Metallopeptidases of Toxoplasma gondii : in silico identification and gene expression

Sandie Escotte-Binet, Antoine Huguenin, Dominique Aubert, Anne-Pascaline Martin, Matthieu Kaltenbach, Isabelle Florent, Isabelle Villena

Parasite **25**, 26 (2018)

https://doi.org/10.1051/parasite/2018025

PARASITE

Available online at:
www.parasite-journal.org

RESEARCH ARTICLE OPEN ACCESS

Metallopeptidases of *Toxoplasma gondii*: *in silico* identification and gene expression

Sandie Escotte-Binet[1,2], Antoine Huguenin[1,2], Dominique Aubert[1,2], Anne-Pascaline Martin[1], Matthieu Kaltenbach[1], Isabelle Florent[3], and Isabelle Villena[1,2,*]

[1] EA 7510, ESCAPE, Laboratory of Parasitology-Mycology, University of Reims Champagne-Ardenne, 51100 Reims, France
[2] Laboratory of Parasitology-Mycology, Toxoplasmosis National Reference Center, Toxoplasma Biological Resource Center, Maison Blanche Hospital, 51100 Reims, France
[3] UMR7245 CNRS-MNHN, National Museum of Natural History, Department Adaptations of the Living, 75005 Paris, France

Received 15 September 2017, Accepted 16 April 2018, Published online 8 May 2018

Abstract – Metallopeptidases are a family of proteins with domains that remain highly conserved throughout evolution. These hydrolases require divalent metal cation(s) to activate the water molecule in order to carry out their catalytic action on peptide bonds by nucleophilic attack. Metallopeptidases from parasitic protozoa, including *Toxoplasma*, are investigated because of their crucial role in parasite biology. In the present study, we screened the *T. gondii* database using PFAM motifs specific for metallopeptidases in association with the MEROPS peptidase Database (release 10.0). In all, 49 genes encoding proteins with metallopeptidase signatures were identified in the *Toxoplasma* genome. An Interpro Search enabled us to uncover their domain/motif organization, and orthologs with the highest similarity by BLAST were used for annotation. These 49 *Toxoplasma* metallopeptidases clustered into 15 families described in the MEROPS database. Experimental expression analysis of their genes in the tachyzoite stage revealed transcription for all genes studied. Further research on the role of these peptidases should increase our knowledge of basic *Toxoplasma* biology and provide opportunities to identify novel therapeutic targets. This type of study would also open a path towards the comparative biology of apicomplexans.

Keywords: *Toxoplasma gondii*, metallopeptidase, endopeptidase, carboxypeptidase, aminopeptidase, enzymatic activity

Résumé – Métallopeptidases de *Toxoplasma gondii* : identification *in silico* et expression génique. Les métallopeptidases sont une famille de protéines dont les domaines restent hautement conservés tout au long de l'évolution. Ces hydrolases nécessitent un ou plusieurs cations métalliques divalents pour activer la molécule d'eau afin de réaliser leur action catalytique sur la liaison peptidique par une attaque nucléophile. Les métallopeptidases provenant de protozoaires parasites, y compris *Toxoplasma*, sont étudiées en raison de leur rôle crucial dans la biologie du parasite. Dans la présente étude, nous avons examiné la base de données de *T. gondii* en utilisant des motifs PFAM spécifiques des métallopeptidases en association avec la base de données de peptidases MEROPS (version 10.0). Quarante-neuf gènes encodant des protéines avec des signatures de métallopeptidases ont été identifiés dans le génome de *Toxoplasma*. L'utilisation d'Interpro Search a permis de découvrir l'organisation des domaines / motifs ainsi que les orthologues présentant la plus haute similarité par BLAST et ont été utilisés pour l'annotation. Ces 49 métallopeptidases de *Toxoplasma* ont été regroupées dans les 15 familles décrites dans la base de données MEROPS. L'analyse expérimentale de l'expression de leurs gènes au stade tachyzoïte a révélé la transcription de tous les gènes étudiés. D'autres recherches sur l'implication de ces peptidases devraient accroitre notre connaissance de la biologie fondamentale de *Toxoplasma* et fournir des opportunités pour identifier de nouvelles cibles thérapeutiques. Une telle étude ouvre également la voie à la biologie comparée des Apicomplexa.

*Corresponding author: ivillena@chu-reims.fr

Introduction

Toxoplasma gondii is an obligate intracellular apicomplexan protozoan parasite that is responsible for toxoplasmosis in humans and animals. Although toxoplasmosis is generally clinically asymptomatic in healthy individuals, it may cause severe complications and become opportunistic in immunocompromized hosts, such as AIDS and transplant patients. It can also cause severe congenital infections. Proteases, including metallopeptidases, play major roles in all organisms, catalyzing a broad spectrum of important biological reactions, including protein metabolism, immune reactions, and tissue remodeling for example. It is not surprising, therefore, that proteases have been found in species from viruses to humans. In parasites, besides basic roles in eukaryotic cell biology and physiology, proteases fulfill specific functions linked to the parasitic way of life, facilitating invasion of host tissues or parasite egress, allowing parasites to digest host proteins, helping parasites to evade the host immune response, and preventing blood coagulation among others [16,37,45–46]. As seen with other apicomplexan parasites such as *Plasmodium, Eimeria* and *Cryptosporidium*, toxoplasmic proteases could thus be considered potential therapeutic targets in light of this involvement in host-parasites interactions [16,37,45–46]. Metallopeptidases represent a very diverse catalytic type of peptidase and are classified in the MEROPS database (https://www.ebi.ac.uk/merops/) based on homologous sets of peptidases containing related sequences that are grouped together in families, which are then grouped in clans based on their related primary structures [56]. All known metallopeptidases have been divided into 16 different clans as described in MEROPS: MA (divided in MA(E) also called gluzincins and MA(M) called metzincins subclans), MC, MD, ME, MF, MG, MH, MJ, MM (with a motif like that of clan MA but bound to plasma membranes), MN, MO, MP, MQ, MS, MT, and M- which includes metallopeptidases which are not yet well characterized. Only a few of these clans are represented in *T. gondii*. Studies on *T. gondii* metallopeptidases remain scarce today, whereas complete surveys of protease homologs in *Plasmodium falciparum* [16,66] and *Eimeria tenella* predicted proteomes [31] have been published. To date, seven metallopeptidases have been experimentally explored in *T. gondii* [5,23,25,29–30,34,67,68–69]. The ToxoDB database (http://toxodb.org/toxo/, Release 29), that gathers *T. gondii* genome and post-genome data for numerous strains, provides an invaluable resource to investigate the most complete set of metallopeptidases for this parasite. Using human and protozoan metallopeptidase sequences and peptidase family domains (PFAM motifs [19]) defined in the MEROPS database, we identified, in ToxoDB, 49 putative toxoplasmic metallopeptidases clustered into 15 families corresponding to 7 clans. Expression of these metallopeptidase genes in the tachyzoite stage was then evaluated by PCR and RT-PCR assays.

Materials and methods

T. gondii *metallopeptidase identification,* in silico *analysis and classification*

In this manuscript, we chose to classify metallopeptidases according to their MEROPS classification in families, beyond their amino-, carboxy- or endopeptidase predicted activity. Putative metallopeptidase *Toxoplasma* genes were identified from the *T. gondii* ToxoDB database, (http://toxodb.org/toxo/, Release 29) using peptidase family domain (PFAM motifs) recorded in the MEROPS Database (https://www.ebi.ac.uk/merops/) for this family of enzymes. The TGME49 genome was chosen as a reference in our study. The domain/motif organization of predicted proteases was studied using the Interpro Search (http://www.ebi.ac.uk/interpro/). At the end of this search, a total of 49 genes encoding proteins with metallopeptidase signature motifs were identified in the *Toxoplasma gondii* ME49 genome. They were subsequently assigned to families and sub-families of metallopeptidase annotations by amino acid sequence comparisons using the BLASTp program in the Washington University (http://www.ebi.ac.uk/Tools/blast) and the BLAST MEROPS server using the MEROPS Database.

The deduced amino acid sequences of these putative *Toxoplasma* metallopeptidases proteins were aligned with sequences from other organisms according to the ClustalW multiple sequence alignment algorithm on the EMBL-EBI website (European Bioinformatics Institute, www.ebi.ac.uk/Tools/clustalw2) using the Blosum62 matrix. The prediction of protein localization sites in parasites was performed by using a computer program Psort II (http://www.psort.org/).

Parasites

The RH *T. gondii* strain (genotype I) was used throughout our experiments. Tachyzoites were obtained by inoculation of *T. gondii* in the intraperitoneal cavity of female Swiss mice. The animal housing facility is accredited according to French regulations (approval No. B 51-454-4). The experimental protocol for inoculation was approved by the local Ethics Committee for Animal Experiments (CEEA RCA No. 56) and is referenced under state law under protocol number 56-2012-16.

Design of specific primers for each metallopeptidase sequence

Primers were designed based on the selective sequences of the RH *T. gondii* genomic DNA (gDNA). Positions of introns in putative metallopeptidase genes were obtained by alignment of gDNA with complementary DNA (cDNA). One pair of primers was designed per gene following two conditions if possible: the pair of primers should flank a genomic region spanning an intron and/or amplify the metallopeptidase catalytic

Table 1. Metallopeptidase primers used for the PCR and RT-PCR.Gene: gene nomenclature in ToxoDB release 29.

