cDNA LIBRARY CONSTRUCTION AND NUCLEOTIDE SEQUENCE CHARACTERIZATION OF SELECTED cDNA CLONES FROM HEMOCYTES OF BLACK TIGER SHRIMP (*PENAEUS MONODON*)

JANTANA WONGSANTICHON

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(MOLECULAR GENETICS-GENETIC ENGINEERING)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2000

ISBN 974-664-391-6

Thesis
entitled

cDNA LIBRARY CONSTRUCTION AND NUCLEOTIDE SEQUENCE CHARACTERIZATION OF SELECTED cDNA CLONES FROM HEMOCYTES OF BLACK TIGER SHRIMP (*PENAEUS MONODON*)

Jantana Wongsantichon
Ms. Jantana Wongsantichon
Candidate

Asst. Prof. Burachai Sonthayanon,
Ph.D.
Major Advisor

Apinunt Udomkit
Asst. Prof. Apinunt Udomkit, Ph.D.
Co-advisor

Sakol Panyim
Prof. Sakol Panyim, Ph.D.
Co-advisor

Prof. Liangchai Limlomwongse, Ph.D.
Dean
Faculty of Graduate Studies

Prapon Wilairat
Prof. Prapon Wilairat, Ph.D.
Chairman
Master of Science Programme in
Molecular Genetics-Genetic engineering
Institute of Molecular Biology and
Genetics

Thesis
entitled

cDNA LIBRARY CONSTRUCTION AND NUCLEOTIDE SEQUENCE CHARACTERIZATION OF SELECTED cDNA CLONES FROM HEMOCYTES OF BLACK TIGER SHRIMP (*PENAEUS MONODON*)

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science (Molecular Genetics-Genetic Engineering)
on
May 31, 2000

Jantana Wongsantichon
Ms. Jantana Wongsantichon
Candidate

Asst. Prof. Burachai Sonthayanon,
Ph.D.
Chairman

Apinunt Udomkit
Asst. Prof. Apinunt Udomkit, Ph.D.
Member

T. Suthipongchai
Asst. Prof. Tuangporn Suthipongchai,
Ph.D.
Member

Sakol Panyim
Prof. Sakol Panyim, Ph.D.
Member

Prof. Liangchai Limlomwongse, Ph.D.
Dean
Faculty of Graduate Studies
Mahidol University

Sakol Panyim
Prof. Sakol Panyim, Ph.D.
Director
Institute of Molecular Biology and
Genetics
Mahidol University

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to Dr. Burachai Sonthayanon, my advisor, for his invaluable advice, guidance, supervision, and encouragement throughout this work. I am really grateful to Dr. Albert J Ketterman and Dr. Sakol Panyim for their kindness, helpful suggestion and discussion.

My sincere appreciation is expressed to Ms. Chanikarn Boonchuoy for training and helping me in DNA sequencing, to Ms. Pongsopee Attasart for technical suggestions, and Mr. Arkhom Sai-ngam for his assistance, concern, and encouragement.

I am particularly indebted to Thailand Research Fund for its financial support through the senior research fellowship to Prof. Sakol Panyim.

Finally, I would like to express my deepest appreciation to my parents and all members in my family for their love, regard, and encouragement throughout my study.

Jantana Wongsantichon

4036147 MBMG/M : MAJOR : MOLECULAR GENETICS-GENETIC ENGINEERING; M. Sc. (MOLECULAR GENETICS-GENETIC ENGINEERING)
KEY WORDS : *PENAEUS MONODON* / cDNA LIBRARY/ ESTs/ SHRIMP / CRUSTACEA

JANTANA WONGSANTICHON: cDNA LIBRARY CONSTRUCTION AND NUCLEOTIDE SEQUENCE CHARACTERIZATION OF SELECTED cDNA CLONES FROM HEMOCYTES OF BLACK TIGER SHRIMP (*PENAEUS MONODON*). THESIS ADVISORS : BURACHAI SONTHAYANON Ph.D., APINUNT UDOMKIT Ph.D., SAKOL PANYIM Ph.D. 110 p. ISBN 974-664-391-6

Shrimp aquaculture are facing devastating viral and bacterial diseases while molecular information regarding crustacean genes in defense mechanism is unavailable. To quickly obtain gene sequences which are expressed in phagocytic cells of the black tiger shrimp, cDNA libraries were constructed and random clones picked up for 5' sequence determination. Directional lambda ZAPII-based cDNA libraries were constructed from mRNA of hemocytes of the black tiger shrimp (*Penaeus monodon*). A few hundred randomly selected plaques were excised into pBluescript phagemids. Each clone was subjected to automated DNA sequencing using dye terminator chemistry. BLAST programs were used to search for sequence similarity with those from online databases. From the first cDNA library, out of 67 ESTs, 32 (48 %) were identified to be of mitochondrial genes, 5 (8 %) were of ribosomal protein genes, 3 (4 %) were of other genes, and 11 (16 %) did not match existing gene data. From the second cDNA library constructed from a larger amount of mRNA, a set of 194 ESTs was determined and analyzed. 123 (64 %) clones showed sequence similarity to previously identified genes and ESTs of many species, 71 (36 %) clones showed no sequence similarity to those of any genes. Frequently found EST identities were mitochondrial genes (14 %), ribosomal proteins (6 %), actin (5 %), and serine proteinase inhibitors (3 %). Several DNA clones for proteins involved in defense mechanism such as serine proteinase inhibitors, and peroxinectin were also found. These two genes were of interest for their involvement in the prophenoloxidase activating system. Some ESTs for protein in signal transduction pathway such as Ran and serine/threonine kinase were also found. In total, over 216 DNA sequences, some complete, have been submitted to Genbank database. The EST data set from our hemocyte cDNA libraries should serve as useful basic information for many fields of biological researches and applications.

4036147 MBMG/M : สาขาวิชา : อณูพันธุศาสตร์และพันธุวิศวกรรมศาสตร์ ; วท.ม. (อณูพันธุศาสตร์และพันธุวิศวกรรมศาสตร์)

จันทนา วงศ์สันติชน : การสร้างห้องสมุด cDNA และการศึกษาคุณลักษณะลำดับรหัสเบสของบรรดา cDNA clone ที่แยกมาได้จากเซลล์เม็ดเลือดของกุ้งกุลาดำ (Penaeus monodon) (cDNA LIBRARY CONSTRUCTION AND NUCLEOTIDE SEQUENCE CHARACTERIZATION OF SELECTED cDNA CLONES FROM HEMOCYTES OF BLACK TIGER SHRIMP, *PENAEUS MONODON*) คณะกรรมการควบคุมวิทยานิพนธ์ : บุรชัย สนธยานนท์, Ph.D., อภินันท์ อุดมกิจ, Ph.D., สกล พันธุ์ยิ้ม, Ph.D. 110 หน้า ISBN 974-664-691-6

การเพาะเลี้ยงกุ้งกำลังเผชิญกับความเสียหายหนักจากโรคที่เกิดจากเชื้อไวรัสและแบคทีเรีย ในขณะที่ข้อมูลของยีนใน Crustacean ที่เกี่ยวกับระบบภูมิคุ้มกันยังไม่มี เพื่อให้เราได้บรรดาลำดับรหัสเบสของยีนที่แสดงออกในเซลล์ที่จับกินสิ่งแปลกปลอมของกุ้งกุลาดำ (*Penaeus monodon*) อย่างรวดเร็ว จึงได้สร้างห้องสมุด cDNA ขึ้น และสุ่มเลือกโคลนเพื่อวิเคราะห์ลำดับรหัสจากปลายด้าน 5' cDNA library ซึ่งอาศัยพาหะ Lambda ZAP II ได้ถูกสร้างขึ้นจาก mRNA ของเม็ดเลือดจากกุ้งกุลาดำ เมื่อทำการคัดเลือก plaques แบบสุ่มและตัดออกมาเป็น pBluescript phagemids โคลนหลายร้อยตัวถูกนำไปหาลำดับรหัสเบสโดยใช้ dye-terminator chemistry ผ่านเครื่องวิเคราะห์อัตโนมัติ เมื่อนำรหัสที่ได้ไปเทียบหาความเหมือนกับลำดับรหัสเบสเท่าที่มีอยู่ในฐานข้อมูล Online โดยใช้โปรแกรม BLAST ชนิดต่างๆ ผลการวิเคราะห์ ESTs จำนวน 67 โคลนจากห้องสมุดแรก พบเอกลักษณ์ดังนี้ คือ Mitochondrial gene 48 %, Ribosomal protein 8 %, ยีนอื่นๆ 4 %, และ 16 % ไม่สามารถระบุได้ ผลการวิเคราะห์ ESTs จำนวน 194 โคลนจากห้องสมุด cDNA ที่สองที่สร้างจาก mRNA ปริมาณสูงกว่า 64 % พบความเหมือนของลำดับรหัสเบสกับยีนหรือ ESTs จากสิ่งมีชีวิตชนิดต่างๆ และ 36 % ไม่ปรากฏความเหมือนกับยีนใดๆ เอกลักษณ์ที่พบบ่อย คือ Mitochondrial gene 14 %, Ribosomal protein 6 %, Actin 5 %, และ Serine proteinase inhibitor 3 % โดยพบโคลนที่เป็นโปรตีนซึ่งเกี่ยวข้องกับระบบภูมิคุ้มกันด้วย เช่น Serine proteinase inhibitors และ Peroxinectin ซึ่งสองยีนนี้มีความน่าสนใจเพราะเกี่ยวข้องใน Prophenoloxidase activating system นอกจากนี้ยังพบรหัสเอกลักษณ์ EST ที่เกี่ยวข้องกับ signal transduction pathway เช่น Ran และ Serine/threonine kinase โดยภาพรวมแล้วข้อมูลรหัสดีเอ็นเอ 216 ชิ้น ได้ถูกส่งเข้าฐานข้อมูล Genbank ข้อมูล EST จากห้องสมุด cDNA ของเราเหล่านี้เชื่อว่าจะเป็นข้อมูลพื้นฐานที่เป็นประโยชน์ต่อการวิจัยทางชีวภาพและการประยุกต์ใช้ในด้านต่างๆ ได้ต่อไป

CONTENTS

CHAPTER

I INTRODUCTION

III MATERIALS AND METHODS

CONTENTS (conts.)

2. Methods

CONTENTS (conts.)

CONTENTS (conts.)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (conts.)

LIST OF ABBREVIATIONS

Amp = ampicillin

bp = base pair

CIP = Calf intestine alkaline phosphatase

cDNA = complementary deoxyribonucleic acid

cps = counts per second

CTAB = cetyl trimethyl ammonium bromide

dATP = Dideoxyadenosine 5'-triphosphate

dCTP = Dideoxycytidine 5'-triphosphate

dGTP = Dideoxyguanosine 5'-triphosphate

dTTP = Dideoxythymidine 5'-triphosphate

dNTPs = dATP, dTTP, dGTP, dCTP

DTT = dithiothreitol

E. coli = *Escherichia coli*

EDTA = ethylene diamine tetraacetic acid

EtBr = ethidium bromide

EtOH = ethanol

IPTG = isopropyl-β-D-thiogalactopyranoside

kb = kilobase (s)

kDa = kilodalton (s)

LB = Luria-Bertani media

mM = millimolar

mRNA = messenger RNA

LIST OF ABBREVIATIONS (conts.)

MOPS = 3-(N-morpholino) propanesulfonic acid

MT = metric tons

µg = microgram

µl = microlitre

ng = nanogram

nmole = nanomole

nt = nucleotide

NBT = nitroblue tetrazolium chloride

OD = optical density

PCR = polymerase chain reaction

pfu = plaque forming unit

RNase = ribonuclease

rRNA = ribosomal RNA

rpm = revolution per minute

SDS = sodium dodecyl sulphate

Tris = tris-(hydroxymethyl)-aminoethane

TEMED = tetramethyl-ethylenediamine

U = unit(s)

v/v = volume/volume

w/v = weight/volume

w/w = weight/weight

CHAPTER I

INTRODUCTION

1. General background of *Penaeus monodon*

Penaeus monodon, known as black tiger shrimp or giant tiger shrimp, is named for its big size (reported length up to 363 millimeters) and dark-colored banded tail. A general morphology, with internal anatomy, is shown in figure 1. This marine organism is native to the Indian Ocean, South China Sea, and west Pacific. It is by far the leading farm-raised marine shrimp species in Asia and Australia. The shrimp belongs to *Penaeidae* family of Decapod crustaceans (1). Abundance of this marine organism's farmings grew very strongly in the late 1980s, boomed by the growing aquaculture industry. A production of the tiger shrimp reached a peak at 726,000 MT in 1994 and steadily declined for several years afterward (2). The global production in 1997 was about 608,000 MT, due to disease problems in the main producing countries. Outbreaks of diseases led to a sudden decline of Chinese production (in 1993), Thai production (in 1996 and 1997), and Ecuadorian production (in 1999). However, the shrimp aquaculture production was expected to increase beyond 800,000 MT level in the near future.

2. Thai shrimp farming

Thailand has been the world's major exporter of frozen shrimps, the major of which is *P. monodon* from some 25,000 farms. However, in 1997, its shipments decreased as a result of viral disease problems. In that year, Thai frozen shrimp exports were only 125,000 MT, about half of the previous year's export. In 1998, the disease problem came under control and Thai exports were back to normal level at around 200,000 MT, generating income at two US billion dollar level. The main importing countries for Thai frozen shrimps have been Japan and the USA (2).

3. Shrimp's diseases

Shrimp aquaculture industry in several countries has significantly expanded, which result in more shrimp catches obtained from the farms rather than from marine fishery. Despite much research on the culturing condition which increase production yield from the ponds, many problems still exist due to the delicately-balanced nature of shrimp culturing condition. A major problem at present is frequent outbreaks of viral and bacterial diseases in many areas around the world. The major disease problems, which cause high mortality and significant economic losses, mostly come from white spot baculovirus (WSBV) and yellow-head virus (YHV) (80, 81).

White spot baculovirus (WSBV) first appeared in East Asia in 1992-3 and rapidly spreaded to several shrimp farming countries in Asia and the Indo-Pacific (3). The common gross clinical signs are white spots, or patches, of various sizes embedded in the carapace. Infected shrimps are lethargic, do not eat, swim slowly on the pond's water surface, before eventually sink to the bottom and die. Moribund animals display a pink to reddish-brown coloration, due to expansion of cuticular chromatophores. Mortality rates of these shrimp populations can reach 100% within 3-10 days, after onset of those signs (4-7).

Yellow-head virus (YHV) was named for the characteristic yellowish discoloration of the head part, called cephalothorax. Symptoms are sudden food consumption, followed by fasting, and eventually, death. This acute disease can cause a 100 % mortality in affected shrimp within 3-5 days after the first occurrence of the clinical signs (8-10).

Horizontal transmission of these viral diseases has been shown to be via water and living carriers, such as wild shrimp (e.g. krill, *Acetes sp.*) and other crustaceans (e.g. crabs), via carcasses of infected shrimps used as feeds (11-12). Thus, the multiple viral reservoirs are able to carry on the infection rounds and rounds and the diseases may persist for a long period under the farming environment. Emergency harvest is required when clinical sign was first noticed, otherwise shrimps in an entire pond would die in a matter of few days.

Other viral diseases in the shrimp's aquaculture are caused by monodon baculovirus (MBV), hepatopancreatic parvovirus (HPV), infectious hypodermal and hematopoietic necrosis virus (IHHNV), lymphoid organ virus (LOV), gill-associated virus (GAV), etc. However, these diseases do not cause high mortality in shrimps although the growth is much retarded (13).

To alleviate the plight of the shrimp's industry, research works have been focused on disease detection methods such as histological observation (14-15), *in situ* hybridization (14, 16-17), and polymerase chain reaction (PCR) (18-19). However, defense mechanism of the shrimp is still not well studied. Molecular basis of diseases is not understood and disease treatment is far from being achieved.

4. Arthropod's immunity

The immune system of invertebrate animals is of innate immune mechanism only. All species studied so far don't seem to have an adaptive immune response. For humoral immunity there was no immunoglobulin antibody either. The innate immune defense is directed against by recognizing foreign materials such as lipopolysaccharides (LPS) or peptidoglycans from bacteria, β-1,3 glucans or mannans from fungi. The proteins that recognize these materials are called pattern recognition proteins (PRP). The arthropod PRPs will work as opsonins, agglutinins, or they can initiate a defense system called prophenoloxidase activating system (proPO-system) (20). The primary group of cells involved are blood cells.

In crustaceans, the circulating blood cells, or hemocytes, can be grouped into three subpopulation : granular, semi-granular, and hyaline cells. The ratio between them is quite different among species. Hyaline calls are considered as phagocytes. Semi-granular cells containing small granules are involved in encapsulation. They react first during an immune response by degranulation. The vesicle contents will be released and stimulate the granular cells to degranulate (21-22). The latter mainly participate in the proPO-system.

The proPO-system is an efficient non-self recognition system in invertebrates (23, 79). It was extensively studied in a freshwater crayfish, *Pacifastacus leniusculus*. Upon activation by microbial cell wall via degranulation of the blood cells, prophenoloxidase activating enzyme (ppA) will become active and convert the zymogen prophenoloxidase (proPO) into active phenoloxidase (PO) through serine proteinase activity. PO will adhere to surfaces of the pathogen, oxidize tyrosines to quinones, and non-enzymatically polymerize insoluble melanin deposits. The reaction is called melanization (20-23). Several ProPOs isolated from several invertebrate animals have molecular masses of 70-80 kDa while the active POs have the masses of 60-70 kDa. The molecular masses of ppA from crayfish, *D. melanogaster*, and *B. mori* are about the same, at 30 kDa (23).

To avoid formation of melanin at inappropriate sites other than the intention sites, proteinase inhibitors are used as the main regulatory factors. Some proteinase inbihitors have been identified in crayfish, such as pacifastin, a Kazal inhibitor, and a protein called serpin (21).

With the activation of proPO, another important component called a cell adhesion molecule is induced to become biologically active. This 76 kDa protein is synthesized in the arthropod blood cells and stored in their granules. It is directly involved in cell to cell communication, by functioning as a cell adhesion molecule. It also involves in cell degranulation and encapsulation. Study of crayfish cDNA revealed that the protein structurally belongs to a family of heme-containing peroxidases. After its peroxidase activity was also found, it was named peroxinectin (24). Within the peroxidase structural domain, there is a putative cell adhesive motif and integrin binding motif 'KGD' close to the C-terminus of the polypeptide sequence (25). The finding is interesting because peroxidase is present in most eukaryotic organisms as a producer of antibacterial factors.

Arthropod immunity has been studied in great detail in *Drosophila*. One aspect widely studied is an induction of antimicrobial peptides and their regulation. These small peptides are important in serving as a broad spectrum antimicrobial agents (26). By 1996, about hundred antimicrobial peptides had been characterized from numerous arthropods. In insects, antimicrobial peptides are predominantly produced in the fat body, which is functional equivalent to mammalian liver. Induction of the antimicrobial synthesis is quite complex, involving several intracellular signaling pathways. The immune system of *Drosophila* can distinguish major types of pathogens (gram-positive bacteria, gram-negative bacteria, and fungi) and can make an effective response to that pathogen group. There are two major classes of antimicrobial peptides : the cyclic peptides with disulfide bridges and the linear polypeptides without cysteines. The most prominent peptides of cyclic antimicrobial peptides are the 4 kDa antibacterial (anti-gram-positive) 'defensins' and the 5 kDa antifungal peptides 'drosomycin'. The linear antimicrobial peptides can be grouped into three families :- cecropins, proline-rich peptides, and glycine-rich peptides (27). A type of gene transcription regulatory sequences called κB-related response elements was found in a promoter region of each immune-inducible antibacterial peptide gene of insects. The consensus sequence of κB-RE is : GGGDNTYYYY, where D = A, T or G (27-28).

A κB element is recognized by a group of is a member Rel-family transcription factors, called NF-κB, which are well-characterized in mammalian and other eukaryotic systems. In *Drosophila*, three members of the Rel family proteins have been identified : Dorsal, Dif (dorsal-related immunity factor), and Relish. (28) In the signal transduction pathway to Dorsal, binding of a ligand Spätzle to a membrane surface receptor 'Toll' can trigger signal transduction through 'Tube' and 'Pelle' proteins. The intracellular signalling leads to the degradation of a inhibitor of κB protein or IκB homologue, Cactus, which is previously complexed with Dorsal (NFκ B-like). The removal of IκB subunit reveals a nuclear localization signal within the 3D structure of Dorsal. After entering the nucleus, the Rel domains of Dorsal bind to various κB-related response elements and thus activates numerous target genes coded for various enzymes, including those of antimicrobial peptides (29-30).

Other works have been done in characterization of lipopolysaccharide binding protein (LPS-BP) (31), shrimp's hemocytes and plasma components (32), characterization of hemolymph clottable proteins (33), induction of antimicrobial peptides (34-35), understanding an anticoagulant with prophenoloxidase system (36), and selection of specific pathogen resistant (SPR) shrimp (37-38).

In spite of much works done incharacterizing a number of defense proteins and the hemocytes, not much molecular mechanism has been studied on this marine organism so far. When compare to other organisms, recent searching of online databases revealed around 3 millions sequence entries of *Homo sapiens*, 1 million entries of *Mus musculus*, 150,000 of *Drosophila melanogaster*, but only 471 nucleotide sequences and 58 protein sequences of *P. monodon* (NCBI; as to May, 2000) (39). Characterizing expressed sequence tags (ESTs) is a practical approach to discover a large collection of putative genes expressed in specific cells and tissues (40-42). The stock of ESTs data would serve primary molecular information, which facilitate researchers in any field of interests to further study them according to their possible roles, including in defense responses and/or signal transduction.

In this thesis, an approach on characterization of ESTs was used to rapidly generate DNA sequence information of *Penaeus monodon*, specifically for those from the hemocytes. This is because hemocytes are phagocytic cells and are involved in elimination of invading pathogens. They are also crucial in the conduction of defense reactions of crustaceans (32, 43-45).

5. Expressed sequence tags (ESTs)

Expressed Sequence Tags (ESTs) are short partial cDNA sequences, few hundred base pairs in length, derived by single pass sequencing of the inserts of randomly selected cDNA clones. These clones can be sequenced at the 5' and 3' end or from both ends, which yield two ESTs for each clone (46-47). The EST derived from 3' end of the insert often aligns with sequence in the 3' untranslated region of the gene, while that from the 5' end of the insert derived from sequence further upstream. Depending on the length of the clone insert, the two ESTs may overlap. Some genes may be represented by only one EST, whereas other, such as human serum albumin, are represented by more than 1,000 (48). One could then identify the identity of genes by a computer search of all existing sequence databases. There are examples of three important uses of the ESTs that are gene identification, gene-based physical-map construction, and the computer-assisted large-scale characterization of the megabases of genomic sequence generated by laboratories participating in the genome project (49). Thus, the database which has a high rate of growth is a division of GenBank called dbEST, or the database of Expressed Sequence Tags (http://www.ncbi.nlm.nih.gov/dbEST). The identification of a cDNA clones was based on a high sequence identity (usually > 35 %) over a relative long range (usually > 200 bp.) (50). Regardless of its benefits, ESTs are subjected to 'expression bias' because the cDNA from which the ESTs are derived are present in libraries in proportion to the abundance of mRNAs in the tissues from which the library was prepared. Thus, genes expressed at very low levels are not likely to be found within EST data set, while abundantly expressed genes are over-represented. One approach to reduce the prevalence of abundant cDNAs is to use a 'subtractive technique' (51-53), in which sequence-tagged cDNAs can be used to identify and remove redundant clones from the cDNA libraries. The number of clones that should be sequenced from an unsubtracted library would be small. Frequency of identification of a new gene from a subtracted library is high and the redundancy is low. So far, no subtraction library for crustacean system has been reported. Recently, researchers in Australia reported on ESTs from mixed-tissues of *P. monodon* (40).

6. Computational analysis

Searching sequence similarity to a known protein is a first clue of a newly sequenced gene. There are a number of software tools for searching sequence databases. All of them use some measure of similarity between sequences to distinguish a biological significant relationship from a chance similarity (54).

BLAST (Basic Local Alignment Search Tool) is a series of sequence similarity search programs designed to explore all of the available DNA and protein sequence databases (54). There are currently five BLAST programs, namely :-

blastp	for comparing an amino acid query sequence against a protein sequence database,
blastn	for comparing a nucleotide query sequence against a nucleotide sequence database,
blastx	for comparing the six-frame conceptual translation products of a nucleotide query sequence (from both strands) against a protein sequence database,
tblastn	for comparing a protein query sequence against a nucleotide sequence database, by dynamically back-translated into all six reading frames,
tblastx	for comparing the six-frame translation of a nucleotide query sequence against the six-frame translation of a nucleotide sequence database.

As of early 2000, WWW access to BLAST currently offers two interfaces, a 'Basic' version with default search parameters and an 'advanced' option, which allows customization of the parameters. The version 2.0 of BLAST allows the introduction of gaps (deletions and insertions) into sequence alignments. This tends to be more biologically meaningful than ungapped results because homologous domains are not broken into several segments (55).

The fundamental unit of BLAST algorithm output is the high-scoring segment pair (HSP). An HSP consists of two sequence fragments of arbitrary but equal length whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold score. A set of HSPs is defined by a scoring system and a cutoff score: this set may be empty if the cutoff score is sufficiently high.

The approach of the BLAST program is first to look for similar segment (HSPs) between the query sequence and a database sequence. Then any matches are evaluated for the statistical significance and finally only those satisfy a user-selectable threshold of significance are reported. The principle equation relating the score of an HSP to its expected frequency of chance occurrence is :

$$E = KN \exp(-\text{lambda } S)$$

where E is the expected frequency of chance occurrence of an HSP having score S; K and Lambda are Karlin-Altschul parameters; N is the product of the query and database sequence lengths, or the size of the search space; and exp is the exponentiation function (56).

After searching for sequence similarity, an alignment between the full-length sequence of a putative gene with the query sequence can be done by using many programs, such as Clustal W and Clustal X (57-58). These programs would reveal overall similarity between new sequence and existing families of sequences.

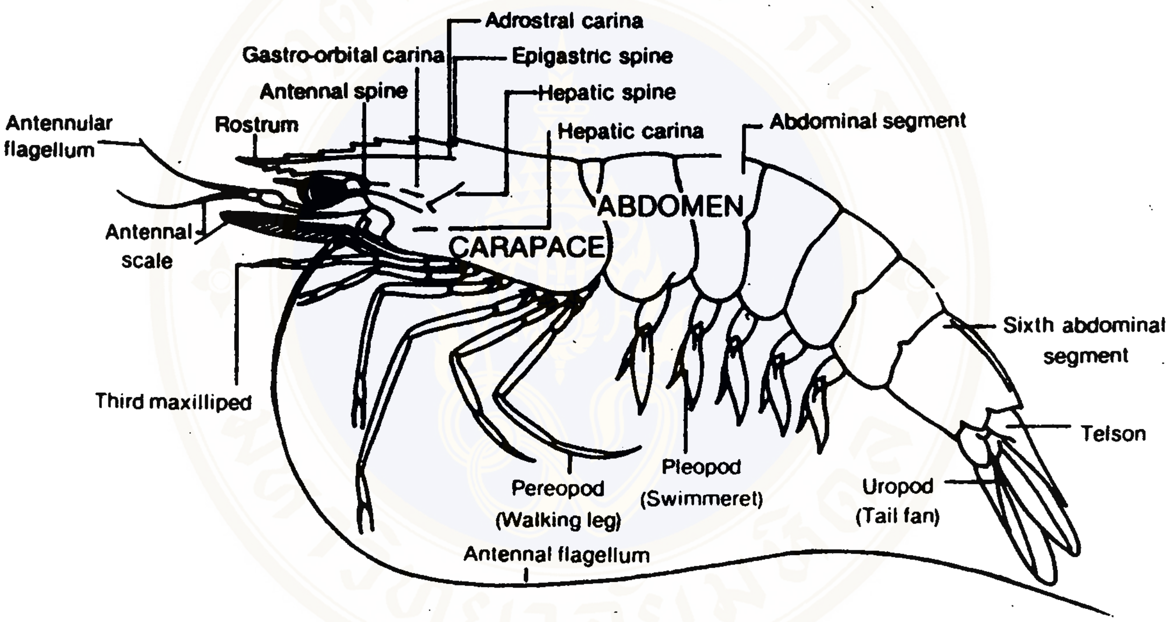


Figure 1 **General morphology of *Penaeus monodon* with external anatomy.** (59)

CHAPTER II

OBJECTIVES

The aim of this thesis was to construct cDNA libraries from hemocytes of *Penaeus monodon*, a main defense cell of the shrimp, and characterize a set of randomly isolated partial cDNA sequences, called expressed sequence tags (5'-ESTs). Information regarding types of putative genes according to their biological cellular roles would be determined.

CHAPTER III

MATERIALS AND METHODS

1. Materials

1.1. Shrimp specimens

Black tiger shrimps *(Penaeus monodon)* were purchased from local shrimp farms. Some were purchased from Shrimp Culture Research Center, Chareon Pokphand Group, Mahachai, Samutsakorn, Thailand. Shrimp hemolymph was collected from live shrimps. It was drawn out from the ventral sinus located at the base of the first abdominal segment of the shrimp with a 23 G gauge needle attached to a 1 ml syringe. About 0.5-1.0 ml was obtained from a 15 g size shrimp. The hemolymph was immediately mixed with Alsever solution in 1:1 ratio and kept at -80 °C before RNA extraction. A total of 85 ml hemolymph was used for RNA preparation for construction of the cDNA library no.1 while 280 ml was used for construction of the library no.2.

1.2. Modified Alsever solution (32)

Alsever solution (AS): 27 mM of sodium citrate; 336 mM of sodium chloride; 115 mM of glucose; 9 mM of EDTA, pH 7.0 was used as an anticoagulant.

1.3. Total RNA isolation

Guanidine thiocyanate was purchased from Sigma and TRIzol® LS reagent from GIBCO BRL, Life Technologies, Inc., USA.

1.4. Poly (A^+) RNA isolation

Oligo(dT)-cellulose was purchased from Pharmacia Biotech, Inc. and PolyATtract® mRNA isolation system IV was purchased from Promega, Inc., USA.

1.5. cDNA library construction

ZAP-cDNA ® Gigapack® III Gold cloning kit was purchased from Stratagene, USA. This kit contains all reagents for a cDNA synthesis, UniZAP™ XR vector (Lambda ZAP® II vector, digested with *EcoR* I and *Xho* I, CIAP treated), *E.coli* XL1-Blue MRF' strain and SOLR™ strain, ExAssist™ interference-resistant helper phage, VCSM13 interference-resistant helper phage, Gigapack® III Gold packaging extracts, λcl857 *Sam7* wild-type lambda control DNA, and VCS257 host strain. The genetic map of the Uni-ZAP XR insertion vector is shown in figure 2.

1.6. Radioactive [α-^{32}P] dATP

[α-^{32}P] dATP (3000 Ci/mmol) was purchased from Amersham Pharmacia Biotech, England.

1.7. Bacterial strains

XL1-Blue MRF' strain : Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173* *endA1* *supE44* *thi-1* *recA1* *gyrA96* *relA1* *lac*[F' *proAB* *lacI*^{q} *Z*Δ*M15* Tn*10*(Tet^{r})] (Stratagene, Inc.)

SOLR strain : e14$^-$ (McrA$^-$)Δ(*mcrCB-hsdSMR-mrr)171 sbcC recB recJ uvrC umuC*::Tn5 (Kan^r) *lac gyrA96 relA1 thi-1 endA1* λR [F' *proAB lacI*^{q}*Z*Δ*M15*] Su$^-$ (nonsuppressing) (Stratagene, Inc.)

