Published in final edited form as:

Mol Biochem Parasitol. 2009 March ; 164(1): 86–94. doi:10.1016/j.molbiopara.2008.11.012.

The cathepsin L of *Toxoplasma gondii* (TgCPL) and its endogenous macromolecular inhibitor, toxostatin[☆]

Robert Huang^a, Xuchu Que^a, Ken Hirata^b, Linda S. Brinen^c, Ji Hyun Lee^c, Elizabeth Hansell^d, Juan Engel^d, Mohammed Sajid^d, and Sharon Reed^{a,b,*}

^aDepartment of Medicine, University of California, San Diego, San Diego, California 92103

^bDepartment of Pathology, University of California, San Diego, San Diego, California 92103

^cDepartments of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA 94143

^dSandler Center for Basic Research in Parasitic Diseases, University of California, San Francisco, San Francisco, CA 94143

Abstract

T. gondii is an obligate intracellular parasite of all vertebrates, including man. Successful invasion and replication requires the synchronized release of parasite proteins, many of which require proteolytic processing. Unlike most parasites, *T. gondii* has a limited number of Clan CA, family C1 cysteine proteinases with one cathepsin B (TgCPB), one cathepsin L (TgCPL) and three cathepsin Cs (TgCPC1, 2, 3). Previously, we characterized toxopain, the only cathepsin B enzyme, which localizes to the rhoptry organelle. Two cathepsin Cs are trafficked through dense granules to the parasitophorous vacuole where they degrade peptides. We now report the cloning, expression, and modeling of the sole cathepsin L gene and the identification of two new endogenous inhibitors. TgCPL differs from human cathepsin L with a pH optimum of 6.5 and its substrate preference for leucine (vs. phenylalanine) in the P2 position. This distinct preference is explained by homology modeling, which reveals a non-canonical aspartic acid (Asp 216) at the base of the predicted active site S2 pocket, which limits substrate access. To further our understanding of the regulation of cathepsins in *T. gondii*, we identified two genes encoding endogenous cysteine proteinase inhibitors (ICPs or toxostatins), which are active against both TgCPB and TgCPL in the nanomolar range. Over expression of toxostatin-1 significantly decreased overall cysteine proteinase activity in parasite lysates, but had no detectable effect on invasion or intracellular multiplication. These findings provide important insights into the proteolytic cascades of *T. gondii* and their endogenous control.

Keywords

Toxoplasma gondii; cysteine proteinases; cathepsin L; inhibitor of cysteine proteinases

[☆]Note: Nucleotide sequence data reported in this paper are available in the DDBJ, EMBL and GenBank[™] databases under the accession numbers **EF452500**, **EF452501**, and **EU304362**

Corresponding Author: Tel: 619-543-6146; fax: 619-543-6614; email: slreed@ucsd.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Toxoplasma gondii is an obligate intracellular protozoan parasite that can invade and replicate in any nucleated cell of multiple vertebrate hosts, including humans [1–3]. Toxoplasmosis causes a range of manifestations from asymptomatic to fatal infection. Primary infection of the fetus, which occurs in approximately 1 in 1,000 live births, causes devastating, and often fatal disease [4]. Reactivation of latent toxoplasmosis most often manifests as toxoplasma encephalitis in AIDS patients. Without treatment, toxoplasma encephalitis is uniformly fatal in this population [5].

Invasion by *T. gondii* is regulated by the sequential release of a set of unique apical complex organelles: micronemes, rhoptries, and dense granules [1]. The majority of these key proteins require proteolytic processing. Cysteine proteinases are likely candidates as they are involved in host cell invasion and/or replication in a number of other Apicomplexa parasites such as *Plasmodium* [6–7] and *Cryptosporidium* [8]. These proteinases also appear to be crucial in the process of invasion of toxoplasma. Unlike most protozoa, *T. gondii* has a limited number of Clan CA, family C1 cysteine proteinases with only one cathepsin B (TgCPB), one cathepsin L (TgCPL), and three cathepsin C's (TgCPC 1, 2, and 3) [9]. We have shown that the cathepsin B, TgCPB, is essential to the invasion and replication of *Toxoplasma* as specific inhibitors or antisense to TgCPB blocked the invasion of host cells and caused abnormal rhoptry morphology [10]. Inhibition of TgCPB also limited *in vivo* infection in a chick embryo model of disseminated toxoplasmosis [11]. The cathepsin Cs are key for intracellular survival of the parasite and degrade peptides within the parasitophorous vacuole [12]. We now report the first expression and characterization of active *Toxoplasma gondii* cathepsin L.

The intracellular control of proteolytic activity within a protozoan is critical. The activity of cysteine proteinases of higher eukaryotes is controlled by a number of endogenous inhibitors, including cystatins and α 2-Macroglobulin. No genes homologous to cystatins have been detected in protozoa, but several protozoa, including *T. cruzi* [13], *T. brucei* [14], *Leishmania* [15], *E. histolytica* [16], and *P. falciparum* [17] synthesize endogenous inhibitors with a novel conserved structure, called Inhibitor of Cysteine Proteinases or ICP. Related proteins have also been identified in bacteria but are absent in higher eukaryotes [18,19]. The structure of the *L. mexicana* ICP [15] and chagasin [20,21] were recently described and have a unique immunoglobulin-like fold. ICPs may inhibit parasite cysteine proteinases as in *T. cruzi* [13] and *T. brucei* [14] or host proteinases as in *Leishmania* [15]. We now report the identification of genes encoding two cysteine protease inhibitors, toxostatin-1 and 2, which inhibit *T. gondii* cathepsin L and B in the nanomolar range. Further understanding of the interactions of toxoplasma cathepsins and these endogenous inhibitors should shed light on their role in the pathogenesis of toxoplasmosis.

2. Materials and methods

2.1. Toxoplasma cultures

Primary human foreskin fibroblasts (HFF) were cultured in Dulbecco's modified Eagle's medium (MEM) supplemented with 10% fetal calf serum (FCS) (Irvine Scientific, Irvine, CA) and penicillin and streptomycin (50 μ g/ml) and maintained subsequently in the same medium with 2% FCS. *T. gondii* RH tachyzoites were maintained by serial passage in HFF monolayers in MEM supplemented with 10% FCS and 20 μ g/ml gentamicin solution at 37°C in a humid 5% CO₂ atmosphere.

2.2. Isolation of the TgCPL Gene from a Toxoplasma cDNA Library

DNA primers were synthesized based upon the partial cathepsin L sequence submitted in Genbank by Hansner et. al [22] ([AF184984.1](#)) to amplify a truncated 501 base pair fragment from genomic DNA (TgCPL5: 5'-CAGGGGCAGTGCGGGAGGTGTTGGGC-3' and TgCPL3: 5'-CCAGGTGTTTTT-GACGAT-CCAATAG-3'). The PCR derived probe was radiolabeled with P32 dCTP with DNA polymerase I (Promega) and used to screen the cDNA RH(EP) *T. gondii* bacteriophage lambda library (NIH AIDS Research and Reference Reagent Program). Positive spots, confirmed using duplicate filters, were cored from the agar plates and re-suspended in SM buffer. Positive phage were subjected to another round of screening, the phagemid rescued, and the DNA sequenced as previously described [23]. The complete sequence is in GenBank under accession number [EU304362](#).

