAAGGATGAGGACGATGATGACGACGACGATCTTGGCGATGAAGACGAAGATGATCTTGA
T**AATGATCTTGAACATGACGAGCTGTAA**

Figure 3.22 Nucleotide sequence of the PCR product amplified with CRT-F (underlined, above) and ORFCRT-R primers (underlined, below). Shared sequence with that of the fragment generated from ORFCRT-R & CRT-F primers is highlighted.

A)

B)
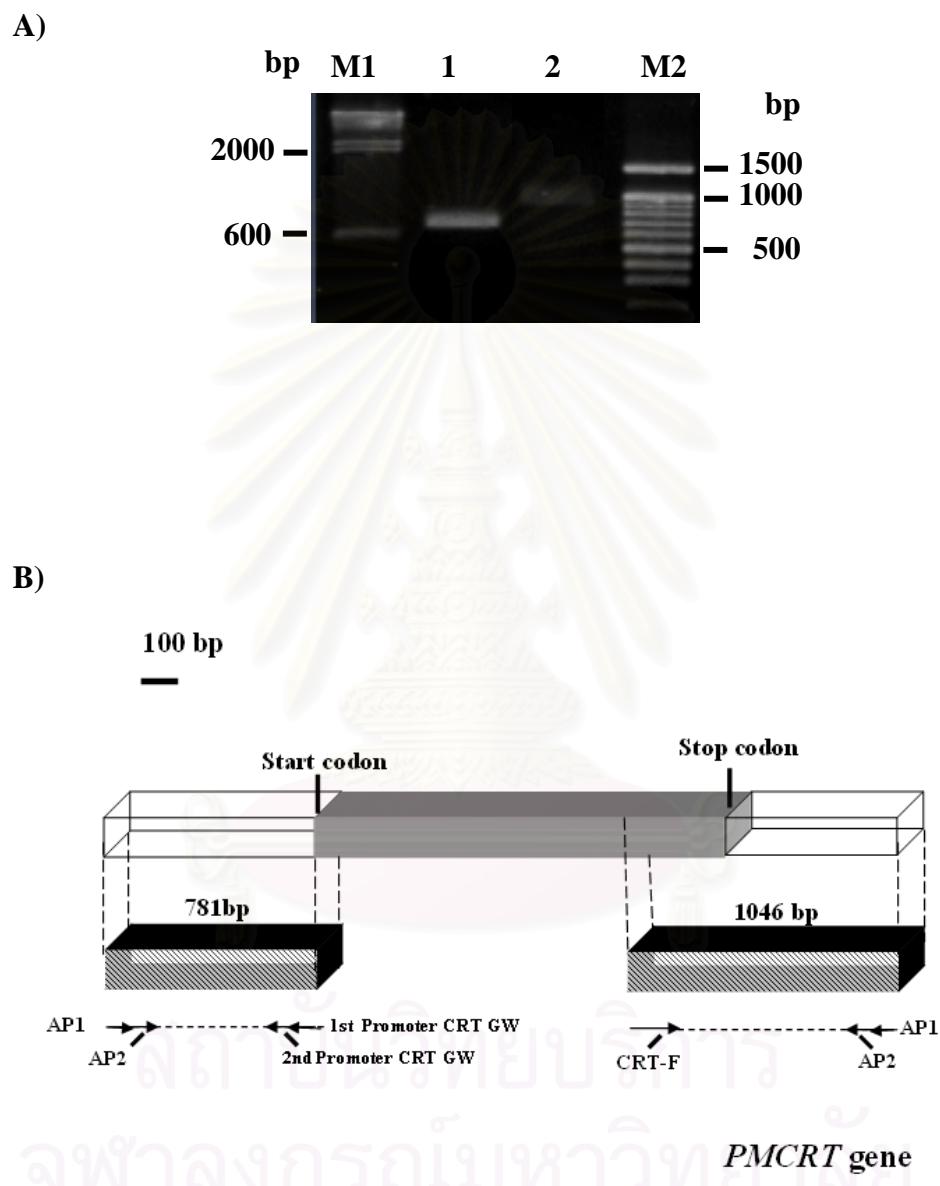
Figure 3.23 Genome walking analysis for isolation of 5´ (lane 1) and 3´ (lane 2) UTRs of *PmCRT*. λ *Hind III* (lane M2) and a 100 bp DNA ladder (lane M1) were used as the markers. Genomic orientation of each amplified fragment (B).

GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGTTCACTATAGGGCACGCGTGGTCAC
GGCCCTGGCTGGTGTAAGGAATTTTCTCACTTTCAAATGGTAAATATATTTGGAAATAGTACTTTGGCATTG
GTTATTCTGCTTTAGATCCCCCTCCGCTCTTACGTCGGACCAAAACAAATACTTGGGCCAATCAGCGACATC
CATGACATAACCCTAACCAATCACAGACGTCCATGTGAGGGAGTTTAACCAATCAGCTTTCTCCAAGTTAGA
AACAACACGCCTATGTTGATATCATAGATGACAGAATCAGGACCGTACTTCCTCGTGGAGATTTAGCTATTG
TGTACATTTCACTGGCATTCAAGAGAAAGAAATCGATGTTTAAAGCAAGAATTACAATTTATGATAATTTC
TACGTACACGAACATAGCCAATGCCTTCAACATAAGAGTTCAACTGAAAAGAGAAACATCATATACTCCTTC
GAAGAGCACTTATAAAGAGATCTGAATCTACTTGTTCACAAATTACATCACTTATGATCCTCTGCCGCGAC
AACCAATCGGCGTTCTCCACTTGTATCAGCTTCCTAGCAAAATCTGCGGCAGTCGACCGTCGGGGAGCAGTG
AAGGTGGGGACTTACTCGAGTGACAGTCAAGAGCGAAGACGGAACG**ATG**AAGACCTGGGTTTTTCTTGCCCT
ATTTGGGGTTGCCCTAGTGGAATCTAAAGTATT**TTTCGAAGAAAGATTCGACAGCCCTGGT**

Figure 3.24 Nucleic acid sequence of the amplified fragment generated by 1st-5´GWCRT (underlined, above) and AP1 (underlined, below) and AP2 (underlined and highlighted) primers. The start codon is illustrated in boldfaced and double-underlined.

ATTGGGACCAACCTGAACACATTCCTGATCCTGATGCCACCAAACCTGAGGACTGGGATGATGAAATGGATG
GCGAGTGGGAACCACCCATGATTGACAATCCTGACTACAAGGGTGAATGGAAGCCTAAGCAGATTGATAACC
CTGATTACAAGGGTCCATGGATTCACCCTGAAATTGACAACCCAGAATACACACCTGACCCAGAGATCTACA
AGTATGATGAGGTCTGTGCTCTTGGTTTGGATCTTTGGCAGGTAAAATCTGGTACTATCTTTGACAACTTCC
TCATCTCAAATGATCCTGAAGAAGCCCGCAAGATTGGTGAAGAGACTTGGGGTGCTACTAAAGATGCAGCTA
AGAAGATGAAGGATGAACAGGATGAAGAGGAGCGAAAGAGAGCAGAGGAAGAAGCTAAGGCAGCTGCTGATG
CTGAGAAGGATGAGGACGATGATGACGACGACGATCTTGGCGATGAAGACGAAGATGATCTTGATAATGATC
TTGAACATGACGAGCTG**TAA**AGTTATATTATTTTATTAAGGAAGTTATTATTTAAATAGCCTGTTGTACTAT
TTAAACATCAAAGTACAATTAACTGAACCTTTTGGTTGTACATTCTTGTAAATACCAAGGGCTTCGGTTAAT
TCTAGTCATGGAATCTTTTGTTGTGGGCAAACTAAAATCCAAGCATTCCAGTAGCTGAAGCTGATTGGAGGT
TCCTTGACAAGAAGCCACCTATTTAGGTGGTATTCATAGTACATCACTCATCATTAATCATCCCATCATCAT
GTATGCAGCACAGTCCTGTATGTTTGTACGAGACAAAAGTGTGTCAGTACTGATGTCAATTTAAACATGATT
CAGTATCGCTGATTGCACAATGTGTGTCTGAGTGAGCGATTTCCAATAAACACAATCAAAAAAAAAAAAAAA
AAAAAAAAAGTACTCTGCGTGGATACCACTGCTTGCCCTATAGTGAGTCGTATTAGAATGGG**ACCAGCCCGG**
GCCGTCGACCACGCGTGCCCTATAGTGAGTCGTATTAC

Figure 3.25 Nucleic acid sequence of the amplified fragment generated by 1st-3´GWCRT (underlined, above) and AP1 (underlined, below) primers. The stop codon is illustrated in boldfaced and double-underlined

```
TCACTATAGGGCACGCGTGGTCACGGCCCTGGCTGGTGTAAGGAATTTTCTCACTTTCAA - 437
ATGGTAAATATATTTGGAAATAGTACTTTGGCATTGGTTATTCTGCTTTAGATCCCCCTC - 377
CGCTCTTACGTCGGACCAAAACAAATACTTGGGCCAATCAGCGACATCCATGACATAACC - 317
CTAACCAATCACAGACGTCCATGTGAGGGAGTTTAACCAATCAGCTTTCTCCAAGTTAGA - 257
AACAACACGCCTATGTTGATATCATAGATGACAGAATCAGGACCGTACTTCCTCGTGGAG - 197
ATTTAGCTATTGTGTACATTTCACTGGCATTCAAGAGAAAGAAATCGATGTTTAAAGCAA - 137
GAATTACAATTTATGATAATTTCTACGTACACGAACATAGCCAATGCCTTCAACATAAG - 77
AGTTCAACTGAAAAGAGAAACATCATATACTCCTTCGAAGAGCACTTATAAAGAGATCTG - 17
                 +1
AATCTACTTGTTCACAAAATTACATCACTTATGATCCTCTGCCGCGACAACCAATCGGCG   44
```

```
TTCTCCACTTGTATCAGCTTCCTAGCAAAATCTGCGGCAGTCGACCGTCGGGGAGCAGTG    104
AAGGTGGGGACTTACTCGAGTGACAGTCAAGAGCGAAGACGGAACGATGAAGACCTGGGT    164
TTTTCTTGCCCTATTTGGGGTTGCCCTAGTGGAATCTAAAGTATTTTTCGAAGAAAGATT    224
CGACAGCCCTGgtaagcttgtatattttatataaaaatccttcgtaatttgttgaattac    284
ttagttaaatggtcgtcattgacaggcccggctaatatgaccggtgattacagacttaac    344
cctcggatattaattgtaaaaaggaatgcttgtatgtttatgtagtgattgtatagtagt    404
tgacgacgtttaaggctttataatgaaattttaaagttggaaatattagttgcttaaact    464
atagctatttatgccgcaagttggacaagtacagaattttttctgtgaaggcgttgccac    524
aagtcatgtcttcatcactatttatgtgaaagtggcatgtcacaacgtcatgtgatttgt    584
gtaatatcttcttgatgatcaagtgtgaggagtcctgaagatgcattttttttttcttgc    644
agATTGGGAGAAAAATTGGGTTCAGTCTGCACACAAGGGGAAGGAGTTTGGACCCTTCAA    704
GTTGACAGCTGGCAAATTTTATGGCGATGCTGAAAAGGATAAGGGAATCCAGACTGGACA    764
Ggtaaaatttctcttgtttagagacttgcttgttgtcactcaactttcagaattttttttt    824

attatgttgagatattttgtgtattcaacacttcatgaatttgcttagaaatcaagtaaa    884
attgagttgaaagtattttcaaacactggtaaactaatggatagttcaaaacagaaggta    944
tgtgaactttcacattttgactccttatttatgtctgcctttatttttttcagGATGCCCG    1004
CTTTTATGGTCTTTCTACGAAGTTTGAGCCCTTCAGTAATAAGGATTCCCCACTTGTCAT    1064
CCAGTTTTCTGTAAAACATGAACAGAACATTGACTGTGGTGGAGGATATCTAAAGGTCTT    1124
CGATTGCTCTTTAGACCAGAAAGACATGCACGGAGAGTCGCCATACCTCATTATGTTTGg    1184
taagtagccttcatagctttagatcttatcctttaattttttaaactttgacttcataag    1244
atgctttaatagtgttgctgtagacatactaatttcttttgtctctttaaaggtcctgaT    1304
ATCTGTGGCCCAGGCACCAAGAAGGTTCATGTAATCTTCAATTACAAGGGTGAGAACCAT    1364
CTGATCAAGAAGGAAATCCGTTGCAAGGATGACGTATTTTCCCATCTGTATACCCTCATT    1424
GTCAATCCTGACAACACCTACGAAGTTCTTATTGACAATGAAAAAGCTCAGTCTGGTGAA    1484
CTCGAGGAGGACTGGGACTTCCTTCCACCCAAGAAGATCAAGGACCCAGAAGCCAAGAAG    1544
CCCGACGATTGGGATGACCGCCCCACCATTGCTGATCCTGACGATACTAAGCCTGAAGAT    1604
TGGGACCAACCTGAACACATTCCTGATCCTGATGCCACCAAACCTGAGGACTGGGATGAT    1664
GAAATGGATGGCGAGTGGGAACCACCCATGATTGACAATCCTGACTACAAGGGTGAATGG    1724
AAGCCTAAGCAGATTGATAACCCTGATTACAAGGGTCCATGGATTCACCCTGAAATTGAC    1784
AACCCAGAATACACACCTGACCCAGAGATCTACAAGTATGATGAGGTCTGTGCTCTTGGT    1844
TTGGATCTTTGGCAGGTAAAATCTGGTACTATCTTTGACAACTTCCTCATCTCAAATGAT    1904
CCTGAAGAAGCCCGCAAGATTGGTGAAGAGACTTGGGGTGCTACTAAAGATGCAGCTAAG    1964
AAGATGAAGGATGAACAGGATGAAGAGGAGCGAAAGAGAGCAGAGGAAGAAGCTAAGGCA    2024
GCTGCTGATGCTGAGAAGGATGAGGACGATGATGACGACGACGATCTTGGCGATGAAGAC    2084
GAAGATGATCTTGATAATGATCTTGAACATGACGAGCTGTAAAGTTATATTATTTTATTA    2144
AGGAAGTTATTATTTAAATAGCCTGTTGTACTATTTAAACATCAAAGTACAATTAACTGA    2204
ACCTTTTGGTTGTACATTCTTGTAAATACCAAGGGCTTCGGTTAATTCTAGTCATGGAAT    2264
CTTTTGTTGTGGGCAAACTAAAATCCAAGCATTCCAGTAGCTGAAGCTGATTGGAGGTTC    2324
CTTGACAAGAAGCCACCTATTTAGGTGGTATTCATAGTACATCACTCATCATTAATCATC    2384
CCATCATCATGTATGCAGCACAGTCCTGTATGTTTGTACGAGACAAAAGTGTGTCAGTAC    2444
TGATGTCAATTTAAACATGATTCAGTATCGCTGATTGCACAATGTGTGTCTGAGTGAGCG    2504
ATTTCCAATAAACACAATCAAAAAAAAAAAAAAAAAAAAAAAAAA    2564
```

Figure 3.26 Organization of the *PMCRT* gene. Nucleotides of exons are capitalized while those of introns are highlighted and shown in lower letters. The TATA box and start and stop codons are bold-italicized and highlighted. The CAAT box and transcription initiation site are double-underlined and bold-capitalized, respectively. The poly A additional signal site is underlined. 5´ UTR of *PMCRT* contained two endoplasmic reticulum stress elements (underline-italicized), an ultraspiracle-TF (bold and double under lines) and a heat shock-TF (boxed letter).

Table 3.1 GC content and length of exons and introns of the *PMCRT* gene

Compositions of gene	Genomic DNA		mRNA	GC content (%)		
Exon	(No. of nucleotides)		(No. of nucleotides)			
1	1-85	(85 bp)	1-85	43.5	%	
2	496-614	(119 bp)	86-204	46.2	%	
3	846-1032	(187 bp)	205-391	42.8	%	
4	1146-1975	(830 bp)	392-1221	46.7	%	
Intron	Genomic DNA		Type	GC content (%)		GT/AG rule
	(No. of nucleotides)					
1	86-496	(411 bp)	1	34.1	%	Yes
2	616-846	(231 bp)	0	28.6	%	Yes
3	1034-1153	(120 bp)	1	30.0	%	Yes
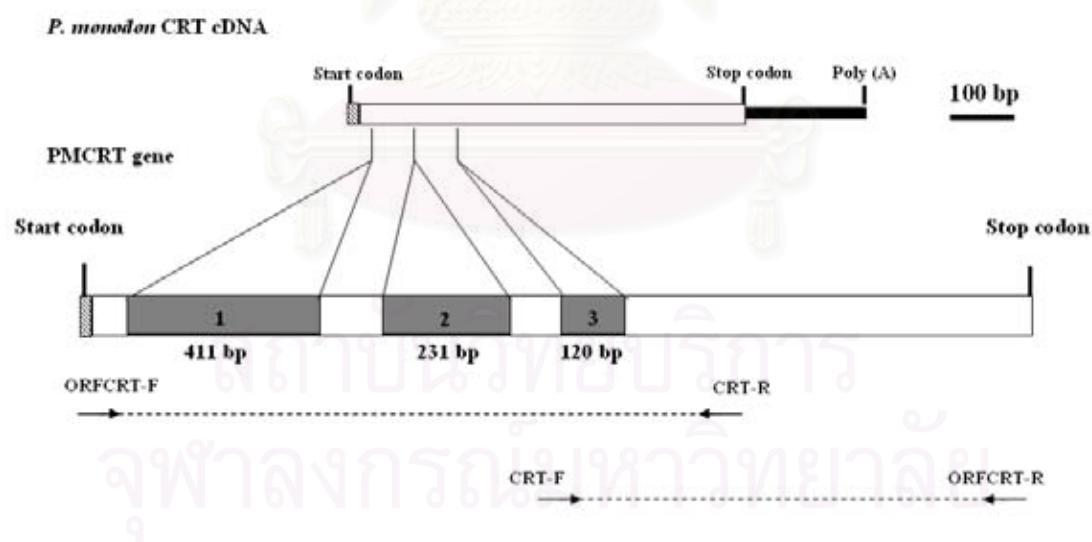
Figure 3.27 Schematic diagrams of *PmCRT* cDNA and gene. Coding regions are represented in open bars. Introns (with numbers) are gray shaded. Primers used for amplification of genomic *PMCRT* are illustrated.

3.5.2 Genomic organization of *Calnexin*

Genomic organization of *PMCNX* was characterized by genome walking analysis. Initially, ORFCNX-F and ORFCNX-R were used along with the adaptor primer (AP1 or AP2) (Figure 2.28). A 514 bp fragment was obtained from ORFCNX-F and AP1 primer whereas ORFCNX-R and AP1 primer generated a 1783 bp product (Figures 2.29-2.30).

Subsequently, primer 1st 5´cont GW-CNX & AP1 and 1st 3´ cont GW-CNX & AP1 were used for the primary genome walking PCR of *PmCNX*. The primary PCR products were diluted 50 fold and used as template for the secondary PCR. A fragments of 1164 bp fragments was obtained from 2nd 5´cont GW-CNX and AP2 (Figure 2.31 and 2.32). However, no amplification band was obtained from amplification of *P. monodon* genomic DNA with 2nd 3´ cont GW-CNX and AP2 primers.

After that the 1st and 2nd GW CNX-new primers were designed from the sequence obtained from 2nd 5´cont GW-CNX and AP2 for further analysis. The 1st GW CNX-new and AP1 were used for the primary PCR whereas the 2nd GW CNX-new and AP2 were used for the secondary genome walking amplification. A fragment of 659 bp in length was obtained (Figure 3.33). Nucleotide sequence of this fragment is shown by Figure 3.34. Many primers were designed and applied for further characterization of *PmCNX* but the amplificatiob product was not obtained.
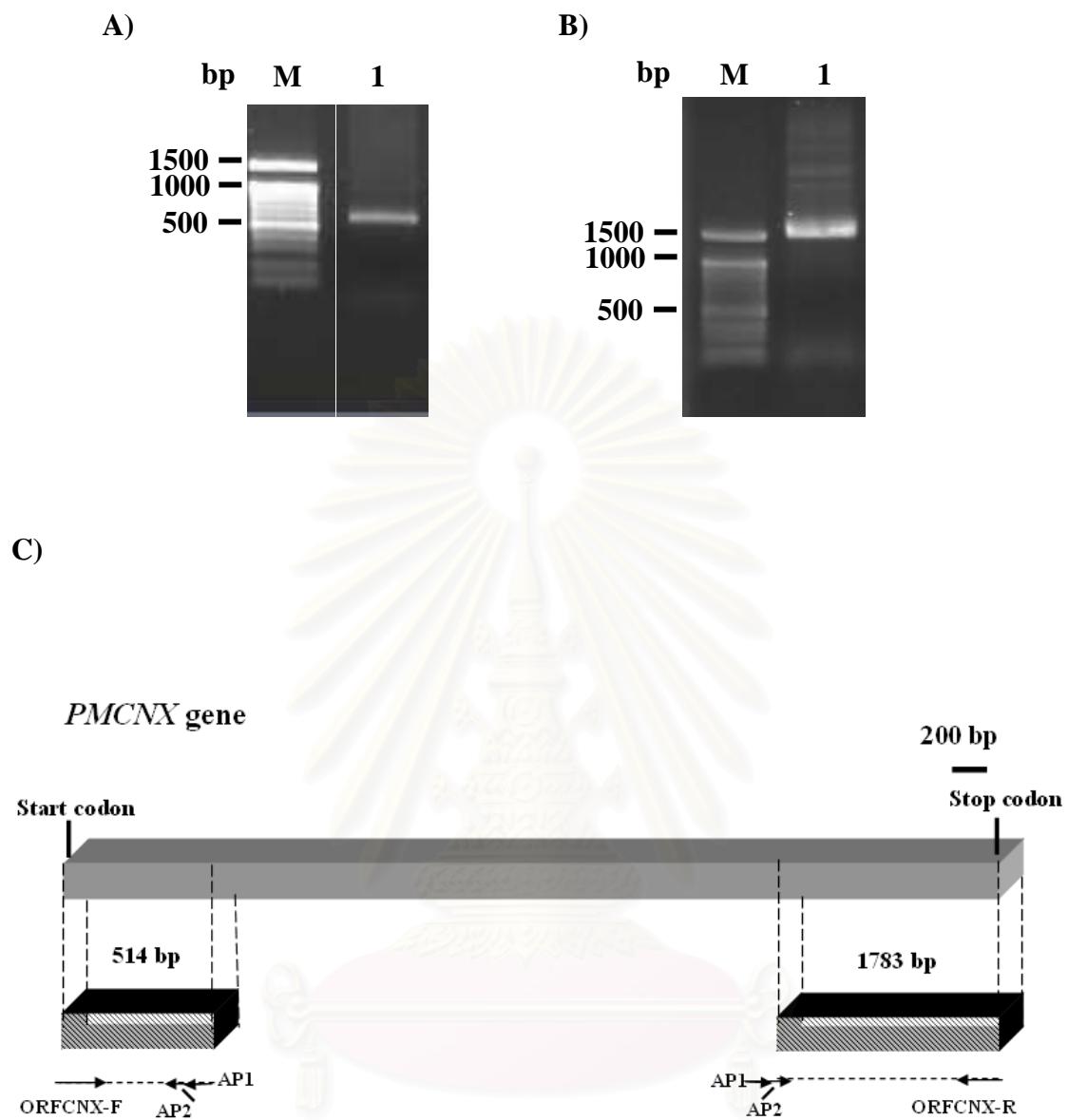
Figure 3.28 Genome walking analysis of *PmCNX* using ORFCNX-F & AP1 (lane 1, A) and ORFCNX-R & AP1 (lane 1, B). A 100 bp DNA ladder (lane M) was used as the markers. Genomic orientation of the amplicon resulted from ORFCNX-F & AP1 and ORFCNX-R & AP1 primers are shown (C).

GAAGCGTGAGTCGTCCATTGACTAAATACCTGAAATACGACCGATTTAGGAGGAGCG**ATG**AAGTCGAGGTGG
CAGAGAAAAGCAGTATTAGCACTGCTAGTTCTTGGCCTGCTGTTACCTTTTGGTATTAAAGCAGATGACGAT
GACGATGAAGAAGCTGTTGTCACGGAAGAGCAAACAGTAAGTTATTGATATCTTGAACAATTTTGGCTTTAA
AGAACAATATGCAGAATTATAAAAAATGGATTTCTAAATCTATTGATATTACATAATGTATGGTGATTTATG
GGAAGAGATTATGATTTCAGATAGTGGATAACCTACTTCTGTGCTTTTATTGGTCAGCAGGGTCTTCAAAAA
TCAATAACCATATGTAGTGGATTCTGGGCCTTGGGGAACCTTTAACTTGGAATATATGACCTTTTTTTTCA
ATAATGATAAGAATTCCACCAGCCCGGGCCGTCGACCAGCCCGGGCCGTCGACCACGCGT**GCCCTATAGTGA**
GTCGTATTAC

Figure 3.29 Nucleotide sequences of the genome walking fragment of *PmCNX* amplified by ORF-F & AP1. The start codon is double-underlined. Primer sequences are boldfaced and underlined. The adaptor sequence is highlighted.

GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGTCGACGGCCCGGGCTGGTATTTTTA
AAAATGAATGAAAAGGGTAAACAGGTGAGACAGGTAGTACGTGTAGAGCCATCTATGTATAAGAAAAAATTT
ATAAAATAAACTCACTGGGGGCATGGCGTGTAAGTACATGCTGTGCCTGTTGGCTTTGGGTTAATAAAACAA
GATGTATTACACTAATTTCATTACTTTTTTTGTAGCAAACTTTTTGGGGGAAAATCTTGAAGCTAACCAATT
ACAAGCCTGTAATGTGGGCACTCTACGTACTTTATATGATTATTCCAACTGGATTAATAATCTTTTGTCTAC
TGAAGAGAGTATATGAGGTATTTTTCTGGTAAAATTTTTTGCTAGATAAGCAACAGGATTACATAAGGGCTA
GTGCAAGAAACTATGTTAGCTAGGGTGAGCATATGAATTGCTGTGCTAGTTCAGGATGTGACATTTGTGTAT
GAGTAATTATATGAAAAGTAGGGGTAGGTTATTAGTAAGAGTTGGTTGTTGCTTATAGTGTTGCTTTATGTA
AAATATATTTTTTTTTTTATTTTTAAAAATCTTTCATACTTTAACATATGAACCAAAGCCTTTTTGATACACA
CCTTCTAGATACAGATTATTTGGCATTCTTTTCAGTGCTTTAATACAAGTTCCCTGTGTGATTGCCCTGCTT
TAGAGTAATAGCCACATTCATCTGTATCTGTGTATTCTTTATTGTAGACTGGAGTTATTCGACGAATCATCA
ACTACTCCAATAAGAATCCATGGCTATATGGTATCTACGTACTCCTAGTGGCCATTCCTGTAGTGTTGATAT
TTGCCTGTTGTTGTGCAGAAGCAAAGGTAGAGTATGATAATAGCTGTCAAATGTAATGATTTCTTTGCTTTT
AAGTGTAAAAACTAAAAGATTTAACATTAATAATAGTGTAGACAAGTCCTTACTGTCACACGTGTTTGAAAA
TGAATGTGCAATATTACTAGAACTTCAAAAATATATATTTTGTGTCAGTTAATTATTTTGGATCTCTCAGGA
CACCAAGGAAGATAAGGATGCAGAGAGGAAAAAGACCGATGCACCTTCTCCTGATGATCCACAATCAGAGCA
AAACAAAGAAGATTCTAGTCAAACTGCAGCAGATGATGATGCTCCTGGAAGTGGGGATGAAGCTGAAGGGGA
TGAAGACAAAGGTGATGAAGAGGGAGAGGAGGAAGAAGATGAGGAAGAGGAGGAGGAAGCAGAGAAAGCTGA
TGCAGCTGAGGTATTGATCCAAAACCAGTTATCTTAACTATTTTCTTGTGTTGCCCTACTTTATATAAATGT
TTGAATCTGCATTTCTCCATCTAATTGTCTTTAAAAGTTTGAATGTTGTTCTCATTTCTCTCAATGTGCAAA
CCACTTTATCAGATTTTTAAAACACTGTTTTGTGTTTCATTTCTTTGCTGATTTCCCAGTTGTACATTTTTT
GAGTACCACATATTGTTCATATTTGTCTAGGGTTGCTTATTTTCAGATATTGTAATAACATAACTGGTGCTG
CATATCCAAGTAAAGAGTACTTTTTTGGTTTTGCCAAGTATTTTTTTTTTTTTTCAAAAACATGATCCATAG
GTGTATACCAATGTTTCAGTAGTGTATCACATACAAGAAAATTTGTACTTTTCTAATATTAATTTTTTTCTTT
CAAGGAGGTCCAGACTCGCACATCACCAAG**GCTGCGCAAGTCCCGGAGAGATTAA**

Figure 3.30 Nucleotide sequences of the genome walking fragment of *PmCNX* amplified by ORF-R & AP1. Primer sequences are boldfaced and underlined. The adaptor sequence is highlighted.

A)

B)
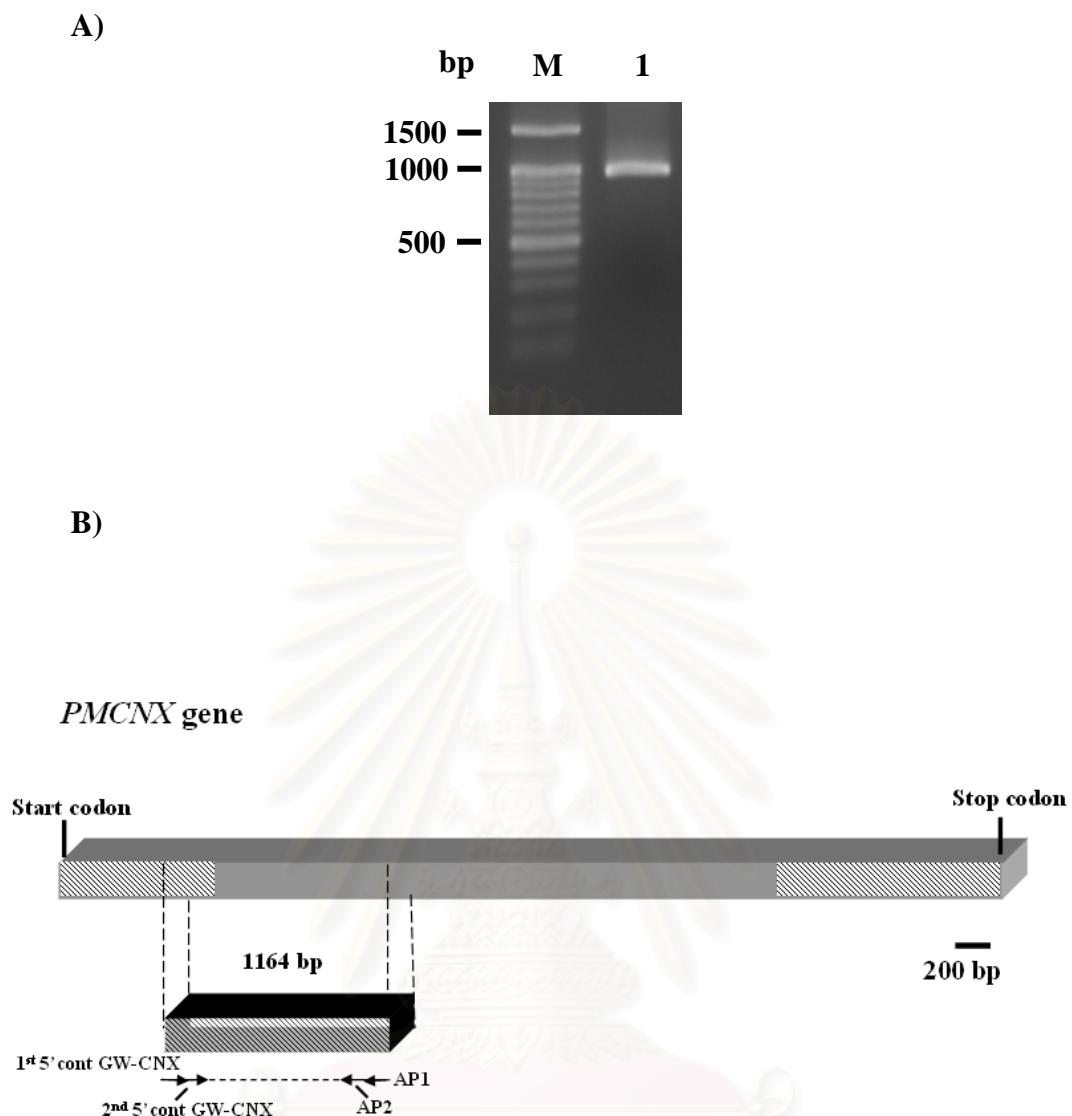
Figure 3.31 Genomic organization of *PMCNX* was further carried out using 2nd-5´ cont GW-CNX and AP2 (lane 1, A). A 100 bp DNA ladder (Lane M, A) was used as the marker. Genomic orientation of the amplified fragment (B).

AACCATATGTAGTGGATTCTGGGCCTTGGGGGAACCTTTAACTTGGAATATATGACCTTTTTTTTTCAATAAT
GATAAGAATTCCACCAGCAGAATATACATAAGGATTTAACCTTCTGTATGAACTTTAGAAATTGTGTACTAG
GACAAAATCAGAACAGAACATTTTTGTAATGAGCTGTACAAAAGAATGGTCGAGTACAAAGTCAGGATGAGA
CCTGGACACAGCTTGAGGGCTTCAGGGGCTTCTTCCTGCTAAGCTGTGCGGAAGTAAGGGAAGAAGCAAACC
TAATGCATTGATCCATATATAGTGGAAATACTTCATTCTAGGTTCCTATAGATTATTATTATTGAAGGTGCT
CATTTGGGATAGGTTATACTTTTAGATTTAGATTTTTATGAAATAATCAATATATAATTGAAATGATTTCAT
GGTTCTTGTGGAAAGGGTATTAAAGCTTATGGGATAAATCTATACATTTGGATTTTTACCAGTGTTCATT
TAGAAGACTTGATCTCTGAACAGCAGTATAGATATATTAAAGTTAGATGAAAATTTAGAGAGTTGAATTTAA
AGGTGACTATGAATTGATTAGATAATTTTACTTGTTTATGCTGTTAGAATATATTAACATTAAATAGCCATG
CAGTAATAACTTTTTAGGTTTGCCAAGAAATGATGTAGCTGTGTGCCTTTGCAAGTCTGGCAAGGTATTTCG
TCATCCTAGCAACCTAACAAATGCTAAAATGACTACAGTACAGTGTTGCCAAACCAGAGAAATTTTATGTAT
TTGCTCAGGAGTAGTGCATTCTAATAATTCATGTTTTAACTATATAGTTCATTTTAAATCCAATATCTTCCC
CTTGCTGATATGATTTTGATGTATCAACATTTTGAAGACTGAATATCATTGTTATAATTTCCTTTTAACTTT
CCATTTATCTTGGTCAGGAATATAGAATGCCTGTTATGAGTTATTTTGAACAGCTTGAAAAGGCATGAACAG
CAGGAATATCAAATCCAGTGTTGGATAATTTTTTTTTTTTTTTTATTTGTCAGGAAGGAGAGGAAGATGAT
GTAGAGGAGGTTGTGTATGCAACACCCAAGGCACTACCAAATGCATATCTGACTGAAAAC⬚CAGCC⬚**CGGGCCG**
TCGACCACGCGT

Figure 3.32 Nucleotide sequences of the genome walking fragment of *PmCNX* amplified by 2nd 5´cont GW-CNX & AP2. Primer sequences are boldfaced and underlined. The adaptor sequence is highlighted.
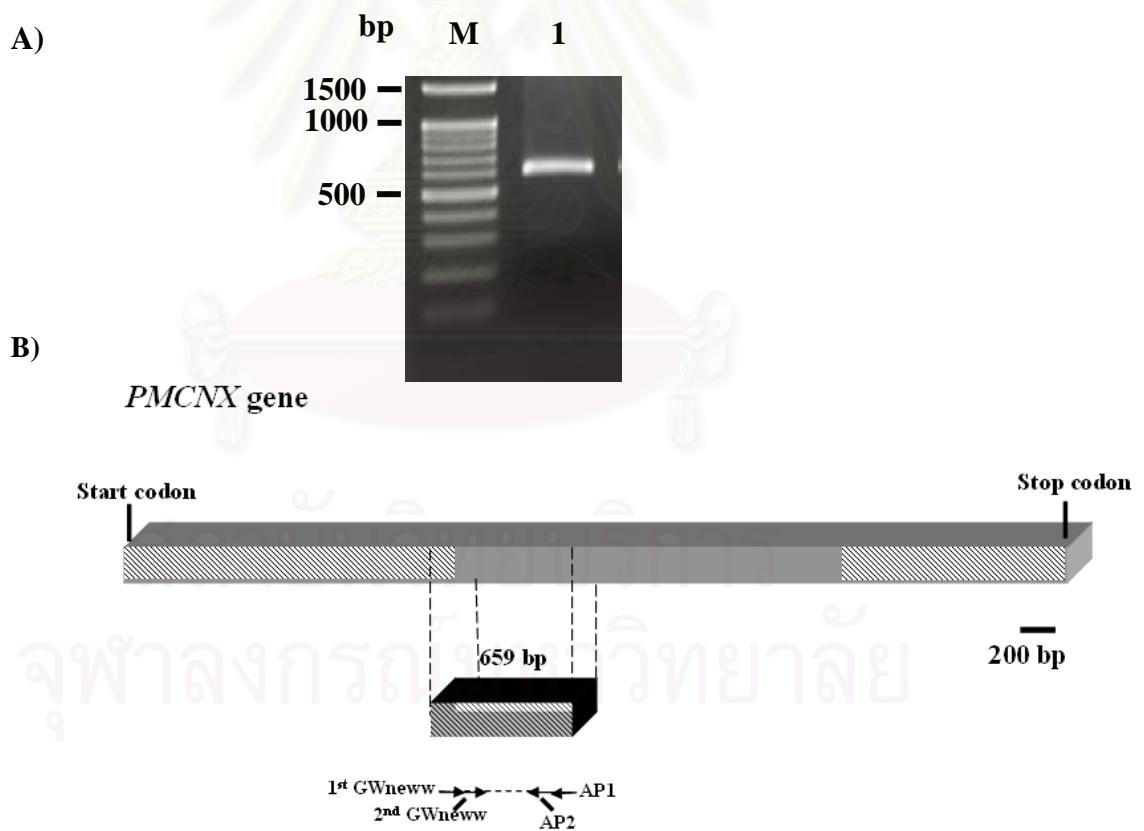
Figure 3.33 Genomic organization of *PMCNX* was further carried out using 2nd GW CNX-new and AP2 (lane 1, A). A 100 bp DNA ladder (Lane M) was used as the marker. Genomic orientation of the amplified fragment (B).

CAAGGCACTACCAAATGCATATCTGACTGAAACGTTTGATGACATAGCTACTTTTGAGAAGACGTGGATCAA
ATCTGAAGCCAAGAAGGACGGTGTTGATGAAAACATTGCTAAATATGATGGTGTTTGGGCAGTAGAACCTGC
TGAACGATTGGCTCTTACTGGGGATCGTGGGTTGGTGTTGAAGTCAAAGGCAAAACATGCAGCAATTGCAGC
ACCACTGAAGAAACCATTTGTATTCAGTAACAAGCCTTTTGTGGTTCAGTATGAAGTTAATTTACAGAATGG
TCAAGAATGTGGTGGAGCATATATCAAACTAATCAGTGCCCAAAAAGGAGCGGGTGGTTCTTATTTGGGACA
TTTGGATCCACGTGGAAAACAGCCACGTGGTCCACTGGTAATAGAAATAGAGTTATCCACATCCTGTAACAG
AAATAGAATTATCCACATCTTATACACCCATCCACCCACCCATCCACTCACCTACCCATCCATCCATCCACT
CAGCCACTCATCCACCCACCCACCTACCCACCCACCCATCCACCGGGCCGTCGACCACGCGTGCCCTATAGT
AATCGAATTCCCGCGGCCGCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTC~~ACCAGCC~~**CGGGCCGT
CGACCACGCGT**

Figure 3.34 Nucleotide sequences of the genome walking fragment of *PmCNX* amplified by 2nd 5′GW-CNX-new & AP2. Primer sequences are boldfaced and underlined. The adaptor sequence is highlighted.

Genomic organization of *PMCNX* gene is not completely characterized at present. In total, 5 complete exons and 3 complete introns were identified and characterized. The complete exons were located at positions 1-124 bp (1st complete exon), 125-504 bp (2nd complete exon), 1385-1510 bp (presuming 3rd complete exon), 1511-1737 (presuming 4th complete exon) and 1738-1788 bp (presuming 5th complete exon) of *PMCNX* mRNA. Genomic organization of the coding sequence between 505-1384 bp is still unknown and further studies are required.

The 1st complete intron was 1245 bp and located between the 1st and 2nd exons. While the remaining introns inserted after the presuming 3rd and 4th complete exons. They were 189 and 426 bp in length, respectively. Moreover other incomplete introns were also found after the 2nd complete exon (296 bp) and before the presuming 3rd complete exon (700 bp) (Figure 3.35).