Peptidase family	[Gene]	Primer Name	Primers	
M1	TGME49_224350	M1	5'-ATCTCGAACTCGAGTCGGTGG-3'	5'-TAGCAAGGCTTCTGGAAGGGG-3'
	TGME49_221310	M1A	5'-CGACCAGGCATTGCTGTTGAG-3'	5'-GTCGTGACCAACCGAGTGAAG-3'
	TGME49_224460	M1B	5'-TTCTGCCCAGGGAAGCAGAAG-3'	5'-GGAACGAAGAGCCCATTGGAG-3'
	TGME49_262575	M1C	5'-AGGAGCGTGGACTACTCACT-3'	5'-TGTCATGAGGATCTGCTGCC-3'
M3	TGME49_226420	M3	5'-TTGGACAACCCCGTTGTGTGG-3'	5'-ACCTCGCCTTCAGCACCAAAG-3'
	TGME49_272670	M3A	5'-TTGCACAGTTCACCGTGCGAG-3'	5'-CACGAAGCTGTAGAAGGCCTG-3'
	TGME49_216150	M3B	5'-CAAGGCGAACGTCCAGACAC-3'	5'-TCGGTCGAATTCGTCGTCCC-3'
M13	TGME49_295640	M13	5'-TCCAGTGGCACTTGGGTCTTG-3	5'-GCTAGTTGAATCCCGCCGTTG-3'
M14	TGME49_265780	M14	5'-CGCATTTTCGTGGTGCCGATG-3'	5'-CCTTGGCGCAGAAGATCGAAG-3'
	TGME49_271870	M14A	5'-CAGTTGCTTCAGACCTACGGG-3'	5'-ACGACTGCGGAGGCGAAAAG-3'
	TGME49_253170	M14B	5'-TACCCTTGGGGCTCGTATGAC-3'	5'-GCCTTCGTGCTCTCTCTGAAC-3'
	TGME49_202910	M14C	5'-CGCCCCAACTCAAAGAAGAGG-3'	5'-GGCAAATGCTTGCGTCTCAGG-3'
M16	TGME49_202680	M16	5'-CCTGGTGTACAGTGCAGAGTG-3'	5'-GTCCACAAATGCGCGCATCAG-3'
	TGME49_253890	M16A	5'-TCGTCCCATCCGTTCGCTTC-3'	5'-AGCGTGACGCTCTCGACAAC-3'
	TGME49_235680	M16B	5'-GCCTCACCTCTTGGAGGTTTG-3'	5'-GTCCTGAAACTCCAGTGCGG-3'
	TGME49_244480	M16C	5'-ATCAAGCAGCACGGAGGCTG-3'	5'-GACGCACAGGCACTTGCTTG-3'
	TGME49_206510	M16D	5'-GCGCGTACTCGGTGTTTTCAG-3'	5'-GGCGCCTTCAACAATGATCGG-3'
	TGME49_257010	M16E	5'-GCTTCCCAGGTGGCATATGAG-3'	5'-CCAGCTCCCCGTCAGATATTG-3'
	TGME49_269890	M16F	5'-GTGCAGAAGTTCCTGGTTCCC-3'	5'-AGCTTGTCCGAAAAGCCTGCC-3
	TGME49_269885	M16G	5'-GTACATCTGGGACAGCCGTCC-3'	5'-GACGACCGTGACGAGGTTTC-3'
	TGME49_236210	M16H	5'-TCGAACGGCAAACCGAAGAGG-3'	5'-GTGTCGCTGTAGCAGGTGTTG-3'
	TGME49_314850	M16I	5'-CAAAGGAGACAGACGACGGTG-3'	5'-GTAGCAGCGGCTTTCTTCGG-3'
	TGME49_214490	M16J	5'-ATCGAGGAGGCGAAAACGCTG-3'	5'-CACTGAACGAGTTCGCGCTG-3'
M17	TGME49_290670	M17	5'-CCCGGGATGTTTCCACGATG-3'	5'-TCCTGGCGCGACGACTTTTG-3'
M18	TGME49_297970	M18	5'-CAACATGCTAGGCACGAGCAG-3'	5'-TTAGGCCGAAGACGGAGACAG-3'
M20	TGME49_213520	M20	5'-GTTCGGGTCACAACAGGATGG-3'	5'-CTTGTCGCCAATTCCCCTGTG-3'
M22	TGME49_274110	M22A	5'-GTCTCTTCGACGCAGGGAAAC-3'	5'-TTACAACCGACTGCATGCGCC-3'
	TGME49_202310	M22B	5'-CGGCAGCACATCATCAAGGG-3'	5'-GAGCAGGATCATTGGGCAGG-3'
M24	TGME49_248850	M24	5'-CGGCGAACTCTTCCATACGAC-3'	5'-TCGACGCCTGTTTCAGTGACC-3'
	TGME49_211330	M24A	5'-TTTCTGTGGACACGGCATCGG-3'	5'-TCGCGATCGTCCATTTGTCGG-3'
	TGME49_257730	M24B	5'-GGGCAGTATCGCCGAGAATAG-3'	5'-CTTAAGAGCGTCCGGTGTCTG-3'
	TGME49_205460	M24C	5'-CTTTCGGAGATTGGCGGCTTC-3'	5'-TTCAAACGCAGTGCGGCGAAC-3'
	TGME49_279390	M24D	5'-CAAGAGGGCTGTAGATCGCAC-3'	5'-GAACTGGGCTACGTATTCGCC-3'
	TGME49_261600	M24E	5'-CCTGATCGTCAACACGGCTTC-3'	5'-CGAGCCAAAGCATTTCGGCAC-3'
	TGME49_233310	M24F	5'-GGAACAGCGCGTGGTGTATG-3'	5'-GAGCTTCGGGGATACTCAGG-3'
	TGME49_221670	M24G	5'-ACGGAATCGTCGCCACGATG-3'	5'-TCCCTTGGCAGTTGTGAGGTG-3'
M28	TGME49_225850	M28	5'-AGAAGACGCTCAAGCGGAGTG-3'	5'-GACATCCATGGCGAGAGAGG-3'
	TGME49_231130	M28A	5'-AACTCGCTCTACGCCAACTC-3'	5'-CGAGTAGTTCCCAGGCTCAC-3'
M41	TGME49_202630	M41A	5'-CGCCAGTCAGACTAAGGCTC-3'	5'-AGAGGTAGACGATCGCTCGAC-3'
	TGME49_200020	M41B	5'-ACTTTCTCCCAGTCGGAGCAG-3'	5'-GTCTCGTTTCTCCTCGAGCAG-3'
	TGME49_259260	M41C	5'-CCATCCGGATCATACCGAACC-3'	5'-CATGTGGCAAACCCGCTTCTC-3'
M48	TGME49_221170	M48	5'-TTCCTCGGGCCTTTCCTGTG-3'	5'-CGCGGAGTTCTTCGTCCTTG-3'
M50	TGME49_266140	M50	5'-ATGGACGACCCGACTTTGCAG-3'	5'-AAAACTCCGAGGACGAGACGG-3'
	TGME49_285670	M50A	5'-AAAGAGGGCAGTGGATCACC-3'	5'-CGCGCAGAGTCCAGATGTTA-3'
M67	TGME49_251500	M67A	5'-GTCAAAACCGGGTCTCTTGCC-3'	5'-ATCCGCTCTTCCAGGTCCTTC-3'
	TGME49_231970	M67B	5'-AACCCAGCGCTGTACGTGTTG-3'	5'-CTCGACCTTCATGATCGCCTG-3'
	TGME49_228190	M67C	5'-TTTCCCCTTCGCGCTTTCGTC-3'	5'-TTTCGCTCCCAGTGCAGAACC-3'
	TGME49_269250	M67D	5'-CCCTTTCCTCGTTTGCCTCTG-3'	5'-AAGTCCGTCATGCGGACCTG-3'
	TGME49_269840	M67E	5'-CTGTAGCGTCGCGTTTGTGTG-3'	5'-GAAAAGACGTCGACGACTCGG-3'
	TGME49_208590	M67F	5'-AAGCTGCAACGTCTGTCTCCC-3'	5'-CCAAAGTCCGAGGCCTTTTCC-3'
M76	TGME49_257110	M76	5'-AAACGCGTGAGGAGAGATGCC-3'	5'-AGATGGTCGCCACTTGCTGTC-3'

domain. All primer pairs were designed using Primer Pro 3.4 software (www.changbioscience.com/primo/primo.html). The primers used to assess metallopeptidase gene expression are listed in Table 1 with their corresponding gene.

Polymerase chain reaction (PCR) of metallopeptidases sequences

Genomic DNA was extracted from purified RH *T. gondii* tachyzoites using the QIAamp® DNA Mini Kit

(Qiagen, Courtaboeuf, France), following the manufacturer's instructions. Amplifications were performed using 1 µL of cDNA or 3 µL of gDNA (10 ng/µL), 50 pmol of each primer and 2U of Taq DNA polymerase (Invitrogen™ Life Technologies) in 50 µL PCR reaction containing 1×PCR Buffer (20 mM Tris–HCl (pH 8.4), 50 mM KCl, 1.5 mM $MgCl_2$), and 200 µM of dATP, dTTP, dGTP, dCTP. The template was subjected to 35 cycles (94 °C for 30 s, 60 °C for 45 s and 72 °C for 2 min) followed by a final 10 min extension at 72 °C. PCR products were analyzed by electrophoresis in 1× TBE buffer on 2% agarose gel stained with 0.5 µg/mL ethidium bromide and photographed under ultraviolet light (Phospho imager, Biorad). SAG1 tachyzoite transcript was used as an RT-PCR positive control (SAG1S 5′-caatgtgcacctgtaggaagc-3′, SAG1R 5′-tgggcaggtgacaacttgatt-3′). A negative control containing all reagents, except cDNA, was also included. The presence of gDNA contamination in cDNA samples was verified by PCR using primer pairs which amplify intron(s)-containing regions.

Reverse transcription-PCR

Total RNA was isolated from tachyzoites using the RNeasy® Mini Kit (Qiagen). Prior to reverse transcriptase (RT)-PCR analysis, total RNA was treated at room temperature for 15 min with RNase-free DNase I (Invitrogen™ Life Technologies). Total RNA samples of 1 µg, denatured at 65 °C for 10 min, were reverse transcribed at 42 °C for 50 min in a total volume of 20 µL using oligo-(dT) 18 as the primer with 200U Superscript™ II reverse transcriptase (Invitrogen™ Life Technologies). Following heat inactivation at 70 °C for 15 min, the reverse transcribed mRNA (cDNA) mixture was incubated with 2U of *Escherichia coli* RNase H at 37 °C for 20 min to remove complementary RNA to the cDNA. A negative control containing all reagents, except total RNA, was also included for each experiment.

Results and discussion

The *T. gondii* reference genome database ToxoDB was screened to identify putative metallopeptidase sequences. In all, 49 metallopeptidases containing PFAM domains that characterize the metallopeptidase enzyme superfamily were identified. The genome of the RH strain (genotype I) shares high similarity to the archived genomes of the *T. gondii* GT1 (genotype I), ME49 (genotype II) and VEG strains (genotype III) [35]. We then decided to investigate metalloprotease expression in total RNA from RH tachyzoites by conventional RT-PCR. To do so, gene-specific primer pairs flanking a region spanning intron(s) were designed in order to amplify fragments of distinct length from cDNA and gDNA templates. Amplifications with these different primers pairs yielded PCR and RT-PCR products of expected sizes for each gene (Figure 1). These results confirmed their transcription and are in agreement with the currently proposed intron-exons gene model boundaries in ToxoDB. In view of the high structural diversity seen in metallopeptidase families, putative metallopeptidases from the *T. gondii* genome database were assigned according to the MEROPS classification, as described below.

In this study, we used the MEROPS Nomenclature system (release 10.0) as described in Rawlings et al. (2016) [56]. In this system, proteases are classified into 8 catalytic superfamilies (aspartic, cysteine, glutamic, metallo-, mixed catalytic type, serine, threonine, and unknown catalytic type peptidases), and metallopeptidases into 16 clans based on their related structures. Metallopeptidases from the *T. gondii* genome database (ToxoDB) were therefore classified based on their domain organization and sequence similarities to metallopeptidases from other organisms. We have found 49 putative metallopeptidases as shown in Table 2 (see also Appendix 1), which details protease MEROPS clans and families, number of metal ions, *T. gondii* ME49 gene ID, chromosomal location, protein length (amino acids), ToxoDB product description, protease homolog with highest BLAST score using BLASTp program combined with MEROPS BLAST server, primer name correspondence, alias, related publication and PFAM ID, and signal peptide presence/prediction. A total of 49 putative metallopeptidases have thus been identified and ascribed to 15 metallopeptidases families described in the MEROPS database: four M1, three M3, one M13, four M14, eleven M16 (the most represented family), one M17, one M18, one M20, eight M24, two M28, three M41, one M48, two M50, six M67 and one M76 peptidase families.

Hereafter, we describe our results concerning each of these 15 families found in *T. gondii* genome.

M1 Peptidase family (Aminopeptidase N family)

M1 family peptidases, also called membrane alanyl aminopeptidase (aminopeptidase N), are dependent on a single zinc ion for activity, and catalyze amino acid cleavage from amino-termini of protein or polypeptide substrates. These aminopeptidases are involved in several biochemical processes, including protein maturation and activation. This M1 family of metallopeptidase enzymes (clan MA(E)) presents 2 key signatures: H**E**xxH$(x)_{18}$E, the active site motif in which the 2 histidines and the last glutamic acid (underlined) bind zinc atom and the first glutamic acid (bold) is involved in catalysis, and an upstream GAMEN motif involved in substrate recognition.

Four *T. gondii* peptidases from ToxoDB can be ascribed to this M1 peptidase family (TGME49_221310 (*Tg110*), TGME49_224460, TGME49_224350, and TGME49_262575) but only three display both typical HExxH$(x)_{18}$E and GAMEN signatures (Figure 2). TGME49_262575 is highly atypical and very small (290 amino acid). It may be incomplete in ToxoDB but is, however, conserved among coccidia which is intriguing and deserves further investigation. Interproscan analysis indicates a leukotriene A4-hydrolase domain (Superfamily domain SSF63737), classified within the M1 peptidase

