Structural basis of surface antigen glycoprotein mediated virulence in *Toxoplasma gondii*

by

Ekaterina Bruic
Doctorate, Medicine, Tver State Medical Academy, Russia, 2003

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biochemistry and Microbiology

Supervisory Committee

Structural basis of surface antigen glycoprotein mediated virulence in *Toxoplasma gondii*

by

Ekaterina Bruic
Doctorate, Medicine, Tver State Medical Academy, Russia, 2003
A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biochemistry and Microbiology

Supervisory Committee

Dr. Martin J. Boulanger, Assistant Professor, Department of Biochemistry and Microbiology
Supervisor

Dr. Caroline Cameron, Associate Professor, Department of Biochemistry and Microbiology
Departmental Member

Dr. Fraser Hof, Assistant Professor, Department of Chemistry
Outside Member

Abstract

Supervisory Committee
Dr. Martin J. Boulanger, Assistant Professor, Department of Biochemistry and Microbiology
Supervisor
Dr. Caroline Cameron, Associate Professor, Department of Biochemistry and Microbiology
Departmental Member
Dr. Fraser Hof, Assistant Professor, Department of Chemistry
Outside Member

Toxoplasma gondii is a eukaryotic, intracellular parasite capable of infecting any vertebrate animal and establishing a life-long latent infection. Despite the prevalence of *T. gondii* infections, the molecular mechanisms by which these parasites gain access to the host cell remain largely unknown. Recent knockout studies have implicated a select group of *T. gondii* surface proteins, termed SRSs (Surface Antigen Glycoprotein Related Sequences), in directing parasite attachment and persistence. Follow-up structural studies with the prototypical SRS antigen, SAG1, revealed a novel fold, termed the SRS fold, and a dimeric structure with a topologically defined basic groove predicted to play a role in ligand binding. While these initial results were very exciting, follow-up work has failed to identify a host cell ligand for SAG1, and no other members out of more than 160 members of the SRS superfamily have been structurally characterized. As a result, conservation of the SRS fold and, more specifically, structural determinants of molecular recognition remain elusive.

While sequence alignments of the SRS superfamily suggested conservation of the SRS fold, several insertions and deletions presented the possibility of localized structural elements that may be essential in molecular recognition. To characterize how these insertions/deletions are represented at the structural level, the X-ray crystal structures of two members of the SRS superfamily, BSR4 and SRS2, were solved. Structural analysis revealed an unexpected degree of diversity in the SRS fold. Divergent connectivity of the beta-strands in studied proteins indicates that the SRS superfamily may be more structurally diverse than previously thought, while structural variations in the beta-strands

and the loops of D1 domain suggests a possible mechanism to recognize diverse host cell ligands, such as heparan sulfate proteoglycans (HSPGs). To probe HSPGs binding and determine the role of homodimerization, the dimer constructs of SRS2 and BSR4 were engineered, produced and tested in a carbohydrate binding macro-array. Selective binding of the SRS2 dimer to heparin was detected during screening and validated using heparin-agarose pull-down and native gel shift assays. Possible molecular mechanism for SRS-HPGS interaction and the implications in *T. gondii* virulence are also discussed.

Table of Contents

List of Tables

List of Figures

List of abbreviations

~	approximately
^{o}C	degrees Celsius
Å	Angstrom
a. a.	amino acid
AIDS	Acquired Immuno Deficiency Syndrome
Ala, A	Alanine
Arg, R	Arginine
Asp, D	Aspartate
AT	Antithrombin
BEVS	Baculovirus Expression Vector System
BLAST	Basic Local Alignment Search Tool
bp	base pairs
BSA	Bovine Serum Albumin
BSR4	Bradyzoite surface antigen
CCP4	Collaborative Computational Project 4
CHO	Chinese hamster ovary cells
COOT	Crystallographic Object-Oriented Toolkit
CSF	Cerebrospinal fluid
Cys , C	Cysteine
DMSO	dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	deoxynucleotidetriphosphate
ECM	extracellular matrix
ESPRIPT	Easy Sequencing in Postscript
ER	endoplasmic reticulum
EtBr	ethidium bromide
EXPASY	Expert Protein Analysis System
FPLC	Fast protein liquid chromatography
GAG	Glucosaminoglycan
GalNac	*N*-Acetylgalactosamine
Glc	glucose
GlcA	glucuronic acid
GlcNac	N-acetylglucosamine
GlcNS	N-sulfated glucosamine
Glu, E	Glutamic acid
GPI	Glycosylphosphatidylinositol
6X His tag	Hexahistidine tag
*Hi*5	High Five Cells derived from *Trichopulsia ni*
HBS	Hepes Buffered Saline
His, H	Histidine
HIV	Human Immunodeficiency Virus
HSPG	heparan sulphate proteoglycans
IdoA	iduronic acid

Ig	Immunoglobulin
IPTG	isopropyl-beta-D-thiogalactopyranoside
K	degrees Kelvin
kDa	kilodalton
LB	Luria-Bertani medium
Lys, K	Lysine
mA	milliamp
mAb	Monoclonal antibody
MCS	Multiple Cloning Site
mg	milligram
MIC	microneme protein
MR	Molecular Replacement
ml	millilitres
mM	millimolar
NCBI	National Centre for Biotechnology Information
Ni-NTA	nickel-nitrilotriacetic acid
ng	nanogram
OD_{600}	Optical Density at 600nm
ORF	Open Reading Frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB ID	protein data base identification
PEG	polyethylene glycol
RNA	ribonucleic acid
r.m.s.d.	root mean square deviation
RON	rhoptry neck protein
rpm	revolutions per minute
RT	room temperature
SAG	surface antigen glycoprotein
SAXS	Small Angle X-ray Scattering
SDS	sodium dodecyl sulphate
*Sf*9	*Spodoptera frugiperda*
SFM	Serum Free Medium
spp.	several species
SRS	SAG1 related sequence
TgAMA-1	*Toxoplasma gondii* apical membrane antigen-1
Th	T-helper
Thr, T	Threonine
Trp, W	Tryptophan
UV	ultraviolet
µg	microgram
µl	microliter
V	volt
v/v	volume to volume ratio
w/v	weight to volume ratio
WT	wildtype

Acknowledgements

I am grateful to my supervisor, Dr. Martin Boulanger, for his support and expertise involved in this project. I would also like to thank my committee members, Dr. Caroline Cameron and Dr. Fraser Hof for their guidance and advice. Special thanks to my co-workers Joanna Crawford and Ognjen Grujic for their help in this project, and to fellow students Jasleen Bains, Adrienne Law, Jeremy Mason and Michelle Tonkin. I also would like to acknowledge the people in the Dr. Boraston's lab for allowing me the use their equipment.

Dedication

“Look deep into nature, and then you will understand everything better.”

~Albert Einstein

Structural biochemistry provides an in depth view into nature. To take this closer look for the first time would not be possible for me without the inspiration and guidance from my former chemistry instructor Dr. Lawrence Lee and my current supervisor Dr. Martin Boulanger, my co-workers and friends Jasleen Bains, Ognjen Grujic, Joanna Crawford, Adrienne Law…and, of course, the love, patience and delicious dinners from my dear family Michael, Anna, Nikolas, Irina and Andrey to whom I dedicate this thesis.

Chapter1: Introduction.

1.1 *Toxoplasma gondii.*

T. gondii is an obligate intracellular parasite capable of infecting any vertebrate animal and establishing chronic disease (Sibley 1995). *T. gondii* is a serious human pathogen with up to 30% of the adult human population infected worldwide. This parasite is also of a high veterinary importance for its distribution across multiple species of pets (Tenter *et al.* 2000), livestock (Dubey 1998b; Dubey *et al.* 1986) and marine animals (Oksanen *et al.* 1998). The ability to infect such a broad range of hosts resulted in *T. gondii* being recognized as the most successful parasite on Earth (Kim and Weiss 2004). Although the mechanism of the parasite's attachment and invasion is currently extensively studied, the precise molecular interactions involved remain largely unknown. Identification and characterization of these interactions at the molecular level will allow development of prophylactic or therapeutic interventions for *T. gondii* infections.

1.2 Nomenclature and ultrastructure of *Toxoplasma gondii*.

T. gondii is a parasitic protozoan, which exists as a single species in the family *Sarcosystidae*, order *Eucoccidiorida*, subclass *Coccidiasina*, class *Conoidasida* and phylum *Apicomplexa*. *T. gondii* is a model for studying the phylum *Apicomplexa* (Kim and Weiss 2004), which also includes *Plasmodium*, the ethiologic agent of malaria, and *Cryptosporidium*, a devastating AIDS pathogen (Kissinger and Kuo 2007).

T. gondii is a unicellular eukaryote with an approximate size of 6µm x 2µm. It consists of various organelles and inclusion bodies (Figure 1). Typical to all parasites of the phylum *Apicomplexa*, *T. gondii* contains one apicoplast (a non-photosynthetic plastid), and the

apical complex, which includes the conoid, micronemes and rhoptries. The conoid is a filamentous structure, capable of rotating, tilting, extending and retracting. Located at the apical end of *T. gondii*, the conoid is used for probing the host cell plasmalemma immediately before penetration (Dubey *et al.* 1998). Micronemes are secretory organelles that contain proteins important for initial recognition of the host cell. Rhoptries are excretory organelles that contain a mixture of proteins that are discharged into the host cell during invasion (Blackman, 2001). Both micronemes and rhoptries occur mostly at the anterior end of the parasite (Dubey *et al.* 1998).

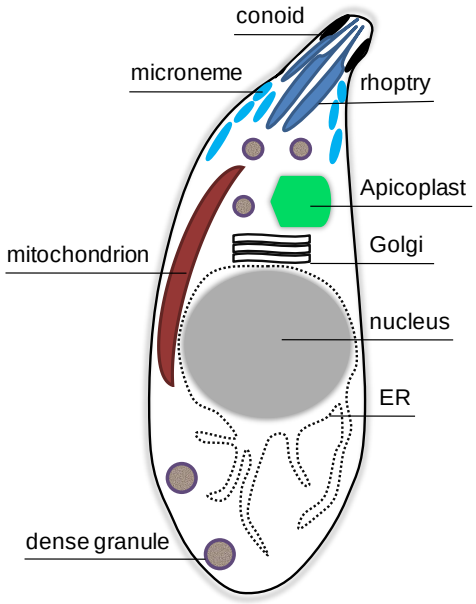

Figure 1. Schematic representation of *Toxoplasma gondii*.
The apical end of the parasite is pointed and includes organelles involved in the invasion process: the conoid, micronemes and rhoptries. Dense granules are mostly found also at the apical end of the parasite and are necessary for the intracellular survival and replication of the parasite. The apicoplast is a non-photosynthetic plastid-like organelle that is present in all members of the phylum *Apicomplexa*. The endoplasmic reticulum (ER) and Golgi are centrally located, while the single mitochondrion is principally found toward the apical end of *T. gondii*. Location of the nucleus depends on the developmental stage of the parasite (Dubey *et al.* 1998).

1.3 *Toxoplasma gondii*: life cycle, infective stages and transmission.

To better understand the molecular mechanisms by which *T. gondii* gains entry into host cells and establishes life-long infection, it is important to consider the parasite's life cycle and, in particular, the infectious stages involved. *T. gondii* has a typical apicomplexan life-cycle, which includes both sexual and asexual reproduction (Figure 2). The sexual cycle occurs only in the intestine of the primary host, which can be any member of the *Felidae* (cat) family. In the cat's intestine, *T. gondii* produces gametes that eventually form oocysts that are then excreted into the environment with the cat's faeces. Sporulation of oocysts occurs one to five days after excretion and results in the formation

of sporozoites. Sporozoites contaminate food and water sources, mediating transmission of *T. gondii* from the specific primary feline host to the non-specific secondary vertebrate host. When sporozoites are ingested by a secondary host, they invade intestinal epithelium and differentiate into tachyzoites by an unknown mechanism. Tachyzoites are rapidly dividing, motile form of *T. gondii* responsible for acute infection. Tachyzoites are capable of actively invading and replicating within multiple cell types of vertebrate hosts, which results in lysis of the infected cell (Dubey 1998a). After death of the host cell, newly released tachyzoites may invade neighbouring cells, often causing extensive tissue damage. Alternatively, tachyzoites that remain extracellular spread via the bloodstream to distant organs (Dubey 1998b; Garweg 2005) and actively transmigrate through the epithelial, blood-brain and placental barriers (Barragan and Sibley 2002). During acute infection, host innate immune response is activated against the tachyzoite form, leading to partial elimination of the parasite. However, some parasites survive, invading tissues of the brain, eyes and muscles, where they undergo stage conversion to bradyzoite form (Aliberti 2005). Bradyzoites are an encysted, slow-replicating form of *T. gondii*, responsible for establishment of a life-long, latent infection. The reverse transformation of bradyzoites back to tachyzoites may occur in immuno-compromised hosts, resulting in reactivated disease and toxoplasmic tissue lesions (Knoll and Boothroyd 1998). In addition to reactivation of existing infection, bradyzoites can induce *T. gondii* infection *de novo* in a new host after ingestion of the bradyzoite-filled tissue cysts (Tenter *et al.* 2000).

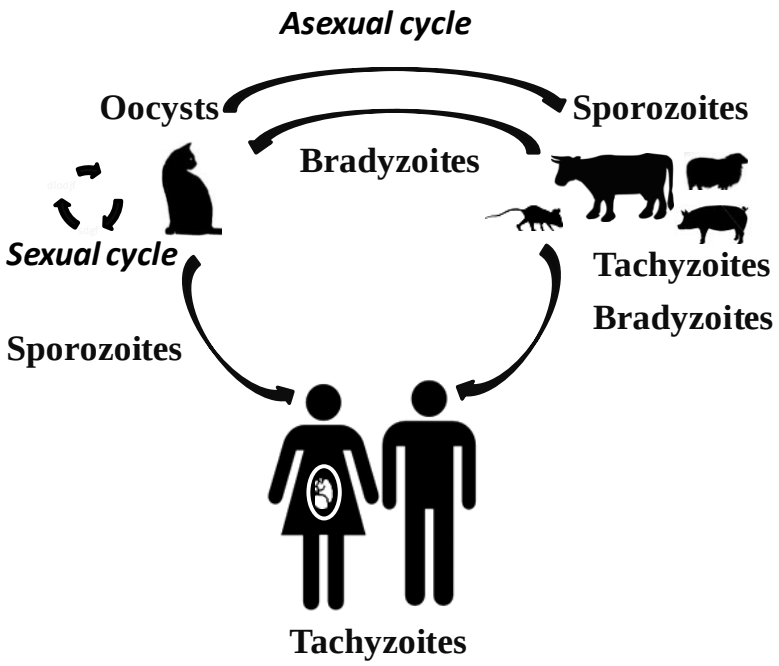

Figure 2. Life cycle and transmission of *Toxoplasma gondii.*
The life cycle of *T. gondii* includes sexual and asexual modes of proliferation and transmission. The sexual cycle takes place in the intestine of members of the cat family and results in the formation of oocysts that are shed into the environment with the cat's faeces. Oocysts then sporulate into highly infectious sporozoites that can be ingested by the secondary host, which can be a mouse, cow, sheep, pig or other vertebrate animal, including a human. Inside the secondary host, *T. gondii* propagates asexually via tachyzoite form, establishing acute infection, which may also result in infection of the fetus. After the acute phase, chronic infection develops due to the formation of bradyzoites, which are concentrated in the tissue cysts and are infectious upon ingestion.

Although all three forms of *T. gondii*, sporozoites, tachyzoites and bradyzoites, are recognized as being infectious, there are differences in the way the infection is established. For example, in the feline model, after ingestion of bradyzoite-filled tissue cysts, infection occurs in almost 100 % of cases, and after only three to ten days cats begin to shed oocysts; while after ingestion of tachyzoites or oocysts only about 30 % of animals shed oocysts after thirteen to eighteen days respectively (Barragan and Sibley 2002). This suggests that some forms of *T. gondii* are more effective at establishing infection than others. Surprisingly, all three infectious forms have very similar ultra

structural features (described in Figure 1) with only minor differences in the appearance of the rhoptries, and the number of micronemes and dense granules (Barragan and Sibley 2002). For example, the rhoptries are labyrinthine in tachyzoites, electron-dense in mature bradyzoites and mixed in sporozoites. Number of micronemes is the highest in bradyzoites, intermediate in sporozoites and lowest in tachyzoites, while dense granules are more numerous in sporozoites and tachyzoites (Barragan and Sibley 2002). Despite the structural similarity, sporozoites, tachyzoites and bradyzoites express unique sets of proteins on their surfaces, by which these forms can be identified (Jung *et al.* 2004; Lekutis *et al.* 2001). Some of these proteins play important role in *T. gondii* virulence and are discussed in details in section 1.6.

1.4 *Toxoplasma gondii* genotypes and infection.

T. gondii is a single species that consists of only three genetic lines, referred to as type I, II and III. Each type has a variety of strains, all of which share a common phenotype of the type they belong to. For example, type I strains are highly virulent and lethal and are mostly responsible for congenital toxoplasmosis (Barragan and Sibley 2002), while types II and III are relatively avirulent and tend to establish latent chronic infection (Grigg *et al.* 2001).

1.5 *Toxoplasma gondii* infection in humans.

Routes of infection

It has been estimated that up to a third of the adult population worldwide has been infected with this parasite (Tenter *et al.* 2000). Humans become infected with *T. gondii* after consuming undercooked meat infected with bradyzoites, after accidental ingestion

of oocysts and sporozoites shed into the environment by cats, or after transmission of tachyzoite form via blood transfusion, organ transplant, and transplacental route from mother to fetus (Figure 2). Toxoplasmosis is the disease caused by *T. gondii* (Joynson and Wreghitt 2001).

Seroprevalence of *T. gondii* infection in the population rises with age, which is attributed to ingestion of infected meat products. *T. gondii* infections can also be linked to a geographic location. One of the best documented *T. gondii* outbreaks is the recent waterborne outbreak in Victoria, Canada in 1994-95, where a total of 100 individuals acquired acute toxoplasmosis and twelve infants acquired congenital toxoplasmosis (Bowie and King 1997). This outbreak was linked to the contamination of a municipal water reservoir with *T. gondii* oocysts (or sporozoites). Analysis of this and other waterborne toxoplasmosis outbreaks revealed that waterborne *T. gondii* infections caused by the oocyst/sporozoite form of the parasite are clinically more severe than infections caused by ingestion of the bradyzoite form of *T. gondii*. It has not been established, however, whether the severity is related to the infective form of *T. gondii* involved in waterborn route of infection (sporozoite) or to the high dosage of oocysts (or sporozoites) ingested by the host (Jones and Dubey 2009).

Course of infection

The course of *T. gondii* infection in humans is affected by the following factors: immunological status, gender and genetic background of the patient; genetic type of the parasite, inoculum size and route of infection (Montoya and Liesenfeld 2004). That is, in immuno-competent healthy individuals, an acute *T. gondii* infection can produce only mild symptoms, such as chills, fever, headache, myalgia, fatigue and swollen lymph

nodes (lymphadenopathy), or it can be asymptomatic. More severe manifestations of acute *T. gondii* infection, such as inflammation of the ileum in the small intestine (ileitis), septic shock, inflammation of the brain (encephalitis), inflammation of the heart (myocarditis), inflammation of the liver (hepatitis), or inflammation of the retina and choroid in the eyes (retinochoroiditis) can occur in individuals with immune hyper-reactivity. After resolution of an acute infection, *T. gondii* establishes life-long infection without any symptoms in the presence of a functional healthy immune system. In immuno-compromised patients with organ transplants, cancers, or AIDS, untreated *T. gondii* infections result in severe manifestations, such as toxoplasmic encephalitis and ocular toxoplasmosis. It is estimated that *T. gondii* causes severe encephalitis in up to 40 % of AIDS patients worldwide (Tenter *et al.* 2000).

Congenital toxoplasmosis

A developing fetus is extremely vulnerable to *T. gondii* infection. Vertical transmission of *T. gondii* from mother to fetus can result in severe disease manifestations including spontaneous abortion, premature birth, or congenital toxoplasmosis, with the risk of fetal infection increasing as pregnancy progresses (Rorman *et al.* 2006). That is, if primary maternal infection occurs during the preconception period, the risk of fetal infection is 1 %, but if a woman becomes infected during the first trimester and the infection is left untreated, the risk of congenital toxoplasmosis elevates to 25 %. Infection during the second and third trimesters correlates with the incidence of fetal infection as high as 30–54 % and 60–65 %, respectively (Rorman *et al.* 2006). Overall, the worldwide incidence of congenital toxoplasmosis is roughly 1 per 1000 births (Tenter *et al.* 2000). The classic triad of congenital toxoplasmosis includes retinochoroiditis, intracranial calcifications,

and hydrocephalus (accumulation of the CSF in the cavities of the brain), while other symptoms may include convulsions, an enlargement of the spleen (splenomegaly) and liver (hepatomegaly), fever, anaemia, jaundice, lymphadenopathy and progressive mental retardation (Tenter *et al.* 2000). Infected newborns without any symptoms of toxoplasmosis at birth are likely to develop eye disease due to toxoplasmosis later in life (Koppe *et al.* 1986).

Link to psychiatric disorders

Recently, latent *T. gondii* infections in immuno-competent individuals, have been linked to an increased incidence of serious psychiatric conditions, such as bipolar disorder and schizophrenia (Yolken and Torrey 2008). The mechanism for altered brain activity during *T. gondii* infection has been recently investigated by Gaskell *et al.* (Gaskell *et al.* 2009). It was found that *T. gondii* produces tyrosine hydroxylase, a crucial enzyme in the production of dopamine. Since dopamine plays important roles in mood, sleep patterns, sociability, attention, motivation and learning, changes in the concentration of dopamine in the brain has been linked to neurological conditions such as Parkinson's disease and attention deficit disorder (Gaskell *et al.* 2009). In addition, the link between altered dopamine production and schizophrenia has long been established (Elevates 2008). Chronically infected with *T. gondii* humans are also found to be at higher risk of traffic accidents due to the delayed reaction times (Flegr 2007).

Based on the clinical discoveries noted above, chronic *T. gondii* infections may present higher risk than previously estimated and further investigation may lead to new treatment options for patients suffering from serious psychiatric conditions and chronically infected with *T. gondii*.

Diagnostics of toxoplasmosis

Current diagnostic methods such as serological methods, PCR, hybridization, isolation and histology, are successful at identifying *T. gondii* infection. However, diagnosing infection in pregnant women remains challenging since it is necessary to establish not only the fact of infection, but also when the primary infection first occurred with regards to gestation. Currently, two tests are used for diagnosis of acute and recent infection in pregnant women. First, the IgM test is performed in order to identify whether infection is present. The second test, IgG-avidity index assay, determines how recently the infection was acquired. That is, low IgG-avidity antibodies indicate recent infection. However, some women have long-lasting, low IgG-avidity antibodies, which results in false positive outcomes (Montoya and Liesenfeld 2004). While this project is not concerned with the development of new diagnostic methods for *T. gondii* infections, the detailed molecular information about the surface antigens specifically expressed on different infective forms of the parasite may result in the development of new diagnostic techniques.

Treatment of Toxoplasma gondii infections

Current treatment options include both pyrimethamine and sulfadiazine for acute, severe toxoplasmosis, while asymptomatic disease is currently recommended left untreated (Montoya and Liesenfeld 2004). The main limitation of widely-used antibiotic treatment strategies is that these drugs do not reach the encysted parasites of chronically infected patients in sufficient concentration, and, therefore, they cannot be used for treatment of chronic *T. gondii* infections.

The current vaccine research is focused on the development of a vaccine that can induce Th1 and humoral (including IgA) responses in order to establish life-long immunity to *T. gondii* (Montoya and Liesenfeld 2004). Therefore, studying *T. gondii* antigens may be beneficial to the development of such a vaccine.

1.6 *Toxoplasma gondii* surface proteins.

T. gondii is a very successful parasite due to its ability to infect any nucleated cell and to establish life-long chronic infection in its host (Ajioka *et al.* 2007). In order to infect, *T. gondii* first adheres to the host cell laterally. This initial adhesion is reversible due to unknown reasons. After successful lateral adhesion, the parasite irreversibly attaches apically. And finally, actively invades the host cell (Carruthers and Boothroyd 2007).

Some of the molecules involved in the parasite's attachment and invasion have been well described. Among them are MICs, a group of proteins secreted to the surface of the parasite by micronemes (for the location of micronemes see Figure 1). Some MICs include recognizable adhesive domains to directly mediate irreversible attachment, while others assemble into protein complexes with other MICs and different proteins to facilitate invasion (Zhou *et al.* 2005). For example, the apical membrane antigen, AMA-1, forms a complex with rhoptry neck proteins, RON2 and RON4, to mediate invasion of the parasite (Alexander *et al.* 2005).

Less is known about the molecular mechanism of *T. gondii* initial reversible adhesion to the host cell. It has been suggested that it involves recognition of the host cell surface receptors by *T. gondii* GPI-anchored surface proteins, collectively termed Surface Antigen Glycoprotein (SAG) Related Sequences, or SRS (Jung *et al.* 2004), which are the focus of this research.

1.6.1 Discovery of SRS superfamily.

The first SRS proteins were discovered through the use of mAbs and radio-iodination. Among the first members of the SRS superfamily to be discovered were SAG1, SAG2A and SAG3 by Burg *et al.* in 1988, Cesbron-Delauw *et al.* in 1994 and Prince *et al.* in 1990, respectively. Subsequently, SAG4A, SRS1 and BSR4 were cloned by Odberg-Ferragut *et al.* in 1996, Hehl *et al.* in 1997 and Knoll and Boothroyd in 1998 (Lekutis *et al.* 2001). Using the *T. gondii* expressed sequence tag (EST) database, more SRSs were discovered, including SAG4 (Manger *et al.* 1998), SRS2, SRS3 and SRS4 (Ajioka *et al.* 1998). By 2001, a total of 21 SRS homologs were identified and subsequently classified based on their homology to the prototypic members of the superfamily, SAG1 and SAG2. The SAG1 family was defined as those genes that include the 12 conserved cysteine residues of SAG1 and 24 to 99 % of sequence identity (Jung *et al.* 2004). Alternatively, the SAG2 family included SRSs that share only a subset of the cysteine residues and only about 20 % identity (Lekutis *et al.* 2001).

With the recent release of the *T. gondii* genome, over 160 unique SRS sequences have been bioinformatically identified, based on the following parameters: 20-99 % sequence identity, conserved cysteine and proline residues, conserved tryptophan residues in the membrane proximal domain, signal peptide and GPI-addition sequences, and significant homology centred around the GATLTI and CSYGAD motifs (Jung *et al.* 2004). A sequence alignment of the representative SRSs in Figure 3 illustrates these conserved regions.

