Published in final edited form as:
Microbes Infect. 2011 December ; 13(14-15): 1199–1210. doi:10.1016/j.micinf.2011.08.013.

Analysis of the Glycoproteome of *Toxoplasma gondii* using Lectin Affinity Chromatography and Tandem Mass Spectrometry

Qilie Luo[1,5], **Rajendra Upadhya**[2], **Hong Zhang**[2], **Carlos Madrid-Aliste**[5], **Edward Nieves**[2,5], **Kami Kim**[3,4], **Ruth Hogue Angeletti**[1,5], and **Louis M. Weiss**[2,4,*]

[1]Department of Developmental and Molecular Biology, Albert Einstein College of Medicine, Bronx, NY 10461

[2]Department of Pathology, Albert Einstein College of Medicine, Bronx, NY 10461

[3]Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY 10461

[4]Department of Medicine, Albert Einstein College of Medicine, Bronx, NY 10461

[5]Laboratory for Macromolecular Analysis and Proteomics, Albert Einstein College of Medicine, Bronx, NY 10461

Abstract

Glycoproteins are involved in many important molecular recognition processes including invasion, adhesion, differentiation, and development. To identify the glycoproteins of *Toxoplasma gondii*, a proteomic analysis was undertaken. *T. gondii* proteins were prepared and fractioned using lectin affinity chromatography. The proteins in each fraction were then separated using SDS-PAGE and identified by tryptic in gel digestion followed by tandem mass spectrometry. Utilizing these methods 132 proteins were identified. Among the identified proteins were 17 surface proteins, 9 microneme proteins, 15 rhoptry proteins, 11 heat shock proteins (HSP), and 32 hypothetical proteins. Several proteins had 1 to 5 transmembrane domains (TMD) with some being as large as 608.3 kDa. Both lectin-fluorescence labeling and lectin blotting were employed to confirm the presence of carbohydrates on the surface or cytoplasm of *T. gondii* parasites. PCR demonstrated that selected hypothetical proteins were expressed in *T. gondii* tachyzoites. This is data provides a large scale analysis of the *T. gondii* glycoproteome. Studies of the function of glycosylation of these proteins may help elucidate mechanism(s) involved in invasion improving drug therapy as well as identify glycoproteins that may prove to be useful as vaccine candidates.

Keywords

glycosylation; Toxoplasma; glycoproteome; membrane proteins; lectin chromatography

[*]To whom correspondence should be addressed: Louis M. Weiss, Department of Pathology, 1300 Morris Park Avenue, Bronx, NY 10461. Tel: 718-430-3475; Fax: 718-430-8939; louis.weiss@einstein.yu.edu.

1. Introduction

Toxoplasma gondii is an obligate intracellular parasite of both humans and domestic animals. Infection in humans is common occurring via food-borne or waterborne transmission and via maternofetal transmission resulting in congenial infection [1, 2]. There are three major life stages of this Apicomplexan: the tachyzoite which is involved in acute infection and dissemination of the parasite in its host, the bradyzoite which is found in tissue cysts and latent infection, and the oocysts which is the sexual stage that develops in the feline gastrointestinal system[3]. Usually people become infected with *T. gondii* after ingestion of uncooked or undercooked meat products containing bradyzoites, or oocyst/sporozotes from contaminated water or soil[4]. This causes an acute infection, due to tachyzoites, which is occasionally symptomatic, but in the majority of cases resolves resulting in latent infection. Latent infection persists due to the formation of cysts, containing bradyzoites, in muscle, neurons and glia. It is capable of causing severe congenital neurological impairment in acquired *in utero* [1, 3]. Infection can reactivate from latent tissue cysts in patients with immune suppression, such as HIV infection, resulting in acute infections most often manifesting as encephalitis [4]. Reported *T. gondii* infection rates can be as high as 70%, depending on the population or geographic area studied[1]. *T. gondii* can infect all warm-blooded mammals although the definitive hosts are members of the cat family.

Glycosylation, both N-linked and O linked oligosaccharides, is one of the most common and important post-translation modifications seen in eukaryotic proteins [5–7]. This post translational modification can have significant effects on protein structure and function and is often a developmentally regulated process. Glycosylation pathways occur in the cytosol, endoplasmic reticulum, and the Golgi complex and involve transport steps, processing glycosidases, and glycosyltransferases[8, 9]. Glycosylation is often found on membrane and secreted proteins, being added onto these proteins during transport and synthesis in the ER (N-linked glycosylation) or during passage through the Golgi (O-linked glycosylation) Oligossacchrides greatly affect the physical properties and biological functions of many proteins playing critical roles in correct protein folding as well as cell-cell interactions.

Several studies had suggested that glycosylation was rarely seen in *T. gondii* proteins and this issue was considered controversial, but recent papers have clearly demonstrated glycoproteins in *T. gondii* RH strain tachyzoites [10–12]. Glycosylation has been demonstrated on proteins in the inner membrane complex associated with motility[10, 12]. Consistent with this observation, tunicamycin treated parasites have been demonstrated to have defects in invasion and motility [10, 12]. The tissue cyst wall (formed by modification of the parasitophorous vacuole by bradyzoites) of *T. gondii* has been demonstrated to react with the lectins *Dolichos biflorus* (DBA) and succinyl Wheat Germ Agglutinin (SWGA) [13]. Examination of the *T. gondii* genome (www.toxodB.org) demonstrates the presence of enzymes for the synthesis of both N-linked (dolichol-linked precursor oligosaccharides) and O-linked (UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferases) modifications[14, 15]. Cell free extracts of *T. gondii* have been demonstrated to have both N-glycosylation and O-glycosylation activity using synthetic peptide substrates [15–17]. In addition, studies using mass spectrometry have defined the presence of $Man_6(GlcNAc)_2$, $Man_7(GlcNAc)_2$, and $Man8(GlcNAc)_2$ N-glycans in *T. gondii* [12]. Interestingly, there was an absence of glycans containing sialic acid, galactose and fucose residues in this analysis [12].

The development of new preventative and therapeutic strategies for pathogens should relay on an improved understanding of the interactions between pathogens and their hosts. Surface proteins are a potential target of many compounds aimed at preventing microbial infections

and many of these proteins are glycosylated. Moreover, because surface proteins are likely to interact with the host immune system, they often become components of effective vaccines, many of which are based on glycoproteins. *T. gondii*, like other Apicomplexa, is surrounded by a triple membrane system, termed the pellicle, consisting of the plasmalemma and inner membrane complex. There are several approaches currently in practice to identify surface proteins. The first approach is based on surface protein prediction by genome analysis using algorithms such as PSORT [18]. The method is rapid but is not fully reliable nor is it quantitative. The second approach employs separation of membrane and cell wall fractions from the cytoplasmic fraction followed by the identification of proteins by 2D-gel electrophoresis or 2D-chromatography coupled to mass spectrometry [5]. This approach is reasonably quantitative. Subcellular extraction can simplify this type of analysis and enhance protein identification.

In this report we demonstrate a simple and efficient approach for determining the glycoproteome of *T. gondii.* A buffer containing detergent and a high salt concentration was utilized to extract surface and organelle glycoproteins using serial lectin affinity chromatography (SLAC). From this fraction, 132 glycoproteins were identified. Several methods were used to confirm the identification and the presence of carbohydrates on the surface or in cytoplasm of *T. gondii*.

2. Materials and methods

2.1 Host cells and parasites

Toxoplasma gondii RH strain was maintained by serial passage in confluent monolayers of human foreskin fibroblasts (HFF) incubated at 37°C under a 5% CO2 atmosphere in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS, Invitrogen, CA), 2mM glutamine and 20mM penicillin streptomycin. Cells were maintained as a contact-inhibited monolayer until required. Infection was maintained by inoculating flasks (T25 or 75) with freshly harvested tachyzoites. Parasites were purified by passing infected cells through a 25 gauge needle and then separated from host cell debris by passage through a 3μm Nucleopore membrane. Parasites were pelleted and then used for either protein or RNA purification. Total RNA was extracted using TRIzol (Invitrogen CA) as per the manufacturer's recommendations.

2.2 Extraction of T. gondii proteins

T. gondii (2×10^{10}) was suspended in 1ml PBS containing 0.5% Nonidt P40 and 500 mM NaCl and protease inhibitor cocktail (Complete Protease Cocktail, Roche Applied Science, Indianapolis, IN) and incubated for 2 hours at 4°C on ice. The tube was then centrifuged at 13600 × g in an Eppendorf microcentrifuge for 16 min and the supernatant collected for lectin affinity chromatography.

