BIOCHEMISTRY

 Volume 46, Number 41 October 16, 2007

Articles

The Solution Structure of Clip Domains from *Manduca sexta* Prophenoloxidase Activating Proteinase-2†

Rudan Huang,‡ Zhiqiang Lu,‡ Huaien Dai,§ David Vander Velde,|| Om Prakash,§ and Haobo Jiang*,‡

Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, Oklahoma 74078, Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506, and Department of Medicinal Chemistry, University of Kansas, Lawrence, Kansas 66045

Received June 1, 2007; Revised Manuscript Received July 16, 2007

ABSTRACT: Clip domains are structural modules found in arthropod serine proteinases and some proteolytically inactive homologues, which mediate extracellular signaling pathways of development and immunity. While little is known about their structures or functions, clip domains are proposed to be sites for interactions of proteinases with their activators, cofactors, and substrates. Here we report the solution structure of dual clip domains from *Manduca sexta* prophenoloxidase activating proteinase-2. Each domain adopts a new mixed α/β fold (a three-stranded antiparallel β-sheet flanked by two α-helices), and the architecture provides structural information on clip domains from a catalytically active proteinase for the first time. Examination of the structure in conjunction with a multiple sequence alignment of the clip domains from different groups suggests a substrate-binding site, a bacteria-interacting region, and a surface for specific interactions. In summary, our results provide insights into the structural basis of clip domain functions and this structure may represent the prototype of group-2 clip domains.

The clip domain was first identified in the horseshoe crab proclotting enzyme (*1*) and was so named because it took the shape of a paperclip in the schematic drawing showing the disulfide linkages. Members of this family are 30−60 residues long, with an overall sequence similarity of 20−40%, characterized by six absolutely conserved Cys residues (*2*). Clip domains, found so far only in arthropods such as insects, crustaceans, and horseshoe crabs, are located at the amino-terminus in serine proteinases (SPs[1]) or serine protease homologues (SPHs). SPHs lack an enzymatic activity because one or more of the catalytic triad residues are substituted. Clip-domain SPs are activated by cleavage between the clip and proteinase domains. After cleavage activation, the clip domain remains covalently attached to

† This work is supported by National Institutes of Health Grants GM58634 (to H.J.), S10-RR022392 (to O.P.), and P20-RR017708 (COBRE-PSF). This article was approved for publication by the Director of Oklahoma Agricultural Experimental Station and supported in part under Project OKLO2450.

* Corresponding author. Department of Entomology and Plant Pathology, 127 Noble Research Center, Oklahoma State University, Stillwater, OK 74078. Tel: (405)-744-9400. Fax: (405)-744-6039. E-mail: haobo.jiang@okstate.edu.

‡ Oklahoma State University.

§ Kansas State University.

|| University of Kansas.

[1] Abbreviations: proPO, prophenoloxidase; PAP2 and proPAP2, proPO activating proteinase-2 and its precursor; PPAE and PPAF, proPO activating enzyme/factor; SP and SPH, serine proteinase and serine proteinase homologue; HPLC, high performance liquid chromatography; SDS, sodium dodecyl sulfate; NMR, nuclear magnetic resonance; TOCSY, total correlation spectroscopy; NOE and NOESY, nuclear Overhauser effect and NOE spectroscopy; HSQC, heteronuclear single quantum coherence; RMSD, root-mean-square deviation; HP21, hemolymph proteinase 21; PDB, Protein Data Bank.

10.1021/bi7010724 CCC: $37.00
Published on Web 09/19/2007

its catalytic or SP-like domain by an interchain disulfide bridge. Genome analysis revealed that clip domains represent the most abundant structural modules in arthropod SPs and SPHs: about 80 genes in *Drosophila melanogaster* and *Anopheles gambiae* encode SP-related proteins, each with 1 to 5 clip domains (*3*, *4*). In the tobacco hornworm *Manduca sexta*, there are at least 17 SPs and 2 SPHs containing 1−2 clip domains (*5*). The clip-domain SPs comprise three groups (1_a, 1_b, and 2) whereas clip-domain SPHs represent a lineage diverged from the former during early evolution of this gene family (*2*, *3*).

Clip-domain SPs mediate developmental signals and defense responses in arthropods. These include the horseshoe crab hemolymph coagulation, *Drosophila* Toll pathway activation during embryonic development and induced synthesis of antimicrobial peptides, as well as mosquito melanotic encapsulation of malaria parasites (*6*−*9*). One of the best studied immune responses in insects and crustaceans involves the limited proteolysis of prophenoloxidase (proPO). Recognition of pathogens by pattern recognition receptors activates a SP cascade via sequential proteolysis. Active phenoloxidase catalyzes quinone formation and melanin production for wound healing and microbe killing (*10*). In some insects, proPO activation is mediated by a proPO activating proteinase (PAP, also known as PPAE) in the presence of an SPH (*11*, *12*). PAP and SPH each contains 1−2 regulatory clip domains at their amino-terminus (*13*−*17*). They are terminal components of the SP cascade, analogous to blood coagulation and complement activation systems in mammals (*18*, *19*). *M. sexta* PAP2 belongs to group-2, the largest and most prevalent subfamily of clip-domain SPs which include *Drosophila* easter and PAPs from moths, beetles, flies, and mosquitoes.

While clip-domain proteins are implicated in a wide range of biological processes, relatively little is understood about molecular functions of clip domains. To study the structural basis of their functions, we expressed and purified the dual clip domains (i.e., clip-1 and clip-2) from *M. sexta* PAP2 and determine their three-dimensional structures by NMR spectroscopy. The structure serves as a basis for structure-based probing of clip domain functions and supports the proposed roles of clip domain in proPO activation.

MATERIALS AND METHODS

Cloning. The 5′ fragment of *M. sexta* PAP-2 cDNA was amplified by PCR using the longest clone as template (*14*). Primers 662 (5′-TTTGC**CATG**GGACAAGCCTGC-3′) and 672 (5′-TGC**AAGCTTA**CGGGCAGCA-3′) correspond to nucleotides 79−93 and 423−442 (reverse complement) of the cDNA except for a few changes to introduce the restriction sites and stop codon. The thermal cycling conditions were 35 cycles of 94 °C, 30 s; 53 °C, 30 s; 72 °C, 40 s, followed by 10 min of incubation at 72 °C. After digestion with *Nco*I and *Hin*dIII, the 364 bp PCR product was inserted to the same sites in plasmid H6pQE60 (*20*) to generate plasmid H12/H6pQE60, which adds the sequence Met-His-His-His-His-His-His-Ala-Met-Gly at the amino terminus. The resulting transformants were examined for induced expression of the recombinant clip domains, correct restriction digestion pattern, and insert sequence.

