Title
The Proteolytic Repertoire of the Planarian Schmidtea mediterranea

Permalink
https://escholarship.org/uc/item/7sk8r6fc

Author
Goupil, Louise

Publication Date
2016

Peer reviewed|Thesis/dissertation

The Proteolytic Repertoire of the Planarian Schmidtea mediterranea

by

Louise Susan Goupil

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

ACKNOWLEDGEMENTS

Chapter 2 is adapted from Louise S. Goupil, Sam L. Ivry, Ivy Hsieh, Brian M. Suzuki, Charles S. Craik, Anthony O'Donoghue, and James H. McKerrow. "Cysteine and aspartyl proteases contribute to protein digestion in the gut of freshwater planaria." *PLoS NTD*. 2016; **10(8)**: e0004893. doi: 10.1371/journal.pntd.0004893. The dissertation author is the first author of this manuscript.

ABSTRACT OF THE DISSERTATION

The Proteolytic Repertoire of the Planarian *Schmidtea mediterranea*

by

Louise S. Goupil

Doctor of Philosophy in Chemical Biology

University of California, San Francisco, 2016

Professor James H. McKerrow, Chair

Proteases perform numerous vital functions in flatworms, many of which are likely to be conserved throughout the phylum *Platyhelminthes*. Within this phylum are several parasitic worms with minimal biochemical characterization due to their complex life-cycles and lack of responsiveness to genetic manipulation. The flatworm *Schmidtea mediterranea*, or planaria*,* is an ideal model organism to study the complex role of protein digestion in flatworms due to its simple life cycle and amenability to techniques like RNA interference (RNAi). In this study, we were interested in better understanding the digestive protease system that exists in the planarian gut. To do this, we developed an alcohol-induced regurgitation technique to isolate the gut enzymes in *S.*

mediterranea. Using a panel of fluorescent substrates, we show that this treatment induces a release of gut proteolytic activity. *Schmidtea* gut proteases have broad yet diverse substrate specificity profiles. Proteomic analysis of the gut contents confirmed the presence of both cysteine and metallo-proteases. However, treatment with class-specific protease inhibitors showed that aspartyl and cysteine proteases are responsible for the majority of protein digestion. Specific RNAi knockdown of the cathepsin B-like cysteine protease (SmedCB) reduced protein degradation *in vivo*. Immunohistochemistry and whole-mount *in situ* hybridization (WISH) confirmed that the full-length and active forms of SmedCB are found in secretory cells surrounding the planaria intestinal lumen. Finally, we show that the knockdown of SmedCB reduces the speed of tissue regeneration. Defining the roles of proteases in planaria can provide insight to functions of conserved proteases in parasitic flatworms, and potentially identify new drug targets in parasites.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

1.1 Planaria: a model flatworm

The family *Platyhelminthes* contains over 25,000 known species of flatworms, and is the fourth largest family in existence.[1] The family includes parasitc cestodes (tapeworms), trematodes (flukes), and the mostly non-parasitic Turbellarians. All flatworms are bilaterally symmetric, unsegmented acoelomates possessing a blind-end gut. Their flattened shape is due to a lack of circulatory and respiratory systems; the increased surface area promotes diffusion of oxygen and nutrients through their bodies. Over half of *Platyhelminthes* are parasitic, including several human parasites. Due to their relevance to human health, as well as their important position in the evolution of Metazoa,[2] members of the family *Platyhelminthes* are a key focus for research.

One of the best-studied groups of *Platyhelminthes* are the free-living planarian flatworms. These worms have been studied for over a century due to their remarkable regenerative capacity, which was first observed in the late 19th century by Thomas Morgan. Morgan calculated that the minimum requirement needed for complete regeneration is a mere 1/279th of total body size.[3] Planarians regenerate using a large number of totipotent stem cells known as neoblasts. Between 20-30% of the all cells in the parenchyma, or connective tissue, are neoblasts. These stem cells are the only mitotically dividing cells in the planarian worm,[4] and are distributed throughout the body except in the pharynx and the region anterior to the eyespots. Neoblasts maintain tissue homeostasis; they continuously provide new cells and maintain body ratios during degrowth (or shrinkage) during periods of food restriction. Under starvation conditions, planaria reduce their total cell number but maintain their appropriate body size and pattern. During tissue loss or injury, a large mitotic peak is observed and neoblasts migrate to the site of injury. Within 2-3 days,

neoblasts begin differentiating into new tissue types and missing body tissues are completely regrown in only a week.[5]

Many studies of planarians have focused on the worm *Schmidtea mediterranea*. These worms are triploblastic organisms belonging to the Lophotrochozozoan clade. *S. mediterranea* are categorized as triclads due to their three-branch digestive system, which connects to a pharynx.[6] The blind-end gut distributes food throughout the body before expelling waste back out of the pharynx. Connective tissue fills the space between the branched gut and the muscles of the body wall. These muscles allow planaria to move with the help of simultaneous slime secretions and ciliated ventral tissue.[7] *S. mediterranea* can reproduce sexually or asexually. Sexual reproduction occurs by cross fertilization of hermaphrodites. The eggs produced are ectolecithal, meaning that the yolk is outside of the egg. Sexual maturity depends on feeding levels; during periods of starvation, sexual organs are resorbed and must grow *de novo* once food is more readily available. Asexual reproduction may also occur by transverse fission.[8,9] *S. mediterranea* worm size ranges from 1mm to 2.5cm depending on food availability, although some species of planaria may be as long as 40-50cm.[10]

There are many advantages of *S. mediterranea* as an organism for study. These planarians have a simple lifecycle and are able to generate large colonies through asexual reproduction. The *S. mediterranea* genome has also been sequenced and is readily available to the public. There are an estimated 30,000 genes distributed on 4 chromosomes of these diploid organisms.[11] Transgenic organisms can be created with some success;[12] most parasitic flatworms are not amenable to this technique. Live worms can be injected with DNA constructs flanked by repeat transposable sequences, along with a secondary vector with a helper to facilitate integration. To date, this is only available for worms that reproduce sexually. Most notably, loss-of-function phenotypes can

be observed by RNA interference (RNAi) in *S. mediterranea*. Large-scale genetic screening has shown that RNAi knockdowns are widespread, efficient, and can produce a variety of phenotypes. One such study showed phenotypes associated with 240 of 1065 genes knocked down. These genes were intended as a representative sample of the planaria genome.[13] Proteomic studies have also been performed in asexual *S. mediterranea* and some 3,000 non-redundant cDNAs have been characterized, including transcription factors, DNA replication/modification molecules, and a variety of membrane-associated proteins.[14] Because of these many advantages, *S. mediterranea* has been proposed as model organism for other flatworms, like the parasite *Schistosoma mansoni*, which is difficult to study due to its complex lifecycle.

1.2 *Schistosoma*: the causative agent of schistosomiasis

The disease schistosomiasis, caused by parasitic flatworms of the genus *Schistosoma*, affects over 200 million people worldwide with an additional 700 million at risk for infection. The annual global morbidity of schistosomiasis is 14,000,[15] but the true measure of the impact of this disease is better represented by disability-adjusted life years (DALYs). This measure of disease burden expresses the number of years of human life and productivity lost due to ill-health, disability, and death. An estimated 3.31 million DALYs are lost annually due to schistosomiasis,[16] second only to malaria as the most burdensome tropical disease.

The *Schistosoma* life cycle relies on an intermediate freshwater snail host and a terminal mammalian host. The eggs of *Schistosoma* worms hatch in freshwater and the resulting miracidia infect snails. The worms develop into cercariae, which are triggered to leave the snail into the surrounding water. Mammals, including humans, become infected when they enter parasite-infested water. *Schistosoma* cercariae use proteolytic secretions to degrade host skin and enter the

bloodstream. In six weeks, they mature into adults. Depending on the species of worm, adults reside mostly in the superior mesenteric veins of the small intestine (*S. japonicum*) or large intestine (*S. mansoni*), or in the venous plexus of the bladder (*S. haematobium*). Adults are always paired: a large female (1.2-2cm) and a slender male (1cm). Both have a cylindrical body with blind-end gut fed by two suckers and protected by a complex tegument. Unlike planaria, *Schistosoma* worms are not hermaphrodites and individual worms possess either male or female reproductive systems. Female worms can produce hundreds to thousands of eggs per day depending on the species. The reproductive potential of a single female is upwards of 500 million eggs during her lifetime.[17] Eggs escape the mammalian hosts by transiting to the lumen of the intestine (*S. japonicum* and *S. mansoni*) or the bladder (*S. haematobium*) to be excreted with feces or urine, respectively. When eggs hatch in freshwater, they release another motile larval form, the miracidium, which infects an intermediate host snail.

The pathology of schistosomiasis is largely due to the presence of eggs trapped inside the host. While the juvenile stage (schistosomula) can cause disease, most patients recover spontaneously after 2-10 weeks once the schistosomulae reach adulthood. Adults escape detection of the immune system, but eggs trigger immune responses resulting in chronic inflammation characterized by granulomata. Trapped eggs can also cause genitourinary lesions,[18] which can increase the transmission of diseases like HIV,[19] especially in Sub-Saharan populations where both diseases are prevalent.

The current drug of choice for schistosomiasis treatment is the trematocide and cestocide praziquantel (acylated quinoline-pyrazine, PZQ). This drug is highly effective at killing adult worms with cure rates above 90%. PZQ can be orally administered and is safe for young children and pregnant women to use.[20] The precise target of PZQ is unknown; parasites experience rapid

muscle contractions followed by paralysis. Despite its many advantages, PZQ has little to no effect on eggs or juvenile worms. Because PZQ is used for mass drug administration, and has been in use for over 40 years, there is concern that resistance will arise.[21] Furthermore, PZQ does not prevent reinfection, which is common in areas where *Schistosoma* worms are endemic. No vaccines currently exist to prevent infection, so PZQ remains the best form of treatment despite its drawbacks.

Aside from targeting the worms directly, another proposed method to reduce *Schistosoma* populations would be to target the intermediate host snail. While snail populations can be reduced, molluscides can have adverse effects on other aquatic organisms, and often fail to completely eliminate the snails themselves. Proper water sanitation can help eliminate transmission of schistosomiasis, but with millions worldwide still without a safe water supply, new therapies for treating this disease must be explored while people are still at risk.[22] New treatment strategies and new drug targets must be identified to help reduce and prevent infection.

There are several drawbacks to studying parasitic *Schistosoma* flatworms directly. Many of these issues stem from the complex lifecycle of these parasites, which require researchers to maintain populations of both freshwater snails and mammalian hosts. The maturation time for *Schistosoma* from egg to adult is 4-6 weeks, so the process of propagating these worms is lengthy as well as expensive and labor-intensive. Furthermore, genetic tools are very limited in these flatworms. Although advances have been made in the development of tools for transgenesis, no transgenic schistosome lines have been reported to date.[23] Loss-of-function studies can be attempted with RNAi, but this technique produces highly variable results.[24,25] There is still no standardized technique. Results from the few studies of gene knockdown are highly variable and

phenotypic results can be difficult to observe or replicate. Because of this, it is difficult to search for new potential targets of therapeutic interest in *Schistosoma* parasites directly.

1.3 Similarities between planaria and schistosomes: the case for a free-living model organism

The free-living planarian *S. mediterranea* and parasitic *Schistosoma* worms share many similarities despite their differences in lifestyle. Like all flatworms, they are bilaterally symmetric acoelomates with a blind-end gut. Several other body structures are conserved between *S. mediterranea* and *S. mansoni*, including the protonephridia essential in water balance.[26] They also possess special glands associated with egg-laying known as Mehlis' gland (schistosomes) or cement glands (planarians). Both worms belong to a subset of Plathelminthes known as neoophora due to their ectoletchical eggs.[27] Recently, a population of neoblast-like stem cells has been identified in *S. mansoni*.[28] These somatic stem cells may give insight as to how schistosomes can survive for decades inside their hosts.

Aside from the conservation in their overall morphology, planarians and *S. mansoni* share many important genes and proteins. One proteomic study found that of the 1604 proteins identified from *S. mediterranea*, 1369 (or 85%) were also found in the *S. mansoni* proteome.[29] This included both serine and metalloproteases. Proteases are potential therapeutic targets, as will be discussed in the following section.

S. mediterranea are a strong candidate for understanding flatworm biology due to its similarity to the parasite *S. mansoni* and other intractable flatworm parasites. However, while the regulation of tissue regeneration has been studied in depth in planaria, much less is known about the importance of proteases. As proteases constitute some 2.5% of the *S. mansoni* proteome,[30] it is

likely that a similar such number of proteases are conserved in the free-living *S. mediterranea*. Increasing our understanding of planaria proteases could provide important insights into flatworm biology conserved in many parasites, as well as identify potential therapeutic targets for treatment.

1.4 Proteases: a diverse family of enzymes

When searching for new therapeutic targets against *Schistosoma* worms, the most promising candidates are proteins found at the host-parasite interface.[31] This includes a variety of proteolytic enzymes that are used by parasites to degrade host proteins.

Proteases are found in all organisms, including both prokaryotes and eukaryotes. This family of enzymes hydrolyzes proteins at a site known as the scissile bond. Convention denotes the residues on the N-terminal side of peptide bond cleavage as "non-prime" and the amino acid residues on the C-terminal side of cleavage as "prime" positions. These sites are also distinguished as "P," part of the protein being cleaved, or "S," a site on the protease enzyme. Proteases can be classified by their cleavage specificity and pH optima, but are most generally described by the nucleophile responsible for peptide hydrolysis. There are five broad groups of proteolytic enzymes found in Metazoans, each named for the amino acid or ion responsible for their catalysis: serine, cysteine, threonine, aspartyl, and metalloproteases. The first three types act via a covalent acyl-enzyme intermediate, while aspartyl and metalloproteases activate a water molecule for noncovalent hydrolysis of the peptide bond.

Because the cleavage of a peptide bond by a protease is irreversible, proteolytic activity is tightly controlled in the cell. Mechanisms of control include spatial and temporal regulation of protease production and release, as well as degradation of proteases. Many proteases are synthesized in an inactive zymogen form and are only activated once this zymogen "tag" is cleaved

autoproteolytically or by a second protease. Cells also employ the use of macromolecular inhibitors to block activity of proteases.[32] Proteases are involved in a range of cellular processes, so targeting proteolytic activity can modulate many pathways throughout the cell.

In *Schistosoma* worms, protease activity ranges from parasite-specific activities like invading through the host skin,[33] to more general purposes, like protein digestion. Several of these proteases are described in detail in the following section. Because parasitic flatworms and their human hosts use homologous proteolytic enzymes, there is a concern that chemotherapeutic agents may induce toxicity in humans. Thus far, this fear has been alleviated for some protease targets thanks to pre-clinical studies showing high tolerance for cysteine protease inhibitors in mammals.[34] One hypothesis is that many mammalian cysteine proteases serve redundant roles, while essential functions overlap less in parasites.[35] It is also possible that due to the high concentration of mammalian lysosomal cysteine proteases (millimolar), the human cells are less affected by any off-target effects on their activity. Most parasites possess proteases in much lower quantities, so the amount of drug required to block parasitic activity is insufficient to abolish host protease activity.[36,37] Furthermore, parasitic proteases are often secreted extracellularly while host homologues are stored intracelluarly, lowering the chance of unwanted inhibition of host enzymes.[38]

1.5 Protease classes involved in *S. mansoni* protein digestion

One of the major functions of proteolytic activity in parasites is digestion of host proteins. Digestion of host albumin and hemoglobin occurs in the blind-end gut of parasites like *S. mansoni,* where degraded proteins provide nutrients for developing schistosomula and adult schistosomes. It is possible that the activity of many of these proteases might be conserved throughout the family

Platyhelminthes; all worms must acquire and break down nutrients acquired from their surroundings.

Flatworms like *S. mansoni* have a diverse number of proteases that work in a cooperative network to degrade proteins during digestion. Several key proteases have been observed in *S. mansoni* including cysteine proteases cathepsin B (SmCB1), cathepsin L (SmCL1 and SmCL2), cathepsin C (SmCC), and legumain (SmAE), as well as the aspartyl protease cathepsin D (SmCD).[39] Many trematodes have similar gut proteases, although in some cases the relative of abundance of cathepsin L is higher than cathepsin B, as in *Fasciola*.[40,41] Previous RNAi knockdown experiments suggest that both classes of proteases contribute to host protein digestion.[42] These enzymes have been identified directly in the gut, as schistosomes can be induced to release their luminal contents and gastrointestinal contents (GIC) when placed in distilled water.[43] Following this osmotic shock, gut contents were analyzed. Competitive labeling with small molecule inhibitors helped identify key proteases. The GIC of adult schistosomes has been estimated to be between pH 6.4-6.8.[44] The optimal environment for cysteine and aspartyl proteases is acidic, such as the gut lumen (cysteine) or in vacuoles of intestinal lining cells (aspartyl). Optimal degradation of the host proteins albumen and hemoglobin by GIC *in vitro* was observed at pH 4,[45] suggesting that proteolysis might occur in both luminal or cellular microenvironments of the gut.

There are several cysteine and aspartyl proteases active in the schistosome GIC. The most abundant and active papain family protease in the schistosome gut and GIC is the cathepsin B protease SmCB1.[46] This enzyme has been immunolocalized to both the gut lumen and cells of the gastrodermis, and it exhibits both exo- and endopeptidase activity.[47] SmCB1 has been used as a serodiagnostic marker of schistosomiasis.[48] It shares redundancy in substrates with the cathepsin

L protease SmCL1, which is localized to gastrodermal cells and the dorsal tegument.[49] Unlike SmCB1 and SmCL1, the cathepsin C protease, SmCC, is solely an exoprotease.[50] Exoproteases play a downstream role in the degradation of host proteins into absorbable peptides. In the network of protease activity in the schistosome gut, SmCC acts after endopeptidases like SmCB1, SmCL, and the aspartyl protease SmCD. Another cysteine protease identified in the GIC of schistosomes known as legumain, or SmAE, is also localized in the gut epithelium as well as the ventral surface of adult males.[51] Its precise role is still unknown. Aspartyl proteases, like cathepsin D (SmCD), produce the primary cleavage of hemoglobin while cysteine proteases do the same with albumin. Inhibiting both aspartyl and cysteine proteases in digestion is synergistic. SmCD is found in the cecum and gastrodermis,[52,53] and is optimally active at pH 3-4.

