THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 16, pp. 12063–12072, April 16, 2010
Printed in the U.S.A.

# Structural and Functional Characterization of SporoSAG

## *A SAG2-RELATED SURFACE ANTIGEN FROM TOXOPLASMA GONDII*[*] [S]

Received for publication, August 12, 2009, and in revised form, January 13, 2010 Published, JBC Papers in Press, February 17, 2010, DOI 10.1074/jbc.M109.054866

**Joanna Crawford[‡], Erika Lamb[§], James Wasmuth[¶1], Ognjen Grujic[‡], Michael E. Grigg[§2], and Martin J. Boulanger[‡3]**

*From the [‡]Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia V8W 3P6, Canada, the [§]Molecular Parasitology Unit, Laboratory of Parasitic Diseases, NIAID, National Institutes of Health, Bethesda, Maryland 20892, and the [¶]Program in Molecular Structure and Function, Hospital for Sick Children, Toronto, Ontario M5G 1L7, Canada*

***Toxoplasma gondii*, the etiological agent of toxoplasmosis, utilizes stage-specific expression of antigenically distinct glycosylphosphatidylinositol-tethered surface coat proteins to promote and establish chronic infection. Of the three infective stages of *T. gondii*, sporozoites are encapsulated in highly infectious oocysts that have been linked to large scale outbreaks of toxoplasmosis. SporoSAG (surface antigen glycoprotein) is the dominant surface coat protein expressed on the surface of sporozoites. Using a bioinformatic approach, we show that SporoSAG clusters with the SAG2 subfamily of the SAG1-related superfamily (SRS) and is non-polymorphic among the 11 haplogroups of *T. gondii* strains. In contrast to the immunodominant SAG1 protein expressed on tachyzoites, SporoSAG is non-immunogenic during natural infection. We report the 1.60 Å resolution crystal structure of SporoSAG solved using cadmium single anomalous dispersion. SporoSAG crystallized as a monomer and displays unique features of the SRS β-sandwich fold relative to SAG1 and BSR4. Intriguingly, the structural diversity is localized to the upper sheets of the β-sandwich fold and may have important implications for multimerization and host cell ligand recognition. The structure of SporoSAG also reveals an unexpectedly acidic surface that contrasts with the previously determined SAG1 and BSR4 structures where a basic surface is predicted to play a role in binding negatively charged glycosaminoglycans. Our structural and functional characterization of SporoSAG provides a rationale for the evolutionary divergence of this key SRS family member.**

*Toxoplasma gondii* is a highly prevalent, obligate, intracellular protozoan parasite that infects nearly one-third of the human population (1, 2). Since its recognition as the causative agent of toxoplasmosis in the late 1930s, many clinical manifestations have been attributed to *T. gondii* infections including lymphadenopathy, ileitis, encephalitis, and/or blinding ocular infections in both children and adults (1, 3–8). *T. gondii* infections can also be lethal to a developing fetus and immunocompromised, cancer, AIDS, and organ transplant patients. A key feature of the ability of *T. gondii* to infect and multiply in virtually any warm blooded animal is a complex life cycle that encompasses both sexual (sporozoite) and asexual (tachyzoite and bradyzoite) stages of development. Sexual replication occurs exclusively in felines, whereas asexual division occurs in all warm blooded hosts. Three main routes of infection exist for *T. gondii*. Congenital infection occurs with maternal transmission of tachyzoites via the placenta to a fetus, whereas ingestion of bradyzoites occurs when tissue cysts in undercooked meat are consumed, and finally, infection with sporozoites occurs when food or water contaminated with oocysts is consumed. A remarkable trait of *T. gondii* is its ability to differentiate from the ingested bradyzoite or sporozoite into the fast replicating tachyzoite, facilitating rapid dissemination throughout the host. Upon challenge by the immune system, the tachyzoite converts to the slow growing, encysted bradyzoite responsible for establishing a chronic, transmissible infection.

Historically, meat containing *T. gondii* tissue cysts was considered the major route of transmission to humans; however, improved farming practices have considerably reduced the likelihood of infected livestock (9). The prevalence of *T. gondii* in felines, however, remains high (10–14). Shedding of *T. gondii* oocysts containing infectious sporozoites by both feral and domestic cats continues to result in widespread environmental contamination, due both to large numbers of parasites shed (as many as 1 million/cat) and to the resistance of oocysts to environmental degradation (15). Localized incidences of waterborne transmission as well as substantial outbreaks due to contaminated water supplies or soil have been extensively documented (16–21). Due to the potentially broad scale of an outbreak, environmental transmission of *T. gondii* through sporozoite ingestion arguably poses the most significant global risk.

Stage conversion in *T. gondii* coincides with a major change in expression of surface antigens belonging to the surface anti-

* This work was supported by Research Grant MOP82915 from the Canadian Institutes of Health Research (CIHR) (to M. J. B.). This work was also supported by the Intramural Research Program of the National Institutes of Health (to M. E. G.) through the NIAID.

*The atomic coordinates and structure factors (codes 2WNK and 2X28) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).*

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. 1 and 2.

[1] Recipient of CIHR Grant MOP84556 (to M. E. G.).

[2] A scholar of the Canadian Institute for Advanced Research (CIFAR) Program for Integrated Microbial Biodiversity. To whom correspondence may be addressed: Molecular Parasitology Unit, Laboratory of Parasitic Diseases, NIAID, National Institutes of Health, 4 Center Dr., Bethesda, MD, 20892. Tel.: 301-402-1609; Fax: 301-402-0079; E-mail: griggm@niaid.nih.gov.

[3] A CIHR New Investigator and a Michael Smith Foundation for Health Research (MSFHR) scholar. To whom correspondence may be addressed: Dept. of Biochemistry and Microbiology, University of Victoria, P. O. Box 3055 STN CSC, Victoria, BC,V8W 3P6, Canada. Tel.: 250-721-7072; Fax: 250-721-8855; E-mail: mboulang@uvic.ca.

gen 1 (SAG1)[4]-related sequences (SRS) family, which are predicted to play a dual role in parasite attachment and regulation of host immunity to establish chronic infection (22–26). Structural requisites of this superfamily are an N-terminal secretion signal, a glycosylphosphatidylinositol (GPI) anchor, and a complement of conserved amino acids that typically include six cysteines per SRS monomer that participate in three disulfide bonds. Sequencing of the *T. gondii* genome has revealed more than 160 SRS family members (27). Recent structural studies of the predominantly tachyzoite-expressed SAG1 and bradyzoite-expressed BSR4 identified a topologically defined groove that is postulated to coordinate host cell surface molecules such as heparin (25, 28).

