Host cell invasion is an essential step during infection by *Toxoplasma gondii*, an intracellular protozoan that causes the severe opportunistic disease toxoplasmosis in humans. Recent evidence strongly suggests that proteins discharged from *Toxoplasma* apical secretory organelles (micronemes, dense granules, and rhoptries) play key roles in host cell invasion and survival during infection. However, to date, only a limited number of secretory proteins have been discovered, and the full spectrum of effector molecules involved in parasite invasion and survival remains unknown. To address these issues, we analyzed a large cohort of freely released *Toxoplasma* secretory proteins by using two complementary methodologies, two-dimensional electrophoresis/mass spectrometry and liquid chromatography/electrospray ionization-tandem mass spectrometry (MudPIT, shotgun proteomics). Visualization of *Toxoplasma* secretory products by two-dimensional electrophoresis revealed ~100 spots, most of which were successfully identified by protein microsequencing or matrix-assisted laser desorption ionization-mass spectrometry analysis. Many proteins were present in multiple species suggesting they are subjected to substantial post-translational modification. Shotgun proteomic analysis of the secretory fraction revealed several additional products, including novel putative adhesive proteins, proteases, and hypothetical secretory proteins similar to products expressed by other related parasites including *Plasmodium*, the etiologic agent of malaria. A subset of novel proteins were re-expressed as fusions to yellow fluorescent protein, and this initial screen revealed shared and distinct localizations within secretory compartments of *T. gondii* tachyzoites. These findings provided a uniquely broad view of *Toxoplasma* secretory proteins that participate in parasite survival and pathogenesis during infection.

The protozoan *Toxoplasma gondii* is a human pathogen that causes severe opportunistic disease (toxoplasmosis) in congenitally infected babies and immunocompromised individuals (e.g. AIDS) (1, 2). Although foodborne transmission via ingestion of infected meat products contributes to the steady rise in age-dependent seroprevalence, waterborne transmission of the highly infectious, feline-derived oocyst stage has led to recent outbreaks (3, 4). The efficiency of waterborne transmission and the availability of drug-resistant strains have raised awareness of toxoplasmosis as a threat to public health from natural outbreaks or the potential malicious contamination of public water sources. Because effective control of outbreak situations will be enhanced by having multiple options for the diagnosis and treatment, the identification and exploitation of novel parasite targets is of acute importance.

* T. gondii belongs to phylum Apicomplexa, which also includes several other notable pathogens such as *Plasmodium* (the agents of malaria), *Cryptosporidium* (the cause of cryptosporidiosis), and *Eimeria* (the cause of coccidiosis). For several aspects of apicomplexan biology, *Toxoplasma* is emerging as an important model organism (5) because it exhibits many of the features and capabilities that define the phylum, yet it is more amenable to experimental manipulation than many of its kin.

*Toxoplasma* is an obligate intracellular parasite and must invade a vertebrate host cell for survival and replication. Invasion initiates the lytic cycle leading to the cell and tissue destruction that is a hallmark feature of *Toxoplasma* pathology. *Toxoplasma* invasion is a rapid (<30 s), dynamic, and complex process that relies on the secretion of numerous secretory proteins from specialized secretory organelles, including micronemes, rhoptries, and dense granules (6). Previous studies revealed that sequential secretions of these secretory proteins are critical events in parasite invasion and establishment of infection (7, 8). Micronemal proteins (MICs)3 are released first upon the parasite apical attachment to a host cell and function in host cell attachment and penetration (reviewed in Ref. 9). Next, the contents of rhoptries are discharged and are thought to be critical for biogenesis of the parasitophorous vacuole (PV) that envelopes the parasite during invasion and its interaction with host cellular organelles (10, 11). Finally, dense granules proteins (DGs) are exocytosed both during and after invasion and are thought to function in intracellular survival and replication (12–14). Recent studies have also suggested that the parasite releases products capable of manipulating the host immune response (15, 16). Also, secretory products and surface antigens (SAGs) have received scrutiny as promising diagnostic markers (17, 18). Despite this progress, only a limited number of secretory proteins have been discovered to date. This point is underscored by the recent bioinformatic identification of >800 genes encoding proteins with a putative secretory signal peptide (www.toxodb.org), yet fewer than 30 *Toxoplasma* secretory proteins have been described to date. Although many of the putative secretory proteins may be retained in internal compartments such as the ER/Golgi or

---

3 The abbreviations used are: MIC, microneme; MZAP, MIC2-associated protein; DG, dense granule; ROP, rhoptry; SUB, subtilisin-like protease; PDI, protein-disulfide isomerase; SAG, surface antigen; Cyp, cyclophilin; ESA, excretory/secretory antigen; PMF, peptide mass fingerprint; MudPIT, multidimensional protein identification technology; MALDI, matrix-assisted laser desorption ionization time-of-flight; GPL, glyco-sphingomyelin; MS, mass spectrometry; HPLC, high pressure liquid chromatography; LC, liquid chromatography; ESI, electrospray ionization; YFP, yellow fluorescent protein; CHAPS, 3-[[(cholamidopropyl)dimethylammonio]-1-propanesulfonic acid; IEF, isoelectric focusing; PBS, phosphate-buffered saline; PV, parasitophorous vacuole; ER, endoplasmic reticulum; Xcorr, cross-correlation number.
Toxoplasma Invasion and Survival

apicoplast (a plastid like organelle), a significant fraction is probably exported to external sites for interaction with the host.

The recent coupling of high resolution protein separation techniques (two-dimensional electrophoresis or multichromatography) with high throughput identification strategies has permitted the wide scale analysis of protein identity, expression, modification, localization, and interactions within whole cells or subcellular fractions (19, 20). With these tools, it is possible to analyze a multitude of cellular proteins from a more global perspective. The recent completion of parasite genome projects (21–23) has fueled proteomic studies of Plasmodium (24, 25), Eimeria tenella (26), Trypanosoma cruzi (27), and Leishmania major (28). The Toxoplasma genome is nearing completion, and two studies have described preliminary mapping and partial identification of the tachyzoite (a rapidly dividing form of the parasite) proteins, either by two-dimensional electrophoresis immunoblotting (29) or by two-dimensional electrophoresis coupled with MALDI-MS (30). Although these studies provided a valuable initial view of the protein complement for the parasite, a relatively small subset of the Toxoplasma proteome was examined.

