TRIBOLIUM CASTANEUM GENES ENCODING PROTEINS WITH THE CHITIN-BINDING TYPE II DOMAIN

by

SINU JASRAPURIA

B.S., Calcutta University, India, 2003
M.S., Calcutta University, India, 2005

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Graduate Biochemistry Group

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2011

Abstract

The extracellular matrices of cuticle and peritrophic matrix of insects are composed mainly of chitin complexed with proteins, some of which contain chitin-binding domains. This study is focused on the identification and functional characterization of genes encoding proteins that possess one or more copies of the six-cysteine-containing ChtBD2 domain (Peritrophin A motif =CBM_14 =Pfam 01607) in the red flour beetle, *Tribolium castaneum*. A bioinformatics search of *T. castaneum* genome yielded previously characterized chitin metabolic enzymes and several additional proteins. Using phylogenetic analyses, the exon-intron organization of the corresponding genes, domain organization of proteins, and temporal and tissue-specificity of expression patterns, these proteins were classified into three large families. The first family includes 11 proteins essentially made up of 1 to 14 repeats of the peritrophin A domain. Transcripts for these proteins are expressed only in the midgut and only during feeding stages of development. We therefore denote these proteins as "Peritrophic Matrix Proteins" or PMPs. The genes of the second and third families are expressed in cuticle-forming tissues throughout all stages of development but not in the midgut. These two families have been denoted as "Cuticular Proteins Analogous to Peritrophins 3" or CPAP3s and "Cuticular Proteins Analogous to Peritophins 1" or CPAP1s based on the number of ChtBD2 domains that they contain. Unlike other cuticular proteins studied so far, TcCPAP1-C protein is localized predominantly in the exocuticle and could contribute to the unique properties of this cuticular layer. RNA interference (RNAi), which down-regulates transcripts for any targeted gene, results in lethal and/or abnormal phenotypes for some, but not all, of these genes. Phenotypes are often unique and are manifested at different developmental stages, including embryonic, pupal and/or adult stages. The

experiments presented in this dissertation reveal that while the vast majority of the CPAP3 genes serve distinct and essential functions affecting survival, molting or normal cuticle development. However, a minority of the *CPAP1* and *PMP* family genes are indispensable for survival under laboratory conditions. Some of the non-essential genes may have functional redundancy or may be needed only under special circumstances such as exposure to stress or pathogens.

TRIBOLIUM CASTANEUM GENES ENCODING PROTEINS WITH THE CHITIN-BINDING TYPE II DOMAIN

by

SINU JASRAPURIA

B.S., Calcutta University, India, 2003
M.S., Calcutta University, India, 2005

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Graduate Biochemistry Group

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2011

Approved by:

Major Professor
Subbaratnam Muthukrishnan

Copyright

Abstract

The extracellular matrices of cuticle and peritrophic matrix of insects are composed mainly of chitin complexed with proteins, some of which contain chitin-binding domains. This study is focused on the identification and functional characterization of genes encoding proteins that possess one or more copies of the six-cysteine-containing ChtBD2 domain (Peritrophin A motif =CBM_14 =Pfam 01607) in the red flour beetle, *Tribolium castaneum*. A bioinformatics search of *T. castaneum* genome yielded previously characterized chitin metabolic enzymes and several additional proteins. Using phylogenetic analyses, the exon-intron organization of the corresponding genes, domain organization of proteins, and temporal and tissue-specificity of expression patterns, these proteins were classified into three large families. The first family includes 11 proteins essentially made up of 1 to 14 repeats of the peritrophin A domain. Transcripts for these proteins are expressed only in the midgut and only during feeding stages of development. We therefore denote these proteins as "Peritrophic Matrix Proteins" or PMPs. The genes of the second and third families are expressed in cuticle-forming tissues throughout all stages of development but not in the midgut. These two families have been denoted as "Cuticular Proteins Analogous to Peritrophins 3" or CPAP3s and "Cuticular Proteins Analogous to Peritophins 1" or CPAP1s based on the number of ChtBD2 domains that they contain. Unlike other cuticular proteins studied so far, TcCPAP1-C protein is localized predominantly in the exocuticle and could contribute to the unique properties of this cuticular layer. RNA interference (RNAi), which down-regulates transcripts for any targeted gene, results in lethal and/or abnormal phenotypes for some, but not all, of these genes. Phenotypes are often unique and are manifested at different developmental stages, including embryonic, pupal and/or adult stages. The

experiments presented in this dissertation reveal that while the vast majority of the CPAP3 genes serve distinct and essential functions affecting survival, molting or normal cuticle development. However, a minority of the *CPAP1* and *PMP* family genes are indispensable for survival under laboratory conditions. Some of the non-essential genes may have functional redundancy or may be needed only under special circumstances such as exposure to stress or pathogens.

TABLE OF CONTENTS

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Name
CHS	Chitin synthase
PMP	Peritrophic matrix protein
Tc	*Tribolium castaneum*
CBD	Chitin-binding domain
ChtBD2	Chitin-binding peritrophin domain type II
CPAP	Cuticular Protein Analogous to Peritrophins
R & R consensus	Rebers and Riddiford consensus
GlcNAc	N-acetylglucosamine
PDB	Protein Data Bank
EST	Expressed Sequence Tag
Bt	*Bacillus thuringiensis*
PCR	Polymerase Chain Reaction
RNAi	RNA interference
IgG	Immunoglobulin G
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Ver	Tryptophan oxygenase (vermilion gene)
PBS	Phosphate buffered saline
NBT/BCIP	Nitroblue tetrazolium salt/5-bromo-4-chloro-3-indolyl phosphate

Acknowledgements

I cannot start my dissertation without acknowledging the valuable contributions of many people. I owe a lot to each of them. I have no words to express my sincere gratitude to my dissertation advisor, Prof. S. Muthukrishnan. He was much more than just an advisor. Throughout the last 5 years at the Kansas State University, he has guided me through each and every aspect of a graduate student career: from patiently explaining the basic concepts of biochemistry to meticulously demonstrating how to troubleshoot a real life experiment. He has taught me how to think like a biochemist.

I would like to thank Prof. Karl Kramer for being there throughout my Ph.D. career. He has always been available for stimulating discussions. He has taken keen interest in my research work and has been a huge support. I am grateful to Prof. Richard Beeman for being patient in clarifying my doubts, making suggestions for troubleshooting or improving the experiment and all the general friendly advice.

It is not possible to talk about my dissertation without mentioning Dr. Yasuyaki Arakane who has played a very important role in my graduate career. I have learnt a lot from the myriad long and fruitful scientific discussions with him. Working with him in the lab also helped me improve my experimental skills.

One cannot be successful in advanced research without a thorough understanding of the fundamental concepts in the field. The graduate coursework at KSU was ideal for the same, especially the course “Advanced Topics in Insect Biochemistry” offered by Prof. Mike Kanost. Also worth mentioning is the course on “Insect Physiology“ offered by the Entomology

Department and taught by Prof. Yoonseong Park and the course “Topics in Bioinformatics“ offered by CIS department and taught by Prof. Doina Caragea and Prof. Susan Brown. These courses gave me an opportunity to learn the ABC's of this field and prepared me well for further research. In the following years, whenever I had trouble with some theoretical concepts, I could always count on them, as they were always welcoming and happy to help me out.

I thank Dr. Khurshida Begum for teaching me *in situ* hybridization techniques, Dr. Yasuaki Hiromasa for mass spectrometry and Dr. Charles Specht for chitin content data analysis. It is a pleasure to thank Prof. Judith H. Willis, University of Georgia for helpful suggestions and reviewing drafts of the manuscripts.

I would like to thank all the past and present members of our group whom I have worked with during my stay at KSU – Radhika Dixit, Sujata Chaudhari, Meera Kumari, Dr. Gunnar Broehan and Dr. Monica Cardosa. It has been great experience to work with all these people, and to interact and support each other in our individual projects. I would specially like to thank Dr. Hans Merzendorfer for the fantastic interaction over the few months we worked together.

Also special thanks to Dr. Neal Dittmer, Dr. Maureen Gorman and Dr. Renata Bolognesi for their valuable advice and friendly discussion. I would also like to thank Shagun Sharma and USDA technician Kathy Leonard for insect husbandry.

I would like to thank my committee members, Prof. S. Muthukrishnan, Prof. Karl Kramer, Prof. Richard Beeman, Prof. Yoonseong Park, Prof. Mike Kanost and my outside chair Prof. Vara Prasad for their guidance, encouragement, time commitment and valuable advices.

I wish to thank my close friends Abhishek, Achira, Avinash, Dipanwita and Devdeep. They have been extremely helpful, supportive and encouraging. Finally, I would like to thank all my friends and family here and back in India for all their love and support.

Dedication

I would like to dedicate my dissertation to my Dad, Mom and my brother Ankur without whose support I could not have come this far and realized my dreams, and also to my wonderful husband Sanjeev for his love and support.

Overview of Dissertation

Chapter 1 is a literature review that introduces the current state of knowledge available on chitin-binding proteins, their structure and functional aspects throughout different arthropods.

Chapter 2 presents a detailed study on the identification and characterization of all potential ChtBD2-domain-containing protein genes in *T. castaneum*, as well as the cloning details of all the ChtBD2-domain-containing protein genes, including the 11 PMP family, 10 CPAP1 family and 8 CPAP3 family members. This work utilizes the different bioinformatics tools including phylogenetic analysis, tandem repeat analyzer, secondary structure predictors, and weblogo analysis for analyzing conserved residues within a domain. Different BLAST searches identified orthologs for these proteins in different species. Most of the work presented in this chapter has appeared as an article in the journal *Insect Biochemistry and Molecular Biology* (Jasrapuria et al., 2010).

In **Chapter 3** I report the functional aspects of Cuticular Proteins Analogous to Peritrophins and demonstrate their function by RNAi-mediated transcript knock down at embryonic, larval, pupal and adult stages of development. I analyzed both temporal and spatial expression patterns throughout different developmental stages. I also expressed and purified some of the recombinant proteins. The manuscript describing this work is being prepared,

In **Chapter 4** I report the functional aspects of Peritrophic Matrix Proteins based on RNAi-mediated down-regulation of transcripts as well as and mRNA levels at different stages and different tissues during development.

Finally, **Chapter 5** summarizes the key contribution this work to the field of *T. castaneum* research and presents some potential experiments and future research directions.

The **Appendices** at the end include the protein sequences, the accession numbers of the different proteins used for Weblogo and phylogenetic tree analysis. Also list of bacterial strains, kits, antibodies used for my research.

As a whole, this dissertation describes three different classes of proteins based on their consensus sequence and tissue-specific expressions as well as insight into their functions by RNAi-mediated transcript knock-down.

CHAPTER 1

Literature review

1.1. *Tribolium castaneum* as a model insect

Insects belong to an extremely diverse group of organisms both in terms of morphology and life history traits, which make them ideal models for comparative studies in the fields of evolutionary biology, population biology, developmental biology and physiology. The knowledge of insect biology is crucial to solve the problems that they cause as agricultural pests and vectors of disease. *T. castaneum* is a powerful model for the studies of gene function. It meets the criteria for a genetic model organism, including small body size, short generation time and large brood sizes. It can be cultured in the normal laboratory environment with minimal effort and the availability of fully sequenced and annotated genome (Richards et al., 2008) allows extensive bioinformatics studies. While *Drosophila melanogaster* is one of the most powerful genetic models, *T. castaneum* is more representative of other insect species than is *Drosophila* (Richards et al., 2008). Systemic RNAi brought about by injection of double-stranded RNA (dsRNA) is an effective tool in many emerging model systems to obtain loss of function phenotypes. Besides the nematode *Caenorhabditis elegans* (Fire et al., 1998) and plants (Waterhouse, 1998), systemic RNAi has been documented in other organisms, including other nematodes (Felix, 2008), flatworms (Sanchez Alvarado and Newmark, 1999), crustaceans (Robalino et al., 2005), chelicerates (Akiyama-Oda and and Oda, 2006; Aljamali et al., 2003; Narasimhan et al., 2004; Soares et al., 2005) and insects (Bucher et al., 2002; Roignant, 2003; Tomoyasu and Denell, 2004), although the mode of action is unknown. The systemic RNAi mechanism is not universal, as the leading insect model, *D. melanogaster*, does not respond to injected dsRNA beyond the syncytial stage during embryonic development and beyond (Roignant, 2003).

1.2. Chitin and its importance in insect exoskeleton and peritrophic matrix organization

Chitin serves as one of the main components of the cell walls of most fungi and of the skeletal polysaccharide of several animal phyla, namely Arthropoda, Annelida, Molluska and Coelenterata. It is also found in eggshells and pharynx of nematodes (Fanelli et al., 2005; Zhang et al., 2005). Chitin is not only a major constituent of the cuticle of insects (Kramer and Muthukrishnan, 1997) but also is a part of the peritrophic matrix (PM), foregut, hindgut, tracheae, eggshells and muscle attachment points (Kramer and Muthukrishnan, 2005). Chitin is a naturally occurring structural homopolymer of β-(1-4) linked N-acetylglucosamine monomers (GlcNAc, $(C_8H_{13}O_5N)_n$, where n>>1). Chitin is primarily composed of poly-GlcNAc, but it can also contain variable amounts of deacetylated glucosamine residues.

Chitin can exist in three crystalline allomorphs α, β and γ. Of these α is most stable and common (chitin chains exhibit antiparallel orientation), while β-chitin with the chitin chains arranged in parallel fashion has lower stability. The γ form of chitin is rare and not well characterized (chitin polymers are a mixture of two parallel and one anti-parallel chains). Chitin is often complexed with proteins, such as cuticular structural proteins and peritrophins. Depending on the type of protein complexed with it and the extent of protein cross-linking, chitin-containing composites can have varying degrees of rigidity, elasticity and water impermeability. While the chitinous cuticle serves mainly to provide a rigid and water-impermeable protective outer layer for the underlying epidermal tissue and to protect the insect from dehydration (Moussian et al., 2007), the PM is believed to have roles in protecting the midgut epithelial cells and in aiding digestion (Bolognesi et al., 2001; Bolognesi et al., 2008; Terra, 2001) and is likely to be heavily hydrated as a result of glycosylation of the residues in the

linker regions of the PMPs, some of which have mucin-like domains (Hegedus et al., 2009; Terra, 2001; Toprak et al., 2009; Wang and Granados, 1997a; Wang et al., 2004). The biochemical basis of this dichotomy is unclear, but differences in the nature of the proteins associated with chitin and the presence of cross-linking mediated by quinones/quinone methides (Andersen, 2010) are likely to contribute to these differences in properties of the two chitin-containing matrices, namely the cuticle and the PM. Another possibility is that the arrangement of chitin fibers in the gut may be different in the PM when compared to that in cuticular chitin, which has multiple layers of α-chitin bundles in the form of stacked laminae. It is also possible that the PM-associated chitin is in the β-form, which leads to less hydrogen bonding between parallel chains of chitin compared to α-chitin (Jang et al., 2004).

The insect cuticle is a rigid and hydrophobic structure made up of an outer lipid waxy layer (the envelope), overlying an epicuticle made up of chitin and proteins cross-linked with catechols. The laminar procuticle is below the epicuticle and is composed of chitin microfibrils and proteins, underlying which is the layer of basal epithelial cells. The chitin polymer secreted by the epidermal and gut cells is assembled into microfibrils (2-6 nm diameter containing 20-400 chitin chains) that arrange into chitin bundles (~20 nm diameter containing ~10 microtubules) to make sheets. In cuticle (but probably not in PM) these sheets are cross-oriented relative to one another at a constant angle to form an apparent helicoidal bundle known as the Bouligand structure (Havemann et al., 2008; Hegedus et al., 2009; Moussian et al., 2005). Like the cuticle, the peritrophic membrane of insects is composed of chitin, proteins, glycoproteins and proteoglycans. The PM secreted by midgut epithelial cells throughout the length of the midgut is termed type I PM and those secreted by localized cardia cells in the anterior midgut are designated as type II PM (Peters, 1992; Tellam et al., 1999; Wigglesworth, 1930). Scanning

electron microscopic (SEM) studies done on *Manduca sexta* have indicated that the microvilli of the brush border cells in the midgut act as templates for the formation of the chitin network formation (Harper and Hopkins, 1997). Chitin microfibrils are secreted as an orthogonal network in the apical region of anterior midgut microvilli. This network moves to the tip of the microvilli where the proteinaceous matrix is added prior to the delamination of the single PM into the lumen (Hopkins and Harper, 2001). This creates two distinct compartments within the midgut, the ecto-peritrophic space and the endo-peritrophic space. Both synthesis and degradation of chitin take place at multiple developmental stages in the cuticle and PM. The new cuticle is usually synthesized as portions of the old endocuticle and PM and tracheae are reabsorbed and digested materials are recycled. A schematic diagram of the cuticle and the gut is shown in Figure 1-1.

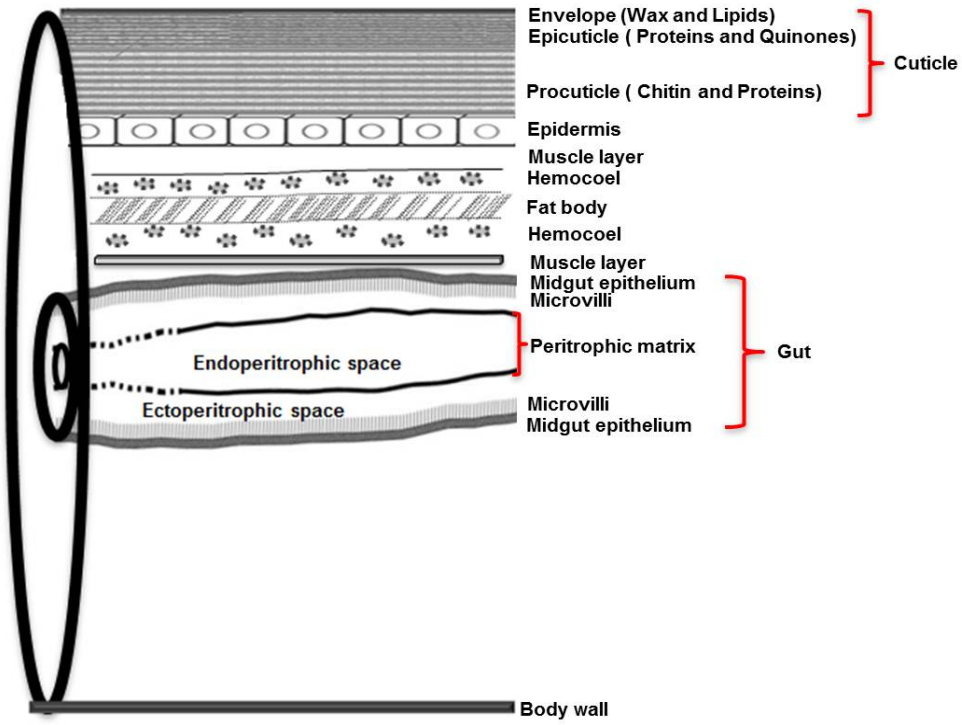


Figure 1-1 Schematic diagram showing the cross-section of different layers in the insect cuticle and gut.

1.3. Chitin-binding proteins in insects

1.3.1. Proteins with Rebers & Riddiford chitin-binding motif

Proteins that bind to chitin influence its mechanical and physicochemical properties. Since chitin is an extracellular matrix polysaccharide, the proteins that have an affinity for chitin are expected to be extracellularly secreted proteins. Insect proteins having chitin-binding ability can be broadly classified into three groups based on their amino acid sequence motifs. The first group consists of a large assortment of insect cuticular protein that belong to the CPR family and contains Rebers & Riddiford motifs, which are stretches of 70 amino acids that have the following consensus sequence:

PDGDYNY+YETSNGIADQETGD+KSQGETRDG++AVDVV+GSYSYVDPDGTTRTVTYT ADDENGFQPVGAHLP; pfam00379;(Karouzou et al., 2007); (Willis, 2010). Two major groups of CPR proteins have been recognized, namely RR-1 associated generally with soft cuticle and RR-2 associated generally with hard cuticle (Iconomidou et al., 2005). This motif is found in about 70% of all cuticle proteins (Karouzou et al., 2007). Modeling studies have suggested that this domain may form a β-sheet-type structure with aromatic side chains that may participate in chitin interaction. The CPR family of proteins is rich in histidine residues and lacks cysteines. The number of cuticular proteins ~100 belonging to CPR family varies in different insect species, as shown in Table 1-1 (Willis, 2010).

Table 1-1 Cuticular proteins having domains other than the peritrophin-A domain.

Name	Species	References
i) CPR (R & R Consensus)		
	Xenopus	(Klein et al., 2002)
	Penaeus japonicas	(Ikeya et al., 2001)
	Tachypleus tridentatus	(Iijima et al., 2005)
Agcp2b (now AgamCPR97)	*Anopheles gambiae*	(Rebers and Willis, 2001)
32 CPR	*Apis mellifera*	(Honeybee Genome Sequencing Consortium, 2006)
62 CPR	*Nasonia vitripennis*	(Werren et al., 2010)
~100 CPR	*D. melanogaster*	(Cornman, 2009)
156 (102 RR-2)	*A. gambiae*	(Togawa et al., 2004)
148 (89 RR-2)	*Bombyx mori*	(Futahashi et al., 2008)
ii) CPF and CPFL		
6 CPF	*Tenebrio/Locusta*	(Andersen et al., 1997)
4 AgamCPFs/AgamCPF1 and AgamCPFL2-7		(He et al., 2007)
11	*A. gambiae*	(Togawa et al., 2004)
5	*B. mori*	(Willis, 2010)
3	*D. melanogaster*	"
3	*Apis mellifera*	"
iii)Tweedle(TWDL)(pfam03103)		
27 member	*D. melanogaster*	(Guan et al., 2006)
12	*A. gambiae*	(Willis, 2010)
4	*B. mori*	"
2	*A. mellifera*	"
2	*N. vitripennis*	"
3	*Tribolium castaneum*	"
		"
iv) CPLCA (pfam04527/IPR007614)		
3	*A. gambiae*	(Willis, 2010)
11	*D. melanogaster*	"
v) CPLCG family		
Dacp1 and Dacp2	*D. melanogaster*	(Qui and Hardin, 1995)
27	*A. gambiae*	(He et al., 2007)
2	*T. castaneum*	"
vi) CPLCW		
9	*A. gambiae*	(Cornman, 2009)
vii) CPLCP		
AgamCPLCP8,10,11,12	*A. gambiae*	(He et al., 2007)
24	*A. gambiae*	(Willis, 2010)
19	*Aedes aegypti*	"
24	*Culex quinquefasciatus*	"
7	*B. mori*	"
5	*D. melanogaster*	"
2	*A. mellifera*	"
3	*N. vitripennis*	"
4	*T. castaneum*	"
viii) Apidermin		
apd-1, apd-2, apd-3	*A. mellifera*	(Kucharski et al., 2007)
3 genes	*N. vitripennis*	(Willis, 2010)
ix) Glycine rich (GGYGG /GGxGG)		
CPG (18 members)	*B. mori*	(Futahashi et al., 2008; Zhong et al., 2006)

1.3.2. Proteins with peritrophin-A motif

The second group of chitin-binding proteins lacks the R&R consensus but has the peritrophin-A domain. The domain is a stretch of 60-70 amino acids that has 6 cysteine residues and has the consensus spacing of cysteines, $CX_{13-20}CX_{5-6}CX_{9-19}CX_{10-14}C\text{-}X_{4-14}$; pfam01607 (Tellam et al., 1999). The first reported peritrophic matrix proteins (PMPs) having the peritrophin-A domain were peritrophin-44 and -48 extracted from the guts of *Lucilia cuprina* larvae (Elvin et al., 1996; Schorderet et al., 1998). Since then, a large number of PMPs has been characterized either by direct extraction from PMs or by analysis of cDNA sequences and more recently by bioinformatics searches of annotated genomes. All PMPs have signal peptides, suggesting these are secreted protein. Some of them have long stretches of serines/threonines, while others have N-linked glycosylation sites (Hegedus et al., 2009), suggesting the possibility that some PMPs may be glycoproteins. The PMP with the largest number of CBDs characterized so far with 19 peritrophin A motifs is the one predicted from a cDNA clone isolated from *Mamestra configurata* (Shi et al., 2004). The midgut of Hessian fly has a small peritrophin with 2 CBDs (Mittapalli et al., 2007). Proteins with 1 CBD have also been identified (Eisemann et al., 2001; Wijffels et al., 2001) and have been suggested to have roles in protecting the ends of chitin chains from chitinase attack (Bolognesi et al., 2005) and also to protect the midgut from pathogen invasion (Hao and Aksoy, 2002). PM proteins having CBD domains can also have mucin domains. The secreted mucins contain domains rich in serine, proline and threonine residues that are potential sites for O-linked glycosylation. These were first reported in *Trichoplusia ni* and termed as “invertebrate intestinal mucins” (IIM) (Wang and Granados, 1997b).

The presence of the peritrophin-A domain is not a bona fide hallmark of PM proteins. Peritrophin-A domain-containing proteins named "gasp" (gene analogous to small peritrophins) (Barry et al., 1999) and "Obstructor" have been reported in cuticle-forming tissues in *D. melanogaster* (Behr and Hoch, 2005). Also some members of the chitinase and chitin deacetylase enzyme families that have a role in chitin metabolism also contain this peritrophin-A domain (Campbell et al., 2008; Dixit et al., 2008; Kramer, 1993; Zhu et al., 2008a). Most of these proteins bind to chitin and their affinity for the insoluble substrate increases in the presence of one or more of the peritrophin-A domains compared to enzymes without these domains (Arakane et al., 2003). A list of different ChtBD2 domain containing proteins in arthropods is shown in Table 1-2.

Table 1-2 List of arthropod proteins containing the peritrophin-A domain(s).

Name	Repeats	Species	References
I) Peritrophins	***PFAM 01607, CBM_14***		
Peritrophin-44	5 ChtBD2 (domain 2 has 8 cysteines) + Mucin	*Lucilia cuprina*	(Elvin et al., 1996)
Peritrophin-48	5 ChtBD2	*L. cuprina*	(Schorderet et al., 1998)
Peritrophin-48	5 ChtBD2	*Chrysomya bezziana*	(Vuocolo et al., 2001)
Peritrophin-15	1 ChtBD2	*C. bezziana*	(Wijffels et al., 2001)
Peritrophin-15	1 ChtBD2	*L. cuprina*	(Wijffels et al., 2001)
GmPro2	1 ChtBD2	*Glossina morsitans*	(Hao and Aksoy, 2002)
Ag-Aper1	2 ChtBD2 + Mucin	*Anopheles gambiae*	(Shen and Jacobs-Lorena, 1999)
Ae-Aper50	5 ChtBD2 + Mucin	*Aedes aegypti*	(Shao et al., 2001)
Ag-Aper14	1 ChtBD2	*A. gambiae*	(Devenport et al., 2005)
McPM1		*Mamestra configurata*	(Shi et al., 2004)
CBP1	12 Tandem ChtBD2	*Trichoplusia ni*	(Wang et al., 2004)
CBP2	10 Tandem ChtBD2		
MdesPERI-A1[1]	2 ChtBD2	*Mayetiola destructor*	(Mittapalli et al., 2007)
Peritrophins[1]	2 ChtBD2	*Tenebrio molitor*	(Ferreira et al., 2007)
Peritrophins[1]	7 CHTBD2	*Spodoptera frugiperda*	"
Peritrophin-95	3 ChtBD2 + Mucin	*L. cuprina*	(Casu et al., 1997)
Peritrophin-55	1 ChtBD2 (8C) + Mucin	*L. cuprina*	(Tellam et al., 2003)
AEIMUC1[1]		*A. aegypti*	(Rayms-Keller et al., 2000)
TnIIM		*T. ni*	(Wang and Granados, 1997b)
PxIIM	6 ChtBD2 + 3MD	*Plutella xylostella*	(Sarauer et al., 2003)
McIIM1	6 ChtBD2 + 2 MD	*M. configurata*	(Shi et al., 2004)
McIIM2	2 ChtBD2 + 1 MD	*M. configurata*	"
McIIM3	3 ChtBD2 + 2 MD	*M. configurata*	"
McIIM4	4 ChtBD2 + 2 MD	*M. configurata*	"
Fm-SOP	3 ChtBD2	*Fenneropenaeus merguiensis*	(Loongyai et al., 2007)
II) Cuticular protein with Peritrophin-A domains			
PL1, PL2, PL3	2, 2 and 1 ChtBD2	*Ctenocephalides felis*	(Gaines et al., 2003)
MPL1	3 ChtBD2+ 1 mucin	*C. felis*	"
MPL2	4 ChtBD2+ 4 mucin	*C. felis*	"
Peri A		*D. melanogaster*	(Barry et al., 1999)
Gasp		*D. melanogaster*	"
	peptide fragment	*A. gambiae*	(He et al., 2007)
Obstructor[A-E], [F-J]	3 ChtBD2	*D. melanogaster*	(Behr and Hoch, 2005)
III) Proteins with Chitin deacetylase Domains			
TnPM-P42	14 cysteine residues	*T. ni*	(Guo et al., 2005)

McCDA1		*M. configurata*	(Toprak et al., 2008)
TcCDA, 2, 3, 4, 5	1 ChtBD2	*T. castaneum*	(Dixit et al., 2008)
IV) Proteins with Chitinase Domains (partial list)			
MsCHT	1 ChtBD2	*M. sexta*	(Kramer, 1993)
TcCHT4, 5,7,8	1 ChtBD2	*T. castaneum*	(Zhu et al., 2008a)
TcCHT10	5 ChtBD2	*T. castaneum*	"
AgCHT4, 5, 7,8	1 ChtBD2	*A. gambiae*	"
AgCHT10	4 ChtBD2	*A. gambiae*	"
DmCHT4, 5,6,7,8, 12	1 ChtBD2	*D. melanogaster*	"
DmCHT10	4 ChtBD2	*D. melanogaster*	"
SfCHI	1 ChtBD2	*S. frugiperda*	(Bolognesi et al., 2005)

[1]Putative PM proteins.cDNA

Undoubtedly, there exist strong non-covalent interactions between chitin and chitin-binding proteins, but the evidence for the presence of covalent interactions is weak. There is evidence for the involvement of the peritrophin-A motif as well as the RR motifs in chitin-binding (Arakane et al., 2004; Rebers and Willis, 2001). However, there have been reports indicating that the mere presence of RR motif does not necessarily confer an affinity for colloidal chitin or "chitin beads" as measured under laboratory conditions (Suderman et al., 2006; Togawa et al., 2004). Model building based on crystallographically characterized templates and ligand-docking experiments have been described to predict some of the interactions between the N-acetylglucosamines and the amino acid residues lining the putative binding pockets (Hamodrakas et al., 2002; Iconomidou et al., 2005; Kramer and Muthukrishnan, 2005), but the precise amino acid residues involved in chitooligosaccharide binding are still unknown for either type of CBD.

1.3.3. Proteins with tachystatin-type chitin-binding motif

The third group of chitin-binding proteins comprises antimicrobial peptides related to tachystatins from horseshoe crab and the agatoxins (calcium channel antagonists) from spider venom. They play roles in immune defense against bacteria, fungi and other pathogens (Kramer and Muthukrishnan, 1997). These proteins have 6 cysteine residues, forming a triple-stranded β-sheet structure with an inhibitory cysteine-knot (ICK) motif unlike the peritrophin-A motif (Fujitani et al., 2007). Tachystatin has been reported to have chitin-binding activity (Osaki et al., 1999). Members representing each of the three groups of chitin-binding proteins have been extracted from the PM or cuticle or isolated from hemocytes.

1.4. ChtBD2/ Chitin-binding domain

There are currently 63 defined families of Carbohydrate Binding Modules (CBM), based on their ligand specificity (see http://www.cazy.org/Carbohydrate-Binding-Modules.html). Lectins are highly specific carbohydrate-binding proteins found in plants (Sauvion et al., 1996).

Many of these lectins including wheat germ agglutinin (WGA) and barley lectin binds specifically to chitin (Chrispeels and Raikhel, 1991). Hevein, a protein found in rubber latex contains a 43 amino acid-long coiled structure consisting of two β sheets and a small α-helix, called the “hevein domain “and is grouped in CBM 18 family, whereas the peritrophin-A domain is included in the CBM14 family (Boraston et al., 2004). The three dimensional structures of hevein as well as its truncated form, hevein-32, in both the free form and in complex with ligands have been determined. These studies have established that aromatic residues at positions 21, 23 and 30 and a serine at position 19 are involved in chitin binding (Aboitiz et al., 2004).

1.4.1. Many proteins have multiple chitin-binding motifs

While hevein has a single chitin-binding motif, the lectin, wheat germ agglutinin, has four hevein motifs. This organization appears to be the case in insects as well. Insect proteins with the number of CBM14 domains ranging from 1 to 19 have been described, particularly in the group of proteins known as peritrophins, which have been extracted from the peritrophic marix of both larvae and adult insects.

1.4.2. Structure of the peritrophin-A-type chitin-binding domain

Tertiary structures based on 2D-NMR studies in solution are available for two insect proteins with the peritrophin-A domain, namely tachycitin (PDB ID: 1DQC) and scarabaecin (PDB ID: 1IYC). The structure of scarabaecin, which is only 36 amino acids long, is stabilized by one disulfide bond and has two anti-parallel β-strands followed by a short α-helix (Hemmi et al., 2003). Tachycitin has 10 cysteines that form five disulfide bonds and its sequence from residues 40 to 60 shares significant secondary and tertiary structure similarity to scarabaecin and to the predicted structure of the C-terminal half of the peritrophin-A motif of several peritrophin-like proteins from a wide range of insect species (Suetake et al., 2000). The 3D structure of the sequence of tachycitin (positions 40 to 60) as determined by NMR is very similar to that of hevein-32 (PDB ID: 1TOW) from positions 20 to 32 (Aboitiz et al., 2004). All of these proteins have or are predicted to have two anti-parallel β sheets followed by a short α-helix in the C-terminal region, which is stabilized by one disulfide bond. This region in hevein has been shown to contain two aromatic amino acids that are critical for binding to chitin (Asensio et al., 1998).

1.4.3. Nomenclature of chitin-binding domains in peritrophic matrix proteins

Several peritrophins have been extracted from the PMs of *Lucilia cuprina* and *Chrysomya bezziana*, *Anopheles gambiae* and *Trichoplusia ni* (Shen and Jacobs-Lorena, 1999; Tellam et al., 1999; Wang and Granados, 1997b) and shown to bind to chitin. More recently, other proteins with multiple chitin-binding domains have been characterized by bioinformatics analysis. They have anywhere from one to 19 CBDs. Based on the number of cysteine residues and the consensus amino acid sequences within the cysteine-containing regions, (Tellam et al., 1999)) have divided them into three groups. The most common are proteins that have one or more repeats of the six-cysteine containing "peritrophin-A domain" with the consensus spacing between adjacent cysteines of $CX_{13-20}CX_{5-6}CX_{9-19}CX_{10-14}C\text{-}X_{4-14}C$ where X is any amino acid except cysteine. Two proteins from *L. cuprina* that have 8 cysteines were extracted from the PM with 6 M guanidine hydrochloride after first extracting the proteins with peritrophin A motifs with 6 M urea. These two proteins of *L. cuprina* named peritrophin-30 and peritrophin-55 presumably have tighter affinity to chitin (data on direct re-binding of extracted and renatured protein to chitin is unavailable) with the cysteine spacing of $CX_{12-13}CX_{20-21}CX_{10-}CX_{12}CX_2CX_8CX_{7-12}C$. This motif is referred to as the peritrophin B motif. Distantly related orthologs were found in several insect orders. Yet another protein was extracted from the insoluble PM material of *L. cuprina* as well as *C. bezziana*, which remained after 6 M guanidine chloride extraction of peritrophins-30 and -55 by treatment for 30 min with 10 mM dithiothreitol and 5% SDS at 95°C. These proteins denoted as peritophin-15 also have 6 cysteines but the consensus spacing of cysteines ($CX_{8-9}CX_{17-21}CX_{10-11}CX_{12-13}CX_{11}C$) is quite different from that of the peritrophin A group (Tellam et al., 1999). This motif has been named the "peritrophin C" motif. The most widely distributed peritrophins have the A-type peritrophin domain.

Representatives are found in all insect orders and other arthropod species including chelicerates and crustaceans. Proteins containing either the peritrophin B or the peritrophin C motif have been found only in dipterans.

The presence of the peritrophin-A domain in a protein does not imply that the corresponding genes are expressed in the gut. For example, *D. melanogaster* expresses a set of genes referred to as "*Obsructors*" encoding proteins with three CBM-14 domains, which are not expressed in the digestive tract. cDNAs encoding peritrophin-like proteins were isolated from cDNA libraries of hindgut and Malphighian tubules of adult *Ctenocephalides felis* fleas, a species that does not produce a PM (Gaines et al., 2003). This protein was also detected immunologically in Maplighian tubules, rectum and tracheae. There are several ways to denote this kind of chitin-binding domain, including ChtBD2 domain = Peritrophin A motif =CBM_14 =Pfam01607 = smart00494. Therefore to avoid confusion, this motif will be hereafter referred to as the "chitin-binding type II" domain (ChtBD2 domain) throughout this to describe these proteins with chitin-binding domains that have six cysteine residues and that occur not only in the PM, but also in other tissues.

1.5. Functional aspects of ChtBD2

The PM provides protection against the abrasive action of food materials and physical protection against the digestive enzymes. It also has a vital role in toxin filtration (e.g. tannins) recycling of the digestive enzymes, and facilitating nutrient and gas exchange. In plants many GlcNAc-specific lectins have been tested for insect toxicity (Harper et al., 1998; Macedo et al., 2003). These proteins bind to chitin or glycan receptors on the surface of cells lining the insect gut and disrupt the integrity of the PM. A novel approach has been proposed to develop strategies for insect control by utilizing chitin-binding molecules to specifically target formation

of the PM. Calcofluor white (CFW), a chemical whitener with chitin-binding properties, was used as a model compound in the diet to inhibit PM formation in *T. ni* and to increase larval susceptibility to baculovirus infection (Wang and Granados, 2000). CFW was also effective in suppressing PM formation in *S. frugiperda* and at the same time in preventing the establishment of a decreasing gradient of proteinases along the midgut tissue (Bolognesi et al., 2001). In the mite *Acarus siro*, combinations of diflubenzuron and CFW were more effective in reducing chitin content of the PM (Sobotnik et al., 2008). Hence, combinations of CFW with other insecticidal compounds affecting chitin synthesis may prove to be a useful strategy for insect control. Disruption of the PM structure following CFW treatment was consistently reported in various lepidopteran species (Bolognesi et al., 2001; Wang and Granados, 2000; Zhu et al., 2007). When larvae of *T. ni* and *S. exigua* were fed a CFW-containing diet, an increase in PM permeability was observed and the larvae became more susceptible to baculoviral infections. Interestingly, a significant amount of proteins was released upon CFW treatment, which may explain the altered permeability (Wang and Granados, 2000; Zhu et al., 2007). Immunization of sheep with protein extracts from the peritrophic matrix of *L. cuprina* larvae resulted in growth retardation and a reduction in lifespan of these insects feeding on the immunized sheep, although the PM structure was not affected (Tellam and Eisemann, 2000). Numerous sugar-binding proteins (lectins) from animals and plants such as galectins, WGA and chitinase-like lectins also bind chitin because of their high preference for GlcNAc. Like CFW, they disrupt PM formation in numerous cases, and therefore have been investigated for their insecticidal potential (Cohen, 2010). PM proteins have been tested as antigens in a vaccine to prevent *Plasmodium* infection in mosquitoes (Ramasamy et al., 1997).

In some cases, additional unrelated proteins that possess one or more chitin-binding domains (CBD) but are devoid of chitinolytic activity enhance degradation of chitin (Vaaje-Kolstad et al., 2005). This system also probably operates in the gut during degradation of PM and increases the porosity of the PM. It may also help in the digestion of chitin-containing prey (Bolognesi et al., 2005; Khajuria et al., 2010).

Inducible chitinolytic enzymes from bacteria cause insect mortality under certain conditions. These enzymes may compromise the structural integrity of the PM barrier and improve the effectiveness of a Bt toxin by enhancing contact of the toxin with its epithelial membrane receptor. For example, five chitinolytic bacterial strains isolated from midguts of *S. littoralis* induced a synergistic increase in larval mortality when combined with a Bt spore-crystal suspension relative to either an individual bacterial strain or a Bt suspension alone (Sneh et al., 1983).

Several GlcNAc-specific lectins from plants have been evaluated for insect toxicity (Harper et al., 1998; Macedo et al., 2003). These proteins appear to disrupt the integrity of the PM by binding to chitin or glycan receptors on the surface of cells lining the insect gut. Moreover, they may bind to glycosylated digestive enzymes and inhibit their activity. Apparently, these proteins bind to the PM, causing developmental abnormalities and reduced survival rates.

A protease from *A. gambiae* with a chitin-binding domain has been described, which may be involved in insect defense (Danielli et al., 2000; Gorman et al., 2000). This 147-kDa protein, sp22D, is expressed in a variety of tissues, most strongly in hemocytes, and is secreted into the hemolymph. Upon bacterial infection, the transcripts for this protein increase by about two-fold, suggesting a role in insect defense. This protein has a multidomain organization that includes

two copies of an *N*-terminal ChtBD2 domain, a *C*-terminal protease domain, and several receptor domains. It binds strongly to chitin and undergoes complex proteolytic processing during pupal to adult metamorphosis. It has been proposed that exposure of this protease to chitin may regulate its activity during tissue remodeling or wounding.

1.6. Bibliography

Aboitiz, N., Vila-Perello, M., Groves, P., Asensio, J.L., Andreu, D., Canada, F.J., Jimenez-Barbero, J., (2004). NMR and modeling studies of protein-carbohydrate interactions: synthesis, three-dimensional structure, and recognition properties of a minimum hevein domain with binding affinity for chitooligosaccharides. *Chembiochem* 5, 1245-1255.

Akiyama-Oda, Y., and Oda, H., (2006). Axis specification in the spider embryo: dpp is required for radial-to-axial symmetry transformation and sog for ventral patterning. *Development (Cambridge, England)* 133, 2347-2357.

Aljamali, M.N., Bior, A.D., Sauer, J.R., and , Essenberg, R.C., (2003). RNA interference in ticks: a study using histamine binding protein dsRNA in the female tick *Amblyomma americanum*. *Insect molecular biology* 12, 299-305.

Andersen, S.O., (2010). Insect cuticle sclerotization: A review. *Insect Mol Biol* 40, 166-178.

Andersen, S.O., Rafn, K., Roepstorff, P., (1997). Sequence studies of proteins from larval and pupal cuticle of the yellow meal worm, *Tenebrio molitor*. *Insect Biochem Mol Biol* 27, 121-131.

Arakane, Y., Hogenkamp, D.G., Zhu, Y.C., Kramer, K.J., Specht, C.A., Beeman, R.W., Kanost, M.R., Muthukrishnan, S., (2004). Characterization of two chitin synthase genes of the red flour beetle, *Tribolium castaneum*, and alternate exon usage in one of the genes during development. *Insect Biochem Mol Biol* 34, 291-304.

Arakane, Y., Zhu, Q., Matsumiya, M., Muthukrishnan, S., Kramer, K.J., (2003). Properties of catalytic, linker and chitin-binding domains of insect chitinase. *Insect Biochem Mol Biol* 33, 631-648.

Asensio, J.L., Canada, F.J., Bruix, M., Gonzalez, C., Khiar, N., Rodriguez-Romero, A., Jimenez-Barbero, J., (1998). NMR investigations of protein-carbohydrate interactions: refined three-dimensional structure of the complex between hevein and methyl beta-chitobioside. *Glycobiology* 8, 569-577.

Barry, M.K., Triplett, A.A., Christensen, A.C., (1999). A peritrophin-like protein expressed in the embryonic tracheae of *Drosophila melanogaster*. *Insect Biochem Mol Biol* 29, 319-327.

Behr, M., Hoch, M., (2005). Identification of the novel evolutionary conserved obstructor multigene family in invertebrates. *FEBS Lett* 579, 6827-6833.

Bolognesi, R., Arakane, Y., Muthukrishnan, S., Kramer, K.J., Terra, W.R., Ferreira, C., (2005). Sequences of cDNAs and expression of genes encoding chitin synthase and chitinase in the midgut of *Spodoptera frugiperda*. *Insect Biochem Mol Biol* 35, 1249-1259.

Bolognesi, R., Ribeiro, A.F., Terra, W.R., Ferreira, C., (2001). The peritrophic membrane of *Spodoptera frugiperda*: secretion of peritrophins and role in immobilization and recycling digestive enzymes. *Arch Insect Biochem Physiol* 47, 62-75.

Bolognesi, R., Terra, W.R., Ferreira, C., (2008). Peritrophic membrane role in enhancing digestive efficiency. Theoretical and experimental models. *J Insect Physiol* 54, 1413-1422.

Boraston, A.B., Bolam, D.N., Gilbert, H.J., Davies, G.J., (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* 382, 769-781.

Bucher, G., Scholten, J., Klingler, M., (2002). Parental RNAi in *Tribolium* (Coleoptera). *Curr Biol* 12, R85-86.

Campbell, P.M., Cao, A.T., Hines, E.R., East, P.D., Gordon, K.H., (2008). Proteomic analysis of the peritrophic matrix from the gut of the caterpillar, *Helicoverpa armigera*. *Insect Biochem Mol Biol* 38, 950-958.

Casu, R., Eisemann, C., Pearson, R., Riding, G., East, I., Donaldson, A., Cadogan, L., Tellam, R., (1997). Antibody-mediated inhibition of the growth of larvae from an insect causing cutaneous myiasis in a mammalian host. *Proc Natl Acad Sci U S A* 94, 8939-8944.

Chrispeels, M.J., Raikhel, N.V., (1991). Lectins, lectin genes, and their role in plant defense. *Plant Cell* 3, 1-9.

Cohen, E., 2010. *Chitin biochemistry: synthesis, hydrolysis and inhibition*. New York, NY: Academic Press.

Cornman, R.S., (2009). Molecular evolution of *Drosophila* cuticular protein genes. *PLoS One* 4, e8345.

Danielli, A., Loukeris, T.G., Lagueux, M., Muller, H.M., Richman, A., Kafatos, F.C., (2000). A modular chitin-binding protease associated with hemocytes and hemolymph in the mosquito *Anopheles gambiae*. *Proc Natl Acad Sci U S A* 97, 7136-7141.

Devenport, M., Fujioka, H., Donnelly-Doman, M., Shen, Z., Jacobs-Lorena, M., (2005). Storage and secretion of Ag-Aper14, a novel peritrophic matrix protein, and Ag-Muc1 from the mosquito Anopheles gambiae. *Cell and tissue research* 320, 175-185.

Dixit, R., Arakane, Y., Specht, C.A., Richard, C., Kramer, K.J., Beeman, R.W., Muthukrishnan, S., (2008). Domain organization and phylogenetic analysis of proteins from the chitin deacetylase gene family of *Tribolium castaneum* and three other species of insects. *Insect Biochem Mol Biol* 38, 440-451.

Eisemann, C., Wijffels, G., Tellam, R.L., (2001). Secretion of the type 2 peritrophic matrix protein, peritrophin-15, from the cardia. *Arch Insect Biochem Physiol* 47, 76-85.

Elvin, C.M., Vuocolo, T., Pearson, R.D., East, I.J., Riding, G.A., Eisemann, C.H., Tellam, R.L., (1996). Characterization of a major peritrophic membrane protein, peritrophin-44, from the

larvae of *Lucilia cuprina*. cDNA and deduced amino acid sequences. *J Biol Chem* 271, 8925-8935.

Fanelli, E., Di Vito, M., Jones, J.T., De Giorgi, C., (2005). Analysis of chitin synthase function in a plant parasitic nematode, *Meloidogyne artiellia*, using RNAi. *Gene* 349, 87-95.

Felix, M.A., (2008). RNA interference in nematodes and the chance that favored Sydney Brenner. *Journal of biology* 7, 34.

Ferreira, A.H., Cristofoletti, P.T., Lorenzini, D.M., Guerra, L.O., Paiva, P.B., Briones, M.R., Terra, W.R., Ferreira, C., (2007). Identification of midgut microvillar proteins from *Tenebrio molitor* and *Spodoptera frugiperda* by cDNA library screenings with antibodies. *J Insect Physiol* 53, 1112-1124.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C., (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806-811.

Fujitani, N., Kouno, T., Nakahara, T., Takaya, K., Osaki, T., Kawabata, S., Mizuguchi, M., Aizawa, T., Demura, M., Nishimura, S., Kawano, K., (2007). The solution structure of horseshoe crab antimicrobial peptide tachystatin B with an inhibitory cystine-knot motif. *J Pept Sci* 13, 269-279.

Futahashi, R., Okamoto, S., Kawasaki, H., Zhong, Y.S., Iwanaga, M., Mita, K., Fujiwara, H., (2008). Genome-wide identification of cuticular protein genes in the silkworm, *Bombyx mori*. *Insect Biochem Mol Biol* 38, 1138-1146.

Gaines, P.J., Walmsley, S.J., Wisnewski, N., (2003). Cloning and characterization of five cDNAs encoding peritrophin-A domains from the cat flea, *Ctenocephalides felis*. *Insect Biochem Mol Biol* 33, 1061-1073.

Gorman, M.J., Andreeva, O.V., Paskewitz, S.M., (2000). Sp22D: a multidomain serine protease with a putative role in insect immunity. *Gene* 251, 9-17.

Guan, X., Middlebrooks, B.W., Alexander, S., Wasserman, S.A., (2006). Mutation of TweedleD, a member of an unconventional cuticle protein family, alters body shape in *Drosophila*. *Proc Natl Acad Sci U S A* 103, 16794-16799.

Guo, W., Li, G., Pang, Y., Wang, P., (2005). A novel chitin-binding protein identified from the peritrophic membrane of the cabbage looper, Trichoplusia ni. *Insect Biochem Mol Biol* 35, 1224-1234.

Hamodrakas, S.J., Willis, J.H., Iconomidou, V.A., (2002). A structural model of the chitin-binding domain of cuticle proteins. *Insect Biochem Mol Biol* 32, 1577-1583.

Hao, Z., Aksoy, S., (2002). Proventriculus-specific cDNAs characterized from the tsetse, *Glossina morsitans morsitans*. *Insect Biochem Mol Biol* 32, 1663-1671.

Harper, M.S., Hopkins, T.L., (1997). Peritrophic membrane structure and secretion in European corn borer larvae (*Ostrinia nubilalis*). *Tissue Cell* 29, 463-475.

Harper, M.S., Hopkins, T.L., Czapla, T.H., (1998). Effect of wheat germ agglutinin on formation and structure of the peritrophic membrane in European corn borer (*Ostrinia nubilalis*) larvae. *Tissue Cell* 30, 166-176.

Havemann, J., Muller, U., Berger, J., Schwarz, H., Gerberding, M., Moussian, B., (2008). Cuticle differentiation in the embryo of the amphipod crustacean *Parhyale hawaiensis*. *Cell Tissue Res* 332, 359-370.

He, N., Botelho, J.M., McNall, R.J., Belozerov, V., Dunn, W.A., Mize, T., Orlando, R., Willis, J.H., (2007). Proteomic analysis of cast cuticles from *Anopheles gambiae* by tandem mass spectrometry. *Insect Biochem Mol Biol* 37, 135-146.

Hegedus, D., Erlandson, M., Gillott, C., Toprak, U., (2009). New insights into peritrophic matrix synthesis, architecture, and function. *Annu Rev Entomol* 54, 285-302.

Hemmi, H., Ishibashi, J., Tomie, T., Yamakawa, M., (2003). Structural basis for new pattern of conserved amino acid residues related to chitin-binding in the antifungal peptide from the coconut rhinoceros beetle *Oryctes rhinoceros*. *J Biol Chem* 278, 22820-22827.

Honeybee Genome Sequencing Consortium, (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*
. *Nature* 443, 931-949.

Hopkins, T.L., Harper, M.S., (2001). Lepidopteran peritrophic membranes and effects of dietary wheat germ agglutinin on their formation and structure. *Arch Insect Biochem Physiol* 47, 100-109.

Iconomidou, V.A., Willis, J.H., Hamodrakas, S.J., (2005). Unique features of the structural model of 'hard' cuticle proteins: implications for chitin-protein interactions and cross-linking in cuticle. *Insect Biochem Mol Biol* 35, 553-560.

Iijima, M., Hashimoto, T., Matsuda, Y., Nagai, T., Yamano, Y., Ichi, T., Osaki, T., Kawabata, S., (2005). Comprehensive sequence analysis of horseshoe crab cuticular proteins and their involvement in transglutaminase-dependent cross-linking. *FEBS J* 272, 4774-4786.

Ikeya, T., Persson, P., Kono, M., Watanabe, T., (2001). The DD5 gene of the decapod crustacean *Penaeus japonicus* encodes a putative exoskeletal protein with a novel tandem repeat structure. *Comp Biochem Physiol B Biochem Mol Biol* 128, 379-388.

Jang, M.K., Kong, B.G., Jeong, Y.I., Lee, C.H., Nah, J.W., (2004). Physicochemical characterization of alpha-chitin, beta-chitin, and gamma-chitin separated from natural resources. *Journal of Polymer Science Part a-Polymer Chemistry* 42, 3423-3432.

Karouzou, M.V., Spyropoulos, Y., Iconomidou, V.A., Cornman, R.S., Hamodrakas, S.J., Willis, J.H., (2007). *Drosophila* cuticular proteins with the R&R Consensus: annotation and

classification with a new tool for discriminating RR-1 and RR-2 sequences. *Insect Biochem Mol Biol* 37, 754-760.

Khajuria, C., Buschman, L.L., Chen, M.S., Muthukrishnan, S., Zhu, K.Y., (2010). A gut-specific chitinase gene essential for regulation of chitin content of peritrophic matrix and growth of *Ostrinia nubilalis* larvae. *Insect Biochem Mol Biol* 40, 621-629.

Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W., Richardson, P., (2002). Genetic and genomic tools for *Xenopus* research: The NIH *Xenopus* initiative. *Dev Dyn* 225, 384-391.

Kramer, K.J., Corpuz, L., Choi, H. K., Muthukrishnan, S., (1993). Sequence of a cDNAand expression of the gene encoding epidermal and gut chitinases of *Manduca sexta*. *Insect Biochem Mol Biol* 23, 691-701.

Kramer, K.J., Muthukrishnan, S., (1997). Insect chitinases: molecular biology and potential use as biopesticides. *Insect Biochem Mol Biol* 27, 887-900.

Kramer, K.J., Muthukrishnan, S., 2005. Chitin metabolism in insects, in: Gilbert, L.I., Iatrou, K., Gill, S. (Eds.), Comprehensive Molecular Insect Science. Biochemistry and Molecular Biology,. Elsevier Press, Oxford, UK,, pp. 111-144.

Kucharski, R., Maleszka, J., Maleszka, R., (2007). Novel cuticular proteins revealed by the honey bee genome. *Insect Biochem Mol Biol* 37, 128-134.

Loongyai, W., Avarre, J.M., C., Lubzens, E., Chotigeat, W., (2007). Isolation and functional characterization of a new shrimp ovarian peritrophin (SOP) from the ovary of *Fenneropenaeus merguiensis*. . *Mar Biotechnol* 9, 624-637.

Macedo, M.L., Damico, D.C., Freire, M.G., Toyama, M.H., Marangoni, S., Novello, J.C., (2003). Purification and characterization of an N-acetylglucosamine-binding lectin from *Koelreuteria paniculata* seeds and its effect on the larval development of *Callosobruchus maculatus* (Coleoptera: Bruchidae) and *Anagasta kuehniella* (Lepidoptera: Pyralidae). *J Agric Food Chem* 51, 2980-2986.

Mittapalli, O., Sardesai, N., Shukle, R.H., (2007). cDNA cloning and transcriptional expression of a peritrophin-like gene in the Hessian fly, *Mayetiola destructor* [Say]. *Arch Insect Biochem Physiol* 64, 19-29.

Moussian, B., Schwarz, H., Bartoszewski, S., Nusslein-Volhard, C., (2005). Involvement of chitin in exoskeleton morphogenesis in *Drosophila melanogaster*. *J Morphol* 264, 117-130.

Moussian, B., Veerkamp, J., Muller, U., Schwarz, H., (2007). Assembly of the *Drosophila* larval exoskeleton requires controlled secretion and shaping of the apical plasma membrane. *Matrix Biol* 26, 337-347.

Narasimhan, S., Montgomery, R.R., DePonte, K., Tschudi, C., Marcantonio, N., Anderson, J.F., Sauer, J.R., Cappello, M., Kantor, F.S., Fikrig, E., (2004). Disruption of *Ixodes scapularis*

anticoagulation by using RNA interference. *Proceedings of the National Academy of Sciences of the United States of America* 101, 1141-1146.

Osaki, T., Omotezako, M., Nagayama, R., Hirata, M., Iwanaga, S., Kasahara, J., Hattori, J., Ito, I., Sugiyama, H., Kawabata, S., (1999). Horseshoe crab hemocyte-derived antimicrobial polypeptides, tachystatins, with sequence similarity to spider neurotoxins. *J Biol Chem* 274, 26172-26178.

Peters, W., 1992. *Peritrophic Membranes*. Berlin: Springer.

Qui, J., Hardin, P.E., (1995). Temporal and spatial expression of an adult cuticle protein gene from *Drosophila* suggests that its protein product may impart some specialized cuticle function. *Dev. Biol*. 167, 416-425.

Ramasamy, R., I. C. Wanniarachchi, K. A. Srikrishnaraj, Ramasamy., M.S., (1997). Mosquito midgut glycoproteins and recognition sites for malaria parasites. *Biochim. Biophysiol.Acta* 1361, 114-122.

Rayms-Keller, A., McGaw, M., Oray, C., Carlson, J.O., Beaty, B.J., (2000). Molecular cloning and characterization of a metal responsive *Aedes aegypti* intestinal mucin cDNA. *Insect Mol Biol* 9, 419-426.