2.3. Serial lectin affinity chromatography (SLAC)

Spin columns of Agarose bound to Con A, WGA, or Jaclin (Vector Laboratories, Burlingame, CA) were equilibrated with buffer A (10 mM Tric Cl, pH 7.5, 150 mM NaCl, 1mM $CaCl_2$, 1mM $MnCl_2$). One ml of the *T. gondii* membrane extraction was first applied to the Con A column, this column was allowed to rotate at 4°C for 2 h and then the ConA column was then spun down (1000 × g for 1 min) and the pass through loaded into WGA column. The WGA column was then handled as noted for the Con A column with the pass-through being loaded onto the Jaclin column. Once the glycoproteins were bound to each column (i.e. ConA, WGA and Jaclin) the columns were washed extensively with buffer A until no proteins could be detected in washes. Elution was then carried out by incubation for 15 min with 200 mM α-methyl mannoside (α-MM) for the Con A column, 500 mM N-

acetyglucosamine (NAGlc) for the WGA column, and 800 mM galactose for the Jaclin column, respectively followed by centrifugation (100 × g for 10 min). The eluates were then used for SDS-PAGE and subsequent mass spectrometry. This workflow is outlined in Figure 1. Changing the order of the columns might change where in the purification a protein appeared, as a protein could bind to multiple lectins, however, one would expect a similar overall list of proteins if the columns were used in a different order.

2.4. SDS-PAGE and in-gel tryptic digestion

Protein eluates (30 μl per lane) were analyzed by electrophoreses in a 7.5 to 17.0 % gradient, 1 mm thick gel using a constant current of 400 mA for 1 h. The resolved proteins were visualized by GelCode Blue (Pierce, IL) and the corresponding bands were excised. The gel bands were destained with 60% acetonitrile solution and then the proteins in the gels were digested with trypsin[19]. Briefly, gel pieces were completely dried down in a vacuum centrifuge, rehydrated with a trypsin solution and allowed to incubate on ice for 45 min. After 45 min, the trypsin supernatant was removed and replaced with approximately 20 μl of digestion buffer without trypsin so that the gel pieces were covered. The gel pieces were kept wet at 37 °C overnight for digestion with mixing.

2.5. LC-MS Analysis

The labeled tryptic peptides were subjected to LC-MS/MS analysis on an LTQ mass spectrometer (Thermo Scientific, San Jose, CA). Chromatographic separation of peptides was performed on a nano HPLC System (LC Packings, San Francisco, CA, USA). The LC eluate from a 75 μm i.d. ×15 cm, PepMap C18 column (Dionex, Marlton, NJ) was directed to a micro-ionspray source. Throughout the LC gradient, MS and MS/MS data were recorded continuously using a 6-sec cycle time. With each cycle, MS data were accumulated for 1 s, followed by three CID acquisitions of 2.5 s each on ions selected by preset selection parameters of the data-dependant acquisition method. In general, the ions selected for CID were the 2 most abundant obtained from the survey MS spectrum, except that singly charged ions were excluded and dynamic exclusion was employed to prevent repetitive selection of the same ions within a preset time. Rolling collision energies were used to adjust automatically for the charge state and the mass/charge value of the precursor ion. Searches were performed using Mascot and a *T. gondii* protein database (http://toro.aecom.yu.edu/biodefense/) [19–21]. In all searches, oxidation of methionine, deamidation of asparagines and glutamine and conversion of glutamine to pyroglutamic acid (N-terminus) were selected as variable modifications and carboxyamidomethylation of cysteine residues was selected as a fixed modification, and a maximum of one missed cleavage. The peptide precursor mass tolerance was +/− 2 Da and the MS/MS product ions, +/− 0.8 Da. The data were also searched against a scrambled decoy database with false discovery rate of 1%.

2.6. Isolation of glycopeptides from intact T. gondii

*I*n other experiments, trypsin was utilized to cut the surface membrane proteins free from *T. gondii,* (2 hr at 37°C), the *T. gondii* were then removed by centrifugation (10,000 × g for 3 min). The supernatant containing the liberated glycoproteins was then removed and was allowed to complete its proteolysis until tryptic peptides were obtained (12 hr at 37°C). The supernatant was diluted in buffer A and Con A conjugated beads (lectin affinity chromatography) were used to capture these glycopeptides. The column was washed three times with buffer A, twice with buffer B (25 mM Tris, 10 mM NaCl, pH7.36) and then the bound glycopeptides were released from the beads by incubation for 2 hr at 37 °C in buffer B containing peptide N-glycosidase F (PNGase F) and the column washed with 1 volume of buffer B. The last wash and PNGase F fractions were pooled and then strong cation exchange chromatography (SCX beads) was used to fractionate the deglycosylated peptide

mixture into eight fractions by step-gradient elution using standard protocols. Nano-LC-MS/MS was then performed to analyze these eight fractions, and obtain amino acid sequences of the released glycopeptides.

2.7. Lectin blotting

For detection of lectin-binding sites on glycoproteins following SDS-PAGE and transfer to PFA membranes, the following HRP conjugated lectins (EY Laboratories) were used: Con A, high affinity for alpha-D-mannosyl residues; wheat germ agglutinin (WGA), affinity for D-N-acetylglucosamine; Jacalin, high affinity for N-acetylgalactosamine-beta-1,3-galactose. After blocking PFA membranes in 1% BSA/PBS containing 0.2 % Tween 20 for 4 h, they were incubated for 1 h at room temperature with lectins conjugated with HRP in PBS/0.2%BSA/0.2 Tween 20 at the following concentrations: 1 μg/mL Con A, 2 μg/mL WGA, 2μg/mL Jacalin, respectively. All incubations were accompanied by control incubations in the presence of corresponding inhibiting sugars: 0.4 M methyl-mannopyranose for Con A, 0.5 M N-acetylglucosamine for WGA, 0.8 M galactose for Jaclin. After rinsing using TBS/0.2% Tween 20, wash three times for 5 min in the same buffer, and reactive bands were detected according to the manufacturer's instructions (Amersham Biosciences, Part of GE Healthcare).

2.8. Lectin-fluorescence labeling

For surfacing labeling of parasites, freshly purified *T. gondii* tachyzoites were suspended in fixation buffer (PBS/3% paraformaldehyde/0.05% glutaraldehyde, pH 7.2) at 10^5 parasites/ml and were then applied to poly-lysine coated (100 μg/ml) glass coverslips for 30 min. After adherence onto coverslips specimens were rinsed in PBS and placed in blocking solution (PBS/1%BSA/50 mM glycine) for 4h. Lectins (ConA, WGA, Jacalin) conjugated with FITC were diluted 1:200 in PBS and incubated with the coverslips for 1 h at 37°C. Specimens were rinsed five times in PBS for 5 min each time and then briefly rinsed in distilled water. For a negative control slides were incubated with anti mouse IgG -FITC at 1:500 for 1 hour and then rinsed five times in PBS. For a positive control slides were incubated with a mAb to SAG1 (p30 a major *T. gondii* surface antigen) at 1:500 for 1 hour followed by washing with PBS, then incubation with secondary anti-mouse IgG-FITC at 1:500 for 1 hour and then washed 5 times in PBS. The preparations were then examined on a Leitz Laborlux S fluorescence microscope.

For the detection of intracellular lectin binding sites, parasites were fixed in PBS/3% paraformaldehyde, and after adherence onto coverslips rinsed in PBS and subsequently permeabilized using 0.1% Triton X100 for 30 min. After fixation the coverslips were rinsed three times in PBS and were placed into blocking solution (PBS/1%BSA/50 mM glycine), washed with PBS and incubated with 1:200 lectin conjugated with FITC in PBS for 1 h at 37°C. Specimens were then rinsed in five times in PBS for 5 min each time and then briefly rinsed in distilled water. The preparations were then examined on a Leitz Laborlux S fluorescence microscope.

2.9. Hydropathy calculations and transmembrane mapping

All identified proteins were analyzed using the ProtParam program (available at http://www.expasy.ch/sprot/sprot-top.html), which allows the calculation of the grand average of hydrophobicity (GRAVY) value for a given protein. The proteins exhibiting positive GRAVY values were recognized as a hydrophobic. All identified proteins were examined by the TMHMM Server (Prediction of transmembrane helices in proteins; http://www.cbs.dtu.dk/services/TMHMM). The analysis was performed to map TMDs [22].

2.10. Polymerase Chain Reaction (PCR)

A total RNA of 5ug was used for the synthesis of cDNA using AMV reverse transcriptase (TAKARA USA BIO) following the protocol supplied with the reagent. One ul of the resultant cDNA was used as a template in a 50ul PCR reaction using the fast start high fidelity PCR system (Roche Applied Science, IN). Hot start amplification was initiated with 2 min 95°C denaturation, followed by 30 cycles (95°C for 15sec, 53°C for 15 sec, and 72°C for 3 min) of amplification which was terminated by a 72°C extension for 10 min. Gene products were cloned into the PCR8GWTOPO vector (Invitrogen, CA) by TOPO cloning and sequenced at the Einstein Sequencing Facility.