Blocking the activity of the major cysteine proteases in schistosome digestion has been shown to reduce worm burdens and ameliorate chronic pathology of schistosomiasis in animal models. The vinyl sulfone K11777 (*N*-methyl-piperazine-phenylalanyl-homophenylalanyl-vinylsulfone phenyl) irreversibly binds to and inhbits cathepsin B and cathepsin L proteases. This drug was originally developed to target the protozoan *Trypanosoma cruzi*, the causative agent of Chaga's disease.[54] K11777 is orally bioavailable and well-tolerated by mammalian hosts.[55] Treatment of mice infected with *S. mansoni* with K11777 resulted in a 92% reduction in egg burden. If given during worm development, adult female and male worm numbers decreased 80% and 79% respectively.[56] These data suggest that targeting proteases like SmCB1 and SmCL1 could be a promising way to both target adult worms and also reduce the major symptoms of schistosomiasis by decreasing inflammation caused by eggs.

Because many proteases are involved in protein degradation in the *S. mansoni* gut, there are many possible therapeutic targets for drug development. However, due to difficulties

examining specific loss-of-function with RNAi in schistosomes and other parasitic flatworms, validating these targets is a roadblock for identifying the optimal target(s). Because free-living planaria like *S. mediterranea* are so amenable to RNAi, it could be possible to use this surrogate worm to evaluate the relative importance of various proteases. Little is known about the gut of this worm, despite over one hundred years of research on this flatworm. We therefore sought to identify the major classes of proteases present and active in the planarian gut, which would provide insight to the conservation of these proteases among parasitic and free-living members of *Platyhelminthes*.

1.6 Multiplex substrate profiling: a tool for analyzing global protease specificity

Because proteases often work in a complex network, it can be difficult to deconvolute this network to determine which protease(s) provide the bulk of the activity and which peptide targets are preferred. It is necessary to not only identify active proteases in flatworm digestion, but to understand their specificity. Multiplex-substrate profiling by mass spectrometry can be employed to examine global specificity of proteases and provide these necessary insights.

Multiplex-substrate profiling by mass spectrometry (MSP-MS) relies on a physicochemically diverse library of peptide substrates that can reveal the activity of any endo- or exopeptidase. MSP-MS provides a direct cleavage assay that employs mass spectrometry to detect degradation products from the library of synthetic peptides. This library is based on the "two-site hypothesis" of protease activity: all cleavage specificity can be derived from two major sites on a peptide. The library of synthetic peptides was designed to produce every combination of neighbor and near-neighbor pairs of amino acids. The current library consists of 124 14-mer peptides for a total of 1,612 cleavage sites.[57] The length of 14 residues is sufficient to allow binding of endopeptidases without forming tertiary peptide structures, which could limit the accessibility of

these peptides to the proteases. The library is incubated with the protease(s) of interest. As peptides are cleaved and time points are collected, the resulting aliquots are quenched and sequenced via liquid chromatography-tandem mass spectrometry (LC-MS/MS). The relative abundance of cleavages at each site can be tabulated to produce an iceLogo plot, which highlights favored and disfavored cleavages at each site.[58] Cleavage sites are identified by comparison with a control that lacks a peptidase. To date, this technique has been used to examine proteases derived from humans,[59,60] mycobacteria,[61] protozoans,[62] fungi,[63] arthropods,[64] and the flatworms *S. mansoni* and *Fasciola hepatica*.[65,66]

MSP-MS can be used to investigate the global specificity of a single proteases or complex mixtures of proteases. In the case of *S. mansoni*, proteolytic secretions from cercarial larvae revealed the presence of serine, cysteine, and metalloproteases. To further examine the contribution by each class of proteases, the secretions were incubated with specific inhibitors for each class of protease. No effect on the overall cleavage pattern was observed when the mixture was treated with the metalloprotease inhibitor EDTA or the cysteine protease inhibitors E-64 and CA-074. However, an elastase (serine protease) specific chloromethyl ketone inhibitor showed a 36% reduction in total bond cleavage and change in the amino acid enrichment at the P1 position.[67] Employing the use of class-specific or protease-specific inhibitors allows MSP-MS to dissect the activities of complex mixtures from live organisms. This type of analysis has also been performed on extracts of adult schistosomes as well as eggs, which has allowed for comparison of protease activity during various life stages. However, MSP-MS has not yet been used to examine the proteolytic activity of schistosome GIC or other parasitic flatworms.

1.7 Techniques to elucidate the biological function of proteases

Along with using MSP-MS to understand the global profile of proteases in flatworms, orthogonal techniques must be employed to determine a more precise biological function of these enzymes. One such technique is RNAi, whereby a loss of function of a particular protease may reveal a phenotype indicative of the enzyme's importance. RNAi was first developed in the nematode *C. elegans* and exploits a mechanism for gene silencing present in many eukaryotes by causing the destruction of specific mRNA molecules.[68] In this technique, double-stranded microRNAs (miRNAs) or small interfering mRNAs (siRNAs), are introduced into a cell, often via microinjection or electroporation. These double-stranded RNAs are cut by the protein Dicer, and the resulting double-stranded fragments unwind. One strand is degraded while the other, known as the guide strand, joins with the RNA-induced silencing complex (RISC). The guide strand is specially designed to base pair with the mRNA corresponding to the gene of interest. The guide strand will produce double-stranded RNA with the mRNA of interest. The catalytically active protein Argonaute will cleave the mRNA, thus decreasing the amount of available mRNA transcript in the cell for protein translation. Thus, RNAi specifically and robustly suppresses the amount of gene expression for a select target. In some organisms, like planaria, RNAi occurs systemically throughout the animal,[69] although the precise mechanism is not well understood. RNAi is highly efficient in *S. mediterranea*. We can therefore exploit the efficiency of RNAi in planaria to observe whether the loss of any protease function will lead to a detrimental phenotype. The effects of RNAi are more variable in *S. mansoni*. A previous study stated that long-term suppression of SmCB1 via RNAi retarded schistosome growth but this result has not been reproduced.[70]

Localization of proteases can also provide insight as to their biological functions. For example, the proteases in *S. mansoni* digestion discussed previously all localized to tissues in and surrounding the schistosome gut, as is expected because they play a role in digestion or absorption of nutrients. Protein localization is achieved via immunohistochemistry. The most commonly used technique is an indirect method whereby antibodies produced against the protein of interest can be used to label the protein. A secondary antibody, fused to a detectable tag, binds to the first antibody. This secondary antibody allows for a variety of tags to be used depending on the desired technique; a fluorescent reporter can be employed for light microscopy, while a gold nanobead may be used for electron microscopy. In the case of proteases, the primary antibody can be generated against antigens in the inactive, zymogen form or the cleaved, active form of the protease. This will not only label the protease, but also provide insights as to the spatial and temporal activation of the protease.

Localization can also provide information about gene expression and transcription levels. *In situ* hybridization uses a labeled strand of complementary DNA or RNA to localize DNA or RNA throughout an entire organism, which is referred to as whole-mount *in situ* hybridization (WISH). To do this in organisms like planaria, organisms are fixed and permeabilized. The probe is then added; often an antigen-labeled base called digoxigenin is used. This probe is then visualized and quantified, providing information about the areas of the organism where transcription and expression of this gene occur. WISH is well established in *S. mediterranea* and can be performed as a high throughput assay using specialized robotics.[71] This protocol has since been adapted to perform WISH on *S. mansoni* for several transcripts as well.[72]

Together with MSP-MS, RNAi and localization techniques can help provide insight as to the biological roles played by proteases in the free-living worm *Schmidtea mediterranea*. They

can be used to evaluate the major proteases involved in flatworm digestion and help determine

whether these roles appear to be conserved with out knowledge of protein digestion in *S. mansoni*.

This will provide new insights into planaria digestion as well as help further establish *S. mediterranea* as a potential model organism for parasitic flatworms.

1.8 References

[1] Littlewood, D.T.J. & Bray R.A. Interrelationships of the Platyhelminthes. Taylor & Francis, London (2001), pp. 127–134.

[2] Adoutte, A., Balavoine, G., Lartillot, N. & de Rosa, R. Animal evolution: the end of the intermediate taxa? *Trends Genet.* 1999; **15,** 104–108.

[3] Morgan, T.H. Experimental studies of the regeneration of *Planaria maculata*. Arch. Entwm. Org. 1898; **7**: 364–397.

[4] González-Estévez, C. & Saló, E. Autophagy and apoptosis in planarians. *Apoptosis*. 2010; **15**: 279–92.

[5] Pellettieri, J. *et al*. Cell death and tissue remodeling in planarian regeneration. *Dev. Biol*. 2010; **338**: 76–85.

[6] Philippe, H., Lartillot N., Brinkmann H. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol Biol Evol*. 2005; **22**:1246–1253.

[7] González-Estévez, C. & Saló, E. Autophagy and apoptosis in planarians. *Apoptosis*. 2010; **15**: 279–92.

[8] Curtis W. The life history, the normal fission, and the reproductive organs of Planaria maculata. *Proc Boston Soc Nat Hist*. 1902; **30**: 515–559.

[9] Newmark P.A., Wang Y., Chong T. Germ cell specification and regeneration in planarians. *Cold Spring Harb Symp Quant Biol*. 2008; **73**:573–581

[10] Hyman L. The Invertebrates: Platyhelminthes and Rhynchocoela— The Acoelomate Bilateria. McGraw-Hill Book Company Inc, New York (1951).

[11] Gentile, L., Cebrià, F. & Bartscherer, K. The planarian flatworm: an in vivo model for stem cell biology and nervous system regeneration. *Dis. Model. Mech* 2001; **4**: 12–9.

[12] González-Estévez, C., Momose, T., Gehring, W. J. & Saló, E. Transgenic planarian lines obtained by electroporation using transposon-derived vectors and an eye-specific GFP marker. *Proc. Natl. Acad. Sci. U. S. A*. 2003; **100**: 14046–5.

[13] Reddien, P. W., Bermange, A. L., Murfitt, K. J., Jennings, J. R. & Alvarado, A. S. Identification of Genes Needed for Regeneration, Stem Cell Function, and Tissue Homeostasis by Systematic Gene Perturbation in Planaria. *Cell*. 2005; **8**: 635–649.

[14] Alvarado, a. S. The Schmidtea mediterranea database as a molecular resource for studying platyhelminthes, stem cells and regeneration. *Development*. 2002; **129**: 5659–5665.

[15] Gryseels, B., Polman, K., Clerinx, J. & Kestens, L. Human schistosomiasis. *Lancet*. 2006; **368**: 1106–18.

[16] Hotez, P. J., Alvarado, M., Basáñez, M. G., Bolliger, I. Bourne, R. *et al*. The Global Burden of Disease Study 2010: Interpretation and Implications for the Neglected Tropical Diseases. *PLoS Negl. Trop. Dis*. 2014; **8**: 10.1371/journal.pntd.0002865.

[17] Gryseels, B., Polman, K., Clerinx, J. & Kestens, L. Human schistosomiasis. *Lancet*. 2006; **368**: 1106–18.

[18] Cheever, A.W., Hoffmann, K.F., Wynn, T.A. Immunopathology of schistosomiasis mansoni in mice and men. *Immunol Today.* 2000; **21**: 465–66.

[19] Secor, W. E. The effects of schistosomiasis on HIV/AIDS infection, progression and transmission. *Curr Opin HIV AIDS.* 2012; **7**: 254-259.

[20] Andrews, P. Praziquantel: mechanisms of anti-schistosomal activity. *Pharmacol. Ther.* 1985; **29**: 129–56.

[21] King, C. H., Muchiri, E. M. & Ouma, J. H. Evidence against rapid emergence of praziquantel resistance in *Schistosoma haematobium*, Kenya. *Emerg. Infect. Dis.* 2000; **6**: 585–594.

[22] Gryseels, B., Polman, K., Clerinx, J. & Kestens, L. Human schistosomiasis. *Lancet.* 2006; **368**: 1106–18.

[23] Alrefaei, Y. N., Okatcha, T. I., Skinner, D. E., Brindley, P. J. Progress with schistosome transgenesis. *Mem. Inst. Oswaldo Cruz.* 2011; **106**: 785–793.

[24] Geldhof, P. Visser, A. Clark, D., Saunders, G., Britton, C. Gilleard, J., *et* al. RNA interference in parasitic helminthes: current situation, potential pitfalls, and future prospects. *Parasitology.* 2007; **134**: 609-619.

[25] Krautz-Peterson, G., Bhardwaj, R., Faghiri, Z., Tararam, C. A., Skelly, P.J. RNA interference in schistosomes: machinery and methodology. *Parasitology.* 2010; **137**: 485-495

[26] Scimone ML, Srivastava M, Bell GW, Reddien PW. A regulatory program for excretory system regeneration in planarians. *Development.* 2011; **138**: 4387–4398.

[27] Collins, J. J. & Newmark, P. A. It's no fluke: the planarian as a model for understanding schistosomes. *PLoS Pathog.* 2013; **9**: 10.1371/journal.ppat.10033396

[28] Collins, J. J., Wang, B., Lambrus, B. G., Tharp, M. E., Iyer, H., Newmark, P. A. Adult somatic stem cells in the human parasite Schistosoma mansoni. *Nature.* 2013; **494**: 476–9.

[29] Bocchinfuso, D. G., Taylor, P., Ross, E., Ignatchenko, A., Ignatchenko, V., Kislinger, T. *et al.* Proteomic profiling of the planarian *Schmidtea mediterranea* and its mucous reveals similarities with human secretions and those predicted for parasitic flatworms. *Mol. Cell. Proteomics.* 2012; **11**: 681–91.

[30] Berriman, M., Haas, B. J., LoVerde, P. T., Wilson, R. A., Dillon, G. P. Cerqueira, G. C. *et al.* The genome of the blood fluke Schistosoma mansoni. *Nature.* 2009; **460**: 352–8.

[31] Figueirdo, B. C., Ricci, N. D., de Assis N. R., de Morais S. B., Fonseca, C. T., Oliveira, S. C. Kicking in the guts: *Schistosoma mansoni* digestive tract proteins are potential candidates for vaccine development. *Front Immunol.* 2015; **6**:22 doi: 10.3389/fimmu.2015.00022

[32] Farady, C. J., Craik, C. S. Mechanisms of macromolecular protease inhibitors. *Chembiochem.* 2010; **11**: 2341–6.

[33] Lim, K. C., Sun, E., Bahgat, M., Bucks, D. , Guy, R., Cullander, C. *et al.* Blockage of skin invasion by schistosome cercariae by serine protease inhibitors. *Am. J. Trop. Med. Hyg.* 1999; **60**: 487–492.

[34] Ang, K. K. H., Ratnam, J., Gut, J., Legac, J., Hansell, E., Mackey, Z. B. *et al.* Mining a cathepsin inhibitor library for new antiparasitic drug leads. *PLoS Negl. Trop. Dis.* 2011; **5**: e1023. 10.1371/journal.pntd.0001023.

[35] McKerrow, J.H., Engel, J.C., Caffrey, C.R. Cysteine protease inhibitors as chemotherapy for parasitic infections. *Bioorg Med Chem.* 1999; **7**: 639–644.

[36] Scory, S., Caffrey, C.R., Stierhof, Y.D., Ruppel, A., Steverding, D. Trypanosoma brucei: killing of bloodstream forms in vitro and in vivo by the cysteine proteinase inhibitor Z-phe-ala-CHN2. *Exp Parasito.l* 1999; **91**: 327–333.

[37] Engel, J. C., Ang, K. K., Chen, S., Arkin, M. R., McKerrow, J. H., Doyle, P. S. Image- based high-throughput drug screening targeting the intracellular stage of Trypanosoma cruzi, the agent of Chagas' disease. *Antimicrob Agents Chemother.* 2010; **54**: 3326–3334.

[38] Renslo, A. R., McKerrow, J. H. Drug discovery and development for neglected tropical diseases. *Nat. Chem. Biol.* 2006; **2**: 701-710.

[39] Delcroix, M., Sajid, M., Caffrey, C. R., Lim, K. C., Dvorák, J., Hsieh, I. *et al.* A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. *J. Biol. Chem.* 2006; **281**: 39316–29.

16

[40] Dalton, J. P., Neill, S. O., Stack, C., Collins, P. Walshe, A. Sekiya, M. *et al.* Fasciola hepatica cathepsin L-like proteases: Biology, function, and potential in the development of first generation liver fluke vaccines. *Int. J. Parasitol.* 2003; **33**: 1173–1181.

[41] Collins, P. R., Stack, C. M., O'Neill, S. M., Doyle, S., Ryan, T., Brennan, G. P. *et al.* Cathepsin L1, the major protease involved in liver fluke (Fasciola hepatica) virulence: Propeptide cleavage sites and autoactivation of the zymogen secreted from gastrodermal cells. *J. Biol. Chem.* 2004; **279**: 17038–17046.

[42] Delcroix, M., Sajid, M., Caffrey, C. R., Lim, K. C., Dvorák, J., Hsieh, I. *et al.* A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. *J. Biol. Chem.* 2006; **281**: 39316–29.

[43] Chappell, C.L., Dresden, M.H. Schistosoma mansoni: proteinase activity of "hemoglobinase" from the digestive tract of adult worms. *Exp Parasitol.* 1986; **61**:160–167.

[44] Sajid, M., McKerrow, J. H., Hansell, E., Mathieu, M.A. Lucas, K. D., Hsieh, I. *et al.* Functional expression and characterization of Schistosoma mansoni cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. *Mol. Biochem. Parasitol.* 2003; **131**: 65–75.

[45] Delcroix, M., Sajid, M., Caffrey, C. R., Lim, K. C., Dvorák, J., Hsieh, I. *et al.* A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. *J. Biol. Chem.* 2006; **281**: 39316–29.

[46] Caffrey CR, Ruppel A. Cathepsin B-like activity predominates over cathepsin L-like activity in adult Schistosoma mansoni and S. japon- icum. Parasitol Res 1997; 83(6):632–5.

[47] Sajid, M., McKerrow, J. H., Hansell, E., Mathieu, M.A. Lucas, K. D., Hsieh, I. *et al.* Functional expression and characterization of Schistosoma mansoni cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. *Mol. Biochem. Parasitol.* 2003; **131**: 65–75.

[48] Ruppel A., Shi Y.E., Wei D.X., Diesfeld H.J.. Sera of *Schistosoma japonicum*-infected patients cross-react with diagnostic 31/32kD proteins of *S. mansoni. Clin Exp Immunol.* 1987; **69**: 291–8.

