CHAPTER 11

CYSTEINE PROTEASES FROM BLOODFEEDING ARTHROPOD ECTOPARASITES

Daniel Sojka,[1] Ivo M.B. Francischetti,[2] Eric Calvo[2] and Michalis Kotsyfakis*,[3]

[1]Laboratory of Vector Immunology, Institute of Parasitology, Biology Centre of the Academy of Sciences of the Czech Republic, Ceske Budejovice, Czech Republic; [2]Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, Maryland, USA; [3]Laboratory of Genomics and Proteomics of Disease Vectors, Institute of Parasitology, Biology Centre of the Academy of Sciences of the Czech Republic, Czech Republic
**Corresponding Author: Michalis Kotsyfakis—Email: kotsyfakis@paru.cas.cz*

Abstract: Cysteine proteases have been discovered in various bloodfeeding ectoparasites. Here, we assemble the available information about the function of these peptidases and reveal their role in hematophagy and parasite development. While most of the data shed light on key proteolytic events that play a role in arthropod physiology, we also report on the association of cysteine proteases with arthropod vectorial capacity. With emphasis on ticks, specifically *Ixodes ricinus*, we finally propose a model about the contribution of cysteine peptidases to blood digestion and how their concerted action with other tick midgut proteases leads to the absorbance of nutrients by the midgut epithelial cells.

INTRODUCTION

Cysteine proteases are involved in different physiological procedures associated with the development and tissue homeostasis of hematophagous arthropod ectoparasites. Most of these proteases belong to the papain-like superfamily of cysteine proteases, while legumains/asparaginyl endopeptidases and caspases that play a role in mosquito biology have also been reported. Here we review the available information about cysteine protease activities in bloodfeeding arthropod ectoparasites with particular emphasis on their varied roles at the molecular level. The chapter is divided into three parts; cysteine protease activities in (i) ticks, (ii) mosquitoes and (iii) other bloodfeeding arthropods.

Cysteine Proteases of Pathogenic Organisms, edited by Mark W. Robinson and John P. Dalton.

CYSTEINE PROTEASES OF TICKS

Ticks are mites (Acari) classified to the suborder Ixodida. Approximately 850 species described worldwide are represented by two major tick families, Ixodidae (hard ticks) and Argasidae (soft ticks).[1] Both groups are obligate bloodfeeding ectoparasites of terrestrial vertebrates and are important vectors of human and animal diseases, transmitting viruses, bacteria (notably *Rickettsia*), protozoa and fungi.[2] Ticks express cysteine peptidases with important roles in two major physiological events that are crucial to the ectoparasitic lifestyle: digestion of host blood and embryogenesis.

Blood feeding and digestion are essential events for ticks. Vertebrate blood is a rich source of nutrients for energy metabolism to support their demanding activities. There are remarkable differences in bloodfeeding strategies between hard and soft ticks.[3] In soft ticks, nymphs and adults of both sexes feed rapidly and drop off the host within half an hour to hours. Feeding and oviposition of female soft ticks are repeated and feeding performance is not related to mating status. Virgin females may conserve nutrients upon fertilization that can occur in up to 150-200 days in *Ornithodoros moubata.*[4] Hard ticks feed only once in each life-cycle stage. Adult females take in a single large blood meal, lay down a single large clutch of eggs and then die.[5] The process of female feeding lasts for several days and consists of the slow feeding period (6-9 days) followed by rapid engorgement (12-24 hrs prior to detachment)[6] in mated females.[7,8]

Despite being bloodfeeding arthropods, ticks are believed to digest blood proteins intracellularly, within the endosomes of their gut cells. Thus, the gut lumen serves as their nutrient storage organ.[9] Digestive gut cells use both receptor-mediated and fluid-phase endocytosis to take up the liquid blood meal from the gut lumen.[5] Proteolytic degradation occurs in the endosomal pH environment, well below the pH 6.3-6.5 of the gut contents.[7] The acidic optimum of proteolysis in tick gut cells is around pH 3.0[10] and is consistent with the activity range of specific cysteine and aspartic protease activities in the midgut, as shown with specific substrates and inhibitors in *Boophilus (Riphicephalus) microplus*.[11]

Tick Cathepsin-Like Cysteine Proteases

Purified enzymes/activities and genes encoding for cysteine and aspartic proteases, both putatively involved in blood digestion, have been reported individually from several tick species in the last decade. Cathepsin L from *B. microplus* (BmCL1) was identified in 2000.[12] Later, RT-PCR and western blot profiling using antibodies against the *Escherichia coli*-expressed recombinant BmCL1 confirmed that the enzyme is solely expressed in the gut of tick feeding stages. BmCL1 is expressed as a 42 kDa precursor protein with a 23 kDa mature peptidase domain. Immunogold electron microscopy localized the antigens to secretory vesicles of gut epithelial cells; however, localization has been proposed from the gut ultrastructure without any additional histochemical markers.[13] Predominant gut cysteine peptidases of 40 kDa and 48 kDa were identified in another hard tick species, *Haemaphysalis longicornis*, by performing inhibition studies against the gelatin substrate.[14] Two mRNAs encoding proteins with homology to cathepsin L were also identified by RT-PCR using cDNA isolated from *H. longicornis* gut.[15] Recently, a gut-associated *H. longicornis* cathepsin L (HlCPL-A) was prepared as recombinant enzyme in *E. coli*, purified as both zymogen and mature enzyme and functionally characterized.[16] Expression of the HlCPL-A gene was found to

be up-regulated during feeding on vertebrate hosts. Recombinant HlCPL-A showed an optimum activity at pH 3.6 against both fluorescent peptidyl and bovine haemoglobin substrates.[16] Moreover, an individual enzyme classed as a cathepsin B, termed "longipain", was described from *H. longicornis* and proposed to have a critical role in the transmission of *Babesia spp.*[17] According to the authors, the longipain message is solely expressed in the gut tissue and up-regulated during tick feeding. Antibodies recognize the 39 kDa mature domain of the enzyme in immunoblots. Longipain appears to be expressed in the lysosomal vacuoles and gut cell surface and possesses the "occluding loop" that is typical of cathepsin B enzymes and necessary for its exopeptidase activity. Longipain expressed in *Pichia pastoris* showed unusually different pH optima for ZFR-AMC and ZRR-AMC peptidyl substrates. RNAi knockdown of longipain demonstrated its role in blood digestion and transmission of *Babesia sp.*[17]

