Asparaginyl Endopeptidase from *Schistosoma mansoni:*

A promising new target for anti-parasitic therapy

by

Mary Alice Mathieu

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Biomedical Science

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco

Date

University Librarian

Acknowledgement

This work was supported by a TDRU grant (AI35707) and the Sandler Family Foundation.

I am indebted to Dr. James H. McKerrow for his direction and support. I have greatly appreciated the opportunity to learn from his rich knowledge in Biomedical Science.

I dedicate this work to my parents, Joan and Donald R. Mathieu.

Table of Contents

Tables and Figures

Page Number

1. Introduction

The parasitic disease, schistosomiasis (Bilharzia), caused by tropical blood flukes of the genus *Schistosoma* is a devastating health problem for developing countries, afflicting an estimated 200-250 million people worldwide (1). According to the World Health Organization, schistosomiasis ranks second behind malaria in terms of socioeconomic and public health importance in tropical and subtropical areas. Effective control of schistosomiasis ideally includes a multidisciplinary approach combining water sanitation, awareness training, immunotherapy and chemotherapy. An understanding of schistosome biochemistry is crucial for clarification of aspects of the disease process that remain obscure. Intervention must be aimed at processes that are critical for parasite development and infectivity. Proteases are considered rational targets for chemotherapeutic intervention and novel anti-parasitic drugs because of their importance in key aspects of schistosome biology. Parasite proteases are essential for host invasion, pathogenicity, nutrient catabolism and immunoevasion. The identification of proteases in the pathway of host blood protein catabolism has become a major focus of research in parasitology. The mechanism of hemoglobin catabolism is thought to involve a cascade of proteases including cysteine proteases (cathepsins L-, B-, and C-like, and asparaginyl endopeptidase) and the aspartic protease cathepsin D. The major goal of my research was to characterize the initial steps in the pathway of hemoglobin degradation by *Schistosoma mansoni.*

I. 2. Schistosome physiology, infectivity and reproduction

Human schistosomiasis is caused by three species belonging to the genus Schistosoma: *S. mansoni, S. japonicum,* and *S. haematobium.* Schistosomes enter the vertebrate host through contact with infested surface waters in rural agricultural and fishing communities. Sequential stages of development define the complex life cycle of the helminth (Figure 1). Infection is initiated when the aquatic larvae (cercariae) penetrate human skin. They migrate through the venous system to the lung capillary bed, into the heart, and on to their final intravascular habitat where they mature to adult worms. *S. japonicum* lives in the superior mesenteric veins, *S. haematobium* in the venous plexus surrounding the urinary bladder, and *S. mansoni* in the inferior mesenteric veins. Although the three species are similar in their basic life cycles, they differ markedly with respect to adult morphology, egg shape and size, intermediate (invertebrate) host, and infectivity. *S. mansoni* is the most widely studied due to its accessibility and the ease of maintaining life cycle stages in a laboratory setting.

S. mansoni causes intestinal and hepatic schistosomiasis. Adult worms have separate sexes: a female measuring 12-15 mm in length, and a male measuring 7-10mm. They mate continuously as the male surrounds the female in its gynocophoric canal. The female produces approximately 300 eggs per day. Adult worms have ventral and anterior suckers that are used to attach to the endothelium of the hepatic portal veins. The eggs pass through the birth pore located above the posterior sucker and can penetrate the host endothelium and intestinal wall where they stimulate the host immune response. Accumulation of parasite eggs and the ensuing immune response are the major cause of

disease pathology. Eggs that reach the small intestine are passed in the feces into fresh water, perpetuating the cycle of infection. The miracidium hatches in fresh water and represents the first larval stage that finds its snail host, *Biomphalaria glabrata.* Within the snail, infection proceeds through the mother sporocyst stage, the daughter sporocyst, and finally the cercaria, which exits from the snail to come into contact with the vertebrate host. Cercariae follow a thermal gradient to find their host. Lipids on the surface of host skin stimulate penetration.

In all stages of the life cycle, schistosomes express proteolytic enzymes to act on several different protein substrates of the infected host. A protease secreted from the cercarial preacetabular gland facilitates skin penetration. Upon entering the skin of the definitive host, this enzyme is secreted to degrade polymers of the dermal extracellular matrix (2). Proteases have also been detected in eggs, which may be secreted to assist egg penetration into tissues of the intestine (3). In addition, proteinase activity may be involved in hatching of the miracidium and infection of the intermediate snail host (4). Finally, the action of proteolytic enzymes is of vital importance in parasite nutrition.

I. 3. Hemoglobin degradation in *S. mansoni*

Schistosomes utilize blood proteins such as hemoglobin (Hb) as a source for amino acids essential for growth, development and reproduction (Figure 2). Red blood cells (RBC) are ingested and lysed by the action of hemolysins in the parasite esophagus. Hemoglobin is released from RBC's into the blind bifurcated cecum where it is catabolized by a cascade of proteases (5). Several proteases have been identified in the

gut lumen or cells: cathepsin B (SmCB1), schistosome legumain, and cathepsin L-, D-, and C-like enzymes. These enzymes are expressed in the helminth gastrodermis and cecum and have pH optima from 4.5 (cathepsin D) to 6.8 (cathepsin B) (6).

There remains some controversy over the identity of the schistosome 'hemoglobinase', but the picture emerging is that a battery of proteases contributes to native hemoglobin digestion. *S. mansoni* legumain (Sm32) and SmCB1 enzymes are immunodiagnostic markers for schistosomiasis and were early candidates for the schistosome 'hemoglobinase'. Legumain was initially considered to be the *S. mansoni* hemoglobinase because its *B-balactosidase*-Sm32 fusion protein expressed in *Esherichia coli* was reported to degrade hemoglobin (7). This hypothesis was challenged on the basis that recombinant Sm32 from insect cells could not degrade hemoglobin and its sequence showed no similarity to any other known protease (8). In addition, the predominant protease activity in schistosome gut regurgitant was shown to be cathepsin B-like, and the cathepsin B inhibitor [L-3-*trans*- (propylcarbamoyl) oxyrane -2-carbonyl]-L-isoleucyl-L-proline (CA-074), blocks the major protease activity in worm vomitus from both *S. mansoni* and *S. japonicum* (9). While cathepsin B may be the primary protease, the cathepsin L-, D-, and C-like enzymes may act downstream in the hemoglobin degradation pathway (10). The role of legumain remains uncertain. Although legumain localization in the schistosome gut indicates that it plays a role in Hb degradation, inhibitors such as CA-074 and E-64, which block Hb degradation, do not inhibit legumain. The focus of this project is the biochemical characterization of schistosome 'legumain' in an effort to understand its role in hemoglobin degradation and to determine whether it could be a useful target for chemotherapeutic intervention.

I. 4. Asparaginyl endopeptidases

Asparaginyl endopeptidases (EC 3.4.22.34), or 'legumains' belong to the C13 family of cysteine proteases. They have been assigned to clan CD, along with the caspases, clostripains, gingipains, and GPI: protein transamidases based on their catalytic dyad in the motif His-Gly-spacer-Ala-Cys (11, 12). The legumain family members are believed to have a protein fold closely related to the caspases (11). The catalytic residues in all known cysteine endopeptidases are a dyad of cysteine and histidine, which may occur in either his-cys or cys-his order in the amino acid sequence. Two blocks of amino acids are conserved between the plant and schistosome legumains that contain active site residues: Glu189-Ala-**Cys**-Glu-Ser-Gly-Ser195, and Val144-Phe-Ile-Tyr-Phe-Thr-Asp-**His**-Gly-Ala-Pro-Gly155 (13). The thiol group of the cysteine residue acts as the nucleophile and histidine is the proton donor/general base. In some families of cysteine proteases a third residue orientates the immidazolium ring of His. In other families, a third residue is needed to form the oxyanion hole. In the case of legumain, a glutamine within the third block of conserved residues may complete the catalytic triad. However, this awaits confirmation by structural studies and site-directed mutagenesis.

Legumains were first characterized in the seeds of leguminous plants and later discovered in the parasitic blood fluke *S. mansoni* (14) and in mammalian cells (15). There is evidence that the plant enzymes process and degrade proenzymes in storage vacuoles (16). The mammalian enzymes function in bacterial antigen processing for MHC class II presenting cells and in the inhibition of osteoclast formation and bone

resorption (17). *S. mansoni* asparaginyl endopeptidase (Sm32) was first identified as a diagnostic marker for schistosomiasis and was proposed as a candidate in the protease pathway of host hemoglobin degradation by the parasite (18). Monoclonal antibodies have localized Sm32 to the gut epithelium of the worm where it may *trans*-process other proteins including proteases that are involved in hemoglobin digestion (19). Although its role in parasite nutrition may be indirect, it is an attractive target for inhibition of "downstream" metabolic processes catalyzed by cysteine proteases. Broad-spectrum inhibitors against cysteine proteinases have been used to treat *S. mansoni*-infected mice, resulting in reduced worm burden and parasite egg production, thus demonstrating that cysteine proteinases are promising targets for antischistosome chemotherapy (20).

Proposal:

I propose that schistosome legumain is primarily an activating protease indirectly involved in hemoglobin degradation by initiating the protease cascade in the schistosome gut. I further propose that specific inhibitors could be designed to interrupt this pathway by mapping the active site specificity of schistosome legumain versus the human homologue. I will construct and apply a positional-scanning synthetic combinatorial library (PS-SCL) to determine substrate specificity of active recombinant asparaginyl endopeptidase from *Schistosoma mansoni*. This work is intended to biochemically characterize schistosome legumain as a first step in elucidating its biological role and in the development of new antiparasitic chemotherapy.

Specific Aim 1: Identification, cloning and expression of the schistosome asparaginyl endopeptidase.

The protein product of the Sm32 gene is inactive due to an asparagine residue at the predicted crucial active site position C197 (11). I will perform site-directed mutagenesis of the inactive gene product, (N197C), to show that C197 is necessary for catalysis. Expression of asparaginyl endopeptidase from *S. mansoni* will be carried out using the *Pichia pastoris* protein expression system. Sensitivity to specific inhibitors of asparaginyl endopeptidase will be used to confirm the presence of recombinant Sm32, and detection during purification will be achieved by immunoblotting using a monospecific rabbit antiserum.

Specific Aim 2: Activation of Sm32 and optimization of an assay for protease activity.

Parameters such as temperature, pH, and buffer ionic strength are specific for individual enzymes. I will attempt expression of active recombinant Sm32 by refining parameters using the *Pichia pastoris* protein expression system. Optimization of assay conditions for Sm32 will be necessary as a preliminary step for the subsequent determination of substrate specificity, and for the acquisition of accurate kinetic data.

Specific Aim 3: Synthesis of a positional-scanning synthetic combinatorial library (PS-SCL) of fluorogenic peptide substrates.

It has previously been reported that schistosome asparaginyl endopeptidase has a strict specificity for asparagine at P1 (24). I will use a combinatorial synthesis approach to generate a library with asparagine fixed at P1 and the P2 and P3 positions positionally randomized. Mass spectrometry will be used to analyze product identity and purity.

Specific Aim 4: Screening of the library and characterization of substrate specificity.

