Expression and Characterization of Digestive Enzymes from Tsetse (*Glossina morsitans morsitans*)

2013

Ruttayaporn Ngasaman

Doctoral Program in Animal and Food Hygiene

Graduate School of Obihiro University of Agriculture and Veterinary Medicine

ツェツェバエ由来消化酵素の発現と性状解析

(*Glossina morsitans morsitans*)

平成 25 年
（2013）

Ruttayaporn Ngasaman

帯広畜産大学大学院畜産学研究科

博士後期課程 畜産衛生学専攻

Table of Contents

List of Figures

List of Tables

Abbreviations

AAT	animal African trypanosomosis
BApNA	Nα-Benzoyl-L-arginine 4-nitroanilide hydrochloride
BMGY	buffered glycerol-complex medium
BMMY	buffered methanol-complex medium
bp	base pair
BSA	bovine serum albumin
cDNA	complementary deoxyribonucleic acid
CO_2	carbondioxide
DMSO	dimethyl sulfoxide
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
GmcathB	*G. m. morstians* cathepsin B
GmcathL	*G. m. morstians* cathepsin L
Gmtrypsin	*G. m. morstians* trpsin-lke serine protease
GST	glutathione-S-transferase
HAT	human African trypanosomosis
Hb	bovine hemoglobin
IPTG	isopropyl β-D-1-thiogalactopyranoside
Km	Michaelis constants
LB	Luria broth

min	minute
ml	milliliter
mM	millimolar
OD	optical density
ORF	open reading frame
PAD	Pichia Adenine Dropout
P. pastoris	*Pichia pastoris*
RNA	ribonucleic acid
RT-PCR	reverse transcriptase polymerase chain reaction
ScathL	*Sarcophaga peregrina* cathepsin L
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
sec	second
T. brasilliensis	*Triatoma brasilliensis*
T. cruzi	*Trypanosoma cruzi*
μg	microgram
μl	microliter
μM	micromolar
WHO	world health organization
Z-Arg-Arg-MCA	benzyloxycarbonyl-L-arginyl-L-argine-4-methylcoumaryl-7-amide
Z-Phe-Arg-MCA	benzyloxycarbonyl-L-phenylalanyl-L-argine-4-methycoumaryl-7-amide

General Introduction

1. Overview of Trypanosomosis

Trypanosomes are unicellular parasitic protozoa belonging to Genus *Trypanosoma*. Different species of trypanosome infect a variety of vertebrates, including domestic animal and human. A number of *Trypanosoma* species and subspecies are economically significant causing human and animal diseases which are an obstacle to human welfare, affecting cattle rearing and agricultural development in Africa, Asia and South America. Presently, WHO has listed 36 countries as being endemic for Human African trypanosomosis (HAT), or African sleeping sickness which is transmitted only in rural sub-Saharan countries. No vaccine is available, disease treatment is difficult and has life-threatening side effects. *Trypanosoma brucei gambiense* accounts for more than 98% of reported cases of sleeping sickness. More than 95% of the cases of *T. b. gambiense* infection is reported from central Africa and in limited areas of west Africa such as the Democratic Republic of the Congo, Angola, Sudan, Central African Republic, Chad, and northern Uganda. By the end of 20th century, some 30,000 cases were reported annually. However, by the eradication program of WHO cooperated with nongovermental organizations and private sectors, in 2012 the number accounted less than 7000 of new reported cases (http://www.who.int/mediacentre/factsheets). Animal African trypanosomosis (AAT) causes mainly by *T. congolense*, *T. b. brucei* and *T. vivax*, occurs in 37 countries resulting in a risk to approximately 60 million cattle. Most cases of AAT are chronic such as loss of condition and emaciation and anemia leading to serious economic losses in livestock. Acute and many untreated cases which may be fatal, can also occur. AAT is found mainly in those

regions of Africa where its biological vector, the tsetse fly, exists. *T. vivax*, has become established in South America and Asia where it is transmitted by biting flies acting as mechanical vectors such as tabanids. Trypanosomes can be infected all domesticated animals, clinical cases have been described in cattle, water buffalo, sheep, goats, camels, horses, donkeys, alpacas, llamas, pigs, dogs, cats and other species. In parts of Africa, cattle are the main species affected, due to the feeding preferences of tsetse flies. More than 30 species in the wild, including ruminants such as white-tailed deer, duikers, antelope and African buffalo, as well as wild equidae, lions, leopards, warthogs, capybaras, elephants, nonhuman primates and various rodents are also known to be susceptible to trypanosome infection. The host preferences of each trypanosome species may differ, but *T. congolense*, *T. vivax* and *T. b. brucei* have a wide host range among domesticated animals. *T. godfreyi* and *T. suis* occur in pigs. *T. simiae* appears to be most important in pigs, but it has also been reported by PCR in camels, horses and cattle. Most trypanosomes must develop for one to a few weeks in tsetse flies, which act as biological vectors. When an infected tsetse fly bites an animal, the parasites are transmitted in the saliva. Trypanosomes can also be spread by mechanically including surgical instruments, needles, syringes and various biting flies including horse flies (tabanidae). After transmission, trypanosomes cause a local infection called a chancre, and then proceed into the bloodstream, from the bloodstream they can infect the brain and spinal fluid where they remain, and cause the degenerative disease. The life cycle of trypanosomes is mostly similar for *T. brucei* (including *T. b. gambiense* and *T. b. rhodesiense*), and *T. congolense.* The exception is that *T. conglense* after has developed in the midgut it goes forward to the proboscis not the salivary glands as *T. brucei* (Figure 1).

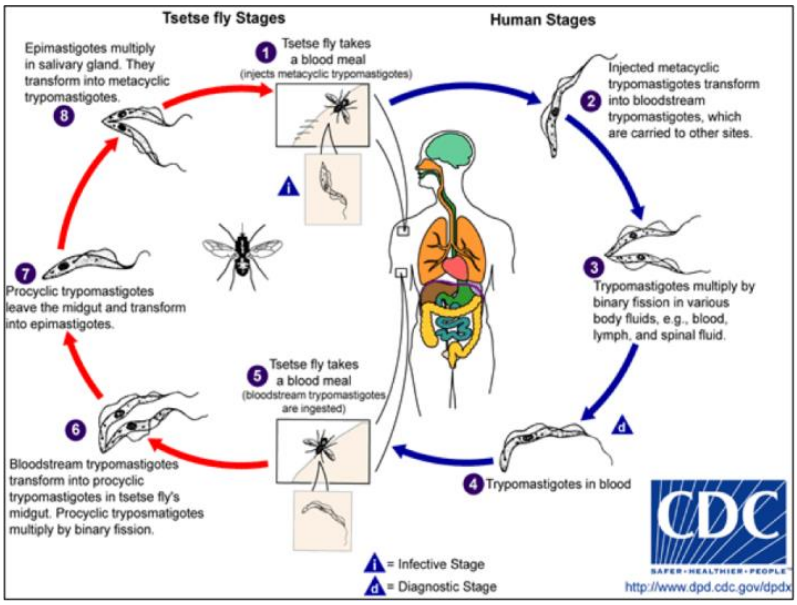


Figure 1. Life cycle of trypanosome

(http://en.wikipedia.org/wiki/African_trypanosomiasis)

2. Biology of Tsetse (*Glossina spp.*)

Tsetse flies are strictly hematophagous and confined almost exclusively to sub-Saharan Africa. They inhabit in rural areas, living in the woodlands and thickets of the savannah and the dense vegetation along streams, they bite during daylight hours. Tsetse flies are classified into one genus, *Glossina* of the family Glossinidae, order *Diptera.* The tsetse genus is generally split into three groups of species based on a combination of distributional, behavioral, molecular and morphological characteristics. The genus includes savannah flies (Subgenus *Morsitans*), forest flies (Subgenus *Fusca*) and riverrine flies (Subgenus *Palpalis*). There are more than 30 species and subspecies under three genus have been identified at present. Although all species of tsetse are believed to be capable of transmitting trypanosomes, only a few species are important vectors (Table 1).

Table 1 Important vector species of *Glossina spp.*

Subgenus (species group)	Species name	Trypanosomosis of	
		Human	Livestock
fusca	*G. nashi*		•
	G. tabaniformis		•
	G. medicorum		•
	G. brevipalpis		•
	G. longipennis		•
palpalis	*G. tavhinoides*	•	•
	G. fuscipes fascipes	•	•
	G. palpalis palpalis	•	•
	G. p. gambiensis		•
morsitans	*G. austeni austeni*		•
	G. longipalpis		•
	G. pallidipes	•	•
	G. swannertoni		•
	G. morsitans morsitans	•	•
	G. m. centralis	•	•
	G. m. submorsitans		•

Glossina morsitans is 7.5 to 14 mm long and is brownish gray in color. Its wings cross like scissors and it has recognizable branched Arista on the antennae. Adults have highly specialized mouthparts used to bite and lap up blood. *G. morsitans* is solely a daytime feeder widely distributes in the region of tropical Africa having a mean annual temperature of varying from 19-28 ˚C. Both sexes respond to CO_2 produced by the exhalation of animals and feed exclusively on blood and produce anticoagulants to prevent clotting. The economic and biological importance of the *G. morsitans* in the fact that its species are the primary vector of *T. brucei* and *T. congolense*, the causative agents of African sleeping sickness in human and nagana disease in animal. The transmission occurs when a tsetse fly bites an infected animal and contracts the protozoan while feeding off of its blood. The parasite then rapidly moves to the mid intestine of the fly and begins to reproduce. From here they move to the fly's salivary gland and remain there to mature. Once in the epimastigote form, they divide and are ready to be transmitted to a mammalian host via saliva injected into the bite.

Female tsetse mates when only a few days old. Egg matures one at a time, ovulation occurs when females are about 1 week old, and the fertilized egg embryonates in the "uterus". The egg hatches within 5 days with the assistance of specialized structures in the female reproductive tract. All three larval instars are retained within the uterus and are nourished by a protein-and lipid-rich secretion ("milk") produced by the female's accessory glands ("milk gland"). Intrauterine larval development takes 5-6 days. At the end of the 3rd larval instar the female undergoes parturition (gives birth to the larva). The larval fly will burrow into the ground and pupates within 30 minutes of deposition. The pupa will

develop for 3-4 weeks after that the adult fly will emerge, then the next (nine-day long) larviposition cycle begins. The life cycle of tsetse is shown in figure 2.

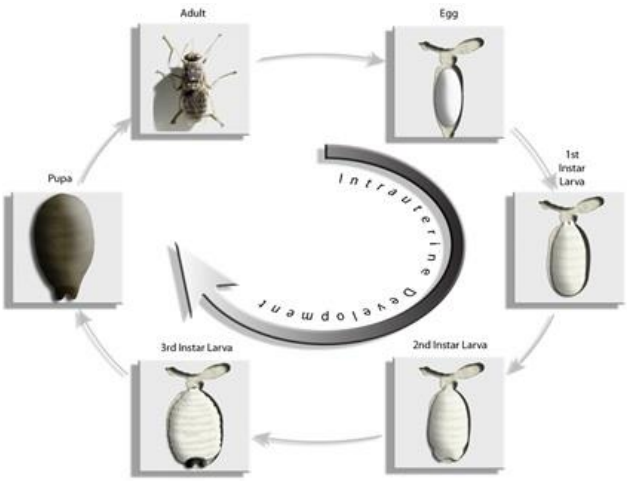


Figure 2. Life cycle of tsetse (*Glossina spp.*)

(http://publichealth.yale.edu/news/slideshows/hat.aspx)

3. Enzyme in General

Enzymes are biocatalyst proteins produced by living tissue that increase the rate of reaction that occur in the tissue. Although enzymes increase reaction rates, they do not change the relative concentrations of reactants and products when equilibrium is stained. The Nomenclature Committee of the International Union of Biochemistry and Molecular Biology has been classified and numbering a systematic name that defines six major classes of enzyme function, each with several subclasses, as described below.

1. Oxidoreductases (EC 1) catalyze a variety of oxidation-reduction reactions and frequently employ coenzymes such as nicotinamide adenine dinucleotide (NAD^+), its phosphate derivative ($NADP^+$), flavin adenine dinucleotide (FAD), or lipoate.

2. Transferases (EC 2) catalyze the transfer of groups such as amino, carboxyl, carbonyl, methyl, acyl, glycosyl or phosphoryl.
3. Hydrolases (EC 3) catalyze the hydrolytic cleavage of bonds between a carbon atom and some other atoms including phosphoric anhydride bonds by addition of water. A number of hydrolases acting on ester, glycosyl, peptide, amide or other bonds are known to catalyse not only hydrolytic removal of a particular group from their substrates, but likewise the transfer of this group to suitable acceptor molecules.
4. Lyases (EC 4) catalyze cleavage of carbon-carbon, carbon-sulfur, and certain carbon-nitrogen bonds (excluding peptide). Lyases differ from other enzymes in that they require only one substrate for the reaction in one direction, but two substrates for the reverse reaction.
5. Isomerases (EC 5) catalyze the conversion of one isomeric form of a compound into another and involve in intracellular oxidation-reduction reaction.
6. Ligases (EC 6) catalyze the formation of bonds between carbon and oxygen, sulfer, nitrogen, and other atoms.

Digestive enzymes are hydrolases, name proteases (also termed peptidase or proteinase) which are enzymes that conduct proteolysis by hydrolysis of the peptide bonds in the polypeptide chain forming protein. Proteases include endopeptidase (EC 3.4.21-24) which cleaves an internal peptide bond and exopeptidase (EC 3.2.11-19) which cleaves a terminal amino acid residue at the end of a polypeptide. Proteases cleave a peptide bond, calls the scisscile bond, between two amino acid residues named P1 and P1'. Residues on the amino-terminal side of the scissile bond (P1) are numbered in the C to N direction,

whereas residues on the carboxyl-terminal side of scissile bond (P1’) are numbered in the N to C direction (Figure 3).

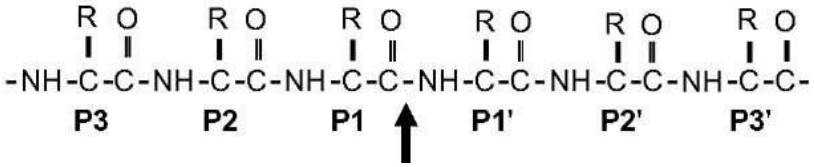


Figure 3 The Schechter and Berger notation for protease cleavage sites. The arrow designates the scissile peptide bond between amino acid residues P1 and P1′ (Schechter and Berger, 1967).

Proteases are currently divided into six subclasses on the basis of their catalytic activities, as shown with specific reagents or the effect of pH.

1. Serine proteases (EC 3.4.21) are enzymes that cleave peptide bonds in proteins, in which serine serves as the nucleophilic amino acid at the active site, that contains histidine, aspartic acid and serine amino acid residues.
2. Cysteine proteases (EC 3.4.22) are also known as thiol proteases have a functional role in protein degradation. They share a common catalytic mechanism that involves a nucleophilic cysteine thiol in the active site.
3. Aspartate proteases (EC 3.4.23) are enzymes that use an aspartate residue for catalysis of their peptide substrates. They have two highly-conserved aspartates in the active site and are optimally active at acidic pH.

4. Metalloproteases (EC 3.4.24) are the most diverse of the catalytic types of proteases. They are characterized by the requirement for a divalent metal activity in their activity.
5. Threonine proteases (EC 3.4.25) are family of proteolytic activity harboring a threonine residue within the active site. It is responsible for functioning proteasome, the large protein-degrading apparatus.
6. Glutamic proteases (MEROPS family G1) are non-eukayote proteases have a unique catalytic dyad consisting of glutamine and glutamate residues.

Proteases can be found in all organisms they are involved in a multitude of physiological reactions from simple digestion of food proteins to highly regulated cascades (e.g., the blood-clotting cascade, the complement system, apoptosis pathways, and the invertebrate prophenoloxidase-activating cascade). Insect proteases have been determined in many species of order Lepidoptera, Hemiptera, and Diptera. They were characterized in many stages and organs of insects such as eggs and embryos, digestive tracts, fat bodies and hemolymph. The pivotal roles are digestion of yolk protein for embryonic development (Raikhel and Dhadialla, 1992), activation of a prophenoloxidase as a protective response to infection and wounding (melanization) (Paskewitz et al., 2006), induction of metamorphosis (Gui et al., 2006; Jones et al., 1993; Lee et al., 2009), and activation of Spaztle-processing enzyme (proSpaztle) (Gorman et al., 2000). Common digestive enzymes of insect are serine proteases (trypsin and chymotrypsin), cysteine proteases (cathepsin), as well as aspartate proteases (cathepsin D). From the molecular study on *Glossina spp.* genome reveals major classes of digestive enzymes are cathepsin B, cathepsin L,

Zinc-metalloprotease, Zinc-carboxypeptidase, chymotrypsin as well as trypsin (Cheeseman and Gooding, 1985; Yan et al., 2001, 2002a). Additionally, there are eleven ORFs harbor the CLIP domain of serine protease, which possess orthologous in the *Drosophila melano-gaster* genome for genes implicated in immunity via the Toll and melanization pathway (Mwangi et al., 2011). Protein in the blood meal has function to stimulate enzyme activity in the midgut. Specifically, trypsin inhibitor found in blood meal serum were believed to stimulate production of trypsin and carboxypeptidase B in the tsetse midgut (Gooding, 1977).

4. Michaelis- Menten Equation

Leonor Michaelis and Maude Menten proposed a simple model that accounts for most of the features of enzyme-catalyzed reactions, as represented below:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P$$

In this model, enzyme (E) reversibly combines with substrate (S) form an enzyme-substrate complex (ES) which subsequently yield the product (P) and regenerating the free enzyme (E). k_1, k_{-1} and k_2 are rate constants. The Michaelis-Menten equation describes how reaction velocity varies with substrate concentration:

$$V_0 = \frac{V_{max}[S]}{K_M + [S]}$$

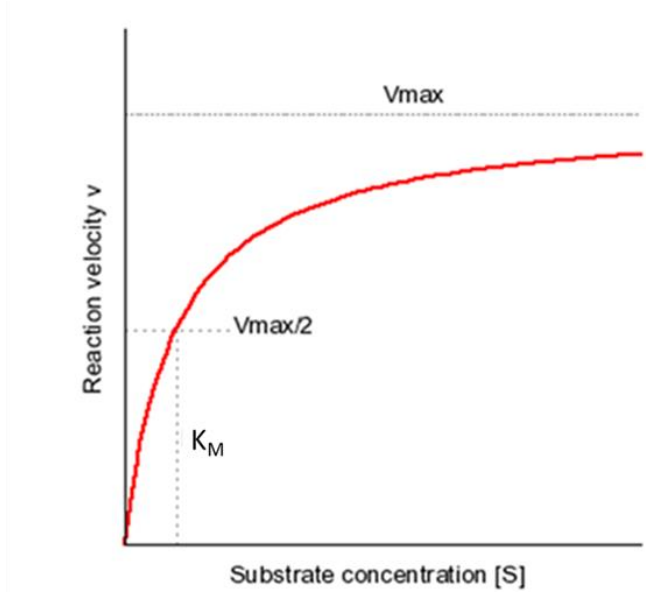


Figure 4. Michaelis-Menten Equation

(http://depts.washington.edu/wmatkins/kinetics/michaelis-menten.html)

Where V_0 = initial reation velocity

V_{max} = maximal velocity

K_M = Michaelis constants = $(k_{-1} + k_2)/ k_1$

[S] Substrate concentration

Important conclusions of Michaelis-Menten kinetics:

1. K_M is a characteristic of an enzyme and its particular substrate, and reflects the affinity of the enzyme for that substrate. K_M is numerically equal to the substrate concentration at which the reaction velocity is equal to ½ V_{max}. K_M does not vary with the concentration of enzymes. Small K_M reflects a high affinity of the enzyme for substrate, because a low concentration of substrate is needed to reach a velocity that is ½ V_{max}. Large K_M reflects a low affinity of enzyme for substrate because a high concentration of substrate is needed for a half-saturate the enzyme.

2. Relationship of velocity to enzyme concentration: The rate of the reaction is directly proportional to the enzyme concentration at all substrate concentrations.
3. Order of reaction: At low concentration of substrate ([S]< K_M), the velocity of the reaction is called first order that is proportional to substrate concentration. When substrate is higher than K_M , the velocity is constant and equal to V_{max} . The rate of reaction is then independent of substrate concentration and is called zero order which respect to substrate concentration.
4. The relevant kinetic parameters of an enzyme under steady-state condition that determine are k_{cat} and K_M. k_{cat} is the catalytic constant for the conversion of substrate to product.
5. Catalytic efficiency (k_{cat}/ K_M) is a property of enzyme refers to as the 'specificity constant', use for comparing the relative rates of an enzyme acting on alternative, competing substrates.

5. Tsetse-trypanosome Interaction

Tsetse flies are important in medical and veterinary entomology because their ability to transmit pathogenic trypanosomes in Africa. However, vector competency of tsetse is in general, lower than that of other vector arthropods, such as mosquitoes and ticks. The most likely reason is they have a robust innate immune system in which most ingested trypanosomes are eliminated in the fly midgut. Indirect evidence suggested tsetse had midgut lectins that were capable of killing trypanosomes (Welburn and Maudlin, 1999). Other information demonstrated that antimicrobial peptides involved in decreasing of trypanosome infection (Hao et al., 2001). Components of prophenoloxidase cascade have

been reported as one factor of trypanosome elimination (Nigam et al., 1997). Digestive enzymes are part of component that capable to stimulate prophenoloxidase cascade (Kim et al., 2002). After uptake the blood meal from infected vertebrate, tsetse flies secrete several digestive enzymes into the midgut lumen to digest the blood protein contents and propose to kill the parasites. Once development in the tsetse midgut, trypanosomes differentiate from bloodstream forms (ingested in the blood meal) into procyclic forms. Differentiation of procyclic trypanosome (PCF) is accompanied by extensive morphological changes, by the loss of the variant surface glycoprotein (VSG) surface coat, and by the acquisition of a new set of coat molecules name procyclins. Procyclins, the products of a small multigene family in *T. brucei*, are glycosyl phosphatidylinositol-anchored proteins containing characteristic amino acid repeats at the C terminus [either EP (EP procyclin, a form of procyclin rich in Glu-Pro repeats) or GPEET (GPEET procyclin, a form of procyclin rich in Gly-Pro-Glu-Glu-Thr repeats)]. Procyclic forms of *T. congolense* have been identified two stage-specific surface molecules. A protease-resistant surface molecule (PRS) is expressed in the early phase of midgut infection in the tsetse (Bütikofer et al., 2002), and species-specific forms of procyclin in the late phase (Utz et al., 2006). It has suggested that procyclin containing both type of amino acid repeat are resistant to tsetse midgut proteases, to protect the parasite surface from digestive enzymes, (Acosta-Serrano et al., 2001) and allowed survival in the midgut (Roditi and Liniger, 2002). Alternative hypotheses are that the procyclin may be involved in protection against tsetse proteases or that they influence transit through the different compartments of the fly and tropism within tsetse (Roditi and Pearson, 1990). These may be the result that trypanosomes try to avoid the tsetse's immune response and the proteolytic activity of digestive enzymes. Therefore, digestive enzymes

are consider as one of the factors that trypanosomes much surpass to success in establishing infection. Beside that there are some additional signals of digestive enzymes can provide the natural trigger on trypanosome differentiation. Speculation, digestive enzyme may retard the establishment trypanosome. More interestingly, trypanosome infection causes a concentration-dependent decrease in trypsin and trypsin-like enzymes in the crude midgut homogenase of *G. m. morsitans*, suggesting trypanosomes might overcome the hostile environment of midgut by inhibit the enzymatic activity (Imbuga et al., 1992). However, interaction between digestive enzymes and trypanosome interactions have not been extensively studied.

