MINIREVIEW

Recent Insights into the Structure and Function of a Larval Proteinase Involved in Host Infection by a Multicellular Parasite

(43233)

JAMES H. MCKERROW,*,[1] GEORGE NEWPORT,† AND ZVI FISHELSON‡

Departments of Pathology and Pharmaceutical Chemistry,† University of California, San Francisco 94143 and the San Francisco Veterans Administration Medical Center,* San Francisco, California 94121, and the Department of Chemical Immunology, Weizmann Institute of Science, Rehovot, Israel*

Tissue invasion by motile cells or organisms is a key step in embryonic development, inflammation, cancer metastasis, and host infection by parasites, fungi, and bacteria (1). In each of these phenomena, tissue invasion is a multifactorial event requiring chemotaxis, motility, adhesion, and tissue degradation. Although each cell or organism has unique properties, any factor identified as being shared in common is of particular importance in helping to define molecular mechanisms of tissue invasion. One of the better characterized factors common to these phenomena is the elaboration of proteinases for tissue degradation. This review will focus on an example of host infection by a multicellular parasite for which proteinase expression is particularly important because of the size of the organism involved and the multiplicity of roles the enzyme can play.

Schistosome Infection

Schistosomes (blood flukes) are trematode parasites that, while infecting a variety of animals, are best known for producing one of the world's major health problems, schistosomiasis (Fig. 1). This disease is endemic throughout much of the tropical world and is estimated to affect over 200 million people (2). The disease itself is produced by a host granulomatous response to eggs laid by adult female worms in the portal venous system. Initiation of infection occurs when an aquatic larval form, the cercaria, invades host skin. Morphologic studies of experimental infections showed that cercariae invade intact skin and tunnel through the epidermis and dermis, eventually entering small blood vessels (Fig. 2). From there, they reach the systemic circulation via the pulmonary vasculature and develop into adults after reaching the portal vein and its tributaries.

Cercariae develop in an intermediate-host freshwater snail and find their human host by following a thermal gradient (3). Their stimulus for invading the host is lipid on the surface of skin (3). Specific 12–14 carbon-free fatty acids (such as linoleic acid) are the key components of skin lipid which induce invasion (4). Concomitant with invasion, the cercaria begins to transform to a new larval form, the schistosomulum. This transformation involves loss of a surface glycocalyx that helps to maintain osmotic balance in fresh water. In the host, this glycocalyx is a powerful activator of complement, and is therefore potentially lethal to the parasite (5, 6). The cercaria also loses its tail, and, within a few days of entering the host, exhibits profound changes in metabolism. These changes all contribute to adaptation for the aquatic to the host environment. As will be discussed below, a serine proteinase released during invasion may function both in facilitating tissue degradation and facilitating metamorphosis from the aquatic to the nonaquatic larva form.

Serine Proteinases Released by Cercariae

Work from a number of laboratories over the past three decades has identified serine proteinase activity as the major proteolytic activity released from invading and transforming larvae (6–11). Several proteolytic species were identified in glandular secretions or conditioned media from transforming larvae. Recent structural analysis has shown that at least some of these are in fact posttranslational derivatives of the same gene product (11, 12). Although evidence to date suggests that there is a single serine proteinase gene in schisto-

[1] To whom requests for reprints should be addressed at Department of Pathology, University of California, School of Medicine, San Francisco, CA 94143-0506.