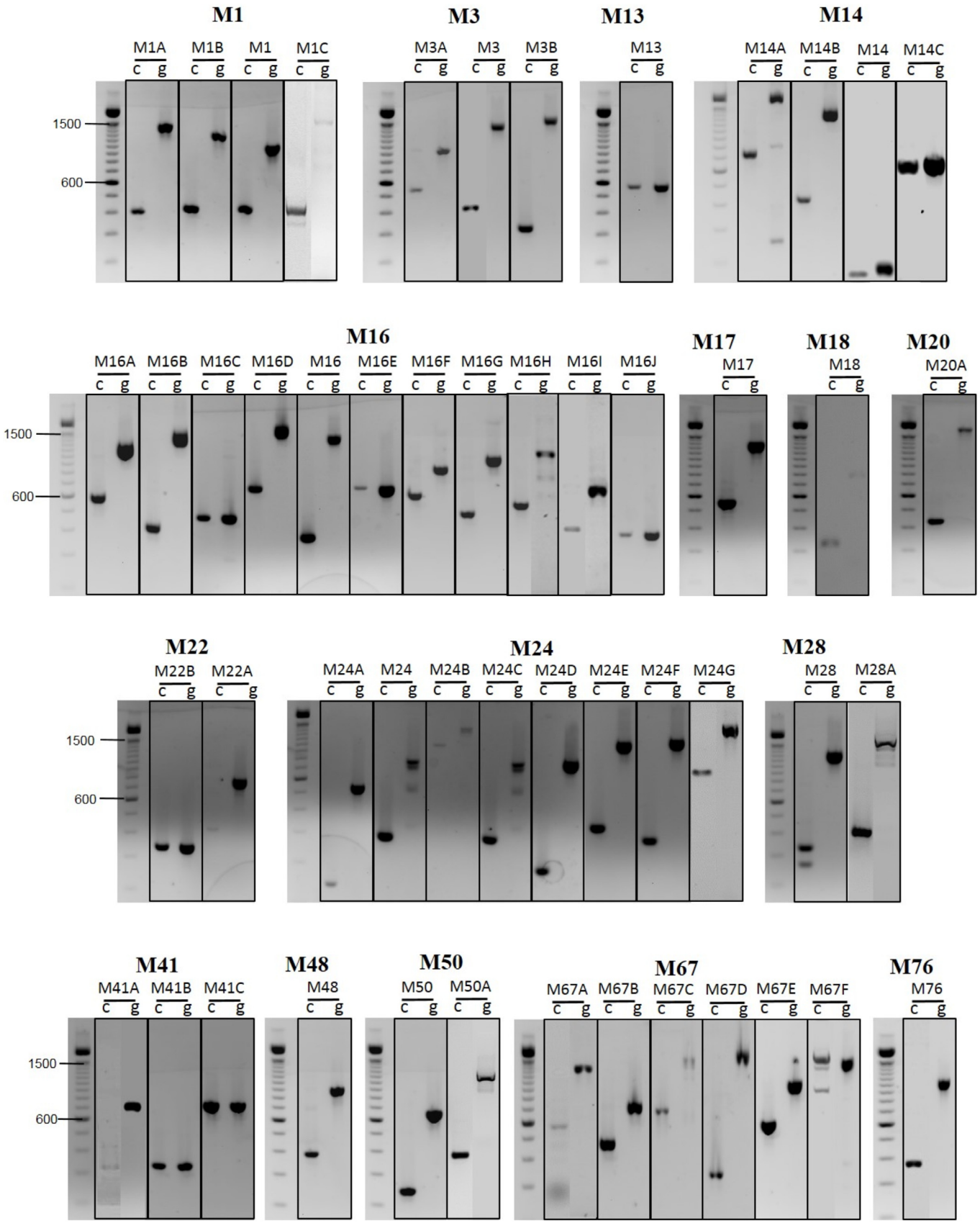

Figure 1. Metallopeptidase gene expression in extracellular toxoplasmic tachyzoites by RT-PCR.
Products of the expected size were observed for all primers, using either cDNA and gDNA as templates. As a further control for the presence of contaminating gDNA, primers of each gene were designed to amplify fragments of distinct length from cDNA(c) and gDNA (g) due to the presence of introns. Molecular size standards are indicated to the left.

Table 2. Metallopeptidase genes identified and classified in the *T. gondii* genome database (strain ME-49, genotype II). We used Pfam motifs (http://pfam.xfam.org/) in association with the MEROPS Database to screen the *T. gondii* database (http://toxodb.org/toxo/, Release 29). The motif organization of predicted peptidases was studied using the InterProScan Search (http://www.ebi.ac.uk/interpro/) and family assignment is based on MEROPS – the peptidase Database – classification (https://www.ebi.ac.uk/merops/).

MEROPS Clan Number of metal ion	MEROPS families	Gene ID ME-49 strain	Chro.	Protein Length (aa)	ToxoDB Product Description (ME-49)	Blast MEROPS %identities/e-score (Families)	Primer Name	Alias	Publication	PFAM	SignalP peptid
MA(E) 1 Zn^{2+}	M1	TGME49_224350	X	1419	aminopeptidase N, putative	92,18%/3,40e-213 (M1)	M1	-	-	PF01433	No
		TGME49_221310	II	1069	aminopeptidase N protein	57,30%/3,6e-123 (M1)	M1A				No
		TGME49_224460	X	970	aminopeptidase N, putative	100%/2,6e-238 (M1)	M1B	Tg110	Berthonneau J. 2000		Yes
		TGME49_262575	VIIb	290	hypothetical protein	39,33%/ 4,7e-10 (M1)	M1C	-	-		No
MA(E) 1 Zn^{2+}	M3	TGME49_226420	X	667	peptidase family M3 protein	99,29%/9,70e-152 (M3B)	M3	-	-	PF01432 PF08439	Yes
		TGME49_272670	VIII	1094	peptidase family M3 protein	81,91%/2,60e-270 (M3A)	M3A	-	-		Yes
		TGME49_216150	XI	506	peptidase family M3 protein	95,47%/2,50e-144 (M3B)	M3B	-	-		No
MA(E) 1 Zn^{2+}	M13	TGME49_295640	Ia	1038	peptidase family M13 protein	98,44%/0 (M13)	M13	-	-	PF01431	Yes
MC 1 Zn^{2+}	M14	TGME49_265780	IX	2803	flagellar/basal body protein	100%/7,50e-55 (M14D)	M14	-	-	PF00246	Yes
		TGME49_271870	VIII	1321	zinc carboxypeptidase superfamily protein	44,44%/2,30e-51 (M14A)	M14A	-	-		Yes
		TGME49_253170	III	2204	zinc carboxypeptidase, putative	47,71%/5,20e-55 (M14B)	M14B	-	-		No
		TGME49_202910	VIIa	318	zinc carboxypeptidase superfamily protein	49,62%/1,30e-67 (M14B)	M14C	-	-		Yes
ME 1 Zn^{2+}	M16	TGME49_202680	VIIa	563	peptidase M16 alpha subunit, putative	100%/3,10e-107 (M16B)	M16	MPPA	-	PF00675 PF05193 PF16187 PF08367	Yes
		TGME49_253890	III	1604	peptidase M16 inactive domain-containing protein	100%/3,8e-105 (M16B)	M16A	-	-		Yes
		TGME49_235680	X	1692	peptidase M16 inactive domain-containing protein	76,28%/2,3e-95 (M16B)	M16B	-	-		No
		TGME49_244480	VI	357	peptidase M16 inactive domain-containing protein	75,00%/5,00e-80 (M16A)	M16C	-	-		No
		TGME49_206510	VIIa	2340	Toxolysine-4	81,50%/2,70e-81 (M16A)	M16D	TLN4	Laliberté J. 2011		No
		TGME49_257010	VIIb	1023	sporozoite developmental protein	100%/1,10e-109 (M16A)	M16E	-	-		Yes
		TGME49_269885	VIII	1645	rhoptry metalloprotease toxolysin TLN1 (TLN1)	93,50%/3,50e-99 (M16A)	M16F	TLN1	Hajagos BE. 2012		Yes
		TGME49_227948	X	1306	peptidase M16 inactive domain-containing protein	90,95%/7,50e-212 (M16C)	M16G	-	-		Yes
		TGME49_236210	X	509	peptidase M16 family potein, putative	66,01%/7,00e-145 (M16B)	M16H	-	-		Yes
		TGME49_314850	XI	2174	hypothtical protein	45,59%/2,70e-06 (M16A)	M16I	-	-		No
		TGME49_214490	X	1353	peptidase M16 inactive domain-containing protein	83,77%/7,50e-235 (M16C)	M16J	-	-		No
MF 2 Zn^{2+}/Mn^{2+}	M17	TGME49_290670	IX	781	leucyl aminopeptidase LAP (LAP)	94,13%/7,00e-184 (M17)	M17	TgLAP	Jia H. 2010	PF00883 PF02789	yes
MH 2 Zn^{2+}	M18	TGME49_297970	II	508	aspartyl aminopeptidase	97,86%/1,70e-246 (M18)	M18	TgAAP	Zheng J. 2016	PF02127	No
MH 2 Zn^{2+}	M20	TGME49_213520	V	514	peptidase M20D, amidohydrolase	96,99%/5,80e-205 (M20D)	M20	-	-	PF07687 PF01546	yes
MG 2 Co^{2+}/Mn^{2+}	M24	TGME49_248850	XII	416	methionine aminopeptidase	100%/8,20e-145 (M24A)	M24	-	-	PF00557 PF16188	No
		TGME49_211330	IV	697	methionine aminopeptidase	95,93%/1,50e-129 (M24A)	M24A	-	-		No
		TGME49_257730	VIIb	484	methionine aminopeptidase, type i, putative	54,30%/2,10e-66 (M24A)	M24B	-	-		yes
		TGME49_305460	IX	480	methionine aminopeptidase 2, putative	61,21%/1,50e-106 (M24A)	M24C	-	-		No
		TGME49_279390	IX	462	proliferation-associated protein 2G4, putative	51,22%/3,00e-83 (M24X)	M24D	-	-		No
		TGME49_261600	VIIb	886	creatinase domain containing protein	35,00%/1,20e-97(M24B)	M24E	TgAPP	Yang M. 2016		No
		TGME49_233310	VIII	599	peptidase D, putative	100%/4,20e-125 (M23B)	M24F	-	-		No
		TGME49_221670	II	1198	transcriptional elongation factor FACT140	39,62%/9,00e-49 (M24X)	M24G				No
MH 2 Zn^{2+}	M28	TGME49_225850	X	1555	peptidase, M28 family protein	42,55%/3,10e-28 (M28X)	M28	-	-	PF04389	No
		TGME49_231130	VIII	711	hypothetical protein	36,60%/7,30e-25 (M28X)	M28A	-	-		No
MA(E) 1 Zn^{2+}	M41	TGME49_202630	VIIa	1188	ATP-dependent metallopeptidase HflB subfamily protein	99,59%/2,00e-125 (M41)	M41A	-	-	PF01434	Yes
		TGME49_300020	XII	1021	ATP-dependent metallopeptidase HflB subfamily protein	100%/3,10e-113 (M41)	M41B	-	-		Yes
		TGME49_259260	VIIb	1250	membrane protein FtsH1	100%/6,50e-120 (M41)	M41C	FtsH1	Karnataki A. 2007		No
MA(E) 1 Zn^{2+}	M48	TGME49_221170	II	432	CAAX metallo endopeptidase	94,51%/1,20e-141 (M48A)	M48	-	-	PF01435 PF16491	No
MM 1 Zn^{2+}	M50	TGME49_266140	IX	378	peptidase, M50 family protein	34,27%/1,40e-23 (M50B)	M50	-	-	PF02163	No
		TGME49_285670	V	260	hypothtical protein	70,00%/6,50e-00 (M50B)	M50A	-	-	PF13398	No
MP 1 Zn^{2+}	M67	TGME49_251500	XII	600	eukaryotic translation initiation factor 3 subunit 3, putative	29,17%/1,60e-32 (M67X)	M67A	-	-	PF01398 PF05021 PF14464	Yes
		TGME49_231970	VIII	2538	pre-mRNA splicing factor PRP8, putative	27,37%/1,50e-01 (M67C)	M67B	-	-		No
		TGME49_228190	X	346	eukaryotic translation initiation factor 3 subunit 5, putative	66,67%/3,30e-45 (M67A)	M67C	-	-		No
		TGME49_269250	VIII	343	Mov34/MPN/PAD-1 family protein	57,00%/2,9e-39 (M67A)	M67D	-	-		No
		TGME49_269840	VIII	314	proteasome regulatory subunit	100%/1,70e-167 (M67A)	M67E	-	-		No
		TGME49_208590	XI	489	Mov34/MPN/PAD-1 family protein	94,05%/2,50e-134 (M67A)	M67F	-	-		Yes
MA(E) 1 Zn^{2+}	M76	TGME49_257110	VIIb	618	hypothetical protein	100%/3,50e-50 (M76)	M76	-	-	PF09768	No

family. A *T. gondii* aminopeptidase, named *Tg110*, and able to cleave L-Arg-AMC, L-Leu-AMC, and L-Tyr-AMC (aminopeptidase substrates) was described experimentally by Berthonneau in 2000 [5]. *Tg110* was identified in cell-free extracts and was purified using high-performance liquid chromatography. Its optimal activity was at pH 7.4 and it was strongly inhibited by classical metallopeptidase inhibitors (EDTA and o-phenanthroline). The purified enzyme exhibited a pI of 4.7 and had an apparent molecular weight of 110 kDa. These features are in agreement with theoretical values for TGME49_224460 (Table 2, see also Appendix 1). Interestingly, *Tg110* was detected in human sera from patients undergoing toxoplasmosis, suggesting involvement in infection response [5]. As of today, no function has been ascribed for any M1 peptidase family from *T. gondii*.