6. Aims of the Study

In the tsetse midgut, trypanosomes must interact with the hostile environment of the midgut lumen, in which proteolytic enzymes and digested products might affect their differentiation and survival. Thus, the midgut is not only the site for synthesis and secretion of digestive enzymes and nutrient absorption, but also a primary interface between tsetse and trypanosome. Hence, the physiology of blood meal digestion may substantially influence trypanosome differentiation and survival in the tsetse fly. In order to understand the mechanism of enzymatic activity in blood meal digestion and tsetse-trypanosome interaction. The general objectives of this study are, (1) to elucidate the mechanism of blood meal digestion in the midgut of tsetse, (2) to find out the method to prevent trypanosome infection in tsetse, and (3) to improve the quality of human health and animal health in Sub-Saharan Africa. The specific objective aims to characterize digestive enzymes from the midgut of tsetse and to investigate their roles on tsetse-trypanosome interaction.

Chapter 1: Expression and characterization of cathepsin B from tsetse (*Glossina morsitans morsitans*)

1.1 Introduction

Cathepsin B (EC 3.4.22.1) is one of the cysteine proteases, a member of the papain family which are inhibited by compound mercury. It contains a conserved active site that is formed by cysteine, histidine and asparagine residues. This enzyme is the most abundant form of cathepsin in mammals and invertebrates such as marine shrimp, parasites and insects, including the tsetse fly (Atkinson et al., 2009; Stephans et al., 2012; Wang et al., 2008a; Yan et al., 2002a). Cathepsin poteases are typically encoded by gene specifying preproprotien that is a signal peptide to facilitate secretion, and proregion that is cleaved off to yield active enzyme. Cathepsin B exhibits endopeptidase activity, which cleaves internal peptide bonds. In exceptional cases, it also displays exopeptidase activity which digest protein from the N- or C- terminal amino acids. Occluding loop which presence of two adjacent histidine residues is responsible for exopeptidase activity of cathepsin B.

Mammalian cathepsin B is involved in protein degradation in lysosomes (Cesen et al., 2012), and participated in various physiological processes such as programmed cell death, antigen processing, hormone activation, and bone turnover (Mort and Buttle, 1997; Zavasnik-Bergant and Turk, 2007). Within the immune response of mammalian, cathepsins are released from the lysosome into the cytoplasm appeared to play a significant role in activation of special proteases called cysteinyl-aspartate cleaving proteases (caspases) and apoptosis (Blomgran et al., 2007). It was reported that, mammalian cathepsin B was essential protease in induction of apoptosis in T and B cells, cancer, rheumatoid arthritis,

selectors for the repertoire of surface peptide/MHC II complexes (Conus and Simon, 2008). In parasite, cathepsin B has been suggested to play several roles in parasitism, including encystment, migration, feeding and immune evasion (Smooker et al., 2010). *Schistosoma mansoni* and *S. japonicum* secreted cathepsin B protease to penetrate the skin of host (Dvorak et al., 2008). FhCB2 from *Fasciola hepatica* is releated into host tissue at early infection, is highly antigenic in animal during infection (Law et al., 2003), and it is resistant to the inhibition by host cystetins (Beckham et al., 2009). Therefore, Cathepsin B of liver fluke is promising as a target for control fasciolosis, both as vaccine and drug target.

Insect cathepsin B participated in many physiological processes including programmed cell death of the fat body during metamorphosis (Lee et al., 2009; Shiba et al., 2001; Yang et al., 2007; Yano et al., 1995) and intracellular protein degradation during embryogenesis (Carnevali et al., 2006; Cho et al., 1999; Ribolla et al., 1993; Yang et al., 2006; Zhao et al., 2005). Cysteine proteases, including cathepsin B, have been identified in several insect species such as silkworm (Kageyama and Takahashi, 1990; Xu and Kawasaki, 2001), western corn rootworm (Kaiser-Alexnat, 2009), *Aedes aegypti* (Cho et al., 1999) and the tsetse fly (Yan et al., 2002b). In the tsetse fly, cathepsin B expression is reported to be induced within a few hours of blood meal and to remain high throughout the blood meal digestion cycle. Thus it was speculated that tsetse cathepsin B might play an important role in the tsetse-trypanosome interaction (Yan et al., 2002b). This study aimed to express enzymatically active tsetse cathepsin B, analyze its enzymatic activity using various substrates and investigate its enzymatic activity upon cultured trypanosome.

1.2 Materials and Methods

Total RNA isolation and cDNA synthesis. Tsetse flies (*Glossina morsitans morsitans*) were kindly provided by Prof. Yuichi Chigusa (Dokkyo Medical University), and maintained at 25°C with 50-55% humidity in National Research Center for Protozoan Diseases, Obihiro University of Agriculture and Veterinary Medicine. The male tsetse flies (48 hr post blood meal) were anesthetized on ice and soaked in ethanol for 2 min before isolating midgut. Four midguts were pooled to produce total RNA by using TRIzol® reagent (Invitrogen, CA, USA) following the manufacturer's instructions. The RNA concentration was determined by measuring absorbance at 260 nm using Nanodrop system (Thermo Fisher Scientific Inc., MA, USA) further cleaned by using RNase-free DNase, stored at -80°C or immediately used for preparing single stranded cDNA template. The single stranded cDNA was synthesized using a Ready-To-Go T-Primed First-Strand Kit (Amersham Biosciences, NJ, USA) according to the manufacturer's instructions.

Reverse transcriptase polymerase chain reaction (RT-PCR). The first strand cDNA was amplified using the sense and antisense primers which were designed to amplify coding region of *G. m. morsitans* cathepsin B (GmcathB) proenzyme as shown in table 2. Ten µl PCR reaction mixture containing 1 µl midgut cDNA template, 2.4 µl PCR master mix, 0.5 µl each of sense and antisense primers and 5.6 µl distilled water, was performed using a thermal cycler (VERITI™ Thermal Cycler, Applied Biosystems, CA, USA). The reaction was subjected to the following conditions: first denature at 94°C for 3 min; 35 cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C 1 min, and final extension at 72°C for 7 min.

Cloning and sequencing. PCR product was separated by agarose gel electrophoresis and extracted using QIAquick Gel Extraction Kit (QIAGEN, MD, USA), cloned into pCR®2.1 vector (Invitrogen, CA, USA), transformed into *E. coli* DH5α. Positive clones were selected by colony PCR. Plasmid clones containing GmcathB gene were extracted using the QIAGEN DNA purification kit (QIAGEN, MD, USA) followed by digestion with restriction enzymes to confirm the corrected insert size. Nucleotide sequences of the cloned gene were determined using M13 sequencing primers with BigDye Terminator v1.1 cycle sequencing kit (Applied Biosystems, CA, USA). The amino acid sequences were deduced using the GENETYX-WIN DNA analysis software system (GENETYX Co., Tokyo, Japan). Homology search of nucleotide and amino acid sequences were performed using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Expression of recombinant GmcathB in *Escherichia coli.* Cloned GmcathB gene was excised by using restriction enzymes, *Eco* RI and *Xho* I, and isolated from electrophoresis gel using QIAquick gel purification kit (QIAGEN, MD, USA). The fragment was sub-cloned into pGEX-4T-1 protein expression vector (GE Healthcare, Little Chalfont, UK), transformed to *E. coli* BL21. Correct insertion of GmcathB gene into the vector plasmid was confirmed by colony PCR and restriction enzyme digestion. *E. coli* containing the recombinant plasmid was grown in LB medium with appropriate antibiotic until OD_{600nm} reach between 0.5 and 0.8. Recombinant protein expression was induced by addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Insoluble GST fusion proteins were lysed by 2% sarkosyl-TNE buffer, and purified by using a Glutathione Sepharose 4B (GE Healthcare, Little Chalfont, UK) following the manufacturer's instruction. The concentration of purified protein was checked by Nanodrop at 280 nm

absorbance and confirmed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Expression of recombinant GmcathB in *Pichia pastoris.* PCR amplified GmcathB gene was digested by *Mly* I and *Kpn* I and the *P. pastoris* expression vector, pPinkα-HC (Invitrogen, CA, USA), was digested with *Stu* I and *Kpn* I making the 5' blunt end. Digested vectors and gene fragments were used for constructing pPinkα-HC-GmcathB, transformed into *E. coli* DH5α and culture on LB agar containing ampicillin (50 μg/ml). The constructs were verified by colony PCR, restriction enzyme analysis and sequencing. Recombinant plasmids (5-10 μg) were linearized by *Spe* I and transfected into *P. pastoris* PichiaPink strain by electroporation at 1.5 kV, 25 μF for 5 msec. Yeasts were incubated at 24-30°C on Pichia Adenine Dropout (PAD) selective agar for 3-5 days until distinct white colony are formed. The composition of gene into *P. pastoris* genome was confirmed by PCR using a combination of the 5' α-factor or 3' *CYC* 1 primer and a primer that hybridized within the inserted genes. The transformed *P. pastoris* were inoculated into 10 ml of buffered glycerol-complex medium (BMGY), and allow to grow the cell at 30°C with 250 rpm agitation until the $OD_{600\ nm}$ reach 2-6. Cells were harvested at 1,500 xg for 5 min at room temperature, and suspended with buffered methanol-complex medium (BMMY) at 1/5 of volume of the original culture in order to induce protein expression. The supernatant from 4-day culture containing secreted recombinant protein was dialyzed in *Phosphate buffered* saline (PBS), concentrated by subject to Amicon ultra centrifugal filter tube (Amicon® Ultra, MA, USA). The BCA protein assay was used to protein quantification (BCA Protein Assay Kit, PIERCE Chemical Company, IL, USA).

Polyclonal antibody production. Antibody against rGmcathB was raised by mixing 100 μg of recombinant GST-fused protein in equal volume of TiterMax Gold adjuvant (CytRx Co., CA, USA) and intraperitoneally injected into mice (female BALB/c, 7 weeks old) with two subsequent boosts at 2 weeks interval. Immune sera were obtained from the mice at 10 days after the 3rd immunization after confirmation of high antibody titre against rGmcathB by ELISA.

Characterization of proteolytic activity. The proteolytic activity was determined by using fluorogenic casein (EnzCheck® Protease Assay Kits-Green Fluorescent, Invitrogen, CA, USA). Briefly, 100 μl of rGmcathB (6 μg/ml) was mixed with 100 μl of 1% v/v BODIPY FL casein (Invitrogen, CA, USA) working solution in MES (2-(*N*-morpholino)ethanesulfonic acid) buffer (Sigma-Aldrich, MO, USA). The mixtures were incubated over a temperature range 20-40 ^{o}C at various pH from 2 to 9, and enzyme amount from 0.2 to 1.0 μg for 24 hr. In the inhibition assay, rGmcathB was pre-incubated with broad spectrum cysteine protease inhibitor E-64 at 10 μM for 30 min at 37 ^{o}C before adding BODIPY FL casein. The fluorescence was detected in a fluorescence microplate reader (Fluoroscan Ascent, Thermo Fisher Scientific Inc., MA, USA) at 485 nm excitation and 518 nm emissions.

Hemoglobin and BSA degradation assay. Bovine hemoglobin (Hb) or bovine serum albumin (BSA) were mixed with rGmcathB that was pre-incubated with 10 mM L-cysteine for enzyme activation, and were treated with rGmcathB at the enzyme-substrate ratio 1:20 in 100 mM sodium acetate buffer (pH 4.0) with total reaction volume 600 μl. The reaction with Hb was incubated at 30°C for 0, 4, 18 and 24 hr, while the reaction with BSA

was incubated for 0, 4, 48 and 72 hr. In the inhibition assay, reaction was done in duplicate using 10 μM of E-64. Results were separated by SDS-PAGE using non-reducing SDS loading buffer, and strained with Coomassie Brilliant Blue.

Determination of enzyme kinetic parameters. The hydrolysis efficiency of rGmcathB was analysed using fluorogenic peptidyl substrates which were purchased from Peptide Institute Inc, Osaka, Japan. The substrate was Z-Arg-Arg-MCA (benzyloxycarbon-yl-L-arginyl-L-argine-4-methycoumaryl-7-amide) (Sigma-Aldrich, MO, USA) which is a selective substrate for cathepsin B. Z-Arg-Arg-MCA stock solution was prepared as 10 mM in dimethyl sulfoxide, and was diluted with deionized water to 1 mM before using. In order to activate rGmcathB, 1 μg of rGmcathB was preincubated with assay buffer (100 mM sodium acetate pH 4.0, 4 mM EDTA, 2.5 mM dithiothreitol) for 15 min at 30 ^{o}C. Then the substrates were added in various final concentrations from 0.025 to 0.25 mM. Total 100 μl of assay were performed in 96-well plates in dark condition. The appearance of 7-amino-methylcoumarin was measured with 355 nm excitation and 460 nm emission by a fluorescence microplate reader (Fluoroscan Ascent, Thermo Fisher Scientific Inc., MA, USA). The hydrolysis of Hb was performed at concentration ranges from 25 to 500 μg/ml in sodium acetate buffer (pH 4.0). Hb solution (200 μl) was incubated with 3 μg of enzyme at 30°C for 60 minutes. The reaction was stopped with 250 μl of 10% trichloroacetic acid, centrifuged 15,000 rpm for 15 min, and the absorbance was read at 366 nm. All assays were repeated three times. K_M and K_{cat} values were obtained by fitting the initial rate to the Michaelis-Menten equation using nonlinear regression analysis (GraphPad Prism 5, MDF Co. Tokyo, Japan).

Effect of rGmcathB on trypanosome. *Trypanosoma congolense* IL 3000 strain was used for analyzing affect of rGmcathB treatment *in vitro*. rGmcath B (2.5 μg) was co-cultured with BSF and PCF (number of parasite, 3.4 x 10^6 and 10.5 x 10^6 trypanosomes/ml, respectively) in phosphate saline glucose (PSG) buffer with a total volume of 500 μl, at 27 ^oC. At 4 h intervals, the culture was taken to count the viability of trypanosome and to prepare thin wet smear. The thin wet smear was air-dried, fixed with ethanol and then stained by Giemsa' s stain (Merk, Damstadt) and examined by Nikon Eclipse E200 microscope.

Results

Cloning and expression of cathepsin B. Computational analysis of GmcathB sequence predicted that the sequence encodes 310 of 340 amino acids of pro-cathepsin B (AF329480_1), including N-terminal signal peptide (13 amino acids, corresponding to amino acid position 18-30 in AF329480_1), pro-peptide (73 amino acids, corresponding to amino acid position 31-103 in AF329480_1), and active mature enzyme without C-terminal 13 amino acids (224 amino acids, corresponding to position 104-327 in AF329480_1) (Figure 5). Inactive pro-enzyme has pro-peptide, which acts as an auto-inhibitor, and it needs to activate the enzyme by self-catalysis. The sequence has a conserved catalytic triad of cysteine protease (Cys116, His286 and Asn306 in AF329480_1) and possesses 13 conserved cysteine residues (Figure 5). The analysis also indicated that Asn134 and Asn199 (in AF329480_1) were potential N-linked glycosylation sites. The amino acid sequence showed 78% identity to cathepsin B of flesh fly (*Sarcophaga peregrina*).

GmcathB proenzyme (nucleotide position in full length GmcathB: 142 to 1,071) was cloned and expressed by both bacterial (*E. coli*) and yeast (*P. pastoris*) expression systems (Figure 6). It has a calculated molecular mass of 35 kDa and PI 7.2. As expected, rGmcathB-GST fusion protein expressed by bacterial expression system was 61 kDa in its molecular mass, which is corresponds to sum of 35 kDa rGmcathB and 26 kDa GST (Figure 6A). rGmcathB-GST from *E. coli* mostly expressed as insoluble form. On the other hand, rGmcathB from *P. pastoris* expression system was resulting at least two major soluble proteins of 35 and 45 kDa, which were recognized by anti-rGmcathB-GST mouse polyclonal antibody (Figure 6B and 6C).

Proteolytic activity of recombinant GmcathB. The proteolytic activity of rGmcathB was only detected in the recombinant protein expressed by *P. pastoris* system. rGmcathB hydrolysed both synthetic dipeptidyl substrate (Z-Arg-Arg-MCA) and fluorogenic casein (BODIPY FL casein). Although rGmcathB showed broad pH range activity from pH 4.0 to 9.0, highest activity was observed at pH4.0 for the two substrates (Figure 7A). Dose dependency of activity was similar between rGmcathB and control enzyme (papain) (Figure 7B). Although clear temperature preference of rGmcathB was observed at 30°C in fluorogenic casein (Figure 7C), such temperature preference was not observed when synthetic dipeptidyl substrate (Z-Arg-Arg-MCA) was used (Figure 7C). Activity was completely inhibited by the specific cysteine protease inhibitor E-64 (Figure 7A and 7B).

Recombinant GmcathB hydrolysed bovine hemoglobin and bovine serum albumin. The ability of rGmcathB to hydrolyse Hb and BSA was performed at pH 4.0 and 30°C according to the result from proteolytic activity of fluorogenic casein (Figure 7A and

7C). rGmcathB (5 μg) hydrolysed Hb and BSA (100 μg) after 18 hr and 48 hr of incubation, respectively. As the results, bands corresponding to Hb monomer (16 kDa) were weakly strained and BSA was digested into small fragments (Figure 8A and 8B). Papain was employed as a positive control for enzymatic activity, and could not hydrolysed Hb and BSA at enzyme to substrate ratio 1:20. However increasing amount of papain (10 μg) hydrolysed both Hb and BSA at 18 hr and 4 hr of incubation, respectively (data not shown). The activity of rGmcathB on both substrates was totally inhibited by E-64.

The kinetic constants of GmcathB. rGmcathB was able to hydrolysed the substrates (Z-Arg-Arg-MCA), and the K_M values obtained from the substrates are 0.29 mM. (Table 3). Using bovine Hb as substrate, K_M value was 0.44×10^{-3} mM. The catalytic efficiency (K_{cat}/ K_M) of rGmcathB for Z-Arg-Arg-MCA was 7.58 $mM^{-1}sec^{-1}$. On the other hand, hydrolytic activity toward Hb was 3.77×10^{3} $mM^{-1}sec^{-1}$, which is approximately 7 times lower than that of papain (2.8×10^{4}) (Table 3).

rGmcathB effected to the viability of trypanosome *in vitro*. Direct treatment of rGmcathB to culture trypanosome resulted in a reduction the number of both BSF (94.1%) and PCF (33.33 %) at 8 h of incubation (Figure 9A). Considerable abnormal cell deaths were observed by thin wet smear. rGmcath B was able to digest whole body of BSF while digested only outer membrane of PCF as shown in figure 9B.

1.4 Discussion

Previous studies have proposed a role of insect cathepsin B in various physiological processes. In addition to the role in food digestion, metamorphosis and embryogenesis, cathepsin B has been reported to play a role in the insect immune system, which might have the potential to kill the trypanosomes (Ahn et al., 2007; Koo et al., 2008). Interest in cathepsin B is increasing and is generally focused on a wide range of pathogenic parasites, which causing malaria, Chagas disease, schistosomiasis and African trypanosomosis as well as their vectors (Cazzulo, 2002; Dvorak et al., 2008; Hansen et al., 2011; Peacock et al., 2012; Pearson, 2001; Yan et al., 2002a). Investigations on the role of systemic responses during trypanosome infection in the tsetse fly suggested that digestive enzymes secreted in the midgut played key roles in both blood meal digestion and tsetse-trypanosome interaction (Bütikofer et al., 2002; Osir et al., 1993). It was reported that midgut proteases cleaved the N-terminus of *T. brucei* procyclin and released the peptide that might influence infection (Acosta-Serrano et al., 2001). Although cathepsin B is one of the tsetse fly midgut digestive enzymes, its enzyme characteristics have not previously been analyzed (Yan et al., 2002a). Gene cloning and subsequent sequence analyses revealed that the deduced amino acid sequence of GmcathB shares most of the common features that have been reported in tsetse cathepsin B (Yan et al., 2002a). GmcathB contained 13 cysteine residues and two possible N-linked glycosylation sites. These N-glycosylation sites may have roles in protein folding, structural stability, substrate recognition and secretion processes. The deduced amino acid sequence of GmcathB contained a catalytic triad (Cys116, His286 and Asn306, the number corresponding to position in AF329480_1), which is responsible for its proteolytic activity. rGmcathB expressed in prokaryotic

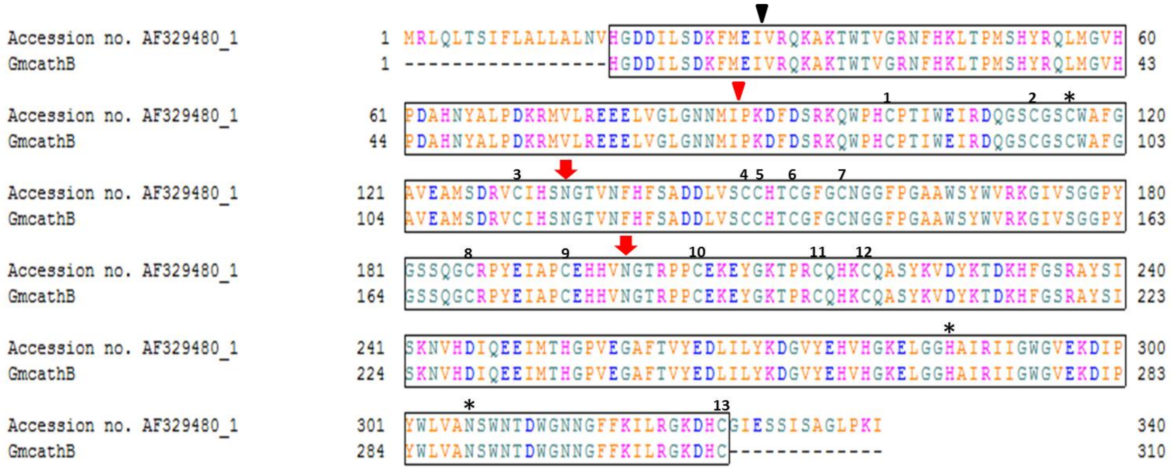
expression system was of insoluble form, and the renatured recombinant protein failed to show enzymatic activity (data not shown). In order to produce soluble and enzymatically active rGmcathB protein, the eukaryotic (*P. pastoris*) expression system was utilized because of its many advantages, including protein processing, protein folding, and the availability of posttranslational modifications. As a result, two major rGmcathB variants, 35 kDa and 45 kDa, were observed. Given that the transcription of gene included the sequence of the AOX1 priming site of the pPinkα-HC vector, 85 amino acid residues with an estimated molecular mass of 10 kDa, the 45 kDa protein would seem to be full-length AOX1-fused rGmcathB. On the other hand, because of the self-catalytic activity of rGmcathB, the 35kDa variant might be a mature enzyme which has proteolytic activity. This is supported by a study on cathepsin B-like protease from the cotton bollworm, which reported that a reducing condition using dithiothreitol plays a role in the activity of enzymes (Zhao et al., 2002).