0037-9727/91/1972-0119$3.00/0

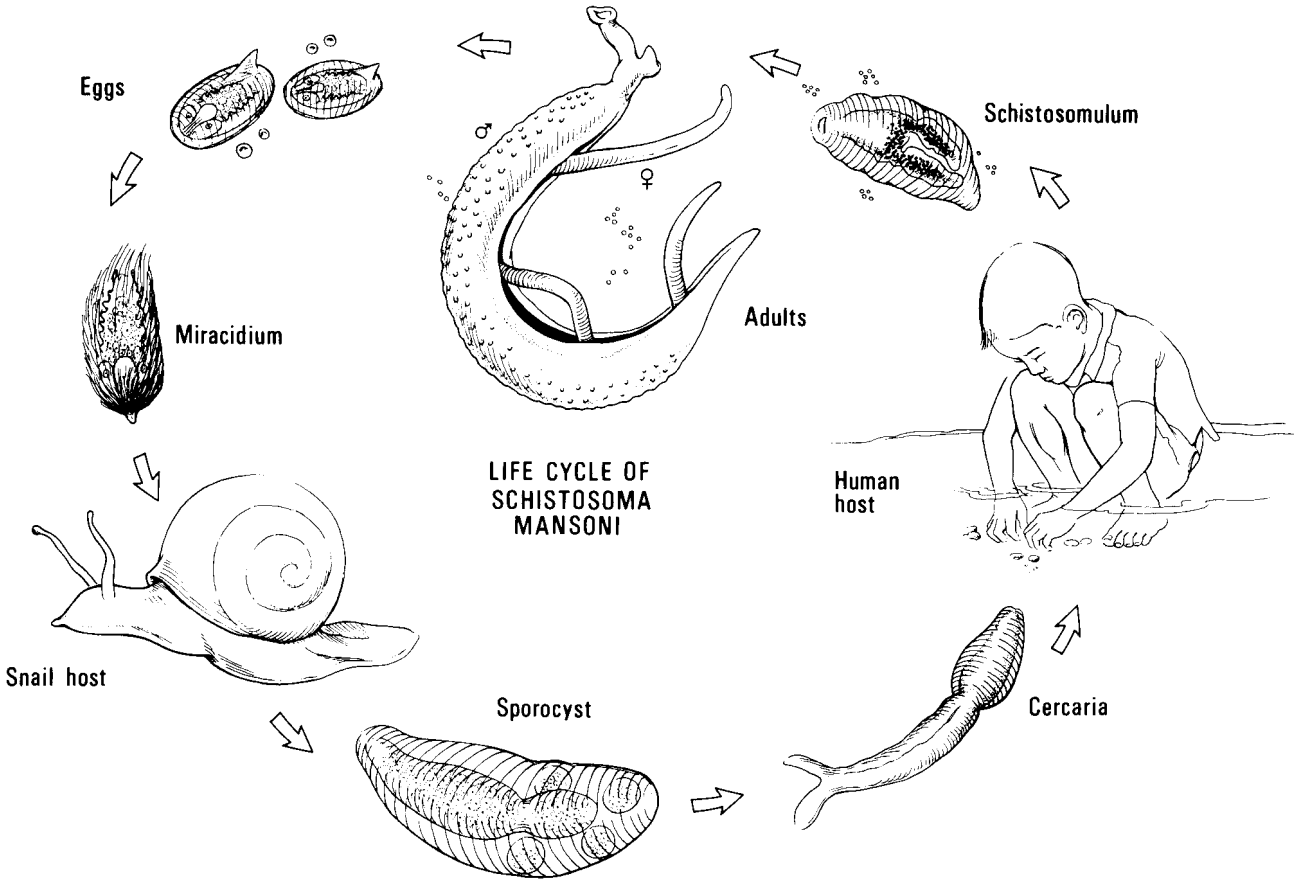

Figure 1. Life cycle of *Schistosoma mansoni*, one of the three major schistosome species causing human disease. Illustration by Charles Hoffman.

somes, the origin of certain biochemical differences between the serine proteinase species remains obscure. Nevertheless, it is clear that the major serine proteinase activity originates from vesicles derived from a group of cells known as the acetabular glands (3). Release of the contents of these cells is initiated upon stimulation by skin lipid and can be detected up to 48 hr after stimulation. As will be discussed below, there is also data from immunologic localization studies that suggests this, or a similar enzyme, may be present even longer.