The GC content of complete exons (41.6-56.9%) were greater than that of complete introns (27.5-31.6%), reflecting a grater thermal stability of the coding regions than that of the noncoding regions (Table 3.2). All of the complete exon-introns boundaries followed the GT/AG rule and interrupt the ORFs between two codons (type 0 introns).

```
genomicPMCNX   ATGAAGTCGAGGTGGCAGAGAAAAGCAGTATTAGCACTGCTAGTTCTTGGCCTGCTGTTA 60
mRNAPMCNX      ATGAAGTCGAGGTGGCAGAGAAAAGCAGTATTAGCACTGCTAGTTCTTGGCCTGCTGTTA 60
               ************************************************************
genomicPMCNX   CCTTTTGGTATTAAAGCAGATGACGATGACGATGAAGAAGCTGTTGTCACGGAAGAGCAA 120
mRNAPMCNX      CCTTTTGGTATTAAAGCAGATGACGATGACGATGAAGAAGCTGTTGTCACGGAAGAGCAA 120
               ************************************************************
genomicPMCNX   ACAGgtaatattgatatcttgaacaattttggctttaaagaacaatatgcagaattataa 180
mRNAPMCNX      ACAG-------------------------------------------------------- 124
               ****
genomicPMCNX   aaaatggatttctaaatctattgatattacataatgtatggtgatttatgggaagagatt 240
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   atgatttcagatagtgaagaacctacttctgtgcttttattggtcagcagggtcttcaaa 300
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   aatcaataaccatatgtagtggattctgggccttgggggaacctttaacttggaatatat 360
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   gacctttttttttcaataatgataagaattccaccagcagaatatacataaggatttaacc 420
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   ttctgtatgaactttagaaattgtgtactaggacaaaatcagaacagaacattttttgtaa 480
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   tgagctgtacaaaagaatggtcgagtacaaagtcaggatgagacctggacacagcttgag 540
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   ggcttcagggggcttcttcctgctaagctgtgcggaagtaagggaagaagcaaacctaatg 600
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   cattgatccatatatagtggaaatacttcattctaggttcctatagattattattattga 660
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   aggtgctcatttgggataggttatacttttagatttagatttttatgaaataatcaatat 720
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   ataattgaaatgatttcatggttcttgtggaaatgggtattaaagcttatgggataaat 780
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   ctatacatttggatttttaccagtgttcatttagaagacttgatctctgaacagcagtat 840
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   agatatattaaagttagatgaaaatttagagagttgaatttaaaggtgactatgaattga 900
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   ttagataattttacttgtttatgctgttagaatatattaacattaaatagccatgcagta 960
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   ataacttttaggtttgccaagaaatgatgtagctgtgtgcctttgcaagtctggcaagg 1020
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   tatttcgtcatcctagcaacctaacaaatgctaaaatgactacagtacagtgttgccaaa 1080
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   ccagagaaattttatgtatttgctcaggagtagtgcattctaataattcatgtttttaact 1140
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   atatagttcattttaaatccaatatcttccccttgctgatatgattttgatgtatcaaca 1200
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   ttttgaagactgaatatcattgttataatttcctttttaactttccatttatcttggtcag 1260
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   gaatatagaatgcctgttatgagttattttgaacagcttgaaaaggcatgaacagcagga 1320
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   atatcaaatccagtgttggataatttttttttttttttttttatttgtcagGAAGGAGAGGA 1380
mRNAPMCNX      ---------------------------------------------GAAGGAGAGGA 135
                                                            ***********
genomicPMCNX   AGATGATGTAGAGGAGGTTGTGTATGCAACACCCAAGGCACTACCAAATGCATATCTGAC 1440
mRNAPMCNX      AGATGATGTAGAGGAGGTTGTGTATGCAACACCCAAGGCACTACCAAATGCATATCTGAC 195
               ************************************************************
genomicPMCNX   TGAAACGTTTGATGACATAGCTACTTTTGAGAAGACGTGGATCAAATCTGAAGCCAAGAA 1500
mRNAPMCNX      TGAAACGTTTGATGACATAGCTACTTTTGAGAAGACGTGGATCAAATCTGAAGCCAAGAA 255
               ************************************************************
genomicPMCNX   GGACGGTGTTGATGAAAACATTGCTAAATATGATGGTGTTTGGGCAGTAGAACCTGCTGA 1560
mRNAPMCNX      GGACGGTGTTGATGAAAACATTGCTAAATATGATGGTGTTTGGGCAGTAGAACCTGCTGA 315
               ************************************************************
```

```
                     ************************************************************
genomicPMCNX  ACGATTGGCTCTTACTGGGGATCGTGGGTTGGTGTTGAAGTCAAAGGCAAAACATGCAGC 1620
mRNAPMCNX     ACGATTGGCTCTTACTGGGGATCGTGGGTTGGTGTTGAAGTCAAAGGCAAAACATGCAGC 375
                     ************************************************************
genomicPMCNX  AATTGCAGCACCACTGAAGAAACCATTTGTATTCAGTAACAAGCCTTTTGTGGTTCAGTA 1680
mRNAPMCNX     AATTGCAGCACCACTGAAGAAACCATTTGTATTCAGTAACAAGCCTTTTGTGGTTCAGTA 435
                     ************************************************************
genomicPMCNX  TGAAGTTAATTTACAGAATGGTCAAGAATGTGGTGGAGCATATATCAAACTAATCAGTGC 1740
mRNAPMCNX     TGAAGTTAATTTACAGAATGGTCAAGAATGTGGTGGAGCATATATCAAACTAATCAGTGC 495
                     ************************************************************
genomicPMCNX  CCAAAAAGGAgcgggtggttcttatttgggacatttggatccacgtggaaaacagccacg 1800
mRNAPMCNX     CCAAAAAGGA------------------------------------------------- 504
                     **********
genomicPMCNX  tggtccactggtaatagaaatagagttatccacatcctgtaacagaaatagaattatcca 1860
mRNAPMCNX     ------------------------------------------------------------

genomicPMCNX  catcttatacacccatccacccacccatccactcacctacccatccatccatccactcag 1920
mRNAPMCNX     ------------------------------------------------------------

genomicPMCNX  ccactcatccacccacccacctacccacccacccatccaccgggccgtcgaccacgcgtg 1980
mRNAPMCNX     ------------------------------------------------------------

genomicPMCNX  ccctatagtaatcgaattcccgcggccgccatggcggccgggagcatgcgacgtcgggcc 2040
mRNAPMCNX     ------------------------------------------------------------

genomicPMCNX  caattc                                                       2046
mRNAPMCNX     ------
```

Remain unstudied region

```
mRNAPMCNX     CGTGTGGATCTTAAAAATTTCCATGACAAAACACCGTACACTATTATGTTTGGA 2100
mRNAPMCNX     CCAGACAAATGTGGCAATGACTTCAAGTTGCATTTCATCTTCAGGCATGTTAATCCTCTT 2160
mRNAPMCNX     ACTGAAGAAATTGAAGAAAAACATGCTAAGAGACCACGTGACAAGATTGAGGAACCATTT 2220
mRNAPMCNX     AAGGACAAGAAGTCTCATTTGTACACATTAGTAATTCGACCAGACAACACCTTTGAAATA 2280
mRNAPMCNX     AGCCTGGATCACGAGGTAATCAATTCAGGAAGCCTTCTGGAGGACTTCACCCCATCTGTC 2340
mRNAPMCNX     AACCCTCCCAAAGAAATCGATGATCCTGAAGACTTTATGCCAGAAGACTGGGATGAAAGA 2400
mRNAPMCNX     GAAAAGATTCCAGATCCAGAAGCCACAAAGCCTGATGACTGGGATGAGGATGCTCCCATG 2460
mRNAPMCNX     CAGATCCCTGATCCAGTGGCTGAGAAACCTAGTGGATGGCTGGATGATGAGCCAGAAATG 2520
mRNAPMCNX     GTGCCAGATCCCACTGCTGAGAAACCTGATGACTGGGATGATGAAATGGATGGTGAGTGG 2580
mRNAPMCNX     GAAGCTCCACTGATCACCAACCCCAAGTGTGTTGATGCACCAGGCTGTGGAGAGTGGAAG 2640
mRNAPMCNX     CCTCCCATGGTGGACAATCCTGAGTTCAAGGGCAAATGGCGTCCACCCATGATTGACAAT 2700
mRNAPMCNX     CCTAATTACCGTGGAAAATGGAAGCCACGAAAGATCCCCAACCCTGACTTCTTTGAAGAC 2760
mRNAPMCNX     CTGGAACCTTTCAAGATGACTGCTATTGATGCTGTTGGTCTGGAATTGTGGTCAATGTCA 2820
mRNAPMCNX     GACAATATCCTCTTTGACAACATACTTGTCACAGACAATGTTGCTGAGGCTTACCAGTTT 2880
mRNAPMCNX     GCTCAAGAAACTTTTGACTTGAAGGTCATGAAGATAGAGAAGGGAC 2926
```

Remain unstudied region

```
genomicPMCNX                                               atttttaaaaatga 2940
mRNAPMCNX                                                   --------------

genomicPMCNX  atgaaaagggtaaacaggtgagacaggtagtacgtgtagagccatctatgtataagaaaa 3000
mRNAPMCNX     ------------------------------------------------------------

genomicPMCNX  aatttataaaataaactcactgggggcatggcgtgtaagtacatgctgtgcctgttggct 3060
mRNAPMCNX     ------------------------------------------------------------

genomicPMCNX  ttgggttaataaaacaagatgtattacactaatttcattacttttttttgtagcaaacttt 3120
```

```
mRNAPMCNX      ------------------------------------------------------------
genomicPMCNX   ttggggaaaatcttgaagctaaccaattacaagcctgtaatgtgggcactctacgtact 3180
mRNAPMCNX      ------------------------------------------------------------


genomicPMCNX   ttatatgattattccaactggattaataatcttttgtctactgaagagagtatatgaggt 3240
mRNAPMCNX      ------------------------------------------------------------

genomicPMCNX   attttttctggtaaaattttttgctagataagcaacaggattacataagggctagtgcaag 3300
mRNAPMCNX      ------------------------------------------------------------

genomicPMCNX   aaactatgttagctagggtgagcatatgaattgctgtgctagttcaggatgtgacatttg 3360
mRNAPMCNX      ------------------------------------------------------------

genomicPMCNX   tgtatgagtaattatatgaaaagtaggggtaggttattagtaagagttggttgttgctta 3420
mRNAPMCNX      ------------------------------------------------------------

genomicPMCNX   tagtgttgctttatgtaaaatatatttttttttttatttttaaaaatctttcatactttaa 3480
mRNAPMCNX      ------------------------------------------------------------

genomicPMCNX   catatgaaccaaagcctttttgatacacaccttctagatacagattatttggcattcttt 3540
mRNAPMCNX      ------------------------------------------------------------

genomicPMCNX   tcagtgctttaatacaagttccctgtgtgattgccctgctttagagtaatagccacattc 3600
mRNAPMCNX      ------------------------------------------------------------

genomicPMCNX   atctgtatctgtgtattctttattgtAGACTGGAGTTATTCGACGAATCATCAACTACTC 3660
mRNAPMCNX      ------------------------AGACTGGAGTTATTCGACGAATCATCAACTACTC 1428
                                      ********************************
genomicPMCNX   CAATAAGAATCCATGGCTATATGGTATCTACGTACTCCTAGTGGCCATTCCTGTAGTGTT 3720
mRNAPMCNX      CAATAAGAATCCATGGCTATATGGTATCTACGTACTCCTAGTGGCCATTCCTGTAGTGTT 1488
               ************************************************************
genomicPMCNX   GATATTTGCCTGTTGTTGTGCAGAAGCAAAGgtagagtatgataatagctgtcaaatgta 3780
mRNAPMCNX      GATATTTGCCTGTTGTTGTGCAGAAGCAAAG----------------------------- 1510
               *******************************
genomicPMCNX   atgatttctttgctttaagtgtaaaaactaaaagatttaacattaataatagtgtagac 3840
mRNAPMCNX      ------------------------------------------------------------

genomicPMCNX   aagtccttactgtcacacgtgtttgaaaatgaatgtgcaatattactagaacttcaaaaa 3900
mRNAPMCNX      ------------------------------------------------------------

genomicPMCNX   tatatattttgtgtcagttaattattttggatctctcgagACACCAAGGAAGATAAGGAT 3960
mRNAPMCNX      ---------------------------------------ACACCAAGGAAGATAAGGAT 1530
                                                       *******************
genomicPMCNX   GCAGAGAGGAAAAAGACCGATGCACCTTCTCCTGATGATCCACAATCAGAGCAAAACAAA 4020
mRNAPMCNX      GCAGAGAGGAAAAAGACCGATGCACCTTCTCCTGATGATCCACAATCAGAGCAAAACAAA 1590
               ************************************************************
genomicPMCNX   GAAGATTCTAGTCAAACTGCAGCAGATGATGATGCTCCTGGAAGTGGGGATGAAGCTGAA 4080
mRNAPMCNX      GAAGATTCTAGTCAAACTGCAGCAGATGATGATGCTCCTGGAAGTGGGGATGAAGCTGAA 1650
               ************************************************************
genomicPMCNX   GGGGATGAAGACAAAGGTGATGAAGAGGGGAGAGGAGGAAGAAGATGAGGAAGAGGAGGAG 4140
mRNAPMCNX      GGGGATGAAGACAAAGGTGATGAAGAGGGGAGAGGAGGAAGAAGATGAGGAAGAGGAGGAG 1710
               ************************************************************
genomicPMCNX   GAAGCAGAGAAAGCTGATGCAGCTGAGgtattgatccaaaaccagttatcttaactattt 4200
mRNAPMCNX      GAAGCAGAGAAAGCTGATGCAGCTGAG--------------------------------- 1737
               ***************************
genomicPMCNX   tcttgtgttgccctactttatataaatgtttgaatctgcatttctccatctaattgtctt 4260
mRNAPMCNX      ------------------------------------------------------------

genomicPMCNX   taaaagtttgaatgttgttctcatttctctcaatgtgcaaaccactttatcagatttttta 4320
mRNAPMCNX      ------------------------------------------------------------

genomicPMCNX   aaacactgttttgtgtttcatttctttgctgatttcccagttgtacatttttttgagtacc 4380
mRNAPMCNX      ------------------------------------------------------------

genomicPMCNX   acatattgttcatatttgtctagggttgcttattttcagatattgtaataacataactgg 4440
mRNAPMCNX      ------------------------------------------------------------

genomicPMCNX   tgctgcatatccaagtaaagagtacttttttggttttgccaagtattttttttttttttttc 4500
mRNAPMCNX      ------------------------------------------------------------

genomicPMCNX   aaaaacatgatccataggtgtataccaatgtttcagtagtgtatcacatacaagaaaatt 4560
mRNAPMCNX      ------------------------------------------------------------

genomicPMCNX   tgtactttctaatattaattttttctttcaagGAGGTCCAGACTCGCACATCACCAAGG 4620
mRNAPMCNX      --------------------------------GAGGTCCAGACTCGCACATCACCAAGG 1762
```

```
                                        *************************
genomicPMCNX      CTGCGCAAGTCCCGGAGAGATTAA 4644
mRNAPMCNX         CTGCGCAAGTCCCGGAGAGATTAA 1788
                  ***********************
```

Figure 3.35 Organization of the *PMCNX* gene. Nucleotide sequences of complete exons are capitalized and those of complete introns are shown in lower letters. Highlighted lower letters are incomplete introns while highlighted capitalized letters indicated un-studied regions. Start and stop codons are boldfaced and underlined. Results from genome walking analysis were successfully identified 5 complete exons (**1st exon:** positions 1-124 in genomic and mRNA sequences; **2nd exon:** positions 1371-1750 and 125-504; **presuming 3rd exon**: positions 3627-3752 and 1385-1510; **presuming 4th exon:** positions 3941-4167 and 1511-1737 and **presuming 5th exon:** positions 4594-4644 and 1738-1788) and 3 complete introns (1st intron inserted between 1st and 2nd exons, presuming 2nd intron was found between 3rd and 4th exons and the last complete intron intervenes the last exon).

Table 3.2 GC content and length of exons and introns in the *PMCNX* gene

Compositions of gene	Genomic DNA		mRNA	GC content (%)	
Exon	(No. of nucleotides)		(No. of nucleotides)		
1	1-124	(124 bp)	1-124	45.2	%
2	1370-1750	(380 bp)	125-504	42.0	%
3	-	-	-	-	-
4	3627-3751	(126 bp)	1385-1510	41.6	%
5	3941-4029	(227 bp)	1511-1737	47.6	%
6	4594-4644	(51 bp)	1738-1788	56.9	%

Intron	Genomic DNA		Type	GC content (%)		GT/AG rule
	(No. of nucleotides)					
1	125-1369	(1245 bp)	0	31.6	%	Yes
2	3752-3940	(189 bp)	0	27.5	%	Yes
3	4030-4593	(426 bp)	0	28.9	%	Yes
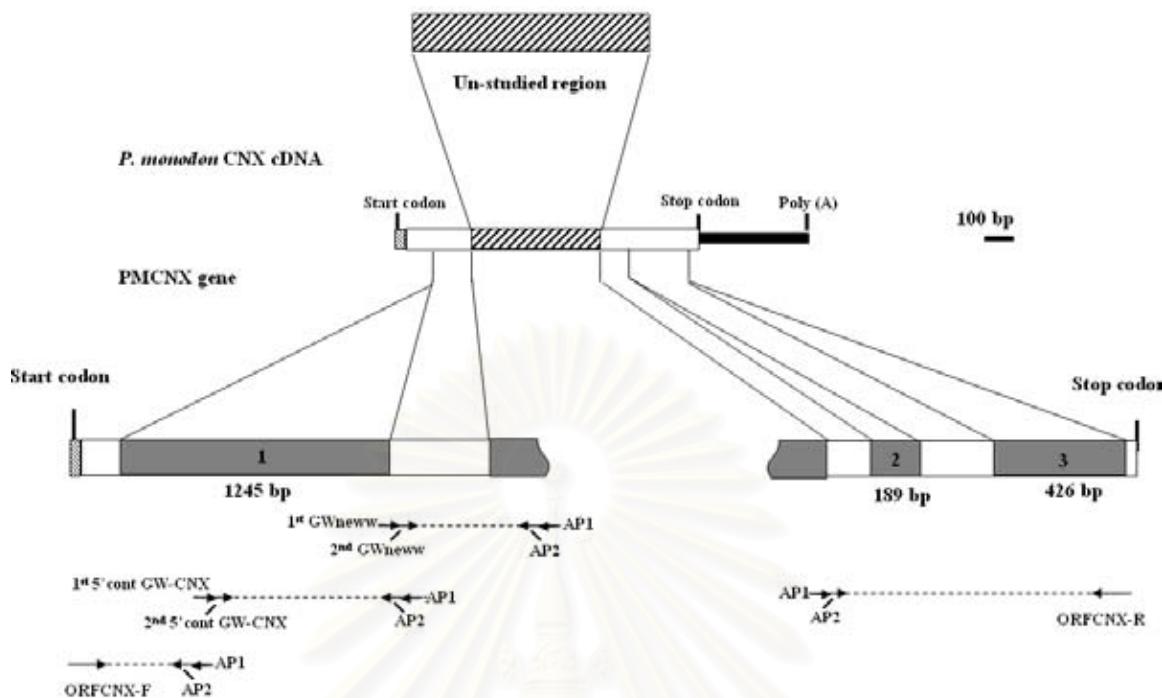
Figure 3.36 Schematic diagrams of *PmCNX* cDNA and gene. Coding regions are illustrated in open bars. Complete introns (with numbers) are gray-shaded and shown in boxes while incomplete intron (s) was illustrated in gray-shaded broken boxes. Unstudied region (s) is marked by a diagonal line box. Primers used for amplification of genomic *PMCNX* are illustrated.

3.5.3 Genomic organization of *ERp57*

The genomic organization of *PMERp57* was also characterized by genome walking analysis. Initially, ORFERp57-F and ORFERp57-R primers were used in combination with the adaptor primer (AP1. A 705 bp fragment was successfully amplified by ORFERp57-F & AP1 whereas a 863 bp product was successfully amplified by ORFCNX-R and AP1 (Figures 3.37-3.39).

Afterwards, primers 1st-5´cont and 1st-3´ cont GW-ERp57 were used to perform the primary genome walking PCR in combination with the AP1 primer. The primary PCR products were diluted 50 fold and used as template in the secondary PCR. The

amplification product of 2350 bp was generated from of 2nd 5´-cont GW-ERp57 and AP2 primers (Figure 3.40) whereas a 2604 bp fragment was generated from 2nd 3´ cont GW-ERp57 and AP2 primers (Figures 3.41 and 3.42).

The primary and secondary PCR were further carried out using additional primers (called 5´- and 3´-next ERp57-GW). Genome walking PCR using 5´-next ERp57-GW and AP1 generated a 1286 bp PCR product while 3´-next ERp57-GW and AP2 generated a 1505 bp product (Figure 3.43). These fragments were cloned and their nucleotide sequences are illustrated by Figures 3.44 and 3.45.

Finally, a fragment of 1543 bp in length was successfully amplified using primers last-ERp57GW R 1st & AP1 and last-ERp57GW R 2nd and AP2 for the primary and secondary PCR, respectively (Figure 3.46) Nucleotide sequence of this fragment is shown by Figure 3.47.
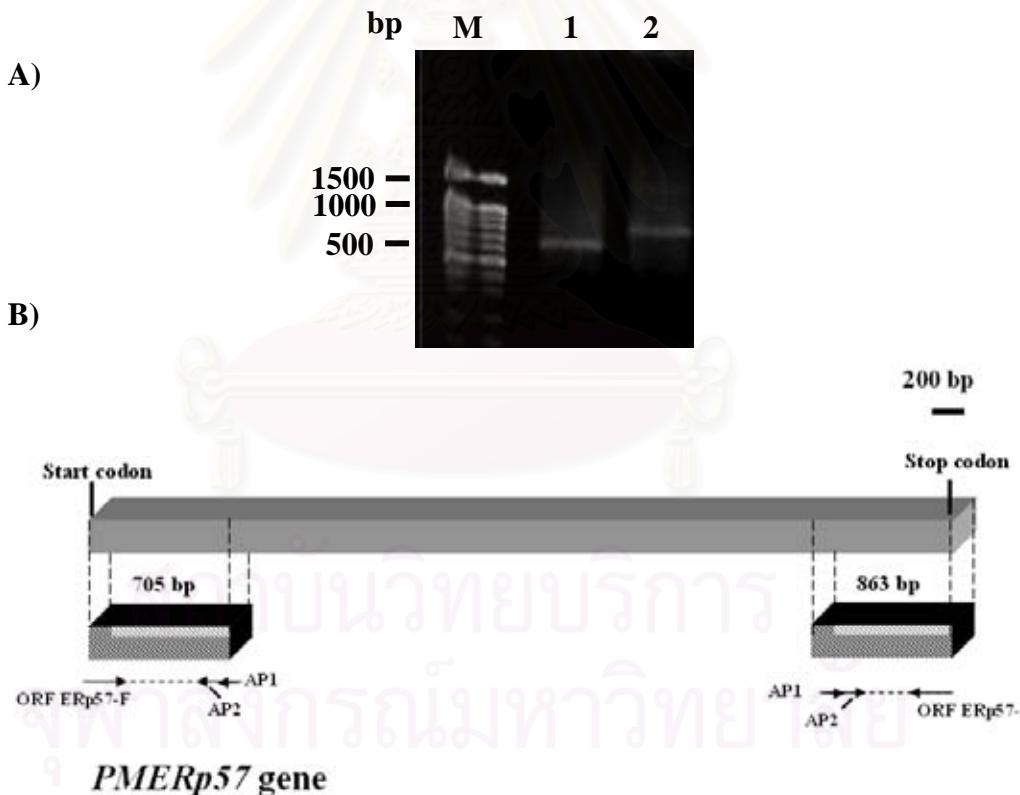
Figure 3.37 Genome walking analysis of *PMERp57* using ORFERp57-F & AP1 (lane 1, A) and ORFERp57-R & AP1 (lane 2, A). A 100 bp DNA ladder (lane M1) and λ *Hind III* (lane M2) were used as the markers. Genomic orientation of each amplified fragment (B).

ATGGCTACGAGATTGTTAATACTACTCCTCTCCCTCGTGGCCGTGGCGCTGGGAGACGATGTCCTGCAATTA
AACGACGCGGATTTCGACGGGAAAGTGGCCAGCTACCACACGGTCCTCGTCATGTTCTACGCCCCGTGGTGA
GCCTTGGGGTCCTGTGGGTCCTTTGGGGTCCTGTGGGGTCCTCTCGCCTAGTTGGTCCTGTGGGTCCTCTCG
TGCCTTGGGGTCACCTTCCCCTTGAGCGAGGGGAGCGCTCGCGGGTGTTTGCGCTGCCCCCTTCCTACGGTC
TCGTATTAGGGGGCCAAACAGCGGGCCTTCATAGCAGGTTTTGAAGCGAATGGCTTGGATGTAGATTAATCT
GTAATATATTGGGTTTGGTTCAAGATAGGGCCATGGATAGACGCCAGAGGCGACGTGTATGCACGCACGCAT
GCTGCTGCTGCATGCGTGCGAGGGCGTCCGCCTTTGGCAGGCCGTCCCGTGGCGGCGTGGCCGGGCTTGGCA
GGAGGCAAAGGGCCGAGCTGGGGACTTGGCCGTTGGGCCTAGGCGCGGCTCAGGCAGCTGTTGGGTTGTGGG
CTTTCGTTTGGTTGGTAATCTCGTAATTAATATTACCAGACTGTCAAAGCTGTAATGTGTGAATACCAGCCC
GGGCCGTCGCGACCAGCCCGGGCCGTCGACCACGCGT**GCCCTATAGTGAGTCGTATTAC**

Figure 3.38 Nucleotide sequences of the genome walking fragment of *PmERp57*
amplified by ORFERp57-F & AP1 primers. Primers are boldfaced and underlined. The
adaptor sequence is highlighted. The start condon is double-underlined.

GTAATACGACTCACTATAGGGCACGCGTGGTCGACGTAAGTTCTTGTGTAGCCTGTTTTGTTTTGTATTTCT
AATATTTGCCTGTATTATATTTAACAACTTACTCATCTCTAACTTCACTTCCATAGAATGGATAACCTCCAA
GCATTCCTCACCAATCTCAAGGCGGGTGAGCTTGAGCCATATCTGAAGTCTGAGGCAGTGCCAACACAAGAT
GGCCCTGTCACTGTTGCTGTGGGTAAGAACTTCAATGAAGTTGTCTCTGATGAGCGTGATGCCCTCATTGAA
TTCTATGCTCCTTGGTGTGGTCACTGCAAGAAATTAGCGCCCACCTATGATGAGCTGGGAGAAGCGGTAAGG
TTTTTCTAATTATTCTTGTGTATTCATTTGAGAATGAATACTGGATAATTACTGTACAATGCATTCTTTCTG
GTAAACACTAATTTCCTTACAGATGAAGGATGAAGATGTAGACATTGTGAAGATGGATGCCACTGCCAATGA
TGTTCCTCCTCAGTACAATGTTCAAGGCTTCCCCACCATCTTCTGGAAACCCAAGGGTGGTGTTCCAAGGAA
TTACAATGTAAGTAGAGCCTATTTTATACCAATTTTTGACATTGAATCTAATATAGGAGTTCATAGACACTG
AGCTTCAATCAAATGCTTGTTCAAAGGCACCCAAGCTCTTTGATATGCAGAATTGAAAAATATTTGATTTGG
TGAACTAAACTTGATCTTAACTTCCCCAGGGTGGCCGGGAACTGGACGATTTTGTCAAGTACATTGCCCAAC
ATTCCACAAATGAACTGAATGGGTATGACCGCAAGGGGAAGGCAAAGAAAGGCAAGAAGACTGAACTT**TGA**

Figure 3.39 Nucleotide sequences of the genome walking fragment of *PmERp57*
amplified by ORFERp57-R & AP1 primer. The AP1 primer is boldfaced and underlined.
The adaptor sequence is highlighted.

A)

B)
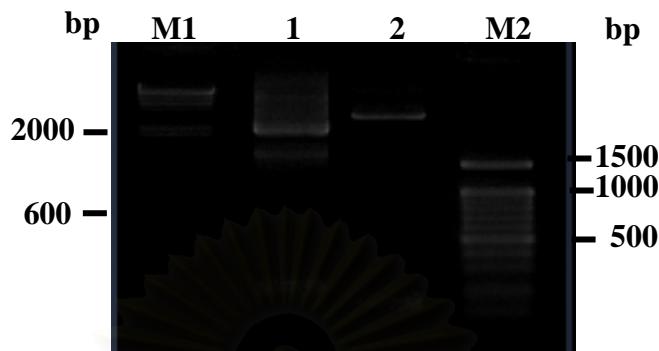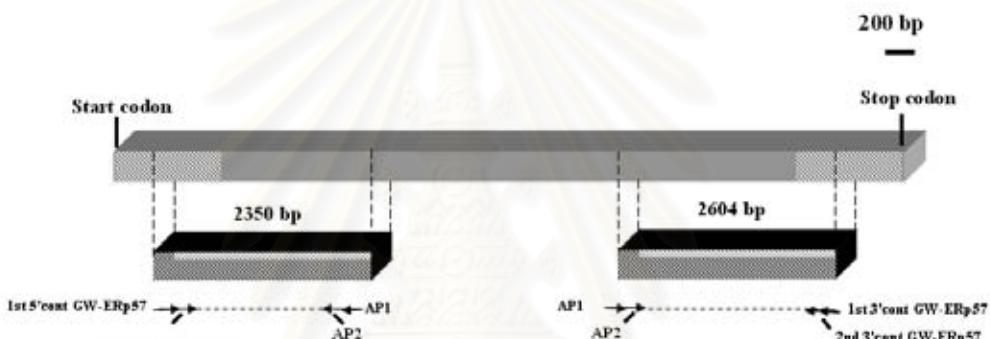
Figure 3.40 Genomic organization of *PMERp57* was further carried out using 2nd-5´ and - 3´ cont GW-ERp57 and the Adaptor primer 2 (AP2) (lanes 1 and 2, respectively) primers. A 100 bp DNA ladder (Lane M1, A) and *λ-Hin*d III (lane M2, A) were used as the markers. Genomic orientation of each amplified fragment (B).

ACCAGACTGTCAAAGCTGTAATGTGTGAATATTTTCCGTCGCAAGTGATATAAATGAAGAGGGTATTTGTTA
GTGATGCACCGGCCCCACCAGACATTCTAAAAACCAGGGTATTGTATGAACCCAAAAATCCAAACTTAACCA
TCTCTACCTTTTTTTTTTTTTATTATTACATCCCCCCCTTCCTTTATCATCACTGTGGCAACTGTCACAATTT
TAAGTGTGGAAGGTTTGACGTGCGATAGTGAGGAATTGCAGAAAAGTCCTTGTTCTTTAAGAAGGGGTGGGG
TACATTTCTCTTGCTCTTGATAATAATTACCTATTCCACTAAATTTGCCTGTAAAATGTATACCAAAGTATA
GTGTTGCACGACTTTAATATGAACTTGCAAAGGTAATACCATAGTTTTTTAATGTATTTGTCCATTGCAGCA
CTGTTTTCCCACAGCCATGTAGTACAAGATAGACTAATGTTACCAGTGTTATATCTGAGTCCTTTATTTTTC
ATTGCAAGTTTTATTTTAAGAATTGCTTATGGAGACTGCACCCCTGAAACTGTCCCTATTACTCCAGCCCCT
CTTGGAGATACATTGATGTATACAAGAAAAAAACTGACCGCCGTGCACTTGTTTCCCTATCGGGGGCTCTG
ACCCCAACCCTCCCCTTTTTGCCATGGAACTAATGTTTTGTACAACATTAATGGACCTATGCTTTGATATTT
ACCTTCCAACACATCTAGGTAAAAAAACTTGCCAAAATGACCCCCCAGTTCCCATACACAGTTTGAATAAGT
GCGTGAACTGGGGGGTGGCAATACACCACCTTTTCTTGGGTGAATGCCCAAAGGGAATGCGTAGTCATGGCA
GTTTCGATTGTTGGGTTAACGTTAGAGTTGGGCACAGATAGAGGTAAATGGACCCTTACTATCTTATAAAGC

```
ATAAAAATATTAATTCCCCATCTAAAAAATAAAAGGGAGGAAAGTTAGGTCTGGATTTGGGTTGATACCAGC
CCCGGGAATCCACTAAGTGAATTCGCGGGCCCCTGCAGGTCCACCATTATGGGGGAAAGCCTCCCACCCGCG
GTGGGAGTGCCAAAGCCTGGGGGAATTATCGATAATTAAAATAGTTTTTCATTTCAGGCTATGACACTTGTG
TGCCTAACAACTTTTATTATTTTTGAGTTGCGATGGAAATGCACAAGTTTCTTGGGTTCATTTGCTATAGTT
TGCATGGTAAAATGAATTGAAAGAGGAAAGCAGTCTTAAGGTGGTTTGTTACAGTTTTAAATACATTAATTA
AGAGGAGCCACTGTTGATTAATGTTTTCAAGTGCATAGTGAACCACAACTGCTTGAGACTACAACATGTGCA
GTCATTCTTGTTTATCCTAGCAAGCTAAACACAGTGGTGACTTATTTTATTTAGGCTTTTTAGCCTTGTTAC
ATCCTTCTGCAGCTACAAAGGTCATGATTTTTTCCTGCACAATTTCCTTTTATGGGCTTGATCAATGCACTA
CCAGATGCAGGGCTTGGTATAAGTGAAACATACTAATTTTGATCAATGAAATCTCTTCGTAATGACATATAG
GTCTTGAAAGCCAAGTCATTGCTATGAGTAACCCATTTCCATAAAGATCTTTTAACTTAATGGGTAACCCCT
TCCCCCTCTAAGGTGGGTTTTGGGCTGGTTCTTACTAAAATTTCATGGAAATATTTTAAAGTTTACATTTGA
TTACAATTTCATTTAATAATTAGATTGGGTTTCCACAGCATAGAAAAACTACGGAAAGTGAGTAATAATTCT
TAGTATCCCTGGCATAAGTGCCAAACTAATCTTCTATTTTGGGTAATTGATTATGAAGTTAATGCACCTTAG
AATGTCCCAGACTAAGGGGTTACTTATGTTAATCCAGTCATTAGTTTTTTTGTTTTTTGTTTTTTTTACTTC
CAAATCGGTACTATAGTTGATAGATCAGGAGGTTTTAGAAAGTGGTAAATAAATGCTGTAGTTGGATGCTCT
TAACTATAAACCCATTTACTAGGCTTATAGTGTAGCATAGCATTTCTGCAACCAAGTTTCTTCTATGGTTGA
ACAAATGCTAGTTAGTGGCACAGCACTTACTTTGAGCGCATGGTATTTAATACATAGGAATATAACTGATCC
TGGAAATGACAGGTTTGTCAAACCTCCACCCCAATCCACCCTTCACAAAGAGAAGGATTTTAATAAGTTGC
TGAAGCAAAATGTGCCTGAATGTATGGGATAGTCTCAGGATTATGCTGGAGAAGTTGTCAGGCATTAAGTTG
ACTTTTTTTTCCAGGTGTGGTCACAGCC**CGGGCCGTCGACCACGCGT**
```

Figure 3.41 Nucleotide sequences of the genome walking fragment of *PmERp57* amplified by 2nd-5´ cont GW-ERp57 & AP2 primers. Primer sequences are boldfaced and underlined. The adaptor sequence is highlighted. The start codon is boldfaced and double-underlined

```
**ACGCGTGGTCGACGGCCCGGGCTG**TGGAAATCAGTTGTGATGGGAGAAGTTGGGATTTATTTCCCTCGGTGT
ACAAGGCAGGGTATGCTAGAAATTGTAACTGACCACCTATGTTGGACCACATCCTATTCGTTACAGAATGTA
TCATTGTTCGCTGGTTGAATACAGTCAAAATGGTCTGATCAGGCACTTAGAGTTGAATGATAGTTTGAAGTG
TGCAGTTTCATACACTGGCTACTCTCCATGGAAGTAATTTGAAGTAAATTAGGTATTTCAGCATCCTGTTAT
TAGGGCTTTTGGTGCTGTATATGAAATGGGCTTTGTCATGTTAAGAAAATTCCCTTAATATTAGGGAAAAGA
GAACTAGGTTTGTCCAAAGCATTTGCTAATAGCAGCACATGTAGGGTATGGAGAACTTTGATACTTGTGGTT
TACAACATTCATTTGGGCCCCATTTGCAGAGTTCGGCTTCAAACTCGTGTGTGATAGCTACAGAGCTTGGTT
GATGGCTTTGGTTTGTATTTTTGGACAAGTTACTTGGCATTAAATTATGTCGGGCACACCGGCAAATCAAAA
TGGTTTGGCTGCTGAGTGCACACTTCACACCAATGTTCTTAAAATTTTGTTGCATCAAGCCATTTCTTTTAT
CAACACATGTTTGGTTACTAACATAATTGTGAAATATGGATATGAATGTAATTTAAACTGGTGGTTAAAAGT
TTCGACTGTATAGCATGGATTCCTCAGACCAAATGTCTGCTTAAAGAATTTTGCACTTTGAATTTATGCAGT
TGATATTAATGGTGGTAAATGCGTGTTGATTAACCCAAATAACTTTTAACACTAATCATGGTTACAAGATTG
AATGTGTGAAGGTTTAGGTTGAAACCTTTCTAGTATATTTCATAATTGGGCAAATTTTGGGTAGTGTAAAAA
ATTTCCACATGTAATTAATGTTTTCCCTTTATCAGGTGGATTGTACTGATGATGGAAAGGACAGCTGTAGCA
GATTTGGTGTGTCTGGCTACCCTACCCTGAAGATCTTCAAGGGAGGAGAGCTCTCTACGGACTACAATGGTC
CACGAGATGCCAGTAAGTGCTTGATGTACTTATGTATCTTGATCACAAATAAGATTTAACACCAAGCCCCCT
TGTCTGTCCTTTTGTGCATATTCTCCCTTTTAAATGTTTCTTCTCCTTCCAGGTGGTATTGTAAAGTACATG
AGGTCACAGGTTGGACCAGCCTCTAAGGAGTTGACATCCGTGGAGGCAGCAGAAGCATTCCTTGGTGCTGCT
GAAGTTGGAGTCGTTTACTTTGGAGGAGATTCCAAACTTAAAGGTTTGTATTCCCATGCTGTTACTTTGAAA
TGGGTATTTGATCAATGCTTTACCAATGAAAAGGGGAAAAGACAATCTAAATGCTATTCATTGCATAAATCT
AGAATTTAAACATTCTTAGTTTTAAACTTCAAAATGCACACTACTGCATATAGCATCCCAAATATTTTTGAA
ACCAGGACAAGTCCAATTACATCTTGCTTCAGACATGTGAGAATGCTATGCCATTAGGTTGGTGGCAGTGAA
CTCTTACACTTTTCATGAGCATCATAGGTTTGCACTTCAGAAGTAACCTTCCAAACAATCCTGTTTCACTTT
GATAACTTTTTTGCAGTGTGCAGTTGTGCTGAAATCTTTTGCCTTGATTTCAGTACAGGTGGTCCATTTTGT
TTTACTATTCCTATAGTATGGGAATATTATTAAAGCAGTATAACCATTTGAAATGGTTTTGTTTCAAGTAAT
TGACCAATGAATCCTGGATTACCTTGAAACATTGAATATATTTCAGATGCTTTCCTAAAGGCTGCTGATAAG
CTGAGGGAATCCATCCGTTTTGCACACTCCCTCGATGCCACTGTTAATGAAAGTATGGGTACAGCGATGTT
GTTGTACTTTTCCGACCGAAACACCTGGAGAACAAATTTGAGCCTTCCTCTGTTGTATTTGAGGGATCGGCA
GACAGGGCTGAGATTGAGTCTTTCATCAAAAGAACTTCCATGGTTTTTCTTGCATGCATACTTGCTCACTC
CATTGCAATAACTTCAAGTACATATTTTGTTTTAGTTTATACTACTACTGTTGCCTTGTTTGATTCTGCAGA
CTATGCAAGGAATCTATTGTAATACGATCTCTCATTGTGGCTTTATATTAACCATCAGAGCTCTTGTCTGAA
```

```
CACTTAAATGCTATGTATATTGAATAGATGCAAGTATTTGACCTGTCGGCTTCAAGATCAATTACTGTTCTT
TACAGCCATGGTTTGGTAGGACACCTAACGCAAGACACTGCTCAGGATTTCAAACCTCCAGTTGTGATTGCT
TACTACAATGTTGATTACATCAAAAATGTTAAGGGTACAAATTACTGGCGCAATCGTGTCCTTAAGGTAAGA
CTTTGTATGTTTGGAAAAAGGTTTACAGAATTCATTTCTTAGCTTTGACATGTGTAGGAGAGTTGACAAGTA
GTTAAAAGGTGTTGTATTTGCAGGTGGCACAAAACTTTGCTGATGACTCAAGTTTGC**CGTTGCCAATAAGGA**
**CGACTTCCAGCA**
```

Figure 3.42 Nucleotide sequences of the genome walking fragment of *PmERp57* amplified by 2nd-3´ cont GW-ERp57 & AP2 primers. Primer sequences are boldfaced and underlined. The adaptor sequence is highlighted.

A)
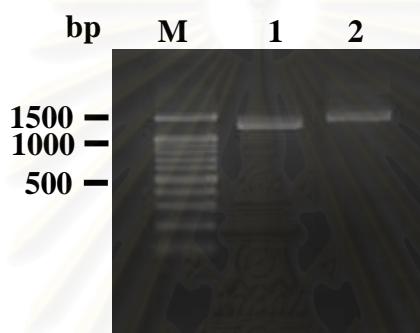
B)
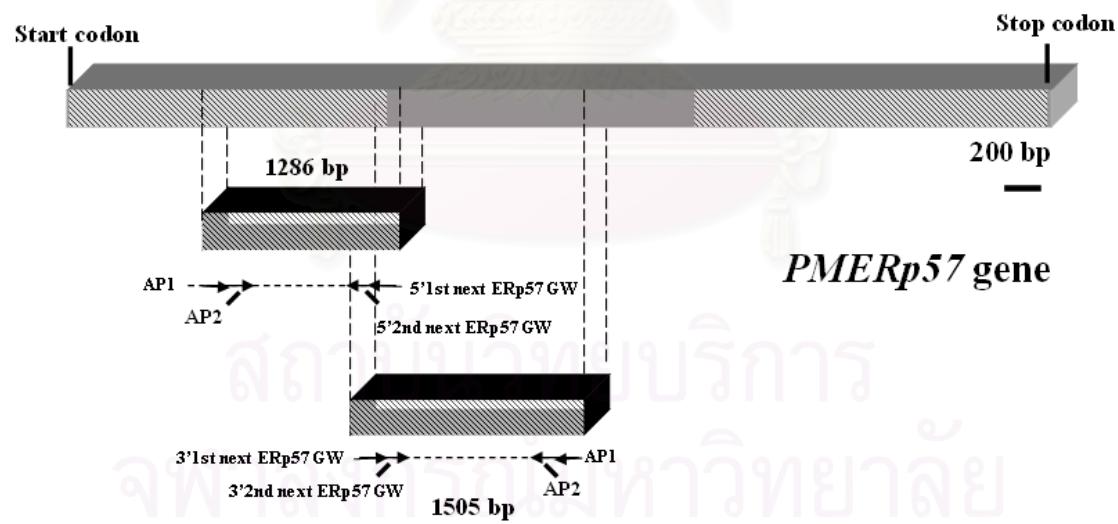
Figure 3.43 Genomic organization of *PMERp57* was further carried out using 2nd-5´ and -3´ next ERp57GW and the Adaptor primer 2 (AP2) (lanes 1 and 2, respectively). A 100 bp DNA ladder was used as the marker (A). Genomic orientation of each amplified fragment (B).

GTAATACGACTCACTATAGGGCACGCGTGGTCGACATGACACTTGTGTGCCTAACAACTTTTATTATTTTTG
AGTTGCGATGGAAATGCACAAGTTTCTTGGGTTCATTTGCTATAGTTTGCATGGTAAAATGAATTGAAAGAG
GAAAGCAGTCTTAAGGTGGTTTGTTACAGTTTTAAATACATTAATTAAGAGGAGCCACTGTTGATTAATGTT
TTCAAGTGCATAGTGAACCACAACTGCTTGAGACTACAACATGTGCAGTCATTCTTGTTTATCCTAGCAAGC
TAAACACAGTGGTGACTTATTTTATTTAGGCTTTTTAGCCTTGTTACATCCTTCTGCAGCTACAAAGGTCAT
GATTTTTTCCTGCACAATTTCCTTTTATGGGCTTGATCAATGCACTACCAGATGCAGGGCTTGGTATAAGTG
AAACATACTAATTTTGATCAATGAAATCTCTTCGTAATGACATATAGGTCTTGAAAGCCAAGTCATTGCTAT
GAGTAACCCATTTCCATAAAGATCTTTTAACTTAATGGGTAACCCCTTCCCCCTCTAAGGTGGGTTTTGGGC
TGGTTCTTACTAAAATTTCATGGAAATATTTTAAAGTTTACATTTGATTACAATTTCATTTAATAATTAGAT
TGGGTTTCCACAGCATAGAAAAACTACGGAAAGTGAGTAATAATTCTTAGTATCCCTGGCATAAGTGCCAAA
CTAATCTTCTATTTTGGGTAATTGATTATGAAGTTAATGCACCTTAGAATGTCCCAGACTAAGGGGTTACTT
ATGTTAATCCAGTCATTAGTTTTTTTGTTTTTTGTTTTTTTTACTTCCAAATCGGTACTATAGTTGATAGAT
CAGGAGGTTTTAGAAAGTGGTAAATAAATGCTGTAGTTGGATGCTCTTAACTATAAACCCATTTACTAGGCT
TATAGTGTAGCATAGCATTTCTGCAACCAAGTTTCTTCTATGGTTGAACAAATGCTAGTTAGTGGCACAGCA
CTTACTTTGAGCGCATGGTATTTAATACATAGGAATATAACTGATCCTGGAAATGACAGGTTTGTCAAACCT
CCACCCCAATCCACCCTTCACAAAGAGAAGGATTTTAATAAGTTGCTGAAGCAAAATGTGCCTGAATGTAT
GGGATAGTCTCAGGATTATGCTGGAGAAGTTGTCAGGCATTAAGTTGACTTTTTTTCCAGGTGTGGTCACTG
CAAGAGATTAAAGCCTGAGTTTGAGAAGGCCTCTAC**CACCTTGAAGGCCAACGACCCTCCCG**

Figure 3.44 Nucleotide sequences of the genome walking fragment of *PmERp57* amplified by 2nd-5′ next ERp57GW & AP2 primers. Primer sequences are boldfaced and underlined. The adaptor sequence is highlighted.

