Available online at www.sciencedirect.com

ScienceDirect

Insect Biochemistry and Molecular Biology 38 (2008) 452–466

Insect Biochemistry and Molecular Biology

www.elsevier.com/locate/ibmb

Domain organization and phylogenetic analysis of the chitinase-like family of proteins in three species of insects

Qingsong Zhu[a], Yasuyuki Arakane[a], Debarshi Banerjee[a], Richard W. Beeman[b], Karl J. Kramer[a,b], Subbaratnam Muthukrishnan[a,*]

[a]*Department of Biochemistry, Kansas State University, 141 Chalmers Hall, Manhattan, KS 66506, USA*
[b]*Grain Marketing and Production Research Center, ARS-USDA, 1515 College Avenue, Manhattan, KS 66502, USA*

Received 7 April 2007; received in revised form 20 June 2007; accepted 27 June 2007

Abstract

A bioinformatics-based investigation of three insect species with completed genome sequences has revealed that insect chitinase-like proteins (glycosylhydrolase family 18) are encoded by a rather large and diverse group of genes. We identified 16, 16 and 13 putative chitinase-like genes in the genomic databases of the red flour beetle, *Tribolium castaneum*, the fruit fly, *Drosophila melanogaster*, and the malaria mosquito, *Anopheles gambiae*, respectively. Chitinase-like proteins encoded by this gene family were classified into five groups based on phylogenetic analyses. Group I chitinases are secreted proteins that are the most abundant such enzymes in molting fluid and/or integument, and represent the prototype enzyme of the family, with a single copy each of the catalytic domain and chitin-binding domain (ChBD) connected by an S/T-rich linker polypeptide. Group II chitinases are unusually larger-sized secreted proteins that contain multiple catalytic domains and ChBDs. Group III chitinases contain two catalytic domains and are predicted to be membrane-anchored proteins. Group IV chitinases are the most divergent. They usually lack a ChBD and/or an S/T-rich linker domain, and are known or predicted to be secreted proteins found in gut or fat body. Group V proteins include the putative chitinase-like imaginal disc growth factors (IDGFs). In each of the three insect genomes, multiple genes encode group IV and group V chitinase-like proteins. In contrast, groups I–III are each represented by only a singe gene in each species.

Keywords: Chitinase; Insect; Phylogenetic analysis; Chitin-binding domain; Glycosylhydrolase; *Tribolium*; *Anopheles*; *Drosophila*; Chitin; Cuticle; Peritrophic membrane

1. Introduction

The insoluble structural polysaccharide, chitin, is a principal structural component of the exoskeleton of insects, the peritrophic matrix or membrane (PM) that lines the midgut, and the cuticular lining of the foregut, hindgut and trachea (Kramer and Koga, 1986; Cabib, 1987; Kramer and Muthukrishnan, 1997, 2005; Lehane, 1997). Although a tough and durable exoskeleton protects insects from environmental injury and predation, it also restricts the insect's growth. The flexible PM, on the other hand, protects the midgut epithelium from mechanical and biological injuries caused by dietary components and pathogens, and also serves as a digestive chamber, regulating the permeability and localization of enzymes and digestion products. During insect growth and development, both the cuticle and PM must be degraded periodically and replaced to allow for growth, maturation and repair.

Chitinolytic enzymes play important roles in shedding of the old cuticle and turnover of the PM. Chitinase genes from many insects have been characterized, including lepidopteran species such as *Manduca sexta* (Kramer et al., 1993), *Bombyx mori* (Kim et al., 1998), *Spodoptera litura* (Shinoda et al., 2001), *Spodoptera frugiperda* (Bolognesi et al., 2005), *Hyphantria cunea* (Kim et al.,

Abbreviations: NAG, β-*N*-acetylglucosaminidase; CHT, chitinase; GlcNAc, *N*-acetylglucosamine; Tc, *Tribolium castaneum*; PM, peritrophic membrane

*Corresponding author. Tel.: +1 785 532 6939; fax: +1 785 532 7278.
E-mail address: smk@ksu.edu (S. Muthukrishnan).

0965-1748
doi:10.1016/j.ibmb.2007.06.010

1998) and *Choristoneura fumiferana* (Zheng et al., 2002); dipteran species such as *Aedes aegypti* (de la Vega et al., 1998), *Anopheles gambiae* (Shen and Jacobs-Lorena, 1997), *Glossina morsitans* (Yan et al., 2002), *Lutzomyia longipalpis* (Ramalho-Ortigao and Traub-Cseko, 2003) and *Chironomus tentans* (Feix et al., 2000); hymenopteran species, *Chelonas* sp. (Krishnan et al., 1994); and coleopteran species such as *Phaedon cochleariae* (Girard and Jouanin, 1999) and *Tenebrio molitor* (Royer et al., 2002; Genta et al., 2006). In most of these species, only one chitinase gene or cDNA has been identified and Southern blot experiments with chitinase cDNA probes have been interpreted to support the existence of a single-copy gene (e.g., Choi et al., 1997). At the protein level, however, multiple chitinases differing in size have been detected in molting fluid secreted by the epidermis and in midgut tissues of lepidopterans. Since all of these chitinases cross-reacted with an antibody to a molting fluid chitinase, it had been assumed that the smaller forms arose as a result of proteolysis of the largest form (Koga et al., 1992). In *B. mori*, the isolation of full-length cDNA clones for chitinases has been reported from several laboratories (Kim et al., 1998; Mikitani et al., 2000; Abdel-Banat and Koga, 2001; Daimon et al., 2003, 2005). The sequence data suggested the presence of at least three chitinase genes in *Bombyx*. A BLAST search of the *Bombyx* EST sequences identified five non-redundant cDNAs, presumably representing five different chitinase or chitinase-like genes (Zhu et al., unpublished data). While four of the chitinases encoded by these ESTs were closely related to other lepidopteran chitinases, one encoded by the gene, *BmChi-h*, displayed a much greater sequence similarity to bacterial and baculoviral chitinases than to insect chitinases, suggesting that an ancestor of *B. mori* acquired this chitinase gene from a bacterium or baculovirus by horizontal gene transfer (Daimon et al., 2005). However, we do not know with certainty the total number of chitinase-like genes in any individual insect species except for *Drosophila melanogaster*. We previously identified 18 chitinase-like genes from the genome of this insect (Zhu et al., 2004). Furthermore, during this investigation, we realized that two chitinase genes previously characterized from *D. melanogaster* were actually part of a single larger gene with multiple catalytic domains, thus reducing the number of *Drosophila* chitinase genes to 16. It is unknown whether other insect species have a correspondingly wide assortment of genes encoding chitinases and chitinase-like proteins.

Previous studies have shown that there are four highly conserved regions or signature sequences in the amino acid sequences of all known insect chitinases (Kramer and Muthukrishnan, 1997; de la Vega et al., 1998; Zhu, 1998). The consensus sequence for conserved region I is KXXXXXGGW, where X is a non-specified amino acid. A second conserved region II has the consensus sequence FDGXDLDWEYP, which is known to be located in or near the catalytic site of the enzyme, with residue E being the putative proton donor in the catalytic mechanism (Watanabe et al., 1994; Lu et al., 2002; Zhang et al., 2002). Consensus sequences for conserved regions III and IV are MXYDXXG and GXXXWXXDXDD, respectively.

With the completion of the genome sequences of several insect species, it is expected that most of the proteins with significant sequence similarity to insect chitinase-like proteins will be identified. The completion of genome sequences from two dipteran species, *D. melanogaster* and *A. gambiae*, and one coleopteran species, *Tribolium castaneum* (all hereafter referred to by their genus names only), provided us with a resource to compare gene families and proteins that are related to chitinases from two different orders of insects and also to investigate the number and domain structures of proteins encoded by these genes as well as their phylogenetic relationships. *Apis mellifera*, *Bombyx mori* and *Aedes aegypti* were omitted from our whole-genome analyses because of the incompleteness of their genome assemblies and annotations at the time this project was conducted.

2. Materials and methods

2.1. Isolation of total RNA from Tribolium and synthesis of first-strand cDNA

The RNeasy Mini kit (Qiagen, Valencia, CA) was used to isolate total RNA from *Tribolium* larvae. Three to four insects (about 10 mg) were collected and total RNA was prepared by following the kit manufacturer's instructions. The animals were weighed and then ground in liquid nitrogen. Extraction buffer was added to the homogenate and the extract was passed through a 20-gauge needle. The tissue lysate was then centrifuged at maximum speed in a microcentrifuge (12,000*g*) for 3 min. The total RNA in the supernatant was recovered and purified using the RNeasy column.

The SuperScript III first-strand synthesis system for RT-PCR (Invitrogen, Carlsbad, CA) was used to synthesize first-strand cDNA by following instructions from the manufacturer. Oligo-$(dT)_{20}$ was used as a primer for reverse transcription and 5 μg of total RNA was used as template. Reverse transcription was carried out at 50 °C for 50 min. Template RNA was removed by treating the cDNA sample with RNase H for 20 min at 37 °C.

2.2. PCR amplification of a putative Tribolium chitinase-like gene fragment

PCR was performed using a P × 2 Thermal Cycler (Thermo Electron, CA) and a pair of degenerate oligonucleotides "DLDW", with the sequence, 5′-TTYGAYGGNYTNGAYYTNGAYTGGGARTAYCCN-3′, and "MDDF", with the sequence, 5′-GGNGCNATGACNTGGGCNATHGAYATGGAYGGYTTY-3′, corresponding to the plus strand of the mRNA encoding the highly conserved region II with the amino acid sequence FDGLDLDWEYP and the negative strand of the mRNA corresponding to the

conserved region IV with the amino acid sequence GAMTWAIDMDDF of family 18 insect chitinases. PCR was carried out for 30 cycles that included the following: annealing (1 min, 55 °C), extension (2 min, 72 °C), and denaturation (45 s at 94 °C) followed by a final extension step (10 min, 72 °C). PCR products were separated on a 1% agarose gel. The amplification product of about 750 bp was excised and purified using a Quantum Prep Freeze 'N Squeeze DNA gel extraction spin column (Bio-Rad, Hercules, CA). This 750 bp DNA fragment was then subcloned into the Topo-vector (Invitrogen, Carlsbad, CA). Six of the positive clones were selected and plasmids were prepared for DNA sequencing. This clone was named *TcCHT5* (see Section 2 for the naming convention used in this study).

