THE E3-UBIQUITIN LIGASE ADAPTOR PROTEIN SKP1 IS GLYCOSYLATED BY AN EVOLUTIONARILY CONSERVED PATHWAY THAT REGULATES PROTIST GROWTH AND DEVELOPMENT

by

KAZI MD. MOSTAFIZUR RAHMAN

(Under the Direction of Christopher M. West)

ABSTRACT

Toxoplasma gondii (Tg) is a protist parasite of warm-blooded animals. *Toxoplasma* depends on a prolyl 4-hydroxylase (PhyA), which hydroxylates Skp1, a subunit of Skp1/Cullin-1/F-box-protein E3 ubiquitin ligases, for optimal proliferation in fibroblasts. In another protist, *Dictyostelium*, Skp1 hydroxyproline is uniquely modified by five sugars via the action of three glycosyltransferases, Gnt1, PgtA and AgtA, which are required for O_2-dependent development. Here we show that TgSkp1 hydroxyproline is also modified by a pentasaccharide, based on mass spectrometry, and that the core trisaccharide depends on homologs of Gnt1 and PgtA. Reconstitution of the glycosyltransferase reactions in extracts with radioactive sugar nucleotides and appropriate Skp1 acceptors confirmed the predicted sugar identities as GlcNAc, Gal and Fuc. Disruption of *gnt1* or *pgtA* resulted in decreased parasite growth which was rescued by genetic complementation. However, AgtA, the α-galactosyltransferase that catalyzes addition of the outer sugars in *Dictyostelium*, is absent in *Toxoplasma*. We predicted two cytoplasmic, phylogenetically co-distributed, retaining glycosyltransferases: Glt1 and Gat1. Their roles in Skp1 glycosylation was indicated by the appearance of expected truncated glycans in *glt1*Δ- and *gat1*Δ-parasites.

Enzymatic assays of cell extracts suggested that Glt1 is a novel Skp1 glucosyltransferase, which was confirmed using purified recombinant Glt1. Treatment with glycosidases indicated that the terminal sugar is α-galactose. Thus the TgSkp1 glycan differs at least by substitution of the fourth sugar, α-galactose, with α-glucose. Plaque assays revealed slow growth intermediate between that of parental and *phyA*Δ cells, suggesting that the role of PhyA depends in part on full glycosylation. Complementation of *glt1*Δ with catalytically active or inactive Glt1 confirmed the involvement of its enzyme activity for optimal growth. The acceptor trisaccharide was a much better substrate when conjugated to Skp1 rather than para-nitrophenol, suggesting that Skp1 is a preferred substrate in cells. A phylogenetic survey indicates that the *Toxoplasma* mechanism is ancestral, and suggests that loss of *glt1* led to its replacement by *agtA* in amoebozoans and liberation of Gat1 to prime glycogen synthesis as glycogenin in fungi and animals. Future studies will exploit the glycan differences to investigate mechanistic aspects of their function(s), and investigate a potential cellular role in nutrient storage.

INDEX WORDS: Cytoplasmic glycosylation, E3 Ubiquitin ligase, Glycosyltransferase, Glycobiology, Mass spectrometry, Parasitology, SCF-Complex, Skp1, *Toxoplasma gondii*

THE E3-UBIQUITIN LIGASE ADAPTOR PROTEIN SKP1 IS GLYCOSYLATED BY AN EVOLUTIONARILY CONSERVED PATHWAY THAT REGULATES PROTIST GROWTH AND DEVELOPMENT

by

KAZI MD. MOSTAFIZUR RAHMAN

B.S., University of Dhaka, Bangladesh, 2007

M.S., University of Dhaka, Bangladesh, 2008

M.S., Yamaguchi University, Japan, 2011

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2017

THE E3-UBIQUITIN LIGASE ADAPTOR PROTEIN SKP1 IS GLYCOSYLATED BY AN EVOLUTIONARILY CONSERVED PATHWAY THAT REGULATES PROTIST GROWTH AND DEVELOPMENT

by

Kazi Md. Mostafizur Rahman

Major Professor:	Christopher M. West
Committee:	R. Lance Wells Silvia N. J. Moreno Natarajan Kannan

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May, 2017

DEDICATION

TO MY MOTHER, BEGUM SHAMSUNNAHAR. THANK YOU FOR YOUR EVERLASTING LOVE AND INSPIRATION

ACKNOWLEDGMENTS

At this very special juncture of my scientific career, I am indebted to many people for their support and guidance. First, I would like to express my sincere gratitude to Dr. Christopher West, my major professor and my mentor. His guidance, training, constant encouragement and the countless enlightening conversations has not only helped me achieve this goal, but will also help me as a scientist in years to come. I would also like to thank my ex-mentor, Dr. Ira Blader, for giving me the opportunity to work on this project and for continuing mentoring and technical support. I would also like to thank my past committee members at the University of Oklahoma Health Sciences Center, Dr. David Dyer, Dr. Noah Butler, Dr. Scott Plafker and Dr. Jens Kreth as well as my committee members from UGA, Dr. Lance Wells, Dr. Silvia Moreno and Dr. Natarajan Kannan for their time and invaluable input into my research.

I would like to thank all the past and present members of the West and Blader labs. I would especially like to acknowledge Dr. Kevin Brown, Hanke van der Wel, and Dr. Msano Mandalasi for being good friends and excellent colleagues.

I also express my gratitude to Dr. Linda Peng Zhao and Dr. Lance Wells for analyzing and explaining all the mass spectrometry data, an essential part of this dissertation research. I am indebted to Dr. Boris Striepen and Dr. Silvia Moreno for sharing reagents and offering advice.

Last but not the least, I would like to express my undying love and gratitude to my mother and to my close family members, for a lifetime of support, encouragement, and education. I give a special thanks to my wife, Shanzida Siddique, and my lovely daughter, Manha Afsin, for without their love and support, the accomplishment of this dissertation would not have been possible.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

ABBREVIATION	MEANING
EST	Expressed sequence tags
GlcNAcT	Polypeptide N-acetyl-α-glucosaminyltransferase
GT	Glycosyl Transferase
HFF	Human Foreskin Fibroblasts
Hyp	*4R*,*2S* (*trans*)-hydroxyproline
MS	Mass Spectrometry
S100	a cytosolic extract prepared as the supernatant after 100,000 × *g* centrifugation
SCF	Skp1/Cullin-1/FBP subcomplex of the Cullin-1/RING ligase class of E3 ubiquitin ligases
SF-tag	a 51-amino acid peptide including 2 Strep-tag II epitopes and a FLAG epitope
TCA	Trichloroacetic acid

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Introduction

Toxoplasma gondii (Tg) is an obligate intracellular protozoan parasite that belongs to the phylum Apicomplexa. It is the most common parasitic infection worldwide. About one-third of the world population is infected with *Toxoplasma* although local rates vary tremendously (1). Infections are usually asymptomatic in healthy individuals as the tissue destructing tachyzoite stage is well controlled by the immune system. In response to a healthy immune response, tachyzoites convert to a slow growing bradyzoite stage that remains latent inside the tissue cysts and can evade the immune system. Diseases could be manifested and even fatal in immunocompromised patients where bradyzoites convert back to tachyzoites and start uncontrolled proliferation. Clinical manifestations are diverse and range from ocular toxoplasmosis (choroiditis, retinochoroiditis), myocarditis, and encephalitis (2). In the developing fetus, which lacks a mature immune system thus allowing uncurbed replication of tachyzoites, hydrocephalus of the newborn and mental disorders can ensue. Currently, no vaccine against *Toxoplasma* exist and drugs have severe side effects in humans and resistance is developing to these drugs (3-5). So there is an urgent need for alternative drugs or drug targets for the treatment of human toxoplasmosis. Ideal drug targets are those pathways that are present only in the parasites and not in the hosts.

SCF (Skp1/Cul1/F-box)-E3 ubiquitin ligases are the largest family of E3 ligases and play crucial roles in ubiquitin-proteasome dependent protein degradation (6). SCF-E3 ligases polyubiquitinate a large number of cellular proteins for degradation by the 26S proteasome, and control diverse cellular processes including the mammalian cell cycle and plant hormone signaling (7, 8). To ensure accurate and timely substrate degradation, SCFs themselves are regulated in a number of ways including neddylation/deneddylation of the scaffold-like protein Cul1, and their substrates are often regulated by posttranslational activation (9).

A novel form of SCF-E3 ligase regulation has been observed in the social amoeba *Dictyostelium discoidium (Dd)*, where the adaptor subunit, Skp1, undergoes prolyl hydroxylation and complex cytosolic O-glycosylation (10). It has been shown that Pro143 of Skp1 in *Dictyostelium* is hydroxylated by a cytosolic prolyl 4-hydroxylase that is subsequently modified by a pentasaccharide by five glycosyltransferase activities of three different cytosolic enzymes. Based on analysis of disruptions of these enzyme genes, Skp1 hydroxylation/glycosylation is required for *Dictyostelium* development at physiological O_2 concentrations. Emerging evidence suggests that Skp1 glycosylation controls the stability of the proteome, via affecting the conformation of Skp1 as well as the activity of SCF E3 ligases through influencing the assembly of SCF complexes (11, 12).

Biochemical studies show the evidence of active proteasome in *Toxoplasma* (13) and bioinformatics analyses have identified candidate genes for encoding major players of the ubiquitin proteasome system including the components of SCF complexes (14). Proteomic studies suggest that *Toxoplasma* encodes the capacity to ubiquitinate a sizeable portion of its proteome and many of those ubiquitinated proteins are regulated in a cell cycle dependent manner (15). However, it is understudied how the ubiquitin proteasome system mediates the regulated

proteolysis in *Toxoplasma* and the role and regulation of SCFs in this process. Interestingly, a bioinformatics analysis of the non-redundant database suggests that the *Dictyostelium*-like Skp1 modification pathway is partially conserved in *Toxoplasma* and among a select group of widely divergent protists, but is lacking in mammals (10). Initial genetic and biochemical evidence indicates that the pathway is functional in *Toxoplasma* and that at least the first prolyl hydroxylation step is important for parasite proliferation (16). The importance of this pathway in *Toxoplasma* growth but its absence in mammalian hosts raised an interest to understand the pathway in detail and to decipher its role in SCF-ligase regulation to utilize this pathway as a potential drug target against *Toxoplasma* infection. In this work, we aimed to determine the nature of Skp1 glycosylation in *Toxoplasma*, establish the glycosylation pathway by genetic and biochemical analysis as well as measure the importance of associated glycosyltransferases in parasite proliferation.

This dissertation is organized into five chapters. Chapter 1 reviews the literature describing the general background on *Toxoplasma* taxonomy, life cycle and biology, general description of SCF-E3 Ub ligases, and their regulation and role in cellular processes. Special focus and details are given to the genetic and biochemical basis of the Skp1 modification pathway in *Dictyostelium,* its potential contribution to SCF regulation as well as its significance in *Dictyostelium* O_2-dependent development, which provided the initial clue for the *Toxoplasma* studies. At the end of this chapter, our current knowledge on SCF-E3 ligases and the potential Skp1 modification pathway in *Toxoplasma* is described. Chapter 2 describes our findings on the nature of Skp1 glycosylation in *Toxoplasma* (a pentasaccharide), and the role of the glycosyltransferases in forming the core trisaccharide and in parasite growth in cell cultures. Chapter 3 and chapter 4 describe the identification, evolutionary conservation and functional characterization of two novel

glycosyltransferases that are responsible for adding the fourth and fifth sugar on *Toxoplasma* Skp1, respectively. The final chapter summarizes the significance and future directions of our research. All together, the work presented in this dissertation shed lights on the nature of Skp1 glycosylation in *Toxoplasma* and characterizes its importance in parasite proliferation.

General Background on *Toxoplasma gondii*

History and Taxonomy

Presence of *Toxoplasma gondii* was first described in the tissues of a North African rodent, *Ctenodactylus gundi*, by Nicolle and Manceaux in 1908 and independently in Brazil by Splendore (17). A fatal case of toxoplasmosis in a child was reported by the ophthalmologist Josef Janků in 1923 and sixteen years later, Wolf, Cowen and Paige were the first to conclusively identify *T. gondii* as a cause of human disease. The life cycle of the parasite was described in detail in 1970 (18). Classical taxonomy places *Toxoplasma gondii* in the phylum Apicomplexa due to the presence of an apical complex, a collection of unique organelles that include the rhoptries, the micronemes, the apical polar ring and the conoid. Important human pathogens such as *Cryptosporidium spp.* and *Plasmodium spp*. also belong to the phylum Apicomplexa (19). In addition to the apical complex, most apicomplexans have an essential chloroplast-like organelle called the apicoplast, which is important for a number of metabolic processes such as fatty acids, isoprenoid and haem synthesis (20). All apicomplexans are obligate intracellular parasites (21). *Toxoplasma* belongs to the class Conoidasida and the family Sarcocystidae along with at least three other closely related animal pathogens: *Hammondia sp., Neospora spp.* and *Sarcocystis spp.* (22, 23).

Life Cycle

Toxoplasma possesses a heterogeneous life cycle requiring the participation of different animal hosts (**Fig.1.1**). Felidae, the cats, are the only definitive hosts, and all other warm-blooded animals are intermediate hosts. There are three main life cycle stages: the tachyzoites, the bradyzoites and the sporozoites. Sexual reproduction of *T. gondii* occurs in the intestinal epithelium of felines, starting with ingestion of host of tissue cysts containing the bradyzoites. The cyst wall is digested by proteolytic enzymes in the stomach and the released bradyzoites invade the intestinal epithelium and convert to the invasive tachyzoite stage. Besides systemic dissemination, some invasive tachyzoites inside the epithelium go through asexual replication by enodyogeny, where two daughter cells are created inside a mother cell by a single round of DNA replication and mitosis followed by the assembly of two daughter cells and cytokinesis, and by schizogeny, which involves the formation of multiple merozoite cells around previously divided nuclei (24, 25). The sexual cycle starts by the formation of male and female gamonts from the merozoites, which fertilize to form macrogamont (zygote) within a thick walled oocyst just under the host cell microvilli in the small intestine, especially the ileum. Oocysts are discharged from the ruptured epithelial cells into the intestinal lumen of the cat and passed to the environment via cat feces approximately three to fifteen days after ingestion of the tissue cysts. Initially, oocysts remain unsporulated, but after one to five days their sporulation results in the formation of two sporocysts, each with four infectious sporozoites. Oocysts are very resistant to adverse environmental conditions and can preserve their infectivity for months or years. The oocysts can be ingested by intermediate hosts including humans, sporozoites invade the intestinal epithelium and convert to the tachyzoites and replicate there. After egress and crossing the lamina propria, tachyzoites rapidly travel to other parts of the body including muscles, brain and retina, where they

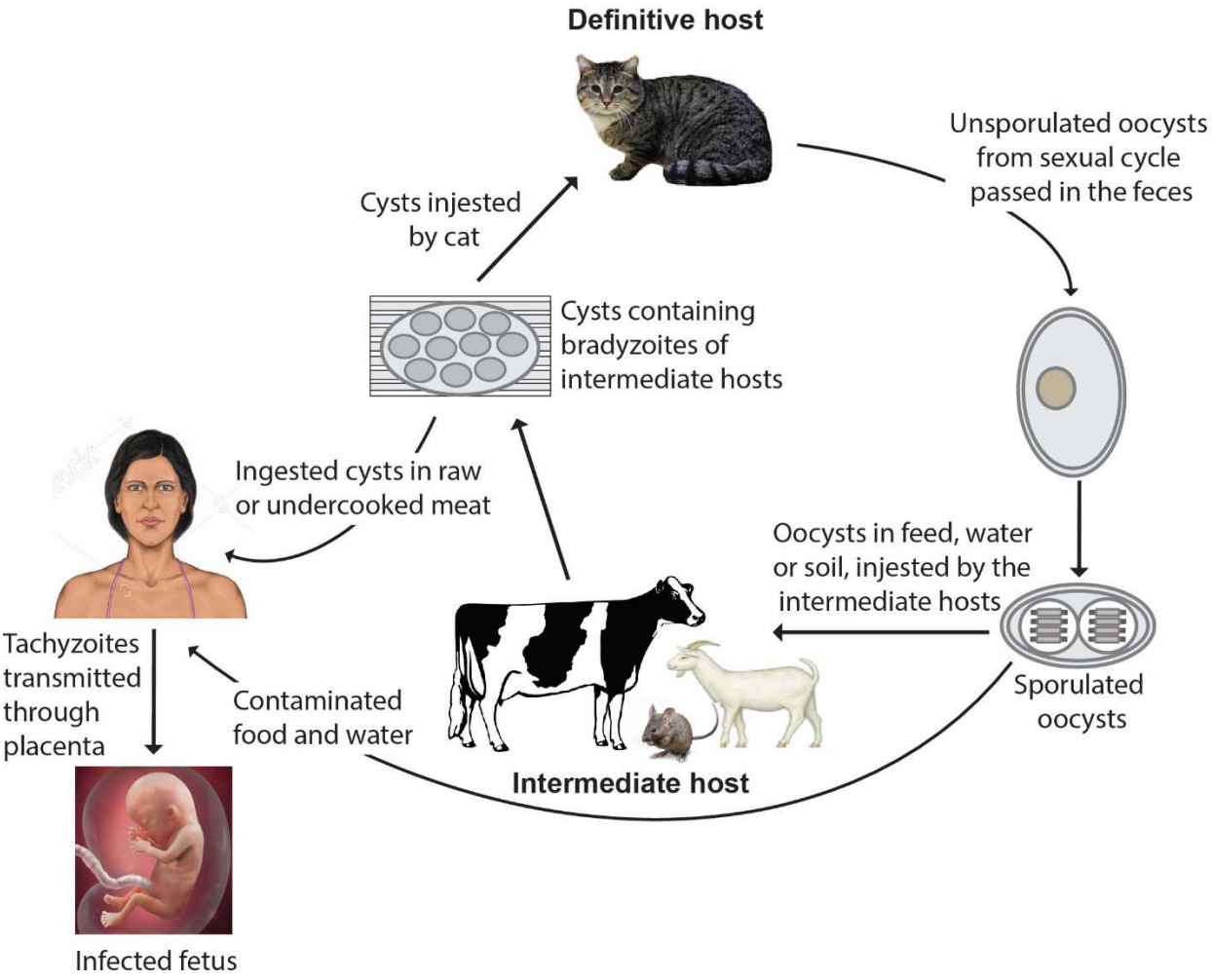


Figure 1.1. Life cycle of *Toxoplasma gondii* *(adapted from (26))*

enter new cells and establish latent infections by conversion to the encysted bradyzoite form. Enclosed in tissue cysts, bradyzoites remain viable in the infected organ for a long time. They have the potential to be passed on to a new host through the consumption of infected animals by other animals (26). In this way, *Toxoplasma* ensures its continuity in the environment.

While the life cycle of *Toxoplasma* has been well described, *in vivo* models have limitations in studying mechanistic aspects of the pathogenesis of this parasite. In order to investigate the molecular basis of stage conversion, host-pathogen relationship, pathogenesis and drug/vaccine development, *in vitro* culture systems have been developed that employ primary cell strains (such as human fibroblast monolayers), immortalized cell lines, or cell lines of various organs (such as organotypic rat brain slices) as hosts to culture the tachyzoite stage of the parasite (27-29). Tachyzoite to bradyzoite conversion occurs spontaneously *in vitro* within rat primary neurons, astrocytes, microglia, primary skeletal muscle cells differentiated to mature myotubes as well as in a rat myoblast cell line L6.C10 at higher rates than in fibroblasts (30-34). Bradyzoite conversion can also be induced in cell culture by abiotic stress factors such as acidic or alkaline pH, drugs or chemicals, heat shock, IFNγ and other inflammatory cytokines (34). Therefore, i*n vitro* culture systems now represent a valuable tool to study the lytic cycle **(Fig.1.2)** of *Toxoplasma* at the molecular and cellular level, which has relevance for understanding *in vivo* pathogenesis.

Cell Biology of T. gondii

As an obligate intracellular parasite, *Toxoplasma* needs an intracellular location for its reproduction, to fulfill most of the metabolic demands as well as to escape from the host's immune system. *Toxoplasma* has achieved the ability to successfully infect and replicate in a great variety of nucleated cells. Infection of cells is an active parasite driven event that is powered by its

cytoskeleton and a number of secretory molecules that aid in cell binding, invasion and formation of the parasitophorous vacuole (PV) (35), (36). Initially, different GPI-anchored surface antigens on the parasite membrane, which includes surface antigen (SAG) and SAG-related sequence (SRS) proteins, bind to the host cell surface matrix and receptors. Initial contact of the parasite to the cell surface by SAG proteins precedes apical attachment that involves sequential secretion of MIC and RON proteins from two apical tubular organelles, micronemes and rhoptries, respectively (**Fig. 1.3**) (37). MIC proteins possess various conserved adhesive domains that serve as ligand for different protein-protein and protein-carbohydrate interactions with the host. MIC and RON proteins create an intimate binding interface with the host as a ring-like structure and without disrupting the host cell membrane, *Toxoplasma* pushes itself along with the host cell membrane through a constriction called the moving junction. Simultaneously or immediately thereafter, ROP proteins from rhoptries are injected at the invasion site into the host cytoplasm, some of which remain associated with the developing PV membrane (PVM) during penetration. The invagination of the host cell membrane ends with the closure of moving junction and creation of the PV.

Upon invasion, many host cell membrane proteins are selectively excluded from the PVM and the parasite inserts novel proteins from rhoptries and another type of organelle known as dense-granule (38). GRA proteins from dense granules and ROP proteins modify the PV in such a way that the parasite can subvert the metabolic functions of the host (39-42). These coordinated events create a nonfusogenic vacuole by conferring unusual biophysical and biochemical properties to the PVM that secure the parasite from fusion with the host cell lysosomal pathway. Yet, the parasites are able to import host lysosomal molecules into the PV by an unconventional way, where microtubule based invagination of the PVM act as conduits to deliver host endolysomes within the PV (39). Moreover, the PVM is permeable for free diffusion of low

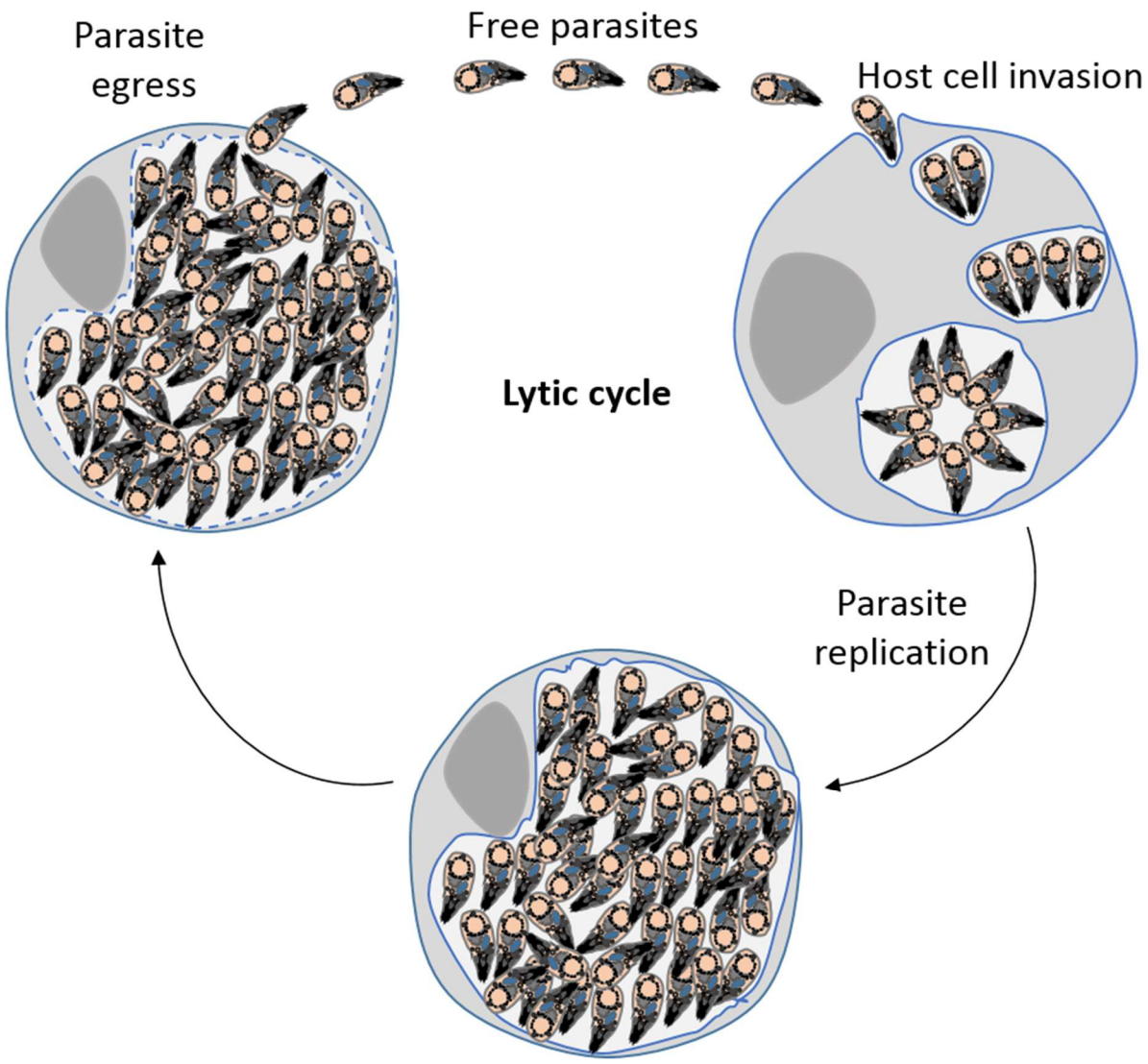


Figure 1.2. The lytic cycle of invasion, replication and egress

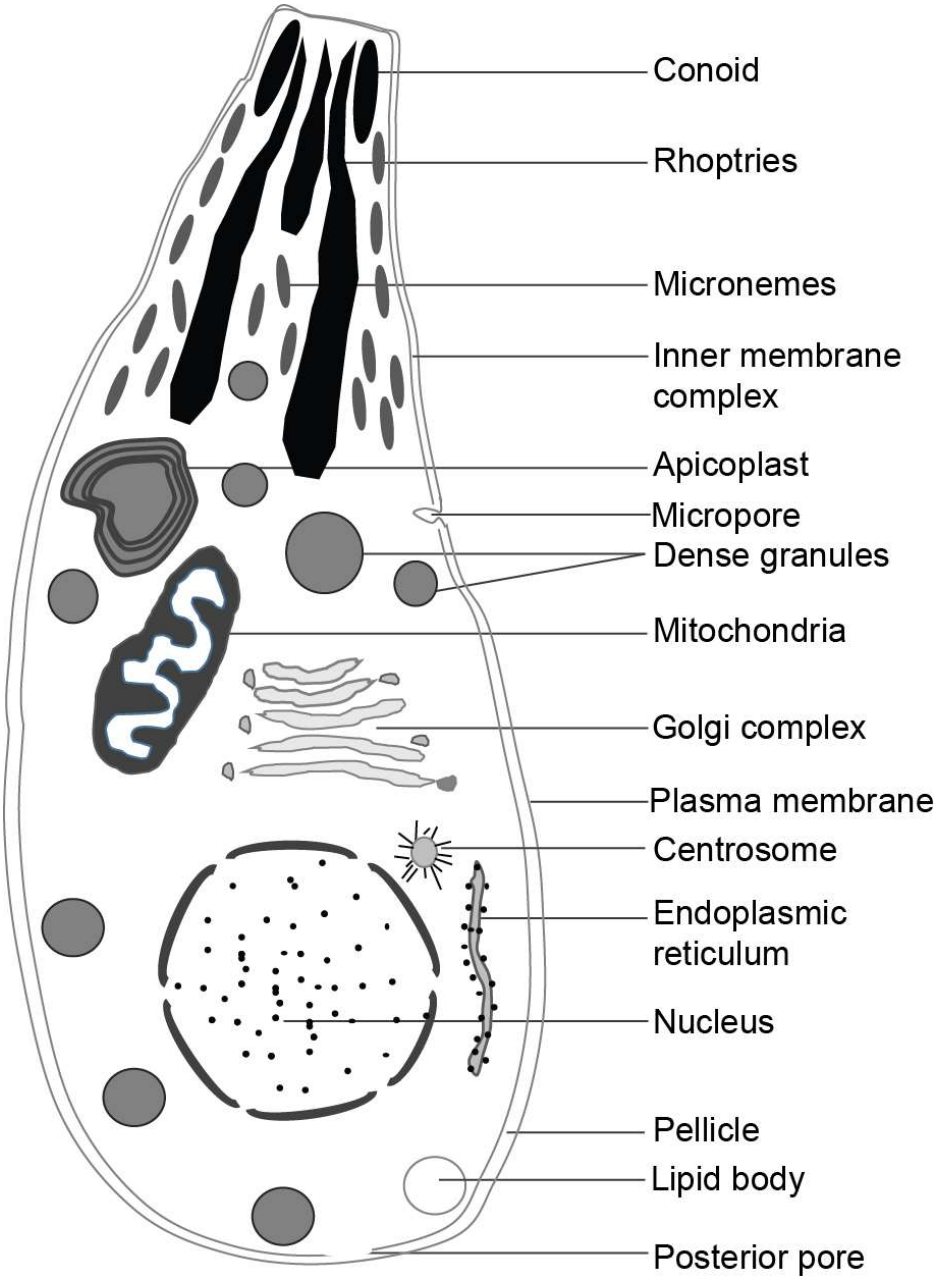


Figure 1.3. Schematic drawing of a tachyzoite cell showing intracellular organelles.

molecular weight molecules, e.g., amino acids, sugars, cofactors from the host cytoplasm. The PV closely associates with host mitochondria and endoplasmic reticulum, the cell lipid biosynthetic apparatus to gain access to the lipid sources (42-44). *Toxoplasma* needs nutrients from the host cell for their own biosynthetic pathways (40, 45). After an initial lag phage of 4 to 6 h, tachyzoites start a number of successive mitotic divisions and produce multiple daughter cells until the host cell is lysed. Parasites are also able to egress actively from the host cell before lysis (46). Released parasites rapidly invade adjacent cells and start new cycles of division. Cell destruction due to parasite burden as well as inflammatory response is the major cause of the pathology of *Toxoplasma* infection. In response to a strong immune reaction, the tachyzoites can transit to the bradyzoite stage by turning on bradyzoite specific protein expression (47). Bradyzoites replicate slowly, they are non-lytic and create host-cell derived tissue cysts inside which they can remain for years in the infected host such as in retina, skeletal and cardiac muscle, and the brain. Bradyzoites might revert to the tachyzoite stage during the immune suppression of the host and start proliferation, cell lysis and spreading to other parts of the body that might cause host death if remained untreated.

SCF-E3 Ubiquitin Ligase Complex

Ubiquitin-dependent protein degradation regulates cellular protein abundance and controls many critical biological processes in the cell (48). Lys48- or Lys11-linked polyubiquitin chains are mostly used to mark proteins for degradation by the 26S proteasome (49). Ubiquitination on the target protein is achieved by the concerted action of an enzyme cascade involving a ubiquitin-activating enzyme (E1), a ubiquitin-conjugating enzyme (E2) and a ubiquitin ligase (E3) (50). At first, the E1 enzyme activates the ubiquitin by adenylation of its C-terminal glycine that forms a

thioester bond with the cysteine of E1 enzyme. Activated ubiquitin is then transferred to the E2 conjugating enzyme that with the aid of E3 ligase transfers the ubiquitin to the epsilon amino group of a lysine on a target protein via an isopeptide linkage. E3 ligases play a central role by interacting with both the ubiquitin-loaded E2-enzyme and the specific substrates from thousands of proteins in that milieu. CRLs [Cullin–RING (really interesting new gene)] comprise the largest known category of E3 ubiquitin ligases. They are multisubunit complexes composed of a cullin, RING box protein, a variable substrate-recognition subunit (SRS), and for most CRLs, an adaptor that links the SRS to the complex (9). CRLs are classified on the type of Cullin protein in the complex. The evolutionarily conserved Cullin family has seven key members in humans that share similar structural architecture: Cul1, Cul2, Cul3, Cul4A, Cul4B, Cul5 and Cul7 (51). Cul1-based CRLs, also called SCF complexes, are the most intensively studied CRLs and contain four subunits: Skp1; Cul1; an F-box protein; and the Rbx1/Roc1/Hrt1 ring H2 finger proteins **(Fig.1.4)** (9). Here, Cul1 serves as a scaffold that binds the Skp1 and RbxI via its N and C-termini, respectively. Skp1 acts an adaptor that connects the F-box protein to the complex (52). The F-box proteins are characterized by a conserved ~40 amino acid long F-box motif that binds with Skp1, followed by a C-terminus protein-protein interaction domain such as WD-40 repeats or leucine rich repeats that bind the substrate to be ubiquitinated. There are a large number of F-box proteins in eukaryotic genomes (~70 in human and ~700 in Arabidopsis) that are thought to be involved in the ubiquitination of a large number of substrates (9, 53).