Although SAG1 and BSR4 show differential expression patterns, they are both members of the SAG1 family, which are phylogenetically divergent from the SAG2 family of SRS antigens (27). Interestingly, paralogs within the SAG2 family share less identity (~20%) when compared with the SAG1 family (~30%). Members of the SAG2 family also differ in terms of open reading frame size, with the smaller SAG2A and SAG2B proteins consisting of a single SAG domain, whereas SAG2C and SporoSAG contain two SAG domains interrupted by a single intron. Despite the increased divergence of the SAG2 family relative to the SAG1 family, it is the SAG2 family that is more conserved across other tissue-dwelling coccidia (27).

To date, only two structural descriptions have been reported for the 160+ SRS superfamily: one for the tachyzoite-expressed SAG1 (25) and the second for the bradyzoite-expressed BSR4 (28). To determine the structural and functional implications of a sporozoite-expressed SRS protein, we report the 1.60 Å resolution crystal structure and immunoreactivity profile of the major GPI-tethered cell surface antigen expressed in sporozoites. The characterization of SporoSAG also provides the first structural view of a *T. gondii* SRS family member from the infectious sporozoite stage and from the SAG2 subfamily.

## EXPERIMENTAL PROCEDURES

*Bioinformatics*—The protein sequence alignment for phylogenetic reconstruction was constructed using the program PROMALS3D (29). The structures of SAG1 (1KZQ), BSR4 (2JKS), and SporoSAG (2WNK) were used to anchor the alignment. Besides the choice of initial sequence alignment method (PROMALS), default parameters were implemented. The resulting alignment was hand-edited in Geneious (30). Due to the prevalence of indels in the sequence alignment, positions where the 50% consensus was a gap were ignored for tree building. A neighbor-joining tree was generated from a protein distance matrix, using PROTDIST and NEIGHBOR from the PHYLIP package (version 3.68) (31). The rate variation between sites was gamma distributed with an alpha value of 0.5. A total of 10,000 bootstraps were run, and only nodes receiving >50% support were resolved in the tree displayed.

*SporoSAG Immunogenicity*—Sera from infected patients, mice, and rabbits were tested for reactivity against insect cell-produced SporoSAG by enzyme-linked immunosorbent assay, as described previously by Kong *et al.* (32). Briefly, infected patient, mouse, and rabbit sera (1:80 dilution) were incubated overnight with 0.3 $\mu$g of baculovirus-expressed recombinant proteins. Positive reactions were detected using horseradish peroxidase-conjugated anti-human, anti-mouse, or anti-rabbit polyvalent IgG, IgA, and IgM detection antibodies (Sigma), developed with 2,2′-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) peroxidase substrate (Kirkegaard & Perry Laboratories, Inc.), and the absorbance was measured at 405 nm after >60 min. Serum samples from a known uninfected volunteer served as a negative control, and serum from a patient infected with Type II strain oocysts was included as the positive control, respectively. For murine and rabbit infections, positive control sera were derived by injecting bradyzoite cysts or sporulated oocysts of the Type II Me49 strain via the oral (bradyzoites, oocysts) and subcutaneous routes (oocysts). To obtain sporulated oocysts, fecal floats from cats shedding *T. gondii* oocysts were incubated in 2% sulfuric acid for 1 week at room temperature to allow for sporulation. Sporulated oocysts were counted, and either 100 or 1000 oocysts were fed orally or injected subcutaneously into mice or rabbits. Sera reactivity against insect cell-produced SAG1 or SRS2 was measured as the positive control for infection with *Toxoplasma*, and reactivity against a generic insect cell-produced protein was included as a negative control.

*Molecular Cloning and Recombinant Virus Production*—The tandem SAG domains of SporoSAG are separated by an intervening intronic sequence therefore requiring a multistep PCR approach to generate a single cDNA template. Each domain was initially amplified separately from *T. gondii* RH genomic DNA (Ex1, forward, 5′-cagactccatgggaCCTGAAGCAACTTCCTGTG-3′; Ex1 (overlap), reverse, 5′-CACCGGAGCCTGTGGTGCCGCTTTGACGAG-3′; Ex2 (overlap), forward, 5′-AAGCGGCACCAAGGCTCCGGTGTGTTCGG-3′; Ex2, reverse, 5′-ctgtctggcggccgcAACGGTTACCAAAACTTTGCAGG-3′). Domain 1 and domain 2 amplicons were then combined, allowing homologous regions (underlined) to anneal and extend, resulting in a single cDNA template. This product was then amplified (Ex1 forward and Ex2 reverse) and cloned into NcoI/NotI-digested pAcGP67b (Pharmingen) in-frame with a C-terminal hexahistidine tag. Expression of the sequence-verified clone was conducted in the baculovirus insect cell system. Primary virus was generated by co-transfection of *Spodoptera frugiperda* (Sf9) insect cells with linearized Sapphire™ baculovirus DNA (Orbigen) and the SporoSAG/pAcGP67b modified clone. The recombinant SporoSAG encoding primary virus was subsequently amplified to generate high titer virus for use in protein production.

*Recombinant Protein Expression and Purification*—A 4-liter volume of High Five insect cells was infected with 2.5 ml of amplified SporoSAG virus/liter of culture. The culture was harvested after 65 h and centrifuged at 450 × *g* to remove cellular debris, and the supernatant was concentrated using tangential flow and buffer-exchanged into buffer A (20 mM HEPES, pH 8, 1 M NaCl, 30 mM imidazole). SporoSAG was selectively purified HisTrapFF nickel affinity with fractions analyzed by SDS-PAGE and pooled based on purity. Removal of the SporoSAG

[4] The abbreviations used are: SAG, surface antigen glycoprotein; SRS, SAG-related sequence; GPI, glycosylphosphatidylinositol; SporoSAG, major sporozoite surface antigen glycoprotein.

hexahistidine tag with thrombin was followed by gel filtration (Superdex 75) in buffer B (20 mM HEPES, pH 8, 150 mM NaCl) and anion exchange chromatography (Source 30Q) in buffers C (20 mM HEPES, pH 8, 10 mM NaCl) and D (20 mM HEPES, pH 8, 500 mM NaCl) to produce sufficiently pure protein for crystallization studies. The purified SporoSAG protein begins with the N terminus sequence GSAMGPEAT (GS remaining after thrombin cleavage and AMG derived from the NcoI site) and extends through LVTVAAALVPR (AAA from the NotI site and LVPR remaining after thrombin cleavage), resulting in a total molecular mass of 25,657 Da.

*Crystallization, Data Collection, and Processing*—Crystallization trials were set up with Crystal screen HT™, Crystal screen 2 HT™, and PEG/ION screen™ (Hampton Research) and Precipitant Synergy™ Primary 64 crystallization screen (Emerald Biosystems) in 96-well plates (Emerald Biosystems). The final drops consisted of 1 μl of SporoSAG (at 25 mg/ml) in 20 mM HEPES, pH 8, 100 mM NaCl, and 1 μl of reservoir solution equilibrated against 100 μl of reservoir solution. Crystals of native SporoSAG were observed within 3 days in 25% polyethylene glycol 3350, 0.1 M Tris, pH 8.5, whereas cadmium-derivatized crystals of SporoSAG grew in 1 week in 20% polyethylene glycol 4000 with 0.2 M $Cd(NO_3)_2$. SporoSAG crystals were cryoprotected in mother liquor plus 20% glycerol for 30 s and flash cooled at 100 K directly in the cryostream. Diffraction data were collected on a Rigaku R-axis IV++ area detector coupled to an MM-002 x-ray generator with Osmic "blue" optics and an Oxford Cryostream 700. Diffraction data for the native SporoSAG crystals to 1.60 Å and cadmium-derivatized crystals to 2.15 Å were processed using Crystal Clear software with d*TREK (33). Data collection and refinement statistics are presented in Table 1.