2.3. 3′- and 5′-RACE

The 3′-Rapid Amplification of cDNA Ends (RACE) kit (Invitrogen, Carlsbad, CA) was used to clone fragments containing 3′-end sequences of *Tribolium* chitinase-like cDNAs. cDNA was synthesized from total RNA isolated as described above, but using the adapter-oligo-$(dT)_{17}$ primer (AP): 5′-GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTT-3′. Two gene-specific primers from the sense strand were designed, based on the known cDNA sequence or predicted cDNA sequence. One of these two forward gene-specific primers and the reverse AP primer were used for amplification of target cDNAs to yield nested PCR products, which were recovered and cloned into the TOPO-vector as described above. Positive clones were selected and the inserts were sequenced.

The 5′-RACE kit (Invitrogen, Carlsbad, CA) was used to clone the 5′-end sequences of chitinase cDNA clones. Three gene-specific primers from the anti-sense strand were designed near the 5′-end of the known cDNA sequences. First-strand cDNA was synthesized using an anti-sense gene-specific primer. Following thermal inactivation of the reverse transcriptase at 70 °C for 15 min, a mixture of RNase H and RNase T1 was used to degrade the RNA template. First-strand cDNA was purified using a S.N.A.P. column following the manufacturer's instructions. Purified cDNA was tailed by terminal deoxynucleotidyl transferase (TdT) and an abridged anchor primer-binding site was created on the 5′-end of the cDNA. Hemi-nested PCR was used to amplify the 5′-ends of the genes with the anchor primer, followed by cloning in the TOPO vector and DNA sequencing of the insert.

2.4. Southern blotting

Ten micrograms of *Tribolium* genomic DNA were digested with restriction enzymes *Ase* I, *Cla* I, *Hinc* II, *Hind* III or *Xba* I (Promega). Digestion products were separated by agarose gel electrophoresis (1%) and transferred onto Hybond N^+ Nylon membrane (Amersham, Piscataway, NJ) under alkaline conditions and hybridized with random primed ^{32}P-labeled probes prepared using Ready-To-Go DNA labeling beads (Amersham, Piscataway, NJ). The probe was a 323 bp fragment spanning nucleotide positions 549–872 of *TcCHT5* cDNA. The probes were amplified by PCR using cDNA as template with appropriate forward and reverse primers. The membrane was hybridized to the probe at 55 °C, washed under moderate stringency with 1 × SSC containing 0.1% SDS for 30 min and exposed to X-ray film overnight using intensifying screens.

2.5. BAC-library screening with an overgo probe

Two partially complementary oligonucleotides, TcCHT5-F9, 5′-ATACGAGTTCGATGGGTTTGATCTG-3′, and TcCHT5-R15, 5′-GGGGTATTCCCAGTCCAGATCAAAC-3′, with ten complementary nucleotides at the 3′-ends were designed based on the conserved region II of *TcCHT5*. The overlapping oligonucleotide ("overgo") probe was synthesized and labeled with α-^{32}P-dATP and α-^{32}P-dCTP using the Klenow fragment of DNA polymerase I. The labeled probe was used in hybridization experiments with the *Tribolium* BAC library filters.

The BAC library used is described in Arakane et al. (2004). Hybridization was carried out overnight at 65 °C and filters were washed with 2 × SSC with 0.1% SDS at room temperature for 30 min. The filters were exposed to X-ray film overnight with an intensifying screen. Positive clones were identified and these clones were grown for Bac DNA isolation and sequencing.

2.6. Mapping of chitinase genes on Tribolium chromosomes and sequencing

BAC clones confirmed to harbor chitinase-like genes were genetically mapped by single-strand conformational polymorphism (SSCP) analysis. We identified SSCP dimorphisms between two highly inbred *T. castaneum* strains, GA-2 and ab2, using primer pairs specific for end-sequences from each BAC. The BACs were mapped onto a whole-genome recombination map at an average resolution of ca. 1.5 cM using a backcross family that consisted of 179 siblings and using a marker set totaling more than 400 unique DNA sequences derived from BACs, cDNAs and other sources. Details of this mapping procedure are given in Lorenzen et al. (2005). In other cases, chitinase-like genes were first mapped to genome sequence scaffolds using "blastn" program (see below). The positions of these scaffolds on the linkage maps were then determined by identifying SSCP-mapped BACs derived from the same scaffolds.

2.7. BLAST searches of the T. castaneum, A. gambiae and D. melanogaster genome databases

The three databases containing fully sequenced genomes, Flybase (http://flybase.bio.indiana.edu/blast/), Beetlebase (http://bioinformatics.ksu.edu/BeetleBase/) and the *A. gambiae* genome database (http://agambiae.vectorbase.org/index.php) were searched for amino acid sequences

related to previously characterized insect chitinases. To increase the probability of identifying new chitinases including those with low sequence similarity, several insect chitinase-like protein sequences differing substantially in sequence were used as query sequences. These included sequences from dipteran, lepidopteran and coleopteran insects. We also included chitinase-like proteins that differed in the developmental stage or tissue of expression in these queries. These included a molting fluid chitinase from *M. sexta* (GenBank accession # P36362) (Choi et al., 1997) as well as *A. aegypti* (GenBank accession # AAB81849) (Shen and Jacobs-Lorena, 1997) and *P. cochleariae* (GenBank accession # CAA77014) (Girard and Jouanin, 1999) chitinases that are expressed in gut tissues of feeding adults and larvae, respectively. Also used as queries were a fat body-specific chitinase from tsetse fly, *G. morsitans* (GenBank accession # AAL65401) (Yan et al., 2002) and *B. mori* chitinase h (GenBank accession # BAC67246). The latter closely resembles bacterial and baculoviral chitinases, and is expressed in both epidermis and midgut and to a lesser extent in fat body (Daimon et al., 2003).

A protein sequence identified from similarity searches to insect chitinases was categorized as a potential chitinase-like protein if it contained upon visual inspection at least three of the four conserved signature regions (see Introduction) found in amino acid sequences of well characterized family 18 chitinases (Kramer and Muthukrishnan, 1997, 2005; de la Vega et al., 1998; Zhu, 1998). Proteins in which a catalytically critical glutamate residue was replaced by other residues were still included in our set of putative chitinase-related proteins (see Section 2). This glutamate residue in conserved region II serves as the proton donor during the cleavage of the glycosidic bond.

2.8. Modular architectural analysis of identified putative chitinases

SMART modular architectural analysis programs (http://smart.embl-heidelberg.de/) (Schultz et al., 1998) were used to predict the domain architecture of the proteins identified as chitinase-like.

2.9. Multiple sequence alignment and phylogenetic analysis

Multiple sequence alignment was performed using the ClustalW program (http://www.ebi.ac.uk/clustalw/) available at the European Bioinformatics Institute web site (Thompson et al., 1994).

To investigate the evolutionary relationship among the putative chitinases-like proteins identified in this work and others reported in the literature, phylogenetic analyses were performed. Phylogenetic trees were constructed using MEGA 3.0 software (Kumar et al., 2004). Protein sequences used for phylogenetic analysis were extracted from GenBank. Amino acid sequences of chitinases used for this analysis were *A. aegypti* (Accession # AAB81849), *A. gambiae* (accession # AAB87764), *B. mori* (accession # AAB47538), *Chironomus tentans* (accession # CAA73685), *Choristoneura fumiferana* (accession # AAM43792), *Glossina morsitans* (accession # AAL65401), *Hyphantria cunea* (accession # AAB47539), *L. longipalpis* (accession # AAN71763), *M. sexta* (accession # P36362), *P. cochleariae* (accession # CAA77014), *S. litura* (accession # BAB12678) and *T. molitor* (accession # CAD31740) as well as *Drosophila, Anopheles* and *Tribolium* chitinase and chitinase-like sequences. *M. sexta* hemocyte aggregation inhibitory protein (MsHAIP) (accession # AAB32418) was also included because of its similarity to chitinases.

3. Results

3.1. Isolation of Tribolium chitinase-like cDNAs

Degenerate primers corresponding to conserved regions II and IV of family 18 chitinases (Kramer and Muthukrishnan, 2005) were used in RT-PCR using total RNA prepared from penultimate and last instar larvae, prepupae, pupae and adults as RNA templates. A predominant 750 bp fragment in an RT-PCR reaction with a prepupal cDNA template was cloned and sequenced. This fragment was found to contain a partial open-reading frame encoding a protein with high sequence similarity to known insect chitinases (Kramer et al., 1993; Kim et al., 1998; Shinoda et al., 2001; Zheng et al., 2002). This partial cDNA (*TcCHT5*) was used in Southern blot analysis and BAC library screening to obtain an estimate of the number of chitinase-related genes and to identify the BACs containing these genes.

Tribolium genomic DNA was digested with restriction enzymes that lack recognition sites within the *TcCHT5* cDNA probe. Hybridization with a labeled probe showed the presence of one strong band and several minor bands in five different restriction enzyme digests (Fig. 1), suggesting the presence in the *Tribolium* genome of a single-copy of the *TcCHT5* gene, as well as several additional chitinase-related genes. To identify the BAC clones containing these chitinase-related sequences, ^{32}P-dATP- and ^{32}P-dCTP labeled "overgo" probes derived from *TcCHT5* and corresponding to a highly conserved region in chitinases were hybridized to a 6× *Tribolium* genomic BAC library. Sixty-four colonies hybridized to the probe. DNA sequencing confirmed the existence of multiple chitinase-like genes in the *Tribolium* genome.

