Title
CHARACTERIZATION OF THE ROLE OF CERCARIAL ELASTASE IN SCHISTOSOME HOST INVASION AND IMMUNE EVASION

Author
Ingram, Jessica

Publication Date
2011

Peer reviewed|Thesis/dissertation

CHARACTERIZATION OF THE ROLE OF CERCARIAL ELASTASE IN SCHISTOSOME HOST INVASION AND IMMUNE EVASION

by

JESSICA INGRAM

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOCHEMISTRY AND MOLECULAR BIOLOGY

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

[redacted]

ACKNOWLEDGEMENTS

Many thanks to Professors Jim McKerrow, Charly Craik and Shaun Coughlin for all their guidance in this project; to Judy Sakanari, K.C. Lim, Giselle Knudsen, Elizabeth Hansell and Anthony O'Donoghue for their continually help and advice; to Mari Nishino, Kyriacos Koupparis, Ryan Swenerton, Shuyi Zhang, Debbie Ruelas and Zac Mackey for making the lab a great place to work these past few years.

The material presented in Chapter 2 is part of a manuscript to be submitted to *PloS Neglected Tropical Diseases* in December 2011 titled: Investigation of the proteolytic functions of an expanded cercarial elastase gene family in *Schistosoma mansoni.* The co-authors are Ingram J, Rafi S, Lambeth L, Hsieh I, Lim KC, Ruelas D, Sakanari J, Craik C, Jacobson M and McKerrow, JH.

The material presented in Chapter 3 is part of a manuscript to be submitted to *Nature Chemical Biology* in December 2011 titled: Multiplexed Profiling of Protease Specificity Using a Physicochemically Diverse Synthetic Peptide Library and Mass Spectrometric Detection. The co-authors are O'Donoghue A, Eroy-Reveles A, Knudsen G, Ingram J, Zhou M, Moccand C, Maltby D, Anderson M, Fitzgerald T, McKerrow JH, Burlingame AL and Craik CS.

The material presented in Chapter 4 is part of a manuscript previously published in *PloS Neglected Tropical Diseases*: Ingram J, Knudsen G, Lim KC, Hansell E, Sakanari J, McKerrow JH. 2011. Proteomic Analysis of Human Skin Treated with Larval Schistosome Peptidases Reveals Distinct Invasion Strategies among Species of Blood Flukes. *PLoS Negl Trop Dis* 5(9): e1337.

CHARACTERIZATION OF THE ROLE OF CERCARIAL ELASTASE IN SCHISTOSOME HOST INVASION AND IMMUNE EVASION

JESSICA INGRAM

ABSTRACT

Schistosomes and related blood flukes are platyhelminth (flatworm) parasites of hosts ranging from fish to humans. Cercarial elastase is an evolutionarily distinct serine protease that is essential for invasion of human skin by the larva of *Schistosoma mansoni*. Several isoforms of the protease exist, but little is known about their genomic organization and regulation. Previous work suggests that the protease is able to degrade a variety of host macromolecules, including dermal and immune components. The expansion of the gene family and the indication that various protease isoforms had different substrate specificities presents two alternative evolutionary scenarios: that gene duplication had allowed for differential regulation and diversified substrate specificity, in order to facilitate multiple functions in host invasion; or that SmCE gene duplication directly enhanced fitness by simply increasing the amount of protease present. Work presented as part of this thesis project indicates that expansion of the cercarial elastase gene family is an example of gene dosage. In addition, work in other schistosomes species, including those that infect humans and those that are avian-specific, suggest that different schistosome lineages have evolved to use different classes of proteases in cercarial invasion.

TABLE OF CONTENTS

Acknowledgements		iii
Abstract		iv
Table of Contents		v
List of Tables		viii
List of Figures		ix
CHAPTER 1		
INTRODUCTION		
1.1	Schistosomiasis	1
1.2	The life cycle of *Schistosoma mansoni*	2
1.3	Invasion of skin by *S. mansoni* cercariae	5
1.4	Cercarial Dermatitis	7
1.5	The role of proteases in skin invasion	7
1.6	Cercarial Elastase	8
1.7	Other proteolytic activities in cercarial secretions	11
1.8	Specific Aims of Thesis Project	13
	Figures for Chapter 1	16
CHAPTER 2		
INVESTIGATION OF THE PROTEOLYTIC FUNCTIONS OF AN EXPANDED CERCARIAL ELASTASE GENE FAMILY IN *SCHISTOSOMA MANSONI*		
2.1	Author Summary	25
2.2	Introduction	25
2.3	Methods	27
2.3.1	Collection of parasite material	27
2.3.2	RNA and protein isolation	29
2.3.3	Quantitative RT-PCR analysis	29
2.3.4	Western blot analysis	30
2.3.5	Activity-based probe labeling	31
2.3.6	Snail section staining	32
2.3.7	Computational modeling	32
2.4	Results	33
2.4.1	All SmCE isoforms are expressed primarily in the intramolluscan and larval stages	33
2.4.2	SmCE is activated prior to exit from the intermediate snail host	35
2.4.3	Computational modeling predicts key differences in the SmCE isoforms substrate-binding site	36
2.5	Discussion	38
	Figures for Chapter 2	42

CHAPTER 3
GLOBAL PROTEOLYTIC ANALYSIS OF *SCHISTOSOMA MANSONI* CERCARIAL SECRETIONS
3.1 Chapter Summary 52
3.2 Introduction 53
3.3 Experimental Procedures 53
3.3.1 Cercarial secretion isolation 53
3.3.2 PLUSS library analysis of purified SmCE 1a/b and SmCE2a 54
3.3.3 Multiplex Peptide Cleavage Assay 55
3.3.4 Peptide Cleavage Site Identification by Mass Spectrometry 56
3.3.5 Heat Map Generation and Motif Analysis 57
3.4 Results 57
3.5 Discussion 59
Figures for Chapter 3 61

CHAPTER 4
PROTEOMIC ANALYSIS OF HUMAN SKIN TREATED WITH LARVAL SCHISTOSOME PEPTIDASES REVEALS DISTINCT INVASION STRATEGIES AMONG SPECIES OF BLOOD FLUKES
4.1 Author Summary 66
4.2 Introduction 67
4.3 Methods 69
4.3.1 Phylogenetic analysis of schistosome cercarial elastase and cathepsin B2 proteins 69
4.3.2 Purification of S. mansoni CB2 and CE 70
4.3.3 Ethics Statement 71
4.3.4 Skin digestion 71
4.3.5 Proteomic/Mass spectrometry analysis 72
4.3.6 Human Collagen I and Complement C3 cleavage and N-terminal Sequencing 74
4.4 Results 75
4.4.1 Identification and phylogeny of cercarial elastase and cathepsin B2 isoforms 75
4.4.2 Comparative proteomic analysis of human skin treated with SmCE and SmCB2 76
4.4.3 In vitro digestion of human collagen I and complement C3
4.5 Discussion 80
Figures for Chapter 4 86

CHAPTER 5
CONCLUSIONS AND FUTURE DIRECTIONS
5.1 Conclusions 102
5.2 Future directions 103
5.2.1 Expression of active, recombinant cercarial elastase 103
5.2.2 Crystallographic studies of cercarial elastase 104
5.2.3 Further functional characterization of secreted proteases

in other schistosome species 105
5.2.4 Vaccine trials with cercarial elastase 106

CHAPTER 6
REFERENCES
6.1 Bibliography 107

LIST OF TABLES

Table 4.1 Substrates of SmCE and SmCB2 identified in *ex vivo* skin
Table S4.1 Complete list of schistosome cathepsin B sequences
Table S4.2 Complete list of schistosome cercarial elastase sequences.

LIST OF FIGURES

Figure 1.1 Phylogeny of *Schistosomatidae* family.
Figure 1.2 Geographic distribution of schistosomiasis
Figure 1.3 A Ugandan boy with severe hepatosplenic schistosomiasis
Figure 1.4 The life cycle of *Schistosoma mansoni*
Figure 1.5 Diagram of human skin
Figure 1.6 Diagram of glandular systems present in the cercarial head
Figure 1.7 Timeline of cercarial invasion of skin
Figure 1.8 Sequence specificity of SmCE
Figure 2.1 All SmCE isoforms show similar mRNA and protein expression patterns throughout the *S. mansoni* life cycle
Figure 2.2 SmCE is activated within the intermediate snail host
Figure 2.3 Protein sequence alignment reveals that SmCE isoforms can be classified as belonging to one of two main groups
Figure 2.4 Modeling of qualitative electrostatic potential of the surface of SmCE1a.1 and SmCE2b shows distinct differences in the areas proximal to the active site
Figure 2.5 Computational modeling predicts key differences in the substrate binding pocket among SmCE isoforms
Figure S2.1 Ponceau stain of anti-SmCE immunoblot
Figure S2.2 Ponceau stain of avidin-HRP immunoblot
Figure S2.3 SmGAPDH is stably expressed throughout the *S. mansoni* life cycle
Figure S2.4 Anti-SmCE antibody and biotin-nVPL-$(OPh)_2$ do not cross-react with uninfected snail tissue
Figure 3.1 Inhibition profiles of *S. mansoni* cercarial secretions treated with various protease inhibitors, as tested against the generic substrate FITC-Casein.
Figure 3.2 Proteolytic activities of cercarial holo-secretions as compared to background activity from the snail environment.
Figure 3.3 P3 - P3’ substrate specificity of SmCE type I isoforms as determined by PLUSS library.
Figure 3.4 P3 - P3’ substrate specificity of SmCE type 2a isoform as determined by PLUSS library.
Figure 4.1 Differential expansion of select peptidase gene families in schistosomes species
Figure 4.2 Human collagen I is preferentially cleaved by SmCB2
Figure 4.3 Human Complement C3 is cleaved by both SmCE and SmCB2
Figure S4.1 Preparative and analytical SmCE SDS-PAGE gels.
Figure S4.2 Preparative and analytical SmCB2 SDS-PAGE gels
Figure S4.3 Phylogenetic analysis of all known schistosome cathepsin B proteins
Figure S4.4 SmCB2 and SjCB2 protein alignment

CHAPTER 1

INTRODUCTION

1.1 Schistosomiasis

Schistosomes and related blood flukes are platyhelminth (flatworm) parasites of hosts ranging from fish to humans (Figure 1.1). The five major species of schistosomes that infect humans--*Schistosoma mansoni*, *Schistosoma japonicum*, *Schistosoma haematobium*, *Schistosoma mekongi*, and *Schistosoma intercalatum*--are distributed across three continents, with some overlap among species (Figure 1.2). In humans, schistosome infection can lead to the disease schistosomiasis (or "bilharzia"). The disease can be divided into two phases: the acute phase ("Katayama fever"), marked by fever and flu-like symptoms, is due to a host immune response to adult parasites and early egg production; and the chronic phase, which is caused by the host's response to parasite eggs deposited in tissues by the adult worm. Immune-induced granuloma formation around the egg can lead to fibrosis and blocked blood flow. Depending on the infecting species, different organ systems can be affected. Recent epidemiological studies have shown that most infected individuals have a spectrum of clinical problems which can include fatigue, stunted development, diarrhea, under-nutrition, infertility and sexual dysfunction [1]. In the most severe cases *S. mansoni* and *S. japonicum* infections, the granulomatous response to eggs results in periportal cirrhosis and portal hypertension (Figure 1.3). In an analogous fashion, eggs of *S. haematobium*, the adults of which resides in

venules surrounding the bladder, can lead to cystitis, urinary tract blockage and, in severe cases, squamous carcinoma of the bladder.

Preventative measures for schistosomiasis center around improved water sanitation, and occasionally, use of molluscides to control the intermediate snail host population. In addition, a drug treatment, praziquantel, has been widely available since the mid-1970s. Due to high rates of re-infection, however, the disease remains a global health problem, with upwards of 200 million people infected [2]. Recent studies also suggest that schistosome infections may impact the etiology and transmission of other infectious diseases, including HIV/AIDS, tuberculosis and malaria [3,4]. It therefore contributes to the high socioeconomic burden imparted on endemic countries by these diseases. Despite various efforts, a vaccine is not currently available for the disease [5].

1.2 The life cycle of *Schistosoma mansoni*

Of all the species of schistosomes that infect humans, *Schistosoma mansoni* is the most well studied in the laboratory setting. Its life cycle, depicted in Figure 1.4, is representative of the other species that infect humans. In *S. mansoni*, adults migrate from the hepatic portal vein into the mesenterics of the intestine where females produce eggs that eventually exit with feces. When released into fresh water, the low osmolarity induces egg hatching, and a motile larval stage, the miracidium, emerges. Miracidia swim through water by ciliary action, and are positively phototropic. Upon coming into contact with their intermediate snail host

(*Biomphalaria glabrata*), miracidia are induced to swim faster and penetrate the snail. The signals that initiate this are unknown.

Upon snail penetration, the miracidium quickly sheds its ciliated plates and transitions into a mother sporocyst. This occurs near the site of initial infection, often in the tentacles of the snail. Germinal cells within the mother sporocyst then develop into daughter sporocysts, which bud off from the mother and migrate to the hepatopancreas. Mother sporocysts will continue to produce daughter sporocysts for six to seven weeks; in addition, daughter sporocysts can also give rise to subsequent generations of daughter sporocysts by asexual reproduction, thus greatly expanding their numbers in the infected snail [6].

Four to six weeks after initial miracidial infection, another larval stage, called the cercaria(e), forms within the daughter sporocyst. Like miracidia, the cercariae are positively phototropic, and light induces their release from the snail into water. The mechanism by which this occurs is unknown, but it has been suggested that proteolysis aids in cercarial egress from the snail [7].

Once cercariae are “shed” from the snail host into an aquatic environment they float up and swim down the water column, thus increasing the likelihood of human contact. While cercariae respond to cues released from skin, including warmth, short chain fatty acids, and short peptides with n-terminal arginines, these do not cause direct chemotaxis; instead, these cues seem to signal the

cercaria to more rapidly change the direction of its swimming, alternating between forward and backward swimming, in a process referred to as chemolinokinesis [8-10]. Upon contact with the human host, cercariae initially creep along the surface of skin before initiating direct penetration. Host fatty acid signaling also responsible for initiating skin penetration, by triggering the release of secretions from the acetabular glands. Secretion release is thought to be mediated by prostaglandin signaling in the parasite that causes contractions of the smooth muscle surrounding the glandular network[11]. Aided by these secretions, the cercariae then actively migrate though skin to enter dermal blood vessels. This process can take anywhere from several minutes to many hours, and also involves shedding of the cercarial tail and the glycocalyx, a protective carbohydrate coat [12]. These two events mark the transition of the cercaria into a juvenile worm, called a “schistosomulum”. The process of skin invasion is more extensively described in Section 1.3.

The schistosomulum continues its migration through the vasculature of the host, tracking with the blood flow. This leads it on a path to the heart, through the lungs, and back through the heart into the aorta. The worm will eventually migrate to the hepatic portal vein (or to the veins of the bladder, in *S. haematobium* infection), where it matures into an adult worm. Adult worms are dimorphic, and exist *in copula*, with the larger male enclosing the thinner female in a groove running the length of its body, termed the gynecophoral canal. Successful mating leads to the production of eggs. Eggs then pass through the

intestinal wall into the lumen of the gut or bladder, and are eventually passed out of the host, thus completing the cycle.

1.3 Invasion of skin by *S. mansoni* cercariae

Human skin is a complex barrier. In addition to the mechanical barrier of structural proteins in the epidermis, basement membrane and dermal extracellular matrix, both the epidermis and dermis are bathed in plasma proteins, including early sentinels of the immune system (Figure 1.5) [13]. Unlike many other pathogens that utilize trauma or an insect bite to gain entry to skin, cercariae are able to directly penetrate their host [12,14].

The cercaria is 300 μm long, 70 μm wide and comprised of roughly 1000 cells [15]. Its body is covered in a hydrated, carbohydrate-rich coat and a series of actin-containing spines that are directed posteriorly. It is distinguished by its long, forked tail, an acetabulum (or “ventral sucker”) and three sets of glands that open via a series of ducts at the anterior end of the larva: the pre- and post-acetabular glands, and the head gland (Figure 1.6). These glands dominant the internal structures of the cercaria, comprising the majority of its volume. Previous research suggests that the secretions from all glands participate in host invasion [16,17].

Chemical signals released from human skin, predominantly medium chain fatty acids, prompt cercarial invasion. This signaling directs cercariae to secrete

vesicles containing a variety of proteins and an adhesive, mucin-like substance from their acetabular glands, which aids in initial attachment to the skin [10]. Concurrently, signaling induces vigorous tail movement; this probing motion aids in entry of the most superficial epidermis which is hydrated in fresh water. Tails are usually shed as the cercaria moves deeper into the skin, away from the point of entry. Tail shedding is thought to initiate shedding of the glycocalyx, the carbohydrate coat that protects the cercaria from the low osmolarity of fresh water, but is no longer needed as the parasite adapts to its definitive host [12]. Shedding of the glycocalyx is also key to evading the host immune response as it is a potent inducer of complement activation.

While cercariae are able to move freely between dead cells at the surface of the skin (stratum corneum), they must degrade cell-cell tight junctions (desmosomes) in the mid-epidermis. Their migration is thus stalled significantly as they approach the basement membrane. The protein matrix barrier of the basement membrane is comprised of cross-linked, fibrillar proteins that must also be degraded in order to reach the underlying dermis, and breaching this barrier can take several hours. Once the cercaria has broken through, it penetrates further into the dermal layer until reaching a blood vessel, which it must also successful breach in order to enter the host vasculature. The contents of the acetabular glands are slowly released during the entry of cercariae. Some secreted material still remains days after entry, as shown in *in vitro* assays and electron micrograph studies of lung stage schistosomulae [18].

1.4 Cercarial Dermatitis

In addition to the pathologies caused by human-specific schistosomes, schistosomes with small mammalian or avian definitive hosts, such as *Trichobilharzia spp*., often cause a dermatological response in humans. This occurs when the larval stage of the parasite attempts to penetrate human skin. The exact factors that induce cercarial dermatitis are unknown. The release of parasite antigens presumably leads to a pronounced human host immune response, characterized by pruritic maculopapular eruptions. Prior exposure to larval infection is necessary for priming the immune system to produce this response. Referred to as cercarial dermatitis, or “swimmer’s itch,” the reaction normally resolves within 24-72 hours. Cercarial dermatitis is wide-spread, and frequent outbreaks occur in both the United States and Europe.

1.5 The role of proteases in skin invasion

Given the substantial protein barriers to invasion of that cercariae, it was apparent to several early investigators that the parasite would require proteolytic activity to breach these barriers. As early as 1921, it was reported that the acetabular glands had a role in “tissue dissolving” during cercarial penetration [19,20]. This theory was subsequently tested in the 1930s, when an enzymatic extract of homogenized cercariae was used to digest frog tissue, and subsequently, human skin. The discovery that cercariae could be induced to secrete the contents of their acetabular glands *in vitro*, in response to human

sebum or linoleic acid, marked the advent of enzymatic purification and biochemical characterization of cercarial secretions [21].

A major proteolytic activity was first purified from acetabular secretions by gel filtration, and subsequently by isoelectric focusing and ion exchange [22-25]. Soon after, advances in molecular cloning allowed for the isolation of cDNA encoding the enzyme corresponding to this activity, which was termed cercarial elastase (SmCE) [26]. Biochemical characterization revealed that the protease has activity against insoluble elastin and other fibrillar macromolecules of skin [27]. Moreover, in studies where a specific, irreversible protease inhibitor was applied to *ex vivo* skin, the majority of cercariae were blocked from invading, suggesting that SmCE had an essential role in *S. mansoni* skin penetration [28].

More recently, proteomic studies have identified the majority of proteins secreted by *S. mansoni* cercariae, from all three glands. These include several classes of histolytic peptidases [29,30]. Commiserate with earlier biochemical studies, proteomic studies confirmed that the most abundant protease in *S. mansoni* secretions is SmCE [31].

1.6 Cercarial Elastase

SmCE is classified as an S1A serine protease. Serine proteases of this type--notably trypsin, chymoptrypsin and mammalian pancreatic elastase—are among the best-characterized proteases, and X-ray crystallographic structures exist for

each. The catalytic mechanism of these proteases is based on an active site serine residue, one of three polar residues that form the catalytic triad--His57, Asp102 and Ser195 (chymotrypsin numbering system). In peptide hydrolysis, Asp102 brings His57 into the correct orientation through hydrogen-bonding. His57 then can then act as a general base, withdrawing a hydrogen from the hydroxyl group of Ser195, increasing its nucleophilicity. Ser195 is then able to initiate a nucleophilic attack on the scissile peptide bond of the protein substrate. As a second step to the reaction, donation of a hydrogen from His57 to the amide nitrogen of the peptide creates a covalent tetrahedral intermediate, which rapidly breaks down, leading to the dissociation of the amine product.

In addition to the residues that make up the catalytic core, accessory residues which line the substrate-binding pocket have key roles in determining substrate specificity, i.e. the preferred residues neighboring the scissile bond in the protein substrate. In chymotrypsin, for example, the presence of hydrophobic residues in the substrate-binding pocket favors binding of aromatics in the position proximal to the scissile bond, towards the amino terminus of the protein (i.e. the P1 position). For a full description of substrate binding nomenclature, see Figure 1.8.

SmCE, which was termed an elastase for to its ability is to cleave insoluble elastin, has been predicted to be structurally similar to trypsin, and to contain a canonical trypsin-like fold [32]. Its substrate preference is similar to that of chymotrypsin, in that it prefers substrates with hydrophobic residues like

phenylalanine in the P1 position. Early studies by McKerrow et al profiled SmCE activity against marcomolecular substrates, including elastin, fibronectin and collagen type IV; this work suggested that the enzyme may have an extended binding site, with strong amino acid preferences for both the P1-P4 binding pockets as well as the corresponding P1'-P4' sites (Figure 1.8) [27,33,34]. This in turn suggests that the protease may have additional activity that unwinds fibrillar substrates, like those present in skin, in order to facilitate cleavage.

In addition to the structural proteins cleaved by SmCE, work by Pleass et al showed that the enzyme was able to cleave human IgE protein *in vitro* [35,36]. This suggested that the protease may have an additional role in host invasion, actively degrading immune components that might target the parasite as it migrates through skin. Moreover, proteomic analysis of *ex vivo* skin exposed to live cercariae identified a number of complement proteins that were degraded during invasion, although this observation was not directly tied to SmCE activity [37].

Previous work identified five isoforms of cercarial elastase in a *S. mansoni* cDNA and genomic libraries [34]. Based on sequence identity, these isoforms were classified as belonging to one of two groups: SmCE1a, b and c; and SmCE 2a and b. Corresponding proteins were identified for SmCE 1a, 1b and 2a; SmCE 1c is a pseudogene, and no protein was identified for SmCE2a. Further chromatographic purification of SmCE1a and 1b allowed for comparative studies on their respective substrate specificities, and each was independently scanned

against a fluorescent tetrapeptide library for the P1-P4 substrate preferences [34]. The results indicated that the two isoforms were isozymes, with subtle differences in substrate specificity. In addition, multiple homologs of SmCE were found in *S. haematobium,* and more recently, a single CE isoforms was identified in *S. japonicum.*

Recent sequencing and annotation of the *S. mansoni* genome indicates that the parasite encodes several additional SmCE isoforms [38]. A total of ten full-length (i.e. encoding the full catalytic core of the enzyme) isoforms are present, although they fall into the same classification scheme (type 1 vs type 2) introduced by Salter et al [34]. Further work was needed to elucidate the function of these additional isoforms. Given the number of isoforms in *S. mansoni*, and the indication that different isoforms of the protease have different substrate specificities, it is possible that they are differentially regulated in order to facilitate multiple functions of infection. An alternative explanation is that all isoforms may have identical substrate preferences and localization, and SmCE gene duplication may simply enhance fitness by directly increasing the amount of transcript present.

1.7 Other proteolytic activities in cercarial secretions

Several additional proteases have been identified in *S. mansoni* cercarial secretions. These include leishmanolysin-like metalloproteases (SmM8s). This class of protease was only present in studies where cercarial secretions were

collected at later time points (longer than 3 hours), and preliminary studies indicate that it localizes the to head gland [30] (*Judy Sakanari, unpublished data*). This is in contrast to SmCE, which localizes to both the pre- and post-acetabular glands, and which appears to be secreted earlier in the invasion process [29,31]. In addition, several other, lower abundance proteases have also been identified in cercarial secretions. These include a calpain-like protease, and a number of exopeptidases. The biological functions of these enzymes has not yet been elucidated.

There is also evidence that other species of schistosomes do not use an SmCE-like protease in host invasion. The zoonotic species *S. japonicum* contains no serine protease activity in its larval secretions [39]. *S. japonicum*, however, does encode several isoforms of cathepsin B2, a cysteine peptidase, which are secreted by the invading parasite [39]. Moreover, orthologs of SjCB2 have been identified in the cercarial secretions of other, non-human schistosome species, including members of the genus *Trichobilharzia* [40]. This leads to the hypothesis that the invasive protease differs between schistosome species, and also suggest that the use of a serine protease in invasion is the exception, rather than the rule, among parasitic schistosomes.

The question arrises as to why the cercarial elastase gene family has expanded to such an extent in *S. mansoni,* while the closely related *S. japonicum* encodes only a single isoform of the protease. Have *S. mansoni* and *S. japonicum* evolved

to use different proteases in skin invasion? Are there additional, shared proteases present in cercarial secretions that also facilitate migration through skin?

1.8 Specific Aims of this thesis to address to differential use of cercarial elastase isoforms in *S. mansoni*, and other schistosome species

1.8.1 Are cercarial elastase isoforms differentially expressed throughout the *S. mansoni* lifecycle?

Using sequence data generated from the *S. mansoni* genome, I will generate primer sets and TaqMan probes specific to SmCE 1(a and b), SmCE 2a and SmCE 2b and monitor their mRNA expression levels throughout the *S. mansoni* life cycle with quantitative PCR. I will then correlate this with total SmCE protein levels through the life cycle, as determined by western blot analysis with an existing polyclonal anti-SmCE antibody. In addition, by using biotinylated diphenyl phosphonates (a set of activity-based probes that covalent label serine proteases) with peptide elements specific for SmCE cleavage, I will track the activity of SmCE throughout the life cycle. Together, this will help determine whether isoforms are differentially regulated, and whether they are regulated on the transcriptional, translational or post-translational level.

1.8.2 Do various SmCE isoforms have different substrate specificities? What other proteolytic activities are present in *S. mansoni* cercarial secretions, and how might these activities contribute to cercarial invasion?

In the absence of recombinant enzyme or purified native enzyme for each class of SmCE isoform, there is not currently a full analysis of isoform substrate specificities. Using a new library developed by the Craik lab at UCSF, termed the Peptide Library for Universal Substrate Specificity (PLUSS), I will compare the prime and non-prime side substrate specificities between native purified SmCE 1 (a and b) and SmCE 2a. Concurrently, by computer modeling, I will compare the peptide-binding sites of SmCE1(a and b), SmCE2a and SmCE2b to look for differences that might influence substrate specificity.

In addition, I will use the PLUSS library to profile the proteolytic specificity of SmCE cercarial secretions as a whole. Then, by treating collected secretions with various classes of protease inhibitors, I will dissect the different activities present, in an attempt to better understand their respective contributions to substrate cleavage.

1.8.3 What are the *in vivo* substrates of *S. mansoni* cercarial elastase? Do other schistosome species, purported to use other proteases during cercarial invasion, degrade a similar host of proteins as they migrate through skin?