Escherichia coli JM109 : [*recA1 supE44 endA1 hsdR17 gyrR17 relA1 thi* Δ (*lac-proAB* F' (*traD36 proAB*$^+$ *lacI*^{q} *lacZ*Δ*M15*)], previously described by Yanisch-Perron *et al.* (60), was employed as a competent host strain for transformation reactions.

1.8. Plasmid vector

pBluescript SK(-) was used as vector in DNA subcloning. The 2958 bp plasmid was derived from pUC19. The SK designation indicates direction of a polylinker in a *lacZ* gene from *Sac* I to *Kpn* I. The circular map and polylinker sequence of pBluescript SK (+/-) phagemid is shown in figure 3 (Stratagene, Inc.).

1.9. Agarose gel DNA purification

GENECLEAN II® kit was purchased from BIO 101 Inc., USA.

1.10. DNA sequencing reaction

ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction kit with Amplitaq® DNA Polymerase FS and ABI PRISM™ BigDye™ Terminator Cycle Sequencing Ready Reaction kit with Amplitaq® DNA Polymerase FS were purchased from PE Applied Biosystems, Inc., USA.

1.11. Enzymes

Lysozyme and RNase A were purchased from Sigma; Klenow (Large fragment of DNA Polymerase I) and T_4 DNA ligase from GIBCO BRL; restriction endonucleases from GIBCO BRL, Promega, and Biolabs. The recognition sites and optimal condition of the restriction enzymes used are shown in Table 1.

Table 1 : Restriction enzymes with their recognition sites, recommended buffers and manufacturers.

Enzymes	Recognition sequence	Buffer	Manufacturer
EcoR I	G^AATTC	Buffer H	Promega
Hind III	A^AGCTT	Buffer E	Promega
Kpn I	GGTAC^C	ReacT 4	GIBCO BRL
Nde I	CA^TATG	NEB buffer 4	Biolab
Pst I	CTGCA^G	NEB buffer 3	Biolab
Xho I	C^TCGAG	Buffer D	Promega

Note

Buffer D: 6 mM Tris-HCl, pH 7.9, 150 mM NaCl, 1 mM DTT, 6 mM $MgCl_2$

Buffer E: 6 mM Tris-HCl, pH 7.4, 100 mM NaCl, 1 mM DTT, 6 mM $MgCl_2$

Buffer H: 90 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM DTT, 10 mM $MgCl_2$

ReacT 4 buffer: 20 mM Tris-HCl, pH 7.4, 50 mM NaCl, 1 mM, 5 mM $MgCl_2$

NEB buffer 3: 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT
(pH 7.9 at 25°C)

NEB buffer 4: 20 mM Tris-HCl, 10 mM Magnesium acetate, 50 mM potassium acetate, 1 mM DTT (pH 7.9 at 25°C)

^ represents the cleavage site of each restriction enzyme.

1.12. PCR primers

M13 (–20) primer	5' GTAAAACGACGGCCAGT 3'
M13 reverse primer	5' GGAAACAGCTATGACCATG 3'
T_7 20-mer primer	5' GTAATACGACTCACTATAGGGC 3'
T_3 22-mer primer	5' AATTAACCCTCACTAAAGGG 3'

1.13. Growth media (61)

LB broth: 1 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 1 % (w/v) NaCl, pH 7.0

LB agar: LB broth with 2 % (w/v) agar

LB top agar: LB broth with 0.7 % (w/v) agarose

NZY broth: 0.5 % (w/v) NaCl, 0.2 % (w/v) $MgSO_4.7H_2O$, 0.5 % (w/v) yeast extract, 1 % (w/v) NZ amine (casein hydrolysate), pH 7.5

NZY agar: NZY broth with 1.5 % (w/v) agar

NZY top agar: NZY broth with 0.7 % (w/v) agarose

Supplemented media: LB broth with 0.2 % (w/v) maltose-10 mM $MgSO_4$

1.14. Chemicals

Cetyl trimethyl ammonium bromide (CTAB)	Sigma
Isopropyl-1-thio-β-D-thio-galactopyranoside (IPTG)	Sigma

5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal)	Sigma
N, N, N', N'-tetramethylenediamine (TEMED)	Fluka
Ethylene diamine tetraacetic acid (EDTA)	Sigma
Water-saturated phenol	Amresco

All other chemicals used in this work were purchased from many suppliers. They were either of analytical or molecular biology grades.

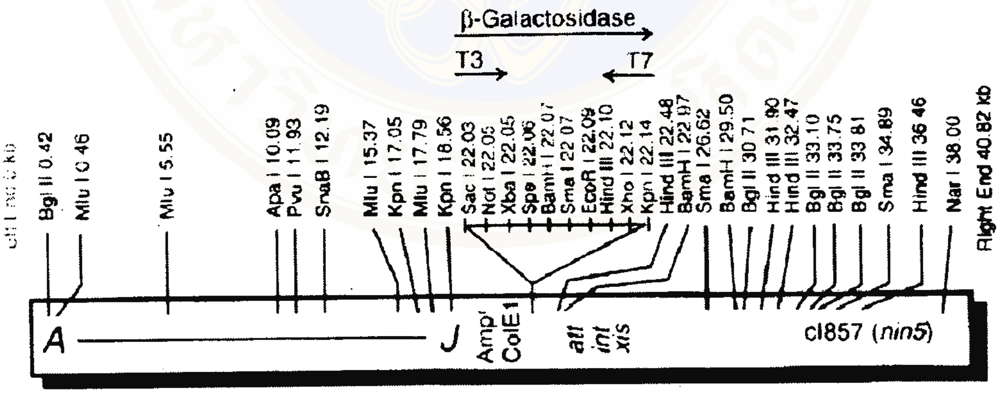


Figure 2 The genetic map of the Uni-ZAP XR insertion vector. Taken from (61).

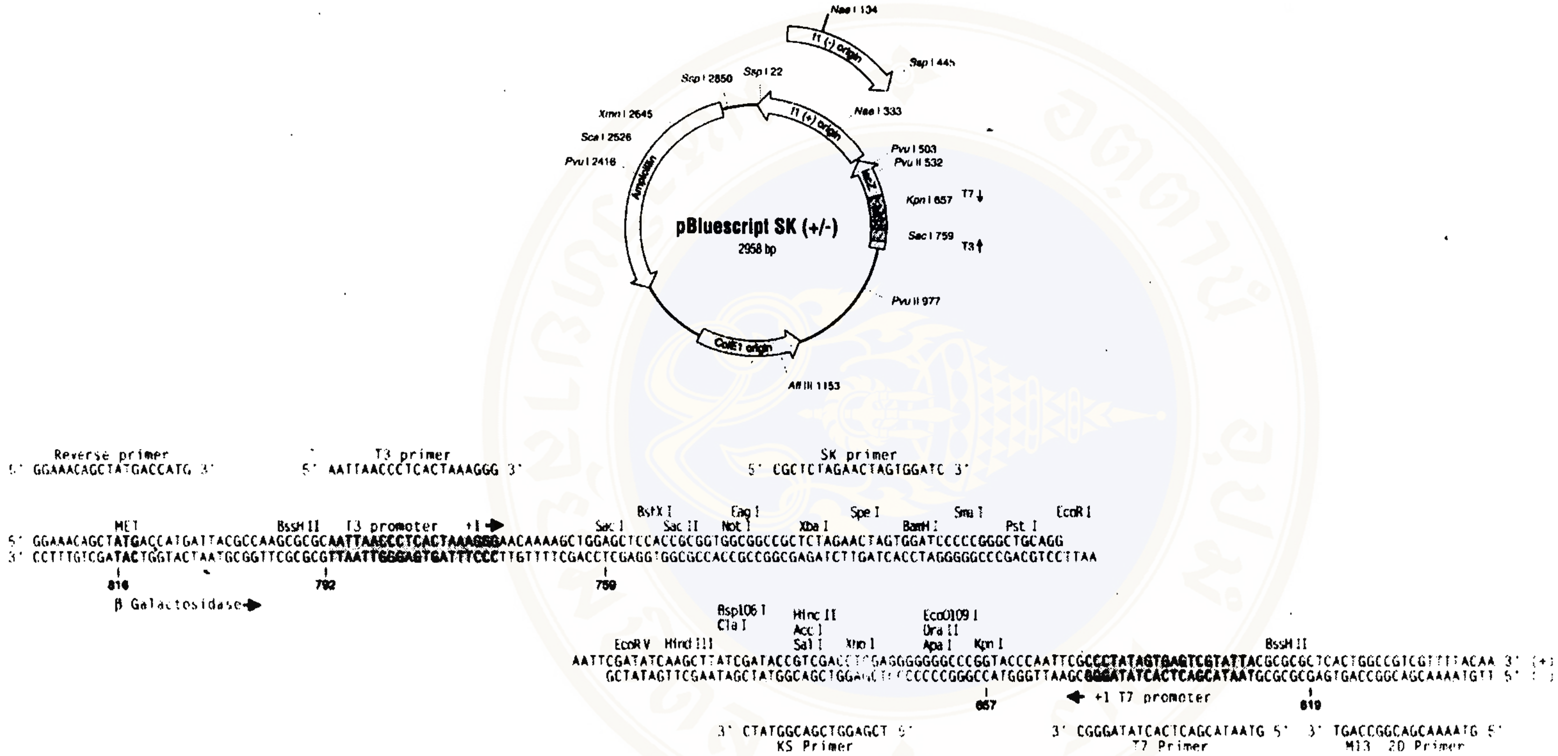


Figure 3 The circular map and polylinker sequence of pBluescript SK (+/-) phagemid. Taken from Stratagene® catalog (61)

2. Methods

2.1 Hemocyte preparation

Hemolymph was collected by using a 23 G gauge needle and a 1 ml syringe, half diluted with a precooled (4 °C) modified AS, pH 7.0 and was centrifuged at 700 x g for 10 min at 4 °C (2,600 rpm in a BECKMAN Avanti™ 30 centrifuge). The supernatant was removed and the cells were immediately extracted for total RNA.

2.2 Total RNA isolation

Method 1: Total RNA isolation by TRIZOL™ LS reagent was used for isolation of total RNA from whole hemolymphs. This procedure was used for construction of the first cDNA library.

A 0.25 ml aliquot of liquid nitrogen-pulverized hemolymph was added with 0.75 ml of TRIzol LS reagent. Cells in the sample suspension were lysed by using a power homogenizer (RW 20 DZM, JANKE & KUNKEL IKA® Labortechnik). The homogenized sample was incubated for 5 minutes at room temperature to permit a complete dissociation of nucleoprotein complexes. After 0.2 ml of chloroform was added, the tube was shaken vigorously by hand for 15 second and incubated for 2-15 minutes at room temperature. The sample was centrifuged at 12,000 x g for 15 min at 4 °C. An aqueous phase was transferred to a new tube and precipitated by mixing with 0.5 ml of isopropanol. The tube was incubated for 10 minutes at room temperature and centrifuged at 12,000 x g for 10 min at 4 °C. A gel-like pellet of RNA precipitate was formed on the side and bottom of the tube. The supernatant was removed and RNA pellet was washed with 1 ml of 75 % ethanol. The sample was mixed by vortexing and centrifuging at 7,500 x g for 5 min at 4 °C. The RNA pellet was air-dried for 5 min and redissolved in DEPC-treated water by passing through a pipette tip and incubating for 10 min at 55 °C.

Method 2: Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction (62), with some modifications, was used for isolation of total RNA from pelleted hemocytes for the construction of the second cDNA library.

Pelleted hemocytes from a 1 ml freshly collected hemolymph was resuspended and lysed in 1 ml of GTC buffer (4.0 M guanidine thiocyanate, 25 mM sodium citrate pH 7.0, 0.5% sarcosyl, 0.1 M 2-mercaptoethanol). A recommended cell concentration was 10^7 cells/ml of GTC (62) whereas a total hemocyte counts of *Penaeus monodon* was reported by around $2x10^7$ cells/ml (63). The cell lysate was sheared 6 times through a 21 G gauge needle. Then, the lysate was added with 0.1 volume of 2 M sodium acetate pH 4.0, 1 volume of water-saturated phenol and 0.3 volume of chloroform-isoamyl alcohol (49:1), followed by mixing for 10-20 sec and incubation on ice for 15 min. The tube was centrifuged at 12,000 rpm for 30 min at 4 °C. The top aqueous phase was transferred to a new tube and an equal volume of isopropanol was added, mixed, and precipitated at –20 °C for 2 hours. The tube was centrifuged at 12,000 rpm for 20 min at 4 °C again. The RNA pellet was redissolved in 0.3 volume of the initial volume of GTC and an equal volume of isopropanol was added and mixed. The tube was incubated at –20 °C for 1 hour and centrifuged at 13,000 rpm for 20 min at 4 °C. The RNA pellet was washed with 75 % ethanol by centrifuging at 6000 rpm for 5 min at 4 °C and allowing to air-dry for 10 min after remove the isopropanol. The RNA pellet was redissolved in DEPC-treated water and stored at –80 °C until use.

2.3 Poly (A^+) RNA isolation procedures

Method 1: mRNA for construction of cDNA library no.1 was purified from total RNA by using oligo(dT)-cellulose as described in Maniatis *et al.* (64)

Method 2: PolyATtract® mRNA isolation system. (Figure 4)

About 0.1-1.0 mg of total RNA was added with RNase-free water to a final volume of 500 µl in a sterile RNase-free tube. The tube was placed in a 65 °C heating block for 10 min. A 3 µl aliquot of the Biotinylated-Oligo(dT) Probe and 13 µl of 20x SSC were added to the RNA. The tube was gently shaken and incubated at room temperature until cooled down. Streptavidin-paramagnetic particles (SA-PMPs) were resuspended by gently flicking the bottom of the tube until completely dispersed. Then, it was captured by placing in a magnetic stand until the SA-PMPs had collected at the side of the tube. The supernatant was carefully removed. The SA-PMPs were washed three times with 0.5 x SSC (0.3 ml per wash), each time captured by using the magnetic stand and supernatant was carefully removed. The washed SA-PMPs were resuspended in 0.1 ml of 0.5 x SSC. The entire content of the annealing reaction was added to the tube containing the washed SA-PMPs. The tube was incubated at room temperature for 10 min. The SA-PMPs were captured by using the magnetic stand and the supernatant was carefully removed. The particles were washed four times with 0.1 x SSC (0.3 ml per wash) by gently flicking the bottom of the tube until all the particles were resuspended. After the final wash, the aqueous phase was removed as much as possible. To elute mRNA, the final SA-PMP pellet was resuspended in 0.1

ml of the RNase-free water, the particles were gently resuspended by flicking the tube. The SA-PMPs were magnetically captured and the eluted mRNA aqueous phase was transferred to a sterile, RNase-free tube. The elution step was repeated by resuspending the SA-PMPs in a 0.15 ml of RNase-free water. The capture step was repeated. The eluate was pooled with a previously eluted RNA (0.25 ml total volume).

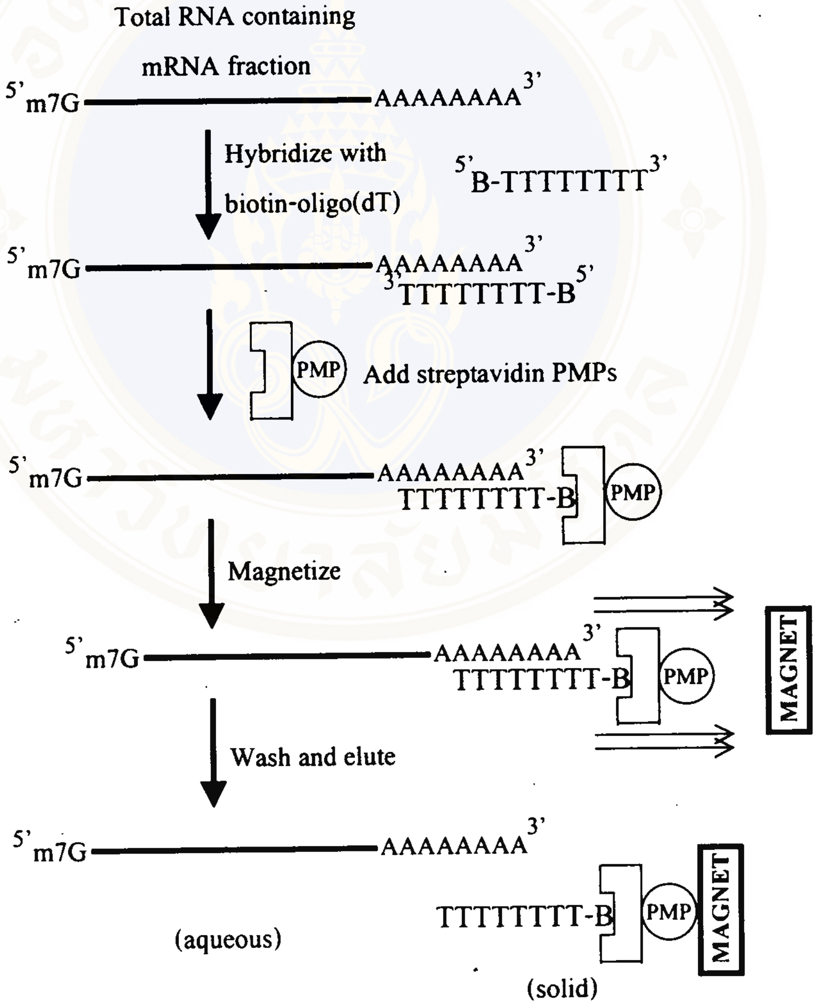


Figure 4 Schematic diagram of the PolyATtract® mRNA isolation procedure, redrawn from (65).

2.4 Formaldehyde agarose gel electrophoresis

For each gel, 1.0 g of agarose was melted in a sterile buffer [10 ml of 10x MOPS buffer (200 mM 3-[N-morpholino]propanesulfonic acid, MOPS; 50 mM sodium acetate; 10 mM EDTA, final pH 6.5-7.0), 85 ml of sterile water] and cooled to around 50 °C. Then, 5.4 ml of 37% (v/v) formaldehyde was added and mixed well by swirling inside a fume hood. A horizontal gel slot (7 cm x 10 cm) was set up in a BioRad chamber. Each total RNA sample (around 3 μg) was dried for 5 min in a vacuum evaporator and then resuspended in 10 μl of formaldehyde gel loading buffer [720 μl of formamide, 160 μl of 10 x MOPS buffer, 260 μl of formaldehyde, 100 μl of sterile water, 100 μl of EtBr (10 mg/ml), 80 μl of sterile glycerol, 80 μl of saturated bromophenol blue in sterile water (small amount of BPB crystal dissolved in water until the presence of an orange pellet was seen)]. The sample was incubated at 65 °C for 15 min and quickly cooled on ice. A 10 μl aliquot of each sample was loaded onto the gel. An electrophoresis was typically conducted at 80 volts for 2 hr in 1x MOPS buffer. The gel was destained with sterile water for 10 min. The pattern of RNAs was visualized under UV light and photographed.

2.5 cDNA library construction

First-strand synthesis: In a RNase-free microcentrifuge tube, 5 μl of 10x first-strand buffer (0.5 M Tris-HCl (pH 8.3), 0.75 M KCl, 0.1 M DDT, and 30 mM $MgCl_2$), 3 μl of first-strand methyl nucleotide mixture (10 mM dATP,dGTP, and dTTP plus 5 mM 5-methyl dCTP), 2 μl of linker-primer (5'-GAGAGAGAGAGAGA

GAGAGAACTAGT**CTCGAG**TTTTTTTTTTTTTTTTTT-3') (1.4 μg/μl) and 1 μl of RNase block ribonuclease inhibitor (40 U/μl) were added and mixed. Then 37.5 μl of mRNA in DEPC-treated water was added and mixed gently. The primer was allowed to anneal for 10 min at room temperature. After that, 1.5 μl of MMLV-RT (50 U/μl) was added. The final volume of 50 μl was mixed gently and spun down. A 5 μl sampling of the first-strand synthesis reaction was transferred to a separated tube containing 0.5 μl of [α-^{32}P]dATP (3000 Ci/mmol) as a first-strand synthesis control reaction. Both tubes were incubated at 37 °C for 1 hour. After 1 hour, the first-strand synthesis reaction was placed on ice while the radioactive first-strand synthesis control was stored at –20 °C.

Second-strand synthesis: To the 45 μl nonradioactive first-strand synthesis reaction on ice, 20 μl of 10x second-strand buffer, 6 μl of second-strand dNTP mixture (10 mM dATP, dGTP, and dTTP plus 26 mM dCTP), 114 μl of sterile distilled water, 2 μl of [α-^{32}P]dATP (3000 Ci/mmol), 2 μl of RNase H (1.5 U/μl) and 11 μl of DNA polymerase I (9.0 U/μl) were added. After gently vortexing and spinning the reaction in a microcentrifuge, the tube was incubated for 2.5 hours at 16 °C. Then the tube was immediately placed on ice.

Blunting of the cDNA termini: 23 μl of blunting dNTP mix (2.5 mM dATP, dGTP, dTTP, and dCTP) and 2 μl of cloned *Pfu* DNA polymerase (2.5 U/μl) (Stratagene, Inc.) were added to the second-strand synthesis reaction and incubated at 72 °C for 30 min. The reaction was removed and 200 μl of phenol-chloroform [1:1 (v/v)] was added and votexed. The tube was spun at maximum speed for 2 min at room temperature. An aqueous layer, containing cDNA, was transferred to a new

tube. An equal volume of chloroform was added and vortexed. The tube was spun at maximum speed for 2 min at room temperature. An aqueous layer was transferred to a new tube. The cDNA was precipitated by adding 20 µl of 3 M sodium acetate and 400 µl of 100% (v/v) ethanol. The reaction was vortexed and precipitated overnight at –20 °C. The microcentrifuge tube was spun at maximum speed for 60 min at 4 °C. The radioactive supernatant was removed and discarded in a radioactive waste container. The pellet was gently washed by adding 500 µl of 70 % (v/v) ethanol. The microcentrifuge tube was spun at maximum speed for 2 min at room temperature. The ethanol was aspirated and the pellet was lyophilized until dry. Then the pellet was resuspended in 9 µl of *Eco*R I adapters (`5′-OH-AATTCGGCACGAG -3′`) and incubated
`3′-GCCGTGCTCp-5′`
at 4 °C for at least 30 min to allow the cDNA to resuspend. 1 µl of this second-strand synthesis reaction was transferred to a separate tube as a control reaction. At this point, the samples of the first- and second-strand synthesis reactions were performed on an alkaline agarose gel as described in section 2.8.

Ligating the *Eco*R I adapters: To the tube containing the blunted cDNA, 1 µl of 10 x ligase buffer (0.5 M Tris-HCL, pH 7.5; 70 mM $MgCl_2$; 10 mM DDT), 1 µl of 10 mM rATP and 1 µl of T4 DNA ligase (4 U/µl) were added. This ligation reaction was spun down and incubated at 4 °C for 2 days, then heat-inactivated by placing the tube in a 70 °C water bath for 30 min.

Phosphorylating the *Eco*R I Ends: After heat inactivation, the ligation reaction was spun for 2 seconds and cooled at room temperature for 5 min. 1 µl of 10x ligase buffer (0.5 M Tris-HCL, pH 7.5; 70 mM $MgCl_2$; 10 mM DDT), 2 µl of 10 mM rATP, 6 µl of sterile water and 1 µl of T4 polynucleotide kinase (10 U/µl) were added. The reaction was incubated for 30 min at 37 °C, then inactivated for 30 min at 70 °C. Condensation in the tube was spun down for 2 seconds. The reaction was equilibrated to room temperature for 5 min.

Digestion with *Xho* I: The reaction was added with 28 µl of *Xho* I buffer supplement and 3 µl of *Xho* I (40 U/µl) then incubated for 1.5 hours at 37 °C. After that, 5 µl of 10 x STE buffer (1 M NaCl, 200 mM Tris-HCl pH 7.5, and 100 mM EDTA) and 125 µl of 100 % (v/v) ethanol were added to the tube and precipitated overnight at –20 °C. Following precipitation, the reaction was spun at maximum speed for 60 min at 4 °C. The supernatant was discarded, the pellet was dried completely and resuspended in 14 µl of 1 x STE buffer. 3.5 µl of column loading dye was added to the sample. The cDNA synthesis scheme was shown in figure 5.

Size Fractionating: The drip-column was prepared as in Stratagene's protocol. Gloves were worn in all assembling steps. The drip column was loaded with a uniform suspension of Sepharose CL-2B gel filtration medium by a Pasteur pipet. The gel was added until ¼ inch below the 'lip of the pipet', the point where the pipet and the syringe were joined. After the drip column was washed by filling the syringe with 10 ml of 1 x STE buffer, the cDNA sample was immediately loaded when around 50 µl of the STE buffer remained above the surface of the resin. Once the sample entered the Sepharose CL-2B gel filtration medium, the connecting tube was gently filled with 3 ml of 1 x STE buffer using a pipettor. A fresh microcentrifuge tube was used to collect each fraction. Three drops per fraction were begun collecting when the leading edge of the dye reached the –0.4 ml gradation on the pipet.

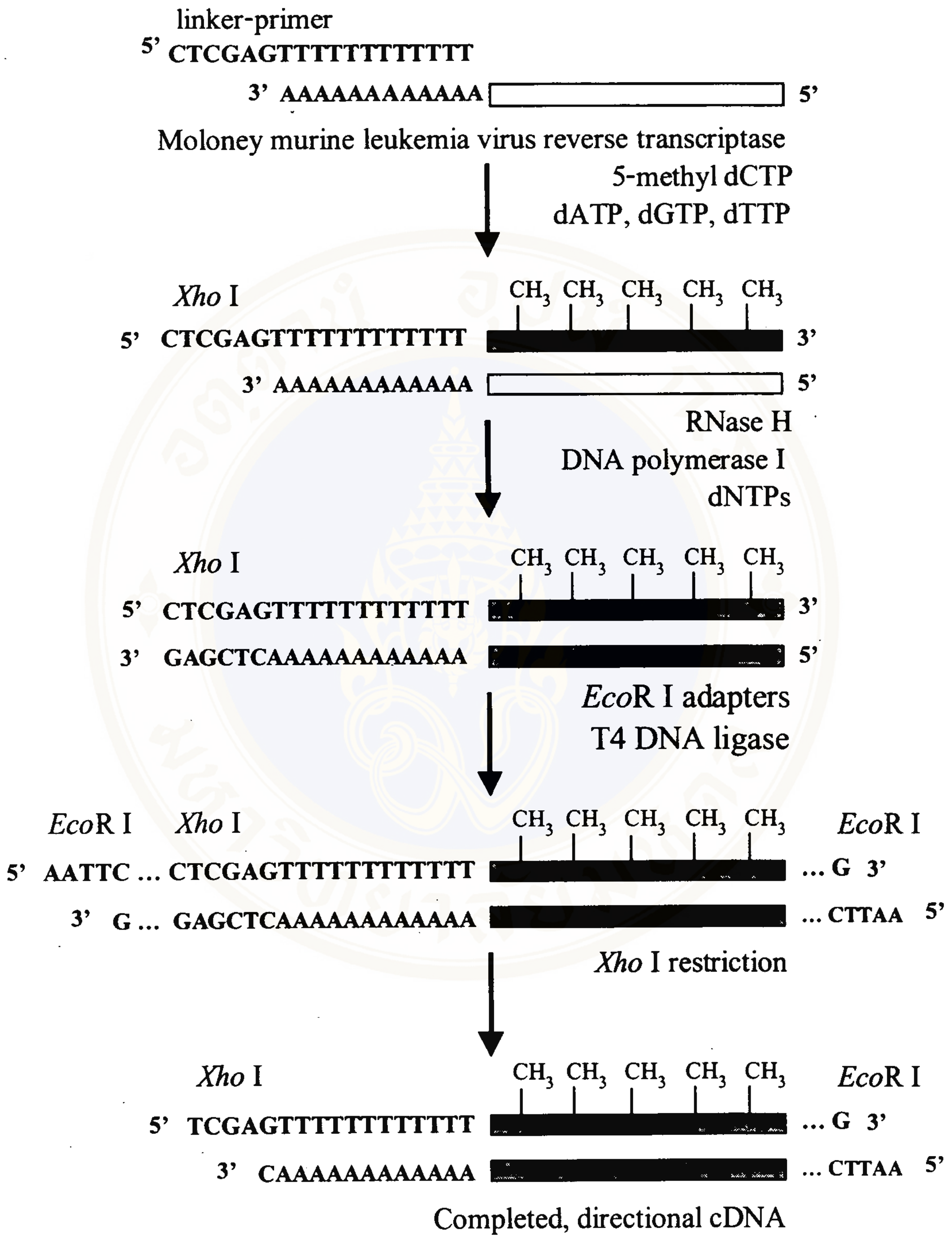


Figure 5 A cDNA synthesis scheme, redrawn from (61)

Fractions were collected until the trailing edge of the dye reached the 0.3 ml gradation. They were monitored for the presence of radioactivity to determine whether the cDNA had eluted successfully. Before processing the fractions and recovering the size-fractionated cDNA, 8 μl of each collected fraction were removed and saved for electrophoresis on 5 % nondenaturing acrylamide gel as described in section 2.9.

Processing the cDNA Fraction: An equal volume of phenol-chloroform [1:1 (v/v)] was added to the remainder of the collected fraction. The tube was vortexed and spun in a microcentrifuge at maximum speed for 2 min at room temperature. The upper aqueous layer was transferred to a fresh microcentrifuge tube. An equal volume of chloroform was added. After vortexing and spinning in a microcentrifuge at maximum speed for 2 min at room temperature, the upper aqueous layer was transferred to a fresh microcentrifuge tube. To each extracted sample, a volume of 100 % (v/v) ethanol in twice the volume of individual sample was added. The tube was precipitated overnight at –20 °C. The sample was spun at maximum speed for 60 min at 4 °C. The supernatant was transferred to another tube. Geiger counter was used to check the level of radioactivity present in the pellet to ensure that the cDNA had been recovered. The pellet was carefully washed with 200 μl of 80 % (v/v) ethanol. The sample was spun at maximum speed for 2 min at room temperature. The ethanol was removed and the pellet was air-dried for 5 min. Using a handheld Geiger counter and the number of counts per second (cps) was detected for each fraction. If <30 cps was detected, each cDNA pellet was resuspended in 3.5 μl of sterile water. If >30 cps was detected, each cDNA pellet was resuspended in 5 μl of sterile water. At this point, Ethidium Bromide plate assay was recommended to quantitate the cDNA before ligating into the Uni-ZAP XR vector.

Ligation of cDNA with the Uni-ZAP XR Vector: To prepare the sample ligation, the following component were added : 2.5 μl of resuspended cDNA (100 ng), 0.5 μl of 10 x ligase buffer, 0.5 μl of 10 mM rATP (pH 7.5), 1.0 μl of the Uni-ZAP XR vector (1 μg/μl), and 0.5 μl of T4 DNA ligase (4U/μl). The reaction was incubated at 4 °C for 2 days.

Packaging: 4 μl (0.1-1.0 μg) of the ligated DNA was immediately added to the packaging extract, which was removed from –80 °C and begun thawing. The tube was stirred, mixed well and quickly spun for 3-5 sec to ensure that all contents were at the bottom of the tube. Then, the tube was incubated at room temperature (22 °C) for 2 hours. A 500 μl of SM buffer (1 litre composed of 5.8 g of NaCl, 2.0 g of $MgSO_4.7H_2O$, 50.0 ml of 2 % (w/v) gelatin, and deionized H_2O to a final volume of 1 litre) and 20 μl of chloroform were added and gently mixed. After briefly spinning the cell debris, the supernatant containing the phage was stored at 4 °C and ready for titering.