To determine the protein sequences encoded by the members of the chitinase-like gene family, we attempted to obtain several corresponding full-length cDNA clones. The availability of shotgun genomic sequences from *Tribolium* in the spring of 2004 allowed a search of this partial genome sequence for sequences related to insect chitinase. The *M. sexta* chitinase protein sequence (Genbank accession A56596) was used as the initial query sequence to search the *Tribolium* shotgun genomic sequences using *tblastn*. Genomic sequences encoding segments of

five new putative chitinases, *TcCHT2*, *TcCHT6*, *TcCHT7*, *TcCHT10* and *TcCHT16*, were identified by this approach. The coding sequences were predicted using Genscan software (http://genes.mit.edu/GENSCAN.html). Primers designed from these five predicted chitinase-like protein-coding regions and from *TcCHT5* were used in an attempt to obtain the full-length cDNA sequences by using 5′- and 3′-RACE. The lengths of the longest cDNA clones ranged from 1218–8254 nucleotides. Clones *TcCHT2*, 5, 6, 7 and 16 contained ORFs corresponding to full-length chitinase-like proteins together with their 5′- and 3′-UTR sequences, and were deemed to represent full-length or near full-length cDNAs. The longest cDNA clone, *TcCHT10,* was 5′-truncated, since it lacked the start codon and a 5′ UTR. Additional cDNA clones for chitinases and chitinase-like family proteins were isolated using 5′ and 3′ primers based on the recently completed genomic sequences (see the following section). The genomic locations, transcript lengths and characteristics of the encoded proteins for all of the *Tribolium* chitinase family genes are listed in Table 1. All cDNA sequences have been deposited in GenBank.

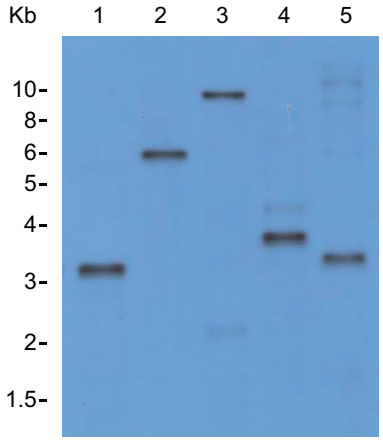

Fig. 1. Southern blot analysis of *T. castaneum* genomic DNA with a *TcCHT5* probe. *Tribolium* genomic DNA (10 μg) was digested with one of the five indicated restriction enzymes. Lane 1, *Ase* I; Lane 2, *Cla* I; lane 3, *Hinc* II; lane 4, *Hind* III and lane 5, *Xba* I. Hybridization was with ^{32}P-labeled *TcCHT5* probe as described in the methods section under moderate stringency. The minor bands probably represent cross-hybridization to other chitinase-like genes.

3.2. BLAST search to identify chitinase-like genes from Tribolium genome database

During the course of this research, the sequencing of the genome of *Tribolium castaneum* was completed. Version 1 of the genome assembly became available to the public in January 2005. The isolation and characterization of full-length cDNA sequences for several chitinase-like proteins (see Table 1) enabled us to deduce the complete sequences of the encoded proteins.

The amino acid sequences of molting-associated chitinases of *M. sexta*, *B. mori, A. aegypti*, *P. cochleariae* and *G. morsitans* as well as the *B. mori* chitinase-h were used to carry out a *tblastn* search of the *Tribolium* genome. This was done to maximize the probability of identifying the full assortment of chitinase-like genes in the database. Ten more putative chitinase-like genes, *TcCHT4, TcCHT8*, *TcCHT9*, *TcCHT11*, *TcCHT12*, *TcCHT13*, *TcCHT14, TcCHT15, TcIDGF2* and *TcIDGF4* were identified in addition to those corresponding to the six fully characterized cDNAs described in Section 3.1. We have also cloned near full-length cDNAs and/or ORF's corresponding to all of these

Table 1
Properties of putative *Tribolium* chitinase-like genes

Gene	GenBank accession #	Predicted protein			Map position[b]	Orientation[c]
		Length[a] (aa)	MW (kDa)	*p*I		
TcCHT2	AY873913	377	42.1	4.6	LG7, 7.0	–
TcCHT4	EF125543	475	50.2	6.4	LG7, 7.0	+
TcCHT5	AY675073	533	60.1	5.8	LGX, 52.7	+
TcCHT6	AY873916	377	40.6	4.5	LG7, 7.0	+
TcCHT7	DQ659247	980	110.9	7.0	LG6, 2.5	–
TcCHT8	DQ659248	496	54.1	4.9	LG7, 7.0	–
TcCHT9	DQ659249	383	42.0	4.6	LG7, 7.0	+
TcCHT10	DQ659250	2700	305.4	6.1	LG9, 52.8	–
TcCHT11	DQ659251	366	39.9	4.8	LG7, 7.0	–
TcCHT12	XM_967709	376	41.4	4.6	LG7, 7.0	+
TcCHT13	DQ659252	377	41.2	4.6	LG7, 7.0	–
TcCHT14	XM_967912	375	41.3	4.5	LG7, 7.0	–
TcCHT15	XM_967984	379	41.6	4.3	LG7, 7.0	–
TcCHT16	AY873915	384	42.0	4.6	LG7, 7.0	+
TcIDGF2	DQ659253	439	49.3	7.0	LG5, 5.5	+
TcIDGF4	DQ659254	431	48.2	7.6	LG5, 5.5	+

[a]Numbers refer to the longest cDNA for the particular gene.
[b]Linkage group, followed by map position in centiMorgans (cM).
[c]Refers to convention used in Genboree genome viewer, Baylor College of Medicine, Human Genome Sequencing Center.

clones with the exception of *TcCHT10* which has an ORF of >8 kb. EST sequences were identified for *TcCHT9* and *TcCHT11*, which corresponded to the full-length cDNAs. The 16 identified putative *Tribolium* chitinase-like genes were assigned gene numbers based on the nomenclature of the previously identified members of the *Drosophila* chitinase-like gene family to which they are most-closely related (Zhu et al., 2004) (Table 1 and Fig. 4). It is to be noted that there is no *CHT1* or *CHT3* gene in the nomenclature because these numbers were originally assigned based on PCR data obtained using genomic DNA of *Drosophila* as the template (de la Vega et al., 1998). *Drosophila CHT1* and *CHT3* were subsequently determined to be portions of a larger chitinase gene, *DmCHT10*. Two other *Tribolium* chitinase family genes are closely related to *Drosophila IDGFs* and these are designated as *TcIDGF2* and *TcIDGF4* based on their sequence similarities to the correspondingly numbered *IDGFs* from *Drosophila*. We did not find any genes in the *Tribolium* genome corresponding to the *DmIDGF1* or *DmIDGF3* genes of *Drosophila*.

3.3. Linkage group assignment of chitinase-like genes in Tribolium

Linkage group assignments, map positions and orientations of the 14 chitinase-like genes and two IDGF-like genes are shown in Table 1. Eleven of the 14 chitinases are clustered within a ~29 kb segment near one end of linkage group 7 (chromosome coordinates not shown). The remaining three are found singly on three different linkage groups, 6, 9 and X. The two IDGF genes are closely located near one end of linkage group 5.

3.4. Characterization of chitinase-like gene families from Anopheles and Drosophila

Insect chitinase sequences were used as query sequences to search the *Anopheles* and *Drosophila* genomic databases in the same manner as described above for *Tribolium*. The results of *tblastn* searches are summarized in Supplementary material; Tables S1 and S2. Thirteen putative chitinase-like proteins were identified from the *Anopheles* genome. The *Anopheles* chitinase-like proteins were classified using a nomenclature based on that of their *Drosophila* counterparts (Zhu et al., 2004). To the extent possible, we have assigned the same number to closely related chitinase-like genes from all three insect species. *Anopheles* has only one gene encoding a protein resembling an imaginal disk growth factor (IDGF), whereas *Drosophila* has six and *Tribolium* has two. The *Anopheles* gene is named *IDGF4* because it has the highest sequence similarity to *Drosophila IDGF4*.

3.5. Domain organization of insect chitinases and chitinase-like proteins

Four conserved motifs (see Section 1) were used as signatures not only to identify potential chitinases or chitinase-like proteins in the available insect genome sequences, but also to begin resolving functional subcategories of such proteins. All three insect species have a large family of genes encoding chitinase-like proteins. The total number of chitinase-related genes in *Drosophila*, *Anopheles* and *Tribolium* are 16, 13 and 16, respectively (Fig. 2). Analysis of the domain organization of the putative chitinase-like proteins in the three species indicates that some of them have more than one catalytic domain as well as more than one ChBD, whereas others lack the S/T-rich linker domain and/or the ChBD. Some similarities are obvious in the number and domain organization of chitinase-related proteins in the three insect species analyzed in detail here. In each of the three insect species, there are two prototype chitinases with a domain organization identical to that of chitinases isolated from molting fluids of lepidopteran insects such as *M. sexta*, *B. mori* and *S. litura*. They all have a signal peptide, a catalytic domain, an S/T-rich linker domain and a ChBD in that order from the N-terminus. These prototypical chitinases are encoded by genes designated *CHT5* and *CHT8*. Chitinases encoded by *CHT4* genes resemble CHT5 chitinases, but they have truncated S/T-rich linker regions. In addition, all three insect genomes contain genes encoding four other chitinases lacking a C-terminal ChBD (chitinases encoded by *CHT 2*, *9*, *11* and *13*). Some additional chitinases lacking ChBD's are found in *Tribolium* and/or *Anopheles* but not in *Drosophila* (*TcCHT6*, *12*, *14*, *15* and *16* in *Tribolium* and *AgCHT6*, *CHT12* and *CHT16* in *Anopheles*). All three species have one large chitinase-like protein with two catalytic domains (*CHT7*), and another very large chitinase with 4 or 5 catalytic domains (*CHT10*). In addition, there is at least one gene encoding an IDGF-like protein in all three species. The amino acid sequence similarities/identities among IDGFs from the three species are greater than the similarities/identities with other conspecific chitinases. In addition, the IDGFs share additional structural similarities, as discussed below.

