Published in final edited form as:

Insect Mol Biol. 2005 August ; 14(4): 433–442.

Gene structure and expression profile of *Manduca sexta* prophenoloxidase-activating proteinase-3 (PAP-3), an immune protein containing two clip domains

Z. Zou and H. Jiang

Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK, USA

Abstract

Prophenoloxidase-activating proteinase-3 (PAP-3) is a component of the defence system in *Manduca sexta*. We have isolated genomic clones and elucidated the organization of this gene. The 3' end of exon 2, the entire exon 3 and the 5' end of exon 4 encode the two amino-terminal clip domains. Southern blot analysis suggested a single copy of the *PAP-3* gene in the genome. We identified several putative immune-responsive elements in the upstream region. The *PAP-3* gene is not highly expressed in the fat body during larval development until the wandering stage begins. The mRNA level is high in the epithelium, fat body and haemocytes. Tissue-specific alternative splicing occurs in the fat body and trachea. A bacterial injection markedly induced the gene expression in the fat body and haemocytes.

Keywords

phenoloxidase; melanization; haemolymph protein; insect immunity; gene regulation

Introduction

To survive and prosper in pathogen-rich environments, insects must possess an efficient defence system in order to prevent and control microbial infection (Gillespie *et al.*, 1997). Sclerotized cuticle is effective in fending off infectious agents in the environment. Plasma proteins, haemocytes and fat body initiate a series of humoral and cellular responses to immobilize and kill invading microorganisms. Quinone production plays an important role in many of these immune mechanisms, including cuticle sclerotization, melanin synthesis, wound healing, encapsulation and free radical-mediated pathogen killing (Nappi & Vass, 2001). Phenoloxidase (PO), which catalyses quinone formation, is synthesized as an inactive zymogen prophenoloxidase (proPO) and activated by a cascade of serine proteinases (SPs), known as the proPO activation system (Ashida & Brey, 1998).

Analogous to the blood clotting and complement systems in vertebrates, the proPO activation cascade is triggered upon recognition of physiological or pathological signals. In the presence of microbial cell wall components (e.g. peptidoglycan, lipopolysaccharide and β -1,3-glucan), circulating or cell surface receptors specifically bind to these pathogen-associated molecular patterns and trigger autoactivation of the first proteinase precursor. This, in turn, leads to sequential cleavage activation of other cascade members. To date, our understanding of this pathway is largely limited to the final step, namely proPO activation by its activating enzyme prophenoloxidase-activating proteinase (PAP). PAP has been isolated and cloned from

Manduca sexta, *Bombyx mori*, *Holotrichia diomphalia* and *Pacifastacus leniusculus* (Jiang *et al.*, 1998,2003a,b; Lee *et al.*, 1998; Satoh *et al.*, 1999;Wang *et al.*, 2001). In *M. sexta* and *H. diomphalia*, PAP converts proPO to active PO only when clip-domain serine proteinase homologues (SPHs) are present.

The clip domain is a structural unit stabilized by three disulphide bonds. It is thought to have regulatory functions (Jiang & Kanost, 2000). While other PAPs contain one clip domain at the amino terminus, the *B. mori* proPO activating enzyme, *M. sexta* PAP-2 and PAP-3 have two – the role of the extra clip domain is not understood. Many arthropod SPs and SPHs contain clip domains. However, little is known about their structure, function or evolution. There is no information on the exon-intron organization and expression regulation of *PAP* genes.

With three terminal proteinases, the proPO activation system in *M. sexta* appears to be quite complex. A bacterial challenge significantly increased the mRNA and protein levels of PAP-1, PAP-2 and PAP-3 (Jiang *et al.*, 1998,2003a,b). Their tissue localization may also change at different developmental stages. For instance, PAP-1 is present in the prepupal cuticle whereas PAP-2 and PAP-3 accumulate in the haemolymph of the same stage insect. Profiling these PAP mRNAs in various tissues and cells should allow us to better understand the transcriptional regulation of these genes in a developmental context. In this paper, we report the exon-intron organization and expression profile of the *M. sexta* *PAP-3* gene.

Results

PAP-3 genomic clones

A *M. sexta* genomic library was screened with the full-length *PAP-3* cDNA and two positive clones, λ 2 and λ 3, were obtained (Fig. 1A). Sequence analysis indicated that a 5' cDNA fragment (115 bp) was absent in the genomic clones. Using the corresponding fragment derived from a PCR, we screened the genomic library again and isolated another positive clone, λ 7. From these three bacteriophages, we isolated DNA, mapped the restriction sites and subcloned ten genomic fragments into a vector. The inserts in the resulting recombinant plasmids, 24.4 kb long in total, were completely sequenced.

Clone λ 7, containing the 5' end of the *PAP-3* gene, did not overlap with clone λ 3 (Fig. 1A). Long-distance PCR using primers located at the 3' end of λ 7 and 5' end of λ 3 did not yield any PCR product (data not shown), indicating that the gap between the genomic clones is long. After the *Bam*HI site near the 3' end of λ 3 (Fig. 1A), there was a 167 bp sequence identical to the 5' end of λ 2 (Fig. 1C). While this result strongly suggested that these two genomic fragments overlap, we identified an extra 633 bp sequence at the 3' end of the λ 3 insert (Fig. 1C, clone 3–3). This fragment did not match the λ 2 sequence. To test whether or not it represents a cloning artefact, we amplified the genomic DNA using primers flanking the 167 bp overlap (Fig. 1A) and obtained a single 954 bp product. Cloning and sequence analysis indicated that the PCR product was identical to the overlap and flanking genomic sequences in λ 2 and λ 3, but does not contain the 633 bp fragment.

Exon–intron organization

A comparison between the genomic and cDNA sequences revealed that the *M. sexta* *PAP-3* gene is composed of eight exons and seven introns (Fig. 1B). Exon 1 includes a 5' untranslated region and a coding region for the nineteen-residue signal peptide of PAP-3 (Fig. 2). Exon 2 and the 5' end of exon 3 code for the first clip domain, whereas the rest of exon 3 and the 5' end of exon 4 encode the second clip domain. The remaining part of exon 4 covers the linker sequence and the first two residues of the SP domain. Most of the catalytic domain is encoded by exon 5, exon 6 and the 5' end of exon 7. The other part of exon 7 (1040 bp) corresponds to

the 3' untranslated region in the *PAP-3* cDNA, which contains the polyadenylation signal (AATAAA) ten nucleotides before the poly(A) tail.