Protein Expression and Purification. Escherichia coli M15 cells (Qiagen) containing H12/H6pQE60 were initially grown in 50 mL of LB medium (100 μg/mL ampicillin) at 30 °C overnight. The culture was transferred into 1 L of M9 minimal medium supplemented with 1 g of $^{15}NH_4Cl$, 3 g of ^{13}C-glucose, and 10 mL of 10×^{15}N, ^{13}C-labeled Bio-Express-1000 (CIL). After growing at 37 °C for 2 h, the culture was added to 20 L of the fresh medium and continued to grow for about 4 h till A_{600} reached 0.7. Expression was induced with 1 mM isoprppyl-β-D-1-thiogalactopyroanoside for 4 h at 37 °C. Cells were harvested by centrifugation at 10000*g* for 20 min at 4 °C, and the pellet (55 g, wet weight) was stored at −80 °C.

The cell pellet was resuspended with 500 mL of buffer A (50 mM phosphate buffer, pH 8.0, 0.9% NaCl) containing 0.1 mg/mL lysozyme. After incubation at room temperature for 30 min, the cell suspension was sonicated on ice (1 min each, 10 times). Following centrifugation at 18000*g* for 30 min, the supernatant was heated at 70 °C for 10 min with agitation and then chilled in an ice/water bath. After removal of the precipitate by centrifugation (18000*g*, 30 min), the supernatant was recovered and loaded onto a Ni-nitrilotriacetic acid agarose column (150 mL bed volume) and washed with 750 mL of buffer A. The bound proteins were eluted by a linear gradient of 0−300 mM imidazole in 1500 mL of buffer A. Fractions containing the PAP-2 clip domains were combined and separated by reversed-phase HPLC on a Source-15 RPC column (4.6 × 100 mm) (Amersham Biosciences). In each run, 100 mL of the pooled sample was loaded and eluted with a 0−30% acetonitrile gradient in 10 mM Tris-HCl, pH 8.5 in 50 min at a flow rate of 1.0 mL/min. Fractions of the first peak, confirmed by electrophoretic analysis to be the monomer migrating as a single band on a nonreducing polyacrylamide gel, were combined and subjected to vacuum centrifugation on a Speed-Vac for 1 h. With the organic solvent removed, the protein sample was dialyzed against 10 mM NH_4HCO_3 and lyophilized.

NMR Spectroscopy. The protein samples for NMR studies were either uniformly ^{15}N-labeled or ^{13}C- and ^{15}N-double labeled, dissolved in 100 mM phosphate buffer (pH 8.0) with 90% H_2O, 10% ^2H_2O. The protein concentration was 1.0 mM. In the $^1H/^2H$ exchange experiment, the lyophilized protein was dissolved in 99.9% ^2H_2O and the 2D $^1H-^{15}N$ HSQC spectrum was collected after 8 h. All NMR spectra were acquired at 298 K on Bruker Avance 800 MHz spectrometers equipped with a 5 mm $^1H/^{13}C/^{15}N$ triple resonance inverse detection cryogenic probe with *xyz* pulse field gradients. Spectral data were processed and interpreted using NMRPipe (*21*) and SPARKY (*22*) software packages, respectively. ^{1}H, ^{13}C, and ^{15}N chemical shift referencing followed a previous report (*23*). No stereospecific assignments were made.

A suite of 2D- and 3D-NMR spectra were obtained for backbone and side-chain chemical shift assignments and structural constraints. These include 2D $^1H-^{15}N$ HSQC and $^1H-^{13}C$ HSQC, 3D CBCA(CO)NH, HNCACB, HNCA, HNCO, CCCONH, HBHA(CO)NH, HCCH-TOCSY, hCCH-TOCSY, ^{15}N-edited $^1H-^1H$ NOESY (mixing time: 100 ms), and ^{13}C-edited ^1H-^1H NOESY (mixing time: 100 ms). An additional 3D ^{15}N-edited $^1H-^1H$ NOESY-HSQC spectrum (mixing time: 100 ms) was collected from the ^{15}N-labeled sample of PAP-2 clip domains in order to identify HN−HN contacts. Sequence-specific backbone ^{1}H, ^{15}N, and ^{13}C resonance assignments were based on the 3D CBCA(CO)-

NH, HNCACB, HNCA, and HNCO. The side chain ^{1}H, ^{15}N, and ^{13}C signals were assigned using the data from 3D HBHA(CO)NH, CCCONH, HCCH-TOCSY, and hCCH-TOCSY. Resonance assignment for Asn and Gln side chain amides was based on 2D $^{1}H-^{15}N$ HSQC and 3D $^{1}H-^{1}H-^{15}N$ NOESY-HSQC. Unambiguous NOE constraints were manually assigned using 3D ^{15}N-edited and ^{13}C-edited NOESY spectra.

Hydrogen bond restraints (O−HN, 1.8−2.1 Å; O−N, 2.7−3.0 Å) were deduced from backbone chemical shift analysis and from the $^{1}H-^{2}H$ exchange experiment. Backbone Φ and Ψ torsion angle restraints were derived from an analysis for H^{α}, C^{α}, C^{β}, C^{γ}, and backbone ^{15}N chemical shifts using TALOS (*24*). Backbone dihedral angle constraints and hydrogen bond constraints were applied only to residues that are clearly engaged in helical or β-strand structures.

Structure Calculations. Only residues Q11−G124 were included in the structure calculation, whereas the remaining part of the protein was more flexible as indicated by the lack of NOE correlations. Structures were calculated from extended backbone conformations as starting models using the program CNS (*25*). NOE-derived interproton distance restraints were classified into approximate distance ranges of 1.8−2.5 Å, 1.8−3.5 Å, 1.8−5.0 Å, and 1.8−6.0 Å corresponding to strong, medium, weak, and very weak NOE cross peaks; an additional 0.5−1.0 Å was added to the upper distance bound of distance restraints involving methyl groups.

Initial structure ensembles were calculated with manually assigned NOE distance restraints and dihedral angles. A total of 1000 structures were calculated, and the best in terms of lowest energy was used as a starting point for further NOE assignment of the remaining peaks. After disulfide bond connections were certified by sulfur atom distances, NOE cross peaks, and energy optimization, disulfide bond relationship and hydrogen bond constraints were added to subsequent calculations. A total of 100 structures were calculated in the last iteration, from which 20 were chosen for further analysis. None of these structures has NOE or dihedral violations greater than 0.5 Å or 5°, respectively.

Neither enough NOE restraints in the link region between clip-1 and clip-2 from residues Pro66−Thr70 nor interdomain NOE restraints were obtained. Since the exact position of the two domains cannot be determined, structures calculated based on full-length template were divided into two single domains for further analysis. Ensemble RMSDs were calculated with MOLMOL (*26*). The structure ensembles were analyzed and validated in terms of geometry and restraint violations with AQUA and PROCHECK-NMR (*27*). Molecular graphics were produced with MOLMOL and PYMOL (http://pymol.sourceforge.net).

