

Published in final edited form as:

Adv Exp Med Biol. 2011 ; 712: 49–61. doi:10.1007/978-1-4419-8414-2_4.

Cathepsin proteases in *Toxoplasma gondii*

Zhicheng Dou¹ and Vern B. Carruthers^{1,2}

¹Department of Microbiology and Immunology, University of Michigan School of Medicine, 5740A Medical Science Building II, 1150 W. Medical Center Dr., Ann Arbor, MI, USA 48109-5620

Abstract

Cysteine proteases are important for the growth and survival of apicomplexan parasites that infect humans. The apicomplexan *Toxoplasma gondii* expresses five members of the C1 family of cysteine proteases, including one cathepsin L-like (TgCPL), one cathepsin B-like (TgCPB), and three cathepsin C-like (TgCPC1, 2 and 3) proteases. Recent genetic, biochemical and structural studies reveal that cathepsins function in microneme and rhoptry protein maturation, host cell invasion, replication, and nutrient acquisition.. Here, we review the key features and roles of *T. gondii* cathepsins and discuss the therapeutic potential for specific inhibitor development.

Keywords

cell invasion; degradation; endopeptidase; endosome; exopeptidase; lysosome; maturation

Introduction

Peptidases play a critical role in protein catabolism by hydrolysis of peptide bonds in the polypeptides.¹ Peptidases are classified into seven categories based on the principal catalytic residue in the active site: Aspartic, Cysteine, Glutamic, Serine, Threonine, Metallo and Mixed, each of which can be further divided into clans and families. Cysteine peptidases, also called thiol peptidases, use the nucleophilic thiol group of cysteine for hydrolysis. Cathepsin peptidases belonging to the C1 family clan CA of “papain-like” cysteine peptidases are widely distributed in eukaryotic organisms. During catalysis, a basic amino acid in the catalytic triad, usually histidine, de-protonates the cysteine thiol group, which attacks the carbonyl carbon group in the substrate for hydrolysis. The hydrolytic cycle is completed when the newly-derived substrate terminal amine accepts the proton from the active site histidine, thus regenerating the pre-hydrolysis status of the active site.²

General roles of cathepsins in biological systems

In eukaryotic cells, cathepsin proteases act classically as lysosomal hydrolases that digest endogenous and exogenous endocytosed polypeptides.³ However, increasingly it is appreciated that cathepsins can also play more specialized roles in higher eukaryotic organisms including spermatogenesis, antigen presentation, tumor invasion, degradation of matrix proteins, and TNF α -induced apoptosis.^{4–6} Upregulated cathepsin L enzyme in tumor cells plays an important role in nonmetastatic tumor cell conversion into a highly invasive metastatic state.^{7–9} In unicellular eukaryotes such as *Plasmodium falciparum*, cathepsin L-like proteases (falcipains 2a, 2b, and 3) are responsible for hemoglobin digestion within the parasite food vacuole during erythrocyte infection.^{10–15} Falcipain 1, which resides in intracellular vesicles, plays a yet-to-be-defined, non-essential role in parasite invasion of

²To whom correspondence should be addressed: vcarruth@umich.edu; Phone +1 734 763 2081; FAX +1 734 764 3562.

erythrocytes,^{16, 17} but it is required for efficient parasite development within infected mosquitoes.¹⁸ In *Trypanosoma b. brucei* and related species, cathepsin L and cathepsin B localize to endolysosomes.¹⁹ Cathepsin L is not essential but it facilitates traversal of the blood-brain barrier by *T. b. rhodesiense* in experimentally infected mice²⁰, possibly by activating a protease-activated receptor.²¹ *T. b. brucei* cathepsin B appears to be required for parasite survival²⁰ perhaps because of its role in degrading endocytosed host transferrin for iron acquisition.^{22, 23} Accordingly, like higher eukaryotes, protozoan parasites appear to use cathepsin endopeptidases for protein degradation and other specialized roles.

Properties of *Toxoplasma gondii* and its cathepsins

T. gondii is a ubiquitous apicomplexan parasite that infects a wide range of warm-blooded animals. It is estimated that almost one third of the human population is infected by this parasite.²⁴ Infection with *Toxoplasma* can lead to encephalitis, chorioretinitis and congenital birth defects, and AIDS and immunocompromised patients are at especially high risk of developing toxoplasmosis. As an obligate intracellular parasite, *T. gondii* must invade host cells to survive and expand the infection. Unlike intracellular bacteria and viruses, *T. gondii* and other related parasites use a unique gliding motility mechanism for invading host cells. *Toxoplasma* parasites lack specialized appendages for motility and instead utilize an intrapellicular actin-myosin system to slide on a substrate or a host-cell surface.²⁵ During cell invasion, two subcellular organelles, micronemes and rhoptries, sequentially discharge their contents at the apical end of the parasite to mediate entry.²⁶ As the parasite invades the host cell, a parasitophorous vacuole (PV) membrane is formed and surrounds the parasite.²⁷ After the parasite finishes entry, other organelles, termed dense granules (DG), secrete proteins into the PV. DG proteins are thought to function in modification of the PV for nutrient acquisition.²⁶ While most microneme and rhoptry proteins are subjected to limited proteolysis (a process termed proteolytic maturation) as they traffic to their respective secretory organelles, DG proteins are not processed and follow a constitutive secretion pathway to the PV.²⁸ Protease inhibitor studies revealed that cysteine proteases are involved in the maturation of microneme and rhoptry proteins and the biogenesis of some subcellular organelles.^{29–33}

The endosomal system in *T. gondii* is generally similar to that of other eukaryotic cells, but it also displays some unique features (Fig. 1). Recently, a dynamic Vacuolar Compartment (VAC) (also termed the Plant-Like Vacuole or PLV) was identified in the *Toxoplasma* endosomal system.^{33, 34} The VAC is often closely associated with the late endosome (LE) in the intermediate apical region of the parasite. The major population of the *T. gondii* cathepsin Ls (TgCPL) and cathepsin B (TgCPB, also termed toxopain-1 or CP-1) is distributed in the VAC^{(30, 33, 35} and Dou and Carruthers, unpublished data), suggesting this organelle may have a role similar to lysosomes. Although a low proportion of extracellular tachyzoites of *T. gondii* show endocytic ability,^{36, 37} it remains possible that endocytosis of parasite membrane proteins or host cell polypeptides and oligopeptides is more active in intracellular replicating parasites with a high nutrient demand.