Plating and Titering: A glycerol stock of XL1-Blue MRF' strain was streaked onto a LB agar plate containing 12.5 μg/ml tetracyclin. The plate was incubated overnight at 37 °C. A single colony was picked to inoculate a 20 ml of LB-tetracyclin medium, supplemented with 10 mM $MgSO_4$ and 0.2 % (w/v) maltose. The inoculated host was shaken overnight at 37 °C (an OD_{600} not more than 1.0). The cells were spun at 500 x g (2,200 rpm) for 10 min. The supernatant was discarded and the cells were resuspended with sterile 10 mM $MgSO_4$ to an OD_{600} of 0.5. Then 1 μl of lambda phage was added to 200 μl of the diluted host cells. The phage and the bacterium was incubated for 15 min at 37 °C. 3 ml of NZY top agar (melted and cooled to 48 °C), 15 μl of 0.5 M IPTG (in water), and 50 μl of X-gal [250 mg/ml (in DMF)] were added. The top agar was plated immediately onto NZY agar plate and allowed to set for 10 min before inverting and incubating at 37 °C overnight.

2.6 *In vivo* excision

Selected white recombinant plaques were picked from agar plates with large-bore 1000-ml micropipette tips. Each plaque was transferred to a tube containing 500 μl SM buffer and 20 μl chloroform. The tube was vortexed to release the phage particles into the SM buffer, then incubated for 1-2 hours at room temperature or overnight at 4 °C. Overnight cultures of XL1-Blue MRF' and SOLR cells were grown in supplemented LB broth at 30 °C. After spinning down, XL1-Blue MRF' and SOLR cells were resuspended with 10 mM $MgSO_4$ at an OD_{600} of 1.0. To excise a single clone, 100 μl of XL1-Blue MRF' cells at an OD_{600} of 1.0, 100 μl of phage stock (containing > $1x10^5$ phage particles), and 1 μl of ExAssist helper phage (>$1x10^6$ pfu/μl) were combined in a microcentrifuge tube. After incubating at 37 °C for 15 min, 500 μl of LB broth was added. The tube was incubated at 37 °C for 2.5 hours with shaking. Then, the tube was heated at 65-70 °C for 20 min and spun down the debris at 10,000 x g for 15 min. 100 μl of supernatant containing phagemid was mixed with 200 μl of SOLR at an OD_{600} of 1.0. After incubating at 37 °C for 15 min, 10 μl of the mixture was streaked on LB agar containing 50 μg/ml ampicillin. The plate was incubated overnight at 37 °C. Colonies appearing on the plates contained the double strand pBluescript phagemid harboring DNA inserts were later characterized by plasmid DNA extraction and restriction endonuclease digestion.

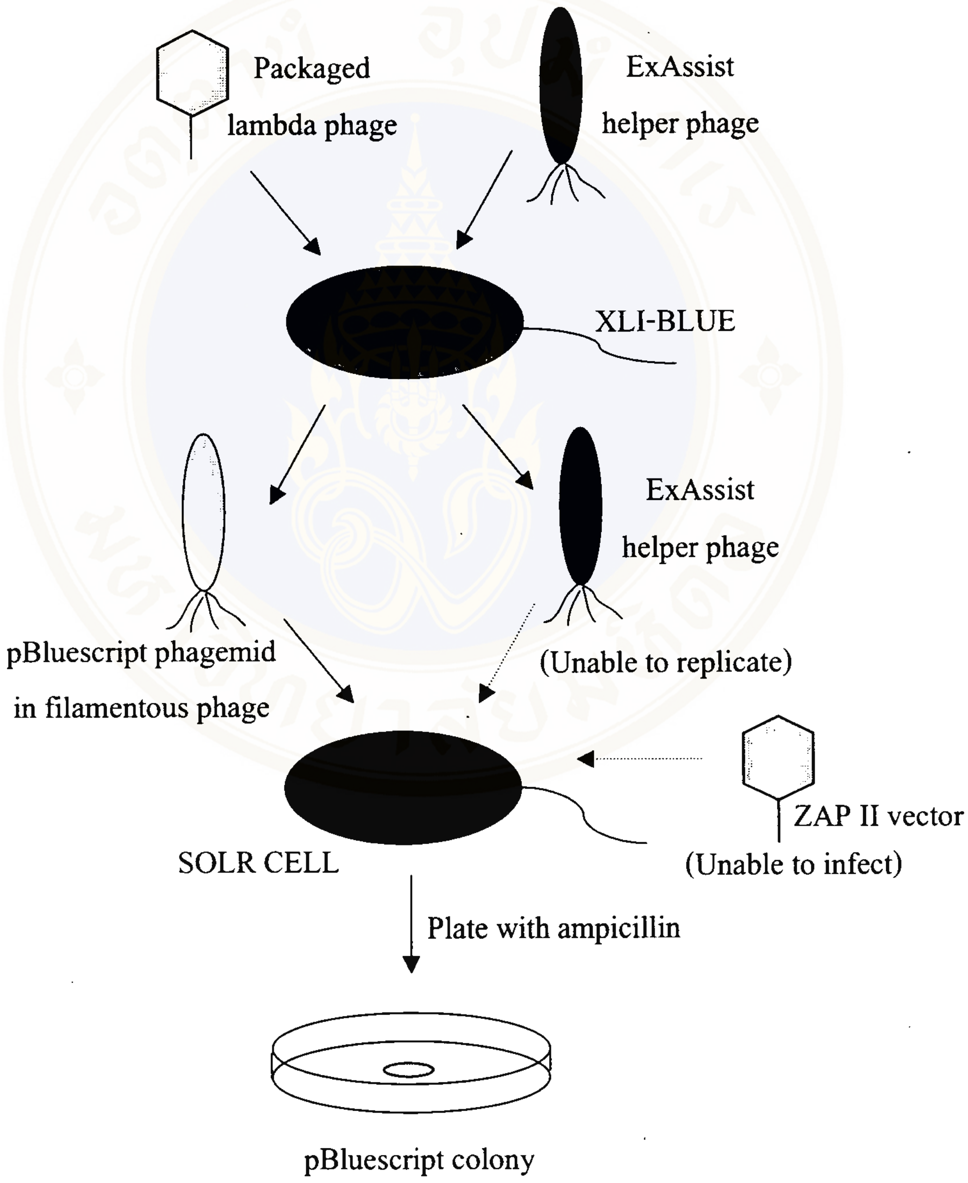


Figure 6 Schematic diagram of the *in vivo* excision protocol, redrawn from Stratagene's protocol.

2.7 Extraction of plasmid DNA using the CTAB method (66)

A single colony of bacterium was inoculated in 3 ml of LB broth containing 100 μg/ml ampicillin and incubated at 37°C with 200 rpm shaking for 16-20 hr. Cell pellet was harvested in 1.5 ml microcentrifuge tubes by centrifugation at 12,000 rpm for 30 sec and then resuspended in 200 μl of STET buffer (8 % w/v sucrose, 0.1 % v/v Triton X-100, 50 mM EDTA, 50 mM Tris-HCl, pH 8.0). To the cell mixture was added 5 μl of freshly prepared lysozyme solution (50 mg/ml) and incubated at room temperature for 20 min. Then, the mixture was boiled for exactly 30 sec and immediately centrifuged at 12,000 rpm for 15 min at room temperature. The pellet was removed with a sterile toothpick. Plasmid DNA and residual low molecular weight RNA was recovered from the supernatant by adding 1/10 volume of 5% (w/v) CTAB. The content was inverted several times and centrifuged at 12,000 rpm for 10 min at room temperature. The pellet was resuspended in 300 μl of 1.2 M NaCl by vigorous vortexing and 2 μl of 10 mg/ml RNase was added, inverted and incubated at 37°C for 30 min. The protein was removed by adding equal volume of chloroform, inverted vigorously and centrifuged at 12,000 rpm for 15 min at room temperature. The clear aqueous phase was transferred to a new tube. The DNA pellet was precipitated with 2 volumes of absolute ethanol at -80°C for 15 min and centrifuged at 12,000 rpm for 10 min at room temperature. The final DNA pellet was washed in 70% ethanol, dried and resuspended in 20 μl sterile water.

2.8 Alkaline agarose gel electrophoresis

For each 100 ml preparation, 0.8 g of agarose was melted in 72 ml of water, allowed to cool to 55 °C. and added with 8 ml of 10 x alkaline buffer (50 ml composed of 3 ml of 5.0 M NaOH, 2 ml of 0.5 M EDTA, 45 ml of sterile water). After swirling to mix, the agarose was poured immediately into the gel apparatus. When the gel was set, the sample was loaded with an equal volume of 2 x loading buffer (1 ml composed of 200 µl of glycerol, 750 µl of water, 46 µl of saturated bromophenol blue, and 5 µl of 5 M NaOH). The gel was run with a 1 x alkaline buffer at 100 mA for 90 min.

2.9 Nondenaturing (5 %) acrylamide gel electrophoresis

Polyacrylamide gel was prepared from the followings in a flask: 1 ml of 10 x TBE buffer, 1.67 ml of a 29:1 (w/w) acrylamide / bis-acrylamide solution, 7.33 ml of sterile deionized water, 5 µl of TEMED, and 50 µl of 10 % (w/v) ammonium persulfate [$(NH_4)_2SO_4$]. The mixture was mixed and immediately filled into the gel apparatus by using a plastic syringe until the glass plate sandwich was full. The comb was inserted into the assembled gel. The gel was let to polymerize at room temperature for 1 hr. The sample was mixed with loading buffer (0.25 % bromophenol blue, 0.25 % xylene cyanol, and 30 % glycerol in water) and loaded into the gel. An electrophoresis was performed in 1 x TBE buffer (89 mM Tris-HCl, pH 8.0, 89 mM boric acid, and 2.5 mM EDTA) at 50 Volts for 2 hours.

2.10 Agarose gel electrophoresis

The analysis of DNA fragments was performed by submarine agarose gel electrophoresis. Separation of the different DNA fragments depends on the molecular size of the DNA, gel concentration and conformation of the DNA. The appropriate gel concentration is dependent on the molecular weight of DNA to be analyzed.

Agarose gels of various percentages (0.8-1.2 %) were prepared by completely dissolving the gel powder upon heating in 1 x TBE (89 mM Tris-HCl, pH 7.4; 89 mM boric acid; 2.5 mM EDTA). The solution was boiled in a microwave oven and allowed to cooled to around 50 °C before pouring into an electrophoresis chamber set, with comb inserted. The DNA sample was mixed with 1/3 volume of loading dye (25% glycerol, 60 mM EDTA, 0.25 % Bromophenol Blue) and loaded into gel slots in a submarine condition. Electrophoresis was performed at 80 volts for 1.5-2.0 hr. The DNA bands in the gel were visualized by staining with 2.5 µg/mL EtBr and photographed under UV light.

2.11 Agarose gel DNA purification

To remove impurities such as small RNAs, proteins, unincorporated nucleotides or primers from DNA sample, the GENECLEAN II® kit was used. A DNA band was excised from an EtBr-stained agarose gel with a flamed razor blade. Three volume of NaI stock solution was added, placed the tube in a 45 °C to 55 °C waterbath. After the agarose was completely dissociated, 5 µl of GLASSMILK® was added, mixed and placed on iced for 5 minutes to allow binding of DNA to the silica matrix and mixed every 1-2 minutes to ensure that of GLASSMILK® stayed suspended. The silica matrix with bound DNA was pelleted in a microcentrifuge for approximately 5 seconds at 14,000 x g. Pellet was washed 3 times with 10 to 50 volumes of ice cold NEW WASH. After the supernatant from the third wash was removed, the tube was spun again for a few seconds to remove the last bit of liquid. DNA from of GLASSMILK® was eluted with TE buffer or sterile water.

2.12 Restriction Endonuclease digestion of plasmid DNA

To check an insert size of each recombinant clone, 1 µl (100-200 ng/µl) of DNA was added with 2 µl of 10 x restriction enzyme buffer (provided with enzymes by manufacturers), 0.1 µl of *Eco*R I (12 U/µl), 0.1 µl of *Xho* I (10 U/µl), and 6.8 µl of sterile DW were added. The tube was quickly centrifuged and incubated at 37 °C overnight before running the agarose gel electrophoresis.

2.13 Blunt-ending of protruding ends

To a 500 µl microcentrifuge tube, 3 µl of 10 x ReacT2 buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM NaCl, pH 8.0), 1 µl of 0.5 mM dATP, 1 µl of 0.5 mM dCTP, 1 µl of 0.5 mM dGTP, 1 µl of 0.5 mM dTTP, 0.5 µg of DNA, 1 µl of Large fragment of DNA polymerase I (diluted to 0.5 U/µl with Klenow Dilution buffer), and sterile water to 30 µl were added. The tube was mixed gently, centrifuged briefly to bring the contents to the bottom and incubated on ice for 20 min. The reaction was terminated by a phenol extraction.

2.14 Blunt-end ligation reaction

To a 0.5 ml microcentrifuge tube, 0.1-1.0 μg of blunt-ended DNA was added with 4 μl of 5 x ligase buffer, 1 μl of T_4 DNA ligase (1 U/μl), and sterile water to final volume of 20 μl. The tube was incubated at 14 °C for 16-24 hours before the plasmid transformation.

2.15 Preparation of *E. coli* competent cells by $CaCl_2$ method (64)

A single colony of *E. coli* JM109 was inoculated into 3 ml of LB broth and incubated at 37 °C with 200 rpm shaking for 16-20 hr. The overnight culture was diluted 1:100 into new LB broth and incubated at 37 °C until the OD_{600} reach 0.4-0.5. The cell culture was chilled on ice for 15 min prior the 10 ml aliquots of the culture was pipetted into 15 ml sterile polypropylene tubes. The cells was pelleted by centrifugation at 3,000 rpm for 10 min at 4 °C, the pellet was suspended in 5 ml of ice-cold 0.1 M $CaCl_2$ and placed on ice for 15-20 min. After centrifugation at 3,000 rpm for 10 min at 4 °C, the pellet was suspended in 1 ml of ice-cold 0.1 M $CaCl_2$ and left on ice for 1 hr to establish competency. Glycerol was added into the cell suspension to give 15% (v/v) final concentration and the cells were kept in 200 μl aliquots at –80 °C until used.

2.16 Transformation of plasmid DNA

A 5 μl aliquot of ligation products was added into a tube containing 200 μl of *E. coli* competent cells. After mixing, the tube was immediately placed on ice for 30 min. The cells were subjected to a heat-shock at 42 °C for 90 sec, and placed on ice for an additional 5 min. Transformed cells were mixed with 800 µl of LB broth and incubated at 37 °C for 1 hr with shaking at 200 rpm. Finally, 100 µl of the transformed culture was spreaded on an LB agar plate containing 100 µg/ml ampicillin and incubated at 37 °C for 16 °C.

2.17 Ethidium bromide plate assay

A 100 ml of 0.8 % (w/v) agarose gel was prepared using Tris-acetate buffer (TAE, 40 mM Tris-acetate, 1 mM EDTA (pH 8.0)). The molten agarose was cooled to 50 °C and 10 µl of EtBr stock solution (10mg/mml) were added. The mixed agarose was poured into 100-mm Petri dishes using 10 ml per plate. The plate was allowed to harden and stored in the dark at 4 °C. A DNA sample of known concentration was used to make seven serial dilutions in 100 mM EDTA as 200, 150, 100, 75, 50, 25, and 10 ng/µl). 0.5 µl of each standard and the cDNA sample were spotted onto the surface of a prepared EtBr plate. All spots were allowed to absorb into the plate for 15 min at room temperature. The lid was removed and the plate was photographed under UV light. The spotted sample of unknown concentration was compared with known standards.

2.18 Automated DNA sequencing

Automated DNA sequencing was performed by using either ABI 310 or ABI 377 DNA sequencers.

Cycle sequencing reaction was performed by using either ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction kit with Amplitaq® DNA Polymerase

FS or ABI PRISM™ BigDye™ Terminator Cycle Sequencing Ready Reaction kit with Amplitaq® DNA Polymerase FS kit according to the protocols as recommended by the manufacturer. The sequencing reaction was performed in 10 μl of total volume containing 4 μl of termination mix, 10 pmol of sequencing primer and 100-300 ng of plasmid DNA template. The reaction was performed in 25 cycles with the following temperatures and time lengths: 96 °C for 10 sec, 50 °C for 5 sec and 60 °C for 4 min. After cycle sequencing, 90 μl of sterile distilled water were added, followed by 0.1 volume of 3 M sodium acetate pH 4.8 and 2.5 volumes of cold absolute ethanol. After centrifugation at 12,000 rpm for 15 min, the pellet was washed with 70% ethanol and briefly dried at room temperature for 5-10 min before resuspended with 2 μl of loading solution (83% of deionized formamide, 8 mM EDTA, pH 8 and 0.8% blue dextran). The reaction mixture was heated at 90 °C for 2 min and immediately cooled on ice before loading into the DNA sequencer.

2.19 Computational analysis of 5' EST data

The 5' EST sequences were compared against nucleotide sequences in all databases by using BLAST program (Basic Local Alignment Search Tool, Version 2.0, developed by Altschul, et al, in 1990) via WWW (54). The multiple alignments were later made using ClustalW and ClustalX program (57-58). Only sequences deemed reliable were reported.

CHAPTER IV

RESULTS

1. Total RNA extraction and poly (A^+) RNA isolation

For cDNA library no.1, an average recovery of total RNA was around 168 μg/ 85 ml of total hemolymph of *Penaeus monodon* by using TRIZOL™ LS reagent, generating around 0.18 μg of poly (A^+) by using a small oligo (dT)-cellulose column method.

For cDNA library no.2, an average recovery of total RNA was around 1.7 mg/ hemocytes from 280 ml of total hemolymph of *Penaeus monodon* by acid guanidinium thiocyanate-phenol-chloroform extraction method, generating around 3.83 μg of Poly (A^+) RNA by using PolyATtract® mRNA isolation system.

Amounts of total RNA and Poly (A^+) RNA were estimated by measuring absorption at 260 nm and 280 nm. A total RNA preparation typically had an A_{260} /A_{280} ratio of 1.8 or higher, indicating a low level of protein contamination. Pattern of total RNA on a formaldehyde agarose gel electrophoresis of both preparations were the same, as shown in figure 7. The predominant band present on the gel was likely 18S ribosomal RNA.

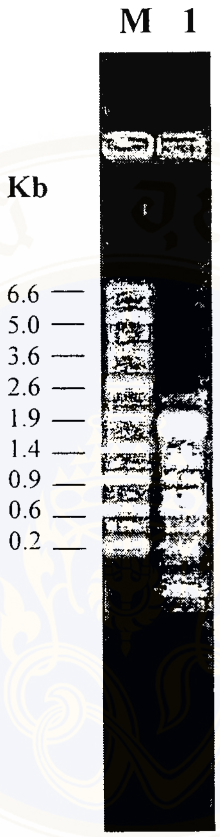


Figure 7 An ethidium bromide-stained gel of a total RNA preparation from hemocytes of *Penaeus monodon*. A 3 μg aliquot of total RNA was electrophoresed on a 1.0 % formaldehyde agarose gel at 80 volts for 90 min.

Lane M : RNA size markers (Promega), 3 μg.

Lane 1 : Total RNA of shrimp's hemocytes, 3 μg. The predominant band at the size of 1.9 kb was likely a mixture of 18 S rRNA and degraded 28 S rRNA.

2. A total hemolymph cDNA library of *P. monodon* (library no.1)

A 0.18 μg of poly (A^+) RNA was used to construct a cDNA library by using ZAP-cDNA® Gigapack® III Gold Cloning kit from Stratagene, Inc. The titer of an unamplified cDNA library was estimated at 1.4 x 10^5 pfu/μg vector.

A number of random plaques was picked from the unamplified library. After *in vivo* excision by single colony excision protocol of Stratagene, the pBluescript SK- phagemids were extracted by CTAB method. The insert sizes were determined by double-digestion with *Eco*R I and *Xho* I, and electrophoresed on 0.8 % agarose gel. From few hundred randomly selected clones, sixty-seven were sequenced by a cycle sequencing protocol using dye terminator chemistry. All of them were subjected to sequencing from the 5' end, by using either T_3 primer or M13 universal reverse primer. The sequencing products were analyzed using ABI 377 automated sequencer, ABI 377 data collection and analysis software. All sequencing chromatograms were also examined visually to ascertain quality of all peaks against the automatic base calling. The vector sequence was removed, the usable sequence at around 500 bases was analyzed for identity.

By using BlastN and BlastX program of NCBI via WWW, 8 out of 67 showed sequence similarity with previously identified genes of other species, and 11 showed no significant matching to those already present in online databases (Table 2, 3 and 9). Each sequence was then examined individually for determination of their significance by using alignment with Clustal X program. Thirty-two clones were confirmed as mitochondrial genes, 10 identified as of *E. coli*, and 6 of lambda sequences. Those sequences were excluded from further analysis.

The results of sequence matching searches for the identification of 8 produced clones are reported in tables 2 and 3. BlastX results in table 2 showed more meaningful than those from BlastN only. The most abundant transcripts observed from this library were mitochondrial genes (47.8 %), and ribosomal proteins (7.5 %). According to a large number of contaminated with mitochondrial genes, *E. coli*, and lambda DNA, a new cDNA library was constructed.

ESTs data set from total hemolymph library was submitted into dbEST (67). Genbank accession numbers and dbEST_id of all clones were shown in table 4.

Table 2 : List of characterized clones from the cDNA library #1. Percent identity and similarity (third and second column from right) of each clone were shown in the protein level, as derived from the Clustal X program. The E value (last column) and putative identity (third column) were based on 'Blast X' searches. Clones identified as ribosomal RNAs and other mitochondrial DNA were not included in this table.

Sequencc identity / Clone	Insert length (kb)	Putative Identification	Matching sequences					
			Organism	Accession No.	Length (aa)	Identity (%)	Similarity (%)	E value
1. Gene / protein expression								
ESTPMH0005	0.6	60S ribosomal protein L14	*Rattus norvegicus*	Q63507	214	51 (42/82)	65 (54/82)	2e-16
ESTPMH0038	0.5	40S ribosomal protein S12	*Sus scrofa*	P46405	132	79 (80/101)	84 (86/101)	7e-42
ESTPMH0095	0.6	40S ribosomal protein S9	*Homo sapiens*	NP_001004.1	194	84 (103/122)	91 (112/122)	2e-52
ESTPMH0140	0.5	60S acidic ribosomal protein P2	*Bos taurus*	P42899	115	64 (43/67)	83 (56/67)	8e-17
ESTPMH0235	0.5	RpL9 gene product [alt 1]	*Drosophila melanogaster*	AAF53049.1	190	57 (22/38)	78 (30/38)	7e-06
2. Cell structure / motility								
ESTPMH0008	1.0	Thymosin beta-11	*Drosophila melanogaster*	CAA21832.1	129	62 (31/50)	68 (34/50)	4e-08
ESTPMH0017	1.3	Calponin homolog	*Schistosoma mansoni*	AAB47536.1	361	50 (57/112)	66 (75/112)	6e-21
3. Metabolism								
ESTPMH0025	1.6	ADP-ribosylation factor 1	*Drosophila melanogaster*	P35676	182	100 (124/124)	100 (124/124)	2e-68

* An overlapping region of homology was shown in parentheses as residues of identical / residues of total compared region.

The clone designation of this library no.1 was PMHxxxx, whereas xxxx was the unique identification number. Each clone's partial sequence was designated with a prefix EST in front of the clone identity. Insert length (kb) was determined from restriction enzyme analysis, with *Eco*R I and *Xho* I. The length (aa) of the matching sequences was derived from the online record.

Table 3 : List of characterized clones from the cDNA library #1. Percent identity (second column from right) of each clone was shown in the nucleotide level, as derived from the Clustal X program. The E value (last column) and putative identity (third column) were based on 'Blast N' searches. Clones identified as ribosomal RNAs and other mitochondrial DNA were not included in this table.

Sequence identity / Clone	Insert length (kb)	Putative Identification	Matching sequences				
			Organism	Accession No.	Length (bp)	Identity (%)	E value
1. Gene / protein expression							
ESTPMH0005	0.6	No significant similarity found	-	-	-	-	-
ESTPMH0038	0.5	No significant similarity found	-	-	-	-	-
ESTPMH0095	0.6	40S ribosomal protein S9 (partial cds)	*Schizosaccharomyces pombe*	AB029515.1	637	83 (60/72)	4e-05
ESTPMH0140	0.5	No significant similarity found	-	-	-	-	-
ESTPMH0235	0.5	No significant similarity found	-	-	-	-	-
2. Cell structure / motility							
ESTPMH0008	1.0	No significant similarity found	-	-	-	-	-
ESTPMH0017	1.3	Neuronal protein (NP25)	*Mus musculus*	AB031291.1	846	91 (34/37)	6e-04
3. Metabolism							
ESTPMH0025	1.6	ADP-ribosylation factor 1	*Locusta migratoria*	U90609.1	2193	80 (256/317)	2e-32

* An overlapping region of homology was shown in parentheses as bp of identical / bp of total compared region.

The clone designation of this library no.1 was PMHxxxx, whereas xxxx was the unique identification number. Each clone's partial sequence was designated with a prefix EST in front of the clone identity. Insert length (kb) was determined from restriction enzyme analysis, with *Eco*R I and *Xho* I. The length (bp) of the matching sequences was derived from the online record.

Table 4 : dbEST_Id and Genbank accession numbers of our ESTs submission from the total hemolymph cDNA library (library #1).

dbEST_Id[a]	User_Id[b]	GenBank_Accn[c]	dbEST_Id	User_Id	GenBank_Accn
3712633	ESTPMH0019	AW330525	3914860	ESTPMH0005	AW497579
3712634	ESTPMH0052	AW330526	3914861	ESTPMH0008	AW497580
3712635	ESTPMH0057	AW330527	3914862	ESTPMH0017	AW497581
3712636	ESTPMH0089	AW330528	3914863	ESTPMH0025	AW497582
3712637	ESTPMH0108	AW330529	3914864	ESTPMH0038	AW497583
3712638	ESTPMH0118	AW330530	3914865	ESTPMH0074	AW497584
3712639	ESTPMH0124	AW330531	3914866	ESTPMH0083	AW497585
3712640	ESTPMH0145	AW330532	3914867	ESTPMH0095	AW497586
3712641	ESTPMH0188	AW330533	3914868	ESTPMH0140	AW497587
3712642	ESTPMH0266	AW330534	3914869	ESTPMH0172	AW497588
3712643	ESTPMH0305	AW330535	3914870	ESTPMH0235	AW497589

(a) received after submission, used for searching of DNA sequences in database of EST.
(b) assigned by our lab.
(c) received after submission, used for searching of DNA sequences in Genbank database.

* The 22 clones listed in this table also included some miochondrial DNAs, which were not listed in Table 2 and 3.

3. A total hemocyte cDNA library of *P. monodon* (library no.2)

The lambda ZAP II-based cDNA library was constructed from 3.83 μg of poly (A^+) RNA from total hemocyte of *P. monodon*. The titer of the library was 1.7 x 10^6 pfu/μg vector. After *in vivo* excision and screening by double digestion with *Eco*R I and *Xho* I of 460 clones, 194 independent clones with an insert size mostly greater than 0.5 kb were sequenced. The 5' ESTs were examined for similarities in the nucleic acid and protein databases using BLAST algorithm (54), 123 out of 194 clones showed significant match to sequences in online databases (Table 9). Twenty-seven of them were considered as mitochondrial genes. The putative identification of the rest was showed in table 5, in protein level only. They were classified according to cellular biological roles (68). By using Blast N alone, many recombinant clones shown no significant homology. Then, sequence determination in translated forms was conducted by BlastX. Clustal X program was used for sequence alignment. No contamination with *E. coli* or lambda DNA was found in this library. All ESTs were submitted into dbEST. Genbank accession numbers and dbEST_id were shown in table 6.

The original cDNA library is almost perfectly directional, since 99 % (122/123) of sequences with similarities in the databases contained matched on the expected strand. Among the 194 clones, insert length was varied from 0.4 kb to 2.6 kb. An average size was 1.0 kb. Size distribution was shown in figure 8 and table 7.

The most abundant transcripts observed from total hemocyte cDNA library were ribosomal proteins (5.7 %), actin (4.1 %), elongation factor (3.6 %), and serine proteinase inhibitor (3.1 %) as showed in table 8.

Table 5 : List of characterized clones from the cDNA library #2. Percent identity and similarity were shown in the protein level only (third and second column from right), as derived from the Clustal X program. The E values (last column) and putative identities (third column) were as obtained from Blat X searches. Clones identified as rRNAs and mitochondrial DNAs were not listed in this table.

