Insect Biochemistry and Molecular Biology 62 (2015) 75–85

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

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

Structural features, evolutionary relationships, and transcriptional regulation of C-type lectin-domain proteins in *Manduca sexta*

Xiang-Jun Rao [a, 1], Xiaolong Cao [b, 1], Yan He [b], Yingxia Hu [b], Xiufeng Zhang [b], Yun-Ru Chen [c], Gary Blissard [c], Michael R. Kanost [d], Xiao-Qiang Yu [e], Haobo Jiang [b, *]

[a] School of Plant Protection, Anhui Agricultural University, Hefei, Anhui 230036, PR China
[b] Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK 74078, USA
[c] Boyce Thompson Institute, Cornell University, Ithaca, NY 14853, USA
[d] Department of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan, KS 66506, USA
[e] Division of Molecular Biology and Biochemistry, School of Biological Sciences, University of Missouri-Kansas City, Kansas City, MO 64110, USA

ARTICLE INFO

Article history:
Received 14 October 2014
Received in revised form
3 December 2014
Accepted 8 December 2014
Available online 29 December 2014

Keywords:
Insect immunity
Pattern recognition
Carbohydrate recognition domain
Expression profiling
Comparative genomics

ABSTRACT

C-type lectins (CTLs) are a large family of Ca^{2+}-dependent carbohydrate-binding proteins recognizing various glycoconjugates and functioning primarily in immunity and cell adhesion. We have identified 34 CTLDP (for CTL-domain protein) genes in the *Manduca sexta* genome, which encode proteins with one to three CTL domains. CTL-S1 through S9 (S for simple) have one or three CTL domains; immulectin-1 through 19 have two CTL domains; CTL-X1 through X6 (X for complex) have one or two CTL domains along with other structural modules. Nine simple CTLs and seventeen immulectins have a signal peptide and are likely extracellular. Five complex CTLs have both an N-terminal signal peptide and a C-terminal transmembrane region, indicating that they are membrane anchored. Immulectins exist broadly in Lepidoptera and lineage-specific gene duplications have generated three clusters of fourteen genes in the *M. sexta* genome, thirteen of which have similar expression patterns. In contrast to the family expansion, CTL-S1~S6, S8, and X1~X6 have 1:1 orthologs in at least four lepidopteran/dipteran/coleopteran species, suggestive of conserved functions in a wide range of holometabolous insects. Structural modeling suggests the key residues for Ca^{2+}-dependent or independent binding of certain carbohydrates by CTL domains. Promoter analysis identified putative κB motifs in eighteen of the CTL genes, which did not have a strong correlation with immune inducibility in the mRNA or protein levels. Together, the gene identification, sequence comparisons, structure modeling, phylogenetic analysis, and expression profiling establish a solid foundation for future studies of *M. sexta* CTL-domain proteins.

1. Introduction

Insect innate immune systems utilize soluble and membrane-bound receptors to recognize pathogen-associated molecular patterns (Kanost et al., 2004; Lemaitre and Hoffmann, 2007). Peptidoglycan recognition proteins, β-1,3-glucanase-related proteins, and an array of lectins bind to polysaccharides, glycoproteins and glycolipids on pathogen surface to induce defense responses (Charroux et al., 2009; Jiang et al., 2010; Weis et al., 1998). Lectins are classified based on the domain architectures and action mechanisms (Gallagher, 1984). C-type lectins (CTLs) constitute one of the largest and most diverse families of lectins in animals. They require Ca^{2+} for maintaining carbohydrate-binding activities and structures (Cambi et al., 2005). Each CTL contains one or more carbohydrate recognition domains (CRDs), known as CTL domains. A CTL domain is composed of β-sheets, α-helices, and loops (Weis et al., 1991). CTL domains may participate in protein interaction and binding to lipids and inorganic surfaces, which does not always require Ca^{2+} (Zelensky and Gready, 2005).

Specificity of CTLs is governed by key residues in the CRDs, which interact with the cognate oligosaccharides through Ca^{2+}

Abbreviations: CTL, C-type lectin; CRD, carbohydrate recognition domain; CTLDP, CTL-domain protein; IML, immulectin; LC, low complexity; proPO, prophenoloxidase; MBL, rat mannose binding lectin; FPKM, fragments per kilobase of exon per million fragments mapped; C, control; I, induced; F, fat body; H, hemocytes; TM, transmembrane; SR, scavenger receptor.

