

# Domain organization and phylogenetic analysis of proteins from the chitin deacetylase gene family of *Tribolium castaneum* and three other species of insects

Radhika Dixit<sup>a</sup>, Yasuyuki Arakane<sup>a</sup>, Charles A. Specht<sup>c</sup>, Chad Richard<sup>a</sup>, Karl J. Kramer<sup>a,b</sup>,  
Richard W. Beeman<sup>b</sup>, Subbaratnam Muthukrishnan<sup>a,\*</sup>

<sup>a</sup>Department of Biochemistry, Kansas State University, 141 Chalmers Hall, Manhattan, Kansas 66506, USA

<sup>b</sup>Grain Marketing and Production Research Center, ARS-USDA, 1515 College Avenue, Manhattan, Kansas 66502, USA

<sup>c</sup>Department of Medicine, University of Massachusetts, Worcester, MA 01605, USA

Received 31 October 2007; received in revised form 5 December 2007; accepted 6 December 2007

## Abstract

A bioinformatics investigation of four insect species with annotated genome sequences identified a family of genes encoding chitin deacetylase (CDA)-like proteins, with five to nine members depending on the species. CDAs (EC 3.5.1.41) are chitin-modifying enzymes that deacetylate the  $\beta$ -1,4-linked *N*-acetylglucosamine homopolymer. Partial deacetylation forms a heteropolysaccharide that also contains some glucosamine residues, while complete deacetylation produces the homopolymer chitosan, consisting exclusively of glucosamine. The genomes of the red flour beetle, *Tribolium castaneum*, the fruit fly, *Drosophila melanogaster*, the malaria mosquito, *Anopheles gambiae*, and the honey bee, *Apis mellifera* contain 9, 6, 5 and 5 genes, respectively, that encode proteins with a chitin deacetylase motif. The presence of alternative exons in two of the genes, *TcCDA2* and *TcCDA5*, increases the protein diversity further. Insect CDA-like proteins were classified into five orthologous groups based on phylogenetic analysis and the presence of additional motifs. Group I enzymes include CDA1 and isoforms of CDA2, each containing in addition to a polysaccharide deacetylase-like catalytic domain, a chitin-binding peritrophin-A domain (ChBD) and a low-density lipoprotein receptor class A domain (LDL<sub>A</sub>). Group II is composed of CDA3 orthologs from each insect species with the same domain organization as group I CDAs, but differing substantially in sequence. Group III includes CDA4s, which have the ChBD domain but do not have the LDL<sub>A</sub> domain. Group IV comprises CDA5s, which are the largest CDAs because of a very long intervening region separating the ChBD and catalytic domains. Among the four insect species, *Tribolium* is unique in having four CDA genes in group V, whereas the other insect genomes have either one or none. Most of the CDA-like proteins have a putative signal peptide consistent with their role in modifying extracellular chitin in both cuticle and peritrophic membrane during morphogenesis and molting.

© 2007 Elsevier Ltd. All rights reserved.

**Keywords:** Chitin; Chitin deacetylase; Insects; Phylogenetic analysis; Groups; *Tribolium*; *Anopheles*; *Drosophila*; *Apis*

## 1. Introduction

Chitin is a naturally occurring polysaccharide that is one of the main components in the cell walls of fungi and exoskeletons of insects and crustaceans. The extracellular matrix of insect exoskeleton is modified in different ways to make the cuticle either very rigid and thick or thin and flexible, or to generate specialized structures such as mandibles with sharp cutting edges. Chitin is not only a constituent of the cuticle of insects (Kramer and Muthukrishnan, 1997) but also is a part of a sac-like structure,

**Abbreviations:** CDA, chitin deacetylase; CE-4, carbohydrate esterase family 4; ChBD, chitin-binding peritrophin-A domain; EST, expressed sequence tags; LDL<sub>A</sub>, low-density lipoprotein receptor class A; ORF, open reading frame; PCR, polymerase chain reaction; PM, peritrophic membrane; RT, reverse transcriptase; SpPGDA, *Streptococcus pneumoniae* peptidoglycan deacetylase; Tc, *Tribolium castaneum*.

\*Corresponding author. Tel.: +1 785 532 6939; fax: +1 785 532 7278.

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

the peritrophic membrane (PM), which is secreted by endodermal midgut epithelial cells. Other insect structures that are ectodermal in origin such as the highly branched tracheal network as well as the foregut and hindgut also are lined with a chitinous cuticle. Without chitin, the underlying cells lose their characteristic shape and assume abnormal structures and dimensions (Moussian et al., 2005; Arakane et al., 2005). Therefore, chitin plays a critical role in modeling tissues of various shapes, sizes and mechanical properties.

Chitin in the procuticle is multi-layered, extensively hydrogen-bonded to adjacent chains, and embedded in a proteinaceous matrix that is cross-linked by oxidized catecholic precursors (Anderson, 2005). Chitin deacetylases (CDAs) are metalloproteins that belong to a family of extracellular chitin-modifying enzymes, which deacetylate chitin, the  $\beta$ -(1,4)-linked *N*-acetylglucosamine homopolymer, to form chitosan, a polymer of  $\beta$ -1,4-linked D-glucosamine residues. This modification may contribute to an affinity for a variety of proteins distinct from those that bind specifically to chitin. Chitosan has been described in nature as a component of the fungal cell wall. The amino group of chitosan is positively charged at neutral to acidic pHs and this unique polymer has gained importance because of its commercial use in the food industry, for water treatment and in wine clarification (Tokuyasu, 1999). Chitosan also has medical applications and is used as a wound dressing that helps to rapidly clot blood. It is also used as an artificial skin for burn patients (Pusateri et al., 2003). Thus, the study of chitin deacetylases becomes especially important because of the biomedical uses and various other applications of chitosan.

CDAs have been isolated and characterized from various bacterial and fungal species and belong to the carbohydrate esterase family 4 (CE-4) of the Carbohydrate Active Enzymes (CAZY) database ([www.cazy.org](http://www.cazy.org)). CE-4 esterases deacetylate different carbohydrate substrates including chitin, acetyl xylan and bacterial peptidoglycan. Members of this family, which also includes the chitooligosaccharide deacetylases and NodB, a nodulation protein from *Rhizobium* (John et al., 1993), have a similar catalytic domain. The structures of the catalytic domain of two bacterial peptidoglycan deacetylases (Blair and Van Aalten, 2004; Blair et al., 2005) and a fungal chitin deacetylase (Blair et al., 2006) have recently been described. This has enabled the prediction of protein folding of those CE-4 esterases, for which only the protein sequences are known (Blair et al., 2006). The biological function of microbial CDAs is most likely to evade lysozyme or chitinase digestion, or for the production of oligosaccharides with regulatory or signaling functions (Vollmer and Tomasz, 2000).

CDAs have not been studied extensively in insects. Guo et al. (2005) were the first to isolate a cDNA encoding a chitin deacetylase-like protein from a midgut cDNA expression library of the cabbage looper, *Trichoplusia ni*. This protein was associated with the PM and had strong

chitin-binding activity even though it lacked a chitin-binding domain motif. Attempts to demonstrate chitin deacetylase activity for this protein were unsuccessful. Luschnig et al. (2006) and Wang et al. (2006) characterized two *Drosophila* genes required for normal tracheal tube morphogenesis by the disruptive effects of transposon insertions into these genes, which resulted in anomalous tracheal tube dimensions and morphology during embryogenesis. These genes denoted as *serpentine* (DmCDA1, CG 32209) and *vermiform* (DmCDA2, CG8756) encoded proteins with a chitin-binding domain (ChBD), a low-density lipoprotein receptor class A domain (LDLa) and a CDA-like catalytic domain. These researchers also reported a third gene encoding a CDA-like protein (*ChLD3*, DmCDA3) that was expressed in epidermis during late stage embryogenesis, but this gene has not been characterized by mutational studies in detail. To our knowledge, an exhaustive search of insect genomes for proteins with CDA domains has not been carried out for any insect species. The recent completion of the genome sequence of the coleopteran *Tribolium castaneum* provided us with an opportunity to undertake such a study together with a comparative analysis of CDA families in three other insect species including two dipteran species, *D. melanogaster* and *Anopheles gambiae*, and a hymenopteran species, *Apis mellifera*. Our results indicate that the number of CDA genes varies from species to species with *T. castaneum* having the greatest number including some recently duplicated genes found clustered in its genome.