Clone	Insert length (kp)	Putative Identification	Matching sequences					
			Organism	Accession No.	Length (aa)	Identity (%)	Similarity (%)	E value
1. Cell signaling communication								
ESTPMC0002	1.072	Ran GTP-binding protein	*Xenopus laevis*	BAA89696.1	215	87 (188/215)	91 (197/215)	e-99
ESTPMC0010	1.3	Ferritin subunit precursor	*Aedes aegypti*	P41822	209	33 (73/212)	51 (114/212)	2e-18
ESTPMC0014	1.0	Calreticulin	*Xenopus laevis*	S29129	411	79 (136/171)	88 (151/171)	1e-73
ESTPMC0231	0.6	Serine/threonine-protein kinase	*Mus musculus*	Q61136	496	55 (33/59)	77 (46/59)	5e-12
2. Gene / protein expression								
ESTPMC0007	0.5	60S ribosomal protein L3	*Drosophila melanogaster*	O16797	416	73 (99/134)	85 (115/134)	3e-50
ESTPMC0311	0.45	60S ribosomal protein L5	*Bombyx mori*	O76190	299	65 (44/67)	79 (54/67)	3e-20
ESTPMC0284	0.8	60S ribosomal protein L7	Caenorhabditis elegans	O01802	244	49 (100/204)	67 (137/204)	1e-56
ESTPMC0337	0.6	60S ribosomal protein L9	*Drosophila melanogaster*	P50882	190	64 (124/192)	80 (155/192)	4e-65
ESTPMC0076	0.9	60S ribosomal protein L12	*Rattus rattus*	P23358	165	74 (63/84)	88 (75/84)	2e-29
ESTPMC0271	2.0	60S ribosomal protein L15	*Orconectes limosus*	Q9XYC2	204	63 (118/187)	72 (135/187)	5e-63
ESTPMC0107	1.0	60S ribosomal protein L35	*Rattus rattus*	P17078	123	64 (77/119)	76 (91/119)	8e-27
ESTPMC0298	0.9	40S ribosomal protein S2	*Drosophila melanogaster*	P31009	267	92 (118/127)	94 (121/127)	7e-70
ESTPMC0124	0.7	40S ribosomal protein S7	*Homo sapiens*	NP_001002.1	194	69 (131/189)	84 (158/189)	6e-69
ESTPMC0342	0.5	40S ribosomal protein S12	*Sus scrofa*	P46405	132	69 (78/113)	82 (93/113)	2e-41
ESTPMC0330	0.5	40S ribosomal protein S19	*Mya arenaria*	Q94613	149	57 (69/120)	75 (91/120)	1e-36
ESTPMC0098	0.7	Elongation factor 1-alpha	*Armadilidium vulgare*	AAC03145.1	364	86 (178/204)	90 (185/204	1e-96
ESTPMC0220	1.3	Elongation factor 1-alpha	*Archaeoattacus edwardsii*	AAC47893.1	413	94 (105/111)	98 (109/111)	5e-58
ESTPMC0290	0.7	Elongation factor 1-beta	*Artemia salina*	P12262	207	71 (49/69)	88 (61/69)	1e-31
ESTPMC0317	0.7	Elongation factor 1-gamma	*Artemia sp.*	P12261	430	54 (113/208)	68 (144/208)	5e-67
ESTPMC0328	0.75	Elongation factor 2 (EF-2)	*Caenorhabditis elegans*	P29691	852	86 (146/169)	92 (157/169)	3e-87
ESTPMC0025	0.85	Eukaryotic translation elongation factor 1 beta 2	*Homo sapiens*	NP_001950.1	225	61 (102/160)	73 (122/160)	3e-33
ESTPMC0159	1.0	Eukaryotic translation elongation factor 1 delta	*Homo sapiens*	NP_001951.1	281	48 (29/60)	63 (38/60)	5e-07
ESTPMC0309	0.8	Eukaryotic translation initiation factor 3 subunit 2	*Drosophila melanogaster*	O02195	326	43 (104/239)	61 (149/239)	2e-49

Clone	Insert length (kp)	Putative Identification	Matching sequences					
			Organism	Accession No.	Length (aa)	Identity (%)	Similarity (%)	E value
ESTPMC0086	0.52	Ubiquitin extension protein	*Drosophila melanogaster*	S23988	128	93 (108/116)	100 (116/116)	1e-57
ESTPMC0022	1.4	PolyA binding protein-interacting protein PAIP1	*Mus musculus*	AAD28259.1	338	32 (55/159)	52 (89/159)	7e-16
ESTPMC0033	0.73	26S proteasome regulatory complex subunit p97 mRNA complete cds	*Homo sapiens*	BAA11226.1	908	85 (79/92)	95 (88/92)	1e-30
ESTPMC0041	1.4	Exon 5 contains a tRNA-Leu	*Caenorhabditis elegans*	AAA81125.1	470	33 (63/190)	51 (98/190)	2e-24
ESTPMC0045	2.0	Heterogeneous nuclear RNP protein	*Drosophila melanogaster*	B41732	345	69 (60/86)	75 (65/86)	9e-21
ESTPMC0105	0.95	TATA box binding protein (TBP)-associated factor, RNA polymerase II, G, 32kD	*Homo sapiens*	NP_003178.1	264	59 (79/131)	78 (104/131)	6e-23
ESTPMC0194	0.9	Small nuclear ribonucleoprotein SM D2	*Caenorhabditis elegans*	Q18786	118	40 (46/114)	54 (62/114)	3e-14
ESTPMC0214	1.4	Proteasome regulator dREG	*Drosophila melanogaster*	AAF19529.1	245	73 (96/131)	92 (122/131)	1e-52
ESTPMC0320	0.6	Multicatalytic endopeptidase complex	*Homo sapiens*	JC1445	269	49 (43/87)	71 (63/87)	3e-18
ESTPMC0401	1.0	ADP/ATP translocase	*Oryctolagus cuniculus*	BAA23777.1	298	76 (111/145)	88 (129/145)	1e-63
ESTPMC0415	0.9	ADP/ATP translocase	*Oryctolagus cuniculus*	BAA23777.1	298	77 (58/75)	87 (66/75)	5e-40
3. Cell division / DNA synthesis								
ESTPMC0006	2.1	Equilbrative nitrobenzylthioinosine-sensitive nucleoside transporter	*Rattus norvegicus*	AAB88049.1	457	44 (62/136)	63 (87/136)	5e-24
ESTPMC0127	1.2	Chromatin assembly factor 1	*Drosophila melanogaster*	Q24572	430	76 (148/192)	83 (161/192)	5e-83
ESTPMC0308	0.65	Component of the spindle pole body, Nuf1p	*Saccharomyces cerevisiae*	NP_010643.1	944	26 (45/173)	47 (82/173)	0.005
ESTPMC0312	0.8	SNR1 (regulate homeotic gene transcription)	*Drosophila melanogaster*	AAC77830.1	370	47 (82/174)	60 (106/174)	3e-34
4. Cell structure / motility								
ESTPMC0357	1.0	Actin	*Kluyveromyces lactis*	P17128	375	56 (108/191)	66 (128/191)	1e-58
ESTPMC0211	0.8	Actin-5C	*Drosophila melanogaster*	P10987	376	72 (81/112)	80 (91/112)	6e-38
ESTPMC0281	1.4	Actin 11	*Limulus polyphemus*	P41341	376	97 (94/96)	99 (96/96)	5e-70
ESTPMC0458	0.95	Actin 46	*Solanum tuberosum*	P93586	336	68 (28/41)	68 (28/41)	7e-15
ESTPMC0226	1.2	Actin, cytoplasmic A3A	*Helicoverpa armigera*	Q25010	376	100 (125/125)	100 (125/125)	e-112
ESTPMC0028	1.8	Actin type 5, cytosolic	*Gallus gallus*	A26559	362	96 (135/140)	98 (138/140)	2e-87
ESTPMC0102	0.8	Actin, nonmuscle	*Halocynthia roretzi*	P53461	376	52 (93/177)	72 (129/177)	5e-46
ESTPMC0011	0.6	Gamma-non muscle actin	*Artemia sp.*	P18603	376	100 (137/137)	100 (137/137)	8e-77
ESTPMC0090	0.8	Profilin	*Drosophila melanogaster*	P25843	126	41 (34/82)	56 (47/82)	9e-10
ESTPMC0112	0.9	Tubulin alpha 1	*Homarus americanus*	Q25008	451	74 (160/215)	79 (172/215)	2e-83
ESTPMC0162	1.1	Thymosin beta-11	*Drosophila melanogaster*	CAA21832	129	40 (51/126)	50 (64/126)	5e-15
ESTPMC0165	1.0	Thymosin beta-11	*Drosophila melanogaster*	CAA21832	129	52 (67/128)	60 (78/128)	2e-26

Clone	Insert length (kp)	Putative Identification	Matching sequences					
			Organism	Accession No.	Length (aa)	Identity (%)	Similarity (%)	E value
ESTPMC0343	1.0	Thymosin beta-11	*Drosophila melanogaster*	CAA21832	129	48 (51/105)	62 (66/105)	5e-21
ESTPMC0167	1.5	Calponin	*Schistosoma japonicum*	AAD11976	361	33 (45/134)	45 (62/134)	6e-12
ESTPMC0174	1.5	Associated to golgi apparatus	*Homo sapiens*	CAA62380	214	66 (119/180)	78 (141/180)	2e-65
ESTPMC0192	0.67	Profilin 2	*Triticum aestivum*	P49233	141	37 (30/79)	55 (45/79)	1e-11
ESTPMC0240	1.1	Coactosin	*Homo sapiens*	AAA88022.1	142	50 (61/121)	62 (76/121)	8e-28
ESTPMC0286	0.95	Coactosin	*Homo sapiens*	AAA88022.1	142	52 (49/93)	64 (61/93)	2e-20
ESTPMC0238	0.8	Selenoprotein W	*Rattus norvegicus*	NP_037159.1	88	62 (51/82)	74 (61/82)	7e-23
ESTPMC0307	0.9	Myosin regulatory light chain, non-muscle	*Drosophila melanogaster*	P40423	174	78 (22/28)	92 (26/28)	6e-06
ESTPMC0341	0.8	Gcap 1 (Granule cell marker protein)	*Mus musculus*	AAA68426.1	85	35 (20/56)	60 (34/56)	0.002
5. Cell / organism defence and homeostasis								
ESTPMC0023	1.0	Serine proteinase inhibitor	*Pacifastacus leniusculus*	S45677	228	50 (56/110)	62 (70/110)	1e-23
ESTPMC0246	0.9	Serine proteinase inhibitor	*Pacifastacus leniusculus*	S45677	228	47 (46/97)	57 (56/97)	9e-16
ESTPMC0299	1.7	Serine proteinase inhibitor	*Pacifastacus leniusculus*	S45677	228	55 (40/72)	63 (46/72)	4e-18
ESTPMC0347	1.5	Serine proteinase inhibitor	*Pacifastacus leniusculus*	S45677	228	43 (14/32)	64 (21/32)	0.002
ESTPMC0396	0.9	Serine proteinase inhibitor	*Pacifastacus leniusculus*	S45677	228	53 (37/69)	60 (42/69)	7e-17
ESTPMC0411	1.2	Serine proteinase inhibitor	*Pacifastacus leniusculus*	S45677	228	40 (22/54)	49 (27/54)	3e-04
ESTPMC0184	0.5	Chelonianin (Basic protease inhibitor) (RTPI)	*Caretta caretta*	P00993	110	52 (24/46)	56 (26/46)	8e-05
ESTPMC0129	0.65	Secretory leukocyte protease inhibitor (antileukoproteinase)	*Homo sapiens*	NP_003055.1	132	35 (32/90)	53 (49/90)	1e-04
ESTPMC0084	0.5	Clathrin-associated protein AP17 delta	*Homo sapiens*	CAA09019	104	93 (29/31)	96 (30/31)	8e-08
ESTPMC0417	0.5	Anti-bacterial protein 11.5 kDa	*Carcinus maenas*	CAB51030.1	84	43 (20/46)	49 (23/46)	2e-04
ESTPMC0155	1.4	T-complex protein 1, theta subunit (contain cytosolic chaperonin)	*Mus musculus*	P42932	548	52 (100/191)	64 (124/191)	5e-49
ESTPMC0252	0.9	Clathrin light chain	*Drosophila melanogaster*	AAC14276.1	228	59 (42/71)	78 (56/71)	6e-19
ESTPMC0265	1.2	Peroxinectin	*Pacifastacus leniusculus*	JC4397	818	67 (133/197)	82 (163/197)	1e-77
ESTPMC0287	1.6	Zinc finger protein	*Mus musculus*	NP_035885.1	580	47 (68/144)	62 (90/144)	9e-35
ESTPMC0412	1.4	T complex protein 1, zeta subunit (chaperonin subunit 6a)	*Mus musculus*	NP_033968.1	531	65 (109/166)	80 (135/166)	2e-56
6. Metabolism								
ESTPMC0003	0.95	Phosphopyruvate hydratase	*Penaeus monodon*	AF100985	434	100 (141/141)	100 (141/141)	1e-79
ESTPMC0185	0.95	ATPase, H^+ transporting, lysosomal (vacuolar proton pump), isoform 2	*Homo sapiens*	NP_001682.1	615	73 (60/82)	90 (74/82)	1e-30
ESTPMC0055	1.3	ADP-ribosylation factor 1 (ARF1)	***Drosophila melanogaster***	P35676	182	100 (87/87)	100 (87/87)	8e-45
ESTPMC0059	0.6	Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	*Homarus americanus*	P00357	333	90 (29/32)	93 (30/32)	8e-11

Clone	Insert length (kp)	Putative Identification	Matching sequences					
			Organism	Accession No.	Length (aa)	Identity (%)	Similarity (%)	E value
ESTPMC0077	1.0	Anti-oxidant protein 2	*Homo sapiens*	NP_004896.1	224	69 (92/133)	80 (107/133)	5e-50
ESTPMC0125	1.5	Glucose-6-phosphate isomerase	*Sus scrofa*	P08059	558	79 (106/134)	91 (123/134)	2e-63
ESTPMC0171	0.8	Methylmalonyl-CoA mutase	*Homo sapiens*	AAA99226.1	750	73 (91/123)	85 (106/123)	5e-48
ESTPMC0229	0.9	Phosphoglyceromutase	*Drosophila melanogaster*	S50326	255	79 (129/163)	89 (146/163)	2e-73
ESTPMC0266	1.2	Adenosylhomocysteinase 1	*Xenopus laevis*	P51893	433	78 (139/177)	88 (157/177)	9e-87
ESTPMC0273	2.0	Selenophosphate synthetase	*Drosophila melanogaster*	O18373	398	90 (149/165)	94 (156/165)	3e-83
ESTPMC0301	0.8	NADH-ubiquinone oxidoreductase B18 subunit	*Bos taurus*	Q02368	137	50 (44/88)	64 (57/88)	3e-18
ESTPMC0441	1.6	Pyruvate kinase	*Xenopus laevis*	Q92122	527	46 (43/92)	64 (60/92)	2e-27
ESTPMC0449	0.6	Nucleoside diphosphate kinase-Z1	*Danio rerio*	AAF20910.1	153	42 (33/77)	60 (47/77)	5e-12
7. Unclassified								
ESTPMC0072	1.0	E2IG5	*Homo sapiens*	AAF09484.1	148	40 (21/52)	65 (34/52)	2e-04
ESTPMC0118	1.1	RP42	*Mus musculus*	AAF04863.1	259	40 (55/136)	57 (79/136)	5e-23
ESTPMC0172	1.6	cDNA EST yk415c12.5 comes from this gene	*Caenorhabditis elegans*	CAB04889.2	183	36 (37/102)	65 (67/102)	2e-19
ESTPMC0255	0.9	BACN4L24.f	*Drosophila melanogaster*	CAB58075.1	135	30 (37/122)	48 (59/122)	3e-10
ESTPMC0276	0.5	HSPC300	*Homo sapiens*	AAF28978.1	110	71 (53/74)	87 (65/74)	3e-25
ESTPMC0300	0.9	ARP2/3 complex	*Caenorhabditis elegans*	Q18491	344	53 (81/152)	78 (119/152)	2e-43
ESTPMC0306	1.7	F45F2.5 gene product	*Caenorhabditis elegans*	AAC48028.1	556	42 (40/95)	66 (63/95)	1e-14
ESTPMC0393	0.5	Hypothetical protein	*Schizosaccharomyces pombe*	CAB54156.1	317	36 (37/102)	50 (52/102)	3e-11
ESTPMC0373	0.8	Balbiani ring protein 3 precursor	*Chironomus tentans*	Q03376	1700	33 (22/66)	54 (36/66)	2e-04

* An overlapping region of homology was shown in parentheses as residues of identical / residues of total compared region.

The clone designation of this library no.2 was PMCxxxx, whereas xxxx was the unique identification number. Each clone's partial sequence was designated with a prefix EST in front of the clone identity. Insert length (kb) was determined from restriction enzyme analysis, with *Eco*R I and *Xho* I. The length (aa) of the matching sequences was derived from the online record.

Table 6 : dbEST_Id and Genbank accession numbers of our ESTs submission from the total hemocyte cDNA library (library #2).

dbEST_Id(a)	User_Id(b)	GenBank_Accn(c)	dbEST_Id	User_Id	GenBank_Accn
4033938	ESTPMC003	AW600672	4033988	ESTPMC172	AW600722
4033939	ESTPMC007	AW600673	4033989	ESTPMC174	AW600723
4033940	ESTPMC011	AW600674	4033990	ESTPMC175	AW600724
4033941	ESTPMC012	AW600675	4033991	ESTPMC180	AW600725
4033942	ESTPMC014	AW600676	4033992	ESTPMC181	AW600726
4033943	ESTPMC016	AW600677	4033993	ESTPMC184	AW600727
4033944	ESTPMC022	AW600678	4033994	ESTPMC185	AW600728
4033945	ESTPMC023	AW600679	4033995	ESTPMC194	AW600729
4033946	ESTPMC028	AW600680	4033996	ESTPMC196	AW600730
4033947	ESTPMC033	AW600681	4033997	ESTPMC211	AW600731
4033948	ESTPMC034	AW600682	4033998	ESTPMC214	AW600732
4033949	ESTPMC037	AW600683	4033999	ESTPMC217	AW600733
4033950	ESTPMC040	AW600684	4034000	ESTPMC220	AW600734
4033951	ESTPMC041	AW600685	4034001	ESTPMC226	AW600735
4033952	ESTPMC045	AW600686	4034002	ESTPMC231	AW600736
4033953	ESTPMC047	AW600687	4034003	ESTPMC237	AW600737
4033954	ESTPMC053	AW600688	4034004	ESTPMC238	AW600738
4033955	ESTPMC055	AW600689	4034005	ESTPMC239	AW600739
4033956	ESTPMC059	AW600690	4034006	ESTPMC240	AW600740
4033957	ESTPMC062	AW600691	4034007	ESTPMC245	AW600741
4033958	ESTPMC063	AW600692	4034008	ESTPMC252	AW600742
4033959	ESTPMC064	AW600693	4034009	ESTPMC255	AW600743
4033960	ESTPMC066	AW600694	4034010	ESTPMC256	AW600744
4033961	ESTPMC068	AW600695	4034011	ESTPMC257	AW600745
4033962	ESTPMC075	AW600696	4034012	ESTPMC258	AW600746
4033963	ESTPMC076	AW600697	4034013	ESTPMC261	AW600747
4033964	ESTPMC077	AW600698	4034014	ESTPMC262	AW600748
4033965	ESTPMC089	AW600699	4034015	ESTPMC265	AW600749
4033966	ESTPMC098	AW600700	4034016	ESTPMC266	AW600750
4033967	ESTPMC102	AW600701	4034017	ESTPMC271	AW600751
4033968	ESTPMC107	AW600702	4034018	ESTPMC273	AW600752
4033969	ESTPMC111	AW600703	4034019	ESTPMC276	AW600753
4033970	ESTPMC113	AW600704	4034020	ESTPMC277	AW600754
4033971	ESTPMC115	AW600705	4034021	ESTPMC279	AW600755
4033972	ESTPMC118	AW600706	4034022	ESTPMC280	AW600756
4033973	ESTPMC121	AW600707	4034023	ESTPMC281	AW600757
4033974	ESTPMC124	AW600708	4034024	ESTPMC284	AW600758
4033975	ESTPMC125	AW600709	4034025	ESTPMC286	AW600759
4033976	ESTPMC127	AW600710	4034026	ESTPMC287	AW600760
4033977	ESTPMC128	AW600711	4034027	ESTPMC289	AW600761
4033978	ESTPMC132	AW600712	4034028	ESTPMC290	AW600762
4033979	ESTPMC135	AW600713	4034029	ESTPMC293	AW600763
4033980	ESTPMC146	AW600714	4034030	ESTPMC295	AW600764
4033981	ESTPMC155	AW600715	4034031	ESTPMC297	AW600765
4033982	ESTPMC162	AW600716	4034032	ESTPMC298	AW600766
4033983	ESTPMC165	AW600717	4034033	ESTPMC299	AW600767
4033984	ESTPMC166	AW600718	4034034	ESTPMC301	AW600768
4033985	ESTPMC167	AW600719	4034035	ESTPMC001	AW600769
4033986	ESTPMC170	AW600720	4034036	ESTPMC002	AW600770
4033987	ESTPMC171	AW600721	4034037	ESTPMC006	AW600771

Table 6 (continued)

dbEST_Id[a]	User_Id[b]	GenBank_Accn[c]	dbEST_Id	User_Id	GenBank_Accn
4034038	ESTPMC009	AW600772	4053528	ESTPMC337	AW618932
4034039	ESTPMC010	AW600773	4053529	ESTPMC339	AW618933
4034040	ESTPMC021	AW600774	4053530	ESTPMC341	AW618934
4034041	ESTPMC025	AW600775	4053531	ESTPMC342	AW618935
4034042	ESTPMC044	AW600776	4053532	ESTPMC343	AW618936
4034043	ESTPMC072	AW600777	4053533	ESTPMC347	AW618937
4034044	ESTPMC079	AW600778	4053534	ESTPMC352	AW618938
4034045	ESTPMC086	AW600779	4053535	ESTPMC357	AW618939
4034046	ESTPMC087	AW600780	4053536	ESTPMC359	AW618940
4034047	ESTPMC090	AW600781	4053537	ESTPMC362	AW618941
4034048	ESTPMC105	AW600782	4053538	ESTPMC371	AW618942
4034049	ESTPMC106	AW600783	4053539	ESTPMC373	AW618943
4034050	ESTPMC112	AW600784	4053540	ESTPMC375	AW618944
4034051	ESTPMC129	AW600785	4053541	ESTPMC376	AW618945
4034052	ESTPMC134	AW600786	4053542	ESTPMC377	AW618946
4034053	ESTPMC140	AW600787	4053543	ESTPMC389	AW618947
4034054	ESTPMC144	AW600788	4053544	ESTPMC391	AW618948
4034055	ESTPMC148	AW600789	4053545	ESTPMC393	AW618949
4034056	ESTPMC152	AW600790	4053546	ESTPMC394	AW618950
4034057	ESTPMC159	AW600791	4053547	ESTPMC396	AW618951
4034058	ESTPMC163	AW600792	4053548	ESTPMC399	AW618952
4034059	ESTPMC177	AW600793	4053549	ESTPMC401	AW618953
4034060	ESTPMC186	AW600794	4053550	ESTPMC402	AW618954
4034061	ESTPMC189	AW600795	4053551	ESTPMC403	AW618955
4034062	ESTPMC192	AW600796	4053552	ESTPMC406	AW618956
4034063	ESTPMC204	AW600797	4053553	ESTPMC410	AW618957
4034064	ESTPMC224	AW600798	4053554	ESTPMC411	AW618958
4034065	ESTPMC229	AW600799	4053555	ESTPMC412	AW618959
4034066	ESTPMC243	AW600800	4053556	ESTPMC415	AW618960
4034067	ESTPMC246	AW600801	4053557	ESTPMC416	AW618961
4034068	ESTPMC300	AW600802	4053558	ESTPMC417	AW618962
4034069	ESTPMC084	AW600803	4053559	ESTPMC419	AW618963
4053513	ESTPMC303	AW618905	4053560	ESTPMC422	AW618964
4053514	ESTPMC306	AW618906	4053561	ESTPMC428	AW618965
4053515	ESTPMC307	AW618907	4053562	ESTPMC434	AW618966
4053516	ESTPMC308	AW618920	4053563	ESTPMC438	AW618908
4053517	ESTPMC309	AW618921	4053564	ESTPMC441	AW618909
4053518	ESTPMC311	AW618922	4053565	ESTPMC445	AW618910
4053519	ESTPMC312	AW618923	4053566	ESTPMC446	AW618911
4053520	ESTPMC313	AW618924	4053567	ESTPMC449	AW618912
4053521	ESTPMC317	AW618925	4053568	ESTPMC451	AW618913
4053522	ESTPMC320	AW618926	4053569	ESTPMC453	AW618914
4053523	ESTPMC322	AW618927	4053570	ESTPMC454	AW618915
4053524	ESTPMC328	AW618928	4053571	ESTPMC455	AW618916
4053525	ESTPMC330	AW618929	4053572	ESTPMC456	AW618917
4053526	ESTPMC333	AW618930	4053573	ESTPMC457	AW618918
4053527	ESTPMC336	AW618931	4053574	ESTPMC458	AW618919

(a) received after submission, used for searching of DNA sequences in database of EST.

(b) assigned by our lab.

(c) received after submission, used for searching of DNA sequences in Genbank database.

* The 194 clones listed in this table also include those of rRNAs and mitochondrial DNAs, which were not listed in Table 5.

Table 7 : Size-distribution of cDNA inserts from the cDNA library # 2.

Size-range (kb)	Number of clones
0.0-0.5	15
0.6-1.0	106
1.1-1.5	46
1.6-2.0	21
2.1-2.5	5
2.6-3.0	1
Total	194

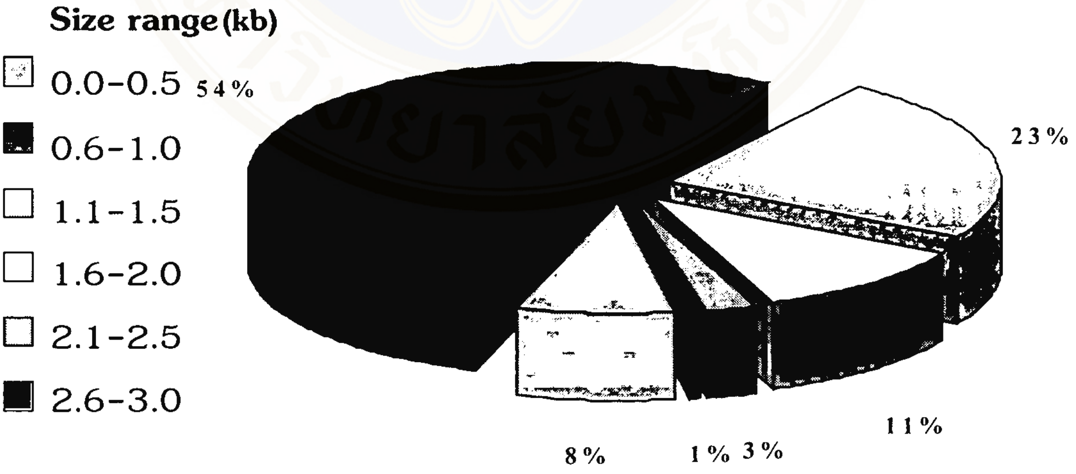


Figure 8 Distribution of the 194 insert sizes from the cDNA library # 2.

Table 8 : Abundant ESTs from *P. monodon*'s hemocyte library (library # 2).

Rank	Putative identity	No. of ESTs	%
1	Ribosomal proteins * (all different)	11	5.7
2	Actin	8	4.1
3	Elongation factor * (2 types) [2 of EF-1 alpha, 2 of identical unclassified EF, and 3 of different EF]	7	3.6
4	Serine proteinase inhibitor	6	3.1
5	Thymosin beta-11	3	1.5
6	Chaperonin * (2 types) [1 of theta-like, and 1 of zeta-like subunit]	2	1.0
7	Coactosin	2	1.0
8	ADP/ATP translocase	2	1.0
9	Profilin	2	1.0

* With the exception of clones for ribosomal proteins, which are all different, and clones for translation elongation factors and chaperonins, which are of 2 types, the rest of the clones for the same identity were likely belong to the same gene.

Table 9 : Summary of ESTs from the two hemocyte cDNA libraries of *P. monodon* (library no.1 and 2, respectively).

EST category	Library #1		Library #2	
	No. of ESTs	Percentage	No. of ESTs	Percentage
Total no. of ESTs	67	100	194	100
ESTs with significant match				
Ribosomal proteins	5	8	11	5.7
Metabolic enzymes	-	-	13	6.7
Mitochondrial genes	32	48	27	13.9
Others	3	4	72	37.1
ESTs without significant match	11	16	71	36.5
Contaminants				
E. coli DNA	10	15	-	-
Lambda vector DNA	6	9	-	-

From similarity searching, ESTPMC0003 was a clone of phosphopyruvate hydratase or enolase, a gene previously reported from our laboratory in 1998 (Accession: AF100985) (69). When the sequence alignment was done as in figure 9, an identity within the open reading frame was 100 % but some sequence discrepancies were found in the 3'UTR. Compare to AF100985, with its 1861 bp full-length, this clone was considered to be truncated at 5' end of around 903 bp. A diagnostic peptide sequence motif (DDLTVTNPK) known to be specific to all enolase active sites was found, figure 10. Upon reexamining the chromatograms of the new clone PMC0003 and those of AF100985 produced by Boonyawan, the sequence differences seemed to be real. It was possible that these could be due to errors of sequencing reactions. Another possibility is that this sequence was transcribed from another enolase gene, with a diverged 3' UTR sequence.

One of the most abundant ESTs was putative actin sequences (4.1 %). Multiple alignment of all 8 putative actin sequences of the hemocytes was done (data not shown). The cDNA sequences were identical and different only in sizes. A schematic pattern of overlapping regions is shown in figure 11. A consensus sequence, around 1,100 bp with poly A^+, was named as PMC-act. After the sequence identity searching was performed again, BLAST search results were obtained as shown in figure 12. Clustal X program was used for sequence alignments of PMC-act with the highest score match. An alignment of PMC-act with gamma non-muscle actin of *O. cuniculus* (accession: X60733.1) showed 90 % identity (Figure 13). PMC-act was translated and aligned with the highest score match from Blast X, actin of *Artemia sp.* (accession: P18603). Two muscle actin cDNAs of *P. monodon* reported earlier from our laboratory in 1998 (70) were also compared at the amino acid level.

Multiple amino acid sequence alignment of actin is shown in figure 14. An identity of hemocyte PMC-act with actin of *Artemia sp.* (99 %) was higher than those compared with muscular *act1* of *P. monodon* (92 %) and *act2* of *P. monodon* (86 %). These suggested that the PMC-act sequence expressed in the shrimp hemocytes was a different type of actin from the abdominal muscle sequences. Some actin amino sequences with various subtypes were retrieved from an online record. Multiple sequence alignment was performed to identify the type of actin in PMC-act. An alignment in figure 15 showed that PMC-act had higher similarity with actin in subtype beta. Base on sequence similarity only, PMC-act might be classified into a beta actin cDNA sequence.

ESTPMC0002 was another sequence of interest since its product relates to a signal transduction process. The clone was identified as a putative RAN GTP-binding protein (Ras-related nuclear protein). The full insert length was obtained by sequencing of the clone PMC0002 from two directions, 5' and 3'. Each direction was confirmed by performing the sequencing reaction twice. A full-length cDNA insert of 1,053 bp with a long run of poly A^+ (19 letters) was observed. Homology searching results by BlastN and BlastX are shown in figure 16. A nucleotide sequence identity between the full-length PMC0002 and RAN GTP-binding protein of *Xenopus laevis*, the highest score match from BlastN, was 78 % (Figure 17). An identity between the amino sequence of PMC0002 and RAN GTP-binding protein of *D. melanogaster*, the highest score match from Blast X, was 86 % (Figure 18). Comparing to RAN GTP-binding protein of various organisms in table 10, PMC0002 sequence can be translated into a predicted protein sequence of 215 amino residues, like those from

Salmo salar and *Danio rerio*, whereas some other eukaryotes including Drosophila have 216 amino residues instead.

Two types of clone identities from this second library, which might be involved in a defense function, were the serine proteinase inhibitors (71) and the peroxinectin (24, 25, 72).

A group of 6 ESTs (ESTPMC0023, ESTPMC0246, ESTPMC0299, ESTPMC0347, ESTPMC0396, and ESTPMC0411) was identified as the putative serine proteinase inhibitors. The BlastX result and the feature of amino acid sequences translated from these clones showed the characteristic domains in serine proteinase inhibitor of the Kazal family. The multiple alignment with a Kazal type serine proteinase inhibitor from *Pacifastacus leniusculus* (accession: S45677) was shown in figure 19. The repeated stretches of conserved regions from an alignment were underlined. PMC0023 seemed to be the most truncated cDNA sequence because the first conserved region of PMC0023 is likely to be the third conserved region instead, when compared to others. It can assumed from an alignment that the putative Kazal type serine proteinase inhibitors of *P. monodon* have at least 5 Kazal domains.

An EST sequence of ESTPMC0265 was identified as putative peroxinectin. Peroxinectin was a cell adhesion molecule, which was found to be important in invertebrate immunity, especially in encapsulation and phagocytosis processes (24, 25, 72). The truncation at 5' end of ESTPMC0265 was found for several hundred residues as shown in figure 20.

Although many a number of truncated cDNAs was found, more than 120 novel cDNA sequences which represented around 60 different identities of *P. monodon* were identified from this study.