Substrate specificity will be determined using the PS-SCL. To rule out any bias in the library, I will screen it using an enzyme with specificity that is known to be

II. Identification of a cDNA encoding an asparaginyl endopeptidase from *S. mansoni* and its expression in *Pichia pastoris*

Materials and Methods

II. 1. Cloning and sequencing of C197 Sm32, and site-directed mutagenesis

Poly(A)+ mRNA was isolated (Pharmacia) from 6-week-old *S. mansoni* worms of mixed sex and converted to single-stranded cDNA using AMV Reverse Transcriptase as described by the manufacturer (Life Technologies). PCR amplification of the C197 Sm32 gene was accomplished with primers designed to encompass the pro- and C-terminal regions of the endopeptidase: forward: 5'-ATACAATTAGATACAAATTATGAAGTATCC-3' and reverse: 5'-AATACCGCAAATTTTTATGATTGCT-3'. For eventual cloning into the *P. pastoris* expression vector pPIC Z A, the forward primer incorporated *Xho*I endonuclease and Kex 2 endopeptidase sites immediately 5' of the gene-specific sequence. The reverse primer incorporated a transcription termination codon and a *Not*I site immediately 3' of the gene-specific sequence. Amplification reactions were performed with Pwo polymerase (Boehringer) with 35 cycles of 94°C, 55°C and 72°C, each for 1 min. PCR products were purified (Qiagen) and digested with *Xho*I and *Not*I. pPIC Z A was similarly digested and ligated to the PCR products. For sequencing, the same PCR products were amplified using the above PCR conditions and cloned into pCR-Blunt II-TOPO according to the manufacturer's (Invitrogen) instructions. A unique *Afl*III restriction site was found for

C197 Sm32 (see below) and was used to distinguish such clones from those of N197 Sm32. The signal peptide sequence of C197 Sm32 was obtained from individual Blunt II-TOPO clones that had been generated using a forward primer directed to the signal peptide (5'-ATGATGCTATTCTCTTTATTTCTTA-3') and the above-described reverse primer. PCR conditions to generate such clones were as described above. The entire open reading frame (ORF) of C197 Sm32 was sequenced in both directions. To generate the N197C mutant Sm32, site-directed mutagenesis was performed by sequential PCR steps (24) using the forward primer 5'-TTATATTGAAGCATGAATCAGG-3' and the appropriate anti-parallel reverse primer. This mutation resulted in the same *Afl*III site as found in C197 Sm32. All recombinant plasmids were propagated in DH5 cells (Life Technologies) and purified (Qiagen).

II. 2. Expression of Sm32 in *P. pastoris*

Procedures to express Sm32 in the X33 strain of *P. pastoris* were as detailed by the manufacturer (Invitrogen). Briefly, recombinant pPIC Z A (10 μg) was linearized with *Sac*I and purified. *P. pastoris* was electroporated at 1.5 kV and 129 Ω in electroporation cuvettes (2 mm gap; BTX). *Pichia* colonies growing under zeocin (Invitrogen) selection were picked for expansion first in 10 ml of yeast extract peptone dextrose medium containing 100 μg/ml zeocin and then in 200-400 ml of the same medium. Expression of recombinant protein was induced by incubation of *Pichia* for 48 h in buffered-minimal medium containing 1% methanol as the sole carbon source. Culture medium was clarified through a 0.45 μm filter, lyophilized and stored at 4°C.

Lyophilized culture media containing recombinant Sm32 were resuspended in 0.1 M citrate-phosphate, pH 6.8 to a volume one-tenth of that of the original. Complete buffer exchange was accomplished using PD10 desalting columns (Pharmacia). Solutions were concentrated 10 times in Centricon 10 centrifugation units (Millipore) and stored at -20°C.

II. 3. Detection and activation of recombinant Sm32, and endopeptidase assays

Recombinant Sm32 was detected by immunoblotting after SDS-PAGE (4-12% gradient gels; Novex) using a monospecific rabbit antiserum developed to recombinant N197 Sm32 expressed in *Escherichia coli* (Ewald Beck, Biochemisches Institut, Universität Giessen, Germany). For activation experiments, enzyme solution was incubated for different time periods at 37°C in the presence of 2 mM dithiothreitol (DTT) (pH was measured to be 6.8) or 0.2 M sodium acetate pH 4.5 and 2mM DTT. The following endopeptidase inhibitors (final concentrations were 14 μM unless otherwise indicated) were tested for their ability to prevent activation of Sm32: iodoacetic acid (IA; 5 μM), E-64, morpholinourea-Phe-homoPhe-vinyl sulfone phenyl, PMSF (2 mM) leupeptin, pepstatin A and 1,10-phenanthroline (2 mM). Endopeptidase activity was monitored by hydrolysis of benzyloxy (Z)-Ala-Ser-Asn-7-amido-4-methylcoumarin (NMec; custom synthesis, Enzyme Systems Products). In our hands, Sm32 degrades this substrate approximately twice as fast as the more commonly used Z-Ala-Ala-Asn-NMec (Caffrey, unpublished observation). Assays were performed in black 96-well microtiter plates and contained 5-30 μl enzyme solution preincubated for 10 min in 100 μl 0.1 M

citrate-phosphate, 2 mM DTT, pH 6.8. Substrate (stocks in dimethyl sulfoxide) in 100 µl of the same buffer (10 µM) was then added and the reaction allowed to proceed for 20-30 min at room temperature. Fluorescence was measured in a Labsystems Fluoroskan II plate reader at excitation and emission wavelengths of 355 and 460 nm, respectively.

Results

II. 4. A second gene encodes Sm32 with cysteine in the active site

Heretofore, only one complete gene sequence has been identified *S. mansoni* asparaginyl endopeptidase (Sm32) (7, 22). This was predicted to encode an inactive enzyme as the active site C197 had been replaced by an asparagine residue (11). Consistent with this prediction, our preliminary attempts to obtain endopeptidase activity from an N197 clone expressed in *P. pastoris* failed. Subsequent isolation and sequencing of a number of clones derived from *S. mansoni* mRNA identified a second gene for Sm32 where C197 was present. The entire ORF is presented in Figure 3. C197 Sm32 has 97.2 and 41.0% identity to the previously published N197 Sm32 (22) and human legumain (21), respectively (Figure 4). C197 Sm32 differs from N197 Sm32 by 12 amino acid residues. Also, C197 Sm32 has two potential *N*-glycosylation sites whereas N197 Sm32 has three, one of which, interestingly, is found at position 197.

II. 5. Expression of Sm32 forms in *P. pastoris* and activation of C197 Sm32

Both N197 and C197 Sm32 were expressed and detectable in *Pichia* cultures using an anti-N197 Sm32 monospecific rabbit serum (Figure 5). In fresh preparations of the N197 Sm32, a single immunoreactive protein was detected at 50 kDa (Figure 5a). For C197 Sm32; predominant protein bands at 50 and 47 kDa were detected with minor species at 42 and 39 (Figure 5b). Very little hydrolysis (<5%) of the fluorogenic

substrate Z-Ala-Ser-Asn-NMec could be detected in fresh C197 Sm32 preparations (Figure 6). Significant endopeptidolytic activity was only evident on incubation of C197 Sm32 preparations at pH 4.5 and correlated with the appearance of a final protein product at 34 kDa (Figure 5b, 2h time point onwards). Prolonged incubation of C197 Sm32 at pH 6.8 did not evoke increased hydrolysis of the fluorogenic substrate from the 5% basal level (Figure II.4) nor conversion of the major 50 or 47 kDa proteins (not shown), thus indicating the acid pH requirement for conversion to the fully active enzymatic form. In contrast to C197 Sm32, N197 Sm32 was incapable of being activated, either at pH 4.5 (Figure 5) or at pH 6.0 (not shown), and did not hydrolyze Z-Ala-Ser-Asn-NMec (Figure 6). These results support the prediction based on sequence homology (10) that Sm32 with N rather than C at position 197 is inactive.

II. 6. Activation of Sm32 requires Cys at position 197 and is an autocatalytic event

Confirmation that the endopeptidolytic activity of Sm32 requires Cys at position 197 was obtained by recombinant expression of N197 Sm32 where the active site Asn at 197 had been mutated to a Cys. Like C197 Sm32, the N197C mutant was capable of being activated, but only at pH 4.5 (Figure 6 and 7) and not at pH 6.8 (not shown). Also, increased hydrolysis of Z-Ala-Ser-Asn-NMec was only associated with the appearance of the final protein product which resolved at 37 kDa (Figure 7, 2h time point and onwards). Tests with endopeptidase inhibitors argue against the possibility that an endopeptidase(s) other than Sm32 in the *Pichia* culture medium activates Sm32 in *trans*. Only the general cysteine endopeptidase inhibitor IA blocked activation of both C197 (Figure 5b) and

N197C mutant Sm32 (Figure 7). Importantly, inhibitors of papain-like cysteine endopeptidases such as E-64, morpholino-Phe-homoPhe-vinyl sulfone phenyl (Figure 5b) and leupeptin (not shown) did not prevent conversion to the final enzyme forms. E-64 has previously been shown to be relatively ineffective against asparaginyl endopeptidase activity in schistosome extracts (24) and pig legumain (21). Likewise, inhibitors of serine (PMSF and leupeptin), aspartic (pepstatin A) and metallo-endopeptidases (1,10-phenanthroline) were ineffective (not shown).

II. 7. The function of the C-terminal extension in Sm32

The immunoblot and enzymatic experiments herein demonstrate that maximal endopeptidolytic activity of C197 and N197C mutant Sm32 is associated with the autocatalytic processing of an approximately 50 kDa protein species to about 35 kDa. Consistent with plant and animal legumains (7, 21, 26), loss of the C-terminal extension as well as the N-terminal pro-domain would account for this conversion. Given the specificity of asparaginyl endopeptidases like Sm32 for Asn at P1, the processing sites for removal of the N-terminal pro-domain and the C-terminal extension of Sm32 as previously proposed (14) (Figure 3) must now be re-evaluated. The identification of a 50 kDa Sm32 precursor and a 35 kDa mature protein is consistent with the detection of native asparaginyl endopeptidase species in *S. mansoni* extracts (19). It is not clear, however, whether the conversion to the fully active enzyme in the worm is a *cis* or *trans* event. The C-terminal extension may perform a function similar to that of the N-terminal

pro-region of papain-like endopeptidases such as cathepsin B and L (27) by inhibiting the active site until activity is required by the cell.

II. 8. Summary

Given the difficulty in obtaining large numbers of schistosomes, it is not practical to isolate proteases directly from parasite extracts for analysis of their biochemical characteristics. I have optimized expression of active recombinant asparaginyl endopeptidase from *S. mansoni* using the *Pichia pastoris* expression system. Two distinct cDNA's with 97.2% identity have been identified that encode for asparaginyl endopeptidase homologues from *S. mansoni.* The first cDNA (N197 Sm32) encodes an asparagine residue at position 197 (*S. mansoni* numbering), where a crucial active site cysteine resides in active asparaginyl endopeptidases. The second cDNA contains the requisite C197 residue (C197 Sm32). Both gene products expressed in *Pichia pastoris* produce an inactive, immunoreactive 50kDa protein of 429 amino acids (22). The mature glycoprotein has an apparent molecular mass of 32kDa.

After successfully producing both recombinant Sm32 species, I have shown by site-directed mutagenesis that C197 is required for protease activity. Incubation at acidic pH for 2.5 hours in the presence of DTT converts C197 to the fully active form by autocatalysis. The N197 form, however, remains inactive. Activity is determined by the ability of legumain to hydrolyze the fluorogenic substrate Z-Ala-Ala-Asn-NHMec. The optimum pH for activity is 6.8 and the optimum temperature is in the range of 37-45°C. N-ethylmaleimide and iodoacetamide inhibit legumain activity, whereas the broad-

III. 1. Determination of substrate specificity of schistosome legumain

The substrate specificity of a protease provides clues to its function *in vivo*. Legumains have a strict specificity for hydrolysis of bonds on the carboxyl side of asparagine. The unique specificity of Sm32 for asparagine at P1 may be advantageous for the elucidation of its precise physiological role. Historically, the amino acid sequence flanking the cleavage site of a protein substrate was used to classify proteolytic enzymes and to develop synthetic substrates. However, the sequence of the natural substrate is not necessarily optimal for binding to the enzyme's active site. It has recently been demonstrated that fluorogenic peptide substrates can be used to assay for peptide sequences preferred by a specific enzyme (28). As an alternative to the labor intensive and costly synthesis of individual substrates, combinatorial approaches such as positional-scanning synthetic combinatorial libraries (PS-SCL) have been successfully used to characterize protease specificity (23).

