

MOLECULAR CLONING AND EXPRESSION OF
Macrobrachium rosenbergii L-TYPE LECTIN IN
Escherichia coli

CHAN CHEW CHIN

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2019

MOLECULAR CLONING AND EXPRESSION OF
Macrobrachium rosenbergii* L-TYPE LECTIN IN *Escherichia
coli

CHAN CHEW CHIN

DISSERTATION SUBMITTED IN PARTIAL
FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF BIOTECHNOLOGY

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2019

UNIVERSITY OF MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **CHAN CHEW CHIN**

Matric No: **SGF160017**

Name of Degree: **MASTER OF BIOTECHNOLOGY**

Title of Project Paper/Research Report/Dissertation/Thesis (“this Work”):

MOLECULAR CLONING AND EXPRESSION OF *Macrobrachium rosenbergii* L-TYPE LECTIN IN *Escherichia coli*

Field of Study:

MOLECULAR BIOLOGY

I do solemnly and sincerely declare that:

- (1) I am the sole author/writer of this Work;
- (2) This Work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya (“UM”), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- (6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate’s Signature

Date:

Subscribed and solemnly declared before,

Witness’s Signature

Date:

Name:

Designation:

MOLECULAR CLONING AND EXPRESSION OF *Macrobrachium rosenbergii* L-TYPE LECTIN IN *Escherichia coli*

ABSTRACT

Macrobrachium rosenbergii, also known as the giant freshwater prawn represents a valuable commodity in aquaculture farming all over the world. The lack of adaptive immunity system in *M. rosenbergii* renders difficulty in development of effective treatment for both pathogenic viral and bacterial infections. Existing disease control relies on effective biosecurity measures such as pathogen-free larvae stock and meticulous cultivation environment management. The effort to elucidate innate defence responses towards pathogens' invasion provides valuable insights for better disease prevention and control. Lectins are recognised as a crucial component in pattern recognition receptor (PRR), facilitating phagocytosis and pathogen clearance from infected hosts. Recently, attention has been placed on L-type lectins regarding on their potential involvement in innate immune response. This project has cloned and expressed the L-type lectin domain protein, *MrLTL*, characterised from *M. rosenbergii*. This protein comprises a signal peptide, L-type lectin domain, a transmembrane region, with a total sequence length of 323 amino acids. Two major carbohydrate-binding motifs, YSN and GDL, were identified. The predictive structure of the protein adopts a beta-sandwich fold solved via homology modelling. Bioinformatics simulation demonstrates preferential binding of mannose-type ligands towards *MrLTL* with extensive hydrogen bond formation. The production of soluble *MrLTL* will provide insights into structure and function of this protein, which in turn serve as an effort to expand the current knowledge on L-type lectin in crustacean species.

Keywords: *Macrobrachium rosenbergii*, innate defence response, pattern recognition receptor, L-type lectin, recombinant expression

PENGLONAN MOLEKUL DAN UNGKAPAN LECTIN JENIS L *Macrobrachium rosenbergii* DALAM *Escherichia coli*

ABSTRAK

Macrobrachium rosenbergii, dikenali sebagai udang galah, merupakan salah satu komoditi yang berharga dalam bidang akuakultur di seluruh dunia. Kekurangan sistem imun penyesuaian dalam *M. rosenbergii* menyebabkan kesulitan dalam penyediaan rawatan yang berkesan untuk jangkitan virus dan bakteria patogenik. Setakat ini, pencegahan penyakit bergantung pada langkah-langkah biosekuriti yang berkesan seperti penggunaan benih udang yang bebas daripada patogen dan pengurusan yang teliti antara pusat-pusat perternakan udang. Usaha-usaha untuk menyelidik sistem imun semula jadi terhadap jangkitan patogen menyumbang kepada perkembangan pengawalan dan pencegahan penyakit. *Lectins* adalah salah satu komponen *pattern recognition receptor (PPR)* yang amat penting, memudahkan fagositosis dan penyingkiran patogen daripada haiwan yang dijangkiti. Baru-baru ini, *lectin* jenis L telah mendapat perhatian umum disebabkan potensi penglibatannya dalam reaksi sistem imun semula jadi. Dalam projek ini, *lectin* jenis L yang berasal daripada *M. rosenbergii*, dinamakan *MrLTL*, telah diklon dan diungkap. Protein ini terdiri daripada peptida isyarat, *domain lectin* jenis L, bahagian transmembran, dengan jumlah jujukannya sepanjang 323 asid amino. Dua *motif* pengikatan karbohidrat, YSN dan GDL, telah ditemui. Struktur ramalan protein ini menerima konformasi *beta-sandwich* yang diramal melalui pemodelan homologi. Simulasi bioinformatik menunjukkan pengikatan pilihan oleh *mannose-type ligands* terhadap *MrLTL* dengan penghasilan ikatan hidrogen yang ekstensif. Kejayaan penghasilan protein larut *MrLTL* akan menyumbang kepada pemahaman struktur dan fungsi protein tersebut, membawa kepada pengembangan pengetahuan terhadap *lectin* jenis L dalam spesies krustasia.

Kata kunci: *Macrobrachium rosenbergii*, imun semula jadi, *pattern recognition receptor*, *lectin* jenis L, Ungkapan rekombinan

University of Malaya

ACKNOWLEDGEMENTS

I would like to express my greatest gratitude to my supervisor Assoc. Prof. Dr. Subha Bhassu for believing in my potential to perform this project. Her critical thoughts, kind advice, and unconditional support constantly motivates me through any challenge encountered during the pursue of this degree. I admire her genuine dedication, not only in guiding her students to attain academic excellence, but also in moulding us to be the better version of ourselves in life.

Also, my heartfelt thanks to all the lab members from AGAGEL for their assistance and generosity in sharing of knowledge. In particular, my warmest appreciation goes to Nurul Huda, for being helpful throughout my study. Her delightful personality cheered my spirit up whenever I was depressed, and those extended hours she spent in the lab to help solving my problem are priceless.

Lastly, my family is what I hold dearest to my heart. Their simply being present puts a sense of security in my mind, reminds me that I am always loved and they always have my back, and is the only reason that I can chase any dream of mine relentlessly.

TABLE OF CONTENTS

Abstract	iii
Abstrak	iv
Acknowledgements	vi
Table of Contents	vii
List of Figures	x
List of Tables	xi
List of Symbols and Abbreviations	xii
List of Appendices	xvi
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: LITERATURE REVIEW	4
2.1 <i>Macrobrachium rosenbergii</i>	4
2.2 Infectious Diseases and control of <i>Macrobrachium rosenbergii</i>	5
2.3 Immune system in prawn	8
2.4 Pattern recognition receptor and Lectin	11
2.5 L-type Lectin	13
2.6 Recombinant protein technologies	15
CHAPTER 3: METHODOLOGY	19
3.1 Bioinformatics characterisation of <i>MrLTL</i> sequence.....	19
3.2 Sample collection.....	19
3.3 Total RNA extraction.....	20
3.4 First strand cDNA synthesis.....	21
3.5 PCR of <i>MrLTL</i> gene.....	21

3.6	Purification of PCR products from agarose gels.....	22
3.7	DNA cloning	23
3.7.1	Restriction endonuclease digestion	23
3.7.2	Ligation process	24
3.8	Transformation of expression host	24
3.9	Induction of recombinant protein expression.....	25
3.10	Optimisation of recombinant protein expression.....	25
3.10.1	Temperature	25
3.10.2	IPTG concentration.....	26
3.10.3	Additives	26
3.11	Cell lysis	26
3.12	SDS-PAGE Analysis	27
3.13	Western Blotting.....	27
CHAPTER 4: RESULTS.....		29
4.1	Bioinformatics characterisation of <i>MrLTL</i> gene sequence.....	29
4.1.1	<i>MrLTL</i> coding sequence and protein features.....	29
4.1.2	<i>MrLTL</i> and <i>MrLTLD</i> ProtParam parameters.....	30
4.1.3	<i>MrLTL</i> structural features.....	31
4.1.4	<i>In silico</i> simulation of binding of ligands on <i>MrLTLD</i>	34
4.2	Amplification of the nucleotide sequence of <i>MrLTL</i>	35
4.3	Cloning of <i>MrLTL</i> and <i>MrLTLD</i> sequences into pET-30(a) expression vector...	36
4.4	Expression of recombinant proteins <i>MrLTL</i> and <i>MrLTLD</i>	38
CHAPTER 5: DISCUSSION.....		41
CHAPTER 6: CONCLUSION		46

REFERENCES	47
APPENDICES.....	55
Appendix A: Medium and Buffers Composition	55
Appendix B: Formula for preparation of stacking and resolution gel for SDS-PAGE. .	59
Appendix C: pET-30(a) vector mapping.....	60
Appendix D: Ramachandran plots of the predicted <i>MrLTL</i> protein structure.....	61

University of Malaya

LIST OF FIGURES

Figure No.	Titles	Page
Figure 4.1	: <i>MrLTL</i> nucleotide sequence and translated putative amino acid sequence.	29-30
Figure 4.2	: <i>MrLTL</i> predicted protein structure.	32
Figure 4.3	: Zoom in view of <i>MrLTL</i> protein's carbohydrate-binding motifs.	33
Figure 4.4	: Structural comparison of L-type lectins of various origin.	33
Figure 4.5	: Illustration of ligand's interaction with carbohydrate-binding motif of <i>MrLTL</i> .	35
Figure 4.6	: Gel image of amplified <i>MrLTL</i> sequence with expected amplicon size of 1247 bp.	36
Figure 4.7	: PCR products of <i>MrLTL</i> and <i>MrLTL</i> ΔD containing restriction endonuclease recognition sites (<i>Pci</i> I & <i>Xho</i> I).	36
Figure 4.8	: Colony PCR of <i>MrLTL</i> and <i>MrLTL</i> ΔD <i>E. coli</i> BL21(DE3) transformants.	37
Figure 4.9	: The pET-30(a) expression vector construct mapping.	37
Figure 4.10	: Schematic illustration of recombinant <i>MrLTL</i> and <i>MrLTL</i> ΔD peptide sequences expressed via pET-30(a) vector.	38
Figure 4.11	: SDS-PAGE image of <i>MrLTL</i> recombinant protein induced with different concentrations of IPTG.	39
Figure 4.12	: SDS-PAGE image of <i>MrLTL</i> ΔD recombinant protein induced with different concentrations of IPTG.	39
Figure 4.13	: SDS-PAGE images of <i>MrLTL</i> (left) & <i>MrLTL</i> ΔD (right) recombinant proteins induced in various induction temperatures.	40
Figure 4.14	: SDS-PAGE image of <i>MrLTL</i> & <i>MrLTL</i> ΔD recombinant protein induced expression with additives (A) 0.5 mM MgCl ₂ + 0.5 mM CaCl ₂ and (B) 3% ethanol.	40

LIST OF TABLES

Table No.	Titles	Page
Table 3.1	: Oligonucleotides and their respective sequences.	22
Table 4.1	: ProtParam physical and chemical values for both <i>MrLTL</i> and <i>MrLTLD</i>	31
Table 4.2	: Autodock binding simulation results of carbohydrate ligands with <i>MrLTLD</i> .	34

University of Malaya

LIST OF SYMBOLS AND ABBREVIATIONS

SYMBOLS

%	:	Percentage
°C	:	Degree Celsius

ABBREVIATIONS

µg	:	Micrograms
µl	:	Microlitre
AHPND	:	Acute hepatopancreatic necrosis disease
BC	:	Blue claw
bp	:	Base pair
Ca ²⁺	:	Calcium cation
CaCl ₂	:	Calcium Chloride
cDNA	:	Complementary DNA
CRD	:	Carbohydrate recognition domain
DNA	:	Deoxyribonucleic acid
dNTP	:	Deoxy nucleoside triphosphate
ERGIC-53	:	ER-Golgi intermediate compartment protein-53kDa
ERGL	:	ER-Golgi intermediate compartment protein-like
ER-resident	:	Endoplasmic reticulum-resident
FAO	:	Food and Agriculture Organisation of United Nations
GAA	:	Global Aquaculture Alliance
GRAVY	:	Grand average of hydropathicity value
GST	:	Glutathione-S-transferase
H-bond	:	Hydrogen band

hr	:	Hour
HRP	:	Horseradish peroxidase
IMD	:	Immune deficiency pathway
IMNV	:	Infectious myonecrosis virus
IPTG	:	Isopropyl- β -D-1-thiogalactopyranoside
LB	:	Luria-Bertani
LGBP	:	Lipopolysaccharide-and-glucan-binding-protein
LPS	:	Lipopolysaccharide
LTL	:	L-type lectin
LTLD	:	L-type lectin domain
mA	:	Milliampere
Man	:	Mannose
ManNac	:	N-Acetyl-D-mannosamine
MBP	:	Maltose-binding protein
MCS	:	Multiple cloning site
mg	:	Milligrams
MgCl ₂	:	Magnesium Chloride
MHPV	:	<i>Macrobrachium</i> hepatopancreatic Parvo-like virus
min	:	Minute
ml	:	Millilitre
mM	:	Millimolar
MMV	:	<i>Macrobrachium</i> muscle virus
Mn ²⁺	:	Manganese cation
M _r	:	Molecular mass
<i>Mr</i> LTL	:	<i>M. rosenbergii</i> L-type lectin
<i>Mr</i> LTLD	:	<i>M. rosenbergii</i> L-type lectin domain

<i>MrNV</i>	:	<i>M. rosenbergii</i> nodavirus
NCBI	:	National Centre for Biotechnology Information
NusA	:	N utilisation substance A
OC	:	Orange claw
OD	:	Optical density
ORF	:	Open reading frame
PAMP	:	Pathogen associated molecular pattern
PCR	:	Polymerase chain reaction
PDB	:	Protein Data Bank
PGN	:	Peptidoglycan
pI	:	Isoelectric point
PRR	:	Pattern recognition receptor
PVDF	:	Polyvinylidene difluoride
RE	:	Restriction endonuclease
RNA	:	Ribonucleic acid
rpm	:	Rotations per minute
s	:	Second
SDS-PAGE	:	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SOC	:	Super optimal broth with catabolite repression
SUMO	:	Small ubiquitin-related modifier
TLR	:	Toll-like receptor
Trx	:	Thioredoxin
TSV	:	Taura syndrome virus
V	:	Voltage
VIP36	:	Vesicular integral membrane protein-36kDa
VIPL	:	Vesicular integral membrane protein-like

w/v	:	Weight per volume
WSSV	:	White spot syndrome virus
x g	:	Gravitational force
XSV	:	Extra small virus
YHV	:	Yellow head virus
βGBP	:	β-glucan-binding-protein

University of Malaya

LIST OF APPENDICES

Appendix No.	Titles	Page
Appendix A	: Medium and buffer compositions	55
Appendix B	: Formula for preparation of stacking and resolution gel for SDS-PAGE	59
Appendix C	: pET-30(a) vector mapping.	60
Appendix D	: Ramachandran plots of the predicted <i>MrLTL</i> protein structure.	61

University of Malaya

CHAPTER 1: INTRODUCTION

Prawns and shrimps are widely adopted and enjoyed as food delicacies around the world. The gradual expansion of shrimp cultivation and production was anticipated in response to increased demand. However, periodic disease outbreaks were observed accompanied by the development of shrimp cultivation. A rough estimation of US\$ 1 billion annual loss was determined (Flegel et al., 2008). It was mentioned in Flegel's work (2008) that viral infection inflicted 60% while bacterial infection contributed to 20% of losses in shrimp cultivation from a survey conducted by Global Aquaculture Alliance (GAA). Commonly encountered pathogenic viruses in penaeid shrimps are white spot syndrome virus (WSSV), Taura syndrome virus (TSV), infectious myonecrosis virus (IMNV), yellow head virus (YHV), and *Vibrio* species for bacterial infections. Disease control of shrimp cultivation relies on biosecurity measures, which emphasises on tight control of culture environments and prevention of potential pathogens invasion. The use of vaccines and antibiotics is not recommended for disease prevention due to the absence of long-lasting adaptive immunity in invertebrates and development of bacteria resistance to antibiotics.

