Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 12220–12225, October 1998
Biochemistry

Pro-phenol oxidase activating proteinase from an insect, *Manduca sexta:* A bacteria-inducible protein similar to *Drosophila* easter

HAOBO JIANG, YANG WANG, AND MICHAEL R. KANOST*

Department of Biochemistry, Kansas State University, Manhattan, KS 66506

Communicated by John H. Law, University of Arizona, Tucson, AZ, August 13, 1998 (received for review June 16, 1998)

ABSTRACT **Activation of pro-phenol oxidase (proPO) in insects and crustaceans is important in defense against wounding and infection. The proPO zymogen is activated by a specific proteolytic cleavage. PO oxidizes phenolic compounds to produce quinones, which may help to kill pathogens and can also be used for synthesis of melanin to seal wounds and encapsulate parasites. We have isolated from the tobacco hornworm, *Manduca sexta*, a serine proteinase that activates proPO, and have cloned its cDNA. The isolated proPO activating proteinase (PAP) hydrolyzed artificial substrates but required other protein factors for proPO activation, suggesting that proPO-activating enzyme may exist as a protein complex, one component of which is PAP. PAP (44 kDa) is composed of two disulfide-linked polypeptide chains (31 kDa and 13 kDa). A cDNA for PAP was isolated from a hemocyte library, by using a PCR-generated probe based on the amino-terminal amino acid sequence of the 31-kDa catalytic domain. PAP belongs to a family of arthropod serine proteinases containing a carboxyl-terminal proteinase domain and an amino-terminal "clip" domain. The member of this family most similar in sequence to PAP is the product of the *easter* gene from *Drosophila melanogaster*. PAP mRNA was present at a low level in larval hemocytes and fat body, but became much more abundant in fat body after insects were injected with *Escherichia coli*. Sequence data and ^{3}H-diisopropyl fluorphosphate labeling results suggest that the same PAP exists in hemolymph and cuticle.**

Phenol oxidase (PO) (EC1.14.18.1) is important in cuticular sclerotization and in defense against pathogens and parasites by insects and crustaceans (1–3). PO catalyzes hydroxylation of monophenols to *o*-diphenols and the oxidation of *o*-diphenols to quinones. The quinones take part in sclerotization and tanning of the cuticle and serve as precursors for synthesis of melanin. PO is synthesized in arthropods as an inactive zymogen, pro-phenol oxidase (proPO), which can be activated by proteolytic cleavage at a specific site near the protein's amino terminus (1–3). Wounding of an insect or exposure to certain microbial polysaccharides leads to activation of a protease that specifically activates proPO, perhaps the last step of a protease cascade (3). The reactions catalyzed by PO then can result in melanization of foreign organisms trapped in capsules or hemocyte nodules (4). Melanization may be an important factor in determining the outcome of an infection. A strain of the mosquito, *Anopheles gambiae*, that is refractory to infection by *Plasmodium cynomolgi* encapsulates the parasite ookinetes in a melanin coat, whereas susceptible strains do not (5, 6). To understand how such melanization is regulated, more information is needed about the biochemical activation of proPO.

ProPO activating enzymes have been isolated from cuticle of *Bombyx mori* (7, 8), a crayfish hemocyte lysate (9), and *Drosophila melanogaster* pupae (10), and from plasma of *Holotrichia diomphalia* (11). All of these enzymes are serine proteinases with masses of ≈30 kDa with specificity for cleavage after arginine residues. The proteinase from *H. diomphalia* required an additional plasma protein that lacked detectable proteinase activity for activation of proPO. Likewise, a proteinase preparation from hemolymph of *Hyalophora cecropia* required another plasma protein factor for activation of proPO (12). Other than the amino terminal sequences of the *H. diomphalia* proteinase (11), no sequence information or molecular cloning of a PAP has been reported, and little is known about how PAP becomes activated.

Two serine proteinase cascades in arthropods have been well characterized. The horseshoe crab clotting pathway has been investigated biochemically (13), and the pathway for determining embryonic dorsal-ventral polarity in *D. melanogaster* has been studied genetically (14). Both of these cascades contain serine proteinases with a domain architecture including a carboxyl-terminal proteinase domain and a small amino-terminal "clip" domain (13, 15). The clip domain may have a regulatory function or may modulate interactions of the proteinases with their substrates or other proteins. We recently identified two proteinases expressed in hemocytes of the tobacco hornworm, *Manduca sexta*, which also contain clip domains (16).

To further the understanding of the biochemistry of the proPO activating system, we have partially purified a proPO-activating proteinase (PAP) from *M. sexta* and used the amino-terminal sequence of the protein to design a strategy for cloning its cDNA. We report here that *M. sexta* PAP is a serine proteinase with an amino-terminal clip domain and that its expression is induced by injection of bacteria.

MATERIALS AND METHODS

Insects and Tissue Collection. *M. sexta* eggs were originally obtained from Carolina Biological Supply. Larvae were reared on an artificial diet as described by Dunn and Drake (17). Hemolymph was collected from chilled day 3 fifth instar larvae, and cell-free hemolymph (plasma) was prepared as described (18). Integuments from pharate pupae at 0–12 h before the brown bar stage were used for preparation of PAP. Insects were dissected to remove hemolymph, gut, and fat body. The integuments (and attached muscles) were washed

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ProPO, pro-phenol oxidase; PAP, pro-PO activating proteinase; DFP, diisopropyl fluorophosphate; IEAR*p*Na, acetyl-Ile-Glu-Ala-Arg-*p*-nitroanilide.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AF059728).

*To whom reprint requests should be addressed. e-mail: kanost@ksu.edu.

twice in chilled buffer A (0.1 M Tris·HCl, pH 7.5/0.5 M NaCl/0.001% 1-phenyl-2-thiourea/0.5 mM glutathione/0.2 mM *p*-aminobenzamidine/0.05% Triton X-100) and stored at −70°C.