## 2. Materials and methods

### 2.1. Insect cultures

The GA-1 strain (Haliscak and Beeman, 1983) of *T. castaneum* was used in all experiments. Insects were reared at 30 °C under standard conditions as described previously (Beeman and Stuart, 1990).

### 2.2. Synthesis of first-strand cDNA

The RNeasy Mini kit (Qiagen, Valencia, CA) was used to isolate total RNA from *Tribolium* tissue samples or whole insects according to the manufacturer's instructions. The total RNA samples were treated with RNase-free DNase I (Ambion, Austin, TX, 2 U/ $\mu$ 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 by following instructions from the manufacturer. Oligo-(dT)<sub>20</sub> was used as a primer for reverse transcription and 2.5–5  $\mu$ g of total RNA was used as template.

### 2.3. Cloning of cDNAs

PCR amplification of chitin deacetylase cDNAs for all nine *Tribolium* CDA genes was carried out using

gene-specific primers which were designed from the available expressed sequence tags (ESTs) or from open reading frames (ORFs) in GLEAN predictions from the *Tribolium* shotgun genome sequences (BeetleBase website: <http://www.bioinformatics.ksu.edu/BeetleBase>). We attempted to obtain near full-length cDNAs covering the ORFs of each CDA using cDNA templates and pairs of gene-specific primers designed in the 5'- and 3'-UTR regions of each CDA gene. cDNAs prepared from insects at different developmental stages were used as templates to maximize the probability of amplifying the target cDNAs.

#### 2.4. RNA ligase-mediated rapid amplification of cDNA ends

To obtain full-length cDNAs of *TcCDA5* and *TcCDA3*, the GeneRacer kit (Invitrogen, Carlsbad, CA) was used as per the manufacturer's protocol. In this procedure, an RNA oligonucleotide of known sequence is ligated to the 5'-ends of decapped mRNA, which is then used in RT-PCR. The 5'-end of mRNA is ligated to the GeneRacer RNA oligomer (5'-CGA CUG GAG CAC GAG GAC ACU GAC AUG GAC UGA AGG AGU AGA AA-3') with a T4 RNA ligase. Reverse transcription is then performed, using the 5'-oligomer-ligated mRNA as template, oligo(dT) as primer and superscript III reverse transcriptase (Invitrogen, Carlsbad, CA).

A touchdown PCR amplification of the resulting reverse transcription was done with TaKaRa ExTaq<sup>TM</sup> DNA polymerase. The first PCR was carried out with the GeneRacer 5'-primer (5'-CGA CTG GAG CAC GAG GAC ACT GA-3') and reverse primer for *TcCDA3* (5'-TGA TGG TTC CAA ATT ACC AGG-3') or *TcCDA5* (5'-CAG GAA AGT CGG GGT AAT C-3'). A second round of nested PCR amplification was done as above using the GeneRacer 5'-nested primer (5'-GGA CAC TGA CAT GGA CTG AAG GAG TA-3') along with a reverse primer for *TcCDA3* (5'-ATG ACG TCG AAT AGA AGC-3') and a reverse primer for *TcCDA5* (5'-CTT GTT GGT TGT GCT GAA G-3').

#### 2.5. DNA sequencing

DNA sequencing was conducted at the DNA sequencing facility at Kansas State University using an automated sequencer (ABI Prism 3700).

#### 2.6. DNA and protein sequence analyses

After cloning full-length or nearly full-length cDNAs of *TcCDAs*, exon–intron organization of each gene was investigated. Protein sequence analysis tools used in this study included those for ORF, MW and pI predictions at the ExPASy Proteomics website (<http://us.expasy.org/>) (Gasteiger et al., 2003). Domains in the protein sequences were identified in BlastP searches at NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Signal peptide prediction was conducted using the SignalP 3.0 server ([www.cbs.dtu.dk/services/](http://www.cbs.dtu.dk/services/)

SignalP/) (Bendtsen et al., 2004). Sequence alignments of multiple proteins were carried out using the ClustalW software from PBIL Expasy tool ([http://npsa-pbil.ibcp.fr/cgi-bin/align\\_clustalw.pl](http://npsa-pbil.ibcp.fr/cgi-bin/align_clustalw.pl)) (Combet, 2000).

#### 2.7. Identification of chitin deacetylases from *D. melanogaster*, *A. gambiae* and *A. mellifera* genome databases

CDA's from other insect species were identified from the "blast" search results of NCBI and four other databases containing completely sequenced genomes: Flybase (<http://flybase.bio.indiana.edu>); BeetleBase website: <http://www.bioinformatics.ksu.edu/BeetleBase>; *A. gambiae* ([www.anobase.org](http://www.anobase.org)) and the *A. mellifera* genome database ([http://racerx00.tamu.edu/bee\\_resources.html](http://racerx00.tamu.edu/bee_resources.html)).

#### 2.8. Phylogenetic analysis of CDAs from four insect species

ClustalW software ([www.ebi.ac.uk/clustalw/](http://www.ebi.ac.uk/clustalw/)) was used to perform multiple sequence alignment prior to phylogenetic analysis. MEGA 3.0 (Kumar et al., 2004) was used to construct the consensus phylogenetic tree using maximum parsimony method and a cut-off value of 50% similarity. To evaluate the branch strength of the phylogenetic tree, bootstrap analysis of 5000 replications was performed. The accession numbers of the *Tribolium* CDAs are shown in Table 1.

#### 2.9. Homology modeling of putative *Tribolium* CDAs

The Geno3D modeling program (<http://geno3d-pbil.ibcp.fr>, Combet et al., 2002) was used to generate the homology models for two putative *Tribolium* CDAs, *TcCDA1* and *TcCDA9*, using a bacterial CDA, *Streptococcus pneumoniae* as the template, (pdb 2CIG; Blair et al., 2005).

### 3. Results

#### 3.1. Identification of putative *Tribolium* CDA genes

Genes for CDA-like proteins in the *Tribolium* genome database (BeetleBase website: <http://www.bioinformatics.ksu.edu/BeetleBase>) and the NCBI database ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) were identified using the *tblastn* program (Tcas V2.0 version of the *Tribolium* genome assembly). The amino acid sequences of the CDA domain of the proteins encoded by the *serpentine* (*DmCDA1*) and *vermiform* (*DmCDA2*) genes from *D. melanogaster* (Luschnig et al., 2006), as well as an *A. gambiae* CDA, were used as query sequences to identify five *Tribolium* CDAs. Using the translated sequences of these five as queries, four additional CDAs were identified in the *Tribolium* genome (*TcCDA1–9*, Table 1).

Table 1  
Properties of identified *Tribolium castaneum* chitin deacetylase cDNAs

cDNA	GenBank accession #	cDNA length (bp) <sup>a</sup>	ORF (aa)	M.W (kDa)	pI	Scaffold #	Linkage group #	Glean #	Locus
TcCDA1	ABU2522	1791	534	62	5.08	1302–1332	LG 5	14100	LOC657802
TcCDA2A	ABU25224	1831	535	66	5.31	1302–1332	LG 5	14101	LOC657879
TcCDA2B	ABU25225	1810	528	60	5.24				
TcCDA3	ABW74145	1580 <sup>b</sup>	517	59	6.07	1125–2371	LG 8	05409	LOC658524
TcCDA4	ABW74146	1628	490	56	5.80	2712–4368	LG 4	07635	LOC658721
TcCDA5A	ABW74147	3694	1231	129	8.56	3548	Unknown	06846	LOC656801
TcCDA5B	ABW74148	3694	1231						
TcCDA6	ABW74149	1246	403	45	4.74	1302–1332	LG 5	13662	LOC661969
TcCDA7	ABW74150	1152	374	43	5.54	1302–1332	LG 5	13661	
TcCDA8	ABW74151	1154	376	43	4.79	1302–1332	LG 5	14147	LOC662008
TcCDA9	ABW74152	1175	381	43	4.71	4979–4226	LG 3	03905	LOC659575

<sup>a</sup>Most of the cDNAs were isolated in our laboratory. Some cDNA sequences were obtained from EST databases.

<sup>b</sup>Full-length cDNA clone not available and the sequence was put together from two overlapping cDNA clones.