Tick Asparaginyl Endopeptidases

The CD clan member *Ixodes ricinus* asparaginyl endopeptidase/legumain (IrAE) is the first member of its class described in arthropods.[18] It was cloned from a gut cDNA library of partially engorged females. Semi-quantitative RT-PCR profiling of IrAE mRNA levels in different tissues from feeding female ticks revealed that the enzyme message was expressed specifically in the gut. Immunofluorescence microscopy with specific serum raised against the *E. coli*-expressed recombinant zymogen localized IrAE in the digestive vesicles of gut cells and to the peritrophic matrix. Expression of the IrAE protein on the gut cell surface was further confirmed by immunogold electron microscopy. Recombinant IrAE produced in *P. pastoris* resembled its ortholog from the helminth parasite *Schistosoma mansoni* (SmAE) in its ability to auto-catalytically activate and to process schistosomal cathepsin B to its mature form. It has a strict specificity for asparagine at P1 and both recombinant and native enzymes show a drop of activity dependent on pH values (the enzyme loses its activity at pH values greater than 6.0).[18] Two comparative studies on the specific inhibition of IrAE and SmAE have been published. A set of novel selective legumain-specific inhibitors, the aza-peptide Michael acceptors and epoxides, show very similar preferences for both the helminth and tick enzymes, suggesting their similar substrate preferences/enzymatic roles.[19,20] Two different legumains/AEs have been also described in *H. longicornis*; HlLgm1[21] and HlLgm2.[22] Based on the collective findings of the authors, the HlLgm1 and HlLgm2 expression profiles[23] are associated with the tick gut (where they have also been localized by immunohistochemistry) suggesting a role in blood feeding. The authors subsequently expressed an active recombinant enzyme in *E. coli*. However, some of the biochemical properties of the recombinant protein, including a pH optimum of about pH 8.0, were quite unusual for legumains and most likely will require further investigation. In a more recent analysis,[24] RNAi knockdown of HlLgm1 and HlLgm2 gene expression resulted in tick rejection from the host before the ticks reached repletion; a significant reduction in engorged tick body weight was also observed. Silencing legumain gene expression by RNAi also had a major impact on embryogenesis by delaying oviposition (characterised by a reduced number of eggs and structurally deformed eggs that failed to hatch). According to Alim et al,[24] HlLgm1 and HlLgm2 play a role in gut cell proliferation and morphology as demonstrated by the disruption of developmental events within the midgut cells and certain damage of midgut tissue in legumain gene-silenced ticks.

Mechanism of Haemoglobin Digestion by Tick Proteases

Based on the numerous individual reports about tick digestive cysteine peptidases and keeping in mind that gut-related aspartic[25,26] and leucine[27] aminopeptidases have been reported from ticks, it is evident that degradation of host proteins in tick gut cells is a complex process that is not performed by a sole "hemoglobinase" enzyme activity. Rather, the digestive process relies on a network of gut-associated cysteine and aspartic proteases analogous to what has been well studied in *S. mansoni*[28,29] and other flatworm parasites.[30] The digestive network consists of key cysteine endopeptidases of the papain family (clan CA) cathepsins B, C and L, asparaginyl endopeptidase (clan CD) and the aspartic peptidase cathepsin D (clan AA). Furthermore, the presence of a multienzyme complex of digestive cysteine and aspartic peptidases was supported by analysis of the first tick gut expressed sequence tag (EST) project by Anderson et al.[31] The analysis of 1,679 ESTs from midgut cDNA libraries of *Dermacentor variabilis* female ticks (the mialome, as it was named) at varying stages of feeding was performed. Fourteen different cysteine peptidase transcripts were found; of the 24 total cysteine peptidase ESTs, 19 ESTs were found in the 6-day-fed female guts, whereas only 5 ESTs were found in the 2-day-fed library, further confirming that expression of midgut cysteine peptidases is induced relatively late during hard tick feeding, in agreement with previous reports on individual cysteine peptidases. Phylogenetic analysis of these sequences identified three major groups of cysteine peptidases: legumain-like, cathepsin B-like and cathepsin L-like.[31]

To fully address the hypothesis that a conserved multienzyme network, comprised of cysteine and aspartic peptidases, functions to digest host blood in ticks and helminthes, a wide screening for the (i) expression and (ii) activity of schistosomal enzyme homologues was performed. A single tick species at a certain bloodfeeding stage was selected; the partially engorged females of *I. ricinus,* the major Lyme disease vector in Europe. (i) Reverse genetics/PCR-based screening with gut cDNA template and primers that amplify conserved protease domains revealed expression of mRNA sequences encoding cathepsins B, C and L (IrCB1, IrCC1, IrCL1), the aspartic peptidase (IrCD1) and the asparaginyl endopeptidase (IrAE1).[18,26] RT-PCR profiling revealed their association to gut tissue and, importantly, their simultaneous expression in tick gut cells upon feeding. (ii) Recent biochemical screening uncovered the entire enzymatic core of hemoglobinolysis in the gut cells of *I. ricinus* feeding females.[32] The activity-based profiling employed enzyme-specific substrates and inhibitors to define endogenous peptidase activities in tick gut tissue extracts (GTE). The overall hemoglobinolytic activity measured in GTE during feeding is highly elevated from the 6th day after attachment to the host, which corresponds to partially engorged ticks. Authentic enzyme activities associated to coding sequences for IrCB1, IrCL1, IrCC1, IrAE1 and IrCD1[26] were clearly detected and their optimal pH working conditions were estimated. Their exclusive role in tick gut hemoglobinolysis was confirmed by zero hemoglobin degradation in assays when a set of inhibitors targeting the full spectrum of the identified peptidases was applied. Moreover, all the endogenous forms of the described enzymes were visualized by activity-based probes on SDS-PAGE gels. Their dominant abundance in GTE was also confirmed by mass spectrometry. For hemoglobinolytic studies, bovine haemoglobin was incubated with GTE in the presence of various protease-specific inhibitors. Haemoglobin fragments were purified and analyzed by mass spectrometry. A specific haemoglobin cleavage map with marked cleavage sites for each studied protease type was created.[32] Collectively, the large amount of data obtained from the reverse genetics approach and the biochemical

