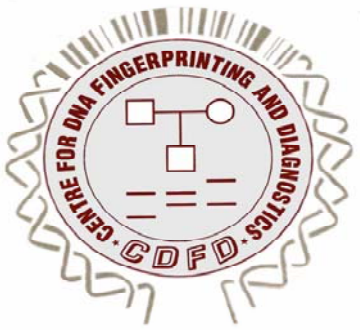

Insights into insect immunity: Analysis of novel immune proteins from silkmoths

Thesis submitted to

Manipal University

for the Degree of

Doctor of Philosophy

By

Archana Gandhe

Centre of Excellence for Genetics and Genomics of Silkmoths

Laboratory of Molecular Genetics
Centre for DNA Fingerprinting and Diagnostics
Hyderabad 500076
October 2007

DECLARATION

The research work embodied in this thesis entitled "Insights into insect immunity: Analysis of novel immune proteins from silkmoths," has been carried out by me at the Centre of Excellence for Genetics and Genomics of Silkmoths, Laboratory of Molecular Genetics, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, under the guidance of Dr. J. Nagaraju. I hereby declare that this work is original and has not been submitted in part or full for any degree or diploma of any other university or institution.

Ms. Archana S. Gandhe
Centre of Excellence for Genetics and Genomics of Silkmoths
Laboratory of Molecular Genetics
Centre for DNA Fingerprinting
and Diagnostics, Hyderabad**.**

CERTIFICATE

This is to certify that this thesis entitled "Insights into insect immunity: Analysis of novel immune proteins from silkmoths," submitted by Ms Archana Gandhe for the degree of Doctor of Philosophy to Manipal University is based on the work carried out by her at the Laboratory of Molecular Genetics, Centre for DNA Fingerprinting and Diagnostics, Hyderabad. This work is original and has not been submitted in part or full for any degree or diploma of any other university or institution.

Dr. J. Nagaraju
Thesis Supervisor
Centre of Excellence for Genetics
and Genomics of Silkmoths
Laboratory of Molecular Genetics
Centre for DNA Fingerprinting
and Diagnostics, Hyderabad.

Dr. Shekhar C Mande
Dean, Academic affairs
Centre for DNA Fingerprinting
and Diagnostics, Hyderabad

Contents

Acknowledgements

This thesis is not just an outcome of my work at CDFD since the past five years but also of many individuals who have helped me along the way to achieve this moment in my life. My experience in CDFD as a PhD student has been a wonderful and cherishing one. The credit for the same goes to several people who directly or indirectly helped me achieve this goal. First and foremost, I thank my guide Dr. Nagaraju for giving me this opportunity to work at the Laboratory of Molecular Genetics, CDFD. Very importantly, he taught me to work independently but his helpful advice was available at any time. This thesis would not have been possible without his kind support, encouragement, trenchant criticism, probing questions and patience. My heartfelt thanks to Savithri Madam for her affectionate and kind gestures that brought a unique home-like touch during my stay at Hyderabad. I am grateful to Dr. Gowrishankar for his valuable inputs in my research. I especially cherish all the interactions with him in 'Genetics Club' seminars, 'Briefings' and other scientific presentations that were very informative and inspiring. My heartfelt thanks to Dr. Shekhar Mande for supporting me not only academically but also morally and administratively. I also thank Dr. Sanjeev Khosla, Dr. Ranjan Sen and Dr. Gayatri for their timely inputs and advice. I am gratified by the valuable suggestions and kind help by Prof. Ramasarma. A special acknowledgement to Dr. Vidya Gupta for her excellent and patient guidance during my first stint in research as a project assistant at NCL. I also thank Dr. Hasnain for his support and for giving me this opportunity at CDFD. I am very thankful to my colleagues Janardhan and Serene who worked with me on this project and enabled me to accomplish this goal. I am indebted to my senior Dr. Sriramana for his excellent advice and continuous help throughout. I also thank Arun for his timely help particularly in using various software programs. My thanks to my lab members and colleagues Jyoti, Padmapriya, Satish, Dr. Prashant, Sunil, Vidya, Reddy, Murali, Eshwar, Jayendra, Dharma Prasad, Sravan, Muthulaxmi, Sobhan babu, Asha, Rebecca, Vijay, Kaliappan, Swarna, Ramesh babu, Rafiq, Rohit, Deena, Manju, Madhu, Subhadra Jyoti, Satyavathi, Gayatri, Premdeepa, Subbaiah, Selvendran, Mahesh, Angel, Sudhir, Naushad, Krishna Prasad, Rajitha, Jamal, Srinivas and Prasad. My special thanks to Chandrasekhar and Vijaybhaskar for their assistance. I appreciate the friendship and company of my friends Abira, Archana, Ghazala, Shubhada and Jaya. I am thankful to Nitin

and his wife Minal for their warm friendship and for all the Ganeshutsavs that I participated at their home. It was indeed a home away from home. I am also thankful to Krishnaveni, Santosh, Pankaj, Nitin, Shafiq, Abhijit and Rajendra for their excellent help in my research. I thank Mr. Prakash, Mr. V.H Rao, Mr. Ravinder, Prashanthi, Punitha and all other administrative staff for their help throughout. I am thankful to the security staff, drivers, maintenance staff and dispatch section people for their assistance. I also thank the research committee at Manipal University for registering me with them for my PhD degree.

And lastly but most importantly I thank my family whose love and unconditional support made me sail through the thick and thin of this long journey. My heartfelt appreciation for Sunil, my life partner for being a constant pillar of strength throughout. But for his patience and constant motivation this thesis could not have been materialized. I thank my father and mother for their selfless love and sacrifice which propelled me in life. I truly believe that this PhD has been a result of their blessings. I am also indebted to my parents in law for their continuous encouragement that enthused confidence in me to carry out this endeavour. And then thanks to my brother Niranjan for his great sense of humor and for teaching me to look at the comical side of every situation. I am thankful to my brother in law, Sujit for his camaraderie right through. And finally, I thank all those people whom I could not mention here due to the space constraint but have nevertheless touched upon my life and helped me in some or the other way to achieve this goal in life.

List of Abbreviations

1. AMPs- anti-microbial peptides
2. PO- phenoloxidase
3. MIC- minimum inhibitory concentration
4. NAM- N-acetyl muramic acid
5. NAG- N-acetyl glucosamine
6. PGN- peptidoglycan
7. LPS-lipopolysaccharide
8. proPO- prophenoloxidase
9. PAPs- proPO activating proteinases
10. PRRs- protein recognition receptors
11. PAMPs- pathogen associated molecular patterns
12. PGRPs- peptidoglycan recognition proteins
13. βGRPs- β glucan recognition proteins
14. GNBPs- gram-negative bacteria binding proteins
15. TLRs- Toll-like receptors
16. RNAi- RNA interference
17. miRNAs- micro RNAs
18. siRNAs- short interfering RNAs
19. RISC- RNA induced silencing complex
20. DXV- Drosophila X virus
21. JAK-STAT- Janus kinase-signal transducer of activation
22. DCV- Drosophila C virus
23. hpi- hours post infection
24. GO-gene ontology
25. DFPs- defence proteins
26. Ig- immunoglobulin
27. PSI-Blast- position-specific iterative Blast
28. ECM- extracellular matrix
29. LLPs- lysozyme-like proteins
30. ALLP- *A. mylitta* lysozyme-like protein
31. BLLP- *B. mori* lysozyme-like protein

32. AL- *A. mylitta* lysozyme
33. BL- *B. mori* lysozyme
34. HL-hen lysozyme
35. MTCC- Microbial Type Culture Collection
36. MW- molecular weight
37. pI- isoelectric pH
38. MOE- Molecular Operating Environment
39. IC- inhibitory concentration
40. CFU- colony forming units
41. IM- inner membrane
42. IBS- insect buffer saline
43. GFP- green fluorescent protein
44. dsRNA- double stranded RNA

Synopsis

Insect immunity has been studied with renewed interest in the past decade due to its close resemblance to mammalian innate immune systems. An understanding of insect immune mechanisms also offers strategies to control spread of insect vector borne diseases in animals. In addition, an efficient management of insect pests of agricultural crops and disease resistance of beneficial insects can also be achieved via a thorough knowledge of its immune system. The dipteran insect, *Drosophila melanogaster* has been widely investigated to study insect immune responses towards diverse pathogens like bacteria, fungi, parasites and viruses. However lepidopteran immune mechanisms remain less understood. The present study analyses the immune responses of two economically important lepidopteran insects, the domesticated silkmoth *Bombyx mori* and the Indian tasar wild silkmoth, *Antheraea mylitta* which are widely cultivated for silk.

Chapter 1 gives a background of insect immunity with reference to the information gathered from studies on the model insect *D. melanogaster*. Both humoral as well as cellular immune responses are discussed. Humoral response is mainly characterized by an array of anti-microbial peptides (AMPs) secreted in to the insect haemolymph. Cellular responses, on the other hand, are mediated by insect blood cells or haemocytes. Haemocyte mediated responses include phagocytosis, encapsulation or nodulation and melanization. The various steps which include: recognition, signal transduction and synthesis of effector molecules involved in an immune response are also described. The two major signalling pathways of anti-microbial peptide induction, Toll and Imd, elucidated in Drosophila have been described. Some latest research papers also throw light on the anti-viral immune mechanism in insects which are hitherto less well known. The anti-viral pathways suggested include RNA interference, Toll pathway and JAK-STAT pathway. Finally advances in lepidopteran immunity are discussed. Several lepidopteran insect species like *B. mori*, *Manduca sexta*, *Hyalophora cecropia* etc are especially studied due to their economic importance, their large body large size which facilitates proteomic and biochemical study and ease of their culture in laboratory. AMPs from these species have been extensively analysed and their structures and mechanism of action have been elucidated. Cecropin, a ubiquitous AMP present in all insects was first identified and its sequence determined from *H. cecropia* and hence designated as cecropin. Finally the research aims and subsequent results addressed in this dissertation are described.

Chapter 2 describes the experimental analysis of *A. mylitta* immune challenged fat-body transcriptome to unravel potential immune genes. *A. mylitta* is an economically important tropical wild silkmoth reared for its silk known as tasar silk. However, very little is known about the immune genes of this organism. Hence, to explore the *A. mylitta* immune response genes, an expressed sequence tag (EST) library from bacteria challenged fat body tissue was generated. Immune repertoire of *A. mylitta* included both known as well as putative immune genes. A total of 719 unique sequences were obtained out of which around 33% showed similarity with immune genes reported from other organisms. A family of three putative novel genes (DFP-1, 2 and 3) was identified that bore similarity with extracellular matrix proteins from vertebrates. Also another putative defence protein (DFP-4) was identified which was similar and yet distinct from cecropin. A novel lysozyme like protein was also detected which lacked the characteristic catalytic amino acid residues of lysozyme required for peptidoglycan hydrolysis activity. Immune upregulation of a few putative immune genes was confirmed by semi-quantitative RT-PCR. The analysis of the *A. mylitta* immune transcriptome thus led to identification of hundreds of putative immune genes which should facilitate their further characterization. The next two chapters describe the characterization of two of the novel putative immune genes from the *A. mylitta* transcriptome.

Chapter 3 includes the functional analysis of a novel lysozyme-like protein family from both *A. mylitta* as well as *B. mori*. They were designated as lysozyme-like proteins (LLPs) owing to their partial similarity with lysozymes. However, lack of characteristic catalytic amino acid residues essential for muramidase activity in LLPs clearly distinguishes them from classical lysozymes. Two LLPs, one from *B. mori* (BLLP1) and the other from *A. mylitta* (ALLP1) expressed in a recombinant system, exhibited a broad-spectrum anti-bacterial activity. Further study of the anti-bacterial mechanism revealed that BLLP1 is bacteriostatic rather than bactericidal against *Escherichia coli* and *Micrococcus luteus*. A substantial increase in haemolymph bacterial load was observed in *B. mori* upon RNA interference mediated in vivo knock-down of BLLP1. Interestingly, as we have shown here that the anti-bacterial mechanism of this protein depends on peptidoglycan binding unlike peptidoglycan hydrolysis or membrane permeabilization as observed with lysozymes and most other anti-microbial peptides! Considering this unique feature of the protein its worth assessing whether its justified in naming the protein as lysozyme-like protein! Literature survey reveals that this is the first report on functional analysis of a novel, non-catalytic lysozyme-like family of anti-bacterial proteins that are quite apart functionally from classical lysozymes.

Chapter 4 gives details of the functional analysis of yet another novel defence protein (DFP-1) identified from *A. mylitta* immune transcriptome. Our results indicated a role for this protein in nodulation response of insects and hence designated as Noduler. Noduler possessed a characteristic reeler domain found in several extracellular matrix vertebrate proteins. Noduler was shown in vitro to bind a wide range of bacteria, yeast and also insect haemocytes. Furthermore, Noduler specifically bound lipopolysaccharide, lipotechoic acid and β-1, 3 glucan components of microbial cell walls. RNA-interference mediated knock-down of Noduler resulted in significant reduction in the number of nodules and consequent increase in bacterial load in larval haemolymph. The results suggest that Noduler is widely conserved and is involved in very early clearance of bacteria by forming nodules of haemocytes and bacterial complexes in insects. The results would promote further studies for understanding of the crucial but hitherto overlooked nodulation mechanism in insects and also provide cues for the study of similar mammalian proteins whose function is not understood.

The results emanated from the studies described in the thesis point towards a vast repertoire of molecules involved in a variety of immune functions in insects. As more and more whole genome sequence data and expressed sequence tags data become available from diverse insect sources many more novel molecules implicated in immune function will emerge to be added to the already existing amazing diversity of immune molecules in insects that has contributed to the evolutionary success of insects as the most diverse and abundant taxon in the animal kingdom.

Chapter-1

General Introduction

1.1 Introduction

'Silk' has been extremely desirable to mankind since ancient times. Silks fall into two main types, the mulberry and the non-mulberry silks. The classification is based on the type of silkworms that are used for silk production. Almost all the varieties of mulberry silks are derived from the domesticated silkworm *Bombyx mori*. The non-mulberry silks often called the 'Wild Silks' are derived from the silkworms which are not domesticated. 'Wild Silks' generally come in natural colours - cream, beige, brown and gold. They are not only user-friendly but also healthy owing to their porous texture and thermal properties. Tribal communities and economically disadvantaged sections of the society are the primary rearers of these silkworms.

Silkmoth species include the domesticated silkmoth *B. mori* of family Bombycidae and wild silkmoths that belong to Saturniidae, *Antheraea mylitta* (Indian tropical tasar silkmoth), *A. roylei* (Indian oak tasar silkmoth), *A. assama* (Indian golden silkmoth), *A. yamamai* (Japanese oak silkmoth), *A. pernyi* (Chinese oak silkmoth) and S*amia cynthia ricini* (Indian castor silkmoth). Silk production based on these moths, especially *B. mori*, *A. pernyi*, *A. mylitta* and *A. assama* plays important role in rural economies of many developing nations.

B. mori is widely reared for its silk. Life cycle of silkworm comprises embryonic, larval, pupal and moth stages spanning for about 35-45 days. Broadly all the silkworm varieties are classified into two different types, diapausing and non-diapausing varieties. The diapause strains hibernate during winter and have one or two life cycles per year (also referred as 'uni' or 'bivoltine'). Non-diapause strains go through multiple generations (hence termed 'multivoltine') without intervening embryonic diapause. The life cycle of silkworm, *B. mori* is shown in Figure 1.1. The larval stage, which is the only feeding stage in the entire life cycle, is comprised of five instars during which larvae feed on mulberry (*Morus alba*) leaves. The larval stage is also the time when silkworms are highly vulnerable to various pathogens such as fungal, protozoan, viral and bacterial pathogens. Diseases of silkworm are classified as pebrine (caused by microsporidia *Nosema* sp.), grasserie (caused by nuclear polyhedrosis virus), flacherie (refers to diseases caused by bacteria and viruses individually or in combination). Amongst the bacterial diseases, staphylococcal and streptococcal infections are also reported to be common amongst silkworm.

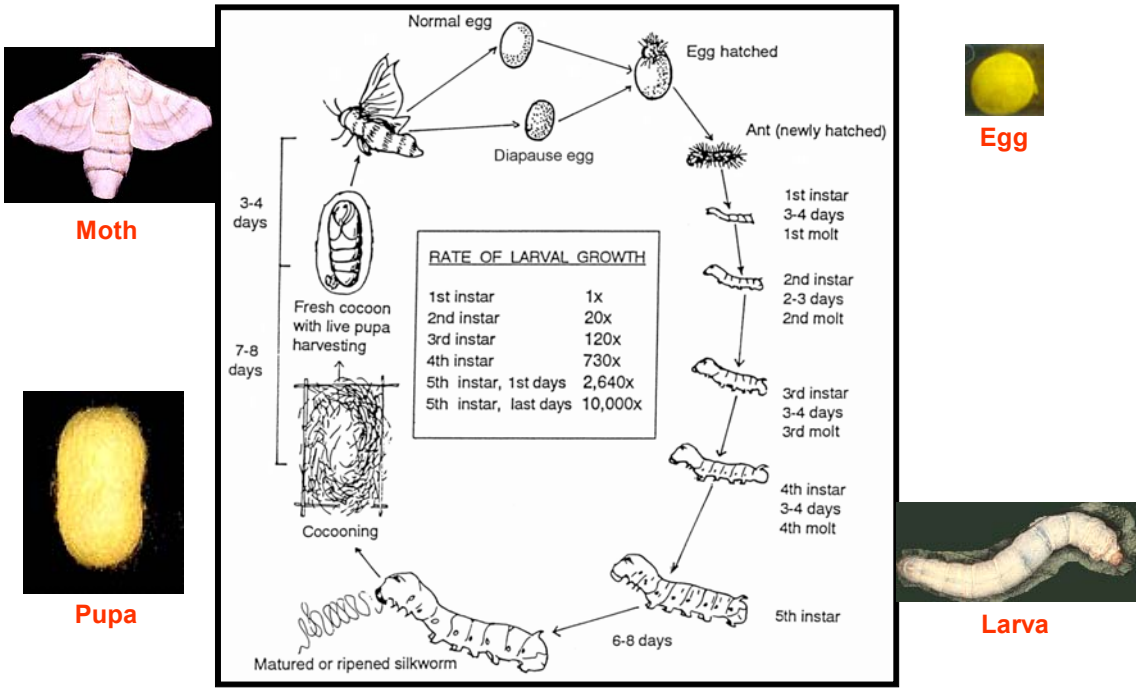

Figure 1.1 Life cycle of the domesticated silkmoth, *Bombyx mori*

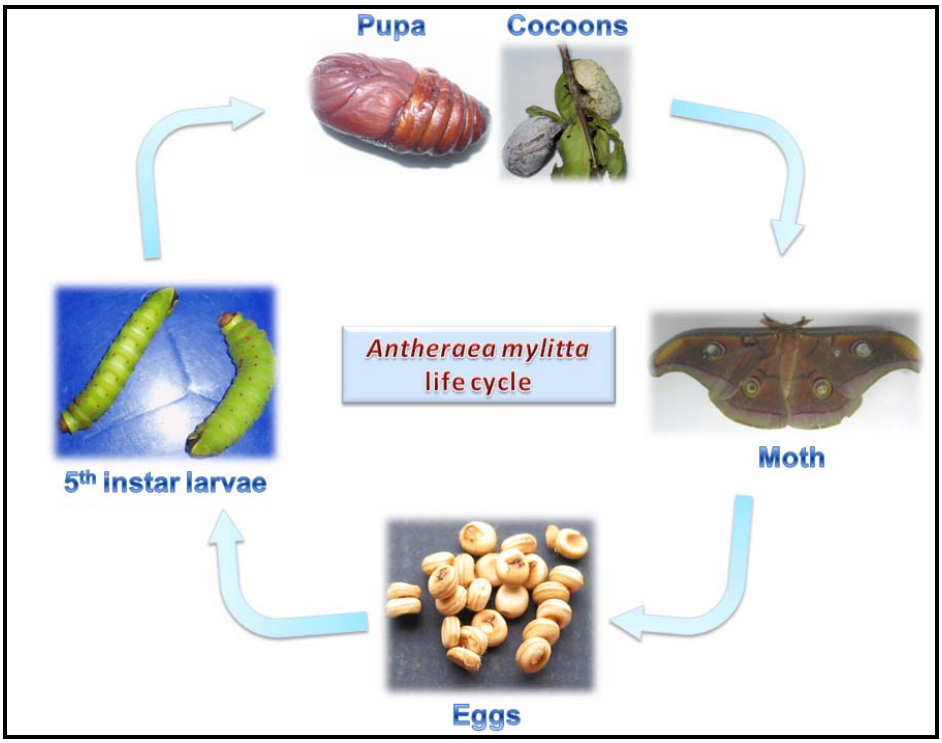

Figure 1.2 Life cycle of the Indian tasar wild silkmoth *Antheraea mylitta*

The Indian tasar silkmoth *A. mylitta* is a natural fauna of tropical India, represented by 19 ecoraces. Tasar culture is inherent in India and practiced by tribals in natural habitats. The life cycle of *A. mylitta* is depicted in Figure 1.2. The common host plants of tasar larvae are *Terminalia tomentosa*, *T. arjuna* and *Shorea robusta*. Tasar silkworms are exposed to harsher environments and wide-ranging pathogens in wild as compared to domesticated silkmoth *B. mori*. However, study of its immunity and causative pathogens is limited posing difficulty for assigning efficient control measures.

Immunity is defined as the ability of an organism to resist infection. The origin of term "immunity" has interestingly come from the Latin word *immunis* meaning exemption from military service, tax payments or other military services. The concept of immunity has fascinated mankind for more than thousand years. The 19th century witnessed a great leap in the field of human immunology which resulted in the advent of vaccines. Since then the impact of therapeutic immunology has been ever-growing and has immensely benefited the mankind. In fact the evolutionary success of an organism depends greatly on its immunity or its ability to overcome infection. Insects are an excellent example of this principle. Insects possess a very effective immune system exemplified by their evolutionary success. Around more than a million species of insects inhabit almost every niche in the environment. The vertebrate immune responses have been widely studied and comprise of innate and adaptive responses. The innate immune responses include the non-specific mechanisms while the adaptive immunity is mediated by T and B cells and is pathogen-specific. The vertebrate immune mechanisms also possess a memory that results into an enhanced response upon reinfection with an organism. However insects have been found to lack T and B cell mediated adaptive responses and possess only the innate arm of defence. Nevertheless, recent evidences suggest presence of adaptive responses in insects. An immunoglobulin family protein called Dscam that exists as more than 30,000 isoforms has been identified to play an immune function in Drosophila as well as Anopheles (DONG *et al.* 2006; WATSON *et al.* 2005). Study of mammalian immunity is very attractive due to the potential applications in prevention and cure of diseases. However, insect immunity also has come into renewed focus in the past decade for its great resemblance with the mammalian innate immune system. Drosophila is used as a model system to study mammalian innate mechanisms. Also, insect vectors play an important role in the transmission of devastating human diseases like malaria, dengue, leishmaniasis etc. Several agricultural crops are affected every year by insect pests like *Heliothis armigera*, *Manduca sexta* and *Spodoptera armigera*. An understanding of the

insect immune system thus offers to provide probable solutions in management of insect vectors and pests. In addition, study of insect immunity will also provide efficient strategies to control infections in beneficial insects like *B. mori* and *Apis mellifera* which are prone to bacterial, viral and fungal infections.

Insects are evolutionarily successful organisms and occupy almost all habitats in nature. An efficient immune system is one of the attributes for this evolutionary success. Insects rely solely on innate immune responses though some recent reports give preliminary evidences for existence of adaptive mechanisms. Innate immune system in insects can be divided into humoral and cellular mechanisms. Humoral mechanism is characterized by a rapid synthesis of an array of anti-microbial peptides (AMPs) by the insect fat body that are subsequently secreted in the haemolymph. The insect fat body is an organ analogous to the mammalian liver and a major site of AMP synthesis. However other tissues like haemocytes, epithelium and mid gut also partly contribute to AMP synthesis. Three types of cellular responses are mediated by haemocytes in Drosophila - 1. Phagocytosis or engulfment of foreign particles by a subset of haemocytes called plasmatocytes. 2. Nodulation or encapsulation refers to aggregation of haemocytes i.e. lamellocytes that entrap pathogens and 3. Phenoloxidase (PO) cascade, a serine protease cascade that leads to formation of melanin at the site of infection. Enzymes essential for this proteolytic cascade are secreted by a haemocyte cell type called crystal cells. Figure 1.3 illustrates different immune mechanisms prevalent in insects.

1.2 Insect immune system

1.2.1 AMPs

Insect anti-microbial molecules are mostly peptides or polypeptides containing less than 150 amino acid residues and are mostly induced upon infection. Insect AMPs are induced to extremely high levels (upto mg concentrations) in haemolymph, which is much above the minimum inhibitory concentrations (MIC) required to kill the microorganism (BULET and STOCKLIN 2005). The molecular mechanism of AMP gene induction has been specifically well studied in *D. melanogaster*. Most of the AMPs have a net positive charge at physiological pH due to a higher content of positively charged amino acid residues like arginine and lysine (BULET *et al.* 2004) and hence can interact by electrostatic attraction with the negatively charged microbial cell membranes (Figure 1.4). A large number of anti-

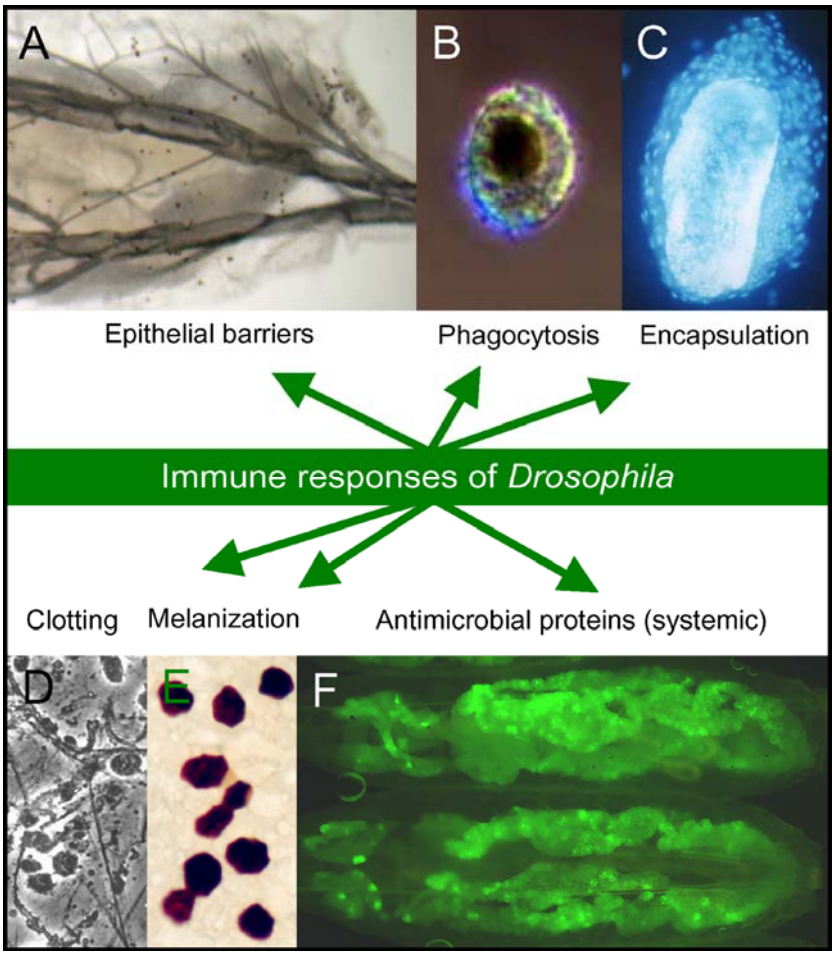

Figure 1.3 Schematic representation of the various innate immune mechanisms in insects. A. Epithelial barriers act as first line of defence, B. A plasmatocyte that mediates phagocytosis, C. Encapsulation of a parasite egg by Drosophila lamellocytes. The parasite egg and lamellocytes are stained with a fluorescent nuclear stain, D. Clot formation that leads to wound healing, E. Crystal cells in contact with the cuticle. The cells are melanized, which is induced upon infection or wounding, F. Anti-microbial peptide synthesis in insect fat body. The fat body appears green due to induction of a reporter gene (green fluorescent protein), which is under the control of anti-microbial peptide gene promoter (Adapted from Govind *et. al.*, PLOS Biology, 2004).

bacterial proteins have been identified across various insect orders, which include cecropins, attacins, lysozymes, defensins, lebocins, apidaecins, abaecins, gloverins, drosocins etc. Relatively less emphasis has gone into the identification of anti-fungal and anti-parasitic molecules. However, a few anti-fungal proteins that have been reported are drosomycins and metchnikowins from Drosophila and heliomicins from lepidopteran *Heliothis virescens*. AMPs are classified into four main groups based on their structure and mode of action a. Linear, amphipathic α-helices b. Cysteine- stabilized peptides with mixed α-helix and β-sheet structures c. Peptides with an over-representation of some amino acids d. Enzymes that digest bacterial cell wall e.g. lysozymes.

a. Linear, amphipathic α-helices

The AMPs belonging to this category are small linear, α-helical peptides that are 29-42 residues in length and lack cysteine residues e.g. cecropins (BULET and STOCKLIN 2005). These peptides possess hydrophobic and hydrophilic α-helical regions (Figure 1.5). Cecropins have been demonstrated to form helical structures only in hydrophobic environments and remain in an unordered state in aqueous solution (STEINER 1982). This structural property of formation of α-helical conformation upon contact with hydrophobic bacterial cell membrane renders the anti-bacterial action leading to membrane disintegration and lysis of the bacterial cell.

b. Cysteine- stabilized peptides, mixed α-helix and β-sheet structures

These include peptides, which contain both α-helical and β-sheet structures that are strengthened by intramolecular disulphide bonds e.g. defensins, heliomicins, drosomycins, termicins and thanatins (Figure 1.5). Defensins are present widely across all insect orders investigated. Defensins are predominantly active against gram-positive bacteria although a few anti-fungal defensins (drosomycins, heliomicins) have also been identified. Defensins kill bacteria by disrupting the permeability barrier of the cytoplasmic membrane leading to leakage of cytoplasmic potassium and partial depolarization of bacterial membrane (COCIANCICH *et al.* 1993). In contrast the anti-fungal defensins have been shown to act by inhibiting spore germination and through hyphal perforation (BULET and STOCKLIN 2005) .

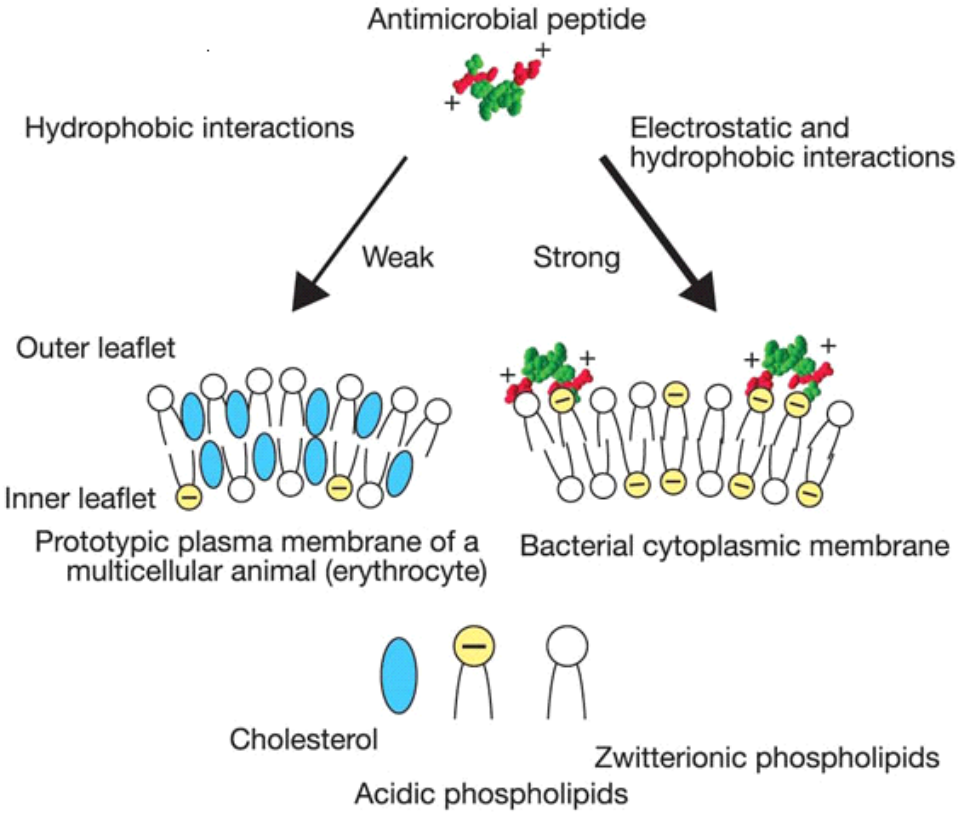

Figure 1.4 Basis of differential action of anti-microbial peptides on eukaryotic and prokaryotic cells. Bacterial membrane is negatively charged due to presence of phospholipids. This ensures a strong interaction between the cationic AMPs and the bacterial membrane. However, the surface of eukaryotic cell is relatively neutral owing to zwitterionic phospholipids and cholesterol and hence interaction of AMPs is weak.

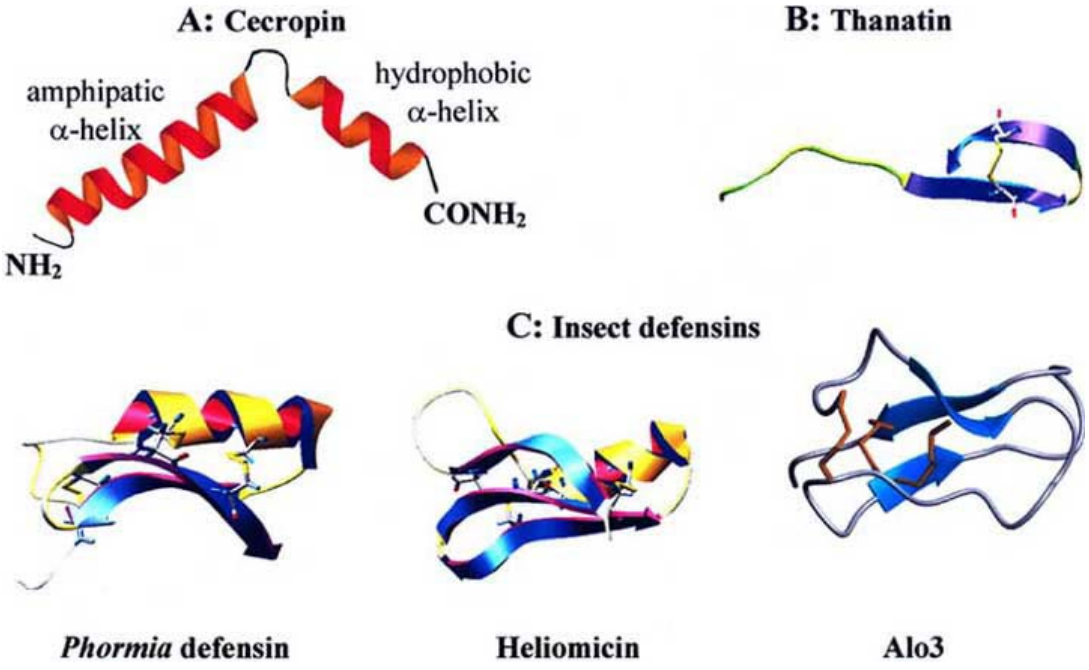

Figure 1.5 Three dimensional structure of five different insect AMPs, A) Cecropin from *Hyalophora cecropia*, B) Thanatin from *Podisus maculiventris*, C) Anti-bacterial defensin from *Phormia terranovae*, anti-fungal heliomicin from *Heliothis virescens* and Alo3 from *Acrocinus longimanus* (Adapted from Bulet *et. al.*, Protein and Peptide Letters, 2005).

c. Peptides with an over-representation of some amino acids

The peptides under this category are 14-40 residues in size, include proline-rich peptides like apidaecins (honey bee), abaecins (bumble bee), lebocins (lepidopteran insects), drosocins and metchnikowins (fruit fly). Proline residues are not only over-represented in these peptides (more than 25 %) but also exist in doublets along with basic amino acids arginine and histidine as PRP or PHP motifs. The shorter proline rich peptides (less than 20 residues) are active against gram-negative bacteria especially belonging to *Enterobacteriaceae* family although gram-positive strains remain mostly non-susceptible (BULET and STOCKLIN 2005). However, the long chain peptides like abaecins and lebocins exert activity against both gram-positive and gram-negative bacteria while metchnikowins show only anti-fungal activity (BULET *et al.* 1999). Studies revealed that unlike other AMPs, small proline rich peptides

such as apidaecins require several hours to kill bacteria. It was also demonstrated that apidaecin is taken up by permease/transporter mediated uptake in the *Escherichia coli* cell (CASTLE *et al.* 1999). Some of these proline rich peptides have been shown to bind the *E. coli* heat shock protein DnaK thereby preventing the chaperone-mediated protein folding leading to death of the bacterial cell (KRAGOL *et al.* 2001).