Isolation of Cuticular PAP. Procedures for extraction and purification of PAP were carried out at 4°C. Frozen integuments from approximately 100 pharate pupae were homogenized in a Waring blender in 700 ml of buffer A. After removing debris by centrifugation (20,000 × *g*, 30 min), PAP in the supernatant (≈600 ml) was precipitated with ammonium sulfate at 40% saturation. The precipitate collected by centrifugation (20,000 × *g*, 20 min) was resuspended in 100 ml of buffer A containing ammonium sulfate at 10% saturation and then centrifuged (245,000 × *g*, 18 h) to remove insoluble material. The ammonium sulfate concentration in the clarified solution was slowly adjusted to 35% saturation. After centrifugation (20,000 × *g*, 20 min), the pellet was dissolved in 50 ml of buffer B (10 mM Tris·HCl, pH 7.5/0.001% 1-phenyl-2-thiourea/0.5 mM glutathione/0.2 mM *p*-aminobenzamidine) and dialyzed against three changes of same buffer (1.5 l for 8 h each).

After centrifugation (27,000 × *g*, 30 min), the solution (≈65 ml) was applied to a DEAE-cellulose column (Sigma, 45 × 2.5 cm) equilibrated in buffer B and washed with 1 l buffer B. Bound proteins were eluted at 2 ml/min with a linear gradient of 0–0.5 M KCl in buffer B (750 ml). Fractions that activated proPO were combined and precipitated with ammonium sulfate at 50% saturation. The precipitate collected by centrifugation was dissolved in 5 ml of buffer C (10 mM potassium phosphate, pH 7.2/0.2 M KCl) and applied to a Sephacryl S-300 column (Pharmacia, 180 × 1.5 cm) equilibrated with buffer C. The column was eluted at 0.25 ml/min with buffer B, and 5 ml fractions were collected. Active fractions were pooled and applied to a hydroxylapatite HT column (Bio-Rad, 10 × 2.5 cm), which was then washed with 100 ml of buffer C. The bound proteins were eluted with a linear gradient of 10–200 mM potassium phosphate (pH 7.2), 0.2 M KCl (160 ml) at 1.0 ml/min. Active fractions (about 38 ml) were combined and dialyzed against buffer D (10 mM potassium phosphate, pH 6.4). Approximately 20% of the dialyzed sample (8 ml) was applied to a cross-linked dextran sulfate column (Sigma, 6 × 1 cm) equilibrated with buffer D. After washing with 20 ml of buffer D, the column was eluted with a gradient of 0–0.6 M NaCl in buffer D (50 ml) at 1.0 ml/min, and 2 ml fractions were collected.

Measurement of proPO Activation and Amidase Activity. ProPO was purified from *M. sexta* larval hemolymph as described (19). Samples of column fractions (2–5 μl) were mixed with 0.1 μg of proPO, diluted to 20 μl with 10 mM Tris·HCl (pH 7.8), and incubated in microplate wells on ice for 60 min. PO activity was assayed by adding 200 μl of 2 mM dopamine in 50 mM sodium phosphate (pH 6.5) to each sample well. Absorbance at 450 nm was then monitored continuously using a microplate reader (Molecular Devices). One unit of PAP activity was defined as the amount of enzyme yielding active phenol oxidase producing an increase of 0.001 absorbance units/min. In the absence of added PAP, the purified proPO had no detectable activity. Amidase activity was assayed by using acetyl-Ile-Glu-Ala-Arg-*p*-nitroanilide (Sigma, A0180) as a substrate. Samples of column fractions (10–20 μl) were mixed with 200 μl of 25 μM peptidyl-*p*-nitroanilide substrate dissolved in 0.1 M Tris·HCl/0.1 M NaCl, pH 7.8. One unit of activity was defined as $\Delta A_{405\ nm}/min = 0.001$.

^{3}H-Diisopropyl Fluorophosphate (DFP) Labeling. Plasma samples (14 μl) or partially purified PAP (18 μl, 2 μg protein) were incubated with ^{3}H-DFP (2 μl, 0.1 mM, 10 mCi/mmol, New England Nuclear; 1 Ci = 37 GBq) at 37°C for 30 min. After addition of SDS sample buffer, the labeled proteins were separated by electrophoresis on a SDS/12% polyacrylamide gel (20). One lane was loaded with ^{14}C-labeled molecular weight standards (GIBCO BRL). After electrophoresis, the gel was treated for fluorography (21). Dried gels were then exposed to x-ray film for 3–7 days at −70°C.

Amino Acid Sequencing. Partially purified PAP (120 μl, 14 μg protein) eluted from the dextran sulfate column was precipitated with 10% trichloroacetic acid and collected by centrifugation. After several washes with anhydrous acetone, the pellet was treated with SDS-sample buffer containing 2-mercaptoethanol and separated on an SDS/10% polyacrylamide gel. The protein was then transferred to a polyvinylidene difluoride membrane and stained with amido black. The 31-kDa polypeptide corresponding to a ^{3}H-DFP-labeled band was subjected to automated Edman degradation by using an Applied Biosystems Model 473 pulse-liquid sequencer.

cDNA Isolation and Sequencing. Lambda phage DNA (0.1 μg) isolated from a *M. sexta* larval hemocyte cDNA library (22) was used as a template in a PCR to amplify a PAP cDNA fragment. Two degenerate primers were designed based on the amino-terminal sequence of cuticular PAP: F1 (5′-ATCTAY-GGIGGICARATHACN GAYYT-3′) and F2 (5′-GAYYTI-GAYGARTTYTGGATG-3′) encode IYGGQITDL and DL-DEFPWM, respectively. Primer R (5′-CATGAGSGGRC-CRCCSGARTCNCC-3′) is the reverse complement of the sequence encoding GDSGGPLM, which is a highly conserved sequence around the reactive site serine residue in serine proteinases of the chymotrypsin family. In a first PCR, primer F1 was used with primer T7, which anneals with a sequence in the cloning vector near the 3′ end of the inserted cDNA, under the following conditions: 94°C, 30 s; 50°C, 40 s; 72°C, and 80 s for 50 cycles. The reaction product (1 μl) was used directly as a template for a second, nested PCR by using primers F2 and R under the same thermal cycling conditions. After agarose gel electrophoresis of the resulting products, a DNA band of the expected size (≈600 bp) was recovered and cloned into the pGem-T vector (Promega). This PCR-derived cDNA clone was used to probe the *M. sexta* hemocyte cDNA library by hybridization (23). Positive clones isolated from the library were purified to homogeneity and subcloned by *in vivo* excision of pBluescript phagemids. The complete nucleotide sequence of the longest cDNA clone was determined from double-stranded plasmid DNA templates by the dideoxynucleotide method using modified T7 DNA polymerase (Amersham Life Science). Sequencing strategies included subcloning of restriction fragments and use of oligonucleotide primers designed from previously determined sequences.