Rebers, J.E., Willis, J.H., (2001). A conserved domain in arthropod cuticular proteins binds chitin. *Insect Biochem Mol Biol* 31, 1083-1093.

Richards, S., Gibbs, R.A., Weinstock, G.M., Brown, S.J., Denell, R., Beeman, R.W., Gibbs, R., Bucher, G., Friedrich, M., Grimmelikhuijzen, C.J., Klingler, M., Lorenzen, M., Roth, S., Schroder, R., Tautz, D., Zdobnov, E.M., Muzny, D., Attaway, T., Bell, S., Buhay, C.J., Chandrabose, M.N., Chavez, D., Clerk-Blankenburg, K.P., Cree, A., Dao, M., Davis, C., Chacko, J., Dinh, H., Dugan-Rocha, S., Fowler, G., Garner, T.T., Garnes, J., Gnirke, A., Hawes, A., Hernandez, J., Hines, S., Holder, M., Hume, J., Jhangiani, S.N., Joshi, V., Khan, Z.M., Jackson, L., Kovar, C., Kowis, A., Lee, S., Lewis, L.R., Margolis, J., Morgan, M., Nazareth, L.V., Nguyen, N., Okwuonu, G., Parker, D., Ruiz, S.J., Santibanez, J., Savard, J., Scherer, S.E., Schneider, B., Sodergren, E., Vattahil, S., Villasana, D., White, C.S., Wright, R., Park, Y., Lord, J., Oppert, B., Brown, S., Wang, L., Weinstock, G., Liu, Y., Worley, K., Elsik, C.G., Reese, J.T., Elhaik, E., Landan, G., Graur, D., Arensburger, P., Atkinson, P., Beidler, J., Demuth, J.P., Drury, D.W., Du, Y.Z., Fujiwara, H., Maselli, V., Osanai, M., Robertson, H.M., Tu, Z., Wang, J.J., Wang, S., Song, H., Zhang, L., Werner, D., Stanke, M., Morgenstern, B., Solovyev, V., Kosarev, P., Brown, G., Chen, H.C., Ermolaeva, O., Hlavina, W., Kapustin, Y., Kiryutin, B., Kitts, P., Maglott, D., Pruitt, K., Sapojnikov, V., Souvorov, A., Mackey, A.J., Waterhouse, R.M., Wyder, S., Kriventseva, E.V., Kadowaki, T., Bork, P., Aranda, M., Bao, R., Beermann, A., Berns, N., Bolognesi, R., Bonneton, F., Bopp, D., Butts, T., Chaumot, A., Denell, R.E., Ferrier, D.E., Gordon, C.M., Jindra, M., Lan, Q., Lattorff, H.M., Laudet, V., von Levetsow, C., Liu, Z., Lutz, R., Lynch, J.A., da Fonseca, R.N., Posnien, N., Reuter, R., Schinko, J.B., Schmitt, C., Schoppmeier, M., Shippy, T.D., Simonnet, F., Marques-Souza, H., Tomoyasu, Y., Trauner, J., Van der Zee, M., Vervoort, M., Wittkopp, N., Wimmer, E.A., Yang, X., Jones, A.K., Sattelle, D.B., Ebert, P.R., Nelson, D., Scott, J.G., Muthukrishnan, S., Kramer, K.J., Arakane, Y., Zhu,

Q., Hogenkamp, D., Dixit, R., Jiang, H., Zou, Z., Marshall, J., Elpidina, E., Vinokurov, K., Oppert, C., Evans, J., Lu, Z., Zhao, P., Sumathipala, N., Altincicek, B., Vilcinskas, A., Williams, M., Hultmark, D., Hetru, C., Hauser, F., Cazzamali, G., Williamson, M., Li, B., Tanaka, Y., Predel, R., Neupert, S., Schachtner, J., Verleyen, P., Raible, F., Walden, K.K., Angeli, S., Foret, S., Schuetz, S., Maleszka, R., Miller, S.C., Grossmann, D., (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 452, 949-955.

Robalino, J., Bartlett, T., Shepard, E., Prior, S., Jaramillo, G., Scura, E., Chapman, R.W., Gross, P.S., Browdy, C.L., Warr, G.W., (2005). Double-stranded RNA induces sequence specific antiviral silencing in addition to nonspecific immunity in a marine shrimp: convergence of RNA interference and innate immunity in the invertebrate antiviral response? *Journal of Virology* 79, 13561-13571.

Roignant, J.Y., Carre, C., Mugat, B., Szymczak, D., Lepesant, J.A., and Antoniewski, C., (2003). Absence of transitive and systemic pathways allows cell-specific and isoform-specific RNAi in *Drosophila*. *RNA (New York, NY)* 9, 299-308.

Sanchez Alvarado, A., Newmark, P.A., (1999). Double-stranded RNA specifically disrupts gene expression during planarian regeneration. *Proceedings of the National Academy of Sciences of the United States of America* 96, 5049-5054.

Sarauer, B.L., Gillott, C., Hegedus, D., (2003). Characterization of an intestinal mucin from the peritrophic matrix of the diamondback moth, *Plutella xylostella*. *Insect Mol Biol* 12, 333-343.

Sauvion, N., Rahbe, Y., Peumans, W., van Damme, E., Gatehouse, J., Gatehouse, A., (1996). Effects of GNA and other mannose binding lectins on development and fecundity of the peach-potato aphid *Myzus persicae*. *Entomol.Exp.Appl*. 79, 285-293.

Schorderet, S., Pearson, R.D., Vuocolo, T., Eisemann, C., Riding, G.A., Tellam, R.L., (1998). cDNA and deduced amino acid sequences of a peritrophic membrane glycoprotein, 'peritrophin-48', from the larvae of *Lucilia cuprina*. *Insect Biochem Mol Biol* 28, 99-111.

Shao, L., Devenport, M., Jacobs-Lorena, M., (2001). The peritrophic matrix of hematophagous insects. *Arch Insect Biochem Physiol* 47, 119-125.

Shen, Z., Jacobs-Lorena, M., (1999). Evolution of chitin-binding proteins in invertebrates. *J Mol Evol* 48, 341-347.

Shi, X., Chamankhah, M., Visal-Shah, S., Hemmingsen, S.M., Erlandson, M., Braun, L., Alting-Mees, M., Khachatourians, G.G., O'Grady, M., Hegedus, D.D., (2004). Modeling the structure of the type I peritrophic matrix: characterization of a *Mamestra configurata* intestinal mucin and a novel peritrophin containing 19 chitin binding domains. *Insect Biochem Mol Biol* 34, 1101-1115.

Sneh, B., Schuster, S., Gross, S., (1983). Biological control of *Spodoptera littoralis* (Boisd.) (Lepidoptera: Noctuidae) by *Bacillus thuringiensis* subsp. entomocidus and *Bracon hebetor* Say (Hymenoptera:Braconidae). *Z. Ang. Entomol* 96, 77-83.

Soares, C.A., Lima, C.M., Dolan, M.C., Piesman, J., Beard, C.B., Zeidner, N.S., (2005). Capillary feeding of specific dsRNA induces silencing of the isac gene in nymphal *Ixodes scapularis* ticks. *Insect Mol Biol* 14, 443-452.

Sobotnik, J., Kudlikova-Krizkova, I., Vancova, M., Munzbergova, Z., Hubert, J., (2008). Chitin in the peritrophic membrane of *Acarus siro* (Acari: Acaridae) as a target for novel acaricides. *J Econ Entomol* 101, 1028-1033.

Suderman, R.J., Dittmer, N.T., Kanost, M.R., Kramer, K.J., (2006). Model reactions for insect cuticle sclerotization: cross-linking of recombinant cuticular proteins upon their laccase-catalyzed oxidative conjugation with catechols. *Insect Biochem Mol Biol* 36, 353-365.

Suetake, T., Tsuda, S., Kawabata, S., Miura, K., Iwanaga, S., Hikichi, K., Nitta, K., Kawano, K., (2000). Chitin-binding proteins in invertebrates and plants comprise a common chitin-binding structural motif. *J Biol Chem* 275, 17929-17932.

Tellam, R.L., Eisemann, C., (2000). Chitin is only a minor component of the peritrophic matrix from larvae of *Lucilia cuprina*. *Insect Biochem Mol Biol* 30, 1189-1201.

Tellam, R.L., Vuocolo, T., Eisemann, C., Briscoe, S., Riding, G., Elvin, C., Pearson, R., (2003). Identification of an immuno-protective mucin-like protein, peritrophin-55, from the peritrophic matrix of *Lucilia cuprina* larvae. *Insect Biochem Mol Biol* 33, 239-252.

Tellam, R.L., Wijffels, G., Willadsen, P., (1999). Peritrophic matrix proteins. *Insect Biochem Mol Biol* 29, 87-101.

Terra, W.R., (2001). The origin and functions of the insect peritrophic membrane and peritrophic gel. *Arch Insect Biochem Physiol* 47, 47-61.

Togawa, T., Nakato, H., Izumi, S., (2004). Analysis of the chitin recognition mechanism of cuticle proteins from the soft cuticle of the silkworm, *Bombyx mori*. *Insect Biochem Mol Biol* 34, 1059-1067.

Tomoyasu, Y., Denell, R.E., (2004). Larval RNAi in *Tribolium* (Coleoptera) for analyzing adult development. *Dev Genes Evol* 214, 575-578.

Toprak, U., Baldwin, D., Erlandson, M., Gillott, C., Hegedus, D.D., (2009). Insect intestinal mucins and serine proteases associated with the peritrophic matrix from feeding, starved and moulting *Mamestra configurata* larvae. *Insect Mol Biol* 19, 163-175.

Toprak, U., Baldwin, D., Erlandson, M., Gillott, C., Hou, X., Coutu, C., Hegedus, D.D., (2008). A chitin deacetylase and putative insect intestinal lipases are components of the Mamestra configurata (Lepidoptera: Noctuidae) peritrophic matrix. *Insect Mol Biol* 17, 573-585.

Vaaje-Kolstad, G., Horn, S.J., van Aalten, D.M., Synstad, B., Eijsink, V.G., (2005). The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. *J Biol Chem* 280, 28492-28497.

Vuocolo, T., Eisemann, C.H., Pearson, R.D., Willadsen, P., Tellam, R.L., (2001). Identification and molecular characterisation of a peritrophin gene, peritrophin-48, from the myiasis fly *Chrysomya bezziana*. *Insect Biochem Mol Biol* 31, 919-932.

Wang, P., Granados, R.R., (1997a). An intestinal mucin is the target substrate for a baculovirus enhancin. *Proc Natl Acad Sci U S A* 94, 6977-6982.

Wang, P., Granados, R.R., (1997b). Molecular cloning and sequencing of a novel invertebrate intestinal mucin cDNA. *J Biol Chem* 272, 16663-16669.

Wang, P., Granados, R.R., (2000). Calcofluor disrupts the midgut defense system in insects. *Insect Biochem Mol Biol* 30, 135-143.

Wang, W., Davison, S., Krell, P.J., (2004). Identification and characterization of a major early-transcribed gene of *Trichoplusia ni* single nucleocapsid nucleopolyhedrovirus using the baculovirus expression system. *Virus Genes* 29, 19-29.

Waterhouse, P.M., Graham, M.W.,Wang, M.B., (1998). Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc. Natl. Acad. Sci.* 95, 13959-13964.

Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K., Beukeboom, L.W., Desplan, C., Elsik, C.G., Grimmelikhuijzen, C.J., Kitts, P., Lynch, J.A., Murphy, T., Oliveira, D.C., Smith, C.D., van de Zande, L., Worley, K.C., Zdobnov, E.M., Aerts, M., Albert, S., Anaya, V.H., Anzola, J.M., Barchuk, A.R., Behura, S.K., Bera, A.N., Berenbaum, M.R., Bertossa, R.C., Bitondi, M.M., Bordenstein, S.R., Bork, P., Bornberg-Bauer, E., Brunain, M., Cazzamali, G., Chaboub, L., Chacko, J., Chavez, D., Childers, C.P., Choi, J.H., Clark, M.E., Claudianos, C., Clinton, R.A., Cree, A.G., Cristino, A.S., Dang, P.M., Darby, A.C., de Graaf, D.C., Devreese, B., Dinh, H.H., Edwards, R., Elango, N., Elhaik, E., Ermolaeva, O., Evans, J.D., Foret, S., Fowler, G.R., Gerlach, D., Gibson, J.D., Gilbert, D.G., Graur, D., Grunder, S., Hagen, D.E., Han, Y., Hauser, F., Hultmark, D., Hunter, H.C.t., Hurst, G.D., Jhangian, S.N., Jiang, H., Johnson, R.M., Jones, A.K., Junier, T., Kadowaki, T., Kamping, A., Kapustin, Y., Kechavarzi, B., Kim, J., Kiryutin, B., Koevoets, T., Kovar, C.L., Kriventseva, E.V., Kucharski, R., Lee, H., Lee, S.L., Lees, K., Lewis, L.R., Loehlin, D.W., Logsdon, J.M., Jr., Lopez, J.A., Lozado, R.J., Maglott, D., Maleszka, R., Mayampurath, A., Mazur, D.J., McClure, M.A., Moore, A.D., Morgan, M.B., Muller, J., Munoz-Torres, M.C., Muzny, D.M., Nazareth, L.V., Neupert, S., Nguyen, N.B., Nunes, F.M., Oakeshott, J.G., Okwuonu, G.O., Pannebakker, B.A., Pejaver, V.R., Peng, Z., Pratt, S.C., Predel, R., Pu, L.L., Ranson, H., Raychoudhury, R., Rechtsteiner, A., Reese, J.T., Reid, J.G., Riddle, M., Robertson, H.M., Romero-Severson, J., Rosenberg, M., Sackton, T.B., Sattelle, D.B., Schluns, H., Schmitt, T., Schneider, M., Schuler, A., Schurko, A.M., Shuker, D.M., Simoes, Z.L., Sinha, S., Smith, Z., Solovyev, V., Souvorov, A., Springauf, A., Stafflinger, E., Stage, D.E., Stanke, M., Tanaka, Y., Telschow, A., Trent, C., Vattathil, S., Verhulst, E.C., Viljakainen, L., Wanner, K.W., Waterhouse, R.M., Whitfield, J.B., Wilkes, T.E., Williamson, M., Willis, J.H., Wolschin, F., Wyder, S., Yamada, T., Yi, S.V., Zecher, C.N., Zhang, L., Gibbs, R.A., (2010). Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* 327, 343-348.

Wigglesworth, V., (1930). The formation of the peritrophic membrane in insects, with special reference to the larvae of mosquitoes. *Q. F. Microsc. Sci.* 73, 583-616.

Wijffels, G., Eisemann, C., Riding, G., Pearson, R., Jones, A., Willadsen, P., Tellam, R., (2001). A novel family of chitin-binding proteins from insect type 2 peritrophic matrix. cDNA sequences, chitin binding activity, and cellular localization. *J Biol Chem* 276, 15527-15536.

Willis, J.H., (2010). Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. *Insect Biochem Mol Biol* 40, 189-204.

Zhang, Y., Foster, J.M., Nelson, L.S., Ma, D., Carlow, C.K., (2005). The chitin synthase genes *chs-1* and *chs-2* are essential for *C. elegans* development and responsible for chitin deposition in the eggshell and pharynx, respectively. *Dev Biol* 285, 330-339.

Zhong, Y.S., Mita, K., Shimada, T., Kawasaki, H., (2006). Glycine-rich protein genes, which encode a major component of the cuticle, have different developmental profiles from other cuticle protein genes in *Bombyx mori*. *Insect Biochem Mol Biol* 36, 99-110.

Zhu, Q., Arakane, Y., Banerjee, D., Beeman, R.W., Kramer, K.J., Muthukrishnan, S., (2008a). Domain organization and phylogenetic analysis of the chitinase-like family of proteins in three species of insects. *Insect Biochem Mol Biol* 38, 452-466.

Zhu, R., Liu, K., Peng, J., Yang, H., Hong, H., (2007). Optical brightener M2R destroys the peritrophic membrane of *Spodoptera exigua* (Lepidoptera: Noctuidae) larvae. *Pest Manag Sci* 63, 296-300.

CHAPTER 2

Bioinformatics of ChtBD2 domain-containing proteins in *Tribolium castaneum*

Title: Genes encoding proteins with peritrophin A-type chitin-binding domains in *Tribolium castaneum* are grouped into three distinct families based on phylogeny, expression and function

Most of the material presented in this chapter was published in Insect Biochemistry and Molecular Biology (2010) **40**: 214-27. The original publication is available at www.elsevier.com.

2.1. Abstract

This study is focused on the characterization and expression of genes in the red flour beetle, *T. castaneum*, encoding proteins that possess one or more six-cysteine-containing chitin-binding domains related to the peritrophin A domain (ChtBD2). An bioinformatics search of the genome of *T. castaneum* queried with ChtBD2 sequences yielded 13 previously characterized chitin metabolic enzymes and 29 additional proteins with signal peptides as well as one to 14 ChtBD2s. Using phylogenetic analyses, these additional 29 proteins were classified into three large families. The first family includes 11 proteins closely related to the peritrophins, each containing one to 14 ChtBD2s. These are midgut-specific and are expressed only during feeding stages. I propose the name "Peritrophic Matrix Proteins" (PMP) for this family. The second family contains eight proteins encoded by seven genes (one gene codes for 2 splice variants), which are closely related to "gasp/obstructor"-like proteins that contain 3 ChtBD2s each. The third family has ten proteins that are of diverse sizes and sequences with only one ChtBD2 each. The genes of the second and third families are expressed in non-midgut tissues throughout all stages of development. I propose the names "Cuticular Proteins Analogous to Peritophins 3" (CPAP3) for the second family that has three ChtBD2s and "Cuticular Proteins Analogous to Peritophins 1 (CPAP1) for the third family that has 1 ChtBD2. Even though proteins of both CPAP1 and CPAP3 families have the "peritrophin A" domain, they are expressed only in cuticle-forming tissues. I determined the exon-intron organization of the genes, encoding these 29 proteins as well as the domain organization of the encoded proteins with ChtBD2s. All 29 proteins have predicted cleavable signal peptides and ChtBD2s, suggesting that they interact with chitin in extracellular locations. Comparison of ChtBD2-containing proteins representing

different arthropod taxa suggests that ChtBD2s are ancient protein domains whose affinity for chitin in extracellular matrices has been exploited many times for a range of biological functions.

2.2. Introduction

There are two main classes of chitin-binding motifs that have been described in insect proteins. The first class contains a sequence consisting of six cysteines that probably form three disulfide bridges. This sequence motif is referred to as the "peritrophin A domain" and is found in numerous proteins extracted from insect PMs. This motif belongs to the CBM14 family of carbohydrate-binding domains (pfam 01607; (Elvin et al., 1996)) also known as the type 2 chitin-binding domain (ChtBD2=smart00494; (Tjoelker et al., 2000). The second general class of chitin-binding sequence motifs is the Rebers and Riddiford consensus sequence (RR consensus; PDGDYNY+YETSNGIADQETGD+KSQGETRDG++AVDVV+GSYSYVDPDGTTRTVTYT ADDENGFQPVGAHLP; pfam00373;(Karouzou et al., 2007); (Willis, 2010), which lacks cysteine residues. Many of these proteins are rich in histidines, which have been postulated to participate in cross-linking reactions of these proteins via quinones derived from N-ß-alanyldopamine and N-acetyldopamine (Andersen, 2010; Kramer et al., 2001). Some of these RR proteins were identified in insect cuticle extracts by proteomic analyses (Willis, 2010).

Proteins with ChtBD2s that have six cysteine residues with a characteristic spacing between them were initially extracted from the insect PM using strong denaturing agents and were appropriately denoted as "peritrophins" (Tellam et al., 1999); (Shen and Jacobs-Lorena, 1999). Later on (Barry et al., 1999) and (Behr and Hoch, 2005) identified cDNAs/genes encoding proteins with this peritrophin-A type chitin-binding motif (ChtBD2s) in cuticle-

forming tissues from *D. melanogaster*. These proteins have three ChtBD2s separated by spacers of characteristic lengths. (Gaines et al., 2003) characterized five cDNA clones from RNA expressed in the hindgut and Malpighian tubules of the cat flea, *Ctenocephalides felis,* which encode proteins with one to four peritrophinA domains. Thus, it is clear that the ChtBD2 is present in proteins that are expressed in locations other than the midgut that secretes the PM. However, the group of proteins isolated from the PM after extraction with strong denaturing agents, collectively called "peritrophins," exhibit a wider variation in the number of ChtBD2s and, consequently, in sizes. Several other proteins with a large number of ChtBD2 repeats have been predicted from genome sequences of insects, even though they have not been extracted directly from PM or gut tissue. The number of ChtBD2s in "peritrophins" can range from one to 19 (reviewed in: (Dinglasan et al., 2009);(Hegedus et al., 2009); (Tellam et al., 1999); (Venancio et al., 2009)). They are also interspersed with serine and threonine-rich mucin-like domains that are likely to be glycosylated. These studies have revealed that the group of ChtBD2-containing proteins is much larger than previously known. However, a bioinformatics study cataloging all of the proteins with one or more ChtBD2s encoded by a single genome had not been undertaken prior to this work. Here, I have used a bioinformatics search to identify in the red flour beetle, *T. castaneum*, all of the genes that encode proteins containing the ChtBD2, which has been associated with the ability to bind to chitin. My studies have revealed that the *T. castaneum* genome encodes a large assortment of proteins with ChtBD2s in addition to the PM-associated peritrophins and enzymes of chitin metabolism such as chitinases and chitin deacetylases, some of which have ChtBD2s. A vast majority of them have signal peptide sequences at the N-terminus consistent with their presumed roles in binding to chitin in extracellular matrices. I propose a new nomenclature for gene families that encode nonenzymatic

proteins with one or more ChtBD2s, the first being Peritrophic Matrrix Proteins (PMPs) for proteins expressed in the gut, and the second "Cuticular Proteins Analogous to Peritophins" (CPAPs) for proteins, expressed in cuticle-secreting tissues.

2.3. Materials and Methods

2.3.1. Insect cultures

The GA-1 strain of *T. castaneum* was used in all experiments. Insects were reared at 30°C in wheat flour containing 5% brewer's yeast under standard conditions as described previously (Beeman and Stuart, 1990).

2.3.2. Identification of the gene encoding proteins with ChtBD2s

To identify genes encoding proteins with one or more ChtBD2s, we conducted a bioinformatics search of the *T. castaneum* genome and EST databases (Beetlebase, http://www.bioinformatics.ksu.edu/Beetlebase; Genboree, NCBI, (www.ncbi.nlm.nih.gov/) and Baylor), using previously characterized chitin-binding domains from insect peritrophins A, B and C with six, eight and six cysteines, respectively, as queries (Tellam et al., 1999). Similarly, several other ChtBD2s from *T. ni*, *L. cuprina*, *A. gambiae* and *D. melanogaster* were also utilized as queries to identify genes encoding proteins with ChtBD2s in *T. castaneum*.

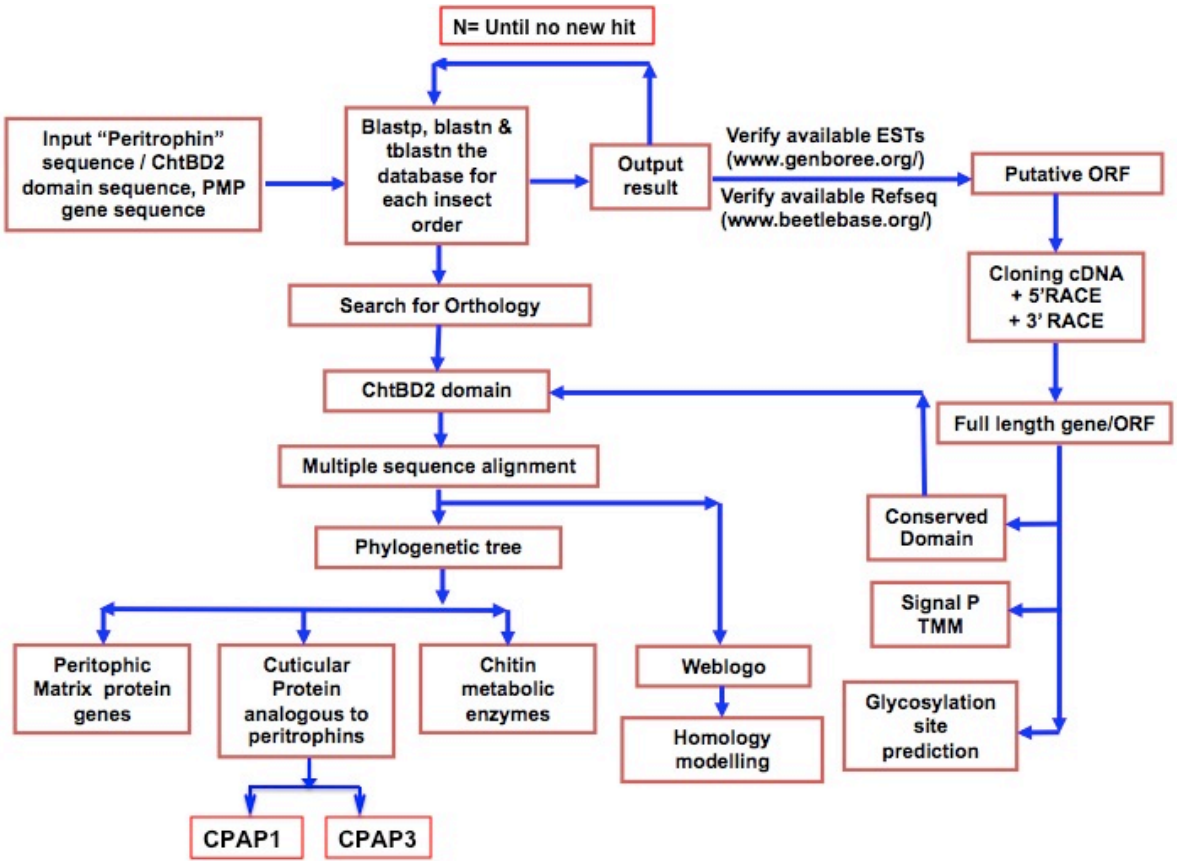


Figure 2-1 Strategy for identification of all the genes encoding ChtBD2 proteins in *T. castaneum* genome.

2.3.3. Cloning of cDNAs

To clone the cDNAs corresponding to the genes identified in the bioinformatics search, gene-specific primers for each gene were designed from available expressed sequence tags (ESTs) or from open reading frames (ORFs) in GLEAN predictions from the *T. castaneum* genome sequences (Beetlebase website: http://www.bioinformatics.ksu.edu/BeetleBase). Total RNA (2.5-5 μg) isolated by RNeasy Mini kit (Qiagen, Valencia, CA) from whole insects at

various developmental stages according to manufacturer's protocol was used to synthesize cDNA. The SuperScript III first-strand synthesis system for RT-PCR (Invitrogen, Carlsbad, CA) was used to synthesize first-strand cDNA by following manufacturer's protocol. DNA fragments were amplified using pairs of gene-specific primers designed from sequences in the 5'- and 3'- UTR regions or within the ORFs of target genes, and were cloned into the pGEMT vector (Promega, Madison, WI). To obtain the full-length cDNA, 5' RACE (RNA ligase-mediated rapid amplification of cDNA ends) was also done using the Gene Racer kit (Invitrogen, Carlsbad, CA) by following the instructions from the manufacturer. DNA sequencing was conducted at the DNA sequencing facility at Kansas State University (ABI Prism 3730). Using this approach, we obtained near full-length cDNAs covering the ORFs of each of the peritrophins and CPAPs.

2.3.4. DNA and protein sequence analysis

After sequencing full-length or near full-length cDNAs of *TcPMPs* and *TcCPAPs*, exon-intron organization of each gene was determined by sequence comparisons to the corresponding genomic sequences. Protein sequence analysis tools used in this study include those for open reading frame (ORF), MW and pI predictions at the ExPASy Proteomics website (http://us.expasy.org/). Domains in the protein sequences were identified *via* Conserved Domain Searches at NCBI (www.ncbi.nlm.nih.gov/) and Scanprosite. Potential sites for O-glycosylation were identified using searches at the NetOGlyc 3.1 server (www.cbs.dtu.dk/services/NetOGlyc/). Signal peptide prediction was conducted using the SignalP 3.0 server (www.cbs.dtu.dk/services/SignalP/). Multiple sequence alignments of proteins were carried out using the ClustalW software at the PBIL website (http://npsa-pbil.ibcp.fr/cgi-

bin/align_clustalw.pl) and Jalview software at the website (http://www.jalview.org/index.html). Secondary structure prediction for the conserved ChtBD2 region was done by Jpred (http://www.compbio.dundee.ac.uk/www-jpred/). Weblogos were constructed at (http://weblogo.berkeley.edu/logo.cgi.); (Schneider and Stephens, 1990).

2.3.5. Phylogenetic analysis of ChtBD2-containing proteins in *T. castaneum*

ClustalW software was used to perform multiple sequence alignments prior to phylogenetic analysis. MEGA 4.0 software (Tamura et al., 2007) was used to construct the consensus phylogenetic tree using the neighbor-joining method and a cut-off value of 50% similarity. To evaluate the branch strength of the phylogenetic tree, a bootstrap analysis of 5000 replications was performed. The accession numbers of the genes encoding *T. castaneum* ChtBD2 containing proteins are listed in Table 2-1.

2.4. Results

2.4.1. Bioinformatics search of *T. castaneum* genome databases

The initial search of the *T. castaneum* genome identified 29 proteins with ChtBD2s. The domains from these proteins were subsequently used in a second search to identify additional proteins with ChtBD2s. This process was repeated until no additional protein sequences with ChtBD2s could be identified. In the end, we identified a total of 49 putative *T. castaneum* genes capable of encoding 50 proteins (one gene codes for two proteins as a result of alternative splicing) with one or more ChtBD2s. Of these, 11 are PMPs, 18 are CPAPs, 13 include the chitinases and chitin deacetylases and 12 are classified as miscellaneous. The basis of these groupings and their descriptions are given in section 2.4.5. Interestingly, our searches yielded proteins with only the peritrophin-A type ChtBD2 that contains six cysteines, but not the

peritrophin B or C domains (with eight and six cysteines respectively). The results of our searches for ChtBD2s-containing proteins, including their predicted molecular weights and number of ChtBD2s, are shown in (Table 2-1).

2.4.2. Exon-intron organization of genes encoding proteins with ChtBD2s

To confirm the amino acid sequences of proteins with ChtBD2s inferred by GLEAN or NCBI gene predictions, I used RT-PCR to amplify the putative full-length coding sequence (CDS) of each cDNA. In many cases, I amplified a cDNA that precisely matched the predicted CDS, confirming the GLEAN/NCBI gene models. However, we found errors in exon-intron assignments in several of these genes. In some cases, the corrected annotations changed the number of ChtBD2s initially indicated by the GLEAN and/or NCBI predictions (Figure 2-2) and (Figure 2-3). The correct sequences of these cDNAs and the corresponding proteins have bseen submitted to NCBI and shown in (Table 2-1).

2.4.2.1. Misannotated gene encoding ChtBD2-containing proteins in the BeetleBase

In the case of the peritrophin, *TcPMP5-B*, comparison of the RT-PCR-derived cDNA with the Glean prediction (Glean_08506) revealed a rare error in genome sequence assembly, rather than the more common type of annotation error. The Glean annotation based on incorrect assembly predicts a protein with seven ChtBD2s, two more than the five that are actually present (Figure 2-2).

In the case of Glean_06098, I was unable to amplify the predicted 6.5 kb full-length cDNA by RT-PCR. Instead, I have confirmed that this gene actually represents an incorrect fusion of two tandem genes on the same strand, encoding related proteins, each with multiple ChtBD2s. This conclusion is based on the following evidence: 3'-RACE detected the predicted

polyadenylated products corresponding to the two genes, *TcPMP9* and *TcPMP14*, listed in (Table 2-1). I was unable to amplify a cDNA that bridged the two ORFs using several combinations of forward and reverse primers in the two genes. We found a putative TATA box between the polyadenylation site of the first gene and the apparent start codon of the second. RT-PCR using a forward primer just downstream of this TATA box and a reverse primer in the 3'-UTR of the proposed *TcPMP14* gene amplified a cDNA that encodes a signal peptide in the first exon that is very similar in sequence to the signal peptide of TcPMP9. The second exon, which follows a short intron of 50 bp, encodes 14 ChtBD2s. Thus, I conclude that there are two tandem genes encoding these two PMPs with an intergenic region of less than 300 bp, which was annotated as an intron in the Glean prediction (Figure 2-3 & Figure 2-12). The gene structure and functional motifs of all ChtBD2-encoding *T. castaneum* genes, omitting those that encode chitinases or chitin deacetylases and a few others that lack a signal peptide, are shown in

Figure 2-4). Of the 50 proteins identified as having 1 or more ChtBD2s, 45 contained a cleavable signal peptide at the N-terminus as analyzed by SignalP software. All except TcCHT7 (Zhu et al., 2008a) lacked trans-membrane domains, suggesting that they are secreted proteins. This inference is consistent with their predicted function in forming complexes with the extracellular polysaccharide chitin.

Table 2-1 Genes encoding proteins with ChtBD2 domain in *T. castaneum.*

	Gene/Splice Variant	GeneBank accession #	Glean #	cDNA length (bp)*	No.of Exons+	ORF (aa)	M.W (kDa)	No.of CBD	p/	LG	Map position (cM)	Citations
Peritrophic Matrix Proteins (PMPs)	*PMP1A*	GU128096	03179	501	4	166	18.7	1	5.3	3	0.0	
	PMP1B	GU128097	15620	671	2	207	22.6	1	5.1	6	15.3	
	PMP1C	GU128098	09231	2052	3	672	74.7	1	5.6	7	3.9	
	PMP2A	GU128099	03274	507	3	168	18.4	2	4.6	3	0.0	
	PMP2B	GU128100	03275	578	3	175	19.7	2	4.1	3	0.0	
	PMP2C	GU128101	09580	1096	2	941	104.4	2	5.3	7	3.9	
	PMP3	GU128102	09232	1652	3	538	58.0	3	3.6	7	3.9	
	PMP5A	GU128103	03273	1119	6	372	41.0	5	4.2	3	0.0	
	PMP5B	GU128104	08506	1560	6	519	79.2	5	4.0	4	55.5	
	PMP9	GU128105	06098	2121	1	706	70.0	9	6.1	10	3.2	
	PMP4	GU128106	06098	3920	2	1306	147.0	14	4.4	10	3.2	
Cuticular Proteins Analogous to Peritrophins (CPAPs)	*CPAP3-A1*	GU128092	11140	714	5	237	26.6	3	5.1	10	15.8	
	CPAP3-A2	GU128093	11141	711	5	236	26.4	3	5.1	10	15.8	
	CPAP3-B	EF125544	11139	948	5	279	32.0	3	5.3	10	15.8	
	CPAP3-C5a	EF125545	01169	1117	5	274	30.0	3	4.8	2	36.4	Jasrapuria et al. 2010
	CPAP3-C5b	EF125546	01169	950	5	237	27.6	3	4.7	2	36.4	
	CPAP3-D1	GU128094	11142	687	3	228	25.1	3	5.8	10	15.8	
	CPAP3-D2	EF125547	01350	807	4	255	28.8	3	5.6	2	49.0	
	CPAP3-E	GU128095	11349	744	3	247	26.9	3	4.5	10	15.8	
	CPAP1-A	GU128083	04733	987	5	328	37.7	1	8.4	1=X	30.1	
	CPAP1-B	GU128084	00587	594	4	197	22.6	1	5.1	2	32.9	
	CPAP1-C	GU128085	00316	1287	2	428	47.5	1	4.9	2	47.2	
	CPAP1-D	GU128086	09263	377	3	114	12.6	1	5.8	7	3.9	
	CPAP1-E	GU128087	09887	360	3	124	14.3	1	4.8	7	38.6	
	CPAP1-F	GU128088	09893	375	3	119	14.0	1	4.5	7	38.6	
	CPAP1-G	GU128089	08877	1086	3	305	35.9	1	6.3	7	38.6	
	CPAP1-H	GU128090	09894	2472	5	823	95.5	1	7.9	7	38.6	
	CPAP1-I	GU139459	12766	2448	4	748	87.3	1	9.2	9	59.0	
	CPAP1-J	GU128091	11101	3948	4	1315	146.4	1	7.4	10	25.3	
Chitin Metabolic Enzymes	*CHT4*	EF125543	09180	1519	7	475	42.1	1	4.6	7	7.0	Zhu et al. 2008a
	CHT8	DQ659248	09624	1564	6	496	54.1	1	4.9	7	7.0	Zhu et al. 2008a
	CHT5	AY675073	01770	1825	8	533	60.1	1	5.8	1=X	52.7	Zhu et al. 2008a
	CHT6	EFA00965	03876	7110	16	2369	267.7	1	6.3	3	67.0	Arakane and Muthkurishnan, 2009
	CHT7	DQ659247	15481	3435	7	980	110.9	1	7.0	6	2.5	Zhu et al. 2008a
	CHT10	DQ659250	12734	8254	15	2700	305.4	5	6.1	9	52.8	Zhu et al. 2008a
	CDA1	ABU2522	14100	1791	5	534	62.0	1	5.1	5	20.0	Dixit et al. 2009
	CDA2A	ABU25224	14101	1831	5	535	66.0	1	5.3	5	20.0	Dixit et al. 2009
	CDA2B	ABU25225	14101	1810	5	528	60.0	1	5.2	5	20.0	Dixit et al. 2009
	CDA3	ABW74145	05409	1580	5	517	59.0	1	6.1	8	50.3	Dixit et al. 2009
	CDA4	ABW74146	07635	1628	7	490	56.0	1	5.8	4	9.7	Dixit et al. 2009
	CDA5A	ABW74147	06846	3694	15	1231	129.0	1	8.6	U	U	Dixit et al. 2009
	CDA5B	ABW74148	06846	3694	15	1231	129.0	1	8.6	U	U	Dixit et al. 2009

* Longest available cDNA

+ Minumum no. of exons

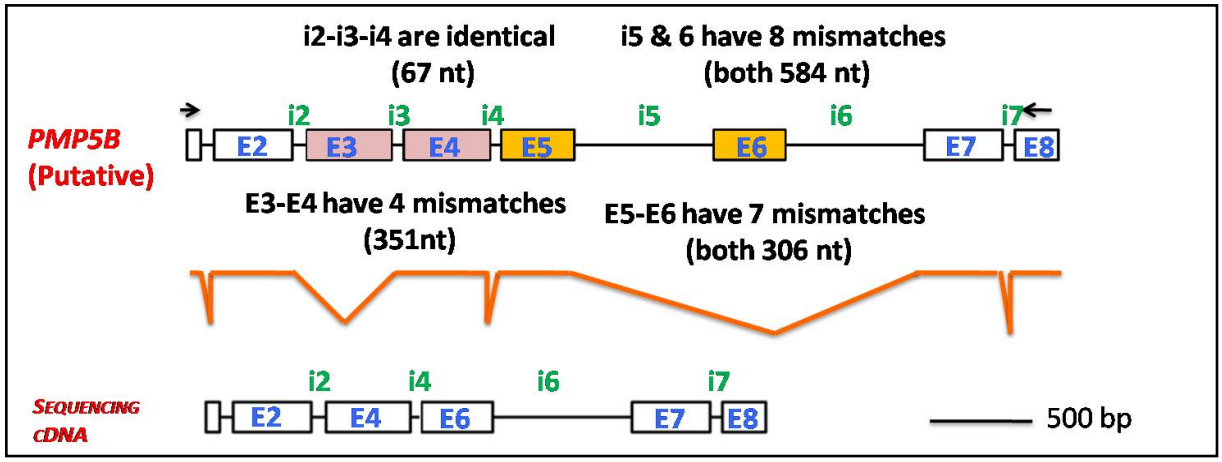


Figure 2-2 Schematic diagram showing mis-annotation of gene in *TcPMP5-B*

Top box shows the genome assembly of TcPMP5-B, the putative gene contains 8 exons shown as E [1-8] and 7 introns shown as i [1-7]. Exon 2 and inton 7 are unique. However the sequences of 417 nucleotides are identical in i2 and i3 with a single mismatch. Further, a stretch of 519 nucleotides between E2-E3 and E3-E4 is also identical with a single nucleotide mismatch. i2 and i4 are identical and 51 nucleotide long, i5 and i6 have 8 mismatches and are 584 nucleotides long. The cloned cDNA has 5 ChtBD2 and 6 exons.

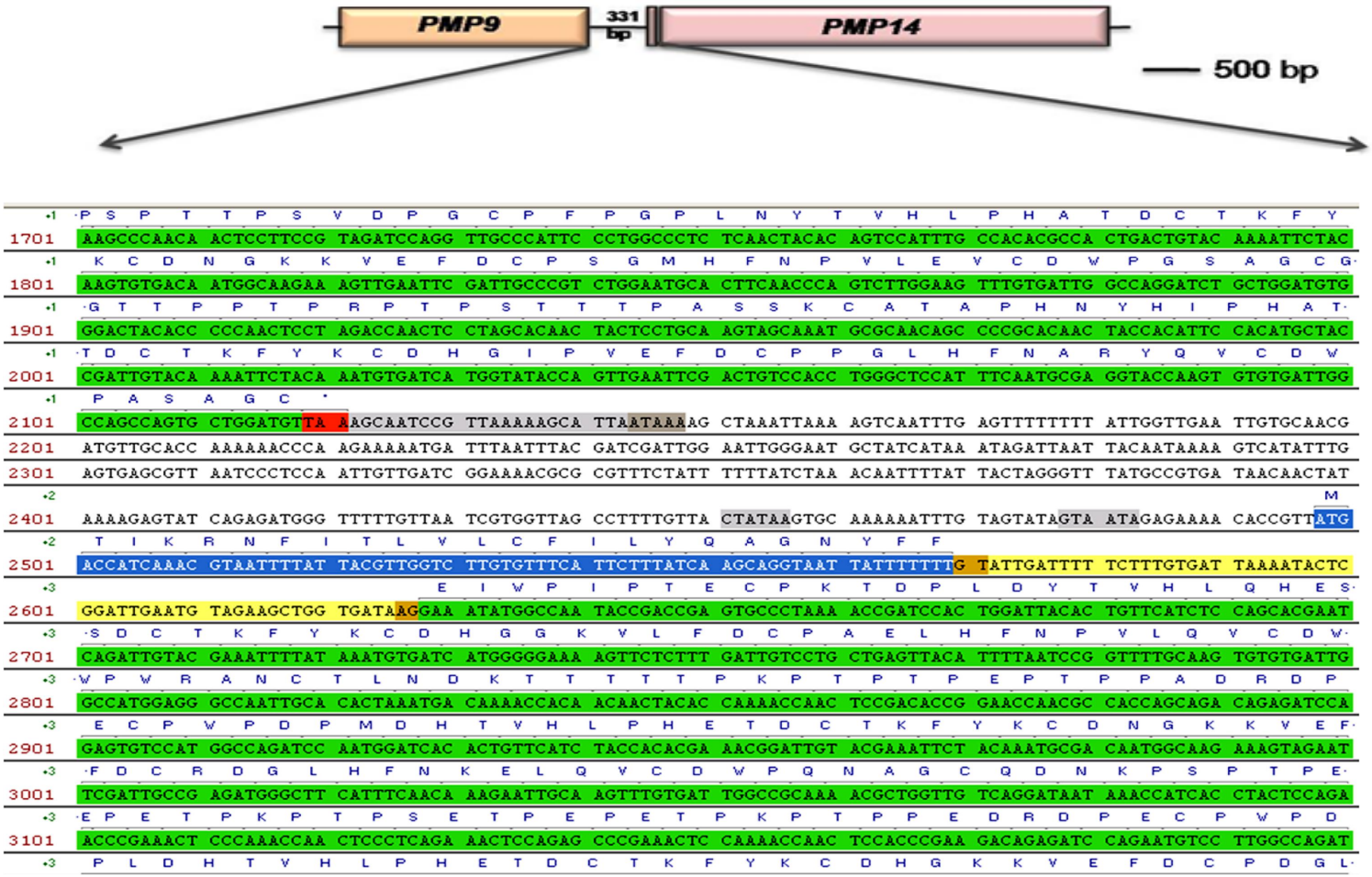


Figure 2-3 Gene organization and deduced amino acid sequence showing the intergenic region between *TcPMP9* and *TcPMP14*

TcPMP9 and TcPMP14 are separated by a 331 base- pair intergenic region. TcPMP9 lacks introns and TcPMP14 has a 52 base-pair intron (shaded in yellow). The nucleotide sequence of the intergenic region is also shown. The red shaded codon is the stop codon for PMP 9 and the blue shaded region is the ORF of the first exon of TcPMP14 starting at ATG. The exon splice junction for TcPMP14 is marked in orange starting at GT and ending at AG

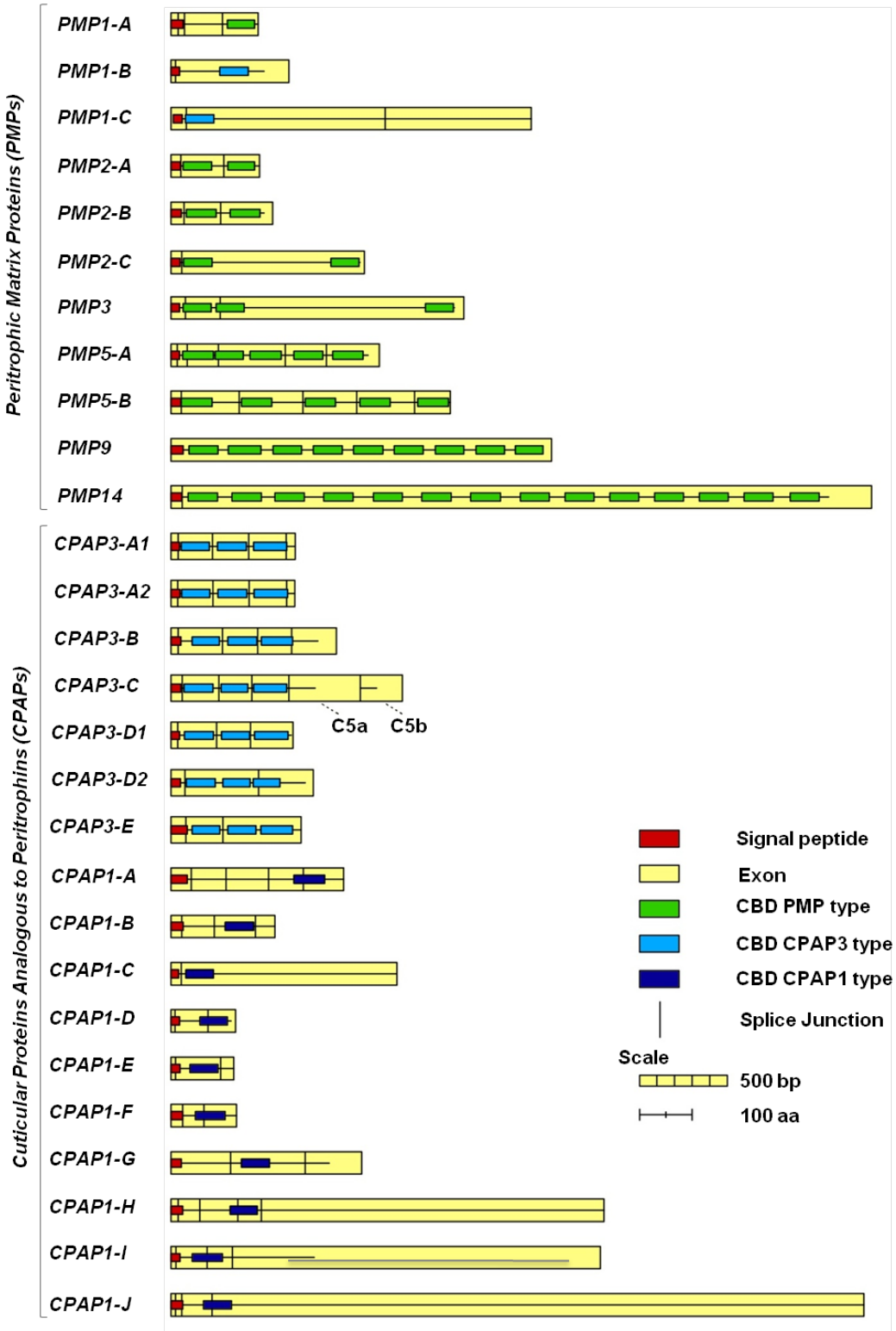

Peritrophic Matrix Proteins (PMPs)
PMP1-A
PMP1-B
PMP1-C
PMP2-A
PMP2-B
PMP2-C
PMP3
PMP5-A
PMP5-B
PMP9
PMP14
Cuticular Proteins Analogous to Peritrophins (CPAPs)
CPAP3-A1
CPAP3-A2
CPAP3-B
CPAP3-C
C5a
C5b
CPAP3-D1
CPAP3-D2
CPAP3-E
CPAP1-A
CPAP1-B
CPAP1-C
CPAP1-D
CPAP1-E
CPAP1-F
CPAP1-G
CPAP1-H
CPAP1-I
CPAP1-J
Signal peptide
Exon
CBD PMP type
CBD CPAP3 type
CBD CPAP1 type
Splice Junction
Scale
500 bp
100 aa

Figure 2-4 Schematic diagram of the exon–intron-domain organizations of *T. castaneum* genes encoding ChtBD2-containing proteins.

Yellow boxes represent the longest cDNA for the particular gene that has been cloned. Red solid box indicates signal peptide and green, light blue and dark blue boxes indicate the ChtBD2s typical of PMP, CPAP3 and CPAP1 families of proteins. The lines connecting boxes represent linker or other protein sequences. Each vertical line represents the splice junction with the introns removed. The two alternative forms of exon 5 in CPAP3-C are indicated as the two yellow boxes labeled C5a and C5b.

2.4.3. Clustering of *PMP* family and *CPAP* family of genes in the *T. castaneum* genome

As mentioned earlier in section 2.4.2, the two genes, *PMP9* and *PMP14* are arranged in tandem in linkage group 10 in the *T. castaneum* genome and their ORFs are separated by only 331 base pairs (Figure 2-3; Figure 2-12). There are two other clusters of *PMP* genes in the *T. castaneum* genome. *TcPMP1-B* and *TcPMP3* are located on linkage group 7 within a stretch of only 977 bp. Furthermore, *TcPMP2-A* and *TcPMP5-A* are in linkage group 3 within a stretch of 1,575 bp (Figure 2-5). These genes may be a result of relatively recent gene duplications. Four of the *CPAP* genes, *TcCPAP3-A1*, *TcCPAP3-A2*, *TcCPAP3-B* and *TcCPAP3-D1* are tightly clustered within a 26 kbp region of linkage group 10 at map position 15.8 cM, while a fifth gene, *TcCPAP3-E*, is ~190 kb away at the same map position. The other two genes *TcCPAP3-C* and *TcCPAP3-D2* are in linkage group 2 at map positions 36.4 cM and 49 cM, indicating that they are widely separated on the same chromosome.

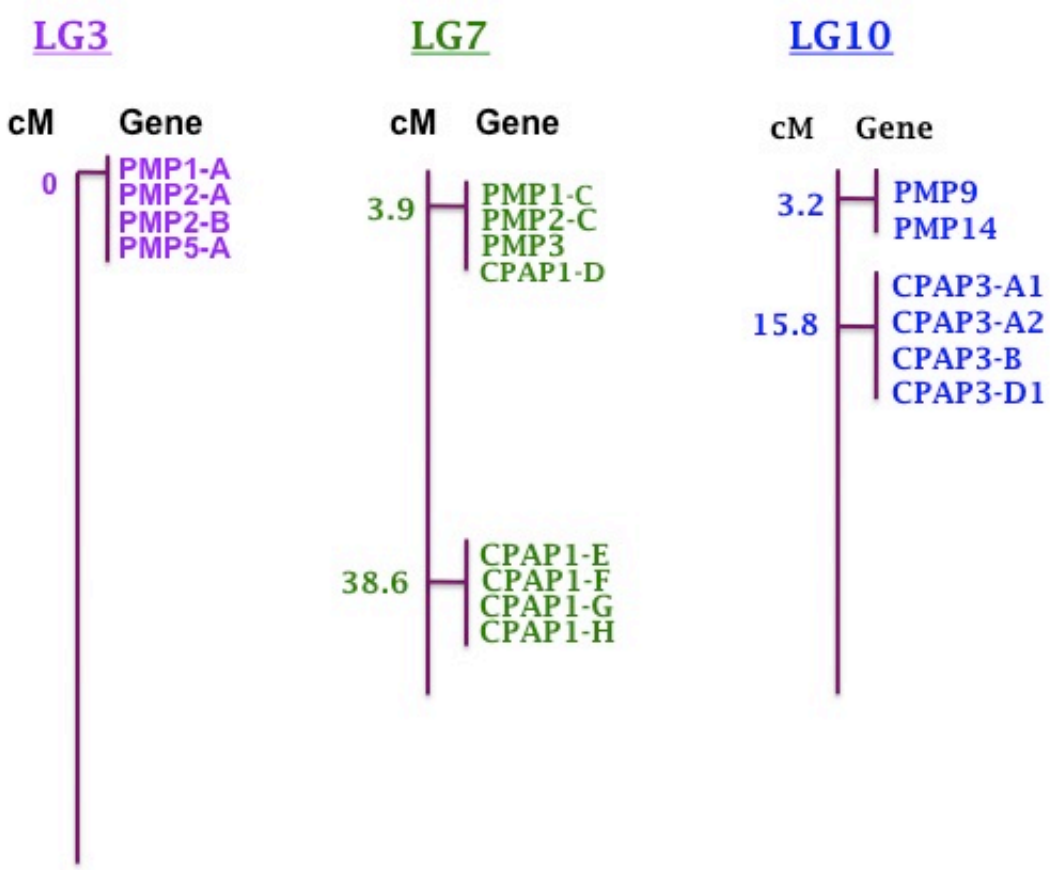


Figure 2-5 Schematic diagram showing linkage in *PMP* family and *CPAP* family of genes.

Each chromosome is represented as vertical line, namely LG3, LG7 and LG10
LG= Linkage Group, cM= centimorgan. The numbers represents the distance in cM. Genes PMP1-A, PMP2-A, PMP2-B and PMP5-A are clustered on LG 3 at 0 cM. On LG7, PMP1-C, PMP2-C, PMP3, CPAP1-D are clustered at 3.9 cM and CPAP1-e, CPAP1-F, CPAP1-G and CPAP1-H are clustered at 38.6 cM. PMP9 and PMP14 are arranged in tandem in linkage group 10 at 3.2 cM and CPAP3-A1, CPAP3-A2, CPAP3-B and CPAP3-D1 are clustered at 15.8 cM on LG 10.

2.4.4. Presence of CPAP family of proteins in different groups of arthropods

The CPAP3 family of proteins is highly conserved through different insect orders, including Diptera, Coleoptera, Hymenoptera, Hemiptera, Phithiraptera and Lepidoptera. It is also found in the subphyla Crustacea and Chelicerata. Like *T. castaneum*, other insect species have additional paralogs of the *TcCPAP3-A* genes. *A. gambiae* has three paralogs of this gene, and *N. vitripennis* and *A. pisum* each have two (data not shown). Similarly, two paralogs of *TcCPAP3-D* genes are found in several species including *A. aegypti, A. gambiae, C. quinquefasciatus, N. vitripennis* and *A. pisum* (data not shown). Regarding the CPAP1 subgroup, there are orthologs in *D. melanogaster* for TcCPAP1-B, TcCPAP1-C, TcCPAP1-D, TcCPAP1-G and TcCPAP1-H. The latter contains a mucin domain and is very rich in serine residues. However, there is nothing known about the nature or functions of these proteins. None of the members of the family of CPAP1 proteins of *T. castaneum* have sequence similarities with one another other than in the ChtBD2 region (Table 2-2).

Table 2-2 Identification of orthologs of TcCPAP proteins in arthropods

		Phylum : Arthropoda								
		Sub-phylum : Hexapoda								
		Class : Insecta								
CPAPs	Accession No.	Diptera	Coleoptera	Hymenoptera	Hemiptera	Phthiraptera	Lepidoptera	Crustacea	Chelicerata	Nematodes
CPAP3-A1	ACY95475	+	+	+	+	+	id	+	-	-
CPAP3-A2	ACY95476	+	+	+	+	+	id	+	-	-
CPAP3-B	ABL73928	+	+	+	+	id	+	+	-	-
CPAP3-C	ABL73929	+	+	+	+	+	+	+	-	-
CPAP3-D1	ACY95477	+	+	+	+	+	+	+	+	-
CPAP3-D2	ABL73931	+	+	+	+	+	id	+	+	-
CPAP3-E	ACY95478	+	+	+	+	+	id	+	-	-
CPAP1-A	ACY95466	-	+	+	+	-	-	-	-	-
CPAP1-B	ACY95467	+	+	+	-	-	id	-	-	-
CPAP1-C	ACY95468	+	+	-	-	-	-	-	-	-
CPAP1-D	ACY95469	+	+	+	-	id	-	-	-	-
CPAP1-E	ACY95470	-	+	-	-	id	-	+	-	-
CPAP1-F	ACY95471	-	+	+	-	-	-	-	-	-
CPAP1-G	ACY95472	+	+	+	-	id	id	-	-	-
CPAP1-H	ACY95473	+	+	-	-	id	-	-	-	-
CPAP1-I	ACZ04319	-	+	-	-	id	id	-	-	-
CPAP1-J	ACY95474	-	+	-	-	id	id	-	-	-

id: insufficient data; "+" if overall similarity had e-values greater than (> e^{-22}) ; "-" if overall similarity had e-values less than (<e^{-22}). Representative members for different insect orders, sub-phylum and phylum: Diptera- Drosophila, Culex, Anopheles, Aedes, Glossina, Saptomyza, Lutzomiya, Culicoides; Lepidoptera- Choristeneura, Bombyx, Spodoptera, Loxostege, Tenebrio, Mamestra, Helicoverpa, Plutella; Coleoptera- Tribolium, Holotrichia, Popillia, Tenebrio; Hymenoptera-Harpegnathus, Nasonia, Solenopsis, Apis, Camponotus; Hemiptera-Pisum; Phithiraptera- Pediculus; Crustacea-Daphnia, Artemia, Lepeophtheirus, Caligus, Rimicaris; Chelicerata- Ixodes; Nematoda- Ascaris, Caenorhabditis.