Role of Secreted Serine Proteinase in Skin Invasion

The fact that the acetabular gland proteinase was released upon stimulation by skin lipid and was active at neutral pH suggested that it may play a role in tissue degradation during cercarial invasion of skin (3). This hypothesis received further support when it was found that the enzyme had a remarkably broad substrate specificity toward host tissue macromolecules, including keratin, fibronectin, laminin, type IV collagen, and elastin (9). When applied to skin, purified enzyme produced spongiosis of the epidermis similar to that seen in the tunnels made by cercariae (13). *In vitro* studies of cercarial degradation of the extracellular matrix showed that matrix degradation by live cercariae was blocked by an α_1-proteinase inhibitor, a potent inhibitor of the enzyme (14). More recently, molecular modeling of the enzyme has led to the prediction of specific peptide-based inhibitors which inhibit cercarial penetration of human skin.

Serine Proteinase Release and Acquisition of Complement Resistance

The complement system, the effector arm of the humoral immune response, plays a major role in host protection against invading microorganisms. It is composed of 19 plasma proteins which interact, in a cascade form, to produce the lytic membrane attack complex (16). Besides their direct lytic activity, complement components can potentiate the phagocytic and cytotoxic activities of leukocytes. Complement is activated via the classical pathway by antibodies that bind to surface antigens of the foreign organisms. Activation also may be initiated, in the absence of antibodies, via the alternative pathway by surface substances such as the lipopolysaccharide of gram-negative bacteria or the glycocalyx coat of cercariae. Indeed, cercariae are rapidly killed *in vitro* by nonimmune sera (for a review, see reference 17). As mentioned before, one of the early