The domain architectures of all chitinase-like proteins in *Tribolium*, *Anopheles* and *Drosophila* are shown in Fig. 2. CHT7 and CHT10 are large chitinases with two or more catalytic domains in the same ORF. *Anopheles* and *Drosophila* CHT10 proteins contain four catalytic domains each, while *Tribolium* CHT10 has five. There is one ChBD in CHT7, which has two catalytic domains, whereas the CHT10s from the three species have either four or five ChBDs. The pattern of distribution of ChBDs between catalytic domains also appears to be conserved among the three CHT10 proteins. Most of the chitinases are predicted to contain a cleavable signal peptide and to be secreted proteins. Some *Anopheles* chitinases appear not to have signal peptides, but this is probably due to our inability to determine the signal peptide-coding region in the absence of 5′-RACE data or EST evidence. *Drosophila*, *Anopheles*, *Apis* and *Tribolium* CHT7 chitinases have a predicted transmembrane segment at the N-terminal region. DmCHT7 and AgCHT7 have, in

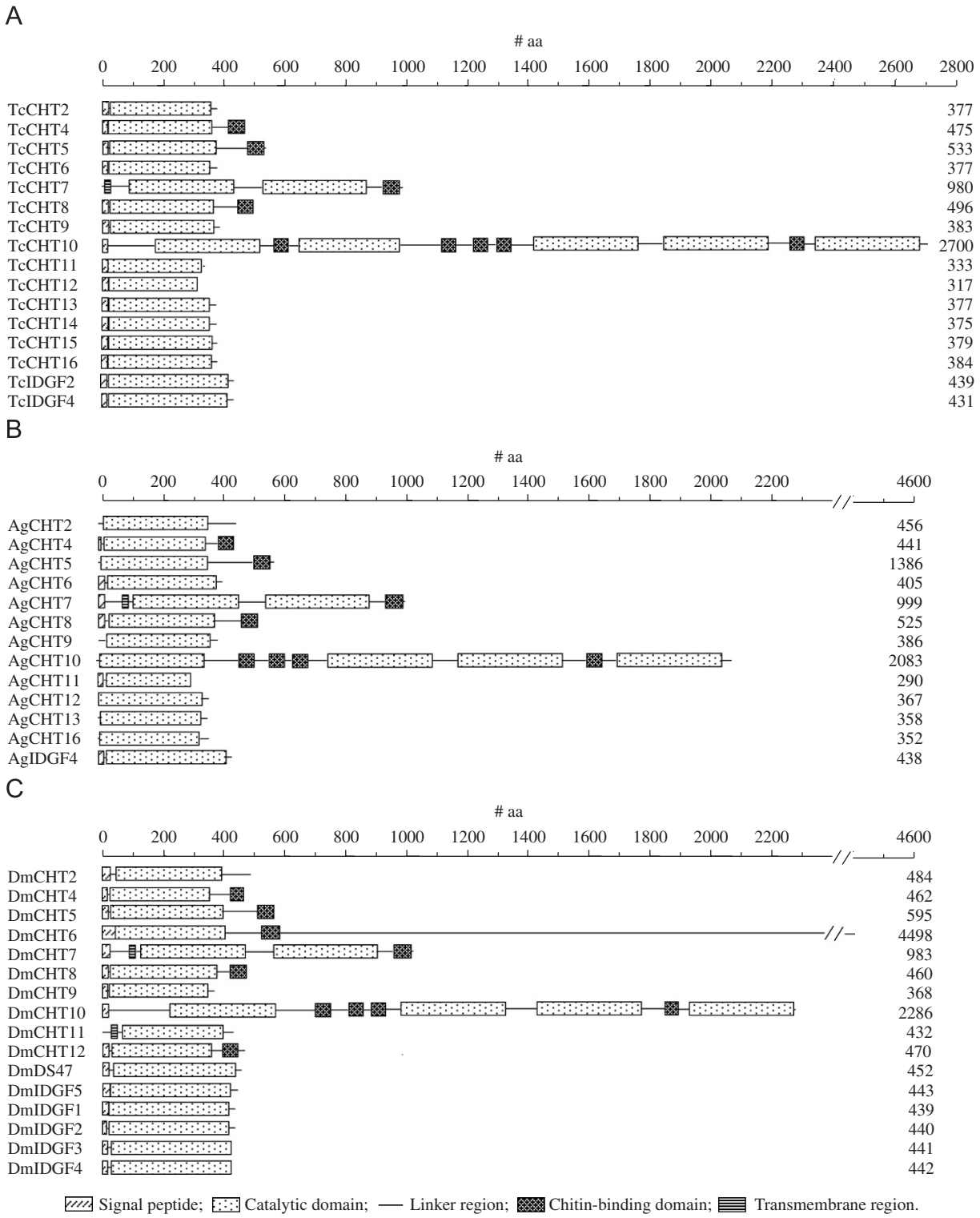

Fig. 2. Domain architecture of putative chitinases and chitinase-like proteins from *Tribolium* (A), *Anopheles* (B) and *Drosophila* (C). The program SMART was used to analyze the identified domains. DmDS47 is an IDGF-like glycoprotein described by Kirkpatrick et al. (1995). TcCHT7, DmCHT7 and AgCHT7 each have a single transmembrane domain before the first catalytic domain.

addition, a putative signal peptide. Thus, all of these proteins may be membrane-anchored, which leaves the importance of the cleavable signal peptide in DmCHT7 and AgCHT7 unclear. No transmembrane motif is apparent in any of the other chitinases.

Both AgCHT10 and DmCHT10, each of which has four catalytic domains, have a glutamate→asparagine (E→N) substitution in the DWEYP motif in the first catalytic unit. The other three catalytic units in each of these large proteins retain the catalytically important E

residue in this motif. TcCHT10 has the same substitution (E→N), but only in the second catalytic unit. The fourth catalytic domain of DmCHT10 and first catalytic domain of TcCHT10 have a different substitution in the DWEYP motif, namely D→A (DmCHT10) or D→H (TcCHT10). Proteins with these substitutions are expected to have very little or no enzymatic activity (Lu et al., 2002). Therefore, all of these "glutamate- or aspartate-mutated" catalytic domains are predicted to be non-catalytic, whereas the other domains are expected to be catalytically active.

3.6. Multiple sequence alignment and phylogenetic trees

Amino acid sequences from the catalytic regions and ChBDs from all of the sequences reported in this paper plus other fully characterized insect chitinases and chitinase-like proteins were aligned using the Clustal W program. Within the catalytic domain, regions I to IV containing the signature sequences of family 18 chitinases were highly conserved. As expected, conserved region II containing a glutamate (proton donor) has the greatest sequence conservation (data not shown). The IDGF alignments revealed that when compared to the chitinolytic proteins, there are additional amino acids between β-sheet 4 (β4) and α-helix 4 (α4), between α helix-c and β sheet-c, and between β sheet-d and β sheet-e in the triose phosphate isomerase (TIM)-like barrel structure of the IDGFs and also *Manduca* hemocyte aggregation inhibitory protein (HAIP) (Varela et al., 2002) (see also Section 4.7). The alignment of this narrow region within the catalytic domain is shown in Fig. 3. The most striking difference between the IDGFs and other chitinase-like proteins is the presence of a stretch of 24 extra amino acid residues between β4 and α4 in the former. These additional sequences are conserved in all IDGFs and HAIP but are absent in family 18 chitinases.

Genes/cDNAs encoding family 18 chitinases have been cloned from seventeen insect species so far. To establish the evolutionary relationships among these insect proteins, a phylogenetic tree was constructed based on amino acid sequences of the catalytic domains (Fig. 4). The insect chitinase-like proteins have been placed into five major groupings. Group I includes most of the prototypical chitinases, all with a single copy each of the catalytic domain, ChBD and an intervening S/T-rich linker region. Several of these enzymes have been shown to be synthesized in epidermal cells and are associated with cuticle digestion and molting in lepidopteran, dipteran and coleopteran species. Group II consists of chitinases with 4–5 catalytic domains, including DmCHT10, AgCHT10 and TcCHT10 [identified in this work and also in Zhu et al. (2004)], as well as a chitinase from another beetle, *T. molitor*, which contains five catalytic domains and five ChBDs (Royer et al., 2002). CHT7 chitinases were placed as members of group III, all of which contain two catalytic domains, one ChBD and a membrane-anchoring region. Group IV is made up of a very divergent group of proteins containing a single catalytic domain. Many, but not all, lack a ChBD. Even though not all members of this group have been fully studied, those that have been characterized are gut- and fat body-specific chitinases. Two subgroups were observed within group IV, one including chitinases primarily from dipteran, hemipteran and lepidopteran species, and the other containing only coleopteran chitinases from *Tribolium* and *Tenebrio* (this work and Genta et al. 2006). Chitinase-like imaginal disc growth factors (IDGF) and *M. sexta* HAIP fall into group V. There were a few outliers that could not be precisely grouped such as *Anopheles* CHT11.