Another class of AMPs belonging to this group are the glycine-rich attacins and gloverins. While attacin is found both in lepidopterans as well as dipterans (DUSHAY *et al.* 2000; HULTMARK *et al.* 1983), gloverin is found exclusively in lepidopteran insects (AXEN *et al.* 1997). They are large, linear proteins ranging from 12-19 kDa in size. Attacins and gloverins have been demonstrated to inhibit synthesis of outer membrane proteins in *E. coli* leading to increased membrane permeabilisation (AXEN *et al.* 1997; CARLSSON *et al.* 1991).

d. Enzymes that digest bacterial cell wall

Lysozymes are muramidases that cleave the β 1, 4 glycosidic linkage between the N-acteyl muramic acid (NAM) and N-acetyl glucosamine (NAG) polymer of peptidoglycan backbone in the bacterial cell wall. Lysozymes are ubiquitous in insects and are upregulated upon infection as against the constitutively expressed vertebrate lysozymes. The insect lysozymes belong to the chicken-type (c-type) lysozyme category. Structural studies have revealed the 3-D conformation of *A. mylitta* and *B. mori* lysozymes to be similar yet distinct from vertebrate lysozymes (JAIN *et al.* 2001; MATSUURA *et al.* 2002). Although lysozymes have been shown to be involved in immunity across insect species, a special adaptation is seen in flies like Drosophila and Musca (HULTMARK 1996). Since, these flies feed on decomposing matter containing microorganisms their lysozymes have been adapted to serve a digestive function and are expressed constitutively in digestive tract.

1.2.2 Cellular responses

a. Phagocytosis

Phagocytosis can be defined as engulfment of pathogens by host haemocytes or blood cells. In Drosophila the macrophage-like cells involved in phagocytosis are termed as

plasmatocytes. This process requires pathogen recognition by the receptors present on plasmatocytes followed by engulfment. Unlike the well-studied AMP synthesis pathways, phagocytosis is less well understood. Only recently, researchers have focused on mechanisms of phagocytosis in Drosophila in an attempt to understand the mammalian phagocytosis process. A Drosophila cell line S2 is highly phagocytic and is being widely used for RNAi screens to identify novel players in phagocytosis. Recently, two transmembrane proteins namely Eater (KOCKS *et al.* 2005) and Nimrod (KURUCZ *et al.* 2007) containing many epidermal growth factor-like repeats were identified in Drosophila using differential expression screens. Knock-down of Eater expression in S2 cells reduced phagocytosis of *E. coli* and *Serratia marcescens* by over 50%. Similarly, suppression of Nimrod expression in plasmatocytes inhibited phagocytosis of *Staphylococcus aureus*. Eater and Nimrod superfamily proteins are conserved in other insects as well as humans.

b. Nodulation/Encapsulation

Nodulation refers to haemocyte aggregation around microorganisms like bacteria and fungi while encapsulation is the haemocyte aggregation around larger pathogens like parasitoids or wasps. Lamellocytes, a subset of haemocytes are involved in nodulation/encapsulation responses in Drosophila. Upon infection these haemocytes exhibit a change in morphology and behaviour thereby getting transformed from freely circulating cells to adherent cells tending to form aggregates. However, the molecular mechanisms underlying these changes have not been well understood. Prior reports suggest involvement of eicosanoids in mediating nodulation (MILLER *et al.* 1994). Eicosanoids are signalling molecules known to trigger macrophage locomotion and morphological changes upon infection in mammals. However, their role in insect immunity and the pathways that mediate nodulation remain to be understood at this stage.

c. Phenoloxidase cascade

A serine protease cascade reminiscent of the coagulation cascade in mammals is activated in insects upon recognition of pathogen-associated molecules such as peptidoglycan (PGN), β-1, 3 glucan or lipopolysachharide (LPS). The last enzyme of this pathway, phenoloxidase (PO) is present as a zymogen, prophenoloxidase (proPO) in the haemolymph, which is then

activated by proteolytic cleavage at a specific site near its amino terminus (KANOST *et al.* 2004). The PO catalyses synthesis of melanin from dopamine that leads to wound healing and the process also generates toxic intermediates such as reactive oxygen species resulting in pathogen killing. The exact molecular pathway of the PO cascade is still not well defined although a series of proPO activating enzymes (PAPs) have been identified (JIANG and KANOST 2000). Endogenous serine protease inhibitors (serpins) have been reported from insects that inhibit various proteases in the pathway (KANOST 1999). Serpins probably control the induction of the pathway in absence of infection and also possibly localizes its activation at site of infection preventing systemic melanization. In Drosophila, crystal cells have been identified to secrete enzymes involved in phenoloxidase pathway.

1.2.3 Steps in insect immunity

Insect immunity can be divided into three steps based on the chronological events that take place in an immune response –

a. Recognition
b. Signalling
c. Effector mechanisms

a. ***Recognition***: How do insects sense infection by invading pathogens and discriminate self from non-self? This is a key step in any immune response and is mediated by a group of proteins termed as pathogen recognition receptors (PRRs) that recognize pathogen associated molecular patterns (PAMPs) such as LPS, PGN and β-1,3 glucan etc. that are specific to pathogens but are absent in the host (HOFFMANN *et al.* 1999). Different PRRs identified from insects include peptidoglycan recognition proteins (PGRPs) (KANG *et al.* 1998), β-1, 3 glucan recognition proteins (βGRPs) (OCHIAI and ASHIDA 1988) also termed as gram-negative bacteria binding proteins (GNBPs) (KIM *et al.* 2000), lectins (LACKIE and VASTA 1988) and hemolins (SUN *et al.* 1990).

PGRPs

PGRPs are a family of proteins that bind PGN, which is a polymer of NAG and NAM present in the bacterial cell wall. Several PGRP isoforms have been identified from Drosophila

(WERNER *et al.* 2000) as well as in *B. mori* (YOSHIDA *et al.* 1996) and *M. sexta* (YU *et al.* 2002). PGRPs are classified into short (S) or long (L) forms and are either secreted or transmembrane proteins. PGRP-LC, a transmembrane PGRP, is required for activation of Imd pathway in Drosophila while PGRP-SA, a secreted isoform is involved in detection of gram-positive bacteria and subsequent activation of Toll pathway (MICHEL *et al.* 2001).

βGRPs

Proteins binding to β-1, 3 glucan components in fungal cell wall have been identified in several insect species (MA and KANOST 2000; OCHIAI and ASHIDA 2000). These proteins also bind LPS and are termed as gram-negative bacteria binding proteins (GNBPs). The βGRPs possess a C-terminal glucanase-like domain, which lacks enzymatic activity due to substitutions at the key catalytic residues. Its N-terminal domain is involved in binding to β-1, 3 glucans. Upon binding to its bacterial and fungal ligands, βGRPs have been shown to activate the PO cascade in many insects.

Lectins

C-type lectins have been reported from several insects to function as recognition proteins in immunity. C-type lectins bind to carbohydrates in a calcium dependent manner. They contain one or two carbohydrate recognition domains that bind to LPS from gram-negative bacteria (KOIZUMI *et al.* 1999; YU and KANOST 2000).

b. Signalling pathways

During the last decade the molecular mechanisms of pathogen recognition and signalling have been well studied, especially in *D. melanogaster*. The two pathways of antimicrobial peptide induction, Toll pathway activated by gram-positive bacteria and fungi (LEMAITRE *et al.* 1996) and the Imd pathway triggered by gram-negative bacteria (LEMAITRE *et al.* 1995) have been especially well elucidated in this model organism (Figure 1.6). The Toll pathway is selectively turned on by gram-positive bacteria and fungi thereby mediating activation and nuclear translocation of two NF-κB transcription factors, DIF and Dorsal, respectively. These transcription factors initiate the induction of various target anti-microbial genes e.g. anti-

fungal *drosomycin* gene. Imd pathway, on the other hand, is triggered by gram-negative bacteria activating a NF-κB transcription factor called Relish that leads to induction of anti-microbial peptide genes like diptericin. Thus it has been well demonstrated in Drosophila that insect immune system is broadly specific and exerts a differential response against different categories of pathogens. It has also been shown that the differential activation of Toll and Imd is achieved by detection of the gram-positive lysine type PGN and the gram-negative diaminopimelic acid type PGN, respectively by the immune system in Drosophila (LEULIER *et al.* 2003). The presence of Toll and Imd pathways has been confirmed in other dipteran insects like mosquitoes (LUNA *et al.* 2006). Also, analogues of the toll-like genes have been reported in *B. mori* (CHENG *et al.* 2007; IMAMURA and YAMAKAWA 2002) and hymenopteran insect *A. mellifera* (EVANS *et al.* 2006). However the presence of these pathways in these insects has not been experimentally proved. The identification of Toll function in Drosophila has led to the elucidation of Toll-like receptors (TLRs) in mammalian immunity as pattern recognition receptors. Furthermore, several components of the insect and mammalian signalling pathways are conserved e.g. Drosophila Toll pathway is similar to the MyD88-dependent interleukin 1 receptor-TLR pathway and Imd pathway is similar to the MyD88-independent (adaptor protein TRIF-dependent) branch (CHERRY and SILVERMAN 2006). As mentioned earlier, the signalling mechanisms that activate nodulation, phagocytosis and PO cascade are relatively less understood. Similarly, the signalling responses and effector mechanisms against viral pathogens is poorly understood at this stage. However, recent reports suggest the role of three main pathways in anti-viral defence namely 1. RNA interference (RNAi) pathway 2. Toll pathway and 3. JAK-STAT pathway.

Although RNAi is widely used as a tool for functional genomics, it originally exists as a possible innate immune mechanism in an organism. RNAi mediates gene silencing through micro RNAs (miRNAs) or small interfering RNAs (siRNAs). In Drosophila siRNAs are generated by the enzyme Dicer-1 while the miRNA precursors are recognized by Dicer-2 (TOMARI and ZAMORE 2005). The small RNAs form a complex with a protein called Argonaute to generate the RNA induced silencing complex (RISC) that directs specific RNA silencing. Recently, it has been demonstrated that RNAi offers protection against two RNA viruses in Drosophila (WANG *et al.* 2006). Furthermore, recent reports by Zambon *et. al.* (ZAMBON *et al.* 2005) suggest that activation of DIF transcription factor by Toll receptors is needed for protection against Drosophila X virus (DXV) infection in Drosophila. However

the intermediate components of Toll pathway are dispensable suggesting that this pathway involves Toll receptor and leads to activation of DIF but via a signalling pathway that is different from the classical Toll pathway. Yet another pathway, the Janus kinase-signal transducer of activation (JAK-STAT) pathway is implicated in anti-viral immunity by Dostert *et. al.* (DOSTERT *et al.* 2005) in their studies on Drosophila C virus (DCV) infections in Drosophila. JAK-STAT pathway is a key cytokine receptor signalling pathway in mammals (MURRAY 2007). Dostert *et. al.* identified that a STAT transcription factor which is activated by JAK kinase is required in the responses against DCV infection in the fruit fly. The Toll mediated DIF activation is not essential for protection against DCV infections in Drosophila unlike DXV infections. In the absence of the JAK-STAT pathway, susceptibility to the virus is increased and increased viral loads observed in the adult flies. These recent reports suggest involvement of the above-mentioned pathways in anti-viral immunity. Nevertheless, these are preliminary reports and the exact nature of effector genes activated by these signalling pathways remains to be elucidated.

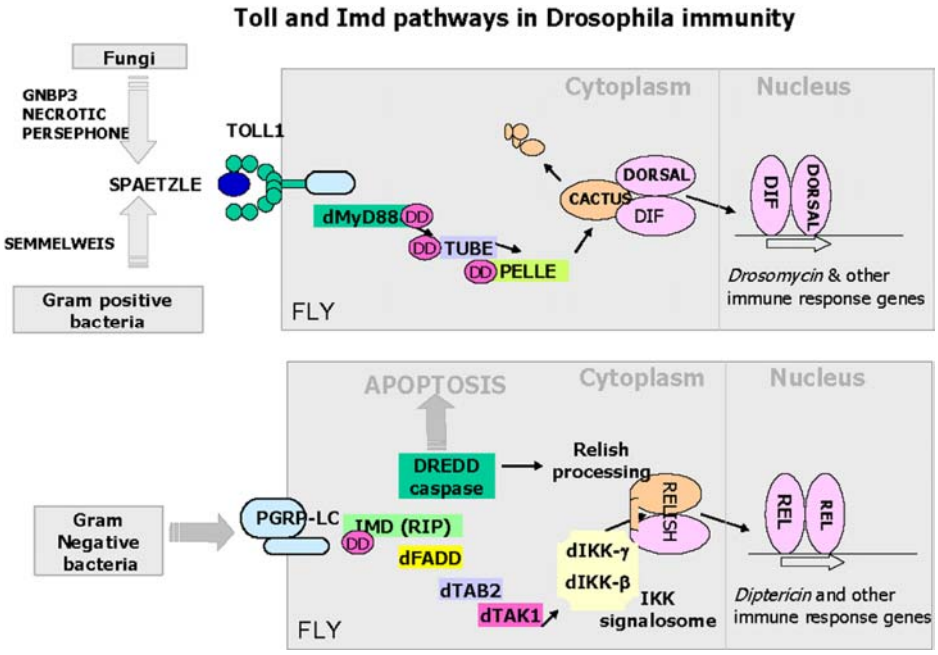

Figure 1.6 Toll and Imd pathways of anti-microbial peptide gene activation in Drosophila. The transmembrane receptor, Toll is activated upon binding to processed (active) Spatzle. Spatzle is processed by a serine protease cascade. Toll activation mediates intracellular signaling through cytoplasmic proteins Tube and Pelle, leading to the degradation of the Cactus and nuclear localization of NF-κB proteins Dorsal and DIF. These transcription factors bind to promoters of target genes, such as *drosomycin*, activating their transcription. Imd pathway of Drosophila is activated upon binding of diaminopimelic acid type PGN (gram-negative bacteria) to the receptor PGRP-LC. A series of components involved in the intracellular signaling then lead to activation and nuclear translocation of the NF-κB protein for the Imd pathway, Relish. Imd pathway is responsible for synthesis of major AMP genes e.g. diptericin (Adapted from Govind *et. al.*, PLOS Biology, 2004).

1.2.4 Immune responses in lepidopteran insects

The overview of the insect immunity presented so far comes mostly from studies in the dipteran insect *D. melanogaster*. However the insect immune mechanism has been also widely investigated in lepidopteran insects and they are next best studied after dipterans amongst all the insect orders. The large body size of most of the lepidopteran insects is especially advantageous in studies of humoral responses as large volumes of haemolymph (1-5 ml/last instar larva) can be easily obtained. In fact the earliest breakthroughs in insect immunity have come from studies in the lepidopteran insects, *Hyalophora cecropia* and *S. c. ricini* (BOMAN *et al.* 1974). Although genetic studies are not viable in lepidopteran insects, the biochemical studies are well suited due to large size of this insects enabling availability of high amount of tissue samples. In the past decade several immune genes have been identified from insects like *Hyphantria cunea, M. sexta, B. mori, A. mylitta, Trichoplusia ni* etc. by large scale gene expression analyses (GANDHE *et al.* 2006b; KANG *et al.* 1996; MITA *et al.* 2003; SHIN *et al.* 1998; ZHU *et al.* 2003). Several AMPs, recognition proteins and components of PO cascade have been identified in lepidopteran insects.

a. AMPs

Cecropin, the first anti-microbial peptide to be sequenced in insects was isolated from *H. cecropia* (STEINER *et al.* 1981). The AMPs isolated from different insect species include lysozymes, attacins (BOMAN *et al.* 1985), lebocins (HARA and YAMAKAWA 1995b), moricins (HARA and YAMAKAWA 1995a) and gloverins (AXEN *et al.* 1997). Lebocins, moricins and gloverins are lepidopteran specific antibacterial proteins. Anti-fungal peptide gallerimycin that bears similarity with drosomycin, an anti-fungal peptide from *D. melanogaster* has been identified from *Galleria mellonella* (SCHUHMANN *et al.* 2003). Similar peptide has also been isolated from *H. virescens* (LAMBERTY *et al.* 2001).

b. Pattern recognition proteins

Discovery of PGRPs in both insects and mammals came from their identification in insect *T. ni* as a pattern recognition protein that binds PGN (KANG *et al.* 1998). Thereafter several PGRP family proteins were identified from *B. mori* (YOSHIDA *et al.* 1996), *M. sexta* (ZHU *et*

al. 2003) etc. In addition, a protein composed of four immunoglobulin domains termed hemolin is a pattern recognition protein specific to lepidopteran insects (SUN *et al.* 1990). Hemolin binds to surfaces of haemocytes and bacterial surface components like LPS and LTA. Recent reports highlight importance of hemolin in mediating the cellular responses of nodulation and phagocytosis (ELEFTHERIANOS *et al.* 2007). Other pattern recognition proteins identified in insects include βGRPs and lectins. Both βGRPs and lectins have been demonstrated to activate the PO cascade upon binding to microbial ligands.

c. PO cascade

The various enzymes involved in the PO cascade have been extensively investigated in *M. sexta* (KANOST *et al.* 2004). The proPO, which is a zymogen form of PO, has been isolated and so are the proPO activating proteinases (PAPs). *M. sexta* PAPs are clip domain proteinases which contain at least one clip domain at the N-terminal and a proteinase domain at the C-terminal end (JIANG *et al.* 1998). The clip-domain is probably involved in the regulation of the enzyme or in modulation of the protease activity against various substrates. It was also shown that PAPs require an additional protein cofactor to generate active PO (KANOST *et al.* 2004). Insect haemolymph also possesses a complex of serine protease inhibitors (serpins). In *M. sexta* 12 isoforms of serpins have been reported (JIANG and KANOST 1997). Serpins differ in their inhibitory activity towards different PAPs suggesting specificity in terms of their action. Serpins probably function in regulating the PO cascade and prevent its auto activation in the absence of infection. They probably also localize the PO reaction at the site of infection and prevent a systemic activation. A model for proposed PO cascade in *M. sexta* has been shown in Figure 1.7.

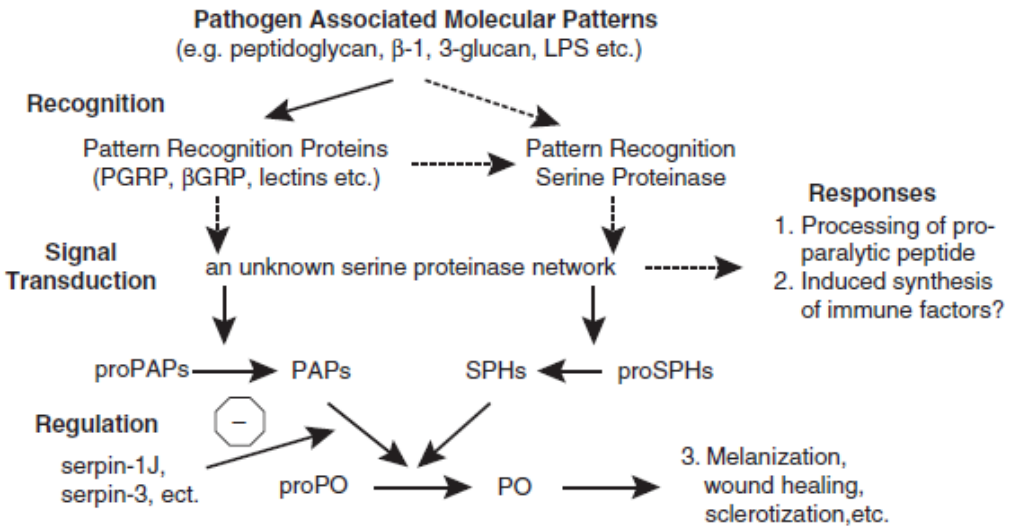

Figure 1.7 A model for *Manduca sexta* PO cascade. The dashed arrows indicate steps not established experimentally (Adapted from Kanost *et. al.*, Immunological Reviews, 2004).

Chapter-2

Analysis of immune transcriptome of Indian tasar wild silkmoth *Antheraea mylitta* to identify potential immune genes

2.1 Introduction

Insect immunity is well studied in dipterans such as fruit flies and mosquito species (AGUILAR *et al.* 2005; JOHANSSON *et al.* 2005; VODOVAR *et al.* 2005; XU *et al.* 2005). Substantial information is also available on genes induced on pathogen challenge in a few lepidopteran species that include the domesticated silkmoth, *B. mori* (MITA *et al.* 2003) Cecropia moth, *H. cecropia* (BOMAN *et al.* 1985) and tobacco hornworm, *M. sexta* (ZHU *et al.* 2003) but in these too the immune response pathways employed to combat pathogen infections remain to be fully characterized.

Abundant genetic resources are now available for *B. mori*, with a 9 X shotgun sequence coverage of its genome and more than 100,000 ESTs in dbEST (NCBI) (MITA *et al.* 2004; MITA *et al.* 2003; WANG *et al.* 2005). With reference to insect immunity, the ESTs have been obtained from baculovirus-infected *B. mori* cultured cells and pupae, but no large scale information on bacteria-induced immune genes is as yet available.

In the present study, we have constructed and analysed an immune transcriptome following bacterial challenge of the Indian tasar wild silkmoth, *A. mylitta,* an economically important lepidopteran cultivated in the wild for silk production. Prior information on immune response genes in wild silkmoths is lacking except for a few peripheral studies. Two proteins from *A. mylitta* - a lysozyme protein, 3-D structure of which is elucidated (JAIN *et al.* 2001) and a protease inhibitor have been characterized (SHRIVASTAVA and GHOSH 2003) earlier. We chose to examine the fat body transcriptome since it is a major immune organ in insects, analogous to the mammalian liver. We generated a total of 1412 ESTs, of which 31% could be ascribed to putative immune functions. We also validated the upregulation of a subset of selected genes from the immune transcriptome by semi-quantitative RT-PCR.

2.2 Materials and Methods

2.2.1 Insects, bacterial inoculation and tissue collection

A. mylitta, 5th instar, day 3 larvae were procured from Regional Research Station, Warangal, Andhra Pradesh. Log phase *E. coli* cells (DH5α), washed and resuspended in saline (0.3 M NaCl, 0.005 M KCl), were injected into the haemocoel of the larvae as described earlier (ABRAHAM *et al.* 1995). At 24 hours post infection (hpi), larvae were dissected to isolate fat body, and the tissue was flash frozen in liquid N_2 and then stored at –70°C till further use.

2.2.2 cDNA library construction and generation of ESTs

Total RNA was extracted from the fat body using Trizol reagent (Invitrogen). The complementary DNA synthesis was carried out using Stratagene ZAP-cDNA® synthesis kit following manufacturer's instructions. Directional cDNA library was constructed by cloning of cDNA fragments into pBluescript II SK (+) vector and electroporation into *E. coli* strain DH10B. Insert-containing plasmid clones were sequenced with RV-M primer (5'GAG CGG ATA ACA ATT TCA CAC AGG 3') with the aid of MegaBACE3000 sequencer (Amersham).

2.2.3 EST processing

Raw sequences obtained from the sequence chromatograms were processed using several programs. A cut-off Phred Quality Value of ≥15 was assigned to extract quality sequences from chromatograms. The quality sequences were screened for the presence of vector sequences using 'Cross Match' program (EWING *et al.* 1998). Then masked vector sequences were automatically removed by in-house developed trimming tool. Sequences shorter than 50 bases were removed. The resulting high-quality sequences were assembled into sequence contigs with the TGICL program (www.tigr.org), which initially makes clusters using MegaBLAST, and thereafter make an assembly using CAP3 for each cluster generated in the first step. A cluster is defined as a unique sequence obtained either by multiple alignment of many sequences that are > 95% similar or derived from a single sequence. A cluster containing ≥ 2 ESTs is termed a contig and that containing only one sequence, a singleton.

The unique putative gene sequences obtained by clustering and assembly were annotated by running BLAST against non-redundant (*nr*) protein database of NCBI. Further, BLAST output was parsed to classify the putative gene transcripts into different functional classes.

2.2.4 Gene Ontology (GO)

Based on the GO annotation of the closely related homologues, ESTs were assigned molecular function, biological process and cellular component from the GO database (LEVASHINA *et al.* 1999). GO annotation generates a dynamic controlled vocabulary that can be applied to all organisms, even while knowledge of gene and protein roles in cells is still accumulating and changing. To this end, the Seqdblite FASTA sequence flat file was downloaded from the GO database. By running BLAST against Seqdblite, closest homologue was identified. From BLAST output, molecular functions, biological process and cellular localization were parsed by building an in-house GO database in MySQL from the GO-term-database flat file, downloaded from Gene Ontology Database Downloads (http://www.godatabase.org/dev/). The Perl-DBI was used to interface with MySQL, to extract the parent terms of each individual GO term that were obtained by parsing BLAST output. The output was then represented graphically.

2.2.5 Analysis of unknown proteins

Proteins that did not show any significant hits in NCBI *nr* database or showed similarity to unknown or hypothetical proteins were characterized by additional computational tools.

a. Domain search

For the sequences showing high similarity to hypothetical and/or unknown proteins, and those showing weak similarity to *nr* protein database, domain search was performed using ProDom (BRU *et al.* 2005). Putative function was assigned based on the type of domains found.

b. Signal peptide and transmembrane domain analysis

Presence of transmembrane domains and signal peptide analysis was done on the transcripts not showing any significant hits in NCBI database. The signal peptide analysis was done by SignalP software (http://www.cbs.dtu.dk/services/SignalP/) and trans-membrane domain analysis was done with TMHMM program (http://www.cbs.dtu.dk/services/TMHMM-2.0/).

c. Functional annotation with PSI-BLAST

The functional annotation of four novel immune upregulated transcripts (DFPs) was done using PSI-BLAST (ALTSCHUL *et al.* 1997).

2.2.6 Expression profile

A. mylitta larvae were differentially challenged- a) Unchallenged, b) Saline-injected, c) *Micrococcus luteus* and d) *E. coli*. Log phase *E. coli* and *M. luteus* bacteria, washed and resuspended in insect saline (0.3 M NaCl, 0.005 M KCl), were injected (30 µl) into a set (n = 4) of 5^{th} instar, day 3, larvae. One set each of saline-injected and uninjected larvae were kept as a control. Four tissues, namely fat body, epidermis, mid gut, and silk gland were dissected out and flash frozen in liquid nitrogen. Total RNA was isolated using Trizol reagent. To remove genomic DNA contamination, total RNA was treated with RNAse free DNAse (NEB) as prescribed by the manufacturer. cDNA was synthesized using MMLV reverse transcriptase (Invitrogen) and oligo dT primers from 1 µg of total RNA. The primers were designed for the selected ESTs by Primer3 software (ROZEN and SKALETSKY 2000).

Semi-quantitative RT-PCR was carried out for all the four differentially challenged tissues using an Eppendorf master cycler under the following conditions- 94°C, 2 min- initial denaturation, 27 cycles (94°C - 30 s, 58°C- 30s, 72°C-2 mins) and a final elongation at 72°C for 10 mins. Actin cDNA was amplified as an endogenous control. PCR reaction components included: 1X buffer, 100µM dNTPs, 1.5 mM MgCl2, 0.5 units Taq polymerase (MBI), 0.5 µM primers. Primer sequences are enlisted in Table 2.7.

2.2.7 Obtaining full-length cDNA by 5′ RACE

We had obtained full-length coding sequences of the DFP-1, 2 and 3 through EST sequencing. To acquire full-length DFP-4 cDNA, we carried out 5′ RACE PCR using the 5′ RACE kit (Clontech). The 5′ ends were amplified by using an adaptor primer and a reverse gene specific primer. PCR was performed for 25 cycles in an Eppendorf master cycler. A 300 bp band was isolated, sequenced and confirmed to be the 5′ DFP-4 sequences.

2.3 Results

As described in Materials and Methods, we constructed a cDNA library from fat body tissues of *E. coli*-challenged *A. mylitta* larvae and randomly sequenced a large number of inserts from the library. By running TGICL program, we obtained 719 clusters from a total of 1412 ESTs, of which 166 were contigs (comprising 859 ESTs) and 553 were singletons. The majority of the EST clusters were 500 to 600 bp, with an average of 524 bp and a maximum of 1994 bp (Figure 2.1). The 1412 EST sequences (accession numbers EB742119- EB743530) can be accessed at the NCBI EST sequence database, dbEST and the 719 clusters can be accessed at URL - http://210.212.212.7:9999/PHP/SILKSAT/uniqueseqs/.

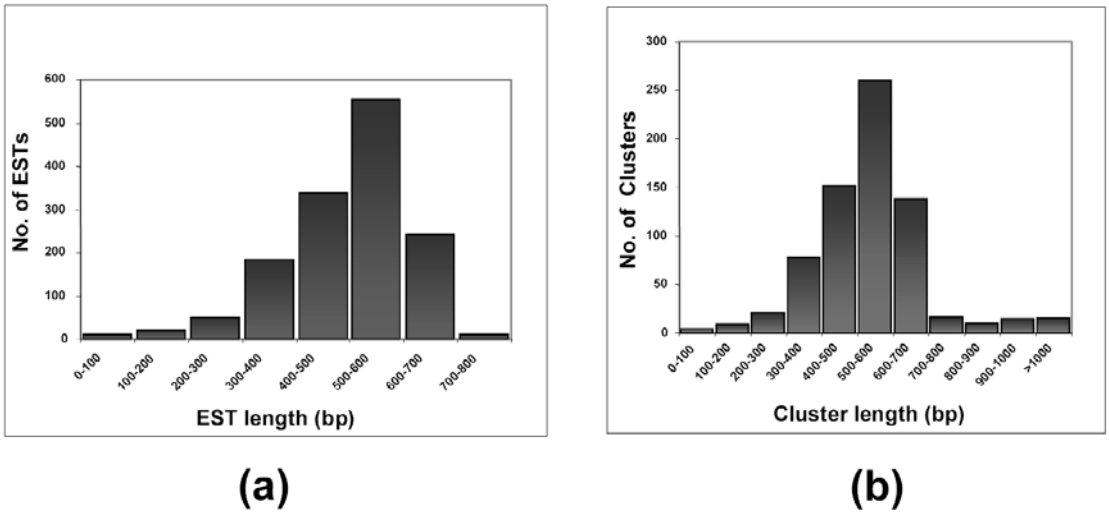

Figure 2.1 Distribution of read lengths of *Antheraea mylitta* ESTs and clusters. (a) Read lengths of 1412 ESTs. (b) Read lengths of clusters

2.3.1 BLAST analysis

The different ESTs were classified into categories such as immune-related, housekeeping, hypothetical insect proteins, hypothetical non-insect proteins based on the homology in NCBI protein BLAST (Table 2.1). Of the 1412 ESTs, 432 (31%) showed similarity to known or putative insect immune proteins. A total of 569 ESTs (39%) were homologous to housekeeping genes and proteins involved in functions other than immunity. Based on the sequence similarity, 679 of the 1412 ESTs were classified as insect-specific (224 of the 719 clusters) and 409 ESTs (204 clusters) were common to both insects and mammals (Figure 2.2 a and 2.2 b). The number of ESTs showing similarity to only the mammalian proteins was 14 and those showing similarity neither to insects nor to mammals were as few as 13.

Table 2.1 Classification of EST functional class categories based on similarity searches with NCBI protein database. Percentage value of each category is in parentheses. *- Hits with relatively less similarity (E-value>1e - 05).

Category	No. of ESTs
Total no. of ESTs	1412 (100%)
Immune-related	432 (31%)
Housekeeping genes and other genes not involved in immunity	569 (39%)
Hypothetical/unknown insect proteins	135 (9.5%)
Hypothetical/unknown non-insect proteins	50 (3.5%)
Weak homology in NCBI*	170 (12%)
No homology in NCBI	56 (3.9%)

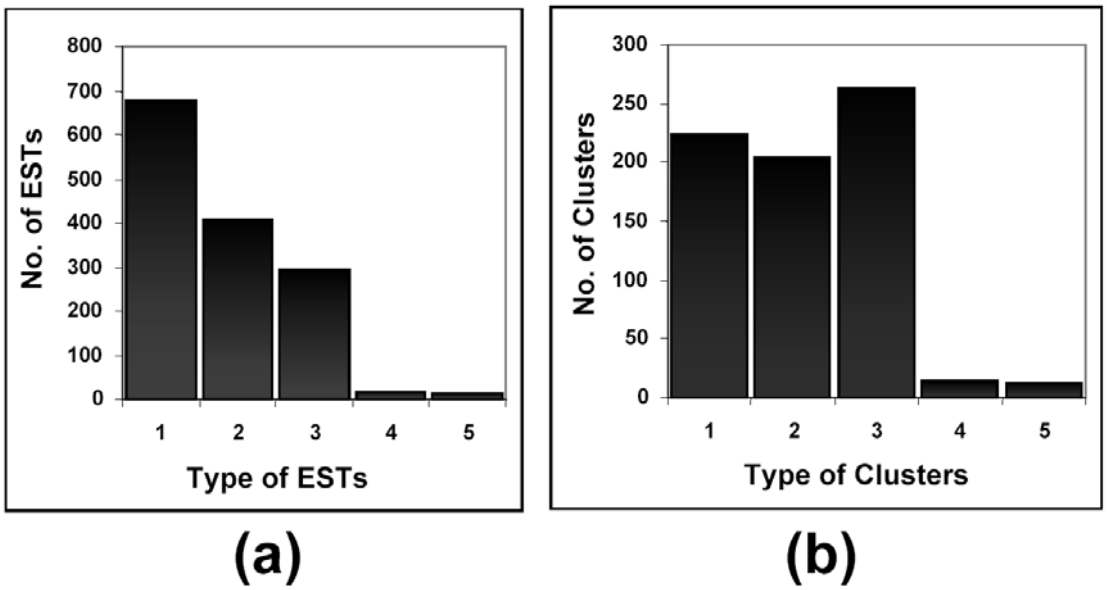

Figure 2.2 Classification of ESTs (a) and Clusters (b) based on the number of sequences showing similarity to insect and or mammalian proteins in NCBI protein nr database. X-axis: 1) Insect, 2) Insect and Mammal, 3) Unknown, 4) Mammal, 5) Neither Insect nor Mammal, Y-axis: Number of sequences.

2.3.2 Gene Ontology

All the ESTs were assigned biological processes, molecular functions and cellular components using Gene Ontology (GO) database. The closest annotated homologue in the GO database was used for assigning these categories. The results of the GO annotation are graphically represented in the Figure 2.3. Based on cellular localization results, most of the gene products were found to be localized in cell (45.4%). In the cell, gene products were most abundant in the intracellular region (66%). A significant portion of proteins was found to localize in the extracellular region (7.7%), revealing that these are likely to be secretory proteins. A majority of proteins from the immune transcriptome was assigned to have the 'binding' function (property of binding macro-molecules) and 67% of these were found to be protein binding. Protein-protein interactions form an integral part of various pathways in immune mechanisms. The next abundant molecular function encountered in the transcriptome was catalytic activity (21.5%). Enzymes also play a major role in the immune system as components of signalling pathways, phenoloxidase cascade, anti-microbial

compounds (lysozyme) etc. In case of biological process, majority of the ESTs belonged to the category of physiological processes (36%). The next abundant category was cellular processes (35%), in which cellular physiological process was about 71%. The abundance of the categories determined by GO highlighted the importance of these processes in immune function.

