Title
Characterization of the protease that facilitates schistosome infection

Author
Salter, Jason Patrick

Publication Date
2001

Peer reviewed|Thesis/dissertation

CHARACTERIZATION OF THE PROTEASE THAT FACILITATES SCHISTOSOME INFECTION

by

Jason Patrick Salter

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA SAN FRANCISCO

Date University Librarian

Degree Conferred: .....

Dedication

This work is dedicated to my Family—my biological family, my friends, and the people who will become so in the future.

Thank you.

Jason

Acknowledgments

The text of this dissertation is in part (Chapter 1) a reprint of the material as it appears in the Journal of Biological Chemistry, 275(49) :38667. Copyright by Journal of Biological Chemistry. Permission granted for reproduction.

CHARACTERIZATION OF THE PROTEASE THAT FACILITATES SCHISTOSOME INFECTION

Jason P. Salter

Abstract

Aquatic larvae (cercariae) of the trematode parasite, *Schistosoma mansoni*, rapidly penetrate human skin by disrupting the epidermal cell layer and degrading host basement membrane and extracellular matrix proteins. Previous work from several laboratories identified two serine proteases potentially involved in cercarial invasion, one "chymotrypsin-like" (cercarial elastase) and the second "trypsin-like". To evaluate the relative roles of these two proteases in larval invasion both were purified, and biochemically characterized. Invasion inhibition assays using selective inhibitors confirmed that cercarial elastase is the enzyme involved in skin penetration. Its ability to degrade skin elastin was confirmed and the three sites of cleavage within elastin help define a new family of elastases.

Genes coding for cercarial elastase isoforms were identified and sequenced. They can be divided into two classes by amino acid and promoter sequence homology. Two of the five genes identified in *Schistosoma mansoni* account for over 90 percent of the enzyme released. Positional scanning synthetic combinatorial libraries demonstrate that the two major isoforms are isoenzymes.

Previous work suggested that cercarial elastase may have extended binding sites on both the amino terminal and carboxy terminal side of peptide bond cleavage. To test this hypothesis, positional scanning synthetic combinatorial libraries and a substrate phage library were used to map selectivity from P4 to P2'. A database of human proteins was then screened for potential substrates that contained the optimal substrate sequence. A number of epidermal extracellular matrix, and basement membrane proteins found in

skin contained this consensus sequence. To confirm these database predictions of biological substrates, potential substrates were digested with pure cercarial elastase *in vitro* and skin sections incubated with pure cercarial elastase.

Cercarial elastase has previously been shown to be immunogenic in humans. In collaboration with Dr. Reda Ramzy (Ain Shams University, Cairo, Egypt) an ELISA was developed using purified native cercarial elastase as antigen. Results of testing the ELISA in endemic villages indicated there is a positive correlation between disease incidence and cercarial elastase antibodies. This suggests that the ELISA will be a useful epidemiological tool to identify exposure to schistosome larvae.

Table of Contents

List of Tables

List of Figures

CHARACTERIZATION OF THE PROTEASE THAT FACILITATES SCHISTOSOME INFECTION

Introduction

I.1. Lifecycle of the schistosome

Schistosomiasis is a disease affecting over 250,000,000 people worldwide. The infectious agent is a family of blood flukes classified as *Trematoda; Digenea; Strigeidida; Schistosomatidae*; *Schistosoma*. The complete lifecycle requires six lifecycle stages and two separate hosts (Figure I.1.). Infection occurs when an aquatic larval form, the cercaria(ae), leaves its secondary host, a snail, and locates the exposed skin of its vertebrate host. The cercaria then penetrates the host's skin reaching the vascular system sometimes as rapidly as five to 10 minutes but often after hours. This is facilitated by the use of highly specialized mechanical and enzymatic factors. Following entry into the skin the cercariae sheds its tail and a protective carbohydrate surface layer called the glycocalyx. After this transformation the larval stage is referred to as a schistosomule or schistosomulum(a). The schistosomule continues its migration through host tissue until a venule or lymphatic vessel is entered. Following the flow of blood or lymph the schistosomule migrates towards the heart, through the lungs, back through the heart, and into the aorta. Depending on the species of schistosome, the schistosomule will migrate either to the hepatic portal vein or to the veins of the bladder. After reaching its final destination the schistosomule beings a period of rapid growth transforming into an adult schistosome worm.

Each cercaria exiting the snail is already committed to develop into a male or female worm. When both sexes are present as developing adult worms they will pair with the larger male worm enclosing the female in a groove called the gynecophoral canal. When growth is complete and pairing is successful the females begin egg laying. Each

egg is comprised of a strong shell that protects a multicellular ciliated life cycle stage called the miracidum(a). Eggs are released individually from the female and travel through the intestine or bladder wall exiting into the lumen of the gut or bladder respectively. Eggs then pass out of the host and into fresh water. A hypotonic environment signals the eggs to "hatch". Rupture of the protective shell releases the miracidum which initiates a search for a secondary host snail .

Upon contact with a snail host the miracidum will penetrate its soft tissue, shed its ciliated plates, and take up residence in the snail tissue near the site of entry. This transformation step completes the miracidum lifecycle stage and the parasite stage is now defined as a mother sporocyst. Germinal cells within the mother sporocyst then develop into daughter sporocysts. Daughters will bud off from the mother and migrate to the snail's hepatopancreas. When the daughter sporocysts reach their destination cercariae begin to develop internally. Daughter sporocysts also expand their numbers through asexual replication until limited by the size of the snail host [1]. When cercarial mature they exit the daughter sporocyst through a birth pore and enter snail tissue. They will then migrate out of the snail and repeat the life cycle (Figure I.1.).

I.2. Disease characteristics

I.2.1. Species differences

There are three predominant and two less common species of schistosomes that infect humans: *S. mansoni*, *S. japonicum*, *S. haematobium* (most common), *S. mekongi*, and *S. intercalatum* (less common). Each species differs subtly in morphology, behavior, and disease characteristics. There are two locations within the body that schistosomes will establish a permanent residence depending on the species. *S. mansoni*, *S. intercalatum*, *S. japonicum*, and *S. mekongi* reside in the portal venous system. *S. haematobium* localizes to the venous system of the bladder. Both locations position the

adult worms adjacent to an excretory system that permit passage of parasite eggs out of the host.

I.2.2. Chronic disease

Paired male and female worms lay eggs that traverse the vessel wall, enter the excretory system, and exit the host. Eggs that fail to complete this process become trapped in host tissue. The eggs initiate a granulomatous response. Granuloma formation is cumulative and proportional to parasite fecundity. This eventually leads to the disruption of normal organ function. Eggs can also travel via the vascular system to distant organs. *S. haematobium* eggs are frequently found in genital tissues of infected patients. This causes significant morbidity but does not affect fertility. Portal venous hypertension resulting from *S. mansoni* infection may lead to splenomegaly and ascites. Additionally, renal lesions from deposits of host immunoglobulin complexed with schistosome antigens are occasionally found. The pathogenesis of schistosomiasis is therefore caused primarily by the immune response to parasite's eggs. This situation results in low mortality but high morbidity that is cumulative and irreversible.

I.2.3. Cercarial dermatitis

A dermatological reaction initiated by cercarial invasion is common in areas where avian or small mammal schistosomes are present. The response rapidly progresses within minutes of infection and is characterized by itching, erythema, and/or papular eruption. Prior exposure to cercarial infection is require for priming of the immune system to produce the response. The rash normally resolves itself in 24 to 72 hours. Schistosomes that infect humans rarely produce the reaction.

I.2.4. Acute schistosomiasis

Acute schistosomiasis is a response to worm development and egg laying. Individuals who have no prior immunity and are experiencing a primary infection are most likely to experience a continuing fever, rigors, sweating, myalgia, and headache. This acute phase of the disease is most commonly reported with infections by *S. japonicum*. Non-immune individuals are most likely to have the strongest response. Symptoms last from 1-30 days with an average duration of 8 days.

I.2.5. Geographic distribution

Geographical distributions of each species are unique but can overlap (Table I.1.). The distribution of schistosomes throughout the world mirrors the presence of a compatible secondary snail host. Snail geographic distributions are more restricted than primary host distributions and therefore are the limit of disease range. Disease transmission requires both hosts to be present.

I.3. Characteristics of cercariae

Primary infection is initiated by penetration of the cercaria through host skin. Cercariae are "heat seeking missiles" that can follow thermal gradients to locate their host [2]. Detection of lipids on host skin signals the parasite to begin penetration. The parasite will halt swimming and initiate gland secretion and probing movements [3].

Cercariae are approximately 150 microns long and 70 microns wide. They are completely covered in sharp spines directed posteriorly. The internal musculature consists of two sets of concentric rings set 90 degrees to each other allowing alternating contractions and extensions of the body. Using a combination of mechanical and enzymatic forces, a cercaria can penetrate under the stratum corneum in 3-7 minutes [4]. Further penetration and entry into a vessel can occur within 5 to 24 hours. To complete this journey, the parasite must penetrate all tissue structures of the skin. These layers

include an epithelial layer, a basal lamina, connective tissue, another basal lamina, and finally an endothelial layer. Additionally, during penetration the parasite must avoid or resist the host's immune system.

Once the cercaria arrives at the dermal/epidermal basement membrane a significant delay is observed [5]. This is thought to result from difficulties in degrading this complex structure. Cercarial extracts and cercarial secretions have been shown to cleave many of the proteins that constitute the connective tissue of human skin.

Cercarial internal organization is dominated by three types of glands; pre-acetabular, post-acetabular, and head glands (Figure I.2). The pre and post acetabular glands represent approximately 40% of the internal volume of the cercariae and release their contents during invasion in a pre-programmed sequence (Figure I.3.). The pre-acetabular glands are used in invasion when the cercariae are localized in the epidermis and at its basement membrane. The post-acetabular glands releases their contents primarily on the surface of skin immediately upon contact to provide an adhesive surface for penetration [6]. The head gland secretes its material late in the penetration process when cercariae have reached the basement membrane of a vessel wall [7].

All three glands have been shown to contain proteases. The pre and post acetabular glands have been shown to contain cercarial elastase, a serine protease [8]. A cysteine protease has been localized to the head gland but has not been identified in cercarial secretions induced by skin contact [9]. Other enzymes have been speculated to exist but have yet to be cloned and or definitively demonstrated to be present in the acetabular glands [10, 11].

The majority of acetabular contents are released within 12 hours of penetration [4]. Within this time frame cercariae are predominantly localized to the epidermis and its basement membrane [12]. During invasion, cercariae disrupt cell-cell contacts between squamous cells of the epidermis and often damage nearby squamous cells themselves. These observations suggest that desmosomes are one of the targets for proteolysis.

Within 30 minutes of cercarial exposure edema is also noted at the site of penetration. This edema formation is correlated with enhanced infectivity of cercariae and is associated with protease activity and cercarial secretions [13, 14].

Given the number of species of schistosomes, the conservation in mechanical and glandular structures between cercariae of different species suggests that the mechanism of invasion is highly adapted and conserved.

I.4. Proteases released from cercariae

The acetabular glands were noted in the earliest observations of the cercaria and surmised to be involved in invasion. In 1919 Lutz reported that the cercarial glands were absent after penetration and in 1921 Cort specifically stated that the glands functioned for dissolving tissue [15, 16]. In 1931 a complete and accurate description of the acetabular glands (2 pairs of pre-acetabular glands and 3 pairs of post-acetabular glands) was published [17]. The earliest test of the "tissue dissolving" hypothesis was reported in 1936. Frog tissue was used as a substrate and homogenized cercariae from *D. flexicaudum* as enzyme extract [18]. This early work set the stage for over 80 years of research involving cercariae and the molecular mechanisms of penetration.

Gorden et. al. were the first to digest human skin sections with enzyme extracts made from homogenized snail "livers" infected with *S. mansoni* [6]. Lewert et. al. were the first to test protease inhibitors and report the results of a skin inhibition assay [19]. But, these first studies used cercariae or infected snail hepatopancreas homogenates. Stirewalt et. al. led the way to enzyme purification with a novel method for secretion collection using the discovery that human sebum or Vaseline could induce pre-acetabular secretion [20]. This technique allowed the collection of large quantities of cercarial secretions for enzyme characterization and further purification.

Electron microscopy studies of cercarial invasion were first reported in 1970 by Bruce et. al. [12]. This was followed by further examination of the individual acetabular

glands, their development, the elemental composition of their contents, and additional studies of the invasion process [7, 21-26].

The first attempts at enzyme purification began with the use of gel filtration columns [27]. Additional techniques to further increase purity included iso-electric focusing and ion exchange [28-30]. In 1985, McKerrow et. al. successfully developed a protocol to purify acetabular glands protease activity to homogeneity [31]. Three years later the cDNA of this enzyme was cloned and named cercarial elastase [32]. Two additional groups subsequently cloned cercarial elastase using genomic PCR, RT-PCR, and a genomic phage library [33-35]. The combination of cloning cercarial elastase and its purification enabled the discovery of specific inhibitors [36]. Using these inhibitors in a skin inhibition assay demonstrated that cercarial elastase indeed was necessary for efficient host invasion [36-38].

Cercarial secretions are composed of pre-acetabular and post-acetabular "gland" (cell) contents. Prior work had identified and cloned a single protease from these secretions: *S. mansoni* cercarial elastase (SmCE) [31, 32]. Additional proteases with unique activities within these secretions have been reported without further purification or cloning [10, 11]. Early reports of fractionated cercarial secretions also demonstrated multiple peaks of activity [28-30]. The use of macromolecular substrates and whole cercarial extracts made it difficult to determine if peaks of activity were the result of unique proteases, multimers, or contamination. Further work was necessary to resolve questions about enzyme numbers, location, and function.

I.5. Specific aims of this thesis designed to address questions about the character and function of cercarial proteases

I.5.1. How many proteases are present in cercarial secretions?

Protease identification will be accomplished by screening chromatographic fractions of cercarial secretions with a variety of small molecular peptide substrates. Each new activity will be purified and its enzymatic properties characterized. N-terminal amino acid sequence data will be used to clone each gene by RT-PCR using degenerate primers. cDNA clones will then be used as probes for Southern blot analysis to determine copy number and confirm enzyme source.

I.5.2. What role do these proteases play in the invasion process?

Specific inhibitors will be identified for each novel enzyme. These inhibitors will be tested in a skin assay to determine their relative role in the invasion process.

I.5.3. What is the stage specificity and localization of the proteases identified?

The cloned enzymes will be expressed in *E. coli* and purified using a histidine tag to provide sufficient material for the production of polyclonal rabbit antiserum. Antiserum will be used for immunohistochemistry and Western blots to determine location and stage specificity of expression for each enzyme.

I.5.4. Transcriptional regulation of cercarial elastase is stage and cell specific. How are the cercarial elastase promoters structured?

To further our understanding of schistosome transcriptional control the non-coding upstream genomic DNA was isolated for multiple genes. These putative promoter regions provided both an opportunity for comparison between different protease enzymes by alignment and provided promoter regions for expression cassettes to be used in expression and transgenic studies.

I.5.5. Cercarial elastase is more effective at degrading macromolecular natural substrates than small synthetic ones. What is the extended substrate selectivity for cercarial elastase?

The extended binding of cercarial elastase will be determined from P4 through P2'. Previous work hinted that cercarial elastase may have an extended binding site on both the prime and non-prime side [39, 40]. A combination of combinatorial synthetic chemical libraries and substrate phage libraries will be used to map the prime and nonprime side selectivity.

I.5.6. Can binding site specificity data be used to identify natural targets?

Sufficient selectivity at more than four positions will allow protein databases to be screened for matches to the consensus sequence. Proteins with one or more consensus sequence will be further screened for appropriate tissue localization and function. Predictions from database screening will be confirmed by direct *in vitro* assays and microscopic analysis of skin sections.

I.5.7. Where is the site of elastin cleavage ?

Fragments of digested native elastin will be purified and N-terminally amino acid sequenced. Sequence data will be compared to the predicted sequence of elastin to identify the site of cleavage.

I.5.8. Cercarial elastase is one of the first proteins released into the host. Can cercarial elastase be used in an ELISA assay to identify individuals with recent exposure to schistosomes?

Native enzyme was used as an antigen to determine exposure of Egyptian school children to schistosomes. Recombinant cercarial elastase will also be tested in these ELISA essays as a functional substitute.

I.6. Figures and tables for introduction

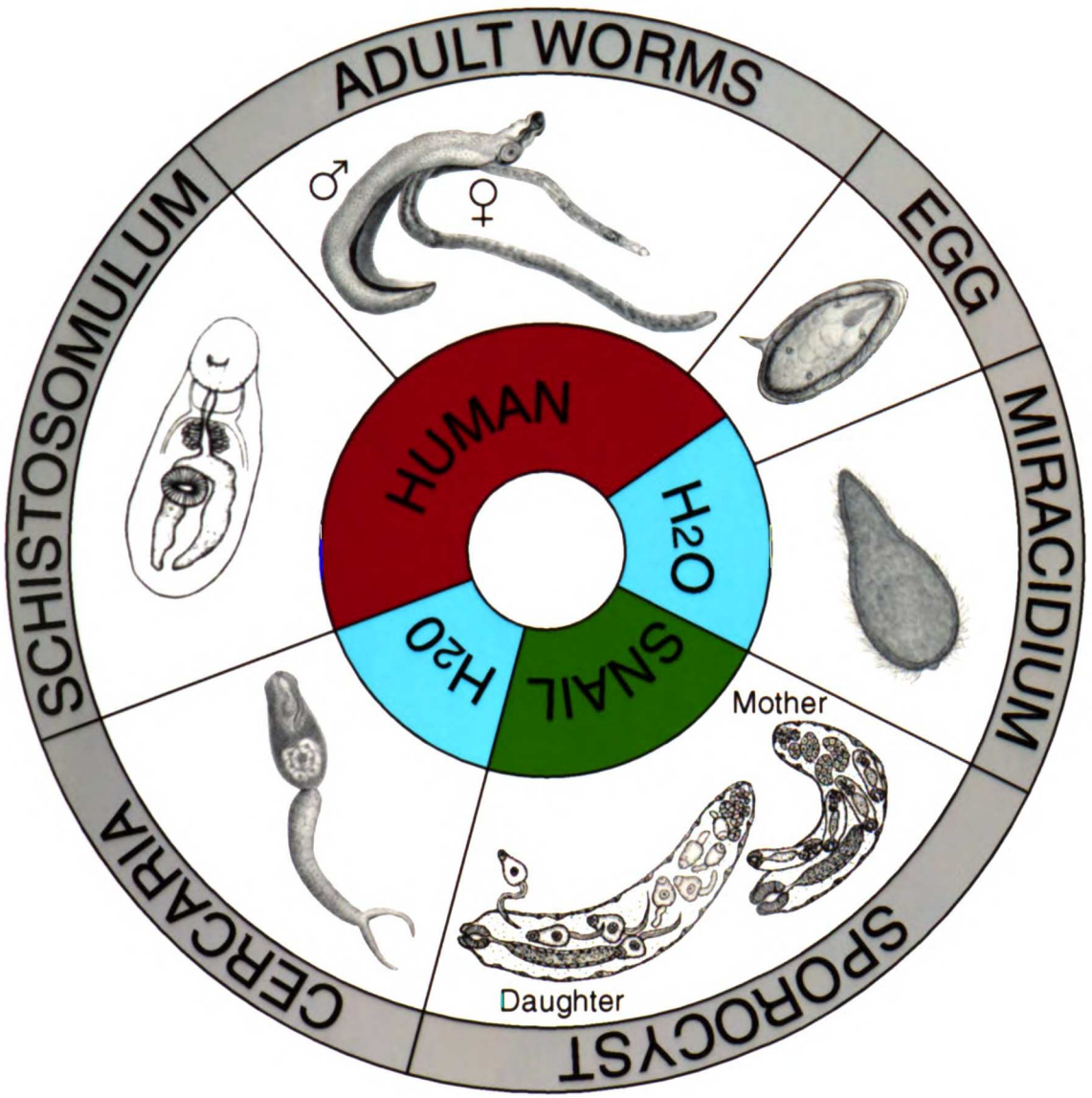


Figure I.1 Life cycle of the schistosome

Lifecycle advances in the clockwise direction. Inner ring denotes the environment of the parasite. Outer ring denotes the life cycle stage name. Images in middle ring are not proportionally sized.

Images from Cort, 1921 [16].

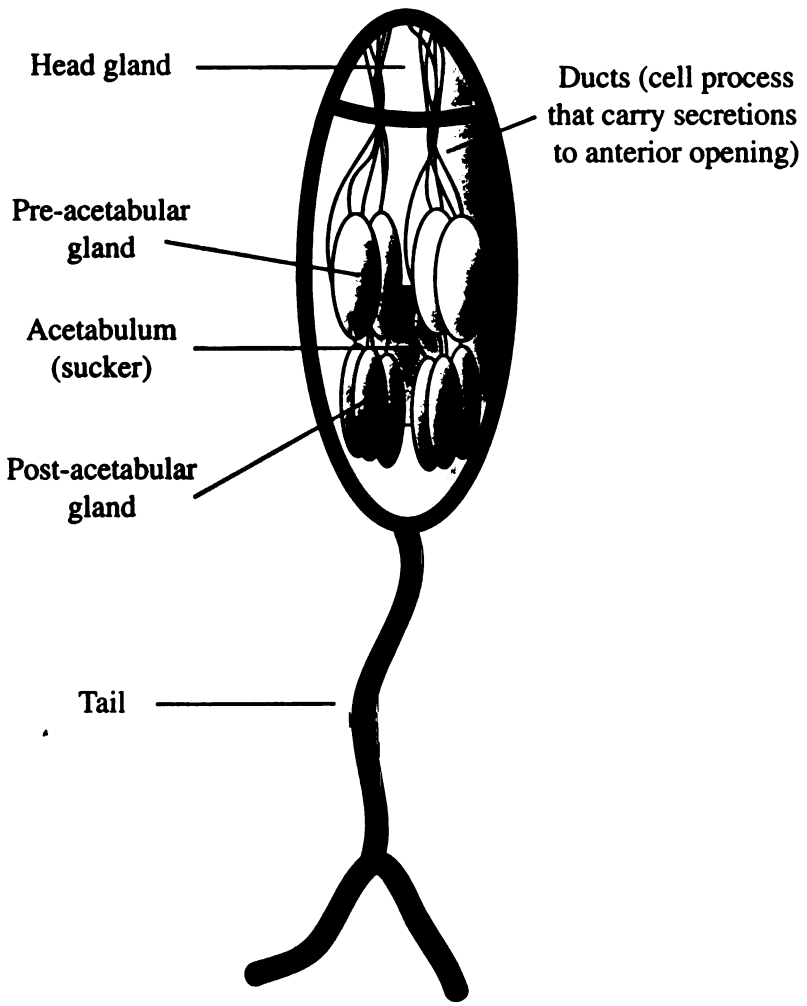


Fig I.2. Schematic of cercaria structures

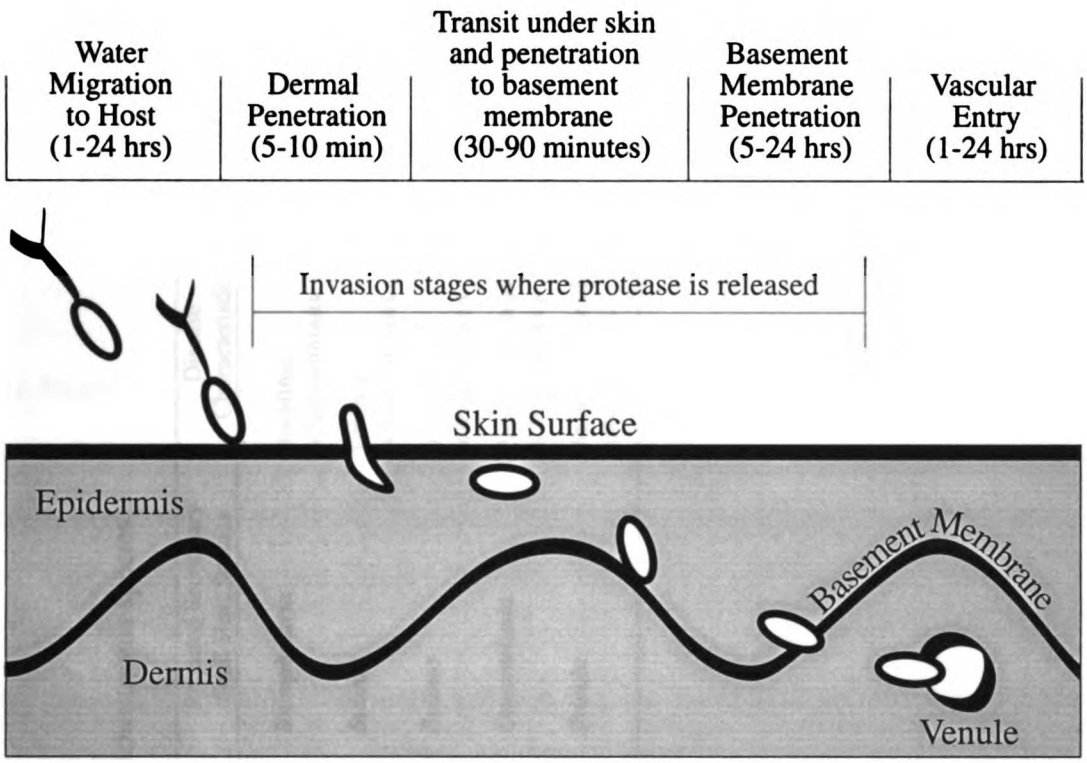


Figure I.3. Five major steps of parasite entry into host skin

Water born cercariae follow a thermal gradient until contact with host skin is detected by lipid receptors. Upon contact they thrust their tails vigorously and begin to secrete their acetabular gland contents. After penetration of the stratum corneum (skin surface) they shed their tails and travel horizontally away from the site of entry. Once away from the site of entry the cercariae migrate deeper into the epidermis towards the basement membrane. Cercariae require significant time to breach this barrier before continuing through the dermis. After penetration of the epidermal / dermal basement membrane the acetabular gland contents have been exhausted. Once in the dermis cercariae locate a venule and enter the circulatory system.