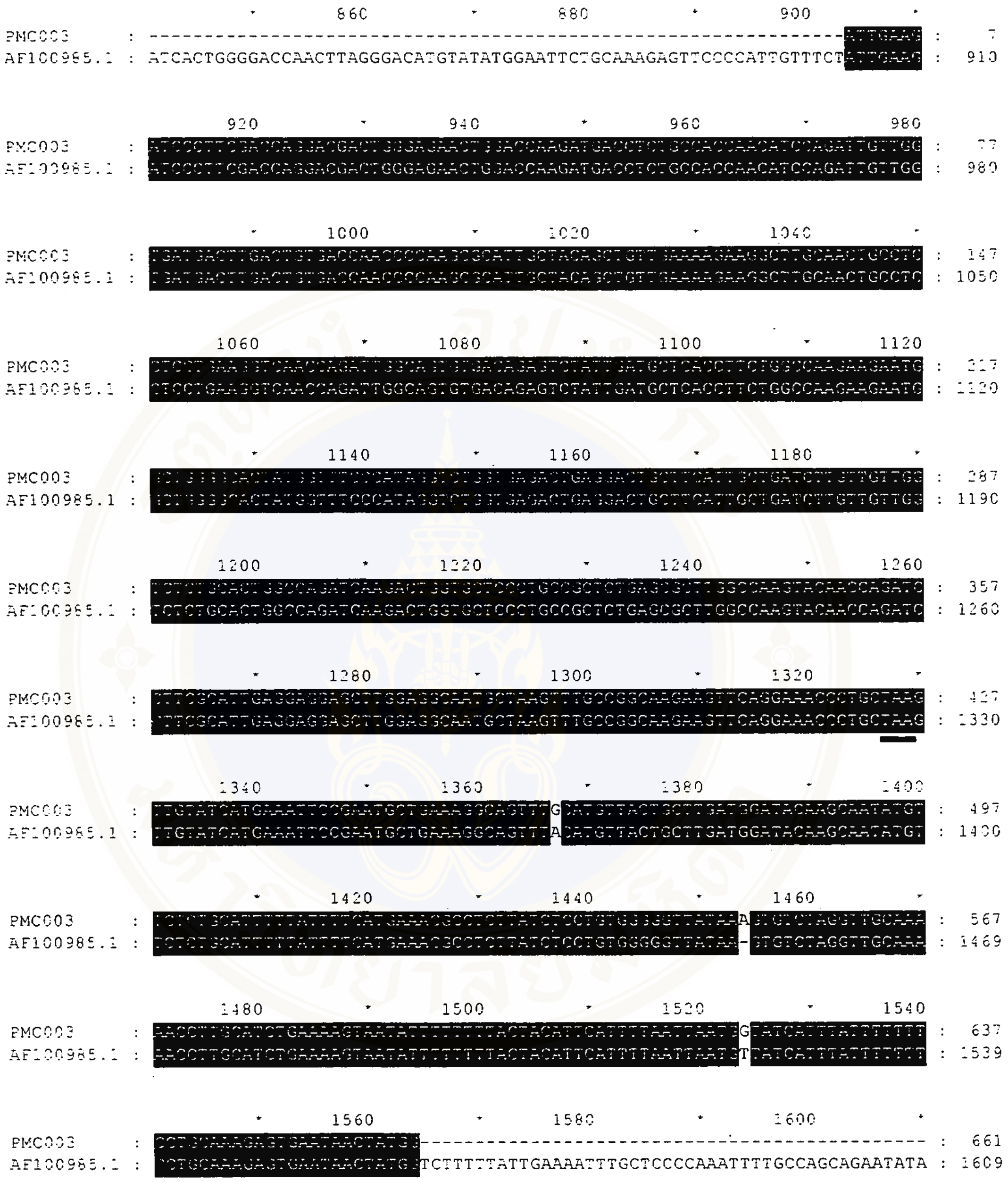
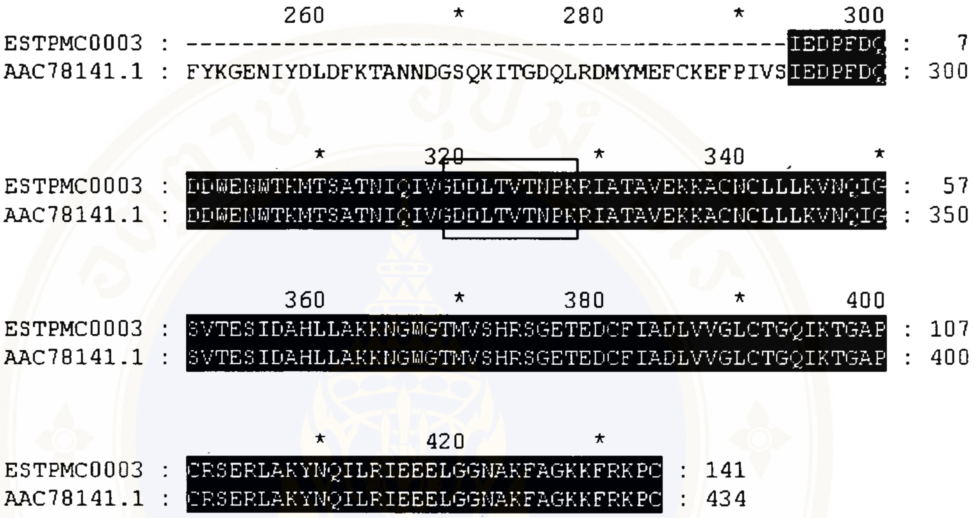


Figure 9 DNA sequence alignment between ESTPMC0003 and phosphopyruvate hydratase of *P. monodon* (Accession: AF100985) previously reported in 1998. The sequence of a newly isolated hemocyte cDNA clone, PMC0003, was compared with part of the muscular cDNA sequence, clone PMM020, coding for phosphopyruvate hydratase (or enolase) earlier reported from our laboratory. Clustal X was used. An overall identity of 100 % in ORF was found. Stop codon was underlined. Only the consensus region was shown.



```
                    260         *       280         *       300
ESTPMC0003 : -------------------------------------------IEDPFDQ :   7
AAC78141.1 : FYKGENIYDLDFKTANNDGSQKITGDQLRDMYMEFCKEFPIVSIEDPFDQ : 300

                 *            320         *       340         *
ESTPMC0003 : DDWENWTKMTSATNIQIVGDDLTVTNPKRIATAVEKKACNCLLLKVNQIG :  57
AAC78141.1 : DDWENWTKMTSATNIQIVGDDLTVTNPKRIATAVEKKACNCLLLKVNQIG : 350

                    360         *       380         *       400
ESTPMC0003 : SVTESIDAHLLAKKNGWGTMVSHRSGETEDCFIADLVVGLCTGQIKTGAP : 107
AAC78141.1 : SVTESIDAHLLAKKNGWGTMVSHRSGETEDCFIADLVVGLCTGQIKTGAP : 400

                 *            420         *
ESTPMC0003 : CRSERLAKYNQILRIEEELGGNAKFAGKKFRKPC : 141
AAC78141.1 : CRSERLAKYNQILRIEEELGGNAKFAGKKFRKPC : 434
```

Figure 10 Amino acid sequence alignment between those of ESTPMC0003 and phosphopyruvate hydratase of *P. monodon* (Accession : AAC78141.1). Deduced partial amino acid sequence of the hemocyte PMC0003 clone was compared with the deduced amino acid sequence from our lab's sequences, as retrieved from Genbank database. Clustal X was used. An overall identity in the aligned region was 100 %. The known motif specific to this protein, DDLTVTNPK, is boxed.

Figure 11 A scheme showing relative overlapping region among 8 hemocyte actin clones. Eight ESTs from 8 different cDNA clones, identified as non-muscle type actin from Blast searches and Clustal X alignments, were aligned and found to possess overlapping DNA region of the same transcript. A consensus sequence of all the clones were thus determined. The consensus sequence obtained from multiple sequence alignment (PMC-act) of all hemocyte actin segments was around 1.1 kb, with the presence of poly (A) in 3'end.

(a)

```
                                                                 Score    E
Sequences producing significant alignments:                      (bits)Value

emb|X60733.1|OCRNAGNMA  O.cuniculus mRNA for gamma-non muscl...  1219   0.0
gb|M18826.1|DROACT8F  D.melanogaster actin gene, complete cd...  1195   0.0
gb|AF035774.1|AF035774  Equus caballus beta actin mRNA, comp...  1187   0.0
gb|M26111.1|GOOACTB  Goose beta-actin mRNA, complete cds          1174   0.0
ref|NM_001101.2|  Homo sapiens actin, beta (ACTB) mRNA            1132   0.0
emb|X00351.1|HSAC07  Human mRNA for beta-actin                    1132   0.0
```

(b)

```
                                                                 Score    E
Sequences producing significant alignments:                      (bits)Value

sp|P18603|ACT4_ARTSX  ACTIN, CLONE 403 >gi|102728|pir||S1145...   667   0.0
sp|Q27250|ACT4_BOMMO  ACTIN, CYTOPLASMIC A4 (ACTIN A3B) (ACT...   667   0.0
gb|AAC27796.1|  (AF079161) cytoplasmic beta actin [Xenopus l...   666   0.0
gb|AAF26678.1|  (AF168615) beta-actin [Rivulus marmoratus]         665   0.0
sp|P53484|ACT1_FUGRU  ACTIN, CYTOPLASMIC 1 (BETA-ACTIN 1) >g...   665   0.0
gb|AAF63665.1|  (AF135499) beta-actin [Platichthys flesus]         665   0.0
```

Figure 12 Identity searching results of the actin consensus sequence. (a) By using Blast N program and **(b)** By using Blast X program of NCBI. Only part of the results with the highest scores are shown.

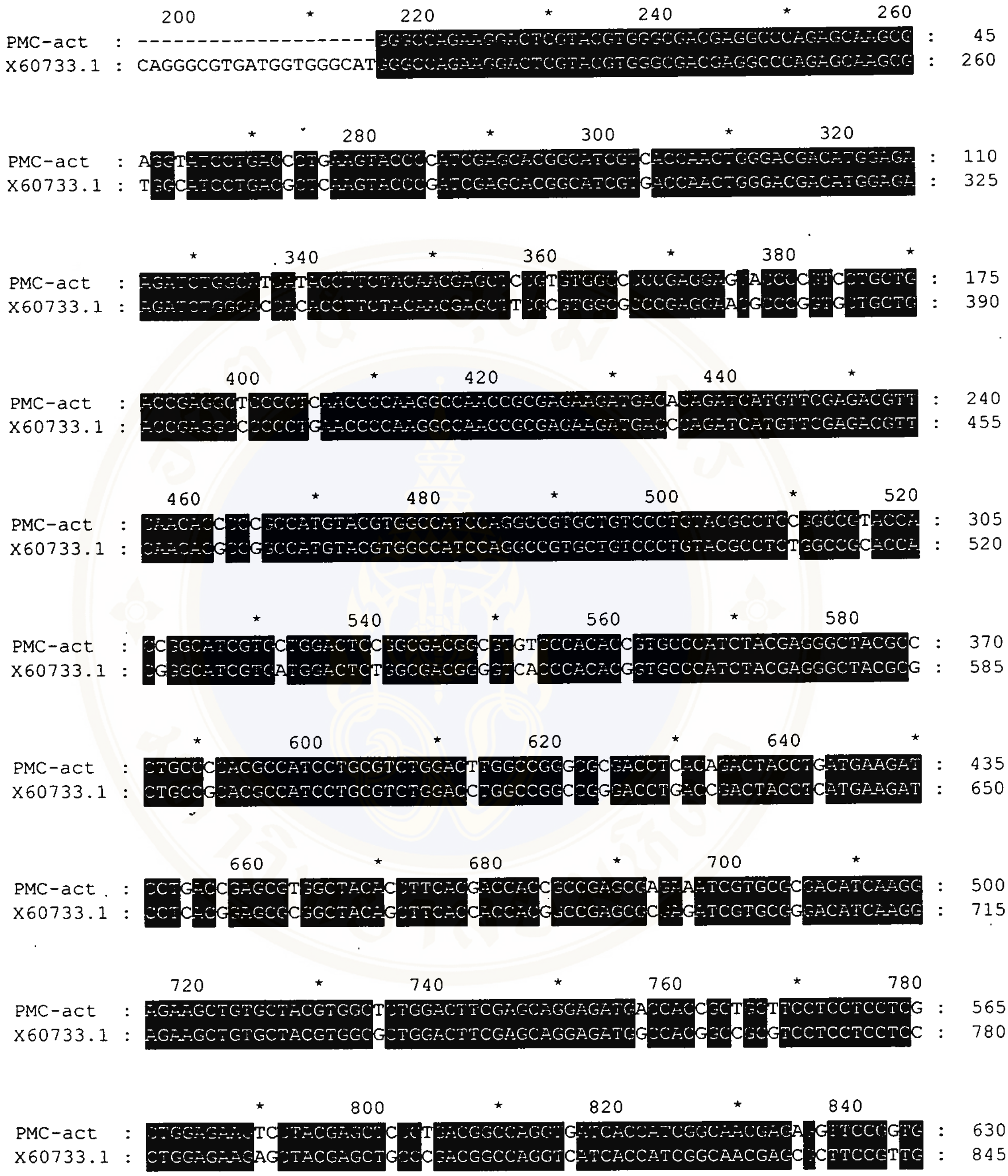


Figure 13 DNA sequence alignment between the hemocyte actin consensus sequence and a gamma (non-muscle) actin of *Oryctolagus cuniculus* (Accession: X60733.1). Clustal X was used. An overall identity of 90 % in an overlapping region within the ORF was found. Stop codon was underlined. Only the consensus region was shown.

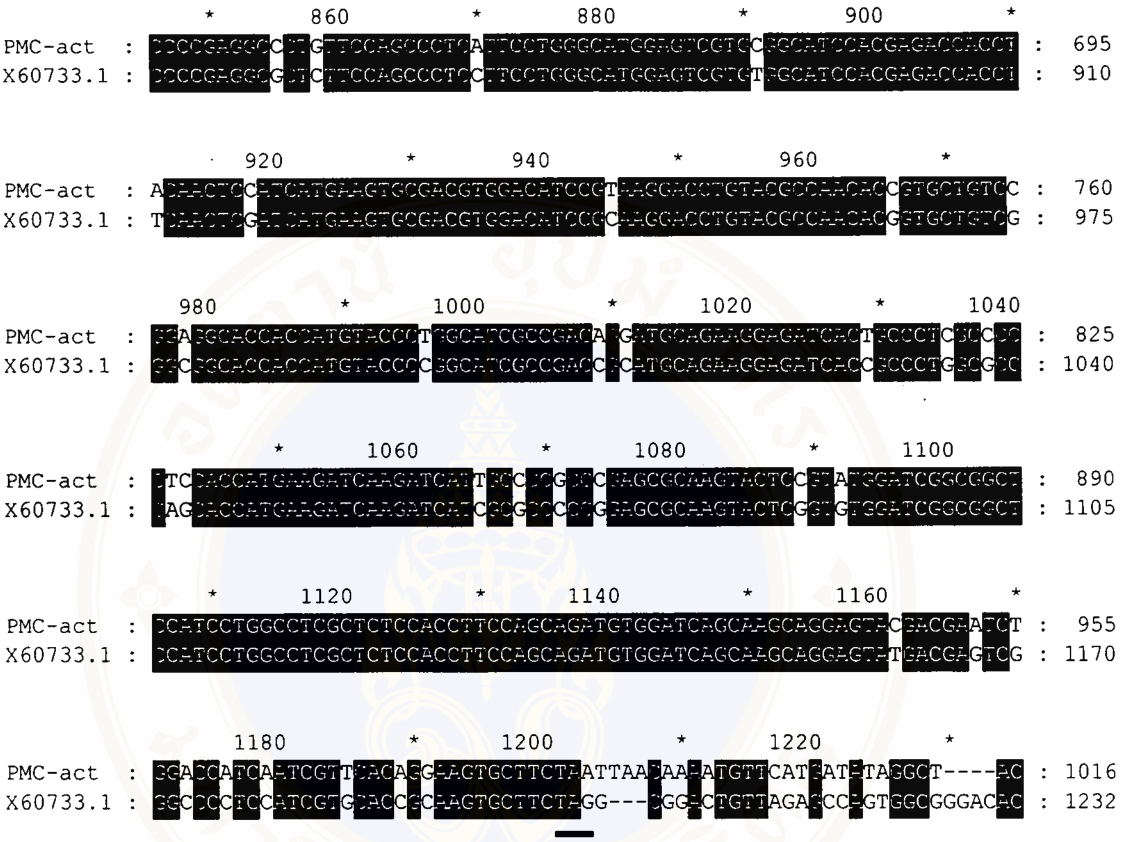


Figure 13 (continued)

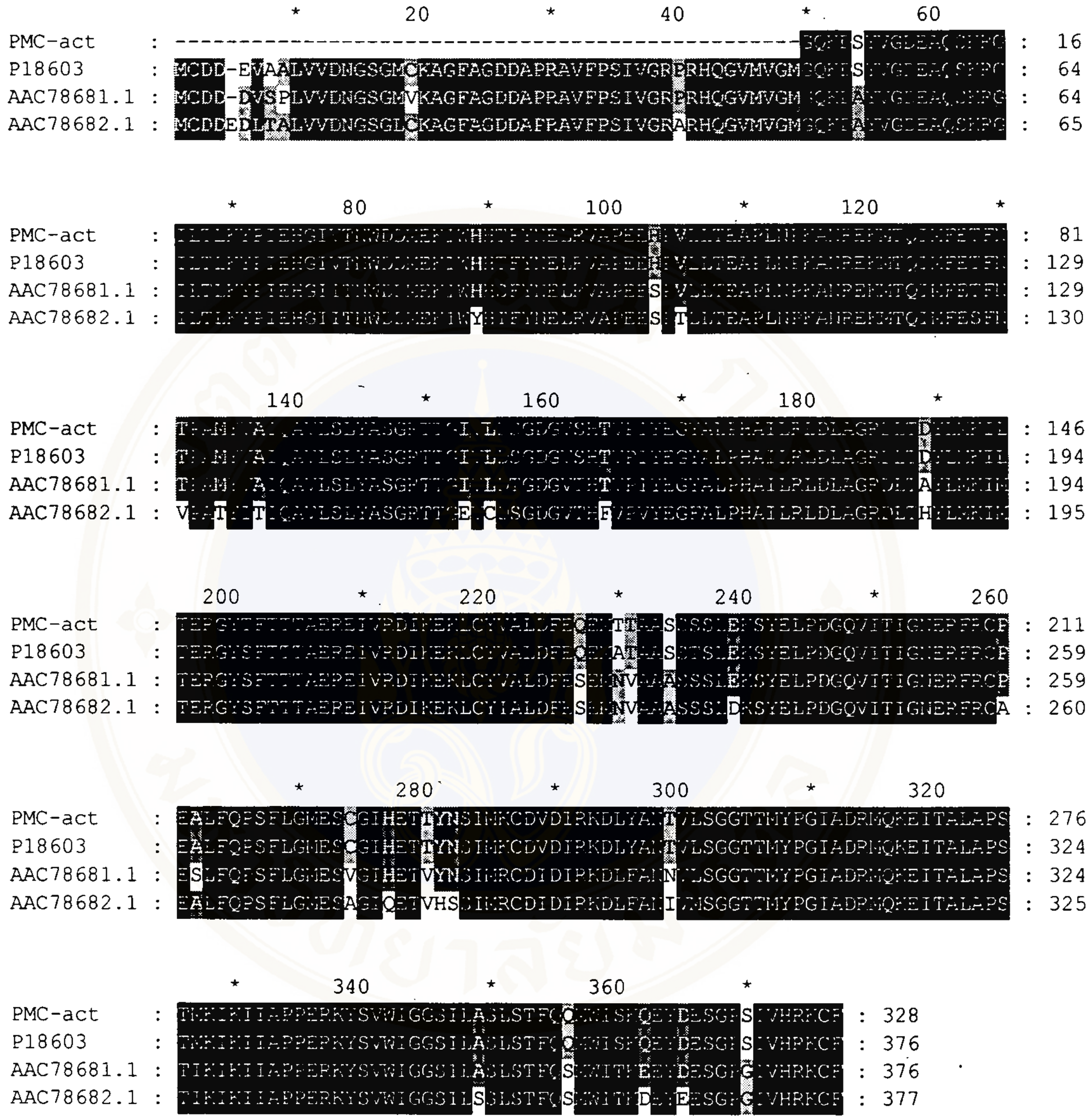


Figure 14 Multiple amino acid sequence alignment of PMC-act with those of actins from *Artemia sp.* (accession: P18603), our laboratory's muscle *act1* (AAC78681.1), and *act2 sequence* from *P. monodon* (AAC78682.1). Clustal X was used. Identities of PMC-act with those sequences were 99 %, 92 % and 86 %, respectively.

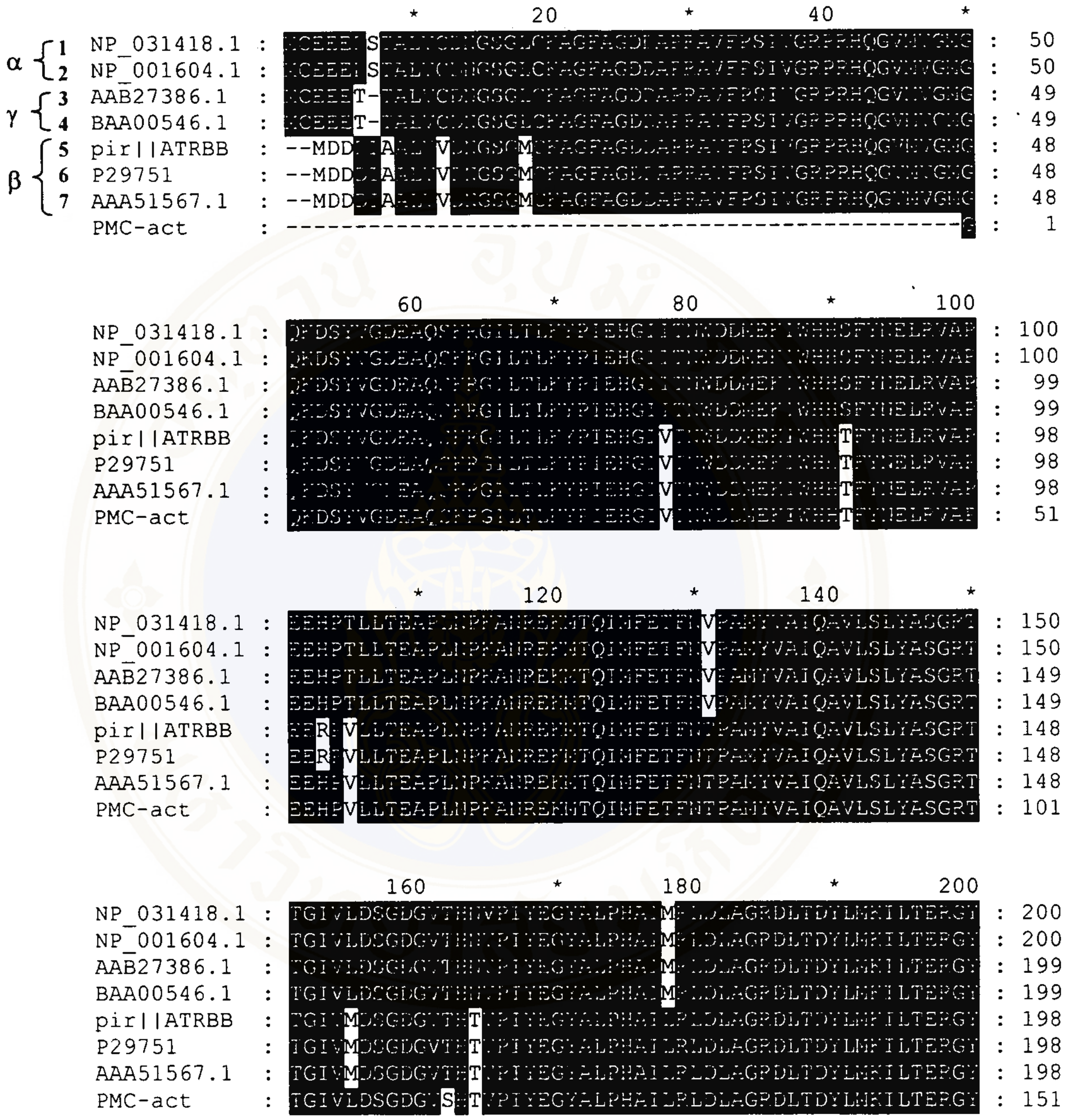
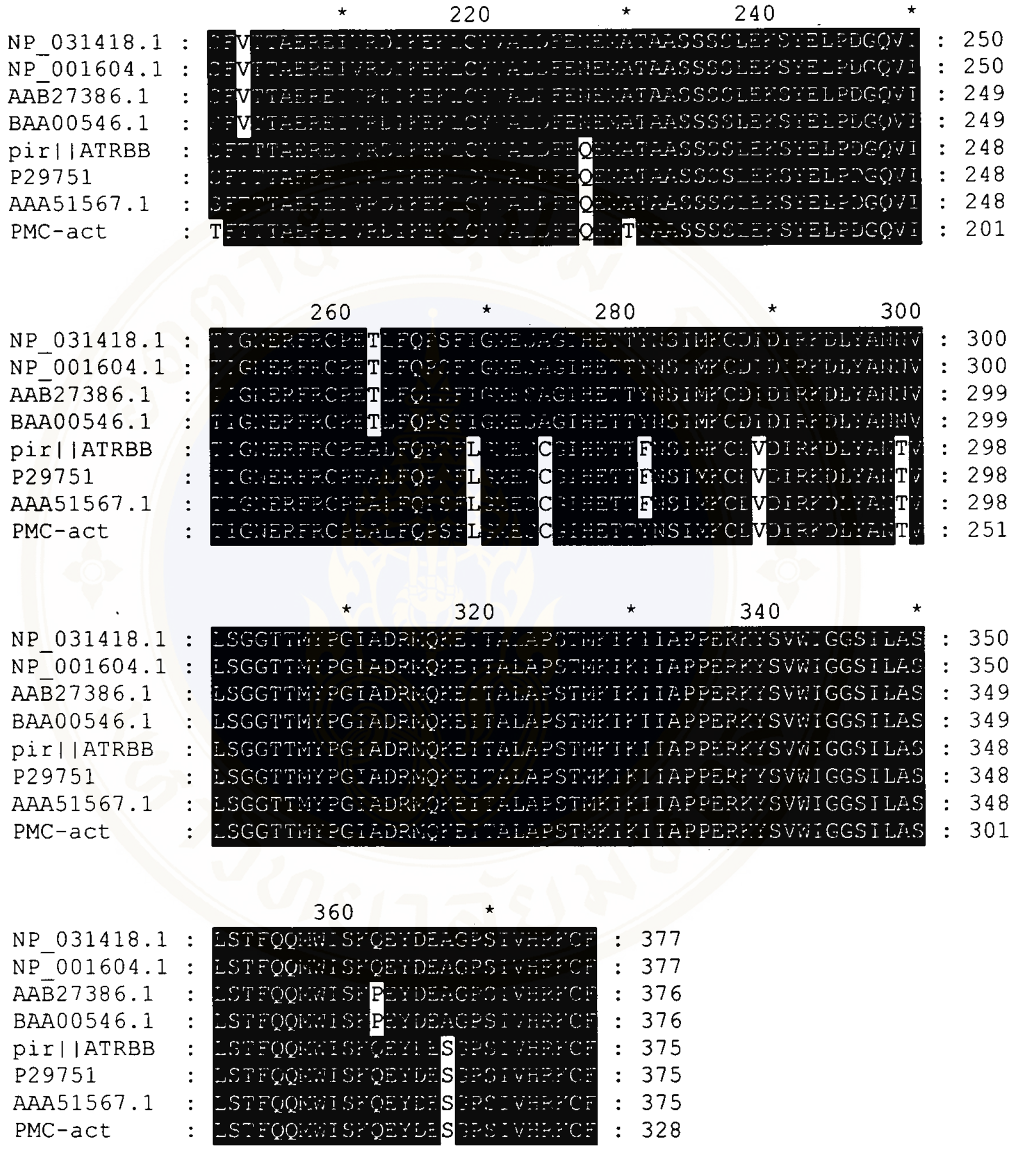


Figure 15 Multiple amino acid sequence alignment of PMC-act with selected actin of subtypes alpha (α), beta (β), and gamma (γ). Sequence 1-7 are from *M. musculus, H. sapiens, G. gallus, H. sapiens, O. cuniculus, O. cuniculus,* and *H. sapiens*; respectively. The sequence identities of PMC-act with α-actin was 94 %, whereas those with β and γ-actin were 94 % and 98 %, respectively.



```
                          *       220          *       240          *
NP_031418.1  : SFVTTAEREIVRDIKEKLCYVALDFENEMATAASSSSLEKSYELPDGQVI : 250
NP_001604.1  : SFVTTAEREIVRDIKEKLCYVALDFENEMATAASSSSLEKSYELPDGQVI : 250
AAB27386.1   : SFVTTAEREIVRDIKEKLCYVALDFENEMATAASSSSLEKSYELPDGQVI : 249
BAA00546.1   : SFVTTAEREIVRDIKEKLCYVALDFENEMATAASSSSLEKSYELPDGQVI : 249
pir||ATRBB   : SFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSYELPDGQVI : 248
P29751       : SFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSYELPDGQVI : 248
AAA51567.1   : SFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSYELPDGQVI : 248
PMC-act      : TFTTTAEREIVRDIKEKLCYVALDFEQEMTTAASSSSLEKSYELPDGQVI : 201

                    260          *       280          *       300
NP_031418.1  : TIGNERFRCPETLFQPSFIGMESAGIHETTYNSIMKCDIDIRKDLYANNV : 300
NP_001604.1  : TIGNERFRCPETLFQPSFIGMESAGIHETTYNSIMKCDIDIRKDLYANNV : 300
AAB27386.1   : TIGNERFRCPETLFQPSFIGMESAGIHETTYNSIMKCDIDIRKDLYANNV : 299
BAA00546.1   : TIGNERFRCPETLFQPSFIGMESAGIHETTYNSIMKCDIDIRKDLYANNV : 299
pir||ATRBB   : TIGNERFRCPEALFQPSFLGMESCGIHETTFNSIMKCDVDIRKDLYANTV : 298
P29751       : TIGNERFRCPEALFQPSFLGMESCGIHETTFNSIMKCDVDIRKDLYANTV : 298
AAA51567.1   : TIGNERFRCPEALFQPSFLGMESCGIHETTFNSIMKCDVDIRKDLYANTV : 298
PMC-act      : TIGNERFRCPEALFQPSFLGMESCGIHETTYNSIMKCDVDIRKDLYANTV : 251

                      *       320          *       340          *
NP_031418.1  : LSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILAS : 350
NP_001604.1  : LSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILAS : 350
AAB27386.1   : LSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILAS : 349
BAA00546.1   : LSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILAS : 349
pir||ATRBB   : LSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILAS : 348
P29751       : LSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILAS : 348
AAA51567.1   : LSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILAS : 348
PMC-act      : LSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILAS : 301

                    360          *
NP_031418.1  : LSTFQQMWISKQEYDEAGPSIVHRKCF : 377
NP_001604.1  : LSTFQQMWISKQEYDEAGPSIVHRKCF : 377
AAB27386.1   : LSTFQQMWISKPEYDEAGPSIVHRKCF : 376
BAA00546.1   : LSTFQQMWISKPEYDEAGPSIVHRKCF : 376
pir||ATRBB   : LSTFQQMWISKQEYDESGPSIVHRKCF : 375
P29751       : LSTFQQMWISKQEYDESGPSIVHRKCF : 375
AAA51567.1   : LSTFQQMWISKQEYDESGPSIVHRKCF : 375
PMC-act      : LSTFQQMWISKQEYLESGPSIVHRKCF : 328
```

Figure 15 (continued)

(a)

```
                                                               Score    E
Sequences producing significant alignments:                    (bits) Value

dbj|AB030945.1|AB030945  Xenopus laevis Xran mRNA for ran GT...   309  1e-81
emb|X66906.1|GGRLPA  G.gallus mRNA for ras-like protein           299  1e-78
gb|AF159256.1|AF159256  Mus musculus Lps/Ran GTPase (lps(d))...   297  4e-78
gb|S83456.1|S83456  GTP-binding protein [mice, C3H/HeJ splee...   297  4e-78
gb|L32751.1|MUSM1RAN  Mouse (clone M1) GTPase (Ran) mRNA, co...   297  4e-78
gb|AF054183.1|AF054183  Homo sapiens GTP binding protein mRN...   289  1e-75
```

(b)

```
                                                               Score    E
Sequences producing significant alignments:                    (bits) Value

gb|AAF48008.1|  (AE003485) CG1404 gene product [Drosophila m...   372  e-102
dbj|BAA89696.1|  (AB030945) ran GTP-binding protein [Xenopus...   361  4e-99
sp|P17080|RAN_HUMAN  GTP-BINDING NUCLEAR PROTEIN RAN (TC4)        360  1e-98
sp|P42558|RAN_CHICK  GTP-BINDING NUCLEAR PROTEIN RAN (TC4) >...   360  1e-98
ref|NP_006316.1||  ras-related nuclear protein; RAN, member ...   360  1e-98
pdb|1QG4|A  Chain A, Canine Gdp-Ran F72y Mutant >gi|5542357|...   360  2e-98
```

Figure 16 Identity searching results of the ESTPMC0002. (a) By using Blast N program and **(b)** By using Blast X program of NCBI. Only few hits with high scores are shown.