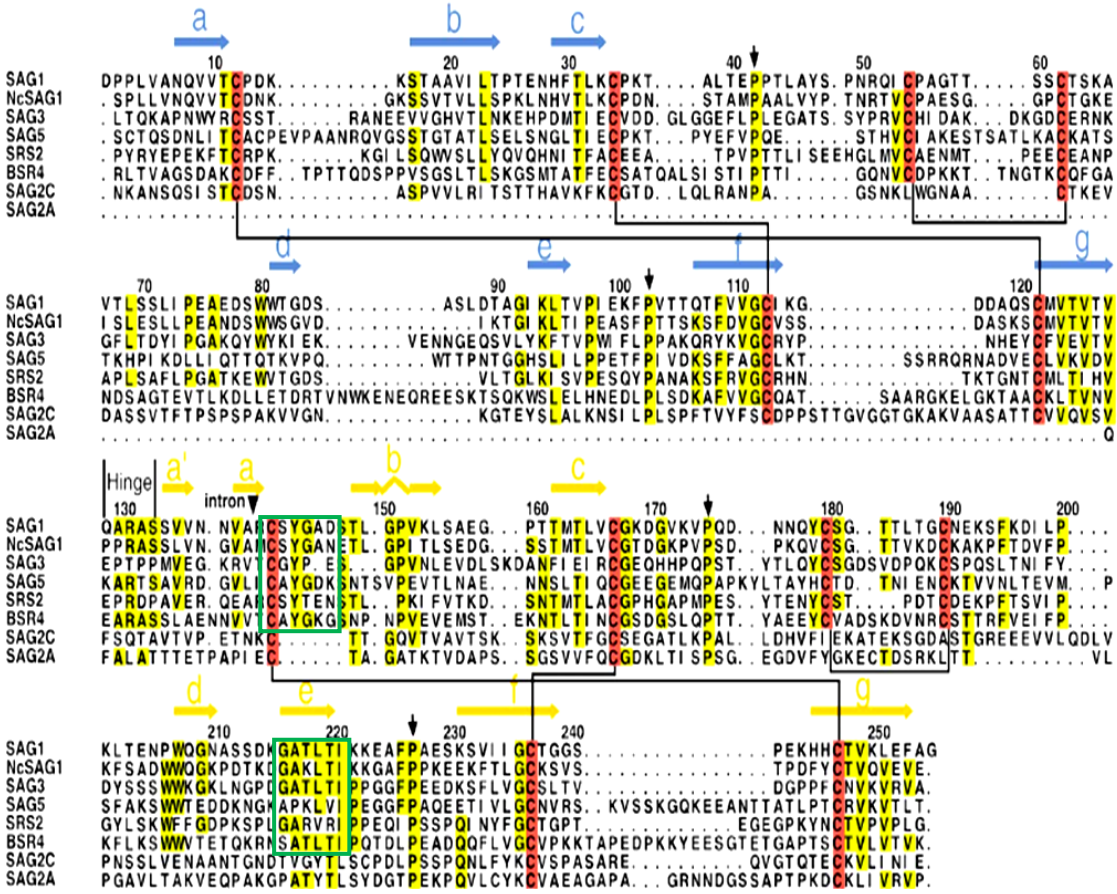

Figure 3. Sequence alignment of the representative members of SRS superfamily.
The disulfide bond-forming cysteine residues are boxed in red. The positions where more than four out of eight sequences are identical are shaded in yellow. The beta-strands in the D1 and D2 domains are blue and orange, respectively. Conserved motifs are boxed in green. Figure adapted with permission from (He *et al.* 2002).

1.6.2 Stage specific expression of SRS proteins.

Many infectious organisms have the ability to express variant antigens on their surface in a stage-specific manner, which is often referred to as antigenic variation. This feature promotes chronic infection in a host thereby increasing the chance of parasite transmission. Antigenic variation is achieved by either DNA recombination, as in African trypanosomes (Vanhamme *et al.* 2001) and *Neisseria* spp., or by activation and deactivation of promoters regulating individual genes in the active expression sites, as in *Candida, Plasmodium* spp. (Deitsch *et al.* 1997).

Although the mechanism of antigenic variation in *T. gondii* has not been established, it is known that each infective form of *T. gondii* (tachyzoite, bradyzoite, and sporozoite) expresses a unique, largely non-overlapping repertoire of the surfaces proteins, many of which are antigens (Jung *et al.* 2004). For example, bradyzoites are covered mainly with BSR4, SRS9 and SRS6, while tachyzoites are dominated by SAG1, SAG2 and SRS1–3. There are also shared SRSs, which are equally present on the surface of tachyzoites and bradyzoites, such as SAG3 (Kim and Boothroyd 2005; Kim *et al.* 2007).

Phylogenetic analysis of SRSs shows that genes expressed in a stage-specific manner cluster together as closest paralogs (Crawford *et al.* 2009a). The neighbour-joining tree in Figure 4 constructed from representative members of each of the three *T. gondii* life cycle stages reveals the evolutionary relationship between these SRSs. Such clustering pattern may indicate stage-specific sequence and, perhaps, structural features that likely play an important role in the biology of *T. gondii* infection, dissemination and pathogenesis.

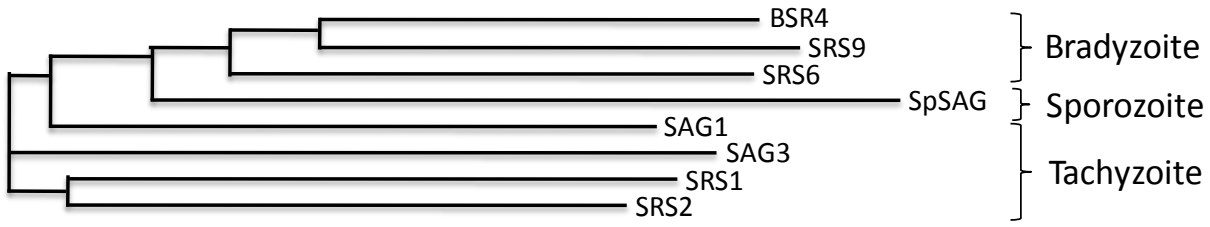

Figure 4. Evolutionary relationship between representative SRS family members from the three life cycle stages of *T. gondii*.
Phylogenetic analyses of representative SRSs using CLUSTALW reveals clustering of SRSs expressed at the same infective stage. Figure adapted with permission from (Crawford *et al.* 2009a).

1.6.3 SRS fold.

Prior to the results described in this thesis, SAG1 was the only known SRS out of the 160+ member superfamily for which a crystal structure had been determined. Structural characterization of SAG1 revealed that the N-terminal (D1) and C-terminal (D2) domains share a similar secondary structure (He *et al.* 2002). Most importantly, individual SAG1 domains did not bear evolutionary relationship to any known protein. Therefore, SAG1 tandem domains were defined as members of a new structural superfamily termed 'the SRS fold' (He *et al.* 2002) described in details in section 3.1. Besides structural definition of the SRS fold, SAG1 study had also established the homodimeric configuration of recombinant SAG1. According to the authors of the SAG1 study, this configuration is the most logical considering the fact that SRSs are attached to the membrane of the parasite via GPI-anchor. The biological relevance of SAG1's dimeric configuration was supported by the evidence that native SAG1 purified as a dimer from the surface of tachyzoites (He *et al.* 2002).

1.6.4 Biological role of SRS proteins.

Although precise biological roles of SRSs have not been identified, it has been shown that some SRSs may play dual roles of adhesins, mediating *T. gondii* attachment to the host cell, and antigens, inducing strong immune responses (Kim and Boothroyd 2005).

SRS adhesins

Among SRS adhesins, tachyzoite expressed SAG1 and SAG3 are the most studied. Several *in vitro* studies demonstrated that SAG1 is able to adhere directly to various host cells, such as bovine kidney cells, human foreskin fibroblasts (Robinson *et al.* 2004), monocytes (Channon *et al.* 1999) and epithelial cells (Velge-Roussel *et al.* 2001). However, the host cell receptor for SAG1 has not yet been identified. To date, SAG3 is the only member of the SRS superfamily for which a putative host cell ligand has been suggested (Jacquet *et al.* 2001). Jacquet *et al.* demonstrated that binding of tachyzoites to the Chinese hamster ovary cells (CHO-K1) involves SAG3, which interacts with the cell-surface sulphated proteoglycans, in particular heparan sulphate proteoglycans (HSPGs), such as heparin and heparan sulphate (Jacquet *et al.* 2001). These interactions between SAG3 and HSPGs were shown to mediate the parasite attachment to target cells (Jacquet *et al.* 2001).

The molecular basis for SAG3- HSPGs was initially proposed by He *et al.* in the SAG1 study, where modeling of SAG3 onto the SAG1 dimer structure revealed that the putative ligand binding groove at the dimer interface of the N-terminal domains accommodates heparin. Moreover, based on the multiple sequence alignment of major SRSs, the authors predicted that this positively charged groove is conserved among SRSs and potentially serves as a sulphated proteoglycan binding site on target cell surfaces (He *et al.* 2002).

SRS antigens

Tachyzoite SAG1 and SAG2A were found to function as antigens (Kim *et al.* 2007) dominating antibody response to *T. gondii* during acute infection (Kim and Boothroyd 2005). Other studies determined that SAG1 induces not only massive antibody response early after oral infection, but also a strong B and T cell-mediated immunity in mouse (Rachinel *et al.* 2004; Velge-Roussel *et al.* 1994) and human infections (Beghetto *et al.* 2003; Brenier-Pinchart *et al.* 2006). Recently, SAG1 was co-crystallized with a monoclonal antibody, obtained after immunization of mice with tachyzoites (Graille *et al.* 2005). The epitope for antibody binding was found on the surface loops of the D1 domain and did not overlap with the putative ligand binding groove identified in the first SAG1 study by He *et al.* The results of these two SAG1 studies rationalized the dual role of SAG1 as adhesin and antigen. Other than SAG1, tachyzoite-specific SRS2 was found to be highly immunogenic in mice and to induce protective immunity against *T. gondii* (Crawford *et al.* 2009b). On the other hand, SRSs expressed on the surface of *T. gondii* bradyzoites were found to lack immune response in natural infection (Kim and Boothroyd 2005; Kim *et al.* 2007).

1.6.5 Orthologs of SRS proteins in other Apicomplexan parasites.

T. gondii is a model member of phylum Apicomplexa, which includes *Plasmodium*, the ethiologic agent of malaria, and *Cryptosporidium*, a devastating AIDS pathogen (Ajioka *et al.* 2007; Donald *et al.* 2007). Several SRS superfamily orthologs have been discovered in *Neospora caninum* (Nc), such as BSR4 and SRS2. NcBSR4 shares 66 % sequence identity with *T. gondii* (Tg) BSR4 and is expressed mostly on the surface of *N. caninum* bradyzoites similarly to TgBSR4 (Risco-Castillo *et al.* 2007). NcSRS2 shares 53 %

sequence identity with TgSRS2 and is also a tachyzoite-specific antigen (Howe *et al.* 1998; Howe and Sibley 1999). Orthologs of the SRS superfamily have also been found in *Sarcocystis* (Sn), such as SnSAG1, SnSAG2, SnSAG3, and SnSAG4 (Howe *et al.* 2005)

It was previously thought that SRS orthologous genes were restricted to tissue-dwelling coccidian, such as *Neospora* and *Sarcocystis* (Jung *et al.* 2004); however, TgSRSs were found to share a vertical relationship with a 6-Cys domain family from *Plasmodium falciparum* (Pf) (Cowman and Crabb 2006). Similarly to TgSRSs, the 6-Cys domain family is expressed on the surface of the parasite in a stage-specific manner and has a predicted domain structure conserved with TgSAG1, the SRS fold. The fact that the SRS fold is conserved among several species may indicate that the SRS superfamily has an essential, conserved function(s) (Cowman and Crabb 2006). Therefore, molecular characterization of the SRS superfamily will allow a more comprehensive understanding of the pathogenesis of not only toxoplasmosis, but also other infectious diseases, such as malaria, caused by *Plasmodium*, and human sarcosporidiosis caused by *Sarcocystis bovihominis* and *Sarcocystis suihominis*.

1.7 Research hypotheses and objectives.

T. gondii is an intracellular parasite with a remarkable ability to infect any vertebrate animal and any nucleated cell. The initial adhesion of *T. gondii* to a host cell is a prerequisite step in establishing infection and is predicted to involve the SRS superfamily. To date, only SAG1 has been structurally characterized revealing a novel SRS fold. A multiple sequence alignment (Figure 3) reveals several insertions and deletions amongst the SRS family members with the longest insertions found in BSR4,

expressed on the surface of bradyzoites, and almost absent in SAG1 and SRS2, expressed on the surface of tachyzoites. Based on this, we hypothesize that:

1. The core SRS fold is conserved among members of the SRS superfamily, though developmentally expressed SRSs are most likely to display structurally divergent loops.

Previous study by Jacquet *et al.* demonstrated possible interactions between HSPGs and SAG3. These interactions were rationalized in the SAG1 structural study, which revealed a positively charged putative ligand binding groove that ideally suited to bind a glycan moiety. Although the multiple sequence alignment suggests that the positively charged residues which participate in the groove formation are conserved among SRSs, the localization of the groove at the dimer interface of N-terminal domains implies that dimerization of SRSs is required for the groove formation. Based on this, we hypothesize that:

2. SRS proteins coordinate HSPGs via homodimeric architecture.

Identification and characterization of the structural features that enable *T. gondii* to adhere to host cells in order to establish infection may contribute to the development of the therapeutic or prophylactic interventions and to limit infectivity of this widespread pathogen.

Chapter 2: Materials and methods.

2.1 Materials.

Primers (IDT); PCR mix and PCR grade water (Pwo Master, Roche); agarose (EMD); ethidium bromide (BioRad); restriction enzymes (NEB); BSA (NEB); digest buffers (NEB); QIAquick® PCR Purification Kit (QIAGEN); Antarctic Phosphatase (NEB); T4 DNA ligase (NEB); T4 DNA ligase buffer (NEB); LB broth (EMD); competent cells (Novagen); QIAprep® Spin Miniprep Kit (QIAGEN); Plasmid Midi Kit® (QIAGEN); pACgp67a and b vectors (Pharmingen); *Spodoptera frugiperda* insect cell line (Orbigen); *Hi*5 insect cell line (Orbigen); tissue culture plates (CellStar); BaculoGold Bright Linearized Baculovirus DNA (Orbigen); Sf900 insect cell media (Gibco); Express Five SFM (Gibco); L-glutamine (Gibco); Cellfectin (Invitrogen); Gentamicin (Gibco); Ampicillin (Fisher Bioreagents); trypan blue stain (Invitrogen); Nickel-NTA agarose beads (QIAGEN); Hepes (SIGMA); sodium chloride (SIGMA); imidazole (SIGMA); Tunicamycin (Calbiochem); DMSO (SIGMA); filter paper (Millipore); Centricon spin concentrators (Millipore); thrombin (Invitrogen); calcium chloride (SIGMA); nickel sulfate (SIGMA); SDS (SIGMA); Precipitant Synergy 64 (Emerald BioSystems); WIZARD I/II (Emerald BioSystems); Index (Hampton Research); Peg Ion (Emerald BioSystems); 96-well crystallization plates (Emerald Biosystems); nitrocellulose membrane (Life Sciences); Tris-HCl (Calbiochem); Tween-20 (Calbiochem); Alexafluor 680/Strepavidin conjugate (Molecular probes/Invitrogen); Biotin-NTA (Molecular probes/Invitrogen); heparin agarose beads (SIGMA); heparin sodium salt (SIGMA); PhastGel (GE); Phast System Separation and Control Unit (Pharmacia); mucin II from

porcine stomach (SIGMA); hyaluronic acid sodium salt from *Streptococcus zoepidemicus* (SIGMA); heparan sulfate sodium salt from bovine kidney (SIGMA); chondroitin sulfate C sodium salt from shark cartilage (SIGMA); glycogen type IX from bovine liver (SIGMA); amylopectin (SIGMA).

Materials Contributed by Others

T. gondii genomic DNA was kindly provided by Dr. Michael Grigg and Dr. Laura Knoll. BSR4 construct was cloned by Dr. Martin J. Boulanger. Vector designed for co-expression of the dimerization coils with protein of interest was a gift from Dr. Erin Adams.

2.2 General methods.

2.2.1 DNA manipulation.

PCR reactions

Each PCR reaction contained 500 nM of each forward and reverse primer, 20 ng of template DNA, 22 µl of PCR grade water and 25 µl of PCR mix (Pwo SuperYield DNA polymerase, 4 mM $MgCl_2$, dATP, dCTP, dGTP, dTTP, each 0.4 mM). PCR product was visualized on a 1 % (w/v) agarose gel with ethidium bromide (EtBr) using the EagleEye II system. All products were purified with QIAquick® PCR Purification Kit according to the manufacturer's protocol.

Restriction digestion

Purified PCR product and the vector were digested with appropriate enzymes (5 U of enzyme per 1 µg of DNA) in 10 x NEBuffer supplemented with 10X bovine serum albumin (BSA). The reaction was incubated for 2 hours (hr) at 37 ° C. Correct digestion was confirmed by separation of products on a 1 % (w/v) agarose gel with EtBr and

visualization using the EagleEye II system. To remove 5' phosphates and prevent recircularization, digested vector was treated with Antarctic Phosphatase at 5 U per 1 µg of the vector DNA, in the presence of 10X Antarctic Phosphatase Buffer for 30 minutes (min) at 37° C, followed by heat inactivation of the enzyme at 65° C for 5 min. All digested DNA fragments were finally purified with QIAquick® PCR Purification Kit according to the manufacturer's instructions.

Ligation

Ligation of the digested and de-phosphorylated vector (100 ng) and digested insert (three molar excess of insert to vector) was carried out using 10 U of T4 DNA ligase and 10x T4 DNA ligase buffer in the final volume of 20 µl, at 16° C, overnight.

Competent cell preparation

An overnight bacterial culture (1 ml of DH5α for plasmid propagation) was inoculated into 100 ml of Luria-Bertani (LB) broth and incubated at 37° C with shaking until the OD_{600} reached approximately 0.55. The resulting culture was then divided into 4 x 25 ml sterile centrifuge tubes, and cells were harvested by centrifugation at 8,000 rpm for 5 min at 4° C. After removal of the supernatant, each pellet of cells was resuspended in 6.25 ml of cold 100 mM $MgCl_2$ and centrifuged at 8,000 rpm for 5 min at 4° C. The supernatant was then removed. Each pellet was resuspended in 12.5 ml of 100 mM $CaCl_2$ and incubated on ice for 45 min. Centrifugation, removal of supernatant and resuspention in $CaCl_2$, and incubation were repeated, and followed by addition of 40 % glycerol to each of the tubes to the final concentration of 15 % glycerol. Cells were then aliquoted (50 µl into 1.5 ml tube), flash frozen in liquid nitrogen and stored at -80° C.

Transformation and isolation of plasmid DNA

Thawed competent cells (50 µl) were incubated with 5 µl of the ligation mixture on ice for 30 min, followed by a heat shock at 42° C for 45 seconds (sec) and incubation on ice for an additional 2 min. LB broth (500 µl) was added to the transformation mixture, which was incubated with shaking at 37° C for 45 min. Bacterial suspension (100 µl) was plated on LB agar plates with selective antibiotic, followed by incubation at 37° C overnight. Successfully transformed plasmids were identified by growing bacterial cultures in 5ml LB media with appropriate antibiotic overnight with shaking at 200 rpm at 37° C, followed by preparation of plasmid DNA using the QIAprep® Spin Miniprep Kit and finally by digestion with appropriate enzymes. Digested plasmids were confirmed by separation of products on a 1 % (w/v) agarose gel with EtBr and visualization using the EagleEye II system. Plasmids which contained appropriate size inserts were then sent for sequencing to the DNA sequencing facility at the Center for Biomedical Research (University of Victoria, BC). DNA sequence data was analysed using BLAST server (Sayers *et al.* 2008). Once the clones were verified, the high quantity and quality of mutation free open reading frame (ORF) plasmid DNA required for protein expression studies in the Baculovirus insect cell system was produced using Plasmid Midi Kit®.

2.2.2 Protein expression and purification.

All proteins of interest discussed in this work were produced using Baculovirus Expression Vector System (BEVS), established for eukaryotic protein expression in Dr. Boulanger's lab.

Media Preparation

Media for *Sf*9 cells was prepared by addition of 100 µl of gentamicin (10 µg/ml final concentration) to 500 ml of SF900 media. Media for *Hi*5 cells was prepared by addition of 200 µl of gentamicin and 90 mL of L-glutamine to 1 L of Express Five SFM.

Target protein virus production

Target genes were first cloned (as described in 2.1.1) into engineered transfer vectors, pACgp67a and b, which contain the gp67 signal sequence in front of the multiple cloning site (MCS) to force secretion of the recombinant protein (Pharmingen 1999). The transfer vector with the gene of interest and linearized baculovirus DNA were then co-transfected into Sf9 insect cells to generate primary virus. Prior to co-transfection, 2 ml of Sf9 cells at $1x10^6$ cells/ml were seeded into a well of a 6-well tissue culture plate (2×10^6 cells per well) and incubated at 27^o C for 30 min without shaking, followed by a wash with 2 ml of fresh Sf900 media. To prepare DNA for co-transfection, 0.5 µg of linearized baculovirus DNA was combined with 2 µg of the transfer vector containing the gene of interest, mixed well, and incubated for 5 min at room temperature (RT) of 23° C. This DNA mixture was added to 100 µl of Sf900 media pre-mixed with 10 µl of Cellfectin, and incubated for 30 min at RT of 23° C. After the 30 min incubation, 1.8 ml of SF900 media was added to the mixture. This mixture was then added to the previously washed Sf9 cells in the 6-well tissue culture plate and incubated for 7 hrs at 27^o C, without shaking. After 7 hr incubation, transfection media was removed. Sf9 cells were washed once with fresh Sf900 media with gentamicin at 10 µg/ml, finally adding 2.5 ml of Sf900 media with gentamicin at 10 µg/ml. The plate was wrapped in Saran wrap to prevent

evaporation placed in a plastic container, and incubated for 7 days. After 7 days, the primary virus (P1) was harvested and used for virus amplification.

Virus amplification

To amplify virus, 40 µl of primary transfection supernatant was added to 50 ml of Sf9 cells at 1.5 x 10^6 cells/ml in Sf900 media containing gentamicin and incubated at 27^o C with shaking. The number of both dead and viable cells was noted every day to maintain viable cell count at 1.5 x 10^6 cells/ml. Viability was checked by mixing cells with trypan blue stain. Once the cell density was below 0.5 x 10^6 viable cells/ml (5-7 days), secondary transfection supernatant (P2) was harvested.

Test protein expression

To determine the optimal amount of the virus to add for maximal expression in *Hi*5 cells, 4 tissue culture plate wells were plated with 2.5 ml of *Hi*5 cells at 1.8 x 10^6 cells/ml, followed by addition of 3 µl, 8 µl or 20 µl of P2 virus to each well and incubation at 27^o C with shaking for 72 hrs, including a negative control well containing no virus. After incubation, the contents of each well were harvested by centrifugation at 3,000 rpm for 4 min. 1.8 ml of the supernatant was next transferred into a fresh 2 ml tube. 200 µl of 10X Binding buffer (200 mM Hepes; 1.5 M NaCl; 200 mM imidazole; pH 8.0) was added, followed by 25 µl of nickel-agarose beads slurry. The mixture was incubated on the rocking platform for 1 hr at 4^o C. To pellet agarose the mixture was centrifuged at 5,000 rpm for 1 min and subsequently the supernatant was removed. In the next step, the beads were washed with 30 µl of 1X binding buffer (20 mM Hepes; 150 mM NaCl; 20 mM imidazole; pH 8.0), centrifuged at 5,000 rpm for 1 min, and the supernatant was removed. Bound protein was then eluted by addition of 20 µl of high imidazole buffer (20 mM

Hepes; 150 mM NaCl; 500 mM imidazole; pH 8.0). To visualize the amount of protein produced at different concentration of virus, 20 µl of supernatant from each tube was removed, mixed with of 4X SDS, heated to 95° C for 5 min. Loaded onto 15 % SDS PAGE and run at 180 V for 1 hr.

2.2.3 Carbohydrate binding macro-array.

The fluorescent label was prepared by mixing 50 µl of 10 mg/ml $NiSO_4 \cdot 6H_2O$ with pre-mixed 1mg Alexafluor680-Streptavidin (Cedarlane/Anaspec) and 1.5 mg Biotin-NTA in 1ml of Tris-Buffer (5mM Tris-HCl, pH 8.0). The mixture was incubated in the dark for 10 minutes at RT. Free nickel, free NTA-Biotin and free Alexafluor680-Streptavidin were separated from the nickel-NTA-Biotin-Strepatavidin-Alexafluor680 complex using desalting pre-packed Sephadex G-25 column (GE Healthcare/Amersham).

One µl of 1% or 5% solutions of carbohydrates, proteoglycans, and glycosaminoglycans were spotted onto a nitrocellulose membrane and air dried overnight.

The membrane then was blocked with 10 mL of blocking buffer (20mM Tris-HCl; 1 % BSA; 0.5 % Tween-20; pH 7.5) with shaking at 18° C in the dark. After 1 hr incubation, 5 ml of fresh blocking buffer with 100 µg of the protein, or just blocking buffer for negative control, was added and incubated for 1 hr with shaking at 18° C in the dark, followed by a double wash with 10 ml of Tris-buffer (20mM Tris-HCl; pH 8.0) to remove all unbound protein. Fluorescent label (3.5 µl) was added in 5 ml of fresh blocking buffer. The mixture was incubated for 40 min with shaking at 18° C in the dark. Final incubation was followed by triple wash with 10 ml of Tris-buffer. The blots were air dried in the dark for 20 min and imaged using the Licor/Odyssey system.

2.2.4 Heparin binding assays.

Heparin-agarose pull-down

Heparin-agarose beads slurry (30 µl) was pre-equilibrated in binding buffer (20mM Hepes; 50mM NaCl; pH 7.5) and added to 10 µg of each protein to a final reaction volume of 600 µl. The mixture was then incubated at RT for 45 min with inverting every 10 min and subsequent centrifugation at 1,600 rpm for 5 min. The supernatant was removed. The beads were washed three times with cold binding buffer and 15 µl of 4X SDS added to the beads. Samples were heated to 95° C for 5 min, centrifuged at 1,600 rpm for 5 min, loaded onto 15 % SDS PAGE and run at 180 V for 1 hr.

Native gel electrophoresis

Proteins of interest (4µg of protein per reaction) were mixed with solubilised heparin at various molar ratios (protein to heparin as 1:2; 1:5; 1:10) in the presence of binding buffer (20mM Hepes; 150mM NaCl; pH 7.5). The reaction mixture was then incubated at RT for 30 min. After incubation, 4 µl of the reaction mixture was loaded into each well of the native gel (PhastGel, 8-25 KDa, GE) using the Phast System Separation and Control Unit (Pharmacia) and run until 268Avh was reached, according to manufacturer instructions. All experiments were performed in triplicate.

2.2.5 Bioinformatics.

Phylogenetic analysis

The neighbour joining tree was constructed using the CLUSTALW (version 1.83) (Thompson *et al.* 1994) by the method of Saitou and Nei (Saitou and Nei 1987).

Multiple sequence alignments

Sequence alignments were done using CLUSTALW (Thompson *et al.* 1994). Conserved regions were identified and illustrated using ESPRIPT (Gouet *et al.* 2003).

