THE SOLUTION STRUCTURE OF CLIP DOMAIN FROM *MANDUCA SEXTA* PROPHENOLOXIDASE ACTIVATING PROTEINASE 2

By

HUANG RUDAN

Bachelor of Science in Biology

University of Science and Technology of China

Hefei, China

2004

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
May, 2007

THE SOLUTION STRUCTURE OF CLIP DOMAIN

FROM *MANDUCA SEXTA* PROPHENOLOXIDASE

ACTIVATING PROTEINASE 2

Thesis Approved:

Thesis Adviser

Dean of the Graduate College

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to my major advisor, Dr. Haobo Jiang for his intelligent support, guidance, inspiration and supervision during my Master's study.

My sincere acknowledgment extends to my other committee members Dr. Robert Matts and Dr. Ramamurthy Mahalingam, whose diligent instruction, assistance and encouragement are invaluable.

I would also like to express my sincere gratitude to the members in Dr. Jiang's lab. They offered me active spirit and lively scientific environment, also provided suggestions and collaboration for this study.

My special appreciation goes to my parents for their encouragement and guidance.

Finally, I would like to thank the Department of Biochemistry and Molecular Biology for this study opportunity.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The clip domain was first identified in the horseshoe crab proclotting enzyme (Muta, *et al.*, 1990) and was so named because it took the shape of a paper clip in a 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 (Jiang and Kanost, 2000). The 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) or serine proteinase homologs (SPHs). SPHs lack catalytic activity because one or more of the catalytic triad residues are substituted. Clip domain proteinases and SPHs are activated by cleavage between the clip domain and proteinase domain. After activation, the clip domain remains covalently attached to 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: around 80 genes in *Drosophila melanogaster* and *Anopheles gambiae* encode SP-like proteins, each with 1 to 5 clip domains (Ross, *et al.*, 2003; Christophides *et al.*, 2002). There are at least 17 SPs and 2 SPHs containing a clip domain in the tobacco hornworm, *Manduca sexta* (Jiang, *et al.*, 2005).

Clip-domain SPs mediate developmental signals and defense responses in arthropods. These include the horseshoe crab hemolymph coagulation, *Drosophila* Toll pathway activation during embryogenesis and induced synthesis of antimicrobial peptides, and

mosquito melanotic encapsulation of malaria parasites (Iwanaga, *et al.*, 1998; Belvin, *et al.*, 1996; Hoffmann, *et al.*, 2002; Christophides, *et al.*, 2004). One of the best studied immune responses in insects and crustaceans involves limited proteolysis of prophenoloxidase (proPO). Active phenoloxidase (PO) catalyzes quinone formation and melanin production for wound healing and microbe killing (Nappi *et al.*, 2005).

In some insects, proPO activation is mediated by a proPO-activating proteinase (PAP, also known as PPAE or PPAFI) in the presence of an SPH (Yu, *et al.*, 2003; Kwon, *et al.*, 2000). PAP and SPH each contain 1-2 regulatory clip domains at their amino-terminus (Jiang, *et al.*, 1998; Jiang, *et al.*, 2003a; Jiang, *et al.*, 2003b; Satoh, *et al.*, 1999; Lee, *et al.*, 1998). These molecules are terminal components of a branched SP cascade, analogous to blood coagulation and complement activation systems in mammals (Kim, *et al.*, 2002; Kanost, *et al.*, 2004). Recognition of pathogens by pattern recognition receptors activates the cascade via sequential proteolysis, and specific protein-protein interactions ensure a localized defense response.

A few biochemical experiments in an SPH have revealed possible roles of clip domains. In the proPO activation reaction, a clip domain in an SPH may provide a binding site for proPO: a mutation (Val111Ala) in *Holotrichia diomphalia* PPAFII (an SPH) abolishes its binding to proPO – this Val residue is strictly conserved among all the clip domains (Jiang, *et al.*, 2000). The inhibition of bacterial growth by the recombinant clip domain of *Pacifastacus leniusculus* PAP indicates its probable interaction with the bacterial surface (Wang, *et al.*, 2001). Deletion of the clip domain in *H. diomphalia* PPAFI (a SP) largely reduced its cleavage activity and amidase activity (Piao, *et al.*,

2005), suggesting that this domain is required for association with its activating proteinase.

The clip-domain SPs (cSPs) comprise three groups (1_a, 1_b and 2) whereas cSPHs represent lineage(s) diverged from cSPs during early evolution of this gene family (Jiang and Kanost, 2000; Ross, *et al.*, 2003). *M. sexta* PAP-2 belongs to group-2, the largest and most prevalent subfamily of cSPs.Group-2 includes *Drosophila* easter and PAPs from moths, beetles, flies and mosquitoes. To date, the only one tertiary structure of a clip domain available is for an SPH, *H. diomphalia* PPAFII (Piao, *et al.*, 2005). The clip domain crystal structure contains several loops, a four-stranded irregular β-sheet, and a prominent central cleft hypothesized to bind proPO. Because significant differences exist between SP- and SPH-associated clip domains (*i.e.*, length, sequence, cleavage site and interdomain relationship), we decided to investigate the structural basis for functions of SP-associated clip domains. In this work, we have determined the structure of dual clip domains from the PAP-2 by NMR spectroscopy. The PAP-2 clip domains adopt a novel α/β mixed fold with a β-sheet and hydrophobic pocket similar to the PPAFII clip domain. Our results provide insights into the structural basis of domain functions, and this structure may represent the prototype of group-2 clip domains.

CHAPTER II

REVIEW OF LITERATURE

The immune system protects organisms from infection with layered defenses of increasing specificity. First of all, physical barriers prevent pathogens from entering the body. If a pathogen breaches these barriers, the innate immune system provides an immediate response. In case, pathogens successfully evade the innate response, vertebrates possess a third layer of protection, the adaptive immune system. In contrast to the adaptive immune system specified only found in vertebrates, innate response is universal found in plants and animals.

Insects account for more than 70% of the animal species on the earth. While insects rely solely on innate immunity to defend against invading pathogens, diverse responses contribute to their evolutionary success. To better understand the role of innate immunity in human defense system and its evolution, people study defense mechanisms in several model insects, such as fruit fly, silkworm and tobacco hornworm. On the other hand, many parasites use insect vectors to transmit human diseases such as malaria. Melanotic encapsulation is a resistance mechanism against pathogen and parasites. For instance, *Plasmodium cynomolgi* ookinetes are encapsulated in the melanin coat in a refractory strain of mosquito, *Anopheles gambiae* (Collins, *et al.*, 1986). In order to disrupt the transmission, efforts have been made to study mosquito immune processes, particularly melanization, a phenoloxidase-mediated reaction.

Phenoloxidase (PO) and prophenoloxidase (proPO)

PO, a copper-containing tyrosinase, converts monophenols to *o*-diphenols and oxidizes *o*-diphenols to *o*-quinones (Mason, 1965). To minimize detrimental effects of PO-generated reactive intermediates to host tissues and cells, arthropod POs known so far are all produced as inactive proenzymes, named proPO, which require specific proteinases for proteolytic activation. In *M. sexta*, proPO contains two polypeptides: proPO-p1 (78 kDa) and -p2 (80 kDa) with 50% sequence identity. Various proPO have been isolated from many other insects, such as *Drosophila melanogaster* (Fujimoto, *et al.*, 1995), *A. gambiae* (Jiang, *et al.*, 1997), *H. diomphalia* (Kwon, *et al.*, 1997) and *Bombyx mori* (Kawabata *et al.*, 1995). Activation of proPO in insects is mediated by a serine proteinase cascade, analogous to the coagulation pathway and complement system in human plasma.

Different cleavage sites have been found in different proPOs. *B. mori* and *M. sexta* proPO are effectively activated by cleavage at Arg^{51} (Satoh, *et al.*, 1999; Yu, *et al.*, 2003). In contrast, Arg^{51} cleavage *H. diomphalia* proPO yields an inactive intermediate; a secondary cleavage at Arg^{161} is required for manifesting a PO activity (Kwon, *et al.*, 2000; Lee, *et al.*, 1998). Sequence alignment indicates that the Arg^{161} equivalent is present in *M. sexta* proPO, but not in *B. mori* proPO.