TTGAAGGCCAACGACCCTCCCGTCTACCTTGCTAAGGTAAGCTGGCAAATTTGGGTGCAGGTTACCTTGTGG
ATTTTAGGTTGTAGATCTTGAAGCAAGTTTTACAGAGTAAAATTTAAGTACAGCTTGTGTACAATAGCTGAT
TACAGTTGTTAGCAAAGCCTAGCAATATATTTGTACTACACTATAACTAACATCTGTTGCAAATGTCTTGAT
TAAAGAATTGTAGGGAAGGTATTTTCATTGATTTTGGCCATGTCCAAGTTCATTTTGCATAACATGAACAGT
GGGTGTGGTTATATTAAACTAAGTATGTATCCCATGTCATGTTTTGACTGAATTTTGGTGTCGTTCTAATTT
TTGACATGAGCTTGCAGCTACTATAAAGCAAATTTCCCCCCTTCACGGCTCTTTATCAGCAACGTAAGAATT
TCTATTGGATATGTGTAGTTAATGCAACACCCATGCTCAGCAGTCTGATGCTGAGATTATATTTTAACTTTA
AATACTAACTTGTTCAAAAGGTTAAACCTTCCATTTGGAAACCCTGCTGAAGGGTGTTGTGCTGTTGCAGTG
GTATAGCAAATTTTTGCACAGATAGGTGTTGGAAGTATGAATTTTGATCTTGACCAGTTGAACATGAAATGT
TGGATTAATGCCCAATTTGAGTTTTGTATGTGTACATTTTAGCGTTTTGTAAACTTTCCAAATTGCTACTGA
AAGCGAAGACCAGGGCTCCCTTCCACAATGAGTAAAATAACATGCCCTGCTTTCCTTGAGTGCTCTTCCCTA
CTAGCAGCCCTACTATACACCTGGTTTCATATCTGCATTTGAGATTTTGCCTACATCTCTTCTGTAGACTTC
TGATTCAGTAGATTGTGATTTTTTTATATTCAAAACCTGTATACTAGATAATTTTTAGTTTGACAATTGGTC
CTTCCAGTACAATGTTTACCCAATTTATTCGTGATAAATCGTACCACTTCCGTCGCAATCACAGCCCAGCTA
TCTTTGATATACTTCATGCCCTCCCCAACTATTGGGACTAGTCAGAGTATTGTGTAAAACCATCACAATTTA
TTAATTTTTGGCATTTGGTGGGCTTGAATACTAAACTGAAGTCTATTCTATTTAGCAATATATTGATAAATG
ACCTGGAGTTTGGTTGCACTTTATTCCAACAAATGCTGGCAAGTCTCGCAACACATCCTGGTGCTGTTCATA
CACATTTGCAAGTTATTGTACAGAATAGTGGGAAAACTATTAATGCATTAAAGTAGTAAATAAAAAAAGTGC
ATTAAATGTGCATGCCTTATGAAAATCAGGCATTTCACTTTTTCAAGCTGTGATTTGTTAACTAATTAACAA
ATTGCAGTAATGAGTTTTGCAGTTATTGGATGGTCAGTGTAACCATAAGGGACATGGAAAAGATGGGTTGAA
CTTTATATATCTATTTTTTATTTTTTTTCCTTTCCCATAACCAGCC**CGGGCCGTCGACCACGCGT**

Figure 3.45 Nucleotide sequences of the genome walking fragment of *PmERp57* amplified by 2nd-3′next ERp57GW & AP2 primers. Primer sequences are boldfaced and underlined. The adaptor sequence is highlighted.

A)

B)
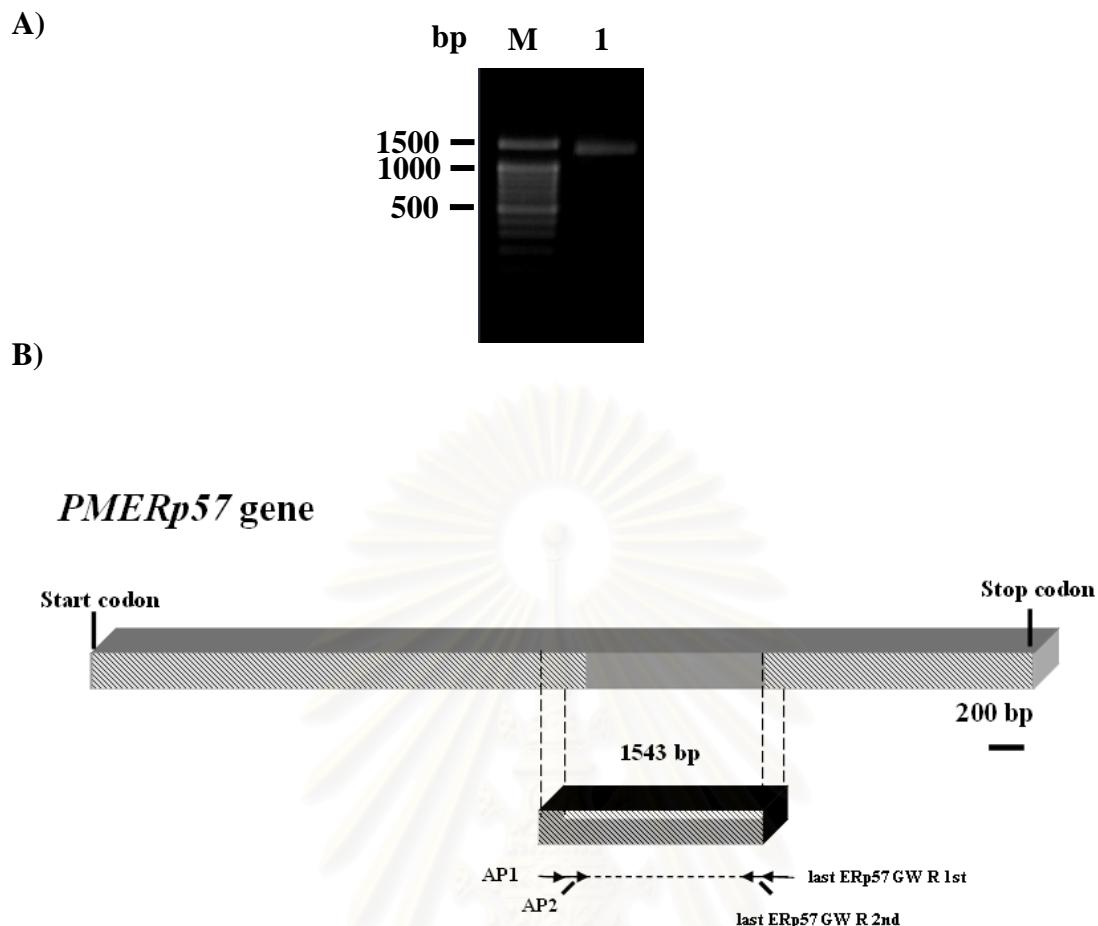
Figure 3.46 Genomic organization of *PMERp57* was further carried out using the last ERp57GW R 2nd & the Adaptor primer 2 (AP2) (lanes 1, A) primers. A 100 bp DNA ladder was used as the marker (A). Genomic orientation of each amplified fragment (B).

GTAATACGACTCACTATAGGGCACGCGTGGTCGACTAAAGTAGTAAATAAAAAAAGTGCATTAAATGTGCAT
GCCTTATGAAAATCAGGCATTTCACTTTTTCAAGCTGTGATTTGTTAACTAATTAACAAATTGCAGTAATGA
GTTTTGCAGTTATTGGATGGTCAGTGTAACCATAAGGGACATGGAAAAGATGGGTTGAACTTTATATATCTA
TTTTTTATTTTTTTTCCTTTCCCATAGACCACTCCAGTTTGTCAAGATTTCTCATCCACTCAATTTCCTATT
AATTTCTCTGCCCAAAGAAAATTAAAACTTTAATTTGTTATACCTTGATTACAAAGCTAGTTTGTCCCTCTT
CCCCCACATTGTATTGGAGAATGCAATCAAGATTTTTTAATGTACACAAAGTAAACCTTTGATTTGCCAAGT
GAGGAATCAAGGTTGAGATTGCTTCAATGAAATTGGACAATGGGACAACTACCATGAAGAGGGAAGTGCATT
TAATGATTTGTTACCACTGAAACTGGAAATCAGTTGTGATGGGAGAAGTTGGGATTTATTTCCCTCGGTGTA
CAAGGCAGGGTATGCTAGAAATTGTAACTGACCACCTATGTTGGACCACATCCTATTCGTTACAGAATGTAT
CATTGTTCGCTGGTTGAATACAGTCAAAATGGTCTGATCAGGCACTTAGAGTTGAATGATAGTTTGAAGTGT
GCAGTTTCATACACTGGCTACTCTCCATGGAAGTAATTTGAAGTAAATTAGGTATTTCAGCATCCTGTTATT
AGGGCTTTTGGTGCTGTATATGAAATGGGCTTTGTCATGTTAAGAAAATTCCCTTAATATTAGGGAAAAGAG
AACTAGGTTTGTCCAAAGCATTTGCTAATAGCAGCACATGTAGGGTATGGAGAACTTTGATACTTGTGGTTT
ACAACATTCATTTGGGCCCCATTTGCAGAGTTCGGCTTCAAACTCGTGTGTGATAGCTACAGAGCTTGGTTG
ATGGCTTTGGTTTGTATTTTTGGACAAGTTACTTGGCATTAAATTATGTCGGGCACACCGGCAAATCAAAAT
GGTTTGGCTGCTGAGTGCACACTTCACACCAATGTTCTTAAAATTTTGTTGCATCAAGCCATTTCTTTTATC
AACACATGTTTGGTTACTAACATAATTGTGAAATATGGATATGAATGTAATTTAAACTGGTGGTTAAAAGTT
TCGACTGTATAGCATGGATTCCTCAGACCAAATGTCTGCTTAAAGAATTTTGCACTTTGAATTTATGCAGTT
GATATTAATGGTGGTAAATGCGTGTTGATTAACCCAAATAACTTTTAACACTAATCATGGTTACAAGATTGA

```
ATGTGTGAAGGTTTAGGTTGAAACCTTTCTAGTATATTTCATAATTGGGCAAATTTTGGGTAGTGTAAAAAA
TTTCCACATGTAATTAATGTTTTCCCTTTATCAGGTGGATTGTACTGATGATGGAAAGGACAGCTGTAGCAG
ATTTGGTGTGTCTGGCTACCCTACCCTGAAG
```

Figure 3.47 Nucleotide sequences of the genome walking fragment of *PmERp57* amplified by the last ERp57GW R and AP2 primers. Primer sequences are boldfaced and underlined. The adaptor sequence is highlighted.

Genomic organization of *Erp57* deduced from nucleotide sequences of genome walking analysis spanned 8254 bp, composing of 10 exons (88-226 bp in length) and 9 introns (93-2787 bp in length) (Figure 3.48). A greter thermal stability in exons (40.8-56.3%) than that in introns (29.0-41.9%) of the PMERp57 gene was observed. Introns 2, 6 and 8 of *PMERp57* interrupt the ORFs between two codon (type 0 intron), whereas other introns interrupt the ORFs within the same codons (type 1 intron). Boundaries of introns 1, 2 and 5 did not follow the GT/AG rule (Figure 3.49 and Table 3.3).

```
ATGGCTACGAGATTGTTAATACTACTCCTCTCCCTCGTGGCCGTGGCGCTGGGAGACGAT    60
GTCCTGCAATTAAACGACGCGGATTTCGACGGGAAAGTGGCCAGCTACCACACGGTCCTC   120
GTCATGTTCTACGCCCCGTGGTgagccttggggtcctgtgtgggtcctttggggtcctgtgg   180
ggtcctctcgcctagttggtcctgtgtgggtcctctcgtgccttggggtcaccttcccccttg   240
agcgaggggagcgctcgcgggtgtttgcgctgccccccttcctacggtctcgtattagggg   300
gccaaacagcgggccttcatagcaggttttgaagcgaatggcttggatgtagattaatct   360
gtaatatattgggtttggttcaagatagggccatggatagacgccagaggcgacgtgtat   420
gcacgcacgcatgctgctgctgcatgcgtgcgagggcgtccgcctttggcaggccgtccc   480
gtggcggcgtggccgggcttggcaggaggcaaagggccgagctggggacttggccgttgg   540
gcctaggcgcggctcaggcagctgttgggttgtgggctttcgtttggttggtaatctcgt   600
aattaatattaccagactgtcaaagctgtaatgtgtgaatattttccgtcgcaagtgata   660
taaatgaagagggtatttgttagtgatgcaccggccccaccagacattctaaaaccagg   720
gtattgtatgaacccaaaaatccaaacttaaccatctctaccttttttttttttattatt   780
acatccccccccttccttatcatcactgtggcaactgtcacaattttaagtgtggaaggt   840
ttgacgtgcgatagtgaggaattgcagaaaagtccttgttctttaagaagggtgggta   900
catttctcttgctcttgataataattacctattccactaaatttgcctgtaaaatgtata   960
ccaaagtatagtgttgcacgactttaatatgaacttgcaaggtaataccatagttttt   1020
aatgtatttgtccattgcagcactgttttcccacagccatgtagtacaagatagactaat   1080
gttaccagtgttatatctgagtcctttattttttcattgcaagtttattttaagaattgc   1140
ttatggagactgcacccctgaaactgtccctattactccagcccctcttggagatacatt   1200
gatgtatacaagaaaaaaactgaccgccgtgcacttgtttccctatcggggctctgac   1260
cccaaccctcccttttttgccatggaactaatgttttgtacaacattaatggacctatgc   1320
tttgatatttaccttccaacacatctaggtaaaaaacttgccaaaatgacccccagtt   1380
cccatacacagtttgaataagtgcgtgaactggggggtggcaatacaccacctttttcttg   1440
ggtgaatgcccaaagggaatgcgtagtcatggcagtttcgattgttgggttaacgttaga   1500
gttgggcacagatagaggtaaatggacccttactatcttataaagcataaaaatattaat   1560
tccccatctaaaaaataaaagggaggaaagttaggtctggatttgggttgataccagccc   1620
```

```
cgggaatccactaagtgaattcgcgggcccctgcaggtccaccattatggggggaaagcct    1680
cccacccgcggtggggagtgccaaagcctgggggaattatcgataattaaaatagtttttc    1740
atttcaggctatgacacttgtgtgcctaacaacttttattattttttgagttgcgatggaa    1800
atgcacaagtttcttgggttcatttgctatagtttgcatggtaaaatgaattgaaagagg    1860
aaaagcagtcttaaggtggtttgttacagtttttaaatacattaattaagaggagccactgt    1920
tgattaatgttttcaagtgcatagtgaaccacaactgcttgagactacaacatgtgcagt    1980
cattcttgtttatcctagcaagctaaacacagtggtgacttatttttatttaggcttttta    2040
gccttgttacatccttctgcagctacaaaggtcatgatttttttcctgcacaatttccttt    2100
tatgggcttgatcaatgcactaccagatgcagggcttggtataagtgaaacatactaatt    2160
ttgatcaatgaaatctcttcgtaatgacatataggtcttgaaagccaagtcattgctatg    2220
agtaacccatttccataaagatcttttaacttaatgggtaaccccttcccctctaaggt    2280
gggtttgggctggttcttactaaaatttcatggaaatattttaaagtttacatttgatt    2340
acaatttcatttaataattagattgggtttccacagcatagaaaaactacggaaagtgag    2400
taataattcttagtatccctggcataagtgccaaactaatcttctattttgggtaattga    2460
ttatgaagttaatgcaccttagaatgtcccagactaaggggttacttatgttaatccagt    2520
cattagttttttttgttttttgttttttttacttccaaatcggtactatagttgatagatc    2580
aggaggttttagaaagtggtaaataaatgctgtagttggatgctcttaactataaaccca    2640
tttactaggcttatagtgtagcatagcatttctgcaaccaagtttcttctatggttgaac    2700
aaatgctagttagtggcacagcacttactttgagcgcatggtatttaatacataggaata    2760
taactgatcctggaaatgacaggtttgtcaaacctccaccccaatccacccttcacaaaa    2820
gagaaggattttaataagttgctgaagcaaaatgtgcctgaatgtatgggatagtctcag    2880
gattatgctggagaagttgtcaggcattaagttgactttttttccaggtGTGGTCACTGC    2940
AAGAGATTAAAGCCTGAGTTTGAGAAGGCCTCTACCACCTTGAAGGCCAACGACCCTCCC    3000
GTCTACCTTGCTAaggtaagctggcaaatttgggtgcaggttaccttgtggattttaggt    3060
tgtagatcttgaagcaagttttacagagtaaaatttaagtacagcttgtgtacaatagct    3120
gattacagttgttagcaaagcctagcaatatatttgtactacactataactaacatctgt    3180
tgcaaatgtcttgattaaagaattgtagggaaggtattttcattgattttggccatgtcc    3240
aagttcattttgcataacatgaacagtgggtgtggttatattaaactaagtatgtatccc    3300
atgtcatgttttgactgaattttggtgtcgttctaatttttgacatgagcttgcagctac    3360
tataaagcaaatttccccccttcacggctctttatcagcaacgtaagaatttctattgga    3420
tatgtgagttaatgcaacacccatgctcagcagtctgatgctgagattatattttaact    3480
ttaaatactaacttgttcaaaaggttaaaccttccatttggaaaccctgctgaagggtgt    3540
tgtgctgttgcagtggtatagcaaattttttgcacagataggtgttggaagtatgaatttt    3600
gatcttgaccagttgaacatgaaatgttggattaatgcccaatttgagttttgtatgtgt    3660
acattttagcgttttgtaaactttccaaattgctactgaaagcgaagaccagggctccct    3720
tccacaatgagtaaaataacatgccctgctttccttgagtgctcttccctactagcagcc    3780
ctactatacacctggtttcatatctgcatttgagattttgcctacatctcttctgtagac    3840
ttctgattcagtagattgtgattttttttatattcaaaacctgtatactagataattttta    3900
gtttgacaattggtccttccagtacaatgtttacccaatttattcgtgataaatcgtacc    3960
acttccgtcgcaatcacagcccagctatctttgatatacttcatgccctcccccaactatt    4020
gggactagtcagagtattgtgtaaaaccatcacaatttattaattttttggcatttggtgg    4080
gcttgaatactaaactgaagtctattctatttagcaatatattgataaatgacctggagt    4140
ttggttgcactttattccaacaaatgctggcaagtctcgcaacacatcctggtgctgttc    4200
atacacatttgcaagttattgtacagaatagtgggaaaactattaatgcattaaagtagt    4260
aaataaaaaagtgcattaaatgtgcatgccttatgaaaatcaggcatttcactttttca    4320
agctgtgatttgttaactaattaacaaattgcagtaatgagttttgcagttattggatgg    4380
tcagtgtaaccataagggacatggaaaagatgggttgaactttatatatctattttttat    4440
ttttttttcctttcccatagaccactccagtttgtcaagatttctcatccactcaatttcc    4500
tattaatttctctgcccaaagaaaattaaaactttaatttgttataccttgattacaaag    4560
ctagtttgtccctcttcccccacattgtattggagaatgcaatcaagatttttaatgta    4620
cacaaagtaaacctttgatttgccaagtgaggaatcaaggttgagattgcttcaatgaaa    4680
ttggacaatgggacaactaccatgaagagggaagtgcatttaatgatttgttaccactga    4740
aactggaaatcagttgtgatgggagaagttgggatttatttccctcggtgtacaaggcag    4800
ggtatgctagaaattgtaactgaccacctatgttggaccacatcctattcgttacagaat    4860
gtatcattgttcgctggttgaatacagtcaaaatggtctgatcaggcacttagagttgaa    4920
tgatagtttgaagtgtgcagtttcatacactggctactctccatggaagtaatttgaagt    4980
aaattaggtatttcagcatcctgttattagggctttttggtgctgtatatgaaatgggctt    5040
tgtcatgttaagaaaattcccttaatattagggaaaagagaactaggtttgtccaaagca    5100
tttgctaatagcagcacatgtagggtatggagaactttgatacttgtggtttacaacatt    5160
catttgggccccatttgcagagttcggcttcaaactcgtgtgtgatagctacagagcttg    5220
gttgatggctttggttttgtattttttggacaagttacttggcattaaaattatgtcgggcac    5280
```

```
accggcaaatcaaaatggtttggctgctgagtgcacacttcacaccaatgttcttaaaat   5340
tttgttgcatcaagccatttcttttatcaacacatgtttggttactaacataattgtgaa   5400
atatggatatgaatgtaatttaaactggtggttaaaagtttcgactgtatagcatggatt   5460
cctcagaccaaatgtctgcttaaagaattttgcactttgaatttatgcagttgatattaa   5520
tggtggtaaatgcgtgttgattaacccaaataacttttaacactaatcatggttacaaga   5580
ttgaatgtgtgaaggtttaggttgaaacctttctagtatatttcataattgggcaaattt   5640
tgggtagtgtaaaaaatttccacatgtaattaatgttttccctttatcAGGTGGATTGTA   5700
CTGATGATGGAAAGGACAGCTGTAGCAGATTTGGTGTGTCTGGCTACCCTACCCTGAAGA   5760
TCTTCAAGGGAGGAGAGCTCTCTACGGACTACAATGGTCCACGAGATGCCAgtaagtgct   5820
tgatgtacttatgtatcttgatcacaaataagatttaacaccaagcccccttgtctgtcc   5880
ttttgtgcatattctcccttttaaatgtttcttctccttccagGTGGTATTGTAAAGTAC   5940
ATGAGGTCACAGGTTGGACCAGCCTCTAAGGAGTTGACATCCGTGGAGGCAGCAGAAGCA   6000
TTCCTTGGTGCTGCTGAAGTTGGAGTCGTTTACTTTGGAGGAGATTCCAAACTTAAAGgt   6060
ttgtattcccatgctgttactttgaaatgggtatttgattcaatgctttaccaatgaaaa   6119
ggggaaaagacaatctaaatgctattcattgcataaatctagaatttaaacattcttagt   6179
tttaaacttcaaaatgcaaacactactgcatatagcatcccaaatatttttgaaaccagg   6237
acaagtccaattacatcttgcttcagacatgtgagaatgctatgccattaggttggtggc   6297
agtgaactcttacacttttcatgagcatcataggtttgcacttcagaagtaaccttccaa   6357
acaatcctgtttcactttgataactttttttgcagtgtgcagttgtgctgaaatcttttgc   6417
cttgatttcagtacaggtggtccattttgttttactattcctatagtatgggaatattat   6477
taaagcagtataaccatttgaaatggttttgtttcaagtaattgaccaatgaatcctgga   6537
ttaccttgaaacattgaatatatttcagATGCTTTCCTAAAGGCTGCTGATAAGCTGAGG   6596
GAATCCATCCGTTTTGCACACTCCCTCGATGCCACTGTTAATGAAAAGTATGGGTACAGC   6656
GATGTTGTTGTACTTTTCCGACCGAAACACCTGGAGAACAAATTTGAGCCTTCCTCTGTT   6716
GTATTTGAGGGATCGGCAGACAGGGCTGAGATTGAGTCTTTCATCAAAAAGAACTTCCAT   6776
GGTTtttcttgcatgcatacttgctcactccattgcaataacttcaagtacatattttgt   6836
tttagtttatactactactgttgccttgtttgattctgcagactatgcaaggaatctatt   6896
gtaatacgatctctcattgtggctttatattaaccatcagagctcttgtctgaacactta   6956
aatgctatgtatattgaatagatgcaagtatttgacctgtcggcttcaagatcaattact   7016
gttctttacagccatggttTGGTAGGACACCTAACGCAAGACACTGCTCAGGATTTCAAA   7076
CCTCCAGTTGTGATTGCTTACTACAATGTTGATTACATCAAAAATGTTAAGGGTACAAAT   7136
TACTGGCGCAATCGTGTCCTTAAGgtaagactttgtatgtttggaaaaaggtttacagaa   7196
ttcatttcttagctttgacatgtgtaggagagttgacaagtagttaaaaggtgttgtatt   7256
tgcagGTGGCACAAAACTTTGCTGATGACTTCAAGTTTGCCGTTGCCAATAAGGACGACT   7316
TCCAGCATGACCTCAATGAATATGGCCTTGATTATGTTCCTGGTGACAAGCCAGTAATTT   7376
GTGCACGTAATGCTAAAGCCCAGAAGTTTGTCATGCAGGAAGAATTTTGgtaagttcttg   7436
tgtagcctgtttttgtttttgtatttctaatatttgcctgtattatatttaacaacttactc   7496
atctctaacttcacttccatagAATGGATAACCTCCAAGCATTCCTCACCAATCTCAAGG   7556
CGGGTGAGCTTGAGCCATATCTGAAGTCTGAGGCAGTGCCAACACAAGATGGCCCTGTCA   7616
CTGTTGCTGTGGGTAAGAACTTCAATGAAGTTGTCTCTGATGAGCGTGATGCCCTCATTG   7676
AATTCTATGCTCCTTGGTGTGGTCACTGCAAGAAATTAGCGCCCACCTATGATGAGCTGG   7736
GAGAAGCGgtaaggttttctaattattcttgtgtattcatttgagaatgaatactggat   7796
aattactgtacaatgcattctttctggtaaacactaatttccttacagATGAAGGATGAA   7856
GATGTAGACATTGTGAAGATGGATGCCACTGCCAATGATGTTCCTCCTCAGTACAATGTT   7916
CAAGGCTTCCCCACCATCTTCTGGAAACCCAAGGGTGGTGTTCCAAGGAATTACAATgta   7976
agtagagcctattttataccaattttttgacattgaatctaatataggagttcatagacac   8036
tgagcttcaatcaaatgcttgttcaaaggcacccaagctctttgatatgcagaattgaaa   8096
aatatttgatttggtgaactaaacttgatcttaacttccccagGGTGGCCGGGAACTGGA   8156
CGATTTTGTCAAGTACATTGCCCAACATTCCACAAATGAACTGAATGGGTATGACCGCAA   8216
GGGGAAGGCAAAGAAAGGCAAGAAGACTGAACTT**TGA**                            8254
```

Figure 3.48 Organization of the PMERp57 gene. Nucleotides of exons are capitalized while those of introns are highlighted and shown in lower letters. Start and stop codons are illustrated in boldfaced.

Table 3.3 GC content and length of exons and introns in the *PMERp57* gene

Compositions of gene	Genomic DNA		mRNA	GC content (%)		
Exon	(No. of nucleotides)		(No. of nucleotides)			
1	1-140	(140 bp)	1-140	56.3	%	
2	2928-3015	(88 bp)	141-228	52.4	%	
3	5691-5811	(121 bp)	229-349	50.4	%	
4	5924-6058	(135 bp)	350-484	47.4	%	
5	6566-6778	(213 bp)	485-697	44.9	%	
6	7034-7161	(128 bp)	698-825	40.8	%	
7	7263-7426	(164 bp)	826-989	43.3	%	
8	7520-7745	(226 bp)	990-1215	49.6	%	
9	7846-7974	(129 bp)	1216-1344	45.0	%	
10	8141-8254	(114 bp)	1345-1458	47.4	%	
Intron	Genomic DNA		Type	GC content (%)		GT/AG rule
	(No. of nucleotides)					
1	141-2927	(2787 bp)	1	41.9	%	No
2	3016-5690	(2605 bp)	0	35.9	%	No
3	5812-5923	(112 bp)	1	37.5	%	Yes
4	6059-6565	(507 bp)	1	33.9	%	Yes
5	6779-7033	(255 bp)	1	34.9	%	No
6	7162-7262	(101 bp)	0	33.7	%	Yes
7	7427-7519	(93 bp)	1	30.1	%	Yes
8	7746-7845	(100 bp)	0	29.0	%	Yes
9	7975-8140	(166 bp)	0	33.7	%	Yes
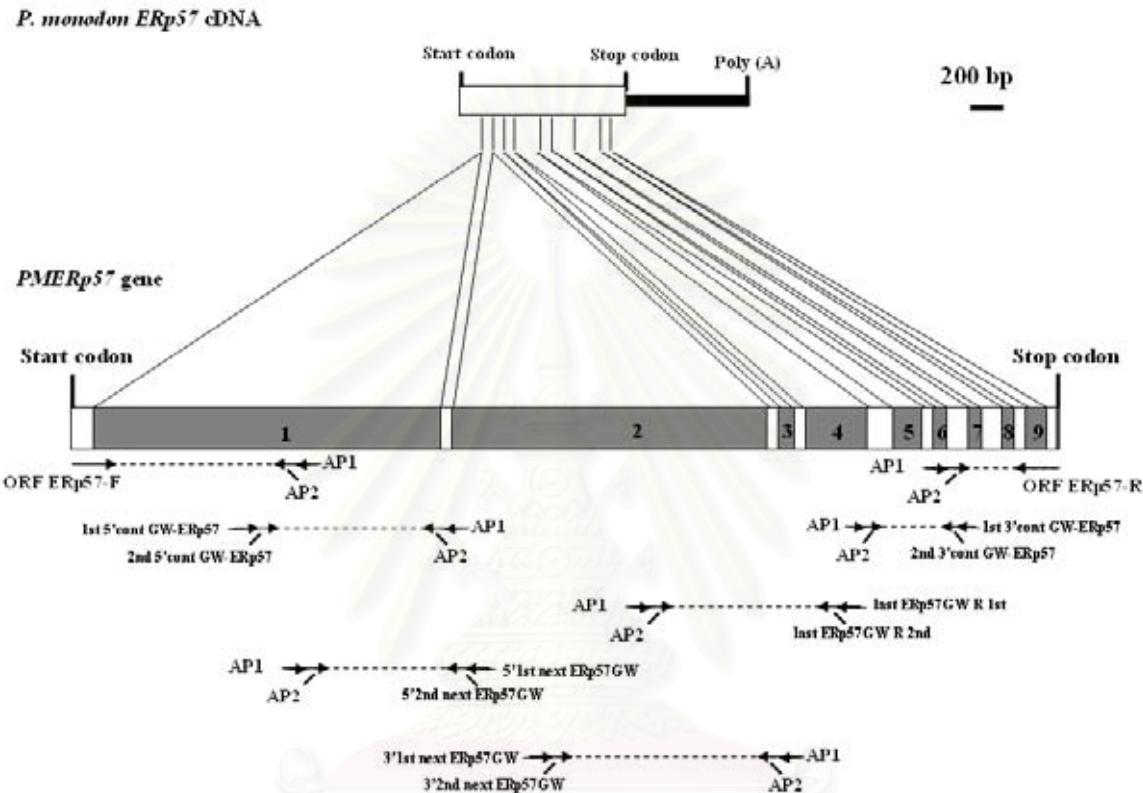
Figure 3.49 Schematic diagrams of *P. monodon* ERp57 cDNA and gene. Coding regions are represented open bars. Introns (with numbers) are grey shaded. Primers used for amplification of the *PMERp57* gene are illustrated.

3.6 Tissue expression analysis of *PmCRT*, *PmCNX* and *PmERp57* analyzed by RT-PCR

Tissue expression analysis of *PmCRT*, *PmCNX* and *PmERp57* was examined revealed in various tissues of a female and testis of male juveniles and broodstock of *P. monodon*.

PmCRT was abundantly expressed in intestine of juvenile shrimp. Lower expression was observed hepatopancreas, ovary, heart, pleopod and gill. Rare expression of *PmCRT* was observed in eyestalk, hemocyte lymphoid organ and antennal gland (Figure 3.50). In broodstock, *PmCRT* was comparably expressed in various tissues but very low expression was found in thoracic ganglion, testis and eyestalk.
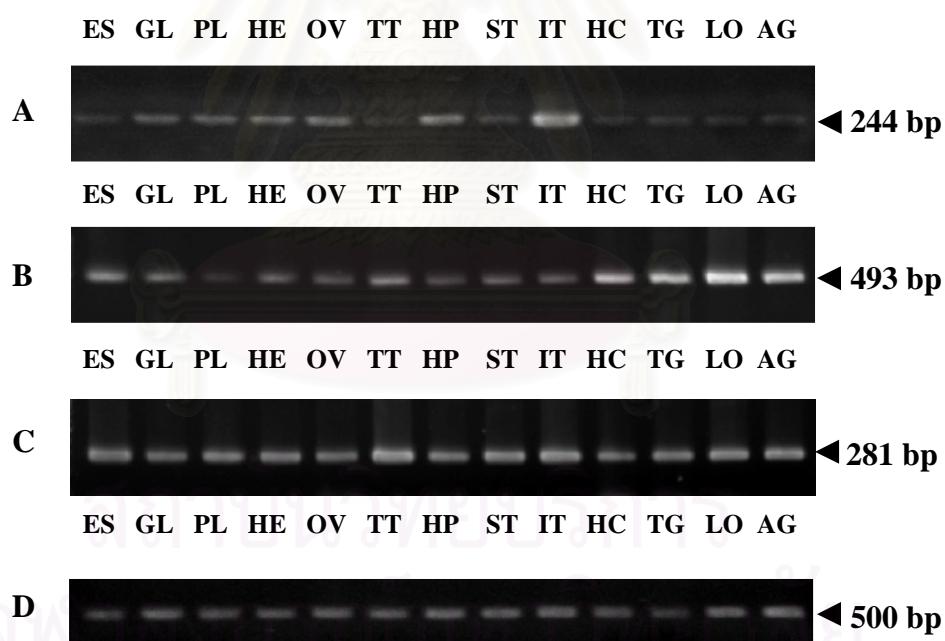
Figure 3.50 1.5% ethidium bromide-stained agarose gel showing results from tissue expression analysis of *PmCRT* (A), *PmCNX* (B), *PmERp57* (C) using the first stand cDNA of eye stalk (ES), pleopod (PL), gill (GL), heart (HE), ovary (OV), testis (TT), hepatopancreas (HP), stomach (ST), intestine (IT), hemoctye (HC), thoracic ganglia (TG), lympiod organ (LO) and antennal gland (AG) of *P. monodon* juveniles (28 cycles). *EF 1-α* was successfully amplified from the same template and included as the positive control (D).

PmCNX was abundantly expressed in hemocyte, thoracic ganglion, lymphoid organ and antennal gland of juvenile *P. monodon*. This transcript was moderately expressed in the remaining tissues except pleopod where low abundant expression was observed (Figure 3.50). *PmCNX* was constitutively expressed in all tissues but low expression was observed in eyestalk, gill and thoracic ganglia (Figure 3.51).

PmERp57 was comparably expressed in al examined tissues of juveniles (Figure 3.50) but the expression of this transcript was quite low in testis of *P. monodon* broodstock (Figure 3.51).
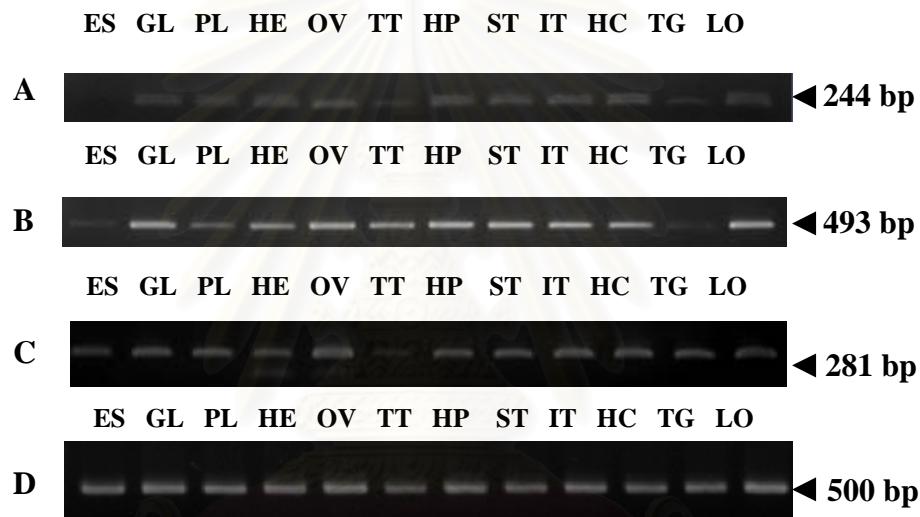
Figure 3.51 The 1.5% ethidium bromide-stained agarose gel showing results from tissue distribution analysis of homologue of *CRT* (A), *CNX* (B), *ERp57* (C) using the first stand cDNA synthesized from total RNA of eyestalk (ES), pleopod (PL), gill (GL), heart (HE), ovary (OV), testis (TT), hepatopancreas (HP), stomach (ST), intestine (IT), thoracic ganglia (TG) and lymphiod organ (LO) of *P. monodon* broodstock (28 cycles). *EF 1-α* was successfully amplified from the same template and included as the positive control (D).

3.7 Semiquantitative RT-PCR of *PmCRT*, *PmCNX* and *PmERp57* upon induction by thermal stress.

3.7.1 Optimization of semi-quantitative RT-PCR conditions

The first stand cDNA of of *PmCRT*, *PmCNX* and *PmERp57* in heamocytes, hepatopancrease and gill of juvenile shrimps under the normal condition and after temperature stress for 0, 1, 3, 6, 12, 24 and 48 hours ($N = 3$ for each group) was subjected to semiquantitative RT-PCR analysis. This technique requires optimization of several parameters including concentration of primers, $MgCl_2$, and the number of PCR cycles.

Primers for the target genes were designed. *EF-1 α* was used as the internal control. Non-quantitative RT-PCR was carried out using 100 ng of the first stand cDNA template from hemocytes, hepatopancreas and gill of *P. monodon* using high stringency conditions (annealing temperature of 65ºC for *PmCRT* and *PmERp57* and 69ºC for *PMCNX*), 1 U of Dynazyme DNA polymerase and 0.2 mM $MgCl_2$ for 30 cycles. The most suitable primer and $MgCl_2$ concentrations and number of cycles were optimized and used for semiquantitative RT-PCR analysis.

3.7.1.1 Optimization of the primer concentration

RT-PCR of each gene was carried out with fixed components except primer concentrations (0.05, 0.10, 0.15, 0.20 and 0.25 μM) as described above. Lower concentrations of primers may result in non-quantitative amplification whereas higher concentrations of primer may leave a large amount of unused primers which could give rise to non-specific amplification products. The suitable concentration of primers for each gene in each tissue is shown in Table 3.4 (Figure 3.52-3.54 A).

3.7.1.2 Optimization of the $MgCl_2$ concentration

The optimal concentration of $MgCl_2$ (between 1.0, 1.5, 2.0 and 2.5 mM) for each primer pair was carefully examined using the amplification conditions with the optimized primer concentration (Figure 3.52-3.54 B). The concentration of $MgCl_2$ that gave the highest yields and specificity for each PCR product was chosen (Table 3.4).
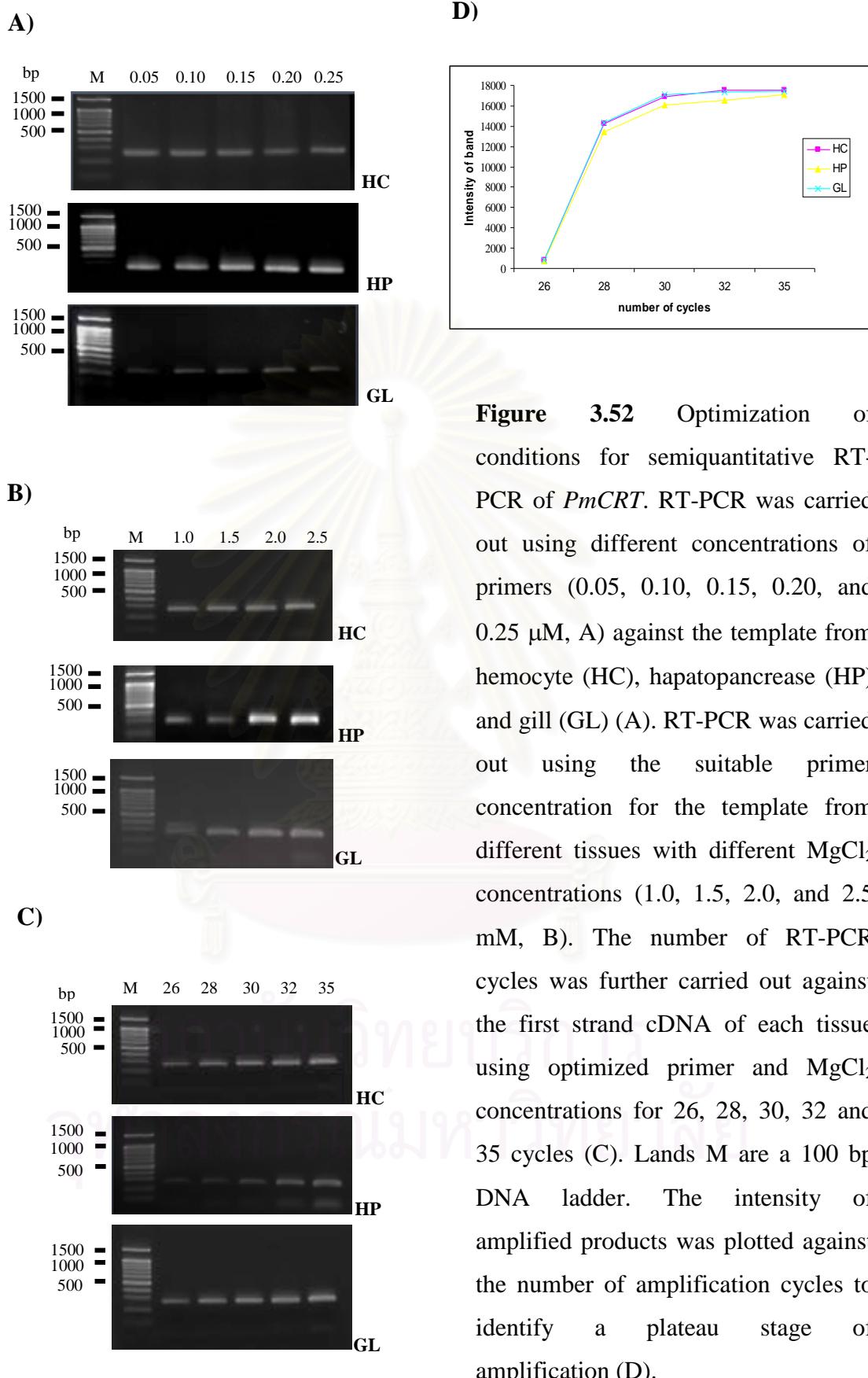
Figure 3.52 Optimization of conditions for semiquantitative RT-PCR of *PmCRT*. RT-PCR was carried out using different concentrations of primers (0.05, 0.10, 0.15, 0.20, and 0.25 μM, A) against the template from hemocyte (HC), hapatopancrease (HP) and gill (GL) (A). RT-PCR was carried out using the suitable primer concentration for the template from different tissues with different MgCl$_2$ concentrations (1.0, 1.5, 2.0, and 2.5 mM, B). The number of RT-PCR cycles was further carried out against the first strand cDNA of each tissue using optimized primer and MgCl$_2$ concentrations for 26, 28, 30, 32 and 35 cycles (C). Lands M are a 100 bp DNA ladder. The intensity of amplified products was plotted against the number of amplification cycles to identify a plateau stage of amplification (D).

A)

B)

C)

D)

Figure 3.53 Optimization of conditions for semiquantitative RT-PCR of *PmCNX*. RT-PCR was carried out using different concentrations of primers (0.05, 0.10, 0.15, 0.20, and 0.25 μM, A) against the template from hemocyte (HC), hapatopancrease (HP) and gill (GL) (A). RT-PCR was carried out using the suitable primer concentration for the template from different tissues with different $MgCl_2$ concentrations (1.0, 1.5, 2.0, and 2.5 mM, B). The number of RT-PCR cycles was further carried out against the first strand cDNA of each tissue using optimized primer and $MgCl_2$ concentrations for 26, 28, 30, 32 and 35 cycles (C). Lands M are a 100 bp DNA ladder. The intensity of amplified products was plotted against the number of amplification cycles to identify a plateau stage of amplification (D).