The discovery of antimicrobial peptides and related protein molecules towards pathogens invasion spikes intense interests in the study of innate immunity of shrimps and related invertebrates. By knowing the mechanism and relationship between these protein molecules and pathogens, identification and production of specific defence molecules to achieve better disease prevention and control are envisaged. Lectin is suggested as a potential component carrying crucial roles in innate defence system. Up-regulation of lectin genes during viral challenge proposes possible intervention of these molecules in defence mechanism. Various c-type lectin variants have been identified and characterised in penaeid shrimps (Li & Xiang, 2013). They are mainly found in

haemocytes and hepatopancreas as pattern recognition receptor (PRR) exhibiting differential affinity towards pathogens carbohydrate surface components. The diversity of pathogen species and dynamic composition of pathogen surface antigen patterns demand a great degree of lectins diversity to execute successive immune response. Unlike adaptive immune system that generates diversified immunoglobins via genetic recombination, lectin system suggests the presence of multigene families, alternative splicing and isoforms, and formation of chimeric structure with different pattern recognition domains contribute to the versatility of this system (Vasta et al., 2007).

Macrobrachium rosenbergii, commonly known as the giant freshwater prawn, is an important aquaculture farming product in Malaysia. The study of innate immune response and interaction towards pathogenic agents provide insights to formulate effective farming strategies and prevent unnecessary economical losses. Till now, very few published works on L-type lectins in shrimp are available. There is no publicly available report of characterised *M. rosenbergii* L-type lectin when this project was initiated. This project involved the cloning of L-type lectin gene (*MrLTL*) from *M. rosenbergii* into an *Escherichia coli* expression system. The successive expression of this gene allows structural and functional characterisations of this L-type lectin. The subsequent purification of this lectin provides insights into the function of this L-type lectin in *M. rosenbergii* innate defence system.

The objectives of this study were:

- To characterise *M. rosenbergii* L-type lectin (*MrLTL*) gene with bioinformatics.
- To clone and express *MrLTL* and a deletion mutant lacking the signal and transmembrane peptides (*MrLTLD*) in *E. coli*.

- To optimise the expression of the recombinant *MrLTL* and *MrLTLD* in *E. coli* through manipulation of different incubation settings, which includes IPTG concentrations (0.1 mM, 0.2 mM, 0.5 mM, and 1.0 mM), incubation temperatures and duration (20 °C (overnight), 25 °C (overnight), 37 °C (4 hours)), and medium additives (0.5 mM MgCl₂ + 0.5 mM CaCl₂ and 3 % ethanol).

University of Malaya

CHAPTER 2: LITERATURE REVIEW

2.1 *Macrobrachium rosenbergii*

Macrobrachium rosenbergii belongs to the order of Decapoda and family of Palaemonidae, and represents the largest of its genus (De Grave et al., 2013). They are commonly known as the giant freshwater prawn, or “udang galah” locally in Malaysia. This species is indigenous in Southeast Asia, northern Oceania and western Pacific islands (New, 2002). Since the introduction of rearing technology of this species in Hawaii in 1965, it has been quickly adopted by many countries, and the production expands steadily per annum (New, 2002). A global production exceeded 200,000 tonnes per year was reported by Food and Agriculture Organisation of United Nations (FAO) in year 2002.

M. rosenbergii habitats a wide range of freshwater areas and tolerates exceptionally well in turbid conditions (New, 2002). They are distinctive from other prawn species with the presence of an elongated second pair of chelipeds. Male prawns can grow up to 320 mm and female prawn 250 mm. Interestingly, male giant freshwater prawns exhibit heterogeneous individual growth (FAO, 2017). They are categorised into three different morphologies, namely blue claw (BC), orange claw (OC) and small males. BC males rest on the hierarchical top, dominating both OC and small males. In addition, OC males will only metamorphose into BC males when their body sizes overtake nearby largest BC males. Male and female prawns shared highly similar external appearances, and sexing of individual can be determined through the small protrusion present only on the ventral side of the first somite on male prawn (New, 2002). While adult prawns thrive in freshwater areas, the survival and development of larvae require brackish water. The mating event happens throughout the year, and increased frequency coincides with summer’s onset and rainy season in temperate and tropical regions, respectively (New, 2002). Berried female prawns carry their eggs in the brood chamber underneath abdomen

segment, and hatching eggs are often dispensed in estuaries, where mild salinity water is available. The development of larvae will pass through 11 stages before metamorphoses into postlarvae, which resembles juvenile prawns and begin to migrate upstream into freshwater areas. Giant freshwater prawns are omnivores and feed on various food sources, such as small aquatic animals, fish fleshes and offal, grains and seeds. Cannibalism is also observed when food is scarce and during moulting interval (FAO, 2004).

Macrobrachium rosenbergii farming industry receives most establishment in Asian countries. In the report published by New (2005), China is the leading producer in the world, with 128 338 tonnes production recorded in the year 2001. Other major producers include Vietnam, India, Thailand, Bangladesh and Taiwan. Marginal production of *M. rosenbergii* was also recorded from countries such as Guatemala, Mexico, Senegal, and USA. Malaysia was ranked number 8th in New's report with a total production of 752 tonnes in the year 2001. In Malaysia, the highest output of giant freshwater prawn was achieved in the year 2000 with 1338 tonnes in total (Banu & Christianus, 2016). Since then, the figure plummeted and fluctuated in last decade with an average of 500 tonnes per year (Banu & Christianus, 2016). Department of Fisheries Malaysia (2017) reported a production of 309 tonnes giant freshwater prawn valued approximately 37.7 million Malaysia Ringgit in 2016. The staggering performance was due to the shortage of quality broodstock supply, low survivability of imported postlarvae, limitation in culture technology, and cost burden on imported larval feeds (Banu & Christianus, 2016).

2.2 Infectious Diseases and control of *Macrobrachium rosenbergii*

Infectious diseases of *M. rosenbergii* are of various origins, including viruses, bacteria, fungi and protozoa. Viral infections are more detrimental than bacterial infections in farmed shrimp industry. Giant freshwater prawns are presumed to be more resistant and

suffer milder impact from infection compared to penaeid shrimps. This might be due to genetic makeup of *M. rosenbergii* with higher disease tolerance, or the fact that giant freshwater prawn aquaculture is operated in a less stocking density condition. In fact, the intensification trend of giant freshwater prawn culture and enhanced world trade in recent years, is now accompanied by a more frequent and severe production loss due to disease infections.

M. rosenbergii nodavirus (*MrNV*) and the associated extra small virus (*XSV*) are the causative agents for white tail disease in *M. rosenbergii* (Bonami & Widada, 2011). They are both icosahedral in shape, non-enveloped, single-stranded RNA viruses, with a size of approximately 27 nm and 15 nm diameter, respectively. Postlarvae and juveniles are the most susceptible, showing signs of opaque muscle and white tail formation, which causes 100% mortality in infected prawns. While viruses are believed to be transmitted through vertical mode where eggs are contaminated by infected berried female prawns, the possibility of horizontal transmission via contaminated food and environmental sources should not be neglected (Bonami & Widada, 2011). Also, *Macrobrachium* muscle virus (*MMV*) and *Macrobrachium* hepatopancreatic Parvo-like virus (*MHPV*) were also reported to infect giant freshwater prawns. The former causes 50%-75% mortality on postlarvae during an introduction to the grow-out pond, while the latter has limited information on its impact (Saurabh & Sahoo, 2007). On the contrary, the two most notorious penaeid shrimp's viruses, namely infectious hypodermal and haematopoietic necrosis virus (*IHHNV*) and white spot syndrome virus (*WSSV*), are rather well tolerated by the giant freshwater prawn. Both viruses were found to infect the prawn without showing clinical symptoms and significant mortality (Saurabh & Sahoo, 2007). Nonetheless, evidence on *IHHNV* infection inflicted 80%-100% mortality on postlarvae and juvenile was first reported in Taiwan (Hsieh et al., 2006). Hazreen (2012) also

indicates detection of IHHNV positive *M. rosenbergii* in wild population of berried female prawns in Malaysia.

On the other hand, the bacterial infection of *M. rosenbergii* caused by *Vibrio* spp. receives most of the attention. *Vibrio* spp. contaminated seafood leads to food-borne sickness upon consumption. They are prevalently found in the water system, and larval survival is vulnerable when their number increased (Saurabh & Sahoo, 2007). In addition, the outbreak of acute hepatopancreatic necrosis disease (AHPND) which struck heavily on Vietnam penaeid shrimp industry was identified as *Vibrio parahaemolyticus* AHPND strain infection (Hien et al., 2016). There is no report on AHPND infection in *M. rosenbergii* to date. The prawn commonly develops melanised spots on their shells as a result of activated melanisation defence mechanism upon bacterial infection. Moulting process then replaces these spotted shells on healthy prawns. The environmental microbiome is responsible for most pathogenic problems when the animal is under stress. Mixed pathogen infection including bacteria, fungi, protozoa is often observed in muscle necrosis and fouling of the prawn (Saurabh & Sahoo, 2007). Protozoa that are readily found on adult prawn include *Zoothamnium* spp., *Epistylis* spp., *Vorticella* spp., and *Acineta* spp.; while common bacteria found are *Vibrio* spp., *Aeromonas* spp., *Edwardsiella* spp., and *Pseudomonas* spp. (Hoa et al., 2014).

Extensive effort was invested to mitigate the adverse impact of infectious diseases in prawn aquaculture. While studies on the pathogenicity and biology of the pathogens are utmost critical, many ventures into the investigation of environmental factors and its impact on prawn susceptibility towards diseases. Vaccines offer trivial benefit to viral infection because adaptive immune system is absent in shrimps, and the use of antibiotics potentially leads to development of resistance bacteria. Since effective therapeutic substance is not available for most shrimp pathogenic diseases, current

endeavours ground on the basic of intensive farming management and tight biosecurity parameters. Preventive measures are by far the most pragmatic and practical actions in disease control. Early detection of pathogens and routine surveillance of prawns' health status are crucial. A vigilant monitoring of every possible aspect ranges from water source, brood and culture stock origin, farm conditions, and feed, prepares owners from any unsolicited infection mishap, limits the likelihood of pathogen's entry into the farm, and facilitates the tracing of pathogen's origin. Feed represents a versatile component for manipulation. Conventional compositions of shrimp feed comprise animal carcasses, fish offal and crustacean, and plant-based ingredient. Innovative practices by addition of immunostimulants into shrimp dietary confer beneficial effects on disease mitigation and their control (Mastan, 2015). Immunostimulants elicit a general immune response by interacting with host's immune components, mount protection against potential pathogenic infection. Various substances have been characterised, and their immunostimulatory properties in shrimps have been demonstrated. They are components derived from microbial source or plant-based such as muramyl dipeptide, chitin, oligosaccharide, polysaccharide, vitamins, and yeast derivatives (Mastan, 2015). Among all, β -glucan, a polysaccharide found in the cell wall of bacteria, fungi and plants found to be most effective in stimulating shrimp's immune response (Mastan, 2015).

2.3 Immune system in prawn

Without the presence of adaptive immune system in crustacean and invertebrate, prawns manoeuvre a wide array of innate immune mechanisms to protect itself from pathogenic invasion. Exoskeleton or shell apparently serves as the first physical barrier in deterring the invasion of pathogenic agent. Interestingly, a study by Tiruvayipati and Bhasu (2016) outlines the interplay between environmental factors, in particular magnesium ion concentration, and the adsorption of *Vibrio parahaemolyticus* on the prawns' shell. An elevated Mg^{2+} concentration in the environment increases the

expression of bacterial gene which is responsible for prawn's shell surface interaction, hence postulates a magnified event of bacteria invasion. Upon breaching of this physical barrier, two major divisions of innate immune response are mounted to safeguard the host, which are cell-mediated and humoral immunity.

Cell-mediated immunity is performed by haemocytes, the free circulating cell population found in crustacean's haemolymph. They are deemed as the homologs of white blood cells, playing vital roles such as phagocytosis, encapsulation, nodule formation and prophenoloxidase activation (Giulianini et al., 2007). Generally, three types of haemocytes are defined based on morphological observation in crustacean (Giulianini et al., 2007; Martin & Graves, 1985). There are hyaline cells, small granular cells and large granular cells. The features separating these cells are based on nucleus to cytoplasm ratio and the number and size of the granule present. Hyaline cells lack of granule and have the highest nucleus to cytoplasm ratio; small granular cells display variable moderate size granules in the cytoplasm; and large granular cells are filled with dense, relatively big size granules. Given the limitation of morphology classification, monoclonal antibodies were raised to categorise haemocytes based on cells' surface antigen or specific cells' components, in attempt to define specific functionality and understand cell's lineage development (Braak et al., 2000; Johansson et al., 2000; Rodriguez et al., 1995). The total and differential percentage of these cell population not only differ across species, but also among individual and sensitive to external stimuli (Giulianini et al., 2007). Hyaline cells, in general, represent the largest haemocyte's population in a host, while small and large granule cells constitute the remaining. While all haemocyte's types are capable of phagocytosis with differential activity observed across species, prophenoloxidase system is exclusive to granule-bearing cells only (Giulianini et al., 2007; Smith & Soderhall, 1983b; Supamattaya et al., 2003). Upon triggering, degranulation occurred and the prophenoloxidase was released, activated

through a series of enzyme cascade reactions and resulted in melanisation (Smith & Soderhall, 1983a).

On the other hand, humoral immunity comprises pathogen recognition by cell-bound or circulating receptors, antimicrobial peptide synthesis, reactive oxygen species production, as well as the clotting cascade. The expression of these components is usually under tight regulation in ambient condition and demonstrates sensitive fluctuation in the host over the course of pathogenic attack. Toll pathway and immune deficiency (IMD) pathway are two important and well-studied immune pathways in *Drosophila* (Chen et al., 2014). The former is responsive towards Gram-positive bacteria and fungi, while the latter is preferentially activated by Gram-negative bacteria. The end product of these pathways is an activated transcription factor, which enters the nucleus and highly promotes the transcription rate and expression of antimicrobial peptides. Penaeidins, crustins, and antilipopolysaccharide factor are some examples of antimicrobial peptides that display significant expression during microbial invasion (Chen et al., 2014). So far, many of the major components involved in the signalling cascade of these pathways were discovered in shrimp, such as Toll-like receptors, Pelle, MyD88, Tube, Dorsal in the Toll pathway; and Relish in IMD pathway (Chen et al., 2014). Because viral infections are a major concern to the industry, the dynamic between viral pathogen and these pathways generates intense attention. Surprisingly, Wang and colleagues (2014) pointed out that white spot syndrome virus (WSSV) can hijack and benefit from Toll or IMD pathways' activation. The virus encodes proteins that can activate Toll pathway and also genes that can be initiated by the activated transcription factor. While immune-related genes are indeed up-regulated after WSSV infection, these viral early genes are simultaneously produced, participate in successive viral replication process. Subsequently, the virus may shutdown the Toll pathway with inhibiting products, hampers the host's ability to synthesise antiviral peptides.

2.4 Pattern recognition receptor and Lectin

The hallmark of an immune system builds on the discrimination of self and non-self molecules, hence signals an intrusion of a foreign entity. In innate immunity, the recognition and binding of the pathogen molecules with host's recognition receptor is an indispensable step for immune system initiation. Crustacean produces a myriad of cell-associated or circulating proteins to serve for this purpose. They are termed pattern recognition receptors (PRRs), while the target recognised is called pathogen associated molecular pattern (PAMP). PAMPs are general features found among pathogens, representing an important property for pathogens' survival and development. These include cell wall constituents such as lipopolysaccharides, β -1,3-glucans, peptidoglycans, and even nucleotide materials from some viruses (Cerenius et al., 2010). Many PRRs have been isolated and cloned from various crustaceans, namely Toll-like receptor (TLR), β -glucan-binding-protein (β GBP), lipopolysaccharide-and-glucan-binding-protein (LGBP), and lectins. These PRRs display preferential affinity towards different PAMPs and trigger an appropriate immune response. The presence of vast array of PRR repertoire, with differences in structural profile and reactivity towards PAMPs, in a sense, compensates the limitations in the lacking of a real adaptive immunity in crustaceans.