Introns 1, 2 and 3 are > 10, 2.6 and 0.8 kb in length, respectively, whereas the average size of introns 4–7 is only 0.4 kb. We compared the 5' and 3' ends of these introns and identified the consensus sequence: 5'-GTR(/G)W(/A)D(/G)K(/T) and B(/T)TY(/T)BCAG-3', where the nucleotide in each set of parentheses appears in five to six out of the seven sequences at that position.

Transcription initiation, sequence variations and copy number

We determined the transcriptional initiation site of the *PAP-3* gene by primer extension. After annealing with RNA from the fat bodies of bacteria-induced larvae, the primer (derived from nucleotides 28–56 of the *PAP-3* coding region) was extended by ninety-six nucleotides by reverse transcriptase (Fig. 3). Therefore, the RNA synthesis started at an A (nucleotide +1), and there was a TCAGT sequence at nucleotides –2 to +3. This motif is typically present within ten nucleotides either side of the transcription initiation site in arthropod genes (Cherbas & Cherbas, 1993). We did not identify a TATA or Goldberg-Hogness box around the '–30 region'. Nevertheless, a perfect TATA sequence (TATAAA) was present at nucleotides –94 to –89 (Fig. 2).

The exons were nearly identical in sequence to the *PAP-3* cDNA clone isolated from the bacteria-induced fat body library (Jiang *et al.*, 2003b). There are nine nucleotide differences in the coding region, most likely caused by allelic variations. Two of these give rise to amino acid residue changes (T¹³⁷ to S¹³⁷, Y³⁸³ to F³⁸³).

Southern blot analysis was used to estimate the copy number of the *PAP-3* gene in the *M. sexta* genome (Fig. 4). The *Bam*HI digest (Fig. 4, lane 1) contained a 6 kb fragment that hybridized with the cDNA probe before exon 4 (Fig. 1A). The > 12 kb band with higher radioactivity probably included exons 4–8. A 2.3 kb *Hind*III fragment (Fig. 4, lane 2) matched the sequence between exon 4 and the 5' end of exon 7 (Fig. 1) and produced a strong hybridization signal. The other three bands (> 10 kb) probably included exons 1, 2–3 and 7–8. With two recognition sites in the *PAP-3* gene (Fig. 1A), *Sac*I yielded three high molecular weight bands that did not separate well (Fig. 4, lane 3). In contrast, *Xba*I generated a much smaller band at 3.1 kb (Fig. 4, lane 4), probably corresponding to exons 7 and 8. The other part of the *PAP-3* gene migrated to the ~11 kb position and hybridized with the probe much more strongly. *Eco*RI digestion generated two strong bands at 2.4 and > 10 kb and two weak bands at 2.7 and ~8 kb (Fig. 4, lane 5). Finally, *Sac*I and *Xba*I double digestion generated at least two bands (Fig. 4, lane 6). The lower one has the same mobility as the 3.1 kb *Xba*I fragment. The upper one (3.6 kb) strongly reacted with the cDNA probe in the region corresponding to exons 3–6. Longer exposure revealed three other bands at 5.5, 7 and 12 kb (data not shown). In summary, the number, size and intensity of these radioactive restriction fragments agree well with those predicted from the restriction enzyme map of the cloned gene (Fig. 1A). There is only one copy of the *PAP-3* gene in the *M. sexta* genome and the sequence variations (Fig. 2) are likely due to allelic differences in the insects used for library constructions.

Putative regulatory elements

In order to identify putative regulatory elements, we searched the 1.8 kb upstream sequence of the *PAP-3* gene by computer analysis (Table 1). Nine NF-κB motifs are present in this region, four on the plus strand and five on the minus strand. The motifs starting at nucleotides –1018 and –427 match nine out of ten positions in the NF-κB consensus sequence, originally derived from *Drosophila melanogaster* (Hultmark, 1993). A perfect match is located on the minus strand from –1756 to –1765. There are five interferon-stimulated response elements (ISREs)

beginning at nucleotides −1724, −944, −292, −213 and −193. In comparison with the thirteen-base mammalian ISRE consensus, all these sequences contain two mismatches. An ISRE was identified in the *Anopheles gambiae* prophenoloxidase gene-1 (Ahmed *et al.*, 1999).

Four GATA boxes exist at positions −1713, −1137, −536 and −145 on the plus strand. We have also identified three ecdysone response elements (EcREs) on the same strand at positions −804, −291 and −6. While the GATA boxes perfectly match the consensus (WGATAA), three of the fifteen nucleotides in the putative EcREs are not consistent with the motif (RRGKTCANTGAMCY) (Cherbas & Cherbas, 1996).

A 151 bp segment in the 5' upstream region is 85% identical in nucleotide sequence to an intron region in a *Bombyx mori* ABC transporter gene (Fig. 5A). It is unclear whether or not these sequences represent a lepidopteran repetitive sequence with a regulatory function. Using BLASTN, we searched the GENBANK database with the PAP-3 intron sequences and identified several high-score matches from *M. sexta* and other insects. For instance, a 68 bp fragment between exons 3 and 4 is 85–93% similar to intron sequences from the *M. sexta* juvenile hormone binding protein, eclosion hormone, serpin-1 and arylphorin genes (Fig. 5B). The significance of these similar sequences is not understood.

Temporal and spatial regulation of transcription

To understand the regulation of PAP-3 transcription, we examined its mRNA levels in different tissues and development stages (Fig. 6). Semi-quantitative RT-PCR using PAP-3 specific primers resulted in PCR products of the expected sizes. Due to a 75 bp insert (Jiang *et al.*, 2003b and data not shown), most of the PAP-3 transcripts in the midgut gave rise to a 438 bp PCR product that migrated slower than the 363 bp one from the fat body, muscle, integument, salivary gland, and nerve tissue (Fig. 6A). Both transcripts were detected in haemocytes, the trachea and Malpighian tubules. There was no major change in relative abundances of the mRNAs between feeding and wandering larvae. Relative intensities indicated that PAP-3 transcripts were present at a low level in the fat body and haemocytes of the naive larvae and, after a bacterial challenge, the mRNA levels increased considerably (data not shown). This result, consistent with our previous RT-PCR data (Jiang *et al.*, 2003b), and the existence of immune-responsive elements in the PAP-3 gene, validated our methods for estimating the mRNA levels. It also agreed well with data from an immunoblot analysis of the haemolymph samples (Jiang *et al.*, 2003b) and immunohistochemical analysis of haemocytes (Fig. 7).