Serine proteinase cascade for proPO activation

The proPO activating proteinase (PAP) is the terminal component of a branched SP cascade. In *M. sexta*, three PAPs have been isolated: PAP1 from integument (Jiang, *et al.*, 1998), PAP2 and PAP3 from hemolymph (Jiang, *et al.*, 2003a; Jiang, *et al.*, 2003b).

Different terminologies are used for such enzymes in different arthropod species: PPAE in *B. mori* (Satoh, *et al.*, 1999) and PPAFI in *H. diomphalia* (Lee, *et al.*, 1998). *M. sexta* PAP1 and *H. diomphalia* PPAFI contain one clip domain at amino-terminus; whereas *M. sexta* PAP2, PAP3, and *B. mori* PPAE has dual clip domains.

Most upstream components of this serine proteinase cascade are unclear until recently. The pathway in *M. sexta* initiates with a hemolymph proteinase 14 (HP14), which autoactivates upon bacterial recognition (Wang and Jiang, 2006). HP14 cleaves proHP21 to form HP21, which then activates PAP2/PAP3 (Gorman *et al.*, 2007).

In some insects, proPO cleaved by PAP alone has no or low PO activity, and full activation require the presence of a cofactor, consisting of serine proteinase homolog (SPH). SPH activation involves another branch of serine proteinase cascade. In *H. diomphalia*, proteinase PPAFIII activates proPPAFII, an SPH (Kim, *et al.*, 2002).

The clip-domain serine proteinases (SPs) and their homologs

The clip-domain serine proteinases belong to the S1A family peptidases, containing a catalytic triad of His^{57}, Asp^{102}, and Ser^{195} (chymotrypsin numbering). In most cases, these residues are present in highly conserved sequence motifs of TAAHC, DIAL and GDSGGP. Substrate binding clefts near the active site largely determine the substrate specificity. The cSPs are classified into two major groups, the first one further is divided into two subgroups (1_a, 1_b) (Jiang and Kanost, 2000; Ross, *et al.*, 2003). PAPs belong to group-2, characterized by a short insertion in a region between the His and Asp residue of the catalytic triad. Two Cys residues in this region may form a disulfide bond.

Clip-domain SPs from different groups have different activation cleavage sites: group-1_a and group-2 members are cleaved after Lys/Arg residue, whereas group-1_b proteinases are cut after other residues (Ross *et al.*, 2003). In *A. gambiae*, cSPs and cSPHs, also known as CLIPs, are divided into four subfamilies (A-D) (Christophides, *et al.*, 2002). Most cSPHs belong to subfamily A, forming a separate lineage. The other three subfamilies B, C, D correspond to cSPs in group 2, 1_b, 1_a, respectively. *A. gambiae* subfamily B members share common features with PAPs. CLIPB14 and CLIPB15 isolated from hemocytes are involved in antibacterial immunity, particularly, in the killing of *Plasmodium* ookinetes (Volz, *et al.*, 2005).

The cSPs and cSPHs represent one of the most abundant groups of proteins in insects. The cSPs mediate horseshoe crab hemolymph clotting (Muta, *et al.*, 1990), *Drosophila* dorsal-ventral axis establishment (Belvin and Anderson, 1996), and melanization in insects and crayfish. Inhibition of serine proteinases by benzamidine leads to reduced induction of some antimicrobial proteins in *M. sexta* suggests their role in stimulating antimicrobial gene expression. In addition to the role in proPO activation as cofactors, cSPHs may be also involved in pattern recognition in crayfish (Huang, *et al.*, 2000) and somatic muscle attachment in *Drosophila* embryos (Murugasu-Oei, *et al.*, 1995).

CHAPTER III

METHODOLOGY

Protein expression and purification

The 5' fragment of *M. sexta* PAP-2 cDNA was amplified by PCR using the longest clone as template [14]. Primers j662 (5'-TTTG**CCATGG**GACAAGCCTGC-3') and j672 (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 restriction sites and a stop codon. The thermal cycling conditions were 35 cycles of 94°C, 30s; 53°C, 30s; 72°C, 40s, followed by 10 min of incubation at 72°C. After digestion with *Nco*I and *Hin*dIII, the 364 bp PCR product containing two clip domains was inserted to the same sites in plasmid H6pQE60 [28] 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.

E. coli strain M15 (Qiagen) containing H12/H6pQE60 was initially grown in 50 ml of LB medium (100 μg/ml ampicillin) at 30°C overnight. The culture was transferred into 1 liter of M9 minimal medium supplemented with 1g $^{15}NH_4Cl$, 3 g ^{13}C-glucose and 10 ml 10×^{15}N, ^{13}C-labeled Bio-Express-1000 (CIL). After growing at 37°C for 2 h, the culture was added to 20 liters of the fresh medium and continued to grow for about 4 h until A_{600} reached 0.7. Expression was induced with 1 mM isoproyl-β-D-1-thiogalactopyroanoside

for 4 h at 37°C. Cells were harvested by centrifugation at 10,000×*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, containing 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 18,000×*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 (18,000×*g*, 30 min), the supernatant was recovered and loaded onto a Ni-NTA agarose column (150 ml bed volume) and washed with 750 ml buffer A. The bound proteins were eluted with 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 C18 (4.6×100 mm)(Amersham Biosciences). In each run, 100 ml of the pooled sample was loaded and then 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. After electrophoretic analysis, fractions containing the monomer 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 then 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 sodium 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 eight hours.

All NMR spectra were acquired at 298 K on a Bruker Avance 800 MHz spectrometer 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 [29] and SPARKY [30] software packages, respectively. ^1H, ^{13}C and ^{15}N chemical shift referencing followed a previous report [31]. No stereo-specific 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 ^1H, ^{15}N and ^{13}C resonance assignments were based on the 3D CBCA(CO)NH, HNCACB, HNCA and HNCO. The side chain ^1H, ^{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 ^1H-^{15}N HSQC and 3D ^1H-^1H-^{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, 1.8-2.1Å; O-N, 2.7-3.0Å) were deduced from

backbone chemical shift analysis and from the 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' and backbone ^{15}N chemical shifts using TALOS [23]. 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 ends of the protein were more flexible as indicated by the lack of NOE correlations. Structures were calculated from extended backbone conformations as starting models using the program CNS [32]. 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 term 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.

Few NOE restraints were obtained in the link region between clip-1 and clip-2 from residues Pro66-Thr70, and inter-domain NOE restraints were not obtained. Since the exact position of the two domains in relation to each other could not be determined, structures calculated based on full-length template were divided into two single domains for further analysis. Ensemble RMSDs were calculated with MOLMOL [33]. The structure ensembles were analyzed and validated in terms of geometry and restraint violations with AQUA and PROCHECK-NMR [34]. 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 code 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).

CHAPTER IV

RESULTS AND DISCUSSIONS

Preparation of Clip

The two clip domains of *M. sexta* PAP-2, connected by a short linker region, were expressed as a soluble recombinant protein fused with an amino-terminal hexahistidine tag which facilitated its capture by Ni-NTA beads. Although the affinity-purified protein migrated as a single predominant band on the reducing SDS-polyacrylamide gel, electrophoresis under nonreducing condition 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 [22], the monomer eluted from the C18 column as a sharp peak followed by broader peaks containing other forms (Figure 1). From 20 liters of *E. coli* culture, we obtained 8.0 mg monomeric PAP-2 clip domains at a purity of over 95%.