*Structure Solution and Refinement*—A total of five cadmium sites were identified and refined using autoSHARP (34). High quality phases of the cadmium derivative were obtained following density modification that enabled building and registering of ~80% of the backbone using ARP/Warp (35). The partially built SporoSAG structure was then used as a search model in molecular replacement using MOLREP (36) with the high resolution native data set. The remaining structure was built manually, and solvent atoms were selected using COOT (37) and refined with REFMAC (38) to an $R_{cryst}$ of 19.0 and an $R_{free}$ of 24.0%. All solvent atoms were inspected manually before deposition. Stereochemical analysis of the refined SporoSAG structure was performed with PROCHECK and SFCHECK in CCP4 (39) with the Ramachandran plot showing excellent stereochemistry with 99.3% of the residues in the most favored and additional allowed conformations and no residues modeled in disallowed orientations. Overall, 5% of the reflections were set aside for calculation of $R_{free}$. Data collection and refinement statistics are presented in Table 1.

**TABLE 1**
**Data collection and refinement statistics**
Values in parentheses are for the highest resolution shell. NA, not applicable.

| | Native SporoSAG | $Cd(NO_3)_2$ SporoSAG |
|---|---|---|
| **Data collection** | | |
| Spacegroup | P6 | $P2_12_12$ |
| *a*, *b*, *c* (Å) | 106.42, 106.42, 31.60 | 102.62, 72.97, 30.66 |
| α, β, γ (degrees) | 90, 90, 120 | 90, 90, 90 |
| Wavelength | 1.5418 | 1.5418 |
| Resolution (Å) | 30.72-1.60 | 28.26-2.15 |
| Measured reflections | 223926 | 158772 |
| Unique reflections | 52494 | 13203 |
| Redundancy | 4.3 (3.5) | 12.0 (9.9) |
| Completeness (%) | 99.3 (95.5) | 100 (98.8) |
| $I/\sigma(I)$ | 18.1 (2.1) | 21.5 (5.3) |
| $R_{merge}$ [a] | 0.033 (0.451) | 0.107 (0.441) |
| Figure of merit | | 0.29724 |
| **Refinement statistics** | | |
| Resolution range (Å) | 23.42–1.60 (1.69–1.60) | 28.30–2.15 (2.20–2.15) |
| $R_{cryst}$ [b]/$R_{free}$ [c] | 0.190/0.240 | 0.236/0.289 |
| No. of atoms | | |
| Protein | 1623 | 1600 |
| Solvent | 310 | 168 |
| Cadmium | NA | 5 |
| *B*-values | | |
| Protein (Å$^2$) | 18.90 | 21.56 |
| Solvent (Å$^2$) | 31.75 | 30.19 |
| Cadmium (Å$^2$) | NA | 37.57 |
| r.m.s.[d] deviation from ideality | | |
| Bond lengths (Å) | 0.010 | 0.019 |
| Bond angles (degrees) | 1.303 | 1.949 |

[a] $R_{merge} = \sum_{hkl}\sum_i|I_{hkl,i} - (I_{hkl})|/\sum_{hkl}\sum_i I_{hkl,i}$, where $(I_{hkl})$ is the average of symmetry-related observations of a unique reflection.
[b] $R_{cryst} = \sum|F_{obs} - F_{calc}|/\sum F_{fobs}$, where $F_{obs}$ and $F_{calc}$ are the observed and the calculated structure factors, respectively.
[c] $R_{free}$ is *R* using 5% of reflections randomly chosen and omitted from refinement.
[d] r.m.s., root mean square.

## RESULTS AND DISCUSSION

*SporoSAG Is a SAG2 Subfamily SRS Protein*—Phylogenetic clustering analyses of the 161 bioinformatically related SRS proteins have previously identified a bifurcated tree that established two evolutionarily distinct sequence clades, or subfamilies, related to SAG1 and SAG2, respectively (27). To determine the SRS subgroup the sporozoite protein SporoSAG most closely aligned with, prototypic sequences of SRS proteins known to cluster with the SAG1 subfamily (SAG1, SAG3, SAG5A, SAG5C, BSR4, SRS2, and SRS9) and with the SAG2 subfamily (SAG2A, SAG2C, SAG2Y, SRS20A, SRS23, SRS37A, SRS47D, and SRS47F) were aligned with SporoSAG using the structure-based PROMALS3D alignment program. The structures of SAG1, BSR4, and SporoSAG were used to anchor the alignment (see supplemental Fig. 1). A neighbor-joining tree was constructed using the PROTDIST and NEIGHBOR programs in the PHYLIP 3.68 suite of phylogenetic software, and 10,000 bootstrap replicates were run. Only nodes receiving greater than 50% support were resolved in the tree (Fig. 1). The bootstrap with neighbor-joining analysis robustly supported the inclusion of the dominant sporozoite surface antigen SporoSAG as a distal member of the SAG2 clade of sequences. A domain-based tree was also generated, and both domains of SporoSAG clustered with the SAG2 subfamily (data not shown).

*SporoSAG Is Not Immunogenic during Natural Infection*—SAG1 and SAG2 are the dominant tachyzoite antigens expressed against which the majority of the humoral immune response is focused (40–44). To determine whether the dominant sporozoite surface antigen SporoSAG is likewise immunodominant during *T. gondii* infections, we evaluated sera reactivity from known *T. gondii*-infected hosts, across a range of species and across routes of infection (Fig. 2). Ten and three serum samples (designated *Oocyst*), collected from patients whose infection was linked epidemiologically to the consumption of water-borne oocysts from outbreaks in Victoria, Canada

(1995) and Santa Isabel do Ivai, Brazil (2001), respectively, were tested for the presence of antibodies that specifically reacted against SporoSAG (18, 19). Control samples comprised serum from pregnant patients whose *T. gondii* infection was detected during routine screening and who were presumably infected via the consumption of undercooked meat (designated *Bradyzoite*) (12, 45). Patient sera were evaluated across a range of titers to SAG1, with higher titers prioritized to maximize the possibility of detecting a reaction against SporoSAG. Sera from an uninfected individual and from a Type II strain oocyst-infected individual were included as the negative and positive controls, respectively. No antibody reactivity (neither IgM, IgA, nor IgG) was detected against SporoSAG in any of the human samples, suggesting that the host immune system does not mount an antibody response to this protein (Fig. 2*A*). This was not the result of a failure of these individuals to mount antibody responses against *Toxoplasma* because all patients possessed anti-SAG1 reactivity. Identical results were obtained using sera acquired from either bradyzoite-infected or sporulated oocyst-infected mice inoculated by peroral injection or via subcutaneous inoculation and developed using polyvalent anti-mouse secondary antibodies to detect IgA, IgM, and IgG (Fig. 2*B*). The same sporulated oocysts were also injected subcutaneously into rabbits and yielded the same results (data not shown). The lack of reactivity to SporoSAG during infection established that this highly expressed, sporozoite-specific antigen is a poor candidate reagent for the development of a test capable of detecting individuals infected by oocysts, a question of considerable epi-