Chromosome organization

Five cathepsin proteins are encoded in the genome of *T. gondii*: one cathepsin L-like protein (TgCPL), one cathepsin B-like protein (TgCPB), and three cathepsin C-like proteins (TgCPC1, 2 and 3) (Table 1). These five cathepsin genes are distributed among four distinct chromosomes. *Tgcpl* is encoded in chromosome Ib and has four exons while *Tgcpb* consists of seven exons and is encoded on chromosome XII. *Tgcp1* and *Tgcp3* are encoded on chromosome IX and therefore could have resulted from intra-chromosomal gene duplication, although they are separated by a considerable distance on the chromosome. *Tgcp1* and *Tgcp3* have nine and fourteen exons, respectively. *Tgcp2* has ten exons on

chromosome III. All *T. gondii* cathepsin genes have homologs among the three sequenced strain types of *T. gondii*, except *Tgcp3*. *Tgcp3* is not present in the genome sequence of the type I reference strain GT1 (www.toxdb.org), probably due to incomplete sequence coverage.

Catalytic residues and motifs

Toxoplasma cathepsins have almost identical amino acid sequences in their catalytic region as other members of the papain family. The conserved cysteine, histidine and asparagine residues form a triad in the active site that catalyzes peptide cleavage (Fig. 2). TgCPL has several conserved motifs including ERFNIN, which is a signature motif within the prodomain of cathepsin L and H proteases, a KNFD motif (also in the prodomain), and a SPV domain in the mature enzyme.³⁵ TgCPB has 12 cysteine residues in the mature form, which may participate in the formation of six disulfide bonds. A conserved motif (GCNKG) that exists in the human cathepsin B is also present in TgCPB. Besides endopeptidase activity, TgCPB also shows exopeptidase activity, which is contributed by an intact positively charged occluding loop that can bind to the terminal carboxylic acid of the target substrate.³⁰

Like human cathepsin C, also called dipeptidyl peptidase I (DPPI), all mature TgCPCs are composed of an N-terminal residual prodomain (exclusion domain), a heavy chain and a light chain. As a dipeptidyl peptidase, the aspartic residue at the N-terminus of the exclusion domain can block formation of the substrate binding cleft with the polypeptide beyond its S2 substrate-binding site.^{38, 39} Unlike other cathepsin-like enzymes, a halide ion is required for its activity, and a conserved tyrosine residue that binds a chloride ion in the crystal structures of the rat and human cathepsin C proteases also exists in *Toxoplasma* CPCs.^{40, 41} In addition, a tyrosine-based motif, YXXΦ (X is any amino acid and Φ is a bulky hydrophobic amino acid) is present in all *Toxoplasma* CPCs.³⁹ TgCPC2 and TgCPC3 also display a dileucine-motif, which in higher eukaryotes participates in endosomal/lysosomal protein targeting.⁴² TgCPC2 has a 100-residue C-terminal extension, the biological role of which remains unknown.

Three-dimensional structure of TgCPL and TgCPB

Recently, Larson *et al.* solved the three dimensional structure of autoactivated TgCPL with its propeptide resident in the active site cleft.³² The structure of mature TgCPL is very similar to previously determined papain-like cysteine protease structures. TgCPL consists of two domains divided by a deep active site cleft; one domain is primarily α -helical and the other contains a β -barrel-like fold and several α -helices.³² The Cys₃₁, His₁₆₇ and Asn₁₈₉ residues (numbered according to ref. ³²) cluster together and form the active site triad. Three disulfide bonds stabilize the TgCPL structure, a feature that is common in the papain-like cysteine proteases. The propeptide occupies the active cleft in a reverse orientation compared to that of the substrate, thus forming a stable interaction that inhibits its endopeptidase activity until activation and propeptide dissociation occurs.³² An aspartic acid residue occupies the S2 substrate-binding site, which is a major specificity-determining site. Although initial homology modeling predicted that the aspartic acid created an unusually shallow S2 substrate-binding site³⁵, the crystal structure of TgCPL revealed that the aspartic acid does not overtly occlude the pocket.³² Accordingly, TgCPL hydrolyzes synthetic substrates with a variety of hydrophobic amino acids in the P2 position.³³

Although a crystal structure of TgCPB is not available, homology modeling studies based on human lysosomal and rat cathepsin B show a glutamic residue at the base of the TgCPB S2 pocket,³⁰ which can interact with positively charged residues, such as arginine at the P2 site. An occluding loop that is close to the substrate-binding cleft blocks the C-terminal end of

the active site. Two positively charged histidine residues in this loop can associate with the C-terminal carboxylate group of the P2' residue to endow this enzyme with exopeptidase activity.⁴³

Activation profile and optimal pH for activity

Recombinant TgCPL and TgCPB can be auto-activated at low pH in the presence of a reducing agent.^{30, 32, 33} Recombinant TgCPL is active on both peptide and polypeptide substrates.³³ Recombinant TgCPC1 and TgCPC2 failed to auto-activate in vitro, indicating that they require exogenous cleavage by another protease molecule for activation.³⁹ The heavy chain and residual prodomain of TgCPC1, which remains associated with the active enzyme, were co-purified from a tachyzoite lysate, indicating that the expected activating proteolysis occurs in vivo.³⁹ Human DPPI protease is *trans*-activated by cathepsin L cleavage at several sites, but it remains to be determined if TgCPCs are activated by TgCPL.⁴⁴

Almost all cathepsin proteases perform optimally at low pH range.⁶ Recombinant TgCPL efficiently cleaves a recombinant proform of *T. gondii* MIC2-Associated Protein (proTgM2AP) at pH 5.5–6.0,³³ and the optimal pH range for digesting a Lys-Gln-Leu-Arg substrate is 5.5–6.5 with maximum activity at pH 6.5.³⁵ Recombinant TgCPB expressed in *E. coli* can self-cleave at pH 6.0, but currently the optimal pH for TgCPB proteolytic activity is unknown. TgCPC1 is active within a broad pH range from 4.5 to 8.0, and shows the highest activity near pH 6.5.³⁹