In the present study, although the highest proteolytic activity of rGmcathB was achieved at pH 4.0, its pH preference was not very strict (Figure 7A). The tsetse midgut lumen is reportedly of high pH (pH 8-10) (Liniger et al., 2003a). This is probably one of the reasons why GmcathB shows catalytic activity over a wide pH range. When fluorogenic casein was used as a substrate, the optimal temperature of the enzyme was 30 °C, which is within the temperature range of the tsetse habitat. Against Z-Arg-Arg-MCA substrate, however, the optimal temperature was somewhat unclear (Figure 7C). In addition, the catalytic efficiency in Z-Arg-Arg-MCA was extremely low (7.58) in comparison to Hb (3.77×10^3) (Table 3). These results suggest that Z-Arg-Arg-MCA is not a suitable substrate for GmcathB. Hydrolytic activity of GmcathB (5 μg) toward Hb and BSA was observed

after 18 and 48 h of incubation, respectively, suggesting that the rGmcathB has higher catalytic activity for Hb than BSA (Figure 8). In contrast, 10 μg of papain had higher catalytic activity for BSA than Hb. These findings were also supported by the K_M values, GmcathB had a higher affinity to Hb than papain (Table 3). In general, the digestion process consists of two steps corresponding to each midgut compartment: endo- and ecto-peritrophic membrane spaces. As a first step of digestion, dietary components, such as proteins, are cleaved by proteases in the endo-peritrophic membrane space (midgut lumen) to release small peptides. Then peptides are digested by peptidase located in the midgut epithelium and ecto-peritrophic membrane spaces (Terra and Ferreira, 1994). Futhermore, proteolytic acitivity of rGmcathB could affect to the viability of BSF than PCF, suggesting PCFs are relatively resistant to GmcathB inside the tsetse midgut. However, induction of abnormal cells was observed from the treatment by rGmcathB in both BSFs and PCFs. From this results refer that Gmcath B may be a good candidate using for killing the parasite. Further investigation is needed to clarify the midgut localization of GmcathB, and involvement of Gmcath B to the transformation of PCF into EMF.

1.5 Summary

A DNA fragment (930 bp) encoding pro-cathepsin B (Accession No. AF329480_1) was cloned and expressed in *E. coli* and *P. pastoris* protein expression system. An active recombinant cathepsin B (rGmcathB) produced by *P. pastoris* exhibited the highest proteolytic activity at pH 4.0 and at 30°C. It has degraded bovine hemoglobin and serum albumin at that optimal condition. Moreover, rGmcathB exhibited hydrolysis preference for bovine hemoglobin more than for Z-Arg-Arg-MCA. The proteolytic activity of rGmcathB was inhibited by specific cysteine protease inhibitor (E-64) confirmed belonging to papain-like cysteine protease family. These results indicated that rGmcathB shows the activity of cathepsin B and have high affinity with blood protein referring a role in blood meal digestion. This study has proved that Gmcath B was able to kill the parasite at early phase of infection and somehow differentiated phase. The result in this study also implied that *P. pastoris* expression system is more suitable for expressing enzymatically active recombinant proteases than *E. coli* expression system because recombinant protein expressed by *E. coli* system was not active. Moreover, this study is the first report to characterize the enzymatic activity of cathepsin B from tsetse.



```
Accession no. AF329480_1     1 MRLQLTSIFLALLALNVHGDDILSDKFMEIVRQKAKTWTVGRNFHKLTPMSHYRQLMGVH  60
GmcathB                      1 -----------------HGDDILSDKFMEIVRQKAKTWTVGRNFHKLTPMSHYRQLMGVH  43

Accession no. AF329480_1    61 PDAHNYALPDKRMVLREEELVGLGNNMIPKDFDSRKQWPHCPTIWEIRDQGSCGSCWAFG 120
GmcathB                     44 PDAHNYALPDKRMVLREEELVGLGNNMIPKDFDSRKQWPHCPTIWEIRDQGSCGSCWAFG 103

Accession no. AF329480_1   121 AVEAMSDRVCIHSNGTVNFHFSADDLVSCCHTCGFGCNGGFPGAAWSYWVRKGIVSGGPY 180
GmcathB                    104 AVEAMSDRVCIHSNGTVNFHFSADDLVSCCHTCGFGCNGGFPGAAWSYWVRKGIVSGGPY 163

Accession no. AF329480_1   181 GSSQGCRPYEIAPCEHHVNGTRPPCEKEYGKTPRCQHKCQASYKVDYKTDKHFGSRAYSI 240
GmcathB                    164 GSSQGCRPYEIAPCEHHVNGTRPPCEKEYGKTPRCQHKCQASYKVDYKTDKHFGSRAYSI 223

Accession no. AF329480_1   241 SKNVHDIQEEIMTHGPVEGAFTVYEDLILYKDGVYEHVHGKELGGHAIRIIGWGVEKDIP 300
GmcathB                    224 SKNVHDIQEEIMTHGPVEGAFTVYEDLILYKDGVYEHVHGKELGGHAIRIIGWGVEKDIP 283

Accession no. AF329480_1   301 YWLVANSWNTDWGNNGFFKILRGKDHCGIESSISAGLPKI                     340
GmcathB                    284 YWLVANSWNTDWGNNGFFKILRGKDHC-------------                     310
```

Figure 5. Alignment of the deduced amino acid sequence of GmcathB from this study with *Glossina morsitans* cathepsin B (Accession No. AF329480_1). To determine the corrected sequence, the deduced amino acid sequence was aligned with the full sequence. The cleavage site of the signal peptide is shown by the black arrowhead. The start of the mature enzyme (amino acid residue from 89 to 340 in AF329480_1) is shown by the red arrow head. The sequence has a catalytic triad of cysteine proteases, Cys116, His286 and Asn306 (asterisks) possess 13 conserved cysteines (numbers 1 - 13), and two potential N-linked glycosylation sites (Asn134 and Asn 199 in AF329480_1).

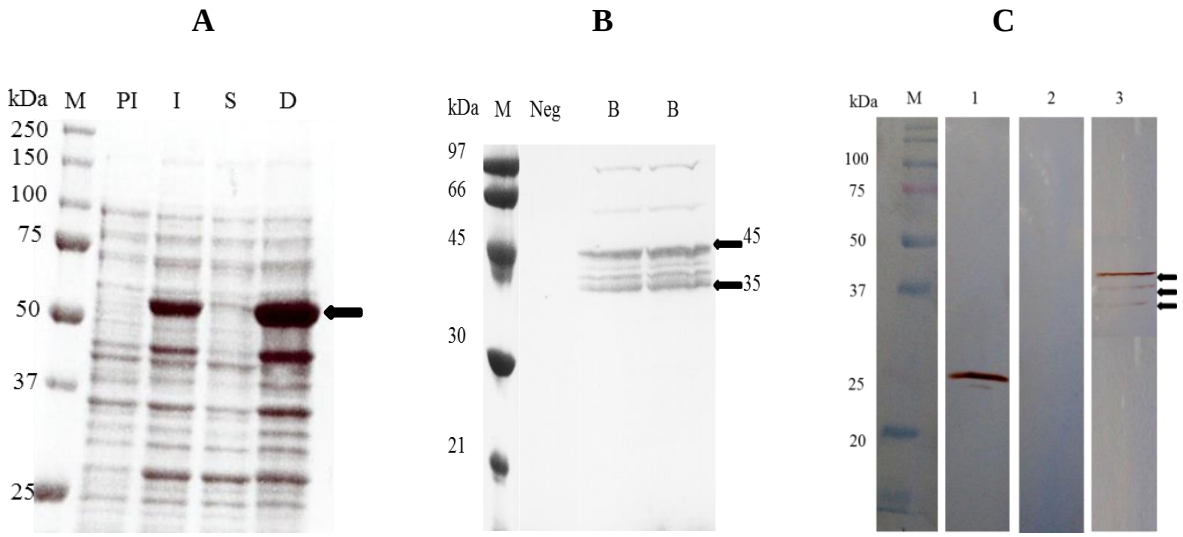


Figure 6. Expression of rGmcathB. A: SDS-PAGE analysis of rGmcathB in the *E. coli* expression system. The glutathione-S-transferase-tagged-rGmcathB protein was expressed as 61 kDa protein (arrow). Total bacterial protein samples from before and after induction of the rGmcathB expression were electrophoresed at lanes PI and I, respectively. Soluble (S) and insoluble (D) fractions of the total protein were electrophoresed. B: rGmcathB was expressed as secreted soluble protein using the *Pichia pastoris* protein expression system, and electrophoresed culture supernatant from before (Neg) and after induction (B), under reducing conditions. The upper band represents the full-length recombinant protein (45 kDa) and the lower band is an active rGmcathB (35 kDa). C: Western blot analysis of rGmcathB from *P. pastoris* with anti-GST-tagged-rGmcathB antibody. Lane 1: positive control (GST protein). Lane 2: negative control (*Pichia* stain 4 that contain pPinkα-HC vector). Lane 3: rGmcathB. The upper arrow indicates full-length recombinant protein, middle and lower arrows show proenzyme and processed mature enzyme after activation with dithiothreitol, respectively. M indicates molecular size marker.

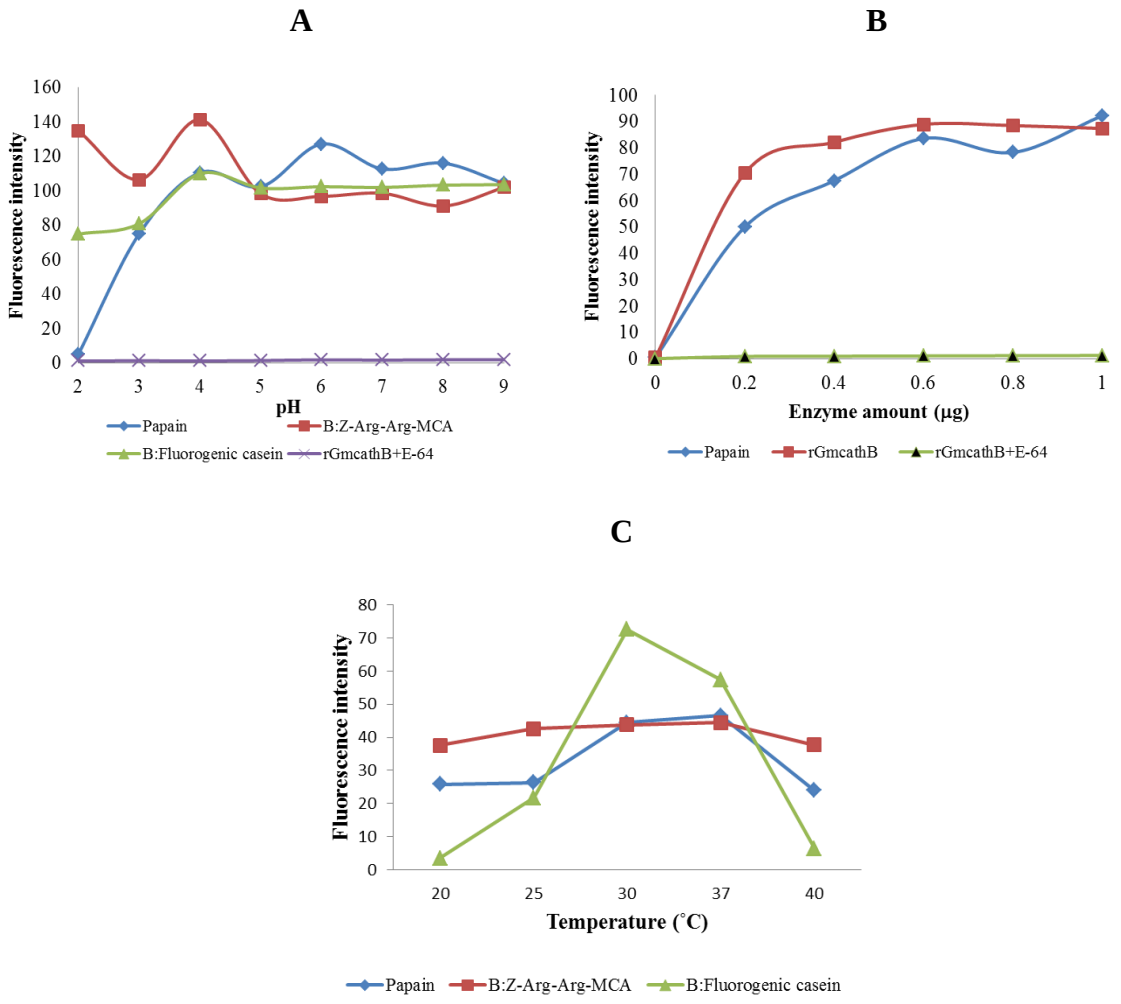
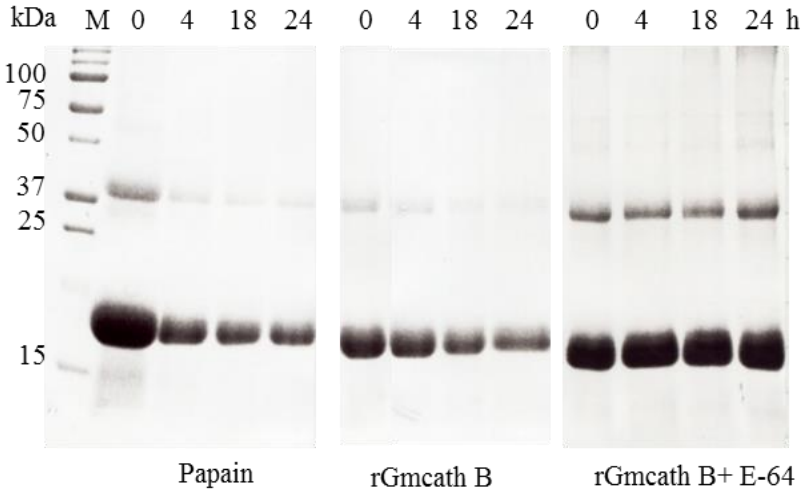


Figure 7. Enzyme characterization of rGmchathB. A: Effect of pH on proteolytic activity of the rGmcathB. Fluorogenic casein and Z-Arg-Arg-MCA were used as substrates. Papain and Z-Arg-Arg-MCA were used for control proteolysis. The amounts of enzyme for fluorogenic casein and Z-Arg-Arg-MCA were 0.8 and 0.5 µg, respectively. E-64 is a specific cysteine protease inhibitor (E-64). B: Dose dependency of rGmcathB and papain. Z-Arg-Arg-MCA was used as a substrate. C: The effect of reaction temperature on the proteolytic activity at pH 4.0.

A



B

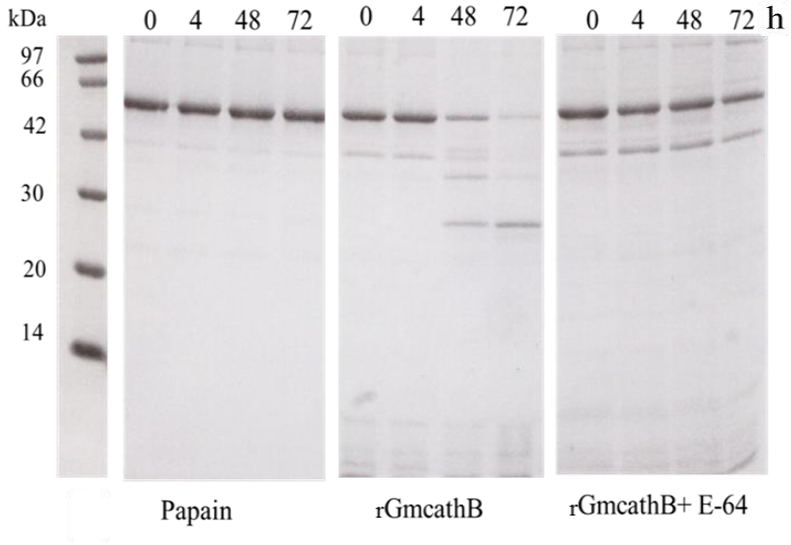
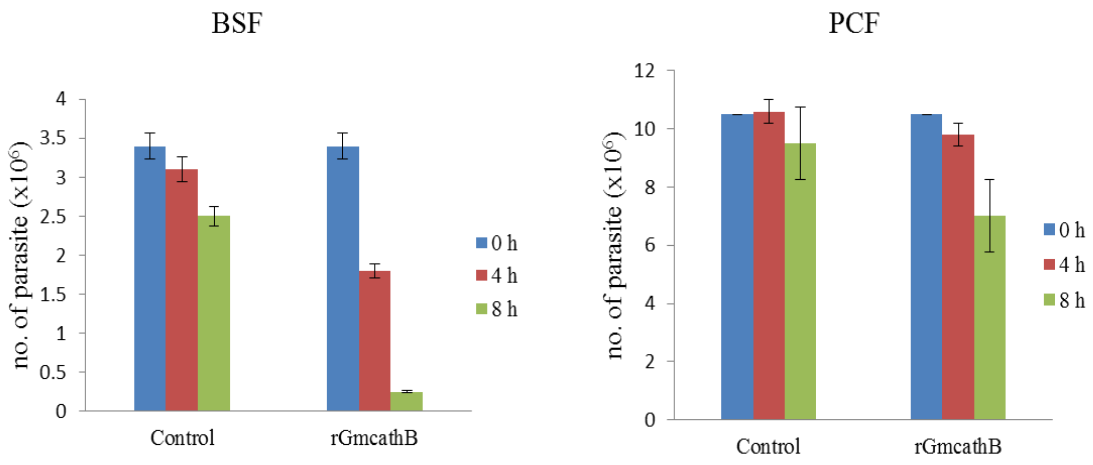


Figure 8. The proteolytic activity of rGmcathB toward bovine hemoglobin (Hb) and bovine serum albumin (BSA) occurred at pH 4.0 at 30 °C. Five μg of rGmcathB degraded Hb (100 μg) after 18 h of incubation (Panel A) and BSA (100 μg) after 48 h of incubation (Panel B). The activity was inhibited by cysteine protease inhibitor E-64.

A



B

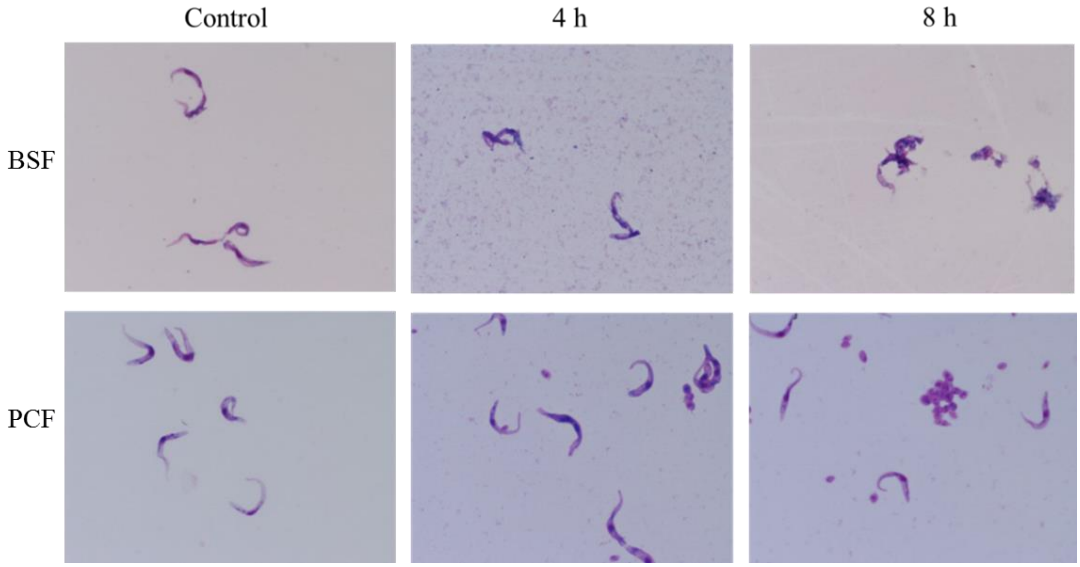


Figure 9. The effect of rGmcathB to *T. congolense*. Panel A: rGmcathB was directly treated to the cultured parasites resulted in a reduction the number of BSF and PCF at 8 h of incubation, 93% and 33.33%, respectively (data not show). Panel B: Thin smear at 8 h of incubation showed that rGmcathB digested the whole body of BSF, and digested outer membrane of PCF and then remained the nucleus.

Table 2. The primer set for RT-PCR of GmcathB

Expression system	Sequence (5'-3')	Target sequence (bp)	Location	Restriction Enzyme
E. coli	F 5' **GAATTC**CATGGAGACGATATATTG R 5' **CTCGAG**TTAGCAGTGGTCCTTGCC	930	52-69 967-981	*Eco* RI *Xho* I
P. pastoris	F 5' **GAGTCG**ACTCCATGGAGACGATATA R 5' **GGTACC**TTAGTGGTCCTTGACGCC	927	52-66 967-978	*Mly* I *Kpn* I

Restriction enzyme sites are indicated by bold typeface. Stop codon is indicated by underline.

Table 3. Kinetic constants for substrate hydrolysis of rGmcath B

Substrate (enzyme)	Enzyme affinity K_M (mM)	Catalytic production k_{cat} (sec^{-1})	Enzyme efficiency k_{cat} / K_M (mM^{-1} sec^{1})
Z-Arg-Arg-MCA (rGmcathB)	0.29	2.20	7.58
Bovine hemoglobin (rGmcathB)	$0.44x10^{-3}$	1.66	$3.77x10^{3}$
Bovine hemoglobin (Papain)	$3.33x10^{-3}$	94.03	$2.8x10^{4}$

Chapter 2: Expression and characterization of cathepsin L from tsetse (*Glossina morsitans morstians*)

2.1 Introduction

Cathepsin L is a lysosomal endopeptidase member of the peptidase C1 family, which is a dimer composed of disulfide-linked heavy and light chains, both produced from a single protein precursor. Cathepsin L is produced as preprocathepsin L, transported via the Golgi apparatus as procathepsin L in secretory vesicles and then stored as mature cathepsin L in lysosome (Katunuma, 1989). The conversion of procathepsin L which is secreted from various cells to mature form occurred by removal of the immature protein possessing an N-terminal pro-region (Ishidoh and Kominami, 1995). Cathepsin L plays important roles in the overall degradation of proteins in lysosome including collagen and elastin, as well as alpha-1 protease inhibitor, a major controlling element of neutrophil elastase activity. This enzyme has been implicated in several pathological processes of human, including myofibril necrosis in myopathy and in myocardial ischemia, and in the renal tubular response to proteinuria. Cathepsin L influences in degrading extracellular protein such as serum protein and cytoplasmic and nuclear protein (Barrett and Kirschke, 1981). It exhibits hormone-regulated expression during ovulation in the rat (Sriraman and Richards, 2004) and acts as a pro-hormone processing in regulated secretory vesicle for the production of peptide neurotransmitters and hormones (Hook et al., 2004).