Amino acid polymorphism maps

Polymorphism maps were created by alignments of sequences from different types and strains of *T. gondii*. Amino acid differences were identified in the sequence alignments as amino acid polymorphic sites. Polymorphic sites were then mapped onto the structures of SAG1, BSR4 and SRS2 and depicted as spheres using PYMOL (Delano 2002).

2.3 Target plasmid constructs and cloning.

2.3.1 BSR4 plasmid.

The coding region of *bsr4* consists of 398 amino acids with a signal-peptidase cleavage consensus site at position between amino acids 37 and 38 (Dyrløv Bendtsen *et al.* 2004). In addition, two conserved SAG domains (Figure 5) were identified in the *bsr4* sequence using BLAST search (Marchler-Bauer and Bryant 2004).

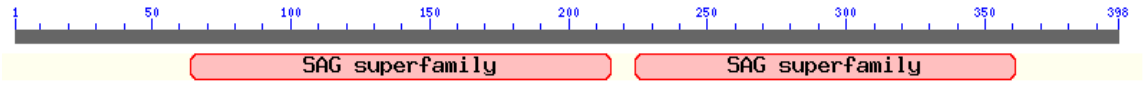

Figure 5. Conserved SAG domains identified in *bsr4* sequence.
Full *bsr4* sequence depicted as a ruler with conserved tandem SAG domains, identified using BLAST (Marchler-Bauer and Bryant 2004).

The *bsr4* sequence was amplified using forward primer: 5' ATCGGATCCCGTGGAGGTGACTTCAAGGC 3'; and reverse primer: 5' AGCTCTAGAGTGATGGTGATGGTGATGGGCTTTGACAGTTACCAGC 3') from *T. gondii* type I, RH, genomic DNA, to contain the 919 bp (amino acids 58 to 363) that

encode only the SAG tandem domains without the targeting sequence. The amplified region was cloned into BamHI and XbaI cut vector pACgp67a.

2.3.2 SRS2 plasmid.

The coding region of *srs2* consists of 372 amino acids with a signal-peptidase cleavage consensus site between amino acids 52 and 53 (Dyrløv Bendtsen *et al.* 2004). Two conserved SAG domains (Figure 6) were identified in the *srs2* sequence using BLAST search (Marchler-Bauer and Bryant 2004). Therefore, the 771 bp DNA product (amino acids 55- 311) utilized in this study encodes only the SAG tandem domains without the targeting sequence.

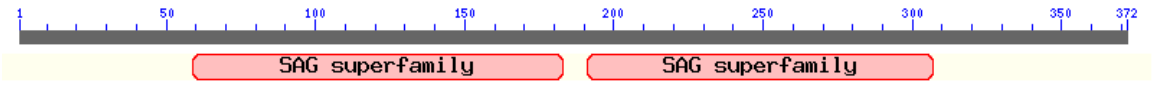

Figure 6. Conserved SAG domains identified in *srs2* sequence.
Full *srs2* sequence depicted as a ruler with conserved tandem SAG domains, identified using BLAST (Marchler-Bauer and Bryant 2004).

The *srs2* from *T. gondii* type I, GT1 genomic DNA was amplified using forward primer: 5’ CAGACTCCATGGGACCGTACAGATACGAGCCTG 3’; and reverse primer: 5’ GTGCCAGTACAACTAGGAGGTGCGGCCGCCAGACAGT 3’). The amplified region was then cloned into NcoI and NotI cut pAcGP67b vector previously modified to contain thrombin sites directly upstream and downstream of the NcoI and NotI sites respectively and a downstream hexahistidine (6X His) tag (Figure 7).

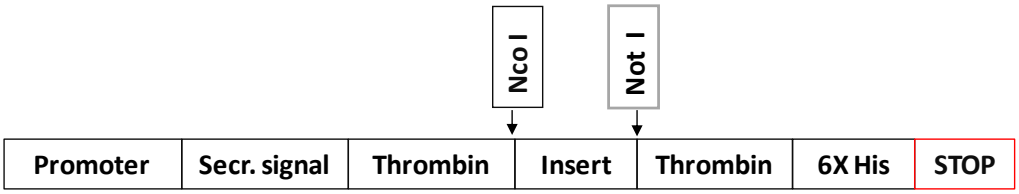

Figure 7. Schematic representation of the modified baculovirus expression vector designed for secreted expression of SRS2.
The baculovirus expression vector contains a polyhedrin promoter for high level recombinant *srs2* expression, a gp67 secretion signal for secreted recombinant protein expression, a cloning site for cloning of recombinant genes (insert), hexahistidine tag (6XHis) for purification and two thrombin sites to facilitate crystallization.

2.3.3 Dimerization plasmids.

To allow for co-expression of the recombinant protein of interest and dimerization domain, previously amplified genes, *bsr4* and *srs2*, were cloned into NcoI and NotI cut pACgp67b vector previously modified to contain a downstream dimerization domain (Blondel and Bedouelle 1991) connected by a glycine-rich linker. The full construct is outlined in Figure 8.

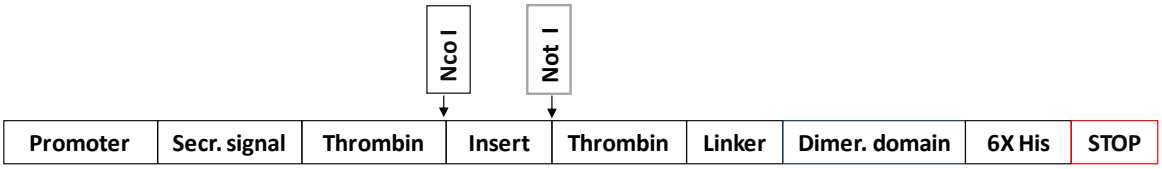

Figure 8. Schematic representation of the modified baculovirus expression vector designed for secreted co-expression of target protein with dimerization domain.
The baculovirus expression vector contains a polyhedrin promoter for high level recombinant protein expression, a gp67 secretion signal for secreted recombinant protein expression, a cloning site for cloning of recombinant genes (insert), a glycine-rich linker for flexibility, a dimerization domain, a hexahistidine tag (6XHis) for purification and two thrombin sites to facilitate crystallization after digestion.

2.4 Target protein expression, purification and crystallization.

2.4.1 Expression, purification and crystallization of BSR4.

Large scale expression and purification of BSR4

In total, 4 L of *Hi*5 cells at 1.8 x 10^6 cells/ml were infected with the amplified BSR4 virus (1 ml of amplified virus per 1 L of culture) and incubated for 60 hrs at 27^o C with shaking at 120 rpm. After 60 hrs, the culture was harvested by centrifugation at 1,500 rpm for 15 min, and the supernatant was filtered sequentially through 5 µm, 1 µm and 0.45 µm filters to remove remaining cellular debris. To reduce volume of the supernatant, tangential flow concentration was utilized. Ni-NTA resin was then manually added to the concentrated supernatant and allowed to batch bind at 4° C for at least 1 hr. Fractions eluted from the Ni-NTA resin were analyzed by SDS-PAGE and those that contained BSR4 were pooled, concentrated using Centricon spin concentrators and injected on to an FPLC Superdex 75 Hi Load 16/60 size exclusion column in the HBS buffer (20 mM Hepes; 50 mM NaCl; pH 7.5). Fractions were analyzed by SDS-PAGE, pooled based on purity and concentrated to 10 mg/ml. Final yield of purified BSR4 from 4 L experiment was 20 mg.

Crystallization of BSR4

Crystallization trials were set with Precipitant Synergy 64, WIZARD I/II, Index and Peg Ion in 96-well plates. The final drops consisted of 1.0 µl protein at 10 mg/ml and 1.0 µl reservoir solution, and were equilibrated against 150 µl of reservoir solution and incubated at 293 K, using a sitting drop method. The crystals of BSR4 grew in 18 % polyethylene glycol (PEG) 8000, 100 mM sodium cacodylate pH 6.5, and 100 mM zinc

acetate to a maximum size of 0.5 x 0.4 x 0.2 mm after 7 days at 293K (Grujic *et al.* 2008).

2.4.2 Data collection, processing, and structure solution for BSR4.

A single BSR4 crystal was looped into cryoprotectant consisting of mother liquor supplemented with a mixture of 10 % glycerol and 10 % ethylene glycol for 10 sec and flash cooled directly in the cryostream (100 K). Diffraction data was collected to 1.9 Å as described in Table 1 (Crawford *et al.* 2009a). All refinement steps were carried out using the CCP4 suite of programs (Collaborative Computational Project 1994). Initial phases were obtained by molecular replacement (MR) with MOLREP (Vagin and Teplyakov 1997), and the individual domains of SAG1 (PDB ID: 1KZQ), pruned with CHAINSAW (Schwarzenbacher *et al.* 2004) served as the search models. The individually docked D1 and D2 domains were used as the starting point for ARP/Warp (Perrakis *et al.* 1999), which built and registered the sequence of approximately 70 % of the backbone. The remaining structure was built manually, with solvent atoms selected using COOT (Emsley and Cowtan 2004). All solvent atoms were inspected manually before deposition. The overall structure of BSR4 was refined with REFMAC (Murshudov *et al.* 1997) to an R_{cryst} of 23.8 % and an R_{free} of 26.8 %. Stereo-chemical analysis of the refined BSR4 structure was performed with PROCHECK and SFCHECK in CCP4, with the Ramachandran plot showing excellent stereochemistry with more than 92 % of the residues in the favored conformations and no residues modeled in disallowed orientations. Overall 5 % of the reflections were set aside for calculation of $R_{free.}$

Table 1. Data collection and refinement statistics for BSR4.

A. *Data collection*	
Spacegroup	$P4_12_12$
Cell dimensions	
a,b,c (Å)	92.05, 92.05, 98.31
α, β, γ (deg.)	90
Resolution (Å)	33.60 – 1.90 (1.97 – 1.90)
Measured reflections	396248
Unique reflections	33894
Average redundancy	11.69 (11.41)
Completeness (%)	100 (100)
I/▽(I)	14.5 (3.8)
R_{merge}[a](%)	0.064 (0.495)
B. *Refinement Statistics*	
Resolution range (Å)	30.87 – 1.90
R_{cryst}[b]	0.238 (0.334)
R_{free}[c]	0.288 (0.355)
No. of atoms	
Protein	2138
Solvent	279
Zn	8
B-values	
Protein (Å^2)	36.18
Solvent (Å^2)	45.80
Zn (Å^2)	45.52
r.m.s. deviation from ideality	
Bond lengths (Å)	0.018
Bond angles (deg.)	1.868

Values in parentheses are for the highest resolution shell

[a] $R_{merge}= \sum_{hkl}|I-\langle I\rangle|/\sum_{hkl}I$, I is the intensity of unique reflection *hkl*, and <I> is the average over symmetry-related observation of unique reflection *hkl*.

[b] $R_{cryst}=\sum|F_{obs}-F_{calc}/\sum F_{fobs}$, F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

[c] R_{free} is R with 5% of reflections omitted from refinement

Small angle X-ray scattering (SAXS)

Synchrotron X-ray scattering data from solutions of BSR4 were collected at the X33 beamline of the EMBL (DESY, Hamburg) (Koch and Bordas 1983) using a Pilatus 500K instrument. A 4.2 mg/ml solution of BSA was measured as a reference and for calibration procedures (Crawford *et al.* 2009a). The scattering patterns were measured with an exposure time of 2 min at 288 K at a wavelength of 1.5 Å. The sample-to-detector distance was set at 2.4 m resulting in scattering vectors ranging from 0.06 $Å^{-1}$ to 0.5 $Å^{-1}$. Three concentrations of BSR4 7.63 mg/ml, 5.61 mg/ml and 4.05 mg/ml were measured to test for consistency and eliminate concentration dependent effects. Background scattering was measured after each protein sample using the buffer solution and then subtracted from the protein scattering patterns after proper normalisation and detector response correction. The values of radii of gyration (R_g) were derived from the Guinier approximation (Guinier and Fournet 1955): $I(q) = I(0) \exp(-q^2R_g^2/3)$, where $I(q)$ is the scattered intensity and $I(0)$ is the forward scattered intensity. The radius of gyration and $I(0)$ are inferred respectively from the slope and the intercept of the linear fit of $Ln[I(q)]$ *vs* q^2 in the q-range $q.R_g<1.12$. The distance distribution function $P(r)$ was calculated on the merged curve by the Fourier inversion of the scattering intensity $I(q)$ using GNOM (Svergun 1992) and GIFT (Bergmann *et al.* 2000). The low-resolution shape of BSR4 was determined *ab initio* from the scattering curve using the program GASBOR (Svergun *et al.* 2001).

2.4.3 Expression, purification and crystallization of SRS2.

Large scale expression and purification of SRS2

A total of 3 L of *Hi*5 insect cells at 1.8 x 10^6 were infected with 1 mL of amplified SRS2 virus per 1 L culture for 60 hours at 27^o C with shaking at 120 rpm. After 60 hours, the culture was harvested as described for BSR4 in 2.4.1. Tangential flow concentration was used for both to reduce the volume and to buffer exchange the sample into Ni affinity column Buffer A (20 mM Hepes; 1 M NaCl; 30 mM imidazole; pH 8.0). The sample was re-filtered through a 0.45 µm filter prior to automatic loading onto an FPLC HisTrapFF affinity column (GE Healthcare), fractions eluted off with increasing concentrations of Buffer B (20 mM Hepes; 1 M NaCl; 500 mM imidazole; pH 8.0) and analyzed by SDS-PAGE. Those containing the protein of interest were pooled, concentrated using Centricon spin concentrators and thrombin digested to remove the polyhistidine tag. SRS2 was digested with 1 µl (1U) of thrombin for every 20 mg of purified protein in the presence of $CaCl_2$ (2.25 mM final concentration) at 4^o C, overnight.

Digested SRS2 was centrifuged at 16,000 x g prior to manual injection onto an FPLC size-exclusion Superdex 75 column (GE Healthcare) in Buffer C (20 mM Hepes; 50 mM NaCl; pH 7.5). Resulting fractions were visualized by SDS-PAGE, pooled, concentrated and buffer exchanged into Buffer D (20 mM Hepes; 10 mM NaCl; pH 8.0). To further purify SRS2, it was injected onto an ion exchange FPLC SOURCE 30 Q column (GE Healthcare) and eluted off the column using Buffer E (20 mM Hepes; 500 mM NaCl; pH 8.0). Purified SRS2 was pooled and concentrated to 36 mg/mL using Centricon spin concentrators. Final yield of purified SRS2 from 3 L experiment was 30 mg.

Crystallization of SRS2

Crystallization trials were set with Precipitant Synergy 64, WIZARD I/II and Index in 96-well plates using the sitting drop method. The final drops consisted of 1.0 µl protein at 36 mg/ml and 1.0 µl reservoir solution and were equilibrated against 150 µl of reservoir solution and incubated at 293 K. SRS2 crystallized in Index screen condition (12 % PEG 3350; 0.1 M Hepes pH 7.5; 0.005 M Cobalt chloride hexahydrate; 0.005 M Nickel chloride hexahydrate; 0.005 M Cadmium chloride dehydrate; 0.005 M Magnesium hexahydrate) at 293 K after 7 days.

2.4.4 Data collection, processing, and structure solution for SRS2.

A single SRS2 crystal was looped into cryoprotectant consisting of mother liquor supplemented with a mixture of 10 % glycerol and 10 % ethylene glycol for 10 sec and flash cooled directly in the cryostream (100 K). Diffraction data was collected to 2.1 Å as described in Table 2. All refinement steps were carried out using the CCP4 suite of programs (Collaborative Computational Project 1994). Initial phases were obtained by MR with MOLREP (Vagin and Teplyakov 1997), and the individual domains of SAG1 (PDB ID: 1KZQ), pruned with CHAINSAW (Schwarzenbacher *et al.* 2004) served as the search models. The individually docked D1 and D2 domains were used as the starting point for ARP/Warp (Perrakis *et al.* 1999). The remaining structure was built manually, with solvent atoms selected using COOT (Emsley and Cowtan 2004). All solvent atoms were inspected manually before deposition. The overall structure of SRS2 was refined as described for BSR4 in 2.4.2.

Table 2. Data collection and refinement statistics for SRS2.

A. *Data collection*	
Spacegroup	$P4_12_12$
a, b, c (Å)	83.63, 83.63, 87.79
α, β, γ (deg.)	90, 90, 90
Wavelength	1.5418
Resolution (Å)	34.14 - 2.20
Measured reflections	106845
Unique reflections	14336
Redundancy	7.45 (7.58)
Completeness (%)	99.7 (98.9)
$I/\sigma(I)$	9.1 (2.1)
R_{merge} [a]	0.106 (0.490)
B. *Refinement Statistics*	
Resolution range (Å)	22.64 - 2.20 (2.26 - 2.20)
R_{cryst} [b] / R_{free} [c]	0.278 / 0.197
No. of atoms	
Protein	1969
Solvent	232
B-values	
Protein ($Å^2$)	26.76
Solvent ($Å^2$)	33.53
r.m.s. deviation from ideality	
Bond lengths (Å)	0.022
Bond angles (deg.)	2.533

Values in parentheses are for the highest resolution shell

[a] $R_{merge}= \sum_{hkl} \sum_i |I_{hkl,i} - [I_{hkl}]| / \sum_{hkl} \sum_i I_{hkl,i}$, where $[I_{hkl}]$ is the 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

2.4.5 Expression and purification of the dimer constructs.

Large scale expression of the dimer constructs

Three litres of Hi5 insect cells were infected with 3.5 mL of amplified BSR4 dimer or SRS2 dimer virus per 1 L culture for 60 hours at 27^{o} C with shaking at 120 rpm. After 60 hrs, the culture was harvested, concentrated and purified on the FPLC HisTrapFF affinity column as described for SRS2 in 2.4.3. Fractions of SRS2 and BSR4 dimers eluted off the column were analyzed by SDS-PAGE. Those containing BSR4 dimer and SRS2 dimer were pooled, concentrated using Centricon concentrators and centrifuged at 16,000 x g, prior to manual injection onto an FPLC size-exclusion Superdex200 column (GE Healthcare) pre-equilibrated in Buffer C (20 mM Hepes; 150 mM NaCl; pH 7.5). Purified proteins were concentrated to 40 mg/ml using Centricon spin concentrators. Final yield of purified SRS2 and BSR4 dimers from 3 L was 25 mg and 20 mg respectively.

Crystallization trials with the dimer constructs

Crystallization trials were set with Precipitant Synergy 64, WIZARD I/II, Index and Peg Ion in 96-well plates using the sitting drop method. The final drops consisted of 1.0 µl protein at 40 mg/ml and 1.0 µl reservoir solution and were equilibrated against 150 µl of reservoir solution and incubated at 293 K. No signs of crystallization were observed.

Chapter 3: Exploring the SRS fold via tachyzoite and bradyzoite expressed SRSs.

3.1 Contributions to the data.

My contributions to the data presented in this chapter include: harvesting and Nickel affinity purification of BSR4; production, complete purification and crystallization of SRS2; structural analysis of SRS2, including rotation in the linker region figure; structural overlays of BSR4, SRS2 and SAG1.

3.2 Introduction.

SRS fold

Structural characterization of the prototypic member of the SRS superfamily, SAG1, revealed a novel structural organization termed the **'**the SRS fold' (He *et al.* 2002)**.** Even though it has been noted that the SRS fold is similar to Immunoglobulin (Ig) beta-sandwich (He *et al.* 2002), the major difference is in the pure anti-parallel strands of the Ig-beta sandwich vs. a mixture of parallel and anti-parallel strands of the SRS fold (Figure 9, right panel). The SRS fold appears to be unique to the apicomplexan parasites and predicted to be present only in *Neospora*, *Sarcocystis* (Jung *et al.* 2004), *Plasmodium* species (Templeton 2007). For example, recent modeling studies showed structural and sequence similarity of *P. falciparum* 230 6-cys domains to SAG1 dimer (Gerloff *et al.* 2005). Such cross-species conservation of the SRS fold indicates conserved, essential functions of these proteins (Cowman and Crabb 2006). Therefore, molecular characterization of the SRS superfamily will allow more comprehensive understanding of the pathogenesis of not only toxoplasmosis, but also other infectious diseases, such as malaria, caused by *Plasmodium*, and human sarcosporidiosis, caused by *Sarcocystis*.

The crystal structure of SAG1 revealed that the N-terminal (D1) domain of ~ 130 residues and the C-terminal (D2) domain of ~ 120 residues share a high degree of structural similarity (2.7 Å r.m.s.d.) despite their low overall sequence homology (He *et al.* 2002). Both domains were found to adopt the four-on-three organization of beta-sheets, pinned together by three disulfide bonds, illustrated in Figure 9 (right panel). The number and positions of disulfide bonds in the SRS fold were bioinformatically predicted to be conserved within SAG1 and SAG2 subfamilies (Jung *et al.* 2004) consistent with a conserved structural core. However, outside of the SRS beta-sandwich core, insertions and deletions of various sizes indicate that SRSs may display structurally divergent interstrand regions. The most extended insertions are observed in the bradyzoite-expressed SRSs, such as BSR4 (Figure 3). On the other hand, in the SRSs expressed on the surface of tachyzoites, such as SAG1 and SRS2, these insertions are absent. Based on these observations, the core SRS fold is hypothetically conserved among members of the SRS superfamily expressed on the same developmental stages of *T. gondii*, while the SRSs expressed on the different stages, such as bradyzoite-expressed BSR4 and tachyzoite-expressed SRS2, are most likely to display structurally divergent loops.

SAG1 dimer

The crystal structure of SAG1 also established that the SAG1 monomers form a parallel homodimer (Figure 9), in which C-termini of each monomer are attached to the cell surface (He *et al.* 2002). In this structure, the N-terminal domains form a dimer interface with a topologically defined contiguous basic groove decorated with the surface exposed loops. A divergent loop structure may endow the SRS superfamily with a structural plasticity that enables *T. gondii* to attach to such a wide variety of host cells.

Identification and characterizing of these structural features are, therefore, crucial to understanding *T. gondii* virulence.

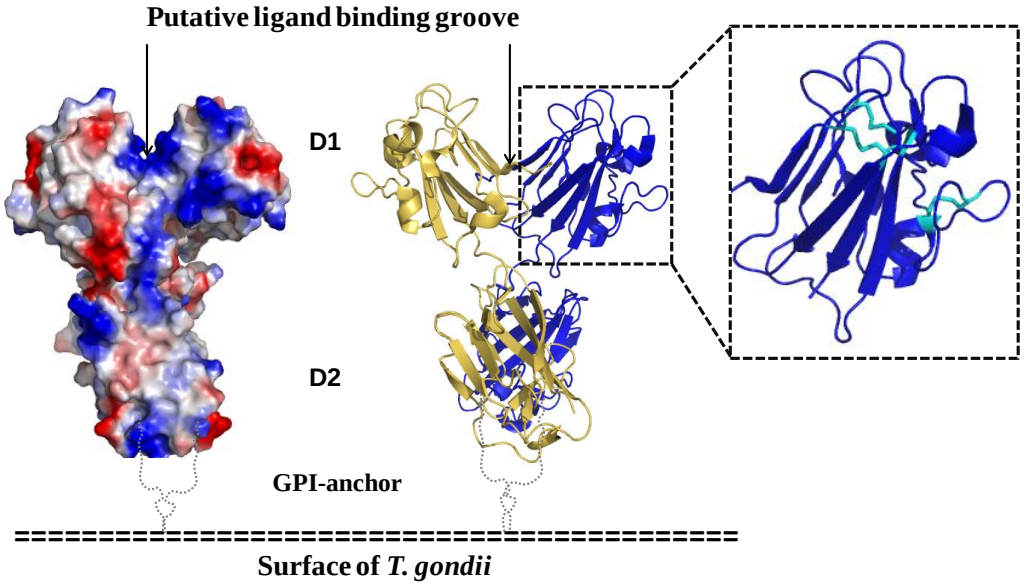

Figure 9. Structure of SAG1.
Surface charge distribution of SAG1 dimer (PDB ID: 1KZQ) with putative ligand binding groove indicated with an arrow (left panel). Ribbon view of the SAG1 dimer (middle panel). One monomer is yellow and the other is blue. The C-terminus was extended with a dotted line to show the location of the GPI-linkage with the membrane of *T. gondii*. Each monomer includes tandem SAG domains characterized by the SRS fold, a close-up view of which is illustrated in the right panel with disulfide bonds highlighted in cyan (He *et al.* 2002).

Although the structure of SAG1 is highly informative, structural characterization of novel SRSs will provide a more complete description of the SRS fold. For example, a high resolution structure of tachyzoite-specific SRS2 and its comparison to SAG1 may reveal the degree of conservation of the SRS fold among SRSs expressed during tachyzoite

stage of *T. gondii*. The structural characterization of bradyzoite-expressed BSR4 will allow identification of the unique features of the bradyzoite stage of the parasite.

3.3 Biochemical characterization of BSR4 and SRS2.

Size exclusion chromatography of BSR4 and SRS2 showed the majority of protein eluted in the fraction corresponding to the 43 KDa, standard III (Figure 10 A and 11 A). This result was unexpected as the predicted molecular weight of BSR4 is 34 KDa and SRS2 is 28 KDa. Such results can be due to heavy glycosylation of both molecules. Resulting fractions were visualized by SDS-PAGE, which revealed SRS2 has a contaminating band at 20 KDa (Figure 11 A). Ion exchange chromatography was used to purify SRS2 to homogeneity (GE Healthcare) (Figure 11 B).

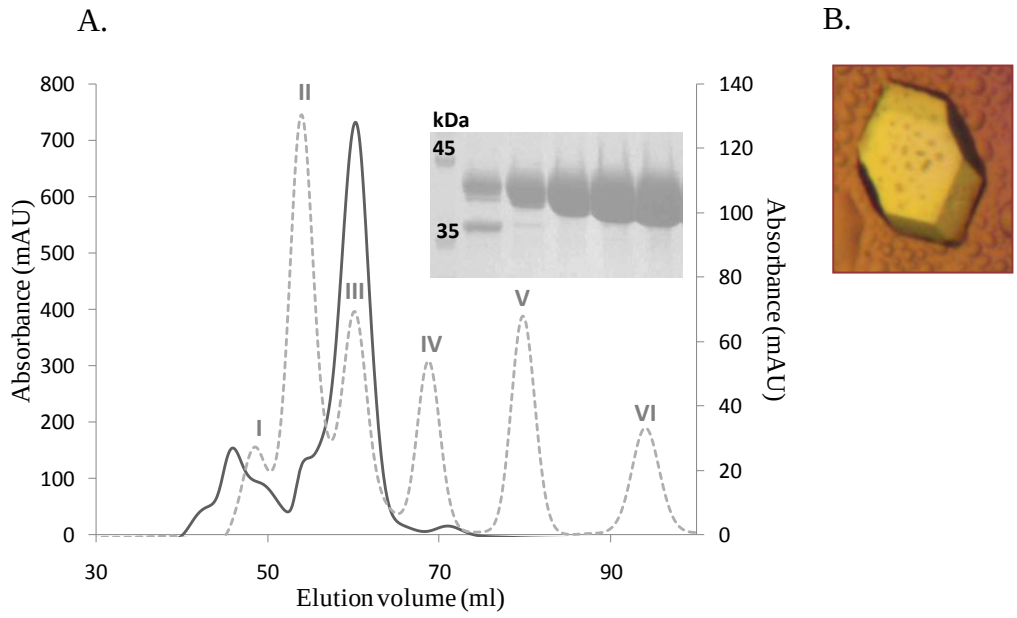

Figure 10. Purification and crystallization of BSR4.
A. Size exclusion chromatography of BSR4. Sx75 column standards are shown as grey dotted line peaks: II (53 ml) 75000 Da; III (59 ml) 43000 Da; IV (68 ml) 29000 Da; V (79 ml) 13700 Da; VI (93 ml) 6500 Da; I (under 50 ml) insoluble conglomerate. BSR4 trace (black solid line) shows that the majority of BSR4 is in the fraction corresponding to standard III, 43000 Da. **B.** Largest BSR4 crystal with dimensions 0.5 x 0.4 x 0.2 mm. Figure adapted with permission from (Grujic *et al.* 2008).

