



Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

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

Sinu Jasrapuria^a, Yasuyuki Arakane^a, Gamal Osman^{b,d}, Karl J. Kramer^{a,c},
Richard W. Beeman^c, Subbaratnam Muthukrishnan^{a,*}

^a Department of Biochemistry, Kansas State University, 141 Chalmers Hall, Manhattan, KS 66506-3702, USA

^b Advanced Genetic Engineering Research Institute, Giza, Egypt

^c USDA-ARS, Center for Grain and Animal Health Research, 1515 College Avenue, Manhattan, KS 66502, USA

^d Biology Department, Umm Al-Qura University, Makkah, Saudi Arabia

ARTICLE INFO

Article history:

Received 27 October 2009

Received in revised form

25 January 2010

Accepted 29 January 2010

Keywords:

Chitin
Chitin-binding domain
Cuticle
Epidermis
Gasp
Midgut
Peritrophic matrix
Peritrophin A domain
Obstructors
Phylogenetic analysis
Red flour beetle
Tribolium castaneum

ABSTRACT

This study is focused on the characterization and expression of genes in the red flour beetle, *Tribolium castaneum*, encoding proteins that possess one or more six-cysteine-containing chitin-binding domains related to the peritrophin A domain (ChtBD2). An exhaustive 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. We 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. We propose the names “Cuticular Proteins Analogous to Peritrophins 3” (CPAP3) for the second family that has three ChtBD2s and “Cuticular Proteins Analogous to Peritrophins 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. We 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 ChtBD2s-containing proteins in different insect species belonging to different orders 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. The differences in the expression profiles of PMPs and CPAPs indicate that even though they share the peritrophin A motif for chitin binding, these three families of proteins have quite distinct biological functions.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Chitin, an extracellular matrix polysaccharide composed of β (1 → 4) linked N-acetylglucosamine residues, is a major component

of the insect exoskeleton, tracheae and midgut peritrophic matrix (PM). In all of these structures, chitin is associated with an assortment of proteins that influence the physical, chemical and biological properties of the extracellular structure. Although the exact chemical nature of the association of the chitinous matrix with proteins is not well understood, progress is being made in cataloging the proteins found in the cuticle and PM. Two main classes of chitin-binding motifs have been identified 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

Abbreviations: CHS, chitin synthase; PMP, peritrophic matrix proteins; Tc, *Tribolium castaneum*; CBDs, chitin-binding domains; ChtBD2, peritrophin A-type chitin-binding domain; CPAP, Cuticular Protein Analogous to Peritrophins; R&R Consensus, Rebers and Riddiford Consensus.

* Corresponding author. Tel.: +1 785 776 2797; fax: +1 785 532 7278.

E-mail address: smk@ksu.edu (S. Muthukrishnan).

0965-1748/\$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ibmb.2010.01.011

Please cite this article in press as: Jasrapuria, S., et al., Genes encoding proteins with peritrophin A-type chitin-binding domains in *Tribolium castaneum* are grouped..., *Insect Biochemistry and Molecular Biology* (2010), doi:10.1016/j.ibmb.2010.01.011

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 = SMART 00494). The second general class of chitin-binding sequence motifs found in insect cuticular proteins is the Rebers and Riddiford Consensus sequence (R&R Consensus; PDGDYNY+YETSNGIADQETGD+KSQGETRDG++A VDVV+GSYSYVDPDGTTRTVTYTADDENGFQPVGAHLP; pfam00379; Rebers and Riddiford, 1988; Anderson, in this issue; Karouzou et al., 2007; Willis, 2010), which is devoid of cysteine residues. Many of these proteins are likely to be cross-linked with one another via quinones derived from N- β -alanyldopamine and N-acetyldopamine (Kramer et al., 2001; Andersen, in this issue). Some of these R&R proteins were identified in insect cuticle extracts by proteomic analyses (He et al., 2007).

Evidence for the involvement of the ChtBD2 motif as well as the R&R Consensus in chitin-binding comes from experiments showing a gain in affinity for chitin following attachment of these sequences to a protein that normally does not bind to chitin (Arakane et al., 2003; Rebers and Willis 2001). Models have been proposed to explain some of the interactions between the N-acetylglucosamines and the amino acid residues lining the putative binding pockets (Kramer and Muthukrishnan, 2005; Hamodrakas et al., 2002; Iconomidou et al., 2005), but the molecular details involved in chitooligosaccharide binding are still unknown for either type of CBD.

Proteins with CBDs 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/genomes encoding proteins with this peritrophin A-type chitin-binding motif, ChtBD2, in cuticle-forming tissues from *Drosophila 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 peritrophin A domains. More recently, Nisole et al. (in press) have cloned a cDNA from *Choristoneura fumiferana*, which encodes a protein homologous to these proteins. This gene was highly expressed in the epidermis and the purified recombinant protein bound to chitin. Thus, it is clear that ChtBD2 is present in proteins that are found in locations other than 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 molecular masses. 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 Tellam et al., 1999; Dinglasan et al., 2009; Venancio et al., 2008; Hegedus et al., 2009). Many of these are also interspersed with serine and threonine-rich mucin-like domains that are likely to be glycosylated (Toprak et al. 2009). These studies have revealed that the group of ChtBD2-containing proteins is much larger than previously known. However, a bioinformatics study cataloging all proteins with one or more ChtBD2s (regardless of their location) encoded by a single genome has not been undertaken so far. In this paper, we have used an exhaustive bioinformatics search to identify in the red flour beetle, *Tribolium castaneum*, all of the genes that encode proteins containing one or more ChtBD2s, which have been associated with the ability to bind to chitin. Our studies have revealed that the *T. castaneum* proteome contains 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, a characteristic consistent with their roles in

binding to chitin in extracellular matrices. We propose a new nomenclature for three gene families that encode non-enzymatic proteins with one or more ChtBD2s, the first being Peritrophic Matrix Proteins (PMPs) for proteins expressed in the midgut, and the second and third as “Cuticular Proteins Analogous to Peritrophins 1 and 3” (CPAP1 and CPAP3) for proteins expressed in cuticle-forming tissues with one and three ChtBD2 domains respectively.

2. Materials and methods

2.1. Insect cultures

The GA1 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). The GA1 strain used in our study is a highly inbred robust strain from which the GA2 strain was obtained and used for DNA sequencing.

2.2. Identification and cloning of the genes encoding proteins with ChtBD2s

Sequence comparisons of previously characterized peritrophins from several insects including *T. castaneum* indicated substantial variations in the spacing between the conserved cysteines of the peritrophin A, B and C motifs with 6, 8 and 10 cysteines respectively. However, there were similarities among the amino acid sequences between these cysteines among peritrophins from the same insect species. This was also true of the “Obstructor” family of proteins (CPAP3 family in this paper). In order to maximize the probability of identifying as many genes as possible encoding *T. castaneum* proteins with ChtBD2s, we conducted an exhaustive bioinformatics search of the *T. castaneum* genome and EST databases (Beetlebase, <http://www.bioinformatics.ksu.edu/Beetlebase>; NCBI, <http://www.ncbi.nlm.nih.gov/projects/genome/seq/BlastGen/BlastGen.cgi?taxid=7070>; and the Baylor College of Medicine's Human Genome Sequencing Center, <http://www.hgsc.bcm.tmc.edu/project-species-i-Tribolium%20castaneum.hgsc?pageLocation=Tribolium%20castaneum>), using previously characterized CBDs from insect peritrophins A, B and C as queries (Tellam et al. 1999). Similarly, several other ChtBD2s from *Trichoplusia ni*, *Lucilia cuprina*, *Anopheles gambiae* and *D. melanogaster* were also utilized as queries to identify genes encoding proteins with ChtBD2s in *T. castaneum*. The sequences of the *T. castaneum* proteins identified in the initial search were used as queries for a second round of “BLAST” search to identify additional PMP or CPAP genes. This process was repeated until no additional genes could be identified.

To clone the corresponding cDNAs, gene-specific primers for each gene were designed from available ESTs or from gene models predicted using the GLEAN software, which are available in Beetlebase. Template cDNAs were prepared from RNA isolated from insects at various developmental stages. DNA fragments were amplified using pairs of gene-specific primers designed from sequences in the 5'- and 3'-UTR regions or within the protein-coding regions (CDSs) of target genes, and were cloned into the pET vector (Promega, Madison, WI) and then sequenced (DNA sequencing facility at Kansas State University). Using this approach, we obtained near full-length cDNAs covering the CDSs of each of the PMPs and CPAPs.

2.3. DNA and protein sequence analyses

After sequencing full-length or near full-length cDNAs of *TcPMPs* and *TcCPAPs*, the 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 (Pôle Bioinformatique Lyonnais) website (http://npsa-pbil.ibcp.fr/cgi-bin/align_clustalw.pl).

2.4. Profiles of gene expression during development

The RNeasy Mini kit (Qiagen, Valencia, CA) was used to isolate total RNA from larvae according to the manufacturer's instructions. Total RNA samples were treated with RNase-free DNase I (Ambion, Austin, TX, 2 U/ μ l) 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 μ g 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 (Supplementary Table S1).