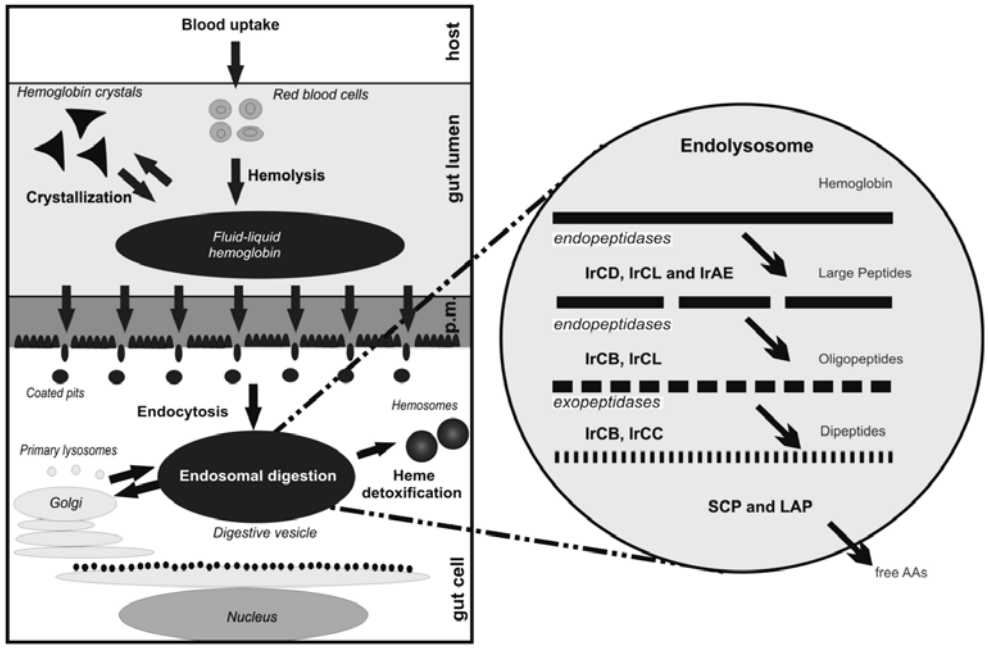

Figure 1. Schematic image of the uptake and digestion of host hemoglobin by *Ixodes ricinus* female ticks. In focus: the model of the proteolytic degradation cascade of host hemoglobin in endosomes of the gut cells as proposed by Horn et al[32] with marked positions of the cysteine and aspartic peptidases identified from the gut cDNA.[26]

mapping of tick gut hemoglobinolysis has resulted in the proposal of a mechanistic model for hemoglobinolysis in *I. ricinus* midgut cells (Fig. 1). Upon fusion of endosomes with lysosomes that carry the proteolytic machinery, IrCD endopeptidases, supported by IrCL and IrAE are responsible for the primary globin cleavage. Production of small secondary fragments is performed primarily by the IrCB. The endopeptidase-cleaved peptides are further processed by the exopeptidase (carboxy-dipeptidase) activity of IrCB and (amino-dipeptidase) IrCC. Single amino acids are then released from the resulting dipeptides by leucine or serine monopeptidases.[32]

Most of the research reviewed here contributes to the understanding of the function of the digestive cysteine proteases in hard ticks. Significantly less has been reported about the soft ticks, although the presence of similar enzyme activities, namely cathepsin B, C and legumain/AE, has been reported in *O. moubata.*[33]

The Role of Cysteine Proteases in Tick Embryogenesis

The second major physiological event in the tick body where cysteine proteases have been reported to play a role is embryogenesis, specifically the degradation of yolk proteins. During embryonic development, ticks rely on yolk reserve substances, mainly vitellin. Degradation is limited to the acid environment of the yolk spheres (specialized organelles that are considered as lysosomal-like organelles) since they contain both substrate (vitellin) and the degradative machinery. There is limited data on the particular enzymes involved in vitellin processing. Except for the reported yolk-degrading aspartic

peptidases, the yolk proteolytic activity has been essentially attributed to a cathepsin L-like enzyme through substrate and inhibitor specificity studies. In the soft tick *O. moubata*, the activity has been demonstrated as two bands of 37 kDa and 39 kDa on gelatin gels with maximum degradation efficiency at pH 3-4. At neutral pH, the protein binds vitellin.[34] In a later study it was shown that the enzyme is stored as an inactive zymogen that can be activated by a low pH environment.[35] However, molecular or biochemical identification of the specific protease activity, e.g., based on cDNA or mass spectrometry data, is still missing. This mechanism seems not to be soft tick specific, as a vitellin-degrading cysteine endopeptidase (VTDCE) has been identified from eggs of the hard tick *B. microplus*, as well.[36] Purified endogenous VTDCE is tightly associated to vitellin, is inhibited by E64 and appears as a dimer of 17 kDa and 22 kDa. Vaccination of bovines with purified VTDCE induces a partial protective immune response against *B. microplus* infestation.[37] A further vitellin-degrading enzyme (RmLCE) was purified and identified from larval extracts of *B. microplus* larvae.[38] It has been proposed that VTDCE is a maternally-derived protease with restricted specificity and that RmLCE is synthesized in larvae to complete hydrolysis of the remaining vitellin peptides. Such an enzyme might help tick survival by providing amino acids for protein catabolism enabling larval survival until the first blood meal is achieved.

Unlike the digestive machinery, the yolk-processing cascades have remained conserved among bloodfeeding insects and ticks. Based on the similarities with gut cells, i.e., receptor-mediated endocytosis followed by subsequent endosomal degradation of vitellin, we speculate on the existence of a similar degradation enzyme matrix of cysteine and aspartic peptidases in guts and eggs/larvae. Different enzyme isoforms, possibly mutated duplicates of ancestral single-copy genes, encoding peptidases with varying substrate specificity for vitellin should be identified and their specific role confirmed by expression profiling, RNAi knock-down and proteomic analyses.

