Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Insect Science

Clip-domain serine proteases as immune factors in insect hemolymph

Michael R Kanost[1] and Haobo Jiang[2]

CLIP proteases are non-digestive serine proteases present in hemolymph of insects and other arthropods. They are composed of one or more amino-terminal clip domains followed by a linker sequence and a carboxyl-terminal S1A family serine protease domain. The genes for CLIP proteases have evolved as four clades (CLIPA, CLIPB, CLIPC, CLIPD), each present as multigene families in insect genomes. CLIP proteases in hemolymph function in innate immune responses. These include proteolytic activation of the cytokine Spätzle, to form an active Toll ligand leading to synthesis of antimicrobial peptides, and specific activation of prophenoloxidase (proPO), required for the melanization response. CLIP proteases act in cascade pathways. In the immune pathways that have been characterized, microbial surface molecules stimulate activation of an initiating modular serine protease, which then activates a CLIPC, which in turn activates a CLIPB. The active CLIPB then cleaves and activates an effector molecule (proSpätzle or prophenoloxidase). CLIPA proteins are pseudoproteases, lacking proteolytic activity, but some can function as regulators of the activity of other CLIP proteases and form high molecular weight immune complexes. A few three dimensional structures for CLIP proteases are now available for structure–function analysis of these immune factors, revealing structural features that may act in specific activation or in formation of immune complexes. The functions of most CLIP proteases are unknown, even in well studied insect species. It is very likely that additional proteins activated by CLIP proteases and acting in immunity remain to be discovered.

Addresses
[1] Department of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan, KS 66506 USA
[2] Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK 74078 USA

Corresponding author: Kanost, Michael R (Kanost@ksu.edu)

Current Opinion in Insect Science 2015, **11**:47–55

This review comes from a themed issue on **Molecular physiology**

Edited by **Naoki Yamanaka** and **Alexander S Raikhel**

For a complete overview see the Issue and the Editorial

Available online 21st September 2015

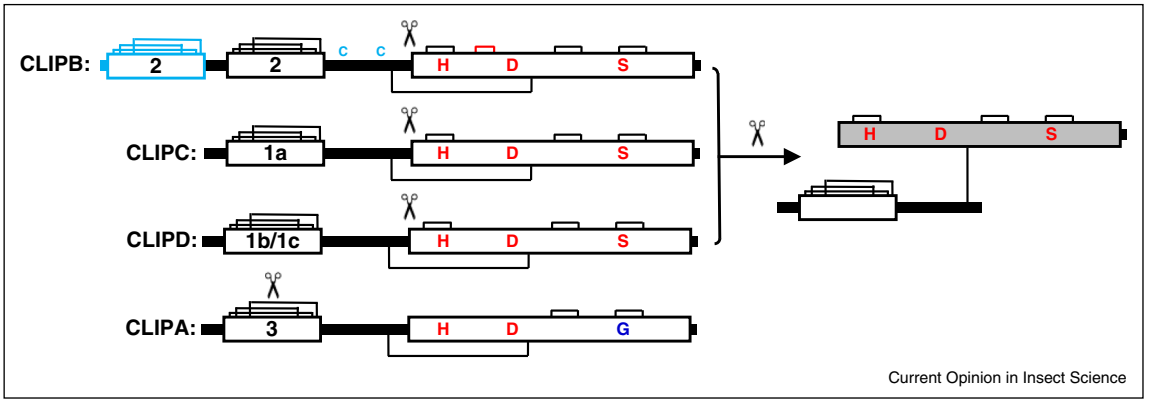
http://dx.doi.org/10.1016/j.cois.2015.09.003

Introduction

Specific proteolysis is the most prominent posttranslational modification of extracellular proteins and a mechanism for regulating their activity [1]. Extracellular serine protease pathways have evolved in animals to stimulate rapid responses to tissue damage and pathogen invasion [2]. Protease cascade pathways offer mechanisms for rapid, local amplification of a small initial signal, with regulation at multiple levels [3]. The proteases circulate as inactive zymogens and become sequentially activated upon recognition of aberrant tissues or microbial polysaccharides. Through specific molecular interactions and limited proteolysis, a localized reaction is rapidly initiated to stop bleeding, dismantle clots, or attack invading microorganisms. After accomplishing their functions, the active enzymes are inactivated by serine protease inhibitors, especially members of the serpin superfamily [4,5].

Protease cascade systems have evolved in innate immune systems of insects [6,7]. Biochemical, genetic, and molecular biological approaches have led to varying degrees of understanding of these pathways in a few insect species. Serine proteases (SPs) containing one or more amino-terminal clip domains [8] function in extracellular pathways that regulate some immune responses of insects (Figures 1 and 2). Clip domains were named because a diagram of the disulfide bond pattern in the N-terminal domain of horseshoe crab clotting enzyme [9] resembled a paper clip (Shun-ichiro Kawabata, personal communication). CLIP proteases (Snake and Easter) also participate in a protease cascade that regulates dorsal–ventral pattern in *Drosophila melanogaster* embryos [10]. The CLIP proteases represent a protein architecture apparently unique to invertebrates. They have been identified in arthropods and molluscs, and they form large gene families in the insect genomes studied so far (ranging from 15 CLIP protease genes in *Bombyx mori* to 42 in *Manduca sexta* and 63 in *Aedes aegypti*) [11–15,16••]. Infections can stimulate activation of CLIP protease zymogens in hemolymph, with specific cleavage at a site at the amino-terminus of the protease domain, creating a two-chain form active form of the enzyme, in which the clip domain and protease domain remain connected by a disulfide bond (Figure 1). Immune cascade pathways containing CLIP proteases lead to activation of prophenoloxidase (proPO) [17–26] or the Toll-ligand Spätzle [19•,27–30] (Figure 2). Once CLIP proteases are activated, they are regulated by serpin inhibitors in hemolymph plasma [31–34]. Some members of the CLIP superfamily contain a protease domain with mutations of one or more of the catalytic triad residues, such that they lack proteolytic activity. Such serine protease homologs (SPH) can function as cofactors required for proPO activation by an