Cathepsin L has been identified in several groups of insects such as Diptera, Lepidoptera, Coleoptera and Hemiptera (Blanco-Labra et al., 1996; Cristofoletti et al., 2005; Liu et al., 2006; Pyati et al., 2009). It plays several roles in insect life events

including digestion process (Cristofoletti et al., 2005), embryo vitellin degradation (Liu et al., 1996), larval mounting and metamorphosis (Wang et al., 2010). Cathepsin L is related to the transformation of granulocytes to macrogranulocytes to enter the fat body, and induce hemocyte apoptosis for further tissue degradation (Zhai and Zhao, 2012). It is also participated in the differentiation of imaginal discs of *Sarcophaga peregrina* (Homma et al., 1994; Homma and Natori, 1996). Cathepsin L of *Drosophila melanogaster* involved in immune function by participating in the degradation of internalized material in phagocytic cell (Tryselius and Hultmark, 1997). Cathepsin L of *Delia coarctata* (DcCathL) was affected the immune related proteolytic activation cascade that leads to produce an active phenoloxidase and melanization (Pyati et al., 2009). In *Fasciola hepatica* and *Trypanosoma brucei*, cathepsin L is secreted to facilitate tissue invasion, migration, and development within the mammalian host (Grab et al., 2009; Stack et al., 2008). Cathepsin L of trypanosome is suspected to act as pathogenic factor in the mammalian host, where they also trigger prominent immune response so it is proved useful as a diagnostic target of trypanosome (Lalmanach et al., 2002).

In tsetse, cathepsin L (Accession no. ABC48937) was analyzed from fat body of adult flies which significant similarity to vitellogenic protein of mosquito (*Aedes aegypti*) that accumulated in oocyst where they activated proenzyme and allow for utilization of yolk protein during oogenesis (Attardo et al., 2006). However, the enzymatic activity of tsetse cathepsin L has not yet well studied. In addition to the reported cDNA sequence constructed from fat body tissue mRNA. This study has expressed cathepsin L from the midgut of tsetse and characterized its biological significance, namely enzymatic activity, blood meal digestion, and tsetse–trypanosome interaction.

2.2 Materials and Methods

Total RNA isolation and cDNA synthesis. Total RNA isolation and cDNA synthesis were performed as described in chapter 1.

Reverse transcriptase polymerase chain reaction (RT-PCR). First-strand cDNA systhesis was amplified using sense and antisense primers designed to amplified the coding region of *G. m.morsitans* TC 31 cathepsin L like protease (Gmcath L: Accession no. DQ294221.1) (Table 4). The reaction was performed as desctibed in chapter 1.

Cloning and Sequencing of GmcathL. GmcathL gene was cloned and sequenced as described in chapter 1. The alignment of similar amino acid sequences and neighbor jointing phylogenetic tree was built by using ClustalX 2.0.11 program.

Expression and purification of recombinant GmcathL in *Escherichia coli.* Expression and purification of recombinant GmcathL in *E. coli* were performed as described in chapter 1.

Expression of recombinant GmcathL in *Pichia pastoris.* Expression of recombinant GmcathL in *P. pastoris* was performed as described in chapter 1.

Polyclonal antibody production. Antibody against rGmcathL was produced as described in chapter 1.

Characterization of proteolytic activity. Proteolytic activity of rGmcathL was determined as described in chapter 1.

Hemoglobin and BSA degradation assay. Hb and BSA degradation assay were performed as described in chapter 1.

Determination of enzyme kinetic parameters. The hydrolysis efficiencies of rGmcathL toward Z-Phe-Arg-MCA were determined as described in chapter 1

Effect of rGmcathL on trypanosome. Effects of rGmcathL on trypanosome were performed as described in chapter 1. The number of BSF and PCF was 3.1 x 10^6 and 10.5 x 10^6 trypanosomes/ml, respectively.

2.3 Results

Cloning and expression of GmcathL. The ORF 1,551 bp extending from nucleotide position 315 to position 1,866 (Acession no. DQ294221) was cloned from the midgut of adult male tsetse. The deduced amino acid sequence encodes 516 of 550 amino acid residues of pro-cathepsin L (Acession no. ABC48937) was named GmcathL. Computational analysis of the GmcathL amino acid sequence predicted that the sequence including pre-propeptide (225 amino acids, corresponding to amino acid position 21-245 in ABC48937), propeptide inhibitor domain (56 amino acids, corresponding to amino acid position 246-301 in ABC48937), and peptidase_C1A domain (217 amino acid positions, corresponding to amino acid position 332-548 in ABC48937). The sequence contained the catalytic residues of C1 family peptidases, Cys-335 and His-498 (in ABC48937) forming a catalytic dyad that has been likened to the catalytic triad of chymotrypsin. Two other residues play an important role in catalysis: a Gln-349 (in ABC48937) preceding the catalytic Cys, believed to help in the formation of the oxyanion hole; and an Asn-518 (in ABC48937) which orients the imidazolium ring of the catalytic His. The analysis also indicated that this sequence possessed 6 conserved cysteine residues (Figure 10) and Asn434 and Asn446 (in ABC48937) are potential sites of N-linked glycosylation. Phyloge-netic analysis determined that the amino acid sequence showed 82 % identity to cathepsin L of flesh fly (*Sarcophaga peregrina*) (Figure 11). GmcathL proenzyme expressed by two

protein expression systems (*E. coli* and *P. pastoris)* (Figure 12). The deduced amino acid sequence had a calculated molecular mass of 59 kDa and PI 5.94. As expected, the molecular mass of the rGmcathL-GST fusion protein expressed by the bacterial expression system was 85 kDa, which is corresponds to the sum of 59 kDa rGmcathL and 26 kDa GST (Figure 12A). rGmcathL-GST from *E. coli* was mostly expressed as insoluble form. In contrast, rGmcathL from the *P. pastoris* expression system resulted in at least three major soluble proteins of 69, 59 and 45 kDa (Figure 12B). The specific proteins were recognized by anti-rGmcathL-GST mouse polyclonal antibody using reduced SDS-buffer, I29 domain was cleaved by DTT remaining active domain with molecular mass 23 kDa (Figure 12C).

Proteolytic activity of recombinant GmcathL. The proteolytic activity of rGmcathL was only detected in the recombinant protein expressed by the *P. pastoris* system. rGmcathL hydrolyzed both synthetic dipeptidyl substrate (Z-Phe-Arg-MCA) and fluorogenic casein (BODIPY FL casein). Although rGmcathL displayed broad pH range activeity toward Z-Phe-Arg-MCA from pH 4.0-7.0, the highest activity for the two substrates was observed at pH 6.0 similar to the activity of papain (Figure 13A). A clear temperature preference of rGmcathL was observed at 30 °C in fluorogenic casein (Figure 13B), while it was observed in a wide range between 25-37 °C when synthetic dipeptidyl substrate (Z-Phe-Arg-MCA) was used (Figure 13B). The dose dependency of activity in rGmcathL and control enzyme (papain) was similar (Figure 13C). Activity was completely inhibited by the specific cysteine protease inhibitor E-64 (Figure 13A, 13B and 13C).

Recombinant GmcathL hydrolyzed bovine hemoglobin (Hb) and bovine serum albumin (BSA). The ability of rGmcathL to hydrolyze Hb and BSA was performed at pH 6.0 and 30 °C according to the results relating to the proteolytic activity of fluorogenic

casein and Z-Phe-Arg-MCA (Fig. 13A and 13B). rGmcathL (4 μg) hydrolyzed BSA and Hb (20 μg) after 1 h and 24 h of incubation, respectively. As a result, band corresponding to Hb (16 kDa) was weakly strained and the band corresponding to BSA (66 kDa) was digested into smaller fragments (Figure 14A and 14B). Papain was employed as a positive control for enzymatic activity. Four μg of papain completely digested Hb (20 μg) at 24 of incubation and ten μg degraded the main band of BSA (100 μg) into several smaller fragments at 4 h of incubation (Figure 14A and 14B). The activity of rGmcathL on both substrates was totally inhibited by E-64.

The kinetic constants of rGmcathL. rGmcathL was able to hydrolyze the substrate Z-Phe-Arg-MCA gave $K_{M,}$ and k_{cat} value of 0.17 mM and 1.86 s^{-1}, respectivey. The consequence value, a catalytic efficiency (k_{cat} / K_M) was 10.9 $mM^{-1}s^{-1}$. The kinetic constants of papain were K_M = 0.065 mM, k_{cat} = 1.61 s^{-1} and catalytic efficiency was 24.8 $mM^{-1}s^{-1}$. In comparison to papain, the kinetic costants of rGmcathL was approximately 2 times lower than that of papain (Table 5).

The effect of rGmcathL to trypanosome. rGmcathL was added to the cultured trypanosome resulting reduction of the number of BSF 38.7% from the initial number (Figure 15A). The abnormal BSF cells were observed at 8 h of incubation (Figure 15B). On the other hand, PCF showed resistance against rGmcathL treatment.

2.4 Discussion

This study has expressed and characterized cathepsin L cysteine protease from the midgut of tsetse. Moreover, this study also investigated the effect of recombinant GmcathL on trypanosome *in vitro*. GmcathL from the midgut of tsetse showed the highest level of identity to the cathepsin L-like protease of *Sarcophaga peregrina.* Similar to other lysosomal cathepsin proteases, GmcathL was synthesized as inactive precursor. As a result, the expression of GmcathL gene including the sequence of the AOX1 priming site of pPinkα-HC vector, 85 amino acid residues with an estimated molecular mass of 10 kDa, the 69 kDa protein would seem to be full-length AOX1-fused rGmcathL. The bands at 59 and 45 kDa are corresponding to pre-propeptide and propeptide, respectively. This sequence was activated by DTT and cleaved off pro-peptide that contained autoinhibitor (I29 domain) at N-terminal, remained peptidase_C1A domain which has protein digestion function. Therefore by Western blot anylysis, only the band of peptidase_C1A domain was shown (Figure 12C).

rGmcathL from *P. pastoris* expression system showed optimal activity at pH 6.0, however it also exhibited the activity with wide range pH from 5.0 to 7.0. The previous studies in other insect reported that activated ScathL (with reducing agent) of *S. peregrina*, was hydrolysed Z-Phe-Arg-AMC with optimal activity at pH 5.5 (Philip et al., 2007). Expression of cathepsin L of *Triatoma brasilliensis*, a main vector of *T. cruzi*, was highly upregulated at pH 5 in the lumen of posterior midgut, which is the main region of blood digestion (Waniek et al., 2012). In addition, the optimal pH of rGmcathL was corresponded to the typical pH optimum of lysosomal cathepsins from arthropods and insects (pH 3.5-6.0). This result was slightly different from the gut pH (7.8 to 10) of teneral tsetse

flies 48 h after first blood meal (Liniger et al., 2003b). The possible reasons of differences are as follows: (i) GmcathL might active in the acidic compartment, for instance lysosome and endosome (pH 4.5- 6.0) of gut cell, or (ii) the activity of Gmcath L adapts to pH value when enzyme mixed with ingested blood meal (pH 7.4) inside the gut lumen. However,it is still unclear whether GmcathL located inside the gut cell or secreted into the gut lumen. Therefore, the further characterization of endogenous GmcathL is needed. Temperature preference of rGmcathL is between 25- 37 ˚C, which is in range of conditions of tsetse habitat (25 ˚C).

The catalytic activity against Z-Phe-Arg-MCA demonstrated a high affinity to arginine at P1 position and substrate containing hydrophobic amino acid in P2 position (phenylalanine). Agreement with the previous results which demonstrated that cathepsin L specificity was largely based on P2 selectivity (Choe et al., 2006; Greenbaum et al., 2002). This is the characteristic feature of the papain family, the preference is for bulky hydrophobic or aromatic residues at the P2 side chain of the substrate to occupy the S2 subsite. An exception is cathepsin B, which readily accepts arginine in this subsite because of the presence of a glutamic acid in the bottom of the pocket. This property can use for differentiation of cathepsins L from other cysteine type cathepsins. Similar to cathepsin L from other source, the activity of GmcathL was inhibited by specific cysteine protease inhibitor E-64.

The hydrolysis activity of rGmcathL upon protein contents in blood meal was in part clarified by higher affinity to BSA than Hb. This result was coincided with the hydrolytic activity of papain which had higher affinity to BSA than Hb. But, contrast with the results of rGmcathB in chapter 1 that exhibited higher catalytic activity to Hb than BSA.

However, this result suggest that GmcathL is involved in blood meal digestion. So far, the evaluation of the effect of rGmcathL to trypanosome was done *in vitro*. The results showed slight decrease of the numner of BSF but not PCF. This suggests that GmcathL affects the viability of BSF and may involve in elimination of parasite at early phase of infection in tsetse midgut.

2.5 Summary

A cDNA encoding a cysteine protease named TC 31 cathepsin L-like protease, which was previously identified in the fat body of adult tsetse, was cloned from a midgut of adult male tsetse flies (48 h post blood meal). The cDNA encoded 516 amino acid residues having the highest identity to cathepsin L of *S. peregrina* (82 %). The amino acid sequence contained conserved amino acid residues, Cys-His forming the catalytic dyad of an active site of C1A cysteine protease family, an Asn orients the imidazolium ring of the catalytic His, and a Gln helps to form the oxylation hole. The recombinant GmcathL (rGmcathL) was successfully expressed by using *P. pastoris* protein expression system. Activation of rGmcathL was occurred in presence of DTT (2.5 mM) and cysteine (10 mM). Proteolytic activity was exhibited the highest at pH 6.0 with broad temperature range from 25 to 37 ^{o}C. The proteolytic activity of rGmcathL was completely inhibited by broad spectrum cysteine protease inhibitor (E-64). The catalytic efficiency (k_{cat}/K_M) of rGmcathL was obtained from hydrolysis of Z-Phe-Arg-MCA indicating rGmcathL exhibited hydrolysis preference for amino acid that has aromatic ring and hydrophobic side chain (Phe) at P2 position. Moreove, rGmcathL was capable of degrading both hemoglobin (Hb) and bovine serum albumin (BSA), but it preferred to cleave BSA as compared to Hb. Co-cultivation of rGmcatL and trypanosome resulted in a reduction the number of BSF, while PCF showed relative resistant to the treatment. The results from this study indicated that GmcathL may involve in bloodmeal digestion and parasite elimination.

```
GmcathL            LGSHYDHYYLDYDSYEHGDIPNVIFELDPDMECVDFPGPGKGHYATFNPMEEFIHPAIQH   220
G.m.m_ABC48937     LGSHYDHYYLDYDSYEHGDIPNVIFELDPDMECVDFPGPGKGHYATFNPMEEFIHPAIQH   240
S.p_BAA76272.1     LGSHYDHYYLDYDSYDHDDIPNEVFELDDNMECMPFPGPGKGHYATFNPMQEFVHPTVDH
A.e_ABE72972.1     LGSHYDHYYLDYDSYEHQDIPADVFQVKTSDPCIGFPGPGNGHYATFNPMQEFIHPRSEE
R.p_AAL34984.1     -------------------MLIPSFDIDPQ---------------------------
H.s_AAA66974.1     ---------------MNPTLILAAFCLGIASATLTF----------------------DH
                                  :    * :
                                                I29
GmcathL            HVDDAFNHFKNKHNIKYHTEEDHEYRKNVFRQNLRFIQSKSR----AQLNYALAINHLAD   276
G.m.m_ABC48937     HVDDAFNHFKNKHNIKYHTEEDHEYRKNVFRQNLRFIQSKSR----AQLNYALAINHLAD   296
S.p_BAA76272.1     HVEHAFKHFKNKHGIDYRTEQEHEYRKNIFRQNLRFINSKNR----GKLSYTLAINHLAD
A.e_ABE72972.1     HLDNEFTRFRYKHGKSYHNEKEHDLRRDIFRQNLRFIHSHNR----AGKGFTVAVNHLAD
R.p_AAL34984.1     ----EWLAFKAMHGKNYRNQFEEIFRMKVFIDNKKKIDEHNAKYELGEASYKMKMNHLGD
H.s_AAA66974.1     SLEAQWTKWKAMHNRLYG-MNEEGWRRAVWEKNMKMIELHNQEYREGKHSFTMAMNAFGD
                        :  ::  *.  *    :.  *  :: .* : *. :.     .  .: : :* :.*
                                                   Peptidase C1A  Q
GmcathL            KTDEELKAIR-GFRSSG-VYNTGKPFPYDVSKLKDQIPDQYDWRLYGAVTPVKDQSVCGS   334
G.m.m_ABC48937     KTDEELKAIR-GFRSSG-VYNTGKPFPYDVSKLKDQIPDQYDWRLYGAVTPVKDQSVCGS   354
S.p_BAA76272.1     KSDDELKGRR-GYKSSG-VFNTGKPFPYNLEKYRDTVPDQYDWRLYGAVTPVKDQSVCGS
A.e_ABE72972.1     RTDEELKALR-GFKSSN-IYNGGQPFPYNPEDFKDELPESLDWRIAGAVTPVKDQSVCGS
R.p_AAL34984.1     LMVHEFKALMNGFKKTPNAERNGKIY----VPSNENLPKSVDWRQRGAVTPVKDQGHCGS
H.s_AAA66974.1     MTSEEFRQVMNGFQNRK--PRKGKVFQ---EPLFYEAPRSVDWREKGYVTPVKNQGQCGS
                      .*::    *::.     . *: :           * . ***  * *****:*. ***
                                                                S2
GmcathL            CWSFGTIGHIEGAYFLKNGGNLVRLSQQALIDCSWQYGNNGCDGGEDFRSYQWMMQMGGV   394
G.m.m_ABC48937     CWSFGTIGHIEGAYFLKNGGNLVRLSQQALIDCSWQYGNNGCDGGEDFRSYQWMMQMGGV   414
S.p_BAA76272.1     CWSFGTIGHLEGAFFLKNGGNLVRLSQQALIDCSWEYGNNGCDGGEDFRAYKWMMEMGGV
A.e_ABE72972.1     CWSFGTAGHIESAYFLKYN-KLMRFSQQALIDCSWGYGNNGCDGGEDFRAYQWMMEVGGI
R.p_AAL34984.1     CWSFSATGSLEGQLFLKTG-RLVSLSEQNLVDCSKTYGNSGCEGGLMNQAFQYVRDNKGI
H.s_AAA66974.1     CWAFSATGALEGQMFRKTG-RLISLSEQNLVDCSGPQGNEGCNGGLMDYAFQYVQDNGGL
                   **:*.: * :*.  * * . .*: :*:* *:***   **.**:**    ::::: :  *:
                                                                        S2
GmcathL            PTEEEYGPYLGQDGYCHAQNATLVAPIKGFVNVTSNDPDAFKIALLNHGPLSVAIDASPK   454
G.m.m_ABC48937     PTEEEYGPYLGQDGYCHAQNATLVAPIKGFVNVTSNDPDAFKIALLNHGPLSVAIDASPK   474
S.p_BAA76272.1     PTEEEYGPYLGQDGYCHAKNVSLVAPITGFFNVTPNDPMALKIALLKHGPISVAIDASPK
A.e_ABE72972.1     PSEEEYGGYLGQDGYCRLENKTLYAAIDGWVNVTSGDAEAMKVALFKHGPLSIAIDAGHK
R.p_AAL34984.1     DTEASY-PYEARENNCRFKEDKVGGTDKGYVDILEASEKDLQSAVATVGPISVRIDASHE
H.s_AAA66974.1     DSEESY-PYEATEESCKYNPKYSVANDTGFVDIPKQ-EKALMKAVATVGPISVAIDAGHE
                    :* .*  * . :  *: :     .   *:.::       :  *: . **:*: ***. :
                                     S2 S2                    N
GmcathL            TFSFYSHGVYYEPSCKNSMDSLDHAVLAVGYGKMN----DEHYWLIKNSWSTYWGNDGYV   510
G.m.m_ABC48937     TFSFYSHGVYYEPSCKNSMDSLDHAVLAVGYGKMN----DEHYWLIKNSWSTYWGNDGYV   530
S.p_BAA76272.1     TFSFYSHGVYYEPTCKNGLDELDHAVLAVGYGTIN----GEDYWLVKNSWSTYWGNDGYI
A.e_ABE72972.1     SFSYYANGVYYEPECRNNLDGLDHAVLAVGYGKLK----GEDYWLIKNSWSNYWGNDGYA
R.p_AAL34984.1     SFQFYSEGVYKEQYCSP--SQLDHGVLTVGYGTE----NGQDYWLVKNSWGPSWGESGYI
H.s_AAA66974.1     SFLFYKEGIYFEPDCSS--EDMDHGVLVVGYGFESTESDNNKYWLVKNSWGEEWGMGGYV
                   :* :* .*:* *  *    . :**.**.****       .:.***:****.  ** .**
                            6    S2
GmcathL            LMS-ARE--------------   516
G.m.m_ABC48937     LMS-ARENNCGVMTTPTFVEM-   550
S.p_BAA76272.1     LMS-ARKNNCGVMTMPTYVEM-
A.e_ABE72972.1     LMA-MKDNNCGLSTDATYVNLK
R.p_AAL34984.1     KIARNHKNHCGIASMASYPVV-
H.s_AAA66974.1     KMAKDRRNHCGIASAASYPTV-
                    ::  :
```

Figure 10. Alignment of amino acid sequence of GmcathL. *Glossina m. morsitans*, (G.m.m ABC48937) and amino acid sequences of Cathepsin L from *Homo sapeins* (H.s_ AAA66974.1), *Rhodnius prolixus* (R.p_AAL34984.1), *Sacophaga peregrine* (S. p_BAA76272.1) and *Aedes aegypti* (A.e_ABE72972.1) were aligned. The catalytic residues of C1A family peptidase Cys and His, forming a catalytic dyad are indicated by red capitals. A Gln (Q) preceding the catalytic Cys to help in the formation of the oxyanion

hole; and an Asn (N) residue orients the imidazolium ring of the catalytic His. The yellow labeled letter indicated the N-glycosylation sites, and six conserved cysteine residues are labeled in gray. Inhibition_29 domain (I29) starts from amino acid position 246-301 (in ABC48937). Active region name “peptidase C1A” are started at PDQYD until the amino acid residue number 548 (in ABC48937). S2 subsites (S2) is the dominant substrate specificity subsites of papain cysteine proteases.

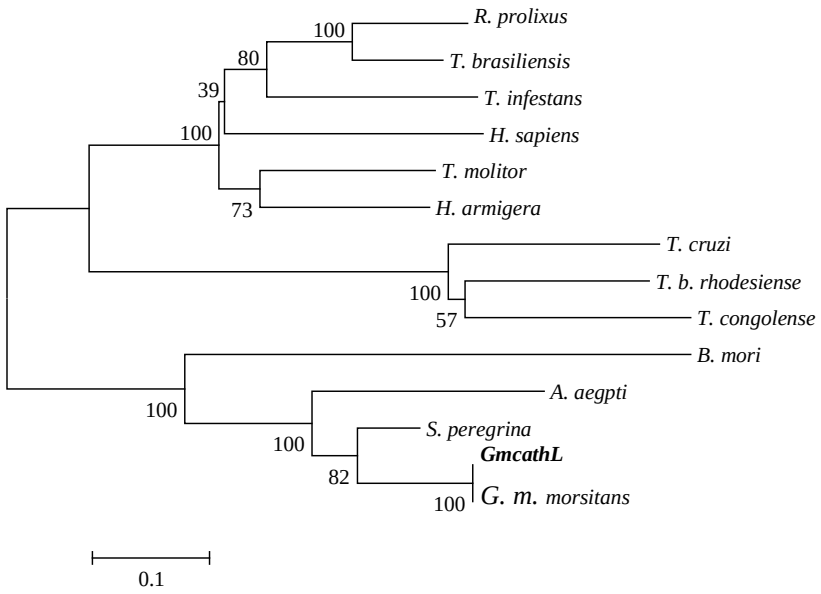


Figure 11: Phylogenetic tree of Gmcath L. Neighbor-joining dendrogram of GmcathL and other cathepsin L-like proteases from different sources. Branch lebels below indicate the percentage of bootstrap proportion (1000 replicates) supporting that brance.