It will also be informative to test what substrates SmCE cleaves *in vivo*, and compare this to previous data obtained from skin exposed to whole cercaria, in order to better understand the role of SmCE in invasion. This will be done using a proteomic approach with isolated, *ex vivo* human skin. A mixture of purified SmCE isoforms will be incubated with ex vivo skin for several hours, and soluble protein will then be harvested, trypsin digested and analyzed by LC/MS/MS. As a control, soluble fractions from skin treated with inhibited SmCE will be collected and similarly analyzed. The peptides generate from both the treated and untreated samples will be compared, and should elucidate some of the *in vivo* targets of SmCE. If new targets are identified, their cleavage will then be further characterized by *in vitro* cleavage assays.

Given the finding that the *S. japonicum* genome encodes only a single SmCE isoform, together with previous data suggesting that *S. japonicum* and *Trichobilharzia* species use a cysteine protease as their invasive enzyme, I will perform a simultaneous proteomic study to determine the *in vivo* targets of cathepsin B2. This will then be used to determine whether these two classes of protease cleave similar substrates in human skin.

CHAPTER 1
FIGURES AND TABLES

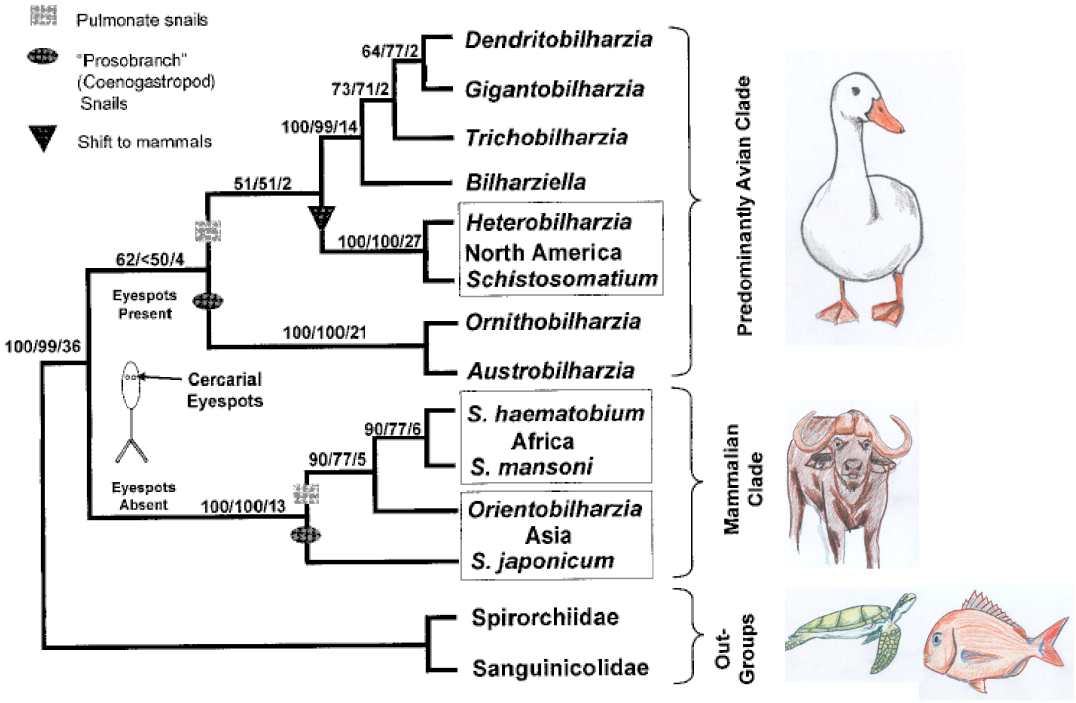


Figure 1.1 Phylogeny of Schistosomatidae family. This tree highlights several methods used to group schistosomes, including the presence or absence of cercarial eyespots, the gastropod intermediate host and the definitive host. Boxed groups indicate a restricted geographical distribution. Adapted from *Snyder and Loker, 2000.*

	Species	Geographical distribution
Intestinal schistosomiasis	*Schistosoma mansoni*	Africa, the Middle East, the Caribbean, Brazil, Venezuela, Suriname
	Schistosoma japonicum	China, Indonesia, the Philippines
	Schistosoma mekongi	Several districts of Cambodia and the Lao People's Democratic Republic
	Schistosoma intercalatum and related *S. guineansis*	Rain forest areas of central Africa
Urogenital schistosomiasis	*Schistosoma haematobium*	Africa, the Middle East

Figure 1.2 Geographic distribution of schistosomiasis (*World Health Organization, 2011*).

Figure 1.3 A Ugandan boy with severe hepatosplenic schistosomiasis. ***Image credit: USAID NTD Program*****.** Upon infection with *S. mansoni,* an immune response against parasite eggs deposited in host tissue leads to granuloma formation, portal hypertension, and fibrosis. In severe cases this leads to enlargement of the spleen and liver, and retention of fluids in the abdominal cavity (ascites).

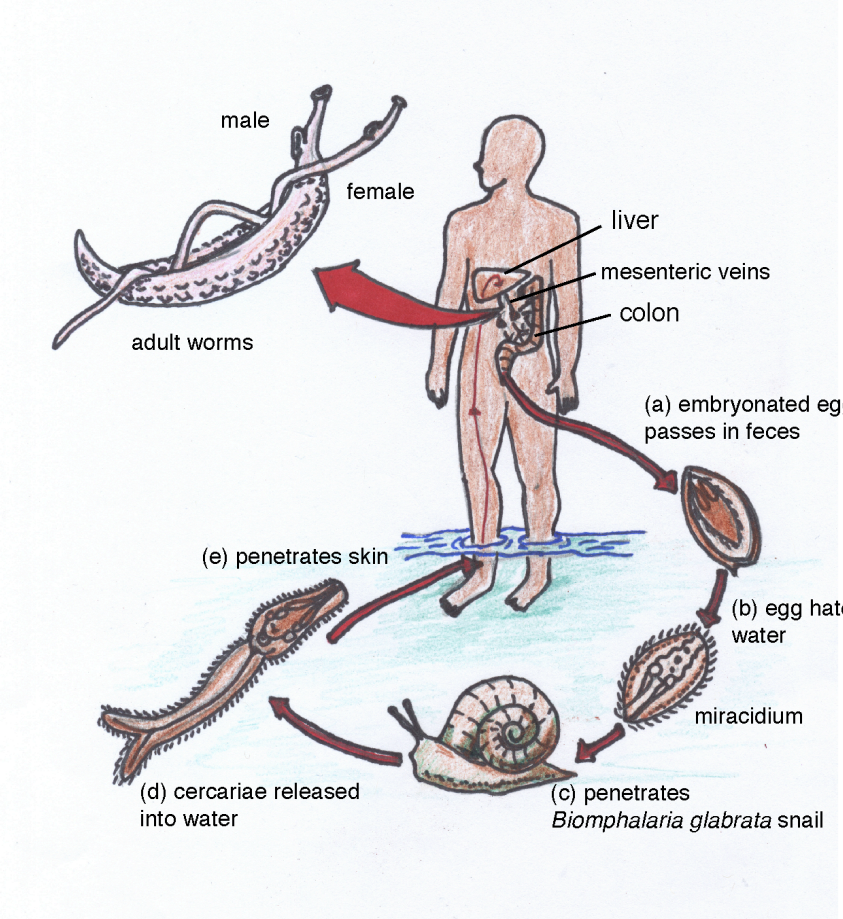


Figure 1.4 The life cycle of *Schistosoma mansoni,* adapted from *Foundations of Parasitology, 8th ed.* Embroynated eggs (a) are released with feces; in fresh water, the egg hatches, releasing the miracidium (b). The miracidium then infects an intermediate snail host (c) where it produces sporocysts that will give rise to a second larval form, the cercaria (d). The cercaria then infects a human host, via the skin (e). Cercariae then transform into schistosomulae, which eventually reach the vasculature system, and mature into adult worms. Adult worms exist *in copula*, and eggs are deposited in the mesenterics and traverse the gut wall into the intestinal lumen, completing the cycle.

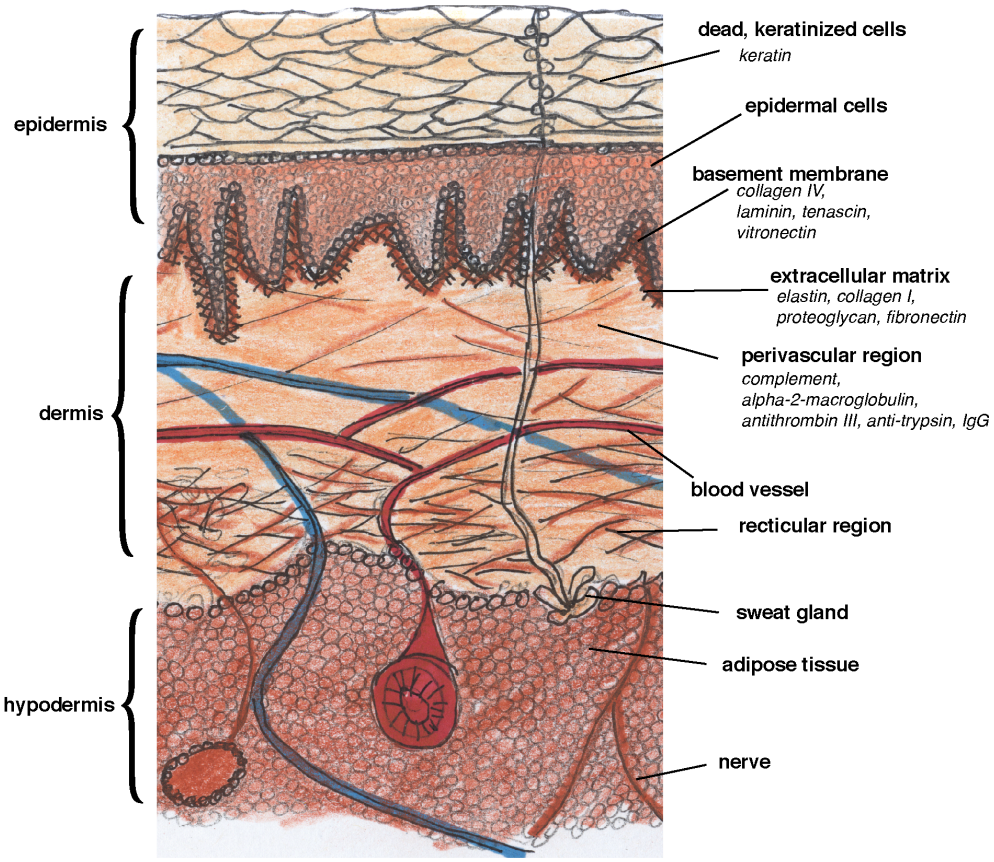


Figure 1.5 Diagram of human skin. Structural features of skin (bold) and the protein components that make up these features (italicized) are listed.

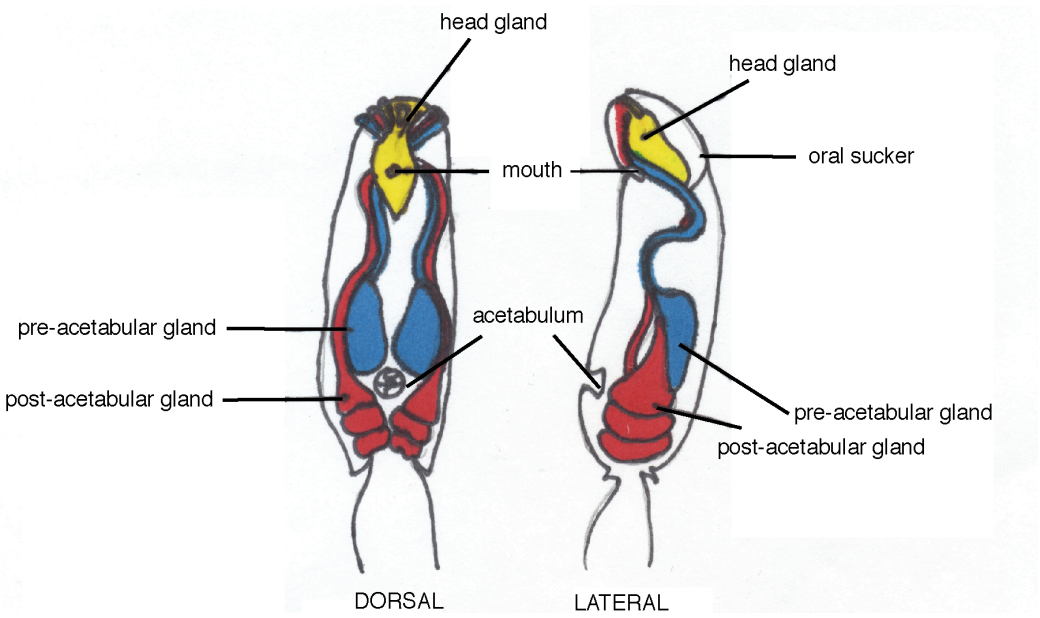
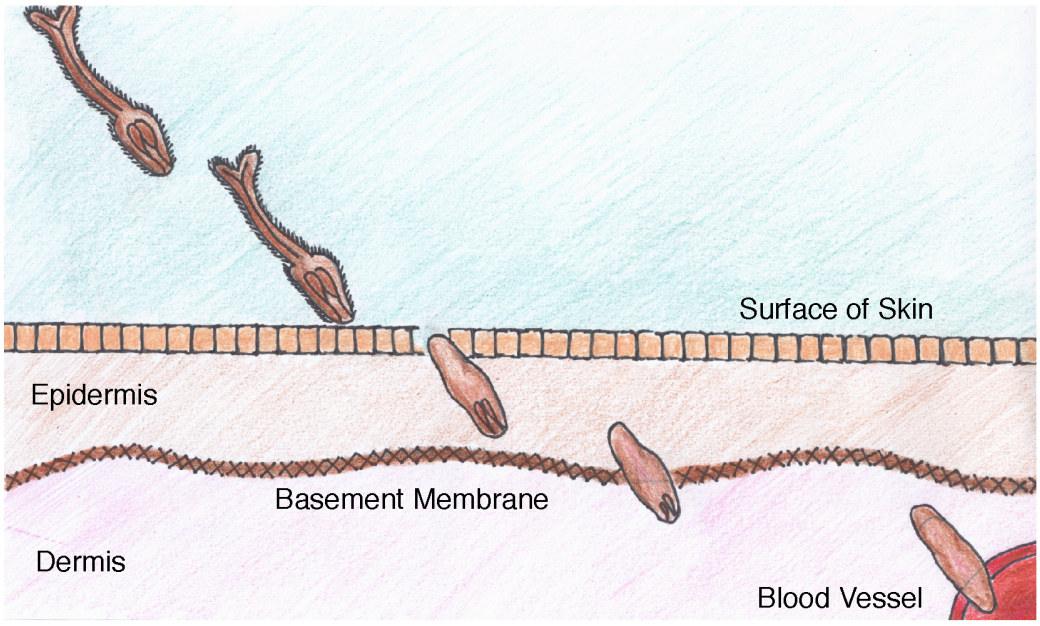


Figure 1.6 Diagram of glandular systems present in the cercarial head. The head of the cercaria includes three sets of glands: the pre- and post-acetabular glands, and the head gland. All three emerge as a series of ducts at the anterior tip of the larva.

Migration in Water	Initial Dermal Penetration	Transit to Basement Membrane	Basement Membrane Penetration	Vascular Entry
1-24hrs	5-10 mins	30-90 mins	5-24 hrs	1-24hrs



Figurere 1.7 Timeline of cercarial invasion of skin, *adapted from the thesis of Jason Salter, 2001.* After being shed from the intermediate snail host, cercariae have up to 24 hours to initiate skin penetration, at which point they begin to exhaust their energy stores. Initial penetration of superficial skin layers is a relatively rapid process on a time scale of minutes. The cercaria slows down significantly as it progresses through the lower layers of the epidermis and the basement membrane; crossing this barrier can take an additional 5-24 hours. Once the parasite has breached the basement membrane, migration through the dermis and eventual invasion of a blood vessel takes several additional hours.

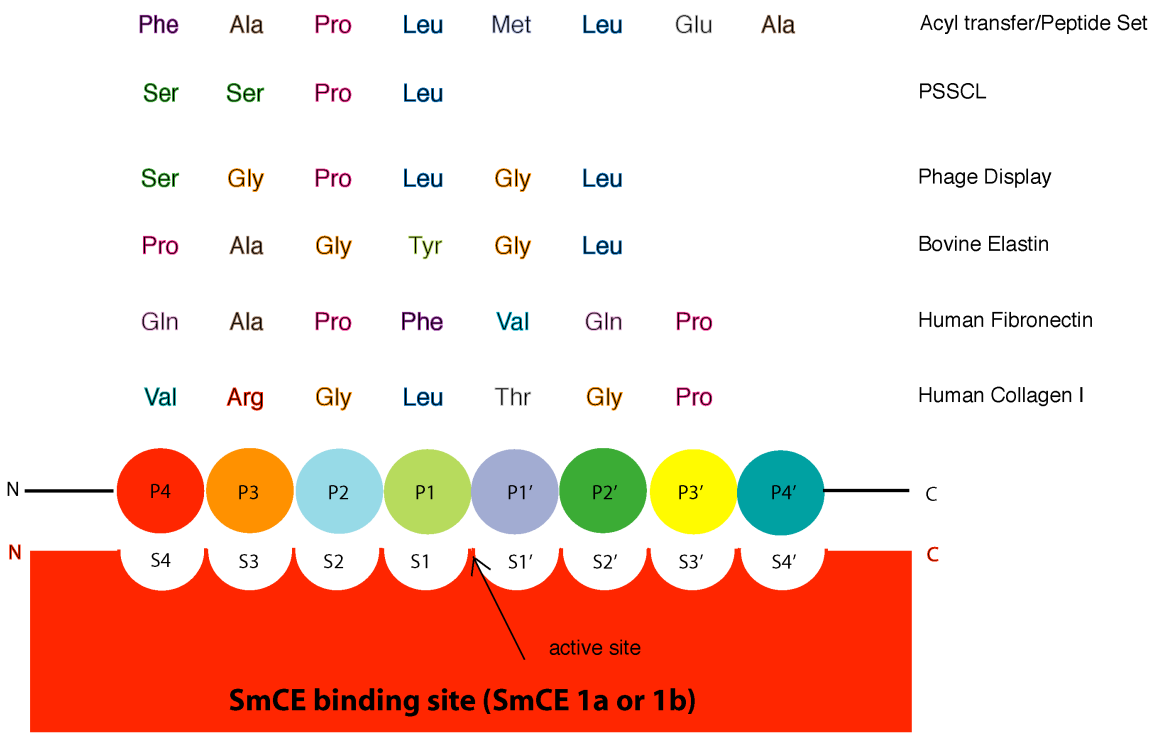


Figure 1.8 Sequence specificity of SmCE, as determined by profiling against multiple substrates. Previous studies has determine an extended binding site (P4-P4') against both macromolecular substrates and smaller, peptide-based libraries. The results of the these studies are shown above. Substrate binding nomenclature centers around the position of the scissile bond of the cleaved peptide. The amino acid position most proximal to the N(amino)-terminus of the scissile bond is the P1 position; the amino acid position most proximal to the C(carboxyl)-terminus of the scissile bond is the P1' position, etc. The corresponding substrate binding sites of the protease are termed S4-S4'.

CHAPTER 2
INVESTIGATION OF THE PROTEOLYTIC FUNCTIONS OF AN EXPANDED CERCARIAL ELASTASE GENE FAMILY IN *SCHISTOSOMA MANSONI*

(This work was submitted to *PLoS Neglected Tropical Diseases*, December 2011)

2.1 Author Summary

Schistosome parasites are a major cause of disease in the developing world. The larval stage of the parasite transitions between an intermediate snail host and a definitive human host in a dramatic fashion, burrowing out of the snail and subsequently penetrating human skin. This process is facilitated by secreted proteases. In *Schistosoma mansoni*, cercarial elastase is the predominant secreted protease, and essential for host skin invasion. Genomic analysis reveals a greatly expanded cercarial elastase gene family in *S. mansoni*. Despite sequence divergence, SmCE isoforms show similar expression profiles throughout the *S. mansoni* life cycle and have largely similar substrate specificities, suggesting that the majority of protease isoforms are functionally redundant and therefore their expansion is an example of gene dosage. However, activity-based profiling also indicates that a subset of SmCE isoforms are activated prior to the parasite's exit from its intermediate snail host, suggesting that the protease may also have a role in this process.

2.2 Introduction

Schistosomes and related blood flukes are platyhelminth parasites of hosts ranging from fish to humans. In humans, schistosome infection leads to the

disease schistosomiasis, which effects upwards of 200 million people worldwide [2]. Prevalent in developing nations, it is ranked second to malaria in terms of overall morbidity caused by parasitic disease [3,4]. The parasite has a complex life cycle, infecting both an intermediate snail host and a definitive human host. At multiple life cycle stages, the parasite must migrate through host tissue, and breach substantial structural barriers, including the extracellular matrix [41]. The process of early human infection is well-characterized: the multi-cellular larval stage, termed cercaria(e), directly penetrates host skin in a process facilitated by secretions from a glandular network that runs the length of the larval body. These secretions contain multiple histolytic proteases [29,30]. In *Schistosoma mansoni*, the most of abundant of these is cercarial elastase (SmCE), an S1A serine protease named for its ability to digest insoluble elastin.

Completion of the *S. mansoni* genome sequence revealed an expanded cercarial elastase gene family, including eight full-length genes. Based on sequence identity, these isoforms were classified as belonging to one of two groups: Group I, comprised of SmCE1a, b and c; and Group 2, comprised of SmCE 2a and b. Corresponding proteins were identified for SmCE 1a, 1b and 2a. SmCE 1c was determined to be a pseudogene, and no corresponding protein has been identified for SmCE2b transcript. Although there is evidence of multiple CE isoforms in the related species *Schistosoma haematobium*, expansion of this family appears to be limited to the human-specific schistosomes: *Schistosoma japonicum* encodes only a single cercarial elastase gene, and there is substantial

evidence that this species, along with the avian-specific schistosomes, utilizes another class of protease in host invasion [34,39,40,42,43].

The expansion of the cercarial elastase gene family in *S. mansoni*, and the sequence divergence among isoforms, suggests two alternative evolutionary scenarios: that gene duplication has allowed for differential regulation or diversified substrate specificity, in order to facilitate multiple functions in host invasion; or that SmCE gene duplication directly enhances fitness by simply increasing the amount of protease present. In the absence of robust genetic tools for studying schistosome biology, we decided to take a biochemical approach to study the functions of SmCE isoforms, using a combination of quantitative PCR, activity-based profiling and computer modeling to identify unique characteristics of the isoforms. Here, we present data that supports the hypothesis that expansion of the SmCE gene family in *S. mansoni* directly increases gene dosage. Moreover, analysis of SmCE isoform expression and activity throughout the parasite life cycle confirms that the protease is present and active very early on in cercarial development, which suggests a potential role for the protease in egress from the intermediate snail host.

2.3 Methods

2.3.1 Collection of parasite material

Parasite material originates from a Puerto Rican isolate of *S. mansoni* maintained by BEI Resources (Rockville, MD), and passaged through the

intermediate snail host, *Biomphalaria glabrata,* and the Golden Syrian hamster *Mesocricetus auratus* (Simonsen Labs) or BALB/C mice (Jackson Labs, Bar Harbor, ME) at the University of California, San Francisco. Animals were maintained and experiments carried out in accordance with protocols approved by the Institutional Animal Care and Use Committee (IACUC) at UCSF. Adult worms were harvested as previously described [44]. Eggs were collected from infected livers harvested from euthanized hamsters. Briefly, livers were mechanically disrupted in a Waring blender on a low setting in 1x PBS/ trypsin. The resulting lysate was incubated at 37°C for 2 hrs, rinsed with 1x PBS, and liver material was allowed to sediment. Sedimented material was then transferred to a Petri dish, swirled, and eggs were collected from the center of the dish with a transfer pipette. Miracidia were collected by inducing egg hatching with light exposure for 30 min in distilled water. Miracidia swim towards the light source, and were collected from this area with a transfer pipette. Miracidia were then used to infect *Biomphalaria glabrata* snails *en masse*. Roughly 100 snails, just under 1 cm each, were incubated with several thousand miracidia overnight. Individual snails were then analyzed for signs of infection. Forty days *p.i.*, whole hepatopancreas were dissected from infected snails. Cercariae were harvested using a previously described method [45]. Collection of lung stage schistosomules was initiated by subcutaneous infection of 6-week-old BALB/C mice with several hundred cercariae. One week post-infection, mice were euthanized and schistosomules were harvested by perfusion.

2.3.2 RNA and protein isolation

All parasite material was stored at -80°C; a portion of each sample was stored in Trizol (Life Tech) for subsequent RNA extraction. Total RNA was isolated from frozen tissue by homogenization and incubation in Trizol at 65°C for 5 min. Total nucleic acid was subsequently collected by phenol/chloroform extraction. The aqueous phase was then loaded onto a Stratagene RNA purification column (Stratagene) and purified following the manufacturer's protocol, with the inclusion of an on-column DNase treatment step. First-strand cDNA was then generated from this material using the first-strand synthesis kit and oligo d(T) primer (Invitrogen). Remaining frozen samples were lysed in 100 mM Tris, pH 8 and 0.1% NP40, briefly sonicated to disrupt tissue, and spun down for 15 min at 16,000 rcf at 4°C; the resulting supernatant was saved as the soluble protein lysates.

2.3.3 Quantitative RT-PCR analysis

The following TaqMan primer/probe sets (LifeTech, Carlsbad, CA) were designed to distinguish between SmCE1(a and b), SmCE2a and SmCE2b mRNA sequences: SmCE1 Forward 5'-TTA AGG TGG CAC CAG GAT ATA TGC-3'; SmCE1 Reverse 5'-TGA GTG TCT GTG CGA TTG GT-3'; SmCE1 Probe 5'-TCG TGC CGA CAT ACA AG-3'; SmCE2a Forward 5'-CCA CCA CTG GGA ATC CTA TTT GT-3'; SmCE2a 5'-CCT GGT GCG GTG ATT TGC-3'; SmCE2a Probe 5'-AAG CGG CGT ATG TGT TC-3'; SmCE2b Forward 5'-GGC ATG CAG ACA CAA ACG T-3'; SmCE2b Reverse 5'-TGT CAC CTG GAC CAG CTA TTT G-3';

SmCE2b Probe 5'-CAG GTC CGA ACT CAA AG-3'. In addition, the following primers/probe were designed against SmGAPDH as a reference gene: SmGAPDH Forward 5'-ACT CAT TTA CGG CTA CAC AAA AGG T-3'; SmGAPDH Reverse 5'-CAG TGG AAG CTG GAA TAA TAT TTT GCA-3'; SmGAPDH Probe 5'-CTC GCC ATA ATT TTG-3'. Reactions were performed in 10 µL 2x ABI Gene Expression Mix, 1 µL Pre-mixed Primer/Probe, 7 µL dH_2O and 1 µL cDNA template. Cycle conditions were as follows: 95ºC for 10 min; 40 cycles of 95ºC for 15s, 60ºC for 1min. Reactions were run on an MX3005P quantitative thermocycler (Stratagene).

2.3.4 Western blot analysis

Bradford assays were performed to quantify total protein amount and an equal protein concentration of lysate from each life cycle stage was added to 4x reduced SDS-PAGE loading dye (LifeTech). A total volume of 15 µL was loaded onto a 10% Bis-Tris SDS-PAGE gel (LifeTech). Bands were then transferred by electroblotting to a PVDF membrane, and visualized by Ponceau staining to check for equal transfer of all samples (Figure S1). Blots were incubated with a 1:1000 dilution of polyclonal SmCE [45], followed by a 1:5000 dilution of anti-rabbit IgG-HRP (GE Healthcare) and 2ml ECL reagent (GE Healthcare). Blots were exposed to Kodak High Sensitivity BioFilm for visualization (Kodak, Rochester, NY).