2.5. 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 and all branches with a value less than 50% were collapsed. The accession numbers of the genes encoding *T. castaneum* ChtBD2-containing proteins are listed in Table 1.

3. Results

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

The initial search of the *Tribolium* genome identified 29 proteins with ChtBD2s. The domains from these proteins were subsequently

Table 1.
Tribolium castaneum genes encoding proteins with ChtBD2 domains.

	Gene/Splice Variant	GeneBank accession #	GLEAN#	cDNA length (bp) ^a	No. of Exons ^b	ORF (aa)	M.W (kDa)	No. of CBD	pI	LG	Map position (cM)	Citations	
Peritrophic Matrix Proteins (PMPs)	PMP1-A	GU128096	03179	501	4	166	18.7	1	5.3	3	0.0		
	PMP1-B	GU128097	15620	671	2	207	22.6	1	5.1	6	15.3		
	PMP1-C	GU128098	09231	2052	3	672	74.7	1	5.6	7	3.9		
	PMP2-A	GU128099	03274	507	3	168	18.4	2	4.6	3	0.0		
	PMP2-B	GU128100	03275	578	3	175	19.7	2	4.1	3	0.0		
	PMP2-C	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		
	PMP5-A	GU128103	03273	1119	6	372	41.0	5	4.2	3	0.0		
	PMP5-B	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		
	PMP14	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		
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 Muthukrishnan, 2010	
	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., 2008	
	CDA2A	ABU25224	14101	1831	5	535	66.0	1	5.3	5	20.0	Dixit et al., 2008	
	CDA2B	ABU25225	14101	1810	5	528	60.0	1	5.2	5	20.0	Dixit et al., 2008	
	CDA3	ABW74145	05409	1580	5	517	59.0	1	6.1	8	50.3	Dixit et al., 2008	
	CDA4	ABW74146	07635	1628	7	490	56.0	1	5.8	4	9.7	Dixit et al., 2008	
	CDA5A	ABW74147	06846	3694	15	1231	129.0	1	8.6	U	U	Dixit et al., 2008	
	CDA5B	ABW74148	06846	3694	15	1231	129.0	1	8.6	U	U	Dixit et al., 2008	

^a Longest available cDNA.

^b Minimum no. of exons.

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, 10 are CPAP1 and eight are CPAP3 family proteins, 13 include the chitinases and chitin deacetylases and eight others are classified as miscellaneous proteins. The rationale for these groupings and their descriptions are given in Section 3.3. Interestingly, our searches yielded only proteins with the peritrophin A-type ChtBD2 that contains six cysteines, but not the peritrophin B or C domains (with eight and ten cysteines, respectively). The results of our searches for ChtBD2-containing proteins, including their predicted molecular masses and number of ChtBD2s (except for the genes discussed in Section 3.8), are shown in Table 1.

3.2. 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, we used RT-PCR to amplify the putative full-length coding sequence (CDS) of each cDNA. In many cases, we 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, these corrections changed the number of ChtBD2s initially indicated by the GLEAN and/or NCBI predictions. We will describe some, but not all, of the examples of these incorrect annotations below. The correct sequences of these cDNAs and the corresponding proteins have been submitted to NCBI (Table 1).

In the case of the peritrophin, PMP5-B, comparison of the DNA sequence of the RT-PCR-derived cDNA with the genomic sequence corresponding to GLEAN_08506 revealed a rare error in genome sequence assembly, rather than the more common type of annotation error. The GLEAN annotation based on the incorrect assembly predicted a protein with seven ChtBD2s. The longest cDNA corresponding to this gene that we amplified encoded a protein that contained only five ChtBD2s. This was further verified by amplifying and sequencing the corresponding region in the genomic DNA. In the case of GLEAN_06098, we were unable to amplify the predicted 6.5 kb full-length cDNA by RT-PCR. Instead, we have confirmed that this gene actually represents an incorrect fusion of two tandem genes on the same strand, which encode 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 1. We were 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 gene, *TcPMP14*. 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 a protein with 14 ChtBD2s. Thus, we 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. Fig. 1 shows the gene structures and functional motifs of all *T. castaneum* genes encoding proteins with ChtBD2s, omitting those that encode chitinases or chitin deacetylases and several others that lack a signal peptide. Of the 50 proteins identified as having one 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 transmembrane domains, suggesting that they are secreted proteins. This inference is consistent with their predicted function in forming complexes with chitin, an extracellular polysaccharide.

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

We attempted to align the amino acid sequences of ChtBD2s from the 50 *Tribolium* proteins with one another and with the ChtBD2s of well-characterized insect peritrophins. During this process, it became clear that 11 *Tribolium* proteins were highly related to *L. cuprina* and *T. ni* peritrophins with ChtBD2s closely matching Tellam's consensus sequence for peritrophin A domains found in PM-associated proteins (Tellam et al., 1999; Wang et al., 2004). This family of eleven proteins was, therefore, tentatively classified as Peritrophic Matrix Proteins (PMPs) or "peritrophins". It should be emphasized that, at present, we lack definitive proof that these are *bona fide* proteins associated with the PM. The number of ChtBD2s in the PMP family of proteins ranges from one to 14 and the lengths of the individual ChtBD2s range from 52 to 56 amino acids when counting from the first to the sixth cysteines. The spacings between adjacent cysteines are in general accordance with the Tellam consensus for peritrophin A domains (Tellam et al., 1999). Even the sequences separating the cysteines in the peritrophin A domains of these PMPs appear to be more closely related to one another as compared to the corresponding sequences in ChtBD2s from the other *Tribolium* proteins (Fig. 2A, see the consensus sequence at the bottom of the alignment). Particularly noteworthy is the lack of variation in the spacing between the second and third cysteines as well as the spacing between the fourth and fifth cysteines in the consensus sequence for this group. PMP1-B has the greatest deviation from this consensus, the most noticeable being the number of amino acid residues between the first and second cysteines. In fact, the amino acid sequence of this protein (along with PMP1-C) is more closely related to the CPAP1 family (see below). These eleven proteins differ in the number of ChtBD2s (from one to 14) and, as a result, in their molecular sizes. In proteins with multiple ChtBD2s, the sequences of the individual ChtBD2s were found to be highly similar. This is especially true of PMP5-B, PMP9 and PMP14 with five, nine and 14 ChtBD2 repeats, respectively.

Of the remaining 39 proteins with ChtBD2s, 8 are grouped into a family that we have denoted as "Cuticular Proteins Analogous to Peritrophins-3" (CPAP3) because of their apparent lack of a function in PM organization and presumed cuticular localization (see below). These eight proteins contain three ChtBD2s each, which are 54 to 56 amino acids long. We have named this family CPAP3 to indicate they have three ChtBD2s. All of these proteins also have a remarkably similar domain organization (see below). Two proteins, TcCPAP3-C5a and TcCPAP3-C5b, arise from alternative splicing of the fifth and 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. In the ChtBD2 motifs of this family of eight proteins, the third ChtBD2 has the greatest variation in sequence and in spacing between cysteines (Fig. 2B). The spacing between the first and second cysteines as well as the spacing between the fourth and fifth cysteines are greater than those in the PMPs, but the spacing between the third and fourth cysteines is shorter (Fig. 2B). These proteins are orthologs of members of the family of *Drosophila* proteins collectively known as "gasps" (genes analogous to small peritrophins) or "obstructors," all of which have three ChtBD2s (Barry et al., 1999; Behr and Hoch, 2005). We have named the individual *T. castaneum* orthologs of this family to reflect their orthology to the corresponding *Drosophila* proteins as well as the number of encoded ChtBD2s (e.g. *T. castaneum* CPAP3-B is the