Data Bank Accession Codes. The coordinates and constraints for the two clip domains have been deposited in the Protein Data Bank with accession codes 2IKD and 2IKE for clip-1 and clip-2, respectively. The chemical shift assignments have been deposited in the Biological Magnetic Resonance Data Bank with accession code 15105 (clip-1) and 15106 (clip-2).

RESULTS AND DISCUSSION

Preparation of the Dual Clip Domains. The clip domains of *M. sexta* PAP-2 were expressed as a soluble recombinant protein fused with an amino-terminal hexahistidine tag which facilitated its capture by Ni-nitrilotriacetic acid beads. While the affinity-purified protein migrated as a single predominant band on the reducing SDS−polyacrylamide gel, electrophoresis under nonreducing conditions revealed at least four bands whose apparent molecular masses correspond to the monomer, dimer, trimer, and tetramer (Figure 1). In order to isolate the monomer, we applied the protein sample onto a reversed phase-HPLC column equilibrated at an alkaline condition. As expected for correctly folded Cys-rich proteins, the monomer came off the C18 column as a sharp peak followed by broad peaks containing other forms. From 20 L of *E. coli* culture, we obtained 8.0 mg of monomeric PAP-2 clip domains at a purity of over 95%. The assigned $^{1}H-^{15}N$ heteronuclear single quantum coherence (HSQC) spectrum shows a good dispersion of $^{1}H-^{15}N$ cross peaks (Figure 2), indicative of a well-defined structure. In addition, the HSQC spectrum with the expected number of cross peaks (i.e., without minor or shifted NH signals) suggests that our protein adopted one conformation.

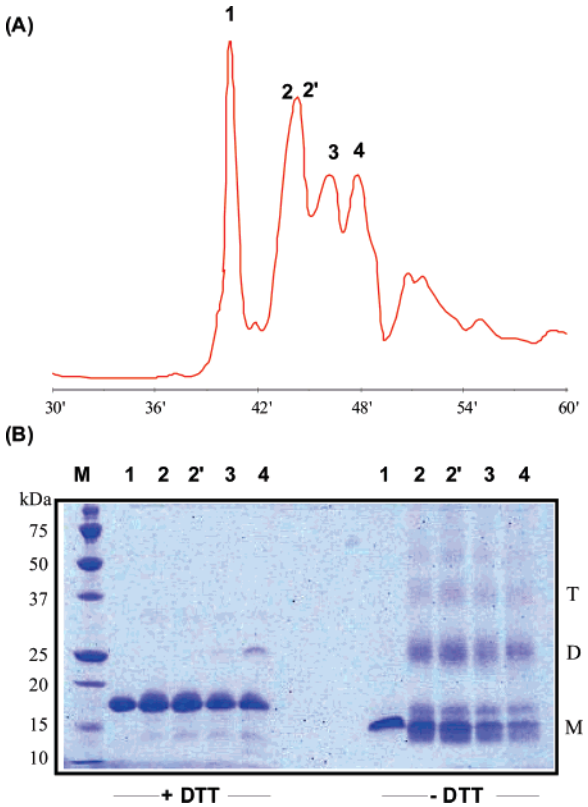

FIGURE 1: Different association states of the PAP-2 dual clip domains separated by reversed phase HPLC. (A) Elution profile: the affinity-purified recombinant protein was applied to a Source-15 RPC column equilibrated with 10 mM Tris-HCl, pH 8.5. A linear gradient of acetonitrile was employed to elute the protein. With absorbance monitored at 214 nm, fractions were manually collected. The first and second halves of peak-2 (labeled 2 and 2′) were collected separately for SDS−polyacrylamide gel electrophoretic analysis. (B) The samples (1, 2, 2′, 3, and 4) and molecular weight markers (M) were treated by SDS-sample buffer with or without dithiothreitol and resolved on a 12% gel. Following electrophoresis, the protein bands were visualized by Coomassie Brilliant Blue staining. The sizes of the markers are indicated on the left, and the positions of monomer (M), dimer (D), and trimer (T) are marked on the right.

Disulfide Linkage. The C^{α} and C^{β} chemical shifts of cysteine residues are diagnostic of disulfide bond formation

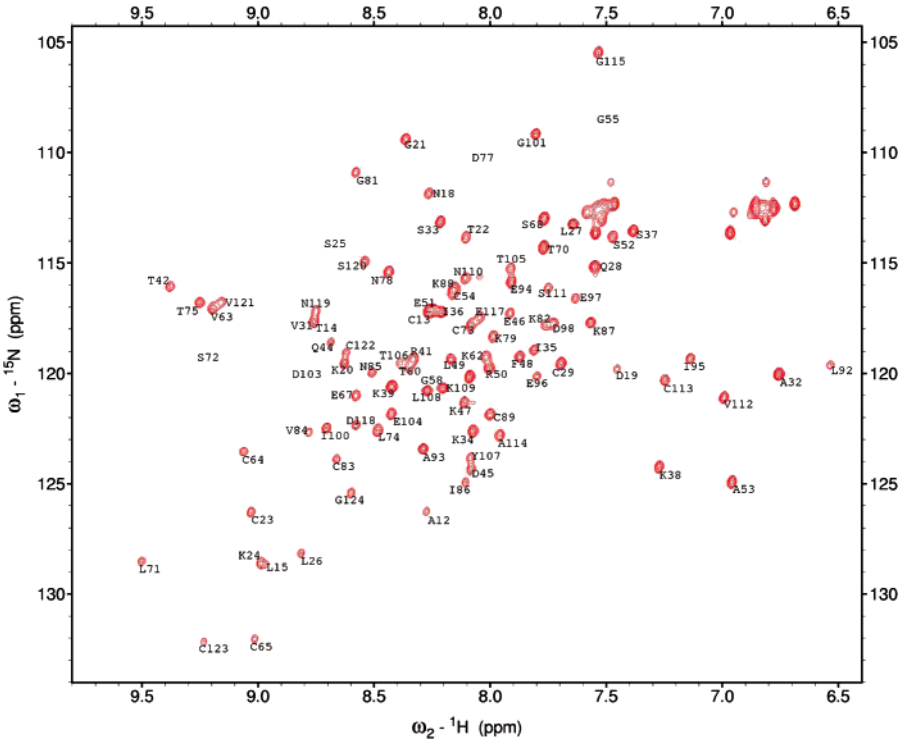

FIGURE 2: $^{1}H-^{15}N$ HSQC spectrum of the dual clip domains labeled with the backbone assignments. The spectrum was collected from a $^{15}N,^{13}C$-labeled sample in 100 mM phosphate buffer (90% H_2O, 10% $^{2}H_2O$) at pH 8.0. Assignments were made from residue Q11 through residue G124. Gly124, located at the carboxyl terminus, is different from the other Gly residues.