Table I.1.
Geographic distribution of Schistosoma species

Species	Locations	Required Secondary Snail Host Species	Disease Characteristics
S. mansoni	Africa, the Middle East, the Caribbean and South America (53 countries)	*Biomphalaria*	Intestinal schistosomiasis
S. haematobium	Africa and the Middle East (54 countries)	*Bulinus*	Urinary schistosomiasis
S. intercalatum	Ten central African countries	*Bulinus*	Intestinal schistosomiasis
S. japonicum	Seven countries in South-East Asia and in the Western Pacific region	*Oncomelania*	Asiatic intestinal schistosomiasis
S. mekong	Mekong river basin	*Tricula*	Asiatic intestinal schistosomiasis

Chapter 1: Schistosome invasion of human skin and degradation of dermal elastin is mediated by a single serine protease

(This work was published in the Journal of Biological Chemistry, Salter et al. 275(49):38667 / reprinted with permission)

1.1. Summary

Aquatic larvae (cercariae) of the trematode parasite, *Schistosoma mansoni*, rapidly penetrate human skin by degrading host proteins including elastin. Two serine proteases, one chymotrypsin-like and the second trypsin-like, have been proposed to be involved. To evaluate the relative roles of these two proteases in larval invasion both were purified, identified by sequence, and then biochemically characterized. The trypsin-like activity was resolved into two distinct serine proteases 76% similar in predicted amino acid sequence. Southern blot analysis, genomic PCR, and immunolocalization demonstrated that the trypsin-like proteases are in fact not from the schistosome, but released with larvae from the snail host *Biomphalaria glabrata*. Invasion inhibition assays using selective inhibitors confirmed that the chymotrypsin-like protease is the enzyme involved in skin penetration. Its ability to degrade skin elastin was confirmed and the three sites of cleavage within elastin help define a new family of elastases.

1.2. Introduction

Infection of a human host by the trematode parasite *Schistosoma mansoni* begins with invasion of intact skin by an aquatic larva, the cercaria(ae) [41]. Exiting the freshwater snail *Biomphalaria glabrata*, cercariae locate a human host by thermal [2] and chemical signals [3], and rapidly penetrate the skin, entering the vascular system in the dermis [6]. *S. mansoni* cercariae are approximately 150 microns long and 70 microns wide and require lysis of skin tissues to migrate into blood vessels. Host macromolecules

representing barriers to cercarial invasion are known to be cleaved by proteolytic activities present in cercarial secretions. These include elastin [42], chondromucoprotein [28], keratin [43], fibronectin, laminin, and collagen IV and VIII [44]. Two distinct serine proteases have been reported in extracts of cercariae or in secretions from cercariae induced upon contact with skin lipids. One is a "chymotrypsin-like" protease with a preference for large hydrophobic side chains at P1 [45]. The second is a "trypsin-like" protease, with preference for positively charged side chains at P1 [11]. The class of proteases responsible for host protein degradation has been demonstrated by several independent studies to be serine proteases [43] but the relative contributions to invasion of the trypsin-like or chymotrypsin-like proteases is not known.

To analyze the relative contributions of each of these proteases to the degradation of host proteins, cercarial secretions were fractionated and the two proteases purified. The trypsin-like activity, that had not been previously purified or sequenced, was purified and a cDNA was cloned by RT-PCR based on amino-terminal amino acid sequence. Specific inhibitors were identified to evaluate the role of each protease in skin invasion. Proteases were localized by immunohistochemistry and specific sites of cleavage in elastin, the most protease-resistant target in host skin, were analyzed.

1.3. Experimental procedures

1.3.1 Chemical Materials

AAPF-pNA and LGR-pNA were purchased from Sigma. FPR-CMK and AAPL-CMK were purchased from ESP (Enzyme Systems Products, Livermore, CA), p32-a-dCTP was purchased from NEN Research Products, IPTG was purchased from Amersham Pharmacia Biotech, Piscataway, NJ.

1.3.2. Biologic Materials

Schistosoma mansoni -- Approximately $5x10^5$ cercariae (Puerto Rican strain) were collected in 750 ml of distilled water from 200-300 infected *Biomphalaria glabrata* snails using a light induction method previously reported [37]. Cercarial secretions were collected using a modification of a previously reported technique [46]. The cercariae were placed in petri dishes coated with linoleic acid (to simulate skin contact) and floated in a 37° C water bath to produce a thermal gradient. After 2 hours the cercariae had released the majority of their gland contents and the conditioned water was collected and filtered to remove cercarial bodies and debris. The secretion sample was then lyophilized and stored at +4° C until further purification.

Biomphalaria glabrata -- Snails were maintained with a diet of organic lettuce and school chalk as a calcium supplement [47]. All snails were housed in the absence of light to increase yields of cercariae during light induction.

E. coli -- BL21 cells (Novagen, Madison, WI) were electroporated with the appropriate plasmid construct and selected overnight using LB 50 μg/ml ampicillin plates. A single colony was picked and grown overnight in 4ml of LB with 100 μg/ml ampicillin. 1 ml of overnight culture was added to 1000 ml LB with 100 μg/ml ampicillin and grown to an OD600 = 0.6. IPTG was then added to100 μM and after four hours of continued incubation the cells were pelleted and frozen at -70° C.

1.3.3. Gel filtration chromatography

Lyophilized secretions from $4x10^6$ cercariae were resuspended in 6 ml of running buffer (200 mM Na-Acetate pH 6.8) and loaded onto a SR 16/100 column (Amersham Pharmacia Biotech, Piscataway, NJ) packed with Sephacryl 200 (Amersham Pharmacia Biotech, Piscataway, NJ). Fractions were eluted at a pumping rate of 25 ml / hour. Four milliliter fractions were collected at +4° C.

1.3.4. Benzamidine affinity chromatography

Lyophilized secretions from 4 x 10^6 cercariae were resuspended in 8 ml of running buffer (100 mM Tris pH 8.0, 500 mM NaCl) and pumped at 1ml / min through a HR 5/5 column packed with benzamidine-sepharose (Amersham Pharmacia Biotech, Piscataway, NJ). The column was washed with 6 ml of washing buffer (10 mM Tris pH 8.0, 50 mM NaCl) and eluted with 6ml of elution buffer (20 mM Na-Acetate pH 4.5, 50 mM NaCl). 300 υl fractions were collected.

1.3.5. Ion exchange chromatography

Fractions from the benzamidine affinity column were diluted 10 fold into running buffer (20 mM Tris pH 8.0) and pumped through a HR5/5 Mono-Q column (Amersham Pharmacia Biotech, Piscataway, NJ) at 1 ml/min. The column was then washed with an additional 10 ml of running buffer. Elution was performed over 30 1 ml fractions using running buffer and a linear salt gradient from 0 to 1.0 M NaCl.

1.3.6. Ni-ATA affinity chromatography

E. coli pellets from 330 ml of inducted culture were thawed and resuspended in 5 ml of binding buffer (8 M urea, 50 mM Tris pH 8.0). Cell disruption was completed with 6 rounds of 20 second sonications on ice. Histidine tagged proteins were then purified

according to the manufacturers instructions (Qiagen, Valencia, CA) using 1ml spin columns.

1.3.7. Enzyme and Inhibitor Assays

All assays were performed with native enzymes in buffer with 100 mM glycine pH 9.0 (the pH optima with small synthetic substrates) with 100 μM of pNA substrate at room temperature unless otherwise noted. Inhibitors were tested in the same buffer system under the same conditions. A 15 minute pre-incubation of inhibitor with enzyme was carried out before substrate was added. A standard assay consisted of 10 μl of sample and 100 μl of assay buffer in nylon 96 well plates (Falcon PVC plates, Becton Dickinson, Franklin Lakes, NJ). Optical density was monitored using a UV-Max spectrophotometer and SoftMax v2.02 software (Molecular Devices, Sunnyvale, CA).

1.3.8. Skin Invasion Assays

Skin invasion assays were performed as previously described [37]. Briefly, human skin samples were fixed on plastic wells over warm media (RPMI-1640) while 200 ml of inhibitor solution (2 mM in DMSO) or controls, including DMSO, were applied to the skin surface and allowed to permeate and dry for 30 minutes. Approximately 3000 cercariae in 3 ml of water were then applied using 15 mm plastic cylinders on the skin. After 120 minutes the skin was fixed in 10% phosphate-buffered formalin, embedded in paraffin, sectioned at 7 μm, and stained with hematoxylin and eosin. Cercariae penetrating the skin were counted as described previously [37]. Multiple sections of three separate skin samples were counted for each inhibitor and control.

1.3.9. Protein Blotting and N-terminal Peptide Sequencing

NU-PAGE gradient gels (Invitrogen, Carlsbad, CA) were used according to the manufactures instructions. Proteins were blotted onto PVDF membranes using the Novex

transfer system. Peptide sequencing was carried out using an ABI Procise 491 Protein Sequencer (Applied Biosystems, Foster City, CA) at the UCSF Biomolecular Resource Center.

1.3.10. Nucleic Acid Purification and cDNA Synthesis

Poly(A)+ mRNA was isolated by poly(T) affinity chromatography (Amersham Pharmacia Biotech, Piscataway, NJ) from the hapatopancreases of 5 infected snails. RNA was converted to cDNA using AMV Reverse Transcriptase as described by the manufacturers (Life Technologies, Rockville, MD). DNA isolation from snails required CsCl purification because of the high proportion of glycoproteins. DNA was isolated from *S mansoni* by standard detergent lysis and phenol / chloroform extraction.

1.3.11. Cloning and Plasmid Construction

A nested PCR strategy was used to clone both BgSP's. Amplification conditions for the primary PCR were 2 cycles of 94° C, 40° C, and 72° C each for 1 minute then 35 cycles of 94° C, 50° C, and 72° C each for 1 minute. Secondary PCR conditions were 25 cycles of 94° C, 50° C, and 72° C each for 1 minute. Primers used for these reactions: BgSP-α 1° forward 5'-ATCGTCGGNGGNAARGARTCNATGC-3', BgSP-β 1° forward 5'-ATGGTCGGWGGWCARGARGCNGTNC-3', BgSP-α 2° forward 5'-CCNAAYAAYCAYAWNTGYGG-3', BgSP-β 2° forward 5'-GCGCCNACNCAYCAYTTYTGYGG-3', the reverse primer for both 1° and 2° reactions was p(dT)15.

The *E. coli* expression construct pET-21a-BgSP-β was assembled by inserting the active portion of BgSP-β into the restriction sites NdeI and XhoI of the vector pET21-a (Invitrogen, Carlsbad, CA). The forward primer 5'-CGCCATATGGTCGGWGGWCARGARGCNGTNC-3' created a NdeI site (underlined) and changed the 1st amino acid to a start codon (I->M). This construct creates an inactive

protease. The reverse primer 5'-GCGCTCGAGTCTGTTGATGACGGTGTTA-3' added an XhoI site (underlined) and deleted the stop codon creating an open reading frame into the pET21-a vector that added a 6 x histidine tag to the C-terminus of the protein.

Genomic PCR was performed using *S. mansoni* or *B. glabrata* DNA. Forward primers were targeted to an internal region overlapping the catalytic serine: BgSP-α 5'-ACGGGGGATGCCGGCGGTC-3' and BgSP-β 5'-CAGGGTGATGCCGGTGGCC-3'. Reverse primers were targeted to unique regions within the 3' UTR of each gene: BgSP-α 5'-TTGGTTCACCGCAGACTT-3' and BgSP-β 5'-CGCCCTTTACTATTCGCAT-3'

1.3.12. Southern Blot

Genomic DNA (10μg) was digested overnight with the individual restriction enzymes EcoRI, PstI, XbaI, and XhoI (New England Biolabs, Beverly, MA). The DNA was transferred to a Hybond-N+ charged nylon membrane (Amersham Pharmacia Biotech, Piscataway, NJ) using alkaline capillary transfer. The DNA was fixed to the membrane by UV cross linking and pre-hybridized for 45 minutes in Rapid-Hybe solution (Amersham Pharmacia Biotech, Piscataway, NJ). 40ng of probe (full active length PCR product) was labeled with P^{32} (NEN Research Products, Wilmington, DE) using a High Prime nick translation kit (Roche, Alameda, CA) according to the manufacturers instructions. Hybridization was carried out at 65° C for one hour. 2X SSC stringency washing was sufficient for the removal of non-specific probe. The membrane was exposed for 3 days at -70° C to Hyperfilm (Amersham Pharmacia Biotech, Piscataway, NJ) using intensifying screens. The membrane was then stripped for 30 minutes in 0.4 M NaOH at 45° C, washed twice for 10 minutes at room temperature in strip solution (200 mM Tris-HCl pH7.0, 0.1 SSC, 0.1 (w/v) SDS), and reused.

1.3.13. Antibody Production and Immunohistochemistry

Recombinant His-tagged protein from *E. coli* was used to produce rabbit antiserum using standard commercial procedures (Corvance, Richmond, CA). Materials prepared for immunohistochemistry were incubated in fixative (2% paraformaldehyde, 1% gluteraldehyde, 0.1 M phosphate buffer pH7.4) for > 2 hours and then embedded in JB4 plastic, 7 micron sections were cut, and antibody localization was visualized using an ABC kit (Vector Labs, Burlingame, CA).

1.3.14. Elastin cleavage site determination

Insoluble elastin was resuspended in 0.1 M glycine pH 9.0 at a concentration of 100 μg/ml. Cercarial elastase was added (1 μg) to 100 μl of protein solution and incubated at 37° C overnight. Digested protein samples were separated using HPLC. Purified digestion fragments were sequenced using an ABI Procise 491 Protein Sequencer (Applied Biosystems) at the Protein and Nucleic Acid Facility, Beckman Center, Stanford University Medical Center.

1.4. Results

1.4.1. Identification of trypsin-like activity within cercarial secretions

In order to identify and isolate a previously reported trypsin-like [11] activity from cercariae, gel filtration was chosen as the first chromatography step to fractionate the complex protein mixture of cercarial secretions (Figure 1.1.). The chromogenic substrate LGR-pNA was used to detect the trypsin activity, and AAPF-pNA was used to detect the chymotrypsin-like activity (SmCE--*Schistosoma mansoni* Cercarial Elastase). The trypsin activity was identified and resolved from the SmCE activity and while it was 30-fold more active than SmCE against its small tri-peptide substrate, it had no significant activity against macromolecular elastin.

Using molecular weight standards as a reference, the trypsin-like activity had an apparent molecular weight of 25 kD while the SmCE activity peak eluted at an apparent molecular weight of 15kD. SmCE has a pI greater than 9.0 and may interact with the column resin, retarding its migration relative to other proteins. At higher salt concentrations the peak of SmCE is shifted to the left (data not shown).

1.4.2. Purification of native trypsin activity

To further characterize this native trypsin activity, benzamidine conjugated to sepharose was employed as a first step affinity column yielding a 1000-fold purification (Table 1.1.). Ion exchange chromatography was employed as the second and final step. The trypsin activity was resolved into two distinct proteases. Sufficient material was generated to obtain unambiguous N-terminal sequences (Figure 1.2) for both peaks of activity. These two activities were designated BgSP-α (*Biomphalaria glabrata* Serine Protease) and BgSP-β corresponding to the order in which they eluted from the column.

1.4.3. Biochemical characterization of BgSP-α and BgSP-β

As shown in Table 1.2. both native BgSP-α and native BgSP-β have a very similar pH optimum and calcium optimum. BgSP-β is significantly more effective at degrading Azocoll (denatured collagen), than BgSP-alpha. BgSP-β will cleave several different peptide substrates that have a charged amino acid in the P1 position, while BgSP-α prefers the substrate LGR-pNA.

1.4.4. Cloning of BgSP-α and BgSP-β cDNAs by RT-PCR

Based on N-terminal sequence data, a set of nested degenerate PCR primers were designed for each protease. RT-PCR was performed on mRNA isolated from the hepatopancreas of infected snails. This is the location where the parasite replicates and develops. A SmCE cDNA had previously been cloned from this source [32]. Once the cercaria leaves the host snail there is relatively little transcription or translation until after it enters the human host.

The nested primers in conjunction with a polyA reverse primer yielded two bands of the predicted size. Cloning and sequencing of these two bands revealed two serine proteases of the chymotrypsin-fold family that are 76% similar (Figure 1.3.). The amino termini matched those of the purified proteases and the S1 subsites were consistent with trypsin-like serine proteases. Both S1 pockets contained an aspartic acid at the position corresponding to trypsin 195D. A BLAST [48] search confirmed that both sequences are unique but similar to other serine proteases with a chymotrypsin tertiary structure. BgSP-α has a putative glycosylation site at position N51. The utilization of this glycosylation site is consistent with the observed molecular weight determined by SDS-PAGE. BgSP-α has an apparent molecular weight 2-3 kD larger than BgSP-β yet their calculated MW's are 26,241 and 26,337 daltons respectively .

1.4.5. BgSP-α and BgSP-β are produced by the host snail *Biomphalaria glabrata*

Southern blot analysis shows that both protease sequences are present only in the genome of the intermediate host snail (Figure 1.4.). Additionally, genomic PCR using *Biomphalaria glabrata* DNA produced a fragment that contained both the protease sequence and the sequence of an intron with the consensus sequence for an acceptor site at the predicted junction (data not shown).

1.4.6. Localization of BgSP-β and SmCE by microscopy

Polyclonal rabbit antiserum was generated against recombinant BgSP-β produced in *E. coli* and purified using a 6X histidine tag. Specific localization of the antiserum was visualized in granules produced by the epithelial cells in the hepatopancreas of the snail. No staining was seen in any stage of cercarial development or in or on free swimming cercariae. In contrast polyclonal anti-sera reactive against SmCE, generated with recombinant SmCE produced in *E. coli*, only localizes within the secretory glands of the cercaria and does not stain the adjoining snail tissue (Figure 1.5.).

1.4.7. Similarity of BgSP's to other serine proteases of Mollusca

A search of the gene database reveals only five proteases that have been identified within the phylum Mollusca [NCBI:txid6447]. Three of these proteases belong to be Furin/Kex family of serine proteases (AF140362--*Lymnaea stagnalis*, AF140361-*Lymnaea stagnalis*, AF107213-*Helix aspersa*) a structurally unrelated serine protease. The two remaining proteases belong to the chymotrypsin family (AA547777-*Biomphalaria glabrata*, X71438-*Haliotis.rufescens*). The amino acid similarities between BgSP-alpha, BgSP-β and these two family members range from 35% to 43%.

1.4.8. BgSP production under various conditions

Two conditions were tested to determine what factors may influence the production of BgSP's within the snail. The total activity of BgSP's within the hepatopancreas of the snail was not effected by feeding. No significant differences were seen over an eight hour time course of feeding following 24 hours of starvation. Total activity within the hepatopancreas was effected by schistosome infection. The displacement and diminution of the snail hepatopancreas by the dividing parasite causes a "crowding out" effect and reduces the total activity of the BgSP's on average 10 fold (data not shown).

1.4.9. Identification of inhibitors for SmCE and BgSP's for use in "chemical knockout"

The tri-peptide inhibitor FPR-CMK was highly effective against the pooled native BgSP's (0.9 nM IC50) and also had some activity against native SmCE (1700 nM IC50). In contrast AAPF-CMK was effective against SmCE (1200 nM IC50) but had no measurable activity against the BgSP's (>10,000 nM IC50) (Table 1.3.). To evaluate the role of these proteases in the invasion process an in vitro human skin invasion assay was used [37, 45]. AAPF-CMK was effective in reducing the number of cercariae penetrating by over 80% (Figure 1.6.), while FPR-CMK produced a 50% reduction in cercarial penetration.

1.4.10. Relative rates of elastin degradation by BgSP, SmCE, and chymotrypsin

Elastin is the skin protein most resistant to proteolysis. It is a major component of the dermal barrier to cercarial invasion [19, 49]. Native BgSP, native SmCE, and sequence grade bovine chymotrypsin were compared on a molar basis against insoluble native bovine elastin. Only SmCE demonstrated significant elastase activity (Figure 1.7.)

1.4.11. Identification of elastin cleavage sites by SmCE

Three cleavage sites localizing to exons 12 (amino acids 202..225) and 13 (226..239) of elastin, were identified (Figure 1.8.). Exon 12 is the same location where macrophage elastase cleaves (R.P. Mecham, personal communication)

1.5. Discussion

Schistosoma mansoni cercariae directly penetrate human skin within five to 10 minutes after contact [4]. During this process 10 glands that comprise approximately 40 percent of the volume of a cercaria release their contents. Previous studies have shown these glands contain and release proteolytic activity [32, 46] and that cercarial invasion is a lytic, not solely a mechanical event [22].

Two proteases were proposed to be involved in skin invasion. The first, a chymotrypsin-like serine protease also known as SmCE (*Schistosoma mansoni* Cercarial Elastase), cleaves a variety of human skin macromolecules and has been studied in detail [32, 36, 44, 45]. The second, a trypsin-like serine protease, has only recently been identified through the use of small synthetic substrates [11]. Secretions from cercariae were collected and assayed for trypsin-like activity using the substrate LGR-pNA. A purification strategy was developed and the activity was found to result from two similar serine proteases designated BgSP-α (*Biomphalaria glabrata* Serine Protease) and BgSP-β.

Southern blot analysis and genomic DNA PCR both demonstrate that the trypsin-like proteases are of snail host origin. The predominance of single bands and their similar arrangements between the two blots suggest that both genes are present in single copies and may be contiguous. Immunohistochemistry localized the protease to secretory vesicles within the epithelium of the snail hepatopancreas. The function of the trypsins for the snail is likely digestion. The trypsin-like activity is released by the snail during the release of cercariae and would therefore contaminate preparations of cercariae or their secretions. Expression of protease activity in the snail was "constitutive" without induction or enhanced release during feeding. However less activity was found in infected snails due to replacement of the snail tissue by the developing parasites.

Elastin is one of the most difficult host macromolecules to degrade and few proteases are "true" elastases capable of degrading native insoluble elastin [50]. To gain insight into the mechanism of elastin degradation by parasite larvae, the site of cleavage by SmCE was determined. Solubilized fragments from native bovine elastin degraded by SmCE were purified and sequenced. Three cleavage sites localizing to two exons (12 and 13 / amino acids 202..239) within bovine elastin were discovered.

To confirm the role of SmCE in the invasion process, and rule out any possible contribution of the BgSP (e.g. being passively carried to skin by cercariae) a human skin invasion assay was used in combination with selective protease inhibitors. One inhibitor (AAPF-CMK) inhibited SmCE alone while the other (LGR-CMK) inhibited both the proteases but was 1900-fold more effective against the BgSP. If BgSP contributed to invasion FPR-CMK should have the predominant effect. If SmCE alone was responsible both inhibitors should be effective but AAPF-CMK would be more active. AAPF-CMK inhibited 80% of cercariae invading, while LGR-CMK inhibited 50%. This is the expected result if the SmCE was acting alone to facilitate cercarial invasion. SmCE is a biologically potent histolytic protease with activity against many of the macromolecular barriers of skin [28, 42-44]. It is one of the few "true elastases" capable of degrading insoluble elastin itself. Its cleavage pattern on elastin suggests exons 12 and 13 (amino acids 202-239) are solvent accessible domains. Exon 12 is also targeted by macrophage elastase and the observations reported here support the conclusion of Mecham et. al. [50] that "elastases" are a more diverse family of proteases than originally thought. Insoluble elastin can be degraded by proteases acting outside the alanine-rich regions targeted by mammalian pancreatic elastases.