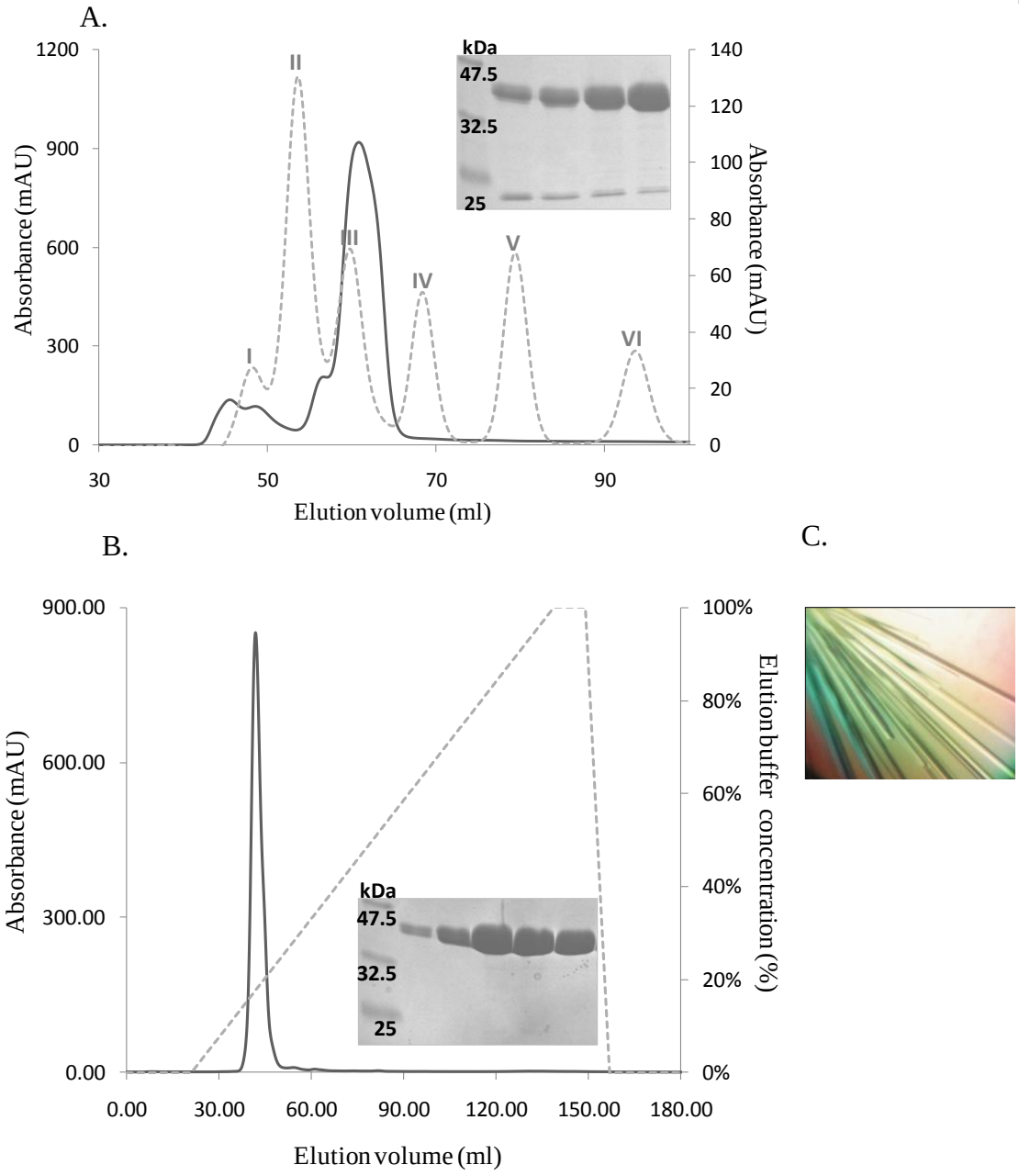

Figure 11. Purification and crystallization of SRS2.
A. Size exclusion chromatography of SRS2. Sx75 column standards are shown as grey dotted line peaks: II (53 ml) 75000 Da; III (59 ml) 43000 Da; IV (68 ml) 29000 Da; V (79 ml) 13700 Da; VI (93 ml) 6500 Da; I (under 50 ml) insoluble conglomerate. SRS2 trace (black solid line) shows that the majority of SRS2 is in the fraction corresponding to standard III, 43000 Da. **B.** Ion exchange purification of SRS2. Source30Q elution gradient (grey dotted line) shows that SRS2 eluted at 20 % concentration of elution buffer E (20 mM Hepes; 500 mM NaCl; pH 8.0). **C.** SRS2 crystals with a maximum size of 0.9 x 0.1 x 0.1 mm.

As can be seen from Figures 10 and 11, both BSR4and SRS2 purified as monomers in contrast to SAG1, which was shown to be a dimer when purified from the natural source (surface of tachyzoites) and a mixture of monomeric and dimeric forms when produced recombinantly (He *et al.* 2002). The reason for this difference is unknown. However, it is possible that these proteins may exist in the dimer form on the surface of *T. gondii* when they are anchored to the membrane via GPI-linker. Recent studies have shown that GPI-anchored proteins or transmembrane proteins with single pass transmembrane helices exist as biological multimers on the cell surface, but show no appreciable multimerization in solution (Caiolfa 2007; Sharma *et al.* 2004). The shift in equilibrium towards the monomeric form is predicted to result from the reduced avidity enhancing effects of the cell membrane.

3.4 Overall structures of BSR4 and SRS2.

SRS2 and BSR4 crystallized as monomers with one molecule in the asymmetric unit of the $P4_12_12$ unit cell. Both structures were solved using molecular replacement based on the SAG1 structure (PDB ID: 1KZQ). The SRS2 structure was refined to a resolution of 2.1 Å and revealed a dumbbell-shaped molecule consisting of two domains oriented in a head-to-tail fashion (Figure 12). Similar to SAG1, D1 and D2 domains of SRS2 are of the similar size with the D1consisting of ~130 amino acids and the D2 of ~120 amino acids. Both domains are also composed of beta-sheet sandwiches with a four-on-three sandwich core and crosslinked by three conserved disulfide bonds.

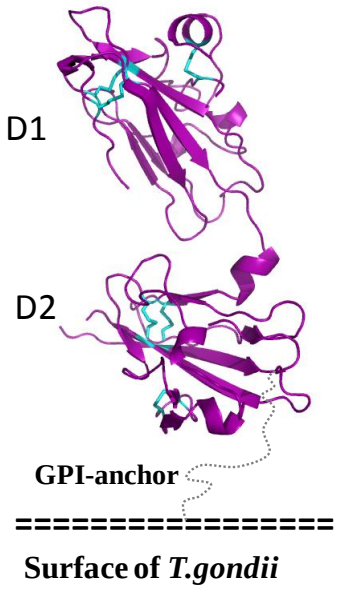

Figure 12. Overall structure of SRS2.
Ribbon view of the SRS2 monomer. Three pairs of disulfide bonds are highlighted in cyan. The C-terminus is extended in a dotted line to indicate the location of the GPI-anchor with the parasite's membrane, and orientation of D1 and D2 domains.

The BSR4 structure was refined to a resolution of 1.90 Å. Both of BSR4's domains adopt a four-on-three beta-sandwich fold composed of antiparallel and parallel strands stabilized by conserved disulfide bonds (Figure 13 A). However, the D1 domain of BSR4 consists of 157 residues and adopts a flat, extended structure (Figure 13 B, left panel), unlike the smaller (~130 residues), globular D1 of SRS2 and SAG1. The upper face of BSR4 D1 beta-sandwich is formed by two central parallel strands (b and d) organized in an upwards fashion that are bookended by two downward facing strands (a and e) to give an overall down-up-up-down topology. The lower face of the beta-sandwich is also comprised of a mixture of parallel and antiparallel strands in which strands "c" and "g" directed down and strand "f" directed up. Three disulfide bonds add significant stability to the core structure. The D2 domain of BSR4 is smaller than the D1 domain and adopts a more compact, globular structure (Figure 13 B, right panel). Unlike the D1 domain, the

organization of the strands in the D2 domain of BSR4, SRS2 and SAG1 is that of a five-on-three beta-sandwich (Crawford *et al.* 2009a).

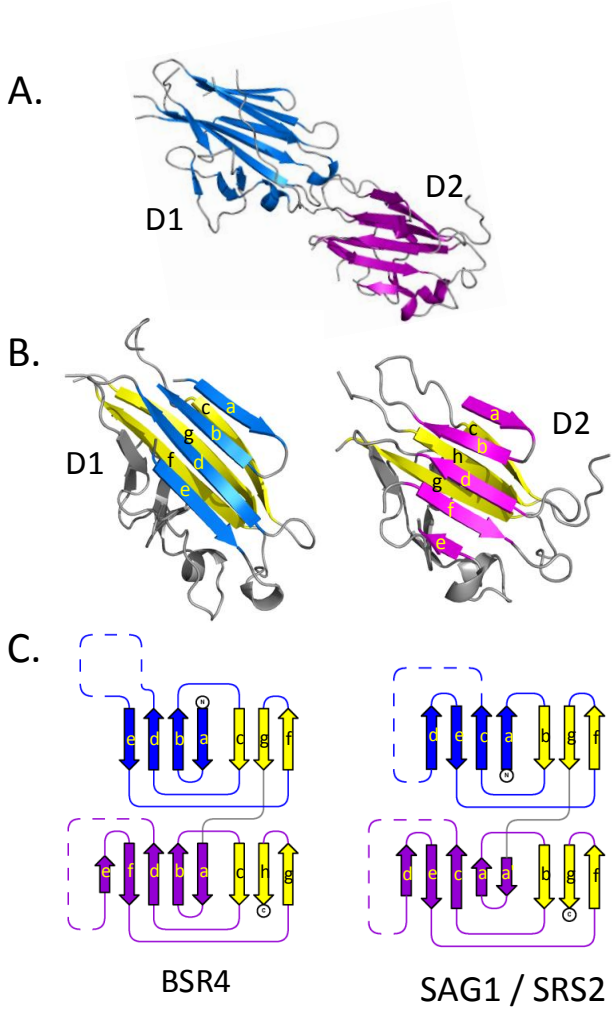

Figure 13. Secondary structure depiction of BSR4 and topology diagrams of SRS2 and SAG1.
A.Ribbon diagram showing the head to tail structure of BSR4 with the N-terminal (D1) domain in blue and the C-terminal (D2) domain in magenta. **B.** Detailed view of the beta-sandwich folds adopted by the D1 and D2 domains. The beta-strands form the upper leaf of the beta-sandwich. The yellow strands depict the lower part of the beta-sandwiches. Structure figures were generated with PYMOL (Delano 2002). **C.** The lower left and right hand panels represent the topology diagrams generated with TopDraw (Bond 2003) for BSR4 and SAG1/SRS2, respectively. Figure adapted with permission from (Crawford *et al.* 2009a).

3.4.1 Analysis of the SRS fold.

Although SRS2 and BSR4 are members of the SAG1 family and share about 30 % sequence identity with SAG1 (30 % and 26.5 % respectively), a detailed comparison of their structures revealed a divergence in the topology of BSR4 D1 domain. In BSR4, strands “a” and “b” are organized in an antiparallel fashion followed by strand “c” that forms part of the lower face of the beta-sandwich (Figure 13 C, left panel). In SAG1 and SRS2, however, strands “a” and “b” are positioned on opposite faces of the beta-sandwich while strand “c” is parallel to strand “a” (Figure 13 C, right panel) (Crawford *et al.* 2009a). Reorganization of the connectivity of strands “a”, “b” and “c” that form the dimer interface affects the surface structure of the D1 domain, which was originally described as a putative ligand binding groove of SAG1. Although the overall topology of the D2 domain is conserved between BSR4 and SAG1/SRS2, the strand designations in BSR4 have been modified to reflect that the first two strands are antiparallel and therefore denoted as “a” and “b” rather than “a” and “a'” as defined in SAG1 (Barragan and Sibley 2002).

In addition to topology analysis, the structures of BSR4 and SRS2 were analyzed using structural overlays of individual SRS domains and compared to SAG1, the prototypic member of the SRS superfamily (Figure 14). All three proteins were found to have conserved disulfide bonds position and number. Overall structures of the D2 domains were also found to be well-conserved. However, the D1 domain of BSR4 displayed longer beta-strands and more extensive surface loop structure (Figure 14, left panel) comparing to the D1 of SRS2 and SAG1(Figure 14, right panel).

Figure 14. Structural overlays of N-terminal domains of SAG1, SRS2 and BSR4.
N-terminal (D1) domains overlay BSR4 (green) and SRS2 (purple) over SAG1 (yellow) using PDB viewer (Guex and Peitsch 1997). BSR4 over SAG1 has r.m.s.d. of 1.14 Å over just 85 residues, while SRS2 over SAG1 has r.m.s.d. of 1.37 Å over 100 residues.

Although both SRS2 and BSR4 share approximately the same sequence identity with SAG1, there are substantial differences in the D1 domain between bradyzoite-expressed, BSR4 and tachyzoite-expressed SRS2 and SAG1. The implications of these findings can be seen considering the evidence of functional importance of D1 domain in SRSs. First, in the original SAG1 study, a putative ligand binding groove was identified at the dimeric interface of D1 domains (Figure 9), suggesting ligand binding role of this domain. Secondly, when SAG1 was co-crystallized with a monoclonal antibody that mimics human immune response to *T. gondii* infection, the immunodominant epitope was also located at the surface loops of D1 domain (Graille *et al.* 2005). Therefore, our findings of the structurally divergent D1 in BSR4 suggests that BSR4 evolved for binding a different ligand than SAG1 and SRS2 and/or plays different biological role than SAG1.

3.4.2 Polymorphic nature and flexibility of SRS linkers.

In BSR4 and SRS2, D1 and D2 domains are oriented in a head-to-tail fashion connected by an interdomain region, identified in this work as a linker. The crystal structures of BSR4 and SRS2 clearly show the linker region to be extended such that the D1 and D2 domains do not interact. A ten-residue linker of SRS2 (Figure 12) positions the D1 and D2 domains further apart and makes the interdomain region of the molecule more solvent accessible, than in BSR4 (Figure 13 A), with a five-residue linker (Figure 15 A). This suggests that the linker of SRS2 may serve as an additional epitope for ligand binding. It may also serve as a cleavage site for a proteolytic enzyme to remove the N-terminal domain in a similar fashion as the major *Plasmodium* surface antigen and adhesin, PfMSP-1 which is cleaved during *Plasmodium* invasion (Bentley 2006). In addition to the unusual length, a helical structure of the SRS2 linker was revealed (Figure 15 B). Presence of the secondary structure in the linker region suggests limited flexibility of this region, which in turn may facilitate ligand binding.

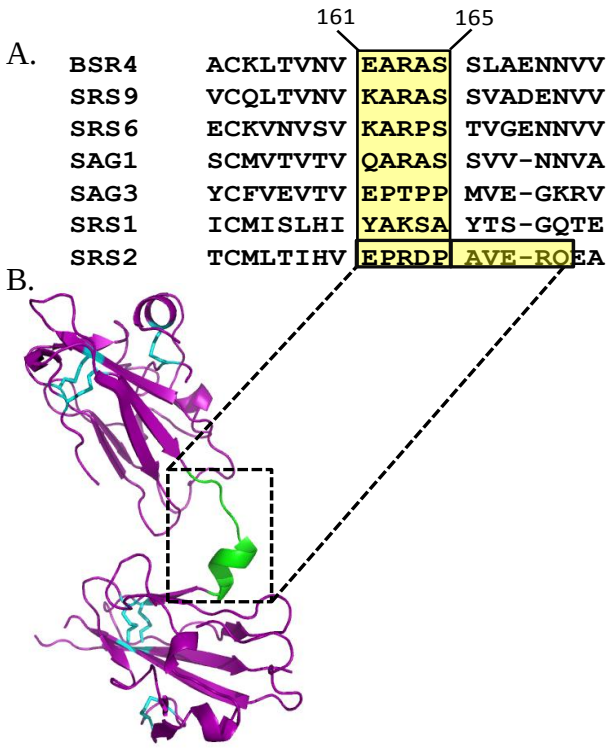

Figure 15. Polymorphic linkers of SRSs.
A. Linker regions of representative SRSs are aligned with CLUSTALW and highlighted in yellow. Figure adapted with permission from (Crawford *et al.* 2009a). **B.** Ribbon view of the SRS2 monomer. The linker region is highlighted in green and boxed.

Rotation in SRS2 linker

SRS2 structure analysis revealed the possibility of 90° rotation in the linker region along the vertical axis of the molecule (Figure 16). Mapping of the SAG1 dimer interface (Figure 16, right panel) onto the surface of the SRS2 molecule, based on their sequence alignment, allowed visualization of the putative dimer interface in SRS2, which in turn revealed that the SRS2 domains would have to rotate about 90° along the vertical axis in order to form a dimer in the same fashion as SAG1.

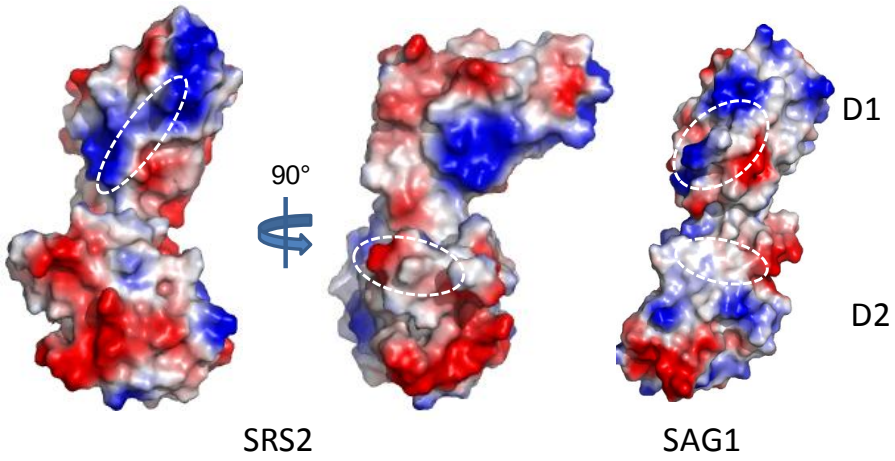

Figure 16. Rotation in the linker region of SRS2.
Electrostatic surface representation of the SRS2 monomer (left panel) and SAG1 monomer (right panel). SAG1 monomer created in PYMOL using SAG1 dimer molecule (PDB ID: 1KZQ) where dimer interface is circled with a white dotted line. SAG1 dimer interface was then projected into the SRS2 monomer to illustrate 90° rotation.

Unique features of SRS2 linker brought our attention to the interdomain region of other SRSs. Comparison of the predicted linker sequences of seven different SRS proteins from bradyzoite and tachyzoite cell stages of *T. gondii* revealed two general subfamilies: alanine rich linkers of BSR4 (EARAS), SAG1 (QARAS), SRS1 (YAKSA), SRS9 (KARAS) and SRS6 (KARPS) form one group while the proline rich linkers of SAG3 (EPTPP) and SRS2 (EPRDPAVERQ) form the second group (Figure 15 A). In general, the linker regions of SRSs appear to provide a structural element, which imparts flexibility to selectively mediate dimerization (Crawford *et al.* 2009a). Interestingly, the structure of the linker, and therefore relative disposition of the D1 and D2 domains, does not show stage specificity. One possible explanation is that the linker provides a structural filter to both facilitate heterodimerization between select SRS antigens while restricting others, creating new ligand binding sites. This suggests that the structure of the linker, and, therefore, relative disposition of the D1 and D2 domains, is not stage-

specific, since SRS2 and SAG1 are both tachyzoite-expressed but divergent in this region. Another possible explanation for such non-stage specific division of SRSs linker regions is that, perhaps, these regions serve as an enzyme binding site. The reason for this is unclear and needs further investigation.

Flexibility of BSR4 linker

The structure of BSR4 in solution was also determined using small angle X-ray scattering (SAXS), which allowed characterization of the structural relationship between the D1 and D2 domains (Crawford *et al.* 2009a). The radii of gyration and D_{max} values were calculated for the three concentrations of BSR4 (Table 3). Normalized spatial discrepancy (NSD) value for each shape was computed by several *ab initio* GASBOR (Svergun *et al.* 2001) calculations, and subsequent comparison with the program DAMAVER (Konarev *et al.* 2003) for each concentration. The NSD values of 0.92-0.05 in all experiments indicated limited conformational variability (Putnam *et al.* 2007).

The overall *ab initio* shape determined by SAXS was utilized for docking of the individual domains of BSR4 (Figure 17) together with the surface calculated based on the crystal structure. It was then determined that while all shapes produced with GASBOR (Svergun *et al.* 2001) indicate a kink between the domains with an interdomain angle of approximately 45°, the χ values for the crystal structures calculated with CRYSOL (Petoukhov *et al.* 2002) (Table 3) suggest that the disposition of the D1 and D2 domains in the crystal structure is lenear. Therefore, the structural relationship between the D1 and D2 domains of BSR4 in crystal structure is different than in solution.

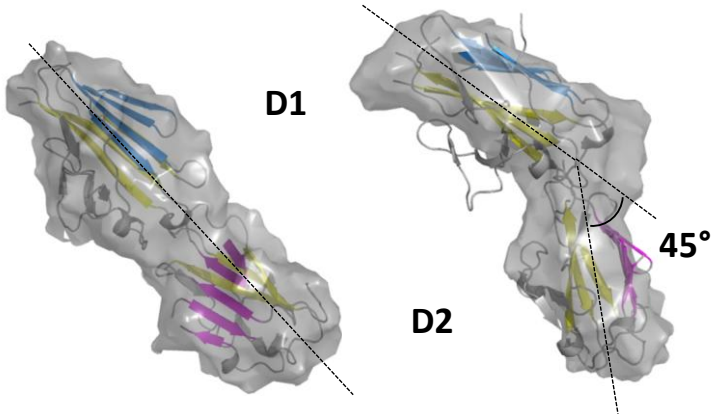

Figure 17. Interdomain flexibility of the BSR4 monomer.
Surface representation of the BSR4 monomer calculated based for the crystal structure and the solution SAXS structure. In the crystal structure, the D1 and D2 domains are organized in a linear head to tail fashion while in the SAXS structure the D2 domain is reoriented such that is it deviates approximately 45° from linearity. Figure adapted with permission from (Crawford *et al.* 2009a).

Table 3. Structural parameters of BSR4 obtained by SAXS.

Sample concentration	**Mw (kDa)**	**Number of a. a.**	**R_G (Å)**	**D_{max} (Å)**	***ab initio* modelling**	
					dχGASB)	**χ (Model)**
BSR4 (7.63 mg/ml)	34	315	26.1⬅0.5	85⬅5	1.79	2.81
BSR4 (5.61 mg/ml)	34	315	26.5⬅0.7	85⬅3	1.28	1.97
BSR4 (4.05 mg/ml)	34	315	26.2⬅0.7	85⬅3	1.25	2.24
BSA	66.3	582	30.7⬅1.0	90⬅2	nd	nd

Table adapted with permission from (Crawford *et al.* 2009a).

3.5 Conclusions.

Prior to my thesis work, the only structure of an SRS protein from *T. gondii* was that of the tachyzoite-expressed SAG1. The high resolution structure of bradyzoite-expressed BSR4 represents the first structural description of the SRS expressed on the highly infectious, cyst-forming bradyzoite stage of *T. gondii*, while tachyzoite-expressed SRS2 structure in combination with BSR4 offers a unique opportunity to analyze the degree of conservation of the SRS fold among SRSs expressed at the different developmental stages of the parasite.

In depth structural analysis of the SRS2 and BSR4 structures revealed that both proteins retained global features of SAG1 adopting a similar dumbbell structure in which N-terminal and C-terminal domains display a beta-sandwich fold with antiparallel and parallel strands stabilized by conserved disulfide bonds. This indicates conservation of the core SRS fold established in SAG1 study and characterized by a mixture of antiparallel and parallel beta-sheets in the four-on-three organization pinned together by three conserved disulfide bonds. However, a detailed comparison of novel structures and SAG1 revealed an unexpected degree of divergence between bradyzoite-expressed BSR4 and tachyzoite-expressed SRS2 and SAG1. In BSR4 structure the connectivity of the beta-strands in forming the overall beta-sandwich of the D1 domain differs from SAG1 indicating that the SRS superfamily may be more structurally diverse than previously thought. Moreover, the beta-strands and the loops of BSR4 D1 domain are far more extensive than that of SRS2 and SAG1. This may result in the formation of the structurally divergent ligand binding groove found in the SAG1 structure.

In addition to expanding existing knowledge of SRS fold, this study also revealed polymorphic nature and flexibility of the interdomain region of SRSs. Unusually extended, helical linker of SRS2 may facilitate ligand binding. Solution X-ray scattering experiments with BSR4 revealed a potential flexibility of its linker region, which may play a role in multimerization of this molecule on the surface of the parasite.

Structural divergence found in bradyzoite-expressed BSR4 and tachyzoite-expressed SAG1 and SRS2 suggests that the stage-specific structural features may have implications in ligand binding. To identify possible ligands for SRS2 and BSR4 and to probe the effect of multimerization on ligand binding, dimer constructs of these proteins and SAG1 were engineered. The next chapter addresses these dimer constructs and the ligand binding assays that were designed and conducted.

Chapter 4: Ligand identification for *Toxoplasma gondii* SRSs.

4.1 Contributions to the data.

My contributions to the data presented in Chapter 4 include: SRS2 dimer protein production and purification; establishment of the carbohydrate screening macro-array, ligand identification and validation.

4.2 Ligand identification: background and current theories.

Since *T. gondii* is an obligate intracellular parasite, it must invade the host cell in order to replicate and establish infection. Host cell invasion by *T. gondii* depends on the initial attachment of the parasite to the host cell and recognition of the desirable cell for successful invasion and replication. Initial attachment and identification of the "right" host cell is followed by apical attachment, active penetration through the plasma membrane, closure and separation. The whole process only takes about 2 minutes, with closure and separation, being rate limiting and requiring 1–2 min to complete, and attachment and invasion occurring in just 15-20 seconds (Carruthers and Boothroyd 2007).