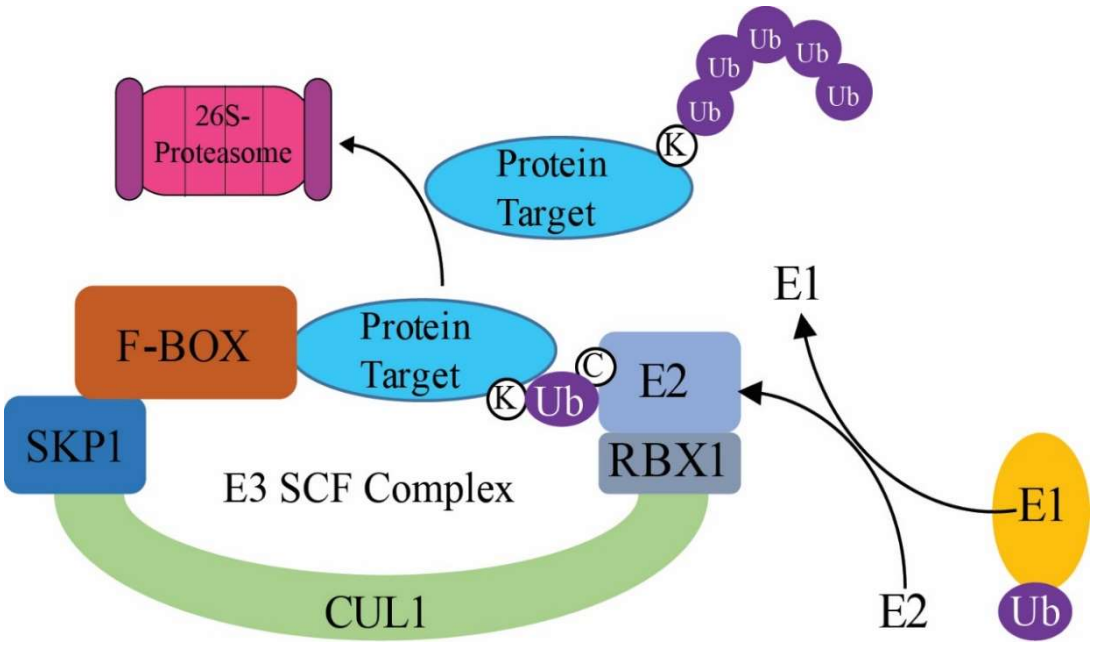


Figure 1.4. Schematic of SCF-E3 ubiquitin ligase complex.

Regulation of SCF-E3 ligases

SCF-E3 ligases play crucial role in regulated proteolysis. To ensure accurate substrate ubiquitination SCFs themselves are known to be regulated by a number of mechanisms, as described in the following sections. Significant overlap in the regulatory mechanism has been observed between SCFs and other CRLs.

Regulation by Cul1 Neddylation

Nedd8 is a ubiquitin-like protein (58% identical to ubiquitin) that is covalently linked to the C-terminal conserved Lys720 on human Cul1 in a process called neddylation (54). Neddylation on Cul1 increases the SCF-E3 ligase activity by stimulating processes such as recruitment of ubiquitin-conjugated E2 enzymes and positioning of the E2 active site for facilitating the transfer of ubiquitin onto target proteins (55, 56). Neddylation is essential in mammalian cells (57), *Caenorhabditis elegans* (58), *Arabidopsis thaliana* (59), and fission yeast (*Schizosaccharomyces pombe*) (60), but not in budding yeast (*Saccharomyces cerevisiae*) (61, 62). The process of neddylation is analogous to ubiquitination although it relies on its own E1-E2-E3 enzyme cascade to form an isopeptide bond between the terminal carboxyl group of NEDD8 with the ϵ-amino group of lysine residue on the Cul1 or other target substrates. Another conserved protein, DCN1, cooperates with RBX1 for maximal neddylation of Cul1, but is not essential for the *in vivo* neddylation reaction (63). Conversely, Nedd8 removal from Cul1 is catalyzed by the COP9 signalosome (CSN), a large octameric complex that negatively regulates the SCF-E3 ligase activity. The CSN5 subunit of this complex has a zinc-metalloprotease active site with isopeptidase activity that removes the Nedd8 conjugate from the Cul1. Paradoxical studies suggest that *in vitro* CSN serves to block the SCF-E3 ligase activity, but *in vivo* CSN actually promotes SCF substrate

degradation (64). It is not fully understood how SCF-E3 ligases are regulated by CSN interaction. From recent studies on the role of the CSN interactions with SCF and other CRLs such as CRL4, it has been proposed that substrate free SCFs/CRL4 are inactivated by CSN mediated deneddylation and upon binding of substrates, the CSN is ejected from the SCFs/CRL4 to sustain neddylation (**Fig. 1.5**) (65-71). Neddylation also regulates the inhibitor CAND1, discussed in the next section.

Regulation by the inhibitor CAND1

SCFs are also regulated by the inhibitor CAND1 (cullin-associated and neddylation-dissociated), a 120 kDa protein that binds to unneddylated Cul1 and sequesters it to an inactive state **(Fig.1.5)** (72). The crystal structure of CAND1 bound to Cul1-Rbx1 complex shows that CAND1 forms an extended structure that wraps around Cul1 while binding to its N- and C-terminal domains (73). CAND1 binding simultaneously blocks the Skp1 binding site and the Nedd8 conjugation site of Cul1. In the simultaneous presence of Nedd8 conjugation enzymes and a Skp1-F-box protein complex an active SCF complex is formed by the removal of CAND1 from Cul1 (71, 74). Interestingly, like CSN, CAND1 also shows a negative activity on SCF function in *vitro* but promotes activity *in vivo* (72, 75-77). To explain this paradox, it was proposed that both CAND1 and CSN function to protect Cul1 (78) or F-box proteins (79-81) from autoubiquitination and degradation in the absence of substrates. Or, alternatively CAND1 in combination with CSN complex redistributes the F-box proteins bound to the SCF complexes, as an exchange factor analogous to GEFs acting on G-proteins (9, 82, 83).

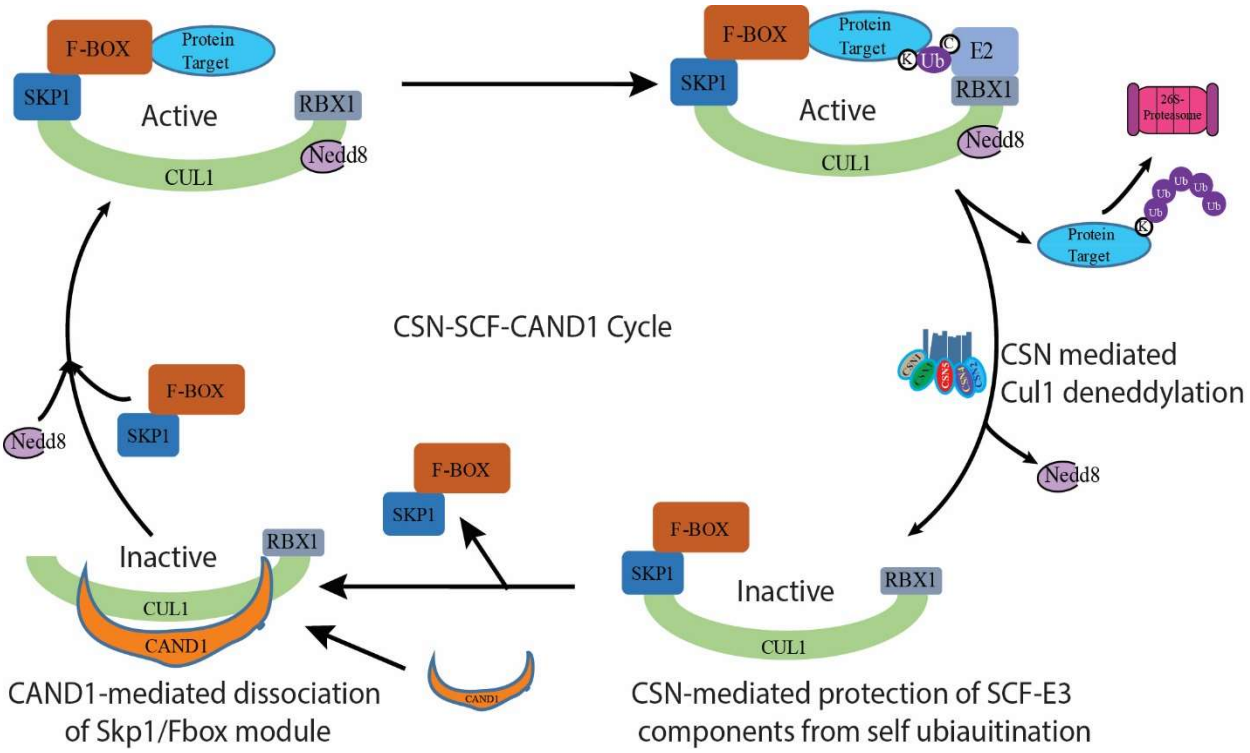


Figure 1.5. Schematic of SCF-E3 ubiquitin ligase Regulation by CSN-CAND1 cycle. A completely assembled SCF (upper right) binds to an E2 ubiquitin-conjugating enzyme and polyubiquitinates a substrate protein (blue). The polyubiquitinated SCF substrate and E2 dissociate from the complex and the substrate becomes degraded by the 26S proteasome. The CSN deneddylates the CUl1 resulting in inactivation of the SCF (lower right) and dissociation of the adaptor-substrate receptor complex from CUL1-RBX1. Binding of CAND1 to CUL1-RBX1 prevents SCF reassembly (lower left) until Nedd8 is covalently bound to CUL1 and the adaptor-substrate receptor complex associates with CUL1-RBX1 (upper left). This renders the CRL active, and a new substrate and E2 associate with the fully reassembled CRL.

Regulation through F-Box proteins

Selective expression of FBPs is important too- if they are not there then they don't operate. It has been demonstrated that SCFs can also be regulated through F-box proteins, the substrate receptor subunit of the complex (84). The F-box proteins might contain a dimerization domain (i.e., D domain), typically at the N-terminus region that might mediate dimerization of the entire SCF complex (85). A number of F-box proteins undergo dimerization *in vivo*, including Cdc4, Pop1 & Pop2, Fbw7, Fbw4, βTrcp1 & βTrcp2 and Skp2 (86-89). Inactivating mutations in the D domains of Fbx4 and Fbw7 results in dimerization deficiency and failure in substrate regulation (90, 91). Dimerization of SCF complexes can be regulated by phosphorylation events within the D domain (91), suggesting that dimerization might be regulated *in vivo*. Additionally, it has been observed that F-box proteins might also undergo autoubiquitination and proteasome mediated degradation when linked to the SCF complex (92-96).

Posttranslational Modification of Substrates

Posttranslational modification of substrates, such as phosphorylation within their SCF recognition motifs called 'phosphodegrons', can enable recognition by their respective F-box proteins as a form of substrate priming (97, 98). Recognition of phosphorylated Sic1p by WD-40 repeats of CDC4 is a well characterized example of phosphodegron utilization during the G1/S transition in *S. cerevisiae* (99, 100). Besides phosphorylation, other posttranslational modifications such as N-glycosylation can be used, as evident from the recognition of the misfolded glycosylated proteins by F-box protein Fbs1 in the cytosol that are retrotranslocated from the ER (101).

Posttranslational Modification of Adaptor Skp1

Neddylation, CSN and CAND1 dependent regulations and posttranslational modification of substrates are thought to be the basic mechanism of all CRL regulation in eukaryotes, but has been well studied only in yeasts/fungi, animals and plants. However, studies have suggested the possibility of CRL1 (SCF) specific regulation via the CRL1-specific subunit Skp1. In the protozoan *Dictyostelium discoideum*, the adaptor subunit Skp1 undergoes prolyl hydroxylation and complex O-glycosylation, which has been proposed to mediate a novel way to regulate SCF-E3 ligases in this protist (12). *Dictyostelium* Skp1 was first discovered as a fucosylated protein through the incorporation of metabolically labeled Fuc (102). Subsequent sugar composition, mass spectrometric, exoglycosidase and NMR studies revealed a pentasaccharide on a hydroxyproline (HyPro-143) of Skp1 with the structure, Galα1-3Galα1-3Fucα1-2Galβ1-3GlcNAcα1-4(*trans*)HyPro143 **(Fig.1.6A)** (103). Biochemical studies suggest that glycosylation enhances the folding order of Skp1 through increasing α-helical and decreasing β-sheet contents of Skp1 and also stabilizes the intrinsically disordered F-box combining region which surrounds Pro143. Glycosylated Skp1 isoforms show less dimerization than less modified isoforms *in vitro*, and glycosylation increases binding to a heterologous F-box protein Fbs1 (8). Based on interactome studies conducted on cell extracts from various pathway mutants, glycosylation also increases the assembly and abundance of SCF complexes containing the three F-box proteins FbxA, FbxD or JcdI ((11) & (unpublished data)), as well as proteasomal regulatory particles suggesting a role in enhancing the activity of SCF complex (12).

The discovery of Skp1 glycosylation in *Dictyostelium* has raised an intense interest to better understand this unusual cytosolic glycosylation process. Our lab has extensively studied and

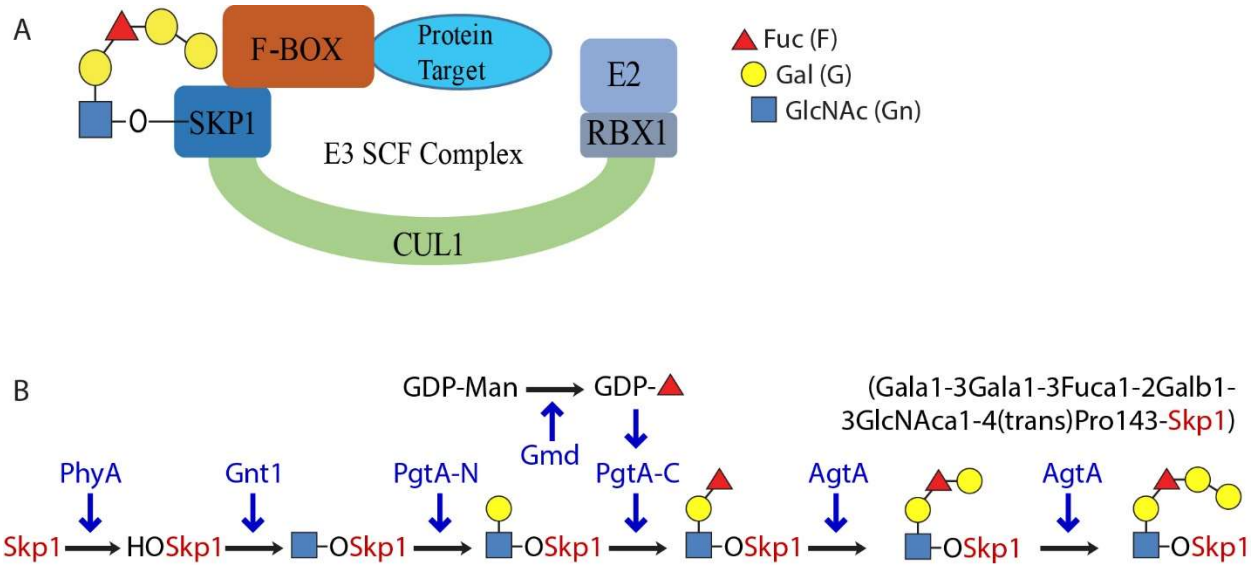


Figure 1.6. Novel regulation of SCF-E3 ligase by Skp1 glycosylation. (A) SCF complex with glycosylated Skp1 (B) Skp1 glycosylation pathway in *Dictyostelium*

discovered the complete glycosylation pathway in *Dictyostelium* (**Fig.1.6B**) (10). The glycosylation event occurs entirely in the cytoplasm and starts with the hydroxylation of Pro143 by a prolyl 4-hydroxylase (PhyA or P4H1) enzyme that lacks the N-terminal signaling sequences for the secretory pathway (104). *Dictyostelium* PhyA is closely related to a mammalian HIF-α type P4H (PHD2) that is involved in the hypoxia-dependent degradation of the transcription factor HIF-1 by hydroxylation of conserved proline residues on its HIF-α subunit. Both P4H1 and PHD2 depend on Fe^{2+}, α-ketoglutarate and ascorbic acid for their activities, and act as direct O_2 sensors due to their high K_m for O_2 (>40%) (10, 105). However, *Dictyostelium* lacks HIF-α and PhyA targets Skp1, whereas PHD2 targets only HIF-α as mammals lack the equivalent of Pro143 on their Skp1. Following hydroxylation of Skp1 by P4H1, three different glycosyltransferases with five enzymatic activities sequentially transfer the pentasaccharide on HyPro-143. Each of these glycosyltransferases have been purified from the cytosol of *Dictyostelium* based on their enzymatic activities and genes encoding these enzymes were identified by molecular approaches (106-112). The first enzyme that transfers the GlcNAc onto HyPro, named as Gnt1, is distantly related to the αGalNAc-transferases that initiate mucin-type O-linked-glycosylation in the Golgi, rather than the O-βGlcNAc transferase in the cytosol. The next galactosylation and fucosylation steps are carried out sequentially by the N- and C-terminal domains of a bifunctional enzyme, PgtA (109). The fourth and fifth sugars are mediated by AgtA, a two-domain UDP-galactose:fucoside α3-galactosyltransferase, in which the galactosyltransferase domain is fused to a β-propeller-like set of 7 WD40 domains, which has Skp1-binding activity (111, 113). Although AgtA is cytosolic, its catalytic domain is similar to a large number of plant Golgi glycosyltransferases that are involved in pectin biosynthesis. It has been shown that core disaccharide was sufficient for nuclear

enrichment of Skp1, but mere prolyl hydroxylation is not sufficient (114). The significance of this effect on compartmentalization is unknown.

While hydroxylation/glycosylation on Skp1 is proposed to regulate the SCF function in *Dictyostelium*, studies in *Saccharomyces cerevisiae* suggests that phosphorylation of Skp1 at S162 might inhibit its interaction with an F-box protein, Met30, which mediates the ubiquitination of transcription activator Met4 involving in the uptake and assimilation of sulfur into methionine and S-adenosyl-methionine (115, 116).

In aggregate, emerging evidence suggest that regulation of protein degradation not only occurs at the substrate level but also involves a number of complex mechanisms controlling CRL activity, some of which are specialized for the regulation of the SCF class itself.

SCF-E3 Ligases in Cellular Regulation

SCF-E3 ligases are the largest family of ubiquitin ligases, and are responsible for promoting the degradation of about ~20% of Ubiquitin-proteasome system (UPS) regulated proteins (117). Components of basic SCF complex are conserved in all eukaryotes from metazoans to protozoans (14, 118-121) and SCF mediated ubiquitination and degradation of proteins plays important roles in diverse cellular processes, as surveyed below.

Role in Metazoan Cell Cycle Regulation

Regulation of the eukaryotic cell cycle is dependent on controlled protein degradation by the ubiquitin-proteasome system. SCF complexes are one of the two major types of ubiquitin ligases (the other one is anaphase-promoting complex or cyclosome (APC/C)) that is responsible for periodic degradation of many cell cycle regulators (122). While the APC/C controls the

metaphase-anaphase transition, SCF controls the S-phase entry and mitotic onset by targeting various regulatory proteins for degradation that include cyclin-dependent kinases (CDKs) and cyclin-dependent kinase inhibitors (CKIs). Primarily three F-box-protein-bound SCF complexes (SCF^{Skp2}, $SCF^{\beta\text{-}TrCP}$ and SCF^{FBW7}) are responsible for cell cycle relevant proteolysis in mammalian cells (6, 123, 124). SCF^{Skp2} levels accumulate during the G1/S transition and this primarily mediates the degradation of CKIs such as $p21^{cip1}$, $p27^{kip1}$, $p57^{kip2}$, pocket protein p130 and the G1-S transition specific cyclin E (122, 123). Hence, SCF^{Skp2} promotes cells to progress through S and G2 phases by increasing the S-phase cyclin/CDK activities. On the other hand, SCF-$^{\beta TrCP}$ primarily regulates progression into mitosis through the ubiquitin-mediated proteolysis of cell cycle inhibitor Emi1 and Wee1 in early mitosis (122, 123). In contrast to SCF^{Skp2} and $SCF^{\beta\text{-}TrCP}$, SCF^{FBW7} primarily degrades the cell cycle activators such as c-Myc, C-Jun, Notch, Mcl-1 and cyclin-E and thereby controls cell-cycle progression, and acts as a general tumor suppressor (124-126).

Role in Plant Hormone Signaling

SCF mediated proteolysis controls a wide range of cellular processes in plants. Signaling by auxin, a major plant hormone that regulates plant growth and development, is regulated by the SCF complex. A family of six F-box proteins, Transport Inhibitor Response1 (TIR1) and Auxin Signaling F-Box proteins (AFBs), act as auxin receptors (127, 128). In the presence of auxin, members of the TIR1/AFB1 bind with a Aux/IAA transcriptional repressor and promote its degradation, resulting in the activation of auxin-responsive genes. Another important plant hormone Jasmonate binds the F-box protein COI1 and enables the degradation of JAZ

transcriptional repressor that allows the transcription of jasmonate responsive genes (129). The list of SCF regulated plant signaling molecules is increasing.

Role in Protozoan O_2-Dependent Development

The biological significance of Skp1 modification on the development of *Dictyostelium* has been extensively studied. *Dictyostelium* is a free living amoeba whose natural habitat is the upper layer of soil rich in decaying organic material. Bacteria are their main food source. When the food supply is abundant they live in a unicellular form but starvation induces the solitary amoebae to aggregate into a multicellular slug, which in response to external cues migrates to the soil surface and culminates into a fruiting body composed of a stalk and a spore-containing sorus. Spores are protected by a tough cell wall. The stalk raises the sorus high enough for the spores to be dispersed to distant locations where they can germinate to renew the cycle. High O_2 of the above ground environment is one of the important factors that promote the decision of slug-to-fruit body formation. At an air-water interface, *Dictyostelium* can grow and form a slug at 2.5% O_2, but ~10% is required for culmination to fruiting bodies (130). Substantial genetic, biochemical and physiological evidence suggests that Skp1 hydroxylation and glycosylation has important role in O_2-dependent development of *Dictyostelium*. First, reverse genetic analysis shows that PhyA is required for culmination to fruiting bodies (111). *PhyA* mutants have a higher O_2 requirement for culmination, suggesting its role as O_2 sensor to drive the process (131). Glycosylation itself is important too as *gnt1*, *pgtA*, and *agtA* mutants also exhibit higher than normal requirements for O_2 to culminate. Skp1 appears to be the only biochemical acceptor substrate of the enzymes in cells based on metabolic labeling studies with [^{3}H] Fuc, and biochemical complementation of mutant cell extracts with the recombinantly expressed missing glycosyltransferase and radioactive sugar

nucleotide donors. A functional role for Skp1 in O_2 sensing is demonstrated by reciprocal effects on O_2-dependence of culmination based on genetically controlled expression levels of Skp1. Overexpressed Skp1 also inhibited sporulation, and the effect of Skp1 overexpression could be countered by mutating its Pro143 or by co-overexpression of PhyA. Furthermore, a double Skp1B/agtA mutant exhibited intermediate O_2-dependence, further supporting their functioning in the same pathway as suggested by the biochemical studies. Biochemical studies also support the involvement of Skp1 glycosylation to the *Dictyostelium* development. As mentioned elsewhere, glycosylation increases the binding of three F-box proteins (FbxA, FbxD and JCDI) and at least one of them, FbxD, is involved in *Dictyostelium* development as FbxD overexpression significantly delays culmination to fruiting bodies (12). It is possible that glycosylation-dependent enhanced binding of different F-box proteins to Skp1 might increase the SCF-dependent degradation of culmination repressors. Another possibility is that increased binding of F-box proteins results in their enhanced autoubiquitionation and degradation and thereby release the culmination promoter for development. In fact, unpublished data (Boland A and West CM) from our group support the second hypothesis where proteasome inhibitor MG132 that accumulates K48-linked polyubiquitinated proteins, partially rescues the culmination of phyA-knock out mutants suggesting the accumulation of any culmination promoter that would ordinarily degrade in phyA- mutants. Additional studies are needed to know how glycosylation of SCF ligases regulates *Dictyostelium* development.

SCF E3 ligases in *Toxoplasma*

Toxoplasma tachyzoites replicate by endodyogeny, in which two daughter cells are formed within the mother during cytokinesis. While a number of maternal cellular components such as

genome-containing organelles (nucleus, apicoplast and mitochondrion) are duplicated and inherited to the daughter parasites, a large number of maternal components are also needed to be degraded and newly synthesize in the daughters (24, 132). The degraded maternal components are used as the building blocks for developing daughter parasites and therefore the degradation must be tightly regulated. Both selective proteolysis by ubiquitin proteasome system and bulk degradation by autophagy are likely involved in degradation of maternal components (133). Like many other protozoan parasites, *Toxoplasma* also shows evidence for an active proteasome (13, 134-137). Most of the well-known eukaryotic proteasome inhibitors block *Toxoplasma* growth and intracellular replication, but don't affect entry or establishment of the parasitophorous vacuole (138). Bioinformatic analyses suggest that the *Toxoplasma* genome encodes many major members of the ubiquitin proteasome system including the components of SCF complex (1 Skp1, 1 Cul1, ~16 F-box proteins, 1 Rbx1) and mammalian-type basic SCF regulatory components (**Table 1.1**) (14). Multiple attempts to knock out Skp1 by our group was unsuccessful indicating that the gene might be essential. Indeed, a genome wide CRISPR-mediated knock out strategy in *Toxoplasma* suggests that many of the predicted components of SCF-E3 ligase complexes including Skp1 might be essential for *Toxoplasma* survival (**Table 1.1**) (139). Recent attempt in mapping of ubiquitome in *Toxoplasma* tachyzoites identified over 500 ubiquitin-modified proteins, with almost 1000 sites, 35% of which are predicted to be regulated by cell-cycle (15, 140). Together, these data suggest an important role for ubiquitin-mediated proteolysis in *Toxoplasma* growth. However, the contributions of SCF E3 ligases in this process and how they are regulated in *Toxoplasma* is unknown.

Interestingly, genome wide bioinformatics analyses predict that *Dictyostelium*-like Skp1 modification occur in *Toxoplasma* as well as many other protists. Pro-143 equivalent on TgSkp1

Table 1.1: List of SCF-related genes in *Toxoplasma*. Human SCF-related protein sequences were collected from Uniprot and searched for homologs against *Toxoplasma* proteome in NCBI non-redundant databases or ToxoDB (Release 29). Phenotypic scores, an indicator of essentiality of the gene of interest for parasite survival, were collected from Reference (139). A Score ≤ 2 suggest that the gene might be essential for parasite survival in monolayer fibroblasts.

Functional group	Proteins	Name	ID	Phenotype score
SCF Complex	Skp1	Putative suppressor of kinetochore protein 1	TGGT1_207680	-4.10
	Cul1	Cullin Family protein 1	TGGT1_289310	-4.44
	RBX1	Ring box protein 1 family protein	TGGT1_213690	-4.21
	F-box	FBXW1	TGGT1_261370	-2.90
		FBXW2	TGGT1_299230	-0.47
		FBXW3	TGGT1_310910	-2.79
		FBXL1	TGGT1_262530	-0.18
		FBXO5	TGGT1_243750	-1.70
		FBXO8	TGGT1_305630	-2.20
		FBXO2	TGGT1_215210	0.49
		FBXO7	TGGT1_275780	-1.40
		FBXO3	TGGT1_225900	-2.31
		FBXO1	TGGT1_310930	-3.02
		FBXO4	TGGT1_228380	1.18
		FBXO6	TGGT1_258900	0.31
		FBXO9	TGGT1_359350	-0.45
		FBXL2	TGGT1_278815	-1.59
		FBXO10	TGGT1_215620	-0.68
		FBXO11	TGGT1_203040	1.55
Ubiquitination	E1	Putative UBA1	TGGT1_290290	-4.25
		Putative UBA7	TGGT1_314890	-4.56
		Putative UBA7	TGGT1_311500	-3.59
	E2	putative ubiquitin conjugating enzyme E2	TGGT1_310040	-2.07
		Putative UBE2A	TGGT1_216130	1.01
		Putative UBE2A	TGGT1_263490	-3.31
		Putative UBE2A	TGGT1_319870	-2.07
		Putative UBE2B	TGGT1_216130	1.01
		Putative UBE2B	TGGT1_263490	-3.31
		Putative UBE2B	TGGT1_319870	-2.07
Neddylation	CAND1	Putative cand1	TGGT1_215040	-0.87
	NEDD8	Putative nedd8	TGGT1_254600	-2.03
	E1	Putative ULA1 (E1 subunit)	TGGT1_274070	-2.29
	E1	Putative UBA3 (E1 subunit)	TGGT1_264880	-2.77
	E2	Putative UBE2F (E2)	TGGT1_280760	-4.48
	E2	Putative UBE2F (E2)	TGGT1_319870	-2.07
	E2	Putative UBE2F (E2)	TGGT1_304770	-2.47
Deneddylation	COP9	Putative CSN8	TGGT1_262040	-0.03
		Putative CSN5	TGGT1_269840	-4.73
		Putative CSN6	TGGT1_269250	-4.99
		Putative CSN2	TGGT1_236220	-1.78
		Putative CSN1	TGGT1_238180	-4.95
		Putative CSN5	TGGT1_308590	-1.10
		Putative CSN3	TGGT1_225450	-0.40
	Deneddylation enzyme (DUBs)	ubiquitin carboxyl-terminal hydrolase UCHL3	TGGT1_263470	-0.11

and certain modification pathway genes are partially conserved in *Toxoplasma*. The genomes of *Toxoplasma* along with a select group of protists harbor the first three enzymes of the *Dictyostelium* HyPro modification pathway, *phyA*, *gnt1* and *pgtA* that are responsible for hydroxylation-dependent core trisaccharide formation on *Dictyostelium* Skp1, but lack the terminal disaccharide modifying gene, *agtA* (**Table-1.2**) (10). Primary studies show that bacterially expressed TgPhyA protein can prolyl hydroxylate both *Toxoplasma* and *Dictyostelium* Skp1s and has similar kinetic properties of DdPhyA including a superimposable dependence on the concentration of its co-substrate alpha-ketoglutarate (16). Interestingly, TgPhyA had a significantly higher apparent affinity for O_2, compared to DdPhyA, but overexpression of TgPhyA in Dd*phyA*-null mutants can restore the defects in DdSkp1 glycosylation as well as O_2-requiement to a level compared to that of overexpressed DdPhyA. Moreover, Tg*phyA*-disrupted parasites show growth defects at atmospheric O_2 levels, which were more severe at low O_2 indicating the importance of *phyA* in parasite O_2 sensing.

Toxoplasma experiences a number of hypoxic and normoxic conditions inside its human host during its life cycle but lacks the HIF-α, the transcription factor that responds to hypoxic conditions in mammals. Due to the cellular/functional conservation of Tg*phyA* in relation to Dd*phyA* and the partial presence of DdSkp1-modifying glyosyltransferases in *Toxoplasma*, it has been proposed that *Toxoplasma* might also possess a *Dictyostelium*-like O_2-sensing pathway involving Skp1 glycosylation. A study in our lab supported the hypothesis by showing that disruption of *phyA* in *Toxoplasma* results in the decreased apparent M_r of Skp1 compared to wild-type Skp1 of about 1000 Da in an SDS-PAGE gel, suggesting that Skp1 from *phyA*Δ mutants lacks a hydroxylation-dependent modification, possibly glycosylation (16). Another study showed evidence of Gnt1 activity in a *Toxoplasma* cytosol preparation that targets *Dictyostelium* Skp1 *in*

Table 1.2: Homologs of DdSkp1 modification pathway-like genes in *Toxoplasma* and other organisms. BLASTP and TBLASTN searches were performed for sequences corresponding to *Dictyostelium* Skp1 and its HyPro-modification pathway related genes (*phyA*, *gnt1*, *pgtA* and *agtA*). The presence (+), dual functional (---------) and reverse orientation of the domains (R) relative to *Dictyostelium* enzyme is shown. No candidates for a homolog of DdAgtA (highlighted in green) were detected in organisms other than amoebazoans using these algorithms or PSI- or PHIBLAST toward its catalytic domain. The *Dictyostelium* Skp1 modification genes are on the top row. *Toxoplasma* and numerous other organisms lack AgtA but possess earlier genes of the pathway.