Here we use complementary approaches to examine a cohort of excreted/secreted antigens (ESA) freely released by Toxoplasma tachyzoites, the life stage responsible for the tissue pathology observed in toxoplasmosis. Our findings reveal that Toxoplasma mobilizes a rich assortment of putative effector molecules that may interact with the host during infection. These include adhesins that presumably contribute to attachment and penetration, proteases that may facilitate invasion or migration into host tissues, metabolic enzymes potentially involved in nutrient acquisition, and a series of hypothetical proteins of unknown function. These findings significantly widen the viewing aperture of the molecular players contributing to Toxoplasma survival and pathogenesis during infection.

EXPERIMENTAL PROCEDURES

Reagents and Chemicals—Porcine trypsin (sequencing grade) was from Promega (Madison, WI). MS calibration mixture II (peptide mass standard kit) was from Applied Biosystems (Framingham, MA). One-dimensional electrophoresis and two-dimensional electrophoresis reagents and immobilized IEF strips were from Bio-Rad or Amersham Biosciences. The MALDI-MS matrix, a-cyano-4-hydroxycinnamic acid, was purchased from Sigma. All the organic solvents are HPLC grade. All other reagents and chemicals were obtained from either Fisher or Sigma and were of the highest purity available.

Cell Culture and Large Scale ESA Preparation—Parasite culture and ESA preparation were performed according to Ref. 31. Briefly, T. gondii strain 2F was propagated in human foreskin fibroblast cells. Freshly egressed tachyzoites were harvested by passage twice through a 20-gauge needle followed by filtration through a 3-mm pore size membrane to remove host cell debris. Parasites were washed twice by centrifugation in D0 medium (Dulbecco’s modified Eagle’s medium, 2 mM glutamine, 10 mM HEPES). Large scale preparation of ESA proteins was performed by incubating ~4 x 10^6 filter-purified tachyzoites in 15 ml of D0 plus 1% (v/v) ethanol at 37 °C for 20 min followed by cooling on ice for 5 min. Parasites were removed by centrifugation (1000 x g, 10 min, 4 °C). The supernatant, which was supplemented with a mixture of proteinase inhibitors, was concentrated to ~400 μl using C-20 concentrators according to the manufacturer’s instructions (Millipore; Billerica, MA).

Two-dimensional Gel Electrophoresis—ESA proteins were separated in the first dimension on 11- or 18-cm immobilized dry strips (pH 3–10, pH 4–7) using an isoelectric focusing system (Amersham Biosciences or Bio-Rad). 120 μg of ESA protein was mixed with rehydration buffer containing 8 M urea, 2% CHAPS, 0.5% carrier ampholytes buffer, and 65 mM dithiothreitol. Rehydration and isoelectric focusing were performed according to the manufacturer’s instructions. Following isoelectric focusing, proteins were reduced and alkylated by successive 15-min treatments with equilibration buffer containing 2% dithiothreitol followed by 2.5% iodoacetamide. Proteins were then resolved in the second dimension on an SDS-polyacrylamide gel (12.5% homogenous gel for the Amersham Biosciences system, 10–20% gradient gel for Bio-Rad). Resolved proteins were either transferred directly to polyvinylidene difluoride membrane for N-terminal sequencing or stained with colloidal Coomassie Blue stain (32) for MALDI-MS analysis.

N-terminal Sequencing—Proteins on a two-dimensional electrophoresis gel were transferred to Immobilon® P membranes (Millipore) and stained with Coomassie Brilliant Blue R-250. Bands were excised and subjected to Edman degradation on a PerkinElmer Life Sciences model 477A gas-phase protein sequencer.

In-gel Digestion, Peptide Extraction, MALDI-MS Analysis, and Data Base Searching—Protein spots from two-dimensional electrophoresis gels were excised, in-gel digested with trypsin (12.5 ng μl^-1 in 50 mM NH_4HCO_3), and solvent-extracted as described previously (33). After extraction, peptide mixtures were dried down and re-dissolved in 2 μl of 50% acetonitrile and 0.3% trifluoroacetic acid, mixed with saturated a-cyano-4-hydroxycinnamic acid (matrix) solution, and deposited on the MALDI-MS target plate according to a general two-layer method (34). The MS analysis of digested peptides was performed with a Voyager DE-STR MALDI-time-of-flight mass spectrometer (PerSeptive Biosystems, Framingham, MA). The spectra were acquired in the reflection-positive mode with delayed extraction, and external peptides or proteins adjacent to the sample spots were used for MS calibration. Internal calibration was refined using selected trypsin autolysis ions. Peptide monoisotopic masses were used to search against Toxoplasma NCBI data base via the MS-Fit algorithm (prospector.ucsf.edu) or to search Toxoplasma protein/EST/genome data bases via an in-house installed Mascot program (35). In both cases, a mass tolerance of ±50 ppm was allowed in the data base searches. Generally, a minimum number of four peptide matches, sequence coverage of at least 15%, and a match to a T. gondii entry were required for true positive match when using MS-fit algorithm, whereas a score higher than the Mascot significant score was regarded as positive match in Mascot data base searches.

MudPIT Analysis and Protein Identification—Lyophilized ESA proteins were reconstituted in 1 ml of 50 mM ammonium bicarbonate solution. The proteins were reduced by adding 10 μl of dithiothreitol (1 M) and incubation at 37 °C for 2 h. Cysteines were alkylated by adding 10 μl of iodoacetic acid (1 M in 1 N NaOH) and incubated in the dark for 30 min at room temperature. The sample was spun through a Microcon YM-3 concentrator until dry and resuspended in 1 μl of 50 mM ammonium bicarbonate buffer for trypsin digestion.