CYSTEINE PROTEASES IN MOSQUITOES

Cysteine Proteases and Mosquito Embryogenesis

An unusual cathepsin-B-like cysteine protease has been cloned and characterized from the yellow fever- and malaria-transmitting mosquito *Aedes aegypti*.[39] It is named vitellogenic cathepsin B (VCB) and it is produced in the fat body of vitellogenic female mosquitoes after blood feeding. Exclusively secreted by their fat body as a latent proenzyme, it is stored in yolk bodies of the developing oocytes in the inactive proenzyme form of 44 kDa.[39] The kinetics of its secretion by the vitellogenic fat body are similar to those of the yolk protein precursors and its synthesis does not take place in the ovaries. Instead, the protein is taken up by the ovaries through endocytosis, again similar to the yolk protein precursors. Upon embryogenesis it is processed to a mature 33 kDa VCB that is active against Z-Arg-Arg-pNA, a cathepsin B-specific substrate. The fat body-secreted, hemolymph form of VCB cannot be activated by acidic pH alone through autocatalysis, but acidic pH was sufficient for the activation of VCB in ovarian extracts, suggesting a physiological mechanism that ensures uptake of VCB by the oocytes in the latent, inactive proenzyme form and ensures its activation only in the mosquito embryo.[39] There, VCB degrades vitellogenin, the major yolk protein precursor. Vitellogenin degradation is inhibited by the thiol protease inhibitor E-64, further confirming the involvement of

cysteine protease activity. Finally, the addition of an anti-VCB antibody to the embryonic extract prevented cleavage of vitellogenin, strongly indicating that VCB is involved in embryonic degradation of vitellin in vivo. In conclusion, VCB, similar to other key proteins in mosquito embryogenesis, is produced by an extraovarian tissue (fat body) and is endocytosed by the follicular cells, where it accelerates yolk protein degradation, a key event for the production of nutrients for the developing embryo.[39] An orthologous or homologous cathepsin in the mosquito *Anopheles gambiae* shows a blood-inducible pattern of expression in the fat body by quantitative real-time PCR (qPCR) analysis with a maximum accumulation of 35- to 74-fold, 24 hrs. after every blood meal.[40]

Finally, a study on the role of cathepsins in the development and maturation of the ovaries of the mosquito *Culex pipiens pallens* showed that cathepsin B- and L-like proteinases gradually accumulate in the developing ovaries after a blood meal and that they require more than 10 min of preincubation under acidic conditions to reach their maximum activities.[41] In contrast, homogenates of degenerating follicles, 3 days after blood feeding, display proteolytic activities without any acid treatment, suggesting that these proteases are already active in the mosquito ovaries at that time point and that they remained as inactive forms in the mosquito ovaries immediately after mosquito blood feeding. Chemical and immunohistochemical analyses also showed that, 3 days after blood feeding, more proteinases were located in the cytoplasm, rather than being associated with yolk granules.[41] In conclusion, these ovarian cysteine proteases from *C. pipiens*, apart from being activated at the onset of embryogenesis, are also activated during oogenesis, presumably to enhance oosorption, i.e., degeneration of some follicles in the developing ovaries. In this way, oosorption may enable females to recycle nutrients to developing oocytes when nutrient availability is poor.

Mosquito Caspases

The development and tissue homeostasis of *Aedes aegypti* is also safeguarded by the regulated activity of its caspases, key enzymes that serve as the initiators and executioners of apoptosis throughout the animal kingdom.[42] These enzymes fall into two categories depending on the size of their prodomain. Those caspases containing long prodomains (caspase recruitment or death effector domains) are called initiator caspases, because they activate the early apoptotic cascades. Those with short prodomains are called effector caspases, as they are involved in downstream apoptotic steps such as cleavage of the substrates/cellular proteins required to execute cell death (protein kinases, chromatin modifying enzymes and DNA repair proteins).[42] There is a key difference between *Drosophila* and vertebrates in the regulation of caspase activity; insect caspase activity is almost constitutive, abundant in all insect cells and inhibited by members of the inhibitors-of-apoptosis family (DIAP1: *Drosophila* inhibitor of apoptosis 1) in the cells that do not have to undergo apoptosis. In vertebrates, caspase activation requires the presence of death signals in the environment of the cells that will undergo apoptosis and members of the vertebrate inhibitors-of-apoptosis family contribute to apoptosis termination.[42] Although most information about insect caspases comes from studies in *Drosophila*, the cloning and functional characterisation of initiator caspases (named *Aedes* Dredd and *Aedes* Dronk, respectively) from the mosquito *Ae. aegypti* has been reported.[43,44] *Aedes* Dredd (*Ae*Dredd) contains two N-terminal death effector domains and the well-conserved caspase catalytic domain. Multiple sequence alignments, homology modeling and functional substrate assays of the recombinant protein suggest that *Ae*Dredd

is an ortholog of *Drosophila* Dredd and human caspase-8, all key players in the death receptor-mediated apoptotic pathway.[43] qPCR analysis revealed low levels of *Ae*Dredd transcripts throughout the body of the mosquito and throughout its development. Transcript levels in callow pupae were five- to-six-fold higher than in other developmental stages. In adult mosquitoes, transcript levels were the highest in the fat body tissues, up to four-fold higher than the levels found in salivary glands and ovaries. *Ae*Dredd transcripts were also two-fold higher in midgut tissues when compared with those in salivary glands and ovaries.[43] Exposure of third instar larvae for 24 or 48 hrs to ecdysone did not significantly change *Ae*Dredd transcription, but 30 min exposure of adult midguts to ultraviolet light induced a six-fold higher *Ae*Dredd transcription level compared with the control.[43]

Aedes Dronc (*Ae*Dronc) is another initiator caspase in *Ae. aegypti* that is predicted to contain an N-terminal caspase recruitment domain and the conserved caspase subunit, which has a unique sequence, SIRCG, surrounding the catalytic Cys.[44] Molecular modeling, phylogenetics, sequence comparison analyses and its substrate specificity show that it is the mosquito homolog of *Drosophila* Dronc and human caspases-2 and -9. *Ae*Dronc appears to be involved principally in insect development; qPCR revealed that *Ae*Dronc transcript levels in third instar larvae were 26-fold higher than those in the early instar stages. Transcript levels in fourth instar larvae were 12-fold higher than those in the third instar larvae, while transcript levels in callow and black pupae were 126 and 53-fold higher, respectively, than those in the third instar larvae.[44] Thus, the highest levels of transcripts are detected in late instar larvae and early and late pupae, which corresponds to tissue reorganization events and pulses of the steroid hormone ecdysone. Indeed, a six-fold increase in *Ae*Dronc transcripts was observed after 24 hrs of exposure of third instart larvae to ecdysone and a 16-fold increase after 48 hrs of exposure. An analysis of total cellular protein from larvae treated with ecdysone showed an increase in both *Ae*Dronc-specific activity and the downstream, effector caspase-like activity. The *Ae*Dronc-specific activity increased at 24 hrs and remained relatively constant up to 48 hrs, whereas the more generic effector caspase-like activity steadily increased up to 48 h.[44] Overall these data demonstrate that exposure of third instar larvae to ecdysone results in a significant increase in both transcript levels and caspase activity, suggesting that *Ae*Dronc may play a role in the tissue reorganization that takes place during mosquito metamorphosis. *Ae*Dronc transcripts were also found in all adult tissues, with the highest levels detected in the fat body, a tissue with high cell turnover rates and the primary immune response tissue in insects.[44]