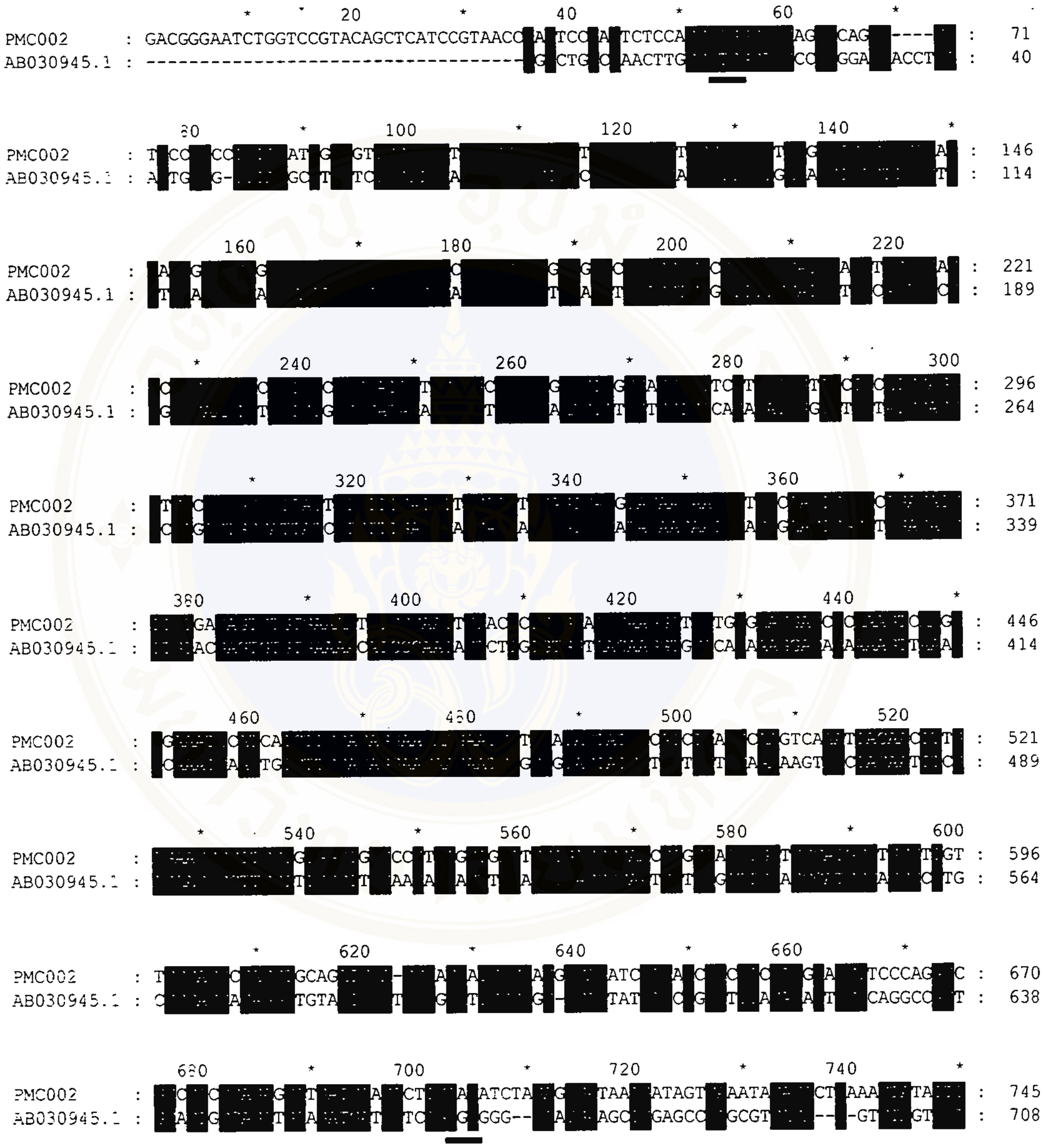


Figure 17 Alignment of PMC0002 full-length DNA sequence with that of Ran GTP-binding protein of *Xenopus laevis* (Accession: AB030945.1). Clustal X was used. An overall identity of 78 % in ORF was found. Start and stop codon were underlined. The full-length PMC0002 was 1072 bp : only 745 bp is shown.

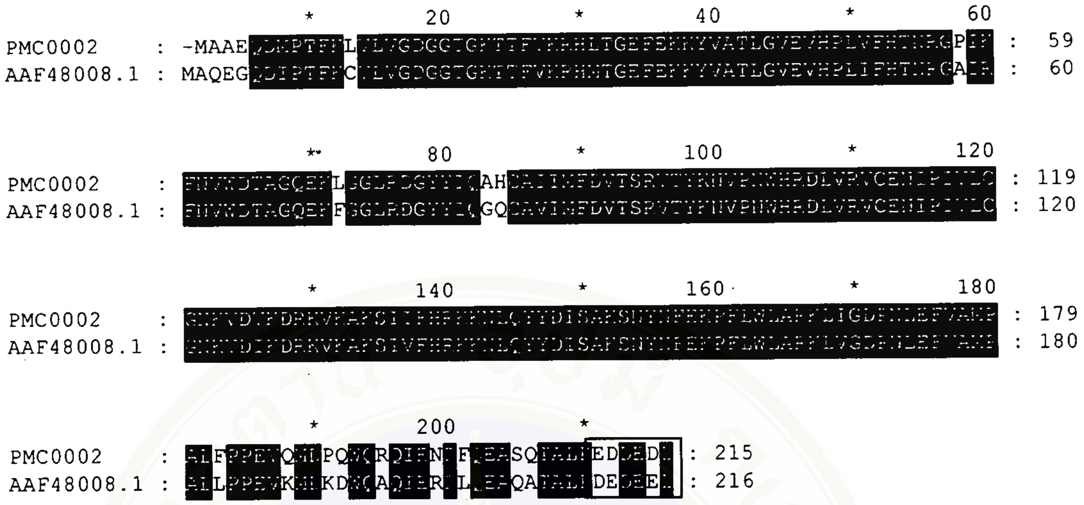


Figure 18 Amino acid sequence alignment of PMC0002 (deduced sequence) and Ran GTP-binding protein of *Drosophila melanogaster* (Acession: AAF48008.1). Clustal X was used. An overall identity of 86% was found, with a gap at the start codon. An acidic conserved C-terminus was boxed. A complete amino acid sequence length from PMC0002 was 215 residues.

Table 10: Comparison of the putative Ran GTP-binding protein (deduced amino acid sequence of PMC0002) with those from few other eukaryotic organisms.

ORGANISM	ACCESSION NO.	LENGTH (AA)	IDENTITY (%)	SIMILARITY (%)
Penaeus monodon	-	215	-	-
Xenopus laevis	BAA89696.1	216	87	91
Gallus gallus	P42558	216	87	91
Drosophila melanogaster	AAF48008.1	216	86	93
Homo sapiens	P17080	216	86	91
Mus musculus	NP_006316.1	216	86	91
Salmo salar	CAA10039.1	215	86	93
Danio rerio	P79735	215	86	92

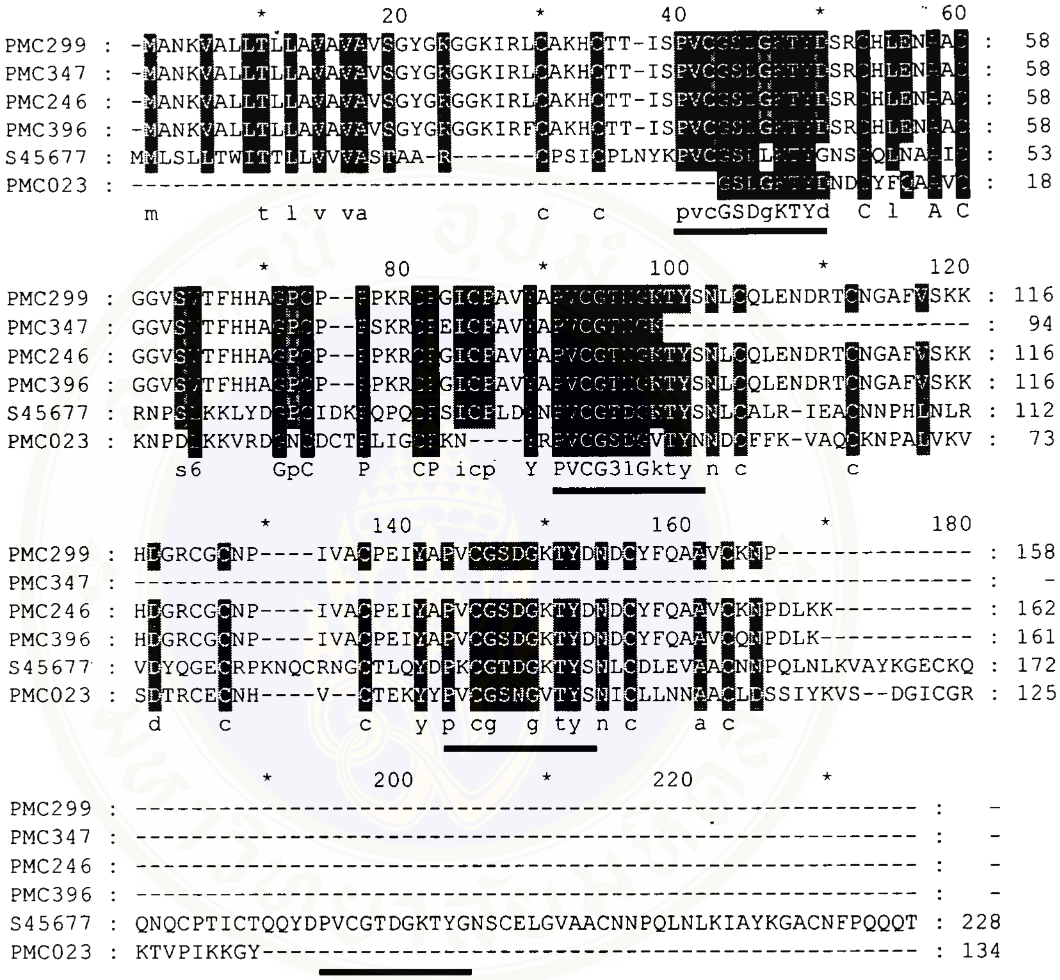


Figure 19 Multiple amino acid sequence alignment of 5 new putative Kazal type serine proteinase inhibitors with that from *Pacifastacus leniusculus* (accession: S45677). Deduced amino acid sequences from 5 partially sequenced cDNA clones (PMCxxxx) were aligned with a previously characterized sequence from crayfish. Four repeated stretches are underlined. The sequence of another clone, PMC0411, was not included because of its low quality nucleotide sequence.

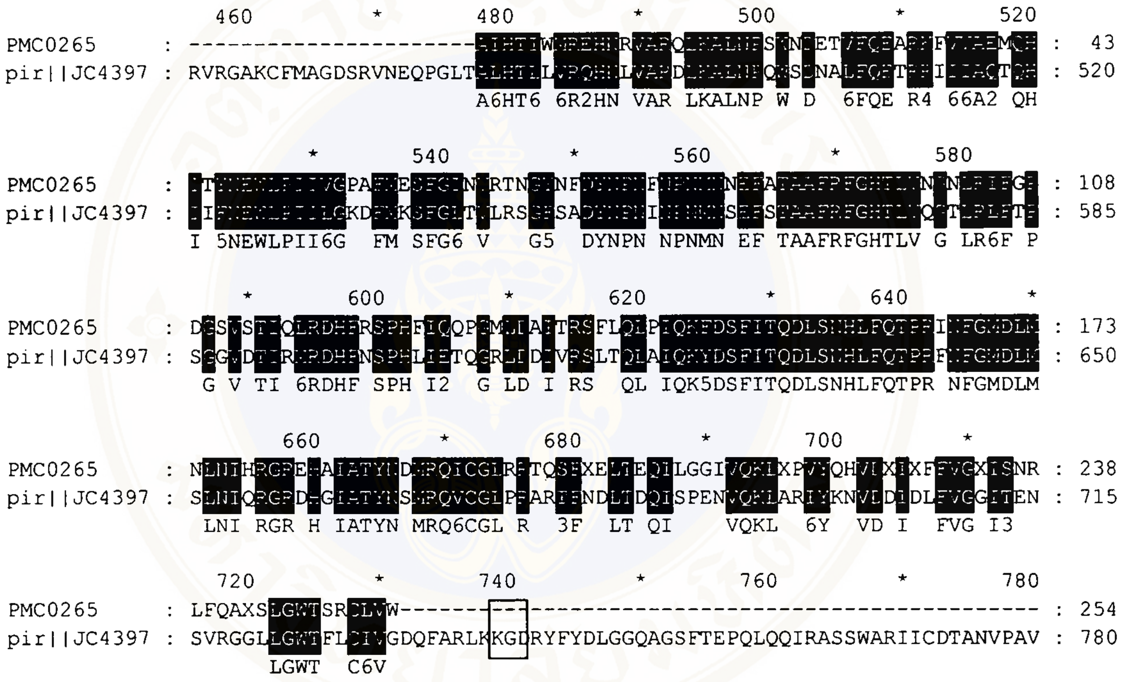


Figure 20 Amino acid sequence alignment of ESTPMC0265 with peroxinectin sequence of *Pacifastacus leniusculus* (accession: JC4397). Clustal X was used. An overall identity of 67 % and similarity 82 %, valued in the same length, was found. A putative cell adhesive and binding motif 'KGD' was boxed.

CHAPTER V

DISCUSSION

1. Total RNA isolation

Only one predominant RNA band at around 1.9 kb was detected on an EtBr staining formaldehyde denaturing gel. This band was likely 18S rRNA. The 28S rRNA and 18S rRNA of insects have been found to be around 3900-4200 bp and 1900-2500 bp in length, respectively (74-75). From figure 7, the 28S rRNA could not be observed. The same result was reported by Boonyawan B. and Wongsombat T. in total RNA preparation of other tissues from *P. monodon* (70, 76). This observation was also found in other insects (77, 78). They found that 28S rRNA can be cleaved by heating at 60 °C and converted into 18S rRNA size products which showed identical mobility to the native 18S rRNA on gel electrophoresis, although the 18 S thermal product is a little heavier than the native 18 S RNA by sucrose gradient centrifugation analysis(78). The heating step was routinely done for RNA denaturation before loading RNA sample into the gel and also was used in this work.

The yield of total RNA isolation in library no.1 was around 2 μg/ml of hemolymph, while in library no.2 was around 6 μg/ml of hemolymph. They were different in their isolation method. Total RNA was isolated from the whole content of hemolymph for library no.1, but it was isolated from solely hemocytes for library no.2. This might be suggested that the centrifugation of total hemolymph in the presence of anticoagulant to sediment hemocytes before total RNA isolation is a better method than isolated from the whole hemolymph as done in library no.1.

2. Poly (A^+) RNA isolation

The poly (A^+) RNA isolation of library no.1, by using an oligo-dT cellulose column, generated the yield of around 0.11 % of total RNA. The better yield of 0.22 % of total RNA was derived in library no.2, by using PolyATtract® mRNA isolation system. This might be the result from a very small amount of 168 μg of total RNA used in library no.1, while the total RNA of library no.2 was 1.7 mg (> around 10 folds). The less of total RNA used as starting material, leads to larger proportion of poly (A^+) RNA loss during isolation. However, it seems not suitable to compare between the yields from two different isolation methods.

3. The quality of cDNA libraries.

A first cDNA library constructed from total hemolymph of *P. monodon* (library no.1) was found not suitable for EST survey, due to a presence of high number of clones with mitochondrial DNA, *E. coli*, and lambda DNA, at around ¾ of the library (72 %) as shown in table 9. The low quality of the cDNA library may be resulted from the very low amount of mRNA for the library construction at only 0.18 μg, which was below an optimized condition for a cDNA synthesis as suggested by manufacturer. The manufacturer's protocol suggests the minimum at 1.5 μg and the maximum at 7.0 μg of mRNA. We decided to construct this library after many small batches of sample were collected, to try the run-through process. Then, second library was cautiously planned to construct with some modification processes, especially in the hemocyte sample preparation, and the amount of starting material -the total RNA.

A second cDNA library of total hemocytes of *P. monodon* was constructed from around 3.83 μg of mRNA. This cDNA library contained an approximately 10^6

independent primary lambda clones, that was the expected titer normally found in an unamplified cDNA library (40, 41, 70). Screening by restriction enzyme digestion, around 50 % of the recombinants could be cleaved at the ligated sites, *Eco*R I and *Xho* I. The uncleavable clones, even have an insert, were ignored from further analysis. To roughly determine the quality of the hemocyte library (library no.2) by multiplying the titer of this library (1.7 x 10^6) with 50 % and dividing by the redundancy of 8 (the most redundant which was found in identical actins), approximately 10^5 genes were represented. The unidirectional orientation of cDNA at 98 % can facilitate the sequencing analysis, which the 5' and 3' of cDNA could be easily determined. Without significant contaminating sequences from other organisms, including *E. coli*, and lambda vector DNA, it was considered a good cDNA library. Some clones of mitochondrial DNAs (13.9 %) could be found, which may resulted from the high A+T content of the *P. monodon* mitochondrial genome (>70 %). The mitochondrial RNA molecules may bind to the oligo (dT) used to isolate the poly(A^+) RNA. A study of other tissue-specific ESTs from *P. monodon* revealed the incorporation of mitochondrial rRNA and other mitochondrial genes at 8.9 % in cephalothorax, 1.8 % in eyestalk, and 59.4 % in pleopod (40). From figure 8, most of cDNA inserts have the size in the range of 0.6-1.0 kb, but they are mainly at around 1.0 kb that is a minimum average size recommended by Adams *et al* (68). A 92 % of cDNA clones containing an insert size ≥ 500 bp, represented a good size-selection library. The 5' truncation of many cDNA clones in this library might be resulted from partial degradation during RNA preparation or the presence of some nuclease activity in the cDNA synthesis reaction. The sequencing of overlapping clones corresponding to different parts of one messenger can be done and allowed the reconstitution of a complete mRNA sequence by contig construction from the partial sequences (82) as in figure 11. Moreover, a full coding sequence of each unique truncated clone can be observed by using 3' or 5' RACE (83). Although EST data was obtained from 5'-sequencing, some of them can be sequenced until the poly A tail. This suggests the completion of the 3'-end of cDNA inserts. The truncation of a cDNA clone at the 5'-end is usually found in a population of a cDNA library. Some approaches could be done to avoid this major drawback such as using a biotinylated CAP trapper (85-88).

4. ESTs data analysis

Similarity searching was conducted by using BlastN and BlastX programs of NCBI on the WWW. The sequence similarity of ESTs with those in online databases were mostly found by BlastX. This is because BlastX compares the query sequence against amino acid sequence database and, in general, protein sequence motifs tend to be more conserved across a wide-range of taxa. At least 120 ESTs of the *P. monodon* hemocytes could be assigned putative identities. For those of 71 ESTs (36.5 %) without significant sequence matching, they might belong to genes whose sequences have diverged considerably among different organisms or they represented completely novel functions in crustaceans. Some might consist mainly of untranslated mRNA regions that are poorly conserved during the course of evolution. Many of the unknown sequences might provide interesting information in the future when we have accumulated more molecular information on the shrimp into the database.

The frequently found ESTs (table 8) are generally considered to represent mRNA sequences from highly expressed genes that play vital biological roles in the cell. The frequency of a cDNA identity provides a relative estimation of the steady-state level of a transcript of that gene compare to others in the particular tissue. In table 8, with the exception of ribosomal proteins and elongation factors, which were composed of various sub-populations involved in protein translation, actins and serine proteinase inhibitors were the redundant transcript expressed in the shrimp hemocytes.

From the sequence identity searching of PMC-act (the consensus sequence of the hemocyte actin clones) in figure 12 and sequence alignment in figure 13, 14 and 15, PMC-act should be classified from the sequence similarity as β-actin. Both β or γ actin classes are the major actins present in cytoplasm and muscle of invertebrates. Among crustaceans, four actin isoforms have been isolated from *Artemia* during its development (84). Actin serves important functions in the maintenance of cell shape, cell division, cell motility, organelle movement and signal transduction. They are ubiquitously expressed in all eukaryotic cells and are a group of highly conserved proteins. Nevertheless, they are produced from distinct genes in a stage and tissue-specific manner. It had been studied in vitro that relating to microbial rocketing motility (96). This might be the reason of its redundancy in hemocytes.

The discovery of EST for putative serine proteinase inhibitor was quite promising. Serine proteinase inhibitor is considered an important component in an arthropod immunity (71). Its likely function in arthropod hemolymph is in protecting the host cells from infection by pathogens or parasites. Some may inhibit fungal or bacterial proteinases. Others probably have role in regulating endogenous proteinases involved in coagulation, prophenol oxidase activation, or cytokine activation (71). In crustaceans, there is a cascade involved in the recognition of microbial polysaccharides called prophenoloxidase activating system. It is so sensitive that picograms per litre of microbial polysaccharides can activate the system (23). The zymogenic prophenoloxidase is activated by proteinases and hence proteinase inhibitors are key regulatory factors to avoid inappropriate activation of the system.

An EST for a putative peroxinectin (ESTPMC0265) was also found from this library. Protein of this gene is the first cell adhesion molecule ever purified from an invertebrate blood, a crayfish named *Pacifastacus leniusculus*. It was found to stimulate encapsulation and endocytosis (24, 25, 72). Peroxinectin is synthesized in the blood cells, stored in granules (89), and released as an inactivated form during degranulation. It can be activated in the presence of LPS or β-1,3-glucans.

In addition to the putative serine proteinase inhibitor and peroxinectin that are relevant to defense system of crustacean, some other putative identities might be indirectly involved in an immune response. Clathrin is important for endocytosis by forming polyhedral structure (73). Ubiquitin and 26 S proteasome are responsible for specific cellular protein degradation (93, 94), involving cell cycle progression, differentiation, and modulation of the immune and inflammatory responses (94). Chaperonin assists in folding of actin, tubulin, and some other cytosolic proteins (95). It was also found to be involved in scarless wound healing, may be by preventing the forming of stress fiber of alpha-smooth muscle actin (97). In contrast, the polymerization of actin with other small number of proteins is required for the reconstitution of microbial rocketing motility in vitro without the presence of myosin (96). Calreticulin, a calcium binding protein found in non-skeleton muscle cells, is known to bind to the molecules of innate immunity, such as C1q, the first subcomponent of complement (98).

Other EST identities such as Ran, ferritin, serine/threonine protein kinase are important in signal transduction pathway.

For the putative Ran GTP-binding protein, ESTPMC0002, this sequence was derived from a gene which constitutes a distinct branch of the superfamily of Ras-related GTP-binding protein, which function as molecular switches cycling between GTP-bound 'on' and GDP-bound 'off' states. Ran protein is located predominantly in the nuclei of eukaryotic cells and is involved in the nuclear import of proteins and export of RNA as well as in control of DNA synthesis and of cell-cycle progression (90, 91). The protein Ran was first cloned in human because of its high sequence similarity to the oncogene Ras. Most of this sequence similarity is found within conserved domains involved in guanine nucleotide binding. Unlike many G-proteins, mammalian Ran does not appear to undergo any post-translation lipid modifications and probably as a result does not bind membranes inside the cell or require lipids for its activity. In human, Ran possesses an acidic–DEDDDL at its C terminus that is conserved across species instead of a consensus prenylation domain (CAAX motif: C = cysteine, A = aliphatic amino acid, X = any amino acid). For the ESTPMC0002, the –EDDEDL was also found at C terminus while –DEDDDL was found in *D. melanogaster*. (Figure 18)

We have found various interesting ESTs from the second hemocyte cDNA library. This suggests a good representation for cDNA clones from this specific tissue. All ESTs described in this study have been submitted to dbEST division of Genbank. A number of 216 DNA sequences were added to the Ganbank record, 22 from total hemolymph library and 194 from total hemocyte library. All of known identities from library no.2 were compared to all existing online records of *P. monodon*'s sequences, at least 60 identities were newly reported. The novel identities such as : Ran GTP-binding protein, ferritin subunit precusor, calreticulin, ubiquitin extension protein, polyA binding protein, 26 S proteasome, proteasome regulator, heterogeneous nuclear RNP, RNA polymerase II, small nuclear ribonucleorotein, multicatalytic endopeptidase complex, ADP/ATP translocase, equilbrative nitrobenzlythioinosine-sensitive nucleoside transporter, chromatin assembly factor, SNR1, cytosolic actin, profilin, tubulin, thymosin, calponin, coactosin, golgi apparatus associated protein, selenoprotein, serine proteinase inhibitor, antileukoproteinase, clathrin associated protein, clathrin light chain, peroxinectin, zinc finger protein, T-complex protein (chaperonin), ADP-ribosylation factor, ATPase, anti-oxidant protein, glucose-6-phosphate isomerase, methylmalonyl-CoA mutase, phosphoglyceromutase, adenosylhomocysteinase, selenophosphate synthetase, pyruvate kinase, nucleoside diphosphate kinase, etc. In addition to these, many ribosomal proteins and elongation factors were newly reported. The ESTs data set will serve as molecular biological information background of *P. monodon*. Some of them might be useful in other fields of studies. However, the identity of ESTs assigned by this study were only deduced from high sequence similarity but the true functions can only be confirmed or demonstrated by biochemical and mutagenesis approaches.

5. Full length characterization of two muscle cDNA clones.

During the course of this work, insert sequence of PMM088 and PMM110 earlier described by Boonyawan was also completely determined by the author using a religation method (Appendix 1), as an additional work. The method is one convenient approach to avoid many complicated steps in subcloning, such as vector preparation and DNA fragment preparation. More than two religation reactions might be needed for a clone with an insert larger than 1.5 kb because there was a limitation on reading distances by the automated sequencer, approximately 500 bp in length could be reliable read per one sequencing reaction.

The truncation of 47 bp (figure 22) from the presumed start codon in the PMM088 sequence might be the result from degradation of RNA prior to the abdominal library preparation. The presence of a minute amount of nuclease in the reverse transcription could also lead to the generation of incomplete cDNA. The difference of nucleotides in some position when compare to a recently reported cytochrome b sequence of *Penaeus monodon* from Australia in the database, indicated a genetic variation between different population of the shrimps. This may be useful serving as a DNA marker for population tracking study.

At present, there was no other nucleotide sequence of calcium binding proteins in online databases, except that of an EST of *P. monodon* from another laboratory (AI253941). However, an alignment in figure 24 showed the 5'-UTR truncation but an open reading frame was complete. This 192 amino residue protein might be involved in signal transduction pathway whereas the three Ca^{2+} binding sites in figure 25 were reported to have high affinity to Ca^{2+} (92). In a cell, the concentration of cytosolic Ca^{2+} is normally maintained at very low (around 0.1 μM) by Ca^{2+} pump that actively export Ca^{2+} across the plasma membrane, mitochondria and endoplasmic reticulum. The intracellular Ca^{2+} will increase in response to some hormonal, neural or other stimuli through specific Ca^{2+} channels. The cellular responses such as Ca^{2+}-dependent enzymes will be activated and calcium binding proteins will play their role in this step of regulating those activity (73).

These two full-length sequences have not been submitted to Genbank, but will be subsequently done in the near future.

CHAPTER VI

CONCLUSION

1. Two lambda ZAPII-based cDNA libraries were constructed from *Penaeus monodon*'s hemocytes. The first cDNA library was constructed from a small amount of total RNA (around 0.18 μg), isolated from snap-frozen total hemolymph of live shrimps. Sequences of 67 cDNA clones were partially determined from their 5' end. Searching online databases revealed that up to 72 % were contaminating of mitochondrial DNA, *E. coli* DNA, and lambda DNA. This library is not suitable for EST research work. Nevertheless, 22 novel DNA sequences of *P. monodon* were obtained and submitted to Genbank.

2. The second cDNA library was constructed by using a larger amount of total RNA (around 3.83 μg), isolated from centrifuged hemocytes of live shrimps. A total of 194 cDNA clones were partially sequenced from their 5' end. Around 49.4 % of ESTs were found to match known genes that were not mitochondrial DNA, while 36.5 % did not match any sequences in current databases. No contamination of *E. coli* or lambda DNA was detected from this library. This library was consideredsuitable for EST research work and gene isolation. All 194 sequences have been submitted to Genbank.

3. Two groups of EST identities - serine proteinase inhibitors and the peroxinectin, from the hemocyte library # 2 were found by chance and sequenced. They were assumed to involve in defense mechanism, especially in the prophenoloxidase activating system. Some other ESTs which might be indirectly involved in immunity

were also found, such as: clathrin, calreticulin, 26 S proteasome protein, ubiquitin, etc. Some ESTs which play their roles in signal transduction pathway such as: Ran, serine/threonine kinase, ferritin, etc., was also found. Other ESTs identified constituted a broad range type of transcripts which are cytoplasmic structural and regulatory proteins, enzymes, nuclear proteins, and a number of proteins of unknown functions.

REFERENCES

1. Baily-Brock JH, Moss SM. *Penaeid* taxonomy, biology and zoogeography, In: Marine Shrimp Culture: Principle and Practices (Fast AW, Lester LJ, ed.). Elsevier Science Publishers, Amsterdam. 1992; 9-23.
2. Food and Agriculture Organization of the United nations' homepage, 2000 (www.fao.org)
3. Chamberlain G. Shrimp whitespot virus confirmed in Central America. From Global Aquaculture Alliance's homepage. (www.gaalliance.org)
4. Kasornchandra J, Boonyaratapalin S, Khongpradit R, and Akpanithanpong U. A bacilliform virus, the causative agent of red disease with white patch in black tiger shrimp (*Penaeus monodon*). Technical paper No.3/1995, National Institute of Coastal Aquaculture, Kaoseng, Songkla, Thailand, 11.
5. Wongteerasupaya C, Vickers JE, Sriurairatana S, *et al.* A non-occluded, systemic baculovirus that occurs in cells of ectodermal and mesodermal origin and causes high mortality in the black tiger prawn, *Penaeus monodon.* Dis. Aquat. Org. 1995; 21: 69-77.
6. Chou HY, Huang CY, Wang CH, Chiang HC, Lo CF. Pathogenicity of a baculovirus infection causing white spot syndrome in cultured penaeid shrimp in Taiwan. Dis. Aquat. Org. 1995; 23: 165-173.
7. Limsuwan C. What kind of white spot kills shrimp ? AAHRI Newsletter Article 1997; 6 (2).

8. Chantanachooklin C, Boonyaratpalin S, Kasornchandra J, *et al.* Histology and ultrastructure reveal a new granulosis-like virus in *Penaeus monodon* affected by "yellow head" disease. Dis. Aquat. Org. 1993; 17: 145-57.

9. Boonyaratpalin S, Supamattaya K, Kasornchandra J, Direkbusaracom S, Aekpanithapong U, and Chantanachooklin C. Non-occluded baculo-like virus, the causative agent of yellow head disease in the black tiger shrimp (*Penaeus monodon*). Fish. Pathol. 1993, 28: 103-9.

10. Wongteerasupaya C, Sriurairatana S, Vickers JE, *et al.* Yellow-head virus of *Penaeus monodon* is an RNA virus. Dis. Aquat. Org. 1995, 22 (1): 45-50.

11. Supamattaya K, Hoffmann RW, Boonyaratpalin S and Kanchanaphum P. Experimental Transmission of White Spot Baculovirus (WSBV) from Black Tiger Shrimp (*Penaeus monodon*) to the Portunid Crab (*Portunus pelagicus*), Mud Crab (*Scylla serrata*) and Krill (*Acetes sp.*) Dis. Aquat. Org. 1998; 32(2): 79-85.