The digested ESA peptide mixtures were buffer-exchanged into 5% acetonitrile, 0.1% acetic acid prior to analysis. The MudPIT experiments were performed according to the manufacturer’s instructions on a ProteomEX work station (Thermo-Finnigan, San Jose, CA), which is composed of an autosampler, two HPLC pumps, a 10-port column-switching valve, and a Deca-XP ion trap mass spectrometer with a micro-ESI interface. For one-dimensional MudPIT, HPLC separation of peptide mixtures was conducted through reverse phase chromatography only, whereas for two-dimensional MudPIT, the HPLC separation was performed through ion exchange chromatography in the first dimension and then reverse phase chromatography in the second dimension. Specifically, in the case of two-dimensional MudPIT, peptides were bound to a strong cation exchanger (BioBasic SCX, 0.32 mm x 10 cm, Thermo Hypersil) and eluted stepwise with five NH_4Cl concentrations (10, 50, 200, 333, and 667 mM NH_4Cl). Fractions containing the peptide of interest were pooled, dried, and re-dissolved in 50% acetonitrile and 0.1% acetic acid.
100, 200, and 500 mM, respectively). Peptides eluted from each salt step were further separated on a reverse phase capillary column (BioBasic C18, 300Å, 5 μm silica, 180 μm × 10 cm) using an acetonitrile gradient (0–60% solvent B in solvent A, A = 0.1% formic acid; B = 100% acetonitrile containing 0.1% formic acid). The mass/charge (m/z) ratios of eluted peptides and fragmented ions were analyzed by an LCQ Deca XP ion trap mass spectrometer. Following each full scan mass spectrum, two MS/MS spectra of the top two most intense peaks were acquired. The dynamic exclusion feature was enabled to obtain MS/MS spectra of co-eluted peptides.

Protein identification was performed by searching the *T. gondii* protein subset of NCBI “nr,” the amino acid TgTwinScan gene predictions, and the clustered EST data bases using the TurboSequest algorithm (36) in the Bioworks 3.1 software package (Thermo Finnigan). The identified peptides were further evaluated using charge-state versus cross-correlation number (Xcorr). The criteria for positive identification of peptides were Xcorr >1.5 for singly charged ions, Xcorr >2.0 for doubly charged ions, and Xcorr >2.5 for triply charged ions (37). Low scoring peptide matches (Xcorr <1.9/2.2/3.1 for +1/+2/+3, respectively) were verified by manual inspection of MS/MS spectra. Additionally, only proteins with at least two peptides of distinct sequence identified were accepted.

*In Silico Analysis of Identified Proteins*—All proteins identified by any of the three proteomic approaches were subjected to multiple *in silico* analyses. Predicted localization of proteins was assigned primarily through previously published localization and subsequently by several algorithms as follows: signal peptide (SignalP 3.0) (38), apicoplast targeting (PATTS) (39), mitochondria targeting (PlasMit) (40), and nuclear localization sequence (PredictNLS) (41). Proteins with predicted signal peptides were also analyzed for the presence of transmembrane domains (TMHMM) (42) and GPI anchor sites (129.194.185.165/dgpi/index_en.html). Protein function was assigned based on published results. Novel proteins were assigned putative function if strong homology over long stretches of the protein (based on BLASTP with default settings against NCBI nonredundant data base) was found to a protein of known function.

**Ortholog Identification**—To identify orthologs to ESA proteins, we followed the approach used to generate the TIGR orthologous gene
alignments (TOGA) (43). ESA amino acid sequences were BLASTed against available apicomplexan TIGR gene indices (44) using WUBLAST tblastn, and for sequence matches with E values of <10^{-5}, the percent conservation of the best high scoring pair was recorded. Putative orthology was assigned if the reciprocal BLAST (blastx) using identical stringency against the Toxoplasma TgTwinScan amino acid gene predictions identified the original query as the top hit. As an alternative approach, we identified the protozoan OrthoMCL (45) ortholog clusters containing identified ESA proteins and retransformed the relative similarity scores among proteins into the original BLAST E values.

Expression and Analysis of YFP Fusion Proteins—Several novel ESAs predicted to be in the secretory pathway were cloned using primers designed against the TgTwinScan gene predictions and C-terminally fused to YFP (in destYFP/sagCAT (46)) using the Gateway cloning system (Invitrogen). Both extracellular and intracellular wild-type parasites were transiently transfected and grown in human foreskin fibroblast cells. Extracellular parasites were isolated by membrane filtration. Parasites were fixed in solution for 15 min with 2.5% formaldehyde, 0.02% glutaraldehyde and washed three times with PBS for 5 min, and extracellular parasites were permitted to settle on polylysine-coated chamber slides for 30 min. Slides were blocked with PBS, 10% fetal bovine serum for 30 min and permeabilized with 0.1% Triton X-100 for 15 min. For mitochondrial staining, isolated extracellular tachyzoites were permitted to settle on polylysine-coated chamber slides for 20 min. For mitochondrial staining, isolated extracellular tachyzoites were stained with the appropriate Alexa-594-conjugated secondary antibody and 4′,6-diamidino-2-phenylindole (5 μg ml^{-1}, Sigma), washed again, and mounted using Mowiol. The following primary antibodies were used at the indicated dilutions: mouse monoclonal anti-GRA2 (Tg17-179; 1:400), rabbit anti-GRA7 (kind gift of Isabelle Coppens; 1:400), rabbit anti-ROP2 (kind gift of Con Beckers; 1:2000) for extracellular parasites, mouse monoclonal anti-ROP2.3 (T34A71C8; 1:1000), and mouse monoclonal anti-AMA1 antibody (B3.90; 1:1000). Images were captured with ×600 or ×1000 magnification using a Nikon Eclipse E800 microscope equipped with a Spot RT CCD camera. Images were deconvoluted using SimplePCI (C-Imaging Systems, Cranberry Township, PA).

RESULTS

Toxoplasma secretory products have been implicated in parasite entry, intracellular survival, and interaction with the host (reviewed in Ref. 47). Regardless of whether they are initially secreted in a soluble or membrane-associated form, many of these products ultimately accumulate in ESA as a result of proteolytic shedding from the parasite surface (48–50).

ESA proteins were recovered from the medium after stimulating parasite secretion with 1% ethanol treatment (51). To maximize ESA product discovery, we employed several strategies that are outlined in Fig. 1. The first approach was to separate proteins by two-dimensional electrophoresis coupled with identification either by N-terminal sequencing (Fig. 1A), which was primarily used for alkaline proteins, or by MALDI-MS and peptide mass fingerprinting (PMF) (Fig. 1B). The second approach was to use multidimensional liquid chromatography to separate tryptic peptides in line with ESI-MS/MS to acquire tandem MS data (Fig. 1C). ESA proteins were identified by matching the acquired N-terminal sequence, MALDI/PMF, or ESI-MS/MS data with sequences in the T. gondii subset of the NCBI nonredundant data base, T. gondii EST, or gene prediction (TgTwinScan) data bases using several search algorithms including BLAST, ProteinProspector (51, 52), Mascot (35), and SEQUEST (36). A subset of the novel putative secretory proteins identified in these screens was then expressed in a YFP fusion construct to test for localization within the parasite secretory pathway.