[49] Brady, C. P., Dowd, A. J., Brindley, P. J., Ryan, T. Day, S. R., Dalton, J. P. Recombinant expression and localization of *Schistosoma mansoni* cathepsin L1 support its role in the degradation of host hemoglobin. *Infect Immun.* 1999; **67,** 368–374.

[50] Butler R., Michel A., Kunz W., Klinkert M. Q. Sequence of Schistosoma mansoni cathepsin C and its structural comparison with papain and cathepsins B and L of the parasite. *Prot Pept Letts.* 1995; **2**:313–20.

[51] Brindley, P. J., Kalinna, B. H., Dalton, J. P., Day, S. R., Wong, J. Y., Smythe, M. L. *et al.* Proteolytic degradation of host hemoglobin by schistosomes. *Mol. Biochem. Parasitol.* 1997; **89**: 1–9.

[52] Bogitsh, B. J., Kirschner, K. F., Rotmans, J. P. Schistosoma japonicum: Immunoinhibitory Studies on Hemoglobin Digestion Using Heterologous Antiserum to Bovine Cathepsin D. *J Parasitol.* 1992; **78**:454-459

[53] Bogitsh, B. J., Kirschner, K.F. *Schistosoma japonicum*: ultrastructural localization of a hemoglobinase using mercury labeled pepstatin. *Exp Parasitol.* 1986; **62**: 211-215.

[54] McKerrow, J.H. Designing drugs for parasitic diseases of the developing world. *PLoS Med.* 2005; **2**: 10.1371/journal.pmed.0020210

[55] Barr, S.C., Warner, K.L., Kornreic, B.G., Piscitelli, J., Wolfe A, Benet, L., *et al*. A cysteine protease inhibitor protects dogs from cardiac damage during infection by *Trypanosoma cruzi. Antimicrob Agents Chemother.* 2005; **49**: 5160– 5161.

[56] Abdulla, M.H., Lim, K.C., Sajid, M., McKerrow, J. H., Caffrey, C. R. Schistosomiasis mansoni: novel chemotherapy using a cysteine protease inhibitor. *PLoS Med.* 2007; **4**: e14 10.1371/journal.pmed.0040014

[57] O'Donoghue, A. J., Eroy-Reveles, A. A., Knudsen, G. M., Ingram, J., Zhou, M. Statnekov, J. B. *et al.* Global identification of peptidase specificity by multiplex substrate profiling. *Nature.* 2012; **9**: 1095–1103.

[58] Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., Gevaert, K. Improved visualization of protein consensus sequences by iceLogo. Nat Meth. 2009; 6: 786–787.

[59] O'Donoghue AJ, Jin Y, Knudsen GM, Perera NC, Jenne DE, Murphy, J. E., *et al.* Global substrate profiling of proteases in human neutrophil extracellular traps reveals consensus motif predominantly contributed by elastase. *PLoS ONE*. 2013; **8**(9): e75141. 10.1371/journal.pone.0075141

[60] Julien, O., Zhuang, M. Wiita, A. P., O'Donoghue, A. J., Knudsen, G. M., Craik, C. S., Wells, J. A. Quantitative MS-based enzymology of caspases reveals distinct protein substrate specificities, hierarchies, and cellular roles. *PNAS* 2016; **14**: e2001-e2010. 10.1073/pnas.1524900113.

[61] Small, J. L., O'Donoghue, A. J. Boritsch, E. V., Tsodikov, A. V. Knudsen, G.M. Vandal, O., Craik, C. S., Ehrt, S. Substrate specificity of MarP, a perplasmic protease required for resistance to acid and oxidative stress in *Mycobacterium tuberculosis*. *J. Biol. Chem.* (2013) **288**: 12489-12499.

[62] Li, H., O'Donoghue, A. J., van der Linden, W. A., Xie, S. C., Yoo, E., Foe, I. T. *et al.* Structure- and function-based design of Plasmodium-selective proteasome inhibitors. *Nature.* 2016; **530,** 233–236.

[63] O'Donoghue, A. J., Knudsen, G. M., Beekman, C., Perry, J. A., Johnson, A. D., DeRisi, J. L., Craik, C. S., Bennett, R. J. Destructin-1 is a collegen degradting endopeptidase secreted by *Pseudogymnoascus destructans*, the causative agent of white-nose syndrome. *PNAS.* 2014; **112**: 7478-7483.

[64] Bibo-Verdugo, B., O'Donoghue, A. J., Rojo-Arreola, L., Craik, C. S., García-Carreño, F. Complementary proteomic and biochemical analysis of peptidases in lobster gastric juice uncovers the functional role of individual enzymes in food digestion. *Mar. Biotechnol.* 2016; **18**: 201–14.

[65] Dvořák, J., Fajtová, P., Ulrychová, L., Leontovyč, A., Rojo-Arreola, L., Suzuki, B. M., *et al.* Excretion/secretion products from *Schistosoma mansoni* adults, eggs and schistosomula have unique peptidase specificity profiles. *Biochimie.* 2016; **122**: 99–109.

[66] Corvo I., O'Donoghue A. J., Pastro L., Pi-Denis N., Eroy-Reveles A., *et al.* Dissecting the Active Site of the Collagenolytic Cathepsin L3 Protease of the Invasive Stage of *Fasciola hepatica. PLoS Negl Trop Dis* 2013; **7**(7): e2269. doi: 10.1371/journal.pntd.0002269

[67] Dvořák, J., Fajtová, P., Ulrychová, L., Leontovyč, A., Rojo-Arreola, L., Suzuki, B. M., *et al.* Excretion/secretion products from *Schistosoma mansoni* adults, eggs and schistosomula have unique peptidase specificity profiles. *Biochimie.* 2016; **122**: 99–109.

[68] Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., Mello, C. C. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans. Nature.* 1998; **391**: 806–811.

[69] Newmark P. A., Reddien, P. W, Cebria, F., Sanchez Alvarado A. Ingestion of bacterially expressed double-stranded RNA inhibits gene expression in planarians. *Proc Natl Acad Sci USA*. 2003; **100**: 11861–11865.

[70] Correnti, J. M., Brindley, P. J., Pearce, E. J. Long-term suppression of cathepsin B levels by RNA interference retards schistosome growth. *Mol. Biochem. Parasitol.* 2005; **143**: 209–15

[71] Pearson, B. J., Eisenhoffer, G. T., Gurley, K. A., Rink, J. C., Miller, D. E., Sánchez Alvarado, A. Formaldehyde-based whole-mount in situ hybridization method for planarians. *Dev. Dyn.* 2009; **238**: 443–50.

[72] Cogswell, A.A., Collins, J.J., Newmark, P.A., Williams, D.L. Whole mount *in situ* hybridization methodology for *Schistosoma mansoni*. *Mol Biochem Parasit*. 2011; **178**:46-50.

Chapter 2 The Proteolytic Repertoire of the Planarian *S. mediterranea*

2.1 Introduction

The family *Platyhelminthes* contains an estimated 25,000 species, including the free-living planaria *Schmidtea mediterranea*. This freshwater flatworm has been an experimental model for over a century due to its regenerative capacity. Worms can fully regenerate missing body tissues from fragments as small as <1/200th of their original size.[1] This is due to the large population of stem cells, or neoblasts, that make up approximately 30% of all adult tissues in the worm.[2] The *S. mediterranea* genome has been sequenced and these worms are highly amenable to the use of RNAi, making them an ideal flatworm for study.[3,4]

While *S. mediterranea* is commonly used as a model for regeneration, it has recently been proposed as a model for parasitic flatworms, much like the free-living *Caenorhabditis elegans* is used as a model for other parasitic nematodes.[5] Over half of the known members of *Platyhelminthes* are human or veterinary parasites, including worms of the genus *Schistosoma*, the causative agents of the disease schistosomiasis. Schistosomiasis affects more than 200 million people worldwide and is the second most common parasitic disease behind malaria.[6] However, these and other parasitic worms are difficult to study due to their complex, multi-host life cycles and their resistance to biochemical tools such as RNA interference (RNAi).

S. mediterranea and the parasite *S. mansoni* share several similarities in their reproductive systems, protonephridia, and stem cell populations.[7] Both are triploblastic,

bilaterally symmetric metazoans and contain nervous, digestive, and secretory systems.[8] In addition, these organisms share over 85% of their predicted proteome, including several families of proteases.[9] For digestion, *S. mansoni* and *S. mediterranea* have a blind-end, bifurcated gut that distributes digested food throughout the body.[10] Adult female *S. mansoni* worms will utilize several distinct proteases to rapidly digest red blood cells at a rate of 500 per minute.[11] While vertebrates rely on serine proteases from the trypsin family (clan PA) for protein digestion, invertebrate helminths generally use cysteine proteases from the clan CA (papain-like proteases) and aspartic proteases from the clan AA. Cysteine and aspartyl proteases are also key to digestion in other invertebrates like nematodes and arthropods. It appears that the preference for digestive serine proteases occurred during the evolution of arthropods or mollusks.[12,13,14]

Three cysteine cathepsin proteases perform the majority of digestive function in helminths: cathepsins B, C, and L. These clan CA proteases are found in many flatworm parasites, including the trematodes *Fasciola hepatica, Clonorchis sinesis*, and *Opisthorchis viverrini*.[11,15] Immunohistochemistry suggests that cathepsins B, C, and L are associated strongly with the gastrodermis, vitellaria, and tegument.[15] These proteases work in concert to degrade hemoglobin and albumin in the acidic helminth gut.[16] Inhibition of cysteine proteases has been shown to kill parasites both *in vitro* and *in vivo*, suggesting that these proteases are important for worm viability.[17,16] Previous work has shown that inhibition of schistosome cysteine cathepsin protease activity in infected mice has lead to a reduced worm and egg burden and an improvement in organ pathology.[18]

Schistosoma flatworms can be induced to regurgitate their gut contents and the cysteine proteases cathepsin B, C, and L, as well as an aspartyl protease (cathepsin D) and an asparaginyl endopeptidase (legumain), have been identified.[19,10] Cathepsin B1, or SmCB1, represents the most abundant cysteine peptidase activity measurable in both adult schistosomes and gastrointestinal content (GIC) extracts. Along with degrading host hemoglobin and albumin, SmCB1 has been shown to degrade several immunoglobulins *in vitro*, suggesting possible roles in immune evasion.[20] Furthermore, SmCB1 has been suggested as a drug target, a potential vaccine target, and a serodiagnositc marker.[21,22] Although schistosomula larvae with cathepsin B reduced by RNAi were still viable, these worms showed a significant decrease in growth compared to control groups.[23] This suggests that SmCB1 plays an important role in digestion such that a lack of activity has a negative effect on the acquisition of nutrients for growth.

Very little is known about the function of proteolytic enzymes in planaria and if conservation in enzymatic function exists between parasitic and free-living flatworms. Given the important role played by *S. mansoni* gut proteases in digestion, we investigated the role of these enzymes in the gut of *S. mediterranea*. Using proteomics and a global protease substrate profiling method, referred to as multiplex substrate profiling by mass spectrometry (MSP-MS), we identified and detected active proteases in *S. mediterranea* gastrointestinal contents. Using RNAi and specific protease inhibitors, we probed the function of several protease families *in vivo* to determine the roles of these enzymes in protein digestion. We further examined a cathepsin B-like cysteine protease and its localization in *S. mediterranea*. We hypothesized that planarians utilize cysteine proteases

21

for digestion and that they perform similar roles in free-living flatworms as in parasitic helminths.

2.2 Results

2.2.1 Several protease families are conserved between *S. mediterranea* and the parasite S. mansoni.

A preliminary BLAST search using sequences from previously identified *S. mansoni* proteases against the *S. mediterranea* database (http://smedgd.neuro.utah.edu) revealed that several protease families are conserved between both parasitic and free-living flatworms (Table 1). Among the conserved protease families were aspartyl and cysteine cathepsin proteases, including cathepsin B, D, L1, and L2/L3, all of which are involved in blood feeding and found in the gut of in *Schistosoma* worms. The aspartyl proteases cathepsin D had two putative homologs, while the cysteine protease cathepsin L1 and L2/L3 each had several putative homologs. Unlike *S. mansoni*, only one gene corresponding to cathepsin B was found in *S. mediterranea*. Interestingly, two well-characterized *S. mansoni* proteases were absent in the *S. mediterranea* genome. These enzymes are the serine protease, cercarial elastase, which is involved in skin invasion, and the gut-associated asparaginyl endopeptidase called legumain.[10]

Table 1.   Homologous *Schistosoma mansoni* proteases found in *Schmidtea mediterranea*.

Protease	Present in *Schimdtea*	Number of Homologs	Function in *S. mansoni*	Localization in *S. mansoni*
Cathepsin B1	Yes	1	Blood feeding	Gut (cecum)
Cathepsin B2	Yes	1	Unknown	Tegument
Cathepsin D	Yes	2	Blood feeding	Gut
Cathepsin L1 (F)	Yes	3	Blood feeding	Gut
Cathepsin L2	Yes	17	Blood feeding	Gut
Cathepsin L3	Yes	20	Blood feeding	Gut
Leucine Aminopeptidase (1&2)	Yes	5	Hatching and unknown gut function	Gastrodermis
Kallikrein	Yes	2	Unknown; suggested to be modulation/evasion of host immune system	Schistosomula released products and male dorsal tubercles
M8 (Invadolysin)	Yes	8	Unknown	Acetabular glands
Legumain (SmAE, asparaginyl endopeptidase)	No	0	Unknown; suggested to be hemoglobin degradation	Gut (cecum)
SmCE (cercarial elastase)	No	0	Invasion through skin	Acetabular glands and pre-acetabular glands

BLAST comparisons of known *S. mansoni* proteases with the *S. mediterranea* genome database (http://smedgd.neuro.utah.edu/). Some proteases, like those in the cathepsin (Clan CA, papain) family, were conserved with 1 or more homologs. As expected, the cercerial protease used by schistosomes for skin invasion is absent. Surprisingly, there appears to be no legumain (SmAE) homolog, which is found in many helminths, including *S. mansoni*.

2.2.2 Cysteine proteases represent the major proteolytic activity in *S. mediterranea* and its gut.

To identify and characterize intestinal tract proteases in *S. mediterranea*, worms were placed in 3% ethanol to induce regurgitation. These worms were starved for a week to avoid contamination by proteases from the *S. mediterranea* food source. Protease activity in the worm regurgitant was assessed using a mixture of internally quenched fluorescent substrates. These substrates consist of 7-mer to 10-mer peptide sequences flanked by 7-methoxycoumarin on the amino terminus and dinitrophenol conjugated to lysine on the carboxyl terminus. Protease activity from replicate *S. mediterranea* tanks was assayed in pH 3.5, 5.5 and 7.5 buffer. Activity was highest at pH 3.5 and lowest at pH 7.5. No activity was detected in the water from worms that were not exposed to ethanol (Fig 1A). The increase in protease activity due to ethanol treatment proved that we had successfully enriched for the gut proteases of *S. mediterranea*. The regurgitant from the replicate tanks was pooled and the substrate specificity profile was uncovered using an unbiased substrate profiling assay consisting of 124 physiochemically diverse peptides of 14 residues each.[24] Cleavage of these peptides was detected by LC-MS/MS sequencing after incubation of regurgitant with the peptide mixture at pH 3.5, 5.5 and 7.5. An aliquot of each reaction was removed after 15, 60, 240, and 1200 minutes. Proteases active at pH 3.5 and 5.5 cleaved at more sites than the proteases that were active in the pH 7.5, indicating a broader substrate specificity (Fig 1B). Proteases active in the pH 3.5 buffer had a preference for all hydrophobic residues except Val and Pro at most positions between P3 and P4′ (Fig 1C). In the pH 5.5 buffer, hydrophobic residues were also found at high
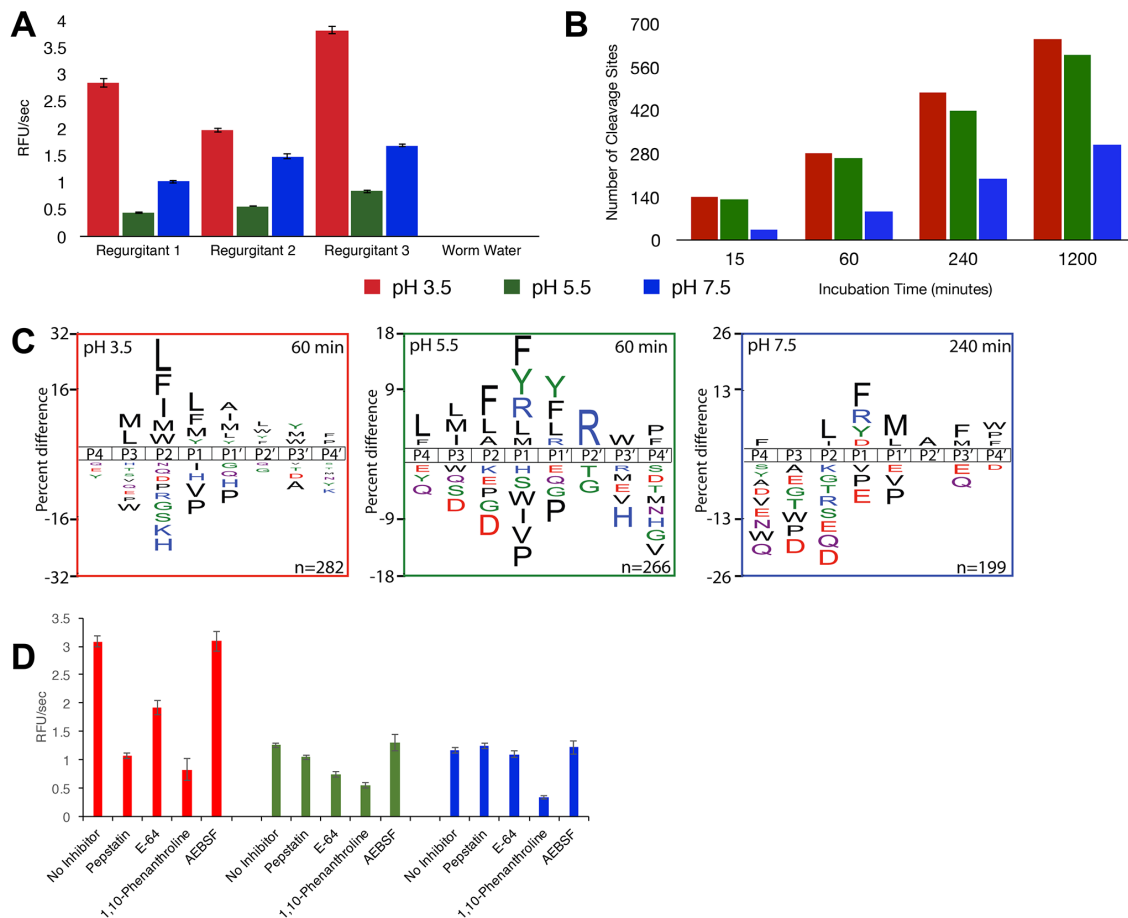
Figure 1: Incubation of worm regurgitant with fluorescent substrates and MSP-MS reveals highest activity at low pH and is dependent on aspartyl, cysteine, and metalloproteases. (A) Replicates of *S. mediterranea* regurgitant were incubated with a library of fluorescent substrates at pH 3.5, 5.5, and 7.5. Each sample showed the highest rate of cleavage at pH 3.5. No detectable activity was found in samples without regurgitant. Fluorescence was measured in relative fluorescence units (RFU) per second. **(B)** The regurgitant was incubated with peptide substrates for MSP-MS and the total number of cleavage sites was detected after 15, 60, 240, and 1200 minutes of incubation with the peptide mixture at pH 3.5, 5.5, and 7.5. **(C)** An iceLogo generated from the pattern of cleavage events after 60 minutes reveals the specificity of protease activity at pH 3.5, 5.5, and 7.5. Amino acids that are most frequently observed at each position (P4-P4') are shown above the axis, while less frequently observed amino acids are shown below. **(D)** The pooled samples of regurgitant were assayed with the same fluorescent library in the presence of several inhibitors, including pepstatin (aspartyl protease inhibitor), E-64 (cysteine protease inhibitor), 1,10-phenanthroline (metalloprotease inhibitor), and AEBSF (serine protease inhibitor). While the effect of inhibition of astpartyl, cysteine, and metalloproteases varied according to pH, inhibition of serine proteases had no effect on the overall amount of activity detected at any pH.