A)

B)

C)

D)

Figure 3.54 Optimization of conditions for semiquantitative RT-PCR of *PmERp57*. RT-PCR was carried out using different concentrations of primers (0.05, 0.10, 0.15, 0.20, and 0.25 μM, A) against the template from hemocyte (HC), hapatopancrease (HP) and gill (GL) (A). RT-PCR was carried out using the suitable primer concentration for the template from different tissues with different $MgCl_2$ concentrations (1.0, 1.5, 2.0, and 2.5 mM, B). The number of RT-PCR cycles was further carried out against the first strand cDNA of each tissue using optimized primer and $MgCl_2$ concentrations for 26, 28, 30, 32 and 35 cycles (C). Lands M are a 100 bp DNA ladder. The intensity of amplified products was plotted against the number of amplification cycles to identify a plateau stage of amplification (D).

3.7.1.3 Optimization of the cycle numbers

The number of amplification cycles was important because the product reflecting the expression level should be measured quantitatively before reaching a plateau amplification phase. At the plateau stage, transcripts initially present at different levels may give equal intensity of the amplification products.

In this experiment, RT-PCR of each gene was performed using the optimized primers and $MgCl_2$ concentrations for 26, 28, 30, 32 and 35 cycles (Figure 3.52-3.54 C and D). The number of cycles that gave the highest yield before the product reached a plateau phase of amplification was chosen (Table 3.4).

Table 3.4 Optimal primer and $MgCl_2$ concentrations and the number of amplification cycles for semiquantitative analysis of *PmCRT*, *PmCNX* and *PmERp57* in hemocytes, hepatopancreas and gill of juvenile *P. monodon*

Transcript	Tissues	Suitable RT-PCR component		
		Primer concentrations (μM)	MgCl2 concentrations (mM)	PCR cycles
PmCRT	hemocyte	0.2	2.5	30
	hepatopancrease	0.2	2.0	30
	gill	0.1	1.5	30
PmCNX	hemocyte	0.05	2.0	30
	hepatopancrease	0.1	1.5	30
	gill	0.2	2.5	30
PmERp57	hemocyte	0.2	2.5	30
	hepatopancrease	0.2	2.0	30
	gill	0.2	1.5	30

3.7.1.4 *Elongation factor 1-α* (*EF-1 α*)

The most suitable primer and $MgCl_2$ concentrations for semiquantitative RT-PCR of *EF-1α* were 0.2 μM and 1.5 mM of $MgCl_2$ with the thermal profile of 94 °C for 5 minutes followed by 23 cycles of 94°C for 30 seconds, 58°C for 45 seconds and 72°C for 45 seconds. The final extension was carried out at 72°C for 7 minutes. The expression level of *EF-1α* of the same tissue among different groups of samples was not significantly different ($P < 0.05$) indicating that the expression level *EF-1* was reasonably suitable to be used as the internal control (Figure 3.55-3.56).

3.7.2 Semi-quantitative RT-PCR analysis

3.7.2.1 *PmCRT*

The relative expression levels of *PmCRT* in hepatopancreas and gill was greater than that of hemocytes ($P < 0.05$). The expression level of *PmCRT* in hemocyte but not hepatopaancreas and gill was significantly increased after temperature stress (Figures 3.57 - 3.62). The highest expression level of *PmCRT* in hemocytes was observed at 0 hour post treatment (0 hpt; 0.863 ± 0.011, $P < 0.05$). The level of *PmCRT* was slightly decreased from that at 0 hpt but still significantly higher than that of control at 1 hpt (0.679 ± 0.021, $P < 0.05$). The expression level of *PmCRT* retuned to the normal levels at 3 (0.476 ± 0.029, $P > 0.05$), 6 (0.428 ± 0.012, $P > 0.05$), 12 (0.479 ± 0.044, $P > 0.05$), 24 (0.478 ± 0.020, $P > 0.05$) and 48 hpt (0.442 ± 0.029, $P > 0.05$) (Figures 3.57 and 3.58; Table 3.5).

3.7.2.2 *PmCNX*

Like *PmCNX*, the relative expression levels of *PmCNX* in hepatopancreas and gill was significantly greater than that of hemocytes ($P < 0.05$). After temperature stress, only the expression level of *PmCNX* in hemocytes was significantly increased (Figures 3.63 - 3.68). *PmCNX* in hemocytes of juvenile *P. monodon* was up-regulated and exhibited the highest expression level at 0 hpt (0.659 ± 0.008, $P < 0.05$). The expression of this gene was slightly decreased from that at 0 hpt but still significantly greater than that of the control (0.362 ± 0.022) at 1 hpt (0.534 ± 0.018, $P < 0.05$). The

expression of *PmCNX* at longer time course (3, 6, 12, 24 and 48 hpt) after treatment was comparable with that of control (Figures 3.63-3.64, Table 3.5).
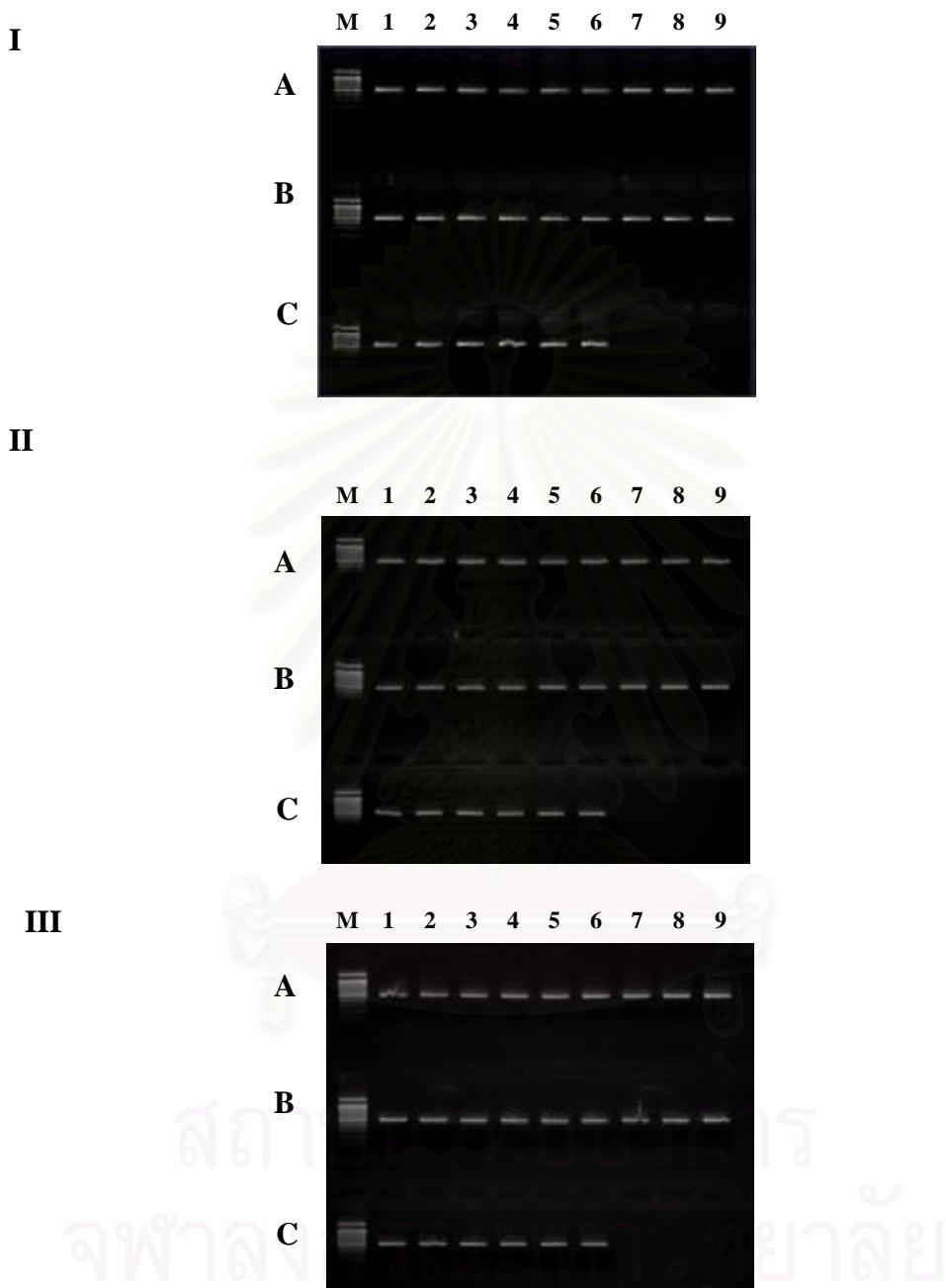
Figure 3.55 A 1.5% ethidium bromide-stained agarose gel showing expression levels of *EF-1α* in hemocyte (I), hepatopancrease (II) and gill (III) of *P. monodon*. The template was from juvenile shrimp under the ambient temperature as the control (lane A 1-3) and that after heat stress at 35 ºC for 0 (lane A, 4-6), 1 (lane A, 7-9), 3 (lane B, 1-3), 6 (lane B, 4-6), 12 (lane B, 7-9), 24 (lane C, 1-3) and 48 (lane, C 4-6) as the treatment. Lane M = 100 bp DNA ladder.

A)
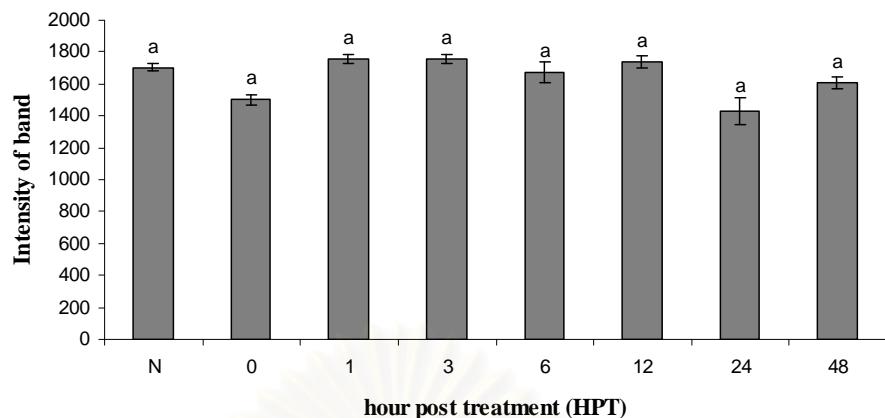
B)
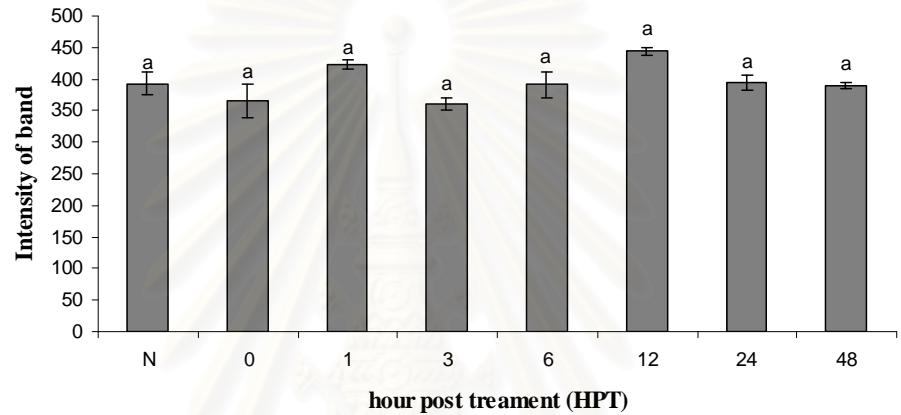
C)
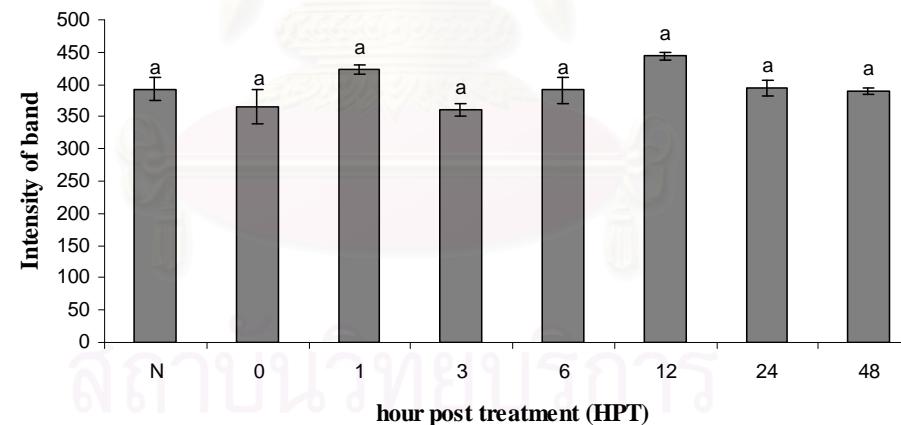
Figure 3.56 Histograms showing the time-course relative expression levels of *EF-1α* in hemocyte (A), hepatopancrease (B) and gill (C) for 0, 1, 3, 6, 12, 24 and 48hours post treatment of thermal stress and normal shrimp as control (N).

3.7.2.3 *PmERp57*

Similarly, the relative expression levels of *PmERp57* in hepatopancreas and gill was greater than that of hemocytes ($P < 0.05$). The expression level of *PmERp57* in hemocyte but not hepatopaancreas and gill was significantly increased after

temperature stress (Figures 3.69 - 3.74). The highest expression level of *PmERp57* in hemocytes was observed at 0 hpt (0.609 ± 0.012, P < 0.05). The level of *PmERp57* was slightly decreased from that at 0 hpt but still significantly higher than that of control at 1 hpt (0.391 ± 0.006, P < 0.05). The expression level of *PmCRT* retuned to the normal levels at 3 (0.345 ± 0.006, P > 0.05), 6 (0.380 ± 0.014, P > 0.05), 12 (0.354 ± 0.044, P > 0.01), 24 (0.439 ± 0.014, P > 0.05) and 48 hpt (0.349± 0.043, P > 0.05) (Figure 3.69 and 3.70; Table 3.5).

M 1 2 3 4 5 6 7 8 9

A

B

C

Figure 3.57 A 1.8% ethidium bromide-stained agarose gel showing the expression level of *PmCRT* in hemocytes of *P. monodon* for 0 (lane A, 4-6), 1 (lane A, 7-9), 3 (lane B, 1-3), 6 (lane B, 4-6), 12 (lane, B 7-9), 24 (lane, C 1-3) and 48 (lane C, 4-6) post treatment. The first strand cDNA template from hemocytes of untreated shrimp was included as the control (lane A, 1-3). Lane M = 100 bp DNA ladder.
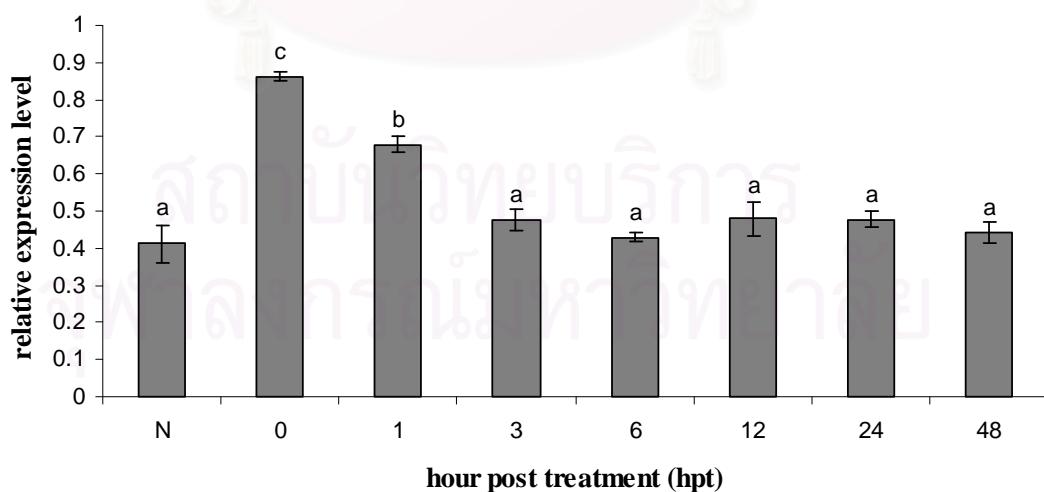
Figure 3.58 Histograms showing time-course relative expression levels of *PmCRT* in hemocyte of juvenile *P. monodon* for 0, 1, 3, 6, 12, 24 and 48 hours post thermal

treatment. Normal shrimp (N) were included as the control. The same letters indicated that the expression levels were not significantly different (P>0.05).
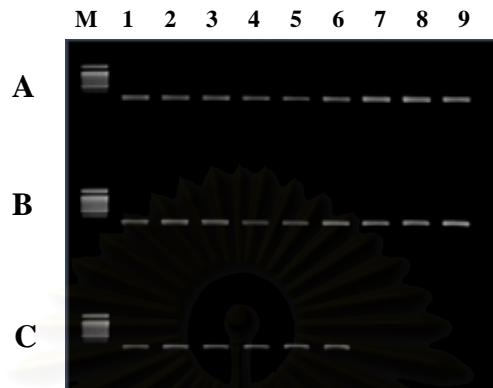
Figure 3.59 A 1.8% ethidium bromide-stained agarose gel showing the expression level of *PmCRT* in hepatopancreas of *P. monodon* for 0 (lane A, 4-6), 1 (lane A, 7-9), 3 (lane B, 1-3), 6 (lane B, 4-6), 12 (lane, B 7-9), 24 (lane, C 1-3) and 48 (lane C, 4-6) post treatment. The first strand cDNA template from hemocytes of untreated shrimp was included as the control (lane A, 1-3). Lane M = 100 bp DNA ladder.
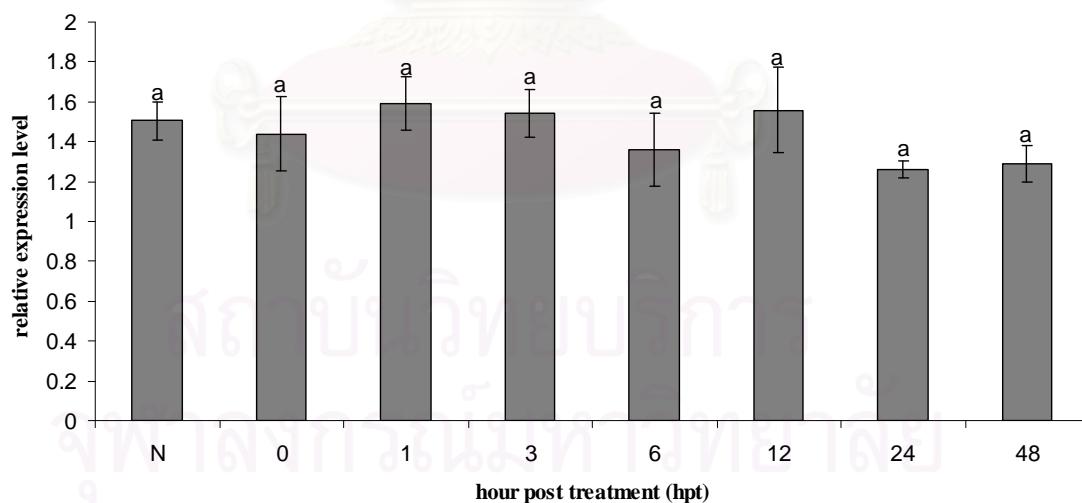
Figure 3.60 Histograms showing time-course relative expression levels of *PmCRT* in hepatopancreas of juvenile *P. monodon* for 0, 1, 3, 6, 12, 24 and 48 hours post thermal treatment. Normal shrimp (N) were included as the control. The same letters indicated that the expression levels were not significantly different (P>0.05).
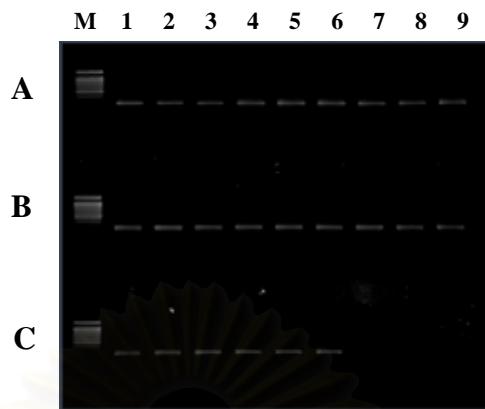
Figure 3.61 A 1.8% ethidium bromide-stained agarose gel showing the expression level of *PmCRT* in gill of *P. monodon* for 0 (lane A, 4-6), 1 (lane A, 7-9), 3 (lane B, 1-3), 6 (lane B, 4-6), 12 (lane, B 7-9), 24 (lane, C 1-3) and 48 (lane C, 4-6) post treatment. The first strand cDNA template from hemocytes of untreated shrimp was included as the control (lane A, 1-3). Lane M = 100 bp DNA ladder.
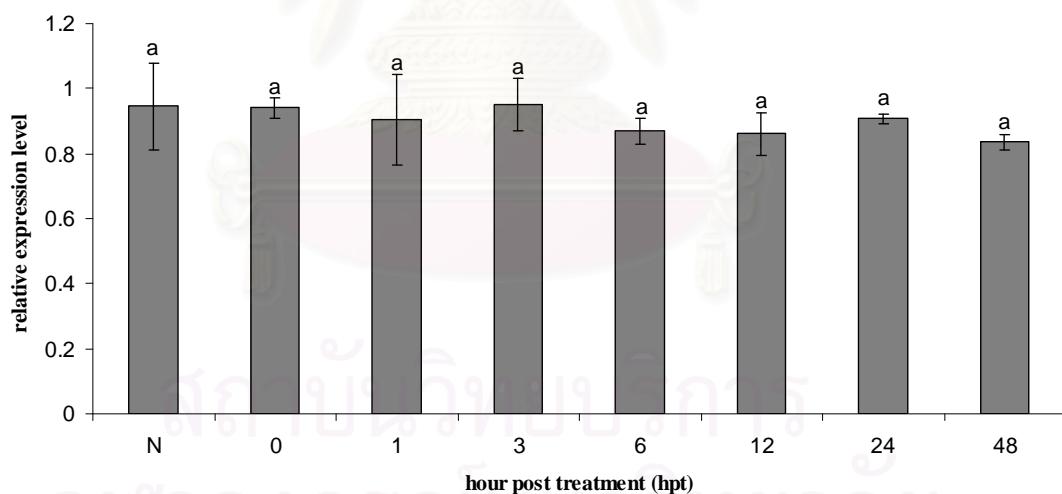
Figure 3.62 Histograms showing time-course relative expression levels of *PmCRT* in gill of juvenile *P. monodon* for 0, 1, 3, 6, 12, 24 and 48 hours post thermal treatment. Normal shrimp (N) were included as the control. The same letters indicated that the expression levels were not significantly different (P>0.05).

Figure 3.63 A 1.8% ethidium bromide-stained agarose gel showing the expression level of *PmCNX* in hemocytes of *P. monodon* for 0 (lane A, 4-6), 1 (lane A, 7-9), 3 (lane B, 1-3), 6 (lane B, 4-6), 12 (lane, B 7-9), 24 (lane, C 1-3) and 48 (lane C, 4-6) post treatment. The first strand cDNA template from hemocytes of untreated shrimp was included as the control (lane A, 1-3). Lane M = 100 bp DNA ladder.
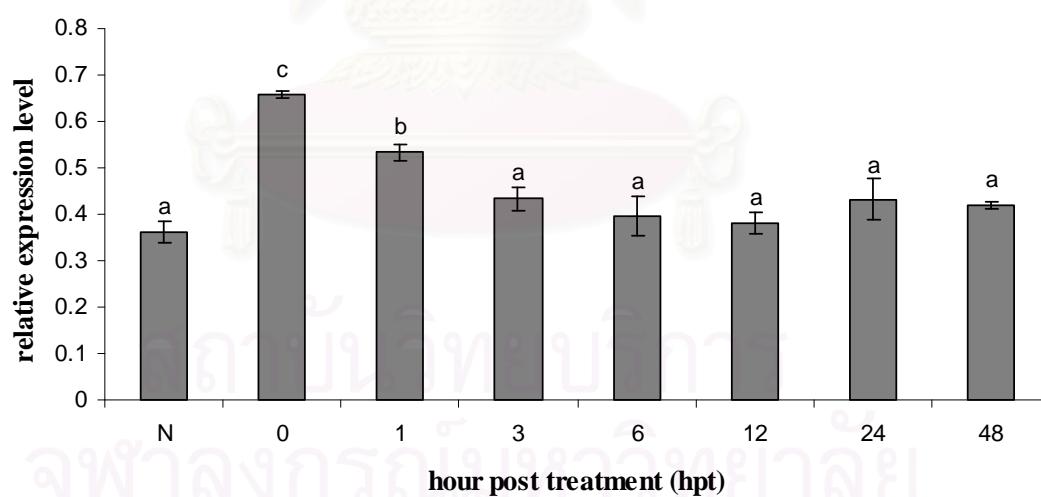
Figure 3.64 Histograms showing time-course relative expression levels of *PmCNX* in hemocyte of juvenile *P. monodon* for 0, 1, 3, 6, 12, 24 and 48 hours post thermal treatment. Normal shrimp (N) were included as the control. The same letters indicated that the expression levels were not significantly different (P>0.05).
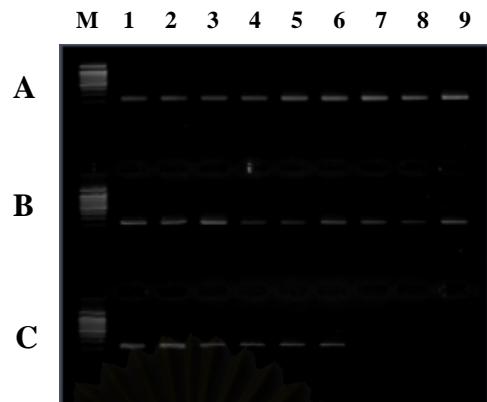
Figure 3.65 A 1.8% ethidium bromide-stained agarose gel showing the expression level of *PmCNX* in hepatopancreas of *P. monodon* for 0 (lane A, 4-6), 1 (lane A, 7-9), 3 (lane B, 1-3), 6 (lane B, 4-6), 12 (lane, B 7-9), 24 (lane, C 1-3) and 48 (lane C, 4-6) post treatment. The first strand cDNA template from hemocytes of untreated shrimp was included as the control (lane A, 1-3). Lane M = 100 bp DNA ladder.
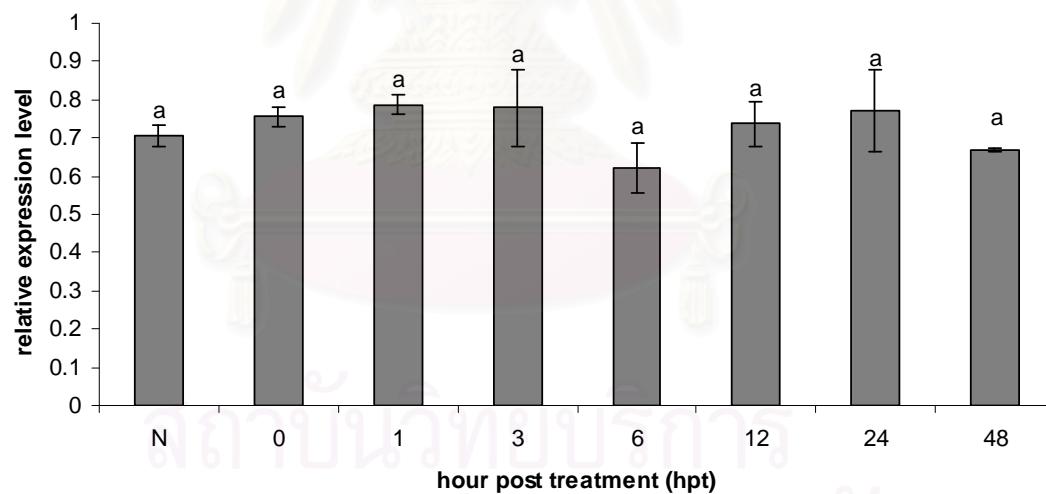
Figure 3.66 Histograms showing time-course relative expression levels of *PmCNX* in hepatopancreas of juvenile *P. monodon* for 0, 1, 3, 6, 12, 24 and 48 hours post thermal treatment. Normal shrimp (N) were included as the control. The same letters indicated that the expression levels were not significantly different (P>0.05).
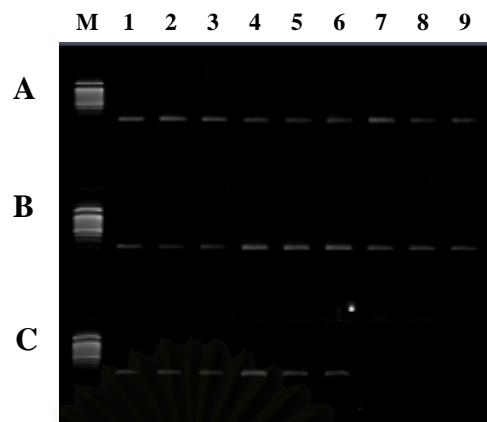
Figure 3.67 A 1.8% ethidium bromide-stained agarose gel showing the expression level of *PmCNX* in gill of *P. monodon* for 0 (lane A, 4-6), 1 (lane A, 7-9), 3 (lane B, 1-3), 6 (lane B, 4-6), 12 (lane, B 7-9), 24 (lane, C 1-3) and 48 (lane C, 4-6) post treatment. The first strand cDNA template from hemocytes of untreated shrimp was included as the control (lane A, 1-3). Lane M = 100 bp DNA ladder.
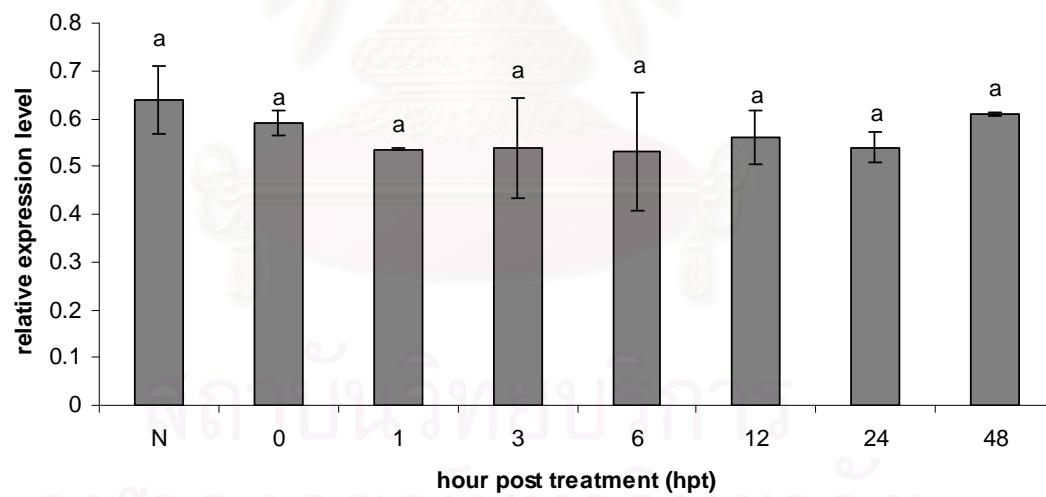
Figure 3.68 Histograms showing time-course relative expression levels of *PmCNX* in gill of juvenile *P. monodon* for 0, 1, 3, 6, 12, 24 and 48 hours post thermal treatment. Normal shrimp (N) were included as the control. The same letters indicated that the expression levels were not significantly different (P>0.05).

Figure 3.69 A 1.8% ethidium bromide-stained agarose gel showing the expression level of *PmERp57* in hemocytes of *P. monodon* for 0 (lane A, 4-6), 1 (lane A, 7-9), 3 (lane B, 1-3), 6 (lane B, 4-6), 12 (lane, B 7-9), 24 (lane, C 1-3) and 48 (lane C, 4-6) post treatment. The first strand cDNA template from hemocytes of untreated shrimp was included as the control (lane A, 1-3). Lane M = 100 bp DNA ladder.
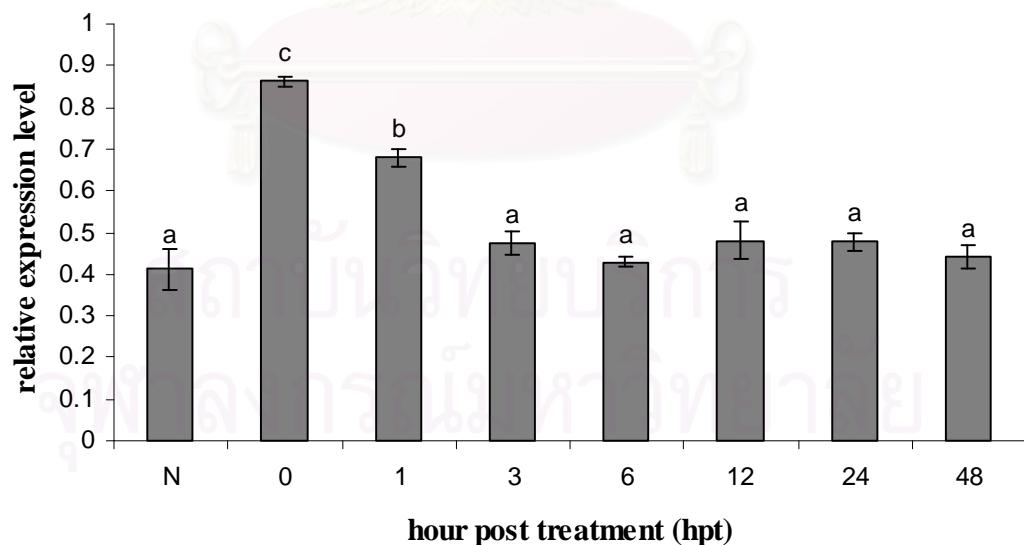
Figure 3.70 Histograms showing time-course relative expression levels of *PmERp57* in hemocyte of juvenile *P. monodon* for 0, 1, 3, 6, 12, 24 and 48 hours post thermal treatment. Normal shrimp (N) were included as the control. The same letters indicated that the expression levels were not significantly different (P>0.05).
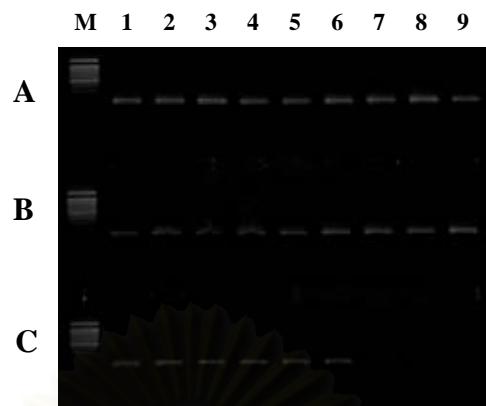
Figure 3.71 A 1.8% ethidium bromide-stained agarose gel showing the expression level of *PmERp57* in hepatopancreas of *P. monodon* for 0 (lane A, 4-6), 1 (lane A, 7-9), 3 (lane B, 1-3), 6 (lane B, 4-6), 12 (lane, B 7-9), 24 (lane, C 1-3) and 48 (lane C, 4-6) post treatment. The first strand cDNA template from hemocytes of untreated shrimp was included as the control (lane A, 1-3). Lane M = 100 bp DNA ladder.
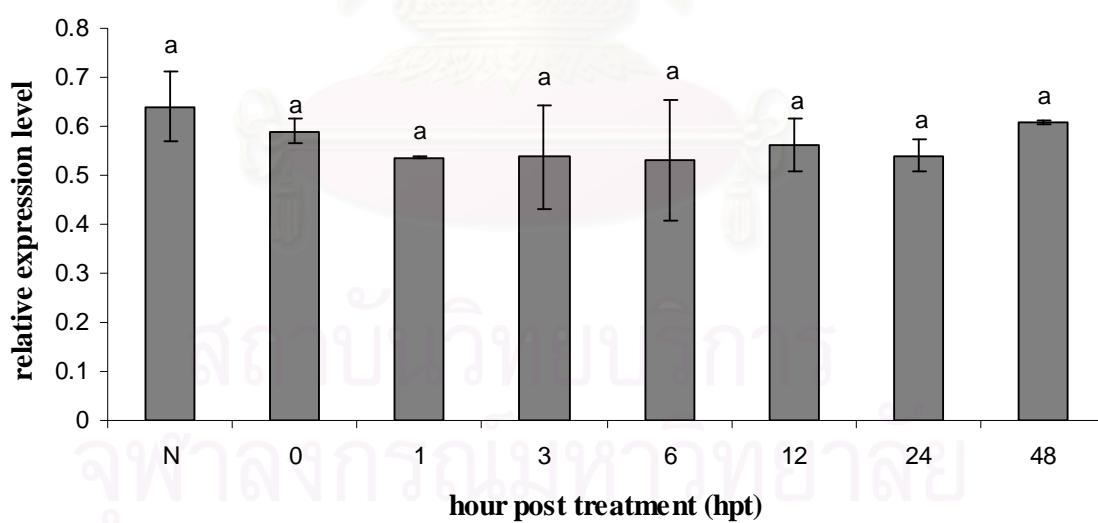
Figure 3.72 Histograms showing time-course relative expression levels of *PmERp57* in hepatopancreas of juvenile *P. monodon* for 0, 1, 3, 6, 12, 24 and 48 hours post thermal treatment. Normal shrimp (N) were included as the control. The same letters indicated that the expression levels were not significantly different ($P > 0.05$).
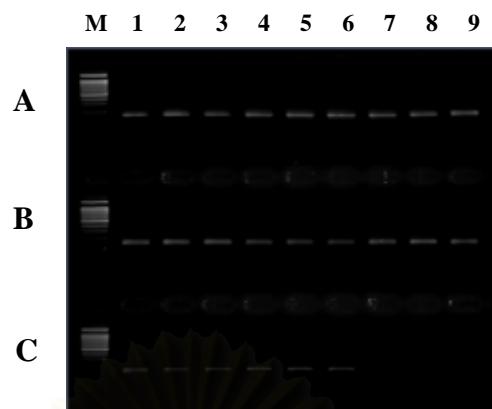
Figure 3.73 A 1.8% ethidium bromide-stained agarose gel showing the expression level of *PmERp57* in gill of *P. monodon* for 0 (lane A, 4-6), 1 (lane A, 7-9), 3 (lane B, 1-3), 6 (lane B, 4-6), 12 (lane, B 7-9), 24 (lane, C 1-3) and 48 (lane C, 4-6) post treatment. The first strand cDNA template from hemocytes of untreated shrimp was included as the control (lane A, 1-3). Lane M = 100 bp DNA ladder.
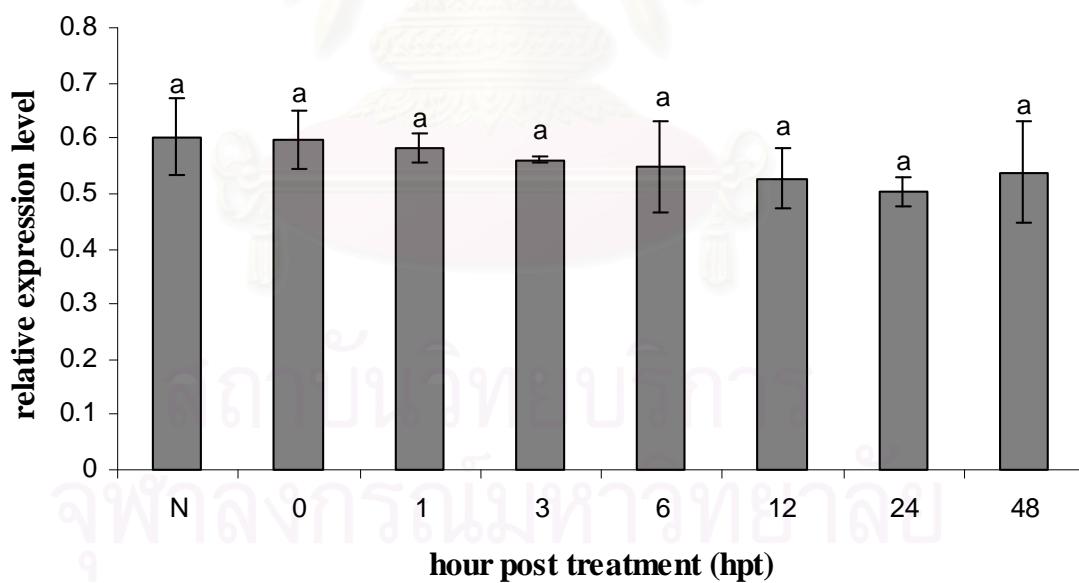

Figure 3.74 Histograms showing time-course relative expression levels of *PmERp57* in gill of juvenile *P. monodon* for 0, 1, 3, 6, 12, 24 and 48 hours post thermal treatment. Normal shrimp (N) were included as the control. The same letters indicated that the expression levels were not significantly different (P>0.05).

Table 3.5 A time-course analysis of relative expression levels of *PmCRT*, *PmCNX* and PmERp57 in hemocytes, hepatopancreas and gill of juvenile *P. monodon* by semiquantitative RT-PCR

Hours post treatment (HPT)	Mean relative expression level								
	Hemocytes			Hepatopancrease			Gill		
	PmCRT	*PmCNX*	*PmERp57*	*PmCRT*	*PmCNX*	*PmERp57*	*PmCRT*	*PmCNX*	*PmERp57*
N	0.412 ± 0.051^a	0.362 ± 0.022^a	0.384 ± 0.010^a	1.504 ± 0.094^a	0.705 ± 0.029^a	0.629 ± 0.012^a	0.945 ± 0.132^a	0.640 ± 0.071^a	0.602 ± 0.07^a
0	0.863 ± 0.011^c	0.659 ± 0.008^c	0.609 ± 0.012^c	1.439 ± 0.184^a	0.756 ± 0.025^a	0.651 ± 0.098^a	0.941 ± 0.031^a	0.590 ± 0.026^a	0.597 ± 0.052^a
1	0.679 ± 0.021^b	0.534 ± 0.018^b	0.392 ± 0.007^b	1.589 ± 0.134^a	0.787 ± 0.026^a	0.566 ± 0.084^a	0.904 ± 0.14^a	0.536 ± 0.002^a	0.583 ± 0.026^a
3	0.476 ± 0.029^a	0.433 ± 0.024^a	0.345 ± 0.010^a	1.544 ± 0.12^a	0.779 ± 0.099^a	0.615 ± 0.066^a	0.951 ± 0.081^a	0.537 ± 0.105^a	0.561 ± 0.006^a
6	0.428 ± 0.012^a	0.396 ± 0.042^a	0.380 ± 0.014^a	1.358 ± 0.181^a	0.621 ± 0.064^a	0.675 ± 0.042^a	0.869 ± 0.041^a	0.530 ± 0.124^a	0.548 ± 0.084^a
12	0.479 ± 0.044^a	0.381 ± 0.024^a	0.353 ± 0.001^a	1.558 ± 0.215^a	0.737 ± 0.058^a	0.647 ± 0.051^a	0.861 ± 0.065^a	0.561 ± 0.055^a	0.527 ± 0.053^a
24	0.478 ± 0.020^a	0.432 ± 0.043^a	0.4399 ± 0.01^a	1.260 ± 0.042^a	0.771 ± 0.108^a	0.572 ± 0.028^a	0.907 ± 0.015^a	0.540 ± 0.032^a	0.505 ± 0.027^a
48	0.442 ± 0.029^a	0.420 ± 0.008^a	0.349 ± 0.004^a	1.288 ± 0.091^a	0.670 ± 0.004^a	0.579 ± 0.015^a	0.835 ± 0.023^a	0.608 ± 0.004^a	0.538 ± 0.092^a

*The expression of *EF-1α* was normalized to 1.00. The relative expression level of the target genes was determined as the ratio between that of the target and *EF-1α*. The same superscripts between different time interval data idicate non-significant differences of eexpression levels (*P*>0.05)

3.8 Quantitative analysis of *PmCRT*, *PmCNX* and *PmERp57* in hemocytes of juvenile shrimp by quantitative real-time PCR

Semiquantitative RT-PCR clearly illustrated that temperature stress resulted in upregulation of *PmCRT*, *PmCNX* and *PmERp57* in hemocytes but not in hepatopancreas and gill. Accordingly, quantitative analysis on expression of those genes in hemocytes of the control and treated shrimp was reanalyzed using real-time PCR.