		Skp1-	P4H1	Gnt1	PgtA		AgtA	
Group	**Organism**	**Pro**	**PhyA**	**α-GlcNAc**	**β1,3Gal**	**α1,2Fuc**	**α1,3Gal-1**	**α1,3Gal-2**
Amoebazoa	*Dictyostelium discoideum* (cellular slime mold)	+(2)	+	+	+-------	-------+	+-------	-------+
	D. purpureum (cellular slime mold)	+	+	+	+-------	-------+	+-------	-------+
	D. fasciculatum (cellular slime mold)	++(2)	+	+	+-------	-------+	+-------	-------+
	Polysphondylium pallidum (cellular slime mold)	++(2)	+	+	+-------	-------+	+-------	-------+
	Acytostelium subglobosum (cellular slime mold)	++(2)	+	+	+-------	-------+	+-------	-------+
	Physarum polycephalum (acellular slime mold)	+	+	+	+-------	-------+	+-------	-------+
	Acanthamoeba castellani (Neff)(soil protozoa)	+	+	+				
Rhizaria	*Reticulomyxa filosa* (fresh water foraminifera)	+(4)	+	+	+	+		
	Bigelowiella natans CCMP2755 (chlorarachnea)	+(?more)	+	+		+		
Alveolates	***Toxoplasma gondii*** (apicomplexan)	+(2)	+	+	+-------	-------+R		
	Hammondia hammondi (apicomplexan)	+	+	+	+-------	-------+R		
	Neospora caninum (apicomplexan)	+	+	+	+-------	-------+R		
	Sarcocystis neurona (apicomplexan)	+	+	+	+-------	-------+R		
	Chromera velia (chromerida)	+	+	+	+-------	-------+R		
	Vitrella brassicaformis (chromerida)	+	+	+	+-------	-------+R		
	Stylonychia lemnae (ciliated)	+	+-------	-------+	+-------	-------+R		
	Oxytricha trifallax (ciliated)	+	+-------	-------+	+-------	-------+R		
	Simbiodinium microadriaticum (dinoflagellate)	+	+	+	+-------	-------+R		
	Karenia brevis (marine unicellar dinoflagellate; red tide)	+	+	+	+-------	-------+R		
Chromalveolate	*Nannochloropsis gaditana* (marine microalga)	+	+-------	-------+	+-------	-------+R		
	Guillardia theta CCMP 2712 (cryptophyte)	+	+-------	-------+		+		
Heterokonts/	*Thalassiosira pseudonana* (diatom)	+	+-------	-------+	+-------	-------+R		
Stramenopiles	*Albugo laibachii* (oomycete)	+	+-------	-------+	+-------	-------+R		
	Pythium ultimum (oomycete)	+	+-------	-------+	+-------	-------+R		
	Aureococcus anophagefferens (non-motile pelagophyte)	+		+	+-------	-------+R		
	Ectocarpus siliculosis (brown alga)	+(2)	+-------	-------+	+-------	-------+R		
	Aphanomyces euteiches, strain: 109 (oomycete)	+(2)	+-------	-------+	+-------	-------+R		
	Phytophtera infestans (oomycete)	+	+-------	-------+				
	P. ramorum (oomycete)	+	+-------	-------+				
	P. sojae (oomycete)	+	+-------	-------+				
	Phaeodactylum tricornutum (marine pennate diatom)	+	**+(2)**	+				
	Fragilariopsis cylindrus (coldwater diatom)	+	+-------	-------+				
	Emiliania huxleyi (marine haptophyte alga)	+	+-------	-------+				
	Chlamydomonas reinhardtii (chlorophyte alga)	+	+	+				
	Volvox carteri (chlorophyte alga)	+	+-------	-------+				
	Coccomyxa subellipsoidea (chlorophyte alga)	+	+-------	-------+				
	Chlorella variabilis (chlorophyte single-cell green alga)	+	+-------	-------+				
Excavates	*Malawimonas californiana*	+	+	+				
	Naegleria gruberi (amoebe flagellate, heterolobosea)	+	+-------	-------+				

vitro suggesting that Skp1-glycosylation pathway might be conserved in *Toxoplasma* (16, 141). In addition, the growth defect of *phyA* mutants in atmospheric O_2 levels in addition to lower O_2 suggests that glycosylation-dependent SCF regulation is important for generalized tachyzoite proliferation. Other life cycle stages were not investigated. However, it is not known about the nature of this glycosylation, and the limited conservation of *Dictyostelium* pathway genes in *Toxoplasma* made it difficult to anticipate the pattern of this glycosylation. It is also unknown if the *Dictyostelium* glycosyltransferase homologues Gnt1 and PgtA are involved in *Toxoplasma* glycosylation process or if any novel glycosyltransferases are also involved and have any role in parasite proliferation. It is also unknown how Skp1 modification pathway regulate the SCF-E3 ligases in *Toxoplasma* and how SCF-E3 ligases regulate parasite proliferation and how Skp1 glycosylation affect the overall SCF-activity. *Toxoplasma* might be a model for studying the pathway in other organisms that harbor *Toxoplasma*-like pathway including plant pathogen *Pythium ultimum* and animal pathogen *Sarcocystis neurona.* Since, the pathway is absent from the mammalian hosts, understanding the Skp1 modification pathway might lead it to be a druggable target for parasite control.

CHAPTER 2

THREE GLYCOSYLTRANSFERASE ACTIVITIES CATALYZE SYNTHESIS OF THE CORE TRISACCHARIDE ON *TOXOPLASMA* SKP1 AND PLAY IMPORTANT ROLES IN PARASITE GROWTH[1]

[1]Kazi Rahman, Peng Zhao, Msano Mandalasi, Hanke van der Wel, Lance Wells, Ira J Blader, Christopher M West. 2016. *Journal of Biological. Chemistry.* 291: 4268-4280

Reprinted here with permission of the publisher.

Abstract

Toxoplasma gondii is a protist parasite of warm-blooded animals that causes disease by proliferating intracellularly in muscle and the central nervous system. Previous studies showed that a prolyl 4-hydroxylase related to animal HIFα prolyl hydroxylases is required for optimal parasite proliferation, especially at low O_2. We also observed that Pro154 of Skp1, a subunit of the Skp1/Cullin-1/F-box protein (SCF)-class of E3-ubiquitin ligases, is a natural substrate of this enzyme. In an unrelated protist, *Dictyostelium discoideum*, Skp1 hydroxyproline is modified by 5 sugars via the action of three glycosyltransferases, Gnt1, PgtA and AgtA, which are required for optimal O_2-dependent development. We show here that TgSkp1 hydroxyproline is modified by a similar pentasaccharide, based on mass spectrometry, and that assembly of the first three sugars is dependent on *Toxoplasma* homologs of Gnt1 and PgtA. Reconstitution of the glycosyltransferase reactions in extracts with radioactive sugar nucleotide substrates and appropriate Skp1 glycoforms, followed by chromatographic analysis of acid hydrolysates of the reaction products, confirmed the predicted sugar identities as GlcNAc, Gal and Fuc. Disruptions of *gnt1* or *pgtA* resulted in decreased parasite growth. Off target effects were excluded based on restoration of the normal glycan chain and growth upon genetic complementation. By analogy to *Dictyostelium* Skp1, the mechanism may involve regulation of assembly of the SCF complex. Understanding the mechanism of *Toxoplasma* Skp1 glycosylation is expected to help develop it as a drug target for control of the pathogen, as the glycosyltransferases are absent from mammalian hosts.

Introduction

Toxoplasma is a world-wide, obligate intracellular apicomplexan parasite that infects most nucleated cells (142). Toxoplasmosis, the disease caused by *Toxoplasma*, is an opportunistic infection in AIDS and other immune-suppressed patients (143). In addition, *in utero* infections can cause mental retardation, blindness, and death (144). *Toxoplasma* is transmitted by digesting parasites from feline feces (as oocysts) or undercooked meat (as tissue cysts). Once in the host, parasites convert to the tachyzoite form that disseminates to peripheral tissues (e.g. brain, retina, and muscle). The resulting immune response and/or drugs can control tachyzoite replication but the parasite survives by converting into slow growing bradyzoites that encyst. Cysts sporadically burst and the released parasites convert to tachyzoites whose unabated growth, as can occur in immune suppressed hosts, results in cell and tissue damage (145). Currently, no *Toxoplasma* vaccine exists, anti-toxoplasmosis drugs have severe side effects, and resistance to these drugs is occurring.

Recently, disruption of the gene for PhyA, the prolyl 4-hydroxylase that hydroxylates Pro154 in Skp1, was observed to reduce tachyzoite proliferation in cell culture and fitness in a competition assay (16). Skp1 is an adaptor in the SCF (Skp1/Cullin-1/F-box protein) class of E3 ubiquitin ligases and its hydroxylation was hypothesized to contribute to O_2-dependent proliferation. That study noted that loss of hydroxylation resulted in increased migration in SDS-PAGE gels suggesting a decrease in M_r of approximately 1000 Da. Previous studies in an unrelated protist, the social soil amoeba *Dictyostelium*, had shown that the Skp1-hydroxyproline (Hyp) could be glycosylated by 5 glycosyltransferase activities encoded by 3 genes, resulting in assembly of a pentasaccharide at the equivalent Pro-residue (10, 113). Since two of these genes, *gnt1* and *pgtA*, have apparent homologs in the *Toxoplasma* genome, we suspected that the gel shift might result

from inability of the glycosyltransferases to modify Skp1 in the absence of formation of the Hyp anchor. In *Dictyostelium*, Dd*gnt1* encodes a polypeptide αGlcNAc-transferase that transfers GlcNAc from UDP-GlcNAc to form GlcNAcα1-O-Skp1 (107). Dd*pgta* encodes a dual function diglycosyltransferase whose N-terminal domain then transfers Gal from UDP-Gal to form a Galβ1-3GlcNAc linkage, and whose C-terminal domain processively transfers Fuc from GDP-Fuc to form a Fucα1-2Gal linkage (109). However, the two domains are switched in the *Toxoplasma* version of the predicted protein (TGGT1_260650), and there is no evidence for *agtA*, the *Dictyostelium* gene that is responsible for addition of the final two sugars, both αGal residues.

If the Skp1 Hyp of *Toxoplasma* can be glycosylated, the importance of hydroxylation for proliferation might be due to consequent loss of glycosylation rather than inability to hydroxylate *per se*. In *Dictyostelium*, hydroxylation alone partially rescues O_2-dependent development (146). Full recovery depends, however, on full glycosylation (147), and glycosylation is required to promote efficient assembly of the Skp1/F-box protein heterodimer, based on interactome studies (12). Therefore, we sought direct evidence for Hyp-dependent glycosylation of *Toxoplasma* Skp1 and the role of the predicted glycosyltransferases, and used disruption strains of the predicted glycosyltransferase genes to test their contribution to parasite proliferation. The findings implicate Skp1 as the functional target of this novel post-translational modification pathway in *Toxoplasma* and indicates that the Skp1 modification pathway is evolutionarily conserved amongst protists.

Experimental Procedures

Parasites, Cell Culture and Plaque Assays– Toxoplasma strain RHΔ*ku80*Δ*hxgprt* (RHΔΔ) was cultured in association with human foreskin fibroblasts (HFFs) using Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% (v/v) fetal bovine serum, 2 mM L-glutamine

and 100 units/ml penicillin/streptomycin (Complete medium) in a humidified CO_2 (5%) incubator at 37°C. RH*Δku80ΔphyA* (RH*phyA*Δ), RH*Δku80Δgnt1* (RH*gnt1*Δ) and RH*Δku80ΔpgtA* (RH*pgtA*Δ) strains were cultured in the same medium supplemented with 25 μg/ml mycophenolic acid (Sigma) and 25 μg/ml xanthine (Sigma). RHΔ/SF and RH*phyA*Δ-1/SF strains where TgSkp1 was tagged with SF-tag and have the chloramphenicol acetyltransferase marker were cultured in the DMEM medium supplemented with 20 μM of chloramphenicol (Sigma). Strains were cloned by limiting dilution in 96-well plates.

To perform cell growth plaque assays, confluent HFF monolayers in 6-well tissue culture plates were infected with freshly lysed-out (see below) parasites at 250 parasites/well, equivalent to a multiplicity of infection of 0.002. After 3 h, unattached parasites were removed by two rinses with phosphate buffered saline (15 mM sodium phosphate, pH 7.4, 135 mM NaCl). After undisturbed incubation in complete medium for 5.5 d, monolayers were fixed with methanol and stained with crystal violet to detect plaques. Plaques ($n \geq 50$) from at least 2 wells were manually encircled and areas were calculated by NIH ImageJ software. Data were presented and statistically analyzed using GraphPad Prism version 6.

TgphyA, Tggnt1 and TgpgtA Disruption Strains– DNAs for gene disruptions were generated from pminiGFP.ht (gift of Dr. Gustavo Arrizabalaga, University of Idaho), in which the *hxgprt* gene is flanked by multiple cloning sites. The approach was modeled after that used for the Tg*phyA* disruption strain RH*phyA*Δ-1, in which exon 1 of Tg*phyA* was replaced with *hxgprt* (16). To generate an independent Tg*phyA* disruption strain, RH*phyA*Δ-2, the complete coding region was replaced with *hxgprt* by double cross-over homologous recombination. First, the 5'-flank and 3'-flank targeting sequences of Tg*phyA* from RHΔΔ were PCR amplified with primer pairs a and

a' and pairs b and b', respectively (**Supplemental Table S2.1**). The 5'-fragment was digested with KpnI and HindIII and inserted into pminiGFP.ht between its KpnI and HindIII sites. The resulting plasmid was digested with XbaI and NotI and ligated to the XbaI and NotI digested 3'-flank. The resulting vector was linearized with KpnI and electroporated into RHΔΔ strain as described (16). Drug-resistant transformants were selected in the presence of 25 μg/ml MPA and 25 μg/ml xanthine, and cloned by limiting dilution. Genomic DNA from three clones was screened by PCR to identify Tg*phyA* disruption strain RH*PhyAΔ-2*, as described in Results. The PCRs were performed on extracts from 2 x 10^6 parasites, using Taq polymerase, and primers as listed in **Supplemental Table S2.1**. Standard conditions included 1.5 mM $MgCl_2$, and reactions were run for 30 cycles of the following standard scheme: 94°C, 30 s; 60°C, 1 min; 68°C, 3 min. Exact conditions were adjusted for specific reactions.

To disrupt Tg*gnt1*, the 5'-flank and 3'-flank targeting sequences were PCR amplified with primer pairs c and c' and d and d' (**Supplemental Table S2.1**), respectively, and inserted into pminiGFP.ht as above. The vector was linearized with SapI, transfected into RHΔΔ, and drug-resistant clones were screened by PCR to generate the Tg*gnt1*-disruption strain (RHg*nt1*Δ). Similarly, the Tg*pgtA* disruption construct was generated by PCR amplification and insertion into pminiGFP.ht of 5'-flank and 3'-flank targeting sequences using primer pairs e and e' and f and f', respectively. After digestion with PacI, the DNA was transfected into RHΔΔ, and RH*pgtA*Δ clones were screened by PCR.

Tggnt1 and TgpgtA Complemented Strains– pminiGFP.ht was used as the backbone for constructing Tg*gnt1* complementation construct after removing its HXGPRT cassette by KpnI and NotI digestion. A 7-kb DNA fragment containing the Tg*gnt1* genomic region was PCR-amplified

using primer pairs c and d' (**Supplemental Table S2.1**), digested with KpnI and NotI, and ligated into the similarly digested pminiGFP.ht. The resultant vector was linearized with KpnI and electroporated into RH*gnt1*Δ. Transformants were selected under 300 µg/ml of 6-thioxanthine (Matrix Scientific), and clones were screened by PCR. In order to complement Tg*pgtA* knock-out, the fosmid clone Rfos01M21 (148), containing a 36-kb fragment of RH strain chromosome VIIb (2039542-2076165), which includes the Tg*pgtA* gene (gift of Dr. Boris Striepen, University of Georgia), was linearized with ScaI and electroporated into RH*pgtA*Δ. Complemented clones were isolated as described for Tg*gnt1*.

Epitope Tagging of Endogenous TgSkp1– To modify the C-terminus of endogenous TgSkp1, the *skp1* genomic locus was modified by the insertion of SF-tag cDNA. A 1.5-kb region upstream of the *skp1* stop codon was PCR amplified using primer pairs g and g'. Using a ligation independent cloning strategy (149), the product was inserted into PacI digested pSF-TAP-LIC-HXGPRT and pSF-TAP-LIC-CAT vectors (from Dr. Vern Carruthers, University of Michigan). 50 µg of the resulting constructs were linearized with EcoRV at position 527 of the insert, and the DNAs were electroporated into RHΔΔ and RH*phyAΔ-1*, respectively. RHΔ/SF transformants were selected under 25 µg/ml MPA/xanthine, and RH*phyA*Δ-1/SF were selected under 20 µM chloramphenicol. Site-specific integration was confirmed by PCR of clones using the primer pairs h and h'. DNA sequencing confirmed that the 3'-end of the Skp1 coding sequence encoded the native C-terminus (…VREENKWCEDA) followed by a peptide containing 2 Strep-II tags and a FLAG tag (in bold): *AKIGSGGREF**WSHPQFEK**GGGSGGGSGGGS**WSHPQFEK**GASGE**DYKDDDDK***. Characteristics of the above strains are summarized in **Table 2.1**.

Table 2.1. *Toxoplasma* strains used in this study.

Strain	Parental strain	Genotype	Gene targeted	Selection marker	Selection drug	Ref.
KU80ΔΔ	RH(I)	Δku80;Δhxgprt	–	–	–	(150)
RHΔ*phyA*-1	KU80ΔΔ	*phyA*Δ;Δku80	*phyA*-exon 1	Hxgprt	MPA, xanthine	(16)
RHΔ*phyA*-2	KU80ΔΔ	*phyA*Δ;Δku80	*phyA*-exons 1-9 (all)	Hxgprt	MPA, xanthine	TR[a]
RHΔ*gnt1*	KU80ΔΔ	*gnt1*Δ;Δku80	*gnt1*-exon 1 (all)	Hxgprt	MPA, xanthine	TR
RHΔ*gnt1*/ complemented	RHΔ*gnt1*	Δku80;Δhxgprt	–	ΔHxgprt	6-thioxanthine	TR
RHΔ*pgtA*	KU80ΔΔ	*pgtA*Δ;Δku80	*pgtA*-exons 1-14 (all)	Hxgprt	MPA, xanthine	TR
RHΔ*pgtA*/ complemented	RHΔ*pgtA*	Δku80;Δhxgprt	–	ΔHxgprt	6-thioxanthine	TR
RHΔΔ/Skp1-SF	KU80ΔΔ	*Skp1*^{SF};Δku80; Δhxgprt;CAT^{+}	Skp1 C-terminus	CAT	Chloram-phenicol	TR
RHΔ*phyA*-1/ Skp1-SF	RH*phyA Δ*-1	*Skp1*^{SF};*phyA*Δ; Δku80;CAT^{+}	*phyA*-exon 1; Skp1 C-terminus	CAT	Chloram-phenicol	TR

TR[a], this report

Purification of Endogenous TgSkp1– Tachyzoites from RHΔΔ, RH*phyAΔ-1*, RH*gnt1Δ* and RH*pgtAΔ* strains were harvested from infected HFF monolayers by scraping and passage through a 27-gauge needle, centrifuged at 2000 × *g* for 8 min at room temperature, resuspended in sterile phosphate-buffered saline and counted on a hemacytometer chamber as described (151). 6 × 10^8 tachyzoites were pelleted, frozen at -80 °C, and subsequently thawed and solubilized in 8 M urea, 50 mM HEPES-NaOH (pH 7.4) supplemented with protease inhibitors (1 mM PMSF, 10 µg/ml aprotinin and 10 µg/ml leupeptin) at 4 °C for 30 min. The lysates were centrifuged at 16,000 × *g* for 15 min at 4 °C, supernatants (S16) were collected and diluted 8-fold in IP buffer (0.2% NP-40 (v/v) in 50 mM HEPES-NaOH, pH 7.4, protease inhibitors) and incubated with 60 µl of rabbit polyclonal anti-TgSkp1 UOK75-Sepharose beads for 1 h at 4 °C. The UOK75 antiserum (16) was first affinity purified against recombinant TgSkp1-Sepharose beads, performed as described for affinity purification of anti-DdSkp1 (12), and then coupled to CNBr-activated Sepharose CL-4B. After centrifuging and resuspending the beads three times with IP buffer and three times with wash buffer (10 mM Tris-HCl, pH 7.5, 154 mM NaCl), bound material was eluted twice with 150 µl of 133 mM triethylamine (pH 11.5) for 10 min and immediately neutralized with 150 µl of 200 mM acetic acid (pH 2.7). The pooled eluates (~400 µl) were divided into two equal parts, concentrated by vacuum centrifugation to ~10 µl, and snap-frozen in liquid nitrogen. To purify SF-tagged TgSkp1, soluble S16 fractions were prepared from RHΔ/SF and RHp*hyA*Δ-1/SF strains as described above, and incubated with 100 µl of mouse anti-FLAG M2-agarose beads (Sigma) for 1 h at 4 °C. The beads were washed as above, and eluted with 300 µl 8 M urea in 25 mM NH_4HCO_3 (natural pH), supplemented with 40 mg urea, and incubated for 15 min at room temperature, and the supernatants harboring TgSkp1-SF was collected at 2400 × *g* for 5 min in room temperature.

Mass Spectrometry of TgSkp1 Peptides– The untagged TgSkp1samples were taken to dryness and solubilized in 100 µl 8 M urea in 25 mM NH_4HCO_3. The untagged and SF-tagged TgSkp1 samples were reduced by addition of 0.5 M TCEP to a final concentration of 5 mM for 20 min at 22 °C, and alkylated by the addition of 0.5 M iodoacetamide to a final concentration of 10 mM for 15 min in the dark. TCEP was then added to a final concentration of 10 mM, and 300 µl of 50 mM NH_4HCO_3 (pH 7.8) was added to dilute urea to 2 M. Samples were treated with 1 mg/ml mass spectrometry grade trypsin (Promega) at a final concentration 8.0 µg/ml and incubated overnight at 37ºC. Peptides were recovered by adsorption to a C18 Zip Tip (OMIX TIP C18 100 µl) and eluted with 0.1% trifluoroacetic acid in 50% (v/v) acetonitrile.

Dry peptides were reconstituted in 15.6 µL of solvent A (0.1 % formic acid) and 0.4 µL of solvent B (0.1% formic acid in 80% acetonitrile), and loaded onto a 75 µm (I.D.) × 115 mm C18 capillary column (YMC GEL ODS-AQ120ÅS-5, Waters) packed in-house with a nitrogen bomb. Peptides were eluted into the nanospray source of an LTQ Orbitrap™ mass spectrometer (Thermo Fisher Scientific) with a 160-min linear gradient consisting of 5-100% solvent B over 100 min at a flow rate of 250 nL/min. The spray voltage was set to 2.0 kV and the temperature of the heated capillary was set to 210 °C. Full scan MS spectra were acquired from *m*/*z* 300 to 2000 at 30k resolution, and MS2 scans following collision-induced fragmentation were collected in the ion trap for the 12 most intense ions. The raw spectra were searched against a *Toxoplasma gondii* protein database (UniProt ATCC 50611/Me49, Sep. 2013) using SEQUEST (Proteome Discoverer 1.3, Thermo Fisher Scientific) with full MS peptide tolerance of 50 ppm and MS2 peptide fragment tolerance of 0.5 Da, and filtered to generate a 1% target decoy peptide-spectrum-match false discovery rate for protein assignments. The spectra assigned as glycosylated TgSkp1 peptides were manually validated.

Cytosolic (S100) Extracts– Parasites were permeabilized as described (152) with slight modifications. Briefly, a pellet of 2.5×10^9 frozen tachyzoites was resuspended in 500 µl of ice-cold water containing 10 µg/ml aprotinin, 10 µg/ml leupeptin, 1 mM PMSF and 0.1 mM *N*-α-*p*-tosyl-L-lysine chloromethyl ketone (TLCK) and incubated for 10 min on ice. The suspension was transferred to a Dounce homogenizer and sheared by 10 strokes, diluted with an equal volume of 100 mM HEPES-NaOH at pH 7.4, 10 mM $MgCl_2$, 10 mM $MnCl_2$, 50 mM KCl, 10 µg/ml aprotinin, 10 µg/ml leupeptin, 1 mM PMSF, and sheared by 20 additional strokes. After confirmation of lysis using phase contrast microscopy, the lysate was centrifuged at 200,000 × *g* at 5°C for 35 min, the supernatant (S100) was immediately desalted over a PD10 column at 5°C into 50 mM HEPES-NaOH (pH 7.4), 5 mM $MgCl_2$, 15% (v/v) glycerol, and 0.1 mM EDTA. Fractions with highest A_{280} values (≥1, 1 cm path length) were snap frozen at -80°C for enzyme assays.

Glycosyltransferase Assays– Skp1-dependent GlcNAcT activity was assayed in S100 fractions by the transfer of [^{3}H] from UDP-[^{3}H]GlcNAc to exogenous *Dictyostelium* HO-DdSkp1 (106). Typically, a 50 µl reaction volume containing 30 µl S100 fraction, 50 pmol of HO-DdSkp1 (11), and 0.5-2.5 µM UDP-GlcNAc (including 1 µCi UDP-[^{3}H]GlcNAc at 37 Ci/mmol, Perkin-Elmer), in 50 mM HEPES-NaOH (pH 7.4), 10 mM $MgCl_2$, 2 mM DTT, 3mM NaF, and protease inhibitors, was incubated at 30 °C for 0, 1 or 3 h. Reactions were stopped by addition of 4×-Laemmli electrophoresis sample buffer, supplemented with 2 µg of soybean trypsin inhibitor (Sigma) as a marker that comigrates with DdSkp1, boiled for 5 min, and resolved by SDS-PAGE (see below). The gel was stained for 1 h with 0.25% (w/v) Coomassie Blue in 45% (v/v) methanol, destained overnight in 5% (v/v) methanol, 7.5% (v/v) acetic acid and rinsed in H_2O for 1 h. Five ~1-mm gel slices including and surrounding the soybean trypsin inhibitor band were excised and

incubated in 7 ml of a scintillation cocktail containing 100 ml Soluene-350 (PerkinElmer) and 900 ml of 0.6 g/l PPO and 0.15 g/l dimethyl-POPOP in scintillation grade toluene. After 5 d, ^{3}H was quantitated by scintillation counting in a Beckman LSC6500 instrument.

GalT activity was assayed similarly except that the donor was 1 μM UDP-[^{3}H]Gal, prepared from a mixture of 1 μCi UDP-[^{3}H]Gal (113) and unlabeled UDP-Gal, and the acceptor was recombinantly prepared GlcNAc-DdSkp1 (11). FucT activity was similarly assayed except that 2 μM GDP-[^{3}H]Fuc, prepared from a mixture of GDP-[^{3}H]Fuc (1 μCi of 20 Ci/mmol, Perkin-Elmer) and unlabeled GDP-Fuc, was used in place of UDP-[^{3}H]Gal, and 2 μM unlabeled UDP-Gal was added to generate Gal-GlcNAc-DdSkp1 acceptor from the added GlcNAc-DdSkp1.

Radioactive Sugar Analyses– The chemical form of the radioactivity incorporated into DdSkp1 was determined by HPAEC (high pH anionic exchange chromatography) analysis after acid hydrolysis as described (106). Briefly, Gnt1 reaction products were resolved by SDS-PAGE and electroblotted onto a 0.45 μm PVDF membrane (EMD Millipore). The membrane was stained with 0.2% Ponceau S in 3% (w/v) TCA, and the Skp1 protein band was excised with a razor blade, submerged in 400 μl of 6 M HCl, and incubated at 100°C for 4 h. Hydrolysates were removed into microtubes, evaporated to dryness under vacuum centrifugation, dissolved in 500 μl H_2O and dried again twice, and reconstituted in 25 μl H_2O. A solution containing 1 nmol of $GalNH_2$ and $GlcNH_2$ in 9 μl H_2O was added to the hydrolysates and chromatographed on PA-1 column on a DX-600 Dionex HPAEC in 16 mM NaOH at 1 ml/min with pulsed amperometric detection. Fractions were collected into EconoSafe (Research Products International) scintillation fluid and counted for ^{3}H incorporation on a liquid scintillation counter. PgtA assay products were similarly hydrolyzed in 4 M TFA and mixed with a standard solution containing 1.5 nmol each of Glc, Gal, Man and Fuc.

Western Blotting– Western blotting was performed as described (16). Briefly, tachyzoite pellets were solubilized in lysis buffer containing 8 M Urea, 50 mM HEPES-NaOH (pH 7.4) supplemented with protease inhibitors at 4°C for 30 min and centrifuged at 16,000 × *g* for 15 min at 4 °C to generate a soluble S16 fraction. After combining with SDS-PAGE sample buffer, proteins were resolved on a 4-12% gradient SDS-PAGE gel (NuPAGE Novex, Invitrogen) and transferred to a nitrocellulose membrane using an iBlot system (Invitrogen). After probing with a 1:500-fold dilution of the UOK75 anti-TgSkp1 antibody, and 1:10,000-fold dilution of alexa-680 labeled goat anti-rabbit IgG secondary antibody (Invitrogen), blots were imaged on a Li-Cor Odyssey infrared scanner.

Results

Toxoplasma Skp1 is Modified by a Pentasaccharide– Our previous study showed that disruption of exon 1 of the Skp1 prolyl 4-hydroxylase gene (*phyA*) in the parental type 1 strain RHΔ*ku80*Δ*hxgprt* (RHΔΔ) resulted in greater mobility of Skp1 during SDS-PAGE, corresponding to an M_r difference of about 1000 (16). To investigate this possibility that this was due to a loss of hydroxylation-dependent glycosylation as occurs in *Dictyostelium* (10), a previously described antiserum raised against recombinant TgSkp1 (16) was used to immunoprecipitate Skp1 from tachyzoite extracts and its tryptic peptides were analyzed by conventional nanoLC/MS in an LTQ-XL Orbitrap mass spectrometer. Peptides covering 75% of the 170 amino acid sequence of Skp1, including ^{145}IFNIVNDFT**P**EEEAQVR containing unmodified Pro154 (*m*/*z* 1011, $[M+2H]^{2+}$) eluting at 88.7 min, were observed. Potential hydroxylated glycopeptides were sought using a theoretical mass list of glycoforms of this peptide containing any combination of 1-8 residues of Hex, dHex, HexNAc, HexUA, and Pent. This search yielded, at 82.7 min elution time, a single

glycopeptide ($[M+2H]^{2+}$, 1437.1464; $[M+3H]^{3+}$, 958.09) with an exact match (within 0.56 ppm) to a glycoform containing 1 HexNAc, 3 Hex, and 1 dHex residues (**Fig. 2.1A**). Similar results were obtained for a Skp1-SF preparation isolated by anti-FLAG immunoprecipitation from a strain in which Skp1 was C-terminally modified by an SF-tag (data not shown). The putative glycopeptide ion was absent from the RH*phyAΔ* strain (summarized in **Table 2.2**), consistent with its identity as the predicted Skp1 glycopeptide.

The putative glycopeptide ion was subjected to MS/MS analysis to confirm its composition and characterize its organization. Fragmentation of the doubly-charged parent ion by CID yielded a series of ions whose mass differences corresponded to loss of a combination of Hex, dHex and/or HexNAc residues (**Fig. 2.1B**), resulting ultimately in the expected hydroxypeptide ion. The predominant ions were consistent with the presence of a linear pentasaccharide whose composition from the non-reducing end is Hex-Hex-dHex-Hex-HexNAc-. Fragmentation also yielded a series of singly-charged peptide and glycopeptide fragment b- and y-ions that confirmed the predicted amino acid sequence of the parent ion, and demonstrated the position of the additional O-atom as occurring on Pro (to yield Hyp), and the attachment of all sugars via Pro154 (**Fig. 2.1C**). These data indicate that a fraction of Skp1 in *Toxoplasma* is modified by a linear pentasaccharide reminiscent of the linear pentasaccharide on DdSkp1.