frequency in most sub-sites, except for P2′, which prefers Arg. In addition, Tyr and Arg are well tolerated at P1 and P1′. Unlike the cleavage sites detected in the pH 3.5 and 5.5 assays, protease activity at pH 7.5 was dominated by exo-peptidases. After 15 minutes incubation, 30 out of 36 cleavage sites detected occurred at the amino or carboxyl termini of the 14-mer peptides. After 60 minutes incubation, 68 out of the 94 cleavage sites had occurred at the termini and the overall substrate preference was for arginine at P1 and norleucine at P1′. Taken together, these data indicate that multiple proteases are present in the gut of *S. mediterranea* that are capable of cleaving diverse set of peptide bonds over a broad pH range.

In order to better understand the proteases responsible for cleaving the peptide substrates, we incubated *S. mediterranea* regurgitant with the class specific protease inhibitors pepstatin-A, E-64, 1,10-phenanthroline, and AEBSF. These compounds are standard inhibitors of aspartyl, cysteine, metallo-, and serine proteases, respectively. Under all conditions tested, AEBSF failed to inhibit activity, indicating that there were no active serine proteases present in the regurgitate (Fig 1D). At each pH, 1,10-phenanthroline reduced activity by more than 50%, while pepstatin and E-64 inhibited protease activity at pH 3.5 and pH 5.5 only.

To identify the specific proteins responsible for the aspartyl-, cysteine- and metallo-protease activity in the *S. mediterranea* regurgitant, we performed a proteomic analysis on the regurgitant proteins. Proteins were digested with trypsin, and the resulting peptides were sequenced with liquid chromatography-tandem mass spectrometry (LC-MS/MS). These peptides were compared to a database of putative proteins from the asexual *S.*

mediterranea strain (http://smedgd.stowers.org/downloads/). Peptides were ranked based on their intensity sums. We identified a total of 122 proteins, including 7 proteases (Supplemental Table 1). Based on peptide count, the two metalloproteases, astacin-2 and astacin-5 appear to be the most abundant proteases in the regurgitant. Recombinant astacin from the parasitic nematode, *Teladorsagia circumcincta* was expressed in *E. coli* and found to be sensitive to 1,10-phenanthroline.[25] Therefore, it is likely that astacin-2 and -5 are the source of the 1,10-phenanthroline-sensitive activity in *S. mediterranea* regurgitant. We also found that one regurgitant sample contained three other metalloproteases: an M12 protease, astacin 1, and two M10 matrix metalloproteinases, mmp1 and mmp2. Although our inhibition studies determined that there was one or more pepstatin-sensitive aspartyl protease in the gut, we were unable to identify this protease by mass spectrometry; it is possible that aspartyl proteases are in lower abundance than metallo- or cysteine proteases. However, we were able to detect two clan CA cysteine proteases; a cathepsin B-like and a cathepsin L1 protease that are likely to be the E-64 sensitive enzymes in the pH 5.5 assay.

2.2.3 Inhibition of cysteine and aspartyl proteases, but not metalloproteases, blocks digestion in *S. mediterranea.*

As noted above, cysteine, aspartyl, and metalloproteases are all active in the regurgitant of *S. mediterranea* and are all likely to play important yet distinct roles in protein digestion. To investigate the multi-enzyme model of digestion, we utilized a chemical knockdown strategy using protease inhibitors to monitor the effect of these protease classes. Starved worms were fed beef liver coated with rhodamine-labeled bovine

serum albumin (Rh-BSA), which has been used previously to assay digestive proteases in *S. mansoni.*[10] Fluorescence generated by cleavage of quenched rhodamine-BSA molecules is proportional to the activity of digestive enzymes; lower fluorescence indicates a decrease in digestion (Fig 2A). Measurements are represented as fluorescence over worm area to correct for differences in worm size.

We first examined the role of cysteine proteases, specifically the two identified in the worm regurgitant: cathepsin B and cathepsin L. While E-64 is the standard pan-cysteine protease inhibitor *in vitro*, its poor cell-permeability limits its effects *in vivo*.[26] However, the vinyl sulfone K11777 is a potent chemical inhibitor of both cysteine proteases cathepsin B and L, is cell-permeable, and has been shown to reduce the parasite burden of schistosomes in mice.[27,18] Chemical inhibition of cysteine cathepsin proteases by K11777 strongly reduced digestion in the planaria gut. Worms pretreated with K11777 were fed RhBSA and fluorescence was imaged as a measurement of proteolytic activity. K11777 treated worms exhibited a 38% reduction in fluorescence compared to untreated controls (Fig 2B, E). This decrease in signal persisted over time, and twenty-four hours post feeding treated worms had a a 42% decrease in signal.

As one or more pepstatin-sensitive aspartic proteases are present in the *S. mediterranea* regurgitant, we used this inhibitor to determine the role of these enzymes in protein digestion. Worms treated with pepstatin resulted in a 16% reduction in RhBSA activity immediately after feeding compared to the untreated worms. After 24 hours, only
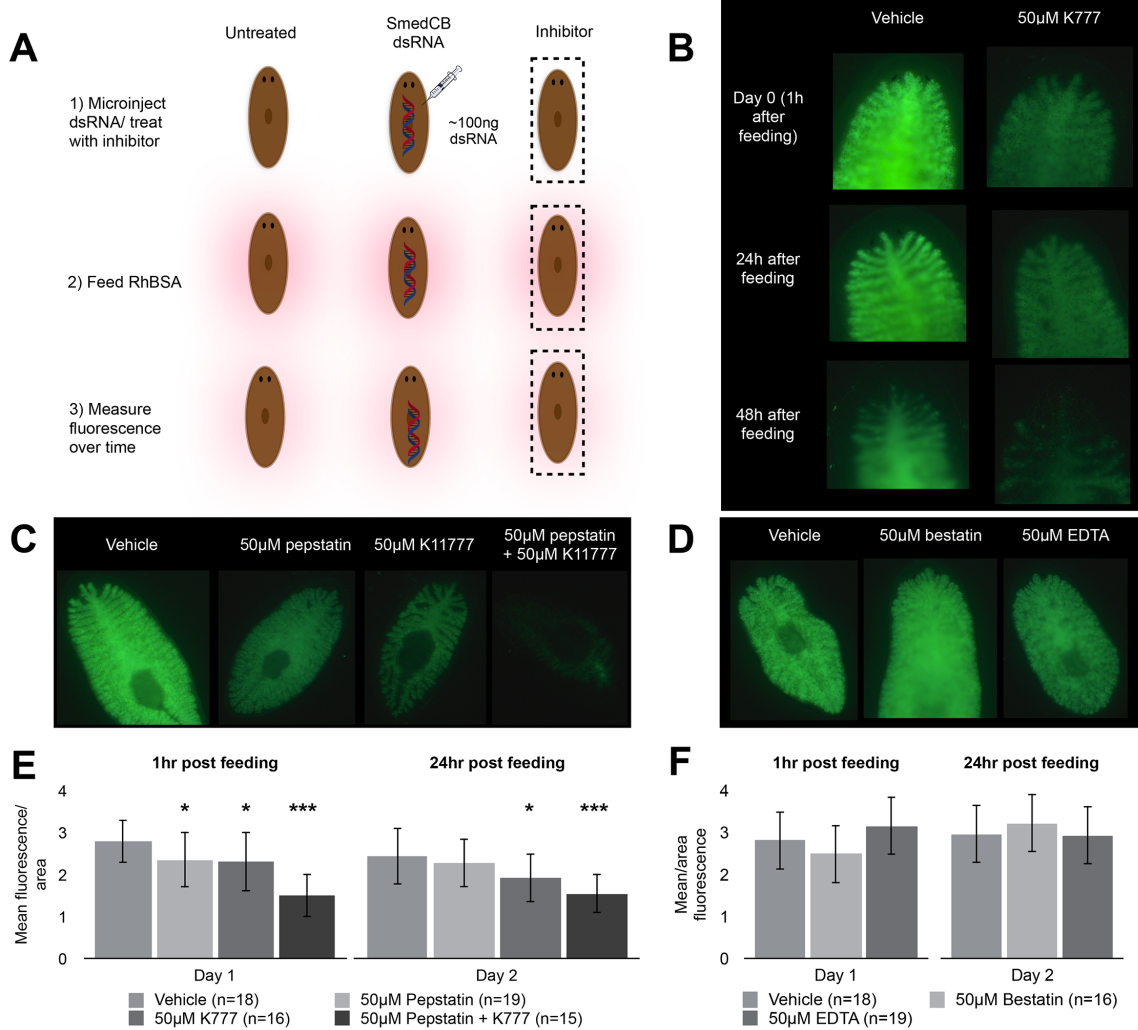
Figure 2: Inhibition of cysteine, aspartyl, but not metalloproteases, inhibits planaria digestion. (A) Animals were treated with 50μM protease-specific chemical inhibitors or dsRNA prior to feeding with rhodamine-labeled bovine serum albumin (RhBSA) and imaging. Fluorescence, due to cleavage of quenched albumin, was imaged **(B).** These representative images show a decrease in fluorescence in worms treated with K11777. Inhibition of aspartyl proteases by pepstatin as seen in **(C)** also decreased digestion. Combined inhibition of cathepsin B, L, and D completely reduced fluorescence. In contrast, inhibition of metalloproteases with bestatin and EDTA **(D)** did not reduce fluorescence. These data are quantified in **(E)** for K11777 and pepstatin and show that the reduction in fluorescence is significant in the presence of these inhibitors. **(F)** Treatment with metalloprotease inhibitors bestatin and EDTA did not lead to a significant change in fluorescence.

a 6% decrease in activity was evident. However, co-treatment with pepstatin and K11777 had a dramatic effect on digestion. Treated worms were on average 46% and 37% less bright than vehicle treated worms after one and twenty-four hours, respectively (Fig 2C, F). The representative image in Figure 2C shows that very little fluorescence, indicative of albumin degradation, was observed in worms where both cysteine and aspartic proteases were inhibited.

Due to the presence of metalloproteases in the worm regurgitant as documented by biochemical assays and the proteome analysis, 1,10-phenathroline was used to examine the effect of reducing the metalloprotease activity. However, 1,10-phenanthroline was highly toxic to the worms and therefore two alternative metalloprotease inhibitors, bestatin and EDTA, were used for the *in vivo* digestion studies. Representative images in Figure 2D show that no decrease in fluorescence due to albumin degradation was detected; therefore, digestion of rhodamine-labeled albumin was unaffected by metalloprotease inhibitors (Fig 2D, F).

2.2.4 Knockdown of SmedCB inhibits digestion.

The reduction in protein digestion in *S. mediterranea* that occurs in the presence of a combined treatment of a cysteine and aspartyl protease inhibitor confirms that these proteases are the major digestive enzymes in the worm gut. Although we could not confirm the exact aspartyl protease present, we predicted that this enzyme was a cathepsin D-like protein found in the genome of planaria. Primer pairs were designed against the mRNA sequence of cathepsin D and the two cysteine proteases, cathepsin B and L. We first used

30

RT-PCR to quantify the levels of cysteine and aspartyl proteases over time following starvation. Worms starved over a three-week period showed a 50-65% reduction in cathepsin D levels and a 65-75% reduction in cathepsin B mRNA levels while cathepsin L levels were reduced by only 20% (Fig 3A). These data suggest that cathepsin B and D expression is regulated by food intake while cathepsin L expression is less affected by changes in feeding conditions.