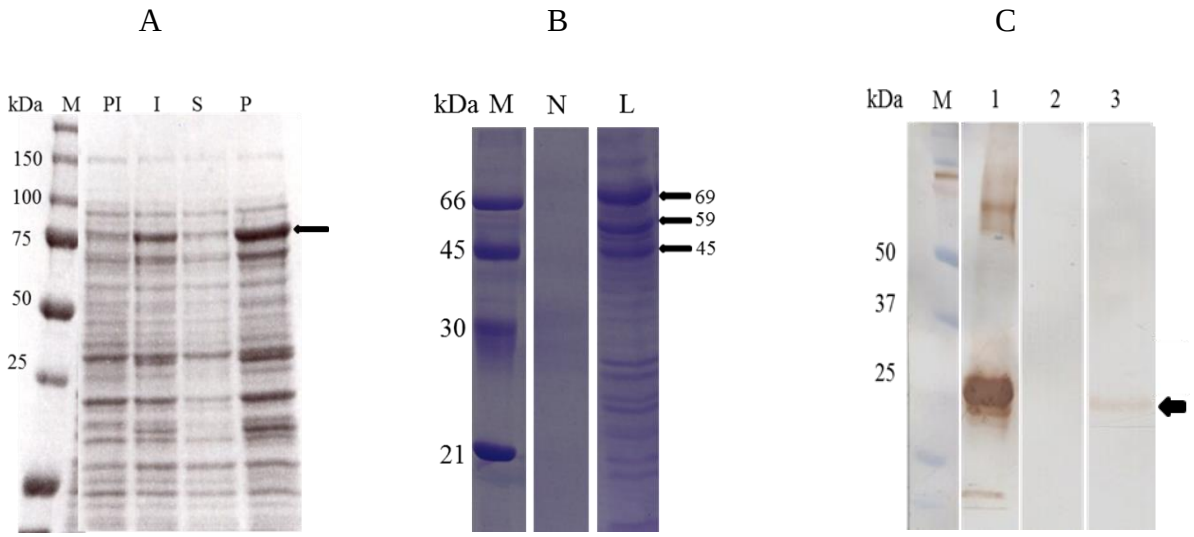
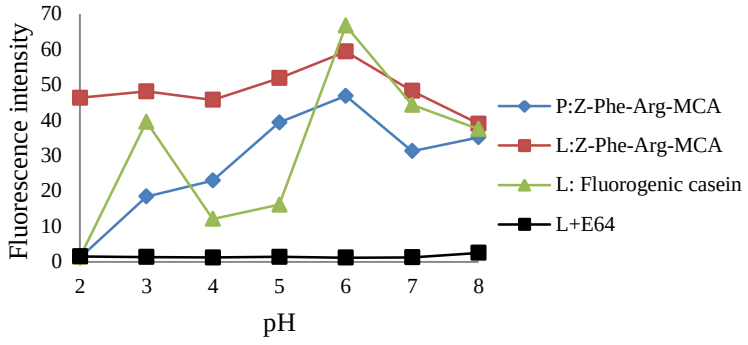
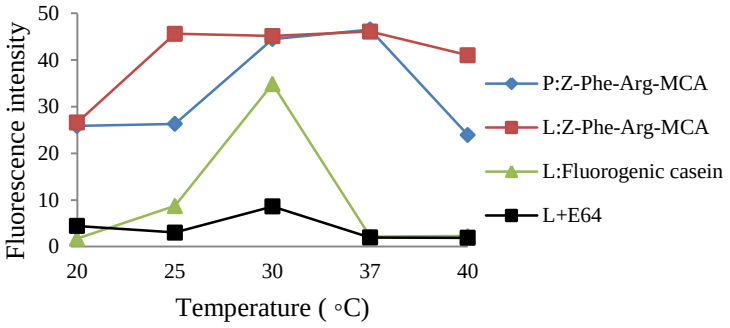


Figure 12. Expression of rGmcathL: A: SDS-PAGE analysis of rGmcathL in the *E. coli* exprression system. The glutathione-S-transferase-tagged-rGmcathL protein was expressed as 85 kDa protein (arrow). Total bacterial protein samples from before and after induction were electrophorested at lane PI and I, respectively. Soluble (S) and insoluble (P) fraction of total protein were electrophorased. B: rGmcathL was expressed as secreted soluble protein using *P. pastoris* protein expression system, and electrophorased culture supernatant from before (N) and after induction (L), under reducing condition. The upper band represents the full-length of recombinant protein (69 kDa), middle band and lower band is pre-propeptide (59 kDa) and propeptide (45 kDa), respectively. C: Western blot analysis of rGmcathL from *P. pastoris* with anti-GST-tagged-rGmcathL antibody. Lane 1: positive control (GST protein). Lane 2: negative control (*Pichia* stain 4 that contain pPinkα-HC vector). Lane 3: rGmcathL. The I29 domain was cleaved off after activation with dithiothreitol remained active domain C1A with molecular mass of 23 kDa (arrow).

A



B



C

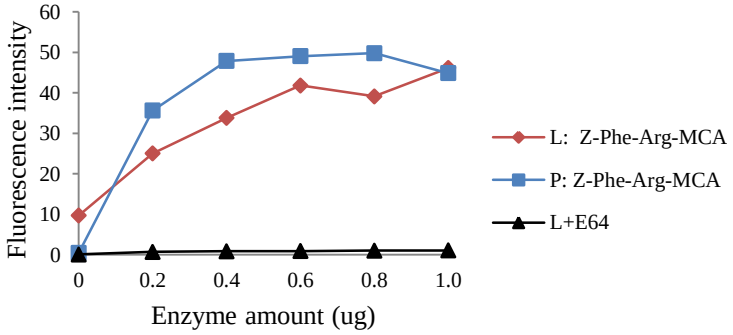
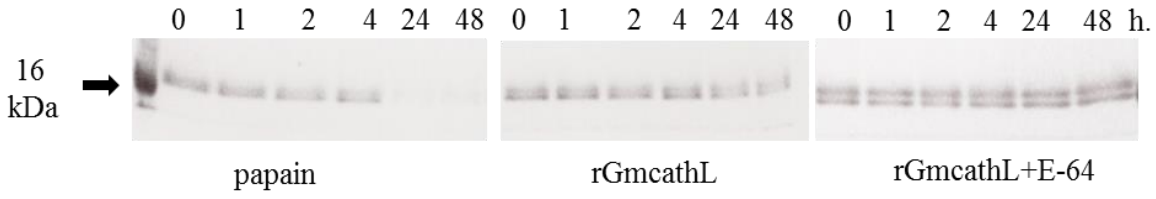


Figure 13. Proteolytic activity of rGmcath L toward Z-Phe-Arg-MCA and fluorogenic casein. rGmcath L exhibited the activity at optimal pH 6.0 (A), temperature range between 25-37 ˚C (B) and on a dose dependent manner (C). The activity of rGmcath L was inhibited by broad spectrum cysteine protease inhibitor (E64). P: papain, L: rGmcath L

A



B

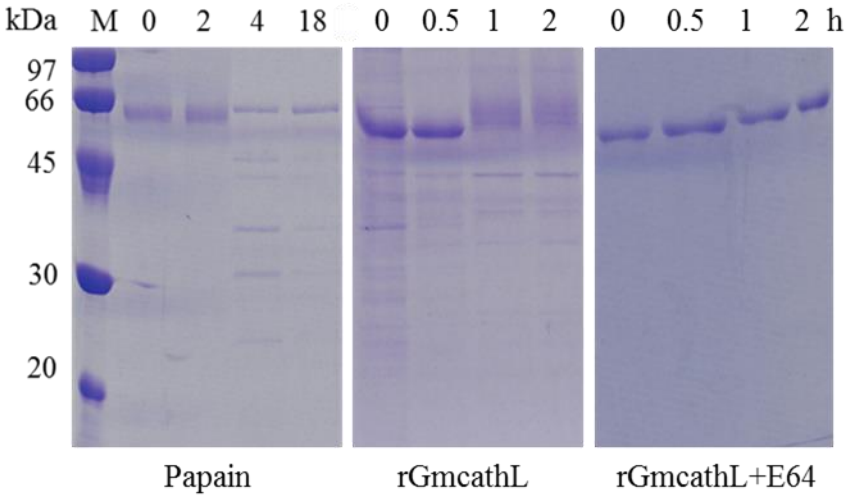
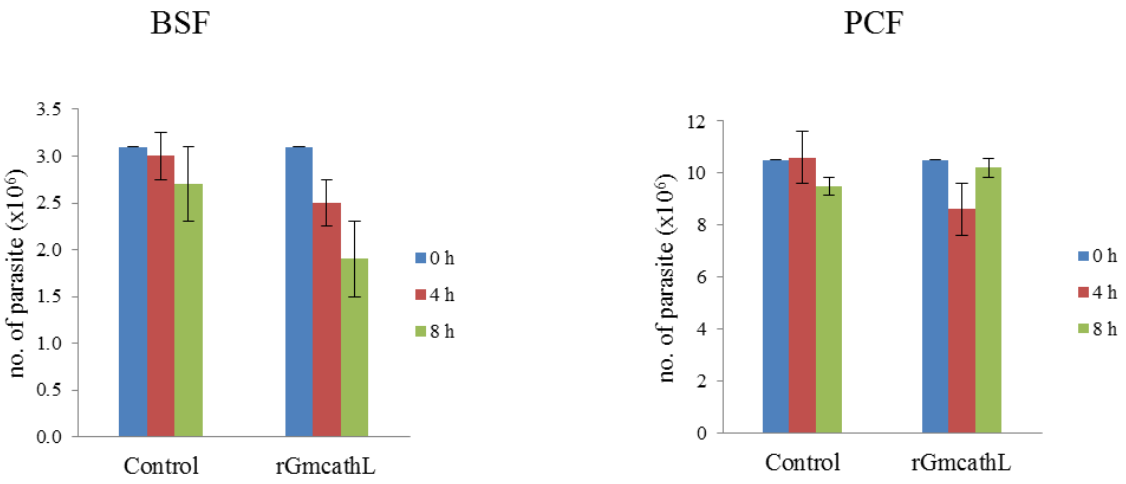


Figure 14. The proteolytic activity of rGmcathL toward bovine hemoglobin (Hb) and bovine serum albumin (BSA) occured at pH 6.0 at 30 °C. A: four µg of rGmcath digested Hb (20 µg) after 24 h of incubaion whereas same amount of papain showed completely digestion at 24 h of incubation. B: BSA (20 µg) were digested by rGmcathL (5 µg) into several small fragments after 1 h of incubation, however ten µg of papain digested BSA after 4 h of incubation. The activity was inhibited by broad spectrum cysteine protease inhibitor E-64.

A



B

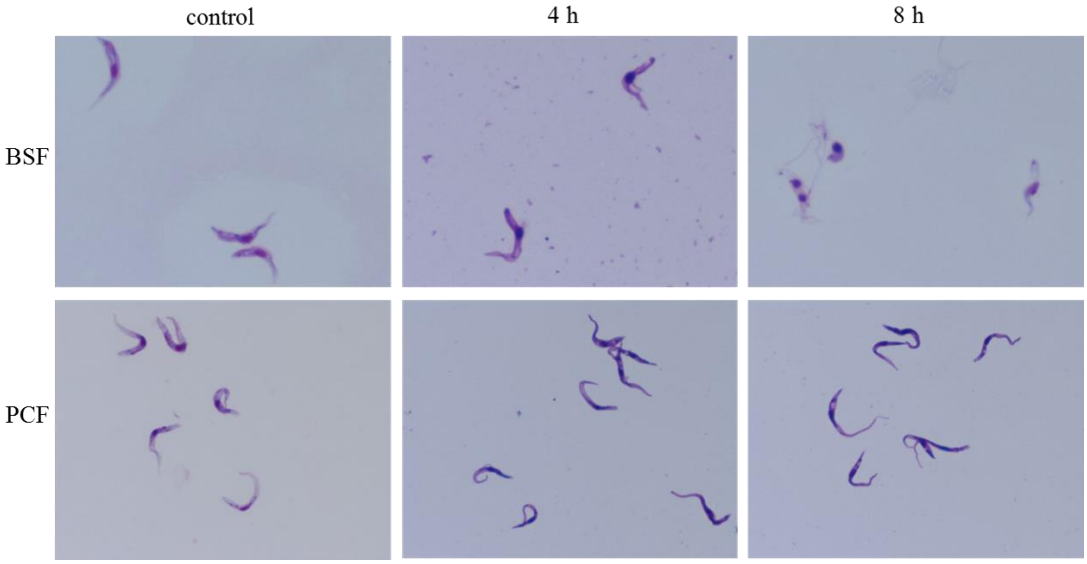


Figure 15. The effect of rGmcathL to *T. congolense*. Panel A: rGmcath L was added to the cultured parasites resulting in a reduction the number of BSF 38.7 %. Panel B: Light microscope photograph of BSF and PCF after treated with rGmcathL at 4 and 8 h of incubation. The results showed that outer membrane of BSF was digested by rGmcathL.

Table 4. The primer set for RT-PCR of GmcathL

Expression system	Sequence (5'-3')	Target sequence (bp)	Location	Restriction Enzyme
E. coli	F 5' **GAATTC**AAACCACCTCGTTGGGAT	1,548	61-78	*EcoR* I
	R 5' **CTCGAG**TTA TTCGCGCGCTGACAT		1594-1608	*Xho* I
P. pastoris	F 5' **CCCGGG**AAACCACCTCGTTGGGAT	1,548	61-78	*Sma* I
	R 5' **GGTACC**TTATTCGCGCGCTGACAT		1594-1608	*Kpn* I

Restriction enzyme sites are indicated by bold typeface. Stop codon is indicated by underline.

Table 5. Kinetic constants for Z-Phe-Arg-MCA hydrolysis of rGmcath L

Enzymes	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1} s^{-1})
rGmcathL	0.17	1.86	10.9
Papain (positive control)	0.065	1.61	24.8

Chapter 3: Expression and characterization of trypsin-like serine protease from tsetse (*Glossina morsitans morsitans*)

3.1 Introduction

Trypsin-like serine proteases (E*C* 3.4.21.4) are ubiquitous in eukaryote such as vertebrate and invertebrate. They are responsible for many essential functions including digestion process, homeostasis, immune response process, and nervous system (Wang et al., 2008b). Trypsin has a catalytic triad composed of histidine, aspartic acid and serine residues. These three amino acid residues form a charge relay that serves to make the active site serine nucleophilic. In the enzymatic reaction of trypsin, it is thermodynamically favorable but requires significant activation energy. Trypsin represents as a principal digestive alkaline endopeptidase in most insects including blood sucking insect, which involves in the initial phase of polypeptide chain digestion. In general, trypsin is produced as an inactive proenzyme trypsinogen then activated at optimal pH from 8 to 9 (Oliveira et al., 2005). Trypsin prefers to cleave peptide chains mainly at the carboxyl side of arginine and lysine (Olsen et al., 2004). It's hydrolysis of peptide bonds break down proteins into smaller peptides which are further hydrolyzed into amino acids by peptidases.

In invertebrate, serine proteases are reported to have a roles in digestion (Muller et al., 1993), hemolymph coagulation (Iwanaga et al., 1998), antimicrobial peptide synthesis (Tong et al., 2005), melanization and the activation of immune pathway in response to the pathogen (Abt and Rivers, 2007). Insect trypsin has been extensively studying in mosquitoes such as *Aedes* spp., *Culex* spp., and *Anopheles* spp.. The most profound role of mosquito trypsin-like serine proteases is digestion of blood meal (Barillas-Mury et al.,

1995). They are also involved in activation of the prophenoloxidase enzymatic cascade that is induced for melanization of mosquitos (Gorman et al., 2000). Futhermore trypsin has been implicated in the establishment of infection. For example in *Aedes aegypti*, trypsin affected replication and dissemination of dengue type 2 flavivirus (Molina-Cruz et al., 2005). Potentially, trypsin activated a chitinase of *Plasmodium gallinaceum* which necessary for ookinetes to escape from the perithrophic matrix (Shahabuddin et al., 1996). Inhibition of trypsin resulted in significant retardation of the parasite growth in mosquito larvae (Casu et al., 1994).

The previous study in tsetse reported that activity level of *Gsp2* gene encoding trypsin-like serine protease (Accession no. AFF91346.1) in the gut lumen increased following blood feeding, and changed significantly in the gut cell throughout the digestion cycle (Yan et al., 2001). Trypsin from the midgut of *Glossina m. morsitans* cleaved the major surface glycoproteins of procyclic forms of *T. brucei* (EP and GPEET procyclin) (Liniger et al., 2003c). In addition, trypsin is capable of inducting transformation of *T. b.* gambiense and *T. b. brucei* from BSF into PCF (Abubakar et al., 2006; Yabu and Takayanagi, 1988). Hence, trypsin may have an important role in tsetse-trypanosome interaction. In this chapter, trypsin-like serine protease precursor, which consisted of 227 amino acid residues, was cloned from the midgut of tsetse. The recombinant protein (rGmtrypsin) was characterized for its enzymatic activity. Futhermore, rGmtrypsin was inferred the role in tsetse-trypanosome interaction, namely the effect on viability of trypanosome and involvement on transformation of PCF into EMF.

3.2 Materials and Methods

Trypanosomes: *Trypanosoma congolense* wild type strain was used, derived from the stock *Trypanosoma congolense* IL 3000.

Total RNA isolation and cDNA synthesis: Total RNA isolation and cDNA synthesis were performed as described in chapter 1.

Reverse transcriptase polymerase chain reaction (RT- PCR): The first strand cDNA was amplified using primers, sense 5'-**GAATTC**CACAATGACCATACAGCC-3 and anti-sense 5'-**CTCGAG**TTAATCACGTAGAGCGGC-3, which were designed to amplified coding region of *G. m. morsitans* trypsin-like serine protease proenzyme gene (Gmtrypsin). The reaction was performed as described in chapter 1.

Cloning and sequencing of Gmtrypsin. Cloning and sequencing of Gmtrypsin was performed as described in chapter 1. The alignment of similar amino acid sequences and neighbor jointing phylogenetic tree was built by using ClustalX 2.0.11 program.

Expression and purification of recombinant Gmtrypsin in *Escherichia coli.* Expression and purification of recombinant Gmtrypsin in *E. coli* were performed as described in chapter 1. However, *E. coli* expression host strain Rosetta (EMD Biosciences Inc., USA) was used instead of BL21.

Characterization of proteolytic activity: Proteolytic activity of rGmtrypsin was measured by releasing of p-nitroanilide (pNA) from specific substrate, Nα-Benzoyl-L-arginine 4-nitroanilide hydrochloride (BApNA) (Sigma-Aldrich, CA, USA)and toward Benzyloxycarbonyl-L-phenylalanyl-L-arginine-4-methylcoumaryl-7-amide (Z-Phe-Arg-MCA) (Sigma-Aldrich, MO, USA). The reactions were performed as described in chapter 1.

Hemoglobin and BSA degradation assay. Hb and BSA degradation assay were performed as described in chapter 1.

Determination of enzyme kinetic parameters. Determination of enzyme kinetic parameters using BApNA substrate was performed as described in chapter 1.

Effect of rGmtrypsin on trypanosome. The effect of rGmtrypsin on trypanosome was performed as described in chapter 1. The number of BSF and PCF is 8.2 x 10^5 and 1.6 x 10^6 trypanosomes/ml, respectively. The transformation of PCF into EMF was further confirmed by Hoechst 33342 (Invitrogen, CA, USA) staining, and observed by confocal microsope scanning.

3.3 Results

Cloning and expression of Gmtrypsin: The cDNA with 681 bp long for a polypeptide of 227 amino acid residues was *Glossina morsitans morsitans* trypsin-like serine protease precursor (AAF91346.1), it was named Gmtrypsin. Computational analysis of Gmtrypsin predicted that the sequence encodes 227 of 255 amino acid of pro-Gmtrypsin (AAF91346.1), including pro-peptide (8 amino acids, corresponding to amino acid position 20-27 in AAF91346.1) and active mature peptide (219 amino acids, corresponding to amino acid position 28-246 in AAF91346.1). Inactive proenzyme has pro-peptide that are cleaved at cleavage motif: GR-IVNGV during limited proteolysis to generate active enzyme. The sequence composed of the characteristic catalytic triad of serine protease, His70-Asp115-Ser209 (number is corresponding to amino acid in AAF91346. 1), six-conserved cysteine residues and a conserved substrate specificity (Asp203-Ser224-Gly226; number is corresponding to amino acid in AAF91346.1). The organization and

alignment of Gmtrypsin with trypsin from other sourses are illustrated in figure 16A. Phylogenetic analysis indicated that Gmtrypsin was close to midgut-specific serine protease 2 of *S. calcitrans* (83%) (Fig. 16B). The molecular mass corresponds to pro-Gmtrypsin is 24 kDa with PI 7.08. As expected, rGmtrypsin-GST fusion protein expressed by bacterial expression system was approximately 50 kDa, which is corresponds to sum of 24 kDa rGmtrypsin and 26 kDa GST (Fig. 16C).

Proteolytic activity of rGmtrypsin: The pH, temperature, and dose dependent profiles were determined by using BApNA and Z-Phe-Arg-MCA substrate. The reaction mixture was incubated with different various buffers ranging from pH 5-9 at various temperatures between 15 to 50 $^{\circ}$C. rGmtrypsin exhibited the highest BApNA hydrolysis at pH 8.0 at 25 $^{\circ}$C (Figure 17A and 17B). Dose dependency of activity was similar between rGmtrypsin and control enzyme, bovine trypsin (Figure 17C). The activity of rGmtrypsin toward Z-Phe-Arg-MCA showed the highest at pH 8.0 at 30°C (data not show).

Recombinant Gmtrypsin hydrolyzed bovine hemoglobin and bovine serum albumin. The proteolytic activity of rGmtrypsin with blood protein contents, bovine hemoglobin (Hb) and bovine serum albumin (BSA), was visualized by SDS-PAGE. Results showed that rGmtrypsin (10 μg) completely digested BSA (100 μg) after 4 h of incubation as shown in the control enzyme, bovine trypsin (Figure 18A). In contrast, rGmtrypsin showed no hydrolytic activity upon Hb (Figure 18B). The activity of rGmtrypsin was completely inhibited by antipain which is a specific inhibitor of trypsin.