The following oligonucleotide primer pairs were used for proteins identified by SLAC (the ID number of the corresponding gene from ToxodB (www.toxodb.org) is listed followed by the primers used (5′-3′) for RT-PCR for each gene): 55.m00103 (TCATTCACCCCGTTTTT TGTGACTCTTGG::CTGACGTGACGCTGCGTACTGACTG); 41.m01274 (TCGGA CTCGGTCGGTCGAAATGTGC::CTCGGACTCACAAGCCAGTGAATACGTCG), 76.m01589 (GTGTGCCTCAGGCACTGGTGGCTC::GCCTCGCTGCATCGTCTCTCGA), 67.m00007 (TGGCGCATCTGGAGATGCCGGCTG::TCGCCTTGCGGAAACGTGTACGTCC), 49.m03169 (CAGGCGATAGCGCGGGGGGACCGC:: TTCGCTTGGTCTCTGGTAGCCCAGCC), 80.m02161(AGGGCGATCGTGGCATCGACGCAGCAG::GTTGTGTTTGCTGCCTGCA GAGCC GCGCA), 46.m01601 (AAGCCACAAGTTTTGTTCGGTCTTC::AACGTATTTCTTC AAAAGGTTGTCAAGGGTGG), 42.m03584 (TCGTCATCCAGATTGGTACTCGTTTCC::CCCC GTGACGGGGAAGTACGCAGTCAGTTGA), 31.m00928 (AGGACGTCACGTCTCTTGTG TGCATTTGG::CAGCACTTGTTGCATTG CGATTCCAGAAGC), 55.m08219 (AAAGTGACCA CGAAAGGGCTTGCTTT TGC:: CATCCGATGTGAAGAAAGTTCGGTAGTTGG), 50.m00023 (GCTGAAACTGCTCTGTACTAC CAGG::GTATTTGAAGTTCGGAGGCAACACAAACGC), 76.m01543 (TGTGTGTGGTACGGACAGGCACG::TATATTGTCATCTTGCTCGCCAGGA CCTGAAGCG), 59.m03403 (ATCCACACGGGCCTGGATCTGC::GTGAGCAGGAGCACCG GCCGGT)

3. Results

3.1 Glycoprotein identification

T. gondii membrane proteins were separated by SLAC followed by SDS-PAGE (Figs.1 and 2). Among the proteins identified from *T. gondii* by SLAC as having glycoepitopes were 9 microneme proteins, 7 dense granule proteins, 15 rhoptry proteins, 17 surface proteins, 19 enzymes, 11 heat shock proteins, 20 other proteins, and 32 hypothetical proteins (Table 1 and Supplemental Table 1). This data has been deposited at EPICdB (http://toro.aecom.yu.edu/cgi-bin/biodefense/main.cgi) and ToxodB. (http://toxodb.org/toxo/) [20]. Of these 132 glycoproteins some bound only to con A; some to both WGA and con A, or con A and Jacalin, or WGA and Jacalin; and some to all three lectins (see Table 1 final column). The majority of these proteins were purified using the Con A column. Of interest, is that one of the 32 novel hypothetical proteins, protein (80.m02347), with a molecular weight of 608.3 kDa was predicted by TMHHD to have a single TM helix domain with the majority of its structure being outside the membrane on the surface of the parasite. Table 1 presents all of the SLAC identified proteins organized

according to presumed function. The majority of these proteins have predicted N-glycosylation sites (http://www.cbs.dtu.dk/services/NetNGlyc/). It was previously believed that glycosylation was rare in *T. gondii* proteins; however, our data provide evidence for a large number of previously unrecognized glycoproteins in this organism. Overall, this work demonstrates that SLAC combined with tandem MS is a powerful approach for glycoproteomic analysis in *T. gondii* and that glycosylation occurs in a significant number of the proteins.

3.2. Glycopeptide Isolation from Intact T. gondii

A total of 30 *T. gondii* glycoproteins were identified using an approach which employed trypsin to “shave” surface proteins off of *T. gondii* followed by completion of the proteolysis of the liberated proteins until tryptic peptides were obtained. The glycopeptides were then purified by Con A lectin affinity chromatography and the bound glycopeptides released from the affinity column by PNGase F and analyzed by mass spectrometry. As expected, most of the identified glyoproteins were putative membrane proteins based on BLAST similarities (Table 2), some, however, were enzymes and hypothetical proteins of unknown function and localization. All of the proteins had N-linked glycoprotein domains (indicated by bold type in Table 2) consistent with the presence of glycoepitopes in these proteins. This simplified approach of “protein shaving”, could prove to have significant utility for the identification of unknown or previously unsuspected *T. gondii* surface membrane proteins.

3.3. Detection of lectin binding in T. gondii using florescence microscopy

FITC-conjugated Con A, WGA, and Jacalin were used to examine the microscopic localization of these lectins in RH strain tachyzoites (i.e. a lectin IFA technique). Con A produced a bright and uniform staining on the entire surface of every tachyzoites (see Fig. 3A) as did WGA and Jacalin. When these organisms were permeablized with 0.1% Triton X-100 diffuse staining of the cytoplasm was seen. No specific structures (i.e. micronemes, rhoptries, dense granules, etc.) were uniquely stained by any of these lectins. This suggests that glycoepitopes binding these lectins are present in both membrane bound and cytoplasmic proteins (Fig. 3A).

3.4. Detection of glycoproteins using lectin-blotting

Further information on the presence of glycoproteins in *T. gondii* was obtained by employing HRP-conjugated lectins Con A, WGA, and Jacalin. Following SDS-PAGE and transfer to nitrocellulose, the same lectins used for fluorescence studies were used to probe the three different fractions obtained after NP-40 extraction of *T. gondii* tachyzoites (Fig. 3B). This lectin overlay procedure demonstrated multiple bands in the solubized membrane preparations. Binding of the lectins to these proteins could be blocked by the addition of the corresponding saccharide used for SLAC elution in the lectin-HRP binding solution. These results demonstrate that the membrane protein preparation contains glycoproteins with both mannose and N-acetylglucosamine modifications.

3.5. PCR validation

To test for the presence of expression of hypothetical proteins identified by mass spectrometric techniques, we chose 13 candidate proteins which were annotated as hypothetical proteins in the *T. gondii* genome database. Of these, 8 were eluted with α-methyl mannoside from the ConA column and 4 were identified in the galactose eluted fraction from the jacalin column. A forward and reverse primer was designed to the beginning and end of the target gene coding region as predicted in the *T. gondii* database. As shown in Fig. 4, out of 13 target genes there were specific PCR products of the predicted size for 12 of the candidate genes confirming their expression at the transcriptional level in

T. gondii RH strain validating the presence and active expression of the majority of the genes corresponding to the hypothetical proteins identified by mass spectrometry.

4. Discussion

The identification of an entire proteome, regardless of its origin, is a daunting task for several reasons. The dynamic range of current instrumentation is limited by fluctuating protein expression levels, which may span more than six orders of magnitude. In addition, the limited sensitivity and ability of contemporary proteomics to characterize proteins with high molecular masses, extreme isoelectric points, or extremes in hydrophobicity precludes complete coverage of a given proteome. One approach to circumvent these difficulties is to reduce sample complexity by fractionation. The enrichment of proteins from a selected part of proteome is expected to facilitate the identification of low abundance proteins and assist characterization of proteins related to specific organelles or structures. Fractionation of tachyzoites using the non-ionic detergent NP-40 and 500 mM NaCl demonstrated that the majority of glycoproteins were identified in the NP-40-soluble, hydrophobic, fraction, which is enriched for membrane proteins; however, it should be appreciated that this method may not detect all of the integral glycoproteins. In addition, while the method used should limit protein complex formation during SLAC it is still feasible that some of the identified proteins may have been purified due to their binding to glycoprotein(s) which bound to the columns.