Predicted TgSkp1 Modifying Glycosyltransferases– BLASTP and TBLASTN searches for sequences corresponding to the three GT genes that catalyze formation of the pentasaccharide on DdSkp1 in ToxoDB (V.7.3) yielded high scoring hits for DdGnt1 and DdPgtA. No candidates for a homolog of DdAgtA were detected using these algorithms or PSI- or PHI-BLAST toward either its catalytic or WD40-repeat domains. The *Toxoplasma* Dd*gnt1*-like sequence is represented

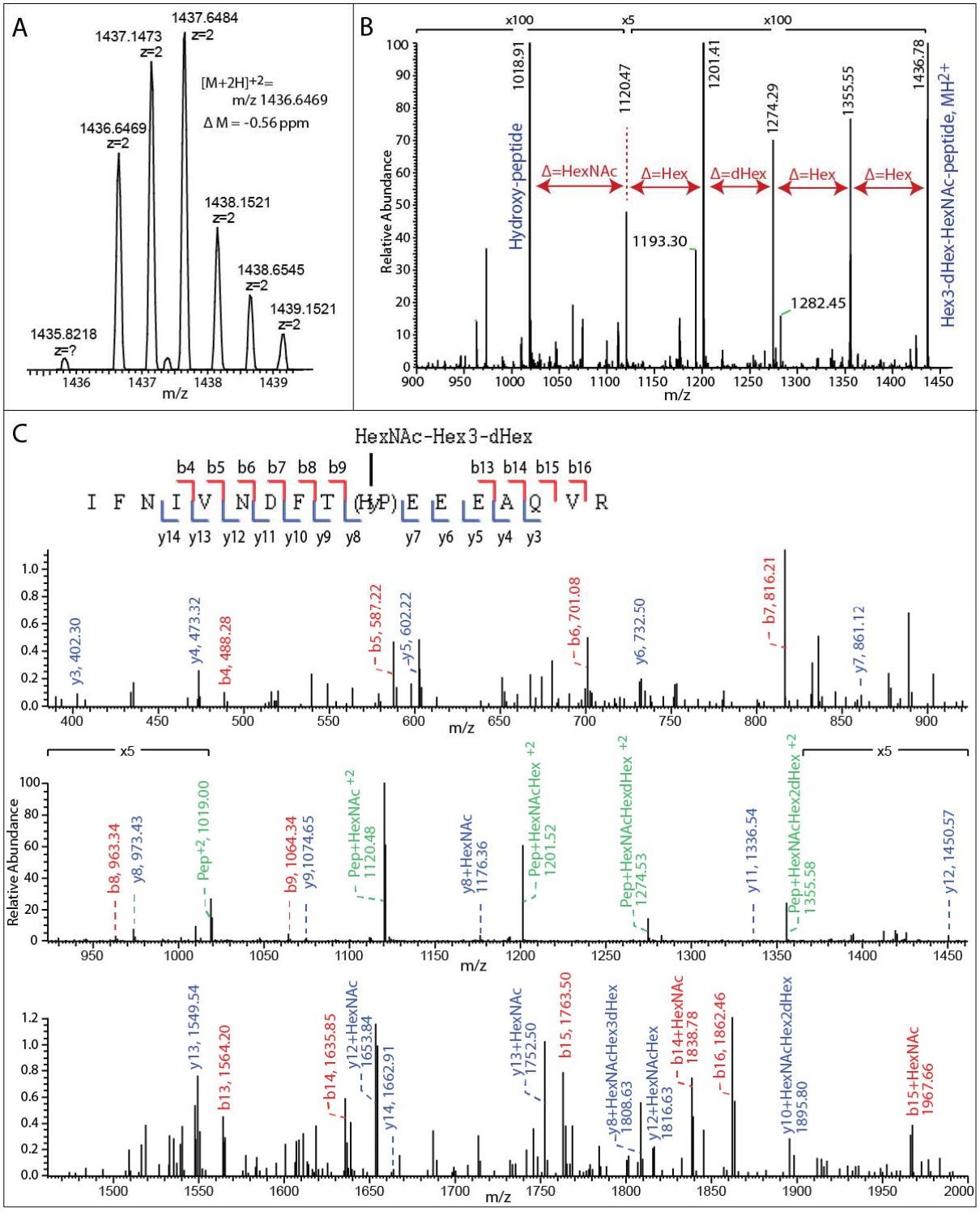


Figure 2.1. Orbitrap MS analysis of the TgSkp1 glycopeptide. RHΔΔ tachyzoites were lysed out of human foreskin fibroblasts (HFFs), urea-solubilized, and immunoprecipitated with bead-bound affinity-purified anti-TgSkp1 (pAb UOK75). The enriched preparation of TgSkp1 was

eluted with triethylamine, reduced and alkylated, trypsinized and analyzed by RP-HPLC on a LTQ-XL Orbitrap MS. Extracted ion chromatograms showed co-elution of a doubly-charged (*m*/*z* 1436.6464) and a triply-charged ion (*m*/*z* 958.0983) corresponding, with a delta mass of 0.56 ppm, to the predicted tryptic TgSkp1 peptide ^{145}IFNIVNDFT(HyP)EEEAQVR^{161} bearing a pentasaccharide with composition Hex3dHex1HexNAc1 (*A*). *B*, CID fragmentation of the doubly-charged precursor ion yields a sequential loss of monosaccharide residues corresponding to Hex, Hex, dHex, Hex and HexNAc, indicating the presence of a linear pentasaccharide. *C*, Inspection of the full CID fragmentation spectrum shows b- (blue annotations) and y- (red annotations) ion series that match the predicted peptide sequence, as illustrated in the inset, and demonstrate that the glycan is linked via a hydroxylated derivative of Pro154. Peptides with residual sugars are annotated in green.

Table 2.2. MS detection of Skp1 glycopeptides in strains. Isoforms of the Skp1 peptide ^{145}IFNIVNDFTPEEEAQVR were detected as described in Fig. 2.1. The distribution of raw ion counts among the detected isoforms are shown for the strains analyzed.

Strain	Expt #	Unmodified peptide			Hyp peptide			Hyp-HexNAc peptide			Hyp-penta-saccharide peptide		
		Occu-pancy*	*m/z***	ΔM (ppm)	Occu-pancy*	*m/z***	ΔM (ppm)	Occu-pancy*	*m/z***	ΔM (ppm)	Occu-pancy*	*m/z***	ΔM (ppm)
RHΔΔ	1	73%	1011.0002	-2.18	0%			0%			27%	1436.6464	-0.92
RHΔΔ	2	64%	1011.0052	2.77	0%			0%			35%	1436.6538	4.23
RHΔΔ/SF	1	78%	1011.0012	-1.19	0%			0%			22%	1436.6456	-1.48
RHphyAΔ-1	1	100%	1011.0018	-0.59	0%			0%			0%		
RHphyAΔ-1/SF	1	100%	1011.0044	1.98	0%			0%			0%		
RHgnt1Δ	1	60%	1010.9994	-2.97	40%	1018.9995	-0.35	0%			0%		
RHgnt1Δ	2	55%	1011.0024	0.00	45%	1019.0004	0.53	0%			0%		
RHpgtAΔ	1	72%	1011.0031	0.69	0%			28%	1120.5431	3.19	0%		

* apparent occupancy based on raw spectral counts.
** values refer to $[M+2H]^{2+}$ ions.

by a one-exon gene model (**Fig. 2.2A**) in three sequenced strains of *Toxoplasma* (GT1, Me49 and Veg). The GT1 (type 1 strain) sequence (TGGT1_315885) exhibited 42% identity and 67% similarity to DdGnt1 over 214 amino acids of the ~250 amino acid catalytic domain (**Supplemental Fig. S2.1**). Like DdGnt1, TGGT1_315885 is predicted to be a cytoplasmic protein owing to the absence of detectable membrane or nuclear targeting motifs. However, at 1510 amino acids, TGGT1_315885 is substantially longer than DdGnt1, 423 amino acids. As revealed by the amino acid sequence alignment (**Supplemental Fig. S2.1**), and illustrated in **Fig. 2.2C**, this results from multiple insertions throughout the length of the protein, a common occurrence in *Toxoplasma* genes as observed, e.g., in TgIF2Kb (153). In addition, a C-terminal sequence referred to as domain A that lies outside of the predicted catalytic domain (146) but is required for DdGnt1 activity is weakly conserved across these predicted proteins (**Fig. 2.2**). The remaining intervening sequences are poorly conserved even among coccidian apicomplexans that have PhyA, Gnt1- and PgtA-like sequences (**Supplemental Fig. S2.1**). The coccidian *Sarcosystis neurona* is predicted to contain even longer insert sequences. In comparison, sequences from *Chromera velia* and *Vitrella brassicaformis*, representatives of the closest known photosynthetic relatives of apicomplexans in the alveolate superphylum (154), largely lack these inserts and more resemble the length of the *Dictyostelium* sequence. While the TgGnt1 coding region remains to be confirmed experimentally, two of the inserts are present in EST sequences derived from mRNA, and a third was detected in a shotgun proteomics screen (**Fig. 2.2C**; **Fig. S2.1**). While the significance of these additional sequences is enigmatic, their low conservation suggests that they fulfill *Toxoplasma*-specific functions that are unlikely to be related to the proposed enzymatic activity.

The *Toxoplasma* Dd*pgtA*-like sequence is represented by a 14-exon gene model (**Fig. 2.2B**). The predicted amino acid sequence of TGGT1_260650, from the type I GT1 strain, is 98%

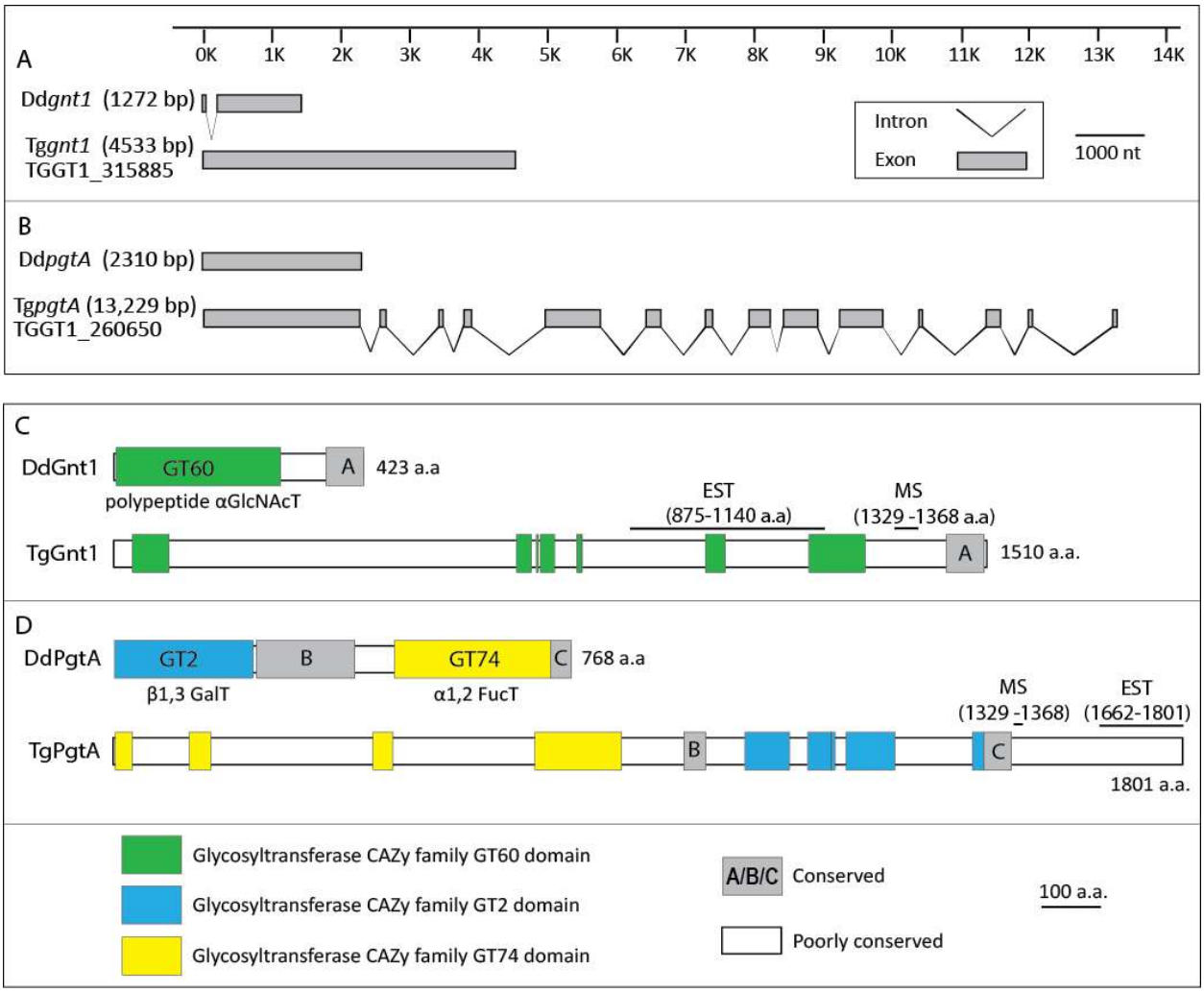


Figure 2.2. Comparative genomic and domain organization of Gnt1 and PgtA from *Dictyostelium* and *Toxoplasma.* *A, B*, exon-intron organization of the *gnt1* and *pgtA* genes. *gnt1* (A) and *pgtA* (*B*) gene models from *Dictyostelium discoideum* are from refs. 8 and 9 and available at dictybase.org. The gene models from *Toxoplasma gondii* (GT1 [type 1] strain) are from http://toxodb.org/toxo/. The length from the start codon to the stop codon in nt is in parentheses. *C, D*, domain organization of Gnt1 (*C*) and PgtA (*D*) proteins. Gnt1- and PgtA-like sequences from *Toxoplasma* and 3 other coccidian apicomplexans were aligned with corresponding sequences from *Dictyostelium discoideum* and 3 other amoebozoa, and sequences from 2 chromerid alveolates, as shown in Figs. S1 and Fig S2. Regions of high conservation among all 10 sequences are shown in color for the glycosyltransferase-like sequences and gray for non-GT-like sequences. Toxoplasma sequences whose expression has been confirmed at the transcriptional (EST) or proteomic (MS) level, as reported at http://toxodb.org/toxo/, are indicated. Diagrams are shown to scale.

and 97% identical to that of the type II ME49 and Type III VEG strain sequences, and the protein is predicted to be cytoplasmic. The *Toxoplasma* candidate is 1801 amino acids long, compared to the 768 amino acid length of DdPgtA, and the order of the two putative glycosyltransferase domains is reversed **(Fig. 2.2D**). The N-terminal CAZy GT2 family sequence of DdPgtA, which encodes a β3-GalT activity, has its sequence homolog in the C-terminal half of the *Toxoplasma* protein, whereas the C-terminal CAZy GT74 family sequence of DdPgtA, which encodes an α2-FucT activity, has its homolog in the N-terminal half of the *Toxoplasma* protein. The GT2-like domain of the predicted TgPgtA is 22% identical and 60% similar to the DdPgtA-βGalT domain over the 191 most conserved amino acids, and the GT74-like domain is 33% identical and 66% similar to the DdPgtA-αFucT domain over 190 amino acids. The amino acid sequence alignment (**Supplemental Fig. S2.2**) of the PgtA-like sequences reveal that, as for DdGnt1, the predicted apicomplexan PgtA-like sequences have numerous multiple inserts relative to the *Dictyostelium* prototype, as illustrated in **Fig. 2.2D**. These insert sequences, several of which occur in the protein based on EST and MS data (**Fig. 2.2D**), tend to diverge even within the coccidian apicomplexans, and are minimal in the chromerids. Thus they are unlikely to be critical for enzymatic activity.

Tggnt1 and TgpgtA are required for TgSkp1 glycosylation– To determine whether Tg*gnt1* and Tg*pgtA* are involved in TgSkp1 glycosylation, their genes were disrupted by double-crossover homologous recombination in the RHΔΔ strain, as described in "Experimental Procedures" and illustrated in **Fig. 2.3A** for Gnt1. Deletion of *gnt1* in recovered clones was demonstrated by loss of a PCR product for *gnt1*-coding DNA, and positive PCR reactions for the insertion of the selection marker *hxgprt* between *gnt1* flanking sequences, as described in **Fig. 2.3B**. To control for off-target genetic modifications, a *gnt1*-disruption clone was complemented with a version of

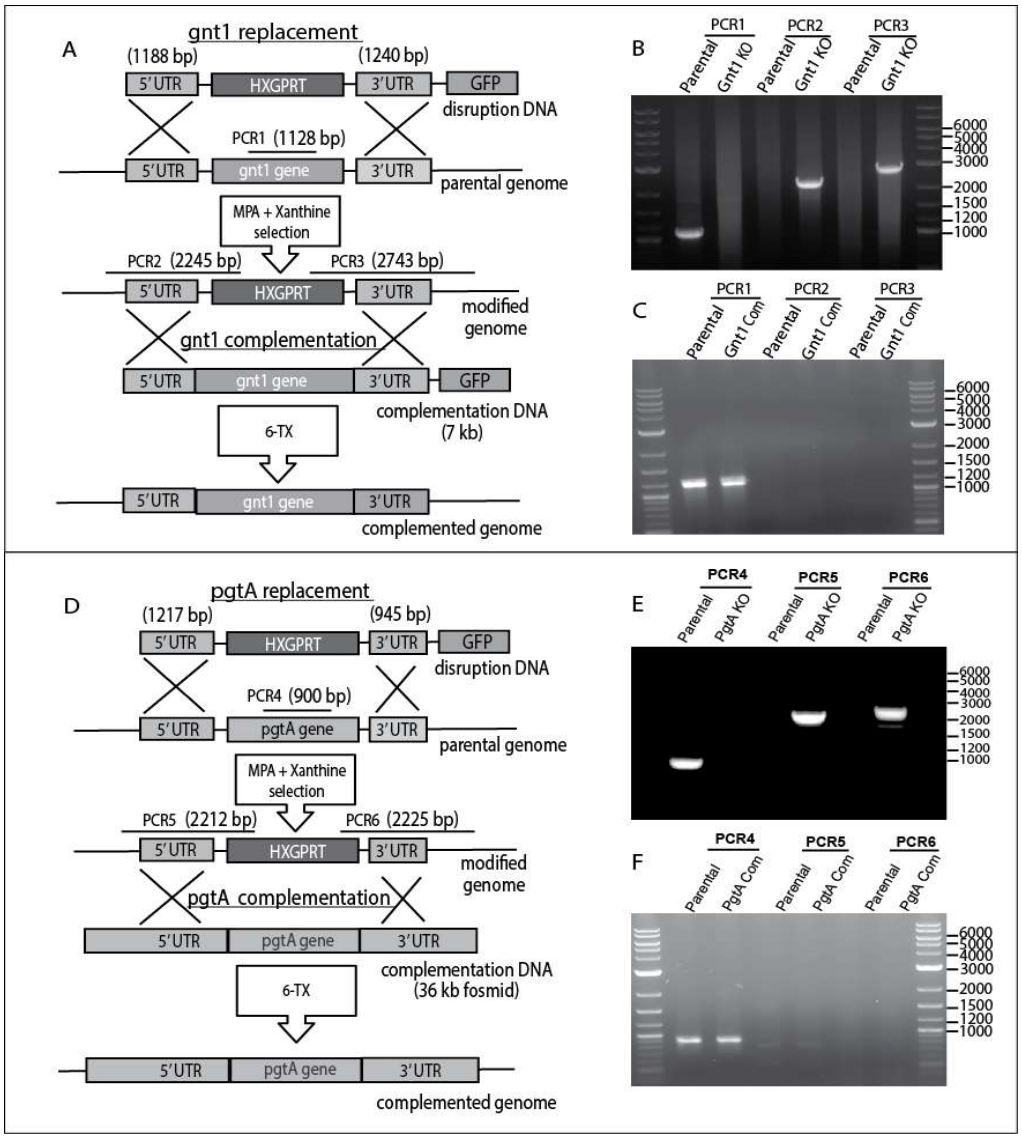


Figure 2.3. Disruption and complementation of Tg*gnt1* and Tg*pgtA*. *A*, strategy for deletion of Tg*gnt1* and its subsequent complementation. The plasmid-derived disruption DNA with homologous targeting sequences was electroporated into parasites. Recovery of *hxgprt*-positive clones that were resistant to MPA and xanthine and were GFP-negative were candidates for double crossover gene replacement. *B*, gene replacement was confirmed by PCR-1, which demonstrated loss of *gnt1* coding DNA, and PCR-2 and -3 which demonstrated that the inserted *hxgprt* DNA was flanked by neighboring *gnt1* DNA. To complement Tg*gnt1* in the disruption strain, a plasmid containing a ~7 kb genomic locus including Tg*gnt1* coding region and 5'- and 3'- untranslated regions (*A*) was transfected. Complemented strains where the *hxgprt* is replaced by Tg*gnt1* locus were counter selected under 6-thioxanthine. *C*, *gnt1* replacement was confirmed by the positive PCR-1, and negative reactions for PCR-2 and PCR-3 which depended on the presence of *hxgprt*. *D-F*, Tg*pgtA* was similarly targeted for disruption and complementation. Strain characteristics are summarized in **Table 2.1.**

the original disruption DNA in which *hxgprt* was replaced by the deleted coding region, and counter-selected for loss of *hxgprt* (**Fig. 2.3A**). The same set of PCRs was used to confirm the desired gene restoration in clonal isolates (**Fig. 2.3C**).

The effect of *gnt1*-deletion on Skp1 glycosylation was evaluated initially by SDS-PAGE and Western blotting. As shown in **Fig. 2.4A**, Skp1 from parasites lacking *gnt1* (lane 3) migrated more rapidly than wild-type Skp1 (lane 1), and similarly to Skp1 from parasites whose *phyA* had been disrupted (lane 2). MS searches for the Skp1 glycopeptide were negative, but a novel hydroxypeptide corresponding to the hydroxylated but non-glycosylated Skp1 was obtained, in addition to the unmodified peptide (**Table 2.2**). Analysis of the complemented strain revealed that normal mobility of Skp1 in SDS-PAGE was at least partially restored. Thus Skp1 HexNAcylation depends on Gnt1 and, by analogy with the *Dictyostelium* example, Gnt1 is expected to directly catalyze addition of the first sugar in α-linkage onto Hyp154.

A similar analysis was performed on PgtA, whose genetic locus was manipulated as described in **Fig. 2.3D-F**. In this case, a recently prepared genomic fosmid (148) was utilized to restore the genomic locus in the *pgtA*-disruption clone. Genomic DNA was utilized owing to the large number of predicted introns and our difficulty in isolating a full-length cDNA using RT-PCR (data not shown), and the fosmid clone was used owing to the predicted large size (13.2 kb) of its genomic locus. As shown in **Fig. 2.4**, lane 7, Skp1 glycosylation also appeared to be affected by the loss of *pgtA* based on SDS-PAGE/Western blotting, and this was confirmed by the accumulation of the HexNAc form of Skp1, the expected acceptor substrate of PgtA, based on MS analysis of tryptic peptides (**Table 2.2**). As expected, restoration of the *pgtA* locus resulted in at least partial recovery of Skp1 glycosylation (**Fig. 2.4**, lane 8). At a minimum, PgtA is thus required for addition of the second sugar to the Skp1 glycan.

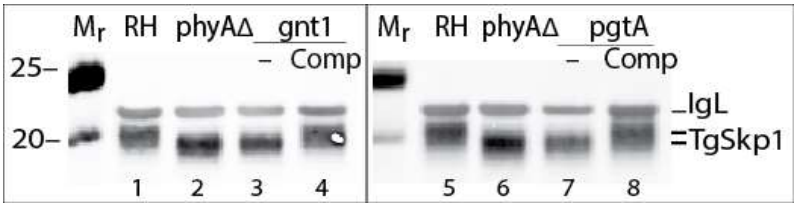


Figure 2.4. Disruption of Tg*gnt1* or Tg*pgtA* affects TgSkp1 glycosylation. Soluble S16 fractions from equivalent numbers (3×10^6 cells) of parental RHΔΔ (RH) and RH*phyA*Δ-1 (*phyA*$^-$), RH*gnt1*Δ, RH*pgtA*Δ and their complemented cells were resolved by 4-12% SDS-PAGE, electroblotted, and probed using anti-TgSkp1 (UOK75) antiserum. Changes in glycosylation inferred from altered gel mobility were confirmed by mass spectrometry (Table 2). Similar results were obtained for independently derived clones of RH*gnt1*Δ and RH*pgtA*Δ.

TgGnt1 has properties of a Skp1 polypeptide UDP-GlcNAc:HO-Skp1 GlcNAcT– A previous study detected a Gnt1-like activity in tachyzoite cytosolic extracts, based on transfer of ^{3}H from UDP-[^{3}H]GlcNAc to $DdHis_{10}$-Skp1 that was recovered from an SDS-PAGE gel (141). This activity depended on addition of DdPhyA indicating dependence on Hyp. Replication of an optimized form of this assay (see Experimental Procedures) revealed time-dependent transfer of ^{3}H to HO-DdSkp1, that was absent from *gnt1*Δ extracts (**Fig. 2.5A**). As shown in **Fig. 2.5B**, no incorporation into endogenous proteins was detected based on analysis of an entire SDS-PAGE gel lane, indicating absence of activity of endogenous GTs from other sources, such as the Golgi, that modify other targets in these cytosolic preparations. The lack of other radiolabeled proteins is inconsistent with the existence of an intermediate TgGnt1 substrate that itself mediates modification of Skp1. Incorporation was reduced as expected after addition of a 10-fold excess of unlabeled UDP-GlcNAc, but not of unlabeled UDP-GalNAc (**Fig. 2.5C**), indicating that the enzyme is selective for the GlcNAc isomer. Since a homolog of Gnt1 that resides in the Golgi transfers GalNAc to proteins (155), and *Toxoplasma* possesses an epimerase that can interconvert UDP-GlcNAc with UDP-GalNAc, the nature of the transferred ^{3}H was confirmed by another method. ^{3}H was found to be incorporated as GlcNAc, based on co-chromatography of ^{3}H released by HCl hydrolysis, which de-N-acetylates GlcNAc to $GlcNH_2$, with a $GlcNH_2$ standard (**Fig. 2.5D**). Although the evidence that TgGnt1 modifies Skp1 in this assay is indirect, its homology with DdGnt1, whose purified recombinant version can αGlcNAcylate HO-DdSkp1 *in vitro* (107), suggests that TgGnt1 also directly αGlcNAcylates TgSkp1.

TgPgtA is a bifunctional glycosyltransferase with GalT and FucT activities– To characterize the role of *pgtA* in extending the Skp1 glycan, the above assay was first modified by

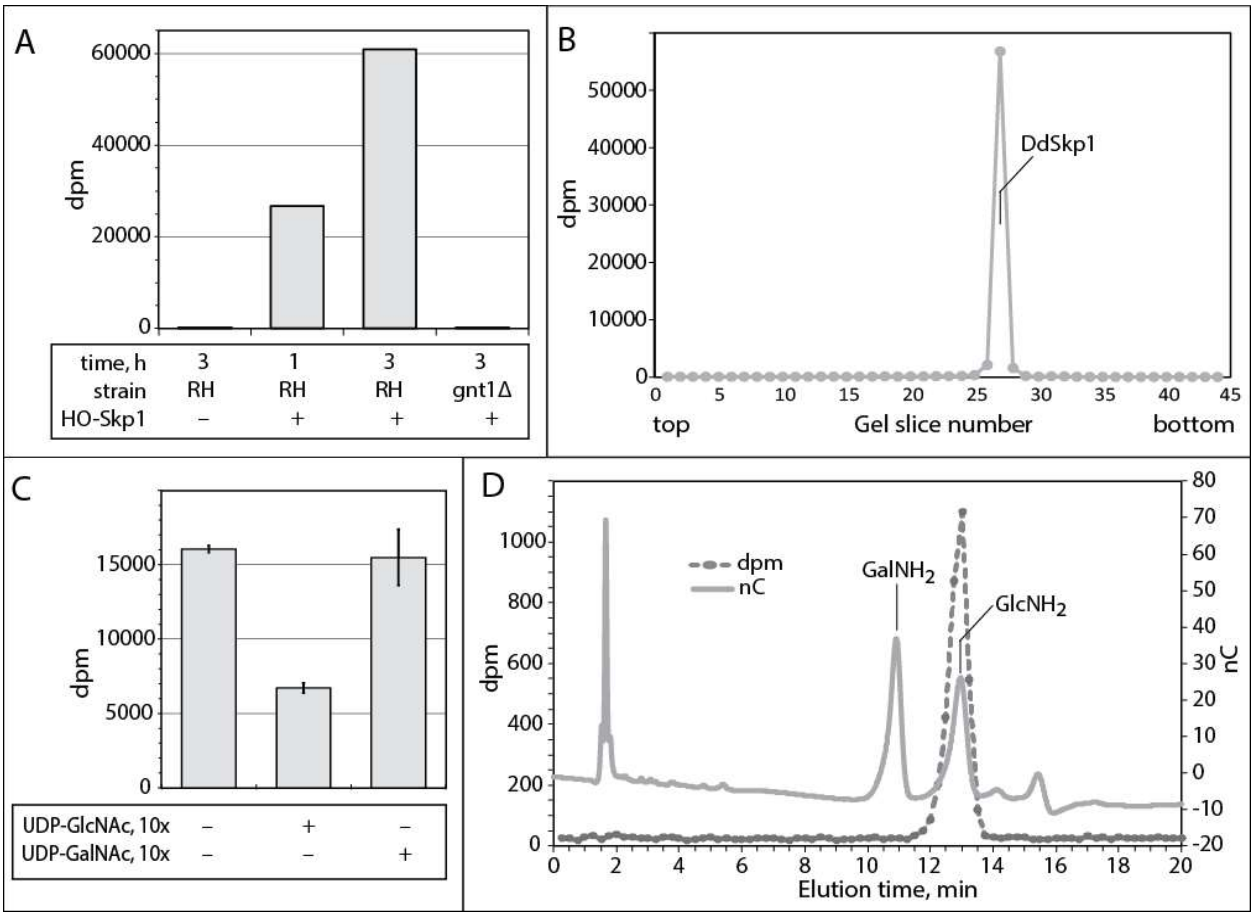


Figure 2.5. TgGnt1 is a Skp1 GlcNAcT. *A*, GlcNAcT activity in S100 cytosolic parasite was assayed based on transfer of [^{3}H] from 0.5 µM UDP-[^{3}H]GlcNAc to exogenous HO-DdSkp1 for 1-3 h as described under "Experimental Procedures". Reactions were loaded onto and separated on an SDS-PAGE gel, and the Coomassie blue stained DdSkp1 bands were excised and subjected to liquid scintillation spectroscopy. The reaction time, presence of HO-Skp1, and source of the extract (RHΔΔ or RH, or RH*gnt1*Δ or *gnt1*Δ) were varied as indicated. *B*, the entire lane from a parallel 3-h reaction (RH, +HO-Skp1) from panel A was analyzed for incorporation of ^{3}H. Incorporation was only detected at the migration position of DdSkp1. *C*, the donor substrate specificity of the GlcNAcT activity was examined by including a 9-fold excess of unlabeled UDP-GlcNAc or UDP-GalNAc to reactions containing 10 µM UDP-[^{3}H]GlcNAc. Incorporation was measured as in panel A. Error bars show standard deviations of the mean of 2 replicates from each of two independent reactions. *D*, analysis of incorporated ^{3}H. The reacted Skp1 band was excised from a PVDF membrane electroblot of the SDS-PAGE gel and subjected to acid hydrolysis in 6 N HCl, and analyzed by high pH anion-exchange chromatography. The hydrolysate was supplemented with $GlcNH_2$ and $GalNH_2$ and chromatographed on a Dionex PA-1 column. Elution of the sugar standards was monitored by a pulsed amperometric detector (nC), and fractions were collected to monitor the elution of ^{3}H by scintillation counting (dpm).

substituting UDP-[^{3}H]Gal for UDP-[^{3}H]GlcNAc, and GlcNAc-Skp1 for HO-Skp1. Time-dependent incorporation of ^{3}H into GlcNAc-Skp1 was observed (**Fig. 2.6A**). Incorporation into the Skp1 band on the SDS-PAGE gel required the inclusion of Skp1, and Skp1 was the only protein that incorporated detectable radioactivity (**Fig. 2.6B**). No incorporation was detected in extracts of *pgtA*Δ cells. Recovery and analysis of the ^{3}H after acid hydrolysis confirmed incorporation as Gal rather than a derivative (**Fig. 2.6C**). Similar findings were observed in a corresponding FucT assay, in which UDP-[^{3}H]Gal was replaced by GDP-[^{3}H]Fuc (**Fig. 2.6D-F**). However, incorporation of ^{3}H depended on the inclusion of UDP-Gal (unlabeled), in contrast to the GalT reaction which did not require GDP-Fuc (**Fig. 2.6A**). This indicated that TgPgtA is, like DdPgtA (109), a processive diglycosyltransferase that catalyzes the sequential addition of Gal and then Fuc, an order consistent with the MS-MS data. Although >95% of incorporation of [^{3}H]Fuc into the Skp1 band depended on Skp1 and *pgtA* (Fig. 6D, E), residual incorporation was observed at this position and elsewhere in the gel indicative of additional *pgtA*-and Skp1-independent FucT activity in the extract. These results are consistent with accumulation of GlcNAc-Skp1, and absence of higher glycosylation states, in *pgtA*Δ cells (**Fig. 2.4** and **Table 2.2**).