2.4.5. Classification of proteins containing one or more ChtBD2s into families

I attempted to align the amino acid sequences of ChtBD2s from these 50 *T. castaneum* proteins with one another and with the ChtBD2s of well-characterized insect peritrophins. During this process, it became clear that one subgroup of 11 *T. castaneum* proteins is highly related to *L. cuprina* and *T. ni* peritrophins with ChtBD2s closely matching Tellam's consensus sequence for PM-associated peritrophin-A domains (Tellam et al., 1999); (Wang et al., 2004). These proteins were therefore tentatively classified as Peritrophic Membrane Proteins (PMPs) or "peritrophins". It should be emphasized that, at present, we lack definitive proof that these proteins are associated with the peritrophic matrix. The number of ChtBD2s in this group of proteins ranges from one to 14, and the lengths of the individual domains range from 52 to 56 amino acids when counting from the first-to-sixth cysteine. The spacings between adjacent cysteines are in general accordance with the Tellam consensus for peritrophin-A domains (Tellam et al., 1999). In proteins with multiple ChtBD2s, the sequences of the individual ChtBD2s were found to be highly similar. This is especially true of TcPMP5-B, TcPMP9 and TcPMP14 with five, nine and 14 ChtBD2s repeats, respectively.

Of the remaining 39 proteins with ChtBD2s, 18 form a large, group (consisting of two families), which I have named "Cuticular Proteins Analogous to Peritrophins" (CPAPs) because of their apparent lack of a function in PM organization and presumed cuticular localization (explained in detail in CHAPTER 3). Within this large group, there are eight proteins containing three CBDs each, which are 54 to 56 amino acids long. I have named this subgroup as CPAP3 family to indicate that each family member has three ChtBD2s. All of these proteins also have a remarkably similar domain organization (Figure 2-4). Two proteins,

TcCPAP-C5a and TcCPAP-C5b, arise from alternative splicing of the fifth (i.e. the last) exon from a pre-mRNA, which results in proteins with different C-terminal sequences but with a common, ChtBD2-containing N-terminal portion. These proteins are homologs of members of the family of *D. melanogaster* proteins collectively known as "gasps" or "obstructors" all of which have three ChtBD2s (Barry et al., 1999; Behr and Hoch, 2005). I have named the individual *T. castaneum* orthologs of this family to reflect their orthology to the corresponding *D. melanogaster* proteins as well as the number of encoded ChtBD2s (e.g. *T. castaneum* CPAP3-B is the ortholog of *D. melanogaster* Obst-B and has three ChtBD2s). There are five fewer CPAP3 orthologs in *T. castaneum* than in *D. melanogaster* (Behr and Hoch, 2005). Each member of the CPAP3 family is conserved in different insect species, and has signature sequence by which one can distinguish a CPAP3-A1 from a CPAP3-C and so on. Multiple sequence alignments of each of the three ChtBD2 of different species of insects and crustaceans are shown in Figure 2-6; Figure 2-7; Figure 2-8.

The remaining 10 *T. castaneum* proteins in the CPAP group contain a single ChtBD2 at either the N-terminus or the C-terminus, and are therefore given the family name CPAP1. Notably, they show much greater conservation of the sequences not only between cysteines but also in the flanking sequences on either side of the ChtBD2 compared to the other CPAP group or the PMP group (Figure 2-9). In addition to the 11 PMPs and 18 CPAPs listed in Table 2-1, we previously characterized six chitinases and seven chitin deacetylases, all of which contain one or more ChtBD2s (Dixit et al., 2008; Zhu et al., 2008a). Apart from these, I have also identified ten proteins with ChtBD2s represented by Glean_02107, 02058, 09553, 11724, 00588, 15245, 05468, 05469 and 13568. The Glean_13568 lacks a predicted signal peptide. Another Glean_12551 represents a very large protein with a predicted molecular weight of of 437.7 kDa,

which has a signal peptide, five von Willebrandt factor type D domains and five interspersed ChtBD2s. This protein is an ortholog of a *D. melanogaster* hemolectin, a hemolymph protein involved in blood clotting and immunity (Goto et al., 2001). Glean_09553, which has a signal peptide, lacks the consensus spacing between the cysteines in two out of the three ChtBD2s. This gene is expressed only in the adult stage but is not expressed in elytra (modified forewings), hindwings or gut (discussed in section 2.4.10.2). These seven proteins most likely have roles other than interactions with chitin in the gut, tracheae or cuticle. Therefore, I have not included these proteins in our analyses. Glean_15245, which has a signal peptide and encodes a protein with multiple ChtBD2s, is in a region where there is a 4 kb gap in the genome assembly and is not analyzed in this chapter because of incomplete annotation. This gene and the encoded protein will be discussed in section 2.4.10.1. These genes have been included in the miscellaneous category, as based on available data, none of them is suitable for inclusion in the aforementioned families of PMP and CPAPs (Table B-3).

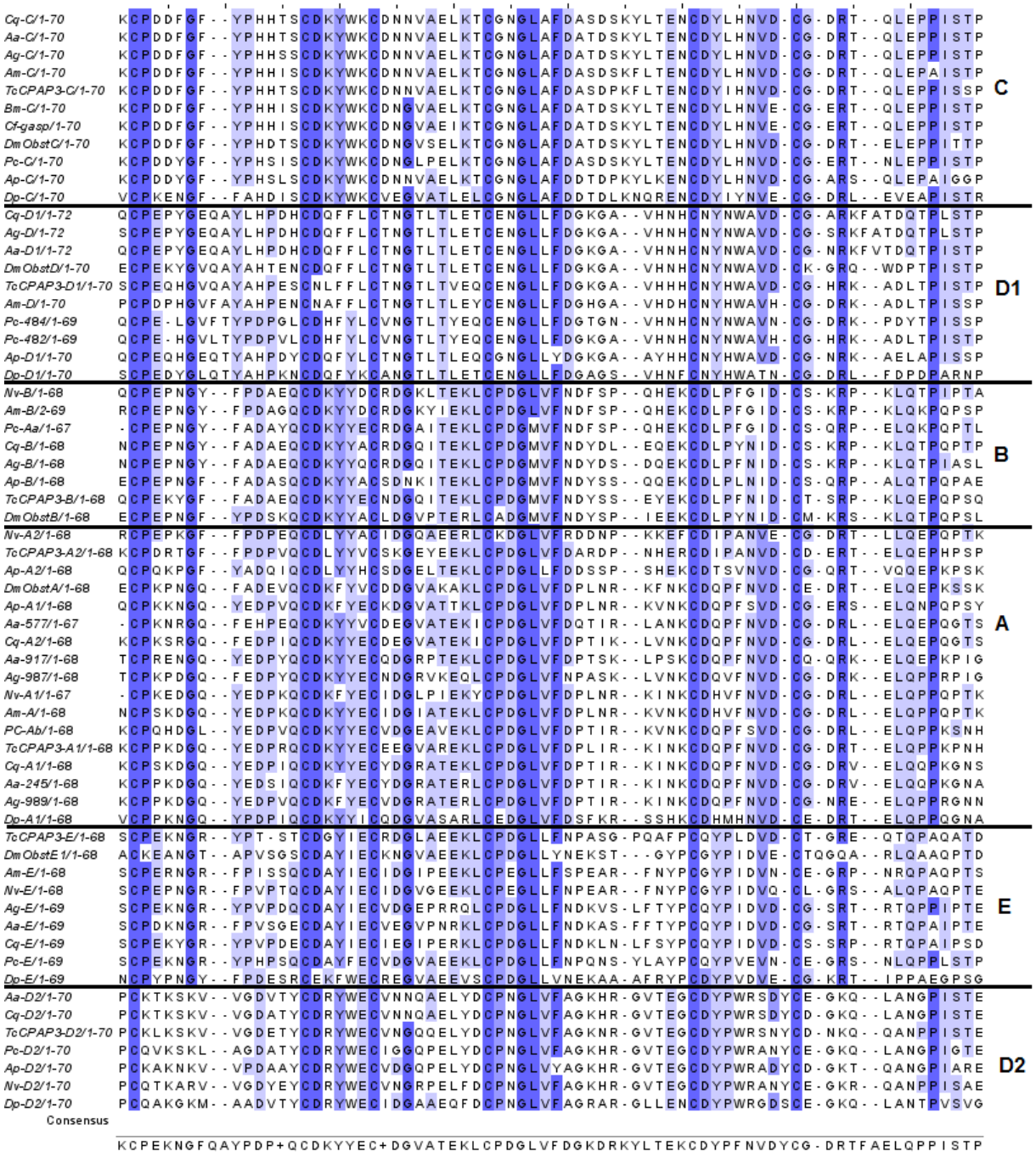


Figure 2-6 Multiple sequence alignment of 1st ChtBD2 of CPAP3 family.

Amino acid sequence alignment of the ChtBD2 (1st ChtBD2 domains of the CPAP3 family of proteins from Drosophila melanogaster (Dm), Anopheles gambiae (Ag), Acyrthosiphon pisum (Ap), Aedes aegypti (Aa), Apis mellifera (Am), Culex. quinquefasciatus (Cq), Nasonia vitripennis (Nv), Pediculus humanus corporis(Pc), Tribolium castaneum (Tc), Bombyx mori (Bm), Choristoneura fumiferana (Cf) and the chelicerate, Daphnia pulex (Dp). Each member of the

CPAP3 family is represented by a black box. Shaded in dark blue are the highly conserved residues. The less conserved amino acids are shown in various shades of blue

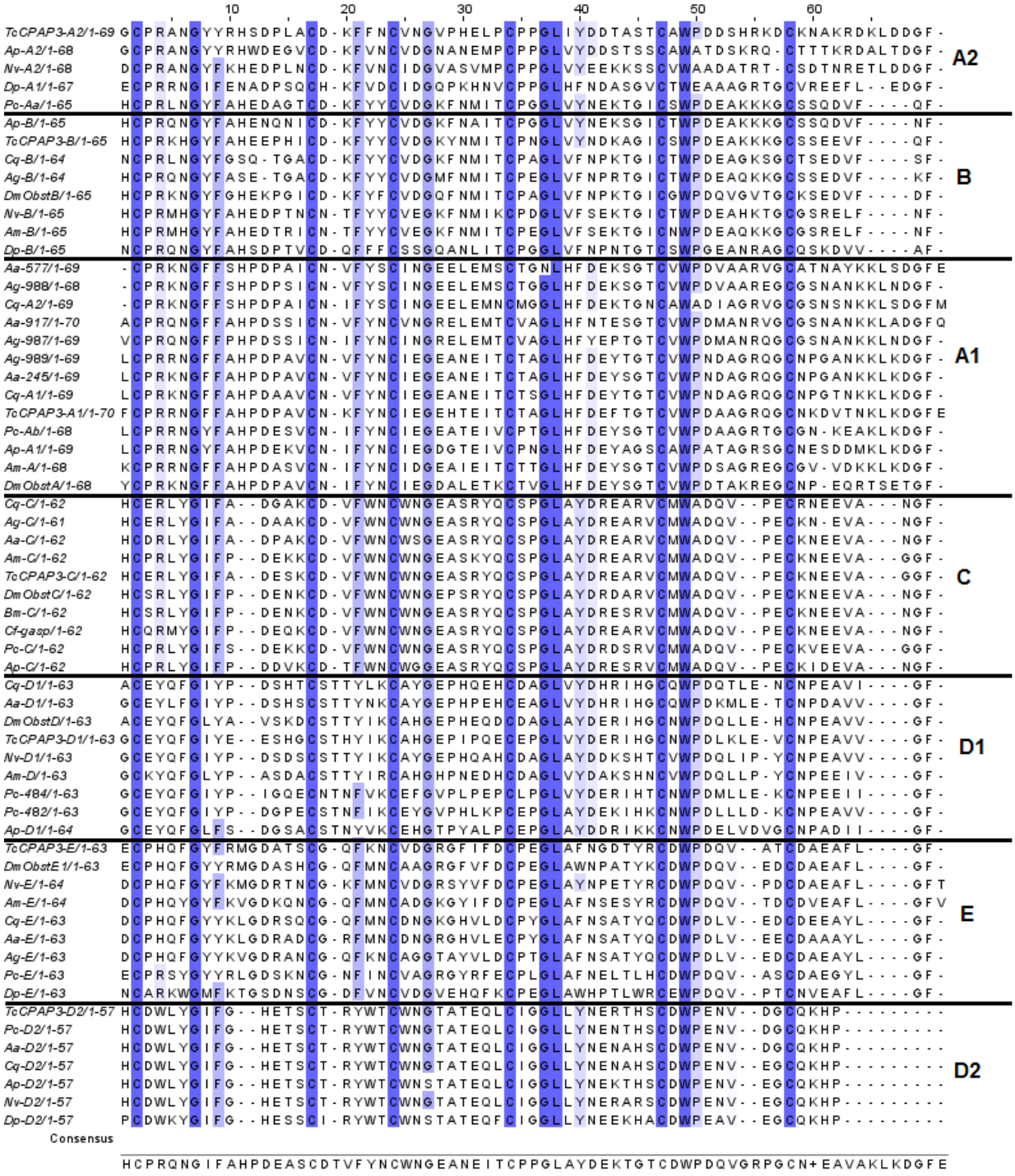


Figure 2-7 Multiple sequence alignment of 2nd ChtBD2 of CPAP3 family.

Amino acid sequence alignment of the ChtBD2 (2nd ChtBD2) domains of the CPAP3 family of proteins from Drosophila melanogaster (Dm), Anopheles gambiae (Ag), Acyrthosiphon pisum

(Ap), Aedes aegypti (Aa), Apis mellifera (Am), Culex quinquefasciatus (Cq), Nasonia vitripennis (Nv), Pediculus humanus corporis(Pc), Tribolium castaneum (Tc), Bombyx mori (Bm), Choristoneura fumiferana (Cf) and the crustacea, Daphnia pulex (Dp). Each member of the CPAP3 family is separated by a black line. Shaded in dark blue are the highly conserved residues including cysteines. The less conserved amino acids are shown in various shades of blue.

.

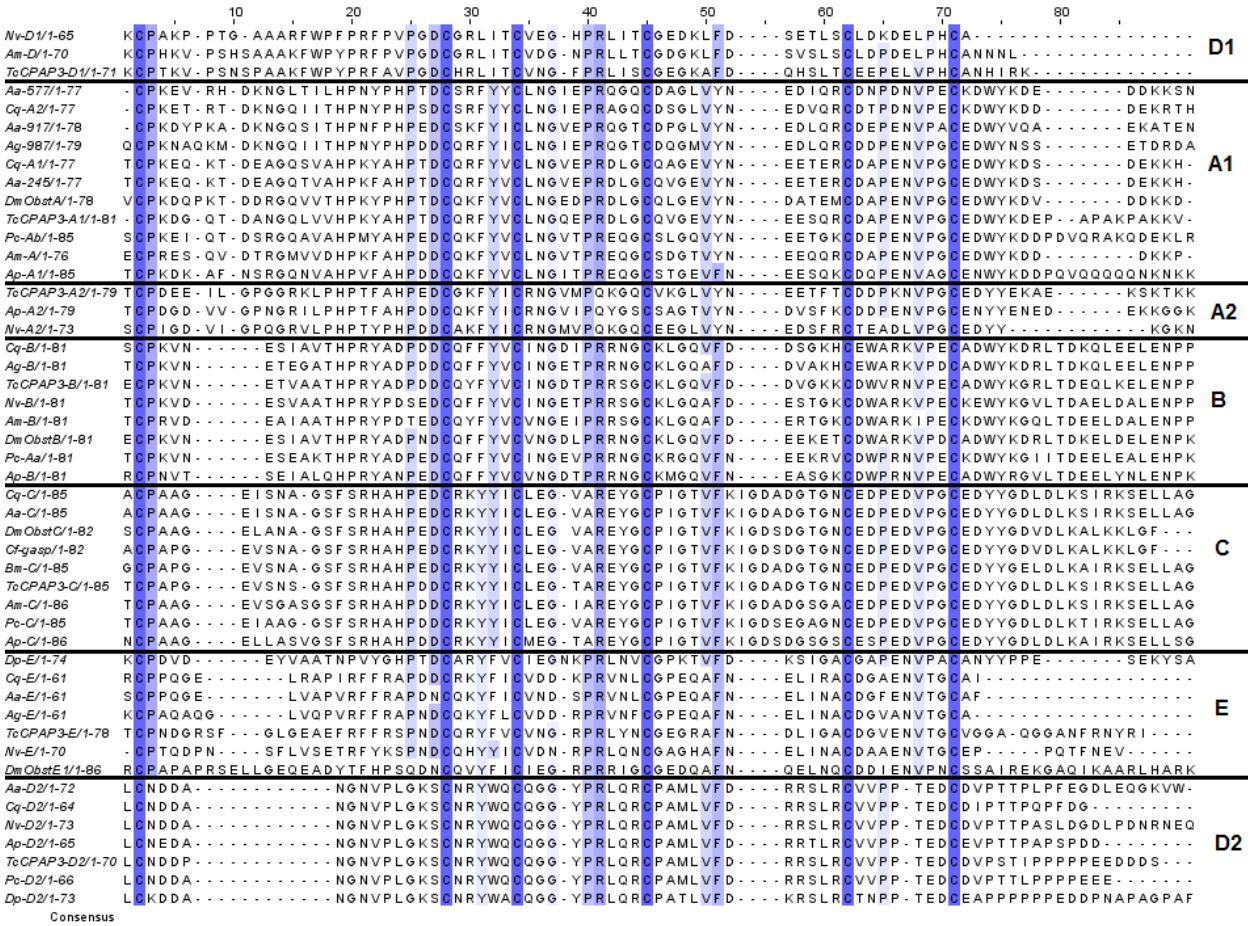


Figure 2-8 Multiple sequence alignment of 3rd ChtBD2 of CPAP3 family.

Amino acid sequence alignment of the ChtBD2 (3rd ChtBD2) domains of the CPAP3 family of proteins from Drosophila melanogaster (Dm), Anopheles gambiae (Ag), Acyrthosiphon pisum (Ap), Aedes aegypti (Aa), Apis mellifera (Am), Culex quinquefasciatus (Cq), Nasonia vitripennis (Nv), Pediculus humanus corporis(Pc), Tribolium castaneum (Tc), Bombyx mori (Bm), Choristoneura fumiferana (Cf) and crustcea Daphnia pulex (Dp). Each member of the CPAP3 family is represented by a black box. Shaded in dark blue are the cysteines. The less conserved amino acids are shown in various shades of blue

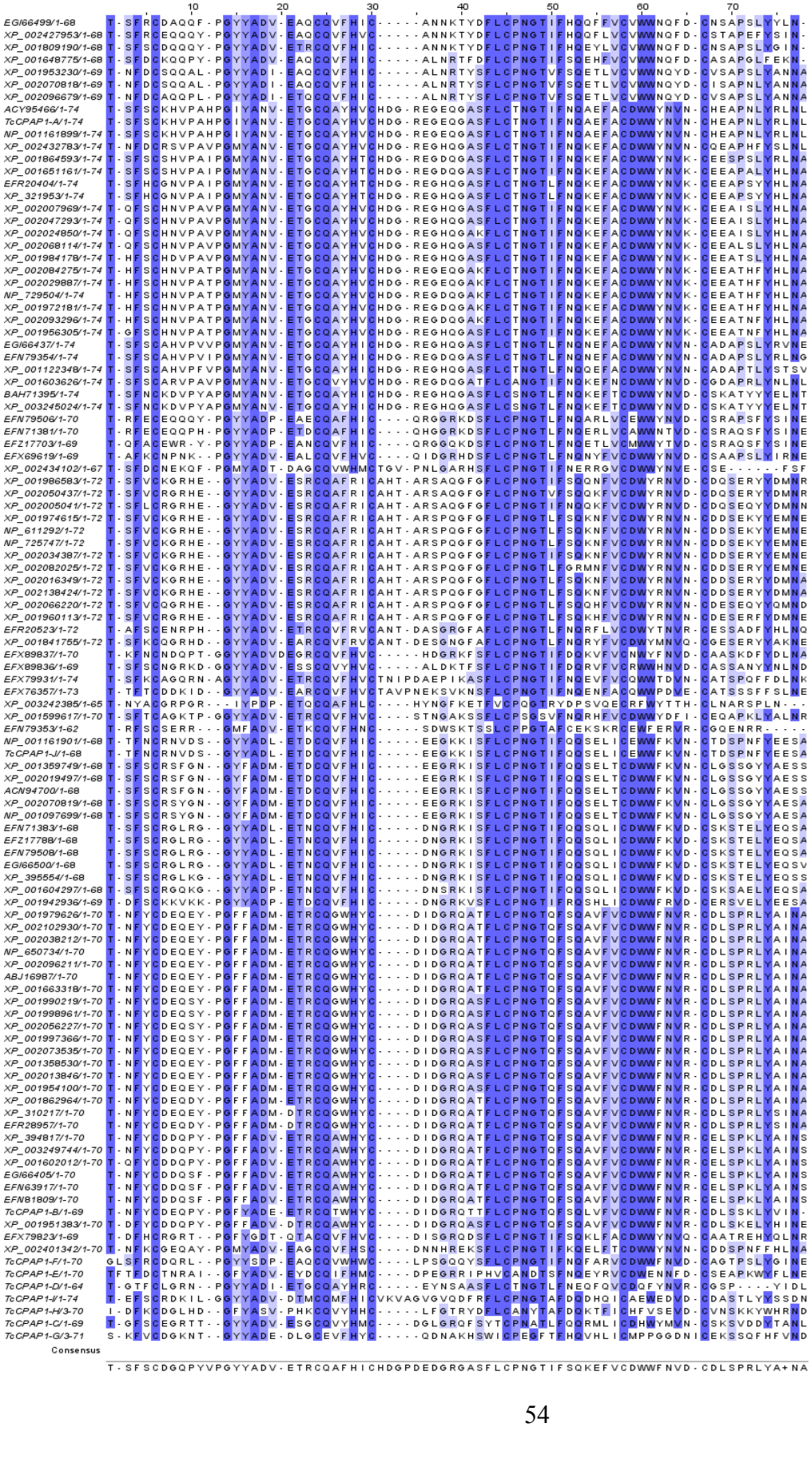


Figure 2-9 Multiple sequence alignment of ChtBD2 of CPAP1 family.

Amino acid sequence alignment of the ChtBD2 of the CPAP1 family of proteins from different insect species. Shaded in dark blue are the highly conserved residues including cysteines. The less conserved amino acids are shown in various shades of blue. The accession numbers for each of these proteins is described in Table B-2.

2.4.5.1. Conserved consensus for ChtBD2

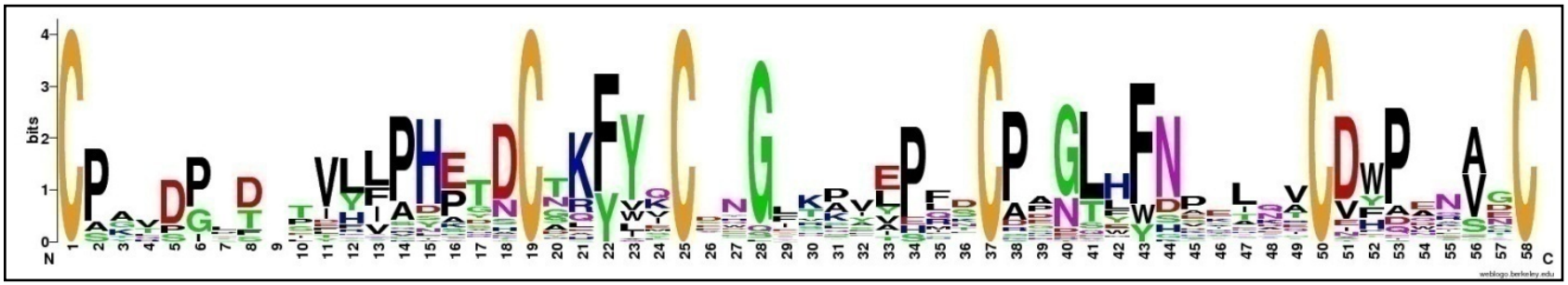
According to (Tellam et al., 1999) the consensus for six- cysteine-containing "peritrophin-A domains" is $CX_{13-20}CX_{5-6}CX_{9-19}CX_{10-14}C\text{-}X_{4-14}C$ where X is any amino acid except cysteine. This consensus was derived from the limited sequence data available at that time. Since then several insect genomes have been fully sequenced and additional sequence information on insect proteins with ChtBD2 domains has become available. I have aligned multiple amino acid sequences of the ChtBD2 domains of members belonging to each of the PMP, CPAP3 and CPAP1 families from a large number of insects belonging to different insect orders and some arthropods,. During this process, it was discovered that in these proteins there are conserved amino acid residues even in the regions flanking the 6-cysteine-containing ChtBD2 domains. Furthermore, there are additional conserved amino acids even between the cysteines of the ChtBD2 domains, which are characteristic of each family and even for subfamilies of the CPAP3 proteins. These results are presented as weblogos in Figures 2-2 through 2-9. As indicated by the weblogo analysis of PMPs from large number of species, there is less conservation of amino acid sequences within the ChtBD2 domains of the PMPs compared to those from subfamiliesCPAP3 or CPAP1 families. This is particularly true of the sequence between the first and second conserved cysteine in the ChtBD2. In addition the distance between the two cysteines is also highly variable in he PMP family. The exception is the lack of variation

in the spacing between the second and third cysteines in the second ChtBD2 domain in CPAP3 family.

In the eight subfamilies of CPAP3 (CPAP3-A through CPAP3-E) proteins, the third ChtBD2 has the greatest variation in sequence and in spacing between cysteines. The spacings between the first and second cysteines as well as the spacing between the fourth and fifth cysteines are greater than those in PMPs, but the spacing between the third and fourth cysteines is shorter (Figure 2-10), This family shows no variation in the number of amino acids between the second and third cysteines as well as between the fourth and fifth cysteines (Figure 2-10). The most noticeable fact is that not only is the ChtBD2 domain conserved but the flanking sequences are also conserved.

Among the three families, in the CPAP1 family the sequences are highly conserved.

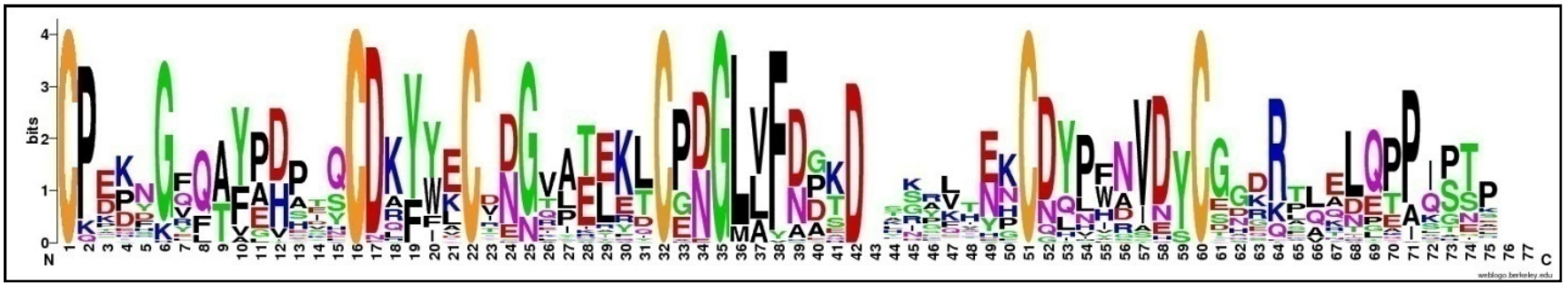
PMP (all ChtBD2)



Consensus of conserved cysteines and spacings: Cpx$_2$dpxdx$_4$Vx$_2$phx$_2$dCxkFx$_2$Cx$_2$Gx$_6$px$_3$CpxglxFx$_5$Cdwpx$_4$C

x = any amino acid

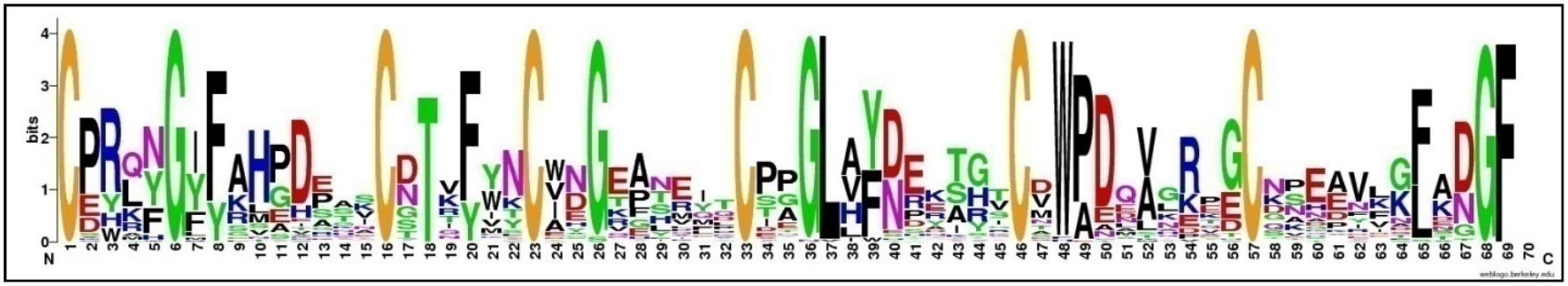
CPAP3 (1st ChtBD2)



Consensus of conserved cysteines and spacings:

:CPEKNGFQAYPDP+QCDKYYEC+DGVATEKLCPDGLVFDGKDRKYLTEKCDYPFNVDYCG-DRTFAELQPPISTP

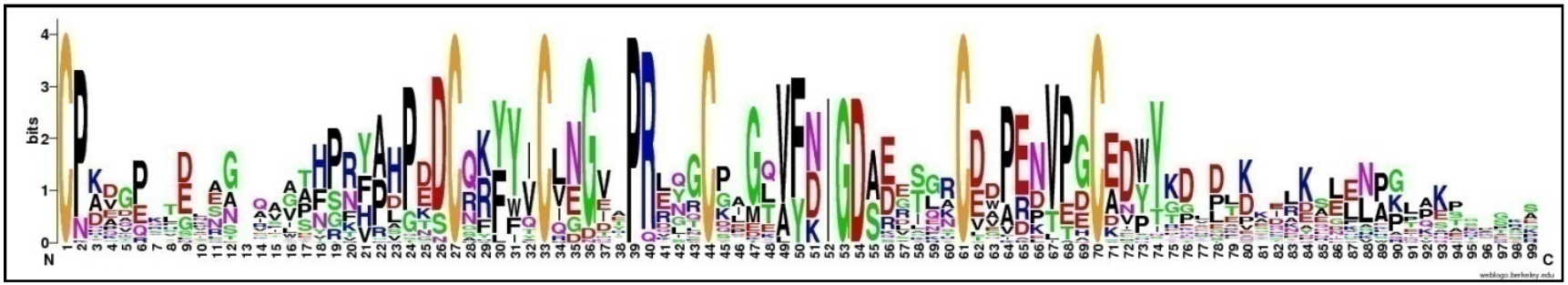
CPAP3 (2nd ChtBD2)



Consensus of conserved cysteines and spacings:

```
CPRQNGIFAHPDEASCDTVFYNCWNGEANEITCPPGLAYDEKTGTCDWPDQVGRPGCN+EAVAKLKDGFE
```

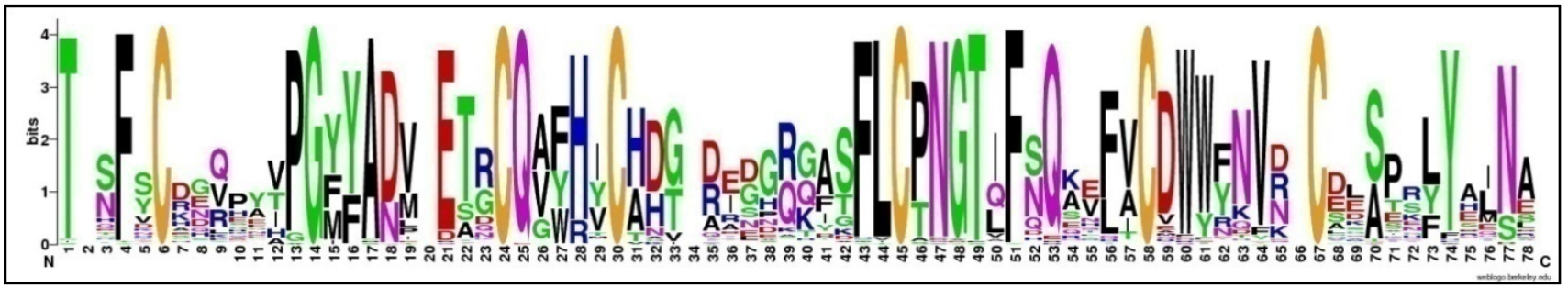
CPAP3 (3rd ChtBD2)



Consensus of conserved cysteines and spacings:

```
CPKDGPKT-D++GQA+THPRYAHPDDCQKYYICLNGVTPRLQGCPAGQVFNIGDAEE+GRCDDPENVPGCEDWYKDL+LDA+LKSEKNP+
```

CPAP1 (only ChtBD2)



Consensus of conserved cysteines and spacings:

T - SF SCDGQPYVPGYYADV - ETRCQAFHICHDGPDEDGRGASFLCPNGTIFSQKEFVCDWWFNVD - CDLSPRLYA+NA

Figure 2-10 Conserved amino acid residues of ChtBD2s of each of the PMP, CPAP3 and CPAP1 families of proteins across arthropods shown as weblogo.

The multiple sequence alignment of ChtBD2s for each of these ChtBD2-containing proteins was carried out using the ClustalW software from PBIL Expasy tool as shown in (Figure 2-6, Figure 2-7,Figure 2-8,Figure 2-9). The X-axis shows the position of occurrence of a particular residue and the Y-axis measures the log frequency of occurrence of the particular residue in bits. Shown in yellow are the cysteines. The less-conserved amino acids have smaller "bit values". At the bottom, the conserved spacing between the six cysteines of the ChtBD2s is indicated as well as the extended amino acid sequence consensus for this group of ChtBD2-s.

2.4.6. Phylogenetic analysis of ChtBD2-containing proteins in *T. castaneum*

To establish the evolutionary relationship among the *T. castaneum* proteins with ChtBD2s and among the domains within the same protein, we carried out a phylogenetic analysis of all ChtBD2s in proteins from *T. castaneum* using the ClustalW program. All ChtBD2s present in nine out of the 11 PMP family proteins are clustered in one large branch in which different PMPs form additional branches, indicating that they are closely related (see the branch shaded in green in (Figure 2-11). The only exceptions are the two ChtBD2s associated with TcPMP1-B and TcPMP1-C, which align with the members of the CPAP3 family. Most of the ChtBD2s within the same protein (e.g. all ChtBD2s of TcPMP9 and ChtBD2s 2 through 14 of TcPMP14) are clustered in the same branch, indicating minimal sequence divergence among the ChtBD2s in the same protein. Occasionally, we can find evidence for sequence divergence of one or more ChtBD2s away from the other domains in the same protein (e.g. the third ChtBD2 of TcPMP5).

The ChtBD2s of the CPAP3 and CPAP1 family proteins form two sub-branches of another distinct branch away from the PMP branch of the evolutionary tree (shaded dark and light blue in (Figure 2-11). One sub-branch contains all *T. castaneum* proteins with a single ChtBD2 (TcCPAP1-A through TcCPAP1- J; shaded dark blue), indicating a common ancestral origin for the ChtBD2s of these proteins of diverse sizes, which share no other sequence similarity. All of the *T. castaneum* proteins with three ChtBD2s (TcCPAP3 family, shaded light blue), which are similar to the *D. melanogaster* obstructor/gasp proteins, form the other sub-branch. Interestingly, the first ChtBD2s of almost all members of the TcCPAP3 family with the exception of TcCPAP-3 C and TcCPAP3-D1 are more closely related to one another than to the

second and third ChtBD2 within the same protein. Similarly, the third ChtBD2 of most of the TcCPAP3 family proteins are also clustered, with TcCPAP3-D2 being the lone exception. ChtBD2s 1 and 2 of different TcCPAP3 proteins are more closely related to one another than to the third ChtBD2 in the same protein, which form a separate branch. The phylogenetic analysis (Figure 2-11) further supports our classification of the ChtBD2-containing proteins of *T. castaneum* into three groups based on the variations in the Tellam consensus sequence for the conserved cysteines of ChtBD2s.

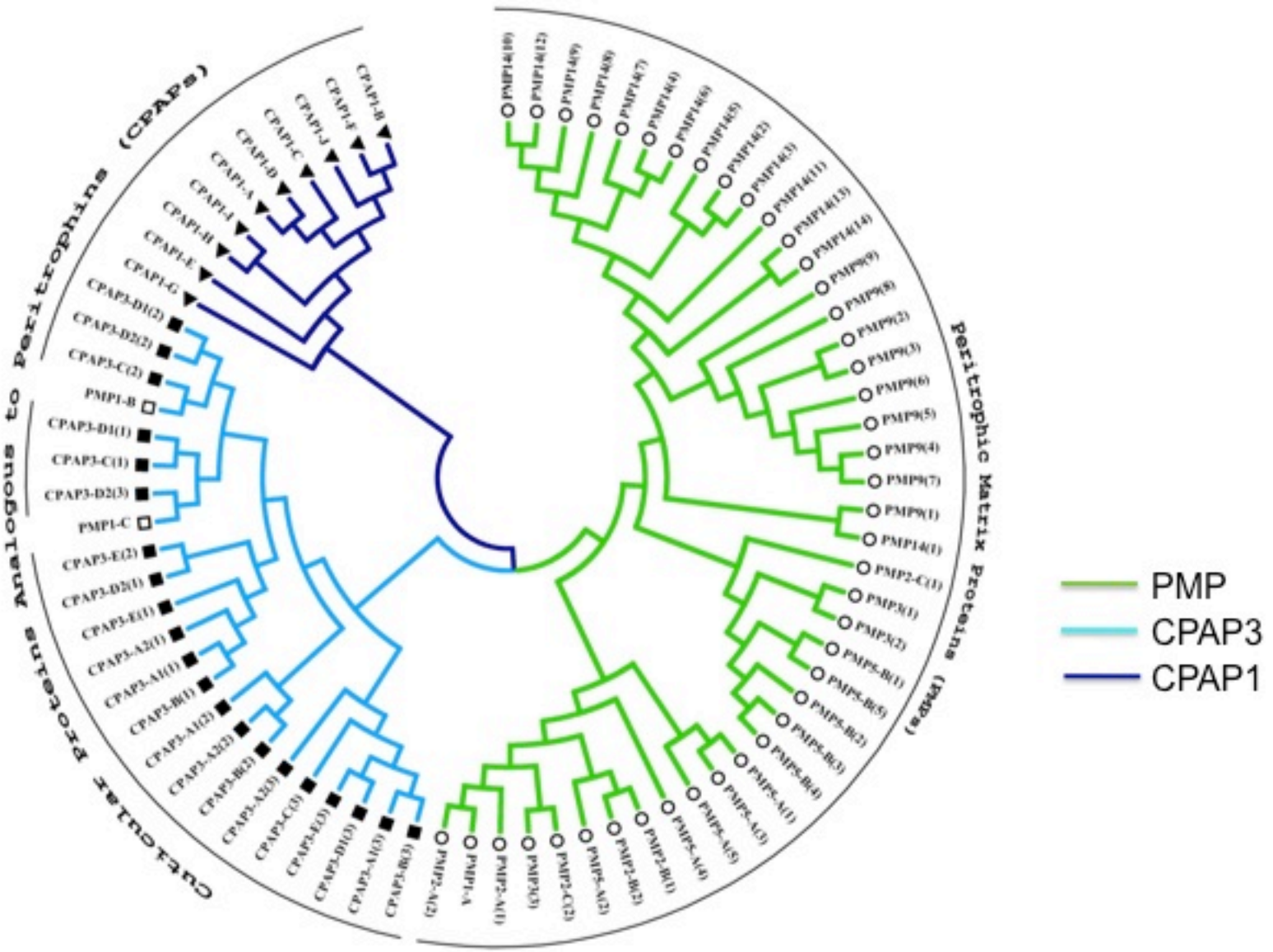

Cuticular proteins Analogous to peritrophins (CPAPs)
Peritrophic Matrix Proteins (PMPs)
CPAP1-B
CPAP1-F
CPAP1-J
CPAP1-C
CPAP1-D
CPAP1-A
CPAP1-I
CPAP1-H
CPAP1-E
CPAP1-G
CPAP3-D1(2)
CPAP3-D2(2)
CPAP3-C(2)
PMP1-B
CPAP3-D1(1)
CPAP3-C(1)
CPAP3-D2(3)
PMP1-C
CPAP3-E(2)
CPAP3-D2(1)
CPAP3-E(1)
CPAP3-A2(1)
CPAP3-A1(1)
CPAP3-B(1)
CPAP3-A1(2)
CPAP3-A2(2)
CPAP3-B(2)
CPAP3-A2(3)
CPAP3-C(3)
CPAP3-E(3)
CPAP3-D1(3)
CPAP3-A1(3)
CPAP3-B(3)
PMP2-A(2)
PMP1-A
PMP2-A(1)
PMP2-C(3)
PMP5-A(2)
PMP2-B(2)
PMP2-B(1)
PMP5-A(4)
PMP5-A(5)
PMP5-A(3)
PMP5-A(1)
PMP5-B(4)
PMP5-B(3)
PMP5-B(2)
PMP5-B(5)
PMP5-B(1)
PMP5(2)
PMP5(1)
PMP2-C(1)
PMP14(1)
PMP9(1)
PMP9(7)
PMP9(4)
PMP9(5)
PMP9(6)
PMP9(3)
PMP9(2)
PMP9(8)
PMP9(9)
PMP14(14)
PMP14(13)
PMP14(11)
PMP14(3)
PMP14(2)
PMP14(5)
PMP14(6)
PMP14(4)
PMP14(7)
PMP14(8)
PMP14(12)
PMP14(10)
PMP
CPAP3
CPAP1

Figure 2-11 Phylogenetic tree of ChtBD2s present in PMP, CPAP1 and CPAP3 families of *T. castaneum* proteins.

Individual ChtBD2s from each protein (number in parenthesis indicates their relative location from the N-terminus) were used to construct a phylogenetic tree as described in the Materials and Methods section. PMPs are indicated by green branches ending with an open circle. The two ChtBD2s of PMP1-B and PMP1-C, which align with the ChtBD2s of the CPAP3 branch end in light blue branches with an open square. The ChtBD2s of the CPAP3 proteins are indicated in light blue branches ending in a closed square. The ChtBD2s of CPAP1 proteins are indicated in dark blue branches ending in a closed triangle.

2.4.7. Peritrophic matrix proteins (PMPs) in *T. castaneum*

The nomenclature for the 11 PMPs is according to the following convention. The number that follows the designated PMP indicates the number of ChtBD2s in the protein. For example, PMP9 indicates a PMP with nine ChtBD2s. The ChtBD2s within this protein (and other PMPs) are numbered from the N-terminus to the C-terminus e.g. TcPMP9 (#1), TcPMP9 (#2) and so on. If there is more than one protein with the same number of ChtBD2s, then an upper case alphabet follows the number (e.g. TcPMP2-A, TcPMP2-B and TcPMP2-C).

The number of ChtBD2s in a PMP ranges from one to 14 and the molecular sizes range from 18.7 to 147 kDa in Table 2-1. All of these proteins have an N-terminal signal peptide and no trans-membrane helices, indicating that they are secreted proteins. Figure 2-4 shows the distribution of the ChtBD2s in the 11 PMPs. There are three PMPs with one ChtBD2, three with two, two with five and one each with three, nine and 14 ChtBD2s. Six of the PMPs (TcPMP2-A, TcPMP2-B, TcPMP5-A, TcPMP5-B, TcPMP9 and TcPMP14) are essentially tandem repeats of ChtBD2s and linker regions. The ChtBD2s and linker regions in TcPMP9 and TcPMP14 are closely related, but the linker in TcPMP14 is slightly longer (37 amino acids in 10 out of 14

ChtBD2s in TcPMP14 versus 21 amino acids in seven out of nine ChtBD2s in TcPMP9). The linkers in both proteins are rich in proline and threonine and also contain a few serines. NetOGlyc 3.1 predicts that these linker regions are likely to be O-glycosylated. The linker regions in TcPMP5-B are much longer (48-65 amino acids) than the linkers in TcPMP9 and TcPMP14, and have long stretches of serines shown in Figure 2-12. They resemble mucin domains but are not predicted to be highly glycosylated by the NetOGlyc 3.1 software. The other four PMPs, PMP1-A, PMP1-B, PMP1-C and PMP2-C, have, in addition to the ChtBD2, sequences encoding proteins for which orthologs have not been detected in either *Drosophila* or *Anopheles* species.

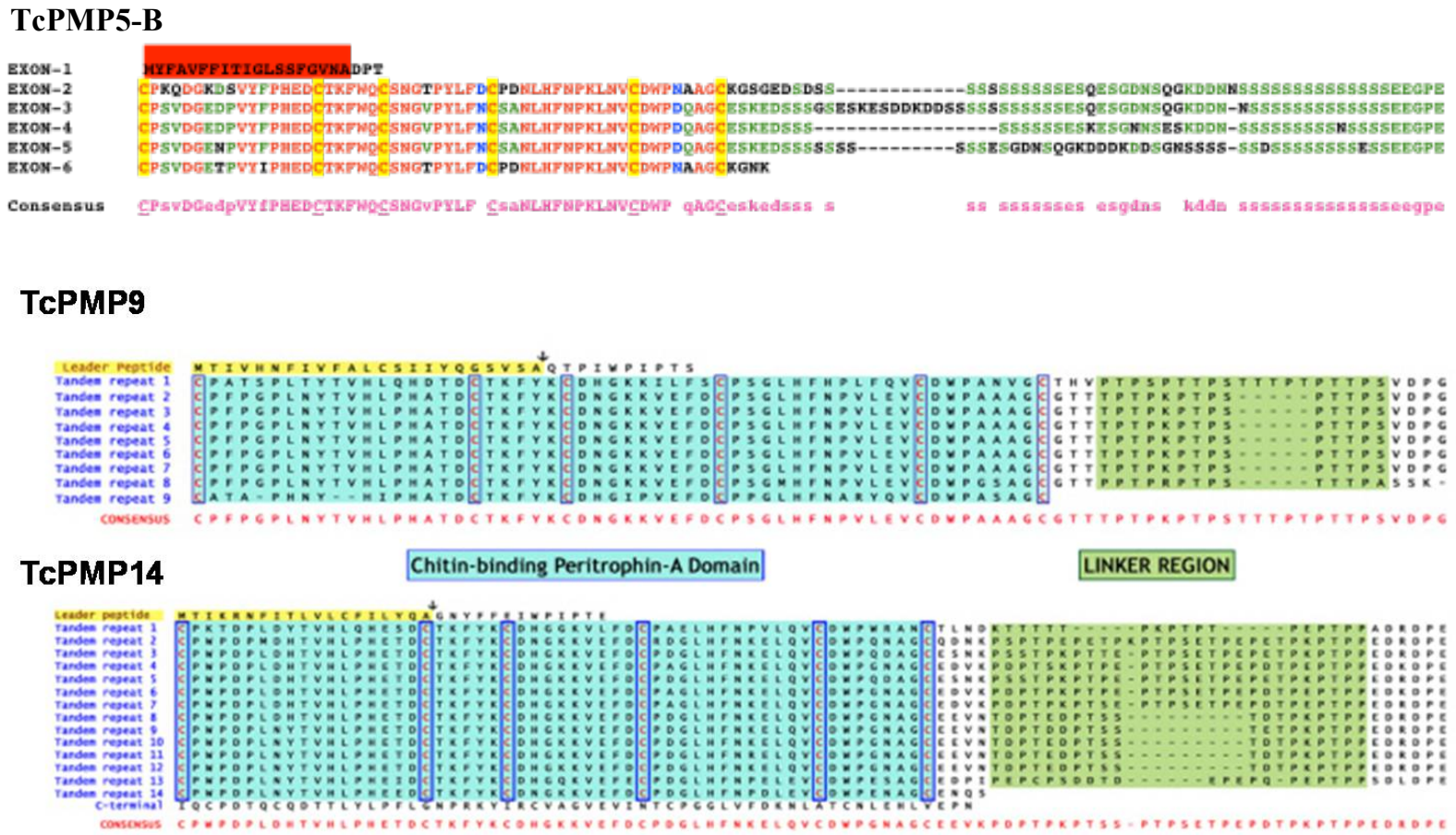


Figure 2-12 Protein sequence alignment of TcPMP5-B, TcPMP9 and TcPMP14

TcPMP5-B: The leader peptide is shaded in yellow. The predicted cleavage site is indicated by black arrow. Each tandem repeat consists of a ChtBD2 domain (shaded in cyan) and a linker region (shaded in green). TcPMP9 and TcPMP14: Each has 9 and 14 tandem repeats respectively. A tandem repeat consists of a ChtBD2 domain shaded in cyan and the linker region shaded in green. The tandem repeats consists of 6 cysteine residues shown in red and boxed and the linker region is rich in threonine, proline and serine residues.

2.4.8. Cuticular proteins analogous to peritrophins (CPAPs)

All of the proteins with ChtBD2s other than the PMPs are expressed in the integument but not in the midgut and, therefore, are unlikely to be associated with the peritrophic matrix (Illustrated in 3 and 4). Since these proteins are predicted to be secreted and to have a peritrophin-like chitin-binding domain, which will allow them to interact with cuticular chitin, we have called them cuticular proteins analogous to peritrophins or CPAPs. We further divide them into two families, labeled CPAP1 and CPAP3, based on the number of ChtBD2s. The phylogenetic tree shown in Figure 2-11 also supports such a grouping.

2.4.8.1. CPAP1 family

The CPAP1 family proteins, which have only one ChtBD2, differ from the CPAP3 family proteins in ways other than the number of ChtBD2s, the most noticeable being the location of the chitin-binding domain itself (Figure 2-4). In different CPAP1 proteins, the ChtBD2 (dark blue boxes) may be located close to the N-terminus, the C-terminus or in the middle, in contrast to CPAP3 proteins, which have their tandem repeats of ChtBD2s close to the N-terminus. Secondly, introns often interrupt the ChtBD2-encoding regions, which is rare in other *CPAP* genes. Nevertheless, the ChtBD2s of the CPAP1 family of proteins are highly related, indicating a common ancestral origin shown in Figure 2-11. Proteins belonging to this group also have N-terminal signal peptides and lack trans-membrane domains, properties consistent with a role that involves interactions with an extra cellular matrix such as chitin.

Of the 10 proteins belonging to this family, TcCPAP1-D, E and F are the smallest, containing between 114 and 124 amino acid residues of which 16-22 amino acids represent the

signal peptide and 53-54 are in the ChtBD2, shown in Figure 2-4. Therefore, the mature protein is almost entirely made up of the ChtBD2 and little else. The other seven proteins have variable lengths of protein-coding regions of which the ChtBD2-coding region represents only a minor part of the overall sequence.

2.4.8.2. CPAP3 family

There are seven genes in *T. castaneum* encoding eight proteins that have a tandem arrangement of three ChtBD2s (Figure 2-4). Orthologs of these genes have been found in *D. melanogaster* and several other species of insects. The *D. melanogaster* genes have been named previously as "*obstructors*" (Behr and Hoch, 2005) or as "*gasps*" ("gene analogous to small peritrophins," (Barry et al., 1999). We propose to name this family of proteins the "CPAP3 family" with the number 3 indicating that they have three ChtBD2s. For ease of comparison to the *D. melanogaster* orthologs, we have assigned individual proteins of this family, names such as TcCPAP3-B, TcCPAP3-C and TcCPAP3-E based on their orthology to the corresponding *D. melanogaster* proteins. In the case of TcCPAP3-A and Tc-CPAP3-D, *T. castaneum* has two paralogs instead of one protein each as in *D. melanogaster*. We have denoted them as TcCPAP3-A1, TcCPAP3-A2, TcCPAP3-D1 and TcCPAP3-D2. In the case of *TcCPAP3-C*, two alternatively spliced transcripts are generated from a single gene using one or the other of the two alternate forms of exon 5. However, *T. castaneum* may be unique among insects with fully sequenced genomes in having two alternatively spliced transcripts for this gene. Since the first four exons code for the signal peptide and the three ChtBD2s, the encoded proteins differ only in the C-terminal regions that follow the three ChtBD2s. These proteins are referred to as TcCPAP3-C5a and TcCPAP3-C5b. In the case of TcCPAP3-E, there is only one gene for this protein in *T. castaneum* and several other insects, whereas there are two orthologs in *D.*

melanogaster (Behr and Hoch, 2005). On the other hand, unlike other insects, *D. melanogaster* has a second group of genes encoding proteins with three ChtBD2s. These proteins, which are placed in a separate subgroup (#2), differ from the first group in having spacers of variable length, but they exhibit extensive sequence similarities with the ChtBD2s of CPAP3 proteins of the first group. There are no orthologs for these genes in *T. castaneum.* (Behr and Hoch, 2005) have reported that *D. melanogaster* is unique among insects in having this additional subgroup. A phylogenetic analysis of CPAP3 proteins from several insects with completed genome sequences is shown in Figure 2-13.

It is clear that CPAP3 proteins form five separate branches, each branch corresponding to one of the five CPAP3 proteins, A through E. It is clear that branches containing the CPAP3-A and CPAP3-D proteins have undergone an expansion in several insects, but branches representing CPAP3-B, C and E have only a single representative with the exception of *D. melanogaster,* which has two orthologs of CPAP3-E.

The linker regions between the first and second ChtBD2s in the CPAP3 family proteins are very similar to the corresponding linker regions in the orthologs from other insect species. This linker, which is 13 amino acids-long in all members of this family, does not contain P, T or S residues, which are found in linkers of many PMPs (e.g. TcPMP9 and TcPMP14). The linker between the second and third ChtBD2s varies in length from six to thirteen amino acid residues among the CPAP3 proteins and is devoid of residues that are associated with O-glycosylation. As indicated in Figure 2-4, the C-terminal amino acid sequences following the ChtBD2s in different CPAP3 proteins are variable in both length and sequence.

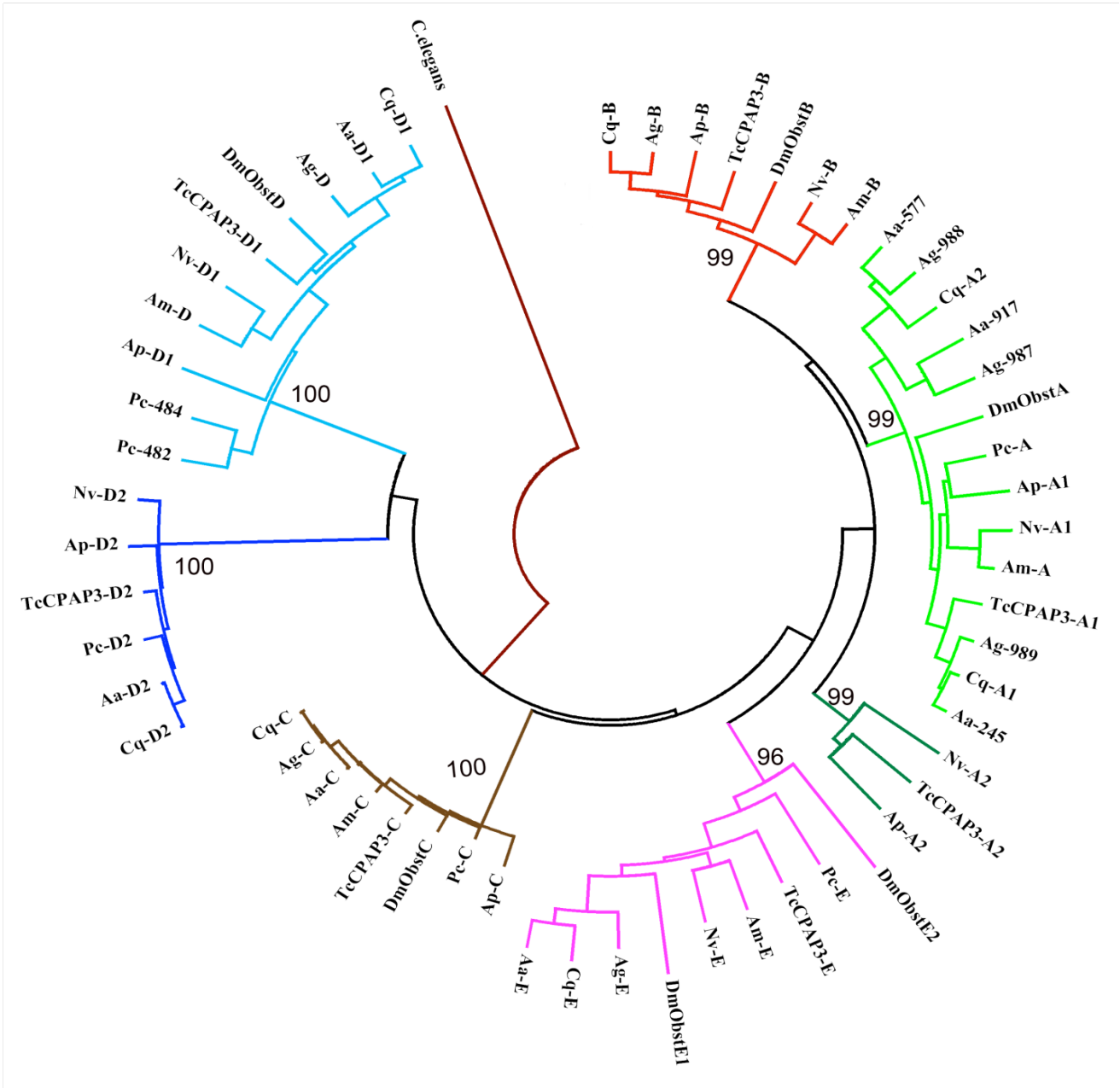

C.elegans
Cq-D1
Aa-D1
Ag-D
DmObstD
TcCPAP3-D1
Nv-D1
Am-D
Ap-D1
Pc-484
Pc-482
100
Nv-D2
Ap-D2
100
TcCPAP3-D2
Pc-D2
Aa-D2
Cq-D2
Cq-C
Ag-C
Aa-C
Am-C
TcCPAP3-C
DmObstC
Pc-C
Ap-C
100
Aa-E
Cq-E
Ag-E
DmObstE1
Nv-E
Am-E
TcCPAP3-E
Pc-E
DmObstE2
96
Ap-A2
TcCPAP3-A2
Nv-A2
99
Aa-245
Cq-A1
Ag-989
TcCPAP3-A1
Am-A
Nv-A1
Ap-A1
Pc-A
DmObstA
99
Ag-987
Aa-917
Cq-A2
Ag-988
Aa-577
99
Am-B
Nv-B
DmObstB
TcCPAP3-B
Ap-B
Ag-B
Cq-B

Figure 2-13 Phylogenetic analysis of the ChtBD2s of the CPAP3 family of proteins from different insect orders.