2.3.5 Activity-based probe labeling

Bradford assays were performed to quantify total protein amount, and an equal protein concentration of lysate from each life cycle stage was used. For each sample, 50 µL lysate was added to 100 µL 50mM Tris, pH 8, 150mM NaCl. This mixture was then incubated for 5 min with 50 µL avidin beads (Pierce Biotechnology, Rockford, IL), pre-equilibrated in 50 mM Tris, pH 8. The beads were pelleted by centrifugation for 30 sec at 3000 rcf, and the soluble portion was moved to a new microfuge tube. The probe biotin-nVPL-O(Ph)$_2$ (biotin-PEG-norleucyl-valyl-prolyl-leucyl-diphenyl phosphonate) was synthesized following a previously described method and added for a final concentration of 5 µM, and incubated for 1 hour at room temperature [46]. To stop the reaction, 400 µL 1M guanidine-HCl was added and the entire reaction was run over a 5 kDa Amicon filter (Millipore, Bedford, MA). The filter was washed in 400 µL 1M guanidine-HCl, 400 µL 50 mM Tris, pH 8 and concentrated to a final volume of 30 µL. Ten microliters was then added to 6 µL SDS-PAGE reducing dye and loaded onto a 10% bis-TRIS SDS-PAGE gel (LifeTech). Bands were then transferred by electroblotting to a PVDF membrane, and visualized by Ponceau staining to check for equal transfer of all samples (Figure S2). Blots were incubated with a 1:2000 dilution of Avidin-HRP (Pierce) and 2 mL ECL reagent (GE Healthcare). Blots were then exposed to Kodak High Sensitivity BioFilm for visualization (Kodak, Rochester, NY).

2.3.6 Snail section staining

The hepato-pancreas containing mature schistosome daughter sporocysts were dissected from infected snails and placed in 4% paraformaldehyde in phosphate buffer pH 7.4. Following an overnight fixation at 4°C, the tissue was washed in 0.1 M phosphate buffer with 3% sucrose, dehydrated in graded acetones, and then infiltrated with glycol methacrylate monomer. The tissue was then transferred to molds containing the complete embedding mixture, and placed under vacuum for 12 hours. The entire procedure was carried out at 4°C. The hardened blocks were then section at 2μ and air-dried. To identify active protease in developing cercariae, the esterase procedures of Li *et al* for alpha-naphthylacetate (with and without sodium fluoride) and naphthol AS-D chloroacetate were performed at pH 8 [47]. To definitively identify activity as due to the serine protease of cercariae, two chloromethyl ketone inhibitors were employed. The first is a known inhibitor of the protease, AAPF-CMK (alyl-alyl-prolyl-phenyl-cholormethyl ketone), and the second is known not to inhibit the protease, FPR-CMK (phenyl-prolyl-argyl-cholormethyl ketone) [45]. Sections were pre-incubated with inhibitors for one hour prior to initiation of the esterase reaction. Cleaved substrate gives a distinctive color reaction (orange/red) visualized by light microscopy.

2.3.7 Computational modeling

We used the structure of bovine alpha-chymotrypsin refined at 1.68 Å resolution (pdb ID: 4CHA) as a template to construct models for the 8 isoforms SmCE1a.1,

SmCE1a.2, SmCE1b, SmCe1c, SmCE2a.1, SmCE2a.2, SmCE2a.3 and SmCE2b. Only chains B and C of chymotrypsin were used. The percent identities for each one with the chymotrypsin structure (chains B and C of 4CHA) are 23.68%, 23.25%, 22.64%, 20.61%, 23.68%, 23.68%, 23.68% and 21.05% respectively. The models were built by using the Prime software (version, 2.2, Schrödinger LLC, New York, NY, 2010). Secondary structure prediction was performed. The sequence alignment generated by Prime was manually modified to allow for the regions of insertions or deletions in the target to fall in the loop regions. A qualitative electrostatic representation of the 2 isoforms SmCE1a.2 (representative of the SmCE1 group) and SmCE2b (most different from rest of the isoforms) was generated using Pymol (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC.). The sequence alignment was generated using ClustalW [48].

2.4 Results

2.4.1 All SmCE isoforms are expressed primarily in the intramolluscan and larval stages

Annotation of the *S. mansoni* genome revealed a total of eight full-length SmCE isoforms; based on amino acid sequence homology, these can be grouped, as can previously identified isoforms, into two major classes [34]. To determine if all SmCE isoforms were expressed at the same parasite life cycle stages, we purified total RNA from all major stages. We then performed quantitative RT-PCR with TaqMan primer/probe combinations that were specific to the different isoform

subsets (Figure 1A). SmCE1a and b are highly similar at the nucleotide sequence level, so a primer/probe set common to both was made, in addition to sets specific for SmCE2a and SmCE2b. Transcript levels were normalized to SmGAPDH, which is stably expressed throughout the entire *S. mansoni* life cycle (Figure S3). SmCE1(a and b), SmCE2a and SmCE2b mRNA expression is confined to the daughter sporocyst and cercarial stages, with significantly more transcript present in the thirty day old daughter sporocysts, which reside in the hepato-pancreas of the intermediate snail host. The relative levels of isoform mRNA expression correlate with the number of genes corresponding to each isoform: SmCE1(a and b) are encoded by the most genes, and are the most abundant SmCE transcripts. SmCE2b, for which no protein has been identified, comprises the smallest portion of SmCE transcript.

Next, we sought to determine the protein expression profile of SmCE. Soluble protein was extracted from all parasite life cycle stages, and SmCE protein was detected by immunoblot with polyclonal antiserum raised against recombinant cercarial elastase [34] (Figure 1B). While not isoform specific, this antibody revealed that the majority of SmCE protein is expressed in the late sporocyst (40 d.p.i) and cercarial stages, but is absent in the juvenile, or lung-stage, parasites one week after initial infection. The antibody does not cross-react with uninfected snail hepato-pancreas (Figure S4). This correlates with previous studies that suggest SmCE is produced very early in cercarial development within the

daughter sporocyst and that the contents of the acetabular glands are rapidly exhausted during the course of skin invasion [49].

2.4.2 SmCE is activated prior to exit from the intermediate snail host

Given the frequent presence of endogenous protease inhibitors and activators, protein levels alone are not always indicative of proteolytic activity, and tracking activity is critical to understanding a given protease's function. SmCE is expressed as a zymogen and contains a short prodomain that must be removed in order for the protein to become active; moreover, the presence of SmSerpinQ, a serpin (serine protease inhibitor) with specificity against cercarial elastase that is itself contained within the acetabular glands offers an additional barrier to protease activity [45] (L. Lopez-Quezada, *unpublished*). We therefore sought to specifically identify where active SmCE is present in the *S. mansoni* life cycle. To this end, we made use of a biotinylated phosphonate probe, biotin-nVPL-O$(Ph)_2$, that had previously been shown to bind to active SmCE [39]. The presence of phenyl leaving groups on this molecule strongly promote nucleophilic attack by the active site serine of SmCE, which in turn leads to a covalent linkage. In this way, only active protease (i.e., protease without a prodomain sequence or in complex with inhibitor) is biotinylated. Lysates from all major life cycle stages were incubated with the inhibitor for an hour at room temperature, run on an SDS-PAGE gel and transferred to PVDF membrane for detection with HRP-conjugated avidin. This labeling revealed that active SmCE was present in both six-week daughter sporocyst and shed cercariae (Figure 2A).

While SmCE is known to be active in the invasive cercariae, where it facilitates transit through host skin, the finding that it was activated prior to exit from the intermediate snail host was surprising. Given the possibility that a non-physiological activating factor was present in the lysate, we further explored the activation of the enzyme in histological cross-sections of snail hepato-pancreas tissue. Snails were sectioned approximately six weeks after infection, and sections were preserved in paraformaldehyde to minimize protein cross-linking and thereby preserve enzyme function. Treatment with a general serine esterase stain led to distinct visual staining in both the acetabular cell bodies and the ducts leading from acetabular cells to the anterior end of developing larvae (Figure 2B). These morphological features are prominent and easily identifiable in the developing cercariae, which is itself contained within the daughter sporocyst [31]. Pre-treatment with AAPF-CMK, a known inhibitor of SmCE, abolished stain reactivity (Figure 1C), whereas treatment with FPR-CMK, to which the SmCE is largely insensitive, had no effect on esterase staining (Figure 1D)[19]. This led to the conclusion that visualized staining with specific to SmCE activity within the developing cercaria.

2.4.3 Computational modeling predicts key differences in the SmCE isoforms substrate-binding site

In order to further explore whether SmCE gene expansion has lead to diversified function of protease isoforms, we next looked at the amino acid sequence

variation between isoforms. Computational models of all eight full-length SmCE isoforms were generated by alignment to bovine chymotrypsin (pdb ID: 4CHA) (Figure 3). Models of SmCE1a.2, SmCE2a.2 and SmCE2b, as representative members of each isoform group, are presented here. While SmCE shares the basic α / ß fold with all chymotrypsin-like serine proteases, there are several notable structural features that are largely unique to SmCE, as has been previously noted [32]. These include a missing disulfide bond that is formed between Cys116 and Cys220 in chymotrypsin, and that defines helps the S3 binding pocket. In this way SmCE more closely resembles the structure of rat mast cell protease II [50]. An additional similarity is the high percentage of charged resides distributed throughout the surface of the protein, which may suggest that the protease is packaged within the acetabular glands in an analogous fashion to mast cell protease packaging within granulocytes (Figure 4) [51].

While our models predict that amino acid variations between isoforms are scattered throughout the surface of the protein, there are several notable structural differences within the substrate-binding pocket that exist between SmCE isoform groups (Figure 5). In protease nomenclature, the position where the amino acid N-terminal to the scissile bond of the substrates binds the active site of the protease is referred to as the S1 pocket, and the next site is the S2, etc. The corresponding C-terminal positions are the S1', S2', etc. In chymotrypsin, Ser189 and Ser190 form the base of the S1 pocket [52]; in all SmCE isoforms, a hydrophobic leucine or isoleucine fills the 189 position (Figure

3; Figure 5a). Position 190 is more varied, with a threonine is SmCE2a, an alanine in SmCE2b and a proline in SmCE1a and 1b. Moreover, while Gly216 is conserved between trypsin, chymotrypsin and SmCE isoforms, amino acid 226 is a leucine rather than a glycine, suggesting that that architecture of the S1 pocket is distinctly hydrophobic in most SmCE isoforms.

The S4 pocket varies significantly between SmCE2b and rest of the SmCE isoforms (Figure 5). SmCE2b is unique in having a phenylalanine rather than a threonine at position 218, again increasing hydrophobicity in this pocket and also potentially leading to a steric clash with substrates including AAPF-CMK (Figure 5a). As depicted in Figure 4, the charges of the prime side substrate-binding pocket also vary significantly between SmCE2b and other SmCE isoforms. The consequences of this are of on-going investigation (*O'Donoghue et al, manuscript in preparation*), but suggest that SmCE2b has a diversified substrate binding preference compared to other SmCE isoforms.

2.5 Discussion

Gene duplication is abundant in all sequenced eukaryotic genomes [53]. Parasites frequently regulate gene expression through gene duplication. This is especially true for protozoan parasites, like *Leishmania* and *Trypanosoma spp*, which contain tandem arrays of duplicated genes throughout their respective genomes [54,55]. This presumably compensates for a lack of transcriptional regulatory control (both species transcribe genes as polycistronic mRNAs), with greater gene

number resulting in greater gene expression. In other parasite species, along with most eukaryotes, gene duplication frequently results in the sub-functionalization or neo-functionalization of the gene, and there are several examples of this that specifically involve protease gene families, including the duplication and functional divergence of the cathepsin L family in the liver fluke *Fasciola hepatica*, and that of several aspartic protease in the nematode *Strongyloides ratti* [56,57].

The expansion of the cercarial elastase gene family by gene duplication in *S. mansoni* appears to involve both an increase in gene dosage for a critical enzyme during a crucial life cycle stage transition—host skin invasion--and may also represent the neo-functionalization of a serine protease as an invasive enzyme. The tight temporal regulation of the expression and activity of all SmCE isoforms throughout the parasite life cycle suggests that they are all share a conserved function. Moreover, the residues of the active site and the surrounding residues that determine the substrate specificity are largely conserved among the isoforms, suggesting that they share a common pool of substrates, most likely those present in human skin [43]. Slight variations in the substrate binding pockets of individual isoforms, however, may expand the range of available substrates in skin.

A notable exception to this is the SmCE 2b isoform. A homolog of this isoform is present in *S. japonicum* and is likely to be the ancestral gene of this family [9].

Despite the presence of transcript in the developing cercariae, no protein corresponding to this isoform has been identified in *S. mansoni*, or *S. japonicum*, although the latter case is an point of ongoing debate [29,30,34,39]. However, the fact that SmCE 2b is actively transcribed and that the catalytic triad has been preserved by evolution suggest that the gene is functional. In addition, computational modeling show the greatest level of divergence in the substrate binding pocket of SmCE 2b compared to the other isoforms, notably in the S4 pocket, suggesting that it may have a different substrate preference than its cohorts.

The zoonotic *S. japonicum,* the avian-specific schistosome species and the more distantly related liver flukes are reported to use papain-like cysteine protease, rather than a serine protease, to facilitate initial invasion of their respective hosts [39,40,43,56]. Based on current genome data, the expansion of the cercarial elastase gene family coincides with increase selectivity of humans as the definitive host of *S. mansoni*. A plausible model is that the initial duplication of the ancestral gene allowed for mutations within the substrate-binding site of the duplicated gene that allowed for greater activity against substrates present in human skin. Subsequent gene duplication events in the *S. mansoni* / *S. haematobium* lineage then increased gene expression of the human substrate-specific isoforms that comprise the group 1 / group 2a SmCE gene families.

This model presupposes that SmCE 2b functions differently in the skin invasion

process, or functions elsewhere in the parasite life cycle. It is possible the SmCE2b has a greater role in egress from the intermediate snail host, since data presented here suggests that at least a subset of SmCE isoforms are activated prior to the parasite exiting the intermediate host. While a previous microarray study by Gobert *et al* suggest the unique up-regulation of SmCE2b in 5 day cultured schistosomules, only trace amounts of SmCE2b transcript were found in the *in vivo* lung schistosomules in this study [58]. Further work, including the development of more sensitive molecular probes to track individual isoform activity and recombinant expression of all SmCE protease isoforms, are sure to lead to further elucidation of the functions of this gene family.

CHAPTER 2
FIGURES AND TABLES

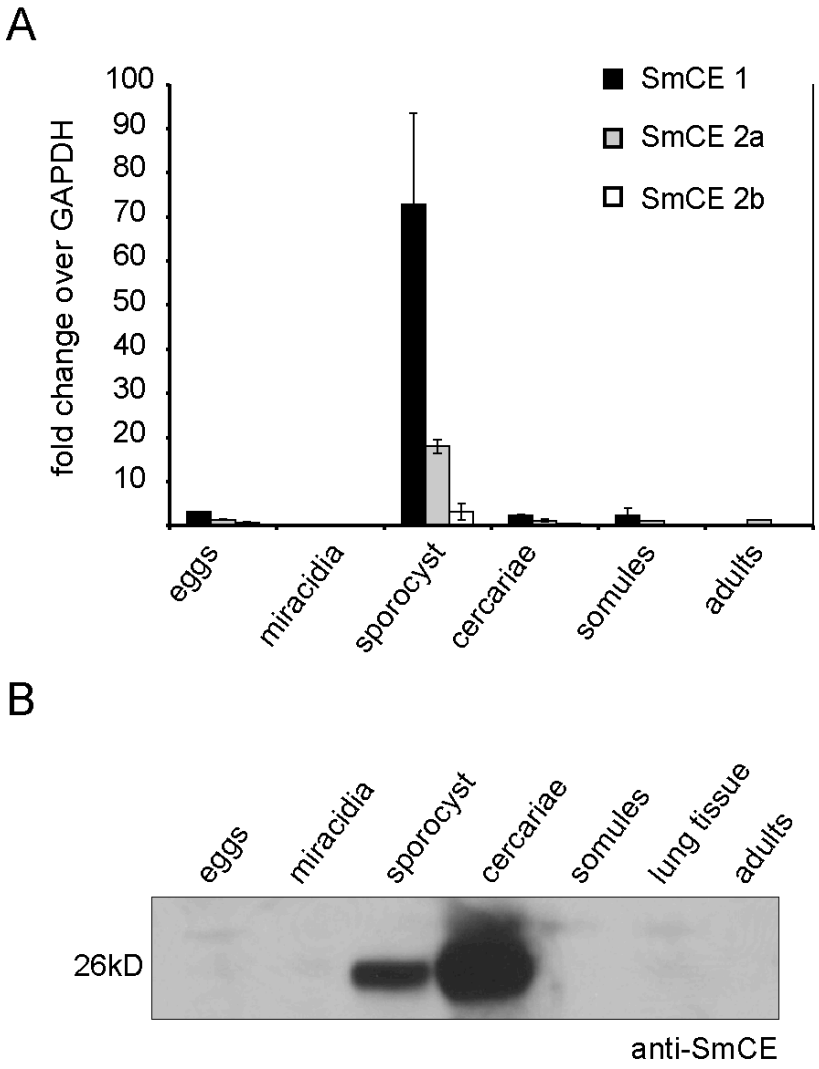


Figure 2.1 All SmCE isoforms show similar mRNA and protein expression patterns throughout the *S. mansoni* life cycle (A) SmCE1 (a and b), SmCE 2a and SmCE2b mRNA levels were monitored by quantitative RT-PCR. (B) Protein levels of SmCE were monitored in lysates from all life cycle stages by immunoblot with a polyclonal antibody raised against recombinant SmCE.

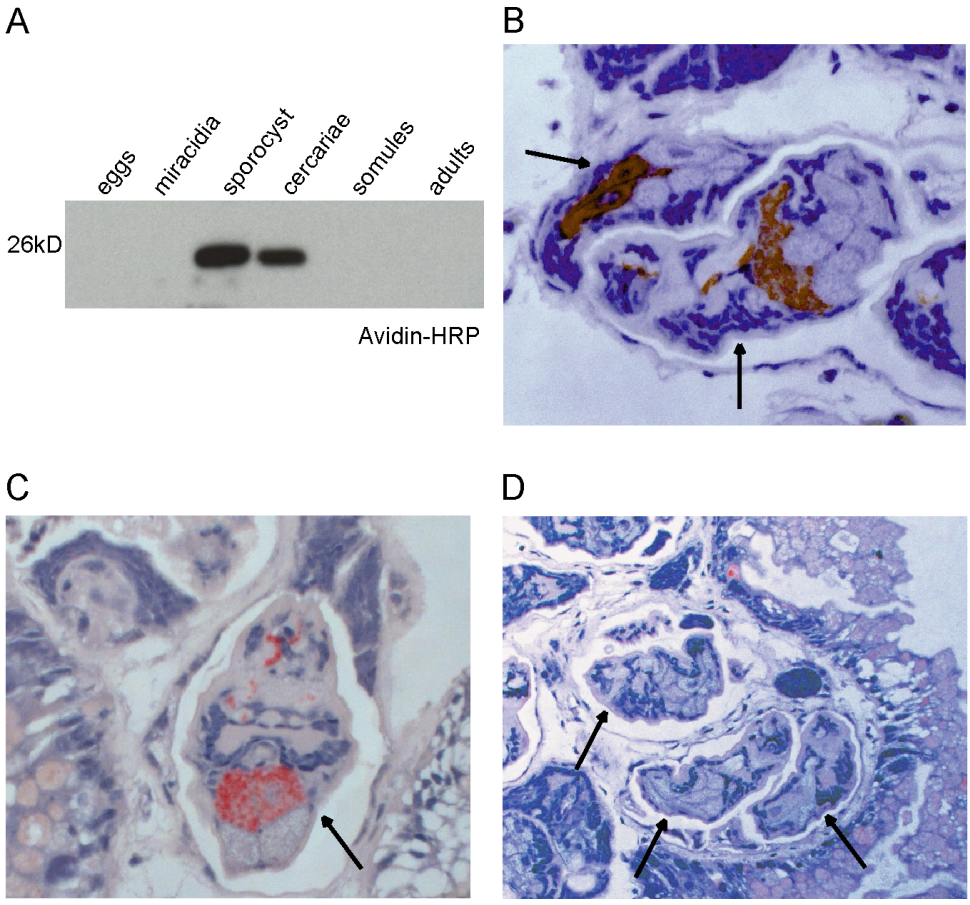


Figure 2.2 SmCE is activated within the intermediate snail host. (A) Avidin-HRP blot of lysates made from various *S. mansoni* life cycle stages incubated with biotin-nVLP-OPh2. (B) Histological section of daughter sporocyst within a patent *B.glabrata* hepatopancreas, treated with general esterase stain. (C) Pre-treatment with FPR-CMK. (D) Pre-treatment with AAPF-CMK. Arrows indicated the developing cercariae within the daughter sporocyst.

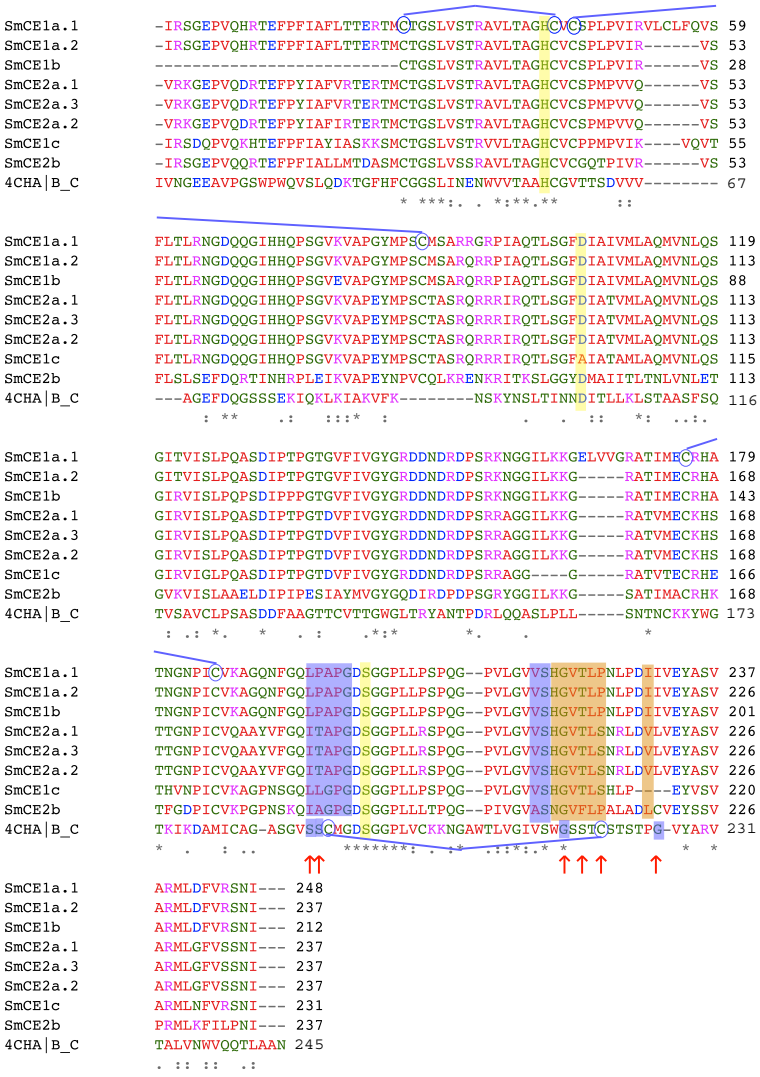


Figure 2.3 Protein sequence alignment reveals that SmCE isoforms can be classified as belonging to one of two main groups. Mature SmCE isoforms amino acid sequences and bovine chymotrypsin (4CHA) B and C chains were aligned using ClustalW. Active site residues are highlighted in yellow. Cysteines forming disulfide bridges are circled. Residues comprising the S1 binding pocket are shaded blue; residues comprising the S4 binding pocket are shaded orange. Residues further described in the binding site models (Figure 2.5) are indicated with red arrows.

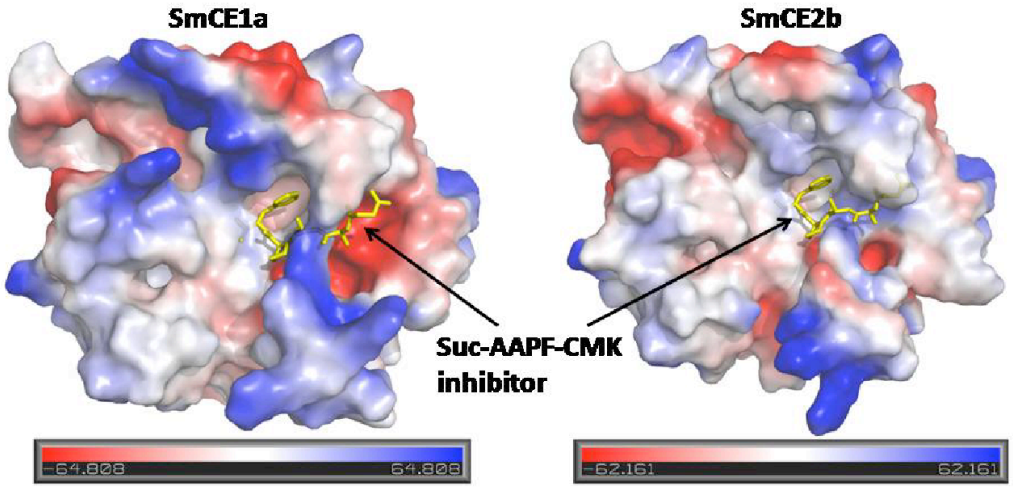


Figure 2.4 Modeling of qualitative electrostatic potential of the surface of SmCE1a.1 and SmCE2b shows distinct differences in areas proximal to the active site. SmCE 1a.1 and SmCE2b were modeled on bovine chymotrypsin (4CHA), and the known SmCE inhibitor, AAPF-CMK (yellow) is docked in the active site of the proteases.