* Corresponding author. Tel.: +1 405 744 9400; fax: +1 405 744 6039.
E-mail address: haobo.jiang@okstate.edu (H. Jiang).

[1] These authors have made equal contribution to this study.

http://dx.doi.org/10.1016/j.ibmb.2014.12.006
0965-1748/

coordination and a network of hydrogen bonds. In the Ca^{2+} binding site-2 of rat mannose-binding lectin A, Glu^{185}, Asn^{187}, Glu^{193}, Asn^{205} and Asp^{206} are implicated in specific interactions (Weis and Drickamer, 1994; Weis et al., 1992). CTLs containing a Glu-Pro-Asn (EPN) motif in the CRD are characteristic of mannose-binding and thus called mannose-type CTLs. CTLs with a Gln-Pro-Asp (QPD) motif are generally galactose-type CTLs (Zelensky and Gready, 2005). Most CTLs contain a single CTL domain. Immulectins (IMLs) from lepidopteran insects have two. CTLs with dual CTLDs are also found in *Tribolium castaneum* (Zou et al., 2007) and crustaceans (Yu and Kanost, 2001), but it is unknown whether they share a common ancestor or arose independently.

Genome-wide analyses in insects revealed a number of genes encoding proteins with one or more CTL domains (Dodd and Drickamer, 2001; Christophides et al., 2002; Waterhouse et al., 2007; Zou et al., 2007; Tanaka et al., 2008). *Drosophila melanogaster*, *Anopheles gambiae*, *Aedes aegypti*, *T. castaneum*, and *Bombyx mori* have 34, 25, 39, 17 and 21 such genes, respectively. Since it is unclear whether these proteins bind carbohydrates or not in the presence or absence of Ca^{2+}, we suggest that they be named CTL-domain proteins (CTLDPs) instead of CTLs. Individual immulectins have been identified and characterized in lepidopteran species (Yu and Kanost, 2008) (Table 1). In *Manduca sexta*, functions of IML-1~4 have been characterized biochemically. IML-1 can induce agglutination of Gram-positive and -negative bacteria and yeast in a Ca^{2+}-dependent manner (Yu et al., 1999). Injection of IML-2 antiserum into *M. sexta* larvae inhibits clearance of a Gram-negative bacterial pathogen, *Serratia marcescens*, and decreases larval survival after bacterial infection (Yu and Kanost, 2003). Recombinant CRD2 of IML-2 directly binds to *Caenorhabditis elegans* and a human filarial nematode *Brugia malayi* and enhances encapsulation and melanization of *C. elegans in vivo* (Yu and Kanost, 2004). CTLD2 of IML-2 interacts with proPO, and the extended loop of CTLD2 is important for ligand binding and proPO activation (Shi and Yu, 2012). IML-3 is translocated into hemocytes in response to microbial stimulation (Ling et al., 2008). IML-4 can bind to immobilized LPS and lipoteichoic acid (LTA) in the absence of Ca^{2+}, but agglutinate *Escherichia coli*, *Staphylococcus aureus* and *Saccharomyces cerevisiae* in a Ca^{2+}-dependent manner (Yu et al., 2006).

To acquire an overview of *M. sexta* CTLDPs, we annotated CTLDP genes in the *M. sexta* genome based on the RNA-Seq data. Multiple sequence alignment and phylogenetic analysis revealed orthologs in other insects and lineage-specific expansion of the immulectin genes in lepidopterans. Analysis of the RNA-Seq reads provided expression patterns of the CTLDP genes in different tissues and stages. Putative immune responsive elements in the promoter regions were identified, and we examined whether presence of these elements correlates with mRNA and protein level changes in larval hemolymph before and after the immune challenge (Zhang et al., 2011, 2014). We also studied sequence conservation and structure–function relationships via molecular modeling and discuss their potential roles insect physiological processes.