Phylogenetic relationships

A molecular phylogenetic analysis of the catalytic domains of papain-like proteases in apicomplexan parasites is shown in Figure 3. Cathepsin L- and B-like proteins are recognized by the presence of an “ERFNIN” motif in the prodomain or an occluding loop, respectively. Six members of the parvapaain family in *Theileria parva* are all cathepsin L-like, and are probably the earliest origin of the apicomplexan papain-like proteases. The occluding loop in cathepsin B may have been gained during evolution to grant it exopeptidase activity. Deletion of this loop by mutagenesis completely eliminates this activity.⁴³ Similarly, the N-terminal residual prodomain in cathepsin Cs (exclusion domain) may also have evolved from the prodomain of cathepsin L to associate with their catalytic heavy and light chains to contribute to their exclusive exopeptidase activity. The activation of human DPPI by cathepsin L and not cathepsin B may also indicate that cathepsin Cs are derived from cathepsin L and co-evolve with cathepsin B.⁴⁴ Among the apicomplexan parasites for which genome sequence is available, a cathepsin B-like protease only exists in *T. gondii*. The early ancestor of apicomplexan parasites, *Perkinsus marinus*⁴⁵ encodes ten cathepsin B-like proteases, implying that apicomplexans (except *T. gondii*) probably lost cathepsin B, perhaps due to redundancy with cathepsin L- and cathepsin C-like proteases.

Localization, physiological functions and regulation

TgCPL is principally expressed within the VAC of tachyzoites and bradyzoites.^{34–35} Its function in the VAC is not known, but since the VAC resembles a lysosome or lytic vacuole, TgCPL is proposed to function in protein degradation within this compartment. The VAC contains internal membrane tubules and vesicles typically seen in multivesicular bodies, a major center for membrane protein turnover. TgCPL is also associated with the residual body in the PV after cell division, where it could contribute to the destruction of mother cell organelles that are not partitioned into daughter cells.³³ Additional subpopulations of TgCPL are localised within small cytoplasmic vesicles throughout the cytoplasm and within the late endosomes (LE), which is characterized by the presence of

TgRab7, vacuolar pyrophosphatase 1 (TgVP1), and immature microneme proteins also known as proMICs.^{46–48} ProMICs display an N-terminal or internal propeptide that is proteolyzed within 15–60 minutes after synthesis.^{33, 49, 50} Parussini *et al* found that TgCPL contributes to the proteolytic maturation of proTgM2AP and proTgMIC3 based on delayed maturation in a TgCPL-deficient strain and correct processing of recombinant proTgM2AP by recombinant TgCPL *in vitro*.³³ Moreover, the propeptide cleavage sites of proTgM2AP and proTgMIC3 contain residues favorable for TgCPL recognition and proteolysis based on screening of a peptide substrate library and mapping of the autocatalytic cleavage site of recombinant TgCPL. Maturation of proTgM2AP and proTgMIC3 was not completely abolished in TgCPL-deficient parasites, suggesting the existence of alternative maturase(s) within the *T. gondii* endocytic pathway.

TgCPL proteolytic activity is regulated in several ways. First, like most papain-like enzymes, TgCPL is initially synthesized as an inactive zymogen with the propeptide sterically occluding the active site cleft.³² This arrangement probably prevents TgCPL activity as it traffics through the proximal secretory system (ER, Golgi), thereby avoiding proteolysis of inappropriate substrates. Second, TgCPL activity is pH regulated, with optimal activity occurring under moderately acidic conditions (pH 5.0–6.5),^{33, 35} which TgCPL probably encounters as it enters the endocytic system and undergoes maturation to remove the autoinhibitory propeptide. Third, *T. gondii* expresses two inhibitors of cysteine proteases termed toxostatin-1 and toxostatin-2.³⁵ Overexpression of toxostatin-1 in *T. gondii* tachyzoites inhibited TgCPL and TgCPB activity by 80–90% in parasite extracts. Although it remains unclear if this inhibitor encounters TgCPL or TgCPB within intact cells, toxostatin-1 and toxostatin-2 could serve to prevent unwanted proteolysis by TgCPL or TgCPB that strays to an inappropriate location within the parasite. Finally, TgCPL processing of its substrates is likely regulated spatially by membrane segregation via delivery of substrates to the VAC or shuffling diminutive quantities of TgCPL to the LE for limited proteolysis of proMICs and possibly other substrates.³³

TgCPB expression and activity have been detected in tachyzoites. It remains uncertain whether TgCPB is expressed in bradyzoite tissue cysts since its transcript was not detected by RT-PCR in cysts recovered from mouse brains,³⁵ but was shown by microarray analysis to increase upon induction of bradyzoite differentiation *in vitro* (www.toxodb.org). TgCPB occupies the rhoptries based on immunoelectron and immunofluorescence microscopical studies,³⁰ although it was not identified in the rhoptry proteome.⁵¹ TgCPB was also detected in the residual body and an electron lucent vacuole that is probably the VAC,³⁰ raising the possibility that TgCPB works in concert with TgCPL at these locations. Parasites treated with a cathepsin inhibitor showed impaired invasion, altered rhoptry morphology, and delayed maturation of TgROP2, implicating TgCPB in ROP protein maturation and parasite invasion.³⁰ An antisense inhibition approach was used to knockdown the expression of TgCPB, resulting in decreases in parasite cell invasion, replication, and parasite tissue burden in a chicken embryo model of infection, indicating that TgCPB contributes to multiple processes during infection.³⁰

TgCPC1 and TgCPC2 are expressed in tachyzoites, with TgCPC1 transcripts being ~20-fold more abundant than TgCPC2.³⁹ None of the TgCPC transcripts including TgCPC3 were detected in bradyzoite tissue cysts. TgCPC3 is expressed in the sporozoite stage based on the presence of expressed sequence tags (www.toxodb.org). TgCPC1 and TgCPC2 have been reported to occupy DGs and the PV during parasite replication.³⁹ Partial gene disruption of TgCPC1 by replacement of exons 3–5 with a selectable marker resulted in ~4-fold upregulation of TgCPC2 mRNA, suggesting that TgCPC2 compensates for the loss of TgCPC1 and that these proteases might function in the same process or pathway.³⁹ A selective inhibitor of cathepsin C partially impaired parasite replication and reduced parasite

tissue burden in experimentally infected chicken embryos, implicating TgCPCs in parasite replication.³⁹ The inhibitor also stabilized the expression of *E. coli* β -lactamase within the parasitophorous vacuole of transgenic parasites, suggesting that TgCPCs function in the degradation of exogenous proteins. Proteins derived from the host endoplasmic reticulum were recently shown to be present within the lumen of the PV in infected dendritic cells, indicating that such proteins are available to contribute to the parasites nutritional needs.⁵²