TgGnt1 and TgPgtA are important for Toxoplasma growth in cell culture– Previous studies revealed that disruption of exon 1 of *phyA* results in a parasite growth defect, which could be detected as reduced plaque areas after 5 d of replication on a fibroblast monolayer (16). To check that no residual *phyA* activity was present, all 9 exons were deleted (RH*phyA*Δ-2). A similarly reduced ability to grow on monolayers was observed, as illustrated in **Fig. 2.7A** and

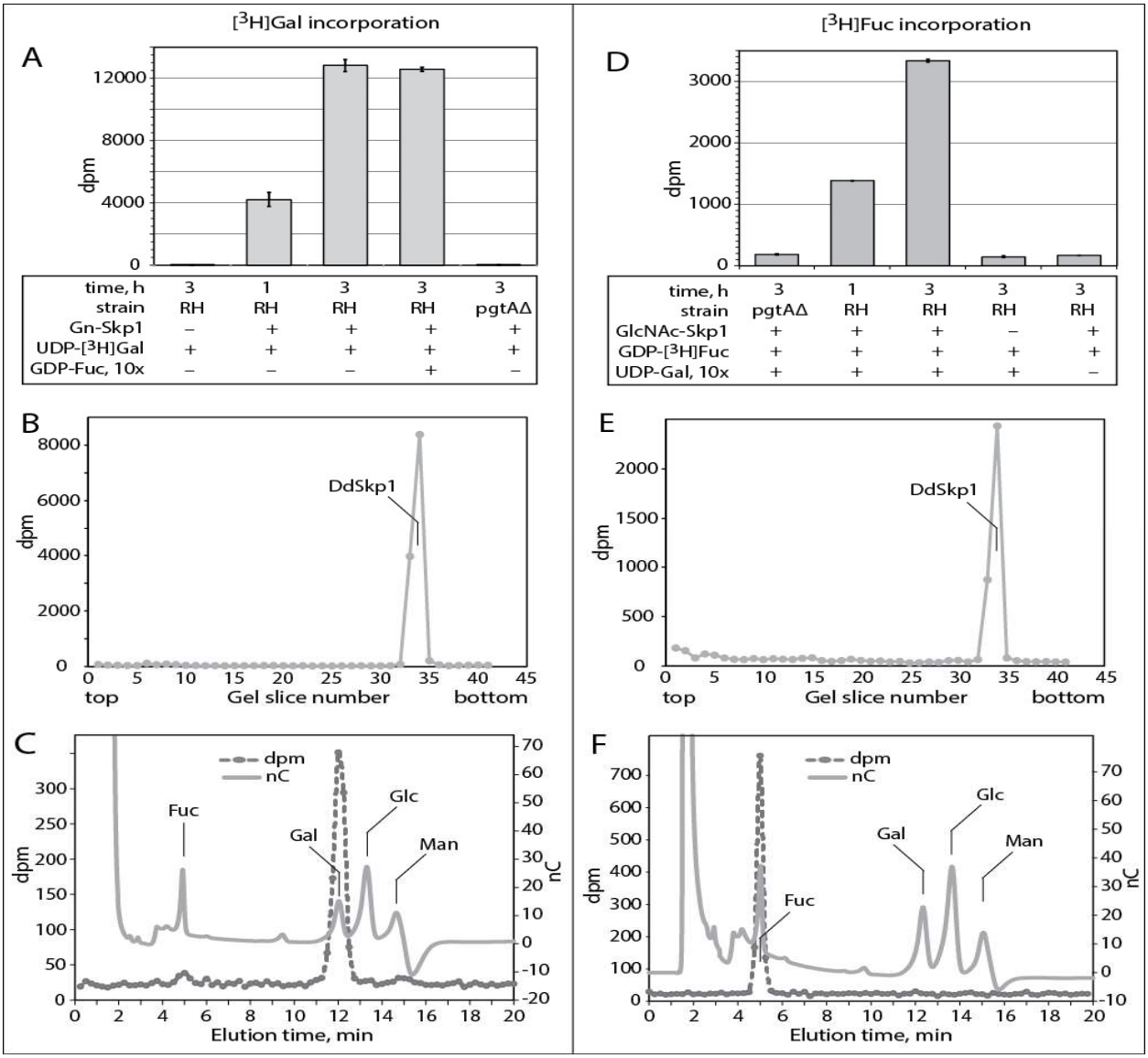


Figure 2.6. PgtA is a Skp1 GalT and FucT. *A*, GalT activity directed toward GlcNAc-Skp1 was assayed as described for GlcNAcT activity in Fig. 5A, except that GlcNAc-Skp1 and UDP-[^{3}H]Gal were used in place of HO-Skp1 and UDP-[^{3}H]GlcNAc. The reaction time, inclusion of GlcNAc-Skp1 and GDP-Fuc, and source of the extract (RHΔΔ or RH, or RH*pgtA*Δ or *pgtA*Δ), were varied as indicated. *B*, the entire lane from a parallel 3-h reaction (RH, +Gn-Skp1) from panel A was analyzed for incorporation of ^{3}H. Incorporation was only detected at the migration position of DdSkp1. *C*, ^{3}H-DdSkp1 from the 3-h GalT reaction in Panel A was isolated as in Fig. 5D and hydrolyzed in 4 M TFA. The hydrolysate was chromatographed on a Dionex PA-1 column with internal standards of Gal, Glc, Man and Fuc, and the elution of ^{3}H was monitored by scintillation counting of collected fractions. *D-F*, FucT activity assays. Reactions were conducted as above except that GDP-[^{3}H]Fuc replaced UDP-[^{3}H]Gal, and the dependence of incorporation on a 10-fold concentration excess of UDP-Gal, GlcNAc-Skp1, PgtA in the extract and time was examined.

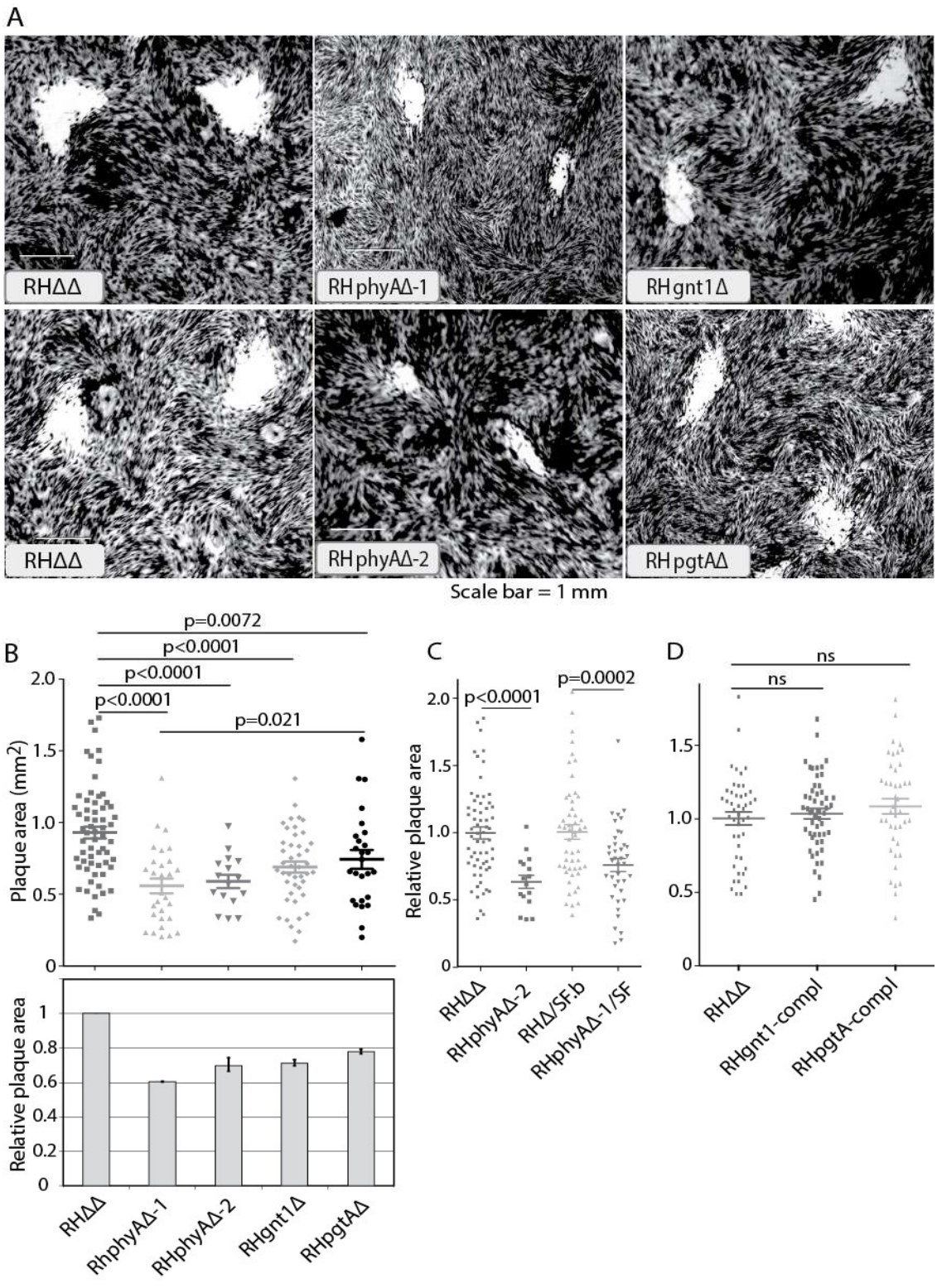

A
RHΔΔ
RHphyAΔ-1
RHgnt1Δ
RHΔΔ
RHphyAΔ-2
RHpgtAΔ
Scale bar = 1 mm
B
p=0.0072
p<0.0001
p<0.0001
p<0.0001
p=0.021
Plaque area (mm2)
2.0
1.5
1.0
0.5
0
Relative plaque area
1
0.8
0.6
0.4
0.2
0
RHΔΔ
RhphyAΔ-1
RHphyAΔ-2
RHgnt1Δ
RHpgtAΔ
C
p<0.0001
p=0.0002
Relative plaque area
2.0
1.5
1.0
0.5
0
RHΔΔ
RHphyAΔ-2
RHΔ/SF.b
RHphyAΔ-1/SF
D
ns
ns
1.5
1.0
0.5
0
RHΔΔ
RHgnt1-compl
RHpgtA-compl

Figure 2.7. Role of Tg*gnt1* and Tg*pgtA* in parasite proliferation. HFF monolayers were inoculated with freshly isolated tachyzoite stage parasites at a multiplicity of infection of 0.002. After 5.5 days, monolayers were stained with crystal violet. *A*, representative images of cleared areas of the host monolayers. *B-D*, the images digitized and plaque areas were calculated. The dot plots show the area distributions and average values ± S.E.M. from a representative of 2 independent experiments. Average parental strain areas ranged from 0.5-1.0 mm^2. *p* values for statistical significance of the differences, based on a one-way ANOVA test, are shown above. ns=not significant. *B*, data from RHΔΔ, RH*phyA*Δ-1 and RH*phyA*Δ-2, generated by different strategies, RH*gnt1*Δ, and RH*pgtA*Δ strains. Bar graph shows average (±S.D.) from 2 independent experiments. *C*, data from strains in which Skp1 was SF-tagged. *D*, data from Tg*gnt1* or Tg*pgtA* complemented (compl) strains.

quantified in **Fig. 2.7B**. Furthermore, deletion of exon 1 in a strain in which Skp1 was C-terminally modified with an SF-epitope tag, which itself did not affect growth, also resulted in slowed growth (**Fig. 2.7C**). To examine the roles *gnt1* and *pgtA*, the plaque-forming abilities of the disruption strains described above were analyzed. As shown in **Fig. 2.7B**, *gnt1*Δ cells exhibited slow growth that was statistically indistinguishable from that of *phyA*Δ cells. *pgtA*Δ cells also exhibited a slow growth phenotype, which was intermediate between that of *gnt1*Δ and parental (RHΔΔ) cells. Complementation of *gnt1* and *pgtA* by gene replacement at their original loci restored normal growth (**Fig. 2.7D**), showing that the growth differences in the original disruption strains could be attributed to the GT targets. Although these findings do not demonstrate directly that the altered modifications of Skp1 are involved in reduced growth, the finding that 3 independent enzymes that share Skp1 as a target substrate exhibit a similar deficiency is consistent with a role for Skp1, a model that has both biochemical and genetic support in *Dictyostelium*.

Discussion

Skp1 isolated from the tachyzoite stage of *Toxoplasma* is partially modified by a glycan chain that consists of 5 sugars and is linked to the hydroxylated form of Pro154. The monosaccharides are organized as a linear pentasaccharide with a sequence, reading from the peripheral non-reducing end, of Hex-Hex-αFuc-βGal-αGlcNAc-, based on mass spectrometry and characterization of the core GTs. Remarkably, these properties match those of the glycan that was previously characterized on Skp1 from an unrelated protist, the social amoeba *Dictyostelium* (10). Genetic disruption of glycosylation rendered a growth defect in host cell monolayers, implicating a role for *Toxoplasma* E3^{SCF}ubiquitin ligases in cell proliferation.

The predicted Skp1 glycan was analyzed at the glycopeptide level owing to the resistance

of the glycan-Hyp linkage to known methods of cleavage. Native Skp1 from parasites was analyzed to avoid overexpression artifacts, but this necessitated application of highly sensitive methods due to the limited amount of material available since *Toxoplasma* can only be grown intracellularly. As is typical for the analysis of glycopeptides by mass spectrometry, their detection required manual inspection of spectra for candidate ions, owing to suppression and abundance issues. The pentasaccharide-peptide was detected in an exhaustive search of the primary spectra for tryptic peptides bearing Pro154 (**Fig. 2.1**), the previously documented hydroxylation site (16), and any combination of up to 8 monosaccharides. Decomposition analysis in the gas phase confirmed the glycan's location at Hyp154 and suggested that it is organized as a linear pentasaccharide with a sequence, from the non-reducing end, of Hex-Hex-dHex-Hex-HexNAc-. The only other peptide isoform detected was non-hydroxylated and therefore non-glycosylated (**Table 2.2**), suggesting limited glycosylation microheterogeneity.

The sugar identities were investigated by highly sensitive incorporation of radioactive sugars mediated by parasite extracts that harbor the biosynthetic enzymes. Cytosolic extracts were observed to specifically incorporate radioactivity from three different radioactive sugar nucleotides into recombinantly generated isoforms of *Dictyostelium* Skp1, which was previously shown to be an excellent substrate for DdPhyA (16). Based on the confirmed identities of the ^{3}H-sugars after incorporation, the core trisaccharide is concluded to consist of Fuc-Gal-GlcNAc, which matches the dHex-Hex-HexNAc- found by MS.

Incorporation of ^{3}H-GlcNAc into HO-Skp1 depended on Tg*gnt1* based on absence of activity in *gnt1*Δ extracts (**Fig. 2.5**). By analogy with its *Dictyostelium* ortholog, TgGnt1 is inferred to transfer GlcNAc in an α-linkage to Hyp154 of TgSkp1. Incorporation of ^{3}H-Gal depended on Tg*pgtA* (**Fig. 2.6**), and selective sensitivity to β3-galactosidase[1] indicates that, as for the

Dictyostelium precedent, Gal is β3-linked to the underlying GlcNAc. The addition of ^{3}H-Fuc, which also depended upon Tg*pgtA*, required prior addition of Gal (**Fig.2. 6**), but the nature of the linkage requires further study to determine whether Fuc is α2-linked as observed in *Dictyostelium*. Thus, although the physical order of the two GT domains is reversed relative to DdPgtA (**Fig. 2.2**), the order of addition of the two sugars is conserved. The existence of the two terminal Hex residues was unexpected owing to the absence of a homolog for AgtA in the *Toxoplasma* genome, and their identities are under current investigation.

Both Tg*gnt1* and Tg*pgtA* were important for efficient plaque-forming ability in tissue culture monolayers cultivated under standard conditions (**Fig. 2.7**). Loss of *gnt1* was as severe as loss of *phyA*, the prolyl 4-hydroxylase that is required for Gnt1 action on Skp1. Loss of *pgtA* resulted in an intermediate effect, i.e., plaque-forming ability was improved relative to cells lacking *phyA*. These effects were specific for the GT genes as normal plaque forming ability was restored upon genetic complementation to the original genotype. Thus, Gnt1 has a relatively minor contribution to growth, with the remainder owing to Gnt1-dependent glycosylation contributed by, at least on Skp1, PgtA and/or the additional unknown GTs. In *Dictyostelium*, biochemical studies indicate that Skp1 is the only substrate for the orthologs of these GTs, and disruption of either *gnt1* or *pgtA* inhibits development in a way that is related to, but less severe than disruption of *phyA* (146). Gene dosage manipulations on Skp1 expression show inverse effects on development consistent with the modification genes acting via Skp1 in this organism, a model that is supported by point mutations that remove the target Pro residue, and from a double mutant between a GT gene and one of two Skp1 genes (156). By analogy, and based on similar phenotypes of Tg*phyA*Δ and Tg*gnt1*Δ, we propose that the modification genes render their effects in a common pathway that affects Skp1. Further studies on Skp1 itself are needed to evaluate this possibility.

The reduced plaque areas are consistent with a role for Skp1 modification in cell cycle progression as demonstrated in yeast and mammalian cells, where SCF-type E3 ubiquitin ligases are important for signaling proteasome-dependent turnover of cell cycle kinase inhibitors (157). However, SCF ligases represent a family of enzymes that includes many F-box proteins with distinct substrate receptor activities (158). Given the large variety of substrates known to be recognized by yeast, plant and human F-box proteins, further studies are needed to evaluate whether other processes required for plaque formation, such as binding of the parasite to host cells, ingress and/or egress are affected. In future studies, it will be interesting to evaluate whether any of these potential mechanisms are more severely affected in low O_2 or altered metabolic states that are anticipated to influence Skp1 modification enzyme activities in cells (147).

PhyA-, Gnt1-, and PgtA-like sequences are selectively conserved across a broad spectrum of unicellular eukaryotes, including representatives of all major protist clades (10). Validation of their shared functions in both *Toxoplasma* and *Dictyostelium*, which are highly diverged, suggests that Skp1 hydroxylation and glycosylation occurred in ancestral eukaryotes prior to their loss in fungi, higher plants, animals, and select protists. The relatively rapid evolution of F-box proteins, with which Skp1 partners for many if not all of its functions, suggests that Skp1 modifications serve an outsized role in environmental regulation of unique lineage and species specific functions in many unicellular organisms.

[1] Rahman, K, Blader, I. J., and West, C. M., unpublished data.

Supplemental TABLE S2.1. Oligonucleotide sequences employed

Targeting sequence amplification:

Tg*phyA*-disruption 2

a) PhyAF1 5'-flank 5'-end 5'-GGGGTACCCGTTGGTGGCAGGAGAAT (Kpn1)
a') PhyAR1 5'-flank 3'-end 5'-GGAAGCTTTGGCAAGGAAAGCCACAA (HindIII)
b) PhyAF2 3'-flank 5'-end 5'-GCTCTAGAGTTCGGTCAAATCTGTCGTTATTT(XbaI)
b') PhyAR2 3'-flank 3'-end 5'-GCGCGGCCGCGTCGAGGTTCAGCAGACTTT (NotI)

Tg*gnt1*-disruption and complementation

c) GntF1 5'-flank 5'-end 5'-GGGGTACCCAGCCTCACACAGACGAAA (KpnI)
c') GntR1 5'-flank 3'-end 5'-GGAAGCTTCTGTTGAATCGCCGAGAAATG (HindIII)
d) GntF2 3'-flank 5'-end 5'-GCTCTAGAGTTCGGTCAAATCTGTCGTTATTT (XbaI)
d') GntR2 3'-flank 3'-end 5'-GCGCGGCCGCGTCGAGGTTCAGCAGACTTT (NotI)

Tg*pgtA*-disruption and complementation

e) PgtF1 5'-flank 5'-end 5'-GGGGTACCGTTAACAGCAGACCGCATTTC (KpnI)
e') PgtR1 5'-flank 3'-end 5'-GGAAGCTTAGAGGAGAAAGGTGAGGAGAA (HindIII)
f) PgtF2 3'-flank 5'-end 5'-GCTCTAGAAAGGATGTCTTGTCGCTGTATC (XbaI)
f') PgtR2 3'-flank 3'-end 5'-GCGCGGCCGCCTGGCATTCCGTTAATCTCTCT (NotI)

Primers used for PCR confirmation of disruption and complementation strains (see Fig. 3.3)

Tg*phyA*-disruption 2

PCR 1:
Forward: 5'- TGCAGGGTCACTTCTGTT
Reverse: 5'- GCAAATGATGTGCAACTCCTC
PCR 2:
Forward: 5'- CTTCGTCGGCTTCAGCATT
Reverse: 5'- TGTGCTGTTGTCCTGGTACGA
PCR 3:
Forward: 5'- AGGACAGAGTTGAGAAGTTGGCGT
Reverse: 5'- TGGTTACCTATGGCAACGG

Tg*gnt1*-disruption

PCR 1:
Forward: 5'- TGAGAGTAGCGAGGTAGATGAG
Reverse: 5'- GAAGGAAGAGCAGGAAGTACAA
PCR 2:
Forward: 5'- GATCGGAAACACACGACAAAC
Reverse: 5'- TGTGCTGTTGTCCTGGTACGA
PCR 3:
Forward: 5'- AGGACAGAGTTGAGAAGTTGGCGT
Reverse: 5'- TCCACGTGGTTGAGAGATTG

Supplemental TABLE S2.1 (Cont.). Oligonucleotide sequences employed

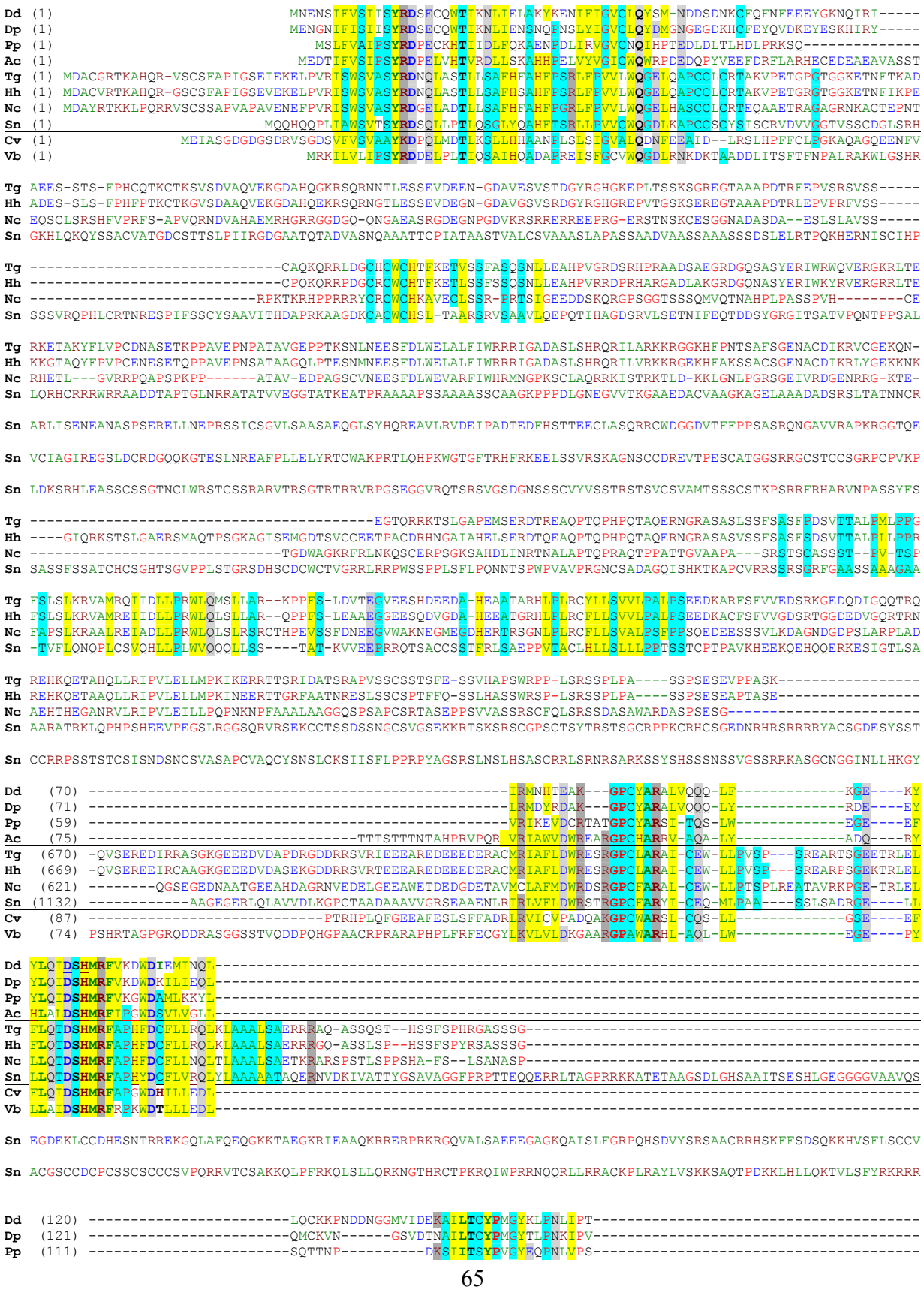
Tg*pgtA*-disruption

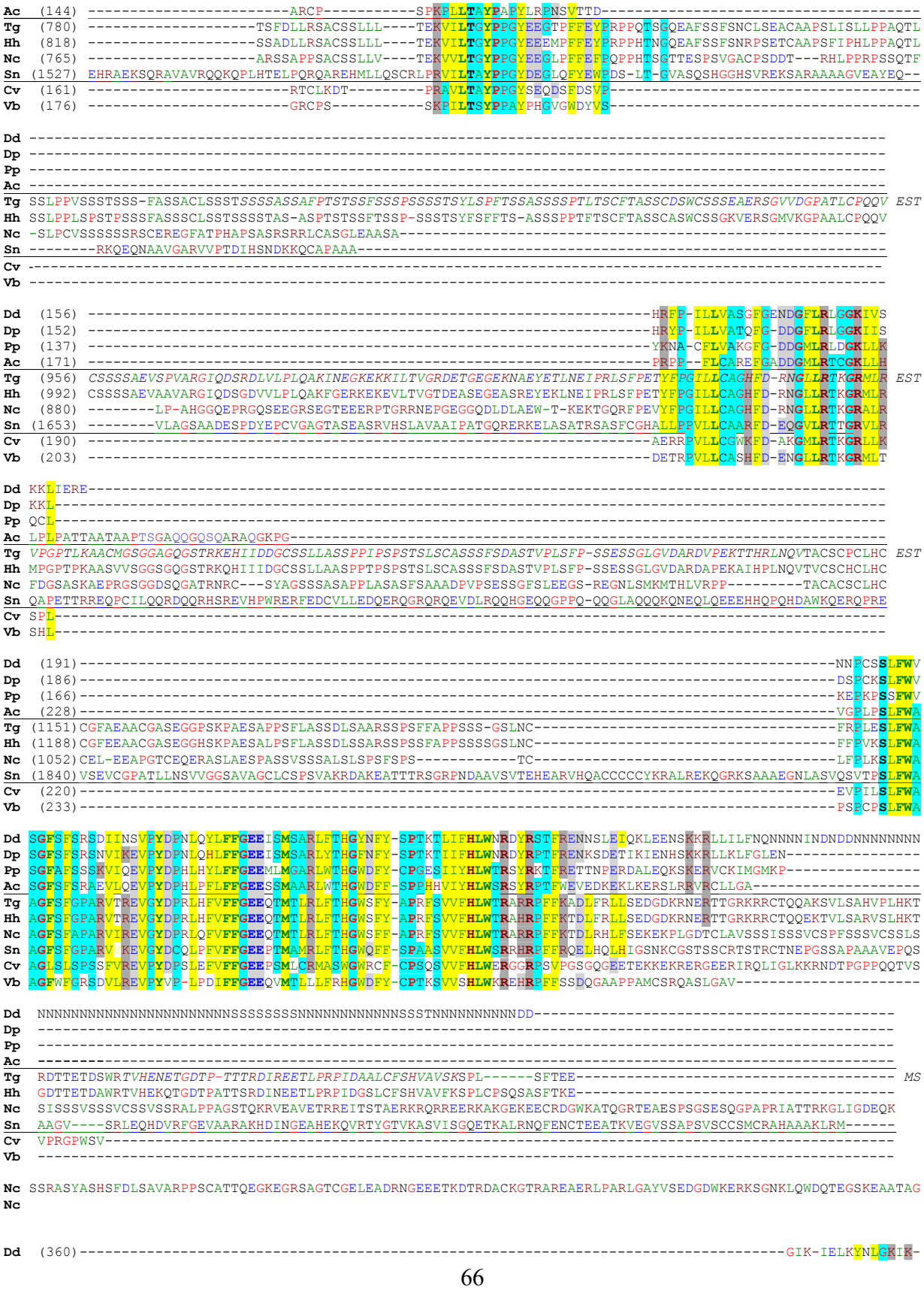
PCR 1:
Forward: 5'- CTCCGTTCTCCGAGTTCTTTC
Reverse: 5'- TTCCTCTCTTCACCTCTCTCT
PCR 2:
Forward: 5'-TGGAGAGAAACATGCAGGAA
Reverse: 5'-TGTGCTGTTGTCCTGGTACGA
PCR 3:
Forward: 5'- AGGACAGAGTTGAGAAGTTGGCGT
Reverse: 5'- CTGTCTCGACTGCCTTCAAATA

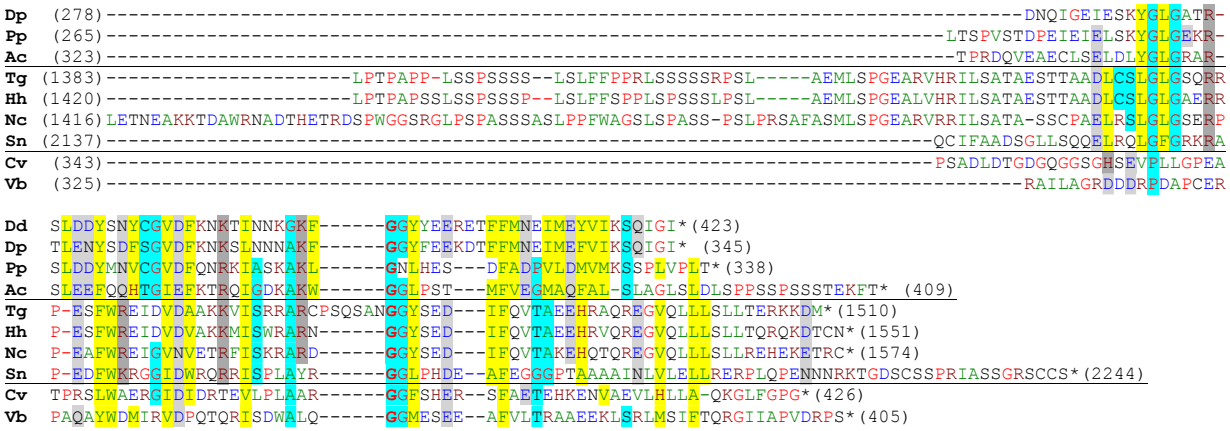
SF tagging of TgSkp1

Amplification primers
g) SKP.LIC.F 5′-TACTTCCAATCCAATTTAGCGCGTGCCTTTGCTTTCACTTGTCT (LIC)
g')SKP.LIC.R 5'-TCCTCCACTTCCAATTTTAGCTGCATCTTCGCACCACTTGTTCTC (LIC)

Tagging confirmation
l) Forward: 5'-TCGCCTCCATGAGTGCGTTGATAA
l')Reverse: 5'-ATTTACAACTCAGTGCTCCAGCCG







MS = sequence supported by mass spectrometry at ToxoDB
EST = sequence supported by EST sequences at ToxoDB

Origin of Sequences:

Amoebozoa–

Dd: *Dictyostelium discoideum,* GI:60471644

Dp: *Dictyostelium purpureum,* jgi|Dicpu1|89733|fgeneshDP_pm.C_scaffold_328000004

Pp: *Physarum polycephalum, contig assembled from multiple overlapping raw genomic fragments accessed at Physarum database (http://genome.fli-leibniz.de/ggl_blast/blast_ggl.pl)*