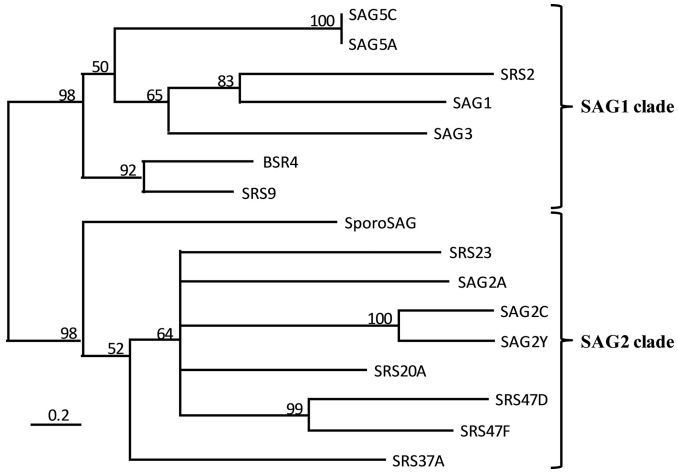



FIGURE 1. **SporoSAG is a SAG2 subfamily SRS protein.** SporoSAG was aligned with SAG1 and SAG2 subfamily sequences using the PROMALS3D program and hand-edited in Geneious. A neighbor-joining tree was constructed using PROTDIST and NEIGHBOR (PHYLIP 3.68), and 10,000 bootstrap replicates were run to identify supported nodes.

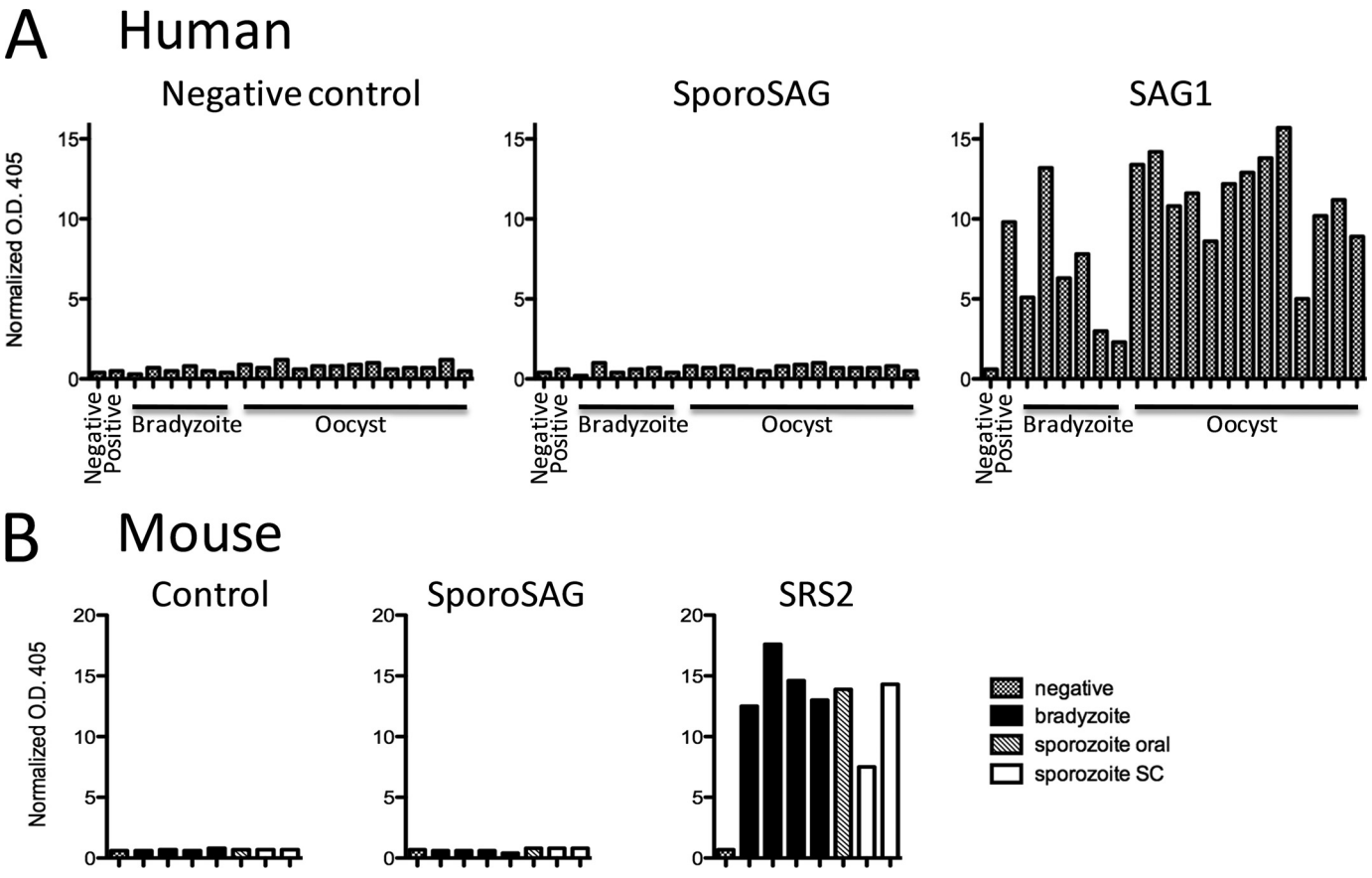



FIGURE 2. ***T. gondii*-infected patients fail to mount antibody responses to insect cell-produced SporoSAG.** *A*, 19 human serum samples from patients with *T. gondii* infections epidemiologically linked to consumption of water-borne oocysts (*Oocyst*) or presumed bradyzoite transmission (*Bradyzoite*) were tested against sera from an uninfected patient (*Negative*) and sera from a known Type II *T. gondii* oocyst infection (*Positive*). *O.D.*, optical density. *B*, serum from mice infected with 20 Type II Me49 bradyzoite cysts perorally or 100 Type II Me49 sporulated oocysts either perorally or subcutaneously were reacted against an irrelevant baculovirus-expressed D2 antigen, recBSR4 protein (*Control*); SporoSAG; and the diagnostic proteins for *T. gondii* infection, SAG1 (human) or SRS2 (mouse). Enzyme-linked immunosorbent assay data for each serum were normalized by dividing the optical density value obtained at 405 nm for each of the *Toxoplasma* proteins by the optical density reading obtained for the irrelevant control D2 protein.