Computer Analysis of Sequence Data. Preliminary sequence editing and analysis was performed using IBI PUSTELL programs. The amino acid sequence deduced from the cDNA was used to search the nonredundant peptide sequence database (National Center for Biotechnology Information) with the BLAST program (24). The most similar proteinase sequences were retrieved and aligned with the PAP sequence using the PILEUP program from the GCG Sequence Analysis Software Package 7.3.1 (25). Phylogenetic trees from such alignments were constructed using the CLUSTALW program (26). The position of the signal peptide cleavage site was predicted using the program of Nielsen *et al.* (27).

Northern Blot Analysis. Fat body and hemocyte total RNA samples were prepared according to Wang *et al.* (28). Samples of total RNA (10 μg) from these tissues were resolved by electrophoresis on an agarose gel containing formaldehyde, transferred to a nitrocellulose membrane, and hybridized with ^{32}P-labeled PAP cDNA. To confirm equal mRNA loading in different samples, a duplicate blot was hybridized with a cDNA for *M. sexta* ribosomal protein S3 (29) as a control.

RESULTS AND DISCUSSION

Purification and Properties of PAP. We chose integuments of pharate pupae as a source for isolating PAP, based on

reports that this tissue contains significant levels of the active enzyme in *B. mori* and *M. sexta* (7, 30). The integument extract contained high levels of PAP activity, which was detected by incubating the extract with purified proPO and then measuring the resulting PO activity. Activation of proPO coincided with the appearance of a slightly lower molecular weight PO detected by Western blot analysis (data not shown), indicating the participation of a proteinase in the activation process.

Proteins in a 10–35% ammonium sulfate fraction of the cuticular extract were separated by chromatography on DEAE-cellulose. The fractions that contained PAP activity eluted between 0.2–0.3 M KCl at pH 7.5. This differed from *B. mori* cuticular PAP, which does not bind to DEAE-cellulose even at pH 8.5 (7). PAP was recovered by ammonium sulfate precipitation and then resolved by gel permeation chromatography on Sephacryl S-300. ProPO was activated by fractions that eluted just after the void volume. Fractions from gel permeation that contained PAP activity were pooled and subjected to hydroxylapatite chromatography. PAP was contained in fractions that eluted between 50–100 mM potassium phosphate. At this point we had recovered ≈30% of the total activity with a 20-fold increase in specific activity.

A sample containing this enriched PAP was dialyzed and then separated on a column of cross-linked dextran sulfate. To detect the amidase activity of PAP in these fractions, we used as a substrate acetyl-Ile-Glu-Ala-Arg (IEAR)-*p*-nitroanilide, which is similar to the amino-terminal side of the putative activation site in *M. sexta* proPO-p1 and proPO-p2 (Leu-Asn-Asn-Arg) (19). Very low IEARase activity was present in the flow-through fractions (4–10), which contained most of the protein, whereas a large peak of IEARase activity was detected in fractions 25–28 (Fig. 1*A*).

In an assay for proPO activation, the flow-through peak (fraction 8) had a small amount of PAP activity, and the IEARase activity peak eluted by the salt gradient (fraction 26) had very little PAP activity (Fig. 1*B*). The total PAP activity recovered in fractions 4–40 accounted for only 5% of that applied to the column. One possible explanation for the activity loss is that the activation of proPO may require two or more components, which were separated by the dextran sulfate chromatography. To test this possibility, aliquots of the column fractions were combined to test for restored activity. A mixture of fractions 8 and 26 activated proPO at a level much higher than the sum of the individual components (Fig. 1*B*) and had ≈3-fold greater IEARase activity (data not shown). We estimate from these data that at least 50% of the total activity loaded could be recovered by combining the flow-through and bound fractions.

These results indicate that the dextran sulfate chromatography separates a prophenoloxidase activating complex into two components, a proteinase in the later peak and a factor in the flow-through peak that alters the activity of the proteinase, making it competent to activate proPO. The elution of the PAP activity near the void volume in Sephacryl S-300 chromatography supports a hypothesis that PAP is part of a high molecular weight species that may be a protein complex. Treatment of fraction 8 with trypsin or by boiling eliminated its ability to increase the PAP activity of fraction 26, indicating that the modulating factor is proteinacious. Serpin-1J, a serine proteinase inhibitor from *M. sexta* plasma that specifically inhibits proPO activation (18), inhibits the IEARase in fraction 26, and inhibits proPO activation by the combined fractions 26 and 8 (data not shown), which is consistent with the suggestion that the enzyme with IEARase activity is PAP. The small peak of PAP activity in the flow through peak may be due to incomplete binding of PAP to the dextran sulfate, leading to some residual activity in the unbound fraction. These results are similar to previous reports in which serine proteinases isolated from plasma of *H. cecropia* (12) and *H. diomphalia* (11) required an additional plasma protein factor before they could activate proPO.