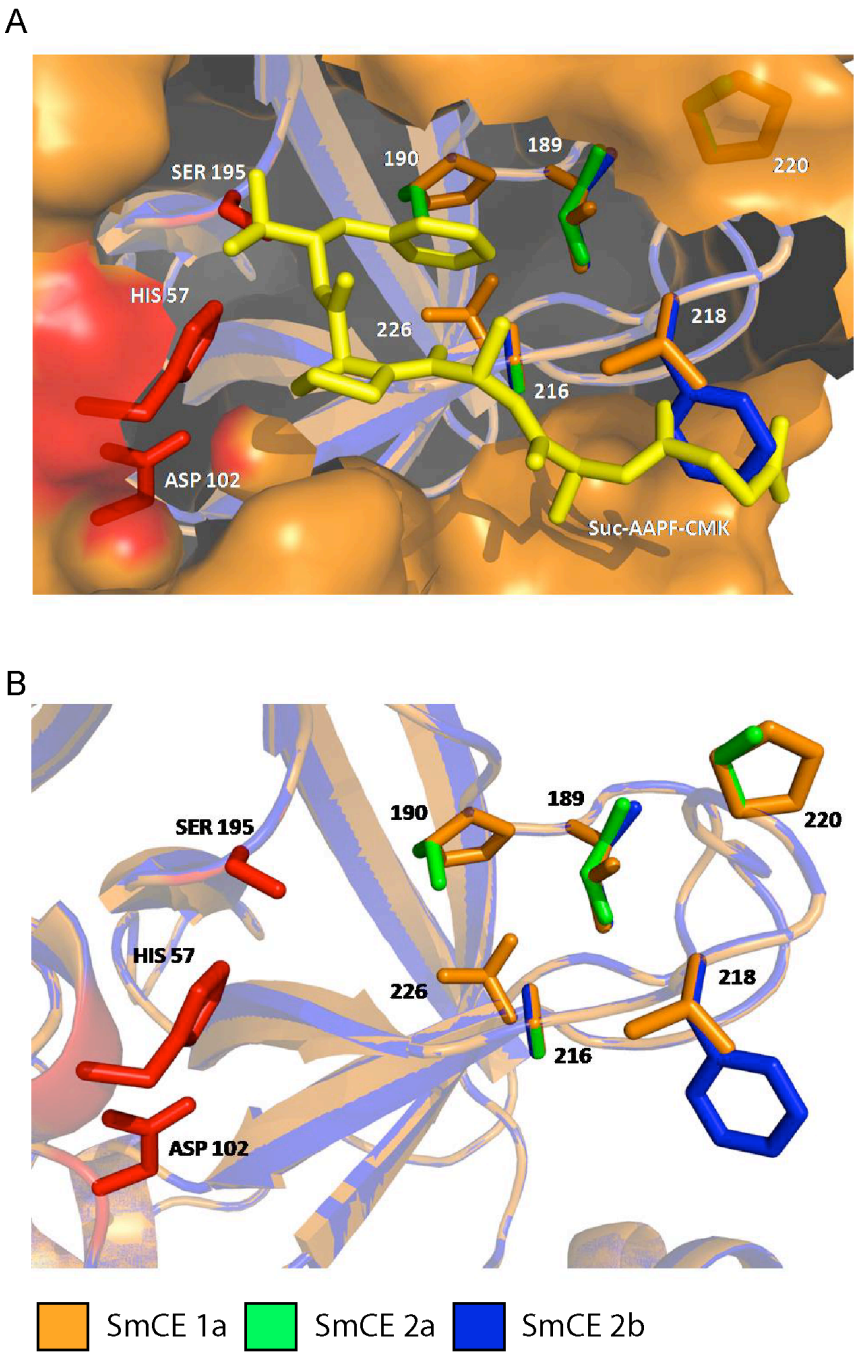


Figure 2.5 Computational modeling predicts key differences in the substrate-binding pocket of among SmCE isoforms. SmCE 1a.1, SmCE 2a.1 and SmCE2b were chosen as representative members of each isoform group and modeled on bovine chymotrypsin (4CHA). Overlaid models of the substrate binding pocket are presented here with both (A) AAPF-CMK (yellow) docked in the active site and (B) the active site without substrate. Residues that are key binding site determinants are presented as ball-and-stick models, and colored according to isoform group. Active site residues are depicted in red. All amino acid positions are presented according to canonical chymotrypsin numbering.

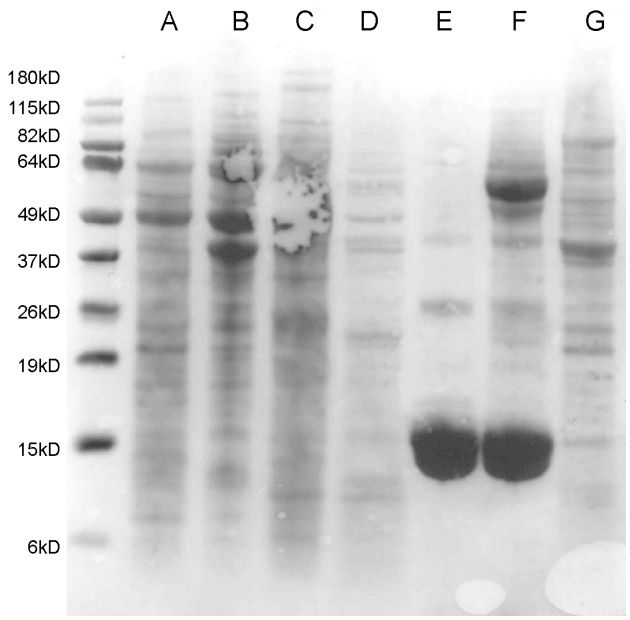


Figure S2.1 Ponceau stain of anti-SmCE immunoblot. A-eggs; B- miracidia; C-daughter sporocyst; D-cercariae; E-lung stage schistosomules; F- lung control; G-adult worms.

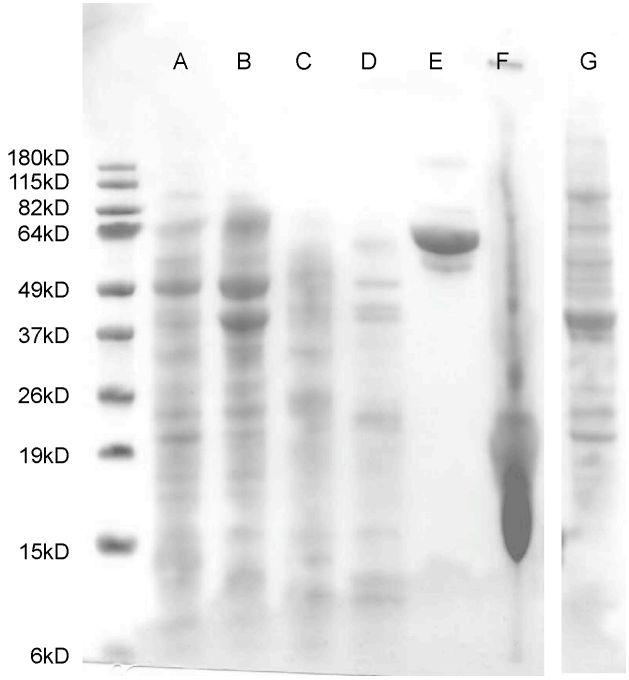


Figure S2.2 Ponceau stain of avidin-HRP immunoblot. A-eggs; B- miracidia; C-daughter sporocyst; D-cercariae; E-lung stage schistosomules; F- lung control;G-adult worms. An additional lung control lane (between lanes F and G) was cropped from the image.

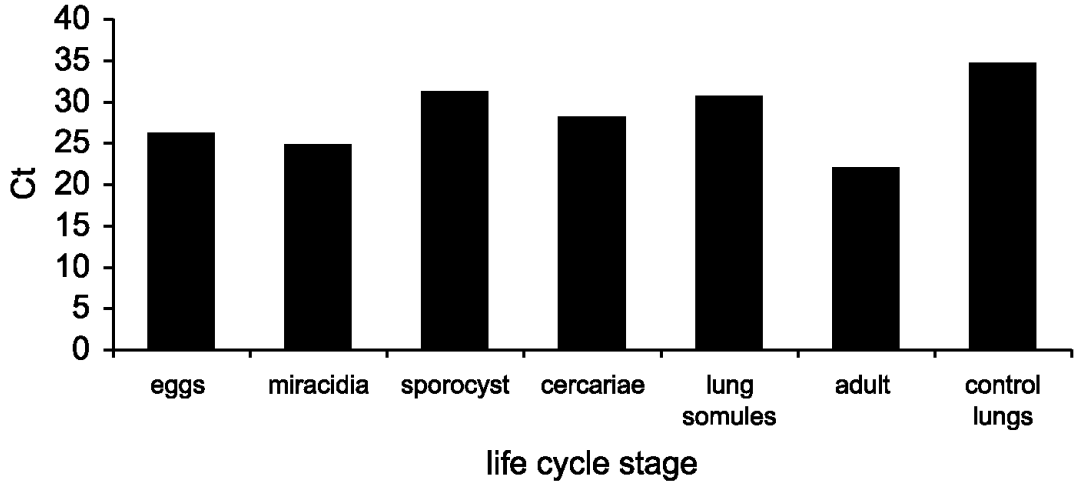


Figure S2.3 SmGAPDH is stably expressed throughout the *S. mansoni* life cycle. Threshold values are plotted for SmGADPH amplification for each life cycle stage. Identical mRNA starting concentrations were used to generate template for each reaction.

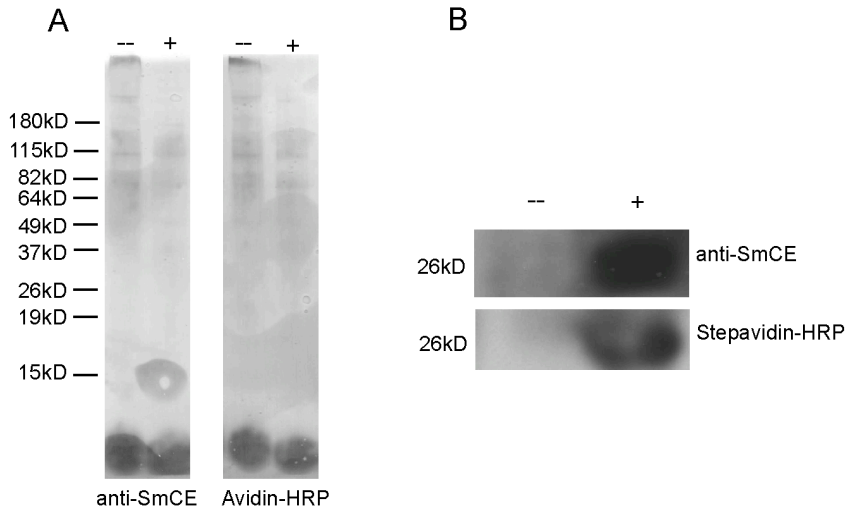


Figure S2.4 Anti-SmCE antibody and biotin-nVPL-$(OPh)_2$ do not cross-react with uninfected snail tissue. A. Ponceau stain of immunoblot and avidin-HRP membrane. B. Immunoblot (top) and avidin-HRP blot (bottom) of uninfected snail hepato-pancreas (-) and 40 d.p.i snail hepato-pancreas (+).

CHAPTER 3
GLOBAL PROTEOLYTIC ANALYSIS OF *SCHISTOSOMA MANSONI* CERCARIAL SECRETIONS

(A portion of this work is from a paper submitted in its entirety to *Nature Chemical Biology*, December 2011)

3.1 Chapter Summary

Multiple proteomic studies of *Schistosoma mansoni* cercarial secretions have identified various classes of proteases as being present. The most of abundant of these is cercarial elastase (SmCE), an S1A serine protease named for its ability to digest insoluble elastin. In addition to SmCE, proteomic studies have identified other proteases present in cercarial secretions. These additional proteases include a leishmanolysin-like metalloprotease and several other, less abundant proteases, including a calpain-like protease and a number of exopeptidases [30]. These proteomic studies disagree on the exact composition of secreted proteins, and the contribution of other proteolytic enzymes in host skin invasion is unknown [29,30]. We therefore sought to define the proteolytic activity of *S. mansoni* larval secreted proteases on a global scale, in order to better understand their respective contributions to host invasion. Using a recently developed peptide library for universal substrate specificity (PLUSS), we were able to determine that SmCE is the dominant proteolytic activity in *S. mansoni* cercarial secretions. This confirms previous work that suggests that SmCE is necessary for successful skin penetration by *S. mansoni* [32].

3.2 Introduction

Presented here is a direct cleavage assay that uses mass spectrometry-based peptide sequencing for detection. This method utilizes a defined library of 124 synthetic 14-mer peptides without additional tags or labels. In this study we wanted to examine whether the depth of information obtained from a small, defined library could be as rich as that derived from exponentially large synthetic or proteome-derived libraries. We demonstrate that this library design has multiple advantages. First, the tag-free design is compatible with profiling of multiple classes of exo- and endo-peptidases and can produce substrate specificity preference data covering both non-prime- and prime-side specificity. The method, because of high information content relative to low background, requires no additional labeling or sample fractionation as are needed for resolving complex proteome-derived libraries of peptides. Peptide degradation kinetics were also monitored by label-free quantitation by integrating parent ion abundance in the MS as a means to rank the best substrates identified within the multiplex assay. Furthermore, use of defined synthetic peptide sequences rather than a proteome-derived sample allows for control of amino acid frequency and optimization of sequence diversity, making it highly reproducible. Finally, through the use of differential inhibition, we also demonstrate an application whereby complex mixtures of multiple proteases can be deconvoluted using this assay.

3.3 Experimental Procedures

3.3.1 Cercarial secretion isolation

Cercariae were shed from several hundred infected *Biomphalaria glabrata* and secretions were collected as previously described [45]. Isolated secretions were lyophilized and re-suspended in 5 mL 50mM Tris, pH 7.5. They were then sonicated for 1 min at 30% output, centrifuged for 5 min at 16,000 rcf at 4°C, and supernatant was saved as the soluble fraction. Reconstituted secretions were split into eight 250 µL samples, to which 225 µL 50 mM Tris, pH 7.5 and 25 µL inhibitor solution was added to a final concentration of 250nM L-3-*trans*-(Propylcarbamoyl)oxirane-2-carbonyl]-L-isoleucyl-L-proline (CAO74) / 250nM N-(trans-Epoxysuccinyl)-L-leucine 4-guanidinobutylamide (E64); 500 nM 1,10-Phenanthroline (1.10P); 25mM Ethylenediaminetetraacetic acid (EDTA); or 500nM succinyl ala-ala-pro-phe-chloromethyl ketone (Suc-AAPF-CMK). All Inhibition reactions were incubated for 30 min at room temperature. Fifty microliters of each inhibited sample was then removed and assayed against 100 µg/ml Enz-chek (Invitrogen) in 50 mM Tris, pH 7.5 (Figure 3.1). This reaction was monitored on a Flex Station fluorimeter (Molecular Devices) with excitation/ emission values of 485/530. The remaining reaction was then increased to a final volume of 470 µL with 50 mM Tris, pH 7.5.

3.3.2 PLUSS library analysis of purified SmCE 1a/b and SmCE2a

SmCE activity was purified from lysate as previously described with the following modifications [34]. Cercariae shed from approximately 50 snails were pelleted by centrifugation at 100 rcf for 1 minute and stored at -20°C. One milliliter of pelleted cercariae was resuspended in 5 ml 300 mM sodium acetate, pH 6.5,

0.1% Triton X-100, 0.1% Tween-20, 0.05% NP40, and sonicated for 1 minute at 40% output. Soluble protein was harvested by centrifugation for 15 minutes at 7, 500 rcf, followed by 0.2 μ filtration. Fractions were again measured for SmCE activity against Suc-AAPF-pNA (succinyl-Ala-Ala-Pro-Phe-p-nitroanilide), and active fractions were run on 10% bis-TRIS polyacrylamide gels (Invitrogen, Carlsbad, CA) according to the manufacturer's specifications, and silver stained [59]. For confirmation of protein identification, bands corresponding to the correct molecular weight of SmCE were excised from the gel, and subjected to in-gel trypsin digestion, followed by LC-MS/MS analysis.

3.3.3 Multiplex Peptide Cleavage Assay

Enzyme assays were performed in 50mM Tris, pH 7.5 using 100 nM of purified enzyme or cercariae shed in a reaction volume of 600 μl. The multiplex assay was prepared using three separate peptide pools, termed A, B, and C containing 52, 52 and 20 peptides respectively. These peptide pools were chosen for good chromatographic resolution by LC-MS/MS, in order to reproducibly detect all 124 intact library peptides under initial conditions and monitor their degradation over the time-course of the experiment. A control sample lacking enzyme was also prepared under identical conditions and quenched at the first and last time points of the assay to account for non-enzymatic degradation of the substrates. Each peptide pool was incubated at room temperature with protease or shed and aliquots were removed and acid quenched to pH 3 or less with formic acid (4% final) after 15, 60, 240 and 1200 minutes.

3.3.4 Peptide Cleavage Site Identification by Mass Spectrometry

Peptide cleavage site identification was performed using peptide sequencing by mass spectrometry. Samples containing approximately 1-3 µg total peptide (calculated as 60 µl of enzyme reaction containing peptide pools at 500 nM concentration) were desalted using C18 zip tips (Millipore), then rehydrated in 0.1% formic acid; 0.1 µg total peptide was injected per LC-MS/MS run. For LC-MS/MS, a linear ion trap LTQ mass spectrometer (Thermo) equipped with an Ultimate HPLC and Famos autoinjector (LC Packings) was used, with a C18 "Magic" column (Michrom Bioresources, Inc., 5 µm bead size, 0.3 x 150 mm Magic, 200 Å). The LC system was operated at 5 µL/min flow rate, and peptides were separated using a linear gradient over 42 min from 2% B to 40% B, with solvent A: 0.1% formic acid in water and solvent B: 0.1% formic acid in 70% acetonitrile, 30% water. On the LTQ instrument, survey scans were taken over 300-1500 m/z, and the top three ions in the survey scan were subjected to a high resolution MS "zoom" scan of the precursor and then a CID fragmentation MS/MS scan.

Mass spectrometry peak lists were generated using in-house software called PAVA, and data were searched using Protein Prospector software v. 5.8.0 (Chalkley, Lynn). Database searches were performed against a defined library of 124 peptides. For estimation of false discovery rate, a decoy database containing randomized sequences of the same 124 entries was concatenated to

the original 124 entries. Data were searched with parent mass and fragment mass tolerances of 0.8 Da. For database searching, peptide sequences were matched with no enzyme specificity requirement, and variable modifications including oxidation of Trp, Pro and Phe, and N-terminal pyroglutamate from glutamine. The Protein Prospector parameters were: minimum protein score of 15, minimum peptide score of 10, and maximum expectation values of 0.1 for protein and 0.1 for peptide matches.

3.3.5 Heat Map Generation and Motif Analysis

A list of all cleavage sites was generated that included the P4 to P4′ residues if present. The frequency of each amino acid at all eight positions was calculated using software written by David T. Barkan (Barkan, 2010) and a heat map was generated using R software.

3.4 Results

In order to look globally at the proteolytic secretions, a laboratory strain of *S. mansoni* cercariae were shed from infected *Biomphalaria glabrata* snails and induced to secrete the contents of their acetabular glands by exposure to human sebum. Glandular secretions were then collected, and soluble protein was isolated by mechanical disruption and centrifugation. Previous proteomic analyses of *S. mansoni* secretions report the presence of several isoforms of a serine protease (SmCE), an aminopeptidease, calpain, several isoforms of a leishmanolysin-like metalloprotease (SmPepM8), and a dipeptidyl peptidase IV

(DPPIV) homolog (5, 6). We therefore chose to incubated the lysate with several broad-class protease inhibitors: E64, CAO74, EDTA, and 1.10P, as well as AAPF-CMK, a known inhibitor of *S. mansoni* cercarial elastase(12). Inhibited lysate was then assayed against FITC-labeled casein to determine the extent of inhibition (Figure 1A). To control for background proteolytic activity from the snail host or other aquatic organisms present in the environment, a mock sample lysate was obtained by “shedding” uninfected *Biomphalaria glabrata*, and was also treated with protease inhibitors. Activity against the casein substrate suggested that the majority of proteolytic activity was specific to cercarial secretions and not background host proteases. This activity was largely inhibited by incubation with Suc-AAPF-CMK, suggesting that a serine protease, most likely cercarial elastase, comprised the majority of proteolysis. Both 1.10P and E64 incubation had minimal effects on casein digestion, whereas incubation with EDTA led to a 10% decrease in casein digestion. It is likely that EDTA has a more pronounced effect because it inhibits multiple proteases in the lysate. as calpain and metalloprotease activity both require divalent cations.

As a macromolecular substrate, casein cleavage was unlikely to reveal the presence of exopeptidases in *S. mansoni* cercarial secretions. For this reason, and to better identify the “proteolytic signatures” or substrate specificities of all secreted proteases, additional inhibited material was incubated with the Peptide Library for Universal Substrate Specificity, or PLUSS, library for various time points, and the resulting cleavage patterns were analyzed by LC/MS/MS (Figure

1B). This analysis revealed significant background proteolysis, with a total of 85 unique cleavages that were not specific to cercarial secretions. A proteolytic signature of both infected and non-infected snail shed was generated using the peptide library (Figure 3.2). Proteases in the non-infected snail “shed“ cleaved 85 unique sites in the peptide library indicating a high background of proteases. In the infected snail shed cleavage was observed at 152 sites. Of these, 47 sites were observed in non-infected snails while 105 were unique to *S. mansoni* infection. When samples were pre-treated with an known cercarial elastase inhibitor, Suc-AAPF-CMK, the number of cleavage sites unique to infected snails was reduced by 70% indicating that the majority of proteolytic activity was derived from cercarial elastase. Native isoforms of cercarial elastase were isolated and profiled using the peptide library under identical buffer conditions (Figures 3.3 and Figure 3.4). A total of 114 unique cleavage sites were observed after a four hour assay. All isoforms have a strict preference for bulky hydrophobic residues in the P1 position, particularly Phe>Tyr>Leu/norLeu. Furthermore, Pro is only found in the P2 position confirming that Suc-AAPF-CMK is a suitable inhibitor of the cercarial elastase isoform group.

3.5 Discussion

Proteases are important virulence factors for many human parasites, and are often pursued as therapeutic targets. A caveat to the current wealth of genomic and proteomic data for many parasitic species is the knowledge that most proteases require post-translational modification to become active, and therefore

the presence and localization of proteolytic activity, rather than that of protein alone, is more indicative of a given protease's function. To address this discrepancy, several global protease or "degradomic" profiling techniques have recently been developed. Here, we used the recently developed PLUSS library to determine the complete set of proteolytic activities in *S. mansoni* cercarial secretions. These data suggest that a cercarial elastase-like activity is the dominant proteolytic activity in *S. mansoni* cercarial secretions, which is commiserate with proteomic analyses that identify SmCE as being a predominant protein. Moreover, a closer analysis of the substrate specificity of two SmCE protease isoforms verifies that Suc-AAPF-CMK is a suitable inhibitor of SmCE activity.

CHAPTER 3
FIGURES AND TABLES

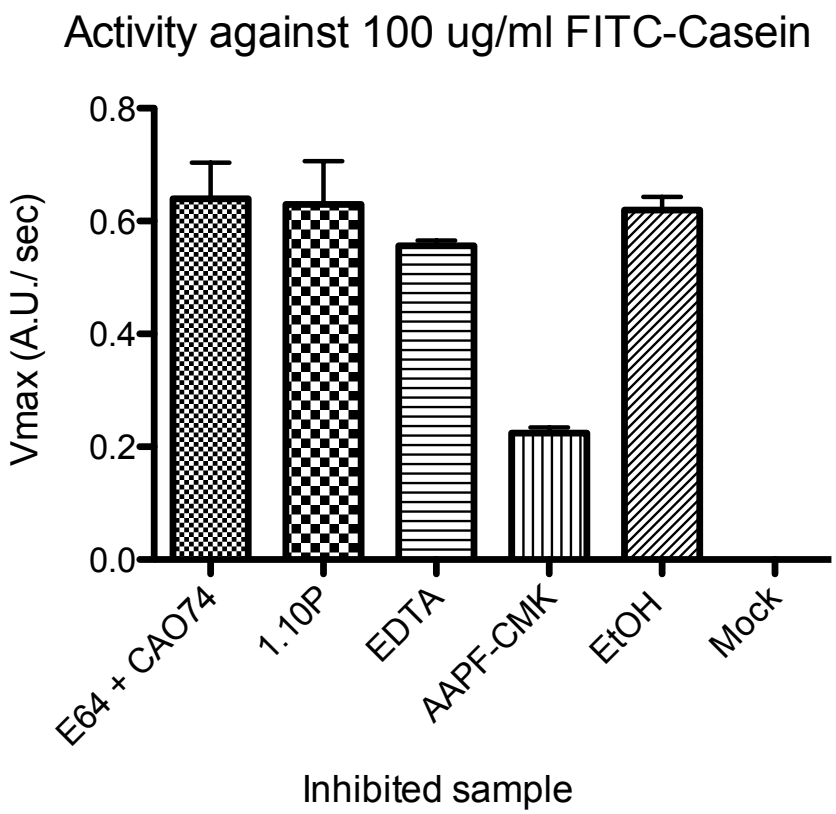


Figure 3.1 Inhibition profiles of *S. mansoni* cercarial secretions treated with various protease inhibitors, as tested against the generic substrate FITC-Casein. Soluble cercarial secretions were incubated with 250nM CAO74 / 250nM E64; 500 nM 1.10P; 25mM EDTA; or 500nM AAPF-CMK for 30 min at room temperature. Fifty microliters of each inhibited sample was then removed and assayed against 100 µg/ml FITC-labeled casein in 50 mM Tris, pH 7.5.

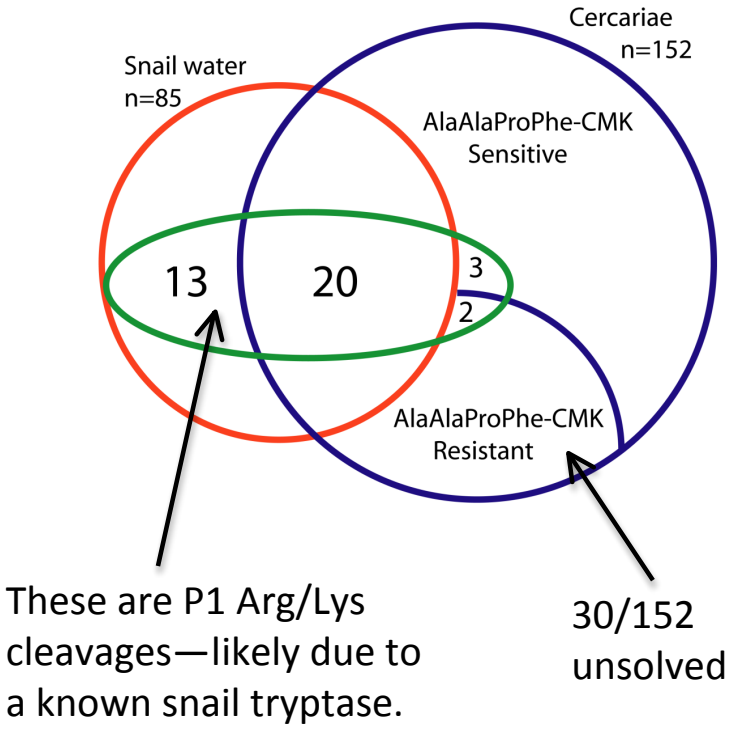


Figure 3.2 Proteolytic activities of cercarial holo-secretions as compared to background activity from the snail environment. Cleavage events were determined by running both cercarial and background samples against the PLUSS library.

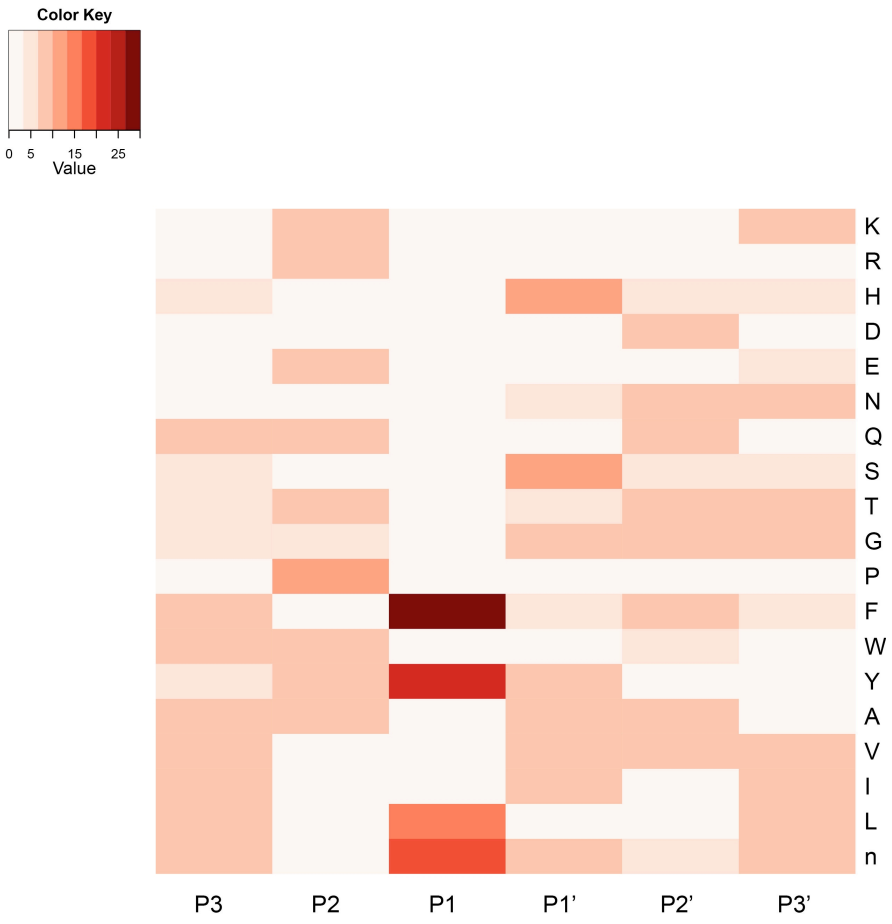


Figure 3.3 P3 - P3' substrate specificity of SmCE type I isoforms as determined by PLUSS library. Color intensity indicates a greater number of peptides cleaved at the indicated amino acid position. The amino acid position is indicated on the horizontal axis, and the amino acid filling that position is indicated on the vertical axis.