Little is known about the substrate specificity of the legumain family at positions other than P1. Knowledge of specific side chains optimal for binding would help to define possible biological functions and serve as a basis for inhibitor design. I have utilized combinatorial libraries to determine the substrate specificity of Sm32 at P1-P3. After confirming a strict preference for asparagine at P1 using a P1 diverse library, I synthesized a tripeptide positional library where asparagine is fixed in P1 and the P2 and P3 positions are randomized, respectively. Recombinant Sm32 was expressed using the *Pichia pastoris* expression system and profiled for P2 and P3 specificity compared to the corresponding human "legumain".

Low molecular weight synthetic inhibitors such as the vinyl sulfones have been designed to target specific proteases based on their substrate preferences. Vinyl sulfones consist of a peptide portion containing side chains that mimic the preferred substrate, attached to an electrophile that becomes irreversibly alkylated when bound to the active site cysteine (Figure 13a). Inactivation follows when a covalent adduct forms between the C-terminal vinyl sulfone moiety of the inhibitor and the active site cysteine of the protease (Figure 13b). Combinatorial libraries of small-molecule inhibitors are screened to probe the specificities of cysteine proteases (40).

We have obtained a P1-asparagine vinyl sulfone inhibitor library from Dr. Matthew Bogyo (Department of Biochemistry, UCSF). Active recombinant Sm32 was screened using the library to look for inhibition of activity against the fluorogenic substrate Z-Ala-Ala-AsnMCA.

III. 2. Solid phase synthesis of peptide substrate library

Rink Amide resin (0.80 meq/g), PyBOP and Fmoc amino acids were purchased from Novabiochem (San Diego) [Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(*O*-t-Bu)-OH, Fmoc-Glu(*O*-t-Bu)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH, Fmoc-Nle-OH, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Ser(*O*-*t*-Bu)-OH, Fmoc-Thr()-*t*-Bu)-OH, Fmoc-Trp(Boc)-OH, Fmoc-Tyr(*O*-*t*-Bu)-OH, Fmoc-Val-OH]. Anhydrous *N,N*-dimethylformamide was purchased from EM Science (Hawthorne, NY) and *O*-(7-azabenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) from PerSeptive Biosystems (Foster City, CA). Diisopropylcarbodiimide (DIC), 1-hydroxybenzotriazole (HOBt), trifluoroacetic acid (TFA), triisopropylsilane (TIS), *N,N*-diethyl-p-phenylenediamine (DIPEA) and collidine were purchased from Aldrich. The FlexChem Solid Phase Chemistry System used for library synthesis was purchased from Robbins Scientific (Sunnyvale, CA). Recombinant *Schistosoma mansoni* asparaginyl-endopeptidase (Sm32) was prepared as described previously (29). Sm32 was activated at 37°C for 3 h in 100 mM sodium acetate at pH 4.5 containing 1 mM DTT. Human asparaginyl-endopeptidase (hAE) was a gift from Dr. H. Chapman (UCSF). Purified recombinant cathepsin L from *Trypanosoma cruzi* (cruzain) was prepared as described previously (30). The tri-peptide vinyl sulfone library was provided by Dr. Matthew Bogyo (Department of Biochemistry, UCSF).

III. 3. Synthesis of Fmoc-Asp-NHMec

Asp-OH (2.4 mmol) and HATU (2.4 mmol), were dissolved in 15 ml of DMF. Collidine was added (5.7 mmol) and the mixture was stirred for 10 minutes. 7-(4-methyl)-coumarylamide (NHMec, 4.2 mmol) was added as solid and the mixture was stirred at room temperature overnight. Water was added dropwise to precipitate the product. The whitish-yellow precipitate was extracted three times with ethyl acetate, washed two times each with sodium carbonate and sodium chloride, and extracted with 0.1 N HCl. The product was dried over magnesium sulfate and filtered. After rotoevaporation, the product was purified on a silica gel column and dried to give 2.8 g of cream-colored product. The identity of the product was confirmed by mass spectrometry.

III. 4. Synthesis of P1-substituted NHMec-Resin

Gram quantities of NHMec-resin with asparagine in P1 were synthesized batchwise by the following method. Rink Amide resin (70 μmol) was swelled in 0.5 ml DMF and deprotected with 20% piperidine in DMF with agitation for 15 minutes. The resin was washed with DMF (5x with 0.5 ml) and was combined with Fmoc-Asp-NHMec (140 μmol), PyBOP (140 μmol), HOBT (140 μmol), and DIPEA (280 μmol). This reaction mixture was agitated overnight, filtered, washed with DMF, and the coupling reaction procedure was repeated a second time. The product was washed (DMF, 3x with 0.5 ml; MeOH, 3x with 0.5 ml). The identity of the product was confirmed by mass

spectrometry. The efficiency of first residue attachment was estimated at 4.2 mmol/g by Fmoc absorbance at 290 nm.

III. 5. Synthesis of P1-Asn-NHMec library

P1-Asn-NHMec-resin was added to 38 wells of a Flexchem 96 well reaction apparatus (~20 mg, 20 μmol). The resin was swelled with 0.5 ml DMF, filtered, and deprotected with 20% piperidine in DMF for 20 minutes. The wells were then filtered and washed (DMF, 3x with 0.5 ml, MeOH, 2x with 0.5 ml) in preparation for the P2 coupling reaction. Each of nineteen amino acids (cysteine was omitted and norleucine was substituted for methionine) were added to wells 1-19 to introduce a fixed P2 position (5 eq each of Fmoc-amino acid, DIC, and HOBt in DMF). In wells 20-34, an isokinetic mixture of Fmoc-amino acids was added to randomize the P2 position in these reactions. Coupling was achieved using 5 eq of HOBt and DIC per well following a previously described protocol (14). The reaction block was agitated for 4 hours, filtered, and washed with DMF and MeOH as above. Deprotection of each reaction was repeated using 20% piperidine in DMF, followed by washing with DMF and MeOH as above. The randomization of P3 in wells 1-19, and fixing of P3 in wells 20-38 was achieved by reversing the protocol described above. The reaction block was agitated for 4 hours and the wells were washed with DMF, MeOH, and methylene chloride (2x with 0.5 ml). The resulting tripeptides were capped using a solution (150 μl/well) of acetic anhydride (400 μl), pyridine (315 μl) and DMF (6.0 ml). The block was agitated for 4 hours and then washed with DMF, MeOH and CH_2Cl_2 as above. To cleave each acylated tripeptide

from the resin, a solution (95:2.5:2.5 TFA/TIS/H_2O, 0.5 ml/well) was added and the block was agitated for 1 hour. The reaction block was then transferred to a deep-well collection chamber and the tripeptides were removed by filtration under reduced pressure with addition of remaining cleavage solution. Contents from each well were transferred to 15 ml falcon tubes and precipitated with ether. Ether was evaporated off and the remaining contents were lyophilized. The amount of tripeptide substrate recovered from each reaction was estimated to be 0.005 mmol on the basis of single substrate yields. Individual substrates were dissolved in DMSO to provide 10 mM stocks for enzymatic assays.

III. 6. Kinetic Assay of Library

Enzymatic assays were carried out in opaque 96-well microtiter plates from Corning Inc. (Corning NY) using a Labsystems Fluoroskan II and Deltasoft 3 software. Each member of the substrate library was assayed in triplicate at a final concentration of 50 μM/well against activated enzyme (20 μl/well) in 100 mM sodium acetate at pH 4.5 containing 1 mM DTT and 1 mM EDTA. Hydrolysis of substrate was monitored fluorimetrically with excitation at 355 nm and emission at 460 nm based on the fluorescence properties of free AMC.

III. 7. P1-diverse library synthesis and assay

Individual P-1 substituted Fmoc amino acid resin was prepared as previously described (31). 7-amino-4-carbamoylmethylcoumarin (ACC) was used as a fluorescent-leaving group in place of AMC. Randomized P_2, P_3, and P_4 positions were incorporated by addition of the isokinetic mixture of 19 amino acids as described above. Substrates from the P1-diverse library were added to 19 wells of a 96-well Microfluor plate (Dynex Technologies, Chantilly, VA) for a final concentration of approximately 0.1 μM. Hydrolysis reactions were initiated by addition of enzyme and monitored on a Perkin-Elmer LS50B luminescence spectrometer with excitation at 380 nm and emission at 460 nm. The synthesis and assay of this library was kindly carried out by Dr. Youngchool Choe (UCSF).

III. 8. Single Substrate Kinetic Assays

Individual tripeptide substrates were obtained as custom syntheses from Enzyme Systems Products, Livermore, CA. The following custom substrates were used for kinetic analysis: Z-Ala-Ala-Asn-NHMec, Ac-Thr-Ala-Asn-NHMec, Ac-Pro-Thr-Asn-NHMec, and Ac-Phe-Tyr-NHMec. Enzyme activity was monitored at 25°C in assay buffer of .50mM sodium acetate, pH 5.5, containing 2mM DTT and 2mM EDTA. The final concentration of substrate ranged from 0.006 to 0.4mM. The DMSO concentration in the assay was less than 5%. Pre-activated enzyme solution was added to each well (100μl total/well). Release of AMC was monitored fluorometrically with an exitation

wavelength of 380nm and emission wavelength at 460nm. GraphPad Prism software (San Diego, CA.) was used to fit the data to a non-linear regression curve. Vmax and Km were determined using a Michaelis-Menten plot with initial enzyme velocity vs. substrate concentration.

III. 9. Screening of Vinyl Sulfone Library

Recombinant Sm32 (100μl/well) was activated as previously described and pre-incubated with each of 38 vinyl sulfone inhibitors (50 μM/well) in a 96-well Microfluor plate. After 1 hour of incubation, 1 μl of 10 mM substrate was added to each well and activity was measured on a Perkin-Elmer LS50B luminescence spectrometer with excitation at 380 nm and emission at 460 nm.

Results

III. 10. Design and construction of the peptide library

As a member of the asparaginyl endopeptidase group, Sm32 was predicted to have specificity for cleavage on the carboxyl side of asparagine residues (16,17). Strict specificity for cleavage after asparagine was confirmed using a P1 diverse library (Figure 8). The fluorogenic tri-peptide substrate library with the general structure Ac-X-X-Asn-amidomethylcoumarin (Figure 9) was therefore generated to determine the substrate specificity of *S. mansoni* asparaginyl endopeptidase. The synthetic method used to fix asparagine in P1 is diagrammed in Figure 10. Randomization of P2 and P3 positions respectively, was achieved using a method that ensures equal representation of all amino acids of an isokinetic mixture. (32). As some amino acids typically provide low coupling yields, each coupling was repeated to increase substitution levels. Substitution levels of amino acid after each coupling were monitored by spectrophotometric quantitation of the fulvene-piperidine adduct resulting from piperidine deprotection of the Fmoc protecting group.