NMR structure determination

Nearly 95% of the backbone and side-chain chemical shifts were assigned for residues 11-124, which are largely well-structured. Residues 1-10 (Met-His-His-His-His-His-His-Gly-Met-Gly, not a part of the first clip domain) were not included in the structure determination – the low signal-to-noise ratio in the NMR spectra for this region

was caused by rapid exchange of backbone NH and water protons and apparent structural flexibility. 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 our protein adopted one conformation. Based on the signal-to-noise ratio of the 3D ^{15}N- and ^{13}C-edited nuclear Overhauser effect spectroscopy (NOESY), a total of 1557 distance constraints were added (Table 1). In addition, 55 torsion angle backbone constraints derived using TALOS [23], 74 hydrogen bond restraints from ^{1}H-^{2}H exchange experiment, as well as confirmed disulfide bond connections (see below) were employed for the structure calculation. Secondary structural elements were determined by a combination of CSI (using H^{α}, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and ^{13}C' chemical shift) [24] and TALOS analysis. Strong $H_{\alpha,i}$-$H_{\beta,i+3}$ NOEs in the predicted regions support the existence of α-helices. The β-strand topology was determined by identification of inter-strand NOE contacts and analysis of slow-exchange amide protons. Results of these experiments indicate that the clip domain fold starts from two residues before Cys-1 and end at one residue after Cys-6 of each clip domain. As few NOEs were observed in the inter-domain region from Pro66 to Thr70, we could not determine the relative positions and orientations of two clip domains. Consequently, we separate the structure ensemble into two parts: clip-1 from Q11-P66 (56 residues) and clip-2 from L71-G124 (54 residues). Structural superposition and statistical evaluations are displayed separately. Figure 3 depicts superposition of backbone C^{α} atoms of the 20 energy-minimized conformers that represent the solution structure of clip-1 and clip-2.

The atomic root-mean-square deviation (RMSD) of the mean coordinates for the backbone nuclei of 20 conformers are 0.83Å and 0.94 Å for clip-1 and clip-2, respectively. These do not include the disordered regions (residues 15-20, 26-29, 39-42, 53-61 in clip-1; residues 75-80, 86-89, 98-101 and 112-119 in clip-2). An analysis of the 20 conformers with PROCHECK-NMR indicates that 97.2% (clip-1) and 99.3% (clip-2) of backbone dihedral angles are in allowed regions of Ramachandran plot (Table 1).

Disulfide bond confirmation

The C^{α} and C^{β} chemical shifts of cysteine residues are diagnostic of disulfide bond formation [25]. 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} shift ranging from 38.9 to 46.8 ppm, which clearly indicates the formation of disulfide bonds.

In the first calculation cycle, disulfide bond connection was not included, and only manually assigned NOE restraints and dihedral angle restraints were used – the preliminary structure shown in the ribbon diagram highly resembles the final set using a template with disulfide bond restraints (Figure 4). We then calculated the distances between sulfur atoms: in a distance smaller than 5Å, Cys-3 in each domain was close to Cys-6 (Cys29-Cys65 in clip-1 and Cys89-Cys123 in clip-2); Cys-1, -2 and -5 are all in proximity in this preliminary model whereas Cys-4 has a higher probability to be near to Cys-2.

Clear NOE cross-peaks from the β protons of Cys residues on ^{13}C-edited NOESY-

HSQC further confirmed the disulfide bond connection between Cys-3 and Cys-6 (Figure 5). We have also identified the other NOE correlations between the H_β from Cys-1 and Cys-5 (*i.e*. Cys13-Cys64 in clip-1 and Cys73-Cys122 in clip-2). The clear differences in H_β chemical shifts of the 12 cysteine residues allow unambiguous assignment for these NOEs and indicate that there is no inter-domain disulfide linkage. Although the overlaps in the H_β region of Cys-2 and Cys-4 prevent direct establishment 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 from previous reports [1, 21].

Solution structure of Clip reveals a novel fold

The clip domains adopt a compact mixed α/β fold, which comprises a three-stranded antiparallel β-sheet flanked by two α-helices (denoted as β1-β2-α1-α2-β3). The whole structure packs around a large internal cavity. The secondary structure elements, as defined by MOLMOL†, are β1 (Gln11-Thr14 in clip-1; Leu71-Leu74 in clip-2), β2 (Gly21-Ser25 in clip-1; Gly81-Asn85 in clip-2), α1 (Asp30-Lys38 in clip-1; His90-Glu97 in clip-2), α2 (Ala43-Ser52 in clip-1; Glu102-Ser111 in clip-2) and β3 (Lys62-Pro66 in clip-1; Ser120-Gly124 in clip-2). The secondary and tertiary structures of the two clip domains, which are 28% identical in sequence, closely resemble each other. One residue in α1 and one in loop-4 lead to the length difference between the two clip domains.

Each domain contains four loop regions: loop-1 between β1-β2, loop-2 between β2-

α1, loop-3 between α1-α2, and loop-4 between α2-β3. Except for loop-4, which is poorly defined due to a paucity of NOE observations, these loops appear to be considerably rigid. We detected many medium/long range NOEs between residues in loop-1, which confine β1-loop1-β2 as a β-hairpin conformation, and this loop forms a special protruding region on the protein surface (Figure 6). Loop-2 is stabilized by hydrophobic contacts with strands β2 and β3, as supported by the following long-range NOEs: Val26-Val63 and Cys29-Val63 in clip-1; Ile86-Val121 and Cys89-Val121 in clip-2. The structural rigidity plays a role in stabilizing the α/β anchor (see below). Loop-3 forms the center turn of the helix-turn-helix sub-structural component. While no inter-residue NOEs (except for sequential NOEs) were found, a conserved Pro residue (Figure 7) may be important in maintaining the position of the loop, separating the α-helices, and avoiding collapse of the internal cavity. This is consistent with the absence of NOE between the α-helices.

The two α helices are antiparallel to each other and almost perpendicular to the β-sheet (Figures 3 and 4). We observed a network of NOEs between the core residues, which indicate the formation of a compact structure named α/β anchor. On one side, close interactions between loop-2 and β3 [Cys29-Val63(clip-1)/Cys89-Val121(clip-2)], disulfide linkage [Cys29-Cys65 (clip-1)/Cys89-Cys123(clip-2)], and NOEs between loop-2 and α1 [Cys29-Ala32(clip-1)/Cys89-Leu92(clip-2)] define the relationship between α1 and the β-sheet around an anchor point of Cys29(clip-1)/Cys89(clip-2). Moreover, the NOE of Ala32-Val63 (clip-1)/Leu92-Val121 (clip-2) fixes the relative orientation of α1 to the β-sheet. On the other side, α2 anchors to the β-sheet directly

through its end residue Ser52 (clip-1)/Ser111 (clip-2), clearly supported by the NOE of Ser52-Cys64 (clip-1)/Ser111-Cys122 (clip-2). The NOE of Leu49-Val63 (clip-1)/Leu108-Val121 (clip-2) fixes the relative orientation of α2 to β-sheet.

The cysteine residues are absolutely conserved in clip domains (Figure 7). Based on the solution structure, the disulfide framework makes a primary contribution to the stability and compactness of the whole structure. The Cys-1 and Cys-5 pair compacts the β-sheet by connecting strand β1 to β3; the Cys-2 and Cys-4 pair contributes to the interaction between loop-4 and the α/β anchor; Cys-3 and Cys-6 pair supports anchoring of α1 to the β-sheet. The preference of cysteine residue in β-strand conformation [26] and relatively high conservation of sequences around the Cys residues [2, 3] lead to the assumption that the β-sheet is a conserved structural feature of clip domains. This is consistent with the presence of a similar β-sheet element in PPAFII clip domain (Figure 8) [21]. In our structure, four of the six Cys residues are the center residues in the β-strands: Cys-1 located in β1, Cys-2 in β2, Cys-5 and Cys-6 in β3. These correspond to the β-strands in PPAFII clip domain: Cys-1 in β0, Cys-2 in β1, Cys-5 and Cys-6 in β3.

One major difference in the clip domain structures of PAP2 and PPAFII lies between Cys-3 and Cys-4: in PAP-2 this region adopts a helix-turn-helix fold whereas in PPAFII it forms a long loop containing a small fragment of β-strand (β2-1). The β2-1 interacts with Region I (corresponding to the proteolytic activation site in cSPs) and prevents its cleavage – PPAFII is cleaved between Cys-3 and Cys-4 [21]. In contrast, PAP-2 cleavage activation occurs at a typical serine proteinase activation site, rather than within the clip domain, and its clip domains are intact. The sequence hypervariability observed in the

entire family of clip domains (Figure 7), suggests that structural variations in other subgroups may exist.

To investigate whether there are any entries in the Protein Data Bank with a similar fold to this ensemble, we submitted the coordinates to DALI servers [27] and did not find any closely related structures. The first structure of SP-associated clip domains, therefore, represents a novel fold.