When samples from fractions 25–28 were reacted with ^{3}H-DFP to label the active site serine residue of serine proteinases, a single radioactively labeled band was detected in fractions 25–28 by SDS/PAGE and fluorography. In the absence of a reducing agent this band had a mobility corresponding to 44 kDa, but under reducing conditions a 31-kDa band was detected (Fig. 2). These results suggest that these fractions contain one detectable serine enzyme that is composed of two disulfide-linked chains. A 31-kDa subunit contains the catalytic site, which is separated from a 13-kDa subunit after treatment with 2-mercaptoethanol. A 31-kDa band was detected as a major protein component of this fraction (≈20% of the total protein) after staining with Coomassie blue. This protein was transferred to a membrane and subjected to Edman degradation. The first 20 residues of the 31-kDa polypeptide were Ile-Tyr-Gly-Gly-Gln-Ile-Thr-Asp-Leu-Asp-Glu-Phe-Pro-Trp-Met-Ala-Leu-Leu-Gly-Tyr

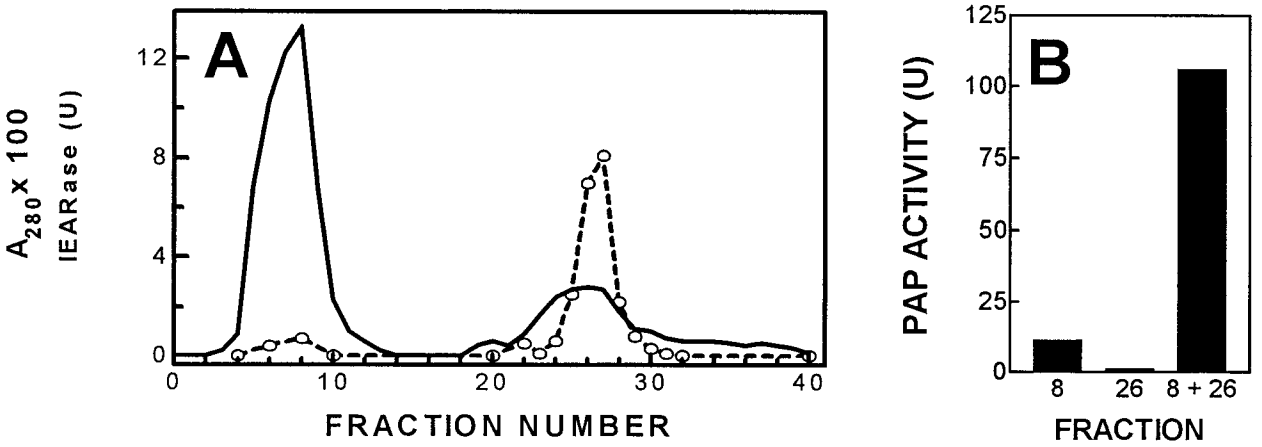

FIG. 1. Activation of proPO by fractions from cross-linked dextran sulfate chromatography of cuticular extract. (*A*) Chromatography of partially purified PAP. After separation of a cuticular extract by ammonium sulfate precipitation, DEAE cellulose chromatography, and Sephacryl S-300 gel permeation chromatography, the fractions containing PAP activity were pooled (8 ml) and loaded onto a dextran sulfate column. After washing (flow-through fractions 1–14), bound proteins were eluted by a linear NaCl gradient, beginning at fraction 15. (—) Absorbance at 280 nm; (○) amidase activity in 20 μl of the fractions assayed by using IEAR-*p*-nitroanilide as a substrate. (*B*) Activation of proPO by column fractions. *M. sexta* proPO (0.1 μg) was incubated with a sample from the flow-through peak (no. 8, 2 μl), a sample from the IEARase activity peak (no. 26, 2 μl), or a mixture of these two samples (no. 8 + no. 26, 2 μl + 2 μl) in 10 mM Tris·HCl (pH 7.8) on ice for 60 min. PO activated during these incubations was then assayed, using dopamine as a substrate.

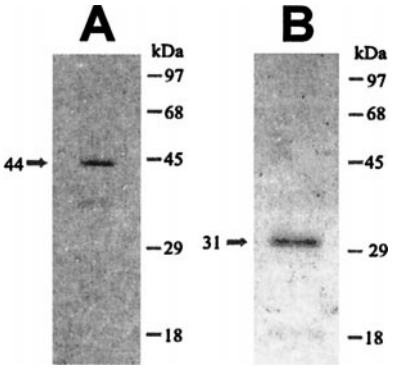

FIG. 2. Labeling and detection of a serine proteinase in a PAP fraction eluted from a cross-linked dextran sulfate column. A fraction containing the highest level of IEAR-*p*-nitroanilide hydrolyzing activity (corresponding to fraction 26 in Fig. 1) was incubated with ^{3}H-DFP to label the active site serine residue of PAP. The reaction mixture was then treated with SDS-sample buffer in the absence of reducing agent (*A*) or including 2-mercaptoethanol (*B*) and subjected to SDS/PAGE. Radioactively labeled proteins were detected by fluorography. The positions of ^{14}C-labeled molecular weight standards are marked on the right.

(Fig. 4). A sequence database search showed that this sequence is very similar to a region at the beginning of a serine proteinase domain in easter from *D. melanogaster* (GenBank accession no. P13582) and protein Sp14D from *A. gambiae* (GenBank accession no. AF007166). Both of these proteins contain an additional amino-terminal clip domain. Proteolytic cleavage of easter or Sp14D at their putative activation site would yield two polypeptides consisting of the clip domain (≈12 kDa) and the serine proteinase (≈31 kDa). This similarity suggested that PAP might have the same type of domain architecture.