Therapeutic potential

T. gondii cathepsins are considered potential therapeutic targets based on genetic and inhibitor studies. For example, genetic disruption of TgCPL diminishes parasite cell invasion and growth (ref. ³³ and Dou and Carruthers, unpublished data). Also, parasite treatment with the cathepsin inhibitor morpholinurea-leucyl-homophenyl-vinyl sulfone phenyl (LHVS, also known as K11017) impairs cell invasion by blocking secretion of adhesive proteins from parasite micronemes.³¹ LHVS principally targets TgCPL based on analysis with a fluorescent derivative of LHVS,³² but our recent findings suggest that it can also inhibit TgCPB (Dou and Carruthers, unpublished data). Antisense inhibition of TgCPB expression or treatment with cathepsin inhibitors diminished parasite replication, cell invasion, and infection in vivo.^{30, 53} Similarly, targeted deletion or chemical inhibition of TgCPC1 reduced parasite replication and infection.³⁹ Nevertheless, it should be noted that none of the *T. gondii* cathepsins have been validated as essential enzymes, and the cathepsin inhibitors that have been tested to date show relatively low potency with effective concentrations in the low- to mid-micromolar range.^{30, 39, 53, 54} It remains unclear whether this is due to poor penetration into *T. gondii* infected cells or partial refractivity to inhibition of the parasite cathepsins within the intracellular environment. Additional genetic ablation studies should provide a clearer picture of the importance of these proteases and their potential for therapeutic development.

Conclusions and future perspectives

The emerging view of *T. gondii* cathepsins is that, like their homologs in other eukaryotes, they function in protein degradation along with playing more specialized roles in the maturation of invasion proteins. However, much remains to be done including the identification of their full range of protein substrates within the parasite, their dependency on one another for activation, and their participation in similar or distinct processes within the endocytic system and parasitophorous vacuole. Additional reverse genetic evidence that TgCPB and the TgCPCs are important to parasite survival would further boost their stock as potential targets. The screening, identification and target validation of small molecule inhibitors with greater potency will also be an important avenue of future work. Recent advances in gene tagging for marker identification should illuminate additional features of the *T. gondii* endosomal system and help identify the pathways taken by cathepsins to reach their subcellular locations.⁵⁵ This goal will also be facilitated by the use of conditionally expressed dominant negative mutants of membrane trafficking determinants.^{56, 57} Finally, now that evidence is emerging showing that host ER proteins gain access to the PV, studies of parasite endocytic uptake within infected cells might shed light on whether *T. gondii*, like its kin the malaria parasite, taps host proteins for nutritional gain.

Acknowledgments

We thank My-Hang Huynh for critically reading the manuscript. Work in the Carruthers lab is supported by grants R01AI063263 (V.B.C.) from the National Institutes of Health (USA) and grant 04R-796 (V.B.C.) from the Stanley Medical Research Institute (USA).

Abbreviations used

DG	dense granule
LE	late endosome
LHVS	morpholinurealeucyl-homophenyl-vinyl sulfone phenyl
PLV	plant-like vacuole
PV	parasitophorous vacuole
VAC	vacuolar compartment

References

1. Lopez-Otin C, Bond JS. Proteases: Multifunctional enzymes in life and disease. *J Biol Chem.* 2008; 283:30433–30437. [PubMed: 18650443]
2. Rawlings ND, Barrett AJ, Bateman A. MEROPS: The peptidase database. *Nucleic Acids Res.* 2010; 38:D227–D233. [PubMed: 19892822]
3. Barrett AJ, Kirschke H. Cathepsin B, cathepsin H, and cathepsin L. *Methods Enzymol.* 1981; 80 Pt C:535–561. [PubMed: 7043200]
4. Kirschke, H.; Barrett, AJ.; Rawlings, ND. *Lysosomal Cysteine Proteases.* 2nd ed. Oxford; New York: Oxford University Press; 1998.
5. Guicciardi ME, Deussing J, Miyoshi H, et al. Cathepsin B contributes to TNF-alpha-mediated hepatocyte apoptosis by promoting mitochondrial release of cytochrome c. *J Clin Invest.* 2000; 106:1127–1137. [PubMed: 11067865]
6. Barrett, AJ.; Rawlings, ND.; Woessner, JF. *Handbook of Proteolytic Enzymes.* San Diego: Academic Press; 2004.
7. Denhardt DT, Greenberg AH, Egan SE, et al. Cysteine proteinase cathepsin L expression correlates closely with the metastatic potential of H-ras-transformed murine fibroblasts. *Oncogene.* 1987; 2:55–59. [PubMed: 3438085]
8. Frade R, Rodrigues-Lima F, Huang S, et al. Procathepsin-L, a proteinase that cleaves human C3 (the third component of complement), confers high tumorigenic and metastatic properties to human melanoma cells. *Cancer Res.* 1998; 58:2733–2736. [PubMed: 9661883]
9. Amuthan G, Biswas G, Zhang SY, et al. Mitochondria-to-nucleus stress signaling induces phenotypic changes, tumor progression and cell invasion. *EMBO J.* 2001; 20:1910–1920. [PubMed: 11296224]
10. Dluzewski AR, Rangachari K, Wilson RJM, et al. *Plasmodium falciparum* : Protease inhibitors and inhibition of erythrocyte invasion. *Exp Parasitol.* 1986; 62:416–422. [PubMed: 3536568]
11. Rosenthal PJ, McKerrow JH, Aikawa M, et al. A malarial cysteine proteinase is necessary for hemoglobin degradation by *Plasmodium falciparum*. *J Clin Invest.* 1988; 82:1560–1566. [PubMed: 3053784]
12. Rosenthal PJ, McKerrow JH, Rasnick D, et al. *Plasmodium falciparum* : Inhibitors of lysosomal cysteine proteinases inhibit a trophozoite proteinase and block parasite development. *Mol Biochem Parasitol.* 1989; 35:177–183. [PubMed: 2671722]
13. Rosenthal PJ, Wollish WS, Palmer JT, et al. Antimalarial effects of peptide inhibitors of a *Plasmodium falciparum* cysteine proteinase. *J Clin Invest.* 1991; 88:1467–1472. [PubMed: 1939639]
14. Rosenthal PJ. *Plasmodium falciparum* : Effects of proteinase inhibitors on globin hydrolysis by cultured malaria parasites. *Exp Parasitol.* 1995; 80:272–281. [PubMed: 7895837]
15. Rosenthal PJ, Olson JE, Lee GK, et al. Antimalarial effects of vinyl sulfone cysteine proteinase inhibitors. *Antimicrob Agents Chemother.* 1996; 40:1600–1603. [PubMed: 8807047]
16. Greenbaum DC, Baruch A, Grainger M, et al. A role for the protease falcipain 1 in host cell invasion by the human malaria parasite. *Science.* 2002; 298:2002–2006. [PubMed: 12471262]