The proteins from Drosophila melanogaster (Dm), Anopheles gambiae (Ag), Acyrthosiphon pisum (Ap), Aedes aegypti (Aa), Apis mellifera (Am), Culex quinquefasciatus (Cq), Nasonia vitripennis (Nv), Pediculus humanus corporis (Pc), Tribolium castaneum (Tc) and Caenorhabditis elegans (C. elegans). The CPAP3 orthologs from different organisms are represented in different colors: A1, light green; A2, dark green; B, red; C, brown; D1, cyan; D2, dark blue; E, pink; and C. elegans in dark red. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 9.98133851 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 141 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4. A bootstrap analysis of 5000 replications was carried out on the tree. Bootstrap values of the major branches are indicated by numbers. The accession numbers of the orthologs from different insect species are listed in Appendix AC.

2.4.9. Chitinases/chitin deacetylases

Included among the proteins with ChtBD2s listed in (Table 2-1) are several chitinases and chitin deacetylases. Among the *T. castaneum* chitinases, there is one with 5 ChtBD2s (chitinase 10) and five with one ChtBD2 (chitinases 4, 5, 6, 7 and 8). The ChtBD2s are at the C-termini with one exception being the largest enzyme, chitinase 10, which has the ChtBD2s interspersed among its five catalytic domains (Zhu et al., 2008b). Similarly, there are seven *T. castaneum* chitin deacetylases including their isoforms, each with one ChtBD2. In contrast to the chitinases, the ChtBD2s are found in N-terminal regions of chitin deacetylases (Dixit et al., 2008). These enzymes and the expression profiles of the genes encoding them have been previously described in detail by (Arakane and Muthukrishnan, 2010b; Dixit et al., 2008; Zhu et al., 2008b) and, therefore, are not a major focus of the current discussion involving non-enzymatic proteins of *T. castaneum* with ChtBD2s.

2.4.10. Other *T. castaneum* proteins that contain ChtBD2s

In addition to the 42 proteins listed in Table 2-1 that are predicted to be putative chitin-interacting proteins, there are two other proteins with ChtBD2s in the predicted proteome of *T. castaneum*. One of them (Glean_15245*)* is in a region where sequence data is incomplete. We could not determine whether the other gene (Glean_09553) is expressed in the gut or a cuticle forming tissue. These rather unique proteins are described in the following sections.

2.4.10.1. Glean_15245

The protein represented by Glean_15245 in Beetlebase has a signal peptide followed by a tandem assembly of five ChtBD2s. However, we could not confirm this prediction by cloning the corresponding full-length cDNA using several combinations of forward and reverse primers designed from the predicted gene sequence. We believe that the presence of a sequence gap of ~ 4 kb in the genome assembly is primarily responsible for this negative outcome. However, we were able to amplify a partial cDNA clone that includes two of the ChtBD2s predicted to occur at the 5'-end of the Glean and two additional ChtBD2s not predicted by the Glean. A comparison of the sequence of this clone with the genome assembly indicates that several introns predicted by the Glean are indeed exons, which increase the number of ChtBD2s in the isolated cDNA. We could also amplify another cDNA with an ORF corresponding to the last three ChtBD2s predicted by the Glean model. If Glean_15245 indeed codes for a single protein, the encoded protein should have seven or more ChtBD2s, including any additional ChtBD2 encoded by exons in the unsequenced region of this gene. Alternatively, there may be two tandem genes encoding proteins with ChtBD2s. We have been unable to amplify that region of the genome to test this possibility. Until that gap is bridged, the annotation of this gene(s) and its protein products will remain incomplete. RT-PCR indicates that this gene is expressed in the carcass

and elytra and hindwings that were dissected out from 3-4 d-old pupae, but not in the midgut, suggesting that it belongs in the CPAP group. It maps to position 2.5 cM in chromosome 6.

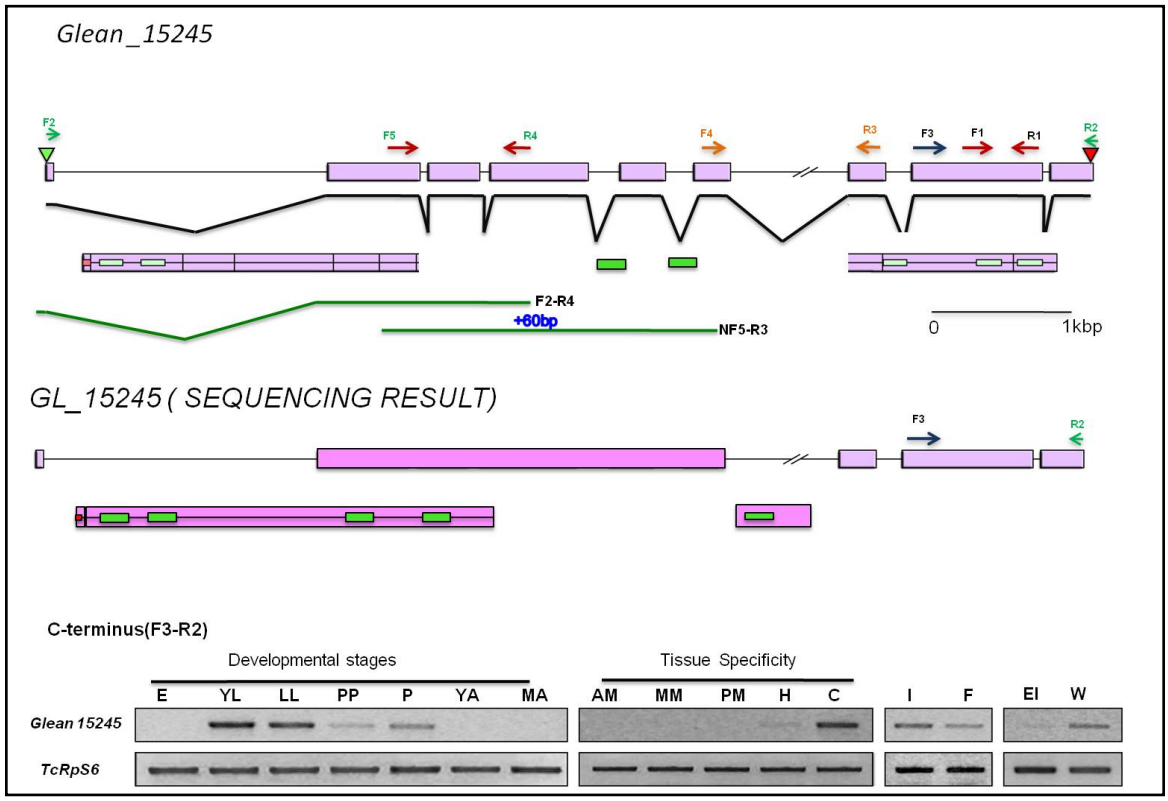


Figure 2-14 Schematic diagram showing gene domain organization and expression profile of Glean_15245.

Glean _15245 is predicted to have 9 exons and 5 ChtBD2 domain, and has a gapped region (sequence not determined between E6 and E7 indicated by a double slash. The sequencing result shows that it has 5 exons and 7-8 ChtBD2s. The unmapped region however could not be amplified with genomic PCR using F4-R3 primers. A polyadenylation signal was also not obtained by 3'RACE using F4 primer. Also a 5'UTR was not found using R3 primer. The gene model obtained after sequencing shows that there is one large 2nd exon with 4 ChtBD2, and small part of the ungapped region that was sequenced also contains a ChtBD2. The bottom panel shows that this gene has expression in the larval and pupal stage, in cuticle forming tissues but not in gut tissues as revealed by RT-PCR data. Different stages of development E, Eggs; YL, penultimate instar or younger larvae; LL, last instar larvae; PP, pharate pupae; P, Pupae; YA, young adult; MA, Mature adult, AM: anterior midgut, MM: middle midgut, PM: posterior midgut, H: hindgut, C: carcass (whole body minus gut). I: integument, F: fatbody, E: elytron, W: hindwing.RT-PCR products for the T. castaneum ribosomal protein 6 (TcRpS6) from the same cDNA templates served as an internal control for loading.

2.4.10.2. Glean_09553

This gene model predicts a protein with a signal peptide and three ChtBD2s. However, only the middle ChtBD2 follows the consensus shown in Figure 2-10. This gene shows trace amount of expression in the larval stages. In the adult it is expressed in the carcass, but not in the midgut or the two cuticle-forming tissues such as elytra and hindwings Figure 2-15. With the available data, we cannot predict whether this protein would be associated with chitin in the cuticle.

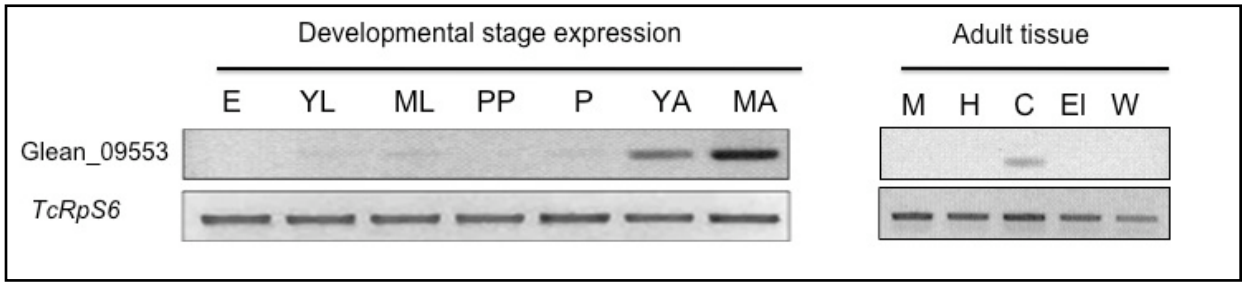


Figure 2-15 Developmental stage expression and tissue-specific expression of Glean 09553 as revealed by RT-PCR

*Glean_09553 shows expression in the adult stage, with very trace amounts in larval stage in cuticle forming tissues but not in gut tissues as revealed by RT-PCR data. Different stages of development E, Eggs; YL, penultimate instar or younger larvae; LL, last instar larvae; PP, pharate pupae; P, Pupae; YA, young adult; MA, Mature adult, AM: anterior midgut, MM: middle midgut, PM: posterior midgut, H: hindgut, C: carcass (whole body minus gut). I: integument, F: fatbody, E: elytron, W: hindwing. RT-PCR products for the T. castaneum ribosomal protein 6 (*TcRpS6*) from the same cDNA templates served as an internal control for loading.*

2.5. Discussion

The goal of this research has been to carry out a comprehensive search for proteins in *T. castaneum* that contain one or more ChtBD2s with the expectation that these proteins, which are predicted to interact with chitin, will have a role in assembly and/or turnover of chitin-containing

structures. This domain was found in several proteins known as peritrophins that were extracted from the PM and shown to bind to chitin tightly (Tellam et al., 1999; Wang et al., 2004). This domain is also found in some but not all chitinases and chitin deacetylases, which degrade or modify chitin (Arakane and Muthukrishnan, 2010a; Dixit et al., 2008; Kramer, 1993; Zhu et al., 2008b), in proteins expressed in embryonic tracheae and epidermis (Barry et al., 1999; Behr and Hoch, 2005) and in the proteins encoded by genes expressed in adult hindgut and Malpighian tubules of cat fleas (Gaines et al., 2003). Removal of ChtBD2s from some of these proteins has been shown to result in lower affinity for chitin and addition of one or more ChtBD2s has been found to increase the affinity for chitin (Arakane et al., 2003; Zhu et al., 2001). It has been proposed that ChtBD2s help to anchor the chitinases and chitin deacetylases on the insoluble polymer, facilitating the catalytic center of the same enzyme to locate a target site in the chitin substrate more efficiently. However, there is very little information on the role of ChtBD2s in proteins that do not have enzymatic functions.

Recently, (Venancio et al., 2009) identified ChtBD2-containing proteins in insect proteomes by searching *D. melanogaster*, *A. aegypti* and *T. castaneum* genomic and EST databases using the six-cysteine sequence motif, $CX_{13-20}CX_{5-6}CX_{9-19}CX_{10-14}C\text{-}X_{4-14}C$ proposed by (Tellam et al., 1999) for the ChtBD2 as the query. They reported the identification of 25 "peritrophins" in the *T. castaneum* proteome, which is considerably less than the 65 proteins that had been predicted from the *D. melanogaster* genome (Venancio et al., 2009). To maximize the probability of identifying proteins with ChtBD2s, in our bioinformatics search of the *T. castaneum* genome, we used as queries ChtBD2 sequences not only from "peritrophins" from several insects but also ChtBD2 from chitinases and chitin deacetylases from *T. castaneum*, which have a slightly different consensus sequence from that of the former set of proteins. We

also utilized ChtBD2s from proteins encoded by *D. melanogaster gasp* and *obstructor* genes as queries (Barry et al., 1999; Behr and Hoch, 2005). The total number (~50) of *T. castaneum* proteins with ChtBD2s that we have identified in the *T. castaneum* genome is more than the number (25) of proteins reported by (Venancio et al., 2009), indicating that our search strategy has been more comprehensive. We verified the ORF predictions and exon/intron assignments for the *T. castaneum* genes encoding proteins with ChtBD2s from three databases (NCBI, Baylor, BeetleBase) by direct cloning of nearly all of the cDNAs. These studies identified new genes and corrected mistakes in several gene models including some that revised the number of ChtBD2s (both an increase and a decrease) in the predicted ChtBD2-containing proteins. Most significantly, we identified four new proteins belonging to the PMP family (TcPMP1-A, TcPMP2-A, TcPMP9 and TcPMP14), four additional members of the *T. castaneum* CPAP3 family and one member of the CPAP1 family, which have not been reported previously. Five other proteins with ChtBD2s were identified, but they were found to lack a predicted signal peptide and were left out of our analysis as they are predicted to be intracellular proteins. One putative extracellular protein, a hemolectin was not included because it is likely to be found in hemolymph and to have roles in clotting and/or immunity. It is not expected to bind to extracellular chitin. In addition, we identified a couple of other proteins, one with three putative ChtBD2s and another with an undetermined number of ChtBD2s. These were not analyzed extensively due to incomplete genomic data or lack of evidence for expression in tissues involved in chitin synthesis. The estimate of the total number of ChtBD2-containing proteins encoded in the *T. castaneum* genome is currently >50 including 13 chitinase and chitin deacetylase isoforms that have been previously characterized (Dixit et al., 2008; Zhu et al., 2008a). There is some uncertainty whether Glean_15245 corresponds to one or two proteins as

discussed in section 2.4.10.1. We believe that we have identified and characterized nearly all of the *T. castaneum* genes encoding proteins with ChtBD2s or related sequences.

Our phylogenetic analysis based on the sequence similarities of the ChtBD2s has resulted in grouping of the *T. castaneum* proteins containing ChtBD2s into three large families, namely the PMP, CPAP1 and CPAP3 families. Our preference for these two namesCPAP1 and CPAP3 instead of the generic term "peritrophins" is to emphasize their different physical localizations within the insect body and their widely different functions, even though all three families of proteins are known / expected to bind to chitin in the extracellular matrix.

Besides the CPAPs studied here, *T. castaneum* has a large assortment of cuticular proteins with the RR consensus (Willis, 2010), which are also thought to be involved in interactions with chitin. They are also likely to be involved in cross-linking with catechols (Andersen, 2010). Our study has indicated that several ChtBD2-containing proteins are secreted into the cuticle by the epidermal tissue. The role of the large assortment of the CPAPs in cuticle structure/function has been largely unexplored until now with the exception of *D. melanogaster* '*gasp*' genes, which have been shown to be expressed in embryonic tracheae (Barry et al., 1999). A role for "Obstructor A" in embryonic development has also been reported (Behr and Hoch, 2005). The *obstructor* family of genes is expressed during early- and mid-embryogenesis in tracheae and epidermis. It is interesting to note that while *D. melanogaster* has 10 genes encoding proteins related to the TcCPAP3 proteins that are divided into two subgroups based on the spacing and size of the linkers, *T. castaneum* has only eight proteins, all belonging to one sub-family. Whereas *T. castaneum* is missing orthologs of five *D. melanogaster* genes (*Obstructors F, G, H, I* and *J* of subgroup II), the beetle has additional genes in subgroup I (Tc*CPAP3-A2* and Tc*CPAP3-D2*). *TcCPAP3-C* generates two isoforms as a result of alternate

splicing of exon5. All other insects studied to date appear to lack the subgroup II members, but compensate by increasing the number of representatives of the first group by 2 or 3 as a result of duplication of genes encoding TcCPAP3-A and TcCPAP3-D.

I have found preliminary evidence for the presence of a couple of the CPAPs identified in this work in pupal and/or adult cuticles of *T. castaneum* by direct proteomic analyses of cuticles (Dittmer et al. unpublished; Arakane et al., unpublished). Whether these genes are expressed in different tissues or cell types and thus influence the properties of specific cuticular structures such as those associated with tracheae, elytra, hindwings and other appendages needs to be investigated. We have preliminary evidence indicating that knocking down transcripts for specific CPAP3 proteins results in molting disruption or joint defects as well as mortality (See CHAPTER 3). The finding that relative abundance of transcripts for specific CPAPs varies between elytra and hindwings, which represent hard and soft cuticle, respectively, raises the possibility that the ratios of specific CPAPs that associate with chitin may influence the physical properties of the cuticle.

Some of the ChtBD2s are associated with proteins with catalytic domains of chitinases (chitinases 5, 7 and 10 with one, one and five ChtBD2s, respectively) and all chitin deacetylases (with one ChtBD2). It is likely that the presence of a ChtBD2 increases the affinity of these enzymes for the chitin-containing matrix and binds them onto the substrate. However, there are at least ten other genes encoding *T. castaneum* proteins with one ChtBD2 belonging to the CPAP1 family. They are expressed in the integument (presumably in epidermal cells) with differing patterns of expression. Nothing is known about their functions because they have no identifiable functional domains. Most of these proteins have a ChtBD2 at one or the other terminus and additional sequences that do not correspond to any characterized domains in the

protein databases. They may be enzymes of unknown function or cuticle proteins that interact with chitin during cuticle assembly or turnover. They may be extracellular proteins based on the presence of a predicted signal peptide and the absence of other membrane spanning segments or other intracellular targeting sequences. The ChtBD2s may improve the affinity of those proteins that transiently interact with chitin during cuticle assembly or turnover. The role of the ChtBD2 in these proteins may be different from those found in CPAP3 proteins with three ChtBD2s, which we propose are very tightly associated with chitin. For example, the CPAP3 family of proteins may participate in the formation of chitin laminae or in stabilization of bundles of chitin fibers.

All of the ChtBD2s present in *T. castaneum* proteins appear to have been derived from a common ancestral CBD sequence with six cysteines belonging to the peritrophin A family. We have found no evidence for the 8- or 6-cysteine containing peritrophin B or C sequence motifs described by (Tellam et al., 1999) in the predicted sequences of *T. castaneum* proteins. We propose that a more inclusive general consensus for the conserved cysteines, which will represent all of the ChtBD2-containing proteins in most insect species described so far, is $CX_{11\text{-}24}CX_{5\text{-}6}CX_{9\text{-}14}CX_{12\text{-}16}C\text{-}X_{6\text{-}8}C$. The individual subgroup members are more closely related to one another than to members of the other groups, and they have consensus sequences that include additional conserved residues besides the six cysteines including some flanking sequences (Figure 2-10). In particular the CPAP1 family members have conserved sequences that extend on either side of the first or the sixth cysteines in the peritrophin A consensus sequence.

Phylogenetically, all of the ChtBD2s that belong to the PMP family are very closely related to one another. This is especially true among the multiple repeats of ChtBD2s

that are found in the same protein (e.g. PMPs 5, 7, 9 and 14), which are nearly identical in sequence both in the ChtBD2 and in the linker regions. We assume that the PMPs with multiple repeats arose from gene duplication and unequal crossover events, resulting in proteins with different numbers of repeats. Supporting this hypothesis is the finding that several *T. castaneum PMP* genes are made up of exons encoding a single ChtBD2 and also that the two genes with the largest number of repeats (*PMP9* and *PMP14*) are separated on the chromosome by only 331 bp of intervening sequence. In addition, these ChtBD2s and the spacers between them are extremely similar both in nucleotide and amino acid sequence, with many of them having identical nucleotide sequences. The longest PMP in *T. castaneum* has 14 ChtBD2s, whereas those from *D. melanogaster* and *A. aegypti* have 17 and 14 ChtBD2s, respectively (Venancio et al., 2009). To date, the PMP that contains the largest number of ChtBD2s is the one from *M. configurata*, which has 19 ChtBD2s (Shi et al., 2004).

The occurrence of multiple genes in each of the three families of proteins with ChtBD2s raises questions about their functions. Are all of these proteins required for survival of the insects as suggested by the finding that insects belonging to different orders of insects have retained all three families and indeed orthologs of each member of the three families? This question is answered in the next chapter by studying the expression profiles of each of these genes encoding proteins with ChtBD2 domains as a function of developmental stages and by RNAi studies in which transcripts for each of these genes were down-regulated by administration of dsRNA.

2.6. Bibliography

Andersen, S.O., (2010). Insect cuticle sclerotization: A review. *Insect Mol Biol* 40, 166-178.

Arakane, Y., Muthukrishnan, S., (2010a). Insect chitinase and chitinase-like proteins. *Cell Mol Life Sci* 67, 201-216.

Arakane, Y., Muthukrishnan, S., (2010b). Insect chitinases and chitinase-like proteins. *Cellular MOl. Life Sci*.

Arakane, Y., Zhu, Q., Matsumiya, M., Muthukrishnan, S., Kramer, K.J., (2003). Properties of catalytic, linker and chitin-binding domains of insect chitinase. *Insect Biochem Mol Biol* 33, 631-648.

Barry, M.K., Triplett, A.A., Christensen, A.C., (1999). A peritrophin-like protein expressed in the embryonic tracheae of *Drosophila melanogaster*. *Insect Biochem Mol Biol* 29, 319-327.

Beeman, R.W., Stuart, J.J., (1990). A gene for lindane + cyclodiene resistance in the red flour beetle (Coleoptera: Tenebrionidae). *J.Econ. Entomol* 83, 1745-1751.

Behr, M., Hoch, M., (2005). Identification of the novel evolutionary conserved obstructor multigene family in invertebrates. *FEBS Lett* 579, 6827-6833.

Dinglasan, R.R., Devenport, M., Florens, L., Johnson, J.R., McHugh, C.A., Donnelly-Doman, M., Carucci, D.J., Yates, J.R., 3rd, Jacobs-Lorena, M., (2009). The *Anopheles gambiae* adult midgut peritrophic matrix proteome. *Insect Biochem Mol Biol* 39, 125-134.

Dixit, R., Arakane, Y., Specht, C.A., Richard, C., Kramer, K.J., Beeman, R.W., Muthukrishnan, S., (2008). Domain organization and phylogenetic analysis of proteins from the chitin deacetylase gene family of *Tribolium castaneum* and three other species of insects. *Insect Biochem Mol Biol* 38, 440-451.

Elvin, C.M., Vuocolo, T., Pearson, R.D., East, I.J., Riding, G.A., Eisemann, C.H., Tellam, R.L., (1996). Characterization of a major peritrophic membrane protein, peritrophin-44, from the larvae of *Lucilia cuprina*. cDNA and deduced amino acid sequences. *J Biol Chem* 271, 8925-8935.

Gaines, P.J., Walmsley, S.J., Wisnewski, N., (2003). Cloning and characterization of five cDNAs encoding peritrophin-A domains from the cat flea, *Ctenocephalides felis*. *Insect Biochem Mol Biol* 33, 1061-1073.

Goto, A., Kumagai, T., Kumagai, C., Hirose, J., Narita, H., Mori, H., Kadowaki, T., Beck, K., Kitagawa, Y., (2001). A *Drosophila* haemocyte-specific protein, hemolectin, similar to human von Willebrand factor. *Biochem J* 359, 99-108.

Hegedus, D., Erlandson, M., Gillott, C., Toprak, U., (2009). New insights into peritrophic matrix synthesis, architecture, and function. *Annu Rev Entomol* 54, 285-302.

Karouzou, M.V., Spyropoulos, Y., Iconomidou, V.A., Cornman, R.S., Hamodrakas, S.J., Willis, J.H., (2007). *Drosophila* cuticular proteins with the R&R Consensus: annotation and classification with a new tool for discriminating RR-1 and RR-2 sequences. *Insect Biochem Mol Biol* 37, 754-760.

Kramer, K.J., Corpuz, L., Choi, H. K., Muthukrishnan, S., (1993). Sequence of a cDNAand expression of the gene encoding epidermal and gut chitinases of *Manduca sexta*. *Insect Biochem Mol Biol* 23, 691-701.

Kramer, K.J., Kanost, M.R., Hopkins, T.L., Jiang, H.B., Zhu, Y.C., Xu, R.D., Kerwin, J.L., Turecek, F., (2001). Oxidative conjugation of catechols with proteins in insect skeletal systems. *Tetrahedron* 57, 385-392.

Schneider, T.D., Stephens, R.M., (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* 18, 6097-6100.

Shen, Z., Jacobs-Lorena, M., (1999). Evolution of chitin-binding proteins in invertebrates. *J Mol Evol* 48, 341-347.

Shi, X., Chamankhah, M., Visal-Shah, S., Hemmingsen, S.M., Erlandson, M., Braun, L., Alting-Mees, M., Khachatourians, G.G., O'Grady, M., Hegedus, D.D., (2004). Modeling the structure of the type I peritrophic matrix: characterization of a *Mamestra configurata* intestinal mucin and a novel peritrophin containing 19 chitin binding domains. *Insect Biochem Mol Biol* 34, 1101-1115.

Tamura, K., Dudley, J., Nei, M., Kumar, S., (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24, 1596-1599.

Tellam, R.L., Wijffels, G., Willadsen, P., (1999). Peritrophic matrix proteins. *Insect Biochem Mol Biol* 29, 87-101.

Tjoelker, L.W., Gosting, L., Frey, S., Hunter, C.L., Trong, H.L., Steiner, B., Brammer, H., Gray, P.W., (2000). Structural and functional definition of the human chitinase chitin-binding domain. *J Biol Chem* 275, 514-520.

Venancio, T.M., Cristofoletti, P.T., Ferreira, C., Verjovski-Almeida, S., Terra, W.R., (2009). The *Aedes aegypti* larval transcriptome: a comparative perspective with emphasis on trypsins and the domain structure of peritrophins. *Insect Mol Biol* 18, 33-44.

Wang, W., Davison, S., Krell, P.J., (2004). Identification and characterization of a major early-transcribed gene of *Trichoplusia ni* single nucleocapsid nucleopolyhedrovirus using the baculovirus expression system. *Virus Genes* 29, 19-29.

Willis, J.H., (2010). Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. *Insect Biochem Mol Biol* 40, 189-204.

Zhu, Q., Arakane, Y., Banerjee, D., Beeman, R.W., Kramer, K.J., Muthukrishnan, S., (2008a). Domain organization and phylogenetic analysis of the chitinase-like family of proteins in three species of insects. *Insect Biochem Mol Biol* 38, 452-466.

Zhu, Q., Arakane, Y., Beeman, R.W., Kramer, K.J., Muthukrishnan, S., (2008b). Characterization of recombinant chitinase-like proteins of *Drosophila melanogaster* and *Tribolium castaneum*. *Insect Biochem Mol Biol* 38, 467-477.

Zhu, X., Zhang, H., Fukamizo, T., Muthukrishnan, S., Kramer, K.J., (2001). Properties of *Manduca sexta* chitinase and its C-terminal deletions. *Insect Biochem Mol Biol* 31, 1221-1230.

CHAPTER 3

Cuticular Proteins Analogous to Peritrophins

3.1. Abstract

This study is focused on functional characterization of *Tribolium castaneum* genes that encode Cuticular Proteins Analogous to Peritrophins (CPAP) containing the chitin-binding domain, ChtBD2, with six characteristically spaced cysteines (Peritrophin A motif=CBM_14=Pfam 01607). These genes are expressed exclusively in cuticle-forming tissues and have been classified into two families, *CPAP1* and *CPAP3* depending on the number of ChtBD2 domains in the proteins. *CPAP1* family has 10 members, each with one ChtBD2 domain and *CPAP3* has eight members, each with three ChtBD2 domains. Individual members of the *CPAP1*and *CPAP3* gene families have distinct developmental patterns of expression. Many of these genes are essential for development, molting, cuticle integrity or proper locomotion and fecundity. RNA interference (RNAi) studies targeting *TcCPAP1-C, TcCPAP1-H, TcCPAP1-J* and *TcCPAP3-C* transcripts result in death at the pharate adult stage. RNAi for other *CPAP3* genes results in different developmental defects including abnormal elytra, hindwings, abnormal gait, or adult/embryonic mortality. Scanning electron microscopic analysis of pharate adults following RNAi for the *TcCPAP3-D* gene shows that elytral cuticle is affected. These results provide experimental support for specialization in the functions of several CPAP proteins in *T. castaneum*. With minor exceptions, many of them serve essential and non-redundant functions in maintaining the structural integrity of the cuticle.

3.2. Introduction

Chitin, a linear polymer of β-1-4 linked N-acetyl-D-glucosamine, is an important component of the insect exoskeleton. It is present in epidermal and tracheal cuticle, and the

peritrophic matrix (PM) that lines the midgut. Insect procuticle is composed mainly of chitin complexed with proteins that contain chitin-binding domains. The two major groups of chitin-binding domains found in insect cuticular proteins have either the Rebers & Riddiford Consensus sequence (R&R Consensus; pfam00379; (Willis, 2010) that lacks cysteine residues, or the cysteine-rich peritrophin-A motif. The latter contains six distinctly spaced cysteines and was found initially in numerous proteins extracted from insect PMs (Tellam et al., 1999). Later, genes encoding proteins with this peritrophin-A type chitin-binding motif, ChtBD2, have been identified in cuticle-forming tissues from *Drosophila melanogaster* and *Ctenocephalides felis* (Barry et al., 1999; Behr and Hoch, 2005; Gaines et al., 2003). By conducting a bioinformatics search of *T. castaneum* genome, we identified 49 putative genes capable of encoding 50 proteins with one or more ChtBD2s (Jasrapuria et al., 2010), which included several enzymes of chitin metabolism and PM proteins with one to 14 ChtBD2 domains. Among these fifty proteins were two distinct subfamilies of cuticular proteins that were expressed only in cuticle-forming tissues with one or three ChtBD2 domains, which we have denoted as CPAP1 (consisting of ten members with one ChtBD2 domain) and CPAP3 families (eight members with three ChtBD2 domains). The CPAP1 family proteins are of variable length and dissimilar in their amino acid sequences, except in the Cht2BD2 domain that typically constitutes only a small fraction of the total length. On the other hand, the CPAP3 family proteins are a collection of three repeats of ChtBD2 domains with small spacers between these domains. These proteins, all of which have three ChtBD2s are encoded by orthologs of members of the family of *D. melanogaster* genes collectively known as "*gasps*" or "*obstructors*," (Barry et al., 1999; Behr and Hoch, 2005). We have named the individual *T. castaneum* orthologs of this sub-family to reflect their orthology to the corresponding *Drosophila* genes and the number of encoded ChtBD2s (e.g. *T. castaneum*

CPAP3-B is the ortholog of *D. melanogaster Obst-B* and encodes a proteins with three ChtBD2s). There are five fewer *CPAP3* orthologs in *T. castaneum* than in *D. melanogaster* (Behr and Hoch, 2005). We have proposed the new name "Cuticular **P**rotein Analogous to **P**eritrophins" for these proteins. The previous name "gene analogous to small peritrophins" (*gasp*) (Barry et al., 1999) suggests that it is similar to peritrophins, but the CPAPs have nothing to do with PM except for having a "Peritrophin-A domain" and they are not necessarily small proteins. The other name " Obstructor" coined by (Behr and Hoch, 2005) as the embryos with a mutant form of this gene showed a "barrier brake down phenotype" (a term that was never defined) is also not very informative (Willis, 2010). Phylogenetic analysis indicated that the members of the CPAP3 family proteins are more closely related to one another and belonged to a different branch of the larger tree that included the CPAP1 proteins as well as a large assortment of peritrophins that have related ChtBD2 domains (Jasrapuria et al., 2010).

In this chapter, I have carried out an extensive functional analysis of *T. castnaeum* genes encoding each of these eighteen proteins belonging to the *CPAP1* and *CPAP3* families in order to determine their role in cuticle formation or integrity. RNAi results indicate that many of these genes have distinct and non-redundant essential functions in development and in maintenance of the structural integrity of the insect cuticle.

3.3. Materials and Methods

3.3.1. Insect cultures

Strains used were *T. castaneum* GA-1 strain and *pu11*, an enhancer trap line in which the gene for green fluorescent protein (GFP) was expressed in the wing and elytral discs at pharate pupal stage (Lorenzen et al., 2003). In *T. castaneum* the different larval instars cannot be

precisely determined, and so the *pu11* strain ensures the identification of animals in the penultimate larval instar and facilitates RNAi experiments at desired developmental stages. Insects were reared at 30°C in wheat flour containing 5% brewer's yeast under standard conditions described previously (Beeman and Stuart, 1990).

3.3.2. Developmental expression profiles of the *CPAP* gene families

The developmental expression patterns of the ten *CPAP1* genes and seven *CPAP3* genes were analyzed by RT-PCR. Total RNA was isolated from whole insects at various stages of development including embryos, young larvae (larval instars 4-5), last instar larvae, pharate pupae, pupae, young adults (± 3 h after adult eclosion) and mature adults (10 d-old) as well as pupae collected from the pupal stage, day-0 to day-5 by using RNeasy Mini kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. First strand cDNA synthesis was carried out using Superscript III first-strand synthesis system (Invitrogen, Carlsbad, CA) using 1μg of total RNA for each reaction. This cDNA served as template for subsequent PCR reaction using the gene-specific primers as listed in Table 3-1.

3.3.3. Profiles of gene expression during development

The RNeasy Mini kit (Qiagen, Valencia, CA) was used to isolate total RNA from mature larvae according to the manufacturer's instructions. The gut of mature larvae was cut into 3 parts, anterior, middle and posterior, and RNA was isolated from separate pools of each of these segments. Total RNA samples were treated with RNase-free DNase I (Ambion, Austin, TX, 2U/μl) for 20 min at 37°C to remove genomic DNA contamination. The Superscript III first –strand synthesis system for RT-PCR (Invitrogen, Carlsbad, CA) was used to synthesize first-strand cDNA according to the manufacturer's instructions. Oligo- $(dT)_{20}$ was used as a primer for

reverse transcription using 1.5-2 μg of total RNA as template. RT-PCR was carried out to check the tissue-specificity of expression of each gene using pairs of gene-specific primers (Table 3-1).

Table 3-1 Sequences of primers used for gene expression analysis.

F = forward primer; R = reverse primer

	Gene	Gene-specific primers 5'--3'
Cuticular Proteins Analogous to Peritrophins (CPAPs)	*TcCPAP1-A*	**F:** CACCAAACGTATGAGGGCTA **R:** GTAAATCCCAGGATGGGCAGG
	TcCPAP1-B	**F:** ATGAATTATTTAACGTGCGTTTTACACACTG **R:** TCATTCATCGTTAAAAATTTCTTCTAATAATTGCTTCG
	TcCPAP1-C	**F:** ATGTGGCGGCTTTTAGGATTTG **R:** TCACTCTTTCTCGGATGGCAC
	TcCPAP1-D	**F:** GGCCATTACATAATCCCCGG **R:** GGAGCAATAAATAATTTAATGAGACTGCG
	TcCPAP1-E	**F:** ATGTTAAGGATCGTAGCTGTATCAG **R:** CTAATTGTTGTCGTGGTAATCGTCC
	TcCPAP1-F	**F:** ATGAATTATTTAACGTGCGTTTTACACACTG **R:** TCGTTAATGCCGTACAGACTCGGA
	TcCPAP1-G	**F:** GAGCTGGAAGAAGAGGAAATCGAGG **R:** GTCTGTGGACTTGTTCAGGTTCTTCC
	TcCPAP1-H	**F:** GACAACGACTACTACAACGACCAC **R:** TCAGAAACTCGTGTAGAGGGC
	TcCPAP1-I	**F:** ATGAAGGGTG TCAGTTTTAT TTTCATAG **R:** CTAGTTTTTTACATTTGCCTTCTTAATTTC
	TcCPAP1-J	**F:** GAGGAACGGAAGAAAACTTGCAAG **R:** CTAGTTTTTTACATTTGCCTTCTTAATTTC
	TcCPAP3-A1	**F:** ATACGAAGACCCGCGACAATGTGA **R:** AACTACAAGTTGGCCATTGGCGTC
	TcCPAP3-A2	**F:** AGACCGTAACTGTCGCACTTCTGA **R:** AAATCTCCTCGTCAGGGCAAGTGA
	TcCPAP3-B	**F:** AGCAGGAAAGCAGGGATGAAGAGT **R:** AACCTCTTCGGAGGAACAGCCTTT
	Tc-CPAP3-C	**F:** ACGCGTGTTTATTGGTGCTCAGTG **R:** CAGCAATTCCAGAACACGTCGCAT
	TcCPAP3-C5a (exon-specific)	**F:** CGACTGCCGCAAGTACTACA **R:** AAAGCAGATACCCGGACCTT
	TcCPAP3-C5b (exon-specific)	**F:** CGACTGCCGCAAGTACTACA **R:** ATCAACGACAAGACGTGCAG
	TcCPAP3-D1	**F:** GCAGTGCGAGAATGGTTTGCTCTT **R:** CGACAGCTTCCGGATTGCAAACTT
	TcCPAP3-D2	**F:** GTGAACGGCCAACAAGAGTT **R:** GGCACAACACACCGTAGAGA
	TcCPAP3-E	**F:** TTAAGAATTGCGTGGATGGGCGTG **R:** CCGTCACAAGCACCAATCAGATCA

3.3.4. Double-stranded RNA synthesis and injection

Double-stranded RNAs (dsRNA) were synthesized from the PCR products using the Ampliscribe^{TM} T7-Flash^{TM} Kit (Epicentre Technologies, Madison, WI) as per the manufacturer's protocol. Unique regions with greatest sequence divergence for each of *CPAP1* and *CPAP3* genes were chosen as templates for synthesis of dsRNAs, whose nucleotide positions in the cDNA sequence and ORF are shown in Table 3-2. In most cases, RNAi experiments were replicated by choosing two non-overlapping regions. To determine the unique functions, if any, of the two isoforms of *TcCPAP3-C* by RNAi, dsRNAs corresponding to the entire sequence of exon 5a or exon 5b were synthesized (Table 3-1). Double-stranded RNA (200 ng per insect; in 0.1 mM potassium phosphate buffer containing 5 mM KCl, pH 7; (Tomoyasu and Denell, 2004) corresponding to each target-gene region was injected into a mixture of penultimate and last instars larvae and pharate pupae (n=40). The dsRNA for *T. castaneum* tryptophan oxygenase gene (ds*TcVer*) was used as a control.

To observe any effect of RNAi on fecundity, oviposition behavior or in the development of the resulting progeny, dsRNAs for each of the ten *CPAP1* genes and seven *CPAP3* genes were also injected into non-virgin adult females (n=20). Two days after injection, dsRNA-treated females were mated with an equal number of untreated males. These insects were maintained under normal conditions as described in (Arakane et al., 2008a). Eggs were collected every 3 d for ~ 1 month. Eggs were dechorionated by treating with 50% Clorox before taking images using a Leica MZFLIII microscope.

.

Table 3-2 Summary of properties of dsRNAs used for RNAi studies.

Gene	Nucleotide position in ORF	Length, bp
TcCPAP1-A		
1	559-732	174
TcCPAP1-B		
1	273-416	144
TcCPAP1-C		
1	907-1050	144
2	222-400	179
TcCPAP1-D		
1	49-377	329
TcCPAP1-E		
1	99-253	155
TcCPAP1-F		
1	17-182	166
TcCPAP1-G		
1	502-699	198
TcCPAP1-H		
1	2159-2472	314
2	146-335	190
TcCPAP1-I		
1	2003-2186	184
TcCPAP1-J		
1	138-324	187
2	3538-3733	196
TcCPAP3-A1		
1	11-114	104
2	700-1149	450
TcCPAP3-A2		
1	210-445	236
TcCPAP3-B		
1	74-501	428
2	505-772	268
TcCPAP3-C		
1	11-344	334
TcCPAP3-C5a		
1	663-949	286
TcCPAP3-C5b		
1	717-831	115
TcCPAP3-D1		
1	72-261	190
2	440-678	239
TcCPAP3-D2		
1	91-283	193
2	516-670	155
TcCPAP3-E		
1	377-620	244

3.3.5. Measurement of transcripts after RNAi for *CPAP* family of genes

RT-PCR experiments were carried out to monitor the effects of dsRNA on levels of the targeted transcripts by using gene-specific primers. Total RNA was isolated from whole insects' 4-d post-injection of dsRNA using RNeasy Mini kit (Qiagen, Valencia, CA)

Three insects were pooled for each RNA extraction. cDNA synthesis and RT-PCR were performed using gene-specific primers. The specificity of RNAi for *CPAP* genes was additionally analyzed by comparing the transcript levels of closely related *CPAP* genes. PCR amplification products using the same cDNA template and a pair of primers for the ribosomal protein S6 were used as internal loading controls.

3.3.6. Scanning electron microscopy (SEM) of elytra after RNAi for *TcCPAP3-D*

Co-injection of dsRNA for *TcCPAP3-D1* and *TcCPAP3-D2* genes (200 ng of each dsRNA per insect) was done at pharate pupal stage. Elytra were dissected from these beetles 10-d after adult eclosion. The elytra were fixed by passing through a series of washing and critical point drying. The elytral samples were coated with 60% gold/ 40% palladium mixture with a DESK II sputter/etch unit (Denton Vacuum, LLC, Moorestown, NJ). Images were taken using the S-3500N SEM equipped with the S-6542 Absorbed Electron Detector (Hitachi Science Systems, Ltd., Hitachinaka, Japan).

3.3.7. Chitin staining of elytra with FITC-CBD after RNAi for *CPAP* genes

Elytra were dissected from pharate adults collected at pupal-day 5 stage after injection of dsRNA at pharate pupae, specific for each of the following genes that show a pharate adult to adult molt arrest: *TcCPAP1-C*, *TcCPAP1-H*, *TcCPAP1-J*, *TcCPAP3-C*, *TcCHS-A* and *TcVer*. Elytra were washed in 1 X PBS, pH 8, heated in 10 M NaOH for 6 h at 95°C, to remove all the

proteins. The resulting solution was neutralized in 1 X PBS, pH 8, then stained with the chitin-binding domain from chitinase of *B. circulans* W12 tagged with fluorescein-isothiocyanate (FITC-CBD, New England BioLabs, Beverly, MA) at 1: 500 dilution in PBS, pH 8 at room temperature for 1 h. After washing off the excess FITC-CBD probe with PBS, the green fluorescence was recorded using GFP filter (470 nm excitation and 515 nm emission wavelengths).

3.3.8. Chitin content analysis of insects after RNAi for *CPAP1* and *CPAP3* family genes

Double stranded RNA for each of the *CPAP1* or *CPAP3* genes was injected into insects at the pharate pupal stage. The resulting insects were collected 5 d after molt at the pharate adult stage just prior to molting into adults. Individual pharate adults were collected in separate 1.5 ml screw-cap microcentrifuge tubes and weighed. The total chitin content of each of these insects was measured using modified Morgan-Elson method as described by (Arakane et al., 2005).

3.3.9. Confocal analysis of dsRNA *CPAP1*-treated insects

Insects injected with dsRNA for *TcCPAP1-H* and/or *TcCPAP1-C* as pharate pupae were collected on pupal day 5. The insects were fixed in 4% p-formaldehyde and passed through a series of sucrose gradients. Cryosectioning was performed as described previously (Klein et al., 1991) on using a Leica CM 1800 instrument. Twenty μm thick, cross-sections of pharate adult (pupal day-5) were stained for chitin with rhodamine-conjugated chitin-binding probe. All incubations for immunofluorescence were performed at room temperature. The sections were rinsed three times for 5 min in 0.2% PBST buffer, then incubated for 1 h in PBST buffer containing 2% BSA, and rinsed. The sections were incubated overnight with chitin stain (1:100)

dilution in 2% BSA. After this treatment the sections were rinsed again with PBST buffer as described above and nuclei were stained with 4', 6-diamidino-2-phenylindole-dihydrochloride (DAPI, 1 mg/ml in PBS buffer). After renewed washing with PBST buffer the sections were covered with Fluromount (Sigma) and sealed with clear nail polish to prevent dehydration. Slides were visualized by using a Zeiss Leica 510 confocal microscope. The instrument has a fully motorized stage, a Plan Apochromat objective (40 x/ 1.4 oil) and differential contrast interference (DIC) was used for taking images. Fluorescence emission imaging of 543 nm line of He-Ne laser was used to excite rhodamine-conjugated chitin binding probe, 488 nm line was used to excite Alexa 488 and 405 nm line for DAPI staining. Analysis of the images was done using the NIH Image J software.

3.3.10. Production of Recombinant TcCPAP1-C and TcCPAP1-H in *Escherichia coli*

For expression of the recombinant proteins, appropriate 3'-fragments of *TcCPAP1-C* and *TcCPAP1-H* cDNA clones were inserted into pProEx HTb vector (Novagen), so as to add an N-terminal 6X-His tag to the C-terminal regions of these proteins. The coding region to be expressed was obtained by PCR using primers mentioned in Table 3-3. The PCR product was digested with the restriction enzymes NcoI and SalI for CPAP1-C and NcoI and NotI for CPAP1-H. The PCR products were purified by agarose gel electrophoresis using the QIAquick ® Gel Extraction kit (Qiagen, Valencia, CA). The purified fragments were then cloned into NcoI/SalI digested pProEx HTb DNA or *CPAP1-C* and NcoI/Not I digested pProEx HTb vector DNA for *CPAP1-H*.

Table 3-3 Sequences of primers used for protein expression in *E. coli*

Gene	Primer sequence 5' to 3'
CPAP1-C-NcoI F	TAT- CCATGGCC CAACTTAACGACACGAAACC
CPAP1-C-SalI R	TAT - GTCGAC TCACTCTTTCTCGGATGG
CPAP1-H-NcoI F	TAT-CCATGGCC CAACCTGTGGGGGCCAAAG
CPAP1-H NotI R	TAT-GCGGCCGCTCA GAAACTCGTGTAGAGGGC

Underlined sequence: Restriction enzyme site

One Shot MAX Efficiency DH5-Alpha™- Competent *E.coli* cells (Invitrogen, Carlsbad, CA) were transformed with *CPAP1-C-pProExHTb* DNA for protein expression. A single colony was picked to inoculate a 5 mL culture in LB medium supplemented with Ampicillin (100 mg /mL). A culture grown overnight was used to inoculate 1 L of LB supplemented with the same antibiotics, which was then incubated at 37° C with shaking at 250 rpm. Upon reaching an OD_{600} of approximately 0.6, the protein expression was induced by the addition of IPTG (Isopropyl β-D-1-thiogalactopyranoside) at a final concentration of 1 mM, and the culture was incubated at different time point of 2 h, 4h, 6 h. Maximum expression was observed at 2 h post-induction after which the protein started to degrade. So the culture 2 h after induction was pelleted by cold centrifugation for 20 min at 1200 X g. The cell pellet was resuspended in lysis buffer (20 mM sodium phosphate buffer pH 8, containing 300 mM NaCl) and sonicated. After sonication, the cell lysate was cold-centrifuged for 20 min at 4000 X g. The protein was recovered in the soluble fraction and was subjected to Ni-NTA purification.

3.3.11. Production of Recombinant TcCPAP1-C and TcCPAP3-C in Hi5 cells

Full-length TcCPAP1-C and TcCPAP3-C cDNAs were amplified from a pharate pupal preparation. The cDNAs were used as templates and pair of primers were designed and used to amplify the full-length coding sequences for each encoded protein. Primers were synthesized at

IDT®Inc. (Coralville, IA). The primers contained restriction enzyme sites, and were used to clone the full-length coding region into the PVL1393 expression vector behind the polyhedron promoter. The reverse primers contained 6-His tag coding sequences. However the reverse designed by IDT for TcCPAP3-C lacked a single nucleotide that caused the His-tag to encode an isoleucine tag due to a frameshift. The PCR products and the PVL1393 vector DNA were cut with desired restriction enzymes at 37ºC for 2 h. Digested product was purified by agarose gel electrophoresis using the QIAquick ® Gel Extraction kit (Qiagen, Valencia, CA). The purified DNA fragments were then cloned into an XbaI/NotI digested PVL1393 for both TcCPAP1-C and TcCPAP3-C.

Sf21 (*Spodoptera frugiperda* cell line) insect cell line from Pharmingen (San Diego, CA) was used for co-transfection. Cells (4 X 10^5) were seeded in 6-well culture plates, according to manufacturer's protocol. The seeded plate was incubated at 27ºC for 5 d, after which the recombinant baculovirus was collected. Virus containing recombinant protein was amplified for 3-4 rounds to increase the virus titer.

Hi-5 cells (*Trichoplusia ni* cell line) were used for the expression of recombinant protein. The cells were incubated for 5 d at 27ºC. The Hi-5 cell medium was collected by centrifugation at 4000 rpm for 10 min. The expression level of protein was checked by SDS-PAGE, followed by Coomassie Brilliant Blue staining.

Table 3-4 Sequences of primers used for protein expression in Hi-5 cell line.

Gene	Primer sequence 5' to 3'
CPAP3-C-XbaI F	TCTAGAATGAAGTCCTACGCGTGTTTATTGGTGCTCAG
CPAP3C-NotI R	GCG GCC GCTTAATGATGATGATGATGATGATTATTATCG GCG GGTTCGGGGGCGTTG
CPAP1-C-XbaI F	TCTAGAATGTGGCGGCTTTTAGGATTTGGAGTC
CPAP1-C-NotI R	GCGGCCGCTCAATGATGATGATGATGATGCTC TTT CTC GGA TGG

Underlined sequence: Restriction enzyme site, Colored in red: Stop codon. Colored in green: start codon.

3.3.12. Purification of rTcCPAP1-C and r TcCPAP1-H by affinity chromatography

To purify the recombinant proteins rHis-CPAP1-C and rHis-CPAP1-H, containing the 6-His tag [$(His)_6$] at the C-terminal, Ni-NTA agarose chromatography was used. The soluble fraction was mixed with Ni-NTA agarose at 4°C for 1 h. Then the mixture of resin and medium was loaded onto a small column and washed with 50 ml of 300 mM NaCl. The proteins were eluted with an imidazole gradient from 20 to 250 mM. The fractions were checked for protein content by SDS-PAGE. The fractions containing recombinant proteins fractions were pooled and concentrated using a 10, 000 Da molecular weight cut-off filter (Amicon, Bedford, MA). The protein bands were cut out of the gel, trypsin digested and sent for mass spectroscopic analysis, ESI-MS/MS (Kansas State University Biotech core facility).

3.3.13. Mass Spectrometry analysis:

In-gel digestion: CBB stained gel pieces were incubated a few times for 10 min in 100 µL of 50% acetonitrile (ACN) at 30 °C. After destaining, the gel pieces were shrunk by addition of 50 µl of 100% ACN for 10 min and solvent was discarded. The gel pieces were dried by speed vacuum concentrator, then incubated with 20 µL of proteomics grade trypsin (Trypsin Gold, Promega, Madison, WI), 200 ng in 20 mM ammonium bicarbonate. Upon rehydration of the gels, an additional 20 µL of 20 mM ammonium bicarbonate and 10% ACN was added, and gel pieces were incubated at 30°C for 17 h. Tryptic peptides were recovered from gel plugs by extraction with 100 µL of 50% ACN in 2% trifluoroacetic acid (TFA) at 30 °C for 30 min.

Extracted peptides were concentrated by speed vacuum concentrator and added to 100 µL of 2% ACN in 0.1% formic acid.

Nano-HPLC and Electrospray Ionization Tandem Mass Spectrometry ESI−MS/MS):

Nano-HPLC was performed automatically using a microcolumn-switching device (Switchos; LC Packings) coupled to an autosampler (Famos: LC Packings) and a nanogradient generator (UltiMate Nano HPLC; LC Packings). Peptide solution (30 µL) was loaded on a C18 reversed-phase capillary column (75 µm ID × 15 cm, PepMap: Dionex) in conjunction with an Acclaim C18 PepMap trapping column (300 µm id×10 mm, Dionex). Peptides were separated by a nano-flow linear ACN gradient using buffer A (0.1% formic acid, 2% ACN) and buffer B (0.1% formic acid, 80% ACN) starting from 5% buffer B to 60% over 45 min at a flow rate of 200 nL/min. Then column was washed by 95% of buffer B for 5 min. The system control software, Hystar 3.2, was used to control the entire process. The eluted peptides were injected into an HCT Ultra Ion Trap Mass Spectrometer (Bruker Daltronics). The mass spectrometer was set up in the data dependent MS/MS mode to alternatively acquire full scans (*m*/*z* acquisition range from 300 to 1500 Da). The four most intense peaks in any full scan were selected as precursor ions and fragmented by collision energy. MS/MS spectra were interpreted and peak lists were generated by Data Analysis 3.4 and Biotools 3.0 software (Bruker Daltronics).

Bioinformatics: Peptide masses were compared to NCBI nr.2011. January using MASCOT 2.2 (http://www.matrixscience.com). The following parameters were used in all searches: the maximum number of missed cleavages allowed was 2; the mass tolerance was 1.2

Da for MS and 0.6 Da for MS/MS. Fixed modification was set on cysteine with carbamidomethylation. Variable modification was done on methionine with oxidation. Positive protein identifications using a threshold of 0.05 were used. Peptides scoring <20 were automatically rejected, ensuring that all protein identifications were based on reliable peptide identifications.

3.3.14. Production of antibodies for the recombinant proteins

The concentrated protein samples were run on single-well 4-12% Bis-Tris-acrylamide gel (Invitrogen, Carlsbad, CA). Bands with the correct size were cut and 0.5 mg of the proteins were sent to Cocalico Biologicals for immunization of rabbits and antibody production.

3.3.15. Immunoblot analysis

The eluted proteins were analyzed by SDS-PAGE on 4-12% Bis-Tris-Acrylamide gels, followed by immunodetection of His-tagged proteins to monitor the presence of rHis-CPAP1-C on the gel. SDS- PAGE was performed according to the standard protocol. Western blotting was performed using the Mini-Trans-Blot® electrophoretic transfer cell (Bio-Rad). Proteins were transferred from the gel onto an Immobilon^{TM}–P membrane (Millipore, Billerica, MA) at 100V for 1 h. Blocking was performed using TBS containing 0.05% Tween-20 and 2% BSA. For immunodetection, a mouse 6X-His monoclonal antibody (GenScript) was used at 1: 2000 dilutions, followed by incubation with a goat anti-mouse IgG conjugated to HRP used at 1: 500 dilutions. Colorimetric detection was achieved using the HRP/H_2O_2 substrate according to manufacturer's protocol (Bio-Rad).

3.4. Results

To determine the precise time for dsRNA injections in RNAi experiments, the expression profiles of *T. castaneum CPAP* genes were determined by RT-PCR using cDNA prepared from RNA isolated at different stages of development.

3.4.1. *CPAP1* genes exhibit divergence in expression profiles

In contrast to the general similarity in expression profiles of the *CPAP3* family of genes (see below), the individual genes of the *CPAP1* family have divergent expression patterns. The most notable difference is that they are expressed at very low levels, if at all, during the embryonic stage. *TcCPAP1-A* and *TcCPAP1-I* are expressed throughout the larval and pupal stages, with little or no expression in embryonic and adult stages. *TcCPAP1-B* and *TcCPAP1-C* have the broadest range of expression, spanning from young larval to mature adult stages. Others have more restricted ranges of expression. *TcCPAP1-D*, *TcCPAP1-F* and *TcCPAP1-G* are expressed in young larval and pupal stages. *TcCPAP1-E* and *TcCPAP1-H* are expressed predominantly in the pupal stage and do not appear to be expressed in earlier developmental stages or at the mature adult stage. *TcCPAP1-J* is expressed in larval and pupal stages but not in embryonic or adult stages (Figure 3-1)**.**

As most of the *CPAP1* genes show expression in the pharate pupal and pupal stages, expression patterns of *CPAP1* genes were also analyzed by RT-PCR from the pharate pupal stage through young adult stage. *TcCPAP1-A*, *TcCPAP1-B*, *TcCPAP1-F* and *TcCPAP1-G* and *TcCPAP1-H* transcripts are observed throughout the developmental stages analyzed. *TcCPAP1-C* is also expressed throughout, except for pupal day-2 when it has low or undetectable expression. *TcCPAP1-D* had low expression on pupal day 1-3. *TcCPAP1-E* and *TcCPAP1-J* show early pupal expression (from pupal day 0 to pupal day 3), but only at very low level.