The importance of M1 family aminopeptidases has been recognized in closely related protozoan species including PF3D7_1311800 (PfA-M1) from *P. falciparum* [1,3,7,14–15,20,24,43,62], NCLIV_048240 (NcAPN1) and NCLIV_048230 (NcAPN3) from *N. caninum* [22], cgd8_3430 from *C. parvum* (strain Iowa II) [49], and

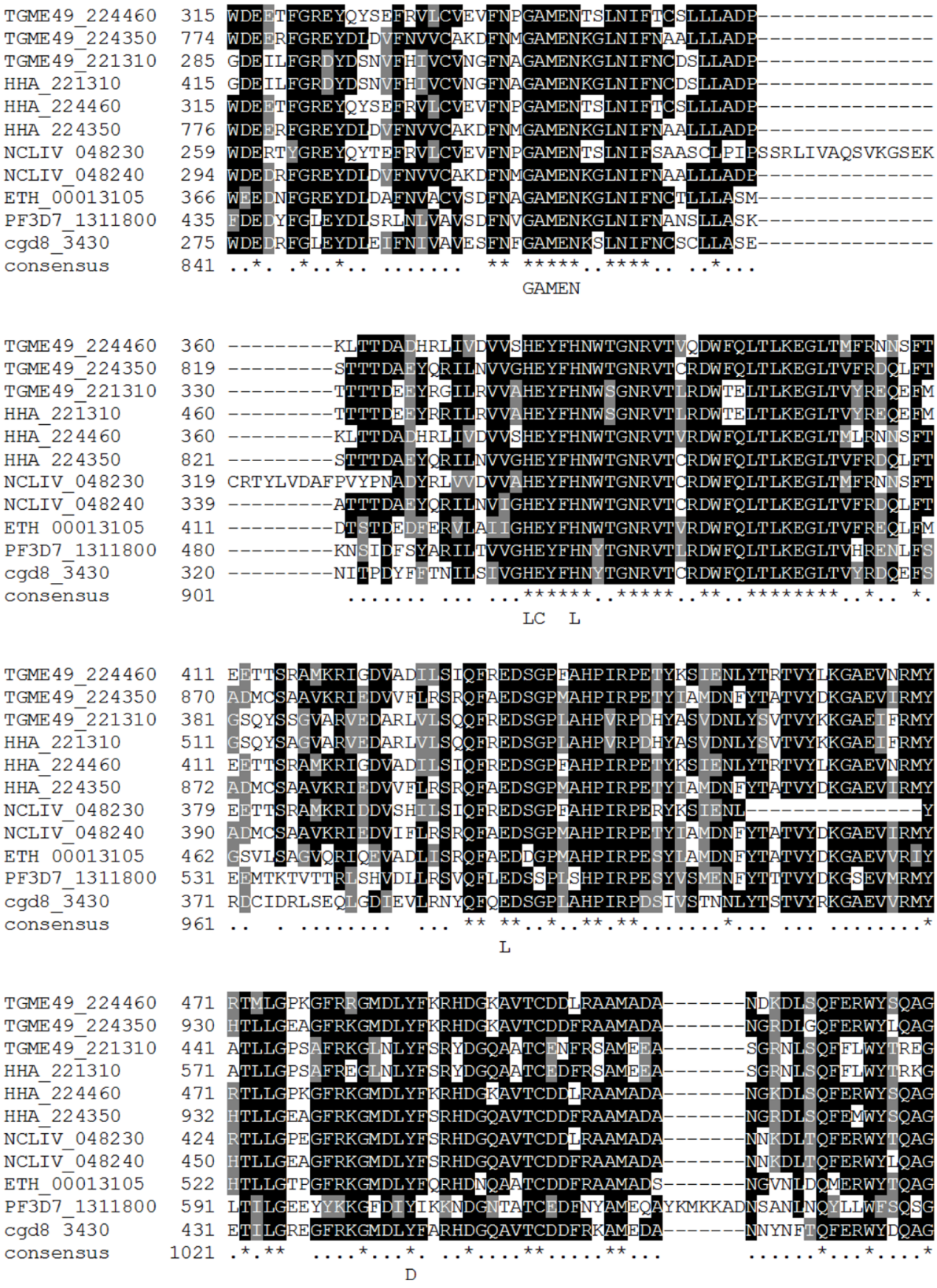

Figure 2. Multiple sequences alignment from *T. gondii* aminopeptidase N (M1 peptidase family) and several selected members of the M1 family of zinc-metallopeptidases: *P. falciparum* (PF3D7_1311800 and PF3D7_1472400), *T. gondii* (TGME49_221310, TGME49_224350, and TGME49_224460), *N caninum* (NCLIV_048240 and NCLIV_048230), and *C. parvum* (cgd8_3430). Amino acid positions identical between these sequences and the *T. gondii* sequence are in darkened letters. Identical (black background) and conserved (grey background) amino acids between all sequences are indicated. The position of the conserved putative zinc ion ligands (L), the conserved glutamyl residue required for catalytic activity (C), and the conserved putative proton donor (D) are indicated in bold on the bottom line. The amino acid numbers for each sequence are indicated on the left. The position of gaps is indicated by full colons. Alignments were performed using the ClustalW2 algorithm (www.ebi.ac.uk/Tools/clustalw2) with the Blosum 62 matrix.

ETH_00013105 and ETH_00013105 from *E. tenella* (strain Houghton) also called EtAPN1 [22] and EtAPN2 [31]. The role of PfA-M1 is largely documented. PfA-M1 is found in various locations in the malaria parasite, such as the cytoplasm, food vacuole, parasitophorous vacuole and nucleus [1,3,14,43]. This M1 aminopeptidase has been mainly involved in parasite metabolism in the last steps of hemoglobin degradation [54] but also in parasite development [48]. EtAPN1 is an active protease during *Eimeria* parasite sporulation [18]. Using bestatin, a well-known broad-spectrum inhibitor of metalloaminopeptidases, on *E. tenella* infected culture *in vitro*, a strong inhibition of parasite development but not of the invasion process was observed [22].

M3 Peptidase family (thimet oligopeptidase and oligopeptidase F families)

M3 peptidases, also belonging to the MA(E) clan, display a highly conserved signature FH**E**xGH$(x)_2$H$(x)_{12}$G$(x)_5$D$(x)_2$ExPS$(x)_3$E, including the H**E**xxH motif, in which **E** (bold) is involved in catalysis and the two underlined H and a C-terminally located E residue act as zinc-binding ligands [56]. This type of endopeptidase only hydrolyzes oligopeptides that contain no more than 20 amino acid residues. The M3 peptidase family is involved in peptide degradation, bioactive neural-peptide synthesis, and cleavage of signal peptides. Most of the M3 peptidases are synthetized without signal peptides, except Mitochondrial Intermediate Peptidase (MIP) which possess a typical amino-terminal mitochondrial leader peptide recognized and cleaved by the mitochondrial processing peptidase. The main role of M3 peptidases is to cleave short peptidic substrates in the cytoplasm, whereas MIP resides in the mitochondrial intermembrane space and cleaves N-terminal octapeptides from proteins during their import into mitochondria. The M3 family is divided into three sub-families: M3A also called thimet oligopeptidases (including neurolysin and MIP), M3B called oligopeptidases F and M3C called Pz-peptidase A [56].

During *T. gondii* asexual development, an oligopeptidase F has been identified by the use of microarrays, and may be involved in the regulation of bradyzoite-specific metabolic pathways, as found in bacteria [13].

Three M3 family peptidases were found in ToxoDB (TGME49_272670, TGME49_226420 and TGME49_216150). By sequence homology, TGME49_272670 would belong to the M3A peptidase sub-family, whereas the two others could belong to the M3B oligoendopeptidase F subfamily (Table 2, see also Appendix 1). These enzymes are predicted to be localized in matrix mitochondria for TGME49_272670 and in the parasite cytoplasm for TGME49_226420 and TGME49_216150 (PSORT II prediction). As of today, no enzyme from the M3 peptidase family has been further experimentally described for any apicomplexan.

M13 Peptidase family (Neprilysin family)

Also belonging to the MA(E) clan, the M13 family (also named neprilysin family) is a large group of zinc-metallopeptidases which present highly conserved sequences, including the HExxH motif and a C-terminally located E residue, in which the underlined amino-acids provide the three zinc ligands, and the catalytically important GENIAD and VNAFY motifs [6]. M13 peptidases are endopeptidases which are responsible for the inactivation and/or activation of peptide signaling events on cell surfaces. Current knowledge suggests that all peptidases in family M13 are restricted to acting on substrates of no more than about 40 residues [56]. These enzymes appear to be synthetized in active forms, without proenzyme forms. The majority of currently described M13 endopeptidases are type II integral transmembrane zinc-metallopeptidases. Homologs are known from all kingdoms of life, but principally so far from bacteria and animals.

As of today, no enzyme from the M13 peptidase family has been described in *T. gondii*, nor in other apicomplexa. In our study, one *T. gondii* M13 peptidase was found in ToxoDB (TGME49_295640) that is predicted to be localized in the mitochondrial matrix space according to PSORT II prediction.

M14 Peptidase family (carboxypeptidase A1, carboxypeptidase E, gamma-D-glutamyl–meso-diaminopimelate peptidase I and cytosolic carboxypeptidase 6 families)

Clan MC contains metallocarboxypeptidases of the M14 family. Within the M14 family, sequence conservation around the zinc ligands and catalytic residues allowed to distinguish four sub-families: M14A, M14B, M14C, and M14D. Most of the carboxypeptidases are synthetized without signal peptides, but with N-terminal propeptides that must be processed to release active enzymes. These carboxypeptidases hydrolyze single C-terminal amino acids from polypeptide chains.

Currently, four *T. gondii* M14 family peptidases have been found in ToxoDB that are thus predicted to display carboxypeptidase functions. Among them, two are indeed characterized by an EC number. TGME49_253170 is characterized as EC 3.4.17.12 (carboxypeptidase M) that is predicted to cleave the amino acids arginine or lysine at the C-terminal of peptidic substrates. In contrast, the TGME49_202910 carboxypeptidase is characterized as EC 3.4.17.1 (carboxypeptidase A), which is predicted to cleave all the other amino acids located at the C-terminal of peptidic substrates except arginine, lysine and proline [56].

M16 Peptidase family (pitrilysin, mitochondrial processing peptidase beta-subunit and eupitrilysin family)

Clan ME includes the M16 peptidase family in which two of the three zinc ligands are present in the motif HxxEH. The complete M16 peptidase family catalytic site

signature is HxxEH$_{74}$E in which the two underlined histidines and the last underlined glutamate are zinc binders and the first glutamate (bold) is involved in the catalytic reaction. This family consists of three subfamilies named M16A, M16B, and M16C, in which the differences lie in the precise architecture of the catalytic sites. Members of the M16A and M16C families are composed of four domains in which only one possesses a zinc binding site. However, the members of the M16B family are heterodimers composed of two identical subunits each of which possesses a zinc binding site. Within the M16B family, MPP peptidases (Mitochondrial Processing Protease) are the most represented enzymes. As their name suggests, they are involved in proteolytic processing in mitochondria. They act with the IMP (Inner Membrane Peptidase) and MIP (Mitochondrial Intermediate Peptidase) to allow protein targeting in the different mitochondrial sub-compartments [21].

In this *in silico* study, 11 proteases were identified, characterized by the HxxEH motif, also called "reverse catalytic signature". Peptidases of the M16A family have been found in different parasites and particularly in *T. gondii*, where they are located in the rhoptries [9].

Two metallopeptidases have been described in *T. gondii* as belonging to the M16A family: toxolysin-1 (TGME49_269885) and toxolysin-4 (TGME49_206510). Toxolysin-1 is a zinc metalloprotease secreted from rhoptries [9]. It presents a pro-domain in its N-terminal region responsible for its targeting to this organelle. By constructing mutants of the gene encoding this protease, Hajagos et al. showed that this protease is not essential for parasite *in vitro* growth nor *in vivo* virulence [23]. Toxolysin-4, stored in micronemes, is released in response to an increase in Ca^{2+} level and could play a role during invasion [34]. This protease appears, in addition, to undergo a complex maturation process as six forms of this protease have been identified ranging from 260 kDa (precursor) to 34 kDa (degradation metabolite) [34].

No protease belonging to the M16B or M16C family has been described to date in *T. gondii*. Two M16C peptidases of *P. falciparum* are particularly well described: falcilysin (PF3D7_1360800) [17,47,52], and PfSPP [63]. Falcilysin has several functions, which is illustrated by at least two different EC classifications in EuPathDB EC.3.4.24.- (metalloendopeptidases) and EC.4.4.1.21 (S-ribosylhomocysteine lyase). This protease is present in the food vacuole, where it appears involved in hemoglobin catabolism [17,47], but additional isoforms generated by alternative splicing are also targeted to the *P. falciparum* apicoplast and mitochondrion, as described by Ralph et al. [55]. Regarding their different destinations, falcilysin may thus be present in three parasitic compartments: the digestive vacuole, the apicoplast by signal peptide cleavage, and the mitochondria by more complex splicing.

M17 Peptidase family (leucyl aminopeptidase family)

The MF clan consists of aminopeptidases that need two cocatalytic metal ions (that could be Zn^{2+} and/or Mn^{2+}) for activity. M17 is the only family represented in this clan; it is composed of leucine aminopeptidases (LAPs) [10]. These metalloexopeptidases catalyze the sequential removal of amino acids from the N-termini of proteins and peptidic substrates [56]. LAPs present two characteristic patterns: VGKG, corresponding to conserved amino acid regions, and NTDAEGRL, important for the active site [41].