2. Materials and methods

2.1. Gene identification, sequence improvement, and feature prediction

Manduca Genome Assembly 1.0 and gene models in *Manduca* Official Gene Set 1.0 and Cufflinks Assembly 1.0 (X et al., 2014) were downloaded from *Manduca* Base (ftp://ftp.bioinformatics.ksu.edu/pub/Manduca/). CTL sequences from *M. sexta* and other insects were used as queries to search Cufflinks 1.0 using the TBLASTN algorithm with default settings. Hits with aligned regions longer than 30 residues and identity over 40% were retained for retrieving corresponding cDNA sequences. Correct open reading frames (ORFs) in the retrieved sequences were identified using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Errors resulting from problematic regions (*e.g.* NNN…) in the genome assembly were corrected after BLASTN search of *Manduca* Oases and Trinity Assemblies 3.0 of the RNA-Seq data (http://darwin.biochem.okstate.edu/blast/blast_links.html). The two genome-independent RNA-Seq assemblies (X et al., 2014) were developed to cross gaps between genome scaffolds/contigs and detect errors in the gene models. The manually improved sequences were incorporated into OGS 2.0. To uncover all genes in a cluster, which were often too similar to distinguish by Cufflinks 1.0, the relevant genome contigs were manually examined to identify exons based on the GT-AG rule and sequence alignment. All improved sequences were further validated by BLASTP homolog search of GenBank (http://www.ncbi.nlm.nih.gov/). Conserved domains and transmembrane regions were identified using SMART (http://smart.embl-heidelberg.de/smart/set_mode.cgi) and InterProScan (http://www.ebi.ac.uk/Tools/pfa/iprscan/). The domain architectures were plotted using DOG 2.0 (http://dog.biocuckoo.org/). Signal peptides were predicted using SignalP4.1 (http://www.cbs.dtu.dk/services/).

Table 1
Functions of some immulectins in lepidopteran insects.

Name	Inducibility	Tissue		agglutination	Ca^{2+} -dependence	Binding	proPO activation	Melanization	Encapsulation	Reference
		H	F							
M. sexta IML-1	Ec, Ml, Sc	No	Yes	Ec, Sa, Sc	Yes	–	Yes	No	Yes	Yu et al., 1999.
M. sexta IML-2	Ec, Ml, Sc	No	Yes	Ec	Yes/no*	Lipid A, LPS, PG, LTA, mannan, laminarin	Yes	Yes	Yes	Shi and Yu, 2012; Yu and Kanost, 2000, 2003, 2004;
M. sexta IML-3	Ec, Ml, Sc	No	Yes	Ec	Yes	LPS, LTA, laminarin	–	No	Yes	Yu et al., 2005; Ling et al., 2008;
M. sexta IML-4	Ec, Ml, Sc	No	Yes	Ec, Sa, Sc	Yes/no*	LPS, LTA, laminarin	Yes	Yes	Yes	Yu et al., 2006.
H. armigera CTL1~8	Bt, Sa, Pp, NPV, 20E	Yes	Yes	Ec, Sa (1,3); Ca (1); – (2, 3–8)	–	–	–	–	–	Chai et al., 2008; Wang et al., 2012.
H. cunea Hdd15	Ec, Ml	–	Yes	–	–	LPS	–	–	–	Shin et al., 2000;
B. mori LBP, MBP, etc.	Ec, Ml, Sc	Yes	Yes	Ec, Smi (LBP); Sma, Ml, Sc (MBP)	– (LBP); + (MBP)	Lipid A (LBP); TA, mannan, PG (MBP)	–	Yes (LBP)	–	Koizumi et al., 1997, 1999; Watanabe et al., 2006; Tanaka et al., 2008;

Bt, *Bacilus thuringiensis* Ml, *Micrococcus luteus*; Sa, *Staphylococcus aureus*; Ec, *Escherichia coli*; Sma, *Serratia marcescens*; Smi, *Salmonella minnesota*, Sc, *Saccharomyces cerevisiae*; Ca, *Candida albicans*; Pp, *Pichia pastoris*; NPV, nucleopolyhedrovirus; 20E, 20-hydroxyecdysone; H: hemocytes; F: fat body; PG, peptidoglycan; LPS, lipopolysaccharide; LTA, lipoteichoic acid; TA, teichoic acid; *: Ca^{2+} required for agglutination of *E. coli*, but not for binding to microbial components.