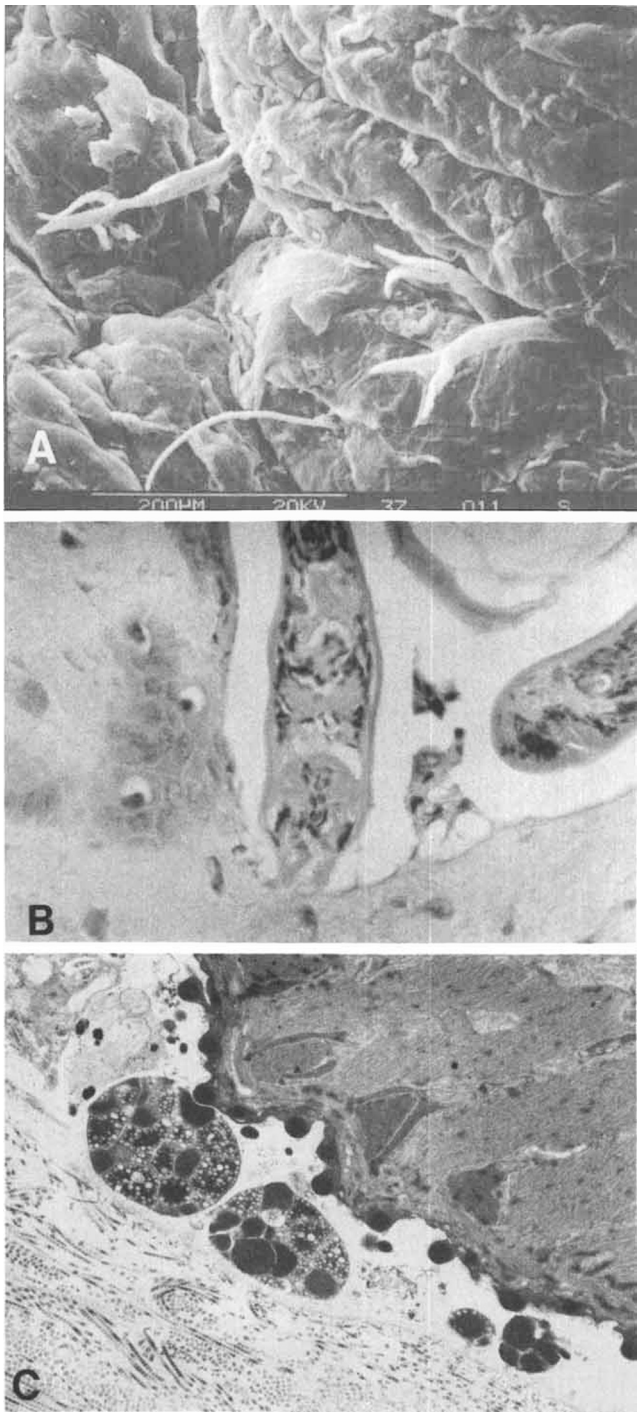

Figure 2. (A) Scanning electron micrograph of cercariae invading surface of human skin. Only forked tails protrude from sites of entry of three cercariae. SEM by Dr. Gerard Apodaca, University of California, San Francisco. (B) Light micrograph of cercariae invading skin, seen in cross-section. Note dissolution of epidermis. One organism can be seen in the process of degrading the extracellular matrix of the dermis (original magnification ×400). (C) Electron micrograph of cercaria (upper right) and ovoid clusters of secretory vesicles, which contain the proteinase. The collagen of the extracellular matrix is in the lower left corner; debris from degraded epidermal cell is in the upper left corner. Cercaria is at the approximate location shown in the light micrograph in Figure 2B (original magnification ×19,000).

stages of larval adaptation to host factors is the shedding of the glycocalyx. Since the latter phenomenon coincides with release of the serine proteinases, it has been suggested that these enzymes facilitate removal of the glycocalyx and acquisition of complement resistance (11). This suggestion received support from experiments showing that the purified serine proteinase, when added exogenously to transforming larvae, accelerated release of glycocalyx and conversion of the larvae to complement resistance (11). The patterns of fragmentation of the glycocalyx released spontaneously and the glycocalyx released by the purified enzyme were identical. Trypsin had no effect on these processes.

Shortly after removal of the glycocalyx, the schistosomula express on their surface a serine proteinase which appears to be identical or similar to the soluble enzyme. The two enzymes have the same molecular size and cross-react immunologically. Furthermore, both enzymes degrade the complement proteins C3, C3b, iC3b, and C9 yielding the same pattern of degradation products (18). This latter finding suggests that the released and membrane bound serine proteinases may further protect the larva against immune damage by cleaving off the complement proteins that become attached to its surface during complement activation.

Structure of the Cercarial Proteinase Gene Provides Clues to Its Evolutionary Relationships in the Serine Proteinase Family

Aside from its importance in host infection, the cercarial proteinase is a valuable link in attempts to map the molecular evolution of enzymes. Serine proteinases are one of the most extensively studied families of enzymes (19). There are two major subgroups within the serine proteinase family. The subtilisin group is represented primarily by prokaryotic serine proteinases, while the trypsin family is represented primarily by eukaryotic proteinases. Both of these groups use the same catalytic triad of amino acids, and this triad is arranged in an identical three-dimensional array. However, both the primary amino acid sequence and three-dimensional folding of the protein are unique to each group.

The cloning of four cDNA encoding the cercarial proteinase has allowed structural analysis of both the gene and the protein (12). Sequence analysis of the cercarial proteinase gene clearly indicated that it was a member of the trypsin family (Fig. 3). As with other members of the family, there was conservation of amino acids around each of the active site amino acids, as well as other important structural motifs.