2.3. Expression of recombinant TgCPL and TgCPB in Pichia

DNA primers were designed to amplify the full-length pro-mature TgCPL (5'-GAA TTC ATG GAC AGC AGC GAG ACG CAC TAC-3' and 5'-GCG GCC GCT CAC ATC ACG GGG AAA GAC GCA TCT-3') or truncated pro-mature protein (5'-GAA TTC TCG TTC CTC ATT CAG TGG CAG GGC-3' and 5'-GCG GCC GCT CAC ATC ACG GGG AAA GAC GCA TCT-3') from the purified TgCPL cDNA library phagemid. Primers for pro-mature TgCPB (5'-CTC GAG AAA AGA ACC CCG GAC GAC TCG TTG TTT CCG CTT-3') and (3'-GCG GCC GCC TAC ATT TCT CTC TCC TCT TCT GC-5') were used to amplify the gene from total cell cDNA. Cloning of the sequences into the pPicZα plasmid (Invitrogen), electroporation into X-33 cells, selection of proteinase expressing clones, and purification were performed as previously described [24].

2.4. Antibody production and immunoblots

Polyclonal antibody to TgCPL was produced by immunizing rabbits three times with 100 µg of recombinant protein mixed with Titermax Gold Adjuvant® (Sigma). The TgCPL antiserum was affinity-purified by adsorption and desorption to recombinant rTgCPL expressed in *E. coli* or *Pichia* on nitrocellulose membranes [10]. The specificity of the TgCPL antiserum was confirmed by immunoblots containing rTgCPL (100–500 ng) and toxoplasma lysate (1.7×10^7 tachyzoites) and detected with rabbit anti-rTgCPL anti-serum (1:1000) and goat-anti-rabbit IgG horseradish peroxidase (HRP, 1:10,000) using SuperSignal™ (Pierce) [10].

Antibodies were produced to toxostatin-1 in rabbits by the same procedure as above. Additional antibodies were also produced by immunizing Rhode Island Red chickens with gelpurified TgICP1 in Freund's complete adjuvant, followed by monthly boosting in Freund's incomplete for 5 months (Robert Sargeant Antibodies, Ramona, CA). IgY was purified from egg yolks by solubilization in a 1:1.5:2 ratio of egg yolk:PBS: chloroform, centrifugation at 3000 rpm for 30 minutes, precipitation with 12% PEG6000, and resuspended in TBS, 0.5% Tween, 1mM EDTA. Monospecific antibody was generated by affinity-purification with rTgICP1 on immunoblots as detailed above [10]. For toxostatin blots, recombinant protein or parasite lysates were electrophoresed, blotted, and detected with polyclonal rabbit or chicken antibody against toxostatin-1 or Au-1 monoclonal antibody (1 µg/ml, Covance Research Products) and probed with goat anti-chicken, rabbit, or mouse-HRP (Zymed), followed by chemiluminescence detection (Super Signal,Pierce).

2.5. Activity and substrate specificity of TgCPL

Proteinase activity was measured based on the liberation of the fluorescent leaving group, 4-amino-7-methylcoumarin (AMC), from synthetic peptide substrates to determine the preferred cleavage of the P1 and P2 sites as previously described [10]. Recombinant TgCPL was activated by pre-incubating with 5mM DTT (dithiothreitol) for 10 minutes in a substrate buffer

of 50 mM sodium citrate, 2mM EDTA, 0.005% Triton X-100 at pH 6.5. The Michaelis constant (K_m) of TgCPL for the synthetic substrates Z-Arg-Arg-AMC, Z-Phe-Arg-AMC, and Z-Lys-Arg was measured using increasing concentrations of synthetic peptide substrates (2.0 to 150 μ M) and determined using the Enzfitter software (Biosoft, Cambridge, United Kingdom).

Native TgCPL was immunoprecipitated from tachyzoite lysates (1.5×10^9) with 1 μ g monospecific rabbit anti-TgCPL followed by binding to protein A/G agarose (Santa Cruz Biotechnology), and resuspended in sample buffer or by binding to FK-29C, a biotinylated inhibitor (10 μ M, MP Bioproducts) followed by streptavidin agarose. Native and recombinant TgCPL were electrophoresed on a 4–20% gradient SDS gel. The N-terminal peptide sequence of recombinant TgCPL was obtained from the protein band on the gel, while the native protein was transferred to a 0.45 μ M polyvinylidenedifluoride membrane (Immobilon, Millipore Corp., Bedford, MA), the mature proteinase band stained with Coomassie blue, excised, and sequenced by Edman degradation in an Applied Biosystems Procise Liquid Pulse Protein Sequenator at the Stanford University Protein and Nucleic Acids Facility.

The pH optimum of TgCPL was determined by comparing the cleavage of the preferred peptide substrate, Z-KQKLR-AMC, in Na_2HPO_4 /citric acid buffer with pH's ranging from 5.0–8.0.

2.6. Homology modeling of TgCPL

Homology modeling of the mature domain of *Toxoplasma gondii* cathepsin L was performed using human cathepsin K (PDB ID 1U9X) as a template. The sequence of human cathepsin K was found to be the most similar to that of *Toxoplasma gondii* cathepsin L based on BLAST searching. ClustalW, with a blosum matrix and penalties of 10.0 for open gaps and 0.05 for gap extensions [25] was used to perform sequence alignment. A three-dimensional model of TgCPL was generated using the program Modeller [26].

2.7. Immunofluorescent and electron microscopy

Confluent human foreskin fibroblasts on Labtek II slides were infected with *T. gondii* tachyzoites (50,000 per well) for 24 hrs at 37°C, washed, fixed with 4% paraformaldehyde and permeabilized as previously described [10]. Slides were incubated with rabbit anti-TgCPL (1:500), anti-ROP2,3,4 monoclonal antibody (1:500, a kind gift of Jean Francois Dubremetz), dense granule GRA3 monoclonal antibodies (1:500, from Dr. Vern Carruthers), anti-microneme antibody at 1:500 (AMA-1, from Drs. Peter Bradley and John Boothroyd), or monoclonal (1:100, Sigma) or rabbit polyclonal antibody to au-1 (1:200, QED, San Diego, CA), and detected with goat-anti-rabbit IgG (1:200, Alexa 594) or goat anti-mouse IgG (Alexa 488).

Immunoelectron microscopy was performed on infected monolayers fixed in 2% paraformaldehyde, 0.1% glutaraldehyde, 0.1M cacodylate buffer, pH 7.4 and cryoprotected with 20% polyvinyl pyrrolidone (Sigma) in 2.3M sucrose as previously described [27]. Sections were incubated with affinity purified rabbit anti-TgCPL at a 1:50 dilution followed by goat anti-rabbit conjugated with 10 nm gold (Ted Pella Inc, Ca.) at a 1:50 dilution for 60 minutes. The sections were then stained with oxalate uranyl acetate and embedded in 1.5% methyl cellulose (Sigma, Mo.), 0.3% aqueous uranyl acetate (Ted Pella Inc., Ca.), and examined with a Philips Tecnai 10 electron microscope.

2.8. Cloning and identification of the toxostatin genes

The *T. gondii* Genome Database (<http://toxodb.org>) was searched for the signature motifs (NPTTGY and/or V/I-X5-G-X8-VRPW) to the chagasin and cystatin family using the BLAST and protein motif program. The putative ICP family genes were identified from EST and cDNA database (TgTwin_Scan). Total cellular RNA from *T. gondii* was isolated using RNazol

reagent (Invitrogen) and transcribed into cDNAs using Superscript II reverse transcriptase and oligo(dT) primer. Toxostatin-1 and 2 were amplified from *T. gondii* RH cDNA and cloned into pQE80L (Qiagen) respectively using primers based on the sequence data from TgTwin_Scan_6575 and 7478 of <http://toxodb.org>. An alignment of the toxostatins with the chagasin protein family was prepared using ClustalW program. The toxostatin sequences were deposited in GenBankTM under accession numbers **EF452500** and **EF452501**. The predicted cleavage site for the signal peptidase was determined according to Von Heijne [28].