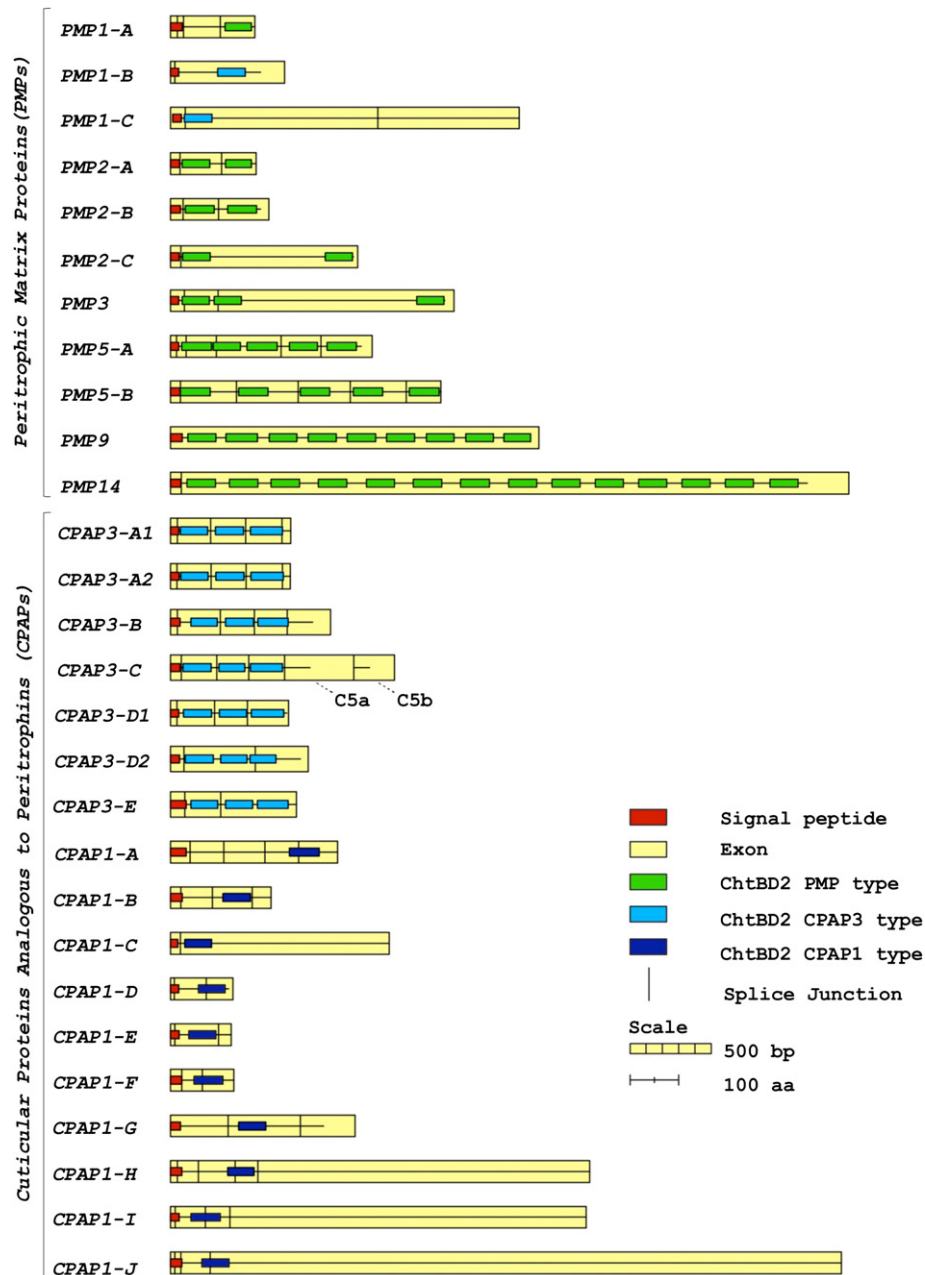


Fig. 1. Schematic diagram of the exon–intron–domain organizations of *T. castaneum* genes encoding ChtBD2-containing proteins. Yellow boxes represent the longest cDNA for the particular gene that has been cloned. Red solid box indicates signal peptide and green, light blue and dark blue boxes indicate the ChtBD2s from PMP, CPAP3 and CPAP1 families of ChtBD2-containing 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*. (For interpretation of the references to colour in figure legends, the reader is referred to the web version of this article.)

ortholog of *D. melanogaster* Obst-B and has three ChtBD2s). There are four fewer CPAP3 orthologs in *T. castaneum* than in *D. melanogaster* (Behr and Hoch, 2005).

A third family of 10 *T. castaneum* proteins includes members with a single ChtBD2 at either the N-terminus or the C-terminus, and is, therefore, given the name CPAP1 family. As in the case of the PMPs, 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. Notably, this family shows much greater conservation of the sequences between cysteines compared to the CPAP3 and PMP families (Fig. 2C).

In addition to the 11 PMP, 8 CPAP3 and 10 CPAP1 proteins listed in Table 1, we previously characterized six chitinases and seven chitin

deacetylases, all of which contain one or more ChtBD2s (Zhu et al., 2008a; Dixit et al., 2008; Arakane and Muthukrishnan, 2010). Apart from these, we have also identified five other proteins with ChtBD2s represented by GLEAN_02107, 02058, 11724, 00588 and 13568 that lack a predicted signal peptide. Another, GLEAN_12551, represents a very large protein of 437.7 kDa with a signal peptide, one ChtBD2 and five von Willebrandt factor type D domains, which is an ortholog of a *Drosophila* 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. Its tissue-specificity is uncertain, but it does not appear to be expressed in elytra (modified forewing), hindwings or gut (data not

A PMP

```

PMP1-C  CPTVDGPD-SVYIFHE-DCHKFWQCSN-G VAYLFNCSATTVPDPSLNVCVHEWDYDC : 54
PMP5-B  CPKQDGKD-SVYIFHE-DCIKFWQCSN-G TPYLFDCPDNLHFNPKLNVCDWPNAAAGC : 54
PMP3    CPARDGAF-PVYIFHE-DCGKFYQCSN-G VAYLQNCPPGLHWNVAKLVCDWPRDAGC : 54
PMP2-C  CPAVDPPPT-PVYIFHESDCSKFYECHE-DG TPHLLECPPEGLDFNPVLNVCDYPEQAGC : 55
PMP9    CPATSPLTYTVHLQHDTDCTKFKYKCDH-G KKILFSCPSGLHFFHPLFQVCDWPNVAVGC : 56
PMP14   CPKTDPLDYTVHLQHESDCTKFKYKCDH-G GKVLFDCEPABLHFNVPVLOVCDWPNVAVGC : 56
PMP1-A  CID---DSTELWHPHADCAKYIECFH-G NSYEMTCPPGLYFSSSSKTCVTADESECC : 52
PMP2-A  CSQL-PLDQEVLFESPIDCSLYYKCYQ-G IFSEEKCPKGLYFSEYNGKCVQAQYSECC : 55
PMP5-A  CAGV-PPGSTYLFPYPGDCIKFYVCEN-G TKRVEDCPGLWFEALQACDHPDNSSGC : 55
PMP2-B  CPY--PSTEITYFPYEGDCTKYWECCYS-G HSYLYTCCPAGLWVHQEISECCDYEPG-DFC : 53
PMP1-B  CKK-----VGIFDPPFYCNKFLVLCVDMG*QAYSSRCEDCLGYNIETGVCQDVKLSGGC : 55
Consensus  C          p    DC KF  C  G          CP GL F          Cd p  C

```

*An insertion QSRL is present at this position only in this CBD

Consensus of conserved cysteines and spacings: CX₁₁₋₁₇CX₅CX₉₋₁₄CX₁₂CX₆₋₇C

B CPAP3 (3rd CBD)

```

CPAP3-B  CP-KVNETVA---ATHPRYADHDDCOYFYVVCINGDTPRPSGCKLQGVFD---DVGKKCDWVRNVPEC : 60
CPAP3-A1 CP-KDQQTANGQLVVHPKYAHFTDCCRFYVCLNGQEPRLDLCQVGEVYN---EESQRCDAPENVVPGC : 64
CPAP3-A2 CP-DEEILGPGGRKLPHTFAHPEDCCKFYICRNQVMPQKGCQVKGLVYN---EETFTCDPKNVVPGC : 64
CPAP3-C  CP-APGEVNSG---SFSRHAHEDDCRKYVICLEG-TAREYGCPICTVFKIGDADGTGNCEDPEDVPGC : 64
CPAP3-E  CP-NDGRSFGLG-EAEFRFRFRSENDCCRYFVVCVNG-RPRLYNGCGEGRAFN---DLIGACDGVENVVGC : 62
CPAP3-D1 CPTKVPNSNSPAAKFWPYPRFAVEGDCHRLITCVNG-FPRLISCGEGKAFD---QHSLLTCEEPELVPHC : 64
CPAP3-D2 CNDDEPNVPLG-----KSNRYWCCQGG-YPRLRQRCPAMLVFD---RRSLRCVVEP-TEDC : 52
Consensus  CP          g          P DC RFW C NG PR C G FD          Cd p Vp C

```

Consensus of conserved cysteines and spacings: CX₁₃₋₂₄CX₅CX₉₋₁₀CX₁₂₋₁₆CX₇₋₈C

C CPAP1

```

CPAP-1A  CKHVPAAHPCHYANVETGCOAYHVCHDGREGEQGSFLCINCTIFNCAEFACDWWYVNV-C : 59
CPAP-1D  C--LGRNPGYYADLETGCOAYHRCEYN---SAASFCLCTNCTLFNEQFQVCDQFYNVVR-C : 53
CPAP-1F  C-DQR-LPGYYSDPEAQCQVWHWCLPSG---QQYSFLCPNGCTIFNCFARVCDWVFNVD-C : 54
CPAP-1B  C-DEQYPYGFYADETRTCCOTWHYCDIDG---RQTTFCLCPNGCTIFNSOLVFCVDFWVFNVR-C : 55
CPAP-1H  C--DGLHDGFYASVPHKCOVYHHC-LFG---TRYDFLCANVYAFDQKTFCHFVSEVD-C : 53
CPAP-1E  C--TNRAIGFYADVEYDCQIFHMCDEPG---RRIPHVCANDTISFNOEYRVCWENNFD-C : 54
CPAP-1I  C--RNVDSGYADLETDCOVFHCCEEGK---KISFLCPNGCTIFNCOSELCEWVFKVN-C : 53
CPAP-1C  C--EGRTTGYYADVESGCQVYHMCDDLG---RQFSYTCPNALFLQORMLICDHWMVNV-C : 54
CPAP-1G  C--DGKNTGYADLEDLGCVEVFMCDQDNA---KHSWICPEGFTFHQVHLICMPPGGDNIC : 54
Consensus  C          G Y          C  H C          C          F          C          C