Description of a potential binding site for PPO

A central hydrophobic cleft in the clip domain of PPAFII was suggested to be the binding site of proPO [21]. Our solution structure reveals an analogous internal cavity for proPO binding, which is formed by α1, loop3, α2, and β3 (Figure 6). The hydrophobic patch mainly consists of aliphatic side-chain of Cys29, Ala32, Ile35, Leu49, Ala53 and Val63 (clip-1)/Cys89, Leu92, Ile95, Leu108, Val112 and Val121 (clip-2). Although we do not have direct experimental evidence to support this notion, the comparison of this model with PPAFII clip domain reveals striking similarities – nearly identical residues are present in the 'clip cleft' region.

Aromatic residues with low thermal flexibility 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). Located in α2 and with the side-chain pointing outward, these aromatic residues could 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.

Positively charged surface from β-sheet portion would facilitate bacterial surface binding

The clip domain of PAP-2 has a protruding region located in the β-hairpin portion which forms a large, positively charged surface (Figure 9). 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.

Since no additional structural data are available for SP-associated clip domains, and residues contributing to this charge property are not conserved, it is unclear whether or not this specific structural feature is a unique feature of the PAP-2 clip domains. A similar positively charged surface may also exist in the crayfish clip domain, in line with its antibacterial activities. Further experiments are needed to test if the clip domains in *M. sexta* PAP-2 could bind bacterial surface components or kill bacteria.

Specific charged surface envisage to function as recognition site

Specific interactions between the clip domains and activator/cofactor may involve an unusual structural surface on the helix-turn-helix element. The charged regions in clip-1 and clip-2 could serve as docking sites for its interacting proteins with complementary electrostatic surfaces. Clip-1 has net positive charges comprised 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 to activator or cofactor. While the nature of interacting proteins is unclear at this moment, high ionic strength significantly slows proPO activation by PAP-2 in the presence of its cofactor (data not shown), suggesting that electrostatic interactions are important in the activation process. Therefore, we propose that electrostatic interactions are driving forces for the initial recognition of the clip domains with their partners and subsequent binding may involve hydrogen bonding and hydrophobic interactions also. The hypervariable sequence between Cys-3 and Cys-4 is likely responsible for recognition and binding specificity.

Flexibility in hinge region and possible function role

The relative positions of α-helices to β-sheet are restricted by two hinges: loop-2 and the end of α2 (Figure 4). The RMSD for superposition of all ordered fragments is higher than that for the β-sheet portion, indicating some flexibility in the hinge regions. While the orientation of individual α-helices to β-sheet is fixed by some hydrophobic interactions, the 'helix-turn-helix' element may move forward or backward relative to the β-sheet. Such movement may facilitate the exposure of the internal cavity for proPO binding. On the other hand, certain level of flexibility provided by loop-2 may assist protein interaction involving α1 – this rigid part becomes prone to conformation changes.

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 proPAP via electrostatic interactions. These specific interactions ensure that the defense response occurs in the vicinity of invading pathogens. PAP 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 activation by the PAP. These processes underscore the importance of clip domains, which remain connected to the catalytic domains after cleavage activation [2].

	'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≤|i-j|≤4)	320	188	130
Long range (|i-j|>4	300	156	144
Hydrongen bonds	74	38	36
Dihedral angle restraints	55	28	27
B. Ensemble statistics evaluation			
Restraint satisfaction (rmsd 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

Table 1. NMR structure determination statistics

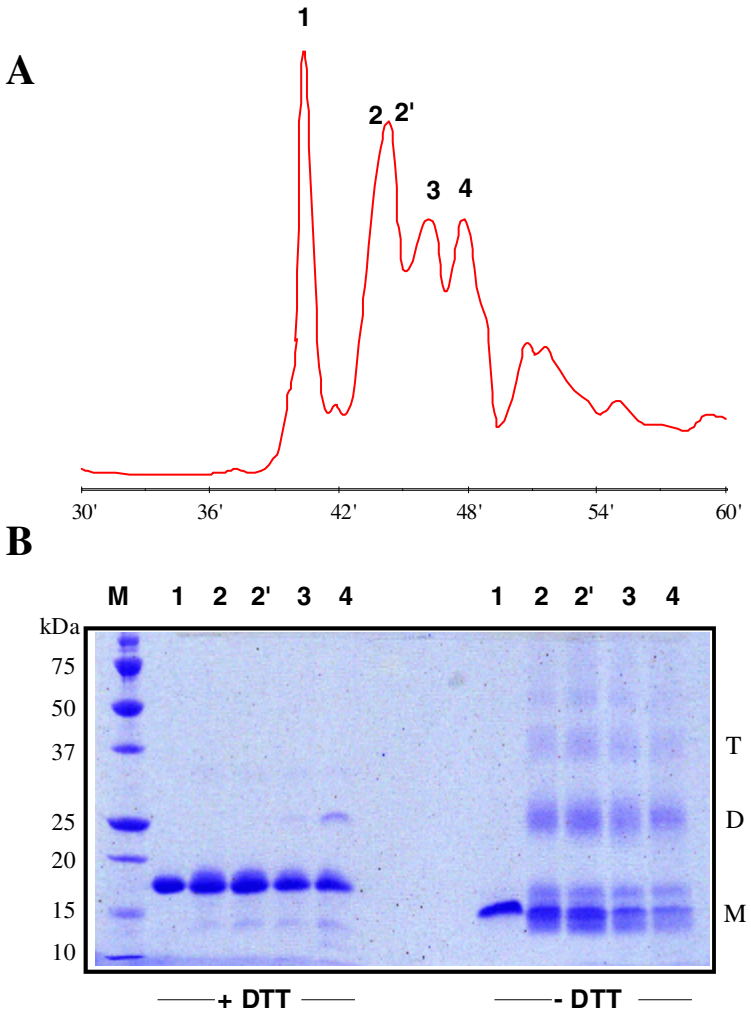

Figure 1. Different association states of the PAP-2 dual clip domains separated by reversed phase HPLC. (A) Elution profile: the affinity-purified recombinant protein was applied to a Source-15 RPC column equilibrated with 10 mM Tris-HCl, pH 8.5. A linear gradient of acetonitrile was employed to elute 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-PAGE 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 position of monomer (M), dimer (D), and trimer (T) are marked on the right.