2.9. Heterologous expression and purification of recombinant toxostatins

The toxostatin coding sequences (without the N-terminal signal region) were amplified from *T. gondii* cDNA with primers incorporating *Sac*I and *Hind*III restriction sites and an N-terminal histidine tag (TgICP1-*Sac*I: 5'-ATA GAG CTC TGC CCG AGC GCG TGC GTC CAC-3' and TgICP1-*Hind*III: 5' - AAT AAG CTT GTC CGT TGC ATG AAT ATG GAC CAC-3'; TgICP2-*Bam*HI: 5'-GATA GGA TCC AGG CAA GGT ACG TCG CCG CGC GCT-3' and TgICP2-*Hind*III: 5'-GAAT AAG CTT GTC GAA GTG TAC GAG AGC GAC GAA G-3'). The cDNA fragments were digested with *Sac*I and *Hind*III, inserted into the linearized vector, pQE80L (Promega), and transformed into *E. coli* JM109 with ampicillin (100 µg/ml) selection. Selected clones were characterized by restriction mapping and sequenced.

For protein expression, *E. coli* JM109 was induced with 1 mM isopropyl-d-thiogalactopyranoside (IPTG) for 4 h, and the recombinant proteins purified by nickel affinity chromatography with imidazole as previously described [10]. Protein purity and concentration were estimated by Coomassie blue staining and immunoblots with His-tag antibody or anti-toxostatin antibodies. Toxostatins were purified to apparent homogeneity (>90%) by SDS-PAGE analysis.

Inhibition of cysteine protease activity was measured by pre-incubation of cathepsin L or cathepsin B with rToxostatin-1 at various dilutions for 30 min at room temperature in 100 mM sodium phosphate, pH 6.0, containing 2 mM EDTA and 1 mM DTT, and subsequent addition of 8 µM Z-Phe-Arg-AMC (for TgCPL) or Z-Arg-Arg-AMC (for TgCPB). The IC₅₀ was calculated as the concentration of inhibitor resulting in 50% inhibition of proteinase activity compared with non-inhibited controls.

2.10. Over-expression of toxostatin-1

The pminiHXGPRT-gra1 vector (NIH AIDS Research and Reference Reagent Program) containing a strong promoter of *T. gondii* *gra*1 gene with C-terminal AU1 epitope tag, was used to drive the overexpression of toxostatin-1. The coding sequence of toxostatin-1 was amplified by PCR from *T. gondii* strain RH cDNA with *Bgl*II and *Avr*II incorporated into primers. The PCR amplicon was digested with *Bgl*II and *Avr*II and ligated in frame into the linearized vector. Plasmid DNA was purified from transformed *E. coli* clone using a Maxiprep kit (Qiagen) and sequenced. Tachyzoites of *T. gondii* ΔHXGPRT strain were electroporated with 50 µg of plasmid DNA (pminiHXGPRT-gra1-TgICP1-au1) and transfectants identified through MPA/X selection (25 µg/ml mycophenolic acid + 50 µg/ml xanthine).

Tachyzoites (5×10^5 control or pminiHXGPRT-gra1-TgICP1) were added to fibroblast monolayers in chamber slides and invasion determined by acridine orange staining at 2 h and intracellular multiplication at 24 h as previously described [10].

3. Results

3.1. Cloning of TgCPL

The sequence of proTgCPL was obtained by screening the *Toxoplasma gondii* RH(EP) λ cDNA with a PCR derived DNA probe based upon a previously submitted truncated *T. gondii* cathepsin L Genbank sequence. Clone TgCPL53 encoded a predicted 421 amino acid zymogen, which includes an 1197 bp 3' c-terminal extension and a 378 bp 5' non-coding region. Analysis of the deduced open reading frame revealed the classic cathepsin L amino acid motifs ERFNIN, KNFD, and SPV. The full-length zymogen consists of a deduced 221 amino acid mature protein, 125 amino acid pro-region, and a 75 amino acid pre-region. A 23 amino acid potential transmembrane domain spans the pre-pro region (SOSUI Signal Program: <http://sosui.proteome.bio.tuat.ac.jp>). Two potential N-glycosylation sites were predicted utilizing the YinOYang 1.2 software algorithm (<http://www.cbs.dtu.dk>). BlastP analysis indicated homology with other protozoal cathepsin L-like genes, including 55% deduced amino acid identity with *Sarcocystis muris*, 44% with *C. parvum*, 38% with Falcipain-3, 36% with Falcipain-2, and 24% with TgCPB (Figure 1).

3.2. Purification and activity of rTgCPL

Pichia pastoris has proven to be a useful system for expressing active recombinant cathepsins [24], which are usually lethal to bacteria. We took advantage of expression of the TgCPL and B as proenzymes with an α -mating factor fusion, which promotes secretion into the media. After concentration and purification by anion exchange column using FPLC, a single band with Mr ~32 kD was detected (Figure 2). Since the predicted Mr of mature TgCPL is 24 kD, this band is consistent with partially processed pro-mature rTgCPL or aberrant electrophoretic motility as seen with cruzain from *T. cruzi* [29]. To confirm this, we obtained the N-terminal sequence of the active, recombinant enzyme: LAGVDWRSR, confirming that the cleavage occurred at the predicted site (see double arrow, Figure 1). The N-terminus of the native enzyme was blocked, but since it has the same electrophoretic mobility as the recombinant enzyme, we assume that the endogenous mature proteinase is similarly processed and lacks the transmembrane domain.

rTgCPL has the greatest affinity for leucine in the P2 site with a Km of $7 \pm 0.7 \mu\text{M}$ (Table 1), while the preferred substrate of human cathepsin L, Z-FR-AMC, had a Km of $15.7 \pm 1.0 \mu\text{M}$. Homology modeling revealed the molecular basis for this substrate specificity (see 3.3). rTgCPL was active in broad pH range from 4 to 8, but maximal at pH 6.5 (Figure 3).

3.3. Homology modeling of TgCPL

The sequence of the mature domain of *Toxoplasma gondii* cathepsin L showed 50% identity with human cathepsin K. Based on a BLAST search of sequences that have known three-dimensional structures, human cathepsin K showed the highest degree of similarity and was therefore used as a modeling template. Comparing the model of TgCPL with the three-dimensional structure of human cathepsin L (PDB ID 1CJL or 1CS8) revealed significant differences in the area of the enzyme active site. In the model of TgCPL, an aspartic acid (Asp 216) is present in the bottom of the pocket (Figure 4). In human cathepsin L, the comparable position is filled by an alanine (Ala 214). The presence of an inflexible aspartic acid in the pocket of TgCPL likely is the basis for the observed substrate preference for leucine in P2 vs. the usual preference of cathepsin Ls for phenylalanine as the larger phenylalanine cannot be accommodated (Figure 5) [30]. The remainder of the S2 region maintains some similarity. Both enzymes have a methionine in the S2 pocket region (*Toxoplasma* Met 74 and human Met 70). Furthermore, modeling of the *Toxoplasma* enzyme shows Leu 163 in this cleft while the human enzyme has Met 161 in the comparable position.