Ac: *Acanthamoeba castellani (Neff strain),* GI:470526279

Apicomplexa–

Tg: *Toxoplasma gondii,* From ToxoDB 7.3: TGGT1_315885

Hh: *Hammondia hammondi,* From ToxoDB 7.3: HHA_315885

Nc: *Neospora caninum,* From ToxoDB 7.3: NCLIV_058510

Sc: *Sarcocystis neurona,* From ToxoDB 7.3:SN3_00900395

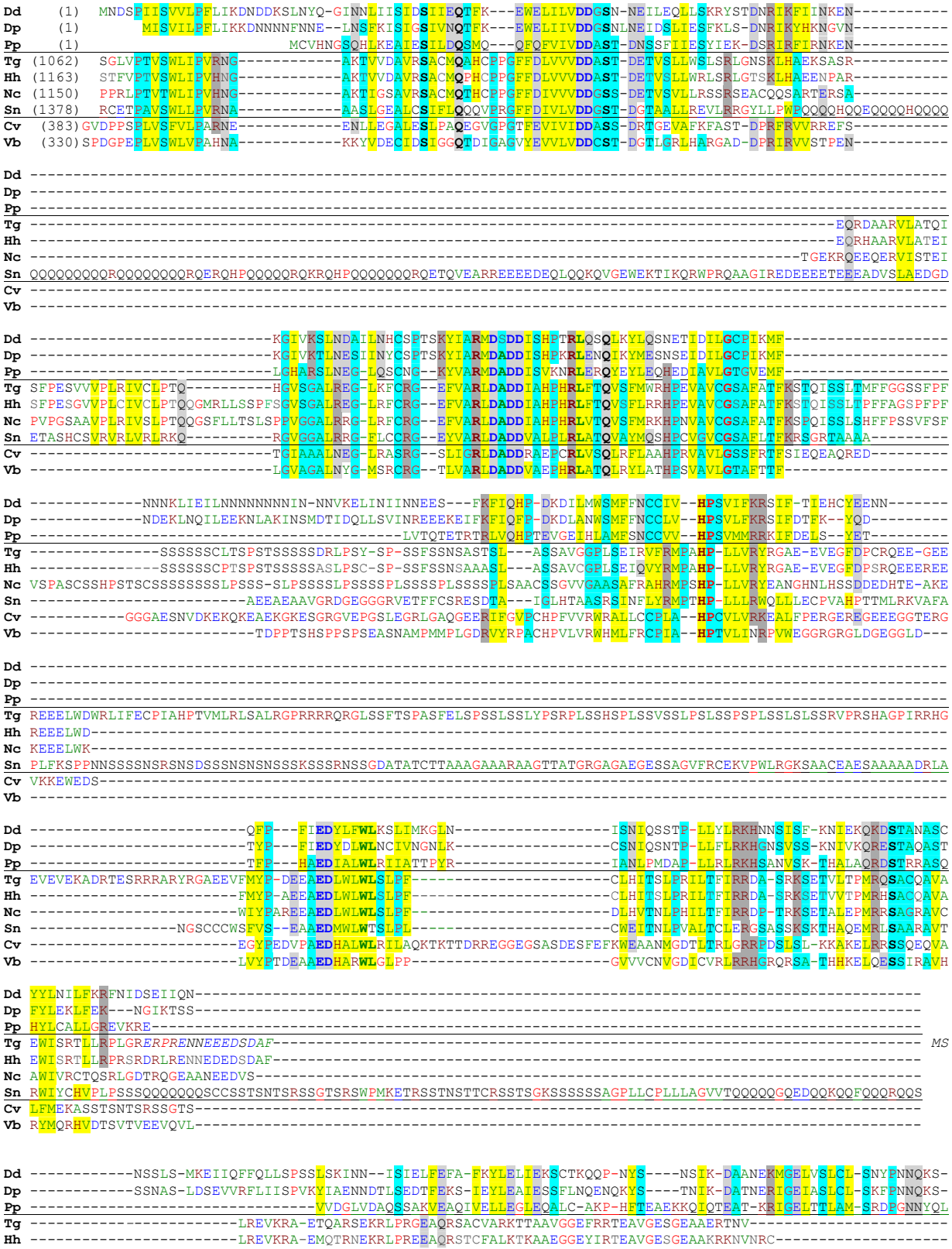
Chromerida–

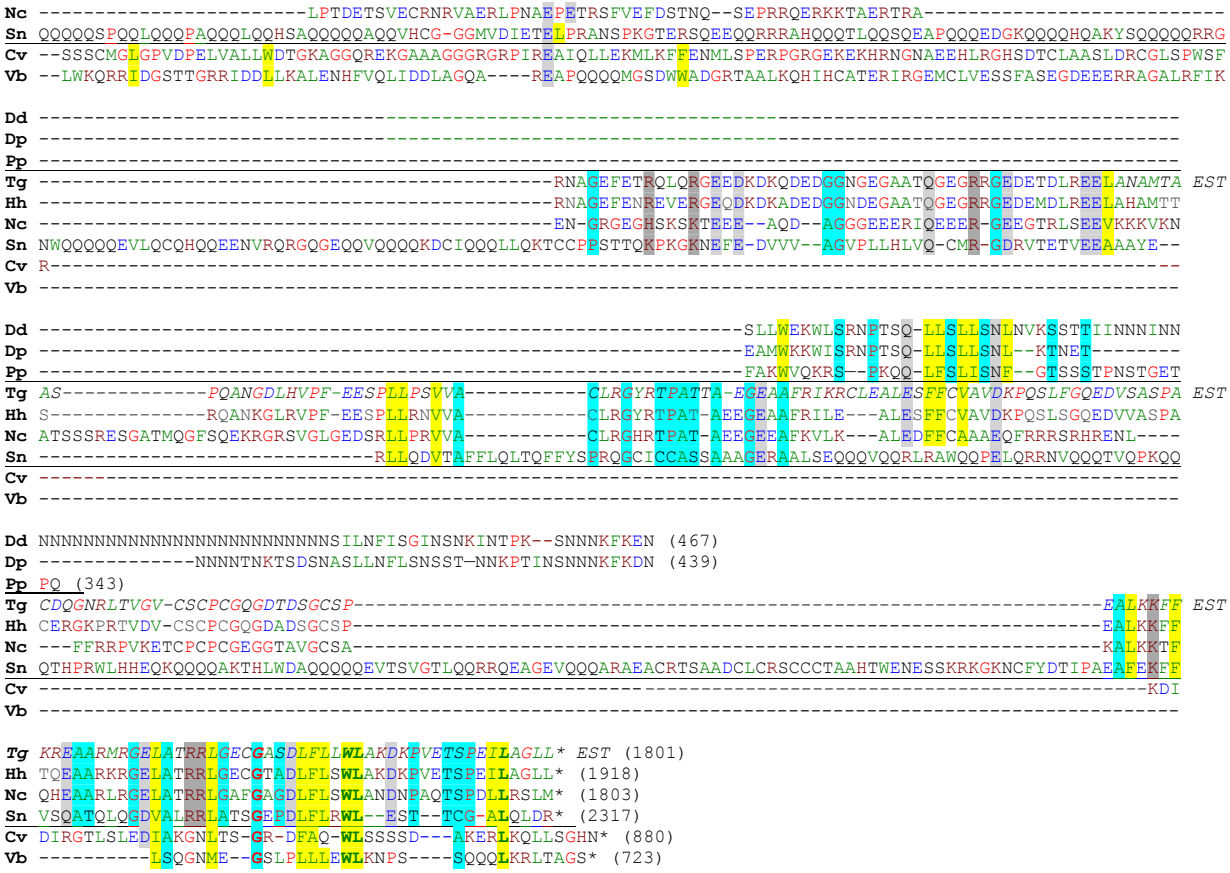
Cv: *Chromera velia,* From CryptoDB, Cvel_16021

Vb: *Vitrella brassicaformis,* from CryptoDB Vbra_13029

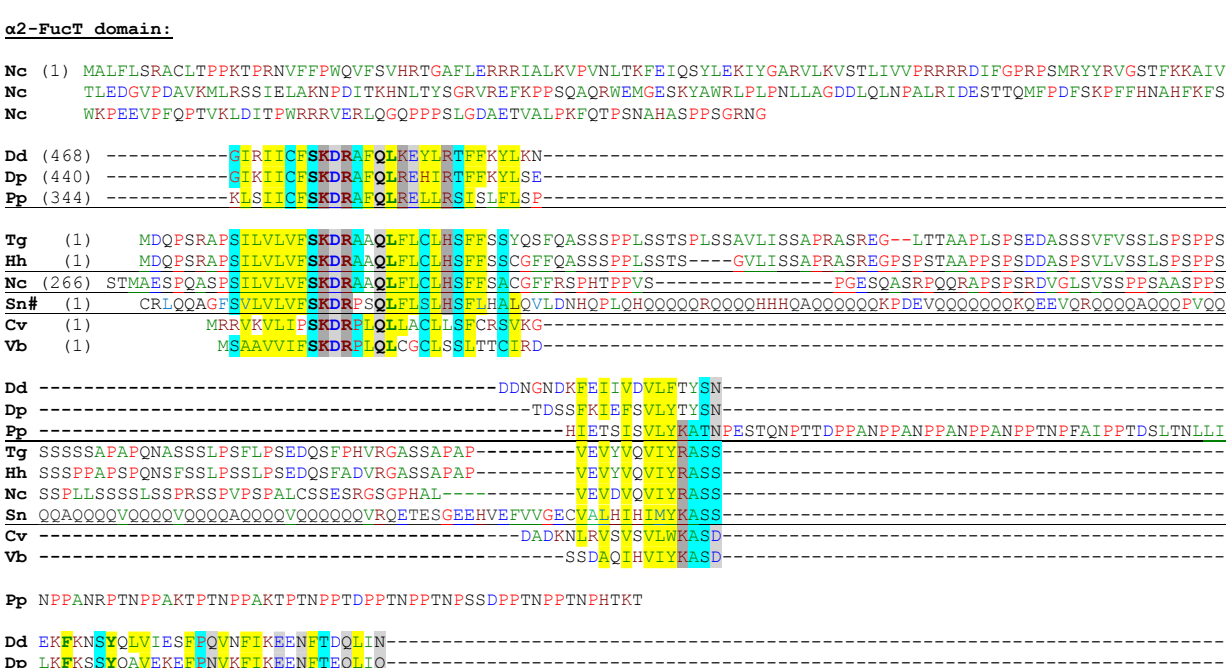
Supplemental Figure S2.1. Alignment of Gnt1-like sequences. The amino acid sequence of *Dictyostelium discoideum* Gnt1 is aligned with the amino acid sequences of predicted Gnt1-like proteins from 3 other amoebozoa, 4 coccidian apicomplexa, and 2 chromerids that also possess *phyA*- and *pgtA*-like sequences. Amino acids are color-coded with respect to chemical similarities that guided the alignments, giving preference to the registration of hydrophobic residues: green, hydrophobic; blue, acidic; dark red, basic; black, polar; bright red, secondary structure breaking (P or G). To facilitate the aligning, positions occupied by identical amino acids across all the organisms are bolded and similar amino acids are highlighted when they have majority representation at a given position. Numbers in parentheses refer to the position of the amino acids to the right. Origins of the sequences are given at the bottom. Sequences whose expression has been confirmed at the transcriptional (EST) or proteomic (MS) level, as reported by http://toxodb.org/toxo/, are annotated at the right.

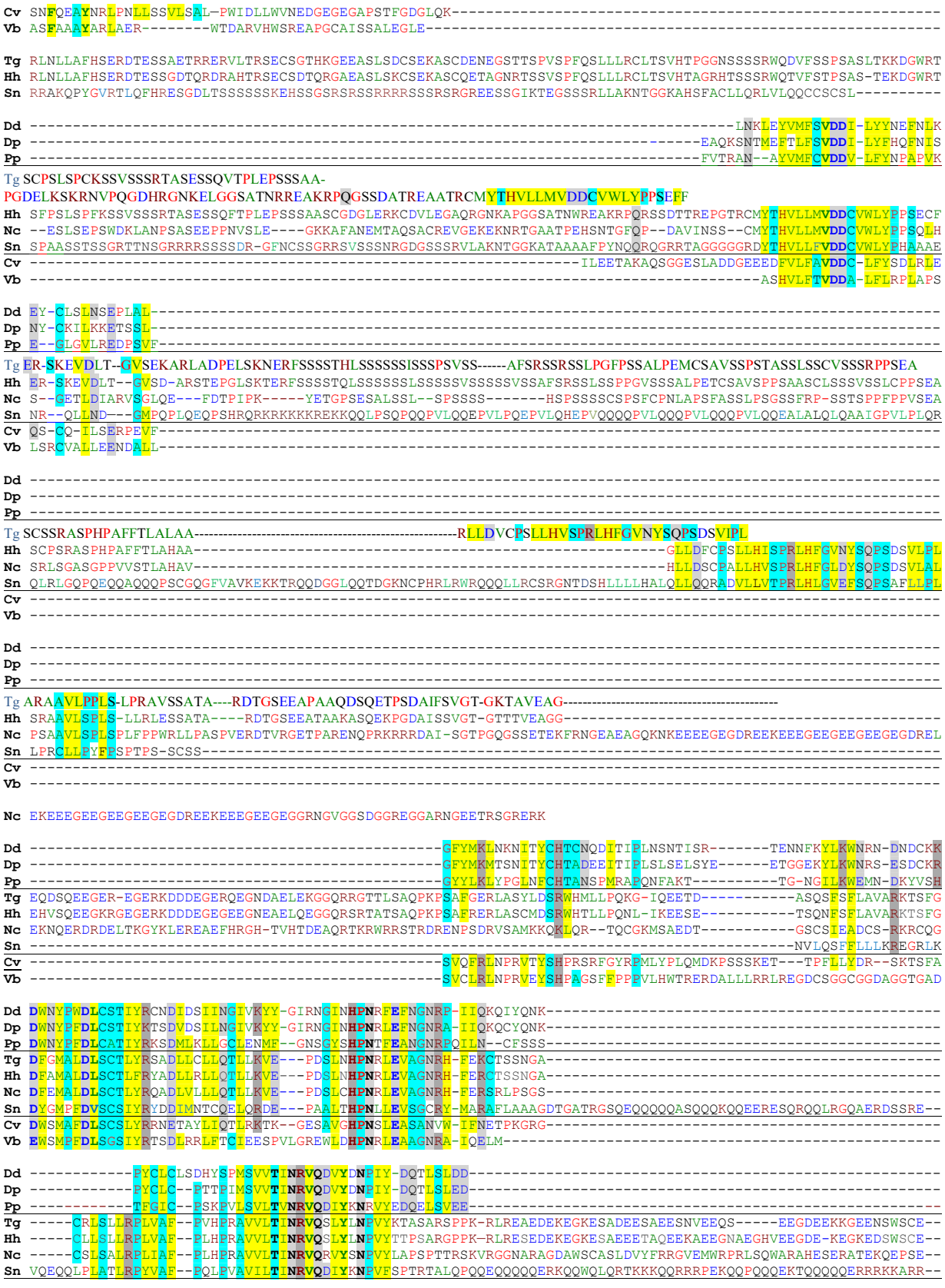
β3-GalT domain

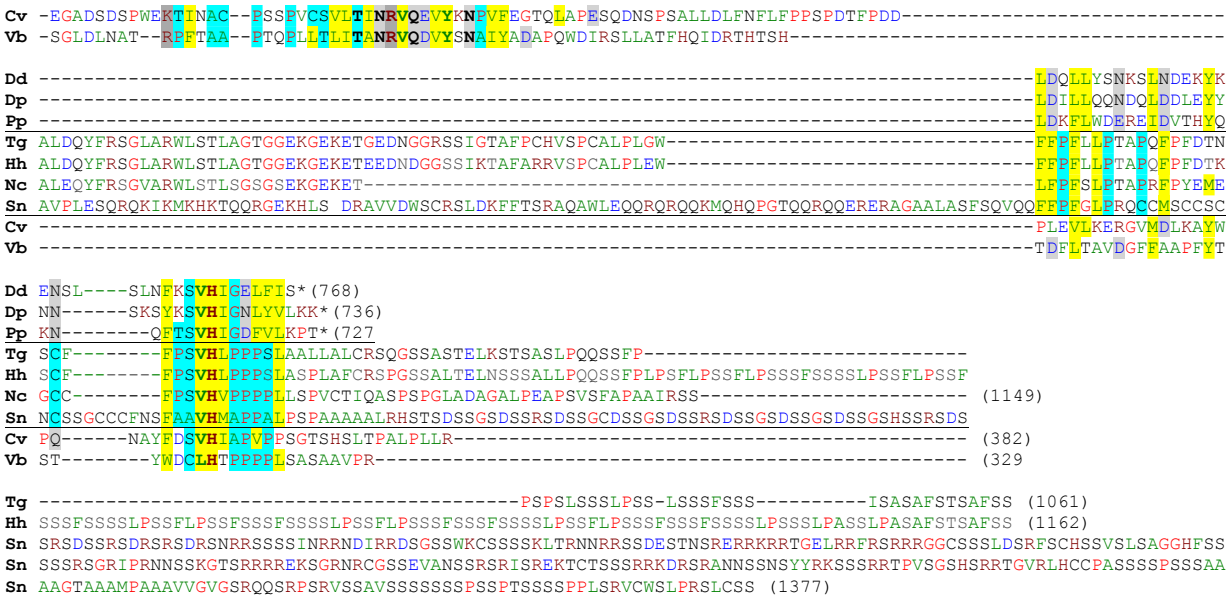




α2-FucT domain:







# = numbering start is arbitrarily assigned
MS = sequence supported by mass spectrometry at ToxoDB
EST = sequence supported by EST sequences at ToxoDB

Origin of Sequences:

Amoebozoa–

Dd: *Dictyostelium discoideum,* GI:60467473
Dp: *Dictyostelium purpureum,* jgi|Dicpu1|159362|GID1.0049988
Pp: *Physarum polycephalum,* (gene locus 131, Phypoly_transcript_04421:Physarum database (*http://genome.fli-leibniz.de/ggl_blast/blast_ggl.pl*)

Apicomplexa–

Tg: *Toxoplasma gondii,* From ToxoDB 7.3: TGGT1_260650
Hh: *Hammondia hammondi,* From ToxoDB 7.3: HHA_260650
Nc: *Neospora caninum,* From ToxoDB 7.3: contig FR823389
Sc: *Sarcocystis neurona,* From ToxoDB 7.3: Contig sneu_scaffold00020

Chromerida–

Cv: *Chromera velia,* From CryptoDB, CVEL_18515
Vb: *Vitrella brassicaformis,* From CryptoDB Vbra_12849

Supplemental Figure S2.2. Alignment of PgtA-like sequences. Sequences related to DdPgtA are aligned as in Fig. S1. Since the order of the two predicted GT domains of the apicomplexan and chromerid *pgtA*-like sequences is reversed relative that of the amoebozoa, N-terminal regions of the amoebozoan proteins are aligned with the C-terminal regions of the other sequences, and the C-terminal regions of amoebozoan PgtA sequences are aligned with N-terminal regions of the other sequences. Sequence origins are listed at the bottom. Sequences whose expression has been confirmed at the transcriptional (EST) or proteomic (MS) level are annotated at the right.

CHAPTER 3

BIOCHEMICAL AND CELLULAR ROLES OF A NOVEL GLUCOSYLTRANSFERASE THAT EXTENDS THE CORE TRISACCHARIDE OF *TOXOPLASMA* SKP1[1]

[1]Kazi Rahman, Msano Mandalasi, Peng Zhao, John H. Kim, Sheikh M. Osman, Hanke van der Wel, Lance Wells, Christopher M West. To be submitted to the *Journal of Biological Chemistry.*

Abstract

Skp1 is an adaptor subunit of the SCF (Skp1/Cullin-1/F-box protein) class of E3 ubiquitin ligases that are important for cell cycle and developmental regulation. Unlike its animal counterparts, Skp1 from *Toxoplasma gondii* (*Tg*) is hydroxylated at Pro154 by an O_2-dependent prolyl-4-hydroxylase (PhyA), and the resulting hydroxyproline is sequentially modified by a chain of 5 sugars. A related modification is also found in the social amoeba *Dictyostelium discoideum (Dd)*, where it regulates SCF assembly and O_2-dependent development. We previously reported that homologous genes assemble a similar core trisaccharide in both organisms. While *Dictyostelium* uses AgtA, a bifunctional galactosyltransferase, to add the terminal disaccharide on DdSkp1, *Toxoplasma* genome lacks AgtA homologue, and therefore, the mechanism of addition of the two peripheral sugars on TgSkp1 remained unknown. Here, we report that *Toxoplasma* diverged from *Dictyostelium* by recruiting a novel glucosyltransferase, named Glt1, to transfer the fourth sugar on its Skp1. Disruption of *glt1* blocked addition of the terminal disaccharide, based on mass spectrometry. Cytosolic extracts catalyzed transfer of [^{3}H]glucose from UDP-[^{3}H]glucose to a synthetic trisaccharide corresponding to the Skp1 glycan in a *glt1*-dependent fashion. Furthermore, recombinant Glt1 catalyzed the same reaction *in vitro*, indicating that it is the direct mediator of Skp1 glucosylation. The enzyme preparation was over 1000-fold more active toward the glycan attached to Skp1, suggesting that substrate recognition involves protein specific determinants. Plaque assays showed that *glt1*-knockout parasites exhibit a modest growth defect intermediate between that of parental and *phyA*-knockout parasites, indicating that glycosylation of hydroxyproline is important for mediating *phyA* function in cells. A genomics bioinformatics survey suggests that Glt1 belongs to the ancestral Skp1 glycosylation pathway and is related to glycosyltransferases in traditional Golgi glycosylation pathways. Understanding Skp1

glycosylation and its regulation in *Toxoplasma* will enable studies of the role of SCF ubiquitin ligases in parasite biology and to develop its potential as a drug target.

Introduction

Skp1 is an adaptor subunit of SCF (Skp1/Cullin-1/F-box protein)-class E3 ubiquitin ligases that target proteins for polyubiquitination and degradation via 26S proteasome (159). In the agent for human toxoplasmosis, *Toxoplasma gondii*, Skp1 is hydroxylated by the cytoplasmic prolyl 4-hydroxylase (*phyA*) at Pro154, and subsequently modified by a linear pentasaccharide that regulates parasite growth in fibroblasts (chapter 2) (16). In the social amoeba *Dictyostelium*, Skp1 is also modified by a pentasaccharide, which regulates SCF assembly and O_2-dependent development (10).

Both in *Toxoplasma* and *Dictyostelium*, the first 3 sugars are added sequentially to Hyp of Skp1 by Gnt1, a polypeptide αGlcNAcT, and PgtA, a processive bifunctional enzyme with β1,3-GalT and α1,2-FucT activities (10, 16). These sugar nucleotide-dependent enzymes are soluble in the cytoplasm and lack a rough endoplasmic reticulum targeting sequence or membrane anchor motifs. The Fuc terminus of the core trisaccharide of *Dictyostelium* Skp1 is further extended by two α,1,3-linked Gal residues by AgtA, a cytosolic glycosyltransferase from CAZy GT family 77. AgtA has a N-terminal catalytic domain and a C-terminal WD40-repeat domain expected to fold into a β-propeller (111). The C-terminal domain serves several functions including an apparent dampening of the function of undermodified Skp1, yet targeting and activating the catalytic domain to glycosylate the trisaccharide form of Skp1, FGGn-Skp1 (4). Data indicate that Skp1 is the sole cellular target of AgtA and bioinformatics analysis indicate that AgtA is restricted to *Dictyostelium* and other amoebazoa (10). The mechanism of addition of the terminal disaccharide

of *Toxoplasma* Skp1 cannot be explained by known glycosylation pathways, suggesting the existence of novel glycosyltransferases residing in the cytoplasmic and/or nuclear compartments of the parasite.

Here, using genomic studies, reverse genetics and mass spectrometry approaches, we have shown that *glt1*, a previously non-annotated gene from *T. gondii*, encodes a cytoplasmic glucosyltransferase from CAZy GT family 32 that modifies the Fuc terminus of TgSkp1 trisaccharide. The enzyme is absent in the amoebazoans, which can have AgtA, but is widespread in a large group of protists that lack AgtA but harbor earlier genes of the pathway.

Experimental Procedures

Parasites strains, Host Cell Culture and Plaque Assays– Strain RHΔ*ku80*Δ*hxgprt* (RHΔΔ) of *Toxoplasma gondii* was cultured on established monolayers of human foreskin fibroblasts (HFFs), hTERT cells, or BJ fibroblasts, as indicated, in the presence of Complete Medium, which consisted of Dulbecco's Modified Eagle's Medium (Corning Inc.) supplemented with 10% (v/v) fetal bovine serum, 2 mM L-glutamine and 100 units/ml penicillin/streptomycin (Corning Inc.) at 37°C in a humidified CO_2 (5%) incubator. RHΔ*ku80*Δg*lt1* (*glt1*Δ) strain was selected and manintained under the same medium supplemented with 25 μg/ml mycophenolic acid (Sigma) and 25 μg/ml xanthine (Sigma). The medium of complemented *glt1*Δ strains was instead supplemented with 250 μg/ml of 6-thioxanthine (Matrix Scientific). Strains were cloned by limiting dilution on BJ HFF cells in 96-well plates. Plaque assays were conducted in medium without selection drug as described (160).

Tgglt1 Disruption and Complementation– DNA for the gene disruption was generated from pminiGFP.ht, in which the *hxgprt* gene is flanked by multiple cloning sites as described (160). Briefly, 5'-flanking and 3'-flanking targeting sequences of *TgGlt1* from RHΔΔ were PCR amplified with primer pairs a and a' and pairs b and b', respectively (**Table 3.1**). The 5'-fragment was digested with ApaI and HindIII and inserted into similarly digested pminiGFP.ht. The 3'-fragment was similarly inserted using XbaI and NotI. The resulting vector was linearized with ApaI and electroporated into strain RHΔΔ. Transformants were selected under mycophenolic acid and xanthine, and GFP negative parasites were cloned by limiting dilution. Genomic DNA was prepared and screened by PCR, according to ref. (160) and as described in Results.

To complement the Tg*glt1*Δ strain, the HXGPRT cassette of pminiGFP.ht was replaced with a ~5kb DNA fragment containing the *Tgglt1* coding region from RHΔΔ as well as ~1 kb each of 5'-flanking and 3'-flanking DNA by the Complementary Annealing Mediated by Exonuclease cloning method (161). Briefly, the vector and insert were PCR amplified separately for 20 cycles by Q5 high fidelity DNA polymerase (primers in **Table 3.1**). The amplified insert was designed with 15-base overhangs that matched the termini of the amplified vector. The gel purified amplicons were mixed at a 1:3 molar vector-to-insert ratio (total ~100 ng), incubated with T4 DNA polymerase (Novagen) at room temperature for 2.5 minutes to generate 5'-overhangs, incubated at 75ºC for 20 mins to heat-inactivate the polymerase, and annealed at 50ºC for 30 mins. 2 µl of the mixture was transformed into *E. coli* Top10 competent cells. The plasmid was electroporated into *glt1Δ* parasites which were subjected to selection in the presence of 6-thioxanthine (160). To complement *glt1Δ* parasites with a triple point mutant of (D348N, D363A, D365A) version of Tg*glt1*, the above construct was altered by site-directed mutagenesis (primers in **Table 3.1**) as described (16).

Table 3.1. Oligonucleotide sequences employed.

Targeting sequence amplification:

Tg*glt1*-disruption

a) Glt1F1 5′-flank 5′-end 5′-CCGGGCCCCAAGCAAACCCTGTCTCTCTAT(ApaI)
a′) Glt1R1 5′-flank 3′-end 5′-GGAAGCTTCAATTCCACACTCAAGCGAATG(HindIII)
b) Glgt1F2 3′-flank 5′-end 5′-GCTCTAGAGAAGATGCGCAGGATTCATCTA (XbaI)
b′) Glt1R2 3′-flank 3′-end 5′-GCGCGGCCGCCTGTTAGGCGATGCTGTCTT (NotI)

Tgglt1 complementation

Vector (forward): GCAGCCACGATTTTCCAAG
Vector (Reverse): TATTGATGAACCCGTTGTTCC
Insert (Forward): ACGGGTTCATCAATAGCTGGCTGACGATCTTTAAGT
Insert (Reverse): GAAAATCGTGGCTGCCCACATAACTCCTGCTTACAGAC

Mutation of Tgglt1 complementation construct:

^{363}DVD^{365} to ^{363}AVA^{365}:
Forward: 5'-ggttttatagcctccatggccacagctgcgtatattcctgaat-3'
Reverse: 5'-attcaggaatatacgcagctgtggccatggaggctataaaacc-3'

D348N:
Forward: 5'-gcaggcgggcaatattcgattttgcacccgg-3'
Reverse: 5'-ccgggtgcaaaatcgaatattgcccgcctgc-3'

PCR confirmation of Tgglt1 disruption and complementation

PCR 1
Forward: 5′-CTGTGACTCAGAACTCCTCAAG
Reverse: 5′-CCCGAAGAACATCCGATTAGAG
PCR 2
Forward: 5′-GCTAGTGTACAGTTCCCACTAAG
Reverse: 5′-AGTCGCGGAACATCTCGTTGAAGT
PCR 3
Forward: 5′-ATTTGCATCCTGAAAGGCTCTCGC
Reverse: 5′-TCGAGCAAGTATCGCCTAATG

Immunoprecipitation of TgSkp1- TgSkp1 was immunoprecipitated from urea-solubilized parasite extracts as previously described (160), except that the sample was first precleared by incubation with 50 µl of ant-rabbit antibody bound beads for 1 h at 4 °C.

Mass Spectrometry of TgSkp1 Peptides- Samples were dissolved in 100 µl 10 mM dithiothreitol in NH_4HCO_3 buffer at 56 °C for 1 h, alkylated with 55 mM iodoacetimide for 45 min in dark, and digested with trypsin (Promega) at 37 °C overnight. The resulting peptides were recovered by addition to a C18 spin column (MicroSpin™ Column, The Nest Group), elution with 0.1% formic acid in 80% acetonitrile, and drying under vacuum. Peptides were reconstituted in 19.5 µL of solvent A (0.1 % formic acid) and 0.5 µL of solvent B (0.1% formic acid in 80% acetonitrile), separated on an Acclaim PepMap RSLC C18 column (75 µm x 15 cm), and eluted into the ion source of an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer (Thermo Fisher Scientific) with a linear gradient consisting of 0.5-100% solvent B over 150 min at a flow rate of 200 nL/min. The spray voltage was set to 2.2 kV and the temperature of the heated capillary was set to 280 °C. Full MS scans were acquired from m/z 300 to 2000 at 120k resolution, and MS2 scans following collision-induced fragmentation were collected in the ion trap for the most intense ions in the Top-Speed mode within a 3-sec cycle using Fusion instrument software (v2.0, Thermo Fisher Scientific). The acquired raw spectra were analyzed using SEQUEST (Proteome Discoverer 1.4, Thermo Fisher Scientific) with a full MS peptide tolerance of 20 ppm and MS2 peptide fragment tolerance of 0.5 Da, and filtered to generate a 1% target decoy peptide-spectrum-match (PSM) false discovery rate for protein assignments. The spectra assigned as glycosylated TgSkp1 peptides were manually validated.

Cytosolic S100 Preparation and Glycosyltransferase Activity Assays– Cytosolic S100 extracts were prepared and desalted as described (6), and assayed for glycosyltransferase activities

as follows. GlcT or GalT activity was assayed by the transfer of [^{3}H] from UDP-[^{3}H]Glc or UDP-[^{3}H]Gal to exogenous FGGn-*Dd*Skp1. Typically, a 50 µl reaction volume containing 30 µl S100 fraction (1 mg protein/ml), 50 pmol of FGGn-*Dd*Skp1(11), 1 µCi UDP-[^{3}H]Glc (27 Ci/mmol, Perkin-Elmer) or UDP-[^{3}H]Gal (113), in 50 mM HEPES-NaOH (pH 7.4), 10 mM $MgCl_2$, 2 mM $MnCl_2$, 2 mM DTT, 3 mM NaF, and protease inhibitors, was incubated at 37 °C for 1.5 or 3 h. Reactions were stopped by addition of 4×-Laemmli electrophoresis sample buffer, supplemented with 2 µg of soybean trypsin inhibitor (Sigma) as a marker that comigrates with *Dd*Skp1, boiled for 3 min, and resolved by SDS-PAGE. Radioactivity at the Skp1 band position was counted as described (6). The chemical form of the radioactivity incorporated into DdSkp1 was determined by HPAEC (high pH anionic exchange chromatography) analysis after acid hydrolysis (6 M HCl at 100 °C for 4 h) as described (6), except that 1.5 nmol each of L-Fuc, D-Glc, D-Gal and D-Man were added as internal standards. Biochemical complementation of Glt1 was performed as described above for the GlcT assay except that 2µci of UDP-^{3}H-Glc was used in a 200 µl reactions containing 170 µl of each of the RHΔΔ or *glt1*Δ desalted parasite extracts (1mg/ml) with 2µl of His_6-Glt1 and incubated for 3h at 37 °C. The reaction products were concentrated by Nanosep 3K concentrator (Pall Corporation) that also removed the unreacted UDP-^{3}H-Glc. The samples were resolved by SDS-PAGE gel and each of the lanes were cut into 36 slices and radioactivity in each of the slice was measured as described above.

Expression and purification of Recombinant Glt1- The predicted coding sequence of Glt1 was codon optimized for expression in *E. coli*, and chemically synthesized by GenScript (Piscataway, NJ) and inserted into the pUC57 vector between its NdeI and BamHI sites **(Supplementary Fig. S3.2)**. After treatment with NdeI and BamHI, the released coding fragment

was inserted into similarly digested pET15b (Invitrogen), which resulted in the full-length 605 amino acid long coding sequence preceded by an N-terminal His_6-tag and TEV protease cleavage site (MGSSHHHHHHSSGRENLYFQGH-). *E. coli* Gold cells expressing His_6-Glt1 were grown to an A_{600} (1 cm path length) of 0.4–0.6 in 1 L LB medium in the presence of 100 µg/ml of carbinicillin, and induced with 0.5 mM isopropyl 1-thio-β-D-galactopyranoside at 22°C. After overnight incubation, cells were pelleted, resuspended in 40 ml 0.1 M Tris-HCl (pH 8.2), 5 mM benzamidine, 0.5 µg/ml pepstatin A, 5 µg/ml aprotinin, 5 µg/ml leupeptin, 0.5 mM phenylmethylsulfonyl fluoride, and 1 mg/ml lysozyme at 4°C, lysed using a probe sonicator (model #FB505, Fisher Scientific) and mixed with 0.5 M $MgCl_2$, 10 mg/ml RNaseA and 5 mg/ml DnaseI to final concentrations of 5 mM, 50 µg/ml and 10 µg/ml, respectively. The lysate was immediately centrifuged at 22,000 × *g* for 30 min at 4°C, and the supernatant (S22) was immediately applied to a 1-ml GE Healthcare Hi-Trap Ni^{2+} column pre-equilibrated at 4°C in 20 mM Tris-HCl (pH 8.0 at 4 °C), 0.5 M NaCl, and 5 mM imidazole. The column was eluted using a 0.005–1.0 M gradient of imidazole in the same buffer. His_6-Glt1 elution was tracked based on Western blotting (16) with anti-His_6 monoclonal antibody (Novagen, Catalogue # 70796-3) and enzymatic activity. Active fractions were pooled and dialyzed against 50 mM HEPES-NaOH, pH 7.4, and applied to a 1-ml Hi-Trap Q-Sepharose column (GE-Healthcare) pre-equilibrated at 4°C in 40 mM Tris-HCl (pH 8.0 at 4°C), 2 mM DTT and 2 mM $MgCl_2$. Protein was eluted with a gradient of NaCl from 0 to 1M prepared in the same buffer supplemented with 10% (v/v) glycerol, aprotinin and leupeptin (as above). Fractions with the highest enzymatic activity were confirmed for the presence of His_6-Glt1 by Western blotting with anti-His_6 antibody, pooled, and frozen as aliquots at -80°C.