III. 11. Screening of the peptide library

Cleavage by a protease of Ac-X-X-Asn-amidomethylcoumarin at the scissile bond liberates the fluorescent-leaving group, AMC, for sensitive and convenient detection of proteolytic activity by fluorimetry. Successful activation of schistosome

legumain was first confirmed by its ability to cleave the authentic substrate (Z-Ala-Ala-Asn-NHMec), and by inhibition of enzyme activity using iodoacetamide (29).

The schistosome enzyme showed a preference for P2: Thr>Ala>Val>Ile, and P3: Ala>Thr>Val>Asn, with an overall broader specificity at P3 than at P2 (figure 11). Neither schistosome nor human legumain accommodated His or Tyr in P2 or P3, and activity is low against peptides containing Asp, Glu, Gln, Phe or Trp. As a confirmation of library diversity and to test its reliability as a measure of substrate specificity, four different enzymes were profiled using the P1-Asn library (schistosome legumain Sm32, schistosome legumain N197C, human legumain hAE, and cruzain from *Trypanosoma cruzi*). Legumain N197C is a recombinant product of a gene isoform identified in *S. mansoni*. The wildtype gene has Asn at position 197. This form is inactive, but the mutant N197C regains catalytic function (29). The results for the schistosome enzymes versus the peptide library were nearly identical as expected (Figure 11). Cruzain, a clan CA cysteine protease has a relatively broad P1 substrate preference and is able to cleave peptides with Asn in P1 (31). However, its P1 and P2 specificities are known and are very different from legumain family enzymes. Cruzain was shown to have a strict specificity for the P2 position (Leu>Phe>Arg), and a relatively relaxed specificity at P3 (Leu>Lys,Arg,Gly>Phe,His,Ala). These results are consistent with previously reported data using peptide substrates (33), as well as structural data with peptide inhibitor co-crystals (34).

The library was also screened using hAE to compare its specificity directly to that of the enzyme from *Schistosoma mansoni*. Both human and schistosome enzymes accommodate threonine and alanine well in P2 and P3, however there are some notable

differences between their optimal substrate sequences. In particular, Thr-Ala-Asn is optimal for schistosome legumain, Pro-Thr-Asn is optimal for human legumain.

III. 12. Single Substrate Kinetic Assays

As a further confirmation of the results of the peptide library screen, single tripeptide AMC substrates predicted by PS-SCL to be good (AAN, TAN, PTN), or poor (FYN), were assayed versus schistosome and human legumains. The results of kinetic analysis shown in Table 1 are consistent with the predictions made by the PS-SCL. Ac-Thr-Ala-Asn-NHMec is preferred slightly over Z-Ala-Ala-Asn-NHMec by the schistosome enzyme. The substrate Ac-Pro-Thr-Asn-NHMec, which is preferred by the human enzyme, was disfavored by Sm32, indicating the possibility of designing selective inhibitors based on discrete biochemical properties of these two enzymes. Both enzymes, however, were able to cleave peptides Z-Ala-Ala-Asn-NHMec and Ac-Thr-Ala-Asn-NHMec effectively. As predicted by the PS-SCL, neither enzyme showed activity versus Ac-Phe-Tyr-Asn-NHMec.

III. 13. Screening of the Vinyl Sulfone Library

The results of the inhibitor screen are in agreement with the predictions made by the substrate library (Figure 11 and 13c). The most effective inhibitors of Z-Ala-Ala-Asn-NHMec cleavage contained the following amino acid residues: P3 position:

IV. Discussion and future directions

Asparaginyl endopeptidases, or 'legumains' are a recently identified group of cysteine endopeptidases. The active site mechanism is distinct from members of the papain family enzymes (cathepsins B and L). Legumains represent one of four protease families in clan CD. Families C11, C13, C14 and C25 have been grouped together based on amino acid sequence and catalytic dyad motif (11). Compared with other cysteine proteases of broader specificity, members of clan CD have a strict requirement for a specific side chain of the P1 amino acid residue. Legumain is the only protease requiring asparagine at P1.

With examples of legumain activity in plants, schistosomes and mammals, there appears to be a conserved need for the hydrolysis of asparaginyl bonds. Research to date describes a function for legumains in the limited proteolysis of other proteins. Plant, helminth and mammalian legumains have similar inhibition profiles, and biological roles in post-translational modification of other proteins or enzymes (11,12,24). Schistosome legumain was first detected as an immunogenic component of soluble schistosome extracts, and was initially considered to be the *S. mansoni* hemoglobinase. The identity of Sm32 was clarified when its nucleotide sequence was found to be closest to an asparaginyl endopeptidase from seeds of the jack bean *Canavalia ensiformis*. Localization of Sm32 in the gut epithelium indicates that it plays a role in schistosome nutrition. However, as in plant legumains, schistosome legumain may process and activate other proteolytic enzymes directly involved in Hb digestion. The exact role of

schistosome legumain remains unknown. As a member of a novel family of cysteine proteases with unique specificity, legumain attracted my interest and motivated my efforts to elucidate the function of asparaginyl endopeptidase from *S. mansoni.*

The *Pichia pastoris* eukaryotic expression system produced correctly folded, active Sm32 in substantial quantities for biochemical characterization. The data presented here revealed that, in addition to the previously identified, inactive form of asparaginyl endopeptidase from *S. mansoni*, there exists an active form of the enzyme. The redundancy of the enzyme in schistosomes may be the result of a gene duplication event. Based on primary amino acid sequences and substrate specificities reported here, the enzymes do not appear to be divergent. Why an inactive gene is kept in the schistosome genome is unknown. The product may serve as a binding protein or even a competitor of the active protease for substrate.

Anti-parasitic research has focused on cysteine proteases as targets for synthetic protease inhibitors. Experimental models using both reversible and irreversible cysteine protease inhibitors have shown that they are effective as treatment against parasite infection (38). The utility of vinyl sulfone-derivatized inhibitors has been demonstrated in murine models of malaria, Chagas' disease and leishmaniasis (36, 37, 39). The inhibitor acts like a "suicide substrate" for the active site nucleophile. The efficacy of such inhibitors relies on the vulnerability of the parasite enzyme compared with that of the host. Although there are mammalian orthologues of the target proteases, there is evidence that localization, concentration and function are often redundant enough to minimize toxicity of protease inhibitors for the host (36). Because mammalian legumain is predominantly lysosomal and unstable at neutral pH, it is not likely to have an

extracellular function (11). Recombinant Sm32 has a near neutral pH optimum of 6.8, and as a bloodstream organism, it is more vulnerable to inhibitors than the homologous host proteases that are in host cell lysosomes. Additionally, the concentration of host proteases may significantly exceed the concentration of comparable parasite targets. Therefore, Sm32 represents a potential target for chemotherapuetic intervention.

Considering the lack of structural data for Sm32, combinatorial chemistry was used to assay for substrate preferences that could be used to design protease inhibitors. As a first step in the determination of substrate specificity for the human and *S. mansoni* asparaginyl endopeptidases, we confirmed the P1-Asn specificity by screening a P1 diverse library (figure 8). We then constructed a positional scanning synthetic combinatorial library (PS-SCL) with asparagine fixed at P1 and the P2 and P3 positions randomized using 19 amino acids. The results of the library screen suggested a similar specificity at P2 and P3 for the human and schistosome legumains. However, a notable difference was the preference for Pro at P3 for the human enzyme. Differentiation of the parasite enzyme from that of the host at P3 supports the possibility of selective inhibition. Based on our results, the P2 and P3 specificities appear to be relatively relaxed, indicating that specificity may be determined by residues on the prime side of the sissile bond, or by structural characteristics of the natural protein substrate.

To validate the results of the PS-SCL, and confirm the diversity of the peptide library, we screened cruzain from *T. cruzi* for P2-P3 substrate preference. The substrate preference of cruzain is known, and is different from the legumains (31,34). Indeed, a very different and consistent profile for cruzain was found, suggesting that any bias in the peptide library is minimal (Figure 11). Further confirmation of the PS-SCL was

demonstrated by the Vmax and Km data obtained using purified single substrates (Table 1). The kinetic data is consistent with the predictions of good or poor peptide substrate sequences from the library screen. Finally, the results were also in agreement with previously reported data for porcine legumain (11).

The inhibitor screen (Figure 13c) reveals that a vinyl sulfone electrophile containing a peptide sequence with asparagine, alanine or threonine in P2 or P3 is most effective against Sm32 activity. Although this result is consistent with the peptide library predictions, complete inhibition is not observed using 50 μM concentrations of inhibitor. Vinyl sulfones are intended to function as irreversible inhibitors, and evidence from labeling experiments (not shown) suggest that they act as reversible inhibitors of Sm32. Although vinyl sulfones could not be used as Sm32-specific inhibitors, there are several functional groups capable of irreversibly alkylating the active site cysteine that could be incorporated into peptides. Epoxides, fluoromethyl ketones and acyloxymethyl ketones are examples of electrophiles that may be effective as irreversible inhibitors of Sm32. Radiolabeled substrate analogues are also available as tools for cysteine protease profiling (40).

The general method of using a PS-SCL to determine substrate specificity is rapid and efficient, and provides information for predicting or confirming biological substrates. Manoury, et.al. (17) reported that microbial tetanus toxin antigen was processed by a mammalian asparaginyl endopeptidase (AEP). The digestion products showed that cleavage occurred after asparagine (IDN, PNN, and FNN). One of these cleavage sites (PNN) is consistent with an optimal sequence predicted by the PS-SCL screen. The other two are acceptable but would not be predicted to be optimal. No data is reported as to

which site might be cleaved initially. Dr. M. Sajid (personal communication) reports that *S. mansoni* asparaginyl endopeptidase *trans*-processes and activates recombinant *S. mansoni* cathepsin B (SmCB1). Cleavage occurs at an asparagine residue (DWN) between the pro and mature regions of SmCB1 (Figure 12). This is an acceptable substrate sequence, but again, not an optimal one. These observations, as well as the strict requirement for asparagine in P1 and relatively relaxed preference at P2 and P3, may indicate that residues on the "prime side" of asparagine (35) also determine legumain substrate specificity. Phage display is currently being employed to map the prime side specificity of *S. mansoni* asparaginyl endopeptidase.

Phage display and PS-SCL techniques, both of which can predict the primary structure of a protein substrate, are limited by the lack of consideration for tertiary structure. Tertiary protein structure and folding patterns may strongly influence substrate preferences, and future plans for Sm32 include production and purification of recombinant enzyme for crystallization. Nevertheless, the methods described here provide a starting point from which we can begin to design peptide inhibitors and suggest biological functions for a given protease. Both phage display and PS-SCL approaches represent a vast source of molecular diversity for potential identification of lead compounds.

Schistosomiasis presents a profound human health problem with severe socioeconomic impact on developing countries. With the growing need for anti-parasitic therapies, identification of potential drug targets such as schistosome asparaginyl endopeptidase is critical. The majority of parasite proteases known prior to the discovery of schistosome legumain show broad specificity and gross catabolism of protein

substrates. There is evidence that legumain performs limited proteolysis of other proteins and has strict specificity for cleavage of asparaginyl bonds. Based on these properties, legumain represents a novel target for the development of specific small molecule inhibitors. Biochemical characterization of Sm32 and other potential targets brings medical science one step closer to controlling the global burden of parasite disease.

