Available online at www.sciencedirect.com
SCIENCE DIRECT
Genomics 87 (2006) 399 – 409
GENOMICS
www.elsevier.com/locate/ygeno

An expansion of the dual clip-domain serine proteinase family in *Manduca sexta*: Gene organization, expression, and evolution of prophenoloxidase-activating proteinase-2, hemolymph proteinase 12, and other related proteinases ☆

Yang Wang [1], Zhen Zou [1], Haobo Jiang *

Department of Entomology and Plant Pathology, Oklahoma State University, 127 Noble Research Center, Stillwater, OK 74078, USA

Received 7 July 2005; accepted 14 October 2005
Available online 1 December 2005

Abstract

Prophenoloxidase-activating proteinases (PAPs) take part in insect defense responses including melanotic encapsulation and wound healing. To understand their gene structure and regulation, we screened a genomic library and isolated overlapping λ clones for *Manduca sexta* PAP-2, hemolymph proteinase 12 (HP12), and HP24. Complete nucleotide sequence analysis indicated that all three genes encode polypeptides with two regulatory clip domains at the amino terminus, a linker region, and a catalytic serine proteinase domain at the carboxyl terminus. Each gene contains eight exons, with introns located at equivalent positions. Similar sequences are present in introns as well as exons, indicating that these genes arose from recent gene duplication and sequence divergence. We analyzed their 5′ flanking sequences and identified putative immune and hormone responsive elements. Reverse transcription–polymerase chain reactions confirmed that PAP-2 and HP12 mRNA levels in the larval fat body and hemocytes increased after a bacterial challenge. However, HP24 expression was barely detected. PAP-2 transcripts in cultured fat body became less abundant after 20-hydroxyecdysone treatment. Thus, PAP-2, HP12, and HP24 mRNA levels are differentially regulated by immune and developmental signals. Comparison with HP15, HP23, and PAP-3 sequences suggested an evolutionary pathway of the dual clip-domain serine proteinases in *M. sexta*.

Keywords: Insect immunity; Hemolymph protein; Melanization; Gene cluster; Tobacco hornworm; Clip domain; Ecdysone

Introduction

Insects are capable of defending themselves against pathogen/parasite infection [7,8,15]. Their immune systems consist of physical barriers (e.g., integument, gut), cellular responses (e.g., phagocytosis, nodulation, and encapsulation), and reactions mediated by plasma proteins [e.g., coagulation, prophenoloxidase (proPO) activation, and effect of antimicrobial peptides]. Serine proteinase cascades coordinate some of these defense mechanisms [10,14]. For instance, the proteolytic activation of proPO yields active phenoloxidase (PO), which generates quinones for melanin synthesis, wound healing, and encapsulation [1,3,18].

In the tobacco hornworm *Manduca sexta*, proPO activation is catalyzed by proPO-activating proteinases (PAP-1, PAP-2, and PAP-3) in the presence of two serine proteinase homologues (SPHs) [9,11,12,25]. The SPHs contain an amino-terminal clip domain and a carboxyl-terminal proteinase-like domain that is catalytically inactive (due to the lack of one or more catalytic residues in a serine proteinase). Precursors of PAPs and SPHs are activated by a serine proteinase cascade triggered on recognition of pathogens. PAPs, also known as PPAEs for proPO-activating enzymes, have been isolated and cloned from other arthropod species, including *Holotrichia diomphalia*, *Bombyx mori*, and *Pacifastacus leniusculus* [16,21,23]. *B. mori* PPAE, *M. sexta* PAP-2, and *M. sexta* PAP-3 contain two clip domains at the amino terminus, but the other PAPs only have one. *M. sexta* SPH-1 and SPH-2

☆ Sequence data from this article have been deposited with the GenBank/EBI Data Libraries under Accession Nos. DQ115323 and DQ115324.

* Corresponding author. Fax: +1 405 744 6039.
E-mail address: haobo@okstate.edu (H. Jiang).

[1] These authors have made equal contributions to this work.

0888-7543/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2005.10.003

associate with immulectin-2, a C-type lectin that binds lipopolysaccharide and localizes PO to the bacterial surface. Additionally, serine proteinase inhibitors of the serpin superfamily negatively regulate PAPs [14].

Our knowledge on transcriptional regulation of PAP gene expression is limited. We recently published the gene structure and expression pattern of PAP-1 and PAP-3 [26,27]. In this paper, we report the discovery of a cluster of dual clip-domain serine proteinase genes in the *M. sexta* genome. We have investigated the structure and evolution of these genes. A search for conserved sequence motifs uncovered putative regulatory elements in their 5′ flanking regions. Tissue specificity, immune inducibility, developmental profile, and hormone responsiveness of the gene expression were surveyed.