### 3.2. Cloning of putative *Tribolium* CDA cDNAs

A search of the available *Tribolium* ESTs for CDA sequences identified partial length transcripts corresponding to the 3'-ends of some, but not all, CDA genes. In particular, no ESTs for *TcCDA3* and *TcCDA7* were detected. Similarly, sequences corresponding to the 5'-regions of *TcCDA5* were not present in the EST databases. In an attempt to amplify full-length cDNAs for all nine CDAs, we designed forward primers corresponding to the regions immediately preceding the translation start codon and reverse primers corresponding to sequences in the 3'-UTR upstream of the putative polyadenylation signal. RT-PCR was carried out using RNA from different developmental stages to improve the chances of obtaining the cDNAs. The clones that contained the ORF and the 5'- and 3'-UTR sequences were deemed to be full-length or near full-length. This procedure yielded full-length clones for all *TcCDAs* except *TcCDA3* and *TcCDA5*. The full-length cDNA sequences corresponding to these latter two genes were obtained using the GeneRacer kit, as described in Section 2. The chromosomal locations, transcript lengths and GenBank accession numbers of the proteins corresponding to the *Tribolium* CDA gene family are listed in Table 1.

### 3.3. Exon–intron organization of *TcCDA* genes

A comparison of the sequences of the nine putative CDA cDNAs from *Tribolium* with the corresponding genomic sequences allowed the determination of the exon–intron organization of each gene (Fig. 1). The 5'- and 3'-untranslated regions of all genes are relatively short (less than 200 nucleotides). *TcCDA6*, *TcCDA7* and *TcCDA8* are highly similar to one another in their intron–exon organization as well as in sizes and sequences of the encoded proteins. The exon–intron assignments were in agreement with those predicted by GLEAN as well as NCBI annotations with the exceptions of *TcCDA3*, *TcCDA5* and *TcCDA7*. Detailed analysis of *TcCDA* genes

and their cDNAs indicated the presence of two alternate exons in both *TcCDA2* and *TcCDA5*. The presence of alternate exons in *TcCDA2* and *TcCDA5* cDNAs was verified by RT-PCR using primers corresponding to the exons flanking the putative alternatively spliced exons. The alternate exons in *TcCDA2* (exon 3) are not of the same length and when translated are about 59% similar in their amino acid sequence. Both encode the six cysteine residues of the chitin-binding peritrophin-A domain (ChBD). Exon 3 of *CDA2a* is 134 bp in length. The presence of an additional stretch of amino acids distinguished it from the other ChBD domain-containing proteins and made it difficult for domain prediction software programs like Pfam and CDD to identify the ChBD domain in *TcCDA2a*. On the other hand, exon 3 of *CDA2b*, which is 21 nucleotides shorter, encodes a ChBD domain that was readily identified. The alternate exons in *CDA5a* and *CDA5b* (exon 12) are of the same size (226 nucleotides).

### 3.4. Linkage group assignments of CDA genes

*TcCDA1*, *TcCDA2*, *TcCDA6*, *TcCDA7* and *TcCDA8* map to linkage group LG5 in two clusters. *TcCDA1* and *TcCDA2* are about 16 kb apart from each other and are separated from the *TcCDA6*, *TcCDA7* and *TcCDA8* cluster by about 935 kb. The distances between *TcCDA6* and *TcCDA7*, and *TcCDA7* and *TcCDA8* are about 3.6 and 3.2 kb, respectively. The other *TcCDAs* belong to different linkage groups: *TcCDA3* maps to LG8, *TcCDA4* to LG4, and *TcCDA9* to LG3. The linkage group of *TcCDA5* has not yet been determined. Table 1 shows the linkage group assignments and the location of each *TcCDA* gene on the *Tribolium* chromosomes.

### 3.5. Sequence analysis and domain organization the *TcCDA* family of proteins

The amino acid sequences of proteins encoded by each of the nine *Tribolium* CDAs including isoforms derived from

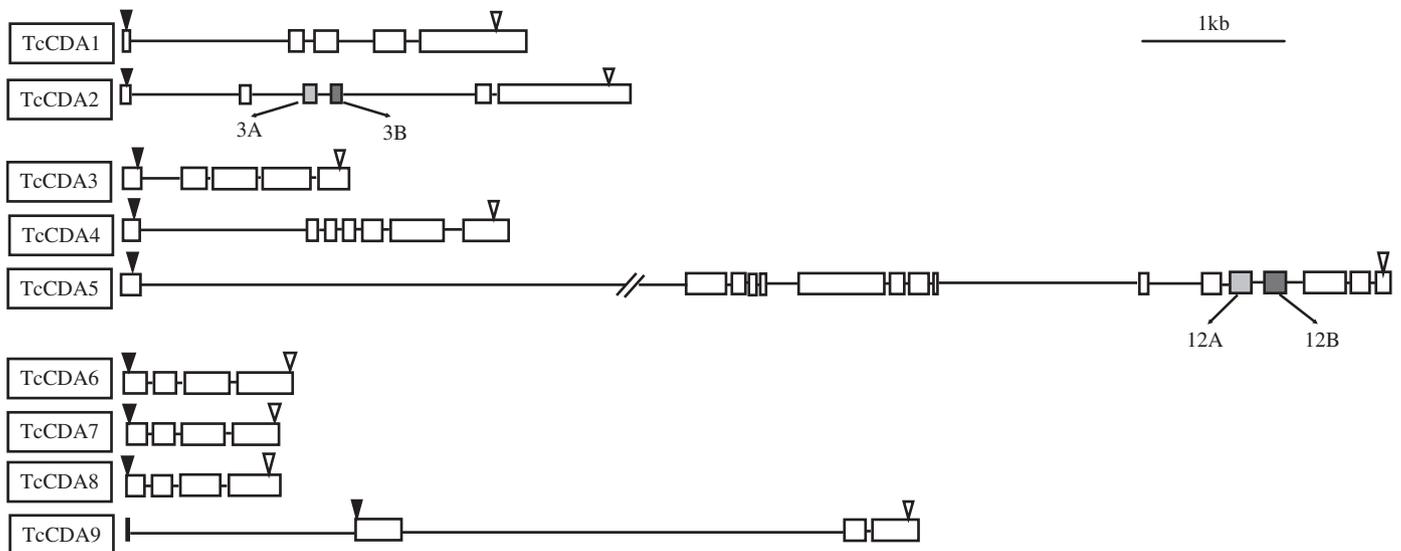


Fig. 1. Schematic diagram of the exon–intron organizations of the putative *Tribolium castaneum* CDA genes. The exon–intron organization of each CDA gene was determined by sequence comparison between genomic sequence and the longest available cDNA sequence. The alternative exons in *TcCDA2* and *TcCDA5* are shown as light gray and dark gray boxed regions. The two forward slashes indicate a 4 kb intron of *TcCDA5*. Open boxes indicate exons and lines indicate the introns. The closed arrowheads represent the start codons and open arrowheads are the stop codons.

alternatively spliced transcripts were analyzed for the presence of a signal anchor or signal peptide (Fig. 2). All TcCDAs were predicted to have a signal peptide and are probably secreted proteins. The predicted pI of the CDAs range from pH 4.71–8.56 with CDA3 having the highest pI. The predicted molecular weights of CDAs range from 43 to 129 kDa. Some of the CDAs are predicted to have a few *N*-glycosylation sites. From the alignment of the TcCDA protein sequences (Fig. 2), several cysteines were found to be conserved. Each CDA was predicted to contain disulfide bridges by the Dipro program ([scratch.proteomics.ics.uci.edu/](http://scratch.proteomics.ics.uci.edu/)). Six of the cysteines are conserved in those CDAs with a chitin-binding ChBD domain (pfam 01607) and another six cysteines are signature amino acids of the low-density lipoprotein receptor domain class A or LDLa domain (pfam 00057). All the TcCDAs have 12 cysteines, eight that appear to flank the catalytic domain (defined below) and four found at the C-terminus of each protein.

The different domain architectures of the putative CDA family of proteins from *Tribolium* generate their diversity in protein size. All have the polysaccharide deacetylase catalytic domain (pfam 01522) found in carbohydrate esterase family 4 (CE-4) of the Carbohydrate Active Enzymes (CAZY) database (Coutinho and Henrissat, 1999). Some of the CDAs (CDA6, CDA7, CDA8 and CDA9) lack the ChBD domain and the LDLa domain. CDA4 and CDA5 lack only the LDLa domain, but retain the ChBD. A schematic diagram depicting the order of the domains is shown in Fig. 4.