2.3.3 Domain search

The sequences, which could not be assigned any function based on homology search in NCBI, were searched for conserved domains in ProDom database. Of the 260 clusters that had no matches with known proteins in NCBI BLAST, we could assign protein domain families to 196 of them based on the ProDom search (Table 2.2). The remaining sequences did not show any hits in ProDom and should be further analysed by other specialized computational tools.

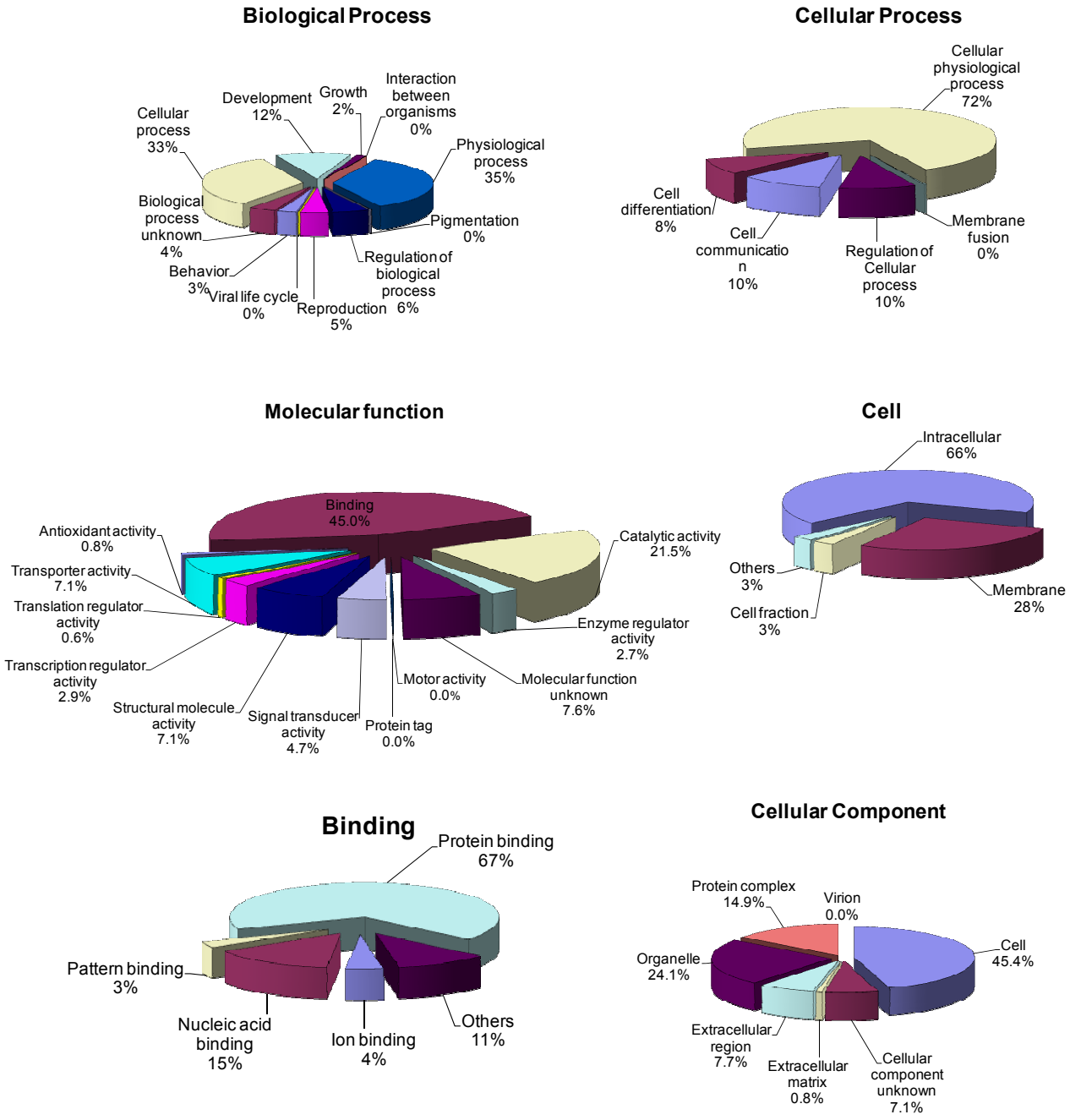

Figure 2.3 Gene Ontology (GO). Annotated GO terms of *Antheraea mylitta* immune challenged fat body EST library. ESTs were classified into different categories based on molecular function, cellular localization and biological process.

Table 2.2 ProDom based functional annotation of clusters, which had no hits, or showed similarity with unknown/hypothetical proteins from NCBI database.
Sequences can be accessed at http://210.212.212.7:9999/PHP/SILKSAT/uniueseqs/.

S.no	Contigs/Singletons	Prodom Domains	Prodom Acc. No
1	amfbcontig0019	No hits	
2	amfbcontig0035	No hits	
3	amfbcontig0038	No hits	
4	amfbcontig0043	Ribosome associated membrane RAMP4	PD054374
5	amfbcontig0051	Phosphoribosyltransferase	PD001439
6	amfbcontig0060	CG14984-A	PD306867
7	amfbcontig0061	No hits	
8	amfbcontig0066	Glycoside hydrolase 29 (alpha-L-fucosidase)	PD886685
9	amfbcontig0070	Aminotransferase class-III	PD000493
10	amfbcontig0072, amfb0294	CG14567 *Drosophila melanogaster*	PD308577
11	amfbcontig0077	CG16926 *Drosophila melanogaster*	PD927175
12	amfbcontig0091	Repressor E1a-stimulated genes	PD152511
13	amfbcontig0092	No hits	
14	amfbcontig0093, amfbcontig0104	Ubiquitin-conjugating enzyme, E2	PD000461
15	amfbcontig0094	Tubulin binding cofactor A	PD010430
16	amfbcontig0095, amfb0124	AICARFT/IMPCHase bienzyme	PD004666
17	amfbcontig0096	No hits	
18	amfbcontig0099	Transmembrane precursor signal TMP21 P23	PD035860
19	amfbcontig0101	Major royal jelly protein	PD014382
20	amfbcontig0108	DEAD/DEAH box helicase	PD000033
21	amfbcontig0111	3-hydroxyacyl-CoA dehydrogenase, NAD-binding	PD003524
22	amfbcontig0115	PF08_0013	PD813653
23	amfbcontig0116	Initiation factor subunit	PD444660
24	amfbcontig0121	Actin binding factor	PD129211
25	amfbcontig0122	No hits	
26	amfbcontig0130	Nudix bifunctional pyrrolidone carboxyl	PD828714
27	amfbcontig0132	Hydrolase Nitrilase Carbon-Nitrogen Amidohydrolase Family	PD052217
28	amfbcontig0134	L27, homolog LIN-7 VELI LIN MUS C	PD154569
29	amfbcontig0136, amfb0145, amfb0354, amfb0391	CHCC Zinc –finger domain	PD690711, PD321788

30	amfbcontig0140	Aldehyde cytosolic dehydrogenase	PDA1A9N1
31	amfbcontig0144	Cysteine Twelve *M. sexta*	PD548193
32	amfbcontig0146	Whey acidic protein core region	PD026912
33	amfbcontig0147	Calcium-binding EF-hand	PD000012
34	amfbcontig0149	Kinase Transferase ATP-binding serine/threonine-protein	PD119917
35	amfbcontig0153,amfb1024	RNA-directed RNA polymerase, P3D protein (polyprotein coat core contains: genome vp3 vp1 p1a vp4 vp2)	PD686877
36	amfbcontig0155	Synthetase metal-binding	PD606972
37	amfbcontig0156	Purine phosphorylase 2	PD006911
38	amfbcontig0157	Histone H3	PD000978
39	amfbcontig0158	No hits	
40	amfbcontig0160, amfb1194,	Factor LPS –induced enriched library endosome	PD015270
41	amfbcontig0161	Helicase ATP-binding hydrolase RNA ATP-dependent	PD035460
42	amfbcontig0162	ADP-specific phospho fructokinase/glucokinase conserved region	PD257594
43	amfbcontig0166, amfb0411	ABC transporter	PD000006
44	amfb0026	No hits	
45	amfb0034	Carbohydrate kinase ,FGGY	PD684127
46	amfb0039	No hits	
47	amfb0047	CG15211-PA RE01453P	PD306644
48	amfb0061	No hits	
49	amfb0065	Synaptobrevin	PD001229
50	amfb0067	GTP-binding elongation factor	PD011419
51	amfb0068	No hits	
52	amfb0081	Peptidase C12, ubiquitin carboxyl-terminal hydrolase 1	PD350662
53	amfb0082	Fasciculation Elongation Zeta Zygin I Coiled Coil	PD011714
54	amfb0095	WD-40 repeat	PD320067
55	amfb0101	Carboxyl esterase activity	PD000713
56	amfb0106	Intra cellular protein transport	PD242575
57	amfb0128	No hits	
58	amfb0135	CG7221-PA	PD622051
59	amfb0144	Receptor 20E ecdysone	PD065649
60	amfb0152, amfb424, afmb0661	ATP dependent RNA helicase	PD035498
61	amfb0168	Aldo/keto reductase	PD000288
62	amfb0173	RNA binding nuclear ribonuclear protein	PD150747
63	amfb0175	Synapse-Associated Sap47	PD022947
64	amfb0178	No hits	
65	amfb0180	Heme Peroxidase	PD564465

66	amfb0185	No hits	
67	amfb0195, amfb0679	NADH-Ubiquinone(oxidoreductase ubiquinone)	PD072136
68	amfb0197	NAD-dependent epimerase /dehydratase	PD291388
69	amfb0199	No hits	
70	amfb0203, amfb1283, amfb0022, amfb1002	RNA-binding region RNP-1 (RNA recognition motif)	PD706470
71	amfb0205	CCR-4 associated factor	PD878878
72	amfb0206	1-(5-Phosphoribosyl)-5-amino-4-imidazole-carboxylate (AIR) carboxylase	PD002193
73	amfb0213, amfb0426, amfb1367	Ras GTPase	PD000015,
74	amfb0214	Flavoprotein Oxidoreductase Dehydrogenase	PD115662
75	amfb0221, amfb0879	beta-3 karyopherin repeat	PD606523
76	amfb0229, amfb1150	Cytochrome b5	PD000612
77	amfb0233	No hits	
78	amfb0241	Kinase receptor HTK32 transferase	PD323132
79	amfb0242	Glutamate-cysteine ligase catalytic subunit	PD235481
80	amfb0244	UPF3 HUPF3A Homolog	PD121682
81	amfb0250	Reticulon	PD186266
82	amfb0263	Short-chain dehydrogenase/reductase SDR	PD246641
83	amfb0276	Glutelin	PD192335
84	amfb0278	Chromosome III B0361.6,(protein of unknown function DUF171)	PD111934
85	amfb0280	Salivary MYS2 precursor signal	PD997477
86	amfb0289	Heavy chain inhibitor inter-alpha-trypsin	PD017272
87	amfb0294	CG33290-PA CG14567-PA	PD308577
88	amfb0296	RNA-binding nuclear ribonucleoprotein binding factor	PD150747
89	amfb0302	CG12489-PA LD18186P	PDA0W1F6
90	amfb0307, amfb0646	ATP-binding MRP homolog nucleotide-binding binding family ATPase chromosome MRP/NBP35 protein	PD092650
91	amfb0311	CG13353-PA	PD315777
92	amfb0312	CG1943-PB CG1943-PA	PD291962
93	amfb0315	DNA Polymerase	PD005357
94	amfb0323	tRNA pseudouridine synthase	PD118232
95	amfb0342	Dehydrogenase 3-hydroxy acyl Co A	PD466280
96	amfb0349	No hits	
97	amfb0353	ATP epsilon synthase	PD026860
98	amfb0372	Oxidoreductase flavoprotein FAD dehydrogenase Acyl-COA	PD000558
99	amfb0374	No hits found	

100	amfb0382	No hits	
101	amfb0387	Myb, DNA-binding(nuclear factor transcription MYB-1 box binding)	PD021927
102	amfb0389	No hits	
103	amfb0394	TDET1 MUS ZGC	PD317441
104	amfb0396	WW/Rsp5/WWP	PD354471
105	amfb0397	No hits	
106	amfb0398	EGF-like immunoglobin	PD000227
107	amfb0401	Eukaryotic initiation factor EIF 3E,	PD336869
108	amfb0411	ABC transporter	PD896973
109	amfb0414	CG14147-PA RE16005P	PD302217
110	amfb0422	Aldose 1 –epimerase	PD127437
111	amfb0423	Luciferase 4-Monooxygenase	PD870069
112	amfb0430, amfb0108, amfb157	RNA binding region RNP-1	PD743635
113	amfb0433	Epoxide hydrolase	PD956892
114	amfb0434	EGF-Like Tolloid Metalloprotease	PD719225
115	amfb0464	CG15092-PA	PD491853
116	amfb0481	Sericin	PD018391
117	amfb0490	No hits	
118	amfb0492	General substrate transporter	PD602677
119	amfb0494	Mitochondrial ATP synthase g subunit	PD015502
120	amfb0501	CG15678-PA SD16010P	PD309707
121	amfb0504	TPR repeat	PD000069
122	amfb0515	Ribosomal transit mitochondrial precursor	PD380986
123	amfb0527	No hits	
124	amfb0533	No hits	
125	amfb0537	Mob1/phocein	PD316458
126	amfb0540	No hits	
127	amfb0544	Perilipin	PD312307
128	amfb0562	Nucleotidyl- Transferase	PD000456
129	amfb0565	No hits	
130	amfb0574	SAICAR synthetase	PD656412
131	amfb0576	Diacylglycerol kinase	PD360309
132	amfb0577	Phosphoribosylformylglycinamidine synthase	PD002258
133	amfb0586	UPF0066 YAEB VIRR uncharacterized enriched product (protein of unknown function)	PD006705
134	amfb0589	No hits	
135	amfb0604	Gamma translocon associated subunit	PD103642
136	amfb0611	RNA binding motif protein 8	PD328190
137	amfb0617	No hits	
138	amfb0618, amfb0824	Farnesoic acid O-methyltransferase activity	PD293778
139	amfb0622	No hits	
140	amfb0625, amfb1187	Ferredoxin	PD001472

141	amfb0638, amfb748	Proteasome component region PCI	PD025198
142	amfb0643	Pyruvate Dehydrogenase	PD186280
145	amfb0649	No hits	
146	amfb0653	Eukaryotic initiation factor EIF 3A	PD014437
147	amfb0655	Peptidylprolyl isomerase, FKBP-type	PD000420
148	amfb0660	Ribosomal protein S8E	PD122856
149	amfb0665	No hits	
150	amfb0696	Death-associated CG12384-PA DAP-1	PD068445
151	amfb0699	Group XII secretory phospholipase A2	PD386805
152	amfb0704	No hits	
153	amfb0708	Glycoprotein	PD722012
154	amfb0714	Mnd1	PD110859
155	amfb0722	GTP-CH-I Cyclohydrolase	PD101974
156	amfb0731	Intracellular chloride channel	PD706961
157	amfb0742	No hits	
158	amfb0757	Peptidase, cysteine peptidase active site	PD585084
159	amfb0759	Exonuclease	PD004947
160	amfb0768	CG17233-PA CG17233-PC	PD719922
160	amfb0771	CG11876-PB CG11876-PC	PD807738
161	amfb0777	Acireductone dioxygenase, ARD	PD006219
162	amfb0778	DCP1-like decapping	PD129069
163	amfb0784	Adenosine deaminase/editase	PD003961
164	amfb0789	Surface ookinete antigen PFS28	PD008013
165	amfb0801	Phenylalanine-4-Hydroxylase, Amino acid-binding ACT	PD973163
166	amfb0812	No hits	
167	amfb0817	Ubiquitin	PD000119
168	amfb0821	Immunoglobulin-like	PD328728
169	amfb0856	No hits	
170	amfb0866	TMS membrane protein/tumour differentially expressed protein	PD593607
171	amfb0873	Serine-hydroxymethyl transferase	PD498026
172	amfb0893	No hits	
173	amfb0895	No hits	
174	amfb0896, amfb0480	Aminotransferase, class V	PD661108
175	amfb0907	No hits	
176	amfb0911, amfb1200	Adenosine/AMP deaminase	PD008716
177	amfb0913	No hits	
178	amfb0916	No hits	
179	amfb0923	Sec61_beta	PD721750
180	amfb0938	No hits	
181	amfb0945	CG33290-PA CG14567-PA	PD308577
182	amfb0953	No hits	
183	amfb0971	Acyl-coA-binding protein, ACBP	PD351532
184	amfb0975	No hits	

185	amfb0984	Nascent polypeptide-associated complex NAC	PD009422
186	amfb0994	No hits	
187	amfb1000	No hits	
189	amfb1011	No hits	
190	amfb1012	ATP-binding RNA hydrolase	PD410733
191	amfb1014	Calponin repeat muscle actin-binding smooth multigene Calmodulin-binding family homolog	PD105641
192	amfb1028	No hits	
193	amfb1035	FMTB wall peptidoglycan-anchor cell	PD851216
194	amfb1038	Carboxylesterase, type B	PD000605
195	amfb1039	No hits	
196	amfb1052	Mov34/MPN/PAD-1	PD363422
197	amfb1056	No hits	
198	amfb1087	No hits	
199	amfb1093	Tudor domain containing repeat kinase	PD291106
200	amfb1099	No hits	
201	amfb1123	Lysozyme C Hydrolase	PD000577
202	amfb1134	Heat shock CG6000-PA RH54517P 67B2 CG12279-PA	PD687240
203	amfb1135	TIA-1 homologue	PD978935
204	amfb1149	CG3706-PA SD20854P C05D2.8 EG:BACR7A4.20	PD104297
205	amfb1162	No hits	
206	amfb1167	CG5080-PB LD34147P CG5080-PA	PD324947
207	amfb1170	CG3884-PA CG32633-PA RE21371P ENSANGP00000022546 IB1	PD293778
208	amfb1212	CG6513-PA CG6513-PB	PDA125G1
209	amfb1220	Cys/Met metabolism pyridoxal-phosphate-dependent enzymes	PD590311
210	amfb1254	No hits	
211	amfb1273	No hits	
212	amfb1276	No hits	
213	amfb1281	No hits	
214	amfb1289	Fibronectin, type III	PD360479
215	amfb1296	No hits	
216	amfb1297	Serine spondin repeat	PD218985
217	amfb1298	No hits	
218	amfb1304	No hits	
219	amfb1308	Sin3 associated polypeptide p18	PD016033
220	amfb1312	Leucine-rich repeat	PD606902
221	amfb1317	No hits	
222	amfb1329	Alpha amylase, catalytic region	PD479378
223	amfb1331, amfb0488	Nuclear ATP-Binding	PD023569
224	amfb1332	No hits	
225	amfb1334, amfb0475	14-3-3 like acetylation phophorylation factor	PD000600
226	amfb1355	Phospholipase/Carboxylesterase	PD352911

227	amfb1356	No hits	
228	amfb1363, amfb0541	Alcohol dehydrogenase super , zinc-containing	PD002174, PD506568
229	amfb1366	Hydrolase CDS L-Homocystein S-Adenosyl ORFs (Glycosyl transferase 8)	PD125323
230	amfb1370	No hits	
231	amfb1373	Protein of unknown function DUF1319 (ORF1 I polyprotein 12and CDS DNA) , TPR repeat	PD087208, PD718959
232	amfb1395	Ribosomal peptide MRP-S26	PD423119
233	amfb1402	Integrase,	PD107985

2.3.4 Signal peptide and transmembrane domains

We further screened the unannotated proteins (no hits or hits with hypothetical proteins in NCBI database) for the presence of signal peptide and absence of transmembrane domains. These characteristics of many of the immune proteins have been utilized to identify probable immune-related genes in large-scale transcriptome studies (ROXSTROM-LINDQUIST *et al.* 2004) although the reliability of this criterion to identify the immune-related genes remains to be experimentally tested. In the present study, 25 out of 260 genes tested fulfilled these criteria and could be considered as potential immune-responsive genes. However, since all the 260 clusters checked are not all full-length sequences, it is possible that we may have missed some others amongst them that represent gene products harbouring signal peptide without transmembrane domain, and hence the actual number may be higher.

2.3.5 Putative immune proteins

A total of 80 clusters were assigned a putative immune function based on their similarity with previously characterized immune response genes and their distribution pattern is shown in Figure 2.4. These putative immune proteins were categorized into different functional groups - i. anti-microbial proteins, ii. pattern recognition receptors (PRRs), iii. proteases and protease inhibitors and iv. putative defence proteins with unknown function. A subset of these putative immune genes (thirty-eight) is represented with the details of their GenBank accession numbers, putative functions, E-values (BLAST) and homologues (Table 2.3). The *A. mylitta* fat body transcriptome also consisted of other accessory proteins implicated in immunity such as antioxidants, ferritins, transferrins, apolipophorins and apoptotic pathway components but their details are not described here.

a. ***Putative anti-microbial proteins***

Attacin-like proteins were the most abundantly expressed transcripts in terms of number of ESTs represented in the immune transcriptome (Table 2.4). Six different types of putative attacins accounted for 34% (148 ESTs) of the immune transcripts. A previous study had reported the presence of four types of attacins in *D. melanogaster* (DOSTERT *et al.* 2005) and a similar study in immunized pupae of *H. cecropia* led to the isolation of six closely related attacins (HULTMARK *et al.* 1983). Second in the list of highly expressed genes were cecropin-like proteins (55 ESTs, ~13%). Three types of cecropin homologues were found in

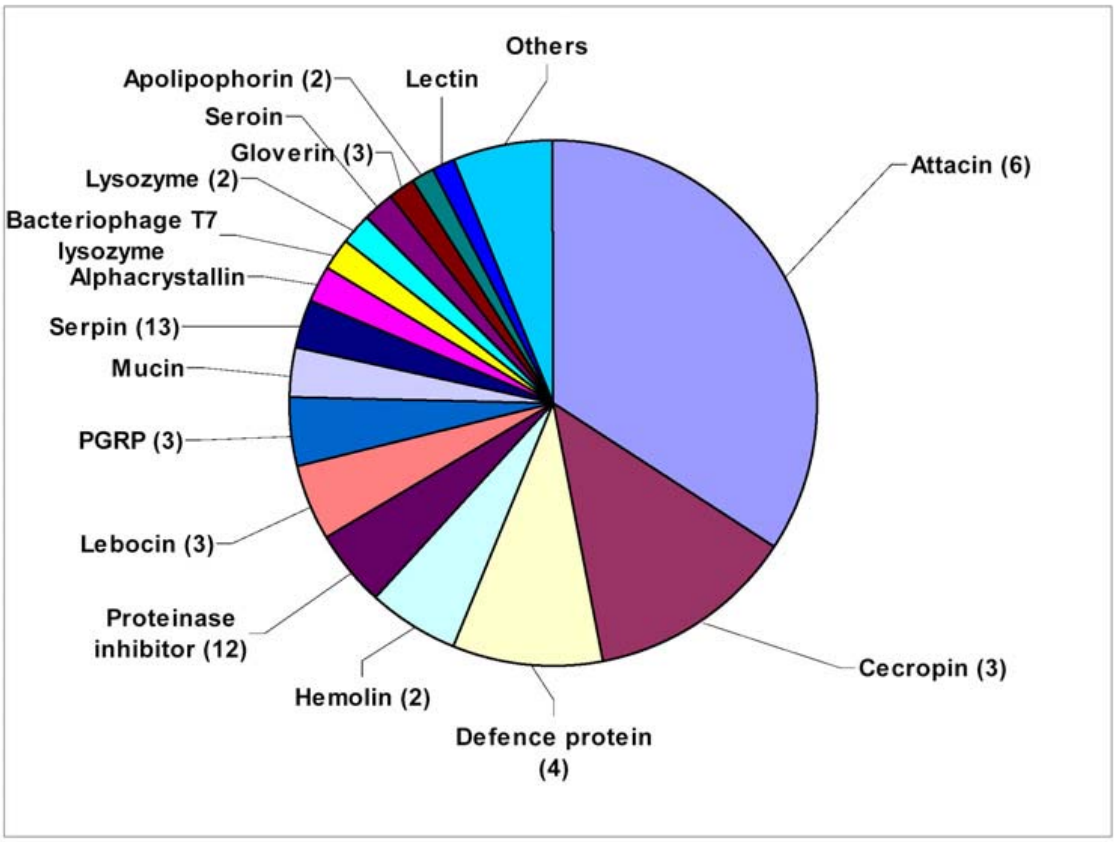

Figure 2.4 Distribution of immune-related transcripts in *Antheraea mylitta* fat body transcriptome. The figure in parentheses indicates the number of isoforms identified in that particular gene family.

A. mylitta transcriptome and a previous study in *H. cecropia* had similarly shown the existence of three types of cecropins (GUDMUNDSSON *et al.* 1991). Three types of putative lysozyme transcripts were detected, one of which was homologous to a bacteriophage T7 lysozyme-like protein reported in *B. mori* that was later named, based on functional analysis, as a peptidoglycan recognition protein (OCHIAI and ASHIDA 1999). The second was homologous to the bacteriolytic lepidopteran lysozymes known to lyse the bacterial cell wall (HULTMARK *et al.* 1980). The third one was a lysozyme-like protein that lacked one or both of catalytic residues essential for muramidase activity. This protein has not been reported in lepidopterans before and is homologous to dipteran lysozyme-like proteins of unknown function. Three types of putative lebocins and gloverins, which are the lepidopteran specific anti-bacterial genes, were also observed in *A. mylitta* transcriptome. Another interesting finding with respect to immune response was the presence of a protein identical to seroin.

This protein is reported to express in the silk glands of *B. mori* and is known to protect the cocoon from microbes (NIRMALA *et al.* 2001). A homologue of anti-fungal protein, gallerimycin, previously characterized from *G. mellonella* and shown to be induced by LPS injection (SCHUHMANN *et al.* 2003) was also found.

b. ***Putative pattern recognition receptors (PRRs)***

PRRs in the insects bind to and detect pathogen associated molecular patterns (PAMPs) like lipopolysaccharide, peptidoglycan, β 1-3 glucan, lipotechoic acid etc (HOFFMANN *et al.* 1999). The putative PRRs identified in the present immune transcriptome analysis were hemolins, peptidoglycan recognition proteins (PGRPs), gram-negative binding proteins (GNBPs), lectins and mucins. *A. mylitta* fat body transcriptome revealed two different proteins resembling hemolin, an immune inducible protein implicated in insect immunity (SUN *et al.* 1990) and it would be worthwhile to study the function and specificities of the different proteins of hemolin family. Three types of PGRP-like proteins were found in the *A. mylitta* immune repertoire in the present study as compared to *D. melanogaster* where 13 PGRP genes are known to be involved in the activation of the various effector pathways in immunity (LU *et al.* 2006).

Table 2.3 A subset of the putative immune genes from the *Antheraea mylitta* transcriptome. The probable function was assigned based on the homology in NCBI-BLAST.

A.p - *Antheraea pernyi*, S.c.r - *Samia cynthia ricini*, H.c - *Hyalophora cecropia*, H.a - *Helicoverpa armigera*, M.s.- *Manduca sexta*, B.m- *Bombyx mori*, L.o - *Lonomia obliqua*, T.n -*Trichoplusia ni*, P.x.- *Plutella xylostella*, C.s.- *Culicoides sonorensis.*

GenBank accession no.	Nucleotide	E-value	Homology	Putative function
DQ666488	977	5e-103	Basic attacin (A.p)	Anti-microbial
DQ666489	984	6e-76	Attacin (S.c.r)	Anti-microbial
DQ666490	429	3e-30	Basic attacin (A.p)	Anti-microbial
DQ666491	924	e-111	Attacin E precursor (H.c)	Anti-microbial
DQ666492	1088	5e-16	Cecropin (H.a)	Anti-microbial
DQ666493	1642	4e-15	Cecropin D (H.c)	Anti-microbial
DQ666494	429	2e-16	Cecropin D (H.c)	Anti-microbial
DQ666495	593	1e-66	Gloverin (M.s)	Anti-microbial
DQ666496	529	8e-54	Gloverin (M.s)	Anti-microbial
DQ666497	1425	0.0	Hemolin (A.p)	Recognition protein
DQ666498	663	2e-88	Hemolin precursor (H.c)	Recognition protein
DQ666499	1145	2e-54	Lebocin (S.c.r)	Anti-microbial
DQ666500	594	8e-11	Lebocin 4 precursor (B.m)	Anti-microbial
DQ666501	713	6e-63	Hypothetical protein (S.c.r)	Anti-microbial

DQ666502	539	4e-50	Hypothetical protein (S.c.r)	Anti-microbial
DQ666503	522	4e-04	Defense protein 4 (L.o)	Unknown
DQ666504	946	2e-47	Lectin 3 (L.o)	Recognition protein
DQ666505	677	1e-70	Peptidoglycan recognition protein (T.n)	Recognition protein
DQ666506	671	2e-82	Peptidoglycan recognition protein 1A (M.s)	Recognition protein
DQ666507	942	2e-66	Serpin 2 (L.o)	Serine protease inhibitor
DQ666508	399	8e-21	Serpin 2 (L.o)	Serine protease inhibitor
DQ666509	607	1e-70	Serpin 2 (L.o)	Serine protease inhibitor
DQ666510	403	4e-40	Serine protease inhibitor 4 (L.o)	Serine protease inhibitor
DQ666511	461	1e-40	Serpin 3 (L.o)	Serine protease inhibitor
DQ666512	301	9e-11	Serpin 2 (B.m)	Serine protease inhibitor
DQ666513	668	1e-35	Serpin 1 (P.x)	Serine protease inhibitor
DQ666514	556	2e-77	Serpin 6 (M.s))	Serine protease inhibitor
DQ666515	448	6e-16	Protease inhibitor 1 (L.o)	Protease inhibitor
DQ666516	377	5e-12	Protease inhibitor 3 (L.o)	Protease inhibitor
DQ666517	485	6e-16	Protease inhibitor 1 (L.o)	Protease inhibitor
DQ666518	584	2e-17	Protease inhibitor 1 (L.o)	Protease inhibitor
DQ666519	615	5e-06	Protease inhibitor 1 (L.o)	Protease inhibitor
DQ666520	579	2e-28	Protease inhibitor 6 (L.o)	Protease inhibitor

DQ666521	569	2e-28	Protease inhibitor 6 (L.o)	Protease inhibitor
DQ666522	592	3e-07	Thiol protease-like (C.s)	Protease inhibitor
DQ666523	400	3e-08	Silk proteinase inhibitor (B.m)	Protease inhibitor
DQ666524	548	1e-36	Kazal-type proteinase inhibitor (M.s)	Protease inhibitor
DQ666525	940	4e-04	Seroin 2 (B.m)	Antimicrobial

Table 2.4 The abundantly expressed genes in *Antheraea mylitta* transcriptome. The number of ESTs indicates the transcript abundance for each of the category.

Gene definition	No of ESTs
Attacin	96
Cecropin D	31
Defense protein 1	30
Ribosomal protein S2, Attacin	26
Elongation factor 1α	24
Hemolin	23
Attacin	22
Cecropin	17
Ferritin	15
Mucin	13

c. Putative proteases and protease inhibitors

Various proteases and protease inhibitors regulate the diverse immune mechanisms like melanization, phagocytosis and induction of anti-microbial peptides. A homologue of a prophenoloxidase activating protease characterized in *B. mori* and *M. sexta* (JIANG *et al.* 2003) was also found in the *A. mylitta* immune transcriptome. Several other classes of putative proteases like cysteine proteases, serine proteases and metallo-proteases were also identified. As many as thirteen distinct serpins (serine protease inhibitors) -like and twelve potential protease inhibitors were detected in the immune transcriptome. Five different serpins have earlier been identified in *M. sexta* and shown to differ in the induction pattern upon immune challenge (KANOST *et al.* 2004). In the light of these studies, further information on serpins and protease inhibitors from the *A. mylitta* immune transcriptome will add to the understanding of various immune pathways in insects.

d. Putative defence proteins with unknown function

Several novel members of known protein families were identified in the present study. Among the new members, one was a putative lysozyme-like protein described in the previous section. An array of putative proteinase inhibitors and proteases were also found. Many of them are new, and their study would enhance our understanding of the mechanisms of proteolytic cascades in insect innate immunity. A few immunoglobulin (Ig) like molecules were identified by ProDom search (Table 2.2). Ig-like molecules- hemolin (SUN *et al.* 1990) and more recently Dscam have been implicated in insect immunity (DONG *et al.* 2006; WATSON *et al.* 2005) and it would be interesting to evaluate the role of these putative Ig-like molecules in insects. Among the potential immune proteins, we describe below in more detail four putative defence proteins (DFPs), for two of which (DFP-1 and DFP-4) we have confirmed the induction upon *E. coli* infection by semi-quantitative RT-PCR.

i. DFP-1, DFP-2 and DFP-3

These three proteins were grouped together, as they are 70-85% similar to each other (DFP-1&DFP-3=85%, DFP-1&DFP-2=79% and DFP-2&DFP-3=77%). All of them have a signal peptide and appear to be secretory proteins. DFP-1 was abundantly expressed in the immune transcriptome suggesting its possible involvement in immunity. In addition, these three

proteins showed high similarity to immune induced unknown proteins from other lepidopterans like *H. cunea*, *S. c. ricini*, *M. sexta* and *Lonomia obliqua* (BAO *et al.* 2003; SHIN *et al.* 1998; VEIGA *et al.* 2005; ZHU *et al.* 2003) (Figure 2.5), as also with some hypothetical proteins from other insects and vertebrates in the NCBI database (Table 2.5). Analysis by position specific iterative BLAST (PSI-BLAST) revealed similarity of DFP-1, 2 and 3 to the vertebrate extracellular matrix proteins (ECM), stromal cell derived factor receptor-2, spondin and reelin and possessed the common domain termed as 'reeler' (Table 2.6). These ECM proteins are involved in the central nervous system signalling and immune mechanisms like 'signalling' and 'pathogen recognition' (HE *et al.* 2004; ZHANG *et al.* 2001). Stromal cell derived factor/receptor complex has been shown to activate JAK/STAT pathway and mediate the migration and proliferation of haematopoietic cells (ZHANG *et al.* 2001). Recently, mindin a protein belonging to the F-spondin family has been shown to act as a pathogen recognition receptor in mice (HE *et al.* 2004). Also, spondin has been shown to be upregulated in *Drosophila* upon bacterial infection by microarray analysis (DE GREGORIO *et al.* 2001). The similarity of DFP-1, 2 and 3 to molecules involved in immune responses in vertebrates further support the immune-related role of these new proteins.

ii. DFP-4

This protein was particularly intriguing as it showed similarity to cecropin, the insect anti-bacterial peptides in the primary sequence BLAST analysis. The multiple alignment of the various cecropins and DFP-4 is shown in Figure 2.6. Based on the SignalP prediction, DFP-4 is likely to be a non-secretory intracellular protein, unlike cecropins that are secreted into the haemolymph. Cecropins are small 5-6 KD peptides whereas DFP-4 is a 17 KD protein with additional unrelated regions at the N and C-termini. The exact role of DFP-4 is not clear and needs to be investigated.

2.3.6 Expression profile

The transcriptome under study most likely represents a plethora of *E. coli*-induced genes in the fat body of the tasar silkmoth. We validated 15 putative immune response genes by semi-quantitative RT-PCR, and their expression profiles are shown in Figure 2.7. All but two genes, a putative protease inhibitor [GenBank: DQ666519] and a seroin gene [GenBank:

DQ666525] were upregulated upon infection. The two DFPs tested, DFP-1 [GenBank: DQ666501] and DFP-4 [GenBank: DQ666503], were induced by both *E. coli* and *M. luteus*. DFP-1 was highly expressed in all the tissues with a more prominent expression in mid gut whereas DFP-4 was exclusively expressed in fat body (Figure 2.7). We analysed the expression pattern of the aforementioned genes in larval tissues differentially challenged with *E. coli* (gram-negative) or *M. luteus* (gram-positive) and compared with challenged or mock-challenged (saline-injected) tissues as negative controls. Most of the genes analysed were expressed more prominently upon infection with *E. coli* than with *M. luteus* suggesting that there may be differential responses towards different pathogens.