TcCPAP1-I had remarkably low expression throughout except for detectable transcripts at pupal day-3 and pupal day-4 (Figure 3-1).

3.4.2. *CPAP3* genes have similar expression profiles throughout development

Many of the *CPAP3* genes were expressed throughout all stages of development, from embryonic to young adult at nearly the same levels with low or undetectable levels at the mature adult stage (Figure 3-1). However, *TcCPAP3-A1* and *TcCPAP3-A2* were expressed at low levels in mature larvae. The notable exception is *TcCPAP3-E* whose expression occurs mainly from the mature larval stage and continues through to the mature adult stage (Figure 3-1).

Expression analysis of the *CPAP3* genes through pharate pupa to young adult shows constitutively expression for all genes except *TcCPAP3-C5b* and *TcCPAP3-E*, which have low levels of transcripts on pupal day 1-2 and pupal day 2-3, respectively (Figure 3-1).

3.4.3. Spatial expression patterns of *CPAP1* and *CPAP3* genes during development

The tissue specificity of expression of *CPAP1* genes encoding proteins with ChtBD2s was assessed. Last instar larvae were dissected to obtain midgut, hindgut and carcass preparations because these tissues express two different chitin synthase (*CHS*) genes that contribute to the synthesis of chitin in either the PM or epidermal cuticle (Arakane et al., 2005; Arakane et al., 2008b). *CHS-B* is expressed in the midgut, while *CHS-A* is expressed in the hindgut and carcass. We anticipated that there might be corresponding differences in the expression profiles of the ChtBD2-containing proteins that bind to chitin present in different extracellular locations. Transcripts for each of these genes were detected using RT-PCR and gene-specific pairs of primers. As shown in Figure 3-2, there are substantial differences in expression between *CPAP1* and *CPAP3* families of genes encoding proteins with ChtBD2s as

well as among members of the same family. Almost all of the genes of the *CPAP3* and *CPAP1* families are expressed in the carcass but not in the midgut. Most of the genes expressed in the carcass are also expressed in the hindgut, consistent with the common ectodermal origin of hindgut and epidermis. The only exception is *TcCPAP3-E*, which has a low level of expression in the anterior and middle portions of the midgut in addition to high expression in the hindgut.

Expression in the midgut might possibly be due to tracheal contamination. The larval carcass was further separated into fat body and integument, and the corresponding RNA preparations were analyzed for transcripts for each gene. These results confirmed that *CPAP3* and *CPAP1* genes that are expressed in the carcass of larvae are also expressed in the integument, which includes the epidermis, muscle and tracheae but not the fat body. There were no detectable transcripts for these genes in the fat body with the exception of *TcCPAP3-A1*, which is expressed at a lower level in the fat body when compared to that in the integument. To further confirm that these genes are expressed in cuticle-forming tissues, we isolated RNA from elytra and hindwings at the pharate adult stage and detected specific transcripts by RT-PCR using gene-specific primer pairs. The details of expression of each member of the *CPAP* family of genes encoding ChtBD2s-containing proteins are described in Figure 3-2.

3.4.3.1. *CPAP1* family genes

The expression patterns of *CPAP1* family genes also show differences in tissue specificity. As expected, none of them are expressed in the larval midgut. *TcCPAP1-H* and *TcCPAP1-J* are expressed in the hindgut but not in the carcass. *TcCPAP1-A*, which is expressed predominantly in the carcass, has a low-level expression in the fat body. All *CPAP1* genes are expressed in the hindgut, but the relative expression in hindgut versus carcass varies among these

genes. Similarly, there are differences in the expression levels in elytra versus hindwings for individual *CPAP1* genes, even though all of them are expressed in both tissues (Figure 3-2).

3.4.3.2. CPAP3 family genes

There are subtle differences in the tissue specificity of expression of *TcCPAP3* genes. *TcCPAP3-E* is expressed at a higher level in the hindgut than in the carcass, whereas most of them are more highly expressed in the carcass than in the hindgut (Figure 3-2). Therefore, TcCPAP3-E may have a special role in hindgut cuticle. *TcCPAP3-A1* is expressed in the hindgut and carcass (Figure 3-2). Fractionation of carcass into integument and fat body revealed that none of the *CPAP3* genes except *TcCPAP-A1* are expressed in the fat body but presumably are expressed in cuticle-forming tissues such as the epidermis.

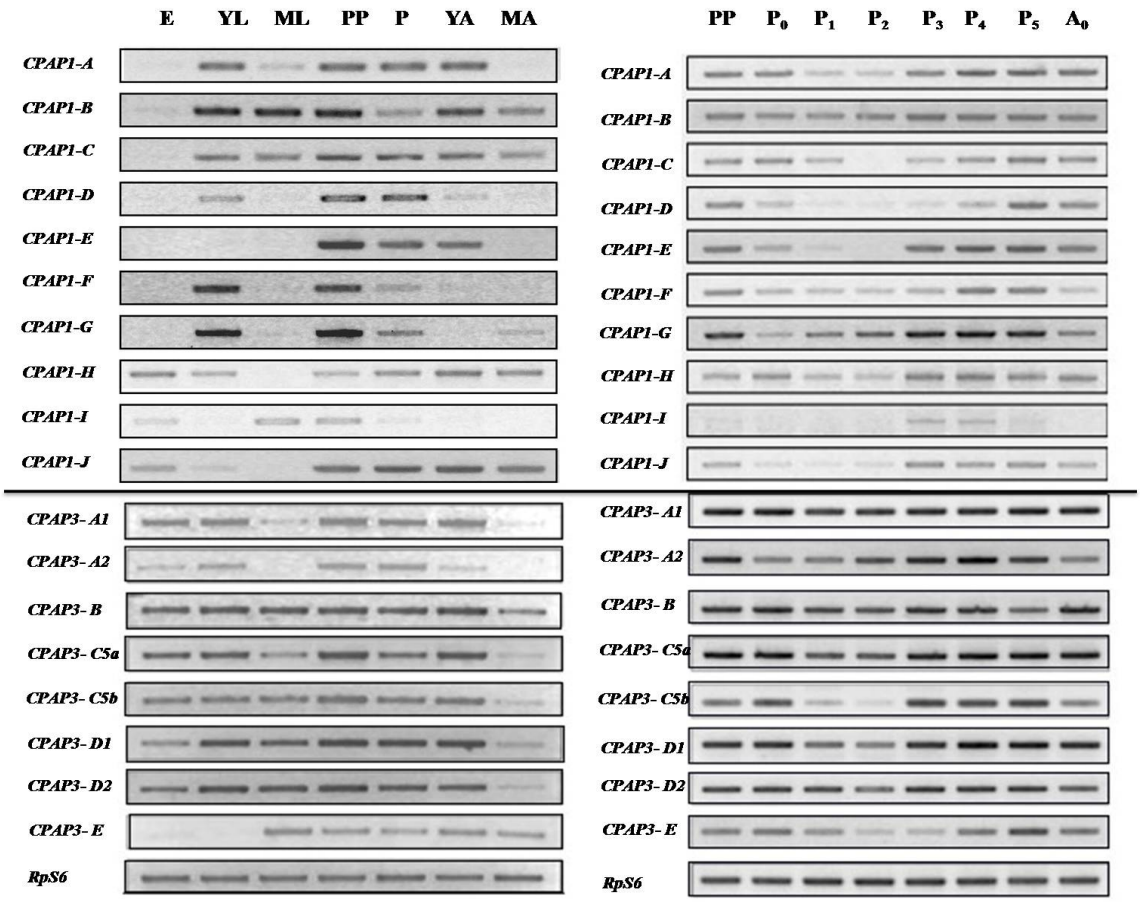


Figure 3-1 Expression profiles of individual members of *CPAP* gene families (*CPAP1* and *CPAP3*) as determined by RT-PCR.

Approximately 500 eggs (14.9 mg) were collected and used for RNA isolation. Pools of four larvae, pharate pupae, pupae from day-0 to day-5, young adult (~ 1-2 h old) or mature adults (> 2 weeks old) were used for preparation of total RNA. cDNAs synthesized from total RNA using oligo-$(dT)_{20}$ primers and reverse transcriptase were used as templates for RT-PCR (28 cycles) using gene-specific primers. RT-PCR products for the T. castaneum ribosomal protein S6 (RpS6) from the same cDNA templates served as an internal control for loading (24 cycles). Left panel: Different stages of development E, eggs; YL, penultimate instar or younger larvae; LL, last instar larvae; PP, pharate pupae; P, pupae; YA, young adult; MA, mature adult; Right panel: Different stages of pupal development: PP, prepupa day 1; $P_{[0-5]}$, 0-day old pupae to 5-day old pupae.

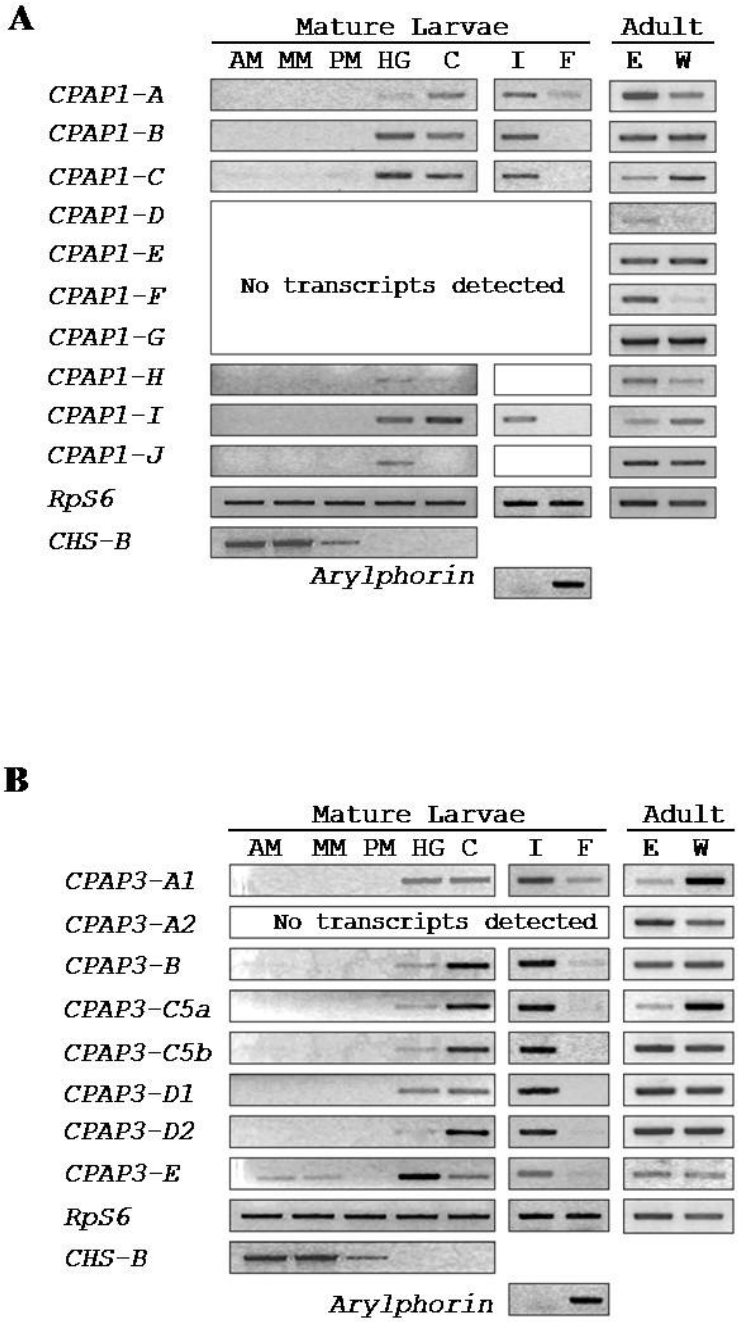


Figure 3-2 Tissue specificity of expression of *CPAP* family genes in *T. castaneum*.

Guts from actively feeding last instar larvae (n =15) were dissected out and divided into 3 parts, to obtain pools of anterior midgut, middle midgut, and posterior midgut and hindgut tissue preparations. A part of the remaining carcass (whole body without gut) was further dissected to obtain larval integument (carcass minus fatbody and muscle) and fatbody preparations. Pupae (4 d after pupation) were dissected to obtain elytron and hindwing preparations. These preparations contain adult tissues with attached remnants of pupal tissue. In addition, 15 actively feeding, 2-week-old adults were dissected to obtain midgut and hindgut preparations.

The remaining tissues were pooled and labeled as the carcass. RT-PCR products for the T. castaneum ribosomal protein 6 (TcRpS6) from the same cDNA templates served as an internal control for loading. Chitin synthase B (ChSB) RT-PCR product served as a check for the absence of midgut tissue in the carcass preparations, while those for arylphorin served to assess any contamination of the larval integument with fat body. AM: anterior midgut, MM: middle midgut, PM: posterior midgut, H: hindgut, C: carcass (whole body minus gut). I: integument, F: fatbody, E: elytron, W: hindwing. NTD = no transcript detected. Tissue-specific expression of CPAP3 family (panel A) and CPAP1 family (panel B) genes. The number of cycles for RT-PCR was 28 cycles except for RpS6 and arylphorin, both of which were amplified using 24 cycles.

3.4.4. *CPAP1* gene is required for pupal-adult molt.

To determine the roles of individual *CPAP* genes of these two families in insect development, we have carried out RNAi experiments mediated by dsRNAs specific for each gene. Our RNAi strategy was designed to down-regulate transcript levels for specific *CPAP* family genes just prior to the stage(s) of peak expression of the target gene and/or prior to each molt. dsRNAs for each of the 10 *CPAP1* genes were injected into groups of insects (n=40) at multiple stages of development namely, penultimate instar larval, pharate pupal or pupal stages depending on their developmental expression profile. The phenotypes of the insects subjected to RNAi as well as insects injected with the positive control dsRNA for *Vermilion* were monitored on a daily basis. Of the ten *CPAP1* family genes**,** *TcCPAP1-C*, *TcCPAP1-H* and*TcCPAP1-J* exhibited lethal phenotypes only at the pharate adult stage, no matter when the dsRNA was injected (larval, pharate pupal and pupal stages). The larval-larval and larval–pupal molts were normal. However, the insects died during adult eclosion. At the time of developmental arrest, incomplete cuticle slippage and abdominal contraction with a lack of elytral expansion were observed as shown in Figure 3-3. The dsRNA *Vermilion*-injected control insects developed into normal adults except for their lacking the dark eye pigment (Lorenzen et al., 2002).

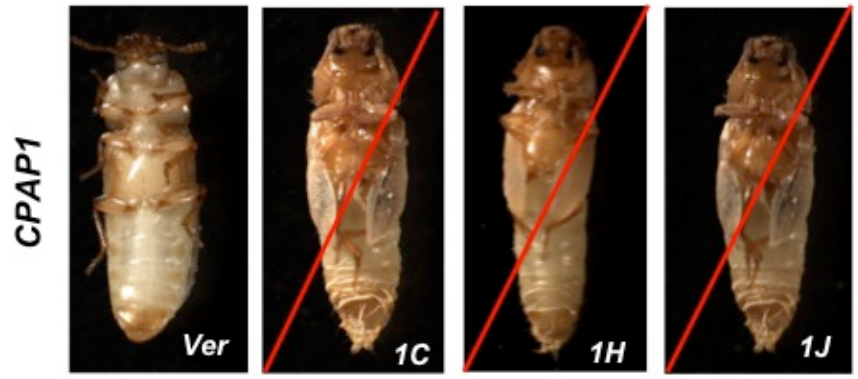


Figure 3-3 Effect of injections of dsRNA for *CPAP1* genes on pupal and adult development of *T. castaneum.*

Shown are lethal phenotypes produced by injection of dsRNAs for the indicated CPAP1 genes along with the Vermilion dsRNA-injected control. Gene-specific dsRNAs for CPAP1 genes were injected into penultimate instar larvae (200 ng per insect; n = 40). Shown are the terminal phenotypes of TcCPAP1-C, TcCPAP1-H and TcCPAP1-J. All three show arrest during the adult molt with 100% mortality.

3.4.5. RNAi for members of *CPAP3* gene family results in diverse phenotypes

Insects injected with dsRNA for ***TcCPAP3-A1*** (n=40) exhibited a lethal phenotype only at the mature adult stage, a week after adult eclosion, no matter when the dsRNA was injected (larval, pharate pupal or pupal stages). These insects were depleted of fat body and had a defecation problem only at the mature adult stage (Figure 3-5).

TcCPAP3-B: Injection of dsRNA (n=40) had no effect on pupation or adult eclosion or adult mortality for up to two months. However, adults had stiff joints and exhibited severe walking defects as indicated by a wobbly gait (see movie). Also 50% of eclosed adults also showed split elytra, which were rough and bumpy. The pronotum was also dimpled as shown in (Figure 3-6).

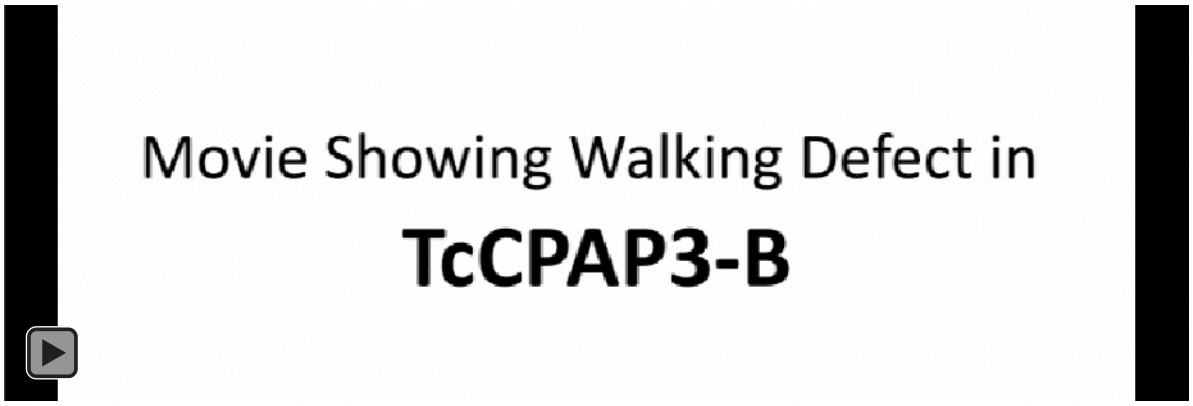


Figure 3-4 Movie showing walking defect in adults after dsRNA *TcCPAP3-B*-treatment of pharate pupae

dsRNA CPAP3-B injected insects showed a wobble gait, compared to dsRNA Ver (control). Their movement was slow and they had split wings, dimpled pronotum and rough elytra.

TcCPAP3-C: dsRNA was injected at various stages (larval, pupal or adult). The insects injected at each stage suffered developmental arrest at the adult molt. There was 100% lethality as none of the insects (n=40) shed their old pupal cuticle and were arrested only during the adult molt shown in Figure 3-6.

On the other hand, injections of dsRNA specific for *TcCPAP3-C5a* and *TcCPAP3-C5b* exons corresponding to the two alternatively spliced forms of the fifth exon of *TcCPAP3-C* gene (Jasrapuria et al., 2010) resulted in developmental arrest in only 5% and 10% of the insects, respectively. The moribund insects had a phenotype similar to the gene-specific, exon-non-specific phenotype observed in case of *TcCPAP3-C*. Ninety-five % of the injected animals became normal adults, indicating that these two alternatively spliced transcripts at least partially

compensated for each other (Figure 3-6).

***TcCPAP3-D1*and*TcCPAP3-D2*:** Neither dsRNA for *TcCPAP3-D1* (n=40) or dsRNA for *TcCPAP3-D2* (n=40) had any effect on larval-larval, larval-pupal or pupal–adult molts (Figure 3-6). However, there were visible differences in the appearance of elytra. The insects injected with dsRNA for *TcCPAP3-D1* failed to fully expand their elytra, which were shorter than normal, and the elytral surface was uneven and wrinkled (Figure 3-6). Co-injection of dsRNAs for the paralogous *TcCPAP3-D1* and *TcCPAP3-D2* genes also did not interfere with larval or pupal development. However, the elytra were more severely affected than the animals injected with dsRNA for either gene alone. The elytra were bumpy, rough and rumpled with creases. Also the pronotum of the insects was dimpled. The dsRNA *Ver*-injected controls developed normally. RT-PCR results to check transcript down-regulation after RNAi show significant knock-down of the *TcCPAP3-D1* and *TcCPAP3-D2* transcripts at the pharate adult stage compared with dsRNA *Ver*-treated controls.

Figure 3-5 RNAi of *TcCPAP3-A1* results in fat body depletion.

Left panel shows dsRNA TcCPAP3-A1 treatment results in adult fat body depletion compared to dsRNA Ver controls, shown by yellow arrows. Also the fecal pellet is attached due to defecation problem in these injected insects, marked with blue arrow.

Figure 3-6 RNAi of *CPAP3* genes.

Shown are lethal phenotypes produced by injection of dsRNAs for the indicated CPAP3 genes along with the Vermilion dsRNA-injected control. Gene-specific dsRNAs for CPAP3 genes were injected into each of the penultimate instar larvae (200 ng per insect; n = 40), pharate pupae (200 ng per insect; n = 40) and pupae (200 ng per insect; n = 40). Most of the CPAP3 family genes showed phenotypes (except TcCPAP3-A2 and TcCPAP3-E). TcCPAP3-A1: shows adult lethality and the resulting insects were unable to shed the fecal pellet. dsRNA CPAP3-B: adults show walking defects with stiff leg joints. dsRNA TcCPAP3CB: show arrest at the adult molt and 100% lethality. TcCPAP3-D1: show rough shortened elytra. dsRNA TcCPAP3-D2: shows a phenotype similar to dsRNA TcCPAP3-D1 but is less severe. dsRNA TcVer (200 ng per insect; n = 40) injected insects were used as controls. All of them had "white" eyes.

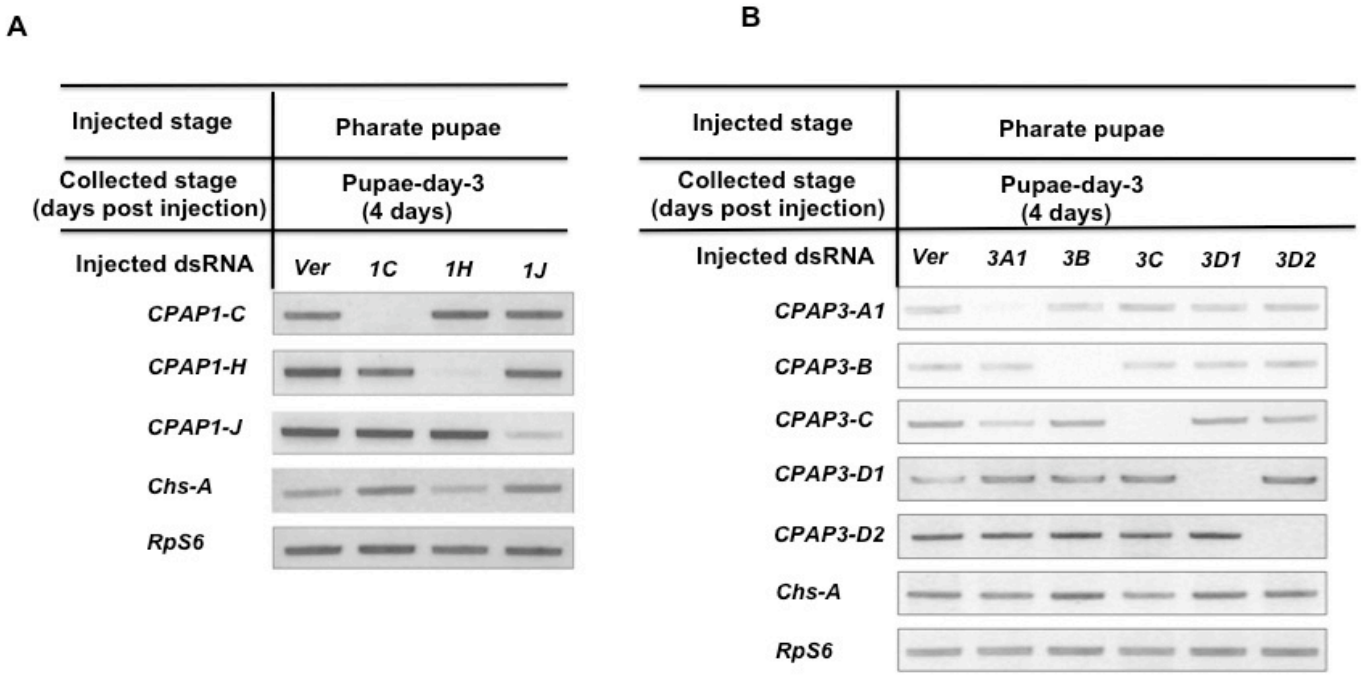


Figure 3-7 RT-PCR analyses to check *CPAP1* and *CPAP3* transcripts after dsRNA injections.

RT-PCR was carried out using RNA isolated from animals four days after dsRNA injections to monitor the extent of depletion of the transcripts of the targeted CPAP gene that showed a phenotype as well as some of the closely related gene(s). The RT-PCR results indicated specific reduction in the transcript level of the targeted CPAP gene with no evidence for depletion of transcripts for other genes of the same family.

3.4.6. Co-injection of dsRNAs for paralogous genes: *TcCPAP3-A1* and *TcCPAP3-A2*; *TcCPAP3-D1* and *TcCPAP3-D2*.

Co-injection of dsRNA for TcCPAP3-*A1* and dsRNA for *TcCPAP3*-A2 had an effect similar to injection of dsRNA for *TcCPAP3-A1* alone, with adult females dying one week after injection. Males were not tested. No effect was observed on larval-larval, larval-pupal and pupal-adult molting. Unlike its *Drosophila* ortholog, the *TcCPAP3-D* gene, is duplicated in *T. castaneum* and several other insect species. Co-injection of dsRNAs for TcCPAP3-*D1* and *TcCPAP3-D2* had a more severe effect than injection of either dsRNA alone. The elytra were very rough and bumpy, and failed to cover the entire length of the abdomen (Figure 3-8). A detailed description of the elytral structure is provided in section 3.4.9.

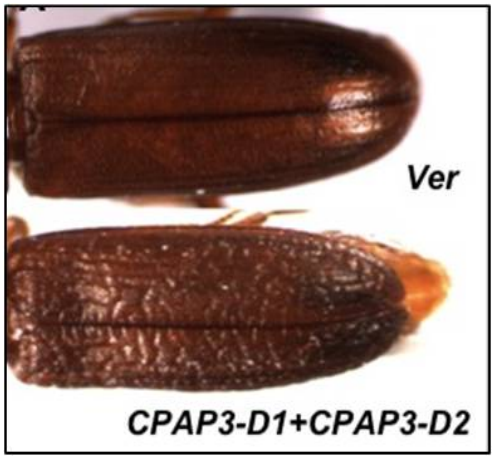


Figure 3-8 Co-injection of dsRNA for *TcCPAP3-D1* and *TcCPAP3-D2* affects elytral cuticle.

dsRNAs for TcCPAP3-D1 and TcCPAP3-D2 were co-injected into pharate pupae (n=40). The resulting insects could molt to otherwise normal adults with abnormal elytra. The elytra were wrinkled and alligator-like, and did not completely cover the abdomen. None of the insects died. dsRNA for TcVer was injected as internal control and the elytra from these insects were normal.

3.4.7. RNAi studies using adult females:

To study the roles of *CPAP* genes in embryonic and early instar larval development,

adult females (n=20, one month-old) were injected with 200 ng of dsRNA for each of the 10 *CPAP1* or 8 *CPAP3* genes. Adult mortality, fertility (as measured by number of eggs laid) and percent hatch of eggs were monitored. Injection of dsRNA for *TcCPAP1-J* into young adult females had no effect on the morphology of the ovaries (Figure 3-11). Injection of dsRNA for *TcCPAP3-D2* resulted in a drastic reduction in the number of eggs laid (Table 3-5). There was no observable effect on adult morphology or survival when dsRNA for *Ver* was injected as a control and the embryos developed normally except for lacking the black eye spot (Figure 3-10). The ovaries of *TcCPAP3-A1* and *TcCPAP3-D2* dsRNA-treated females were abnormal when compared to *TcVer* controls. dsRNA *TcCPAP3-A1*-treated females showed 100% mortality within two weeks of injection. A graph showing the survival of females is shown in Figure 3-12.

Statistical significance of the number of eggs collected: A two tailed t-test was performed on the number of eggs obtained following dsRNA- treatment for each gene through 6 batches. *TcCPAP3-A1* was not considered for this analysis as females were dying and no eggs were obtained after batch 2. The data is considered to be statistically significant if the null hypothesis is rejected. The null hypothesis is rejected if the p-value (probability of obtaining a test static) is less than 0.05. Number of eggs collected from *TcVer* injected insects is taken as control. A graph is plotted with the obtained p-value shows that the null- hypothesis is accepted and there is no significant differences in the number of eggs obtained, except for *TcCPAP3-D2* in which case the null hypothesis is rejected and the reduction in the number of eggs obtained is significant enough (Figure 3-9). The reason for reduction is not clear, however, the dissected ovaries are morphologically different compared to controls and has almost very little number of eggs, no fat body depletion and lethality of these insects was observed.

Table 3-5 Number of eggs laid after parental RNAi for *CPAP genes.*

Batch	*TcVer*	*TcCPAP1-A*	*TcCPAP1-B*	*TcCPAP1-C*	*TcCPAP1-D*	*TcCPAP1-E*	*TcCPAP1-F*	*TcCPAP1-G*	*TcCPAP1-H*	*TcCPAP1-I*	*TcCPAP1-J*
1	158	93	43	202	122	139	125	147	92	55	87
2	160	134	127	182	123	158	147	161	121	138	110
3	166	143	96	206	159	146	115	114	141	199	115
4	210	169	149	208	209	194	178	197	193	189	121
5	200	202	160	225	147	176	192	193	189	179	232
6	153	115	189	165	125	101	108	115	110	105	138

Batch	*TcVer*	*TcCPAP3-A1*	*TcCPAP3-A2*	*TcCPAP3-B*	*TcCPAP3-C*	*TcCPAP3-D1*	*TcCPAP3-D2*	*TcCPAP3-E*
1	199	120	138	234	177	229	21	205
2	218	9	238	221	233	201	49	212
3	224	females died	230	238	237	259	23	239
4	148		165	187	228	218	17	158
5	226		223	143	217	218	13	246
6	231		229	223	228	226	80	215
7	213		210	198	198	253	73	235

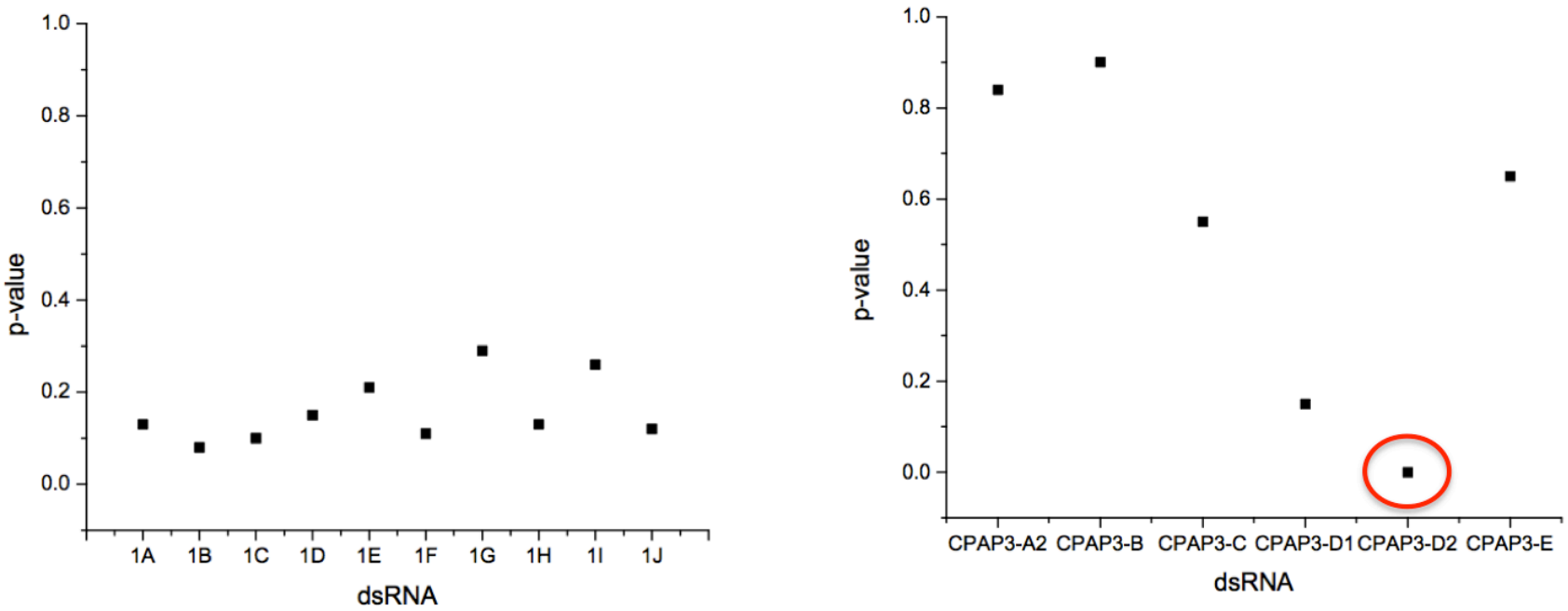


Figure 3-9 Statistical analysis of the number of eggs by t-test

The t-test statictical analysis shows that the reduction in the number of eggs for TcCPAP3-D2 treated insects is significant (red circle). The X-axis refers to the dsRNA treated insects and Y-axis shows the p-value. A p-value of less than 0.05 is considered significant, as it rejects the null hypothesis that the sample does not differ from the TcVer. 1A: TcCPAP1-A; 1B: TcCPAP1-B; 1C: TcCPAP1-C; 1D: TcCPAP1-D; 1E: TcCPAP1-E; 1F: TcCPAP1-F; 1G: TcCPAP1-G; 1H: TcCPAP1-H; 1I: TcCPAP1-I; 1J: TcCPAP1-J.

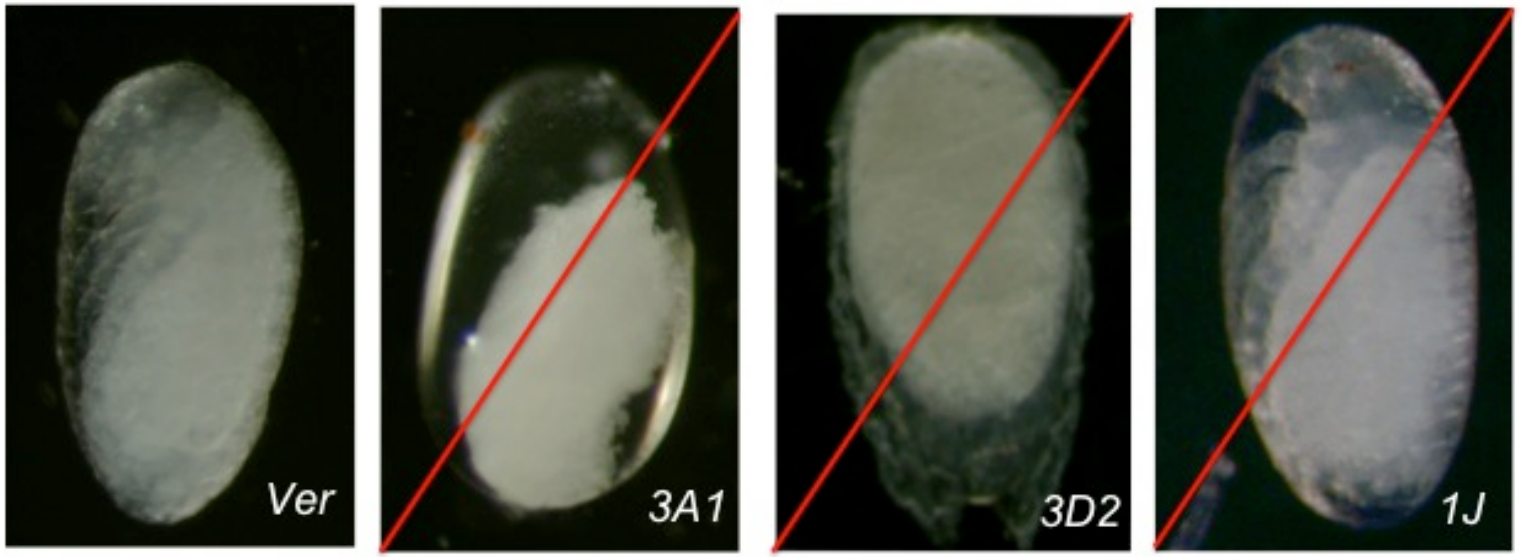


Figure 3-10 Observed RNAi phenotypes at embryonic stage.

dsRNA injected females for TcVer control laid normal eggs and the embryos had normal body segmentation and visible appendages just before hatching. TcCPAP3-A1: the body segmentation was not visible presumably due to failure of embryonic development as it was covered with yolk. dsRNA TcCPAP3-D2 embryos were also lacked body differentiation and the outside chorionic membrane was fragile and was detached on Clorox treatment. dsRNA for TcCPAP1-J treatment resulted in embryos with visible appendages and the red eye spot , but the head was not bent down like as in in TcVer control. Instead they held up their heads. All these embryos showed embryonic arrest and never hatched. Red slash line indicates lethal phenotypes.

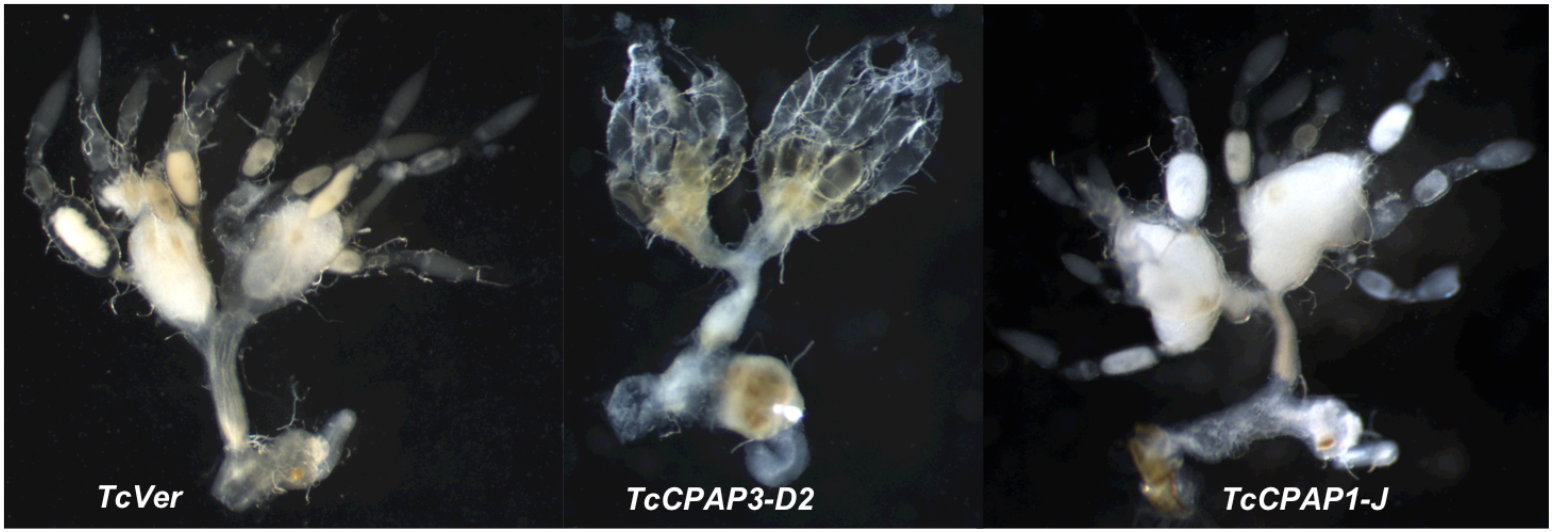


Figure 3-11 Dissected ovaries after RNAi of adult females

Ovaries of dsRNA TcCPAP3-A1-injected females lacked proper differentiation into ovarioles and did not contain eggs. In animals treated with dsRNA for TcCPAP3-D2 the ovaries had ovarioles but did not contain properly developed eggs. After TcCPAP1-J dsRNA-treatment, the adult females had normal ovaries.

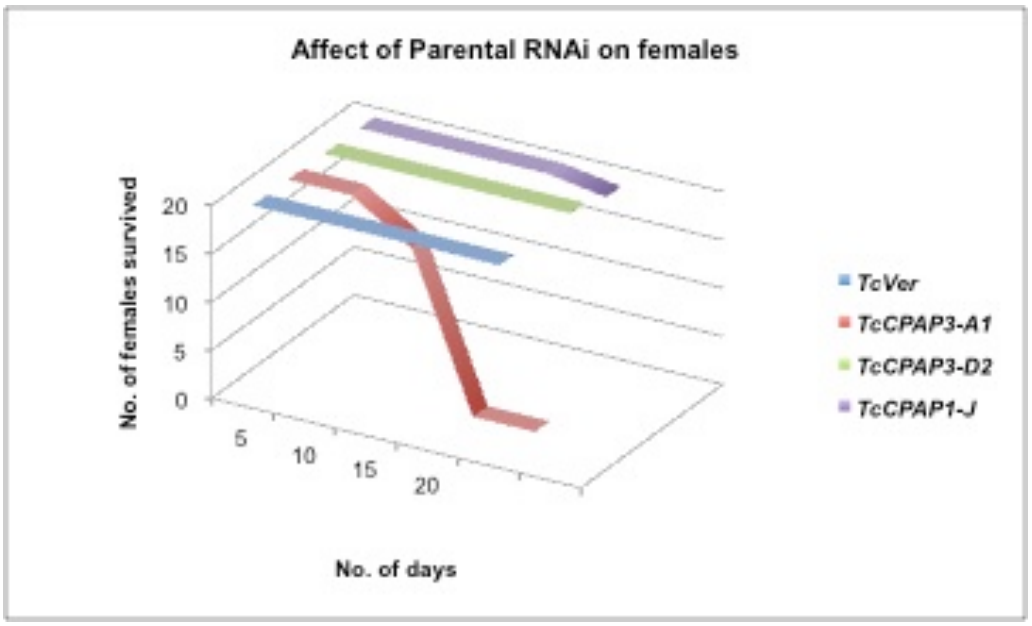


Figure 3-12 Number of females surviving parental RNAi.

The number of surviving females (n=20) is plotted on Y-axis and the number of days after dsRNA-treatment is plotted on the X-axis. All females died within 20 d after dsRNA TcCPAP3-A1 injection (red bar). TcCPAP3-D2 dsRNA-injected females showed normal survival rate (green bar). TcVer control (blue bar). TcCPAP1-J dsRNA-treated females also showed near-normal survival (only one death after 20 d).

3.4.8. No phenotypes observed after knockdown of other *CPAP* family genes

Injection of dsRNAs corresponding to *TcCPAP3-A2*, *TcCPAP3-E, TcCPAP1-A*, *TcCPAP1-B*, *TcCPAP1-D, Tc CPAP1-E, TcCPAP1-F, TcCPAP1-G and Tc CPAP1-I* at different stages of development (penultimate instar larvae, pharate pupae and pupae) did not result in any observable abnormalities (data not shown). Injection of dsRNA for the above genes into young adult females also had no observable effects on their survival, fecundity or egg hatch.

3.4.9. dsRNA for *TcCPAP3-D* affects elytral cuticle

Insects injected with dsRNA for *TcCPAP3-D1* and *TcCPAP3-D2* completed larval-larval, larval-pupal and pupal-adult molts, but the resulting adults had rough elytra. The structural defects on the adult elytra were more severe when the insects were co-injected with dsRNAs for *TcCPAP3-D1* and *TcCPAP3-D2* compared to animals injected with dsRNA for either gene alone. To probe the structural details of these defects, the elytra from a 10-d-old mature adults co-injected with dsRNAs for *TcCPAP3-D1* and *TcCPAP3-D2* at the pharate pupal stage were further analyzed by SEM. SEM analysis showed that the elytra of treated insects treated with dsRNAs for *TcCPAP3-D1* and *TcCPAP3-D2* were not properly expanded and had creases and folds on their dorsal surfaces. The ventral surfaces of elytra of the treated animals were very rough, and not compact showing fibrous structures (Figure 3-13).

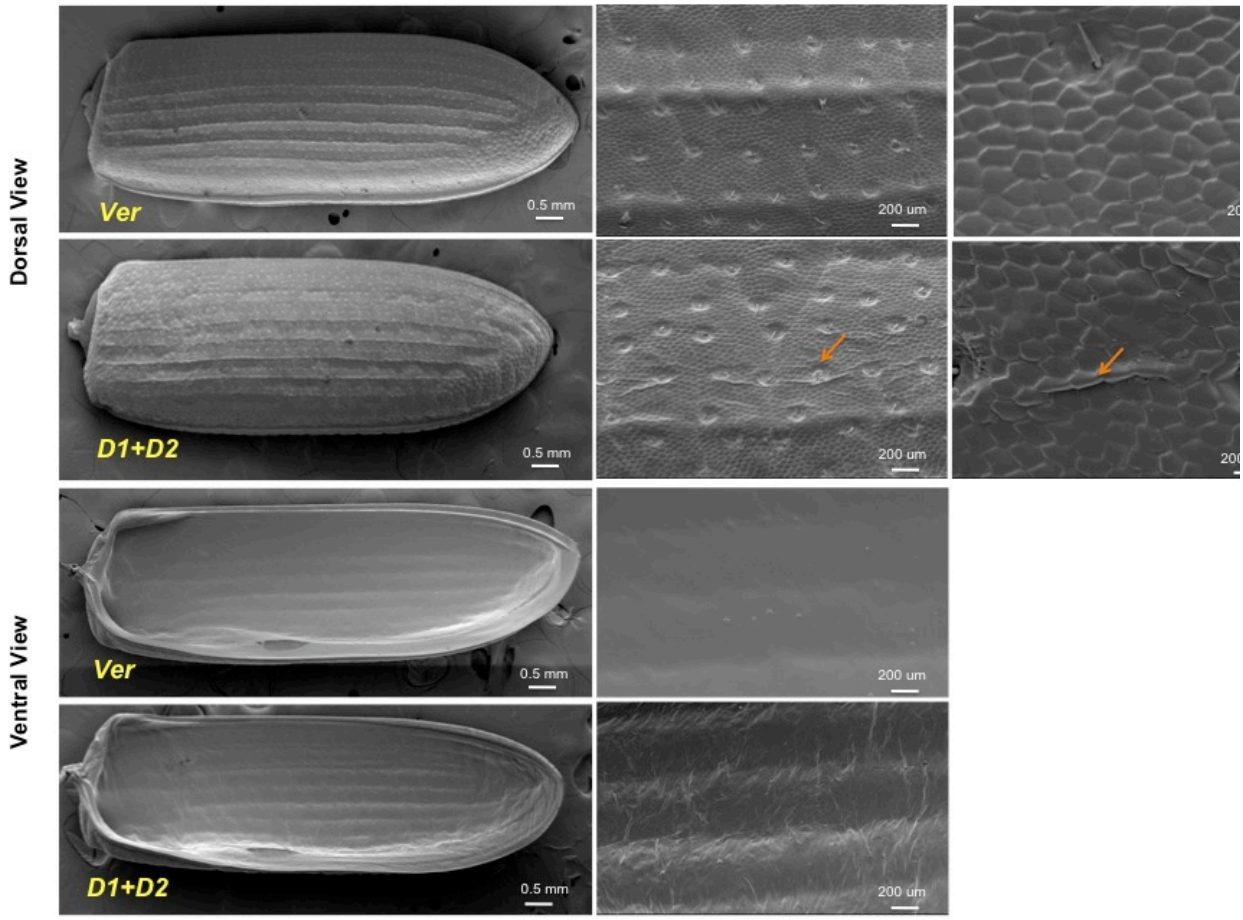


Figure 3-13 SEM analysis of elytra after co-injection of dsRNAs for *TcCPAP3-D1* and *TcCPAP3-D2.*

dsRNAs for TcCPAP3-D1 andTcCPAP3-D2 were co-injected (200 ng per insect; n = 40) into penultimate instar larvae. dsRNA for TcVer was injected as control. dsRNA TcCPAP3-D1 and TcCPAP3-D2 injected insects had rough elytra, but the effect was most severe after co-injection of dsRNAs for both TcCPAP3-D1 and TcCPAP3-D2.

3.4.10. Chitin content analysis

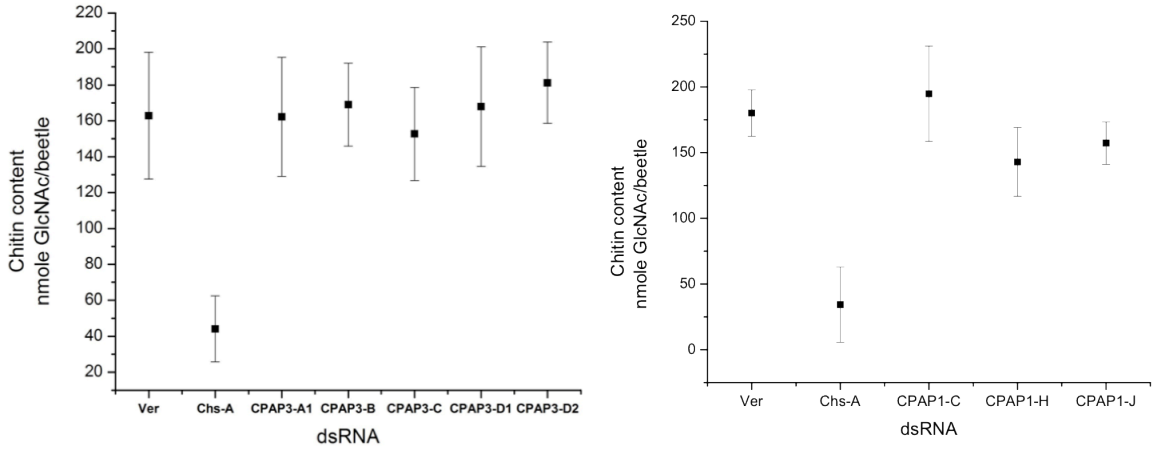
Biochemical analysis: To determine whether the *CPAP* family genes had any effect on chitin content either by promoting chitin synthesis or by protecting chitin from degradation by chitinases, we compared chitin levels in whole animals after RNAi for the *CPAP* family genes that produced observable morphological defects. To test this, chitin content of dsRNA-treated pupae collected 5 d after pupation was analyzed using a modified Morgan-Elson method (Arakane et al., 2005; Bolognesi et al., 2005; Reissig et al., 1955). When compared to the *dsRNA Ver*-treated control, there was a 45% reduction in the chitin content of insects after treatment with dsRNA *CPAP1-H*. This suggests that the reduction in chitin content in these animals might be responsible for the failure to molt and complete lethality at the pharate adult stage. However, there was no significant difference in the chitin content of 5-d old pupae following treatment with dsRNA for the other *CPAP1* genes including *TcCPAP1-C* and *TcCPAP1-J*. (the stage when molt arrest is observed). Biochemical analysis of chitin content for the *CPAP3* genes also did not show any reduction in the chitin content unlike *TcCPAP1-H* as shown in Figure 3-14.

FITC-CBD staining: To further confirm the reduction in chitin, the adult elytra were dissected from d-5 pupae and stained with fluorescein isothiocyanate (FITC)-conjugated chitin-binding protein from *Bacillus circulans* WL-12, after heating with 10 M NaOH at 95°C. Elytra of dsRNA *Ver*-injected insects were used as positive controls and those of dsRNA *Chs-A*-treated animals as a check for loss of chitin. Elytra of *dsRNA for TcCPAP1-H*-treated animals totally disintegrated after NaOH treatment and lost structural integrity, as did like the elytra following dsRNA *TcChs-A* treatment. Elytra of dsRNA *TcCPAP1-C* and dsRNA *TcCPAP1-J*-treated animals did not show any visible reduction

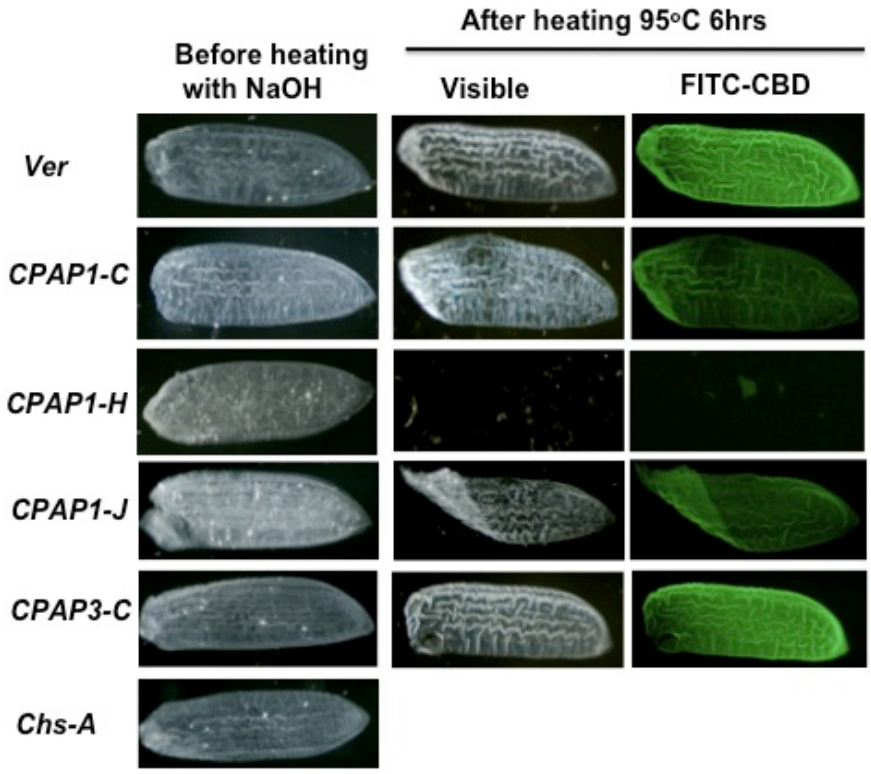
in chitin staining. However, their morphological structural features were altered when compared to control samples as shown in Figure 3-14.

Confocal microscopy: To localize the site of chitin depletion, confocal microscopy was performed on cryosections of dsRNA *Ver* and *dsRNA TcCPAP1-H*-treated insects after staining with rhodamine conjugated chitin-binding probe (chitin stains red). In comparison to the *Vermilion* control, in *dsRNA TcCPAP1-H*-treated insects, we observed a specific reduction in chitin staining in the new adult elytral cuticle, But, the old pupal cuticle did not show a significant change in staining (Figure 3-14).

A.



B.


After heating 95◦C 6hrs
Before heating with NaOH
Visible
FITC-CBD
Ver
CPAP1-C
CPAP1-H
CPAP1-J
CPAP3-C
Chs-A

C.

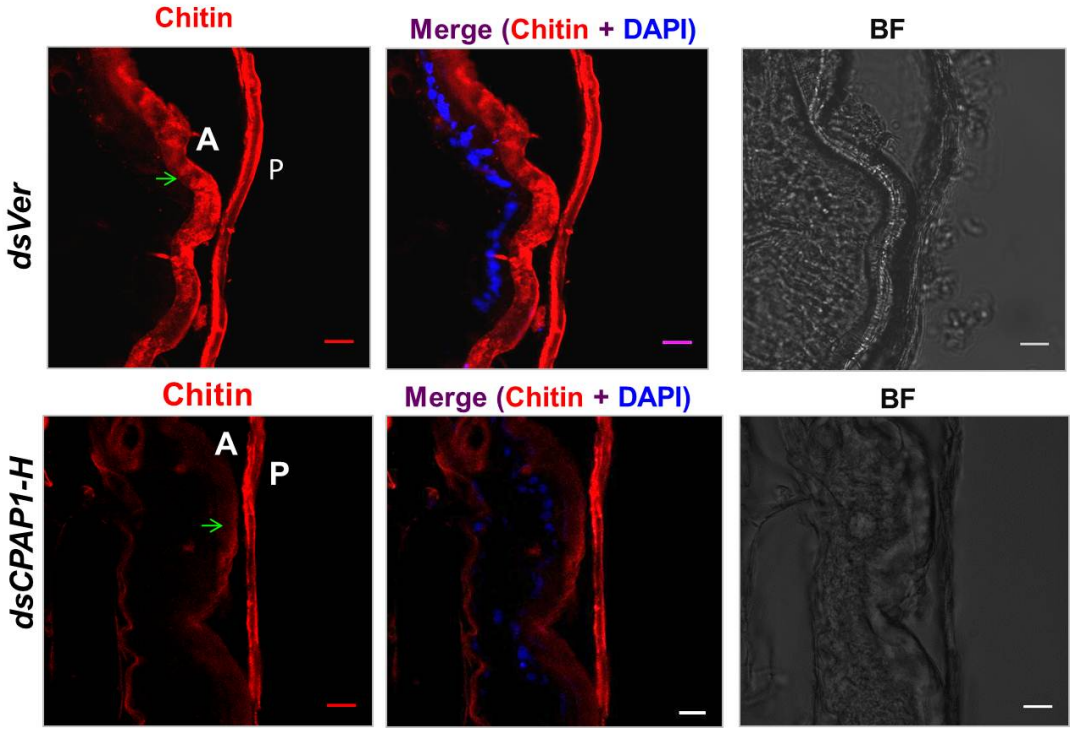
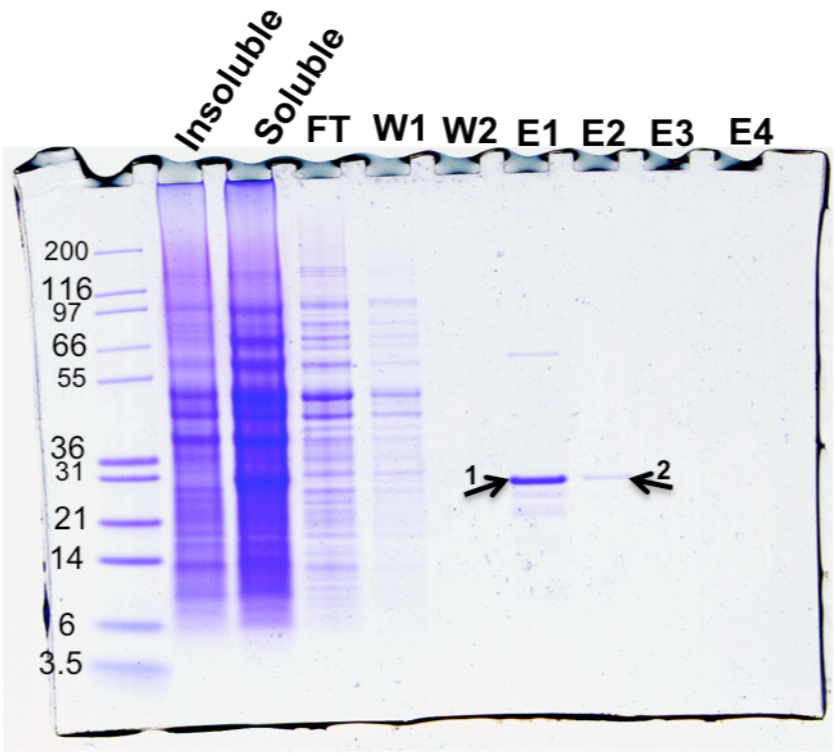

Chitin
Merge (Chitin + DAPI)
BF
dsVer
A
P
Chitin
Merge (Chitin + DAPI)
BF
dsCPAP1-H
A
P

Figure 3-14 Chitin levels after RNAi for *CPAP* genes.