Glt1 Enzyme Activity Assays– Hydrolysis of UDP-Glc and other sugar nucleotides was conducted using the UDP-Glo assay (Promega) as described (162). Briefly, 10 µl of purified His_6-Glt1 (after Q-column purification) was incubated in the presence of 50 µM sugar nucleotides in 20 µl reactions containing 50 mM HEPES-NaOH (pH 7.4), 2 mM $MnCl_2$, 5 mM DTT at 37°C for 16 h, and activity was quantitated based on conversion of the UDP reaction product to ATP. Glt1 glycosyltransferase activity was assayed as the transfer of ^3H from UDP-[6-^3H]Glc to synthetic acceptors or FGGn-*Dd*Skp1. Synthetic acceptors were Fucα1-*para*-nitrophenol (F-pNP)(Sigma), GlcNAcα1-Bn (112), Fucα1,2Galβ1,3GlcNAcα1-pNP (FGGn-pNP) (112), Galβ1,3GlcNAcα1-Bn, Galβ1,3GlcNAcβ1-Bn (163), Fucα1,2Galβ1-Bn (FG-Bn) (112), and Galβ1,3Fucα1-Bn (112). The standard reaction consisted of 1 µl of purified His_6-Glt1, and 2 mM synthetic acceptor or 1 µM FGGn-*Dd*Skp1, 50 mM HEPES-NaOH (pH 7.4), 2 µM UDP-Glc (0.5 µCi UDP-[6-^3H]Glc), 2 mM $MnCl_2$, 50 mM NaCl, and 5 mM DTT, in a final reaction volume of 30 µl. Reactions were incubated for 1 h at 37 °C.

Reactions using synthetic acceptor substrates were terminated by addition of 1 ml of ice-cold 1 mM sodium EDTA (pH 8.0), and applied to a 0.5-ml SepPak C_{18} cartridge (Waters), which was washed 6× with 5 ml H_2O and eluted with 5 ml MeOH. Samples were added to 8 ml of Biosafe II scintillation counting fluid (RPI) and radioactivity was quantitated in a Beckman L6500 scintillation counter. Activity from a reaction lacking an acceptor substrate was subtracted to correct for non-specific background.

Reactions using Skp1 as an acceptor substrate were stopped by sequential addition of 90 µl ice-cold 10 mM sodium EDTA (pH 8.0), 10 µl 100 mg/ml of BSA, and 90 µl of 10% (v/v) TCA. The precipitate was collected on a glass fiber filter under vacuum, and washed 5× with 1.5 ml of

ice-cold 10% TCA and 5× with 1.5 ml acetone. The filter was transferred to a tube containing 5 ml of Biosafe NA scintillation fluid (RPI) and counted as above.

Phylogenetic Analysis- For phylogenetic inferences of Glt1, 40 homologous protein sequences from different organisms were at first aligned by MUSCLE, uploaded in BioEdit sequence alignment editor V 7.2.5 and again manually refined to remove most of the gaps and ambiguous sites, which resulted in sequence alignment based on 201 amino acids of *Toxoplasma* Glt1 (**Fig. 3.2**). For phylogenetic inferences, MEGA7 software (164) was used to estimate the best substitution model as well as estimate for the Maximum Likelihood trees. The evolutionary history was inferred based on the Le_Gascuel_2008 model (165). The tree with the highest log likelihood (-13717.5614) was shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 1.3090)). The rate variation model allowed for some sites to be evolutionarily invariable ([+*I*], 3.5435% sites). The tree is drawn to scale, with branch lengths measuring the number of substitutions per site. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 179 positions in the final dataset. To determine the reliability of each node of the tree, 1000 repetitions were compared using the bootstrapping method (166). A complete manual alignment of several predicted Glt1 sequences along with some characterized and uncharacterized CAZy GT32 family proteins is shown in **Supplemental fig. S3.1**.

Results

Bioinformatics-based Prediction of Candidate TgSkp1 Modifying Glycosyltransferases- *Toxoplasma* Skp1 is modified by a pentasaccharide, but lacks a homolog of AgtA, the enzyme that catalyzes addition of the final two sugar residues *in Dictyostelium* Skp1. To identify a candidate protein(s) that could replace AgtA, we searched databases for predicted glycosyltransferases that reside in the cytoplasm, and exist only in protists whose genomes harbor predicted Skp1 modification pathway enzymes, *phyA, gnt1*, and *pgtA,* but lack *agtA* as outlined in **Fig 3.1**. Briefly, the predicted proteome (8460 proteins) of the *T. gondii* Type I GT1 strain was searched using i) the SUPERFAMILY server, which assigns protein domains at the SCOP 'superfamily' level using hidden Markov models; ii) dbCAN, an automated Carbohydrate-active enzyme Annotation database which utilizes a CAZyme signature domain-based annotation based on a CDD (conserved domain database) search, literature curation, and a hidden Markov model; and iii) the Pfam database, which resulted in 45 putative glycosyltransferases. The 45 sequences were scanned by SignalIP 4.1 and TMHMM servers for signal sequences or transmembrane domains, which yielded 10 candidate cytoplasmic glycosyltransferases. Among those, TgGnt1 and TgPgtA were already identified as TgSkp1 modifying glycosyltransferases (160). The remaining 8 sequences were subjected to BlastP analysis against the NCBI non-redundant database to search for their phylogenetic co-distribution with *gnt1* and *pgtA*. This yielded two candidates, one from CAZy GT32 family, TGGT1_205060, and another from CAZy GT8 family. The GT8 sequence, referred to as Gat1, is the subject of an independent report (Chapter 4).

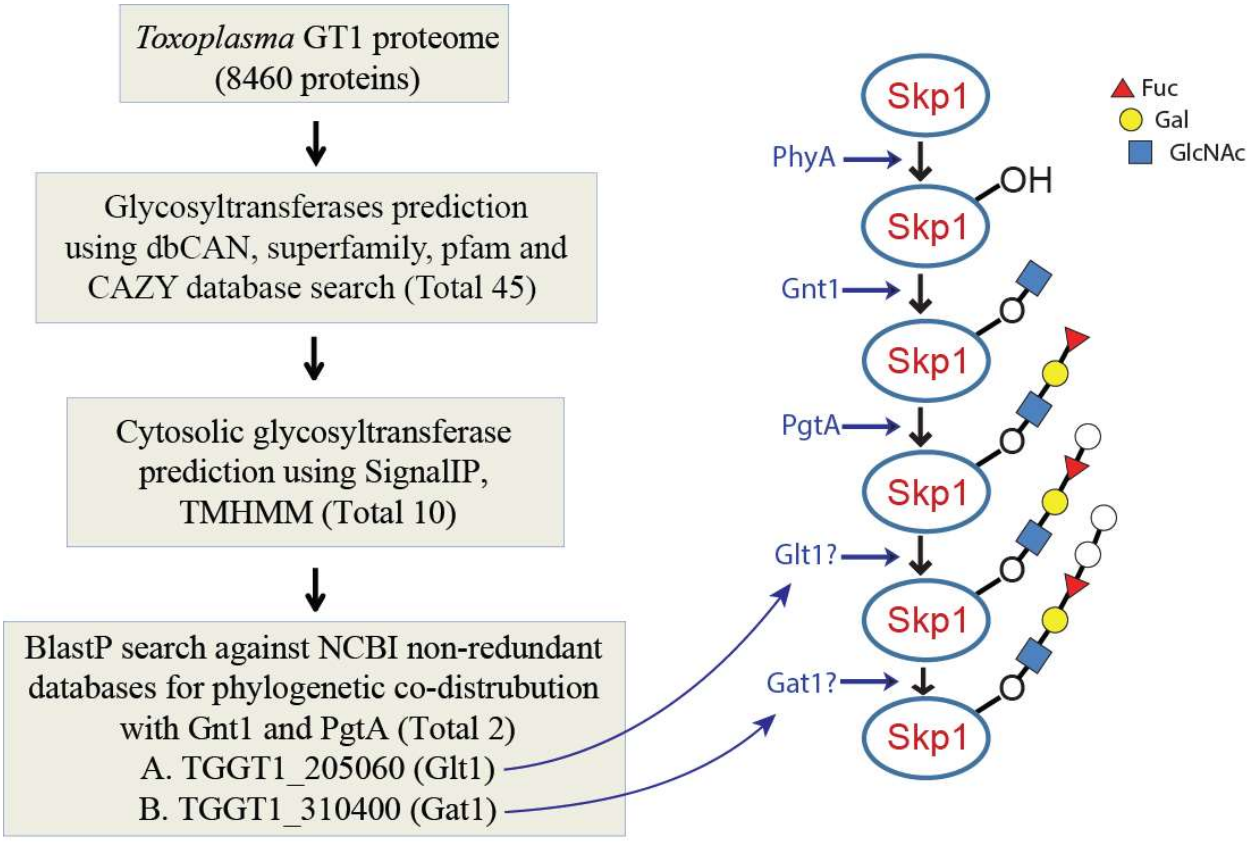


Figure 3.1. Bioinformatics Strategy used to Identify Putative TgSkp1 Modifying Glycosyltransferases. Left panel shows the steps utilized in bioinformatic analysis to identify potential terminal GT candidates. Right panel shows the Skp1-glycosylation pathway in *Toxoplasma*. Arrows from left-to-right show at which step of the pathway putative GTs are predicted to work.

Glt1 Sequence Characteristics and Phylogenetic Position– In the *Toxoplasma* database ToxoDB V29, TGGT1_205060 (type I GT1 strain) is annotated as a 605 amino acid protein encoded by five exons, that is conserved in ME49 (type II) and VEG (type III) strains with 99% identity. The sequence has a DxD motif typical of and essential for CAZy GT32 family and superfamily A glycosyltransferases (167). To investigate the evolutionary origin of TGGT1_205060, a BLASTP search was conducted against the NCBI non-redundant protein database, which resulted ~50 homologous sequences at $E<10^{-5}$. All of the protozoan sequences are from organisms that are predicted to have *Toxoplasma*-like Skp1 modification pathway (i.e., harbor *phyA*, *gnt1* and *pgtA* but lack *agtA*). Given the proclivity of *Toxoplasma* proteins to harbor apicomplexan-specific insertions (153, 160), we searched again using the putative ortholog from the chromerida *Vitrealla brassicaformis*, a close relative of apicomplexa that is predicted to express the *Toxoplasma*-like Skp1 modification pathway, but lacks apicomplexan-specific insertions within their proteins (168). This new strategy considerably increased the number of non-redundant hits below $E<10^{-5}$. We aligned 40 homologous sequences from representative bacteria, protists and other eukaryotes (**Fig. 3.2**). A phylogenetic analysis shows that Glt1-like sequences from protists formed a distinct clade (red color) (**Fig. 3.3**), suggesting the existence of a common ancestor that uniquely seeded protozoan phylogeny. In addition to several motifs conserved among all of the sequences, a glutamate at position 393 of TGGT1_205060 is conserved in all the sequences associated with Skp1 modification and, in addition, the bacterium *Treponema lecithinolyticum*. However, the bioinformatic analysis did not predict the activity of TGGT1_205060, which is hereafter referred to as *glt1*.

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                      10         20         30         40         50         60         70         80
Tg          IPPLLHFVWL GGH-PPPFF- ETIRQSWAVH NPDLIQALWT DAHVTEWEMS DIKAFRKESC PGAKSDIARL LILCHYGGIY
Hh          IPPLLHFVWL GGH-PPPFF- EKIKQSWAIH NPDLIQALWT DAHVTEWEMF DIKAFRKESC PGAKSDIARL LILCHYGGIY
Nc          IPPLLHFVWL GGHRPPSFF- DEVQQSWAVH NPDLIQALWT DAHVIEWERT DIKTFRKESC LAARSDIARL LILCHYGGIY
Sn          IPPLLHFVWC GKEPLPDFF- LQFKQSWKK- NSLLIHALWR DAEIKGWRQR FAHMVEEEER LGAKSDLVRL LLLHTYGGVY
Cv          IPKRLHQIWL GPKEIPSQC- KEFMETWKRR HPEWEYKLWR HADVESLKLR NRSAYEEAEN FAHKSDILRL ELLSQFGGVY
Sm          IPRRIHQIWL GPRRWPEPC- ERFAGEWKAR HPTWEYRLWT DDDAGEELRG HAKAYEAADN PAEKSDILRL AIILRHGGLY
Vb          IPKIIHQIWL GGNQIPPEC- IPWMESWKRY HPDWEYKLWQ DGDLENLPLL PRALIASASN PAEKADILRL ELLRLFGGLY
Ac          IPFIIHQIWL GPHPIPAFC- LQQMETWRQI HPQWEYKLWT DTQVSTLKLQ NKEHFDLAGN YGEKSDILRY EILLQFGGIY
To          IPKILHFIWL GGNPLPRFTS LPCIESWRKH HTGWRFQIWT EADVERLEMH AYSYALKIGN YGLASDVLRL EILSIFGGVY
Bn          IPKIIHQIWF GSNPLPENFE TRFQETWKGT HPHWRYILWR DGDIEEHPLI LKHLLDDAQS SVEKSDIWRL AVLYEMGGVY
Ot          IPKKIHFIWL GEKEKPEYFK VHVYGSWTSK DTYSEIITWG EKDISELDLI NEIIEDKTLN PAFRADALRL EILYQEGGAY
Rf          IPKRIHQIWL GDKAIPECY- LTWQQEWKQK HQGWEYYFWD DEHVLTLPLL SQQMWPICDN FGEQSDLLRL FLLYQFGGVY
Ng          IPRVFHHIWL GSP-LPERF- QALRATWLRH HPEWTHRLWT DADVDSICLR NQHAYSAAPN YGQKSDILRY ELLERHGGVY
Sl          IPKKIHFIWI GKNDLPQNFR DFYQNQWKQT HSDYEFKIWT DEDIKDMIFT NQIIKDPEFN SGLRADALRL EILQTYGGIY
Pu          IPHVIHQIWL GPHPIPVDC- LAWMQTWKRL HPAWEYK--- ---------- ---AFDTATN FGEKSDILRY EILEKYGGVY
Aa          VPRKLHQIWL GPKPPPDAH- ---AAAWRAL HPDWEYKLWR DADVAALGLE NAAAFSAATN WGEASDIARY EILLRFGGVY
Es          IPLKIHHIWL GSP-LPEAF- ARLRESWLAR HPEWEVRLWT DADVDAFGLE NRGAYDAAQN FGQKSDILRY EVLLRHGGLY
Ap          IPKRIHQIWL GPA-PMDEF- HSYIQSWKNH HPDWEYTLWT EREIENWKLQ NQAAYDMATN FGEKSDILRY EILETFGGLY
Gt          IPRIIHHIWI GGK-LPGKF- QSLRDEWILL HPAWKHYLWD DESIQEFSSN PAGSYASASN YGEKSDILRL EVLNRFGGVY
Sc          IPQRV-WQWK ADDKFPSSF- RTYQKTWSGS YPDYQYSLIS DDSIIPFLEN LIQAFKLMPG NILKADFLRY LLLFARGGIY
Eh          IPKIIHQTGR SRC-VSAKF- HEVMRAWMDR FPGWAYRFHD DAAMESLLAR RHAVVQCVHS MTMKADMWRY LALWRYGGIY
Pp          IPRLIHQSWK SVYRIPTRF- HPWMKSWVEF HPTWTYVFWT DADNLRLFEL LLHVAKAVRK -VSLADMARY ALLHQVGGLY
Sp          IPRILHQSWK SADAIPEIF- APWMRSWVQH HPTWTYVFWD DAANLALFAR HYAVASSVGK -IHLADMTRY ALLHRFGGVY
Mb          VPKILHQIWL GTETPPCDWV TTWSRDFRAA HPDWTYHLWT QSELDRLPLR NRAAYLHEEH PANKADIARY ELLYHYGGVY
Ob          IPKIIHQTWK DKM-VPRKL- TSWVKSWVKN HPDWEYWLWT DASAREMLAD KLSTFDNYPE NIRRADALRY FVLYEFGGVY
Sr          IPKTIHQIWL GNNQAPCQWM AAWMHDYVQA NPEWTYRLWR TEDIEALHLE NSDIFALEPH PANKADIARY EILYRYGGVF
Ea          IPKLFHQSWS SNE-LPAKF- ERWSATCRRA HPDWEWVLWT DKDNEELVKT HLKTYKALPG VIYRADLVRN LYMHMFGGVY
Cm          FPPIIHQIWL GNRPFPSSF- SRSQESFRTK NPDMYYALWR DNDIACLIRR FAPVYAKITR VATTSDLARY LVLYVFGGFY
Ss          IPNIIHQSWK NNE-VPARY- KRWIQSWKEL HPNWRYILWT DHTNRKLVLR QIDIYDSLPV NIARADISRF LYMYSYGGVY
Cs          IPQIIHQSWK DEN-VPQGF- KPWQDSWKKN HPGWEYRLWT DVDNRKLVLD HLDTYDALPK PIMKADSVRY LYMYHVGGVY
Bd          IPHVLHQSWK NNT-LPLKF- SKWQKSWLDL HQDWEYKLWT DNDNMELCRD HFERFQEFLT NINRADTARY MYMHLYGGFY
Pb          IPKIIHQVWK TDT-IPDRW- KEPADMVRRM NPDFEIRLWS DEEALKFLQR FAATYLSYQY DIQRADAIRY FLLHHYGGVY
Fp          IPKIIHYCWF GGGPISPES- RKCIESWKKY CPDYKIIEWN EQNFEISQNR YQQAYEA-KK YAFVSDYVRL AVLYRYGGIY
Be          IPKVIHYCWL SGDPIPEKL- QRCMDSWKKF LPDYEFVLWD LERFSQWVKE A---FEA-RK YAFAADYIRL YAVYNYGGIY
Tl          IPHKIHQIWI GSD-LPEKY- YDWCNSWKKL NPEWEYKLWQ EQDILDLLDG PKSIFLKSKN FGAKSDIARY VILQKEGGVY
Ts          IPKIIHQIWI GDSEIPKQC- LQYIDSWKKH HSDWEYRLWT DENMIKLQN- -QELYDSTTN IRQKADIARY ELLYRFGGVY
Sh          IPPLIHQTWK NAD-VPPKW- RTWSETWRRH HPDWEYRLWT DADNRAFLEQ HLPVYDGYAE PIKRADAIRY FLLDHFGGLY
Sn          IPKVIHFIWV GPNPFPQES- VQNVYSWVEN HPNWTIKFWT DRKRLQLISE FRDYYNDSDN YAEKSDLLRY EILNQEGGLY
Ls          IPKIIHQTWR DHH-LPVP-- KAWPESWLRH NPDWEYRLWT DDDLLAFVQQ CETLYLSYPK PVQRADMARY LILHHHGGVY
Ps          IPKIIHYCWF GGKPLPDEV- KKYMETWKKY CPYYEITEWN ESNFDVNQNQ YREAYEA-GK WAFVSDYARL KILYDYGGIY

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                      90        100        110        120        130        140        150        160
                                           a.a393 (TgGlt1)
Tg          ADVDMEAIKP LPPCLTVFMG MQRPDAVELG NALIGCSSGH ELIRFILQRV TNDEATNVIE RTGPGLLTRA TLAWLRDQLK
Hh          ADVDMEAIKP LPPCLTVFMG MQRPDAVELG NALIGCSSGH ELIRFILQRV ANDEATNVIE RTGPGLLTRA TLAWLRDQLK
Nc          ADADMEAIRP LPPCLTVFMG MQRPDAVELG NALIGCSSGH ALIRFILRHV PDDEADNAIE RTGPGLLTRA TLLWLRDQLK
Sn          ADVDMELVRK LPNCFTFVAG AQREDAVELG NALLACTPRH ELLTYIITQI TNREAMQVIS STGPGLLTKA VMEWLQTRAP
Cv          VDVDYECVGS LRELCGFFCG AANVGGVELN NGLMGASPGH PFVSVLMDRI PIPSASSPFA TTGPGLLTRA VCRLGGLGSK
Sm          VDVDFECLKP LDVLHSFYCG LSNVGAVEVN NGLFAATAGH PLVSYLCDHL GKPGQAAFLA TTGPGFFTRG LVKGLRRDLG
Vb          VDVDFECLQP MDRLHDFYCG AASVGVFEIN NGLIAAKPGS ALVWYLLERI PNEPFAPFIA TTGPGFFTRG IMRYFRDRFA
Ac          VDVDFKCLRL FQDLLSFITG ISNTDVVELN NGLIACTRNH PIVRELVASL QSNEFMQTIS QTGPGLLTKT FMGWKSGSKK
To          VDIDYLCISP LDDLIHFFCG ASNAGCVELN NGIMACKEGH QILSNMMRSI DGPSPIEVIE HSGPGLLTRE LCRWLVSEGN
Bn          ADVDFECVRN LTGLHSFYAG LSNTGTVEIN NGIFGAAEKH PLCYTLLKNM TGVDTMKIIA RTGPGHFTRV VMNYILFPSD
Ot          LDTDMSGIYS LNDLLDFIIG LSNTKAFELN NAFIASCPGH PLLKHLMETL IPYQKLNIIA VSGPGFMTQQ IFKYLNDNKK
Rf          IDMDFECILP LDVLFSFMIG LSNTSLFEVN NAFMASKPFH PFIKHLWDNL QRHLVSTIWK -TGPVYVSIQ LWKHWNDRWS
Ng          VDVDMECVRP LDDLHSFYAG FSNTGTVELN NGIIGSIPGH PILRQLIDRI RELEAFVTIE ATGPGLFTRG VAALVLQQGG
Sl          IDIDMAVVKK LDELLDFIIG VSNTQAFELN NAIIASIPNH PILQYLIDNL INPKQVNVIQ MSGPGFMTQY VFRYLNSNKT
Pu          ADVDVACVQA FDPLLSFIAG MANTGNVEIS NSVMLSTAHH AILRQLIDTI PTSSAMDTIA RTGPGLLTRT FMGWSTEEGE
Aa          ADMDFEPLRP LDALRDFFVG FSNVGAVEIN NGLIGAAPGH ALLAALVERV SQADADWTIE HTGPGLITRT FC-------P
Es          VDVDFLCLGS FDDLHEFYAG VSNTGTFELN NGLIGCRPGH PIMRDIVNSA -----TETIV KTGPGVFTRA VMAWMVSSAQ
Ap          VDVDFECLKP FDQLHEFYAG LSNTRSVEIN NALIGCVPNH PILKAVISAI NDMPKLGTIA RTGPAVFTVQ IMGWEGHRPA
Gt          VDVDFKCIRS FHDI----LG LSNVGHIEVN NGLIGSAPQH PLLWHLVPST TGTKFVNSFK QTGPGYFTRM MISAWKEKLL
Sc          SDMDTMLLKP IDSWPGLVIG WSEWRRIQFC QWTIQAKPGH PILRELILNI KNVDGSDIMN WTGPGIFSDI IFEYMNNVLD
```

```
Eh        ADMDVKP-RT LSNAALFFID ----SVRSLT QYFMAATAQS PLMESALKAA GVNRFMYIAR TTGPGALKRG FVTFMESNVY
Pp        VDADFECLQP FDDLHPLVHS VLLEDSAALC NALMASAPRH PFWLSVLDNI RERLKSDAVE LTGPRMVKQT YL-------S
Sp        VDCDFESTSP LDDLLPFVHA VLLETDAILC NAVLASRRGH PFWLQVLDAI AGGAKQDPVS LTGPRLVEAV YHEFDAERVA
Mb        VDADSLWLRP LEPLTGLFFG TPQQEPHMYA VGVIGAAPHH PLFGSAIHGI EKLEGMDVME VAGPGYFTDT VFHYMHAFGG
Ob        ADMDMESFKS IDPLSSCFLG LDSNFEHIVI NAIMGCRKNH PGLELVIRNL RVVERVARWQ VTGPDFVTAT LR-------V
Sr        IDADSMWIKS LDPLTGLFFG APSTVTPILA VGVIGCAQHH PFMKKVMESL PDFSHMWNVL STGPHFLTLH YRQYQKNNLD
Ea        ADLDVECLLP SDHLRKAFFG MGTDFDNSVP NAWMASTPGH PAMAHVVRTL RFTQGVARWE ATGPDFATAA LR-------V
Cm        SDTDTVCLQP IRKWPQAMVG WQDMVRQIQV SVWLASTPLH PFFLLLVESV EGPENWSAEA ITGPAKLFYT IGEYSEKEGH
Ss        VDLDFRALRR FDTLKCVVLG MENNLDDAIP NAIMASVPGH PFWLLCVRKV GFRGTETNFS --GSFVLKEA YQDWCRIHNQ
Cs        ADLDFESIRS LEDVRQVVLA EMTEWDQKIP NAWLASTKGH PFWLYCIQQI CAATNTDRWD TTGPVMLFHA VEAYKKVKGS
Bd        ADLDVECLQS HAPI-ASIGG MSRDYEHNIP NPWLGSAPGH PFWIYLLEKI NATLSSQVEV STGPVVLYKA LKEFEQEKML
Pb        MDLDVGCRMP LSRLNGTSAG FLAPSPLGFS NDVMASIPGH PLAQMLIDSL LFTKYPTVFV STGPAFFSYY IGRYVRSLPA
Fp        LDTDVELVRP LDELKGFISM APSPRTLLVN TGSVGAEPGC EMIGKMLAAY FIQETGEPDL RTCTQRDTPL FTKAGLQQKD
Be        MDMDVEVVRP FDDLSPYILG LE--SEKGVE AGVFGAERYS LFLKKILSYY KEDGTFDMRP --LPSVLFEI LTDNYVLRES
Tl        CDTDFECIKP LEKICTMFAG IIFSEKPEIA NGIIGTTPAH PFLSVILEEL LSKDTDTILS VTGPAFLTRT IL-------E
Ts        VDIDMECFKN IEPLVEFFVG TE--DDFYFS NELMGCVPYH ELMCELVEGI KSDYNLTIDE QSGPIYITKY LL-------P
Sh        VDMDFECLRP MAGLKELVLG WNRRLPRIVG NAFMASRPGH PFLHPCAHGV DDSLEEHVRA LADPRIRLHR LPQWDDDDLA
Sn        VDHDVKCFKP FTPFFDLYCG HQPVSSISAN NNIIGSIPRH PILAKCIANV PGDDKDSIIY RVGHRSFDDA VKELGQKSED
Ls        ADIDTECMAP LDVIRIVFSA HRLGMDLVLS NAVMASPKGH PFWPYLCKAM LFTPLDTALT SPSPRFLTHH YWSWFSLSRA
Ps        MDTDMEVCKP LDRLYAFWSG FE--SESQIP TGIMASCRDN DLLEYLLSYY FKNEHGIYDT TTNVTTITRM VKDKYDIELN
```