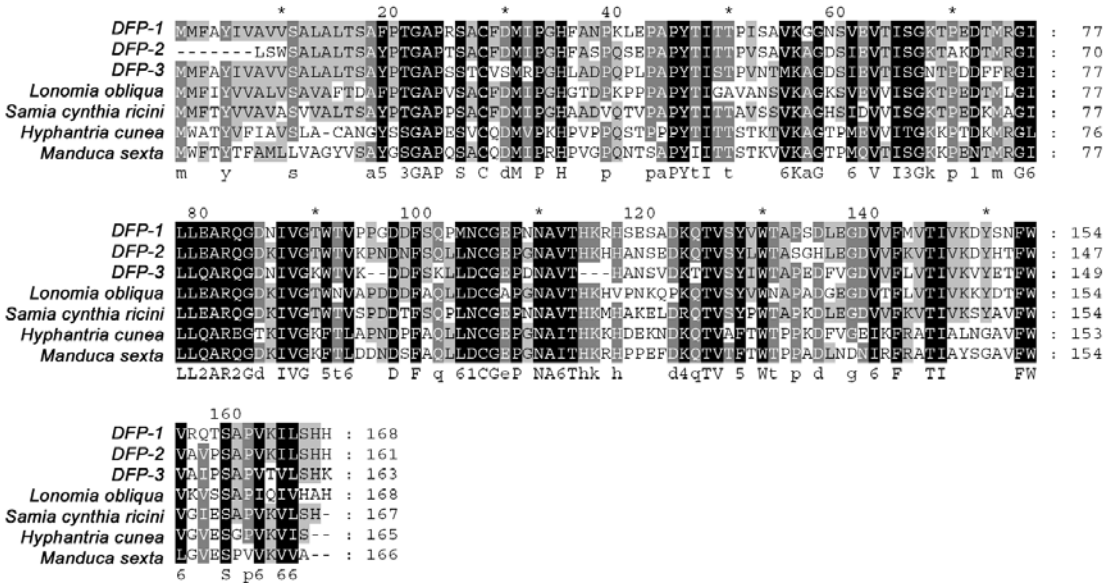

Figure 2.5 Multiple sequence alignment of DFP-1, 2 and 3 amino acid sequences with homologues from other lepidopterans. *Lonomia obliqua* [GenBank: AAV91350], *Manduca sexta* [GenBank: AAO21507], *Samia cynthia ricini* [GenBank: BAD05929], *Hyphantria cunea* [GenBank: AAD09280] protein sequences, respectively. Black and grey shades indicate the identical and similar amino acid residues, respectively.

Table 2.5 Proteins similar to DFPs as determined by protein BLAST in NCBI. The *Bombyx mori* proteins similar to DFPs were obtained by blastp analysis at the website-http://silkworm.genomics.org.cn/. The corresponding gene number has been mentioned. The hits were grouped into 3 categories based on their origin - lepidopteran insects, insects other than Lepidoptera and vertebrates.

Protein	Lepidoptera			Other insects			Vertebrates		
	Organism	Function	Accession No.	Name	Function	Accession No.	Name	Function	Accession No.
DFP-1, DFP-3	*B. mori* *Samia cynthia ricini* *L. obliqua*	Unknown	Bmb041147 BAD05929 AAV91350	*A. gambiae* *D. melanogaster*	Unknown	ENSANGP00000 00354 CG8399-PA	*Pan troglodytes* *Mus musculus*	Predicted-stromal cell receptor factor -2	XP_517470 AAH27770
DFP-2	*B. mori* *Samia cynthia ricini* *L. obliqua*	Unknown	Bmb024557	Same as DFP-1 and 3	Same as DFP-1 and 3	Same as DFP-1 and 3	Same as DFP-1 and 3	Same as DFP-1 and 3	Same as DFP-1 and 3
DFP-4	*L. obliqua* *B. mori* *H. armigera*	Cecropin	AAV91462 BAE53371 AAX51193	None	-	-	No significant similarity	-	-

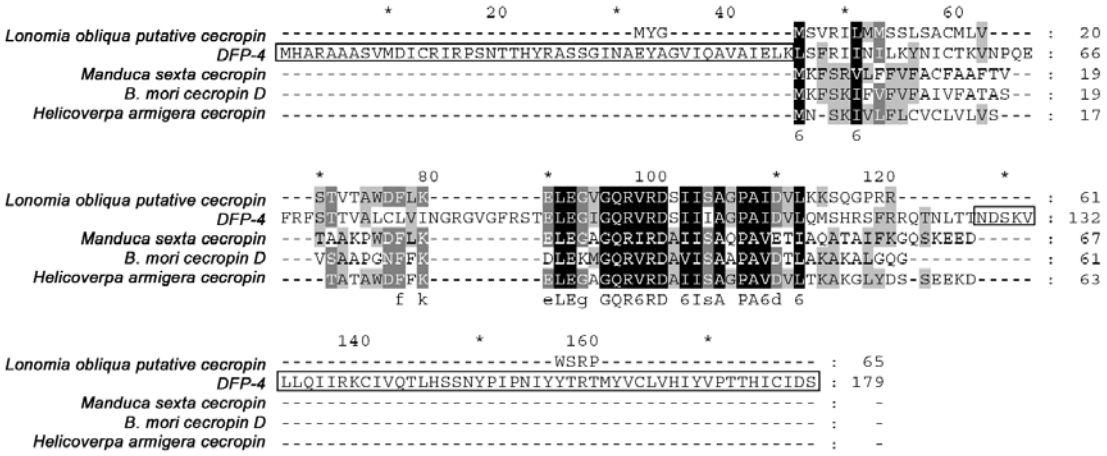

Figure 2.6 Multiple sequence alignment of DFP-4 with cecropins from other lepidopterans. *Lonomia obliqua* putative cecropin [GenBank: AAV91462], *Manduca sexta* cecropin [GenBank: AAO74638], *Bombyx mori* cecropin D [GenBank: BAE53371], *Helicoverpa armigera* cecropin [GenBank: AAX51193.1]. The unique regions of DFP-4 (no homology to cecropins) are boxed.

Table 2.6 Proteins similar to putative defence proteins (DFPs) revealed by PSI-BLAST analysis. The GenBank accession numbers are given in parentheses.

Novel proteins from *A. mylitta*	**Description and function of proteins similar to DFPs.**
DFP-1, DFP-2, DFP-3	1. Stromal cell derived factor receptor 2 homologue *Homo sapiens* [GenBank: NM_001013660.1] (catecholamine catabolism) reeler domain 2. Spondin *Bos taurus* [GenBank: Q9GLX9] *H. sapiens* [GenBank: Q9HCB6] Extracellular matrix, cell adhesion protein that promotes the attachment of spinal cord and sensory neuron cells 3. Reelin precursor *H. sapiens* [GenBank: P78509] Extracellular matrix serine protease. Enzymatic activity is important for the modulation of cell adhesion
DFP-4	**Cecropin D precursor** *L. obliqua* [GenBank: AAV91462]

Table 2.7 Primer sequences used for RT-PCR profile. * GenBank accession numbers are not available for these sequences but the sequences can be accessed at http://210.212.212.7:9999/PHP/SILKSAT/uniqueseqs/.

Gene	GenBank accession number	Forward primer (5´-3´)	Reverse primer (5´-3´)
Attacin1	DQ666490	TTCCACCCTGCGTTAAAGTC	ACAGAATCGGTGCATCAGC
*Attacin 2**	amfb1278	CGGCGAGCTGGTAAAGTATC	CAGAACCGATTGCGCTTAAC
Cecropin1	DQ666492	TCGATGCCTACAGTTTCGAG	CTTCCTGTTCGTTTGCCTTC
Cecropin 2	DQ666493	CTTCCCTTTAGCCAATGCTG	TGTTGTTCGTATTCGCTTGC
DFP-1	DQ666501	TTGTCCGCTGACTCAGAATG	TACAATCAGCGGCAAGACAC
DFP-4	DQ666503	TCGTTCGACAGAACTTGAAGG	TCAGATTTGTTTGCCGTCTG
Gloverin	DQ666495	TGCCTCCTGGATACGAAAAG	TTACCGAATAACCCGTCGTC
Hemolin	DQ666498	CTGACGGAAACCTCAGCTTC	TGGCACTTCTTCATCGACAG
Lebocin 1	DQ666499	AACAGCCATAACGGTTCGTC	TTGGTCGTCGATACGACTTG
Lebocin 2	DQ666500	GAAGGCAAGCTTGAGTGGTC	CCTCAGTGTTGTTCGTGCAG
Lectin	DQ666504	CGTGCGACGATGTATTTGAG	GTCCGAACCGATGAAACTTG
PGRP-1	DQ666506	ACGAAGGATCTGGTTGGTTG	ACGCCGCATCTAAGTAAAGC
*PGRP-2**	amfbc0145	AACCTCAACGTGACCAGGAC	GCAGCACACGTGAGCAATAC
Seroin	DQ666525	TGATGTTCCCTATGCCACAC	ACCGGAGAAATGTTCGTTTG
*Serpin**	amfb0553	TGACTTTGGGTGCAACTGAC	ATGCATATGGTCCCTCGTTC
Proteinase inhibitor	DQ666519	GACCGTTAATGGCTGCAATC	CTTGCCGGTATCATTTCCAG
Actin		GGCATGGGACAGAAGGACT	TAGTGACGATTCCGTGTTCG

2.4 Discussion

The 1412 EST sequences reported in the present study represent 719 novel genes from *A. mylitta* and have tremendously added to the previously available data of a handful number of genes from this organism. Even though the cDNA library was not normalized, a significant

proportion (one-third) of the transcriptome was represented by putative immune-related genes.

Most of the previously known immune gene categories were identified such as AMPs (attacins, cecropins, lysozymes, gloverins), recognition proteins (PGRPs, lectins, hemolins, GNBPs and mucins) and serine proteases and protease inhibitors. The data generated in the present study conformed to the trend of anti-microbial genes being present as multiple gene families as observed in many other insect species and highlighted the essentiality of these genes in the organism (CHENG *et al.* 2006). A novel class of AMPs, which resembled lysozymes but lacked the catalytic amino acid residues essential for muramidase activity, was found. The characterization of these novel AMPs has been described in the next chapter.

Two pathways of differential immune induction have been identified in *Drosophila*. Gram-positive bacteria and fungi induce the Toll pathway whereas gram-negative bacteria evoke the Imd pathway (HEDENGREN-OLCOTT *et al.* 2004) but no protein similar to any of the Toll or Imd pathway components was found in the transcriptome. Apart from Drosophila, the presence of Toll and Imd pathway has been experimentally confirmed in another dipteran insect, Anopheles. Also, orthologues of the two pathways have been found in hymenopteran insect *Apis mellifera* by in-silico analysis. Among the lepidopterans, in *B. mori*, two transcription factors belonging to Rel/NFκB family were shown to be involved in upregulation of AMP genes (TANAKA *et al.* 2005) but no other Toll or Imd pathway component has been identified. The absence of Toll and Imd pathway components in the immune transcriptome suggests three possibilities -1. The lepidopteran Toll and Imd pathway components have diverged considerably from the dipteran/hymenopteran counterparts and hence are not identified by homology searches, 2. The components of the pathway were not identified in this library possibly due to their less abundance with respect to other immune genes or 3. Lepidopteran insects lack Toll and Imd possessing entirely different pathways that have not been discovered yet.

The immune response in insects is dynamic and different effector genes are likely expressed at different time points during infection, contributing to the ability of the insects to ward off infections in spite of the absence of adaptive immunity. The current transcriptome represents genes likely expressed upon *E. coli* infection in the *A. mylitta* fat body at 24 hrs post infection possibly missing out on other genes not active in this period.

The present study has thus increased the repertoire of lepidopteran-specific putative immune response genes by several hundred-fold. This will be a valuable resource for lepidopteran-specific immune studies in particular and insect immune studies in general.

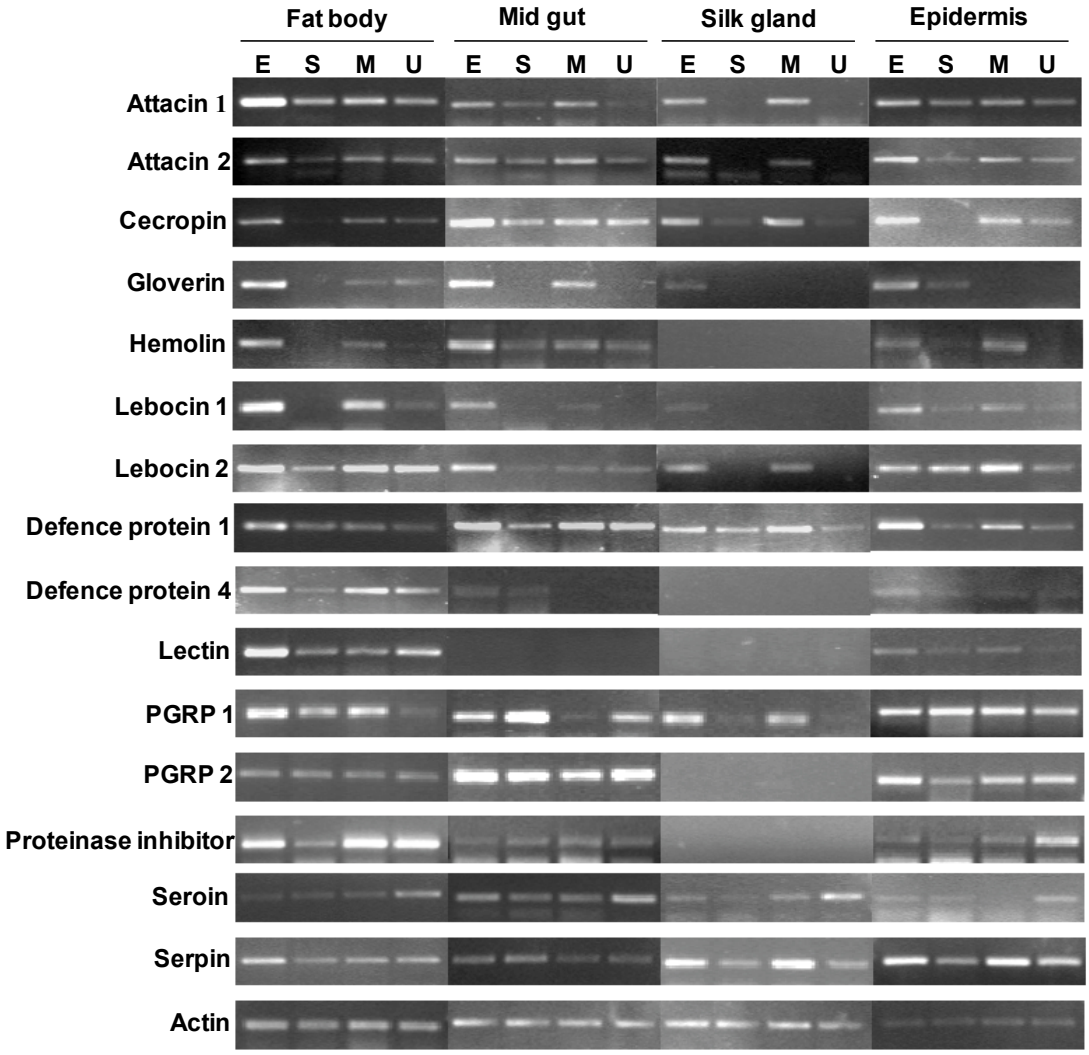

Figure 2.7 RT-PCR profiles of 15 putative immune genes from 4 differentially challenged tissues of 5th instar larvae. E- *Escherichia coli* challenged, M- *Micrococcus luteus* challenged, S- saline-injected, U- unchallenged. Actin was used as an endogenous control.

Chapter-3

Novel anti-bacterial proteins from silkmoths that resemble lysozymes but lack muramidase activity

3.1 Introduction

The immune repertoire in insects comprises mainly of the innate mechanisms, manifested by both cellular and humoral defence mechanisms though recent evidences suggest the presence of adaptive responses as well (DONG *et al.* 2006; SADD and SCHMID-HEMPEL 2006; WATSON *et al.* 2005). The highlight of the humoral defence is a rapid release of different types of AMPs in the insect haemolymph within few hours upon infection. Collectively, these AMPs are known to mount an effective immune defence against invading pathogens (YAMAKAWA and TANAKA 1999). Lysozyme is a widespread AMP occurring in insects, vertebrates, plants and microorganisms. Lysozymes are muramidases that hydrolyse the β-1, 4-glycosidic linkages in the N-acetyl glucosamine and N-acetyl muramic acid residues in the peptidoglycan layer of the bacterial cell and cause their lysis. Lysozymes are upregulated upon infection in the lepidopteran insects unlike their constitutively expressed vertebrate counterparts.

A c-type lysozyme has been previously characterized from the domesticated and wild silkmoth, *B. mori* and *A. mylitta* respectively (JAIN *et al.* 2001; LEE and BREY 1995). In the present study, we report novel AMPs from these two silkworm species. These novel proteins are unique in that they share 50-60 % similarity at the amino acid level with the c-type insect lysozymes but lack characteristic catalytic activity of peptidoglycan hydrolysis exhibited by lysozymes. Hence we chose to designate them as lysozyme-like proteins (LLPs). Nevertheless, we found these LLPs exhibit profound anti-bacterial activity towards a wide range of gram-positive and gram-negative bacteria and are upregulated upon bacterial infection. An immune related function is thereby suggested. Further, the in vivo RNAi mediated knock-down of *B. mori* lysozyme-like protein 1 (BLLP1) resulted in an enhanced bacterial load in haemolymph. We further demonstrated that BLLP1 exhibits bacteriostatic effects against *E. coli* and *M. luteus* and the inhibition was rescued when peptidoglycan was added externally. These results tempt us to hypothesize that the mechanism of inhibition is related to binding of peptidoglycan by BLLP1, leading to growth inhibition rather than hydrolysis of peptidoglycan or membrane permeabilization. To the best of our knowledge this is the first report on natural occurrence of non-bacteriolytic, anti-bacterial lysozyme-like proteins that adds yet another novelty to the already existing diversity of AMPs.

3.2 Materials and Methods

3.2.1 Isolation of *LLP* and *lysozyme* cDNAs

A. mylitta lysozyme (*AL*) cDNA was isolated by designing primers based on the protein sequence reported earlier (JAIN *et al.* 2001) and cDNA clone of *B. mori* lysozyme (LEE and BREY 1995) was a kind gift from Paul Brey. Partial sequence of *A. mylitta* lysozyme-like protein 1 (*ALLP1*) was identified from an *E. coli*-challenged fat-body EST library (GANDHE *et al.* 2006b) and subsequently the full length sequence was obtained by 5′ RACE PCR using the 5′ RACE kit (Clontech). The 5′ end sequence was amplified by using an adaptor primer and a reverse gene specific primer. PCR was performed for 25 cycles on an eppendorf master cycler. A 650 bp band was isolated, sequenced and confirmed to be the 5' *ALLP1* sequence. *BLLP1* was identified from Silkbase (MITA *et al.* 2003) by BLAST search analysis with *ALLP1* as a query. *BLLP2* and *BLLP3* were identified by tBLASTx analysis at SilkDB with *BLLP1* as a query sequence (WANG *et al.* 2005). Primers were designed by Primer-3 software (ROZEN and SKALETSKY 2000) and their sequences were as follows (5′-3′)- ***AL* Forward *(F)*-** AAACGTTTCACCAGATGCG ***AL* Reverse *(R)*-**ACAGTCGCTAATATCTGG
***BL F*-**AAAACGTTCACGAGATGCG ***BL R*-**GCAGCTGCTAATATCAGG
ALLP1 R (for 5′ RACE)-CCTTCGAACTCTTCGCGTTA **Adaptor** (Forward) primer was supplied by Clontech. ***BLLP1 F*-** AAGGTCTTCACGAGATGCCAAC
***BLLP1 R*-** GCATCTGGAGATGTCTGGTAGGTTCTTC. Following PCR conditions were used- 94°, 2 min-initial denaturation, 35 cycles (94°- 30 s, 60°- 30 s, 72°-2 min) and a final elongation at 72° for 10 min. *Actin* cDNA was amplified as an endogenous control. PCR reaction components included: 1X buffer, 100 μM dNTPs, 1.5 mM $MgCl_2$, 0.5 units Taq polymerase (MBI), 0.5 μM primers.

3.2.2 Bacterial strains

BL21-CodonPlus strain (Stratagene) was used for protein expression. For the anti-bacterial assays following strains were procured from Microbial Type Culture Collection (MTCC), Chandigarh, India: *Klebsiella pneumoniae subsp.pneumoniae* (MTCC no. 39), *Serratia marcescens* (MTCC no. 86), *Pseudomonas fluorescens* (MTCC no. 103), *Bacillus thuringiensis subsp. Kurstaki* (MTCC no. 868).

3.2.3 Molecular weight (MW) and isoelectric pH (pI)

The MW and pI of the lysozymes were predicted using the Expasy software. The SignalP software was used to predict the signal peptide (BENDTSEN *et al.* 2004) (www.cbs.dtu.dk/services/SignalP/).

3.2.4 Homology modelling

Homology models of ALLP1 and BLLP1 were constructed with the BL tertiary structure (PDB: 1GD6) as a template. As a first step, the protein sequences were aligned with CLUSTAL X and subsequently the model was generated with MODELLER. The model was validated using PROCHECK and the modelled ALLP1 and BLLP1 structures were superimposed with BL tertiary structure using the Molecular Operating Environment (MOE) software.

3.2.5 Phylogenetic analysis

Phylogenetic analysis was done by comparison of mature full-length protein sequences of 35 insect lysozymes with hen lysozyme (HL) as an outgroup. All the sequences were aligned using ClustalX 1.8 and manually edited using GeneDoc Version 2.6.002. A neighbour-joining (NJ) tree was constructed using HYPHY (POND *et al.* 2005) and was used as a base tree to test the best-fit model of evolution in HYPHY. Akaike Information Criterion (AIC) was used to select the best-fit model, which selected WAG (Whelan and Goldman) model of amino acid substitution (WHELAN and GOLDMAN 2001). A maximum likelihood tree was constructed with the program TREE-PUZZLE 5.2 (SCHMIDT *et al.* 2002) that utilised 50,000 quartet-puzzling steps. The GenBank accession numbers of the sequences used in phylogenetic analysis are mentioned in Table 3.1.

3.2.6 Real time PCR

A. mylitta and *B. mori,* 5th instar, day 3 larvae were challenged with log phase *E. coli* or *M. luteus*. Briefly, log phase *E. coli* (DH5α) or *M. luteus* cells, washed and resuspended in saline

Table 3.1 GenBank accession numbers of insect lysozymes, LLPs and hen lysozyme used in phylogenetic reconstruction

Serial No.	Lysozyme	GenBank accession number
1	*Bombyx mori*	AAB40947
2	*Antheraea assama*	AAS50157
3	*Antheraea mylitta*	Q7SID7
4	*Trichoplusia ni*	P50718
5	*Manduca sexta*	AAB19535
6	*Galleria mellonella*	P82174
7	*Hyphantria cunea*	P50717
8	*Hyalophora cecropia*	P05105
9	*Samia cynthia ricini*	BAB20806
10	*Heliothis virescens*	AAD00078
11	*Spodoptera exigua*	AAP03061
12	*Aedes aegypti (A)*	CAC19819
13	*Aedes aegypti (E)*	AAU09087
14	*Aedes albopictus*	AAV90643
15	*Anopheles stephensi*	AAO74844
16	*Anopheles gambiae c-1*	AAY24699
17	*Anopheles gambiae c-2*	AAT51797
18	*Anopheles gambiae c-3*	AAT51798
19	*Anopheles gambiae c-4*	AAY24700

20	*Anopheles gambiae c-5*	AAY21239
21	*Anopheles gambiae c-7*	AAY21240
22	*Drosophila melanogaster CG11159-PA*	AAF57434
23	*Drosophila melanogaster CG16799-PA*	AAF57435
24	*Drosophila melanogaster CG7798-PA*	AAF58041
25	*Drosophila melanogaster P*	P29615
26	*Drosophila pseudoobscura GA14129-PA*	EAL32771
27	*Drosophila pseudoobscura GA11118-PA*	EAL29964
28	*Drosophila pseudoobscura GA10804-PA*	EAL25565
29	*Drosophila melanogaster CG16756-PA*	AAF46034
30	*Musca domestica 1*	Q7YT16
31	*Musca domestica 2*	AAQ20047
32	*Gallus gallus*	630460A

(0.3 M NaCl, 0.005 M KCl), were injected (10 µl and 30 µl into *B. mori* and *A. mylitta*, respectively) into the haemocoel of a set of larvae (n = 4). A set of larvae, injected with sterile saline (mock infection) and a set of uninjected larvae were used as control. Fat body total RNA was isolated by Trizol method (Invitrogen) from the differentially challenged larvae and it was treated with RNAse free DNAse (MBI) to remove genomic DNA contamination. Subsequently, cDNA was synthesized from 1µg of total RNA with oligo dT primers and MMLV reverse transcriptase (Invitrogen). For quantitative PCR, the components included the cDNA template; primers (0.5 µM) and SYBR Green master mix (Eurogentec). PCR was carried out in ABI Prism 7000 real time PCR cycler under following conditions- 50°C- 2 min, 95°C-10 min and 40 cycles of 95°C-15 s and 60°C-1 min. The fold increase of lysozyme transcripts over the unchallenged levels was calculated by the comparative Ct method (Applied Biosystems). The data obtained was normalised against the endogenous actin control, which was also analysed under similar conditions. The data presented is an

average of two independent reactions. Primers were designed by Primer Express software (ABI) and their sequences (5'-3') were as follows:

AL F-GAGACAAGGCTTCGACGAGAG,

AL R-TTGTTCACTTTACCAACTTTATCCGTAT;

ALLP1 F- CGATGGAAATTGTGCTCTAAAGG, ***ALLP1 R***-TGCATCTGGCCACCCATT;

BL F-GGCTCGAAGGACTACGGATTG, ***BL R***- CCGGACTGGCGCCTTT;

BLLP1 F- AGGGCGGAAATTGCAACAT, ***BLLP1 R***- ATACCCGTTTTGCGCATCTAA

Actin F-GGCATGGGACAGAAGGACT, ***Actin R***-TAGTGACGATTCCGTGTTCG

3.2.7 Tissue expression profile

The spatial gene expression was studied in the tissues of mid gut, epidermis, silk gland and head from differentially challenged *B. mori* and *A. mylitta* 5th instar larvae by semi-quantitative RT-PCR. Four sets of larvae were differentially challenged as described in the previous section. In addition, the expression pattern of the *BL* and *BLLP1* transcripts was determined in various embryonic stages (45, 60, 72, 96 and 120 hour embryos) of *B. mori* but the same analyses could not be carried out in *A. mylitta* due to difficulty in obtaining the samples from the wild silkmoth. Total RNA was isolated from the whole embryos and cDNA was synthesized from 1 μg of RNA as described above. The primers used were- ***AL F***-AAGACCTCGACCCCTGGAAA, ***AL R*** GCGAGTGTTGGCAGTGGTTTA, ***ALLP1 F***-GTAAAGGCGGGAAATGTGACA, ***ALLP1 R***- TCGGGTAGTAGCTGACCTTTG, ***BL F***-TTGTCCTCTGCGTTGGTTCTG, ***BL R***-GTGTTCGTCTTGGACGTCTCA; ***BLLP1 F***-CTTTGATCCCGACTTGGGTA, ***BLLP1 R***- CCAGGTCTTCGTTCAGGAGA, actin was used as an endogenous control in the RT-PCR analysis. PCR was carried out on an eppendorf PCR master cycler under following conditions-94°C, 2 min-initial denaturation, 30 cycles (94°C-30 s, 62°C-30 s, 72°C-2 min) and a final elongation at 72°C for 10 min. PCR reaction components included: 1X buffer, 100 μM dNTPs, 1.5 mM $MgCl_2$, 0.5 units Taq polymerase (MBI), 0.5 μM primers.

3.2.8 Recombinant expression, purification and refolding

AL and *BL* were cloned into pET-23a (+) and *ALLP1* and *BLLP1* were cloned in pET-28a (+) *E. coli* expression vectors (Novagen). The NdeI and XhoI sites were utilised to clone *ALLP1* and *BLLP1* downstream to the N-terminal vector encoded histidine tag containing a thrombin cleavage site for removal of histidine tag. NdeI and NotI sites were used to clone *AL* and *BL* wherein an N-terminal histidine tag was incorporated in the protein by including it in the forward primer. The restriction sites were incorporated in the primers and then the PCR product was digested and ligated to the vector digested with identical restriction enzymes. The clones were confirmed by sequencing, recombinant plasmid was isolated and transformed into BL21-CodonPlus strain (Stratagene) of *E. coli*. The protein expression in BL21-CodonPlus cells was induced by adding 0.8 mM IPTG to a log phase culture. The cells were harvested 6 hours post induction and the protein was purified from the inclusion bodies under denaturing conditions. The N-terminal histidine-tagged recombinant proteins were purified under denaturing conditions (6 M guanidium hydrochloride) by affinity chromatography with Ni-NTA Agarose (Qiagen) using the manufacturer's protocol and the purified protein was incubated with 5 mM DTT at 4°C overnight. The denatured and reduced lysozymes were then refolded to obtain a soluble active protein by dilution with refolding buffer containing 1 mM GSSG (oxidised glutathione) and 0.1 mM GSH (reduced glutathione), 264 mM NaCl, 11 mM KCl, 55 mM Tris-HCl, pH 8.2; 1.1 mM EDTA (LOPEZ-ZAVALA *et al.* 2004). For ALLP1 and BLLP1, 1 mM $CaCl_2$ was added and EDTA was omitted from the refolding buffer (KOSHIBA *et al.* 1999). The resulting protein was concentrated using centricon columns (Millipore) with a 10 KD cut -off membrane, dialysed against Milli Q water overnight. For ALLP1 and BLLP1, dialysis was carried out with 1 mM $CaCl_2$ and 20 mM Tris-HCl, pH-8. The N-terminal histidine tag of ALLP1 and BLLP1 was removed by thrombin (Novagen) digestion according to the manufacturer's protocol and the cleaved protein was purified subsequently by size exclusion chromatography using Superdex-200 (Amersham) columns. Protein estimation was carried out by Bradford's method (BRADFORD 1976) .

3.2.9 CD spectroscopy of ALLP1 and BLLP1

Far UV (194-240nm) CD spectra of ALLP1 and BLLP1 were obtained in 0.1 M ammonium acetate buffer (pH 5.5) and the baseline value was subtracted. The secondary structure was estimated using the inbuilt program in Jasco J-810 spectropolarimeter.

3.2.10 Muramidase assay

Muramidase assay was carried out with *M. luteus* substrate (Sigma) in 50 mM sodium phosphate buffer, pH 6.5 and the absorbance at 450 nm was recorded after every 30 sec.(SHUGAR 1952). 1 unit of lysozyme is defined as the amount that decreases the absorbance of the substrate solution by 0.001 absorbance units /ml/ min.

3.2.11 Peptidoglycan binding assay

Peptidoglycan binding assay was performed as described earlier (KANG *et al.* 1998) with slight modifications. Briefly, *E. coli* and *M. luteus* peptidoglycan (gift from Dr. Bruno Lemaitre, Global Health Institute, EPFL, Lausanne, Switzerland) were washed and resuspended in PBS (40µl) and 10µg of ALLP1 and BLLP1 were added to it. After 30-min incubation, the reaction components were pelleted down, the pellet was washed once with PBS and the supernatant and wash fractions were collected. The pellets were resuspended in 40 µl PBS, boiled for 5 min after addition of 2X-SDS/PAGE loading buffer and analysed by Western blot. Appropriate controls were kept for the experiment. Anti-histidine primary antibody (Qiagen) at 1:2000 dilution and anti-mouse IgG secondary antibody conjugated to horseradish peroxidase (Amersham) at 1:2000 dilution was used. The detection was done with ECL Plus detection reagents (Amersham).

3.2.12 Anti-bacterial assays

a. Radial diffusion assay

The antibacterial activity of the purified proteins was tested in a radial diffusion assay (STEINBERG and LEHRER 1997) . The log phase bacteria were incorporated in an underlay gel

of 1% agarose in citrate phosphate buffer (9 mM sodium phosphate, 1 mM sodium citrate, pH-6.0) supplemented with 3% tryptic soy broth (TSB). Wells of 3 mm diameter were bored into this underlay gel and lysozymes and LLPs were added to each well and allowed to diffuse for 3 hours after which an overlay of 1% agarose in citrate phosphate buffer supplemented with 6%TSB was overlaid on the lower gel and incubated overnight at the ambient temperature. The zone of bacterial inhibition was measured after incubation. The bacterial strains used were-gram-negative: *E. coli* MG1655, *Salmonella typhi*, *K. pneumoniae* and *P. fluorescens* and gram-positive: *S. aureus*, *M. luteus*, and *B. thuringiensis*. The inhibitory concentration (IC) was calculated from the inhibition zone data using a formula (method (a)) valid for inhibition zones smaller than critical diameter as described earlier by Hultmark and colleagues (HULTMARK *et al.* 1982). Method (a) was used in the present study because a curved plot of log n versus d^2 was obtained with the series of concentrations used or in case of a few bacterial species the inhibition zones were obtained only with the highest concentration used in the assay.

$$IC = \frac{0.468\, n}{a\, d^2}$$

Wherein a is the thickness of agar layer (cm), d is the inhibition zone diameter (cm) and n is the concentration in nanomole of the anti-microbial protein utilised for the assay, and IC is expressed in μM.

b. Kinetics of bacterial inhibition

The kinetics and bacteriostatic or bactericidal nature of the anti-bacterial activity of BL and BLLP1 was determined against *E. coli* and *M. luteus* by incubating the cells at a concentration of approximately 10^4-10^5 cells/ml with 10 μM of the peptide in 1 X PBS (pH 6.5). Samples were withdrawn at 1, 2, 4 and 8 hours post incubation with the AMPs and plated onto LB agar to determine the viable colony forming units (cfu/ml). A control in which cells were incubated with equivalent amount of sterile water was also kept in parallel. For PGN competition assay, similar experiment was carried out but BL and BLLP1 were pre-incubated for 15 min with 32 μg of *M. luteus* peptidoglycan (Sigma) and then assayed for kinetics of inhibition against *E. coli* and *M. luteus* as mentioned above.

3.2.13 Inner membrane (IM) permeabilization assay

IM permeabilization of *E. coli* strain MG1655 lacY: Tn10dKan which is a lactose permease deficient strain was estimated by analysing its β-galactosidase activity after incubation with BL and BLLP1 (STEINBERG and LEHRER 1997). The cells were grown in poor LB broth (1% bactotryptone, 0.9%NaCl) in the presence of 0.1 mM IPTG to log phase and 100 µl of bacterial suspensions (10^6cfu/ml) were incubated at 37°C with BL and BLLP1 respectively (10 µM). The samples were withdrawn at 1, 4 and 8 hours post incubation, centrifuged and the supernatants were stored at -20°C. The β-galactosidase activity of the samples was assayed with ONPG as the substrate and the absorbance was read at 420 nm. As a control, cells incubated with sterile water were assayed for β-galactosidase activity.

3.2.14 Preparation of dsRNA and RNAi

BLLP1 cDNA (357 bp) was cloned into pCRII-TOPO vector (Invitrogen) and amplified with M13 forward and reverse primers. This template with flanking T7 and SP6 promoters was utilized for in vitro transcription reaction and sense and antisense RNA strands were generated with the T7 and SP6 Megascript kits (Ambion). The DNA template was removed from the transcripts by DNAse treatment and the RNA products were subsequently purified by Trizol extraction (Invitrogen) followed by isopropanol precipitation. The complementary single stranded RNAs were dissolved in DEPC treated water, combined in equimolar amounts in 1X insect buffer saline (IBS, composition: NaCl –160 mM, KCl- 10 mM, $CaCl_2$-4 mM) and annealed by heating to 95ºC and slow cooling overnight at room temperature. Similarly, dsRNA specific to green fluorescent protein (GFP) was synthesized as a non-specific control. The dsRNA formation was confirmed by agarose gel electrophoresis and the concentration was determined spectrophotometrically. To determine the in vivo role of BLLP1 in immunity, we knocked-down the BLLP1 expression by RNAi and then assessed the effects of knock-down on immune function as described below.