Recently, it was shown that *Ae*Dredd interacts with a mosquito FADD caspase adaptor, named *Aedes* FADD, that is required for antibacterial immunity in *Ae. aegypti*.[45] Moreover, caspase-like activity was detected in the invaded midgut cells of *Ae. aegypti* when infected with *Plasmodium gallinaceum*.[46] Currently, little is known about the biochemical pathways involved in controlling apoptosis in the malaria-infected mosquito midgut, but anacaspase-7 is one of the three apoptosis-related molecules that has been identified in cDNA libraries enriched in mosquito midgut sequences expressed immediately after midgut invasion by a malaria parasite.[47] Anacaspase-7 shares 40% identity and 60% homology with the *Drosophila* caspase DECAY and immunolocalization experiments revealed a putative protease present only in gut cells after invasion by large numbers of ookinetes. Increased apoptosis has also been detected in malaria-infected mosquito ovaries. Activation of a caspase-like molecule in *Anopheles gambiae* ovaries upon *Plasmodium yoelli nigeriensis* infection has been detected with FAM-VAD.fmk in approximately 25% of the ovarian follicles, but its molecular identity remains unknown.[48]

CYSTEINE PROTEASES IN OTHER BLOODFEEDING ARTHROPODS

Tsetse Fly Cathepsin B

A cathepsin B enzyme (GmCatB) is involved in protein digestion in the midgut of *Glossina morsitans morsitans* (tsetse fly), the vector of African trypanosomes.[49] The cDNA for GmCatB was shown as bloodmeal-induced in a subtractive suppression hybridization experiment and it encodes a protein of 340 amino acids with a predicted molecular mass of 38.2 kDa.[49] The first 19 amino acids of the protein are hydrophobic, suggesting a clear secretory signal peptide sequence at its N-terminus. Further structural analysis indicates that the following 65 amino acid residues (from position 20 to 84) correspond to the propeptide of the enzyme, demonstrating that the protease is expressed as an inactive zymogen.[49] This prodomain is likely proteolytically removed for the protease to be activated, resulting in a mature protein of 28.6 kDa. GmCatB transcription is constitutive and further induced throughout the digestion cycle within a few hours following ingestion of the first bloodmeal. It is also parasite-responsive, as its expression is two- to three-fold higher in trypanosome-infected flies, not only in their midgut but in their fat body as well.[49] Analysis of GmCatB transcripts showed high levels of expression in larvae and pupae, while in adult flies, although it was found to be preferentially expressed in the midgut, transcripts corresponding to GmCatB could be detected in the proventriculus and fat body tissues. No transcripts were detected in tsetse fly salivary glands.[49]

Cathepsin-Like Proteases of Triatomids

Using specific substrates and inhibitors, cathepsin B- and cathepsin L-like activities were also identified in the gut extracts of the bloodsucking bug *Triatoma infestans*, the vector of *Trypanosoma cruzi* (the etiologic agent of Chagas' disease).[50] More specifically, small intestine extracts from unfed bugs hydrolysed ZPhe-Arg-pNA; this activity was optimal at pH 5 and low but variable outside this pH range. It decreased during the first 2 days after feeding, but then increased to a maximum value at 5 and 10 days post feeding.[50] The activity was inhibited by both E-64 and CA-074, which are known chemical inhibitors of cysteine proteases, but inhibition was always less when using the specific cathepsin B inhibitor CA-074.[50] A cDNA encoding a cathepsin B-like proteinase (CatB1) was cloned, revealing an open reading frame of 996 nucleotides that encodes for a deduced protein of 332 amino acids with a predicted molecular weight of 36.3 kDa. The CatB1 protein contains a 16-residue secretory signal peptide, while a propeptide cleavage site was predicted between residues Thr69 and Leu70. The calculated mass of the mature protein is approximately 27 kDa, while the occluding loop, characteristic for cathepsin B-like enzymes, was found between the two Cys residues at positions 118 and 138. The deduced proenzyme sequence showed 48–60% identity to cathepsin-B enzymes from different arthropods and 60% identity to that from humans, while the lowest identity was observed with the fat body-cathepsin B of the dipteran *Ae. aegypti*.[50] The CatB1 gene was expressed at low, constitutive levels in unfed and fed *T. infestans* fifth instar larvae.[50] In addition, the cDNA encoding a cathepsin L-like proteinase from *T. infestans* (CatL1) was also cloned;[50] the open reading frame of 984 nucleotides encodes for a theoretical protein precursor with a molecular weight of 36.5 kDa. The CatL1 protein also possesses a 16-residue N-terminal signal peptide and its

propetide contains 95 amino acids. As a result, the predicted mature enzyme consists of 217 amino acid residues and its estimated molecular weight is 23 kDa. The CatL1 precursor has 55–60% identity to cathepsin-L like enzymes from different insects and 68% identity to that of *Rhodnius prolixus*.[50] The three amino acid residues of the catalytic domain, CHN and the GCNGG motif were conserved in both cathepsins from *T. infestans*, but ERYNIN and KNFD motifs occurred only in the CatL1 sequence, defining it as a cathepsin L-like cysteine protease.[50]