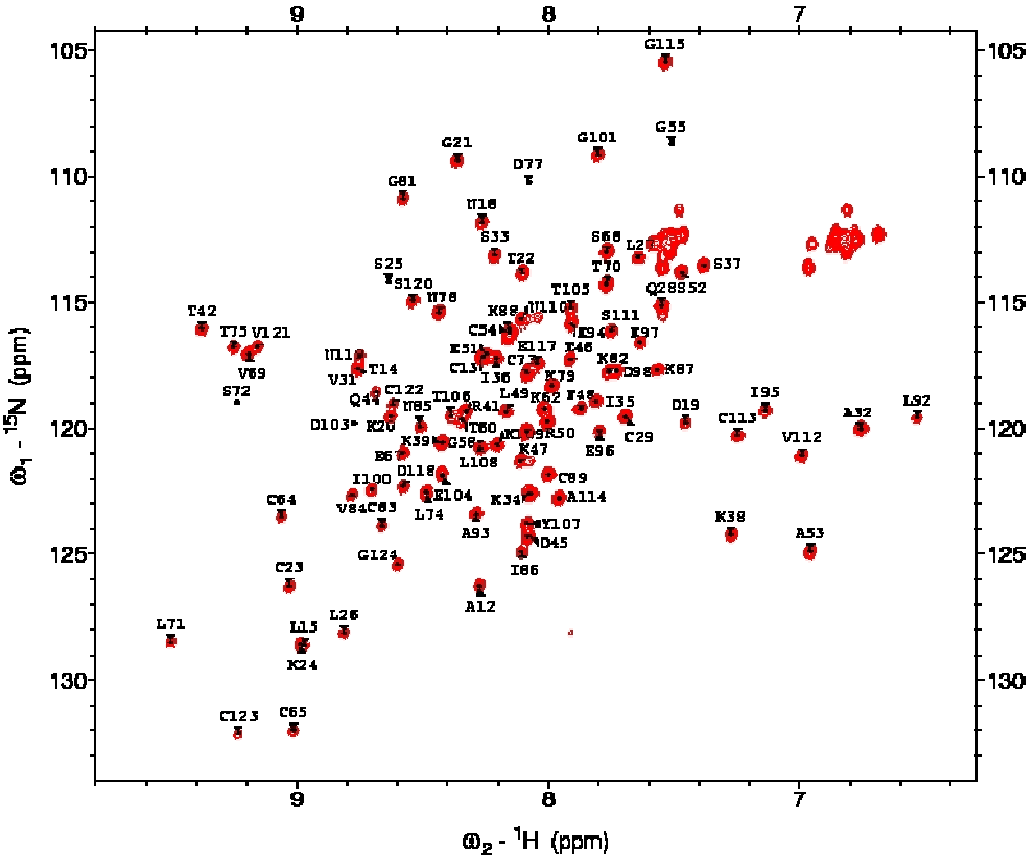

Figure 2. ^{1}H-^{15}N HSQC spectrum of the dual clip domains labeled with backbone assignments. The spectrum was collected from a ^{15}N, ^{13}C-labeled sample in 100 mM phosphate buffer (90% H_2O, 10% ^2H_2O) at pH 8.0. Assignments were made from residue Gln11 through residue G124. Gly124, located at the carboxyl terminus, is different from the other Gly residues.

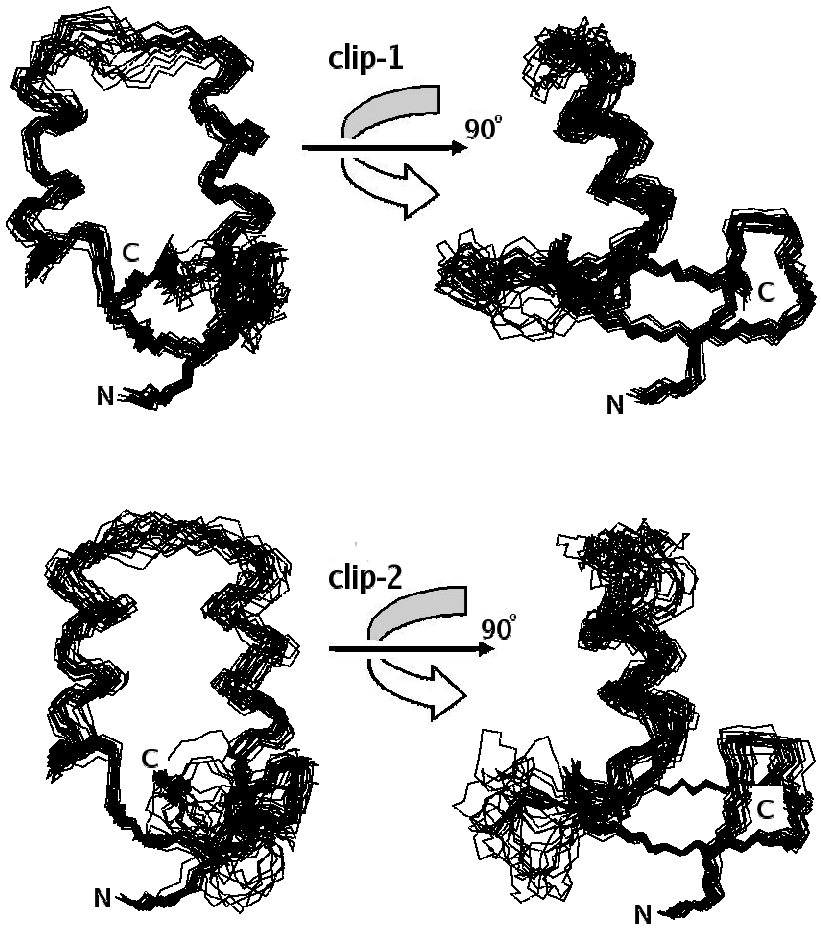

Figure 3. The solution structure of PAP-2 clip-1 and clip-2 presented as a bundle of 20 low-energy conformers. The well-ordered regions comprise residues11-14, 22-25, 30-38, 43-52, 62-65 in clip-1 and 71-74, 82-85, 90-97, 102-111, 120-124 in clip-2.

(A) (B)

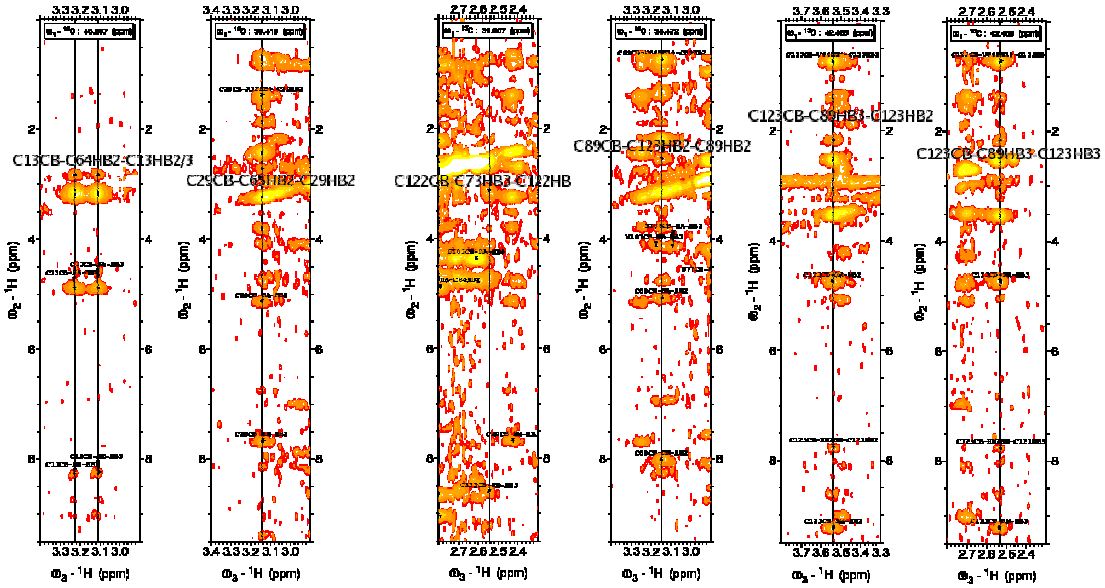

Figure 4. Spectral slides from ^{13}C-edited NOESY-HSQC showing NOE cross-peaks between H_β of the Cys residues in the two clip domains. (A) The cross-peak of H_β between Cys-1 (Cys13) and Cys-5 (Cys64) (a-1, Cys13-C) and between Cys-3 (Cys29) and Cys-6 (Cys65) (a-2, Cys29-C) in clip-1. (B) The cross-peak of H_β between Cys-1 (Cys73) and Cys-5 (Cys122) (b-1, Cys122-C) and between Cys-3 (Cys89) and Cys-6 (Cys123) (b-2, Cys89-C; b-3 and b-4, Cys123-C for H2 and H3, respectively) in clip-2.