3.2.15 In vivo bacteria clearance assay

The ability of clearance of bacteria injected in larvae systemically was assessed and compared with that obtained from RNAi-mediated BLLP1 knock-down larvae. For bacterial

clearance assay, dsRNA specific for BLLP1 (1 μg per larva) or 1X IBS (control) or GFP-dsRNA (control) was injected to a set of 5 larvae followed by an *E. coli* MG1655 injection (approximately 10^6 cells), 6 hours later. 18-hour post *E. coli*-challenge, the haemolymph was aseptically collected in pre-chilled eppendorf tubes containing phenylthiourea crystals and the haemocytes were pelleted by centrifugation at 2000 rpm. 100 μl of the supernatant plasma was plated on LB agar plates, incubated overnight at 37°C and colony-forming units (cfu/ml) were estimated. As a control, the clearance of bacteria in saline-injected and GFP-dsRNA injected larvae was analysed.

3.3 Results

3.3.1 Isolation of *LLP* and *lysozyme* cDNAs

A full-length cDNA of a lysozyme-like transcript from an immune challenged fat body library of *A. mylitta* was obtained as described in experimental procedures. The deduced amino acid sequence revealed 52% similarity at the amino acid level with *A. mylitta* c-type lysozyme (AL) but there was a substitution at the catalytically essential aspartate-50 residue by a tyrosine residue (Figure 3.1). We designated the new protein as *A. mylitta* lysozyme-like protein 1 (ALLP1). The BLAST search of *B. mori* genomic (WANG *et al.* 2005) and EST database (MITA *et al.* 2003) with *ALLP1* as a query sequence, led to the identification of three additional *B. mori* lysozyme-like transcripts (*BLLPs*) which too lacked both the catalytic amino acid residues. We designated them as *BLLP1*, *2* and *3* respectively.

3.3.2 Comparison of gene and protein structure of LLPs and lysozymes

The gene structure of *BL* and the three *BLLP*s as obtained from the *B. mori* genome sequence and its comparison with hen lysozyme (*HL*) gene is shown in Figure 3.2. Gene structure comparison suggests an overall correlation between *BL*, *BLLPs* and *HL* with reference to location of exons/introns and amino acid residues interrupted at the introns, with the only exception of *BLLP2*, which is encoded by a single exon. The similarity in the gene organization between *BL*, *BLLPs* and *HL* suggests a common origin for lysozymes and LLPs. The amino-acid comparison of LLPs with c-type lysozymes revealed that LLPs lack either one or both of the catalytic amino acids with the exception of BLLP2 (Figure 3.1). However

the substrate binding residues were found to be conserved. Of the 12 substrate binding sites identified in insect lysozymes (JAIN *et al.* 2001), 4/12 were conserved in all the three LLPs-ALLP1, BLLP1, BLLP3 while 7/12 were partially conserved (conserved in either one or two of the three LLPs). Comparison of the MW/pI of the silkworm lysozymes, LLPs and HL is represented in Table 3.2. BLLP2 is unique with respect to other LLPs and resembled an *Anopheles* lysozyme (c-6) and a *Drosophila* lysozyme that comprises of multiple lysozyme domains. A signal peptide was predicted with the SignalP 3.0 software in all the silkworm lysozyme-like proteins suggesting their secretory nature. We generated an in-silico tertiary structure of ALLP1 and BLLP1 based on the BL 3-D structure (PDB entry: 1GD6). The superimposition of ALLP1 and BLLP1 with BL (Figure 3.3) revealed that the overall structure and positions of the four-disulphide bonds is conserved but a few variations especially in the secondary structural elements were found. Thus, the amino acid substitutions at catalytic site do not seem to alter the 3-D structure of LLPs though it abolishes the muramidase activity.

3.3.3 Phylogenetic analysis clusters LLPs and lysozymes into distinct groups

Phylogenetic analysis clustered silkmoth LLPs with a few similar dipteran proteins from *Drosophila* and *Anopheles*, which also incidentally lacked one or both the catalytic amino acid residues (Figure 3.4). However, conventional dipteran and lepidopteran lysozyme sequences with conserved catalytic amino acid residues formed two distinct groups in the phylogenetic tree and did not group with LLPs. BLLP2 and hen lysozyme did not group with any other insect lysozymes or LLPs.

In the present study, through functional analysis of ALLP1 and BLLP1, we confirmed them as novel antibacterial proteins that are distinct from insect lysozymes.

Figure 3.1 Comparison of LLP and lysozyme sequences from silkmoths. Multiple alignment of amino acid sequences of *Antheraea mylitta* LLP and lysozyme (ALLP1 and AL, respectively), *Bombyx mori* LLPs and lysozyme (BLLP1, BLLP2, BLLP3 and BL, respectively) and hen lysozyme (HL). The position of two catalytic residues, Glu-32 and Asp-49 is indicated by (*) symbol and the corresponding substituted residues in LLPs are highlighted in green and red colour, respectively. Fully conserved residues are shaded in grey.

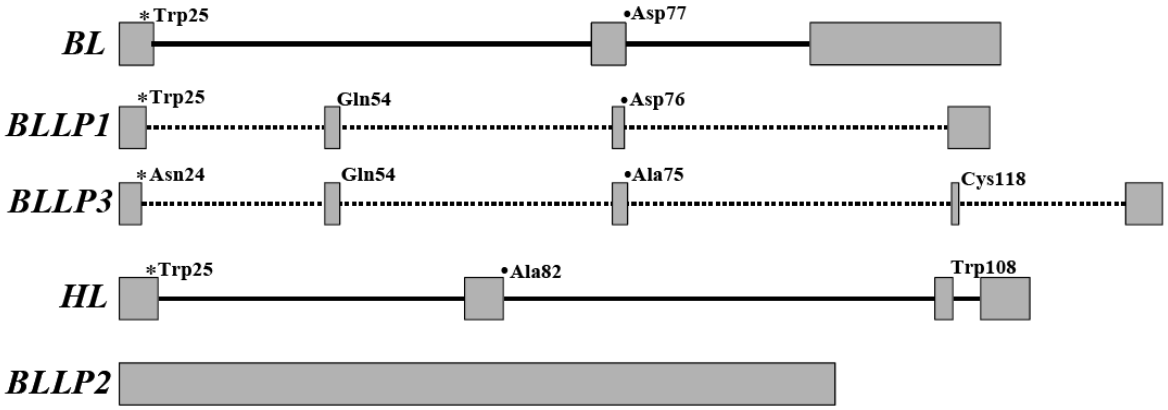

Figure 3.2 Gene structure comparison of *BL*, *BLLP1*, *BLLP2* and *BLLP3* with *HL*. The conserved position of introns in the different genes is marked with identical symbols and the corresponding amino acid residue at the exon-intron junction is shown. Exons are represented as grey boxes and the introns as lines intervening the exons. The introns of unknown length are represented as dashed lines and are not drawn according to scale.

Table 3.2 Molecular weight (MW) and isoelectric point (pI) of silkmoth lysozymes, LLPs and HL. Predicted pI and molecular weight (MW) of mature proteins

Lysozyme/LLP	Amino acids	MW (kDa)	pI
AL	120	13.99	8.81
ALLP1	158	18.56	9.82
BL	119	13.75	9.06
BLLP1	119	14.15	9.28
BLLP2	250	29.1	7.16
BLLP3	162	18.89	9.51
HL	129	14.31	9.32

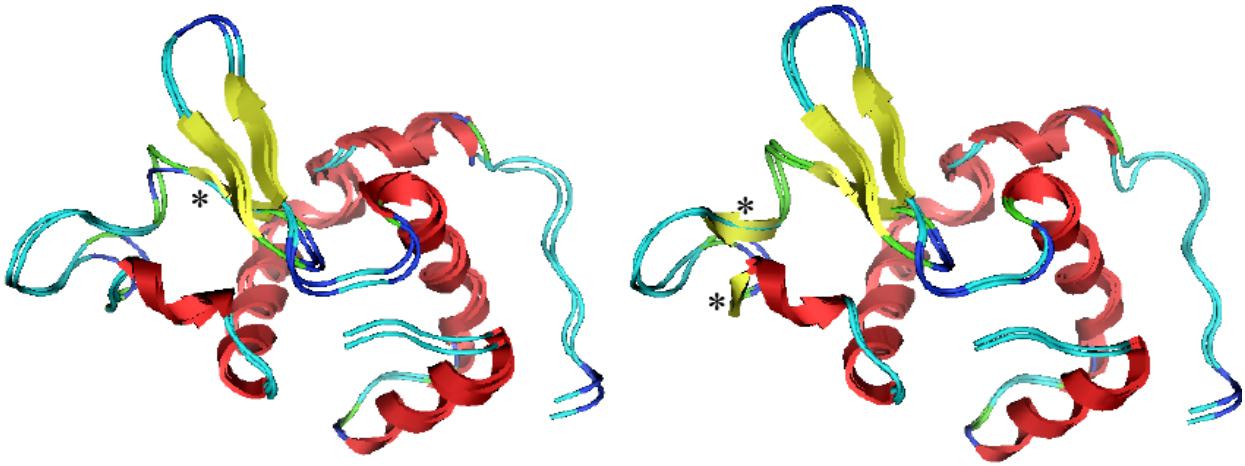

Figure 3.3 Superimposition of BL (PDB: 1GD6) with ALLP1 and BLLP1 protein tertiary structures, generated by homology modelling. ALLP1 (a) and BLLP1 (b) structures aligned with the BL tertiary structure with a r.m.s deviation of 0.39A° and 0.29A°, respectively. Helices, β-sheets and loops are represented in red, yellow and blue colours, respectively. The position of one absent β-strand in ALLP1 and presence of two extra β-strands in BLLP1 as compared to BL are marked with (*) symbol. ALLP1 structure comprised of six α-helical regions (residues 5-15, 19-33, 73-77, 81-95, 97-99, 101-106) and two anti-parallel β-sheets (residues 40-43, 46-50) and BLLP1 consisted of six α-helical elements (residues 9-15, 19-33, 74-78, 82-96, 98-100, 102-107) and five β-strands (residues 40-43, 46-50, 54-56, 60-63, 71-73) with interspersed loops.

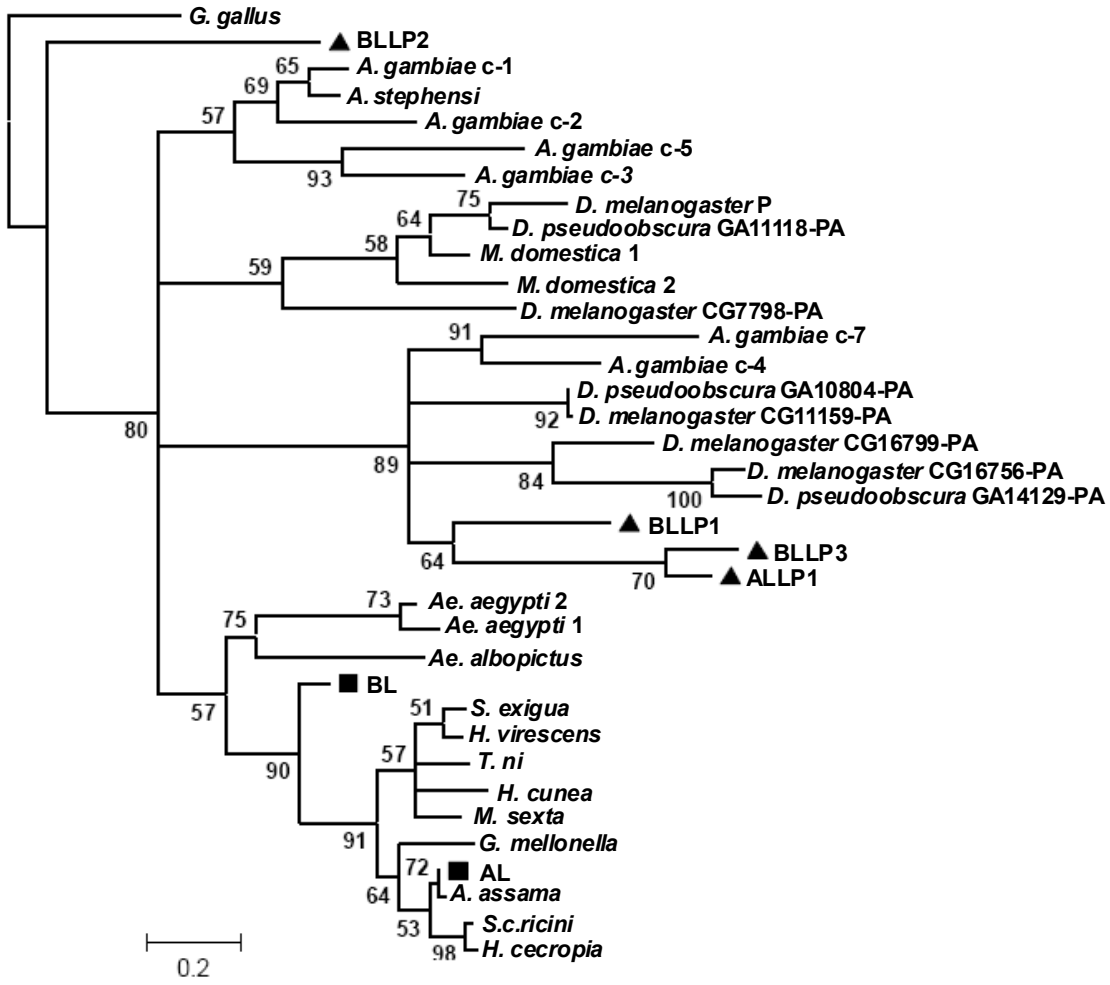

Figure 3.4 Maximum likelihood phylogenetic tree constructed from 35 insect full-length lysozyme sequences with hen lysozyme (*Gallus gallus*) as an outgroup. BLLP2 partial sequence that aligned with insect lysozymes and silkmoth LLPs was taken for analysis. The bootstrap values are represented as a percentage at the branch nodes. The branch lengths are proportional to the amino acid substitutions. The GenBank accession numbers for all the sequences can be found in the Table 3.1. The silkmoth lysozymes and LLPs are marked with closed squares and triangles respectively. Scale shows 0.2 amino acid substitutions per site.

3.3.4 LLPs are upregulated upon bacterial infection

Expression profile of *ALLP1* and *BLLP1* with *AL* and *BL* in *E. coli* and *M. luteus* challenged 5th instar larval fat body tissues was examined quantitatively. *AL* and *ALLP1* were moderately upregulated (3-4 fold) upon *M. luteus*-challenge, and highly induced (15-20 fold) upon *E. coli*-challenge (Figure 3.5). *BL* and *BLLP1* also exhibited similar changes but induction of *BLLP1* was less (2.7 fold) upon *E. coli*-challenge. We also confirmed their immune upregulation in other tissues from 5th instar immune-challenged *A. mylitta* and *B. mori* larvae by semi-quantitative RT-PCR (Figure 3.6 a and b). The expression pattern of *BLLP1* and *BL* in different embryonic stages of *B. mori* as analysed by semi-quantitative RT-PCR is shown in Figure 3.6 c. *BL* was not expressed at the 45-hour stage, increasing gradually in later stages. *BLLP1* was expressed prominently throughout the five stages analysed suggesting a probable function during embryonic development.

3.3.5 Protein expression, purification and refolding

AL, BL, ALLP1 and BLLP1 were expressed as inclusion bodies in BL-21 CodonPlus *E. coli* expression system and were purified under denaturing conditions by affinity chromatography. The purified proteins were then refolded and 6x histidine tag was cleaved followed by re-purification with Superdex-200 gel filtration. However the presence of histidine tag was not found to alter the activity of proteins. Hence the histidine tagged proteins were used for functional assays. The homogeneity and specificity of recombinant LLPs was confirmed by gel filtration, SDS-PAGE (Figure 3.7 a) and b) and western blot. The estimated MW of ALLP1 and BLLP1 were 29.47 kDa and 19.61 kDa respectively, which is slightly higher than the predicted MW of both the recombinant proteins. This is also true with the silkworm lysozymes, wherein estimated MW was found to be higher than the predicted MW. ALLP1 is a larger protein than BLLP1 and possesses 39 extra amino acids at the C-terminal end. Refolding was assessed by CD spectroscopy and the estimated secondary structure was in agreement with the in-silico predicted structure (Figure 3.8).

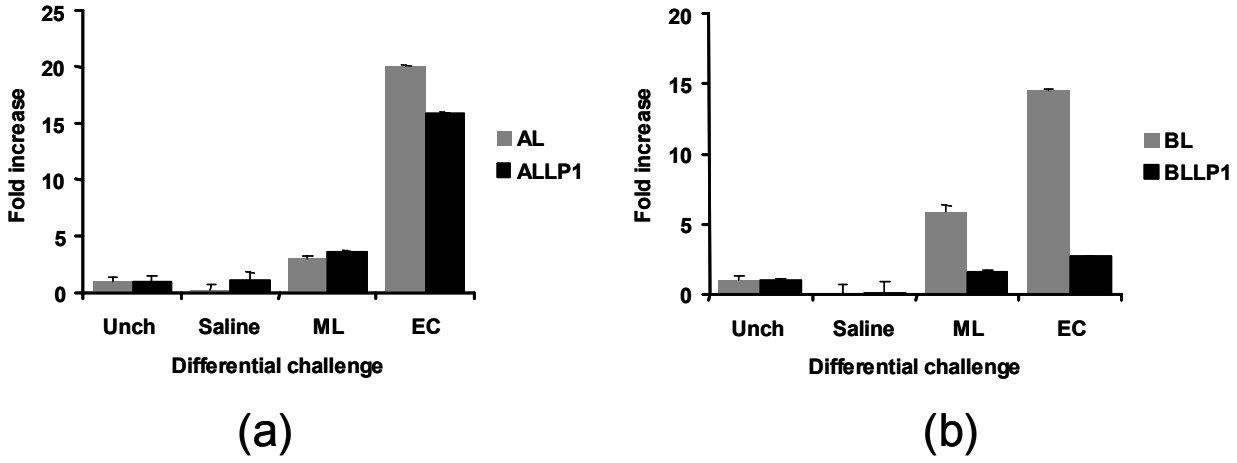

Figure 3.5 Real time PCR profile of (a) *AL* and *ALLP1* and (b) *BL* and *BLLP1* in fat body from differentially challenged 5th instar, day 3 larvae. The Y-axis represents the fold increase of the transcripts over the unchallenged levels. The values reported are an average of two independent experiments and the standard deviation is represented as the Y error bar. X-axis represents the four differentially challenged larval groups -1. Unch-unchallenged, 2. Saline, 3. ML-*Micrococcus luteus* and 4. EC- *Escherichia coli*.

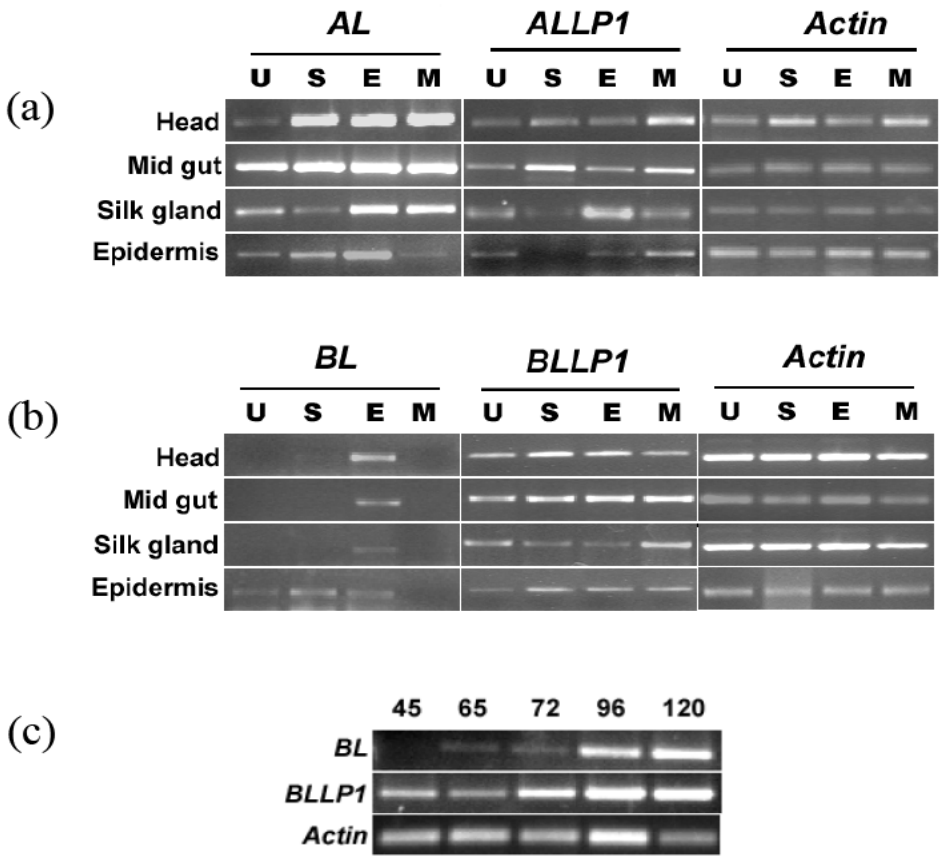

Figure 3.6 Tissue specific expression profiles of *AL*, *ALLP1* (a) and *BL*, *BLLP1* (b) in four tissues from differentially challenged 5th instar silkworm larvae. U-unchallenged, S-saline-injected, E-*Escherichia coli*-challenged, M-*Micrococcus luteus*-challenged (c) Expression profile of *BL* and *BLLP1* in different embryonic stages by semi-quantitative RT-PCR. The number above the panel shows the time in hours when RNA was extracted post oviposition. Actin expression was used as an endogenous control.

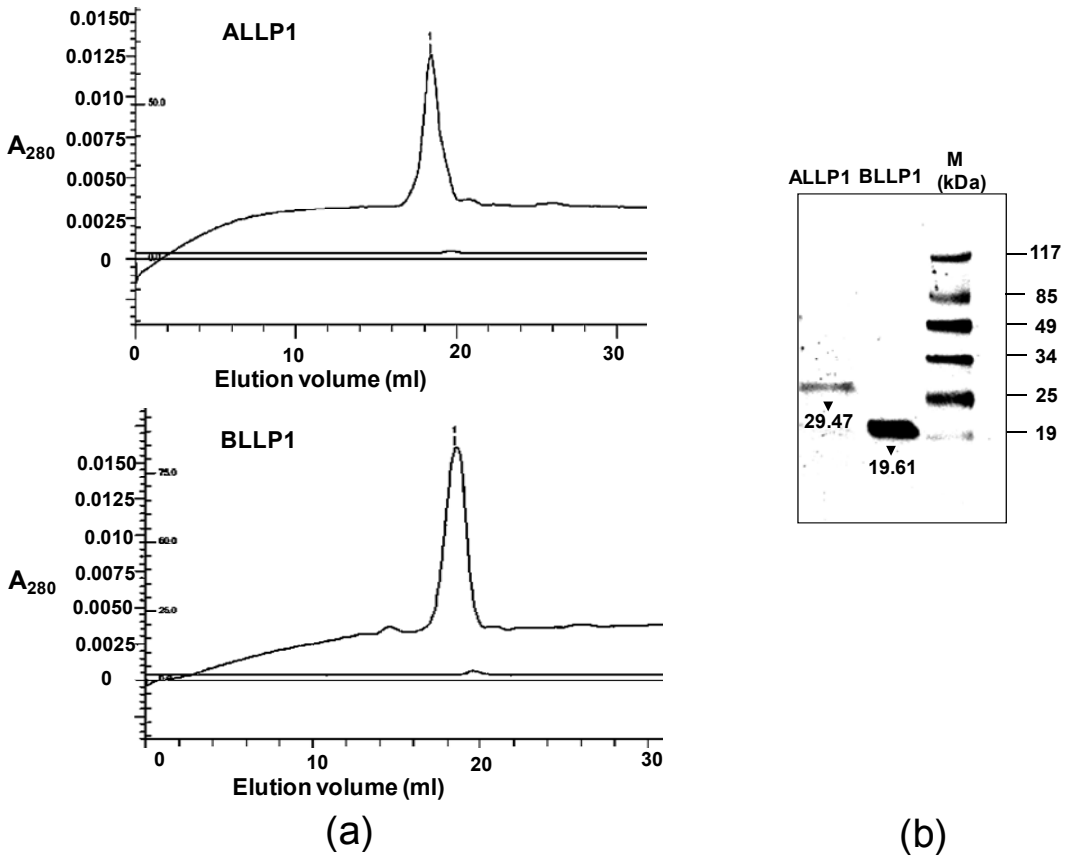

Figure 3.7 Purification of ALLP1 and BLLP1. (a) Size exclusion chromatography of ALLP1 and BLLP1 using Superdex-200 columns (Amersham). The Ni-NTA affinity chromatography purified proteins were further purified by size exclusion chromatography. (b) The corresponding protein peaks were collected and analysed on SDS-PAGE. The calculated MW (in kDa) of the two proteins is indicated. M- Protein molecular weight standards.

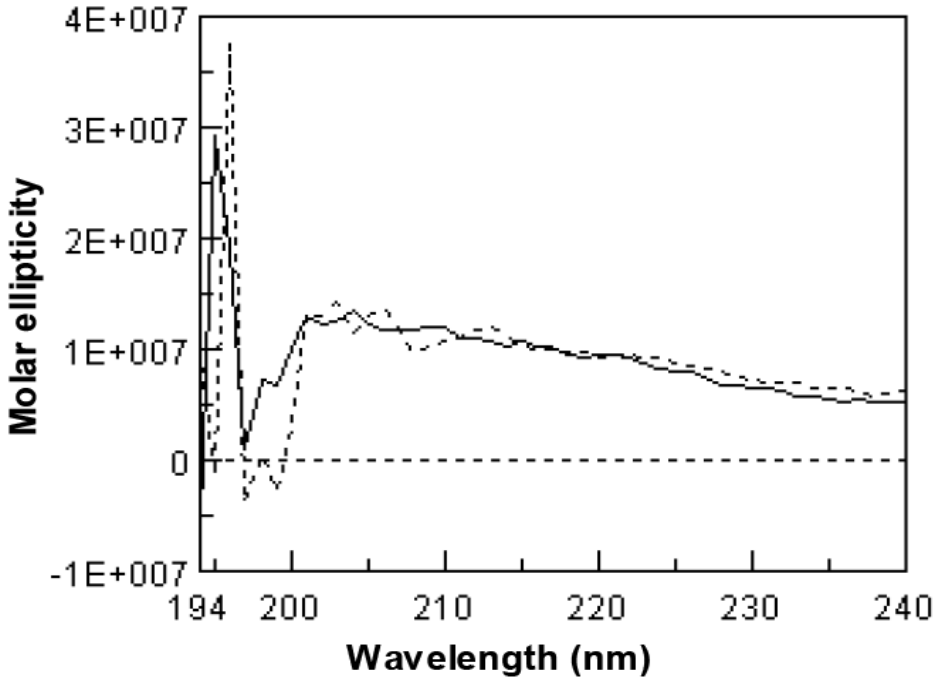

Figure 3.8 CD spectroscopy of ALLP1 and BLLP1. Far UV (194-240nm) CD spectra of ALLP1 (continuous line) and BLLP1 (dashed line) were obtained in 0.1 M ammonium acetate buffer (pH 5.5) and the baseline value was subtracted. The secondary structure was estimated using the inbuilt program in Jasco J-810 spectropolarimeter. Secondary structure estimated was as follows-ALLP1: Helix-13.7%, Beta-19.7% and Turn- 33%; BLLP1: Helix: 0, Beta: 26.5% and Turn- 73.5%. The estimated secondary structure is in agreement with the in-silico predicted structure suggesting correct folding of the recombinant proteins.

3.3.6 LLPs lack muramidase activity but exhibit substrate binding property

The spectrophotometric assay for muramidase activity utilizes the lyophilized *M. luteus* as a substrate and the activity is monitored by a decrease in the turbidity upon addition of the enzyme. Assay results (Table 3.3) show that ALLP1 and BLLP1 lack muramidase activity even at ten fold higher protein concentration over AL and BL, which exhibited a specific activity of 15,657 and 11,571 U/mg respectively. This confirmed the ablation of muramidase activity in LLPs as predicted from the substitutions at the catalytic residues (MALCOLM *et al.* 1989).

Table 3.3 Muramidase activity of silkmoth lysozymes (AL, BL) and lysozyme-like proteins (ALLP1, BLLP1).

Lysozyme/LLP	Specific activity (U/mg)
AL	15658±1117
BL	11572±241
ALLP1	No activity
BLLP1	No activity

Both ALLP1 and BLLP1 were observed to bind peptidoglycan, the characteristic lysozyme substrate, when analysed by western blot analysis (Figure 3.9). This confirms that the LLPs bind to peptidoglycan although they are unable to hydrolyse it unlike classical lysozymes.

3.3.7 LLPs exhibit anti-bacterial activity

Both ALLP1 and BLLP1 inhibited a wide range of bacterial species including both gram-positive and gram-negative bacteria in the radial diffusion assay (Figure 3.10). Similar to the silkworm lysozymes, ALLP1 and BLLP1 also showed a higher activity against the gram-positive bacteria than the gram-negative bacteria. However the only exception to this trend is

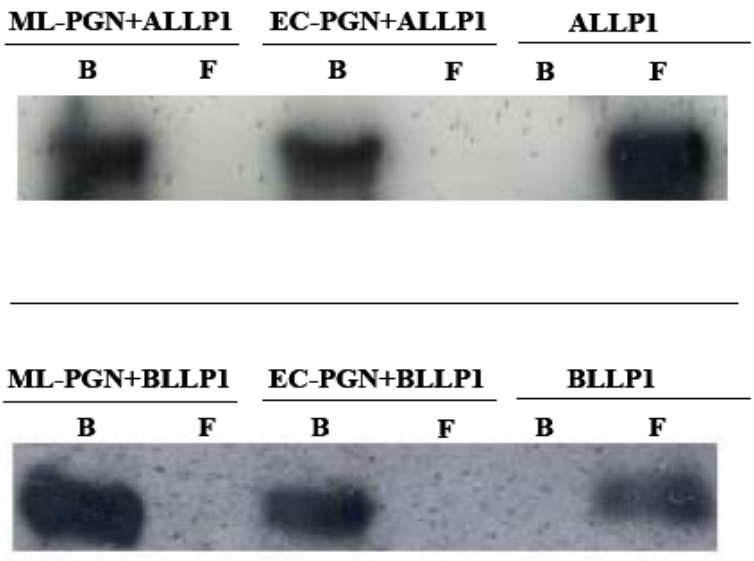

Figure 3.9 Peptidoglycan binding assay of ALLP1 and BLLP1 with *Micrococcus luteus* and *Escherichia coli* insoluble peptidoglycan. The bound (B) and free (F) fractions were obtained as described in the text. All the samples were run on a 12% SDS-PAGE gel and ALLP1 and BLLP1 proteins were detected by western blot using anti-histidine antibodies. A control containing only ALLP1 or BLLP1 without peptidoglycan was treated in the same manner as the test samples and bound and free fractions were analysed. EC- *E. coli*, ML- *M. luteus*.

E. coli against which all the four proteins exhibited activity comparable to that against the other gram-positive bacteria. The inhibitory concentration values of all the four proteins are presented in Table 3.4.

Since bacterial growth inhibition indicates either bacteriostatic or bactericidal action, the viability of *E. coli* (gram-negative bacterial species) and *M. luteus* (gram-positive bacterial species) was determined at different time points after treatment with BL and BLLP1. Figure 3.11 a and b indicate the kinetics of anti-bacterial action of BL and BLLP1 against *E. coli* and *M. luteus*, respectively. BLLP1 exerts a bacteriostatic effect on both the bacteria and no killing was observed even after 8-hour incubation as against BL, which reduced the viable bacterial count, by 2-3 log scale (Figure 3.11).

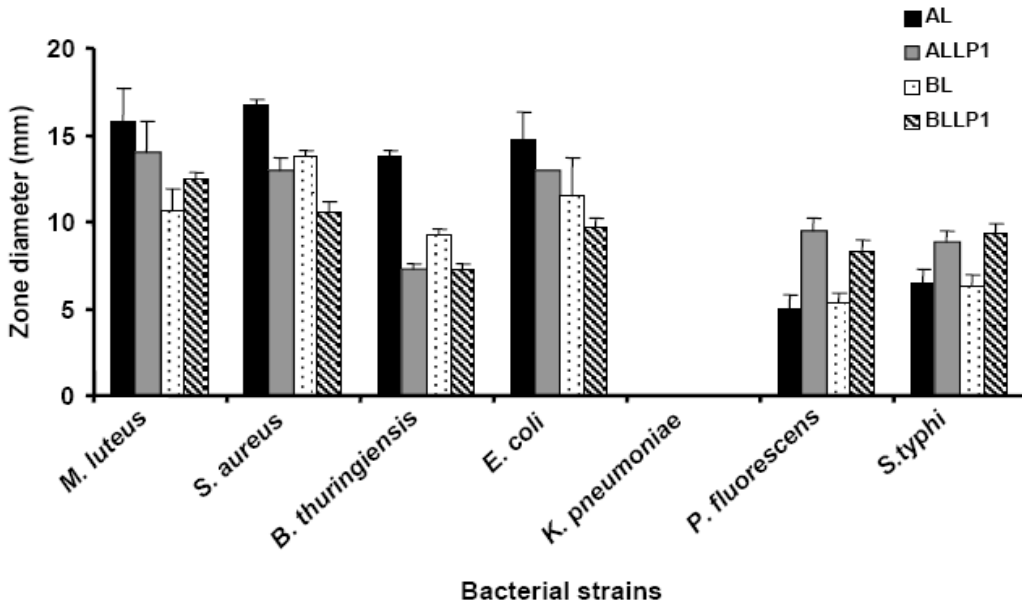

Figure 3.10 Radial diffusion anti-bacterial assay of AL, ALLP1, BL, and BLLP1. Anti-bacterial activity of the purified proteins was assayed against three gram-positive (*Micrococcus luteus*, *Staphylococcus aureus*, *Bacillus thuringiensis*) and four gram-negative (*Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas fluorescens, Salmonella typhi*) bacteria. The zone of inhibition is plotted on the Y-axis and the different bacterial species are indicated on X-axis.

Table 3.4 Inhibitory concentration (IC) of lysozymes and LLPs. The IC value was calculated from the inhibition zone data as described in the experimental procedures. The IC value against various gram-positive and gram-negative bacteria tested is mentioned in the table below.

Bacteria	**IC (μM)**			
Gram-positive	AL	ALLP1	BL	BLLP1
Micrococcus luteus	0.57	0.77	1.62	1.06
Staphylococcus aureus	0.49	0.93	0.81	1.67
Bacillus thuringiensis	0.81	5.20	2.40	5.20
Gram-negative				
Escherichia coli	0.67	0.93	1.29	2.11
Pseudomonas fluorescens	23.4	2.22	17.33	3.29
Salmonella typhi	6.68	2.72	8.5	2.34

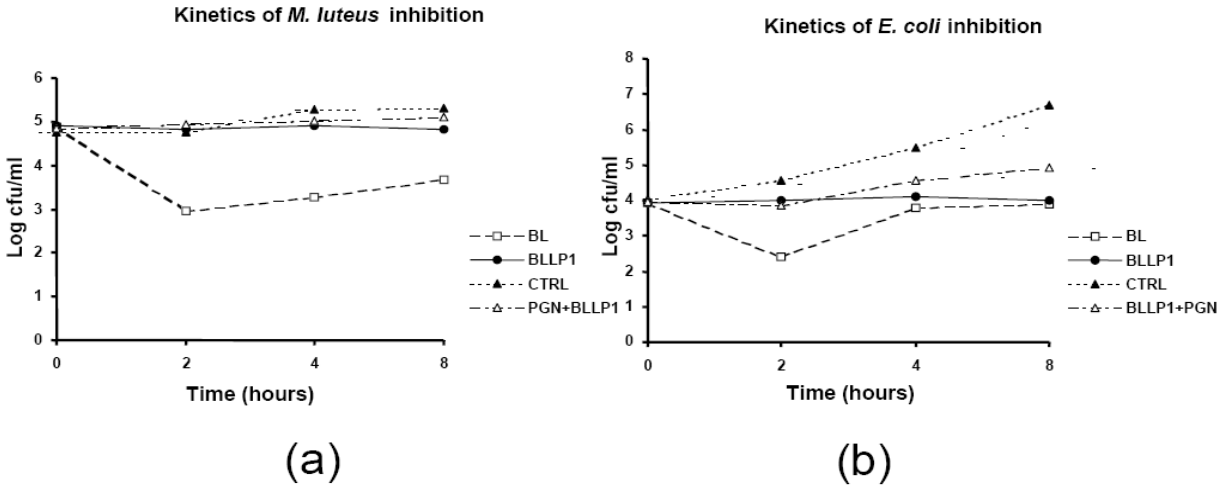

Figure 3.11 Kinetics of viability of *Escherichia coli* (a) and *Micrococcus luteus* (b) at different time points post incubation with BL and BLLP1. The log_{10} value of colony forming units (cfu/ml) is plotted on the Y-axis and the time period of incubation with the proteins is shown on the X-axis. A control in which the bacteria were incubated with equivalent amount of PBS and a competition assay in which BLLP1 was pre-incubated with peptidoglycan was also done.