To view a wide profile of extracellular products, ESA proteins were initially resolved on broad pH range IEF strips (pH 3–10). Because only a minority of species occupied the alkaline region of the gel, these proteins were identified by a combination of N-terminal sequencing (Fig. 2A) and MALDI/PMF (Fig. 2B). Note that different electrophoresis systems were used in Fig. 2, A–C. Twelve spots were subjected to N-terminal microsequencing and matched to their cognate genes by BLAST (TABLE ONE). Seven of the 12 spots were SAG proteins, a family of glycosylphosphatidylinositol (GPI)-anchored proteins that abundantly occupy the parasite surface. SAG1 and SAG2 each migrated as three distinct species. Each of the SAG1 species possessed the same N-terminal sequence, suggesting that the charge heterogeneity is not because of differential cleavage of the N terminus. On the other hand, N-terminal sequences of the SAG2 species were slightly offset in a manner that would alter both the size and charge of the protein, potentially resulting in the distinct migration of these species. Their presence in the ESA suggests that SAG proteins are shed from the parasite surface, possibly contributing to their well known hyper-immunogenicity and pro-inflammatory properties (53). One novel protein, termed p40, was identified by N-terminal sequencing. This product is discussed in greater detail below.

As shown in Fig. 2B, several additional proteins in the alkaline region were identified by MALDI/PMF including GRA2, a dense granule pro-

| N-terminal sequences of alkaline ESA proteins | Protein | Empirical M_{r}/pl | Theoretical M_{r}/pl | N-terminal sequence |
|-----------------------------------------------|---------|-------------------|-------------------|-----------------|
| SAG1                                          | 30/6.6  | 30/6.3            | SDPPPL            |
| SAG1                                          | 30/6.3  | 30/6.3            | SDPPLVANQVVT      |
| SAG1                                          | 30/6.3  | 30/6.3            | SDPPLV            |
| SAG2                                          | 13/8.0  | 14/7.6            | TPAPIETAGATKTV     |
| SAG2                                          | 14/6.3  | 14/6.4            | ETPAPIEXTAGA       |
| SAG2                                          | 15/6.1  | 15/6.4            | STXETTPAPIEXT      |
| SAG3                                          | 39/6.0  | 35/6.5            | XKSXSKITYF         |
| MIC11 C terminal                              | 9/0.8.2 | 6.7/8.2           | KFDTLKTLAKGALK      |
| p40                                           | 34/8.4  | 40/8.4            | KMPAVVKVK          |
| Cyp18                                         | 13/6.0  | 18/6.1            | ENAGVRKAYMDI       |
| Cyp20                                         | 15/8.4  | 19/8.7            | XNFXVFDI           |
| SOD^a                                         | 23/7.5  | 23/6.9            | VFTLPLLPAHDALAPXSS |

^a SOD indicates superoxide dismutase.
tein necessary for maintaining the intravacuolar network of membranes within the parasitophorous vacuole (13). Also present in the alkaline region was ROP9 (p36), a protein of unknown function (54) and the only rhoptry-derived protein identified in this study. A parallel study of the rhoptry proteome by Bradley et al. (120) provides a comprehensive analysis of this unique subset of secretory products.

Because the migration of ESA proteins was strongly skewed toward the acidic region of the broad pH range gels, we used high resolution narrow pH range IEF strips (pH 4–7) to further separate the protein mixture (Fig. 2C). These narrow range gels revealed 100 distinct spots, of which 55 were successfully matched to 35 distinct proteins by MALDI/PMF analysis (TABLE TWO). Identified proteins were classified into several main categories, including MIC proteins (MIC1, -2, -4, -5, -6, -8, -10, and -11, M2AP, AMA1, and SUB1), dense granule proteins (GRA1, -2, -5, and -7 and T. gondii protease inhibitor-1), surface antigens (SAG1, -2, and -3 and SRS1), and other secretory proteins (PDI and Cyp18). Also, two-dimensional electrophoresis/MS revealed the extensive proteolytic modifications that proteins such as M2AP, MIC2, MIC11, MIC4, and SUB1 undergo. In addition to proteins identified previously, several novel hypothetical proteins were discovered as a result of searching against the Toxoplasma data bases using PMF data (TABLE TWO). In general, this gel-based approach provided the

| Localization | Name | NCBI GI no. | Method* | Putative function | Ref. |
|--------------|------|-------------|---------|------------------|------|
| Microneme    | MIC1 | CAA96466    | P       | Lactose-binding lectin | 57, 89 |
|              | MIC2 | AAB63303    | P       | Glycosaminoglycan binding | 59, 68, 90 |
|              | MIC4 | AAD33906    | P       | Host cell binding | 58 |
|              | MIC5 | CAA70921    | P       | Unknown | 60 |
|              | MIC6 | AAD20818    | P       | Escorter | 65 |
|              | MIC8 | AAK19757    | P       | Host cell binding | 50 |
|              | MIC10| AAG32024    | P       | Unknown | 62 |
|              | SUB1 | AAK94670    | P       | Unknown | 63 |
| Dense granule| GRA1 | P13403      | P       | Unknown | 92 |
|              | GRA2 | P13404      | P       | Biogenesis of intravacuolar membranes | 13, 93 |
|              | GRA5 | Q07828      | P       | Unknown | 94 |
|              | GRA7 | AAC48314    | P       | Immune protection and parasite virulence | 14, 95 |
|              | TgP1-1| AAF32427    | P       | Serine protease inhibitor | 71, 72 |
| Rhoptry      | ROP9 | CAC04420    | P       | Unknown | 54 |
| Surface      | SAG1(P30) | CAA32245 | P, N | Immune modulation | 81, 96 |
|              | SAG2(P22) | A44968    | P, N | Unknown | 97 |
|              | SAG3 | AAO72428    | P, N | Host cell attachment | 78 |
|              | SUS1 | AAC47762    | P       | N/A | 98 |
| Others (ER, cytosolic, etc) | PDI | CAC28361    | P       | Immunodominant antigen in mucosal immune response | 82, 99 |
|              | Cyp18 | AAA17997    | P, N | PPase; Immune response modulator | 83, 100 |
|              | Cyp20 | AAA17998    | P, N | Peptidyl-prolyl-cis-isomerase | 83 |
|              | Enolase | AAG60329   | P       | Glycolytic enzyme | 101 |
|              | Mu1 adaptin | AAM77470 | P       | Vesicular transport | 103 |
|              | HSP70 | AAC72001    | P       | Chaperone; protection against cell damage associated with invasion | 104, 105 |
|              | SOD | AAC63943    | N       | Antioxidant | 106 |
|              | Peroxiredoxin | AAG25678 | P       | Antioxidant | 107 |
|              | Aldolase | AAN75043   | P       | Glycolysis; motility | 108 |
|              | Catalase | Q9XZD5     | P       | Antioxidant | 109 |
| Novel proteins | P40 | 27723986    | N       | Unknown | N/A |
|              | A30/9.0 | TgEST_100121304; TgTwinScan_2359 | P       | Host-cell binding | N/A |
|              | A70/6.0 | Tg9x_994290_hs_1 TgTwinScan_0069 | P       | Unknown | N/A |
|              | A40/4.5 | Tg9x_994635_hs_1 | P | Unknown | N/A |
|              | A50/7.0 | Tg9x_994250_hs_3 | P | Unknown | N/A |
|              | A11/6.5 | TgTwinScan_1327 | P | Unknown | N/A |
|              | A25/6.8 | TgTwinScan_4636 | P | Unknown | N/A |