Lectins are found abundant in almost every organism (Wang & Wang, 2013). They are defined as carbohydrate-binding proteins which harbour various carbohydrate recognition domain (CRD) and can be categorised based on their CRD, affinity towards specific carbohydrates moieties and protein domain structure. Their biological roles are diverse and essential, including protein trafficking and sorting, cell signalling, cell communication and pathogen recognition (Wang & Wang, 2013). Lectins commonly assemble into oligomers *in vivo* with the presence of more than one CRDs (Vasta et al., 2007). Seven types of lectins were identified in shrimps, namely C-type, L-type, M-type, P-type, fibrinogen-like domain lectin, galectins, and calnexin/calreticulin (Wang &

Wang, 2013). C-type lectin is the most studied, while others are fairly investigated. C-type lectin describes a Ca^{2+} -dependant manner in carbohydrate binding; L-type lectin consists of CRD homologous to leguminous lectin; M-type lectin is a type II transmembrane protein with glycoside hydrolase family 47 protein group; P-type lectin refers to either cation-dependent or independent CRD specific towards mannose-6-phosphate; fibrinogen-like domain lectin recognises acetyl groups in both carbohydrate and non-carbohydrate compounds; galectin shows similarity in overall topology to L-type lectin's CRD and binds to galactose; and calnexin interacts with N-linked glycans and represents as the prototype of a small group of ER-resident chaperone protein (Drickamer, 2014). Moreover, lectins can be classed into intracellular and extracellular proteins. Calnexin, M-type, L-type, and P-type lectins belong to the former; while C-type, galectins, and some fibrinogen-like domain lectins belong to the latter (Drickamer, 2014).

The perspective of lectin as an immune-related protein is durable owing to its ability to recognise specific carbohydrate moiety and agglutinate most of the bacterial cells. Indeed, the agglutination of the foreign object by lectins in host system resembles antibody opsonisation in adaptive immunity. Many studies have demonstrated the immune functions of lectins in shrimp model, emphasising on C-type lectins. *PmCLec* was cloned from black tiger shrimp *Penaeus monodon*. Purified *PmCLec* protein agglutinates Gram-positive bacteria but does not show affinity towards Gram-negative bacteria, and the knockdown shrimps displayed higher cumulative mortality towards *Vibrio harveyi* infection (Wongpanya et al., 2017). In the work by Wang and colleagues (2009), a C-type lectin *FcLec3* from *Fenneropenaeus chinensis* (Chinese white shrimp) was cloned. The recombinant *FcLec3* recognises muramic acid and peptidoglycan, agglutinates both Gram-positive and negative bacteria, and could interact with WSSV VP28 envelop protein. In addition, *FcLec4* isolated from the same species was postulated to function as an opsonin with β -integrin as the receptor (Wang et al., 2013). Suppression

of *FcLec4* expression delayed bacterial clearance, and the knockdown of β -integrin retards the opsonic activity of this lectin. Moreover, *MrLec* from *M. rosenbergii* agglutinates both Gram-positive and negative bacteria in a Ca^{2+} -dependent manner, and exhibits significant up-regulation in gene expression level after challenged with pathogens *Vibrio parahaemolyticus* and WSSV (Feng et al., 2016). While most lectins' functions were fulfilled as PRR to recognise invading objects and stimulate host's immune system, exceptions were observed. There are reports on C-type lectins isolated from *Fenneropenaeus chinensis*, denoted as *Fc-hsL* (Lai et al., 2013) and *FcCTL* (Sun et al., 2008); two from *Eriocheir sinensis* termed *rEsLecA* and *rEsLecG* (Jin et al., 2013), were demonstrated to directly inhibit the growth of both Gram-positive and negative bacteria.

2.5 L-type Lectin

L-type lectin (LTL) shares CRD resemblances to those present abundantly in seeds of leguminous plants, hence the designated name. The L-type lectin domain (LTL D) is composed of two antiparallel β -sheets motifs connected by loops and turns, devoided of α -helices. Plant lectins are formerly recognised as haemagglutinin due to the ability to agglutinate red blood cells (Lagarda-Diaz et al., 2017). They constitute a large and diverse group of highly homologous proteins. Despite sugar-binding specificity and affinity differences, their tertiary structures often exhibit substantial similarity (Sharon & Lis, 2002). Plant lectins share nearly identical three-dimensional structure consisting of antiparallel β -sheets referred as jelly roll fold, with carbohydrate-binding pocket indented on the protein's subunit surface. Metal-binding sites for divalent cation, Ca^{2+} and Mn^{2+} , are located in proximity to the carbohydrate-binding site (Sharon & Lis, 2002). Plant lectins commonly exist in dimers or tetramers, and their assembly determines the multivalency and subsequent sugar-binding specificity and strength of the protein. Concanavalin A is the first plant lectin being crystallised with three-dimensional structure

identified (Loris et al., 1998). The Asp-Asn-Arg residue triad forms the critical carbohydrate-binding motif in the CRD domain of Concanavalin A. Additionally, a cis-peptide bond is always found preceding the Asp residue of the activity triad (Loris et al., 1998). There are about 100 members of plant lectins that have been isolated and characterised, with variety of functions being elucidated. In addition to the proposed antimicrobial properties as well as deterrent against insects, plant lectins were shown to assist in symbiosis relationship by anchoring nitrogen-fixing rhizobia to leguminous plants (Sharon & Lis, 2002; Sharon & Lis, 2004). Moreover, there are robust investigations on antitumor properties of plant lectin such as Concanavalin A via targeting programmed cell death and induced-apoptosis in malignant cells (Fu et al., 2011; Liu et al., 2010).

LTL identified in animal are intracellular type 1 membrane-bound proteins, with luminal LTLD take roles in protein trafficking, sorting and targeting (Xu et al., 2014). Vesicular integral membrane protein-36kDa (VIP36), ER-Golgi intermediate compartment protein-53kDa (ERGIC-53), ERGL, and VIPL are L-type lectins involved in early secretory pathway in mammalian animals (Kamiya et al., 2007; Xu et al., 2014). The interactive function of these LTLs in secretory pathway is postulated in the work of Kamiya's team (2008). VIPL is a resident protein in endoplasmic reticulum, which binds to newly folded proteins and protect them from endogenous degradation. ERGIC-53 chauffeurs between ER and ERGIC, acts as a cargo receptor and transports secretory glycoproteins toward Golgi apparatus. VIP36 is mainly found on cis Golgi, and serves as quality checkpoint to prevent misfolded glycoproteins from escaping out from the cell. In general, animal's LTLs interact with N-glycan of glycoproteins and demonstrate strong binding affinity towards mannose-type residues in dependent on Ca^{2+} (Huang et al., 2014).

A few studies have investigated and reported the potential roles of L-type lectin in immune response of different aquatic animals. There are three LTL genes, corresponding to ERGIC-53, VIP36, and VIPL, were characterised in channel catfish, and significant up-regulation of these genes were observed during artificial challenge by bacterium *Edwardsiella ictalurid* (Zhang et al., 2012). Oriental river prawn *Macrobrachium nipponense* L-type lectin, MnLTL1, has a deduced protein length of 323 amino acids with a luminal LTLD and a 23-a.a. type 1 transmembrane region (Xiu et al., 2015). This LTL displayed significant fluctuation during bacterial challenge. Xu and colleagues (2014) isolated an LTL gene and a disintegrin and metalloprotease-like protein (ADAM) gene from *Marsupenaeus japonicas*, denoted as *MjLTL1* and *MjADAM*, respectively. Interestingly, the authors suggested that *MjADAM* with proteolytic activity cut the ectodomain of *MjLTL1*, which subsequently released into extracellular space and serves as an opsonin for enhanced phagocytosis. Moreover, two LTLs, *EsERGIC-53* and *EsVIP36*, originated from Chinese mitten crab, *Eriocheir sinensis* have been shown to facilitate *Vibrio parahaemolyticus* clearance in the host (Huang et al., 2014). These proteins can bind to lipopolysaccharide (LPS), peptidoglycan (PGN), mannose (Man), and N-Acetyl-D-mannosamine (ManNAc). On the contrary, a down-regulation of L-type lectin during WSSV challenge was detected in *Procambarus clarkii*, the red swamp crayfish (Dai et al., 2016). The research team showed that the recombinant L-type lectin, termed *PcL*-lectin, can interact with the envelope protein VP24 of WSSV and increased the viral multiplication in the host, indicating a possible involvement of *PcL*-lectin in WSSV replication. Collectively, these reports offer a likelihood of L-type lectins involvement in immune system, and the detail mechanism is yet to be elucidated.

2.6 Recombinant protein technologies

Recombinant cloning is a powerful tool to obtain a desired protein in high quantity and purity. Instead of direct extraction and purification from its original host, recombinant

cloning technology utilises the synthesis machinery of an expression host, usually a bacterium, to manufacture a particular protein in relatively high yield within a significantly short period. The whole process can be categorised into three major stages, namely recombinant vector construction, expression host transformation, and recombinant protein expression.

The choice of an expression host is so critical that it directly affects the entire cloning process, transformation strategies, downstream purification process, final product activity and quality, and hence, the costs required (Demain & Vaishnav, 2009). Generally, a simple manipulative host, vector's features which facilitate downstream purification in combination with less demanding incubation requirements are the ideal conditions. *Escherichia coli*, the single cell bacterium, is the system of choice most of the time. The growth parameters of *E. coli* are well-established, the least expensive, simplest and most manipulative for production. However, prokaryotic system suffers from small heterologous protein size, unfavourable for disulphide bond formation and the absence of post-translational modifications. Eukaryotic cells are chosen to overcome the limitations of prokaryotic system. Yeasts, insect cells, and mammalian cells are used to manufacture more complex recombinant proteins when accurate folding and glycosylation are essential for a functional product. While both yeast and mammalian cells perform post-translational modifications, the former adopts a different glycosylation pattern renders some recombinant product less effective compared to those produced via mammalian cells (Demain & Vaishnav, 2008). However, the high cost of maintaining mammalian cell line and relatively lower yield are the factors impeding the wide implementation of this system.

Plasmid represents the most convenient way to introduce a foreign gene into an expression host. Commercial plasmid vectors come with variety of features tailored

towards customised needs and consideration. An expression vector generally composed of a strong transcription promoter site; a multiple cloning site (MCS); an antibiotic resistance gene; and different fusion protein tags. An MCS is a short nucleotide sequence incorporated with multiple restriction endonuclease recognition sites, which in turn facilitates directional insertion into the plasmid. After that, the host transformed with recombinant plasmid need to be screened to identify the plasmid. The screening process is commonly achieved through plating of the transformed host on antibiotic-supplemented agar plate, and only those carrying the antibiotic gene vector will form colonies on the plate. Nucleotide sequencing can then be performed to identify the correct insertion of the gene of interest among these recombinant plasmids.

Fusion tags offer various applications in the recombinant protein expression process (Novagen, 2003). Tags such as glutathione-S-transferase (GST), thioredoxin (Trx), N utilisation substance (NusA) are highly soluble polypeptide by itself, and fusion with these tags has been shown to enhance the solubility of the fused protein. Furthermore, fusion of catalytic enzymes including thioredoxin, DsbA, and DsbC can be used to facilitate proper disulphide bond formation, or incorporation of a signal peptide to export the fusion protein towards the periplasmic space, which is favourable for disulphide bond formation. To simplify downstream purification process, His-tag short peptide is a convenient tag reacts strongly with nickel ion affinity column. During a purification process, endogenous proteins will be easily eluted from the affinity column whereas His-tag proteins are retained.

Most of the time, overexpression of a recombinant protein within the bacterial host may lead to formation of insoluble inclusion bodies. Inclusion bodies are aggregation of misfolded proteins, resulted from overburden of cell synthesis machinery, instability of heterologous protein in the bacterial cytoplasmic environment, or the fact that disulphide

bond is not permissible for proper folding (Rosano & Ceccarelli, 2014). By far, expression of the recombinant protein under a lower incubation temperature has been shown to suppress inclusion bodies formation. Bacterial cells grow slower under sub-optimal temperature, thus alleviates the stress from intense protein synthesis and permits more time for the newly translated recombinant protein to fold properly.

In conclusion, any efforts are welcomed to ensure a sustainable development and expansion of the shrimp industry. Study of the host's immune response towards pathogens will offer meaningful information to strategise effective intervention during a disease outbreak. Lectins are good candidates due to their critical roles as pattern recognition receptors in the shrimps' immune systems. These proteins can be studied in detail *in vitro* by employing recombinant technology to produce them in large quantity and high quality, especially in study of their structural characterisations and functional analysis.

CHAPTER 3: METHODOLOGY

Preparation and compositions of media and chemicals are included in Appendix.

3.1 Bioinformatics characterisation of *MrLTL* sequence

The gene sequence of *MrLTL* was retrieved from NCBI GenBank with accession ID: KM268862.1. ORFfinder online tool (<https://www.ncbi.nlm.nih.gov/orffinder/>) was used to deduce the potential protein coding sequence. The conserved domain and motif region of the putative protein were analysed using online tools Conserved domain search (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>), Prosite (<https://prosite.expasy.org/>), and InterPro (<https://www.ebi.ac.uk/interpro/>). Signal peptide was predicted through SignalP-4.1 (<http://www.cbs.dtu.dk/services/SignalP/>) and transmembrane region (<http://www.cbs.dtu.dk/services/TMHMM/>) via TMHMM Server v.2.0. Protein parameters were calculated via ProParam (<https://web.expasy.org/protparam/>). Swiss-Model workspace (<https://swissmodel.expasy.org/interactive>) was used to generate *MrLTL* protein structure.

3.2 Sample collection

Healthy live *Macrobrachium rosenbergii* was purchased from a local seafood restaurant (Unique Seafood PJ23, Lot 9B-3, Jalan Kemajuan, Section 13, Petaling Jaya, 46200, Selangor) and dissected in the laboratory. Tissues included eyes, gills, pleopod, stomach, hepatopancrease, muscle, heart, and intestine were collected in sterile 1.5 ml microcentrifuge tubes. Haemolymph was also collected from beneath the carapace by using a 26 G needle attached syringe (3 ml), filled with 0.5 ml 10 % (w/v) sodium citrate as anticoagulant. The tissues were snap frozen with liquid nitrogen and stored in a -80 °C freezer. Working bench and apparatus were sterilised with RNaseZap (Ambion, USA) prior used.

3.3 Total RNA extraction

Total RNA was extracted from hepatopancrease tissue using *TransZol Up Plus* RNA kit (TransGen Biotech, China) following manufacturer's instructions. Briefly, 50-100 mg of frozen tissue was grounded into powder in liquid nitrogen in a pre-chilled mortar. They are transferred quickly into a 1.5 ml microcentrifuge tube filled with 1 ml *TransZol Up* solution. The tissue powder was homogenised with pipetting up and down and incubated at room temperature for 5 min. Then, 0.2 ml chloroform was added into the tube, vortexed for 30 s and incubated at room temperature for 3 minutes. The mixture was centrifuged at 10000 x g for 15 min at 4 °C. After centrifugation, the mixture was separated into three distinct layers, consisting of an upper transparent liquid and a lower pink coloured solution, separated by a cloudy layer in the middle. The transparent aqueous layer was transferred into an RNase-free tube without carryover of the intermediate layer. An equal volume of absolute ethanol was added and mixed gently. The mixture was then transferred into a spin column and centrifuged at 12000 x g for 30 s at room temperature. The flow-through was discarded. The spin column was washed with 500 µl of CB9 solution and centrifuged at 12000 x g for 30 s at room temperature. The flow-through was discarded. This step was repeated once. Next, the spin column was washed with 500 µl of WB9 solution and centrifuged at 12000 x g for 30 s at room temperature. The flow-through was discarded. This step was repeated once. The spin column was then centrifuged at 12000 x g for 2 min to remove any solution residue, and air-dried for 5 min. The spin column was placed into a new 1.5 ml RNase-free tube. RNase-free water, 50 µl, was added onto the spin column matrix and incubated at room temperature for 1 min. The RNA was eluted at 12000 x g centrifugation for 1 min. The elution step was repeated once with another 50 µl RNase-free water. The extracted total RNA was stored in a -80 °C freezer.