The mechanism of lysozyme action is well studied. It involves peptidoglycan hydrolysis that results in osmotic lysis of bacteria. The bacterial membrane is the site of anti-microbial action of many other vertebrate and invertebrate AMPs as exemplified by cecropins, magainins and defensins (BULET *et al.* 2004). These AMPs permeabilize the inner membrane (IM) of bacteria by forming pores leading to rapid death of the organism. IM permeabilization of lysozymes and LLPs was assayed by measuring the β-galactosidase activity of *E. coli* samples incubated with BL and BLLP1 (Figure 3.12). β-galactosidase is located in the cytosol and hence will only be detected if the bacterial cell membrane is permeabilized.

Figure 3.12 shows that β-galactosidase activity increased in bacterial samples incubated with BL reaching a maximum at 4 hours post incubation with BL. Though the mechanism of lysozyme action involves peptidoglycan hydrolysis, the cells subsequently lyse due to the osmotic pressure as an indirect effect of peptidoglycan hydrolysis thereby releasing the cytoplasmic contents. However, no β-galactosidase activity was detected in the bacterial samples incubated with BLLP1 even at 8 hours post incubation indicating that they neither hydrolyse the peptidoglycan like conventional lysozymes nor do they permeabilize the cell membrane like many other AMPs as has been demonstrated before (STEINBERG and LEHRER 1997).

The ability of LLPs to bind the peptidoglycan substrate and its higher activity against gram-positive bacteria pointed towards involvement of peptidoglycan binding as a possible mechanism of anti-bacterial action. We observed that the inhibitory activity of BLLP1 against *E. coli* and *M. luteus* was actually rescued in the presence of peptidoglycan indicating that growth inhibition primarily involves peptidoglycan binding (Figure 3.11). Bacteria-binding proteins opsonise bacteria and agglutinate them and this may cause error in counting the bacteria by colony count method. However, this possibility was ruled out, as agglutination of bacteria was not observed post treatment with BLLP1 when visualised by phase contrast microscopy. Thus at this stage it indicates that peptidoglycan binding is indeed involved in the anti-bacterial action though it is not known whether peptidoglycan binding itself renders the anti-bacterial effect or some other process is inhibited consequent to peptidoglycan binding.

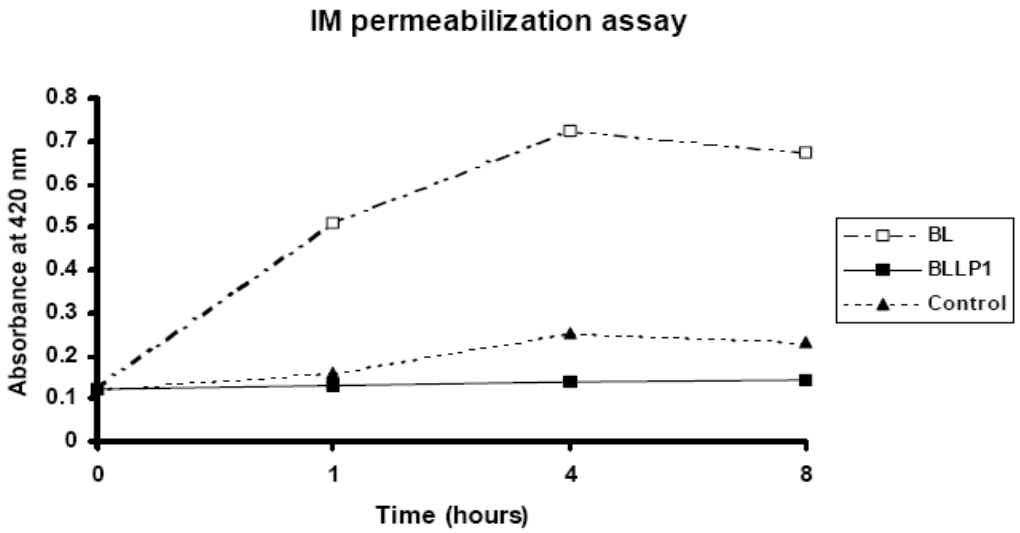

Figure 3.12 Kinetics of inner membrane (IM) permeabilization of *Escherichia coli* by BL and BLLP1. The IM permeabilization was determined by a spectrophotometric assay as explained in experimental section by measuring the β-galactosidase activity of *E. coli* samples incubated with BL and BLLP1 using a spectrophotometric assay. The absorbance is plotted on Y-axis, which is proportional to the β-galactosidase activity and the incubation time of bacterial samples with BL and BLLP1 is shown on X-axis.

3.3.8 In vivo functional analysis by RNAi

Previous reports have shown that immunity is mediated by secretion in to the haemolymph of recognition proteins and effector anti-microbial proteins upon immune challenge. Hence we analysed the clearance of *E. coli* from haemolymph in BLLP1 knock-down insects as compared to larvae injected with non-specific dsRNA (GFP-dsRNA) or saline prior to *E. coli*-challenge. Figure 3.13 a indicates that *E. coli* load was 3-4 fold higher in BLLP1 knock-down larvae as compared to the control larvae. The BLLP1-dsRNA induced knock-down of BLLP1 was confirmed at molecular level by RT-PCR and is shown in Figure 3.13 b. This indicates the involvement of BLLP1 in immunity and point towards its occurrence in haemolymph as predicted from the signal peptide studies.

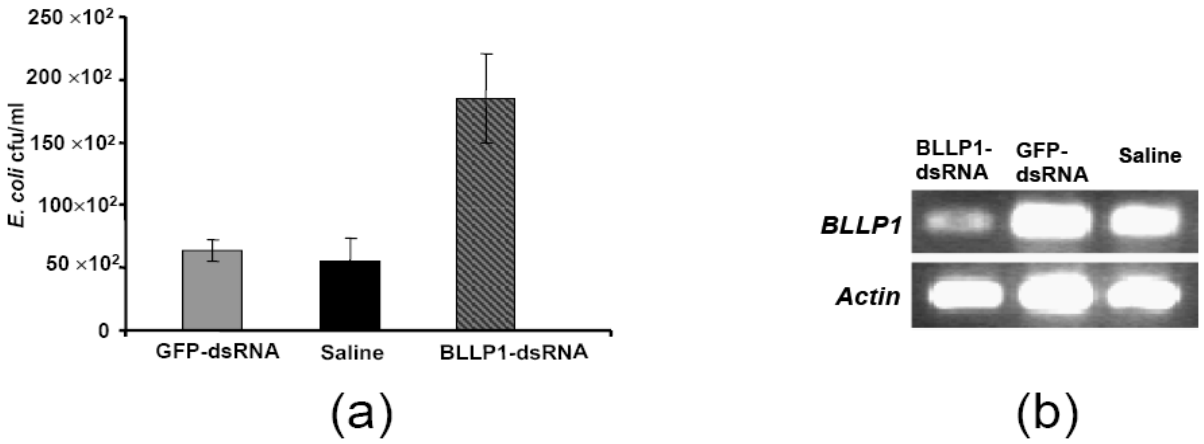

Figure 3.13 In vivo functional analysis of BLLP1 by RNAi. (a) Bacteria clearance assay: Colony forming units/ml of *Escherichia coli* in the hemolymph of BLLP1-dsRNA injected and control (GFP-dsRNA injected and saline) larvae. The data represented is an average of two independent experiments with a set of 5 larvae in each group. The standard deviation is represented as Y error bars. (b) RT-PCR analysis from fat body of BLLP1 knock-down (BLLP1-dsRNA injected) and control (saline-injected and GFP-dsRNA injected) larvae. Actin is used as an endogenous control.

3.4 Discussion

The present study focuses on the characterization of a novel family of antibacterial proteins, ALLP1 and BLLP1 in the present study and also identified two more LLPs, BLLP2 and BLLP3 in *B. mori*, which need to be further, analysed functionally. BLLP3 probably resembles ALLP1 and BLLP1 functionally due to high degree of similarity among them. Moreover, the EST database of *B. mori* (MITA *et al.* 2003) revealed that BLLP3 was expressed in the fat body of larval and pupal stages and also found in the baculovirus-

challenged fat body tissue of *B. mori*. However BLLP2 was very different from the other LLPs and c-type lysozymes and resembled the multidomain lysozyme-like proteins, c-6 from *A. gambiae* and one *Drosophila* lysozyme. Also, it was found to be expressed in the pheromone gland in the *B. mori* EST database suggesting that it may be involved in a different function altogether.

The insect immune repertoire elicits an impressive array of anti-microbial proteins that enable a strong effector response to successfully ward off infections during their lifetime. The immune-upregulation and the in vitro and in vivo functional analyses suggest that LLPs are anti-bacterial proteins, which play an important role in immune response. There are a few evidences in literature demonstrating the anti-bacterial nature of catalytically inactive mutant lysozymes (IBRAHIM *et al.* 2001; NASH *et al.* 2006). We however present a first report of naturally occurring muramidase deficient anti-bacterial c-type lysozyme-like proteins from insects. Ibrahim *et. al.* have shown earlier that a catalytically inactive Asp-52-Ser hen mutant

lysozyme possesses bactericidal activity against *S. aureus* and *B. subtilis* (IBRAHIM *et al.* 2001) and recently Nash *et. al.* have demonstrated the in vivo killing properties of Asp-52-Ser, a mutant mouse lysozyme against *S. aureus*, *K. pneumoniae* and *P. aeruginosa* (NASH *et al.* 2006). Our results on the other hand show that BLLP1 inhibits the growth of *E. coli* and *M. luteus* and does not kill them. Previous work on non-catalytic lysozyme mutants has led to three probable mechanisms to explain the antibacterial action: 1. lysis of cells due to binding of lysozyme to a potential target inducing autolysin mediated death, 2. membrane permeabilization leading to cell lysis and 3. non-lysis mediated mode due to interference of membrane function as a consequence of lysozyme binding at the surface (IBRAHIM *et al.* 2001). Our study corroborates the last mechanism as exemplified by BLLP1, which neither permeabilizes the inner membrane nor causes lysis. Moreover, the ability of the catalytically inactive LLPs to bind peptidoglycan and their higher activity towards gram-positive bacteria point towards peptidoglycan as the possible site of anti-bacterial action. The rescue of the bacterial inhibition upon incubation with peptidoglycan indeed points towards involvement of peptidoglycan binding as a requirement for anti-bacterial mechanism. Though the site of action of LLPs as well as lysozymes appears to be peptidoglycan, the mechanism appears to be different as the LLPs lack peptidoglycan hydrolysis property. Peptidoglycan layer is surrounded by an outer membrane in gram-negative bacteria making them less-accessible to

AMPs acting on peptidoglycan and this trend also is well reflected by LLPs which are preferentially more active against gram-positive bacteria. We speculate that the bacterial inhibition is due to the binding of LLPs to bacterial surface and subsequently interfering with its normal physiology. One possibility is that the LLPs bind peptidoglycan precursors and hence compete with the peptidoglycan synthesis machinery thereby affecting its growth. Our hypothesis is supported by a recent report indicating bacteriostatic and bactericidal properties of human peptidoglycan recognition proteins, another class of peptidoglycan binding proteins (LU *et al.* 2006) . However, the mechanisms of anti-bacterial action remain speculative at this juncture and merit further analysis.

The roles of several immune proteins have been analysed by RNAi mediated knock-down in the recent studies (FOLEY and O'FARRELL 2004; KAMBRIS *et al.* 2006) . We, in this study show an increased proliferation of bacteria in BLLP1 knock-down larvae. Insect immune repertoire comprises of several anti-bacterial proteins and hence effects of knock-down of a single AMP on the bacterial proliferation are intriguing. However different immune pathways exhibit a complex network and involve cooperative roles of different immune factors. It is possible that BLLP1 is involved in other functions in addition to the anti-bacterial effector function. For example, human peptidoglycan recognition proteins have recently been shown to function as both recognition and anti-bacterial proteins (LU *et al.* 2006) . The immune role of LLPs is established by these analyses but a further study is required to probe its exact role in the immune system. A strong expression of transcripts at the embryonic stages also suggests a possible role during development. Many immune related protein molecules, for example, hemolin has been shown to have a dual role in immunity as well as development (BETTENCOURT *et al.* 2002) and hence other functions cannot be ruled out.

The current studies present a first report of a functional analysis of a novel class of lysozyme-like proteins lacking muramidase activity from insects. The functional annotation of this new group of lysozymes likely provides cues for analysis of their homologues in other organisms. For example, two human c-type lysozyme-like proteins lacking catalytic residues have been recently reported (ZHANG *et al.* 2005) whose function might be elucidated in the light of the insect LLPs reported in the present study.

Chapter-4

Noduler, a novel immune upregulated protein mediates nodulation response in insects

4.1 Introduction

Insect immunity has gained importance in the past decade as a model to study mammalian innate immunity due to several parallels between them (HOFFMANN 2003). Insect immune system possesses only an innate immune arm that constitutes both humoral and cellular responses. Humoral response is characterized by a rapid activation of the phenoloxidase cascade (CERENIUS and SODERHALL 2004) and synthesis of an array of antimicrobial proteins in the fat body within hours upon microbial challenge that are subsequently secreted in insect haemolymph (BULET *et al.* 2004; HULTMARK *et al.* 1980; LEMAITRE and HOFFMANN 2007; YAMAKAWA and TANAKA 1999). Cellular response, on the other hand, is mediated by haemocytes and is an immediate response triggered within minutes of pathogen exposure (LAVINE and STRAND 2002). Cellular defence mechanisms constitute nodulation, encapsulation and phagocytosis. Nodulation is quantitatively the most important defence mechanism against bacterial, fungal and even viral infections in insects and other invertebrates (MILLER *et al.* 1994; TRUDEAU *et al.* 2001). Nodulation is mediated by haemocyte aggregate formation around bacteria and fungi (LAVINE and STRAND 2002). Encapsulation is similar to nodulation and refers to haemocyte aggregation around larger pathogens like parasitoids and nematodes. Although knowledge on antimicrobial peptide synthesis and phenoloxidase pathways is accumulating over the years, the cellular responses involved in phagocytosis and nodulation are relatively unknown. Only recently two transmembrane proteins namely Eater (KOCKS *et al.* 2005) and Nimrod (KURUCZ *et al.* 2007) containing multiple epidermal growth factor-like repeats were identified on the surface of haemocytes in Drosophila. These proteins were shown to be major mediators of bacterial phagocytosis in Drosophila. However no knowledge on players involved in nodulation is reported so far. Our initial analysis of *A. mylitta* immune transcriptome identified a novel immune protein upregulated in haemolymph upon bacterial infection referred to as defence protein 1 (DFP-1) (GANDHE *et al.* 2006a). In the present study we assessed the functional role of DFP-1 in immune response and named it as Noduler due to its involvement in nodule formation. Noduler was shown to bind several microorganisms and their ligands. Binding to microbial ligands is a criterion exhibited by insect pathogen recognition receptors (PRRs) that bind pathogen associated molecular patterns (PAMPs) (STEINER 2004; YU *et al.* 2002). In vivo RNAi mediated knock-down of Noduler especially affected the nodulation response thus suggesting a role for this protein in immunity. Similar proteins have been reported

earlier from other lepidopterans but their role in immunity was not understood (SHIN *et al.* 1998; ZHU *et al.* 2003).

4.2 Materials and methods

4.2.1 Insects and bacteria

A. mylitta, 5th instar, day 3 larvae were procured from Regional Research Station, Warangal, India and Central Tasar Research and Training Institute, Ranchi, India.

4.2.2 Isolation of Noduler cDNA sequence

A. mylitta Noduler cDNA sequence was identified from an *E. coli* challenged fat body transcriptome (GANDHE *et al.* 2006b). The protein sequence was derived from the cDNA sequence and signal peptide prediction was performed using signalP program.

4.2.3 Multiple sequence alignment and phylogeny

Phylogenetic analysis was done by comparison of Noduler with similar protein sequences obtained by blastp (E-value≤1e-05). Reelin from *Homo sapiens* was used as an outgroup. Only the reeler domains of the proteins were used for the analysis. All the sequences were aligned using ClustalX 1.8 and manually edited using GeneDoc Version 2.6.002. A neighbour-joining (NJ) tree was constructed using HYPHY and was used as a base tree to test the best-fit model of evolution in HYPHY. Akaike Information Criterion (AIC) was used to select the best-fit model, which selected WAG model of amino acid substitution. A maximum likelihood tree was constructed with the program TREE-PUZZLE 5.2 that utilized 50,000 quartet-puzzling steps.

4.2.4 Recombinant expression of Noduler in insect cells

Noduler was expressed in Sf9 cells by constructing a recombinant baculovirus using the BAC-TO-BAC (Invitrogen) system. N-terminal 6x histidine tagged mature Noduler protein

was expressed by amplification using the following primers: Forward-5′-CGGGATCCATGCATCACCATCACCATCACTTCCCTACTGGTGCACCA-3′, Reverse-5′-CGCTCTAGATTAATGGTGACTTAAAATTTTTACGGGTGCCGAGGTTTG-3′. The PCR product was digested with BamHI and XbaI and cloned into pFASTBAC vector (Invitrogen). The Noduler encoding recombinant baculoviruses were subsequently obtained using manufacturer's protocol (Invitrogen). Sf9 cells were harvested 72 hours post infection (pi) with recombinant baculovirus. The protein was purified under native conditions using BD TALON^{TM} metal affinity resin (Clontech) according to the manufacturer's instructions. The purified protein was analysed by 15% reducing SDS-PAGE gels and western blot with anti-histidine and Noduler-specific peptide antibody. Anti-Noduler antibody was generated in rabbits against an internal peptide sequence of Noduler - DKQTVSYVWTAPSDLEGDVVF. This peptide did not show similarity with any other protein in the NCBI database except Noduler and its homologues. The ability of this antibody to detect Noduler was confirmed against recombinant Noduler expressed in insect cells. However the Noduler peptide exhibits 67% and 81% sequence identity with corresponding peptides of Noduler homologue 1 and 2, respectively. Hence cross reactivity of anti-Noduler antibody against the two Noduler homologues in *A. mylitta* is possible.

4.2.5 Immunoblot analysis

Haemolymph was collected from *A. mylitta* 5th instar, day 3, larvae in pre-chilled polypropylene tubes containing a few crystals of phenylthiourea at 30 mins, 1, 4, 8 and 24 hrs post *E. coli* challenge. The haemocytes were removed by centrifugation at 2000 rpm and equal amount of plasma samples from each time point were subjected to western blot. The blots were probed with a 1000 fold dilution of Noduler-specific rabbit primary antibody. Detection was done with goat anti-rabbit IgG (1:2000 dilution) conjugated to horseradish peroxidase and ECL detection reagent (Amersham).

4.2.6 Binding assays with microorganisms, microbial cell wall components and insect cells

Binding assays with bacteria, yeast, peptidoglycan (PGN) and curdlan were carried out using a protocol by Kang *et. al.* (KANG *et al.* 1998) with a few modifications. Curdlan, *S. aureus*

PGN and *M. luteus* PGN were purchased from Sigma. *E. coli* PGN was a kind gift from Dr. Bruno Lemaitre. We further repurified this *E. coli* PGN as described by Glauner *et. al.* (GLAUNER 1988). Briefly, *E. coli* PGN was treated with trypsin (200 μg) at 37°C for 30 mins to remove any contaminating lipoprotein followed by extraction with 8% boiling SDS for 30 mins. The pellet obtained was washed several times with water and used for binding assays. For the binding assays, log phase bacterial culture (3 ml) was pelleted down and resuspended in 20 μl PBS. Also, for curdlan and PGN binding experiments, 1 mg of each of the components was utilized and resuspended in 20 μl PBS. 1 μg of recombinant Noduler protein was added to the substrates. After 30-min incubation, the reaction components were pelleted down, the pellet was washed four times with PBS and the supernatant and wash fractions were collected. The pellets were resuspended in 40 μl PBS, boiled for 5 min after addition of 2X SDS-PAGE loading buffer and analysed by western blot. Anti-histidine primary antibody (Qiagen) at 1:2000 dilution and anti-mouse IgG secondary antibody conjugated to horseradish peroxidase (Amersham) at 1:2000 dilution was used. The detection was done with ECL Plus detection reagents (Amersham).

Binding of Noduler to lipopolysaccharide (LPS) and lipotechoic acid (LTA) was determined by a competition assay wherein Noduler binding to *E. coli* and *S. aureus* was competed by adding 10 μl (50 μg) of soluble LPS (*E. coli* 0111:B4, Sigma) and LTA (*S. aureus*, Sigma), respectively. To remove possible PGN contamination from *E. coli* LPS (Sigma), we treated it with lysozyme for 12 hrs at 25°C according to a protocol by Werner *et. al.* (WERNER *et al.* 2000). Activity of lysozyme was confirmed on pure PGN and also it was not inhibited in the presence of LPS. Further, LPS was reextracted with phenol by a method described by Hirschfeld *et. al.* (HIRSCHFELD *et al.* 2000). The binding assay samples were assayed by western blot as described earlier. The band intensity of Noduler in the supernatant fraction in the presence or absence of competition (LPS and LTA) was estimated with ImageQuant 5.2 software. Amount of Noduler bound to soluble LPS or LTA was obtained by subtracting the band intensity of the supernatant fraction in the absence of LPS or LTA from that obtained in their presence. Noduler binding to LPS and LTA was expressed as a percentage value of the amount of bound Noduler as compared to the total protein (bound+unbound fraction). The data shown is an average of four independent experiments and the error bars indicate the standard deviations. As a negative control a histidine-tagged bacterial Hsp70 protein was used in all the binding assays.

4.2.7 Preparation of dsRNA

The dsRNA specific to Noduler was synthesized corresponding to entire protein coding region (507 bp) using primers, Forward primer, 5' ATGATGTTCGCGTACATAGTAGCTG 3', Reverse primer, 5' TTAATGGTGACTTAAAATTTTTACGGGTG 3' (Figure 4.1 a), and cloned into pCRII-TOPO vector (Invitrogen) followed by amplification with M13 forward and reverse primers. This template with flanking T7 and SP6 promoters was utilized for in vitro transcription reaction and sense and antisense RNA strands were generated with the T7 and SP6 Megascript kits as prescribed by the manufacturer (Ambion). The DNA template was removed from the transcripts by DNAse treatment and the RNA products were subsequently purified by Trizol extraction (Invitrogen) followed by isopropanol precipitation. The complementary single stranded RNAs were dissolved in DEPC treated water, combined in equimolar amounts in 1X insect buffer saline (IBS, composition: NaCl –160 mM, KCl- 10 mM, $CaCl_2$- 4 mM) and annealed by heating to 95 ºC and slow cooling overnight at room temperature. Similarly, dsRNA specific to full length green fluorescent protein (GFP) was synthesized as a non-specific control. The dsRNA formation was confirmed by agarose gel electrophoresis and the concentration was determined spectrophotometrically.

4.2.8 RNAi mediated knock-down and bacteria clearance assay

The in vivo functional analysis of Noduler was carried out by dsRNA mediated Noduler knock-down in 5^{th} instar larvae. The ability of Noduler knock-down larvae to clear both gram-positive (*S. aureus*) and gram-negative (*E. coli*) bacteria that were injected systemically was subsequently assayed. Noduler-dsRNA (100 µg) was injected into a set of larvae (n=8). GFP-dsRNA (non-specific-dsRNA) or saline injected larvae (n=8) were maintained as control. 24-hrs post dsRNA injection, control and test larvae were injected with approximately 10^7 cells of log phase *E. coli* and *S. aureus* bacteria, separately, resuspended in 1X IBS. The larval haemolymph was collected aseptically; 30 min post-bacterial injection in phenylthiourea (PTU) and 100 µl of the ten fold diluted haemolymph samples were plated on LB agar plates. The colony forming units (CFU) were enumerated after overnight incubation of the LB plates. The data presented is an average of values obtained from 8 larvae in each set. The statistical significance was determined using t-test.

4.2.9 RNAi mediated Noduler knock-down and its effect on i. nodule formation, ii. phenoloxidase activity and iii. phagocytosis

i. Noduler-dsRNA or GFP-dsRNA (100 μg each) or saline was injected to day 3, 5th instar larvae followed by an *E. coli* MG1655 or *S. aureus* injection (approx 10^7 cells/larva) 6 hours later. Eighteen hrs post *E. coli* challenge, the larvae were dissected under a saturated solution of phenylthiourea prepared in 1X insect buffer saline (IBS) and nodule formation was assessed as described by Eleftherianos *et al* (ELEFTHERIANOS *et al.* 2007). The nodules were counted manually using a stereomicroscope. The data presented is an average of 8 larvae in each group and statistical analysis was done using t-test.

ii. Noduler-dsRNA or GFP-dsRNA (100 μg each) was injected to day 3, 5th instar larvae followed by an *E. coli* injection (approx 10^7 cells/larva) 24 hours later. One hour post *E. coli* injection, the Noduler knocked-down as well as control larvae were bled aseptically to collect the haemolymph. The phenoloxidase activity of the haemolymph samples was estimated by a spectrophotometric method using dopamine as a substrate as described earlier (JIANG *et al.* 2003). The data presented is an average of 4 larvae in each group and statistical analysis was done using t-test.

iii. Noduler-dsRNA (100 μg) or GFP-dsRNA (100 μg) was injected in to day 3, 5th instar larvae and 24 hours later haemolymph was collected aseptically in chilled Grace insect medium (1:1). A total of 10^6 hemocytes were incubated with heat killed FITC-labelled *E. coli* cells (Molecular probes) for 30 mins at 28°C. The fluorescence of extracellular particles was quenched with Trypan blue (0.2 % final concentration) before analysis. The fluorescence intensity of the phagocytosed particles was measured with FACS analysis (BD FACSVANTAGE flow cytometer) and percentage of phagocytosis was determined as described previously (KOCKS *et al.* 2005). Four independent experiments were performed and the average values determined for both Noduler knock-down and control larvae.

4.2.10 Noduler protein expression in nodule extract and tissues

Nodules and different tissues (fat body, mid gut, silk gland and epidermis) were collected from *A. mylitta* larvae challenged with *E. coli* MG1655 at 18 hours pi. A total of 5 mg of the nodules and various tissues were washed with 1X IBS twice to remove haemolymph traces and then were crushed in 1X IBS. Protein estimation was done by Bradford's method and equal amount of protein was loaded on a SDS-PAGE gel followed by immunoblot analysis with Noduler-specific antibody. The band intensity was measured by densitometric analysis. As a loading control the blot was probed with α-tubulin antibody. Also, the presence of *E. coli* MG1655 in the nodules was confirmed by streaking the nodule extract on LB agar plates followed by overnight incubation at 37˚ C. The bacterial colonies were confirmed to be *E. coli* MG1655 by morphological and biochemical analysis.

4.3 Results

4.3.1 A reeler domain spans a major part of Noduler protein

Noduler transcript (GenBank: ABG72705) was recently identified as one of the most abundant ESTs in *A. mylitta* immune transcriptome and its immune upregulation was confirmed by semi-quantitative RT-PCR (GANDHE *et al.* 2006b). Two more highly similar Noduler homologues, Nod homologue 1 (80% similarity at nucleotide and amino acid level) and Nod homologue 2 (83% and 85% similarity at nucleotide and amino acid level, respectively) were also identified in immune transcriptome of *A. mylitta* (Figure 4.1 a). Noduler is a 168 amino acid (aa) protein with a predicted secretory signal at the N-terminal end (Figure 4.1 b). A reeler domain spans almost entire length (aa residues 28-155) of the protein (Figure 4.1 b). Noduler homologues have been reported to be immune upregulated in other lepidopterans (one homologue identified in each of *S. c. ricini*, *L. obliqua*, *B. mori*, *M. sexta* and *H. cunea*) (BAO *et al.* 2003; SHIN *et al.* 1998; VEIGA *et al.* 2005; ZHU *et al.* 2003). Among the other insects, three Noduler homologues were found in *A. gambiae*, four in *D. melanogaster*, two in *D. pseudoobscura*, three in *Aedes aegypti*, three in *Tribolium*

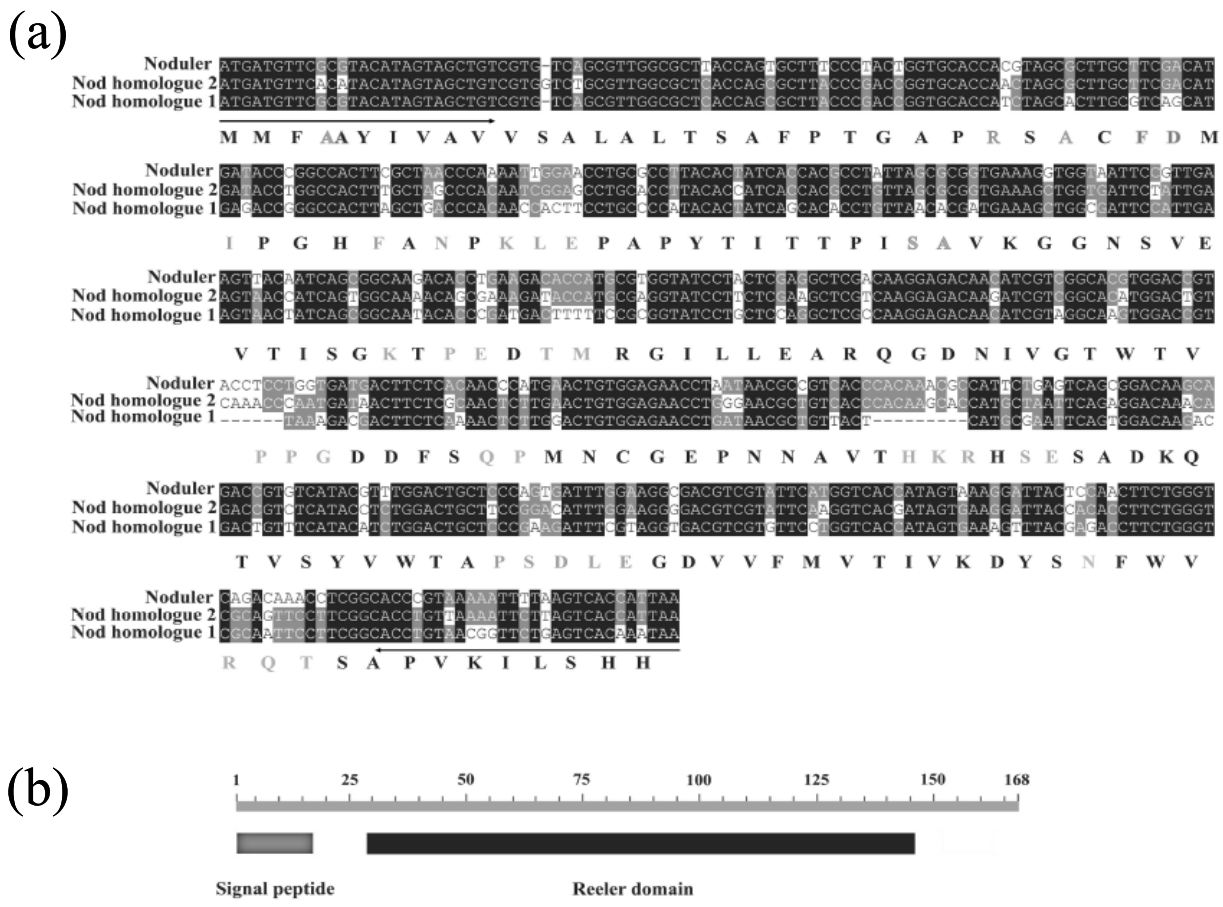

Figure 4.1 Noduler and its two homologues in *Antheraea mylitta* (a) The cDNA sequence comparison of Noduler and its two homologues, Nod homologue 1 and Nod homologue 2. The forward and backward arrows indicate the regions used for designing forward and reverse primers, respectively for preparation of double stranded RNA for RNAi experiments. The full length protein sequence of Noduler is shown below the cDNA sequence alignment. The amino acid residues completely conserved amongst all the three proteins are indicated in black colour while the partially conserved residues are shown in grey colour. (b) Schematic representation of the organization of the Noduler from *A. mylitta*. The scale represents numbered amino acid residues.

castaneum and one in *A. mellifera* (HOLT *et al.* 2002; NENE *et al.* 2007) However, no functional analysis has been done to assign a role for these proteins in immunity. Noduler also showed similarity with many vertebrate extracellular matrix (ECM) proteins that include

stromal cell derived factor receptor 2 (SDR2), ferric chelate reductase (FCR), reelin and many predicted or putative protein sequences. The common feature of all these vertebrate proteins is the presence of an N-terminal reeler domain. The vertebrate proteins showing similarity to Noduler were much larger in size and possessed other domains in addition to reeler unlike insect homologues, which were of similar size with no additional protein domains. Recently, mindin, an ECM protein from mice has been shown to function as a pattern recognition receptor (HE *et al.* 2004). However, mindin possesses an N-terminal spondin domain unlike the reeler domain present in Noduler. A phylogenetic analysis of all the reeler domains from Noduler and its insect and vertebrate homologues was done (Figure 4.2). All the lepidopteran sequences clustered together while the remaining insect sequences formed three different groups. All the mammalian sequences grouped into a separate cluster. A mammalian protein of unknown function, stromal cell derived factor receptor 2 (SDR2), showed highest similarity with Noduler among all the vertebrate proteins. Thus, proteins similar to Noduler seem to be widespread in animal kingdom but their function remains unknown till date.