### 3.6. Characterization of CDA-like gene families from *Drosophila*, *Anopheles* and *Apis*

The nine *Tribolium* CDA protein sequences were used as queries to search the *Drosophila*, *Anopheles* and *Apis*

genomic databases. Only six genes encoding putative CDA-like proteins were identified in the *Drosophila* genome and five in each of the genomes of *Anopheles* and *Apis* (see Fig. 4). Homologs of *TcCDA6*, *TcCDA7* and *TcCDA8*, were not detected in the other insect species. While those of *TcCDA1–5* were found in all three species, only *Drosophila* encodes a homolog of TcCDA9. Counterparts to the alternate exons of exon 3 of *TcCDA2* and exon 12 of *TcCDA5* were found for the respective genes of the other insect species, and they are spliced at similar positions in each gene (data not shown).

*TcCDA5* and its orthologs in the other species encode the largest member of the CDA protein family. A long stretch of about 300–600 amino acids unique to these CDAs does not correspond to any well-characterized protein domain or motif in the protein databases. All chitin deacetylase-like proteins from the four insect species are predicted to have signal peptides and are probably secreted proteins. The only exception is DmCDA3 (DmChLD3), which may not be secreted (data not shown).

### 3.7. Sequence alignment of catalytic domains of CDAs

The catalytic domain sequences of putative CDA-like proteins predicted from the *Drosophila*, *Anopheles*, *Apis* and *Tribolium* genomes were aligned using the ClustalW program with previously defined deacetylase domains of insect, fungal and bacterial proteins belonging to the CE-4 family (Fig. 3). Recently, Blair et al. (2005, 2006) described several signature motifs that make up the active sites of the deacetylase domain of CE-4 family proteins. Motif 1 (TFDD) includes one aspartate that coordinates with zinc or cobalt and an adjacent second aspartate that interacts with the acetate released from the substrate.

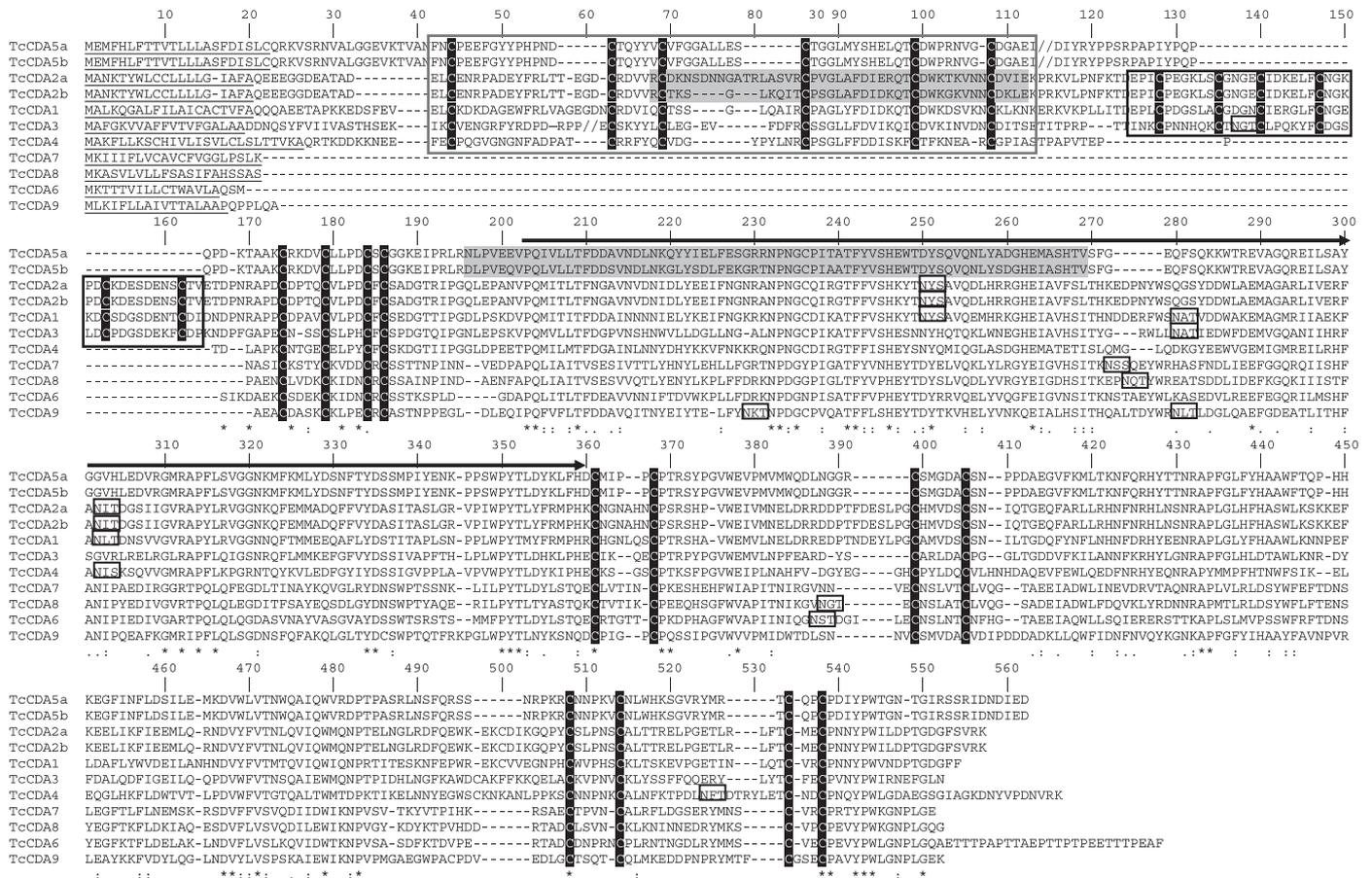


Fig. 2. Amino acid sequence alignment of CDAs from *Tribolium castaneum*. The multiple sequence alignment of CDA family of protein was carried out using the ClustalW software from PBIL ExPASy tool. ([http://npsa-pbil.ibcp.fr/cgi-bin/align\\_clustalw.pl](http://npsa-pbil.ibcp.fr/cgi-bin/align_clustalw.pl)) (Combet, 2000). The underlined region indicates predicted signal peptide. The chitin-binding domain is represented as a gray box, the LDL domain is boxed in dark black and the long horizontal arrow region is the catalytic domain. The alternatively spliced regions of *CDA2a* and *CDA2b* and *CDA5a* and *CDA5b* are highlighted in gray. Shaded black regions are the cysteine residues; three residue black-boxed regions are the predicted *N*-glycosylation sites. The conserved regions are indicated by asterisks below the aligned residues and two forward slashes (//) indicate additional amino acids (seven amino acids for TeCDA3 and 651 amino acids for TeCDA5a and TeCDA5b).

Motif 2 (H[S/T]xxHP) contains two histidines that bind to a metal and a serine or threonine that forms a hydrogen bond with the second histidine in motif 2 stabilizing the loop. Motifs 3 and 4 line up on opposite sides of the active site groove. Motif 5 contains a leucine and a tryptophan that help to form the hydrophobic pocket where the methyl group of the acetate product interacts.

Alignment of multiple CDAs from insects and deacetylases/esterases belonging to the CE-4 family from other species reveals that the insect CDA domains have all of the five signature motifs except for some substitutions of critical amino acids (Fig. 3). The TFDD motif 1 is replaced by TFNG in all isoforms of CDA2 from all of the four insect species. In TeCDA7 and TeCDA8, this motif is replaced by TVSE. These substitutions are predicted to render the protein enzymatically inactive because the first aspartate in the TFDD motif has been implicated in the reaction mechanism to abstract a proton from the Zn<sup>2+</sup>-bound water molecule to create a hydroxyl nucleophile (Blair et al., 2005). In the second motif, H(S/T)xxHP, the first histidine critical for binding zinc is replaced by either

phenylalanine or glutamate in CDA2 and CDA4, respectively, in all species studied here (Fig. 3). TeCDA6 has an asparagine in this position. These proteins are also predicted to be enzymatically inactive. In the third motif, RxPY, CDA4s as well as CDAs 6, 7 and 8 have Y to F or Y to Q substitutions with the exception of DmCDA4. This Y residue is implicated in H-bonding with acetate and mutation of this residue to alanine has been shown to inactivate the *S. pneumoniae* peptidoglycan deacetylase enzyme (Blair et al., 2005).