To examine the specific role(s) played by SmedCB, we used RNAi to assess whether loss of function was detrimental to worm viability and survival. Worms were first starved for two weeks before injection with ~100ng SmedCB dsRNA for three consecutive days. This protocol resulted in 80% reduction in SmedCB mRNA levels after fourteen days when compared to untreated worms (Fig 4A). This RNAi knockdown was specific to SmedCB and did not significantly impact mRNA levels of related proteases cathepsin L or cathepsin D (Fig 3B). Using the standard cathepsin B fluorescent substrate, z-Arg-Arg-*AMC*,[28] protease activity was reduced in worm lysates by 81% fourteen days after SmedCB mRNA knock-down. (Fig 4B). Activity using this substrate was confirmed to be derived from cathepsin B because CA-074, a highly selective cathepsin B inhibitor, reduced total activity by 71% (Fig 5). In addition, protein hydrolysis using rhodamine-labeled albumin was reduced by 20% for 48 hours after feeding (Fig 4C). Taken together, these data confirm that SmedCB specifically plays a central role in protein digestion in *S. mediterranea*, much like cathepsin B1 does in *Schistosoma mansoni*.
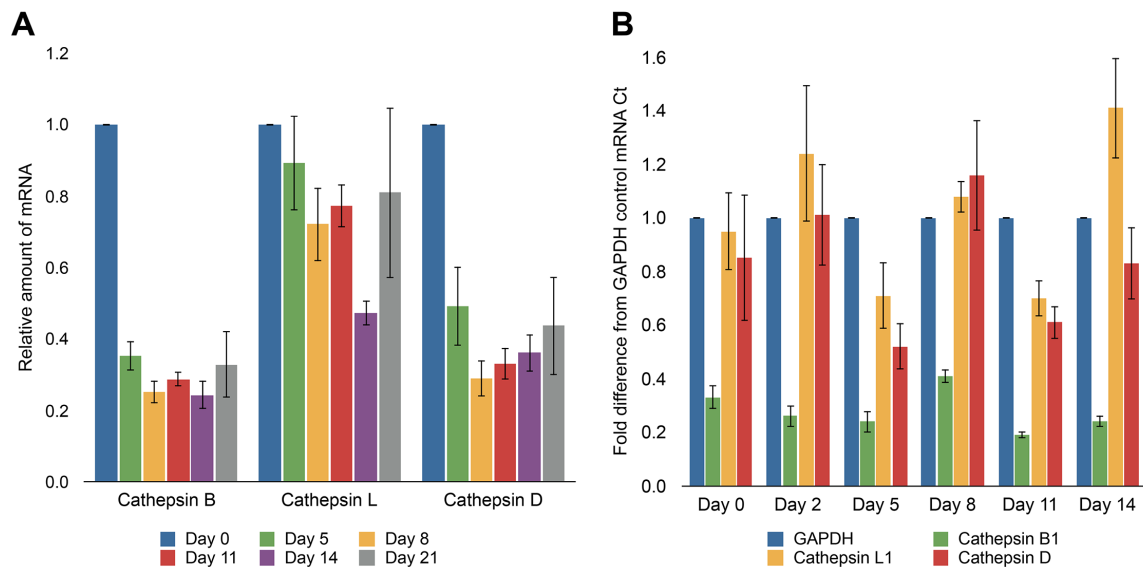
Figure 3: *S. mediterranea* **cathepsin protease mRNA is reduced after starvation while dsRNA treatment with SmedCB is specific to cathepsin B. (A)** RNA was extracted from three sets of 10 to 15 worms starved for three weeks. mRNA levels for three major cathepsin protease genes (B, L, and D) were quantified using RTPCR; levels were compared to recently fed worms (day 0) to measure the relative change in transcript levels over time. Starved worms demonstrated a 65-75% reduction in cathepsin B mRNA levels and 50-65% reduction in cathepsin D mRNA. Cathepsin L mRNA was reduced, on average, 20% with the exception of day 14, which had a 50% reduction. **(B)** Worms injected with ~100ng of SmedCB dsRNA for three consecutive days, followed by amputation (day 0) and three weeks of starvation, showed a marked reduction of SmedCB mRNA when analyzed via RTPCR using GAPDH mRNA levels as a baseline control. SmedCB levels decreased by ~80% after two weeks. Cathepsin L and D showed some fluctuation in mRNA levels, but remained within 50% of baseline levels and there was no continued trend observed over time. This indicates that fluctuations in levels are due to variations in worm populations chosen for RNA extraction rather than significant off-target effects of SmedCB dsRNA.
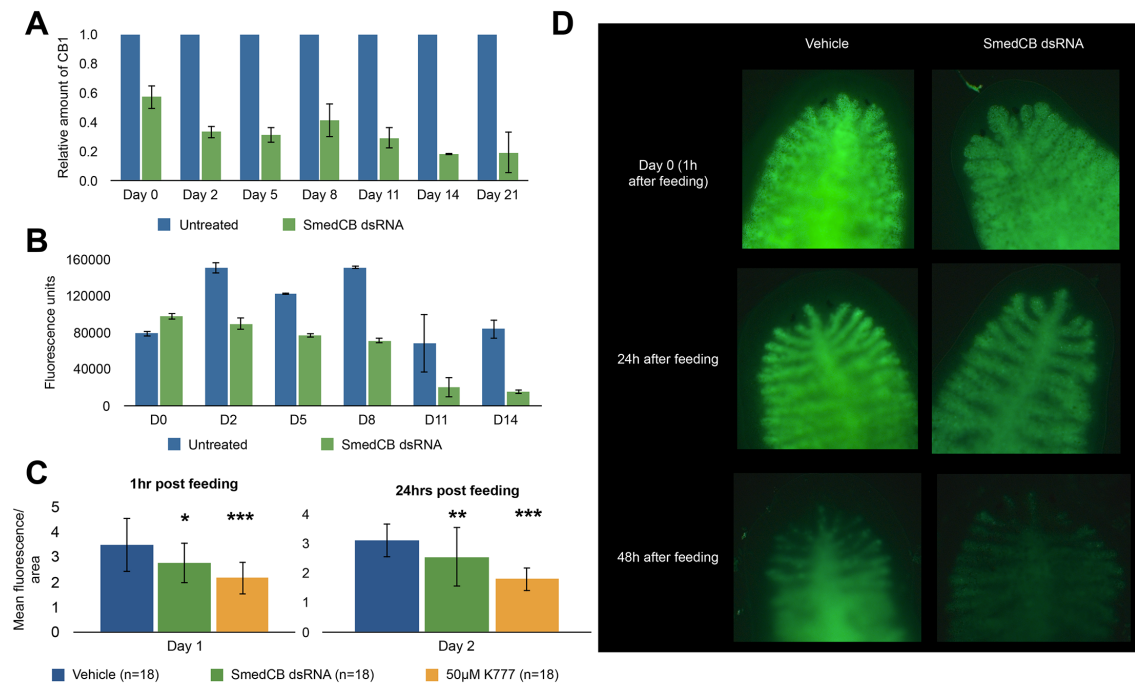
Figure 4: SmedCB RNAi treatment reduces cathepsin B mRNA levels and proteolytic activity, inhibiting digestion. A) Worms injected with ~100ng of SmedCB dsRNA for three consecutive days, followed by amputation (day 0) and three weeks of starvation, showed a marked reduction of SmedCB mRNA when analyzed via RTPCR in comparison to untreated worms. The initial three days of injection prior to day 0 causes a 45% reduction that increases to over 80% after two weeks of starvation during regeneration. SmedCB dsRNA treated worms also showed a marked reduction of SmedCB protease activity via Z-RR-AMC cleavage **(B)**. Treated worms exhibited between 50-75% less fluorescence from Z-RR-AMC cleavage than untreated worms over two weeks. **C)** Fluorescence from digestion of Rh-BSA is decreased in worms treated with SmedCB dsRNA. This decrease is less dramatic than treatment with the cathepsin B and cathepsin L inhibitor K11777, but is statistically significant. Representative images of this fluorescence can be observed in **D**.
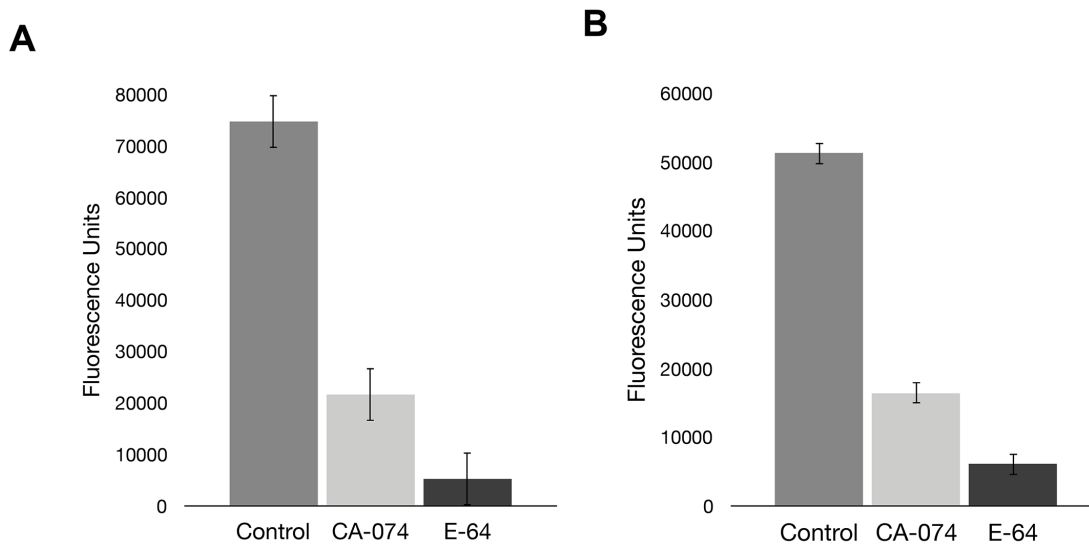
Figure 5: Total fluorescence of Z-RR-AMC and Z-FR-AMC cleavage by planaria lysate is reduced by inhibition of cathepsin activity. (A) Whole worm lysate was preincubated with 50μM E-64, a pan-cysteine protease inhibitor, before assaying with the fluorescent probe Z-RR-AMC. Pretreatment with E-64 reduced fluorescence by 93%. Inhibition of cathepsin B protease alone by CA-074 caused a 71% decrease in activity, indicating that the majority of proteolytic cleavage is caused by cathepsin B-like proteases. **(B)** Similar effects were observed when Z-FR-AMC, a probe cleaved by both cathepsin B and cathepsin L proteases, was used.

2.2.5 SmedCB is localized to the worm gut.

Schistosoma mansoni expresses two isoforms of cathepsin B; SmCB1 is found in the gut and performs a digestive function, while SmCB2 is localized to the tegument where it serves an unknown purpose. We hypothesized that SmedCB would also be found in the gut of *S. mediterranea*, as indicated by its presence in the worm vomit, and would perform a role in digestion.

In order to determine the localization of SmedCB in *S. mediterranea,* we first needed to confirm that this protein was detectable in a whole worm lysate. Using the activity based probe DCG-04,[29] which specifically targets cysteine proteases, we observed a biotinylated band on an SDS-PAGE gel at 27 kDa. Mass spectrometry sequencing of the excised band confirmed that its identity as SmedCB (Supplemental Table 2). Knowing that SmedCB was present in whole worm lysate and not just in the regurgitation, we developed antibodies that could detect procathepsin B and mature cathepsin B. Peptides encoding a region of the propeptide, as well as the catalytic domain, were synthesized for antibody development (Fig 6A). Both antibodies specifically labeled SmedCB in the planaria lysate (Fig 6B). Immunohistochemistry using the zymogen SmedCB antibody labeled cells surrounding the intestinal lumen (Fig 6C, D). Localization of the catalytic region of SmedCB, found in both the full-length zymogen as well as the catalytically activated protein, was performed using an antibody generated against recombinant protein expressed in *Escherichia coli*. Immunohistochemistry showed that mouse antisera against the catalytic domain of SmedCB also labeled cellular vesicles surrounding the intestinal lumen (Fig 6E, F). Length of starvation did not effect localization of SmedCB (Fig 6G).

A

MKIILISFVLLFSLVNC**DRTKWESLTKSVKG**QIPIHTPLSFDLINYAQTTWKAGPT
TRFQTVSDIRKVLGVLKDPNNFKLPKKSSGLNAIR**LPASFDSRIQWPKC**KSINEI
RDQSNCGS**C**WAFGAVEAMTDRYCIHSNGTKNPRISAEDLLTCCGFRCGDGCN
GGFPSGAWSFWVSHGLVTGGEYGAHLGCQDYAFPKCSHHVIGPYPNCTGEF
PTPKCKKTCQA**GYAKTYKEDKQFGKT**SYSVGGREEEIMQEILTNGPVEAAFSV
YADFPNYKSGVYQHVSGAMLGG**H**AIKILGWGVENNTPYWLVA**N**SWNPTWGD
KGYFKIKKGTNECGIEDEVVAGIPKL

Underlined: Protein sequence used in mouse antibody generation against catalytic CB1
Pink: Synthesized peptide used for rabbit antibodies against catalytic CB1
Green: Synthesized peptide used for rabbit antibody against zymogen CB1
Red: Catalytic triad (Cys-His-Asn)

Figure 6: Immunofluorescence labeling and electron microscopy of zymogen and catalytic SmedCB shows strong labeling in cells lining the intestinal lumen. A) Several antibodies were generated against *Schmidtea* cathepsin B, including one against the catalytic region only (underlined region), and one against the zymogen tag and catalytic region (a combination of the green and pink peptides). **(B)** Western blot with antibody used to localize SmedCB (1:5,000). Lane 1, recombinant prep of SmedCB from *E. coli*. Band at 70kDa is a dimer of SmedCB, lower band is monomeric SmedCB. Lanes 2-5 contain planaria lysate at 15µL, 7.5 µL, 5µL, and 15µL, respectively. Higher band (~37kDa) is zymogen, lower band (~27kDa) is catalytic domain. Lanes 1-4 use the mouse-derived antibody underlined in **A**, while lane 5 was blotted against the rabbit-derived antibody from the pink and green peptides in **A**. **(C)** Labeling of paraffin-embedded *S. mediterranea* cross sections of worms starved for one week with (1:100) anti-SmedCB zymogen pro-peptide antibody showed strongest signal in the cells lining the intestinal lumen. Comparison with the negative control **(D)** anti-rabbit secondary (1:100) only shows that the labeling on the border of the planaria cross sections is due to drying artifact. This pattern is also seen in worms with the anti-SmedCB catalytic domain only antibody (1:100) **(E)**. The negative control **(F)** with anti-mouse secondary only (1:100), shows labeling only on the edges of the cross sections due to drying artifact. Worms were also starved for two weeks prior to labeling **(G)** as opposed to one week in **C** and **E**. SmedCB labeling remained the same.

36

Figure 6 continued: Immunofluorescence labeling and electron microscopy of zymogen and catalytic SmedCB shows strong labeling in cells lining the intestinal lumen

(H) Cross sections of *S. mediterranea* worms (starved one week) were labeled with (1:100) anti-SmedCB zymogen pro-peptide antibody. Intestinal lumen seen as white space in the center of the image. **(I)** Immuno-gold labeling of antibody target, visualized as black dots, shows vesicle-like structures are heavily labeled. Other tissue, like the rounded, gray lipid droplets in **(J)** did not show labeling.

Immunohistochemistry using this antibody labeled cells surrounding the intestinal lumen (Fig 6C, D). Previous histological analysis of planaria had identified two types of intestinal cells: "phagocytes" that absorb food for intracellular digestion, and secretory "goblet cells" that release digestive enzymes into the intestinal lumen.[30,31,32] Electron microscopy revealed that the proSmedCB localizes to vesicular structures within these intestinal cells (Fig 6H, I) but not neighboring cells, like lipid droplets (Fig 6J).

SmedCB localization was also confirmed via whole-mount *in situ* hybridization (WISH) to identify SmedCB RNA expression. The WISH protocol previously established,[33] and modified,[34] was used to maximize signal sensitivity. Planaria starved for four or eight days prior to sample preparation showed similar expression patterns despite the difference in worm feeding (Fig 7A, B). *SmedCB* is highly expressed throughout the branched intestine of the worm, confirming the labeling seen with immunohistochemistry. Furthermore, there appear to be punctae of SmedCB labeling in the mesenchyme, especially in the head where the gut signal is less pronounced. In regenerating worm fragments, labeling of the growing branched intestine as well as the punctae seen in intact worms is observed, although there are several differences in *SmedCB* expression. Planaria were cut into three segments: heads, tails, and pharynxes. Amputations were performed three, five, and seven days prior to treatment to assess the changes in *SmedCB* expression during regeneration. Head segments exhibit concentrated labeling near the blastema, the region of newly regenerated body tissue, throughout the time course. (Fig 7C). Regenerating pharynxes show some increased signal near the blastema, but this is only seen on worms that have been recently amputated. Worms that have healed over five and
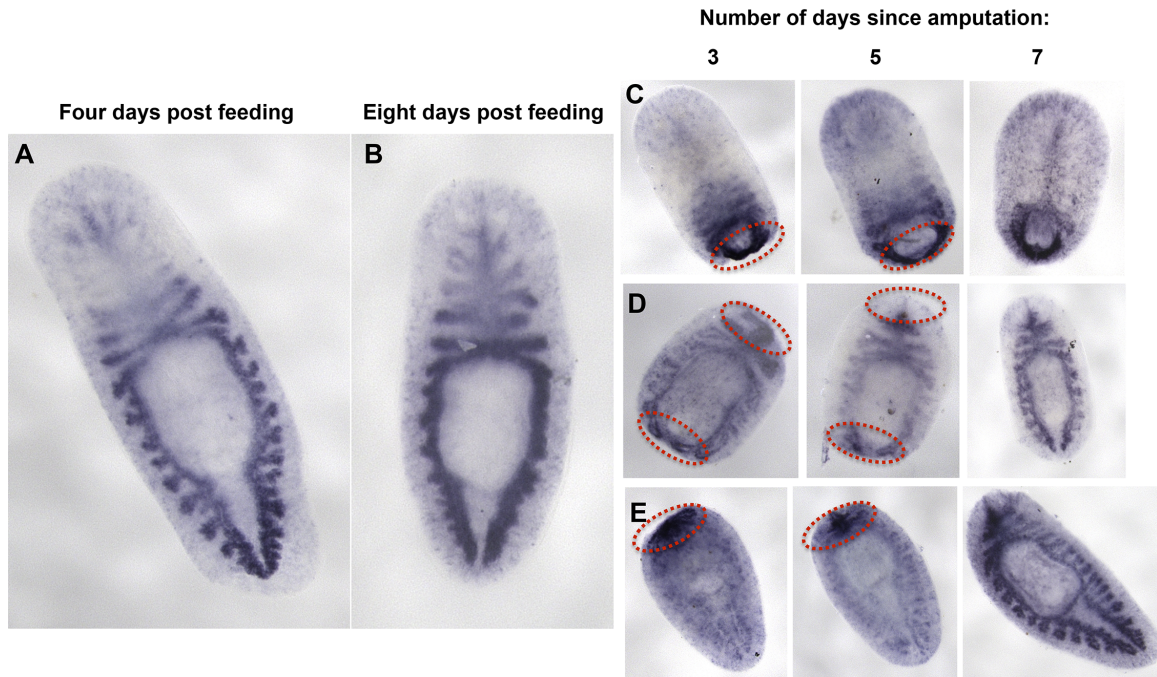
Figure 7: Whole-mount *in situ* hybridization of *SmedCB* shows labeling throughout the worm gut in intact and regenerating animals. An *in situ* DIG-12-UTP riboprobe against *SmedCB* shows labeling throughout the branched intestine. This pattern is observed in worms fixed four days after feeding **(A)** and eight days after feeding **(B)**. For regenerating worms, animals were fixed three, five, or seven days after amputation. Some regenerating animals have strong labeling near the (newly regenerated tissue, circled in red) as well as the newly-forming gut. Amputated heads **(C)** have labeling in the blastema up to seven days after amputation, while pharynxes **(D)** only show labeling near the blastema for three days. Tails **(E)** exhibit an intermediate phenotype with blastema labeled five days after amputation. All animals were treated with 1:100 dilution of the same riboprobe.

seven days no longer have an intense level of *SmedCB* near the blastema sites (Fig 7D). The blastema labeling in tail segments is observed at three and five days after amputation, but *SmedCB* appears to be localized to the same area as intact worms after a week of regeneration: the branched intestine (Fig 7E).

2.2.6 Inhibition of SmedCB and cathepsin L slows *S. mediterranea* regeneration.

Because of the interest in the regenerative ability of *S. mediterranea*, we also examined whether SmedCB might be required for regeneration as well as digestion. Another protease, transmembrane matrix-metalloproteinase A (Smed mt-mmpA), has been shown to modulate cell migration and delay new tissue growth.[35] Cathepsin B may modulate tissue growth as well given its localization to the regenerating blastema during *in situ* hybridization. We first tested whether SmedCB activity changed after feeding and amputation by incubating worm lysate with the activity-based probe DCG-04. This probe contains an electrophilic "warhead", which becomes covalently attached to the nucleophilic residue of cysteine proteases. Specificity of the probe for various protease targets is achieved via a linker region between the warhead and a biotin tag, which is used to visualize protease labeling.[36] Following digestion, an increase in SmedCB labeling was observed for 2-6 days before returning to baseline during worm starvation, suggesting that cathepsin B was active when food was present, but inactive once digestion was finished (Fig 8A). Interestingly, a similar trend was observed when worms were amputated following digestion (Fig 8B). *S. mediterranea* worms take 7-8 days to fully regenerate; high SmedCB activity during this period suggests a possible use for cathepsin B during worm growth and regeneration.