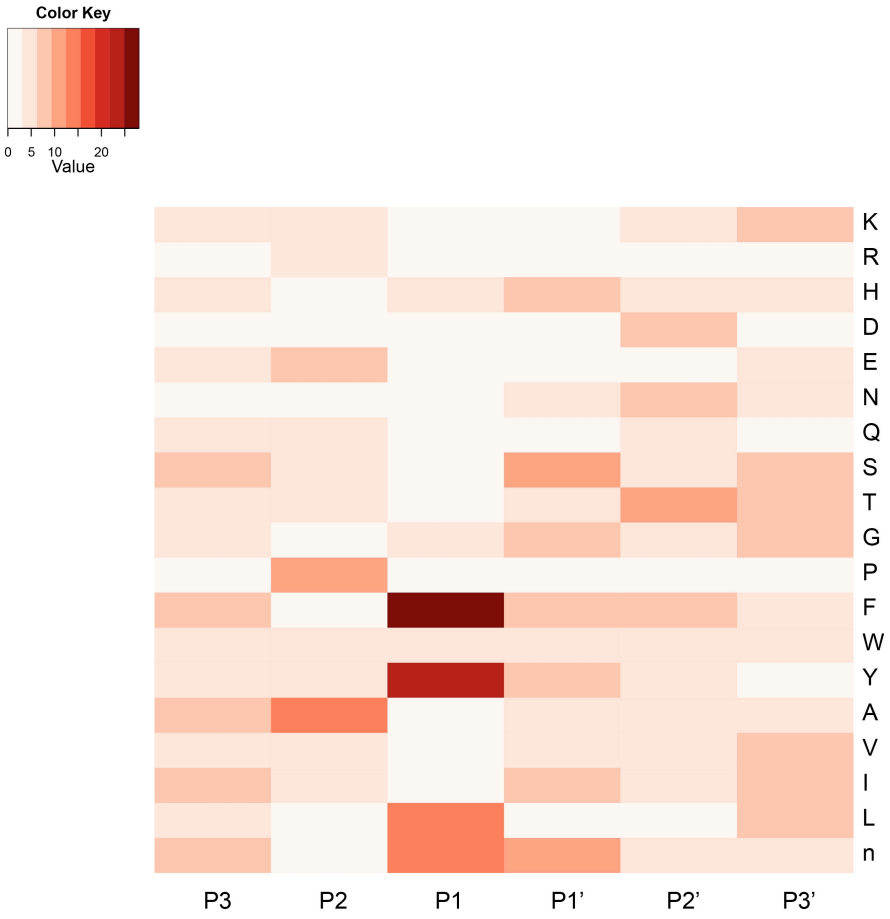


Figure 3.4 P3 - P3' substrate specificity of SmCE type 2a isoform as determined by PLUSS library. Color intensity indicates a greater number of peptides cleaved at the indicated amino acid position. The amino acid position is indicated on the horizontal axis, and the amino acid filling that position is indicated on the vertical axis.

CHAPTER 4

PROTEOMIC ANALYSIS OF HUMAN SKIN TREATED WITH LARVAL SCHISTOSOME PEPTIDASES REVEALS DISTINCT INVASION STRATEGIES AMONG SPECIES OF BLOOD FLUKES

(This work was published in PLoS Neglected Tropical Diseases, *Ingram et al*., and reprinted with permission)

4.1 Author Summary

Schistosome parasites are a major cause of disease in the developing world, but the mechanism by which these parasites first infect their host has been studied at the molecular level only for *S. mansoni*. In this paper, we have mined recent genome annotations of *S. mansoni* and *S. japonicum*, a zoonotic schistosome species, to identify differential expansion of peptidase gene families that may be involved in parasite invasion and subsequent migration through skin. Having identified a serine peptidase gene family in *S. mansoni* and a cysteine peptidase gene family in *S. japonicum*, we then used a comparative proteomic approach to identify potential substrates of representative members of both classes of enzymes from *S. mansoni* in human skin. The results of this study suggest that while these species evolved to use different classes of peptidases in host invasion, both are capable of cleaving components of the epidermis and dermal extracellular matrix, as well as proteins involved in the host immune response against the migrating parasite.

4.2 Introduction

Human skin is a formidable barrier for much of the microbial world. In addition to the mechanical barrier of structural proteins in the epidermis, basement membrane and dermal extracellular matrix, both the epidermis and dermis are bathed in plasma proteins, including early sentinels of the immune system[13]. In order to successfully breach this barrier, an invading pathogen must degrade protein matrices while minimizing the immune response that it elicits. To this end, many invading organisms utilize insect bites or other mechanical trauma to facilitate their entry into skin, but the multi-cellular larvae of the schistosome blood fluke—the causative agent of the disease schistosomiasis—have the remarkable ability to directly penetrate host skin and gain access to dermal blood vessels [12,14].

The invasive larva(e)—termed cercaria(e)—is 300 μm long, 70 μm wide and comprised of roughly 1000 cells [60]. Upon direct contact with the surface of human skin, cercariae begin to secrete vesicles containing a variety of proteins and an adhesive, mucin-like substance [10]. Proteomic studies identified the majority of proteins secreted by *S. mansoni* cercariae. These include histolytic peptidases [29,30]. The most abundant peptidase in *S. mansoni* secretions is an S1A serine peptidase, termed cercarial elastase (SmCE) (GenBank: AAC46967.1) that has activity against insoluble elastin and other fibrillar macromolecules of skin [33]. Biochemical and immunolocalization studies have confirmed SmCE activity in cercarial secretions [27,45]. Moreover, applying an irreversible serine peptidase

inhibitor to *ex vivo* skin before exposure to cercariae blocks the majority of larvae from invading, suggesting that this serine peptidase has an essential role in skin penetration [28].

While the serine peptidase, cercarial elastase, plays a key role in *S. mansoni* skin invasion, the zoonotic species *S. japonicum* has no serine peptidases in its larval secretions [39]. *S. japonicum*, however, encodes a number of isoforms of cathepsin B2 (SjCB2) (GenBank: CAA50305.1), a cysteine peptidase, which are secreted by the invading parasite [39]. Moreover, orthologs of SjCB2 have been identified in the cercarial secretions of other, non-human schistosome species, including members of the genus *Trichobilharzia* [40]. This led us to the hypothesis that the primary invasive peptidase differs between schistosome species, with *S. mansoni,* a human-specific schistosome species, utilizing cercarial elastase, and *S. japonicum*, a zoonotic schistosome species, utilizing cathepsin B2. Given that *T. regenti* also appears to utilize cathepsin B2 for skin invasion, these observations suggest that the use of a serine peptidase in invasion is the exception, not the rule, among parasitic schistosomes. The use of cercarial elastase may reflect unique properties required by *S. mansoni* to preferentially infect human hosts.

To confirm that cathepsin B2 is also capable of facilitating skin invasion, we used a proteomic approach to identify potential substrates in host skin, for both *S. mansoni* cercarial elastase and *S. mansoni* cathepsin B2 (GenBank:

CAC85211.2), a close homolog of *S. japonicum* CB2. Although RNAi has been developed as a tool in juvenile and adult schistosome worms, it is currently unavailable for the intramolluscan and cercarial stages of development [61]. We therefore chose to use a proteomic approach to validate the roles of these peptidases in skin invasion.

We found that the vast majority of cleaved proteins resulting from human skin exposure to either purified SmCE or SmCB2 overlap, suggesting that both enzymes are capable of facilitating parasite migration through skin. However, we also identified several potential substrates in skin that appear to be cleaved by only one of the two enzymes. Candidate substrates were further validated by *in vitro* cleavage of purified human skin proteins with either peptidase. Together, these observations suggest that more than one mechanism of skin penetration may have evolved as an adaptation specific to the schistosome-host relationship.

4.3 Methods

4.3.1 Phylogenetic analysis of schistosome cercarial elastase and cathepsin B2 proteins

To determine the number of cercarial elastase and cathepsin B2 protein isoforms in schistosome species, all full-length protein sequences (*i.e.,* those possessing the full catalytic core of the peptidase) were collected from both GenBank (NCBI) and *S. japonicum* and *S. mansoni* genome annotation websites (Sanger Institute GeneDB). ClustalW (DNA Databank of Japan), was then used to perform

multiple sequence alignments and to construct phylogenetic trees. A Blosum protein weight matrix was used to score the alignment, with a gap open penalty of 10, a gap extension penalty of 0.20, and gap distance penalty of 5. Bootstrapping values were calculated using the p-distance method, with a count of 100. The resulting phylogenetic tree was visualized with the program Dendroscope.

4.3.2 Purification of S. mansoni CB2 and CE

S. mansoni cercariae were shed from *Biomphalaria glabrata* using a light induction method as previously described [28]. SmCE activity was purified from lysate as previously described with the following modifications [34]. Cercariae shed from approximately 50 snails were pelleted by centrifugation at 100 rcf for 1 minute and stored at -20°C. One milliliter of pelleted cercariae was resuspended in 5 ml 300 mM sodium acetate, pH 6.5, 0.1% Triton X-100, 0.1% Tween-20, 0.05% NP40, and sonicated for 1 minute at 40% output. Soluble protein was harvested by centrifugation for 15 minutes at 7, 500 rcf, followed by 0.2 μ filtration. Fractions were again measured for SmCE activity against AAPF-pNA (Ala-Ala-Pro-Phe-p-nitroanilide), and active fractions were run on 10% bis-TRIS polyacrylamide gels (Invitrogen, Carlsbad, CA) according to the manufacturer's specifications, and silver stained [59]. For confirmation of protein identification, bands corresponding to the correct molecular weight of SmCE were excised from the gel, and subjected to in-gel trypsin digestion, followed by LC-MS/MS peptide

sequencing, described below. Active site titration was performed using the synthetic peptide inhibitor AAPF-CMK (Ala-Ala-Pro-Phe-chloromethylketone).

Recombinant SmCB2 was expressed in *Pichia pastoris* as previously described [62]. Media containing secreted protein underwent 0.2 μ filtration and lyophilization. SmCB2 activity was purified as previously described [62]. Fractions were monitored for SmCB2 activity against 5 μM ZFR-AMC (Z-Phe-Arg-7-amino-4-carbamoylmethylcoumarin) in citrate-phosphate buffer, pH 5.3 supplemented with 4 mM DTT. Enzyme concentration was measured by active site titration using the cysteine peptidase inhibitor CAO74 (N-(L-3-trans-propylcarbamoyloxirane-2- carbonyl)-L-isoleucyl-L-proline).

4.3.3 Ethics Statement

The human skin sample was taken in compliance with protocols approved by the Committee on Human Research at the University of California, San Francisco. Written informed consent was obtained for the operation and use of tissues removed.

4.3.4 Skin digestion

Excised human skin was stored at –80°C. For digestion experiments, skin was thawed, dissected into eight 150-170 mg sections, and placed in 1.5 ml microfuge tubes. To each of these skin sections 100 μl of digestion solution containing either peptidase or inhibited peptidase at 1.8 μM was added, along

with corresponding controls. SmCE reaction buffer consisted of 100 mM TRIS-HCl, pH 8; SmCB2 reaction buffer consisted of 100 mM sodium acetate, pH 5.5, 4 mM DTT. Inhibited SmCE was prepared by incubating 1.8 μM SmCE with 2 μM AAPF-CMK for one hour at room temperature; inhibition was monitored against AAPF-pNA, prior to its addition to skin. Similarly, inhibited SmCB2 digestion solution was prepared by incubating 1.8 μM SmCB2 with 2 μM CAO74 for one hour at room temperature, with full inhibition monitored by activity against ZFR-AMC, prior to its addition to skin. Inhibitor alone digestion solutions were prepared to control for human skin peptidase activity using either 2 μM AAPF-CMK in 100 mM Tris, pH 8.0, or 2 μM CAO74 in 100 mM sodium acetate, pH 5.5, 4 mM DTT. After addition of digestion solution to skin samples, the reaction mix was vortexed briefly, and then incubated for five hours at 37ºC. Following incubation, reactions were centrifuged for 20 minutes at 16,000 rcf at 4°C, and the resulting supernatant was saved as the soluble fraction. Fifteen microliters were removed for analysis on a bis-TRIS 4-20% acrylamide gel. Gels were silver-stained and stored at 4°C.

4.3.5 Proteomic/Mass spectrometry analysis

Proteomic analysis of skin digestion samples was performed by LC-MS/MS on two independent preparations as follows. Representative preparative gels are shown in Supplementary Figures 1 and 2, and contain replicate lanes of approximately 20 μg total protein for each of the skin digestion solutions. Each pair of sample lanes was cut into ten protein bands, and diced into 1-2 mm

cubes, then subjected to in-gel trypsin digestion, following a previously published protocol [29]. The resulting peptides were extracted and analyzed by on-line liquid chromatography/mass spectrometry using an Eksigent nanoflow pump and a Famos autosampler that were coupled to a quadrupole-orthogonal-acceleration-time-of-flight hybrid mass spectrometer (QStar Pulsar or QStar Elite, Applied Biosystems, Foster City, CA). Peptides were fractionated on a reversed-phase column (C18, 0.75x150 mm) and a 5-50% B gradient was developed in 35 min at a 350 nl/min flow rate. Solvent A was 0.1% formic acid in water, solvent B was 0.1% formic acid in acetonitrile. Data were acquired in information-dependent acquisition mode: 1 sec MS surveys were followed by 3 sec CID experiments on computer-selected multiply charged precursor ions. Peak lists were generated using Analyst 2.0 software (Applied Biosystems) with the Mascot script 1.6b20 (Matrix Science, London, UK).

Database searches were performed using ProteinProspector v. 5.7.1 (http://prospector2.ucsf.edu) [63]. Searches were performed using the SwissProt databank (August 10, 2010, 519,348 entries). For false discovery rate estimation, this database was concatenated with randomized sequences generated from the original database [64]. Search parameters included selecting trypsin as the digestion enzyme, allowing one missed cleavage but no non-specific cleavages. Peptide modifications that were searched included carbamidomethyl (Cys) as the only fixed modification, and up to two variable modifications from among the following: oxidation (Met), acetyl (N-term), oxidized

acetyl (N-term), pyroglutamate (Gln), Met-loss (N-term), and Met-loss+acetyl (N-term). Mass accuracy settings were 200 ppm for precursor and 300 ppm for fragment masses. Data reported in Supplementary Table 3 has a Protein Prospector minimum score cutoff of 22 (protein), 15 (peptide) and maximum expectation values of 0.01 (protein) and 0.05 (peptide), resulting in a 2% false discovery rate.

4.3.6 Human Collagen I and Complement C3 cleavage and N-terminal Sequencing

Lyophilized type I human skin collagen (Calbiochem) was resuspended in 17.5 mM acetic acid for a final concentration of 1mg/ml. Human complement C3 (Calbiochem) was purchased as a 1.2 mg/ml stock. For SmCB2 digestion, 180 nM enzyme was added to 50 µl collagen I or 25 µl complement C3 in 50 mM sodium acetate, pH 5.5, 4 mM DTT and incubated at 37°C for 1-22 hours. For SmCE digestion, 180 nM enzyme was added to 50 µl collagen I or 25 µl Complement C3 in 50 mM Tris, pH 8.0 and incubated at 37°C for 1-22 hours. Both enzymes were also pre-incubated with 1 mM CAO74 (SmCB2) or 1 mM AAPF-CMK (SmCE) for one hour at room temperature prior to their addition to collagen. As a control, collagen was incubated in 50 mM sodium acetate, pH 5.5, 4 mM DTT or 50 mM Tris, pH 8.0 for 22 hours at 37°C. To stop the reaction, 15 µl reduced SDS-PAGE loading dye (Invitrogen) was added, and a sample of each reaction was run on a 4-20% Tris-Glycine SDS PAGE gel (Invitrogen). Bands were then electroblotted onto PVDF membrane (Biorad, Foster City, CA) and

visualized by Coommassie Blue staining. N-terminal sequence of selected bands was determined using Edman chemistry on an Applied Biosystems Procise liquid-pulse protein sequenator at the Protein and Nucleotide Facility, Stanford University.

4.4 Results

4.4.1 Identification and phylogeny of cercarial elastase and cathepsin B2 isoforms

To outline the molecular evolution of larval peptidases in schistosomes, all previously reported orthologs were re-examined (Figure 1). In addition to the previously identified full-length cercarial elastase isoforms in *S. mansoni*--SmCE1a (GenBank: AAM43939.1), SmCE1b (GenBank: CAA94312.1), SmCE1c (GenBank: AAC46968.1), SmCE2a (AAM43941.1) and SmCE2b (GenBank: AAM43942.1) and *Schistosoma haematobium* cercarial elastase (GenBank: AAM4394)--sequencing and annotation of the full *S. mansoni* genome revealed three additional full-length genes [34,42] (Figure 1A). In marked contrast, the *S. japonicum* genome contains only a single cercarial elastase isoform (Sjp_0028090). No cercarial elastase genes have been detected in any *Trichobilharzia* species.

Both *S. mansoni* and *S. japonicum* encode a number of cathepsin B genes (Supplementary Figure 3). We chose to focus on the cathepsin B2 isotype, since a proteomic analysis of *S. japonicum* cercarial secretions identified a peptide

sequence common to this subset [39](Figure 1B). Notably, while the *S. mansoni* genome encodes only a single cathepsin B2 isoform, *S. japonicum* encodes four CB2 isoforms. In one of these isoforms, SjCB(Y)2d (GenBank: CAX71091.1), the nucleophilic cysteine of the active site is mutated to tyrosine, which may diminish, if not eliminate, its catalytic activity. Three of the four SjCB2 isoforms (SjCB2b (GenBank: CAX71088.1), SjCB2c (GenBank: CAX71090.1) and SjCB (Y)2d correspond to the peptide sequence identified in proteomic analysis of *S. japonicum* cercarial secretions [39]. A full list of schistosome cercarial elastase and cathepsin B isoforms is provided as supplementary material (Supplementary Tables 1 and 2).

4.4.2 Comparative proteomic analysis of human skin treated with SmCE and SmCB2

A previous proteomic study generated a list of proteins that were released as soluble peptides from *ex vivo* human skin upon treatment with live *S. mansoni* cercariae, indicating that they are actively degraded during cercarial migration through skin [37]. These included many of the structural components of skin, including extracellular matrix proteins, proteins involved in cell-cell adhesion and multiple serum proteins. To identify specific substrates of CE in skin, and to compare these to potential substrates of cathepsin B2, we treated *ex vivo* skin with either peptidase. Since active, recombinant *S. japonicum* cathepsin B2 is not currently available, and purifying sufficient amounts of native peptidase from *S. japonicum* was not feasible, we used *S. mansoni* cathepsin B2 as model

peptidase in our analysis. *S. mansoni* cathepsin B2 has high homology to the *S. japonicum* cathepsin B2 (90% sequence identity and 94% sequence similarity for the mature peptidase, see Supplementary Figure 4), including the active site and substrate binding pocket, and therefore is likely to display highly similar biochemical properties and substrate specificity [62]. SmCE was purified directly from *S. mansoni* cercariae, and the protein composition of proteolytically active fractions was determined by mass spectrometric analysis as a mixture of SmCE1a, 1b and 2a isoforms, but not SmCE2b. This is consistent with the isoform composition of previous proteomic analysis of *S. mansoni* cercarial secretions [29,42]. Active SmCB2 was expressed in recombinant form in *P. pastoris* and purified as previously described [62]. To ensure that equimolar amounts of active enzyme were added to skin samples, an active site titration was first performed for both SmCE and SmCB2 with respective covalent inhibitors.

In comparison to control samples treated with inhibited peptidase, multiple skin proteins migrated through an SDS-PAGE gel as smaller fragments, *i.e.* fragments less than the predicted molecular weight of the full-length protein, upon addition of active SmCE or SmCB2. These were thus identified as substrates of the specific enzyme and included multiple extracellular matrix proteins (Table 1). Addition of both SmCE and SmCB2 to skin led to the cleavage of collagen VI, which is found in interstitial tissue, and collagen XII, a collagen located in the basement membrane of the epidermis [65]. Only samples incubated with active SmCB2 showed cleavage of collagens I, III and XVIII. In addition to collagen,

several other components of the extracellular matrix were degraded upon treatment with either peptidase, including vitronectin, fibronectin, and galectin. Both vimentin and talin-1, cytoskeletal proteins that are associated with desmosomes, were cleaved upon addition of either peptidase. Two additional extracellular matrix components, tenascin-X and thrombospondin-1, were uniquely cleaved upon addition of SmCB2.

Another subset of extracellular proteins identified as substrates of SmCE and SmCB2 were derived from blood plasma that bathes the dermis. These included components of the coagulation cascade, e. g. fibrinogen, antithrombin-III, as well as proteins involved in the host immune response, e. g. complement C3, complement factor D. Addition of either active SmCE or SmCB2 led to the digestion of gelsolin, an actin assembly protein that exists intracellularly and in plasma. Addition of active SmCB2 also led to the digestion of both kininogen-1 and fibrinogen, both of which are members of the coagulation cascade. Complement C3, an integral component of both the classical and alternative complement activation pathways was cleaved upon addition of either SmCE and SmCB2; complement C4A and complement D proteins, respective members of the classical and alternative complement activation pathways, were cleaved by SmCB2 alone.

In addition to the extracellular proteins identified, many cytosolic proteins were also cleaved by either SmCE or SmCB2. A complete list of peptides identified is provided as a supplementary table (Supplementary Table 3).

4.4.3 In vitro digestion of human collagen I and complement C3

To corroborate proteomic identification of substrates in skin, candidate substrates were selected for *in vitro* digestion with either SmCE or SmCB2. Type I collagen was of particular interest, given that lower molecular weight peptides of the protein were only found in skin samples treated with SmCB2, suggesting it is cleaved by SmCB2 but not SmCE. To test this with purified protein, type I human collagen was treated with either SmCB2 or SmCE for up to 22 hours at 37°C, and cleavage of the protein was determined by SDS-PAGE analysis (Figure 2). While the majority of collagen I was degraded after 5 hours with SmCB2 (Figure 2A), SmCE treatment resulted in the appearance of discrete lower molecular weight bands only after 22 hours of enzyme treatment (Figure 2B). This confirms that SmCE shows reduced activity against type I collagen relative to SmCB2, even *in vitro*. To confirm that the two peptidases cleaved collagen at unique sites, candidate lower molecular weight bands resulting from peptidase treatment were submitted for N-terminal sequencing, and the resulting amino acid sequence was mapped onto the full protein to determine cleavage sites (Figure 2C). Consistent with previous analysis of SmCE substrate specificity, *in vitro* digestion of collagen I revealed that peptide bond cleavage only occurred following a leucine residue (VRGL/TGPI) [32,34]. In comparison, SmCB2 cleavage occurred following an

arginine residue (GER/GGP), which is consistent with its reported activity, including a level of “promiscuity” in its amino acid preference in the P2 substrate binding pocket, relative to other types of cathepsins [62,66].

Complement C3 was also of particular interest as a potential substrate of both SmCE and SmCB2, given its role in the host immune response against the parasite [67]. Purified complement C3 was treated with SmCB2 or SmCE. Discrete lower molecular weight bands were visible within 1 hour of treatment with either peptidase, in comparison to inhibited peptidase controls (Figure 3A, B). N-terminal sequencing of selected fragments again revealed that both SmCE and SmCB2 digested the protein in a manner consistent with their known specificities, with an arginine in the P1 position (RR/SVQ) for SmCB2 and and a tyrosine in the P1 position (TMY/HAK) for SmCE (Figure 3C).

4.5 Discussion

In *S. mansoni*, the most abundant peptidase in cercarial secretions is a serine peptidase, termed cercarial elastase (SmCE) for its ability to degrade insoluble elastin [27,30]. In addition to proteomic analysis, biochemical and immunolocalization studies have detected SmCE activity in cercarial secretions and confirmed that the enzyme is able to cleave such substrates as type IV collagen (basement membrane collagen), fibronectin, laminin and immunoglobulin *in vitro* [33,36,45]. Here, we have shown that SmCE cleaves

additional substrates in skin, including several types of collagen, other extracellular matrix proteins, and components of the complement cascade.

Recent sequencing and annotation of the *S. mansoni* genome suggests a unique role for cercarial elastase. An expanded gene family was identified with ten individual genes that encode multiple isoforms of the peptidase. Even without a complete genome, multiple orthologs of SmCE have been also been found in *S. haematobium*, a related human-specific species of schistosome common throughout North Africa and the Middle East [34]. This is not the case for the zoonotic *S. japonicum*, a schistosome species that infects humans and other mammals throughout southeast Asia. The *S. japonicum* genome contains only a single gene encoding cercarial elastase. This gene corresponds to the cercarial elastase "2b" isoform in *S. mansoni*, for which minimal transcript is made relative to other CE isoforms (Ingram and McKerrow, *unpublished*). While one report suggested that CE was detected by immunofluorescence in *S. japonicum* secretions, no cercarial elastase protein was detected in a high resolution mass spectrometric proteomic analysis of *S. japonicum* acetabular secretions, and no cercarial elastase-like activity was identified by direct biochemical assays [39,42]. *Trichobilharzia regenti*, an avian schistosome that is capable of invading human skin, but not establishing a successful infection in humans, encodes a cysteine peptidase, cathepsin B2 (TrCB2 (GenBank: ABS57370.1)), which has elastinolytic properties and localizes to the acetabular glands of the parasite [40]. *S. japonicum* also encodes a cathepsin B2 ortholog, and transcript is expressed

in the developing larval stage of the parasite. Moreover, proteomic analysis has identified cathepsin B2 as being present in *S. japonicum* cercarial secretions [42]. Notably, *S. japonicum* has 40- fold higher cathepsin B activity in its acetabular secretions, relative to *S. mansoni* secretions [39]. It is therefore likely that in *S. japonicum* cercariae, cathepsin B2, not cercarial elastase, is the predominant invasive enzyme.

The differential use of these two classes of peptidases raises the question of how their respective pH optima are achieved in schistosome secretions. SmCB2 is maximally active under acidic, reducing conditions [68]. Since the influence of *S. japonicum* cercarial secretions on the local environment of skin is unknown, SmCB2 incubations were performed under acidic conditions to ensure optimal peptidase activity. SmCE activity is optimal in a slightly alkaline environment, and *S. mansoni* secretions are also alkaline; therefore all SmCE incubations were performed at pH 8 [31]. Certainly, for *S. mansoni*, the evolutionary selection is most likely coordination of the pH of the acetabular gland secretions and the pH optimum of the peptidase. The pH optimum of the cercarial elastase is 8, and the pH of the secretions is also alkaline [41]. As *S. mansoni* cercariae migrate through skin, a microenvironment is created by the secreted material, which allows for optimal activity of the peptidase. The situation is less clear for *S. japonicum* and the *Trichobilharzia* cercariae. While some activity of the cathepsin B2 is likely to continue at neutral, or even alkaline pH, the pH optimum is slightly acidic [62]. The situation is reminiscent of the secretion of cathepsin B by

macrophages into tissue compartments of vertebrates. Secreted human cathepsin B is known to degrade extracellular matrix proteins in human tissue, where it has been reported to facilitate tumor invasion and metastasis [69]. The pH optimum of mammalian cathepsin B is also slightly acidic [70]. It is not known if the microenvironment around migrating macrophages is acidic or when that enzyme is released; however, it appears that there is sufficient cathepsin B activity to cause tissue degradation.