References

1. Mahmoud, A.A.F., and Wahab, M.F.A. (1990) Tropical and Geographical Medicine (Warren, K.S., and Mahmoud, A.A.F., Eds.) pp.458-473, McGraw-Hill. New York.
2. Salter, J.P., Lim, K.C., Hansell, E., Hsieh, I., and McKerrow, J.H. (2000) *J. Biol. Chem.* **49,** 38667-38673
3. Smith, M.A. (1974) *International Journal of Parasitology* **4,** 681-683
4. Monroy, F.P., Dresden, M.H. (1996) *International Journal of Parasitology* **1,** 109-112
5. Brindley, P. J., Kalinna, B. H., Dalton, J. P., Day, S. R.,Wong, J. Y., Smythe, M. L., and McManus, D. P. (1997) *Mol. Biocem. Parisitol.* **89**, 1-9
6. Wasilewski, M.M., Lim, K.C., Phillips, J., McKerrow, J.H. (1996) *Molecular and Biochemical Parasitology* **81**, 179-189
7. Davis, A.H., Nanduri, J., and Watson, D.C. (1987) *J. Biol. Chem.* **262**, 12851-12855
8. Gotz, B., Klinkert, M.-Q. (1993) *Biochem. J.* **290** 801-806
9. Caffrey, C.R., Rheinberg, C.E., Moné, H., Jourdane, J., Li, Y.L.,Ruppel, A. (1997) *Parasitology Research* **83,** 37-41
10. Brindley, P. J., Kalinna, B. H., Dalton, J. P., Day, S. R., Wong, J. Y., Smythe, M. L., and McManus, D. P. (1997) *Mol. Biocem. Parisitol.* **89**, 1-9
11. Chen, J-M., Rawlings, N. D., Stevens, R., and Barrett, A. J. (1998) *FEBS. Lett.* **441,** 361-365
12. Chen, J-M., Fortunato, M., and Barrett, A. J. (2000) *Biochem. J.* **352**, 327-33

13. Rawlings, N.D. and Barrett, A.J. (1994) *Methods Enzymol.* **244**, 461-486

14. Dalton, J. P., and Brindley, P. J. (1998) in: *Handbook of Proteolytic Enzymes* (Barrett, A.J., Rawlings, N.D., and Woessner, J. F., eds.) pp.749-754, Academic Press, London

15. Kembhavi, A.A., Buttle, D.J., Knight, C.G., and Barrett, A.J. (1993) *Arch. Biochem. Biophys.* **303**, 208-213.

16. Shimada, T., Hiraiwa, N., Nishimura, M., and Hara-Nishimura, I. (1994) *Plant Cell Physiol.* **35**, 713-718

17. Manoury, B., Hewitt, E.W., Morrice, N., Dando, P. M., and Watts, C. (1998) *Nature* **396**, 695-699

18. Brindley, P. J., Kalinna, B. H., Dalton, J. P., Day, S. R., Wong, J. Y., Smythe, M. L., and McManus, D. P. (1997) *Mol. Biocem. Parisitol.* **89**, 1-9

19. El Maenawy, M. A., Aji, T., Phillips, N. F. B., Davis, R. E., Salata, R. A., Malhotra, I., McClain, D., Aikawa, M., and Davis, A. H. (1990) *Am. J. Trop. Med. Hyg.* **43**, 67-78

20. Wasilewski, M. M., Lim, K. C., Phillips, J., and McKerrow, J. H. (1996) *Mol. Biochem. Parasitol.* **81**, 179-189

21. Chen, J.M., Dando, P.M., Rawlings, N.D., Brown, M.A., Young, N.E., Stevens, R.A., Hewitt, E., Watts, C., and Barrett, A.J. (1997) *Journal of Biological Chemistry* **272**, 8090-8098

22. Klinkert, M.-Q., Ruppel, A., Felleisen, R., Link, G., and Beck, E. (1989) *Mol. Biochem. Parasitol.* **33**, 113-122

23. Backes, B. J., Harris, J. L., Leonetti, F., Craik, C. S., and Ellman, J. A. (2000) *Nature Biotechnology* **18**, 187-193

24. Dalton, J.P., Hola-Jamriska, L., and Brindly, P.J. (1995) *Parasitology* **111**, 575-580

25. Cormack, B. (1999) in: *Short Protocols in Molecular Biology* (Ausubel, A., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., Eds.), pp. 8-21 John Wiley, New York

26. Ishii, S., Abe, Y., Matsushita, H., and Kato, I. (1990) *J. Protein Chem* **9**, 294-295

27. Cygler, M., and Mort, J.S. (1997) *Biochimie* **79**, 645-652

28. Rano, T.A., Timkey, T., Peterson, E.P., Rotonda, J., Nicholson, D.W., Becker, J.W., Chapman, K.T., Thornberry, N.A. (1997) *Chemistry and Biology* **2**, 149-55

29. Caffrey, C.R., Mathieu, M.A., Gaffney, A.M., Salter, J.P., Sajid, M., Lucas, K.D., Franklin, C., Bogyo, M., and McKerrow, J. H. (2000) *FEBS. Lett.* **466**, 244-248

30. Eakin, A. E., McGrath, M. E., McKerrow, J. H., Fletterick, R. J., and Craik, C. S. (1993) *J. Biol. Chem.* **286**, 6115-6118

31. Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) *PNAS.* **97**, 7754-7759

32. Furka, A., Sebestyen, F., Asgedom, M., and Dibo, G. (1991) *Int. J. Peptide and Prot. Res.* **37**, 487-493

33. Freymann, D. M., Wenck, M. A., Engel, J. C., Feng, J., Focia, P. J., Eakin, A. E., and Craig, S. P. (2000) *Chem. And Biol.* **12**, 957-968

34. Nguyen, J. T., Inouye, H., Baldwin, M.A., Fletterick, R. J., Cohen, F. E., Prusiner, S. B., and Kirschner, D. A. (1995) *J. Mol. Biol.* **252**, 412-422

35. Schechter, I., and Berger, A. (1968) *Biochem. Biophys. Chem. Commun.* **27**, 157-162

36. Selzer, P.M., Pingel, S., Hsieh, I., Ugele, B., Chan, VJ., Engel, J.C., Bogyo, M., Russell, D.G., Sakanari, J.A., McKerrow, J.H. (1999) *Proceedings of the National Academy of Sciences of the United States of America* **96**, 11015-22.

37. Engel, J.C., Doyle, P.S., Hsieh, I., and McKerrow, J.H. (1998) *Journal of Experimental Medicine* **188**, 725-734

38. McKerrow, J.H., Engel, J.C., and Caffrey, C.R. *Bioorganic and Medicinal Chemistry* **6**, 1-6

39. Rosenthal, P., Olson J.E., Lee, K., and Palmer, J.T. (1996) *Antimicrob. Agents Chemother.* **40** 1600-1603

40. Bogyo, M., Verhelst, S., Bellinggard-Dubouchaud, V., Toba, S., Greenbaum, D. (1999) *Chemistry and Biology* **7**, 27-38

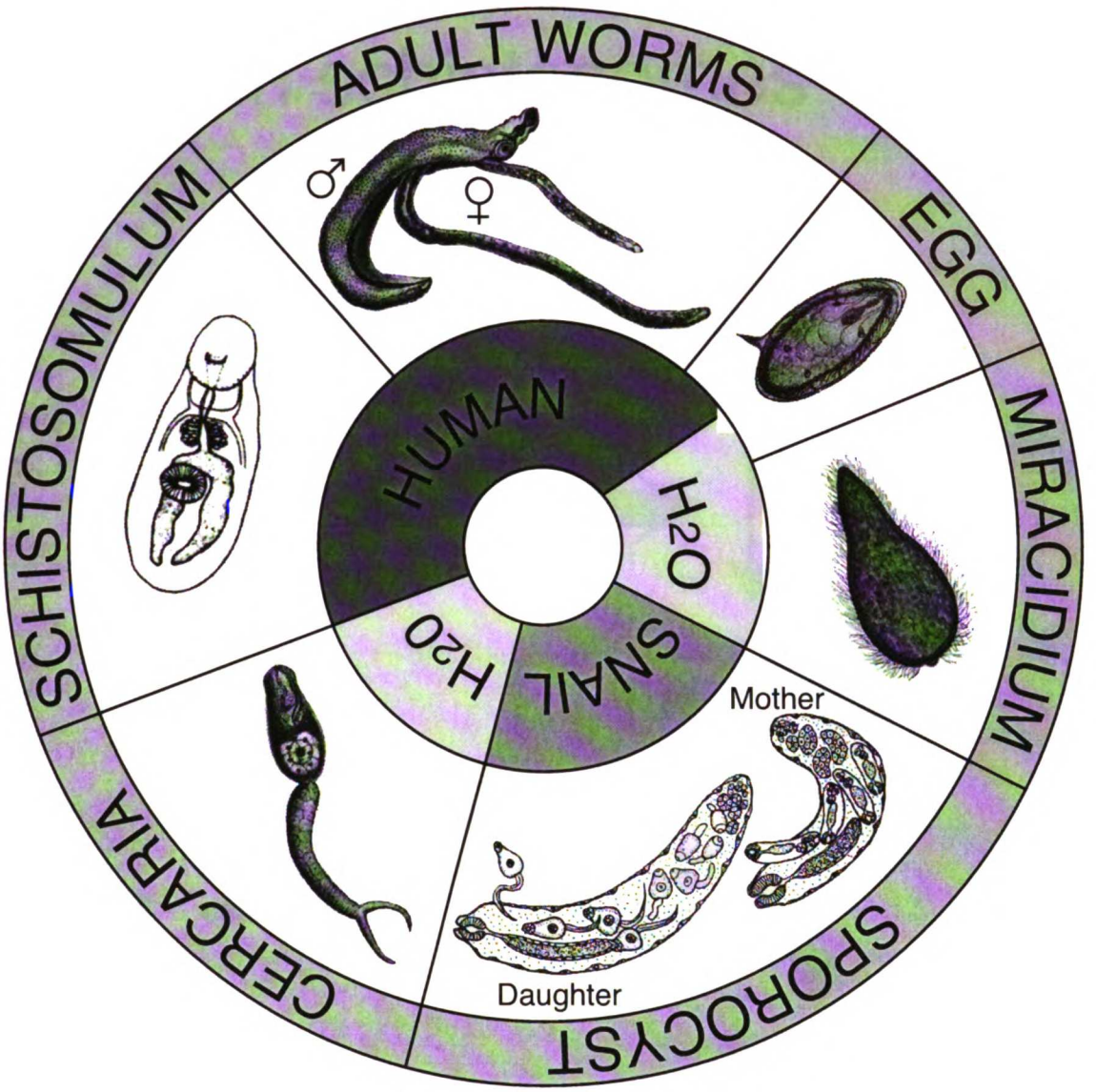


Figure 1 Life cycle of the schistosome

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.