The standard amplification curve of the target (*PmCRT*, *PmCNX* and *PmERp57*) and *EF-1α* were constructed (Figure 3.75). Due to large differences between expression levels of *PmCRT*, *PmCNX* and *PmERp57* and *EF-1α*, quantitative real-time PCR was carried out using 50 ng and 11.25 ng of the first strand cDNA template for the target and the control, respectively. Results from real-time PCR revealed that the expression level of *EF-1α* between different groups of specimens was not statistically significant indicating that *EF-1α* was an appropriate house-keeping gene. Its suitable used as the internal control ($P > 0.05$, Figure 3.76).

The relative expression level of *PmCRT* at 0-12 hpt was significantly greater than that of the control ($P < 0.05$). Levels of *PmCRT* mRNA at 0 (3.363 ± 0.400) and 1 hpt (3.355 ± 0.204) were comparable ($P > 0.05$). Although the expression level of this transcript was obviously decreased at 3 (0.743 ± 0.181) 6 (0.626 ± 0.156) and 12 hpt (0.801 ± 0.233), its expression level was still significant greater than that of the control (0.104 ± 0.015) ($P < 0.05$). The expression level *PmCRT* returned to the normal state at 24 and 48 hpt, ($P > 0.05$) (Figure 3.77 and Table 3.6).

PmCNX in hemocytes of *P. monodon* was also upregulated following the temperature stress. The relative expression level of this transcript was 0.071 ± 0.018. The highest relative expression level of *PmCNX* was observed at 0 hpt (0.658 ± 0.12, $P < 0.05$). Expression of *PmCNX* was significantly reduced from 0 hpt for approximately 2 fold at 1 hpt (0.332 ± 0.005) and still further decreased for about 3 fold at 3 (0.1886 ± 0.034), 6 (0.223 ± 0.074) and 12 hpt (0.224 ± 0.068), respectively. Interestingly, the expression level of *PmCNX* at 24 ((0.034 ± 0.001) and 48

(0.034±0.001 and 0.038±0.007) hpt was significant lower than that of the control (P <
0.05) (Figure 3.78 and Table 3.6).

A)

B)

C)

D)
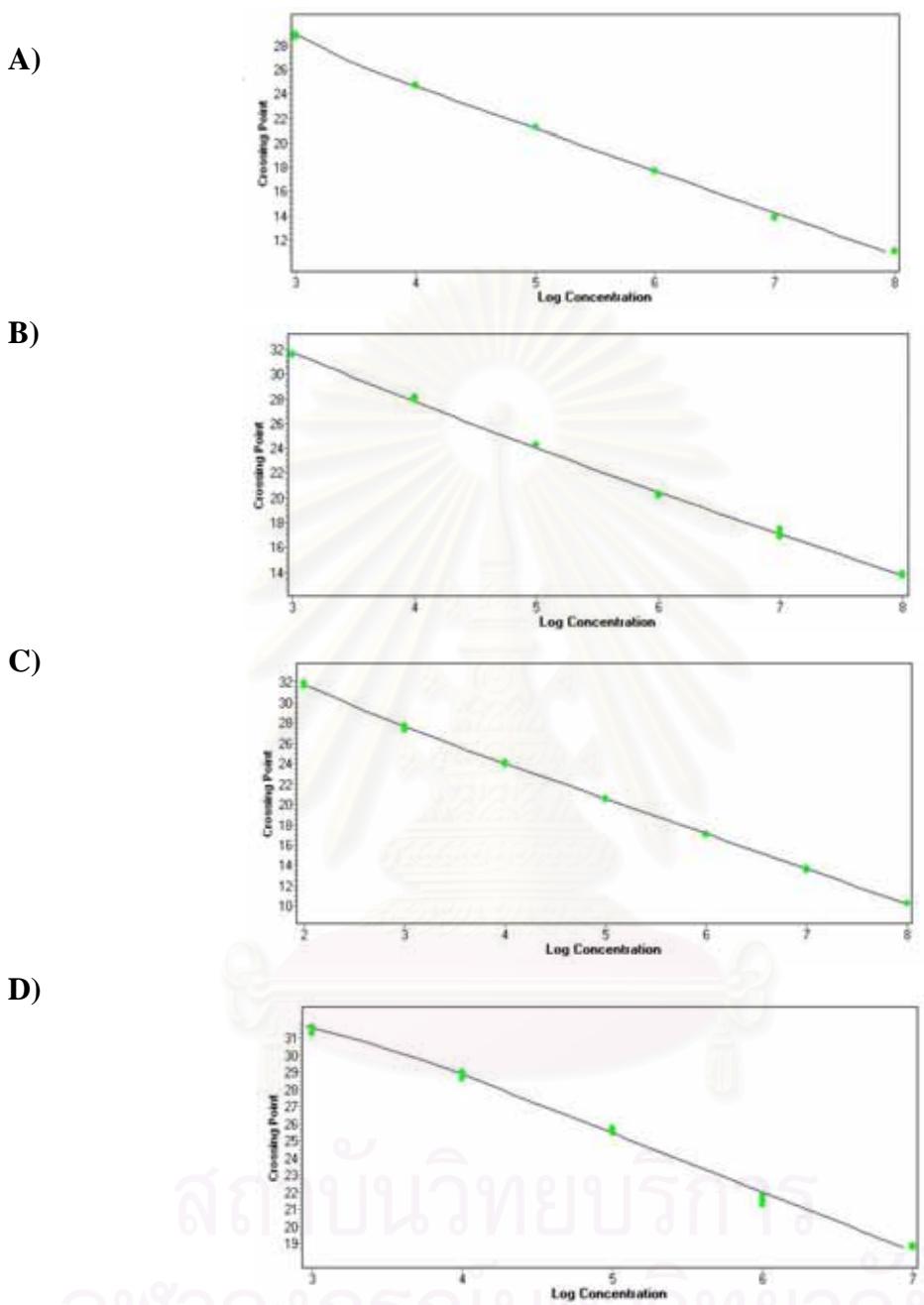
Figure 3.75 The standard amplification curves of various genes examined by real-
time PCR. The standard curve of *EF-1α* (A; r^2 for standard curve = 0.9967, efficiency
for the amplification = 1.941), *PmCRT* (B; r^2 = 0.9985, efficiency for the
amplification = 1.997), *PmCNX* (C; r^2 = 0.974, efficiency for the amplification =
1.948) and *PmERp57* (D; r^2 = 0.9755, efficiency for the amplification = 1.951). The
abscissa reveals log copy number concentrations of each gene (10^3 to 10^8 copy,
respectively).
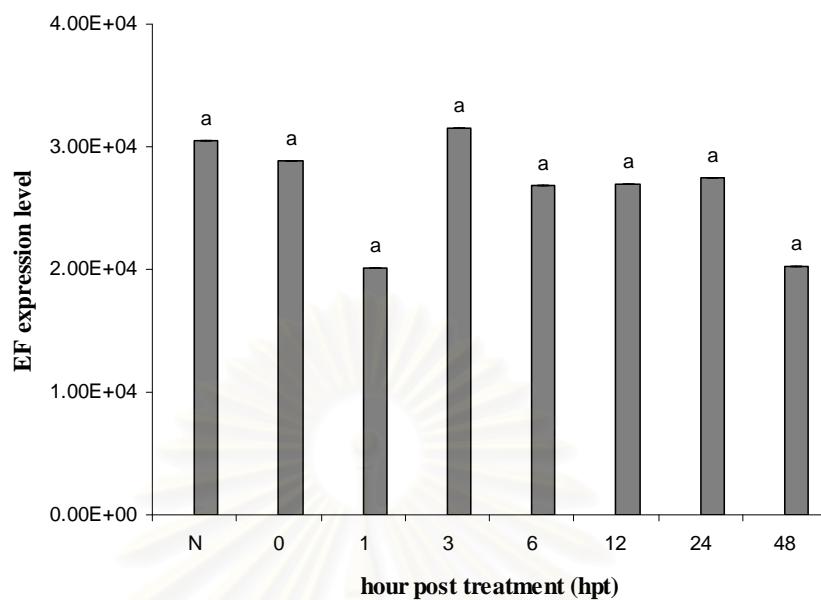
Figure 3.76 Real-time PCR analysis illustrating the absolute expression level (copy number) of *EF-1α* in hemocytes of juvenile *P. monodon* at the normal conditions (N) and at 0, 1, 3, 6, 12, 24 and 48 hours post temperature stress. The same letters indicated that the expression levels were not significantly different ($P > 0.05$).
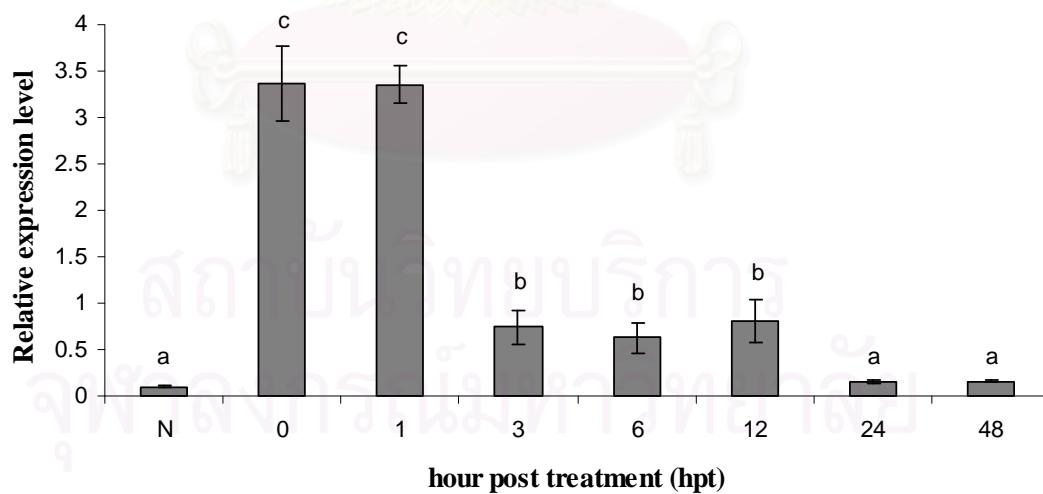
Figure 3.77 Real-time PCR analysis illustrating the relative expression of *PmCRT* in hemocytes of juvenile *P. monodon* at the normal conditions (N) and at 0, 1, 3, 6, 12, 24 and 48 hours post temperature stress. The same letters indicated that the expression levels were not significantly different ($P > 0.05$).
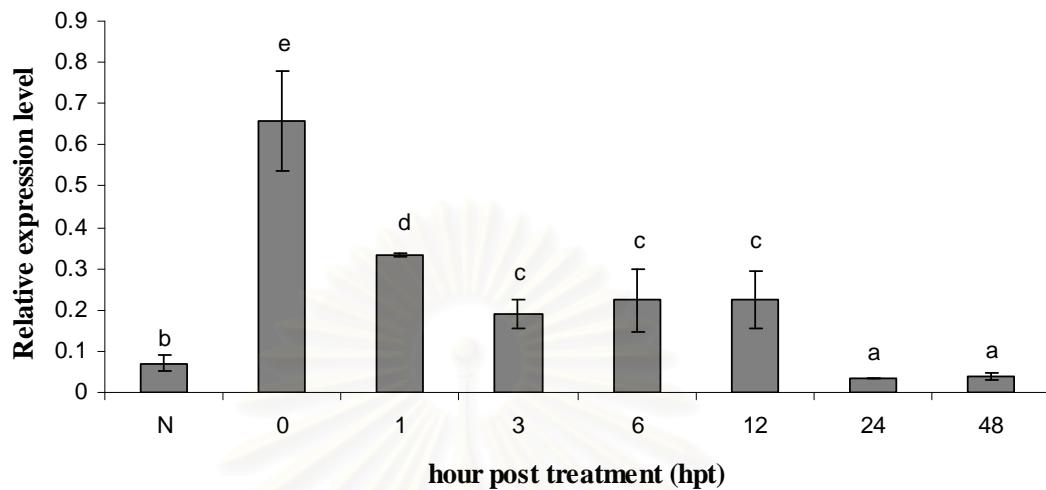
Figure 3.78 Real-time PCR analysis illustrating the relative expression of *PmCNX* in hemocytes of juvenile *P. monodon* at the normal conditions (N) and at 0, 1, 3, 6, 12, 24 and 48 hours post temperature stress. The same letters indicated that the expression levels were not significantly different (P > 0.05).
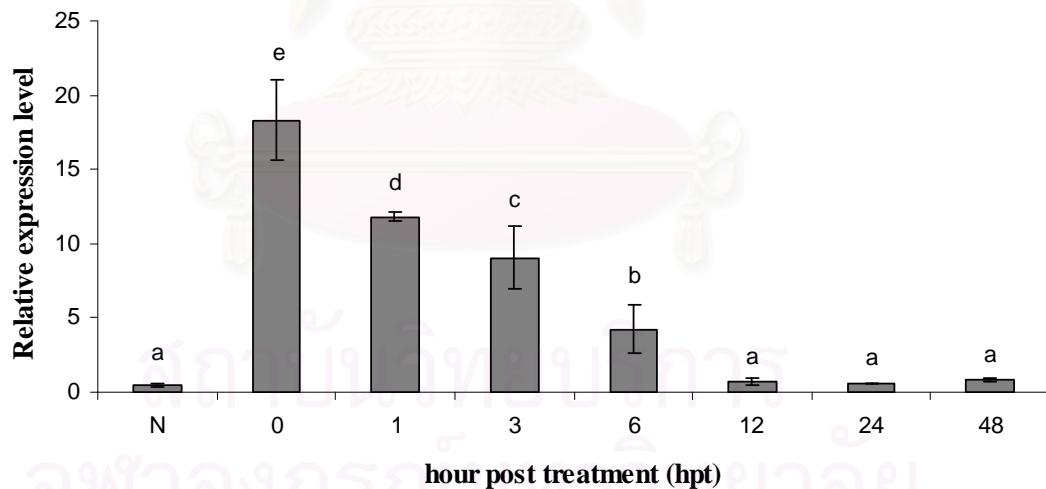
Figure 3.79 Real-time PCR analysis illustrating the relative expression of *PmERp57* in hemocytes of juvenile *P. monodon* at the normal conditions (N) and at 0, 1, 3, 6, 12, 24 and 48 hours post temperature stress. The same letters indicated that the expression levels were not significantly different (P > 0.05).

Quantitative real-time PCR illustrated that *PmERp57* was more temperature-sensitive than *PmCRT* and *PmCNX*. *PmERp57* in hemocytes of *P. monodon* was rapid upregulated following the temperature stress. The relative expression level of this transcript at 0 hpt (18.324±2.691) was approximately 37 times greater than that of the control (0.482±0.11, P < 0.05). A time course reduction of *PmERp57* mRNA was observed at 1 (11.812±0.274), 3 (9.048±2.07) and 6 (4.254±1.647) hpt but the expression level was still greater than that of the control for about 24, 18 and 8 times, respectively. Levels of *PmERp57* at 12 (0.732±0.213), 24 (0.595±0.054) and 48 (0.858±0.161) hpt were comparable with that of the control (P > 0.05) (Figure 3.79 and Table 3.6).

Table 3.6 A time-course relative expression levels of PmCRT, PmCNX and PmERp57 in hemocytes of normal and temperature stress shrimp using quantitative real-time PCR

Hour Post Treatment (HPT)	Mean relative expression level		
	PmCRT	*PmCNX*	*PmERp57*
N	0.104 ± 0.015^a	0.071 ± 0.018^b	0.482 ± 0.11^a
0	3.363 ± 0.400^c	0.658 ± 0.120^e	18.324 ± 2.691^e
1	3.355 ± 0.204^c	0.332 ± 0.005^d	11.812 ± 0.274^d
3	0.743 ± 0.181^b	0.1886 ± 0.034^c	9.048 ± 2.07^c
6	0.626 ± 0.156^b	0.223 ± 0.074^c	4.254 ± 1.647^b
12	0.801 ± 0.233^b	0.224 ± 0.068^c	0.732 ± 0.213^a
24	0.155 ± 0.017^a	0.034 ± 0.001^a	0.595 ± 0.054^a
48	0.157 ± 0.011^a	0.038 ± 0.007^a	0.858 ± 0.161^a

*The expression of *EF-1α* was normalized to 1.00. The relative expression level of the target genes was determined as the ratio between that of the target and *EF-1α*. The same superscripts between different time interval data are not significantly different (*P*>0.05)

3.9 *In vitro* expression of interesting genes using the bacterial expression system

3.9.1 Construction of recombinant plasmid in cloning and expression vector

Before construction of recombinant plasmid, the ORF amplified primers of each interesting gene that covered from start to stop codon were designed from RACE-full length cDNA. The ORF PCR products of *PmCRT*, *PmCNX* and *PmERp57* were 1221, 1788 and 1458 bp in length and encoded 406, 595 and 465, respectively (Figure 3.80). The ORF sequence of *PmCRT* significantly matched that of *Apis mellifera* (E-value = 2e-157) while *PmCNX* still matched with *Aedes aegypti* (E-value = 1e-170) and another protein similar to *ERp57* of *Bombyx mori* (E-value = 8e-155). All of them were ligated, cloned into pGEM-T easy vector and transformed into *E. coli* JM109 (Figure 3.81). The recombinant plasmid of the positive clone was re-sequenced to confirm the orientation and nucleotide sequence of a partial recombinant clone. Recombinant plasmid in the cloning vector was used as the template for amplification of the fragment that inserted into expression vectors.
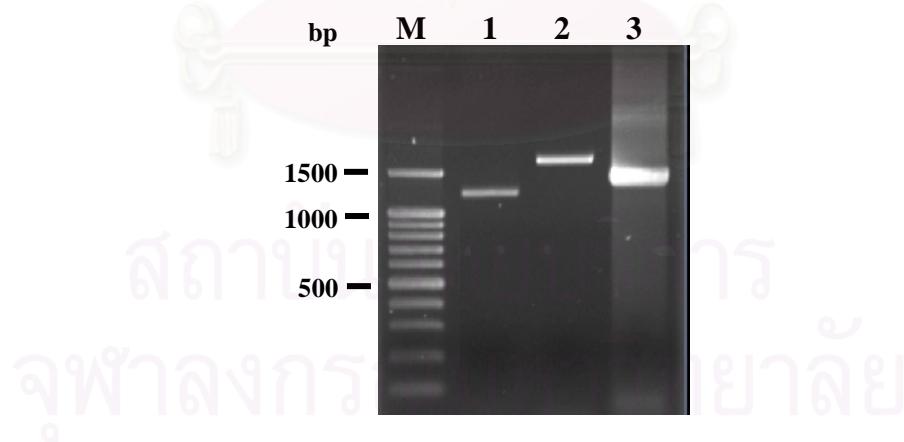
Figure 3.80 ORF amplification of interesting gene including PmCRT (lane 1), PmCNX (lane 2) and PmERp57 (lane 3). A 100 bp DNA ladder (lane M) was used as the marker.

PmCRT were expressed in two forms, providing full length cDNA and its cDNA without signal peptide (mature protein). The amplification products of further and latter forms were 1221 bp and 1173 bp in length. The nucleic acid sequence for mature protein of PmCNX that cut both signal peptide and transmembrane domain was 1359 bp in length. While PCR product of mature PmERp57 protein was 1410 bp (Figure 3.82). All of nucleotide sequence and deduced amino acid were aligned to its ORF (Figure 3.83-3.88). Two forms of PmCRT and mature PmCNX amplification products were digested with *Bam* HI and *Eco* RI, eluted and ligated into pGEX 4T-1 expression vector. The amplification product of mature PmERp57 product was digested with *Nde* I and *Bam* HI before eluttion and ligation with pET 15-b expression vector. After that four ligated expression vectors were transformed into *E. coli* JM109 and re-sequenced to confirm the orientation and nucleotide sequence. The last, they were transformed into *E. coli* BL21 (DE3) codon+ RIPL.
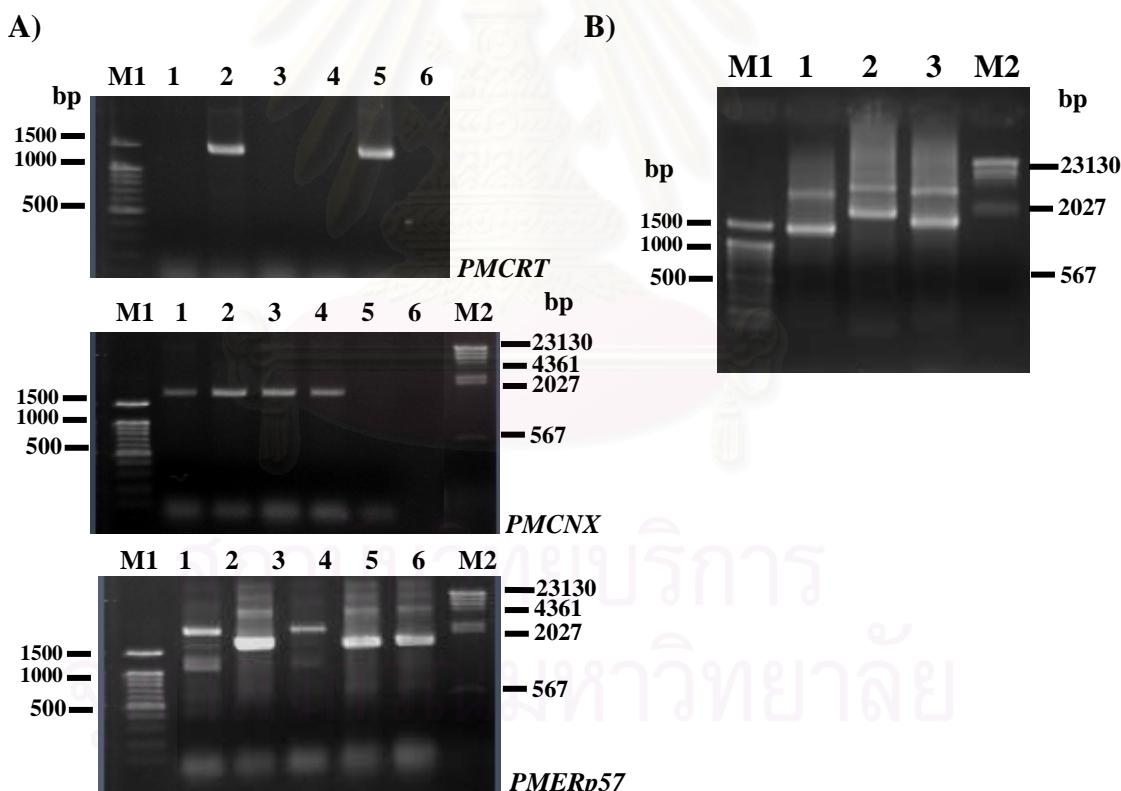
Figure 3.81 Colony PCR for determining sizes of positive clone of interesting ORF (A). The orientation of PmCRT (1), PmCNX (2) and PmERp57 (3) that insert into pGEM-T easy vector were checked by colony PCR (B). The 100 bp (M1) and λ *Hind* III (M2) markers were used as molecular ladder.

Figure 3.82 Amplification product of interesting genes consisting *PmCRT* (full length and mature cDNA are in lane 1 and 2, respectively), *PmCNX* (lane 3) and *PmERp57* (lane 4) that were inserted into expression vector. A 100 bp DNA ladder (lane M) was used as the marker.

```
ORF-PMCRT      ATGAAGACCTGGGTTTTTCTTGCCCTATTTGGGGTTGCCCTAGTGGAATCTAAAGTATTT 60
full-PMCRT     ATGAAGACCTGGGTTTTTCTTGCCCTATTTGGGGTTGCCCTAGTGGAATCTAAAGTATTT 60
maturePMCRT    -----------------------------------------------ATGAAAGTATTT 12
                                                               ********

ORF-PMCRT      TTCGAAGAAAGATTCGACAGCCCTGATTGGGAGAAAAATTGGGTTCAGTCTGCACACAAG 120
full-PMCRT     TTCGAAGAAAGATTCGACAGCCCTGATTGGGAGAAAAATTGGGTTCAGTCTGCACACAAG 120
maturePMCRT    TTCGAAGAAAGATTCGACAGCCCTGATTGGGAGAAAAATTGGGTTCAGTCTGCACACAAG 72
               ************************************************************

ORF-PMCRT      GGGAAGGAGTTTGGACCCTTCAAGTTGACAGCTGGCAAATTTTATGGCGATGCTGAAAAG 180
full-PMCRT     GGGAAGGAGTTTGGACCCTTCAAGTTGACAGCTGGCAAATTTTATGGCGATGCTGAAAAG 180
maturePMCRT    GGGAAGGAGTTTGGACCCTTCAAGTTGACAGCTGGCAAATTTTATGGCGATGCTGAAAAG 132
               ************************************************************

ORF-PMCRT      GATAAGGGAATCCAGACTGGACAGGATGCCCGCTTTTATGGTCTTTCTACGAAGTTTGAG 240
full-PMCRT     GATAAGGGAATCCAGACTGGACAGGATGCCCGCTTTTATGGTCTTTCTACGAAGTTTGAG 240
maturePMCRT    GATAAGGGAATCCAGACTGGACAGGATGCCCGCTTTTATGGTCTTTCTACGAAGTTTGAG 192
               ************************************************************

ORF-PMCRT      CCCTTCAGTAATAAGGATTCCCCACTTGTCATCCAGTTTTCTGTAAAACATGAACAGAAC 300
full-PMCRT     CCCTTCAGTAATAAGGATTCCCCACTTGTCATCCAGTTTTCTGTAAAACATGAACAGAAC 300
maturePMCRT    CCCTTCAGTAATAAGGATTCCCCACTTGTCATCCAGTTTTCTGTAAAACATGAACAGAAC 252
               ************************************************************

ORF-PMCRT      ATTGACTGTGGTGGAGGATATCTGAAGGTCTTCGATTGCTCTTTAGACCAGAAAGACATG 360
full-PMCRT     ATTGACTGTGGTGGAGGATATCTGAAGGTCTTCGATTGCTCTTTAGACCAGAAAGACATG 360
maturePMCRT    ATTGACTGTGGTGGAGGATATCTGAAGGTCTTCGATTGCTCTTTAGACCAGAAAGACATG 312
               ************************************************************

ORF-PMCRT      CACGGAGAGTCGCCATACCTCATTATGTTTGGTCCTGATATCTGTGGCCCAGGCACCAAG 420
full-PMCRT     CACGGAGAGTCGCCATACCTCATTATGTTTGGTCCTGATATCTGTGGCCCAGGCACCAAG 420
maturePMCRT    CACGGAGAGTCGCCATACCTCATTATGTTTGGTCCTGATATCTGTGGCCCAGGCACCAAG 372
               ************************************************************

ORF-PMCRT      AAGGTTCATGTAATCTTCAATTACAAGGGTGAGAACCATCTGATCAAGAAGGAAATCCGT 480
full-PMCRT     AAGGTTCATGTAATCTTCAATTACAAGGGTGAGAACCATCTGATCAAGAAGGAAATCCGT 480
maturePMCRT    AAGGTTCATGTAATCTTCAATTACAAGGGTGAGAACCATCTGATCAAGAAGGAAATCCGT 432
               ************************************************************
```

```
ORF-PMCRT    TGCAAGGATGACGTATTTTCCCATCTGTATACCCTCATTGTCAATCCTGACAACACCTAC 540
full-PMCRT   TGCAAGGATGACGTATTTTCCCATCTGTATACCCTCATTGTCAATCCTGACAACACCTAC 540
maturePMCRT  TGCAAGGATGACGTATTTTCCCATCTGTATACCCTCATTGTCAATCCTGACAACACCTAC 492
             ************************************************************

ORF-PMCRT    GAAGTTCTTATTGACAATGAAAAAGCTCAGTCTGGTGAACTCGAGGAGGACTGGGACTTC 600
full-PMCRT   GAAGTTCTTATTGACAATGAAAAAGCTCAGTCTGGTGAACTCGAGGAGGACTGGGACTTC 600
maturePMCRT  GAAGTTCTTATTGACAATGAAAAAGCTCAGTCTGGTGAACTCGAGGAGGACTGGGACTTC 552
             ************************************************************

ORF-PMCRT    CTTCCACCCAAGAAGATCAAGGACCCAGAAGCCAAGAAGCCCGACGATTGGGATGACCGC 660
full-PMCRT   CTTCCACCCAAGAAGATCAAGGACCCAGAAGCCAAGAAGCCCGACGATTGGGATGACCGC 660
maturePMCRT  CTTCCACCCAAGAAGATCAAGGACCCAGAAGCCAAGAAGCCCGACGATTGGGATGACCGC 612
             ************************************************************

ORF-PMCRT    CCCACCATTGCTGATCCTGACGATACTAAGCCTGAAGATTGGGACCAACCTGAACACATT 720
full-PMCRT   CCCACCATTGCTGATCCTGACGATACTAAGCCTGAAGATTGGGACCAACCTGAACACATT 720
maturePMCRT  CCCACCATTGCTGATCCTGACGATACTAAGCCTGAAGATTGGGACCAACCTGAACACATT 672
             ************************************************************

ORF-PMCRT    CCTGATCCTGATGCCACCAAACCTGAGGACTGGGATGATGAAATGGATGGCGAGTGGGAA 780
full-PMCRT   CCTGATCCTGATGCCACCAAACCTGAGGACTGGGATGATGAAATGGATGGCGAGTGGGAA 780
maturePMCRT  CCTGATCCTGATGCCACCAAACCTGAGGACTGGGATGATGAAATGGATGGCGAGTGGGAA 732
             ************************************************************

ORF-PMCRT    CCACCCATGATTGACAATCCTGACTACAAGGGTGAATGGAAGCCTAAGCAGATTGATAAC 840
full-PMCRT   CCACCCATGATTGACAATCCTGACTACAAGGGTGAATGGAAGCCTAAGCAGATTGATAAC 840
maturePMCRT  CCACCCATGATTGACAATCCTGACTACAAGGGTGAATGGAAGCCTAAGCAGATTGATAAC 792
             ************************************************************

ORF-PMCRT    CCTGATTACAAGGGTCCATGGATTCACCCTGAAATTGACAACCCAGAATACACACCTGAC 900
full-PMCRT   CCTGATTACAAGGGTCCATGGATTCACCCTGAAATTGACAACCCAGAATACACACCTGAC 900
maturePMCRT  CCTGATTACAAGGGTCCATGGATTCACCCTGAAATTGACAACCCAGAATACACACCTGAC 852
             ************************************************************

ORF-PMCRT    CCAGAGATCTACAAGTATGATGAGGTCTGTGCTCTTGGTTTGGATCTTTGGCAGGTAAAA 960
full-PMCRT   CCAGAGATCTACAAGTATGATGAGGTCTGTGCTCTTGGTTTGGATCTTTGGCAGGTAAAA 960
maturePMCRT  CCAGAGATCTACAAGTATGATGAGGTCTGTGCTCTTGGTTTGGATCTTTGGCAGGTAAAA 912
             ************************************************************

ORF-PMCRT    TCTGGTACTATCTTTGACAACTTCCTCATCTCAAATGATCCTGAAGAAGCCCGCAAGATT 1020
full-PMCRT   TCTGGTACTATCTTTGACAACTTCCTCATCTCAAATGATCCTGAAGAAGCCCGCAAGATT 1020
maturePMCRT  TCTGGTACTATCTTTGACAACTTCCTCATCTCAAATGATCCTGAAGAAGCCCGCAAGATT 972
             ************************************************************

ORF-PMCRT    GGTGAAGAGACTTGGGGTGCTACTAAAGATGCAGCTAAGAAGATGAAGGATGAACAGGAT 1080
full-PMCRT   GGTGAAGAGACTTGGGGTGCTACTAAAGATGCAGCTAAGAAGATGAAGGATGAACAGGAT 1080
maturePMCRT  GGTGAAGAGACTTGGGGTGCTACTAAAGATGCAGCTAAGAAGATGAAGGATGAACAGGAT 1032
             ************************************************************

ORF-PMCRT    GAAGAGGAGCGAAAGAGAGCAGAGGAAGAAGCTAAGGCAGCTGCTGATGCTGAGAAGGAT 1140
full-PMCRT   GAAGAGGAGCGAAAGAGAGCAGAGGAAGAAGCTAAGGCAGCTGCTGATGCTGAGAAGGAT 1140
maturePMCRT  GAAGAGGAGCGAAAGAGAGCAGAGGAAGAAGCTAAGGCAGCTGCTGATGCTGAGAAGGAT 1092
             ************************************************************

ORF-PMCRT    GAGGACGATGATGACGACGACGATCTTGGCGATGAAGACGAAGATGATCTTGATAATGAT 1200
full-PMCRT   GAGGACGATGATGACGACGACGATCTTGGCGATGAAGACGAAGATGATCTTGATAATGAT 1200
maturePMCRT  GAGGACGATGATGACGACGACGATCTTGGCGATGAAGACGAAGATGATCTTGATAATGAT 1152
             ************************************************************

ORF-PMCRT    CTTGAACATGACGAGCTGTAA 1221
full-PMCRT   CTTGAACATGACGAGCTGTAA 1221
maturePMCRT  CTTGAACATGACGAGCTGTAA 1173
             *********************
```

Figure 3.83 Alignment between the full length cDNA of *PmCRT* from ORF amplification (*ORF-PmCRT*) and amplification fragments for *in vitro* expression, providing full-length ORF (*full-PmCRT*) and mature proteins with out signal peptide (*maturePmCRT*).

```
ORF-PMCRT     MKTWVFLALFGVALVESKVFFEERFDSPDWEKNWVQSAHKGKEFGPFKLTAGKFYGDAEK 60
full-PMCRT    MKTWVFLALFGVALVESKVFFEERFDSPDWEKNWVQSAHKGKEFGPFKLTAGKFYGDAEK 60
maturePMCRT   ---------------MKVFFEERFDSPDWEKNWVQSAHKGKEFGPFKLTAGKFYGDAEK 44
                             *******************************************

ORF-PMCRT     DKGIQTGQDARFYGLSTKFEPFSNKDSPLVIQFSVKHEQNIDCGGGYLKVFDCSLDQKDM 120
full-PMCRT    DKGIQTGQDARFYGLSTKFEPFSNKDSPLVIQFSVKHEQNIDCGGGYLKVFDCSLDQKDM 120
maturePMCRT   DKGIQTGQDARFYGLSTKFEPFSNKDSPLVIQFSVKHEQNIDCGGGYLKVFDCSLDQKDM 104
              ************************************************************

ORF-PMCRT     HGESPYLIMFGPDICGPGTKKVHVIFNYKGENHLIKKEIRCKDDVFSHLYTLIVNPDNTY 180
full-PMCRT    HGESPYLIMFGPDICGPGTKKVHVIFNYKGENHLIKKEIRCKDDVFSHLYTLIVNPDNTY 180
maturePMCRT   HGESPYLIMFGPDICGPGTKKVHVIFNYKGENHLIKKEIRCKDDVFSHLYTLIVNPDNTY 164
              ************************************************************

ORF-PMCRT     EVLIDNEKAQSGELEEDWDFLPPKKIKDPEAKKPDDWDDRPTIADPDDTKPEDWDQPEHI 240
full-PMCRT    EVLIDNEKAQSGELEEDWDFLPPKKIKDPEAKKPDDWDDRPTIADPDDTKPEDWDQPEHI 240
maturePMCRT   EVLIDNEKAQSGELEEDWDFLPPKKIKDPEAKKPDDWDDRPTIADPDDTKPEDWDQPEHI 224
              ************************************************************

ORF-PMCRT     PDPDATKPEDWDDEMDGEWEPPMIDNPDYKGEWKPKQIDNPDYKGPWIHPEIDNPEYTPD 300
full-PMCRT    PDPDATKPEDWDDEMDGEWEPPMIDNPDYKGEWKPKQIDNPDYKGPWIHPEIDNPEYTPD 300
maturePMCRT   PDPDATKPEDWDDEMDGEWEPPMIDNPDYKGEWKPKQIDNPDYKGPWIHPEIDNPEYTPD 284
              ************************************************************

ORF-PMCRT     PEIYKYDEVCALGLDLWQVKSGTIFDNFLISNDPEEARKIGEETWGATKDAAKKMKDEQD 360
full-PMCRT    PEIYKYDEVCALGLDLWQVKSGTIFDNFLISNDPEEARKIGEETWGATKDAAKKMKDEQD 360
maturePMCRT   PEIYKYDEVCALGLDLWQVKSGTIFDNFLISNDPEEARKIGEETWGATKDAAKKMKDEQD 344
              ************************************************************
```

Figure 3.84 Alignment of deduced amino acid sequences between of *PmCRT* from ORF amplification (*ORF-PmCRT*) and amplification fragments for recombinant protein expression, providing full-length ORF (*full-PmCRT*) and mature proteins with out signal peptide (*maturePmCRT*).

```
ORF-PMCNX     ATGAAGTCGAGGTGGCAGAGAAAAGCAGTATTAGCACTGCTAGTTCTTGGCCTGCTGTTA 60
maturePMCNX   ------------------------------------------------------------
ORF-PMCNX     CCTTTTTGGTATTAAAGCAGATGACGATGACGATGAAGAAGCTGTTGTCACGGAAGAGCAA 120
maturePMCNX   --------------ATGGATGACGATGACGATGAAGAAGCTGTTGTCACGGAAGAGCAA 45
                            *****************************************

ORF-PMCNX     ACAGAAGGAGAGGAAGATGATGTAGAGGAGGTTGTGTATGCAACACCCAAGGCACTACCA 180
maturePMCNX   ACAGAAGGAGAGGAAGATGATGTAGAGGAGGTTGTGTATGCAACACCCAAGGCACTACCA 105
              ************************************************************

ORF-PMCNX     AATGCATATCTGACTGAAACGTTTGATGACATAGCTACTTTTGAGAAGACGTGGATCAAA 240
maturePMCNX   AATGCATATCTGACTGAAACGTTTGATGACATAGCTACTTTTGAGAAGACGTGGATCAAA 165
              ************************************************************

ORF-PMCNX     TCTGAAGCCAAGAAGGACGGTGTTGATGAAAACATTGCTAAATATGATGGTGTTTGGGCA 300
maturePMCNX   TCTGAAGCCAAGAAGGACGGTGTTGATGAAAACATTGCTAAATATGATGGTGTTTGGGCA 225
              ************************************************************

ORF-PMCNX     GTAGAACCTGCTGAACGATTGGCTCTTACTGGGGATCGTGGGTTGGTGTTGAAGTCAAAG 360
maturePMCNX   GTAGAACCTGCTGAACGATTGGCTCTTACTGGGGATCGTGGGTTGGTGTTGAAGTCAAAG 285
              ************************************************************

ORF-PMCNX     GCAAAACATGCAGCAATTGCAGCACCACTGAAGAAACCATTTGTATTCAGTAACAAGCCT 420
maturePMCNX   GCAAAACATGCAGCAATTGCAGCACCACTGAAGAAACCATTTGTATTCAGTAACAAGCCT 345
              ************************************************************

ORF-PMCNX     TTTGTGGTTCAGTATGAAGTTAATTTACAGAATGGTCAAGAATGTGGTGGAGCATATATC 480
maturePMCNX   TTTGTGGTTCAGTATGAAGTTAATTTACAGAATGGTCAAGAATGTGGTGGAGCATATATC 405
              ************************************************************

ORF-PMCNX     AAACTAATCAGTGCCCAAAAAGGACGTGTGGATCTTAAAAATTTCCATGACAAAACACCG 540
maturePMCNX   AAACTAATCAGTGCCCAAAAAGGACGTGTGGATCTTAAAAATTTCCATGACAAAACACCG 465
              ************************************************************

ORF-PMCNX     TACACTATTATGTTTGGACCAGACAAATGTGGCAATGACTTCAAGTTGCATTTCATCTTC 600
maturePMCNX   TACACTATTATGTTTGGACCAGACAAATGTGGCAATGACTTCAAGTTGCATTTCATCTTC 525
              ************************************************************
```

```
ORF-PMCNX      AGGCATGTTAATCCTCTTACTGAAGAAATTGAAGAAAAACATGCTAAGAGACCACGTGAC 660
maturePMCNX    AGGCATGTTAATCCTCTTACTGAAGAAATTGAAGAAAAACATGCTAAGAGACCACGTGAC 585
               ************************************************************

ORF-PMCNX      AAGATTGAGGAACCATTTAAGGACAAGAAGTCTCATTTGTACACATTAGTAATTCGACCA 720
maturePMCNX    AAGATTGAGGAACCATTTAAGGACAAGAAGTCTCATTTGTACACATTAGTAATTCGACCA 645
               ************************************************************


ORF-PMCNX      GACAACACCTTTGAAATAAGCCTGGATCACGAGGTAATCAATTCAGGAAGCCTTCTGGAG 780
maturePMCNX    GACAACACCTTTGAAATAAGCCTGGATCACGAGGTAATCAATTCAGGAAGCCTTCTGGAG 705
               ************************************************************

ORF-PMCNX      GACTTCACCCCATCTGTCAACCCTCCCAAAGAAATCGATGATCCTGAAGACTTTATGCCA 840
maturePMCNX    GACTTCACCCCATCTGTCAACCCTCCCAAAGAAATCGATGATCCTGAAGACTTTATGCCA 765
               ************************************************************

ORF-PMCNX      GAAGACTGGGATGAAAGAGAAAAGATTCCAGATCCAGAAGCCACAAAGCCTGATGACTGG 900
maturePMCNX    GAAGACTGGGATGAAAGAGAAAAGATTCCAGATCCAGAAGCCACAAAGCCTGATGACTGG 825
               ************************************************************

ORF-PMCNX      GATGAGGATGCTCCCATGCAGATCCCTGATCCAGTGGCTGAGAAACCTAGTGGATGGCTG 960
maturePMCNX    GATGAGGATGCTCCCATGCAGATCCCTGATCCAGTGGCTGAGAAACCTAGTGGATGGCTG 885
               ************************************************************

ORF-PMCNX      GATGATGAGCCAGAAATGGTGCCAGATCCCACTGCTGAGAAACCTGATGACTGGGATGAT 1020
maturePMCNX    GATGATGAGCCAGAAATGGTGCCAGATCCCACTGCTGAGAAACCTGATGACTGGGATGAT 945
               ************************************************************

ORF-PMCNX      GAAATGGATGGTGAGTGGGAAGCTCCACTGATCACCAACCCCAAGTGTGTTGATGCACCA 1080
maturePMCNX    GAAATGGATGGTGAGTGGGAAGCTCCACTGATCACCAACCCCAAGTGTGTTGATGCACCA 1005
               ************************************************************

ORF-PMCNX      GGCTGTGGAGAGTGGAAGCCTCCCATGGTGGACAATCCTGAGTTCAAGGGCAAATGGCGT 1140
maturePMCNX    GGCTGTGGAGAGTGGAAGCCTCCCATGGTGGACAATCCTGAGTTCAAGGGCAAATGGCGT 1065
               ************************************************************

ORF-PMCNX      CCACCCATGATTGACAATCCTAATTACCGTGGAAAATGGAAGCCACGAAAGATCCCCAAC 1200
maturePMCNX    CCACCCATGATTGACAATCCTAATTACCGTGGAAAATGGAAGCCACGAAAGATCCCCAAC 1125
               ************************************************************

ORF-PMCNX      CCTGACTTCTTTGAAGACCTGGAACCTTTCAAGATGACTGCTATTGATGCTGTTGGTCTG 1260
maturePMCNX    CCTGACTTCTTTGAAGACCTGGAACCTTTCAAGATGACTGCTATTGATGCTGTTGGTCTG 1185
               ************************************************************

ORF-PMCNX      GAATTGTGGTCAATGTCAGACAATATCCTCTTTGACAACATACTTGTCACAGACAATGTT 1320
maturePMCNX    GAATTGTGGTCAATGTCAGACAATATCCTCTTTGACAACATACTTGTCACAGACAATGTT 1245
               ************************************************************

ORF-PMCNX      GCTGAGGCTTACCAGTTTGCTCAAGAAACTTTTGACTTGAAGGTCATGAAGATAGAGAAG 1380
maturePMCNX    GCTGAGGCTTACCAGTTTGCTCAAGAAACTTTTGACTTGAAGGTCATGAAGATAGAGAAG 1305
               ************************************************************

ORF-PMCNX      GGACAGACTGGAGTTATTCGACGAATCATCAACTACTCCAATAAGAATCCATGGCTATAT 1440
maturePMCNX    GGACAGACTGGAGTTATTCGACGAATCATCAACTACTCCAATAAGAATCCATAA------ 1359
               **************************************************    

ORF-PMCNX      GGTATCTACGTACTCCTAGTGGCCATTCCTGTAGTGTTGATATTTGCCTGTTGTTGTGCA 1500
maturePMCNX    ------------------------------------------------------------
ORF-PMCNX      GAAGCAAAGGACACCAAGGAAGATAAGGATGCAGAGAGGAAAAAGACCGATGCACCTTCT 1560
maturePMCNX    ------------------------------------------------------------
ORF-PMCNX      CCTGATGATCCACAATCAGAGCAAAACAAAGAAGATTCTAGTCAAACTGCAGCAGATGAT 1620
maturePMCNX    ------------------------------------------------------------
ORF-PMCNX      GATGCTCCTGGAAGTGGGGATGAAGCTGAAGGGGATGAAGACAAAGGTGATGAAGAGGGA 1680
maturePMCNX    ------------------------------------------------------------
ORF-PMCNX      GAGGAGGAAGAAGATGAGGAAGAGGAGGAGGAAGCAGAGAAAGCTGATGCAGCTGAGGAG 1740
maturePMCNX    ------------------------------------------------------------
ORF-PMCNX      TCCAGACTCGCACATCACCAAGGCTGCGCAAGTCCCGGAGAGATTAA 1788
maturePMCNX    -----------------------------------------------
```