*A: **Biochemical Analysis.** Chitin content of 5 d-old pupae was analyzed using a modified Morgan–Elson method. dsRNA Ver was injected as a negative control, and dsRNA TcChs-A-treated animals served as positive controls for depletion of chitin. When compared to control insects, the dsRNA TcCPAP1-H-treated animals had reduced chitin content. dsRNA TcCPAP1-C and dsRNA TcCPAP1-J injected animals had no chitin reduction. The dot represents mean and line represents error bar for each gene. GlcNAc =N-acetylglucosamine.*
*B: **FITC-CBD** staining of elytral cuticle after digestion with NaOH to remove cuticular proteins. Pharate pupae were treated with dsRNAs for TcCPAP1-C, TcCPAP1-H, TcCPAP1-J and TcChs-A. The elytra were dissected out from 5-d-old pupae. The elytra from dsRNA for CPAP1-H and Chs-A-treated animals were fragile and were dissolved by the treatment (10 M NaOH at 95°C for 6 h). In contrast, elytra from dsRNA TcCPAP1-C and TcCPAP1-J-treated insects lost their original shape after NaOH treatment, but showed normal staining by FITC-conjugated chitin-binding protein (FITC-CBD) when compared with elytra from dsRNA Ver-treated control insects. Elytra from dsRNA TcCPAP3-C-treated pupae did not show any effect.*
*C: **Confocal microscopic analysis.** Pharate adult cuticles from dsRNA TcCPAP1-H and dsRNA Ver–treated (control) insects were stained with rhodamine-conjugated chitin-binding probe (red). An apparent decrease in chitin staining of pharate adult elytral cuticle (green arrows) following RNAi for TcCPAP1-H compared to Ver (control) was observed. Scale bar = 10 μm. P: Pupal cuticle, A: Adult elytra. Plan Approchromat 40X oil.*

3.4.11. Expression and purification of C-terminal fragment of TcCPAP1-C protein in *E. coli* and mass spectroscopic analysis

The TcCPAP1-C and TcCPAP1-H proteins contain codons rarely used by *E. coli*. Therefore, they were produced in a bacterial expression system using the Rosetta 2 (DE3) strain, which is engineered to supply tRNAs for those rare codons. rHis-CPAP1-C protein was abundant in the soluble fraction and could be purified easily by passage through a Ni-NTA column. These proteins were used for antibody production.



1

Matched peptides shown in **Bold Red**

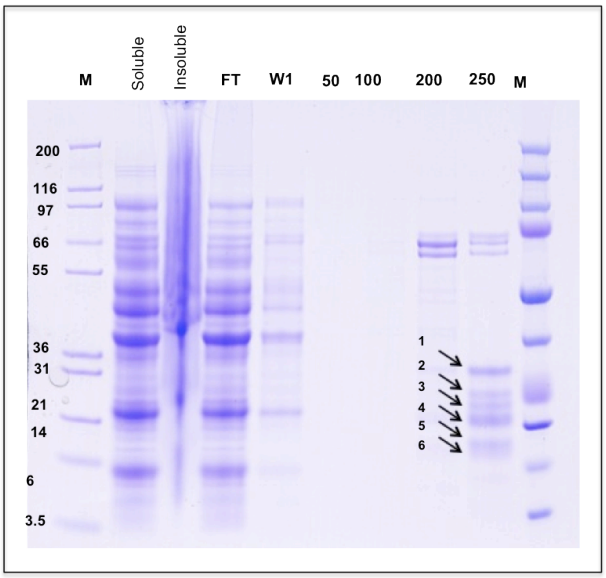
```
  1 QLNDTKPNEN LVVSTVNFES KFKATTPVYP KSVELTSEEP SEVGVLPPKD
 51 TNSTETDVVP PLSIDFEPPV VDSNFKPGDH SEEIPPELPS KFYQPPKFEP
101 DYTELAKLKG GFKALSSGEW EDLRKKFLIP DYEFPLETVS RPSYDSVLSS
151 FQVGPVPSEK E
```

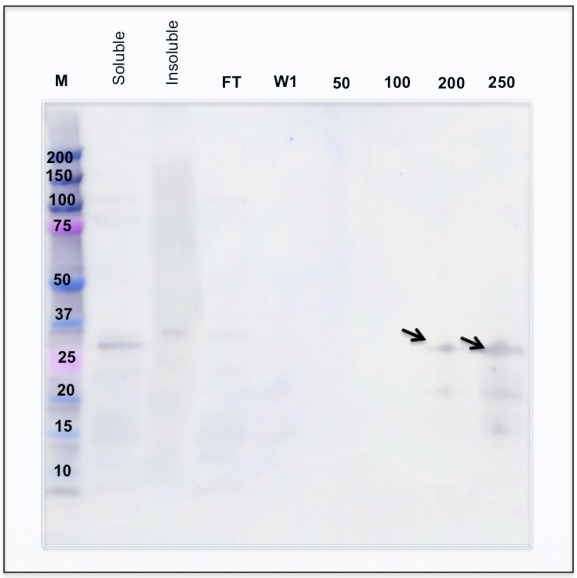
2

```
  1 QLNDTKPNEN LVVSTVNFES KFKATTPVYP KSVELTSEEP SEVGVLPPKD
 51 TNSTETDVVP PLSIDFEPPV VDSNFKPGDH SEEIPPELPS KFYQPPKFEP
101 DYTELAKLKG GFKALSSGEW EDLRKKFLIP DYEFPLETVS RPSYDSVLSS
151 FQVGPVPSEK E
```

Figure 3-15 SDS-PAGE analysis after Ni-NTA purification and mass-spec analysis of C-terminal fragment of TcCPAP1-C protein.

The C-terminal fragment of rHis-CPAP1-C protein was expressed in E. coli and the recombinant protein in the cell lysate was purified by Ni-NTA column. The predicted molecular weight of the protein is 27 kDa. Right panel shows the results of mass spectrometry analysis that confirms the identity of the purified band.





1

```
  1 MSYYHHHHHH DYDIPTTENL YFQGAMAQPV GAKAVEKKPE IKEEETSESA
 51 KAAFKPSVVI KTPLRQKYVN EPQEEFAPVS SRPQTPPRPS LKPSKEFLGP
101 RTTQKPKLPE PDPLDIEDEY DVTLNDALNP TLPNLPVRGF PTGFSNANNE
151 YNNYNSYQRP RYVLEGANSD YYQAKPTRQR YEAVPYSNPQ EHFVNSEADY
201 KGQYSTVNRA NFRPRLTQAL YTSF
```

2

```
  1 MSYYHHHHHH DYDIPTTENL YFQGAMAQPV GAKAVEKKPE IKEEETSESA
 51 KAAFKPSVVI KTPLRQKYVN EPQEEFAPVS SRPQTPPRPS LKPSKEFLGP
101 RTTQKPKLPE PDPLDIEDEY DVTLNDALNP TLPNLPVRGF PTGFSNANNE
151 YNNYNSYQRP RYVLEGANSD YYQAKPTRQR YEAVPYSNPQ EHFVNSEADY
201 KGQYSTVNRA NFRPRLTQAL YTSF
```

3

```
  1 MSYYHHHHHH DYDIPTTENL YFQGAMAQPV GAKAVEKKPE IKEEETSESA
 51 KAAFKPSVVI KTPLRQKYVN EPQEEFAPVS SRPQTPPRPS LKPSKEFLGP
101 RTTQKPKLPE PDPLDIEDEY DVTLNDALNP TLPNLPVRGF PTGFSNANNE
151 YNNYNSYQRP RYVLEGANSD YYQAKPTRQR YEAVPYSNPQ EHFVNSEADY
201 KGQYSTVNRA NFRPRLTQAL YTSF
```

4

```
  1 MSYYHHHHHH DYDIPTTENL YFQGAMAQPV GAKAVEKKPE IKEEETSESA
 51 KAAFKPSVVI KTPLRQKYVN EPQEEFAPVS SRPQTPPRPS LKPSKEFLGP
101 RTTQKPKLPE PDPLDIEDEY DVTLNDALNP TLPNLPVRGF PTGFSNANNE
151 YNNYNSYQRP RYVLEGANSD YYQAKPTRQR YEAVPYSNPQ EHFVNSEADY
201 KGQYSTVNRA NFRPRLTQAL YTSF
```

5

```
  1 MSYYHHHHHH DYDIPTTENL YFQGAMAQPV GAKAVEKKPE IKEEETSESA
 51 KAAFKPSVVI KTPLRQKYVN EPQEEFAPVS SRPQTPPRPS LKPSKEFLGP
101 RTTQKPKLPE PDPLDIEDEY DVTLNDALNP TLPNLPVRGF PTGFSNANNE
151 YNNYNSYQRP RYVLEGANSD YYQAKPTRQR YEAVPYSNPQ EHFVNSEADY
201 KGQYSTVNRA NFRPRLTQAL YTSF
```

6

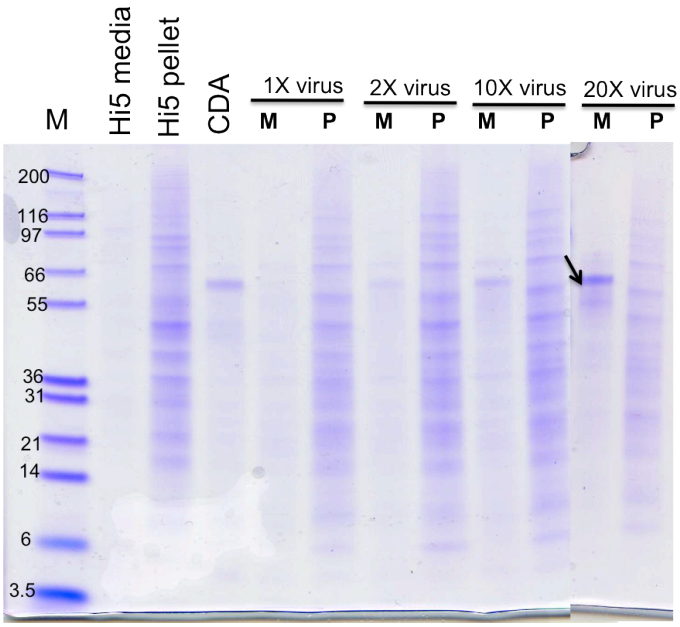
```
  1 MSYYHHHHHH DYDIPTTENL YFQGAMAQPV GAKAVEKKPE IKEEETSESA
 51 KAAFKPSVVI KTPLRQKYVN EPQEEFAPVS SRPQTPPRPS LKPSKEFLGP
101 RTTQKPKLPE PDPLDIEDEY DVTLNDALNP TLPNLPVRGF PTGFSNANNE
151 YNNYNSYQRP RYVLEGANSD YYQAKPTRQR YEAVPYSNPQ EHFVNSEADY
201 KGQYSTVNRA NFRPRLTQAL YTSF
```

Figure 3-16 SDS-PAGE analysis and immunoblotting after Ni-NTA purification and mass-spec analysis of C-terminal fragment of TcCPAP1-H.

Left panel shows the results of affinity purification by Ni-NTA column chromatography of the C-terminal fragment of rHis-CPAP1-H protein expressed in E. coli. The gel was stained with Coomassie Brilliant Blue. The right panel shows the western blot of the same using anti-His antibody (1: 2000) dilution. The gel bands marked [1-6] were analysed by mass-spectrometric analysis. The predicted size of the protein is 27 kDa. The arrow corresponds to the band number 1 on the left. The mass spectrometric analysis data for each of the six bands is shown at the bottom. The matched peptides are shown in red. Mass spectrometric data confirmed that the smaller bands were the degraded products of the protein of interest.

3.4.12. Expression and purification of full length TcCPAP1-C and TcCPAP-3C protein in Hi-5 insect cell line and mass spectrometric analysis.

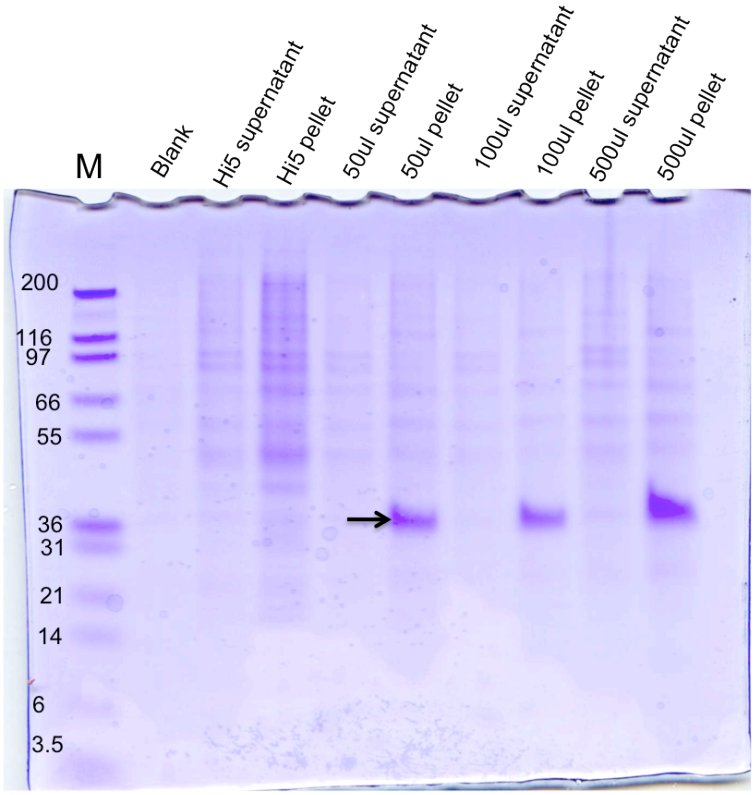
Monolayer cultures of Hi-5 cells were used as host cells to express the two proteins encoded in the recombinant baculoviruses. TcCPAP1-C has a signal peptide and was secreted into the medium. Protein expression was further confirmed by MS analysis. TcCPAP3-C is also a secretd protein and contains a signal peptide, but the frame shift at the C-terminus, led to the change of the His-tag (which was intended) to an Iso-Leu tag. This protein was never secreted into the medium.



Sequence Coverage: **4%**

Matched peptides shown in **Bold Red**

```
     1 MWRLLGFGVV LAAVAAPNNQ VPKTGFSCEG RTTGYYADVE SGCQVYHMCD
    51 GLGRQFSYTC PNATLFQQRM LICDHWYMVN CSKSVDDYTA NLRIGHKEMP
   101 FVDDNEANPY HRTPRPDLLS HPSQSEYDII YRTGRAQLGT NLNLVGAESD
   151 PKNSTTSTTD EPAYSLPSHW STEYNKQVTT TKPKPAKKGK PTTVRVNYQS
   201 NYKATTPVFP QSVEVTEAPD LELVPPSGST ESSVNFESRF KATTPVFPLS
   251 VDVTEAPNLG LLPPLPYQLN DTKPNENLVV STVNFESKFK ATTPVYPKSV
   301 ELTSEEPSEV GVLPPKDTNS TETDVVPPLS IDFEPPVVDS NFKPGDHSEE
   351 IPPELPSKFY QPPKFEPDYT ELAKLKGGFK ALSSGEWEDL RKKFLIPDYE
   401 FPLETVSRPS YDSVLSSFQV GPVPSEKEHH HHHH
```



Sequence Coverage: **62%**

Matched peptides shown in **Bold Red**

```
     1 MK[SYACLLVL SALIYSVLGQ ENFK]CPDDFG FYPHHTSCDK YWK[CDNNVAE
    51 LKTCGNGLAF DASDPK]FLTE NCDYIHNVDC GDRTQLEPPI SSPHCERLYG
   101 IFADESK[CDV FWNCWNGEAS RYQCSPGLAY DR]EARVCMWA DQVPECK[NEE
   151 VAGGFTCPAP GEVSNSGSFS R]HAHPDDCRK [YYICLEGTAR EYGCPIGTVF
   201 KIGDADGTGN CEDPEDVPGC EDYYGDLDLK] SIR[KSELLAG LQSSGSSR]SH
   251 QPATKSKPRP APASR[NAPEP ADNNIIIIII KRPLQI]
```

Figure 3-17 Expression of full-length recombinant protein CPAP1-C in Hi-5 insect cell line.

Expression of recombinant protein in the baculovirus-Hi-5 insect cell line system. Left panel shows the expression of full-length rHisCPAP1-C protein. The bandindicated by the black arrow was confirmed by mass-spectrocopic analysis to be the CPAP!-C protein. Right panel shows the expression of the rCPAP3-C protein. The identity of this band was also confirmed by mass spectroscopic analysis.

Immunoblot analysis of TcCPAP1-C.

3.4.13. Immunoblot analysis of TcCPAP1-C

The antibody raised using the C-terminal fragment of TcCPAP1-C expressed in the insect cell line was used for immunoblot analysis. The antibody detected the antigen used for immunization at the 10 ng protein level and also detected the full-length protein expressed in Hi-5 cell medium.

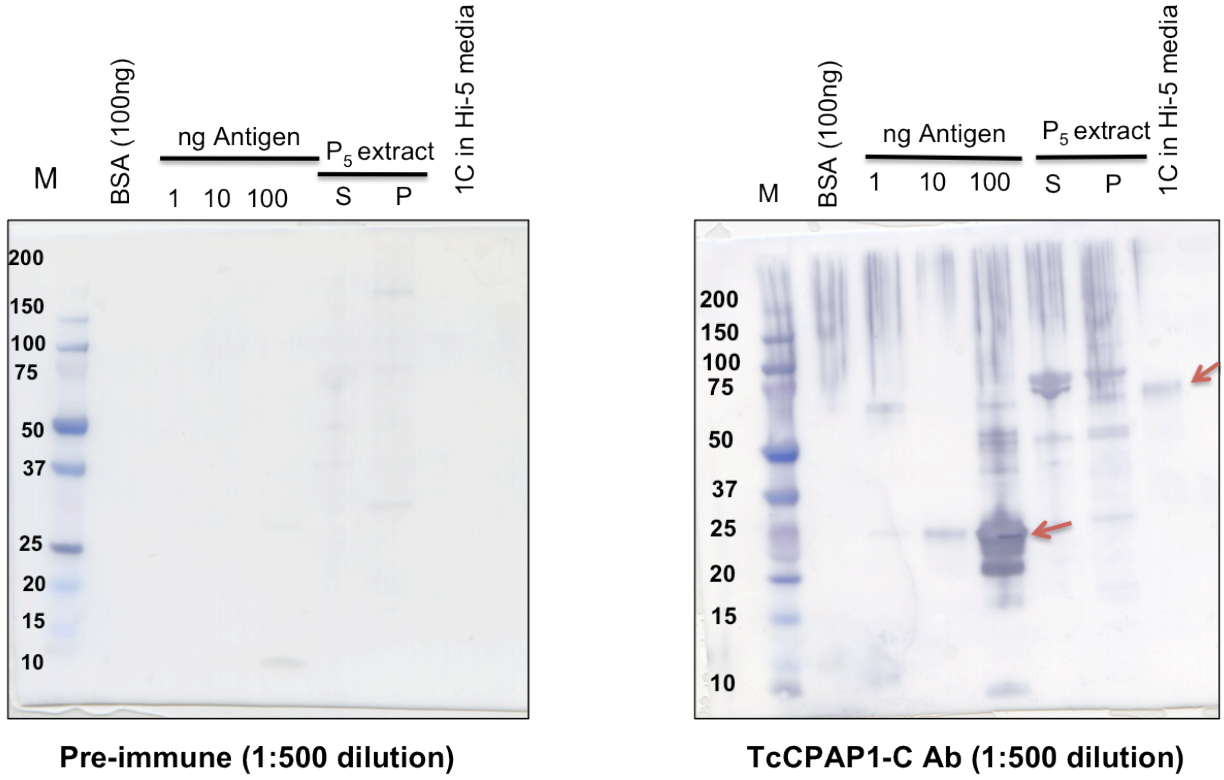


Figure 3-18. Immuno-blot analysis of TcCPAP1-C.

Left panel shows the western blot stained with pre-immune serum (1: 500 dilution) and right panel shows a duplicate blot stained with the C-terminal TcCPAP1-C antibody raised in a rabbit M: protein standard markers; BSA: bovine serum albumin 100 ng; antigen in different concentration of 1 ng, 10 ng, 100 ng; P5 extract fractions S and P were obtained by homogenization of untreated day-5 pupae with PBS. S: supernatant after 1X PBS treatment P: pellet fraction of above; 1C in Hi-5 media: medium from Hi-5 cells expressing the full-length. Red arrows indicate the protein of interest (full length in Hi-5 medium).

3.4.14. TcCPAP1-C protein is localized in the exocuticle

dsRNAVer treated insect were used for immunolocalization of TCPAP1-C in the insect cuticle. TcCPAP1-C co-localizes with chitin in the exocuticle (merged image yellow signal).. The protein was localized in the exocuticle (green signal), and was not seen throughout the underlying layers of the pupal procuticle. A similar preferential distribution of TcCPAP1-C in the exocuticle was also observed in the forming adult elytral cuticle at this stage. dsRNATcCPAP1-C injected insects were used as control for the specificity of the antibody and showed near complete loss of TcCPAP1-C protein following RNAi and no apparent reduction in chitin in the forming elytral cuticle.

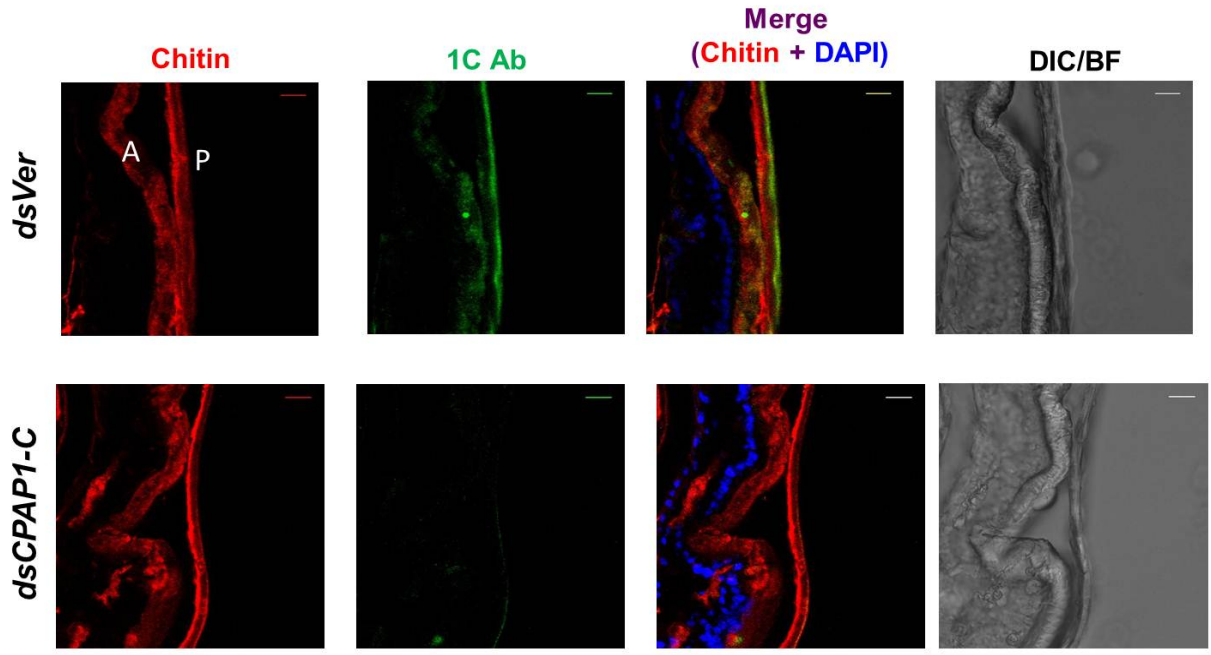


Figure 3-19 TcCPAP1-C protein is localized in exocuticle

Cross-cryosections (20 µm) of ventral abdominal segment of dsRNA TcCPAP1-C and dsRNA Ver-injected (control) animals were incubated with anti-CPAP1-C (dilution 1: 1000) antibodies that were detected by Alexa Fluor_488 conjugated anti-rabbit IgG antibodies (green). Chitin was stained with Rhodamine-conjugated chitin-binding probe (red). Nuclei were stained with DAPI (blue). CPAP1C-colocalizes with chitin in the pupal exocuticle. Scale bar = 10 µm. P: Pupal cuticle, A: Adult elytron. Plan Approchromat 40X oil.

3.5. Discussion

3.5.1. CPAP proteins are involved in cuticle organization

The main focus of this work has been the elucidation of the functions of CPAP1 and CPAP3 families of proteins. Several lines of evidence point to a role for CPAP1 and CPAP3 families of proteins in the assembly of epidermal cuticle and maintenance of its structural integrity. First, all of the CPAP proteins are predicted to have a cleavable leader peptide, do not have membrane-spanning helices and are expected to be secreted proteins consistent with a role in interactions with extracellular chitin. Second, *CPAP* family genes are expressed only in epidermal tissues, which also synthesize cuticular chitin, and notably not in gut tissue, which makes chitin needed for PM assembly. Conversely, *PMP* genes are expressed only in the gut tissue and not in epidermal tissue. Thus, the tissue specificity of expression of both *CPAP* families of genes also supports a role for these proteins in cuticle assembly, rather than in PM assembly. Thirdly, the CPAP1 and CPAP3 proteins are expressed at larval intermolt periods and during pupal stages when cuticle deposition and/or tissue remodeling are extensive. On the other hand, their expression levels decline in mature adult stages when new cuticle synthesis is essentially complete. Finally, CPAP1 and CPAP3 proteins have one and three peritrophin-A domains, respectively, which have been implicated in chitin binding. CPAP proteins are probably chitin-binding proteins (Hegedus et al., 2009; Wang et al., 2004; Wijffels et al., 2001).

. Among the different classes of chitin-binding proteins found in insects, the peritrophins and the CPAP proteins are unique in having the six-cysteine containing peritrophin-A motif originally described by (Tellam et al., 1999). Several peritrophins as well as enzymes with one or more peritrophin-A domains have been shown to bind to chitin (Hegedus et al., 2009; Wang and Granados, 2000; Wijffels et al., 2001). There is good evidence suggesting that

this motif is involved in chitin binding, because a short 8.1 kDa fragment from *M. sexta* chitinase containing this 58 amino acid-long motif can bind to chitin by itself as well as increase the affinity of proteins for chitin when this domain is engineered into them. This increase in affinity for chitin also appears to depend upon the number of CBDs added to these proteins (Arakane et al., 2003). Therefore, it is reasonable that CPAPs with one or three ChtBD2 domains may have an affinity for chitin, with the latter expected to bind to chitin more strongly. So far, the results of experiments designed to test the chitin-binding affinity of CPAP3 proteins have been mixed. (Gaines et al., 2003) reported that a CPAP3 protein, PL1, extracted from fleas failed to bind to chitin beads. On the other hand, (Nisole et al., 2010) expressed a CPAP3 protein from the spruce budworm in *E. coli*, extracted the recombinant protein from inclusion bodies with lauroyl sarcosine and tested the solubilized fractions for chitin binding. While a majority of the protein did not bind to chitin, a small fraction was retained by the chitin column and could be eluted only with hot SDS. These authors concluded that only a part of the recombinant protein might have been properly folded and exhibited the chitin-binding affinity. We have attempted to express some of the CPAP proteins using a baculovirus system, which we have used successfully to express large amounts of *M. sexta* chitinase with a peritrophin-A domain with good chitin-binding affinity(Arakane et al., 2003). Unfortunately, we have not been able to express these CPAP proteins in adequate amounts in Hi-5 cells to test for the chitin-binding properties directly. However, we have raised antibodies using unique parts of two CPAP proteins, TcCPAP1-H and TcCPAP1-C, lacking the peritrophin-A domain, and tested their ability to be detected by the original antigens used to raise them. We have also expressed some CPAP proteins in Hi-5 cells using recombinant baculoviruses. The availability of antibodies for these proteins may allow us to localize individual CPAP1 and CPAP3 proteins within insect tissues and cuticle.

3.5.2. Cuticles with different compositions of CPAP proteins display divergent physicochemical properties.

The physical and chemical properties of the PM are quite different from those of the epidermal cuticle even though both are primarily composed of chitin and proteins. While the PM is a highly hydrated extracellular matrix, the cuticle is essentially impermeable to water. The PM is porous and allows movement of enzymes and products of digestion through the pores whose sizes may vary along the length of the PM (Bolognesi et al., 2008). It has been proposed that cuticular chitin is in the α-form with anti-parallel chitin chains and that the PM-associated chitin is in the β-form with parallel chitin chains. Perhaps, the most interesting difference between these chitin-containing structures is the nature of the proteins associated with them. The PM-associated chitin binding proteins are mostly of the PMP family but also include some chitin-metabolizing enzymes and mucins. It is likely that the nature of the CBDs as well as cross-linking of these proteins to chitin influence the properties of the cuticle, which range from very flexible (e.g. the wing cuticle) to very hard (e.g. the head capsule and the elytron). These variations are likely to be the result of differences in the composition of the assortment of chitin-binding proteins, which may vary from tissue to tissue. It has been proposed that hard cuticle is associated with proteins with the RR2 motif, whereas soft cuticle is associated with proteins with RR1 motif (Willis, 2010). In our study we have found that the relative ratios of specific CPAP proteins in the elytron versus the wing vary substantially. For example, TcCPAP3-A1 is more abundant in the wing tissue, whereas the related TcCPAP-A2 protein is more abundant in the elytron. Similarly TcCPAP3-5A is more abundant in the wing tissue, but the protein product of the other alternatively spliced mRNA from the same gene, TcCPAP3-5b, is equally abundant in both tissues. In further support of the hypothesis that the presence/absence of specific CPAP

proteins influence the physiochemical properties of the cuticle is the finding that elytra isolated from some of the RNAi experiments described in this work (e.g. TcCPAP1-H) are fragile and deficient in chitin as is the case with elytra from animals depleted of chitin by treatment with dsRNA for *TcChs-A*. Elytra from animals with reduced expression of *TcCPAP1-J* and *TcCPAP1-C* were also fragile and had lost most of their structural integrity, while other proteins such as TcCPAP3-C did not exhibit such a phenotype even though all of them had nearly the same amount of chitin as control animals. Thus, many and perhaps all CPAP proteins play unique roles in constructing specific cuticles, which may determine the flexibility/rigidity of the structure.

3.5.3. *CPAP* genes have diverse roles in molting, locomotion and egg hatching

In this work we have taken advantage of the remarkably successful use of dsRNA-mediated transcript down-regulation in *T. castaneum* (Arakane et al., 2005; Tomoyasu and Denell, 2004; Zhu et al., 2008c) to investigate the role of individual *CPAP1* and *CPAP3* genes in this beetle. The presence of ten *CPAP1* genes and eight *CPAP3* genes in the *T. castaneum* genome complicates such analysis because of similarities in the nucleic acid sequences of the genes within the same family as well as between CPAP1 and CPAP3 families because of their shared peritrophin A domains. However, we were able to design gene-specific and alternatively spliced transcript-specific dsRNAs for our RNAi experiments by judicious choice of the regions in the individual genes from which these dsRNAs were generated. We verified that the RNAi was indeed gene-specific with no evidence of down-regulation of transcripts for even closely related genes or alternatively spliced mRNAs. These experiments have revealed that many of the *CPAP3* genes are indeed indispensable for the survival or maintenance of normal morphology and cuticular functions. Among the eight genes in the *CPAP3* group, RNAi for six

genes individually yielded detectable phenotypes ranging from molt arrest, elytral abnormalities, walking defects, reduction in fecundity, fat body depletion and reduced egg hatching. Some combinations of dsRNAs for recently duplicated genes result in synergistic effects. These results are summarized in Table 1.3. The finding that orthologs of all of these genes with the possible exception of *TcCPAP3-A2* are found in insect genomes of different orders (Jasrapuria et al., 2010) further strengthens our conclusion that each member of the CPAP3 family of proteins serves distinctly different and vital functions.

In contrast, we could identify only three genes belonging to the *CPAP1* family (*TcCPAP1-C, TcCPAP1-H* and *TcCPAP1-J*) that yielded visible phenotypes after transcript depletion. The phenotypes observed following RNAi of each of these three genes is molt arrest. In the case of *TcCPAP1-H,* there was also a reduction in chitin content. Confocal microscopy confirmed the loss of chitin predominantly in the newly formed adult cuticle (e.g. in the elytra). Among the *CPAP1* and *CPAP3* gene families, *TcCPAP1-H* is the only one that appears to affect chitin content. RNAi for the other seven *CPAP1* family genes did not result in any apparent morphological defects or insect mortality.

3.5.4. CPAP3 proteins may affect cuticle integrity by affecting content of cuticular chitin and/or its organization

The most common defect seen in the RNAi experiments is developmental arrest at the time of adult molting when there is extensive tissue remodeling and cuticle deposition. Rough elytra as well as defective walking might be secondary consequences of cuticular defects. We suspect that even though the chitin content of many cuticles are unaffected, the organization of the chitin into the characteristic laminar architecture may be affected in specific parts or all cuticles in the insect in many of these RNAi experiments. Loss of laminar architecture has been

demonstrated by depletion of the cuticle organizing protein, Knickkopf, in *D. melanogaster* (Moussian et al., 2007). This protein also protects chitin from degradation by chitinases (Chaudhari *et al.* unpublished data). Unlike, proteins studies so far, TcCPAP1-C protein is not uniformly distributed throughout the procuticle, but is localized in the exocuticle. This finding, suggests that this protein could contribute to the unique properties of the exocuticle such as rigidity and sclerotization. Whether TcCPAP1-H plays a similar role in protecting or maintaining chitin remains to be determined.

Knock-down of *TcPAP3-A1* is adult-lethal, so its effect on embryo development is not surprising. The other two genes, *TcCPAP1-J* and *TcCPAP3-D2*, affect embryonic development as well. Following parental RNAi for *TcCPAP1-J*, the females exhibit no reduction in fecundity and lay a normal number of eggs, but the eggs fail to hatch. This phenotype is similar to that reported following RNAi for *Chs-A,* which leads to a failure of egg hatch and a drastic reduction in chitin content of the embryos, most likely in the developing larval cuticle (Arakane et al., 2008b). *TcCPAP3-D2* genes appear to be involved in some early developmental events, because there is no evidence of embryonic germ bands in the animals subjected to RNAi for these genes.

In summary, the differences in the developmental expression profiles of individual *CPAP3* and *CPAP1* genes as well as the differences in expression in different cuticle-forming tissues (e.g. elytra *versus* wing) and the diverse phenotypes resulting from down-regulation of individual *CPAP* family genes suggest that these genes have distinct, non-redundant functions. This hypothesis is further supported by our finding that down-regulation of transcripts for individual genes and, in some cases, alternatively spliced mRNAs results in lethality and/or morphological defects, which, in most cases, cannot be compensated for by other members of the same family. Since all of these proteins are believed to bind chitin, they

probably interact with this matrix polymer and/or other proteins in distinctly different ways. One might speculate that specific CPAP proteins are involved at different levels of organization of chitin chains into bundles, laminae and stacked laminae. The presence of three ChtBD2 domains in the CPAP3 family of proteins makes them particularly attractive candidates for non-covalent and/or covalent interactions with chitin in nearby and adjacent chains. In addition, the CPAP proteins may interact with other cuticular proteins for higher degrees of order of matrix organization.

3.6. Bibliography

Arakane, Y., Muthukrishnan, S., Kramer, K.J., Specht, C.A., Tomoyasu, Y., Lorenzen, M.D., Kanost, M., Beeman, R.W., (2005). The *Tribolium* chitin synthase genes *TcCHS1* and *TcCHS2* are specialized for synthesis of epidermal cuticle and midgut peritrophic matrix. *Insect Mol Biol* 14, 453-463.

Arakane, Y., Specht, C.A., Kramer, K.J., Muthukrishnan, S., Beeman, R.W., (2008a). Chitin synthases are required for survival, fecundity and egg hatch in the red flour beetle, *Tribolium castaneum*. *Insect Biochem Mol Biol* 38, 959-962.

Arakane, Y., Zhu, Q., Matsumiya, M., Muthukrishnan, S., Kramer, K.J., (2003). Properties of catalytic, linker and chitin-binding domains of insect chitinase. *Insect Biochem Mol Biol* 33, 631-648.

Barry, M.K., Triplett, A.A., Christensen, A.C., (1999). A peritrophin-like protein expressed in the embryonic tracheae of *Drosophila melanogaster*. *Insect Biochem Mol Biol* 29, 319-327.

Beeman, R.W., Stuart, J.J., (1990). A gene for lindane + cyclodiene resistance in the red flour beetle (Coleoptera: Tenebrionidae). *J.Econ. Entomol* 83, 1745-1751.

Behr, M., Hoch, M., (2005). Identification of the novel evolutionary conserved obstructor multigene family in invertebrates. *FEBS Lett* 579, 6827-6833.

Bolognesi, R., Arakane, Y., Muthukrishnan, S., Kramer, K.J., Terra, W.R., Ferreira, C., (2005). Sequences of cDNAs and expression of genes encoding chitin synthase and chitinase in the midgut of *Spodoptera frugiperda*. *Insect Biochem Mol Biol* 35, 1249-1259.

Bolognesi, R., Terra, W.R., Ferreira, C., (2008). Peritrophic membrane role in enhancing digestive efficiency. Theoretical and experimental models. *J Insect Physiol* 54, 1413-1422.

Gaines, P.J., Walmsley, S.J., Wisnewski, N., (2003). Cloning and characterization of five cDNAs encoding peritrophin-A domains from the cat flea, *Ctenocephalides felis*. *Insect Biochem Mol Biol* 33, 1061-1073.

Hegedus, D., Erlandson, M., Gillott, C., Toprak, U., (2009). New insights into peritrophic matrix synthesis, architecture, and function. *Annu Rev Entomol* 54, 285-302.

Jasrapuria, S., Arakane, Y., Osman, G., Kramer, K.J., Beeman, R.W., Muthukrishnan, S., (2010). Genes encoding proteins with peritrophin A-type chitin-binding domains in *Tribolium castaneum* are grouped into three distinct families based on phylogeny, expression and function. *Insect Biochem Mol Biol* 40, 214-227.

Klein, U., Löffelmann, G., Wieczorek, H., (1991). The midgut as a model system for insect K+-transporting epithelia: immunocytochemical
localization of a vacuolar-type H+ pump. *J. Exp. Biol*. 161, 61-75.

Lorenzen, M.D., Berghammer, A.J., Brown, S.J., Denell, R.E., Klingler, M., Beeman, R.W., (2003). piggyBac-mediated germline transformation in the beetle *Tribolium castaneum*. *Insect Mol Biol* 12, 433-440.

Lorenzen, M.D., Brown, S.J., Denell, R.E., Beeman, R.W., (2002). Cloning and characterization of the Tribolium castaneum eye-color genes encoding tryptophan oxygenase and kynurenine 3-monooxygenase. *Genetics* 160, 225-234.

Moussian, B., Veerkamp, J., Muller, U., Schwarz, H., (2007). Assembly of the *Drosophila* larval exoskeleton requires controlled secretion and shaping of the apical plasma membrane. *Matrix Biol* 26, 337-347.

Nisole, A., Stewart, D., Bowman, S., Zhang, D., Krell, P.J., Doucet, D., Cusson, M., (2010). Cloning and characterization of a *Gasp* homolog from the spruce budworm, *Choristoneura fumiferana*, and its putative role in cuticle formation. *J Insect Physiol* 56, 1427-1435.

Reissig, J.L., Storminger, J.L., Leloir, L.F., (1955). A modified colorimetric method for the estimation of N-acetylamino sugars. *J Biol Chem* 217, 959-966.

Tellam, R.L., Wijffels, G., Willadsen, P., (1999). Peritrophic matrix proteins. *Insect Biochem Mol Biol* 29, 87-101.

Tomoyasu, Y., Denell, R.E., (2004). Larval RNAi in *Tribolium* (Coleoptera) for analyzing adult development. *Dev Genes Evol* 214, 575-578.

Wang, P., Granados, R.R., (2000). Calcofluor disrupts the midgut defense system in insects. *Insect Biochem Mol Biol* 30, 135-143.

Wang, W., Davison, S., Krell, P.J., (2004). Identification and characterization of a major early-transcribed gene of *Trichoplusia ni* single nucleocapsid nucleopolyhedrovirus using the baculovirus expression system. *Virus Genes* 29, 19-29.

Wijffels, G., Eisemann, C., Riding, G., Pearson, R., Jones, A., Willadsen, P., Tellam, R., (2001). A novel family of chitin-binding proteins from insect type 2 peritrophic matrix. cDNA sequences, chitin binding activity, and cellular localization. *J Biol Chem* 276, 15527-15536.

Willis, J.H., (2010). Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. *Insect Biochem Mol Biol* 40, 189-204.

Zhu, Q., Arakane, Y., Beeman, R.W., Kramer, K.J., Muthukrishnan, S., (2008c). Functional specialization among insect chitinase family genes revealed by RNA interference. *Proc Natl Acad Sci U S A* 105, 6650-6655.

CHAPTER 4

Peritrophic Matrix Proteins

4.1. Abstract

The biological functions of the 11 Peritrophic Matrix Proteins (PMPs) of the red-flour beetle, *T. castaneum*, were analyzed using gene-specific RNAi. Each of these proteins contains from one to 14 copies of the chitin-binding domain, ChtBD2, with six characteristically-spaced cysteines (Peritrophin A motif=CBM_14=Pfam 01607). These genes are expressed exclusively in the larval and adult gut tissues but are not expressed in tissues that are epidermal in origin. The transcripts for individual members of this family of genes are differentially expressed along the length of the midgut with the genes that contain a larger number of ChtBD2s being expressed in the more posterior parts of the midgut. Developmental expression profiles for PMPs indicate that transcripts for all of these genes accumulate in actively feeding larval and adult stages. Transcripts for some *PMP* genes are detected in the embryonic stage, but transcripts for none of the *PMP* genes are detectable in the pupal stages. Gene-specific transcript knock-down by RNAi shows that both *TcPMP3* and *TcPMP5-B* are required for both larval-pupal and pupal-adult molts. The resulting animals exhibit growth arrest and 100% mortality. Parental RNAi for *TcPMP3* causes adult female mortality. Following RNAi for *TcPMP5-B,* there is no reduction in number of eggs laid, but the eggs are morphologically defective and embryos fail to hatch. There is no effect on fecundity of adult females. Absence of visible phenotypes after specific knock-down of transcripts for other 9 *PMP* genes indicates that most of them have redundant or non-essential functions.

4.2. Introduction

The name “peritrophic membrane”, which means “surrounding the food” was coined by (Balbani, 1890). The term is still in use, but the word ‘membrane‘ is misleading because unlike true biological membranes, the peritrophic membrane is not made up of lipid bilayers. It is a complex, macroscopic, extracellular matrix primarily composed of chitin and proteins, and lines the entire midgut of insects. So the more general term “peritrophic matrix” (PM) is a more accurate description of this structure. It is a semi-permeable, non-cellular structure that divides the gut lumen into two compartments, the ectoperitrophic space and the endoperitrophic space. In insects peritrophic matrices are classified into two types based on their site of synthesis. Type 1 PMs are produced constitutively by midgut epithelial cells lining the entire length of the midgut. Type II PMs are produced by a ring of cardia cells (specialized cells at the anterior portion of the midgut) in response to a feeding stimulus. The type I PM has been extensively studied in dipteran adults and is found in Coleoptera, Dictyoptera, Hymenoptera, Odonata, Orthoptera, larval stages of Lepidoptera and adult hematophagous Diptera (Hegedus et al., 2009; Peters, 1992). The type II PM consists of laminar layers and is found in Isoptera, Embiodea, some Lepidoptera and the larvae of Diptera (Hegedus et al., 2009; Wigglesworth, 1930).

The PM is composed of mixture of chitin, proteins, glycoproteins and proteoglycans. Chitin is an integral part of the insect PM. The proteins that are tightly bound to chitin have been termed “peritrophins” (Tellam et al., 1999). Electron microscopic studies have revealed that the insect PM may have a lattice structure with regularly spaced grids between the crisscrossing chitin fibers (Hegedus et al., 2009; Hopkins and Harper, 2001). The PM is thought to be a porous and semi-permeable, allowing the movement of digestive enzymes and products of digestion to cross the PM barrier (but not the undigested food bolus) and the movement of

digestive enzymes produced by the midgut epithelial cells to the food retained within the PM. In mosquitoes, ingestion of a blood meal induces PM production (Peters, 1992). The PM may have a role in enhancement of the digestive processes and in protection of the gut from mechanical disruption as well as from attack by toxins and pathogens. It has been proposed that the peritrophins associated with the PM influence the permeability of the latter to dietary components, digestive enzymes and products of digestion either directly or indirectly through their association with chitin and other proteins (Hegedus et al., 2009; Tellam et al., 1999; Terra, 2001).

(Tellam et al., 1999) divided proteins associated with PM into four classes based on their ease of extraction from the PM. Class I proteins are loosely associated with the PM and can be easily removed upon treatment with physiological buffers. Class II proteins are the ones that can be removed by mild detergents that disrupt weak ionic interactions between the PM and bound proteins. Class III proteins are released from the PM by chaotropic solvents that include 6 M urea or 6 M guanidine HCl. Class IV proteins are those that are unextractable even by harsh treatment and presumably are covalently linked to other proteins and/or chitin. The class III PM proteins are collectively known as "peritrophins" or PMPs. At least some of the class IV proteins are likely to be related to peritrophins with great affinity to chitin, but this has not been clearly established.

The peritrophins have been divided into two sub-groups according to the presence or absence of "mucin"-like domains that are rich in proline, serine and threonine, which are potential sites of O-glycosylation (Dinglasan et al., 2009; Hegedus et al., 2009; Venancio et al., 2009). These proteins have been referred to as either "non-mucin peritrophins" (or simply "peritrophins") or "mucin-like peritrophins." Proteins with mucin domains are often highly

glycosylated both in mammals and in invertebrates, and they lubricate and protect surfaces of epithelial cells (Perez-Vilar and Hill, 1999). The mucins can be either membrane-bound or secreted. In insects, most, if not all, of the mucin-like peritrophins are secreted proteins because they associate with the PM, which is an acellular structure that is physically detached from the midgut epithelial cells, which secrete them. Rigorous data on glycosylation exists for only some peritrophins, while a vast majority of the proteins grouped together as "mucin-like peritrophins" have not been clearly shown to be glycosylated. It is likely that this distinction between the two classes of peritrophins may not be clear-cut and perhaps, not warranted.

While the PM is generally present in feeding stages of several orders of insects, there are several insect orders where no PM has been demonstrated (e.g. Hemiptera and adult Lepidoptera; reviewed in (Hegedus et al., 2009). Production of a PM is often initiated by the feeding stimulus. Some insects cease production of peritrophic matrix during periods of starvation and molts. Thus, there appears to be some regulation of the production of the PM and, by inference, the genes encoding the constituents of the PM including chitin and peritrophins. However, there are no detailed data on the expression of genes encoding peritrophins during development and in different tissues. In the past it had often been assumed that all genes with peritrophin-A domains are expressed in the gut. However, more recent studies have revealed that this notion is incorrect and have led to the discovery of genes encoding proteins with peritrophin-A domains that are expressed in cuticle-forming tissues (Barry et al., 1999; Behr and Hoch, 2005; Jasrapuria et al., 2010).

The total number of genes encoding peritrophins in different species of insects analyzed so far by bioinformatics searching of fully sequenced insect genomes varies widely (Hegedus et al., 2009). Further, such bioinformatics studies do not provide information on whether these

genes encoding proteins with one or more peritrophin-A domains are, in fact, PMPs as defined by their actual or potential association with the PM. In *T. castaneum,* there are more than 50 genes that encode proteins with one or more peritophin-A domains. Some of these are enzymes such as chitinases and chitin deacetylases, which have a peritrophin A domain in addition to their respective catalytic domains. In this chapter, I provide experimental evidence that only11 out of >50 of these genes capable of encoding proteins with peritrophin-A domains are *bona fide* peritrophin genes based on their tissue specificity and developmental pattern of expression. I further provide evidence that the expression of these ten peritrophin genes varies in different parts of midgut tissue. The results of RNA interference studies to analyze the functions of individual PMPs are also presented.

4.3. Materials and methods

4.4. Insect cultures

Strains used were *T. castaneum* GA-1 and *pu11*, an enhancer trap line in which the gene for green fluorescent protein (GFP) is expressed in the wing and elytral discs at the end of last larval instar (Lorenzen et al., 2003). In *T. castaneum,* the number of larval instars is indeterminate. We have used the *pu11* strain to ensure that the RNAi experiments are targeted at the penultimate (or earlier) larval instar stage. Insects were reared at 30°C in wheat flour containing 5% brewer's yeast under standard conditions as described previously (Beeman and Stuart, 1990).

4.4.1. Developmental expression profiles of the *T. castaneum PMP* gene families

The developmental expression patterns of the eleven *PMP* genes were analyzed by RT-PCR. Total RNA was isolated from whole insects at various stages of development including

embryos, young larvae (larval instars 4-5), last instar larvae, pharate pupae, pupae, young adults (± 3 h after adult eclosion) and mature adults (10 d-old) as well as pupae collected from pupal stages, day-0 to day-5 by using RNeasy Mini kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. First strand cDNA synthesis was carried out using Superscript III first-strand synthesis system (Invitrogen, Carlsbad, CA) using 1μg of total RNA for each reaction. This cDNA served as template for subsequent PCR reaction using the gene-specific primers listed in Table 4-1.

Table 4-1 Sequences of primers used for gene expression analysis.

F: Forward Primer; R: Reverse Primer

	Gene	Gene-specific primers 5'--3'
Peritrophic Matrix Proteins (PMPs)	*TcPMP1-A*	**F:** ATGAAAGCCGTTATTACTTTAGCACTC **R:** CTAACATTGAGAGCTGGTTTTACAACATTC
	TcPMP1-B	**F:** ATGAAGCAAACCGTGTTCTTAATTCTT **R:** TCAATGCAATAAATCTTTATTTTATACTTTTCTCG
	Tc PMP1-C	**F:** TCCAAATCATTT GTGTCCGACTG **R:** GCTCAGTAGGTTCTTCAGTTG
	Tc PMP2-A	**F:** ATGAAAGGTGTTGTAGTGTTTACAAT **R:** TTATTTGCACTGTGAAGGGTC
	TcPMP2-B	**F:** GAAACACCGTCAGTGCATTATG **R:** GTATTGAATGTAACTTGCTAGTAGTGTG
	TcPMP2-C	**F:** CGCACGAAAGCGACTGTAGCAAAT **R:** AGGTTTCCAGGTACTTGTGGGCTT
	TcPMP3	**F:** GACGACCATAGAAGATGAAAGTG **R:** TTAGACATCACATTCGCCAC
	Tc PMP5-A	**F:** TGTGCCCTCCTAGCATTGTTGAC T **R:** TTAGACATCACATTCGCCAC
	TcPMP5-B	**F:** GTCCTGAATGTCCCTCAGTAGATGGAG **R:** CCTTTACATCCTGCGGCGTTTG
	TcPMP9	**F:** CGTTACAACACACACTTAAAATGAC **R:** TTGGGGTTGTAGTCCCACAT
	TcPMP14	**F:** ACACTTTAACCCAGACTTGG **R:** TTAATTTGGCTCAACCAAGTGTTC

4.4.2. Profiles of tissue-specific gene expression during development

The RNeasy Mini kit (Qiagen, Valencia, CA) was used to isolate total RNA from mature larvae according to the manufacturer's instructions. The gut of the mature larvae was cut into 3 parts, anterior, middle and posterior, and RNA was isolated from separate pools of each of these segments. Total RNA samples were treated with RNase-free DNase I (Ambion, Austin, TX, 2U/μl) for 20 min at 37°C to remove genomic DNA contamination. The Superscript III first -strand synthesis system for RT-PCR (Invitrogen, Carlsbad, CA) was used to synthesize first-strand cDNA according to the manufacturer's instructions. Oligo-$(dT)_{20}$was used as a primer for reverse transcription using 1.5-2 μg of total RNA as template. RT-PCR was carried out to check the tissue-specificity of expression of each gene using pairs of gene-specific primers listed in Table 4-1.

4.4.3. Double-stranded RNA synthesis and injection

Double-stranded RNAs (dsRNA) were synthesized from the PCR products using the Ampliscribe^{TM} T7-Flash^{TM} Kit (Epicentre Technologies, Madison, WI) as per the manufacturer's protocol. For each *PMP* gene, unique region(s) with the greatest sequence divergence from other related *PMP* genes were chosen as templates for synthesis of dsRNAs. The primers for the first round of PCR had flanking T7 promoter sequences, which were then used as templates for *in vitro* transcription with T7 RNA polymerase. In most cases RNAi experiments were replicated by choosing two non-overlapping regions for each gene. The positions targeted for dsRNA synthesis for each of these genes are shown in Table 4-2 (all nucleotides are numbered from the translation start site). Double-stranded RNA (200 ng per insect in 0.1 mM potassium phosphate buffer containing 5 mM KCl, pH 7) corresponding to each target-gene region was injected into a mixture of penultimate and last instar larvae and pharate pupae (n=40); (Tomoyasu and Denell,

2004). The dsRNA for the *T. castaneum tryptophan oxygenase* gene (ds*TcVer*) was used as a control for monitoring non-specific effects of dsRNA injection and for following the effectiveness of the RNAi (Lorenzen et al., 2003). After injection, insects were kept at 30°C for visual monitoring of phenotypes and other analyses.

To observe any effect of RNAi on fecundity, oviposition behavior and ovarian morphology or changes in the development of the resulting progeny, dsRNAs for each of the eleven *PMP* genes were also injected into adult females (n=20). Two days after injection, dsRNA-treated females were mated with an equal number of untreated males. These insects were maintained under normal conditions as described in (Arakane et al., 2008b). Batches of eggs were collected every 3 d for ~ 1 month.

Table 4-2 Summary of properties of dsRNA used for RNAi studies.