```

Consensus of conserved cysteines and spacings: CX₁₄₋₁₆CX₅CX₉₋₁₃CX₁₂CX₇₋₈C

Overall consensus for all CBDs for the three sub-groups: CX₁₁₋₂₄CX₅CX₉₋₁₄CX₁₂₋₁₇CX₆₋₈C

Fig. 2. Amino acid sequence alignment of the ChtBD2s (first ChtBD2) of each of the PMP, CPAP3 and CPAP1 families of proteins of *T. castaneum*. The multiple sequence alignment of ChtBD2s for each of these ChtBD2-containing proteins was carried out using the ClustalW software from PBIL Expaty tool (http://npsa-pbil.ibcp.fr/cgi-bin/align_clustalw.pl). Shaded in black are the cysteines and other residues conserved in all proteins of that family. The less-conserved amino acids are shown in various shades of gray. A: ClustalW alignment of the conserved amino acid residues within the first ChtBD2 of the different PMP proteins. An insertion of the sequence, QSRL, is found only in TcPMP1-B at the position indicated by * and is not shown in the alignment, but this insertion is taken into account in building the consensus sequence. At the bottom, the conserved spacing between the six cysteines of the ChtBD2s alone is indicated. The numbers on the right indicate the number of amino acids in each ChtBD2. B: ClustalW alignment of the third ChtBD2 from different CPAP3 proteins in *T. castaneum*. C: ClustalW alignment of ChtBD2s from CPAP1 proteins of *T. castaneum*. The decreasing % similarity is indicated by decreasing shades of gray.

shown). These seven proteins most likely have roles other than interactions with chitin in the gut, tracheae or cuticle. Therefore, we have not included these proteins in our analyses. Finally, there is yet another GLEAN_15245, which has a signal peptide and encodes a protein with multiple ChtBD2s, in a region of the genome where there is a 4 kb gap in the genome assembly (this gene and the encoded protein will be discussed further in Section 3.8.1).

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

To establish the evolutionary relationship among the ChtBD2 domains of the three families of *T. castaneum* proteins including the domains within the same protein discussed above, we carried out a phylogenetic analysis of all ChtBD2s (using only the sequence

between the first and the sixth cysteines of the peritrophin A motif and excluding the linker regions between adjacent ChtBD2s) in proteins from *T. castaneum* using the ClustalW program. The tree appears to contain three distinct branches. The ChtBD2s of the 11 PMP family of 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 Fig. 3). The only exceptions are the two ChtBD2s associated with PMP1-B and PMP1-C, which align with the CPAP3 family. Most of the ChtBD2s within the same protein (e.g. all ChtBD2s of PMP9 and from the 2nd through the 14th ChtBD2 of TcPMP14) are clustered in the same branch, indicating minimal sequence divergence among the ChtBD2s in the same protein. Occasionally, we 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).

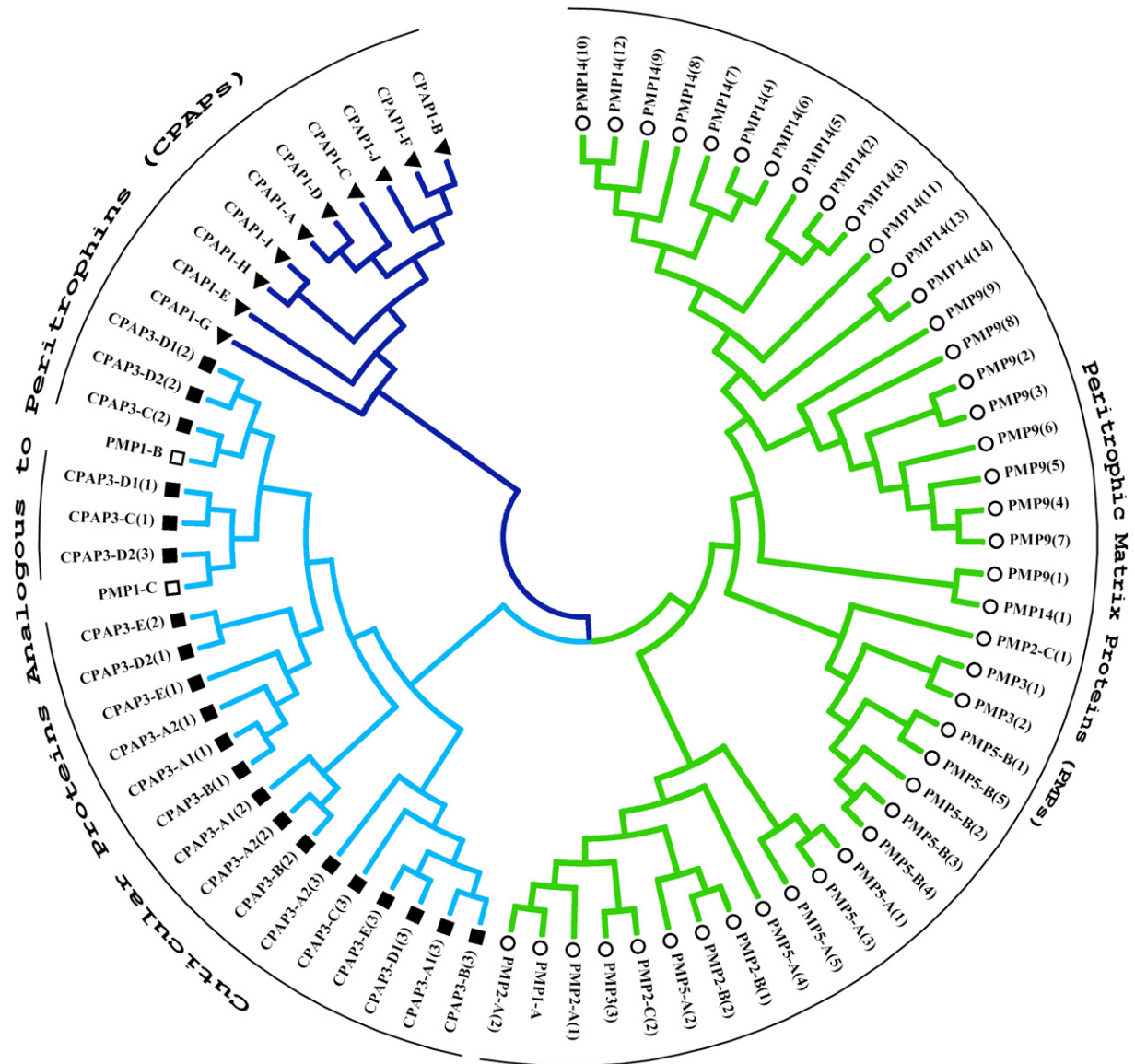


Fig. 3. Phylogenetic tree of ChtBD2s present in PMP, CPAP1 and CPAP3 families of *T. castaneum* proteins. Individual ChtBD2s from each protein (number in parenthesis indicates their relative location from the N-terminus) were used to construct a phylogenetic tree as described in the Materials and Methods 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. (For interpretation of the references to colour in figure legends, the reader is referred to the web version of this article.)

The ChtBD2s of the CPAP1 and CPAP3 families form two other branches of the evolutionary tree (shaded in dark and light blue in Fig. 3). One branch contains all *Tribolium* proteins with a single ChtBD2 belonging to the CPAP1 family (TcCPAP1-A through TcCPAP1-I, shaded in 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 in light blue), which are similar to the *Drosophila* obstructor/gasp proteins, form a second branch. Interestingly, the first ChtBD2s of almost all members of the TcCPAP3 family with the exception of CPAP3-C and CPAP3-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 CPAP3 proteins are also clustered, with CPAP3-D2 being the lone exception. ChtBD2 #1 and #2 of different CPAP3 proteins are more closely related to one another other than to ChtBD2 #3, which form a separate branch. The phylogenetic analysis further supports

our classification of the ChtBD2-containing proteins of *T. castaneum* into three families based on variations in the Tellam consensus sequence for the conserved cysteines of ChtBD2s (Fig. 1).

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

The nomenclature for the 11 PMPs is according to the following convention. The number that follows the designation PMP indicates the number of ChtBD2s in the protein. For example, TcPMP9 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 letter 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 (Table 1). All of these

proteins have an N-terminal signal peptide and no transmembrane helices, indicating that they are secreted proteins. Fig. 1 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. They resemble mucin domains but are not predicted to be highly glycosylated by the NetOGlyc 3.1 software. The other four PMPs, TcPMP1-A, TcPMP1-B, TcPMP1-C and TcPMP2-C, have no similarity to any proteins encoded by the genomes of either *D. melanogaster* or *A. gambiae*, if the ChtBD2 regions are omitted.