3.4 First strand cDNA synthesis

The total RNA extracted from hepatopancrease was used for first strand cDNA synthesis. GoScript™ Reverse Transcription System (Promega, USA) kit was used following the manufacturer's instructions. Firstly, 4 µl of total RNA (up to 5 µg per reaction) and 1 µl Oligo(dT)₁₅ primer was mixed, heated in a 70 °C heat block for 5 min and immediately placed on ice for at least 5 min. Reverse transcription reaction mix with 15 µl in total was prepared, consisting of a final concentration of 1 X GoScript™ reaction buffer, 1.5 mM MgCl₂, 0.5 mM of nucleotide mix, 1 µl GoScript™ Reverse Transcriptase, 1 µl gDNA remover, and 6.8 µl of nuclease-free water. The RNA mix and reverse transcription reaction mix were combined. The reaction mixture was then incubated at 25 °C for 5 min, followed by 42 °C for 1 hour, and lastly 70 °C for 15 min. The cDNA synthesised was stored at -20 °C freezer.

3.5 PCR of *MrLTL* gene

The described polymerase chain reaction (PCR) reaction mix and protocol were applied to all PCR reaction unless otherwise mentioned. GoTaq® Flexi DNA polymerase components were used (Promega, USA). The PCR reaction mix consisted of a final concentration of 1 X GoTaq® Flexi buffer (proprietary formulation, Promega, USA), 1.5 mM MgCl₂, 0.2 mM of dNTP, 0.2 µM of each forward and reverse primer, 1.25 u GoTaq® Flexi DNA polymerase, 0.5 µl cDNA template, and nuclease-free water. The PCR protocol was conducted at initial denaturation step at 95 °C for 3 min, followed by 35 thermal cycles of 95 °C for 30 s, annealing step for 30 s (annealing temperature tabulated at Table 3.1), extension step at 72 °C for 1 min (1000 bp/1 min), and a final extension at 72 °C for 5 min. Primer set 1 was used to directly amplify *MrLTL* gene from cDNA template. The amplified product was then subjected to second PCR reaction using both primer sets 2 and 3. Primer set information is tabulated in Table 3.1. Primer sets 2 & 3 were designed with incorporated restriction endonuclease sites. PCR products were

analysed on a 2% (w/v) agarose gel electrophoresed in 1 x TAE buffer, at 80 V, 180 mA, for 40 min. The sequences were identified via nucleotide sequencing service (Sanger sequencing) provided by MyTACG Bioscience Enterprise.

Table 3.1: Oligonucleotides and their respective sequences.

Primer set	Primer ID	Primer sequence	Annealing temperature (°C)	Amplicon size (bp)
1	LTLF6	5' GCA AGT CCA GGC AAA ACA AC 3'	52 °C	1247 bp
	LTLR1252	5' GGC ATG GAC GAT GTC TTA AC 3'		
2	<i>Mr</i> LTL_F	5' GCG GGC <u>A▼CA TGT*</u> TGG GAG TAG TAA ATG TTG 3'	63 °C	995 bp
	<i>Mr</i> LTL_R	5' CTG GCA <u>C▼TC GAG⁺</u> TTA ATA GAA CCT CTT CCG 3'		
3	<i>Mr</i> LTL_D_F	5' GCG GGC <u>A▼CA TGT*</u> TGG ATT ATA TGA AGA GAG 3'	63 °C	698 bp
	<i>Mr</i> LTL_D_R	5' CGG GCA <u>C▼TC GAG⁺</u> TTA GAG GTC ATA TAA TTT G 3'		
4	T7ProF	5' TAA TAC GAC TCA CTA TAG GG 3'	47 °C	Depend on insert size
	T7TerR	5' GCT AGT TAT TGC TCA GCG G 3'		

*Restriction endonuclease recognition site for *Pci*I.

⁺Restriction endonuclease recognition site for *Xho*I.

▼Restriction endonuclease cut site.

3.6 Purification of PCR products from agarose gels.

PCR product was purified from agarose gels using NucleoSpin® Gel and PCR Clean-Up kit (Macherey-Nagel, Germany). After electrophoresis, the DNA band with expected size was excised and placed into a 1.5 ml microcentrifuge tube. For every 100 mg of 2% (w/v) agarose gel, 400 µl NTI buffer was added. The tube was incubated at 50 °C until the gel slice is completely dissolved. A filter column was placed in a 2 ml collection tube, and 700 µl of the gel mixture was loaded. The tube was centrifuged at 11000 x g for 30 s and the flow-through was discarded. This step was repeated until all the remaining sample was loaded. Next, 700 µl of Buffer NT3 was added into the column, centrifuged at 11000 x g for 30 s, and flow-through was discarded. This step was repeated once. A 1-minute

centrifugation at 11000 x g was performed to completely remove any remaining buffer solution. The column was transferred into a new 1.5 ml microcentrifuge tube. Next, 15 µl of elution buffer was added into the column and incubated at room temperature for 1 min. Lastly, the PCR product was eluted at 11000 x g centrifugation for 1 min. Concentrations of all purified PCR products from agarose gels were determined by using a NanoDrop™ 2000 spectrophotometer (ThermoFisher Scientific, US). The purity of all purified PCR products were assessed based on the 260/280 ratio and 260/230 ratio by using the same spectrophotometer.

3.7 DNA cloning

3.7.1 Restriction endonuclease digestion

Two sequence regions of *MrLTL* gene were cloned into the expression vector, pET-30(a) (Novagen, Germany). There were the full coding sequence *MrLTL* (sequence position at nucleotide 25 to 996), and L-type lectin domain only sequence *MrLTL*D (sequence position at nucleotide 97 to 768) (Figure 4.1). The deletion mutant, *MrLTL*D sequence lacks both signal peptide (M¹ (Met) to I¹⁶ (Ile)) and transmembrane region (V²⁹¹ (Val) to A³¹³ (Ala)). These sequences were screened online using Webcutter 2.0 (<http://www.firstmarket.com/cutter/cut2.html>) to identify the suitable restriction endonuclease (RE) to use for digestion. Primers used to amplify both *MrLTL* and *MrLTL*D sequences were incorporated with respective RE site. A 25 µl of double RE digestion reaction mix was set up, which included a final concentration of 0.5 µg nucleic acid product, 1 x NE 3.1 buffer (composes of 100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl₂, 100 µg/ml BSA, pH 7.9), and 5 u each for both restriction enzymes. In general, the nucleic acid product was digested with two different RE, in a step-wise manner, in which second RE will be added only after inactivation of the first RE. *MrLTL* and *MrLTL*D were digested with *Pci*I and *Xho*I, while vector pET-30(a) was digested with *Nco*I and *Xho*I. *Nco*I RE site is found in and cut within both *MrLTL* and *MrLTL*D

sequences. Hence, *PciI* was used instead to generate compatible overhangs (5'-A[▼]CA TGT-3') to *NcoI* (5'-C[▼]CA TGG-3'). The digestion profile for *PciI* is 37 °C for 60 min and inactivation at 80 °C for 20 min; *XhoI* is 37 °C for 15 min and inactivation at 65 °C for 20 min; and *NcoI* is 37 °C for 15 min and inactivation at 80 °C for 20 min. All RE treated samples were purified from agarose gel prior ligation step. All restriction endonucleases were purchased from NEB, USA.

3.7.2 Ligation process

T4 DNA ligase (3 u/μl) (Promega, USA) and 2 X rapid ligation buffer (60 mM Tris-HCl (pH 7.8), 20 mM MgCl₂, 20 mM DTT and 2 mM ATP) (Promega, USA) were used. The concentrations of insert sequences and pET-30(a) vector were determined using a NanoDrop™ 2000 spectrophotometer. A total 10 μl of ligation reaction mix were set up, which included 5 μl of 2 X rapid ligation buffer, 1 μl of T4 DNA ligase, 2 μl of pET-30(a) vector and 1 μl of insert DNA sample. The amount of vector to insert is in the ratio 1:3, calculated via following equation (Promega, USA):

$$\frac{ng\ of\ vector\ \times\ kb\ size\ of\ insert}{kb\ size\ of\ vector} \times insert : vector\ ratio = ng\ of\ insert\ needed$$

The reaction mix were incubated at room temperature for 3 hr. To achieve maximum ligation performance, the incubation duration can be extended to overnight at 4 °C.

3.8 Transformation of expression host

The recombinant plasmids pT30-*MrLTL* (pET-30(a) with *MrLTL* insert) and pT30-*MrLTL*LD (pET-30(a) with *MrLTL*LD insert) were introduced into the expression host *E. coli* BL21(DE3) (Genotype: *fhuA2* [*lon*] *ompT* *gal* (λ DE3 = λ *sBamHI*o Δ *EcoRI*-B *int*::(*lacI*::*PlacUV5*::T7 *gene1*) *i21* Δ *inl5*) [*dcm*] Δ *hdsS*) (NEB, USA) with the following procedures. First, a vial containing 50 μl of competent BL21(DE3) cells was thawed on ice. Then, 1 μl of recombinant plasmid was added and mixed gently with few tapping.

The vial was then incubated on ice for 30 min without mixing. The cells were heat shocked at 42 °C for 10 s without mixing, and immediately returned on ice and rested for at least 5 min. Next, 950 µl of SOC medium (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, and 20 mM glucose) was added and the vial was incubated at 37 °C for 1 hr with agitation at 220 rpm. Then, 200 µl of the mixture was plated on a LB agar plate supplemented with 30 µg/ml kanamycin. After overnight incubation, colonies formed were screened through PCR using primer set 4 (Table 3.1), and nucleotide sequences of the recombinant plasmids were identified.

3.9 Induction of recombinant protein expression

The colonies with the correct recombinant plasmids pT30-*Mr*LTL and pT30-*Mr*LTLΔ were used as inoculum. An initial inoculum was prepared by culturing the colony in 5 ml LB media with 30 µg/ml kanamycin and incubated at 37 °C, 220 rpm overnight. On the next day, the 1 ml culture was used to inoculate a 50 ml LB media with 30 µg/ml kanamycin. When OD₆₀₀ of the culture reached 0.6-1.0 at around 4 hr incubation, IPTG solution was added into the medium with a final concentration of 1.0 mM. The cell culture was continued incubating at 37 °C, 220 rpm for another 4 hr. After that, the cells were collected through centrifugation at maximum speed for 10 min and the supernatant was discarded. The harvested cells were analysed with SDS-PAGE as described in section 3.12.

3.10 Optimisation of recombinant protein expression.

3.10.1 Temperature

Bacteria colonies containing the recombinant plasmids pT30-*Mr*LTL and pT30-*Mr*LTLΔ were inoculated into 5 ml LB media with 30 µg/ml kanamycin and the expression of the recombinant protein was induced with 0.2 mM IPTG as described in section 3.9 with altered induction temperature and duration at 20 °C (overnight), 25 °C

(overnight), and 37 °C (4 hr), respectively. The harvested cells were analysed with SDS-PAGE as described in section 3.12.

3.10.2 IPTG concentration.

Bacteria colonies containing the recombinant plasmids pT30-*MrLTL* and pT30-*MrLTL*D were inoculated into 5 ml LB media with 30 µg/ml kanamycin and the expression of the recombinant protein was induced with altered IPTG concentrations of 0.1 mM, 0.2 mM, 0.5 mM, and 1.0 mM IPTG, as described in section 3.9. The cells were incubated at 25 °C, for overnight, at 220 rpm. The cells were harvested and analysed with SDS-PAGE as described in section 3.12.

3.10.3 Additives

Two additives were tested: (A) 0.5 mM MgCl₂ + 0.5 mM CaCl₂ and (B) 3% ethanol. Bacteria colonies containing the recombinant plasmids pT30-*MrLTL* and pT30-*MrLTL*D were inoculated into 5 ml LB media with 30 µg/ml kanamycin and the expression of the recombinant protein was induced with 0.2 mM IPTG as described in section 3.9 with the additives supplemented separately into the LB medium. The cells were incubated at 25 °C, for overnight, at 220 rpm. The cells were harvested and analysed with SDS-PAGE as described in section 3.12.

3.11 Cell lysis

Cells collected from step 3.9 were resuspended in 1 ml of lysis buffer (50 mM Tris, 150 mM NaCl, 1 % (v/v) Triton-X 100, 5 % (v/v) glycerol, pH 8.0) per 50 ml cell culture. The cells were lysed with 5 - 8 cycles of freeze-thaw process using liquid nitrogen. After that, 5 µg/ml of DNase and 5 µg/ml RNase were added into the mixture and incubated on ice for 30 min to reduce viscosity if necessary. The mixture was then centrifuged at 12,000 x g for 10 min at 4 °C to separate both insoluble debris and supernatant. Both fractions were collected and stored at -20 °C.

3.12 SDS-PAGE Analysis

SDS-PAGE was used to separate the expressed recombinant protein. A 12% separating gel and 5 % stacking gel were prepared. For 10 ml tested sample in pellet form, 100 μ l of 1 x SDS gel-loading buffer was added, and heated at 100 °C for 3 min. For tested sample in solution form, 75 μ l of liquid sample was mixed with 25 μ l of 4 x SDS gel-loading buffer and heated at 100 °C for 3 min. Next, 10 μ l of the denatured sample was loaded in each well and the gel was electrophoresed at 180 V for 40-50 min. After the electrophoresis, the gel was stained in Coomassie Blue Staining solution (Bio-Rad, California) for 4 hr, and destained in destaining solution overnight. The gel image was documented with a digital camera (ASUS ZenFone Max Pro M1).

3.13 Western Blotting

Western blotting device (TE70X Semi-dry Blotter, Hoefer, USA) was set up by following the manufacturer's instructions. Polyacrylamide gel to be blotted and blotting papers were neutralised by soaking in Towbin buffer for 15 min. The PVDF membrane was first activated in methanol and then soaked in Towbin buffer for 15 min to neutralise it. The transfer stack was arranged in an order as the blotting papers was placed at the bottom, and then the PVDF membrane, polyacrylamide gel placed above the membrane, and lastly another layer of blotting papers on the top as a complete transfer stack. The blotting process was electrophoresed at 45 mA for 1 hr. A complete transfer of the protein markers on the PVDF membrane indicates a successful blotting process. After the transfer process, the PVDF membrane was soaked in 5% skimmed milk in TBST buffer at 4 °C overnight for blocking step. The membrane was rinsed three times with TBST buffer. A 1: 2000 dilution of anti-His Tag Antibody, HRP conjugate (Invitrogen, USA) in 5 % skimmed milk + TBST buffer was added on the membrane and incubated at 4 °C overnight. After the incubation, the membrane was rinsed three times with TBST buffer.

Pierce™ 1-Step Ultra TMB Blotting solution (Invitrogen, USA) was then applied on the membrane and incubated for 30 min for colour development.

University of Malaya

CHAPTER 4: RESULTS

4.1 Bioinformatics characterisation of *MrLTL* gene sequence.

4.1.1 *MrLTL* coding sequence and protein features.

The gene sequence of *MrLTL* retrieved from NCBI GenBank has 1633 nucleotides (Figure 4.1). A protein coding region of 972 bp nucleotides was predicted, which encodes for a putative protein of 323 amino acids. One conserved domain was found on the putative protein sequence, namely L-type lectin domain ranges from amino acid D²⁵ (Asp) to L²⁴⁸ (Leu). There is a predicted signal peptide from amino acids M¹ (Met) to I¹⁶ (Ile) and a transmembrane region from amino acids V²⁹¹ (Val) to A³¹³ (Ala). Two carbohydrate binding motifs Y¹³⁵-S¹³⁶-N¹³⁷ (Tyr-Ser-Asn) and G²³²-D²³³-L²³⁴ (Gly-Asp-Leu) were found. In addition, there are three amino acids showing carbohydrate binding affinity, which are S⁶⁸ (Ser), D¹⁰² (Asp), and H¹⁶² (His), respectively. Four amino acids were deduced with metal binding affinity, particularly calcium ion, which include D¹³³ (Asp), Y¹³⁵ (Tyr), N¹³⁷ (Asn), and D¹⁶⁵ (Asp).