In the current manuscript we present several lines of evidence that indicate the presence of a significant number of glycoproteins in *T. gondii* tachyzoites. This data confirmed that a significant number of proteins involved in invasion and motility are probably glycoproteins. This is consistent with observations from other investigators [10–13, 15, 16, 23] as well as the presence of the enzymatic machinery for both N-linked and O-linked glycosylation in the *T. gondii* genome (www.toxodB.org)[14]. Both lectin overlay (i.e. lectin blotting) and lectin IFA techniques confirmed the presence of glycoproteins and provided independent validation of the SLAC results. In addition, it has been previously reported that tunicamycin treatment of *T. gondii* decreased the number of proteins purified by ConA affinity chromatography [12]. Many of the glyoproteins that were identified were membrane-associated proteins. We identified several surface antigens (SAGs and SRS domain proteins) by SLAC as being probable glycoproteins. By microscopy, ConA, WGA and jacalin fluorescent conjugates labeled the surface of *T. gondii*.

T. gondii has several specialized organelles associated with invasion and its ability to establish a parasitophorous vacuole in which it replicates within its host cell [1, 3, 24]. The apical end of this parasite contains an elaborate cytoskeletal structure and regulated secretory organelles, the micronemes and rhoptries, which function in host cell invasion discharging their contents from the apical end of the parasite. Microneme proteins are released first, upon contact with the host cell and are thought to function in host cell and are thought to function in host cell recognition and attachment. The content of rhoptries are released next and may function in the formation of the parasitophorous vacule (PV). Another secretory vacuole, the dense granules, discharge from the apical, lateral, and posterior surfaces of the parasite. Dense granule proteins are released next during the formation of the parasitophorous vacuole. Several of the glycoproteins we identified were from these organelles, suggesting that this post translational modification may be important in the process of invasion and establishment of the parasitophorous vacuole. This is consistent with data on the ability of tunicamycin to inhibit invasion[10, 12]. In addition, many of the identified proteins in these structures had TMDs, suggesting that they were membrane associated. Microneme proteins (MIC) 1, 2, 3, 4, 6 were identified as glycoproteins and MIC 2 and MIC6 also had predicted TMDs. Fifteen rhoptry proteins

(ROPs and RONs) were identified as glycosylated and four have at least one TMD. This is consistent with a previous publication that also found evidence for glycosylation in RON2 and AMA [12], which are key components in the moving junction formed during cell invasion. Seven dense granule proteins (GRA) proteins were identified as glycoproteins and four GRAs have two TMD. Consistent with these findings, it had been previously demonstrated that GRA2 is O-glycosylated[25], Several cytoskeletal components were also identified as glycoproteins consistent with data that demonstrated that membrane anchor myosin XIV (GAP50) and TgMyoA were glycoproteins and involved in the effect of tunicamycin on invasion [10, 12]. These data are suggestive that glycoproteins are components of the glideosome. Our data confirmed that myosin, despite lacking an obvious ER motif, is consistently purified by lectin affinity chromatography. In other eukaryotes there are examples of proteins being glycosylated without such ER motifs [26, 27]. Overall, our data suggest that glycoproteins are key constituents in host-parasite interactions during invasion and establishment of the parasitophorous vacuole.

Heat shock- or stress induced activation of a set of heat shock protein genes is a characteristic of eukaryotic and prokaryotic cells. The heat shock proteins have been implicated as chaperons for protein folding and transport [28–30]. Heat shock proteins fall into several subfamilies, the low molecular weight (16–35 kDa) or small heat shock proteins (smHsps), the hsp60 family, the hsp70 family (68–78 kDa), and the high molecular weight (89–110 kDa) heat shock protein families (hsp90 and hsp100) [28, 29]. In recent years, it has become clear that heat shock proteins are not just restricted to stress responses, but are also regulated developmentally [28, 29]. In protozoan systems, a link between heat shock proteins and development and intracellular survival has been established [29]. We identified 11 heat shock proteins as possible glycoproteins. In examination of the identified protein gene predictions, there appear to be two different types of hsp90 one with TMD and one without TMD, five isotypes of hsp70 of which one has TMD and two hsp60s one with and one without a TMD. Consistent with our data Tomavo and his collaborators also have demonstrated that a number of heat shock proteins were present in their analysis of N-glycosylation in *T. gondii* [12].

Overall, this work clearly demonstrates that SLAC combined with tandem MS is a powerful approach for glycoproteomics in *T. gondii* and that glycosylation is not a rare modification but occurs in a significant number of the proteins of this important human and veterinary pathogen.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by NIH/NIAID grant AI31744 (LMW), NIH/NIAID contract HHSN266200400054C, NIH/NIAID R01AI087625 (KK) and and RC4AI092801 (KK)

References

1. Weiss, LM.; Kim, K. Toxoplasma gondi: The Model Apicomplexan: Perspectives and Methods. Elsevier (Academic Press); Amsterdam; Boston: 2007.
2. Tenter AM, Heckeroth AR, Weiss LM. *Toxoplasma gondii*: from animals to humans. Int J Parasitol. 2000; 30:1217–1258. [PubMed: 11113252]
3. Dubey JP. Advances in the life cycle of *Toxoplasma gondi*i. Int J Parasitol. 1998; 28:1019–1024. [PubMed: 9724872]

4. Wong SY, Remington JS. Biology of *Toxoplasma gondii*. AIDS. 1993; 7:299–316. [PubMed: 8471191]
5. Marino K, Bones J, Kattla JJ, Rudd PM. A systematic approach to protein glycosylation analysis: a path through the maze. Nat Chem Biol. 2010; 6:713–723. [PubMed: 20852609]
6. Vagin O, Kraut JA, Sachs G. Role of N-glycosylation in trafficking of apical membrane proteins in epithelia. Am J Physiol Renal Physiol. 2009; 296:F459–469. [PubMed: 18971212]
7. Tian E, Ten Hagen KG. Recent insights into the biological roles of mucin-type O-glycosylation. Glycoconj J. 2009; 26:325–334. [PubMed: 18695988]
8. Bill, RM.; Revers, L.; Wilson, IBH., editors. Protein glycosylation. Kluwer Academic Publishers; Boston: 1998.
9. Taylor, ME.; Drickamer, K., editors. Introduction to glycobiology. Oxford University Press; Oxford; New York: 2006.
10. Luk FC, Johnson TM, Beckers CJ. N-linked glycosylation of proteins in the protozoan parasite *Toxoplasma gondii*. Mol Biochem Parasitol. 2008; 157:169–178. [PubMed: 18096254]
11. Garenaux E, Shams-Eldin H, Chirat F, Bieker U, Schmidt J, Michalski JC, Cacan R, Guerardel Y, Schwarz RT. The dual origin of *Toxoplasma gondii* N-glycans. Biochemistry. 2008; 47:12270–12276. [PubMed: 18975916]
12. Fauquenoy S, Morelle W, Hovasse A, Bednarczyk A, Slomianny C, Schaeffer C, Van Dorsselaer A, Tomavo S. Proteomics and glycomics analyses of N-glycosylated structures involved in *Toxoplasma gondii*--host cell interactions. Mol Cell Proteomics. 2008; 7:891–910. [PubMed: 18187410]
13. Zhang YW, Halonen SK, Ma YF, Wittner M, Weiss LM. Initial characterization of CST1, a *Toxoplasma gondii* cyst wall glycoprotein. Infect Immun. 2001; 69:501–507. [PubMed: 11119543]
14. Samuelson J, Banerjee S, Magnelli P, Cui J, Kelleher DJ, Gilmore R, Robbins PW. The diversity of dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. Proc Natl Acad Sci U S A. 2005; 102:1548–1553. [PubMed: 15665075]
15. Stwora-Wojczyk MM, Kissinger JC, Spitalnik SL, Wojczyk BS. O-glycosylation in *Toxoplasma gondii*: identification and analysis of a family of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. Int J Parasitol. 2004; 34:309–322. [PubMed: 15003492]
16. Dieckmann-Schuppert A, Bause E, Schwarz RT. Glycosylation reactions in *Plasmodium falciparum*, *Toxoplasma gondii*, and *Trypanosoma brucei brucei* probed by the use of synthetic peptides. Biochim Biophys Acta. 1994; 1199:37–44. [PubMed: 8280751]
17. Stwora-Wojczyk MM, Dzierszinski F, Roos DS, Spitalnik SL, Wojczyk BS. Functional characterization of a novel *Toxoplasma gondii* glycosyltransferase: UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-T3. Arch Biochem Biophys. 2004; 426:231–240. [PubMed: 15158673]
18. Nakai K, Horton P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci. 1999; 24:34–36. [PubMed: 10087920]
19. Che FY, Madrid-Aliste C, Burd B, Zhang H, Nieves E, Kim K, Fiser A, Angeletti RH, Weiss LM. Comprehensive proteomic analysis of membrane proteins in *Toxoplasma gondii*. Mol Cell Proteomics. 2011; 10:M110.00745. [PubMed: 20935347]
20. Madrid-Aliste CJ, Dybas JM, Angeletti RH, Weiss LM, Kim K, Simon I, Fiser A. EPIC-DB: a proteomics database for studying Apicomplexan organisms. BMC Genomics. 2009; 10:38. [PubMed: 19159464]
21. Dybas JM, Madrid-Aliste CJ, Che FY, Nieves E, Rykunov D, Angeletti RH, Weiss LM, Kim K, Fiser A. Computational analysis and experimental validation of gene predictions in *Toxoplasma gondii*. PLoS One. 2008; 3:e3899. [PubMed: 19065262]
22. Chen Y, Yu P, Luo J, Jiang Y. Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT. Mamm Genome. 2003; 14:859–865. [PubMed: 14724739]
23. Odenthal-Schnittler M, Tomavo S, Becker D, Dubremetz JF, Schwarz RT. Evidence for N-linked glycosylation in *Toxoplasma gondii*, Biochem J. 1993; 291:713–721. [PubMed: 8489499]