Considering its roles as a centre for intermediary metabolism and in haemolymph protein synthesis in insects, we closely monitored the PAP-3 mRNA level in the fat body at different developmental stages. The PAP-3 gene was transcribed at a low level in this tissue throughout the last two larval feeding stages. There was a small, daily increase in the first five days in the fourth instars and a decrease thereafter (Fig. 6B). The mRNA was almost undetectable after day 3 of the fifth instar. However, after the insects entered the wandering stage, PAP-3 mRNA level largely increased and then gradually decreased. The transcripts became abundant again in the early pupal stages and reduced subsequently. The mRNA was still detectable in the fat bodies from late wandering larvae, late pupae and adults.

Discussion

Although it is not well studied, transcriptional regulation of immune genes in lepidopteran insects is generally considered to be similar to that in *Drosophila*. In *Drosophila*, transcription factors of the Rel family (i.e. Dorsal, Dif and Relish) control the expression of many immune genes (Engstrom, 1998; Harshman & James, 1998). They form homodimers or heterodimers, bind to their recognition sequences (NF- κ B motifs), and induce the synthesis of antimicrobial peptides (e.g. dipterecin, cecropin and drosomycin) and other acute-phase proteins (Hoffmann,

2003). The nuclear translocation of these transcription factors are regulated by the Toll and Imd pathways. Pathogen recognition directly or indirectly stimulates these intracellular signalling pathways. For example, *Drosophila* PGRP-SA, Persephone (a clip-domain SP), and cleaved spätzle are components of an extracellular signalling pathway that activates the Toll receptor (Michel *et al.*, 2001;Choe *et al.*, 2002;Ligoxygakis *et al.*, 2002;Ramet *et al.*, 2002;Takehana *et al.*, 2002). NF- κ B motifs and their binding proteins exist in other insects. *Cecropia* immune-responsive factor was identified in nuclear extracts from induced *Hyalophora cecropia* pupae (Sun & Faye, 1992). Responsive elements similar to the *Drosophila* NF- κ B motif (GGGRAYYYYY) are present in the 5' flanking region of the *M. sexta* haemolin and lysozyme genes (Wang *et al.*, 1995;Mulnix & Dunn, 1994).

In this work, we elucidated the structure of *M. sexta* PAP-3, an immune-responsive SP involved in melanization and other defence mechanisms. In particular, nine NF- κ B motifs were identified, three of which closely resemble the *Drosophila* consensus (Table 1). Like other immune responsive elements (e.g. ISREs, GATA boxes), the role of these putative recognition sequences has to be further investigated.

Many arthropod SPs and SPHs involved in immunity and development contain clip domains (Jiang & Kanost, 2000). One of the five *Drosophila* SPs/SPHs has such a regulatory unit (Ross *et al.*, 2003). So far, the structure and function of clip domains are poorly understood and their evolutionary history has not been closely examined. We found that the *Drosophila* SPs/SPHs with multiple clip domains are different in domain subgroups and linker sizes from *B. mori* proPO-activating enzyme (PPAE), and *M. sexta* PAP-2 and PAP-3 (Satoh *et al.*, 1999;Jiang *et al.*, 2003a,b). Perhaps PPAE-like SPs evolved after Lepidoptera and Diptera diverged from their common ancestor ~330 Ma (Gaunt & Miles, 2002), and there was a subsequent expansion of dual clip-domain SPs in lepidopteran insects. In this study, we discovered that each clip domain in PAP-3 did not evolve as a complete exon unit(s). While such an exon-domain relationship impedes domain shuffling/duplication, mutations at the splicing junctions may have contributed to the sequence diversity within the two domains. Exon duplication is responsible for the domain organization of *Drosophila* masquerade, whose clip domains 1, 2, 3–4 and 5 are separately encoded by four exons (<http://www.flybase.org/>).

The PAP-3 mRNA levels varied in different tissues of *M. sexta*. In the fifth instar larvae, PAP-3 transcripts were more abundant in the haemocytes and trachea than in the fat body. After the larvae entered the wandering stage, however, the fat body became a major source of PAP-3 expression. Additionally, the size of the PCR product was 75 bp smaller in the fat body than that in some other tissues (Fig. 6A, Jiang *et al.*, 2003b). At this moment, we do not understand the significance or mechanism of such tissue-specific regulation of PAP-3 transcription/splicing. In fact, the 75 bp sequence was not identified in the sequenced regions of λ 7 and λ 3.

The transcription profiles of PAP-3 and PAP-1 (Zou, unpubl. data) are somewhat similar in the fat body at different developmental stages. Their up-regulation during the wandering stage (days 1–4) and down-regulation in the middle-to-late pupae stage suggest that hormonal signals might play a role the transcriptional control of PAP-1 and PAP-3 expression. Indeed, upon treatment with 20-hydroxyecdysone, the PAP-1 mRNA level in the cultured larval fat body reduced in a concentration-dependent manner (Z. Zou, unpubl. data). We identified three ecdysone responsive elements in both genes. Nonetheless, we did not detect a significant change in PAP-3 mRNA level after the cultured fat body was incubated with the moulting hormone (Z. Zou & H. Jiang, data not shown). The hormonal regulation of immune proteins is worth exploring in the future.

Experimental procedures

Insect rearing and bacterial injection

M. sexta larvae were hatched from eggs (Carolina Biological Supply Co, Burlington, NC) and reared according to Dunn & Drake (1983). For bacterial induction, day 2 fifth instar larvae were injected with formalin-killed *Escherichia coli* XL1-blue (2×10^8 cells/larva). Haemolymph and fat body samples were collected 24 h later for RNA preparation.