1 -	GGCAAGTCCAGGCAAAACAACGTAATGGGAGTAGTAAATGTTGTGCTATTT	- 51
1 -	M G V V N V V L F	- 9
52 -	TTGTCCTTTGTGTCGTTATATCGCCTCAAGCATGGAATACCTACGATTAT	- 102
10 -	L S F V S C I S P Q A W N T Y D Y	- 26
103 -	ATGAAGAGAGAACATTCTTTGGTCAAACCATAACCAAGGGGTTGGCACATCG	- 153
27 -	M K R E H S L V K P Y Q G V G T S	- 43
154 -	ATACCCTACTGGGATTTCTTAGGGTCTACCATGGTAACAAGTAATTATGTA	- 204
44 -	I P Y W D F L G S T M V T S N Y V	- 60
205 -	CGACTTACGGGAGACATTCAAAGCCAGCGTGGGGCCATTTGGAATAAGGTC	- 255
61 -	R L T G D I Q S Q R G A I W N K V	- 77
256 -	CCTCTCTCAGTCCGCAACTGGGAGATGCAGATCCATTTCAAAGTCCATGGA	- 306
78 -	P L S V R N W E M Q I H F K V H G	- 94
307 -	AGAGGGAAGGACTTGTTCGGTGATGGCTTTGCCTTCTGGTATGTGAAGGAG	- 357
95 -	R G K D L F G D G F A F W Y V K E	- 111

Figure 4.1: *MrLTL* nucleotide sequence and translated putative amino acid sequence. The amino acid was highlighted in colours indicating different predicted protein features: Green: L-type lectin domain; Yellow: carbohydrate-binding site; Pink: calcium-binding site; Grey: transmembrane helix.

```

358 - CCAATGCAAACCTGGTGATGTATTTGGGAGCAGAGATTTCTTCACTGGCTTA - 408
112 - P M Q T G D V F G S R D F F T G I - 128

409 - GCCATTATTGCCGACACTTACAGCAATCATAATGGAGTCCACAATCACGGT - 459
129 - A I I A D T Y S N H N G V H N H G - 145

460 - CACCCATATATCTCGGCAATGGTGAACAATGGCACCTTGCACCTACGACCAT - 510
146 - H P Y I S A M V N N G T L H Y D H - 162

511 - GACCCGATGGCACCCACACCCAGCTTTCAGGATGTGTGGCTAAATTTAGA - 561
163 - D R D G T H T Q L S G C V A K F R - 179

562 - AATTTGGATCATGATACCTTCTTATCAATCAAATATGTGCATGATACATTA - 612
180 - N L D H D T F L S I K Y V H D T L - 196

613 - ACGGTGTCAGTTGATATAGACAACAAGATGGCATATAAGGATTGCTTCACA - 663
197 - T V S V D I D N K M A Y K D C F T - 213

664 - GTAAATGGTGTTTTCTTGCCAACTGGTTATTACTTTGGTGTGTCTGCTGCT - 714
214 - V N G V F L P T G Y Y F G V S A A - 230

715 - ACAGGTGACCTAAGTGATGCCCATGATATCGTCTCTCTCAAATTATATGAC - 765
231 - T G D L S D A H D I V S L K L Y D - 247

766 - CTCACAACCCAGATGATGATATTTTTAGAAGATCGAGCTAATATCATGCCT - 816
248 - T T T P D D D I L E D R A N I M P - 264

817 - TCCGCTTATACATGGAGCCTCCAAGAGATCATGTTGATGACCCCAAGCCA - 867
265 - S A L Y M E P P R D H V D D P K P - 281

868 - TCTTCGTTGTCTGTTTTGGAAGCAGCTAGTGCTCCTGGTGTGTTGGTGTCTGT - 918
282 - S S L S V W K Q L V L L V F G V C - 298

919 - GCCATATGTGGTGCAGTCTTTATAGGGGGTCTCTTTTACGTCAAGCATAAA - 969
299 - A I C G A V F I G G L F Y V K H K - 315

970 - GAACAGCAACGGAAGAGGTTCTATTAA - 997
316 - E Q Q R K R F Y * X - 332

998 - TTTTCTAGAGGGTGTGTTAAAGATTATTGTAGATATTGTTTTTCCACATGGAAAGAAC
ATGGTTGGGTTTTATTTTGTGTCATCACTTTATATTTTATATTAGACTAAATGAATAATTAAGCCTAGC
CATTATAGTCAGTTATTTATGGAATAGTTAACGTTAAATTTATCTTATACCCCTAAGAAAAAAAAATCTTA
CTTCACATGGAATTAATAGAACATTTGTAATGTTAAGACATCGTCCATGCCACATGAAAGTTATTTTGT
GGTATTTCTGTGATCATCATCAGTAGCCAGCAATGGGGATAAAAAATGACAGTTACAAGAAATACCATG
TTTGTAGTCATTTTAAATTTTCTTTTGTAGCTTGCATATCAAGAGGTAAAAATTTGAAGGAAGAGGTTA
TAACACTCAGGTTCTTAGTTCTCTTGTATATACAAAGAGGCCGTATTTATTGAGATTGAATTCATGAATTC
GTGTATGGTATTACTTTCACTTGTGAGTGAATTTAAAGCACAACTTACATAACACAGTACATTTAATTTT
ATAATAGAGACTGTATCTGACTTCCGTCTTTTACACTGAAAATTTTACCTGATGTATACCTGTATAATA
CTGTAGTAAGGAA - 1633

```

Figure 4.1, continued.

4.1.2 *MrLTL* and *MrLTLD* ProtParam parameters.

MrLTL is a moderate-sized protein with 323 amino acids and with molecular mass (M_r) of 36 kDa (Table 4.1). It has an isoelectric point (pI) slightly acidic than water. Despite being a transmembrane protein, it has a negative 0.173 of Grand average of

hydropathicity value (GRAVY). A negative GRAVY value implies hydrophilicity of the peptide sequence. *MrLTL*D domain coding sequence was also analysed. The *MrLTL*D has 226 amino acids with position ranges from D²⁵ (Asp) to L²⁴⁸ (Leu) of the full length *MrLTL* protein. The removal of signal peptide and transmembrane sequence reduced the *M_r* to 25 kDa and GRAVY value to -0.283. The addition of fusion tag (His₆-tag) from the pET-30(a) expression vector has added another 5 kDa to the recombinant proteins and further decreases the GRAVY value by roughly 1.5-fold. In both cases, the pI value remained rather constant.

Table 4.1: ProtParam physical and chemical values for both *MrLTL* and *MrLTL*D.

Parameters	<i>MrLTL</i>	<i>MrLTL</i> D domain
Number of amino acids	323 (368)*	226 (270)
Molecular mass (Da)	36446.34 (41416.76)	25591.75 (30449.02)
Theoretical pI (pH)	6.18 (6.16)	6.07 (6.09)
Grand average of hydropathicity (GRAVY)	-0.173 (-0.295)	-0.283 (-0.446)

*Figures in parentheses are the values of recombinant proteins containing the fusion tag (His₆-tag).

4.1.3 *MrLTL* structural features.

A predictive *MrLTL* protein structure was built through homology modelling with Swiss-Model workspace. A canine VIP36 protein (PDB:2DUP) was selected as the template with sharing sequence identity of 56.15% with *MrLTL*. The predictive structure of *MrLTL* covered amino acids 26 to 271. The model shows 94.26% of Ramachandran favour with two outliers, Lys⁹⁷ and Asp²⁵² (Appendix D). The protein structure comprises only beta-strands. There are 18 strands arranged antiparallely, connected by loops and turns, which can be grouped into two major beta-sheets and folded into a beta-sandwich conformation (Figure 4.2). Figure 4.3 provides a closer look at the active motifs and the

metal-binding site. The folding brings both functional motifs and reactive amino acids in proximity. Carbohydrate binding motifs are present on the loop area of the protein's surface. The metal binding amino acids are located near to and adjoining with the YSN motif. Figure 4.4 compares L-type lectins from different organisms and found that all of them adopt a similar beta-sandwich conformation.

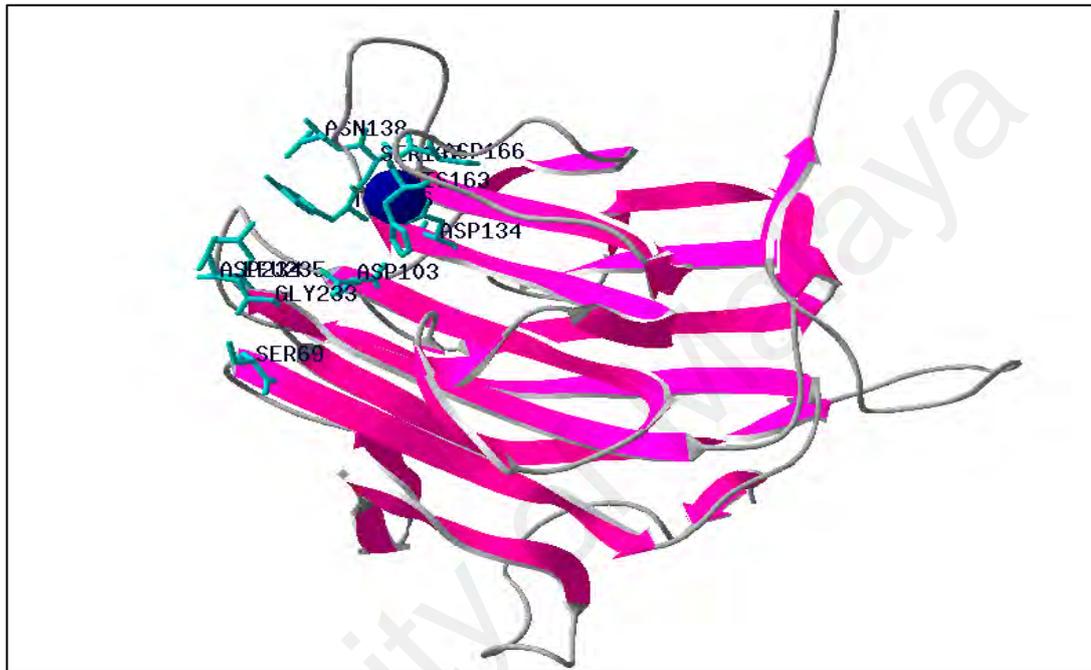


Figure 4.2: *MrLTL* predicted protein structure. This model is built via Swiss-Model workspace online server. The template used for modelling is canine VIP36 (PDB: 2DUP) with sharing sequence identity of 56.15%. Ramachandran plot showed 94.26% favour. Carbohydrate binding motifs: Y¹³⁵-S¹³⁶-N¹³⁷ and G²³²-D²³³-L²³⁴; carbohydrate recognition amino acids: S⁶⁸, D¹⁰², and H¹⁶²; metal binding amino acids: D¹³³ and D¹⁶⁵; are highlighted in backbone structure and coloured in cyan. The blue sphere is calcium ion.

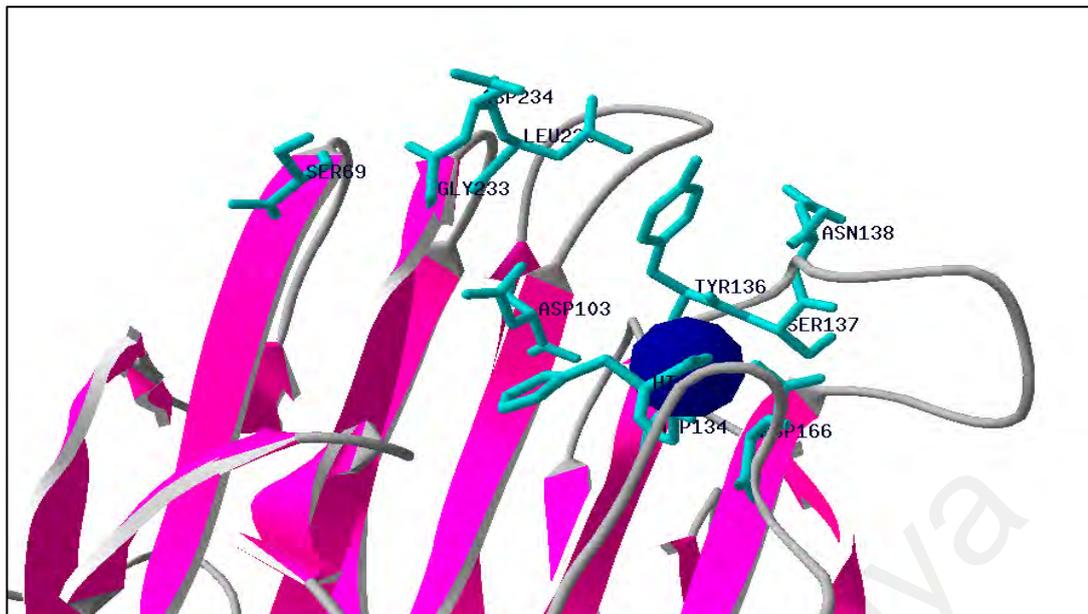


Figure 4.3: Zoom in view of *MrLTL* protein's carbohydrate-binding motifs. The metal binding site was found near to and coordinated with the YSN motif. Carbohydrate binding motifs: Y¹³⁵-S¹³⁶-N¹³⁷ and G²³²-D²³³-L²³⁴; carbohydrate recognition amino acids: S⁶⁸, D¹⁰², and H¹⁶²; metal binding amino acids: D¹³³ and D¹⁶⁵; are highlighted in backbone structure and coloured in cyan. The blue sphere is calcium ion.

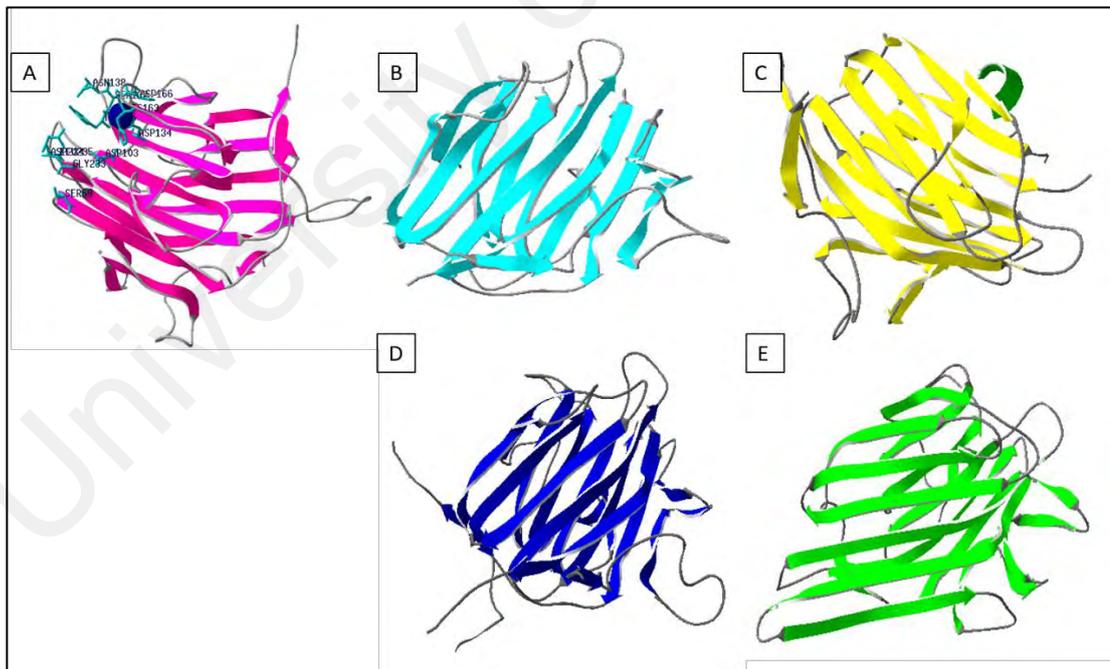


Figure 4.4: Structural comparison of L-type lectins of various origin. A: *MrLTL* predicted model; B: *Homo sapiens* ERGIC-53 protein (PDB: 3WHT); C: *Rattus norvegicus* ERGIC-53 protein (PDB: 1GV9); D: *Canis lupus* VIP36 protein (PDB: 2DUP); E: *Canavalia ensiformis* concanavalin A (PDB: 1NLS).

4.1.4 *In silico* simulation of binding of ligands on *MrLTLTD*.

Ten saccharide molecules (Table 4.2) were used to simulate the binding of ligand with *MrLTLTD* via Autodock 4.2.6. These carbohydrates were selected as they are ubiquitous and serve as the building block for many glycan in biological system. Autodock 4.2.6 applies algorithm to access the lowest binding energy, which also means the most stable interaction; and potential H-bond forms between the examined ligand and receptor (target protein). Among the runs, N-acetyl-D-glucosamine showed the weakest binding energy. Mannose-type compounds averagely showed stronger binding energy compared to other compounds. As many reports have experimentally demonstrated the preferentially binding of animal's L-type lectin towards high mannose-type glycan, a similar prediction was generated by the software. Figure 4.5 illustrates the fitting of a compound, alpha-D-Man-(1,3)-alpha-D-Man-(1,2)-alpha-D-Mannose, into the reaction pocket flanked by both YSN and GDL motifs. This compound was detected to form 8 H-bonds, the highest among tested compounds, with the carbohydrate binding motif with the lowest mean binding energy of – 4.73.