Figure 1



Domain architecture of CLIP serine proteases. CLIP proteases contain one or more amino-terminal clip domains connected by a linker sequence to a carboxyl-terminal serine protease domain. The protease zymogen is activated by specific cleavage at the beginning of the catalytic domain. After this cleavage, the clip domain and protease domain remain connected by an interchain disulfide bond. CLIP proteases that have active sites with an intact catalytic triad (H, D, S) fall into three groups based on sequence alignments, known as clades B, C, and D. CLIPB proteases contain one or two amino-terminal clip domains from sequence type 2. CLIPB proteases include *Manduca* PAP1, PAP2, PAP3, HP8, *Bombyx* PPAE and BAEEase, *Holotrichia* PPAF1 and PPAF3, *Tenebrio* SPE, *Drosophila* SPE, MP1, MP2, easter, and Grass, *Aedes* IMP1, IMP2, TMP, B5, B29, B35, and *Anopheles* B4, B8, B9, B14. *Manduca* PAP2, PAP3, and *Bombyx* PPAE have two clip domains and two extra Cys residues in the linker (shown in light blue). CLIPC proteases, containing a single clip domain from group 1a include *Manduca* HP6 and *Drosophila* Persephone and Spirit. CLIPD proteases contain one clip domain from type 1b or 1c. At this time, there are no members of the CLIPD family with a known function. CLIPA pseudoproteases, known as serine protease homologs, have an amino terminal clip domain from type 3 and a protease-like domain in which the active site serine residue is changed to glycine, and therefore these proteins lack protease activity. CLIPA proteins are apparently activated by a specific cleavage in the clip domain. *Manduca* SPH1a, SPH2, and *Anopheles* CLIPA8 are examples of CLIPAs (for simplicity, additional Cys residues in some of the linkers and protease domains are not indicated).

active CLIP protease [35], and they can also negatively regulate the melanization response [36].

In *D. melanogaster*, *Ae. aegypti*, and *Anopheles gambiae*, genetic and RNAi analyses have demonstrated that certain CLIP proteases have a role in regulation of melanization or the Toll pathway (Figure 2), but substrates of the proteases and the identity of their activating enzymes are generally not yet known. Exceptions are *D. melanogaster* SPE, which activates proSpätzle [27,29], *D. melanogaster* SP7 (MP2), which can activate proPO [20•], and *An. gambiae* CLIPB9, which can activate *M. sexta* proPO [37]. In two beetles, *Tenebrio molitor* and *Holotrichia diomphalia* [33], and in a moth, *M. sexta* [34], biochemical studies have led to more detailed understanding of cascade pathways including CLIP proteases, which activate proPO and proSpätzle. In these systems, direct activation of downstream proteins by specific proteases has been achieved using purified proteins (Figure 2). However, it is clear that much remains to be learned about the functions of hemolymph CLIP proteases, with biochemical functions known for only a few members of this family.

Structure of CLIP proteases

Phylogenetic analyses of CLIP proteases have revealed four distinct clades in this family named A, B, C, and D [13,16••]. Detailed understanding of structure and function of CLIP proteases is currently limited by a lack of three-dimensional structural data for most members of this family. Available structures include the crystal structures of the *D. melanogaster* Grass zymogen, a CLIPB (Protein Data Bank accession 2XXL) [38••], the protease domain of *H. diomphalia* proPO-activating factor-1 (PPAF1), also a CLIPB (2OLG) [39], clip domain pseudoprotease *H. diomphalia* PPAF2, a CLIPA (2B9L) [40••], and the two clip domains from *M. sexta* proPO-activating protease-2 (PAP2, a CLIPB), solved by NMR spectroscopy (2IKD, 2IKE) [41].

The carboxyl-terminal catalytic or protease-like domains of CLIP proteases adopt a chymotrypsin fold consisting of two adjacent β-barrel-like structures arranged perpendicularly to each other (Figure 3). Each unit contains six antiparallel β-strands, with hydrophobic residues holding the barrels together at their interface. The catalytic residues (His, Asp, Ser) are in the cleft formed between the two barrels. Following chymotrypsin nomenclature, surface loops 30, 60, and 140 connect the secondary structure elements and control the access of protein substrates. CLIPBs possess an additional loop closed by a disulfide bond that does not exist in the other groups of CLIP proteases (Figures 1 and 3). This protruding structure may block access to the activation cleavage site, enhancing the specificity of zymogen activation [38••,39].

Clip domains are ~35–55 residue sequences with a conserved pattern of three disulfide bonds (Pfam accession