40
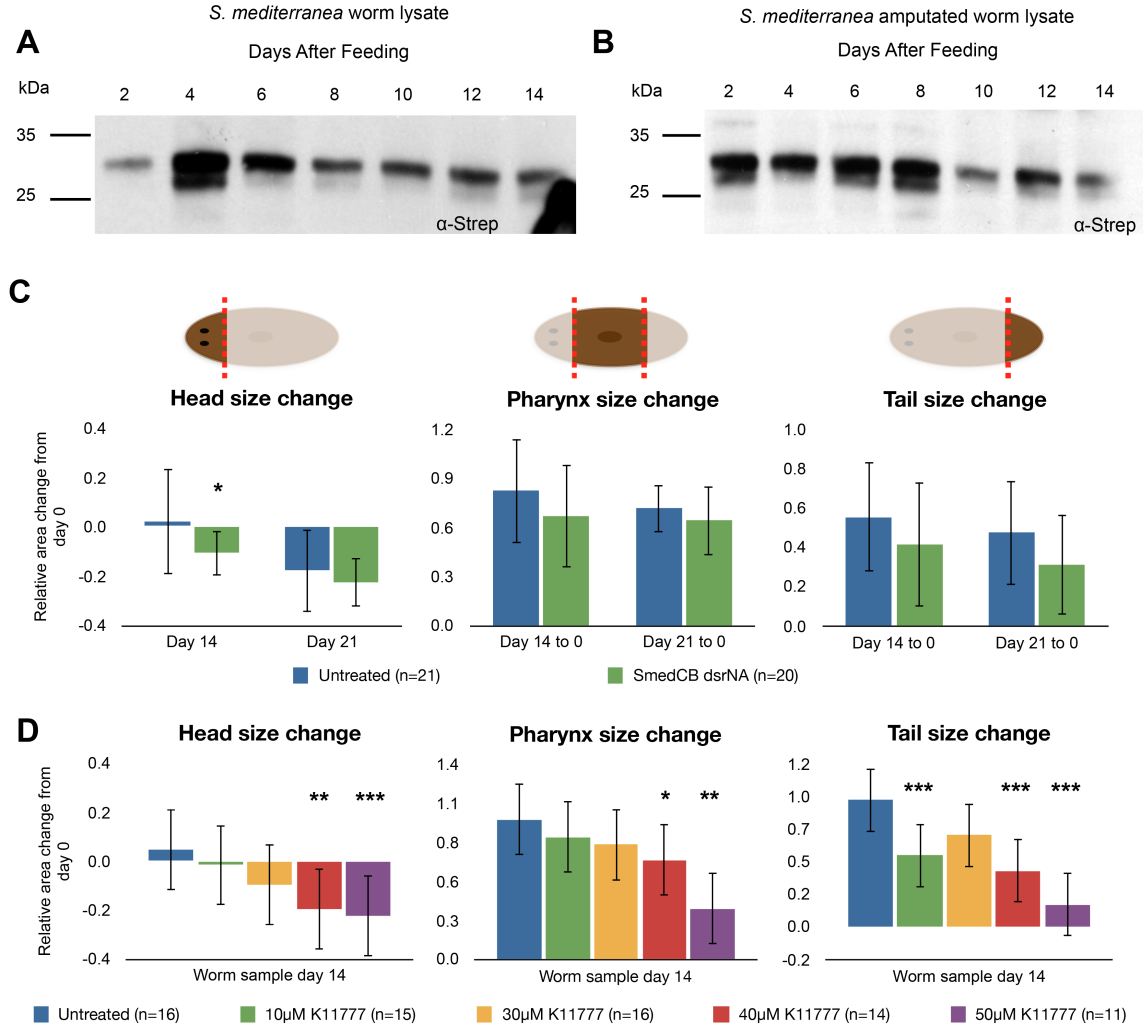
Figure 8: SmedCB shows a transient spike of protease activity following digestion and amputation, while chemical inhibition of both cathepsin B and L leads to a significant decrease in regeneration. (A) Whole worm lysate from *S. mediterranea* was labeled with the activity-based probe DCG-04 and resolved using SDS-PAGE. Western blot against the biotin tag of DCG-04 was used to analyze activity levels. *S. mediterranea* lysate showed an increase in cathepsin B labeling following digestion (day 4-6) before returning to base levels by day 14. **B)** When *S. mediterranea* worms were amputated following feeding, the increase in DCG-04 signal extended to day 8 before returning to baseline. The large upper band was quantified as SmedCB activity; five micrograms of total protein was used in each reaction with $10\mu M$ DCG-04. **(C)** Worms treated with SmedCB dsRNA did not show any significant defects in regeneration compared to untreated controls, but treatment with K11777 **(D)** showed a highly significant decrease in growth compared to vehicle worms. This effect was dose dependent. Relative changes in area are reported to account for any differences in initial size of worms selected. A negative value indicates that fragments shrank over time (heads), while positive values show relative growth.

41

Figure 8 continued: SmedCB shows a transient spike of protease activity following digestion and amputation, while chemical inhibition of both cathepsin B and L leads to a significant decrease in regeneration.

K11777 treated worms, on average, grew less (pharynxes, tails) and shrank more (heads), than untreated worms. These results were highly significant at 40 and 50μM treatment in all fragments, but not significant at lower concentrations.

SmedCB RNAi treated worms did not show major defects in regeneration, although regenerating worm segments did grow less than untreated worms (Fig 8C). This was quantified through the relative size increase of regenerating worms. Relative size increase was measured comparing the change between day 14 and day 0 worms divided by initial size to normalize for any variation in the size of selected animals. This trend of decreased growth in RNAi worms was consistent over two and three weeks, but it was not statistically significant. Therefore, despite the change in active protease levels observed in regenerating worm lysates, SmedCB alone does not appear to play an essential role in growth rates of *S. mediterranea* during regeneration.

We next examined whether inhibition of cathepsin B and L in tandem had any effect on regeneration and found that in contrast to RNAi of SmedCB alone, chemical inhibition of both proteases greatly reduced growth of treated worm fragments (Fig 8D). This effect was dose dependent. In general, head fragments have fewer neoblasts than pharynxes and tails.[37] Therefore, even vehicle treated heads do not grow much larger than their original size after amputation. Untreated heads had a relative size increase of only 0.045, as compared to 1.0 for pharynxes and 0.93 for tails. This means that while pharynxes and tails increased their size by almost 100% over two weeks of regeneration, heads tended to remain the same size. K11777 treated worms had significant reductions in growth and, in the case of heads, increased shrinking. Heads treated with 40μM K11777 were almost 20% smaller than when they were amputated. Pharynx and tail relative size increase stood at 0.72 and 0.41, respectively. Compared to vehicle treated worms, all three regions of worms had a significantly lower relative size increase when treated with K11777. These data

43

suggest that while knockdown of SmedCB alone is not enough to hinder to regeneration, chemical inhibition of SmedCB and cathepsin L proteases inhibits planaria ability to grow following amputation.

2.3 Discussion

Proteases perform many vital functions in both free-living and parasitic flatworms. For free-living flatworms like *Schmidtea mediterranea*, effective protein digestion is essential to growth and reproduction. Maturity of asexual planarians depends on size, which directly correlates with feeding; the more an animal is fed, the more it will divide and reproduce.[38] The major digestive enzymes in the family *Platyhelminthes* are cysteine proteases, in contrast to vertebrates, which predominantly use serine proteases. Several proteases often function as a network for protein digestion. The parasitic worm *Schistosoma mansoni* uses a proteolytic cascade of cysteine and aspartyl proteases to degrade host albumin and hemoglobin into amino acids. *S. mansoni* relies on the cysteine protease cathepsin B for digestion of albumin, although cathepsin L, legumain, and the aspartic protease cathepsin D are involved in later steps of albumin processing.[10] In contrast to the role of cathepsins B and L in albumin degradation, cathepsin D plays the primary role in hemoglobin digestion by schistosomes.[39] Interestingly, while homologs of cathepsin B, L, and D were found in free-living *Schmidtea* worms, no homolog for legumain was identified (Table 2). It is therefore presumed that legumain performs a "parasitic" function, although its precise role in the host-parasite relationship is unknown.[40] Legumain has been

implicated in the digestion of host hemoglobin in flatworm parasites [10,41] as well as other blood-feeding ecto-parasites, like ticks.[42]

We were able to induce regurgitation in *S. mediterranea* to examine the major proteases present and active in the flatworm gut lumen. We used peptide substrates and class specific protease inhibitors to determine that aspartyl, cysteine, and metalloproteases were present and active in the worm regurgitant. Treatment of live worms with a combination of aspartyl and a cysteine protease inhibitor reduced protein degradation in the gut by 46%. While the physiological pH of the *S. mediterranea* gut is unknown, *S. mansoni* regurgitant has been estimated to be pH 6.0-6.8 [20] while the *C. elegans* gut pH varies from pH 3.6 to pH 6.0 depending on specific location [43]. Previous work found that optimal degradation of albumin and hemoglobin by *S. mansoni* proteases, occurred at pH 4.0 and not pH 6.0.[10] Pepsin-type aspartyl proteases generally have little or no activity above pH 5.5 and therefore the pH of *S. mediterranea* is more acidic than pH 5.5 since pepstatin has such a profound effect on albumin degradation. In support of this, the number of cleavages sites generated by proteases active at pH 3.5 and 5.5 were considerably greater than at pH 7.5. In fact, only metalloproteases with exo-peptidase activity were detected a neutral pH and these enzymes are likely to play a role in generating single amino acids from peptide termini in the later stages of degradation. Taken together, we hypothesize that protein degradation is initiated at pH 5.5 or lower by a combination of aspartyl and cysteine proteases and these peptides are further processed by exo-peptidases in a region of the gut where the pH is closer to neutral.

Feeding assays with *S. mediterranea* confirmed the importance of cysteine and aspartyl proteases *in vivo*. Knockdown of cathepsin B activity through RNAi significantly decreased protein digestion in worms. This mirrors the function of SmCB1 in *S. mansoni*. Concurrent inhibition of SmedCB and cathepsin L activity using the inhibitor K11777 resulted in a further decrease in digestive ability. This implies some redundancy in the activities of cathepsin B and L. When SmedCB is specifically knocked down with RNAi, cathepsin L can compensate for some of the loss in digestive ability. However, when both proteases were chemically inhibited, digestion of rhodamine-labeled albumin was profoundly decreased. Aspartyl proteases also act in the process of digestion in *S. mediterranea*. Although no aspartyl proteases were detected in the worm regurgitant by mass spectrometry, it is possible that cathepsin D is not secreted directly into the intestinal lumen. Previous work in *S. japonicum* found that cathepsin D localized in digestive vacuolar compartments lining the gastrodermis, where it aids in the breakdown of host hemoglobin.[44, 39] The primary site of action of cathepsin D in these worms is thought to be in the gastrodermal lysosome or endosome. Blocking the action of aspartyl proteases with pepstatin resulted in a significant decrease in digestion in live *S. mediterranea*, and concurrent inhibition of cysteine and aspartyl proteases saw an almost complete loss of digestive ability. This suggests that initial digestion might take place in the intestinal lumen via the action of cysteine proteases and the remaining fragments are taken up into cells where further degradation by aspartyl proteases occurs. Inhibition of both cysteine and aspartyl proteases showed a dramatic decrease in digestive ability, suggesting that both of these classes of proteases are critical for proper digestion in the worm gut. Inhibition of

metalloproteases had no effect on digestive ability, suggesting that the metalloproteases detected in the worm regurgitant are not involved directly in the breakdown of food.

SmedCB RNA expression was visible throughout the branched intestine. This suggests that SmedCB transcription occurs throughout the gut epithelium followed by packaging of the translated protease in adjacent intestinal cells. This is confirmed by antibody localization of translated protein in secretory vesicles of intestinal cells. This suggests that SmedCB performs a similar role to SmCB1 in *S. mansoni* and that vesicles are used to store full-length SmedCB before cleavage of the zymogen form and secretion into the intestinal lumen. In regenerating worms, *SmedCB* is expressed in the newly formed gut as well as near the blastema during early stages of growth. It is unclear whether this expression is due solely to the formation of the gut near the blastema, or if SmedCB plays other important roles during tissue remodeling, requiring its increased presence and activity at the site of regeneration.

While SmedCB alone does not impact the ability of *S. mediterranea* worms to regenerate, inhibiting the activity of both cathepsin B and L did result in a significant decrease in growth rate. During regeneration, old structures are broken down through both autophagy and apoptosis.[45] Apoptosis occurs in two waves: an initial localized response near the wound site and followed by a systemic response. Even preexisting tissues undergo apoptosis in order to maintain their correct proportions within the worm.[46] Perhaps cysteine proteases are involved in tissue remodeling and turnover during regeneration and their absence slows the rate of this process. The lack of severe morphological effects during

regeneration in cathepsin B and L inhibited worms suggests that while these proteases may play a small role in growth, they are not entirely essential to worm survival.

We have exploited the ease of use of *Schmidtea* to confirm the conservation of a major cysteine protease in free-living and parasitic flatworms. We have found that not only is cathepsin B involved in digestion in free-living worms, but it also plays a role in growth and regeneration. Schistosome parasites also have neoblast-like stem cells, so there may be other aspects of worm growth or life cycle alterations that are conserved between planarians and *Schistosoma*.[7]

2.4 Methods

2.4.1 Colony care. A clonal line[47] of diploid, asexual *Schmidtea mediterranea* was used in all experiments.[48] Worms were maintained as described previously at room temperature (20-22°C) in 1x Montjuïc salts (worm water), pH 7.2, and were fed every two weeks with organic beef liver.[49] Unless otherwise stated, animals were starved for two weeks prior to use in experiments.

2.4.2 Generation of worm regurgitant. Worms were starved for one week and washed several times in 1x Montjuïc salts before the addition of 3% EtOH for one hour to induce regurgitation. Treatment with low-percentage ethanol does not lead to long-term damage of worms.[50] Control samples were treated with water. Worm regurgitant was collected and filtered using a 50mm Filter Unit (Nalgene®), then concentrated 50-fold using at an

Amicon® Ultra 10K MWCO Centrifugal Filter at 8,000 x *g* and 4°C. The concentrated

vomit was stored at -80°C.

2.4.3 Proteomic analysis of *S. mediterranea* regurgitant. Protein identification in *S.*

mediterranea regurgitant was performed using peptide sequencing by mass spectrometry.

10μg total protein was digested with trypsin, extracted, desalted using C18 zip-tips

(Rainin), and lyophilized. Liquid chromatography-tandem mass spectrometry was

performed as previously described.[51] Algorithms in the BLAST2Go program (v2.7.2)

were used to search for proteins with shared sequence features to proteins in *S.*

mediterranea vomit. Searches were conducted against the SwissProt or NCBInr

databases using the National Center for Biotechnology Information Server (July 27,

2015, blast.ncbi.nlm.nih.gov/Blast.cgi) and the top ranking hits reported in Supplemental

Table 1.

2.4.4 Protease assays. Protease activity in regurgitant from three replicate tanks were

compared using a mixture of 7 internally quenched fluorescent substrates available from

CPC Scientific, Sunnyvale, California (Table 2). These substrates were chosen based on

their diverse sequence composition to enable detection of multiple protease classes.

Worm regurgitant was diluted 7.5-fold in assay buffer containing 2.5 μM of each

substrate, 100 mM NaCl, 2 mM DTT, 0.01% Tween-20 and 20 mM Citrate-Phosphate

buffer at pH 3.5, 5.5 or 7.5. For inhibitor assays, the regurgitant from three replicate tanks

were combined at equal volume and incubated with 10 μM Pepstatin-A, 10 μM E-64, 1

mM 1,10-Phenanthroline, 0.5 mM AEBSF or 1% DMSO. Assays were run for 1 hour at

room temperature in black round bottom microplates using an excitation wavelength of

330 nm and emission wavelength of 400 nm. Activity was reported as change in

fluorescent units per second.

Table 2: List of internally quenched fluorescent substrates used to detect cleavages in *S. mediterranea* **regurgitant.**

CPC Catalog #	Sequence
AMYD-112	Mca-His-Gln-Lys-Leu-Val-Phe-Phe-Ala-Lys(DNP)-NH2
AMYD-114	Mca-Glu-Val-Lys-Met-Asp-Ala-Glu-Phe-Lys(DNP)-NH2
AMYD-109	Mca-Ser-Glu-Val-Asn-Leu-Asp-Ala-Glu-Phe-Arg-Lys(DNP)-Arg-Arg-NH2
MMPS-024	Mca-Arg-Pro-Lys-Pro-Tyr-Ala-Nva-Trp-Met-Lys(DNP)-NH2
SUBS-017	Mca-Gly-Lys-Pro-Ile-Leu-Phe-Phe-Arg-Leu-Lys(DNP)-DArg-NH2
AMYD-111	Mca-Arg-Pro-Pro-Gly-Phe-Ser-Ala-Phe-Lys(DNP)
CASP-060	Mca-Val-Asp-Gln-Met-Asp-Gly-Trp-Lys-(DNP)-NH2

The MSP-MS assay was performed as previously described with minor

modifications.[24] Worm regurgitant (500ng/mL) was assayed with an equal molar mixture

of 124 tetradecapeptides (500nM each) in a total reaction volume of 300μL. Assays were

performed at pH 3.5, 5.5, or 7.5 and incubated for 15, 60, 240, or 1200 minutes before

quenching with concentrated formic acid to a final pH of 2.5. Samples were desalted with

C18 zip-tips (Rainin) and analyzed by LC-MS/MS sequencing. Mass spectrometry and

data analysis were performed as described previously.[52] To compare substrate specificity,

iceLogo software (http://iomics.ugent.be/icelogoserver/)[53] was used to generate the

specificity signature for amino acids at ±4 positions adjacent to the identified cleavage site.

2.4.5 Identification of major cysteine proteases and cathepsin B in *S. mediterranea*.
Sequences of *Schistosoma* proteases were obtained via BLAST and used to manually identify homologs found using the *Schmidtea mediterranea* database (SmedDb; http://planaria.neuro.utah.edu).[4] Hits were checked via BLAST; reciprocal best hits were scored as putative homologs. The major cathepsin B homolog was isolated from planaria lysate (~50 planarians in lysis buffer; 100mM Tris pH 7.5, 200mM NaCl, 1% NP-40, 0.1% SDS, 1XTBS). Lysate was separated on a MonoQ 10/100 GL column (GE Healthcare Life Sciences) and assayed for protease activity with Z-RR-AMC (BACHEM). Fractions of cathepsin B activity were pooled and separated on a Superdex 200 10/300 GL column (GE Healthcare Life Sciences) which had been tested with a Gel Filtration Calibration Kit LMW (GE Healthcare) to determine at what volume the cathepsin B homolog would elute. The resulting samples with proteolytic activity were run on a 10% SDS-PAGE gel, silver stained, and cut into bands for analysis on PE-Biosystems Voyager Elite STR MALDI-TOF. The band corresponding to the most proteolytic activity and correct size was determined to be the cathepsin B homolog identified by SmedDb.

2.4.6 Protease activity assays of whole worm lysates. To measure protease activity of worm lysates, planarians were ground with mortar and pestle for thirty seconds and incubated with lysis buffer (100mM Tris pH 7.5, 200mM NaCl, 1% NP-40, 0.1% SDS, 1XTBS) for one hour on ice with occasional vortexing. Samples were spun in a microcentrifuge at maximum speed, 4°C for twenty minutes. The supernatant was saved and activity was measured by adding assay mix (5mMDTT, 50mM sodium citrate, pH5.5)

51

containing 50μM fluorescent peptide, either Z-FR-AMC or Z-RR-AMC (BACHEM). Fluorescence was measured (excitation 360nm, absorbance 460nm) using a FlexStation fluorometer (Molecular Devices) and SoftMax Pro 4.8 software. Both kinetic and endpoint assays were used. Chemical inhibitors of cathepsin B, K11777 (UCSF, CDIPD) and CA-074 (Sigma), were added at 50μM concentration two hours prior to addition of fluorescent peptides for select experiments. Protease activity was also measured with activity-based probes DCG-04 and BMV109, both gifts from Matthew Bogyo.[29, 36] Probes were added to 1μM final volume in samples containing 5mM DTT, pH 5.5 for 1 hour at 37°C before imaging Cy5 levels via Typhoon Trio (GE Healthcare Life Sciences).