Analysis of the nucleotide sequence and organization of the schistosome cercarial proteinase gene has also provided clues to its molecular evolution, as well as to evolutionary modifications in intron-exon boundaries within the serine proteinase gene family. The

```
Sch    MSNRWFVVVVTLFTYCLTFERVSTWL
RatI   *LRFLVFASLV*YGHSTQDFPETNAR
RatII  MIRTLLLSAL*GA*SCGYP*Y*VQHDVS

                          20                  40                  60
Sch    IRSGEPVQHAEFPF-----IAFLTTERTMCTGSLVSTRAVLTAGHCVCSPLPVIintronRVSFL
RatI   VVG*AEARR-NSW*SQISLQYLSGGSWYHT*G*T*IRRNW*M*A****S-----------
RatII  VVG*QEASP-SNW*WQVSLQYLSSGKWHHT*G****ANNW***A***IS-----------
Pig    VVG*TEAQR-NSW*SQISLQYRSGSSWAHT*G*T*IRQNW*M*A****D-----------
Vc     *VG*TDAPR-GKY*YQVSL
Cf     *V**TDATL-G***YQLSFQN
Sg     VVG*TRAAQ-G****

                         80                  100                 120
Sch    TLRNGAQQGIHHQPSGVKVAPGYMPSCMSARQRRPIAQTLS--GFDIAIVMLAQMVNLQS
RatI   SQMTF*VVVGD*NL*QNDGTEQ*VSVQKIMVHPTWNSNNVA-A*Y***LLR***S*T*NN
RatII  NS*TYRVLLGR*SL*TSESGSLAVQVSKLVVHEKWN**K**NN*N***L*K**SP*A*T*
Pig    RELTFRVVVGE*NLNQNNGTEQ*VGVQKIVVHPYWNTDDVA-A*Y***LLR***S*T*N*

                         140                 160                 180
Sch    GIRVISLPQPSKIPPPGTGVFIVGTGR---DDDNDRDPSRLNGGILKKGRATIMECRHAT
RatI   YVQLAV***EGT*LANNNPCY*T*W**TRT*GQLSQTLQQAYLPSVDYSICSSSSYWGS*
RatII  K*QTAC**PAGT*LPNNYPCYVT*W**LQT*GAT*DVLQQGRLLVVDYATCSSASWWGS*
Pig    YVQLGV**RAGT*LANNSPCY*T*W*LTRT*GQLAQTLQQAYLPTVDYAICSSSSYWGS*

                         200                 220                 240
Sch    NGNPICVKAGQNFGQLPAPGDSGGPLLPSLQGPVLGVVSHGVTLPNLPDIIVEYASVARM
RatI   VKTTTM*C**GDGVRSGCQ*******HCL---VNGQYSV****SFVSSMGCNVSKKPTVF
RatII  VKTTNM*C**GDGVTSSCN*******NCQ--ASNGQWQV**I*FFGSTLGCNYPRKPSVF
Pig    VKNNSM*C**GNGVRSGCQ*******HCL---VNGQYAV****SFVSRLGCNVTRKPTVF
Vc           QIC*FTKL*EGAC******

Sch    VDFVRSNI
RatI   TRVSAYISWMNNVIAYT
RatII  TRVSNYIDWINSVIAKN
Pig    TRVSAYISWINNVIASNSN
```

Figure 3. Comparison of the predicted schistosome elastase (Sch) amino acid sequence to those of the most closely related vertebrate and invertebrate serine proteinases. Abbreviations used in this figure: RatI, II, rat pancreatic elastase I and II; PIG, porcine elastase; Vc, hornet proteinase II; Cf, crayfish trypsin; and Sg, *Streptococcus griseus* proteinase. Asterisks denote regions of identity and dashes indicate gaps. Components of the catalytic triad (histidine 46, aspartate 106, and serine 202) are shown in bold lettering, and the numbering system is arbitrary, with each space representing a position.

alignment shown in Figure 3 also suggests that the schistosome enzyme evolved by a series of additions (positions 50–60) and deletions (positions 16–20) which map to potential splice sites in the genome. Interestingly, the predicted exon encoding amino acids 151–159, which is the one where the schistosome enzyme diverges the most from other serine proteinases, has a stretch rich in aspartate residues (D), a feature reminiscent of calcium-binding domains.

Genomic Southern analyses using a cDNA probe encoding all of the translated region of the mature elastase generally results in a single hybridizing fragment, which suggests that the enzyme is encoded by a single-copy gene (B. Wilk, unpublished results). This indicates that the broad specificity of the encoded enzyme is an intrinsic property, and is not due to the presence of a family of closely related proteins of differing specificities. Using synthetic oligonucleotides as primers, it was determined by the polymerase chain reaction (PCR) that the coding sequence of the elastase gene is approximately 150 nucleotides longer than the corresponding portion of the mRNA. This discrepancy is accounted for by a single intron near the 5′ end of the gene. The sequence of the intron contains consensus eukaryotic splicing signals (i.e., 5′exon-GT and 3′ AG-exon). All other eukaryotic serine proteinase genes sequenced to date contain more than one intron, with the positions of these intervening sequences being placed at conserved positions.