17. Sijwali PS, Rosenthal PJ. Gene disruption confirms a critical role for the cysteine protease falcipain-2 in hemoglobin hydrolysis by *Plasmodium falciparum*. *Proc Natl Acad Sci U S A*. 2004; 101:4384–4389. [PubMed: 15070727]
18. Eksi S, Czesny B, Greenbaum DC, et al. Targeted disruption of *Plasmodium falciparum* cysteine protease, falcipain 1, reduces oocyst production, not erythrocytic stage growth. *Mol Microbiol*. 2004; 53:243–250. [PubMed: 15225318]
19. Caffrey CR, Hansell E, Lucas KD, et al. Active site mapping, biochemical properties and subcellular localization of rhodesain, the major cysteine protease of *Trypanosoma brucei rhodesiense*. *Mol Biochem Parasitol*. 2001; 118:61–73. [PubMed: 11704274]
20. Abdulla MH, O'Brien T, Mackey ZB, et al. RNA interference of *Trypanosoma brucei* cathepsin B and L affects disease progression in a mouse model. *PLoS Negl Trop Dis*. 2008; 2:e298. [PubMed: 18820745]
21. Grab DJ, Garcia-Garcia JC, Nikolskaia OV, et al. Protease activated receptor signaling is required for african trypanosome traversal of human brain microvascular endothelial cells. *PLoS Negl Trop Dis*. 2009; 3:e479. [PubMed: 19621073]
22. Mackey ZB, O'Brien TC, Greenbaum DC, et al. A cathepsin B-like protease is required for host protein degradation in *Trypanosoma brucei*. *J Biol Chem*. 2004; 279:48426–48433. [PubMed: 15326171]
23. O'Brien TC, Mackey ZB, Fetter RD, et al. A parasite cysteine protease is key to host protein degradation and iron acquisition. *J Biol Chem*. 2008; 283:28934–28943. [PubMed: 18701454]
24. Montoya JG, Liesenfeld O. Toxoplasmosis. *Lancet*. 2004; 363:1965–1976. [PubMed: 15194258]
25. 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]
26. Carruthers VB. Proteolysis and *Toxoplasma* invasion. *Int J Parasitol*. 2006; 36:595–600. [PubMed: 16600244]
27. Hakansson S, Charron AJ, Sibley LD. *Toxoplasma* evacuoles: A two-step process of secretion and fusion forms the parasitophorous vacuole. *EMBO J*. 2001; 20:3132–3144. [PubMed: 11406590]
28. V K, Qi H, Beckers CJ, et al. The protozoan parasite *Toxoplasma gondii* targets proteins to dense granules and the vacuolar space using both conserved and unusual mechanisms. *J Cell Biol*. 1998; 141:1323–1333. [PubMed: 9628889]
29. Shaw MK, He CY, Roos DS, et al. Proteasome inhibitors block intracellular growth and replication of *Toxoplasma gondii*. *Parasitology*. 2000; 121(Pt 1):35–47. [PubMed: 11085223]
30. 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]
31. Teo CF, Zhou XW, Bogyo M, et al. Cysteine protease inhibitors block *Toxoplasma gondii* microneme secretion and cell invasion. *Antimicrob Agents Chemother*. 2007; 51:679–688. [PubMed: 17145790]
32. Larson ET, Parussini F, Huynh MH, et al. *Toxoplasma gondii* cathepsin l is the primary target of the invasion inhibitory compound LHVS. *J Biol Chem*. 2009
33. Parussini F, Coppens I, Shah PP, et al. Cathepsin L occupies a vacuolar compartment and is a protein maturase within the endo/exocytic system of *Toxoplasma gondii*. *Mol Microbiol*. 2010 in press.
34. Miranda K, Pace DA, Cintron R, et al. Characterization of a novel organelle in *Toxoplasma gondii* with similar composition and function to the plant vacuole. *Mol Microbiol*. 2010
35. Huang R, Que X, Hirata K, et al. The cathepsin L of *Toxoplasma gondii* (TgCPL) and its endogenous macromolecular inhibitor, toxostatin. *Mol Biochem Parasitol*. 2009; 164:86–94. [PubMed: 19111576]
36. Nichols BA, Chiappino ML, Pavesio CE. Endocytosis at the micropore of *Toxoplasma gondii*. *Parasitol Res*. 1994; 80:91–98. [PubMed: 8202461]
37. Botero-Kleiven S, V F, Lindh J, et al. *Receptor-mediated* endocytosis in an apicomplexan parasite (*Toxoplasma gondii*). *Exp Parasitol*. 2001; 98:134–144. [PubMed: 11527436]