To examine whether we could detect activation of PAP in plasma during activation of ProPO by bacteria, we used ^{3}H-DFP to label serine enzymes in larval plasma samples treated for different periods of time with *Micrococcus lysodeikticus*. As observed previously (16), at least five radioactive bands were detected, including a band at 31 kDa (Fig. 3). The intensity of the 31-kDa band increased significantly during the exposure to bacteria, suggesting that the proteinase was activated by this treatment, which leads to activation of proPO (18).

cDNA and Deduced Protein Sequence. We used the amino terminal sequence of the 31-kDa PAP polypeptide to design two primers that were used with a primer that anneals to the vector in a hemocyte cDNA library and a primer based on a conserved sequence around the active site serine in a nested PCR strategy (see *Materials and Methods*) to amplify a PAP cDNA fragment. Using DNA isolated from the hemocyte cDNA library as template, we obtained a 0.6-kb product and cloned it into a plasmid vector. This PCR product contained at its 5′ end a sequence encoding the amino terminus of the 31-kDa polypeptide. We used the 0.6-kb cDNA fragment to screen the hemocyte cDNA library to obtain a full-length clone. From 2×10^5 plaques screened, seven positive plaques were isolated and purified to homogeneity.

The longest PAP cDNA (1513 nucleotides) contained an ORF spanning nucleotides 106–1254 (Fig. 4). The deduced amino acid sequence of PAP contains 383 amino acid residues, including a predicted 19-residue secretion signal peptide. The molecular mass of the mature protein (proPAP) beginning at residue 20 is 39,532 Da, which is smaller than the 44-kDa DFP-labeled band detected in SDS/PAGE without reduction of disulfide bonds (Fig. 2). This apparent discrepancy might be due to posttranslational modifications such as glycosylation. However, the sequence does not contain an N-linked glycosylation site. The calculated isoelectric point of proPAP is 5.95.

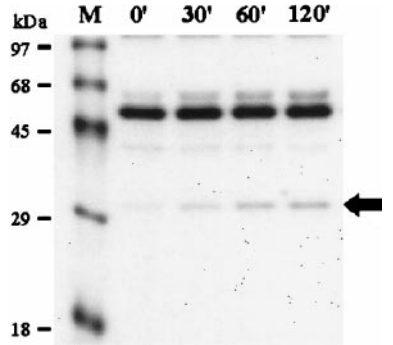

FIG. 3. Detection of ^{3}H-DFP-labeled plasma proteins after treatment with *Micrococcus lysodeikticus*. Larval plasma samples (14 μl) were preincubated with *M. lysodeikticus* (1 μl, 1 μg/μl) and 0.1 M glutathione (2 μl) for 0, 30, 60, and 120 min, then allowed to react with ^{3}H-DFP for 30 min. After separation by SDS/PAGE, labeled proteins were detected by fluorography. Lane M contains ^{14}C-protein molecular mass markers.

```
AGCGAGTCCAGCGCTCTTTAATTTATTTATATAAAAATATTAAAAAAAATATTGCATAGAA   60
AAGTTGTTATTTATTAAGTTATGATAAAGTGTTAAAGTGAACAAAATGAGGAATCACATT  120
                                          M  R  N  H  I        5
GTGTTCATTGTGTTTGCAGTTTATTGGACATGTGTCTTTTCACAGTCATGTACAACACCA  180
 V  F  I  V  F  A  V  Y  W  T  C  V  F  S  Q  S  C  T  T  P   25
CAAGGGGCCGATAGTAACTGTATCTCCCTTTACGAGTGCCCTCAATTACTGTCGGCTTTT  240
 Q  G  A  D  S  N  C  I  S  L  Y  E  C  P  Q  L  L  S  A  F   45
GAGCAACGTCCTCTGCCTAGCCCCGTTGTAAACTACTTGCGCAAGTCACAGTGTGGTTTC  300
 E  Q  R  P  L  P  S  P  V  V  N  Y  L  R  K  S  Q  C  G  F   65
GATGGATATACTCCTCGTGTGTGCTGCGGACCCTTACCTCAACAGGCCAGCAGGCCGCAA  360
 D  G  Y  T  P  R  V  C  C  G  P  L  P  Q  Q  A  S  R  P  Q   85
CCCACGCCGGCCCCTGTACCAACACGCGCTCCACCTGTAAACCCGGGTGGTGTAGACCCA  420
 P  T  P  A  P  V  P  T  R  A  P  P  V  N  P  G  G  V  D  P  105
ACTTATGATGAAGATTCATCACCAGCACCCAGGAATCAGTGCGGTGTGGACATGAATGGA  480
 T  Y  D  E  D  S  S  P  A  P  R  N  Q  C  G  V  D  M  N  G  125
GACCGCATATACGGAGGACAAATCACAGACCTCGACGAGTTTCCATGGATGGCGTTGCTT  540
 D  R ‖I  Y  G  G  Q  I  T  D  L  D  E  F  P  W  M  A  L  L  145
GGATATTTGACTCGTACTGGCTCTACTACGTATCAATGCGGAGGTGTATTGATCAACCAG  600
 G  Y  L  T  R  T  G  S  T  T  Y  Q  C  G  G  V  L  I  N  Q  165
CGATATGTATTGACCGCCGCTCATTGTACTATTGGCGCTGTGGAGAGAGAGGTCGGGAAG  660
 R  Y  V  L  T  A  A ✱H  C  T  I  G  A  V  E  R  E  V  G  K  185
CTAATAACAGTGCGACTTGGAGAGTACGACACCCAAAACTCTGTGGACTGCGTGGACGAT  720
 L  I  T  V  R  L  G  E  Y  D  T  Q  N  S  V  D  C  V  D  D  205
GTTTGTGCTGACCCTCCACAAAACATTCCAATTGAAGCAGCCTACCCTCATTCAGGGTAC  780
 V  C  A  D  P  P  Q  N  I  P  I  E  A  A  Y  P  H  S  G  Y  225
AGTGACAACAACAAGAACAGGAAAGATGACATTGCCCTTGTTAGGCTTACGAGACGAGCT  840
 S  D  N  N  K  N  R  K  D ✱D  I  A  L  V  R  L  T  R  R  A  245
CAGTACACATATTATGTTAAACCTATATGTCTGGCCCACAACAACGAGCGGCTAGCTACT  900
 Q  Y  T  Y  Y  V  K  P  I  C  L  A  H  N  N  E  R  L  A  T  265
GGCAATGACGTTTTCGTGGCCGGTTGGGGAAAAACCTTGAGCGGCAAGAGCAGCCCAATT  960
 G  N  D  V  F  V  A  G  W  G  K  T  L  S  G  K  S  S  P  I  285
AAATTGAAGTTAGGTATGCCAATTTTCGACAAAAGCGACTGCGCATCAAAATACAGGAAT 1020
 K  L  K  L  G  M  P  I  F  D  K  S  D  C  A  S  K  Y  R  N  305
TTAGGAGCAGAATTGACTGACAAACAGATATGCGCTGGTGGAGTATTCGCCAAAGATACT 1080
 L  G  A  E  L  T  D  K  Q  I  C  A  G  G  V  F  A  K ➡D  T  325
TGCAGAGGTGACTCCGGCGGTCCTCTCATGCAGAGAAGACCAGAGGGAATTTGGGAGGTC 1140
 C  R  G  D ✱S  G  G  P  L  M  Q  R  R  P  E  G  I  W  E  V  345
GTAGGCATAGTATCCTTTGGCAATAGGTGTGGACTCGATGGATGGCCAGGCGTTTACTCT 1200
 V  G  I  V  S  F  G  N  R  C  G  L  D  G  W  P  G  V  Y  S  365
TCGGTTGCGGGCTATTCTGACTGGATTCTATCTACACTCAGATCAACCAATGTGTGAATA 1260
 S  V  A  G  Y  S  D  W  I  L  S  T  L  R  S  T  N  V  *     383
GTGTTAGACATTTGTTATTCTTCTTGTGTTGTGTCGTAAGTAGATTAATTAATGGTTGCT 1320
CTCTGACACGCGATAACTCAGTGATTGTGATAATCTTTGACGCTTTCTATGTTTACATGT 1380
TAGTGTTACGTTCTGTCTACTAGTTTGTTTTGTAAATAATTGAACTGATATGTAACTGCT 1440
CAATTACATCTTTCATTATTTGTATTAAAAAATGTTATTTTTAATCAAAAAAAAAAAAAA 1500
AAAAAAAAAAAA                                                  1512
```