Gene/dsRNA #	Nucleotide position in ORF	Length, bp
TcPMP1-A		
1	299-501	203
TcPMP1-B		
1	351-550	200
TcPMP1-C		
1	880-293	587
TcPMP2-A		
1	1-507	507
TcPMP2-B		
1	127-356	229
2	357-561	204
TcPMP2-C		
1	274-472	199
TcPMP3		
1	901-1223	323
2	1360-1617	257
TcPMP5-A		
1	268-576	308
TcPMP5-B		
1	227-601	375
2	1050-1549	500
TcPMP9		
1	(-20)-549	569
2	1939-2128	190
TcPMP14		
1	1-250	250
2	3678-3920	243

4.4.4. *In situ* hybridization

In situ hybridization was performed to study the site of expression of the two gut-specific genes *TcPMP9* and *TcPMP14*. Microtome section of guts of last instar larvae of *T. castaneum* were subjected to *in situ* hybridization with complementary DNA probes specific for *TcPMP9* and *TcPMP14*. Last instar larval guts were dissected in phosphate-buffered saline (PBS) and fixed in Bouin's fixative (saturated picric acid: formaldehyde: glacial acetic acid) at 4^{o}C overnight or Carnoy's fixative (ethyl alcohol: chloroform: acetic acid = 6: 3: 1 V) at room temperature for 15 min. Tissues were gradually dehydrated through an ascending series of ethanol/chloroform gradients (70%, 90%, 100% of ethyl alcohol in chloroform) and embedded in parablast at 60^{0}C for 1 h. Serial sections of 7-10 µm were made, deparaffinized in xylene and rehydrated in an ethanol series (100%, 90%, 70% ethyl alcohol). The tissues were washed 3 times in PBST (PBS and 0.2% Triton-X 100) for 10 min at room temperature (RT). The samples were then treated with proteinase K (10 µg/mL) for 10 min to partially digest the tissues and improve the accessibility of probes into the cells. The reaction was stopped by incubation with PBST-glycine (2 µg/ml) for 5 min followed by two washes with PBST. The tissues were refixed in 4% paraformaldehyde for 15 min at RT followed by washing with PBST. Pre-hybridization was performed without the probe in hybridization solution (50% formamide, 5 X SSC, 50 µg/ml heparin, 0.1% Tween 20 and 100 µg/ml salmon sperm DNA) at 48^{o}C for 20-30 min. Tissues were hybridized with digoxigenin (DIG)-labeled or/and biotin labeled single-stranded sense or antisense DNA probes in hybridization solution at 48^{o}C for 20-30 h. Sense DNA probe was used as the negative control. The DIG-labeled DNA probes were prepared with the PCR DIG Probe Synthesis Kit (Roche Molecular Biochemicals, Mannheim, Germany) and biotin-labeled DNA

probe was synthesized by PCR with biotin-11-dUTP (Biotium, Inc.,). Approximately 200 bp-long DNA fragments specific to each *TcPMP* gene were amplified from cDNA templates using forward and reverse primers, respectively (Table 4-1). The PCR products were purified using the Quick PCR purification kit (Genescript, Piscataway, NJ) and the purified dsDNA was used for asymmetric PCR, which was performed to generate the antisense or sense probe. Before hybridization probes were denatured by boiling for 40-60 min. After hybridization tissues were washed with hybridization solution at 48^{o}C for 2 h followed by washing 3 X with PBST for 10 min. The tissues were blocked in 1% bovine serum albumin (BSA, Fisher) by incubating at room temperature for 1-2 h. The blocking reagent was removed from the tissue and the tissues were incubated with anti-digoxigenin antibody conjugated with alkaline phosphatase (Roche Molecular Biochemicals, Mannheim, Germany, 1:1000 dilution in 1% blocking reagent) at RT for 1 h. The tissues were then washed in alkaline phosphatase (AP) buffer (100 mM NaCl/50 mM $MgCl_2$/100 mM Tris pH 9.5/0.1% Tween 20) three times for 10 min each at room temperature. The tissues were stained with nitroblue tetrazolium salt/5-bromo-4-chloro-3-indolyl phosphate (NBT/BCIP, 1:50 dilution in AP buffer). Staining was monitored in a dissection microscope. The color reaction was stopped by repeated washes with PBS and then mounted in 100% glycerol. For double labeling, tissues were incubated with streptavidin-horseradish peroxidase (1:100 in PBST) at RT for 1 h after incubation with anti-DIG AP. After several washes tissue was incubated with fluorescent dye Alexa Flour 488 (Molecular Probe) at RT for 10 min. Photographs were taken using a digital camera attached to the compound microscope and pictures were edited using Adobe Photoshop version 9.0.

4.4.5. Real time and RT PCR analyses of *PMP* transcripts after dsRNA injections.

RT-PCR experiments were carried out to monitor the effects of dsRNA administration on levels of the targeted transcripts by using gene-specific primers. Total RNA was isolated from whole insects 4-d post-injection of dsRNA using RNeasy Mini kit (Qiagen, Valencia, CA). Three insects were pooled for each RNA extraction. cDNA synthesis and RT-PCR were performed using gene-specific primers. The specificity of RNAi for the *PMP* genes was additionally analyzed by comparing the transcript levels of closely related *PMP* genes. RT-PCR products using the same cDNA template and a pair of primers for the ribosomal protein S6 were used as internal loading controls.

For the two genes, *TcPMP9* and *TcPMP14*, which are linked and separated only by a small intergenic region (Jasrapuria et al., 2010), real-time PCR was conducted using the Takara SYBR Ex Taq® premix reagent with the "shuttle PCR" program. The PCR reaction generated a melt curve to rule out the possibility of primer-dimer formation. The Ct values were determined and used for comparative quantitative analysis using the $2^{\Delta\Delta Ct}$ mehod (Wong and Medrano, 2005). The standard curves for each of these genes were normalized using *TcRpS6* amplification products as the reference standards.

4.4.6. Chitin staining of the gut

Midguts from last instar larvae were dissected out and chitin associated with the PM was stained using a chitin binding domain from a chitinase of W12 tagged with fluorescein-isothiocyanate (FITC-CBD, New England Biolabs, Beverly, MA) as described earlier (Arakane et al., 2004). Midguts were fixed in 10% formaldehyde/PBS (10 mM sodium phosphate buffer, pH 8 containing 100 mM NaCl) for 1 h on ice, followed by washing three times with PBS. The fixed guts were stained with FITC-CBD probe using 1: 500 dilution in PBS, pH 8 at RT. The

extra probe was washed and guts were photographed using GFP-filters with a Leica MZ FLIII fluorescence microscope.

4.5. Results

4.5.1. *PMP* genes are expressed in midgut tissues

To determine whether the expression of the *PMP* genes is confined to gut tissue, we separated gut, hindgut and carcass of larvae and adults, and analyzed RNA extracted from these tissues for presence of transcripts for each *PMP* gene. The larval midgut tissue was further divided into anterior, middle and posterior thirds prior to RNA extraction to investigate possible differences in the expression of specific *PMP* genes along the length of the midgut as reported previously for genes encoding chitin synthase B and some gut chitinases (Arakane et al., 2004; Bolognesi et al., 2001).

As expected, *PMP* genes were expressed in the gut tissues of both larvae and mature adults shown in Figure 4-1. Transcript levels for individual *PMP* genes were not uniform along the length of the larval midgut. Some genes were expressed only in the anterior and middle midgut (*TcPMP2-B, TcPMP3* and *TcPMP5-A*). Some were expressed only in the posterior midgut (*TcPMP1-B* and *TcPMP9*and *TcPMP14*). Others were expressed in all three segments of the midgut tissue (*TcPMP1-A, Tc PMP2-A* and *TcPMP2-C*). Only *TcPMP14, TcPMP5-B* and *TcPMP2-A* were expressed in the hindgut in the larval stage, whereas in the adult hindgut we could detect transcripts for only *TcPMP1-B* and *TcPMP14*. These results indicate that there are gene-specific as well as stage-specific differences in the expression patterns of individual genes within the family of *PMP* genes. However, transcripts for none of the *PMP* genes were detected in larval or adult carcass tissues, which are responsible for the secretion of the cuticle.

4.5.2. *PMP* genes are expressed mostly during feeding stages during development.

To determine if there are developmental stage-specific differences among the *PMP* family genes, the expression profiles of the 11 *PMP* genes were determined by RT-PCR using cDNAs prepared from RNAs isolated at different stages of insect development using gene-specific primers (Table 4-1). Transcripts were detected in the larval stages for all the *PMP* genes. *TcPMP1-A*, *TcPMP2-A*, *TcPMP5-B*, *TcPMP9* and *TcPMP14* were the only *PMP* genes expressed during the embryonic stage.

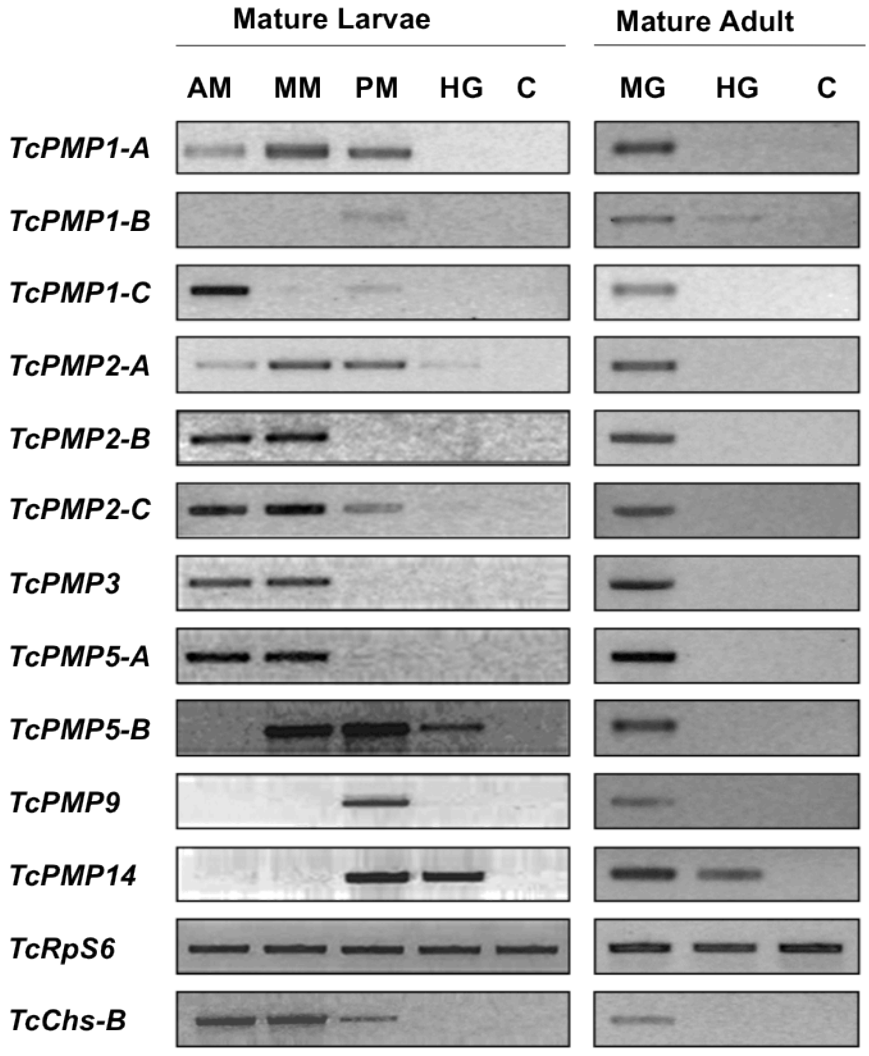


Figure 4-1 Tissue specificity of expression of *PMP* genes in *T. castaneum.*

Guts from actively feeding last instar larvae (n = 15) were dissected out and divided into 3 parts to obtain anterior midgut, middle midgut, and posterior midgut and hindgut tissue preparations. In addition, 15 actively feeding 2-week-old adults were dissected to obtain midgut and hindgut preparations. The remaining tissues were pooled and labeled as the carcass tissue. RT-PCR products for the T. castaneum ribosomal protein S6 (TcRpS6) amplified from the same cDNA templates served as an internal control for loading. Chitin synthase B (Chs-B) RT-PCR product served as a marker for monitoring the contamination of the carcass preparations by remnants of midgut tissue. AM, anterior midgut; MM, middle midgut; PM, posterior midgut; H, hindgut and pyloric valve; C, carcass (whole body minus gut). The number of cycles for RT-PCR was 28 for all genes except for RpS6 (24 cycles).

Among these genes, only the transcript level of *TcPMP1-A* was nearly the same in the embryonic as well as larval stages. At the pharate pupal stage, the transcript levels for the *PMP* genes had declined to undetectable levels except for very low levels of *TcPMP1-B* and *TcPMP5-B* transcripts. At the young adult stage, transcripts for *TcPMP2-A*, *TcPMP3, TcPMP5-B* and *TcPMP14* become detectable with transcripts for *TcPMP2-A* being the most abundant. All of the *PMP* genes were expressed at mature adult stages, when the PM is fully developed (Figure 4-2). However, the typical pattern of expression of *PMP* family genes is that they are expressed during actively feeding larval and adult stages.

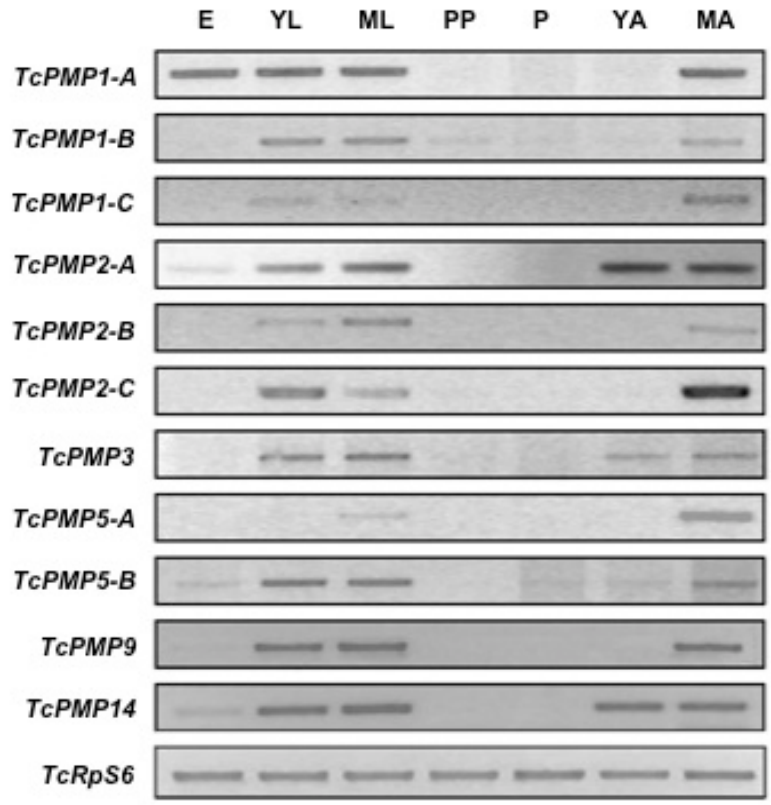


Figure 4-2 Expression profiles of *PMP* family genes as determined by RT-PCR.

Approximately 500 eggs (14.9 mg total weight) were collected and used for RNA isolation. Pools of four larvae, pharate pupae, pupae from day-0 to day-5, young adults (~ 1-2 h old), and mature adults (> 2 weeks-old) were used for preparation of total RNA at these stages. cDNAs synthesized from total RNA using oligo-$(dT)_{20}$ primers and reverse transcriptase were used as templates for RT-PCR (28 cycles) using gene-specific primers. RT-PCR products for the T. castaneum ribosomal protein S6 (RpS6) from the same cDNA templates served as an internal control for loading (24 cycles). The different stages of development are: E, Eggs; YL, penultimate instar or younger larvae; LL, last instar larvae; PP, pharate pupae; P, Pupae; YA, young adult; MA, Mature adult. The young adults were collected ~ 3 h after adult eclosion and mature adults were *10 d-old.*

4.5.3. Specificity of dsRNA-mediated depletion of *PMP* gene transcripts

Most of the *PMP* genes have highly similar nucleotide sequences, especially in the regions coding for the peritrophin-A domains. To minimize cross knock-down of transcripts of

non-target *PMP* genes, dsRNAs were synthesized corresponding to regions that included unique regions in the 5'-UTR or 3'-UTR regions.

dsRNAs for *TcPMP2-A, TcPMP3, TcPMP5-A, TcPMP5-B, TcPMP9* and *TcPMP14* were injected into insects in the penultimate larval instar (200 ng per insect, n=40). At this stage all of these genes were expressed except *TcPMP5-A,* which had very low transcript levels. RT-PCR analyses of RNA isolated 3 d after injection showed that each dsRNA resulted in depletion of transcripts for the target gene without any apparent cross-knock down of transcripts of other *PMP* genes (Figure 4-3). In most cases, the transcript levels were severely depressed with the lone exception being *TcPMP5-B*.

4.5.4. RNAi of *PMP* family genes

For subsequent larval injections, the *pu11* strain of *T. castaneum* was used, which allows for precise discrimination between penultimate and last instar larvae. This strain is an enhancer trap line in which the gene for GFP is expressed in elytra and wings, which is easily detectable under fluorescence microscopy. All of the 11 *PMP* genes are expressed at the larval stage and therefore, each of these genes was targeted with gene-specific dsRNAs at larval and pharate pupal stages to detect phenotypes. All insects except those injected with dsRNAs for *TcPMP3* and *TcPMP5-B* developed into normal adults without displaying any morphological or molting defects or mortality at larval, pupal or at adult eclosion stages. The results of the RNAi experiments with dsRNAs for *TcPMP3* and *TcPMP5-B* are described below.

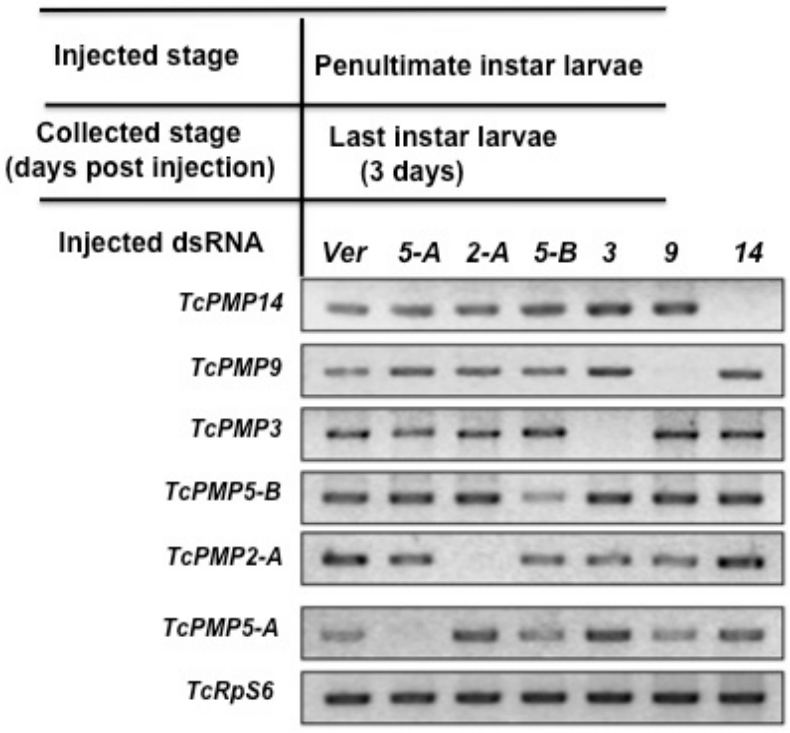


Figure 4-3 RT-PCR analyses to determine the target specificity of dsRNAs for *PMP* genes.

RT-PCR was carried out using RNA isolated four days after dsRNA injections to monitor the extent of depletion of the transcripts of the targeted PMP gene as well as some of the closely related gene(s) of the PMP family. The RT-PCR results indicated significant reduction in the transcripts of the targeted PMP gene and with no evidence of depletion of transcripts of other genes of the same family.

4.5.4.1. RNAi for *TcPMP3* gene affects larval to pupal molt

All of the animals injected with *TcPMP3* dsRNA (n=40) showed a lethal phenotype at the pharate pupal stage and pharate adult stage. The larval-larval molt was normal. The animals did not show any dorsal split and also slippage was not observed. However, the presence of a growing pupa inside the larval cuticle was evident as indicated by the presence of gin traps and

pupal eyes indicated by arrows in Figure 4-5 and the presence of a second cuticle underneath. When these insects were dissected and examined, a newly synthesized pupal cuticle was observed under the larval cuticle. When the larval cuticle was peeled off, the underlying pupal cuticle was revealed (data not shown). Similarly the injected pharate pupae were arrested just before adult molt and could not molt into adults. These insects had depleted fat body and died of desiccation. The insects remained in this stage for 3-4 d before they died. *TcVer* dsRNA-injected insects were used as controls and did not show any phenotypic abnormalities developing into healthy adults except for the loss of eye pigmentation (Figure 4-5).

4.5.4.2. RNAi for *TcPMP5-B* affects larval to pupal molt and pupal to adult molt.

RNAi for both *TcPMP3* and *TcPMP5-B* injection of dsRNA resulted in two terminal phenotypes at the larval-pupal molt and the pupal-adult molts. When dsRNA for *TcPMP5-B* was injected into penultimate instar larvae, they were able to molt into last instar larvae, but were arrested at pharate pupal stage and there was partial pupal development without ecdysis. Injection of dsRNA for *TcPMP5-B* into pharate pupae prevented the pupal-adult molt and the resulting insects were arrested at the pharate adult stage. In this case, apolysis and partial slippage of the old pupal cuticle were observed, but these insects failed to complete ecdysis. At the time of death, the adult cuticle was visible under the old pupal cuticle, but the animals failed to release the old pupal cuticle that remained attached to the abdomen and there was no abdominal contraction and elytral expansion. Also there was no fat body depletion (data not shown). The phenotypes resulting from injection of dsRNA for *TcPMP5-B* into penultimate instar larvae of *T. castaneum* are shown in Figure 4-5. *TcPMP5-B* transcripts were not observed at the developing pupal stage after 28 cycles of PCR. However, by increasing the PCR cycle number, transcripts were detected throughout the pupal stage.

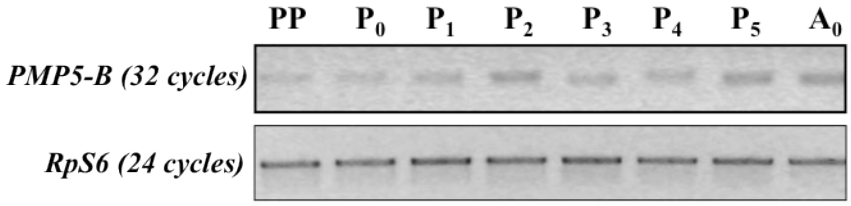


Figure 4-4 Expression of *TcPMP5-B* gene during pupal stages

cDNA templates were synthesized from different stages of pupal development: PP, pharate pupae day 1; P [0-5], 0-day old pupae to 5-day old pupae; YA, young adult. PMP5-B gene does not show detectable transcripts at pupal stage with 28 cycles (data not shown), but low transcripts were observed at 32 cycles of PCR. RT-PCR products for the T. castaneum ribosomal protein S6 (RpS6) from the same cDNA templates served as an internal control for loading (24 cycles).

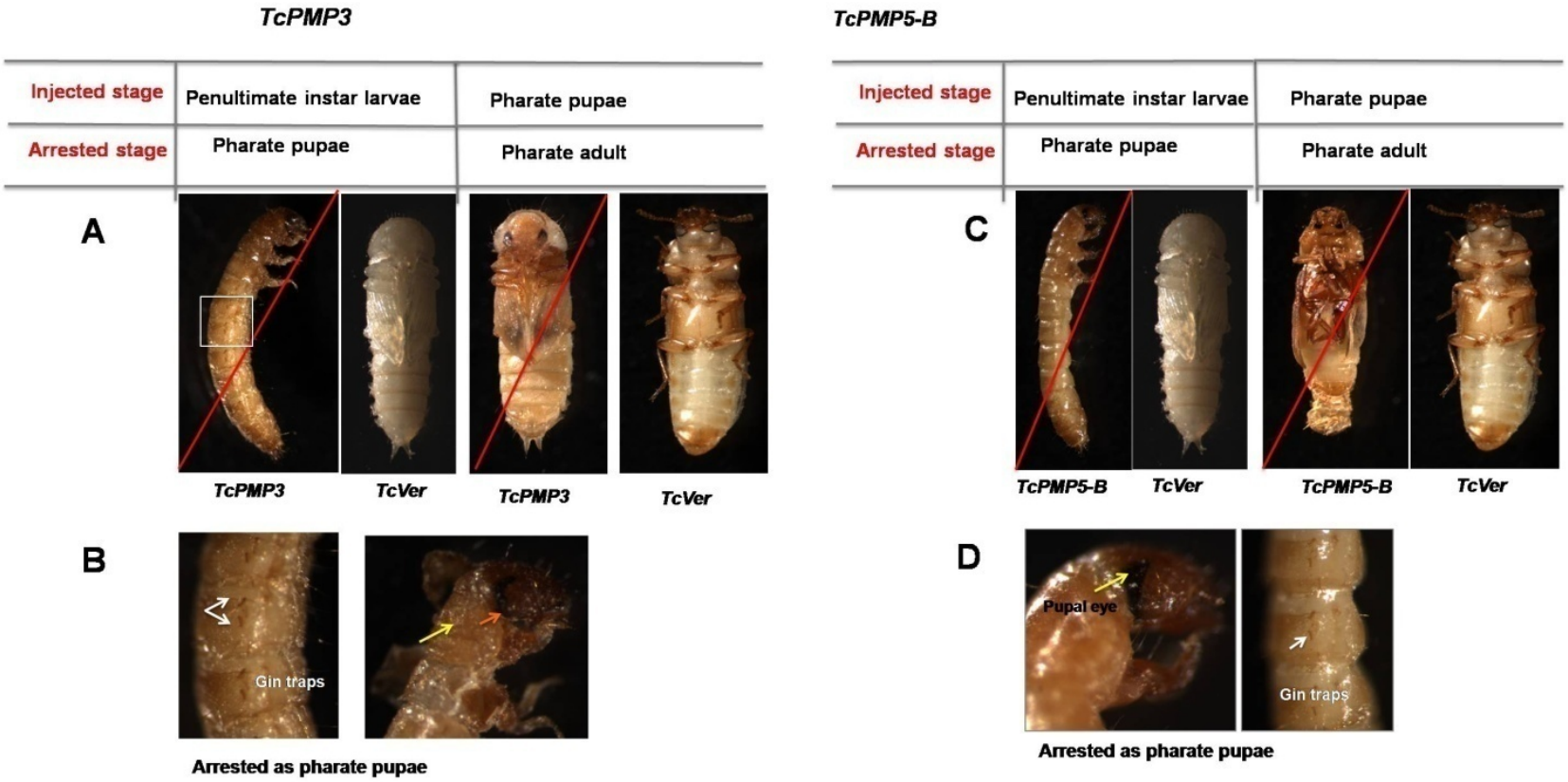


Figure 4-5 Phenotypes of dsRNA *TcPMP3* and *dsRNA TcPMP5-B*-injected animals.

Shown are lethal phenotypes produced by injection of dsRNAs for dsTcPMP3 and dsRNA for TcPMP5-B along with the Vermilion dsRNA-injected control. Gene-specific dsRNAs for the PMP genes were injected into penultimate instar larvae (200 ng per insect; n = 40), or pharate pupae (200 ng per insect; n = 40). Left Panel A shows arrested growth at pharate pupal stage following dsRNA TcPMP3 injection into last instar larvae. Left Lower panel B shows the gin traps suggesting that the arrested animal is undergoing pupal development beneath the larval cuticle. Right panel C shows the effect of RNAi for TcPMP5-B. The penultimate instar larvae could undergo larval to larval molt, but were arrested at pharate pupal stage and the pharate pupae injected with dsRNA for

TcPMP5-B could molt but failed to complete adult eclosion. D, Enlarged view of the developed pupal eye in the dsRNA. TcPMP5-B-treated larvae that were arrested at pharate pupal stage with visible gin traps.

4.5.5. RNAi for *TcPMP9* and *TcPMP14*

TcPMP9 and *TcPMP14* genes are arranged in tandem in linkage group 10 in the *T. castaneum* genome and their CDSs are separated by only 331 base pairs. As described in an earlier chapter, this two-gene cluster had previously been annotated as a single glean model, 08506. Further studies reported in (Jasrapuria et al., 2010) have shown that these are actually two separate genes. However, the proximity and absence of a long intergenic region between these genes suggested that these two genes may be coordinately regulated and perhaps have redundant functions. Therefore, we carried out a detailed quantitative study of their expression.

Quantitative analysis of transcript levels of these two genes was done throughout all developmental stages shown in Figure 4-6. As expected, the expression patterns of these two genes are very similar. Both genes are expressed predominantly during feeding stages (larval and adult stages), with very little expression during non-feeding (pupal) stages. *TcPMP14* transcripts were detected in young adults at a time when *TcPMP9* transcripts were barely detectable at the young adult stage (1-2 h after adult eclosion), indicating a slightly earlier appearance of the transcripts for this gene during adult eclosion. These data are consistent with the RT-PCR data reported in Figure 4-2. It seems likely that *PMP9* and the down-stream gene *PMP14* are not coordinately expressed.

4.5.5.1. *TcPMP9* and *TcPMP14* transcripts are expressed exclusively in the posterior midgut

RT-PCR data described in Fig. 4-2 had indicated that the *TcPMP14* gene was expressed in both the posterior midgut and the hindgut, whereas the closely linked *TcPMP9* was expressed only in the posterior midgut but not in the hindgut. *In situ* hybridization data further confirmed the RT-PCR data. *TcPMP9* and *TcPMP14* transcripts were expressed in gut tissues in the late larval stage. *TcPMP9* was expressed only in the anterior part of the posterior midgut but not in the posterior part of the posterior midgut or in other parts of the midgut, whereas *TcPMP14* was expressed in both the anterior and posterior parts of the posterior midgut tissue but not in the hindgut tissue or in the pyloric valve. Therefore, *TcPMP9*-expressing cells represent a subset of those expressing *TcPMP14*. Positive signals for both transcripts were found in the columnar cells of the posterior midgut (Figure 4-7).

4.5.5.2. Co-injection of dsRNA for *TcPMP9* and *TcPMP14* does not affect normal development

The closely linked *TcPMP9* and *TcPMP14* genes are expressed in the columnar cells of the posterior midgut with roughly parallel expression patterns. We considered the possibility that they might have redundant functions. RNAi experiments involving a mixture of dsRNAs for both *TcPMP9* and *TcPMP14* genes (200 ng per insect) and penultimate instar larvae did not result in any visible abnormalities or mortality. The qPCR data further confirmed the RT-PCR data. The depletion of transcripts for *TcPMP9* and *TcPMP14* were 85% and 97%, respectively, relative to controls (Figure 4-8). dsRNA-mediated RNAi did not result in any observable phenotype, even though substantial depletion of transcripts for both genes could be demonstrated by qRT-PCR.

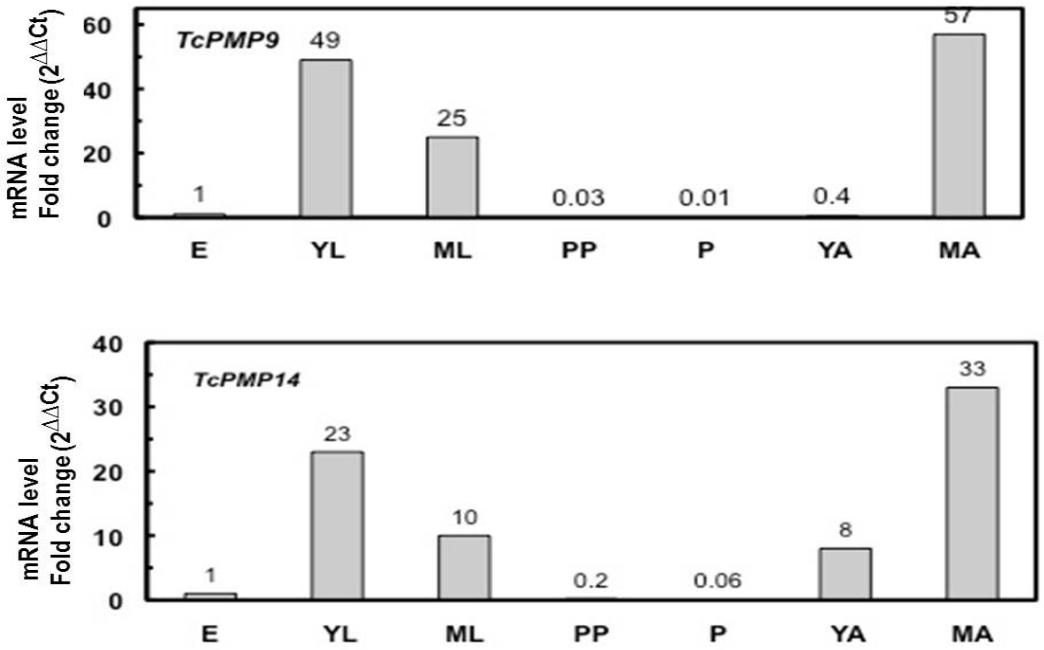
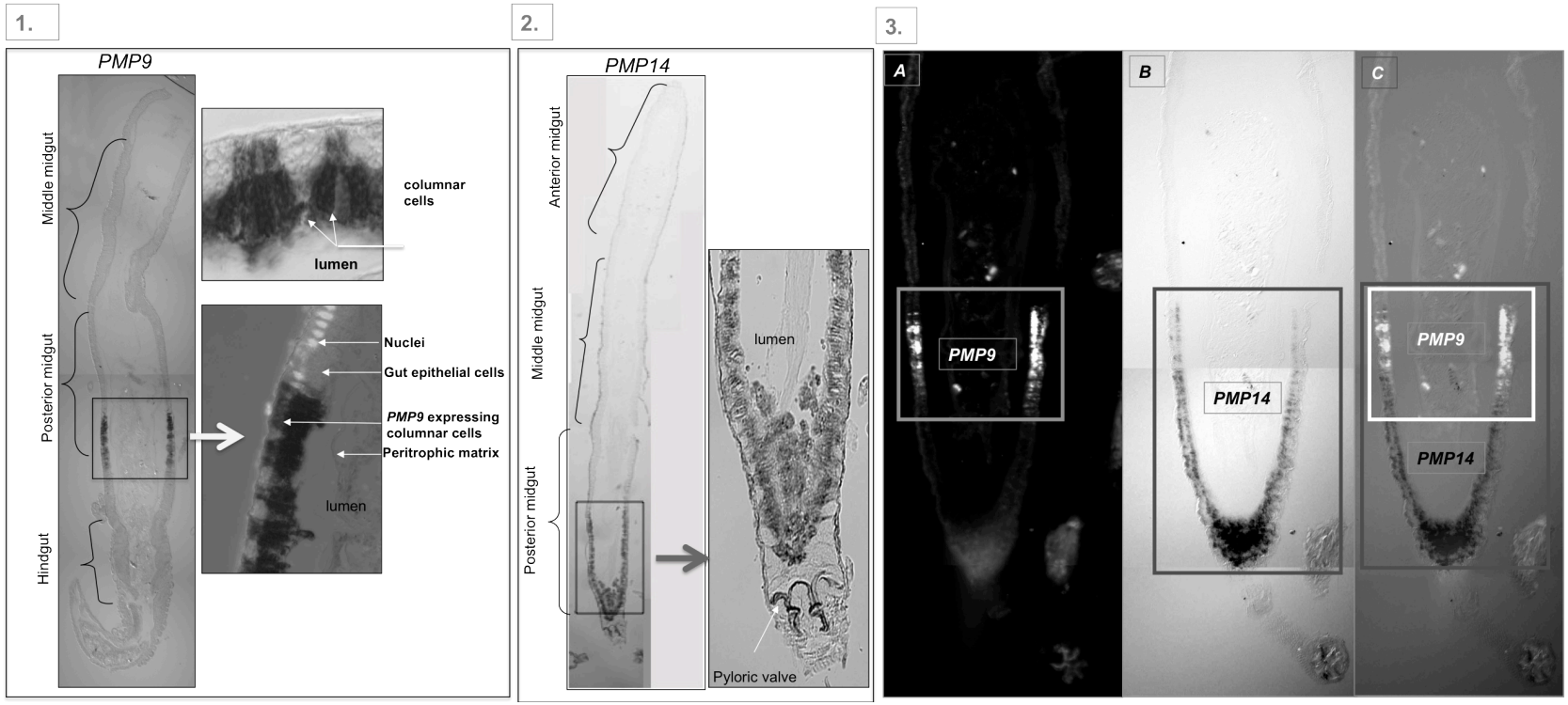


Figure 4-6 Real-time PCR analysis of transcript levels of the two linked *PMP* genes, *TcPMP9* and *TcPMP14 during*

developmental stages.

The cDNAs used for real-time PCR analysis were prepared from total RNA extracted at different stages of development from whole insects of T. castaneum. Data are presented as the average of two separate experiments. The cDNA template was utilized to amplify RT-PCR products of the housekeeping gene TcRpS6 using a pair of primers and these values were utilized to normalize differences in template levels between samples. Steady-state transcript levels for TcPMP9 and TcPMP14 at different developmental stages are shown relative to the levels in embryonic stage (E). Different stages of development: E, Eggs; YL, penultimate instar or younger larvae; LL, last instar larvae; PP, pharate pupae; P, Pupae; YA, young adult; MA, Mature adult. The young adults were collected at ~3 h after adult eclosion, and mature adults 10 d after adult eclosion.



Dr. K. Begum

Figure 4-7 *in situ* hybridization analysis of cell type and tissue expressing *TcPMP9* and *TcPMP14* genes in the gut.

All pictures are shown with the anterior side of the gut at the top and posterior side at the bottom. 1. TcPMP9 transcripts were localized in the posterior midgut by DIG (digoxigenin) labeled single-stranded negative strand DNA probes. Inset shows positive signal in the columnar cells on the lumenal side of the gut. 2. TcPM14 transcripts were localized in the posterior midgut by DIG (digoxigenin) labeled ss-DNA probes. For negative control, a sense probe was used and no signal was detected (data not shown) 3. Double labeling: TcPMP9DNA probe was labeled with biotin and TcPMP14DNA probe was labeled with DIG. In situ hybridization showed that both the genes were expressed in posterior midgut in columnar cells. TcPMP9-expressing cells are a subset of those expressing TcPMP14that are located in more anterior locations relative to those expressing TcPMP14.

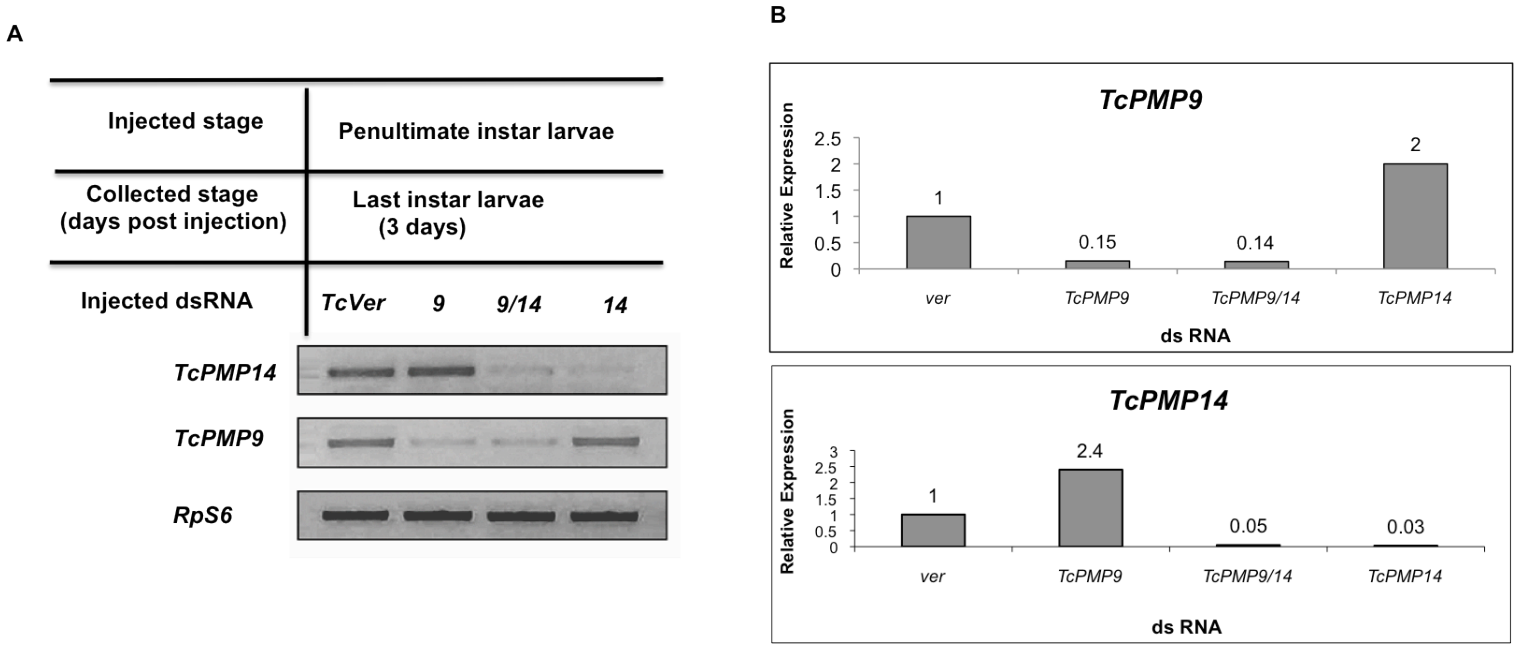


Figure 4-8 Transcript levels after co-injection of dsRNAs for *TcPMP9* and *TcPMP14*

dsRNAs for TcPMP9, TcPMP14 or a mixture of these two (200 ng per insect; n=40) were injected into penultimate instar larvae. dsRNA TcVer was injected as a control dsRNA to monitor non-specific effects. RT-PCR and qRT-PCR of T. castaneum RpS6 transcripts amplified with the same cDNA template served as an internal control for equalizing template concentrations. The levels of TcPMP9 and TcPMP14 transcripts as determined by qRT-PCR are shown relative to the levels of expression of these two genes in TcVer dsRNA-treated controls.

4.5.6. RNAi of adult females

To study the role of *PMP* genes in embryonic development and early instar larval development, adult females (n=20, one month-old) were injected with 200 ng of dsRNA for four of the *PMP* genes, *TcPMP3, TcPMP5-B, TcPMP9* or *TcPMP14*. Adult mortality, fertility (as measured by number of eggs laid) and percent hatch of eggs were monitored. Injection of dsRNA for *TcPMP3* resulted in 100% lethality in adult females with severely depleted fat bodies (Figure 4-9). Eggs from females after RNAi for *TcPMP5-B* were morphologically compromised and never hatched. There is no female mortality after RNAi for *TcPMP5-B*.

4.5.7. FITC-CBD staining of the gut after RNAi for *PMP* genes

We considered the possibility that the chitin content of the PM could be influenced by the presence/absence of specific PMP proteins. The guts dissected from larvae injected with dsRNAs for *TcPMP3, TcPMP5-B*, *TcPMP9, TcPMP14* or *TcPMP9/TcPMP14* was stained with FITC-CBD probe to detect chitin in the PM. No obvious reduction in the chitin content was observed in this experiment. dsRNA for *Chs-B* was used as a control to monitor the effectiveness of this technique because this gene is essential for the synthesis of PM-associated chitin (Arakane et al., 2004). As expected, chitin staining was substantially reduced after RNAi for *Chs-B*. There was no apparent loss of chitin staining after RNAi for any of the *PMP* genes (Figure 4-10). *TcVer* is a positive control and PM from these animals stained well with FITC-CBD.

TcPMP5-B

Morphologically defective

TcVer

TcPMP3

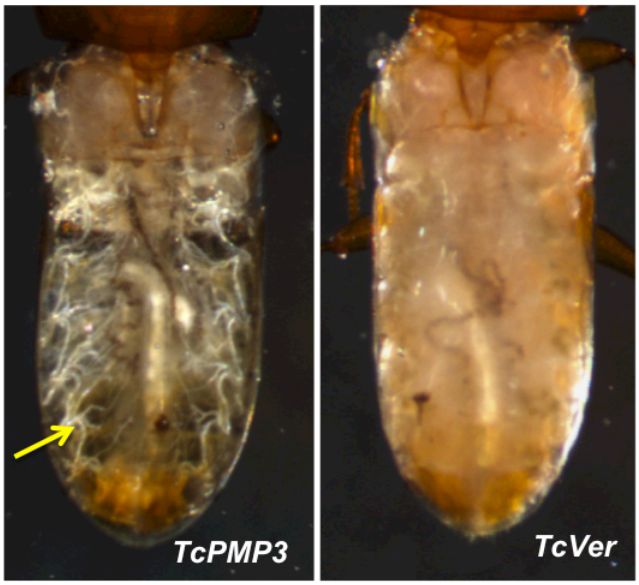


Figure 4-9 Parental RNAi for *TcPMP3* and *TcPMP5-B*

dsRNAs (200 ng per insect) for TcPMP3 or TcPMP5-B were injected into 2-weeks-old mature adults. The resulting adults from the dsRNA TcPMP3 treatment died 10-15 d after injection. They had severely depleted fat bodies and the few resulting eggs never hatched. dsRNA TcPMP5-B-treated females laid normal numbers of eggs compared to dsRNA TcVer-treated females. However, the eggs were morphologically defective, as they did not have proper shape and never hatched.

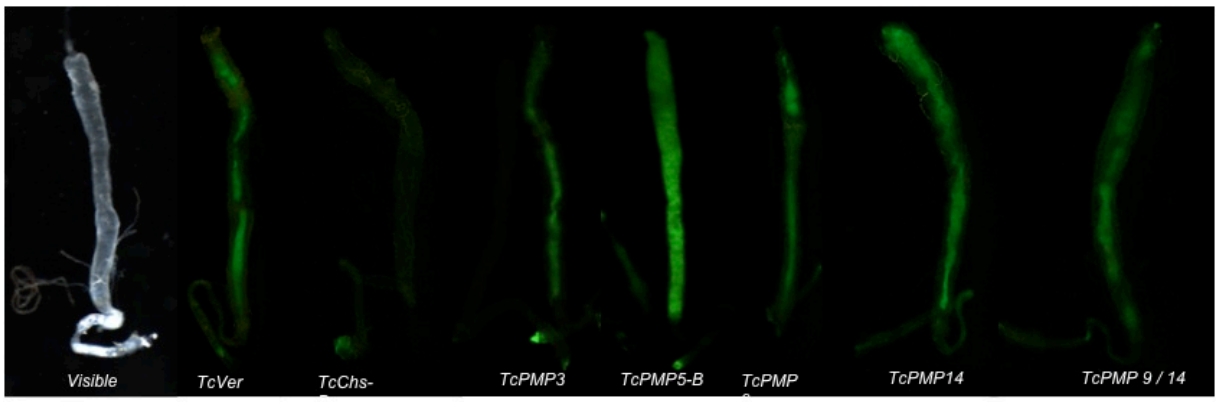


Figure 4-10 Staining of PM-associated chitin in midguts of larvae subjected to RNAi for *PMP* genes.

Groups of penultimate instar larvae were injected with the indicated dsRNA (200 ng/insect; n =20). After 3 d, the insects were dissected to obtain the gut tissue, washed with PBS and stained with FITC-CBD as described in "Materials and Methods" section. The guts were visualized under a fluorescence microscope and photographed. One representative gut from each group is shown.

4.6. Discussion

Using a bioinformatics approach, we previously identified >50 genes capable of encoding proteins with one or more peritrophin-A domains. Some of these proteins could be identified as enzymes of chitin metabolism, including chitinases and chitin deacetylases, with a peritrophin-A domain, which increases their affinity for insoluble chitin (Arakane et al., 2003). Phylogenetic analyses of the remaining proteins with ChtBD2 domains indicated that they form three clades of related proteins. We named one clade as 'peritrophins" because they had the greatest sequence similarity to previously characterized peritrophins extracted from insect PMs. However, the proteins encoded by the *PMP* genes have not been unambiguously identified as peritrophins because of the difficulties of extracting many PM proteins, especially those with more than 4 peritrophin-A domains (Dinglasan et al., 2009).

For this report we relied on the tissue-specificity of detection or localization of transcripts corresponding to these putative *PMP* genes as evidence for their identity as PMPs. As expected, all of the eleven putative *PMP* genes were indeed expressed in midgut tissues. In addition, these genes were not expressed in the carcass, a tissue that is involved in the synthesis of cuticle-associated chitin. Finally, these PMP genes were expressed only during feeding stages when PM could be easily detected in larval and adult stages. Collectively, these observations lend support to the notion that these genes indeed encode PMPs. Our attempts to express these proteins in *E. coli* and obtain enough quantities in soluble form to carry out biochemical studies have been largely unsuccessful. The tiny amounts of PM that can be isolated from *T. castaneum* larvae have so far precluded the extraction and identification of some of these proteins as bona fide PMPs. We are currently attempting to express these proteins in a baculovirus/insect cell system because there have been some reports of successful expression of PMPs using this system (Wang et al., 2004).

Conversely, the genes encoding the proteins representing the other two clades of the phylogenic tree are not expressed in the gut tissue (see CHAPTER 2; (Jasrapuria et al., 2010)). Instead, they are expressed in cuticle-forming tissues and during pupal stages when no *PMP* genes are expressed. These results of tissue-specificity and developmental patterns of expression provide experimental support for the separation of the three families.

4.6.1. *PMP* gene expression pattern varies along the length of the midgut

Insects with a type I PM are predicted to synthesize components of this extracellular matrix (primarily chitin and PMPs) along the entire length of the midgut with all columnar cells lining the midgut contributing to this process. It was natural to expect that all *PMP* genes would be expressed uniformly along the entire length of the midgut. Previous studies (Bolognesi et al.,

2005) have indicated that *Chs-B*, the gene encoding the chitin synthase protein responsible for the synthesis of PM-associated chitin, is expressed in the anterior and middle parts of the midgut and only low levels of transcripts were detected in the posterior midgut. The present study with the *T. castaneum* larval gut divided into roughly three equal parts (anterior, middle and posterior midgut) led to the surprising finding that there are distinct differences in the patterns of expression of individual *PMP* genes in these three sections of the midgut. The most notable finding is that the two genes with the largest number of peritrophin-A domains (*PMP9* and *PMP14*) are expressed in the most posterior parts of the midgut, while generally the genes encoding PMPs with fewer peritrophin domains are predominantly expressed in the anterior parts of the midgut (one exception is PMP1B). There are differences even in the expression of the two *PMP* genes whose expression is confined to the posterior part of the gut. *PMP9* is expressed only in a subset of the cells that express *PMP14*. The most posterior part of the midgut expresses *PMP14* but not *PMP9*, indicating some biological function unique to this protein with a large number of peritrophin-A domains in the posterior midgut. Several other insect species including *Aedes aegypti*, *Mamestra configurata*, *B. mori* and *M. sexta* have proteins with 10 or more peritrophin-A domains. Perhaps this localization pattern is a feature characteristic of many, if not all, insects with a PM.

The significance of the variation in number of ChtBD2s among the family of PMPs from the same species as well as the variation in number of ChtBD2s in PMPs among species from different orders of insects is unclear. It is possible that the larger the number of ChtBD2s, the tighter is the binding to chitin. Consistent with this idea, (Dinglasan et al., 2009) were able to identify only twelve proteins with one to four ChtBD2s in proteins extracted from the *A. gambiae* PM with 0.5% Triton X-100 or 2% SDS even though genes encoding proteins with

more ChtBD2s are present in the *A. gambiae* genome. It is likely that PMPs with a large number of ChtBD2s are bound very tightly to chitin in the PM and, therefore, are not extracted even under rather harsh conditions.

The differences in the location of expression of different *PMP* genes in the cells lining different parts of the midgut suggest that this may have biological significance. It is possible that the PMPs secreted by epithelial cells in a certain part of the midgut associate immediately with the lattice-like chitin network being assembled on the microvilli of the brush border cells in the same region of the midgut (Hopkins and Harper, 2001). If this indeed is the case, there could be differences in the properties of the PM in different parts of the midgut depending on the local differences in the nature of the protein(s) associated with the PM. If TcPMP9 and TcPMP14 are confined to only the posterior part of the PM in the gut, the permeability of the PM to gut contents in the posterior midgut could differ dramatically from that in the anterior parts of the PM where genes encoding PMPs with fewer ChtBD2 domains are expressed. This would be consistent with the model proposed by (Bolognesi et al., 2008) that the permeability of the PM to digestive enzymes and by inference, partial digestion products of dietary components, could differ with the anterior part of the PM being more porous. Such a distribution could help to establish a differential flux of gut contents between the endo- and ectoperitrophic spaces. Indeed, the PM in the anterior midgut is more jelly-like and fragile in its physical properties compared to the tougher and more opaque PM in the posterior midgut. Direct measurement of the permeability of the PM in different parts of the gut may provide experimental support for this hypothesis.

It has been proposed that proteins with single ChtBD2 cap the ends of chitin fibrils, whereas PMPs with multiple ChtBD2s cross-link adjacent chitin chains (Wang and Granados,

2000). It is interesting to note that both the number and spacing of the ChtBD2 repeats varies among the members of the PMP family in *T. castaneum*. Whether these differences influence their binding to chitin or the properties of the PM remains to be investigated. It should be pointed out that none of the *T. castaneum* proteins identified in this study have actually been shown to bind to chitin. We are currently in the process of expressing and purifying recombinant representatives of each class of ChtBD2-containing proteins to address this issue.

RNAi experiments to knockdown the *AaCHS2* transcripts of *Ae. aegypti* showed that it is required in female mosquitoes for the *de novo* synthesis of the PM after a blood meal (Kato et al., 2006). We have carried out RNAi experiments with each of the 11 *PMP* genes to determine their functions. In only two instances could we observe phenotypes when dsRNAs for *PMP* genes were injected into penultimate instar larvae (*PMP3* and *PMP5-B*). Larvae injected with dsRNA for one of these two genes (*PMP3*) shrank in size and did not shed their larval exuvia. The presence of anatomical features characteristic of pupae such as gin traps and compound eyes indicated that the insects had undergone pupal differentiation without molting. When injected with dsRNA for *PMP5-B*, the lresulting insects failed to complete larval-pupal molt as well as pupal-adult molt. It is interesting that this is the only *PMP* gene with expression in the pupal stage. Parental RNAi for the same two genes also yielded lethal phenotypes or failure of embryonic development, suggesting that these two genes are essential at multiple stages in development. When females were injected with dsRNA for *TcPMP5-B*, there was no lethality. However the eggs were misshaped and never hatched. The failure to see any visible phenotypes when transcripts for the other *PMP* genes were depleted may indicate that these genes may have redundant functions in such a way that most of the PMP proteins compensate for the loss of other

PMP proteins. Alternatively, PMPs may perform functions that do not affect survival or fitness unless the animal is placed under conditions of stress or immune challenge.

4.7. Bibliography

Arakane, Y., Hogenkamp, D.G., Zhu, Y.C., Kramer, K.J., Specht, C.A., Beeman, R.W., Kanost, M.R., Muthukrishnan, S., (2004). Characterization of two chitin synthase genes of the red flour beetle, Tribolium castaneum, and alternate exon usage in one of the genes during development. Insect Biochem Mol Biol 34, 291-304.

Arakane, Y., Specht, C.A., Kramer, K.J., Muthukrishnan, S., Beeman, R.W., (2008a). Chitin synthases are required for survival, fecundity and egg hatch in the red flour beetle, *Tribolium castaneum*. *Insect Biochem Mol Biol* 38, 959-962.

Arakane, Y., Zhu, Q., Matsumiya, M., Muthukrishnan, S., Kramer, K.J., (2003). Properties of catalytic, linker and chitin-binding domains of insect chitinase. *Insect Biochem Mol Biol* 33, 631-648.

Balbani, E., (1890). Etudes anatomiques et histologiques sur le tube digestif des. *Crytops*. *Arch*. *Zool*. *Exp*. *Gen*. 8, 1-82.

Barry, M.K., Triplett, A.A., Christensen, A.C., (1999). A peritrophin-like protein expressed in the embryonic tracheae of *Drosophila melanogaster*. *Insect Biochem Mol Biol* 29, 319-327.

Beeman, R.W., Stuart, J.J., (1990). A gene for lindane + cyclodiene resistance in the red flour beetle (Coleoptera: Tenebrionidae). *J.Econ*. *Entomol* 83, 1745-1751.

Behr, M., Hoch, M., (2005). Identification of the novel evolutionary conserved obstructor multigene family in invertebrates. *FEBS Lett* 579, 6827-6833.

Bolognesi, R., Arakane, Y., Muthukrishnan, S., Kramer, K.J., Terra, W.R., Ferreira, C., (2005). Sequences of cDNAs and expression of genes encoding chitin synthase and chitinase in the midgut of *Spodoptera frugiperda*. *Insect Biochem Mol Biol* 35, 1249-1259.

Bolognesi, R., Ribeiro, A.F., Terra, W.R., Ferreira, C., (2001). The peritrophic membrane of *Spodoptera frugiperda*: secretion of peritrophins and role in immobilization and recycling digestive enzymes. *Arch Insect Biochem Physiol* 47, 62-75.

Bolognesi, R., Terra, W.R., Ferreira, C., (2008). Peritrophic membrane role in enhancing digestive efficiency. Theoretical and experimental models. *J Insect Physiol* 54, 1413-1422.

Dinglasan, R.R., Devenport, M., Florens, L., Johnson, J.R., McHugh, C.A., Donnelly-Doman, M., Carucci, D.J., Yates, J.R., 3rd, Jacobs-Lorena, M., (2009). The *Anopheles gambiae* adult midgut peritrophic matrix proteome. *Insect Biochem Mol Biol* 39, 125-134.

Hegedus, D., Erlandson, M., Gillott, C., Toprak, U., (2009). New insights into peritrophic matrix synthesis, architecture, and function. *Annu Rev Entomol* 54, 285-302.

Hopkins, T.L., Harper, M.S., (2001). Lepidopteran peritrophic membranes and effects of dietary wheat germ agglutinin on their formation and structure. *Arch Insect Biochem Physiol* 47, 100-109.

Jasrapuria, S., Arakane, Y., Osman, G., Kramer, K.J., Beeman, R.W., Muthukrishnan, S., (2010). Genes encoding proteins with peritrophin A-type chitin-binding domains in *Tribolium castaneum* are grouped into three distinct families based on phylogeny, expression and function. *Insect Biochem Mol Biol* 40, 214-227.

Kato, N., Mueller, C.R., Fuchs, J.F., Wessely, V., Lan, Q., Christensen, B.M., (2006). Regulatory mechanism of chitin biosynthesis and roles of chitin in peritrophic matrix formation in the midgut of adult *Aedes aegypti*. *Insect Biochem. Mol. Biol.,* 36, 1-9.

Lorenzen, M.D., Berghammer, A.J., Brown, S.J., Denell, R.E., Klingler, M., Beeman, R.W., (2003). piggyBac-mediated germline transformation in the beetle *Tribolium castaneum*. *Insect Mol Biol* 12, 433-440.

Perez-Vilar, J., Hill, R.L., (1999). The structure and assembly of secreted mucins. *The Journal of biological chemistry* 274, 31751-31754.

Peters, W., 1992. *Peritrophic Membranes*. Berlin: Springer.

Tellam, R.L., Wijffels, G., Willadsen, P., (1999). Peritrophic matrix proteins. *Insect Biochem Mol Biol* 29, 87-101.

Terra, W.R., (2001). The origin and functions of the insect peritrophic membrane and peritrophic gel. *Arch Insect Biochem Physiol* 47, 47-61.

Tomoyasu, Y., Denell, R.E., (2004). Larval RNAi in *Tribolium* (Coleoptera) for analyzing adult development. *Dev Genes Evol* 214, 575-578.