38. Turk D, Janjic V, Stern I, et al. Structure of human dipeptidyl peptidase I (cathepsin C): Exclusion domain added to an endopeptidase framework creates the machine for activation of granular serine proteases. *EMBO J.* 2001; 20:6570–6582. [PubMed: 11726493]
39. Que X, Engel JC, Ferguson D, et al. Cathepsin cs are key for the intracellular survival of the protozoan parasite, *Toxoplasma gondii*. *J Biol Chem.* 2007; 282:4994–5003. [PubMed: 17164247]
40. Fruton JS, Mycek MJ. Studies on beef spleen cathepsin C. *Arch Biochem Biophys.* 1956; 65:11–20. [PubMed: 13373403]
41. McDonald JK, Reilly TJ, Zeitman BB, et al. Cathepsin C: A chloride-requiring enzyme. *Biochem Biophys Res Commun.* 1966; 22:771–775. [PubMed: 5970511]
42. Storch S, Pohl S, Braulke T. A dileucine motif and a cluster of acidic amino acids in the second cytoplasmic domain of the batten disease-related CLN3 protein are required for efficient lysosomal targeting. *J Biol Chem.* 2004; 279:53625–53634. [PubMed: 15469932]
43. Illy C, Quraishi O, Wang J, et al. Role of the occluding loop in cathepsin B activity. *J Biol Chem.* 1997; 272:1197–1202. [PubMed: 8995421]
44. Dahl SW, Halkier T, Lauritzen C, et al. Human recombinant pro-dipeptidyl peptidase I (cathepsin C) can be activated by cathepsins L and S but not by autocatalytic processing. *Biochemistry.* 2001; 40:1671–1678. [PubMed: 11327826]
45. Leander BS, Clopton RE, Keeling PJ. Phylogeny of gregarines (Apicomplexa) as inferred from small-subunit rDNA and beta-tubulin. *Int J Syst Evol Microbiol.* 2003; 53:345–354. [PubMed: 12656194]
46. Harper JM, Huynh MH, Coppens I, et al. A cleavable propeptide influences *Toxoplasma* infection by facilitating the trafficking and secretion of the TgMIC2-M2AP invasion complex. *Mol Biol Cell.* 2006; 17:4551–4563. [PubMed: 16914527]
47. Brydges SD, Harper JM, Parussini F, et al. A transient forward-targeting element for microneme-regulated secretion in *Toxoplasma gondii*. *Biol Cell.* 2008; 100:253–264. [PubMed: 17995454]
48. El Hajj H, Papoin J, Cerede O, et al. Molecular signals in the trafficking of *Toxoplasma gondii* protein MIC3 to the micronemes. *Eukaryot Cell.* 2008; 7:1019–1028. [PubMed: 18390648]
49. Rabenau KE, Sohrabi A, Tripathy A, et al. TgM2AP participates in *Toxoplasma gondii* invasion of host cells and is tightly associated with the adhesive protein TgMIC2. *Mol Microbiol.* 2001; 41:537–547. [PubMed: 11532123]
50. Brydges SD, Sherman GD, Nockemann S, et al. Molecular characterization of TgMIC5, a proteolytically processed antigen secreted from the micronemes of *Toxoplasma gondii*. *Mol Biochem Parasitol.* 2000; 111:51–66. [PubMed: 11087916]
51. Bradley PJ, Ward C, Cheng SJ, et al. Proteomic analysis of rhoptry organelles reveals many novel constituents for host-parasite interactions in *Toxoplasma gondii*. *J Biol Chem.* 2005; 280:34245–34258. [PubMed: 16002398]
52. Goldszmid RS, Coppens I, Lev A, et al. Host ER-parasitophorous vacuole interaction provides a route of entry for antigen cross-presentation in *Toxoplasma gondii*-infected dendritic cells. *J Exp Med.* 2009; 206:399–410. [PubMed: 19153244]
53. Que X, Wunderlich A, Joiner KA, et al. Toxopain-1 is critical for infection in a novel chicken embryo model of congenital toxoplasmosis. *Infect Immun.* 2004; 72:2915–2921. [PubMed: 15102804]
54. Shaw MK, Roos DS, Tilney LG. Cysteine and serine protease inhibitors block intracellular development and disrupt the secretory pathway of *Toxoplasma gondii*. *Microbes Infect.* 2002; 4:119–132. [PubMed: 11880042]
55. Huynh MH, Carruthers VB. Tagging of endogenous genes in a *Toxoplasma gondii* strain lacking Ku80. *Eukaryot Cell.* 2009; 8:530–539. [PubMed: 19218426]
56. Agop-Nersesian C, Naissant B, Ben Rached F, et al. Rab11A-controlled assembly of the inner membrane complex is required for completion of apicomplexan cytokinesis. *PLoS Pathog.* 2009; 5:e1000270. [PubMed: 19165333]
57. Breinich MS, Ferguson DJ, Foth BJ, et al. A dynamin is required for the biogenesis of secretory organelles in *Toxoplasma gondii*. *Curr Biol.* 2009; 19:277–286. [PubMed: 19217293]