FIG. 4. Nucleotide and deduced amino acid sequence of a PAP cDNA. Amino acid residues, shown in one letter code, are aligned with the second nucleotide of each codon. A putative secretion signal peptide is underlined. The proteolytic activation site is indicated with "‖." The double-underlined sequence after the activation site was confirmed by Edman degradation sequencing of the 31-kDa polypeptide isolated from the cuticular extract. The residues of the serine proteinase active site triad are indicated with "✱." An Asp residue predicted to be at the bottom of the substrate binding pocket is labeled with an arrow.

The predicted site for proteolytic activation of proPAP is between Arg-127 and Ile-128. This was confirmed by the finding that the sequence from Ile-128 to Tyr-147 is identical to the amino-terminal sequence of the 31-kDa chain of the activated enzyme (Fig. 4). Cleavage of proPAP after Arg-127 would yield an amino-terminal chain consisting of residues 20–127, with a mass of 11,560 Da and a carboxyl-terminal chain (residues 128–383; 27,990 Da), which has a sequence typical of a serine proteinase from the chymotrypsin family, including the conserved His-Asp-Ser catalytic triad. PAP contains an Asp residue at a position predicted to be at the bottom of the substrate binding pocket (Fig. 4), which suggests that it has specificity for cleavage after basic residues. This is consistent with its role in activation of proPO by cleavage after a specific Arg residue.

Sequence Comparisons. In searches of amino acid sequence databases, PAP was found to be most similar to a family of arthropod proteins containing a carboxyl-terminal serine proteinase domain and an amino-terminal clip domain. The clip domain was first identified in the horseshoe crab (*Tachypleus tridentatus*) proclotting enzyme (31). An alignment of the sequence of PAP with the most similar members of this family is shown in Fig. 5, and an evolutionary tree showing the relationship of PAP to all of the known members of this family is shown in Fig. 6. The clip domain extends from the amino terminus of the mature protein to the activation site, which is the beginning of the proteinase domain. These sequences contain 14 conserved Cys residues. The topology of the