Table 4.2: Autodock binding simulation results of carbohydrate ligands with *MrLTLTD*.

Ligand	PubChem CID	Number in Cluster	Mean Lowest Binding Energy	Number of H-bond
Alpha-D-Glucose	79025	78	-3.82	5
Beta-D-Glucose	64689	65	-3.63	3
D-Mannose	18950	78	-3.45	5
N-acetyl-beta-D-mannosamine	11096158	18	-4.28	3
N-acetyl-D-mannosamine	439281	26	-4.35	6
N-acetyl-D-galactosamine	35717	14	-3.64	4
N-acetyl-D-glucosamine	439174	17	-2.82	4
N-acetylmuramic acid	5462244	13	-3.47	5
Alpha-1,2-Mannobiose	11099946	3	-4.17	7
Alpha-D-Man-(1,3)-alpha-D-Man-(1,2)-alpha-D-Mannose	56927890	69	-4.73	8

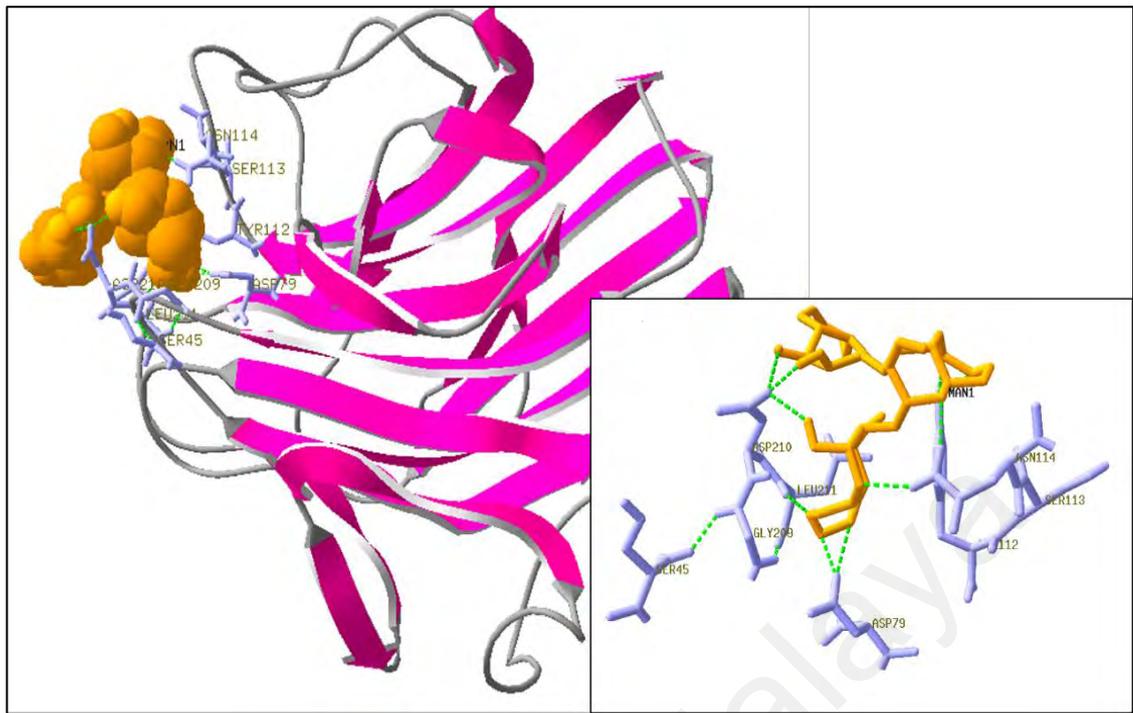


Figure 4.5: Illustration of ligand's interaction with carbohydrate-binding motif of *MrLTL*. Orange CPK model: Alpha-D-Man-(1,3)-alpha-D-Man-(1,2)-alpha-D-Mannose; Silver ball-and-stick model: carbohydrate-binding motif. Right bottom image: Schematic ball-and-stick model drawings of alpha-D-Man-(1,3)-alpha-D-Man-(1,2)-alpha-D-Mannose (orange) bonds with amino acid residuals of the carbohydrate-binding motif (silver) through H-bond (green dot-line).

4.2 Amplification of the nucleotide sequence of *MrLTL*.

Primers were designed (Table 3.1) to amplify the gene sequence from the cDNA sample extracted from *M. rosenbergii* hepatopancreas tissues. An expected amplicon size of 1247 bp was amplified (Figure 4.6). The 1247 bp amplicon was used as the template for the following PCR to produce only the complete coding sequence of *MrLTL* (995 bp) and domain coding sequence *MrLTL*D (698 bp) (Figure 4.7). Restriction endonuclease recognition sites (*Pci*I & *Xho*I) were incorporated into both *MrLTL* and *MrLTL*D during the PCR.

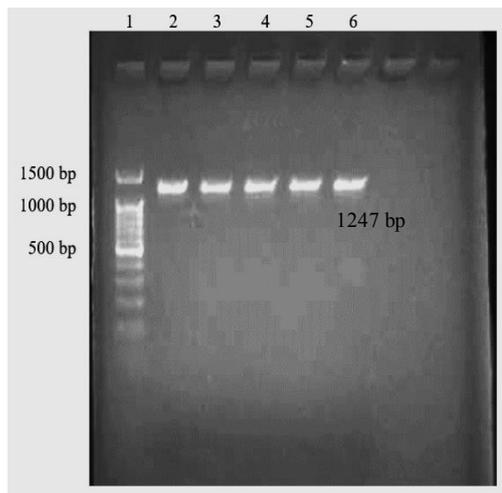


Figure 4.6: Gel image of amplified *MrLTL* sequence with expected amplicon size of 1247 bp. Lane 1: DNA ladder; lanes 2-6: *MrLTL* amplicon replicates.

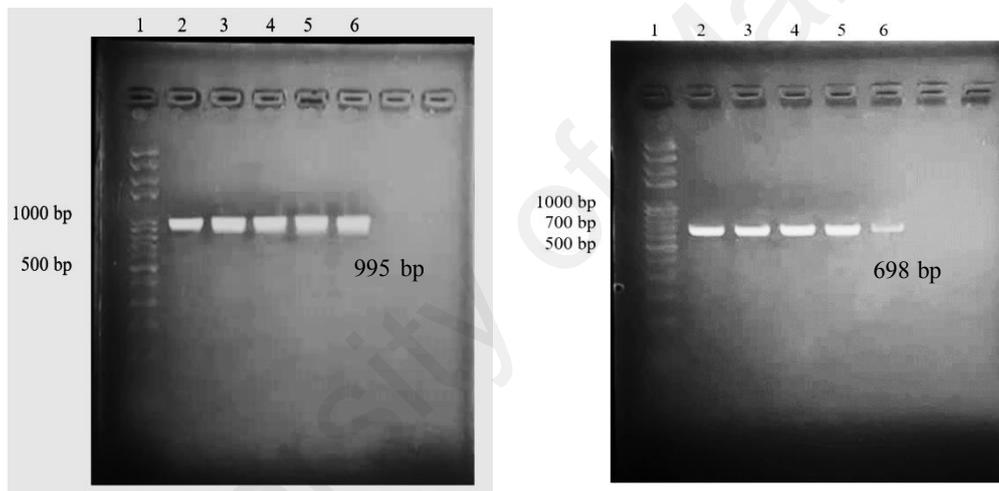


Figure 4.7: PCR products of *MrLTL* and *MrLTLD* containing restriction endonuclease recognition sites (*PciI* & *XhoI*). Left image: lane 1: DNA ladder; lanes 2-6: *MrLTL* PCR products with expected size of 995 bp. Right image: lane 1: DNA ladder; lanes 2-6: *MrLTLD* PCR products with expected size of 698 bp.

4.3 Cloning of *MrLTL* and *MrLTLD* sequences into pET-30(a) expression vector.

Both the *MrLTL* and *MrLTLD* coding sequences were cloned into *NcoI* and *XhoI* MCS site of pET-30(a) expression vector (Figure 4.9). The pET-30(a) vector introduces a fusion His6-tag and S-tag to the inserted gene at N-terminal end (Figure 4.9). The recombinant plasmid was then introduced into *E. coli* BL21(DE3). Recombinant plasmid

extracted from the transformed colonies were selected by PCR amplification for nucleotide sequencing to determine the orientation of the insert in the plasmid and no mismatch nucleotide was introduced (Figure 4.8). The positive clone with correct in-frame insertion was then used in recombinant protein expression. The expressed recombinant proteins are expected to adopt an architect as illustrated in Figure 4.10.

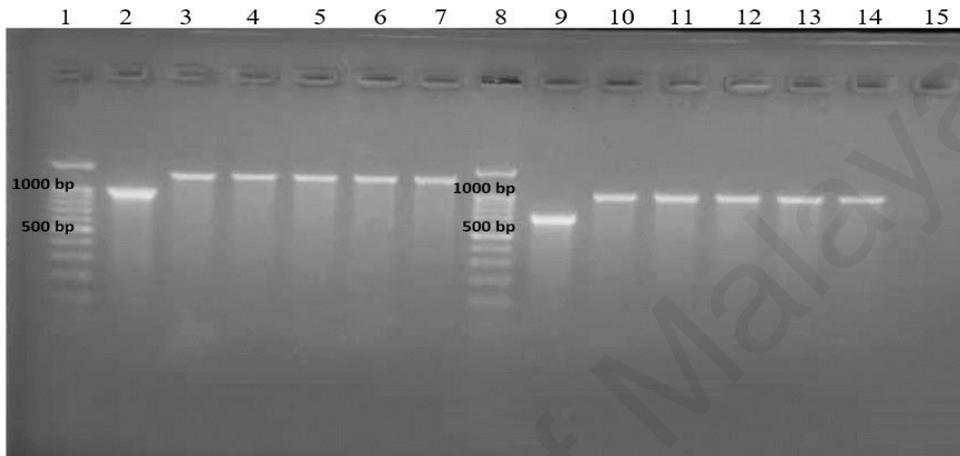


Figure 4.8: Colony PCR of *MrLTL* and *MrLTL*D *E. coli* BL21(DE3) transformants. Lanes 1 and 8: DNA ladder. Lane 2: *MrLTL* coding sequence amplified with *MrLTL*_F & R primers (995 bp); lanes 3-7: *MrLTL* positive clones amplified with T7ProF & T7TerR primers (\approx 1300 bp); lane 9: *MrLTL*D coding sequence amplified with *MrLTL*D_F & R primers (698 bp); lanes 10-15: *MrLTL*D positive clones amplified with T7ProF & T7TerR primers (\approx 1000 bp).



Figure 4.9: The pET-30(a) expression vector construct mapping. Red box highlighted the MCS region. T7 promoter sequence and ribosome binding site (rbs) greatly enhance the translational efficiency of cloned recombinant sequence. *Lac* operator allows controlled expression through IPTG. Both His-tag and S-tag permit convenient detection of the expressed recombinant protein.

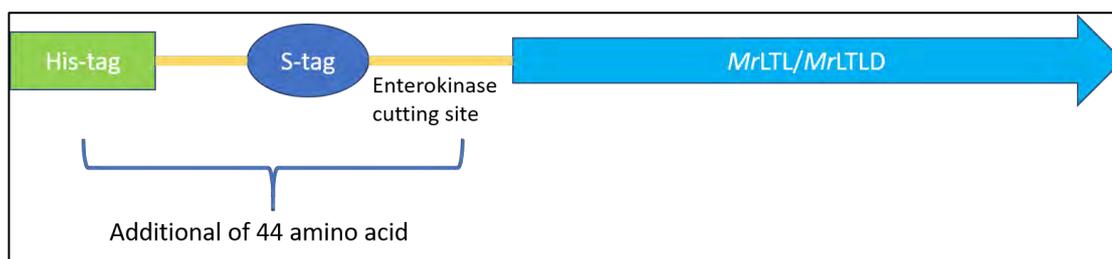


Figure 4.10: Schematic illustration of recombinant *MrLTL* and *MrLTL D* peptide sequences expressed via pET-30(a) vector. There are two fusion tags, His₆ and S-tag precede the inserted recombinant peptide at the N-terminal end. Both tags comprise 44 amino acids and can be removed by enzymatic cutting at the enterokinase site.

4.4 Expression of recombinant proteins *MrLTL* and *MrLTL D*.

Both *MrLTL* and *MrLTL D* clones were induced under the stipulated condition in (Section 3.9). After the induction, cells were harvested and proceeded with lysis protocol. The lysed cells were separated into insoluble (pellet) and soluble (supernatant) fractions. SDS-PAGE was then performed to separate the expressed recombinant proteins. In addition, the expressed recombinant proteins were identified by western blotting using His-tag antibody conjugated to HRP.

Figure 4.11 shows that *MrLTL* protein was not expressed in different IPTG concentrations added. Bands with estimated M_r of 40 kDa was observed on the SDS-PAGE gel but later western blotting proved negative of the presence of this protein. Besides, *MrLTL* protein was also not detected in the SDS-PAGE gel images of different induction temperatures (Figure 4.13) and with the addition of additives into the medium (Figure 4.14).

In contrast, *MrLTL D* domain region of the protein was successfully expressed in all the induction settings. A distinct band with M_r of 30 kDa was observed on SDS-PAGE gel. The authenticity of the band was further confirmed by western blotting using His-tag antibody (Figure 4.12). All expressed *MrLTL D* proteins were found in the pellet fraction instead of supernatant fraction, which indicates that the protein was in insoluble form.

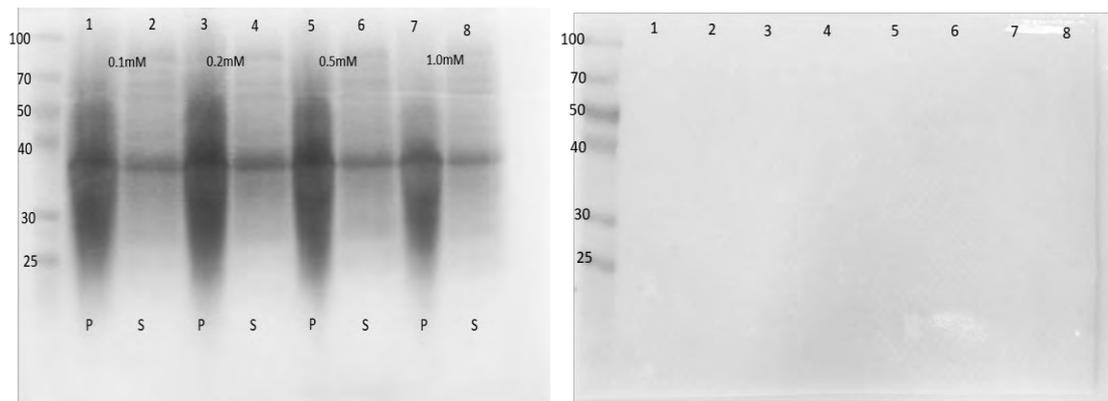


Figure 4.11: SDS-PAGE image of *MrLTL* recombinant protein induced with different concentrations of IPTG. *MrLTL* recombinant protein has an estimated molecular mass of 41 kDa. Picture on the right is a complementary western blotting image for recombinant protein detection. The induction was performed at 37°C for 4 hr with agitation of 180 rpm. Leftmost lane: Protein's ladder; lane 1: 0.1 mM IPTG, pellet; lane2: 0.1 mM IPTG, supernatant; lane 3: 0.2 mM IPTG, pellet; lane 4: 0.2 mM IPTG, supernatant; lane 5: 0.5 mM IPTG, pellet; lane 6: 0.5 mM IPTG, supernatant; lane 7: 1.0 mM IPTG, pellet; lane 8: 1.0 mM IPTG, supernatant.