*Identified either by MALDI/PMF (P) or N-terminal microsequencing (N).
### Proteins Identified by MudPIT

NCBInr is the non-redundant (nr) protein data base compiled by National Center for Biotechnology Information. EST is expressed sequence tag. The following abbreviations are used: *A.t.*, *Arabidopsis thaliana*; *C.p.*, *C. parvum*; *D.d.*, *Dictyostelium discoideum*; *E.t.*, *E. tenella*; *G.h.*, *Gossypium hirsutum*; *H.s.*, *Homo sapiens*; *P.y.*, *P. yoelii*; *P.f.*, *P. falciparum*. G3PDH indicates glyceraldehyde-3-phosphate dehydrogenase; G6PDH indicates glyceraldehyde-6-phosphate dehydrogenase; SOD indicates superoxide dismutase.

| Sequence ID | Name/best NCBI BLAST hit (E \(< 10^{-10}\)) | Putative function | Ref. | Predicted targeting | nr | Twinb- \(b\) scan | ESTb |
|-------------|---------------------------------------------|------------------|------|---------------------|----|-----------------|------|
| Micronemes  |                                            |                  |      |                     |    |                 |      |
| gi 2062142  | MIC1                                        | Lactose-binding lectin | 57, 89 | S                   | 10 | 9               | 10   |
| gi 1923217  | MIC2                                        | Glycosaminoglycan binding | 59, 68, 90 | S                   | 10 | 9               | 10   |
| gi 4929214  | MIC4                                        | Host cell binding  | 58   | S                   | 11 | 11              |      |
| gi 1806541  | MIC5                                        | Unknown           | 60   | S                   | 5  | 5               | 5    |
| gi 4708427  | MIC6                                        | Escorter          | 65   | S                   | 5  | 5               | 5    |
| gi 11141615 | MIC10                                       | Unknown           | 62   | A(S)                | 5  | 7               | 7    |
| gi 23267147 | MIC11                                       | Unknown           | 91   | S                   | 9  | 9               | 10   |
| gi 13991884 | M2AP                                        | MIC2 accessory protein | 64, 66 | S                   | 5  | 5               | 6    |
| gi 2293476  | AMA1                                        | Unknown           | 61   | S                   | 10 | 11              | 11   |
| gi 15419013 | SUB1                                        | Proteolysis       | 63   | S                   | 23 | 22              |      |
| Rhoptries   |                                            |                  |      |                     |    |                 |      |
| gi 9909994  | ROP9                                        | Unknown           | 54   | C(S)                | 3  | 2               | 4    |
| Dense granules |                                       |                  |      |                     |    |                 |      |
| gi 129322   | GRA1                                        | Unknown           | 92   | S                   | 8  | 4               | 4    |
| gi 2506908  | GRA2                                        | Biogenesis of intravacuolar membranes | 13, 93 | S                   | 3  | 3               | 2    |
| gi 2498422  | GRA3                                        | Unknown           | 111  | S                   | 2  | 2               | 3    |
| gi 3023908  | GRA7                                        | Unknown           | 14, 95 | S                   | 7  | 6               | 8    |
| gi 2499220  | NTP1                                        | Unknown           | 112, 113 | S                 | 16 | 16             | 16   |
| gi 2499221  | NTP2                                        | Unknown           | 112, 113 | S                 | 16 | 9               | 16   |
| gi 2702317  | NTP3                                        | Unknown           | 112, 113 | S                 | 15 | 9               | 15   |
| Surface     |                                            |                  |      |                     |    |                 |      |
| gi 37778533 | SAG1                                        | Immune modulation | 81, 96 | S                   | 9  | 6               | 9    |
| gi 37778535 | SAG2                                        | Unknown           | 97   | S                   | 4  | 4               |      |
| gi 3778537  | SAG3                                        | Host cell attachment | 78   | S                   | 5  | 2               | 3    |
| gi 1718389  | SR51                                        | Unknown           | 98   | S                   | 2  | 2               | 2    |
| gi 2305260  | SR52                                        | Unknown           | 98   | S                   | 2  | 2               | 2    |
| Secretry    |                                            |                  |      |                     |    |                 |      |
| gi 11125362 | Ferredoxin NAPD oxidoreductase              | Unknown           | 114  | Api                 | 2  |      |      |
| gi 14530190 | PDI                                         | Chaperone, mucosal antigen | 82   | S                   | 2  | 2               | 2    |
| gi 22035892 | G6PDH                                       | Glycolytic enzyme | S    | 5                   | 2  | 5               |      |
| Cytoplasm   |                                            |                  |      |                     |    |                 |      |
| gi 1695772  | LDH1                                        | Glycolytic enzyme | 115  | C                   | 4  | 4               | 4    |
| gi 13377044 | G3PDH                                       | Glycolytic enzyme | 116  | C                   | 2  | 2               | 3    |
| gi 1703160  | Actin                                       | Cytoskeleton/motility | 117  | C                   | 3  | 3               | 3    |
| gi 44889871 | eIF2a kinase A                              | Translation/stress response | 118  | C                   | 3  | 2               |      |
| gi 27817640 | Ser/Thr phosphatase 2C                     | Cytoskeleton/motility | 110, 119 | C                | 2  | 2               |      |
| gi 3746358  | SOD                                         | Antioxidant       | 106  | C                   | 2  | 2               | 2    |
| Secretry (predicted) |                           |                  |      |                     |    |                 |      |
| TgTwinScan_2357 | Microneme protein 5 (E.t.) | Adhesion/invasion | Novel protein | S | 2 | 2 |
| TgTwinScan_2358 | Microneme protein 5 (E.t.) | Adhesion/invasion | Novel protein | S | 2 | 2 |
| TgTwinScan_2359 | Microneme protein 5 (E.t.) | Adhesion/invasion | Novel protein | S | 2 | 2 |
| TgTwinScan_2361 | Microneme protein 5 (E.t.) | Adhesion/invasion | Novel protein | S | 2 | 2 |
| TgTwinScan_4000 | Secreted insulinase-like peptidase (C.p.) | Proteolysis | Novel protein | S | 10 | 2 |
| TgTwinScan_1683 | Amine oxidoreductase (H.s.) | Unknown | Novel protein | S | 2 | 2 |
| TgTwinScan_2661 | p40 | Unknown | Novel protein | S | 12 | 11 |
| TgTwinScan_203 | Hypothetical protein (P.f.) | Unknown | Novel protein | S | 2 | 2 |
| TgTwinScan_6350 | Hypothetical protein (P.f.) | Unknown | Novel protein | S | 7 | 4 |
| TgTwinScan_2489 | No hit in nr | Unknown | Novel protein | S | 2 | 2 |
| TgTwinScan_1114 | No hit in nr | Unknown | Novel protein | S | 2 | 2 |
advantage of visualizing the complexity and heterogeneity of ESA products but was limited by relatively low sensitivity because many of the low abundance species were not successfully identified.