Previous studies demonstrated the ability of *T. gondii* to quickly infect any nucleated cell *in vitro*, as well as to infect any vertebrate animal (Barragan and Sibley 2002; Dubey 1998b; Sibley 1995). It has been suggested that such ability of *T. gondii* to attach to and invade a broad range of host cells can be attributed to the recognition of either different host cell receptors by various parasite attachment molecules or ubiquitously expressed surface exposed molecules on the host cell (Ortega-Barria and Boothroyd 1999a). Both of these theories of initial receptor recognition are not mutually exclusive since the surface

of *T. gondii* is covered with the superfamily of more than 160 surface proteins, and host cells often express surface exposed ligands which are conserved across species.

Toxoplasma gondii surface attachment molecules

Several studies have demonstrated that *T. gondii* surface coat proteins are possible mediators of the initial parasite attachment to the host cell. Studies of the most abundant tachyzoite antigen SAG1 have shown that SAG1 mAbs block tachyzoite invasion indicating that SAG1 might function as a specific attachment ligand (Grimwood and Smith 1992; Grimwood and Smith 1996; Mineo *et al.* 1993; Velge-Roussel *et al.* 2001). On the other hand, the fact that mutants lacking SAG1 are able to attach and invade faster than the wild type *in vitro* (Carruthers and Boothroyd 2007), and SAG3-deficient parasites exhibit only 50 % of the attachment capacity (Dzierszinski *et al.* 2000), suggest that other SRSs are also involved in the attachment of *T. gondii* to host cells. However, there was no evidence of direct interactions between SRSs and host cell ligands to this date.

Host cell surface receptors

The glycosylated nature of the host cell receptors for *T. gondii* was originally inferred from BSA-glucosamide competitively blocking *in vitro* infection by tachyzoites with SAG1 more effectively than those without SAG1 (Mineo *et al.* 1993). Furthermore, Boothroyd and co-workers demonstrated the role of a carbohydrate binding activity of *T. gondii* tachyzoites in erythrocyte agglutination and infection of human fibroblasts and epithelial cells (Ortega-Barria and Boothroyd 1999b). In this study, agglutination of the erythrocytes appeared to be reversed upon the addition of soluble glycoconjugates, such as heparin, fucoidan, and dextran sulfate (Ortega-Barria and Boothroyd 1999b).

Furthermore, epithelial cells deficient in the biosynthesis of surface proteoglycans were found to be more resistant to *T. gondii* infection (Ortega-Barria and Boothroyd 1999b).

In subsequent studies, a variety of host cell heparan sulphate proteoglycans (HSPGs) were found to be involved in parasite attachment. For example, HSPG-deficient mutant host cells and wild-type cells enzymatically treated to remove HSPGs were partially resistant to parasite invasion, demonstrating the importance of sulphated proteoglycan recognition. Furthermore, excess of soluble heparin or chondroitin sulphate in solution with *T. gondii* inhibited parasite attachment to serum-coated glass (Carruthers *et al.* 2000). Also, Jacquet *et al.* found that only highly sulphated cell surface proteoglycans interact with *T. gondii* surface adhesin SAG3 *in vitro* (Jacquet *et al.* 2001).

Role of HSPGs in Toxoplasma gondii infection

Structural characterization of SAG1 revealed possible molecular mechanism of HSPG binding (He *et al.* 2002). According to the authors, a positively charged groove at the interface of membrane distal domains of the SAG1 dimer is a putative HSPG-binding site. The basic nature of the groove should attract negatively charged HSPGs, and the distal location, should perfectly position it for interaction with a host cell surface ligand. Most importantly, modeling of SAG3 onto the SAG1 dimer structure revealed that this basic groove is conserved on SAG3 and appears to accommodate heparin (He *et al.* 2002), which was found to bind SAG3 *in vitro* (Jacquet *et al.* 2001). On the other hand, Bishop *et al.* argued against the role of HSPGs in *T. gondii* attachment and demonstrated that the parasite uses HSPGs as a modulator of replication or a post-invasion energy source rather then the attachment factor (Bishop *et al.* 2005).

While the role of HSPGs in *T. gondii* infection is much debated, the importance of the interaction between *T. gondii* surface proteins and HSPGs in the parasite's life cycle is generally accepted. Structural characterization of the representative SRSs, complemented with functional binding studies of these proteins, will shed light on the molecular basis for SRS adhesin-mediated *T. gondii* attachment to host cell ligands. SRS2 and BSR4 were considered as objects of our ligand identification studies due to the fact that detailed structural information of these molecules (Chapter 3) may be utilized to rationalize mechanism of the ligand binding.

4.3 Strategies for ligand identification.

4.3.1 Screening a library of natural and synthetic mammalian glycans.

Based on the knowledge of the sugar binding capability of some SRSs, multiple strategies of ligand identification have been developed in this project. A library of natural and synthetic mammalian glycans with amino linkers, printed onto N-hydroxysuccinimide (NHS)-activated glass microscope slides (SCHOTT Nexterion), was probed at the Consortium for Functional Glycomics (see appendix for the full list of the Core H, v4 glycans). However, the screening of this Core H printed array did not reveal carbohydrate ligands for monomeric forms of BSR4 or SRS2. This could be due to the limited scope of the array (e.g. Heparin, IdoA (2S)-GlcNS (6S) is not included in the array). In addition, only monomeric forms of the proteins were tested, but SAG1 studies have indicated the putative ligand binding site is present on the dimeric form of the SRS adhesins (He *et al.* 2002).

To overcome limitations of this initial approach, a carbohydrate macro-array was developed and both monomeric and engineered dimeric forms of SRS2 and BSR4 were probed.

4.3.2 Carbohydrate macro-array screening.

A carbohydrate macro-array enables screening of the precise sugars to which the parasite's proteins binds. To identify potential candidates of host cell ligands that might interact with the surface adhesins from *T. gondii,* an extensive literature review was conducted. The two main criteria for inclusion were the ability of a carbohydrate to bind nitrocellulose membrane and its physiological relevance. Based on these criteria, the following blot candidates were initially chosen: laminin, heparin, heparan sulfate, mucin IS, mucin II, thyroglobulin, chondroitin sulphate, hyaluronic acid, collagen II, collagen IV, glycogen and amylopectin as a negative control. In the course of experiments, laminin, both collagens and mucin IS were excluded due to "high background" illumination. Finalized carbohydrate macro-array was probed with SAG1, SRS2 and BSR4 monomers and dimers. Once a potential ligand was identified, binding was supported by pull-down/SDS-PAGE and native gel analysis.

4.3.3 Dimerization of SRS adhesins.

There is an increasing body of evidence that GPI-anchored surface proteins exist as biological multimers on the cell surface, but not in solution (Caiolfa 2007; Sharma *et al.* 2004). *In vitro* studies of the recombinantly expressed proteins that are predicted to exist as GPI-anchored proteins on the cell membrane *in vivo* have shown the shift in equilibrium towards the monomeric form due to the lack of avidity enhancing effects of the cell membrane. This could be the case for *T. gondii* GPI-anchored surface proteins

such as SAG1, which have been shown to exist as the dimers when purified from the parasite surface, while as monomers or mixture of monomers and dimers when recombinantly expressed (He *et al.* 2002). Most importantly, structural characterization of SAG1 revealed that dimerization of SRSs molecules may be required for ligand binding (He *et al.* 2002), and thus, dimerization of SRS molecules in solution may be beneficial in ligand identification binding assays.

To facilitate dimerization of the proteins of interest in solution SRS2 and BSR4 were co-expressed with the dimerization coils as described in 2.3.3 of this work. The dimerization coils were engineered to include a leucine zipper (Blondel and Bedouelle 1991) and a glycine-rich linker connected to the C-terminal amino acid of the SRS protein as shown in Figure 18. Leucine zippers are part of a naturally occurring helical domain (Gcn4) from *Saccharomyces cerevisiae* which comprises 33 residues to form a parallel coiled-coil (helical) dimer (O'Shea *et al.* 1991). Structural characterization of Gcn4 leucine zipper by O'Shea *et al.* revealed that "the contacts between the helices include ion pairs and an extensive hydrophobic interface that contains a distinctive hydrogen bond. The conserved leucines make side-to-side interactions (as in a handshake) in every other layer of the dimer interface" (O'Shea *et al.* 1991). The glycine-rich linker, also included in the dimerization construct for this study, provides flexibility to allow monomers movement in order to achieve the best position for dimerization.

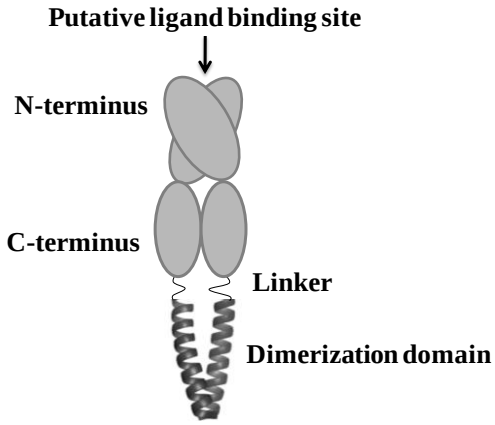

Figure 18. Schematic representation of SRS dimerization *in vitro*.
Dimerization *in vitro* may be achieved by co-expressing the recombinant protein of interest with the dimerization domain (leucine-zipper) connected to the C-terminus of the protein via a flexible linker. Upon dimerization of the protein, the putative ligand binding site is expected to form.

4.4 Biochemical characterization of the dimer constructs.

Engineering BSR4 and SRS2 dimer constructs to contain a dimerization domain (leucine zipper) allowed expression and purification of these molecules as homodimeric species, which is illustrated in Figures 19 and 20. The FPLC Sx200 trace in Figures 19 and 20 shows that both BSR4 and SRS2 dimers eluted off the column in the fraction corresponding to the standard I with MW is 75 KDa. This size of the product is as expected since the total predicted MW for BSR4 dimer in solution is ~ 86 KDa and for SRS2 dimer is ~ 72 KDa. The observed variations in the sizes of products can be attributed to glycosylation of the molecules.

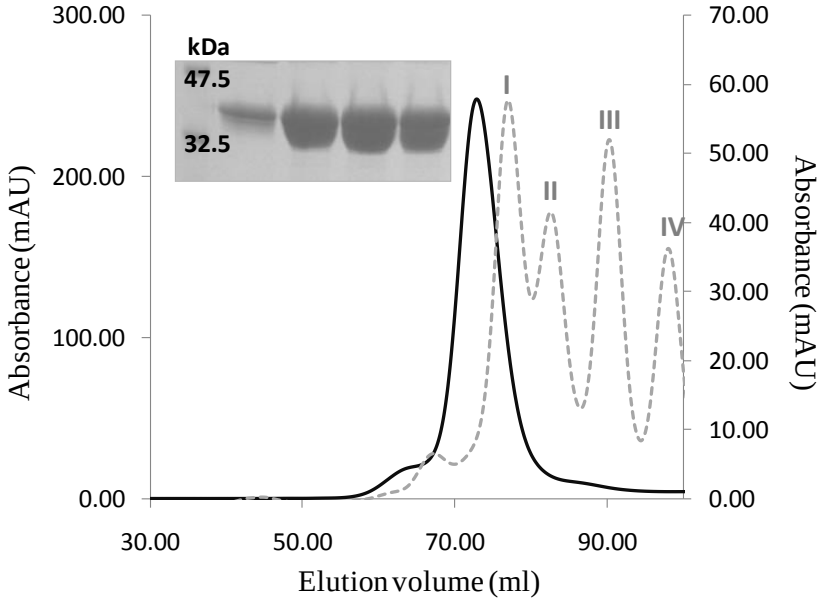

Figure 19. Size exclusion chromatography with BSR4 dimer.
Sx200 16/60 gel filtration analysis showing BSR4 trace (black line) and column standards (grey dotted line): I (77 ml) 75000 Da; II (82.5 ml) 43000 Da; III (90 ml) 29000 Da; IV (98 ml) 13700 Da. SDS PAGE analysis of the purified BSR4 dimer visualizes the band of approximate size 45 KDa, which corresponds to the size of the BSR4 monomer with the dimerization domain, the linker and the His-tag (exact MW of 43.7 KDa).

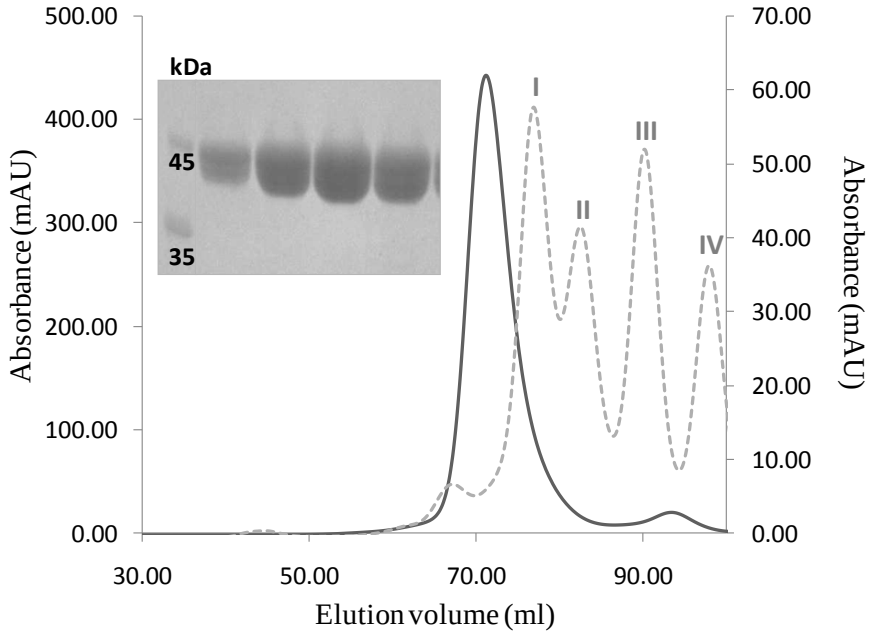

Figure 20. Size exclusion chromatography with SRS2 dimer.
Sx200 16/60 gel filtration analysis showing SRS2 trace (black line) and column standards (grey dotted line): I (77 ml) 75000 Da; II (82.5 ml) 43000 Da; III (90 ml) 29000 Da; IV (98 ml) 13700 Da. SDS-PAGE analysis of the purified SRS2 dimer visualizes a monomeric form of the protein as the band of approximate size 43 KDa, which is not as predicted size of the SRS2 monomer with the dimerization domain, the linker and the His-tag (exact MW of 36 KDa), but can be attributed to glycosylation of SRS2.

4.5 Ligand identification.

Model of ligand screening macro-array

A hypothetical model of binding event between a carbohydrate blotted onto the membrane and a dimeric form of proteins of interest labelled with the biotin-streptavidin-AlexaFluor 680 complex via His-tag-Ni-NTA is illustrated in the Figure 20. That is, the protein of interest is depicted according to the SAG1 dimer structure (He *et al.* 2002), based on the hypothesis that the putative ligand binding groove, formed by D1 (N-terminal domains) may be conserved among SRSs and that dimerization is required for the ligand binding. According to this model, a fluorescently labelled protein is detected when bound to carbohydrate, which is attached to the nitrocellulose membrane, while all unbound protein is washed off the blot, as described in the experimental procedure in section 2.2.3.

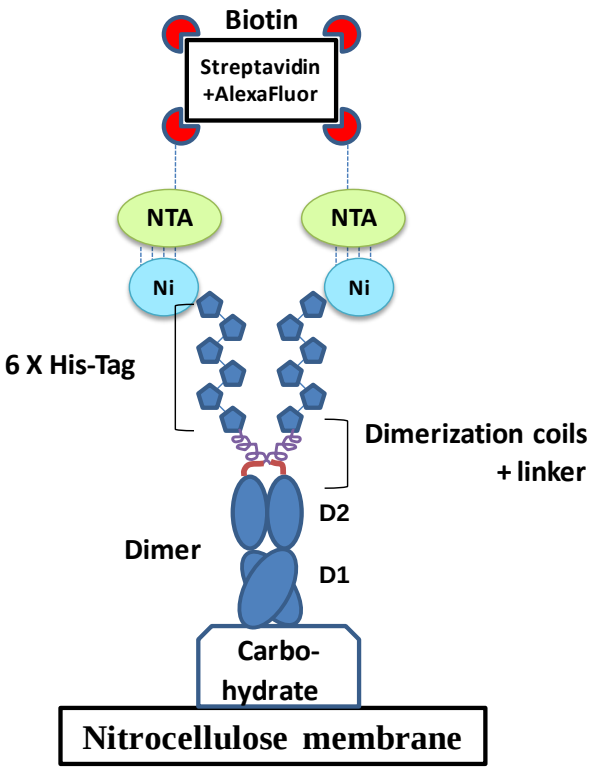

Figure 21. Schematic representation of carbohydrate binding macro-array with SRS dimers.
Purified recombinant SRS2 dimer containing His-tag was combined with a Ni^{2+}-NTA-Biotin-Streptavidin-AlexaFluor680 complex. This entire complex is expected to specifically bind carbohydrates, blotted onto the nitrocellulose membrane, and to be detected using the Licor/Odyssey system.

Results of carbohydrate binding macro-array

The initial screening of the potential SRS ligands using carbohydrate binding macro-array revealed that out of all SRSs tested, SRS2 dimer was found to bind heparin (Figure 22). Furthermore, with an increase in the concentration of heparin from 1 to 5 %, intensity of the spot increased significantly, indicating large amount of fluorescently labeled SRS2 dimer bound to heparin at higher concentration.

1. Heparin 1%	2. Heparin 5%	3. Heparan SO4 1%	4. Thyroglobulin 1%
5. Glycogen 1%	6. Glycogen 5%	7. Chondroitin SO4 1%	8. Chondroitin SO4 5%
9. Mucin II 1%	10. Mucin II 5%	11. Hialuronic acid 1%	12. Amylopectin 1%
13. SRS2 DIMER 1/2 dilution (1mg/ml)	14. SRS2 DIMER 1/10 dilution	15. SRS2 DIMER 1/50 dilution	16. Blank

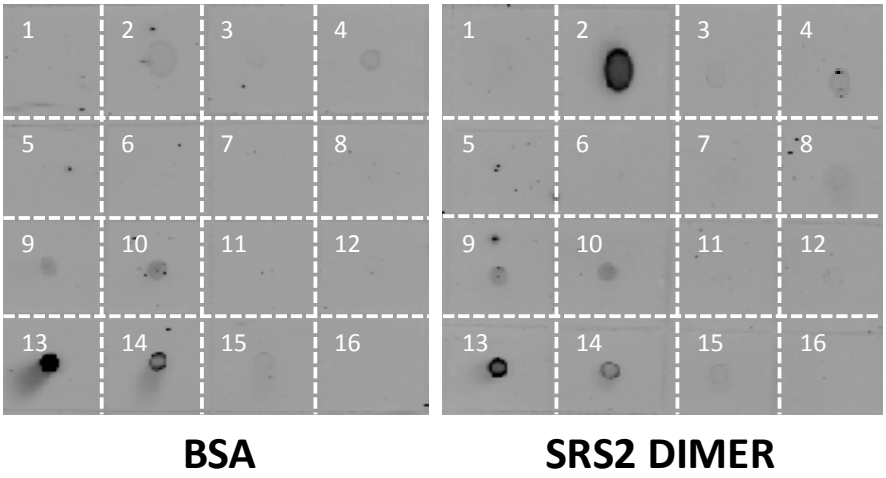

Figure 22. Results of carbohydrate binding macro-arrays.
A. Carbohydrate binding macro-array layout. **B.** Finalized carbohydrate macro-array included heparin sodium salt, mucin II from porcine stomach, heparan sulfate sodium salt from bovine kidney, chondroitin sulfate C sodium salt from shark cartilage, glycogen type IX from bovine liver, amylopectin, and hyaluronic acid sodium salt from *Streptococcus zoepidemicus*. BSA (left panel) serves as a negative control, where no protein was added during incubation, while SRS2 dimer blot (right panel) represents the blot to which protein was added.

Validating heparin binding

Once the potential ligand was identified the binding was confirmed by heparin-agarose beads pull-down experiment (Figure 23). Based on the fact that that SAG1 has been previously shown not to bind the heparin column (Jacquet *et al.* 2001), SAG1 dimer was used as a negative control in this experiment.

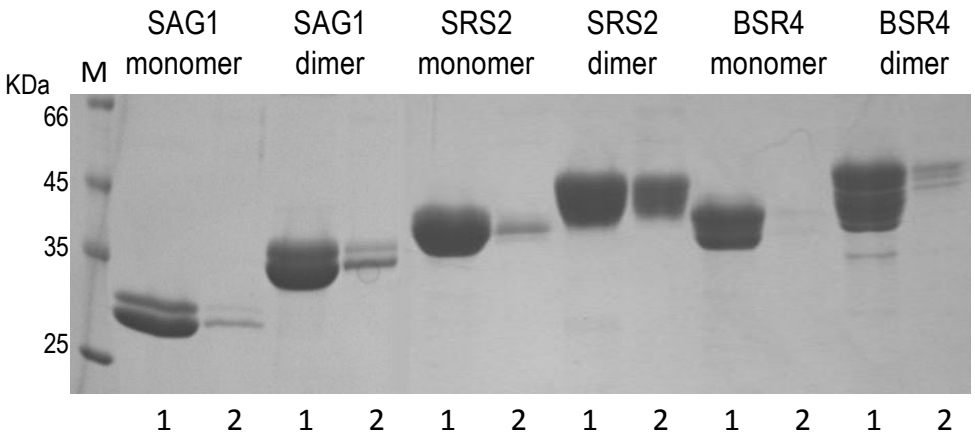

Figure 23. Results of heparin-agarose pull-down.
SDS-PAGE analysis of SAG1, SRS2 and BSR4 binding to heparin-agarose beads. "1" is a 10 µg of protein used in the experiment, while "2" is a pulled-down fraction that was eluted off the beads. High level of binding to heparin-agarose beads is detected for SRS2 dimer.

The results of the pull-down experiment show that the majority of the SRS2 dimer is retained by heparin beads. On the other hand, only residual amounts of monomeric SAG1, SRS2 and BSR4 and dimeric BSR4 and SAG1 were observed in the eluted fraction. To confirm direct interaction between heparin and SRS2 dimer, a soluble form of heparin with average MW of 18 KDa was tested at increasing molar ratios for specific binding to the native form of the protein and visualized using native gel electrophoresis (Figure 24).

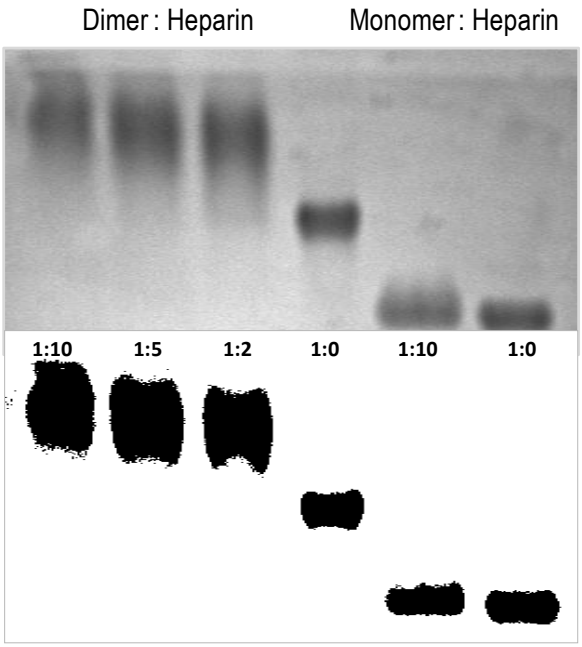

Figure 24. Confirmation of SRS2 dimer binding to heparin on native gel.
Native gel analysis of SRS2 dimer vs. monomer binding to soluble heparin at increasing molar ratios (upper panel). Highest contrast image of the same native gel (lower panel) reveals an increasing shift of the SRS2 dimer + heparin complex as heparin concentration increases, indicating a conformational change that occurs in SRS2 dimer form upon binding to heparin.

The native gel electrophoresis experiment has shown that as the concentration of heparin increases, there is an ascending shift of the protein-ligand complex observed, indicating a conformational change that occurs in SRS2 dimer form upon heparin binding.

4.6 Conclusions.

In this work, for the first time a direct interaction between *T. gondii* SRS and heparin was presented. Tachyzoite-specific SRS2 was shown to function as a heparin-binding SRS, similar to SAG3 which was shown bind heparin column previously (Jacquet *et al.* 2001). Our experiments have also confirmed inability of SAG1 to bind heparin, reported in the same SAG3 study. Our findings may be in correlation with the work of Carruthers *et al.* who found that soluble heparin in solution with *T. gondii* inhibits attachment of the

parasite to the serum-coated glass (Carruthers *et al.* 2000), indicating that the parasite attachment is heparin-dependent. This form of attachment is known for other parasites, such as *Plasmodium* (Pinzon-Ortiz *et al.* 2001) and viruses, such as respiratory syncytial virus (RSV) (Krusat and Streckert 1997). The fact that only the dimeric form of SRS2 was found to bind heparin may be an indicator of the functional advantages of multivalency in the SRS family. Structural basis for this phenomenon is discussed in the next chapter.

Chapter 5: Overview of the structural and functional findings in the context of *Toxoplasma gondii* virulence.

5.1 Connecting structural and functional data.

Studies of SAG-null mutants have shown that the parasites lacking SAG1 are able to attach and invade faster than wild type when characterized *in vitro* (Carruthers and Boothroyd 2007), while SAG3-deficient parasites exhibit only 50 % of the attachment capacity (Dzierszinski *et al.* 2000) suggesting that there are other SRSs involved in the attachment of *T. gondii* to the host cells. In this work, direct binding of tachyzoite specific SRS2 dimer to heparin was observed. Binding to heparin column was initially reported in SAG3 study (Jacquet *et al.* 2001). Structural basis for the interaction between heparin and SRSs was proposed by He *et al.* In their study, modeling of SAG3 onto the SAG1dimer structure revealed a highly basic groove which mey attract and accommodate negatively charged molecules, such as heparin (He *et al.* 2002).

Putative ligand binding groove in SRS2 dimer

In order to visualize this putative ligand binding groove in SRS2, the individual SRS2 domains were overlaid onto the SAG1 dimer to generate an SRS2 dimer (Figure 25). The resulting structure displayed a high degree of conservation reflected by average r.m.s.d of 1.13 Å over 401 residues.

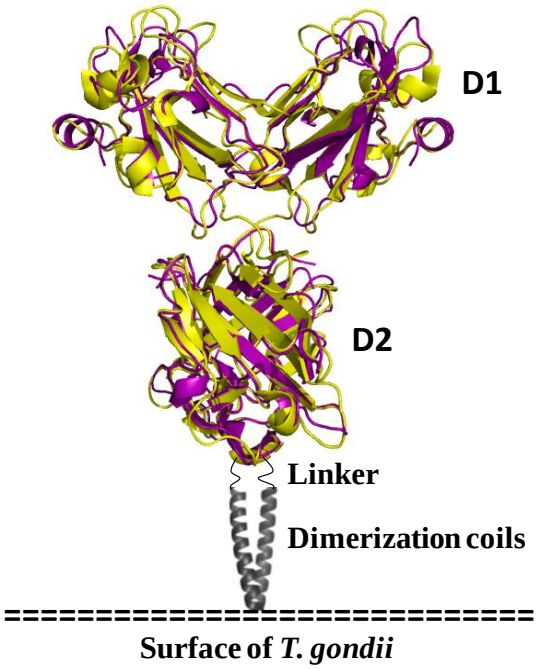

Figure 25. SRS2 dimer model.
The model was created by structural overlays of SRS2 individual domains D1 and D2 (purple) onto SAG1 dimer (yellow) (PDB ID: 1KZQ), using SWISS-PDB viewer (Guex *et al.* 1996), PYMOL (Delano 2002) and COOT (Emsley and Cowtan 2004). Average r.m.s.d is 1.13 Å over 401 residues. Each SRS2 monomer was co-expressed with a dimerization coil (grey helix), which makes side-to-side interactions with the coil of the other monomer, facilitating dimerization in solution.