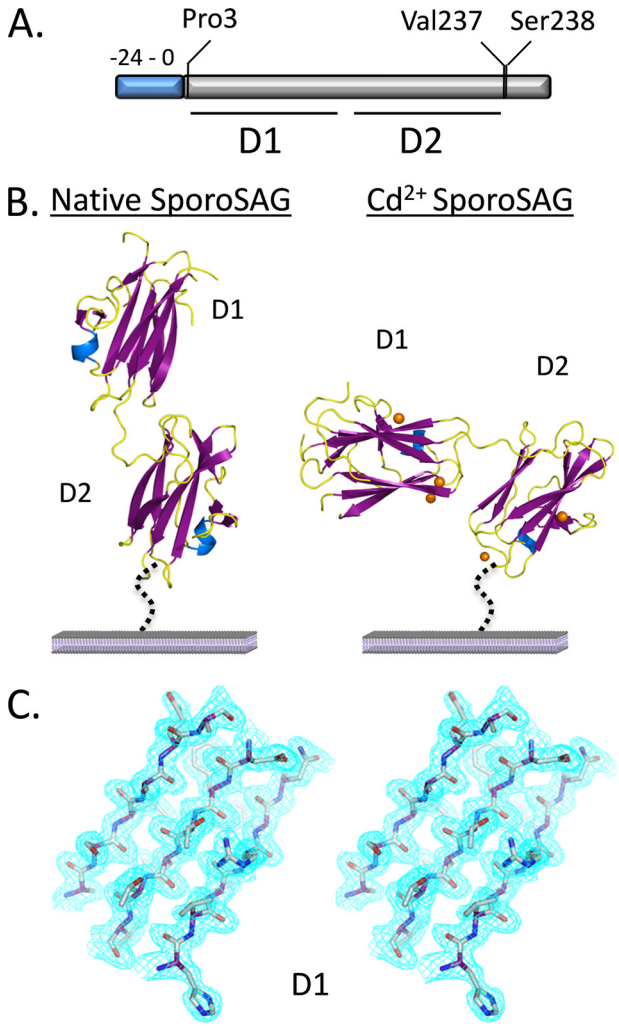



FIGURE 3. **Overall structure of SporoSAG.** *A*, gene construct of *T. gondii* Type II SporoSAG. The insect cell-produced construct extends from Pro-3 to Val-237. The *blue region* denotes the signal sequence cleaved in the mature protein, and *Ser238* denotes the predicted GPI anchor site. *B*, secondary structure representation of SporoSAG in the native P6 form (*left panel*) and the cadmium (*orange spheres*)-bound $P2_12_12$ form (*right panel*). The SporoSAG monomers are presented with the c-terminal end, which is modified by a GPI anchor in the parasite, directed toward the cell surface membrane. *C*, cross-eyed stereo view of a $2F_o - F_c$ σA-weighted electron density map of the D1 domain of SporoSAG contoured at 1.2 σ. All structure figures were prepared with PyMOL (48).

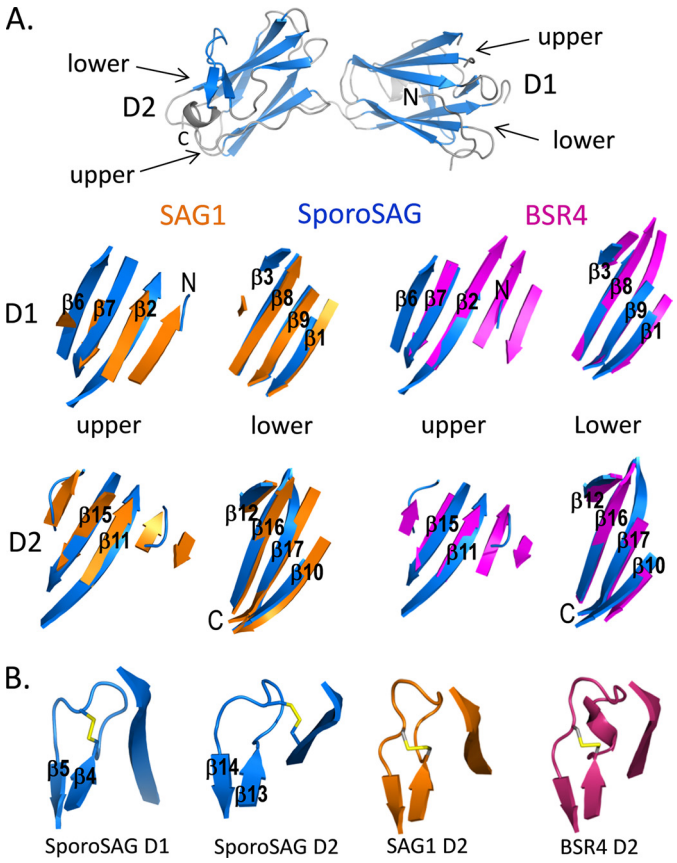
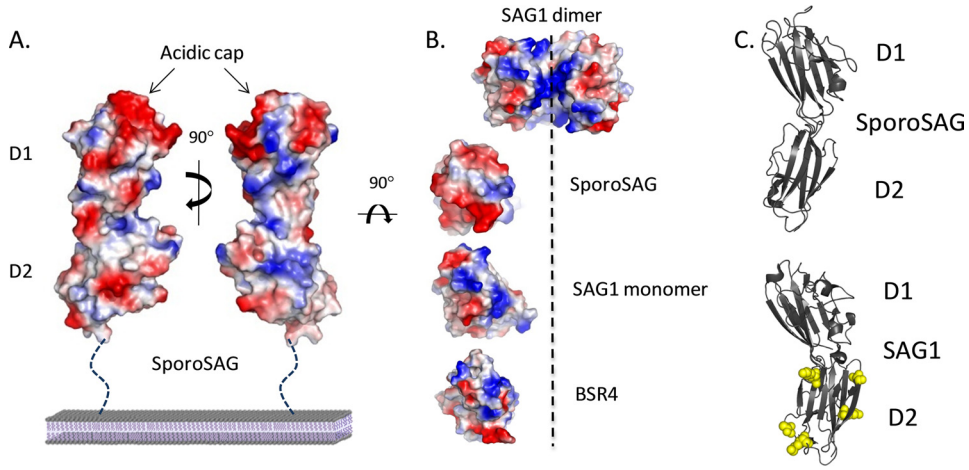



FIGURE 4. **Structural divergence in the SRS fold.** *A*, *top panel*, secondary structure depiction of native SporoSAG highlighting the strand structure (*blue*) of the SRS β-sandwich fold. *Bottom panel*, structural superposition of the upper and lower β-sheets of the β-sandwich fold from SporoSAG (*blue*), SAG1 (*orange*), and BSR4 (*magenta*). The upper β-sheets of both the D1 and the D2 domains of SporoSAG show significant structural divergence when compared with SAG1 and BSR4, whereas the lower β-sheets are largely conserved. *B*, secondary structure view highlighting the unexpectedly structurally equivalent disulfide bond in the D2 domain of SporoSAG (*blue*) relative to the D1 domain of SporoSAG, the D2 domain of SAG1 (*orange*) and the D2 domain of BSR4 (*magenta*), which show a spatially conserved disulfide bond. The new disulfide-bonding pattern in SporoSAG is due to a sequence-shifted cysteine (Cys-220) that appears to be selected for in members of the SAG2 subfamily. Secondary structure assignment is correlated to sequence in supplemental Fig. 1.

demiologic importance. We next sought to solve the SporoSAG structure to compare its evolutionary relationship and structural features with the solved structures for the tachyzoite (SAG1) and bradyzoite (BSR4) proteins that belong to the SAG1 subfamily of SRS proteins.