In this *in silico* study, one LAP was found in the ToxoDB, referenced as TGME49_290670 and previously described by Jia et al. in 2010 [25]. This exopeptidase is localized in the cytoplasm of parasites and appears to be involved in free amino acid pool regulation. In 2015, Zheng et al. demonstrated that a *T. gondii* leucine aminopeptidase gene knockout influenced the growth of *T. gondii* without completely blocking parasite development, virulence or enzymatic activity [68]. We have found in the ToxoDB database that this LAP has an ortholog (NCLIV_042660) in the *N. caninum* genome, an expected situation considering the phylogenetic proximity of the two species.

Interestingly, we can also note that one M17 was identified in the *P. falciparum* genome as PF3D7_1446200, and this has been studied extensively [42,44,57,59]. This protease is expressed in all intra-erythrocytic stages and particularly at the trophozoite stage where protein synthesis increases [8,38]. It appears involved in the regulation of free amino acid spool [59]. Bestatin, a broad spectrum aminopeptidase inhibitor, prevents the growth of *P. falciparum* parasites *in vitro* and PNAP, a PfA-M17 specific inhibitor, blocks the malaria parasite development at ring stage, suggesting that this enzyme could play additional roles in the early erythrocytic development of the parasite [24]. Another apicomplexan LAP has also been characterized in *C. parvum* (CpLAP) that may also play an important role in free amino acid pool regulation [27]. Interestingly, TgA-M17 has a signal peptide contrary to PfA-M17 and TgA-M17 has wider substrate specificity than PfA-M17. While in the malaria parasite PfA-M17 is mainly described as a hemoglobinase, it could fulfill other roles [24]. Of note, TgA-M17 is currently related to glutathione metabolism (Kyoto Encyclopedia of Genes and Genomes KEGG metabolism) [26].

M18 Peptidase family (Aminopeptidase I family)

M18 is part of the MH clan, and contains metallopeptidases that require two cocatalytic metal ions Zn^{2+}. This family consists of aspartyl aminopeptidases (AAP), forming dodecameric complexes in humans, and exclusively cleaving aspartic or glutamic amino acids located at the N-termini of proteins and peptide chain [56]. As few AAP have been described in the literature, there is limited data on their enzymatic activities.

One *T. gondii* M18 has been identified in ToxoDB: TGME49_297970. This M18 peptidase, also called *Tg*AAP, is localized in the cytoplasm of the parasite and appears involved in parasite replication and growth [69].

In *P. falciparum*, Teuscher et al. [61] described *Pf*M18AAP octomers in the cytosol of parasites (synthesized during erythrocytic stages). *Pf*M18AAP (PF3D7_0932300) is exported and appears to act in synergy with other malarial aminopeptidases in order to achieve degradation of proteins such as hemoglobin. Antisense-mediated inhibition of PfM18AAP resulted in a lethal phenotype [61]. However, the involvement of this metallopeptidase in parasite survival remains controversial since Dalal and Klemba (2007) [14] were able to delete the gene without finding any deleterious effects, concluding that the protein function was not essential. This metallopeptidase is, however, considered a potential therapeutic target [58]. *Pf*M18AAP has also been shown to bind *in vitro* to human erythrocyte spectrin (spectrin binding region of 33 amino acids only present in *P. falciparum*), showing multiple enzymatic functions in the parasite and the erythrocytic host [36].

Recently, screening of inhibitors against malarial M1, M17 and M18 families have been tested using inhibitors present in the "Malaria Box", allowing the identification of two potential inhibitors: MMV020750 and MMV666023 of PfA-M1 and PfA-M17, respectively [50].

M20 Peptidase family (Glutamate carboxypeptidase, peptidase T, Xaa-His dipeptidase, carboxypeptidase Ss1 families)

Clan MH also presents the M20 family. This family encompasses glutamate carboxypeptidases and is characterized by the presence of two cocatalytic zinc metal ions, like in the M17 family [56]. Family M20 is currently divided into four separate sub-families: M20A, M20B, M20C and M20D.

One M20 peptidase was found in the *T. gondii* genome: TGME49_213520, but no experimental data concerning this or another *T. gondii* M20 peptidase has been described in the literature.

M24 Peptidase family (Methionyl aminopeptidase 1 and aminopeptidase P families)

Clan MG contains exopeptidases that required two cocatalytic ions of cobalt and/or manganese, and contains only family M24. Peptidases belonging to this M24 peptidase family are also called methionine aminotransferase and cleave methionine residues at the N-terminal level. The M24 family has been divided into two subfamilies: M24A (methionyl aminopeptidase) and M24B (X-pro aminopeptidase and X-pro dipeptidase). Most members of family M24 are cytosolic, and do not require proteolytic activation.

As of today, eight *T. gondii* methionine aminopeptidases have been found in ToxoDB with signature PFAM PF00557. Only one publication describes a *Toxoplasma* M24 peptidase (ToxoDB accession number: TGME49_261600) also called *Tg*APP (aminopeptidase P) and a recombinant form of this *Tg*APP has been expressed to evaluate its enzymatic parameters [67]. Deletion of the *Tg*APP gene in the parasite, through a CRISPR/Cas9 system, resulted in growth inhibition, thus indicating the importance of *Tg*APP as a potential therapeutic target [68].

Four M24 peptidases [62] have been published in *P. falciparum*, named *Pf*MetAP1a, *Pf*MetAP1b, *Pf*MetAP1c [11] and *Pf*MetAP2 [12], but only *Pf*MetAP1b was cloned, overexpressed, purified, and used to screen a compound library for inhibitors [11]. Interestingly, M24 peptidases have different localizations in *P. falciparum* parasites: *Pf*MetAP1a is present in mitochondria, *Pf*MetAP1b is present in cytosol and, *Pf*MetAP1c and *Pf*MetAP2 are in the apicoplast [16].

Other apicomplexa like *Cryptosporidium parvum*, *Eimeria tenella* or *Neospora caninum* encode for 5, 7 and 8 M24 metallopeptidases in their genomes, respectively (see also Table 3).

Besides the already mentioned and mostly studied malarial aminopeptidase PfA-M1 and PfA-M17 [28,53], these M24 malarial aminopeptidases also constitute very promising potential new targets for antimalarial drug development [62].

M28 Peptidase family (aminopeptidase S, glutamate carboxypeptidase II, IAP aminopeptidase and aminopeptidase Ap1 families)

This family, included into clan MH, is composed of aminopeptidases and carboxypeptidases featuring two cocatalytic zinc ions [56].

Only two *T. gondii* M28 peptidases have been found in ToxoDB: TGME49_225850 and TGME49_231130. At the present time, no protein of this family has been experimentally described in *T. gondii* nor in other apicomplexa in the literature.

M41 Peptidase family (FtsH endopeptidase family)

Clan MA(E), mentioned above, also includes family M41. Proteases of the M41 family are ATP-dependent metalloproteinases, also called FtsH peptidases [56]. These peptidases present the H**E**xxH motif and a third zinc ligand, which is a downstream aspartate. An ATPase domain follows the peptidase domain. In many bacteria, their activity increases as the temperature rises or during osmotic stress. These proteases thus play a role in protection against environmental stress [40].

In 2007, Karnataki et al. [29] identified a membrane-associated AAA (ATPases associated with diverse cellular activities) protease in *T. gondii* of the FtsH1 type (M41 peptidase family), corresponding to TGME49_259260. FtsH1 is an integral membrane protein which is targeted to the *T. gondii* apicoplast. From pulse-chase assays, the authors showed that two cleavages occurred within this protein sequence: a first one in the N-terminal part and a second one in the C-terminal part, allowing specific apicoplast targeting of this FtsH1 [29–30]. The authors suggested that the roles of FtsH1 in *T. gondii* could include protein

Table 3. Comparative study of the metallopeptidase repertoires for *T. gondii* (*Tg*) strain ME49, *N. caninum* (*Nc*) strain Liverpool, *H. hammondii* strain HH34, *E. tenella* (Et) strain Houghton, *P. falciparum* (*Pf*) strain 3D7, and *C. parvum* (*Cp*) strain Iowa. Metallopeptidases are indicated by their EupathDB accession numbers and are classified into MEROPS families using PFAM domains and Blast similarity searches.

MEROPS families	PFAM domain	*Tg* (ME49)	*Nc* (Liverpool)	*H. Hammondii (H.H34)*	*Et* (Hougton)	*Pf* (3D7)	*Cp* (Iowa II)
M1	PF01433	TGME49_224350 TGME49_221310 TGME49_224460 TGME49_262575*	NCLIV_048240 NCLIV_048230	HHA_224460 HHA_221310 HHA_224350 HHA_262575*	ETH_00013105 ETH_00015595* ETH_00038820*	PF3D7_1311800 PF3D7_1472400**	cgd8_3430
M3	PF01432 PF08439	TGME49_216150 TGME49_226420 TGME49_272670	NCLIV_046540 NCLIV_060010 NCLIV_034640	HHA_226420 HHA_216150 HHA_272670	ETH_00003860	PF3D7_1337000 PF3D7_1005700	-
M13	PF01431	TGME49_295640	NCLIV_002180	HHA_295640	-	-	-
M14	PF00246	TGME49_202910 TGME49_271870 TGME49_253170 TGME49_265780	NCLIV_035270 NCLIV_007710 NCLIV_039720 NCLIV_022010	HHA_271870 HHA_202910 HHA_253170 HHA_265780	ETH_00001580 ETH_00040020	PF3D7_0103400	cgd4_4160 cgd1_370
M16	PF00675 PF05193 PF16187 PF08367	TGME49_235680 TGME49_244480 TGME49_206510 TGME49_202680 TGME49_269885 TGME49_236210 TGME49_257010 TGME49_253890 TGME49_214490 TGME49_314850 TGME49_227948	NCLIV_008220 NCLIV_019040 NCLIV_036700 NCLIV_044230 NCLIV_050470 NCLIV_050050 NCLIV_045460 NCLIV_022220 NCLIV_051650 NCLIV_029950 NCLIV_051650	HHA_235680 HHA_244490 HHA_214490 HHA_202680 HHA_269885 HHA_257010 HHA_236210 HHA_253890 HHA_206510	ETH_00038595 ETH_00007345 ETH_00032950 ETH_00001730 ETH_00011835 ETH_00012015 ETH_00018155 ETH_00003350 ETH_00033920	PF3D7_1440200 PF3D7_1118300 PF3D7_0933600 PF3D7_1360800 PF3D7_1121800 PF3D7_0523100	cgd3_4280 cgd6_5520 cgd5_3400 cgd3_4210 cgd3_4240 cgd2_2760 cgd2_930 cgd3_4220 cgd3_4250 cgd7_2080 cgd3_4270 cgd8_2720 cgd3_4180 cgd2_920 cgd1_1680 cgd3_4200 cgd3_4260 cgd3_4170 cgd5_2660 cgd2_4270
M17	PF00883 PF02789	TGME9_290670	NCLIV_042660	HHA_290670	ETH_00012380	PF3D7_1446200	cgd5_2600
M18	PF02127	TGME49_297970	NCLIV_006860	HHA_297970	ETH_00026985	PF3D7_0932300	cgd3_3610
M20	PF07687 PF01546	TGME49_213520	NCLIV_069500 NCLIV_013170	HHA_213520	ETH_00002825	-	-
M24	PF00557 PF16188	TGME49_248850 TGME49_211330 TGME49_257730 TGME49_305460 TGME49_279390 TGME49_221670 TGME49_261600 TGME49_233310	NCLIV_064990 NCLIV_028520 NCLIV_005190 NCLIV_043950 NCLIV_011650 NCLIV_025890 NCLIV_038400 NCLIV_033100	HHA_233310 HHA_257730 HHA_248850 HHA_261600 HHA_305460 HHA_221670 HHA_211330 HHA_279390	ETH_00020895 ETH_00023050 ETH_00006870 ETH_00004260 ETH_00020890 ETH_00029300 ETH_00007755	PF3D7_0517400 PF3D7_0804400 PF3D7_1454400 PF3D7_1428300 PF3D7_1434600 PF3D7_1015300 PF3D7_0527300	cgd2_2480 cgd3_2390 cgd7_1930 cgd1_2700 cgd4_2910
M28	PF04389	TGME49_225850 TGME49_231130	NCLIV_046980 NCLIV_031500	HHA_231130 HHA_225850	ETH_00034400	-	-
M41	PF01434	TGME49_202630 TGME49_300020 TGME49_259260	NCLIV_022310 NCLIV_027270 NCLIV_064680	HHA_259260 HHA_300020 HHA_202630	ETH_00025015 ETH_00010985	PF3D7_1464900 PF3D7_1119600 PF3D7_1239700	cgd1_3360
M48	PF01435	TGME49_221170	NCLIV_004750	HHA_221170	ETH_00017305	-	cgd6_70
M50	PF02163 PF13398	TGME49_266140 TGME49_285670	NCLIV_048090 NCLIV_014700	HHA_285670 HHA_266140	ETH_00009130	PF3D7_1305600 PF3D7_1349700	-
M54	PF07998	-	-	-	-	-	cgd3_3240
M60	PF03272 PF13402	-	-	-	-	-	cgd5_1990
M67	PF01398 PF05021 PF14464	TGME49_251500 TGME49_231970 TGME49_228190 TGME49_269250 TGME49_269840 TGME49_208590	NCLIV_031970 NCLIV_036720 NCLIV_037150 NCLIV_045240 NCLIV_053280 NCLIV_066600	HHA_231970 HHA_269250 HHA_271440 HHA_308590 HHA_251500 HHA_269840 HHA_228190	ETH_00011185 ETH_00013695 ETH_00015225	PF3D7_0912900 PF3D7_0918300 PF3D7_1368100	cgd6_3270 cgd7_1080 cgd7_1970 cgd7_2900
M76	PF09768	TGME49_257110	NCLIV_029840	HHA_257110	-	PF3D7_1441700	-
Peptidases number		**49**	**47**	**48**	**33**	**29**	**38**
Peptidase family number		**15**	**15**	**15**	**13**	**11**	**11**
Clan number		**8**	**8**	**8**	**8**	**8**	**7**

surveillance, chaperone activity, and import [29]. Its function, however, has not yet been fully determined.