### 3.8. Phylogenetic analysis of chitin deacetylases

To investigate the evolutionary relationship among the putative *T. castaneum* CDAs, phylogenetic analyses were carried out with the orthologs of other insect species including representatives belonging to coleopteran, dipteran and hymenopteran lineages. A minimum evolution method was used to construct the phylogenetic tree shown in Fig. 4 using the catalytic domain sequences of Fig. 3. Trees based on the Unweighted Pair Group Method with

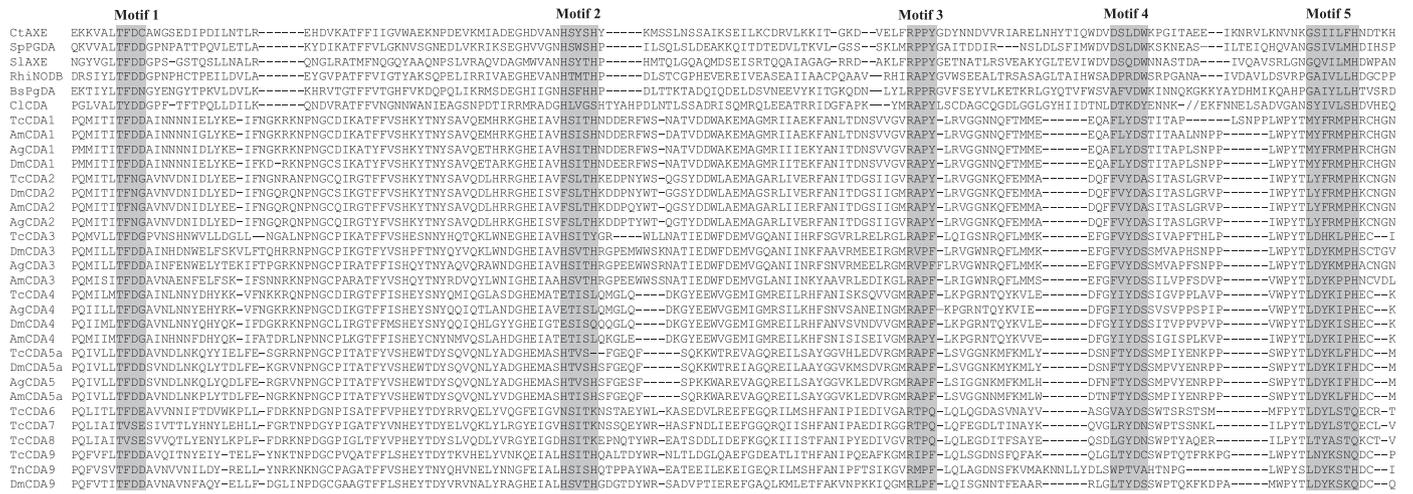


Fig. 3. Amino alignment of catalytic domains of CE4 family enzymes from insects, fungi and bacteria. Partial amino acid sequences from *Tc* (*Tribolium castaneum*), *Dm* (*Drosophila melanogaster*), *Ag* (*Anopheles gambiae*), *Am* (*Apis mellifera*), CtAXE (acetyl xylan esterase of *Clostridium thermocellum*), SpPgdA—peptidoglycan deacetylase of *Streptococcus pneumoniae*, Sl AXE—acetyl xylan esterase of *Streptococcus lividans*, NodB—nodulation B protein from *Rhizobium meliloti*, BsPgDA—peptidoglycan deacetylase of *Bacillus subtilis*, ClCDA—chitin deacetylase of *Colletotrichum lindemuthianum* were aligned using the ClustalW software from PBL ExPasy tool ([http://npsa-pbil.ibcp.fr/cgi-bin/align\\_clustalw.pl](http://npsa-pbil.ibcp.fr/cgi-bin/align_clustalw.pl)) (Combet, 2000). The gray shaded regions are the alternatively spliced isoforms of CDA5 in all the insect species. Boxed regions are the motifs 1, 2, 3, 4 and 5 required for catalytic activity. Motif 1 (TFDD) includes one aspartate which coordinates zinc (cobalt in some cases) and an adjacent second aspartate which binds the acetate released from the substrate. Motif 2 (H[S/T]xxHP) contains the two histidines that bind to a metal and a serine or threonine that forms a hydrogen bond with the second histidine in motif 2. The third motif, RXPY/F, is part of the active site groove and has multiple roles including binding of the product acetate, binding zinc and coordinating the catalytic aspartate residue. In the fourth motif, DxxDW, tryptophan is the most critical residue and forms the other side of the active site groove. Motif 5 which includes a leucine and a histidine residue forms a hydrophobic pocket that binds the acetate methyl group and a histidine that forms a hydrogen bond with the product acetate.

Arithmetic Mean (UPGMA), maximum parsimony as well as Neighbor Joining methods were also constructed (data not shown). All three of the topologies were consistent with one another and grouped the proteins into five major classes, CDA groups I through V. Each group/branch corresponds to a unique gene or gene cluster.

Group I includes CDAs 1 and 2, all with a ChBD, an LDLA domain and a catalytic domain. These genes have been shown to be expressed in tracheal lining cells as well as in epidermal cells in *Drosophila* (Luschnig et al., 2006; Wang et al., 2006). The CDA2 pre-mRNAs undergo alternative splicing and/or exon skipping to yield four transcripts that code for four CDA isoforms in *Drosophila* and *Anopheles*, whereas *Apis* and *Tribolium* have only two isoforms derived from a single gene as a result of alternative splicing.

Group II consists of only one CDA from each species, CDA3. This protein also has a single copy of each of the three domains. CDA3 differs substantially from CDA1 and CDA2 isoforms in overall amino acid sequence, while the catalytic domains are less divergent. CDA3 is 38% identical in the complete amino acid sequence to CDA1 and CDA2, whereas CDA1 is 58% identical to CDA2. Group III has a single representative in each of the four insect species and includes CDA4, which has a chitin-binding domain and a catalytic domain but lacks an LDLA domain. Group IV consists of orthologs of the largest CDA protein (CDA5). These have a chitin-binding domain encoded by exon 2 followed by a glutamine-rich region that

bears no significant similarity to any characterized protein motif in the databases. The CDA catalytic domain is positioned towards the C-terminus. At least three of the insect species have more than one isoform of CDA5 in this group. The presence of additional isoforms is due to alternative splicing and exon skipping during the processing of transcripts for these genes. *Tribolium* and *Apis* have two isoforms similar to DmCDA5a and DmCDA5c, and to our knowledge, *A. gambiae* has only one isoform of the protein. This may be due to a partial annotation of genes in the *Anopheles* database.

Group V is made up of a divergent group of proteins containing only a catalytic domain. All of these proteins lack the N-terminal ChBD and the LDLA domain. Transcripts for the genes of this group are found in the larval gut but not in the carcass suggesting they are expressed predominantly in the gut (data not shown). Two subgroups were observed within group V. One subgroup, which included TcCDA9, has an ortholog only in *Drosophila*, whereas the other subgroup consisted of paralogs from *Tribolium* only and not from other insects of dipteran and hymenopteran lineages.