Results

Isolation of cDNA and genomic clones for PAP-2, HP12, HP23, and HP24

We screened the *M. sexta* genomic library using a full-length PAP-2 cDNA probe [11] and isolated eight positive clones. λ11 and λ12 were selected for subcloning and sequence determination. λ11 contained exons 3–8 of the PAP-2 gene, whereas λ12 included exons 3–8 of *HP24* and exon 1 of *PAP-2* (Fig. 1A). To fill the gap in the PAP-2 sequences, we used two pairs of primers in exon 1 and exon 2 to amplify genomic DNA by nested PCR and obtained a 1-kb fragment (G1) containing the 3′ end of exon 1, the entire intron 1, and the 5′ end of exon 2. Similarly, a seminested PCR resulted in a 1.4-kb genomic fragment (G1.4) containing the 3′ end of exon 2, the entire intron 2, and the 5′ end of exon 3 of *PAP-2*. Perfect matches with the ends of λ11 and λ12 indicated that the PCR-derived fragments represented the gap between the λ clones.

The corresponding exons in λ12 were 67% identical in sequence to those in λ11. Based on a sequence comparison with PAP-2 cDNA, we predicted the exon–intron junctions of *HP24*. A BLASTX search of GenBank with the assembled exon sequences indicated that this partial gene encoded a serine proteinase with two clip domains. We designed a pair of primers in exons 3 and 8 of *HP24* and amplified a 960-bp cDNA fragment of HP24 and, thus, demonstrated that *HP24* is an active gene. Comparison of the genomic and cDNA sequences confirmed the predicted exon–intron structure. To isolate a full-length HP24 cDNA, we screened the cDNA libraries via RecA-mediated homologous pairing and biotin–streptavidin interaction [13]. Among the captured positive clones, 14 were PAP-2, 1 was HP12, and none were HP24.

We used the HP15 probe [13] and isolated a full-length cDNA clone for HP15 and for HP23. Highly similar sequences of HP15 and HP23 (Table 1) seem to be responsible for the cross-hybridization.

To isolate genomic clones for *HP12* and the rest of *HP24*, we amplified an HP12 cDNA fragment corresponding to exons 1 and 2 of *PAP-2*, labeled the product with [^{32}P]dCTP, screened the genomic library, and isolated four positives. Based on the Southern blot analysis, λ1 and λ3 were selected for subcloning and sequence determination. λ3 contained the entire *HP12* and exons 1–3 of *HP24*, whereas λ1 included exons 2 through 8 of *HP24* and exon 1 of *PAP-2* (Fig. 1A). Thus far, we have de termined the complete nucleotide sequence of a 35-kb region containing the HP12–HP24–PAP-2 gene cluster.

We searched the silkworm genome databases and the only dual clip-domain serine proteinase gene found was *PPAE*. The

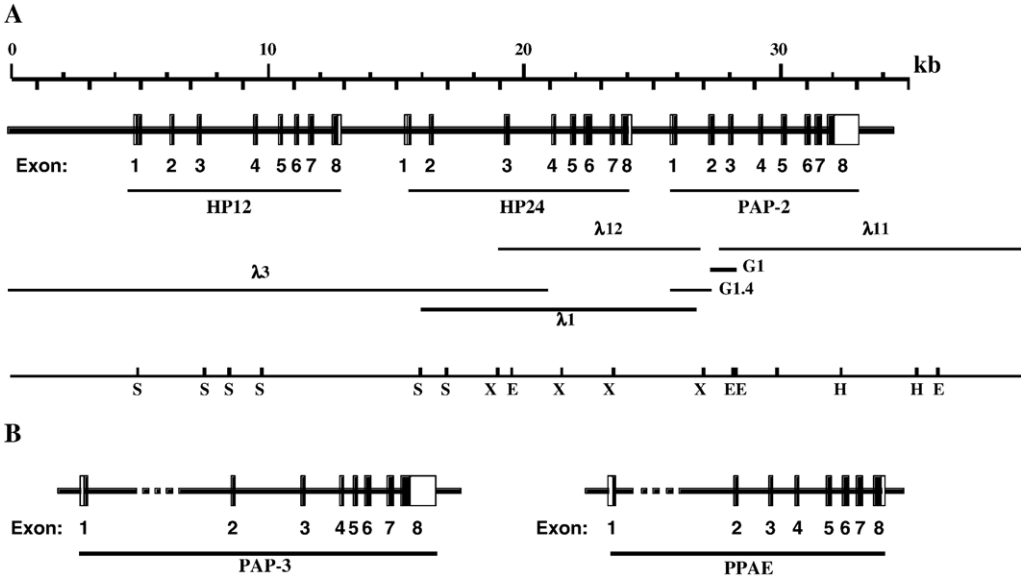

Fig. 1. Structures of *M. sexta HP12*, *HP24*, *PAP-2*, *PAP-3*, and *B. mori PPAE*. (A) Exon–intron organization, cloning, and restriction map of a 34.4-kb *M. sexta* genomic fragment containing *HP12*, *HP24*, and *PAP-2*. Locations of exons 1–8 (vertical bars) and introns 1–7 (horizontal lines) are indicated. Thin lines denote the genomic clones (λ1, λ3, λ11, and λ12) and PCR products (G1 and G1.4) analyzed by restriction digestion, subcloning, and sequencing. S, *Sal*I; X, *Xba*I; E, *Eco*RI; H, *Hin*dIII. (B) *M. sexta PAP-3* and silkworm *PPAE* structures [24,26]. The same scale is used for both panels. The 5′- and 3′-untranslated regions in exons 1 and 8 are marked by open boxes. Dashed lines represent unidentified genomic regions.