Three *T. gondii* M41 peptidases have been identified in ToxoDB: TGME49_202630, TGME49_200020 and TGME49_259260, among which only the latter has been described in the literature [29].

The *P. falciparum* genome encodes for three M41 peptidases. One of them (PF3D7_1239700) was identified as a AAA+/FtsH protease homolog (Pf FtsH1), exhibiting an ATP- and Zn^{2+}-dependent protease activity and it has been localized in the *P. falciparum* mitochondria [60].

M48 Peptidase family (Ste24 endopeptidase and HtpX peptidase families)

Also belonging to clan MA(E), the M48 family is divided into two sub-families: M48A (ste24 endopeptidase) and M48B (HtpX peptidase) [56].

Only one *T. gondii* metallopeptidase was identified in ToxoDB for this M48 peptidase family, TGME49_221170, but no protein of this family has been published to date. Other apicomplexa such as *C. parvum*, *E. tenella* or *N. caninum* also encode for one M48 metallopeptidase in their genomes, but *P. falciparum* does not seem to encode this enzyme.

M50 Peptidase family (S2P peptidase and sporulation factor SpoIVFB families)

The M50 peptidase family consists of metalloendopeptidases with a single zinc in their active site, characteristic of clan MM. They form a distinct family of polytopic membrane metalloproteases containing 4 to 8 transmembrane domains. The M50 family presents a conserved 3 transmembrane domain core structure, containing the H**E**xxH motif within the first transmembrane domain of the core, and a second highly conserved motif called NxxPxxxxxxDG present in the third transmembrane domain; the three underlined amino-acids being the three zinc-ligands [56]. This M50 family has been divided into two sub-families: M50A (S2P protease) and M50B (sporulation factor SpoIVFB) [32–33].

As of today, no protein of this family has been described for *T. gondii* in the literature. Only two predicted proteases have been found in the genome of *T. gondii*: TGME49_266140 and TGME49_285670.

Plasmodium parasites encode in their genome two M50B-like proteases (PFAM13398): PF3D7_1305600 and PF3D7_1349700, according to Deu et al. (2017) [16], but lack the NxxPxxxxxxDG motif. In all invasive stages, the protein is in close proximity to the nucleus.

M67 Peptidase family (Poh1 peptidase, JAMM-like protein and AMSH deubiquitinating peptidase families)

Clan MP contains a single family, M67 which presents divergent sequences divided into three sub-families: M67A (Poh1 peptidase component of the 26S proteasome), M67B (archean JAMM-like proteins), and M67C (AMSH deubiquitinating peptidase) [56]. The feature of their catalytic site motif is HxH, where the two underlined histidines provide zinc ligands together with an aspartate C-terminal to this motif; a glutamate N-terminal to this motif is a catalytic residue [56].

Six *T. gondii* peptidases have been identified in ToxoDB as belonging to this M67 family, none of which has been described in the literature to date.

However, two publications have described the proteasome of the malaria parasite, proposing enzymes involved in this pathway as promising drug targets for chemotherapeutic intervention as well as experimental evidence for metalloproteases in the proteasome complex [2,64]. In *T. gondii*, one publication described proteolytic activities in the proteasome, without indication of the presence of metalloprotease [51].

M76 Peptidase family (Atp23 peptidase family)

These enzymes contain a H**E**xxH motif, in which E (bold) is a catalytic residue and the two H (underlined) are zinc-ion ligands (clan MA(E)), but the third zinc ligand has not yet been identified. The M76 peptidase family consists of endopeptidases whose functions are to achieve the synthesis of ATP from ADP and phosphate, a process occurring in mitochondria [56]. Only one *T. gondii* enzyme was found in ToxoDB: TGME49_257110, with a predicted localization within mitochondria. Yet, no member of this protease family has been described to date in the *T. gondii* literature.

The enigma of M22 Peptidase family

During this study, we identified proteins ascribed to the "M22 peptidase family" in the Eupath database, including two members in the *T. gondii* genome, TGME49_274110 and TGME49_202310. While studying them, we however discovered that this family has been retracted from the MEROPS database, because there is a lack of experimental evidence to support peptidase activity as a general property of this family. The only evidence for any proteolytic activity in M22 was attributed to the O-sialoglycopeptidase from *Pasteurella haemolytica*. Homologs are almost universally distributed, but peptidase activity for members of this family has never been found. Structural studies have shown that members of "M22" have a very different fold to any known metallopeptidase (Rawlings, personal communication), and therefore they have been retracted from the MEROPS Database. Since the M22 domain signature continues to be present in the Eupath database and EMBL-EBI (Interpro service), we thought it was important to mention here that they are not members of the metallopeptidase superfamily, the focus of this current review.

Conclusions

Metallopeptidases are of great importance in basic cell functions but also in specific cell functions. It is therefore

necessary to inventory them for *T. gondii* as a way to better understand the biology of this parasite as well as the complexity of hosts and host-cell interactions. Also, with the aim of eventually undertaking a comparative study of apicomplexan genetic inheritance, it is worth mentioning that currently, *T. gondii* is the organisms that has the largest genome and encodes the highest number of genes, among all currently known apicomplexa.

At present, seven metalloproteases have been studied experimentally and described in *T. gondii*: an aminopeptidase N (family M1, aminopeptidase N) [5], two toxolysins (family M16, pitrilysin) [23,34], a leucine aminopeptidase (family M17, leucyl aminopeptidase) [25], an aspartyl aminopeptidase (family M18, aminopeptidase I) [69], a X-prolyl aminopeptidase (family M24, aminopeptidase P) [67], and a FtsH1 peptidase (family M41, FtsH peptidase) [29–30]. Out of these seven metalloproteases, only two have been shown to be involved in the invasion process of *T. gondii* within the host cell: toxolysins-1 and −4. The other metallopeptidases could be involved to various extents in a variety of metabolic pathways of *T. gondii*.

Overall, 49 metallopeptidases (7 published and 42 putative) containing various typical metallopeptidase signature motifs were identified in this study. Expression analysis of the corresponding 49 metallopeptidase genes in tachyzoite stages revealed the presence of transcripts for all of them, even if at low levels for some, such as M18 or M67 members for example. However, it would be interesting to adopt a quantitative PCR approach for each metallopeptidase, and thus to determine the expression levels of each.

Metalloproteases can be used to modify/degrade the host but also to activate some parasite proteins and they can be involved in egress, in invasion probably acting primarily as maturases, and in interactions with the host cell. *T. gondii* is also able to cross the basal membrane composed of laminin, Type III, IV and VII collagens, as well as glycosaminoglycans in order to diffuse in all organisms [4]. In addition, *T. gondii* must pass within the extracellular matrix composed of elastin and glycoproteins.

On the basis of the *in silico* study describing all putative and/or published metalloproteases in the *T. gondii* genome, we noted that some metalloprotease families were completely absent in the currently known apicomplexan peptidase families, and that some families were present only in one apicomplexan species: for example, peptidase family M54 and M60 are only present in *C. parvum*.

In conclusion, several families of metalloprotease are not represented in an identical manner depending on the parasite's biology, physiology or host interaction, and could be potential therapeutic targets.

According to our comparative survey of metallopeptidases in 6 representative apicomplexan species (*T. gondii*, *N. caninum*, *H. hammondii*, *E. tenella*, *P. falciparum* and *C. parvum*), *T. gondii* together with *N. caninum* and *H. hammondii* contain the most numerous and diverse repertoire (49, 47, 48), followed by *C. parvum* (38), *E. tenella* (33) and then *P. falciparum* (29) (Table 3). This result is consistent with the recent observations by Woo et al., 2015 [65], indicating that the *T. gondii* genome would be currently the least reduced one – among all currently known apicomplexan genomes – compared to the genome that has been inferred for the apicomplexan common ancestor [39]. Besides having the largest number of metallopeptidases, *T. gondii, N. caninum* and *H. hammondii* also have the most diverse representation of metallopeptidases families (15), *P. falciparum* and *C. parvum* having the most reduced diversity (11 families), and *E. tenella* and intermediary status (13 families). The *C. parvum* repertoire is rather atypical with reduced diversity in terms of metallopeptidase families (11) but one of the largest sets of metallopeptidases (38), a situation that is due to remarkable expansion of the M16 family members in this species, the biological function of which will certainly deserve further investigations.

Interestingly, this comparative inventory reveals only two families that are evenly represented in the 6 representative species in terms of members: the M17 and M18 families, which each have a single member in the 6 species. For all the other metallopeptidase families there are many members (up to 20 for M16 in *C. parvum*) to none, possibly reflecting specific functions in the biology or host-parasite interactions of these species.

Thus, beyond its importance in providing novel putatively relevant targets for *T. gondii* chemotherapy, this inventory of *T. gondii* metallopeptidases provides the groundwork for functional investigations of their functions in parasite biology and host-parasite interactions of the diversity of apicomplexan parasites.

Acknowledgements. The authors would like to thank Prof J. Depaquit for his advice and helpful discussions. This work was supported by the *Région Champagne-Ardenne*, France. Anne-Pascaline Bouleau is the recipient of a grant from the *Région Champagne-Ardenne*.

Conflict of interest

The authors declare that they have no conflict of interest.

Appendix 1

Metallopeptidase genes identified and classified in the *T. gondii* genome database (strain GT1, genotype I).
We used Pfam motifs (http://pfam.xfam.org) in association with the MEROPS Database to screen the *T. gondii* database (http://toxodb.org/toxo/, Release 29). The motif organization of predicted peptidases was studied using the InterProScan Search (http://www.ebi.ac.uk/interpro/) and family assignment is based on MEROPS – the peptidase Database – classification (https://www.ebi.ac.uk/merops/).