3.9. Homology modeling

Blair et al. (2005) described the crystal structure of a peptidoglycan deacetylase from *S. pneumoniae* (SpPgdA) and compared this to the crystal structure of another member of the CE-4 family, a *B. subtilis* peptidoglycan

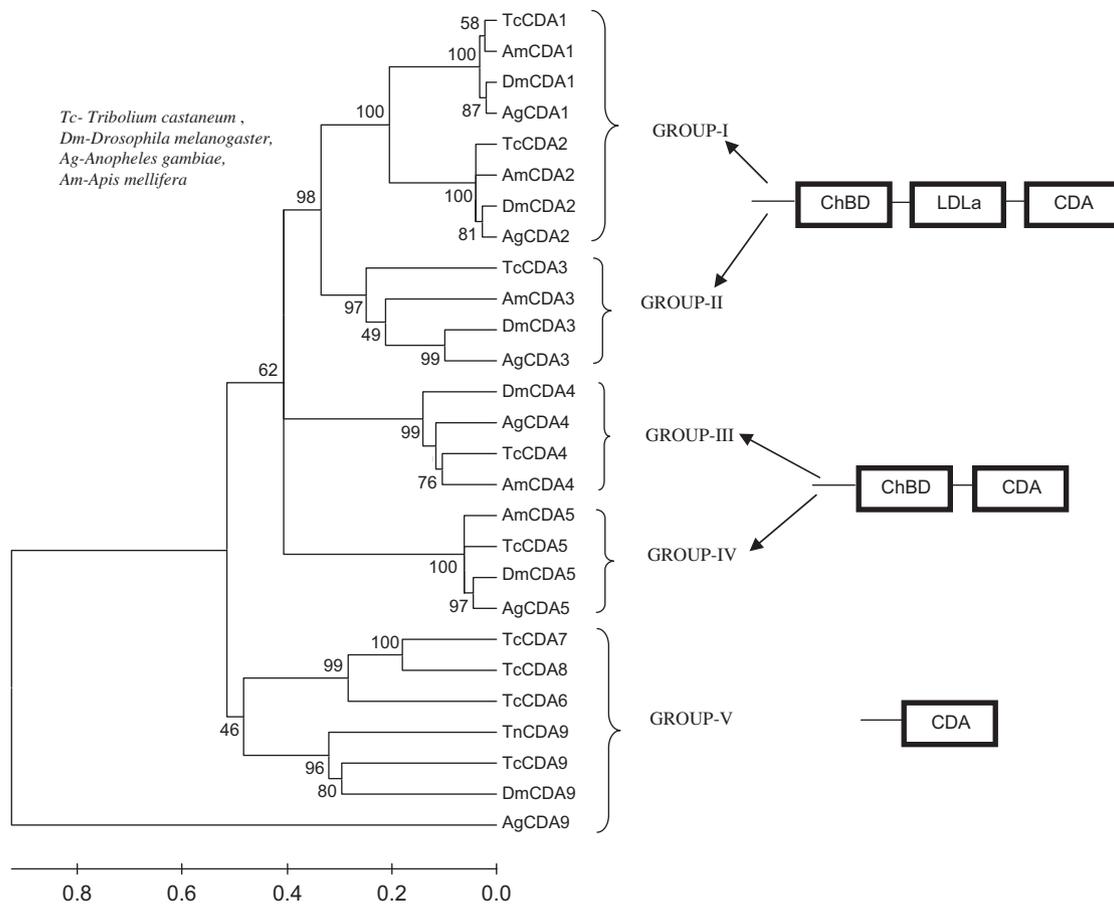


Fig. 4. Phylogenetic tree of putative CDAs from *Tribolium*, *Drosophila*, *Apis* and *Anopheles*. A consensus phylogenetic tree was constructed using the software MEGA 3.0 (Kumar et al., 2004). A bootstrap analysis of 5000 replications was carried out on the trees inferred from the minimum evolution method. Bootstrap values are shown in the cladogram if the values are higher than 50%. Four different insect species belonging to different insect orders were used to construct the phylogenetic tree. *Tc*—*Tribolium castaneum* (coleopteran), *Dm*—*Drosophila melanogaster* and *Ag*—*Anopheles gambiae*, which are dipterans, and *Am*—*Apis mellifera*, a hymenopteran. The sequences were deduced from their corresponding cDNA sequences and the catalytic domain regions were used for constructing the phylogenetic tree. *DmChLD3* is referred to as *DmCDA3* and the *Drosophila serpentine* gene is referred to as *DmCDA1* and *vermiform* gene as *DmCDA2*.

deacetylase (Blair and Van Aalten, 2004). Both enzymes assume a tertiary structure consisting of a  $(\beta/\alpha)_8$  fold, similar to the TIM barrel structure. The homology models of TcCDA1 and TcCDA9, which have the critical residues of motifs 1 and 2, are shown in Fig. 5. Both proteins share about 40% amino acid sequence identity in the catalytic domain region to SpPGDA, which was used as the template. The side chains of critical catalytic residues (an aspartate and two histidine residues) are indicated. They appear to be close to each other in a configuration similar to that observed in the crystal structure of the SpPGDA template. Thus, the insect CDAs not only exhibit sequence similarity to CDAs from other species but also appear to have similar secondary and tertiary structures.

#### 4. Discussion

CDAs are members of the family of carbohydrate esterases (CAZY family CE 4) and several of these enzymes have been characterized, which were purified from fungal

and bacterial pathogens. They apparently function as protective agents against host defense enzymes such as lysozymes and chitinases that target microbial cell wall components. The CE4 family includes several members with varying substrate specificities such as acetylxyylan esterases that act on the *O*-acetyl groups of xylan, as well as chitin deacetylases and peptidoglycan deacetylases that remove acetyl groups from *N*-acetylglucosamine or *N*-acetylmuramic acid residues in chitin and peptidoglycans, respectively.

Previously, CDAs were thought to be restricted to fungi and bacteria until Aruchami et al. (1986) reported their presence in arthropods. Recently, Guo et al. (2005) reported the isolation of a *T. ni* cDNA for a peritrophic membrane-associated protein, TnPm-P42, that contains a CDA-like domain. Recombinant TnPm-P42 had chitin-binding activity, but CDA activity was not detected. These researchers also reported ESTs for proteins closely related to TnPm-P42 in a number of other insects and some acari species. Three *Drosophila* genes encoding proteins with

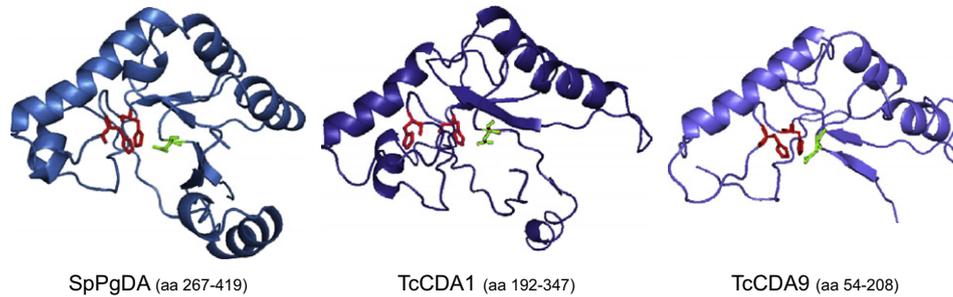


Fig. 5. Homology model of the catalytic domain of *TcCDA1* and *TcCDA9*. The Geno3D modeling program (<http://geno3d-pbil.ibcp.fr>; Combet et al., 2002) was used to generate the homology model of *TcCDA1* and *TcCDA9* using the crystal structure of bacterial *Streptococcus pneumoniae* peptidoglycan deacetylase as a template (pdb-2c1g). The catalytic domain (amino acid # 267–419) of SpPGDA was also modeled. The catalytically critical residues, aspartate of motif-1 (MT-1) is in green and two histidines of motif-2 (MT-2) are in red for each model.

CDA-like domains were identified by Luschnig et al. (2006) and Wang et al. (2006). Two of these CDAs were, in fact, shown to be associated with chitin in tracheal tubules and to be required for tracheal tube morphogenesis. The results reported here indicate the presence of somewhat larger families of five to nine genes in *Drosophila*, *Anopheles* and *Tribolium*, which encode proteins with CDA-like domains. The total number of proteins encoded by these families of genes is even larger because two of the genes, *CDA2* and *CDA5*, give rise to multiple transcripts as a result of alternative splicing. While all proteins of this family share the common CDA-like domain, they differ in amino acid sequence, size and presence or absence of additional domains including ChBD and LDLa domains. They also differ in the pattern of expression during development and in the tissue specificity of expression as measured by RT-PCR on RNA templates (unpublished data), suggesting different biological functions.

We have also been able to amplify full-length or near full-length cDNA clones for all CDAs from *Tribolium* using a variety of approaches and the sequences of these clones have allowed us to accurately determine the exon–intron splice points. While most of the ORFs predicted by GLEAN/NCBI annotations have been confirmed by the cDNA clones reported here, we find substantial differences from the GLEAN/NCBI annotations in the amino acid sequences predicted for *TcCDA3*, *TcCDA5* and *TcCDA7*. These studies have also allowed us to confirm the presence of alternatively spliced transcripts for *TcCDA2* and *TcCDA5* as reported in *Drosophila*, although we have been unable to detect by RT-PCR, *TcCDA5* transcripts that correspond to *DmCDA5b* transcripts predicted to arise as a result of the use of alternate promoters or exon skipping. A more detailed analysis of genes from other insect species is required to establish whether *Drosophila* is unique in having these additional transcripts.