Given the unavailability of active, recombinant SjCB2 or sufficient amounts of *S. japonicum* cercariae from which to purify the native enzyme, we chose to perform our proteomic study with SmCB2, which displays high sequence homology (90% amino acid sequence identity for the mature peptidase) to its *S. japonicum* ortholog. We therefore hypothesized that it is likely to display similar biochemical characteristics, including similar substrate specificity. While we cannot say conclusively that SjCB2 is the protease facilitating *S. japonicum* cercarial invasion, we believe that our study, along with previous work from other groups, supports the proposed role for cathepsin B2 in host skin protein degradation [39,40].

This conclusion, that *S. japonicum* uses a cathepsin B2 peptidase for skin invasion, while *S. mansoni* uses a serine peptidase (SmCE), has implications for the evolution of the human host-parasite relationship in schistosomiasis. A plausible model is that the cathepsin B2 family first emerged as the functional cercarial peptidase during trematode evolution. In contrast, the “humanized”

parasites such as *S. mansoni* appear to have switched to a serine peptidase for cercarial invasion. This model is supported by the notable expansion of the serine peptidase gene family from the single 2b gene found in *S. japonicum* to the multiple isoforms expressed in *S. mansoni* [34,42]. While the genome of the other "humanized" parasite, *S. haematobium*, has not been completed, it is already clear from EST analysis that more abundant serine peptidase isoforms are present in that genome [34].

What is the advantage of a larval serine peptidase for the "humanized" schistosomes? It is interesting to note that by BLAST analysis, some of the proteins with highest homology to cercarial elastase are mammalian mast cell peptidases, which are present in skin [39]. It is therefore possible that cercarial elastase evolved by convergence to resemble a human peptidase, in order to evade detection by the host immune system. Previous work shows that *S. mansoni* cercariae migrate through skin at a much slower rate than their *S. japonicum* counterparts [71]. Despite this, an inflammatory response to *S. japonicum* cercariae occurs more frequently than to *S. mansoni* cercariae [71,72]. Cathepsin B2 is a likely target of the inflammatory response, given that many cysteine peptidases are allergenic [73]. Perhaps the rapid transit of non-humanized cercariae through skin precludes the need for an invasive enzyme that mimics a host peptidase. Other aspects of immune evasion, such as the elimination of complement factors and immunoglobulin, may be common to both species. C3

and C4 components bind to the tegument of schistosomes, but are degraded by both SmCE and SmCB2 [67,74].

The results reported here show that S. *mansoni* cathepsin B2 (a model for *S. japonicum* cathepsin B2) and *S. mansoni* cercarial elastase are both capable of degrading proteins in skin that act as a barrier to cercarial invasion. Many skin proteins are substrates for both enzymes, but cathepsin B2 appears to cleave a broader range of substrates, and therefore may be a more effective invasive enzyme than cercarial elastase.

CHAPTER 4
FIGURES AND TABLES

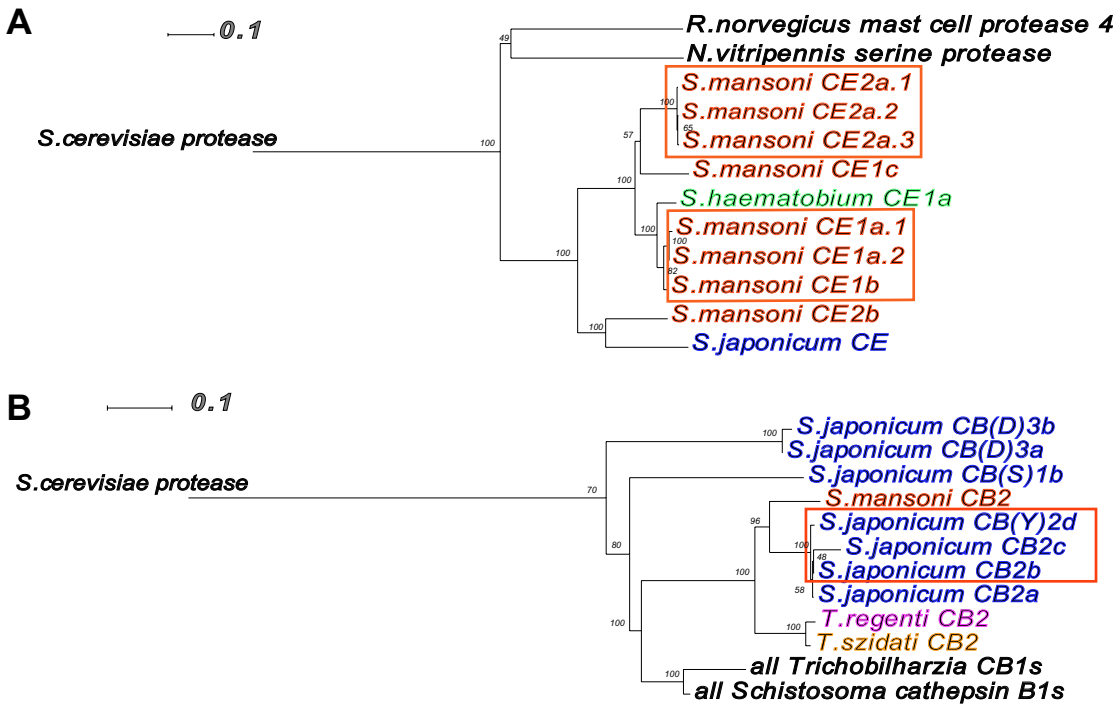


Figure 4.1 Differential expansion of select peptidase gene families in schistosomes species. Phylogenetic analysis of cercarial elastase (A) and cathepsin B2 (B) protein sequences reveals expansion of each gene family in different lineages of schistosomes. Boxes indicate proteins previously determined to be present in cercarial secretions, as determined by LC MS/MS.

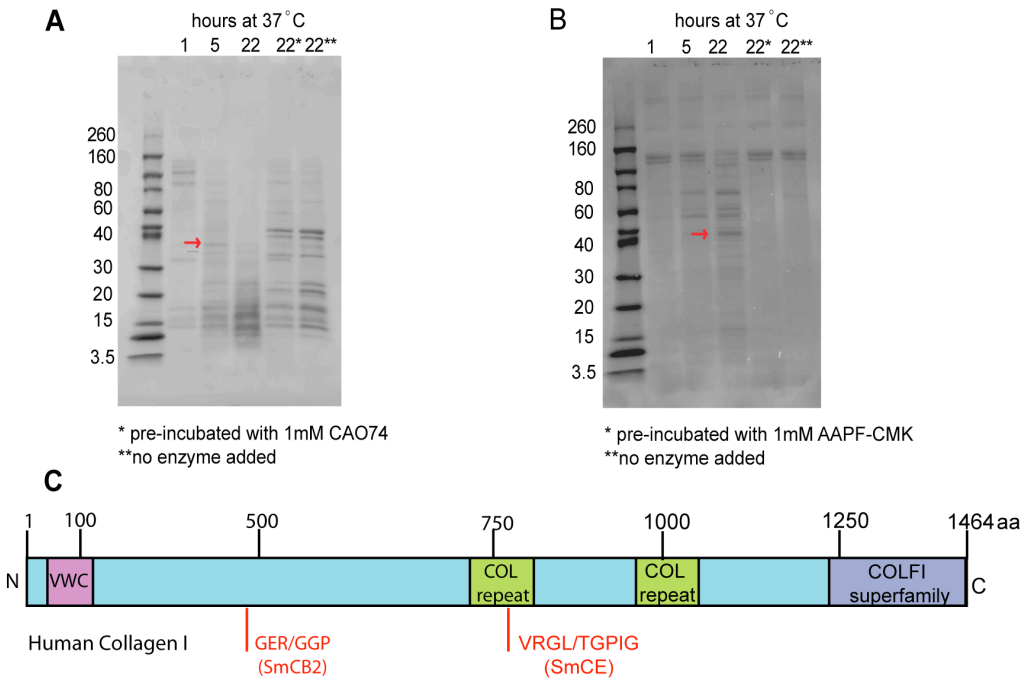


Figure 4.2 Human collagen I is preferentially cleaved by SmCB2. *In vitro* cleavage of human collagen I confirms *ex vivo* analysis, and reveals differential cleavage by CE and CB2. (A) SmCB2 digestion of collagen I (B) SmCE digestion of collagen I. Digestion reactions were performed for 1-22 hours at 37°C . * indicates pre-incubation with (A) 1 mM CAO74 or (B) 1 mM AAPF-CMK. ** indicates no peptidase control. Arrows indicate bands submitted for Edman degradation. (C) Schematic of human collagen I, indicating cleavage sites as determined by Edman degradation (VWC-Von Willenbrand Factor Type C domain; COL- collagen triple helix repeat). (MGI-macroglobulin-1; A2M-alpha-2-macroglobulin; ANA-.Anaphylatoxin homologous domain; NTR-netrin-like domain).

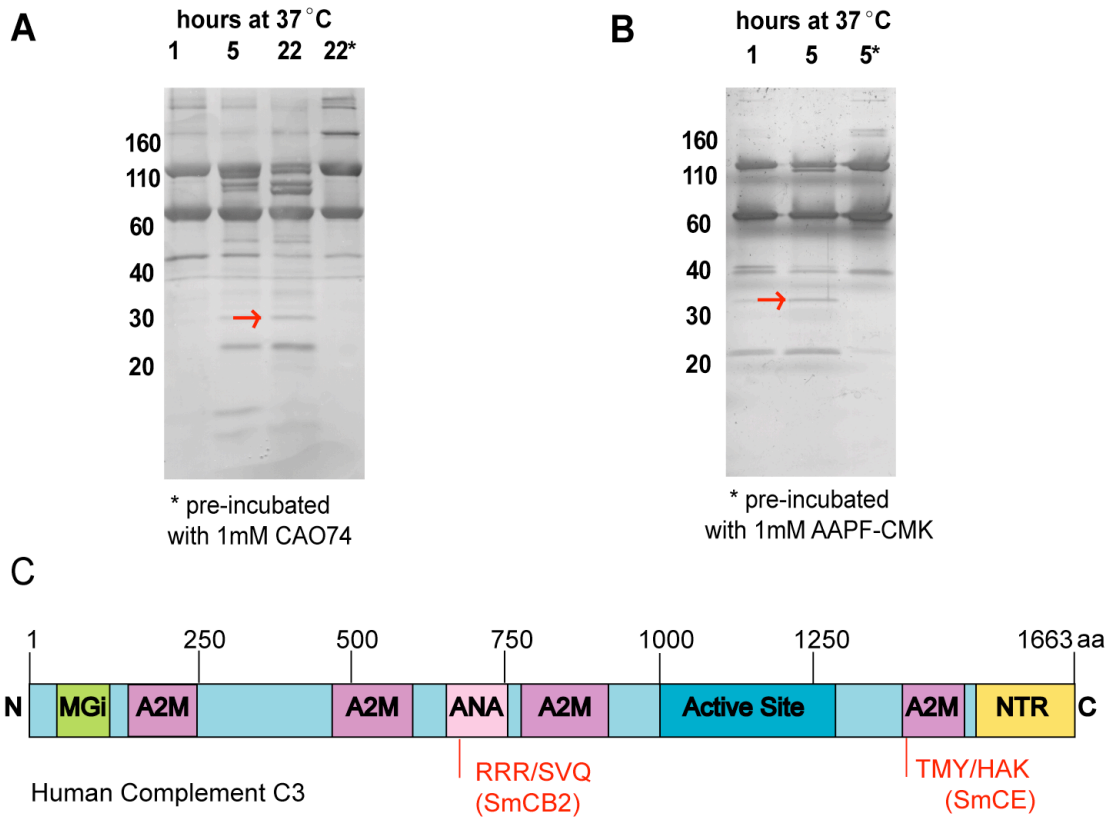


Figure 4.3 Human Complement C3 is cleaved by both SmCE and SmCB2. *In vitro* cleavage of human complement C3 protein confirms *ex vivo* analysis, and reveals differential cleavage by CE and CB2. (A) SmCB2 digestion of complement C3. (B) SmCE digestion of complement C3. Digestion reactions were performed for 1-22 hours (SmCB2) or 1-5 hours (SmCE) at 37°C . * indicates pre-incubation with (A) 1 mM CAO74 or (B) 1 mM AAPF-CMK. ** indicates no peptidase control. Arrows indicate bands submitted for Edman degradation. (C) Schematic of human complement C3, indicating cleavage sites as determined by Edman degradation.

Table 4.1. Substrates of SmCE and SmCB2 identified in *ex vivo* skin

Category		Substrate of		
Accession Number	**Protein Name**	**SmCE**	**SmCB2**	**SmCE and SmCB2**
Extracellular				
P02741	C-reactive protein		x	
P01859	Ig gamma-1 chain C region		x	
P01042	Kininogen-1		x	
P07996	Thrombospondin-1		x	
P17931	Galectin-3	x	x	x
Extracellular Matrix				
P07355	Annexin A2		x	
P98160	Basement membrane heparan sulfate proteoglycan		x	
P16070	CD44 antigen		x	
P02452	Collagen alpha-1(I) chain		x	
P39060	Collagen alpha-1(XVIII) chain		x	
P02751	Fibronectin	x	x	x
P98095	Fibulin-2		x	
Q15063	Periostin		x	
P24821	Tenascin		x	
P22105	Tenascin-X	x	x	x
Q15582	Transforming growth factor-beta-induced protein ig-h3		x	
P04004	Vitronectin	x	x	x
P12111	Collagen alpha-1(XII) chain	x	x	x
Extracellular immune component				
P01024	Complement C3	x	x	x
P0C0L4	Complement C4-A		x	
P00746	Complement factor D		x	
P01023	Alpha-2-macroglobulin	x	x	x
Interstitial tissue				
P02461	Collagen alpha-1(III) chain		x	
P12109	Collagen alpha-1(VI) chain	x	x	x
P12111	Collagen alpha-3(VI) chain	x	x	x
Keratinocytes				
P35527	Keratin, type I cytoskeletal 9		x	
P31944	Caspase-14		x	
Plasma				
P01008	Antithrombin-III		x	
P00488	Coagulation factor XIII A chain		x	

P02671	Fibrinogen alpha chain		**x**	
P00738	Haptoglobin		**x**	
Q14624	Inter-alpha-trypsin inhibitor heavy chain H4		**x**	
P02768	Serum albumin	**x**	**x**	**x**
P07339	Cathepsin D		**x**	

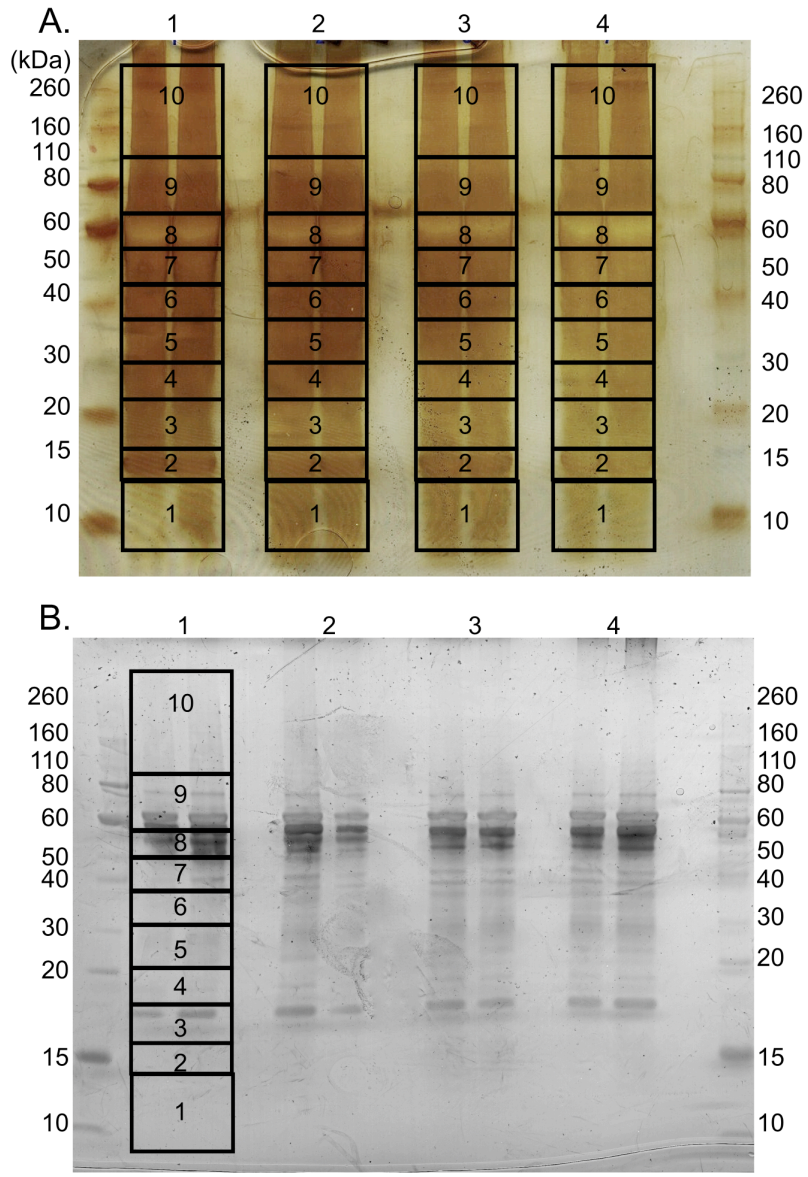


Figure S4.1 Preparative and analytical SmCE SDS-PAGE gels. (A) Preparative SDS-PAGE gel with duplicate lanes loaded with skin lysate, treated with (1) 180 nM SmCE, (2) 1.8 µM SmCE, (3) 2 µM AAPF-CMK followed by 2 µM SmCE, or (4) no enzyme. (B) The same samples loaded at the analytical scale, at 1/10 concentration compared to (A).

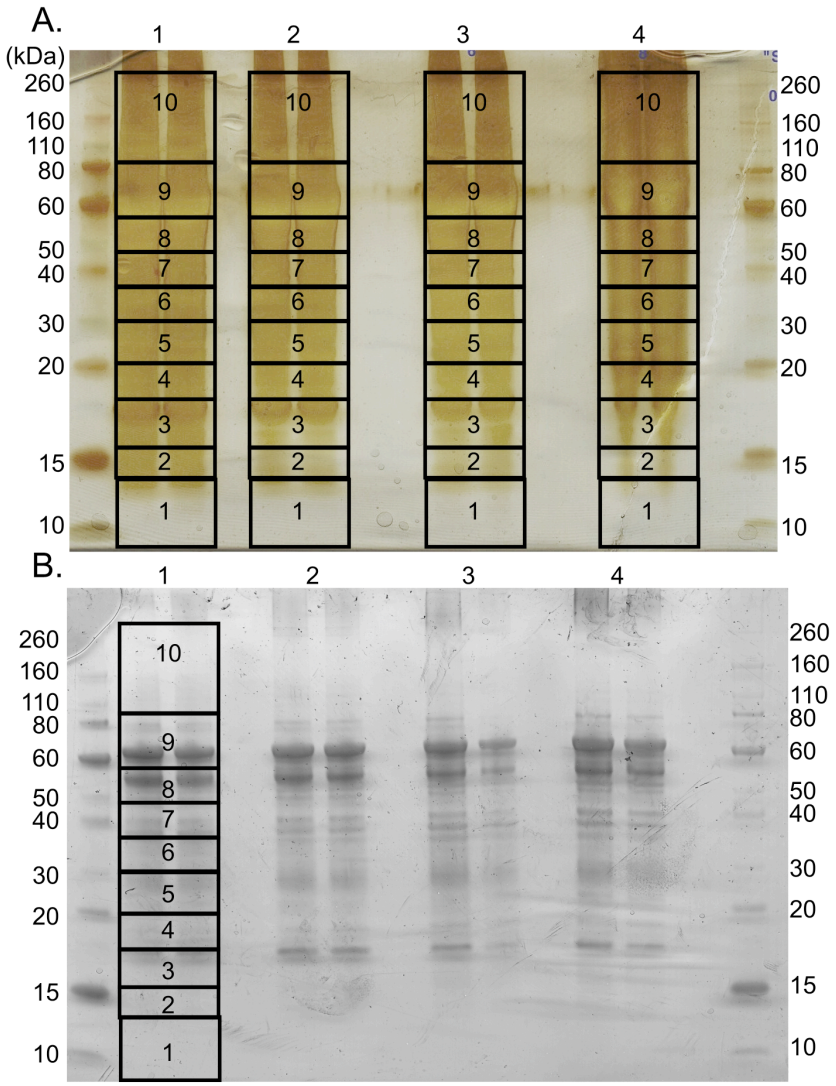


Figure S4.2 Preparative and analytical SmCB2 SDS-PAGE gels. (A) Preparative SDS-PAGE gel with duplicate lanes loaded with skin lysate, treated with (1) 180 nM SmCB2, (2) 1.8 μM SmCB2, (3) 1.8 μM CA074 + 1.8 μM SmCB2, or (4) 1.8 uM CA074. (B) The same samples loaded at the analytical scale, at 1/10 concentration compared to (A).

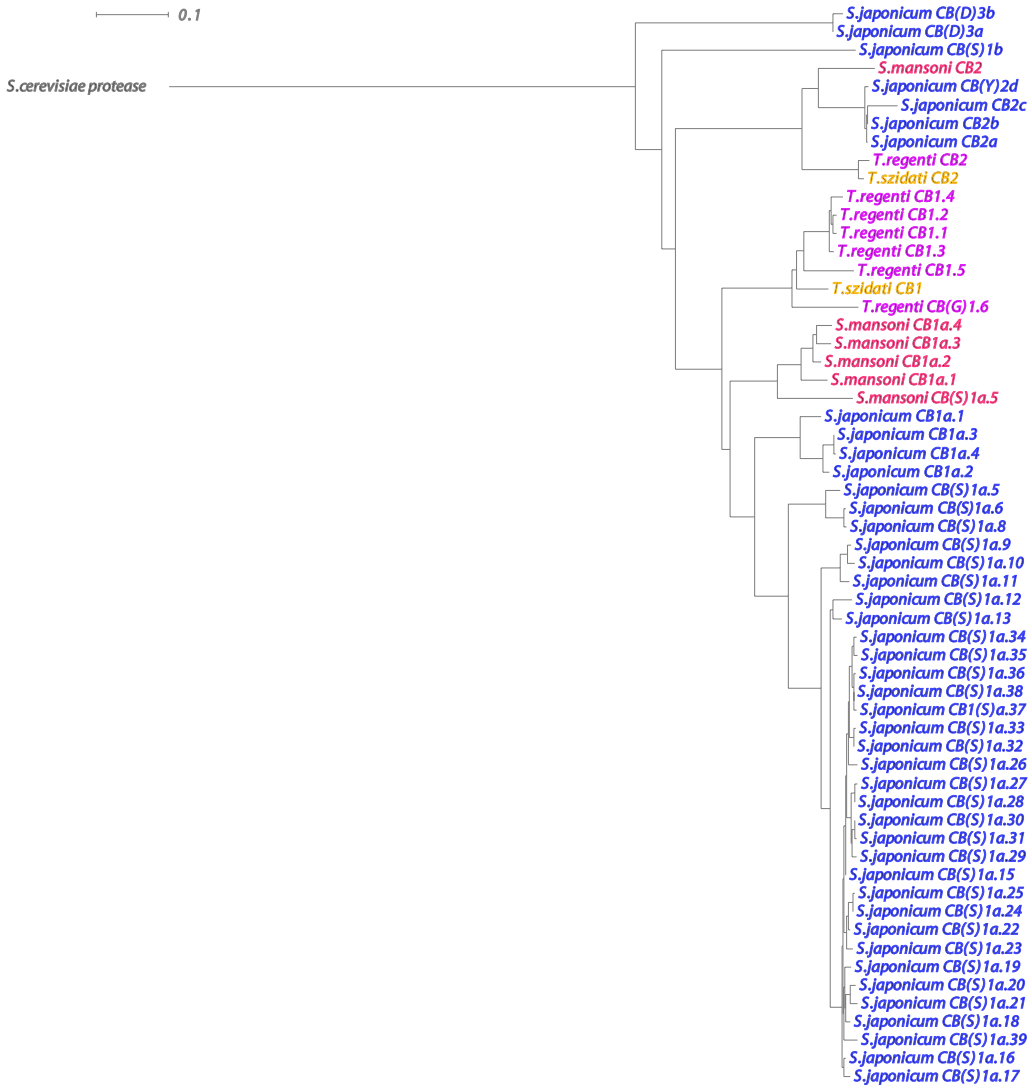


Figure S4.3 Expanded phylogenetic analysis of all known schistosome cathepsin B proteins

Figure S4.4 SmCB2 and SjCB2 protein alignment. ClustalW alignment of SmCB2 (GenBank: CAZ31207.1) and SjCB2 (GenBank: AAO59414.2) protein sequences shows high sequence identity.

Table S1. Complete list of schistosome cathepsin B sequences. All available schistosome cathepsin B isoforms from *S. mansoni* and *S. japonicum* were compiled from GenBank and their respective annotation websites. For each, the name used for this study, previous identifiers, and the full amino acid sequence are listed. Active site residues are highlighted in red. *S. mansoni* sequences are highlighted in pink, and *S. japonicum* sequences are highlighted in blue.