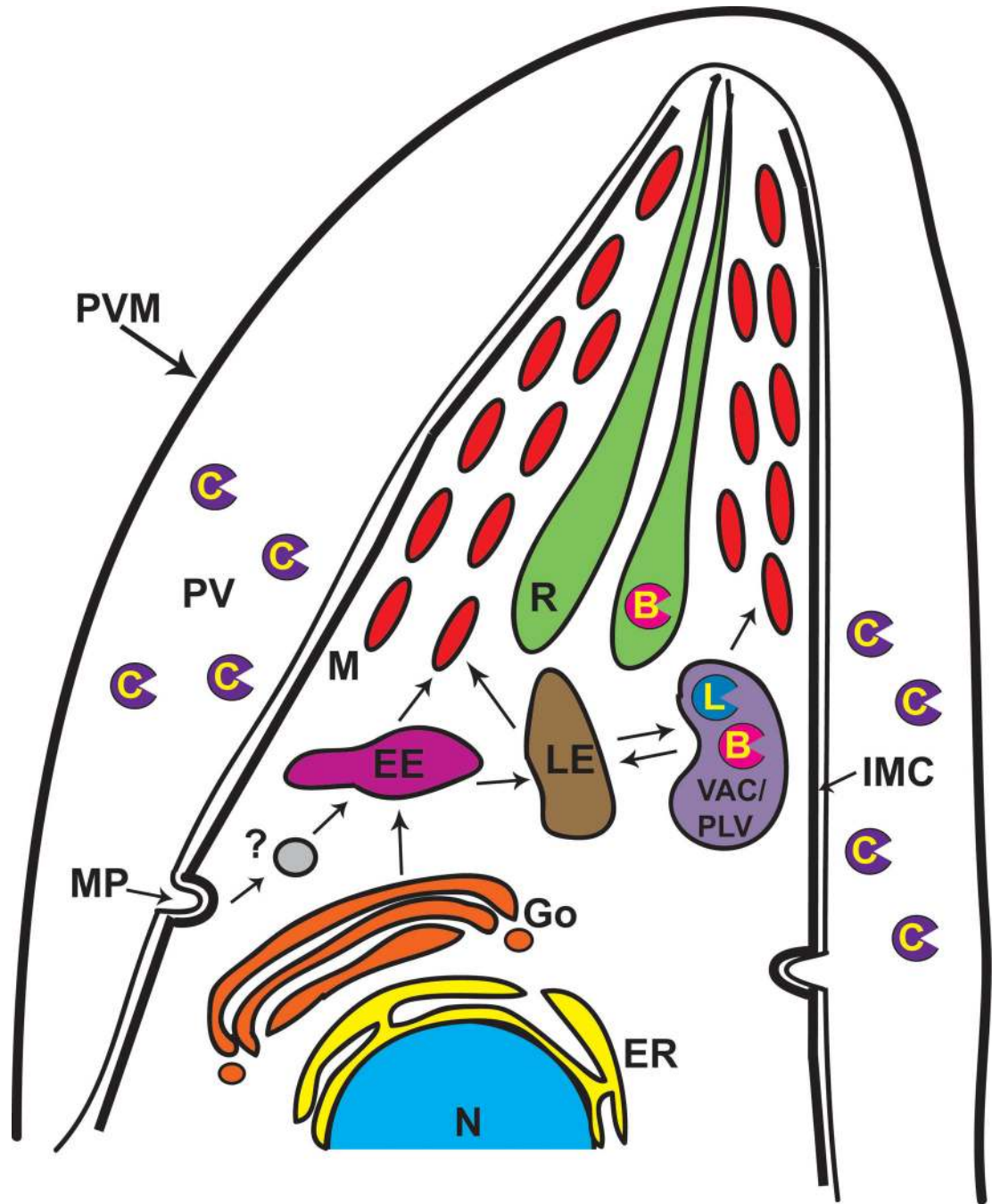


Figure 1. The endosomal system of *T. gondii* and the subcellular locations of cathepsins. TgCPL and TgCPB are predominantly expressed in the VAC and a diminutive amount of TgCPL is seen in the late endosome (LE) where it has been implicated in the maturation of pro-microneme proteins.^{33, 34} TgCPB was also reported to be distributed in the rhoptry to function in the processing of pro-rhoptry proteins.³⁰ Immunofluorescence microscopical studies revealed that TgCPC1 protein is secreted into the PV after cell invasion, where it may digest exogenous proteins to meet the parasite's nutrient needs.³⁹ Exogenous polypeptides may also be endocytosed and trafficked to the VAC for nutrient acquisition. Abbreviations used:

B, cathepsin B-like protease; C, cathepsin C-like protease; EE, early endosome; ER, endoplasmic reticulum; Go, Golgi apparatus; IMC, inner membrane complex; L, cathepsin-L like protease, LE, late endosome; M, microneme; MP, micropore; N, nucleus; PLV, plant-like vacuole; PV, parasitophorous vacuole; PVM, parasitophorous vacuole membrane; R, rhoptry; VAC, vacuolar compartment.

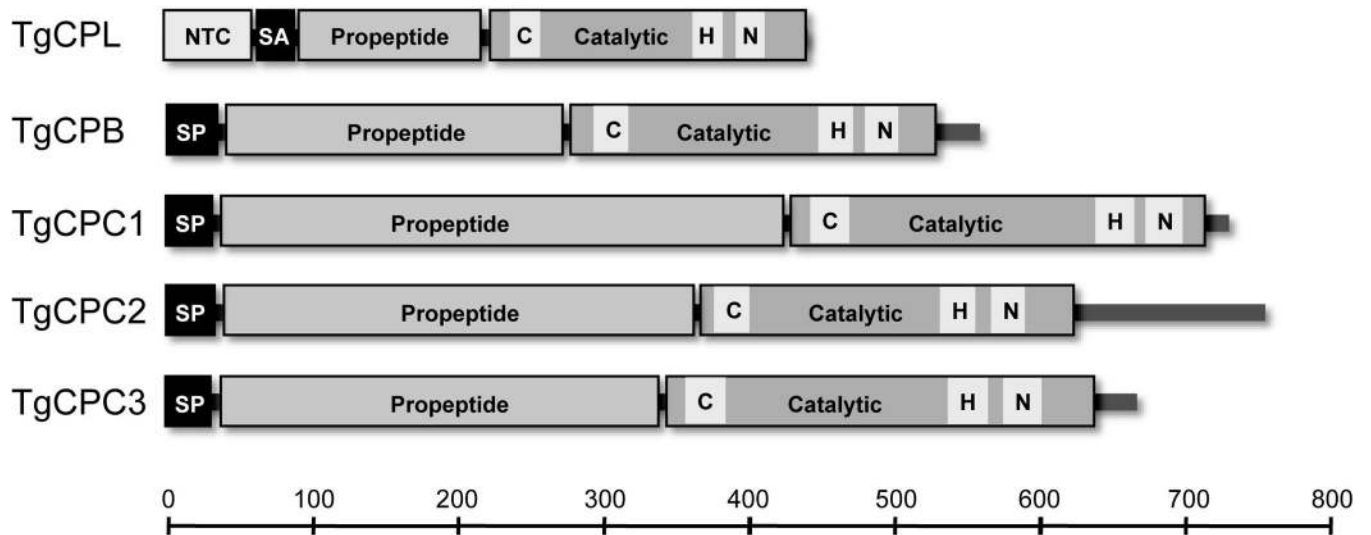


Figure 2. Schematic representations of *T. gondii* cathepsin-like proteases. Domains are represented by rectangles whereas joining regions or extensions are depicted as elongated bars. The positions of catalytic triad residues (C, H, and N) are indicated. Abbreviations used: NTC, N-terminal cytosolic; SA, signal anchor; SP, signal peptide. Scale is in amino acids.