As a complementary approach to two-dimensional electrophoresis/MS, the chromatography-based proteomic method MudPIT was used to analyze ESA proteins. ESA proteins were trypsin-digested, subjected to two-dimensional LC-MS/MS analysis, and identified by searching the un-interpreted product ion spectra against the nonredundant data base using the TurboSequest algorithm (36). A total of 62 proteins were identified (TABLE THREE). Strikingly, this approach resulted in the identification of additional proteins that were not detected on two-dimensional electrophoresis gels such as GRA3, NTP1, NTP2, NTP3, SRS2, and others. Moreover, several novel proteins emerged, including most notably a secreted metalloprotease of the insulinase family and orthologous proteins. The absence of targeting signals in 30% of the ESA proteins (Fig. 4) suggests that some cytosolic proteins are released into the ESA as a result of inadvertent parasite lysis. As expected, the majority (58%) of ESA proteins contained a putative signal sequence or a signal anchor sequence, many of which have been shown previously to localize to the either surface of the parasite or secretory organelles.

As an initial screen to determine the intracellular locations of novel secretory proteins, we selected five genes (TwinScan_0203, -1327, -2359, -2489, and -2661) predicted to lack a transmembrane segment or secretory anchor sequence, and orthologous proteins. The absence of targeting signals in 30% of the ESA proteins (Fig. 4) suggests that some cytosolic proteins are released into the ESA as a result of inadvertent parasite lysis. As expected, the majority (58%) of ESA proteins contained a putative signal sequence or a signal anchor sequence, many of which have been shown previously to localize to the either surface of the parasite or secretory organelles.

As an initial screen to determine the intracellular locations of novel secretory proteins, we selected five genes (TwinScan_0203, -1327, -2359, -2489, and -2661) predicted to lack a transmembrane segment or GPI anchor and cloned them in-frame with a C-terminal YFP-encoding expression plasmid. After transient transfection into tachyzoites, expression and localization were assessed by fluorescence microscopy. In formaldehyde-fixed parasites (Fig. 5). Despite multiple attempts, no expression of TwinScan_2489-YFP was observed. Together with the observation that transfection of this construct yielded morphologically defective parasites not seen with the other constructs, we conclude that expression of TwinScan_2489-YFP was toxic or otherwise poorly tolerated. 

Experimental methods

To assess localization and potential function, ESA protein sequences were analyzed by a series of algorithms designed to detect targeting signals and orthologous proteins. The absence of targeting signals in 30% of the ESA proteins (Fig. 4) suggests that some cytosolic proteins are released into the ESA as a result of inadvertent parasite lysis. As expected, the majority (58%) of ESA proteins contained a putative signal sequence or a signal anchor sequence, many of which have been shown previously to localize to the either surface of the parasite or secretory organelles.

### Table Three—Continued

| Sequence ID           | Name/best NCBI BLAST hit (E < 10⁻⁵) | Putative function                  | Ref. | Predicted targetinga | nrb | Twinb-scann | ESTb |
|-----------------------|------------------------------------|-----------------------------------|------|----------------------|-----|-------------|------|
| TgTwinScan_1327       | No hit in nr                        | Unknown                           | Novel protein | S  | 4  | 5  |
| TgTwinScan_3416       | No hit in nr                        | Unknown                           | Novel protein | S  | 2  | 2  |
| Cytoplasmic (predicted)|                                    |                                    |                  |    |    |    |
| TgTwinScan_0458       | Elongation factor 1-α (C.p.)        | Translation/cytoskeleton/motility  | Novel protein   | C  | 7  | 7  |
| TgTwinScan_3553       | Elongation factor 1-α (C.p.)        | Translation/cytoskeleton/motility  | Novel protein   | C  | 7  | 7  |
| TgTwinScan_3857       | Perforin-like protein (P.L.)        | Unknown                           | Novel protein   | C(S)| 6  |
| TgTwinScan_0397       | Chitinase-like protein (G.h.)       | Unknown                           | Novel protein   | C(S)| 2  |
| TgTwinScan_0436       | Similar to dynein heavy chain (D.d.)| Cytoskeleton                       | Novel protein   | C  | 2  |
| TgTwinScan_1519       | No hit in nr                        | Unknown                           | Novel protein   | C(TM)| 2 |
| TgTwinScan_2092       | LepA homolog (A.t.)                 | Unknown                           | Novel protein   | C  | 2  |
| TgTwinScan_6665       | RAD50 (C.p.)                        | DNA repair                        | Novel protein   | Nuc| 2  |
| TgEST_100120118       | Insulinase-like peptidase (C.p.)    | Unknown                           | Novel protein   | C  | 2  |
| TgTwinScan_2426       | Hypothetical protein (C.p.)         | Unknown                           | Novel protein   | C  | 2  |
| TgTwinScan_5841       | Hypothetical protein (P.y.)         | Unknown                           | Novel protein   | Nuc| 2  |
| TgEST_95058115        | Hypothetical protein (P.y.)         | Unknown                           | Novel protein   | C  | 2  |
| TgTwinScan_4526       | No hit in nr                        | Unknown                           | Novel protein   | Nuc| 2  |
| TgTwinScan_3652       | No hit in nr                        | Unknown                           | Novel protein   | Mito/Nuc| 2 |
| TgTwinScan_6494       | No hit in nr                        | Unknown                           | Novel protein   | Nuc| 2  |
| TgTwinScan_5465       | No hit in nr                        | Unknown                           | Novel protein   | C  | 2  |
| Unknown               |                                    |                                   |                  |    |    |    |
| TgEST_95059642        | No hit in nr                        | Unknown                           | Novel protein   | ρd| 3  |