24. Dubey JP, Lindsay DS, Speer CA. Structures of *Toxoplasma gondii* tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts. Clin Microbiol Rev. 1998; 11:267–299. [PubMed: 9564564]

25. Mercier C, Lecordier L, Darcy F, Deslee D, Murray A, Tourvieille B, Maes P, Capron A, Cesbron-Delauw MF. Molecular characterization of a dense granule antigen (Gra 2) associated with the network of the parasitophorous vacuole in *Toxoplasma gondii*. Mol Biochem Parasitol. 1993; 58:71–82. [PubMed: 8384696]

26. Blachly-Dyson E, Stevens TH. Yeast carboxypeptidase Y can be translocated and glycosylated without its amino-terminal signal sequence. J Cell Biol. 1987; 104:1183–1191. [PubMed: 3032983]

27. Silve S, Monod M, Hinnen A, Haguenauer-Tsapis R. The yeast acid phosphatase can enter the secretory pathway without its N-terminal signal sequence. Mol Cell Biol. 1987; 7:3306–3314. [PubMed: 3313013]

28. Tutar L, Tutar Y. Heat shock proteins; an overview. Curr Pharm Biotechnol. 2010; 11:216–222. [PubMed: 20170474]

29. Shonhai A. Plasmodial heat shock proteins: targets for chemotherapy. FEMS Immunol Med Microbiol. 2010; 58:61–74. [PubMed: 20041948]

30. Csermely, P.; Korcsmáros, T.; Sulyok, K., editors. Stress responses in biology and medicine: stress of life in molecules, cells, organisms, and psychosocial communities. Blackwell Pub. (New York Academy of Sciences); Boston, Mass: 2007.

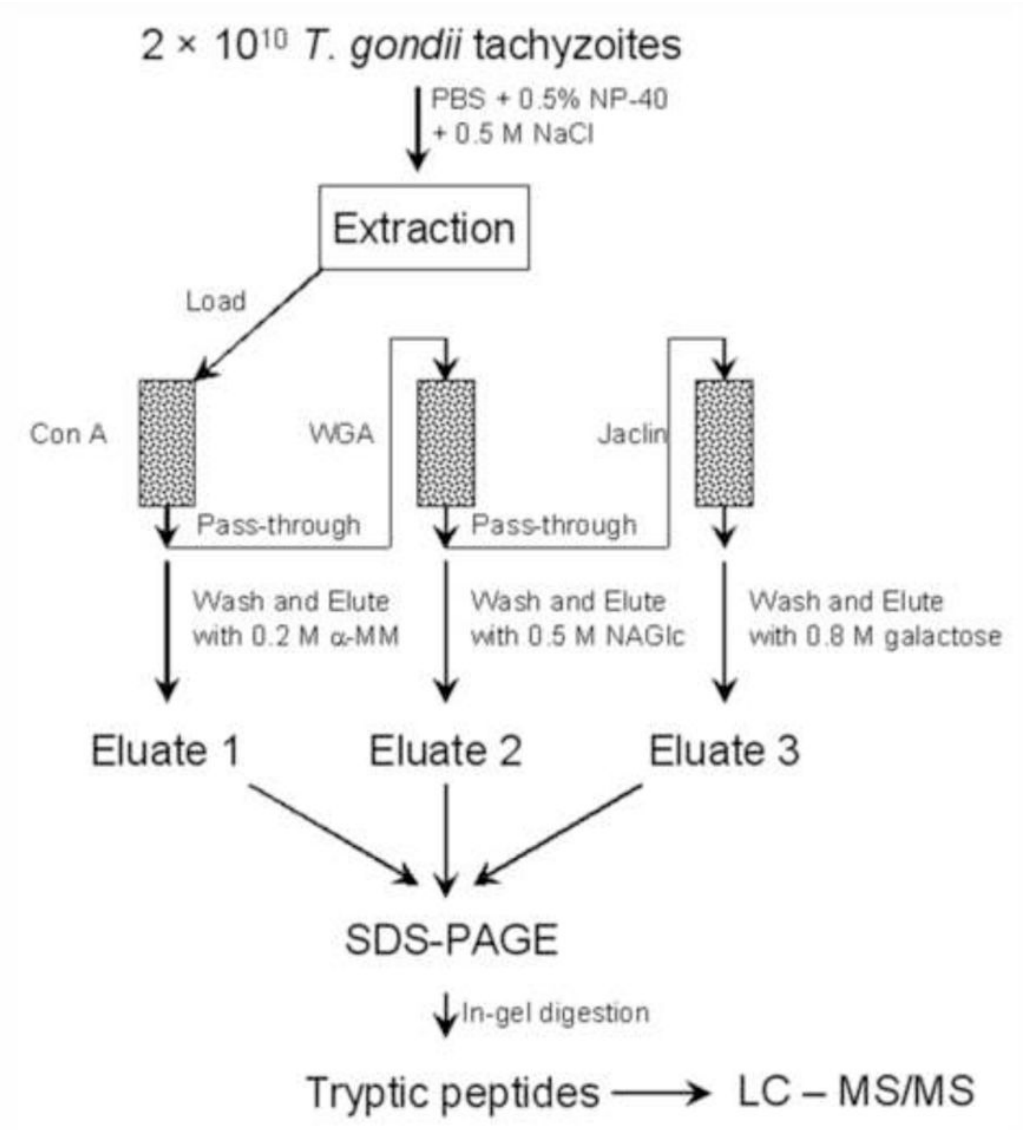

Figure 1.
Schema for glycoprotein enrichment for proteomic analysis of *T. gondii* by LC-MS/MS.

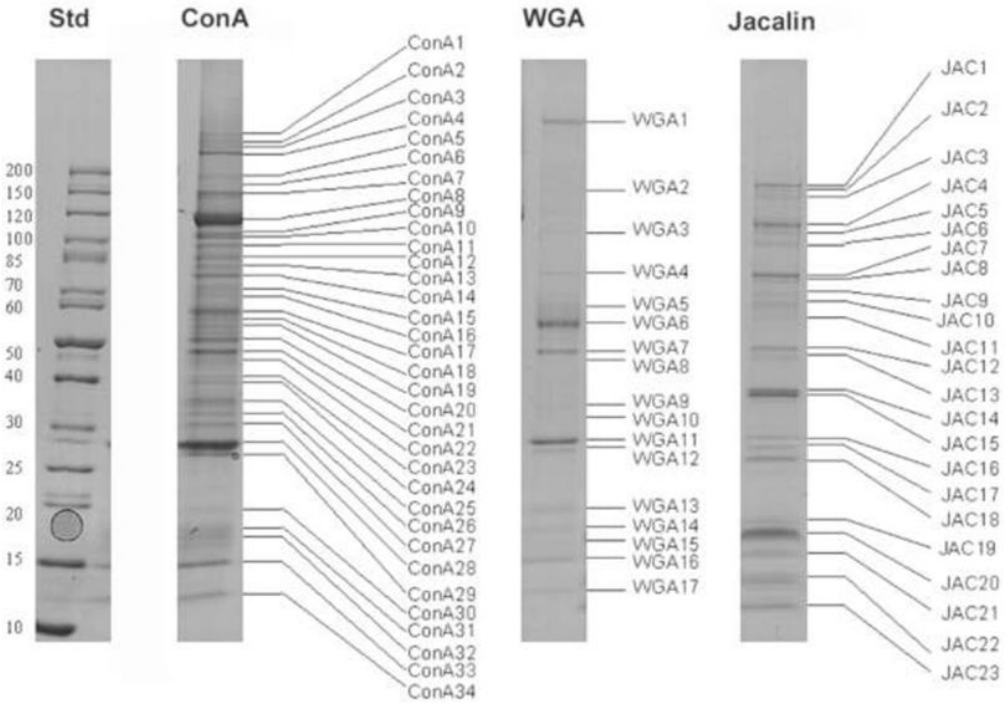

Figure 2. SDS-PAGE of *T. gondii* glycoproteins isolated by lectin affinity chromatography

The protein molecular weight standards are shown on the left. The gel band numbers correspond to proteins identified in Table S1 and Table 2. Std: molecular weight standard.