1.6 Figures and Tables for Chapter 1

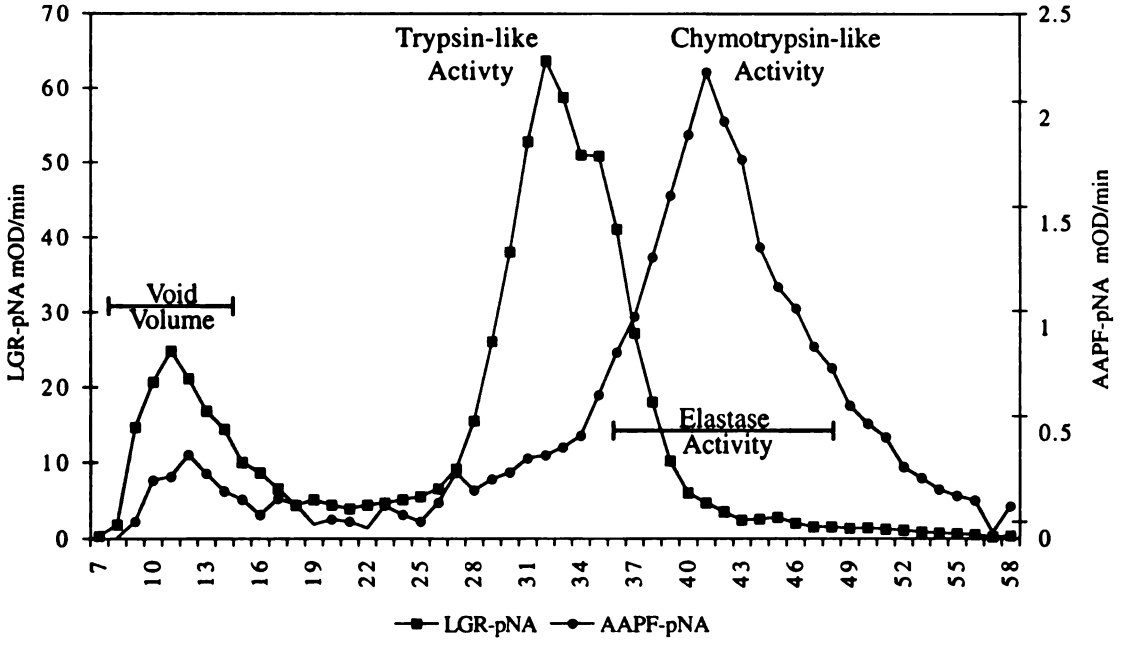


Figure 1.1. Identification of trypsin activity within cercarial secretions fractionated by gel filtration chromatography.

Trypsin activity elutes from the column at its monomeric molecular weight (MW) of 25 kD. Chymotrypsin activity (cercarial elastase) interacts with the Sephacryl 100 resin and has an apparent MW of 15 kD vs. an actual MW of 25 kD. The bar under the chymotrypsin-like activity peak indicates the fractions with activity against native elastin and the bar above the first peak indicates the void volume.

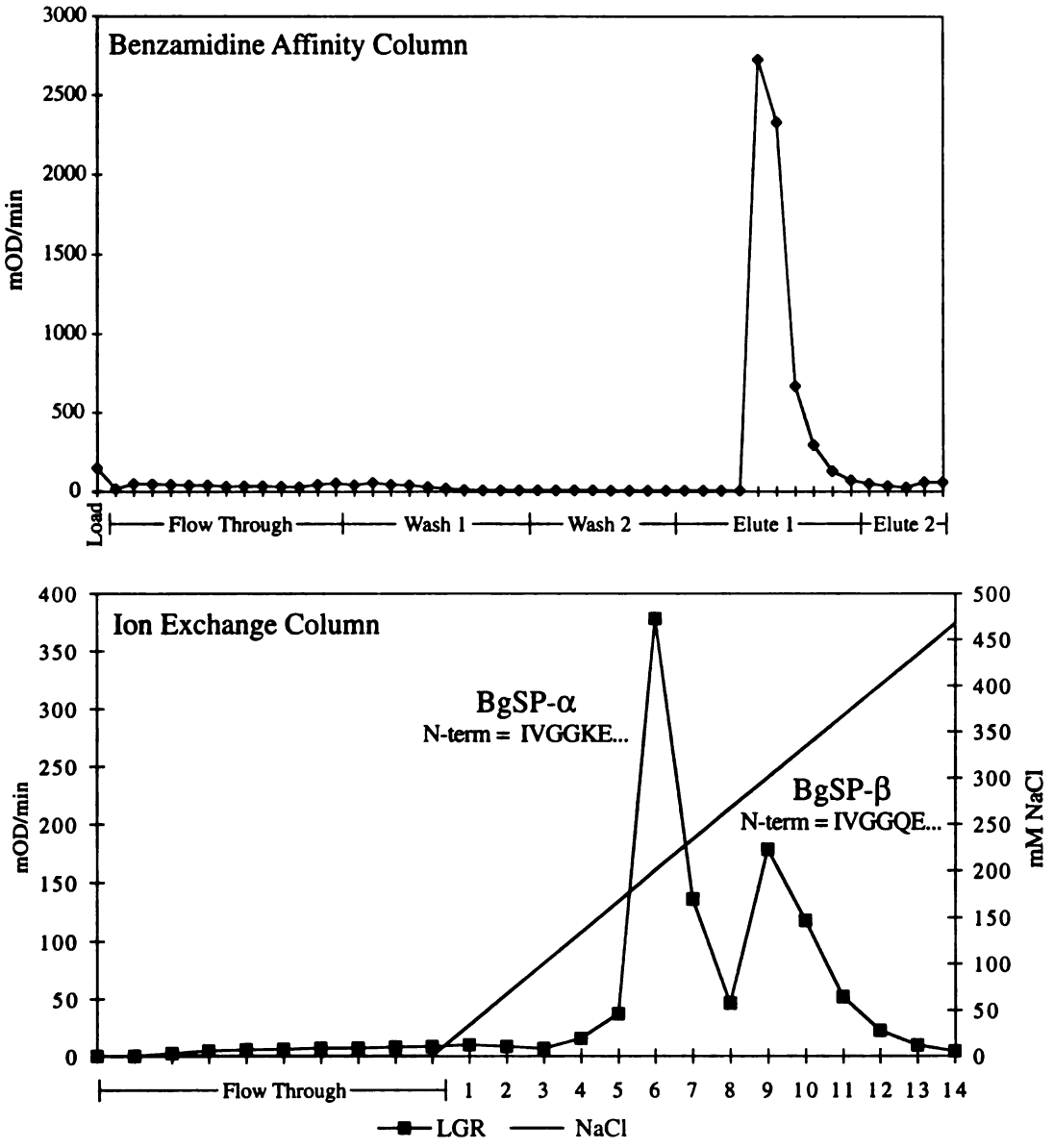


Figure 1.2. Purification of native trypsin activity by affinity and ion exchange chromatography.

Benzamidine sepharose yields over 1000X purification in the first step with a single peak of activity (top graph). Peak is diluted into ion exchange running buffer and fractionated with a Mono-S ion exchange column (lower graph). Two enzymes are revealed using ion exchange and both N-terminally sequences are shown. These enzymes are referred to as BgSP-α and BgSP-β.

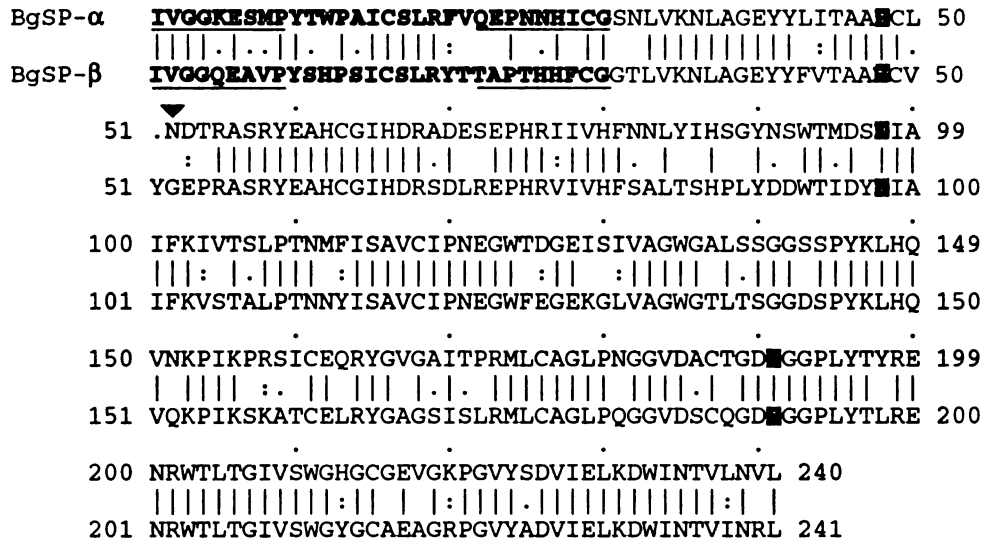


Figure 1.3. Comparison of BgSP-α and BgSP-β.

Bold sequence represents N-terminal sequence data collected from native purification. Underlined sequence indicates regions where degenerate primers were designed for cloning by RT-PCR. Bold and shaded amino acids represent the catalytic core of the serine protease. Solid vertical lines represent identity and dots represent similarities. Inverted triangle indicates the location of a possible glycosylation site for BgSP-α.

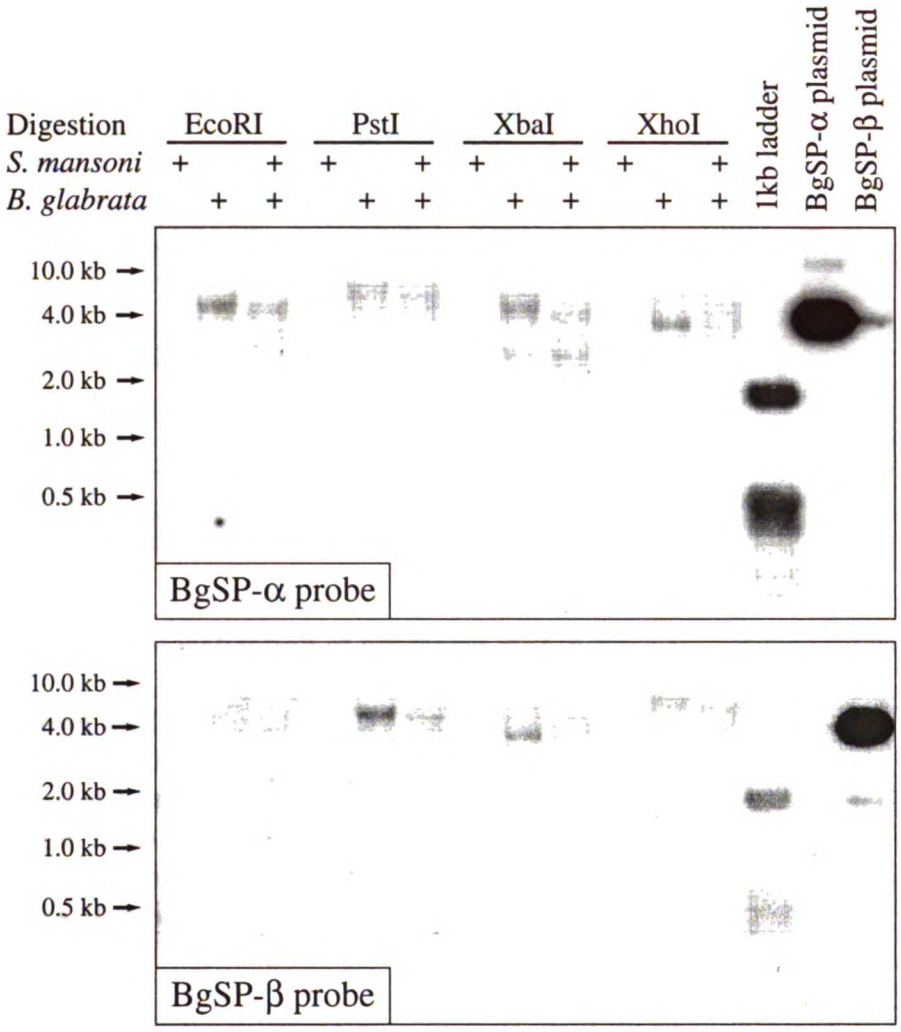


Figure 1.4. Southern blot of *S. mansoni*, *B. glabrata*, and *B. glabrata* infected with *S. mansoni*.

cDNA probes of the mature proteases were labeled and hybridized to the same blot (BgSP-α top image/ BgSP-β bottom image) containing four different digestions of genomic DNA purified from parasite, intermediate host, or infected intermediate. Plasmids containing the mature cDNA sequences were used as positive controls. BgSP-α has an XhoI site within its cDNA.

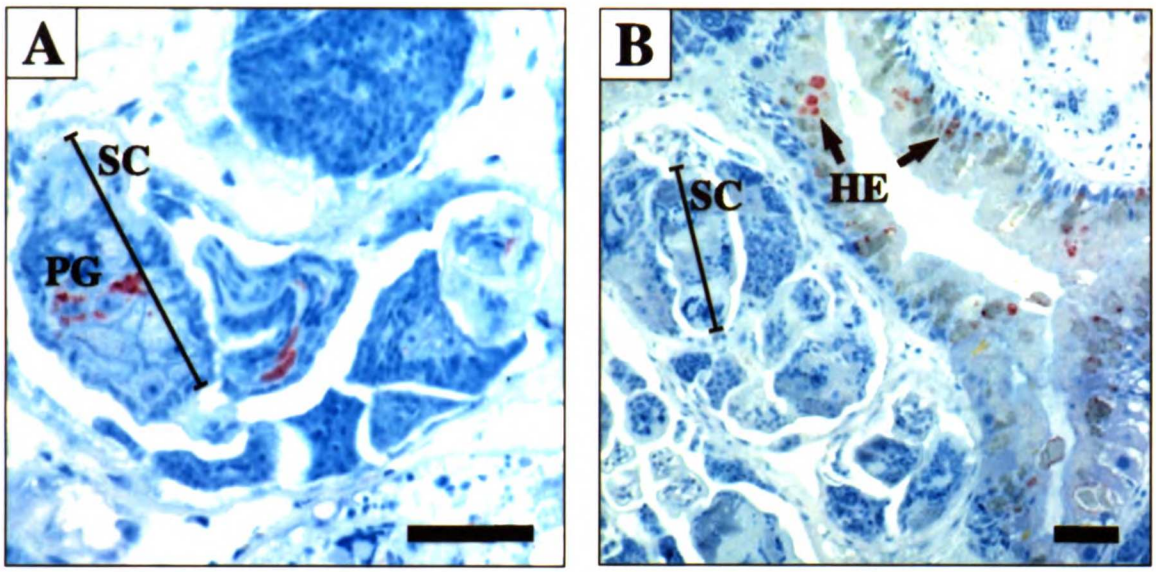


Figure 1.5. Immunolocalization of proteases within the infected host *B. glabrata*.

Image A shows anti-cercarial elastase serum (identified by orange-brown color of peroxidase reaction) localizing within the pre-acetabular glands (PG) of a cercaria (SC with hashed bar) developing within a daughter sporocyst that is within the host snail. Image B shows anti-BgSP serum reacting within vesicles of host snail secretory epithelium (HE) but not within parasitic cercaria (SC with hashed bar). Solid bars are 50mm.

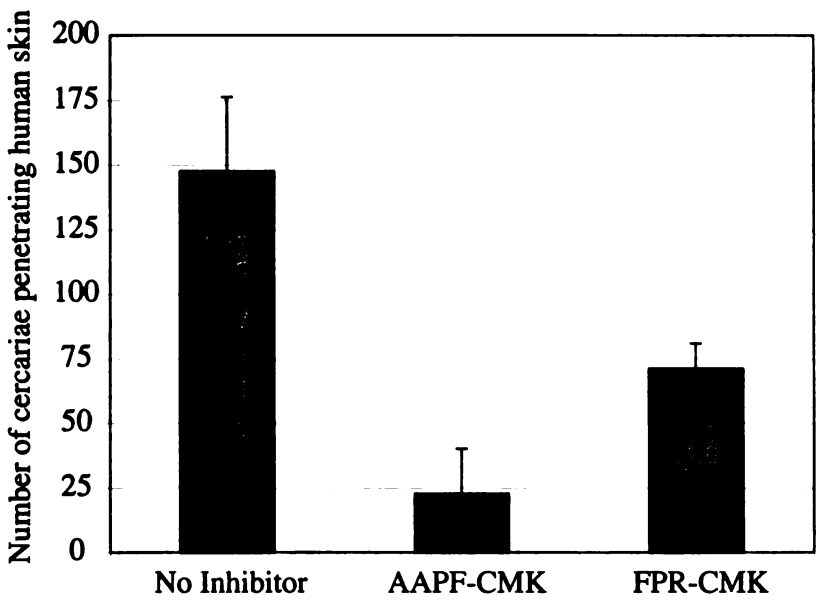


Figure 1.6. Inhibition of cercarial penetration into human skin by topical application of selective protease inhibitors.

Inhibition of cercarial elastase by the selective inhibitor AAPF-CMK (ala-ala-pro-phe chloromethyl ketone) or inhibition of BgSP-α and β with FPR-CMK (phe-pro-arg chloromethyl ketone) as described in 1.3.8.

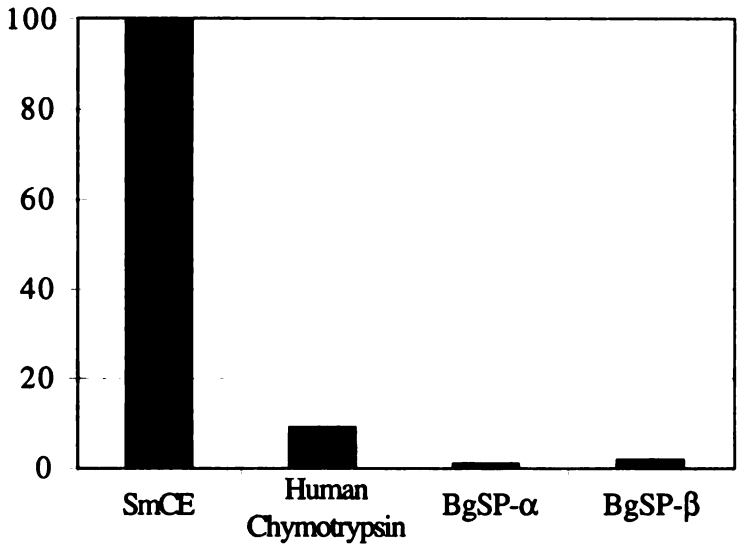


Figure 1.7. Elastin degradation by native SmCE

Human Chymotrypsin, native BgSP-α, and native BgSP-β. Insoluble elastin labeled with congo red dye was incubated with each enzyme for 17 hours. The amount of dye released into the supernatant was quantified and normalized to a scale of 100.

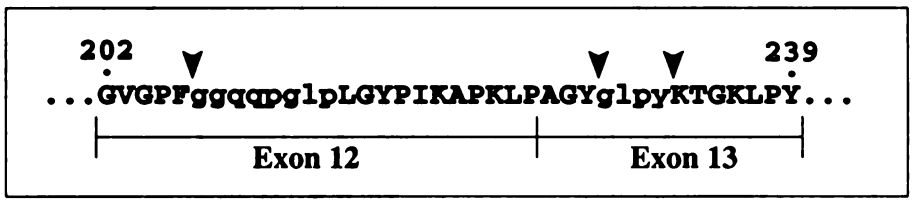


Figure 1.8. Locations of proteolytic cleavage within elastin by cercarial elastase

Arrows mark location of proteolytic cleavage as determined from sequence analysis of fragments (F206/G207, Y228/G229, and Y232/K232). Single letter amino acid sequence only depicts exons 12 and 13 from the 36 exons of elastin (24). Lower case letters represent the sequenced portions of the peptide fragments generated (described in 1.3.14.).

Table 1.1
Purification of native snail trypsins

Step	Volume (ml)	Protein Concentration (mg/ml)	Total Protein (mg)	Enzyme Concentration (mU/ml)	Total Activity (TA) (mU)	% Yield (TA/TA)	Specific Activity (SA) (mU/mg)	Purification (SA/SA)
Starting material	4.5	6.4	29.0	2.0	8.9	--	0.3	--
Benzamidine	0.3	0.10	0.031	39	12	130	380	1300
Mono-Q	1.0	0.001	0.001	1.1	1.1	10	1100	3600

Table 1.2

Biochemical characterization of BgSP's and relative rates of cleavage normalized to preferred substrate

Enzyme characteristics	BgSP-alpha	BgSP-beta
pH optimum with LGR-pNA	9.0	9.0
pH optimum with Azocoll	8.5	8.5
Kcat (LGR-pNA, sec^{-1})	14	12
Km (LGR-pNA, μM^{-1})	110	18
Kcat / Km ($sec^{-1}M^{-1}$)	1.2×10^5	6.5×10^5
Small Substrate Hydrolysis Normalized to LGR activity (mOD/min)		
LGR-pNA	100	100
FR-pNA	1.3	61
AFK-pNA	1.3	12
RR-pNA	0.8	7.4

LGR=Leucine-Glycine-Arginine, FR=Phenylalanine-Arginine, AFK=Alanine-Phenylalanine-Arginine, RR=Arginine-Arginine, pNA=p-nitroanilide

Table 1.3
IC50s for inhibitors of native SmCE and pooled native BgSP's
and their predicted effect on cercariae invading human skin

	IC50 (nM)		Expected Inhibition	
	AAPF-pNA	LGR-pNA	(if SmCE is required)	(if BgSP's are required)
AAPF-CMK	1200	>10000	++	None
FPR-CMK	1700	0.9	+	+++

+++ = >95% inhibition, ++ = >75% inhibition, = + >50% inhibition

Chapter 2: Cercarial elastase is encoded by a gene family conserved across multiple species of schistosomes

(this work has been submitted for publication in the journal of Molecular and Biochemical Parasitiology, J.P. Salter first author, others H. Albrecht, C. Franklin, K.C. Lim, Y. Choe, C.S. Craik, and J.H. McKerrow)

2.1. Abstract

Water borne cercaria(ae) of the trematode genus *Schistosoma* rapidly penetrate host skin. A single serine protease activity, cercarial elastase, is deposited in advance of the invading parasite by holocytosis of vesicles from ten large acetabular gland cells. Cercarial elastase activity is a composite of multiple isoforms. Genes coding for the isoforms can be divided into two classes by amino acid and promoter sequence homology. Two of the five genes identified in *Schistosoma mansoni* account for over 90 percent of the enzyme released. The remaining genes produce little protein or are silent. Positional scanning synthetic combinatorial libraries demonstrate that the two major isoforms have similar substrate specificity and are therefore isoenzymes. The closely related *Schistosoma haematobium* and distantly related *Schistosomatium douthitti* also contain multiple homologous cercarial elastase genes suggesting that gene duplication may have occurred early in *Schistosoma* evolution and that this duplication has been conserved.

2.2. Introduction

Cercaria(ae), the aquatic infective larval stage of schistosomes, are highly adapted to rapidly penetrate the skin of the host upon contact. Enzymatic hydrolysis of host proteins is required for successful entry into the host vascular system [45]. Two gland

systems, the pre-acetabular and post-acetabular glands, release proteases and comprise the majority of the volume of the cercarial head. Each gland cell releases protease at the leading edge of the invading parasite through long microtubule lined ducts. The post-acetabular glands are also responsible for depositing mucin, providing a "sticky" surface on the skin for the parasite to being penetration. Considering the diverse set of macromolecular barriers the cercariae must breach during its invasion, we previously investigated the possibility that multiple enzyme activities were required. Only a single "chymotrypsin-like" activity, cercarial elastase, was found to be present in acetabular gland secretions and required for invasion [38].

We examined the complement of genes coding for cercarial elastase in *Schistosoma mansoni* and found a family of isoforms that can be divided into two classes by amino acid and promoter sequence homology. This family of genes is also conserved in two other schistosome species *Schistosoma haematobium* and *S. douthitti*. The two most highly expressed *S. mansoni* isoforms comprise >90% of the released activity and are virtually identical in biochemical properties.

2.3. Materials and methods

2.3.1. Collection of cercarial secretions

Approximately 3 - $5x10^5$ *S. mansoni* cercariae (Puerto Rican strain) were collected in 750 ml of distilled water from 200–300 infected *Biomphalaria glabrata* snails using a light induction method previously reported (Figure 2.1) [37]. Secretions were collected by placing cercariae in petri dishes coated with linoleic acid (to simulate skin contact) and floated in a 37° C water bath to produce a thermal gradient. After 2 hours the conditioned water was filtered to remove cercarial bodies and debris. The secretions were then lyophilized and stored at +4° C. Each batch of collected material constitutes one "snail shed".

2.3.2. Protease chromatography

Cercarial secretions from four snail sheds were pooled and resuspended in 6 ml of gel filtration buffer (200 mM Na-acetate pH 6.5) and centrifuged for 30 minutes at 10000 RPM in an SA-34 rotor (+4 ° C). The supernatant was loaded onto a SR 16/100 column (Amersham Pharmacia Biotech) packed with Sephacryl 200 (gel filtration resin); (Amersham Pharmacia Biotech). The column was run at a rate of 25 ml/hr and 4 ml fractions were collected overnight at +4° C. Fractions were assayed using 10 µl of sample and 100 µl of assay buffer (100 mM glycine pH 9.0, 100 µm succinyl-alanyl-alanyl-prolyl-phenylalanine p-nitroanilide (AAPF-pNA)). Absorbance kinetics were determined using a UV-Max spectrophotometer and SoftMax Version 2.02 software (Molecular Devices).