Library screening, subcloning, sequencing and computer analysis

To screen the genomic library, full-length or PCR-derived PAP-3 cDNA was labelled with α - ^{32}P -dCTP using Multiprime DNA Labelling System (Amersham Pharmacia Biotech, Bucks, UK). The intact cDNA was retrieved from the PAP-3 clone by restriction enzyme digestion, whereas the 5' PCR fragment (115 bp) was amplified using primers j662 (5'-CGA CGC TGA GGT AAC ACG T-3', nucleotides 2–20) and j663 (5'-TTG TCC ACT AAC AAA GCA G-3', reverse complement of nucleotides 98–116). A *M. sexta* genomic library, generously provided by Dr Yucheng Zhu at the Southern Insect Management Research Unit (USDA-ARS), was screened with the cDNA probes at 5×10^5 cpm/ml by a standard protocol (Sambrook & Russell, 2001). Positive plaques were purified to homogeneity by repeated screening, and λ DNA was prepared from the amplified bacteriophages using Wizard Lambda Preps DNA Purification System (Promega, Madison, WI). Restriction maps were determined by single and double digestions of the DNA samples with *Eco*RI, *Sac*I, and *Xba*I. Following agarose gel electrophoresis and capillary transfer, the separated digestion products on the nitrocellulose membrane were hybridized with the cDNA probes and visualized by autoradiography. PAP-3 gene fragments were subcloned into the same sites in pBluescript-(KS). The resulting recombinant plasmids were sequenced using a BigDye v2.0 Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, Foster City, CA). Sequence editing, assembly, and analysis were performed using MacVector Sequence Analysis Software (Version 6.5, Oxford Molecular Ltd, Oxford, UK).

Determination of the PAP-3 transcription initiation site

Primer extension was carried out as described previously (Sambrook & Russell, 2001) in order to determine the transcription starting position. An oligonucleotide (j396, 5'-CCA CTA ACA AAG CAG AAA TAC GAA GCT AA-3'), corresponding to the reverse complement of nucleotides 28–56 of the PAP-3 coding region, was terminally labelled with γ - ^{32}P -ATP using T4 polynucleotide kinase (Promega). The primer (1.0×10^5 cpm/ml) was added to 15 μg total RNA, incubated at 60 °C for 15 min, and then slowly cooled to below 30 °C in 45 min. Annealed primer-RNA complexes were extended with MMLV reverse transcriptase (200 U/ μl , 1 μl) for 1 h at 42 °C. The extension products as well as the ^{35}S -labelled DNA size standards were separated on 8% polyacrylamide gels containing 7 M urea and visualized by autoradiography.

Southern blot analysis

M. sexta genomic DNA was extracted from a single fifth instar larva. Aliquots of the DNA sample (15 μg each) were incubated with restriction enzymes (*Bam*HI, *Eco*RI, *Hind*III, *Sac*I, *Xba*I, *Sac*I-*Xba*I) at 37 °C for 6 h. After electrophoretic separation on a 1% agarose gel and capillary transfer on to GeneScreen-Plus nitrocellulose membrane (NEN Life Science Products, Boston, MA), the digested genomic fragments on the membrane were hybridized with α - ^{32}P -labelled full-length PAP-3 cDNA.

Immunocytochemistry

At 1 and 24 h after day 2 fifth instar larvae were injected with *E. coli* XL1-blue (1×10^7 /larva), haemocytes were collected, washed and fixed on multiwell slides as described by Willott *et*

al. (1995). These cells were incubated with 1 : 1000 diluted PAP-3 antiserum overnight at 4 ° C. After removing excess antibodies with Tris-buffered saline, the haemocytes were reacted with 1 : 500 diluted FITC-conjugated goat-anti-rabbit IgG (Bio-Rad, Hercules, CA) at room temperature for 1 h. Following a washing step, cells were observed and photographed with phase contrast or fluorescence optics under an Olympus BH-2 microscope.

RNA extraction and RT-PCR analysis

RNA samples were extracted from various tissues of *M. sexta* at different developmental stages (see Fig. 5 legend for details), using Micro-to-Midi Total RNA purification system (Invitrogen Life Technologies, Carlsbad, CA). Similarly, fat body and haemocyte RNA samples were isolated from naïve and bacteria-challenged *M. sexta* larvae. First-strand cDNA synthesis was performed using 2–4 µg RNA, 10 pmol oligo(dT)₁₇, and 200 U MMLV reverse transcriptase (Invitrogen Life Technologies) at 37 ° C for 1 h. *M. sexta* ribosomal protein S3 transcripts were used as an internal standard to control the template amount in a preliminary PCR experiment. Relative levels of *PAP-3* cDNA in the normalized samples were determined by semiquantitative PCR using j699 (5'-CGT GTT GTT ATT AGC TTC GTA TTT CT-3') and j700 (5'-CAT CCC CCC AGC CTC TAC-3' for *PAP-3*). The thermal cycling conditions were: 94 ° C for 30 s; 50 ° C for 30 s; 72 ° C for 60 s. PCR cycle numbers were empirically chosen to show comparable band intensity and avoid reaction saturation. After electrophoretic separation on a 1.3% agarose gel, intensities of the PCR products were quantified and compared using Digital Science 1D Gel Analysis Software (Kodak, Rochester, NY).

GENBANK accession numbers

The nucleotide sequences reported in this paper have been submitted to the GENBANK™/EBI Data Bank with the accession numbers AY789466 and AY789467.

Acknowledgements

This work was supported by the National Institutes of Health Grant GM58634 (to H.J.). We thank Drs Melcher and Dillwith for their helpful comments on the manuscript. This article was approved for publication by the Director of Oklahoma Agricultural Experimental Station and supported in part under project OKLO2450.