Cathepsin B activity was present also in the midgut of *R. prolixus*, the main triatomine vector of Chagas' disease in Central America and the Andean region, not only before its feeding but also for 35 days after the blood meal. There was a 10-fold increase in this activity from unfed to 6 days post feeding of bugs and cathepsin B activity was higher in mated females than males.[51] Cathepsin B was localized to small, Golgi-derived vesicles in the intestinal cells of the small intestine and to lysosomes in cells of all midgut regions.[52] Apart from cathepsin B, cathepsin D activity has also been reported in the midgut of this insect based on the enzymatic activities of crude midgut extracts or their inhibition by specific inhibitors,[53,54] however, amino acid or nucleotide sequences of any of these proteases have yet to be identified. The cloning of a 1.2 kb cDNA encoding for a cysteine protease from *R. prolixus* (RpCat) that shows a high similarity to cathepsin L-like enzymes has also been reported.[55] The RpCat gene is expressed only in the midgut of adults, but not in the salivary glands or other tissues and in the 1st to 4th nymph instars, but not in 5th instar nymphal stages, indicating that its expression may be regulated by hormonal conditions that are modified before transformation into an adult.[55] The open reading frame of the cDNA encodes a zymogen of 316 amino acids that includes an N-terminal 99-residue propeptide. The enzyme has a C-terminus that could direct it to a secretory route, suggesting its participation in blood digestion.[55] The estimated molecular weights of the zymogen and the mature protease are 33.7 kDa and 23.8 kDa, respectively. The motif, ERFININ, that is characteristic for cathepsin-L like cysteine proteases was also identified in the prosegment region between amino acids 29 and 48.[55] Moreover, it has been demonstrated that insect cathepsin-like enzymes are involved in the proteolytic activation of canatoxin that results in production of entomotoxic peptide(s) involved in the deleterious effect of the protein in *R. prolixus*.[56] Canatoxin is a toxic protein isolated from the jackbean *Canavalia ensiformis*. The presence of canatoxin-like proteins in other leguminous seeds, including many edible ones, are suggestive of a physiological role, perhaps related to plant defense. Canatoxin toxicity has also been assessed in various insects, including *Manduca sexta*, *Schistocerca americana*, *Drosophila melanogaster*, *Ae. aegypti, R. prolixus* and *Callosobruchus maculatus*. Canatoxin was given in their diets and found to be lethal when fed to insects relying on cathepsins as their main digestive enzymes, such as *C. maculatus* and *R. prolixus*.[57] In contrast, insects with trypsin-based digestion (the other four tested) were not affected.[57]

Copepod Cysteine Proteases

Cysteine proteases are also present in the bloodfeeding parasitic copepod *Phrixocephalus cincinnatus*.[58] Substrate specificity, pH profile and inhibitor sensitivity indicate that the proteolytic activity found in these copepods could be partially attributed to cysteine proteases that they are mostly similar to mammalian cathepsins B, L and H.[58]

CONCLUSION

Cysteine peptidases play an important role in extracellular and intracellular protein degradation and processing in a wide range of organisms from bacteria to mammals. They have been studied in various species including bloodfeeding arthropods (Table 1). In ticks they have been shown to be the key digestive enzymes in the gut epithelium

Table 1. Cysteine proteases that have been identified and functionally characterized from bloodfeeding arthropods

Name	Species	Specificity	MW (kDa)	Tissue	Ref.
Ticks					
BmCL1	*Boophilus (Rhipicephalus) microplus*	Cathepsin L	42	Gut	12, 13
HlCPL-A	*Haemaphysalis longicornis*	Cathepsin L	29	Gut	16
IrCL1	*Ixodes ricinus*	Cathepsin L	30	Gut predominantly, also salivary glands, ovary, mal. glands	26, 32
IrCB1	*Ixodes ricinus*	Cathepsin B	32	Gut	26, 32
Longipain	*Haemaphysalis longicornis*	Cathepsin B	39	Gut	17
IrCC	*Ixodes ricinus*	Cathepsin C	23-25	Gut predominantly, also salivary glands, ovary, mal. glands	26, 32
IrAE	*Ixodes ricinus*	Legumain/AE	38-40	Gut	18, 32
HlLgm1	*Haemaphysalis longicornis*	Legumain/AE	38	Gut	21
HlLgm2	*Haemaphysalis longicornis*	Legumain/AE	36	Gut	22
Other arthropods					
Vitellogenic cathepsin B	*Aedes aegypti*	Cathepsin B	33	Adult ovaries, fat body	39
Aedes Dredd	*Aedes aegypti*	Initiator caspase	55 (Estimated, containing the death domains)	Abundant, higher expression in early pupae and adult fat body	43
Aedes Dronk	*Aedes aegypti*	Initiator caspase	52 (Estimated, containing the caspase recruitment domain)	Abundant, higher expression in late instar larvae, early and late pupae and adult fat body	44

continued on next page

Table 1. Continued

Name	Species	Specificity	MW (kDa)	Tissue	Ref.
GmCatB	*Glossina morsitans morsitans*	Cathepsin B	28.6 (Estimated)	Abundant, high levels in larvae and pupae, expressed in all adult tissues apart from the salivary glands	49
CatB1	*Triatoma infestans*	Cathepsin B	27 (Estimated)	Low, constitutive levels in unfed and fed fifth instar larvae	50
CatL1	*Triatoma infestans*	Cathepsin L	23 (Estimated)	Not known	50
RpCat	*Rhodnius prolixus*	Cathepsin L	23.8 (Estimated)	Only in adult midguts and in 1st to 4th nymph instars	55

The name, the species in which it was characterized, specificity, molecular weight of the mature enzyme (kDa) and tissue specificity are provided for each cysteine protease, as well as the corresponding reference in the chapter. 'Estimated' means that the gene sequence was used for molecular weight calculation, but there are no experimental data to support the calculation.

and to have specific roles in embryogenesis (i.e., degradation of yolk proteins). The evolutionarily older gut-associated cysteine and aspartic peptidase matrix, best described from bloodfeeding helminths, was conserved up to the insects and in Coleoptera, Diptera and Hemiptera, cysteine proteases have been considered as targets for pest control. However, these organisms have also evolved serine peptidase-based extracellular machinery not present in ticks. Cysteine proteases, such as cathepsin B and L, have also been shown to have important roles in embryogenesis and tissue remodelling during insect metamorphosis, while they can also participate in nutrient recycling and availability to the developing organism. Moreover, caspases have been proposed as important players in the apoptotic machinery of bloodfeeding arthropods. Finally, there is accumulating evidence that cysteine proteases may play a role in disease transmission as well, but this hypothesis remains to be tested as to whether it applies to pathogen transmission by different bloodfeeding arthropods.