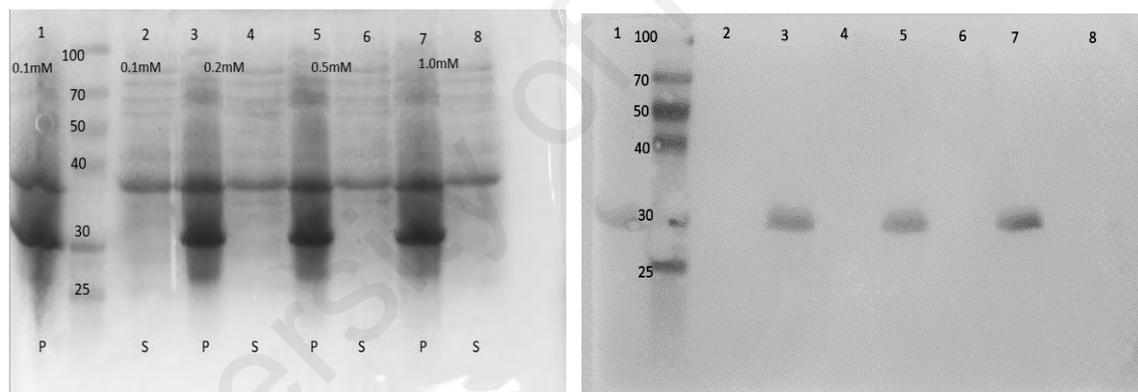


Figure 4.12: SDS-PAGE image of *MrLTLD* recombinant protein induced with different concentrations of IPTG. *MrLTLD* recombinant protein has an estimated molecular mass of 30 kDa. Picture on the right is a complementary western blotting image for recombinant protein detection. The induction was performed at 37°C for 4 hr with agitation of 180 rpm. Lane 1: 0.1 mM IPTG, pellet; lane protein's ladder; lane 2: 0.1 mM IPTG, supernatant; lane 3: 0.2 mM IPTG, pellet; lane 4: 0.2 mM IPTG, supernatant; lane 5: 0.5 mM IPTG, pellet; lane 6: 0.5 mM IPTG, supernatant; lane 7: 1.0 mM IPTG, pellet; lane 8: 1.0 mM IPTG, supernatant.

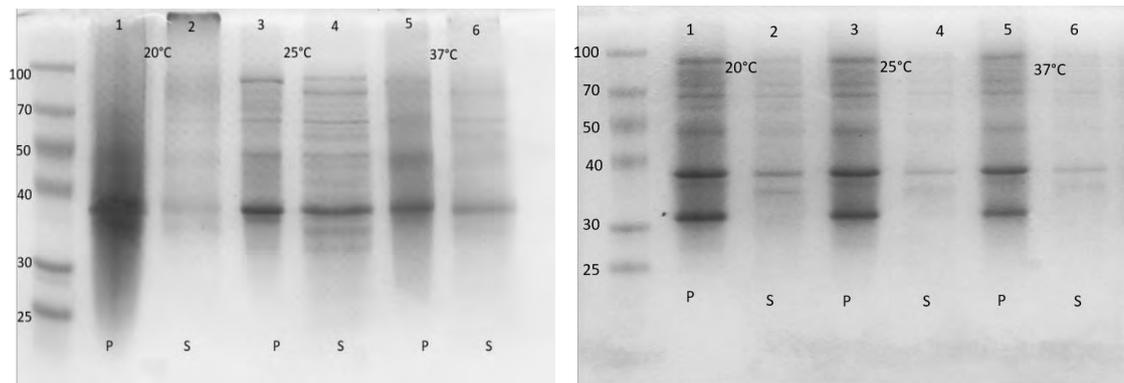


Figure 4.13: SDS-PAGE images of *MrLTL* (left) & *MrLTLTD* (right) recombinant proteins induced in various induction temperatures. The induction temperature and induction duration were 20 °C (overnight), 25 °C (overnight), and 37 °C (4 hr) with agitation 180 rpm. Leftmost lane of both images: Protein marker; P: insoluble fraction; S: soluble fraction.

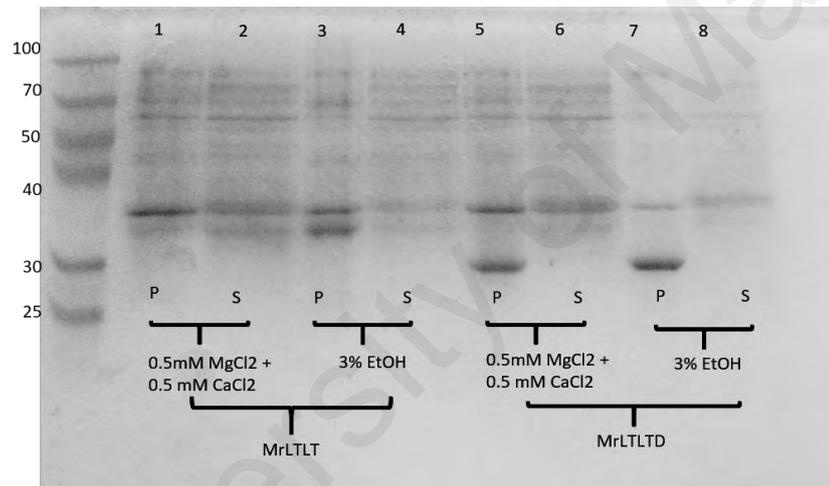


Figure 4.14: SDS-PAGE image of *MrLTL* & *MrLTLTD* recombinant protein induced expression with additives (A) 0.5 mM MgCl₂ + 0.5 mM CaCl₂ and (B) 3% ethanol. Protein induction was performed at 25 °C for overnight with agitation of 180 rpm. Leftmost lane: Protein markers; lanes 1 & 2: *MrLTL* recombinant protein induced with 0.5 mM MgCl₂ + 0.5 mM CaCl₂; lanes 3 & 4: *MrLTL* recombinant protein induced with 3% ethanol; lanes 5 & 6: *MrLTLTD* recombinant protein induced with 0.5 mM MgCl₂ + 0.5 mM CaCl₂; lanes 7 & 8: *MrLTLTD* recombinant protein induced with 3% ethanol. P: insoluble fraction; S: soluble fraction.

CHAPTER 5: DISCUSSION

Unlike plant lectins, animal L-type lectins are not secreted but reside at the membrane organelles within a cell (Gupta et al., 2012). The prediction of a transmembrane sequence present at the c-terminal end of the *MrLTL* protein indicates that this protein is also membrane-anchored, as observed from other animals' homologues. So far, the exact cellular localisation of *MrLTL* is unknown and animal LTLs were found to be compartmentalised at endoplasmic reticulum and Golgi apparatus based only on mammalian vertebrate animal experiments (Kamiya et al., 2008). In the view of substantial sequence and structural similarities, we can deduce that *MrLTL* may adopt a similar cellular localisation and engage in early secretory pathway as their vertebrate counterparts. On the other hand, multiple studies have suggested that crustacean's LTLs function as PRR in the host immune responses by recognising the carbohydrate moieties presented on the pathogen surface (Huang et al., 2014; Xiu et al., 2015; Xu et al., 2014). This require easy access of the LTL with the selective ligands, in which cell plasma membrane is the most possible destination of the LTL protein. Otherwise, a mechanism must exist to transport those ligands into the cell as if intracellular membrane organelles are the major reservoir for crustacean's LTL. Xu and colleagues (2014) reported an interesting observation in which LTL was trimmed and released extracellular as humoral opsonin instead of confined within a cell. In either case, an immunohistochemistry assay will be the method of choice to discover the cellular localisation of *MrLTL* and provides insight into its respective mechanism of action.

The structure of L-type lectin domain (LTL_D) was first described in Concanavalin A protein with a beta-sandwich conformation. The homology modelling showed that *MrLTL* protein indeed adopts a beta-sandwich fold (Figure 4.2). However, the structures of both signal peptide and transmembrane region were not built in the homology

modelling. This is due to difficulty in extraction of native transmembrane region in a heterologous host, and hence lack of available reference template in databases. Additionally, a comparison of LTL proteins among different organisms (Figure 4.4) revealed a striking similarity of structural conformation, with differences in the number of beta-strands makeup. This indicates that the beta-sandwich structure is well conserved and might trace the evolutionary origin of this protein back to plants.

LTLD is devoid of any enzymatic activity with the presence of carbohydrate-binding motif to exert carbohydrate affinity. Two carbohydrate-binding triads, YSN and GDL, were identified in *MrLTL* protein. The separated motifs may predict multivalence of *MrLTL*, where multiple ligands can react to these motifs independently. Interestingly, the homology modelling result suggests a different condition. Both YSN and GDL motifs were brought into proximity by structural folding, with other carbohydrate-reactive residuals, S⁶⁸, D¹⁰², and H¹⁶², located around (Figure 4.3). This centralised arrangement suggests that, instead of function separately, these reactive motifs and residuals work synergistically in recognising and binding towards a particular ligand one at a time. The accessibility of carbohydrate-binding motifs towards ligands is further enhanced since they are spanned on the apex of the *MrLTL* protein. Moreover, the space created between both motifs is wide, hence permits larger entry for ligands; and explains the ability of LTL to react with complex carbohydrate macromolecules such as lipopolysaccharide, peptidoglycan, and high mannose-type residuals (Huang et al., 2014). A Ca²⁺ ion interacted with D¹³³ and D¹⁶⁵ residuals and coordinated with the tyrosine (Y¹³⁵) and asparagine (N¹³⁷) residuals of the YSN motif. Indeed, the Ca²⁺ positions and stabilises the proper orientation of the carbohydrate-binding motif and hence is critical for LTL activity (Sato et al., 2007).

Gene of interest can be introduced into a vector either by TA cloning or directional cloning. The former is a rather simpler protocol but suffers from low recombinant rate due to self-ligation and disorientation of the insert. In this experiment, *MrLTL* was cloned into the *NcoI* and *XhoI* restriction endonuclease (RE) sites of vector pET-30(a) through directional cloning. Directional cloning utilises the complementary overhang produced by restriction endonucleases to eliminate self-ligation of vector and ensures proper orientation of the cloned insert. Respective RE sites were incorporated to *MrLTL* sequence via primer design. Moreover, in order to enhance RE accuracy and activity, a nucleotide oligomer precedes the RE site was also introduced (Novagen, 2003). However, *NcoI* RE site was found in the sequence of *MrLTL*. To avoid disruption of the nucleotide sequence, *PciI* restriction endonuclease was selected to replace *NcoI* in the digestion of the *MrLTL* nucleotide sequence. Although both restriction endonucleases recognise different RE sites, they produce compatible overhangs which make complementary pairing of both the insert and the vector possible.

A number of factors affect the overall outcome of the expressed recombinant protein. Overexpression of the recombinant protein in soluble form is often the goal. Overexpression applies the maximum product to cellular mass ratio while soluble protein is essential for structural crystallisation and protein's activity study. The full-length *MrLTL* protein was unable to be expressed in different induction conditions, including different IPTG concentrations (Figure 4.11), different induction temperatures (Figure 4.13) and additives in the media (Figure 4.14). The disorientation of insert sequence cloned into the plasmid or absence of the recombinant plasmid in the transformed bacterial colony is ruled out for the failure in expressing the recombinant full-length *MrLTL* protein, due to the fact that all transformed bacterial colonies have their plasmid nucleotide sequences screened prior protein expression. Alternatively, the removal of signal peptide permits the expression of the domain region *MrLTL*LD to be expressed

readily (Figure 4.12). It seems that, by unknown reason, the signal peptide sequence is either impeding the translation of this protein or labelling the nascent translated peptide as a prominent target for degradation, and thus no detectable level of *MrLTL* protein was observed.

The expression of domain region *MrLTLD*, instead of the complete sequence, is sufficient for structural and functional studies of this prawn's lectin. *MrLTLD* is readily expressed in standard incubation condition with its authenticity assured by anti-His tag antibody in western blotting (Figure 4.12). Prior to the protein expression, *MrLTLD* was analysed to possess a negative GRAVY value, which implies an intrinsic hydrophilic nature (Table 4.1). Nonetheless, *MrLTLD* was found only in pellet fraction of the cell as insoluble aggregates, which is undesired in this study. Formation of inclusion bodies or insoluble aggregates is a prevailing challenge faced by recombinant protein expression. Many communications were accentuated on the resolution to circumvent this problem (Sorensen & Mortensen, 2005). A series of attempts were performed in this project to enhance the solubility of the expressed *MrLTLD*. The optimisation of IPTG concentration (Figure 4.12) is meant to reduce the inducer potency and determine the minimal IPTG concentration required for consequent expression assay, and 0.1 mM was determined to work as effective as 1.0 mM IPTG. Secondly, *MrLTLD* was incubated at sub-optimal temperature to decrease host's translational rate, attenuate the burden of overexpression of heterologous recombinant protein, and prescribe extended time for proper protein folding. The expression of recombinant *MrLTLD* was slowed in which the bacterial clones incubated at 20 °C and 25 °C took overnight to demonstrate a similar thickness of *MrLTLD* protein band on SDS-PAGE compared with those cultured at 37 °C for four hours (Figure 4.13). Addition of osmolytes and cofactor has been reported to enhance solubility of recombinant protein expression (Blackwell & Horgan, 1991; Lebendiker & Danieli, 2014; Oganessian et al., 2007; Sorensen & Mortensen, 2005). The former acts as

a chemical chaperone by creating a favourable environment (Rabbani & Choi, 2017) while the latter participates in the folding process and stabilises the native conformation for certain proteins (Rosano & Ceccarelli, 2014). Cations, Mg^{2+} and Ca^{2+} , were supplied to *MrLTL*D expression as Ca^{2+} was found important for L-type lectin's motif orientation and activity, and 3% ethanol was reported to mimic an osmotic stressed environment for the host (Figure 4.14) (Chhetri et al., 2015). Despite all the manipulations, no detectable level of soluble *MrLTL*D was expressed.

There are a few suggestions the author would like to offer to enhance the solubility of the recombinant protein *MrLTL*D. Different strains of expression hosts are available for exploitation. These include C41/C43(DE3) and *trxB*-negative *E. coli* strains. The former accommodates the aggregation-prone transmembrane protein with proliferated intracellular membrane, while the latter permits disulphide bond formation in cytosol (Sorensen & Mortensen, 2005). Fusion tags are potent components to increase the chances of soluble recombinant protein expression. To name a few, maltose-binding protein (MBP), NusA, TrxA, and small ubiquitin-related modifier (SUMO) are gaining popularity due to their accumulated success as solubility enhancers (Lebendiker & Danieli, 2014). The use of fusion tags requires prudent consideration. Some fusion tags are relatively larger in size, such as MBP, which will increase the fusion protein's molecular size drastically and render difficult synthesis; unpredicted interaction between fusion tag and protein; and tags removal options need not to be exhaustive. Alternatively, the insoluble *MrLTL*D will need to be resorted to denaturation and refolding process which inevitably reduce final yield of active protein and incur additional cost.

CHAPTER 6: CONCLUSION

In conclusion, a *Macrobrachium rosenbergii*'s L-type lectin gene was characterised using bioinformatics approach. An open reading frame containing 972 nucleotides was determined which codes for a putative protein of 323 amino acids. The protein has a molecular mass of 36 kDa. It comprises a short signal peptide, a L-type lectin domain and a small transmembrane region. The structure of *MrLTL* was built through homology modelling with canine VIP36 protein as reference template. The protein adopts a beta-sandwich conformation, which is well-conserved across different organisms. Two carbohydrate-binding motifs, YSN and GDL, and cation Ca^{2+} metal binding amino acid residuals were identified. Based on structural prediction and ligand interaction simulation tools, these motifs were proposed to function synergistically and react preferentially to mannose-type molecules. The domain region, *MrLTLD*, was readily expressed in standard condition as a recombinant protein using *E. coli* BL21(DE3) strain and vector pET-30(a). The protein was produced in insoluble form. Various approaches were developed to enhance the solubility of the expressed recombinant *MrLTLD*, which include minimal IPTG concentration (0.1 mM), reduced incubation temperature at 25°C, and cations and ethanol additives into the culture medium. However, no apparent improvement was noticed.

This project has proposed a few suggestions to further enhance the solubility for recombinant *MrLTLD* expression. The production of soluble *MrLTLD* will contribute to structural and functional studies of this protein, which in turn help to answer inquiries such as pathogen-host recognition, ligand-motif interaction, and immune pathway mechanism. An intriguing prospect would be delving into the protein-protein interaction of this lectin, hence provides us a glimpse into the early triggering pathway of a complex invertebrate's innate immune response, with shrimp as the model.