The single intron in the schistosome elastase gene separates codons for isoleucine 58 and arginine 59. This location is close to a predicted site based on other serine proteinase genes, and alignment of the schistosome enzyme with homologous from rats requires the use of gaps (Fig. 3). The "missing" introns at positions encoding amino acids phenylalanine 14, serine 101, and aspartate 155 also occur in regions where gaps are introduced in the schistosome enzyme. Assuming that

the schistosome elastase gene has lost introns during its evolution, the observed findings are consistent with the theory that parasitic organisms tend to lose unnecessary genomic baggage. The rationale is that genetic information is metabolically costly, and that the complicated life histories of many parasites require discrete genomic programs for surviving in vastly different habitats, and may thus require the presence of a high number of single-copy genes, associated with the life cycle. For example, the gene coding for the schistosome hypoxanthine guanine phosphoribosyltransferase (HGPRTase), has the same number of introns as HGPRTase genes of higher eukaryotes (e.g., human), but several of the schistosome introns are extraordinarily small (20). Also of note in this regard is the fact that the fraction of repetitive DNA in the schistosome genome is much lower than that of most eukaryotes (21).

Analysis of the complete nucleotide sequence of four cDNA revealed that the 5′ untranslated region of the composite proteinase cDNA is unusually long (> 254 base pair (bp)) and remarkably AT rich (70%), and contains an 11-base stretch (positions 12–22 5′ of the ATG codon) that significantly complements a conserved portion of the eukaryotic 18s rRNA. Recent isolation of a fifth cDNA clone by Dr. Johnny Railey at University of California, San Francisco (unpublished results) has demonstrated that the 3′ untranslated region is 149-bp long and contains a tetranucleotide sequence (TCAG) repeated six times in tandem. The significance of this repeat is unknown, although the same sequence is repeated three times in tandem in the 5′ flanking region of the μ-chain C-region germ line gene, near the switch region recombination site. A consensus polyA addition site (AAUAAA) is present 122 bp from the end of the open reading frame (J. Railey, personal communication).

It has recently been demonstrated that, like nematodes and trypanosomes, schistosomes can undergo trans-splicing (F. Rottman, personal communication). This is a process in which mRNA transcribed from discontiguous parts of the genome are subsequently joined in a splicing reaction. Attempts to demonstrate whether or not the schistosome proteinase transcriptional unit subsequently acquires a spliced leader (using oligonucleotide primers for the described spliced leader) have yielded negative results, suggesting that the processing of the schistosome proteinase gene occurs by conventional mechanisms, or that it acquires a yet-to-be-described leader (B. Wilk, unpublished results).

Computer Modeling of the Active Site of the Cercarial Proteinase and Design of Specific Peptide Inhibitors

Regions of homology between the cercarial proteinase and other serine proteinases allowed molecular modeling of the enzyme based upon the crystal structure of alphalytic proteinase of *Streptococcus griseus*

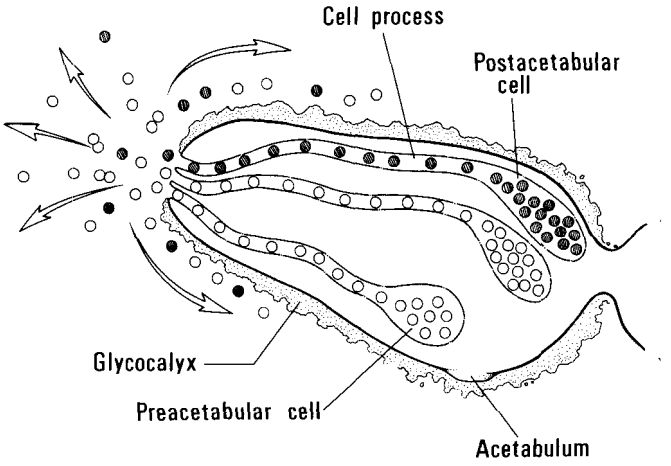

Figure 4. Model of vesicle secretion by pre- and postacetabular cells. Proteinase is present in both types of vesicles and diffuses in advance of cercariae, as well as along their surfaces. See Figure 2B for electron micrograph of secretory vesicles. Illustration by Charles Hoffman.