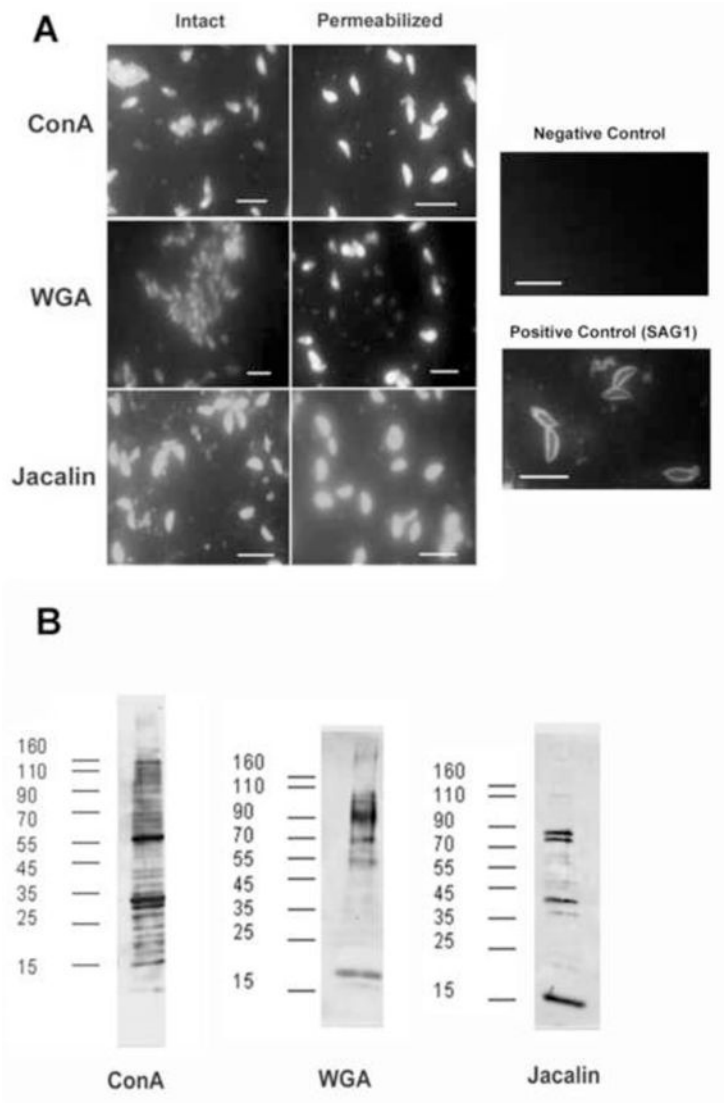

Figure 3. Lectin binding to *T. gondii*

A. Fluorescence microscopy of lectin-stained *T. gondii*. Con A-FITC, WGA-FITC, and Jacalin-FITC was employed to examine both fixed and permeabized (0.2% Triton X100) *T. gondii*. There was no change in staining with permabilization. All three lectins bound to the parasite consistent with the presence of glycoproteins. Negative control: antimouse-FITC, Positive control: mAb to SAG1 (p30) with secondary anti-mouse IgG-FITC. Bar length in each panel is 5 μm.

B. Lectin overlay with ConA,, WGA and Jacalin. Blots were probed with horseradish peroxidase conjugated Con A, WGA, and jacalin, respectively. Numerous bands were present demonstrating the presence of both N- and O-glycoroteins in *T. gondii* tachyzoites. No bands were seen when the sugars binding each lectin (e.g. α-methyl mannoside for ConA, N-acetylglucosamnine for WGA and galactose for jacalin) were added to the elution buffer.

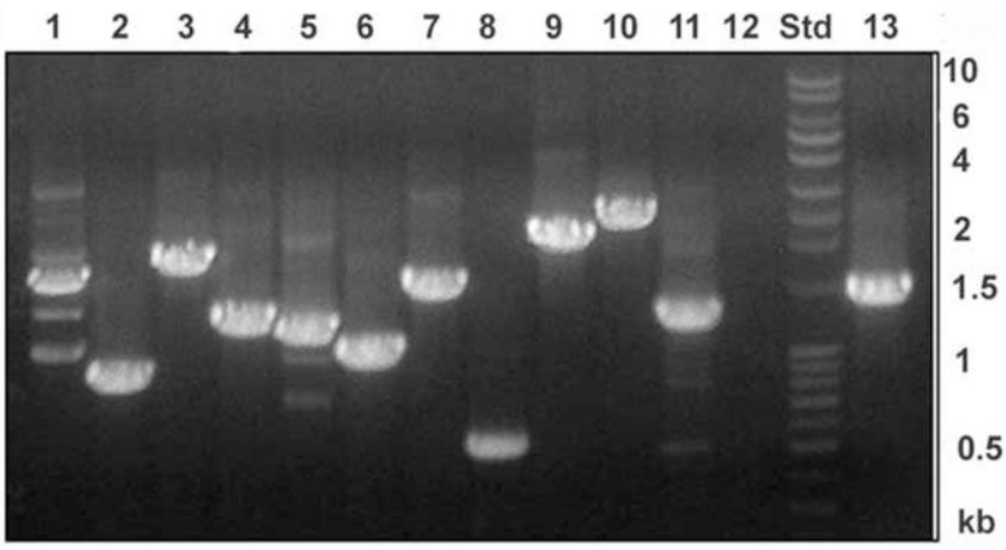

Figure 4. RT-PCR of the genes corresponding to hypothetical glycoproteins identified by mass spectrometry

Lanes 1–8 and lane 13 are the gene products in which the corresponding protein was eluted from Con A column by α methyl mannoside and the lanes 9–12 belongs to gene products corresponds to proteins eluted from jacalin column by galactose. Lane 1: 55.m00103, Lane 2: 41.m01274, Lane 3: 76.m01589, Lane 4: 67.m00007, Lane 5: 49.m03169, Lane 6: 80.m02161, Lane 7: 46.m01601, Lane 8: 42.m03584, Lane 9: 31.m00928, Lane 10: 55.m08219, Lane 11: 50.m00023, lane 12:76.m01543, Lane 13: 59.m03403. Std: DNA ladder.

Table 1

The glycoproteins identified from *T. gondii* using lectin affinity chromatography.

gene id	Release 6 Name	Protein Name	Percent Coverage	MOWSE	Mw (kDA)	Gel Band
MICRONEME PROTEINS						
NR-2062142	TGME49_091890	microneme TgMIC1	46	4723	48.6	ConA21 WGA7 JAC12
20.m00002	TGME49_001780	microneme protein, putative TgMIC2	11	49	124.8	ConA12
641.m00002	TGME49_119560	MIC3 microneme protein TgMIC3	26	400	37.9	ConA13
25.m00006	TGME49_008030	micronemal protein 4 TgMIC4	23	303	63	ConA17 WGA5 JAC9
NR-4704627	TGME49_018520	microneme protein, putative TgMIC6	19	50	36.6	ConA29 WGA12 JAC17
NR-118500931	TGME49_004130	perforin-like protein 1; PLP1	44	5922	124.6	ConA8
20.m03849	TGME49_004130	membrane-attack complex/perforin domain-containing protein; PLP1	18	95	117.3	ConA16
55.m00005	TGME49_055260	apical membrane antigen 1, putative AMA1	38	112	63	ConA16 JAC9
20.m00387	TGME49_004050	subtilase family serine protease, putative TgSUB1	5	42	132.6	JAC1
DENSE GRANULE PROTEINS						
42.m00015	TGME49_027620	28kd antigen GRA2 (or p28)	30	313	19.8	JAC18
NR-161913	TGME49_027620	28kd antigen; GRA2 (or p28)	20	70	28	JAC22
NR-22652337	TGME49_027280	dense granule protein, putative GRA3	17	396	24.2	ConA29
42.m00013	TGME49_027280	Dense granule protein 3 GRA3	16	61	22.2	WGA12
63.m00002	TGME49_075440	dense granule antigen GRA6	27	49	24	ConA27
20.m00005	TGME49_003310	granule antigen protein GRA7	86	701	25.8	ConA28
NR-2062409	TGME49_003310	29kD excretory dense granule protein GRA7	34	88	25.8	ConA29 JAC16
RHOPTRY PROTEINS						
NR-897823	TGME49_109590	rhoptry protein ROP1	18	1323	42.6	ConA20
NR-563627	TGME49_015780	rhoptry protein 2 ROP 2 (ROP2A)	32	1585	64	ConA19
NR-52788873	TGME49_095110	rhoptry protein 4 ROP4 (ROP4/7 locus)	28	289	64	ConA17 JAC11
NR-134035971	TGME49_108080	rhoptry protein 5 ROP5	13	80	60.8	ConA22 JAC10