References

- Ahmed A, Martin D, Manetti AG, Han SJ, Lee WJ, Mathiopoulos KD, Muller HM, Kafatos FC, Raikhel A, Brey PT. Genomic structure and ecdysone regulation of the prophenoloxidase 1 gene in the malaria vector *Anopheles gambiae*. *Proc Natl Acad Sci USA* 1999;96:14795–14800. [PubMed: 10611292]
- Ashida, M.; Brey, PT. Recent advances on the research of the insect prophenoloxidase cascade. In: Brey, PT.; Hultmark, D., editors. *Molecular Mechanisms of Immune Responses in Insects*. London: Chapman & Hall; 1998. p. 135-172.
- Cherbas L, Cherbas P. The arthropod initiator: the capsite consensus plays an important role in transcription. *Insect Biochem Mol Biol* 1993;23:81–90. [PubMed: 8485519]
- Cherbas, L.; Cherbas, P. Molecular aspects of ecdysteroid hormone action. In: Gillbert, LI.; Tata, JR.; Atkinson, BG., editors. *Metamorphosis: Postembryonic Reprogramming of Gene Expression in Amphibian and Insect Cells*. San Diego, California: Academic Press; 1996. p. 175-221.
- Choe KM, Werner T, Stoven S, Hultmark D, Anderson KV. Requirement for a peptidoglycan recognition protein (PGRP) in Relish activation and antibacterial immune responses in *Drosophila*. *Science* 2002;296:359–362. [PubMed: 11872802]
- Dunn P, Drake D. Fate of bacteria injected into naïve and immunized larvae of the tobacco hornworm, *Manduca sexta*. *J Invert Pathol* 1983;41:77–85.
- Engstrom, Y. Insect immune gene regulation. In: Brey, PT.; Hultmark, D., editors. *Molecular Mechanisms of Immune Responses in Insects*. London: Chapman & Hall; 1998. p. 211-244.

- Gaunt MW, Miles MA. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol Biol Evol* 2002;19:748–761. [PubMed: 11961108]
- Gillespie JP, Kanost MR, Trenczek T. Biological mediators of insect immunity. *Annu Rev Entomol* 1997;42:611–643. [PubMed: 9017902]
- Harshman LG, James AA. Differential gene expression in insects: transcriptional control. *Annu Rev Entomol* 1998;43:671–700. [PubMed: 9444758]
- Hoffmann JA. The immune response of *Drosophila*. *Nature* 2003;426:33–38. [PubMed: 14603309]
- Hultmark D. Immune reactions in *Drosophila* and other insects: a model for innate immunity. *Trends Genet* 1993;9:178–183. [PubMed: 8337755]
- Jiang H, Kanost MR. The clip-domain family of serine proteinases in arthropods. *Insect Biochem Mol Biol* 2000;30:95–105. [PubMed: 10696585]
- Jiang H, Wang Y, Kanost MR. Pro-phenoloxidase activating proteinase form an insect, *Manduca sexta*: a bacteria-inducible protein similar to *Drosophila* easter. *Proc Natl Acad Sci USA* 1998;95:12220–12225. [PubMed: 9770467]
- Jiang H, Wang Y, Yu XQ, Kanost MR. Prophenoloxidase-activating proteinase-2 from hemolymph of *Manduca sexta*: a bacteria-inducible serine proteinase containing two clip domains. *J Biol Chem* 2003a;278:3552–3561. [PubMed: 12456683]
- Jiang H, Wang Y, Yu XQ, Zhu Y, Kanost M. Prophenoloxidase-activating proteinase-3 (PAP-3) from *Manduca sexta* hemolymph: a clip-domain serine proteinase regulated by serpin-1J and serine proteinase homologs. *Insect Biochem Mol Biol* 2003b;33:1049–1060. [PubMed: 14505699]
- Ligoxygakis P, Pelte N, Hoffmann JA, Reichhart JM. Activation of *Drosophila* Toll during fungal infection by a blood serine protease. *Science* 2002;297:114–116. [PubMed: 12098703]
- Michel T, Reichhart JM, Hoffmann JA, Royet J. *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. *Nature* 2001;414:756–759. [PubMed: 11742401]
- Mulnix AB, Dunn PE. Structure and induction of a lysozyme gene from the tobacco hornworm, *Manduca sexta*. *Insect Biochem Mol Biol* 1994;24:271–281. [PubMed: 7517269]
- Nappi AJ, Vass E. Cytotoxic reactions associated with insect immunity. *Adv Exp Med Biol* 2001;484:329–348. [PubMed: 11419001]
- Ramet M, Manfrulli P, Pearson A, Mathey-Prevot B, Ezekowitz RA. Functional genomic analysis of phagocytosis and identification of a *Drosophila* receptor for *E. coli*. *Nature* 2002;416:644–648. [PubMed: 11912489]
- Ross J, Jiang H, Kanost MR, Wang Y. Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationship. *Gene* 2003;304:117–131. [PubMed: 12568721]
- Sambrook J, Russell DW. *Molecular Cloning, a Laboratory Manual*. 3rd edn.. New York: Cold Spring Harbor Laboratory Press; 2001.
- Satoh D, Horii A, Ochiai M, Ashida M. Prophenoloxidase-activating enzyme of the silkworm, *Bombyx mori*. Purification, characterization, and cDNA cloning. *J Biol Chem* 1999;274:7441–7453. [PubMed: 10066809]
- Sun SC, Faye I. *Cecropia* immunoresponsive factor, an insect immunoresponsive factor with DNA-binding properties similar to nuclear-factor kappa B. *Eur J Biochem* 1992;204:885–892. [PubMed: 1541299]
- Takehana A, Katsuyama T, Yano T, Oshima Y, Takada H, Aigaki T, Kurata S. Overexpression of a pattern-recognition receptor, peptidoglycan-recognition protein-LE, activates imd/relish-mediated antibacterial defense and the prophenoloxidase cascade in *Drosophila* larvae. *Proc Natl Acad Sci USA* 2002;99:13705–13710. [PubMed: 12359879]
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680. [PubMed: 7984417]
- Wang R, Lee SY, Cerenius L, Söderhäll K. Properties of the prophenoloxidase activating enzyme of the freshwater crayfish, *Pacifastacus leniusculus*. *Eur J Biochem* 2001;268:895–902. [PubMed: 11179955]

- Wang Y, Willott E, Kanost MR. Organization and expression of the hemolin gene, a member of the immunoglobulin superfamily in an insect, *Manduca sexta*. *Insect Mol Biol* 1995;4:113–123. [PubMed: 7551193]
- Willott E, Lowenberger C, Christensen BM, Kanost MR. Monoclonal antibodies against *Manduca sexta* hemocytes bind *Aedes aegypti* hemocytes: characterization of six monoclonal antibodies that bind hemocytes from both species. *Dev Comp Immunol* 1995;19:451–461. [PubMed: 8773196]

Figure 1.