Figure 3.85 Alignment between the full length cDNA of *PmCNX* from ORF amplification (*ORF-PmCNX*) and mature proteins amplification with out signal peptide (*maturePMCmX*) for *in vitro* expression.

```
ORF-PMCNX     MKSRWQRKAVLALLVLGLLLPFGIKADDDDDEEAVVTEEQTEGEEDDVEEVVYATPKALP 60
maturePMCNX   -----------------------MDDDDDEEAVVTEEQTEGEEDDVEEVVYATPKALP 35
                                     *********************************

ORF-PMCNX     NAYLTETFDDIATFEKTWIKSEAKKGGVDENIAKYDGVWAVEPAERLALTGDRGLVLKSK 120
maturePMCNX   NAYLTETFDDIATFEKTWIKSEAKKGGVDENIAKYDGVWAVEPAERLALTGDRGLVLKSK 95
              ***********************************************************

ORF-PMCNX     AKHAAIAAPLKKPFVFSNKPFVVQYEVNLQNGQECGGAYIKLISAQKGRVDLKNFHDKTP 180
maturePMCNX   AKHAAIAAPLKKPFVFSNKPFVVQYEVNLQNGQECGGAYIKLISAQKGRVDLKNFHDKTP 155
              ***********************************************************

ORF-PMCNX     YTIMFGPDKCGNDFKLHFIFRHVNPLTEEIEEKHAKRPRDKIEEPFKDKKSHLYTLVIRP 240
maturePMCNX   YTIMFGPDKCGNDFKLHFIFRHVNPLTEEIEEKHAKRPRDKIEEPFKDKKSHLYTLVIRP 215
              ***********************************************************

ORF-PMCNX     DNTFEISLDHEVINSGSLLEDFTPSVNPPKEIDDPEDFMPEDWDEREKIPDPEAAKPDDW 300
maturePMCNX   DNTFEISLDHEVINSGSLLEDFTPSVNPPKEIDDPEDFMPEDWDEREKIPDPEAAKPDDW 275
              ***********************************************************

ORF-PMCNX     DEDAPMQIPDPVAEKPSGWLDDEPEMVPDPTAEKPDDWDDEMDGEWEAPLITNPKCVDAP 360
maturePMCNX   DEDAPMQIPDPVAEKPSGWLDDEPEMVPDPTAEKPDDWDDEMDGEWEAPLITNPKCVDAP 335
              ***********************************************************

ORF-PMCNX     GCGEWKPPMVDNPEFKGKWRPPMIDNPNYRGKWKPRKIPNPDFFEDLEPFKMTAIDAVGL 420
maturePMCNX   GCGEWKPPMVDNPEFKGKWRPPMIDNPNYRGKWKPRKIPNPDFFEDLEPFKMTAIDAVGL 395
              ***********************************************************

ORF-PMCNX     ELWSMSDNILFDNILVTDNVAEAYQFAQETFDLKVMKIEKGQTGVIRRIINHSNKNPWLY 480
maturePMCNX   ELWSMSDNILFDNILVTDNVAEAYQFAQETFDLKVMKIEKGQTGVIRRIINHSNKNP--- 452
              ********************************************************

ORF-PMCNX     GIYVLLVAIPVVLIFACCCAEAKDTKEDKDAERKKTDAPSPDDPQSEQNKEDSSQTAADD 540
maturePMCNX   ------------------------------------------------------------
ORF-PMCNX     DAPGSGDEAEGDEDKGDEEGEEEEDEEEEEEAEKADAAEEVQTRTSPRLRKSRRD 595
maturePMCNX   ------------------------------------------------------
```

Figure 3.86 Alignment of deduced amino acid sequences between of *mCNX* from ORF amplification (*ORF-PmCNX*) and mature proteins with out signal peptide (*maturePmCRT*) for recombinant protein expression.

```
ORF-PMERp57     ATGGCTACGAGATTGTTAATACTACTCCTCTCCCTCGTGGCCGTGGCGCGGAGACGATGT 60
mature-PMERp57  --------------------------------------------ATG-GGAGACGATGT 14
                                                            * **********

ORF-PMERp57     CCTGCAATTAAACGACGCGGATTTCGACGGGAAAGTGGCCAGCTACGACACGGTCCTCGT 120
mature-PMERp57  CCTGCAATTAAACGACGCGGATTTCGACGGGAAAGTGGCCAGCTACGACACGGTCCTCGT 74
                ************************************************************

ORF-PMERp57     CATGTTCTACGCCCCATGGTGTGGTCACTGCAAGAGATTAAAGCCTGAGTTTGAGAAGGC 180
mature-PMERp57  CATGTTCTACGCCCCATGGTGTGGTCACTGCAAGAGATTAAAGCCTGAGTTTGAGAAGGC 134
                ************************************************************

ORF-PMERp57     CTCTACCACCTTGAAGGCCAACGACCCTCCCGTCTACCTTGCTAAGGTGGATTGTACTGA 240
mature-PMERp57  CTCTACCACCTTGAAGGCCAACGACCCTCCCGTCTACCTTGCTAAGGTGGATTGTACTGA 194
                ************************************************************

ORF-PMERp57     TGATGGAAAGGACAGCTGTAGCAGATTTGGTGTGTCTGGCTACCCTACCCTGAAGATCTT 300
mature-PMERp57  TGATGGAAAGGACAGCTGTAGCAGATTTGGTGTGTCTGGCTACCCTACCCTGAAGATCTT 254
                ************************************************************

ORF-PMERp57     CAAGGGAGGAGAGCTCTCTACGGACTACAATGGTCCACGAGATGCCAGTGGTATTGTAAA 360
mature-PMERp57  CAAGGGAGGAGAGCTCTCTACGGACTACAATGGTCCACGAGATGCCAGTGGTATTGTAAA 314
                ************************************************************

ORF-PMERp57     GTACATGAGGTCACAGGTTGGACCAGCCTCTAAGGAGTTGACATCCGTGGAGGCAGCAGA 420
mature-PMERp57  GTACATGAGGTCACAGGTTGGACCAGCCTCTAAGGAGTTGACATCCGTGGAGGCAGCAGA 374
                ************************************************************

ORF-PMERp57     AGCATTCCTTGGTGCTGCTGAAGTTGGAGTCGTTTACTTTGGAGGAGATTCCAAACTTAA 480
mature-PMERp57  AGCATTCCTTGGTGCTGCTGAAGTTGGAGTCGTTTACTTTGGAGGAGATTCCAAACTTAA 434
                ************************************************************

ORF-PMERp57     AGATGCTTTCCTAAAGGCTGCTGATAAGCTGAGGGAATCCATCCGTTTTGCACACTCCCT 540
mature-PMERp57  AGATGCTTTCCTAAAGGCTGCTGATAAGCTGAGGGAATCCATCCGTTTTGCACACTCCCT 494
                ************************************************************

ORF-PMERp57     CGATGCCACTGTTAATGAAAGTATGGGTACAGCGATGTTGTTGTACTTTTCCGACCGAA 600
mature-PMERp57  CGATGCCACTGTTAATGAAAGTATGGGTACAGCGATGTTGTTGTACTTTTCCGACCGAA 554
                ************************************************************

ORF-PMERp57     ACACCTGGAGAACAAATTTGAGCCTTCCTCTGTTGTATTTGAGGGATCGGCAGACAGGGC 660
mature-PMERp57  ACACCTGGAGAACAAATTTGAGCCTTCCTCTGTTGTATTTGAGGGATCGGCAGACAGGGC 614
```

```
                    ************************************************************
ORF-PMERp57        TGAGATTGAGTCTTTCATCAAAAAGAACTTCCATGGTTTGGTAGGACACCTAACGCAAGA 720
mature-PMERp57     TGAGATTGAGTCTTTCATCAAAAAGAACTTCCATGGTTTGGTAGGACACCTAACGCAAGA 674
                    ************************************************************
ORF-PMERp57        CACTGCTCAGGATTTCAAACCTCCAGTTGTGATTGCTTACTACAATGTTGATTACATCAA 780
mature-PMERp57     CACTGCTCAGGATTTCAAACCTCCAGTTGTGATTGCTTACTACAATGTTGATTACATCAA 734
                    ************************************************************
ORF-PMERp57        AAATGTTAAGGGTACAAATTACTGGCGCAATCGTGTCCTTAAGGTGGCACAAAACTTTGC 840
mature-PMERp57     AAATGTTAAGGGTACAAATTACTGGCGCAATCGTGTCCTTAAGGTGGCACAAAACTTTGC 794
                    ************************************************************
ORF-PMERp57        TGATGACTTCAAGTTTGCCGTTGCCAATAAGGACGACTTCCAGCATGACCTCAATGAATA 900
mature-PMERp57     TGATGACTTCAAGTTTGCCGTTGCCAATAAGGACGACTTCCAGCATGACCTCAATGAATA 854
                    ************************************************************
ORF-PMERp57        TGGCCTTGATTATGTTCCTGGTGACAAGCCAGTAATTTGTGCACGTAATGCTAAAGCCCA 960
mature-PMERp57     TGGCCTTGATTATGTTCCTGGTGACAAGCCAGTAATTTGTGCACGTAATGCTAAAGCCCA 914
                    ************************************************************
ORF-PMERp57        GAAGTTTGTCATGCAGGAAGAATTTTCAATGGATAACCTCCAAGCATTCCTCACCAATCT 1020
mature-PMERp57     GAAGTTTGTCATGCAGGAAGAATTTTCAATGGATAACCTCCAAGCATTCCTCACCAATCT 974
                    ************************************************************
ORF-PMERp57        CAAGGCGGGTGAGCTTGAGCCATATCTGAAGTCTGAGGCAGTGCCAACACAAGATGGCCC 1080
mature-PMERp57     CAAGGCGGGTGAGCTTGAGCCATATCTGAAGTCTGAGGCAGTGCCAACACAAGATGGCCC 1034
                    ************************************************************
ORF-PMERp57        TGTCACTGTTGCTGTGGGTAAGAACTTCAATGAAGTTGTCTCTGATGAGCGTGATGCCCT 1140
mature-PMERp57     TGTCACTGTTGCTGTGGGTAAGAACTTCAATGAAGTTGTCTCTGATGAGCGTGATGCCCT 1094
                    ************************************************************
ORF-PMERp57        CATTGAATTCTATGCTCCTTGGTGTGGTCACTGCAAGAAATTAGCGCCCACCTATGATGA 1200
mature-PMERp57     CATTGAATTCTATGCTCCTTGGTGTGGTCACTGCAAGAAATTAGCGCCCACCTATGATGA 1154
                    ************************************************************
ORF-PMERp57        GCTGGGAGAAGCGATGAAGGATGAAGATGTAGACATTGTGAAGATGGATGCCACTGCCAA 1260
mature-PMERp57     GCTGGGAGAAGCGATGAAGGATGAAGATGTAGACATTGTGAAGATGGATGCCACTGCCAA 1214
                    ************************************************************
ORF-PMERp57        TGATGTTCCTCCTCAGTACAATGTTCAAGGCTTCCCCACCATCTTCTGGAAACCCAAGGG 1320
mature-PMERp57     TGATGTTCCTCCTCAGTACAATGTTCAAGGCTTCCCCACCATCTTCTGGAAACCCAAGGG 1274
                    ************************************************************
ORF-PMERp57        TGGTGTTCCAAGGAATTACAATGGTGGCCGGGAACTGGACGATTTTGTCAAGTACATTGC 1380
mature-PMERp57     TGGTGTTCCAAGGAATTACAATGGTGGCCGGGAACTGGACGATTTTGTCAAGTACATTGC 1334
                    ************************************************************
ORF-PMERp57        CCGACATTCCACAAATGAACTGAATGGGTATGACCGCAAGGGGAAGGCAAAGAAAGGCAA 1440
mature-PMERp57     CCGACATTCCACAAATGAACTGAATGGGTATGACCGCAAGGGGAAGGCAAAGAAAGGCAA 1394
                    ************************************************************
ORF-PMERp57        GAAGACTGAACTTTGA 1456
mature-PMERp57     GAAGACTGAACTTTGA 1410
                    ***************
```

Figure 3.87 Alignment between the full length cDNA of *PmERp57* from ORF amplification (*ORF-PmERp57*) and mature proteins amplification with out signal peptide (*maturePmERp57*) for *in vitro* expression.

```
ORF-PMERp57      MATRLLILLLSLVAVALGDDVLQLNDADFDGKVASYDTVLVMFYAPWCGHCKRLKPEFEK 60
mature-PMERp57   ---------------MGDDVLQLNDADFDGKVASYDTVLVMFYAPWCGHCKRLKPEFEK 44
                                :*******************************************
ORF-PMERp57      ASTTLKANDPPVYLAKVDCTDDGKDSCSRFGVSGYPTLKIFKGGELSTDYNGPRDASGIV 120
mature-PMERp57   ASTTLKANDPPVYLAKVDCTDDGKDSCSRFGVSGYPTLKIFKGGELSTDYNGPRDASGIV 104
                 ***********************************************************
ORF-PMERp57      KYMRSQVGPASKELTSVEAAEAFLGAAEVGVVYFGGDSKLKDAFLKAADKLRESIRFAHS 180
mature-PMERp57   KYMRSQVGPASKELTSVEAAEAFLGAAEVGVVYFGGDSKLKDAFLKAADKLRESIRFAHS 164
                 ***********************************************************
ORF-PMERp57      LDATVNEKYGYSDVVVLFRPKHLENKFEPSSVVFEGSADRAEIESFIKKNFHGLVGHLTQ 240
mature-PMERp57   LDATVNEKYGYSDVVVLFRPKHLENKFEPSSVVFEGSADRAEIESFIKKNFHGLVGHLTQ 224
                 ***********************************************************
ORF-PMERp57      DTAQDFKPPVVIAYYNVDYIKNVKGTNYWRNRVLKVAQNFADDFKFAVANKDDFQHDLNE 300
mature-PMERp57   DTAQDFKPPVVIAYYNVDYIKNVKGTNYWRNRVLKVAQNFADDFKFAVANKDDFQHDLNE 284
                 ***********************************************************
ORF-PMERp57      YGLDYVPGDKPVICARNAKAQKFVMQEEFSMDNLQAFLTNLKAGELEPYLKSEAVPTQDG 360
mature-PMERp57   YGLDYVPGDKPVICARNAKAQKFVMQEEFSMDNLQAFLTNLKAGELEPYLKSEAVPTQDG 344
                 ***********************************************************
ORF-PMERp57      PVTVAVGKNFNEVVSDERDALIEFYAPWCGHCKKLAPTYDELGEAMKDEDVDIVKMDATA 420
mature-PMERp57   PVTVAVGKNFNEVVSDERDALIEFYAPWCGHCKKLAPTYDELGEAMKDEDVDIVKMDATA 404
                 ***********************************************************
ORF-PMERp57      NDVPPQYNVQGFPTIFWKPKGGVPRNYNGGRELDDFVKYIARHSTNELNGYDRKGKAKKG 480
mature-PMERp57   NDVPPQYNVQGFPTIFWKPKGGVPRNYNGGRELDDFVKYIARHSTNELNGYDRKGKAKKG 464
                 ***********************************************************
ORF-PMERp57      KKTEL 485
mature-PMERp57   KKTEL 469
                 *****
```

Figure 3.88 Alignment of deduced amino acid sequences between of *PmERp57* from ORF amplification (*ORF-PmERp57*) and mature proteins with out signal peptide (*maturePmERp57*) for recombinant protein expression.

3.9.2 Production of recombinant proteins

3.9.2.1 Time course study of the protein expression levels

Expression levels of the three recombinant proteins, i.e. mature PmCRT (71.35 kDa), PmCNX (77.95 kDa) and PmERp57 (52.31 kDa), were followed in time course to find out the optimal induction time yielding maximum amounts of the expressed proteins.

As shown in Figures 3.89-3.91, all of the recombinant proteins were expressed within the first hour. Their expression levels increased in line with the induction time. However, they are maximized at the hours of 3-4 and the proteins seemed not to be affected by proteolysis even when the induction time was extended to 14 hr.

Western blotting was also carried out to confirm the recombinant expression of target proteins. All of them were recognized by the antibodies specific to the attached sequence tag (Figures 3.89b, 3.90b and 3.91b).
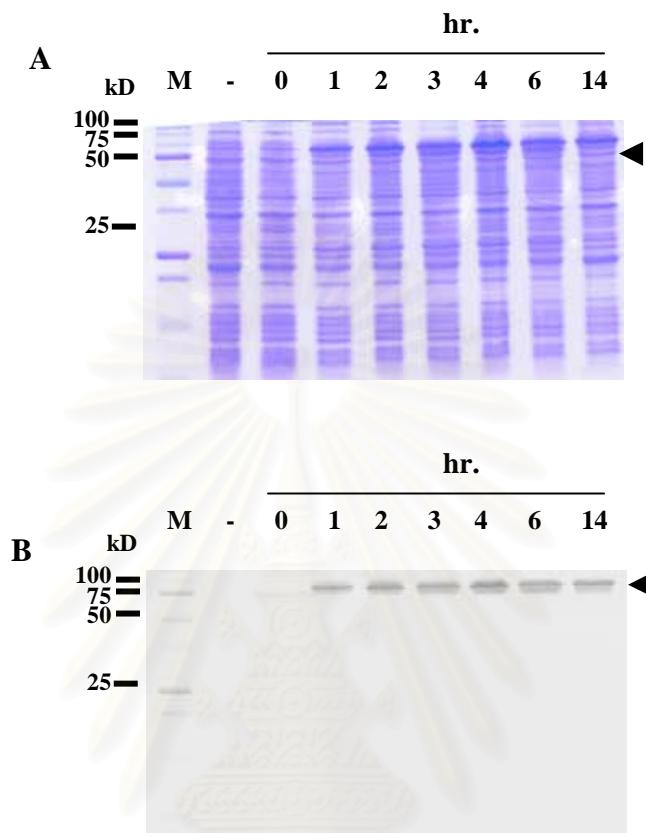
Figure 3.89 Time course study of recombinant expression of mature PmCRT in *E. coli* BL21 (DE3) codon+ RIPL induced for 0, 1, 2, 3, 4, 6 and 14 hr analyzed by SDS-PAGE (A) and western blotting (B). Loaded samples were whole harvested cells taken from a grown culture after the specified induction times. Lanes were specified according to the induction times in hours. Lane "-" is a negative control derived from the host cell containing pGEX 4T-1 vector without insert.
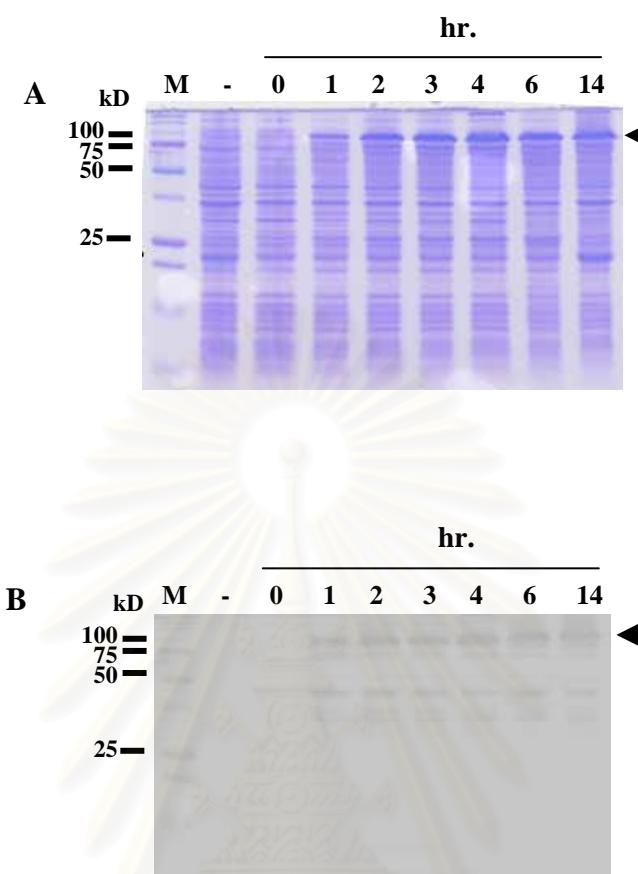
Figure 3.90 Time course study of recombinant expression of mature PmCNX in *E. coli* BL21 (DE3) codon+ RIPL induced for 0, 1, 2, 3, 4, 6 and 14 hr analyzed by SDS-PAGE (A) and western blotting (B). Loaded samples were whole harvested cells taken from a grown culture after the specified induction times. Lanes were specified according to the induction times in hours. Lane "-" is a negative control derived from the host cell containing pGEX 4T-1 vector without insert.
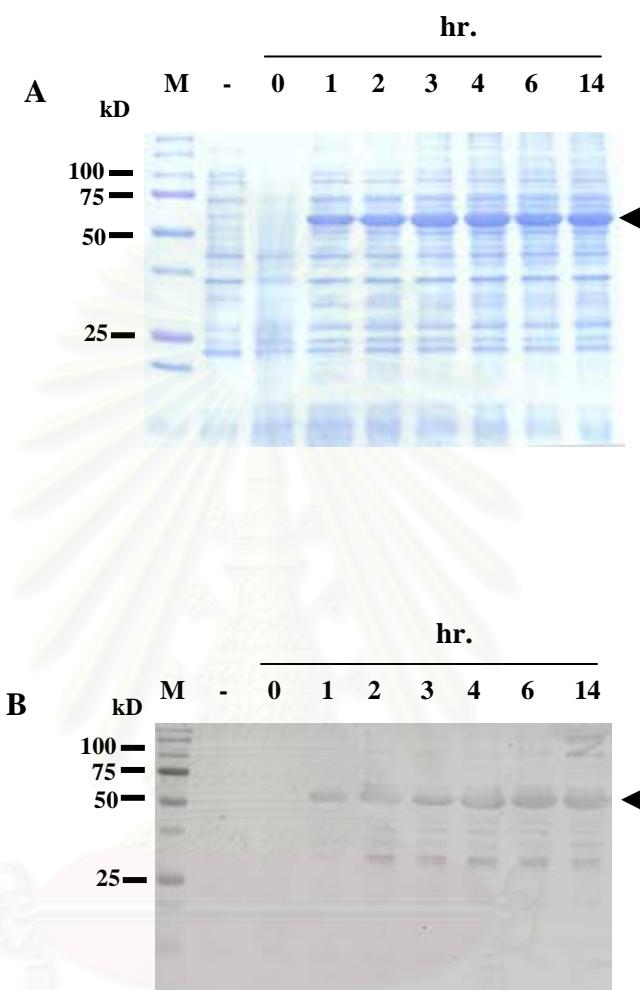
Figure 3.91 Time course study of recombinant expression of mature PmERp57 in *E. coli* BL21 (DE3) codon+ RIPL induced for 0, 1, 2, 3, 4, 6 and 14 hr analyzed by SDS-PAGE (A) and western blotting (B). Loaded samples were whole harvested cells taken from a grown culture after the specified induction times. Lanes were specified according to the induction times in hours. Lane "-" is a negative control derived from the host cell containing pET15b vector without insert.

3.9.2.2 Optimization of induction temperature

After cell disruption, soluble and insoluble fractions were separated to identify the forms of expressed proteins in the host cell. At the 37 ºC induction temperature, recombinant PmCRT and PmCNX were expressed only in soluble form (Figure 3.92). In contrast, some of the expressed PmERp57 molecules existed in inclusion body (insoluble) form in the host cell when cells were induced at 37 ºC. Induction of ERp57 expression at 20 ºC was therefore tried to reduce the protein production rate in the cell, allowing more time for the protein to fold and become soluble. As seen in Figure 3.93, lowering the induction temperature to 20 ºC resulted in the lower protein expression level. Moreover, the proportions between the soluble and insoluble PmERp57 amounts produced at the two temperatures were comparable. Therefore, induction at 37 ºC resulted in a higher total amount of the target protein.
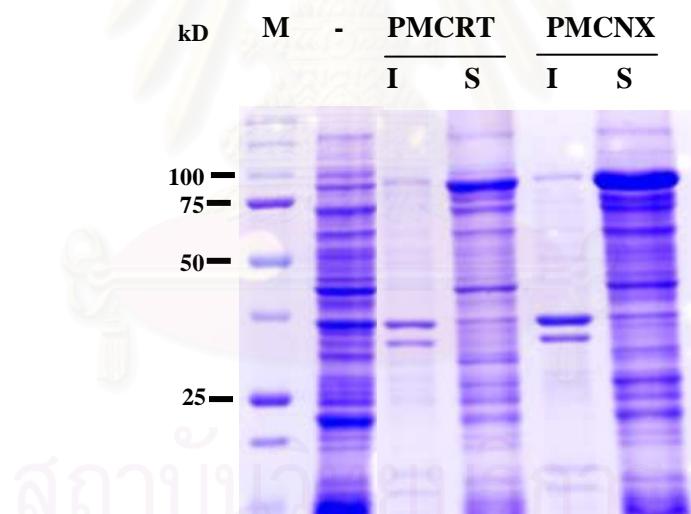
Figure 3.92 SDS-PAGE analyses of cell fractionations from recombinant expression of PmCRT and PmCNX. Cultured cells were induced by IPTG for 3 hr at 37ºC. Cells were then harvested and disrupted. Soluble and insoluble fractions were separated by centrifuging at 19,000 g. Lanes "S" and "I" are soluble and insoluble fractions, respectively.
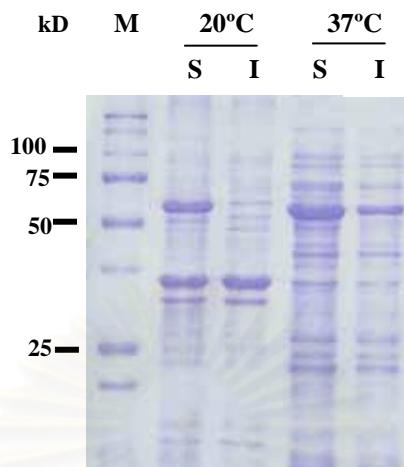
Figure 3.93 SDS-PAGE analysis of cell fractionation from recombinant expression of PmERp57. Cultured cells were induced by IPTG for 3 hr at 37 or 20 ℃. Cells were then harvested and disrupted. Soluble and insoluble fractions were separated by centrifuging at 19,000 g. Lanes "S" and "I" are soluble and insoluble fractions, respectively.

3.9.3 Peptide fingerprints of the recombinant proteins by mass-spectrometry

ESI-LC-MS/MS was performed to confirm the expression. The result showed that PmCRT was significantly matched with *P. monodon* EST-translated peptides similar to calreticulin (Figure 3.94A). The highest match was a peptide similar to calreticulin of *Galleria mellonella* (Score = 131; % coverage = 36%; Queries matched = 17). For PmERp57, the protein was significantly matched with *P. monodon* EST-deduced peptides similar to protein disulfide isomerase. The most significant hit was that from *Tribolium castaneum* (Score = 609, %coverage = 42%; Queries matched = 21) (Figure 3.94B). In contrast, PMCNX was not significantly matched with any deduced peptide in the *P. monodon* EST library (Figure 3.94C). This is because the C-terminal transmembrane domain (117 amino acid residues) was excluded for the recombinant PMCNX and the peptide similar to CNX in the database was only the last C-terminal 80 amino acids of the protein.
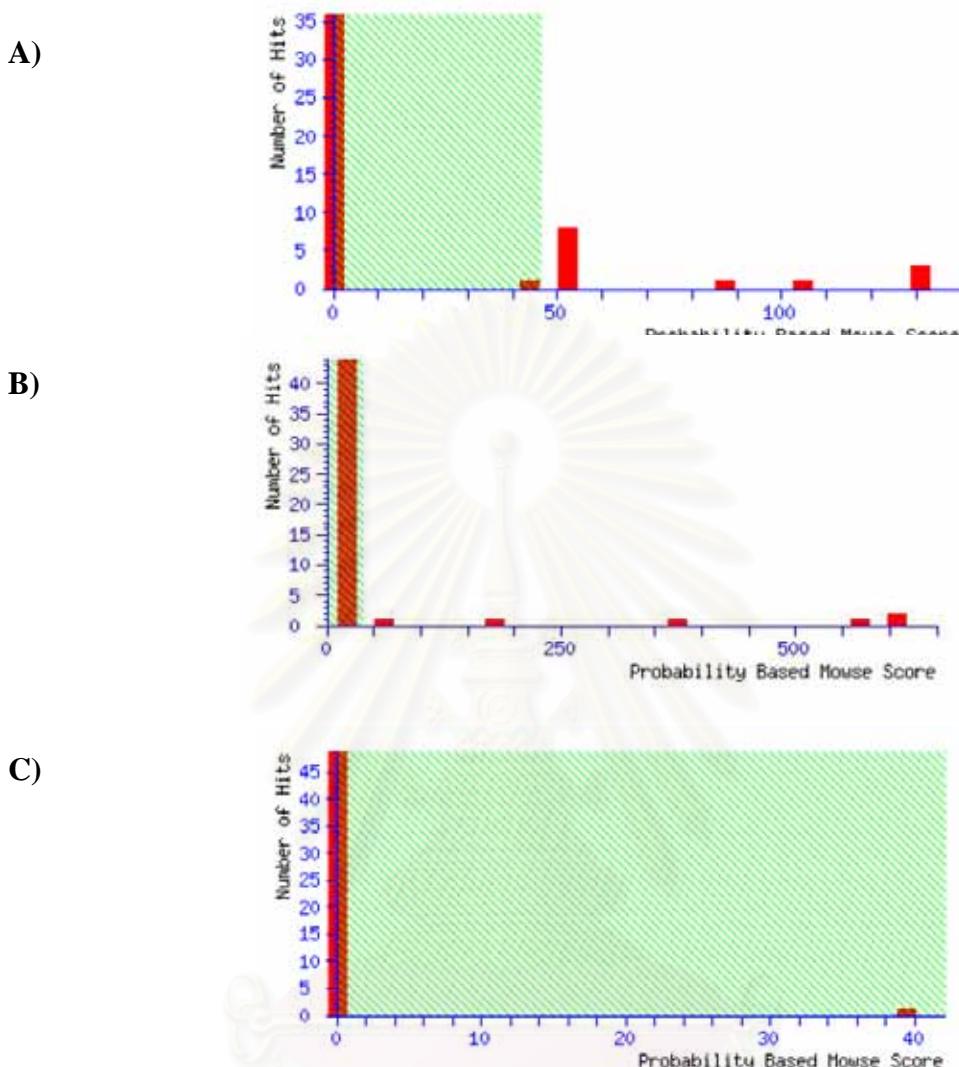
Figure 3.94 Identifications of peptide mass fingerprints of PmCRT (A) PmERp57 (B)and PmCNX (C) by ESI-LC MS/MS. The Mowse score is -10log (P), where P is the probability that the observed match is a random event. The height of each bar represents the number of protein(s) matched within a score range. The matches falling into the shaded area, where the probability of the random event is more than 5 %, are considered to be insignificant. Samples were prepared by in-gel tryptic digestion of the interested SDS-PAGE gel slices.

3.9.4 Purification of the recombinant proteins

Purification of the recombinant PmCRT and PmCNX, which were tagged with a GST sequence, was carried out, using the GST-glutathione system. Both proteins were specifically bound to the column. Some of the bound PmCRT and PmCNX proteins were washed out in the wash buffer containing no reduced glutathione. However, most of them were eluted with 10 mM of reduced glutathione (Figure 3.95 and 3.96). Yields of PmCRT and PmCNX purified from 300-ml bacterial cell cultures were 23.9 and 59.1 mg, respectively (Table 3.7).

For PmERp57, the purification was performed through the His-Ni system. From a SDS-PAGE gel, some of the PmERp57 protein began to be eluted together with other protein contaminants from the column at 500 mM imidazole in the wash buffer (Figure 3.97). However, most of the protein was found to come out with 500 mM imidazole. Yield of PmERp57 purified from 300-ml cultures was 98.6 mg (Table 3.7).
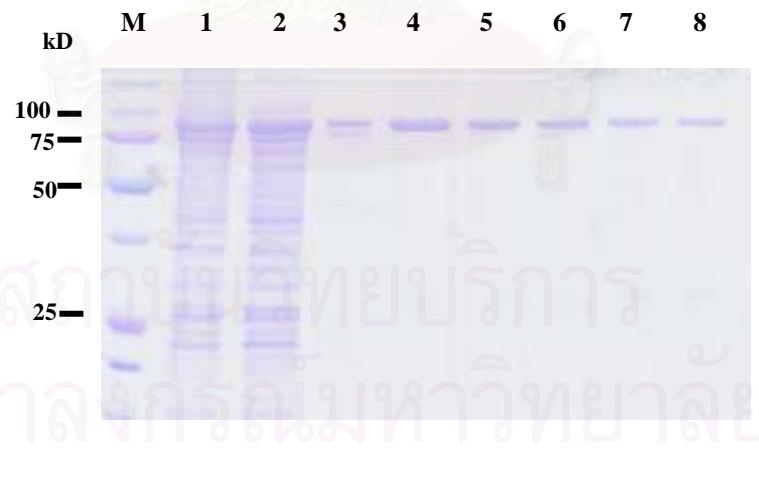
Figure 3.95 SDS-PAGE analysis of recombinant PmCRT purification, using a GSTrap FF column. Lane 1 is the unpurified soluble fraction prepared from an induced cell culture. Lanes 2 and 3 are the unbound flow through and wash fractions, respectively. Lanes 4-8 are the eluted (10 mM reduced glutathione) fractions 1-5, respectively).
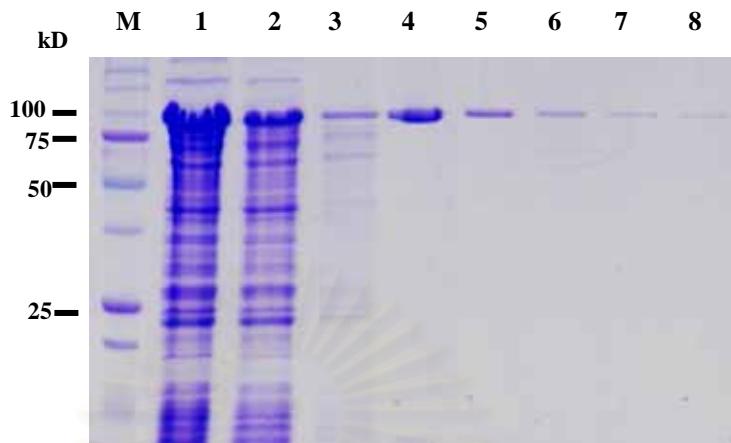
Figure 3.96 SDS-PAGE analysis of recombinant PmCNX purification, using a GSTrap FF column. Lane 1 is the unpurified soluble fraction prepared from an induced cell culture. Lanes 2 and 3 are the unbound flow through and wash fractions, respectively. Lanes 4-8 are the eluted (10 mM reduced glutathione) fractions 1-5, respectively).
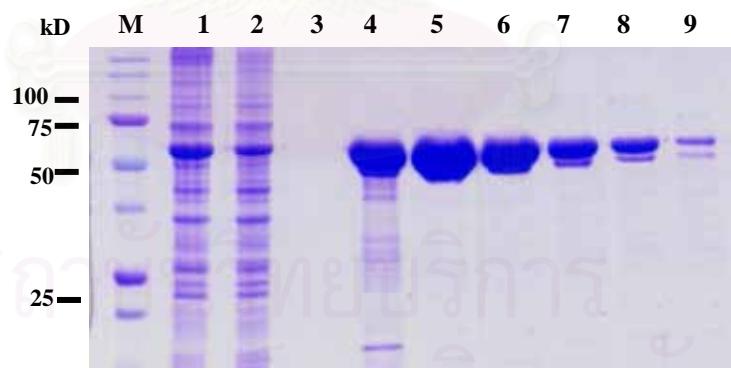
Figure 3.97 SDS-PAGE analysis of recombinant PmERp57 purification, using a His GraviTrap column. Lane 1 is the unpurified soluble fraction prepared from an induced cell culture. Lanes 2 and 3 are the unbound flow through and wash (with 80 mM

imidazole) fractions, respectively. Lanes 4-8 are the eluted (500 mM imidazole) fractions 1-5, respectively.

3.9.5 *In vitro* activity assays of PmCRT, PmERp57 and PmCNX

3.9.5.1 Calcium binding assay

Assay of Ca^{2+} binding activities of PmCRT, PmCNX and PmERp57 was first performed using a gel shift assay technique modified from Wilson *et.al.* (1980), which has been used to characterize calmodulin, a Ca^{2+} binding protein. In this assay, the proteins were incubated with Ca^{2+} before resolved by SDS-PAGE. Results from the SDS-PAGE showed no electrophoretic mobility shifts when the proteins were incubated with Ca^{2+} (Figure 3.98).

Consequently, calcium binding activity of the three proteins was then assayed by SDS-free PAGE in the native condition. Binding of calcium could be detected if the binding results in change of their conformation, causing the band mobility to shift. However, this technique is not applicable to PmCRT and PmERp57 since no mobility shift could be observed. No conclusion could be made. Binding of Ca^{2+} may not cause any conformational change to the proteins; alternatively, the expressed recombinant PmCRT and PmERp57 did not exhibit the activity (Figure 3.99). In contrast, the recombinant PmCNX seemed to bind to Ca^{2+} in the assay as a mobility shift could be observed for both the folded (bottom band) and unfolded (top band) PmCNX. However, the shift was somewhat indistinct and this assay is not a standard method for testing the Ca^{2+} binding activity of CRT and CNX. Therefore, a more reliable technique Ca^{45} autoradiography overlay assay is required to confirm these results.

Table 3.7 Summary of preparation and purification of recombinant PmCRT, PmCNX and PmERp57

	Recombinant protein		
	PmCRT	PmCNX	PmERp57
Size of protein (kDa)	71.35	77.95	53.90
Type and site of fusion tag	GST-tag/ N-terminal	GST-tag/ N-terminal	His-tag/ N-terminal
Form of the expressed protein	Soluble protein	Soluble protein	Soluble protein
Culture size	300 ml	300 ml	300 ml
Induction conditions	37ºC, 3 hr	37ºC, 3 hr	37ºC, 3 hr
Weight of bacterial wet cell (g/l)	5.09	6.7	7.42
Total yield (mg)	7.17	17.73	29.57
Yield (mg protein/l of culture)	23.9	59.11	98.57
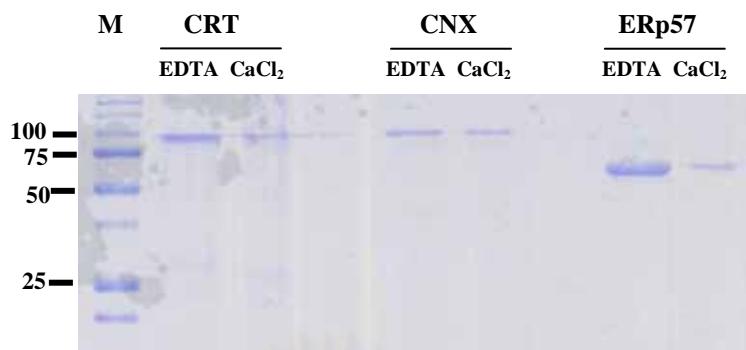
Figure 3.98 SDS-PAGE gel mobility shift assay of PmCRT (lanes 1 and 2), PmCNX (lanes 3 and 4) and PmERp57 (lanes 5 and 6). Ca^{2+} binding (lanes 2, 4 and 6) was performed by an on-ice incubation of the interested protein with 1 mM $CaCl_2$ for 30 min. For the controls (lanes 1, 3 and 5), proteins were incubated with 1 mM EDTA for 30 min. Protein bands were detected by Coomassie blue staining. Lane marked M contained molecular mass standard proteins
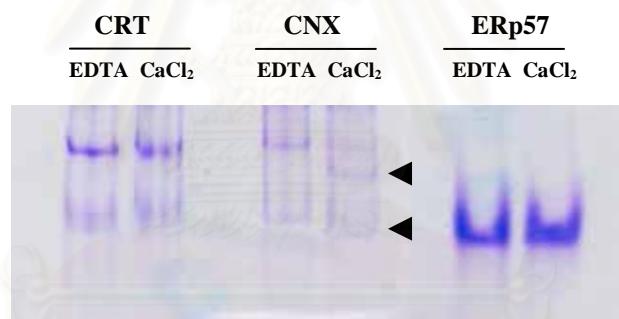
Figure 3.99 To analyze calcium-induced electrophoretic mobility of the PmCRT, PmCNX and PmERp57 on native-PAGE, one microgram of the eluted interesting protein and 1 mM of either EDTA (lane1 = PmCRT+EDTA, lane3 = PmCNX+EDTA and lane 5 = PmERp57+EDTA) or CaCl2 (lanes 2, 4 and 6 are PmCRT, PmCNX and PmERp57 incubated with 50mM $CaCl_2$, respectively) were resolved. Protein bands were detected by Coomassie blue staining. Lane marked M contained molecular mass standard proteins. The arrow heads show the mobility shift band. The upper and lower bands in lane 1-4 correspond to the unfolded and folded form of the according proteins (see Figure 3.100).

3.9.5.1 ERp57 binding assay of PmCRT and PmCNX

Interactions of PmCRT and PmCNX with PmERp57 were studied by a mobility shift assay on a native PAGE gel. When either PmCRT or PmCNX was combined and incubated with PmERp57, a protein band shift can be observed, suggesting that they complexed with PmERp57 (Figure 3.100). Moreover, these complexes could be disrupted when heated, resulting in two separated proteins. Interestingly, incubation of PmCRT with PmERp57 resulted in disappearance of the unfolded PmCRT (the upper band in lane 1). This suggests that PmERp57 may be able to form a complex with unfolded PmCRT or that it may refold PmCRT and then form the complex. Nevertheless, these results need further confirmation by other methods. In particular, complex formation of PmCRT or PmCNX with PmERp57 could be confirmed by mass spectrometry, protein cross-linking or immunoprecipitation.
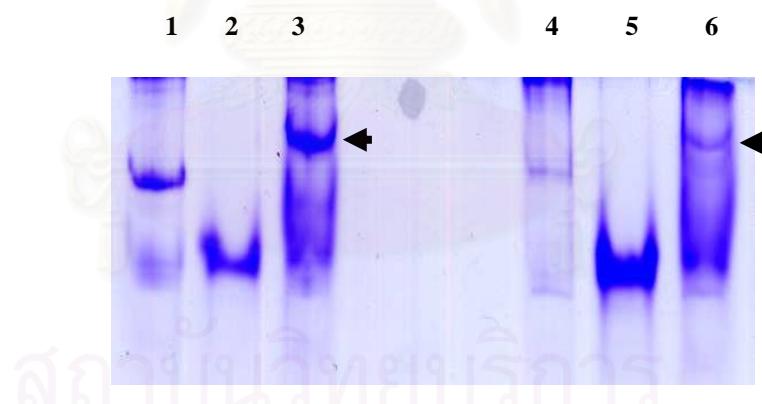
Figure 3.100 Gel mobility shift assay of PmCRT and PmCNX interactions with PmERp57. One μg of PmCRT and PmCNX mixed with 1 μg PmERp57 (lane 3 = PmCRT mixed PmERp57 while lane 6 = PmCNX incubated with PmERp57) were fractionated in a 10% (w/v) native PAGE gel and stained by Coomassie blue. Each protein was loaded on the gel as control for band shift comparison (PmCRT = lane 1, PmCNX = lane4 and PmERp57 = lanes 2 and 5). The arrow heads indicate the mobility shift band. The lower bands in lanes 3 and 6 are the excess uncomplexed PmERp57.