3.5.1. Clustering of PMP genes in the *T. castaneum* genome

As mentioned earlier in Section 3.2, the two genes, *TcPMP9* and *TcPMP14*, are arranged in tandem in linkage group 10 in the *T. castaneum* genome and their CDSs are separated by only 331 base pairs. There are two other clusters of PMP genes in the *T. castaneum* genome. *TcPMP1-B* and *TcPMP3* are paired on linkage group 7 within a stretch of only 977 bp. Furthermore, *TcPMP2-A* and *TcPMP5-A* are paired on linkage group 3 within a stretch of 1,575 bp (see Table 1). These pairs may be a result of relatively recent gene duplications.

3.6. CPAP1 family

The CPAP1 family of proteins, which have only one ChtBD2, differ from the CPAP3 family of proteins in ways other than the number of ChtBD2s, the most noticeable being the location of the chitin-binding domain itself (Fig. 1). In different CPAP1 proteins, the ChtBD2s (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 rarely occurs in other CPAP genes. Nevertheless, the ChtBD2s of the CPAP1 family of proteins are closely related, indicating a common ancestral origin (see Fig. 3). Proteins belonging to this family also have N-terminal signal peptides and lack transmembrane domains, properties consistent with a role that involves interactions with an extracellular polysaccharide 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 present in the ChtBD2 (Fig. 1). 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 in addition to the ChtBD2-coding region. In these proteins, the ChtBD2 represents only a minor part of the overall sequence. Regarding the CPAP1 family, there are orthologs in *Drosophila* only for TcCPAP1-C and TcCPAP1-H. The latter contains a mucin-like domain and is rich in serine residues. However, there is nothing known about the nature or functions of these proteins in the fruit fly. These proteins have no sequence similarities other than the ChtBD2 region.

3.7. CPAP3 family

There are seven genes in *T. castaneum* encoding eight proteins that have a tandem arrangement of three ChtBD2s (Fig. 1). Orthologs of

these genes have been found in *Drosophila* and several other species of insects. The *Drosophila* genes have been named previously as “obstructors” (Behr and Hoch, 2005) or as “gasps” (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 *Drosophila* 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 *Drosophila* proteins. In the case of TcCPAP3-A and TcCPAP3-D, *T. castaneum* has two paralogs instead of one protein each as in *Drosophila*. We have denoted them as TcCPAP3-A1, TcCPAP3-A2, TcCPAP3-D1 and TcCPAP3-D2. Like *T. castaneum*, other insect species have additional paralogs of the CPAP3-A genes. *A. gambiae* has three corresponding genes, and *Nasonia vitripennis* and *Acyrtosiphon pisum* have two (data not shown). Similarly, two paralogs of CPAP3-D genes are found in several other species including *Aedes aegypti*, *A. gambiae*, *Culex quinquefasciatus*, *N. vitripennis* and *A. pisum* (data not shown). 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. 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. Similar alternative splicing of pre-mRNA in *D. melanogaster* results in two isoforms of obstructor C (accession numbers NP_649611 and NP_001097694). In the case of TcCPAP3-E, there is only one ortholog of this protein in *T. castaneum* and several other insects, whereas there are two orthologs in *D. melanogaster* as a result of alternative splicing (Behr and Hoch, 2005). 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 (group #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 the group #2 genes in *T. castaneum*. Behr and Hoch (2005) have reported that *Drosophila* 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 Fig. 4. It is clear that CPAP3 proteins form five separate branches, each branch corresponding to one of the five CPAP3 proteins, A through E. Branches containing the CPAP3-A and CPAP3-D proteins have undergone an expansion in several insect species, but branches representing CPAP3-B and CPAP3-E have only a single representative with the exception of *Drosophila*, which has two variants of CPAP3-E as a result of alternative splicing.

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 (Table 1). The other two genes are in linkage group 2 at map positions 36.4 cM and 49 cM, indicating that they are widely separated on the same chromosome.

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, lacks proline, threonine or serine residues, which are found in linkers of many PMPs (e.g. TcPMP9 and TcPMP14). The linkers between the second and third ChtBD2s range from six to 13 residues in length among the CPAP3 proteins and lack residues associated with O-glycosylation. As indicated in Fig. 1, the C-terminal sequences following the ChtBD2s in different CPAP3 proteins are variable in both length and sequence.

3.8. Chitinases and chitin deacetylases

Included among the proteins with ChtBD2s listed in Table 1 are several chitinases and chitin deacetylases. Among the *T. castaneum*

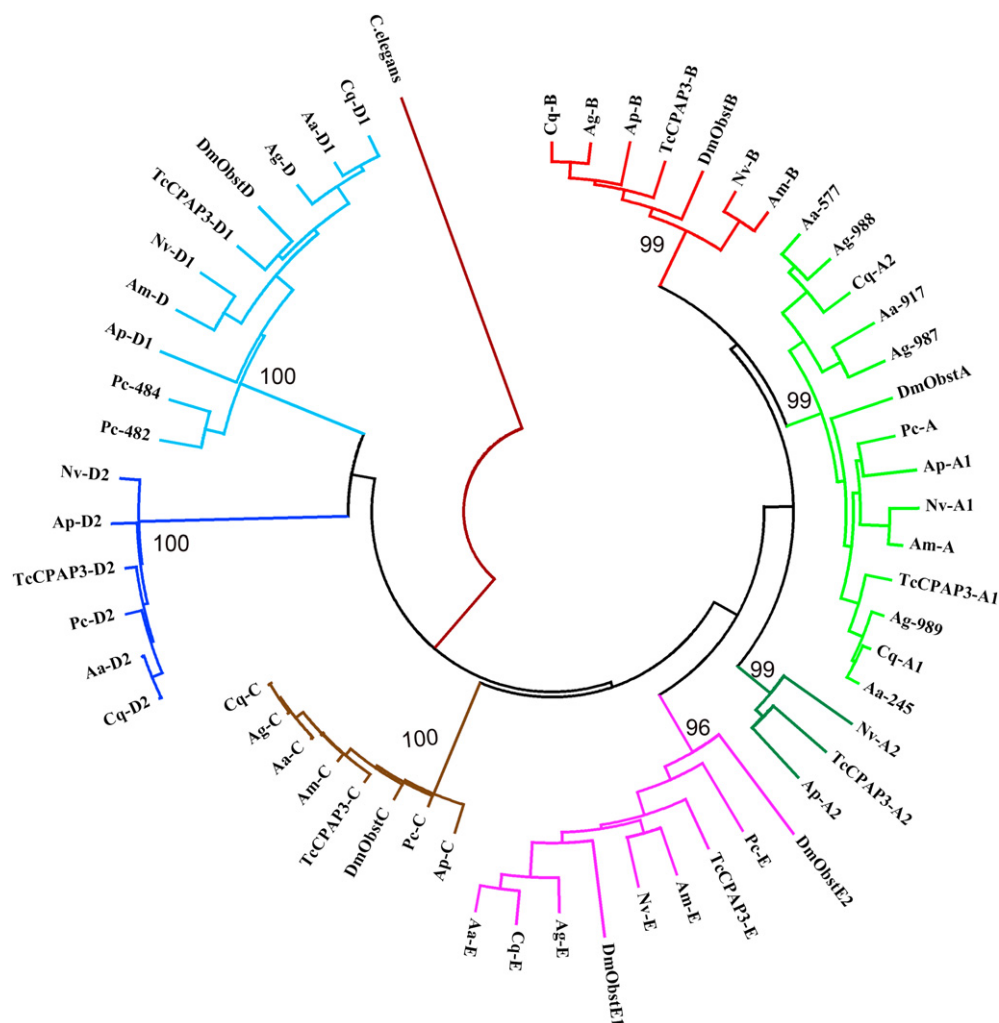


Fig. 4. Phylogenetic analysis of the ChtBD2 sequence of the CPAP3 family of proteins from *Drosophila melanogaster* (Dm), *Anopheles gambiae* (Ag), *Acyrtosiphon pisum* (Ap), *Aedes aegypti* (Aa), *Apis mellifera* (Am), *Culex quinquefasciatus* (Cq), *Nasonia vitripennis* (Nv), *Pediculus humanus corporis* (Pc), and *Tribolium castaneum* (Tc) along with a ChtBD2 domain from *Caenorhabditis elegans* (C. elegans) protein, NP_502145.2 as the outlier. The CPAP3 orthologs 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. A bootstrap analysis of 5000 replications was carried out on the tree. The bootstrap values of the major branches are indicated by the numbers. The accession numbers of the ortholog from different insect species are listed in [Supplementary Table 2](#). (For interpretation of the references to colour in figure legends, the reader is referred to the web version of this article.)

chitinases, there is one with 5 ChtBD2s (chitinase 10) and five with one ChtBD2 (chitinases 4, 5, 6, 7 and 8). The ChtBD2s are near the C-termini with one exception being the largest enzyme, chitinase 10, which has its ChtBD2s interspersed among its five catalytic domains (Zhu et al., 2008a). 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 (Dixit et al., 2008). These enzymes and the expression profiles of the genes encoding them have been previously described in detail by Zhu et al. (2008b,c), Arakane and Muthukrishnan (2010) and Dixit et al. (2008). Therefore, they are not a major focus of the current discussion involving non-enzymatic ChtBD2-containing proteins of *T. castaneum*.