Fractions with high AAPF-pNA activity were pooled and buffer exchanged into running buffer (20 mM MES pH 6.8) using PD-10 buffer exchange columns (Amersham Pharmacia Biotech). The sample was then loaded onto an HR 5/5 Mono-Q column at 1 ml/min. Elution of protein from the column was accomplished with a combination of discontinuous steps or gradients of elution buffer (20 mM MES pH 6.8 and 1.0 M NaCl). There were three elution steps separated by linear gradients (fractions 0—20 / 0–125 mM NaCl, fractions 21—25 / 225—300 mM NaCl, fractions 26—30 / 300—1000 mM NaCl). Fractions were assayed as described above.

Peaks of activity from the ion exchange column were analyzed for purity by SDS-PAGE. NuPage-polyacrylamide 4-12% gradient gels (Invitrogen, Carlsbad, CA) were used according to the manufacture's instructions. Gels were stained with Coomassie blue to visualize protein and determine purity.

2.3.3. Immunoblot analysis and antibodies

Proteins were separated by SDS-PAGE as described above and transferred onto polyvinylidene difluoride membranes (PVDF) using the Novex transfer system (Invitrogen). Polyclonal rabbit sera raised against pure recombinant histidine tagged cercarial elastase expressed in *E. coli*, was used to probe the blot. Anti-rabbit HRP conjugated secondary antibody was used for detection with standard ECL reagents (Amersham Pharmacia Biotech).

2.3.4. Promoter Identification by PCR based Genomic DNA Walking

Genomic libraries were constructed with the overnight digestion of genomic 10 μg DNA and 3 μg of pUC19 (Stratagene) plasmid using individual restriction enzymes (AvrI, BamHI, BclI, EcoRI, HindIII, KpnI, PstI, SacI, SalI, SpeI, XbaI, and XhoI); (New England Biolabs). Matching digestions of genomic DNA and plasmid DNA (acting as a universal linker) were mixed and purified using a spin prep column (Qiagen). Purified samples were ligated overnight at +14° C to link the plasmid to the genomic fragments. First round PCRs of nested reactions were set up in volumes of 50 μl using 2 μl of library DNA as template and the primer sets (5'GATTACGCCAAGCTTGCATGCCTG, 5'-TGCGATGAACGGGAATTCAG for cercarial elastase or 5'-GATTACGCCAAGCTTGCATGCCTG, 5'-TGAAGCTATTCTTCGTACATGATAA for GAPDH). Thermocycling conditions were 45 sec @ 94° C, 1min @ 60° C, 3min @ 72° C, for 30 cycles. Second round PCRs were set up in volumes of 100 μl using 0.5 μl of first round PCR product as template and the primer set (5'GATTACGCCAAGCTTGCATGCCTG, 5'GCACAGGTTCACCACTACG for cercarial elastase or 5'GATTACGCCAAGCTTGCATGCCTG, 5'AATAACTCCGGTGTCTTAAGTCATA for GAPDH). Thermocycling conditions were 45 sec @ 94° C, 1 min @ 60° C , 3 min @ 72° C for 30 cycles. Taq polymerase was used in all PCR reactions.

Second round PCR products were resolved with 1% agarose gels. Individual DNA bands were excised, purified with gel extraction columns (Qiagen), and cloned using a TOPO vector (Invitrogen). Plasmid containing inserts of correct size were sequenced.

2.3.5. Isoform identification by RT-PCR

A primer was designed to bridge the deletion region within group 2 promoters. This primer was specific for group 2 transcripts and was used in RT-PCR to amplify the downstream expressed genes.

Poly(A) mRNA was isolated by poly(T) affinity chromatography (Amersham Pharmacia Biotech) from the hapatopancreases of five infected snails. RNA was converted to cDNA using avian myeloblastosis virus reverse transcriptase (Life Technologies, Inc.) as described by the manufacturer. PCR was performed using 5 μl of first strand cDNA as template and the primer set (5'TTGCAACATTCACATAGACAT, dT15). Thermocycling conditions were 30 sec @ 94° C, 30 sec @ 55° C, 1 min @ 72° C, for 30 cycles using Taq polymerase. PCR products were resolved with 1% agarose gels. Individual DNA bands were excised, purified with a gel extraction columns (Qiagen), and cloned into a TOPO vector (Invitrogen). Plasmid containing inserts of correct size were bi-directionally sequenced at the Biomolecular Resource Center of the University of California, San Francisco.

2.3.6. Amino-terminal peptide sequencing

Proteins were blotted onto PDVF membranes and visualized by Coomassie Blue staining. Peptide sequencing was carried out using an ABI Procise 491 Protein Sequencer (Applied Biosystems, Inc) at the Biomolecular Resource Center of the University of California, San Francisco.

2.3.7. Isoform identification by degenerate PCR

Degenerate PCR primers targeted to conserved regions of serine proteases were used to identify cercarial elastase homologues as previously described [51]. Genomic DNA or cDNA were used as template material. PCR was performed using 0.5 μg of template DNA and the primer set (5'CCITTICTITAICGIRAIRA, 5'CCICTR(Y or R)ICCICCIGGIRA). Thermocycling conditions were 30 sec @ 94° C, 1min @ 45 ° C, 1 min @ 72° C, for 35 cycles using Taq polymerase. Individual DNA bands were excised, purified, and plasmids containing inserts of the predicted size were sequenced.

2.3.8. Isoform identification by hybridization of a phage library

A lambda phage library was made according to the manufacturers instructions (Stratagene) using mRNA from snail hepatopancreas infected with *S. haematobium* sporocysts. A radioactive probe from the cDNA of SmCE-1a was used to screen the library. Positive plaques were serially diluted and probed again to ensure individual sequences. Pure colonies were then recovered and sequenced.

2.3.9. Positional scanning–synthetic combinatorial library screening

Production of the diverse chemical libraries have been described in detail previously [52]. Position P1 was scanned using the P1 diverse library. Positions P4 through P2 were scanned using the P1-Leu library. 1 μg of cercarial elastase was used per well with a final sublibrary substrate concentration of 0.1 μM for the P1-Leu library and 0.04 μM for the P1-diverse library. Reactions were started with the addition of enzyme and monitored with a Perkin–Elmer LS50B luminescence spectrometer. Excitation was 380 nm, emission was 450 nm or 460 nm, and was monitored for 1 hour. Reactions were performed at 25° C in reaction assay buffer (50 mM Tris at pH 8.0, 100 mM NaCl, and 0.01% Tween-20). The P1 diverse library was assayed with an equal mixture of isoforms

SmCE-1a and SmCE-1b due to limited enzyme availability. The P1 position results were separately confirmed with individual substrates for both enzymes (data not shown).

2.3.10. Isoform and species sequence relationship

All available amino acid sequence information was aligned using the GCG pileup program (Oxford Molecular Group) and visually inspected for misalignments. A consensus sequences generated from this alignment was used to determine the nearest neighbors with the BLAST program [48]. This sequences was used in combination with the schistosome sequences to calculate a phylogenetic tree using the protein parsimony method of the PHYLIP software package [53].

2.4. Results

2.4.1. Purification of cercarial elastase reveals multiple isoforms

Cercarial elastase purification requires two steps. Gel filtration of cercarial secretions produces a single broad peak of elastase activity that is detected with the tetra-peptide substrate AAPF-pNA (Figure 2.2.A). Ion exchange chromatography was used to purify cercarial elastase isoforms (Figure 2.2.B). A shallow salt gradient revealed three peaks of activity. These peaks correlated with 25 kD protein products (Figure 2.2.C). Polyclonal antiserum generated from pure recombinant non-glycosylated cercarial elastase of bacterial origin was used in an immunoblot analysis of these three peaks. The high level of reactivity for each of the bands suggested strong sequence similarity (Figure 2.2.D). The lower 16 kD band observed in each lane is a degradation product resulting from autoproteolysis [31].

2.4.2. Identification of isoform promoters reveal two classes of cercarial elastase

A PCR strategy was adopted to clone regions of genomic DNA upstream of cercarial elastase genes (Figure 2.3). Multiple libraries of genomic DNA, digested with individual restriction enzymes and ligated with universal linkers, were constructed. Two assumptions were made about the protease gene family. First, a reverse primer homologous to the N-terminal portion of the gene family would bind all the isoforms. Second, restriction sites in the non-coding upstream region would be heterogeneous. PCR amplification using a forward primer homologous to the universal linker and the cercarial elastase reverse primer yielded multiple products that varied with each library (Figure 2.3.A.). Cloning and sequencing of the individual PCR products confirmed that the bands were upstream genomic DNA from multiple related genes. As a control, reverse primers to the GAPDH gene were also used in a separate experiment. GAPDH has been

previously shown to be a single copy gene [54, 55]. The PCR reactions yielded single products for each library, consistent with it being a single copy gene (Figure 2.3.B).

Two genomic clones of cercarial elastase, designated EL1 and EL2, have been reported by Pierrot et. al. [33]. EL1 was described as the genomic clone of the previously published cercarial elastase cDNA sequence (SmCE-1a). EL2 was reported to contained a D125A mutation in the catalytic triad, yielding an inactive protease, and also having a predicted transcript that was undetectable. An alignment of the predicted cDNA from EL1 and the first published cDNA (SmCE-1a) revealed 25 single base pair substitutions suggesting that these genes may be isoforms. A unique SpeI restriction site is present in SmCE-1a but absent in EL1 (SmCE-1b). Digestion of RT-PCR products with SpeI demonstrated that both transcripts are produced and confirmed the genes are isoforms and not strain variants (data not shown). Exhaustive attempts to identify a transcript by RT-PCR with sporocyst cDNA of the predicted catalytically dead gene (EL2) were similarly unsuccessful. Table 2.1 summarizes previously submitted DNA sequences in public databases and their nomenclature relationships to this paper.

Alignment of the upstream non-coding DNA sequences revealed that one of the isoform promoters contained a 10 bp deletion just upstream of the first coding region (Figure 2.4B). The three sequences also have a high degree of conservation centering on the TATA box. This region extends 36 bp upstream, 55 bp downstream, and is 80% identical. Additionally, three amino acids in the first coding region of this fragment did not match the known elastase sequence indicating another isoform. A forward primer bridging the deletion region allowed for the full length cloning of the mRNA downstream of this promoter by RT-PCR. Two PCR products were generated and sequenced. Both products revealed genes coding for cercarial elastase but significantly divergent from those previously published. These results bring the total number of cercarial elastase genes in *S. mansoni* to five (Figure 2.5).

2.4.3. Not all gene isoforms are transcribed and / or translated

N-terminal sequencing of the purified enzymes have confirmed the translation of three of five cercarial elastase genes. Isoforms SmCE-1a and SmCE-1b are the most abundant species present in cercarial extracts comprising over 90% of the protein and activity. Isoform SmCE-2a constitutes a minor component to cercarial secretions. Isoform SmCE-2b has detectable levels of mRNA transcript, but unsuccessful attempts to purify this isoform indicate its abundance is below the levels of our detection methods. The last predicted isoform, SmCE-1c (EL2), could not be detected by RT-PCR using specific primers.

2.4.4. Positional scanning, combinatorial substrate libraries demonstrate the two major isoforms are isoenzymes

Substrate preferences between the two major species of cercarial elastase, SmCE-1a and SmCE-1b, were compared using a combinatorial library that scanned substrate positions P4, P3, and P2. Both enzymes prefer the same sets of amino acids for each of the positions scanned indicating the isoforms are enzymatic equivalents (Figure 2.6).

Position P4 is selective for serine or threonine but will tolerate 18 other amino acids. Position P3 showed no specific preference. The P2 position is the most selective position of the enzyme preferring only proline. The selectivity for P2 in SmCE-1b is slightly more broad than SmCE-1a with alanine also accepted in the position. The P1 position for SmCE-1a and SmCE-1b were scanned as an enzyme mixture due to the high enzyme requirements of the P1 diverse library. Both isoforms where then tested with individual tetra-peptide substrates for each P1 amino acid that had activity and were found to be similar (data not shown).

2.4.5. *S. haematobium* and *S. douthitti* contain homologous cercarial elastase genes

A variety of methods were used to identify sequences within other species of schistosomes. Degenerate PCR using *S. haematobium* cDNA template produced two similar but unique sequences, identifying two translated isoforms designated ShCE-1a and ShCE-1b. A phage cDNA library was also screened, using a radiolabeled SmCE-1a probe, and identified a single full length mRNA for ShCE-1a (Figure 2.5). All the phage clones sequenced were identical suggesting the relative abundance of the two *S. haematobium* transcripts may differ significantly. The same PCR strategy using *S. douthitti* genomic DNA also produced two fragments, SdCE-1a and SdCE-1b, of which SdCE-1b contained a unique intron that does not share a location with any introns present in *S. mansoni* cercarial elastase sequences (Figure 2.7.).

2.4.6. Molecular phylogeny of cercarial elastase gene family

The phylogenetic relationship of all nine cercarial elastase amino acid sequences were calculated (Figure 2.8.). Each set of isoforms from each species clusters together. *S. mansoni* and *S. haematobium* are closely linked while the *S. douthitti* enzymes branch from the *S. haematobium* cluster. The most closely related serine proteases from other organisms were trypsin-related proteases from *Metarhizium anisopliae* (AF130865) and human elastase IIIA precursor (A29934).

2.5. Discussion

Schistosome cercariae must penetrate the formidable barrier of skin. To do so they have evolved a potent serine protease capable of degrading many macromolecular targets [31]. Previous studies have confirmed the necessity for this protease activity in cercarial invasion [45]. Recent work has confirmed that it is the sole histolytic activity present in cercarial secretions [38]. As such, one would expect it to be conserved across schistosome species. The studies reported here confirm that it is highly conserved between *S. mansoni* and *S. haematobium*, and it is identifiable in the more distantly related *S. douthitti*.

The cercarial elastase activity is encoded by a family of closely related genes. The presence of gene duplication and the resultant protease isoforms in part may reflect selection for redundancy in a key gene required for transmission from snail to human host. Comparison of the dominantly expressed cercarial elastase isoforms in *S. mansoni* demonstrated a high degree of similarity in their preferred substrates. This suggests that these two genes are not providing substrate cleavage diversity but may be playing a role to increase the total amount of protein produced in the acetabular glands through duplication.

The protease isoforms fall into two families identifiable by differences in their promoters. Expression of cercarial elastase is strictly limited to the sporocyst lifecycle stage when cercariae are developing [32, 34]. However, cercarial elastase in *S. mansoni* has been shown to be expressed in both pre- and post-acetabular cells. These two cell compartments, while contributing to cercarial secretions, produce different products. The pre-acetabular cells have cercarial elastase co-localized in vesicles containing high levels of calcium, whereas the post-acetabular glands contain protease and mucopolysaccharides [56, 57]. Perhaps the two promoter families represent gene products that localize to these two separate compartments.

Alternatively, the gene redundancy could be explained by a hypothesis we call the "gene in waiting hypothesis". The duplication and subsequent divergence of these genes will ultimately alter the enzymes specificity for its substrate. This reservoir of genes could provide protection for the parasite against host adaptation by allowing new genes to quickly replace the current highly expressed genes through changes in transcription levels. This hypothesis can explain why there are multiple isoforms yet only two highly conserved genes accounting for the majority of protein produced. This hypothesis is also testable. Additional data on schistosomes infecting a variety of hosts would be predicted to show differential levels of gene transcripts and products with conservation of the genes themselves.

High sequence conservation around the TATA box makes it impossible to identify specific potential transcription factor binding sites but does suggest that the assembly of the transcription machinery would involve most of this region. A better understanding of the schistosome's transcriptional regulation will be possible as more upstream sequence information is obtained to allow comparisons between and within species.

We have shown cercarial elastase activity to be the sum of multiple protease isoforms. Protease gene duplications are observed in the three schistosome species and provide an opportunity to document the evolution of a gene family in the genus *Schistosoma*. Positional scanning combinatorial substrate libraries were used to demonstrate the proteases expressed from the gene family are isoenzymes.

2.6 Figures and tables for chapter 2

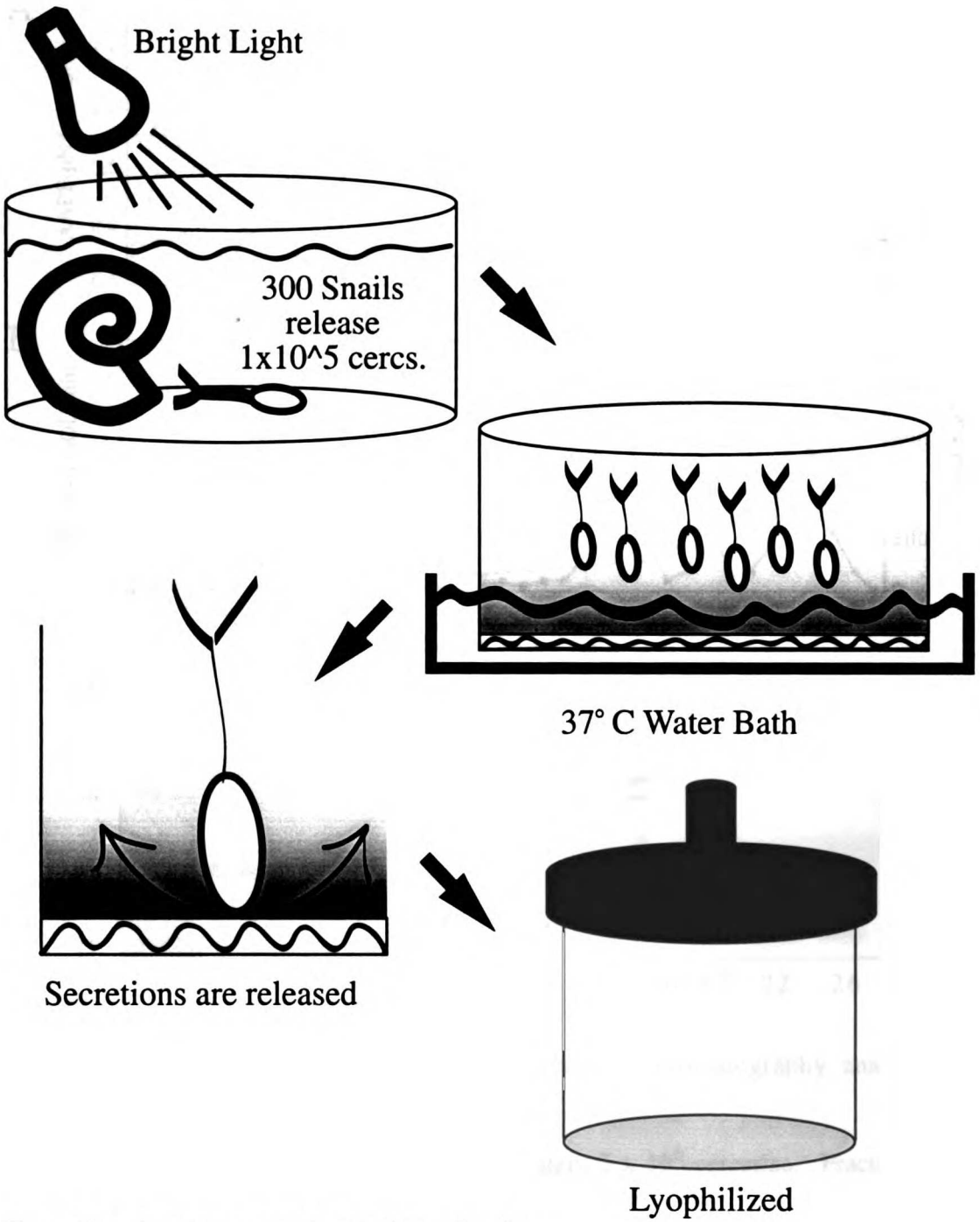


Figure 2.1. Steps in cercarial secretion collection

Light is used to induce the cercariae to migrate of out of the snail host. Cercariae then follow a thermal gradient and contact the bottom of the petri dish. The surface of the dish is coated in lipid. This induces the cercariae to begin their penetration program and secrete their gland contents. The bodies are filtered and the conditioned water is lyophilized.

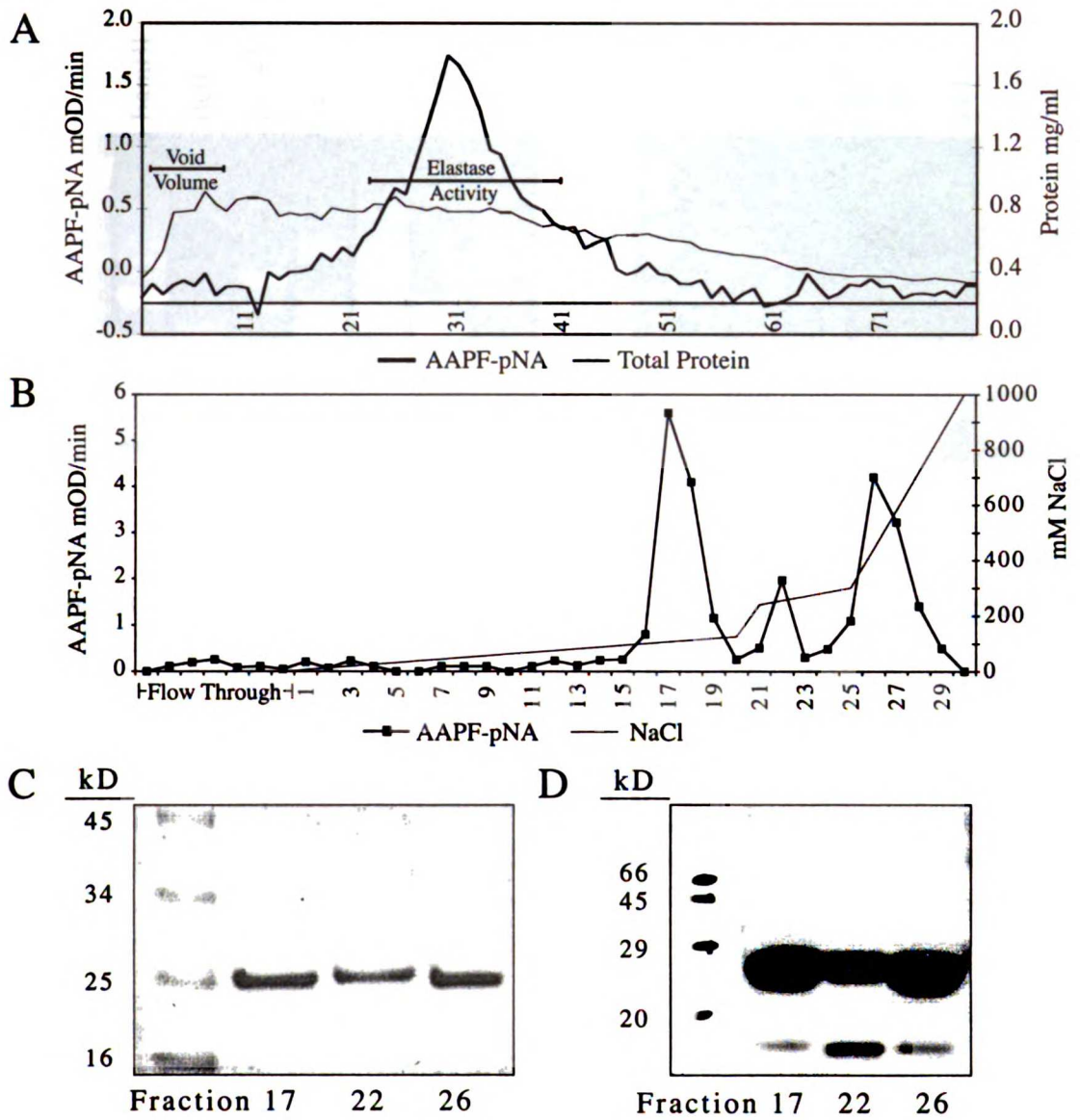


Figure 2.2. *Schistosoma mansoni* cercarial elastase chromatography analyzed by Coomassie and Western blot.

(A) Gel filtration of secretions from approximately 2 x 10^6 cercariae. Fractions were measured for AAPF-pNA activity and protein concentration. Void volume contains unfractionated high molecular weight proteins. Elastase activity indicates fractions capable of degrading insoluable elastin. (B) Ion exchange of fractions with high AAPF-pNA activity from A. Samples were pooled and desalted before loading. A step gradient of salt was used to elute the activity. Fractions were assayed with AAPF-pNA. FT indicates flow through of loaded sample. (C) SDS-PAGE folloed by Coomassie blue staining of activity peaks from panel B. (D) Immunoblot using antibodies reactive to recombinant cercarial elastase. Activity peaks from panel B are displayed.