and pancreatic elastase (15). A deep hydrophobic substrate binding cleft was visualized. This correlated with the chymotrypsin-like specificity of the cercarial proteinase for large hydrophobic amino acids in the P-1 site of peptide substrates. However, in contrast to chymotrypsin, there was minimal steric hindrance for large hydrophobic amino acids at the P-4 site (the fourth amino acid from the site of cleavage of a protein or peptide substrate). This more "tolerant" substrate binding site may correlate with the broad substrate specificity toward connective tissue macromolecules exhibited by the cercarial proteinase versus chymotrypsin. Interestingly, the two enzymes for which the cercarial proteinase shared greatest sequence identity were pancreatic elastases I and II, both of which also have a broad substrate specificity.

Molecular modeling of the enzyme allowed very specific chloromethyl ketone-derivatized peptide inhibitors to be designed, which confirmed the predictions of the model in assays with purified enzyme (15). These inhibitors were also used to test the hypothesis that the enzyme plays a role in cercarial invasion by demonstrating that they also inhibited cercarial invasion of human skin (15).

Cellular Localization and Biosynthesis of the Cercarial Proteinase

A cDNA probe (12) and monospecific antisera (22) to the cercarial proteinase were used to study its biosynthesis, storage, and release in developing cercariae. Messenger RNA for the enzyme was detected in the earliest stages of cercarial differentiation (12). Large cells with prominent nucleoli were identified as some of the first to differentiate from the primordial embryonic mass of cells in the sporocyst (the stage of the parasite in the intermediate host snail). Message was immediately

translated into protein and stored in cytoplasmic vesicles (22, 23). As the cercariae developed, cytoplasmic processes from these cells extended to the anterior end of the cercarial head to form a "duct" system (18). Prior to emergence of cercariae from the intermediate host snail, vesicles containing the enzyme filled the cytoplasm of the cell bodies, as well as the "ducts," compressing other organelles to a thin rim. Therefore, cercariae were already "primed and loaded" with enzyme before their release from the snail. Immunologic localization even suggested that some enzyme may leak out while the cercariae are still in the snail, raising the possibility that the enzyme may also function to aid in cercarial emergence from the snail (shedding).

Although the term acetabular glands has been used to describe the cells in which the enzyme is synthesized, these are not, in fact, true epithelial glands. The cells appear to be of mesenchymal, rather than epithelial, origin (23). They have no desmosomes or cytokeratins, but contain vimentin, a marker of mesenchymal cells. Rather than the enzyme being secreted into an intercellular lumen and an epithelial lined duct as in mammalian glands, cytoplasmic evaginations form a pseudoduct system by which secretions are delivered to the anterior end of the organism. Strong muscular contraction, initiated upon stimulation by skin lipid, results in the rupture of the apical cytoplasmic blebs and the release of enzyme containing vesicles. Immunolocalization of the enzyme within the host (23) shows that it diffuses from ruptured vesicles ahead of the advancing cercaria, as well as along the surface of the cercaria itself, during penetration (Fig. 4). This is consistent with the dual role it probably plays in tissue degradation and glycocalyx release, two early events required for successful infection of the human host.

This work was supported by U.S. National Institute of Allergy and Infectious Disease Grant AI20452, the John D. and Catherine T. MacArthur Foundation, and the Edna McConnell Clark Foundation.