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

In addition to the 42 putatively chitin-interacting proteins listed in [Table 1](#), there are two other proteins with ChtBD2s in the predicted proteome of *T. castaneum*. One of these (GLEAN_15245) is in a region where sequence data are incomplete. Transcripts for the other gene (GLEAN_09553) are not detectable in the midgut or

cuticle-forming tissue. These exceptional proteins are described in the following sections.

3.9.1. 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 within the transcription unit of this predicted gene is the cause of this apparent annotation error. However, we were able to amplify a partial cDNA clone that includes two of the ChtBD2s at the 5'-end of the ORF and two additional ChtBD2s not predicted by the gene model. Inspection of the sequence of this clone indicated that several introns predicted by the GLEAN annotation are, in fact, ChtBD2-containing exons. We also amplified another cDNA encoding the last three ChtBD2s predicted by the GLEAN model. If GLEAN_15245 indeed encodes a single protein, this protein should have at least seven 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 across the relevant gap in the genome assembly to test this possibility. Until this gap is bridged, the correct annotation of the region of GLEAN_15245 will remain incomplete. RT-PCR indicates that this gene is expressed in the carcass and in elytra and hindwings that were dissected from 3–4-d-old pupae, but not in larval and adult midgut (data not shown). It maps to position 2.5 cM in chromosome 6.

3.9.2. GLEAN_09553

This gene model predicts a protein with a signal peptide and three ChtBD2s. However, only the middle ChtBD2 matches the consensus shown in Fig. 2. This gene shows no expression in the larval stages. In the adult, it is expressed in the carcass, but not in the midgut or in either of the two cuticle-forming tissues tested, namely elytra and hindwings (data not shown). With the available data, we cannot predict whether this protein would be associated with chitin in the general body cuticle.

3.10. Tissue-specificity of expression of genes encoding ChtBD2-containing proteins in *T. castaneum*

We assessed the tissue-specificity of expression of genes encoding proteins with ChtBD2s in *T. castaneum*. We dissected last instar larvae 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, 2008). *CHS-B* is expressed in the midgut, while *CHS-A* is expressed in the hindgut and carcass. We also divided the midgut tissue into anterior, middle and posterior thirds prior to RNA isolation because we have previously observed differences in levels of transcripts for CHSs and chitinases in these gut segments (Arakane et al., 2005; Bolognesi et al., 2005). We anticipated that there might be corresponding differences in the expression profiles of the ChtBD2-containing proteins that bind to chitin in different extracellular locations. Transcripts for each of these genes were detected using RT-PCR and gene-specific pairs of primers. As shown in Fig. 5, there are substantial differences in expression among the three families of genes encoding proteins with ChtBD2s as well as among members of the same family. The most striking finding from these studies is that all of the *PMP* genes are expressed in the midgut, while none are expressed in the carcass. In contrast, almost all of the genes of the *CPAP3* and *CPAP1* families have exactly the complementary expression profile, namely expression 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 midgut in addition to high expression in the hindgut, possibly 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 the *CPAP3* and *CPAP1* genes that are expressed in the carcass of larvae are also expressed in the integument, which includes the epidermis and trachea 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 compared to 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 family of ChtBD2s-containing genes are described below.

3.10.1. *PMP* genes

Among the genes encoding *PMPs*, there are differences in expression levels in different regions of the midgut. Transcripts for some of the *PMPs* with one to five ChtBDs (*TcPMP1-A*, *2-A*, *2-B*, *2-C*, *3* and *5-A*) are detected predominantly in the anterior and middle midgut regions and to a lesser extent in the posterior midgut (*TcPMP1-A*, *2-A* and *2-C*), whereas *TcPMP5-B* is expressed in the middle and posterior midgut regions, and to a lesser extent in the hindgut. *TcPMP1-C* is expressed at high levels in the anterior midgut and at a low level in the posterior midgut. Transcripts for *TcPMP9* and *TcPMP14*, which have the largest number of ChtBD2s, are undetectable in the anterior and middle midgut, but are detected in the posterior midgut. *TcPMP14* is also expressed in the hindgut. Thus, there are qualitative differences in the sites of expression of different genes encoding *PMPs* within the midgut and possibly quantitative differences as well (Fig. 5A). All of the *PMP* genes are expressed in the midgut during both larval and adult stages, indicating there are no larval- or adult-specific *PMPs* in *T. castaneum*. Furthermore, there is no evidence of expression of *PMP* genes in the carcass.

3.10.2. *CPAP3* genes

There are subtle differences in the tissue specificities of expression of *TcCPAP3* genes. Unlike most of these genes, *TcCPAP3-E* is expressed at a higher level in the hindgut than in the carcass (Fig. 5B). Therefore, *TcCPAP3-E* may have a special role in hindgut cuticle. *TcCPAP3-A1* is expressed equally in the hindgut and carcass (Fig. 5B). Separation of carcass into integument and fat body revealed that none of the *TcCPAP3* genes except *TcCPAP3-A1* are expressed in the fat body, but presumably are expressed in cuticle-forming tissues such as the epidermis. Measurement of transcripts in adult elytra and hindwings indicated that all of the *TcCPAP3* genes including *TcCPAP3-A2*, which is not expressed in last instar larval tissue, are expressed in these cuticle-forming tissues. *TcCPAP3-A1* and *TcCPAP3-C5a* genes show significant differential expression in the elytra versus hindwings. Combined with the data showing that these genes are not expressed in the midgut, these results suggest a role for these proteins exclusively in interactions with cuticular chitin, but not with PM chitin.

3.10.3. *CPAP1* genes

The expression patterns of *TcCPAP1* genes also show differences in tissue-specificity. As expected, none of them are expressed in the larval midgut and all of them are expressed in larval and/or adult cuticle-forming tissues. *TcCPAP1-H* and *TcCPAP1-I* are expressed in the hindgut, but not in the carcass in the larval stage. *TcCPAP1-A*, which is expressed predominantly in the carcass, has a low level of expression in the fat body. Many of the *TcCPAP1* genes are expressed in the hindgut, but the relative expression in hindgut versus carcass varies among the genes. Similarly, there are differences in the expression levels in elytra versus hindwings for individual *TcCPAP1* genes, even though all of them are expressed in both tissues.

4. Discussion

The goal of this research was to carry out a comprehensive search for *T. castaneum* proteins 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 originally 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). It has been proposed that the peritrophins, particularly those with mucin domains, influence the permeability of the PM to dietary components, digestive enzymes and products of digestion in the endo- and ectoperitrophic space either directly or indirectly through

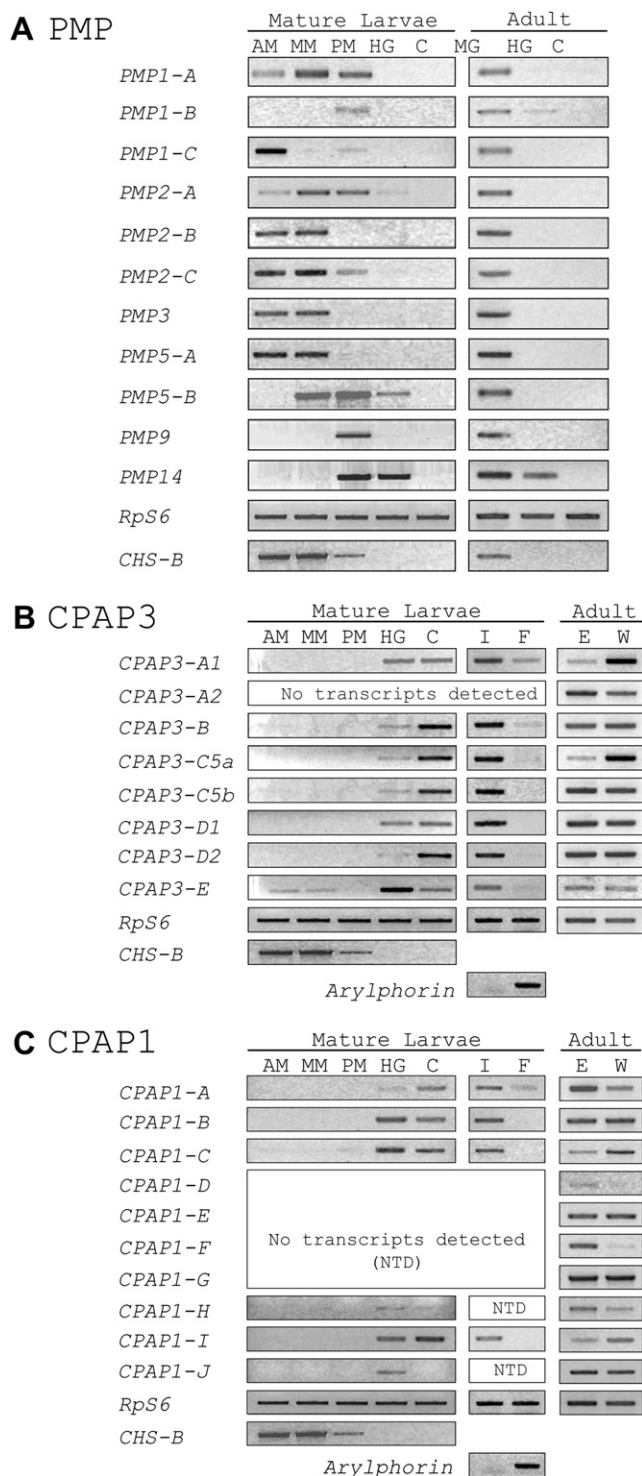


Fig. 5. Tissue-specificity of expression of *T. castaneum* genes encoding proteins with ChtBD2. Actively feeding last instar larvae ($n = 15$) were dissected to obtain anterior midgut, middle midgut, posterior midgut and hindgut tissue preparations. A part of the remaining carcass was further dissected to obtain larval integument (carcass minus fat body and muscles) and fat body preparations. Pupae (4 days 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 control 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.