a The following abbreviations are used: S, signal; A, anchor; C, cytoplasmic; Mito, mitochondrial; Api, apicoplast. (?) indicates conflicting information from other sources.
b Number of distinct peptide hits.
c Because of high sequence similarity between genes, a unique hit cannot be designated based on peptide hits.
d ? indicates incomplete open reading frame.
Because apicomplexan parasites share many core biological features, including an obligate intracellular lifestyle, we were interested to what degree the Toxoplasma-secreted ESA proteins were conserved across the phylum. Emulating the approach used to generate the TIGR orthologous gene alignments (TOGA), we assigned reciprocal BLAST hits against the TIGR gene indices for seven apicomplexan parasites as putative orthologs, identifying orthologs to 33 ESA proteins in one or more Apicomplexa (Fig. 6). As expected, the very closely related coccidian Neospora caninum showed the highest degree of conservation both in number of orthologs and their sequence homology. In general, after normalization with respect to the number of unique sequences in each TIGR gene index (lowest, Cryptosporidium parvum 910; highest, Plasmodium yoelii 9812), the number of orthologs identified decreased according to their phylogenetic distance from T. gondii (55). Because representation in EST collections like the TIGR gene indices is skewed by stage-specificity and abundance of transcripts, the false negative rate of this analysis is expected to be high, making a failure to detect orthologs in the other Apicomplexa less informative. To compensate for this and to expand our analysis to additional protozoa, we consulted a collection of orthologous clusters generated using OrthoMCL developed by David Roos’ laboratory at the University of Pennsylvania. By using OrthoMCL, which utilizes a wider array of sequences, we identified dozens of additional orthologs, particularly in more distantly related organisms. By using both of these approaches, we successfully identify orthologs in one or more apicomplexan parasites for 80% (35/44) of the predicted secretory ESA proteins, indicating that a number of ESA proteins are conserved within the phylum.

**DISCUSSION**

In this study we substantially widen the perspective of *Toxoplasma* proteins that form the interface of the parasite with its host during infection. The efficiency with which *Toxoplasma* initiates infection, avoids clearance by the innate immune system of the host, and persists indefinitely in host tissues is determined in large part by the surface of the parasite and secretory products. Although the precise function of most ESA products remains poorly understood, the current expanded view of these components is an important prelude to understanding mechanisms underlying *Toxoplasma*-host interactions contributing to infection and disease.

A number of cytosolic proteins were identified in the ESA. Because it was necessary to prepare ESA samples in defined, serum-free medium, some inadvertent parasite lysis was expected. Although most of these proteins are likely authentic cytosolic proteins, it is also possible that a subset may be incorrectly assigned due to inaccuracies of the gene assembly programs used to analyze the *Toxoplasma* genome, particularly the difficulty of correctly predicting the initial exon containing the signal sequence.

As expected, the largest subgroup of secretory proteins identified in the ESA is derived from the micronemes. In fact, the only known tachyzoite MIC protein not detected in this study was MIC3. This high coverage rate suggests that many of the novel ESA proteins may also be derived from MICs, a notion supported by the partial localization of recombinant TgTwinScan_2359 and TgTwinScan_2661 within these organelles. MICs are thought to participate in parasite adhesion and invasion of host cells (reviewed in Ref. 56). Consistent with their putative role in host cell attachment, many of the MIC proteins possess recognizable adhesive domains found in vertebrate proteins such as thrombospondin, integrins, and epidermal growth factor (9). Moreover, several studies have documented the cell binding activities of adhesive MICs, including MIC1, MIC2, MIC4, and MIC3 (49, 57–59). However, several other MIC proteins (MIC5, MIC10, M2AP, SUB1, and AMA1) do not contain obvious adhesive sequences, suggesting these proteins have
alternative or accessory functions (60–64). Another feature of *Toxoplasma* MICs is they often assemble into protein complexes and work in concert. For example, MIC6 partners with MIC1 and MIC4 (65), MIC8 escorts MIC3 (50), and MIC2 accompanies M2AP (64, 66, 67). Among these, the MIC2-M2AP complex appears to be the most abundant on the two-dimensional electrophoresis map of the ESA based on the Coomassie-stained two-dimensional electrophoresis profile (Fig. 2), a finding that is consistent with recent evidence that this adhesion complex plays a central role in parasite adhesion and invasion (59, 66, 68).