2.4.7 *In vivo* RNAi. *In vitro* double-stranded RNA (dsRNA) was synthesized from a PCR template of cathepsin B using T3 and T7 polymerases (Promega). The dsRNA was injected as described.[54] Worms were injected with three 33.2nL pulses of dsRNA (100ng total) once a day over three consecutive days. To determine the efficiency of knockdown, quantitative real-time PCR (qPCR) was performed using an Mx3005P QPCR System (Agilent Technologies) and LightCycler® SYBR Green Master I (Roche). mRNA levels of cathepsin genes were compared to a standard internal control clone, GAPDH (Accession number: AY067285.1) to normalize RNA starting material. The efficiency and specificity of the primers was tested via serial dilutions of primers and cDNA template according to the protocol.[55]

Primers used: GAPDH: 5'-AGCTCCATTGGCGAAAGTTA-3', 5'-CTTTTGCTGCACCAGTTGAA -3'; CB1: 5'-CTAGATTCCAAACCGTTTCGGACA-

3', 5'-CAAGCTGCCAAAGAAAAGTTCAGG-3'; CL1: 5'-CAAGGCTATCCAGCAAATGG-3', 5'-GAATCAACGCATTTGCAAT-3'; CD: 5'-GGCGAAATCACAATTGGAAC-3', 5'-ATTTTCCATTCGATACGGCA-3'.

2.4.8 Digestion and feeding assays. Planaria were fed pureed liver with tetramethylrhodamine conjugated bovine serum albumin (RhBSA, Molecular Probes®). Worms were starved one week prior to feeding; RNAi treated worms were fed one week after injections were completed. Unless otherwise stated, worms treated with chemical inhibitors (K11777, pepstatin) received drug treatment for one hour prior to feeding. Worms were fed in the dark for one hour before being transferred into worm water (with or without appropriate drug) without food for one hour. Live worms were placed on slides on ice for a few minutes to stop worm movement before imaging. Images were taken using an Axiovert 40 CFL microscope and Axiovision Rel. 4.8.2 software (Zeiss). Fluorescence was quantified using ImageJ. Live worms were imaged again after 24 or 48 hours. Measurements were evaluated by paired t-tests; differences were considered significant with $P<0.05$.

2.4.9 Antibodies. The catalytic region of *S. mediterranea* cathepsin B (based on the reference sequence mk4.000308.13.01) was cloned into pET28a and expressed in *Escherichia coli*. Protein was purified under reducing conditions (8M urea) on Ni-NTA Agarose (Qiagen) and dialyzed before mouse immunization. Sera was pooled and purified with NAb Protein G Spin Columns (Pierce). Three peptide regions of cathepsin B were

selected and synthesized by New England Peptide, and combined to immunize rabbits by Covance. The sequence and position of these peptides is shown in Figure 6A. Antibodies were purified with NAb Protein G Spin Columns (Pierce).

2.4.10 Whole-mount *in situ* hybridization. DNA template for *in vitro* transcription of anti-sense RNA probes was amplified from a PCR sequence containing the full length SmedCB1 gene. Probes were synthesized in an *in vitro* transcription with DIG-12-UTP (Roche) and precipitated with lithium chloride and ethanol according to the Roche suggested protocol. Probes were then resuspended and stored; animals between 2-5mm were processed for WISH as previously described with modifications.[33, 34] All animals were treated with 1:100 dilution of *SmedCB* probe.

2.4.11 Immunofluorescence and electron microscopy. Animals were killed by treatment with 5% N-Acetyl L-Cysteine for five minutes at room temperature, fixed in 4% formaldehyde, and dehydrated in 50% MeOH. Planarians were embedded in paraffin, sliced, and treated with antibodies described above. For immunofluorescence, anti-rabbit and anti-mouse secondary antibodies with Alexa Fluor 488nm (LifeTechnonolgies) were used at 1:100. All imaging was performed using AxioImage.M1, Axiovision Rel. 4.8.2 software (Zeiss). Worms for electron microscopy were fixed and processed using the same procedures as the immunofluorescence. Worms were treated with the anti-zymogen antibody described above and an anti-rabbit 10nm gold preadsorbed secondary antibody (AbCam).

2.4.12 Regenerative assays. To measure the effect of cathepsin B RNAi on regeneration, planarians were amputated one full day after injections were complete. Worms treated with cathepsin B inhibitors were also amputated pre- and post-pharyngeally and immediately placed in worm water containing the appropriate drug. Unless otherwise indicated, images for size quantitation were taken immediately after amputation (Day 0) and again on days 5, 8, 11, and 14. Worms saved for RNA extraction for qPCR were also frozen at -80°C on these days. Live images were taken using an AxioZoom.V16 microscope and Axiovision Rel. 4.8.2 software (Zeiss). Phenotypes were scored by measuring body areas with ImageJ. Each planaria amputated fragment (head, pharynx, and tail) was measured for change in total area and blastema size. Changes in area were evaluated with paired t-tests; differences were considered significant with $P<0.05$.

2.5 Acknoweldgements

2.6 References

[1] Morgan, T.H. Experimental studies of the regeneration of *Planaria maculata*. *Arch. Entwm. Org.* 1898; **7**: 364–397.

[2] Reddien PW, Alvarado A. S. Fundamentals of planarian regeneration. *Annu. Rev. Cell. Dev.* Biol. 2004; **20**: 725–757.

[3] Reddien, P.W., Bermange, A.L., Murfitt, K.J., Jennings, J.R., Alvarado, A.S. Identification of genes needed for regeneration, stem cell function, and tissue homeostasis by systematic gene perturbation in planaria. *Cell*. 2005; **8**: 635-649.

[4] Collins, J. J., Newmark, P. A. It's no fluke: the planarian as a model for understanding schistosomes. *PLoS Pathog.* 2013; **9**:(7) e1003396. doi:10.1371/ journal.ppat.1003396

[5] Alvarado, A. S. The *Schmidtea mediterranea* database as a molecular resource for studying platyhelminthes, stem cells and regeneration. *Development*. 2002; 129: 5659–5665.

[6] Gryseels, B., Polman, K., Clerinx, J., Kestens, L. Human schistosomiasis. *Lancet.* 2006; **368**: 1106–18.

[7] Collins, J. J., Wang, B., Lambrus, B.G., Tharp, M. Iyer, H., Newmark, P.A. Adult somatic stem cells in the human parasite *Schistosoma mansoni*. *Nature*. 2013; **494**: 476–9.

[8] Nishimura, O., Hirao, Y., Tarui, H., Agata, K. Comparative transcriptome analysis between planarian *Dugesia japonica* and other platyhelminth species. *BMC Genomics*. 2012; **13**: 289.

[9] Bocchinfuso, D. G., Taylor, P., Ross, E., Ignatchenko, A., Ignatchenko, V., Kislinger, T. et al. Proteomic profiling of the planarian *Schmidtea mediterranea* and its mucous reveals similarities with human secretions and those predicted for parasitic flatworms. *Mol. Cell. Proteomics*. 2012; **11**: 681–91.

[10] Delcroix, M., Mohammed, S., Caffrey, C. R., Lim, K. C., Dvořák, J., Hsieh, I. et al. A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. J. *Biol. Chem.* 2006; **281**: 39316–29.

[11] Sajid M, McKerrow J.H. Cysteine proteases of parasitic organisms. *Mol Biochem Parasitol.* 2002; **120**: 1–21.

[12] Brown, A., Burleigh, J.M., Billett, E. E., Pritchard, D. I. An initial characterization of the proteolytic enzymes secreted by the adult stage of the human hookworm *Necator americanus*. *Parasitol.* 1995; **110**: 555-563.

[13] Williamson, A.L., Brindley, P.J., Abbenante, G., Prociv, P. Berry, C., Girdwood, K. et al. Cleavage of hemoglobin by hookworm cathepsin D aspartic proteases and its potential contribution to host specificity. *FASEB J.* 2002; **16**: 1458-1460.

[14] Orr, G. L., Strickland, J. A. & Walsh, T. A. Inhibition of Diabrotica larval growth by a multicystatin from potato tubers. *J. Insect Physiol.* 1994; **40**: 893–900.

[15] Robinson, M. W., Dalton, J. P., Donnelly, S. Helminthpathogen cathepsin proteases: it's a family affair. *Trends Biochem. Sci.* 2008; **33**: 601–8.

[16] Caffrey, C.R., Scory, S., Steverding, D. Cysteine proteinases of trypanosome parasites: Novel targets for chemotherapy. *Curr. Drug Targets.* 2000; **1**: 155–162.

[17] Rosenthal, P.J. Cysteine proteases of malaria parasites. Int J Parasitol. 2004; 34: 1489–1499.

[18] Abdulla, M.-H., Lim, K.-C., Sajid, M., McKerrow, J. H., Caffrey, C. R. Schistosomiasis mansoni: novel chemotherapy using a cysteine protease inhibitor. *PLoS Med.* 2007; **4**, e14 10.1371/journal.pmed/0040014

[19] Chappell, C. L., and Dresden, M. H. *Schistosoma mansoni*: proteinase activity of "hemoglobinase" from the digestive tract of adult worms. *Exp. Parasitol.* 1986; **61**: 160–167.

[20] Sajid, M., McKerrow, J. H., Hansell, E., Mathieu, M. A., Lucas, K. D., Hsieh, I. et al. Functional expression and characterization of *Schistosoma mansoni* cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. *Mol. Biochem. Parasit.* 2003; **131**: 65–75.

[21] Brindley, P. J., Kalinna, B. H., Dalton, J. P., Day, S. R., Wong, J. Y. M., Smythe, M. L., McManus, D. P. Proteolytic degradation of host hemoglobin by schistosomes. *Mol. Biochem. Parasit.* 1997; **89**: 1-9.

[22] Ruppel, A., Shi, Y. E., Wei, D. X., Diesfeld, H. J. Sera of *Schistosoma japonicum*-infected patients cross-react with diagnostic 31/32kD proteins of *S. mansoni. Clin Exp Immunol.* 1987; **69**: 291–8.

[23] Correnti, J. M., Brindley, P. J., Pearce, E. J. Long-term suppression of cathepsin B levels by RNA interference retards schistosome growth. *Mol. Biochem. Parasit.* 2005; **143**: 209–15.

[24] O'Donoghue, A. J., Eroy-Reveles, A. A., Knudsen, G. M., Ingram, J., Zhou, M., Statnekov, J. B. et al. Global identification of peptidase specificity by multiplex substrate profiling. *Nature.* 2012; **9**: 1095–1103.

[25] Stepek, G., McCormack, G., Winter, A. D., Page, A. P. A highly conserved, inhibitable astacin metalloprotease from *Teladorsagia circumcincta* is required for cuticle formation and nematode development. *Int. J. Parasitol.* 2014; **45**: 345–355.

[26] Falgueyret, J. P., Black, W. C., Cromlish, W., Desmarais, S., Lamontange, S., Mellon, C. et al. An activity-based probe for the determination of cysteine cathepsin protease activities in whole cells. *Anal. Biochem.* 2004; **335**: 218–227.

[27] Choy, J. W., Bryant, C., Calvet, C. M., Doyle, P. S., Gunatilleke, S. S., Leung, S. S. et al. Chemical–biological characterization of a cruzain inhibitor reveals a second target and a mammalian off-target. *Journal of Organic Chemistry.* 2013; **9**: 15–25.

[28] Loukas, A., Selzer, P. M., Maizels, R. M. Characterisation of Tc-cpl-1, a cathepsin L-like cysteine protease from *Toxocara canis* infective larvae. *Mol. Biochem. Parasit.* 1998; **92**: 275–289.

[29] Greenbaum, D., Medzihradszky, K. F., Burlingame, A., Bogyo, M. Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem. Biol.* 2000; **7**: 569–581.

[30] Bowen I.D. The fine structure of the planarian *Polycelis tenuis* Iijima. *Protoplasma.* 1974; **79**: 1–17.

[31] Garcia-Corrales P, Gamo J. The ultrastructure of the gastrodermal gland cells in the freshwater planarian *Dugesia gonocephala* s.l. *Acta Zool.* 1986; **67**: 43–51.

[32] Ishii S. Electron microscopic observations on the Planarian tissues. *Fukushima J Med Sci.* 1965; **12**: 67–87.

[33] Pearson, B. J., Eisenhoffer, G. T., Gurley, K. A., Rink, J. C., Miller, D. E., Alvarado, A. S. Formaldehyde-based whole-mount *in situ* hybridization method for planarians. *Dev. Dyn.* 2009; **238**: 443–50.

[34] King, R. S., Newmark, P. A. *In situ* hybridization protocol for enhanced detection of gene expression in the planarian *Schmidtea mediterranea.* BMC Dev. Biol. 2013; **13**: 8.

[35] Isolani, M. E., Abril, J. F., Saló, E., Deri, P., Bianucci, A. M., Batistoni, R. Planarians as a model to assess in vivo the role of matrix metalloproteinase genes during homeostasis and regeneration. *PLoS One.* 2013; **8**(2): e55649. doi: 10.1371/journal.pone.0055649

[36] Verdoes, M., Bender, K. O., Segal, E., van der Linden, W. A., Syed, S. Withana, N. P. et al. Improved quenched fluorescent probe for imaging of cysteine cathepsin activity. *J. Am. Chem. Soc.* 2013; **135**: 14726–30.

[37] Bode, A., Salvenmoser, W., Nimeth, K., Mahlknecht, M., Adamski, R. M., Rieger, R. et al. Immunogold-labeled S-phase neoblasts, total neoblast number, their distribution, and evidence for arrested neoblasts in *Macrostomum lignano. Cell Tissue Res.* 2006; **325**: 577–87.

[38] Miller, C. M., Newmark, P. A. An insulin-like peptide regulates size and adult stem cells in planarians. *Int. J. Dev. Biol.* 2012; **56**: 75–82.

[39] Brindley P.J., Kalinna, B. H., Wong, J. Y. M., Bogtish, B. J., King, L. T., Smyth, D. J. et al. Proteolysis of human hemoglobin by schistosome cathepsin D. *Mol Biochem Parasit.* 2001; **112**: 03–12.

[40] Caffrey, C. R., Mathieu, M. A., Gaffney, A. M., Slater, J. P.,Sajid, M., Lucas, K. D. et al. Identification of a cDNA encoding an active asparaginyl endopeptidase of *Schistosoma mansoni* and its expression in Pichia pastoris. *FEBS Letters.* 2002; **466**: 244–248.

[41] Robinson, M. W., Menon, R., Donnelly, S. M., Dalton, J. P., Ranganathan, S. An integrated transcriptomics and proteomics analysis of the secretome of the helminth pathogen *Fasciola hepatica*: Proteins associated with invasion and infection of the mammalian host. *Mol. Cell. Proteomics.* 2009; **8**: 1891–1907.

[42] Sojka, D., Zdeněk, F., Horn, M., Hajdušek, O., Caffrey, C. R., Mares, M. et al. Profiling of proteolytic enzymes in the gut of the tick *Ixodes ricinus* reveals an evolutionarily conserved network of aspartic and cysteine peptidases. *Parasit. Vectors.* 2008; **1**: 7.

[43] Chauhan, V. M. et al. Mapping the pharyngeal and intestinal pH of *Caenorhabditis elegans* and real-time luminal pH oscillations using extended dynamic range pH-sensitive nanosensors. *ACS Nano.* 2013; **7**: 5577–5587.

[44] Bogitsh, B. J., Krischner, K. F. *Schistosoma japonicum*: ultrastructural localization of a hemoglobinase using mercury labeled pepstatin. *Exp. Parasitol.* 1986; **62**: 211–215.

[45] González-Estévez, C., Saló, E. Autophagy and apoptosis in planarians. *Apoptosis.* 2010; **15**: 279–92.

[46] Pellettieri, J., Fitzgerald, P. Watanabe, S., Mancuso, J., Green, D.R., Alvarado, A. S. Cell death and tissue remodeling in planarian regeneration. *Dev. Biol.* 2010; **338**: 76–85.

[47] Sánchez Alvarado, A., Newmark, P.A., Robb, S. M., Juste, R. The *Schmidtea mediterranea* database as a molecular resource for studying platyhelminthes, stem cells and regeneration. Development. 2002; 129: 5659–5665.

[48] Benazzi M., Ballester, R., Baguña, J., Puccinelli, I. The fissiparous race of the planarian *Dugestia lugubris* S.L. found in Barcelona (Spain) belongs to the biotype G: comparative analysis of the karyotypes. *Caryologia*. 1972; **25**: 59-68.

[49] Cebrià, F., Newmark, P. A. Planarian homologs of netrin and netrin receptor are required for proper regeneration of the central nervous system and the maintenance of nervous system architecture. *Development*. 2005; **132**: 3691–3703.

[50] Stevenson, C. G., Beane, W. S. A low percent ethanol method for immobilizing planarians. *PLoS One*. 2010; **5**, e15310, 10.1371/journal.pone.0015310

[51] O'Donoghue, A. J., Knudsen, G. M., Beekman, C., Perry, J. A., Johnson, A. D., DeRisi, J. L. et al. Destructin-1 is a collagen-degrading endopeptidase secreted by *Pseudogymnoascus destructans*, the causative agent of white-nose syndrome. *PNAS*. 2005; **112**: 7478-7483.

[52] Bibo-Verdugo, B., O'Donoghue, A. J., Rojo-Arreola, L., Craik, C. S., García-Carreño, F. Complementary proteomic and biochemical analysis of peptidases in lobster gastric juice uncovers the functional role of individual enzymes in food digestion. *Mar. Biotechnol.* 2015; **18**: 201-214.

[53] Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., Gevaert, K. Improved visualization of protein consensus sequences by iceLogo. *Nat Meth*. 2009; **6**: 786–787.