their association with chitin and other proteins (Tellam et al., 1999; Terra, 1996, 2001; Hegedus et al., 2009; Wang et al., 2004). This domain is also found in some but not all chitinases and chitin deacetylases, which degrade or modify chitin (Kramer et al., 1993; Zhu et al., 2008b; Dixit et al., 2008; Arakane and Muthukrishnan, 2010) as well as in proteins expressed in embryonic tracheae and epidermis (Barry et al., 1999; Behr and Hoch, 2005). Removal of ChtBD2s from some of these proteins has been shown to result in lower affinity for chitin, while addition of one or more ChtBD2s has been found to increase the affinity for chitin (Zhu et al., 2001; Arakane et al., 2003). It has been proposed that ChtBD2s help to anchor the chitinases and chitin deacetylases on the insoluble polymer, facilitating the positioning of the catalytic center of the same enzyme at a target site in the chitin substrate. The presence of a ChtBD2 distinct from the catalytic site eliminates the need for the enzyme to dissociate from the substrate after hydrolysis and probably increases the processivity and the catalytic proficiency of the enzyme against crystalline chitin. However, there is very little information on the role of ChtBD2s in proteins that do not have enzymatic functions.

Recently, Venancio et al. (2008) identified ChtBD2-containing proteins in insects by querying *D. melanogaster*, *A. aegypti* and *T. castaneum* genomic and EST databases with the six-cysteine ChtBD2 sequence motif, CX_{13–20}CX_{5–6}CX_{9–19}CX_{10–14}C-X_{4–14}C proposed by Tellam et al. (1999) for the ChtBD2 as the query. They identified 25 “peritrophins” in the *T. castaneum* genome, which is considerably fewer than the 65 proteins that had been predicted from the *D. melanogaster* genome (Venancio et al., 2008). To maximize the probability of identifying proteins with one or multiple ChtBD2s, in our bioinformatics search of the *T. castaneum* genome we utilized 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 for 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. (2008), indicating that our search strategy has been more comprehensive. We verified the CDS 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 (either increased or decreased) in the predicted ChtBD2-containing proteins. Most significantly, we identified four new proteins belonging to the PMP family (PMP1-A, PMP2-A, PMP9 and PMP14), four additional members of the *T. castaneum* CPAP3 family (CPAP3-B, CPAP3-C[5a], CPAP3-[5b], and CPAP3-D1) and one member of the CPAP1 family (CPAP1-I), which had not been reported previously. Five other proteins with ChtBD2s were identified, but they lacked a predicted signal peptide and were left out of our analysis as they were 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 two other proteins, one with three ChtBD2s and another with an undetermined number of ChtBD2s

AM: anterior midgut, MM: middle midgut, PM: posterior midgut, H: hindgut, C: carcass (whole body minus gut), I: integument, F: fat body, E: elytron, W: hindwing. NTD = No Transcript Detected. Tissue-specificity of expression of genes for PMPs (Panel A), CPAP3 (panel B) and CPAP1 (panel C). The number of cycles for RT-PCR was 28 except for RpS6 and arylphorin, which was 24 cycles.

(GLEAN_09553 and GLEAN_15245). 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 (Zhu et al., 2008b, c; Dixit et al., 2008). There is some uncertainty whether GLEAN_15245 corresponds to one or two proteins as discussed in Section 3.8.1. 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 similarity of the ChtBD2s has resulted in subdivision of the ChtBD2-containing *T. castaneum* proteins into three large families, namely the PMP, CPAP1 and CPAP3 families. Our preference for these names 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 have at least one peritrophin A domain and are known or expected to bind to chitin in the two major chitin-containing extracellular matrices, namely the cuticle and the PM. 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; Moussian, 2010), the PM is believed to have roles in protecting the midgut epithelial cells and in aiding digestion (Terra 1996, 2001; Bolognesi et al., 2001, 2009) and is likely to be heavily hydrated as a result of glycosylation of several residues in the linker regions of the PMPs, some of which have mucin-like domains (Terra, 2001; Wang and Granados, 1997; Wang et al., 2004; Hegedus et al., 2009; Toprak et al., 2009). 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, in this issue) are likely to contribute to the differences in properties of the cuticle and the PM. Another possibility is that the arrangement of chitin fibers in the gut-associated PM may be different 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 another form (perhaps β), which leads to less hydrogen bonding between parallel chains of chitin compared to α -chitin (Jang et al., 2004). The PM is thought to be porous, allowing the movement of digestive enzymes and products of initial digestion to cross the PM barrier, which separates the food bolus from the midgut epithelial cells. Electron microscopic studies of European corn borer and tobacco hornworm have revealed that at least some of the insect PMs may have a lattice structure with regularly spaced grids between the criss-crossing chitin fibers (Harper and Hopkins, 1997; Hegedus et al., 2009). The PM is also less organized in the anterior midgut when compared to the posterior midgut. While the anterior part may be a gel-like structure, the posterior part is mechanically stronger and probably less permeable to larger molecules (Terra, 2001). The nature of the proteins associated with the PM may influence its properties in the regions of the midgut where they associate with chitin in the PM.

The cuticle, on the other hand, is a rigid and hydrophobic structure made up of a waxy layer (the envelope) and epicuticle overlaying a laminar procuticle made up of chitin and protein cross-linked via oxidized catechols. Besides the CPAP1 and CPAP3 proteins studied here, *T. castaneum* has a large assortment of cuticular proteins containing R&R Consensus motifs (Dittmer and Beeman, personal communication), which are also thought to be involved in interactions with chitin. The cuticular proteins are also likely to be involved in cross-linking with catechols (Andersen, in this issue). 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 and function has been largely unexplored until now. An exception is the *Drosophila* ‘*gasp*’ gene (ortholog of *TcCPAP3-C*), which has been reported to be expressed in embryonic tracheae (Barry et al., 1999). A role for *D. melanogaster* “*obstructor A*” (ortholog of *TcCPAP3-A*) in embryonic development has also been reported (cited in Behr and Hoch, 2005). The *obstructor* family of genes is expressed during early- and mid-stage embryogenesis in tracheae and epidermis. It is interesting to note that while *D. melanogaster* has 10 genes encoding eleven 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 belonging to the CPAP3 family. Whereas *T. castaneum* is missing five orthologs of *D. melanogaster* genes (Obstructors F, G, H, I and J of subgroup II), the beetle has additional genes (*CPAP3-A2* and *CPAP3-D2*). All of the other insect species studied to date appear to lack the *Drosophila*-specific subgroup II *obstructors* (*CPAP3s*), but have additional orthologs of *CPAP3-A* and *CPAP3-D*, which belong to subgroup I. The phylogenetic analysis shown in Fig. 4 suggests that in *Drosophila* species, additional orthologs of the latter two genes have been lost.

The finding that in *T. castaneum*, the CPAP3 family of eight proteins is expressed at several developmental stages in the integument indicates that they may be involved in organization of epidermal cuticle. Even though there are no biochemical studies to indicate that these proteins do bind to chitin, the similarity in the number, organization and spacing of the ChtBD2s in orthologs from several insect species (Fig. 3; Barry et al., 1999; Behr and Hoch, 2005) suggests that they serve similar critical functions in cuticle assembly. Interestingly, these genes are not expressed in the midgut, ruling out a role for these proteins in the assembly of the chitin-containing PM. Perhaps, those functions are fulfilled by the PMP family of proteins. The reason for the conservation of a large number of very similar proteins in the CPAP3 family in several insect orders is unclear. We have found preliminary evidence for the presence of some of the CPAPs identified in this work in pupal and/or adult cuticles of *T. castaneum* by direct proteomic analyses of cuticles (Dittmer and Arakane, unpublished data). 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 down-regulation of transcripts for specific CPAP3 proteins results in molting disruption or leg joint defects as well as mortality (Jasrapuria, unpublished data). 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 4, 5, 6, 7 and 8 with one ChtBD2s, and chitinase 10 with five ChtBD2) and several 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 *T. castaneum* genes that encode 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 other than the ChtBD2s. 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 tightly associated with chitin. For example, the CPAP3 family of proteins may participate in the formation of chitin laminae or stabilization of bundles of chitin fibers.