Structure of the *Manduca sexta* PAP-3 gene. (A) Restriction map of PAP-3 genomic inserts in the positive clones $\lambda 2$, $\lambda 3$ and $\lambda 7$. Arrow heads and vertical bars indicate the start and end points of the clones. B, *Bam*HI; E, *Eco*RI; H, *Hind*III; S, *Sac*I; X, *Xba*I. The underlined *Hind*III site denotes its presence in the cDNA. The 4.51 kb fragment at the 5' end of $\lambda 3$ contains two *Hind*III sites (in parentheses) at positions 1110 and 1500. Horizontal bars mark the PAP-3 genomic fragments used for subcloning. The segment '3-3' represents the 3' end of $\lambda 3$ whose sequence does not match $\lambda 2$. Dotted lines, regions not sequenced; broken lines, 3' end of $\lambda 7$ and 5' end of $\lambda 3$, sequenced. (B) Exon-intron organization of the PAP-3 gene. Numbered vertical bars denote the exons, with the non-coding regions shown as open boxes. S, signal peptide; C1 and C2, clip domains 1 and 2; L, linker region; P, proteinase domain. The same scale and starting site are used for the restriction map and exon-intron structure. (C) Sequence of the *Bam*HI-*Bgl*II fragment at the 3' end of $\lambda 3$. The mismatching region (clone 3-3, panel A), starting two nucleotides before the *Eco*RI site, is shown in italics. The five nucleotides AGATC (underlined) resemble the recognition sequence (AGATCT) of *Bgl*II, a restriction enzyme used for the library construction.