According to the Carbohydrate-Active enZYmes (CAZY) database (http://afmb.cnrs-mrs.fr/CAZY/index.html), ChBDs of insect chitinases belong to carbohydrate-binding module family 14. The amino acid sequences

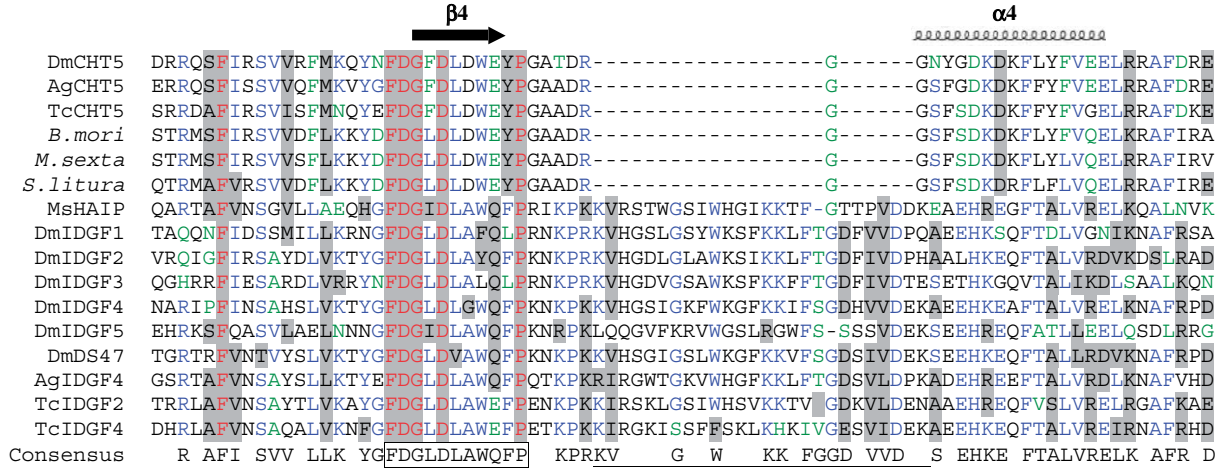

Fig. 3. Multiple sequence alignment of a sample of catalytic domains from typical insect chitinases (group I) and imaginal disc growth factors (IDGFs, group V). The sequences were aligned with the ClustalW program. The conserved region II is boxed. The extra loop sequences of the IDGFs between β-sheet 4 and α-helix 4 structures of the TIM barrel are underlined in the consensus sequence.

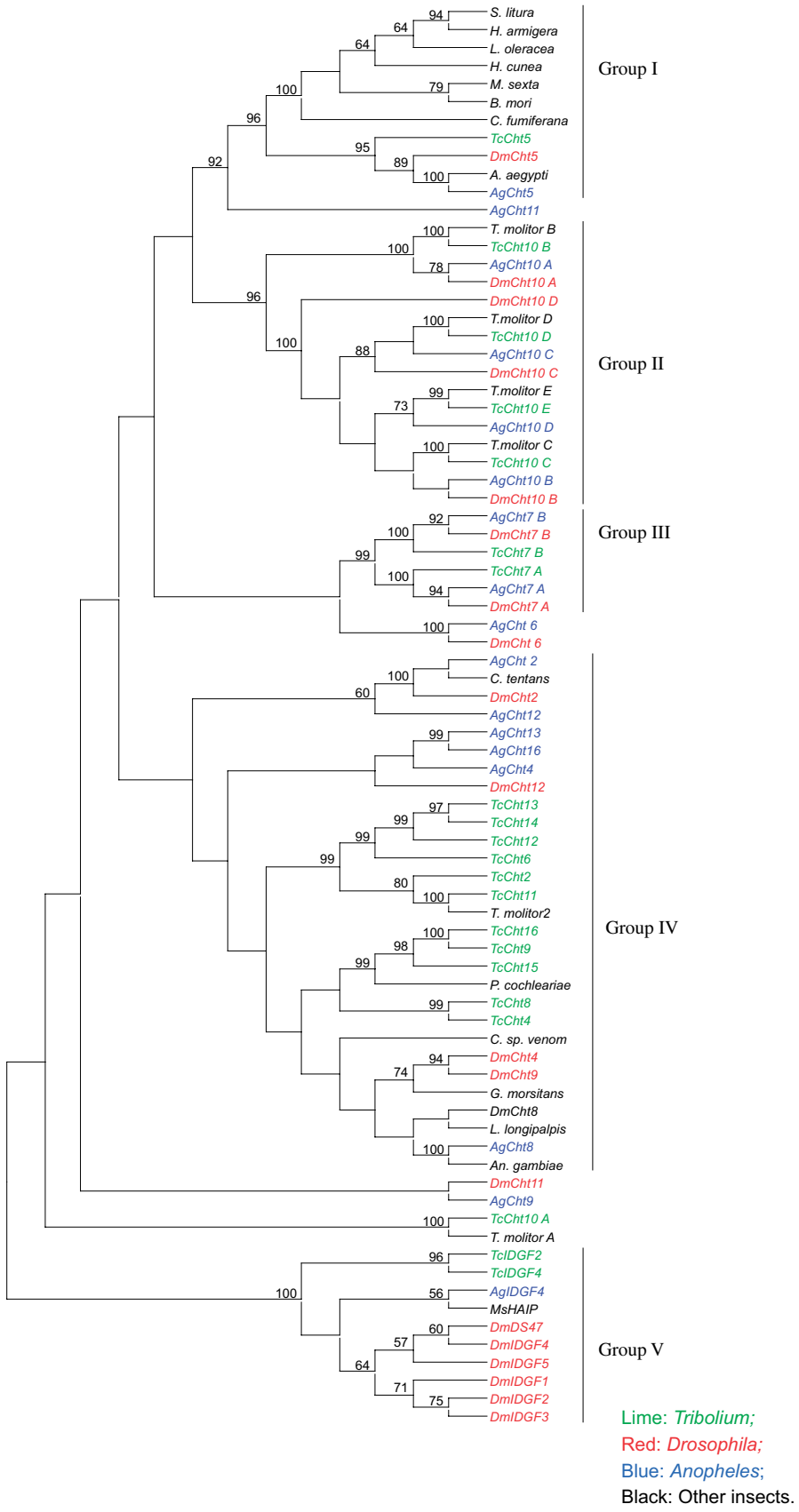

Fig. 4. Phylogenetic analysis of catalytic domain sequences of putative chitinase and chitinase-like proteins from *Drosophila*, *Anopheles*, *Tribolium* and other insects. A consensus phylogenetic tree was constructed using the software MEGA 3.0 (Kumar et al., 2004). A bootstrap analysis of 5000 replications was carried out on the trees inferred from the minimum evolution method. Bootstrap values are shown in the cladogram if the values are higher than 50%. The letters A through E denote multiple catalytic domains in the same gene.

of the ChBDs from 11 previously characterized chitinases and the putative chitinase-like proteins from this work also were aligned (data not shown). There are six cysteine residues in all of these ChBDs. Four of these, which are located in the middle of this domain, are completely conserved in all of the sequences. The other two, near the ends of the domain, were highly, but not universally conserved. The phylogenetic tree (Supplementary material Fig. S1) generated with ChBD sequences was consistent with the tree generated with catalytic domain sequences (Fig. 4). However, because group V chitinase-like proteins (and some members of group IV) contain no ChBD, the ChBDs were classified into only four groups instead of five.

4. Discussion

4.1. Multiple chitinases are represented in insect genomes and transcriptomes

Chitinases and their cDNAs have now been characterized from several insect species. Most of the previous studies involving lepidopteran, dipteran and coleopteran insects have identified only one or two genes encoding chitinase-like proteins, consistent with two major locations of chitin degradation in these insects, the cuticle and PM. PCR amplifications using genomic DNA templates and primers corresponding to conserved regions had led to the identification of 3–4 different chitinase-like gene sequences in dipterans (*Aedes*, *Anopheles* and *Drosophila*) (de la Vega et al., 1998). With the completion of the *Drosophila*, *Anopheles* and *Tribolium* genomic sequences, we were able to expand on those studies and carry out an extensive search of the genomes of these insects for chitinase-related genes. The genome assemblies of silkworm and honeybee were not yet finished at the time this study was conducted, so these insects were not included in the analysis. The computational analyses identified 13 and 16 putative chitinase or chitinase-like genes from *Anopheles* and *Tribolium,* respectively, and 16 chitinase-like genes from *Drosophila.* We had previously reported the presence of 18 chitinase-like genes in *Drosophila* (Zhu et al., 2004), which included two genes, *DmCHT1* and *DmCHT3*, identified by de la Vega et al. (1998). Further analysis revealed that these two sequences were actually derived from a single larger gene, *DmCHT10*, which has multiple catalytic domains. To avoid confusion, the nomenclature of the chitinase-like genes used in this paper excludes these two numbers.

Our studies have established that there is a high degree of complexity in the chitinase-like gene families in both coleopteran and dipteran insects. The system we used for gene numbering represents the most complete correspondence of nomenclature possible in view of the large number of chitinase-like genes and the evolutionary divergence among the 17 species examined.

cDNAs for all of the putative *Drosophila* chitinase-like genes have been obtained, confirming that all genes of the chitinase family in this species are transcribed. EST databases have been searched for *Anopheles* and *Tribolium* chitinase-like cDNAs. We found at least one EST sequence corresponding to every chitinase-like gene identified in the *Anopheles* genome, which allowed the determination of exon–intron organization and protein sequences of most members of this family of genes. We have cloned the full-length cDNAs corresponding to many of the chitinase-like genes of *Tribolium*. In other cases EST sequences were identified from available EST databases, which matched the coding sequences predicted from genomic sequences. In some cases, we were able to amplify the entire ORF with partial 5′-UTR and 3′-UTR sequences using primers designed from genomic sequences and mRNA derived from insects at different developmental stages. The presence of these sequences in the EST databases of *Drosophila, Anopheles* and *Tribolium* suggests that these putative chitinase or chitinase-like genes may all be functional and are not pseudogenes.