MEROPS Clan Number of metal ion	MEROPS families	Gene ID GT1 strain	Chro.	Protein Length (aa)	ToxoDB Product Description (GT-1)	Blast MEROPS %identities/e-score (Families)	Primer Name	Alias	Publication	PFAM	Signa pept
MA(E) 1 Zn^{2+}	M1	TGGT1_224350A	X	1219	putative aminopeptidase N	65,6%/5,9e-178 (M1)	M1	·	·	PF01433	No
		TGGT1_221310	II	1069	aminopeptidase N protein	97,9%/3,7e-156 (M1)	M1A	-	-		No
		TGGT1_224460	X	970	putative aminopeptidase N	97,0%/8,5e-155 (M1)	M1B	Tg110	Berthonneau J. 2000		Yes
		TGGT1_262575	VIIb	290	hypothetical protein	41,0%/3,8e-13 (M1)	M1C	·	·		No
MA(E) 1 Zn^{2+}	M3	TGGT1_226420	X	667	peptidase family M3 protein	95,2%/2,9e-179 (M3B)	M3	·	·	PF01432 PF08439	Yes
		TGGT1_272670	VIII	1094	peptidase family M3 protein	67,1%/9,4e-200 (M3A)	M3A	·	·		No
		TGGT1_216150	XI	506	peptidase family M3 protein	97,9%/8,7e-190 (M3B)	M3B	·	·		No
MA(E) 1 Zn^{2+}	M13	TGGT1_295640	Ia	1038	peptidase family M13 protein	87,6%/5,8e-304 (M13)	M13	·	·	PF01431	Yes
MC 1 Zn^{2+}	M14	TGGT1_265780	IX	2805	flagellar/basal body protein	84,6%/3,2e-46 (M14B)	M14	·	·	PF00246	yes
		TGGT1_271870	VIII	1377	zinc carboxypeptidase superfamily protein	73,9%/1,7e-110 (M14A)	M14A	·	·		No
		TGGT1_253170	III	2203	putative zinc carboxypeptidase	92,9%/3,0e-134 (M14B)	M14B	·	·		No
		TGGT1_202910	VIIa	319	zinc carboxypeptidase superfamily protein	93,8%/1,2e-167 (M14A)	M14C	·	·		yes
ME 1 Zn^{2+}	M16	TGGT1_202680	VIIa	563	putative peptidase M16, alpha subunit	100,0%/5,1e-145(M16B)	M16	MPPA	·	PF00675 PF05193 PF16187 PF08367	yes
		TGGT1_253890	III	1604	peptidase M16 inactive domain-containing protein	92,2%/0e-0(M16)	M16A	·	·		yes
		TGGT1_235680	X	1692	peptidase M16 inactive domain-containing protein	100,0%/3,0e-161(M16B)	M16B	·	·		No
		TGGT1_244480	VI	306	peptidase M16 inactive domain-containing protein	92,9%/3,3e-77(M16A)	M16C	·	·		No
		TGGT1_206510	VIIa	2435	toxolysin TLN4	100,0%/2,9e-129(M16A)	M16D	TLN4	Laliberté J. 2011		yes
		TGGT1_257010	VIIb	1023	sporozoite developmental protein	100,0%/1,2e-141(M16A)	M16E	·	·		yes
		TGGT1_269885A	VIII	1291	rhoptry metalloprotease toxolysin TLN1	99,5%/2,9e-133(M16A)	M16F	TLN1	Hajagos BE. 2012		yes
		TGGT1_227948	X	1306	peptidase M16 inactive domain-containing protein	100,0%/5,6e-299(M16C)	M16G	·	·		yes
		TGGT1_236210	X	509	putative peptidase M16 family potein	68,6%/1,9e-188(M16B)	M16H	·	·		yes
		TGGT1_314850	XI	2174	hypothetical protein	41,5%/1,0e-08(M16A)	M16I	·	·		No
		TGGT1_214490	X	1353	peptidase M16 inactive domain-containing protein	99,8%/0e-0(M16C)	M16J	·	·		No
MF 2 Zn^{2+}/Mn^{2+}	M17	TGGT1_290670	IX	781	leucyl aminopeptidase LAP	100,0%/8,0e-248(M17)	M17	TgLAP	Jia H. 2010	PF00883 PF02789	yes
MH 2 Zn^{2+}	M18	TGGT1_297970	II	508	aspartyl aminopeptidase	100,0%/0,0e-0(M18)	M18	TgAAP	Zheng J. 2016	PF02127	No
MH 2 Zn^{2+}	M20	TGGT1_213520	V	514	peptidase M20D, amidohydrolase	100,0%/3,6e-268(M20D)	M20	·	·	PF07687 PF01546	yes
MG 2 Co^{2+}/Mn^{2+}	M24	TGGT1_248850	XII	416	methionine aminopeptidase	100,0%/5,5e-183 (M24)	M24	·	·	PF00557 PF16188	No
		TGGT1_211330	IV	697	methionine aminopeptidase	100,0%/8,5e-174 (M24)	M24A	·	·		No
		TGGT1_257730	VIIb	484	putative methionine aminopeptidase, type i	99,6%/7,0e-165 (M24)	M24B	·	·		Yes
		TGGT1_305460	IX	480	putative methionine aminopeptidase 2	97,8%/1,5e-210 (M24)	M24C	·	·		No
		TGGT1_279390	IX	462	putative proliferation-associated protein 2G4	100,0%/3,7e-372 (M24)	M24D	·	·		No
		TGGT1_261600	VIIb	886	creatinase domain-containing protein	100,0%/8,9e-173 (M24B)	M24E	TgAPP	Yang M. 2016		No
		TGGT1_233310	VIII	599	putative peptidase D	99,2%/2,6e-181 (M24)	M24F	·	·		No
		TGGT1_221670	II	1198	transcriptional elongation factor FACT140	40,2%/1,0e-56 (M24)	M24G	·	·		No
MH 2 Zn^{2+}	M28	TGGT1_225850	X	1555	peptidase, M28 family protein	35,5%/6,6e-40 (M28)	M28	·	·	PF04389	No
		TGGT1_231130	VIII	711	hypothetical protein	41,2%/2,9e-33 (M28)	M28A	·	·		No
MA(E) 1 Zn^{2+}	M41	TGGT1_202630	VIIa	1188	ATP-dependent metallopeptidase HflB subfamily protein	100,0%/1,5e-156 (M41)	M41A	·	·	PF01434	Yes
		TGGT1_300020	XII	1005	ATP-dependent metallopeptidase HflB subfamily protein	100,0%/4,1e-138 (M41)	M41B	·	·		Yes
		TGGT1_259260	VIIb	1250	membrane protein FtsH1	100,0%/2,5e-155 (M41)	M41C	FtsH1	Karnataki A. 2007		No
MA(E) 1 Zn^{2+}	M48	TGGT1_221170	II	432	CAAX metallo endopeptidase	99,6%/5,0e-188 (M48)	M48	·	·	PF01435 PF16491	No
MM 1 Zn^{2+}	M50	TGGT1_266140	IX	378	peptidase, M50 family protein	40,7%/1,4e-42 (M50B)	M50	·	·	PF02163 PF13398	No
		TGGT1_285670	V	260	hypothetical protein	63,6%/0,22 (M50B)	M50A	·	·		No
MP 1 Zn^{2+}	M67	TGGT1_251500	XII	600	putative eukaryotic initiation factor-3, subunit 3	53,7%/6,2e-84 (M67)	M67A	·	·	PF01398 PF05021 PF14464	Yes
		TGGT1_231970	VIII	2538	pre-mRNA processing splicing factor PRP8	28,9%/0,078(M67D)	M67B	·	·		No
		TGGT1_228190	X	346	putative eukaryotic initiation factor-3, subunit 5	78,0%/2,8e-61(M67A)	M67C	·	·		No
		TGGT1_269250	VIII	343	Mov34/MPN/PAD-1 family protein	66,4%/1,4e-57(M67)	M67D	·	·		No
		TGGT1_269840	VIII	314	proteasome regulatory subunit	100,0%/1,1e-271(M67A)	M67E	·	·		No
		TGGT1_308590	XI	489	Mov34/MPN/PAD-1 family protein	100,0%/3,5e-179(M67A)	M67F	·	·		Yes
MA(E) 1 Zn^{2+}	M76	TGGT1_257110	VIIb	526	hypothetical protein	98,7%/1,4e-156(M76)	M76	·	·	PF09768	No

References

1. Allary M, Schrével J, Florent I. 2002. Properties, stage-dependent expression and localization of *Plasmodium falciparum* M1 family zinc-aminopeptidase. Parasitology, 125, 1–10.
2. Aminake MN, Arndt H-D., Pradel G. 2012. The proteasome of malaria parasites: a multi-stage drug target for chemotherapeutic intervention? International Journal for Parasitology. Drugs and Drug Resistance, 2, 1–10.
3. Azimzadeh O, Sow C, Gèze M, Nyalwidhe J, Florent I. 2010. *Plasmodium falciparum* PfA-M1 aminopeptidase is trafficked via the parasitophorous vacuole and marginally delivered to the food vacuole. Malaria Journal, 9, 189.
4. Barragan A, Sibley LD. 2003. Migration of *Toxoplasma gondii* across biological barriers. Trends in Microbiology, 11, 426–430.
5. Berthonneau J, Rodier MH, El Moudni B, Jacquemin JL. 2000. *Toxoplasma gondii*: purification and characterization of an immunogenic metallopeptidase. Experimental Parasitology, 95, 158–162.
6. Bland ND, Pinney JW, Thomas JE, Turner AJ, Isaac RE. 2008. Bioinformatic analysis of the neprilysin (M13) family of peptidases reveals complex evolutionary and functional relationships. BMC Evolutionary Biology, 8, 16.
7. Bounaadja L, Schmitt M, Albrecht S, Mouray E, Tarnus C, Florent I. 2017. Selective inhibition of PfA-M1, over PfA-M17, by an amino-benzosuberone derivative blocks malaria parasites development *in vitro* and *in vivo*. Malaria Journal, 16, 382.

8. Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J, DeRisi JL. 2003. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. PLoS Biology, 1, E5.
9. Bradley PJ, Ward C, Cheng SJ, Alexander DL, Coller S, Coombs GH, Dunn JD, Ferguson DJ, Sanderson SJ, Wastling JM, Boothroyd JC. 2005. Proteomic analysis of rhoptry organelles reveals many novel constituents for host-parasite interactions in *Toxoplasma gondii*. The Journal of Biological Chemistry, 280, 34245–34258.
10. Burley SK, David PR, Taylor A, Lipscomb WN. 1990. Molecular structure of leucine aminopeptidase at 2.7 Å resolution. Proceedings of the National Academy of Sciences of the United States of America, 87, 6878–6882.
11. Chen X, Chong CR, Shi L, Yoshimoto T, Sullivan DJ, Liu JO. 2006. Inhibitors of *Plasmodium falciparum* methionine aminopeptidase 1b possess antimalarial activity. Proceedings of the National Academy of Sciences of the United States of America, 103, 14548–14553.
12. Chen X, Xie S, Bhat S, Kumar N, Shapiro TA, Liu JO. 2009. Fumagillin and fumarranol interact with *P. falciparum* methionine aminopeptidase 2 and inhibit malaria parasite growth *in vitro* and *in vivo*. Chemistry and Biology, 16, 193-202.
13. Cleary MD, Singh U, Blader IJ, Brewer JL, Boothroyd JC. 2002. *Toxoplasma gondii* asexual development: identification of developmentally regulated genes and distinct patterns of gene expression. Eukaryotic Cell, 1, 329–340.
14. Dalal S, Klemba M. 2007. Roles for two aminopeptidases in vacuolar hemoglobin catabolism in *Plasmodium falciparum*. Journal of Biological Chemistry, 282, 35978-35987.
15. Dalal S, Ragheb D, Schubot F, Klemba M. 2013. A naturally variable residue in the S1 subsite of M1 family aminopeptidases modulates catalytic properties and promotes functional specialization. Journal of Biological Chemistry, 288, 26004-26012.
16. Deu E. 2017. Proteases as antimalarial targets: strategies for genetic, chemical, and therapeutic validation. FEBS Journal, 284, 2604-2628.
17. Eggleson KK, Duffin KL, Goldberg DE. 1999. Identification and characterization of falcilysin, a metallopeptidase involved in hemoglobin catabolism within the malaria parasite *Plasmodium falciparum*. Journal of Biological Chemistry, 274, 32411–32417.
18. Fetterer RH, Miska KB, Barfield RC. 2005. Partial purification and characterization of an aminopeptidase from *Eimeria tenella*. Journal of Parasitology, 91, 1280–1286.
19. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research, 44, D279-285.
20. Florent I, Derhy Z, Allary M, Monsigny M, Mayer R, Schrével J. 1998. A *Plasmodium falciparum* aminopeptidase gene belonging to the M1 family of zinc-metallopeptidases is expressed in erythrocytic stages. Molecular and Biochemical Parasitology, 97, 149–160.
21. Gakh O, Cavadini P, Isaya G. 2002. Mitochondrial processing peptidases. Biochimica et Biophysica Acta, 1592, 63–77.
22. Gras S, Byzia A, Gilbert FB, McGowan S, Drag M, Silvestre A, Niepceron A, Lecaille F, Lalmanach G, Brossier F. 2014. Aminopeptidase N1 (EtAPN1), an M1 metalloprotease of the apicomplexan parasite *Eimeria tenella*, participates in parasite development. Eukaryotic Cell, 13, 884–895.
23. Hajagos BE, Turetzky JM, Peng ED, Cheng SJ, Ryan CM, Souda P, Whitelegge JP, Lebrun M, Dubremetz J-F., Bradley PJ. 2012. Molecular dissection of novel trafficking and processing of the *Toxoplasma gondii* rhoptry metalloprotease toxolysin-1. Traffic, 13, 292–304.
24. Harbut MB, Velmourougane G, Dalal S, Reiss G, Whisstock JC, Onder O, Brisson D, McGowan S, Klemba M, Greenbaum DC. 2011. Bestatin-based chemical biology strategy reveals distinct roles for malaria M1- and M17-family aminopeptidases. Proceedings of the National Academy of Sciences of the United States of America, 108, E526-E534.
25. Jia H, Nishikawa Y, Luo Y, Yamagishi J, Sugimoto C, Xuan X. 2010. Characterization of a leucine aminopeptidase from *Toxoplasma gondii*. Molecular and Biochemical Parasitology, 170, 1–6.
26. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Research, 42, D199-205.
27. Kang J-M., Ju H-L., Sohn W-M., Na B-K. 2011. Molecular cloning and characterization of a M17 leucine aminopeptidase of *Cryptosporidium parvum*. Parasitology, 138, 682–690.
28. Kannan Sivaraman K, Paiardini A, Sienczyk M, Ruggeri C, Oellig CA, Dalton JP, Scammells PJ, Drag M, McGowan S. 2013. Synthesis and structure-activity relationships of phosphonic arginine mimetics as inhibitors of the M1 and M17 aminopeptidases from *Plasmodium falciparum*. Journal of Medicinal Chemistry, 56, 5213–5217.
29. Karnataki A, Derocher AE, Coppens I, Feagin JE, Parsons M. 2007. A membrane protease is targeted to the relict plastid of *toxoplasma* via an internal signal sequence. Traffic, 8, 1543–1553.
30. Karnataki A, DeRocher AE, Feagin JE, Parsons M. 2009. Sequential processing of the *Toxoplasma* apicoplast membrane protein FtsH1 in topologically distinct domains during intracellular trafficking. Molecular and Biochemical Parasitology, 166, 126–133.
31. Katrib M, Ikin RJ, Brossier F, Robinson M, Slapetova I, Sharman PA, Walker RA, Belli SI, Tomley FM, Smith NC. 2012. Stage-specific expression of protease genes in the apicomplexan parasite, *Eimeria tenella*. BMC genomics, 13, 685.
32. Kinch LN, Ginalski K, Grishin NV. 2006. Site-2 protease regulated intramembrane proteolysis: sequence homologs suggest an ancient signaling cascade. Protein Science, 15, 84-93.
33. Koussis K, Goulielmaki E, Chalari A, Withers-Martinez C, Siden-Kiamos I, Matuschewski K, Loukeris TG. 2017. Targeted deletion of a *Plasmodium* site-2 protease impairs life cycle progression in the mammalian host. PLoS One, 12, e0170260.
34. Laliberté J, Carruthers VB. 2011. *Toxoplasma gondii* toxolysin 4 is an extensively processed putative metalloproteinase secreted from micronemes. Molecular and Biochemical Parasitology, 177, 49–56.
35. Lau YL, Lee WC, Gudimella R, Zhang G, Ching XT, Razali R, Aziz F, Anwar A, Fong MY. 2016. Deciphering the draft genome of *Toxoplasma gondii* RH strain. PLoS One, 11, e0157901.
36. Lauterbach SB, Coetzer TL. 2008. The M18 aspartyl aminopeptidase of *Plasmodium falciparum* binds to human erythrocyte spectrin *in vitro*. Malaria Journal, 7, 161.
37. Li H, Child MA, Bogyo M. 2012. Proteases as regulators of pathogenesis: examples from the *Apicomplexa*. Biochimica et Biophysica Acta, 1824, 177–185.