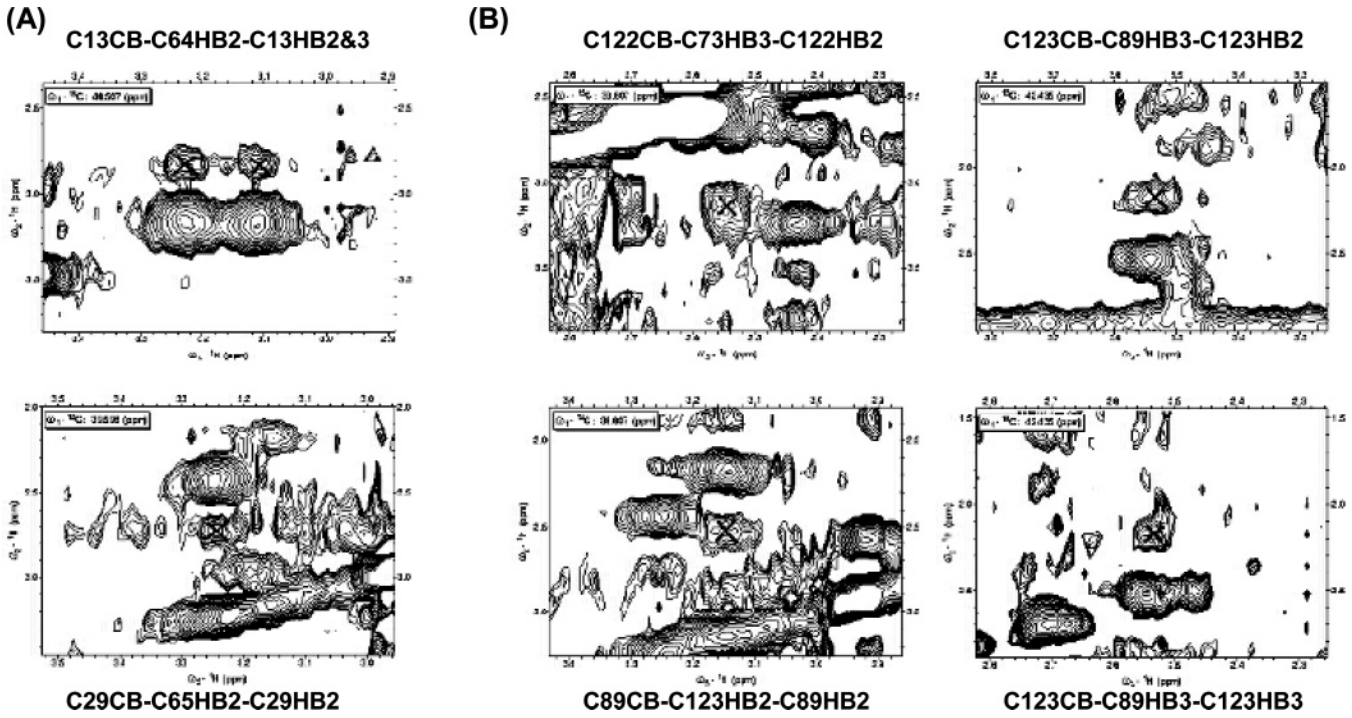

FIGURE 3: Spectral slides from ^{13}C-edited NOESY-HSQC showing NOE cross peaks between H_β of the Cys residues in clip-1 (A) and clip-2 (B). The cross peaks (marked "×") reveal the disulfide linkages (Cys-1 and Cys-5; Cys-3 and Cys-6) in each domain.

(*28*). The C^β shift is particularly characteristic: 24−35 ppm in the reduced state and 32−52 ppm when oxidized. Twelve Cys residues in the dual clip domains have the C^β shifts ranging from 38.9 to 46.8 ppm, which indicate the formation of disulfide bonds.

Clear NOE cross peaks from the β protons of Cys residues on ^{13}C-edited NOESY-HSQC were obtained for two disulfide linkages in each clip domain (Figure 3). One showed relationship between Cys-1 and Cys-5 (i.e., Cys13−Cys64 in clip-1 and Cys73−Cys122 in clip-2), and the other was between Cys-3 and Cys-6 (i.e., Cys29−Cys65 in clip-1 and Cys89−Cys123 in clip-2). Significant differences in H_β chemical shifts of the 12 cysteine residues allow unambiguous assignment for these NOEs and indicate that there is no interdomain disulfide linkage. Although the overlaps in the H_β region of Cys-2 and Cys-4 prevent direct establishment

Table 1: NMR Structure Determination Statistics

	"full-length" residue 11−124	clip-1 residue 11−66	clip-2 residue 71−124
A. Restraint Statistics			
NOE distance restraints			
all	1557	823	690
intraresidue	383	191	185
sequential ($\|i-j\|=1$)	532	288	231
medium range ($2\leq \|i-j\| \leq 4$)	320	188	130
long range ($\|i-j\|>4$	300	156	144
hydrogen bonds	74	38	36
dihedral angle restraints	55	28	27
B. Ensemble Statistics Evaluation			
rmsd[a] (from the mean coordinates)			
backbone heavy atoms (Å)		0.83 ± 0.20	0.94 ± 0.29
all heavy atoms (Å)		1.61 ± 0.21	1.85 ± 0.35
Ramachandran plot statistics (PROCHECK)			
residues in most favored regions (%)		69.0	79.4
residues in additional allowed regions (%)		24.9	17.3
residues in generously allowed regions (%)		3.3	2.6
residues in disallowed regions (%)		2.8	0.7

[a] Rmsd to the mean structure for residues in secondary structured elements (residues 11−14, 22−25, 30−38, 43−52, 62−65 in clip-1 and 71−74, 82−85, 90−97, 102−111, 120−124 in clip-2).

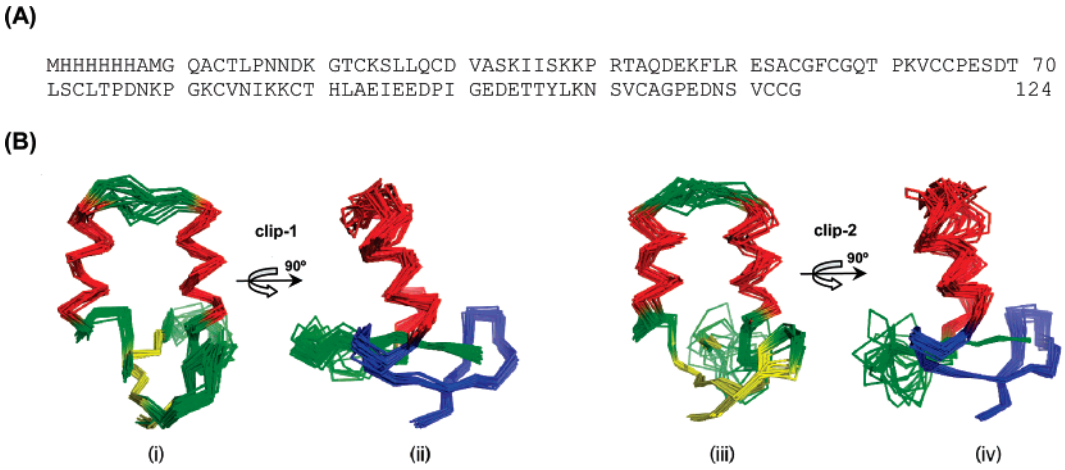

FIGURE 4: Sequence (A) and solution structure (B) of the recombinant dual clip domains. The solution structures of PAP2 clip-1 (i, ii) and clip-2 (iii, iv) presented as a bundle of the 20 low-energy conformers. The models (i, iii) are colored corresponding to 2° structural elements: red for α-helices and yellow for *β*-strands, and they (ii, iv) gradually change color from blue (N-terminus) to red (middle) and green (C-terminus).