The arrow headed bands were cut to confirm using ESI-LC MS/MS. The results showed that all of these bands were real complex of PmCRT/ PmCNX and PmERp57. The PmCRT and PmERp57 complex was significantly matched with *P. monodon* EST-translated peptides similar to both calreticulin and protein disulfide isomerase (Figure 3.101A). The PmCRT and PmERp57 were the highest match to calreticulin of *Apis mellifera* (Score = 154; % coverage = 7%; Queries matched = 4) and protein disulfide isomerase of *Tribolium castaneum* (Score = 373; % coverage = 31%; Queries matched = 9), respectively. For PmCNX-PmERp57 complex band, PmCNX was significantly matched to calnexin of *Mus musculus* (Score = 88; % coverage = 2%; Queries matched = 1) when compared with NCBInr (Figure 3.101B). Wile PmERp57 in PmCNX-PmERp57 complex band was significantly matched with *P. monodon* EST-deduced peptides similar to protein disulfide isomerase. The most significant hit was that from *Tribolium castaneum* (Score = 501, %coverage = 39%; Queries matched = 12) (Figure 3.101C).

A)

B)
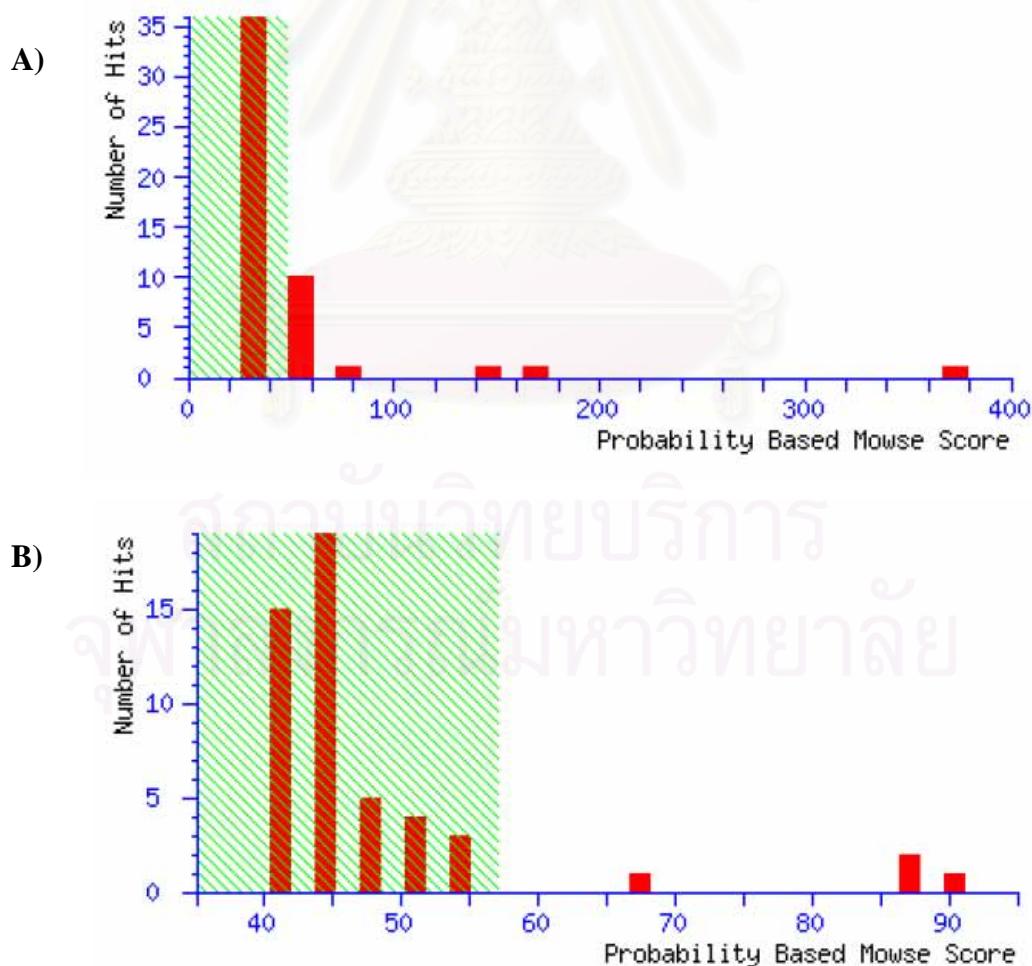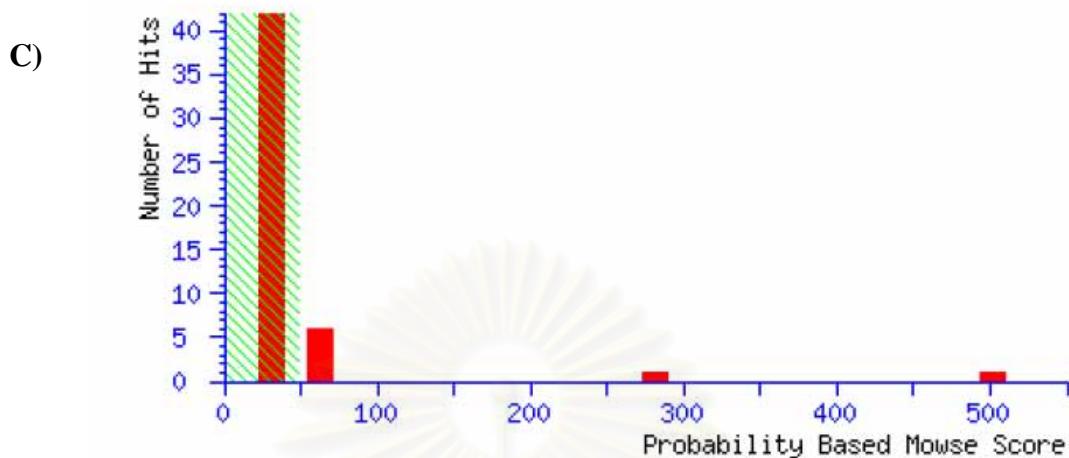
Figure 3.101 Identifications of peptide mass fingerprints of PmCRT-PmERp57 complex (A), PmCNX (B) and PmERp57 (C) in their complex band by ESI-LC MS/MS. The Mowse score is -10log (P), where P is the probability that the observed match is a random event. The height of each bar represents the number of protein(s) matched within a score range. The matches falling into the shaded area, where the probability of the random event is more than 5 %, are considered to be insignificant. Samples were prepared by in-gel tryptic digestion of the interested SDS-PAGE gel slices.

CHAPTER IV

DISCUSSION

The basic information on nucleotide sequences and genomic organization of *CRT, CNX* and *ERp57* genes and functional studies of their products are essential for understanding the involvement of these genes on homeostasis (e.g. Ca^{++} metabolism and stress) in shrimp.

The full length cDNA of *PmCRT, PmCNX* and *PmERp57* were characterized and reported for the first time in *P. monodon*. The translated amino acid sequence of PmCRT contained the signal sequence (MKTWVFLALFGVALVES), which play a role in its localization at the ER. Like other eukaryote CRTs, the deduced amino acid of PmCRT contained three typical conserved domains (Michalak *et al.*, 1999), which are N-, P- and C-domains. For PmCNX, its translated amino acid sequence contained the signal sequence, a luminal domain containing two calreticulin-signature motifs and two four-time repeated motifs (IXDXEXXKPE(/D)DWD and GxWxPPxIxNPxYx), a hydrophobic transmembrane region and highly charged cytosolic domain. PmERp57 possessed a four-domain structure containing a, a′, b and b′ domains. Like ERp57 in other species, two a and a' thioredoxin regions of PmERp57 contained a CGHC motif. These distinguishable domains found in *P. monodon* CRT, CNX and CRT to those from other species reinforce the fact that the domains are conserved for the proteins and essential for their functions.

From phylogenetic analysis, the tree topology supports the lack of gene duplication of *CRT*. PmCRT was allocated within the invertebrate group and evolutionary closely related to that of the Chinese shrimp (*F. chinensis*). In contrast, invertebrate *CNX* is rather diverse. PmCNX is distantly related to other invertebrate CNX. For the ERp57, phylogenetic analysis clearly illustrated lineage separation of vertebrate and invertebrate ERp57 with the exception of *S. purpuratus* ERp57 that was misplaced in the vertebrate group. Invertebrate ERp57, including PmERp57 contain an ER retention signal of KXEL while that of most vertebrates are QDEL (Mazzarella *et al.*, 1994). PmERp57 was closely similar to the mud crab PDI, which also contains an ER retention motif and the a, a′, b, b′ pattern (Maattanen *et al.*, 2006). This further indicated that ERp57 and PDI are members of the same protein family.

The length of *CRT* transcripts and genes in various species is not significantly different. Generally, the genomic DNA sequence of *CRT*s is approximately 3.0-4.6 kb in length. The genomic organization of human and mouse *CRT* consist of 9 exons. In contrast, the *PmCRT* gene contains only 4 exons. This data preliminary showed that evolution of the *CRT* gene may follow the general acceptance that an increase in genome size (3.3×10^9, 3.45×10^9 and 2.53×10^9 bp in human, mouse and *P. monodon*, respectively, www.genomesize.com) and gene number in different eukaryotic lineages is paralleled by a general decrease in genome compactness and an increase in the number and size of introns (Patthy, 1999). This theoretical phenomenon may respond to more numbers of intron of *CRT* in human and mouse than *P. monodon*.

The genomic structure of *CNX* was only reported in *Arabidopsis*. It contained 6 exons and 5 introns of 80-233 bp in length. (Boyce *et al.*, 1994). In this study, the genomic organization *PmCNX* was partially characterized. *PmCNX* contained at least 6 exons and only 3 complete introns (189 – 1245 bp in length) were successfully identified. Accordingly, the actual numbers of exons and introns of *PmCNX* is still inconclusive and should be further characterized.

The genomic structure of *ERp57* has not been reported in any species. Two forms of *PmERp57* shared the same ORF but different length of the 3´UTR (529 and 713 bp for the short and long forms, respectively) were found. The nucleotide sequence of the long form except the additional 3´UTR region of 179 bp in length was identical to that of the short form. Accordingly, length polymorphism of this gene should be generated through the alternative splicing process. Southern blot analysis and genome walking at the 3´UTR should be further carried out to indentify whether *PmERp57* is encoded from a single locus.

Generally, CRT has two main functions as a chaperone and regulator of Ca^{2+} homeostasis (Michalak *et al.*, 2002). Accumulation of unfolded or misfolded proteins in the ER and depletion of intracellular Ca^{2+} store cause "ER stress", leading to upregulation of the *CRT* expression. Expression of *CRT* is regulated through endoplasmic reticulum stress elements (ERSE) located in its promoter region (Waser *et al.*, 1997; Yoshida *et al.*, 1998). ERSE sequences, possessing the consensus motif CCAATTNNNNNNNNNCCACG, are necessary for induction of the *CRT* expression through the unfolded protein (UPR) (Kokame *et al.*, 2001). Promoter deletion assays

revealed that the responsive Ca^{2+} homeostasis antagonist (CHA) is controlled by ERSE sequences (Waser *et al.*, 1997). In the present work, two ERSEs were located in the promoter region of *PmCRT*. The number of ERSE of *PmCRT* was the same as that of the rainbow trout (Kales *et al.*, 2007), while three ERSEs were found in the promoter region of mouse and human *CRT* (Yoshida *et al.*, 1998).

It is interesting to note that the *PmCRT* promoter contained the classical heat shock element (NGAAN) like that of *heat shock protein 70* (Abravaya *et al.*, 1991), suggesting that *PmCRT* is responsive to heat. In addition, a general ultra spiracle transcription factor binding site (USP), which has been found in insects, was found in the promoter region of *PmCRT*. This USP site binds specifically to juvenile hormone (JH) to regulate of cellular and tissue development and to form stage specific morphological structures (Jones and Sharp, 1997). This suggests that PmCRT is associated with molting. These results showed that PmCRT might play multiple functions in *P. monodon* just like those reported in other organisms.

Similar to those in other species (Nakamura *et al.*, 1997; Munoz-Gotera *et al.*, 2001; Tutuncu *et al.*, 2004; Huang *et al.*, 1993; Kozaki *et al.* 1994; Iida *et al.* 1996; Marcus *et al.* 1996), *PmCRT, PmCNX* and *PmER57* were expressed in various tissues in broodstocks and juveniles. In particular, *PmCRT* was abundantly expressed in ovaries of both juvenile and broodstock. PmCRT may play the important role in oocyte of *P. monodon*. CRT has been thought to have a role in signal transduction during or after sperm-egg interactions at fertilization (Tutuncu *et al.*, 2004).

In this study, temporal response of the *PmCRT*, *PmCNX* and *PmERp57* to heat stress was clearly observed in hemocytes. The heat treatment caused early upregulation of these genes after the treatment. In contrast, the expression levels of these genes were relatively stable in gill and hepatopancreas. This indicated that shrimp hemocytes provide vital cellular protection against stress (Gourdon *et al.*, 2000).

Several experiments showed that a particular heat-responsive gene in different tissues responds to temperature stress differently in either expression patterns or time of response (Luana *et al.*, 2007). For instance, heat shock protein 70 (Hsp70) was rapidly expressed in brain, salivary gland, imaginal discs and hindgut while other

tissues expressed Hsp70 more slowly at the relatively constant levels during temperature stress of *Drosophilla melanogaster* larvae (Krebs and Feder, 1997).

Semiquantitative RT-PCR illustrated that expression levels of *PmCRT*, *PmCNX* and *PmERp57* in gill and hepatopancreas were not significantly altered upon temperature stress. In contrast levels of these genes in hemocytes were immediately increased during and early after the treatment. Quantitative real-time PCR further confirmed the results of these genes in hemocytes of *P. monodon*. As a result, both semiquantitative and quantitative real-time PCR can be used to monitor effects of temperature stress on expression of these genes. However, the former is simpler and more convenient than the latter approach.

The relative expression levels of *PmERp57* during 6 hrs after the treatment were much greater than that of *PmCRT* and *PmCNX*. In addition, those of *PmCRT* during the same period were also substantially higher than that of *PmCNX*. Therefore, *PmERp57* was the most suitable gene that can be applied to indicate the temperature stress status in *P. monodon*. The possible reason to explain strong responses of *PmERp57* upon temperature stress is that *PmERp57* contained two thioredoxin domains and the major function of proteins containing the thioredoxin motif is environmental stress responses (Zeller and Klug, 2006). In various organisms, overexpression of protein containing thioredoxin motif helps maintaining homeostasis after the stress induction (Matsuo *et al.*, 2001). In addition, *ERp57* is a member of protein disulfide isomerase (PDI) protein family. Members of this family play an important role on response to diverse ER stresses (Jacquier-Sarlin and Polla, 1996).

Previous studies in other species had also demonstrated the inducible effects of temperature stress on the *CRT* expression at the mRNA level (Conway *et al.*, 1995; Nguyen *et al.*, 1996). The promoter region of *CRT* contains classical the heat shock element like that of heat shock protein 70 (*hsp70*). Moreover, expression pattern of *P. monodon hsp70* in response to thermal stress (Lo *et al.*, 2004) was similar to that of *PmCRT*, *PmCNX* and *PmERp57*. Results from the present study indicated that these proteins should be regarded as heat inducible genes/proteins.

CNX could be crucial for an organism to survive under heat stress. In a *CNX*⁻ mutant of *C. elegens*, the survival rate under heat stress was lower than that of the

wild type (Lee *et al.*, 2005). In addition, due to their role as protein-folding chaperones, any stresses, which cause protein misfolding, may induce expression of these chaperones in order to attenuate the protein misfolding event (Frickel *et al.*, 2004; Grillo *et al.*, 2007).

Results from this study also suggest that increasing expression levels of *PmCRT*, *PmCNX* and *PmERp57* are associated with stress status. Upregulated levels of these genes during temperature stress particularly *PmERp57* and *PmCRT* may be applied as a stress bioindicator in *P. monodon*. In Thailand, domestication of *P. monodon* has been carried out for a period of time (Withyachumnarnkul et al., 1998). The basic knowledge on stress responses of *PmERp57* and *PmCRT* can be directly used for the selection of a stress tolerance strain from different families of domesticated *P. monodon*.

For further functional analysis, PmCRT, PmCNX and PmERp57 were expressed *in vitro* using the bacterial expression system. Recombinant PmCRT and PmCNX were successfully expressed using pGEX 4T-1 and expressed as GST-tagged proteins. These recombinant proteins were expressed in the soluble form similar to sCRT and sCNX of dog that were expressed in *E. coli* using pGEX-3X as the vector (Ihara *et al.*, 1999). Expression and purification of GST-fused PmCRT and PmCNX were performed under standard conditions and PmCNX (59.11 mg/l) gave a better yield than PmCRT (23.9 mg/l).

Alternatively, recombinant PmERp57 cloned into the pET15b expression vector was also successfully expressed in the soluble form at a low culturing temperature (20ºC). The yield of recombinant PmERp57 (98.57 mg/l) was much greater than that of PmCRT and PmCNX. Typically, the size GST-fused proteins were increased from the expected size of an expressed protein alone for approximately 20 kDa. This may cause a lower yield of PmCRT and PmCNX compared to PmERp57 which was expressed as the His-tagged form. Nevertheless, purification of western blot analysis of GST-fused PmCRT and PmCNX give more specific results than His-tagged PmERp57.

In this study, Ca^{2+} binding activity of the recombinant PmCRT, PmCNX and PmERp57 were assayed by gel mobility shift on both SDS-PAGE and native-PAGE

gels. No mobility shift from the SDS-PAGE approach could be because of the presence of EDTA in the sample loading buffer, which may remove the Ca^{2+} in sample. Since this method relies on a unique property of calmodulin, it may not be applicable to some other Ca^{2+}-binding proteins. Autoradiography and fluorescent overlay assays which are commonly used to assay the Ca^{2+} binding activity of CRT and CNX should be used instead.

The native-PAGE approach was performed on the assumption that Ca^{2+} binding might cause change in the protein conformation. However, conclusion could be made only for PmCNX. Binding of PmCNX to Ca^{2+} caused its conformation to be more compact. Like CNX from other species, its Ca^{2+} binding are sequence motif-dependent rather than conformation-dependent. However, the shift was somewhat indistinct and this assay is not a standard method for testing the Ca^{2+} binding activity of CRT and CNX. Therefore, a more reliable technique Ca^{45} autoradiography overlay assay is required to confirm these results.

Furthermore, interactions of PmCRT and PmCNX with PmERp57 were demonstrated. Interestingly, PmERp57 seemed to be able to form a complex with unfolded PmCRT. The complex formation of PmCRT-PmERp57 and PmCNX-PmERp57 were confirmed by mass spectrometry.

Stress is strongly related to shrimp health (Zeller and Klug, 2006). Stressed shrimp have higher mortality rates and are more sensitive to bacterial and viral infection (Luana *et al.*, 2007). In adults, physical stressors (e.g. light and temperature) suppress reproductive maturation of shrimp broodstock. The knowledge on isolation, characterization and functional analysis of *PmCRT*, *PmCNX* and *PmERp57* genes in *P. monodon* strongly suggest the significant involvement of these Ca^{2+} homeostasis-related genes on temperature stress of *P. monodon*. Expression levels of *PmCRT* and *PmERp57* may be used as bioindicators for the health status of *P. monodon*. Recombinant *PmCRT*, *PmCNX* and *PmERp57* were successfully expressed *in vitro* allowing further functional analysis of these proteins in the calcium metabolism regulation and stress response pathways of *P. monodon*.

CHAPTER V

CONCLUSION

1. The full length cDNA of *PMCRT* contained 1682 nucleotides, whose 1221 bp were ORF encoding 406 amino acid residues. The deduced protein has a theoretical molecular weight (MW) of 46.76 kDa and a theoretical isoelectric point (*pI*) of 4.3. Genomic *PMCRT* was 3043 bp in length and consisted of 4 exons and 3 introns.

2. The full length cDNA of a *PMCNX* homologue was 2509 bp long. Its ORF was 1788 bp long and encoded 595 amino acid residues. The deduced protein had a theoretical *pI* and MW of 4.4 and 67.4 kDa, respectively. Genomic organization of *PMCNX* was partially characterized in this study. Only 4 exons and 3 introns could be established.

3. Two forms of *PMERp57* transcripts were isolated. They contained an identical ORF of 1458 bp but different length of the 3' UTR (529 and 713 bp) The deduced protein contained 485 amino acid residues and has a theoretical MW and *pI* of 53.9 kDa and 5.48, respectively. Genomic *PMERp57* spanned 8254 bp and was composed of 10 exons and 9 introns.

4. All genes were expressed in eye stalk, pleopod, gill, heart, ovaries, testis, hepatopancreas, stomach, intestine, hemocytes, thoracic ganglia, lympiod organ and antennal gland of *P. monodon* juveniles and broodstocks. Their transcripts were abundant in gill, hemocytes, ovaries, testis, hepatopancreas and intestine.

5. From the semi-quantitative and real-time PCR studies, 3-hr heat treatment at 35 ℃ upregulated the expression of *PMCRT*, *PMCNX* and *PMERp57* in hemocytes within the first hour after treatment. After that, the effect was attenuated ($P < 0.05$). Their expression was not affected in hepatopancrease and gill.

6. Recombinant proteins of PMCRT, PMCNX and PMERp57 were produced using an *E. coli* system. The proteins were expressed in the soluble form. By growing a 300-ml culture, 7.17, 17.73 and 29.57 mg of PMCRT, PMCNX and PMERp57 were produced, respectively.

7. From the calcium binding assay using native PAGE, PMCNX could bind to Ca^{2+} while PMCRT and PMERp57 did not. The result suggested the sequence motif dependence of calcium binding. However, additional studies are required to affirm the result.

8. The native-PAGE gel shift assay demonstrated the interaction of PMERp57 with PMCRT or PMCNX. Similar to the Ca^{2+} binding study, this result needs further confirmation by other techniques.

REFERENCE

Ahearn, G.A., Mandal, P.K., and Mandal, A. 2004. Calcium regulation in crustaceans during the molt cycle: a review and update. Comp Biochem Physiol A Mol Integr Physiol 137: 247-257.

Allende, J.E., and Allende, C.C. 1995. Protein kinases. 4. Protein kinase CK2: an enzyme with multiple substrates and a puzzling regulation. Faseb J 9: 313-323.

Andrin, C., Pinkoski, M.J., Burns, K., Atkinson, E.A., Krahenbuhl, O., Hudig, D., Fraser, S.A., Winkler, U., Tschopp, J., Opas, M., Bleackley, R.C., and Michalak, M. 1998. Interaction between a Ca^{2+}-binding protein calreticulin and perforin, a component of the cytotoxic T-cell granules. Biochemistry 37: 10386-10394.

Arosa, F.A., de Jesus, O., Porto, G., Carmo, A.M., and de Sousa, M. 1999. Calreticulin is expressed on the cell surface of activated human peripheral blood T lymphocytes in association with major histocompatibility complex class I molecules. J Biol Chem 274: 16917-16922.

Ashby, M.C., and Tepikin, A.V. 2001. ER calcium and the functions of intracellular organelles. Semin Cell Dev Biol 12: 11-17.

Baksh, S., Burns, K., Andrin, C., and Michalak, M. 1995a. Interaction of calreticulin with protein disulfide isomerase. J Biol Chem 270: 31338-31344.

Baksh, S., and Michalak, M. 1991. Expression of calreticulin in *Escherichia coli* and identification of its Ca^{2+} binding domains. J Biol Chem 266: 21458-21465.

Baksh, S., Spamer, C., Heilmann, C., and Michalak, M. 1995b. Identification of the Zn^{2+} binding region in calreticulin. FEBS Lett 376: 53-57.

Basu, S., and Srivastava, P.K. 1999. Calreticulin, a peptide-binding chaperone of the endoplasmic reticulum, elicits tumor- and peptide-specific immunity. J Exp Med 189: 797-802.

Baumann, O., and Walz, B. 2001. Endoplasmic reticulum of animal cells and its organization into structural and functional domains. Int Rev Cytol 205: 149-214.

Bergeron, J.J., Brenner, M.B., Thomas, D.Y., and Williams, D.B. 1994. Calnexin: a membrane-bound chaperone of the endoplasmic reticulum. Trends Biochem Sci 19: 124-128.

Bjorkman, T., and Cleland, R.E. 1991. The role of extracellular free-calcium gradients in gravitropic signalling in maize roots. Planta 185: 379-384.

Blaustein, M.P., and Golovina, V.A. 2001. Structural complexity and functional diversity of endoplasmic reticulum Ca(2+) stores. Trends Neurosci 24: 602-608.

Boston, R.S., Viitanen, P.V., and Vierling, E. 1996. Molecular chaperones and protein folding in plants. Plant Mol Biol 32: 191-222.

Brothers, S.P., Janovick, J.A., and Conn, P.M. 2006. Calnexin regulated gonadotropin-releasing hormone receptor plasma membrane expression. J Mol Endocrinol 37: 479-488.

Burns, K., Duggan, B., Atkinson, E.A., Famulski, K.S., Nemer, M., Bleackley, R.C., and Michalak, M. 1994. Modulation of gene expression by calreticulin binding to the glucocorticoid receptor. Nature 367: 476-480.

Burns, K., Opas, M., and Michalak, M. 1997. Calreticulin inhibits glucocorticoid- but not cAMP-sensitive expression of tyrosine aminotransferase gene in cultured McA-RH7777 hepatocytes. Mol Cell Biochem 171: 37-43.

Carafoli, E., and Brini, M. 2000. Calcium pumps: structural basis for and mechanism of calcium transmembrane transport. Curr Opin Chem Biol 4: 152-161.

Chen, F., Hayes, P.M., Mulrooney, D.M., and Pan, A. 1994. Identification and characterization of cDNA clones encoding plant calreticulin in barley. Plant Cell 6: 835-843.

Chevet, E., Jakob, C.A., Thomas, D.Y., and Bergeron, J.J. 1999a. Calnexin family members as modulators of genetic diseases. Semin Cell Dev Biol 10: 473-480.

Chevet, E., Wong, H.N., Gerber, D., Cochet, C., Fazel, A., Cameron, P.H., Gushue, J.N., Thomas, D.Y., and Bergeron, J.J. 1999b. Phosphorylation by CK2 and MAPK enhances calnexin association with ribosomes. Embo J 18: 3655-3666.

Conway, E.M., Liu, L., Nowakowski, B., Steiner-Mosonyi, M., Ribeiro, S.P., and Michalak, M. 1995. Heat shock-sensitive expression of calreticulin. *In vitro* and *in vivo* up-regulation. J Biol Chem 270: 17011-17016.

Coppolino, M.G., and Dedhar, S. 1998. Calreticulin. Int J Biochem Cell Biol 30: 553-558.

Coppolino, M.G., Woodside, M.J., Demaurex, N., Grinstein, S., St-Arnaud, R., and Dedhar, S. 1997. Calreticulin is essential for integrin-mediated calcium signalling and cell adhesion. Nature 386: 843-847.

Corbett, E.F., Oikawa, K., Francois, P., Tessier, D.C., Kay, C., Bergeron, J.J., Thomas, D.Y., Krause, K.H., and Michalak, M. 1999. Ca^{2+} regulation of interactions between endoplasmic reticulum chaperones. J Biol Chem 274: 6203-6211.

David, V., Hochstenbach, F., Rajagopalan, S., and Brenner, M.B. 1993. Interaction with newly synthesized and retained proteins in the endoplasmic reticulum

suggests a chaperone function for human integral membrane protein IP90 (calnexin). J Biol Chem 268: 9585-9592.

Dedhar, S., Rennie, P.S., Shago, M., Hagesteijn, C.Y., Yang, H., Filmus, J., Hawley, R.G., Bruchovsky, N., Cheng, H., Matusik, R.J., and et al. 1994. Inhibition of nuclear hormone receptor activity by calreticulin. Nature 367: 480-483.

Dupuis, M., Schaerer, E., Krause, K.H., and Tschopp, J. 1993. The calcium-binding protein calreticulin is a major constituent of lytic granules in cytolytic T lymphocytes. J Exp Med 177: 1-7.

Dwarakanath, B.S., Verma, A., Bhatt, A.N., Parmar, V.S., and Raj, H.G. 2008. Targeting protein acetylation for improving cancer therapy. Indian J Med Res 128: 13-21.

Eggleton, P., Reid, K.B., Kishore, U., and Sontheimer, R.D. 1997. Clinical relevance of calreticulin in systemic lupus erythematosus. Lupus 6: 564-571.

Ellerman, D.A., Cohen, D.J., Da Ros, V.G., Morgenfeld, M.M., Busso, D., and Cuasnicu, P.S. 2006a. Sperm protein "DE" mediates gamete fusion through an evolutionarily conserved site of the CRISP family. Dev Biol 297: 228-237.

Ellerman, D.A., Myles, D.G., and Primakoff, P. 2006b. A role for sperm surface protein disulfide isomerase activity in gamete fusion: evidence for the participation of ERp57. Dev Cell 10: 831-837.

Ellgaard, L., Riek, R., Herrmann, T., Guntert, P., Braun, D., Helenius, A., and Wuthrich, K. 2001. NMR structure of the calreticulin P-domain. Proc Natl Acad Sci U S A 98: 3133-3138.

Fadel, M.P., Dziak, E., Lo, C.M., Ferrier, J., Mesaeli, N., Michalak, M., and Opas, M. 1999. Calreticulin affects focal contact-dependent but not close contact-dependent cell-substratum adhesion. J Biol Chem 274: 15085-15094.

Ferrari, D.M., and Soling, H.D. 1999. The protein disulphide-isomerase family: unravelling a string of folds. Biochem J 339: 1-10.

Fliegel, L., Burns, K., MacLennan, D.H., Reithmeier, R.A., and Michalak, M. 1989a. Molecular cloning of the high affinity calcium-binding protein (calreticulin) of skeletal muscle sarcoplasmic reticulum. J Biol Chem 264: 21522-21528.

Fliegel, L., Burns, K., Opas, M., and Michalak, M. 1989b. The high-affinity calcium binding protein of sarcoplasmic reticulum. Tissue distribution, and homology with calregulin. Biochim Biophys Acta 982: 1-8.

Fraser, S.A., Michalak, M., Welch, W.H., and Hudig, D. 1998. Calreticulin, a component of the endoplasmic reticulum and of cytotoxic lymphocyte granules, regulates perforin-mediated lysis in the hemolytic model system. Biochem Cell Biol 76: 881-887.

Freedman, R.B., Klappa, P., and Ruddock, L.W. 2002. Protein disulfide isomerases exploit synergy between catalytic and specific binding domains. EMBO Rep 3: 136-140.

Frickel, E.M., Frei, P., Bouvier, M., Stafford, W.F., Helenius, A., Glockshuber, R., and Ellgaard, L. 2004. ERp57 is a multifunctional thiol-disulfide oxidoreductase. J Biol Chem 279: 18277-18287.

Fuller, J.R., Pitzer, J.E., Godwin, U., Albertino, M., Machon, B.D., Kearse, K.P., and McConnell, T.J. 2004. Characterization of the molecular chaperone calnexin in the channel catfish, *Ictalurus punctatus*, and its association with MHC class II molecules. Dev Comp Immunol 28: 603-617.

Ghebrehiwet, B., and Peerschke, E.I. 2004. cC1q-R (calreticulin) and gC1q-R/p33: ubiquitously expressed multi-ligand binding cellular proteins involved in inflammation and infection. Mol Immunol 41: 173-183.

Goldberger, R.F., Epstein, C.J., and Anfinsen, C.B. 1964. Purification and properties of a microsomal enzyme system catalyzing the reactivation of reduced ribonuclease and lysozyme. J Biol Chem 239: 1406-1410.

Gray, A.J., Park, P.W., Broekelmann, T.J., Laurent, G.J., Reeves, J.T., Stenmark, K.R., and Mecham, R.P. 1995. The mitogenic effects of the B beta chain of fibrinogen are mediated through cell surface calreticulin. J Biol Chem 270: 26602-26606.

Grillo, C., D'Ambrosio, C., Consalvi, V., Chiaraluce, R., Scaloni, A., Maceroni, M., Eufemi, M., and Altieri, F. 2007. DNA-binding activity of the ERp57 C-terminal domain is related to a redox-dependent conformational change. J Biol Chem 282: 10299-10310.

Hammond, C., and Helenius, A. 1995. Quality control in the secretory pathway. Curr Opin Cell Biol 7: 523-529.

Hebert, D.N., Foellmer, B., and Helenius, A. 1995. Glucose trimming and reglucosylation determine glycoprotein association with calnexin in the endoplasmic reticulum. Cell 81: 425-433.

High, S., Lecomte, F.J., Russell, S.J., Abell, B.M., and Oliver, J.D. 2000. Glycoprotein folding in the endoplasmic reticulum: a tale of three chaperones? FEBS Lett 476: 38-41.

Ho, S.C., Rajagopalan, S., Chaudhuri, S., Shieh, C.C., Brenner, M.B., and Pillai, S. 1999. Membrane anchoring of calnexin facilitates its interaction with its targets. Mol Immunol 36: 1-12.

Hofer, A.M., Curci, S., Machen, T.E., and Schulz, I. 1996. ATP regulates calcium leak from agonist-sensitive internal calcium stores. Faseb J 10: 302-308.

Jakob, C.A., Bodmer, D., Spirig, U., Battig, P., Marcil, A., Dignard, D., Bergeron, J.J., Thomas, D.Y., and Aebi, M. 2001. Htm1p, a mannosidase-like protein, is involved in glycoprotein degradation in yeast. EMBO Rep 2: 423-430.

Jaworski, D.C., Higgins, J.A., Radulovic, S., Vaughan, J.A., and Azad, A.F. 1996. Presence of calreticulin in vector fleas (Siphonaptera). J Med Entomol 33: 482-489.

Jeffery, J., Kendall, J.M., and Campbell, A.K. 2000. Apoaequorin monitors degradation of endoplasmic reticulum (ER) proteins initiated by loss of ER Ca(2+). Biochem Biophys Res Commun 268: 711-715.

John, L.M., Lechleiter, J.D., and Camacho, P. 1998. Differential modulation of SERCA2 isoforms by calreticulin. J Cell Biol 142: 963-973.

Johnson, S., Michalak, M., Opas, M., and Eggleton, P. 2001. The ins and outs of calreticulin: from the ER lumen to the extracellular space. Trends Cell Biol 11: 122-129.

Kales, S.C., Bols, N.C., and Dixon, B. 2007. Calreticulin in rainbow trout: a limited response to endoplasmic reticulum (ER) stress. Comp Biochem Physiol B Biochem Mol Biol 147: 607-615.

Kemmink, J., Darby, N.J., Dijkstra, K., Nilges, M., and Creighton, T.E. 1997. The folding catalyst protein disulfide isomerase is constructed of active and inactive thioredoxin modules. Curr Biol 7: 239-245.

Khalife, J., Trottein, F., Schacht, A.M., Godin, C., Pierce, R.J., and Capron, A. 1993. Cloning of the gene encoding a *Schistosoma mansoni* antigen homologous to human Ro/SS-A autoantigen. Mol Biochem Parasitol 57: 193-202.

Khanna, N.C., and Waisman, D.M. 1986. Development of a radioimmunoassay for quantitation of calregulin in bovine tissues. Biochemistry 25: 1078-1082.

Koivunen, P., Salo, K.E., Myllyharju, J., and Ruddock, L.W. 2005. Three binding sites in protein-disulfide isomerase cooperate in collagen prolyl 4-hydroxylase tetramer assembly. J Biol Chem 280: 5227-5235.

Kokame, K., Kato, H., and Miyata, T. 2001. Identification of ERSE-II, a new cis-acting element responsible for the ATF6-dependent mammalian unfolded protein response. J Biol Chem 276: 9199-9205.

Krause, K.H., and Michalak, M. 1997. Calreticulin. Cell 88: 439-443.

Kwiatkowski, D., and Bate, C. 1995. Inhibition of tumour necrosis factor (TNF) production by antimalarial drugs used in cerebral malaria. Trans R Soc Trop Med Hyg 89: 215-216.

Labriola, C., Cazzulo, J.J., and Parodi, A.J. 1999. *Trypanosoma cruzi* calreticulin is a lectin that binds monoglucosylated oligosaccharides but not protein moieties of glycoproteins. Mol Biol Cell 10: 1381-1394.

Leach, M.R., Cohen-Doyle, M.F., Thomas, D.Y., and Williams, D.B. 2002. Localization of the lectin, ERp57 binding, and polypeptide binding sites of calnexin and calreticulin. J Biol Chem 277: 29686-29697.

Lenter, M., and Vestweber, D. 1994. The integrin chains beta 1 and alpha 6 associate with the chaperone calnexin prior to integrin assembly. J Biol Chem 269: 12263-12268.

Li, F., Mandal, M., Barnes, C.J., Vadlamudi, R.K., and Kumar, R. 2001. Growth factor regulation of the molecular chaperone calnexin. Biochem Biophys Res Commun 289: 725-732.

Luana, W., Li, F., Wang, B., Zhang, X., Liu, Y., and Xiang, J. 2007. Molecular characteristics and expression analysis of calreticulin in Chinese shrimp *Fenneropenaeus chinensis*. Comp Biochem Physiol B Biochem Mol Biol 147: 482-491.

Maattanen, P., Kozlov, G., Gehring, K., and Thomas, D.Y. 2006. ERp57 and PDI: multifunctional protein disulfide isomerases with similar domain architectures but differing substrate-partner associations. Biochem Cell Biol 84: 881-889.

Mattson, M.P., and Spaziani, E. 1986. Regulation of crab Y-organ steroidogenesis in vitro: evidence that ecdysteroid production increases through activation of cAMP-phosphodiesterase by calcium-calmodulin. Mol Cell Endocrinol 48: 135-151.

Mazzarella, R.A., Gold, P., Cunningham, M., and Green, M. 1992. Determination of the sequence of an expressible cDNA clone encoding ERp60/calregulin by the use of a novel nested set method. Gene 120: 217-225.

McCauliffe, D.P., Yang, Y.S., Wilson, J., Sontheimer, R.D., and Capra, J.D. 1992. The 5'-flanking region of the human calreticulin gene shares homology with the human GRP78, GRP94, and protein disulfide isomerase promoters. J Biol Chem 267: 2557-2562.

McCauliffe, D.P., Zappi, E., Lieu, T.S., Michalak, M., Sontheimer, R.D., and Capra, J.D. 1990. A human Ro/SS-A autoantigen is the homologue of calreticulin and is highly homologous with onchocercal RAL-1 antigen and an aplysia "memory molecule". J Clin Invest 86: 332-335.

Meldolesi, J., Krause, K.H., and Michalak, M. 1996. Calreticulin: how many functions in how many cellular compartments? Como, April 1996. Cell Calcium 20: 83-86.

Mery, L., Mesaeli, N., Michalak, M., Opas, M., Lew, D.P., and Krause, K.H. 1996. Overexpression of calreticulin increases intracellular Ca^{2+} storage and decreases store-operated Ca2+ influx. J Biol Chem 271: 9332-9339.

Michalak, M., Burns, K., Andrin, C., Mesaeli, N., Jass, G.H., Busaan, J.L., and Opas, M. 1996. Endoplasmic reticulum form of calreticulin modulates glucocorticoid-sensitive gene expression. J Biol Chem 271: 29436-29445.

Michalak, M., Corbett, E.F., Mesaeli, N., Nakamura, K., and Opas, M. 1999. Calreticulin: one protein, one gene, many functions. Biochem J 344: 281-292.

Michalak, M., Milner, R.E., Burns, K., and Opas, M. 1992. Calreticulin. Biochem J 285, 681-692.

Michalak, M., Robert Parker, J.M., and Opas, M. 2002. Ca^{2+} signaling and calcium binding chaperones of the endoplasmic reticulum. Cell Calcium 32: 269-278.

Michikawa, T., Hamanaka, H., Otsu, H., Yamamoto, A., Miyawaki, A., Furuichi, T., Tashiro, Y., and Mikoshiba, K. 1994. Transmembrane topology and sites of N-glycosylation of inositol 1,4,5-trisphosphate receptor. J Biol Chem 269: 9184-9189.

Miyawaki, A., Llopis, J., Heim, R., McCaffery, J.M., Adams, J.A., Ikura, M., and Tsien, R.Y. 1997. Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. Nature 388: 882-887.

Muller-Taubenberger, A., Lupas, A.N., Li, H., Ecke, M., Simmeth, E., and Gerisch, G. 2001. Calreticulin and calnexin in the endoplasmic reticulum are important for phagocytosis. Embo J 20: 6772-6782.

Nakamura, M., Moriya, M., Baba, T., Michikawa, Y., Yamanobe, T., Arai, K., Okinaga, S., and Kobayashi, T. 1993. An endoplasmic reticulum protein, calreticulin, is transported into the acrosome of rat sperm. Exp Cell Res 205: 101-110.

Nakatsuji, T., Lee, C.Y., and Watson, R.D. 2008. Crustacean molt-inhibiting hormone: Structure, function, and cellular mode of action. Comp Biochem Physiol A Mol Integr Physiol.

Nakatsukasa, K., Nishikawa, S., Hosokawa, N., Nagata, K., and Endo, T. 2001. Mnl1p, an alpha -mannosidase-like protein in yeast *Saccharomyces cerevisiae*, is required for endoplasmic reticulum-associated degradation of glycoproteins. J Biol Chem 276: 8635-8638.

Nakhasi, H.L., Pogue, G.P., Duncan, R.C., Joshi, M., Atreya, C.D., Lee, N.S., and Dwyer, D.M. 1998. Implications of calreticulin function in parasite biology. Parasitol Today 14: 157-160.

Nakhasi, H.L., Singh, N.K., Pogue, G.P., Cao, X.Q., and Rouault, T.A. 1994. Identification and characterization of host factor interactions with cis-acting elements of rubella virus RNA. Arch Virol Suppl 9: 255-267.

Nemere, I. 2005. The 1,25D3-MARRS protein: contribution to steroid stimulated calcium uptake in chicks and rats. Steroids 70: 455-457.

Ohsako, S., Janulis, L., Hayashi, Y., and Bunick, D. 1998. Characterization of domains in mice of calnexin-t, a putative molecular chaperone required in sperm fertility, with use of glutathione S-transferase-fusion proteins. Biol Reprod 59: 1214-1223.

Ohtani, H., Wakui, H., Ishino, T., Komatsuda, A., and Miura, A.B. 1993. An isoform of protein disulfide isomerase is expressed in the developing acrosome of spermatids during rat spermiogenesis and is transported into the nucleus of mature spermatids and epididymal spermatozoa. Histochemistry 100: 423-429.

Okumura, T. 2006. Effects of cyclic nucleotides, calcium ionophore, and phorbol ester on vitellogenin mRNA levels in incubated ovarian fragments of the kuruma prawn *Marsupenaeus japonicus*. Gen Comp Endocrinol 148: 245-251.

Opas, M., Dziak, E., Fliegel, L., and Michalak, M. 1991. Regulation of expression and intracellular distribution of calreticulin, a major calcium binding protein of nonmuscle cells. J Cell Physiol 149: 160-171.

Opas, M., Szewczenko-Pawlikowski, M., Jass, G.K., Mesaeli, N., and Michalak, M. 1996. Calreticulin modulates cell adhesiveness via regulation of vinculin expression. J Cell Biol 135: 1913-1923.