*Overall Structure*—The cloned SporoSAG construct begins at Pro-3 and extends through Val-237, one residue short of the predicted GPI anchor site, Ser-238, that tethers the mature protein to the outer membrane of *T. gondii* (Fig. 3*A*). Production of SporoSAG in insect cells was followed by extensive purification using nickel affinity, size exclusion, and ion exchange chromatography and resulted in a protein that migrated as a single band on an SDS-PAGE gel with apparent molecular mass of 30 kDa (supplemental Fig. 2*A*). Initial crystallization trials produced a P6 crystal form that diffracted to 1.60 Å. Attempts to solve the SporoSAG structure using molecular replacement with the individual core domains of SAG1 (25) and BSR4 (28, 46) as search models were unsuccessful. Additional rounds of crystallization produced a cadmium-derivatized $P2_12_12$ crystal form from which we were able to generate high quality phases using cadmium single-wavelength anomalous dispersion. It is noteworthy that SporoSAG adopts a monomer in solution, as shown by gel filtration chromatography, and crystallizes as a monomer in both crystal forms (supplemental Fig. 2, *B* and *C*). The absence of an observed dimer for SporoSAG contrasts with the structures of SAG1, which crystallized as a dimer (25), and BSR4 that showed a clear dimer along the crystallographic 2-fold axis (28), suggesting that SporoSAG may exist as a monomer on the cell surface of the parasite. However, the physiological GPI tether that is absent in our recombinant construct may be required to promote dimerization of SporoSAG on the parasite cell surface.

The overall structure of SporoSAG is comprised of tandem β-sandwich domains tethered by a short linker. Despite the




SporoSAG acidic cap

| SporoSAG | SAG1 | BSR4 |
|---|---|---|
| Glu7 | Pro13 | Asp17/Arg6 |
| Glu9 | Gln120 | Ala153 |
| Glu33 | Ala38 | Ile54 |
| Glu34 | Leu39 | Ser55 |
| Glu72 | Ser87 | Glu107 |
| Asp81* | Lys33* | Lys117 |

SAG1 basic groove

| SporoSAG | SAG1 | BSR4 |
|---|---|---|
| Thr8 | Lys15 | Arg6 |
| Gly10 | Lys16 | Pro27 |
| Asp81* | Lys33* | Lys117 |
| Ser30 | Lys36 | Thr43 |
| Arg25 | Lys94* | Thr43 |

FIGURE 5. **Implications for ligand binding.** *A*, surface electrostatic calculations of native SporoSAG oriented with the D2 domain proximal to the membrane and the D1 domain distal to the membrane as predicted for SAG1 (25) and BSR4 (28). Note the acidic patch that caps the D1 domain of SporoSAG. Of the six acidic residues in SporoSAG, none are conserved in SAG1, and only two are conserved in BSR4. *B*, top view of the SAG1 dimer highlighting the basic groove and the same top view of the SporoSAG, SAG1, and BSR4 monomers along the SAG1 dimer axis. None of the five basic residues in SAG1 are conserved in SporoSAG, and only two are conserved in BSR4. Of particular interest is the spatial substitution of the central Lys-33 in SAG1 and Lys-117 in BSR4 with Asp-81 in SporoSAG. *C*, ribbon diagram showing the total lack of polymorphism in SporoSAG among all alleles PCR-amplified and DNA-sequenced for representative strains among the 11 haplogroups of *T. gondii* strains when compared with SAG1 and BSR4 that display significant polymorphism (*yellow spheres* represent polymorphic residues). Note that the polymorphisms do not align with the charged groove of BSR4.

incorporation of two proline residues (Pro-131 and Pro-134), the linker appears to be flexible as evidenced by its ability to accommodate distinct relative orientations of the N-terminal (D1) and C-terminal (D2) domains in the two crystal forms (Fig. 3*B*). In the P6 crystal form (Fig. 3*B*), the tandem domains are organized in a head-to-tail fashion with the linker adopting a bowed-out conformation, whereas in the $P2_12_12$ form (Fig. 3*B*), the linker adopts a more linear structure with the D2 domain reoriented and positioned orthogonally to the long axis of the D1 domain. Ultimately, the orientations of the SRS domains in these two crystal forms reflect the potential for reorganization on the parasite cell surface. Despite these spatial differences, the overall structures of the individual domains are well ordered, displaying clear electron density (Fig. 3*C*), and are largely conserved with root mean square deviations between the D1 domains and between the D2 domains for the two crystal forms of 0.42 and 0.38 Å, respectively.

*Structural Divergence in the SRS β-Sandwich Fold*—Our bioinformatic analysis clearly shows that SporoSAG is a member of the SAG2 subfamily. To determine how this sequence-based classification translates to the structural level, we compared the structure of SporoSAG with SAG1 (25) and BSR4 (28). A DALI search revealed that SporoSAG is the most structurally divergent of the three. SporoSAG possesses z-scores of 12.3 and 10.3, respectively, when compared with the D1 domains of SAG1 and BSR4. The D2 domains are slightly more divergent with z-scores of 11.4 and 10.0 relative to SAG1 and BSR4, respectively. By contrast, the z-scores between BSR4 and SAG1 are 20.5 and 17.8 for the D1 and D2 domains, respectively. A detailed analysis reveals that this structural divergence is primarily localized to the upper leaves of each SRS domain (Fig. 4*A*). This observation is intriguing when analyzed in the context of SAG1 and BSR4 where the upper leaves of D1 form part of an extended groove that participates in dimerization and is predicted to coordinate host cell ligands (25, 28).