Kinetic properties of rGmtrypsin: The kinetic parameters of rGmtrypsin were analyzed by fitting the velocity ($V_{0)}$ data from various concentration of BApNA (20-100

μM) to the Michealis-Menten equation and statistically analyzed by Student's *t*-test using GraphPad Prism software. The results were compared with activity of bovine trypsin (BvT) incubated at the same condition (pH 8.0 and 25 ◦C). As shown in table 6, the apparent K_M value of rGmtrypsin (21.43 μM) was approximately 1.5 times higher than that of BvT (14.14μM). Turn over number (k_{cat}) of rGmtrypsin and BvT are similar (1.47 and 1.36 molecules per second, respectively). Consequently, the catalytic efficiency (k_{cat}/K_M) of rGmtrypsin (63.5 $mM^{-1}s^{-}1$) was lower than BvT (104 $mM^{-1}s^{-1}$) about 0.6 times (*p* =0. 0830, Cl 95%).

rGmtrypsin effects on viability of trypanosome. Co-culture of trypanosome with rGmtrypsin resulted in a reduction of the number of parasites. At 8 h of incubation, the number of viable BSFs and PCFs were disappeared and 31.25 % reduction from the original number of trypanosomes, respectively (Figure 19A). Abnormal shrinky BSFs were observed by thin smear at 4 h of treatment with rGmtrypsin. While the morphology of PCFs did not show any significant change from the control (Figure 19B). To assess the involvement of rGmtrypsin in transformation from PCF to EMF, rGmtrypsin was incubated with PCF trypanosomes in PSG buffer, and parasites were observed over time. However, the transformation was not observed within 24 h of incubation (Figure 20).

3.4 Discussion

Proteolysis (coagulation and melanization) are generally accepted as part of the immune reaction that eliminate the parasites from the insect (Aksoy, 2003). Among various digestive enzymes, trypsin is a principal alkaline endopeptidase that is previously identified in the midgut of the tsetse and its activity increases throughout the digestion cycle (Yan et al., 2001). Furthermore, inhibition of trypsin or trypsin-like enzymes by D-glucosamine increases the susceptibility of tsetse flies to trypanosomes infection (Osir et al., 1993). Interestingly, trypsin stimulated the transformation of trypanosome *in vitro* (Yabu and Takayanagi, 1988). The expression of trypsin-like serine protease gene (SERF4) in *Anopheles dirus* mosquito changed according to *P. falciparum* infection suggested its role in parasite-specific innate immunity (Sriwichai, 2012). However, the molecular study on the function of trypsin in tsetse has not been well established. This chapter intended to characterize trypsin in the midgut of tsetse and studied on its biological importance on tsetse-trypanosome interaction.

The trypsin proenzyme was cloned from the midgut of tsetse flies consisting of 227 amino acids with a predicted molecular mass of 24.4 kDa, and named as Gmtrypsin. Gmtrypsin contained propeptide and mature peptide with cleavage site at Arg8-Iso9 (corresponding to number in GmT: figure 15). Eight amino acid residues were cleaved in order to produce mature enzyme with 219 amino acid residues (23.5 kDa). Six-conserved cysteine residues present in the sequence formed the three-disulfide bonds which typically found in invertebrate serine proteases. The presence of conserved Asp-Ser-Gly (DSG) was important for arginine (Arg) substrate specificity at P1 position.

Mature Gmtrypsin displayed maximal activity upon BApNA and Z-Phe-Arg-MCA at alkaline pH 8.0 and at 25-30 ˚C. This optimal pH was similar to the condition of posterior midgut of tsetse (pH 7.9+/-0.4) (Liniger et al., 2003a) while optimal temperature seemed to be the same condition for tsetse habitat. These results assumed that Gmtrypsin might active in posterior portion of midgut which is the main part of blood meal digestion, nutrient absorption and killing of trypanosomes. In general serine proteases (trypsin and chymotrypsin) of hematophagous insects in dipteran group were activated at alkaline condition of posterior midgut and served as a main digestive enzyme in digestion process (Chougule et al., 2005). Blood protein hydrolysis revealed that rGmtrypsin had high affinity for only BSA but not for Hb whereas bovine trypsin (BvT) which was used as a control enzyme showed high proteolytic activity to both Hb and BSA. The kinetic values evaluated by Michaelis-Menten equation indicated that rGmtrypsin had a lower affinity and catalytic efficiency to BApNA as compared to BvT, but they showed the same speed of BApNA hydrolysis. Catalytic efficiency of rGmtrypsin was 0.6 times lower than BvT. However, no significant difference was observed between hydrolytic acitivity of BvT and rGmtrypsin toward BApNA (p =0. 0830, Cl 95%).

Co-cultivation of trypanosomes with rGmtrypsin resulted in decreasing BSFs and PCFs viability. rGmtrypsin showed significant effect to the viability of BSF (p =0.003 by 2-WAY ANOVA), while the effect on PCF was not significant (p=0.274 by 2-WAY ANOVA). This decreased viability of BSFs would have an impact on prevalence of parasite infection to tsetse. It was reported that digestive enzymes of sand fly could kill the early transforming amastigote *Leishmania* (Pimenta et al., 1997). From the previous study, trypsin cleaved the major surface glycoproteins of PCFs of *T. brucei*, namely EP and

GPEET procyclin (Liniger et al., 2003c). In this chapter only BSFs cell death was observed. This might be the results that rGmtrypsin has no significant effect to PCFs. Moreover, transformation of PCF into EMF was evaluated by differentiating the position of the nucleus and kinetoplast as well as distance between both organelles. However, the transformation was not observed. Thus Gmtrypsin was not involved in this process. In conclusion, Gmtrypsin is involved in two important processes of tsetse, (1) digestion of protein content in blood meal (BSA) and (2) elimination of early stage of trypanosomes that enter the midgut (BSF). Therefore, trypsin may have an advantage for using to interfere the development of trypanosome infections in tsetse.

3.5 Summary

cDNA (681 bp) encoding the trypsin-like serine protease precursor (Accession no. AAF91346.1) was cloned from the midgut of tsetse. The protein (Gmtrypsin) consists of 227 amino acids with a predicted molecular mass of 24.4 kDa and a PI value of pH 7.08. An active recombinant protein (rGmtrypsin) was expressed in *E. coli* exhibited the highest proteolytic activity at alkaline pH 8.0 and at 25-30 $^{\circ}$C on a dose dependent manner. It had a high affinity to BSA, cleaved Z-Phe-Arg-MCA and BApNA while showing preference for arginine at P1 residues. Proteolytic analysis using BApNA revealed catalytic efficiency (k_{cat}/K_M) of Gmtrypsin 0.6 times lower than that of bovine trypsin (BvT). Moreover, *in vitro* co-cultivated trypsnosomes (*T. congolense*) with rGmtrypsin resulted in a significant reduction of the number of BSFs, whereas PCFs are relatively resistant against rGmtrypsin. The transformation of PCFs into EMFs was not observed within 24 h. The results from this chapter indicated that Gmtrypsin is involved in the blood meal digestion and might therefore be useful in elimination of the trypanosomes in the midgut of tsetse. However, localization of Gmtrypsin in the midgut of tsetse should be studied to fully understand the active area of this enzyme and how involvement in tsetse-trypanosome interaction.

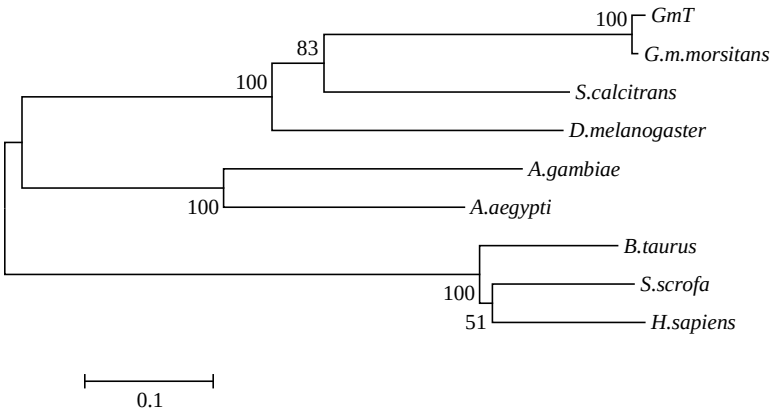
A

```
                                                  ⬇
GmT                 --------------------HN---DHTAGRIVNGVETTIEKRPYQVSLQSAVSGNHFCG
G.m.m_AAF91346.1    -MFRYFLVFVALSTVVAGLLHN---DHTAGRIVNGVETTIEKRPYQVSLQSAVSGSHFCG     56
S.c_AAC39131.1      -MLRFVILFALVSTSLAGVSRN---DFYG-RIVNGVATTIEEHPYQVSLQG-LSGSHFCG
D.m_ABC86406.1      RMNRLLLSVVALVALAASCHGNPGLDFPFGRIVNGEDVDIENYPYQVSVQT-TKGSHFCG
B.t_CAA38513.1      --MHPLLILAFVGAAVAFPSDD------DDKIVGGYTCAENSVPYQVSLNA---GYHFCG
                                         :        :**.*     :. *****::    * ****
                                  2
GmT                 GSIISEDIIVTAAHCVSGSNPSQLKVRLGSTYNNEGGIVVGVKALKYHEKFNNDVLWHDI
G.m.m_AAF91346.1    GSIISEDIIVTAAHCVSGSNPSQLKVRLGSTYNNEGGIVVGVKALKYHEKFNNDVLWHDI     116
S.c_AAC39131.1      GSIISEDIVVTAAHCMQSHSASEFKVRLGSTQYNTGGELVEVKAFKFHENYNSGTMKNDV
D.m_ABC86406.1      GSLIDSETVLTAAHCMQSYAASELQVRVGSTSRSSGGEVVTVRAFKYHEGYNSKLMINDV
B.t_CAA38513.1      GSLINDQWVVSAAHCYQYHIQVRLGEYN-IDVLEGGEQFIDASKIIRHPKYSSWTLDNDI
                    **:*..: :::****       .:         . *  .: .  :  *  :..  : :*:

GmT                 AVLKLEKPVKQSSTIRYIEMAKKVPKTGTPAVVSGWGTKCFLTC-PLSPVLMEVEVTFLQ
G.m.m_AAF91346.1    AVLKLEKPVKQSSTIRYIEMAKKVPKTGTPAVVSGWGTKCFLTC-PLSPVLMEVEVTFLE     175
S.c_AAC39131.1      AVIKLARPVKESATIRFVKLADKTPATGTPAVVTGWGTTCFMACNTLPKTLQKVVVDIVD
D.m_ABC86406.1      AIIKLSSPVRQTSKIRAIELADSEAVSGTNAVVSGWGTTCFLFC-SSPDTLQKVEVDLLH
B.t_CAA38513.1      LLIKLSTPAVINARVSTLLLPSACASAGTECLISGWG-NTLSSGVNYPDLLQCLVAPLLS
                     ::**  *.  .: :  : :..  . :** .:::*** . :      .  *  : . ::
                        3                 4        D 5                   S G  6
GmT                 REDCASKTYLYG-DKIKETMVCG--YATAKDSCQGDSGGPFVADGKLVGVVSWGQGCAMD
G.m.m_AAF91346.1    REDCASKTYLYG-DKIKETMVCG--YATAKDSCQGDSGGPFVADGKLVGVVSWGQGCAMD     232
S.c_AAC39131.1      EKTCASSEYKYG-SKIKPTMVCA--YAEDKDACQGDSGGPLVAGGKLVGVVSWGKGCALP
D.m_ABC86406.1      YKDCAADTYNYGSDSILETMVCA--TGEKKDACQGDSGGPLVADNKLVGVVSWGSGCAWT
B.t_CAA38513.1      HADCEAS---YP-GQITNNMICAGFLEGGKDSCQGDSGGPVACNGQLQGIVSWGYGCAQK
                       * :.   *  ..*  .*:*.      **:********.....:* *:**** ***

GmT                 GYPGVYSDVAALRD----------
G.m.m_AAF91346.1    GYPGVYSDVAALRDWVLENAQKL-       255
S.c_AAC39131.1      AIPGVYADVPSLRTWIEKTAKEL-
D.m_ABC86406.1      GYPGVYADVASLRSWIVDTTDSL-
B.t_CAA38513.1      GKPGVYTKVCNYVDWIQETIAANS
                    . ****:.*
```

B



C

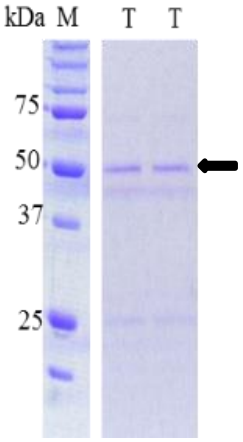


Figure 16. Characterization of Gmtrypsin. Panel A: Alignment of deduced amino acid sequence Gmtrypsin: (GmT) with G.m.m._AAF91346.1, and other insect trypsins namely *S. calcitrans* (S.c_AAC39131.1), *D. melanogaster* (D.m_ABC86406.1), and *B. taurus* (B.t_CAA38513.1). The black headed arrow indicates the cleavage site, His-Asp-Ser which forming the catalytic triad are shaded by yellow color, six-conserved cysteine residues are in gray numbering 1-6, and conserved substrate specificities are indicated by DSG. Panel B shows the phylogenetic tree by using a neighbor joining analysis based on a Clustal X 2.0.11 alignment software. Panel C: The molecular mass of rGmtrypsin (T) showed the band of 50 kDa (arrow).

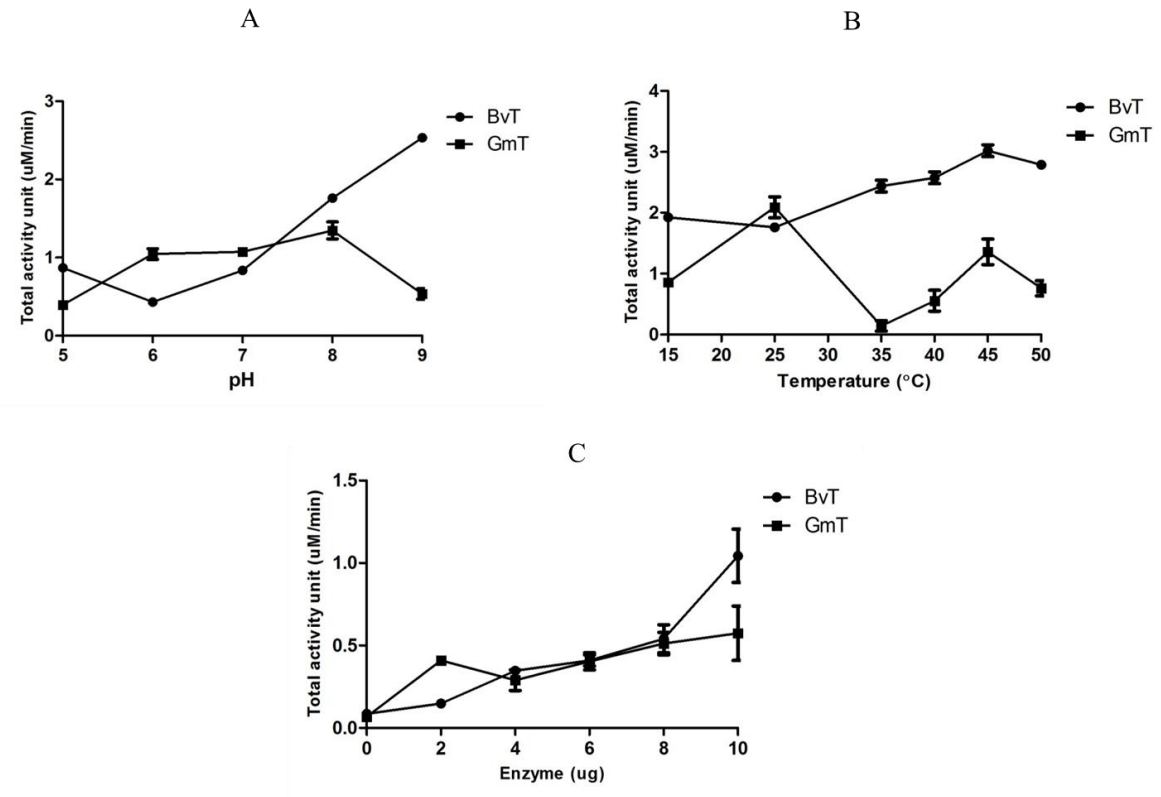


Figure 17. Enzyme characterization of rGmtrypsin (rGmT) using BApNA substrate. A: Effect of pH on proteolytic activity of rGmtrypsin. B: The effect of reaction temperature on proteolytic activity at pH 8.0. C: Dose dependency of rGmtrypsin. Total activity unit was defined as the amount of releasing product pNA (uM) per minute.

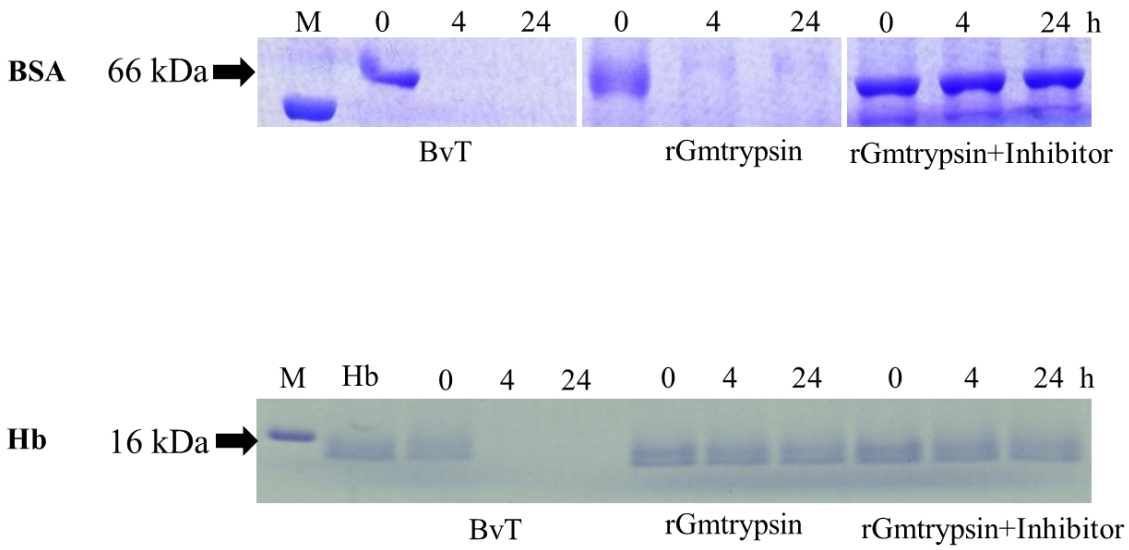


Figure 18. The proteolytic activity of rGmtrypsin toward bovine hemoglobin (Hb) and bovine serum albumin (BSA). The reactions were done at pH 8.0 at 25 ˚C then collected the hydrolytes at 0, 4, 24 h. Ten μg of rGmtrypsin completely degraded BSA (100 μg) at 4 h of incubation similar to the proteolytic activity of bovine trypsin (BvT). In contrast to the proteolytic acitivity upon BSA, rGmT did not digest the monomer Hb. The activity of rGmT was inhibited by trypsin inhibitor, anitipain.

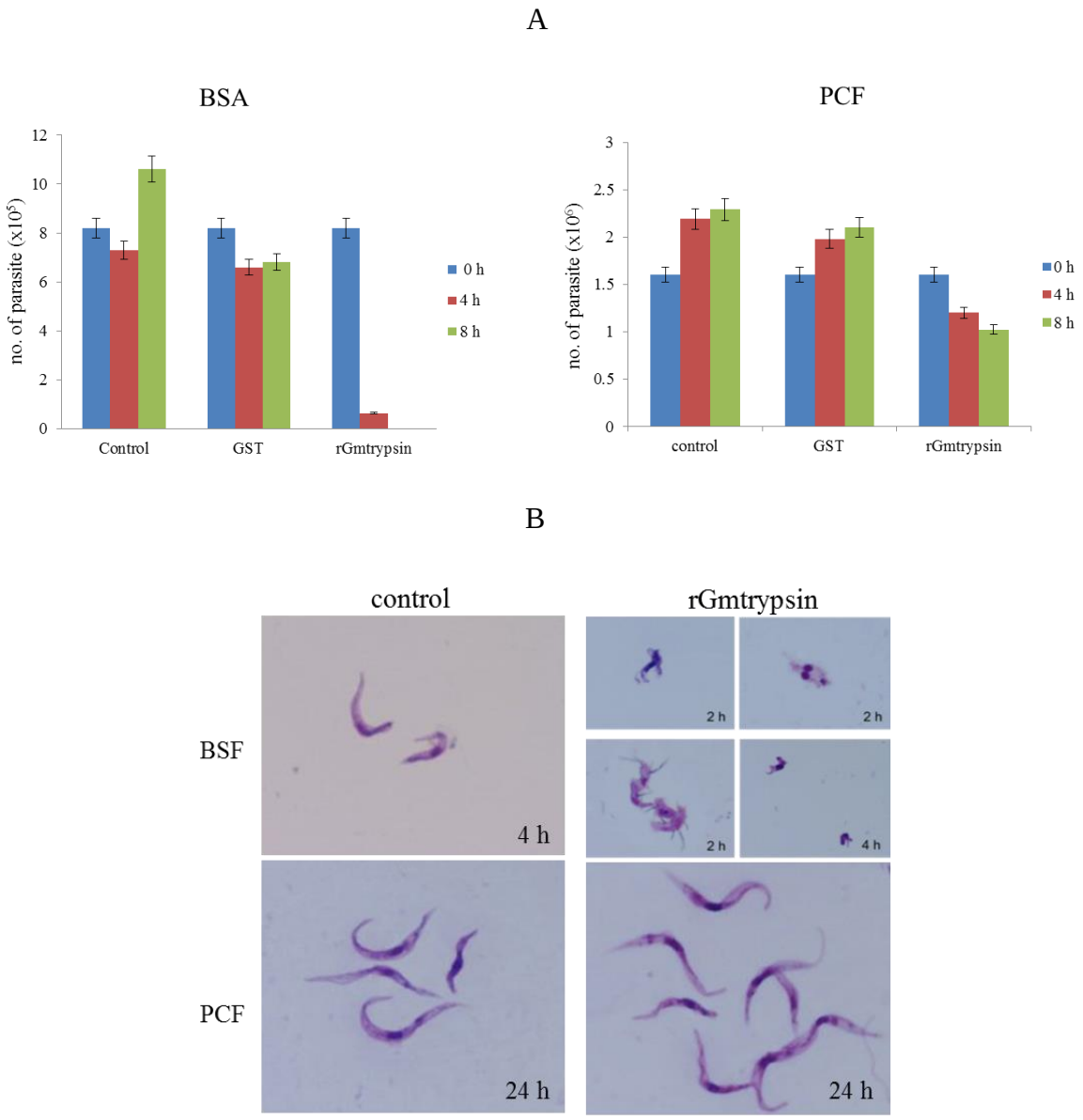


Figure 19. The effect of rGmtrypsin to *T. congolense*. A: rGmtrypsin was co-cultured with parasites resulted in a significant reduction the viable of BSF. B: Light microscope photograph of BSF and PCF after treated with rGmtrypsin at different time points, the aberrant phenotype of BSFs were observed at 2 to 4 h of incubation with rGmT.