The apparent shape of the space available in the S3 regions of the two enzymes is visibly different, with the human enzyme having more room than the *Toxoplasma* enzyme. In addition, the TgCPL's S3 region contains more abundant opportunity for charge-stabilizing interactions with Glu 64, Glu 73, Gln 67 and the main chain carbonyl group of Gly 65 lining the available surface of the S3 area. In contrast, human cathepsin L shows Tyr 72, Leu 69, Glu 63 and the main chain carbonyl of Gly 61 in comparable positions. The difference in the relative depth and breadth of this region in the two enzymes is particularly obvious around *Toxoplasma* Gly 65/human Gly61, however there are no sequence inserts or deletions in this region that account for the predicted difference of the S3 pocket.

3.4. Localization of TgCPL

We produced a polyclonal antibody, which reacts with native TgCPL, for localization within the tachyzoite. TgCPL was found primarily in the apical end of the tachyzoite (Figure 6A), but did not co-localize with any apical organelles, including rhoptries, dense granules, or micronemes (data not shown). Electron microscopy confirmed that TgCPL localized to a small vesicle population (Figure 6B). Although TgCPL has a putative transmembrane domain, it is cleaved from the mature enzyme (Figure 1), and no membrane localization was detected.

3.5. Analysis of the toxostatin genes

We queried the *T. gondii* genome database for signature motifs homologous to that of chagasin, a cysteine protease inhibitor first described in *T. cruzi* [13]. We identified two putative chagasin family genes, toxostatin-1 and 2. The derived amino acid sequences consist of 177 and 258 amino acid residues with calculated molecular masses (without signal peptides) of Mr 17 kDa and 25 kDa, respectively. Toxostatin-1 and 2 show only low homology to each other (28% identity and 42% similarity) with a 15% gap. Only toxostatin-2 contains a 48 aa N-terminal extension and a 51 aa internal insertion resembling serine proteinase inhibitors (serpins) from plants. According to an alignment of selected sequences using the CLUSTAL W algorithm, the overall sequence homology of toxostatin-2 to the chagasin family of *T. cruzi*, *E. histolytica*, *Leishmania major* or *Plasmodium* is less than 15%. However, the amino acid sequence of toxostatin-1 has about 22% identical (49% similarity) to a C-terminal domain of an ICP-like inhibitor sequence from *P. berghei* (Figure 7).

Expression levels of toxostatin-1 and 2 mRNA differed in parasite developmental stages based on SAGE and EST data. Toxostatin-1 is expressed in high level with more than 110 EST tags found from all stages including tachyzoites, bradyzoites and sporozoites while toxostatin-2 is expressed in relatively low levels with 7 EST tags found only from tachyzoite cDNA (<http://www.toxodb.org/toxo/home.jsp>.)

3.6. Purification and activity of recombinant toxostatin-1

We over expressed the toxostatins heterologously in *E. coli*, resulting in approximately 45% of total cellular proteins (data not shown). The recombinant proteins were primarily soluble and were purified by nickel affinity chromatography to 95% homogeneity. The recombinant proteins have apparent molecular weights of 17 kDa and 25 kDa respectively (Figure 8A), similar to the theoretical value calculated from its predicted amino acid sequence.

To investigate the function of the toxostatins, purified recombinant toxostatin-1 was used to test inhibitory activity against cathepsins from *T. gondii*. Recombinant toxostatin-1 effectively inhibited the peptidase activity of rTgCPL and rTgCPB in the nanomolar range (IC₅₀=24.0 nM for rTgCPL and 31.4 nM for rTgCPB). Toxostatin-1 also inhibited human cathepsin L (IC₅₀=9.9 nM), more efficiently than human cathepsin B (IC₅₀=146.5 nM).

The role of toxostatin-1 in *T. gondii* was further tested by over-expression of toxostatin-1 with an au-1 tag under control of the strong *gal* promoter (Figure 8B). Localization studies of epitope tagged toxostatin-1 revealed signal throughout the cell, likely reflecting aberrant trafficking from over expression (data not shown). The cathepsin activity (both cathepsin B and L) was reduced by >90% in toxostatin-1 transfected tachyzoite lysates. To determine if inhibition of the cathepsin activity affected invasion or intracellular multiplication, monolayers were infected with wild type RH or toxostatin-1-over expressing strains. Neither the amount of invasion at 2 h or multiplication at 24 hr was affected (data not shown).

Discussion

Toxoplasmosis remains a major cause of congenital infection and causes serious complications in infected, immunocompromised hosts. While standard treatment is generally accessible and relatively inexpensive, the frequency of adverse events from sulfa drug therapy, some of which can prove fatal, necessitates the search for safe alternatives. Parasitic cysteine proteinases are key enzymes encompassing a broad-range of biological functions including evasion of the host immune defenses, host cell/tissue invasion, and proteolytic processing of precursor proteins. These enzymes have been targeted for chemotherapy using synthetic peptide inhibitors, as host cathepsin homologues are biochemically and structurally distinct and have greater redundancy than parasite cathepsins.

Cysteine proteinases are also critical enzymes in the invasion and replication of several *Apicomplexa* protozoa, including *Toxoplasma gondii* [10], *Eimeria* [31], and *Plasmodia* [6, 32,33]. *Toxoplasma* is unique among protozoa with only a limited number of Clan CA, family C1 cathepsins, including one Cathepsin B [10], one L, and three Cs [12].

The TgCPL protein is synthesized as a zymogen consisting of a 75 amino acid presequence, a 124 amino acid pro-sequence, and a 222 amino acid mature region (Figure 1). TgCPL contains the highly conserved ERFNIN amino acid motif within the pro-region, which is characteristic of cathepsin L-like cysteine proteinases. A BlastP homology search of TgCPL resulted in the highest match to the cathepsin L gene from the apicomplexan *Sarcocystis muris* (55% deduced amino acid identity), which has been reported to be secreted into the parasitophorous vacuole from the dense granules [22]. Previous studies of a human cathepsin L demonstrated the necessity of a carboxy-terminal amino acid motif (S-X-P-X-V) for protein secretion [34]. TgCPL contains a similar motif (S-F-P-V-M) at the carboxy-terminal. Interestingly, when Chauhan et. al. removed the second non-specific amino acid from the pentamer motif, they found that human cathepsin L was still secreted (S-Y-P-V). Proteomics data did not indicate the presence of TgCPL from secreted *T. gondii* proteins, which included several dense granule proteins of unknown function [35].

A potential transmembrane domain was predicted spanning the 3' end of the pre-region and the 5' end of the pro-region of TgCPL using the SOSUI prediction program [36], but this is cleaved from the mature enzyme, which would explain the lack of membrane localization by fluorescent or immunoelectron microscopy (Figure 6). The potential transmembrane domain within the *P. falciparum* genes falcipain-2 and falcipain-3 is responsible for the proper targeting to the plasmodium plasma membrane through the endoplasmic reticulum and secretory pathway [37]. Additionally, a unique bipartite motif from both the cytoplasmic and luminal portions of the falcipain-2 prodomain has been demonstrated to be essential for targeting the cathepsin into the food vacuole [37]. There is no homology to the same region in TgCPL, and the role, if any, of the TgCPL transmembrane domain and pro-region as a trafficking regulator is currently unknown.