REFERENCES

- Banu, R., & Christianus, A. (2016). Giant freshwater prawn *Macrobrachium rosenbergii* farming: A review on its current status and prospective in Malaysia. *Journal of Aquaculture Research & Development*, 7(4), 1-5.
- Berman, H. M., Westbrook, J., Feng, Z., Gillilan, G., Bhat, T. N., Weissig, H., ... Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, 28, 235-242.
- Blackwell, J. R. & Horgan, R. (1991). A novel strategy for production of a highly expressed recombinant protein in an active form. *Federation of European Biochemical Societies*, 295(1,2,3), 10-12.
- Bonami, J. R., and Widada, J. S. (2014). Viral disease of giant freshwater prawn *Macrobrachium rosenbergii*: A review. *Journal of Invertebrate Pathology*, 106, 131-142.
- Braak, C. B. T., Botterblom, M. H. A., Knaap, W. P. W., & Rombout, J. H. W. M. (2000). Characterisation of different morphological features of black tiger shrimp (*Penaeus monodon*) haemocytes using monoclonal antibodies. *Fish & Shellfish Immunology*, 10, 515-530.
- Cerenius, L., Jiravanichpaisal, P., Liu, H. P., & Soderhall, I. (2010). Crustacean immunity. *Advances in Experimental Medicine and Biology*, 708, 239-259.
- Chen, Y. H., Li, X. Y., & He, J. G. (2014). Recent advances in researches on shrimp immune pathway involved in white spot syndrome virus genes regulation. *Journal of Aquaculture Research & Development*, 5(3), 1-13.
- Chhetri, G., Kalita, P., & Tripathi, T. (2015). An efficient protocol to enhance recombinant protein expression using ethanol in *Escherichia coli*. *MethodsX*, 2, 385-391.
- Dai, Y. J., Wang, Y. Q., Zhao, L. L., Qin, Z. D., Yuan, J. F., Qin, Q. W., ... Lan, J. F. (2016). A novel l-type lectin was required for the multiplication of WSSV in red swamp crayfish (*Procambarus clarkii*). *Fish & Shellfish Immunology*, 55, 48-55.

- De Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., ... Hulo, N. (2006). ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research*, 1(34), 362-365.
- De Grave, S., Shy, J., Wowor, D. & Page, T. (2013). *Macrobrachium rosenbergii*. The IUCN red list of threatened species 2013: e. T1978A2503520. Retrieved on November 16, 2018 from <http://www.iucnredlist.org/details/full/197873/0>
- Deacon, A., Gleichmann, T., Kalb, A. J., Price, H., Raftery, J., Bradbrook, G., ... Helliwell, J. R. (1997). The structure of Concanavalin A and its bound solvent determined with small-molecule accuracy at 0.94 Å resolution. *Journal of the Chemical Society, Faraday Transactions*, 93, 4305-4312.
- Demain, A. L., & Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances*, 27(3), 297-306.
- Department of Fisheries Malaysia. (2017). *2016 Annual fisheries statistics*. Retrieved on July 23, 2018 from <https://www.dof.gov.my/index.php/pages/view/3049>
- Drickamer, K. (2014, Jan). *A genomic resource for animal lectins*. Retrieved on July 23, 2018 from <http://www.imperial.ac.uk/research/animallecins/default.html>
- Feng, J. L., Huang, X., Zhang, Y., Li, T. T., Hui, K. M., & Ren, Q. (2016). A C-type lectin (*MrLec*) with high expression in intestine is involved in innate immune response of *Macrobrachium rosenbergii*. *Fish & Shellfish Immunology*, 59, 345-350.
- Flegel, T. W., Lightner, D. V., Lo, C. F. & Owens, L. (2008). Shrimp disease control: Past, present, and future. In M. G. Bondad-Reantaso, C. V. Mohan, M. Crumlish & R. P. Subasinhge (Eds)., *Diseases in Asian aquaculture VI* (pp. 355-378). Fish Health Section, Asian Fisheries Society, Manila: Philippines.
- Fu, L. L., Zhou, C. C., Yao, S., Yu, J. Y., Liu, B., & Bao, J. K. (2011). Plant lectins: Targeting programmed cell death pathways as antitumor agents. *The International Journal of Biochemistry & Cell Biology*, 43, 1442-1449.

- Giulianini, P. G., Bierti, M., Lorenzon, S., Battistella, S., & Ferrero, E. A. (2006). Ultrastructural and functional characterization of circulating hemocytes from the freshwater crayfish *Astacus leptodactylus*: Cell types and their role after *in vivo* artificial non-self-challenge. *Micron*, 38, 49-57.
- Hazreen, N. M. K., Kua, B. C., Bhassu, S., & Othman, R. Y. (2012). Detection and genetic profiling of infectious hypodermal and haematopoietic necrosis virus (IHHNV) infections in wild berried freshwater prawn, *Macrobrachium rosenbergii* collected for hatchery production. *Molecular Biology Reports*, 39(4), 3785-3790.
- Hien, N. T., Huong, N. T. L., Chuong, V. D., Nga, N. T. V., Quang, P. H., Hang, B. T. V., & Long, N. V. (2016). Status of acute hepatopancreatic necrosis disease (AHPND) and other emerging diseases of penaeid shrimps in Viet Nam. In R. V. Pakingking Jr., E. G. T. de Jesus-Ayson, & B. O. Acosta (Eds.), *Addressing Acute Hepatopancreatic Necrosis Disease (AHPND) and other transboundary diseases for improved aquatic animal health in Southeast Asia: Proceedings of the ASEAN regional technical consultation on EMS/AHPND and other transboundary diseases for improved aquatic animal health in Southeast Asia* (pp. 88-95). Philippines: Aquaculture Department, Southeast Asian Fisheries Development Centre.
- Hoa, T. T. T., Hoang, D. T. H., & Phuong, N. T. (2014). *Study on diseases in giant freshwater prawns (Macrobrachium rosenbergii): A review*. Retrieved on August 10, 2018 from https://www.researchgate.net/publication/266448376_Study_on_Diseases_in_Giant_Freshwater_Prawns_Macrobrachium_rosenbergii_A_Review
- Hsieh, C. Y., Chuang, P. C., Chen, L. C., Tu, C., Chien, M. S., Huang, K. C., ... Tsai, S. S. (2006). Infectious hypodermal and haematopoietic necrosis virus (IHHNV) infections in giant freshwater prawn, *Macrobrachium rosenbergii*. *Aquaculture*, 258(1-4), 73-79.
- Huang, Y., Tan, M. J., Wang, Z., Yin, S. W., Huang, X., Wang, W., & Ren, Q. (2014). Cloning and characterization of two different L-type lectin genes from the Chinese mitten crab *Eriocheir sinensis*. *Development and Comparative Immunology*, 46, 255-266.
- Jin, X. K., Li, S., Guo, X. V., Cheng, L., Wu, M. H., Tan, S. J., ... Wang, Q. (2013). Two antibacterial C-type lectins from crustacean, *Eriocheir sinensis*, stimulated cellular encapsulation *in vitro*. *Development and Comparative Immunology*, 41, 544-552.

- Johansson, M. W., Keyser, P., Sritunyalucksana, K., & Soderhall, K. (2000). Crustacean haemocytes and haematopoiesis. *Aquaculture*, 191, 45-52.
- Kamiya, Y., Kamiya, D., Yamamoto, K., Nyfeler, B., Hauri, H., & Kato, K. (2008). Molecular basis of sugar recognition by the human L-type lectins ERGIC-3, VIPL, and VIP36. *The Journal of Biological Chemistry*, 283(4), 1857-1861.
- Lagarda-Diaz, I., Guzman-Partida, A. M., & Vazquez-Moreno, L. (2017). Legume lectins: Proteins with diverse application. *International Journal of Molecular Sciences*, 18, 1242-1250.
- Lai, X. F., Kong, J., Wang, Q. Y., Wang, W. J., & Meng, X. H. (2013). Identification and molecular characterization of a C-type lectin-like protein from Chinese shrimp (*Fenneropenaeus chinensis*). *Molecular Biology Reports*, 40, 2223-2230.
- Lebendiker, M. & Danieli, T. (2014). Production of prone-to-aggregate proteins. *Federation of European Biochemical Societies Letters*, 588, 236-246.
- Li, F. H. & Xiang, J. H. (2013). Recent advances in researches on innate immunity of shrimp in China. *Development and Comparative Immunology*, 39, 11-26.
- Liu, Z. Y., Li, X. F., Ding, X. P., & Yang, Y. (2010). *In silico* and experimental studies of Concanavalin A: Insights into its antiproliferative activity and apoptotic mechanism. *Applied Biochemistry and Biotechnology*, 162, 134-145.
- Loris, R., Hamelryck, T., Bouckaert, J., & Wyns, L. (1998). Legume lectin structure. *Biochimica et Biophysica Acta*, 1383, 9-36.
- Martin, G. G., & Graves, B. L. (1985). Fine structure and classification of shrimp hemocytes. *Journal of Morphology*, 185, 339-348.
- Mastan, S. A. (2015). Use of immunostimulants in aquaculture disease management. *International Journal of Fisheries and Aquatic Studies*, 2(4), 277-280.

- New, N. B. (2002). Farming freshwater prawns. A manual for culture of giant river prawn (*Macbrachium rosenbergii*). *FAO Fisheries Technical Paper*, 428, 1-26.
- New, N. B. (2005). Freshwater prawn farming: Global status, recent research and a glance at the future. *Aquaculture Research*, 36, 210-230.
- Novagen. (2003). *pET system manual*. Retrieved on May 11, 2018 from <http://lifeserv.bgu.ac.il/wp/zarivach/wp-content/uploads/2017/11/Novagen-pET-system-manual-1.pdf>.
- Oganesyan, N., Ankoudinova, I., Sung, H. K., & Kim, R. (2007). Effect of osmotic stress and heat shock in recombinant protein overexpression and crystallization. *Protein Expression and Purification*, 52(2), 280-285.
- Petersen, T. N., Brunak, S., Heijne, V. G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Method*, 8, 785-786.
- Rabbani, G. & Choi, I. (2017). Roles of osmolytes in protein folding and aggregation in cells and their biotechnological applications. *International Journal of Biological Macromolecules*, 109, 483-491.
- Rodriguez, J., Boulo, V., Mialhe, E., & Bachere, E. (1995). Characterisation of shrimp haemocytes and plasma components by monoclonal antibodies. *Journal of Cell Science*, 108, 1043-1050.
- Rosano, G.L., & Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Frontiers in Microbiology*, 5, 1-17.
- Sambrook, J. & Russell, D.W. (2001). *Molecular cloning a laboratory manual* (3rd ed.). New York, NY: Cold Spring Harbor Laboratory Press.
- Satoh, T., Cowieson, N. P., Hakamata, W., Ideo, H., Fukushima, K., Kurihara, M., ... Wakatsuki, S. (2007). Structural basis for recognition of high mannose type glycoproteins by mammalian transport lectin VIP36. *The Journal of Biological Chemistry*, 282, 28246-55.

- Satoh, T., Suzuki, K., Yamaguchi, T., & Kato, K. (2014). Structural basis for disparate sugar-binding specificities in the homologous cargo receptors ERGIC-53 and VIP 36. *Plos ONE*, 9 (2), 1-7.
- Saurabh, S., & Sahoo, P. K. (2008). Major diseases and the defence mechanism in giant freshwater prawn, *Macrobrachium rosenbergii* (de Man). *Proceedings of the National Academy of Sciences*, 78, 103-121.
- Sharon, N., & Lis, H. (2002). How proteins bind carbohydrates: Lessons from legume lectins. *Journal of Agricultural and Food Chemistry*, 50, 6586-6591.
- Sharon, N., & Lis, H. (2004). History of lectins: From hemagglutinins to biological recognition molecules. *Glycobiology*, 14(11), 53-62.
- Smith, V. J., & Soderhall, K. (1983a). β -1,3 glucan activation of crustacean hemocytes *in vitro* and *in vivo*. *The Biological Bulletin*, 164, 299-314.
- Smith, V. J., & Soderhall, K. (1983b). Induction of degranulation and lysis of haemocytes in freshwater crayfish, *Astacus astacus* by components of the prophenoloxidase activating system *in vitro*. *Cell and Tissue Research*, 233, 295-303.
- Sorensen, H. P. & Mortensen, K. K. (2005). Soluble expression of recombinant proteins in cytoplasm of *Escherichia coli*. *Microbial cell Factories*, 4(1), 1-8.
- Sun, Y. D., Fu, L. D., Jia, Y. P., Du, X. J., Wang, Q., Wang, Y. H., ... Wang, J. X. (2008). A hepatopancreas-specific C-type lectin from the Chinese shrimp *Fenneropenaeus chinensis* exhibits antimicrobial activity. *Molecular Immunology*, 45, 348-361.
- Supamattaya, K., Ruangsri, J., Itami, T., Chitiwan, V., Phromkunthong, W., & Muroga, K. (2002). Morphology and immunological roles of hemocytes and fixed phagocytes in black tiger shrimp *Penaeus monodon*. *Fish Pathology*, 38(2), 33-39.
- Tiruvayipati, S., & Bhassu, S. (2016). Host, pathogen and the environment: The case of *Macrobrachium rosenbergii*, *Vibrio parahaemolyticus* and magnesium. *Gut Pathogens*, 8(15), 1-8.

- Vasta, G. R., Ahmed, H., Tasumi, S., Odom, E., & Saito, K. (2007). Biological roles of lectins in innate immunity: Molecular and structural basis for diversity in self/non-self recognition. *Advances in Experimental Medicine and Biology*, 598, 389-406.
- Velloso, L. M., Svensson, K., Schneider, G., Petterson, R. F., & Lindqvist, Y. (2002). Crystal structure of the carbohydrate recognition domain of P58/Brgic-53, a protein involved in glycoprotein export from the endoplasmic reticulum. *Journal of Biological Chemistry*, 277(18), 15979-15984.
- Wang, P. H., Huang, T. Z., Zhang, X. B., & He, J. G. (2014). Antiviral defense in shrimp: From innate immunity to viral infection. *Antiviral Research*, 108, 129-141.
- Wang, X. W., & Wang, J. X. (2013). Diversity and multiple functions of lectins in shrimp immunity. *Development and Comparative Immunology*, 39, 27-38.
- Wang, X. W., Xu, W. T., Zhang, X. W., Zhao, X. F., Yu, X. Q., & Wang, J. X. (2009). A C-type lectin is involved in the innate immune response of Chinese white shrimp. *Fish & Shellfish Immunology*, 27, 556-562.
- Wang, X. W., Zhao, X. F., & Wang, J. X. (2013). C-type lectin binds to β -Intergrin to promote hemocytic phagocytosis in an invertebrate. *Journal of Biological Chemistry*, 289(4), 2405-2414.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., ... Schwede, T. (2018) SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*. 46(1), 296-303.
- Wongpanya, R., Sengprasert, P., Amparyup, P., & Tassanakajon, A. (2017). A novel C-type lectin in black tiger shrimp *Penaeus monodon* functions as a pattern recognition receptor by binding and causing bacterial agglutination. *Fish & Shellfish Immunology*, 60, 103-113.
- Xiu, Y. J., Wang, Y. H., Jing, Y. T., Qi, Y. K., Ding, Z. F, Meng, Q. G., & Wang, W. (2015). Molecular cloning, characterization, and expression analysis of two different types of lectins from the oriental river prawn, *Macrobrachium nipponense*. *Fish & Shellfish Immunology*, 45, 465-469.

Xu, S., Wang, L., Wang, X. W., Zhao, Y. R., Bi, W. J., Zhao, X. F., & Wang, J. X. (2014). L-type lectin from the kuruma shrimp *Marsupenaeus japonicas* promotes hemocyte phagocytosis. *Development and Comparative Immunology*, 44, 397-405.

Zhang, H., Peatman, E., Liu, H., Feng, T. T., Chen, L. Q., & Liu, Z. J. (2012). Molecular characterization of three L-type lectin genes from channel catfish, *Ictalurus punctatus* and their responses to *Edwardsiella ictaluri* challenge. *Fish & Shellfish Immunology*, 32, 598-608.

University of Malaya