Hemoglobin Degradation in schistosomes

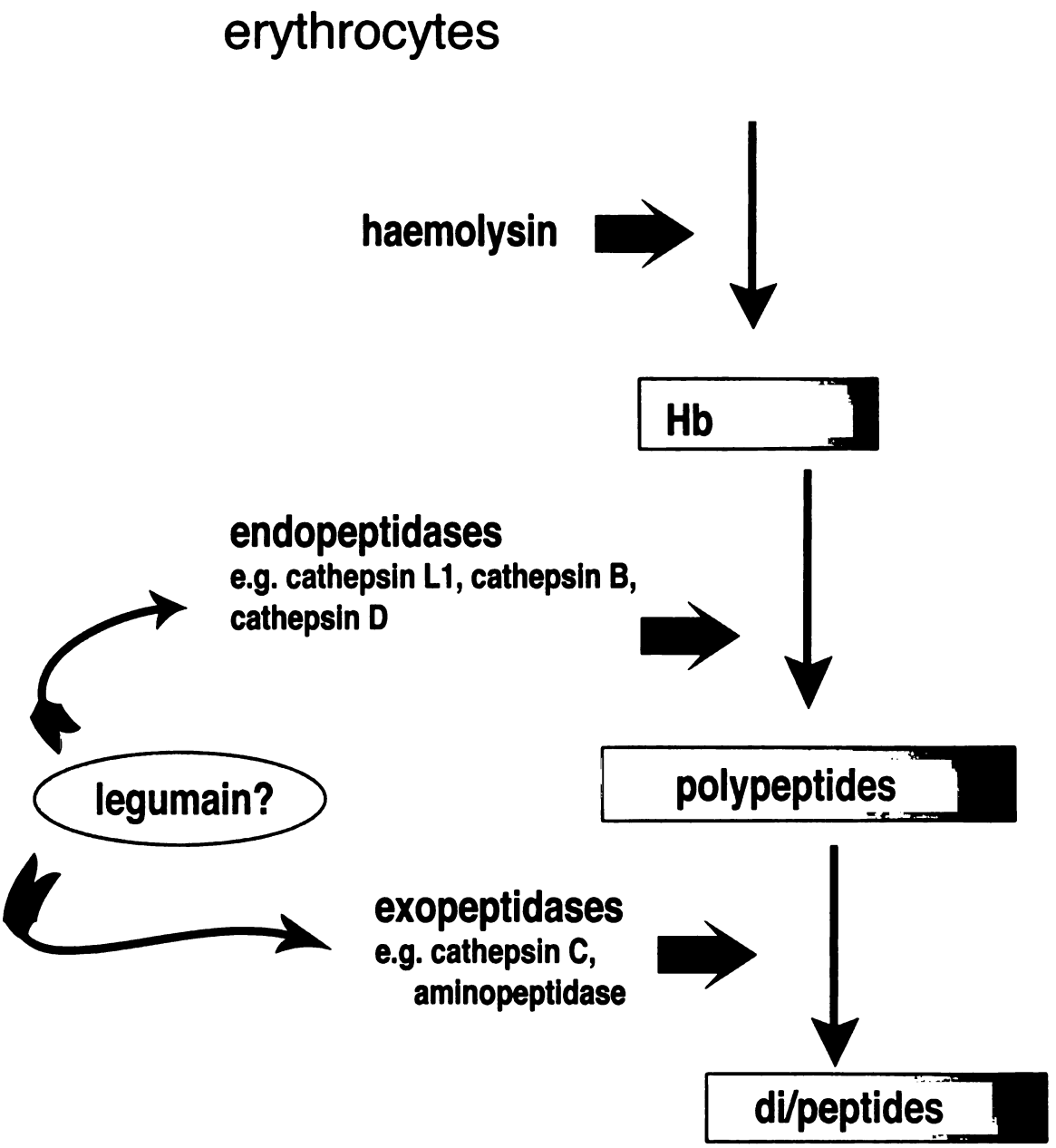


Figure 2 Schematic representation of the schistosome proteolytic pathway and the proteinases hypothesized to accomplish catabolism of host hemoglobin. Schistosome legumain may process other proteinases from inactive to active forms.

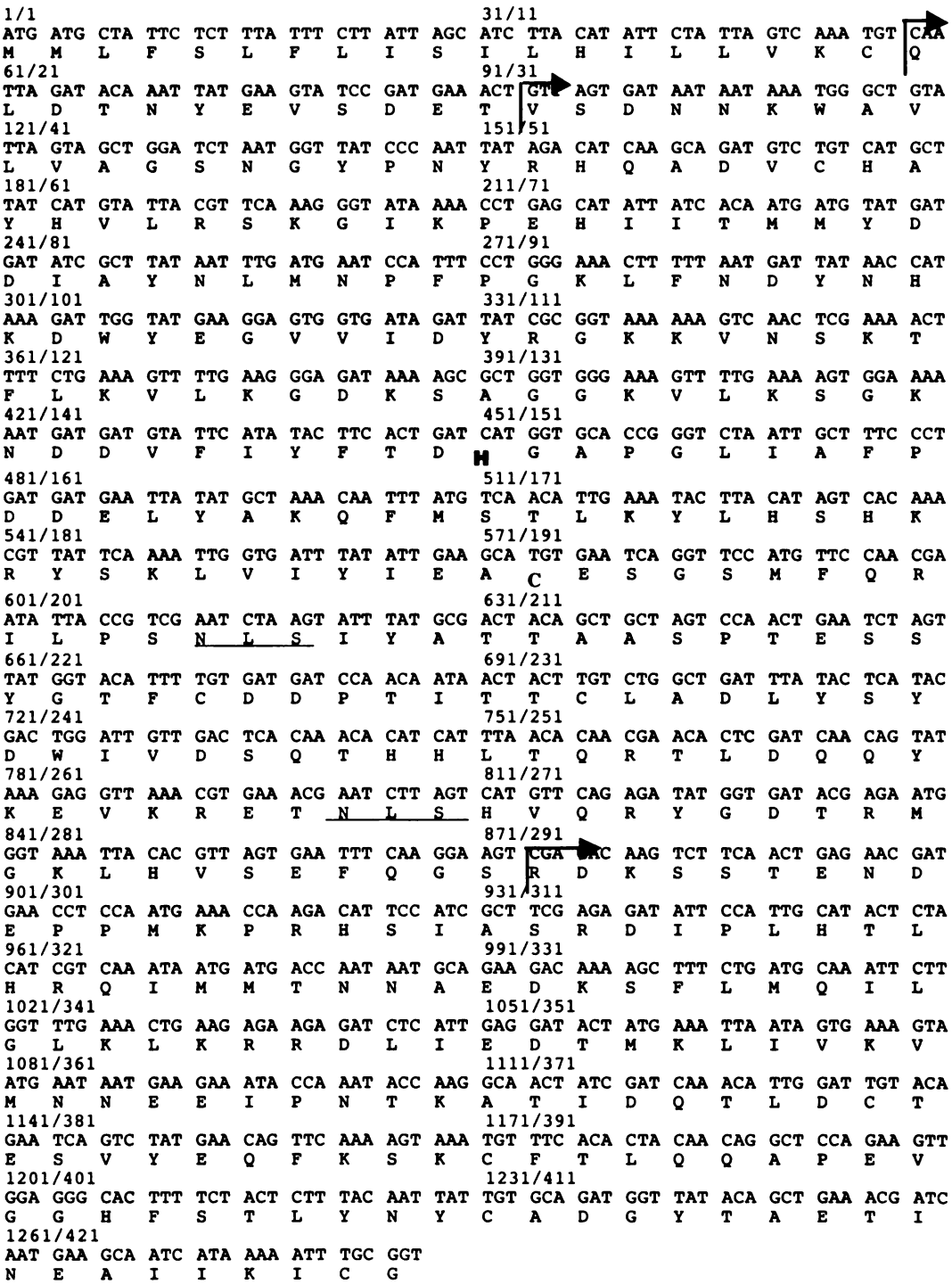


Figure 3 Nucleotide and deduced amino acid sequence of C197 Sm32. Active site histidine and cysteine residues are presented in bold; potential glycosylation sites are underlined and arrows indcate the putative N-termini of the pro-domain, mature protease and C-terminal extension.

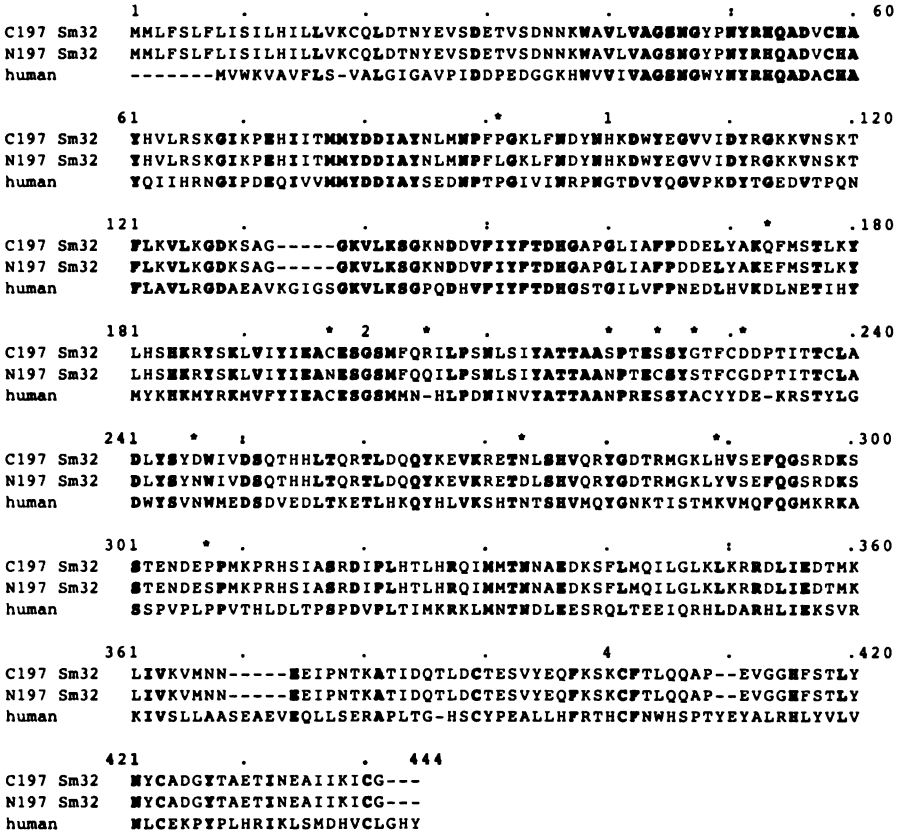


Figure 4 Amino Acid sequence alignment of C197 with N197 Sm32 and human legumain. Amino acid residues differing between the two schistosome sequences are indicated by astericks; areas where all three sequences are identical are shaded.

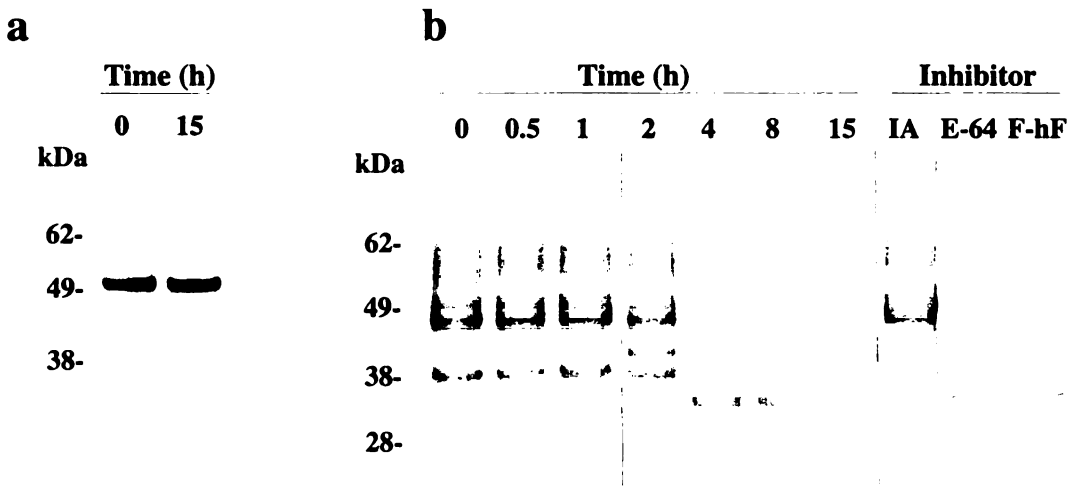


Figure 5 Conversion of C197, but not N197 Sm32, to the final active enzyme product. As described in section 2, Sm32 proteins expressed in *P. pastoris* supernatants were incubated at 37 degrees celsius in the presence of DTT at pH 4.5 for different time periods. After SDS-PAGE and blotting onto PVDF membrane, supernatants were reacted with a monospecific anti-N197 Sm32 rabbit serum. a: N197 Sm32 after incubation at pH 4.5 for 0 and 15 hours. b: as for a, but with C197 Sm32; inhibitors were incubated for 15 hours with C197 Sm32. IA, iodoacetic acid; F-hF, morpholinourea-Phe-homoPhe-vinyl sulfone phenyl.