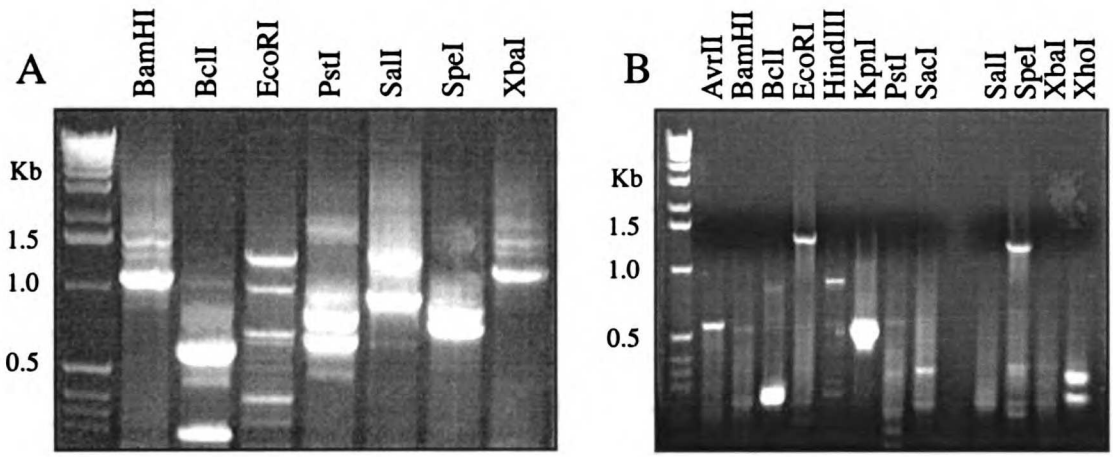
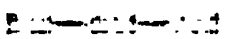


Figure 2.3. Genomic fragments amplified from *Schistosoma mansoni* libraries

Uniquely restricted genomic libraries were used as template DNA in nested PCR reactions. Forward primers were homogous to a linker ligated to the restricted fragments and reverse primers were homologous to cercarial elastase (A) or GAPDH (B). Both gels show the second round of the nested PCR amplifications. Note: If muliple copies of a gene are present in the genome and restriction site locations are heterogenous--then multiple bands would be expected. GAPDH, known to be a single copy gene in *Schistosoma mansoni*, was included as a control.

A

	Sequence information	Transcribed (RT-PCR or) cDNA)	Translated (Purified & Sequenced)	Relative Abundance %
Group 1 cercarial elastases				
CE-1a		√	√	45
CE-1b		√	√	45
CE-1c		-	-	0
Group 2 cercarial elastases				
CE-2a		√	√	10
CE-2b		√	?	<5

B

```
          1                                                    50
CE-1a     ctgactgact gacttacaca catgcaaaat cactcactca ctcacatgta
CE-1b     taataacaca cactgactga ctgacacaca tgcgaaatct ctcacatgta
CE-2a     ttgagtggag tgcaggttgg cttcctcaca .atgcaccga cacacgtgca

          51                                                   100
CE-1a     ctcaacactc atggaatatt ggtgcaTATA Aatag..gta aatgtgttgc
CE-1b     ctcaacagtc atggaatatt ggcgcaTATA Aataggtgtg aatgtgttgc
CE-2a     ctcga.agtc acggaatatc ttcacaTATA Aata..agtg aatgtgttgc

          101                                                  150
CE-1a     aacattgta. gtggttgatt acaacgctca catacataat cacatcgATG
CE-1b     aacattgta. gtggttgatt gcaacggtca catacataat cacatacATG
CE-2a     aagattgtag gtagttgatt gcaacagtca cataga~~~~ ~~~~~~cATG
```

Figure 2.4. *Schistosoma mansoni* cercarial elastases and their transcription / translation properties

(A) Two groups of cercarial elastase have been assigned according to upstream DNA sequence and gene sequence homology. Thin lines represent non-coding regions and blocks represent coding regions. Empty gaps between blocks represent predicted intron positions. Line and block lengths are not proportional. (B) Alignment of Group 1 and Group 2 non-coding upstream DNA. Shaded regions indicate indentity within the three squences. The TATA box and ATG start codon are outlined and the identifing Group 2 deletion is underlined.

```
SmCE-1a  MSNRWRFVVV VTLFTYCLTF ERVSTWLIRS GEPVQHPAEF PFIAFLTTER 50
SmCE-1b  MSNRWRF.LV VTLFTYCLTF ERVSTWLIRS GEPVQHRTEF PFIAFLTTER
SmCE-1c  MSNRWRF.LV VTLFIYCLTF KRVSTWLIRS DQPVQKHTEF PFIAYIASKK
SmCE-2a  MLN.GRTFLM VTLFTYCLTF ERVSTWLVRK GEPVQDRTEF PYIAFVRTER
SmCE-2b  MLNGWTF.LI VTLFTYYLTC QCVSTWLIRS GEPVQQRTEF PFIALLMTDA
ShCE-1a  MLNRWRF.LV VTLFTYCLTV ERVSTWLIRS GEPVQQRTEF PFIAFLTTER

SmCE-1a  TMCTGSLVST RAVLTAGHCV CSPLPVIRVS FLTLRNGDQQ GIHHQPSGVK 100
SmCE-1b  TMCTGSLVST RAVLTAGHCV CSPLPVIRVS FLTLRNGDQQ GIHHQPSGVK
SmCE-1c  SMCTGSLVST RVVLTAGHCV CPPMPVIKVT FLTLRNGDQQ GIHHQPSGVK
SmCE-2a  TMCTGSLVST RAVLTAGHCV CSPMPVVQVS FLTLRNGDQQ GIHHQPSGVK
SmCE-2b  SMCTGSLVSS RAVLTAGHCV CGQTPIVRVS FLSLSEFDQR TINHRPLEIK
ShCE-1a  TMCTGSLVST RAVLTAGHCV CSPLPVVRVS FLTLRNGDQQ GIHHQPSGVV

SmCE-1a  VAPGYMPSCM SARQRRPIAQ TLSGFDIAIV MLAQMVNLQS GIRVISLPQP 150
SmCE-1b  VAPGYMPSCM SARRGRPIAQ TLSGFDIAIV MLAQMVNLQS GITVISLPQA
SmCE-1c  VAPEYMPSCT ASRQRRRIRQ TLSGFAIATA MLAQMVNLQS GIRVIGLPQA
SmCE-2a  VAPEYMPSCT ASRQRRRIRQ TLSGFDIATV MLAQMVNLQS GIRVISLPQA
SmCE-2b  VAPEYNPVCQ LKRENKRITK SLGGYDMAII TLTNLVNLET GVKVISLAAE
ShCE-1a  VAPGYMPSCM SARQRRPIQQ TLSGFDIAIV SLAQLVNLQT GIRVLSLPQP

SmCE-1a  SDIPPPGTGV FIVGYGRDDN DRDPSRKNGG ILKKGRATIM ECRHATNGNP 200
SmCE-1b  SDIPTPGTGV FIVGYGRDDN DRDPSRKNGG ILKKGRATIM ECRHATNGNP
SmCE-1c  SDIPTPGTDV FIVGYGRDDN DRDPSRRAGG ILKKGRATVT ECRHETHVNP
SmCE-2a  SDIPTPGTDV FIVGYGRDDN DRDPSRRAGG ILKKGRATVM ECKHSTTGNP
SmCE-2b  LDIPIPESIA YMVGYGQDIR DPDPSGRYGG ILKKGSATIM ACRHKTFGDP
ShCE-1a  TDIPRPGTPV FIVGYGRDDN DRDPSRRNGG ILKKGRATIM ECRHATNGDP

SmCE-1a  ICVKAGQNFG QLPAPGDSGG PLLPSLQGPV LGVVSHGVTL PNLPDIIVEY 250
SmCE-1b  ICVKAGQNFG QLPAPGDSGG PLLPSPQGPV LGVVSHGVTL PNLPDIIVEY
SmCE-1c  ICVKAGPNSG QLLGPGDSGG PLLPSPQGPV LGVVSHGVTL SHLP....EY
SmCE-2a  ICVQAAYVFG QITAPGDSGG PLLRSPQGPV LGVVSHGVTL SNRLDVLVEY
SmCE-2b  ICVKPGPNSK QIAGPGDSGG PLLLTPQGPI VGVASNGVFL PALADLCVEY
ShCE-1a  ICVKSGQNFG QLPAPGDSGG PLLPSPQGPV LGVVSHGVTL PNLPDIIVEY

SmCE-1a  ASVARMLDFV RSNI 265
SmCE-1b  ASVARMLDFV RSNI
SmCE-1c  VSVARMLNFV RSNI
SmCE-2a  ASVARMLGFV SSNI
SmCE-2b  SSVPRMLKFI LPNI
ShCE-1a  ASVARMLDFV RSNI
```

Figure 2.5. Alignment of full length cercarial elastases protein sequences

Six cercarial elastase family proteases are aligned by amino acid sequence. The pro-domain / active domain junction is indicated by a solid vertical bar. Shaded amino acids indicate the catalytic triad of the proteases. Horizontal bars above and below the aligned sequences indicate predicted sites of auto-proteolysis.

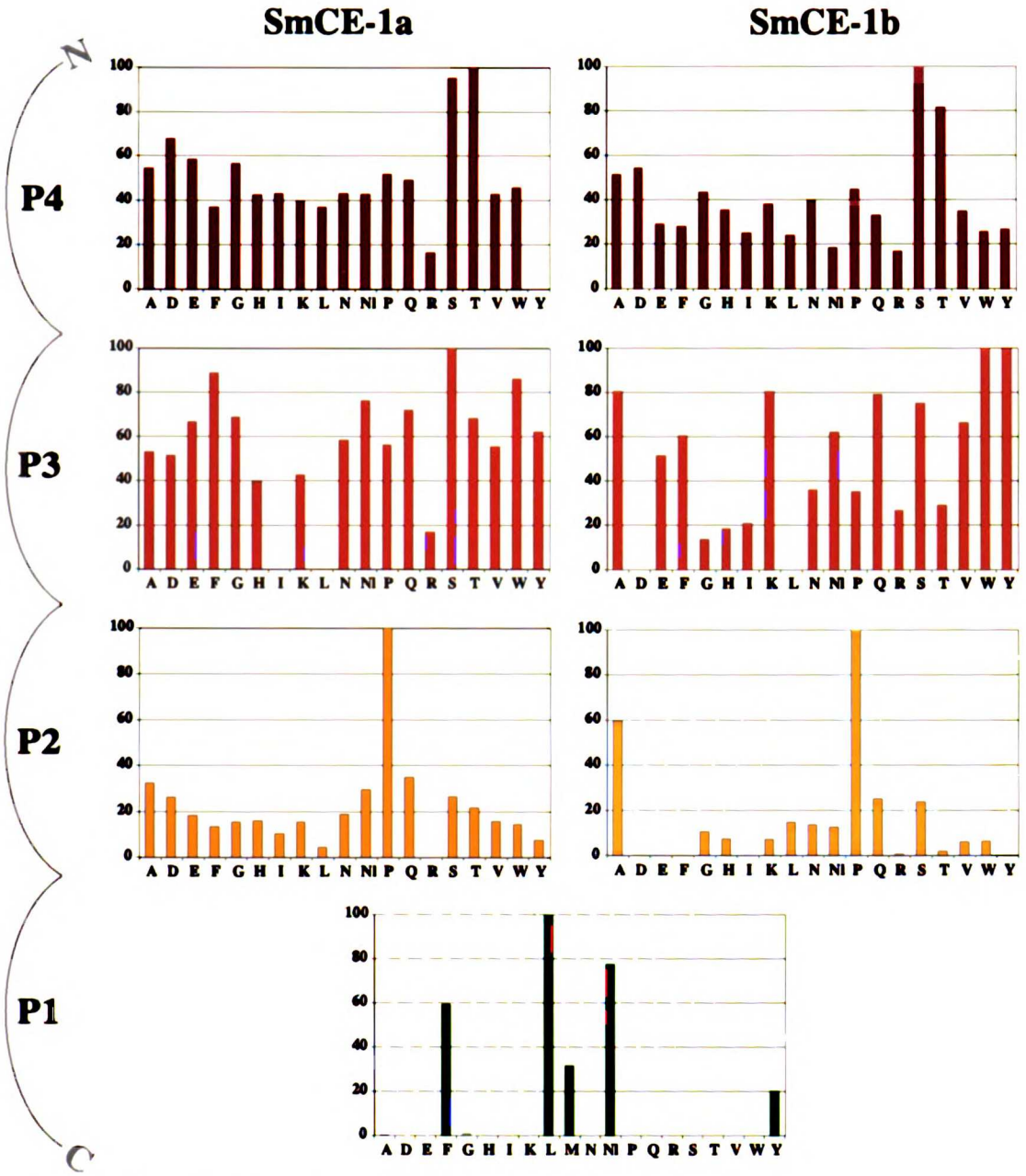


Figure 2.6. Positional scanning, combinatorial substrate libraries, of SmCE-1a and SmCE-1b.

P1-P4 represent four amino acid side chains recognized by protease on the amino terminal side of peptide bond cleavage. Letters on x-axis are standard single letter amino acid codes with Nl = norleucine. Bars indicate level of cleavage for each library member as judged by release of AMC from substrate (see methods). The single P1 histogram represents the combined activity of SmCE-1a and SmCE-1b.

```
ShCE-1a  DQQGIHHQPS GVVVAPGYMP SCMSARQRRP IQQTLSGFDI AIVSLAQLVN 50
ShCE-1b  DQQGIHHQPS GVVVAPGYMP SCMSARQRRP IKQTLSGFDI AIVSLAQLVN
SdCE-1a  DQQGIHHQPS GVVVAPEYMP SCASSRQRRR VKQTLSGFDI AIVLLAEMVN
SdCE-1b  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

ShCE-1a  LQTGIRVLSL PQPTDIPRPG TPVFIVGYGR DDNDRDPSRR NGGILKKGRA 100
ShCE-1b  LQTGIRVLSL PQPTDIPRPG TPVFIVGYGR DDNDRDPQRK NGGILKKGRA
SdCE-1a  LQTGIKVLSL PQPTDIPAPG TPVFIVGYGR DDNDRDPSRR NGGILKKGRA
SdCE-1b  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~GRA

ShCE-1a  TIMECRHATN GDPICVKSGQ NFGQLPAPGD SGGPLLPSPQ GPV 143
ShCE-1b  TIMECRHATN GDPICVKSGQ NFGQLPAPGD SGGPLLPSP~ ~~~
SdCE-1a  TVMESKHSTT GNPICVQGRY VFGQIPAPGD SGGPLLPSPQ NSI
SdCE-1b  TVMESKHSTT GNPICVQGRY VFGQIPAPGD SGGPLLPSPQ NSI
```

Figure 2.7. Alignment of cercarial elastase isoform fragments from *Schistosoma haematobium* and *Schistosomatium douthitti*

The shaded amino acid indicates part of the catalytic triad of the proteases. The boxed region highlights a unique and highly charged loop.

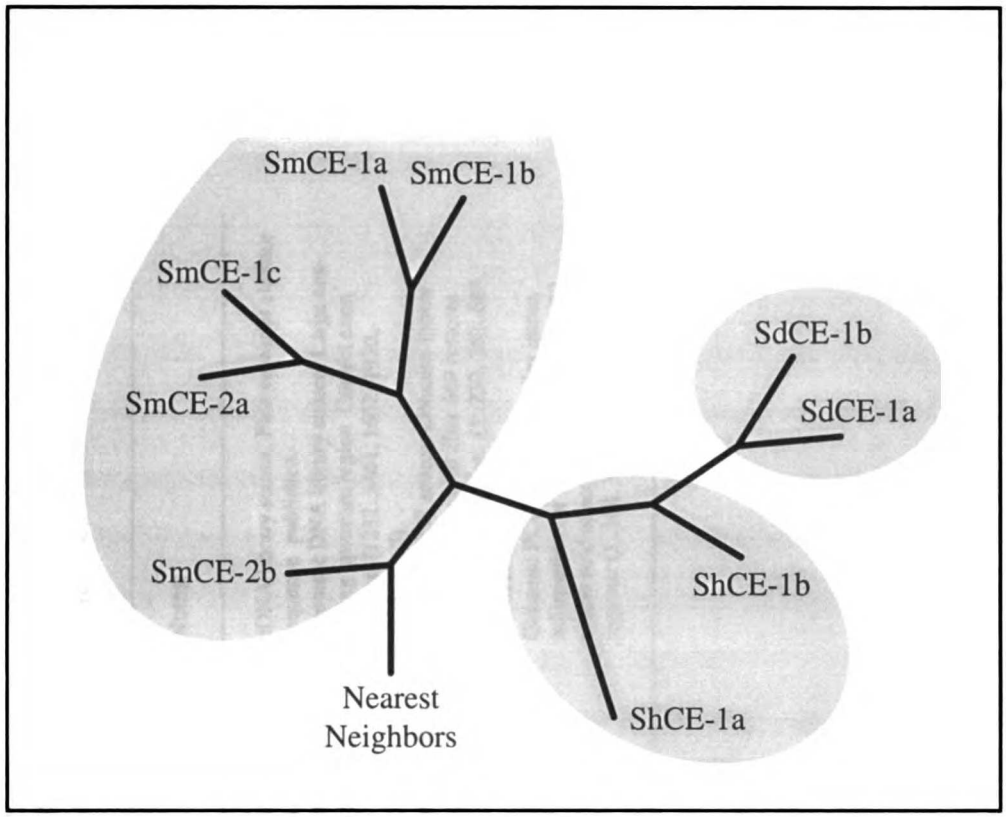


Figure 2.8. A phylogenetic plot of all cercarial elastase isoforms identified to date

Shaded ovals encompass the three species of schistosome examined. Nearest neighbors branch represents the relationship of closely related proteases from other species.

Table 2.1
Cercarial elastase sequences in public databases

New Nomenclature	Previous Nomenclature	DNA Accession Number	Reference	Notes:
CE-1a	Cercarial elastase	J03946	J. Biol. Chem. 263 (26), 13179-13184 (1988)	cDNA library source. First cercarial elastase sequence published.
CE-1b	EL1	U31768	Mol. Biochem. Parasitol. 75 (1), 113-117 (1995)	Genomic DNA library source. Large non-coding upstream region. Correct exon regions: (1232..1461, 1602..1920, 2525..2767)
CE-1c	EL2	Z70296	Mol. Biochem. Parasitol. 75 (1), 113-117 (1995)	Genomic PCR source. Contains intron structure. AT added after 669 restores correct exon regions: (1..230, 362..680, 935..1165)
CE-1a	HP1	U31769	Submitted to database but unpublished	Genomic PCR source, contains intron information for CE-1a, missing first 22 amino acid of sequence. Correct exon regions: (1..168, 310..627, 1235..1477)

Chapter 3: Predicting extracellular matrix targets of a parasite "invasion" protease by chemical and phage display mapping of active site specificity

3.1. Introduction

Aquatic larvae (cercariae) of the trematode parasite *Schistosoma mansoni* rapidly penetrate human skin. This process requires the release of a parasite protease (cercarial elastase) to degrade extracellular matrix and basement membrane proteins [37]. Host macromolecules representing barriers to cercarial invasion are known to be cleaved by proteolytic activity present in cercarial secretions. These include elastin [42], chondromucoprotein [28], keratin [43], fibronectin, laminin, and collagen IV and VIII [44].

Cercarial elastase has been shown to efficiently cleave macromolecular substrates yet slowly hydrolyzes small synthetic peptide substrates [38, 45]. Previous work hinted that cercarial elastase may have extended binding sites on both the prime (amino terminal) and non-prime (carboxy terminal) side of peptide bond cleavage [39, 40]. To confirm these data a substrate phage library was used to map selectivity at two prime side positions (S1' and S2') and one non-prime position (S3). Two positional scanning–synthetic combinatorial libraries were used to map specificity at S4 through S1.

Substrate specificity predicted by a combination of phage-display and chemical library data demonstrates that selectivity extends from P4 to P2'. A database of human proteins was then screened for peptides that contained the optimal substrate sequence. A number of epidermal extracellular matrix, and basement membrane proteins found in skin contained this consensus sequence. Direct biochemical assays of cercarial elastase versus several of these proteins, coupled with observations of skin damage following exposure to purified cercarial elastase, allowed testing of predictions from active site specificity mapping.

3.2. Experimental procedures

3.2.1. Substrate phage library generation

M13 phage with both histidine tags and randomized cleavage sites engineered into the pIII-substrate fusion protein were prepared as described previously [58]. The plasmid containing the pIII gene was randomized using two primers that amplified the complete DNA molecule. An overlapping restriction site (SphI) was used to circularize the PCR products.

PCR was performed using 0.1 μg of pBSHisGIII.1 plasmid [58] as template and the primer set (5'-CATGGGCATGCAGGA WCCNNSCCGCTGNNSNNS GCAGGGCCCGGAGGCGGT CCATTCGGTTTGAA and 5'-CCTGCATGCCCATGATGATGATG). Thermocycling conditions where 15 sec @ 94° C, 30 sec @ 65° C, 4 min @ 72° C, for 25 cycles. eLONGase polymerase (Gibco BRL) was used in 10 identical 100 μl reactions that were pooled after thermocycling yielding approximately 10 μg of PCR product. The DNA was purified with 4 spin columns (Qiagen) and digested overnight with SphI. The DNA was purified again with spin columns and diluted to a concentration <10 μg/ml for ligation. Proper dilution is absolutely necessary to favor recircularization. Ligation was performed at 14° C overnight. A final purification step with a spin column was used to remove the ligation buffer and enzymes. The DNA was lyophilized and resuspended in 10 μl water to concentrate.

The plasmid DNA library was electroporated into high efficiency eletrocompetent Epicurian Coli® XL1-Blue MRF9 cells (Stratagene). A small portion of the transformed cells were used to determine the total number of transformants by growth on LB plates containing 60 μg/ml ampicillin (LB-AMP). The yield for five pooled electroporations was $5x10^6$ individual clones that provided over 99% completeness of the library. The electroporated cells were allowed to grow at 37° C with shaking in LB containing 100

μg/ml ampicillin to an A600 of 0.25 and were then infected with VCSM13 helper phage at a multiplicity of infection of 100 phage/cell for the production of recombinant phage. The culture was allowed to grow for 6 hours. Phage particles were precipitated by addition of polyethylene glycol-8000 to 5% and NaCl to 500 mM. Phage were resuspended in 50 mM Tris, pH 7.4, 100 mM NaCl.

3.2.2. Phage substrate library screening

Three hundred microliters of nickel(II)-nitrilotriacetic acid resin (Qiagen) was washed with 20 ml of activity buffer (50 mM Tris, pH 8.0, 100 mM NaCl, 0.1% Tween 20). Phage particles ($1x10^9$) were added to the washed Ni(II) resin and allowed to bind with gentle stirring by a magnetic stir bar for 2 hours (Figure 3.2.). The Ni(II) resin was then washed with 60 ml of activity buffer to remove unbound phage. Cercarial elastase (SmCE-1a) was added to a final concentration of 10 nM in a volume of 5 ml and incubated for 2 hours. The cleaved phage were separated from the resin with a wash of 5 ml of buffer. The cleaved phage were amplified by the infection of male strain XL1-Blue MRF9 cells with the addition of VCSM13 helper phage to form recombinant phage which were then used for the next round of cleavage selection. After four rounds of cleavage selection, XL1-Blue MRF9 cells were infected with the eluted phage and plated onto LB-AMP. Fifty-one individual colonies were selected and grown in 5 ml of LB-AMP, and plasmid DNA was isolated (Qiagen) and sequenced across the cleavage site.

3.2.3. Positional scanning–synthetic combinatorial library screening

Production of the diverse chemical libraries have been described in detail previously [52]. Position P1 was scanned using the P1 diverse library. Positions P4 through P2 were scanned using the P1-Leu library. 1 μg of cercarial elastase was used per well with a final sublibrary substrate concentration of 0.1 μM for the P1-Leu library and 0.04 μM for the P1-diverse library. Reactions were started with the addition of enzyme

and monitored with a Perkin–Elmer LS50B luminescence spectrometer. Excitation was 380 nm, emission was 450 nm or 460 nm, and was monitored for 1 hour. Reactions were performed at 25° C in reaction assay buffer (50 mM Tris at pH 8.0, 100 mM NaCl, and 0.01% Tween-20). The P1 diverse library was assayed with an equal mixture of isoforms SmCE-1a and SmCE-1b due to limited enzyme availability. The P1 position results were separately confirmed with individual substrates for both enzymes (data not shown).

3.2.4. Motif searching and target identification

Proteins that contained the optimal cercarial elastase recognition sequence were identified using the web site http://expasy.cbr.nrc.ca/cgi-bin/sprot-search-ac. Two patterns were use to survey the degree of specificity for a strict and liberal interpretation of the optimal clip site. The search patterns used were [ST]-X-P-[LFMY]-X-[LG]) and [ST]-[FGLVW]-P-[LFM]-[FGLVWY]-L. The syntax has possible multiple amino acids at any one position enclosed in brackets "[ ]" and individual amino acids are separated by dashes " - ". Results were sorted for their biologic relevance (tissue localization and extracellular expression). All proteins known to be structural components of the skin were included in Table 3.1. A large number of protocadherins also contained potential cleavage sites and are listed in Table 3.2.