All of the ChtBD2s present in *Tribolium* 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 10-cysteine-containing peritrophin B or C sequence motif described by Tellam et al. (1999) in the predicted sequences of *T. castaneum* proteins. The consensus of conserved cysteines for all ChtBD2 sequences in *T. castaneum* proteins appears to be CX_{11–24}CX₅CX_{9–14}CX_{12–16}C-X_{6–8}C, which differs slightly from the one proposed by Tellam et al. (1999), which is CX_{15–17}CX_{5–6}CX₉CX₁₂CX_{6–7}C. Combining these two consensus sequences, 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_{11–24}CX₅CX_{9–14}CX_{12–16}CX_{6–8}C. The individual members of one family are more closely related to one another than to members of other families, and they have consensus sequences that include additional conserved residues besides the six cysteines (Fig. 2).

Phylogenetically, all of the ChtBD2s that belong to the PMP family are 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 cross-over events, resulting in proteins with different numbers of repeats. Supporting this hypothesis is the finding that some *T. castaneum* PMP genes are made up of exons encoding identical ChtBD2s and also that the two genes with the largest number of nearly identical repeats (*TcPMP9* and *TcPMP14*) 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 sequences, 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., 2008). To date the PMP that contains the largest number of ChtBD2s is the one from *Mamestra configurata*, which has a total of 19 ChtBD2s (Shi et al., 2004).

The significance of the variation in number of ChtBD2s among the 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 *Anopheles* genome. It is likely that the PMPs with a large number of ChtBD2s are bound too tightly to chitin in the PM and therefore are not extracted even under rather harsh conditions. We have found some preliminary evidence that suggests that the pattern of expression of the different PMP genes along the length of the midgut varies, with the genes encoding proteins with larger number of ChtBD2s being expressed in the more posterior portions of the midgut (Fig. 5 and our unpublished data), while those with fewer ChtBD2s are expressed in the more anterior parts. 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 et al., 2004). It is interesting to note

that both the number and spacing of 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 been shown to actually 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.

Finally, the expression profiles also provide experimental support for the classification of the ChtBD2-containing proteins into different families. The PMP family of proteins is expressed only during larval and adult feeding periods, but not during life stages when the gut is apparently not involved in digestion (data not shown). Thus, PMPs are likely to be *bona fide* components of the PM. On the other hand, the CPAP3 proteins are expressed at all developmental stages including the embryonic and pupal stages, and are detected in the integument, but not in the midgut, supporting the idea that these are cuticular proteins. The finding that transcripts for all of the CPAP3 genes are detected in elytra and hindwings further supports this conclusion. Our RT-PCR results regarding the expression of PMP and CPAP3 genes in the midgut are in agreement with a recent microarray analysis of genes expressed in the *T. castaneum* midgut, which reveals that several PMP genes are expressed in the midgut at high levels (PMP5-B, PMP2-B and PMP3), whereas the CPAP3 family of genes are only expressed at low or undetectable levels (Morris et al., 2009). The CPAP1 family, on the other hand, has a widely varying expression pattern in non-midgut tissues, most likely in the epidermal cells. In the future, studies of the properties of purified proteins and/or their ChtBD2s together with the results from RNAi experiments will shed further light on the functions of *Tribolium* proteins containing one or more ChtBD2s.

Acknowledgments

We thank Kathy Leonard for beetle husbandry and Dr. Qingsong Zhu for some of the preliminary work on CPAP3. We thank Neal Dittmer for his critical review of this paper and for valuable suggestions. This project was supported by NSF grants IBN-0316963 and IOS-615818. This is contribution no. 10-115-J from the Kansas Agricultural Experiment Station. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

Appendix. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ibmb.2010.01.011.

References

- Andersen, S.O., 2010. Insect cuticle sclerotization : A review. in this issue.
- Arakane, Y., Muthukrishnan, S., 2010. Insect chitinases and chitinase-like proteins. *Cellular Mol. Life Sci.* doi:10.1107/s00018-0161-9.
- 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., 2008. 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, T.A., Christensen, A.C., 1999. A peritrophin-like protein expressed in the embryonic trachea of *Drosophila melanogaster*. *Insect Biochem. Mol. Biol.* 29, 319–327.
- Baylor website <<http://www.hgsc.bcm.tmc.edu/project-species-i-Tribolium%20castaneum>>.

- 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. BeetleBase website <<http://www.bioinformatics.ksu.edu/BeetleBase>>.
- 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., 2009. 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., Jacobs-Lorena, M., 2009. The *Anopheles gambiae* adult midgut peritrophic matrix proteome. *Insect Biochem. Mol. Biol.* 39, 125–134.
- Dixit, R., Arakane, Y., Specht, A.A., Richard, C., Kramer, K.J., 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., Eiseman, C.H., Tellam, R. L., 1996. Characterization of a major peritrophic membrane protein, peritrophin-44, from the larvae of *Lucilia cuprina*. *J. Biol. Chem.* 271, 8925–8935.
- Gaines, P.J., Walmsley, S.J., Wisniewski, 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, hemolactin, similar to human von Willebrand factor. *Biochem. J.* 359, 99–108.
- 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.
- Harper, S.M., Hopkins, L.T., 1997. Peritrophic membrane structure and secretion in European corn borer larvae (*Ostrinia nubilalis*). *Tissue & Cell* 29, 463–475.
- He, N., Botelho, J.M.C., McNall, R.J., Belozero, 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. *Ann. Rev. Entomol.* 54, 285–302.
- 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.
- Jang, M., Kong, B., Jeong, Y., Lee, C.H., Nah, J., 2004. Physicochemical characterization of a chitin separated from natural resources. *J. Polym. Sci. Part A Polym. Chem.* 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.
- Kramer, K.J., Corpuz, L., Choi, H.K., Muthukrishnan, S., 1993. Sequence of a cDNA and 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., Xu, R., Zhu, Y.C., Kerwin, J.L., Turecek, F., 2001. Oxidative conjugation of catechols in insect skeletal systems. *Tetrahedron* 57, 385–392.
- 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*, vol. 4. Elsevier Press, Oxford, UK, pp. 111–144 (chapter 3).
- Morris, K., Lorenzen, M.D., Hiromasa, Y., Tomich, J.M., Oppert, C., Elpidina, E.N., Vinokurov, K., Jurat-Fuentes, J.L., Fabrick, J., Oppert, B., 2009. *Tribolium castaneum* larval gut transcriptome and proteome: a resource for the study of the coleopteran gut. *J. Proteome Res.* 8, 3889–3898.
- Moussian, B., 2010. Tailoring of a biological coat: recent advances in insect cuticle differentiation. *Insect Biochem. Mol. Biol.*, in this issue.
- Moussian, B., Veerkamp, J., Müller, 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., Cusson, M. Cloning and characterization of a Gasp homolog from the spruce budworm, *Choristoneura fumiferana*, and its putative role in cuticle formation. *J. Insect Physiol.*, in press, doi:10.1016/j.jinsphys.2009.12.006.
- Pfam website. www.ebi.ac.uk/goldman-srv/pandit/pandit.cgi.
- Rebers, E.J., Riddiford, M.L., 1988. Structure and expression of a *Manduca sexta* larval cuticle gene homologous to *Drosophila* cuticle proteins. *J. Mol. Biol.* 203, 411–423.
- Rebers, J.E., Willis, J.H., 2001. A conserved domain in arthropod cuticular proteins binds chitin. *Insect Biochem. Mol. Biol.* 31, 1083–1093.
- 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., Altig-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., Wjiffels, G., Willadsen, P., 1999. Peritrophic matrix proteins. *Insect Biochem. Mol. Biol.* 29, 87–101.
- Terra, W.R., 1996. Evolution and function of insect peritrophic membrane. *J. Brazil. Assoc. Adv. Sci.* 48, 317–324.
- Terra, W.R., 2001. The origin and functions of the insect peritrophic membrane and peritrophic gel. *Arch. Insect Biochem. Physiol.* 47, 47–61.
- Toprak, U., Baldwin, D., Erlandson, M., Gillott, C., Hegedus, D.D., 2009. Insect intestinal mucin and serine proteases associated with the peritrophic matrix from feeding, starved and molting *Mamestra configurata* larvae. *Insect Mol. Biol.* doi:10.1111/j.1365-2583.2009.00966.x.
- Venancio, T.M., Cristofolletti, P.T., Ferreira, C., Verjovski-Almeida, S., Terra, W.R., 2008. The *Aedes aegypti* larval transcriptome: a comparative perspective with emphasis on trypsin and the domain structure of peritrophins. *Insect Mol. Biol.* 18, 33–44.
- Wang, P., Granados, R.R., 1997. An intestinal mucin is the target substrate for a baculovirus enhancer. *Proc. Natl. Acad. Sci. U.S.A.* 94, 6977–6982.
- Wang, P., Guoxan, L., Granados, R.R., 2004. Identification of two new peritrophic membrane from larval *Trichoplusia ni*: structural characteristics and their functions in the protease rich insect gut. *Insect Biochem. Mol. Biol.* 34, 215–227.
- Willis, J. 2010. Cuticular proteins: lessons from whole genome sequences. *Insect Biochem. Mol. Biol.*, in this issue.
- 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, 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.
- 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.