New name	Other identifiers	Protein Sequence
S.mansoni CB2	SmCB2	MNQYSCYLLQLYIIILLSYGTLNEIDARRHKRMYQPLSMELINFINYEANTTWKAAPTTRFRTVSDIRRM
	Smp_141610	LGALPDPNGEQLETLCTGYISDELPKSFDARVEWPHCPSISEIRDQSSCGS**C**WAFGAVEAMSDRICIKSK
	gi|18181863|emb|CAC85211.2|	GKHKPFLSAENLVSCCSSCGMGCNGGFPHSAWLYWKNQGIVTGDLYNTTNGCQPYEFPPCEHHVIGPLPS
	cathepsin B endopeptidase [Schistosoma mansoni]	CDGDVETPSCKTNCQPGYNIPYEKDKWYGEKVYRIHSNPEAIMLELMRNGPVEVDFEVYADFPNYKSGVY
		QHVSGALLGG**H**AVRLLGWGEENNVPYWLIA**N**SWNSDWGDKGYFKIVRGKNECGIESDVNAGIPKIKN
S.mansoni CB1a.4	SmCB1a.4	MNYLDMLTSVLCIASLITHLDAHISIKNEKFKPLSDDIIS
	Smp_067060	YINEHPNAGWRAEKSNRFHSLDDARIQMGARREEPDLRRK
	gi|22531389|emb|CAD44625.1|	RRPTVDHNEWNVEIPSNFDSRKKWPGCKSIATIRDQSRCG
	cathepsin B1 isotype 2 [Schistosoma mansoni]	S**C**WAFGAVEAMSDRSCIQSGGKQNVELSAVDLLSCCESCG
		LGCEGGILGPAWDFWVKEGIVTGSSKENHTGCEPYPFPKC
		EHHTKGKYPPCGSKIYKTPRCKQTCQKKYKTPYTQDKHRG
		KSSYNVKNDEKAIQKEIMKYGPVEASFTVYEDFLNYKSGI
		YKHITGEALGG**H**AIRIIGWGVENKTPYWLIA**N**SWNEDWGE
		NGYFRIVRGRDECFIESEVIAGQIN
S.mansoni CB1a.2	SmCB1a.2	MLTSILCIASLITFLEAHISVKNEKFEPLSDDIISYINEHPNAGWRAEKSNRFHSLDDARIQMGARREEP
	Smp_103610	DLRRTRRPTVDHNDWNVEIPSSFDSRKKWPRCKSIATIRDQSRCGS**C**WAFGAVEAMSDRSCIQSGGKQNV
	gi|227284436|emb|CAY16997.1|	ELSAVDLLSCCESCGLGCEGGILGPAWDYWVKEGIVTGSSKENHTGCEPYPFPKCEHHTKGKYPPCGSKI
	gi|256052329|ref|XP_002569725.1|	YKTPRCKQTCQKKYKTPYTQDKHRGKSSYNVKNDEKAIQKEIMKYGPVEAGFTVYEDFLNYKSGIYKHIT
	gi|256090368|ref|XP_002581167.1|	GETLGG**H**AIRIIGWGVENKTPYWLIA**N**SWNEDWGENGYFRIVRGRDECSIESEVTAGRIN
	gi|22531387|emb|CAD44624.1|	
	cathepsin B1 isotype 1 [Schistosoma mansoni]	
S.mansoni CB1a.3	SmCB1a.3	MLTSILCIASLITFLEAHISVKNEKFEPLSDDIISYINEHPNAGWRAEKSNRFHSLDDARIQMGARREEP
	gi|160950|gb|AAA29865.1|	DLRRKRRPTVDHNDWNVEIPSNFDSRKKWPGCKSIATIRDQSRCGS**C**WSFGAVEAMSDRSCIQSGGKQNV
	cathepsin B [Schistosoma mansoni]	ELSAVDLLTCCESCGLGCEGGILGPAWDYWVKEGIVTASSKENHTGCEPYPFPKCEHHTKGKYPPCGSKI
		YNTPRCKQTCQRKYKTPYTQDKHRGKSSYNVKNDEKAIQKEIMKYGPVEASFTVYEDFLNYKSGIYKHIT
		GEALGG**H**AIRIIGWGVENKTPYWLIA**N**SWNEDWGENGYFRIVRGRDECSIESEVIAGRIN
incomplete sequence	SjCB4	MNLLPIINLILLILIKNSCYALHQYDRQFPPLSDELITFINKQPNIEWKADRTTRFTSIHHAKSMMGVLL
	gi|29840882|gb|AAP05883.1| similar to X70968	NRVDQHKLHHPIIHHNDINIKLPKYFDSRKYWKNCSSIRTIRDQSSCGS**C**WAFGAVESMSDRICIHSKGR
	cathepsin B in Schistosoma japonicum	ISIELSAVNLLSCCSRCGFGCNGGIPGMAWDYWKDEGIVTGGSNETHTGCQPYPFPECIHHSTSINHSSC
		EVKYYSTPECYQTCQPDYAIQYENDKYYGKSSYYVTSDEVSIMKEILLNGPVEATFYVYDDFLNYKTGVY
		KYVTGSLLGG**H**AIRITWLGCIHIESYTILVMC
S.japonicum CB1a.1	SjCB1a.1	MLKIAVYIVSLFTFLEAHVTTRNNQRIEPLSDEMISFINEHPDAGWKADKSDRFHSLDDARILMGARKED
	gi|11167|emb|CAA50305.1|	AEMKRNRRPTVDHHDLNVEIPSQFDSRKKWPHCKSISQIRDQSRCGS**C**WAFGAVEAMTDRICIQSGGGQS
	cathepsin B [Schistosoma japonicum]	AELSALDLISCCKDCGDGCQGGFPGVAWDYWVKRGIVTGGSKENHTGCQPYPFPKCEHHTKGKYPACGTK
		IYKTPQCKQTCQKGYKTPYEQDKHYGDESYNVQNNEKVIQRDIMMYGPVEAAFDVYEDFLNYKSGIYRHV
		TGSIVGG**H**AIRIIGWGVEKRTPYWLIA**N**SWNEDWGEKGLFRMVRGRDECSIESDVVAGLIKT
S.japonicum CB(S)1a.34	SjCB(S)1a.34	MLKIAVYIVSLFTLLEAHVTTRNNERIEPLSDEMISFINEHPNAGWKADKSDRFHSVDDARFLLGGRKED
	gi|226474184|emb|CAX71578.1|	PNLREKRRPTVDHHDLNVEIPSHFDSRKKWPRCKSISQIRDQSQCGS**S**WAVSAVGAMSDRICIQSGGKQS
	Cathepsin B-like cysteine proteinase precursor [Schistosoma japonicum]	VELSAVDLISCCKYCGSGCDGGFLGPSWDYWVLRGIVTGGSKENHTGCRPYPFPKCDHFVKGKYRACGDK
		LYKTPQCNQTCQKGYNTSYEQDKHYGGFSYNVLSVESVIQKDIMVHGPVEAYLEIYEDFLNYKSGIYRYT
		TGKYISG**H**AVRLIGWGVENGTAYWLAA**N**TWNEDWGEKGYFRIVRGRNECLIESEIAAGLIKS

S.japonicumCB(S)1b	SjCB(S)1b	MYISILSIYLLLYLKVCNSKQWETFFDEQIRFLNNHPSSGLKASKHNRFTAISDVYSALEYYGEKQFRHH
	gi|226466816|emb|CAX69543.1|	ILPIISHDDDNILLPDYFDSREQWKNCPSIKRIYDQSQCYSSWAMASVAAISDRICIQTNGTVKVELSAI
	Cathepsin B-like cysteine proteinase precursor [Schistosoma japonicum]	ELVSCCSKCAVGCNFGYSESAWYYWVENGLVTGESNGNNSGCLPYPFPKCDHGSSDSYPMCGYVVYTPPV
		CNGTCRPGYPIPYNDDKHFGKSAYQVKQNESDIRREIMLYGPVEASIFIYDDFVDYKSGVYKHLTGRLIT
		IQSVRIIGWGIENGIPYWLCANSWNEEWGLNGFFKILRGSNECEIEAFVNAGRVDMK
S.japonicum CB(D)3b	SjCB(D)3b	MSCLVVILLLNIICNCELNAVENEHIEPLFGKLIEYVNRNPKFGWKAGTNHRFRSSKDIEKMFRKYIEIE
	gi|226466652|emb|CAX69461.1|	NIQTKHIKTISHNSINMEIPRSFDARYHWINCSTIRQIHDESLCRADWAIATVDSISDRICIRSNGRISV
	Cathepsin B-like cysteine proteinase precursor [Schistosoma japonicum]	QLSARDAISCGFSPGCFHGSEVEVLVYWITYGIVTGGSYEDQSGCQPYPLPKCSYHPESRFLDCNNNTFE
		FPQCTNECQDGYNKTYDDDKFYGERIYNVYGTQEDIQKEILMNGPVIASISVNTDFLVYKSGVYLPTPRS
		RNLGWITLRIIGWGYEGKIPYWLCANSWNEEWGDNGYVKIQRGVQAGYIESYVRAPIPKM
S.mansoni CB1a.1	SmCB1a.1	MLISVLYIASLISHLEAHISIKNEKFEPLSDDIISYINEHPNAGWRAEKSNRFHSLDDARFQLGARREEP
	Smp_179950	DLRRTRRPTVDHNDWNVEIPSSFDSRKKWPRCKSIATIRDQSRCGSCCAFGAVEAMSERSCIQSGGKQNV
	gi|256090364|ref|XP_002581165.1|	ELSAVDLEGIVTGSSKENNTGCEPYPFPKCEHFTKGQYPPCGSKIYKTPRCKTTCQKRYKTSYAQDKHRA
	gi|238667006|emb|CAZ37404.1|	IQKEIMKYGPVEASFTVYEDFLNYKSGIYKHITGETLGGHAIRIIGWGVENKTPYWLIANSWNEDWGENG
	cathepsin B-like peptidase (C01 family) [Schistosoma mansoni]	YFRIVRGRDECSIESEVTAGRIN
S.japonicum CB(D)3a	SjCB(D)3a	MFRKYIETENIQTKHIKTISHNSINMEIPRSFDARYHWINCSTIRQIHDESLCRADWAIATVDSISDRIC
	gi|56755295|gb|AAW25827.1|	IRSNGRISVQLSARDAISCGFSPGCFHGSEVEVLVYWITYGIVTGGSYEDQSGCQPYPLPKCSYHPESRF
	SJCHGC06356 protein [Schistosoma japonicum]	LDCNNNTFEFPQCTNECQDGYNKTYDDDKFYGERIYNVYGTQEDIQKEILMNGPVIASISVNTDFLVYKS
		GVYLPTPRSRNLGWITLRIIGWGYEGKIPYWLCANSWNEEWGANGYVKIQRGVQAGYIESYVRAPIPKM
S.japonicum CB(S)1a.12	SjCB(S)1a.12	MLKIAVYIVSLFTLLEAHVTKRNNQRIEPLSDEMISFINKHPNAGWKADKSDRFHSVDDARILLGGRKED
	gi|56756114|gb|AAW26235.1|	SNLRQKRRPTVDHHDLNVEIPSHFDSRKKWPRCKSISQIRDQSRCASSWAVSSVGAMSDRICIQSGGKQS
	unknown [Schistosoma japonicum]	VELSAIDLISCCKNCGSGCDGGYFLPSWDYWVSHGIVTGGSKENHTGCRPYPFPKCDHFVKGKYRACGDK
		LYETPQCKQTCQKGYNTSYEQDKHYGGFSYNVLSVESVIQKDIMMHGPVEAYLEIYEDFLNYKSGIYRYT
		TGKYISGHAVRLIGWGVENGTAYWLAANTWNEDWGEKGYFRIVRGRNECLIESEIAAGLIKS
S.japonicum CB(S)1a.26	SjCB(S)1a.26	MLKIAVYIVSLFNLLEAHVTTRNNQRIEPLSDEMISFINKHPNAGWKADKSDRFHSVDDARILLGGRRED
	gi|56758864|gb|AAW27572.1|	PNLREKRRPTVDHHDLNVEIPSHFDSRKKWPRCKSISQIRDQSQCGSSWAVSAVGAMSDRICIQSGGKQS
	unknown [Schistosoma japonicum]	VELSAVDLISCCKYCGSGCDGGFLGPSWDYWVLRGIVTGGSKENHTGCRPYPFPKCDHFVKGKYRACGDK
		LYKTPQCNQTCQKGYNTSYEQDKHYGGFSYNVLGIESVIQKDIMMHGPVEAYLEIYEDFLNYKSGIYRYT
		TGKYISGHAVRLIGWGVENGTAYWLAANTWNEDWGEKGYFRIVRGRNECLIESEIAAGLIKS
S.japonicum CB(S)1a.9	SjCB(S)1a.9	MLNIAFCIVSLFTLLEAHVTTRNNQRIEPLSDEMISFINKHPNAGWKADKSDRFHSVDDARILLGGRKED
	gi|56758716|gb|AAW27498.1|	PNLRQKRRPTVDHHDLNVEIPSHFDSRKKWPRCKSISQIRDQSRCASSWAVSAVAAMSDRICIQSGGKQS
	unknown [Schistosoma japonicum]	VELSAIDLISCCENCGSGCDGGVTGYSWDYWVKHGIVTGGSKENHTGCRPYPFPKCDHFVKGKYRACGDK
		LYKTPQCKQTCQKGYNTSYEQDKHYGGFSYSVIGVESAIQKEIMMYGPVEAYLEIYEDFLNYKSGIYRYT
		TGKYISGHAVRLIGWGVENGTAYWLAANTWNEDWGEKGYFRIVRGRDECLIESFIVAGQIKS
S.japonicum CB(S)1a.23	SjCB(S)1a.23	MLNIAFCIVSLFTLLEAHVTTRNNERIEPLSDEMISFINEHPNAGWKADKSDRFHSVDDARILLGGRRED
	gi|56757646|gb|AAW26973.1|	PNLREKRRPTVDHHDLNVEIPSHFDSRKKWPRCKSISQIRDQSQCGSSWAVSAVGAMSDRICIQSGGKQS
	unknown [Schistosoma japonicum]	VELSAVDLISCCKYCGSGCDGGFLGPSWDYWVLRGIVTGGSKENHTGCRPYPFPKCDHFVKGKYRACGDK
		LYKTPQCNQTCQKGYNTSYEQDKHYGGFSYNVLSGESVFQKDIMMHGPVEAYLEIYEDFLNYKSGIYRYT
		TGKYISGHAVRLIGWGVENGTAYWLAANTWNEDWGEKGYFRIVRGRNECSIESEIAAGLIKS

S.japonicumCB(S)1a.11	SjCB(S)1a.11	MLNIAFCIVSLFTLLEAHVTTRNNQRIEPLSDEMILFINKHPNAGWKADKSDRFHSVDDARILLGGRRED
	gi|56757271|gb|AAW26807.1|	PNLREKRRPTVDHHDLNVEIPSHFDSRKKWPRCKSISQIRDQSRCAS**S**WAVSAVGAMSDRICIQSGGKQS
	unknown [Schistosoma japonicum]	VELSAIDLISCCKNCGSGCDGGVTGYSWDYWVKHGIVTGGSKENHTGCRPYPFPKCDHFVKGKYRACGDK
		LYKTPQCKQTCQKGYNTSYEQDKHYGEFSYNVIGVESVIQKEIMMYGPVEAYLHIYEDFLNYKSGIYRYT
		TGQFISG**H**AVRLIGWGVENGTSYWLAA**N**TWNEDWGEKGYFRIVRGRDECLIESFIVAGQIKS
S.japonicum CB(S)1a.32	SjCB(S)1a.32	MLKIAVYIVSLFTLLEAHVTTRNNERVEPLSDEMISFINEHPNAGWKADKSDRFHSVDDARILLGGRRED
	gi|56756907|gb|AAW26625.1|	PNLREKRRPTVDHHDLNVEIPSHFDSRKKWPRCKSISQIRDQSQCGS**S**WAVSAVGAMSDRICIQSGGKQS
	unknown [Schistosoma japonicum]	VELSAVDLISCCKYCGSGCDGGFLGPSWDYWVLRGIVTGGSKENHTGCRPYPFPKCDHFVKGKYRACGDK
		LYKTPQCNQTCQKGYNTSYEQDKHYGGFSYNVLSVESVIQKDIMMHGPVEAYLEIYEDFLNYKSGIYRYT
		TGKYISG**H**AVRLIGWGVENGTAYWLAA**N**TWNEDWGEKGYFRIVRGRNECLIESEIAAGLIKS
S.japonicum CB(S)1a.17	SjCB(S)1a.17	MLNIAFCIVSLFTLLEAHVTTRNNQRIEPLSDEMISFINKHPNAGWKADKSDRFHSVDDARNLLGGRRED
	gi|56756475|gb|AAW26410.1|	PNLRQKRRPTVDHHDLNVEIPSHFDSRKKWPRCKSISQIRDQSQCGS**S**WAVSAVGAMSDRICIQSGGKQS
	unknown [Schistosoma japonicum]	VELSAVDLISCCKYCGSGCDGGFLGPSWDYWVLRGIVTGGSKENHTGCRPYPFPKCDHFVKGKYRACGDK
		LYKTPQCKQTCQKGYNTSYEQDKHYGGFSYNVLSVESVIQKDIMMHGPVEAYLEIYEDFLNYKSGIYRYT
		TGKYISG**H**AVRLIGWGVENGTAYWLAA**N**TWNEDWGEKGYFRIVRGRNECLIESEIAAGLIKS
S.japonicum CB(S)1a.29	SjCB(S)1a.29	MLKIAVYIVSLFNLLEAHVTTRNNERIEPLSDEMISFINKHPNAGWKADKSDRFHSVDDARILLGGRKED
	gi|56756410|gb|AAW26378.1|	PNLRQKRRPTVDHHDLKVEIPSHFDSRKKWPRCKSISQIRDQSQCGS**S**WAVSAVGAMSDRICIQSGGKQS
	unknown [Schistosoma japonicum]	VELSAVDLISCCKYCGSGCDGGFLGPSWDYWVLRGIVTGGSKENHTGCRPYPFPKCDHFVKGKYRACGDK
		LYKTPQCKQTCQKGYNTSYEQDKHYGGFSYNVLSVESVIQKDIMMHGPVEAYLEIYEDFLNYKSGIYRYT
		TGQFISG**H**AVRLIGWGVENGTAYWLAA**N**TWNEDWGEKGYFRIVRGRNECSIESEIAAGLIKS
S.japonicum CB(S)1a.39	SjCB(S)1a.39	MLNIAFCIVSLFTLLEAHVTTRNNERIEPLSDEMISFINEHPNAGWKADKSDRFHSVDDARILLGGRKED
	gi|56756380|gb|AAW26363.1|	PNLRQRRRPTVDHHDLNVEIPSHFDSRKKWPRCKSISQIRDQSQCGS**S**WAVSAIGAMSDRICIQSGGKQS
	unknown [Schistosoma japonicum]	VKLSAVDLISCCENCGSGCDGGFLGPSWDYWVLRGIVTGGSKENHTGCRPYPFPKCDHFVKGKYRACGDK
		LYKTPQCNQTCQKGYNTSYEQDKHYGGFSYNVLSVESVIQKDIMMHGPVEAYLEIYEDFLNYKSGIYRYT
		TGQFISG**H**AVRLIGWGVENGTAYWLAA**N**TWNEDWGEKGYFRIVRGRNECLIESEIAAGLIKS
S.japonicum CB(S)1a.21	SjCB(S)1a.21	MLNIAFCIVSLFTLLEAHVTKRINQRIEPLSDEMISFINKHPNAGWKADKSDRFHSVDDARILLGGRKED
	gi|56755451|gb|AAW25905.1|	PNLRQKRRPTVDHHDLKVEIPSHFDSRKKWPRCKSISQIRDQSQCGS**S**WAVSAVGAMSDRICIQSGGKQS
	unknown [Schistosoma japonicum]	VELSAVDLISCCKYCGSGCDGGFLGPSWDYWVLRGIVTGGSKENHTGCRPYPFPKCDHFVKGKYRACGDK
		LYKTPQCKQICQKGYNTSYEQDKHYGGFSYNVLSVESVIQKDIMMHGPVEAYLEIYEDFLNYKSGIYRYT
		TGQFISG**H**AVRLIGWGVENGTAYWLAA**N**TWNEDWGEKGYFRIVRGRNECSIESEIAAGLIKS
S.japonicum CB(S)1a.22	SjCB(S)1a.22	MLNIAFCIVSLFTLLGAHVTTRNNERIEPLSDEMISFINKHPNAGWKADKSDRFHSVDDARILLGGRRED
	gi|56754499|gb|AAW25437.1|	PNLREKRRPTVDHHDLNVEIPSHFDSRKKWPRCKSISQIRDQSQCGS**S**WAVSAVGAMSDRICIQSGGKQS
	unknown [Schistosoma japonicum]	VELSAVDLISCCKYCGSGCDGGFLGPSWDYWVLRGIVTGGSKENHTGCRPYPFPKCDHFVKGKYRACGDK
		LYKTPQCNQTCQKGYNTSYEQDKHYGGFSYNVLSVESVIQKDIMMHGPVEAYLEIYEDFLNYKSGIYRYT
		TGKYISG**H**AVRLIGWGVENGTAYWLAA**N**TWNEDWGEKGYFRIVRGRNECLIESEIAAGLIKS
S.japonicum CB(S)1a.13	SjCB(S)1a.13	MISFINKHPNAGWKADKSDRFHSVDDARILLGGRREDPNLREKRRPTVDHHDLNVEIPSHFDSRKKWPRC
	gi|56754307|gb|AAW25341.1|	KSISQIRDQSRCAS**S**WAVSAVGAMSDRICIQSGGKQSVELSAIDLISCCKNCGSGCDGGVTGYSWDYWVS
	unknown [Schistosoma japonicum]	HGIVTGGSKENHTGCRPYPFPKCDHFVKGKYRACGDKLYKTPQCKQTCQKGYNTSYEQDKHYGGFSYNVL
		SVESVIQKDIMMHGTVEAYLEIYEDFLNYKSGIYRYTTGQFISG**H**AVRLIGWGVENGTAYWLAA**N**TWNED
		WGEKGYFRIVRGRNECLIESEIAAGLIKS

S.japonicum CB(S)1a.10	SjCB(S)1a.10	MLNIAFCIVSLFTLLEAHVTTRNNQRIEPLSDEMILFINKHPNAGWKADKSDRFHSVDDARILLGGRRED
	gi|56752925|gb|AAW24674.1|	PNLRQKRRPTVDHHDLNVEIPSHFDSRKKWPRCKSISQIRDQSRCAS**S**WAVSAVAAMSDRICIQSGGKQS
	unknown [Schistosoma japonicum]	VELSAIDLISCCKNCGSGCDGGVTGYSWDYWVKHGIVTGGSKENHTGCRPYPFPKCDHFVKGKYRACGDK
		LYKTPQCKQTCQKGYNTSYEQDKHYGGFSYSVIGVESAIQKEIMMYGPVEAYLQIYEDFLNYKSGIYRYT
		TGKYISG**H**AVRLIGWGVENGTSYWLAA**N**TWNEDWGEKGYFRIVRGRDECLIESFIVAGQIKS
S.japonicum CB(S)1a.18	SjCB(S)1a.18	MLNIAFCIVSLSTLLEAHVTTRNNQRIEPLSDEMISFINEHPNAGWKADKSDRFHSVDDARILLGGRRED
	gi|56752809|gb|AAW24616.1|	PNLREKRRPTVDHHDLKVEIPSHFDSRKKWPRCKSISQIRDQSQCGS**S**WAVSAVGAMSDRICIQSGGKQS
	unknown [Schistosoma japonicum]	VELSAVDLISCCKYCGSGCDGGFLGPSWDYWVLRGIVTGGSKENHTGCRPYPFPKCDHFVKGKYRACGDK
		LYKTPQCKQTCQKGYNTSYEQDKHYGGFSYNVLSVESVIQKDIMMHGPVEAYLEIYEDFLNYKSGIYRYT
		TGQFISGHAVRLIGWGVENGTAYWLAANTWNEDWGEKGYFRIVRGRNECSIESEIAAGLIKS
S.japonicum CB(S)1a.15	SjCB(S)1a.15	MISFINKHPNAGWKADKSDRFHSVDDARILLGGRREDPNLREKRRPTVDHHDLNVEIPSHFDSRKKWPRC
	gi|56752787|gb|AAW24605.1|	KSISQIRDQSQCGS**S**WAVSAVGAMSDRICIQSGGKQSVELSAVDLISCCKYCGSGCDGGFLGPSWDYWVL
	unknown [Schistosoma japonicum]	RGIVTGGSKENHTGCRPYPFPKCDHFVKGKYRACGDKLYKTPQCNQTCQKGYNTSYEQDKHYGGFSYNVL
		SVESVIQKDIMMHGPVEAYLEIYEDFLNYKSGIYRYTTGQFISG**H**AVRLIGWGVENGTAYWLAA**N**TWNED
		WGEKGYFRIVRGRNECLIESEIAAGLIKS
S.mansoni CB(S)1a.5	SmCB(S)1a.5	MLISVLCIASLITHLEAHISIKNEKFEPLSHDIISYINKH
	Smp_158420	LDARREESDLRRKRRPIVDHNDWNVEIPSNFDSRKKWPGC
		KSIATIRDQSRCGS**S**WAFGAVEAMSDRSCIQSGGKQNVEL
		SAVDLLSCCEHCGDGFEGGFPALAWDYWVKEGIVTGSSKE
		NHTSCQPYPFPKCEHHTKGKYPACFEEIYKTPNCENTCQK
		SYKTPYAQDKHRGKSRYNVKNDEKAIQKEIMKYGPVEANF
		IVYEDFLNYKSGIYKHITGKLVSW**H**AIRIIGWGVENNTPY
		WLIP**N**SWNEDWGENGNFRILRGRHECSIESEVTAGRINE
incomplete sequence	Smp_085180	MNILNRYLKIDQYVNQNPGFGWKAGINSRFRNLKDTKKLF
		KNNIKIKDLGGKRIQTISHRNVNMVIPHTFDARDHWVNCS
		TIKQIHDECCCRADWVSEKIYNVYADQEDIQKEILMNGPV
		IASILVKVDFLVYKSGVYFPTPKSSNLGWINLRIIGWGYE
		GKTPYWLCANSWSKEWGENGYVKVRRGVQAGYIESYVRAP
		IPKI
incomplete sequence	Smp_085010	MILLLITTIIIVLFVDNENCHLIQQNKTYFHPLSDQLITF
		INKHAFGAVESMSDRICIHSKNKISVELSAINLLSCCTRC
		GFGCRGGIPGMAWDYWKYEGIVTGGSNETHTGCQPYPFPE
		CNHHSSSKSYPPCESYYFPTPECHETCQDDYGKPYKKDKF
		YGKSSYNVASEEISIMKEILLNGPVEGGFYVYEDFLNYKS
		GVYKHITGSYLGGHAIRIIGWGIQQNHIPYWLCANSWNNQ
		WGDQGYFKILRGTNECGIESMVTAGLPNLHK
incomplete sequence	Smp_105370	MSDQHGLYLTSRVISRKYPLLPREHYTELWAVASAASISD
		RTCIQTNGTMKVQLSAIELISCSKNKLGCQIGFSEFSWDY
		WLKNGLVTGDPTGCLPYPFPKCDHRSSNSYPKCGYITYTA
		PPCTKTCRSGYPIPYKADKHYGRVIYSLRPNESDIRKEIM
		MNGPVEAGIFVHSDFLNYKSGVYRHITGQLVTIHSVRIIG
		WGIENDIPYWLCANSWNEDWGLNGYFKILRGSNECEIESF
		VNAGKVDNKT

Table S2. Complete list of schistosome cercarial elastase sequences. All available full-length sequences (*i. e.* those possessing the full catalytic core) cercarial elastase sequence from *S. mansoni*, *S. haematobium* and *S. japonicum* were compiled from GenBank and their respective annotation websites. For each, the name used for this study, previous identifiers, and the full amino acid sequence are listed. Active site residues are highlighted in red. *S. mansoni* sequences are highlighted in pink, *S. haematobium* sequences are highlighted in purple and *S. japonicum* sequences are highlighted in blue.