Phylogenetically, the insect CDAs can be divided into five groups. Group I includes genes encoding the two closely related proteins, *CDA1* that has a single isoform and *CDA2* that has multiple isoforms in all four insect species. Groups II, III and IV have one representative gene from each insect species except that *CDA5*, which belongs

to group IV, has multiple isoforms arising from alternative splicing. Group V has four genes, but three of these are found only in *Tribolium* as a tightly linked cluster with intergenic regions of only 3.2–3.4 kb, indicating that these genes represent recent gene duplication events unique to *Tribolium*. The gut-specific CDA-like protein from *T. ni*, TnPm-P42, belongs to this group (Guo et al., 2005). Because the two dipteran and one hymenopteran species analyzed here do not have these additional genes, this duplication event involving group V genes must have happened after the coleopteran lineage had separated from other arthropod lineages. Consistent with this interpretation, the amino acid sequences of TcCDAs 6, 7 and 8 share a high degree of similarity (about 65% to each other) and a much lower degree of similarity to other CDAs including *TcCDA9* (50% similarity).

The presence of representatives of all five groups of CDAs and the conservation of very similar patterns of alternative splicing in *CDA2* and *CDA5* in different orders of insects suggest that these specialized CDA isoforms had evolved long before the divergence of insect orders. The biological necessity for having multiple CDA-like proteins in insects is unclear. However, the differences in the tissue specificity and the developmental patterns of expression of some CDAs offer clues concerning their function. The recently reported CDA-like protein, TnPm-P42, found to be associated with the PM falls into group V (Guo et al., 2005), whose members lack ChBD and LDLa domains. This gene is expressed only in gut tissue and only during feeding stages, suggesting that TnPm-P42 may be involved in modification of chitin in the PM or as a defensive agent against microbes in the gut contents. On the other hand, *Drosophila* *CDA1* and *CDA2* are expressed in tracheal and epidermal cells (Luschnig et al., 2006; Wang et al., 2006). Our preliminary data from RT-PCR analysis of RNA from different tissues indicate that CDAs 6, 7, 8, and 9 are expressed in the gut but not in the carcass (data not shown). The presence of a representative member of each subgroup of CDAs in all four insect species studied here (Figs. 3 and 4) suggests that each of these CDAs serves an indispensable and non-redundant function.

The studies of Luschnig et al. (2006) and Wang et al. (2006) have suggested that CDA1 and CDA2 associate with chitin. The corresponding proteins are apically secreted by tracheal cells and are co-localized with chitin in the tracheal lumen. Loss of function mutants of these genes exhibit long and tortuous distal tracheal tubes, suggesting that a proper ratio of chitin to chitosan is critical for normal tracheal morphology. These mutations also appear to affect the epidermal cuticle because in these mutants, the developing embryos have an abnormal shape with the epidermal cuticle devoid of the laminar structure of the chitin matrix and are bloated, suggesting loss of rigidity or integrity of the exoskeleton.

CDAs differ among themselves in their domain organization. While CDAs 6, 7, 8 and 9 have only the CDA domain, the other CDAs have a variable assortment of additional domains including a ChBD, LDLa receptor domain and, in the case of CDA5, a unique 300–600 amino acid-long region rich in glutamine residues. It is likely that these domains help to bind these proteins to chitin or to some other extracellular ligand. The presence of multiple copies LDLa domain in LDL receptors is thought to play a central role in mammalian cholesterol metabolism. The low-density lipoprotein receptor class-A protein, which has multiple LDLa domains, is said to bind and transport LDL into cells by endocytosis. In contrast, insect CDAs have only one LDLa domain. The crystal structure of the LDL receptor has indicated that this domain participates in  $\text{Ca}^{2+}$  coordination (Fass et al., 1997). However, the nature of the ligand for the LDLa domain of CDAs is unknown. It is also unclear why *Tribolium* has four different CDAs belonging to group V, whereas other species have at most one representative. These proteins may have a role in insect immunity or may alleviate the inhibitory effect of chitooligosaccharides on the activity of gut chitinases needed for molting. The nature of the extra region at the N-terminus of CDA5 is currently unknown, but the corresponding regions in CDA5s from other insect species are similar in amino acid sequence, suggestive of a conserved function.

Blair and Van Aalten (2004), Blair et al. (2005, 2006) have determined the crystal structures of two bacterial peptidoglycan deacetylases and a chitin deacetylase belonging to the CE4 family. Their studies have shown that the catalytic domain contains a substrate-binding cleft lined by several conserved motifs. One motif, TFDD, contains the aspartate (first Asp), which acts as the base that abstracts a proton from a tightly bound water molecule predicted to be involved in catalysis. The second motif, H(S/T)xxHP, has the two His residues that bind to the zinc atom. The third motif, RXPY/F, has multiple roles including binding of the product acetate, binding zinc, and coordinating with the catalytic Asp residue. Our homology models for TcCDA1 and TcCDA9 based on the crystal structure of *Streptococcus* peptidoglycan deacetylase indicate that the zinc-binding Asp and two His residues are present in their substrate-binding pockets. It is interesting

to note that not all of the TcCDAs contain the catalytically critical first aspartate residue in the sequence TFDD, which is required for acetate binding and the second aspartate is needed to coordinate the zinc atom (Fig. 3). In the TcCDA2 isoforms from all of the insect species analyzed here (Fig. 3), the motif TFDD is replaced by TFNG. The replacement of the first aspartate by asparagine has been shown to inactivate the *Streptococcus* peptidoglycan deacetylase (Blair et al., 2005), but in the case of peptidoglycan deacetylase of *B. subtilis*, which is enzymatically active, the second aspartate is replaced by an asparagine residue. It was speculated by Blair et al. (2005) that the *O*-lactoyl group of the substrate itself could interact with zinc, in place of the second aspartate in the TFDD motif. Whether the TcCDAs lacking both of the Asp residues in the TFDD motif are enzymatically active awaits purification and assay of these proteins. It is conceivable that several insect CDA-like proteins are devoid of enzymatic activity and may simply bind to chitin and thus influence the mechanical or permeability properties of chitin-containing structures such as the cuticle or the peritrophic membrane.

The significance of the large number of conserved cysteine residues that flank the putative catalytic domain of insect CDA-like proteins, as shown in Fig. 2 for *Tribolium*, remains to be determined. The prediction that they form disulfide bridges suggests the possibility that a unique folding of the catalytic domain brings other amino acid residues into play for catalysis.

Even though enzymes belonging to the CE4 family share the CDA domain, they differ in their substrate preferences. For example, the peptidoglycan deacetylase from *B. subtilis* deacetylates *N*-acetylmuramic acid residues and not *N*-acetylglucosamine residues in the substrate (Blair and Van Aalten, 2004; Blair et al., 2005). Some CDAs have an exo-type of action, whereas others remove acetyl groups with a random endo-type of attack (Tsigos et al., 2000; Tokuyasu et al., 2000). Some enzymes fully deacetylate the substrate, while others leave a partially deacetylated product. The extent of deacetylation of oligosaccharide substrates also depends on the length of the substrate. While chitotriose and chitotetraose are fully deacetylated by CDA from *Colletotrichum lindemuthianum*, chitobiose is deacetylated only at the non-reducing end. Hekmat et al. (2003) have determined that four subsites (–2, –1, 0 and +1) in this enzyme bind to four GlcNAc residues in the substrate with different affinities, which may influence which residue is deacetylated. While the first deacetylation is relatively fast, subsequent deacetylations are quite slow, which results in a partially deacetylated product perhaps as a result of non-productive binding and inhibition of enzyme activity. It will be interesting to study whether TcCDAs differ in their extent of deacetylation and susceptibility to inhibition by partially deacetylated products. Alternatively, CDAs may function in ameliorating the inhibitory effects of chitooligosaccharides on the activity of chitinases (Arakane et al., 2003).