Despite the structural bifurcation of D1 and D2, the position of the three key disulfide bonds that serve as primary determinants for the SRS fold are spatially conserved between SporoSAG, SAG1, and BSR4. Two pairs of disulfide bonds are responsible for stitching the upper and lower leaves of the β-sandwich and show less than 0.4 Å between the three structures. Sequence alignment, however, predicted a reorganization in the third disulfide bond of the SporoSAG D2 domain as Cys-220 was significantly shifted in sequence relative to SAG1. Intriguingly, the repositioned cysteine aligns with one of the cysteines in SAG2A, consistent with our bioinformatic assessment of SporoSAG as a member of the SAG2 subfamily. However, it is also evident from the sequence alignment that the third disulfide bond is not absolutely conserved and is even predicted to be absent in the D2 domain of some SAG2A family sequences, specifically SAG2C and SAG2Y. A comparative structural analysis reveals that despite the deviation in Cys-220 in the SporoSAG primary sequence, the local structure is largely unperturbed relative to SAG1 (25) and BSR4 (28). This is reflected in a pair of antiparallel strands in SAG1 (Cys-182 → Glu-192) and BSR4 (Cys-218 → Thr-230), which are connected by a short loop stabilized by a disulfide bond (SAG1, Cys-182 and Cys-190; BSR4, Cys-218 and Cys-228) (Fig. 4*B*). Although the interstrand loop is conserved in SporoSAG (Leu-174 → Glu-182), the stabilizing disulfide bond is reorganized such that Cys-179 bonds with Cys-218 on β16, thereby tethering the loop directly to the core β-sandwich (Fig. 4*B*). It appears then that at least in this region of SporoSAG, that sequence, rather than structural divergence, may be the essential feature.

*Substituting the Basic Groove for an Acidic Cap*—Unambiguous definition of the most biologically relevant orientation of SporoSAG on the parasite cell surface is difficult and beyond the scope of the structural studies presented here. Definitive

orientation is complicated by the absence of the GPI anchor in the recombinant protein and the observed flexibility in the interdomain linker. Combined, these features may promote multiple orientations of SporoSAG on the cell surface with distinct potential for multimerization and ligand recognition. However, to develop testable hypotheses designed to correlate structure with function, we compared the native P6 crystal form of SporoSAG with the high resolution crystal structures of SAG1 (25) and BSR4 (28). The P6 form was chosen because it most closely approximates the domain orientation observed in SAG1 (25) and BSR4 (28). For the sake of completeness, we have also incorporated a discussion, albeit more limited, of the $P2_12_12$ SporoSAG crystal form.

An electrostatic surface representation of SporoSAG oriented perpendicular to the membrane shows a predominantly acidic surface at the membrane-distal region of D1 (Fig. 5*A*). The cap of SporoSAG D1 is dominated by acidic residues Glu-7, Glu-9, Glu-33, Glu-34, Glu-72, and Asp-81, which are chemically distinct from the spatially analogous residues in SAG1 (25) and BSR4 (28) (Fig. 5, *A*, *left table*, and *B*). One feature that is particularly interesting is the substitution of Asp-81 in SporoSAG for Lys-33 in SAG1 (Lys-117 in BSR4) that forms the core of the basic groove predicted to play a role in coordinating negatively charged sulfated proteoglycans such as heparin (25). Intriguingly, Asp-81 is located on strand $\beta$7 that forms part of the divergent upper sheet of the SporoSAG D1 domain (Fig. 4*B*). In contrast to Asp-81, the majority of acidic residues in SporoSAG D1 tend to be more apically located than the analogous residues in SAG1 and BSR4. This, along with the observation the SporoSAG adopts a monomer in solution (supplemental Fig. 2), may suggest a functional role that relies on molecular crowding to generate a broad acidic surface rather than the dimerization observed in SAG1 and BSR4. We cannot, however, discount the possibility that SporoSAG adopts a form on the cell surface that more closely approximates the $P2_12_12$ crystal. In this form, the acidic surface would be less likely to project outwards. Instead, a more weakly charged surface would be exposed that would likely require formation of a contorted multimer to form a contiguous electrostatic surface.

To assess the degree of polymorphism in the acidic cap of SporoSAG D1, we analyzed the sequences of the 11 haplogroups of *T. gondii* strains that currently define the population genetic structure. Intriguingly, no polymorphisms were observed, indicating a biological pressure to maintain its primary sequence (Fig. 5*C*). This total lack of diversity is sharply contrasted by the large number of alleles found among *T. gondii* strains for other SRS genes, including *bsr4* and *sag3*. Only the D1 domain of SAG1 has previously been shown to be devoid of polymorphism among extant strains of *T. gondii*, further suggesting that SporoSAG has an important role in the biology of *Toxoplasma* infection, consistent with the previously published results indicating that SporoSAG possesses a demonstrated role in host cell attachment and invasion (47). To date, however, no ligand has been identified for SporoSAG. The extensive acidic surface combined with the lack of a defined basic surface in either crystal form suggests that SporoSAG may coordinate a family of ligands distinct to the tachyzoite-expressed SRS proteins SAG1 and SAG3 that are predicted to bind a negatively charged sulfated proteoglycan (25).

*Conclusions*—In this study, we have undertaken a detailed immunological, bioinformatic, and structural characterization of SporoSAG from *T. gondii*. Despite abundant expression of SporoSAG on the infectious sporozoite, sera from patients do not possess antibodies against SporoSAG. This may be due to the speed at which sporozoites convert to tachyzoites during infection; hence, their limited persistence is insufficient to elicit an immune response. Our bioinformatic analysis clearly shows that SporoSAG is a member of the SAG2 subfamily, consistent with our structural analysis demonstrating that SporoSAG is significantly divergent from SAG1 (25) and BSR4 (28), the only other structures of SRS proteins solved to date. A detailed comparison revealed that the SRS fold is structurally divergent in the upper leaves of the D1 and D2 domains, whereas the lower leaves are structurally conserved. This divergent feature is not predicted at the level of primary sequence and provides intriguing insight to understanding the complexities of the SRS fold that may be critical for ligand binding and multimerization. Furthermore, our data show that the SporoSAG displays a negatively charged surface at the tip of the D1 domain that contrasts with the positively charged surface of SAG1 and BSR4. Whether the acidic surface of SporoSAG is functionally relevant for ligand coordination remains to be determined. Finally, a total lack of allelic polymorphism in SporoSAG among extant *Toxoplasma* lineages provides a compelling new dimension for this highly conserved, abundantly expressed protein that heralds a fresh new perspective in the evolution of functional diversity within the SRS superfamily.