Ostwald, T.J., and MacLennan, D.H. 1974. Isolation of a high affinity calcium-binding protein from sarcoplasmic reticulum. J Biol Chem 249: 974-979.

Parekh, A.B. 2000. Calcium signaling and acute pancreatitis: specific response to a promiscuous messenger. Proc Natl Acad Sci U S A 97: 12933-12934.

Park, B.J., Lee, D.G., Yu, J.R., Jung, S.K., Choi, K., Lee, J., Lee, J., Kim, Y.S., Lee, J.I., Kwon, J.Y., Lee, J., Singson, A., Song, W.K., Eom, S.H., Park, C.S., Kim, D.H., Bandyopadhyay, J., and Ahnn, J. 2001. Calreticulin, a calcium-binding molecular chaperone, is required for stress response and fertility in *Caenorhabditis elegans*. Mol Biol Cell 12: 2835-2845.

Persson, S., Rosenquist, M., and Sommarin, M. 2002. Identification of a novel calreticulin isoform (Crt2) in human and mouse. Gene 297: 151-158.

Plakidou-Dymock, S., and McGivan, J.D. 1994. Calreticulin--a stress protein induced in the renal epithelial cell line NBL-1 by amino acid deprivation. Cell Calcium 16: 1-8.

Putney, J.W., Jr., and Ribeiro, C.M. 2000. Signaling pathways between the plasma membrane and endoplasmic reticulum calcium stores. Cell Mol Life Sci 57: 1272-1286.

Rauch, F., Prud'homme, J., Arabian, A., Dedhar, S., and St-Arnaud, R. 2000. Heart, brain, and body wall defects in mice lacking calreticulin. Exp Cell Res 256: 105-111.

Ribeiro, C.M., McKay, R.R., Hosoki, E., Bird, G.S., and Putney, J.W., Jr. 2000. Effects of elevated cytoplasmic calcium and protein kinase C on endoplasmic reticulum structure and function in HEK293 cells. Cell Calcium 27: 175-185.

Ritter, C., Quirin, K., Kowarik, M., and Helenius, A. 2005. Minor folding defects trigger local modification of glycoproteins by the ER folding sensor GT. Embo J 24: 1730-1738.

Roderick, H.L., Lechleiter, J.D., and Camacho, P. 2000. Cytosolic phosphorylation of calnexin controls intracellular Ca(2+) oscillations via an interaction with SERCA2b. J Cell Biol 149: 1235-1248.

Rojiani, M.V., Finlay, B.B., Gray, V., and Dedhar, S. 1991. *In vitro* interaction of a polypeptide homologous to human Ro/SS-A antigen (calreticulin) with a highly conserved amino acid sequence in the cytoplasmic domain of integrin alpha subunits. Biochemistry 30: 9859-9866.

Rooke, K., Briquet-Laugier, V., Xia, Y.R., Lusis, A.J., and Doolittle, M.H. 1997. Mapping of the gene for calreticulin (Calr) to mouse chromosome 8. Mamm Genome 8: 870-871.

Ruoslahti, E., and Reed, J.C. 1994. Anchorage dependence, integrins, and apoptosis. Cell 77: 477-478.

Russell, S.J., Ruddock, L.W., Salo, K.E., Oliver, J.D., Roebuck, Q.P., Llewellyn, D.H., Roderick, H.L., Koivunen, P., Myllyharju, J., and High, S. 2004. The primary substrate binding site in the b' domain of ERp57 is adapted for endoplasmic reticulum lectin association. J Biol Chem 279: 18861-18869.

Schrag, J.D., Bergeron, J.J., Li, Y., Borisova, S., Hahn, M., Thomas, D.Y., and Cygler, M. 2001. The Structure of calnexin, an ER chaperone involved in quality control of protein folding. Mol Cell 8: 633-644.

Schreiber, K.L., Bell, M.P., Huntoon, C.J., Rajagopalan, S., Brenner, M.B., and McKean, D.J. 1994. Class II histocompatibility molecules associate with calnexin during assembly in the endoplasmic reticulum. Int Immunol 6: 101-111.

Shago, M., Flock, G., Leung Hagesteijn, C.Y., Woodside, M., Grinstein, S., Giguere, V., and Dedhar, S. 1997. Modulation of the retinoic acid and retinoid X receptor signaling pathways in P19 embryonal carcinoma cells by calreticulin. Exp Cell Res 230: 50-60.

Shrivastava, A., and Calame, K. 1994. An analysis of genes regulated by the multi-functional transcriptional regulator Yin Yang-1. Nucleic Acids Res 22: 5151-5155.

Smith, M.J. 1992. A C. elegans gene encodes a protein homologous to mammalian calreticulin. DNA Seq 2: 235-240.

Smith, M.J., and Koch, G.L. 1989. Multiple zones in the sequence of calreticulin (CRP55, calregulin, HACBP), a major calcium binding ER/SR protein. Embo J 8: 3581-3586.

Sontheimer, R.D., Nguyen, T.Q., Cheng, S.T., Lieu, T.S., and Capra, J.D. 1995. The unveiling of calreticulin--a clinically relevant tour of modern cell biology. J Investig Med 43: 362-370.

Sousa, M., and Parodi, A.J. 1995. The molecular basis for the recognition of misfolded glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. Embo J 14: 4196-4203.

Subramanian, K., and Meyer, T. 1997. Calcium-induced restructuring of nuclear envelope and endoplasmic reticulum calcium stores. Cell 89: 963-971.

Tassanakajon, A., Klinbunga, S., Paunglarp, N., Rimphanitchayakit, V., Udomkit, A., Jitrapakdee, S., Sritunyalucksana, K., Phongdara, A., Pongsomboon, S., Supungul, P., Tang, S., Kuphanumart, K., Pichyangkura, R., and Lursinsap, C. 2006. *Penaeus monodon* gene discovery project: the generation of an EST collection and establishment of a database. Gene 384: 104-112.

Taylor, S.C., Ferguson, A.D., Bergeron, J.J., and Thomas, D.Y. 2004. The ER protein folding sensor UDP-glucose glycoprotein-glucosyltransferase modifies substrates distant to local changes in glycoprotein conformation. Nat Struct Mol Biol 11: 128-134.

Terasaki, M., Jaffe, L.A., Hunnicutt, G.R., and Hammer, J.A., 3rd. 1996. Structural change of the endoplasmic reticulum during fertilization: evidence for loss of membrane continuity using the green fluorescent protein. Dev Biol 179: 320-328.

Tharin, S., Dziak, E., Michalak, M., and Opas, M. 1992. Widespread tissue distribution of rabbit calreticulin, a non-muscle functional analogue of calsequestrin. Cell Tissue Res 269: 29-37.

Tjoelker, L.W., Seyfried, C.E., Eddy, R.L., Jr., Byers, M.G., Shows, T.B., Calderon, J., Schreiber, R.B., and Gray, P.W. 1994. Human, mouse, and rat calnexin cDNA cloning: identification of potential calcium binding motifs and gene localization to human chromosome 5. Biochemistry 33: 3229-3236.

Tutuncu, L., Stein, P., Ord, T.S., Jorgez, C.J., and Williams, C.J. 2004. Calreticulin on the mouse egg surface mediates transmembrane signaling linked to cell cycle resumption. Dev Biol 270: 246-260.

Van Leeuwen, J.E., and Kearse, K.P. 1996. The related molecular chaperones calnexin and calreticulin differentially associate with nascent T cell antigen receptor proteins within the endoplasmic reticulum. J Biol Chem 271: 25345-25349.

Vassilakos, A., Michalak, M., Lehrman, M.A., and Williams, D.B. 1998. Oligosaccharide binding characteristics of the molecular chaperones calnexin and calreticulin. Biochemistry 37: 3480-3490.

Verboomen, H., Wuytack, F., Van den Bosch, L., Mertens, L., and Casteels, R. 1994. The functional importance of the extreme C-terminal tail in the gene 2 organellar Ca(2+)-transport ATPase (SERCA2a/b). Biochem J 303: 979-984.

Verkhratsky, A., and Toescu, E.C. 1998. Calcium and neuronal ageing. Trends Neurosci 21: 2-7.

Wada, I., Rindress, D., Cameron, P.H., Ou, W.J., Doherty, J.J., 2nd, Louvard, D., Bell, A.W., Dignard, D., Thomas, D.Y., and Bergeron, J.J. 1991. SSR alpha and associated calnexin are major calcium binding proteins of the endoplasmic reticulum membrane. J Biol Chem 266: 19599-19610.

Ware, F.E., Vassilakos, A., Peterson, P.A., Jackson, M.R., Lehrman, M.A., and Williams, D.B. 1995. The molecular chaperone calnexin binds Glc1Man9GlcNAc2 oligosaccharide as an initial step in recognizing unfolded glycoproteins. J Biol Chem 270: 4697-4704.

Waser, M., Mesaeli, N., Spencer, C., and Michalak, M. 1997. Regulation of calreticulin gene expression by calcium. J Cell Biol 138: 547-557.

Watanabe, D., Yamada, K., Nishina, Y., Tajima, Y., Koshimizu, U., Nagata, A., and Nishimune, Y. 1994. Molecular cloning of a novel Ca(2+)-binding protein (calmegin) specifically expressed during male meiotic germ cell development. J Biol Chem 269: 7744-7749.

White, T.K., Zhu, Q., and Tanzer, M.L. 1995. Cell surface calreticulin is a putative mannoside lectin which triggers mouse melanoma cell spreading. J Biol Chem 270: 15926-15929.

Williams, D.B. 2006. Beyond lectins: the calnexin/calreticulin chaperone system of the endoplasmic reticulum. J Cell Sci 119: 615-623.

Xu, G., Fang, Q.Q., Sun, Y., Keirans, J.E., and Durden, L.A. 2005. Hard tick calreticulin (CRT) gene coding regions have only one intron with conserved positions and variable sizes. J Parasitol 91: 1326-1331.

Yamamoto, S., and Nakamura, M. 1996. Calnexin: its molecular cloning and expression in the liver of the frog, Rana rugosa. FEBS Lett 387: 27-32.

Yoshida, H., Haze, K., Yanagi, H., Yura, T., and Mori, K. 1998. Identification of the cis-acting endoplasmic reticulum stress response element responsible for transcriptional induction of mammalian glucose-regulated proteins. Involvement of basic leucine zipper transcription factors. J Biol Chem 273: 33741-33749.

Zapun, A., Darby, N.J., Tessier, D.C., Michalak, M., Bergeron, J.J., and Thomas, D.Y. 1998. Enhanced catalysis of ribonuclease B folding by the interaction of calnexin or calreticulin with ERp57. J Biol Chem 273: 6009-6012.

Zhang, J., Wu, J., Huo, R., Mao, Y., Lu, Y., Guo, X., Liu, J., Zhou, Z., Huang, X., and Sha, J. 2007. ERp57 is a potential biomarker for human fertilization capability. Mol Hum Reprod 13: 633-639.

Zhu, J. 1996. Ultraviolet B irradiation and cytomegalovirus infection synergize to induce the cell surface expression of 52-kD/Ro antigen. Clin Exp Immunol 103: 47-53.

Ziegler, A., Weihrauch, D., Towle, D.W., and Hagedorn, M. 2002. Expression of Ca^{2+}-ATPase and Na^+/Ca^{2+}-exchanger is upregulated during epithelial Ca^{2+} transport in hypodermal cells of the isopod *Porcellio scaber*. Cell Calcium 32: 131-141.

APPENDICES

Appendix A

Table A1 Raw materials and relative expression level data of PMCRT in hemocytes of thermal stress treated juvenile *P. monodon* using semi-quantitative RT-PCR

Sample Group		Densities of band		Ratio of gene/EF-1α	Average	STD
		PMCRT Density	EF-1α			
Control	N-1	646.66	1676.86	0.385637441		
	N-2	641.77	1722.96	0.372481079	0.41155	0.0507
	N-3	815.77	1711.88	0.47653457		
0 hpt	0-1	1279.59	1465.47	0.873160147		
	0-2	1304.82	1536.66	0.849127328	0.86261	0.011
	0-3	1297.62	1499.2	0.865541622		
1 hpt	1_1	1232.71	1781.48	0.691958372		
	1_2	1147.52	1759.33	0.652248299	0.67902	0.0207
	1_3	1194.97	1724.7	0.692856729		
3 hpt	3_1	794.71	1788.39	0.444371753		
	3_2	831.75	1752.56	0.474591455	0.47576	0.0286
	3_3	881.42	1733.99	0.508318964		
6 hpt	6_1	714.63	1702.1	0.419851948		
	6_2	711.05	1600.87	0.444164735	0.42826	0.0123
	6_3	722.1	1716.13	0.420772319		
12 hpt	12_1	797.64	1737.93	0.458959797		
	12_2	755.1	1704.18	0.443086998	0.47931	0.0444
	12_3	950.09	1772.95	0.535880877		
24 hpt	24_4	661.54	1348.88	0.49043651		
	24_5	704.42	1434.32	0.491117742	0.47776	0.0202
	24_6	683.32	1512.67	0.451731045		
48 hpt	48_1	668.98	1585.48	0.42194162		
	48_2	672.93	1588.73	0.423564734	0.44174	0.0294
	48_3	789.79	1646.38	0.479713067		

Table A2 Raw materials and relative expression level data of PMCNX in hemocytes of thermal stress treated juvenile *P. monodon* using semi-quantitative RT-PCR

Sample Group		Densities of band		Ratio of gene/EF-1α	Average	STD
		PMCNX Density	EF-1α			
Control	N-1	593.82	1676.86	0.354126164		
	N-2	670.42	1722.96	0.38910944	0.36194	0.022
	N-3	586.48	1711.88	0.342594107		
0 hpt	0-1	951.32	1465.47	0.649156926		
	0-2	1018.68	1536.66	0.662918277	0.65922	0.008
	0-3	997.83	1499.2	0.665574973		
1 hpt	1_1	990.149	1781.48	0.555801356		
	1_2	929.4	1759.33	0.528269284	0.53383	0.018
	1_3	892.4	1724.7	0.51742332		
3 hpt	3_1	750.73	1788.39	0.419779802		
	3_2	812.97	1752.56	0.463875702	0.43327	0.024
	3_3	721.59	1733.99	0.416144268		
6 hpt	6_1	583.13	1702.1	0.342594442		
	6_2	691.29	1600.87	0.431821447	0.39586	0.042
	6_3	709.05	1716.13	0.413168		
12 hpt	12_1	613.12	1737.93	0.352787512		
	12_2	650.59	1704.18	0.381761316	0.38056	0.024
	12_3	721.83	1772.95	0.407135001		
24 hpt	24_4	650.59	1348.88	0.482318664		
	24_5	613.12	1434.32	0.427463885	0.43176	0.043
	24_6	583.13	1512.67	0.385497167		
48 hpt	48_1	650.59	1585.48	0.410342609		
	48_2	680.43	1588.73	0.428285486	0.4195	0.008
	48_3	691.29	1646.38	0.419884838		

Table A3 Raw materials and relative expression level data of ERp57 in hemocytes of thermal stress treated juvenile *P. monodon* using semi-quantitative RT-PCR

| Sample Group | | Densities of band | | Ratio of gene/EF-1α | Average | STD |
		PMERp57 Density	EF-1α			
Control	N-1	646.66	1676.86	0.385637441		
	N-2	641.77	1722.96	0.372481079	0.3845	0.01
	N-3	676.77	1711.88	0.39533729		
0 hpt	0-1	876.93	1465.47	0.598395054		
	0-2	930.37	1536.66	0.605449481	0.6096	0.01
	0-3	936.844	1499.2	0.624895945		
1 hpt	1_1	689.23	1781.48	0.386886185		
	1_2	683.57	1759.33	0.388539956	0.3919	0.01
	1_3	690.54	1724.7	0.400382675		
3 hpt	3_1	605.37	1788.39	0.338499992		
	3_2	603.12	1752.56	0.344136577	0.3452	0.01
	3_3	612.01	1733.99	0.352948979		
6 hpt	6_1	617.49	1702.1	0.36278127		
	6_2	619.71	1600.87	0.38710826	0.3803	0.01
	6_3	670.88	1716.13	0.390926095		
12 hpt	12_1	609.55	1737.93	0.350733344		
	12_2	622.36	1704.18	0.365196165	0.3536	0.01
	12_3	611.48	1772.95	0.344894103		
24 hpt	24_4	614.87	1348.88	0.455837436		
	24_5	611	1434.32	0.425985833	0.4395	0.01
	24_6	660.4	1512.67	0.436579029		
48 hpt	48_1	548.99	1585.48	0.346261069		
	48_2	551.63	1588.73	0.347214442	0.3495	0
	48_3	584.4	1646.38	0.35496058		

Table A4 Raw materials and relative expression level data of PMCRT in hepatopancrease of thermal stress treated juvenile *P. monodon* using semi-quantitative RT-PCR

Sample Group		Densities of band		Ratio of gene/EF-1α	Average	STD
		PMCRT Density	EF-1α			
Control	N-1	606.85	373.58	1.624417795		
	N-2	571.8	393.21	1.454184787	1.5040449	0.0937
	N-3	588.68	410.65	1.433532205		
0 hpt	0-1	498.57	393.46	1.267142785		
	0-2	470.76	340.03	1.384466077	1.4395659	0.1839
	0-3	605.67	363.31	1.667088712		
1 hpt	1_1	687.11	413.6	1.661291103		
	1_2	721.29	427	1.689203747	1.5889868	0.1342
	1_3	604.76	426.95	1.416465628		
3 hpt	3_1	593.87	351.81	1.688041841		
	3_2	551.65	362.37	1.522339046	1.5439223	0.1204
	3_3	527.32	370.99	1.421386021		
6 hpt	6_1	464.91	384.92	1.207809415		
	6_2	479.29	374.87	1.278549897	1.3583191	0.1812
	6_3	657.06	413.61	1.588597955		
12 hpt	12_1	604.97	448.59	1.348603402		
	12_2	672.76	447.08	1.504786615	1.5581407	0.2153
	12_3	795.7	436.95	1.821032155		
24 hpt	24_4	480.11	381.61	1.258116926		
	24_5	523.47	400.18	1.308086361	1.259966	0.0422
	24_6	489.92	403.66	1.213694694		
48 hpt	48_1	453.36	384.92	1.17780318		
	48_2	515.11	393.46	1.309180095	1.2881553	0.0908
	48_3	541.64	393.21	1.37748277		

Table A5 Raw materials and relative expression level data of PMCNX in hepatopancrease of thermal stress treated juvenile *P. monodon* using semi-quantitative RT-PCR

Sample Group		Densities of band		Ratio of gene/EF-1α	Average	STD
		PMCNX Density	EF-1α			
Control	N-1	273.26	373.58	0.73146314		
	N-2	280.81	393.21	0.714147656	0.7050936	0.02851
	N-3	275	410.65	0.669670035		
0 hpt	0-1	285.6	393.46	0.725867941		
	0-2	265.59	340.03	0.78107814	0.7560197	0.025
	0-3	276.52	363.31	0.761113099		
1 hpt	1_1	312.52	413.6	0.755609284		
	1_2	347.37	427	0.813512881	0.7873574	0.02626
	1_3	338.55	426.95	0.792949994		
3 hpt	3_1	234.38	351.81	0.666211876		
	3_2	321.91	362.37	0.888346166	0.7791634	0.09938
	3_3	290.46	370.99	0.782932155		
6 hpt	6_1	220.57	384.92	0.573028162		
	6_2	220.09	374.87	0.587110198	0.6213864	0.06432
	6_3	291.19	413.61	0.704020696		
12 hpt	12_1	362.48	448.59	0.808042979		
	12_2	304.69	447.08	0.681511139	0.7367057	0.05795
	12_3	314.85	436.95	0.720562993		
24 hpt	24_4	304.22	381.61	0.797201331		
	24_5	351.14	400.18	0.877455145	0.7713602	0.10832
	24_6	258.11	403.66	0.639424268		
48 hpt	48_1	257.2	384.92	0.668190793		
	48_2	262.2	393.46	0.666395568	0.6701107	0.00444
	48_3	265.71	393.21	0.675745785		

Table A6 Raw materials and relative expression level data of ERp57 in hepatopancrease of thermal stress treated juvenile *P. monodon* using semi-quantitative RT-PCR

Sample Group		Densities of band		Ratio of gene/EF-1α	Average	STD
		PMERp57 Density	EF-1α			
Control	N-1	232.66	373.58	0.622784946		
	N-2	244.09	393.21	0.620762442	0.6291917	0.011527
	N-3	264.47	410.65	0.644027761		
0 hpt	0-1	206.18	393.46	0.524017689		
	0-2	243.83	340.03	0.717083787	0.6508533	0.098279
	0-3	258.48	363.31	0.711458534		
1 hpt	1_1	230.88	413.6	0.558220503		
	1_2	283.2	427	0.66323185	0.565936	0.083787
	1_3	203.38	426.95	0.476355545		
3 hpt	3_1	206.8	351.81	0.587817288		
	3_2	252.9	362.37	0.697905456	0.6145092	0.065978
	3_3	206.94	370.99	0.557804793		
6 hpt	6_1	240.31	384.92	0.624311545		
	6_2	256.34	374.87	0.683810388	0.6747277	0.04163
	6_3	296.17	413.61	0.716061024		
12 hpt	12_1	281.48	448.59	0.627477206		
	12_2	269.43	447.08	0.602643822	0.6474436	0.051382
	12_3	311.2	436.95	0.712209635		
24 hpt	24_4	225.38	381.61	0.590602972		
	24_5	236.11	400.18	0.590009496	0.5719467	0.028444
	24_6	216.05	403.66	0.535227667		
48 hpt	48_1	225.1	384.92	0.584796841		
	48_2	220.44	393.46	0.560260255	0.578756	0.014611
	48_3	232.47	393.21	0.591210803		

Table A7 Raw materials and relative expression level data of PMCRT in gill of thermal stress treated juvenile *P. monodon* using semi-quantitative RT-PCR

Sample Group		Densities of band		Ratio of gene/EF-1α	Average	STD
		PMCRT Density	EF-1α			
Control	N-1	480.38	431.61	1.1129955		
	N-2	426.82	481.63	0.8861989	0.9450268	0.1320393
	N-3	417.5	499.47	0.835886		
0 hpt	0-1	490.26	507.45	0.9661247		
	0-2	514.88	538.81	0.9555873	0.9411518	0.0308872
	0-3	509.99	565.56	0.9017434		
1 hpt	1_1	460.48	541.44	0.8504728		
	1_2	434.6	556.64	0.780756	0.9036099	0.139842
	1_3	604.76	560.17	1.0796008		
3 hpt	3_1	447.04	517.6	0.8636785		
	3_2	551.65	527.76	1.0452668	0.9508648	0.0814009
	3_3	500.37	530.25	0.9436492		
6 hpt	6_1	438.72	520.53	0.8428333		
	6_2	427.34	507.47	0.842099	0.8688242	0.040835
	6_3	465	504.59	0.9215403		
12 hpt	12_1	432.08	492.24	0.8777832		
	12_2	471.16	509.73	0.9243325	0.8611494	0.0652366
	12_3	414.95	531.08	0.7813324		
24 hpt	24_4	403.29	454.08	0.8881475		
	24_5	399.84	439.24	0.9102996	0.9066629	0.0151977
	24_6	415.09	450.43	0.9215416		
48 hpt	48_1	391.84	475.3	0.8244056		
	48_2	390.62	478.51	0.8163257	0.8353512	0.0234951
	48_3	403.37	466.15	0.8653223		

Table A8 Raw materials and relative expression level data of PMCNX in gill of thermal stress treated juvenile *P. monodon* using semi-quantitative RT-PCR

Sample Group		Densities of band		Ratio of gene/EF-1α	Average	STD
		PMCNX Density	EF-1α			
Control	N-1	290.83	431.61	0.6738259		
	N-2	335.77	481.63	0.6971534	0.6400738	0.0711271
	N-3	274.33	499.47	0.5492422		
0 hpt	0-1	316.41	507.45	0.6235294		
	0-2	312.5	538.81	0.5799818	0.5900628	0.0265972
	0-3	320.49	565.56	0.5666773		
1 hpt	1_1	289.57	541.44	0.5348146		
	1_2	299.46	556.64	0.5379779	0.5360255	0.0015268
	1_3	299.85	560.17	0.5352839		
3 hpt	3_1	236.53	517.6	0.4569745		
	3_2	254.53	527.76	0.4822836	0.5371942	0.1052816
	3_3	356.5	530.25	0.6723244		
6 hpt	6_1	253.12	520.53	0.4862736		
	6_2	212.49	507.47	0.4187243	0.5299437	0.1237216
	6_3	345.56	504.59	0.6848332		
12 hpt	12_1	283.02	492.24	0.5749634		
	12_2	251.8	509.73	0.493987	0.5609501	0.0547142
	12_3	326.03	531.08	0.6139		
24 hpt	24_4	229.22	454.08	0.5048009		
	24_5	236.43	439.24	0.5382706	0.5399513	0.0322173
	24_6	259.8	450.43	0.5767822		
48 hpt	48_1	288.54	475.3	0.6070692		
	48_2	293.07	478.51	0.6124637	0.6078982	0.0037679
	48_3	281.63	466.15	0.6041618		

Table A9 Raw materials and relative expression level data of ERp57 in gill of thermal stress treated juvenile *P. monodon* using semi-quantitative RT-PCR

Sample Group		Densities of band		Ratio of gene/EF-1α	Average	STD
		PMERp57 Density	EF-1α			
Control	N-1	298.86	431.61	0.6924307		
	N-2	264	481.63	0.5481386	0.6020563	0.0704
	N-3	282.5	499.47	0.5655995		
0 hpt	0-1	269.43	507.45	0.5309489		
	0-2	345.77	538.81	0.641729	0.5970953	0.0523
	0-3	349.86	565.56	0.6186081		
1 hpt	1_1	333.27	541.44	0.6155253		
	1_2	313.77	556.64	0.5636857	0.5825426	0.0256
	1_3	318.41	560.17	0.5684167		
3 hpt	3_1	294.38	517.6	0.5687403		
	3_2	294.21	527.76	0.5574693	0.5611081	0.0059
	3_3	295.41	530.25	0.5571146		
6 hpt	6_1	256.54	520.53	0.4928438		
	6_2	333.27	507.47	0.6567285	0.5485329	0.0838
	6_3	250.29	504.59	0.4960265		
12 hpt	12_1	229.97	492.24	0.4671908		
	12_2	269.43	509.73	0.528574	0.5272468	0.0531
	12_3	311.2	531.08	0.5859757		
24 hpt	24_4	225.38	454.08	0.4963443		
	24_5	236.11	439.24	0.5375421	0.5045131	0.0267
	24_6	216.05	450.43	0.4796528		
48 hpt	48_1	311.2	475.3	0.6547444		
	48_2	220.44	478.51	0.46068	0.5380422	0.092
	48_3	232.47	466.15	0.4987021		

Appendix B

Table B1 Relative expression level data of PMCRT in hemocytes of thermal stress treated juvenile *P. monodon* using real time-PCR

Sample Group		Concentration		Ratio of gene/EF-1α	Average	STD
		PMCRT Density	EF-1α			
Control	N-1	3195	30650	0.104241436		
	N-2	2995	25250	0.118613861	0.103821	0.015
	N-3	3150	35550	0.088607595		
0 hpt	0-1	96200	28400	3.387323944		
	0-2	88300	23550	3.749469214	3.3626241	0.3998
	0-3	102550	34750	2.951079137		
1 hpt	1_1	73250	21250	3.447058824		
	1_2	69600	19900	3.497487437	3.355436	0.2039
	1_3	60250	19300	3.121761658		
3 hpt	3_1	21950	35650	0.615708275		
	3_2	22200	33500	0.662686567	0.7425047	0.1805
	3_3	24250	25550	0.949119374		
6 hpt	6_1	16900	34600	0.488439306		
	6_2	15850	26700	0.593632959	0.625803	0.156
	6_3	15350	19300	0.795336788		
12 hpt	12_1	20150	35900	0.561281337		
	12_2	19500	19000	1.026315789	0.8005911	0.2328
	12_3	21250	26100	0.814176245		
24 hpt	24_4	4755	27750	0.171351351		
	24_5	3920	25300	0.154940711	0.1547742	0.0167
	24_6	4065	29450	0.13803056		
48 hpt	48_1	3140	18600	0.168817204		
	48_2	2965	19450	0.152442159	0.1568662	0.0105
	48_3	3390	22700	0.149339207		

Table B2 Relative expression level data of PMCNX in hemocytes of thermal stress treated juvenile *P. monodon* using real time -PCR

Sample Group		Concentration		Ratio of gene/EF-1α	Average	STD
		PMCNX	EF-1α			
Control	N-1	1930	30650	0.062969		
	N-2	2320	25250	0.0918812	0.07135419	0.0179
	N-3	2105	35550	0.0592124		
0 hpt	0-1	18850	28400	0.6637324		
	0-2	18250	23550	0.7749469	0.657977038	0.12
	0-3	18600	34750	0.5352518		
1 hpt	1_1	7105	21250	0.3343529		
	1_2	6495	19900	0.3263819	0.33173027	0.0046
	1_3	6455	19300	0.334456		
3 hpt	3_1	5895	35650	0.1653576		
	3_2	5770	33500	0.1722388	0.188592163	0.0345
	3_3	5830	25550	0.22818		
6 hpt	6_1	5540	34600	0.1601156		
	6_2	5450	26700	0.2041199	0.223052579	0.0742
	6_3	5885	19300	0.3049223		
12 hpt	12_1	5845	35900	0.1628134		
	12_2	5660	19000	0.2978947	0.224322881	0.0683
	12_3	5540	26100	0.2122605		
24 hpt	24_4	917.5	27750	0.0330631		
	24_5	890.5	25300	0.0351976	0.033783558	0.0012
	24_6	974.5	29450	0.03309		
48 hpt	48_1	730	18600	0.0392473		
	48_2	876.5	19450	0.0450643	0.038434256	0.0071
	48_3	703.5	22700	0.0309912		

Table B3 Relative expression level data of ERp57 in hemocytes of thermal stress treated juvenile *P. monodon* using real time RT-PCR

Sample Group		Concentration		Ratio of gene/EF-1α	Average	STD
		PMERp57	EF-1α			
Control	N-1	1.82E+04	30650	0.592169657		
	N-2	1.22E+04	25250	0.481188119	0.481555264	0.1104
	N-3	1.32E+04	35550	0.371308017		
0 hpt	0-1	5.23E+05	28400	18.39788732		
	0-2	4.94E+05	23550	20.97664544	18.32388502	2.6905
	0-3	5.42E+05	34750	15.5971223		
1 hpt	1_1	2.56E+05	21250	12.02352941		
	1_2	2.37E+05	19900	11.90954774	11.81188927	0.2739
	1_3	2.22E+05	19300	11.50259067		
3 hpt	3_1	2.74E+05	35650	7.671809257		
	3_2	269500	33500	8.044776119	9.048385602	2.0697
	3_3	292000	25550	11.42857143		
6 hpt	6_1	100400	34600	2.901734104		
	6_2	100750	26700	3.77340824	4.254408415	1.6467
	6_3	117500	19300	6.088082902		
12 hpt	12_1	18100	35900	0.504178273		
	12_2	17600	19000	0.926315789	0.731620626	0.213
	12_3	19950	26100	0.764367816		
24 hpt	24_4	16950	27750	0.610810811		
	24_5	16150	25300	0.638339921	0.594651829	0.0536
	24_6	15750	29450	0.534804754		
48 hpt	48_1	18300	18600	0.983870968		
	48_2	17750	19450	0.912596401	0.857559608	0.161
	48_3	15350	22700	0.676211454		

Restriction mapping of pGEM® T-easy Vector

Xmn I 2009

Sca I 1890

Nae I 2707

f1 ori

Amp^r

pGEM®-T Easy
Vector
(3015bp)

lacZ

ori

T7 ↓	1 start
Apa I	14
Aat II	20
Sph I	26
BstZ I	31
Nco I	37
BstZ I	43
Not I	43
Sac II	49
EcoR I	52
Spe I	64
EcoR I	70
Not I	77
BstZ I	77
Pst I	88
Sal I	90
Nde I	97
Sac I	109
BstX I	118
Nsi I	127
	141
↑ SP6	

T7 Transcription Start

```
5'...TGTAA TACGA CTCAC TATAG GGCGA ATTGG GCCCG ACGTC GCATG CTCCC GGCCG CCATG
3'...ACATT ATGCT GAGTG ATATC CCGCT TAACC CGGGC TGCAG CGTAC GAGGG CCGGC GGTAC
        T7 Promoter                        Apa I   Aat II  Sph I     BstZ I   Nco I
```

```
GCGGC CGCGG GAATT CGATT 3'                 ATCAC TAGTG AATTC GCGGC CGCCT GCAGG TCGAC
CGCCG GCGCC CTTAA GCTA  (cloned insert) 3' TTAGTG ATCAC TTAAG CGCCG GCGGA CGTCC AGCTG
   Not I                                    Spe I   EcoR I      Not I     Pst I   Sal I
      Sac II  EcoR I                                        BstZ I
   BstZ I
```

SP6 Transcription Start

```
CATAT GGGA GAGCT CCCAA CGCGT TGGAT GCATA GCTTG AGTAT TCTAT AGTGT CACCT AAAT ...3'
GTATA CCCT CTCGA GGGTT GCGCA ACCTA CGTAT CGAAC TCATA AGATA TCACA GTGGA TTTA ...5'
                                                            SP6 Promoter
   Nde I      Sac I       BstX I       Nsi I
```

Restriction mapping of pET 15b

pET-15b sequence landmarks

T7 promoter	463-479
T7 transcription start	452
His•Tag coding sequence	362-380
Multiple cloning sites	
(*Nde*1 - *Bam*H I)	319-335
T7 terminator	213-259
lacI coding sequence	(866-1945)
pBR322 origin	3882
bla coding sequence	4643-5500

pET-15b cloning/expression region

Restriction mapping of pGEX 4T-1

pGEX-4T-1 (27-4580-01)

Appendix D

Mass spectrometric analysis of full length protein CRT

```
gi|000213038  OV-N-N01-1242-W calreticulin [Galleria mellonella] frame +2
gi|000213092  OV-N-N01-1251-W calreticulin [Galleria mellonella] frame +2
gi|000224869  OV-N-S01-1559-W calreticulin [Galleria mellonella] frame +1
gi|000083215  HC-N-N01-1081-LF PREDICTED: similar to calreticulin [Apis mellifera] fr
gi|000227491  OV-N-S01-2066-W PREDICTED: similar to calreticulin [Apis mellifera] fra
gi|000236012  TT-N-S01-0724-W PREDICTED: similar to hopscotch CG1594-PA [Apis mellife
gi|000063700  HPO-N-S01-0060-LF Penaeus monodon, complete mitochondrial genome frame
gi|000063934  HPO-N-S01-0112-LF frame -1
gi|000069544  HC-H-S01-0136-LF C-type lectin [Penaeus monodon] frame -1
gi|000077410  HC-N-N01-0637-LF frame -1
gi|000087916  HC-N-N01-11653-LF frame -1
gi|000132056  HC-N-N01-8143-LF frame +2
gi|000151043  HPA-N-N01-0163-LF Penaeus monodon, complete mitochondrial genome frame
gi|000054484  GlEp-N-S01-0503-LF NADH dehydrogenase subunit 5 [Penaeus monodon] frame
gi|000146335  HC-V-S01-0310-LF frame +1
gi|000023269  GL-H-S01-1057-LF PREDICTED: similar to Y43E12A.2 [Tribolium castaneum]
gi|000235076  TT-N-S01-0506-W ICP4 protein [Gallid herpesvirus 3] frame +2
gi|000031277  GL-N-STH01-0106-LF unknown [Schistosoma japonicum] frame -2
gi|000163372  HPa-N-N03-1063-LF conserved hypothetical protein [Aedes aegypti] frame
gi|000109045  HC-N-N01-4028-LF frame +1
```

Mass spectrometric analysis of mature CRT

```
gi|000091748   HC-N-N01-12532-LF calreticulin [Anopheles albimanus] frame +2
gi|000218638   OV-N-S01-0416-W PREDICTED: similar to CG18572-PA, isoform A [Triboliun
gi|000056284   GlEp-N-S01-0821-LF frame -1
gi|000114671   HC-N-N01-5048-LF frame -2
gi|000037784   GlEp-N-N01-0430-LF frame +2
gi|000168082   HPa-N-N03-1967-LF mitochondrial ribosomal protein, L9, putative [Aedes
gi|000011077   ES-N-S02-0264-W frame +1
gi|000083308   HC-N-N01-10827-LF frame -1
gi|000103873   HC-N-N01-3076-LF frame +1
gi|000239248   TT-N-ST02-0063-LF PREDICTED: hypothetical protein [Rattus norvegicus]
gi|000030059   GL-N-STC02-0063-LF frame -2
gi|0000005374  AG-N-N01-1221-W ENSANGP00000014086 [Anopheles gambiae str. PEST] frame
gi|000192946   LP-N-N01-0800-LF frame -1
gi|000079968   HC-N-N01-10156-LF ENSANGP00000023397 [Anopheles gambiae str. PEST] fra
gi|000115715   HC-N-N01-5233-LF frame -2
gi|000197184   LP-N-S01-0384-LF frame -3
gi|000206303   OV-N-N01-0088-W PREDICTED: similar to chromosome 21 open reading frame
gi|000100133   HC-N-N01-14204-LF frame -2
gi|000103675   HC-N-N01-3037-LF transfer RNA-Glu-Pro aminoacyl synthetase frame +1
gi|000208659   OV-N-N01-0491-W frame +3
```

Mass spectrometric analysis of mature CNX

```
gi|000058254   GlEp-N-S01-1169-LF frame -3
gi|000083124   HC-N-N01-10795-LF SJCHGC01974 protein [Schistosoma japonicum] frame -3
gi|000131885   HC-N-N01-8110-LF PREDICTED: similar to hypothetical protein MGC4293 [F
gi|0000001197  AG-N-N01-0275-W G protein s alpha subunit [Litopenaeus vannamei] frame
gi|000151364   HPA-N-N01-0230-LF Penaeus monodon, complete mitochondrial genome frame
gi|000157664   HPa-N-N02-0754-LF Penaeus monodon, complete mitochondrial genome frame
gi|000056680   GlEp-N-S01-0895-LF Malic enzyme, NAD binding domain containing protein
gi|000059457   GlEp-N-S01-1410-LF frame +3
gi|000128122   HC-N-N01-7435-LF hypothetical protein LOC562615 [Danio rerio] frame -1
gi|000111662   HC-N-N01-4512-LF frame +2
gi|000028201   GL-N-S01-1131-LF unnamed protein product [Kluyveromyces lactis] frame
gi|000191497   LP-N-N01-0514-LF frame +1
gi|000023089   GL-H-S01-1022-LF frame +1
gi|000158120   HPa-N-N03-0047-LF Litopenaeus vannamei high density lipoprotein/1,3-be
gi|000015070   ES-N-S03-0553-W frame -1
gi|000126885   HC-N-N01-7208-LF frame +3
gi|0000006009  BT-N-S01-0039-W frame +3
gi|000214741   OV-N-N01-1543-W PREDICTED: similar to CG10077-PA, isoform A, partial [
gi|000048149   GlEp-N-N01-2251-LF frame -2
gi|000041369   GlEp-N-N01-1099-LF frame -2
```

Mass spectrometric analysis of CRT-ERp57 complex

```
Database      : GI_Pmonodon  (240000 sequences; 46497254 residues)
Timestamp     : 16 Mar 2009 at 06:44:36 GMT
Protein hits  : gi|000210181  OV-N-N01-0752-W PREDICTED: similar to CG8983-PA, isoform A [Tribolium
                gi|000073449  HC-H-S01-0813-LF PREDICTED: similar to CG8983-PA, isoform A [Tribolium
                gi|000083215  HC-N-N01-1081-LF PREDICTED: similar to calreticulin [Apis mellifera] :
                gi|000176244  HPa-N-N04-1273-LF frame -3
                gi|000055758  GlEp-N-S01-0727-LF unnamed protein product [Tetraodon nigroviridis] f:
                gi|000041162  GlEp-N-N01-1065-LF ENSANGP00000011246 [Anopheles gambiae str. PEST] f:
                gi|000100935  HC-N-N01-2411-LF protein disulfide-isomerase like protein ERp57 [Bomby
                gi|000113545  HC-N-N01-4855-LF SJCHGC09076 protein [Schistosoma japonicum] frame +1
                gi|0000008671 BT-N-S01-0624-W frame +1
                gi|000035232  GL-ST-S02-0452-LF hypothetical protein XF0051 [Xylella fastidiosa 9a5c
```

Mass spectrometric analysis of CNX-ERp57 complex

- PmERp57 in CNX-ERp57 complex analysis

```
Database      : GI_Pmonodon  (240000 sequences; 46497254 residues)
Timestamp     : 16 Mar 2009 at 06:14:38 GMT
Protein hits  : gi|000210181  OV-N-N01-0752-W PREDICTED: similar to CG8983-PA, isoform A [Tribolium
                gi|000073449  HC-H-S01-0813-LF PREDICTED: similar to CG8983-PA, isoform A [Tribolium
                gi|000186374  IN-N-S01-0691-LF protein disulfide isomerase [Aedes aegypti] frame +2
                gi|000234252  TT-N-S01-0338-W ARP1 actin-related protein 1 homolog A, centractin alp
                gi|000034751  GL-ST-S02-0371-LF hypothetical protein eimer1508.tmp20 [Eimeria tenell
                gi|0000003990 AG-N-N01-0906-W frame -3
                gi|000152284  HPA-N-N01-0414-LF Penaeus monodon, complete mitochondrial genome frame
                gi|000131262  HC-N-N01-7993-LF Tubulin alpha-3 chain (Alpha-III tubulin) frame -3
                gi|000232714  TT-N-S01-0030-W programmed cell death protein [Aedes aegypti] frame -1
                gi|000176161  HPa-N-N04-1256-LF unknown [Schistosoma japonicum] frame +1
```

- PmCNX in CNX-ERp57 complex analysis

```
Database      : NCBInr 20070216 (4626804 sequences; 1596079197 residues)
Timestamp     : 16 Mar 2009 at 06:24:03 GMT
Protein hits  : gi|84402     glutathione transferase (EC 2.5.1.18) - fluke (Schistosoma japonicum)
                gi|6671664   calnexin [Mus musculus]
                gi|108868255 calnexin [Aedes aegypti]
                gi|118784733 ENSANGP00000021843 [Anopheles gambiae str. PEST]
                gi|39645929  Protein disulfide isomerase associated 4 [Danio rerio]
                gi|73978405  PREDICTED: similar to Protein disulfide-isomerase A4 precursor (Protein
                gi|18621163  polyprotein [Sugarcane mosaic virus]
                gi|90589387  Beta-lactamase [Flavobacterium johnsoniae UW101]
                gi|3293544   unknown [Streptomyces lividans]
                gi|110636554 preprotein translocase, secA subunit [Cytophaga hutchinsonii ATCC 3340(
```

BIOGRAPHY

Mr. Apiruck Watthanasurorot was born on July 19, 1984 in Nonthaburi. He graduated with the degree of Bachelor of Science from the Department of Zoology (Biology), Kasetsart University in 2005. He has enrolled a Master degree program at the Program in Biotechnology, Chulalongkorn University since 2008.

Publications related with this thesis

1. **Watthanasurorot, A.**, Visudtiphole, V., Klinbunga, S. and Menasveta, P. (2008). Clonign and characterization of cdna and genomic organization of *calreticulin* in the black tiger shrimp *Penaeus monodon*. 34th Congress on Science and Technology of Thailand, 31 October-2 November 2008, Queen Sirikit National Convention Hall, Bangkok, Thailand (Oral presentation).

2. **Watthanasurorot, A.**, Visudtiphole, V., Klinbunga, S. and Menasveta, P. (2008). Cloning and Expression of Gene Encoding ERp57 Protein of the Black Tiger Shrimp *Penaeus monodon.* The 20th Annual Meeting and International conference of the Thai Society for Biotechnology, 14–17 October 2008, Maha Sarakham, Thailand (Poster presentation).