[54] Oviedo, N. J., Nicolas, C. L., Adams, D. S., Levin, M. Gene knockdown in planarians using RNA interference. *CSH Protoc*. 2008; **3**, e10 10.1101/pdb.prot5054

[55] Nolan, T., Hands, R. E., Bustin, S. A. Quantification of mRNA using real-time RT-PCR. *Nat. Protoc*. 2006; **1**: 1559–1582.

Chapter 3 Discussion and Conclusions

3.1 Protease networks in parasites, like planaria, employ cysteine and aspartyl proteases

Many parasitic flatworms rely on networks of proteases to digest host proteins. As previously discussed, *Schistosoma* worms employ the use of cysteine and aspartyl proteases to break down major host blood proteins like hemoglobin and albumin. The majority of proteolytic activity in the schistosome gut is provided by the cysteine protease SmCB1, a cathepsin B protease. SmCL1 (cathepsin L, a cysteine protease), SmCD (cathepsin D, an aspartyl protease), and SmAE (asparaginyl endopeptidase, a cysteine protease known as legumain) also aid in digestion. Once these proteases have made their initial cuts, metalloproteases, including a leucine aminopeptidase, act as exopeptidases to degrade peptides into absorbable amino acids. This network is similar in more distantly related flatworms, such as the liver fluke *Fasciola hepatica*. This trematode causes the disease fasciolosis. Unlike *Schistosoma*, adult *Fasciola* rely more heavily on cathepsin L-like proteases to digest hemoglobin. Juvenile *Fasciola* worms secrete cathepsin L, cathepsin B, and asparginyl endopeptidases.[1] After three weeks of development, this expression pattern changes to largely favor cathepsin L-like proteases.[2] Both cathepsin L and B proteases likely facilitate transit of juvenile *Fasciola* across the gut wall. RNAi against these enzymes prevents invasion and migration of juveniles in rats.[3]

Like *Schistosoma*, several parasitic nematodes also rely on a network of cysteine, aspartyl, and/or metalloproteases for digestion of host proteins. Roundworm parasites, like

the blood-feeding nematode *Anclyostoma caninum*,[4] *Haemonchus contortus*,[5] and the hookworm *Necator americanus*[6] all initiate cleavage of hemoglobin by cysteine and aspartyl proteases, followed by exopeptidase degradation using metalloproteases and aminopeptidases. *H. contortus* has an especially high number of cathepsin B-like protease genes. Cathepsin B proteases represent approximately 16% of all intestinal transcripts in *H. contortus*.[7] However, only a few of the 23 protease gene products identified have been characterized.[8]

Some protozoan parasites also use a network of proteases in digestion of hemoglobin, including *Plasmodium falciparum*, a causative agent of malaria. Cysteine proteases (falcipains), aspartic proteases (plasmepsins), a metalloprotease (falcilysin), and aminopeptidases all function in the digestion of hemoglobin in the *Plasmodium* food vacuole.[9] Knockouts of either falcipains or plasmepsins lead to an increased sensitivity to pepstatin[10,11] or cysteine protease inhibitors,[12] respectively. This indicates that these two classes of proteases play complementary or redundant roles in digestion.

Finally, ectoparasites such as ticks use similar proteases during digestion. Cathepsins B and L, apartyl proteases, leucine aminopeptidases, and legumains all contribute to degradation of host hemoglobin by ticks. Cathepsin L is expressed only in feeding stages. Cathepsin B can be found at other times during the tick lifecycle, but it is highly upregulated during feeding and is solely expressed in gut tissue. A multienzyme network similar to *Schistosoma* has been observed in hard ticks. Some of these enzymes, like legumain, also have roles in embryogenesis.[13]

Despite being a free-living organism that does not digest host proteins, the proteolytic activity in the planaria gut is similar to that of many parasites. The cysteine protease SmedCB (cathepsin B) provides the bulk of digestive activity, although some cysteine protease activity is attriubted to a cathepsin L-like protease. We also identified activity of an aspartyl protease, and there is a cathepsin D protease homolog in the planarian genome. While metalloprotease activity was not detected *in vivo* during digestion, there do appear to be secreted metalloproteases in the planarian gut, as evidenced by MSP-MS. A homolog to the leucine aminopeptidase observed in *Schistosoma* is present in the planarian genome. The one major difference is that no asparginyl endopeptidase was found in the planarian genome. Although free-living worms do not need to digest host proteins, their network of digestive proteases appears to be relatively well conserved with many parasites beyond the platyhelminths. Further studies can now be done to help relate the evolutionary conservation of this protease network and provide insights into these important classes of enzymes.

3.2 Further development of genetic tools for flatworms is still needed

To fully develop an organism as a model system, there must be 1) an annotated genome and transcriptome, 2) method(s) of transgenesis, and 3) genome editing to enable rapid genetic analyses. To date, *S. mediterranea* has an annotated genome available and transgenic techniques are developed for sexual worms, although they have not yet been fully optimized. However, despite the advantages of planaria, it and other parasitic worms remain genetically intractable without tools to directly edit the genome. One such tool

would be the use of a CRISPR/Cas system, which has been demonstrated in several organisms, including the roundworm *C. elegans*. Clustered regularly interspaced short palindromic repeats (CRISPR)/ Cas9 has allowed researchers to inactivate genes in several organisms. The CRISPR system was first identified in bacteria, where it acts as an adaptive immune system that identifies and cleaves foreign DNA using a DNA endonuclease.[14] To edit the genome directly, the system can be simplified to the Cas9 nuclease and a single guide RNA (sgRNA), which can form a complex and target DNA of interest.[15] So far, this system has been used successfully to study the biological function of gene products in organisms ranging from yeast to mice. Applying this genome editing technique to planaria and parasitic flatworms would greatly increase our ability to study specific biological or parasitic functions of these worms.

However, there are several major steps that must occur before CRISPR/Cas9 can be successfully implemented in platyhelminths. First, there are many important regulatory elements that must be better characterized in order for the Cas9 nuclease to be specific and efficient. Second, there must be a system in place to enrich for transgenic organisms. In the case of parasites, there are several additional challenges to working with organisms that develop inside a host; gene disruption would most likely require identification through a selectable cointegration marker.[16] Finally, the creation of an efficient sgRNA is not trivial. Half of all sgRNAs in *C. elegans* tested have had very low efficiencies.[17,18] Algorithms currently exist to help predict sgRNA efficiency,[19] but these may not necessarily hold true for flatworms.

As there are currently no transgenic parasitic flatworms, it seems most likely that CRISPR/Cas9 systems will be first and most easily employed in *S. mediterranea*. The addition of this tool to study planaria will facilitate new ways to knockin sequences of interest, while current genes can only be knocked down by RNAi. Optimization of CRISPR/Cas9 for a flatworm system will help the transition to parasites as better genome annotation and transgenic enrichment techniques develop. Increasing genetic tools for planaria could lead to new tools in parasitic helminths as well to examine parasite-specific targets of interest.

3.3 References

[1] Beckham, S. A., Piedrafita, D., Phillips, C. I., Samarawickrema, N., Law, R. H., Smooker, P. M., *et al.* 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.* 2009; **41**: 1601-1612.

[2] McVeigh, P., Maule, A. G., Dalton, J. P., Robinson, M. W. *Fasciola hepatica* vireulence-associated cysteine peptidases: a systems biology perspective. *Microbes Infect.* 2012; **14**: 301-310.

[3] McGonigle, L., Mousely, A., Marks, N. J., Brennan, G. P., Dalton, J. P., Spithill, T. W., *et al*. The silencing of cysteine proteases in *Fasciola hepatica* newly excysted juveniles using RNA interference reduces gut penetration. *Int. J. Parasitol.* 2008; **38**: 149-155.

[4] Williamson, A. L., Lecchi, P., Turk, B. E., Choe, Y., Hotez, P. J., McKerrow, J. H. *et al*. A multi-enzyme cascade of hemoglobin proteolysis in the instestine of blood-feeding hookworms. *J. Biol. Chem.* 2004; **279**: 35950-35957.

[5] Fetterer, R. H., Rhoads, M. L. The *in vitro* uptake and incorporation of hemoglobin by adult *Haemonchus contortus*. *Vet Parasitol.* 1997; **69**: 77-87.

[6] Ranjit, N., Zhan, B. Hamilton, B., Stenzel, D., Lowther, J., Pearson, M. *et al*. Proteolytic degradation of hemoglobin in the intestine of the human hookworm *Necator americanus*. *J. Infect. Dis.* 2009; **199**: 904-912.

[7] Jasmer, D. P., Mitreva, M. D., McCarter, J. P. mRNA sequences for *Haemonchus contortus* intestinal cathepsin B-like cysteine proteases display an extreme abundance and diversity compared with other adult mammalian parasitic nematodes. *Mol. Biochem. Parasitol.* 2004; **137**: 297-305

[8] Muleke, C. I., Ruofeng, Y. Lixin, X., Xinwen, B., Xiangrui, L. Cloning and sequence analysis of *Haemonchus contortus* Hc58cDNA. *DNA Seq.* 2007; **18**: 176-183.

[9] Goldberg, D. E. Hemoglobin degradation. *Curr. Top. Microbiol. Immunol.* 2005; **295**: 275-291.

[10] Rosenthal, P. J. Falcipains and other cysteine proteases of malaria parasites. *Adv. Exp. Med. Biol.* 2011; **712**: 30-48.

[11] Sijwali, P. S., Rosenthanl, P. J. Gene disruption confirms a critical role for the cysteine protease falcipain-2 in hemoglobin hydrolysis by *Plasmodium falciparum. Prot. Natl. Acad. Sci. USA.* 2004; **101**: 4384-4389.

[12] Moura, P. A., Dame, J. B., Fidock, D. A. Role of *Plasmodium falciparum* digestive vacuole plasmepsins in the specificity and antimalarial mode of action of cysteine and aspartic proteases inhibitors. *Antimicrob. Agents Chemother.* 2009; **53**, 4968-4978.

[13] Sojka, D., Francischetti, I.M.B., Calvo, E., Kotsyfakis, M. Cysteine proteases from bloodfeeding arthropod ectoparasites. *Adv. Exp. Med. Biol.* 2011; **712**: 177-191.

[14] Marraffini, L.A., Sontheimer, E.J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archea. *Nat. Rev. Genet.* 2010; **11**: 181–90.

[15] Jinek, M. K. Chylinski, I., Fonfara, M., Hauer, J. A. Doudna *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* 2012; **337**: 816-821.

[16] Ward, J. D. Rendering the intractable more tractable: tools from *Caenorhabditis elegans* ripe for import into parasitic nematodes. *Genetics.* 2015; **201**: 1279–1294.

[17] Arribere, J. A., R. T. Bell, B. X. H. Fu, K. L. Artiles, P. S. Hartman *et al.* Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in *Caenorhabditis elegans. Genetics.* 2014; **198**: 837–846.

[18] Paix, A., A. Folkmann, D. Rasoloson, and G. Seydoux. High efficiency, homology-directed genome editing in *Caenorhabditis elegans* using CRISPR-Cas9 ribonucleoprotein complexes. *Genetics.* 2015; **201**: 47–54.

[19] Doench, J. G., E. Hartenian, D. B. Graham, Z. Tothova, M. Hegde *et al.*Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* 2014; **32**: 1262–1267.

Appendix 1: Supplemental Tables

Supplemental Table 1: Proteins identified in *Schmidtea* regurgitant. Samples of worm regurgitant were collected, concentrated, and trypsinized for peptide sequencing by mass spectrometry. Thirty proteins were found in all three replicates, thirty-two in at least two replicates, and sixty proteins were found in only one replicate. A representative sample of the top 62 proteases, including astacin 2, 5, and cathepsins B and L, are listed here.

#	Smed Acc. #	Putative protein function	Total peptides in all 3 replicates
1	4969	PKD1L-2	78
2	7593	slc15a-3-1	53
3	11291	c-Jun N-terminal kinases	49
4	12146	X1.D.A5.1 [Smediterranea]	13
5	8192	PROG-1	18
6	9286	GLI pathogenesis related-1, partial	61
7	20173	tolloid like-1 or astacin 2	21
8	11574	raptor [Smediterranea]	31
9	17678	hypothetical protein, partial [Smediterranea]	11
10	1786	GLI pathogenesis related-1, partial [Smediterranea]	11
11	12224	peptidoglycan recognition protein-1	10
12	1063	calreticulin-1 [Smediterranea]	11
13	9757	SMG-1 [Smediterranea]	10
14	8443	glutathione S transferase-1 [Smediterranea]	25
15	3272	slc30a-10 [Smediterranea]	16
16	13968	secreted frizzled-related protein 1 [Smediterranea]	64
17	16029	Ast5 protein [Smediterranea]	16
18	13828	forkhead box J1-like protein 1 [Smediterranea]	12
19	3136	SPT16 [Smediterranea]	12
20	6448	peptidoglycan recognition protein-1	8
21	19244	hypothetical protein, partial [Smediterranea]	14
22	18755	FKBP2, partial [Smediterranea]	10
23	13267	X1.C3.3 [Smediterranea]	17
24	6574	notchless-1 [Smediterranea]	4
25	11598	insulin receptor-like 1 protein [Smediterranea]	10
26	19955	netrin receptor-like protein	22
27	21608	ELAV2	37
28	18124	slc13a-5	14
29	9985	ATFl1, partial [Smediterranea],	7
30	15211	tyramine beta-hydroxilase, partial [Smediterranea]	4
31	21352	blastula protease 10-like	28
32	6032	trpm8 channel-associated factor homolog	12
33	2785	16 kda calcium-binding protein (egg antigen sme16)	28
34	14056	peptidase inhibitor partial	6
35	11174	annexin a7	8
36	2715	purine nucleoside phosphorylase-like	6
37	21885	peroxiredoxin-6	13
38	4838	protein dj-1-like	8
39	7481	malate cytoplasmic	12
40	12927	ubiquitin-conjugating enzyme e2 l3-like	6

#	Smed Acc. #	Putative protein function	Total peptides in all 3 replicates
41	5595	elongation factor 1-alpha 1	8
42	10179	tnf receptor-associated factor 6	4
43	21067	sim protein, partial (3%C, 36% I)	4
44	14458	secreted peptide prohormone 18 [Smediterranea]	4
45	5488	apaf1 [Smediterranea]	2
46	5499	peptidoglycan recognition protein-1	2
47	8351	slc16a-2 [Smediterranea]	4
48	17679	pmp-10, partial [Smediterranea]	2
49	3057	roboA [Smediterranea]	2
50	13357	SET domain-1 [Smediterranea]	2
51	18278	Tor [Smediterranea]	2
52	11974	inhibin-1 [Smediterranea] 6%C 35%I	2
53	19497	slc20a-2	2
54	6864	slc16a-26 [Smediterranea]	2
55	12149	slc25a-24 [Smediterranea]	2
56	1296	beta 1,3 galactosyltransferase-1 [Smediterranea]	2
57	17233	X1.C.A4.1, partial [Smediterranea]	2
58	5468	sh3bp2 [Smediterranea]	2
59	16525	Piwi-like protein 2	2
60	12986	cathepsin B-like protease [Smediterranea]	2
61	11587	slc35c-2 [Smediterranea]	2
62	22651	cathepsin L-like protease [Smediterranea]	2

Supplemental Table 2: 27kDa band labeled by DCG-04 is SmedCB. Whole worm lysate was labeled with the activity-based probe DCG-04 and separated via SDS-PAGE. A 27kDa labeled band was excised, tryspinized, and sequenced via mass spectrometry. Samples of worm regurgitant were collected, concentrated, and trypsinized for mass spectrometry. The top 25 peptides found are listed; * denotes carbamidomethyl group, ' indicates oxidation of methionine residues.

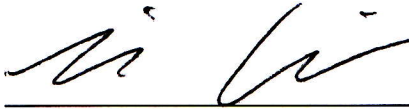| #	Unique pep	% Cov	Peptide	Protein MW	Species	Protein Name
1	7	22.4	LPASFDSR	37946	SCHMD	SmedCB
2	7	22.4	DQSNC*GSC*WAFGAVEAMTDR	37946	SCHMD	SmedCB
3	7	22.4	ISAEDLLTC*C*GFR	37946	SCHMD	SmedCB
4	7	22.4	SGVYQHVSGAM'LGGHAIK	37946	SCHMD	SmedCB
5	7	22.4	SGVYQHVSGAMLGGHAIK	37946	SCHMD	SmedCB
6	7	22.4	KGTNEC*GIEDEVVAGIPK	37946	SCHMD	SmedCB
7	7	22.4	GTNEC*GIEDEVVAGIPK	37946	SCHMD	SmedCB
8	7	10.2	IDNKPLEIEGR	95828	SCHMD	Piwi-like protein 2
9	7	10.2	GTATPTNYNVLEDTR	95828	SCHMD	Piwi-like protein 2
10	7	10.2	VLGGFGTTLQR	95828	SCHMD	Piwi-like protein 2
11	7	10.2	LPGAFC*FDGR	95828	SCHMD	Piwi-like protein 2
12	7	10.2	NSGPTFFETNSFK	95828	SCHMD	Piwi-like protein 2
13	7	10.2	NFNLALEDFK	95828	SCHMD	Piwi-like protein 2
14	7	10.2	GNGFPVPQIIYVIVK	95828	SCHMD	Piwi-like protein 2
15	6	31.2	ELPLTESLALTIDR	28557	ECOLI	2,3-bisphosphoglycerate-dep. PGM
16	6	4	VVGAM'QLYSVDR	192019	DUGJA	Clathrin heavy chain
17	6	4	RPISADSAIMNPISK	192019	DUGJA	Clathrin heavy chain
18	6	4	LHIIEVGQTPTGNQTFTK	192019	DUGJA	Clathrin heavy chain
19	6	4	LHIIEVGQTPTGNQTFTK	192019	DUGJA	Clathrin heavy chain
20	6	4	VVGAMQLYSVDR	192019	DUGJA	Clathrin heavy chain
21	6	4	SHNLTEDVTFWK	192019	DUGJA	Clathrin heavy chain
22	6	4	TLQIFNIELK	192019	DUGJA	Clathrin heavy chain
23	5	7.2	NEFHENLGK	92940	SCHMD	Piwi-like protein 1
24	5	7.2	DGVGDSQLAFTK	92940	SCHMD	Piwi-like protein 1
25	5	7.2	VLNENNVVSVYR	92940	SCHMD	Piwi-like protein 1

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

_____

Author Signature

08/08/16

Date