4.3.2 Noduler is upregulated upon infection in haemolymph

Immunoblot analysis of bacteria challenged haemolymph samples at different time points pi was done with Noduler-specific peptide antibody (Figure 4.3 a). A loading control of the immunoblot samples is indicated in Figure 4.3 b. Noduler was confirmed to be secreted into the haemolymph as predicted from the signal peptide analysis. Noduler was present constitutively in the insect haemolymph and was upregulated several fold upon infection. Upregulation of Noduler was detected as early as 30 mins pi and maximum expression was seen at around 4 hrs pi. A loading control (Figure 4.3 b) of the immunoblot analysis samples indicates intensely stained protein bands around 85 kDa in size that are differentially expressed at different stages of infection. These bands most likely represent the storage proteins (MW ~ 85 kDa) that are accumulated in haemolymph of holometabolous insects like silkworm in the final instar larvae (FUJII *et al.* 1989; SILHACEK *et al.* 1994). Thus the increase in quantity seen in this protein band post infection is rather due to the advancement in

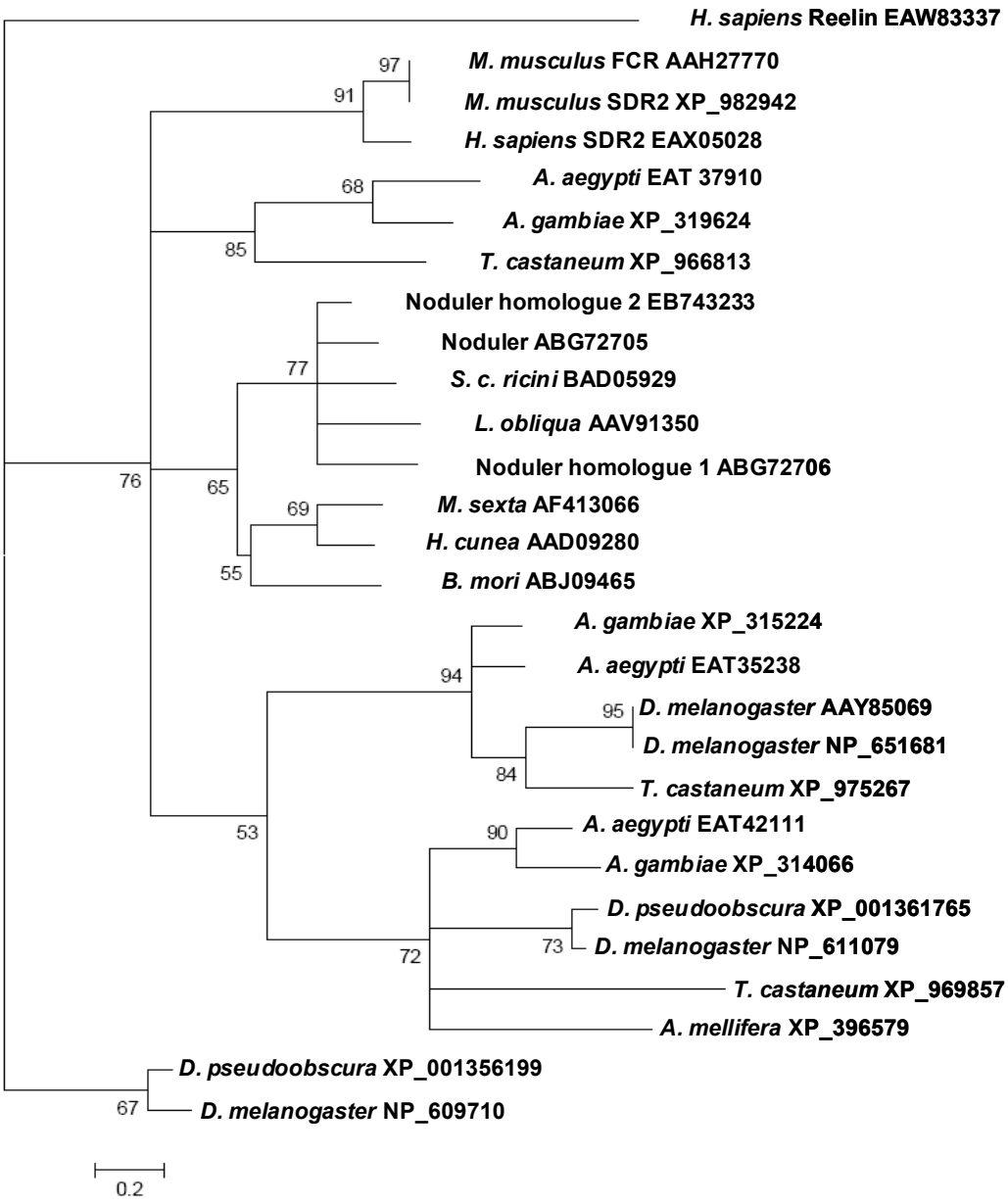

Figure 4.2 The phylogenetic tree of Noduler and its homologues. Phylogenetic comparison of Noduler with similar protein sequences obtained by blastp (E-value≤1e-05). Reelin from *Homo sapiens* was used as an outgroup. Only the reeler domains of the proteins were used for analysis. The bootstrap values are represented as a percentage at the branch nodes. The branch lengths are proportional to the amino acid substitutions. Scale shows 0.2 amino acid substitutions per site. SDR2- stromal cell derived factor receptor 2, FCR-ferric-chelate reductase.

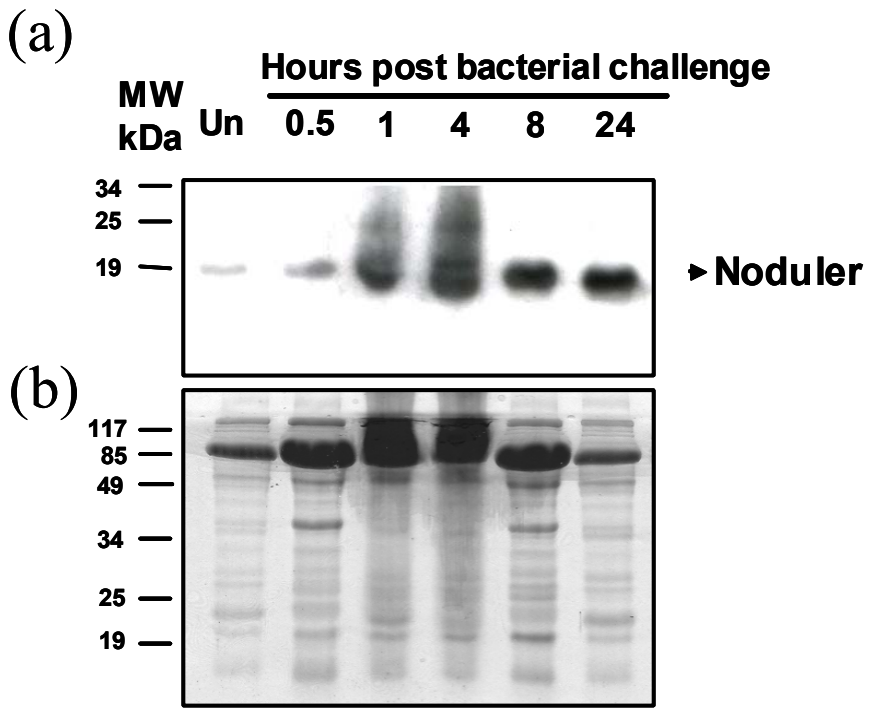

Figure 4.3 Immune upregulation of Noduler in hemolymph. (a) Immunoblot analysis of *Escherichia coli* challenged *Antheraea mylitta* hemolymph samples at different time points post infection using the Noduler-specific antibody. Un- Unchallenged. (b) Coomassie stained gel of the above samples as a loading control for immunoblot analysis. Protein molecular weight (MW) standards are indicated.

developmental stages of the larvae. Other possible candidates present in this region could also be the stress proteins that are synthesized in larval haemolymph post infection (DE MORAIS GUEDES *et al.* 2005).

4.3.3 Recombinant expression of Noduler in Sf9 cells

Noduler was expressed in Sf9 cells by baculovirus mediated expression system as a recombinant protein with an N-terminal 6x histidine tag (Figure 4.4 a). The recombinant Noduler was purified as a native protein by affinity chromatography. The homogeneity and specificity of the purified Noduler was confirmed by SDS-PAGE (Figure 4.4 b) and western blot analysis (Figure 4.4 c). The observed molecular weight of the recombinant Noduler was consistent with the expected one (19 kDa).

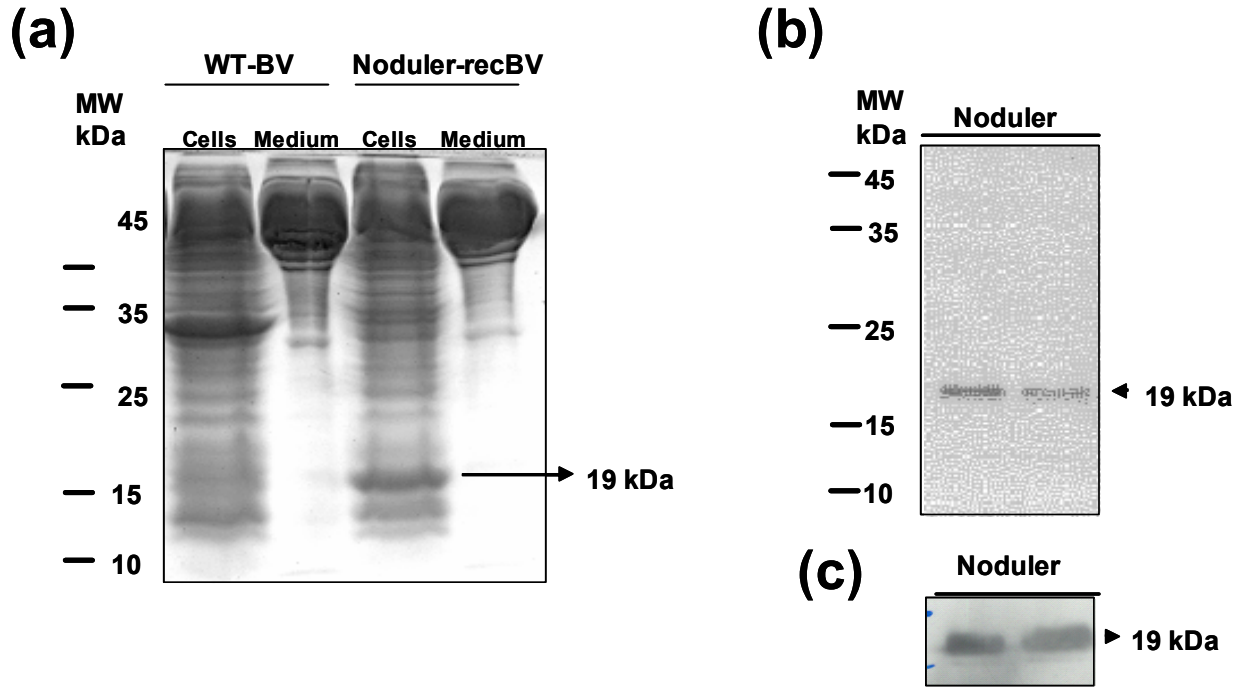

Figure 4.4 Expression and purification of recombinant Noduler expressed in Sf9 cells. (a) Expression of Noduler in Sf9 cells as analysed by SDS-PAGE. Recombinant Noduler expressed as a 19 kDa protein in Sf9 cells infected with recombinant baculoviruses (Noduler-recBV). No expression was seen in Sf9 cells incubated with wild type baculoviruses (WT-BV), which was used as control. (b) SDS-PAGE analysis of recombinant Noduler purified by affinity chromatography. (c) Western blot analysis of purified recombinant Noduler protein with Noduler-specific antibody.

4.3.4 Recombinant Noduler binds to a wide range of microorganisms as well as haemocytes

Several anti-microbial proteins are secreted into haemolymph upon infection. However, Noduler was distinct in that it did not possess anti-bacterial activity as revealed by radial diffusion and growth inhibition assays. To discern its possible role in bacterial recognition, we analysed the binding affinity of Noduler to gram-positive and gram-negative bacteria. Noduler was found to bind gram-positive (*B. subtilis*, *B. megaterium*, *M. luteus* and *S. aureus*) and gram-negative (*E. coli*, *K. pneumoniae*) bacteria tested (Figure 4.5 a). Noduler also bound strongly to yeast (*Saccharomyces cerevisiae*) suggesting that it binds a wide variety of microorganisms (Figure 4.5 a). We further identified the specific microbial cell wall components that bind to Noduler. Noduler was found to strongly bind β-1, 3 glucans of yeast (Figure 4.5 b) and LPS and LTA (Figure 4.5 c) cell wall components in gram-negative and gram-positive bacteria, respectively. However, Noduler did not bind to either gram-

positive (lysine-type) or gram-negative (diaminopimelic acid-type) PGN (Figure 4.5 b) indicating its specificity in binding bacterial ligands. We also demonstrated binding of Noduler to *A. mylitta* haemocytes and a *B. mori* cell line (BmN) (Figure 4.5 b). However, Noduler did not bind to Sf9 cell line derived from *Spodoptera frugiperda*. The prokaryotic heat shock protein used as a negative control in all the binding assays did not bind to any of the microbial ligands or insect cells.

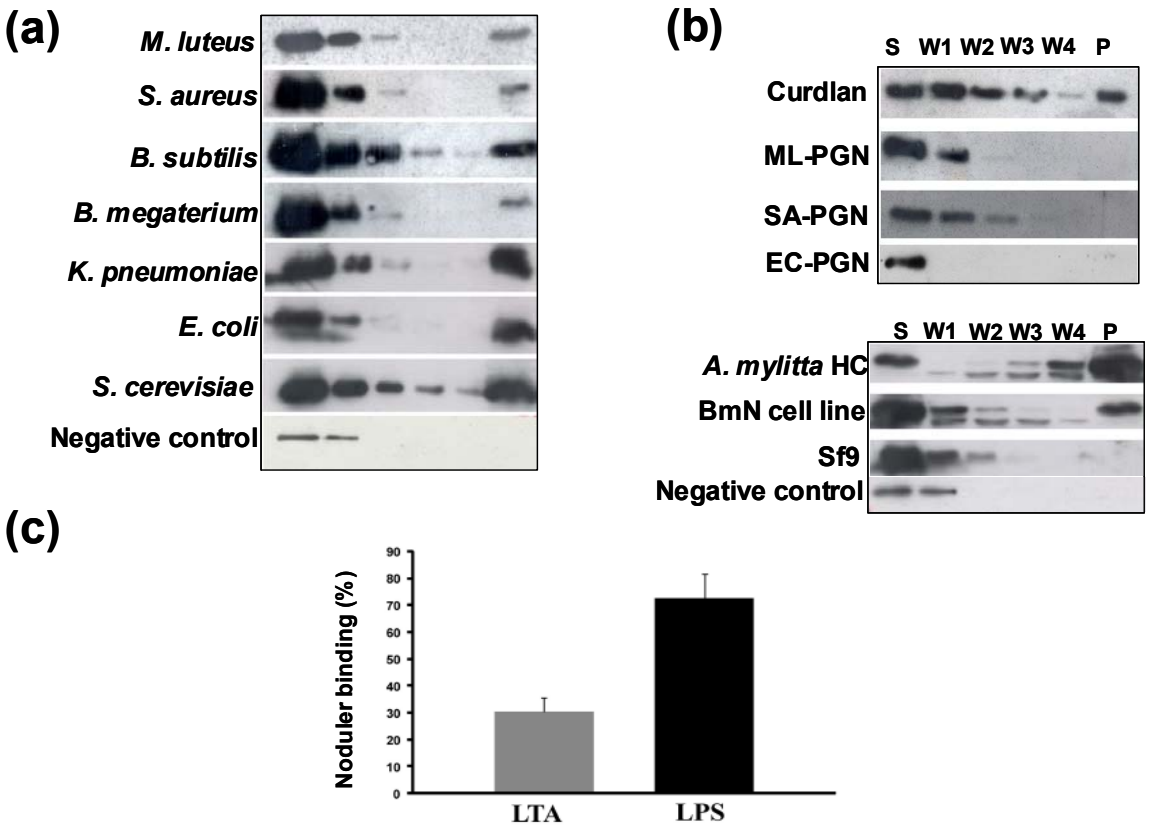

Figure 4.5 Binding assays. (a) Binding of recombinant Noduler to gram-positive bacteria (*Bacillus* sp., *Micrococcus luteus*, and *Staphylococcus aureus*), gram-negative bacteria (*Escherichia. coli*, *Klebsiella pneumoniae*) and yeast (*Saccharomyces cerevesiae*) as analysed by western blot as mentioned in text. W1, W2, W3 and W4 - Wash fractions 1, 2, 3 and 4 respectively, S- supernatant, P- pellet. (b) Binding assay of the recombinant Noduler with curdlan, peptidoglycan (PGN) and insect cells. ML-PGN: PGN from *M. luteus*, SA-PGN: PGN from *S. aureus*. EC-PGN: PGN from *E. coli*, curdlan was from *Alcaligenes faecalis*. *A. mylitta* HC- *A. mylitta* hemocytes. (c) LPS (lipopolysaccharide) and LTA (lipotechoic acid) binding assay of Noduler. LPS (LPS from *E. coli* 0111:4) was purified to remove possible PGN contamination and Noduler binding was determined by a competition assay as described in the text. Y-axis indicates the extent of Noduler binding to LPS or LTA as a percentage of the total amount of the Noduler. LTA (LTA from *S. aureus*) binding assay of Noduler was carried out similar to LPS binding assay. A histidine tagged Hsp70 protein was used as a negative control in all the binding assays.

4.3.5 In vivo knock-down of Noduler by RNAi results in increased bacterial load

The in vivo role of Noduler in immunity was assessed by this experiment. Noduler protein was knocked-down in *A. mylitta* 5th instar, day 3 larvae by injection of Noduler-dsRNA followed by gram-positive (*S. aureus*) and gram-negative (*E. coli*) bacteria independently. Haemolymph bacterial load was estimated 30 mins pi. Around 10 fold higher staphylococcal load ($P < 0.001$) was observed in haemolymph of Noduler knock-down larvae in comparison with the control (Figure 4.6 a). *E. coli* load in haemolymph was found to be 4-5 fold higher ($P < 0.001$) in Noduler knock-down larvae (Figure 4.6 b) as compared to control larvae (saline injected or GFP-dsRNA injected). These results established the in vivo immune related function of Noduler protein and suggested its involvement in early immune mechanisms activated as early as 30 minutes pi. Western blot analysis of haemolymph from Noduler-dsRNA injected larvae showed a complete depletion of Noduler as compared to GFP-dsRNA injected or saline-injected larvae (Figure 4.6 c) indicating that dsRNA mediated knock-down of Noduler can be achieved successfully in silkmoths. To our knowledge, other than phenoloxidase, so far no other molecules related to immune response is identified to be actively involved at very early stage of bacterial infection.

4.3.6 Noduler is involved in nodulation responses against bacteria and phenoloxidase pathway

Further, we analysed the effect of Noduler knock-down on immediate early immune mechanisms of phagocytosis, phenoloxidase cascade and nodulation. Noduler knock-down larvae showed a drastically reduced number of nodules as compared with the control larvae ($P < 0.001$). A three to five fold reduction in the number of nodules per insect upon infection with *S. aureus* (Figure 4.7 a and c) or *E. coli* (Figure 4.7 b and c) was observed in the Noduler depleted larvae as compared to saline injected or GFP-dsRNA injected larvae. Phenoloxidase pathway is known to be involved in nodulation mechanism and is responsible for the melanisation of nodules. Hence the role of Noduler-depletion on the phenoloxidase activity was also analysed. Decrease of phenoloxidase activity in the haemolymph of Noduler knock-down larvae challenged with *E. coli* was significant ($P < 0.001$) compared to control larvae (Figure 4.8). This suggests that Noduler is indeed involved in activation of the phenoloxidase pathway. Noduler knock-down larvae did not show any difference in

phagocytosis activity as compared to the control (GFP-dsRNA injected) larvae ruling out its involvement in phagocytosis (Figure 4.9).

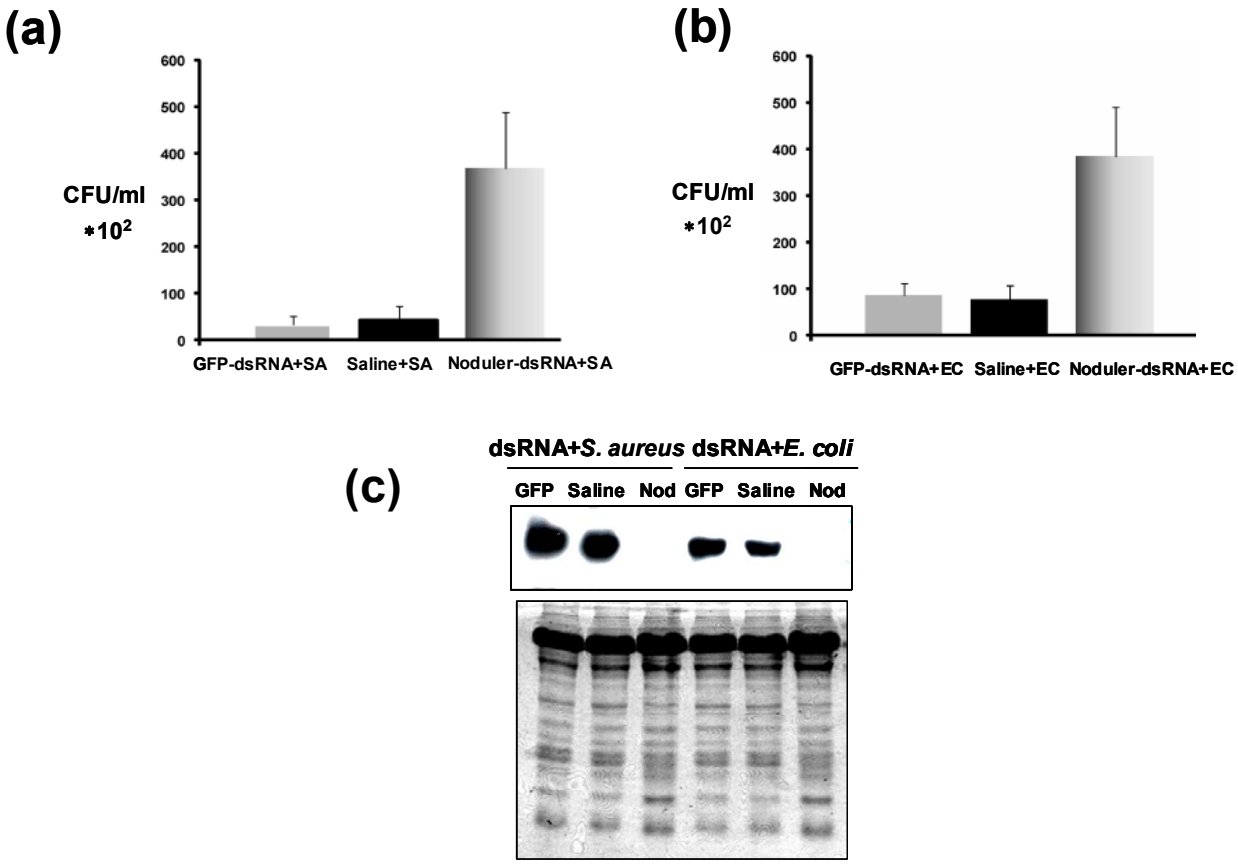

Figure 4.6 Bacteria clearance assay. RNAi mediated Noduler knock-down larvae exhibited decreased clearance of *Staphylococcus aureus* (a) and *Escherichia coli* (b) injected into 5th instar larval haemocoel compared to GFP-dsRNA or saline injected larvae. Y-axis shows the average of colony forming units (CFU)/ml of hemolymph (n=8) and error bars indicate standard deviation of three independent experiments. The statistical significance was estimated by t-test (P < 0.001). SA- *S. aureus*, EC- *E. coli*. (c) Molecular analysis of RNAi mediated knock-down of Noduler. Western analysis of hemolymph samples from Noduler-dsRNA injected larvae showed complete knock-down of Noduler protein as compared to GFP-dsRNA injected or saline injected larvae. Lower panel shows the Coomassie stained SDS-PAGE gel of hemolymph samples utilized for western blot as a loading control. Nod-Noduler.

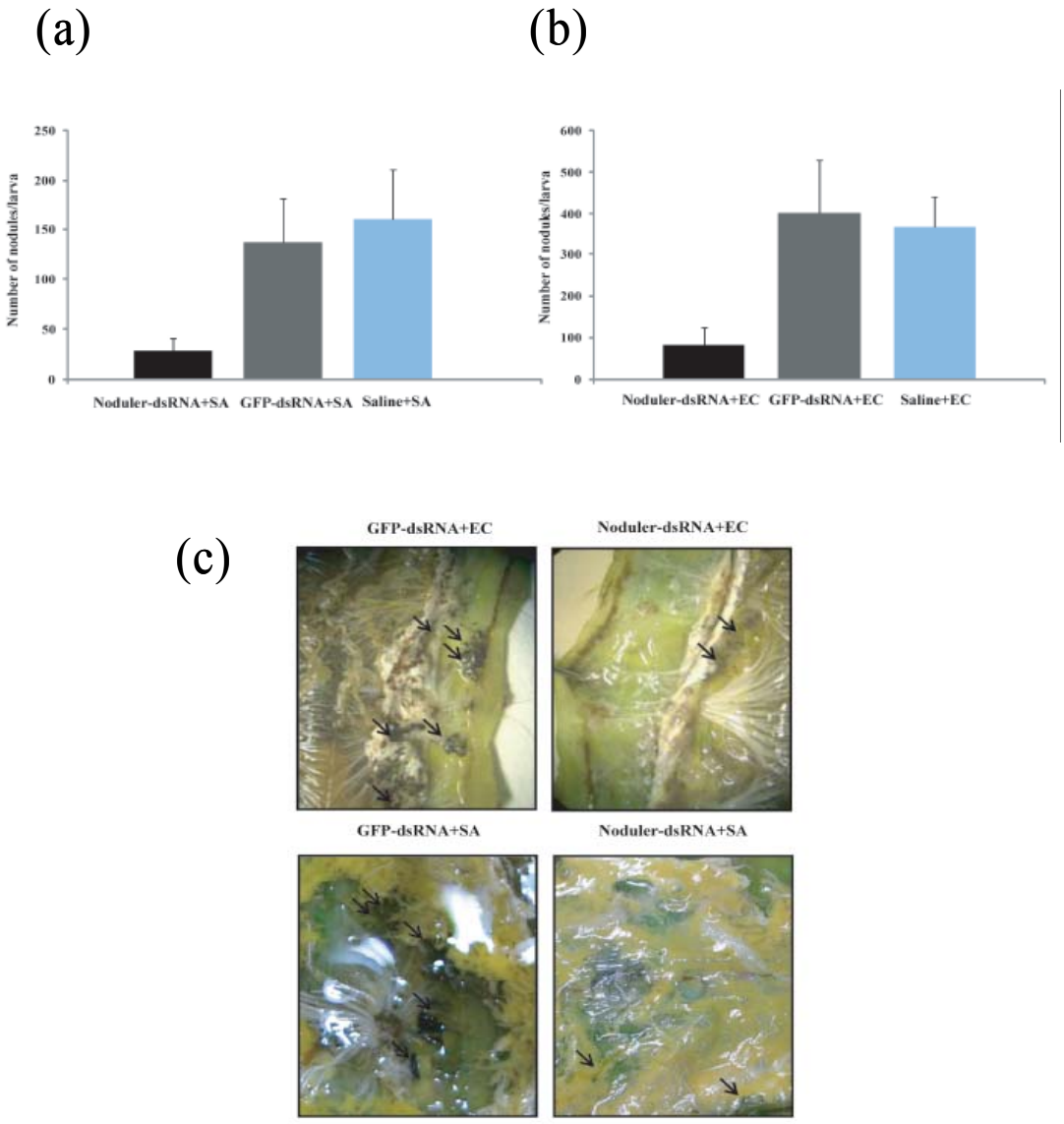

Figure 4.7 Effect of RNAi mediated Noduler depletion on nodulation response.

Nodulation was reduced significantly (P<0.001) in larvae injected with Noduler-dsRNA prior to *Staphylococcus aureus* (a) or *Escherichia coli* (b) injection as compared to GFP-dsRNA or saline injected larvae prior to bacterial injection. The average number of nodules per larva (n=8) in each group is plotted on Y-axis and the error bars represent standard deviation of three independent experiments. (c) Formation of melanised nodules in Noduler knock-down (Noduler-dsRNA+ *S. aureus*/*E. coli*) and control larvae (GFP-dsRNA + *S. aureus*/*E. coli*). The black arrows indicate the nodules. SA- *S. aureus*, EC- *E. coli*.

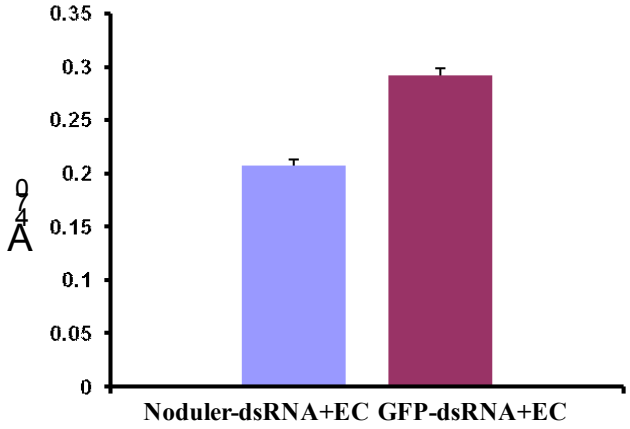

Figure 4.8 Effect of Noduler knock-down on phenoloxidase activity in the 5th instar larval haemolymph. Y-axis represents the absorbance at 470 nm which is directly proportional to phenoloxidase activity and error bars indicate the standard deviation (n=4). Noduler knock-down larvae showed a significant reduction (P < 0.001) in phenoloxidase activity as compared to control (GFP-dsRNA injected) larvae. EC- *E. coli*.

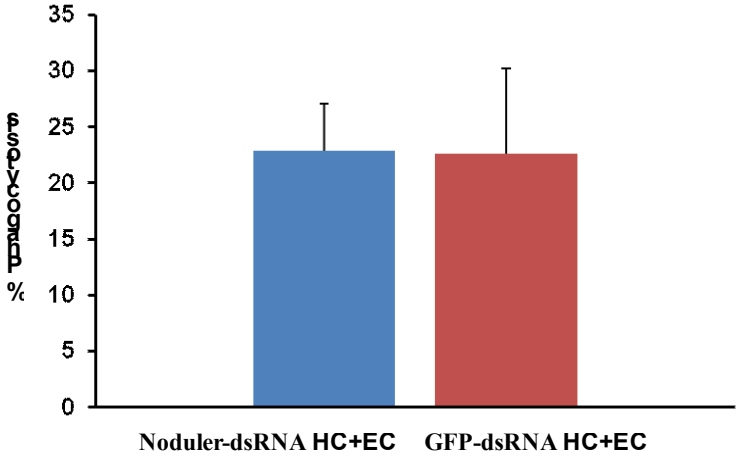

Figure 4.9 Effect of Noduler depletion on phagocytosis. In vitro phagocytosis assay was carried out upon haemocytes (HC) collected from Noduler-dsRNA injected and GFP-dsRNA injected 5th instar larvae. Haemocytes were incubated with FITC-labelled *E. coli* (EC) cells and % phagocytosis was determined by FACS analysis. Y-axis denotes the % phagocytosis activity of haemocytes and the error bars indicate standard deviation (n=4). No significant difference in phagocytosis was seen between Noduler depleted and control haemocytes.

4.3.7 Noduler protein promotes the assembly of bacteria and haemocytes to form nodules

To ascertain that Noduler is indeed involved in the formation of nodules, Noduler protein in nodules was analysed by SDS-PAGE of nodule extract along with other tissues (silk gland, mid gut, epidermis and fat body) from bacteria challenged larvae and probed with Noduler-specific antibody (Figure 4.10 a). Noduler was indeed found to be concentrated in the nodules and was at four-fold higher concentration as compared to fat body, which is a major organ of immune peptide synthesis (Figure 4.10 a). Among the other tissues analysed, Noduler was not detected in silk gland or epidermis though a faint band of slightly higher molecular weight was observed in mid gut, which could be the pro-Noduler with intact signal peptide. The presence of the injected bacterial strain (*E. coli* MG1655) was also confirmed in the nodules by streaking the nodule extract on a LB plate. The bacterial colonies obtained on the LB plate after overnight incubation were confirmed to be *E. coli* MG1655 by gram staining, morphology and biochemical methods.

4.3.8 Noduler probably traps bacteria and haemocytes during nodulation

Our experimental data suggest that Noduler protein is a facilitator of nodulation. We propose that Noduler initiates nodulation by binding both microorganisms as well as host haemocytes. The enhanced concentration of Noduler in the nodules, sequestration of bacteria in nodules, binding of Noduler to bacteria and haemocytes and significant reduction in number of nodules upon Noduler knock-down all of them suggest that Noduler is a key molecule in nodulation. Its co-occurrence in nodules along with haemocytes and bacteria further strengthens this proposal. A schematic representation of the proposed mechanism of Noduler action is shown in Figure 4.10 b.

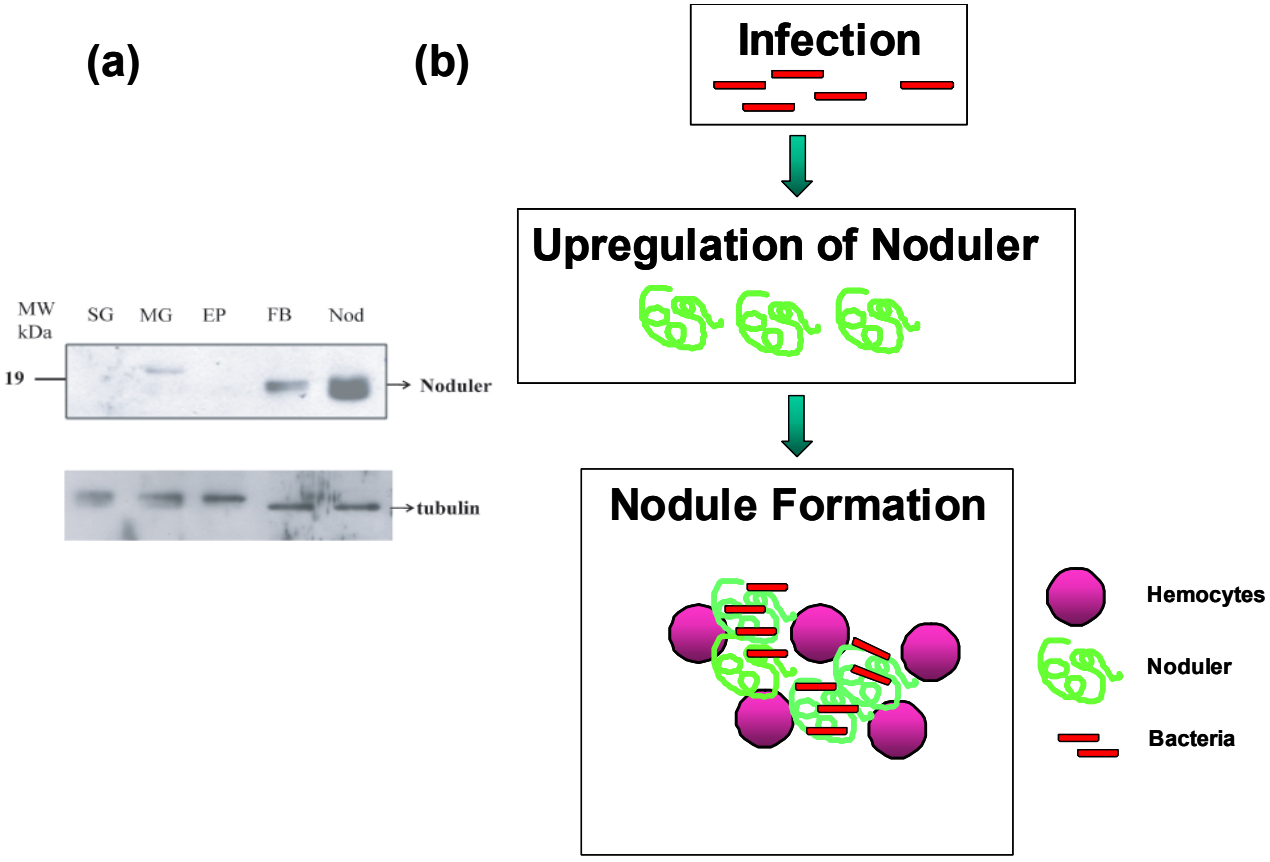

Figure 4.10 Noduler protein is concentrated in nodules. (a) Immunoblot analysis of nodule and other tissue extracts with Noduler-specific antibody. Nodules and various tissues were collected 18-hours post *Escherichia coli* injection and the extracts were prepared as indicated in text. Equal amount of total protein (21 μg) from each tissue was loaded. The loading control (α-tubulin) is shown in lower panel. SG- silk gland, MG- mid gut, EP- epidermis, FB- fat body, NOD- nodules. (b) The suggested role for Noduler in the formation of nodules.

4.4 Discussion

Nodulation is a predominant cellular defence mechanism in insects and other invertebrates against diverse pathogens (LORD *et al.* 2002; MILLER *et al.* 1994). However it is very poorly understood as yet in comparison with other immune responses. The only information available so far is from eicosanoids which are shown to mediate nodulation in many insect species (BEDICK *et al.* 2001; FRANSSENS *et al.* 2005; MILLER *et al.* 1994). Role of eicosanoids is well documented in mammalian immunity and they have shown to have profound effects on macrophage locomotion, phagocytosis and cell shape changes (MILLER *et al.* 1994) and probably exert similar effects on insect haemocytes. However eicosanoids are signalling molecules not specific to nodulation alone but mediate other cellular and humoral responses in insects as well (MILLER *et al.* 1994). In the present study we have identified a key player that distinctively mediates nodulation responses in insects.