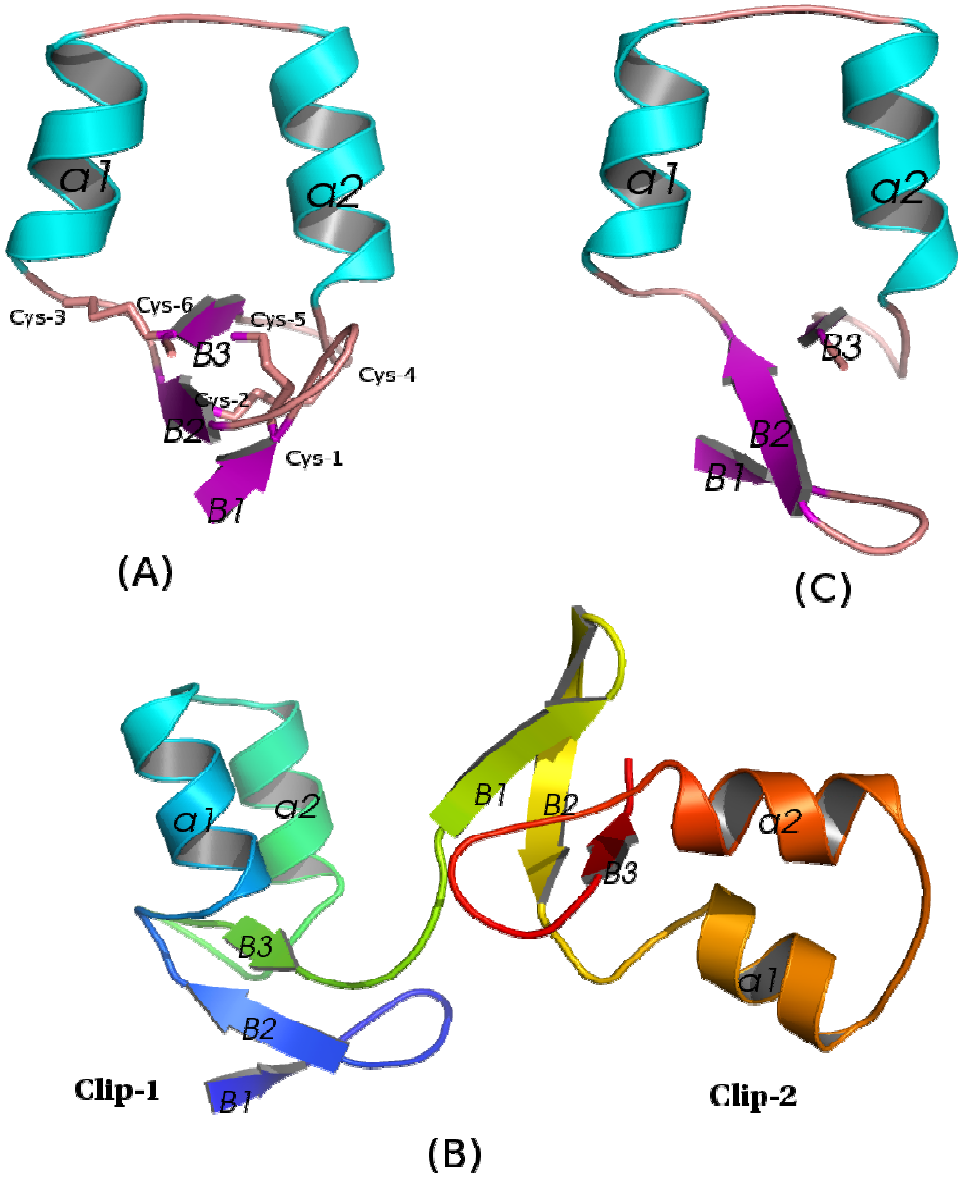

Figure 5. Ribbon diagram of the PAP-2 clip domains generated using PYMOL. (A) The averaged and energy-minimized clip-1 structure with the Cys residues and disulfide linkages shown. (B) A possible conformation of the full-length structure (residues 11-124). (C) Structural model of clip-1 calculated without disulfide bond constraints. In Panels A and C, the α-helices are colored in blue and β-strands in purple.

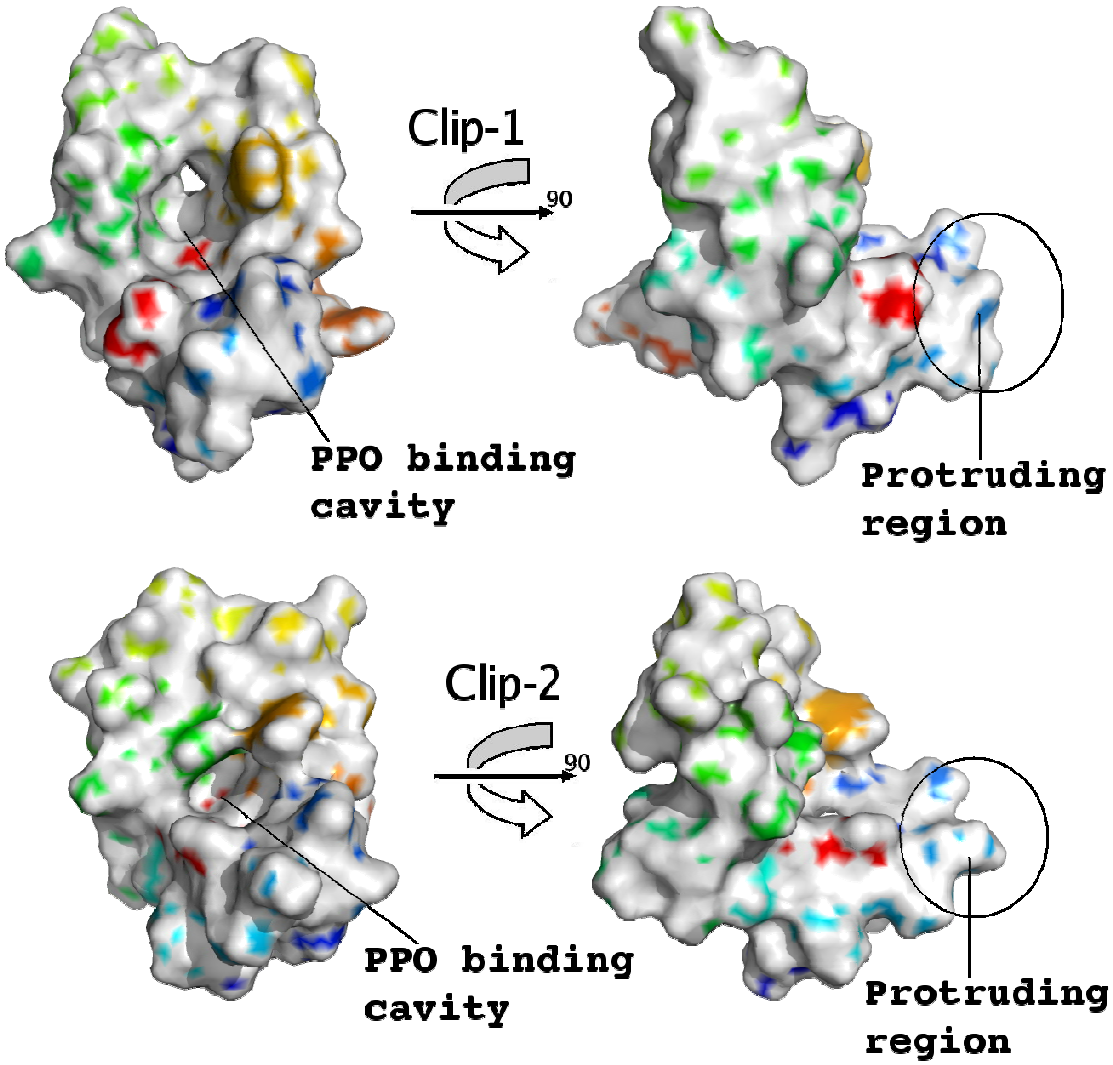

Figure 6. Surfaces of clip-1 (upper panel) and clip-2 (lower panel) displayed using PYMOL. These surfaces correspond to the orientations shown in Figure 2. Two surfaces on the left show the putative proPO-binding cavity; the ones on the right show the protruding regions.

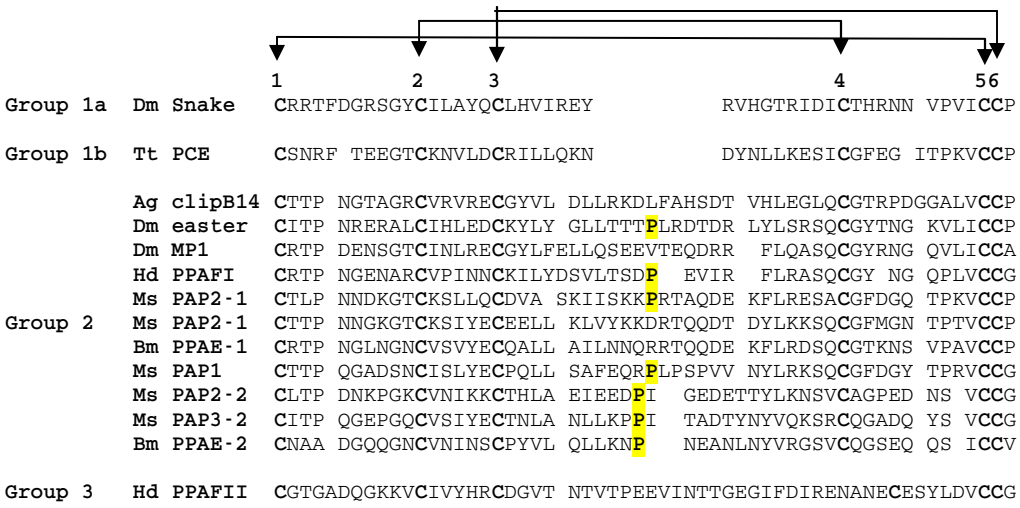

Figure 7. Alignment of the clip domain sequences from different subgroups of arthropod cSPs and cSPHs. The absolutely conserved Cys residues are shown in bold, whereas a commonly found Pro residue between Cys-3 and Cys-4 of group-2 cSPs is shaded yellow. Dm: *D. melanogaster*; Tt, *Tachypleus tridentatus*; Ag: *A. gambiae*, Ms, *M. sexta*; Bm, *Bombyx mori*; Hd, *H. diomphalia*.