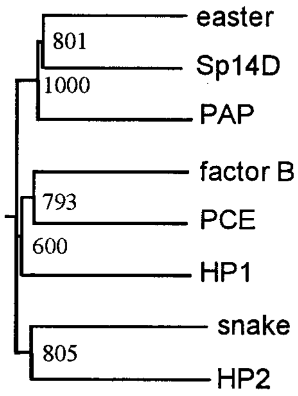

FIG. 6. Phylogenetic tree of arthropod serine proteinases containing a clip domain. The tree is based on an alignment similar to that shown in Fig. 5, but also including *Drosophila* snake (p05049), *M. sexta* hemocyte proteinase 2 (HP2) (AF017664), and horseshoe crab clotting factor B (D14701). Numbers at nodes indicate bootstrap values (from 1,000 trials).

```
                                                                                         80
PAP                                                  QSCTTPQGADSNCISLYECPQLLSAFEQRPLPSPVVNYL
SP14D                                      LAGQTVLALELGQCVNPVGEAGKCVLFRECQPLVDIYNKPVNTPDDTQFL
EASTER                                    FYFPNEAAQVPNYGRCITPNRERALCIHLEDCKYLYGLLTTTPLRDTDRLYL
HP1      DYSELQHPAWEDIIEEGGMHIGGGRSKRFIQLNPNQGNVAYQCTLPSGKQGHCRHLRFCIQE       DFKQDFVKF
PCE                                               QFVFPDEEECSNRFTEEGTCKNVLDCRIL       LQKNDYNLL

                                                                      *      *    *       160
PAP      RKSQCGFDGYTPRVCCGPLPQQASRPQPTPAPVPTRAPPVNPGGVDPTYDEDSSPAPRNQCGV DMNGDR||IYGGQITDLD
SP14D    TESRCGLYERKTLVCCAG  VRSKGKTSLPE             SPN                CGV QLT DR||VLGGQPTKID
EASTER   SRSQCGYTNGKVLICCPDRYRESSSETTPPP             KPNVTSNSLLPLPGQCGN ILS NR||IYGGMKTKID
HP1      MNYVCVIQGQSMGVCCP     EDTTVGGPEGLAGDLPATAPQEGEDETLLKINQAQNRGCGLSTRAQGR||VFGSRPANPR
PCE      KESICGFEGITPKVCCP     KSSHV       ISSTQAPPETTTTERPPKQIPPNLPEVCGIHNTTTTR||IIGGREAPIG

           **        *     **     *   **                   *****        +          +     240
PAP      EFPWMALLGYLTRTGSTTYQCGGVLINQRYVLTAA H CTIGAVEREVGK  LITVRLGEYDTQNSVDCVDDV      CADP
SP14D    EFPWTALIEYEKPNGRFGFHCGGSVINERYILTAA H CIT  SIPRGWK  VHRVRLGEWDLSSTTDQEDDF      YADA
EASTER   EFPWMALIEYTKSQGKKGHHCGGSLISTRYVITAS H CVNGKALPTDWR  LSGVRLGEWDTNTNPDCEVDVRGMKDCAPP
HP1      EWPWMAS   ITPEGFEQ YCGGVLITDRHVLTAA H C       TRRWEANELYVRLGEYDFKRT NDTR
PCE      AWPWMTAV YIKQGGIRSVQCGGALVTNRHVITAS H CVVNSAGTDVMPADVFSVRLGEHNLYSTDDDSN

                  *                   **                 * *                 ***      *  320
PAP      PQNIPIEAAYPHSGYSDNNKNRKD D IALVRLTRRAQYTYYVKPICLAHNN   ERLATGNDVFVAGWGKT LSGKSSPIK
SP14D    PIDLDIEKIIVHPGYNLQDKSHHN D IALIRFNREINYSSTIRAICLPLSNSLRNRKHAGLSSYAAGWGKT ETASASQKK
EASTER   HLDVPVERTIPHPDYIPASKNQVN D IALLRLAQQVEYTDFVRPICLPLDVNLRSATFDGITMDVAGWGKT EQLSASNLK
HP1      SYNFRVVEKVQHVDFEIS   NYHH D IAILKLDKPAIFNTYVWPICLPPP GLSIENETVT   VIGWGTQWYGGPHSHVL
PCE      PIDFAVTSVKHHEHFVLA   TYLN D IAILTLNDTVTFTDRIRPICLPYR KLRYDDLAMRKPFITGWGTTAFNGPSSAVL

                                        **      *    **  ****              * ** *     *    * 400
PAP      LKLGMPIFDKSDCASKYRNLGAELTDKQICAGGVF AKDTCRGD S GGPLM    QRRPEGIWEVVGIVSFG NRCGLDGWP
SP14D    LKVELTVVDVKDCSPVYQRNGISLDSTQMCAGGVR GKDTCSGD S GGPLM     RQMTGSWYLIGVVSFGPQKCGAPGVP
EASTER   LKAAVEGSRMDECQNVYSSQDILLEDTQMCAGGKE GVDSCRGD S GGPLIGLDTNKVNTYYFLAGVVSFGPTPCGLAGWP
HP1      MEVSFPIWTHQNCIEVHTN    SIFDESICAGGHEGGRDACQGD S GGPLMYQM    PSGRWAVVGIVSWG VRCGEPNHP
PCE      REVQLPIWEHEACRQAYEK DLNITNVYMCAGFADGGKDACQGD S GGPMMLPV    KTGEFYLIGIVSFG KKCALPGFP

          * *  *     **          422
PAP      GVYSSVAGYSDWILSTLRSTNV
SP14D    GVYTNVAEYVDWIKDNIY
EASTER   GVYTLVGKYVDWIQNTIES
HP1      GIYTRVDKYIGWIMENARF
PCE      GVYTKVTEFLDWIAEHMV
```

FIG. 5. Sequence alignment of *M. sexta* PAP with other arthropod serine proteinases that have a clip domain. The catalytic triad residues are boxed. Residues conserved in all of the sequences are labeled with an asterisk. A vertical double line marks the putative activation cleavage sites. Conserved cysteines, whose disulfide bonding pattern has been determined in PCE (31), are connected with single lines. A pair of cysteines present only in PAP and easter is labeled with +. *Anopheles gambiae* SP14D (AF007166), *Drosophila* easter (P13582), *M. sexta* hemocyte proteinase 1 (HP1) (AF017663), and horseshoe crab proclotting enzyme (PCE) (P21902).

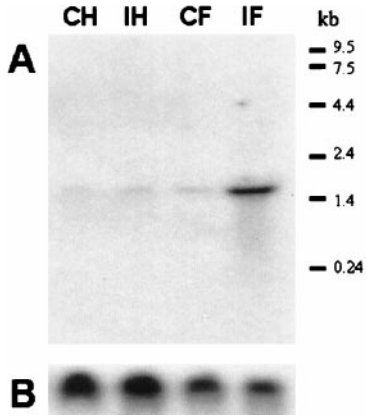

FIG. 7. Northern blot analysis of PAP mRNA level. RNA samples (10 μg) from hemocytes of control insects (CH), hemocytes collected 24 h after injection of *E. coli* (IH), fat body from control insects (CF), or fat body collected 24 h after injection of *E. coli* (IF) were separated by electrophoresis on a 1% agarose gel containing formaldehyde. The RNA was transferred to a nitrocellulose membrane and then hybridized with $_{32}$P-labeled *M. sexta* PAP cDNA (*A*). The positions of RNA standards are shown at the side of each blot. Approximately equal RNA loading of each lane was confirmed by probing a duplicate blot with ^{32}P-labeled cDNA for *M. sexta* ribosomal protein S3 (*B*) (29).

disulfide bonds formed by these residues in proclotting enzyme has been determined (31). There are three disulfide bridges within the serine proteinase domain and three within the clip domain. A seventh disulfide bond links the clip domain to the serine proteinase domain, such that the two domains remain covalently bound after cleavage of the peptide bond at the activation site. This is consistent with the difference in observed size of ^{3}H-DFP-labeled PAP with and without reduction of disulfides (Fig. 2).