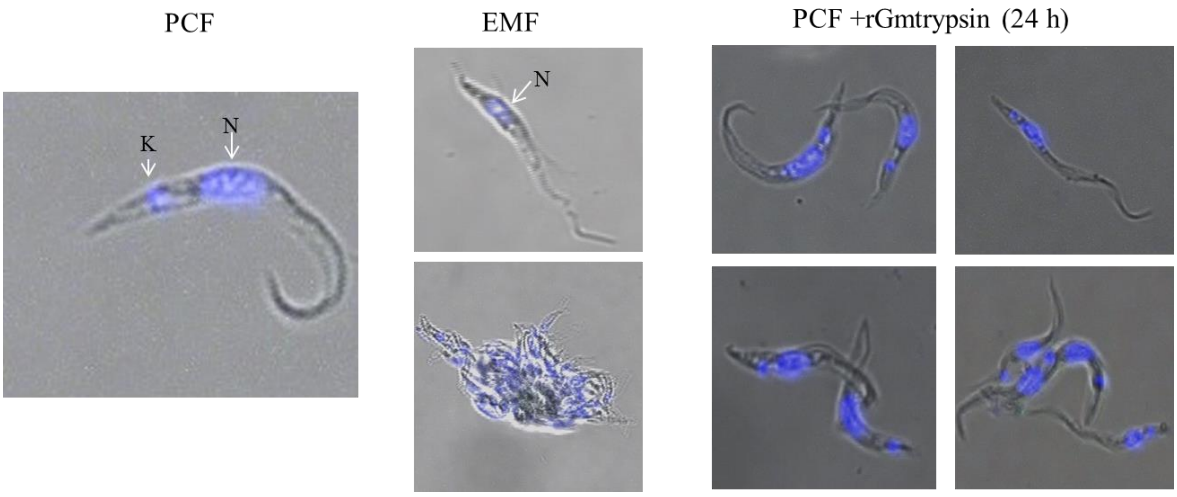


Figure 20. The effect of rGmtrypsin on transformation of PCF into EMF. To assess the involvement of rGmtrypsin in the transformation of PCF into EMF, rGmtrypsin was co-cultured with PCF in PSG buffer pH 8.0 then incubated at 27 ˚C and parasites were observed over 24 h. The transformation from PCF into EMF was confirmed by distinguish the location of kinetoplast using confocal laser scanning microscopy.

Table 6: The kinetic constants for BApNA hydrolysis of rGmtrypsin

Enzyme	Vmax (uM/min)	K_M (uM)	K_{cat} (s^{-1})	K_{cat}/K_M ($mM^{-1} s^{-1}$)
BvT	1.35	14.1	1.47	104.0
rGmtrypsin	0.89	21.4	1.36	63.5

General Discussion

Tsetse-trypanosome interaction is a complex phenomenon that is mediated by various factors from both parasites and vector. In order to describe this interaction, it is convenient to devide the trypanosome life cycle into two stages, (1) establisment into PCF in the midgut of tsetse, (2) maturation: trypanosome must leave the midgut then differentiate into epimastigote form followed by infective metacyclic form in the salivary gland or probossis of tsetse. When tsetse flies are challenged by pathogen (trypanosomes), inate immune response has been implicated in preventing or limiting the establishment of gut infection. Inside the midgut of tsetse, trypanosomes are also interacted with other fates including fluctuating temperature, lack of nutrients as well as proteases.

In general, tsetse flies produce proteases mainly in posterior midgut to play a main role in blood meal digestion that produces necessary nutrients for several physiological processes. The proteases of tsetse may provide triggers or signals that contribute to differentiation of trypanosome *in vivo.* In addtion, they are considered as one of the crucial barriers against trypanosomes during their establishment within the midgut of tsetse that can severely impact the outcome of infection. Slender BSFs are rapidly killed by proteases in digestive lumen of tsetse, while stumpy forms survive and differentiate into procyclic form (Sbicego et al., 1999). Protease treatment resulted in cell dealth of trypanosomes then they were eliminated in the initial phase of development, remaining stumpy BSFs transformed into procyclic forms (PCFs) that expressed a new surface coat (Gibson and Bailey, 2003). To establish the infection, PCFs must penetrate the peritrophic matrix (PM) and migrate to attach the salivary gland epithelium to undergo differentiation into the

mammal infective metacyclic forms (MCFs). The PM is a chitin and glycoprotein layer that lined the invertebrate midgut, act as a physical barrier protecting the midgut epithelium from abrasive food particles, digestive enzymes, and pathogens. The PM defines an endoperitrophic space containing the ingested food and ectoperitrophic space which is the space between the PM and midgut epithelium. Major functions of the PM are compartmentalization of digestive processes which allows efficient nutrient acquisition and reuse of hydrolytic enzymes. The PM acts as a barrier against pathogens, it is not fully understood how trypanosomes cross tsetse PM. It had been reported that *T. brucei* secreted proteases to degrade the extracellular matrix (ECM) of human such as collagen, fibronectin, and laninin to migrate into the tissue (Huet et al., 1992). The ECM of human has function similar to insect PM. Similarly, ookinete of *Plasmodium* secretes a chitinase in order to cross the PM of mosquito (Huber et al., 1991). From those previous reports, it is implied that proteases from both insect vector and parasite are necessary for infection. Therefore, it is important to characterize the function of proteases from tsetse in order to clarify tsetse-trypanosome interaction.

In tsetse, at least seven digestive enzymes have been characterized from different organs namely, midgut, fat body and salivary gland (Attardo et al., 2006; Yan et al., 2001, 2002a). Nevertheless roles of those enzymes on blood meal digestion and effects to trypanosome have not been examined yet. Therefore this study has expresed three digestive enzymes of two protease groups, cysteine proteases (GmcathB and GmcathL) and serine protease (Gmtrypsin) from midgut of tsetse. Further characterization of enzymatic activities and the role on tsetse-trypanosome interaction have done by using recombinant protein of these proteases. In relation to other insects, GmcathB and GmcathL showed certain degree

of similarity to proteases express in *Sarcophaga peregrina* (fresh fly), while Gmtrypsin is homologous to trypsin of *Stomoxys calcitrans* (stable fly). *Sarcophaga* cathepsin B is participated in decomposition of larval fat body during metamorphosis (Takahashi et al., 1993). *Sarcophaga* cathepsin L is involved in differentiation of imaginal discs, through proteolysis of components of basement membranes (Homma et al., 1994). It is reported that cathepsin B and cathepsin L of lepidopteran insect are involved in histolysis of silk gland during metamorphosis (Shiba et al., 2001) and are related to the transformation of prohemocytes in fat body, then induce hemocyte apoptosis for further tissue degradation (Zhai and Zhao, 2012). Cathepsin B and L of blood feeding insect are also involved in embryonic vitellin degradation (Cho et al., 1999). Cathepsin L has insecticidal towards larvae of the tomato moth, *Lacanobia oleracera* (Philip et al., 2007). In stable fly, expression of trypsin occurred predominantly in the opaque region which responds for secretion of digestive enzyme through the digestion cycle (Lehane et al., 1998). This enzyme is produced as zymogens and stored within the cell in zymogen granules (Moffatt and Lehane, 1990), that mainly immunolocalized in ectoperotrophic space in the midgut lumen of fed and blood-fed stable fly (Jordao et al., 1996).

All enzymes in this study are produced as zymogen (inactive proenzyme) which need to activate in order to an active mature enzymes. Immature protein of GmcathB and GmcathL are encoded at N-terminal, which cleave off to activate the enzyme, implying the N-terminal proregion act as autoinhibitor. In addition to its role in autoinhibition, proregion of cathepsin L is a recognition site for mannose phosphorylation (Cuozzo et al., 1995) and is neccessary for proper folding of enzyme (Tao et al., 1994). The processing of zymogen of cathepsin B and cathepsin L into a catalytically active enzyme usually occur in

the lysosome by other proteases or by self-catalysis incertain conditions such as low pH or presence of glycosaminoglycan. The presence of potential N-glycosylation in GmcathB and GmcathL suggests that they are glycosylated and transported to the lysosome via the mannose-6-phosphate receptor. Gmtrypsin contained propeptide and mature petide but not included a putative signal peptide (pro-region) because of the primer designing. Propeptide of Gmtrypsin was cleaved off then amino acid brough up to an active mature peptide.

Enzyme analysis revealed that a recombinant GmcathB (rGmcathB) and a recombinant GmcathL (rGmcathL) expressed in yeast (*P. pastoris*) efficiently hydrolyzed the synthetic substrates specific for cathepsin B and L, respectively. While, active soluble recombinant Gmtrypsin (rGmtrypsin) was successfully expressed by using *E. coli* expression system. rGmcathB and rGmcathL exhibited a high activity againt synthetic specific substrates over a broad range of pH between 4.0 to 8.0 and temperature values between 25 to 37 ◦C. However, both recombinant proteases showed the highest activity toward fluorogenic casein and systhetic specific substrates at pH 4.0, 30 ◦C (rGmcathB) and at pH 6.0, 30 ◦C (rGmcathL). Enzymatic activity of rGmtrypsin against BApBA also exhibited at broad range of pH (6.0-8.0), wheras the temperarure was specified at 25 ◦C. The result that all recombinant proteins showed activity over wide range of pH might because of the pH condition in posterior midgut of tsetse was neutral to alkaline (pH 8.0-10.0). There are probably have some adjustments of ingested blood meal, protease and condition in the midgut lumen. However, the optimal temperature for all recombinant proteins in this study were similar for tsetse habitat (25 ◦C). Although, Hb and BSA degradation assay indicated GmcathB, GmcathL and Gmtrypsin are involved in blood protein digestion, they showed different affinity to the substrates. rGmcathB has a high

affinity for Hb more than for BSA, while rGmcathL and rGmtrypsin showed higher affinity for BSA than Hb. Control enzyme for GmcathB and GmcathL, papain exhibited higher affinity for BSA than for Hb. Bovine trypsin which used as control for Gmtrypsin showed identical affinity for Hb and BSA. In this study, only rGmcathL displayed a affinity property toward both substrates similar to the control enzyme (Papain). Optimal pH of rGmtrypsin was corresponds to pH of posterior midgut, suggesting Gmtrypsin may secreted into the gut lumen of tsetse fly and the proteolytic activity associated with gut lumen condition. Trypsin is the most abundant digestive protease express in the midgut of hematophagous insect upon blood feeding which similar to the herbivore insects (lepidopteran larval), it's midgut harbors serine proteases contribute to 95 % of total digestive activity (Srinivasan et al., 2006).

Moreover, all the recombinant proteases killed BSF trypanosomes *in vitro* as shown in previous reports. rGmcathB and rGmtrypsin decreased some number of viable PCF by 33.3 % and 36.3 %, respectively. This suggests that PCF trypanosomes which already change the variable surface glycoprotein (VSG) into procyclin are more resistant to digestive enzyme in tsetse midgut than BSF. The recombinant proteases is this study are not involved in transformation of PCF into EMF under *in vitro* condition mimicking insect habitat. Therefore, proteases in this study play a role in blood meal digestion and elimination of trypanosome at early stage of infection.

General Summary

Tsetse flies (*Glossina morsitans morsitans*) are strict blood feeders and are capable of trypanosome transmission. In particular, *Trypanosoma brucei* and *T. congolense* are considered as a major cause of human African trypanosomosis (HAT) or sleeping sickness and animal African trypanosomosis (AAT) or nagana disease. Digestive enzymes in tsetse fly midgut act as one of the crucial barriers against trypanosomes during their establishment within the midgut of tsetse. However, some additional signals of digestive enzymes are thought to modulate the development of African trypanosome. The objective of this study was to characterize the enzymatic activity and to investigate the role on tsetse-trypanosome interaction of digestive enzymes of tsetse. This study has expressed and characterized three digestive enzymes from the midgut of tsetse namely cathepsin B (GmcathB) and cathepsin L (GmcathL) cysteine protease and trypsin-like serine protease (Gmtrypsin). Although, these enzymes have been identified as tsetse midgut digestive enzymes for a while, their enzyme characteristic have not been analyzed yet. Active recombinant protein of GmcathB and GmcathL were successfully expressed in yeast (*P. pastoris*) expression system. While, active recombinant protein of Gmtrypsin was obtained from bacterial (*E. coli*) expression system. Activity profiling with specific substrate and fluorogenic casein substrate indicated that GmcathB and GmcathL are acidic proteases, whereas Gmtrypsin is alkaline protease. All three recombinant proteins displayed their activity in wide range of temperature, which seem to be in range of the temperature of tsetse habitat (25 $^{\circ}$C). They have a catalytic efficiency for their specific substrates two times lower than the control enzymes. As these recombinant proteases have shown the ability in blood protein (Hb and BSA) digestion and

elimination of BSF trypanosomes. This suggest that GmcathB, GmcathL and Gmtrypsin from the midgut are involved in blood meal digestion and elimination of trypanosome at early phase infection. In addtion, rGmcathB and rGmtrypsin appeared to reduce PCF viability, although the percentage was not as high as reduction of BSF. Assuming, GmcathB and Gmtrypsin may serve as valuable tool for probing the physiology of different developmental forms. This study has proved that *P. pastoris* expression system is suitable for expressing enzymatically active recombinant protein from cysteine proteases. However, further investigation of digestive enzymes localization and involvement of digestive enzymes in transformation of insect stage trypsanosme into infective metacyclic from are needed to carify the tsetse-trypanosome interaction.

和文要約

Trypanosoma brucei と *T. congolense* はそれぞれヒトのアフリカトリパノソーマ病（アフリカ睡眠病）および動物のアフリカトリパノソーマ病（ナガナ病）の主要病原体として重要である。これらのトリパノソーマはツェツェバエ（*Glossina morsitans morsitans*）によってのみ媒介される。ツェツェバエ中腸の消化酵素は感染動物の血液と共に中腸内に取り込まれたトリパノソーマがツェツェバエへの感染を成立させる際に主要な阻害因子として機能するが、それとは逆に消化酵素による刺激がトリパノソーマの中腸内での生存・増殖・分化を調節しているという報告もある。そこで本研究では中腸内でのツェツェバエ vs トリパノソーマ相互作用に直接関与していることが予想される中腸内消化酵素について、その詳細な酵素活性解析とツェツェバエ vs トリパノソーマ相互作用における役割を明らかにすることを目的とした。本研究で解析した中腸由来消化酵素は、既に遺伝子情報が明らかとなっていたシステインプロテアーゼファミリーのカテプシンＢ（GmcathB）とカテプシンＬ（GmcathL）、ならびにトリプシン様セリンプロテアーゼ（Gmtrypsin）の３種類であった。これらの消化酵素はいずれもツェツェバエ中腸からクローニングされたが、未だその酵素活性については詳細に解析されていない。そこで活性型消化酵素の作製をめざして種々の組換え蛋白質発現系を検討した結果、GmcathB と GmcathL は酵母（*P. pastoris*）発現系、Gmtrypsin は大腸菌（*E. coli*）発現系を用いたことで酵素活性を有する組換え蛋白質を発現することができた。次に、それぞれの消化酵素に特異的な蛍光基質、広域基質であ

る蛍光標識カゼイン、血液蛋白のウシヘモグロビン（Hb）ならびにウシ血清アルブミン（BSA）を用いて酵素活性を解析した。その結果、GmcathB と GmcathL は酸性領域、Gmtrypsin は塩基性領域で至適活性を呈し、いずれも広範囲の温度域で消化活性を示した。これらの特性は各酵素の中腸内における局在やツェツェバエの生息環境（２５℃付近）における酵素活性維持と関係していることが示唆された。また、いずれの消化酵素も対照として用いた消化酵素（パパインまたはトリプシン）の特異基質に対する活性を１００％とした場合、５０％程度の活性であったが、血液蛋白質（Hb および BSA）の両方またはどちらか一方を消化し、血流型トリパノソーマ（BSF）に対する傷害性を示したことから中腸内での血液餌消化と感染初期におけるトリパノソーマ殺滅に関与していることが示唆された。加えて、rGmcathB と rGmtrypsin ではプロサイクリック型トリパノソーマ（PCF）に対する弱い傷害性も認められたことから、これらの消化酵素はトリパノソーマの発育ステージ、特に BSF および PCF とツェツェバエとの相互作用を研究する上で鍵となる分子であると考えられる。本研究では酵母発現系が活性を有するツェツェバエ消化酵素の発現に適していることも明らかとなった。以上をまとめると、本研究によってこれまで詳細に解析されていなかったツェツェバエ中腸由来消化酵素の酵素活性特性が明らかになり、ツェツェバエ vs トリパノソーマ相互作用への関与と血液餌消化のメカニズム解析を行っていくための基礎となる知見を得ることができた。しかしながら将来、本研究で解析した消化酵素がトリパノソーマの発育

分化に及ぼす影響や、臓器・組織における局在と血液餌消化における役割をさらに詳細に明らかにしていく必要がある。

Acknowledgements

I would like to acknowledge the financial and academic support of Obihiro University of Agriculture and Veterinary Medicine (OUAVM) for my doctoral study. This study was financially supported by a grant from the Graduate School of Obihiro University of Agriculture and Veterinary Medicine, the Global COE Program, Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), and Japan and the Japan Society for the Promotion of Science (JSPS).

I am very grateful to have learned and done my research work at the Research Unit for Vector Biology, National Research Center for Protozoan Diseases (NRCPD), OUAVM.

I would like to give my sincerest appreciation to my main supervisor, Prof. Noboru Inoue for giving me the opportunity to study doctoral course at the OUAVM. This study will not be possible without his endless guidance, supervision and support.

I am truly indebted to my co-supervisors, Prof. Shin-ichiro Kawazu, Prof. Xuenan Xuan, Prof. Naoaki Yokoyama and Assoc. Prof. Yoshifumi Nishikawa, for their useful advices and comments to improve my research work; to all the other professors at NRCPD for their special reviews of my progress reports and beneficial advices during graduate student seminars at the NRCPD.

I would like to also extend my heartfelt gratitude to my co-labmates, Mrs. Mo Zhou, Mr. Keisuke Suganuma, Ms. Nguyen Thu Thuy, Ms. Shino Yamasaki and Mr. Kenedy Miyoro Mochabo for helping and giving me some insights in the experiments done in this study; to other staffs and students, both current and past, of the Research Unit for

Advanced Preventive Medicine, Research Unit for Vector Biology and NRCPD for their assistance, friendship and companionship throughout the period of my Ph.D. study.

Finally, I am deeply grateful for my family and friends in Thailand for their support, prayers and love; and most of all to God Almighty for guiding me always.

References

Abt, M., Rivers, D.B., 2007. Characterization of phenoloxidase activity in venom from the ectoparasitoid *Nasonia vitripennis* (Walker) (Hymenoptera: Pteromalidae). J Invertebr Pathol 94, 108-118.

Abubakar, L.U., Bulimo, W.D., Mulaa, F.J., Osir, E.O., 2006. Molecular characterization of a tsetse fly midgut proteolytic lectin that mediates differentiation of African trypanosomes. Insect Biochem Mol Biol 36, 344-352.

Acosta-Serrano, A., Vassella, E., Liniger, M., Kunz Renggli, C., Brun, R., Roditi, I., Englund, P.T., 2001. The surface coat of procyclic *Trypanosoma brucei*: programmed expression and proteolytic cleavage of procyclin in the tsetse fly. Proc Natl Acad Sci U S A 98, 1513-1518.

Ahn, J.E., Guarino, L.A., Zhu-Salzman, K., 2007. Seven-up facilitates insect counter-defense by suppressing cathepsin B expression. Febs J 274, 2800-2814.

Aksoy, S., 2003. Control of tsetse flies and trypanosomes using molecular genetics. Vet Parasitol 115, 125-145.

Atkinson, H.J., Babbitt, P.C., Sajid, M., 2009. The global cysteine peptidase landscape in parasites. Trends Parasitol 25, 573-581.

Attardo, G.M., Strickler-Dinglasan, P., Perkin, S.A., Caler, E., Bonaldo, M.F., Soares, M.B., El-Sayeed, N., Aksoy, S., 2006. Analysis of fat body transcriptome from the adult tsetse fly, *Glossina morsitans morsitans*. Insect Mol Biol 15, 411-424.

Barillas-Mury, C.V., Noriega, F.G., Wells, M.A., 1995. Early trypsin activity is part of the signal transduction system that activates transcription of the late trypsin gene in the midgut of the mosquito, *Aedes aegypti*. Insect Biochem Mol Biol 25, 241-246.

Barrett, A.J., Kirschke, H., 1981. Cathepsin B, Cathepsin H, and cathepsin L. Methods Enzymol 80 Pt C, 535-561.

Beckham, S.A., Piedrafita, D., Phillips, C.I., Samarawickrema, N., Law, R.H., Smooker, P.M., Quinsey, N.S., Irving, J.A., Greenwood, D., Verhelst, S.H., Bogyo, M., Turk, B., Coetzer, T.H., Wijeyewickrema, L.C., Spithill, T.W., Pike, R.N., 2009. A major cathepsin B protease from the liver fluke *Fasciola hepatica* has atypical active site features and a potential role in the digestive tract of newly excysted juvenile parasites. Int J Biochem Cell Biol 41, 1601-1612.

Blanco-Labra, A., Martinez-Gallardo, N.A., Sandoval-Cardoso, L., Delano-Frier, J., 1996. Purification and characterization of a digestive cathepsin D proteinase isolated from *Tribolium castaneum* larvae (Herbst). Insect Biochem Mol Biol 26, 95-100.

Blomgran, R., Zheng, L., Stendahl, O., 2007. Cathepsin-cleaved Bid promotes apoptosis in human neutrophils via oxidative stress-induced lysosomal membrane permeabilization. J Leukoc Biol 81, 1213-1223.

Bütikofer, P., Vassella, E., Boschung, M., Renggli, C.K., Brun, R., Pearson, T.W., Roditi, I., 2002. Glycosylphosphatidylinositol-anchored surface molecules of *Trypanosoma congolense* insect forms are developmentally regulated in the tsetse fly. Mol Biochem Parasitol 119, 7-16.

Carnevali, O., Cionna, C., Tosti, L., Lubzens, E., Maradonna, F., 2006. Role of cathepsins in ovarian follicle growth and maturation. Gen Comp Endocrinol 146, 195-203.

Casu, R.E., Jarmey, J.M., Elvin, C.M., Eisemann, C.H., 1994. Isolation of a trypsin-like serine protease gene family from the sheep blowfly *Lucilia cuprina*. Insect Mol Biol 3, 159-170.

Cazzulo, J.J., 2002. Proteinases of *Trypanosoma cruzi*: patential targets for the chemotherapy of Changas desease. Curr Top Med Chem 2, 1261-1271.

Cesen, M.H., Pegan, K., A., S., Turk, B., 2012. Lysosomal pathways to cell death and their therapeutic application. Cell Res. 318, 1245-1251.

Cheeseman, M.T., Gooding, R.H., 1985. Proteolytic enzymes from tsetse flies, *Glossina morsitans* and *Glossina palpalis* (Diptera: Glossinidae). Insect Biochem 15, 677-680.

Cho, W.L., Tsao, S.M., Hays, A.R., Walter, R., Chen, J.S., Snigirevskaya, E.S., Raikhel, A.S., 1999. Mosquito cathepsin B-like protease involved in embryonic degradation of vitellin is produced as a latent extraovarian precursor. J Biol Chem 274, 13311-13321.

Choe, Y., Leonetti, F., Greenbaum, D.C., Lecaille, F., Bogyo, M., Bromme, D., Ellman, J.A., Craik, C.S., 2006. Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. J Biol Chem 281, 12824-12832.

Chougule, N.P., Giri, A.P., Sainani, M.N., Gupta, V.S., 2005. Gene expression patterns of *Helicoverpa armigera* gut proteases. Insect Biochem Mol Biol 35, 355-367.