Haemocyte aggregation, the first step in nodulation, results in entrapment of the invading microorganisms (LAVINE and STRAND 2002). Such an event requires recognition molecules that interact with both haemocytes as well as invading microorganisms. We hypothesize that Noduler mediates nodulation by virtue of its binding property to both bacteria as well as haemocytes. Our hypothesis stems from the following experimental observations: Upregulation of Noduler within minutes of bacterial infection, reduction in number of nodules upon Noduler silencing and subsequent increase in bacterial load, binding of Noduler to bacteria, yeast and insect haemocytes, and the co-occurrence of bacteria, Noduler and haemocytes in the nodules. However, we did not observe haemocyte aggregation under in vitro conditions upon addition of exogenous Noduler and bacteria to haemocytes suggesting that additional players are probably involved in this mechanism. Also, since nodules are generally encountered adhering to the tissues, the phenomenon probably requires an in vivo milieu in the form of a tissue support. Nodules generally appear dark in colour due to phenoloxidase activity at site of nodule formation (LAVINE and STRAND 2002). A significant reduction in the overall phenoloxidase activity in the haemolymph upon Noduler silencing was observed. Earlier reports have shown that insect pattern recognition proteins (PGRPs, lectins, βGRPs, hemolin) bind to microbial cell wall components and trigger responses such as phagocytosis, nodulation and activation of phenoloxidase cascade. Noduler was indeed shown to bind microbial cell wall components and was involved in nodulation as well as

phenoloxidase cascade mechanisms. However, as shown in the present study, Noduler did not affect phagocytosis.While its role in nodulation could be due to its ability to bind bacteria as well as haemocytes, its exact mechanism of activating the phenoloxidase cascade remains to be studied in future. Immune response of encapsulation, which refers to aggregation of haemocytes around large size invaders like parasites, is in principle similar to nodulation (LAVINE and STRAND 2002). Hence, it is quite possible that Noduler is involved in encapsulation reactions and thus merits further study.

Two additional homologues of Noduler were also identified in *A. mylitta*, which is consistent with previous observations of gene expansion of immune gene families in insects, potentially enabling diversified pathogen defence (CHENG *et al.* 2006; CHRISTOPHIDES *et al.* 2002). Both the Noduler homologues are highly similar to Noduler suggesting that they are involved in a similar function. Furthermore, high level of sequence similarity at both nucleotide and amino acid level makes the detection of effect of knock-down of each of the three Noduler family proteins in *A. mylitta* very difficult at molecular level. Thus it seems most likely that the RNAi mediated phenotype of Noduler in the present study is a result of knock-down of all the three Noduler family proteins. Noduler-like proteins have been identified to be immune upregulated in other insects but their function was unknown (SHIN *et al.* 1998). Kanost *et. al.* observed that the Noduler-like protein from locust copurified with apolipohorin-III, a plasma protein identical in size and charge (ZHU *et al.* 2003). He further adds that the immune role attributed to apolipophorin-III (WHITTEN *et al.* 2004) might be due to the copurified Noduler-like proteins. Our experimental results have indeed shown Noduler to have an immune function and hence the immune role of apolipophorin-III needs to be revisited. In the present study, Noduler was purified from a recombinant expression system thus ruling out the possibility of a contaminating haemolymph protein mediating the effects. Also, we have shown the functional role of Noduler by in vivo RNAi based experiments, which further confirms its role in nodulation and activation of phenoloxidase cascade.

A detailed understanding of insect immune system assumes importance for designing novel disease control strategies against insect-vector borne diseases. Towards this goal, several large-scale genome and transcriptome projects have been initiated in insects in the past decade (ABRAHAM *et al.* 2004; CHRISTOPHIDES *et al.* 2002). Noduler was found to be evolutionarily conserved amongst insects and it would be worthwhile to analyse its role in other insects. Several parallels have been observed between insect and mammalian innate

immune pathways recently. However, nodulation like response is not reminiscent of any of the vertebrate innate immune mechanisms. Nevertheless, several mammalian extra cellular matrix proteins of unknown function show resemblance with Noduler. Thus it will be interesting to explore their role in vertebrate immune pathways in view of the results presented here.

Chapter-5

Concluding remarks

Concluding remarks

Diverse organisms like insects and mammals are now known to share many molecular, cellular and developmental processes. A prominent example is the evolutionary conservation of innate immune system amongst the vertebrates and invertebrates. Due to the absence of adaptive immune system, insects provide an ideal model to study innate immune mechanisms in higher organisms. Innate immunity refers to the first line of defence against infections that comprises physical, molecular and cellular responses. Insects rely on the humoral and cellular responses for protection against pathogenic microorganisms. The present study addresses to identify immune genes from the relatively less studied lepidopteran wild silkmoth *A. mylitta* and the domesticated silkmoth *B. mori*. Several potential hitherto unreported immune genes from *A. mylitta* were found in the immune challenged fat body transcriptome in addition to the known genes. A subset of genes was confirmed to be upregulated upon infection. A novel class of lysozyme-like proteins which lack the muramidase activity but still exhibit anti-bacterial property were identified in both the silkmoths. This adds to yet another member of anti-microbial proteins in insect immune repertoire. Lysozyme-like proteins exhibit bacteriostatic action which can be rescued by adding external peptidoglycan. Noduler, another novel immune induced protein from *A. mylitta* binds to both gram-positive and gram-negative bacteria and insect haemocytes. Noduler is a novel pattern recognition protein in insects and triggers nodulation and activation of phenoloxidase cascade upon binding to microorganisms. Noduler homologues are present all across insect orders and also in mammals. The immune repertoire of *A. mylitta* and the subsequent identification and functional characterization of anti-bacterial lysozyme-like proteins and Noduler protein comprise the major findings of this study, revealing new aspects of insect immunity.

References

References

ABRAHAM, E. G., S. ISLAM, P. SRINIVASAN, A. K. GHOSH, J. G. VALENZUELA *et al.*, 2004 Analysis of the Plasmodium and Anopheles transcriptional repertoire during ookinete development and midgut invasion. J Biol Chem **279:** 5573-5580.

ABRAHAM, E. G., J. NAGARAJU, D. SALUNKE, H. M. GUPTA and R. K. DATTA, 1995 Purification and partial characterization of an induced antibacterial protein in the silkworm, Bombyx mori. J Invertebr Pathol **65:** 17-24.

AGUILAR, R., A. E. JEDLICKA, M. MINTZ, V. MAHAIRAKI, A. L. SCOTT *et al.*, 2005 Global gene expression analysis of Anopheles gambiae responses to microbial challenge. Insect Biochem Mol Biol **35:** 709-719.

ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFFER, J. ZHANG, Z. ZHANG *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:** 3389-3402.

AXEN, A., A. CARLSSON, A. ENGSTROM and H. BENNICH, 1997 Gloverin, an antibacterial protein from the immune hemolymph of Hyalophora pupae. Eur J Biochem **247:** 614-619.

BAO, Y., K. MEGA, Y. YAMANO and I. MORISHIMA, 2003 cDNA cloning and expression of bacteria-induced Hdd11 gene from eri-silkworm, Samia cynthia ricini. Comp Biochem Physiol C Toxicol Pharmacol **136:** 337-342.

BEDICK, J. C., H. TUNAZ, A. R. NOR ALIZA, S. M. PUTNAM, M. D. ELLIS *et al.*, 2001 Eicosanoids act in nodulation reactions to bacterial infections in newly emerged adult honey bees, Apis mellifera, but not in older foragers. Comp Biochem Physiol C Toxicol Pharmacol **130:** 107-117.

BENDTSEN, J. D., H. NIELSEN, G. VON HEIJNE and S. BRUNAK, 2004 Improved prediction of signal peptides: SignalP 3.0. J Mol Biol **340:** 783-795.

BETTENCOURT, R., O. TERENIUS and I. FAYE, 2002 Hemolin gene silencing by ds-RNA injected into Cecropia pupae is lethal to next generation embryos. Insect Mol Biol **11:** 267-271.

BOMAN, H. G., I. FAYE, P. VON HOFSTEN, K. KOCKUM, J. Y. LEE *et al.*, 1985 On the primary structures of lysozyme, cecropins and attacins from Hyalophora cecropia. Dev Comp Immunol **9:** 551-558.

BOMAN, H. G., I. NILSSON-FAYE, K. PAUL and T. RASMUSON, JR., 1974 Insect immunity. I. Characteristics of an inducible cell-free antibacterial reaction in hemolymph of Samia cynthia pupae. Infect Immun **10:** 136-145.

BRADFORD, M. M., 1976 A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem **72:** 248-254.

BRU, C., E. COURCELLE, S. CARRERE, Y. BEAUSSE, S. DALMAR *et al.*, 2005 The ProDom database of protein domain families: more emphasis on 3D. Nucleic Acids Res **33:** D212-215.

BULET, P., C. HETRU, J. L. DIMARCQ and D. HOFFMANN, 1999 Antimicrobial peptides in insects; structure and function. Dev Comp Immunol **23:** 329-344.

BULET, P., and R. STOCKLIN, 2005 Insect antimicrobial peptides: structures, properties and gene regulation. Protein Pept Lett **12:** 3-11.

BULET, P., R. STOCKLIN and L. MENIN, 2004 Anti-microbial peptides: from invertebrates to vertebrates. Immunol Rev **198:** 169-184.

CARLSSON, A., P. ENGSTROM, E. T. PALVA and H. BENNICH, 1991 Attacin, an antibacterial protein from Hyalophora cecropia, inhibits synthesis of outer membrane proteins in

Escherichia coli by interfering with omp gene transcription. Infect Immun **59:** 3040-3045.

CASTLE, M., A. NAZARIAN, S. S. YI and P. TEMPST, 1999 Lethal effects of apidaecin on Escherichia coli involve sequential molecular interactions with diverse targets. J Biol Chem **274:** 32555-32564.

CERENIUS, L., and K. SODERHALL, 2004 The prophenoloxidase-activating system in invertebrates. Immunol Rev **198:** 116-126.

CHENG, T., P. ZHAO, C. LIU, P. XU, Z. GAO *et al.*, 2006 Structures, regulatory regions, and inductive expression patterns of antimicrobial peptide genes in the silkworm Bombyx mori. Genomics **87:** 356-365.

CHENG, T. C., Y. L. ZHANG, C. LIU, P. Z. XU, Z. H. GAO *et al.*, 2007 Identification and analysis of Toll-related genes in the domesticated silkworm, Bombyx mori. Dev Comp Immunol.

CHERRY, S., and N. SILVERMAN, 2006 Host-pathogen interactions in drosophila: new tricks from an old friend. Nat Immunol **7:** 911-917.

CHRISTOPHIDES, G. K., E. ZDOBNOV, C. BARILLAS-MURY, E. BIRNEY, S. BLANDIN *et al.*, 2002 Immunity-related genes and gene families in Anopheles gambiae. Science **298:** 159-165.

COCIANCICH, S., A. GHAZI, C. HETRU, J. A. HOFFMANN and L. LETELLIER, 1993 Insect defensin, an inducible antibacterial peptide, forms voltage-dependent channels in Micrococcus luteus. J Biol Chem **268:** 19239-19245.

DE GREGORIO, E., P. T. SPELLMAN, G. M. RUBIN and B. LEMAITRE, 2001 Genome-wide analysis of the Drosophila immune response by using oligonucleotide microarrays. Proc Natl Acad Sci U S A **98:** 12590-12595.

DE MORAIS GUEDES, S., R. VITORINO, R. DOMINGUES, K. TOMER, A. J. CORREIA *et al.*, 2005 Proteomics of immune-challenged Drosophila melanogaster larvae hemolymph. Biochem Biophys Res Commun **328:** 106-115.

DONG, Y., H. E. TAYLOR and G. DIMOPOULOS, 2006 AgDscam, a Hypervariable Immunoglobulin Domain-Containing Receptor of theAnopheles gambiae Innate Immune System. PLoS Biol **4:** e229.

DOSTERT, C., E. JOUANGUY, P. IRVING, L. TROXLER, D. GALIANA-ARNOUX *et al.*, 2005 The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of drosophila. Nat Immunol **6:** 946-953.

DUSHAY, M. S., J. B. ROETHELE, J. M. CHAVERRI, D. E. DULEK, S. K. SYED *et al.*, 2000 Two attacin antibacterial genes of Drosophila melanogaster. Gene **246:** 49-57.

ELEFTHERIANOS, I., F. GOKCEN, G. FELFOLDI, P. J. MILLICHAP, T. E. TRENCZEK *et al.*, 2007 The immunoglobulin family protein Hemolin mediates cellular immune responses to bacteria in the insect Manduca sexta. Cell Microbiol **9:** 1137-1147.

EVANS, J. D., K. ARONSTEIN, Y. P. CHEN, C. HETRU, J. L. IMLER *et al.*, 2006 Immune pathways and defence mechanisms in honey bees Apis mellifera. Insect Mol Biol **15:** 645-656.

EWING, B., L. HILLIER, M. C. WENDL and P. GREEN, 1998 Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res **8:** 175-185.

FOLEY, E., and P. H. O'FARRELL, 2004 Functional dissection of an innate immune response by a genome-wide RNAi screen. PLoS Biol **2:** E203.

FRANSSENS, V., G. SIMONET, A. BRONCKAERS, I. CLAEYS, A. DE LOOF *et al.*, 2005 Eicosanoids mediate the laminarin-induced nodulation response in larvae of the flesh fly, Neobellieria bullata. Arch Insect Biochem Physiol **59:** 32-41.

FUJII, T., H. SAKURAI, S. IZUMI and S. TOMINO, 1989 Structure of the gene for the arylphorin-type storage protein SP 2 of Bombyx mori. J Biol Chem **264:** 11020-11025.

GANDHE, A. S., K. P. ARUNKUMAR, S. H. JOHN and J. NAGARAJU, 2006a Analysis of bacteria-challenged wild silkmoth, Antheraea mylitta (lepidoptera) transcriptome reveals potential immune genes. BMC Genomics **7:** 184.

GANDHE, A. S., A. K. P, S. H. JOHN and N. J, 2006b Analysis of bacteria-challenged wild silkmoth, Antheraea mylitta (Lepidoptera) transcriptome reveals potential immune genes. BMC Genomics **7:** 184.

GLAUNER, B., 1988 Separation and quantification of muropeptides with high-performance liquid chromatography. Anal Biochem **172:** 451-464.

GUDMUNDSSON, G. H., D. A. LIDHOLM, B. ASLING, R. GAN and H. G. BOMAN, 1991 The cecropin locus. Cloning and expression of a gene cluster encoding three antibacterial peptides in Hyalophora cecropia. J Biol Chem **266:** 11510-11517.

HARA, S., and M. YAMAKAWA, 1995a Moricin, a novel type of antibacterial peptide isolated from the silkworm, Bombyx mori. J Biol Chem **270:** 29923-29927.

HARA, S., and M. YAMAKAWA, 1995b A novel antibacterial peptide family isolated from the silkworm, Bombyx mori. Biochem J **310 (Pt 2):** 651-656.

HE, Y. W., H. LI, J. ZHANG, C. L. HSU, E. LIN *et al.*, 2004 The extracellular matrix protein mindin is a pattern-recognition molecule for microbial pathogens. Nat Immunol **5:** 88-97.

HEDENGREN-OLCOTT, M., M. C. OLCOTT, D. T. MOONEY, S. EKENGREN, B. L. GELLER *et al.*, 2004 Differential activation of the NF-kappaB-like factors Relish and Dif in Drosophila melanogaster by fungi and Gram-positive bacteria. J Biol Chem **279:** 21121-21127.

HIRSCHFELD, M., Y. MA, J. H. WEIS, S. N. VOGEL and J. J. WEIS, 2000 Cutting edge: repurification of lipopolysaccharide eliminates signaling through both human and murine toll-like receptor 2. J Immunol **165:** 618-622.

HOFFMANN, J. A., 2003 The immune response of Drosophila. Nature **426:** 33-38.

HOFFMANN, J. A., F. C. KAFATOS, C. A. JANEWAY and R. A. EZEKOWITZ, 1999 Phylogenetic perspectives in innate immunity. Science **284:** 1313-1318.

HOLT, R. A., G. M. SUBRAMANIAN, A. HALPERN, G. G. SUTTON, R. CHARLAB *et al.*, 2002 The genome sequence of the malaria mosquito Anopheles gambiae. Science **298:** 129-149.

HULTMARK, D., 1996 Insect lysozymes. Exs **75:** 87-102.

HULTMARK, D., A. ENGSTROM, K. ANDERSSON, H. STEINER, H. BENNICH *et al.*, 1983 Insect immunity. Attacins, a family of antibacterial proteins from Hyalophora cecropia. Embo J **2:** 571-576.

HULTMARK, D., A. ENGSTROM, H. BENNICH, R. KAPUR and H. G. BOMAN, 1982 Insect immunity: isolation and structure of cecropin D and four minor antibacterial components from Cecropia pupae. Eur J Biochem **127:** 207-217.

HULTMARK, D., H. STEINER, T. RASMUSON and H. G. BOMAN, 1980 Insect immunity. Purification and properties of three inducible bactericidal proteins from hemolymph of immunized pupae of Hyalophora cecropia. Eur J Biochem **106:** 7-16.

IBRAHIM, H. R., T. MATSUZAKI and T. AOKI, 2001 Genetic evidence that antibacterial activity of lysozyme is independent of its catalytic function. FEBS Lett **506:** 27-32.

IMAMURA, M., and M. YAMAKAWA, 2002 Molecular cloning and expression of a Toll receptor gene homologue from the silkworm, Bombyx mori. Biochim Biophys Acta **1576:** 246-254.

JAIN, D., D. T. NAIR, G. J. SWAMINATHAN, E. G. ABRAHAM, J. NAGARAJU *et al.*, 2001 Structure of the induced antibacterial protein from tasar silkworm, Antheraea mylitta. Implications to molecular evolution. J Biol Chem **276:** 41377-41382.

JIANG, H., and M. R. KANOST, 1997 Characterization and functional analysis of 12 naturally occurring reactive site variants of serpin-1 from Manduca sexta. J Biol Chem **272:** 1082-1087.

JIANG, H., and M. R. KANOST, 2000 The clip-domain family of serine proteinases in arthropods. Insect Biochem Mol Biol **30:** 95-105.

JIANG, H., Y. WANG and M. R. KANOST, 1998 Pro-phenol oxidase activating proteinase from an insect, Manduca sexta: a bacteria-inducible protein similar to Drosophila easter. Proc Natl Acad Sci U S A **95:** 12220-12225.

JIANG, H., Y. WANG, X. Q. YU and M. R. KANOST, 2003 Prophenoloxidase-activating proteinase-2 from hemolymph of Manduca sexta. A bacteria-inducible serine proteinase containing two clip domains. J Biol Chem **278:** 3552-3561.

JOHANSSON, K. C., C. METZENDORF and K. SODERHALL, 2005 Microarray analysis of immune challenged Drosophila hemocytes. Exp Cell Res **305:** 145-155.

KAMBRIS, Z., S. BRUN, I. H. JANG, H. J. NAM, Y. ROMEO *et al.*, 2006 Drosophila immunity: a large-scale in vivo RNAi screen identifies five serine proteases required for Toll activation. Curr Biol **16:** 808-813.

KANG, D., G. LIU, H. GUNNE and H. STEINER, 1996 PCR differential display of immune gene expression in Trichoplusia ni. Insect Biochem Mol Biol **26:** 177-184.

KANG, D., G. LIU, A. LUNDSTROM, E. GELIUS and H. STEINER, 1998 A peptidoglycan recognition protein in innate immunity conserved from insects to humans. Proc Natl Acad Sci U S A **95:** 10078-10082.

KANOST, M. R., 1999 Serine proteinase inhibitors in arthropod immunity. Dev Comp Immunol **23:** 291-301.

KANOST, M. R., H. JIANG and X. Q. YU, 2004 Innate immune responses of a lepidopteran insect, Manduca sexta. Immunol Rev **198:** 97-105.

KIM, Y. S., J. H. RYU, S. J. HAN, K. H. CHOI, K. B. NAM *et al.*, 2000 Gram-negative bacteria-binding protein, a pattern recognition receptor for lipopolysaccharide and beta-1,3-glucan that mediates the signaling for the induction of innate immune genes in Drosophila melanogaster cells. J Biol Chem **275:** 32721-32727.

KOCKS, C., J. H. CHO, N. NEHME, J. ULVILA, A. M. PEARSON *et al.*, 2005 Eater, a transmembrane protein mediating phagocytosis of bacterial pathogens in Drosophila. Cell **123:** 335-346.

KOIZUMI, N., M. IMAMURA, T. KADOTANI, K. YAOI, H. IWAHANA *et al.*, 1999 The lipopolysaccharide-binding protein participating in hemocyte nodule formation in the silkworm Bombyx mori is a novel member of the C-type lectin superfamily with two different tandem carbohydrate-recognition domains. FEBS Lett **443:** 139-143.

KOSHIBA, T., T. HAYASHI, I. MIWAKO, I. KUMAGAI, T. IKURA *et al.*, 1999 Expression of a synthetic gene encoding canine milk lysozyme in Escherichia coli and characterization of the expressed protein. Protein Eng **12:** 429-435.

KRAGOL, G., S. LOVAS, G. VARADI, B. A. CONDIE, R. HOFFMANN *et al.*, 2001 The antibacterial peptide pyrrhocoricin inhibits the ATPase actions of DnaK and prevents chaperone-assisted protein folding. Biochemistry **40:** 3016-3026.

KURUCZ, E., R. MARKUS, J. ZSAMBOKI, K. FOLKL-MEDZIHRADSZKY, Z. DARULA *et al.*, 2007 Nimrod, a putative phagocytosis receptor with EGF repeats in Drosophila plasmatocytes. Curr Biol **17:** 649-654.

LACKIE, A. M., and G. R. VASTA, 1988 The role of galactosyl-binding lectin in the cellular immune response of the cockroach Periplaneta americana (Dictyoptera). Immunology **64:** 353-357.

LAMBERTY, M., A. CAILLE, C. LANDON, S. TASSIN-MOINDROT, C. HETRU *et al.*, 2001 Solution structures of the antifungal heliomicin and a selected variant with both antibacterial and antifungal activities. Biochemistry **40:** 11995-12003.

LAVINE, M. D., and M. R. STRAND, 2002 Insect hemocytes and their role in immunity. Insect Biochem Mol Biol **32:** 1295-1309.

LEE, W. J., and P. T. BREY, 1995 Isolation and characterization of the lysozyme-encoding gene from the silkworm Bombyx mori. Gene **161:** 199-203.

LEMAITRE, B., and J. HOFFMANN, 2007 The host defense of Drosophila melanogaster. Annu Rev Immunol **25:** 697-743.

LEMAITRE, B., E. KROMER-METZGER, L. MICHAUT, E. NICOLAS, M. MEISTER *et al.*, 1995 A recessive mutation, immune deficiency (imd), defines two distinct control pathways in the Drosophila host defense. Proc Natl Acad Sci U S A **92:** 9465-9469.

LEMAITRE, B., E. NICOLAS, L. MICHAUT, J. M. REICHHART and J. A. HOFFMANN, 1996 The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in Drosophila adults. Cell **86:** 973-983.

LEULIER, F., C. PARQUET, S. PILI-FLOURY, J. H. RYU, M. CAROFF *et al.*, 2003 The Drosophila immune system detects bacteria through specific peptidoglycan recognition. Nat Immunol **4:** 478-484.

LEVASHINA, E. A., E. LANGLEY, C. GREEN, D. GUBB, M. ASHBURNER *et al.*, 1999 Constitutive activation of toll-mediated antifungal defense in serpin-deficient Drosophila. Science **285:** 1917-1919.

LOPEZ-ZAVALA, A. A., E. DE-LA-RE-VEGA, S. A. CALDERON-ARREDONDO, K. D. GARCIA-OROZCO, E. F. VELAZQUEZ *et al.*, 2004 Biophysical characterization of an insect lysozyme from Manduca sexta. Protein Pept Lett **11:** 85-92.

LORD, J. C., S. ANDERSON and D. W. STANLEY, 2002 Eicosanoids mediate Manduca sexta cellular response to the fungal pathogen Beauveria bassiana: a role for the lipoxygenase pathway. Arch Insect Biochem Physiol **51:** 46-54.

LU, X., M. WANG, J. QI, H. WANG, X. LI *et al.*, 2006 Peptidoglycan recognition proteins are a new class of human bactericidal proteins. J Biol Chem **281:** 5895-5907.

LUNA, C., N. T. HOA, H. LIN, L. ZHANG, H. L. NGUYEN *et al.*, 2006 Expression of immune responsive genes in cell lines from two different Anopheline species. Insect Mol Biol **15:** 721-729.

MA, C., and M. R. KANOST, 2000 A beta1,3-glucan recognition protein from an insect, Manduca sexta, agglutinates microorganisms and activates the phenoloxidase cascade. J Biol Chem **275:** 7505-7514.

MALCOLM, B. A., S. ROSENBERG, M. J. COREY, J. S. ALLEN, A. DE BAETSELIER *et al.*, 1989 Site-directed mutagenesis of the catalytic residues Asp-52 and Glu-35 of chicken egg white lysozyme. Proc Natl Acad Sci U S A **86:** 133-137.

MATSUURA, A., M. YAO, T. AIZAWA, N. KOGANESAWA, K. MASAKI *et al.*, 2002 Structural analysis of an insect lysozyme exhibiting catalytic efficiency at low temperatures. Biochemistry **41:** 12086-12092.

MICHEL, T., J. M. REICHHART, J. A. HOFFMANN and J. ROYET, 2001 Drosophila Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. Nature **414:** 756-759.

MILLER, J. S., T. NGUYEN and D. W. STANLEY-SAMUELSON, 1994 Eicosanoids mediate insect nodulation responses to bacterial infections. Proc Natl Acad Sci U S A **91:** 12418-12422.

MITA, K., M. KASAHARA, S. SASAKI, Y. NAGAYASU, T. YAMADA *et al.*, 2004 The genome sequence of silkworm, Bombyx mori. DNA Res **11:** 27-35.

MITA, K., M. MORIMYO, K. OKANO, Y. KOIKE, J. NOHATA *et al.*, 2003 The construction of an EST database for Bombyx mori and its application. Proc Natl Acad Sci U S A **100:** 14121-14126.

MURRAY, P. J., 2007 The JAK-STAT signaling pathway: input and output integration. J Immunol **178:** 2623-2629.

NASH, J. A., T. N. BALLARD, T. E. WEAVER and H. T. AKINBI, 2006 The peptidoglycan-degrading property of lysozyme is not required for bactericidal activity in vivo. J Immunol **177:** 519-526.

NENE, V., J. R. WORTMAN, D. LAWSON, B. HAAS, C. KODIRA *et al.*, 2007 Genome sequence of Aedes aegypti, a major arbovirus vector. Science **316:** 1718-1723.

NIRMALA, X., K. MITA, V. VANISREE, M. ZUROVEC and F. SEHNAL, 2001 Identification of four small molecular mass proteins in the silk of Bombyx mori. Insect Mol Biol **10:** 437-445.

OCHIAI, M., and M. ASHIDA, 1988 Purification of a beta-1,3-glucan recognition protein in the prophenoloxidase activating system from hemolymph of the silkworm, Bombyx mori. J Biol Chem **263:** 12056-12062.

OCHIAI, M., and M. ASHIDA, 1999 A pattern recognition protein for peptidoglycan. Cloning the cDNA and the gene of the silkworm, Bombyx mori. J Biol Chem **274:** 11854-11858.

OCHIAI, M., and M. ASHIDA, 2000 A pattern-recognition protein for beta-1,3-glucan. The binding domain and the cDNA cloning of beta-1,3-glucan recognition protein from the silkworm, Bombyx mori. J Biol Chem **275:** 4995-5002.

POND, S. L., S. D. FROST and S. V. MUSE, 2005 HyPhy: hypothesis testing using phylogenies. Bioinformatics **21:** 676-679.

ROXSTROM-LINDQUIST, K., O. TERENIUS and I. FAYE, 2004 Parasite-specific immune response in adult Drosophila melanogaster: a genomic study. EMBO Rep **5:** 207-212.

ROZEN, S., and H. SKALETSKY, 2000 Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol **132:** 365-386.

SADD, B. M., and P. SCHMID-HEMPEL, 2006 Insect immunity shows specificity in protection upon secondary pathogen exposure. Curr Biol **16:** 1206-1210.

SCHMIDT, H. A., K. STRIMMER, M. VINGRON and A. VON HAESELER, 2002 TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics **18:** 502-504.

SCHUHMANN, B., V. SEITZ, A. VILCINSKAS and L. PODSIADLOWSKI, 2003 Cloning and expression of gallerimycin, an antifungal peptide expressed in immune response of greater wax moth larvae, Galleria mellonella. Arch Insect Biochem Physiol **53:** 125-133.

SHIN, S. W., S. S. PARK, D. S. PARK, M. G. KIM, S. C. KIM *et al.*, 1998 Isolation and characterization of immune-related genes from the fall webworm, Hyphantria cunea, using PCR-based differential display and subtractive cloning. Insect Biochem Mol Biol **28:** 827-837.

SHRIVASTAVA, B., and A. K. GHOSH, 2003 Protein purification, cDNA cloning and characterization of a protease inhibitor from the Indian tasar silkworm, Antheraea mylitta. Insect Biochem Mol Biol **33:** 1025-1033.

SHUGAR, D., 1952 The measurement of lysozyme activity and the ultra-violet inactivation of lysozyme. Biochim Biophys Acta **8:** 302-309.

SILHACEK, D. L., S. G. MILLER and C. L. MURPHY, 1994 Purification and characterization of a flavin-binding storage protein from the hemolymph of Galleria mellonella. Arch Insect Biochem Physiol **25:** 55-72.

STEINBERG, D. A., and R. I. LEHRER, 1997 Designer assays for antimicrobial peptides. Disputing the "one-size-fits-all" theory. Methods Mol Biol **78:** 169-186.

STEINER, H., 1982 Secondary structure of the cecropins: antibacterial peptides from the moth Hyalophora cecropia. FEBS Lett **137:** 283-287.

STEINER, H., 2004 Peptidoglycan recognition proteins: on and off switches for innate immunity. Immunol Rev **198:** 83-96.

STEINER, H., D. HULTMARK, A. ENGSTROM, H. BENNICH and H. G. BOMAN, 1981 Sequence and specificity of two antibacterial proteins involved in insect immunity. Nature **292:** 246-248.

SUN, S. C., I. LINDSTROM, H. G. BOMAN, I. FAYE and O. SCHMIDT, 1990 Hemolin: an insect-immune protein belonging to the immunoglobulin superfamily. Science **250:** 1729-1732.

TANAKA, H., M. YAMAMOTO, Y. MORIYAMA, M. YAMAO, S. FURUKAWA *et al.*, 2005 A novel Rel protein and shortened isoform that differentially regulate antibacterial peptide genes in the silkworm Bombyx mori. Biochim Biophys Acta **1730:** 10-21.

TOMARI, Y., and P. D. ZAMORE, 2005 Perspective: machines for RNAi. Genes Dev **19:** 517-529.

TRUDEAU, D., J. O. WASHBURN and L. E. VOLKMAN, 2001 Central role of hemocytes in Autographa californica M nucleopolyhedrovirus pathogenesis in Heliothis virescens and Helicoverpa zea. J Virol **75:** 996-1003.

VEIGA, A. B., J. M. RIBEIRO, J. A. GUIMARAES and I. M. FRANCISCHETTI, 2005 A catalog for the transcripts from the venomous structures of the caterpillar Lonomia obliqua: identification of the proteins potentially involved in the coagulation disorder and hemorrhagic syndrome. Gene **355:** 11-27.

VODOVAR, N., M. VINALS, P. LIEHL, A. BASSET, J. DEGROUARD *et al.*, 2005 Drosophila host defense after oral infection by an entomopathogenic Pseudomonas species. Proc Natl Acad Sci U S A **102:** 11414-11419.

WANG, J., Q. XIA, X. HE, M. DAI, J. RUAN *et al.*, 2005 SilkDB: a knowledgebase for silkworm biology and genomics. Nucleic Acids Res **33:** D399-402.

WANG, X. H., R. ALIYARI, W. X. LI, H. W. LI, K. KIM *et al.*, 2006 RNA interference directs innate immunity against viruses in adult Drosophila. Science **312:** 452-454.

WATSON, F. L., R. PUTTMANN-HOLGADO, F. THOMAS, D. L. LAMAR, M. HUGHES *et al.*, 2005 Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science **309:** 1874-1878.

WERNER, T., G. LIU, D. KANG, S. EKENGREN, H. STEINER *et al.*, 2000 A family of peptidoglycan recognition proteins in the fruit fly Drosophila melanogaster. Proc Natl Acad Sci U S A **97:** 13772-13777.

WHELAN, S., and N. GOLDMAN, 2001 A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol **18:** 691-699.

WHITTEN, M. M., I. F. TEW, B. L. LEE and N. A. RATCLIFFE, 2004 A novel role for an insect apolipoprotein (apolipophorin III) in beta-1,3-glucan pattern recognition and cellular encapsulation reactions. J Immunol **172:** 2177-2185.

XU, X., Y. DONG, E. G. ABRAHAM, A. KOCAN, P. SRINIVASAN *et al.*, 2005 Transcriptome analysis of Anopheles stephensi-Plasmodium berghei interactions. Mol Biochem Parasitol **142:** 76-87.

YAMAKAWA, M., and H. TANAKA, 1999 Immune proteins and their gene expression in the silkworm, Bombyx mori. Dev Comp Immunol **23:** 281-289.

YOSHIDA, H., K. KINOSHITA and M. ASHIDA, 1996 Purification of a peptidoglycan recognition protein from hemolymph of the silkworm, Bombyx mori. J Biol Chem **271:** 13854-13860.

YU, X. Q., and M. R. KANOST, 2000 Immulectin-2, a lipopolysaccharide-specific lectin from an insect, Manduca sexta, is induced in response to gram-negative bacteria. J Biol Chem **275:** 37373-37381.

YU, X. Q., Y. F. ZHU, C. MA, J. A. FABRICK and M. R. KANOST, 2002 Pattern recognition proteins in Manduca sexta plasma. Insect Biochem Mol Biol **32:** 1287-1293.

ZAMBON, R. A., M. NANDAKUMAR, V. N. VAKHARIA and L. P. WU, 2005 The Toll pathway is important for an antiviral response in Drosophila. Proc Natl Acad Sci U S A **102:** 7257-7262.

ZHANG, K., R. GAO, H. ZHANG, X. CAI, C. SHEN *et al.*, 2005 Molecular cloning and characterization of three novel lysozyme-like genes, predominantly expressed in the male reproductive system of humans, belonging to the c-type lysozyme/alpha-lactalbumin family. Biol Reprod **73:** 1064-1071.

ZHANG, X. F., J. F. WANG, E. MATCZAK, J. A. PROPER and J. E. GROOPMAN, 2001 Janus kinase 2 is involved in stromal cell-derived factor-1alpha-induced tyrosine phosphorylation of focal adhesion proteins and migration of hematopoietic progenitor cells. Blood **97:** 3342-3348.

ZHU, Y., T. J. JOHNSON, A. A. MYERS and M. R. KANOST, 2003 Identification by subtractive suppression hybridization of bacteria-induced genes expressed in Manduca sexta fat body. Insect Biochem Mol Biol **33:** 541-559.

Publications

- **Gandhe A S**, John S H, Nagaraju J (2007) Noduler, a novel immune upregulated protein mediates nodulation response in insects. Journal of Immunology (In Press).

- **Gandhe AS**, Janardhan G, Nagaraju J (2007) Immune upregulation of novel antibacterial proteins from silkmoths (Lepidoptera) that resemble lysozymes but lack muramidase activity. Insect Biochemistry and Molecular Biology 37: 655-666.

- **Gandhe AS**, Arunkumar KP, John SH, Nagaraju J (2006) Analysis of bacteria-challenged wild silkmoth, *Antheraea mylitta* (lepidoptera) transcriptome reveals potential immune genes. BMC Genomics 7: 184.