3.2.5. Digestion of human skin proteins and skin sections

Pure proteins predicted by database search to be biological substrates (elastin, laminin, and fibronectin (Sigma)) were resuspended in 0.1 M glycine pH 9.0 at a concentration of 1 mg/ml. Cercarial elastase was added (1 μg) to 100 μl of protein solution and incubated at room temperature for 6 hours. Digested protein samples were separated using NU-PAGE gradient gels according to the manufactures instructions (Invitrogen), or in the case of elastin, sequenced (see 1.3.14). Molecular weights of the digested fragments were compared against the sizes of predicted fragments.

Human skin samples were prepared as frozen sections and placed onto glass microscope slides. Cercarial elastase (0.1 μg) in 100 μl of digestion buffer (50 mM Tris at pH 8.0, 100 mM NaCl, and 0.01% Tween-20) was placed on top of the sample and incubated at 37° C overnight. Slides were fixed in methanol and stained with hematoxylin and eosin.

3.2.6. Microscopy of cercarial invasion

Approximately 3000 cercariae in 3 ml of water were applied, using 15 mm plastic cylinders, to human skin samples. After 120 minutes the skin was fixed in 10% phosphate-buffered formalin, embedded in paraffin, sectioned at 7 μm, and stained with hematoxylin and eosin.

3.3. Results

3.3.1. Phage display reveals P3, P1', and P2' substrate selectivity

A phage library was designed to explore P3', P1, and P2 selectivity of cercarial elastase. A PCR strategy was used to efficiently generate the library and is summarized in Figure 3.1. Using the information obtained from the chemical substrate libraries a randomized sequence was inserted in between a histidine tag attached to the PIII gene of the M13 phage (S/T X P L X X). Proline and leucine, found to be optimal by chemical libraries, were chosen to fix the register of the enzyme at P2 and P1 respectively. P4 contained either serine or threonine, again from predictions of the chemical libraries, and P3, P1', and P2' were completely randomized (Figure 3.1.A). Four rounds of library selection were performed using the most abundant of the two isoenzymes (CE-1a). Fifty-one clones were sequenced and the data combined with chemical substrate library data creating a picture of amino acid selectivity for six positions (Figure 3.3.).

Position P2' is highly selective. Leucine is the preferred amino acid at this position with glycine having a weak interaction. The chi test value for this distribution is highly significant (0.000003). P1' does not have a distribution that is significantly different from a random distribution but inspection of the most common amino acids reveals a strong bias toward bulky and hydrophobic amino acids. This suggests that there is selectivity at P1' for a class of amino acids but not an individual amino acid.

Out of 51 clones 28 contained serine and 23 contained threonine at the P4 position suggesting little preference for one amino acid over the other. Additionally, amino acids selective for P3 were equally represented when serine or threonine was in P4 suggesting selectivity for P3 and P4 are independent (data not shown).

3.3.2. Extracellular matrix and epidermal cell proteins contain optimal substrate motifs

Characterization of binding site preference for P4-P2' allowed prediction of consensus sequences for cercarial elastase substrate preference. The SWISS-PROT and TrEMBL databases were searched with these sequences to identify potential biological substrates. Proteins expressed in the skin through which the cercariae migrate were identified (Table 3.1.). A degree of selectivity for each clip site is indicated. It was calculated by taking the product of each amino acid's positional preference and adjusting it to a percentage of the maximum possible score.

Given the uncertainty at positions P3 and P1' different search parameters were used to query the database. The most general search pattern used ([ST]-X-P-[LFMY]-X-[LG]) yielded 1,970 human proteins whereas the most restrictive pattern ([ST]-[FGLVW]-P-[LFM]-[FGLVWY]-L) returned 114 human hits out of a search space of >50,000 proteins.

3.3.3. Cleavage of pure skin proteins

All three of the proteins putative biologic substrates tested, elastin, fibronectin, and laminin are cleaved by cercarial elastase. The fragment sizes are consistent with cleavage at the predicted sites (data not shown). Elastin cleavage site is shown in Figure 1.8.

3.3.4. Microscopy of cercarial invasion

Figure 3.4 shows that following invasion by cercariae the epidermal cell layer is disrupted by dissolution of cell-cell contacts and some loss of cells themselves. The epidermal-dermal basement membrane is also degraded. These results are consistent with other published data from both electron microscopic and light microscopic studies [7, 22, 49]. Direct effects of cercarial elastase alone could be visualized on the extracellular

matrix (Figure 3.4). Total loss of matrix proteins with the exception of type I collagen was observed.

3.4. Discussion

Identifying the biologic targets of a novel protease remains a largely unmet goal of biochemistry. The proteolytic mechanism of schistosome cercarial invasion represents a unique opportunity to test integration of diverse chemical libraries, phage display, and algorithm search technologies. One can combine microscopic and ultra-structural observations of invasion, knowledge of the structure and composition of mammalian skin, and protease inhibition studies, to validate information on subsite selectivity from P4 to P2' (Figure 3.4).

The demonstration of selectivity across six subsites confirms previous results suggesting that the substrate binding site of cercarial elastase is quite extended [39, 40]. This may explain why cercarial elastase can degrade macromolecular substrates even those like elastin that are generally resistant to proteolysis. While cercarial elastase and chymotrypsin share P1 specificity, chymotrypsin degrades small peptide substrates synthetic peptide substrates much more efficiently than cercarial elastase while only cercarial elastase degrades elastin.

The subsite amino acid preferences for cercarial elastase provides a consensus sequence to screen human protein databases for putative biological substrates. Proteins containing these sequences were examined for consistency with knowledge about the timing, location, and effects of cercarial invasion.

Elastin, laminin, and fibronectin were all predicted to be substrates and were cleaved by cercarial elastase *in vitro*. Type I collagen, previously shown *in vitro* to be resistant to cercarial elastase cleavage, did not show up as a predicted substrate [31]. As a more direct biological assay, skin sections were digested with cercarial elastase. Extracellular matrix proteins were completely degraded except for type I collagen, again as predicted by the search results. The epidermal cell layer and basement membrane were also observed to be lost around the invading cercariae. This is consistent with degradation

of type IV collagen and laminin in the basement membrane and cadherins that link epidermal cells.

3.5. Figures and tables for chapter 3

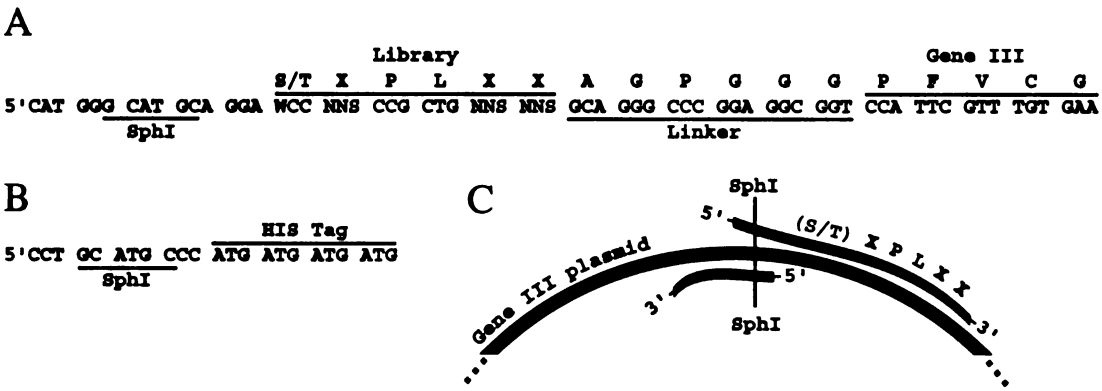


Figure 3.1. Primer design and PCR strategy for phage substrate library construction

(A) Randomized forward primer overlaps gene III, inserts a randomized sequence, a linker, and creates a unique restriction site (SphI) for circularization.
(B) Reverse primer creates same unique restriction site (SphI) and is homologous to gene III expression plasmid.
(C) Cartoon of the two primer's alignment and orientation to the gene III expression plasmid. SphI restriction site overlap for PCR product circularization is indicated with vertical bar.

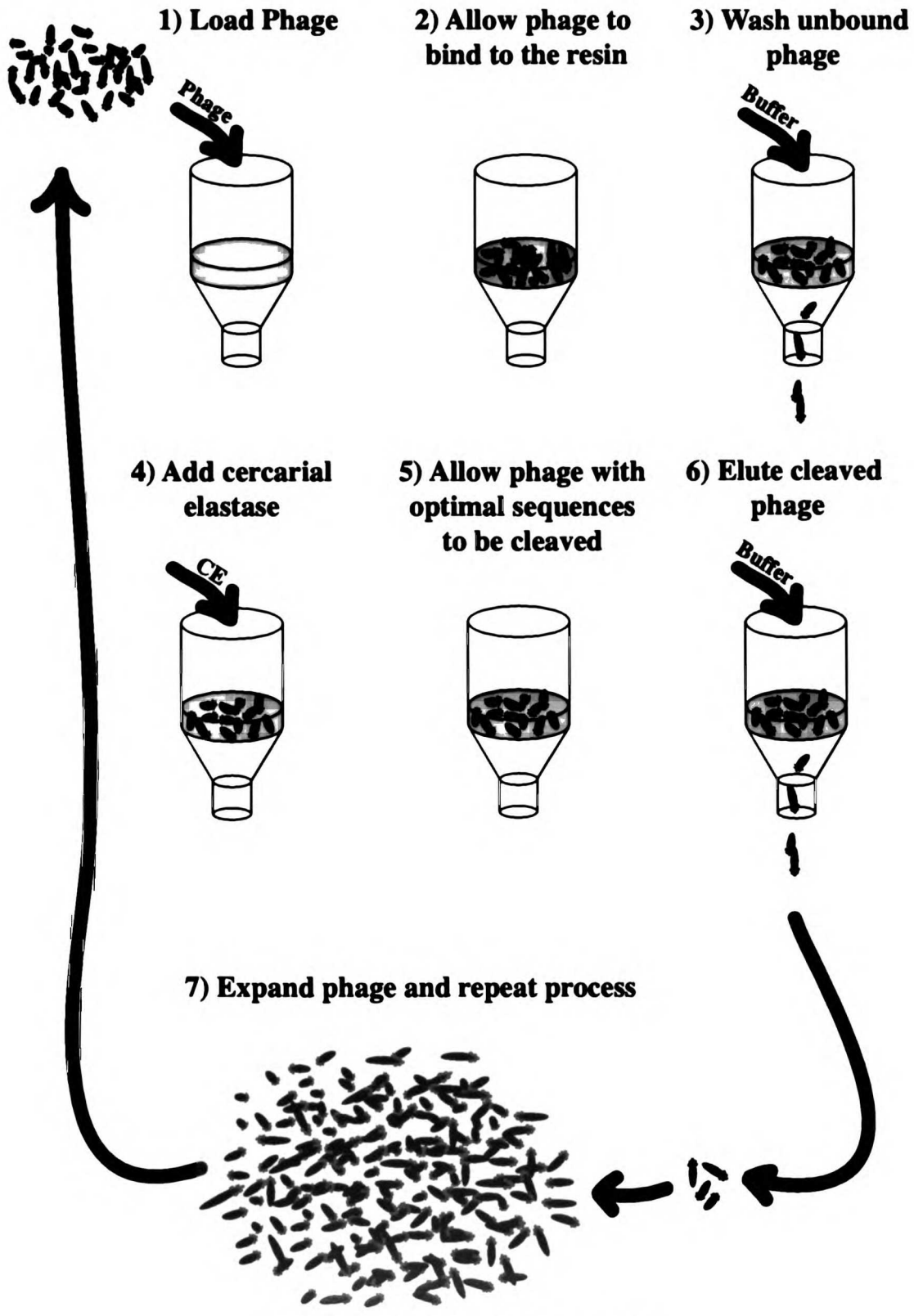
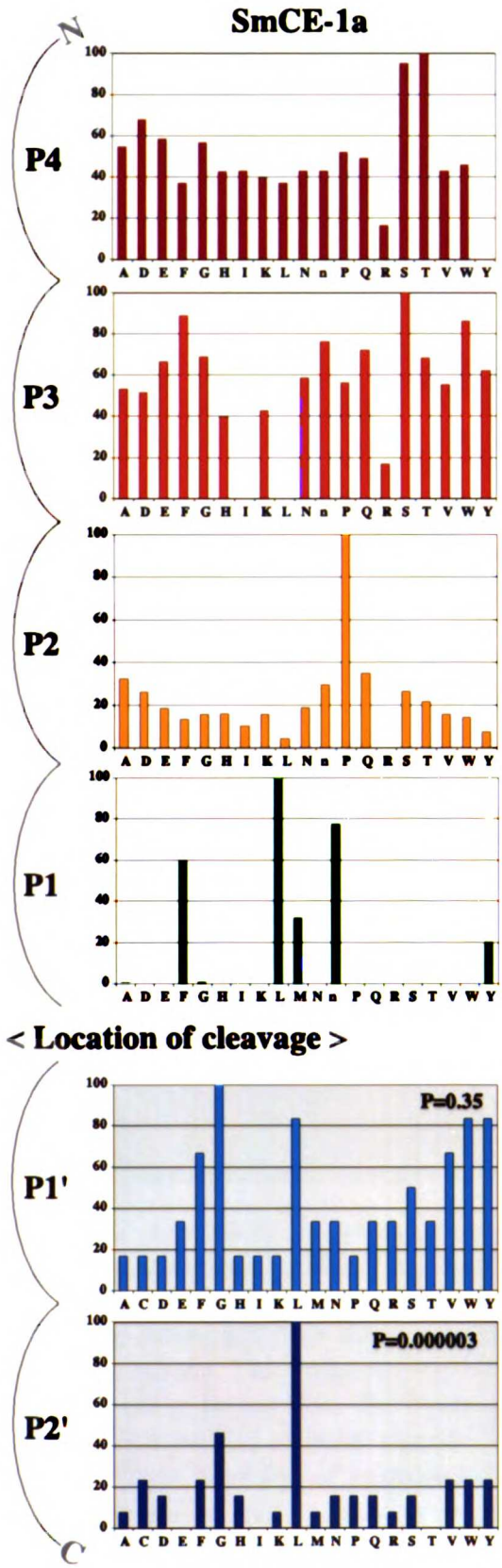


Figure 3.2. Cycle of phage selection



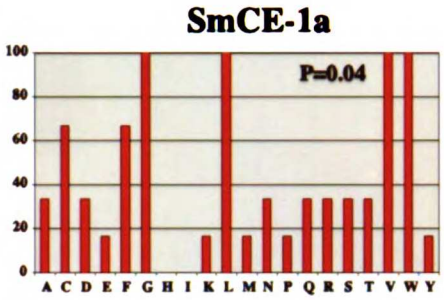


Figure 3.3. Amino acid preferences using combinatorial chemical and phage substrate libraries

Bars indicate normalized relative preferences for each amino acid at a given substrate position. Six substrate positions (4 upstream and 2 downstream of the peptide bond cleavage) are indicated on the left, using the prime / non-prime nomenclature (P4-P2'). Single letter amino acid coding is used with "n" representing norleucine. Results generated using chemical libraries are indicated by histograms with clear backgrounds. Results generated using a phage library, randomized positions (S/T X P L X X), are indicated by histograms with a shaded background. Phage library results were generated from 51 sequenced clones. Chi test values are displayed in the upper right corner for the phage display data.

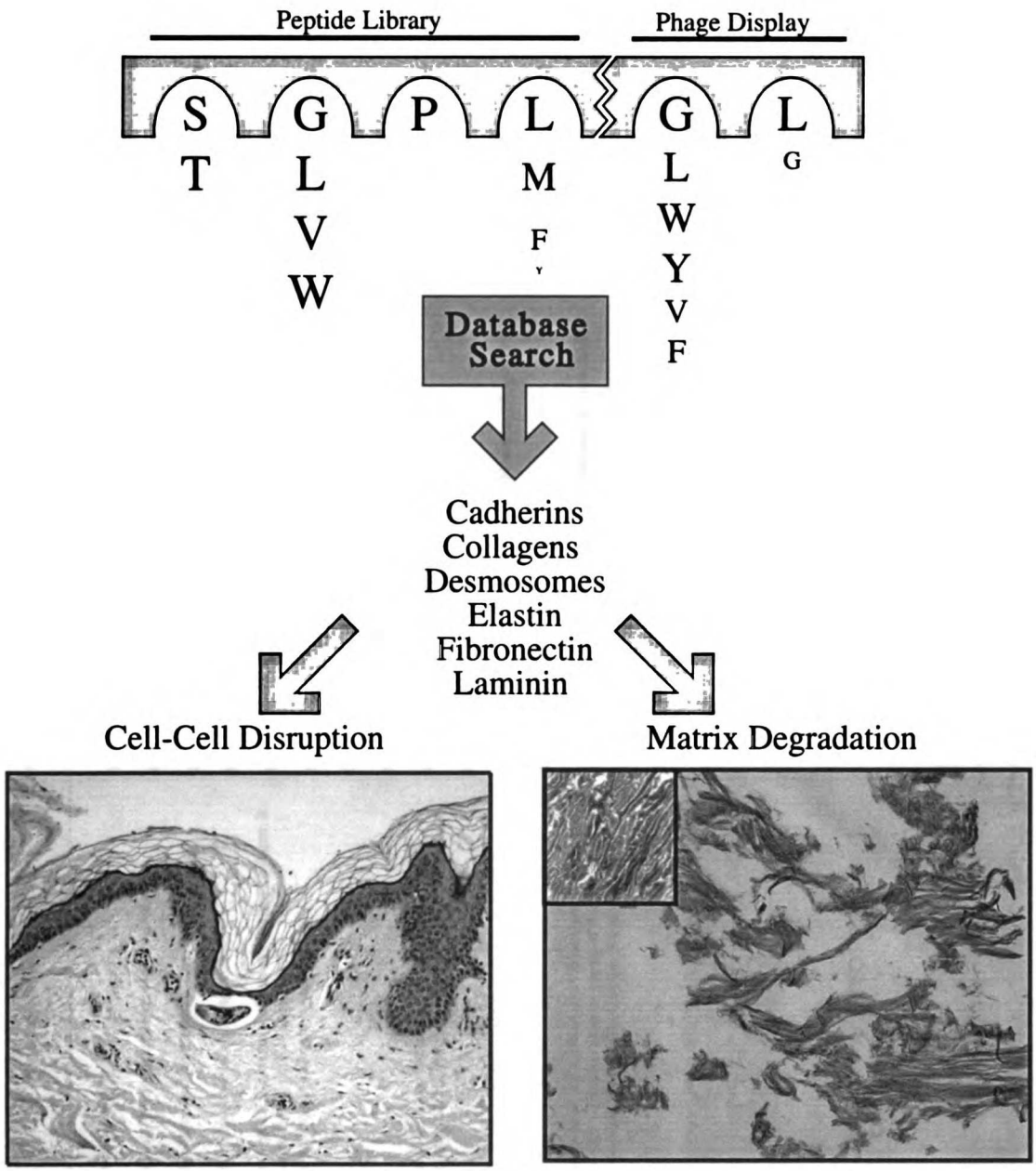


Figure 3.4. Continuity between subsite selectivity determination and macroscopic observations of cercarial invasion and cercarial elastase degradation

Single letter amino acid code size is proportional to the relative preference of that amino acid at that subsite. The image on the left shows a cercaria at the epidermal / dermal junction invading human skin. The figure on the right shows a section of human dermis incubated with purified cercarial elastase. The inset image is a control section incubated with buffer only. Note loss of epidermal cell layer (dark gray) and basement membrane around cercaria on left, consistent with degradation of connective and basement membrane proteins. Note loss of all matrix components except type I collagen on right.

Table 3.1. Epidermal and dermal extracellular matrix and basement membrane proteins containing potential cleavage sites

Accession Number[a]	Protein Name	Protein Description	Cut #	Location	Sequence	Domain of predicted cleavage	Positional amino acid preference[b]					% Score[c]
							P4	P3	P1	P1'	P2'	
Q03001	Bullous Pemphigoid Antigen 1	A hemidesmosomal protein.	1	340	TAPLKL		100	25	100	25	100	6.3
Q14126	Desmoglein 2	Component of intercellular desmosome junctions.	1	998	SNPLEG	Desmoglein Repeat 5	95	25	100	25	46	2.7
P53420	Collagen Alpha 4 (IV)	Forms the framework of basement membranes.	1	360	TPPLPL		100	25	100	25	100	6.3
P12111	Collagen Alpha 3 (VI)	Collagen VI acts as a cell-binding protein.	1	100	TEPLAL	Nonhelical Region	100	25	100	25	100	6.3
P20849	Collagen Alpha 1 (IX)	Collagen type IX is a minor cartilage nonfibrillar collagen. It is associated with type II collagen fibrils.	1	230	SVPFEL	Nonhelical Region 4	95	100	50	25	100	11.9
Q99715	Collagen Alpha 1 (XII)	Interacts with type I collagen containing fibrils.	1	710	SIPLAG	Fibronectin Type-III	95	25	100	25	46	2.7
			2	1166	SSPLVG	Fibronectin Type-III	95	25	100	75	46	8.2
			3	1463	SEPLKG	Fibronectin Type-III	95	25	100	25	46	2.7
O00261	Collagen Type (XIV)	Distributed in tissues containing banded type I collagen fibrils. (Lethias et al., 1993; Oh et al., 1993).	1	288	SDPLVG		95	25	100	75	46	8.2
			2	385	SAPFQL		95	25	50	50	100	5.9
P39060	Collagen Alpha 1 (XVIII)	Type XVIII collagen have been reclassified as heparan sulfate proteoglycans (HSPGs) and located in the basement membrane.	1	462	TPPLAG	Nonhelical Region #1	100	25	100	25	46	2.9
			2	385	SAPFQL		95	25	50	50	100	5.9
P02751	Fibronectin	Fibronectins bind cell surfaces and various compounds. The "extra domain" is present in some forms of fibronectin and absent in others. These differences are due to alternative splicing.	1	1347	STPLRG	Fibronectin Type-III #8	95	25	100	50	46	5.5
			2	1711	SQPLIG	Fibronectin Type-Iii #12 (Extra Domain 2)	95	50	100	25	46	5.5
			3	1982	SEPLIG	Fibronectin Type-III #15	95	25	100	25	46	2.7
P98095	Fibulin-2	Binds to fibronectin and other ligands. Component of both basement membranes and other connective tissues.	1	672	TIPLPL	EGF-Like 1, Calcium-Binding	100	25	100	25	100	6.3
P24043	Laminin Alpha-2	Primarily a mesenchymal laminin subunit. (Leivo and Engvall, 1988; Pauls-son et al., 1991).	1	3081	SIPFRG	Laminin G-Like #5	95	25	50	50	46	2.7
Q16363	Laminin Alpha-4	Endothelial or mesenchymal origin.	1	1399	SSPLFL	Laminin G-Like #3	95	25	100	75	100	17.8
			2	1779	SKPFTG	Laminin G-Like #5	95	25	50	50	46	2.7
Q9UHI2	Laminin Beta 1 Related Protein	Predominates in the sarcolemmal basement membrane.	1	684	SQPLQG		95	50	100	50	46	10.9
Q92675	Melanoma-Associated Chondroitin Sulfate Proteoglycan		1	163	SRPLRG		95	50	100	50	46	10.9
			2	1838	SVPLRL		95	100	100	50	100	47.5

[a] Accession numbers are for the SWISS-PROT and TrEMBL databases and can be accessed at the web site http://www.expasy.ch/sprot/

[b] Nomenclaure is that of XXX. Values are derived from figure XX and represent the enzyme's relative affinity for the substrate's amino acid at the particular position

[c] Score values are the product of the individual amino acid affinities divided by the maximum possible value

Table 3.2. Protocadherins containing potential cleavage sites

Accession #	Protein Name	Cut #	Location	Sequence	Positional amino acid preference					Score
					P4	P3	P1	P1'	P2'	
Q9NRT9	PROTOCADHERIN 10	1	293	SPPLLG	95	25	100	50	46	5.5
Q9NPG4	PROTOCADHERIN 12	1	613	SRPFLL	95	50	50	50	100	11.9
O14917	PROTOCADHERIN 68	1	409	SVPFKL	95	100	50	25	100	11.9
Q9Y5I1	PROTOCADHERIN ALPHA 11	1	326	TPPMAG	100	25	50	25	46	1.4
		2	774	SLPLGL	95	25	100	100	100	23.8
Q9Y5H7	PROTOCADHERIN ALPHA	1	326	TFPLSG	100	50	100	25	46	5.8
Q9Y5I4	PROTOCADHERIN ALPHA C2.	1	407	TLPFRL	100	25	50	50	100	6.3
Q9NYQ7	PROTOCADHERIN FLAMINGO 1	1	1668	TGPLLL	100	50	100	50	100	25.0
Q9NYQ6	PROTOCADHERIN FLAMINGO 2	1	1595	TGPLLL	100	50	100	50	100	25.0
Q9Y5D7	PROTOCADHERIN GAMMA A12	1	788	SEPLLL	95	25	100	50	100	11.9
Q9Y5G7	PROTOCADHERIN GAMMA A6	1	389	SLPFEL	95	25	50	25	100	3.0
Q9Y5F7	PROTOCADHERIN GAMMA C4	1	792	SAPMAG	95	25	50	25	46	1.4

[a] Accession numbers are for the SWISS-PROT and TrEMBL databases and can be accessed at the web site http://www.expasy.ch/sprot/

[b] Nomenclaure is that of XXX. Values are derived from figure XX and represent the enzyme's relative affinity for the substrate's amino acid at the particular position

[c] Score values are the product of the individual amino acid affinities divided by the maximum possible value

Chapter 4: Anti-cercarial elastase antibodies as a marker of exposure to schistosome cercariae

4.1 Introduction

Schistosome infection of a human host begins with invasion of intact skin by an aquatic larva, the cercaria(ae) [41]. Cercarial elastase is one of the first proteins released in the host during invasion and therefore one of first proteins the immune system "sees". Cercarial elastase has previously been shown to be immunogenic, eliciting both an IgM and IgG immune response in humans [59]. The enzyme is only expressed during the cercarial lifecycle stage thus restricting immune exposure to the period of invasion[32].