12. Kanchanaphum P, Wongteerasupaya C, Sitidilokratana N, *et al* TW. Experimental transmission of white spot syndrome virus (WSSV) from crabs to shrimp *Penaeus monodon*. Aquat. Org. 1998; 34: 1-7.

13. Lightner, DV. A handbook of pathology and diagnostic procedures for diseases of penaeid shrimp, section 3 : viruses, published by The World Aquaculture Society, USA. 1996.

14. Lo CF, Ho CH, Chen CH, *et al.* Detection and tissue tropism of white spot syndrome baculovirus (WSBV) in captured brooders of *Penaeus*

monodon with a special emphasis on reproductive organs. Dis. Aquat. Org. 1997; 30: 53-72.

15. Nash G, Arkarjamorn A, and Withyachumnarnkul V. Rapid diagnosis of Yellow-head disease in black tiger shrimp. AAHRI Newsletter Article 1993; 2 (1).

16. Marl J, Bonami JR, and Lightner DV. Taura syndrome of penaeid shrimp: cloning of viral genome fragments and development of specific gene probes. Dis. Aquat. Org. 1998; 33: 11-17.

17. Tang KF, Lightner DV. A yellow head virus gene probe: nucleotide sequence and application for *in situ* hybridization. Dis. Aquat. Org. 1999; 35 (3): 165-73.

18. Cowley JA, Dimmock CM, Spann KM, and Walker PJ. Detection of Australian gill-associated virus (GAV) and lymphoid organ virus (LOV) of *Penaeus monodon* by RT-nested PCR. Dis. Aquat. Org. 2000; 39 (3): 159-67.

19. Limsuwan, C. Reducing the effects of white-spot baculovirus using PCR screening and stressors. AAHRI Newsletter Article 1997; 6(1).

20. Söderhäll K, Thörnqvist P-O. Crustacean immunity. A short review. Dev. Biol. Stand. 1997; 90: 45-51.

21. Holmblad T, Söderhäll K. Cell adhesion molecules and antioxidative enzymes in a crustacean, possible role in immunity. Aquaculture 1999; 172: 111-23.

22. Bachère E, Mialhe E, Noël D, Boulo V, Morvan A, Rodriguez J. Knowledge and research prospects in marine mollusc and crustacean immunology. Aquaculture 1995; 132: 17-32.

23. Söderhäll K, Cerenius L. Role of the prophenoloxidase-activating system in invertebrate immunity. Curr. Opin. Immunol. 1998; 10: 23-28.

24. Johansson MW, Lind MI, Holmblad T, Thörnqvist P-O, Söderhäll K. Peroxinectin, a novel cell adhesion protein from crayfish blood. Biochem. Biophys. Res. Commun. 1995; 216: 1079-87.

25. Johansson MW. Cell adhesion molecules in invertebrate immunity. Dev. Comp. Immunol. 1999; 23: 303-15.

26. Boman HG. Peptide antibiotics and their role in innate immunity. Annu. Rev. Immunol. 1995; 13: 61-92.

27. Hoffmann JA, Reichhart JM, and Hetru C. Innate immunity in higher insects. Curr. Opin. Immunol. 1996; 8: 8-13.

28. Hoffmann JA, and Reichhart JM. Drosophila immunity (Review). Trends in cell biology 1997; 7: 309-16.

29. Anderson KV. Toll signaling pathways in the innate immune response. Curr. Opin. Immunol. 2000; 12: 13-19.

30. Kopp EB, and Medzhitov R. The Toll-receptor family and control of innate immunity. Curr. Opin. Immunol. 1999; 11: 13-18.

31. Vargas-Albores F, Jimenez-Vega F, and Söderhäll K. A plasma protein isolated from brown shrimp (*Penaeus califoriensis*) which enhances the activation of prophenoloxidase system by β–1,3-glucan. Dev. Comp. Immunol. 1996; 20: 299-306.

32. Rodriguez J, Boulo V, Mialhe E, and Bachère E. Characterisation of shrimp haemocytes and plasma components by monoclonal antibodies. J. Cell Sci. 1995, 108: 1043-50.

33. Yeh MS, Huang CJ, Leu JH, Lee YC, and Tsai IH. Molecular cloning and characterization of a hemolymph clottable protein from tiger shrimp (*Penaeus monodon*). Eur. J. Biochem. 1999; 266: 624-33.

34. Song YL, and Hsieh YT. Immunostimulation of tiger shrimp (*Penaeus monodon*) hemocytes for generation of microbicidal substances: analysis of reactive oxygen species. Dev. Comp. Immunol. 1994; 18: 201-9.

35. Destoumieux D, Bulet P, Loew D, Van Dorsselaer A, Rodriguez J, and Bachère E. Penaeidin: A new family of antimicrobial peptides in shrimp *Penaeus vannamei* (Decapoda). J. Biol. Chem. 1997; 272: 28398-406.

36. Vargas-Albores F, Guzman MA, and Ochoa JL. An anticoagulant solution for haemolymph collection and prophenoloxidase studies of penaeid shrimp (*Penaeus califoniensis*). Comp. Biochem. Physiol. 1993; 106A: 299-303.

37. Bachère E, Mialhe E, and Rodriguez J. Identification of defence effectors in the haemolymph of crustaceans with particular reference to the shrimp *Penaeus japonicus (Bate)*: prospects and applications. Fish & Shellfish Immunol. 1995; 5: 597-612.

38. Bachère E, Cedeno V, Rousseau C, *et al.* Transgenic crustaceans. In World Aquaculture 1997; 28: 51-55.

39. National Center for Biotechnology Information (NCBI) 's homepage. http://www.ncbi.nlm.nih.gov/

40. Lehnert SA, Wilson KJ, Byrne K, and Moore SS. Tissue-specific expressed sequence tags from Black tiger shrimp *Penaeus monodon*. Mar. Biotechnol. 1999; 1: 465-76.

41. Davoli R, Zambonelli P, Bigi L, Fontanesi L, and Russo V. Analysis of expressed sequence tags of porcine skeletal muscle. Gene. 1999; 233: 181-88.

42. Fan J, Minchella DJ, Day SR, McManus DP, Tiu WU, and Brindley PJ. Generation, identification, and evaluation of expressed sequence tags from different developmental stages of the Asian blood fluke *Schistosoma japonicum*. Biochem. Biophys. Res. Commun. 1998; 252 (2): 348-56.

43. Gargioni R, and Barracco MA. Hemocytes of the Palaemonids *Macrobrachium rosenbergii* and *M. acanthurus*, and the Penaeid *Penaeus paulensis*. J. Morphol. 1998; 236: 209-21.

44. Johansson MW, and Söderhäll K. Exocytosis of the prophenoloxidase activating system from crayfish haemocytes. Journal of Comparative Physiology. 1985; 156: 175-81.

45. Martin GG, and Graves L. Fine structure and classification of shrimp hemocytes. J. Morphol. 1985; 185: 339-48.

46. Adams MD, Kelley JM, Gocayne JD, *et al.* Complementary DNA sequencing: Expressed sequence tags and human genome project. Science. 1991; 252: 1651-6.

47. Wolfsberg TG, Landsman D. A comparison of expressed sequence tags (ESTs) to human genomic sequences. Nucleic Acids Res. 1997; 25(8): 1626-1632.

48. Schuler GD, Boguski MS, Stewart EA, *et al.* A gene map of the human genome. Science 1996; 274 (5287): 540-46.

49. Marra MA, Hillier L, and Waterston RH. Expressed sequence tags – ESTablishing bridges between genomes. TIG 1998; 14 (1): 4-7.

50. Anderson I, Brass A. Searching DNA databases for similarities to DNA sequences: when is a match significant ? Bioinformatics 1998; 14: 349-56.

51. Gong Z, Yan T, Liao J, Lee SE, He J and Hew CL. Rapid identification and isolation of zebrafish cDNA clones. Gene 1997; 201: 87-98.

52. Ghosh S. A novel ligation mediated-PCR based strategy for construction of subtraction libraries from limiting amount of mRNA. Nucleic acid res. 1996; 24(4): 795-796.

53. Rivolta MN and Wilcox ER. A novel and simple methodology to generate subtracted cDNA libraries. Nucleic Acids Res. 1995; 23(13): 2565-2566.

54. Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ. Basic local alignment search tool. J. Mol. Biol. 1990; 215: 403-10.

55. Benson DA, Boguski MS, Lipman DJ, Ostell J, and Francis BF. Genbank. Nucleic Acid Res. 1998; 26(1): 1-7.

56. Karlin, Samuel and Altschul SF. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc. Natl. Acad. Sci. (USA) 1990; 87: 5873-7.

57. Thompson JD, Higgins DG, and Gibson TJ. CLUSTAL W : Improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22(22): 4673-80.

58. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, and Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multi alignment aided by quality analysis tools. Nucleic Acids Res. 1997; 25 (24): 4876-82.

59. Motoh H. Biology and ecology of *Penaeus monodon.* In proceeding of the first international conference on the culture of penaeid prawns/shrimps (Taki Y, Premavara JH, and Lobera J, eds.) Aquaculture department, Southeast Asia Fisheries Development Center. 1985.

60. Yanisch-Perron C, Vieira J, Messing J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 1985; 33: 105-19.

61. Stratagene. Instruction manual for cDNA synthesis kit, ZAP cDNA® synthesis kit, and ZAP cDNA® gigapack® III gold cloning kit: Preparation of media and reagents. 1998, p 45-48.

62. Chomczynski P, Sacchi N. Single step method of RNA isolation by acid guanidinium thiocyanate-chloroform extraction. Anal. Biochem. 1987; 161: 156-9.

63. Owens L, O'neill A. Use of a clinical cell flow cytometer for differential counts of prawn *Penaeus monodon* haemocytes. Dis. Aquat. Org. 1997, 31 (2): 147-153.

64. Sambrook J, Fritsch EF, and Maniatis T. Molecular cloning, Laboratory manual. New York: Cold Spring Harbor laboratory, 1989.

65. Promega. Technical manual : PolyATtract® mRNA isolation systems. USA. Revised 4/1995.

66. Del Sal G, Manfioletti G, Schneider C. A one-tube plasmid DNA mini preparation suitable for sequencing. Nucleic Acids Res. 1988; 16: 9878.

67. Boguski MS, Lowe TMJ, and Tolstoshev SH. dbEST- database for 'expressed sequence tags'. Nature Genet. 1993; 4: 332-33.

68. Adams MD., et al. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature. 1995; 377(suppl): 3-17.

69. Boonchuoy C, Boonyawan B, Panyim S, and Sonthayanon B. A cDNA sequence of phosphopyruvate hydratase (enolase) from Black Tiger Prawn *Penaeus monodon*. Asia Pac. J. Mol. Biol. Biotechnol. 1999; 7(1): 89-94.

70. Boonyawan B. Sequence characterization of randomly isolated cDNA from black tiger prawn (*Penaeus monodon*). M.Sc. (Molecular Genetics-Genetic Engineering) thesis 1998, Institute of Molecular Biology and Genetics, Mahidol University.

71. Kanost MR. Serine proteinase inhibitors in arthropod immunity. Dev. Comp. Immunol. 1999; 23: 291-301.

72. Johansson MW. Peroxinectin, a cell-adhesive peroxidase. Plant Peroxidase Newsletter 1997; 9: 3-5.

73. Lehninger AL, Nelson DL, Cox MM. Principle of biochemistry, second edition (1993). Worth publishers, Inc. USA.

74. Tautz D, Hancock JM, Webb DA, and Dover GA. Complete sequences of the rRNA genes of *Drosophila melanogaster*. Mol. Biol. Evol. 1988; 5(4): 366-76.

75. Amako D, Kwon OY, and Ishikawa H. Nucleotide sequence and presumed secondary structure of the 28S rRNA of pea aphid: implication for diversification of insect rRNA. J. Mol. Evol. 1996; 43(5): 469-75.

76. Wongsombat T. Molecular cloning of a cDNA encoding neuropeptide hormone in the CHH/MIH/GIH family from *Penaeus monodon*. M. Sc. (Molecular genetics-genetic engineering) thesis 2000, Institute of Molecular Biology and Genetics, Mahidol University.

77. Greenberg JR. Synthesis and properties of ribosomal RNA in *Drosophila*. J. Mol. Biol. 1969; 46: 85-98.

78. Ishikawa H, and Newburgh RW. Studies of the thermal conversion of 28S RNA of *Galleria mellonella* (L.) to an 18S product. J. Mol. Biol. 1972; 64: 135-44.

79. Söderhäll K. Prophenoloxidase activating system and melanization – a recognition mechanism of arthropods ? Dev. Comp. Immunol. 1982; 6: 601-11.

80. Flegel TW. Special topic review : major viral diseases of the black tiger prawn (*Penaeus monodon*) in Thailand. World J. Microbiol. Biotechnol. 1997; 13: 433-42.

81. Flegel TW, Boonyaratpalin S, Withyachumnarnkul B. Progress in research on yellow-head virus and white spot virus in Thailand. In Flegel TW, MacRae I (eds) Disease in Asian aquaculture III, 1997. Asian Fisheries Soc, Manila, 285-96.

82. Cooke R, et al. Further progress towards a catalogue of all *Arabidopsis* genes: analysis of a set of 5000 non-redundant ESTs. The Plant Journal 1996; 9(1): 101-24.

83. Udomkit A, Chooluck S, Sonthayanon B, and Panyim S. Molecular cloning of a cDNA encoding a member of CHH/MIH/GIH family from *Penaeus monodon* and analysis of its gene structure. Journal of Experimental Marine Biology and Ecology. 2000; 244: 145-56.

84. Macias MT, ans Sastre L. Molecular cloning and expression of four actin isoforms during Artemia development. Nucleic Acids Res. 1990; 18(17): 5219-25.

85. Edery I, Chu LL, Sonenberg N, and Pelletier J. An efficient strategy to isolate full-length cDNAs based on an mRNA cap retention procedure (CAPture). Mol. Cell. Biol. 1995; 15(6): 3363-71.

86. Carninci P, Kvam C, Kitamura A, *et al*. High-efficiency full-length cDNA cloning by biotinylated CAP trapper. Genomics 1996; 37(3): 327-36.

87. Suzuki Y, Yoshitomo-Nakagawa K, Maruyama K, Suyama A, and Sugano S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene 1997; 200(1-2): 149-56.

88. Seki M, Carninci P, Nishiyama Y, Hayashizaki Y, and Shinozaki K. High-efficiency cloning of Arabidopsis full-length cDNA by biotinylated CAP trapper. Plant J. 1998; 15(5): 707-20.

89. Liang Z, Lindblad P, Beauvais A, Johansson MW, Latge J-P, Hall M, Cerenius L, and Söderhäll K. Crayfish α–macroglobulin and 76 kD protein; their biosynthesis and subcellular localization of the 76 kD protein. Journal of Insect Physiology. 1992; 38: 987-95.

90. Scheffzek K, Klebe C, Fritz-Wolf K, Kabsch W, and Wittinghofer A. Crystal structure of the nuclear Ras-related protein Ran in its GDP-bound form. Nature. 1995; 374(6520): 378-81.

91. Moore MS. Ran and nuclear transport. J. Biol. Chem. 1998; 273(36): 22857-60.

92. Takagi T, Konishi K. Amino acid sequence of α chain of sarcoplasmic calcium binding protein obtained from shrimp tail muscle. J. Biochem. 1984; 95: 1603-1615.

93. Kornitzer D, Ciechanover A. Modes of regulation of ubiquitin-mediated protein degradation. Journal of Cellular Physiology 2000; 182 (2): 1-11.

94. Ciechanover A, Orian A, Schwartz AL. The ubiquitin-mediated proteolytic pathway : Mode of action and clinical implications. J. Cell. Biochem. 2000; 77(S34): 40-51.

95. Yokota SI, Yanagi H, Yura T, Kubota H. Upregulation of cytosolic chaperonin CCT subunits during recovery from chemical stress that causes accumulation of unfolded proteins. Eur. J. Biochem. 2000; 267(6): 1658-64.

96. Borisy GG, Svitkita TM. Actin machinery : pushing the envelope. Curr. Opin. Cell. Biol. 2000; 12(1): 104-12.

97. Darden DL, Hu FZ, Ehrlich MD, *et al.* RNA differential display of scarless wound healing in fetal rabbit indicates downregulation of a CCT chaperonin subunit and upregulation of a glycophorin-like gene transcript. J. Pediatr. Surg. 2000; 35(3): 406-19.

98. Kishore U, Sontheimer RD, Sastry KN, *et al*. Release of calreticulin from neutrophils may alter C1q-mediated immune functions. Biochem J 1997; 322 (Pt 2): 543-50.

APPENDIX

Appendix 1 : Full length characterization of two muscle cDNA clones.

Two abdominal muscle cDNA clones earlier characterized and initially identified as those for a cytochrome b and a calcium binding protein based on BLAST search with their 5'-EST by Boonyawan (1998) (70), PMM088 and PMM110, were selected for full-length sequence determinations. Cytochrome b sequence could serve as a useful DNA marker for phylogenetic study of crustaceans, as well as for population migration study of shrimps in the seas. For the calcium binding protein, it might play a role in physiology of the shrimp muscles.

Using cycle sequencing protocol with dye terminator chemistry (PE Applied Biosystems), approximately 600 bases per sample were clearly readable by the data analysis software. After visually examining the chromatograms, an average length of usable sequences of around 500 base pair long (including vector sequences) was found. The two clones were sequenced from both 3'and 5' directions, using M13 universal forward and reverse primers. An overlap of 3' and 5' sequencing was earlier used to generate a full-length sequence by Contig Assembly program. However it was found difficult to accurately determine the sequence of several nucleotide bases in this overlapping region, due to their far distance from the primers used. Restriction maps of the two clones were then analyzed from the known DNA sequences by Vector NTI program. From the map, an internal region of PMM088 was then digested in two ways, one with *Eco*R I and *Nde* I double digestion and another with *Xho* I and *Nde* I double digestion, each was then re-ligated of its ends after a blunting reaction. The two re-ligated forms were subjected to sequencing for the middle region of the cDNA insert by using the same M13 universal reverse and forward primers and/or T3 and T7 primers instead. For PMM110, which has an internal *Eco*R I site in the cDNA insert, it was double digested with *Pst* I and *Hin*d III, and with *Xho* I and *Hin*d III. The site of *Pst* I is located beside the *Eco*R I within the polylinker region (figure 3). After blunting and re-ligating their ends, both re-ligated forms were subjected to sequencing in the same fashion as PMM088. All of sequencing data was then used for sequence alignments to find the consensus sequence, which became the representative full-length sequence of each clone after each base position was verified. The restriction map and sequencing directions for complete sequence determination of both clones is shown in figure 21.

Similarity searching on existing DNA sequence data in all available databases was conducted by using BLAST program. Multiple sequence alignments were later performed to compare the sequence identity of the sequence with those from other organisms, using Clustal X programs. Results of BLAST scores and percent sequence identities are shown in Table 11.

We found that the sequence of clone PMM088, 1090 bp long, was truncated around 16 amino acids (47 bp) from the presumed start codon (ATG) at the 5' end. The nucleotide sequence was compare either to a cytochrome b sequence of *Penaeus monodon*, which was reported by another laboratory, or a sequence of *Drosophila yakuba*. The 3' end was apparently complete due to the presence of the poly A^+ tail of the cDNA clone. However, the poly (A) tract found in the clone was located just right after the termination codon (TAA), which was quite unusual. If the missing 16 amino acid residues in the N terminus was assumed, the 378 amino acid polypeptide with a molecular mass of around 40 kDa could be predicted. The GC content within the ORF was 33 %. Similarity of PMM088 sequence to a recently published mitochondrial DNA sequence of *P. monodon* (AF217843) was 98 % (Figure 22). A 99 % amino acid sequence identity was found when compared to a translated protein (AAF43377) of the same mitochondrial DNA, with 1 amino acid difference at position 175 of the complete cytochrome b polypeptide (Figure 23). When compare with the *D. yakuba* sequences, the nucleotide sequence and amino acid sequence identities were 74 % and 77 %, respectively. The full-length sequence of PMM088 was also listed in Appendix 2 with omission of the poly (A) tract.

The sequence of PMM110, 993 bp long, could be translated into a 192 amino acid sequence of calcium binding protein. Alignment of the DNA sequence with a *P. monodon* EST of a putative calcium binding protein (AI253941) showed a 94 % nucleotide sequence identity, in a 622 nucleotide region (Figure 24). When the deduced amino acid sequence was compared to a sarcoplasmic calcium binding protein of *Penaeus sp.* (P02636), a 94 % protein sequence identity was found, with 6 residue differences in the 192 amino acid full-length sequence (Figure 25). Three putative Ca^{2+} binding sites were boxed (93). The full-length sequence of PMM110 was listed in Appendix 3, with omission of the poly (A) tract.

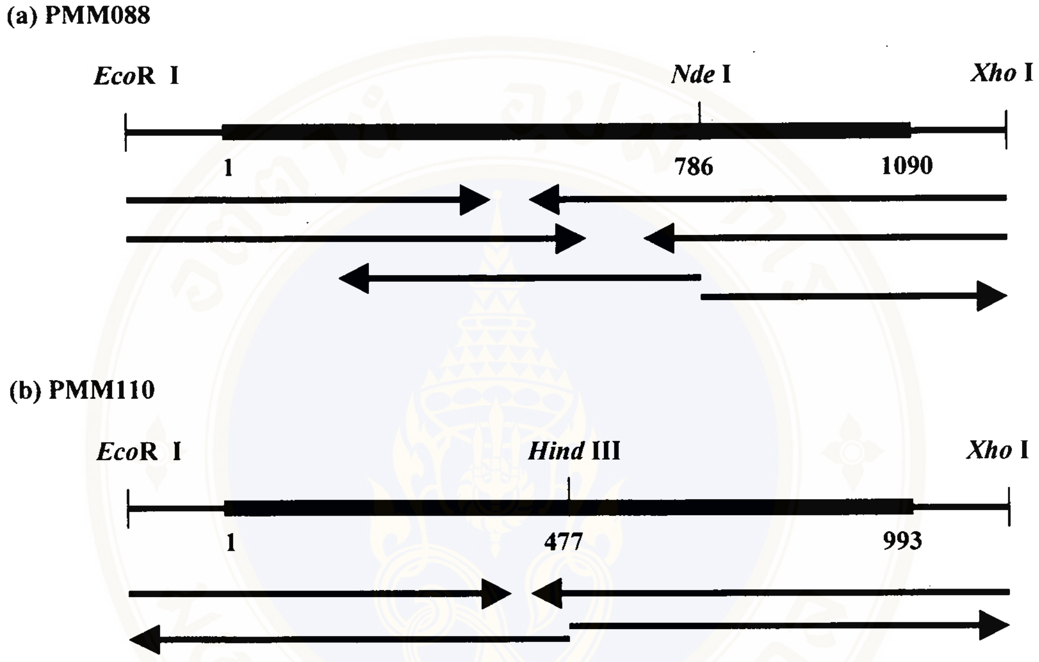


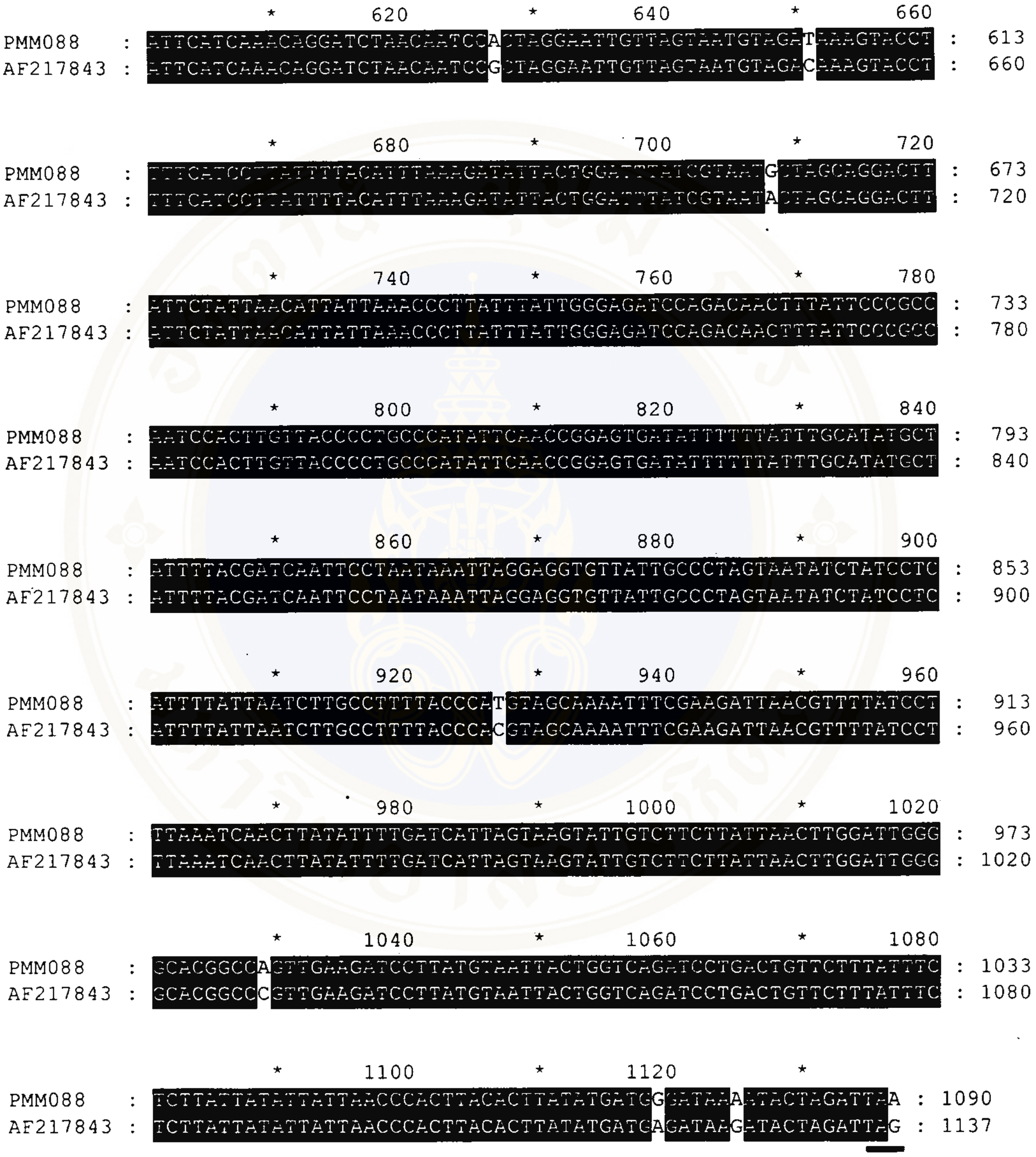
Figure 21 Restriction map and all sequencing directions of (a) PMM088 and (b) PMM110 cDNA fragments in this experiment. The region between *Eco*R I and position 1 is *Eco*R I adapter and the region between the end of cDNA, position 1090 in PMM088 and position 993 in PMM110, and *Xho* I is poly $(A)^+$.

Table 11 : Significant matching of PMM088 and PMM110 sequences with those from various online databases. BLASTN and BLASTX version 2.0.11, available via WWW from NCBI, were used to compare each full-length query sequence with records in databases. E value was the BLAST scores obtained. DNA sequence identity and protein similarity was obtained from alignment with Clustal X program. An overlapping region of homology was shown in parentheses as letters of identical per letters of compared region.

Clone Number	**level**	**Gene characterization from highest score**		**E value**	**Blast score**	**Sequence**	
		Gene/Organism	**Accession**			**Identity**	**Similarity**
PMM088 (1090 bp)	nt	Cytochrome b / *Penaeus monodon*	AF217843.1	0.0	2157	98 % (1077/1089)	-
		Cytochrome b / *Drosophila yakuba*	NC_001322.1	2e-34	153	74 % (806/1090)	-
	aa	Cytochrome b / *Penaeus monodon*	AAF43377	e-164	578	99 % (361/362)	99 % (361/362)
		Cytochrome b / *Drosophila yakuba*	P07704	e-133	475	77 % (281/362)	90 % (326/362)
PMM110 (993 bp)	nt	Calcium binding protein (EST) / *Penaeus monodon*	AI253941	e-129	466	94 % (583/615)	-
	aa	Calcium binding protein / *Penaeus sp.*	P02636	e-105	381	94 % (181/192)	97 % (188/192)
		Calcium binding protein / *Pontastacus leptodactylus*	P05946	3e-91	335	81 % (157/192)	89 % (174/192)

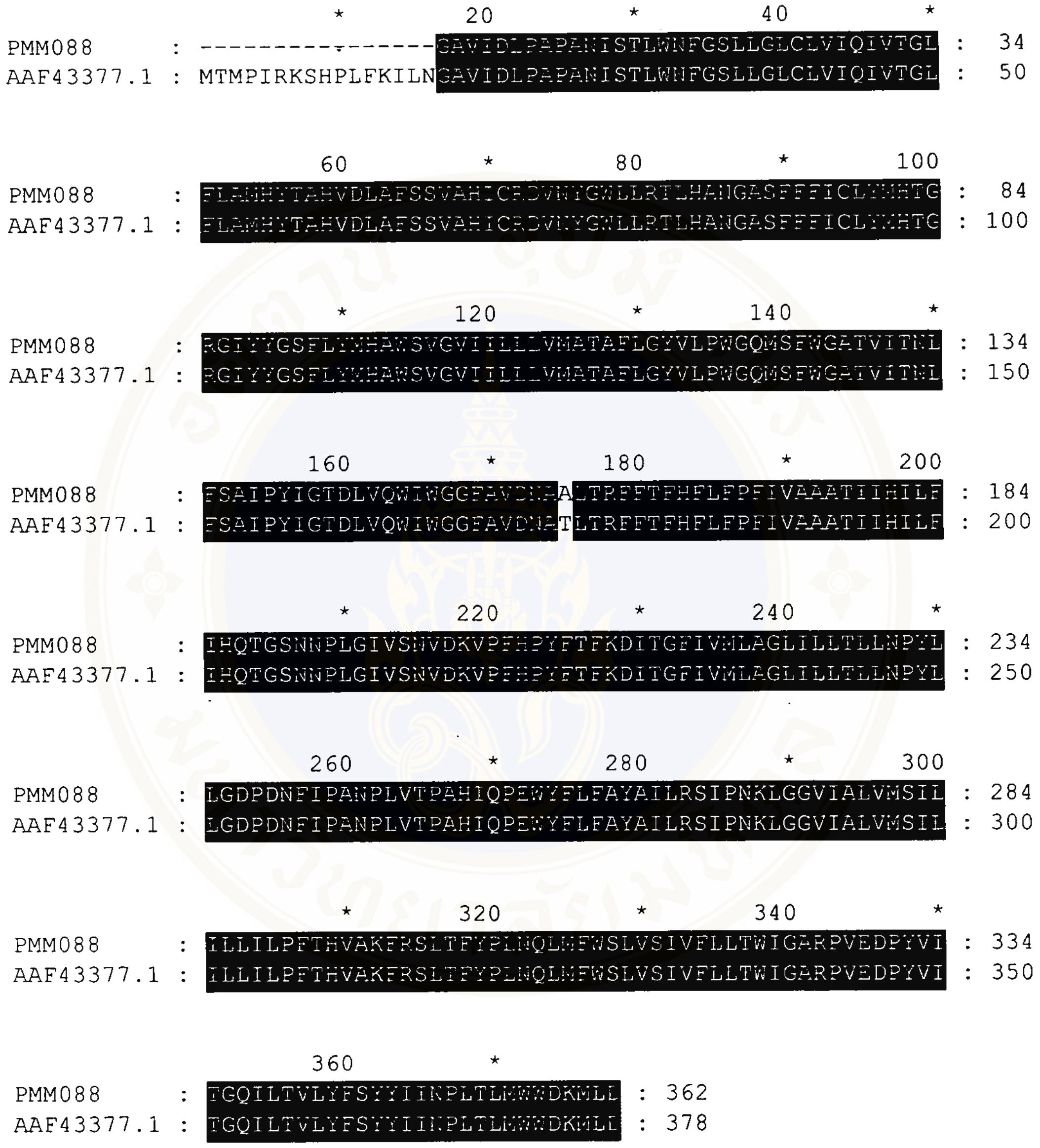
```
                      *        20         *        40         *        60
PMM088    : -----------------------------------------------TGGAGCTGTTATT :  13
AF217843  : ATGACAATACCAATTCGTAAATCGCACCCTTTATTTAAAATTTTAAATGGAGCTGTTATT :  60

                      *        80         *       100         *       120
PMM088    : GACCTCCCAGCACCAGCTAACATTTCAACTTTGTGAAACTTTGGTTCTCTCCTTGGTTTA :  73
AF217843  : GACCTCCCAGCACCAGCTAACATTTCAACTTTGTGAAACTTTGGTTCTCTACTTGGTTTA : 120

                      *       140         *       160         *       180
PMM088    : TGCTTAGTTATTCAAATCGTTACTGGACTCTTTTTAGCTATACACTATACTGCTCATGTT : 133
AF217843  : TGCTTAGTTATTCAAATCGTTACTGGACTCTTTTTAGCTATACACTATACTGCTCATGTT : 180

                      *       200         *       220         *       240
PMM088    : GACCTAGCATTCTCTAGTGTAGCACATATTTGTCGAGATGTAAATTATGGATGACTTCTC : 193
AF217843  : GACCTAGCATTCTCTAGTGTGGCACATATTTGTCGAGATGTAAATTATGGATGACTTCTC : 240

                      *       260         *       280         *       300
PMM088    : CGAACGTTACATGCTAATGGAGCGTCTTTCTTCTTCATTTGTCTGTATATACATACCGGA : 253
AF217843  : CGAACGTTACATGCTAATGGAGCGTCTTTCTTCTTCATTTGTCTGTATATACATACCGGA : 300

                      *       320         *       340         *       360
PMM088    : CGAGGTATTTATTATGGATCGTTTTTATATATACATGCTTGATCAGTAGGTGTCATTATT : 313
AF217843  : CGAGGTATTTATTATGGATCGTTTTTATATATACATGCTTGATCAGTAGGTGTCATTATT : 360

                      *       380         *       400         *       420
PMM088    : CTTCTACTAGTTATAGCGACTGCCTTCTTAGGGTATGTGTTACCTTGAGGACAAATATCA : 373
AF217843  : CTTCTACTAGTTATAGCGACTGCCTTCTTAGGATATGTGTTACCTTGAGGACAAATATCA : 420

                      *       440         *       460         *       480
PMM088    : TTCTGGGGAGCCACTGTAATTACGAATTTATTTTCAGCTATTCCATATATTGGTACAGAT : 433
AF217843  : TTCTGGGGAGCCACTGTAATTACGAATTTATTTTCAGCTATTCCATATATTGGTACAGAT : 480

                      *       500         *       520         *       540
PMM088    : CTTGTACAATGAATTTGAGGTGGATTTGCAGTTGATAATGCTGCATTAACTCGATTCTTC : 493
AF217843  : CTTGTACAATGAATTTGAGGTGGATTTGCAGTCGATAATGCTACATTAACTCGATTCTTC : 540

                      *       560         *       580         *       600
PMM088    : ACTTTTCATTTTTTATTCCCGTTTATCGTAGCAGCAGCAACTATTATTCATATTCTTTTT : 553
AF217843  : ACTTTTCATTTTTTATTCCCGTTTATCGTAGCAGCAGCAACTATTATTCATATTCTTTTT : 600
```

Figure 22 DNA sequence alignment of PMM088 full-length DNA sequence with that of cytochrome b of *P. monodon* (Accession: AF217843). Clustal X was used. An overall identity of 98 % was found, without any gap. Start and stop codon are underlined.