Putative ligand binding groove in BSR4 dimer

To visualize the putative ligand binding groove in BSR4 modelling was not required due to the fact that the analysis of the molecular packing of BSR4 in the tetragonal unit cell revealed a symmetry-related molecule oriented to form a crystallographic dimer of BSR4 (Crawford *et al.* 2009a). Overall, the dimer interface is formed by ten backbone hydrogen bonds with six contributed by the upper portion of the beta-sandwich (strand "a") and four from the lower portion (strand "c"). The contiguous strands in BSR4 show unambiguous electron density (Figure 26 B) including the connectivity between the “a” and “b” strands that form the basis of the structurally divergent topology relative to SAG1. The D2 domains in BSR4 are positioned away from each other and do not

contribute to the dimer formation, in contrast to SAG1, where the D2 domains contribute nearly 40 % of the buried surface area.

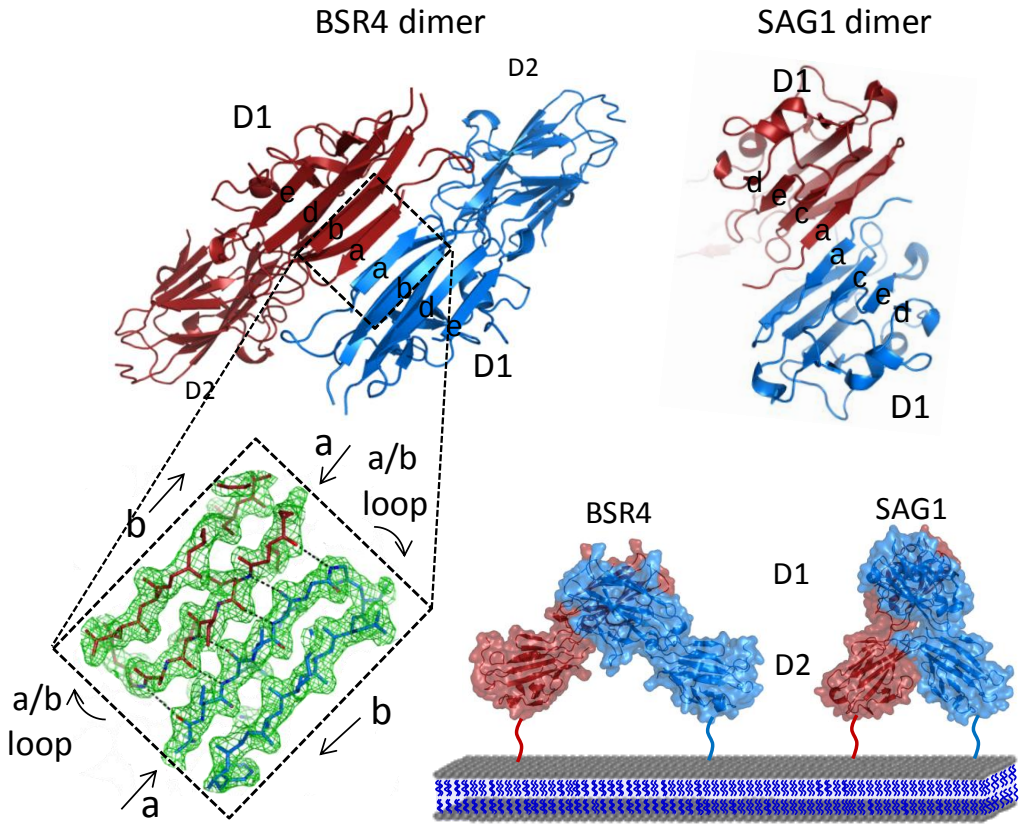

Figure 26. BSR4 dimer.
A. Organization of two symmetry related BSR4 monomers (blue and burgundy) to form the BSR4 dimer (left panel) and the SAG1 dimer (right panel). **B.** Omit electron density map contoured at 1.5 σ in right hand panel shows the ordered residues of the "a" and "b" strands and hydrogen bonds that form the interface of the symmetry related "a" strands that form the BSR4 dimer. Note the clear electron of the "a/b" loop which highlights the unique topology of BSR4 as compared with SAG1. Figure adapted with permission from (Crawford *et al.* 2009a).

Putative ligand binding groove comparison

Surface electrostatic maps next were calculated for SRS2, BSR4 and SAG1 dimers (Figure 27), which allowed a direct comparison of the putative ligand binding groove. It then became evident that in SRS2 and SAG1 the shape of the groove is conserved. In BSR4, on the other hand, the topologically formed surface cleft is very shallow (Figure 27, left panel), which is consistent with the initial characterization of the D1 domain that adopts a flat structure illustrated in Figure 13. The different shape of the groove can be explained by the relatively short strands of the beta-sandwich core structure in SAG1 and SRS2 D1 domains, resulting in a compact, twisted structure that relies on interstrand loops to form the sides of the of the deep groove. In BSR4 the beta-strands are about 40 % longer that those in SAG1 and SRS2, resulting in an extended, flatter surface of the putative ligand binding site. Furthermore, the groove in SAG1 and SRS2 is lined only with basic residues consistent with the idea of them binding negatively charged molecules. The groove in SRS2, however, has higher number of positively charged residues than SAG1, which may rationalize binding of heparin to SRS2 dimer and not to SAG1dimer. In BSR4, Arg^{6} and Lys^{117} combine to create a horseshoe type structure of positive charge that partially encircles a pocket of negative charge formed by Asp^{17} and Glu^{45}. Lys^{15} is also present on the surface but adopts a flattened structure such that the methylene side-chain is exposed and the amino group is directed away from the central core of the dimer interface. The flatter surface of BSR4 combined with the discrete positively and negatively charged regions strongly suggests that BSR4 binds a ligand structurally and chemically distant from SAG1 and SRS2.

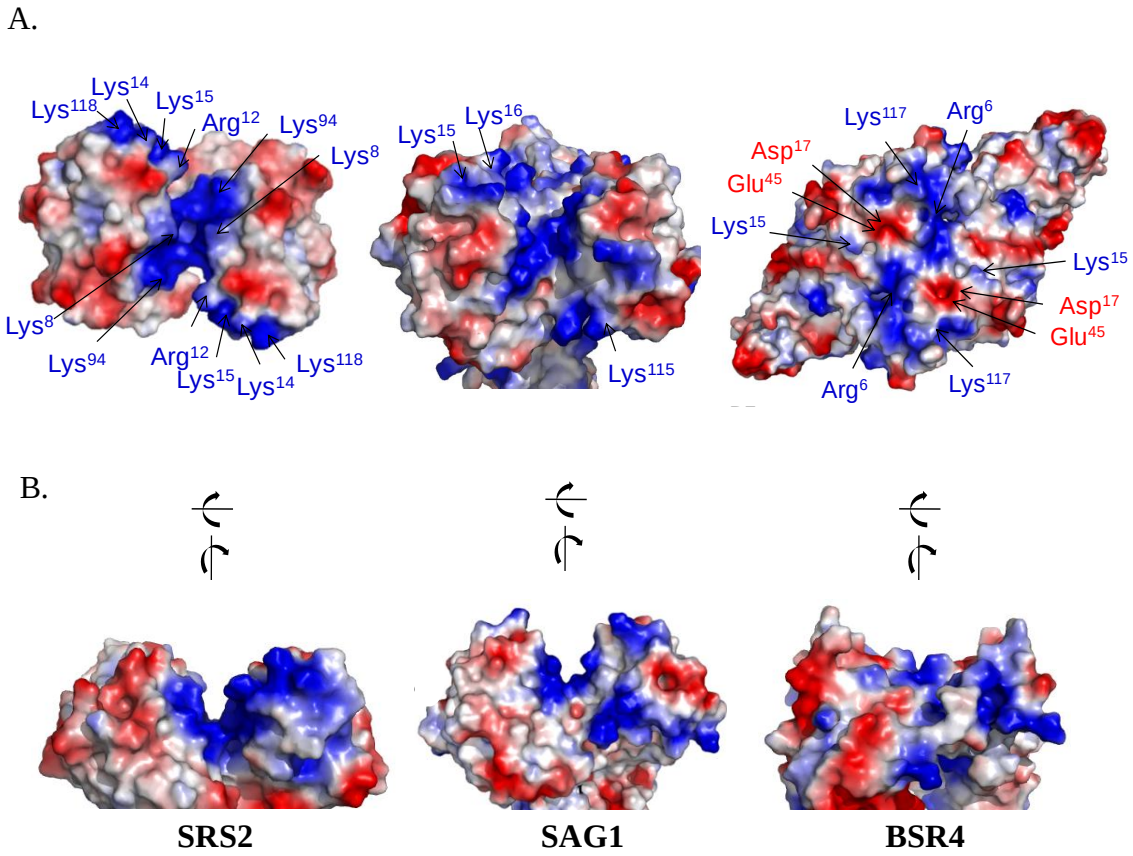

Figure 27. Structural implications for ligand binding.
A. Isoelectric surface potential calculated for the SRS2, SAG1 and BSR4 dimers and displayed in surface format from a top view perspective. The topologically defined surface of the BSR4 dimer is composed of equally acidic and basic residues, while in SRS2, only basic residues are involved in the groove formation. Conserved residues present in the groove of all three proteins are shown in SAG1. **B.** Side views of the electrostatic surfaces of the dimerized D1 domains of SRS2, SAG1 and BSR4 highlighting the architectural and electrostatic difference in the topologically defined surface groove. Isoelectric surface potential for BSR4 adapted with permission from (Crawford *et al.* 2009a).

5.2 Heparin and microbial virulence.

In this work, various carbohydrates were tested in the nitrocellulose based binding assays described in detail in Chapter 4. The dimer form of the tachyzoite-specific surface protein, SRS2, was found to selectively bind heparin. Heparin is a highly sulfated form of the heparan sulfate (HS) glucosamioglycan (GAG), which is attached to the protein core to form cell surface HSPGs (Bishop *et al.* 2007). *In vivo*, heparin is synthesized and stored within the secretory granules of mast cells, especially in the gastrointestinal tract and skin (Gandhi and Mancera 2008). In addition, it has been shown that the heparin-like molecules, highly modifed blocks of heparan sulfate (Carey 1997), are also present in liver hepatocytes (Lyon *et al.* 1994) and oligodendrocyte-type-2 astrocyte progenitor cells (Stringer *et al.* 1999). Native heparin consists of various disaccharide repeating units, including GlcA-GlcNAc, GlcA-GlcNS, IdoA-GlcNS, IdoA(2S)-GlcNS, IdoA-GlcNS(6S), IdoA(2S)-GlcNS(6S), of which the most common unit is IdoA(2S)-GlcNS(6S) (Gandhi and Mancera 2008). Heparin is the most negatively charged biomolecule and contains, on average, 2.4 sulfate groups per disaccharide (Rostand and Esko 1997). Due to chain length variations, the molecular weight of native heparin varies from 3 to 50 kDa, while commercially available heparin utilized in these studies has an average molecular weight of 18 kDa.

Multiple biological roles of heparin include anti-coagulation, cell growth, microbial attachment, and pro-inflammatory reactions (Gandhi and Mancera 2008). Heparin has been shown to have an anti-microbial activity due to the ability to inhibit or block microbs entry into host cells. This has been demonstrated for herpes simplex virus (HSV) (WuDunn and Spear 1989) and respiratory syncytial virus (RSV) (Krusat and Streckert

1997). In addition, broad range of pathogenic organisms has been shown to bind heparin and heparan sulfate to mediate infection, such as *Helicobacter pylori* (Dubreuil et al. 2002), *Bordetella pertussis* (Locht *et al.* 1993), and *Plasmodium falciparum* (Frevert *et al.* 1993). However, the role of heparin in *T. gondii* infection is yet to be determined.

Heparin has also been widely used in multiple *in vitro* studies to demonstrate and characterize molecular interactions between HSPGs and their microbial receptors (Barragan *et al.* 2000; Harper *et al.* 2004; Jenkins *et al.* 2006; McCormick *et al.* 1999). Heparin binding by SRS2 described here is one of the first steps to understanding interactions between negatively charged heparin-like HSPGs and *T. gondii* surface receptors.

5.3 Polymorphism maps.

Structural information obtained in this study can be utilized not only to discuss a basis for *T. gondii* ligand binding, but also to create polymorphism maps for these proteins. The fact that type I strains are highly virulent and lethal and are mostly responsible for congenital toxoplasmosis (Barragan and Sibley 2002), while type II and III are relatively avirulent and tend to establish latent chronic infection (Grigg *et al.* 2001) is attributed to the inter-typic polymorphism found in many genes of this parasite.

It is beyond the scope of this project to analyze a large body of polymorphism data for *T. gondii* surface proteins. However, the structural information about SRSs expressed on the surface of different developmental stages of *T. gondii* can be combined with the amino acid polymorphism data to create polymorphism maps for these proteins (He *et al.* 2002). These maps allow localization of the regions in the molecule with a high degree of

polymorphism. On the surface of the parasitic antigens, these regions typically occur as a result of host immune pressure (Grigg *et al.* 2001).

Polymorphism maps of BSR4 and SRS2

Mapping amino acid polymorphism data (Tables 4, 5 and 6 of Appendix), onto the structures of BSR4, SRS2 and SAG1 revealed major differences in the polymorphism pattern between SRSs expressed on the surface of tachyzoite (SAG1 and SRS2) and bradyzoite (BSR4) form of *T. gondii* (Figure 27). In BSR4, polymorphic sites are numerous and located mainly on the solvent exposed loops of D1 domain, while in SRS2 and SAG1 polymorphic sites are rare and located mostly in the D2 domain. Such observations were not expected since SAG1 and SRS2 are major antigens on the surface of tachyzoites responsible for acute virulence and the dissemination of the parasite via blood stream. This means that SAG1 and SRS2 are under constant pressure from the host immune system and are expected to display a high degree of polymorphism in their surface exposed loops of the D1 domain. Why are SAG1 and SRS2 highly conserved in the regions that are suggested to be under immune pressure? Considering that the main goal of *T. gondii* is to establish life-long, latent infection and not to kill the host, attracting immune response to the tachyzoite stage of the parasite prevents extensive tissue damage during acute infection and host death, providing evolutionary advantage to the parasite. As a result, perhaps, SRS2 and SAG1 have evolved to attract immune response to the tachyzoite form of the parasite to ensure host survival during acute infection. The polymorphism pattern of BSR4 is very different from SAG1 and SRS2 since the most of the polymorphic sites in BSR4 are concentrated in the surface loops of its D1 domain. Other SRSs, expressed on the surface of bradyzoites, such as SAG2C and

SAG2D, display a pattern similar to BSR4 with prevalence of the polymorphic sites at D1 domains (He *et al.* 2002), which suggests a stage-specific character of the polymorphism pattern among SRSs.

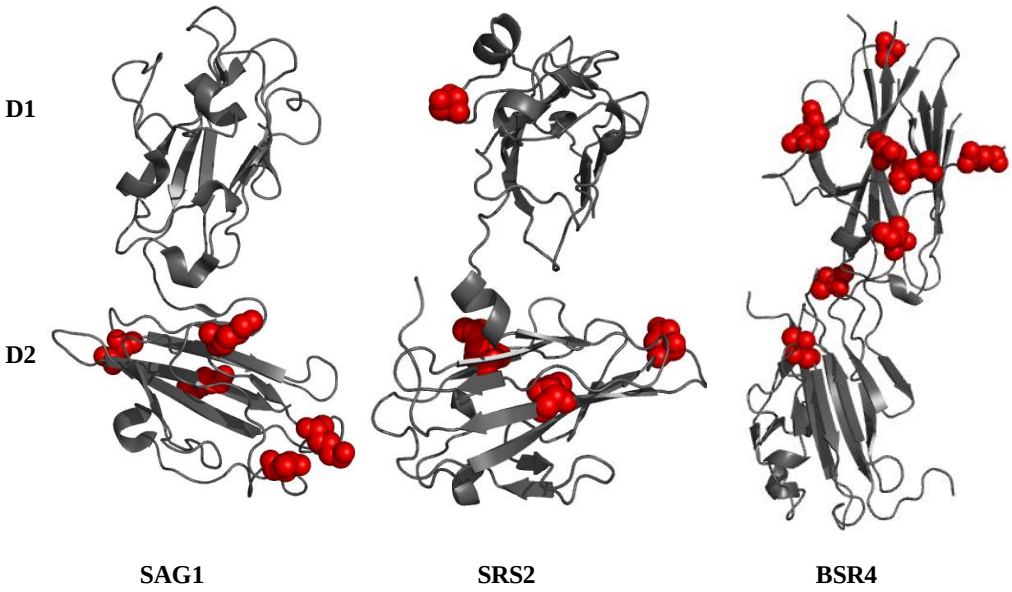

Figure 28. Amino acid polymorphism maps in SAG1, SRS2 and BSR4.
Mapping of amino acid polymorphic sites onto the structures of SAG1, SRS2 and BSR4 reveals extensive polymorphisms in BSR4 D1, but high degree of conservation in SAG1 and SRS2 D1 domains.

5.4 Conclusions.

T. gondii's success as a pathogen depends on its ability to adhere to a host cell and to persist in the face of a fully competent immune system. The parasitic surface proteins, such as SRSs, are perfectly located not only to mediate initial attachment of the parasite to the host cell, but also to interact with the host immune system.

In this thesis, it was hypothesized that the core SRS fold is conserved among members of the sequence-defined SRS superfamily, and that the developmentally expressed SRSs are most likely to display structurally divergent loops. Structural characterization of SRSs expressed at the different developmental stages revealed a high degree of conservation of the core SRS fold in all structurally characterized proteins. At the same time, the bradyzoite-expressed BSR4 displayed divergent loop structure in the N-terminal domain, which indicated that SRS2 and BSR4 may bind differential ligands. Since previous study by He *et al.* has shown that the putative ligand binding groove is present on SAG1 dimer, dimer forms of SRS2 and BSR4 were successfully engineered, produced and tested for carbohydrate binding. SRS2 was found to selectively bind heparin in a dimer dependent manner. However, it cannot be concluded whether heparin is a physiological ligand for SRS2 and further investigation is needed to determine a role of heparin and other HSPGs in *T. gondii* virulence *in vivo.*

5.5 Future directions.

The experiments described here have contributed to our knowledge of the basic structural features of the SRS superfamily of *T. gondii* surface proteins. Structural characterization of BSR4 and SRS2 allowed us to look at the recently identified SRS fold with respect to

stage-specificity. At the same time, carbohydrate binding assays permitted identification of one of the SRS ligands, heparin. However, the exact molecular interactions between these SRSs and their ligands can only be described after co-crystallization of these compounds with discrete units of heparin.

In order to study SRS-mediated *T. gondii* adhesion, various adhesion assays with cultured cells, tissue sections, or plastic surfaces coated with HSPGs can be developed. Inhibition of *T. gondii* adherence by the addition of HSPGs had been done previously by several authors (Carruthers *et al.* 2000; Jacquet *et al.* 2001; Mineo and Kasper 1994), however, direct attachment of *T. gondii* surface proteins to the host cells has not been visualized to this date (Rostand and Esko 1997). The next series of experiments will likely include fluorescently labelling representative SRSs, such as SAG1, SRS2 and BSR4 in order to probe various tissues and cells.

Robinson *et al.* have previously examined a direct interaction between SAG1 and MDBK (Madin-Darby bovine kidney) cells and human foreskin fibroblasts using immuno-blotting (Robinson *et al.* 2004). However, visualization of the direct SRS-host cell interaction using confocal microscopy will provide valuable insights into the mechanism of SRS binding to the host cells. Once binding is detected, host modified cell cultures will be used to confirm the mechanism of binding and to examine how removal of HSPGs from target cells affects adhesion.

Mutagenesis studies can also be conducted in order to confirm that the SRS2-ligand interaction is mediated via the putative ligand binding groove originally proposed in the SAG1 study by He *et al.* Basic residues inside this groove can be mutated to examine the effects of this change in their charge on binding to the ligand, while complementation

studies with SRS2 mutant will allow to determine role of specific amino acids in *T. gondii* virulence.

These strategies will allow identifying, visualizing and confirming *T. gondii* ligands in various tissues. As a result of these studies, the mechanism of *T. gondii* attachment will be revealed, which may lead to the development of molecular inhibitors of *T. gondii* infection.

Bibliography

Ajioka J, Boothroyd J, Brunk B, Hehl A, Hillier L, Manger I, Marra M, Overton G, Roos D, Wan K. 1998. Gene Discovery by EST Sequencing in Toxoplasma gondiiReveals Sequences Restricted to the Apicomplexa. Genome research 8(1):18.

Ajioka JA, Soldati D, Soldati A. 2007. Toxoplasma: Molecular and Cellular Biology: Horizon Scientific Press.

Alexander DL, Mital J, Ward GE, Bradley P, Boothroyd JC. 2005. Identification of the Moving Junction Complex of Toxoplasma gondii: A Collaboration between Distinct Secretory Organelles. PLoS Pathog 1(2):e17.

Aliberti J. 2005. Host persistence: exploitation of anti-inflammatory pathways by *Toxoplasma gondii.* Nature Reviews Immunology 5(2):162-170.

Barragan A, Fernandez V, Chen Q, von Euler A, Wahlgren M, Spillmann D. 2000. The Duffy-binding-like domain 1 of Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) is a heparan sulfate ligand that requires 12 mers for binding. Blood 95(11):3594-3599.

Barragan A, Sibley LD. 2002. Transepithelial migration of Toxoplasma gondii is linked to parasite motility and virulence. Journal of Expimental Medicine 195(12):1625-33.

Beghetto E, Spadoni A, Buffolano W, Del Pezzo M, Minenkova O, Pavoni E, Pucci A, Cortese R, Felici F, Gargano N. 2003. Molecular dissection of the human B-cell response against Toxoplasma gondii infection by lambda display of cDNA libraries. International Journal for Parasitology 33(2):163-173.

Bentley GA. 2006. Functional and immunological insights from the three-dimensional structures of Plasmodium surface proteins. Current Opinion in Microbiology 9(4):395-400.

Bergmann A, Fritz G, Glatter O. 2000. Solving the generalized indirect Fourier transformation (GIFT) by Boltzmann simplex simulated annealing (BSSA). Journal of Applied Crystallography 33:1212-1216.

Bishop JR, Crawford BE, Esko JD. 2005. Cell Surface Heparan Sulfate Promotes Replication of Toxoplasma gondii. Infection and Immunity 73(9):5395-5401.

Bishop JR, Schuksz M, Esko JD. 2007. Heparan sulphate proteoglycans fine-tune mammalian physiology. Nature 446(7139):1030-1037.

Blondel A, Bedouelle H. 1991. Engineering the quaternary structure of an exported protein with a leucine zipper. Protein Engineering 4(4):457-461.

Bond CS. 2003. TopDraw: a sketchpad for protein structure topology cartoons. Bioinformatics 19(2):311-2.

Bowie WR, King AS. 1997. Outbreak of toxoplasmosis associated with municipal drinking water. Lancet 350(9072):173.

Brenier-Pinchart M-P, Villena I, Mercier C, Durand F, Simon J, Cesbron-Delauw M-F, Pelloux H. 2006. The Toxoplasma surface protein SAG1 triggers efficient in vitro secretion of chemokine ligand 2 (CCL2) from human fibroblasts. Microbes and Infection 8(1):254-261.

Caiolfa VR, Zamai, M., Malengo, G., Andolfo, A., Madesn, C.D., Sutin, J., Digman, M.A., Gratton, E., Blasi, F. and Sidenius, N. 2007. Monomer–dimer dynamics and distribution of GPI-anchored uPAR are determined by cell surface protein assemblies Journal of Cell Biology 179(5):1067-1082.

Carey D. 1997. Syndecans: multifunctional cell-surface co-receptors. Biochemical Journal 327(Pt 1):1.

Carruthers V, Boothroyd JC. 2007. Pulling together: an integrated model of Toxoplasma cell invasion. Current Opinion in Microbiology 10(1):83-89.

Carruthers VB, Hakansson S, Giddings OK, Sibley LD. 2000. Toxoplasma gondii Uses Sulfated Proteoglycans for Substrate and Host Cell Attachment. Infect. Immun. 68(7):4005-4011.

Channon JY, Suh EI, Seguin RM, Kasper LH. 1999. Attachment Ligands of Viable Toxoplasma gondii Induce Soluble Immunosuppressive Factors in Human Monocytes. Infection and Immunity 67(5):2547-2551.

Collaborative Computational Project N. 1994. The CCP4 suite: programs for protein crystallography. p 760-763.

Cowman AF, Crabb BS. 2006. Invasion of red blood cells by malaria parasites. Cell 124(4):755-766.

Crawford J, Grujic O, Bruic E, Czjzek M, Grigg M, Boulanger M. 2009a. Structural Characterization of the Bradyzoite Surface Antigen (BSR4) from Toxoplasma gondii, a Unique Addition to the Surface Antigen Glycoprotein 1-related Superfamily. Journal of Biological Chemistry 284(14):9192.

Crawford J, Lamb E, Wasmuth J, Grujic O, Grigg ME, Boulanger MJ. 2009b. Structural and functional characterization of SporoSAG: A SAG2 related surface antigen from *Toxoplasma gondii.*

Deitsch K, Moxon E, Wellems T. 1997. Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. Microbiology and Molecular Biology Reviews 61(3):281-293.

Delano WL. 2002. The PyMOL Molecular Graphics System

Donald RGK, Louis MW, Kami K. 2007. Toxoplasma as a model system for apicomplexan drug discovery. Toxoplasma Gondii. London: Academic Press. p 505-539.

Dubey JP. 1998a. Advances in the life cycle of Toxoplasma gondii. International Journal of Parasitology 28(7):1019-24.

Dubey JP. 1998b. Comparative Infectivity of Toxoplasma gondii Bradyzoites in Rats and Mice. The Journal of Parasitology 84(6):1279-1282.

Dubey JP, Lindsay DS, Speer CA. 1998. Structures of Toxoplasma gondii Tachyzoites, Bradyzoites, and Sporozoites and Biology and Development of Tissue Cysts. Clinical Microbiology Reviews 11(2):267-299.

Dubey JP, Murrell KD, Fayer R, Schad GA. 1986. Distribution of Toxoplasma gondii tissue cysts in commercial cuts of pork. Journal of the American Veterinary Medical Association 188(9):1035-7.

Dubreuil JD, Giudice GD, Rappuoli R. 2002. Helicobacter pylori Interactions with Host Serum and Extracellular Matrix Proteins: Potential Role in the Infectious Process. Microbiology and Molecular Biology Reviews 66(4):617-629.