of the linkages, we, based on the existing information, deduce the disulfide bond connectivity as 1−5, 2−4, 3−6 (Cys13−Cys64, Cys23−Cys54, and Cys29−Cys65 in clip-1; Cys73−Cys122, Cys83−Cys113, and Cys89−Cys123 in clip-2). This is consistent with the linkage pattern in other clip domains (*1*, *29*).

Overall Structure. Residues 1−10 (Met-His-His-His-His-His-His-Gly-Met-Gly, not a part of clip-1) were not included in the structural determination. The solution structure of dual clip domains was calculated by CNS (*25*) using 1557 distance restraints, 55 dihedral angle restraints, 74 hydrogen bond restraints, and the known disulfide bond connections. Superimposing one of the two clip domains led to highly random orientations of the other. Twenty-two sequential or intraresidue NOEs were assigned to residues in the linker region (P66−T70), but no medium- or long-range NOE. There was no obvious peak remaining unassigned for P66−T70 on ^{15}N-NOESY-HSQC or ^{13}C-NOESY-HSQC. Fewer than ten obvious peaks were left unassigned in other parts of the protein. These data indicated a flexible interdomain relationship. Consequently, we separated the structure ensemble into two parts: clip-1 (Glu11−Pro66, 56 residues) and clip-2 (Leu71−Gly124, 54 residues). Experimental restraints and structural statistics are described in Table 1. Residues 11−14, 21−25, 30−38, 43−52, 62−66 in clip-1 and residues 71−74, 81−85, 90−97, 102−111, 120−124 in clip-2 adopt a well-defined conformation. The root-mean-square deviations (rmsd) of the 20 lowest energy conformers superimposed onto the mean coordinates are 0.83 (clip-1)/0.94 (clip-2) for the backbone atoms and 1.61 Å (clip-1)/1.85 Å (clip-2) for all heavy atoms (Table 1; Figure 4).

Comparison of our clip domain structures with the entries in the Protein Data Bank (PDB) using DALI (*30*) revealed no statistically significant similarities, suggesting that the PAP2 clip domains represent a new fold (Figure 4B and Figure 5A). Both adopt a compact mixed α/*β* fold which comprises a three-stranded antiparallel *β*-sheet flanked by two α-helices (denoted as *β*1-*β*2-α1-α2-*β*3). The two clip domains, which are 28% identical in sequence, closely resemble each other in secondary and tertiary structures.

One special feature of the clip domain structure is the relationship between α-helices and *β*-sheet. The two helices

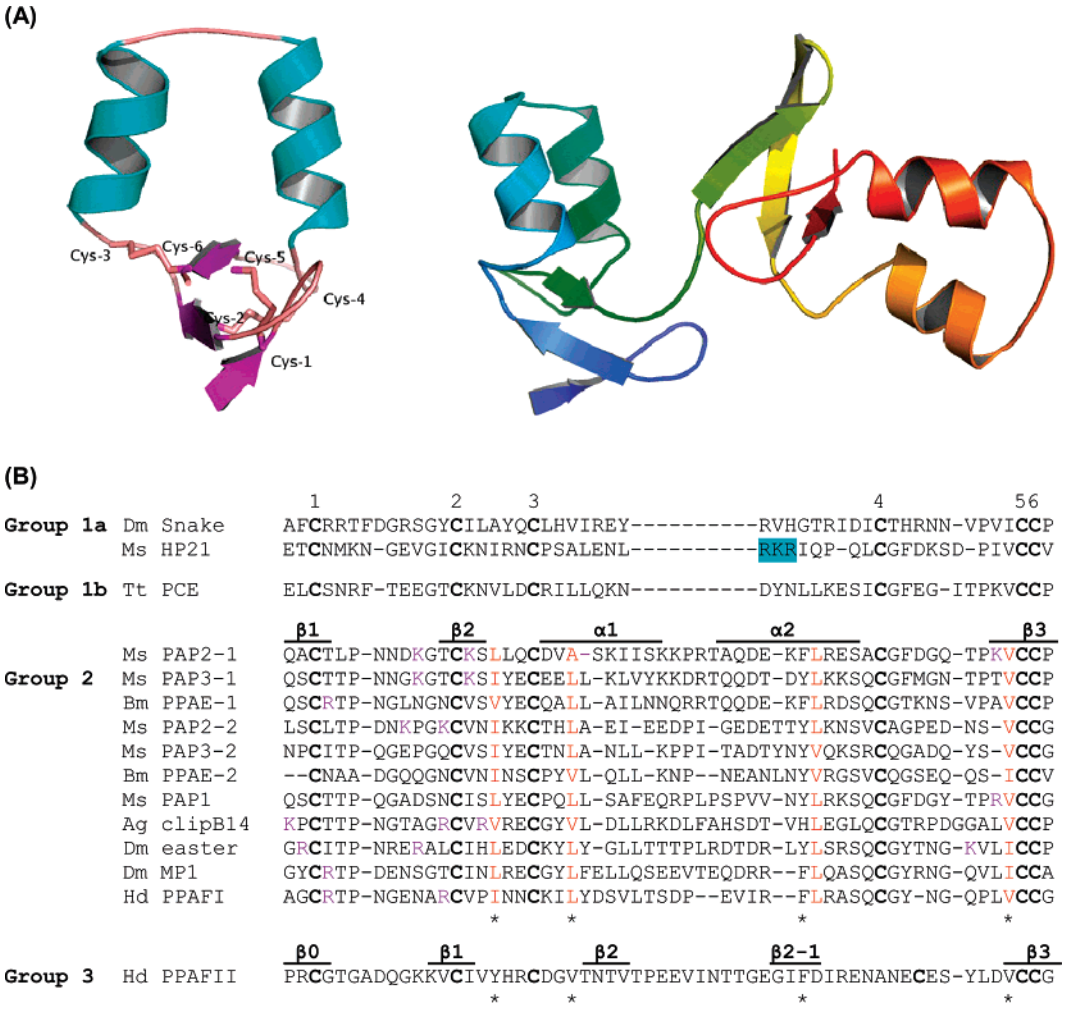

FIGURE 5: Ribbon diagram of the PAP2 clip domains and sequence alignment of the clip domains from different groups. (A) The single clip domain structure with the disulfide linkages shown (left panel) and a possible conformation of the full-length structure (right panel). (B) Multiple sequence alignment of clip domains. The secondary structure of PAP2 and PPAFII clip domains is displayed. The absolutely conserved Cys residues are shown in bold; conserved residues important for hydrophobic packing in group-2 clip domains are marked red; equivalent residues around the central cavity in PAP2 and PPAFII clip domains are indicated by *. The positively charged patch in HP21 clip domain is highlighted in blue. Positively charged residues in or near three strands are shown in purple. Dm: *D. melanogaster*. Tt: *Tachypleus tridentatus*. Ag: *A. gambiae*. Ms: *M. sexta*. Bm: *Bombyx mori*. Hd: *H. diomphalia*.