Previous attempts to measure human exposure to cercariae required both a "cercariometric" method to determine cercarial concentrations in water and a quantitative measure of human water contact [60, 61]. Limitations and errors in direct observation of human contact with infected waters, and therefore difficulty in quantifying exposure presented considerable difficulties in analysis of risk of populations to infection in endemic areas and difficulties assessing efficacy of vaccine or drug programs. In collaboration with Dr. Reda Ramzy (Professor, Ain Shams University) an ELISA was developed using purified native cercarial elastase as antigen. I spent one month in Cairo working with Dr. Ramzy's laboratory on developing this assay and provided purified antigen for these studies. This assay was used to test the hypothesis that serum reactivity to cercarial elastase would positively correlate with exposure to cercariae [62].

Results of using the ELISA in endemic villages indicate there is a positive correlation between disease incidence and cercarial elastase antibodies consistent with the ELISA being a measure of exposure. To increase consistency and reduce costs of antigen production, recombinant cercarial elastase was tested as a substitute for native cercarial elastase.

4.2 Methods

4.2.1. Human subjects

Endemic sera: Blood was collected from 306 schoolchildren living in 4 separate villages endemic for schistosomiasis annually for two years. Urine and stool samples were collected from all subjects and examined for the presence of schistosome eggs. The project was approved by the Ethical Review Board of the Faculty of Medicine of Ain Shams University. Informed consent for participation in the study was obtained and persons with positive urine or stool examinations results were referred to Kafr Shorafa, Ministry of Health (MOH) Heath Center for treatment.

Nonendemic control sera: Sera was collected from 101 people living in Cairo with no history of residence in rural areas and no exposure to cercariae.

4.2.2. Native protein purification

Cercarial elastase was purified as previously described in sections 2.3.1 and 2.3.2. All fractions from the ion exchange column were pooled, inactivated with 20 μM of Suc-Ala-Ala-Pro-Phe-chloromethyl ketone, and then lyophilized.

4.2.3. Parasitological methods

Schistosoma haematobium: Urine samples (10ml) collected early in the morning were centrifuged. Sediments were microscopically examined and eggs counted.

Schistosoma mansoni: Fecal egg counts were determined by the modified Ritchie formol-ether concentration technique [63].

4.2.4. Recombinant protein generation

Expression plasmid -- The *E. coli* expression construct pET-SmCE-1a was assembled by inserting the active portion of SmCE-1a, amplified by PCR, into the

restriction sites NdeI (blunted) and XhoI of the vector pET21-a. The forward primer 5'-TGCGTAGTGGTGAACCTGTGCA-3' created a 5' end in frame with a blunted NdeI site in the vector. This arrangement also changed the 1st amino acid to a start codon (I->M). The resulting protein an inactive protein. The reverse primer 5'-CAGATCCTCGAGAATATTGGAGCGTACAAAATCC-3' added an XhoI site (underlined) and removed the stop codon creating an open reading frame into the pET21-a vector that added a 6X histidine tag to the C-terminus of the protein.

E. coli -- BL21 cells were electroporated with the plasmid construct and selected overnight using LB 50 μg/ml ampicillin plates. A single colony was picked and grown overnight in 4 ml of LB with 100 μg/ml ampicillin. 1 ml of overnight culture was added to 1000 ml LB with 100 μg/ml ampicillin and grown to an OD600 = 0.6. IPTG was then added to100 μM and after four hours of continued incubation the cells were pelleted and frozen at -70° C.

Ni-ATA affinity chromatography -- E. coli pellets from 330 ml of inducted culture were thawed and resuspended in 5 ml of binding buffer (8 M urea, 50 mM Tris pH 8.0). Cell disruption was completed with 6 rounds of 20 second sonications on ice. Histidine tagged proteins were then purified according to the manufacturers instructions (Qiagen, Valencia, CA) using 1 ml spin columns. Pooled protein was dialyzed into 100 mM glycine pH 10.7 overnight, then pH 10 for 2 hours, then pH 9 for 2 hours using a slide-A-lyzer dialysis cassettes (Pierce). Dialyzed protein was then lyophilized for stable storage.

4.2.5. Serologic Tests

Cercarial elastase antigen (250 ng of native or recombinant) was diluted in 50 μl of 0.06 M carbonate buffer, pH 9.6, and added to polyvinyl microtiter plates. After incubation for 3 hours at 37° C the plates were left overnight at 4° C. The plates were washed once with phosphate-buffered saline (PBS), and then blocked with 0.01 M PBS pH 7.4, 0.05% Tween 20 (v/v), 5% fetal calf serum (BLOCK) for 1 hour at 37° C. Sera

diluted 1:100 in BLOCK were incubated in duplicate wells for 2 hours at 37° C. Plates were washed again four times with BLOCK. Peroxidase-conjugated goat anti-human IgM or IgG antibody diluted 1:20000 in BLOCK was added to the wells for 1 hour at room temperature. Plates were washed again four times with BLOCK. Antibody was detected using the substrate *o*-phenylenediamine with H_2O_2. The reactions was carried out at room temperature and stopped after 15 min with 4 M H_2SO_4. Optical density was read and compared to a blank at 490 nm with an ELISA reader. Data were normalized to a scale of 0-100 based on the difference between the positive and negative controls' reactivity.

4.2.6. Western blot analysis

Recombinant cercarial elastase (1 μg/well) was run into 10% SDS-PAGE gels and transferred to nitrocellulose using standard methods. The blot was incubated with BLOCK overnight at 4° C. The blot was rinsed with PBS and sera diluted in BLOCK was added for 2 hours at room temperature. Three rounds of washing in PBS + 0.05% Tween 20 (PBST) were then performed. Peroxidase-conjugated goat anti-human antibody diluted 1:10000 in BLOCK was added to the membrane for 1 hour at room temperature. Blots were washed again four times with TBST. Antibody was detected using the substrate 3, 3'-diaminobenzidine. The reactions was carried out at room temperature and stopped after 15 min by water rinses.

4.3. Results

4.3.1. Native cercarial elastase ELISA

The prevalence of schistosome infection in schoolchildren was determined annually for two years (Table 4.1.). These same subjects were screened using the ELISA developed with native cercarial elastase (Table 4.2.). A positive correlation (0.68) was

observed between the change in level of infection among these schoolchildren and the change in reactivity to cercarial elastase (Figure 4.1.)

4.3.2. Failure of recombinant cercarial elastase to substitute for native protein in the ELISA

ELISA plates were prepared with recombinant cercarial elastase. Pooled sera from both positive and negative populations were assayed. Negative sera produced a signal in equal strength to that of positive sera. Replacement of standard reagents and variations in the binding conditions of the antigen did not improve the signal to noise ratio.

4.3.3. Western blot using recombinant cercarial elastase and pooled sera

A Western blot of recombinant cercarial elastase was prepared and probed with pooled positive control sera (Figure 4.2.). The blot demonstrated strong reactivity against recombinant cercarial elastase. Additional higher molecular weight bands were observed indicating that recombinant protein may be forming multimers that do not dissociate during SDS-PAGE.

4.3.4. Western blot using recombinant cercarial elastase and individual sera

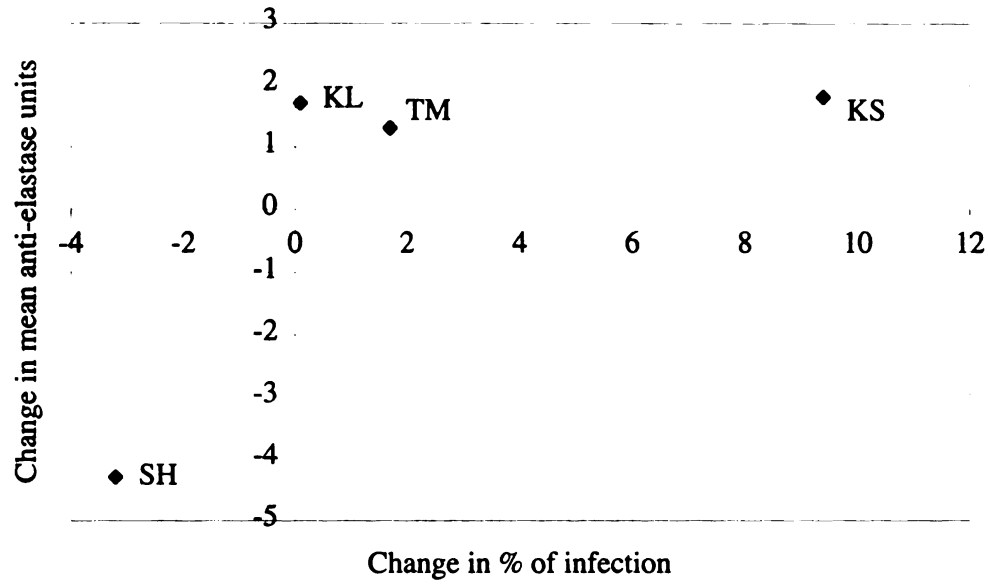
A western blot of recombinant cercarial elastase was prepared. Strips from this blot were probed with sera from individuals with confirmed schistosome infections and with sera from unexposed individuals (Figure 4.3.). The blot demonstrated stronger reactivity against recombinant cercarial elastase with sera from exposed subjects versus unexposed subjects. A high number of additional bands on each strip suggests that recombinant cercarial elastase has not been completely purified from its *E. coli* source, and that individual serum antibodies reactive to *E. coli* contaminants.

4.4. Discussion

Native cercarial elastase inactivated by a chloromethyl keytone shows great promise as an ELISA antigen to identify individuals exposed to infectious schistosome larvae. Recombinant cercarial elastase failed to replace the native protein preparations as a source of antigen. This difference could result from a variety of technical issues. Purity of the recombinant protein could be improved to increase the anti-cercarial elastase signal above background. Recombinant enzyme was also difficult to keep in solution. Precipitation of protein could have led to incomplete coating of ELISA plates. The isoelectric point of cercarial elastase is high (>9.0) and could have interfered with binding to the charged plates. Further exploration of buffers that would stabilize the protein and provide appropriate electrostatic interactions could improve the ELISA results. A folding protocol for the recombinant protein may also be required for maximal reactivity shown with the native enzyme.

The data in Figure 4.1 are consistent with the hypothesis that reactivity can be used as a marker for transmission of the disease. Therefore native protease is now being used to complete the epidemiological study.

4.5. Figures and tables for chapter 4



Village	Change AB	Change Infect	AB Year 1	AB Year 2	Infect Year 1	Infect Year 2
KL	1.7	0.1	8	9.7	1.9	2
KS	1.8	9.4	13.3	15.1	0	9.4
SH	-4.3	-3.2	19.3	15	13.6	10.4
TH	1.3	1.7	7.6	8.9	1.6	3.3

Correlation:	0.68

Figure 4.1. Change in elastase antibody reactivity units vs. change in level of infection in villages of the Nile delta

Four villages are plotted by their change in percent of schoolchildren infected with schistosomes and their change in mean anti-cercarial elastase reactivity units. Infection was determined by microscopic identification of eggs in stool. The table above contains raw data for the graph. The correlation coefficient for the two variables is shown in the lower left-hand box.

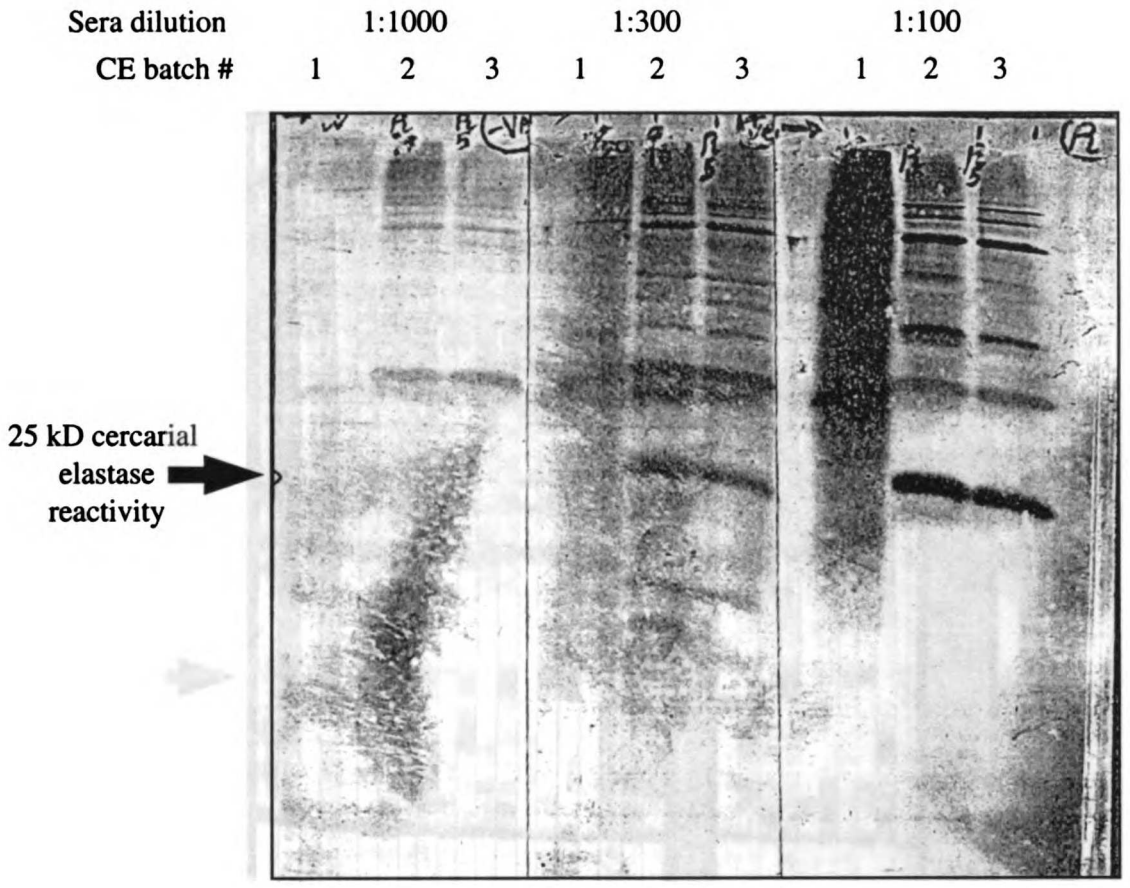


Figure 4.2. Western blot of recombinant cercarial elastase protein with pooled human sera

Three batches of recombinant cercarial elastase were tested in a Western blot assay using three dilutions of pooled human sera from patients with confirmed schistosome infections.

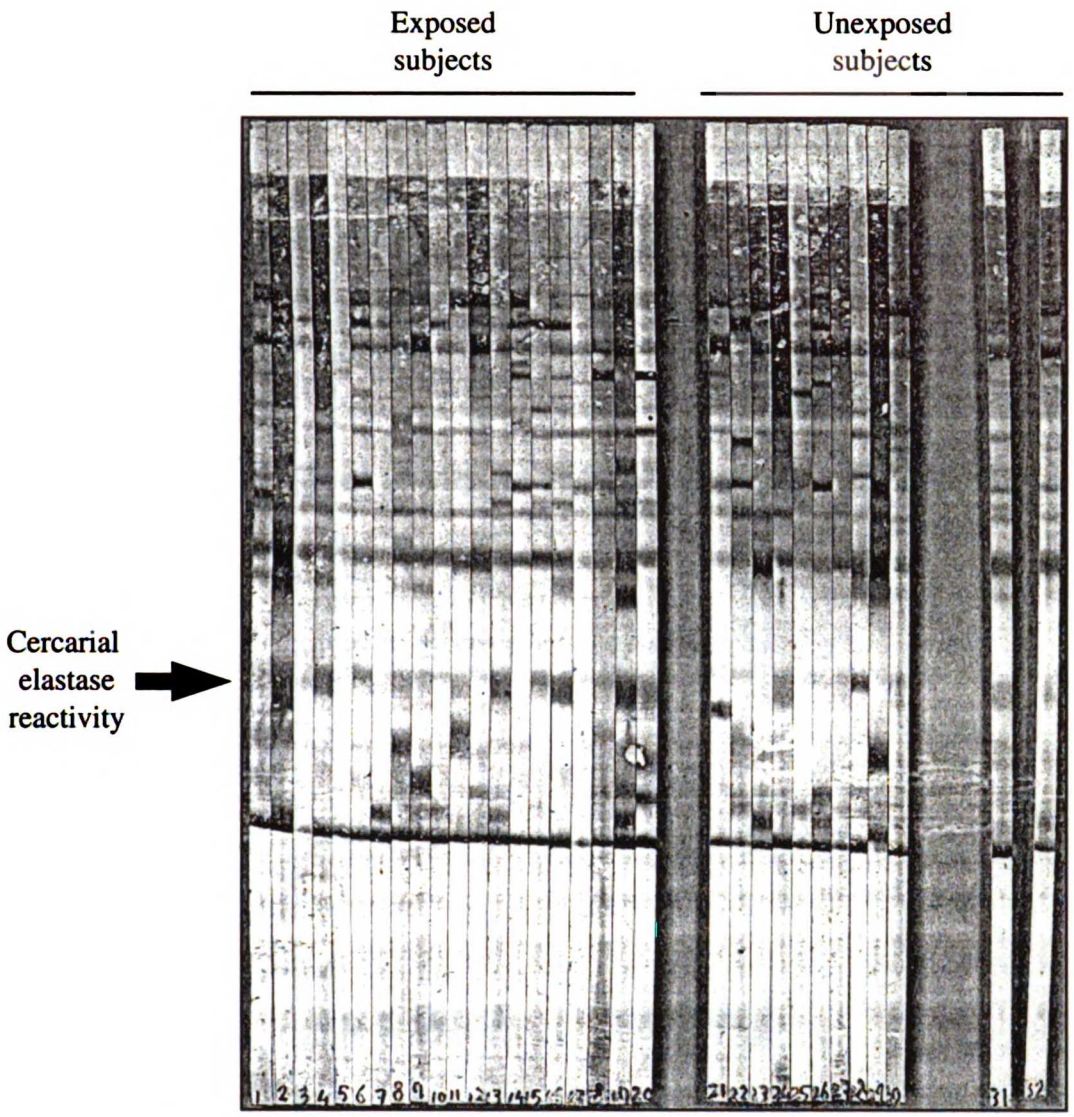


Figure 4.3. Western blot of recombinant cercarial elastase protein with individual human sera

Recombinant cercarial elastase was tested in a strip Western blot assay using sera from individuals who have had schistosome exposure and those who have not.

Table 4.1.

Prevalence rates of *S. haematobium* and *S. mansoni* infection among Tahoria schoolchildren as presented by village

Village	No. Tested		*S. haematobium* Postive		*S. mansoni* Positive	
Year 1	S.h.	S.m.	No	%	No	%
KL	52	42	1	1.9	0	0
KS	35	33	0	0.0	1	3.0
SH	140	148	19	13.6	1	0.7
TH	65	60	1	1.6	0	0.0
Total	292	283	21	7.2	2	0.7
Year 2						
KL	50	51	1	2.0	1	1.9
KS	32	34	3	9.4	3	8.8
SH	163	157	17	10.4	0	0.0
TH	61	58	2	3.3	0	0
Total	306	300	23	7.5	4	1.3

Table 4.2.

Anti-cercarial elastase antibody units among Tahoria schoolchildren as presented by village

Village	No Tested	ELISA				
		Positive				
	Year 1	No	%	Mean units	SD	Media
KL	51	12	32.5	8.0	14.5	0.0
KS	36	11	30.5	13.3	19.3	4.8
SH	175	82	46.8	19.3	21.6	12.8
TH	66	13	19.7	7.6	13.9	0.0
	Year 2					
KL	51	11	21.6	9.7	16.4	1.0
KS	36	15	41.6	15.1	18.7	7.6
SH	173	63	36.4	15.0	20.5	7.4
TH	66	15	22.7	8.9	14.7	0.0

Chapter 5: Future directions

5.1. Expression of recombinant active protease

While sufficient recombinant protein can be produced for antibody production and ELISA, ten years and multiple expression systems failed to produce active recombinant cercarial elastase. *B. subtilis*, *Pichia*, *S. cervesiae*, insect cells, and *E. coli* have all been used as sources of material. Expression constructs varying the arrangement and composition of pre, pro, and active domains were also examined. Combinations of fusion proteins and epitope tags were also explored to aid in purification and activation. The gene has also been recloned and sequenced several times.

Two areas of investigation warrant further exploration. High efficiency *in vitro* transcription and translation (Rapid Translation System - RTS) and protein refolding. RTS recently been successful in express large quantities of recombinant proteins known to be difficult to produce (Roche molecular biochemicals product literature). This system also lends itself to production of cercarial elastase in the presence of reversible inhibitors thereby avoiding auto-proteolysis [31]. Exploration of an array of refolding conditions would be a second approach. *E. coli* is a very efficient system for producing large quantities of insoluble recombinant protein. This provides sufficient starting material for testing a large number of different refolding conditions. In parallel with this approach, denatured native protein could be used as an internal control. Highly specific and sensitive radio-labeled inhibitors could monitor the regeneration of activity from the denatured native protein. This strategy could guide the search for appropriate refolding conditions. Promising leads would then be applied to recombinant products. This strategy also reduces the likelihood that a bottleneck results from improperly produced recombinant protein and focuses the search on refolding conditions.

5.2. Crystallization and point mutations

When active recombinant protease production is achieved, analysis through mutagenesis and crystallization will be possible. Cercarial elastase contains 6 insertions when compared with other classical serine proteases like trypsin. The role of these insertions are unknown. Kinetic studies of mutant enzymes will also be explored. Solution of cercarial elastase's crystal structure will provide additional insight into its efficient cleavage of host macromolecules including elastin. This question is especially interesting considering that few proteases cleave elastin and the site of cleavage by cercarial elastase is unique.

5.3. Transfection and genetic manipulation of schistosomes

5.3.1. Introduction

Cercarial elastase is expressed at high levels in a stage and cell specific manner. As such it represents an ideal model to pioneer analysis of transcriptional regulation in schistosomes. Transient and stable transduction techniques for schistosomes do not exist. The successful development of these techniques will greatly accelerate and improve the efficiency of schistosome research. The development of a complete suite of transgenic techniques follows an ordered path of technical innovation, where each advance is used for the development of the next technique.

5.3.2. Promoter cloning and expression vector construction

A series of expression cassettes are required to monitor the development of new transfection techniques. A reporter construct that contains a strong promoter driving the expression of a highly sensitive reporter gene is required. Additionally, control plasmids with deletions in promoter regions are needed to provide convincing evidence of successful transfection. The choice of promoter source is not restricted to the schistosome

itself. The only requirement for these constructs is strong expression. The use of endogenous promoters does reduce the chance of incompatible transcriptional regulation between species.

5.3.3. Transient transfection of nucleic acids and promoter characterization

Two approaches to achieve transient transfection can be taken. There are independent of the mode of delivery and the lifecycle stage targeted. A DNA construct containing a promoter, reporter gene, and poly adenylation signal or a RNA transcript can be used to develop a transfection system. RNA transfection eliminates doubts about promoter function, tissue and stage specificity, and ease of synthesis. DNA constructs could easily replace RNA usage following the successful detection of the reporter gene in a RNA transfection protocol.

The major hurdle in developing a transfection system is the first detection of reporter activity. Once this has been established, optimization of the system can be pursued to increase efficiencies both in transfection methods and in expression cassette construction. Promoter bashing experiments would define the minimal number of DNA elements necessary for efficient expression.

5.3.4. Expression constructs for screening and selection of transduced schistosomes

Construction of selection and screening constructs will follow the completion of promoter bashing experiments. These will be used to determine the behavior of episomal DNA and for the generation of stable integrations of DNA: transgenic schistosomes.