New Name	Other identifiers	Protein Sequence
S.mansoni CE2a.1	SmCE2a.1	MLNGRTFLMVTLFTYCLTFERVSTWLVRKGEPVQDRTEFPYIAFVRTERTMCTGSLVSTRAVLTAG**H**CVC
	gi|21217531|gb|AAM43941.1|AF510339_1 elastase 2a [Schistosoma mansoni]	SPMPVVQVSFLTLRNGDQQGIHHQPSGVKVAPEYMPSCTASRQRRRIRQTLSGF**D**IATVMLAQMVNLQSG IRVISLPQASDIPTPGTDVFIVGYGRDDNDRDPSRRAGGILKKGRATVMECKHSTTGNPICVQAAYVFGQ ITAPGD**S**GGPLLRSPQGPVLGVVSHGVTLSNRLDVLVEYASVARMLGFVSSNI
S.mansoni CE1a.1	SmCE1a.1	MSNRWRFLVVTLFTYCLTFERVSTWLIRSGEPVQHRTEFPFIAFLTTERTMCTGSLVSTRAVLTAG**H**CVC
	gi|1103829|gb|AAC46967.1| gi|1588494|prf||2208426A elastase [Schistosoma mansoni]	SPLPVIRVLCLFQVSFLTLRNGDQQGIHHQPSGVKVAPGYMPSCMSARRGRPIAQTLSGF**D**IAIVMLAQM VNLQSGITVISLPQASDIPTPGTGVFIVGYGRDDNDRDPSRKNGGILKKGELVVGRATIMECRHATNGNP ICVKAGQNFGQLPAPGD**S**GGPLLPSPQGPVLGVVSHGVTLPNLPDIIVEYASVARMLDFVRSNI
S.haematobium CE1a	ShCE1a	MLNRWRFLVVTLFTYCLTVERVSTWLIRSGEPVQQRTEFPFIAFLTTERTMCTGSLVSTRAVLTAG**H**CVC
	gi|21217535|gb|AAM43943.1|AF510341_1 elastase 1a [Schistosoma haematobium]	SPLPVVRVSFLTLRNGDQQGIHHQPSGVVVAPGYMPSCMSARQRRPIQQTLSGF**D**IAIVSLAQLVNLQTG IRVLSLPQPTDIPRPGTPVFIVGYGRDDNDRDPSRRNGGILKKGRATIMECRHATNGDPICVKSGQNFGQ LPAPGD**S**GGPLLPSPQGPVLGVVSHGVTLPNLPDIIVEYASVARMLDFVRSNI
S.mansoni CE2b	SmCE2b	MLNGWTFLIVTLFTYYLTCQCVSTWLIRSGEPVQQRTEFPFIALLMTDASMCTGSLVSSRAVLTAG**H**CVC
	gi|21217533|gb|AAM43942.1|AF510340_1 gi|256071597|ref|XP_002572126.1| gi|238657278|emb|CAZ28357.1| Smp_006520 elastase 2b [Schistosoma mansoni] cercarial elastase (S01 family) [Schistosoma mansoni]	GQTPIVRVSFLSLSEFDQRTINHRPLEIKVAPEYNPVCQLKRENKRITKSLGGY**D**MAIITLTNLVNLETG VKVISLAAELDIPIPESIAYMVGYGQDIRDPDPSGRYGGILKKGSATIMACRHKTFGDPICVKPGPNSKQ IAGPGD**S**GGPLLLTPQGPIVGVASNGVFLPALADLCVEYSSVPRMLKFILPNI
S.mansoni CE2a.2	SmCE2a.2	MLNGRTFLMVTLFTYCLTFERASTWLVRKGEPVQDRTEFPYIAFIRTERTMCTGSLVSTRAVLTAG**H**CVC
	gi|238657277|emb|CAZ28356.1| gi|256071595|ref|XP_002572125.1| cercarial elastase (S01 family) [Schistosoma mansoni] Smp_006510	SPMPVVQVSFLTLRNGDQQGIHHQPSGVKVAPEYMPSCTASRQRRRIRQTLSGF**D**IATVMLAQMVNLQSG IRVISLPQASDIPTPGTDVFIVGYGRDDNDRDPSRRAGGILKKGRATVMECKHSTTGNPICVQAAYVFGQ ITAPGD**S**GGPLLRSPQGPVLGVVSHGVTLSNRLDVLVEYASVARMLGFVSSNI

S.mansoni CE2a.3	SmCE2a.3	MLNGRTFLMVTLFTYCLTFERASTWLVRKGEPVQDRTEFPYIAFVRTERTMCTGSLVSTRAVLTAG**H**CVC
	gi|227300202|emb|CAY18611.1| gi|256067329|ref|XP_002570606.1| cercarial elastase (S01 family) [Schistosoma mansoni] SmCE_112090	SPMPVVQVSFLTLRNGDQQGIHHQPSGVKVAPEYMPSCTASRQRRRIRQTLSGF**D**IATVMLAQMVNLQSG IRVISLPQASDIPTPGTDVFIVGYGRDDNDRDPSRRAGGILKKGRATVMECKHSTTGNPICVQAAYVFGQ ITAPGD**S**GGPLLRSPQGPVLGVVSHGVTLSNRLDVLVEYASVARMLGFVSSNI
S.mansoni CE1a.2	SmCE1a.2	MSNRWRFLVVTLFTYCLTFERVSTWLIRSGEPVQHRTEFPFIAFLTTERTMCTGSLVSTRAVLTAG**H**CVC
	gi|227280958|emb|CAY19049.1| gi|256048737|ref|XP_002569482.1| cercarial elastase (S01 family) [Schistosoma mansoni] Smp_119130	SPLPVIRVSFLTLRNGDQQGIHHQPSGVKVAPGYMPSCMSARQRRPIAQTLSGF**D**IAIVMLAQMVNLQSG ITVISLPQASDIPTPGTGVFIVGYGRDDNDRDPSRKNGGILKKGRATIMECRHATNGNPICVKAGQNFGQ LPAPGD**S**GGPLLPSPQGPVLGVVSHGVTLPNLPDIIVEYASVARMLDFVRSNI
S.japonicum CE	SjCE	MFSQYILLLVTLLINSLTFQHVSTWLIRSGEPVREQSEFPFIALLMTETSMCTATLISAKAVITAG**H**CVCGTKSINR
	Sjc_0028090 Cercarial protease precursor (EC 3.4.21.-)	IAFLALSDFDHRAVNHDASEIKIPPEYEPTCQLKRENKRVTQSFGGY**D**MAIVLLTKMVNLESGIKVISLPSESDIPM PASIVYIVGYGADVKDPDTSGRYGGILKKGRATVMKCRHRTIGNPICIQPGPDEKQISGPGD**S**GGPLLLSPQGP IIGVASNGLFLPNLGDLCVEYSSVARLLQFILSNI
S.mansoni CE1b	Smp_115980	CTGSLVSTRAVLTAG**H**CVCSPLPVIRVSFLTLRNGDQQGI
		HHQPSGVEVAPGYMPSCMSARQRRPIAQTLSGF**D**IAIVML AQMVNLQSGIRVISLPQPSDIPPPGTGVFIVGYGRDDNDR DPSRKNGGILKKGRATIMECRHATNGNPICVKAGQNFGQL PAPGD**S**GGPLLPSPQGPVLGVVSHGVTLPNLPDIIVEYAS VARMLDFVRSNI
S.mansoniCE1c	gi|1103831|gb|AAC46968.1|	MSNRWRFLVVTLFIYCLTFKRVSTWLIRSDQPVQKHTEFPFIAYIASKKSMCTGSLVSTRVVLTAGHCVC
	or gi|1588495|prf||2208426B elastase	PPMPVIKVQVTFLTLRNGDQQGIHHQPSGVKVAPEYMPSCTASRQRRRIRQTLSGFAIATAMLAQMVNLQ SGIRVIGLPQASDIPTPGTDVFIVGYGRDDNDRDPSRRAGGGRATVTECRHETHVNPICVKAGPNSGQLL GPGDSGGPLLPSPQGPVLGVVSHGVTLSHLPEYVSVARMLNFVRSNI

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Conclusions

The research presented in the proceeding chapters confirms that multiple cercarial elastase isoforms are present in the *Schistosoma mansoni* genome, and that the majority of these isoforms are expressed in the parasite during the transition between the intermediate snail host and the definitive human host. Moreover, cercarial elastase is activated prior to the cercaria's emergence from *Biomphalaria glabrata*, suggesting that the protease may have an additional role in this process. The various SmCE isoforms (of which ten are full length, comprising the full catalytic core of the enzyme) fall into three major groups based on sequence identity: "Group 1," "Group 2a" and "Group 2b." Computational modeling of these isoforms suggests subtle differences in the substrate binding site of the protease, namely in the S4 pocket. However, in experiments where native, purified protease isoforms were run against a universal substrate library, these predicted differences were not visible. Together, the similar expression profiles and similar substrate preferences of the isoforms suggest that their expansion in *S. mansoni* is an example of gene dosage, providing the parasite with ample protease for its critical journey through skin.

In addition, a survey of additional schistosome species suggests that the expansion of the cercarial elastase gene family is unique to *S. mansoni*. The closely related, zoonotic species *S. japonicum* contains only a single isoform of

cercarial elastase in its genome, which previous work suggests is not secreted by cercariae[39,42]. Analysis of the *S. japonicum* genome reveals the expansion of the cathepsin B protease family, and active cathepsin B2 protein is present in *S. japonicum* cercarial secretions. In a proteomic study of *ex vivo* human skin treated with either purified *S. mansoni* cercarial elastase or cathepsin B2, a number of dermal substrates were identified for both proteases. The list of substrates showed significant overlap between the two proteases, and include both structural proteins of skin, like those that form the extracellular matrix, and perivascular skin proteins, including components from the complement cascade. This study was the first to show that cercarial elastase is able to cleave both structural and immune components in the context of intact skin, and further supports cercarial elastase's crucial role in early schistosome invasion and immune evasion. It also gives further insight into the evolution of schistosomes by suggesting that *S. japonicum* and other, non-human schistosomes use a different class of protease (cathepsin B) to facilitate invasion of their respective hosts.

5.2 Future directions

5.2.1 Expression of active, recombinant SmCE

To facilitate future crystallographic and vaccine studies, as well as more detailed substrate specificity analysis, it will be critical to have high yield expression of recombinant SmCE isoforms. Previous and on-going work suggest that *E. coli* expression, using optimized codon sequence, may prove to be the most fruitful

expression system. Currently, bacterial expression of a N-terminal GST or MBP-tagged, catalytically dead SmCE 1a mutant (an active site serine to alanine substitution) gives high yield, soluble expression (*Jessica Ingram and Alberto Rascon, unpublished*). This strategy will likely need to be adapted for expression of active protease, in which case identification of a specific, reversible inhibitor of the enzyme is likely to be useful. Once a robust protocol is developed for expression a single isoform, expression of additional CE isoforms would provide the opportunity for better biochemical and structural analysis of their respective substrate binding sites, and would allow for confirmation of the structural models presented in this thesis. Of particular interest is the SmCE 2b isoform, which cannot be purified from native material and which is the most divergent isoform in terms of amino acid sequence.

5.2.2 Crystallographic studies of cercarial elastase

Kinetic data suggests that purified SmCE has a preference for larger, macromolecular substrates over smaller peptide substrates [27]. This in turn suggests that the process may contain an extended substrate binding site, analogous to those present in some collagenases [75]. Structural data is needed to confirm this; in particular, a crystallographic structure of SmCE in complex with a macromolecular inhibitor, e.g. the bacterial serine protease inhibitor ecotin, would be useful in determining the importance of extended binding site contacts in SmCE. Structural studies could also be used to identify exosites or regulatory

domains of the protease that are not evident by sequencing gazing or computational modeling.

5.2.3 Further functional characterization of secreted proteases in other schistosome species

While data presented here indicates that *S. mansoni* and *S. japonicum* use different proteases to facilitate skin invasion (cercarial elastase and cathepsin B2, respectively), direct functional evidence to confirm that cathepsin B2 is critical to *S. japonicum* invasion is still needed. In order to confirm this, experiments could be performed where a topical inhibitor specific to cathepsin B2 (for instance, CAO74) could be applied to *ex vivo* skin. Skin could subsequently be treated with live *S. japonicum* cercariae, and their progress through skin could be monitored in comparison with cercariae penetrating untreated skin. Another approach, currently being developed in the McKerrow lab, is to use matrigel as a substitute for dermal extracellular matrix. Inhibitors can be directly mixed with matrigel, to which live cercariae can be added; their ability to migrate out of the gel into the surrounding liquid can be assessed quantitatively, and potentially offers a more high-throughput method of assessing migration inhibition.

In this way, using broad class protease inhibitors, various species of schistosomes could be rapidly screened in order to determine the exact nature of their proteolytic secretions, and which proteases are essential for parasite invasion.

5.2.4 Vaccine trials with cercarial elastase

Although an effective drug for the the treatment of schistosomiasis is widely available, a high rate of repeat infection persists. Therefore, vaccine development is a high priority. SmCE is among the first proteins that the host comes in contact with upon infection, and is critical for successful migration through skin. Although the protease is not innately immunogenic [76], it is therefore likely that a host immune response mounted against SmCE--in particular, a neutralizing response that diminishes the activity of the enzyme--could greatly slow down the transit of cercariae through skin, with the prediction that slowed migration would lead to increased parasite death in skin. Increased parasite death would result in the release of additional parasite antigens that could promote an adaptive immune response by the host, in a process that could resemble the cercarial dermatitis induced by infection of human skin with non-human schistosome species.

Availability of recombinant enzyme of high purity will be critical for these studies; however, proof-of-principle studies could be initiated with native protease. This could include identification and development of a neutralizing antibody against the protease by *in vitro* methods, i.e. phage display and antibody engineering. An identified antibody could then be delivered passively to a mice to see if it offers a protective effect against invading *S. mansoni* cercariae. Identification of such an antibody, and its binding epitope on SmCE, would further support more directed SmCE vaccination studies in animals.

REFERENCES

1. King, C.H., Dickman, K. & Tisch, D.J. Reassessment of the cost of chronic helmintic infection: a meta-analysis of disability-related outcomes in endemic schistosomiasis. *Lancet* **365**, 1561–1569 (2005).
2. Chitsulo, L., Engels, D., Montresor, A. & Savioli, L. The global status of schistosomiasis and its control. *Acta Trop.* **77**, 41–51 (2000).
3. Brown, M. et al. Schistosoma mansoni, nematode infections, and progression to active tuberculosis among HIV-1-infected Ugandans. *Am. J. Trop. Med. Hyg.* **74**, 819–825 (2006).
4. Booth, M. et al. Micro-geographical variation in exposure to Schistosoma mansoni and malaria, and exacerbation of splenomegaly in Kenyan school-aged children. *BMC Infect. Dis.* **4**, 13 (2004).
5. Hotez, P.J., Bethony, J.M., Diemert, D.J., Pearson, M. & Loukas, A. Developing vaccines to combat hookworm infection and intestinal schistosomiasis. *Nat. Rev. Microbiol.* **8**, 814–826 (2010).
6. DiConza, J.J. & Hansen, E.L. Multiplication of transplanted Schistosoma mansoni daughter sporocysts. *J. Parasitol.* **58**, 181–182 (1972).
7. Yoshino, T.P., Lodes, M.J., Rege, A.A. & Chappell, C.L. Proteinase activity in miracidia, transformation excretory-secretory products, and primary sporocysts of Schistosoma mansoni. *J. Parasitol.* **79**, 23–31 (1993).
8. Haeberlein, S. & Haas, W. Chemical attractants of human skin for swimming Schistosoma mansoni cercariae. *Parasitol Res* **102**, 657–662 (2007).
9. Brachs, S. & Haas, W. Swimming behaviour of Schistosoma mansoni cercariae: responses to irradiance changes and skin attractants. *Parasitol Res* **102**, 685–690 (2007).
10. Haas, W. et al. Recognition and invasion of human skin by Schistosoma mansoni cercariae: the key-role of L-arginine. *Parasitology* **124**, 153–167 (2002).
11. Wang, Y.S. et al. [A study on mechanism of skin penetration by cercaria of S. japonicum and screening of preventive drugs]. *Hua Xi Yi Ke Da Xue Xue Bao* **20**, 18–20 (1989).
12. Stirewalt, M.A. & Dorsey, C.H. Schistosoma manonsi: cercarial penetration of host epidermis at the ultrastructural level. *Exp. Parasitol.* **35**, 1–15 (1974).
13. Elias, P.M. Stratum corneum architecture, metabolic activity and interactivity with subjacent cell layers. *Exp. Dermatol.* **5**, 191–201 (1996).
14. Fukuyama, K., Tzeng, S. & McKerrow, J. The epidermal barrier to Schistosoma mansoni infection. *Current problems in …* (1983).
15. Dorsey, C.H., Cousin, C.E., Lewis, F.A. & Stirewalt, M.A. Ultrastructure of the Schistosoma mansoni cercaria. *Micron* **33**, 279–323 (2002).
16. GORDON, R.M. & GRIFFITHS, R.B. Observations on the means by which the cercariae of Schistosoma mansoni penetrate mammalian skin, together with an account of certain morphological changes

observed in the newly penetrated larvae. *Ann Trop Med Parasitol* **45**, 227–243 (1951).
17. Crabtree, J.E. & Wilson, R.A. Schistosoma mansoni: an ultrastructural examination of skin migration in the hamster cheek pouch. *Parasitology* **91 (Pt 1)**, 111–120 (1985).
18. Keene, W.E., Jeong, K.H., McKerrow, J.H. & Werb, Z. Degradation of extracellular matrix by larvae of Schistosoma mansoni. II. Degradation by newly transformed and developing schistosomula. *Lab. Invest.* **49**, 201–207 (1983).
19. Cort, W.W. *The cercaria of the Japanese blood fluke, Schistosoma japonicum Katsurada, (University of California publications in zoology)*. 507 (University of California press: 1919).
20. Davis, D. Report on the preparation of an histolytic ferment present in the bodies of cercariae. *J. Parasitol.* (1936).
21. Stirewalt, M. CHEMICAL BIOLOGY OF SECRETIONS OF LARVAL HELMINTHS*. *Annals of the New York Academy of Sciences* (1963).
22. Gazzinelli, G., Ramalho-Pinto, F.J. & Pellegrino, J. Purification and characterization of the proteolytic enzyme complex of cercarial extract. *Comp. Biochem. Physiol.* **18**, 689–700 (1966).
23. Dresden, M.H. & Asch, H.L. Proteolytic enzymes in extracts of Schistosoma mansoni cercariae. *Biochim. Biophys. Acta* **289**, 378–384 (1972).
24. Campbell, D.L., Frappaolo, P.J., Stirewalt, M.A. & Dresden, M.H. Schistosoma mansoni: partial characterization of enzyme(s) secreted from the preacetabular glands of cercariae. *Exp. Parasitol.* **40**, 33–40 (1976).
25. Landsperger, W.J., Stirewalt, M.A. & Dresden, M.H. Purification and properties of a proteolytic enzyme from the cercariae of the human trematode parasite Schistosoma mansoni. *Biochem. J.* **201**, 137–144 (1982).
26. Newport, G.R. et al. Cloning of the proteinase that facilitates infection by schistosome parasites. *J. Biol. Chem.* **263**, 13179–13184 (1988).
27. McKerrow, J., Pino-Heiss, S. & Lindquist, R. Purification and characterization of an elastinolytic proteinase secreted by cercariae of Schistosoma mansoni. *Journal of Biological …* (1985).
28. Lim, K.C. et al. Blockage of skin invasion by schistosome cercariae by serine protease inhibitors. *Am. J. Trop. Med. Hyg.* **60**, 487–492 (1999).
29. Knudsen, G.M., Medzihradszky, K.F., Lim, K.-C., Hansell, E. & McKerrow, J.H. Proteomic analysis of Schistosoma mansoni cercarial secretions. *Mol. Cell Proteomics* **4**, 1862–1875 (2005).
30. Curwen, R.S., Ashton, P.D., Sundaralingam, S. & Wilson, R.A. Identification of novel proteases and immunomodulators in the secretions of schistosome cercariae that facilitate host entry. *Mol. Cell Proteomics* **5**, 835–844 (2006).

31. Fishelson, Z. et al. Schistosoma mansoni: cell-specific expression and secretion of a serine protease during development of cercariae. *Exp. Parasitol.* **75**, 87–98 (1992).
32. Cohen, F.E. et al. Arresting tissue invasion of a parasite by protease inhibitors chosen with the aid of computer modeling. *Biochemistry* **30**, 11221–11229 (1991).
33. McKerrow, J.H., Keene, W.E., Jeong, K.H. & Werb, Z. Degradation of extracellular matrix by larvae of Schistosoma mansoni. I. Degradation by cercariae as a model for initial parasite invasion of host. *Lab. Invest.* **49**, 195–200 (1983).
34. Salter, J.P. et al. Cercarial elastase is encoded by a functionally conserved gene family across multiple species of schistosomes. *J. Biol. Chem.* **277**, 24618–24624 (2002).
35. Pleass, R.J., Kusel, J.R. & Woof, J.M. Cleavage of human IgE mediated by Schistosoma mansoni. *Int. Arch. Allergy Immunol.* **121**, 194–204 (2000).
36. Aslam, A. et al. Proteases from Schistosoma mansoni cercariae cleave IgE at solvent exposed interdomain regions. *Mol. Immunol.* **45**, 567–574 (2008).
37. Hansell, E. et al. Proteomic analysis of skin invasion by blood fluke larvae. *PLoS Negl Trop Dis* **2**, e262 (2008).
38. Berriman, M. et al. The genome of the blood fluke Schistosoma mansoni. *Nature* **460**, 352–358 (2009).
39. Dvorák, J. et al. Differential use of protease families for invasion by schistosome cercariae. *Biochimie* **90**, 345–358 (2008).
40. Dolecková, K. et al. The functional expression and characterisation of a cysteine peptidase from the invasive stage of the neuropathogenic schistosome Trichobilharzia regenti. *Int. J. Parasitol.* **39**, 201–211 (2009).
41. McKerrow, J.H. & Salter, J. Invasion of skin by Schistosoma cercariae. *Trends Parasitol.* **18**, 193–195 (2002).
42. Schistosoma japonicum Genome Sequencing and Functional Analysis Consortium The Schistosoma japonicum genome reveals features of host-parasite interplay. *Nature* **460**, 345–351 (2009).
43. Ingram, J. et al. Proteomic Analysis of Human Skin Treated with Larval Schistosome Peptidases Reveals Distinct Invasion Strategies among Species of Blood Flukes. *PLoS Negl Trop Dis* **5**, e1337 (2011).
44. Abdulla, M.-H. et al. Drug Discovery for Schistosomiasis: Hit and Lead Compounds Identified in a Library of Known Drugs by Medium-Throughput Phenotypic Screening. *PLoS Negl Trop Dis* **3**, e478 (2009).
45. Salter, J.P., Lim, K.C., Hansell, E., Hsieh, I. & McKerrow, J.H. Schistosome invasion of human skin and degradation of dermal elastin are mediated by a single serine protease. *J. Biol. Chem.* **275**, 38667–38673 (2000).

46. Brown, C.M. et al. Peptide length and leaving-group sterics influence potency of peptide phosphonate protease inhibitors. *Chem. Biol.* **18**, 48–57 (2011).
47. Li, C.Y., Lam, K.W. & Yam, L.T. Esterases in human leukocytes. *J. Histochem. Cytochem.* **21**, 1–12 (1973).
48. Larkin, M.A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
49. HACKEY, J.R. & Stirewalt, M.A. Penetration of host skin by cercariae of Schistosoma mansoni. I. Observed entry into skin of mouse, hamster, rat, monkey and man. *J. Parasitol.* **42**, 565–580 (1956).
50. Remington, S.J., Woodbury, R.G., Reynolds, R.A., Matthews, B.W. & Neurath, H. The structure of rat mast cell protease II at 1.9-A resolution. *Biochemistry* **27**, 8097–8105 (1988).
51. Sali, A., Matsumoto, R., McNeil, H.P., Karplus, M. & Stevens, R.L. Three-dimensional models of four mouse mast cell chymases. Identification of proteoglycan binding regions and protease-specific antigenic epitopes. *J. Biol. Chem.* **268**, 9023–9034 (1993).
52. Perona, J.J. & Craik, C.S. Structural basis of substrate specificity in the serine proteases. *Protein Sci.* **4**, 337–360 (1995).
53. Lynch, V.J. Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. *BMC Evol Biol* **7**, 2 (2007).
54. Rogers, M.B. et al. Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania. *Genome Research* (2011).doi:10.1101/gr.122945.111
55. Jackson, A.P. Tandem gene arrays in Trypanosoma brucei: Comparative phylogenomic analysis of duplicate sequence variation. *BMC Evol Biol* **7**, 54 (2007).
56. ROBINSON, M., DALTON, J. & DONNELLY, S. Helminth pathogen cathepsin proteases: it's a family affair. *Trends in Biochemical Sciences* **33**, 601–608 (2008).
57. Mello, L.V., O'Meara, H., Rigden, D.J. & Paterson, S. Identification of novel aspartic proteases from Strongyloides ratti and characterisation of their evolutionary relationships, stage-specific expression and molecular structure. *BMC Genomics* **10**, 611 (2009).
58. Gobert, G.N. et al. Transcriptional Changes in Schistosoma mansoni during Early Schistosomula Development and in the Presence of Erythrocytes. *PLoS Negl Trop Dis* **4**, e600 (2010).
59. Mortz, E., Krogh, T.N., Vorum, H. & Görg, A. Improved silver staining protocols for high sensitivity protein identification using matrix-assisted laser desorption/ionization-time of flight analysis. *Proteomics* **1**, 1359–1363 (2001).
60. Mair, G.R., Maule, A.G., Fried, B., Day, T.A. & Halton, D.W. Organization of the musculature of schistosome cercariae. *J. Parasitol.* **89**, 623–625 (2003).

61. Stefanić, S. et al. RNA interference in Schistosoma mansoni schistosomula: selectivity, sensitivity and operation for larger-scale screening. *PLoS Negl Trop Dis* **4**, e850 (2010).
62. Caffrey, C.R. et al. SmCB2, a novel tegumental cathepsin B from adult Schistosoma mansoni. *Molecular & Biochemical Parasitology* **121**, 49–61 (2002).
63. Chalkley, R.J. et al. Comprehensive analysis of a multidimensional liquid chromatography mass spectrometry dataset acquired on a quadrupole selecting, quadrupole collision cell, time-of-flight mass spectrometer: II. New developments in Protein Prospector allow for reliable and comprehensive automatic analysis of large datasets. *Mol. Cell Proteomics* **4**, 1194–1204 (2005).
64. Elias, J.E. & Gygi, S.P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods* **4**, 207–214 (2007).
65. Lunstrum, G.P., Morris, N.P., McDonough, A.M., Keene, D.R. & Burgeson, R.E. Identification and partial characterization of two type XII-like collagen molecules. *J Cell Biol* **113**, 963–969 (1991).
66. Choe, Y. et al. Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. *J. Biol. Chem.* **281**, 12824–12832 (2006).
67. Castro-Borges, W., Dowle, A., Curwen, R.S., Thomas-Oates, J. & Wilson, R.A. Enzymatic shaving of the tegument surface of live schistosomes for proteomic analysis: a rational approach to select vaccine candidates. *PLoS Negl Trop Dis* **5**, e993 (2011).
68. Sloane, B.F. & Honn, K.V. Cysteine proteinases and metastasis. *Cancer Metastasis Rev.* **3**, 249–263 (1984).
69. Cavallo-Medved, D. et al. Live-cell imaging demonstrates extracellular matrix degradation in association with active cathepsin B in caveolae of endothelial cells during tube formation. *Exp. Cell Res.* **315**, 1234–1246 (2009).
70. Werle, B., Ebert, W., Klein, W. & Spiess, E. Cathepsin B in tumors, normal tissue and isolated cells from the human lung. *Anticancer Res.* **14**, 1169–1176 (1994).
71. He, Y.-X., Salafsky, B. & Ramaswamy, K. Comparison of skin invasion among three major species of Schistosoma. *Trends Parasitol.* **21**, 201–203 (2005).
72. Jenkins, S.J., Hewitson, J.P., Jenkins, G.R. & Mountford, A.P. Modulation of the host's immune response by schistosome larvae. *Parasite Immunol.* **27**, 385–393 (2005).
73. Kikuchi, Y. et al. Crucial commitment of proteolytic activity of a purified recombinant major house dust mite allergen Der p1 to sensitization toward IgE and IgG responses. *J. Immunol.* **177**, 1609–1617 (2006).
74. Ruppel, A., McLaren, D.J., Diesfeld, H.J. & Rother, U. Schistosoma mansoni: escape from complement-mediated parasiticidal

mechanisms following percutaneous primary infection. *Eur. J. Immunol.* **14**, 702–708 (1984).
75. Perona, J.J., Tsu, C.A., Craik, C.S. & Fletterick, R.J. Crystal structure of an ecotin-collagenase complex suggests a model for recognition and cleavage of the collagen triple helix. *Biochemistry* **36**, 5381–5392 (1997).
76. Ramzy, R.M. et al. Evaluation of a stage-specific proteolytic enzyme of Schistosoma mansoni as a marker of exposure. *Am. J. Trop. Med. Hyg.* **56**, 668–673 (1997).

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.

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.

Jessica Ingram
Author Signature

12/14/2011
Date