are antiparallel to each other and almost perpendicular to the β-sheet. Stabilizing factors between the α1 and β-sheet regions include disulfide linkage Cys29−Cys65 (clip-1)/Cys89−Cys123 (clip-2) and packing interactions through the side-chains of Cys29-Ala32-Val63 (clip-1)/Cys89-Leu92-Val121 (clip-2). On the other side, α2 and β-sheet interface resides in Leu49−Val63 (clip-1)/Leu108−Val121 (clip-2). The network of hydrophobic interactions anchors the two α-helices and β-sheet together. Strong sequence conservation between group-2 clip domains at these residue positions suggests that this hydrophobic packing arrangement is a general feature of the fold (Figure 5).

Structure Relatives. To date, the only tertiary structure of clip domain is from a group-3 SPH, *Holotrichia diomphalia* PPAFII (PDB code 2B9L) (*29*). In the crystal structure, PPAFII clip consists of an irregular, four-stranded β-sheet and several loops, one of which contains β2-1 fragment pairing with the linker region. A comparison with our solution structures reveals a somewhat similar β-sheet with three strands located in similar regions. Four of the six cysteines are center residues in the β-strands: Cys-1 (β1), Cys-2 (β2), Cys-5 (β3), Cys-6 (β3) in PAP2 clip domains and, equivalently, Cys-1 (β0), Cys-2 (β1), Cys-5 (β3), Cys-6 (β3) in PPAFII clip. The preference of cysteine residue in β-strand conformation (*31*) and relatively high conservation of sequences around the Cys residues in clip domains (*2*, *3*) lead to the hypothesis that the β-sheet is a common structural feature of clip domains.

Broadly defined as $CX_{3-10}CX_{4-10}CX_{8-36}CX_{5-17}CC$, clip domains vary greatly in their sequence, structure, and associating domains. The PPAF-II clip domain belongs to group-3 which is least similar to the prevailing group-2. While PPAF-II has no catalytic activity and its clip domain (cleaved between Cys-3 and Cys-4) associates with the proteinase-like domain for a catalysis-unrelated function (*29*), the PAP2 clip domains associate with the catalytic domain and are not cleaved. Consistent with these, there are significant differences between the structures. A major one lies between Cys-3 and Cys-4: in PAP2 this region adopts a helix−turn−helix fold whereas in PPAFII it forms a long loop containing a small piece of β-strand (β2-1). This major difference provides the structural basis for those functional differences: the β2-1 in PPAFII clip domain interacts with region I in PPAFII proteinase-like domain (corresponding

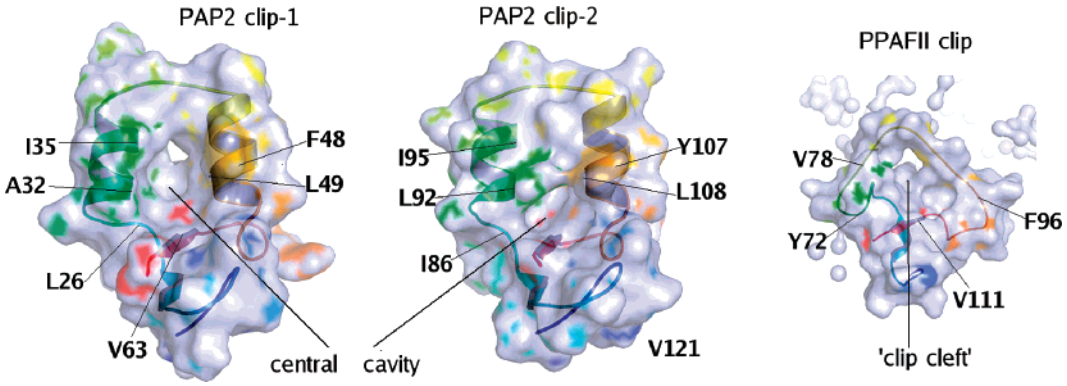

FIGURE 6: Surface comparison of the clip domains in PAP2 and PPAFII. Similar residues forming hydrophobic core of the central cavity for proPO binding are indicated.

to the proteolytic activation site in clip-domain SPs) and prevents its cleavage at the typical position as seen in PAP2.

Probable ProPO Binding Site. A prominent central cleft in PPAFII clip was hypothesized to bind proPO (*29*): a mutation (Val111Ala) abolished the binding, and this Val residue is conserved among all the clip domains. Residues Tyr72, Val78, and Val111 form a hydrophobic core of the cleft. Equivalent residues in PAP2 clips (Leu26, Ala32, and Val63 in clip-1; Ile86, Leu92, and Val121 in clip-2) are located in a large internal cavity resembling the cleft in PPAFII clip (Figure 6). We, therefore, suggest that the hydrophobic cavity in PAP2 clips serves the same purpose. In the solution structure, the outer rim of the cavity is lined by α1, α2, and β3, whereas the bottom of the cavity is composed of the loop between Cys-2 and Cys-3. The rmsd for superposition of all ordered fragments is higher than that for the β-sheet portion, indicating some flexibility in the positions of α-helices relative to the β-sheet. Such flexibility may facilitate the binding site exposure, and we expect a change in cavity size upon proPO binding. This seems to be consistent with the cleft flexibility found in PPAFII clip, the outer rim of which is lined by the β2-1-containing loop and β3 strand.

Aromatic residues with low thermal flexibility may provide control for binding specificity. While highly hydrophilic clip domains contain few aromatic residues, we found one around the cavity of both clip-1 (Phe48) and clip-2 (Tyr107), probably equivalent to Phe96 in PPAFII clip. Located in α2 and with the side-chain pointing outward in the solution structures, these aromatic residues may direct the recognition and binding of proPO. Perhaps a clip domain first interacts with proPO via the aromatic residues, and strong binding follows through interactions between hydrophobic patches.