Conus, S., Simon, H.U., 2008. Cathepsins: key modulators of cell death and inflammatory responses. Biochem Pharmacol 76, 1374-1382.

Cristofoletti, P.T., Ribeiro, A.F., Terra, W.R., 2005. The cathepsin L-like proteinases from the midgut of *Tenebrio molitor* larvae: Sequence, properties, immunocytochemical localization and function. Insect Biochem Mol Biol 35, 883-901.

Cuozzo, J.W., Tao, K., Wu, Q.L., Young, W., Sahagian, G.G., 1995. Lysine-based structure in the proregion of procathepsin L is the recognition site for mannose phosphorylation. J Biol Chem 270, 15611-15619.

Dvorak, J., Mashiyama, S.T., Braschi, S., Sajid, M., Knudsen, G.M., Hansell, E., Lim, K.C., Hsieh, I., Bahgat, M., Mackenzie, B., Medzihradszky, K.F., Babbitt, P.C., Caffrey, C.R., McKerrow, J.H., 2008. Differential use of protease families for invasion by schistosome cercariae. Biochimie 90, 345-358.

Gibson, W., Bailey, M., 2003. The development of *Trypanosoma brucei* within the tsetse fly midgut observed using green fluorescent trypanosomes. Kinetoplastid Biol Dis 2, 1.

Gooding, R.H., 1977. Digestive processes of haematophagous insects. XII. Secretion of trypsin and carboxypeptidase B gy *Glossina morsitans morsitans* Westwood (Diptera: Glossinidae). Can J Zool 55, 215-222.

Gorman, M.J., Andreeva, O.V., Paskewitz, S.M., 2000. Molecular characterization of five serine protease genes cloned from *Anopheles gambiae* hemolymph. Insect Biochem Mol Biol 30, 35-46.

Grab, D.J., Garcia-Garcia, J.C., Nikolskaia, O.V., Kim, Y.V., Brown, A., Pardo, C.A., Zhang, Y., Becker, K.G., Wilson, B.A., de, A.L.A.P., Scharfstein, J., Dumler, J.S., 2009. Protease activated receptor signaling is required for African trypanosome

traversal of human brain microvascular endothelial cells. PLoS Negl Trop Dis 3, e479.

Greenbaum, D.C., Arnold, W.D., Lu, F., Hayrapetian, L., Baruch, A., Krumrine, J., Toba, S., Chehade, K., Bromme, D., Kuntz, I.D., Bogyo, M., 2002. Small molecule affinity fingerprinting. A tool for enzyme family subclassification, target identification, and inhibitor design. Chem Biol 9, 1085-1094.

Gui, Z.Z., Lee, K.S., Kim, B.Y., Choi, Y.S., Wei, Y.D., Choo, Y.M., Kang, P.D., Yoon, H.J., Kim, I., Je, Y.H., Seo, S.J., Lee, S.M., Guo, X., Sohn, H.D., Jin, B.R., 2006. Functional role of aspartic proteinase cathepsin D in insect metamorphosis. BMC Dev Biol 6, 49.

Hansen, G., Heitmann, A., Witt, T., Li, H., Jiang, H., Shen, X., Heussler, V.T., Remnenberg, A., Hilgenfeld, R., 2011. Structural basis for the regulation of cysteine-protease activity by a new class of protease inhibitor in *Plasmodium.* Structure 19, 919-929.

Hao, Z., Kasumba, I., Lehane, M.J., Gibson, W.C., Kwon, J., Aksoy, S., 2001. Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis. Proc Natl Acad Sci U S A 98, 12648-12653.

Harvey, R., Ferrier, D. (2011). Lippincott's Illustrated Reviews: Biochemistry. 5th ed. China. Wolters Kluwer Health.

Homma, K., Kurata, S., Natori, S., 1994. Purification, characterization, and cDNA cloning of procathepsin L from the culture medium of NIH-Sape-4, an embryonic cell line

of *Sarcophaga peregrina* (flesh fly), and its involvement in the differentiation of imaginal discs. J Biol Chem 269, 15258-15264.

Homma, K., Natori, S., 1996. Identification of substrate proteins for cathepsin L that are selectively hydrolyzed during the differentiation of imaginal discs of *Sarcophaga peregrina*. Eur J Biochem 240, 443-447.

Hook, V., Yasothornsrikul, S., Greenbaum, D., Medzihradszky, K.F., Troutner, K., Toneff, T., Bundey, R., Logrinova, A., Reinheckel, T., Peters, C., Bogyo, M., 2004. Cathepsin L and Arg/Lys aminopeptidase: a distinct prohormone processing pathway for the biosynthesis of peptide neurotransmitters and hormones. Biol Chem 385, 473-480.

Huber, M., Cabib, E., Miller, L.H., 1991. Malaria parasite chitinase and penetration of the mosquito peritrophic membrane. Proc Natl Acad Sci U S A 88, 2807-2810.

Huet, G., Richet, C., Demeyer, D., Bisiau, H., Soudan, B., Tetaert, D., Han, K.K., Degand, P., 1992. Characterization of different proteolytic activities in *Trypanosoma brucei brucei*. Biochim Biophys Acta 1138, 213-221.

Imbuga, M.O., Osir, E.O., Labongo, V.L., Darji, N., Otieno, L.H., 1992. Studies on tsetse midgut factors that induce differentiation of bloodstream *Trypanosoma brucei brucei* in vitro. Parasitol Res 78, 10-15.

Ishidoh, K., Kominami, E., 1995. Procathepsin L degrades extracellular matrix proteins in the presence of glycosaminoglycans in vitro. Biochem Biophys Res Commun 217, 624-631.

Iwanaga, S., Yamasaki, N., Kimura, M., 1998. Chymotrypsin inhibitor from *Erythrina variegata* seeds: involvement of amino acid residues within the primary binding loop in potent inhibitory activity toward chymotrypsin. J Biochem 124, 663-669.

Jones, G., Venkataraman, V., Manczak, M., 1993. Transcriptional regulation of an unusual trypsin-related protein expressed during insect metamorphosis. Insect Biochem Mol Biol 23, 825-829.

Jordao, B.P., Lehane, M.J., Terra, W.R., Ribeiro, A.F., Ferreira, C., 1996. An immunocytochemical investigation of trypsin secretion in the midgut of the stablefly, *Stomoxys calcitrans*. Insect Biochem Mol Biol 26, 445-453.

Kageyama, T., Takahashi, S.Y., 1990. Purification and characterization of a cysteine proteinase from silkworm eggs. Eur J Biochem 193, 203-210.

Kaiser-Alexnat, R., 2009. Protease activities in the midgut of Western corn rootworm (*Diabrotica virgifera virgifera LeConte*). J Invertebr Pathol 100, 169-174.

Katunuma, N., 1989. Mechanisms and regulation of lysosomal proteolysis. Revis Biol Celular 20, 35-61.

Kim, M.S., Baek, M.J., Lee, M.H., Park, J.W., Lee, S.Y., Soderhall, K., Lee, B.L., 2002. A new easter-type serine protease cleaves a masquerade-like protein during prophenoloxidase activation in *Holotrichia diomphalia* larvae. J Biol Chem 277, 39999-40004.

Koo, Y.D., Ahn, J.E., Salzman, R.A., Moon, J., Chi, Y.H., Yun, D.J., Lee, S.Y., Koiwa, H., Zhu-Salzman, K., 2008. Functional expression of an insect cathepsin B-like counter-defence protein. Insect Mol Biol 17, 235-245.

Lalmanach, G., Boulange, A., Serveau, C., Lecaille, F., Scharfstein, J., Gauthier, F., Authie, E., 2002. Congopain from *Trypanosoma congolense*: drug target and vaccine candidate. Biol Chem 383, 739-749.

Law, R.H., Smooker, P.M., Irving, J.A., Piedrafita, D., Ponting, R., Kennedy, N.J., Whisstock, J.C., Pike, R.N., Spithill, T.W., 2003. Cloning and expression of the major secreted cathepsin B-like protein from juvenile *Fasciola hepatica* and analysis of immunogenicity following liver fluke infection. Infect Immun 71, 6921-6932.

Leak, S.G.A. (2007). Tsetse Biology and Ecology: Their Role in the Epidemiology and Control of Trypanosomosis. Oxfordshire, UK. CPI Anatomy Rowe, Eastbourne.

Lee, K.S., Kim, B.Y., Choo, Y.M., Yoon, H.J., Kang, P.D., Woo, S.D., Sohn, H.D., Roh, J.Y., Gui, Z.Z., Je, Y.H., Jin, B.R., 2009. Expression profile of cathepsin B in the fat body of *Bombyx mori* during metamorphosis. Comp Biochem Physiol B Biochem Mol Biol 154, 188-194.

Lehane, S.M., Assinder, S.J., Lehane, M.J., 1998. Cloning, sequencing, temporal expression and tissue-specificity of two serine proteases from the midgut of the blood-feeding fly *Stomoxys calcitrans*. Eur J Biochem 254, 290-296.

Liniger, M., Acosta-Serrano, A., Van Den Abbeele, J., Kunz Renggli, C., Brun, R., Englund, P.T., Roditi, I., 2003a. Cleavage of trypanosome surface glycoproteins by alkaline trypsin-like enzyme(s) in the midgut of *Glossina morsitans*. Int J Parasitol 33, 1319-1328.

Liniger, M., Acosta-Serrano, A., Van Den Abbeele, J., Kunz Renggli, C., Brun, R., Englund, P.T., Roditi, I., 2003b. Cleavage of trypanosome surface glycoproteins by

alkaline trypsin-like enzyme(s) in the midgut of *Glossina morsitans*. International Journal for Parasitology 33, 1319-1328.

Liniger, M., Acosta-Serrano, A., Van Den Abbeele, J., Kunz Renggli, C., Brun, R., Englund, P.T., Roditi, I., 2003c. Cleavage of trypanosome surface glycoproteins by alkaline trypsin-like enzyme(s) in the midgut of *Glossina morsitans*. Inter J Parasitol 33, 1319-1328.

Liu, J., Shi, G.P., Zhang, W.Q., Zhang, G.R., Xu, W.H., 2006. Cathepsin L function in insect moulting: molecular cloning and functional analysis in cotton bollworm, *Helicoverpa armigera*. Insect Mol Biol 15, 823-834.

Liu, X., McCarron, R.C., Nordin, J.H., 1996. A cysteine protease that processes insect vitellin. Purification and partial characterization of the enzyme and the proenzyme. J Biol Chem 271, 33344-33351.

Marquardt, W. C., Black, W. C., Higgs, S., Freier, J. E., James, A. A., Hagedorn, H. H., Kondratieff, B., Hemingway, J., Moore, C. G. (2004). Biology of Disease Vectors. 2nd ed. China. Elsevier Academic Press.

Maudlin, I., Holmes, P. H., Miles, M. A. (2003). The Trypanosomiases. Oxfordshire, UK. Cromwell Press, Trowbridge.

Moffatt, M.R., Lehane, M.J., 1990. Trypsin is stored as an inactive zymogen in the midgut of *Stomoxys calcitrans*. Insect Biochem 20, 719-723.

Molina-Cruz, A., Gupta, L., Richardson, J., Bennett, K., Black, W.t., Barillas-Mury, C., 2005. Effect of mosquito midgut trypsin activity on dengue-2 virus infection and dissemination in *Aedes aegypti*. Am J Trop Med Hyg 72, 631-637.

Mort, J.S., Buttle, D.J., 1997. Cathepsin B. Int J Biochem Cell Biol 29, 715-720.

Muller, H.M., Crampton, J.M., della Torre, A., Sinden, R., Crisanti, A., 1993. Members of a trypsin gene family in *Anopheles gambiae* are induced in the gut by blood meal. Embo J 12, 2891-2900.

Mwangi, S., Murungi, E., Jonas, M., Christoffels, A., 2011. Evolutionary genomics of *Glossina morsitans* immune-related CLIP domain serine protease and serine protease inhibitors. Infect, Genet Evol 11, 740-745.

Nigam, Y., Maudlin, I., Welburn, S., Ratcliffe, N.A., 1997. Detection of phenoloxidase activity in the hemolymph of tsetse flies, refractory and susceptible to infection with *Trypanosoma brucei rhodesiense*. J Invertebr Pathol 69, 279-281.

Oliveira, M.G., De Simone, S.G., Xavier, L.P., Guedes, R.N., 2005. Partial purification and characterization of digestive trypsin-like proteases from the velvet bean caterpillar, *Anticarsia gemmatalis*. Comp Biochem Physiol B Biochem Mol Biol 140, 369-380.

Olsen, J.V., Ong, S.E., Mann, M., 2004. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol Cell Proteomics 3, 608-614.

Osir, E.O., Imbuga, M.O., Onyango, P., 1993. Inhibition of *Glossina morsitans* midgut trypsin activity by D-glucosamine. Parasitol Res 79, 93-97.

Paskewitz, S.M., Andreev, O., Shi, L., 2006. Gene silencing of serine proteases affects melanization of Sephadex beads in *Anopheles gambiae*. Insect Biochem Mol Biol 36, 701-711.

Peacock, L., Cook, S., Ferris, V., Bailey, M., Gibson, W., 2012. The life cycle of *Trypanosoma* (Nannomonas) *congolense* in the tsetse fly. Parasit Vectors 5, 109.

Pearson, T.W., 2001. Procyclins, proteases and proteomics: dissecting trypanosomes in the tsetse fly. Trends Microbiol 9, 299-301.

Philip, J.M., Fitches, E., Harrison, R.L., Bonning, B., Gatehouse, J.A., 2007. Characterisation of functional and insecticidal properties of a recombinant cathepsin L-like proteinase from flesh fly (*Sarcophaga peregrina*), which plays a role in differentiation of imaginal discs. Insect Biochem Mol Biol 37, 589-600.

Pimenta, P.F., Modi, G.B., Pereira, S.T., Shahabuddin, M., Sacks, D.L., 1997. A novel role for the peritrophic matrix in protecting *Leishmania* from the hydrolytic activities of the sand fly midgut. Parasitology 115 (Pt 4), 359-369.

Pyati, P.S., Bell, H.A., Fitches, E., Price, D.R., Gatehouse, A.M., Gatehouse, J.A., 2009. Cathepsin L-like cysteine proteinase (DcCathL) from *Delia coarctata* (wheat bulb fly): basis of insecticidal activity. Insect Biochem Mol Biol 39, 535-546.

Raikhel, A.S., Dhadialla, T.S., 1992. Accumulation of yolk proteins in insect oocytes. Annu Rev Entomol 37, 217-251.

Ribolla, P.E.M., Daffre, S., De Bianchi, A.G., 1993. Cathepsin B and acid phosphatase activities during *Musca domestica* embryogenesis. Insect Biochem Mol Biol 23, 217-223.

Roditi, I., Liniger, M., 2002. Dressed for success: the surface coats of insect-borne protozoan parasites. Trends Microbiol 10, 128-134.

Roditi, I., Pearson, T.W., 1990. The procyclin coat of African trypanosomes (or the not-so-naked trypanosome). Parasitol Today 6, 79-82.

Sbicego, S., Vassella, E., Kurath, U., Blum, B., Roditi, I., 1999. The use of transgenic *Trypanosoma brucei* to identify compounds inducing the differentiation of blood-stream forms to procyclic forms. Mol Biochem Parasitol 104, 311-322.

Schechter, I., Berger, A., 1967. On the size of the active site in proteases. I. Papain. Biochem Biophys Res Commun 27, 157-162.

Shahabuddin, M., Lemos, F.J., Kaslow, D.C., Jacobs-Lorena, M., 1996. Antibody-mediated inhibition of *Aedes aegypti* midgut trypsins blocks sporogonic development of *Plasmodium gallinaceum*. Infect Immun 64, 739-743.

Shiba, H., Uchida, D., Kobayashi, H., Natori, M., 2001. Involvement of cathepsin B- and L-like proteinases in silk gland histolysis during metamorphosis of *Bombyx mori*. Arch Biochem Biophys 390, 28-34.

Smooker, P.M., Jayaraj, R., Pike, R.N., Spithill, T.W., 2010. Cathepsin B proteases of flukes: the key to facilitating parasite control? Trends Parasitol 26, 506-514.

Srinivasan, A., Giri, A.P., Gupta, V.S., 2006. Structural and functional diversities in lepidopteran serine proteases. Cell Mol Biol Lett 11, 132-154.

Sriraman, V., Richards, J.S., 2004. Cathepsin L gene expression and promoter activation in rodent granulosa cells. Endocrinology 145, 582-591.

Sriwichai, P., Rongsiryam, Y., Jariyapan, N., Sattabongkot, J., Apiwathnasorn, C., Nacapunchai, D., Paskewitz, S., 2012. Cloning of a Trypsin-like Serine Protease and Expression Patterns During *Plasmodium falciparum* Invasion in the Mosquito, *Anopheles dirus* (Peyton and Harrison). Arch Insect Biochem Physiol 80, 151-165.

Stack, C.M., Caffrey, C.R., Donnelly, S.M., Seshaadri, A., Lowther, J., Tort, J.F., Collins, P.R., Robinson, M.W., Xu, W., McKerrow, J.H., Craik, C.S., Geiger, S.R., Marion,

R., Brinen, L.S., Dalton, J.P., 2008. Structural and functional relationships in the virulence-associated cathepsin L proteases of the parasitic liver fluke, *Fasciola hepatica*. J Biol Chem 283, 9896-9908.

Stephans, A., Rojo, L., Araujo-Bernel, S., Garcia-Cerreno, F., Muhlia-Almazan, A., 2012. Cathepsin B from the white shrimp *Litopenaeus vannamei:* cDNA sequence analysis, tissues-specific expression and biological activity. Comp Biochem Physiol B Biochem Mol Biol 161, 32-40.

Takahashi, N., Kurata, S., Natori, S., 1993. Molecular cloning of cDNA for the 29 kDa proteinase participating in decomposition of the larval fat body during metamorphosis of *Sarcophaga peregrina* (flesh fly). FEBS Lett 334, 153-157.

Tao, K., Stearns, N., Dong, J., Wu, Q., Sahagian, G., 1994. The proregion of cathepsin L is required for proper folding, stability, and ER exist. Arch Biochem Biophys 311, 19-27.

Terra, W.R., Ferreira, C., 1994. Insect digestive enzymes: Properties, compartmentalization and fuction. . Comp. Biochem. Physiol. B. Biochem. Mol. Biol. 109, 1-62.

Tong, Y., Jiang, H., Kanost, M.R., 2005. Identification of plasma proteases inhibited by *Manduca sexta* serpin-4 and serpin-5 and their association with components of the prophenol oxidase activation pathway. J Biol Chem 280, 14932-14942.

Tryselius, Y., Hultmark, D., 1997. Cysteine proteinase 1 (CP1), a cathepsin L-like enzyme expressed in the *Drosophila melanogaster* haemocyte cell line mbn-2. Insect Mol Biol 6, 173-181.

Utz, S., Roditi, I., Kunz Renggli, C., Almeida, I.C., Acosta-Serrano, A., Butikofer, P., 2006. *Trypanosoma congolense* procyclins: unmasking cryptic major surface glycoproteins in procyclic forms. Eukaryot Cell 5, 1430-1440.

Wang, L.F., Chai, L.Q., He, H.J., Wang, Q., Wang, J.X., Zhao, X.F., 2010. A cathepsin L-like proteinase is involved in moulting and metamorphosis in *Helicoverpa armigera*. Insect Mol Biol 19, 99-111.

Wang, X., Liu, B., Wang, G., Tang, B., Xiang, J., 2008a. Molecular cloning and functional analysis of cathepsin B in nutrient metabolism during larval development in *Meretrix meretrix*. Aquaculture 282, 41-46.

Wang, Y., Luo, W., Reiser, G., 2008b. Trypsin and trypsin-like proteases in the brain: proteolysis and cellular functions. Cell Mol Life Sci 65, 237-252.

Waniek, P.J., Pacheco Costa, J.E., Jansen, A.M., Costa, J., Araujo, C.A., 2012. Cathepsin L of *Triatoma brasiliensis* (Reduviidae, Triatominae): sequence characterization, expression pattern and zymography. J Insect Physiol 58, 178-187.

Welburn, S.C., Maudlin, I., 1999. Tsetse-trypanosome interactions: rites of passage. Parasitol Today 15, 399-403.

WHO Fact Sheet on Trypanosomiasis, Human African (sleeping sickness), 2013. Available at: http://www.who.int/mediacentre/factsheets/fs259/en/. Accessed June 2 , 2013.

Xu, Y., Kawasaki, H., 2001. Isolation and expression of cathepsin B cDNA in hemocytes during metamorphosis of *Bombyx mori*. Comp Biochem Physiol B Biochem Mol Biol 130, 393-399.

Yabu, Y., Takayanagi, T., 1988. Trypsin-stimulated transformation of *Trypanosoma brucei gambiense* bloodstream forms to procyclic forms in vitro. Parasitol Res 74, 501-506.

Yan, J., Cheng, Q., Li, C.B., Aksoy, S., 2001. Molecular characterization of two serine proteases expressed in gut tissue of the African trypanosome vector, *Glossina morsitans morsitans*. Insect Mol Biol 10, 47-56.

Yan, J., Cheng, Q., Li, C.B., Aksoy, S., 2002a. Molecular characterization of three gut genes from *Glossina morsitans morsitans*: cathepsin B, zinc-metalloprotease and zinc-carboxypeptidase. Insect Mol Biol 11, 57-65.

Yan, J., Cheng, Q., Narashimhan, S., Li, C.B., Aksoy, S., 2002b. Cloning and functional expression of a fat body-specific chitinase cDNA from the tsetse fly, *Glossina morsitans morsitans*. Insect Biochem Mol Biol 32, 979-989.

Yang, X.M., Hou, L.J., Dong, D.J., Shao, H.L., Wang, J.X., Zhao, X.F., 2006. Cathepsin B-like proteinase is involved in the decomposition of the adult fat body of *Helicoverpa armigera*. Arch Insect Biochem Physiol 62, 1-10.

Yang, X.M., Hou, L.J., Wang, J.X., Zhao, X.F., 2007. Expression and function of cathepsin B-like proteinase in larval hemocytes of *Helicoverpa armigera* during metamorphosis. Arch Insect Biochem Physiol 64, 164-174.

Yano, T., Takahashi, N., Kurata, S., Natori, S., 1995. Regulation of the Expression of Cathepsin B in *Sarcophaga Peregrina* (Flesh Fly) at the Translational Level During Metamorphosis. Eur J Biochem 234, 39-43.

Zavasnik-Bergant, T., Turk, B., 2007. Cysteine proteases: destruction ability versus immunomodulation capacity in immune cells. Biol Chem 388, 1141-1149.

Zhai, X., Zhao, X.F., 2012. Participation of haemocytes in fat body degradation via cathepsin L expression. Insect Mol Biol 21, 521-534.

Zhao, X.F., Wang, J.X., Li, F.X., Sueda, S., Kondo, H., 2005. Analysis of substrate specificity and endopeptidyl activities of the cathepsin B-like proteinase from *Helicoverpa armigera*. Protein J 24, 219-225.

Zhao, X.F., Wang, J.X., Xu, X.L., Schmid, R., Wieczorek, H., 2002. Molecular cloning and characterization of cathepsin B-like proteinase from the cotton boll worm, *Helicoverpa armigera.* Insect Mol Biol 11, 567-575.