-1804 GGTGAAATTCCTTCTATTCTCTCGACCTA**AGAAGTCCCC**AGCAAGTGTCTATACATCCGAGTGATTTCATACAGTAT**GGTAAAAAATGATAAT**CA
 ---- 0.5 kb (-1205 to -1704)
 -1204 TCACGAAACAATAATGTTAAAAATCTATTGGTCAACATATCCAATATCGCTTTATTACTACGT**AGATAA**GACTAACGACAAACATTGATCCAACCT
 -1104 ATACAGCTATAGTATCGCCACTCGCCAAATATCTAAATGAATGACGAACGCACTGTAACAAACAACTAAGCGACTATCAACGATGGG**CAATCCCA**GAGAA
 -1004 GATATTGAAGTACACCGGTTTGATCTTGGCAATTAATAATATGGGACT**GTATCAGTTT**ACATTAAGTTGTAATTAACCTGCGG**GGGATCTC**ACTGTATC
 -904 GTGTCTACAAGATCTCTGTGTTGTATAGAATAATAATAGACGCGCAATTAATGTTTAAATAATTAATATTGTGCGATGTAGCTGTGAGTCCATA
 -804 **AGT**TT**GAATG**CACTTGTGTTCCAGGATTTCTGTTGTA**AGAAATCC**CTGATACAGCTCGAGGGCTTTGCGAGTGTACCGAGATGGTGTGACTAGAAAAACCAA
 -704 **AACAAGTTAC**GAGATTTTCAT**GGAGGAATCTT**GAACCTTTGGACAGACTTTACTTAATCTGCCTAACGAAACACATCTAACCTAACCTCTAACATCTGCA
 -604 TATAGTGCTTATAGTCGAATCATCTTATAACTTTATACGAATTTTGCTATTGATTGCCCACTTGTCTG**AGATAA**AAAAATTAATTTATTTTAAATTCAGAA
 -504 TTTTGATATTGATCATTCTAAACAAACATCAAAACACGAACCTTTCGCATTTTATAAATTTAATTAAGCAACACCG**GGAATCTC**CGCCGCTCTTTTC
 -404 TTATGAGAACTGGTTTCTGAACCTGGCGGTAGAGTTAATATTGTGCGAATCTAAATAATGTATAACATTTTACAGGTTGAATATATTTCATTCATAAATTT
 -304 **AATTTCAAAATTCaAAGaTtATTGcACTCC**CAACATCATTTTATACATAAGATGTACATTTTAAACGCGGTGACATAATTTGGACCGGTGTC**GGGcACGGa**
 -204 **AAAGTTTCaAGaGAGAGAACT**ATACAAAGTACCTTATAAGTATATTTAATTT**AGATAA**AAACCTCTGAGACGATACGTCATCTTAAATGGCGGTA
 -104 AACACGTAGG**TATAAAT**TTGCATCGACACGCTCTATCTGCAAACTACGAACCTTGTCCAT**AAAGATGac**CCAGCGAGCGGTAGACCGGTGGCACTGGC**AG**
 -4 **GTTCAGTGcgCgT**CAACCCGACGCTGAGGTAAACGCTATTGTGATTAAATGTAATGTAATACGAAAAAATGGCAAATTTAGTTTACGTGTTGTTA
 -1
 97 TTAGCTTCGTATTCTGCTTGTGTAGTGGACGTGAGTATTTCATCGTTTCCTCTAAACTATTTAAGTTTTCGCTAATTCATTGTGTTTCATACCTTA
 L A S Y F C F V S G (signal peptide)
 ---- 2.1 kb (+197 to +2296)
 2297 ACTATGCTTCCGAAATATAAGACATAAGACGAGTCTGCGAGCTC 2341 (4.32 kb) - - - - - Unknown gap - - - - -
 - - - Unknown gap - - - - - (4.51 kb) 2342 **GAATTC**CAATGAAACACGACCTGTTTTTTTTTTCATATGGCTCAAAAGTAAAAA
 ---- 1.0 kb (+2397 to +3396)
 3397 TTACACCCATGAATCTTGACTGATGAATGCTTTATAATATATGCTCATGATATTATTGGAACTTTCCAGAAATCATGCACGACGCCCAACACCGGTA
 Exon 2: Q S C + T T P N N G
 3497 AAGGCACGTGAAGTCGATATACGAGTGCAGAAAGCTTCTCAAACCTGGTGTACAAGAGGACAGGACGACGAGGACCGGATTTATTGAAGAGTCCCA
 K G T C + K S I Y E C + E E L L K L V Y K K D R T Q Q D T L K K S Q
 3597 GTGCGGCTTCATGGGCAACACCAACCGTAAGTACTAAACACATACCCCTTATTATAGACTTTTTCCTAAGGACAGAGCAAGAGCTGTGATAACAA
 C + G F M G N T P T
 ---- 2.5 kb (+3697 to +6196)
 6197 TTTCCAGCTAAACATTCACATGCATTATCAATTTCCAGGTTTCTGCGGCAACCTTGCATCAGCCACAGGGAGAGCCAGGCAATGCGTCAGCATT
 Exon 3: V C + C + P N P C + I T P Q G E P G Q C + V S I
 - - - - - clip domain 1 - - - - - || - - - - - clip domain 2 - - - - -
 6297 ACGAATGCACTAATCTCGCAATTTGCTGAAGCCACCAATACCGCGGATACGTATAAATATGTGCAGAAAGTCTAGGTGAGGTTAGTTTATACGGAC
 Y E C + T N L A N L L K P P I T A D T Y N Y V Q K S R
 ---- 1.2 kb (+6397 to +7596)
 7597 CTAATCATTGGACACACATTTTCAGATGCCAAGGGGACAGTCAAGTGTCTGCTGCGGATCAGCGCAAACTTCCCTTCCAGTGGCGACTGCAAGG
 Exon 4: C + Q G A D Q Y S V C + C + G S A P N F P S G D C K
 - - - - - clip domain 2 - - - - - || - - - - - linker region - - - - -
 7697 CTTCCGTTTCCGCTTCCCGCAGATCCCAAGTCCAGGTGCTGCGGTGAGACAGCAGAGTCGGGAACAAATCATTGGTGAATGATTAACCTTTCTA
 A S V S A F P P D P K S K C C G V D S R V G N K || I I
 - - - - - linker region - - - - - || - - - - - catalytic domain - - - - -
 ---- 0.3 kb (+7797 to +8096)
 8097 TTCTAAGCGATTGACGCCAAATCTTTTCAGGTGGAATGCGACAGAGTACGACCAATACCCATGGTTAACCATCATCGAGTACGTGAAACCGGCCCAT
 Exon 5: G G N A T D V D Q Y P W L T I I E Y V K T G P I
 8197 TAAGCTGTTGTGTGGTGGCTCTCATCAGTGGAAATACGTGCTGACAGCAGGACACTGTCTCACGGGACCGGTTTGGCAATIGGCACCGCTGAGTG
 K L L C G G V L I S G K Y V L T A G H C L T G P V L K I G T P
 ---- 0.3 kb (+8297 to +8596)
 8597 TCCATATACCCGAACACGAGCAGTATGTAACTTTCTTATACCTAAATTTGTCAGAAACAAACGCTCCGCTGGGAGAATACAATACGAAAAACGAC
 Exon 6: T N V R L G E Y N T K N D
 8697 GCGCGGACTCGCTGACGGTAGAGGCTGGGGGATGGACTGCACAGAGGGAGCGGTATCGTGCCCATAGAGAAGACAATACCGCACCGGAGTACAACC
 G A D C V T V E A G G M D C T E G A V I V P I E K T I P H P E Y N
 8797 CCATCAGCAGCAGGAGGAGGAGACGACATAGGACTCATCAGGCTGAAGGAGATGGCGCGTTTACTGGTATAGTTTGTATTAACTTTAATAGATT
 P I S R T R R N D# I G L I R L K E M A P F T
 ---- 0.6 kb (+8897 to +9496)
 9497 TTACCGTGATTATTATTATTAATAAATTTGCACTTTGCGAGATTTCATCCGTCCTCTGCTGCCAAGCTTGGACCTGACGCGAGGCTCCGCGGTGAAC
 Exon 7: D F I R P I C L P S L D L T Q A P P V N
 9597 TTCACGCTGTACGCGCGCGTGGGAGCTGTACAGACCTCACAGCCGACGAGCAACGTCAACTGCAGCTCCAGCTGCCGTTTCATCAGTTACGAGGTGA
 F T L Y A A G W G A V S T S Q P S S N V K L H V Q L P F I S Y E
 ---- 0.3 kb (+9697 to +9996)
 9997 CAATGAATAATCCATCCATTCTCGAGCGATGTCAGCGCTGACGCGGTACAGAACACAGACAGATCGAGTTGTGGGAGAAGCAAGTGTGCGCGGCGG
 Exon 8: R C Q P S Y A V Q N R Q I E L W E K Q V C A G
 10097 GCGAGGCTGGCAAGACTCTCGCAAGGGTGACTCGGAGGTCCCTCATGTTCGAAACCGGCGAGCTTACGAAGTATCGGCACTCTCTTTCGGACC
 G E A G K D S C K G D S# G G P L M F E N G Q T Y E V I G I V S F G P
 10197 GACCGCGTGTGAATGCAAGATATACCGCGGTGACACTAAAGTACACTCTTCAAAAGATTGGATAAATAGTAACATAAAGCCGTAGGTTTGTATAATG
 T P C G M Q D I P G V Y T K V H S Y K D W I I S N I K P *
 10297 TGGTCATTGTAATGAGATTACGTAACAACTATTTCAGTTGCAAGTGTCTTAATTAATCCATATTTACGTCGCTATAAGGTAAAAATCTTACAGGTCAAA
 CTATTATAGCAATAAATATTAATGAAGATAATTAATGATCAAGAAATAAAGTAAGTTATCGTTATGTTTACGGAATGTAG
 10497 ACATTGAAATAAACTATTTCGATTCAACAAAGTCTCGAAACGTCGGGAGAAAAATGTAAATACAAAAACCGGATTAATTCGCAAAAAAATAGTTTAT
 TCAATATAGGTAATACGACATAAAGGATGAACCTCTTAAATATAAATAGTGTAGTGTTTTATATTTTATACATTAAGTTTATTCAAAG
 10697 TAAACAAAAATCTTCTCACTATAGATAGCATTATACTCAATATACTCATTTTGTTTAAATGCTTACTAAAGTCAATTCGAATGTGTTTATATTT
 TACGCTATTATATGTCATAATTAATAATACAGTTTAAGTTATCTTCAAAGAAACACTTATTTAATCTAATCATGACTTAAATGGTGGTGTGCT
 10897 GATGTTAAACCCGCCATAGTGGCTCAGCTAGTGTGTCGCTCCAGGATCAGCTGTGTATATCCGTTCCAACTGGCGGGATTAATGTATCGACT
 10997 GTCGAAGGGCAATCATCTCTGCTCTAAGTTTAAATATAAATTTCAACTTTGTAGCTCCACGCGTTCTTTAAAAAATGGCTCTGACAGACAGAG
 11097 AGACTGACAAAGGTGATCCTGTAAGGGTTCCTTTTGTAGTACGAAATAATTTTGAACCTTATTAGCAATGATTAAGGTTAATTAATTTATCAATA
 11197 AAGTTACACTATTACGATATTTAAAAAATGTTGGGATGGGATGATCTCAATAGCCCTCTTGATAATACTATAATTTTATTAAGATTCTAAATAAG
 11297 ATTATGTGTAATAAACTATTTTATAAGGTGTTCTCTTTTATGATTGACTCAACTGGCCCGCTAGATGTGCTCTTCTCAACTATGAATATTC

Figure 2.