The function of the 1307bp TgCPL 3'UTR is also unknown. In other eukaryotes, the 3' UTR is capable of binding proteins, including endonucleases, to regulate transcription levels. In two human cathepsin B genes, protein binding within the 3' UTR has been shown to stabilize the stem-loop structure [38]. In the closely related protozoan, *Plasmodium falciparum*, gene expression can be upregulated through elements within the 3' flanking sequences [39]; however, equivalent regulatory sequences are not present in the TgCPL 3' UTR. Additionally, miRNA's can inhibit translation by binding to the 3'UTR in higher eukaryotes [40]. Cis-regulated regions in the 5' UTR's of the enolase genes [41] and the nucleoside triphosphate hydrolase gene [42] of *T. gondii* have been previously reported. No evidence has been definitively presented linking the binding of protein, presence of regulatory elements, or miRNA in 3' UTR gene regulation of *T. gondii*.

The preferred substrate of most cathepsin Ls, including human, is for Phe in the P2 position. TgCPL is unusual in its preference for Leu in the P2 position (Table 1). These differences in substrate preferences can be explained by homology modeling (Figure 4 and Figure 5). The aspartic acid side chain at the base of the TgCPL S2 pocket is less flexible and cannot rotate out of the pocket to accommodate large incoming groups, such as Phe. Leucine, in contrast, is shorter and more flexible. Limited binding of Phe likely occurs through interactions with the hydrophobic walls of the pocket, which provide enough stabilizing interactions to accommodate partial insertion. In the human enzyme, the small and hydrophobic alanine moiety can readily accommodate Phe, Val or Leu at P2.

Steric considerations for substrate preferences can be illustrated by making use of known structures of another papain super family cysteine protease, cruzain [29], in complex with small molecule inhibitors. These complex structures provide a template for the comparison of potential binding of substrates to TgCPL. Superimposition of the three-dimensional structures of different inhibitor-bound complexes of cruzain on the model structure of TgCPL allow for the approximation of the positioning of various moieties in the toxo enzyme's active site region. Superimposition of cruzain bound to an inhibitor (PDB ID 1F2A) [29] with Phe in the P2 position (rms deviation of protein superimposition 0.952 Å) on the TgCPL model clearly illustrates that Phe is too large of a side chain to sit in this pocket. Steric clash is evident with Asp 216 at the base of the pocket (Figure 5A). In contrast, superimposition of cruzain bound to an inhibitor with Leu in the P2 position (PDB ID 1EWP) (rms deviation of protein superimposition = 0.965 Å) reveals the more ready accommodation of a Leu side chain. Similar probing of the structure of human cathepsin L reveals that there is indeed adequate space for Phe and Leu to fit in the S2 pocket with an alanine at the base (Figure 5B).

Protein modeling predicts that the mature region of TgCPL should be exposed to the extracellular space, raising the possibility of its potential role in host cell invasion and/or immunoevasion [28]. While immunofluorescent imaging did not show membrane-associated staining, antigen presentation could be a transient event dictated by the stage of infection. Carruther's group found that cysteine proteinase inhibitors blocked microneme protein secretion [43], and TgCPL may be linked to microneme protein processing (V. Carruthers, personal communication).

Roles for cathepsin L-like cysteine proteinases in Apicomplexa have been best defined in *Plasmodium*. Falcipain-2 and 3 are critical for hemoglobin degradation in the food vacuoles of *P. falciparum* [6]. Disruption of falcipain-3 is lethal, while knock-out of falcipain-2 results in accumulation of hemoglobin in food vacuoles [32]. Disruption of falcipain-1, reduced oocyst production by 70–90%, suggesting an important function of this cysteine proteinase in the parasite's development in the mosquito midgut [7]. The cathepsin L gene in *Cryptosporidium* is the earliest known lineage in the cathepsin L-like family [44], but the gene has not been cloned or further characterized.

In evaluating potential inhibitors of TgCPL, we identified two proteinaceous inhibitors of cysteine proteinases (ICPs). Functional homology has been identified between proteinase peptidase inhibitors of both protozoa (*Trypanosoma cruzi* [13], *T. brucei* [14], *Leishmania* [15], *P. falciparum* [17]) and bacteria (*Pseudomonas aeruginosa* [18,19]) with a universal inhibition of clan CA, family C1 cysteine peptidases, despite low sequence similarities between these inhibitors [19]. Comparison of the toxostatin sequence with those of several ICP inhibitors from other organisms (Figure 7) shows that these proteins typically contain three regions with conserved sequence elements: NPTTGY(F)xW at positions 24–34, GxGG at positions 59–62 and LxYxRPW(F) at positions 80–86, respectively. The first motif of cruzipain has been shown to interact with the catalytic cysteine of the target protease [21]. This was confirmed in a recent mutagenesis study of chagasin in which the L2 loop, containing the NPPTGY motif, was critical to inhibit its native target, cruzain [45]. The key inhibitory residues may depend on the target enzyme, however, as mutants in Loop 4 were critical for leishmanal ICP interactions with papain [46], while the chagasin L6 loop with LxYxRPW was important to inhibit human cathepsin L [45]. Neither toxostatin nor falstatin have the typical first motif (NPTTG) of chagasin. Instead, toxostatin and falstatin contain a conserved element of GxGYxW(F/L) at the position of first motif. Toxostatins also lack the second motif but contain the third motif similar to the chicken cystatin and rat kininogen. The co-crystallization of chagasin with falcipain-2 [21] and NMR structure with human cathepsin L [20] confirmed the immunoglobulin folds and tripartite binding, suggesting an evolutionarily conserved scaffold [21]. We found that toxostatins were active against multiple cathepsins, including human, similar to the findings with chagasin [13] and falstatin [17], suggesting that the tripartite binding may be more important for their broad inhibitory activity than specific residues.

Based on their broad specificity, ICPs may inhibit either parasite or host cathepsins. Interestingly, chagasin [13], the leishmanial ICP [15], and falstatin [17] are down-regulated at the time of highest expression of clan CA cysteine proteinases. In contrast, mRNA for toxostatin-1 was highest in tachyzoites. Excess chagasin disrupts *in vitro* invasion of *T. cruzi* [47], while disruption of leishmanial ICP limits infection in mice [15]. Despite significant inhibition of overall cathepsin activity by the over expression of toxostatin-1, we could not demonstrate an effect on host cell invasion or parasite multiplication *in vitro*. This may be due to the lack of colocalization of the over expressed toxostatin-1 in the same subcellular compartments as TgCPB or L. Alternatively, toxostatin-1 may act on a host cathepsin or be involved in intracellular inhibition of TgCPL or TgCPB and not play a significant role in invasion. Further understanding of the mechanism of action of the toxostatins may not only clarify the role of TgCPL in the pathogenesis of toxoplasmosis, but may identify unique sequences, which lead to better inhibitor design.

Abbreviations

ICP, Inhibitor of Cysteine Proteinase; rTgCPB, recombinant *Toxoplasma gondii* cathepsin B; TgCPC, *Toxoplasma gondii* cathepsin C; TgCPL, *Toxoplasma gondii* cathepsin L.

Acknowledgements

This work was supported by NIAID (AI41093 S.R., AI35707 J.E.), the University of California University-wide AIDS Research Program (ID04-SD-079, S.R.), the Rockefeller Brothers Fund (S.R.), and the UCSD Center for AIDS Research (X.Q.). We thank Drs. Fran Gillin and Charles Davis for their helpful comments, and Ivy Hsieh for her technical EM support.