ACKNOWLEDGEMENTS

This work was supported by grant No. Z60220518, grant No. IAA600960910, grant No. IAA600960811, and grant No. KJB600960911 from the Grant Agency of the Academy of Sciences of the Czech Republic and the Research Center No. LC06009 from the Ministry of Education, Youth and Sports of the Czech Republic. I.M.B.F., E.C. and M.K. were supported by the Intramural Research Program of the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health. We thank NIAID intramural editor Brenda Rae Marshall for assistance and Dr Mark W. Robinson and Professor John P. Dalton for critical comments on the chapter.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest regarding this manuscript.

REFERENCES

1. Nava S, Guglielmone AA, Mangold AJ. An overview of systematics and evolution of ticks. Front Biosci 2009; 14:2857-2877.
2. de la Fuente J, Estrada-Pena A, Venzal JM et al. Overview: Ticks as vectors of pathogens that cause disease in humans and animals. Front Biosci 2008; 13:6938-6946.
3. Grandjean O, Aeschlimann A. Contribution to the study of digestion in ticks: histology and fine structure of the midgut ephithelium of Ornithodorus moubata, Murray (Ixodoidea, Argasidae). Acta Trop 1973; 30:193-212.
4. Obenchain FD, Galun R. Physiology of Ticks, 1st ed. Oxford: Pergamon Press, 1982.
5. Harrison FW, Foelix RF. Chelicerate arthropoda. New York ; Chichester: Wiley; 1999.
6. Tarnowski BI, Coons LB. Ultrastructure of the midgut and blood meal digestion in the adult tick Dermacentor variabilis. Exp Appl Acarol 1989; 6:263-289.
7. Sauer JR, Hair JA. Morphology, Physiology and Behavioral Biology of Ticks. New York: Halsted Press, 1986.
8. Kaufman WR. Tick-host interaction: a synthesis of current concepts. Parasitol Today 1989; 5:47-56.
9. Sonenshine DE. Biology of Ticks. Oxford: Oxford University Press, 1991.
10. Reich CI, Zorzopulos J. Boophilus microplus: characterization of larval proteases. Exp Parasitol 1978; 44:1-6.
11. Zorzopulos J, Reich CI, Galassi N. Boophilus microplus: characterization of larval phosphomonoesterases and isolation of subcellular fractions with high phosphatase activity. Exp Parasitol 1978; 45:128-138.
12. Renard G, Garcia JF, Cardoso FC et al. Cloning and functional expression of a Boophilus microplus cathepsin L-like enzyme. Insect Biochem Mol Biol 2000; 30:1017-1026.
13. Renard G, Lara FA, de Cardoso FC et al. Expression and immunolocalization of a Boophilus microplus cathepsin L-like enzyme. Insect Mol Biol 2002; 11:325-328.
14. Mulenga A, Sugimoto C, Onuma M. Characterization of proteolytic enzymes expressed in the midgut of Haemaphysalis longicornis. Jpn J Vet Res 1999; 46:179-184.
15. Mulenga A, Sugimoto C, Ingram G et al. Molecular cloning of two Haemaphysalis longicornis cathepsin L-like cysteine proteinase genes. J Vet Med Sci 1999; 61:497-502.
16. Yamaji K, Tsuji N, Miyoshi T et al. Hemoglobinase activity of a cysteine protease from the ixodid tick Haemaphysalis longicornis. Parasitol Int 2009; 58:232-237.
17. Tsuji N, Miyoshi T, Battsetseg B et al. A cysteine protease is critical for Babesia spp. transmission in Haemaphysalis ticks. PLoS Pathog 2008; 4:e1000062.
18. Sojka D, Hajdusek O, Dvorak J et al. IrAE: an asparaginyl endopeptidase (legumain) in the gut of the hard tick Ixodes ricinus. Int J Parasitol 2007; 37:713-724.
19. Gotz MG, James KE, Hansell E et al. Aza-peptidyl Michael acceptors. A new class of potent and selective inhibitors of asparaginyl endopeptidases (legumains) from evolutionarily diverse pathogens. J Med Chem 2008; 51:2816-2832.
20. Ovat A, Muindi F, Fagan C et al. Aza-peptidyl Michael acceptor and epoxide inhibitors—potent and selective inhibitors of Schistosoma mansoni and Ixodes ricinus legumains (asparaginyl endopeptidases). J Med Chem 2009; 52:7192-7210.
21. Abdul Alim M, Tsuji N, Miyoshi T et al. Characterization of asparaginyl endopeptidase, legumain induced by blood feeding in the ixodid tick Haemaphysalis longicornis. Insect Biochem Mol Bio 2007; 37:911-922.
22. Alim MA, Tsuji N, Miyoshi T et al. HlLgm2, a member of asparaginyl endopeptidases/legumains in the midgut of the ixodid tick Haemaphysalis longicornis, is involved in blood-meal digestion. J Insect Physiol 2008; 54:573-585.