In the present study, 10 DG proteins (GRA1, GRA2, GRA3, GRA5, GRA7, and GRA8, *T. gondii* protease inhibitor-1, and NTPase I, II, and III) were recovered from the ESA fraction (TABLES TWO and THREE). DG proteins are secreted at the highest rate soon after the parasite is fully within the vacuole (68, 70). In most cases, the precise functions of DG proteins remain unclear, partly because they lack significant homology with proteins of known function. Despite this, several observations strongly suggest that DGs participate in the modification of PV and its interaction with the host cytosol. For example, almost all of the known DG proteins are membrane-associated proteins, which is in accordance with recent protein genetic evidence that GRA2 and GRA6 play important roles in the biogenesis of intravacuolar membranes (13). GRA7 was suggested to be involved in immune protection, and a decrease in GRA7 synthesis paralleled the loss of parasite virulence (14). The *T. gondii* serine protease inhibitor-1 (71) was recently confirmed as a DG protein, presumably playing an important role in protecting against the proteolytic enzymes and modulating the immune response during the acute...
phase of infection (72). NTPase isoforms are apyrases, which were demonstrated to function in intracellular replication of the parasite based on antisense interruption of expression (73). Also, the full scale activation of NTPase with dithiols may contribute to parasite exit from the host cell (74). It is likely that several of the novel secretory products identified herein are also derived from the DGs. Among these may be proteins of low antigenicity that would have been invisible to the antibody-based screens previously used to identify and characterized DG products.

It was somewhat unexpected to find significant amounts of SAGs in the ESA, given that these proteins are normally tethered to the parasite surface via a GPI anchor (75). At this time it is unclear how these SAGs are released into the ESA, and several possible explanations can be made. We speculated that the SAGs are soluble in the ESA because a small fraction of these proteins may have failed to acquire a GPI anchor, were liberated by a parasite phospholipase, or were proteolytically released from the surface. Alternatively, it is possible that SAGs are released by surface membrane blebbing, as has been observed for Toxoplasma (76) and other pathogenic protozoa (77). As for the function of SAGs in invasion, a SAG3-null mutant (78) showed an approximate 2-fold reduction in host cell invasion and partially attenuated infectivity in mice. This result was explained by a reduction in the capacity to attach the host cell. In addition, SAGs are one of the main targets for humoral and cell-mediated immune responses during initial infection (79–81).

FIGURE 6. Phylogenetic conservation of secretory ESA proteins among apicomplexan parasites. Only proteins with at least one putative ortholog are listed. Reciprocal BLAST hits against the TIGR gene indices of six other apicomplexan parasites with an E value of <10−5 were assigned as putative orthologs with percent positive (identical or chemically similar) conservation in the best high scoring pair shown and indicated in color code. The corresponding OrthoMCL ortholog clusters were used as comparison and to expand ortholog identification to additional protozoa. For the OrthoMCL clusters, the negative logarithm of the BLAST expect values are color-coded to indicate the extent of conservation. Additionally, the number of genes in each species (i.e., paralogs) belonging to the orthologous gene cluster is shown. Putative orthologs identified by one but not the other method are indicated by X. Predicted localization is also specified: microneme (Mic), rhoptry (Rho), dense granule (DG), surface (Sur), Apicoplast (Api), other secretory (OSec), unknown secretory (USec), and mitochondrial (Mito). Species abbreviations: N. caninum (Nc); Sarcocystis neurona (Sn); E. tenella (Et); C. parvum (Cp); Theileria annulata (Ta); Plasmodium berghei (Pb); P. yoelii (Py); P. falciparum (Pf); Plasmodium chabaudi (Pc); Plasmodium knowlesi (Pk); Tetrahymena tetraurelia (Tt); and Cyanidioschyzon merolae (Cm).
tified as one of two major immunophilins in *T. gondii* (83). Cyp18 is mainly located in the ER, but a subset of intracellular parasites shows significant Cyp18 staining in the PV. In addition to its role as a foldase, Sher and co-workers (16) recently demonstrated that Cyp18 stimulates dendritic cells to secrete interleukin-12, thereby contributing to the robust pro-inflammatory response often seen during *Toxoplasma* infection.

Several novel secretory proteins were identified in our analysis of ESA fractions. The sequences of some of these proteins are highly conserved in the Apicomplexa, implying that they fulfill parallel roles in the biology of these parasites. For example, we identified a novel insulin-like metalloprotease (TwinScan_4000), which may contribute to tissue damage or migration/dissemination during infection. Also, MudPIT revealed a family of proteins containing adhesive cysteine-rich Apple/PAN domains that are highly analogous to that of *E. tenella* MIC5 (84). Although genes encoding this family are tandemly linked in the genome, analysis of EST sequences suggests that they are stage-regulated because only one member (TwinScan_2359) is expressed in tachyzoites. A YFP fusion to this tachyzoite-specific product localizes to an apical compartment partially overlapping with micronemes and rhoptries. Additional studies will be required to determine its precise localization and whether this protein is mobilized to the parasite surface during invasion where it could contribute to host receptor binding via one or more of its Apple/PAN domains.

Apicomplexan-specific proteins that are conserved across the phyllum likely play fundamentally important roles the biology of these intra-cellular parasites. Our phylogenetic analysis of ESA products revealed extensive conservation of AMA1 and MIC2 (TRAP family), which is consistent with recent studies (67, 68, 85, 86) strongly implicating these proteins in cell invasion. ROP9 is also well conserved implying that this protein may play a central, albeit still shrouded, role in the Apicomplexa.

In summary, the current study illuminates a diverse legion of secretory proteins that contribute to parasite survival and pathogenesis during infection. Novel parasite-specific proteins identified in this screen significantly expand the number of potential targets for therapeutic intervention or diagnostic gain. We anticipate that ongoing in-depth functional analysis of identified novel proteins would further unravel the invasive and pathogenic mechanisms used by *Toxoplasma* during infection.

**Acknowledgments**—We gratefully acknowledge the Applied Biosystems Mass Spectrometry Facility at The Johns Hopkins School of Medicine, which was funded by a National Center for Research Resources shared instrumentation Grant 1S10-RR14702, The Johns Hopkins Fund for Medical Discovery, and the Institute for Cell Engineering. We thank Claudia Bordon for expert technical assistance. We also thank Feng Chen, ZhongGiang Chen, and David Roos for help with OrthoMCL; Dawn Chen for assistance with TurboSequest searches; Boris Striepen, Gary Ward, and Jean-Francois Dubremetz for providing reagents; and Mac Haymh for critically reading the manuscript.

**REFERENCES**

1. Dubey, J. P. (1994) *J. Am. Vet. Med. Assoc.* 205, 1593–1598
2. Dubey, J. P. (1998) *Int. J. Parasitol.* 28, 1019–1024
3. Aramini, J. J., Stephen, C., Dubey, J. P., Engelstoft, C., Schwantje, H., and Ribble, C. S. (1999) *Epidemiol. Infect.* 122, 305–315

4 V. B. Carruthers, unpublished data.