Dyrløv Bendtsen J, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. Journal of Molecular Biology 340(4):783-795.

Dzierszinski F, Mortuaire M, Cesbron-Delauw MF, Tomavo S. 2000. Targeted disruption of the glycosylphosphatidylinositol-anchored surface antigen SAG3 gene in Toxoplasma gondii decreases host cell adhesion and drastically reduces virulence in mice. Molecular Microbiology 37(3):574-82.

Elevates CT. 2008. Antischizophrenic drugs. Science and psychiatry: groundbreaking discoveries in molecular neuroscience.: American Psychiatric Pub. p 113.

Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. Acta Crystallographica section D 60:2126-2132.

Flegr J. 2007. Effects of Toxoplasma on human behavior. Schizophrenia bulletin 33(3):757.

Frevert U, Sinnis P, Cerami C, Shreffler W, Takacs B, Nussenzweig V. 1993. Malaria circumsporozoite protein binds to heparan sulfate proteoglycans associated with

the surface membrane of hepatocytes. Journal of Experimental Medicine 177(5):1287.

Gandhi NS, Mancera RL. 2008. The Structure of Glycosaminoglycans and their Interactions with Proteins. Chemical Biology and Drug Design 72(6):455-482.

Garweg JG. 2005. Determinants of immunodiagnostic success in human ocular toxoplasmosis. Clinical and Experimental Medicine 27(3):61-68.

Gaskell EA, Smith JE, Pinney JW, Westhead DR, McConkey GA. 2009. A Unique Dual Activity Amino Acid Hydroxylase in *Toxoplasma gondii*. PLoS ONE 4(3):e4801.

Gerloff DL, Creasey A, Maslau S, Carter R. 2005. Structural models for the protein family characterized by gamete surface protein Pfs230 of Plasmodium falciparum. PNAS 102(38):13598-13603.

Gouet P, Robert X, Courcelle E. 2003. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res 31(13):3320.

Graille M, Stura EA, Bossus M, Muller BH, Letourneur O, Battail-Poirot N, Sibaï G, Gauthier M, Rolland D, Le Du M-H and others. 2005. Crystal Structure of the Complex between the Monomeric Form of Toxoplasma gondii Surface Antigen 1 (SAG1) and a Monoclonal Antibody that Mimics the Human Immune Response. Journal of Molecular Biology 354(2):447-458.

Grigg ME, Bonnefoy S, Hehl AB, Suzuki Y, Boothroyd JC. 2001. Success and Virulence in Toxoplasma as the Result of Sexual Recombination Between Two Distinct Ancestries. Science 294(5540):161-165.

Grimwood J, Smith JE. 1992. Toxoplasma gondii: the role of a 30-kDa surface protein in host cell invasion. Expimental Parasitology 74(1):106-11.

Grimwood J, Smith JE. 1996. Toxoplasma gondii: the role of parasite surface and secreted proteins in host cell invasion. International Journal for Parasitology 26(2):169-173.

Grujic O, Grigg ME, Boulanger MJ. 2008. Insect-cell expression, crystallization and X-ray data collection of the bradyzoite-specific antigen BSR4 from Toxoplasma gondii. Acta Crystallographica Section F 64(Pt 5):425-7.

Guex N, Diemand A, Schwede T, Peitsch M. 1996. Swiss PDB Viewer.

Guex N, Peitsch M. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18(15):2714-2723.

Guinier A, Fournet F. 1955. Small angle scattering of C-rays: Wiley Interscience.

Harper J, Hoff E, Carruthers V. 2004. Multimerization of the Toxoplasma gondii MIC2 integrin-like A-domain is required for binding to heparin and human cells. Molecular & Biochemical Parasitology 134(2):201-212.

He XL, Grigg ME, Boothroyd JC, Garcia KC. 2002. Structure of the immunodominant surface antigen from the Toxoplasma gondii SRS superfamily. Nature Structural Biology 9(8):606-11.

Howe DK, Crawford AC, Lindsay D, Sibley LD. 1998. The p29 and p35 Immunodominant Antigens of Neospora caninum Tachyzoites Are Homologous to the Family of Surface Antigens of Toxoplasma gondii. Infection and Immunity 66(11):5322-5328.

Howe DK, Gaji RY, Mroz-Barrett M, Gubbels M-J, Striepen B, Stamper S. 2005. Sarcocystis neurona Merozoites Express a Family of Immunogenic Surface Antigens That Are Orthologues of the Toxoplasma gondii Surface Antigens (SAGs) and SAG-Related Sequences. Infection and Immunity 73(2):1023-1033.

Howe DK, Sibley LD. 1999. Comparison of the major antigens of Neospora caninum and Toxoplasma gondii. International Journal for Parasitology 29(10):1489-1496.

Jacquet A, Coulon L, De Nève J, Daminet V, Haumont M, Garcia L, Bollen A, Jurado M, Biemans R. 2001. The surface antigen SAG3 mediates the attachment of Toxoplasma gondii to cell-surface proteoglycans. Molecular and Biochemical Parasitology 116(1):35-44.

Jenkins C, Wilton JL, Minion FC, Falconer L, Walker MJ, Djordjevic SP. 2006. Two Domains within the Mycoplasma hyopneumoniae Cilium Adhesin Bind Heparin. Infect. Immun. 74(1):481-487.

Jones JL, Dubey JP. 2009. Waterborne toxoplasmosis–Recent developments. Experimental Parasitology.

Joynson DHM, Wreghitt TG. 2001. Toxoplasmosis: a comprehensive clinical guide: Cambridge University Press.

Jung C, Lee CY, Grigg ME. 2004. The SRS superfamily of Toxoplasma surface proteins. International Journal for Parasitology 34(3):285-96.

Kim K, Weiss LM. 2004. Toxoplasma gondii: the model apicomplexan: Elsevier. 423-432 p.

Kim S-K, Boothroyd JC. 2005. Stage-Specific Expression of Surface Antigens by Toxoplasma gondii as a Mechanism to Facilitate Parasite Persistence. Journal of Immunology 174(12):8038-8048.

Kim S-K, Karasov A, Boothroyd JC. 2007. Bradyzoite-Specific Surface Antigen SRS9 Plays a Role in Maintaining Toxoplasma gondii Persistence in the Brain and in

Host Control of Parasite Replication in the Intestine. Infection and Immunity 75(4):1626-1634.

Kissinger J, Kuo C. 2007. Evolution and Comparative Genomics of Toxoplasma gondii. In: Ajioka JW, Soldati D, editors. Toxoplasma: molecular and cellular biology: Routledge. p 211.

Knoll LJ, Boothroyd JC. 1998. Isolation of developmentally regulated genes from Toxoplasma gondii by a gene trap with the positive and negative selectable marker hypoxanthine-xanthine-guanine phosphoribosyltransferase. Molecular and Cellular Biology 18(2):807-14.

Koch M, Bordas J. 1983. J. Nucl. Instrum. Methods 208:461-469.

Konarev PV, Volkov VV, Sokolova AV, Koch MH, Svergun DI. 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography 36:1277-1282.

Koppe J, Loewer-Sieger D, de Roever-Bonnet H. 1986. Results of 20-year follow-up of congenital toxoplasmosis. The Lancet 327(8475):254-256.

Krusat T, Streckert H. 1997. Heparin-dependent attachment ofrespiratory syncytial virus (RSV) to host cells. Archives of virology 142(6):1247-1254.

Lekutis C, Ferguson DJP, Grigg ME, Camps M, Boothroyd JC. 2001. Surface antigens of Toxoplasma gondii: variations on a theme. International Journal for Parasitology 31(12):1285-1292.

Locht C, Bertin P, Menozzi F, Renauld G. 1993. The filamentous haemagglutinin, a multifaceted adhesin produced by virulent Bordetella spp. Molecular Microbiology 9:653-653.

Lyon M, Deakin JA, Gallagher JT. 1994. Liver heparan sulfate structure. A novel molecular design. Journal of Biological Chemistry 269(15):11208-11215.

Manger ID, Hehl A, Parmley S, Sibley LD, Marra M, Hillier L, Waterston R, Boothroyd JC. 1998. Expressed Sequence Tag Analysis of the Bradyzoite Stage of Toxoplasma gondii: Identification of Developmentally Regulated Genes. Infection and Immunity 66(4):1632-1637.

Marchler-Bauer A, Bryant SH. 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Research 32(Web Server Issue):W327.

McCormick C, Tuckwell D, Crisanti A, Humphries M, Hollingdale M. 1999. Identification of heparin as a ligand for the A-domain of Plasmodium falciparum thrombospondin-related adhesion protein. Molecular & Biochemical Parasitology 100(1):111-124.

Mineo JR, Kasper LH. 1994. Attachment of Toxoplasma gondii to host cells involves major surface protein, SAG-1 (P30). Molecular Immunology 31(17):1353-63.

Mineo JR, McLeod R, Mack D, Smith J, Khan IA, Ely KH, Kasper LH. 1993. Antibodies to Toxoplasma gondii major surface protein (SAG-1, P30) inhibit infection of host cells and are produced in murine intestine after peroral infection. Journal of Immunology 150(9):3951-3964.

Montoya JG, Liesenfeld O. 2004. Toxoplasmosis. The Lancet 363(9425):1965-1976.

Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallographica section D 53:240-255.

O'Shea EK, Klemm JD, Kim PS, Alber T. 1991. X-ray Structure of the GCN4 Leucine Zipper, a Two-Stranded, Parallel Coiled Coil. Science 254(5031):539-544.

Oksanen A, Tryland M, Johnsen K, Dubey JP. 1998. Serosurvey of Toxoplasma gondii in North Atlantic marine mammals by the use of agglutination test employing whole tachyzoites and dithiothreitol. Comparative Immunology, Microbiology and Infectious Diseases. 21(2):107-114.

Ortega-Barria E, Boothroyd JC. 1999a. A Toxoplasma lectin-like activity specific for sulfated polysaccharides is involved in host cell infection. Journal of Biological Chemistry 274(3):1267-76.

Ortega-Barria E, Boothroyd JC. 1999b. A Toxoplasma lectin-like activity specific for sulfated polysaccharides is involved in host cell infection. J Biol Chem 274(3):1267-76.

Perrakis A, Morris R, Lamzin VS. 1999. Automated protein model building combined with iterative structure refinement. Nature structural biology 6:458-463.

Petoukhov MV, Eady NA, Brown KA, Svergun DI. 2002. Addition of missing loops and domains to protein models by x-ray solution scattering. Biophysical Journal 83(6):3113-25.

Pharmingen. 1999. Baculovirus Expression Vector System; Instruction Manual.

Pinzon-Ortiz C, Friedman J, Esko J, Sinnis P. 2001. The Binding of the Circumsporozoite Protein to Cell Surface Heparan Sulfate Proteoglycans Is Required for PlasmodiumSporozoite Attachment to Target Cells. Journal of Biological Chemistry 276(29):26784.

Putnam CD, Hammel M, Hura GL, Tainer JA. 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys 40(3):191-285.

Rachinel N, Buzoni-Gatel D, Dutta C, Mennechet FJD, Luangsay S, Minns LA, Grigg ME, Tomavo S, Boothroyd JC, Kasper LH. 2004. The Induction of Acute Ileitis by a Single Microbial Antigen of Toxoplasma gondii. Journal of Immunology 173(4):2725-2735.

Risco-Castillo V, Fernández-García A, Zaballos A, Aguado-Martínez A, Hemphill A, Rodríguez-Bertos A, Álvarez-García G, Ortega-Mora LM. 2007. Molecular characterisation of BSR4, a novel bradyzoite-specific gene from Neospora caninum. International Journal for Parasitology 37(8-9):887-896.

Robinson SA, Smith JE, Millner PA. 2004. Toxoplasma gondii major surface antigen (SAG1): in vitro analysis of host cell binding. Parasitology 128(04):391-396.

Rorman E, Zamir CS, Rilkis I, Ben-David H. 2006. Congenital toxoplasmosis--prenatal aspects of Toxoplasma gondii infection. Reproductive Toxicology 21(4):458-472.

Rostand KS, Esko JD. 1997. Microbial adherence to and invasion through proteoglycans. Infection and Immunity 65(1):1-8.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution 4(4):406-25.

Sayers E, Barrett T, Benson D, Bryant S, Canese K, Chetvernin V, Church D, DiCuccio M, Edgar R, Federhen S. 2008. Database resources of the national center for biotechnology information. Nucleic Acids Res.

Schwarzenbacher R, Godzik A, Grzechnik SK, Jaroszewski L. 2004. The importance of alignment accuracy for molecular replacement. Acta Crystallographica section D D60:1229-1236.

Sharma P, Varma R, Sarasij R, Gousset K, Krishnamoorthy G, Rao M, Mayor S. 2004. Nanoscale organization of multiple GPI-anchored proteins in living cell membranes. Cell 116(4):577-589.

Sibley LD. 1995. Invasion of vertebrate cells by Toxoplasma gondii. Trends in Cell Biology 5(3):129-132.

Stringer SE, Mayer-Proschel M, Kalyani A, Rao M, Gallagher JT. 1999. Heparin Is a Unique Marker of Progenitors in the Glial Cell Lineage. Journal of Biological Chemistry 274(36):25455-25460.

Svergun DI. 1992. Determination of the regularization Parameter in Indirect-Transform Methods using perceptual criteria. Journal of Applied Crystallography 25:495-503.

Svergun DI, Petoukhov MV, Koch MH. 2001. Determination of domain structure of proteins from X-ray solution scattering. Biophysical Journal 80(6):2946-53.

Templeton TJ. 2007. Whole-genome natural histories of apicomplexan surface proteins. Trends in Parasitology 23(5):205-212.

Tenter AM, Heckeroth AR, Weiss LM. 2000. Toxoplasma gondii: from animals to humans. International Journal for Parasitology 30(12-13):1217-1258.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22):4673-80.

Vagin A, Teplyakov AJ. 1997. MOLREP: an automated program for molecular replacement, Appl. Crystallog.

Vanhamme L, Pays E, McCulloch R, Barry J. 2001. An update on antigenic variation in African trypanosomes. Trends in Parasitology 17(7):338-343.

Velge-Roussel F, Chardes T, Mevelec P, Brillard M, Hoebeke J, Bout D. 1994. Epitopic analysis of the Toxoplasma gondii major surface antigen SAG 1. Molecular and biochemical parasitology 66(1):31-38.

Velge-Roussel F, Dimier-Poisson I, Buzoni-Gatel D, Bout D. 2001. Anti-SAG1 peptide antibodies inhibit the penetration of Toxoplasma gondii tachyzoites into enterocyte cell lines. Parasitology 123(03):225-233.

WuDunn D, Spear PG. 1989. Initial interaction of herpes simplex virus with cells is binding to heparan sulfate. Journal of Virology 63(1):52-58.

Yolken RH, Torrey EF. 2008. Are some cases of psychosis caused by microbial agents? A review of the evidence. Molecular Psychiatry 13(5):470-479.

Zhou XW, Kafsack BFC, Cole RN, Beckett P, Shen RF, Carruthers VB. 2005. The opportunistic pathogen Toxoplasma gondii deploys a diverse legion of invasion and survival proteins. Journal of Biological Chemistry 280(40):34233.

Appendix

Table 4. SRS2 amino acid polymorphism.

SRS2	Amino acid number						
Type (strain)	113	261	271	295	304	338	342
Type II/III (Me49/VEG)	S	F	M	D	T	G	S
Type I (RH)	N	F	R	E	T	S	F
Type I (GT1/CAST)	N	F	R	E	T	S	S
Type X	N	F	M	D	T	G	S

Table 5. SAG1 amino acid polymorphism.

SAG1	Amino acid number									
Type (strain)	25	28	178	234	246	253	281	292	294	308
Type I (RH)	L	V	D	T	K	K	K	E	A	I
Type I (GT1)	R	V	D	T	K	K	K	K	A	I
Type II (Me49)	R	A	N	S	N	N	Q	K	S	T
Type III (VEG)	R	A	N	S	N	N	Q	K	S	T

Table 6. BSR4 amino acid polymorphism.

BSR4	Amino acid number															
Type (strain)	30/31	32/33	55/56	59/60	76/77	93/94	114/115	116/117	130/131	137/138	154/155	184/185	201/202	203/204	307/308	342-3/343-4
Type II (Me49; Prugniaud)	V	F	V	S	A	R	N	E	V	G	I	T	T	S	T	DK
Type I (RH; GT1)	V	F	V	S	A	R	N	E	V	G	I	T	T	S	T	DK
Type III (CEP)	L	L	L	A	T	S	T	G	K	D	T	S	K	L	N	YE

Table 7. **Glycomics mammalian core-H array.**

ID No.	Glycan Structure
1	Gal⇨-Sp8
2	Glc⇨-Sp8
3	Man⇨-Sp8
4	GalNAc⇨-Sp8
5	Fuc⇨-Sp8
6	Fuc⇨-Sp9
7	Rh⇨-Sp8
8	Neu5Ac⇨-Sp8
9	Neu5Ac⇨-Sp11
10	Neu5Ac⇦-Sp8
11	Gal⇦-Sp8
12	Glc⇦-Sp8
13	Man⇦-Sp8
14	GalNAc⇦-Sp8
15	GlcNAc⇦-Sp0
16	GlcNAc⇦-Sp8
17	GlcN(Gc)⇦-Sp8
18	Gal⇦1-4GlcNAc⇦1-3(Gal⇦1-4GlcNAc⇦1-6)GalNAc⇨-Sp8
19	GlcNAc⇦1-3(GlcNAc⇦1-4)(GlcNAc⇦1-6)GlcNAc-Sp8
20	[3OSO3][6OSO3]Gal⇦1-4[6OSO3]GlcNAc⇦-Sp0
21	[3OSO3][6OSO3]Gal⇦1-4GlcNAc⇦-Sp0
22	[3OSO3]Gal⇦1-4(Fuc⇨1-3)[6OSO3]Glc-Sp0
23	[3OSO3]Gal⇦1-4Glc⇦-Sp8
24	[3OSO3]Gal⇦1-4[6OSO3]Glc⇦-Sp0
25	[3OSO3]Gal⇦1-4[6OSO3]Glc⇦-Sp8
26	[3OSO3]Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦-Sp8

27	[3OSO3]Gal⇦1-3GalNAc⇨-Sp8
28	[3OSO3]Gal⇦1-3GlcNAc⇦-Sp8
29	[3OSO3]Gal⇦1-4(Fuc⇨1-3)GlcNAc-Sp0
30	[3OSO3]Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp8
31	[3OSO3]Gal⇦1-4[6OSO3]GlcNAc⇦-Sp0
32	[3OSO3]Gal⇦1-4[6OSO3]GlcNAc⇦-Sp8
33	[3OSO3]Gal⇦1-4GlcNAc⇦-Sp0
34	[3OSO3]Gal⇦1-4GlcNAc⇦-Sp8
35	[3OSO3]Gal⇦-Sp8
36	[4OSO3][6OSO3]Gal⇦1-4GlcNAc⇦-Sp0
37	[4OSO3]Gal⇦1-4GlcNAc⇦-Sp8
38	6-H2PO3Man⇨-Sp8
39	[6OSO3]Gal⇦1-4Glc⇦-Sp0
40	[6OSO3]Gal⇦1-4Glc⇦-Sp8
41	[6OSO3]Gal⇦1-4GlcNAc⇦-Sp8
42	[6OSO3]Gal⇦1-4[6OSO3]Glc⇦-Sp8
43	Neu5Ac⇨2-3[6OSO3]Gal⇦1-4GlcNAc⇦-Sp8
44	[6OSO3]GlcNAc⇦-Sp8
45	Neu5Ac(9Ac)⇨-Sp8
46	Neu5Ac(9Ac)⇨2-6Gal⇦1-4GlcNAc⇦-Sp8
47	Man⇨1-3(Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp12
48	Man⇨1-3(Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp13
49	GlcNAc⇦1-2Man⇨1-3(GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp13
50	Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp12
51	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-N(LT)AVL
52	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp8
53	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp12

54	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp13
55	Fuc⇨1-2Gal⇦1-3GalNAc⇦1-3Gal⇨-Sp9
56	Fuc⇨1-2Gal⇦1-3GalNAc⇦1-3Gal⇨1-4Gal⇦1-4Glc⇦-Sp9
57	Fuc⇨1-2Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦-Sp8
58	Fuc⇨1-2Gal⇦1-3GalNAc⇨-Sp8
59	Fuc⇨1-2Gal⇦1-3GalNAc⇨-Sp14
60	Fuc⇨1-2Gal⇦1-3GalNAc⇦1-4(Neu5Ac⇨2-3)Gal⇦1-4Glc⇦-Sp0
61	Fuc⇨1-2Gal⇦1-3GalNAc⇦1-4(Neu5Ac⇨2-3)Gal⇦1-4Glc⇦-Sp9
62	Fuc⇨1-2Gal⇦1-3GlcNAc⇦1-3Gal⇦1-4Glc⇦-Sp8
63	Fuc⇨1-2Gal⇦1-3GlcNAc⇦1-3Gal⇦1-4Glc⇦-Sp10
64	Fuc⇨1-2Gal⇦1-3GlcNAc⇦-Sp0
65	Fuc⇨1-2Gal⇦1-3GlcNAc⇦-Sp8
66	Fuc⇨1-2Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0
67	Fuc⇨1-2Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0
68	Fuc⇨1-2Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0
69	Fuc⇨1-2Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp8
70	Fuc⇨1-2Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0
71	Fuc⇨1-2Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0
72	Fuc⇨1-2Gal⇦1-4GlcNAc⇦-Sp0
73	Fuc⇨1-2Gal⇦1-4GlcNAc⇦-Sp8
74	Fuc⇨1-2Gal⇦1-4Glc⇦-Sp0
75	Fuc⇨1-2Gal⇦-Sp8
76	Fuc⇨1-3GlcNAc⇦-Sp8
77	Fuc⇨1-4GlcNAc⇦-Sp8
78	Fuc⇦1-3GlcNAc⇦-Sp8
79	GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-3GlcNAc⇦-Sp0
80	GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0
81	[3OSO3]Gal⇦1-4(Fuc⇨1-3)Glc-Sp0
82	GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦-Sp0

83	GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦-Sp8
84	GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-4Glc⇦-Sp0
85	GlcNAc⇦1-3Gal⇦1-3GalNAc⇨-Sp8
86	GalNAc⇨1-3(Fuc⇨1-2)Gal⇦-Sp8
87	GalNAc⇨1-3(Fuc⇨1-2)Gal⇦-Sp18
88	GalNAc⇨1-3GalNAc⇦-Sp8
89	GalNAc⇨1-3Gal⇦-Sp8
90	GalNAc⇨1-4(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦-Sp8
91	GalNAc⇦1-3GalNAc⇨-Sp8
92	GalNAc⇦1-3(Fuc⇨1-2)Gal⇦-Sp8
93	GalNAc⇦1-3Gal⇨1-4Gal⇦1-4GlcNAc⇦-Sp0
94	GalNAc⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0
95	GalNAc⇦1-4GlcNAc⇦-Sp0
96	GalNAc⇦1-4GlcNAc⇦-Sp8
97	Gal⇨1-2Gal⇦-Sp8
98	Gal⇨1-3(Fuc⇨1-2)Gal⇦1-3GlcNAc⇦-Sp0
99	Gal⇨1-3(Fuc⇨1-2)Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0
100	Gal⇨1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc-Sp0
101	Gal⇨1-3(Fuc⇨1-2)Gal⇦1-4Glc⇦-Sp0
102	Gal⇨1-3(Fuc⇨1-2)Gal⇦-Sp8
103	Gal⇨1-3(Gal⇨1-4)Gal⇦1-4GlcNAc⇦-Sp8
104	Gal⇨1-3GalNAc⇨-Sp8
105	Gal⇨1-3GalNAc⇨-Sp16
106	Gal⇨1-3GalNAc⇦-Sp8
107	Gal⇨1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp8
108	Gal⇨1-3Gal⇦1-3GlcNAc⇦-Sp0
109	Gal⇨1-3Gal⇦1-4GlcNAc⇦-Sp8
110	Gal⇨1-3Gal⇦1-4Glc⇦-Sp0
111	Gal⇨1-3Gal⇦-Sp8
112	Gal⇨1-4(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦-Sp8

113	Gal⇨1-4Gal⇦1-4GlcNAc⇦-Sp0
114	Gal⇨1-4Gal⇦1-4GlcNAc⇦-Sp8
115	Gal⇨1-4Gal⇦1-4Glc⇦-Sp0
116	Gal⇨1-4GlcNAc⇦-Sp8
117	Gal⇨1-6Glc⇦-Sp8
118	Gal⇦1-2Gal⇦-Sp8
119	Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0
120	Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0
121	Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦-Sp0
122	Gal⇦1-3(Fuc⇨1-4)GlcNAc-Sp8
123	Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦-Sp8
124	Gal⇦1-4GlcNAc⇦1-6GalNAc⇨-Sp8
125	Gal⇦1-3(GlcNAc⇦1-6)GalNAc⇨-Sp8
126	Gal⇦1-3(GlcNAc⇦1-6)GalNAc-Sp14
127	Gal⇦1-3(Neu5Ac⇨2-6)GalNAc⇨-Sp8
128	Gal⇦1-3(Neu5Ac⇨2-6)GalNAc⇨-Sp14
129	Gal⇦1-3(Neu5Ac⇦2-6)GalNAc⇨-Sp8
130	Gal⇦1-3(Neu5Ac⇨2-6)GlcNAc⇦1-4Gal⇦1-4Glc⇦-Sp10
131	Gal⇦1-3GalNAc⇨-Sp8
132	Gal⇦1-3GalNAc⇨-Sp16
133	Gal⇦1-3GalNAc⇦-Sp8
134	Gal⇦1-3GalNAc⇦1-3Gal⇨1-4Gal⇦1-4Glc⇦-Sp0
135	Gal⇦1-3GalNAc⇦1-4(Neu5Ac⇨2-3)Gal⇦1-4Glc⇦-Sp0
136	Gal⇦1-3GalNAc⇦1-4Gal⇦1-4Glc⇦-Sp8
137	Gal⇦1-3Gal⇦-Sp8
138	Gal⇦1-3GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0
139	Gal⇦1-3GlcNAc⇦1-3Gal⇦1-4Glc⇦-Sp10
140	Gal⇦1-3GlcNAc⇦-Sp0
141	Gal⇦1-3GlcNAc⇦-Sp8
142	Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0