1. Sordat B. Is there a common definition for invasiveness? Invasion Metastasis **10**:178–192, 1990.
2. Cline B. Schistosome (blood fluke) diseases. In: Goldsmith R, Heyneman D, Eds. Tropical Medicine and Parasitology. San Mateo, CA: Appleton and Lange, pp434–458, 1989.
3. Stirewalt MA. *Schistosoma manson*: Cercaria to schistosomule. Adv Parasitol **12**:115–182, 1974.
4. Haas W, Schmitt R. Chemical stimuli for penetration of *Schistosoma mansoni* cercariae. Naturwiss **65**:110–111, 1978.
5. Samuelson JC, Caulfield JP. The cercarial glycocalyx of *Schistosoma mansoni*. J Cell Biol **101**:1423–1434, 1985.
6. Marikovsky M, Arnon R, Fishelson Z. Proteases from invasive larvae of the trematode parasite *Schistosoma mansoni* promote resistance to killing by complement. J Immunol **141**:273–278, 1988.
7. Gazzinelli G, Pellgrino J. Elastolytic activity of *Schistosoma mansoni* cercarial extract. J Parasitol **50**:591–592, 1964.
8. Landsperger WJ, Stirewalt MA, Dresden MH. Purification and properties of a proteolytic enzyme from the cercariae of the human trematode parasite *Schistosoma mansoni*. Biochem J **201**:137–144, 1982.
9. McKerrow JH, Jones P, Sage H, Pino-Heiss S. Proteinases from invasive larvae of the trematode parasite *Schistosoma mansoni* degrade connective tissue and basement membrane macromolecules. Biochem J **231**:47–51, 1985.
10. McKerrow JH, Pino-Heiss S, Lindquist R, Werb Z. Purification and characterization of an elastinolytic proteinase secreted by cercariae of *S. mansoni*. J Biol Chem **260**:3703–3707, 1985.
11. Marikovsky M, Fishelson Z, Arnon R. Purification and characterization of proteases secreted by transforming schistosomula of *Schistosoma mansoni*. Mol Biochem Parasitol **30**:45–54, 1988.
12. Newport GR, McKerrow JH, Hedstrom R, Petitt M, McGarrigle L, Barr PJ, Agabian N. Cloning of the proteinase that facilitates infection by schistosome parasites. J Biol Chem **263**:13179–13184, 1988.
13. McKerrow JH, Sakanari J, Brown M, Brindley P, Railey J, Weiss N, Resnick SD. Proteolytic enzymes as mediators of parasite invasion of skin. In Maibach H, Lowe NJ, Eds. Models in Dermatology. Basel: Karger, Vol **4**: pp276–284, 1989.
14. McKerrow JH, Keene WE, Jeong K, Werb Z. Degradation of extracellular matrix by larvae of *Schistosoma mansoni*. I. Degradation by cercariae as a model for initial parasite invasion of host. Lab Invest **49**:195–200, 1983.
15. Cohen FE, Gregoret LM, Amiri P, Aldape, K, Railey J, McKerrow JH. Arresting tissue invasion of a parasite with synthetic protease inhibitors chosen by computer modeling. Biochemistry (in press).
16. Muller-Eberhard HJ. Molecular organization and function of the complement system. Annu Rev Biochem **57**:321–347, 1988.
17. Fishelson Z. Complement and parasitic trematodes. Parasitol Today **5**:19–25, 1989.
18. Parizide M, Ghendler Y, Arnon R, Fishelson Z. Resistance of the parasite *Schistosoma mansoni* to immune attack. FASEB J **4**:A2228, 1990.
19. Neurath H. The versatility of proteolytic enzymes. J Cell Biochem **32**:35–49, 1986.
20. Craig SP, Muralidhar MG, McKerrow JH, Wang CC. Evidence for a class of very small introns in the gene for HGPRTase in *S. mansoni*. Nucleic Acids Res **17**:1635, 1989.
21. Simpson AJG, Sher A, McCutchan T. The genome of *Schistosoma mansoni*: Isolation of DNA, its size, and repetitive sequences. Mol Biochem Parasitol **6**:125–137, 1982.
22. Marikovsky M, Arnon R, Fishelson Z. *Schistosoma mansoni*: Localization of the 28 kDa secreted protease in cercaria. Parasite Immunol (in press).
23. Fishelson Z, Amiri P, Friend DS, Marikovsky M, Petitt M, and McKerrow JH. Cell-specific expression and secretion of a serine protease during development of a metazoan parasite. 1991, submitted for publication.