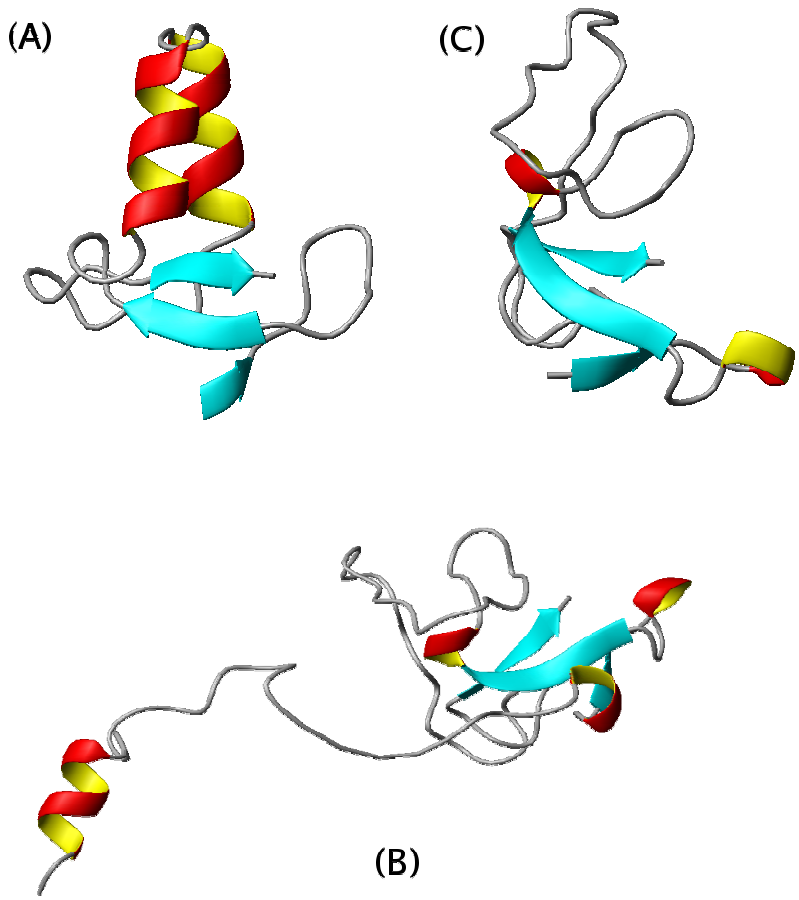

Figure 8. A structural comparison of the clip domains in PAP-2 and PPAFII using MOLMOL. (A) Ribbon diagram of the PAP-2 clip-1. (B) Structure of the PPAFII clip domain along with the amino-terminal extension (PDB 2B9L). (C) Ribbon diagram of PPAFII clip domain with its β-sheet highlighted.

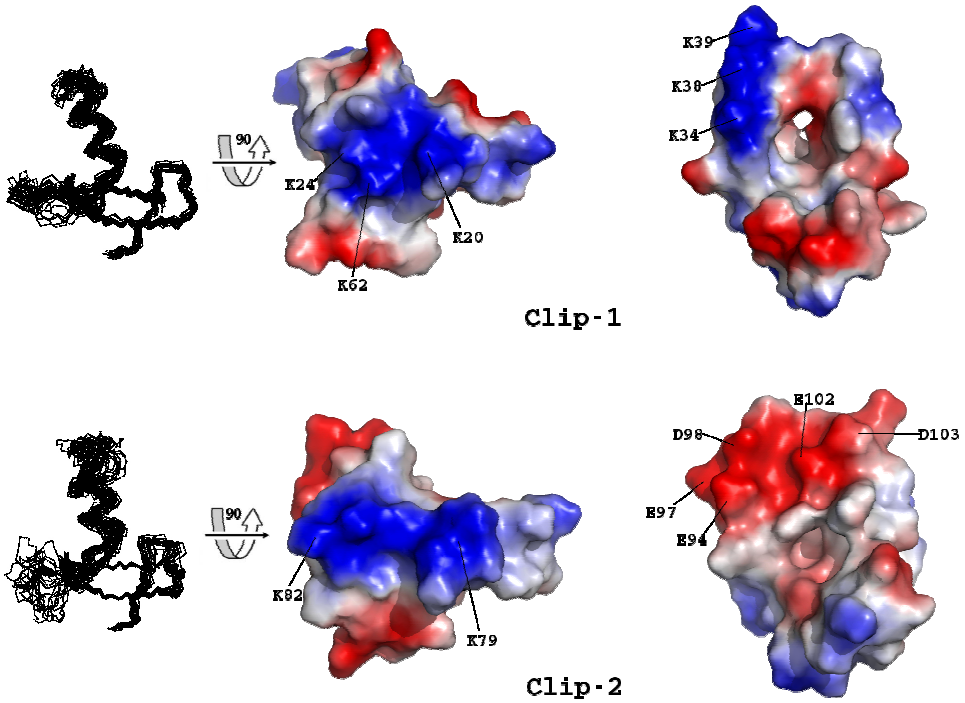

Figure 9. Electrostatic surface of clip domain-1 (upper row) and -2 (bottom row) in PAP-2 displayed using PYMOL. The electrostatic surface on β-sheet portion (middle panel), orientation shown in relative to the left panel. The electrostatic surface on 'helix-turn-helix' part (right panel) corresponds to the orientation shown in Figure 2, left panel. The surfaces are color-coded according to electrostatic potential: red, -10 kT; white, 0kT; blue, +10 kT.

REFERENCES

1. Arolas, J.L., Popowicz, G.M., Bronsoms, S., Aviles, F.X., Huber, R., Holak, T.A. and Ventura, S. (2005). Study of a major intermediate in the oxidative folding of leech carboxypeptidase inhibitor: contribution of the fourth disulfide bond. *J. Mol. Biol.* 352, 961-975

2. Belvin, M. and Anderson, K. (1996). A conserved signaling pathway: the *Drosophila* Toll-dorsal pathway. *Ann. Rev. Cell. Dev. Biol*. 12, 393-416

3. Bhattacharyya, R., Pal, D. and Chakrabarti, P. (2004). Disulfide bonds, their stereospecific environment and conservation in protein structures. *Protein Eng. Des. Sel.* 17(11), 795-808

4. Brunger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R.W. et al. (1998). Crystallography and NMR system: a new software suit for macromolecular structure determination. *Acta Crystallog. Sect.* D, 54, 905-921

5. Christophides, G.K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., Brey, P.T., Collins, F.H., Danielli, A., Dimopoulos, G. et al. (2002). Immunity-related genes and gene families in *Anopheles gambiae*. *Science* 298, 159-165

6. 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

7. Collins, F.H., Sakai, R.K., Vernick, K.D., Paskewitz, S., Seeley, D.C., Miller, L.H. (1986). Genetic selection of a *Plasmodium-refractory* strain of the malaria vector *Anopheles gambiae*. *Science* 234, 607-610

8. 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

9. Delaglio, F., Grzesiek, S., Vuisfer, 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

10. Fujimoto, K., Okino, N., Kawabata, S., Iwanaga, S. and Ohnishi, E. (1995). Nucleotide sequence of the cDNA encoding the proenzymes of phenol oxidase A1 of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* 92, 7769-7773