4.2. Evolution and phylogeny of chitinase-family proteins

The physical clustering of related genes on chromosomes suggests that repeated cycles of gene duplication and functional divergence gave rise to the current number and variety of chitinase-like genes from an ancestor that had as few 5 chitinase precursor genes. This inference is based on the observation that the phylogeny of the catalytic domain contains 5 distinct branches that are well correlated to the physical groupings of the constituent genes (see Figs. 2 and 3). Each branch corresponds to a unique physically isolated gene or gene cluster with the exception of *TcCHT10*, which contains five catalytic domains. Translations of domains B-E from this gene are clustered with group II chitinases of *Drosophila* and *Anopheles*, but domain A falls into an outlier group along with domain A of another multi-catalytic domain chitinase from another coleopteran, *Tenebrio molitor.* The group III enzyme TcCHT7 contains two catalytic domains, a ChBD and a membrane anchoring domain. Group IV chitinases from *Tribolium* consists of 11 chitinases clustered on LG7. The clustering of multiple chitinases from the same insect in distinct sub-branches of the phylogenetic tree is consistent with the notion that many of these genes arose from more recent gene duplications.

We have characterized many members of chitinase-like gene families from three insects, *Tribolium*, *Drosophila* and *Anopheles*, which belong to two different orders, *Coleoptera* and *Diptera*. We believe that all or nearly all members of the chitinase family in each of these three species have been identified. We have used multiple probe sequences of chitinases and IDGFs characterized from related as well as evolutionarily divergent species with substantially different amino acid sequences. Although the number of chitinase-like genes is somewhat variable, all of the three species analyzed here contain a rather large chitinase-like gene family consisting of from 13 to 16 members. Phylogenetic analysis using the program MEGA3 indicated that insect

chitinase-like proteins fall into five related groups, each having at least one representative in each of the three insect species examined in this study (Figs. 4 and S1). One prototypical insect chitinase-like gene belonging to group I was identified in all species of insects analyzed here. About half of the well-characterized insect chitinase-like genes/cDNAs belong to this group. Group I proteins have a typical multi-domain architecture that includes an N-terminal catalytic domain, an S/T-rich linker domain and a C-terminal chitin-binding domain (ChBD). The catalytic domains are highly conserved and their sizes are very similar among all members of the group I chitinases, regardless of the taxonomic order of origin. The catalytic domains contain about 370 amino acids. This size conservation may be dictated by the need for the catalytic domain to assume a $\beta_8\alpha_8$ barrel structure, as has been revealed by crystal structures of several family 18 chitinases and IDGFs (Terwisscha van Scheltinga et al., 1994; van Aalten et al., 2000; Fusetti et al., 2002; Varela et al., 2002). These studies indicate that a minimum of approximately 370 amino acid residues are needed to generate the $\beta_8\alpha_8$ barrel structure.

The ChBDs found in insect chitinases all belong to the family 14 carbohydrate-binding module (Coutinho and Henrissat, 1999). Their amino acid sequences are less well conserved than those of the catalytic domains, but, nonetheless, phylogenetic relationships among the ChBD's of the different groups of chitinases correspond to those observed with the catalytic domains of the same groups, suggesting co-evolution of these two domains. ChBDs, which are only 50–60 amino acid residues long, include six cysteine residues whose relative locations are very highly conserved. The proposed function of the ChBD is to help localize the enzyme on the insoluble substrate, which enhances the efficiency of chitin degradation (Linder and Teeri, 1997). Carbohydrate-binding module family 14 also includes members found in several microbial chitinases. These ChBDs can be found at either the N- or C-terminus and may be present as a single copy or as multiple repeats. These domains are not only cysteine-rich, but they also have several highly conserved aromatic residues (Shen and Jacobs-Lorena, 1999). The cysteine residues help to maintain protein folding by forming disulfide bridges, whereas the aromatic residues probably interact with saccharides in the ligand-binding pocket. The aromatic residues in the cellulose ChBD of cellulases have been proposed to contribute to binding at the surface of cellulose (Simpson and Barras, 1999). There is a linker region between the catalytic domain and ChBD in many of these proteins. The linker regions in most members of group I, II and III chitinases are rich in serine and threonine residues, and are predicted to be *O*-glycosylated. The length of the linker region varies substantially among chitinases.

4.3. Group I chitinases

Group I chitinases most likely code for chitinases involved in molting, because many proteins in this group were obtained from the molting fluid and/or integument of several lepidopteran, dipteran and coleopteran species. All of the group I chitinase-like genes are highly expressed in the epidermis. Transcripts for these chitinases were also detected in the gut of *M. sexta* by northern blot analysis (Kramer et al., 1993) and in guts of *H. armigera* and *S. frugiperda* by RT-PCR (Zheng et al., 2002; Ahmad et al., 2003; Bolognesi et al., 2005). In many of these species, only one or two chitinase genes have been characterized, and therefore it is unclear whether there was cross-hybridization of the probe to other transcripts or whether sequence conservation in the regions chosen for the design of gene-specific primers led to the production of a mixture of RT-PCR amplification products of nearly the same size. It is interesting to note that transcripts for a molting-associated chitinase cDNA were detected in epidermis but not in the midgut of *C. fumiferana*, even though a chitinase was detected in gut tissues of *C. fumiferana* using a chitinase antibody (Zheng et al., 2002). The reason for this discrepancy is unclear. It is likely that gut and epidermis express different tissue-specific chitinases, but this point has not been rigorously examined due to difficulties associated with obtaining epidermal and gut tissues free of contamination with other tissues such as trachea. Nevertheless, in all cases, group I chitinases and their transcripts appeared before ecdysis and disappeared after ecdysis, suggesting that the main function of these enzymes is cuticle chitin degradation and/or PM turnover.

4.4. Group II chitinases

Group II chitinases are unusually large and contain four or five catalytic domains and four or five ChBDs, some of the latter being clustered between the catalytic domains. The last three chitinase units (catalytic domain plus ChBD) of AgCHT10, DmCHT10, TcCHT10 and *T. molitor* chitinase (Royer et al., 2002) share significant amino acid sequence similarity and more than 50% amino acid sequence identity. Domain duplication and shuffling during the evolution of group II proteins might have led to the formation of larger proteins with chitinolytic and chitin-binding activities in a common ancestor of Lepidoptera and Coleoptera. Coleoptera differ from Diptera in the number of group II chitinase units, Coleoptera having five (Fig. 2) and Diptera only four. Interestingly, the first catalytic domain of coleopteran chitinase-like proteins belonging to this class (*Tribolium* TcCHT10 and *Tenebrio* chitinase) showed very high sequence similarity to each other but had much lower sequence similarity to other catalytic units within the same protein, suggesting that coleopterans acquired this unit more recently in evolution than lepidopterans. The first and second units of coleopteran group II chitinases are predicted to lack chitinase activity due to the substitution of critical residues in conserved region II (FDGLDWEYP). In the first catalytic unit, the second aspartic acid residue in this conserved region in both TcCHT10 and *Tenebrio* chitinase is replaced

by histidine. Loss of enzymatic activity was observed in *Manduca* chitinase when the equivalent D144 residue was replaced by non-acidic residues (Lu et al., 2002). The proton donor glutamate in conserved region II is replaced by glutamine in the second unit in both coleopteran chitinases. Site-directed mutagenesis of this glutamate to a glutamine (or other amino acids) resulted in proteins devoid of activity in *M. sexta* chitinase (Lu et al., 2002). Substitution of critical catalytic residues was also observed in the dipteran group II chitinases. The first unit of DmCHT10 and AgCHT10 and fourth unit of DmCHT10 are also predicted to lack enzymatic activity because of similar substitutions. The finding that both dipteran and coleopteran chitinase gene families include a chitinase with multiple catalytic domains and ChBDs suggests that this chitinase serves a function that cannot be fulfilled by other chitinases. Our RNAi studies with *Tribolium* have supported this conclusion (Zhu et al., unpublished data). It is possible that these large chitinases with multiple catalytic domains and ChBDs facilitate the melting of crystalline chitin microfibrils that contain multiple chains of chitin into single unbundled chains, thus making them more accessible to the active sites of its own catalytic domain(s) and those of other chitinases.

A 3.7 kb cDNA from *B. mori* predicted to encode an ecdysteroid-inducible chitinase-like protein with two family 18 catalytic domains in which the proton donor glutamate was replaced with other residues was reported by Takahashi et al. (2002). However, the transcript corresponding to this cDNA had a size of about 10 kb, suggesting that this cDNA probably represents a truncated form of a cDNA derived from a gene encoding a member of group II chitinases.

4.5. Group III chitinases

Group III insect chitinases possess two catalytic domains. TcCHT7 and its orthologs in *Drosophila* and *Anopheles*, DmCHT7 and AgCHT7, contain two catalytic domains and one C-terminal ChBD. The amino acid sequences reveal highly conserved motifs characteristic of family 18 chitinases. The first catalytic domains of CHT7 from all three insects share greater sequence similarity among themselves than when compared to the second catalytic domain, suggesting somewhat distinct functions for each catalytic domain. Hard tick (*Haemaphysalis longicornis*) chitinase is the only biochemically well characterized protein that contains two family 18 chitinase catalytic domains. It has a domain architecture similar to that of TcCHT7, containing two catalytic domains and one chitin-binding peritrophin A domain. The purified enzyme with a size of 108 kDa does have chitinase activity (You et al., 2003). Immunohistochemical analysis showed that this chitinase is present in the space between the old and new cuticle in molting nymphs, suggesting a role in molting physiology. Most insect chitinases are predicted to have cleavable N-terminal signal peptides and to be extracellularly localized because they lack other targeting signals. However, CHT7 from *Tribolium*, *Drosophila, Apis* and *Anopheles* were predicted to contain an N-terminal membrane-anchored region and to be membrane proteins based on the transmembrane helices prediction program, TMHMM version 2 (http://www.cbs.dtu.dk/services/TMHMM/). CHT7 proteins from *Drosophila* and *Anopheles* have, in addition, a cleavable leader peptide. The reason why the *Tribolium* and *Apis* paralogs do not have the leader peptide is unclear. A rather unique chitinase, ChiA, with dual catalytic domains and triple substrate-binding domains, was found in the bacterium, *Pyrococcus kodakaraensis* KOD1 (Tanaka et al., 1999). The N-terminal and C-terminal catalytic domains function independently from each other with the former functioning as an exochitinase and releasing only chitobiose from colloidal chitin and the latter as an endochitinase releasing *N*-acetylchitooligosaccharides of varying lengths including chitobiose. A synergistic effect in chitin degradation was observed when both of the catalytic domains were present in a single protein. The substrate-binding domains apparently play an important role in insoluble chitin binding and hydrolysis (Tanaka et al., 1999). It will be interesting to determine whether the insect chitinases of group III have properties similar to *Pyrococcus* chitinase.