PAP is most similar in sequence to easter (43% identity; 49% similarity), which is the last proteinase in a cascade that functions in determining dorsal-ventral polarity of the *D. melanogaster* embryo (14) and Sp14D (GenBank accession no. AF007166)(42% identity, 49% similarity) from *A. gambiae*. The function of Sp14D has not yet been reported, but this similarity suggests that Sp14D may be a mosquito PAP. We do not know whether *M. sexta* PAP is orthologous or paralogous to *D. melanogaster* easter. PAP and easter contain an extra pair of Cys residues in the proteinase domain, not present in the other sequences (Fig. 5).

PAP Gene Expression. In northern blot analysis, the PAP cDNA hybridized to a 1.5-kb band in RNA samples from fat body and hemocytes (Fig. 7). The level of PAP mRNA increased significantly in fat body and hemocytes after larvae were injected with *E. coli*. This result indicates that greater capacity to activate proPO may be a defensive response to infection. Alternatively, increased PAP gene expression may be required to regenerate proPAP to normal levels in plasma after the enzyme has been depleted through activation and subsequent inactivation by plasma proteinase inhibitors.

We thank Karl Kramer and Gerald Reeck for helpful comments on the manuscript. This work was supported by the National Institutes of Health Grant GM41247. Contribution number 98-482-J from the Kansas Agricultural Experiment Station.

1. Söderhäll, K., Cerenius, L. & Johansson, M. W. (1996) in *New Directions in Invertebrate Immunology*, eds. Söderhäll, K., Iwanaga, S. & Vasta, G. (SOS, Fair Haven, NJ), pp. 229–253.
2. Sugumaran, M. (1996) in *New Directions in Invertebrate Immunology*, eds. Söderhäll, K., Iwanaga, S. & Vasta, G. (SOS, Fair Haven, NJ), pp. 355–374.
3. Ashida, M. & Brey P. T. (1998) in *Molecular Mechanisms of Immune Responses in Insects*, eds. Brey, P. T.& Hultmark, D. (Chapman & Hall, New York), pp. 135–172.
4. Gillespie, J., Kanost, M. R. & Trenczek, T. (1997) *Annu. Rev. Entomol.* **42,** 611–643.
5. Collins, F., Sakal, R., Vernick, K., Paskewitz, S., Miller, D.C., Collins, W., Campbell, C. & Gwadz, R. (1986) *Science* **234,** 607–610.
6. Vernick, K. D. (1998) in *Molecular Mechanisms of Immune Responses in Insects*, eds. Brey, P. T. & Hultmark, D. (Chapman & Hall, New York), pp. 261–309.
7. Dohke, K. (1973) *Arch. Biochem. Biophys.* **157,** 210–221.
8. Ashida, M. & Dohke, K. (1980) *Insect Biochem.* **10,** 47–47.
9. Aspán A., Sturtevant, J. Smith, V. J. & Söderhäll, K. (1990) *Insect Biochem.* **20,** 709–718.
10. Chosa, N., Fukumitsu, T., Fujimoto, K. & Ohnishi, E. (1997) *Insect Biochem. Mol. Biol.* **27,** 61–68.
11. Lee, S. Y., Kwon, T. H., Hyun, J. H., Choi, J. S., Kawabata, S.-I., Iwanaga, S. & Lee, B. L. (1998) *Eur. J. Biochem.* **254,** 50–57.
12. Andersson, K., Sun, S., Boman, H. G. & Steiner, H. (1989) *Insect Biochem.* **19,** 629–637.
13. Kawabata, S., Muta, T. & Iwanaga, S. (1996) in *New Directions in Invertebrate Immunology*, eds. Söderhäll, K., Iwanaga, S. & Vasta, G. (SOS Publications, Fair Haven, NJ), pp. 255–283.
14. Belvin, M. & Anderson, K. (1996) *Annu. Rev. Cell. Dev. Biol.* **12,** 393–416.
15. Smith C. L. & DeLotto, R. (1992) *Protein Sci.* **1,** 1225–1226.
16. Jiang, H., Wang, Y. & Kanost, M. R. *Insect Mol. Biol.* in press.
17. Dunn, P. & Drake, D. (1983) *J. Invertebr. Pathol.* **41,** 77–85.
18. Jiang, H. & Kanost, M. R. (1997) *J. Biol. Chem.* **272,** 1082–1087.
19. Jiang, H., Wang, Y., Ma, C. & Kanost, M. R. (1997) *Insect Biochem. Mol. Biol.* **27,** 835–850.
20. Laemmli, U. K. (1970) *Nature (London)* **227,** 680–685.
21. Skinner, M. K. & Griswold, M. D. (1983) *Biochem. J.* **209,** 281–284.
22. Jiang, H., Wang, Y. & Kanost, M. R. (1994) *J. Biol. Chem.* **269,** 55–58.
23. Sambrook, J. Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual.* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
24. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
25. Genetics Computer Group (1991) *Program Manual for the GCG Package* (Genetics Computer Group, Madison, WI), version 7.
26. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680.
27. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997) *Protein Eng.* **10,** 1–6.
28. Wang, Y., Willott, E. & Kanost, M. R. (1995) *Insect Mol. Biol.* **4,** 113–123.
29. Jiang, H., Wang, Y. & Kanost, M. R. (1996) *Insect Mol. Biol.* **5,** 31–38.
30. Aso, Y., Kramer, K. J., Hopkins, T. L. & Lookhart, G. L. (1985) *Insect Biochem.* **15,** 9–17.
31. Muta T., Hashimoto, R., Miyata, T., Nishimura, H., Toh, Y. & Iwanaga, S. (1990) *J. Biol. Chem.* **265,** 22426–22433.