11. Goddard, T.D. and Kneller, D.G. (2002). SPARKY 3. University of California, San Francsico

12. Gorman, M.J., Wang, Y., Jiang, H., and Kanost, M.R. (2006). *Manduca sexta* hemolymph proteinase 21 activates prophenoloxidase activating proteinase 3 in an insect innate immune response proteinase cascade. *J. Biol. Chem.* In-press

13. Hoffmann, J.A. and Reichhart, J.M. (2002). *Drosophila* innate immunity: an evolutionary perspective. *Nature Immunol.* 3, 121-126

14. Holm, L. and Sander, C. (1997). Dali/FSSP classification of three-dimensional protein folds. *Nucl. Acids Res.* 25, 231-234

15. Huang, T.-S., Wang, H., Lee, S.Y., Johansson, M.W., Soderhall, K., Cerenins, L. (2000). A cell adhesion protein from the crayfish *Pacifastacus leniusculus*, a serine proteinase homologue similar to *Drosophila* masquerade. *J. Biol. Chem.* 275, 9996-10001

16. 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

17. Jiang, H. and Kanost, M.R. (2000). The clip-domain family of serine proteinases in arthropods. *Insect. Biochem. Mol. Biol.* 30, 95-105

18. Jiang, H., Wang, Y., Korochkina, S.E., Benes, H., Kanost, M.R. (1997). Molecular cloning of cDNAs for two pro-phenol oxidase subunits from the malaria vector, *Anopheles gambiae*. *Insect. Biochem. Mol. Biol.* 27, 693-699

19. 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

20. Jiang, H., Wang, Y., Yu, X.-Q. and Kanost, M.R. (2003a). Prophenoloxidase-activating proteinase-2 from hemolymph of *Manduca sexta*. *J. Biol. Chem.* 278, 3552-3561

21. Jiang, H., Wang, Y., Yu, X.Q., Zhu, Y. and Kanost, M.R. (2003b). Prophenoloxidase-activating proteinase-3 (PAP-3) from *Manduca sexta* hemolymph: a clip-domian serine proteinase regulated by serpin-1J and serine proteinase homologs. *Insect Biochem. Mol. Biol.* 33, 1049-1060

22. 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

23. Kanost, M.R., Jiang, H. and Yu, X-Q. (2004). Innate immune responses of a lepidopteran insect, *Manduca sexta*. *Immunol. Rev*. 198, 97-105.

24. Kawabata, T., Yasuhara, Y., Ochiai, M., Matsuura, S., and Ashida, M. (1995) A copper-containing protein homologous to arthropod hemocyanin. *Proc. Natl. Acad. Sci. U.S.A.* 92, 7774-7778

25. Kim, M.S., Baek, M.J., Lee, M.H., Park, J.W., Lee, S.Y., Soderhall, K. and Lee, B.L. (2002). A new easter-type serine proteinase cleaves a masquerade-like protein during prophenoloxidase activation in *Holotrichia diomphalia* larvae. *J. Biol. Chem.* 274, 7441-7453

26. Koradi, R., Billeter, M. and Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macro-molecular structures. *J. Mol. Graph.* 14, 51-55

27. 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

28. Kwon, T.H., Lee, S.Y., Lee, J.H., Choi, J.S., Kawabata, S., Iwanaga, S., and Lee, B.L. (1997). Purification and characterization of prophenoloxidase from the hemolymph of coleopteran insect, *Holotrichia diomphalia* larvae. *Mol. Cells*, 28, 90-97

29. Laskowski, R., Rullman, J., MacArthur, M., Kaptein, R. and Thornton, J. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, 8, 477-486

30. Lee, E., Linder, M.E. and Gilman, A.G. (1994). Expression of G-protein α subunit in *Escherichia coli. Meth. Enzymol.* 236, 146-163

31. 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 proteinase is induced by lipopolysaccharide or 1,3-β-glucan in coleopteran insect, *Holotrichia diomphalia* larvae. *Eur. J. Biochem.* 257, 615-621

32. Mason, H. S. (1965). Oxidases *Annu. Rev. Biochem.*, 34, 595-634

33. Murugasu-Oei, B., Rodrigues, V., Yang, X., and Chia, W. (1995). Masquerade: a novel secreted serine proteinase-like molecule is required fro somatic muscle attachment in the *Drosophila* embryo. *Genes Dev.* 9, 139-154

34. 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

35. 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

36. 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 proteinase and functional roles of the clip domains. *EMBO J.* 24, 4404-4414

37. Ross, J., Jiang, H., Kanost, M. and Wang, Y. (2003). Serine proteinases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationships. *Gene*, 304, 117-131

38. Satoh, D., Horii, A., Ochiai, M. and Ashida, M. (1999). Prophnoloxidase-activating enzyme of the silkworm, *Bombyx mori*. *J. Biol. Chem.* 274, 7441-7453

39. Sharma, D. and Rajarathnam K. (2000). ^{13}C NMR chemical shifts can predict disulfide bond formation. *J. Biomol. NMR*, 18, 165-171

40. Volz, J., Osta, M.A., Kafatos, F.C., and Muller, H-M. (2005). The roles of two clip domain serine proteinases in innate immune responses of the malaria vector *Anopheles gambiae*. *J. Biol. Chem.*, 280, 40161-40168

41. Wang, Y., Jiang, H. (2006). Interaction of beta-1, 3-glucan with its recognition protein activates hemolymph proteinase 14, an initiation enzyme of the prophenoloxidase activation system in *Manduca sexta*. *J. Biol. Chem.* 281, 9271-9278

42. Wang, R., Lee, S.Y., Cerenins, L. and Söderhäll, K. (2001). Properties of the prophenoloxidase activating enzyme of the freshwater crayfish, *Pacifastacus leniusculus*. *Eur. J. Biochem.*, 268, 895-902

43. 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

44. Wishart, D.S. and Sykes, B.D. (1994). The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR*, 4, 171-180

45. 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

VITA

Huang Rudan

Candidate for the Degree of

Master of Science

Thesis: THE SOLUTION STRUCTURE OF CLIP DOMAIN FROM *MANDUCA SEXTA* PROPHENOLOXIDASE ACTIVATING PROTEINASE 2

Major Field: Biochemistry and Molecular Biology

Biographical:

Personal Data:

Born in Jiujiang, China, on December 27, 1982, the daughter of Mr. Huang Weijia and Wu Changming

Education:

I passed the secondary school from the No.1 Middle School, Jiujiang, China on June 2000. Completed my B.S. degree in Biology from University of Science and Technology of China, Hefei, China. Completed the Requirements for the Master of Science degree with a major in Biochemistry and Molecular Biology at Oklahoma State University in May, 2007.

Experience:

I worked as a research assistant from October 2004 to May 2007, under the guidance of Dr. Haobo Jiang in the department of Biochemistry and Molecular Biology, Oklahoma State University. The title of the research project was the solution structure of clip domains from *Manduca sexta* prophenoloxidase activating proteinase 2.

Name: Huang Rudan Date of Degree: May, 2007

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: THE SOLUTION STRUCTURE OF CLIP DOMAIN FROM *MANDUCA SEXTA* PROPHENOLOXIDASE ACTIVATING PROTEINASE 2

Pages in Study: 37 Candidate for the Degree of Master of Science

Major Field: Biochemistry and Molecular Biology

Scope and Method of Study:

Clip domains are structural modules commonly found in arthropod serine proteinases and their homologs, which mediate extracellular signaling pathways of development and immunity. While little is known about their structures or functions, clip domains are proposed to be the 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 protease-2 by NMR.

Findings and Conclusions:

Each domain adopts a mixed α/β structure with a three-stranded antiparallel β-sheet flanked by two α-helices. The architecture provides structural insights into serine proteinase-associated clip domains for the first time. This novel fold includes a probable substrate-binding site, a putative bacteria-interacting region, and a remarkable charged surface for possible specific association with activator/cofactor. These results support the proposed roles of clip domains in prophenoloxidase activation and serve as a basis for structure-based probing of clip domain functions.

ADVISER'S APPROVAL: _____