4.6. Group IV chitinases

Group IV represents the most divergent class of chitinases. It includes several chitinases expressed in the gut or fat body. Chitinases and cDNAs belonging to this group have been isolated from guts of several insects. Shen and Jacobs-Lorena (1997) isolated a chitinase cDNA from *A. gambiae*, which is expressed only in the adult gut but not in larval gut, larval carcass or adult carcass. Secretion of this chitinase into the gut was triggered by blood feeding. A midgut chitinase was also induced by blood-feeding of the sand fly, *L. longipalpis* (Ramalho-Ortigao and Traub-Cseko, 2003). It was detected only in the blood-fed midgut and reached its peak at approximately 72 h post-blood meal. A larval gut-specific chitinase cDNA was isolated from feeding larvae of the beetle, *P. cochleariae* (Girard and Jouanin, 1999). A chitinase that is expressed only in adult fat body but not from larvae and pupae was isolated from tsetse fly, *G. morsitans* (Yan et al., 2002), Unlike *A. gambiae* and *L. longipalpis* gut-specific chitinases, which contained an N-terminal catalytic domain and a C-terminal ChBD, the mustard beetle (*P. cochleariae*) gut-specific chitinase contained only a single catalytic domain. There was no C-terminal ChBD and the protein lacked a serine/threonine-rich linker region. The tsetse fly chitinase has a catalytic domain and the ChBD but lacks the S/T-rich linker domain. Recently, a gut-specific chitinase and its cDNA clone from a beetle, *T. molitor,* have been reported (Genta et al., 2006), which is most closely related to TcCHT11 (>70% amino acid sequence identity) and is classified as a group IV chitinase. This chitinase, which

lacks both the ChBD and an S/T-rich linker domain, has a very low activity against colloidal chitin but hydrolyzes chitooligosaccharides very efficiently. In the case of *Tribolium*, we found at least three chitinases belonging to this group, all of which were expressed at high levels in the larval gut (Zhu et al., data not shown). Thus, many members of this group appear to be expressed in larval and adult guts, particularly in response to feeding. Even though not all members of this family of chitinases have been analyzed extensively, based on the limited data currently available, we predict that most if not all of the group IV chitinases are likely to be expressed in the gut or fat body. These enzymes also appear to have diverse arrangements of the sub-domains with many of them lacking either the ChBD and/or an S/T-rich linker domain.

4.7. Group V proteins

The putative chitinase-like imaginal disc growth factor (IDGFs) genes identified here and in previous studies with *Drosophila* and *Anopheles* all belong to glycosylhydrolase family 18 and fall into group V in the phylogenetic tree (Fig. 4). Included in this group is another protein, HAIP from the hemolymph of *M. sexta*, which inhibits hemocyte aggregation (Kanost et al., 1994). The crystal structure of *Drosophila* IDGF2 and homology modeling of other *Drosophila* chitinase-like proteins suggest that all of these proteins exhibit the $(\beta/\alpha)_8$ TIM barrel structure characteristic of some glycosylhydrolases (Varela et al., 2002; Zhu et al., 2004), indicating that this structure is important for binding and/or hydrolysis of polymeric carbohydrate substrates (Davies and Henrissat, 1995). Although the insect IDGFs contained all four of the conserved regions characteristic of family 18 proteins, the *Drosophila* and *Anopheles* proteins but not the *Tribolium* proteins (see below) lack the glutamate residue that has been identified as the proton donor in the catalytic mechanism. Many of the IDGFs also lack this E residue (typically substituted by a Q residue) and are predicted to be devoid of chitinase activity. The IDGF proteins are probably carbohydrate-binding proteins lacking enzymatic activity. Most likely, these proteins bind to chitin or other carbohydrates containing *N*-acetylglucosamine (GlcNAc). For example, the binding of IDGFs to cell surface glycoproteins may be involved in cell–cell communication, cell proliferation or insect immunity. It is believed that IDGFs have evolved from an ancestral chitinase and acquired a new growth-promoting function (Kawamura et al., 1999).

Another remarkable feature of all identified IDGFs and *Manduca* HAIP is that these proteins have extra loop sequences between β4 and α4, αc and βc and βd and βe components in the TIM barrel structure (Zhu et al., companion paper). Especially interesting is the finding that the extra sequence between β4 and α4 is highly conserved. The precise biological function of these extra loop sequences is unknown. In the crystal structure of *Drosophila* IDGF2, this extra loop amino acid stretch assumes an α-helical structure that probably interferes with substrate binding and/or catalysis (Varela et al., 2002). We propose that these extra sequences could be used as signature motifs for the identification of other insect IDGF-like proteins.

In the soil fungus, *Trichoderma virens*, three genes encoding 42 kDa chitinases, *Tv-ech1*, *Tv-ech2* and *Tv-ech3*, were identified (Kim et al., 2002). These genes are regulated in response to different environments and/or developmental stages or are expressed in different subcellular locations. Although multiple chitinases have been identified in molting fluid, integument and fat body of several insects, it has not been established whether they have distinctive roles in different tissues during different stages of development. *B. mori* was the first species from which multiple chitinase genes were identified (Kim et al., 1998; Mikitani et al., 2000; Abdel-Banat and Koga, 2001; Daimon et al., 2003, 2005). In this paper, we have demonstrated that in addition to *Drosophila*, another dipteran and a coleopteran species also have a large family of chitinases with different domain organizations. Here, we have attempted to identify all members of families of chitinase-like proteins from *Tribolium*, *Anopheles* and *Drosophila* (Zhu et al., 2004). Possible biological functions of individual chitinase-like genes will be addressed in a separate publication (Zhu et al., unpublished data).

A phylogenetic analysis of the amino acid sequences of members of the insect chitinase families suggests that all of these proteins evolved from a common ancestor that preceded the separation of the dipteran and coleopteran lineages. This is evident from the finding that group I (for which we have the most representatives) includes chitinases from eleven insect species of widely divergent lepidopteran and coleopteran lineages, which share a high degree of conservation of amino acid sequence and domain organization. The data also support the hypothesis that duplication of domains to form larger enzymes with multiple catalytic domains and ChBDs or loss of critical residues in the catalytic domain to form catalytically inactive IDGFs and related proteins also occurred early during evolution of insects because of the retention of a similar distribution of the members within the chitinase-like protein families in several species belonging to the dipteran and coleopteran orders. However, group IV chitinase-like proteins, which consist of mostly gut and fat body chitinases, may have evolved more recently as shown by a clustering of dipteran and coleopteran chitinases into two separate sub-groups within group IV proteins. It will be interesting to carry out a similar analysis of other chitinase-like protein families from Lepidoptera and other orders when their genome sequences become available so that a better understanding of chitinase-like gene evolution can be achieved.

Acknowledgments

This research was supported in part by NSF Grants #IBN-0316963 and 0615818. This is Kansas State Agricultural Experiment Station Contribution #07-210-J.

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 US Department of Agriculture or Kansas State University.

We are grateful to Sue Hass for help with rearing insects, Barbara Van Slyke for help with Bac library screening, Drs. Marce Lorenzen, Wanlong Li, Yoonseong Park and Srini Kambhampati for helpful suggestions and several protocols used throughout this work.

There is no conflict of interest.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ibmb.2007.06.010.

References

Abdel-Banat, B.M., Koga, D., 2001. A genomic clone for a chitinase gene from the silkworm, *Bombyx mori*: structural organization identifies functional motifs. Insect Biochem. Mol. Biol. 31, 497–508.

Ahmad, T., Rajagopal, R., Bhatnagar, R.K., 2003. Molecular characterization of chitinase from polyphagous pest *Helicoverpa armigera*. Biochem. Biophys. Res. Commun. 310, 188–195.

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.

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.

Cabib, E., 1987. The synthesis and degradation of chitin. Adv. Enzymol. Relat. Areas Mol. Biol. 59, 59–101.

Choi, H.K., Choi, K.H., Kramer, K.J., Muthukrishnan, S., 1997. Isolation and characterization of a genomic clone for the gene of an insect molting enzyme, chitinase. Insect Biochem. Mol. Biol. 27, 37–47.

Coutinho, P.M., Henrissat, B., 1999. Carbohydrate-active enzymes: an integrated database approach. In: Gilbert, H.H., Davies, G.J., Henrissat, H., Svensson, B. (Eds.), Recent Advances in Carbohydrate Bioengineering. Royal Society of Chemistry, Cambridge, UK, pp. 3–12 (Carbohydrate-active enzymes family server URL: 〈http://afmb.cnrs.mrs.fr/~pedro/CAZY/db.html〉).

Daimon, T., Hamada, K., Mita, K., Okano, K., Suzuki, M.G., Kobayashi, M., Shimada, T., 2003. A *Bombyx mori* gene, *BmChi-h*, encodes a protein homologous to bacterial and baculovirus chitinases. Insect Biochem. Mol. Biol. 33, 749–759.

Daimon, T., Katsuma, S., Iwanaga, M., Kang, W., Shimada, T., 2005. The *BmChi-h* gene, a bacterial-type chitinase gene of *Bombyx mori*, encodes a functional exochitinase that plays a role in the chitin degradation during the molting process. Insect Biochem. Mol. Biol. 35, 1112–1123.