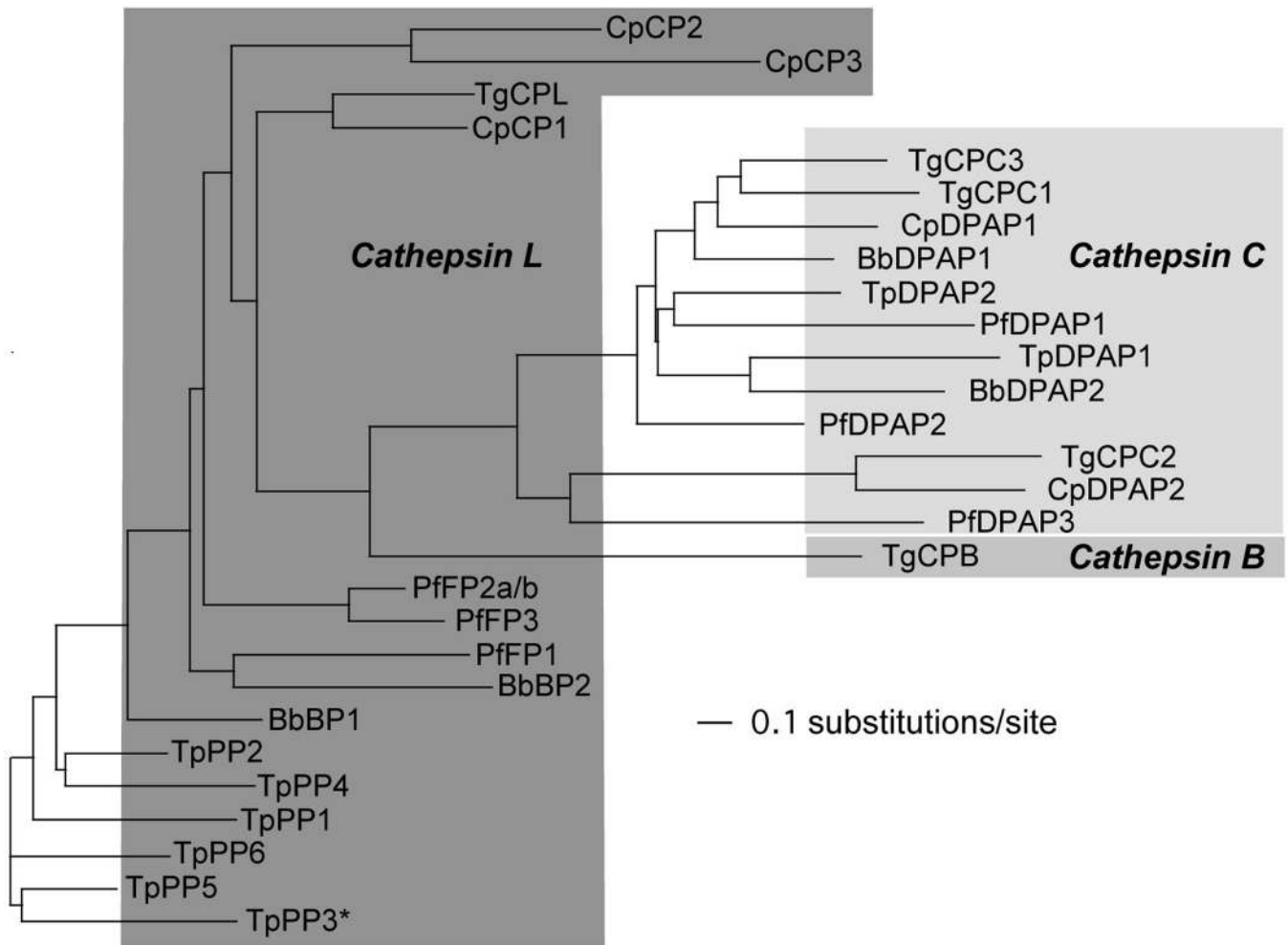




Figure 3. Sequence-based relationships among *T. gondii* cathepsins. (A) Molecular phylogenetic relationships among cathepsin proteases in apicomplexan parasites. The tree was generated by neighbor-joining analysis using POWER (<http://power.nhri.org.tw/power/home.htm>). Subgroups are shaded according to their similarity to cathepsins based on homology and the presence of an “ERFNIN” motif (cathepsin L-like) or occluding loop (cathepsin B-like). The analysis was restricted to apicomplexan parasites with complete or nearly complete genome sequences. One species from each genera was selected based on the maximal genome sequence coverage. *Plasmodium* SERA proteins were excluded due to their substantial divergence. The *Plasmodium falciparum* PfFP2a and PfFP2b are identical in sequence and thus are shown in the same dendrite. Note that the *Babesia* and *Theileria* proteases have not been systematically named previously and hence are designated here according to the nomenclature adopted for the *Plasmodium* and *Cryptosporidium* cathepsins. An asterisk indicates a protease missing at least one amino acid involved in catalysis. Abbreviations used: Bb, *Babesia bovis*; BP, Bovipain; Cp, *Cryptosporidium parvum*; CP, Cryptopain; Pf, *Plasmodium falciparum*; FP, falcipain; Tg, *Toxoplasma gondii*; Tp, *Theileria parva*; PP, parvapain. (B) Multiple sequence alignment of apicomplexan cathepsin catalytic-proximal sequences. Sequences obtained from the MEROPS database (<http://merops.sanger.ac.uk>) include only the catalytic domain beginning six amino acids upstream of the catalytic cysteine and ending two amino acids downstream of the active site asparagine. Sequence alignment was compiled using ClustalW (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>). Fully conserved residues are indicated by asterisks, and highly similar and similar residues are indicated by two dots and one dot, respectively. TpPP3 lacks the conserved cysteine residue in the catalytic site, which is replaced with an alanine.

Table I

Features of *T. gondii* cathepsin-like proteases.

Protease	Length (aa)	Catalytic domain ¹	Mature MW	Activity ³	Chromosome	Exons	Optimal pH	Activation	Stage Expression	Substrate specificity
TgCPL	422	206–420	30 kDa	Endo	Ib	4	5.5–6.0 (polypeptide) 6.5 (synthetic peptides)	Autoactivation in vitro	Tachyzoite and bradyzoite	P2 hydrophobic
TgCPB	569	275–532	28 kDa	Endo/Exo	XII	7	ND	Autoactivation in vitro	Tachyzoite and bradyzoite (microarray analysis)	P2 hydrophobic
TgCPC1	733	412–709	35 kDa ²	Exo	IX	9	6.5	No autoactivation in vitro	Tachyzoite	17-fold higher activity for Gly-Arg than human cathepsin-C specific substrate Gly-Phe
TgCPC2	753	357–630	44 kDa ²	Exo	III	10	ND	No autoactivation in vitro	Tachyzoite	ND
TgCPC3	622	335–647	32 kDa ²	Exo	IX	14	ND	ND	Sporozoite (expressed sequence tags)	ND

¹Numbered from the initiator methionine.

²MW is based on the size of the catalytic domain without further processing into the heavy and light chains, if occurring.

³Endo, endopeptidase; Exo, exopeptidase