143	Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp8
144	Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-4Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0
145	Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-4Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-4Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0
146	Gal⇦1-4[6OSO3]Glc⇦-Sp0
147	Gal⇦1-4[6OSO3]Glc⇦-Sp8
148	Gal⇦1-4GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦-Sp8
149	Gal⇦1-4GalNAc⇦1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦-Sp8
150	Gal⇦1-4GlcNAc⇦1-3GalNAc⇨-Sp8
151	Gal⇦1-4GlcNAc⇦1-3GalNAc-Sp14
152	Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0
153	Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0
154	Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0
155	Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4Glc⇦-Sp0
156	Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4Glc⇦-Sp8
157	Gal⇦1-4GlcNAc⇦1-6(Gal⇦1-3)GalNAc⇨-Sp8
158	Gal⇦1-3(Gal⇦1-4GlcNAc⇦1-6)GalNAc⇨-Sp8
159	Gal⇦1-3(Gal⇦1-4GlcNAc⇦1-6)GalNAc-Sp14
160	Gal⇦1-4GlcNAc⇦-Sp0
161	Gal⇦1-4GlcNAc⇦-Sp8
162	Gal⇦1-4Glc⇦-Sp0
163	Gal⇦1-4Glc⇦-Sp8
164	GlcNAc⇨1-3Gal⇦1-4GlcNAc⇦-Sp8
165	GlcNAc⇨1-6Gal⇦1-4GlcNAc⇦-Sp8
166	GlcNAc⇦1-2Gal⇦1-3GalNAc⇨-Sp8
167	GlcNAc⇦1-3(GlcNAc⇦1-6)GalNAc⇨-Sp8
168	GlcNAc⇦1-3(GlcNAc⇦1-6)Gal⇦1-4GlcNAc⇦-Sp8
169	GlcNAc⇦1-3GalNAc⇨-Sp8
170	GlcNAc⇦1-3GalNAc⇨-Sp14
171	GlcNAc⇦1-3Gal⇦-Sp8

172	GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0
173	GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp8
174	GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0
175	GlcNAc⇦1-3Gal⇦1-4Glc⇦-Sp0
176	GlcNAc⇦1-4-MDPLys
177	GlcNAc⇦1-4(GlcNAc⇦1-6)GalNAc⇨-Sp8
178	GlcNAc⇦1-4Gal⇦1-4GlcNAc⇦-Sp8
179	$(GlcNAc⇦1-4)_6$⇦-Sp8
180	$(GlcNAc⇦1-4)_5$⇦-Sp8
181	GlcNAc⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp8
182	GlcNAc⇦1-6(Gal⇦1-3)GalNAc⇨-Sp8
183	GlcNAc⇦1-6GalNAc⇨-Sp8
184	GlcNAc⇦1-6GalNAc⇨-Sp14
185	GlcNAc⇦1-6Gal⇦1-4GlcNAc⇦-Sp8
186	Glc⇨1-4Glc⇦-Sp8
187	Glc⇨1-4Glc⇨-Sp8
188	Glc⇨1-6Glc⇨1-6Glc⇦-Sp8
189	Glc⇦1-4Glc⇦-Sp8
190	Glc⇦1-6Glc⇦-Sp8
191	G-ol-Sp8
192	GlcA⇨-Sp8
193	GlcA⇦-Sp8
194	GlcA⇦1-3Gal⇦-Sp8
195	GlcA⇦1-6Gal⇦-Sp8
196	KDN⇨2-3Gal⇦1-3GlcNAc⇦-Sp0
197	KDN⇨2-3Gal⇦1-4GlcNAc⇦-Sp0
198	Man⇨1-2Man⇨1-2Man⇨1-3Man⇨-Sp9
199	Man⇨1-2Man⇨1-3(Man⇨1-2Man⇨1-6)Man⇨-Sp9
200	Man⇨1-2Man⇨1-3Man⇨-Sp9
201	Man⇨1-6[Man⇨1-2Man⇨1-3]Man⇨1-6[Man⇨1-2Man⇨1-3]Man⇦1-4GlcNAc⇦1-

	4GlcNAc⇦-Sp12
202	Man⇨1-2Man⇨1-6(Man⇨1-3)Man⇨1-6[Man⇨1-2Man⇨1-2Man⇨1-3]Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp12
203	Man⇨1-2Man⇨1-2Man⇨1-3(Man⇨1-2Man⇨1-3(Man⇨1-2Man⇨1-6)Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp12
204	Man⇨1-3(Man⇨1-6)Man⇨-Sp9
205	Man⇨1-3(Man⇨1-2Man⇨1-2Man⇨1-6)Man⇨-Sp9
206	Man⇨1-6(Man⇨1-3)Man⇨1-6[Man⇨1-2Man⇨1-3]Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp12
207	Man⇨1-6(Man⇨1-3)Man⇨1-6(Man⇨1-3)Man⇦1-4GlcNAc⇦1-4 GlcNAc⇦-Sp12
208	Man⇦1-4GlcNAc⇦-Sp0
209	Neu5Ac⇨2-3Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc-Sp0
210	[3OSO3]Gal⇦1-4(Fuc⇨1-3)[6OSO3]GlcNAc-Sp8
211	Fuc⇨1-2[6OSO3]Gal⇦1-4GlcNAc-Sp0
212	Fuc⇨1-2Gal⇦1-4[6OSO3]GlcNAc-Sp8
213	Fuc⇨1-2[6OSO3]Gal⇦1-4[6OSO3]Glc-Sp0
214	Neu5Ac⇨2-3Gal⇦1-3GalNAc⇨-Sp8
215	Neu5Ac⇨2-8Neu5Ac⇨2-8Neu5Ac⇨2-8Neu5Ac⇨2-3(GalNAc⇦1-4)Gal⇦1-4Glc⇦-Sp0
216	Neu5Ac⇨2-8Neu5Ac⇨2-8Neu5Ac⇨2-3(GalNAc⇦1-4)Gal⇦1-4Glc⇦-Sp0
217	Neu5Ac⇨2-8Neu5Ac⇨2-8Neu5Ac⇨2-3Gal⇦1-4Glc⇦-Sp0
218	Neu5Ac⇨2-8Neu5Ac⇨2-3(GalNAc⇦1-4)Gal⇦1-4Glc⇦-Sp0
219	Neu5Ac⇨2-8Neu5Ac⇨2-8Neu5Ac⇨-Sp8
220	Neu5Ac⇨2-3(6-O-Su)Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp8
221	Neu5Ac⇨2-3(GalNAc⇦1-4)Gal⇦1-4GlcNAc⇦-Sp0
222	Neu5Ac⇨2-3(GalNAc⇦1-4)Gal⇦1-4GlcNAc⇦-Sp8
223	Neu5Ac⇨2-3(GalNAc⇦1-4)Gal⇦1-4Glc⇦-Sp0
224	Neu5Ac⇨2-3(Neu5Ac⇨2-3Gal⇦1-3GalNAc⇦1-4)Gal⇦1-4Glc⇦-Sp0
225	Neu5Ac⇨2-3(Neu5Ac⇨2-6)GalNAc⇨-Sp8
226	Neu5Ac⇨2-3GalNAc⇨-Sp8
227	Neu5Ac⇨2-3GalNAc⇦1-4GlcNAc⇦-Sp0
228	Neu5Ac⇨2-3Gal⇦1-3[6OSO3]GlcNAc-Sp8

229	Neu5Ac⇨2-3Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦-Sp8
230	Neu5Ac⇨2-3Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0
231	Neu5Ac⇨2-3Gal⇦1-3(Neu5Ac⇨2-3Gal⇦1-4)GlcNAc⇦-Sp8
232	Neu5Ac⇨2-3Gal⇦1-3[6OSO3]GalNAc⇨-Sp8
233	Neu5Ac⇨2-3Gal⇦1-3(Neu5Ac⇨2-6)GalNAc⇨-Sp8
234	Neu5Ac⇨2-3Gal⇦-Sp8
235	Neu5Ac⇨2-3Gal⇦1-3GalNAc⇦1-3Gal⇨1-4Gal⇦1-4Glc⇦-Sp0
236	Neu5Ac⇨2-3Gal⇦1-3GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0
237	Fuc⇨1-2[6OSO3]Gal⇦1-4Glc-Sp0
238	Neu5Ac⇨2-3Gal⇦1-3GlcNAc⇦-Sp0
239	Neu5Ac⇨2-3Gal⇦1-3GlcNAc⇦-Sp8
240	Neu5Ac⇨2-3Gal⇦1-4[6OSO3]GlcNAc⇦-Sp8
241	Neu5Ac⇨2-3Gal⇦1-4(Fuc⇨1-3)[6OSO3]GlcNAc⇦-Sp8
242	Neu5Ac⇨2-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0
243	Neu5Ac⇨2-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0
244	Neu5Ac⇨2-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp8
245	Neu5Ac⇨2-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-3Gal⇦-Sp8
246	Neu5Ac⇨2-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp8
247	Neu5Ac⇨2-3Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0
248	Neu5Ac⇨2-3Gal⇦1-4GlcNAc⇦-Sp0
249	Neu5Ac⇨2-3Gal⇦1-4GlcNAc⇦-Sp8
250	Neu5Ac⇨2-3Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0
251	Fuc⇨1-2Gal⇦1-4[6OSO3]Glc-Sp0
252	Neu5Ac⇨2-3Gal⇦1-4Glc⇦-Sp0
253	Neu5Ac⇨2-3Gal⇦1-4Glc⇦-Sp8
254	Neu5Ac⇨2-6GalNAc⇨-Sp8
255	Neu5Ac⇨2-6GalNAc⇦1-4GlcNAc⇦-Sp0
256	Neu5Ac⇨2-6Gal⇦1-4[6OSO3]GlcNAc⇦-Sp8
257	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦-Sp0

258	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦-Sp8
259	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0
260	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0
261	Neu5Ac⇨2-6Gal⇦1-4Glc⇦-Sp0
262	Neu5Ac⇨2-6Gal⇦1-4Glc⇦-Sp8
263	Neu5Ac⇨2-6Gal⇦-Sp8
264	Neu5Ac⇨2-8Neu5Ac⇨-Sp8
265	Neu5Ac⇨2-8Neu5Ac⇨2-3Gal⇦1-4Glc⇦-Sp0
266	Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦1-3Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦-Sp0
267	Neu5Ac⇦2-6GalNAc⇨-Sp8
268	Neu5Ac⇦2-6Gal⇦1-4GlcNAc⇦-Sp8
269	Neu5Gca2-3Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦-Sp0
270	Neu5Gca2-3Gal⇦1-3GlcNAc⇦-Sp0
271	Neu5Gca2-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0
272	Neu5Gc⇨2-3Gal⇦1-4GlcNAc⇦-Sp0
273	Neu5Gc⇨2-3Gal⇦1-4Glc⇦-Sp0
274	Neu5Gc⇨2-6GalNAc⇨-Sp0
275	Neu5Gc⇨2-6Gal⇦1-4GlcNAc⇦-Sp0
276	Neu5Gc⇨-Sp8
277	Gal⇦1-3(Neu5Ac⇨2-3Gal⇦1-4GlcNAc⇦1-6)GalNAc⇨-Sp14
278	Gal⇦1-3GalNAc⇨-Sp14
279	Gal⇦1-3GlcNAc⇦1-3Gal⇦1-3GlcNAc⇦-Sp0
280	Gal⇦1-4(Fuc⇨1-3)[6OSO3]GlcNAc-Sp0
281	Gal⇦1-4(Fuc⇨1-3)[6OSO3]Glc-Sp0
282	Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-3Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦-Sp0
283	Gal⇦1-4GlcNAc⇦1-3Gal⇦1-3GlcNAc⇦-Sp0
284	Neu5Ac⇨2-3Gal⇦1-3GlcNAc⇦1-3Gal⇦1-3GlcNAc⇦-Sp0
285	Neu5Ac⇨2-3Gal⇦1-4GlcNAc⇦1-3Gal⇦1-3GlcNAc⇦-Sp0
286	[3OSO3][4OSO3]Gal⇦1-4GlcNAc⇦-Sp0

287	[6OSO3]Gal⇦1-4[6OSO3]GlcNAc⇦-Sp0
288	6-H2PO3Glc⇦-Sp10
289	Gal⇨1-3(Fuc⇨1-2)Gal⇦-Sp18
290	Gal⇦1-3(Neu5Ac⇨2-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-6)GalNAc⇨-Sp14
291	Gal⇦1-3Gal⇦1-4GlcNAc⇦-Sp8
292	Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp12
293	Gal⇦1-4GlcNAc⇦1-3(Gal⇦1-4GlcNAc⇦1-6)Gal⇦1-4GlcNAc-Sp0
294	Gal⇦1-4GlcNAc⇦1-3(GlcNAc⇦1-6)Gal⇦1-4GlcNAc-Sp0
295	Gal⇦1-4GlcNAc⇨1-6Gal⇦1-4GlcNAc⇦-Sp0
296	Gal⇦1-4GlcNAc⇦1-6Gal⇦1-4GlcNAc⇦-Sp0
297	GalNAc⇨-Sp15
298	GalNAc⇦1-3Gal⇦-Sp8
299	GlcA⇦1-3GlcNAc⇦-Sp8
300	GlcNAc⇦1-2Man⇨1-3(Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp12
301	GlcNAc⇦1-2Man⇨1-3(GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp12
302	GlcNAc⇦1-3Man-Sp10
303	GlcNAc⇦1-4GlcNAc⇦-Sp10
304	GlcNAc⇦1-4GlcNAc⇦-Sp12
305	HOOC(CH3)CH-3-O-GlcNAc⇦1-4GlcNAc⇦-Sp10
306	Man⇨1-6Man⇦-Sp10
307	Man⇨1-6(Man⇨1-3)Man⇨1-6(Man⇨1-3)Man⇦-Sp10
308	Man⇨1-2Man⇨1-2Man⇨1-3(Man⇨1-2Man⇨1-6(Man⇨1-3)Man⇨1-6)Man⇨-Sp9
309	Man⇨1-2Man⇨1-2Man⇨1-3(Man⇨1-2Man⇨1-6(Man⇨1-2Man⇨1-3)Man⇨1-6)Man⇨-Sp9
310	Neu5Ac⇨2-3Gal⇦1-3(Neu5Ac⇨2-3Gal⇦1-4GlcNAc⇦1-6)GalNAc⇨-Sp14
311	Neu5Ac⇨2-3Gal⇦1-3(Neu5Ac⇨2-6)GalNAc⇨-Sp14
312	Neu5Ac⇨2-3Gal⇦1-3GalNAc⇨-Sp14
313	Neu5Ac⇨2-3Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp12

314	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp12
315	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp12
316	Neu5Ac⇨2-8Neu5Ac⇦-Sp17
317	Neu5Ac⇨2-8Neu5Ac⇨2-8Neu5Ac⇦-Sp8
318	Neu5Gc⇦2-6Gal⇦1-4GlcNAc-Sp8
319	Gal⇦1-3GlcNAc⇦1-2Man⇨1-3(Gal⇦1-3GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp19
320	Neu5Ac⇨2-3Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Neu5Ac⇨2-3Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp12
321	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Neu5Ac⇨2-3Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp12
322	Fuc⇨1-3(Gal⇦1-4)GlcNAc⇦1-2Man⇨1-3(Fuc⇨1-3(Gal⇦1-4)GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp20
323	Neu5Ac(9Ac)a2-3Gal⇦1-4GlcNAc⇦-Sp0
324	Neu5Ac(9Ac)a2-3Gal⇦1-3GlcNAc⇦-Sp0
325	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-3Gal⇦1-3GlcNAc⇦-Sp0
326	Neu5Ac⇨2-3Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦1-3Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦-Sp0
327	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0
328	Gal⇨1-4Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4Glc⇦-Sp0
329	GalNAc⇦1-3Gal⇨1-4Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4Glc⇦-Sp0
330	GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0
331	GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0
332	(Neu5Ac⇨2-3-Gal⇦1-3)(((Neu5Ac⇨2-3-Gal⇦1-4(Fuc⇨1-3))GlcNAc⇦1-6)GalNAc-Sp14
333	GlcNAc⇨1-4Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0
334	GlcNAc⇨1-4Gal⇦1-4GlcNAc⇦-Sp0
335	GlcNAc⇨1-4Gal⇦1-3GlcNAc⇦-Sp0
336	GlcNAc⇨1-4Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4Glc⇦-Sp0
337	GlcNAc⇨1-4Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦-Sp0
338	GlcNAc⇨1-4Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦-Sp0

339	GlcNAc⇨1-4Gal⇦1-3GalNAc-Sp14
340	Man⇨1-3(Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp12
341	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp12
342	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-6Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp12
343	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-2Man⇨1-3Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp12
344	Gal⇦1-4GlcNAc⇦1-2Man⇨1-3Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp12
345	Gal⇦1-4GlcNAc⇦1-2Man⇨1-6Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp12
346	Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp12
347	GlcNAc⇦1-2Man⇨1-3(GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4(Fuc⇨1-6)GlcNAc⇦-Sp22
348	Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4(Fuc⇨1-6)GlcNAc⇦-Sp22
349	Gal⇦1-3GlcNAc⇦1-2Man⇨1-3(Gal⇦1-3GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4(Fuc⇨1-6)GlcNAc⇦-Sp22
350	Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦1-2Man⇨1-3(Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp19
351	[6OSO3]GlcNAc⇦1-3Gal ⇦1-4GlcNAc-⇦-Sp0
352	KDN⇨2-3Gal⇦1-4(Fuc⇨1-3)GlcNAc-Sp0
353	KDN⇨2-6Gal⇦1-4GlcNAc-Sp0
354	KDN⇨2-3Gal⇦1-4Glc-Sp0
355	KDN⇨2-3Gal⇦1-3GalNAc⇨-Sp14
356	Fuc⇨1-2Gal⇦1-3GlcNAc⇦1-2Man⇨1-3(Fuc⇨1-2Gal⇦1-3GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp20
357	Fuc⇨1-2Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Fuc⇨1-2Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp20
358	Fuc⇨1-2Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-2Man⇨1-3(Fuc⇨1-2Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNA⇦-Sp20
359	Gal⇨1-3Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Gal⇨1-3Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp20
360	Man⇨1-3(Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp12
361	Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦1-2Man⇨1-3(Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4(Fuc⇨1-6)GlcNAc⇦-Sp22
362	Neu5Ac⇨2-6GlcNAc⇦1-4GlcNAc-Sp21

363	Neu5Ac⇨2-6GlcNAc⇦1-4GlcNAc⇦1-4GlcNAc-Sp21
364	Fuc⇨1-2Gal⇦1-3GlcNAc⇦1-3(Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-6)Gal⇦1-4Glc-Sp21
365	Gal⇦1-4GlcNAc⇦1-2(Gal⇦1-4GlcNAc⇦1-4)Man⇨1-3(Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp21
366	GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp20
367	Gal⇨1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Gal⇨1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp20
368	Gal⇨1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-2Man⇨1-3(Gal⇨1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp20
369	GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-3GlcNAc⇦1-2Man⇨1-3(GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-3GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp20
370	Gal⇨1-3(Fuc⇨1-2)Gal⇦1-3GlcNAc⇦1-2Man⇨1-3(Gal⇨1-3(Fuc⇨1-2)Gal⇦1-3GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp20
371	Fuc⇨1-2Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦1-2Man⇨1-3(Fuc⇨1-2Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp19
372	Neu5Ac⇨2-3Gal⇦1-4GlcNAc⇦1-3GalNAc-Sp14
373	Neu5Ac⇨2-6Gal⇦1-4GlcNAc⇦1-3GalNAc-Sp14
374	Neu5Ac⇨2-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-3GalNAc⇨-Sp14
375	(GalNAc⇦1-4GlcNAc⇦1-2Man⇨1-6)GalNAc⇦1-4GlcNAc⇦1-2Man⇨1-3Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp12
376	Gal⇦1-3GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-4Glc-Sp0
377	Gal⇦1-3GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc-Sp0
378	Gal⇦1-3GlcNAc⇦1-3(Gal⇦1-3GlcNAc⇦1-3Gal⇦1-4GlcNAc⇦1-6)Gal⇦1-4Glc⇦-Sp0
379	Gal⇦1-3GlcNAc⇦1-3(Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-6)Gal⇦1-4Glc-Sp21
380	Fuc⇨1-2Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦1-3(Gal⇦1-4GlcNAc⇦1-6)Gal⇦1-4Glc-Sp21
381	Fuc⇨1-2Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦1-3(Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-6)Gal⇦1-4Glc-Sp21
382	Gal⇦1-3GlcNAc⇦1-3(Gal⇦1-3GlcNAc⇦1-3Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-6)Gal⇦1-4Glc-Sp21
383	Gal⇦1-4GlcNAc⇦1-2(Gal⇦1-4GlcNAc⇦1-4)Man⇨1-3(Gal⇦1-4GlcNAc⇦1-2(Gal⇦1-4GlcNAc⇦1-6)Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp21
384	GlcNAc⇦1-2(GlcNAc⇦1-4)Man⇨1-3(GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp21
385	Fuc⇨1-2Gal⇦1-3GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-4Glc⇦-Sp0

386	Fuc⇨1-2Gal⇦1-3GalNAca1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦-Sp0
387	Gal⇦1-3GlcNAc⇦1-3GalNAc⇨-Sp14
388	Neu5Ac⇨2-3(GalNAc⇦1-4)Gal⇦1-4GlcNAc⇦1-3GalNAc⇨-Sp14
389	GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-3GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦-Sp0
390	Gal⇨1-3Gal⇦1-3GlcNAc⇦1-2Man⇨1-3(Gal⇨1-3Gal⇦1-3GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp19
391	Gal⇨1-3Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦1-2Man⇨1-3(Gal⇨1-3Gal⇦1-3(Fuc⇨1-4)GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp19
392	Neu5Ac⇨2-3Gal⇦1-3GlcNAc⇦1-2Man⇨1-3(Neu5Ac⇨2-3Gal⇦1-3GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp19
393	Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp12
394	GlcNAc⇦1-2Man⇨1-3(Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp12
395	Neu5Ac⇨2-3Gal⇦1-3GlcNAc⇦1-3GalNAc⇨-Sp14
396	Fuc⇨1-2Gal⇦1-4GlcNAc⇦1-3GalNAc⇨-Sp14
397	Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-3GalNAc⇨-Sp14
398	GalNAc⇨1-3GalNAc⇦1-3Gal⇨1-4Gal⇦1-4GlcNAc⇦-Sp0
399	Gal⇨1-4Gal⇦1-3GlcNAc⇦1-2Man⇨1-3(Gal⇨1-4Gal⇦1-3GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp19
400	Gal⇨1-4Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Gal⇨1-4Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-LVANKT
401	Gal⇨1-3Gal⇦1-4GlcNAc⇦1-3GalNAc⇨-Sp14
402	Gal⇦1-3GlcNAc⇦1-6Gal⇦1-4GlcNAc⇦-Sp0
403	Gal⇦1-3GlcNAc⇨1-6Gal⇦1-4GlcNAc⇦-Sp0
404	GalNAc⇦1-3Gal⇨1-6Gal⇦1-4Glc⇦-Sp8
405	GlcNAc⇦1-6(GlcNAc⇦1-3)GalNAc⇨-Sp14
406	Gal⇨1-3(Fuc⇨1-2)Gal⇦1-4(Fuc⇨1-3)Glc⇦-Sp21
407	Neu5Ac⇨2-6Gal⇦1-3GlcNAc⇦1-3(Gal⇦1-4GlcNAc⇦1-6)Gal⇦1-4Glc-Sp21
408	Gal⇦1-3GalNAc⇦1-4(Neu5Ac⇨2-8Neu5Ac⇨2-3)Gal⇦1-4Glc⇦-Sp0
409	Neu5Ac⇨2-3Gal⇦1-3GalNAc⇦1-4(Neu5Ac⇨2-8Neu5Ac⇨2-3)Gal⇦1-4Glc⇦-Sp0
410	Gal⇨1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦1-3GalNAc⇨-Sp14
411	GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦1-3GalNAc⇨-Sp14

412	GalNAc⇨1-3GalNAc⇦1-3Gal⇨1-4Gal⇦1-4Glc⇦-Sp0
413	Fuc⇨1-2Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-3GalNAc⇨-Sp14
414	Gal⇨1-3(Fuc⇨1-2)Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-3GalNAc-Sp14
415	GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-3GalNAc-Sp14
416	Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-2Man⇨1-3(Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4(Fuc⇨1-6)GlcNAc⇦-Sp22
417	Fuc⇨1-2Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Fuc⇨1-2Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4(Fuc⇨1-6)GlcNAc⇦-Sp22
418	GlcNAc⇦1-2Man⇨1-3(GlcNAc⇦1-2(GlcNAc⇦1-6)Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp19
419	Fuc⇨1-2Gal⇦1-3GlcNAc⇦1-3GalNAc-Sp14
420	Gal⇨1-3(Fuc⇨1-2)Gal⇦1-3GlcNAc⇦1-3GalNAc-Sp14
421	GalNAc⇨1-3(Fuc⇨1-2)Gal⇦1-3GlcNAc⇦1-3GalNAc-Sp14
422	Gal⇨1-3Gal⇦1-3GlcNAc⇦1-3GalNAc-Sp14
423	Fuc⇨1-2Gal⇦1-3GlcNAc⇦1-2Man⇨1-3(Fuc⇨1-2Gal⇦1-3GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4(Fuc⇨1-6)GlcNAc⇦-Sp22
424	Gal⇨1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(Gal⇨1-3(Fuc⇨1-2)Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4(Fuc⇨1-6)GlcNAc⇦-Sp22
425	Gal⇦1-3GlcNAc⇦1-2Man⇨1-3(Gal⇦1-3GlcNAc⇦1-2(Gal⇦1-3GlcNAc⇦1-6)Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-Sp19
426	Fuc⇨1-2Gal⇦1-3GlcNAc⇦1-3(Gal⇦1-4GlcNAc⇦1-6)Gal⇦1-4Glc-Sp21
427	Gal⇦1-4GlcNAc⇦1-3Gal⇦1-4(Fuc⇨1-3GlcNAc⇦1-6)Gal⇦1-4Glc-Sp21
428	GlcNAc⇦1-2Man⇨1-3(GlcNAc⇦1-4)(GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNac⇦1-4GlcNAc-Sp21
429	GlcNAc⇦1-4(GlcNAc⇦1-2)Man⇨1-3(GlcNAc⇦1-4)(GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp21
430	GlcNAc⇦1-2Man⇨1-3(GlcNAc⇦1-4)(GlcNAc⇦1-6(GlcNAc⇦1-2)Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp21
431	GlcNAc⇦1-4(GlcNAc⇦1-2)Man⇨1-3(GlcNAc⇦1-4)(GlcNAc⇦1-6(GlcNAc⇦1-2)Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp21
432	Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(GlcNAc⇦1-4)(Gal⇦1-4GlcNAc⇦1-2)Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp21
433	Gal⇦1-4GlcNAc⇦1-4(Gal⇦1-4GlcNAc⇦1-2)Man⇨1-3(GlcNAc⇦1-4)(Gal⇦1-4GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp21
434	Gal⇦1-4GlcNAc⇦1-2Man⇨1-3(GlcNAc⇦1-4)(Gal⇦1-4GlcNAc⇦1-6(Gal⇦1-4GlcNAc⇦1-2)Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc-Sp21

435	Gal⇦1-4Gal⇦-Sp10
436	Gal⇦1-6Gal⇦-Sp10
437	Neu5Ac⇨2-3Gal⇦1-4GlcNAc⇦1-3Gal⇦-Sp8
438	GalNAc⇦1-6GalNAc⇦-Sp8
439	[6OSO3]Gal⇦1-3GlcNAc⇦-Sp0
440	[6OSO3]Gal⇦1-3[6OSO3]GlcNAc-Sp0
441	Fuc⇨1-2Gal⇦1-4GlcNAc⇦1-2(Fuc⇨1-2Gal⇦1-4GlcNAc⇦1-4)Man⇨1-3(Fuc⇨1-2Gal⇦1-4 GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-N
442	Fuc⇨1-2Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-2(Fuc⇨1-2Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-4)Man⇨1-3(Fuc⇨1-2Gal⇦1-4(Fuc⇨1-3)GlcNAc⇦1-2Man⇨1-6)Man⇦1-4GlcNAc⇦1-4GlcNAc⇦-N