gene id	Release 6 Name	Protein Name	Percent Coverage	MOWSE	Mw (kDA)	Gel Band
551.m00238	TGME49_108080	rhoptry protein 5 ROP5	28	2991	60.9	ConA17 WGA5
83.m02145	TGME49_095110	rhoptry protein ROP7 (ROP4/7 locus)	16	415	103.8	ConA18
42.m03584	TGME49_027810	rhoptry protein 11 ROP11	7	48	58	ConA21
NR-71559154	TGME49_112270	rhoptry protein 13 ROP13	9	51	44.8	ConA25
55.m08219	TGME49_062730	rhoptry protein 16 ROP16	17	1271	76.2	JAC13
NR-84618297	TGME49_005250	rhoptry protein 18 ROP18	21	276	62.3	ConA19
583.m00597	TGME49_110010	rhoptry neck protein 1 TgRON1	8	43	118.8	ConA13
NR-71559160	TGME49_100100	rhoptry neck protein 2 TgRON2	31	389	155.5	ConA7
583.m00636	TGME49_111470	rhoptry neck protein 5 TgRON5	27	190	56.2	ConA22 JAC13
583.m00011	TGME49_114500	subtilisin-like protease TgSUB2 TgSUB2	9	57	141.5	ConA8 JAC7
33.m02185	TGME49_014080	toxofilin	21	83	27.1	JAC17
SURFACE PROTEINS						
76.m01626	TGME49_085870	SRS domain-containing surface antigen, putative; SRS20A	68	459	34.9	ConA27
44.m00008	TGME49_033450	GPI-anchored surface protein SRS29A (or SRS1)	21	100	44.2	ConA22
NR-123186979	TGME49_033460	major surface antigen 1 SRS29B (SAG1 or p30)	25	55	32.5	ConA26
NR-50082488	TGME49_033460	unnamed protein product SRS29B (SAG1 or p30)	66	701	26.7	ConA27
NR-22219177	TGME49_033460	unnamed protein product; SRS29B (SAG1 or p30)	86	29783	29.8	ConA28
NR-129348	TGME49_033460	major surface antigen P30 precursor; SRS29B (SAG1 or p30)	55	2728	26.7	WGA11 JAC16
NR-10723	TGME49_033460	major surface antigen P30 precursor; SRS29B (SAG1 or p30)	69	6669	34.8	ConA29
44.m00010	TGME49_033480	SRS domain-containing protein SRS29C (or SRS2 or p35)	16	105	39.1	ConA25
NR-2305260	TGME49_033480	SAG1-related sequence 2 SRS29C (or SRS2 or p35)	20	367	39.3	ConA26
NR-5901701	TGME49_033480	P35 surface protein, putative; SRS29C (or SRS2 or p35)	11	54	28.5	ConA26 WGA9
NR-161926	TGME49_071050	surface antigen P22; SRS34A (SAG2A or p22)	48	3605	19	ConA31 JAC19
57.m01840	TGME49_067140	SRS domain-containing protein; SRS38B	7	68	40.9	ConA26
583.m00001	TGME49_108840	conserved hypothetical protein; SRS51 (or SRS3)	19	270	38	ConA28
583.m05672	TGME49_115320	SRS domain-containing protein; SRS52A	33	47	34.2	ConA28
NR-13447088	TGME49_108010	surface protein, putative	40	844	41.7	ConA25
641.m01520	TGME49_119550	transmembrane protein, putative	14	58	76.1	ConA14

gene id	Release 6 Name	Protein Name	Percent Coverage	MOWSE	Mw (kDA)	Gel Band
129.m00256	TGME49_099110	cleft lip and palate transmembrane protein 1, putative	9	65	68	ConA15
ENZYMES						
50.m03211	TGME49_048160	ATP-dependent DNA helicase II, 70 kDa subunit, putative; Ku70	1, 8	304	94.5	ConA4
52.m01559	TGME49_053030	alpha-glucosidase II, putative	18	140	183.2	ConA4
42.m00006	TGME49_028170	serine/threonine protein phosphatase, putative	25	218	170.7	ConA5 JAC2
113.m00789	TGME49_097650	serine/threonine protein phosphatase, putative	3	50	69	ConA15
NR-133990372	TGME49_028170	serine/threonine protein phosphatase, putative	40	2048	168.2	ConA7
44.m02735	TGME49_032600	patatin-like phospholipase domain-containing protein	8	56	70.8	ConA8 JAC5
583.m05329	TGME49_110080	long-chain-fatty-acid-CoA ligase, putative	10	64	138.6	ConA12
38.m01100	TGME49_019130	glutathione reductase, putative	24	58	64	ConA14
80.m02253	TGME49_089940	uroporphyrinogen decarboxylase, putative	4	50	56.2	ConA17
33.m00007	TGME49_015260	carbamoyl phosphate synthetase II	4	44	186.9	ConA17
83.m00004	TGME49_094200	putative glucose-6-phosphate-1-dehydrogenase	23	80	62.7	ConA18 WGA6
41.m01273	TGME49_020940	ribosomal RNA large subunit methyltransferase J, putative	3	68	82.2	ConA18
27.m00003	TGME49_011680	putative protein disulfide isomerase	33	97	52.8	ConA19 WGA6
55.m04808	TGME49_059010	vacuolar ATP synthase subunit D, putative	33	106	44.9	ConA22
20.m03680	TGME49_001840	eukaryotic aspartyl protease, putative; TgASP1	23	153	66.9	ConA27
28.m00308	TGME49_012310	vacuolar ATP synthase 16 kDa proteolipid subunit, putative	16	103	17.5	ConA32
57.m01692	TGME49_064650	SRS domain-containing, N-acetylglucosamine-phosphate mutase, putative	29	3655	324.2	WGA1
TgTigrScan_3843	TGME49_078850	glucose-6-phosphate dehydrogenase, putative	12	196	197.8	JAC5
44.m02735	TGME49_032600	patatin-like phospholipase domain-containing protein	12	81	70.8	JAC5
HEAT SHOCK PROTEINS						
49.m00060	TGME49_044560	heat shock protein 90, putative; HSP90	23	70	99.1	ConA1 WGA3 JAC5
80.m00001	TGME49_088380	heat shock protein 90; HSP90	9	43	89.5	ConA3
583.m00009	TGME49_111720	heat shock protein 70, putative; HSP70	16	301	73.2	ConA9 WGA4 JAC4
59.m00003	TGME49_073760	heat shock protein 70, putative; HSP70	22	508	72.8	ConA14 WGA4 JAC8

gene id	Release 6 Name	Protein Name	Percent Coverage	MOWSE	Mw (kDA)	Gel Band
NR-3323502	TGME49_073760	heat shock protein, putative; HSP70	23	508	72.3	ConA14 JAC7
50.m00085	TGME49_051780	heat shock protein 70kD, putative; HSP70	17	43	70.6	ConA14 WGA4 JAC8
NR-12248795	TGME49_073760	hsp70; HSP70	39	421	73.4	ConA15
42.m03533	TGME49_026830	DnaK family protein; HSP70	33	617	101.6	ConA7
49.m00030	TGME49_040600	TCP-1/cpn60 chaperonin family protein, putative	30	84	72.2	ConA15
50.m00006	TGME49_047550	heat shock protein, putative; HSP60	11	86	60.9	ConA17 WGA5 JAC9
55.m00016	TGME49_058390	DnaJ protein, putative	19	61	44.6	ConA23
OTHERS						
NR-2209250	TGME49_035470	myosin A, putative; TgMyoA	30	379	93.2	ConA11
25.m00007	TGME49_009030	actin	39	224	41.9	JAC13
59.m00006	TGME49_070240	cyst matrix protein	12	48	62.4	ConA15
46.m01699	TGME49_036540	RRM domain-containing protein	21	377	65.7	ConA18
76.m00016	TGME49_094800	elongation factor 1-alpha, putative	23	304	49	ConA20
50.m05680	TGME49_050770	eukaryotic translation initiation factor 4A	20	48	46.6	ConA20
50.m00069	TGME49_049270	thioredoxin, putative	18	148	46.9	ConA22
42.m00047	TGME49_025800	ABC transporter, putative	35	46	35.6	ConA24
38.m00002	TGME49_018410	60S acidic ribosomal protein P0	28	111	34.1	ConA25
64.m00002	TGME49_076140	ADP-ribosylation factor, arf, putative	37	147	21	ConA30
49.m00025	TGME49_039890	SCP-like extracellular domain-containing protein	15	178	26.6	ConA31
55.m00221	TGME49_063700	40s ribosomal protein S14, putative	25	91	16.3	ConA32, JAC20
541.m01147	TGME49_105010	RNA binding protein, putative	24	84	24.4	ConA32
583.m00614	TGME49_109740	small nuclear ribonucleoprotein, putative	17	43	28.2	ConA33
59.m03439	TGME49_069180	crooked neck-like protein 1, putative	6	53	93.7	WGA6
76.m01689	TGME49_087170	kinesin motor domain-containing protein, putative	14	49	142.3	WGA7
42.m00069	TGME49_026410	elongation factor 1-beta, putative	28	64	36	WGA10
641.m01588	TGME49_120600	glycine-rich protein 2, putative	30	323	23	WGA15
80.m02347	TGME49_092020	cysteine repeat modular protein, putative	6, 7	593	608.3	ConA1
641.m01513	TGME49_119340	kelch motif domain-containing protein	18	77	44	JAC15