Green florescent protein (GFP) expression constructs are the obvious choice for screening by passive detection of expression. While GFP works in many systems, toxicity of the protein and auto-florescence will be major concerns.

The three most common selectable markers are the neomycin resistance gene, thymidine kinase gene, and the zeocin resistance gene. In addition to building expression

constructs with these genes both the snail host and mouse hosts will need to be examined for tolerance to the corresponding drugs used for selection.

5.3.5. Directed targeting constructs for gene knock-outs with positive and negative selectable markers

The zenith of transgenic technology is the targeted insertion of DNA within the genome of an organism. These targeting constructs generally require large flanking regions of homologous genomic DNA to achieve targeted insertion. This necessitates the availability of a high quality genomic library. Additionally, these vectors require both positive and negative selectable markers. This requirement may be avoided if GFP could be used to screen large numbers schistosomes following a negative round of selection to eliminate random insertions.

5.3.6. Additional techniques for the maintenance, transfer, storage, and cross breeding of transgenic schistosomes

The generation of transgenic animals will produce many independent lines of schistosome requiring lifecycle maintenance. The technology for both asexual expansion [64, 65], single sex selection, and cryopreservation [66] have all been developed. These techniques are not commonly used and will require practice and standardization. The ability to serially passage schistosomes from one snail to another will provide an opportunity to expand and distribute transgenic schistosomes [67, 68]. The generation of single sex infections from differing transgenic lines will allow for the cross breeding of homozygous strains within the mammalian host.

5.3.7. Transfection technology developed and tested

A rapid and efficient method for promoter isolation has been used to clone three types of schistosome promoters. The method is explained in section 2.3.4. and

graphically depicted in Figure 5.1. The GAPDH promoter was chosen as a high output ubiquitous promoter. GAPDH is one of the most frequent EST's consistent with a strong promoter. As a single copy gene there is also no possibility of differential regulation from another isoform. GAPDH is also universally required and will be expressed in all cells. A genomic upstream section 1500 bp in length has been cloned and attached to three reporter genes (Figure 5.2).

Cercarial elastase is a gene that is stage and cell specific. It is highly expressed in four large cells in the cercariae and should provide a strong promoter to drive GFP expression, allowing for visual screening of free swimming cercariae. The two promoters that in combination express 90% of cercarial elastase have been cloned and attached to two reporter genes (Figure 5.2).

Actin is a multi-copy gene family (>6 members) with predicted diverse stage and tissues expression patterns for each member. Identification and characterization of these promoters would provide insight into the control and regulation of a gene family with schistosomes. Three promoters have been cloned and three more fragments have been identified (Figure 5.3). One promoter has been attached to two reporter genes (Figure 5.2).

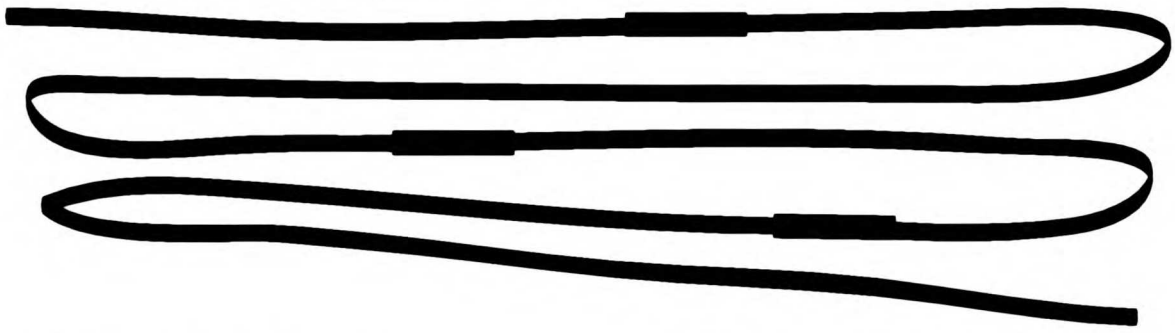
A system for generating T7 promoter linked reporter genes by PCR was also constructed. This system can produce three different reporter constructs through a single pair of primers and three different templates (Figure 5.2). PCR product was used to produce capped RNA. This RNA was shown to be functional in a cell culture (human) transfection system.

Combinations these constructs were tested using multiple transfection methods and multiple lifecycle stages. Primary cultures of disrupted sporocysts were transfected by lipofection or electroporation. Adult worms were transfected using ballistics. Infected snails were injected with a solution of lipofectamine and sealed. None of these systems produce positive signals above background. Further optimization and improved control

constructs will be necessary to optimize techniques as possible candidates for schistosome transfection.

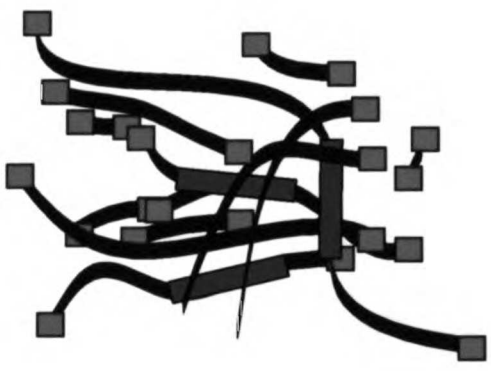
5.4. Figures for chapter 5

1) Three isoforms (green) located within genomic DNA (black)

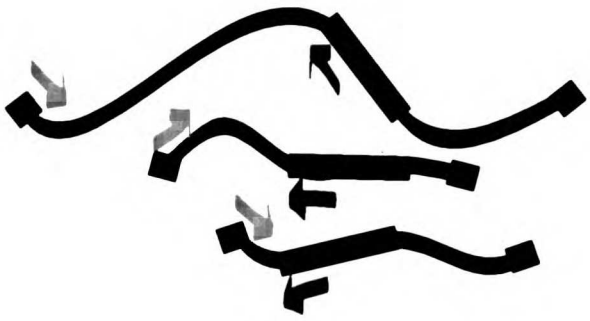


2) Genomic digestion with single restriction enzyme

3) Ligation of linker (red) to free ends of genomic DNA



4) PCR amplification with primers binding to linker (blue) and isoforms (pink)



5) DNA agarose gel of PCR products separate individual upstream regions

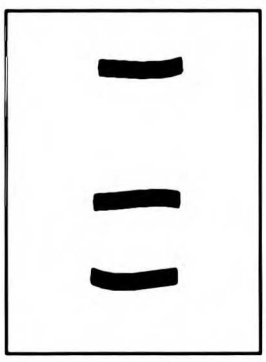


Figure 5.1. Strategy for identification of upstream genomic DNA sequences

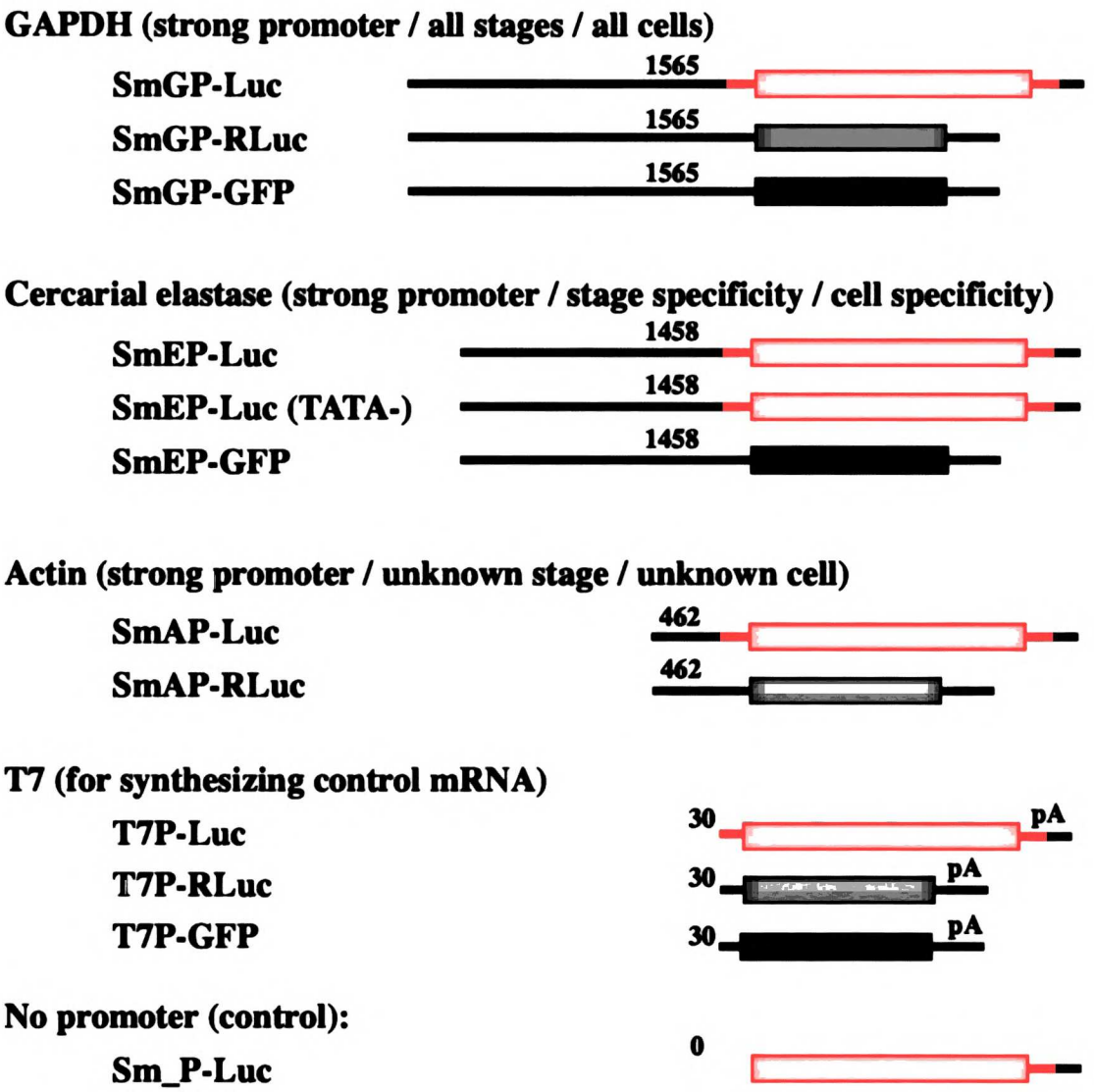


Figure 5.2 Expression constructs built for transgenic projects (as of 4/2/00)

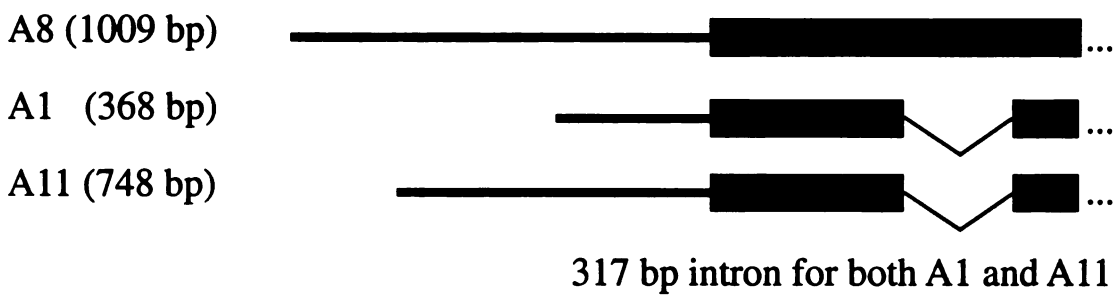


Figure 5.3. Cartoons of three actin promoters and the beginnings of their coding regions

References

1. DiConza, J.J. and E.L. Hansen, *Multiplication of transplanted Schistosoma mansoni daughter sporocysts*. J Parasitol, 1972. **58**(1): p. 181-2.

2. Cohen, L.M., H. Neimark, and L.K. Eveland, *Schistosoma mansoni: response of cercariae to a thermal gradient*. J Parasitol, 1980. **66**(2): p. 362-4.

3. Haas, W., et al., *Schistosoma mansoni cercariae: stimulation of acetabular gland secretion is adapted to the chemical composition of mammalian skin*. J Parasitol, 1997. **83**(6): p. 1079-85.

4. Stirewalt, M.A. and J.R. Hackey, *Penetration of host skin by cercariae of schistosoma mansoni. I. observed entry into skin of mouse, hamster, rat, monkey and man*. Journal of Parasitology, 1956. **42**: p. 565-580.

5. Lewert, R.M. and C.L. Lee, *Studies on the passage of helminth larvae through host tissues*. J. Infect. Dis, 1954. **95**: p. 13-51.

6. Gordon, R. and R. Griffiths, *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. and Parasitol., 1951. **45**: p. 227-243.

7. Crabtree, J.E. and R.A. Wilson, *Schistosoma mansoni: an ultrastructural examination of skin migration in the hamster cheek pouch*. Parasitology, 1985. **91**(Pt 1)): p. 111-20.

8. Marikovsky, M., R. Arnon, and Z. Fishelson, *Schistosoma mansoni: localization of the 28 kDa secreted protease in cercaria*. Parasite Immunol, 1990. **12**(4): p. 389-401.

9. Dresden, M.H., C.K. Sung, and A.M. Deelder, *A monoclonal antibody from infected mice to a Schistosoma mansoni egg proteinase*. J Immunol, 1983. **130**(1): p. 1-3.

10. Chavez-Olortegui, C., M. Resende, and C.A. Tavares, *Purification and characterization of a 47 kDa protease from Schistosoma mansoni cercarial secretion*. Parasitology, 1992. **105**(Pt 2)): p. 211-8.

11. Dalton, J.P., et al., *The cysteine proteinases of Schistosoma mansoni cercariae.* Parasitology, 1997. **114**(Pt 2)): p. 105-12.

12. Bruce, J.I., et al., *Migration of Schistosoma mansoni through mouse tissue. Ultrastructure of host tissue and integument of migrating larva following cercarial penetration.* Am J Trop Med Hyg, 1970. **19**(6): p. 959-81.

13. Teixeira, M.M., et al., *Mechanisms of the inflammatory response induced by extracts of Schistosoma mansoni larvae in guinea pig skin.* J Immunol, 1993. **151**(10): p. 5525-34.

14. Fallon, P.G., et al., *Enhancement of Schistosoma mansoni infectivity by intradermal injections of larval extracts: a putative role for larval proteases.* J Infect Dis, 1996. **173**(6): p. 1460-6.

15. Lutz, A., *Schistosomum Mansoni and Schistosomatosis observed in Brazil.* Instituto Osmaldo Cruz-Memorias, 1919. **11**: p. 109-140.

16. Cort, W.W., *The development of the Japanese blood-fluke, Schistosoma japonicum Katsurada, in its final host.* Amer. Jour. Hyg., 1921. **1**: p. 1-38.

17. Mohammed, A.S., *The secretory glands of the cercariae of S. hematobium.* Ann. Trop. Med. & Parasitol., 1931. **25**: p. 99.

18. Davis, D.J., *Report on the preparation of an histolytic ferment present in the bodies of cercariae.* J. Parasitology, 1936. **22**: p. 108-110.

19. Lewert, R.M. and C.L. Lee, *Quantitative studies of the collagenase-like enzymes of cercariae of Schistosoma mansoni and the larvae of Strongyloides ratti.* J. Infect. Dis, 1956. **99**(1): p. 1-14.

20. Stirewalt, M.A. and F.J. Kruidenier, *Activity of the acetabular secretory apparatus of cercariae of Schistosoma mansoni under experimental conditions.* Experimental Parasitology, 1961. **11**: p. 191-211.

21. Dorsey, C.H., *Schistosoma mansoni: ultrastructure of cercarial escape glands.* Exp Parasitol, 1974. **36**(3): p. 386-96.

22. Stirewalt, M.A. and C.H. Dorsey, *Schistosoma manonsi: cercarial penetration of host epidermis at the ultrastructural level*. Exp Parasitol, 1974. **35**(1): p. 1-15.

23. Dorsey, C.H., *Schistosoma mansoni: development of acetabular glands of cercaria at ultrastructural level*. Exp Parasitol, 1975. **37**(1): p. 37-59.

24. Dorsey, C.H., *Schistosoma mansoni: description of the head gland of cercariae and schistosomules at the ultrastructural level*. Exp Parasitol, 1976. **39**(3): p. 444-59.

25. Davies, T.W., *Schistosoma mansoni: the structure and elemental composition of pre-acetabular penetration gland cell secretion in pre-emergent cercariae*. Parasitology, 1983. **87**(Pt 1)): p. 55-60.

26. Dorsey, C.H. and M.A. Stirewalt, *Schistosoma mansoni: fine structure of cercarial acetabular glands*. Exp Parasitol, 1971. **30**(2): p. 199-214.

27. Gazzinelli, G., F.J. Ramalho-Pinto, and J. Pellegrino, *Purification and characterization of the proteolytic enzyme complex of cercarial extract*. Comparative Biochemistry and Physiology, 1966. **18**: p. 689-700.

28. Dresden, M.H. and H.L. Asch, *Proteolytic enzymes in extracts of Schistosoma mansoni cercariae*. Biochim Biophys Acta, 1972. **289**(2): p. 378-84.

29. Campbell, D.L., et al., *Schistosoma mansoni: partial characterization of enzyme(s) secreted from the preacetabular glands of cercariae*. Exp Parasitol, 1976. **40**(1): p. 33-40.

30. Landsperger, W.J., M.A. Stirewalt, and M.H. Dresden, *Purification and properties of a proteolytic enzyme from the cercariae of the human trematode parasite Schistosoma mansoni*. Biochem J, 1982. **201**(1): p. 137-44.

31. McKerrow, J.H., et al., *Purification and characterization of an elastinolytic proteinase secreted by cercariae of Schistosoma mansoni*. J Biol Chem, 1985. **260**(6): p. 3703-7.

32. Newport, G.R., et al., *Cloning of the proteinase that facilitates infection by schistosome parasites*. J Biol Chem, 1988. **263**(26): p. 13179-84.

33. Pierrot, C., A. Capron, and J. Khalife, *Cloning and characterization of two genes encoding Schistosoma mansoni elastase*. Mol Biochem Parasitol, 1995. **75**(1): p. 113-7.

34. Pierrot, C., et al., *Schistosoma mansoni elastase: an immune target regulated during the parasite life-cycle*. Parasitology, 1996. **113**(Pt 6)): p. 519-26.

35. Price, H.P., M.J. Doenhoff, and J.R. Sayers, *Cloning, heterologous expression and antigenicity of a schistosome cercarial protease*. Parasitology, 1997. **114**(Pt 5)): p. 447-53.

36. Ring, C.S., et al., *Structure-based inhibitor design by using protein models for the development of antiparasitic agents*. Proc Natl Acad Sci U S A, 1993. **90**(8): p. 3583-7.

37. Lim, K.C., et al., *Blockage of skin invasion by schistosome cercariae by serine protease inhibitors*. Am J Trop Med Hyg, 1999. **60**(3): p. 487-92.

38. Salter, J.P., et al., *Schistosome invasion of human skin and degradation of dermal elastin are mediated by a single serine protease*. J Biol Chem, 2000. **275**(49): p. 38667-73.

39. Qasim, M.A., et al., *Interscaffolding additivity. Association of P1 variants of eglin c and of turkey ovomucoid third domain with serine proteinases*. Biochemistry, 1997. **36**(7): p. 1598-607.

40. Schellenberger, V., C.W. Turck, and W.J. Rutter, *Role of the S' subsites in serine protease catalysis. Active-site mapping of rat chymotrypsin, rat trypsin, alpha-lytic protease, and cercarial protease from Schistosoma mansoni*. Biochemistry, 1994. **33**(14): p. 4251-7.

41. Stirewalt, M., *Schistosoma mansoni: Cercaria to Schistosomule*. Advances in Parasitology, 1974. **12**: p. 115-182.

42. Gazzinelli, G. and J. Pellegrino, *Elastolytic Activity of Schistosoma mansoni cercarial extract*. Journal of Parasitology, 1964. **50**(4): p. 591-592.

43. Tzeng, S., et al., *Degradation of purified skin keratin by a proteinase secreted from Schistosoma mansoni cercariae*. J Parasitol, 1983. **69**(5): p. 992-4.

44. McKerrow, J.H., et al., *Proteinases from invasive larvae of the trematode parasite Schistosoma mansoni degrade connective-tissue and basement-membrane macromolecules*. Biochem J, 1985. **231**(1): p. 47-51.

45. Cohen, F.E., et al., *Arresting tissue invasion of a parasite by protease inhibitors chosen with the aid of computer modeling*. Biochemistry, 1991. **30**(47): p. 11221-9.

46. Stirewalt, M.A. and B.E. Austin, *Collection of secreted protease from the preacetabular glands of cercariae of Schistosoma mansoni*. J Parasitol, 1973. **59**(4): p. 741-3.

47. Mishkin, E.M. and E.H. Jokinen, *Effects of environmental calcium on fecundity and cercarial production of Biomphalaria glabrata (Say) infected with Schistosoma mansoni Sambon*. J Parasitol, 1986. **72**(6): p. 885-90.

48. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*. Nucleic Acids Res, 1997. **25**(17): p. 3389-402.

49. Fukuyama, K., et al., *The epidermal barrier to Schistosoma mansoni infection*. Curr Probl Dermatol, 1983. **11**: p. 185-93.

50. Mecham, R.P., et al., *Elastin degradation by matrix metalloproteinases. Cleavage site specificity and mechanisms of elastolysis*. J Biol Chem, 1997. **272**(29): p. 18071-6.

51. Sakanari, J.A., et al., *Serine proteases from nematode and protozoan parasites: isolation of sequence homologs using generic molecular probes*. Proc Natl Acad Sci U S A, 1989. **86**(13): p. 4863-7.

52. Harris, J.L., et al., *Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries*. Proc Natl Acad Sci U S A, 2000. **97**(14): p. 7754-9.

53. Felsenstein, J., *Counting phylogenetic invariants in some simple cases*. J Theor Biol, 1991. **152**(3): p. 357-76.

54. Waine, G.J., et al., *Cloning, molecular characterization, and functional activity of Schistosoma japonicum glyceraldehyde-3-phosphate dehydrogenase, a putative vaccine candidate against schistosomiasis japonica*. Infect Immun, 1993. **61**(11): p. 4716-23.

55. Charrier-Ferrara, S., D. Caillol, and V. Goudot-Crozel, *Complete sequence of the Schistosoma mansoni glyceraldehyde-3-phosphate dehydrogenase gene encoding a major surface antigen*. Mol Biochem Parasitol, 1992. **56**(2): p. 339-43.

56. Stirewalt, M.A. and M. Walters, *Schistosoma mansoni: histochemical analysis of the postacetabular gland secretion of cercariae*. Exp Parasitol, 1973. **33**(1): p. 56-72.

57. Dresden, M.H. and H.L. Asch, *Calcium carbonate content of the preacetabular glands of Schistosoma mansoni cercariae*. J Parasitol, 1977. **63**(1): p. 163-5.

58. Harris, J.L., et al., *Definition and redesign of the extended substrate specificity of granzyme B*. J Biol Chem, 1998. **273**(42): p. 27364-73.

59. Toy, L., et al., *The immune response to stage-specific proteolytic enzymes of schistosoma mansoni*. Molecular Paradigms for Eradicating Helminthic Parasites, 1987: p. 85-103.

60. Butterworth, A.E., et al., *Immunity after treatment of human schistosomiasis mansoni. I. Study design, pretreatment observations and the results of treatment*. Trans R Soc Trop Med Hyg, 1984. **78**(1): p. 108-23.

61. Dalton, P.R., *A socioecological approach to the control of Schistosoma mansoni in St Lucia*. Bull World Health Organ, 1976. **54**(5): p. 587-95.

62. 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, 1997. **56**(6): p. 668-73.

63. Knight, W.B., et al., *A modification of the formol-ether concentration technique for increased sensitivity in detecting Schistosoma mansoni eggs*. Am J Trop Med Hyg, 1976. **25**(6): p. 818-23.

64. Jourdane, J., A. Theron, and C. Combes, *Demonstration of several sporocysts generations as a normal pattern of reproduction of Schistosoma mansoni*. Acta Trop, 1980. **37**(2): p. 177-82.

65. Ivanchenko, M.G., et al., *Continuous in vitro propagation and differentiation of cultures of the intramolluscan stages of the human parasite Schistosoma mansoni*. Proc Natl Acad Sci U S A, 1999. **96**(9): p. 4965-70.

66. Cooper, L.A., F.A. Lewis, and S. File-Emperador, *Re-establishing a life cycle of Schistosoma mansoni from cryopreserved larvae*. J Parasitol, 1989. **75**(3): p. 353-6.

67. Cohen, L.M. and L.K. Eveland, *Schistosoma mansoni: characterization of clones maintained by the microsurgical transplantation of sporocysts*. J Parasitol, 1988. **74**(6): p. 963-9.

68. Cohen, L.M. and L.K. Eveland, *Schistosoma mansoni: long-term maintenance of clones by microsurgical transplantation of sporocysts*. Exp Parasitol, 1984. **57**(1): p. 15-9.