Davies, G., Henrissat, B., 1995. Structures and mechanisms of glycosyl hydrolases. Structure 3, 853–859.

de la Vega, H., Specht, C.A., Liu, Y., Robbins, P.W., 1998. Chitinases are a multi-gene family in *Aedes*, *Anopheles* and *Drosophila*. Insect Mol. Biol. 7, 233–239.

Feix, M., Gloggler, S., Londershausen, M., Weidemann, W., Spindler, K.D., Spindler-Barth, M., 2000. A cDNA encoding a chitinase from the epithelial cell line of *Chironomus tentans* (Insecta, Diptera) and its functional expression. Arch. Insect Biochem. Physiol. 45, 24–36.

Fusetti, F., von Moeller, H., Houston, D., Rozeboom, H.J., Dijkstra, B.W., Boot, R.G., Aerts, J.M., van Aalten, D.M., 2002. Structure of human chitotriosidase. Implications for specific inhibitor design and function of mammalian chitinase-like lectins. J. Biol. Chem. 277, 25537–25544.

Genta, F.A., Blanes, L., Cristofoletti, P.T., do Lago, C.L., Terra, W.R., Ferreira, C., 2006. Purification, characterization and molecular cloning of the major chitinase from *Tenebrio molitor* larval midgut. Insect Biochem. Mol. Biol. 36, 789–800.

Girard, C., Jouanin, L., 1999. Molecular cloning of a gut-specific chitinase cDNA from the beetle *Phaedon cochleariae*. Insect Biochem. Mol. Biol. 29, 549–556.

Kanost, M.R., Zepp, M.K., Ladendorff, N.E., Andersson, L.A., 1994. Isolation and characterization of a hemocyte aggregation inhibitor from hemolymph of *Manduca sexta* larvae. Arch. Insect Biochem. Physiol. 27, 123–136.

Kawamura, K., Shibata, T., Saget, O., Peel, D., Bryant, P.J., 1999. A new family of growth factors produced by the fat body and active on *Drosophila* imaginal disc cells. Development 126, 211–219.

Kim, M.G., Shin, S.W., Bae, K.S., Kim, S.C., Park, H.Y., 1998. Molecular cloning of chitinase cDNAs from the silkworm, *Bombyx mori* and the fall webworm, *Hyphantria cunea*. Insect Biochem. Mol. Biol. 28, 163–171.

Kim, D.J., Baek, J.M., Uribe, P., Kenerley, C.M., Cook, D.R., 2002. Cloning and characterization of multiple glycosyl hydrolase genes from *Trichoderma virens*. Curr. Genet. 40, 374–384.

Kirkpatrick, R.B., Matico, R.E., McNulty, D.E., Strickler, J.E., Rosenberg, M., 1995. An abundantly secreted glycoprotein from *Drosophila melanogaster* is related to mammalian secretory proteins produced in rheumatoid tissues and by activated macrophages. Gene 153, 147–154.

Koga, D., Funakoshi, T., Mizuki, K., Ide, A., Kramer, K.J., Zen, K.C., Choi, H., Muthukrishnan, S., 1992. Immunoblot analysis of chitinolytic enzymes in integuments and molting fluids of the silkworm, *Bombyx mori*, and the tobacco hornworm, *Manduca sexta*. Insect Biochem. Mol. Biol. 22, 305–311.

Kramer, K.J., Koga, D., 1986. Insect chitin: physical state, synthesis, degradation and metabolic regulation. Insect Biochem. 16, 851–877.

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.S. (Eds.), Comprehensive Molecular Insect Science, vol. 4. Elsevier, New York, pp. 111–144.

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.

Krishnan, A., Nair, P.N., Jones, D., 1994. Isolation, cloning, and characterization of new chitinase stored in active form in chitin-lined venom reservoir. J. Biol. Chem. 269, 20971–20976.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinf. 5, 150–163.

Lehane, M.J., 1997. Peritrophic matrix structure and function. Annu. Rev. Entomol. 42, 525–550.

Linder, M., Teeri, T.T., 1997. The roles and function of cellulose-binding domains. J. Biotechnol. 57, 15–28.

Lorenzen, M.D., Doyungan, Z., Brown, S.J., Savard, J., Stuart, J., Shippy, T., Snow, K., Beeman, R.W., 2005. Genetic linkage maps of the red flour beetle, *Tribolium castaneum*, based on bacterial artificial chromosomes and expressed sequence tags. Genetics 170, 741–747.

Lu, Y., Zen, K.C., Muthukrishnan, S., Kramer, K.J., 2002. Site-directed mutagenesis and functional analysis of active site acidic amino acid residues D142, D144 and E146 in *Manduca sexta* (tobacco hornworm) chitinase. Insect Biochem. Mol. Biol. 32, 1369–1382.

Mikitani, K., Sugasaki, T., Shimada, T., Kobayashi, M., Gustafsson, J.A., 2000. The chitinase gene of the silkworm, *Bombyx mori*, contains a novel Tc-like transposable element. J. Biol. Chem. 275, 37725–37732.

Ramalho-Ortigao, J.M., Traub-Cseko, Y.M., 2003. Molecular characterization of Llchit1, a midgut chitinase cDNA from the leishmaniasis vector *Lutzomyia longipalpis*. Insect Biochem. Mol. Biol. 33, 279–287.

Royer, V., Fraichard, S., Bouhin, H., 2002. A novel putative insect chitinase with multiple catalytic domains: hormonal regulation during metamorphosis. Biochem. J. 366, 921–928.

Schultz, J., Milpetz, F., Bork, P., Ponting, C.P., 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. USA 95, 5857–5864.

Shen, Z., Jacobs-Lorena, M., 1997. Characterization of a novel gut-specific chitinase gene from the human malaria vector *Anopheles gambiae*. J. Biol. Chem. 272, 28895–28900.

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

Shinoda, T., Kobayashi, J., Matsui, M., Chinzei, Y., 2001. Cloning and functional expression of a chitinase cDNA from the common cutworm, *Spodoptera litura*, using a recombinant baculovirus lacking the virus-encoded chitinase gene. Insect Biochem. Mol. Biol. 31, 521–532.

Simpson, H.D., Barras, F., 1999. Functional analysis of the carbohydrate-binding domains of *Erwinia chrysanthemi Cel5* (endoglucanase Z) and an *Escherichia coli* putative chitinase. J. Bacteriol. 181, 4611–4616.

Takahashi, M., Kiuchi, M., Kamimura, M., 2002. A new chitinase-related gene, BmChiR1, is induced in the *Bombyx mori* anterior silk gland at molt and metamorphosis by ecdysteroid. Insect Biochem. Mol. Biol. 32, 147–151.

Tanaka, T., Fujiwara, S., Nishikori, S., Fukui, T., Takagi, M., Imanaka, T., 1999. A unique chitinase with dual active sites and triple substrate binding sites from the hyperthermophilic archaeon *Pyrococcus kodakaraensis KOD1*. Appl. Environ. Microbiol. 65, 5338–5344.

Terwisscha van Scheltinga, A.C., Kalk, K.H., Beintema, J.J., Dijkstra, B.W., 1994. Crystal structures of hevamine, a plant defence protein with chitinase and lysozyme activity, and its complex with an inhibitor. Structure 2, 1181–1189.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

van Aalten, D.M., Synstad, B., Brurberg, M.B., Hough, E., Riise, B.W., Eijsink, V.G., Wierenga, R.K., 2000. Structure of a two-domain chitotriosidase from *Serratia marcescens* at 1.9 Å resolution. Proc. Natl Acad. Sci. USA 97, 5842–5847.

Varela, P.F., Llera, A.S., Mariuzza, R.A., Tormo, J., 2002. Crystal structure of imaginal disc growth factor-2. A member of a new family of growth-promoting glycoproteins from *Drosophila melanogaster*. J. Biol. Chem. 277, 13229–13236.

Watanabe, T., Uchida, M., Kobori, K., Tanaka, H., 1994. Site-directed mutagenesis of the Asp-197 and Asp-202 residues in chitinase A1 of *Bacillus circulans WL-12*. Biosci. Biotechnol. Biochem. 58, 2283–2285.

Yan, J., Cheng, Q., Narashimhan, S., Li, C.B., Aksoy, S., 2002. Cloning and functional expression of a fat body-specific chitinase cDNA from the tsetse fly, *Glossina morsitans morsitans*. Insect Biochem. Mol. Biol. 32, 979–989.

You, M., Xuan, X., Tsuji, N., Kamio, T., Taylor, D., Suzuki, N., Fujisaki, K., 2003. Identification and molecular characterization of a chitinase from the hard tick *Haemaphysalis longicornis*. J. Biol. Chem. 278, 8556–8563.

Zhang, H., Huang, X., Fukamizo, T., Muthukrishnan, S., Kramer, K.J., 2002. Site-directed mutagenesis and functional analysis of an active site tryptophan of insect chitinase. Insect Biochem. Mol. Biol. 32, 1477–1488.

Zheng, Y., Zheng, S., Cheng, X., Ladd, T., Lingohr, E.J., Krell, P.J., Arif, B.M., Retnakaran, A., Feng, Q., 2002. A molt-associated chitinase cDNA from the spruce budworm, *Choristoneura fumiferana*. Insect Biochem. Mol. Biol. 32, 1813–1823.

Zhu, X., 1998. Gene cloning, expression and characterization of C-terminal truncated forms of *Manduca sexta* chitinase. Master Thesis, Department of Biochemistry, Kansas State University, Manhattan, KS.

Zhu, Q., Deng, Y., Vanka, P., Brown, S.J., Muthukrishnan, S., Kramer, K.J., 2004. Computational identification of novel chitinase-like proteins in the *Drosophila melanogaster* genome. Bioinformatics 20, 161–169.