References

1. Black MW, Boothroyd JC. Lytic cycle of *Toxoplasma gondii*. Microbiol Mol Biol Rev 2000;64:607–623. [PubMed: 10974128]

2. Dobrowolski JM, Sibley LD. Toxoplasma invasion of mammalian cells is powered by the actin cytoskeleton of the parasite. *Cell* 1996;84:933–939. [PubMed: 8601316]
3. Montoya JG, Liesenfeld O. Toxoplasmosis. *Lancet* 2004;363:1965–1973. [PubMed: 15194258]
4. McAuley J, Boyer KM, Patel D, et al. Early and longitudinal evaluations of treated infants and children and untreated historical patients with congenital toxoplasmosis: the Chicago Collaborative Treatment Trial. *Clin Infect Dis* 1994;18:38–72. [PubMed: 8054436]
5. Luft BJ, Remington JS. Toxoplasmic encephalitis in AIDS. *Clin Infect Dis* 1992;15:211–222. [PubMed: 1520757]
6. Dahl EL, Rosenthal PJ. Biosynthesis, localization, and processing of falcipain cysteine proteases of *Plasmodium falciparum*. *Mol Biochem Parasitol* 2005;139:205–212. [PubMed: 15664655]
7. Eksi S, Czesny B, Greenbaum DC, Bogyo M, Williamson KC. Targeted disruption of *Plasmodium falciparum* cysteine protease, falcipain 1, reduces oocyst production, not erythrocytic stage growth. *Mol Microbiol* 2004;53:243–250. [PubMed: 15225318]
8. Ojcius DM, Perfettini JL, Bonnin A, Laurent F. Caspase-dependent apoptosis during infection with *Cryptosporidium parvum*. *Microbes Infect* 1999;1:1163–1168. [PubMed: 10580271]
9. Kissinger JC, Gajiria B, Li L, Paulsen IT, Roos DS. ToxoDB: accessing the *Toxoplasma gondii* genome. *Nucleic Acids Res* 2003;31:234–236. [PubMed: 12519989]
10. Que X, Ngo H, Lawton J, et al. The cathepsin B of *Toxoplasma gondii*, Toxopain-1, is critical for parasite invasion and rhoptry protein processing. *J Biol Chem* 2002;277:25791–25797. [PubMed: 12000756]
11. Que X, Wunderlich A, Joiner KA, Reed SL. Toxopain-1 is critical for infection in a novel chicken embryo model of congenital toxoplasmosis. *Infect Immun* 2004;72:2915–2921. [PubMed: 15102804]
12. Que X, Engel JC, Ferguson D, Wunderlich A, Tomavo S, Reed SL. Cathepsin Cs are key for the intracellular survival of the protozoan parasite, *Toxoplasma gondii*. *J Biol Chem* 2007;282:4994–5003. [PubMed: 17164247]
13. Monteiro AC, Abrahamson M, Lima AP, Vannier-Santos MA, Scharfstein J. Identification, characterization and localization of chagasin, a tight-binding cysteine protease inhibitor in *Trypanosoma cruzi*. *J Cell Sci* 2001;114:3933–3942. [PubMed: 11719560]
14. Santos CC, Coombs GH, Lima AP, Mottram JC. Role of the *Trypanosoma brucei* natural cysteine peptidase inhibitor ICP in differentiation and virulence. *Molec Microbiol* 2007;66:991–1002. [PubMed: 17944830]
15. Besteiro S, Coombs GH, Mottram JC. A potential role for ICP, a leishmanial inhibitor of cysteine peptidases, in the interaction between host and parasite. *Molec Microbiol* 2004;54:1224–1236. [PubMed: 15554964]
16. Riekenberg S, Witjes B, Saric M, Bruchhaus I, Scholze H. Identification of EhICP1, a chagasin-like cysteine protease inhibitor of *Entamoeba histolytica*. *FEBS Lett* 2005;579:1573–1578. [PubMed: 15757643]
17. Pandey K, Singh N, Arastu-Kapur S, Bogyo M, Rosenthal PJ. Falstatin, a cysteine protease inhibitor of *Plasmodium falciparum*, facilitates erythrocyte invasion. *PLoS Pathogens* 2006;2:1031–1041.
18. Rigden DJ, Mosolov VV, Galperin MY. Sequence conservation in the chagasin family suggests a common trend in cysteine proteinase binding by unrelated protein inhibitors. *Protein Sci* 2002;11:1971–1977. [PubMed: 12142451]
19. Sanderson SJ, Westrop GD, Scharfstein J, Mottram JC, Coombs GH. Functional conservation of a natural cysteine peptidase inhibitor in protozoan and bacterial pathogens. *FEBS Lett* 2003;542:12–16. [PubMed: 12729889]
20. Ljunggren A, Redzynia I, Alvarez-Fernandez M, et al. Crystal structure of the parasite protease inhibitor chagasin in complex with a host target cysteine protease. *J Mol Biol* 2007;371:137–153. [PubMed: 17561110]
21. Wang SX, Pandey KC, Scharfstein J, et al. The structure of chagasin in complex with a cysteine protease clarifies the binding mode and evolution of an inhibitor family. *Structure* 2007;15:535–543. [PubMed: 17502099]

22. Hansner T, Freyer B, Mehlhorn H, Ruger W. Isolation and characterization of a cDNA clone encoding a thiol proteinase of *Sarcocystis muris* cyst merozoites (Apicomplexa). *Parasitol Res* 1999;85:749–757. [PubMed: 10431744]
23. Reed SL, Bouvier J, Pollack AS, et al. Cloning of a virulence factor of *Entamoeba histolytica*: pathogenic strains possess a unique cysteine proteinase gene. *J Clin Invest* 1993;91:1532–1540. [PubMed: 8473498]
24. Que X, Kim SH, Sajid M, et al. A surface amebic cysteine proteinase inactivates Interleukin-18. *Infect Immun* 2003;71:1274–1280. [PubMed: 12595442]
25. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680. [PubMed: 7984417]
26. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993;234:779–815. [PubMed: 8254673]
27. Peters PJ. Cryo-immunogold Electron Microscopy. *Current Protocols in Cell Biology* 1999;4.7.1–4.7.12.
28. Emanuelsson O, Nielsen H, Brunak S, von HG. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* 2000;300:1005–1016. [PubMed: 10891285]
29. McGrath ME, Eakin AE, Engel JC, McKerrow JH, Craik CS, Fletterick RJ. The crystal structure of cruzain: a therapeutic target for Chagas disease. *J Mol Biol* 1995;247:251–259. [PubMed: 7707373]
30. Brinen LS, Hansell E, Cheng J, Roush WR, McKerrow JH, Fletterick RJ. A target within the target: probing cruzain's P1' site to define structural determinants for the Chagas' disease protease. *Structure* 2000 Aug;8:831–840. [PubMed: 10997902]
31. Fuller AL, McDougald LR. Reduction in cell entry of *Eimeria tenella* (Coccidia) sporozoites by protease inhibitors, and partial characterization of proteolytic activity associated with intact sporozoites and merozoites. *J Parasitol* 1990;76:464–467. [PubMed: 2199652]
32. Sijwali PS, Koo J, Singh N, Rosenthal PJ. Gene disruptions demonstrate independent roles for the four falcipain cysteine proteases of *Plasmodium falciparum*. *Mol Biochem Parasitol* 2006;150:96–106. [PubMed: 16890302]
33. Sijwali PS, Rosenthal PJ. Gene disruption confirms a critical role for the cysteine protease falcipain-2 hemoglobin hydrolysis by *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 2004;101:4384–4389. [PubMed: 15070727]
34. Chauhan SS, Ray D, Kane SE, Willingham MC, Gottesman MM. Involvement of carboxy-terminal amino acids in secretion of human lysosomal protease cathepsin L. *Biochem* 1998;37:8584–8594. [PubMed: 9622510]
35. Zhou XW, Kafsack BF, Cole RN, Beckett P, Shen RF, Carruthers VB. The opportunistic pathogen *Toxoplasma gondii* deploys a diverse legion of invasion and survival proteins. *J Biol Chem* 2005;280:34233–34244. [PubMed: 16002397]
36. Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 1998;14:378–3. [PubMed: 9632836]
37. Subramanian S, Sijwali PS, Rosenthal PJ. Falcipain cysteine proteases require bipartite motifs for trafficking to the *Plasmodium falciparum* food vacuole. *J Biol Chem* 2007;282:24961–24969. [PubMed: 17565983]
38. Tam SW, Cote-Paulino LR, Peak DA, Sheahan K, Murnane MJ. Human cathepsin B-encoding cDNAs: sequence variations in the 3'-untranslated region. *Gene* 1994;139:171–176. [PubMed: 7509303]
39. Oguariri RM, Dunn JM, Golightly LM. 3' gene regulatory elements required for expression of the *Plasmodium falciparum* developmental protein, Pfs25. *Mol Biochem Parasitol* 2006;146:163–172. [PubMed: 16439031]
40. Ullu E, Tschudi C, Chakraborty T. RNA interference in protozoan parasites. *Cell Microbiol* 2004;6:509–519. [PubMed: 15104593]
41. Kibe MK, Coppin A, Dendouga N, et al. Transcriptional regulation of two stage-specifically expressed genes in the protozoan parasite *Toxoplasma gondii*. *Nucleic Acids Res* 2005;33:1722–1736. [PubMed: 15784612]