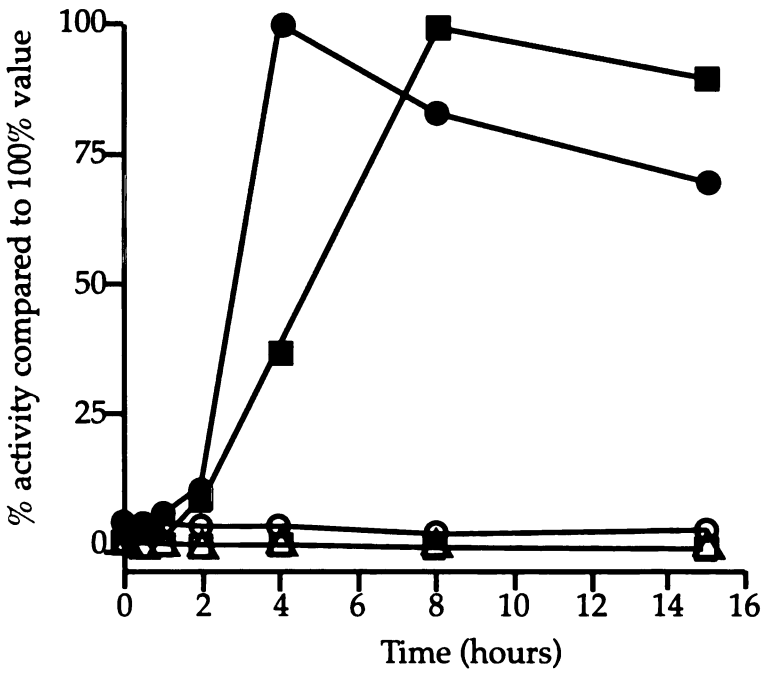


Figure 6 Conversion to final Sm32 products at acid pH is associated with increased endopeptidolytic activity. As described for the above experiments (Figure III.3) N197, C197, or N197C mutant Sm32 was incubated at pH 6.8 or 4.5 for different time periods for up to 15 hours. Endopeptidolytic activity for various Sm32 preparations was measured using the fluorogenic substrate Z-Ala-Ser-Asn-NMec. Incubation at pH 4.5; N197 (△), C197 (●) and N197C mutant (■). Incubation at ph 6.8; C197 (○), and N197C mutant (□).

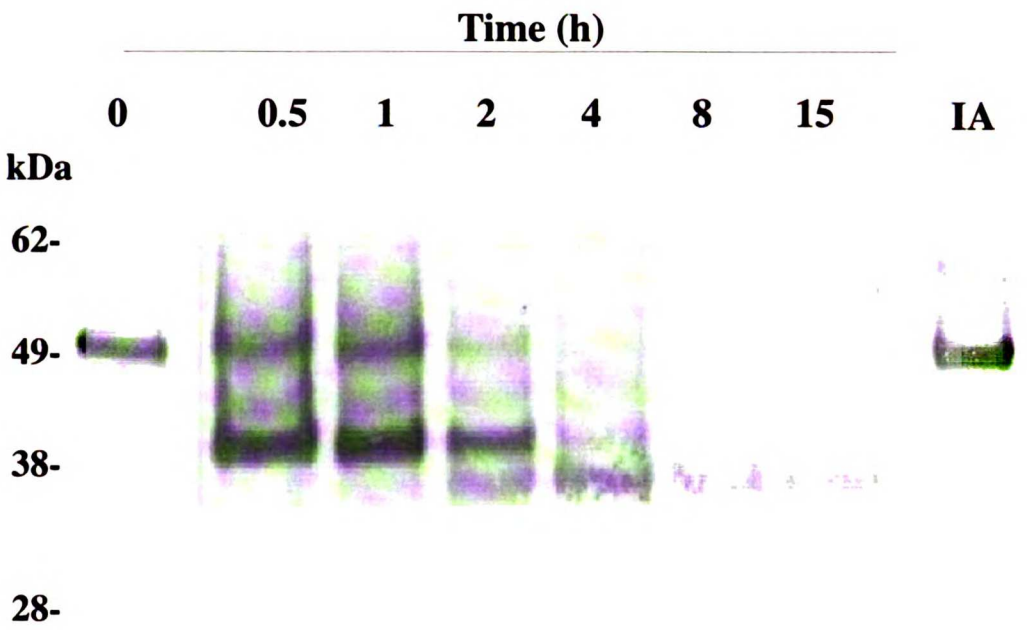


Figure 7 Mutating N197 to C restores the ability of N197 Sm32 to convert to the final active Sm32 product. Experimental design was as for C197 Sm32.

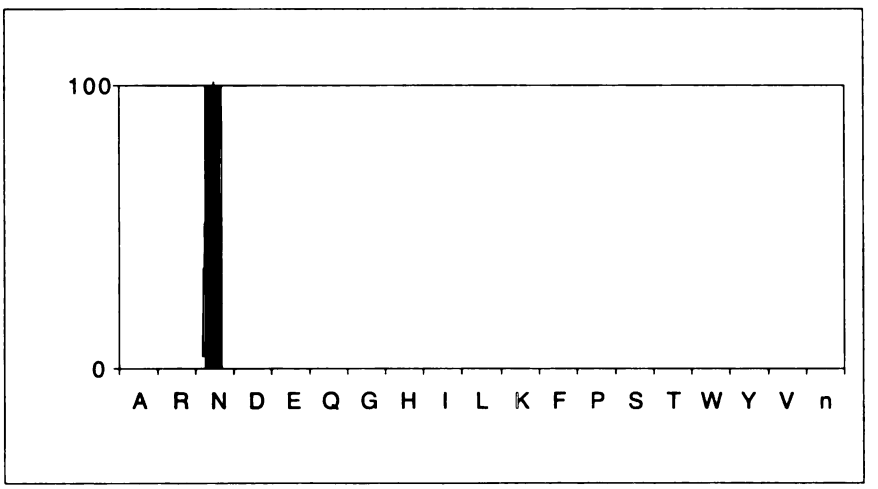


Figure 8
Strict specificity for Asn in P1 of *S. mansoni* asparaginyl endopeptidase is confirmed by screening a P1 variable library. The library consists of 19 wells, each containing a compound with a different amino acid represented at P1, but randomized at P2-P4 with isokinetic mixtures of amino acids. The x-axis provides the special address of the amino acid as represented by the one-letter code (i.e., A=alanine). Enzyme was activated for 2.5 hours at pH 4.5 in 100mM sodium acetate buffer containing 10mM DTT. 100μg of enzyme solution was added to each of the 10 wells. The y-axis represents the rate of AMC production expressed as a percentage of the maximum rate observed in each experiment. Production of AMC was monitored continuously on a Perkin-Elmer LS50B with excitation at 380 nm and emission at 460nm.

Tripeptide PS-SCL

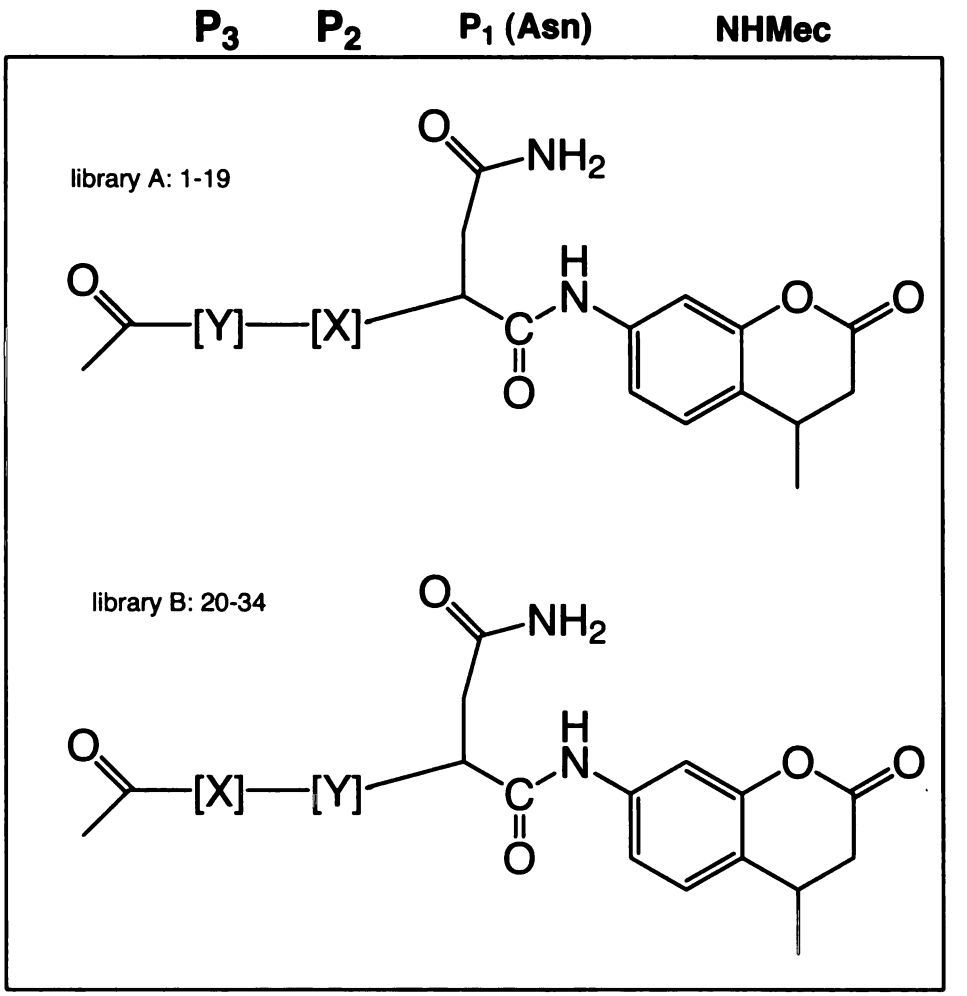


Figure 9

P_1-fixed Asn tripeptide library. The combinatorial positional-scanning library consists of two sublibraries of 19 members each. Libraries are amino-terminal acylated and carboxy-terminal tagged with the fluorophore NHMec. Each library has Asn fixed at P1 where 'X' represents a known amino acid and 'Y' represents an isokinetic mixture of the 19 amino acids indicated previously