Fungal CDAs have been proposed to play a defensive function against plant chitinases and chitosanases with an endo-type of action (Hekmat et al., 2003). Insect CDAs may have a role in defense in addition to a role in chitin modification. Chitin deacetylation could determine the type of proteins associated with polysaccharide in the insect cuticle, PM or tracheal lining depending on the extent of deacetylation and thus influence the properties of the cuticle, PM or tracheae. CDAs with different substrate specificities working together may fully deacetylate chitin or peptidoglycans of invading microbes, making them susceptible to other host defenses in addition to commonly deployed lysozyme and chitinases.

The enzymatic deacetylation of chitin has received substantial attention because of the commercial use of deacetylated chitin (chitosan) in water treatment, in the food industry as a biopesticide, and as a substance that boosts the ability of plants to defend against fungal infections ([www.epa.gov/pesticides/biopesticides/factsheet128930](http://www.epa.gov/pesticides/biopesticides/factsheet128930)). There are other CDA applications in the medical industry such as in the fabrication of artificial skin and as a blood clotting enhancing agent. Also, Kean et al. (2005) discovered that trimethylated chitosans are effective as non-viral gene delivery vectors that increase the transfection efficiency in cells. From a developmental point of view, the studies of Luschnig et al. (2006) and Wang et al. (2006) have opened up a new field of investigation on the roles of CDAs in modifying extracellular matrix chitin in tubule/cuticle morphogenesis. The finding that mutations in CDA genes are embryonic lethal suggests that CDAs play crucial roles during development, especially during embryogenesis and molting.

Continued identification and characterization of novel genes required for processes unique to insects will expand our options for selective biopesticide design. The chitin-containing insect exoskeleton, peritrophic matrix and tracheal lining represent unique physiological adaptations and also are promising targets for biopesticide disruption. We have identified some interesting genes in several insect species encoding CDAs that interact with and modify chitin and are essential for developmental and other processes. Future research will address further the physiological functions of each of the CDAs in *Tribolium* and other insect species.

## Acknowledgments

The authors are grateful to Drs. Ping Wang and Stefan Luschnig for critical reviews of this manuscript. We thank Jonna Voorhees for the supply of insects. Support was provided by the NSF grants IBN-0316963 and IOS 0615818. Mention of a proprietary product does not constitute a recommendation or endorsement by the USDA. The USDA is an equal opportunity/affirmative action employer and all agency services are available without discrimination. This is contribution #08-133-J of the Kansas Agricultural Experiment Station.

There is no conflict of interest.

## References

- Anderson, S.O., 2005. Cuticular sclerotization and tanning. In: Gilbert, L.I., Iatrou, K., Gill, S.S. (Eds.), *Molting and Metamorphosis*. Japan Scientific Societies Press, Tokyo, pp. 133–155.
- 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.
- Arakane, Y., Muthukrishnan, S., Kramer, K.J., Specht, C.A., Tomoyasu, Y., Lorenzen, M.D., Kanost, M.K., 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.
- Aruchami, M., Sundara-Rajulu, G., Gowri, G., 1986. Distribution of deacetylase in arthropod species. *Nat. Technol.*, 263–265.
- 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>>.
- Bendtsen, J.D., Nielsen, H., Heijne, G.V., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340, 783–795.
- Blair, D.E., Van Aalten, D.M., 2004. Structures of *Bacillus subtilis* PdaA, a family 4 carbohydrate esterase, and a complex with *N*-acetylglucosamine. *FEBS Lett.* 570, 13–19.
- Blair, D.E., Schuttelkopf, A.W., MacRae, J.I., Van Aalten, D.M., 2005. Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15429–15434.
- Blair, D.E., Hekmat, O., Schuttelkopf, A.W., Shrestha, B., Tokuyasu, K., Withers, S.G., Van Aalten, D.M., 2006. Structure and mechanism of chitin deacetylase from the fungal pathogen *Colletotrichum lindemuthianum*. *Biochemistry* 45, 9416–9426.
- Combet, C., Blanchet, C., Geourjon, C., Deléage, G., 2000. NPS: network protein sequence analysis. *Trends Biochem. Sci.* 25, 147–150.
- Combet, C., Jambon, M., Deleage, G., Geourjon, C., 2002. Geno3D: automatic comparative molecular modeling of protein. *Bioinformatics* 18, 213–214.
- Coutinho, P.M., Henrissat, B., 1999. Carbohydrate-active enzymes: an integrated database approach. In: Gilbert, H.J., Davies, G., Henrissat, B., Svensson, B. (Eds.), *Recent Advances in Carbohydrate Bioengineering*. The Royal Society of Chemistry, Cambridge, UK, pp. 3–12.
- Fass, D., Blacklow, S., Kim, P.S., Berger, J.M., 1997. Molecular basis of familial hypercholesterolemia from structure of LDL receptor module. *Nature* 388, 691–693.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A., 2003. *ExPASy*: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31, 3784–3788.
- Guo, W., Li, G., Pang, Y., Wang, P., 2005. A novel chitin-binding protein identified from the peritrophic membrane of the cabbage looper, *Trichoplusia ni*. *Insect Biochem. Mol. Biol.* 35, 1224–1234.
- Haliscak, J.P., Beeman, R.W., 1983. Status of malathion resistance in five genera of beetles infesting farm-stored corn, wheat, and oats in the United States. *J. Econ. Entomol.* 76, 717–722.
- Hekmat, O., Tokuyasu, K., Withers, S.G., 2003. Subsite structure of the endo-type chitin deacetylase from a Deuteromycete, *Colletotrichum lindemuthianum*: an investigation using steady-state kinetic analysis and MS. *Biochem. J.* 374, 369–380.
- John, M., Rohrig, H., Schmidt, J., Wieneke, U., Schell, J., 1993. *Rhizobium* NodB protein involved in nodulation signal synthesis is a chitoooligosaccharide deacetylase. *Proc. Natl. Acad. Sci. U.S.A.* 90, 625–629.
- Kean, T., Roth, S., Thanou, M., 2005. Trimethylated chitosans as non-viral gene delivery vectors: cytotoxicity and transfection efficiency. *J. Control. Release* 103, 643–653.

- Kramer, K.J., Muthukrishnan, S., 1997. Insect chitinases: molecular biology and potential use as biopesticides. *Insect Biochem. Mol. Biol.* 27, 887–900.
- Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* 5, 150–163.
- Luschnig, S., Bätz, T., Armbruster, K., Krasnow, M.A., 2006. *Serpentine* and *Vermiform* encode matrix proteins with chitin-binding and deacetylation domains that limit tracheal tube length in *Drosophila*. *Curr. Biol.* 16, 186–194.
- Moussian, B., Schwarz, H., Bartoszewski, S., Nüsslein-Volhard, C., 2005. Involvement of chitin in exoskeleton morphogenesis in *Drosophila melanogaster*. *Journal of Morphology* 264, 117–130.
- Pusateri, A.E., McCarthy, S.J., Gregory, K.W., Harris, R.A., Cardenas, L., McManus, A.T., Goodwin Jr., C.W., 2003. Effect of a chitosan-based hemostatic dressing on blood loss and survival in a model of severe venous hemorrhage and hepatic injury in swine. *J. Trauma* 4, 177–182.
- Tokuyasu, K., 1999. Utilization of chitin and chitosan in food industry. *J. Jpn. Soc. Food Sci. Technol.* 46, 356–360.
- Tokuyasu, K., Mitsutomi, M., Yamaguchi, I., Hayashi, K., Mori, Y., 2000. Recognition of chitooligosaccharides and their *N*-acetyl groups by putative subsites of chitin deacetylase from a deuteromycete, *Colletotrichum lindemuthianum*. *Biochemistry* 39, 8837–8843.
- Tsigos, I., Martinou, A., Kafetzopoulos, D., Bouriotis, V., 2000. Chitin deacetylases: new, versatile tools in biotechnology. *Trends Biotechnol.* 18, 305–312.
- Vollmer, W., Tomasz, A., 2000. The *PgdA* gene encodes for a peptidoglycan *N*-acetylglucosamine deacetylase in *Streptococcus pneumoniae*. *J. Biol. Chem.* 275, 20496–20501.
- Wang, S., Jayaram, A.S., Hemphala, J., Senti, K.A., Tsarouhas, V., Jin, H., Samakovlis, C., 2006. Septate-junction-dependent luminal deposition of chitin deacetylases restricts tube elongation in the *Drosophila* trachea. *Curr. Biol.* 16, 180–185.