42. Nakaar V, Bermudes D, Peck KR, Joiner KA. Upstream elements required for expression of nucleoside triphosphate hydrolase genes of *Toxoplasma gondii*. *Mol Biochem Parasitol* 1998;92:229–239. [PubMed: 9657328]
43. Teo CF, Zhou XW, Bogyo M, Carruthers VB. Cysteine protease inhibitors block *Toxoplasma gondii* microneme secretion and cell invasion. *Antimicrob Agents Chemo* 2007;51:679–688.
44. Sajid M, McKerrow JH. Cysteine proteases of parasitic organisms. *Mol Biochem Parasitol* 2002;120:1–21. [PubMed: 11849701]
45. dos Reis FCG, Smith BP, Santos CC, et al. The role of conserved residues of chagasin in the inhibition of cysteine peptidases. *FEBS Lett* 2008;582:485–490. [PubMed: 18201565]
46. Smith BO, Picken NMC, Westrop GD, et al. The structure of *Leishmania mexicana* ICP provides evidence for convergent evolution of cysteine peptidase inhibitors. *J Biol Chem* 2006;281:5821–5828. [PubMed: 16407198]
47. Santos CC, Sant'Anna C, Terres A, Cunha-e Silva NL, Scharfstein J, Lima APC. Chagasin, the endogenous cysteine protease inhibitor of *Trypanosoma cruzi*, modulates parasite differentiation and invasion of mammalian cells. *J Cell Sci* 2005;118:901–915. [PubMed: 15713748]

TgCPL	-----MDSSETHYVSFLNGEDDGLENGELHQRRGVRAGRSPSPFVVTTRTYF	47
SmPT1	-----MTTRTYF	7
Cparvum	-----MDIGNN-----VEEHQEIYI	14
PfCP3	-----MEYHMEYSPNEVIKQEREVFVGKEKSGSKFKRKRISFIVLTVSICFMFALM	51
PfCP2	-----MDYNMDYAPHEVISQQGERFVDKYVDRKILKNKSLVVISLSVSVGVFV	51
TgCPB	MEGRKSRVLGTLPLFAALAILLLGCMYTRAAGTPDDSLFLPLESDTSVDPRESFSAEDV	60
* ↓		
TgCPL	WKKFLRQNFNT--ARAWIA-----LVAAVSLAFASFLLQWQGEDDRAVFP-----	91
SmPT1	WRKLFGRGRYTT-SKRFLA-----VVAGAATAVLLLAFLLLRQQHSHLLDGADMGYD	59
Cparvum	SGFYIALINGTNQOREFNK-----KLKNIIATLIAIFVLVVTVSLYITN-----	60
PfCP3	LFYFTRNENKTLFTNSLSNNINDDYIINSLKSESQKGFIVSKLEELISSYDKEKKMRT	111
PfCP2	LFYFTPNRKSDFKNSSVENNDYIINSLKSPNGKGFIVSKIDEALSFYDSKKNDIN	111
TgCPB	LNAFVSPESVESLFDISIVAEQVAVATSGNLTESAPRDRDSAPERRSDGGTLRGHPEDAGEL	120
* ↓		
TgCPL	-----PSPVED--HQPANIWEWKEAHFQDAF	116
SmPT1	EAS-----PSSITDGDADYPEKIWEWKDHHFQSQF	89
Cparvum	-----NTSKIDDFVPGDYVDPATREYRKSF	86
PfCP3	TGAENNNMNGIDDKDNKSVSVFNKNGNLKVNNNNQVSYSNLFDTKFLMDNLETVNLV	171
PfCP2	KYNEGNN-----NNADFKGLSLFKENTPSNNFIHN-KDYFINFFDNKFLMNAEHINQF	165
TgCPB	LRLLLDASEDMELWKANFFRHLTHSMRHIVRDSVLVSEKAFPSEQAGEAGRDGDLKKAL	180
* ↓		
TgCPL	SSFQAMAKSYATEEEKQRYAIFKNLVLVYIHTHQQ-GYSYSLKKN-HFGDLSRDEFRR	174
SmPT1	YQFQRDNHKKFYATEEERLKYAIFKNLTYIHNHMQ-GYSYVLKMN-K.GD.TLEEFQ	147
Cparvum	EEFKKKYHKVYSSMEENQRFYIKQNMFIKTNSQ-GFSYVLEMN-E.GD.SKEEFMA	144
PfCP3	YIFLKNNKQYETSEEMQKRFIFSENRYKIELHNKKTNSLYKRGMN-K.GD.SPEEFRS	230
PfCP2	YMFIKTNKQYNSPNEKRFQVFLQNAHKVNMHNNKNSLYKKELN-R.AD.TYHEFKN	224
TgCPB	EETGEDVWESRPASSNAVALIKKKMEKQAGKTGEGAGHTWEPEVSLR.RY.SLKDAKK	240
* ↓		
TgCPL	KYLGFKKSR--NLKSHHLGVATELLNVLPSL-----PAGVDWRSRCCVTPVKDQRDC	225
SmPT1	RYLGYKKP--DLRTPPREVDTTLESVEDNDI-----PTHVDWRQRC.TSVKD.GD.	197
Cparvum	RFTGYIKDSKDDERVFSSRSVASESEEEFVP-----PNSINWVEAGC.NPIRN.KD.	197
PfCP3	KYLNKLTGHGPFKTLSPFVSSEYANYEDVIKKYPADAKLDRIAYDWRHLGG.TPVKD.AL.	290
PfCP2	KYLSLRSSKPLKN-SKYLQDMQYEEVKKYR-GEENFDHAAAYDWRHLHG.TPVKD.KN.	282
TgCPB	LMGTFLVNTKVEGFPTPKGMPLPAKEFENATEFVPAHFADARTAFACKDV.GHVRD.GD.	300
* ↓		
TgCPL	GSCWAFSTTGALEGACAK-TGK-LVSLSEQELMDCSRAB--GNQSCSGGEMNDAFYQVL	281
SmPT1	SATGAMEGVYCAK-TGK-.VNL.Q.QLVD.SRFL--GNQ.G.D..RMEF.FEYVV	253
Cparvum	SAAALGATCAQ-TNRG.PSL.E.QFVD.SKQN--GNFG.D..TMGL.FOYAI	254
PfCP3	SSVGSVESQYAIR-KKA-.FLF.E.ELVD.SVK---NNG.Y..YITN.FDDMI	344
PfCP2	SSIGSVESQYAIR-KNK-.ITL.E.ELVD.SFK---NYG.N..LINN.FEDMI	336
TgCPB	ASTEAFNDRLCIRSQGR.MPL.A.HTTPS.CNAIHCAISFG.N..QPGM.WRWF	360
* ↓		
TgCPL	DSGGICS-----EDAYPYLARD-EECRAQSCCKVVKILG--	314
SmPT1	ENGGICS-----GENY.YMRKD-GVCKSSQ.TSVATITG--	286
Cparvum	KNKYLCT-----NDY.YFAEE-KTCMDSF.ENYIEIPVKA	289
PfCP3	DLGLGCS-----QDDY.YVSNLPETCNLKR.NERYTIKS--	378
PfCP2	ELGGICP-----DGDY.YVSDAPNLCNIDR.TEKYGIKN--	370
TgCPB	RKGVTGDFDALGKGTTCWPYEVFPFAHAKAPF.DCDATLVPRKTPK.RKDCEEQAYA	420
* ↓		
TgCPL	-----FKDVPRRSEAMKAALAK-SPVSIADQMPFQFYHEGVFDASCGT	360
SmPT1	-----YRSVPRRSEKSM.TALALRS.V.VAIQANQAA.QF.YD.IFDAPC.T	333
Cparvum	-----YKYVFPNINAL.TALAKYG.I.VAIQADQTP.QF.KS.VEDAPC.T	336
PfCP3	-----YVSIP--DDKF.EALRYLG.I.ISIAASDD-.AF.RG.FYDGECA	421
PfCP2	-----YLSVP--DNKL.EALRFLG.I.ISVAVSDD-.AF.KE.IFDGEC.D	413
TgCPB	DNVHPFDQDTHKATSAYSLSRSDV.RDMMTHG.V.GAFMVYED-.LS.KS.VYKHVS.L	479
* ↓		
TgCPL	DLD-HGVLLVGYGTDKESKDD-----FWIMKNSWGTGWGRDGYMYMAMHK-GEEGQC	410
SmPT1	NLD-.GVLLV.YSAETAGQGD-----YWIMK...GAA..KG.YMLMAMHK-.PAGQC	383
Cparvum	KVN-.GVVLV.YDMDEDTNKE-----YWLVR...GEA..EK.YIKLALHS-.KKGTC	386
PfCP3	APN-.AVILV.YGMKDIYNEDTGRMEKFYYYI IK...GSD..EG.YINLETDEN.YKKTC	480
PfCP2	ELN-.AVMLV.FGMKEIVNPLTKKGEKHYYYI IK...GQQ..ER.FINIETDES.LMRKC	472
TgCPB	PVGG.AIKII.WGTENGEE-----YWHAV...NTY..DG.QFKIAMGQC.IDGEM	529
* ↓		
TgCPL	GLLLDASFVPM-----	421
SmPT1	GVLLDGSFPVA-----	394
Cparvum	GILVEPVYPVINQSI-----	401
PfCP3	SIGTEAYVPLLE-----	492
PfCP2	GLGTDAFIPLIE-----	484
TgCPB	VAGEAAWQETEGVVNGEELPGQRAAGARAGAHAEEREEM	569