Putative Bacterial Interaction Surface. The fact that the clip domain has the same disulfide bond linkages as defensin leads to the suggestion that clip domains might interact with bacterial surfaces. Indeed, Söderhäll and co-workers showed that the clip domain of *Pacifastacus leniusculus* PAP inhibits bacterial growth (*32*). Examination of PAP2 clip structures reveals a large, positively charged surface around the β-sheet (Figure 7, middle panel). This surface is composed of the side-chains of Lys20, Lys24, Lys62 in clip-1 and Lys79, Lys82 in clip-2. The high net positive charge could facilitate electrostatic attraction and association to the polyanionic surfaces of bacterial cells. The positive charged patches on the PAP2 clip domains are not derived from equivalent residues but from residues located in or around the three β-strands. Similar positive charge residues are found in other group-2 members (Figure 5B). Since no additional structural data are available for SP-associated clip domains, we are unclear whether or not such a structural feature is unique for the PAP2 clip domains.

Possible Activator Docking Site. Several lines of indirect evidence suggest that clip domains may act as an interaction site for activators and cofactors that form complexes in a SP cascade. For instance, the clip domain of *H. diomphalia* PPAFI is essential for the rapid activation of its catalytic domain (*29*). High ionic strength significantly slows proPO activation by PAP2 *in vitro* (data not shown), suggesting that electrostatic interactions are important in the activation process. By examining the surface of PAP2 clip structure, we observed a predominant charged surface on the helix region (Figure 7, right panel). The charged regions in clip-1 and clip-2 could serve as docking sites for its interacting proteins with complementary electrostatic surfaces. Interestingly, clip-1 has net positive charges composed of Lys34, Lys38, and Lys39 side-chains whereas clip-2 has net negative charges contributed by Glu94, Glu97, Asp98, Glu102, and Asp103. Due to the differences in charge properties and surface characteristics, these two sites may have different specificity in binding. The current model of proPO activation pathway in *M. sexta* hemolymph indicates that hemolymph proteinase 21 (HP21) functions as the direct activator for proPAP2 (*33*, *34*). HP21 is a group-1_a clip-domain serine proteinase. While the structure of HP21 clip domain is unknown, a positively charged patch is found in the similar region (Figure 4B). Complementary electrostatic patches on HP21 clip domain and PAP2 clip-2 may direct protein recognition and interaction.

The hypervariable sequence between Cys-3 and Cys-4 is likely responsible for recognition and binding specificity. Consistent with the large structural difference in this region between PPAFII and PAP2 clip domains, we expect structural variations in the other subgroups.

Biological Implications. What could be the roles of PAP-associated clip domains in the proPO activation? Based on the solution structure, we propose that, upon bacterial challenge, SPs may attract and later bind to the microbial surface through their clip domains which serve as anchors for forming immune protein complexes. Activating proteinases associate with proPAP2 via electrostatic interactions. These specific interactions ensure that the defense response

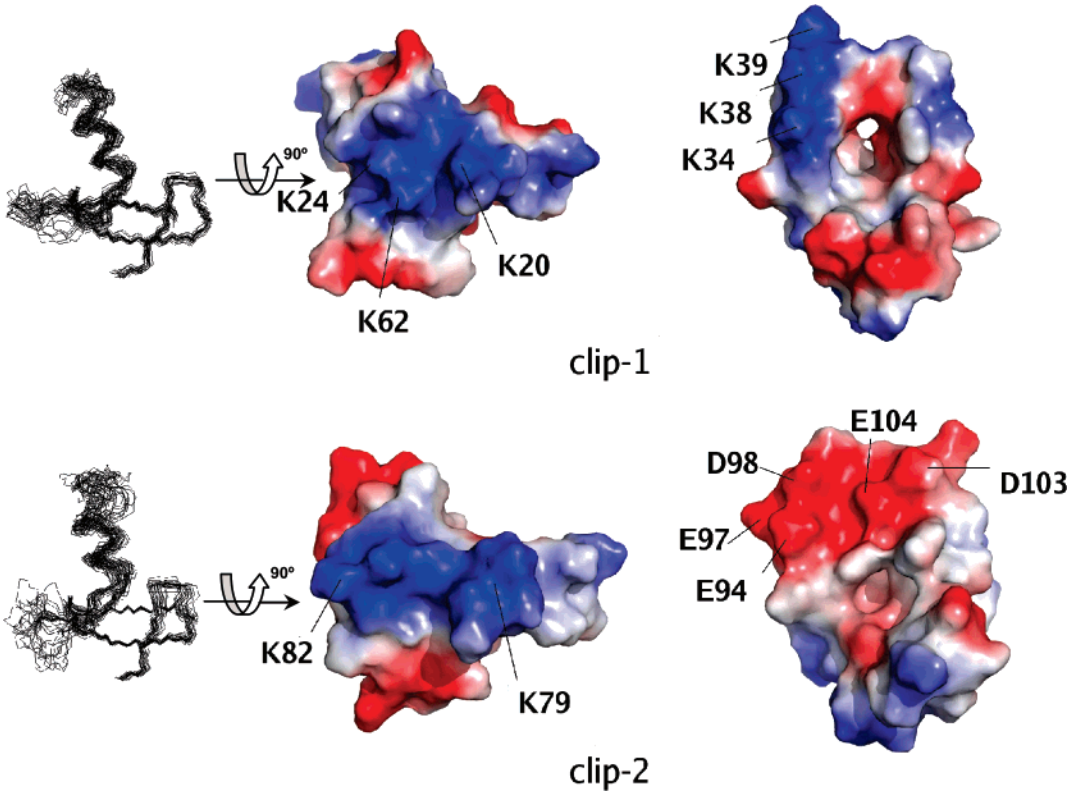

FIGURE 7: Electrostatic surface of clip domain-1 (upper row) and clip domain-2 (bottom row) in PAP2. The electrostatic surface on the β-sheet portion (middle panel), orientation shown in relation to the left panel. The electrostatic surface on the helix part (right panel) corresponds to the orientation shown in Figure 4B, left panel. The surfaces are color-coded according to electrostatic potential: red, −10 kT; white, 0 kT; blue, +10 kT.

occurs in the vicinity of invading pathogens. PAP2 precursor is activated by specific proteolytic cleavage in a manner similar to the activation of chymotrypsinogen. During proPO activation, clip domain may bind to the substrate through its large, hydrophobic cavity and direct the proPO cleavage by PAP2. These processes underscore the importance of clip domains, which remain connected to the catalytic domains after cleavage activation (*2*).

ACKNOWLEDGMENT

We thank Drs. Michael Massiah, Junpeng Deng, Michael Kanost, and Jack Dillwith for their helpful comments on the manuscript.