gene id	Release 6 Name	Protein Name	Percent Coverage	MOWSE	Mw (kDA)	Gel Band
HYPOTHETICAL PROTEINS						
20.m03804	TGME49_003520	hypothetical protein	6	43	256.2	ConA4
42.m00026	TGME49_023920	hypothetical protein	14	61	214.7	ConA6
50.m07132	TGME49_047210	hypothetical protein, conserved	7	57	50.7	ConA6
55.m00103	TGME49_058870	hypothetical protein	17	94	88.1	ConA9
80.m02216	TGME49_089500	hypothetical protein	34	2170	100.2	ConA11
20.m00331	TGME49_002200	hypothetical protein	22	84	80.1	ConA13
583.m05446	TGME49_111710	hypothetical protein	22	345	133	ConA15
49.m00054	TGME49_043930	hypothetical protein	12	1155	101.3	ConA16 JAC8
41.m01274	TGME49_020950	hypothetical protein	16	146	35.4	ConA17
76.m01589	TGME49_085180	hypothetical protein	22	158	66.1	ConA18
41.m01337	TGME49_022100	hypothetical protein	11	82	128.2	ConA19
67.m00007	TGME49_079100	hypothetical protein	38	214	47.7	ConA20
49.m03169	TGME49_039740	hypothetical protein	20	63	44.7	ConA22
80.m02161	TGME49_088650	hypothetical protein	43	1289	47.8	ConA23
583.m05736	TGME49_116250	hypothetical protein	32	308	44.7	ConA24
20.m08222	TGME49_003990	hypothetical protein, conserved	17	91	25.5	ConA24
80.m02287	TGME49_090730	hypothetical protein, conserved	48	1777	37.2	ConA25
46.m01601	TGME49_034380	hypothetical protein	17	144	39.4	ConA26
31.m00928	TGME49_013280	hypothetical protein	25	359	20.8	ConA30
41.m00025	TGME49_022880	hypothetical protein	3	47	27.5	ConA31
583.m05298	TGME49_109600	hypothetical protein	4	45	106.5	WGA3
NR-95007087	TGME49_095100	hypothetical protein	5	43	152.2	WGA10
583.m00652	TGME49_112420	hypothetical protein	5	57	85.9	WGA12 JAC13
69.m00143	TGME49_079420	hypothetical protein	33	2572	144.6	JAC1
42.m03456	TGME49_025860	hypothetical protein	15	604	102.4	JAC3
50.m00023	TGME49_045610	hypothetical protein, conserved	20	378	79.1	JAC13
76.m01543	TGME49_083540	hypothetical protein, conserved	34	2959	47.1	JAC15
59.m03403	TGME49_068760	hypothetical protein	18	320	32	JAC17
583.m05643	TGME49_114840	hypothetical protein, conserved	9	131	228.8	ConA3

gene id	Release 6 Name	Protein Name	Percent Coverage	MOWSE	Mw (kDA)	Gel Band
50.m03374	TGME49_050820	hypothetical protein, conserved	6	52	259.3	ConA3 WGA4
31.m00856	TGME49_012790	hypothetical protein	6	268	230.5	JAC3
NR-35187725	TGME49_015980	hypothetical protein, conserved	10	339	20.2	WGA14

Table 2

The glycoproteins identified from T. gondii using “membrane shaving” PNGF technique

EPICdB ID	ToxodB version 6.0	Identified Peptide	Protein Name
TgGlmHMM_0006	TGME49_037860	QM**NISL**AAA SAFR	DNA polymerase I, putative
TgTigrScan_4836	TGME49_118470	AGNVGL**NLT**R	DNA repair protein, putative
TgTigrScan_1309	TGME49_117700	**NQT**TASSAQLR	enoyl-CoA hydratase, putative
TgGlmHMM_1305	TGME49_017680	GV**NVT**IDR	hypothetical protein, conserved
TgGlmHMM_0380	TGME49_097110	**NLT**NVYMNAFAGTQPSR	kinesin motor domain containing protein
541.m01146	TGME49_104990	AHAQSV**NAT**STLPQ	HSEK guanylate binding protein, putative
49.m00054	TGME49_043930	TMNSEGVISDGLQSQLPV**NHT**R	hypothetical protein TGME49_043930
TgTigrScan_5612	TGME49_014560	AIGE**NGS**CMHKSTSPAR	hypothetical protein TGME49_014560
TgTigrScan_5939	TGME49_019320	**NYT**SEALR	GAP50 membrane anchor for myosin XIV
TgGlmHMM_4009	TGME49_058770	QRE**NGS**GSSQLPGAANGR	UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase T2
TgTigrScan_7263	TGME49_005410	MEALS**NES**RSSFHDVCDETAQK	RNA pseudouridylate synthase, putative
TgGlmHMM_2200	TGME49_112480	MEA**NSS**SSSLPSVQK	uracil phosphoribosyltransferase
129.m00256	TGME49_099110	N**NTT**LYVHVR	cleft lip and palate transmembrane protein 1, putative
TgTigrScan_4899	TGME49_000250	CWKTSKCVFMoHFNNDGCTLSGI**NAT**AQTDANSK	PAN domain-containing protein
TgGlmHMM_0029	TGME49_087480	**NAS**PSATSGLLKQLK	hypothetical protein TGME49_087480
TgTigrScan_5651	TGME49_110750	G**NET**ALMPK	hypothetical protein, conserved
TgTigrScan_0391	TGME49_042890	E**NQS**ANACIENNR	hypothetical protein TGME49_042890
TgTigrScan_5004	|TGME49_082030	EQIVEK**NDT**LELHDR	hypothetical protein TGME49_082030
57.m03967	TGME49_066660	**NLT**ARQEGLSK	hypothetical protein TGME49_066660
TgTigrScan_1934	TGME49_047260	GGAQ**NAS**EAIRTESDK	retinoblastoma-binding protein, putative
TgTigrScan_7907	TGME49_115760	TNTEG**NAT**PVDSQSSPPSK	hypothetical protein, conserved
TgGlmHMM_0680	TGGT1_018880	**NLS**TTPSAVQTEER	conserved hypothetical protein
TgTigrScan_4198	TGME49_002980	LATQQVTAQTS**NVS**QALGDRK	hypothetical protein TGME49_002980
TgTwinScan_3338	TGME49_080730	**NRS**LCAGGAAAEAAVAQK	nucleotide-binding protein, putative
TgTigrScan_4829	TGME49_118370	AMNPSSPPHPNHPVD**NVS**Q	hypothetical protein TGME49_118370
TgTigrScan_2810	TGME49_066670	LASQSMPTDAE**NTS**FTLQGGSVGMGLGGRER	hypothetical protein TGME49_066670
TgGlmHMM_0694	TGME49_002540	TEQGTALLTGAPPSANELEAASAMANPT‘“**NSS**”’R	3′,5′-cyclic nucleotide phosphodiesterase, putative
TgTigrScan_5174	TGME49_079340	NQQETEM**NGS**PHNAAR	hypothetical protein TGME49_079340

EPICdB ID	ToxodB version 6.0	Identified Peptide	Protein Name
TgTigrScan_4790	TGME49_120590	EME**NVS**GDARAGGIGASDK	glycosyl hydrolases family 35 domain-containing protein
TgTigrScan_7969	TGME49_095710	SPTSQGN**NAS**AVCRASRAPSGQEAEMR	ubiquitin-transferase domain-containing protein