Figure 1. Alignment of TgCPL with related cathepsins

Deduced amino acid sequences of the cathepsin Ls from *Toxoplasma gondii* (TgCPL and TgCPB), *Sarcocystis muris* (SMPT1), *C. parvum* (Cparvum), and *P. falciparum* (PfCP2 and 3) were aligned with Clustal W. The 22 amino acid potential transmembrane domain is

underlined (59–82). The predicted pre/pro region cleavage site is marked with a single arrow and the pro/mature site with a double arrow. Conserved motifs (ERFNIN, KNFD, and SPV) are italicized and bold. The active site cysteine, histidine, and asparagine are shown in bold (229, 351, and 387). The potential N-glycosylation sites are marked with stars (*, positions 56 and 266).

Figure 2. Purified, active rTgCPL

rTgCPL was purified from concentrated *Pichia* media by anion exchange chromatography. Lane 1: Purified rTgCPL stained with Coomassie blue. Lane 2: Immunoblot of rTgCPL (500 ng) with monospecific rabbit antiserum. Lane 3: Tachyzoite lysate (10^7) reacted with rabbit monospecific antiserum.

Figure 3. pH optimum of active TgCPL

Points represent the mean $\pm$ S.E. of cleavage of the preferred peptide substrate, Z-KQLR-AMC in Relative Fluorescence Units (RFU)/sec at different pHs.

Figure 4. Structure of the active site pocket of TgCPL superimposed on the structure of human cathepsin L

The aspartic acid (Asp 216) can be seen at the base of the S2 pocket in TgCPL which is an alanine (Ala 214) in the human cathepsin.

Figure 5. Modeling of substrate specificity of TgCPL and cruzain

- A. Phenylalanine in the P2 position of a substrate shows poor fit in active site pocket of TgCPL.
- B. Accommodation of leucine in the P2 position of substrates in cruzain with alanine in the S2 site.

Figure 6. Localization of TgCPL

A. Localization of TgCPL (red) with insert of DIC. B. Immunoelectron micrograph of TgCPL in radiolucent vesicle (arrow). Size bar (0.5 microns) is shown below.

Figure 7. Amino acid sequence alignment of members of the chagasin family with toxostatin-1 and 2

Identical and semi-conserved amino acids are shown with asterisks and dots, respectively. The following sequences were from the NCBI protein database: *Trypanosoma cruzi* chagasin (**AJ299433**), *Entamoeba histolytica* EhICP1 (**Q6KCA4**), *Leishmania major* ICP (**AJ548878**). *Plasmodium* ICPs sequences were from Pandey et.al. [17] and the malaria genome project (<http://www.plasmodb.org/plasmo/>). Identical amino acids are shaded black, conservative amino acid changes are shaded grey. Conserved motifs are overlined. The N-terminal extension and the insertion in *Toxoplasma* and *Plasmodium* ICPs were removed before comparisons.

Figure 8. Recombinant and over expressed toxostatins

A. Purified, recombinant toxostatin-1 and 2 are stained with Coomassie blue. B. Immunoblot of lysates of wild type (RH) and tachyzoites over-expressing toxostatin-1 (Toxo-1++) detected with anti-AU-1 to the epitope tag.

Table 1

Enzyme Kinetics of TgCPL

Substrate	K _m (uM)	V _{max} (RFU/sec)
LR	7.0 ± 0.7	120.6 ± 3.6
FR	15.7 ± 1.0	148.0 ± 4.4
RR	17.6 ± 1.2	14.8 ± 0.3