Nucleotide sequence and structural features of the *Manduca sexta* PAP-3 gene. Nucleotides in the 5' flanking region are assigned negative numbers. Nucleotide 1 is assigned based on the primer extension results (Fig. 3). Exon sequences are underlined with the encoded amino acid sequences listed below translated exons, using a one-letter code under the second nucleotide of each codon. While some regions of the intron sequences (marked '-') are not shown, their sizes and positions are indicated. GATA boxes (6-nucleotide) and ISRE sites (13-nucleotide), bold and double underlined; NF-κB motifs (10-nucleotide) and EcRE (15-nucleotide), bold and single underlined; TATA box (6-nucleotide), bold italic. The mismatches in these motifs are in lower case. Sequences of the 3' end of λ7 (4.32 kb), gap, and the 5' end of λ3 (4.51 kb)

are not shown. The *SacI* and *EcoRI* recognition sites (in bold) are numbered consecutively as 2336–2341 and 2342–2347, respectively. Single nucleotide polymorphic sites are in bold italic on the DNA sequence. Among them, nonsynonymous substitutions are further indicated on the affected amino acid residues (bold and underlined). The Cys residues in the clip domain and the catalytic residues in the SP domain are indicated with ‘+’ and ‘#’, respectively.

Figure 3.

Determination of the transcription initiation site in the *Manduca sexta* PAP-3 gene. A primer, complementary to nucleotides 28–56 of the PAP-3 coding region near the 3' end of exon 1, was terminally labelled with γ - ^{32}P -dATP and annealed to the total RNA from fat bodies from bacteria-injected larvae (15 μg). After annealing to RNA, the primer was extended with MMLV reverse transcriptase. The set of sequencing reactions (ACGT) on the left of the primer extension lane for use as a sizing ladder was from dideoxynucleotide sequencing of single-stranded M13 mp18 DNA using –40 primer. The arrow indicates the 125 bp extension product from fat body RNA isolated from the induced larvae.

Figure 4.

Southern blot analysis of *Manduca sexta* genomic DNA using ^{32}P -labelled *PAP-3* cDNA. Samples of the genomic DNA (15 μg) were digested with *Bam*HI (lane 1), *Hind*III (lane 2), *Sac*I (lane 3); *Xba*I (lane 4), *Eco*RI (lane 5), *Sac*I and *Xba*I (lane 6). After separation by agarose gel electrophoresis and transfer to a nitrocellulose membrane, the DNA fragments were hybridized with the full-length cDNA probe. The positions and sizes of the molecular markers are indicated on the right.

B

Figure 6.

Profiling of *PAP-3* gene transcription by RT-PCR. (A) Expression of the *PAP-3* gene in different tissues from day 3 fifth instar larvae or day 3 wandering larvae. N, nerve tissue; S, salivary gland; Mt, Malpighian tubule; T, trachea; Mg, midgut; H, haemocytes; F, fat body; I, integument; Mu, muscle; l, 438 bp, long PCR product with the 75 bp insert; s, 363 bp, short PCR product; l/s, a mixture of l and s. Two forms of the *PAP-3* cDNA clones (Jiang *et al.*, 2003b) were used as templates to amplify l and s. (B) changes of *PAP-3* mRNA level in fat bodies from *Manduca sexta* at different developmental stages. *M. sexta* ribosomal protein S3 (rpS3) transcripts were normalized in all these samples.

Figure 7.

Immunolocalization of PAP-3 protein in granular cells from the bacteria-injected larvae. The same microscopic field of fixed haemocytes was photographed using fluorescence microscopy. The haemocytes were reacted with a polyclonal antiserum to PAP-3 and FITC-labelled secondary antibodies. (A) Haemocytes from uninjected control larvae. (B) Haemocytes collected at 24 h after injection of *Micrococcus luteus*. Scale bar = 50 μ m. The bright spots in the granular haemocytes represent PAP-3 synthesized and stored after the immune challenge.

Table 1.
Sequence analysis of the 5' flanking region of the *Manduca sexta* PAP-3 gene

Motif name and consensus*	Sequence found**	Strand	Location	Matched
NF-κB: GGGRAYYYYY	GGGGACTTCT	–	–1765	10/10
	GGcAATTCCC	+	–1018	9/10
	GcGGAgTCTC	+	–922	8/10
	aGGAATTCaT	–	–760	8/10
	GGtAACTTgT	–	–694	8/10
	GaGGAaTCTT	+	–683	8/10
	GGaGATTCTC	+	–427	9/10
	tGaAACTTTT	–	–196	8/10
	GGtcACTCTT	–	–35	8/10
	TGATAA	+	–1713	6/6
GATA: WGATAA	AGATAA	+	–1137	6/6
	TGATAA	+	–536	6/6
	AGATAA	+	–145	6/6
ISRE: GGAAANNGAAANN	GGtAAAAaAAATG	+	–1724	11/13
	GtAAACTGAtACA	–	–944	11/13
	tGAAtTTGAAATT	–	–292	11/13
	GGgcACGGAAAAG	+	–213	11/13
	aGAAGGAGAAACT	+	–193	11/13
EcRE: RRGKTCANTGAMCYY	AGtTTgAATGcACTT	+	–804	12/15
	AAGaTtATTGcACTC	+	–291	12/15
	AGGTTcAGTGcgCgT	+	–6	12/15

* R: A,G; Y: C,T; W: A,T; K: G,T; M: A,C. W: A,T; N: A,C,G,T.

** the nucleotides in lower case do not match with the consensus.

ISRE, interferon-stimulated response elements; EcRE, ecdysone response element.