Venancio, T.M., Cristofoletti, P.T., Ferreira, C., Verjovski-Almeida, S., Terra, W.R., (2009). The *Aedes aegypti* larval transcriptome: a comparative perspective with emphasis on trypsins and the domain structure of peritrophins. *Insect Mol Biol* 18, 33-44.

Wang, P., Granados, R.R., (2000). Calcofluor disrupts the midgut defense system in insects. *Insect Biochem Mol Biol* 30, 135-143.

Wang, W., Davison, S., Krell, P.J., (2004). Identification and characterization of a major early-transcribed gene of *Trichoplusia ni* single nucleocapsid nucleopolyhedrovirus using the baculovirus expression system. *Virus Genes* 29, 19-29.

Wigglesworth, V., (1930). The formation of the peritrophic membrane in insects, with special reference to the larvae of mosquitoes. *Q. F. Microsc. Sci.* 73, 583-616.

Wong, M.L., Medrano, J.F., (2005). Real-time PCR for mRNA quantitation. *Biotechniques* 39, 75-85.

CHAPTER 5

Summary & Discussion

5.1. General Conclusion

This is the first comprehensive study of all of the proteins containing ChtBD2-(peritrophin-A) domain(s) in a single insect species focusing on bioinformatics identification and functional characterization. These proteins are predicted to interact with chitin, and have a role in assembly and/or turnover of chitin-containing structures. We chose the model beetle *Tribolium castaneum* because of its well-annotated genome and the ease with which RNA interference studies can be carried out in this beetle.

Our study has established that this chitin-binding motif is ancient being present in almost all organism that utilize chitin with the lone exception being fungi. The ChtBD2 domain is found in proteins not only across different insect orders including Diptera, Coleoptera, Hemiptera, Hymenoptera, Lepidoptera and Phthiraptera but also in crustaceans, chelicerates and nematodes. The ChtBD2 motif with six characteristically spaced cysteines has been repeatedly utilized for fulfilling several functions that involve protein-chitin interactions. Almost all of them are predicted to be secretory proteins consistent with the presence of chitin exclusively in extracellular matrices. There are several groups of proteins with the ChtBD2 motif, all of which appears to have important roles in chitin metabolism and organization into extracellular matrices.

In *T. castaneum* there are >50 proteins with the ChtBD2 domain, which can be divided into four groups. The first group encompasses enzymes of chitin metabolism including 6 chitinases and 7 chitin deacetylases (including CDA5 with two isoforms, each with one ChtBD2), which have one or more ChtBD2 domains (Arakane and Muthukrishnan, 2010a; Dixit et al., 2008; Kramer, 1993; Zhu et al., 2008b). In these proteins, the function of the ChtBD2 domain appears to be to promote catalysis by anchoring the enzyme onto the insoluble chitin

substrate. This tight association of enzyme and substrate appears to influence the catalytic efficiency of these enzymes.

The remaining ~50 proteins with ChtBD2 domains in *T. castaneum* have been further grouped into three families-: Peritrophic Matrix Proteins (PMPs) expressed in the midgut tissues, and two families of Cuticuar Protein Analogous to Peritrophins (CPAP1 and CPAP3) expressed in cuticle-forming tissues. These three groups of proteins are also found in all insect species whose genomes have been fully annotated, indicating highly conserved and essential functions for all three families. Most notable is the conservation of orthologs of several members of the CPAP3 family in a range of taxa, suggesting these are essential genes.

PMPs and CPAP3 families of proteins are essentially tandem arrays of the ChtBD2 domains separated by spacers or linkers. All of them are presumed to be secreted proteins and are thought to bind to chitin in the PM or the cuticle. In these extracellular matrices they are likely to serve a structural function arranging the chitin fibrils into the characteristic lattice-like structures in the case of the PM or into the multi-layered epidermal cuticle in which the chitin bundles are highly ordered in helicoidal laminar arrays. The functions of the CPAP1 family of proteins are unknown. They may have enzymatic functions or may be involved in interactions with other cuticular proteins.

Our studies have further refined and extended the sequence features of the ChtBD2 domains beyond the original sequence motif proposed by (Tellam et al., 1999) which was based on the smaller set of peritrophins and related proteins available at that time. The revised ChtBD2 consensus sequences proposed now include not only the characteristically spaced six-cysteine motif but also sequences between these cysteines and sometimes even include flanking

sequences. Thus, one can identify the PMP, CPAP3 and CPAP1 family members by comparing them to the consensus sequences for each family of ChtBD2 protiens proposed in this work.

The consensus sequence for each of the three ChtBD2 sequences among the CPAP3 family proteins from different species is so well conserved that not only is it possible to identify the CPAP3 proteins from other families of ChtBD2–containing proteins (i.e. PMP or CPAP1 families) but also to identify its relationship to members of individual sub-families of the larger CPAP3 family by comparing their signature sequences. Even more striking is the finding that in the CPAP1 family of proteins that has a single ChtBD2 domain, the consensus includes a stretch of 15 amino acids flanking the six cysteines. The PMP family of proteins has the most variation among the ChtBD2 sequence even though multiple repeats within the same PMP protein in a given species may be highly similar or even identical. PMPs have repeats and linker regions that are rich in serines and threonine residues and are excellent candidates for O-glycosylation.

The presence of orthologs of genes encoding the CPAP3 family of proteins and possibly other families of ChtBD2-containing proteins in insects belonging to several orders of insects and even some arthropods suggests that there is an evolutionary need for conservation of so many families and individual members of these families of proteins for cuticle assembly/integrity. One of the unanswered questions is: why are there so many proteins with related sequences? Even more intriguing is the question: why do some species including *D. melanogaster* and *A. gambiae* have gene duplications resulting in an increased number of proteins within the same CPAP3 family, whereas other species do not.

Chitin binding: The number of ChtBD2 domains varies widely in the family of PMPs from one to nineteen in different insect species (from 1 to 14 in *T. castaneum*). The genes encoding them are not expressed uniformly along the length of the PM, but there is a distinct

overrepresentation of proteins with higher numbers of ChtBD2s in the posterior parts of the midgut. Perhaps these differences in the site of expression within the midgut are related to the different permeability requirements of the PM along the length of the midgut as proposed by (Bolognesi et al., 2008). Furthermore, the PM is a fragile and hydrophilic structure, whereas the cuticle is rigid and hydrophobic. The biological significance of an abundance of CPAP proteins with very similarly-spaced ChtBD2 domains is unclear. These domains are found exclusively in the cuticle where chitin is arranged as fibrils instead of the lattice network of PM chitin. The CPAP3 proteins may bind to different chitin bundles or mediate intra and inter-chain interactions among chitin bundles and the presence of three ChtBD2 domains may be involved in these interactions. CPAP1 may have roles like capping chitin fibers as proposed for PMPs with one ChtBD2 domain (Tellam et al., 1999).

The RNAi studies presented here provide the first experimental evidence for essential roles for CPAP3 and some PMP and CPAP1 proteins in molting, structural integrity, survival and flexibility of limb joints. While the functions of many members of the PMP family and CPAP1 are unknown at present, it is clear that different CPAP proteins influence the physicochemical properties of different cuticle-forming structures. Additional studies are needed to determine the precise locations of individual members of each CPAP family of proteins and PMPs in wild type *T. castaneum* and in animals subjected to RNAi for specific genes. That information will provide further insights into their function.

5.2. Future directions

With good progress occurring in regard to functional analysis of the PMP and CPAP proteins from RNAi studies, we are better able to choose an appropriate target gene or protein associated with insect chitin metabolism to achieve selective control of pest insects. Potentially,

all of the proteins containing ChtBD2 domains are useful targets for selective insect control. Either over-expression strategies or antisense constructs introduced into crop plants may enable selective control of pest insects while sparing beneficial species. Future studies will be required to characterize more fully the role of each of these proteins in the chitin metabolic pathway. Results from these studies will provide critical insight into cuticle organization.

We are currently in the process of expressing and purifying recombinant representatives of each class of ChtBD2-containing proteins to address this issue. In the future, the availability of pure protein and molecular probes for specific ChtBD2 proteins will facilitate a better understanding of chitin metabolism and its regulation.

Specific approaches for future experiments include but are not limited to the following:

1. Express and purify TcPMP3 and TcPMP5-B, both of which contain a signal peptide and show lethal phenotypes after RNAi. Localization of representative proteins uniquely expressed in the anterior, middle and posterior mid-guts using immunostaining along with permeability studies with FITC-labeled beads of different sizes will provide experimental evidence for the hypothesis that expression of PMPs with different numbers of ChtBD2 domains helps to generate a gradient throughout the length of the midgut, which influences its permeability.

2. Determine the chitin-binding ability of these proteins, and also determine which crystalline form of chitin they bind. Does the physical state of chitin differ in the cuticle and PM?

3. Determine the tissue-specificity of the *T. castaneum CPAP1 and CPAP3* gene and protein expression in epidermis and cuticle by *in situ* hybridization and immunohistochemistry.

4. Express and purify *TcCPAP3* (the single CPAP3 gene required for pupal-adult molting). The chitin-binding ability of the respective proteins can be tested using the recombinant proteins expressed *in vitro*. Protein expression studies could help in understanding the biological significance of the presence of large numbers of CPAP3 and CPAP1 family members in *T. castaneum.*

5. Conduct permeability experiments on the gut after RNAi of *PMP3* and *PMP5-B* to gather information about the porosity and also to test the hypothesis that the larger the number of ChtBD2 domains in a PMP protein, the more posterior is its expression and localization.

6. RNAi of some of the *CPAP* genes does not result in any changes in chitin content, indicating that these are most likely structural proteins. TEM and immunogold staining may provide insight onto the structural deformities after RNAi and also provide information about the role of these proteins in cuticle organization.

7. Determine the secondary structure of the full-length protein or just the ChtBD2 domain by circular dichroism spectroscopy and NMR.

8. Some *CPAP* and *PMP* genes are embryo-lethal. It would be worth investigating the role of some of these proteins in the ovaries by whole-mount immunohistochemistry. Also *in situ* hybridization studies of some of the *CPAP* genes have revealed expression in embryonic tracheae in *D. melanogaster*. It will be worth investigating whether such defects are due to improper tracheal morphogenesis. Also the chitin/chitosan content of the embryo after RNAi treatment should be determined.

5.3. Bibliography

Arakane, Y., Muthukrishnan, S., (2010a). Insect chitinase and chitinase-like proteins. *Cell Mol Life Sci* 67, 201-216.

Bolognesi, R., Terra, W.R., Ferreira, C., (2008). Peritrophic membrane role in enhancing digestive efficiency. Theoretical and experimental models. *J Insect Physiol* 54, 1413-1422.

Dixit, R., Arakane, Y., Specht, C.A., Richard, C., Kramer, K.J., Beeman, R.W., Muthukrishnan, S., (2008). Domain organization and phylogenetic analysis of proteins from the chitin deacetylase gene family of *Tribolium castaneum* and three other species of insects. *Insect Biochem Mol Biol* 38, 440-451.

Kramer, K.J., Corpuz, L., Choi, H. K., Muthukrishnan, S., (1993). Sequence of a cDNAand expression of the gene encoding epidermal and gut chitinases of *Manduca sexta*. *Insect Biochem Mol Biol* 23, 691-701.

Tellam, R.L., Wijffels, G., Willadsen, P., (1999). Peritrophic matrix proteins. *Insect Biochem Mol Biol* 29, 87-101.

Zhu, Q., Arakane, Y., Beeman, R.W., Kramer, K.J., Muthukrishnan, S., (2008b). Characterization of recombinant chitinase-like proteins of *Drosophila melanogaster* and *Tribolium castaneum*. *Insect Biochem Mol Biol* 38, 467-477.

Appendix A - Deduced protein sequence

A.1.TcPMP1-A

>gi|268309026|gb|ACY95479.1| peritrophic matrix protein 1-A [*Tribolium castaneum*]

MKAVITLALLFHITVVTPRATIGDDITKHNNNNNNTIDNNNTYNNHSTYNNNNTINNN
NTYNNNNTINNNNTYNNHNTYNNNNNTYNNHNTINNNDPCTPDPKPDPR**CIDDSTEL**
WPHPADCAKYIECFHGNSYEMTCPPGLYFSSSSKTCVTADESECCKTSSQC

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.2.TcPMP1-B

>gi|268309028|gb|ACY95480.1| peritrophic matrix protein 1-B [*Tribolium castaneum*]

MKQTVFLILFVAFSHASPLIDYRLTLQPPKNCTSAGLLCETCSELVACVENNDHTFRKD
HVQTCPSGQKCVKGGCSSDSDPFCDGVADLAFP**CKKVGIFPDPFYCNKFVLCVDMGQ**
SRLQAYSSRCEDGLGYNIETGVCDVKLSGGCKVDEFPVPLCGKAGDSGALEGKPMEY
YTCEEYSSGKRVLYPVIDVCPNAETYENYKCVSK

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.3.TcPMP1-C

>gi|268309030|gb|ACY95481.1| peritrophic matrix protein 1-C [*Tribolium castaneum*]

MKSFVITLLGVSALAFASPPNHL**CPTVDGPDSVYIPHEDCHKFWQCSNGVAYLFNCS**
ATTVFDPSLNVCVHEWDYDCKPNPSPTLIPTKPTELPTEKPTEEPSEESSEVTTKPTRPS
KPTRPTKPSRPTRPSRPTRPSKPTRPSRPTRPSKPTRPPCSSTRPTRPRPTKPQPTKVPTKPP
TEKPTEKPTEPPTEKPTEEPTEPPTEKPTEEPTEPPTEKPTEKPTEEPTEPPTEKPTEKPTEEP
TEPPTEKPTEKPTEEPTEPPTEKPTEKPTEEPTEPPTEKPTEKPTEEPTEPPTEKPTEKPTEEP
TKPPTEKPTEKPTKPPTEKPTEKPTEEPTEPPTEKPTQPPTQKPTQPPTEKPTEETTERTSHE
SSPSNKPTPPPTRPTRPDSSESDDILHQVLNQIIIPGKAMNQIIIIPEKAMNQIIIILGKAMNQI
IIILEKAMNQIIIIPGKAMNQIIIIPEKAMNQIIIIPGKAMNQIIIIPEKAMNLTIITPEKAMNQII
IIPEKAMNQTIIIPEKAMNQIIIIPEKAMNQIIIIPGKAMNQIIIIPEKAMNQIIIIPEKAMNQIIII
PEKAMNLTIITPEKAMNQIIIIPEKAMNQTIIIPGKAMNQIIIIPGKAMNQIIIIPEKAMNQIIII
PEKAMNQTIIIPEKAMNPTIIIITIIRGSSGKEEH

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.4.TcPMP2-A

>gi|268309032|gb|ACY95482.1| peritrophic matrix protein 2-A [*Tribolium castaneum*]

MKGVVVFTILALALGVRAGDPL**CSGLPLDQEVLFPSPIDCSLYYKCYQGIFSEEKCP KGLYFSEYNKGCVEAQYSEC**MGGTGTPPSPTTSTTTTEPPTETPDPVDPR**CIDNETSY WPHVSVCHLYIECYGGKSYEMTCPPGTYFSDQHKKCVEASQSEC**CISDPSQCK

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.5.TcPMP2-B

>gi|268309034|gb|ACY95483.1| peritrophic matrix protein 2-B [*Tribolium castaneum*]

MKNQICSLVLVIFALSCARARQVATPDPGPT**CPYPSTEIIYFPYEGDCTKYWECYSGH SYLYTCPAGLWWHQEISECDYPGDFC**TDGTTQTDWTETTDSTPTIGPTTTNGDLPD**CT GTGDDPVYYPYPGDCTKYYECANGRLYTYNCPPDLWWHQEISECDYPGDFC**VPDT RGRN

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.6.TcPMP2-C

>gi|268309036|gb|ACY95484.1| peritrophic matrix protein 2-C [*Tribolium castaneum*]

RVRIK**MIKVLFFFGVTLCGVLA**DGNEIV**CPAVDPPTPVYFPHESDCSKFYECHDGTP HLLECPEGLDFNPELNVCDYPEQAGC**RGKTTSEISSSASPTTSTSHSDSTSSTSAKPSTP TPKPTTESTSNPTSHSKSTSTSSRPTHPTKSTHPTKHPTHSTSTRPTHSTTEKPTSTWKPTS STHKTTTDDNSSSDSDSDSSDSDDSSKSDDDSDSSDSKSDHHHHHDSNKSDDSDSSDSKS NHHHEHHNHHHNHHHDHNHHNHHGSDHHHQHHHNSDQSDDDLTESGDSKLDKNILY DILFGGIDGLAGEE**CTSDAISSVKAHPYDCTKYLNCYNGAYLVKTCPEKYMFHSIVRI CVPPSYASC**NK

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.7.TcPMP3

>gi|268309038|gb|ACY95485.1| peritrophic matrix protein 3 [*Tribolium castaneum*]

MKVFITILSLFSLGLASSPSDS**CPARDGAFPVYLPHEDCGKFYQCSNGVAYLQNCPP
GLHWNVAKLVCDWPRDAGC**EDKNEENSL**CPAVDGPFPVYLPHEDCGKFYQCSNG
VAHLFDCPAGLHWNVNKLVCDWPHDAGC**EGKNEETTTTTTTTSTTTTTTTTPEEITT
TTSKTTEGSGDDETTEVSKTTTDDNEGTTSSVTVTEETTTNDDNEESTTENPVTEETTTT
ATDDNEETTSSVTVTEETTTNDDNEESTTVDDITENPVTEETTKDDGSEEISTTEDDNGQ
ETTIDDNTTENPVTEDEEGSTSIDDEELTTTDDEDLTTTDDDSDVTEATIIDDDDKKTTEE
TVTEETTDNDEATEEPVTEDNSGNDETTDDGETTEEPVTEDNSGNDETTDDGETTEEPV
TEDDSENVTEDGSDEEPITTEKPTTTTTETPPPPTTTTTTTEKPTTTTTEKPTTTQEPTTQQ
QLSDEEIIAAI**CQASPDDIFLTAHPSNCQKYAVCMTGSYVIQTCPSGYHFSSSSMACIK
SSGGEC**DV

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.8.TcPMP5-A

>gi|268309040|gb|ACY95486.1| peritrophic matrix protein 5-A [*Tribolium castaneum*]

MNAFTLLVLFAVSAFAEKLESDPL**CAGVPPGSTYLFPYPGDCTKFYVCENGTKRVE**
DCPSGLWFNEALQACDHPDNSGCHPIV**CPPSIVDFYPYPEDCTKYIECYHGNPETHT**
CPDNLWFNSVEKRCTDPSSSGCGEHSSSVEPTWSTPNPI**CWGVLPGQTVLRPYPGDC**
NKFYECYGSRQTEMNCPPHLYFNEARQMCDWPDVSGCDDTTETPNPNPTSTITPPTT
PSGNDDPR**CANGNNDYWPDPDCTKFVECYHGHGYIMDCPSGLYFDSVDKKCEDPS**
EADCGRTTPTPDPWTTTKSSDWTNDPD**CPFPSADRYLFPYPGDCTKFLECWNGEKV**
AQECPAGLWFNPNLLVCDYPYHSGCKYGEEEQEV

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.9.TcPMP5-B

>gi|268309042|gb|ACY95487.1| peritrophic matrix protein 5-B [*Tribolium castaneum*]

MYFAVFFITIGLSSFGVNADPT**CPKQDGKDSVYFPHEDCTKFWQCSNGTPYLFDCP**
DNLHFNPKLNVCDWPNAAGCKGSGEDSDSSSSSSSSSSESQESGDNSQGKDDNNSSSS
SSSSSSSSSEEGPE**CPSVDGEDPVYFPHEDCTKFWQCSNGVPYLFNCSANLHFNPKL**
NVCDWPDQAGCESKEDSSSGSESKESDDKDDSSSSSSSSSSESQESGDNSQGKDDNNS
SSSSSSSSSSSEEGPE**CPSVDGEDPVYFPHEDCTKFWQCSNGVPYLFNCSANLHFNP**
KLNVCDWPDQAGCESKEDSSSSSSSSESKESGNNSESKDDNSSSSSSSSNSSSSEEGPE
CPSVDGENPVYFPHEDCTKFWQCSNGVPYLFNCSANLHFNPKLNVCDWPDQAGCE
SKEDSSSSSSSSSESGDNSQGKDDDKDDSGNSSSSSDSSSSSSSESSEEGPE**CPSVDGE**
TPVYIPHEDCTKFWQCSNGTPYLFDCPDNLHFNPKLNVCDWPNAAGCKGNK

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.10.TcPMP9

>gi|268309044|gb|ACY95488.1| peritrophic matrix protein 9 [*Tribolium castaneum*]

MTIVHNFIVFALCSIIYQGSVSAQTPIWPIPTS**CPATSPLTYTVHLQHDTDCTKFYKCD**
HGKKILFSCPSGLHFHPLFQVCDWPANVGCTHVPTPSPTTPSTTTPTPTTPSVDPG**CPF**
PGPLNYTVHLPHATDCTKFYKCDNGKKVEFDCPSGLHFNPVLEVCDWPAAAGCGT
TTPTPKPTPSPTTPSVDPG**CPFPGPLNHTVHLPHATDCTKFYKCDNGKKVEFDCPSGL**
HFNPVLEVCDWPAAAGCGTTTPTPKPTPSPTTPSVDPG**CPFPGPLNYTVHLPHATDCT**
KFYKCDNGKKVEFDCPSGLHFNPVLEVCDWPAAAGCGTTTPTPKPTPSPTTPSVDPG
CPFPGPLNHTVHLPHATDCTKFYKCDNGKKVEFDCPSGLHFNPVLEVCDWPAAAG
CGTTTPTPKPTPSPTTPSVDPG**CPFPGPLNHTVHLPHATDCTKFYKCDNGKKVEFDCP**
SGLHFNPVLEVCDWPAAAGCGTTTPTPKPTPSPTTPSVDPG**CPFPGPLNHTVHLPHAT**
DCTKFYKCDNGKKVEFDCPSGLHFNPVLEVCDWPAAAGCGTTTPTPKPTPSPTTPSV
DPG**CPFPGPLNYTVHLPHATDCTKFYKCDNGKKVEFDCPSGMHFNPVLEVCDWPG**
SAGCGTTPPTPRPTPSTTTPASSK**CATAPHNYHIPHATDCTKFYKCDHGIPVEFDCPPG**
LHFNARYQVCDWPASAGC

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.11.TcPMP14

>gi|268309046|gb|ACY95489.1| peritrophic matrix protein 14 [*Tribolium castaneum*]

MTIKRNFITLVLCFILYQAEAGDKEIWPIPTE**CPKTDPLDYTVHLQHESDCTKFYKCD**
HGGKVLFDCPAELHFNPVLQVCDWPWRANCTLNDKTTTTTPKPTPTPEPTPPADRDP
E**CPWPDPMDHTVHLPHETDCTKFYKCDNGKKVEFDCRDGLHFNKELQVCDWPQ**
NAGCQDNKPSPTPEPETPKPTPPEDRDPE**CPWPDPLDHTVHLPHETDCTKFYKCDHG**
KKVEFDCPDGLHFNKELQVCDWPQDAGCESNKPSSTPKPTTEPTPSETPEPETPKPTPP
EDRDPE**CPWPDPLDHTVHLPHETDCTKFYKCDHGKKVEFDCPAGLHFNKELQVCD**
WPGNAGCEDVKPDPTSKPTPEPTPSETPEPDTPEPTPPEDKDPE**CPWPDPLDHTVHLPH**
ETDCTKFYKCDHGKKVEFDCPDGLHFNKELQVCDWPQDAGCESNKPSSTPKPTPEP
TPSETPEPETPKPTPPEDRDPE**CPWPDPLDHTVHLPHETDCTKFYKCDHGKKVEFDCP**
AGLHFNKELQVCDWPGNAGCEDVKPDPTPKPTPEPTPSETPEPDTPEPTPPEDKDPE**CP**
WPDPLDHTVHLPHETDCTKFYKCDHGKKVEFDCPAGLHFNKELQVCDWPGNAGC
EDVKPDPTPKPTSEPTPSETPEPDTPEPTPPEDRDPE**CPWPDPLDHTVHLPHETDCTKFY**
KCDHGKKVEFDCPDGLHFNKELQVCDWPGNAGCEEVNTDPTEDPTSSTDTPKPTPPE
DRDPE**CPWPDPLNYTVHLPHETDCTKFYKCDHGKKVEFDCPDGLHFNKELQVCD**
WPGNAGCEEVNTDPTDDPTSSTETPKPTPPEDRDPE**CPWPDPLNYTVHLPHETDCTKF**
YKCDHGKKVEFDCPDGLHFNKELQVCDWPGNAGCEEVNTDPTEDPTSSTDTPKPTP
PEDRDPE**CPWPDPLNYTVHLPHETDCTKFYKCDHGKKVEFDCPDGLHFNRELQVC**
DWPGNAGCEEVNTDPTEDPTSSTDTPKPTPPEDRDPE**CPWPDPLNYTVHLPHETDCT**
KFYKCDHGKKVEFDCPDGLHFNKELQVCDWPGNAGCEEVNTDPTEDPTSSTDTPKP
TPPEDRDPE**CPWPDPLNYTVHLPHEIDCTKFYKCDHGQKVEFECPDGLHFNPELEV**
CDWPESAGCEDPIPEPCPSDDTDEPEPQPEPTPPSDLDPE**CPWPDPLNYTVHLPHEEDC**
TKFYKCDNGKKVEFDCPDGLHFNPDLEVCDWPENAGCENQSIQCPDTQCQDTTLY

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

.

A.12.TcCPAP3-A1

>gi|268309018|gb|ACY95475.1| cuticular protein analogous to peritrophins 3-A1 [*Tribolium castaneum*]

MKFALIALVLVSAANAQFK**CPPKDGQYEDPRQCDKYYECEEGVAREKLCPDGLVF DPLIRKINKCDQPFNVDC**GDRTELQPPKPNHF**CPRRNGFFAHPDPAVCNKFYNCIEG EHTEITCTAGLHFDEFTGTCVWPDAAGRQGC**NKDVTNKLKDGFE**CPKDGQTDANG QLVVHPKYAHPTDCQRFYVCLNGQEPRDLGCQVGEVYNEESQRCDAPENVPGC**EDWYKDEPAPAKPAKKV

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.13.TcCPAP3-A2

>gi|268309020|gb|ACY95476.1| cuticular protein analogous to peritrophins 3-A2 [*Tribolium castaneum*]

MKTVTVALLIFVTGSSAQFK**CPDRTGFFPDPVQCDLYYVCSKGEYEEKLCPDGLVF DARDPNHERCDIPANVDC**DERTELQEPHPSPG**CPRANGYYRHSDPLACDKFFNCVNG VPHELPCPPGLIYDDTASTCAWPDDSHRKDC**KNAKRDKLDDGFT**CPDEEILGPGGR KLPHPTFAHPEDCGKFYICRNGVMPQKGQCVKGLVYNEETFTCDDPKNVPGC**EDY YEKAEKSKTKKA

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.14.TcCPAP3-B

>gi|119387886|gb|ABL73928.1| obstructor B [*Tribolium castaneum*]

MTRLNITILLLLTIGLAAAARKHQKQESRDEEYEATDQ**CPEKYGFFADAEQCDKYYE**
CNDGQITEKLCPDGMVFNDYSSEYEKCDLPFNIDCTSRPKLQEPQPSQH**CPRKHGYF**
AHEEPHICDKFYYCVDGKYNMITCPNGLVYNDKAGICSWPDEAKKKGCSSEEVFQF
E**CPKVNETVAATHPRYADPDDCQYFYVCINGDTPRRSGCKLGQVFDDVGKKCDW**
VRNVPECADWYKGRLTDEQLKELENPPTPKPRPTKVSRRKPRPPRPTQVEEEEEK

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.15.TcCPAP3-C5a

>gi|119387888|gb|ABL73929.1| obstructor C1 [*Tribolium castaneum*]

MKSYACLLVLSALIYSVLGQENFK**CPDDFGFYPHHTSCDKYWKCDNNVAELKTCG**
NGLAFDASDPKFLTENCDYIHNVDCGDRTQLEPPISSPH**CERLYGIFADESKCDVFWN**
CWNGEASRYQCSPGLAYDREARVCMWADQVPECKNEEVAGGFT**CPAPGEVSNSGS**
FSRHAHPDDCRKYYICLEGTAREYGCPIGTVFKIGDADGTGNCEDPEDVPGCEDYY
GDLDLKSIRKSELLAGLQSSGSSRSHQPATKSKPRPAPASRNAPEPADNN

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.16.TcCPAP3-C5b

>gi|119387890|gb|ABL73930.1| obstructor C2 [*Tribolium castaneum*]

MKSYACLLVLSALIYSVLGQENFK**CPDDFGFYPHHTSCDKYWKCDNNVAELKTCG**
NGLAFDASDPKFLTENCDYIHNVDCGDRTQLEPPISSPH**CERLYGIFADESKCDVFWN**
CWNGEASRYQCSPGLAYDREARVCMWADQVPECKNEEVAGGFT**CPAPGEVSNSGS**
FSRHAHPDDCRKYYICLEGTAREYGCPIGTVFKIGDADGTGNCEDPEDVPGCEDYY
KDVDLKALKKLGY

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.17.TcCPAP3-D1

>gi|268309022|gb|ACY95477.1| cuticular protein analogous to peritrophins 3-D1 [*Tribolium castaneum*]

MRVFLFVAISILSANAGVVLQDAPS**CPEQHGVQAYAHPESCNLFFLCTNGTLTVEQC ENGLLFDGKGAVHHHCNYHWAVDC**GHRKADLTPISTPG**CEYQFGIYEESHGCSTHY IKCAHGEPIPQECEPGLVYDERIHGCNWPDLKLEVC**NPEAVVGFK**CPTKVPSNSPAA KFWPYPRFAVPGDCHRLITCVNGFPRLISCGEGKAFDQHSLTCEEPELVPHC**ANHIR K

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.18.TcCPAP3-D2

>gi|119387892|gb|ABL73931.1| obstructor D [*Tribolium castaneum*]

MRSAVCVLVLAFFGCVRAQQLKGNQEDP**CKLKSKVVGDETYCDRYWECVNGQQE**
LYDCPNGLVFAGKNRGVTEGCDYPWRSNYCDNKQQANPPISTEHCDWLYGIFGHET
SCTRYWTCWNGTATEQLCIGGLLYNERTHSCDWPENVDGCQKHPL**CNDDPNGNVPLG**
KSCNRYWQCQGGYPRLQRCPAMLVFDRRSLRCVVPPTEDCDVPSTIPPPPPEEDDDS
NIPQNNPKNKQHNFNLPPGAIPVPDKNRPRN

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.19.TcCPAP3-E

>gi|268309024|gb|ACY95478.1| cuticular protein analogous to peritrophins 3-E [*Tribolium castaneum*]

MTREDLSSIVTMETISIIVTIVTLIAIQGLAQRNLGPSS**CPEKNGRYPTSTCDGYIECR DGLAEEKLCPDGLLFNPASGPQAFPCQYPLDVDC**TGREQTQPAQATDE**CPHQFGYF RMGDATSCGQFKNCVDGRGFIFDCPEGLAFNGDTYRCDWPDQVATC**DAEAFLGFT **CPNDGRSFGLGEAEFRFFRSPNDCQRYFVCVNGRPRLYNCGEGRAFNDLIGACDGV ENVTGC**VGGAQGGANFRNYRI

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.20.TcCPAP1-A

>gi|268309000|gb|ACY95466.1| cuticular protein analogous to peritrophins 1-A [*Tribolium castaneum*]

MAISSTTDWSKIRYIMFVLFFISALSASTVATGLTTSGKQKRQLFREHLLPQHGGKIPD
GAFDPHRISLNKLSQNGIFSPDGYHLAQRHPDRHHNPKQHLPNTLKVHHYPDNRFVAPQ
GQQHGVQTIDTTYLIREPVQYPLFPPLRHYTPLPFFPQYSLKLLRHFKYDKVNDHHDGS
WNEIELPTKHQTYEGYAVGPKLRQHEKEQQQDLSKIPGIPGKDYPLFHSVPPTSFS**CKHV**
PAHPGIYANVETGCQAYHVCHDGREGEQGASFLCTNGTIFNQAEFACDWWYNVN
CHEAPNLYRLNLDPAKNPYVPKRKPEEELHFVPVEKY

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.21.TcCPAP1-B

>gi|268309002|gb|ACY95467.1| cuticular protein analogous to peritrophins 1-B [*Tribolium castaneum*]

MNYLTCVLHTVIVILGYFLITDGFEQKSSVKKARASAPPPDEPSNNVGSEYDERNNAL
DTKYVPPGFVSPSVLEYLELGKSIPGRPGTDYPILGRVPYTNFY**CDEQPYPGFYADEET**
RCQTWHYCDIDGRQTTFLCPNGTQFSQLVFVCDWWFNVRCDLSSKLYVINSRLYGR
PKLDPTRPHRTITKQLLEEIFNDE

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.22.TcCPAP1-C

>gi|268309004|gb|ACY95468.1| cuticular protein analogous to peritrophins 1-C [*Tribolium castaneum*]

MWRLLGFGVVLAAVAAPNNQVPKTGFS**CEGRTTGYYADVESGCQVYHMCDGLGR QFSYTCPNATLFQQRMLICDHWYMVNC**SKSVDDYTANLRIGHKEMPFVDDNEANPY HRTPRPDLLSHPSQSEYDIIYRTGRAQLGTNLNLVGAESDPKNSTTSTTDEPAYSLPSHW STEYNKQVTTTKPKPAKKGKPTTVRVNYQSNYKATTPVFPQSVEVTEAPDLELVPPSGS TESSVNFESRFKATTPVFPLSVDVTEAPNLGLLPPLPYQLNDTKPNENLVVSTVNFESKF KATTPVYPKSVELTSEEPSEVGVLPPKDTNSTETDVVPPLSIDFEPPVVDSNFKPGDHSEEI PPELPSKFYQPPKFEPDYTELAKLKGGFKALSSGEWEDLRKKFLIPDYEFPLETVSRPSYD SVLSSFQVGPVPSEKE

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.23.TcCPAP1-D

>gi|268309006|gb|ACY95469.1| cuticular protein analogous to peritrophins 1-D [*Tribolium castaneum*]

MKALVVACLLILSVHGGHYIIPGHSPYDAYHDLHLPHSPPLYPTLASVPPTGFT**CLGRN PGYYADIETGCQAYHRCEYNSAASFLCTNGTLFNEQFQVCDQFYNVRC**GSPYIDL

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.24.TcCPAP1-E

>gi|268309008|gb|ACY95470.1| cuticular protein analogous to peritrophins 1-E [*Tribolium castaneum*]

MLRIVAVSVVCFILTDCHKIHKRWVDIENATFTFD**CTNRAIGFYADVEYDCQIFHMC DPEGRRIPHVCANDTSFNQEYRVCDWENNFDC**SEAPKWYFLNELTYATDPPKDVED DYHDNN

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.25.TcCPAP1-F

>gi|268309010|gb|ACY95471.1| cuticular protein analogous to peritrophins 1-F [*Tribolium castaneum*]

MSLLPHVFVAVLCGFTVHLVFGQLDGYIPGQDYPIYSEVPQGLSFR**CDQRLPGYYSDP EAQCQVWHWCLPSGQQYSFLCPNGTIFNQFARVCDWWFNVDC**AGTPSLYGINEDL YRIPEYNHHQQ

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.26.TcCPAP1-G

>gi|268309012|gb|ACY95472.1| cuticular protein analogous to peritrophins 1-G [*Tribolium castaneum*]

MTMLYLTIVPLVIMAHLIHGFPKTVDPELPNYPRQVHSRGFNQKAPSPFTSQVQEYED
TRPSATQQQPANYPFVYQNAIPVKQKVQQVVRGNKPPQYANQKDALREKNLKELEEEE
IEEPDRLSQLLPNSKFV**CDGKNTGYYADEDLGCEVFHYCQDNAKHSWICPEGFTFHQ**
VHLICMPPGGDNICEKSSQFHFVNDYLYKPVNLEEYQQKPNVSLRYSDRFYPEHYYEE
REVDPENRQPNQHRNSVRVHQQEEPEQVHRPTFRPEPALSQVFRSPNEVNIPLQQRRPQ
YLTQRFFSQK

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.27.TcCPAP1-H

>gi|268309014|gb|ACY95473.1| cuticular protein analogous to peritrophins 1-H [*Tribolium castaneum*]

MKRPPGLWILLSAGVLWLSLQDVYGIRATALIFGKSTTTTTSTEATSEAETSPDETATE
ASTESSNVTKPTLTGVPQVDYIWDPNLPKELNGYNLSEYPFYERVPEDIDFK**CDGLHDG**
FYASVPHKCQVYHHCLFGTRYDFLCANYTAFDQKTFICHFVSEVDCVNSKKYWHR
NDALYKAATTTTAKPVVVYTPPPQPHPAAAPAAVNPRRPGGRRRPMRRRRPQYDYYD
YADDYDDYYEERPPRSEARRRKRPRPRPRPVYEDDYEEYEDERYERRGPGRRRAGDRR
RNNNKDRRKYEDEDKYEDDKLDDDRKSNDRRPNNKDRRKNEDEEDDKLEEDRKTEDK
RTSERRHHKRRPPHDYSEEEEEDRRRSRPKGGSRRPVEDDYEEEEAPQNDKPSAPENVG
DKPLIKPTSGSSIYDRPRAAPKIKPPVPKNEASKYAYKPVVTKAPPVEDEEYYDDYEETP
PKTHRKNERHTTRREEERERKPSPPRLRASQSSRDDVSYEEDVPRRSQKPKYGSHHRERP
RNKKPPPPVEDYYDDYEEESKETPAKAEKPKEKSTTTTTTTTTTTTTTTTPTTTTTTTTT
TTPPPPSTTTERELRPESIIRIVKRPFLPSRGGNPYGPRGIQPVGAKAVEKKPEIKEEETSES
AKAAFKPSVVIKTPLRQKYVNEPQEEFAPVSSRPQTPPRPSLKPSKEFLGPRTTQKPKLPE
PDPLDIEDEYDVTLNDALNPTLPNLPVRGFPTGFSNANNEYNNYNSYQRPRYVLEGANS
DYYQAKPTRQRYEAVPYSNPQEHFVNSEADYKGQYSTVNRANFRPRLTQALYTSF

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.28.TcCPAP1-I

>gi|268373732|gb|ACZ04319.1| CPAP1-I [*Tribolium castaneum*]

MKGVSFIFIGLLGFSCAQFLNNRPFPTYSLENMPDTEFS**CRDKILGGYYADVDTMCQ**
MFHICVKVAGVGVQDFRFLCPNGTAFDQDHQICAEWEDVDCDASTLYYSSDNFDL
YRIGSGFESKAVKYGEDEETFALQRAETGDARLNRDHQAQRVNQQKEQAFRKNRPQN
NNNNNNDREIFKGSSSSNFFNNRNGGKETDDDYDDNTNANQNNDNFQKKKLLRKQHR
RPVQNDNNNQQNRRPQQQNNRPQNNQQQNNRPQQQQPNNQQQPNNRAQNNNNQQQ
QNNRPQNNNQQQQNNRPQNNQQQNNRPQQPNNYQQQNNRPQQQQPNNYQQQNSRPQ
NNNQQQQNNYQQQNNRPQSNNQQQQQPNNYQQQNNRPQSNNQQQQQSNNYQQQNN
RPQNNNQQQQNNRPQQQYVEAQTYRPNRPTGFANNFAGSSYVPTTTRPTTTTFPETVKP
RNGQRQYNNADNFRQNQNSPTPSYPLSTPAPFKQTQQYNSPPNRQTTQNNYNNGQYNP
TTQAPQRETENYPANFQKQNYNKATENYPQRTQTPRNNNNNNNQFNSKPTEVNNRNQ
YTTPVPKVTQYNTHYDVPKIKPTKQNDNYPTTQIKNDYSTQFSKKPTTPFKQNDNYPTT
FAPKFNSNNNNFPSTFAPRTNNAYTQISQQTTFYNSNNNQQQYSTPRTTSFTQYTPTVPKI
TPTTPVARSNRFDETQYDDGSYNPKYDRNDDEFLKTAHSQNIASSRNEYSKSTKNVQST
QKPQYESPRPFSVSPNTPKATDFPKSTPKPVNNVKKVKDVSYDYAYYDTNVGSEPEYEI
DTEIKKANVKN

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

A.29.TcCPAP1-J

>gi|268309016|gb|ACY95474.1| cuticular protein analogous to peritrophins 1-J [*Tribolium castaneum*]

MARYHLLKLGLAILVYFTRSMGCLETEERKFNEEDLVFYQGVTGKPGVDFPVLSHIPR
TTFN**CRNVDSGYYADLETDCQVFHICEEGKKISFLCPNGTIFQQSELICEWWFKVNC**
TDSPNFYEESAEQLREDNARRKASRRVLQNHGAVMRTEEHSSVSASQNGRKFNVGTTF
QETNDLARAPSSDFVRDQKKKGRKPVQFQTQSIEERKSFDDQKVVPSKNHRANERINQN
GAVPQHQEVRNNSARSYNQFGTTRQEVVPNNQRNGGRKVVQGSQDQNQKFETNAKGF
GTSVVPNRQEKEQHSNPNFQQYSNFHNQNSGFGTTSDFKKMVPSQFNTQQSQQERKNI
KTFGTKEVVSNQFNNGQSAQNTFSNSTNREEEINPFFNNGFQDSNTNSAESTDKKSNVS
NFSQSFSHRNFNVDEKRTTFSSSTTTTPRSTTKVPPTSSNLQNHNQDYHQIKNLNIEGKG
GTIKQTSTVPLTRNDLVNRRLEPNDSDESQITQETSSFYKNSFNRIDQTIKTTSNPHPNHRS
TYSEIKEPKFHSTKAYTSLSSVTPQYLTTQPINNGRPFVGSQVGTTLRGFGTKPPTTTPVV
PTEIFNVGRTDKSLKPIALSTLSTFTPLPPLDLTVPSQGSFNTGTPSPKTESSTRVIDHATVY
GKFVKSGATTPTTGRTKILLSKSVFGTTPKTSAEILSTSDSYNRGNSQVRFFGKVQTTTKI
PGVQIISKSDGTTTKPAEILSTSNSYNRGNSQVQFFGKVQTTANLNSISTTQKPFSSSTPTV
LPLSFGTTSSTIGTTVKPTVPGFFVTKNVIPKSEIARPRPFQLPPEPDFGQPIFGTTQNLVTT
LKPTYPTYKISSASPFSGPSTPEPTLPTPFATTQVLPSTFETVDSMITALAEMANSNNYTEN
PRPGLVIPPSAGPQTLHTLAQYFANALDGIATEKEDEEDGDEENLTKVELEKKKLTELLT
QMTMNRYNELFPEDESTTTVVPNDLEGQHSKSPATPAPKVRQLAKVFTQALSSYLDDP
ATFKKVLEEVRPTEPPALDETTDLPDDELLNFSDADTKSSYSPFFPPKLPAVRPTWGYLIA
YNTTPHNLDVKNSIGGTEENLQGADSQSFVSQFNKLQNQKESTTRSSVLLPEDHWTTSP
DATKLWKTAFSFDPASINHNFGTTEPTIGTTTEEVVPAREVRYELHALPKLDLNSTQVHG
ILIDFMNTTKQDEDNRLQRILRKLNTTEDEFLAKMKEIEANPLTRRLILLLISECGNSTQA
VPLHTLLIQNSTEVLQTSSSTHGFKNIDPSLDNDSQDARALQLLNSLYTIASRFGK

Colored in **red** is the signal peptide and in **blue** is the ChtBD2 domain.

Appendix B - Accession numbers of proteins

Table B-1 Accession numbers of CPAP3 orthologs from different insect species

Protein	Species	Accession No.
Cq-B	*Culex quinquefasciatus*	XP_001847921.1
Ag-B	*Anopheles gambiae*	XP_318884
Ap-B	*Acyrthosiphon pisum*	XP_001947981.1
TcCPAP3-B	*Tribolium castaneum*	EF125544
DmObstB	*Drosophila melanogaster*	NP_609339
Nv-B	*Nasonia vitripennis*	NP_001136346
Am-B	*Apis mellifera*	XP_392261
Aa-577	*Aedes aegypti*	XP_001660249.1
Ag988	*Anopheles gambiae*	XP_309182
Cq-A2	*Culex. quinquefasciatus*	XP_001842245
Aa-917	*Aedes aegypti*	XP_001651678.1
Ag987	*Anopheles gambiae*	AGAP000987
Pc-A	*Pediculus humanus corporis*	XP_002429481.1
Ap-A1	*Acyrthosiphon pisum*	NP_001156724.1
Nv-A1	*Nasonia vitripennis*	XP_001607914.1
Am-A	*Apis mellifera*	XP_001120217
DmObstA	*Drosophila melanogaster*	NP_608378
TcCPAP3-A1	*Tribolium castaneum*	GU128092
Cq-A1	*Culex quinquefasciatus*	XP_001842244.1
Aa-245	*Aedes aegypti*	XP_001662353.1
Ag-989	*Anopheles gambiae*	XP_309184
Nv-A2	*Nasonia vitripennis*	XP_001608254.1
TcCPAP3-A2	*Tribolium castaneum*	GU128093
Acp-A2	*Acyrthosiphon pisum*	XP_001947844.1
DmObstE2	*Drosophila melanogaster*	NP_723116
Phc-E	*Pediculus humanus corporis*	XP_002427031.1
TcCPAP3-E	*Tribolium castaneum*	GU128095
Ap-E	*Apis mellifera*	XP_397120
Nv-E	*Nasonia vitripennis*	XP_001603894.1
DmObstE1	*Drosophila melanogaster*	NP_608957
Ag-E	*Anopheles gambiae*	XP_310082
Cq-E	*Culex. quinquefasciatus*	XP_001845871.1
Aa-E	*Aedes aegypti*	XP_001662027.1
Ap-C	*Acyrthosiphon pisum*	XP_001944961.1
Pc-C	*Pediculus humanus corporis*	XP_002425178.1
DmObstC	*Drosophila melanogaster*	NP_649611
TcCPAP3-C	*Tribolium castaneum*	EF125545
Am-C	*Apis mellifera*	XP_001121720
Aa-C	*Aedes aegypti*	XP_001663971
Ag-C	*Anopheles gambiae*	XP_319536
Cq-C	*Culex quinquefasciatus*	XP_001848925
Cq-D2	*Culex quinquefasciatus*	XP_001868521
Aa-D2	*Aedes aegypti*	XP_001662442
Pc-D2	*Pediculus humanus corporis*	XP_002429485
Ap-D2	*Acyrthosiphon pisum*	XP_001947475
Nv-D2	*Nasonia vitripennis*	XP_001607911
TcCPAP3-D2	*Tribolium castaneum*	EF125547
Pc-484	*Pediculus humanus corporis*	XP_002429484.1
Pc-482	*Pediculus humanus corporis*	XP_002429482
Ap-D1	*Acyrthosiphon pisum*	NP_001156306
Am-D	*Apis mellifera*	XP_393551
Nv-D1	*Nasonia vitripennis*	XP_001608252
TcCPAP3-D1	*Tribolium castaneum*	GU128094
DmObstD	*Drosophila melanogaster*	NP_523418
Ag-D	*Anopheles gambiae*	XP_560209
Aa-D1	*Aedes aegypti*	XP_001660247
Cq-D1	*Culex quinquefasciatus*	XP_001842247
C.elegans	*Caenorhabditis elegans*	NP_490942

Table B-2 Accession numbers of CPAP1 proteins from different insect orders and sub-phylum Chelicerata; Crustacea.

Accession No.	*Species*
ACY95466	*Tribolium castaneum*
XP_002432783	*Pediculus Human corporis*
EGI66437	*Acromyrmex echinatior*
XP_001864593	*Culex quinquefasciatus*
EFN79354	*Harpegnathos saltator*
XP_001651161	*Aedes aegypti*
EFR20404	*Anopheles darlingi*
XP_321953	*Anopheles gambiae*
XP_002007969	*Drosophila mojavensis*
XP_002047293	*Drosophila virilis*
XP_001956305	*Drosophila ananassae*
BAH71395	*Acyrthosiphon pisum*
XP_003245024	*Acyrthosiphon pisum*
XP_002084275	*Drosophila simulans*
XP_001984178	*Drosophila grimshawi*
NP_729504	*Drosophila melanogaster*
XP_002029887	*Drosophila sechellia*
XP_001603626	*Nasonia vitripennis*
XP_001972181	*Drosophila erecta*
XP_002024850	*Drosophila persimilis*
XP_001122348	*Apis mellifera*
EFN79353	*Harpegnathos saltator*
XP_003242385	*Acyrthosiphon pisum*
XP_001602012	*Nasonia vitripennis*
XP_002401342	*Ixodes scapularis*
XP_001986583	*Drosophila grimshawi*
EFN79506	*Harpegnathos saltator*
XP_394817	*Apis mellifera*
XP_002005041	*Drosophila mojavensis*
XP_002034387	*Drosophila sechellia*
XP_003249744	*Apis mellifera*
XP_002066220	*Drosophila willistoni*
XP_002050437	*Drosophila virilis*
EFR20523	*Anopheles darlingi*
NP_001161901	*TcCPAP1-J*
EFX69619	*Daphnia pulex*
XP_001990219	*Drosophila grimshawi*
XP_395554	*Apis mellifera*
XP_001974615	*Drosophila erecta*
>NP_611292	*Drosophila melanogaster*
XP_001998961	*Drosophila mojavensis*
XP_001599617	*Nasonia vitripennis*

XP_002056227	*Drosophila virilis*
XP_002073535	*Drosophila willistoni*
XP_001663318	*Aedes aegypti*
EFN71383	*Camponotus floridanus*
XP_001960113	*Drosophila ananassae*
XP_001997366	*Drosophila grimshawi*
EFZ17788	*Solenopsis invicta*
XP_001358530	*Drosophila pseudoobscura pseudoobscura*
XP_002016349	*Drosophila persimilis*
XP_002013846	*Drosophila persimilis*
EFN79508	*Harpegnathos saltator*
NP_725747	*Drosophila melanogaster*
XP_001954100	*Drosophila ananassae*
XP_002138424	*Drosophila pseudoobscura pseudoobscura*
EGI66500	*Acromyrmex echinatior*
XP_001862964	*Culex quinquefasciatus*
ABJ16987	*Drosophila melanogaster*
XP_002082025	*Drosophila simulans*
EFX89837	*Daphnia pulex*
XP_002096211	*Drosophila yakuba*
XP_001951383	*Acyrthosiphon pisum*
EGI66405	*Acromyrmex echinatior*
NP_650734	*Drosophila melanogaster*
XP_001809190	*Tribolium castaneum*
EFX89836	*Daphnia pulex*
XP_002038212	*Drosophila sechellia*
EFN63917	*Camponotus floridanus*
XP_001979626	*Drosophila erecta*
XP_002102930	*Drosophila simulans*
XP_310217	*Anopheles gambiae*
EFR28957	*Anopheles darlingi*
EFN71381	*Camponotus floridanus*
EFN81809	*Harpegnathos saltator*
XP_001648775	*Aedes aegypti*
EFX79823	*Daphnia pulex*
XP_001953230	*Drosophila ananassae*
EFX79931	*Daphnia pulex*
XP_001604297	*Nasonia vitripennis*
XP_002070818	*Drosophila willistoni*
XP_001841755	*Culex quinquefasciatus*
EFZ17703	*Solenopsis invicta*
XP_001359749	*Drosophila pseudoobscura pseudoobscura*
XP_002019497	*Drosophila persimilis*

XP_002434102	*Ixodes scapularis*
XP_002096679	*Drosophila yakuba*
ACN94700	*Drosophila miranda*
EGI66499	*Acromyrmex echinatior*
XP_002070819	*Drosophila willistoni*
NP_001097699	*Drosophila melanogaster*
XP_002427953	*Pediculus humanus corporis*
EFX76357	*Daphnia pulex*
TcCPAP1-B	*Tribolium castaneum*
TcCPAP1-C	*Tribolium castaneum*
TcCPAP1-D	*Tribolium castaneum*
TcCPAP1-E	*Tribolium castaneum*
TcCPAP1-F	*Tribolium castaneum*
TcCPAP1-G	*Tribolium castaneum*
TcCPAP1-H	*Tribolium castaneum*
TcCPAP1-I	*Tribolium castaneum*

Table B-3 Miscellaneous genes containing ChtBD2 domain.

GeneBank accession #	Glean #	cDNA length (bp)	Region within the chromosome	No.of Exons	ORF (aa)	M.W (kDa)	No.of ChtBD2	p*I*	LG	Remarks
XP_972309	02107	1549	17,640,633 to 17, 652, 675	6	516.0	59.2	1	8.9	U	PKC domain
XP_001809190	02058	1767	6,743, 929 to 6, 748, 458	10	588.0	64.4	1	4.5	U	Incomplete Consensus
XP_972801.1	09553	1065	3187764 to 3188958	2	354.0	40	2	9	7.0	Incomplete Consensus
XP_001807896	11724	3675	19489041 to 19482018	4	1224.0	138.1	2	5	9.0	Incomplete Consensus
XP_973295.1	13568	1374	4579662 to 4574340	5	457.0	50.6	1	8.9	1=X	no signal Peptide
XP_971785	00588	1353	11710215 to 11708394	3	450.0	52	1	8	2.0	Incomplete Consensus
XP_971239	12551	11784	15696741 to 15727414	34	3927.0	437.6	1	5.6	9.0	VFW, hemolectin
XP_971219.2	15245	4747	1934021to 1923064	9	1458.0	155	7	6.4	6.0	4 kb gaps
XP_971761.1	05468	669	15809107 to 15808439	1	222	24	1	4.5	8	Incomplete Consensus
XP_971708.1	05469	1099	15806342 to 15805623	2	223	24	1	4.7	8	Incomplete Consensus

PKC: Protein Kinase domain; VFW: Von Willebrand domain; the grey shaded genes are partially cloned.

Appendix C - Bacterial strain, cell line, antibodies and kits

Table C-1 Bacterial strains and insect cell lines

Strain	Genotype/Origin	Application
Maximum Efficiency ® DH5α	F- \80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rk-, mk+) phoA supE44 ⌊- thi-1 gyrA96 relA1	Amplification of plasmid DNA
Rosetta 2 (DE3)	F– *ompT hsdS*$_B$(r_B– m_B)–) *gal dcm* (DE3) pARE2[3] (Cam^R)	Expression of recombinant proteins
sf21	Pupal ovarian tissue of *Spodoptera frugiperda (Vaughn et al., 1977)*	Expression of recombinant proteins
High-five	Ovarian cells of *Trichoplusia ni (Davis et al., 1992)*	Expression of recombinant proteins

Table C-2 Primary antibodies

Antibody	p/m	Host	Dilution	Reference
Anti-CPAP1C	Polyclonal	Rabbit	1: 1000	(Jasrapuria, unpublished)
Anti-Hexa- His	Monoclonal	Mouse	1: 1000	Genscript

Table C-3 Secondary antibodies

Antibody	Conjugate	Host	Dilution	Manufacturer
Anti-rabbit	AP	Goat	1: 1000	Promega
Anti-rabbit	Alexa 488	Goat	1: 1000	Invitrogen
Anti-mouse	AP	Goat	1: 1000	Promega

Table C-4 Kits

The following kits were used according to the manufacturer's protocol

Kit	Application	Manufacturer
RNA-easy Mini kit	**Total RNA isolation**	**Qiagen**
SuperScript III First-Strand Synthesis System	**First strand cDNA synthesis**	**Invitrogen**
QIAquick Gel Extraction Kit	**Purification of DNA from agarose gels**	**Qiagen**
QIAprep Spin Miniprep Kit	**Plasmid purification**	**Qiagen**
DIG DNA Labelling Mix	**DNA-Probe synthesis**	**Roche**
AmpliScribe T7-Flash transcription kit	**Synthesis of dsRNA**	**Epicentre Technologies**
4-12% Bis-Acrylamide	**SDS-gel run and stain**	**Invitrogen**

C.1. Bibliography

Davis, T.R., Trotter, K.M., Granados, R.R., Wood, H.A., (1992). Baculovirus expression of alkaline phosphatase as a reporter gene for evaluation of production, glycosylation and secretion. *Biotechnology (N Y)* 10, 1148-1150.

Vaughn, J.L., Goodwin, R.H., Tompkins, G.J., McCawley, P., (1977). The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae). *In Vitro* 13, 213-217.

Appendix D - Permission to release copyrighted material

In

Chapter 2

Genes encoding proteins with peritrophin A-type chitin-binding domains in *Tribolium castaneum* are grouped into three distinct families based on phylogeny, expression and function.

Jasrapuria, S., Arakane, Y., Osman, G., Kramer, K.J., Beeman, R.W., Muthukrishnan, S., (2010).

Insect Biochem Mol Biol 40, 214-227.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Aug 11, 2011

This is a License Agreement between Sinu Jasrapuria ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Sinu Jasrapuria
Customer address	Kansas State University Manhattan, KS 66502
License number	2725930652541
License date	Aug 11, 2011
Licensed content publisher	Elsevier
Licensed content publication	Insect Biochemistry and Molecular Biology
Licensed content title	Genes encoding proteins with peritrophin A-type chitin-binding domains in*Tribolium castaneum*are grouped into three distinct families based on phylogeny, expression and function
Licensed content author	Sinu Jasrapuria,Yasuyuki Arakane,Gamal Osman,Karl J. Kramer,Richard W. Beeman,Subbaratnam Muthukrishnan
Licensed content date	March 2010
Licensed content volume number	40
Licensed content issue number	3
Number of pages	14
Start Page	214
End Page	227
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic

Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	TRIBOLIUM CASTANEUM GENES ENCODING PROTEINS WITH THE CHITIN-BINDING TYPE II DOMAIN
Expected completion date	Aug 2011
Estimated size (number of pages)	241
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
Total	0.00 USD