23. Alim MA, Tsuji N, Miyoshi T et al. Developmental stage- and organ-specific expression profiles of asparaginyl endopeptidases/legumains in the ixodid tick Haemaphysalis longicornis. J Vet Med Sci 2008; 70:1363-1366.
24. Alim MA, Tsuji N, Miyoshi T et al. Legumains from the hard tick Haemaphysalis longicornis play modulatory roles in blood feeding and gut cellular remodelling and impact on embryogenesis. Int J Parasitol 2009; 39:97-107.
25. Boldbaatar D, Sikalizyo Sikasunge C, Battsetseg B et al. Molecular cloning and functional characterization of an aspartic protease from the hard tick Haemaphysalis longicornis. Insect Biochem Mol Biol 2006; 36:25-36.
26. Sojka D, Franta Z, Horn M et al. Profiling of proteolytic enzymes in the gut of the tick Ixodes ricinus reveals an evolutionarily conserved network of aspartic and cysteine peptidases. Parasit Vectors 2008; 1:7.
27. Hatta T, Kazama K, Miyoshi T et al. Identification and characterisation of a leucine aminopeptidase from the hard tick Haemaphysalis longicornis. Int J Parasitol 2006; 36:1123-1132.
28. Caffrey CR, McKerrow JH, Salter JP et al. Blood 'n' guts: an update on schistosome digestive peptidases. Trends Parasitol 2004; 20:241-248.
29. Delcroix M, Sajid M, Caffrey CR et al. A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. J Biol Chem 2006; 281:39316-39329.
30. Williamson AL, Brindley PJ, Knox DP et al. Digestive proteases of blood-feeding nematodes. Trends Parasitol 2003; 19:417-423.
31. Anderson JM, Sonenshine DE, Valenzuela JG. Exploring the mialome of ticks: an annotated catalogue of midgut transcripts from the hard tick, Dermacentor variabilis (Acari: Ixodidae). BMC Genomics 2008; 9:552.
32. Horn M, Nussbaumerova M, Sanda M et al. Hemoglobin digestion in blood-feeding ticks: mapping a multipeptidase pathway by functional proteomics. Chem Biol 2009; 16:1053-1063.
33. Grunclova L, Horn M, Vancova M et al. Two secreted cystatins of the soft tick Ornithodoros moubata: differential expression pattern and inhibitory specificity. Biol Chem 2006; 387:1635-1644.
34. Fagotto F. Yolk degradation in tick eggs: I. Occurrence of a cathepsin L-like acid proteinase in yolk spheres. Arch Insect Biochem Physiol 1990; 14:217-235.
35. Fagotto F. Yolk degradation in tick eggs: II. Evidence that cathepsin L-like proteinase is stored as a latent, acid-activable proenzyme. Arch Insect Biochem Physiol 1990; 14:237-252.
36. Seixas A, Dos Santos PC, Velloso FF et al. A Boophilus microplus vitellin-degrading cysteine endopeptidase. Parasitology 2003; 126(Pt 2):155-163.
37. Seixas A, Leal AT, Nascimento-Silva MC et al. Vaccine potential of a tick vitellin-degrading enzyme (VTDCE). Vet Immunol Immunopathol 2008; 124:332-340.
38. Estrela A, Seixas A, Termignoni C. A cysteine endopeptidase from tick Rhipicephalus (Boophilus) microplus) larvae with vitellin digestion activity. Comp Biochem Physiol B Biochem Mol Biol 2007; 148:410-416.
39. Cho WL, Tsao SM, Hays AR et al. Mosquito cathepsin B-like protease involved in embryonic degradation of vitellin is produced as a latent extraovarian precursor. J Biol Chem 1999; 274:13311-13321.
40. Nirmala X, Marinotti O, James AA. The accumulation of specific mRNAs following multiple blood meals in Anopheles gambiae. Insect Mol Biol 2005; 14:95-103.
41. Uchida K, Ohmori D, Ueno T et al. Preoviposition activation of cathepsin-like proteinases in degenerating ovarian follicles of the mosquito Culex pipiens pallens. Dev Biol 2001; 237:68-78.
42. Cooper DM, Granville DJ, Lowenberger C. The insect caspases. Apoptosis 2009; 14:247-256.
43. Cooper DM, Pio F, Thi EP et al. Characterization of Aedes Dredd: a novel initiator caspase from the yellow fever mosquito, Aedes aegypti. Insect Biochem Mol Biol 2007; 37:559-569.
44. Cooper DM, Thi EP, Chamberlain CM et al. Aedes Dronc: a novel ecdysone-inducible caspase in the yellow fever mosquito, Aedes aegypti. Insect Mol Biol 2007; 16:563-572.
45. Cooper DM, Chamberlain CM, Lowenberger C. Aedes FADD: a novel death domain-containing protein required for antibacterial immunity in the yellow fever mosquito, Aedes aegypti. Insect Biochem Mol Biol 2009; 39:47-54.
46. Zieler H, Dvorak JA. Invasion in vitro of mosquito midgut cells by the malaria parasite proceeds by a conserved mechanism and results in death of the invaded midgut cells. Proc Natl Acad Sci USA 2000; 97:11516-11521.
47. Abraham EG, Islam S, Srinivasan P et al. Analysis of the Plasmodium and Anopheles transcriptional repertoire during ookinete development and midgut invasion. J Biol Chem 2004; 279:5573-5580.
48. Ahmed AM, Hurd H. Immune stimulation and malaria infection impose reproductive costs in Anopheles gambiae via follicular apoptosis. Microbes Infect 2006; 8:308-315.
49. Yan J, Cheng Q, Li CB et al. Molecular characterization of three gut genes from Glossina morsitans morsitans: cathepsin B, zinc-metalloprotease and zinc-carboxypeptidase. Insect Mol Biol 2002; 11:57-65.
50. Kollien AH, Waniek PJ, Nisbet AJ et al. Activity and sequence characterization of two cysteine proteases in the digestive tract of the reduviid bug Triatoma infestans. Insect Mol Biol 2004; 13:569-579.

51. Houseman JG, Downe AER. Activity cycles and the control of four digestive proteinases in the posterior midgut of Rhodnius prolixus Stål (Hemiptera: Reduviidae). J Insect Physiol 1983; 29:141-148.
52. Billingsley PF, Downe AER. Ultrastructural localisation of cathepsin B in the midgut of Rhodnius prolixus Stål (Hemiptera: Reduviidae) during blood digestion. Int J Insect Morph Embryol 1988; 17:295-302.
53. Terra WR, Ferreira C, Garcia ES. Origin, distribution, properties and functions of the major Rhodnius prolixus midgut hydrolases. Insect Biochem 1988; 18:423-434.
54. Houseman JG, Downe AER. Characterization of an acidic proteinase from the posterior midgut of Rhodnius prolixus Stål (Hemiptera: Reduviidae). Insect Biochem 1982; 12:651-655.
55. Lopez-Ordonez T, Rodriguez MH, Hernandez-Hernandez FD. Characterization of a cDNA encoding a cathepsin L-like protein of Rhodnius prolixus. Insect Mol Biol 2001; 10:505-511.
56. Ferreira-DaSilva CT, Gombarovits ME, Masuda H et al. Proteolytic activation of canatoxin, a plant toxic protein, by insect cathepsin-like enzymes. Arch Insect Biochem Physiol 2000; 44:162-171.
57. Carlini CR, Oliveira AE, Azambuja P et al. Biological effects of canatoxin in different insect models: evidence for a proteolytic activation of the toxin by insect cathepsinlike enzymes. J Econ Entomol 1997; 90:340-348.
58. Perkins PS, Haley D, Rosenblatt R. Proteolytic enzymes in the blood-feeding parasitic copepod, Phrixocephalus cincinnatus. J Parasitol 1997; 83:6-12.