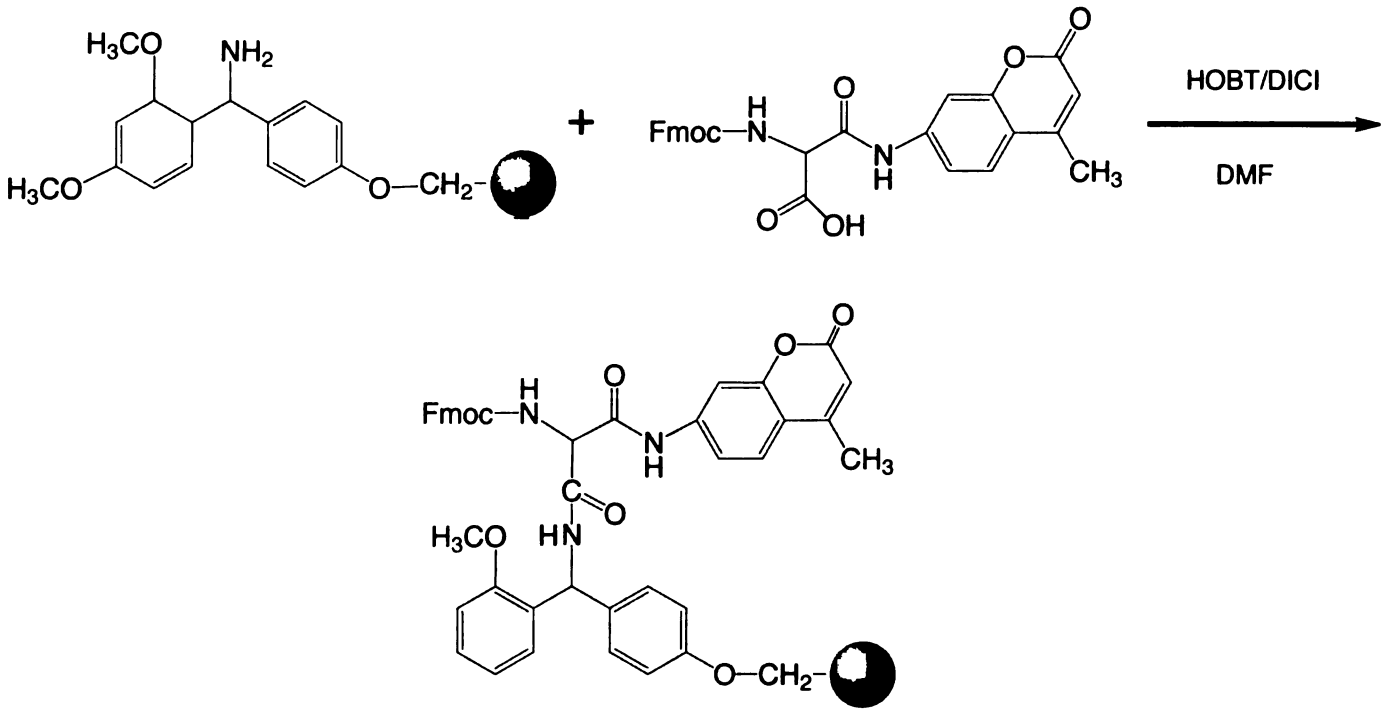


Figure 10
Synthesis of P1-Asn substituted resin. The carboxylic acid residue of Asp-NHMec is hydrolyzed to form an amide linkage resulting in the placement of Asn at P1 attached to the polymeric support resin.

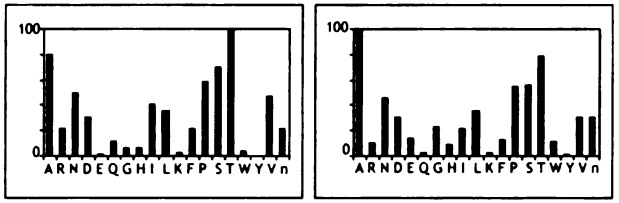
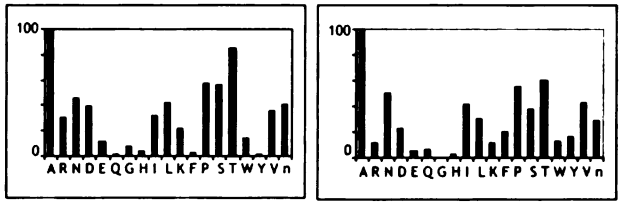
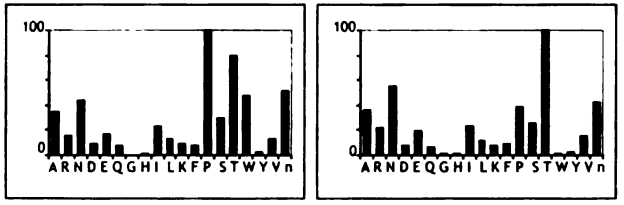
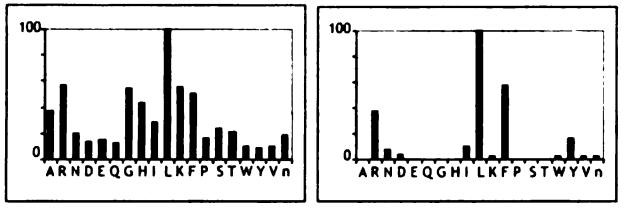
Enzyme	P3	P2	Optimal Sequence
Sm 32 C197			TAN/AAN
Sm32 N197C			AAN/TAN
hAE (human)			PTN
Cruzain			LLN

Figure 11

Profiles of schistosome and human legumains, and cruzain. Protease specificity was determined by the hydrolysis of P1-fixed Asn library. The y-axis represents the rate of AMC production expressed as a percentage of the maximum rate observed in each experiment.

Figure 12 *S. mansoni* asparaginyl endopeptidase *trans*-processes and activates recombinant SmCB1.

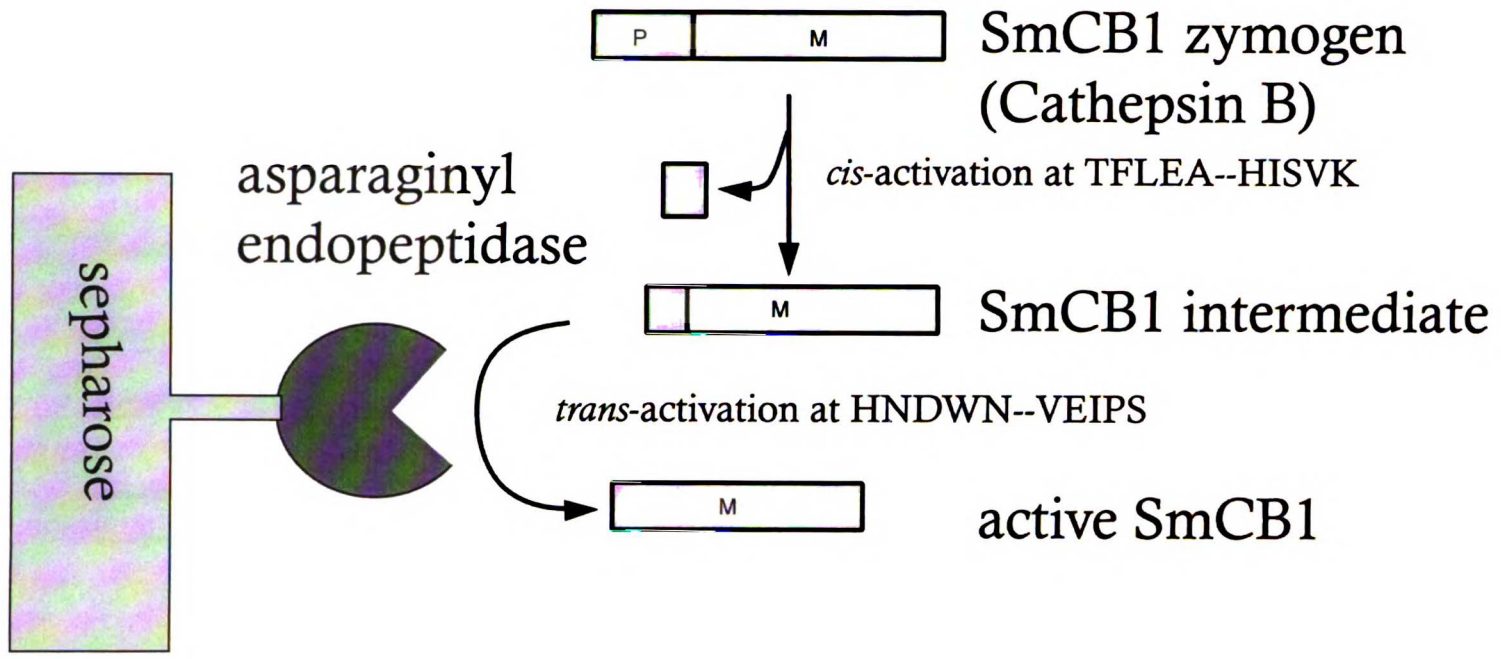


Figure 13a Cbz-R-R-Asn vinyl sulfone

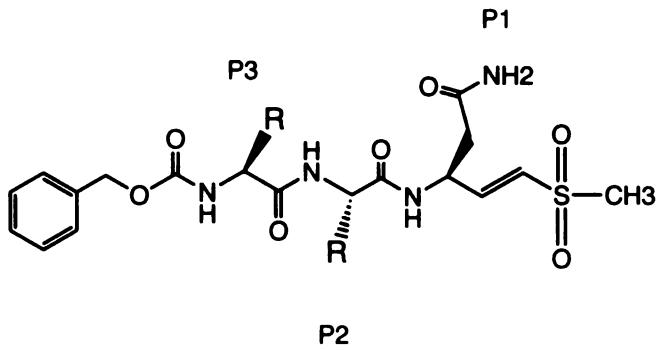


Figure 13b Covalent Enzyme-inhibitor Adduct

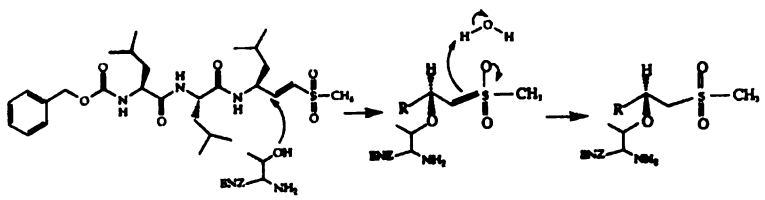
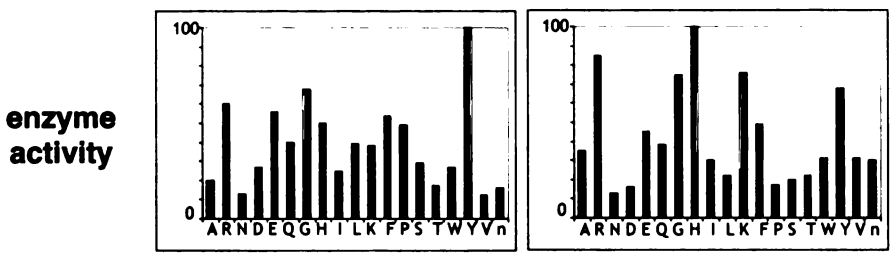


Figure 13c Profiling Protease Inhibitor Specificity



Recombinant Sm32 was pre-incubated with each of 38 vinyl sulfone inhibitors. Inhibition of enzyme activity was measured by loss of ability to cleave the fluorogenic substrate Z-Ala-Ala-AsnMCA (y axis). most effective inhibition=least activity.

Table 1. *Km* and *Vmax* for schistosome and human legumains with selected synthetic peptide MCA substrates

	Schistosome legumain		human legumain	
	Km (μM)	*Vmax* ($\mu M\ s^{-1}$)	*Km* (μM)	*Vmax* ($\mu M\ s^{-1}$)
Z-Ala-Ala-Asn-MCA	90 ± 9	2.9 ± 0.3	80 ± 6	4.4 ± 0.5
Ac-Thr-Ala-Asn-MCA	70 ± 13	3.2 ± 0.4	128 ± 10	5.7 ± 0.6
Ac-Pro-Thr-Asn-MCA	286 ± 40	6.8 ± 0.5	122 ± 13	6.9 ± 0.6
Ac-Phe-Tyr-Asn-MCA	n.a.	n.a.	n.a.	n.a.