## REFERENCES


1. Jackson, M. H., and Hutchison, W. M. (1989) *Adv. Parasitol.* **28,** 55–105
2. McCabe, R. E., and Remington, J. S. (1983) *Eur. J. Clin. Microbiol.* **2,** 95–104
3. Luft, B. J., Brooks, R. G., Conley, F. K., McCabe, R. E., and Remington, J. S. (1984) *JAMA* **252,** 913–917
4. Luft, B. J., and Remington, J. S. (1992) *Clin. Infect. Dis.* **15,** 211–222
5. McDonald, J. C., Gyorkos, T. W., Alberton, B., MacLean, J. D., Richer, G., and Juranek, D. (1990) *J. Infect. Dis.* **161,** 769–774
6. Nussenblatt, R. B., and Belfort, R., Jr. (1994) *JAMA* **271,** 304–307
7. Silveira, C., Belfort, R., Jr., Burnier, M., Jr., and Nussenblatt, R. (1988) *Am. J. Ophthalmol.* **106,** 362–364
8. Grigg, M. E., Ganatra, J., Boothroyd, J. C., and Margolis, T. P. (2001) *J. Infect. Dis.* **184,** 633–639
9. Tenter, A. M., Heckeroth, A. R., and Weiss, L. M. (2000) *Int. J. Parasitol.* **30,** 1217–1258
10. Dubey, J. P., Saville, W. J., Stanek, J. F., and Reed, S. M. (2002) *J. Parasitol.* **88,** 802–803
11. Dubey, J. P., Weigel, R. M., Siegel, A. M., Thulliez, P., Kitron, U. D., Mitchell, M. A., Mannelli, A., Mateus-Pinilla, N. E., Shen, S. K., Kwok, O. C., *et al.* (1995) *J. Parasitol.* **81,** 723–729
12. Dubey, J. P., Zhu, X. Q., Sundar, N., Zhang, H., Kwok, O. C., and Su, C. (2007) *Vet. Parasitol.* **145,** 352–356
13. Sharif, M., Daryani, A., Nasrolahei, M., and Ziapour, S. P. (2009) *Trop. Anim. Health Prod.* **41,** 183–187
14. Smith, K. E., Zimmerman, J. J., Patton, S., Beran, G. W., and Hill, H. T. (1992) *Vet. Parasitol.* **42,** 199–211
15. Dubey, J. P. (2004) *Vet. Parasitol.* **126,** 57–72
16. Bahia-Oliveira, L. M., Jones, J. L., Azevedo-Silva, J., Alves, C. C., Oréfice, F., and Addiss, D. G. (2003) *Emerg. Infect. Dis.* **9,** 55–62
17. Benenson, M. W., Takafuji, E. T., Lemon, S. M., Greenup, R. L., and Sulzer, A. J. (1982) *N. Engl. J. Med.* **307,** 666–669

18. Bowie, W. R., King, A. S., Werker, D. H., Isaac-Renton, J. L., Bell, A., Eng, S. B., and Marion, S. A. (1997) *Lancet* **350,** 173–177
19. de Moura, L., Bahia-Oliveira, L. M., Wada, M. Y., Jones, J. L., Tuboi, S. H., Carmo, E. H., Ramalho, W. M., Camargo, N. J., Trevisan, R., Graça, R. M., da Silva, A. J., Moura, I., Dubey, J. P., and Garrett, D. O. (2006) *Emerg. Infect. Dis.* **12,** 326–329
20. Dubey, J. P., Sharma, S. P., Juranek, D. D., Sulzer, A. J., and Teutsch, S. M. (1981) *Am. J. Vet. Res.* **42,** 1007–1010
21. Teutsch, S. M., Juranek, D. D., Sulzer, A., Dubey, J. P., and Sikes, R. K. (1979) *N. Engl. J. Med.* **300,** 695–699
22. Carruthers, V., and Boothroyd, J. C. (2007) *Curr. Opin. Microbiol.* **10,** 83–89
23. Dzierszinski, F., Mortuaire, M., Cesbron-Delauw, M. F., and Tomavo, S. (2000) *Mol. Microbiol.* **37,** 574–582
24. Grimwood, J., and Smith, J. E. (1996) *Int. J. Parasitol.* **26,** 169–173
25. He, X. L., Grigg, M. E., Boothroyd, J. C., and Garcia, K. C. (2002) *Nat. Struct. Biol.* **9,** 606–611
26. Mineo, J. R., and Kasper, L. H. (1994) *Exp. Parasitol.* **79,** 11–20
27. Jung, C., Lee, C. Y., and Grigg, M. E. (2004) *Int. J. Parasitol.* **34,** 285–296
28. Crawford, J., Grujic, O., Bruic, E., Czjzek, M., Grigg, M. E., and Boulanger, M. J. (2009) *J. Biol. Chem.* **284,** 9192–9198
29. Pei, J., Kim, B. H., and Grishin, N. V. (2008) *Nucleic Acids Res.* **36,** 2295–2300
30. Drummond, A. J., Kearse, M., Heled, J., Moir, R., Thierer, T., Ashton, B., Wilson, A., and Stones-Havas, S. (2006) *Geneious*, version 2.5, Biomatters Ltd., Auckland, New Zealand
31. Felsenstein, J. (1989) *Cladistics* **5,** 164–166
32. Kong, J. T., Grigg, M. E., Uyetake, L., Parmley, S., and Boothroyd, J. C. (2003) *J. Infect. Dis.* **187,** 1484–1495
33. Pflugrath, J. W. (1999) *Acta Crystallogr. D Biol. Crystallogr.* **55,** 1718–1725
34. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) *Methods Mol. Biol.* **364,** 215–230
35. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6,** 458–463
36. Vagin, A., and Teplyakov, A. (1997) *J. Appl. Crystallogr.* **30,** 1022–1025
37. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
38. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255
39. (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50,** 760–763
40. Bessières, M. H., Le Breton, S., and Séguéla, J. P. (1992) *Parasitol. Res.* **78,** 222–228
41. Khan, I. A., Eckel, M. E., Pfefferkorn, E. R., and Kasper, L. H. (1988) *J. Infect. Dis.* **157,** 979–984
42. Lau, Y. L., and Fong, M. Y. (2008) *Exp. Parasitol.* **119,** 373–378
43. Partanen, P., Turunen, H. J., Paasivuo, R. T., and Leinikki, P. O. (1984) *J. Clin. Microbiol.* **20,** 133–135
44. Prigione, I., Facchetti, P., Lecordier, L., Deslée, D., Chiesa, S., Cesbron-Delauw, M. F., and Pistoia, V. (2000) *J. Immunol.* **164,** 3741–3748
45. Jones, J. L., Kruszon-Moran, D., and Wilson, M. (2003) *Emerg. Infect. Dis.* 9, 1371–1374
46. Grujic, O., Grigg, M. E., and Boulanger, M. J. (2008) *Acta Crystallogr Sect. F Struct. Biol. Cryst. Commun.* **64,** 425–427
47. Radke, J. R., Gubbels, M. J., Jerome, M. E., Radke, J. B., Striepen, B., and White, M. W. (2004) *Mol. Microbiol.* **52,** 93–105
48. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific LLC, San Carlos, CA

# Structural and Functional Characterization of SporoSAG: A SAG2-RELATED SURFACE ANTIGEN FROM TOXOPLASMA GONDII

Joanna Crawford, Erika Lamb, James Wasmuth, Ognjen Grujic, Michael E. Grigg and Martin J. Boulanger

*J. Biol. Chem. 2010, 285:12063-12070.*
*doi: 10.1074/jbc.M109.054866 originally published online February 17, 2010*

---