38. Llinás M, Bozdech Z, Wong ED, Adai AT, DeRisi JL. 2006. Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. Nucleic Acids Research, 34, 1166-73.
39. Lorenzi H, Khan A, Behnke MS, Namasivayam S, Swapna LS, Hadjithomas M, Karamycheva S, Pinney D, Brunk BP, Ajioka JW, Ajzenberg D, Boothroyd JC, Boyle JP, Dardé ML, Diaz-Miranda MA, Dubey JP, Fritz HM, Gennari SM, Gregory BD, Kim K, Saeij JP, Su C, White MW, Zhu XQ, Howe DK, Rosenthal BM, Grigg ME, Parkinson J, Liu L, Kissinger JC, Roos DS, Sibley LD. 2016. Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic *Toxoplasma gondii* genomes. Nature communications, 7, 10147.
40. Lüdke A, Krämer R, Burkovski A, Schluesener D, Poetsch A. 2007. A proteomic study of *Corynebacterium glutamicum* AAA+ protease FtsH. BMC Microbiology, 25, 6.
41. Marcilla A, De la Rubia JE, Sotillo J, Bernal D, Carmona C, Villavicencio Z, Acosta D, Tort J, Bornay FJ, Esteban JG, Toledo R. 2008. Leucine aminopeptidase is an immunodominant antigen of *Fasciola hepatica* excretory and secretory products in human infections. Clinical and Vaccine Immunology, 15, 95-100.
42. Maric S, Donnelly SM, Robinson MW, Skinner-Adams T, Trenholme KR, Gardiner DL, Dalton JP, Stack CM, Lowther J. 2009. The M17 leucine aminopeptidase of the malaria parasite *Plasmodium falciparum*: importance of active site metal ions in the binding of substrates and inhibitors. Biochemistry, 48, 5435–5439.
43. McGowan S, Porter CJ, Lowther J, Stack CM, Golding SJ, Skinner-Adams TS, Trenholme KR, Teuscher F, Donnelly SM, Grembecka J, Mucha A, Kafarski P, DeGori R, Buckle AM, Gardiner DL, Whisstock JC, Dalton JP. 2009. Structural basis for the inhibition of the essential *Plasmodium falciparum* M1 neutral aminopeptidase. Proceedings of the National Academy of Sciences of the United States of America, 106, 2537–2542.
44. McGowan S, Oellig CA, Birru WA, Caradoc-Davies TT, Stack CM, Lowther J, Skinner-Adams T, Mucha A, Kafarski P, Grembecka J, Trenholme KR, Buckle AM, Gardiner DL, Dalton JP, Whisstock JC. 2010. Structure of the *Plasmodium falciparum* M17 aminopeptidase and significance for the design of drugs targeting the neutral exopeptidases. Proceedings of the National Academy of Sciences of the United States of America, 107, 2449–2454.
45. McKerrow JH. 1989. Parasite proteases. Experimental Parasitology, 68, 111-115.
46. McKerrow JH, Caffrey C, Kelly B, Loke P, Sajid M. 2006. Proteases in parasitic diseases. Annual Review of Pathology, 1, 497-536.
47. Murata CE, Goldberg DE. 2003. *Plasmodium falciparum* falcilysin: an unprocessed food vacuole enzyme. Molecular and Biochemical Parasitology, 129, 123–126.
48. Nankya-Kitaka MF, Curley GP, Gavigan CS, Bell A, Dalton JP. 1998. *Plasmodium chabaudi chabaudi* and *P. falciparum*: inhibition of aminopeptidase and parasite growth by bestatin and nitrobestatin. Parasitology Research, 84, 552-558.
49. Padda RS, Tsai A, Chappell CL, Okhuysen PC. 2002. Molecular cloning and analysis of the *Cryptosporidium parvum* aminopeptidase N gene. International Journal for Parasitology, 32, 187–197.
50. Paiardini A, Bamert RS, Kannan-Sivaraman K, Drinkwater N, Mistry SN, Scammells PJ, McGowan S. 2015. Screening the medicines for malaria venture 'Malaria Box' against the *Plasmodium falciparum* aminopeptidases, M1, M17 and M18. PLoS One, 10, e0115859.
51. Paugam A, Creuzet C, Dupouy-Camet J, Roisin MP. 2001. Evidence for the existence of a proteasome in *Toxoplasma gondii* : intracellular localization and specific peptidase activities. Parasite, 8, 267-73.
52. Ponpuak M, Klemba M, Park M, Gluzman IY, Lamppa GK, Goldberg DE. 2006. A role for falcilysin in transit peptide degradation in the *Plasmodium falciparum* apicoplast. Molecular Microbiology, 63, 314-334.
53. Poreba M, McGowan S, Skinner-Adams TS, Trenholme KR, Gardiner DL, Whisstock JC, To J, Salvesen GS, Dalton JP, Drag M. 2012. Fingerprinting the substrate specificity of M1 and M17 aminopeptidases of human malaria, *Plasmodium falciparum*. PLoS One, 7, e31938.
54. Ragheb D, Dalal S, Bompiani KM, Ray WK, Klemba M. 2011. Distribution and biochemical properties of an M1-family aminopeptidase in *Plasmodium falciparum* indicate a role in vacuolar hemoglobin catabolism. Journal of Biological Chemistry, 286, 27255–27265.
55. Ralph SA. 2007. Subcellular multitasking − multiple destinations and roles for the *Plasmodium* falcilysin protease. Molecular Microbiology, 63, 309–313.
56. Rawlings ND, Barrett AJ, Finn R. 2016. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Research, 44, D343-350.
57. Skinner-Adams TS, Stack CM, Trenholme KR, Brown CL, Grembecka J, Lowther J, Mucha A, Drag M, Kafarski P, McGowan S, Whisstock JC, Gardiner DL, Dalton JP . 2010. *Plasmodium falciparum* neutral aminopeptidases: new targets for anti-malarials. Trends in Biochemical Sciences, 35, 53–61.
58. Spicer T, Fernandez-Vega V, Chase P, Scampavia L, To J, Dalton JP, Da Silva FL, Skinner-Adams TS, Gardiner DL, Trenholme KR, Brown CL, Ghosh P, Porubsky P, Wang JL, Whipple DA, Schoenen FJ, Hodder P. 2014. Identification of potent and selective inhibitors of the *Plasmodium falciparum* M18 aspartyl aminopeptidase (PfM18AAP) of human malaria via high-throughput screening. Journal of Biomolecular Screening, 19, 1107–1115.
59. Stack CM, Lowther J, Cunningham E, Donnelly S, Gardiner DL, Trenholme KR, Skinner-Adams TS, Teuscher F, Grembecka J, Mucha A, Kafarski LL, Bell A, Dalton JP. 2007. Characterization of the *Plasmodium falciparum* M17 leucyl aminopeptidase. A protease involved in amino acid regulation with potential for antimalarial drug development. Journal of Biological Chemistry, 282, 2069–2080.
60. Tanveer A, Allen SM, Jackson KE, Charan M, Ralph SA, Habib S. 2013. An FtsH protease is recruited to the mitochondrion of *Plasmodium falciparum*. PloS One, 8, e74408.
61. Teuscher F, Lowther J, Skinner-Adams TS, Spielmann T, Dixon MWA, Stack CM, Donnelly S, Mucha A, Kafarski P, Vassiliou S, Gardiner DL, Dalton JP, Trenholme KR. 2007. The M18 aspartyl aminopeptidase of the human malaria parasite *Plasmodium falciparum*. Journal of Biological Chemistry, 282, 30817–30826.
62. Trenholme KR, Brown CL, Skinner-Adams TS, Stack C, Lowther J, To J, Robinson MW, Donnelly SM, Dalton JP, Gardiner DL. 2010. Aminopeptidases of malaria parasites: new targets for chemotherapy. Infectious Disorders Drug Targets, 10, 217–225.
63. Van Dooren GG, Su V, D'Ombrain MC, McFadden GI. 2002. Processing of an apicoplast leader sequence in *Plasmodium falciparum* and the identification of a putative leader cleavage enzyme. Journal of Biological Chemistry, 277, 23612-23619.
64. Wang L, Delahunty C, Fritz-Wolf K, Rahlfs S, Helena Prieto J, Yates JR, Becker K. 2015. Characterization of the

26S proteasome network in *Plasmodium falciparum*. Scientific Reports, 5, 17818.
65. Woo YH, Ansari H, Otto TD, Klinger CM, Kolisko M, Michálek J, Saxena A, Shanmugam D, Tayyrov A, Veluchamy A, Ali S, Bernal A, del Campo J, Cihlár J, Flegontov P, Gornik SG, Hajdušková E, Horák A, Janouškovec J, Katris NJ, Mast FD, Miranda-Saavedra D, Mourier T, Naeem R, Nair M, Panigrahi AK, Rawlings ND, Padron-Regalado E, Ramaprasad A, Samad N, Tomčala A, Wilkes J, Neafsey DE, Doerig C, Bowler C, Keeling PJ, Roos DS, Dacks JB, Templeton TJ, Waller RF, Lukeš J, Oborník M, Pain A. 2015. Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. eLife, 4, e06974.
66. Wu Y, Wang X, Liu X, Wang Y. 2003. Data-mining approaches reveal hidden families of proteases in the genome of malaria parasite. Genome Research, 13, 601–616.
67. Yang M, Zheng J, Jia H, Song M. 2016. Functional characterization of X-prolyl aminopeptidase from *Toxoplasma gondii*. Parasitology, 143, 1443–1449.
68. Zheng J, Jia H, Zheng Y. 2015. Knockout of leucine aminopeptidase in *Toxoplasma gondii* using CRISPR/Cas9. International Journal for Parasitology, 45, 141–148.
69. Zheng J, Cheng Z, Jia H, Zheng Y. 2016. Characterization of aspartyl aminopeptidase from *Toxoplasma gondii*. Scientific Reports, 6, 34448.

Cite this article as: Escotte-Binet S, Huguenin A, Aubert D, Martin A-P, Kaltenbach M, Florent I, Villena I. 2018. Metallopeptidases of *Toxoplasma gondii*: *in silico* identification and gene expression. Parasite **25**, 26