REFERENCES

1. Muta, T., Hashimoto, R., Miyata, T., Nishimura, H., Toh, Y., and Iwanaga, S. (1990) Proclotting enzyme from horseshoe crab hemocytes. cDNA cloning, disulfide locations, and subcellular localization, *J. Biol. Chem. 265*, 22426−22433.
2. Jiang, H., and Kanost, M. R. (2000) The clip-domain family of serine proteinases in arthropods, *Insect Biochem. Mol. Biol. 30*, 95−105.
3. Ross, J., Jiang, H., Kanost, M. R., and Wang, Y. (2003) Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationships, *Gene 304*, 117−131.
4. Christophides, G. K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., Brey, P. T., Collins, F. H., Danielli, A., and Dimopoulos, G. et al. (2002) Immunity-related genes and gene families in *Anopheles gambiae*, *Science 298*, 159−165.
5. Jiang, H., Wang, Y., Gu, Y., Guo, X., Zou, Z., Scholz, F., Trenczek, T. E., and Kanost, M. R. (2005) Molecular identification of a bevy of serine proteinases in *Manduca sexta* hemolymph, *Insect Biochem. Mol. Biol. 35*, 931−943.
6. Iwanaga, S., Kawabata, S., and Muta, T. (1998) New types of clotting factors and defense molecules found in horseshoe crab hemolymph: their structures and functions, *J. Biochem. 123*, 1−15.
7. Belvin, M. P., and Anderson, K. V. (1996) A conserved signaling pathway: the *Drosophila* Toll-dorsal pathway, *Annu. Rev. Cell Dev. Biol. 12*, 393−416.
8. Hoffmann, J. A., and Reichhart, J. M. (2002) *Drosophila* innate immunity: an evolutionary perspective, *Nat. Immunol. 3*, 121−126.
9. Christophides, G. K., Vlachou, D., and Kafatos, F. C. (2004) Comparative and functional genomics of the innate immune system in the malaria vector *Anopheles gambiae*, *Immunol. Rev. 198*, 127−148.
10. Nappi, A. J., and Christensen, B. M. (2005) Melanogenesis and associated cytotoxic reactions: applications to insect innate immunity, *Insect Biochem. Mol. Biol. 35*, 443−459.
11. Yu, X. Q., Jiang, H., Wang, Y., and Kanost, M. R. (2003) Nonproteolytic serine proteinase homologs are involved in prophenoloxidase activation in the tobacco hornworm, *Manduca sexta*, *Insect Biochem. Mol. Biol. 33*, 198−208.
12. Kwon, T. H., Kim, M. S., Choi, H. W., Joo, C. H., Cho, M. Y., and Lee, B. L. (2000) A masquerade-like serine proteinase homologue is necessary for phenoloxidase activity in the coleopteran insect, *Holotrichia diomphalia* larvae, *Eur. J. Biochem. 267*, 6188−6196.
13. Jiang, H., Wang, Y., and Kanost, M. R. (1998) Pro-phenol-oxidase activating proteinase from an insect, *Manduca sexta*: a bacteria-inducible protein similar to *Drosophila* easter, *Proc. Natl. Acad. Sci. U.S.A. 95*, 12220−12225.
14. Jiang, H., Wang, Y., Yu, X. Q., and Kanost, M. R. (2003) Prophenoloxidase-activating proteinase-2 from hemolymph of *Manduca sexta*, *J. Biol. Chem. 278*, 3552−3561.
15. Jiang, H., Wang, Y., Yu, X. Q., Zhu, Y., and Kanost, M. R. (2003) 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. 33*, 1049−1060.

16. Satoh, D., Horii, A., Ochiai, M., and Ashida, M. (1999) Prophenoloxidase-activating enzyme of the silkworm, *Bombyx mori*, *J. Biol. Chem. 274*, 7441−7453.
17. Lee, S. Y., Cho, M. Y., Hyun, J. H., Lee, K. M., Homma, K. I., Natori, S., Kawabata, S. I., Iwanaga, S., and Lee, B. L. (1998) Molecular cloning of cDNA for pro-phenol-oxidase-activating factor I, a serine protease is induced by lipopolysaccharide or 1,3-β-glucan in coleopteran insect, *Holotrichia diomphalia* larvae, *Eur. J. Biochem. 257*, 615−621.
18. Kim, M. S., Baek, M. J., Lee, M. H., Park, J. W., Lee, S. Y., Söderhäll, K., and Lee, B. L. (2002) A new easter-type serine protease cleaves a masquerade-like protein during prophenoloxidase activation in *Holotrichia diomphalia* larvae, *J. Biol. Chem. 274*, 7441−7453.
19. Kanost, M. R., Jiang, H., and Yu, X. Q. (2004) Innate immune responses of a lepidopteran insect, *Manduca sexta*, *Immunol. Rev. 198*, 97−105.
20. Lee, E., Linder, M. E., and Gilman, A. G. (1994) Expression of G-protein α subunit in *Escherichia coli*, *Methods Enzymol. 236*, 146−163.
21. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multi-dimensional spectral processing system based on UNIXpipe, *J. Biomol. NMR 6*, 277−293.
22. Goddard, T. D., and Kneller, D. G. (2002) SPARKY 3, University of California, San Francisco.
23. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., and Oldfield, E. (1995) ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR, *J. Biomol. NMR 6*, 135−140.
24. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J. Biomol. NMR 13*, 289−302.
25. Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., and Grosse-Kunstleve, R. W. et al. (1998) Crystallography and NMR system: a new software suit for macromolecular structure determination, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 54*, 905−921.
26. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macro-molecular structures, *J. Mol. Graphics 14*, 51−55.
27. Laskowski, R. A., Rullmannn, A, J., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, *J. Biomol. NMR 8*, 477−486.
28. Sharma, D., and Rajarathnam, K. (2000) ^{13}C NMR chemical shifts can predict disulfide bond formation, *J. Biomol. NMR 18*, 165−171.
29. Piao, S., Song, Y. L., Kim, J. H., Park, S. Y., Park, J. W., Lee, B. L., Oh, B. H., and Ha, N. C. (2005) Crystal structure of a clip-domain serine protease and functional roles of the clip domains, *EMBO J. 24*, 4404−4414.
30. Holm, L., and Sander, C. (1997) Dali/FSSP classification of three-dimensional protein folds, *Nucleic Acids Res. 25*, 231−234.
31. Bhattacharyya, R., Pal, D., and Chakrabarti, P. (2004) Disulfide bonds, their stereospecific environment and conservation in protein structures, *Protein Eng., Des. Sel. 17*, 795−808.
32. Wang, R., Lee, S. Y., Cerenius, L., and Söderhäll, K. (2001) Properties of the prophenoloxidase activating enzyme of the freshwater crayfish, *Pacifastacus leniusculus*, *Eur. J. Biochem. 268*, 895−902.
33. Gorman, M. J., Wang, Y., Jiang, H., and Kanost, M. R. (2007) *Manduca sexta* hemolymph proteinase 21 activates prophenoloxidase activating proteinase 3 in an insect innate immune response proteinase cascade, *J. Biol. Chem. 282*, 11742−11749.
34. Wang, Y., and Jiang, H. (2007) Reconstitution of a branch of *Manduca sexta* prophenoloxidase activation cascade *in vitro*: Snake-like hemolymph proteinase 21 cleaved by HP14 activates prophenoloxidase-activating proteinase-2 precursor, *Insect Biochem. Mol. Biol. 37*, 1015−1025.

BI7010724