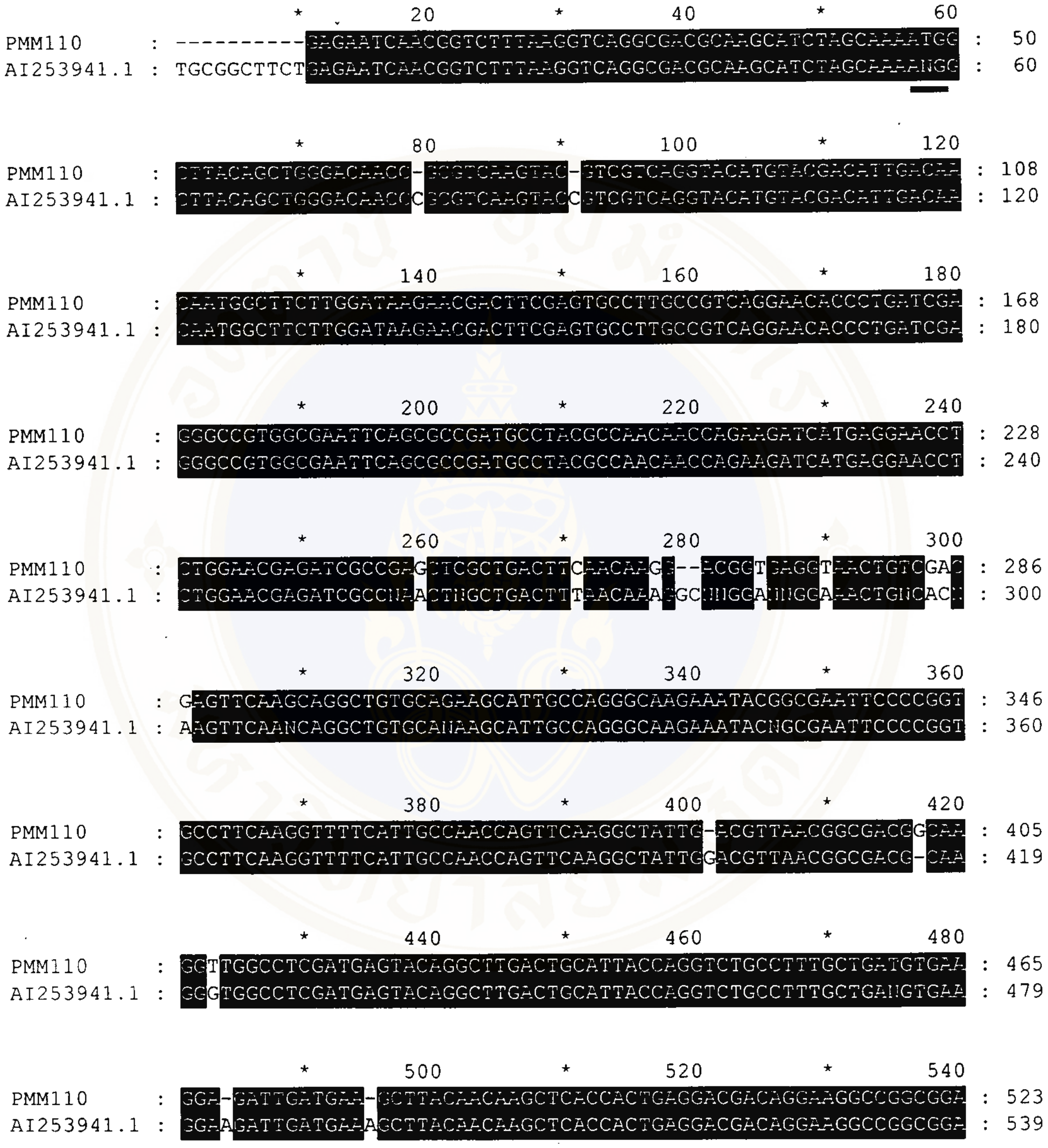
```
                          *        620         *        640         *        660
PMM088    : ATTCATCAAACAGGATCTAACAATCCACTAGGAATTGTTAGTAATGTAGATAAAGTACCT :  613
AF217843  : ATTCATCAAACAGGATCTAACAATCCGCTAGGAATTGTTAGTAATGTAGACAAAGTACCT :  660

                          *        680         *        700         *        720
PMM088    : TTTCATCCTTATTTTACATTTAAAGATATTACTGGATTTATCGTAATGCTAGCAGGACTT :  673
AF217843  : TTTCATCCTTATTTTACATTTAAAGATATTACTGGATTTATCGTAATACTAGCAGGACTT :  720

                          *        740         *        760         *        780
PMM088    : ATTCTATTAACATTATTAAACCCTTATTTATTGGGAGATCCAGACAACTTTATTCCCGCC :  733
AF217843  : ATTCTATTAACATTATTAAACCCTTATTTATTGGGAGATCCAGACAACTTTATTCCCGCC :  780

                          *        800         *        820         *        840
PMM088    : AATCCACTTGTTACCCCTGCCCATATTCAACCGGAGTGATATTTTTTATTTGCATATGCT :  793
AF217843  : AATCCACTTGTTACCCCTGCCCATATTCAACCGGAGTGATATTTTTTATTTGCATATGCT :  840

                          *        860         *        880         *        900
PMM088    : ATTTTACGATCAATTCCTAATAAATTAGGAGGTGTTATTGCCCTAGTAATATCTATCCTC :  853
AF217843  : ATTTTACGATCAATTCCTAATAAATTAGGAGGTGTTATTGCCCTAGTAATATCTATCCTC :  900

                          *        920         *        940         *        960
PMM088    : ATTTTATTAATCTTGCCTTTTACCCATGTAGCAAAATTTCGAAGATTAACGTTTTATCCT :  913
AF217843  : ATTTTATTAATCTTGCCTTTTACCCACGTAGCAAAATTTCGAAGATTAACGTTTTATCCT :  960

                          *        980         *       1000         *       1020
PMM088    : TTAAATCAACTTATATTTTGATCATTAGTAAGTATTGTCTTCTTATTAACTTGGATTGGG :  973
AF217843  : TTAAATCAACTTATATTTTGATCATTAGTAAGTATTGTCTTCTTATTAACTTGGATTGGG : 1020

                          *       1040         *       1060         *       1080
PMM088    : GCACGGCCAGTTGAAGATCCTTATGTAATTACTGGTCAGATCCTGACTGTTCTTTATTTC : 1033
AF217843  : GCACGGCCCGTTGAAGATCCTTATGTAATTACTGGTCAGATCCTGACTGTTCTTTATTTC : 1080

                          *       1100         *       1120         *
PMM088    : TCTTATTATATTATTAACCCACTTACACTTATATGATGGGATAAAATACTAGATTAA : 1090
AF217843  : TCTTATTATATTATTAACCCACTTACACTTATATGATGAGATAAGATACTAGATTAG : 1137
```

Figure 22 (continued)



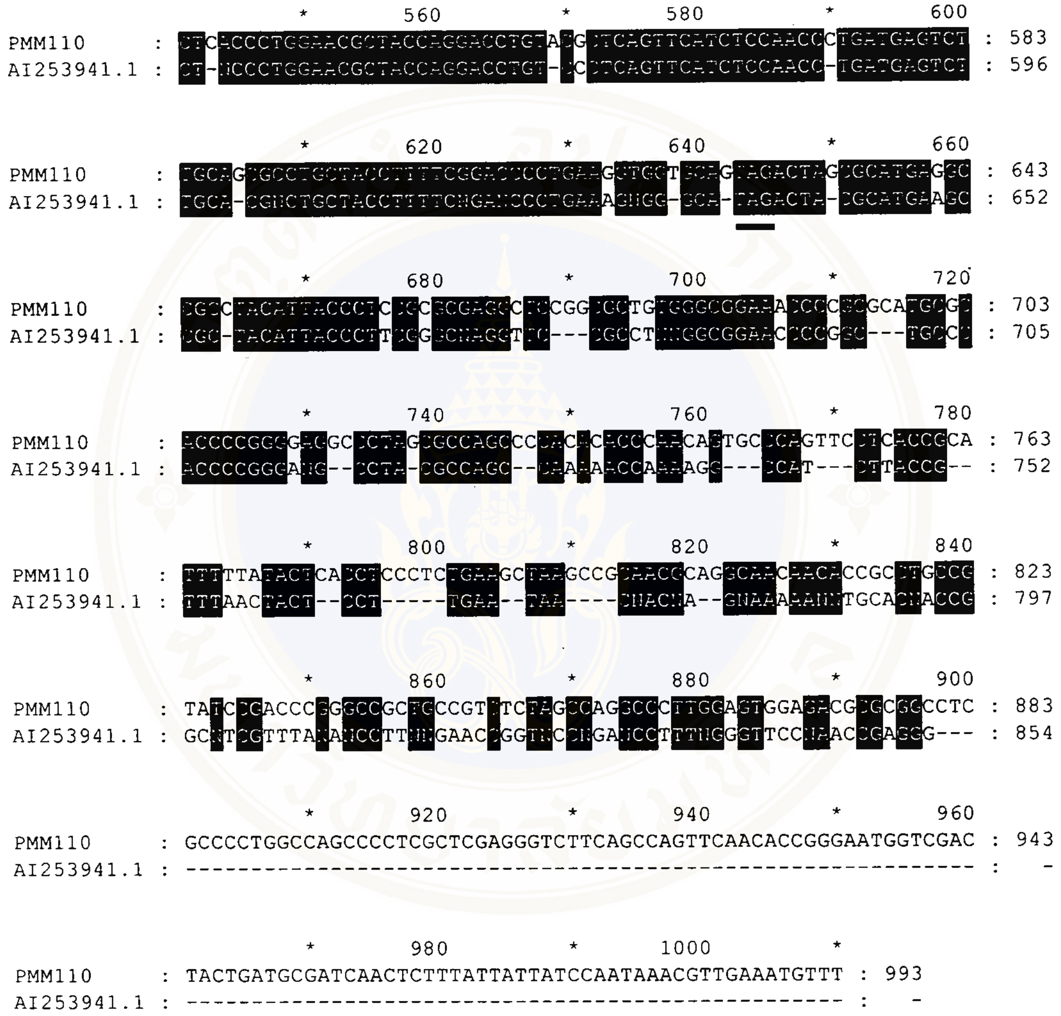
```
                      *        20         *        40         *
PMM088     : ----------------GAVIDLPAPANISTLWNFGSLLGLCLVIQIVTGL :  34
AAF43377.1 : MTMPIRKSHPLFKILNGAVIDLPAPANISTLWNFGSLLGLCLVIQIVTGL :  50

                     60         *        80         *       100
PMM088     : FLAMHYTAHVDLAFSSVAHICRDVNYGWLLRTLHANGASFFFICLYMHTG :  84
AAF43377.1 : FLAMHYTAHVDLAFSSVAHICRDVNYGWLLRTLHANGASFFFICLYMHTG : 100

                      *       120         *       140         *
PMM088     : RGIYYGSFLYMHAWSVGVIILLLVMATAFLGYVLPWGQMSFWGATVITNL : 134
AAF43377.1 : RGIYYGSFLYMHAWSVGVIILLLVMATAFLGYVLPWGQMSFWGATVITNL : 150

                    160         *       180         *       200
PMM088     : FSAIPYIGTDLVQWIWGGFAVDNAALTRFFTFHFLFPFIVAAATIIHILF : 184
AAF43377.1 : FSAIPYIGTDLVQWIWGGFAVDNATLTRFFTFHFLFPFIVAAATIIHILF : 200

                      *       220         *       240         *
PMM088     : IHQTGSNNPLGIVSNVDKVPFHPYFTFKDITGFIVMLAGLILLTLLNPYL : 234
AAF43377.1 : IHQTGSNNPLGIVSNVDKVPFHPYFTFKDITGFIVMLAGLILLTLLNPYL : 250

                    260         *       280         *       300
PMM088     : LGDPDNFIPANPLVTPAHIQPEWYFLFAYAILRSIPNKLGGVIALVMSIL : 284
AAF43377.1 : LGDPDNFIPANPLVTPAHIQPEWYFLFAYAILRSIPNKLGGVIALVMSIL : 300

                      *       320         *       340         *
PMM088     : ILLILPFTHVAKFRSLTFYPLNQLMFWSLVSIVFLLTWIGARPVEDPYVI : 334
AAF43377.1 : ILLILPFTHVAKFRSLTFYPLNQLMFWSLVSIVFLLTWIGARPVEDPYVI : 350

                    360         *
PMM088     : TGQILTVLYFSYYIINPLTLMWWDKMLL : 362
AAF43377.1 : TGQILTVLYFSYYIINPLTLMWWDKMLL : 378
```

Figure 23 Amino acid sequence alignment between PMM088 and cytochrome b of *P. monodon* (Accession: AAF43377.1). ClustalX was used. An overall identity of 99 % was found, without any gap.



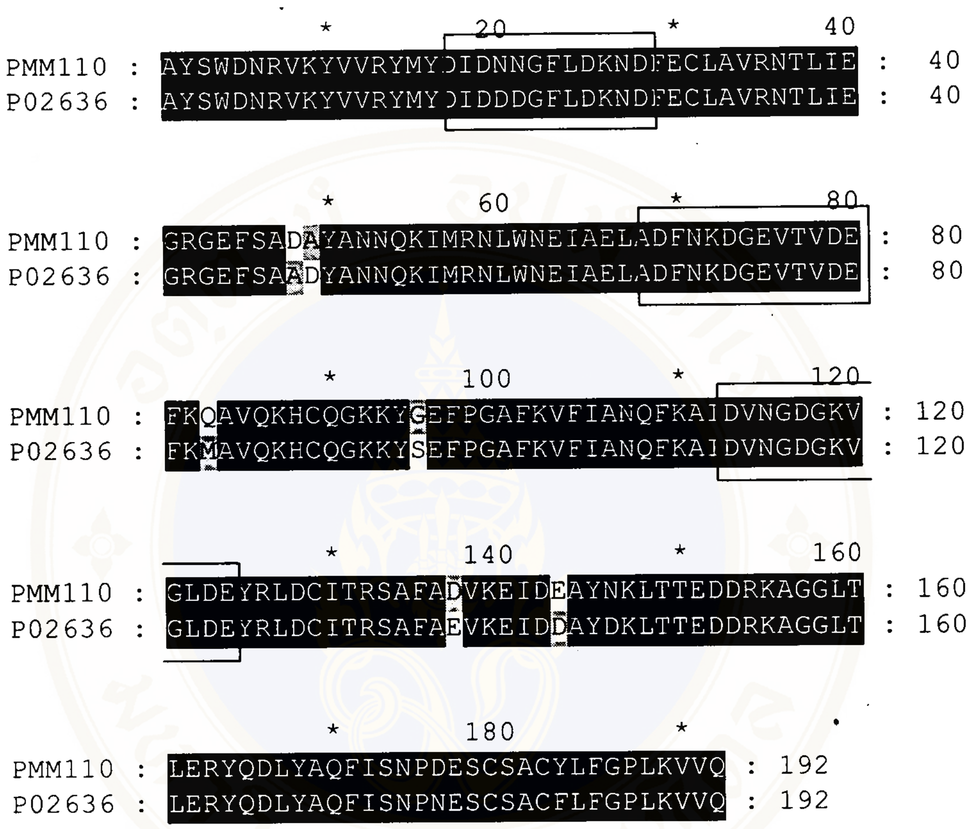
```
                      *        20         *        40         *        60
PMM110     : ----------GAGAATCAACGGTCTTTAAGGTCAGGCGACGCAAGCATCTAGCAAAATGG :  50
AI253941.1 : TGCGGCTTCTGAGAATCAACGGTCTTTAAGGTCAGGCGACGCAAGCATCTAGCAAAANGG :  60

                      *        80         *       100         *       120
PMM110     : CTTACAGCTGGGACAACC-GCGTCAAGTAC-GTCGTCAGGTACATGTACGACATTGACAA : 108
AI253941.1 : CTTACAGCTGGGACAACCCGCGTCAAGTACCGTCGTCAGGTACATGTACGACATTGACAA : 120

                      *       140         *       160         *       180
PMM110     : CAATGGCTTCTTGGATAAGAACGACTTCGAGTGCCTTGCCGTCAGGAACACCCTGATCGA : 168
AI253941.1 : CAATGGCTTCTTGGATAAGAACGACTTCGAGTGCCTTGCCGTCAGGAACACCCTGATCGA : 180

                      *       200         *       220         *       240
PMM110     : GGGCCGTGGCGAATTCAGCGCCGATGCCTACGCCAACAACCAGAAGATCATGAGGAACCT : 228
AI253941.1 : GGGCCGTGGCGAATTCAGCGCCGATGCCTACGCCAACAACCAGAAGATCATGAGGAACCT : 240

                      *       260         *       280         *       300
PMM110     : CTGGAACGAGATCGCCGAGCTCGCTGACTTCAACAAGG--ACGGTGAGGTAACTGTCGAC : 286
AI253941.1 : CTGGAACGAGATCGCCNAACTNGCTGACTTTAACAAAGGCNNGGANNGGAAACTGNCACN : 300

                      *       320         *       340         *       360
PMM110     : GAGTTCAAGCAGGCTGTGCAGAAGCATTGCCAGGGCAAGAAATACGGCGAATTCCCCGGT : 346
AI253941.1 : AAGTTCAANCAGGCTGTGCANAAGCATTGCCAGGGCAAGAAATACNGCGAATTCCCCGGT : 360

                      *       380         *       400         *       420
PMM110     : GCCTTCAAGGTTTTCATTGCCAACCAGTTCAAGGCTATTG-ACGTTAACGGCGACGGCAA : 405
AI253941.1 : GCCTTCAAGGTTTTCATTGCCAACCAGTTCAAGGCTATTGGACGTTAACGGCGACG-CAA : 419

                      *       440         *       460         *       480
PMM110     : GGTTGGCCTCGATGAGTACAGGCTTGACTGCATTACCAGGTCTGCCTTTGCTGATGTGAA : 465
AI253941.1 : GGGTGGCCTCGATGAGTACAGGCTTGACTGCATTACCAGGTCTGCCTTTGCTGANGTGAA : 479

                      *       500         *       520         *       540
PMM110     : GGA-GATTGATGAA-GCTTACAACAAGCTCACCACTGAGGACGACAGGAAGGCCGGCGGA : 523
AI253941.1 : GGAAGATTGATGAAAGCTTACAACAAGCTCACCACTGAGGACGACAGGAAGGCCGGCGGA : 539
```

Figure 24 DNA sequence alignment of PMM110 full-length DNA sequence with that of calcium binding protein EST of *P. monodon* (Accession: AI253941). ClustalX was used. An overall identity of 92 % in ORF was found. Start and stop codon are underlined.



```
                      *        560         *        580         *        600
PMM110     : CTCACCCTGGAACGCTACCAGGACCTGTACGCTCAGTTCATCTCCAACCCTGATGAGTCT : 583
AI253941.1 : CT-NCCCTGGAACGCTACCAGGACCTGT-CCCTCAGTTCATCTCCAACC-TGATGAGTCT : 596

                      *        620         *        640         *        660
PMM110     : TGCAGCGCCTGCTACCTTTTCGGACCCCTGAAGGTGGTGCAGTAGACTAGTGCATGAGGC : 643
AI253941.1 : TGCA-CGNCTGCTACCTTTTCNGANCCCTGAAAGNGG-GCA-TAGACTA-TGCATGAAGC : 652

                      *        680         *        700         *        720
PMM110     : CGCCTACATTACCCTCCGCCCGAGGCTCCGGCGCTGTGGGCGGAAACCCCGCGCATGCGC : 703
AI253941.1 : CGC-TACATTACCCTTCGGCCNAGGTTC---CGCCTNNGGCGGAACCCCGGC---TGCCC : 705

                      *        740         *        760         *        780
PMM110     : ACCCCGGGGAGGCCCTAGCGCCAGCCCCACACACCCAACAGTGCCCAGTTCCTCACCGCA : 763
AI253941.1 : ACCCCGGGANG--CCTA-CGCCAGC--CAAAAACCAAAAGG---CCAT---CTTACCG-- : 752

                      *        800         *        820         *        840
PMM110     : TTTTTATACTCACCTCCCTCTGAAGCTAAGCCGCAACGCAGGCAACAACACCGCCTGCCG : 823
AI253941.1 : TTTAACTACT--CCT-----TGAA--TAA----CNACNA--GNAAAAANNTGCACNACCG : 797

                      *        860         *        880         *        900
PMM110     : TATCCGACCCGGGCCGCTGCCGTCTCTAGCCAGGCCCTTGGAGTGGAGACGCGCGGCCTC : 883
AI253941.1 : GCNTCGTTTANANCCTTNNGAACCGGTNCCNGANCCTTTNGGGTTCCNAACCGAGGG--- : 854

                      *        920         *        940         *        960
PMM110     : GCCCCTGGCCAGCCCCTCGCTCGAGGGTCTTCAGCCAGTTCAACACCGGGAATGGTCGAC : 943
AI253941.1 : ------------------------------------------------------------ :   -

                      *        980         *       1000         *
PMM110     : TACTGATGCGATCAACTCTTTATTATTATCCAATAAACGTTGAAATGTTT : 993
AI253941.1 : -------------------------------------------------- :   -
```

Figure 24 (continued)



```
                            *        20         *        40
PMM110  : AYSWDNRVKYVVRYMYDIDNNGFLDKNDFECLAVRNTLIE  :  40
P02636  : AYSWDNRVKYVVRYMYDIDDDGFLDKNDFECLAVRNTLIE  :  40

                            *        60         *        80
PMM110  : GRGEFSADAYANNQKIMRNLWNEIAELADFNKDGEVTVDE  :  80
P02636  : GRGEFSAADYANNQKIMRNLWNEIAELADFNKDGEVTVDE  :  80

                            *       100         *       120
PMM110  : FKQAVQKHCQGKKYGEFPGAFKVFIANQFKAIDVNGDGKV  : 120
P02636  : FKMAVQKHCQGKKYSEFPGAFKVFIANQFKAIDVNGDGKV  : 120

                            *       140         *       160
PMM110  : GLDEYRLDCITRSAFADVKEIDEAYNKLTTEDDRKAGGLT  : 160
P02636  : GLDEYRLDCITRSAFAEVKEIDDAYDKLTTEDDRKAGGLT  : 160

                            *       180         *
PMM110  : LERYQDLYAQFISNPDESCSACYLFGPLKVVQ  : 192
P02636  : LERYQDLYAQFISNPNESCSACFLFGPLKVVQ  : 192
```

Figure 25 Amino acid sequence alignment between PMM110 and calcium binding protein of *Penaeus sp.* (Accession: P02636) Clustal X was used. An overall identity of 94 % was found, without any gap. Boxes show putative Ca^{2+} binding sequences.

Appendix 2 : A complete nucleotide sequence of a cDNA insert in PMM088.

Length = 1090 letters

```
TGGAGCTGTT ATTGACCTCC CAGCACCAGC TAACATTTCA ACTTTGTGAA ACTTTGGTTC
TCTCCTTGGT TTATGCTTAG TTATTCAAAT CGTTACTGGA CTCTTTTTAG CTATACACTA
TACTGCTCAT GTTGACCTAG CATTCTCTAG TGTAGCACAT ATTTGTCGAG ATGTAAATTA
TGGATGACTT CTCCGAACGT TACATGCTAA TGGAGCGTCT TTCTTCTTCA TTTGTCTGTA
TATACATACC GGACGAGGTA TTTATTATGG ATCGTTTTTA TATATACATG CTTGATCAGT
AGGTGTCATT ATTCTTCTAC TAGTTATAGC GACTGCCTTC TTAGGGTATG TGTTACCTTG
AGGACAAATA TCATTCTGGG GAGCCACTGT AATTACGAAT TTATTTTCAG CTATTCCATA
TATTGGTACA GATCTTGTAC AATGAATTTG AGGTGGATTT GCAGTTGATA ATGCTGCATT
AACTCGATTC TTCACTTTTC ATTTTTTATT CCCGTTTATC GTAGCAGCAG CAACTATTAT
TCATATTCTT TTTATTCATC AAACAGGATC TAACAATCCA CTAGGAATTG TTAGTAATGT
AGATAAAGTA CCTTTTCATC CTTATTTTAC ATTTAAAGAT ATTACTGGAT TTATCGTAAT
GCTAGCAGGA CTTATTCTAT TAACATTATT AAACCCTTAT TTATTGGGAG ATCCAGACAA
CTTTATTCCC GCCAATCCAC TTGTTACCCC TGCCCATATT CAACCGGAGT GATATTTTTT
ATTTGCATAT GCTATTTTAC GATCAATTCC TAATAAATTA GGAGGTGTTA TTGCCCTAGT
AATATCTATC CTCATTTTAT TAATCTTGCC TTTTACCCAT GTAGCAAAAT TTCGAAGATT
AACGTTTTAT CCTTTAAATC AACTTATATT TTGATCATTA GTAAGTATTG TCTTCTTATT
AACTTGGATT GGGGCACGGC CAGTTGAAGA TCCTTATGTA ATTACTGGTC AGATCCTGAC
TGTTCTTTAT TTCTCTTATT ATATTATTAA CCCACTTACA CTTATATGAT GGGATAAAAT
ACTAGATTAA
```

Appendix 3 : A complete nucleotide sequence of a cDNA insert in PMM110.

Length = 993 letters

```
GAGAATCAAC GGTCTTTAAG GTCAGGCGAC GCAAGCATCT AGCAAAATGG CTTACAGCTG
GGACAACCGC GTCAAGTACG TCGTCAGGTA CATGTACGAC ATTGACAACA ATGGCTTCTT
GGATAAGAAC GACTTCGAGT GCCTTGCCGT CAGGAACACC CTGATCGAGG GCCGTGGCGA
ATTCAGCGCC GATGCCTACG CCAACAACCA GAAGATCATG AGGAACCTCT GGAACGAGAT
CGCCGAGCTC GCTGACTTCA ACAAGGACGG TGAGGTAACT GTCGACGAGT TCAAGCAGGC
TGTGCAGAAG CATTGCCAGG GCAAGAAATA CGGCGAATTC CCCGGTGCCT TCAAGGTTTT
CATTGCCAAC CAGTTCAAGG CTATTGACGT TAACGGCGAC GGCAAGGTTG GCCTCGATGA
GTACAGGCTT GACTGCATTA CCAGGTCTGC CTTTGCTGAT GTGAAGGAGA TTGATGAAGC
TTACAACAAG CTCACCACTG AGGACGACAG GAAGGCCGGC GGACTCACCC TGGAACGCTA
CCAGGACCTG TACGCTCAGT TCATCTCCAA CCCTGATGAG TCTTGCAGCG CCTGCTACCT
TTTCGGACCC CTGAAGGTGG TGCAGTAGAC TAGCGCATGA GGCCGCCTAC ATTACCCTCC
GCGCGAGGCT CCGGCGCTGT GGGCGGAAAC CCCGCGCATG CGCACCCCGG GGAGGCCCTA
GCGCCAGCCC CACACACCCA ACAGTGCCCA GTTCCTCACC GCATTTTTAT ACTCACCTCC
CTCTGAAGCT AAGCCGCAAC GCAGGCAACA ACACCGCCTG CCGTATCCGA CCCGGGCCGC
TGCCGTCTCT AGCCAGGCCC TTGGAGTGGA GACGCGCGGC CTCGCCCCTG GCCAGCCCCT
CGCTCGAGGG TCTTCAGCCA GTTCAACACC GGGAATGGTC GACTACTGAT GCGATCAACT
CTTTATTATT ATCCAATAAA CGTTGAAATG TTT
```

BIOGRAPHY

NAME	Miss Jantana Wongsantichon
DATE OF BIRTH	27 March 1974
PLACE OF BIRTH	Phichitr, Thailand
INSTITUTES ATTENDED	Chulalongkorn University, 1991-1995 Bachelor of Science (Medical Technology) Mahidol University, 1997-2000 Master of Science (Molecular Genetics-Genetic Engineering)
RESEARCH SCHOLARSHIP	Thailand Research Fund's Senior Research Fellowship to Prof. Sakol Panyim