```
          ....|....| ....|....| ....|....| ....|....| .
                 170        180        190        200
Tg        SNTTEPPIFF YPVPNHRRKE LREGKVQTEH LESHHWRQTW Q
Hh        SNIKKPPVFF YPVPNHRRKE LREGKVQTER LESHHWRQTW Q
Nc        NNVIRPPIFF YPVPNHRRKD LLEGKVQAEL LDSHHWRQTW Q
Sn        DWLRGPPCVF YPVPNHAR-N TLRKRGDEYI KTRHHWKETW K
Cv        ACSVEPASVF CPIPNVLRRE MRHEDGVRRL KEEHHWMGTW Q
Sm        PMLVCPPEVF YPLPNTERGV PDAEFRMLSS LAVHHWFRTW N
Vb        AAASAPPRFF FPTPNTQHRD QHMQPGERHY GECHHWRRTW R
Ac        DVIALPRDYF YSIPNHLS-- --HSVAQEKL DFKHFWAKTW Q
To        RVLVYPAAVF HPFPNNLR-- KFEEFIERD- ---HLWGSSW Q
Bn        SSSDGPCGYF YPFPNSYK-- QFASEGKVMN SKDHLWEASW Q
Ot        HILITPKQFF YPLSNEVRDE --HPYTDRYA YVQHHWGVSW I
Rf        TILVLPTSVL YPFPNHMR-- TPSPPAAGAV PEVHHWARSW Q
Ng        IVLLAPPVWF YPIPNRPV-A YIEYIDQQQY NDSHMWEVNW Q
Sl        NVLIAPKEYF YPLSNELRQN ADFAMIKLPV N--HYWARSW M
Pu        RAIALPIEYF SPLPNRIRHA AKSNAAAAAS DASHHWAKSW Q
Aa        RAVCLPYEIF Y--------- KAGLLIGDGP AVRHLWGRSW Q
Es        EEIFHPPSYL YPVPNNAAAA GQEAIQRART QPGHYWAYSW D
Ap        TVVILPYKAL YPLCNDDTES REDNWQTIDL KTVHLYHMTA H
Gt        EWIVLPVRYF YSLPNNHG-D DVGQMGAQSS DDKHMFSGSW K
Sc        DVMVLPITSF SP-------- RADPLGRKAC FEKTHFSGFN Q
Eh        GTSITPGLYI G-AGDHNRLG SHPPIVQEAC LNEHHWQCTW C
Pp        DMVVFTSEYF YPAYWNIERQ RFDPLGRKAC AWLHHWQCTW C
Sp        SVTVFPEEYF YPAYWNLGRD HTTPAEARQR YPDGLWPALW P
Mb        DLLVQPLHVF R---WEGR-- KLTNIQQKAC KSLHHWIHTY L
Ob        PVTLIPAHIF YPGGWWKSDP YMYQFGYSTG DPLAAIKQCL R
Sr        GVYLAPAEYF FPFFWMRT-- TKKSFNATKC KQLTYWSHTW T
Ea        PKTVFPTFYF FPGGWWGTDS ---GPEEEAC I--HLFGQSW R
Cm        AVEVLPFQYI YPYSWQRDWA ESSNPDEELC LAKTFWTHSW S
Ss        EVRLEPKGVL YPYSWVWESK PHPIFNEDQC KARTYWTHTW -
Cs        GLTILPPGVI YPIDWRRTNP QSGGFDPEEC KKRTYWSHTW E
Bd        PITYIAPEMI LPYDWHNNSA VIFPIVFQIL SALKMAVRNL E
Pb        TFAKIPERIY STNSWHGDDV RMHRTPR--- ---HYYQGSW Q
Fp        GFLVLPTDCF SPFDYVTE-- YLNGKIEVT- ---HHFAASW H
Be        MMYLFPSDYL TA------KS RIYLKELKKK YIKHYWEISW F
Tl        TDVVFPTHYF YPFPNFCREE ------EQNC QANHRAKLLA S
Ts        QVTVVEQEIF FP-------- RHFGLELAGP RWEERWTLNW D
Sh        LSGVAPELYT YPLPRYLDVR KFERNVRRRL VSIFLFGCLV F
Sn        RFALYAHHYY A-STWFEDES EVNSTLDRAK RWLDDWVLDW D
Ls        KQIDVNISIL IPLERCFETL TEDGKIKKTD N--HHFAGSW L
Ps        NNAIFPFEYF CA------KS
```

Sequence IDs:

Tg: *Toxoplasma gondii* (EPR61400.1)

Hh: *Hammondia hammondi* (XP_008887433.1)
Nc: *Neospora caninum* (CEL66241.1)
Sn: *Sarcocystis neurona* (EupathDB:SRCN_1642)
Cv: *Chromera velia* (EupathDB:Cvel_11229)
Sm: *Symbiodinium microadriaticum* (OLP91655.1)
Vb: *Vitrella brassicaformis* (EupathDB:Vbra_8325)
Ac: *Albugo candida* (CCI39677.1)
To: *Thalassiosira oceanica* (EJK66288.1)
Bn: *Bigelowiella natans* (JGI:aug1.98_g20090)
Ot: *Oxytricha trifallax* (J9J850)
Rf: *Reticulomyxa filosa* (X6P0J2)
Ng: *Nannochloropsis gaditana* (W7UA89)
Sl: *Stylonychia lemnae*(A0A078A4I0)
Pu: *Pythium ultimum* (K3WJQ7)
Aa: *Aureococcus anophagefferens*
Es: *Ectocarpus siliculosus* CBJ34172.1
Ap: *Aphanomyces astaci* (XP_009836009.1)
Gt: *Guillardia theta* XP_005836918.1
Sc: *Saccharomyces cerevisiae (Och1)* (P31755)
Eh: *Emiliania huxlei* (R1FXS0)
Pp: *Phytopthora parasitica* (V9EJ41)
Sp: *Saprolegnia parasitca* (A0A067C728)
Mb: *Monosiga brevicollis* (A9UVV5)
Ob: *Octopus bimaculoidus* (A0A0L8GQY7)
Sr: *Salpingoeca rosetta* (F2UBT4)
Ea: *Exophiala aquamarina* (A0A0D2C1M8)
Cm: *Cyanidioschyzon merolae* (M1V7Z2)
Ss: *Surgaldieria sulphuraria* (M2XR17)
Cs: *Coccomyxa subellipsoidea* (I0YXN8)
Bd: *Batrachochytrium dendrobatidis* (A0A177WTF4)
Pb: *Plasmodiophora brassicae* (A0A0G4IJC2)
Fp: *Faecalibacterium prausnitzii* (D4K9D8)
Be: *Bacteroides eggerthii* (R5JR07)
Tl: *Treponema lecithinolyticum* (U2LYB9)
Ts: *Turicibacter sp.* (A0A0X8G3I9)
Sh: *Streptomyces himastatinicus* (D9WNN9)
Sn: *Simkania negevensis* (F8L667)
Ls: *loktanella Sp.* (A0A0Q0ZJY1)
Ps: *Phascolarctobacterium sp.* (A0A0Q0ZJY1

Figure 3.2. Alignment of Glt1-like sequences. The Glt1-like sequence from *Vitrella brassicaformis* (EupathDB V.30) was used to search for related sequences in publicly accessible databases using BLASTP or TBLASTN. All the non-redundant protozoan, metazoan as well as representative bacterial sequences below E^{-5} were screened and 41 representative bacterial, protozoan and eukaryotic sequences were initially aligned with MUSCLE, edited and realigned in BioEdit sequence alignment editor. The top-panel shows Glt1-like sequences from protists that have *Toxoplasma*-like Skp1 modification genes, the middle-panel contains other highest scoring sequences from eukaryotes that may or may not have Skp1-modification pathway, and the bottom-panel shows highest scoring bacterial sequences. Amino acids are color-coded with respect to chemical similarities that were used as the basis for manual alignment giving preference to the registration of hydrophobic residues: green, hydrophobic; blue, acidic, dark red, basic; bright red, structure breaking (P or G). To facilitate comparison of the sequences, positions occupied by identical amino acids across all the organisms are bolded. The position of a conserved glutamate at position 393 of *Toxoplasma* Glt1 and in other Glt1-like sequences is highlighted in yellow. Sequences IDs are given at the bottom.

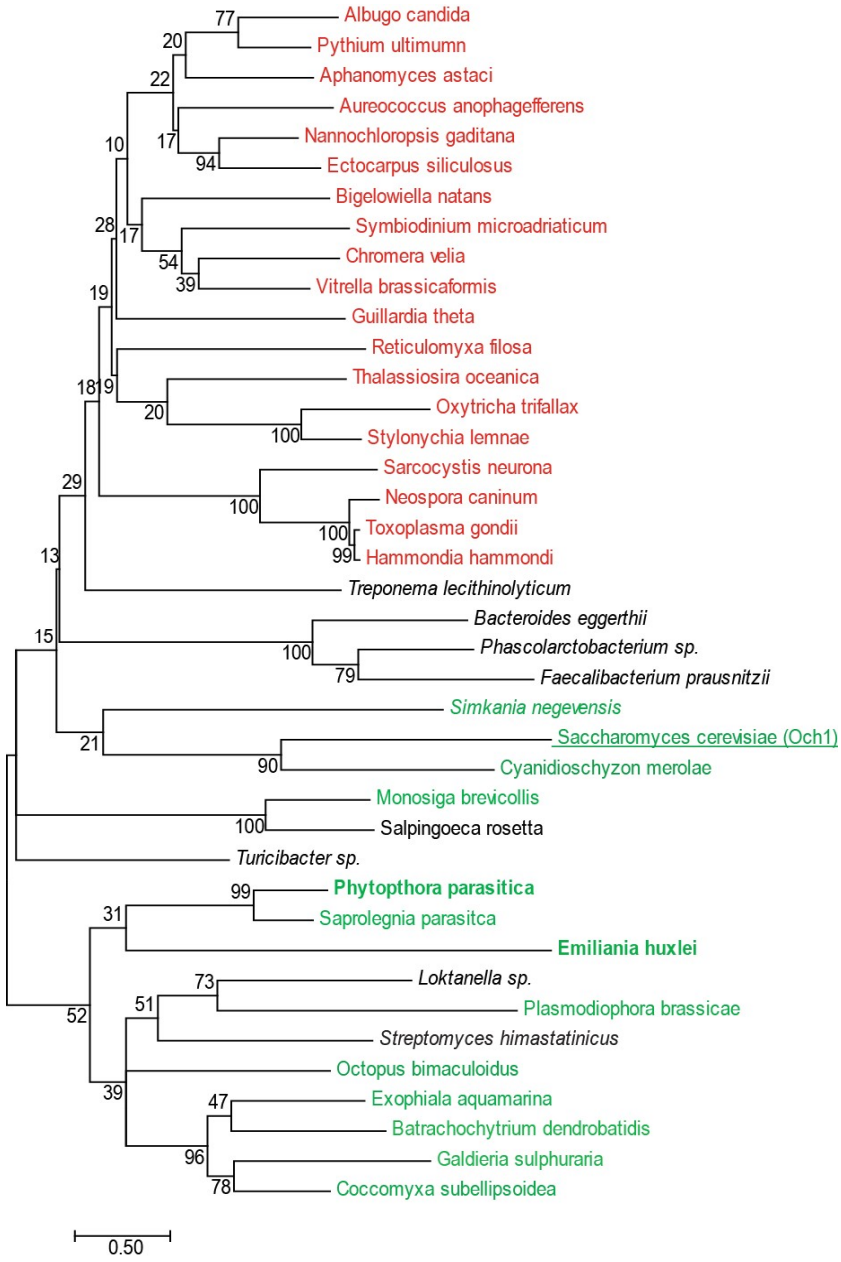


Figure 3.3. Phylogenetic distribution of Glt1. An unrooted tree showing the phylogenetic position of TGGT1_205060. The tree includes highest scoring sequences from Fig. 3.2. Sequences from organisms whose genomes encode *Toxoplasma*-like Skp1 modification genes are in red. Sequences that have predicted signal sequences or transmembrane domains are in green. All the bacterial sequences are in italic. Protists that harbor partial Skp1 modification pathway genes (*phyA* and *gnt1* only) are in bold. The only glycosyltransferase-like sequence whose similarity exceeds the Expect value criterion, and whose enzymatic activity is known, is yeast Och1, a Golgi N-glycan α1,6-mannosyltransferase, is underlined.

Glt1 is Required for TgSkp1 Glycosylation– To determine whether *glt1* is involved in Skp1 glycosylation, we disrupted the gene from RHΔΔ strain by double-crossover homologous recombination as illustrated in **Fig. 3.4A**. *glt1* deletion mutants were confirmed by negative PCR reactions for *glt1*-coding DNA, and positive PCR products for the insertion of the selection marker *hxgprt* between *glt1* flanking sequences, as described in **Fig. 3.4B**. To control for off-target genetic modifications, a complementation construct was made with a version of the original disruption DNA. The construct was transfected into a *glt1* disruption clone-1 to replace the *hxgprt* and complemented strains were counter-selected for loss of *hxgprt* (**Fig. 3.4A**). The same set of PCRs were used to confirm the desired gene restoration in clonal isolates (**Fig. 3.4B**). Previous studies have shown that DXD motif is important for the functional activity of many glycosyltransferases. Glt1 has a highly conserved DXD motif at positions 363 and 365. To test the importance of the expected glycosyltransferase activity of Glt1, a complementation construct was constructed that mutated these two Asp residues to Ala, and in addition mutated the highly conserved Asp348 to Asn (D363A, D365A and D348N) (**Fig. 3.4AB**).

An initial SDS-PAGE and Western blotting analysis showed that Skp1 from parasites lacking *glt1* (lane 3) migrated slightly rapidly than wild-type Skp1 (lane 1 and lane 6) but slower than the Skp1 from parasites whose *phyA* had been disrupted (lane 2) **(Fig. 3.5A)**. Analysis of the complemented strains indicated that normal mobility of Skp1 in SDS-PAGE was restored by complementation with wild-type (lane 4) but not mutant *Glt1* (lane 5). To confirm an effect on Skp1 glycosylation, Skp1 was immunoprecipitated from these samples and its tryptic peptides were analyzed by conventional nanoLC/MS in an Orbitrap Fusion mass spectrometer. Extracted Ion Chromatogram in **Fig. 3.5B** shows that the peptide (^{145}IFNIVNDFT**P**EEEAQVR^{161}) bearing the known modification site, Pro154, was found bearing either the full length pentasaccharide, or

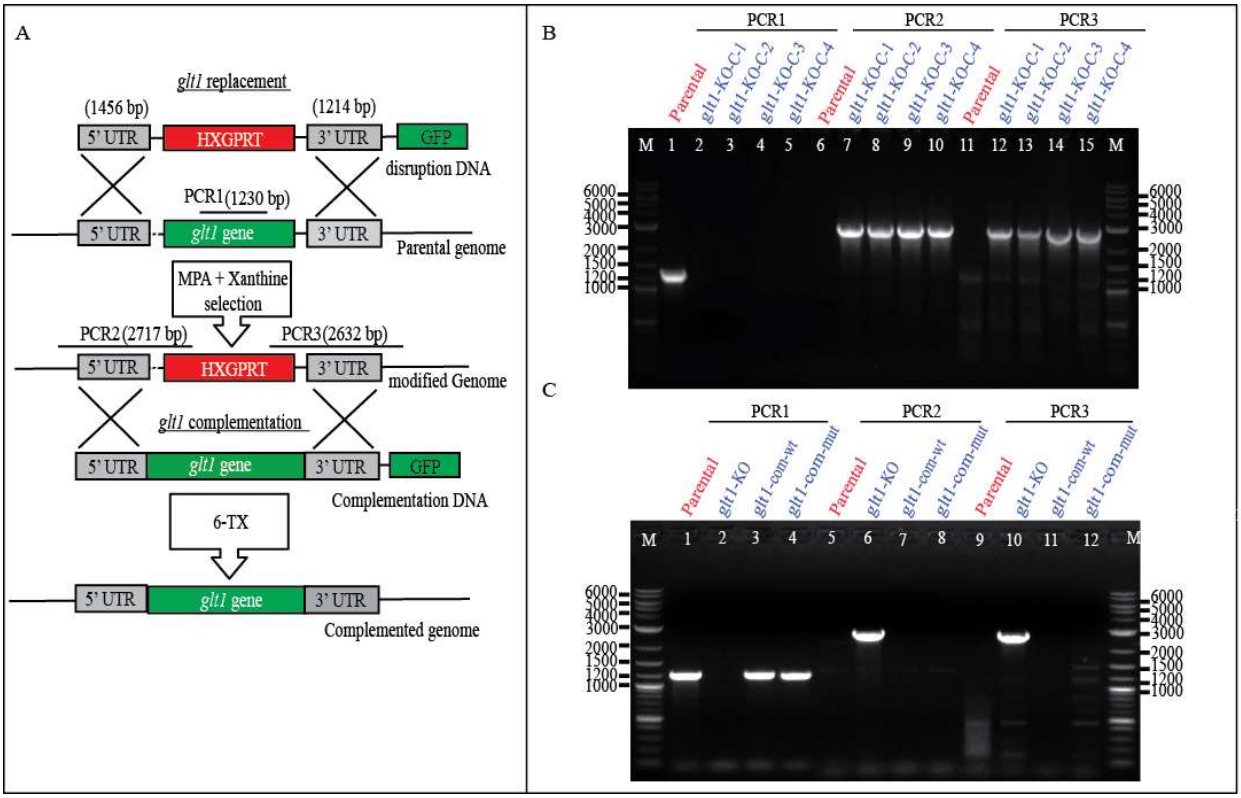


Figure 3.4. Disruption and complementation of Tgglt1. **(A)** Strategy for deletion of *glt1*-coding DNA and its subsequent complementation. The plasmid-derived disruption DNA with homologous targeting sequences was electroporated into parasites. Recovery of *hxgprt*-positive clones that were resistant to mycophenolic acid and xanthine and were GFP-negative were candidates for double crossover gene replacement. **(B)** Gene replacement was confirmed by PCR-1, which demonstrated loss of *glt1*-coding DNA, and PCR-2 and -3 which demonstrated that the inserted *hxgprt* DNA was flanked by neighboring *glt1* DNA. To complement *glt1* in the disruption strain, a plasmid containing a ~5 kb genomic locus including the *glt1* coding region and 5'- and 3'- untranslated regions **(A)** was transfected. Complemented strains where the *hxgprt* is replaced by *glt1* locus were counter selected under 6-thioxanthine. The same strategy was applied to complement with the triple mutant (D363A, D365A and D348N). **(C)** Restoration of wild-type or mutated *glt1* was confirmed by the positive PCR-1, and negative reactions for PCR-2 and PCR-3 which depended on the presence of *hxgprt*. Strain characteristics are summarized in **Table 3.2.**

TABLE 3.2. *Toxoplasma* strains used in this study.

Strain	Parental strain	Genotype	Gene targeted	Selection marker	Selection drug	Ref.
RHΔΔ	RH(I)	Δku80;Δhxgprt	–	–	–	(150)
*phyA*Δ	RHΔΔ	*phyA*Δ;Δku80	*phyA*-exon 1	Hxgprt	MPA/xanthine	(16)
*Glt1*Δ	KU80ΔΔ	*Glt1*Δ;Δku80	*Glt1*-exon 1-5 (all)	Hxgprt	MPA, xanthine	TR
*Glt1*Δ/*Glt1*-Com-Wild	*Glt1*Δ	Δku80;Δhxgprt	–	ΔHxgprt	6-Thioxanthine	TR
*Glt1*Δ/*Glt1*-Com-Mut	*Glt1*Δ	Δku80;Δhxgprt; gtB(D363A, D365A, D348A)	–	ΔHxgprt	6-Thioxanthine	TR

TR: this report

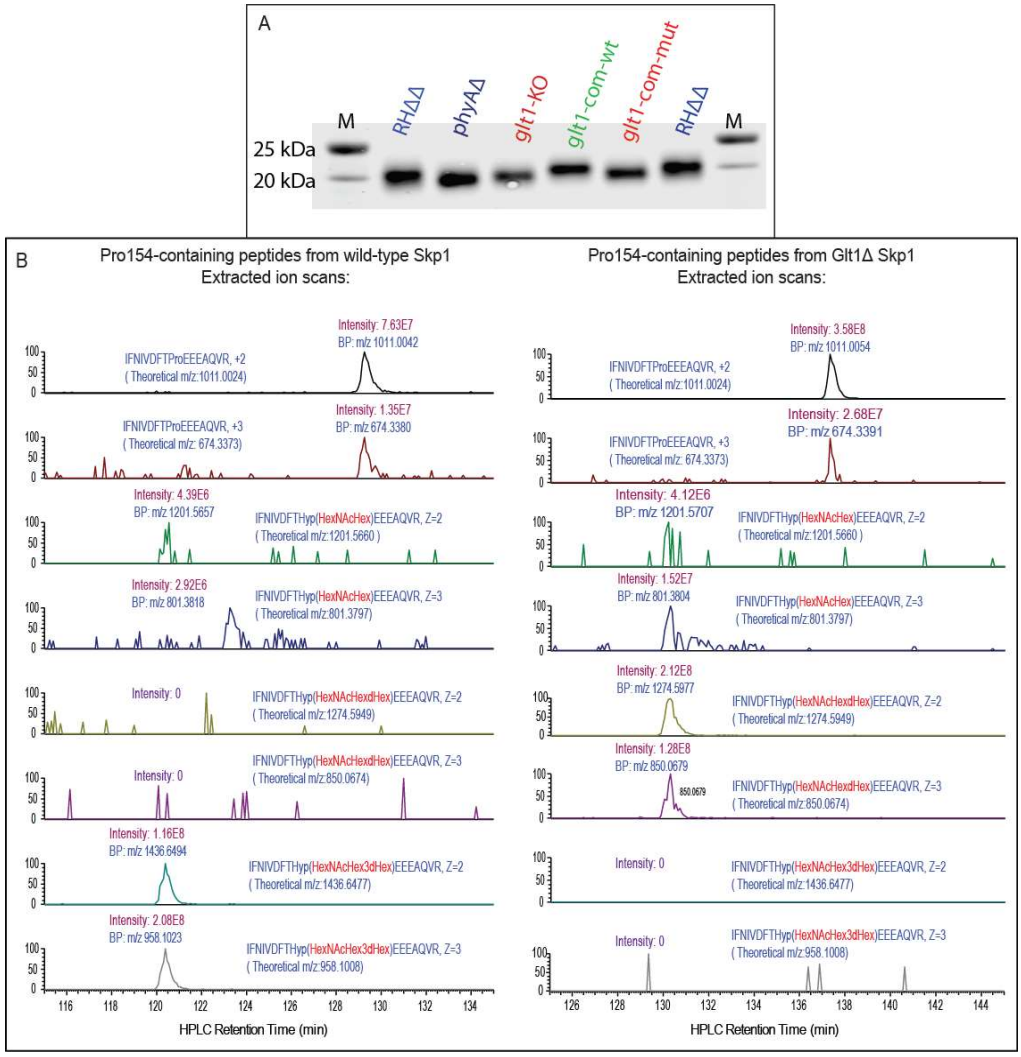


Figure 3.5. Glt1 knock-out parasites accumulate trisaccharide on TgSkp1. **(A)** Disruption of Tg*glt1* affects apparent M_r of TgSkp1. Soluble S16 fractions from equivalent numbers (3×10^6 cells) of parental RHΔΔ, *phyA*Δ and *glt1*Δ or complemented cells were resolved by 4-12% SDS-PAGE, electroblotted, and probed using anti-TgSkp1 (UOK75) antiserum. **(B)** Glt1 is required to add the 4th sugar to the Skp1 trisaccharide. TgSkp1 purified from extracts of RHΔΔ and g*lt1*Δ tachyzoites by immunoprecipitation was trypsinized and analyzed by reverse phase nano-HPLC on an LTQ-XL Orbitrap MS. The extracted ion scans showed *m*/*z*-values consistent with unmodified Skp1 peptide, and a very low abundance of core disaccharide bearing glycopeptide with doubly- and triply-charged ions from both wild-type (left panel) or *glt1*Δ (right panel) samples. In addition, the wildtype sample showed doubly- and triply-charged ions eluting at an earlier position for the fully modified Skp1 glycopeptide with a HexNAcHex3Fuc composition, but no mono-, tri-or tetra-saccharide bearing glycoforms were observed. *glt1*Δ purified extracts showed doubly- and triply-charged ions corresponding to a glycopeptide containing a tri-saccharide with the composition FucHexHexNAc, but no other mono-, tetra- or pentasaccharide bearing glycoforms. See Table 3.3 for a summary of all data. Linda Peng Zhao and Lance Wells are acknowledged for performing this experiment.

no modification, in the parental RHΔΔ strain (**left-panel**). In addition, a very low abundance of the disaccharide bearing form, corresponding to GGn-Skp1, was observed. From *glt1* knock-out parasites we predominantly detected the hydroxy-trisaccharide (dHex-Hex-HexNAc-) form of this peptide, ($[M+2H]^{2+}$, 1274.59; $[M+3H]^{3+}$, 850.06) and completely unmodified form of this peptide as well as a very low abundance of disaccharide modified peptide (**Fig. 3.5B-right panel** and **Table-3.3**). No other glycoforms were observed. The putative trisaccharide-bearing glycopeptide ion (**Fig. 3.6A)** from the *glt1Δ* strain was subjected to MS/MS analysis to confirm its composition and characterize its organization. CID fragmentation of the doubly charged ion yielded a series of ions resulting from sequential loss of a dHex, Hex and HexNAc, representing the loss of core Fuc, Gal and GlcNAc from Skp1 and ultimately resulted in the hydroxypeptide ion (**Fig. 3.6B**). A series of b- and y-ions also confirmed the identity of the peptide (**Fig. 3.6C**). In a clone in which *glt1Δ* was complemented with the original *glt1* genomic DNA, the trisaccharide was replaced with the full pentasaccharide, indicating that loss of the terminal disaccharide was specifically due to disruption of *glt1* (**Table 3.3**). In contrast, g*lt1Δ* strains complemented with the triply mutated (D363A, D365A, D348N) *glt1* genomic DNA failed to rescue glycosylation, demonstrating the importance of its predicted glycosyltransferase activity (**Table 3.3**). These data suggest that the glycosyltransferase activity of Glt1 is responsible for addition of the 4^{th} sugar, previously shown to be a Hex attached to the third sugar, a Fuc residue, but does not show whether the activity acts directly on FGGn-Skp1, or indirectly. The absence of the fifth sugar is expected as it is thought to be anchored to the 4^{th} sugar.

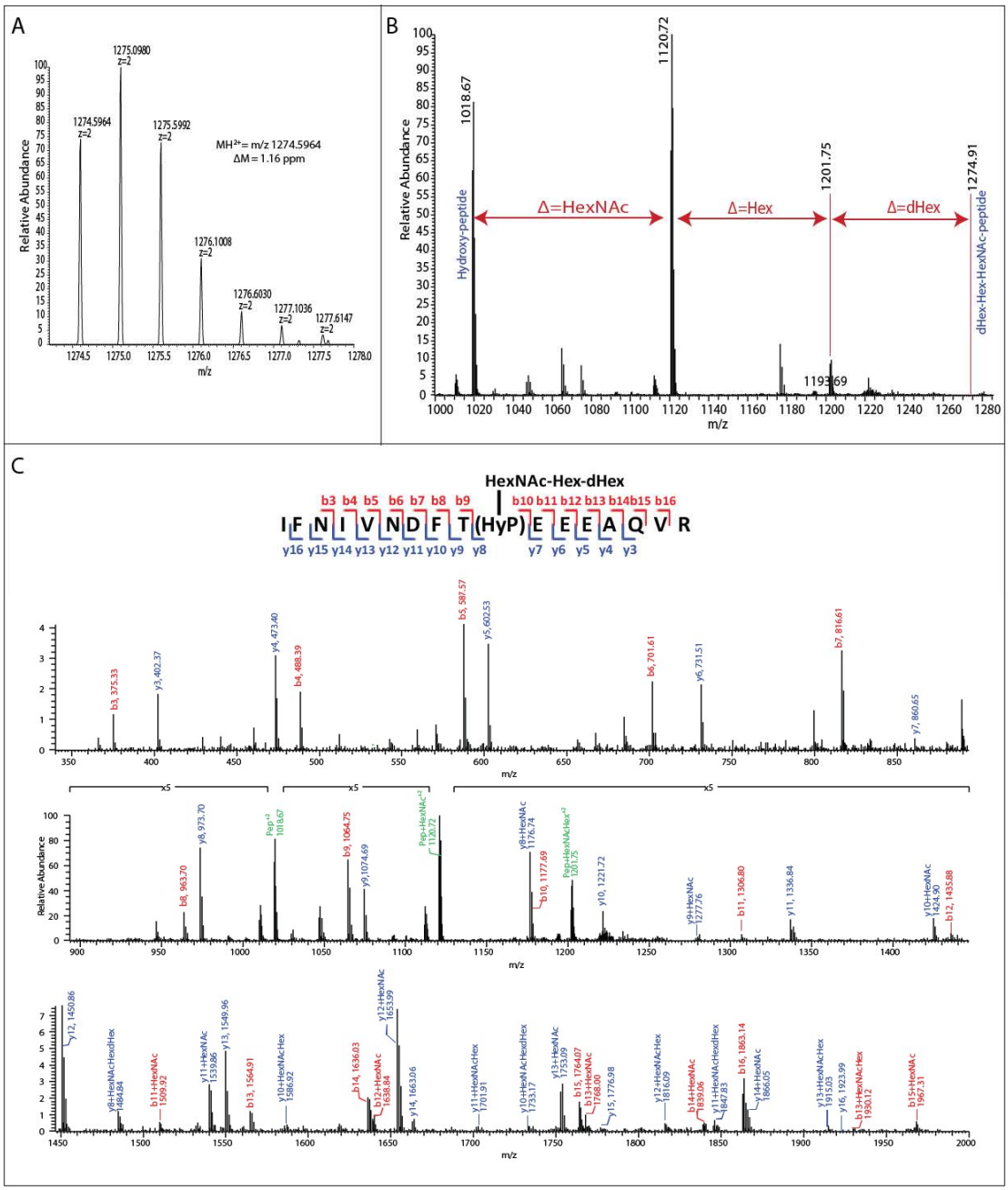


Figure 3.6. MS-MS of doubly charged glycopeptide of endogenous Skp1 from glt1Δ strain. **(A)** Doubly charged ions (m/z 1274.5954) corresponding to HexNAc-Hex-dHex bearing IFNIVFTPEEAQVR. **(B)** CID fragmentation of the doubly-charged precursor ion yields a sequential loss of monosaccharide residues corresponding to dHex, Hex and HexNAc, indicating the presence of a linear trisacchararide. **(C)** Inspection of the full CID fragmentation spectrum shows b- (blue annotations) and y- (red annotations) ion series that match the predicted peptide sequence, as illustrated in the inset, and demonstrate that the glycan is linked via a hydroxylated derivative of Pro154. Peptides with residual sugars are annotated in green. Linda Peng Zhao and Lance Wells are acknowledged for performing this experiment.

Table 3.3. MS detection of Skp1 glycopeptides in strains. Isoforms of the Skp1 peptide ^{145}IFNIVNDFTPEEEAQVR were detected as described in Fig. 3.5 and 3.6. The distribution of raw ion counts among the detected isoforms are shown for the strains analyzed.

Strain	Unmodified peptide			Hyp-HexNAc-Hex peptide			Hyp-HexNAc-Hex-dHex peptide			Hyp-penta-saccharide peptide		
	Occu-pancy*	*m/z***/ ***	ΔM**/ *** (ppm)	Occu-pancy*	*m/z***/ ***	ΔM**/ *** (ppm)	Occu-pancy*	*m/z***/ ***	ΔM**/ *** (ppm)	Occu-pancy*	*m/z***/ ***	ΔM**/ *** (ppm)
*Glt1*Δ	45%	1011.0042/674.3391	1.78/2.67	3%	1201.5630/801.3804	-2.47/0.90	51%	1274.5964/850.0674	1.16/2.06	0%		
*Glt1*Δ/*Glt1*-Com-Wild	27%	1011.0042/674.3380	1.78/1.04	1%	1201.5657/801.3818	-0.22/2.65	0%			71%	1436.6494/958.1023	1.17/1.52
*Glt1*Δ/*Glt1*-Com-Mut	36%	1011.0020/674.3370	-0.40/-0.44	5%	1201.5651/801.3817	-0.72/2.52	59%	1274.5948/850.0659	-0.09/0.30	0%		

* apparent occupancy based on raw spectral counts.
** values refer to $[M+2H]^{2+}$ ions.
*** values refer to $[M+3H]^{3+}$ ions.

Endogenous Glt1 has Skp1 Glucosyltransferase Activity– To characterize the predicted activity of Glt1, exogenous radioactive sugar nucleotide donors and Skp1 glycoforms were introduced into soluble parasite extracts in an attempt to recapitulate reactions in the cell. Since *Toxoplasma* and *Dictyostelium* apparently share the same core trisaccharide on the Skp1 (Chapter 2), and their Skp1 sequences are highly conserved and serve as substrates for each other's PhyA, Gnt1 and PgtA enzymes, we used recombinantly expressed FGGn-*Dd*Skp1 as a surrogate *in vitro* acceptor substrate (113). *In vitro* reactions were performed with cytosolic extracts in the presence of UDP-^{3}H-Glc or UDP-^{3}H-Gal (**Fig. 3.7A**). The parental but not the *glt1*Δ parasite extract catalyzed incorporation of ^{3}H from either UDP-^{3}H-Gal (left panel) or UDP-^{3}H-Glc (right panel) in time- and FGGn-Skp1-dependent fashion, suggesting that the incorporation is mediated by Glt1. Interestingly, UDP-Glc was about six times more efficiently incorporated than UDP-Gal (**Fig. 3.7A)**. To characterize the products, the reaction products were subjected to SDS-PAGE and transferred to nitrocellulose. Regions corresponding to Skp1 as well as neighboring regions of the lane were subjected to acid hydrolysis and analyzed by HPAEC in the presence of internal standard sugars. Radioactivity from both reactions co-chromatographed with glucose (**Fig. 3.7B,** and data not shown). Therefore, Glt1 is inferred to utilize UDP-Glc as its substrate, which could be generated from UDP-Gal in the extract by the action of TGGT1_225880, a putative UDP-Glc 4-epimerase.

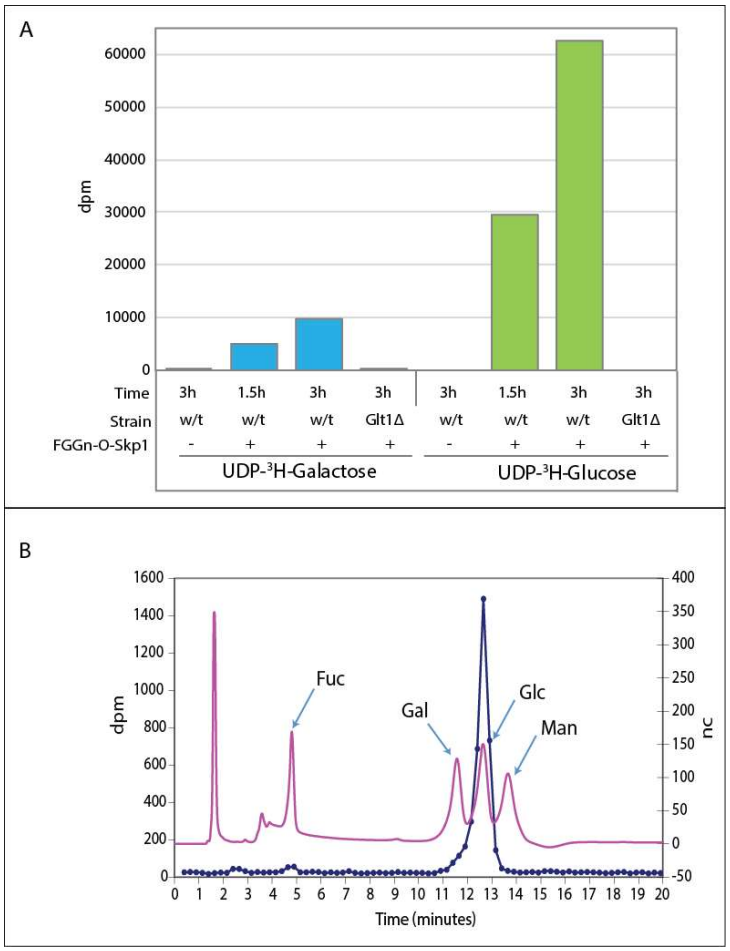


Figure 3.7. Glt1 has Skp1 glucosyltransferase activity- (**A)** Glycosyltransferase activity in *glt1*Δ and normal parasite extracts. Crude desalted cytosolic extracts from normal (RHΔΔ) and *glt1*Δ mutant parasites were incubated with UDP-[^{3}H]Gal or UDP-[^{3}H]Glc with or without FGGn-DdSkp1 for 1.5 or 3 h as described under "Experimental Procedures". Samples were separated on an SDS-PAGE gel, and the DdSkp1 band was excised and analyzed for incorporation of radioactivity. **(B)** High pH anion-exchange chromatography of acid-hydrolyzed reaction products. A reaction containing UDP-[^{3}H]Gal was resolved by SDS-PAGE and transferred to nitrocellulose, and the Ponceau-stained DdSkp1 band was hydrolyzed in 6 M HCl. The supernatant was dried and dissolved in water containing Fuc, Gal, Glc and Man and chromatographed on a Dionex PA-1 column. The elution profile compares the retention times of radioactivity, as determined by analysis of fractions in a scintillation counter, with internal monosaccharide standards as detected by HPEAC-PAD. Similar result was obtained for Glc reaction products (not shown).

Activity of Recombinant Glt1– To determine is Glt1 is capable of directly glycosylating FGGn-Skp1, *Toxoplasma* His_6-Glt1 was prepared in and purified from *E. coli* (sequence in **Supplemental Fig. S3.2**). Although His_6-Glt1 was mostly insoluble, the soluble fraction could be purified over a Ni^{+2}-column and a Q anion exchange column. Western blotting against anti-His_6 monoclonal antibody revealed co-purification of an anti-His_6-reactive band at the expected M_r value of 69.5kDa, and glycosyltransferase activity catalyzing the transfer of [^{3}H]Glc from UDP-[^{3}H]Glc to FGGn-pNP (**Fig. 3.8A**). Activity was time **(Fig. 3.9B)** and enzyme concentration-dependent (not shown), and was stable on ice for at least 96 h. Activity showed a strict requirement for $MnCl_2$ relative to $MgCl_2$ (**Fig. 3.8C**), increased progressively as pH was raised from 6.5 to 8.5 (**Fig. 3.8D**), and was insensitive to NaCl concentration from 0-0.4 M (**Fig 3.8B**). Activity was not saturated by FGGn-pNP up to 50 mM, indicative of a high K_m. For further studies, the activity was assayed at pH 8.0 and 50 mM NaCl. A screening for determination of Glt1 sugar donor specificity shows that Glt1 most efficiently hydrolyze UDP-Glc as expected and fails to utilize UDP-Gal (**Fig. 3.9A**). The enzyme transfers [^{3}H]Glc from UDP-[^{3}H]Glc to the synthetic compounds Fucα1,2Galβ1-Bn and Fucα1,2Galβ1,3GlcNAcα1-pNP in time dependent manner (**Fig. 3.9B**) and shows that the latter, which corresponds to the full length core trisaccharide defined in *Dictyostelium* Skp1, is a substantially better acceptor for His_6-Glt1. GlcNAcα1-Bn and Galβ1,3GlcNAcα1-Bn, representing the core mono- or di-saccharide, respectively, showed no acceptor activity, suggesting that the Glc is transferred directly on the Fuc-terminus of FGGn-Skp1 (**Fig. 3.9C**), consistent with previous MS^2 data (160). Interestingly, enzymatic activity was not detected against Fucα1-pNP, (**Fig. 3.9C**). A comparison of the same FGGn-trisaccharide on pNP or DdSkp1 suggests that the Skp1 carrier renders the trisaccharide to be a 4000-fold better acceptor than pNP, based on the relative concentrations tested (**Fig. 3.9D**). These characteristics indicate

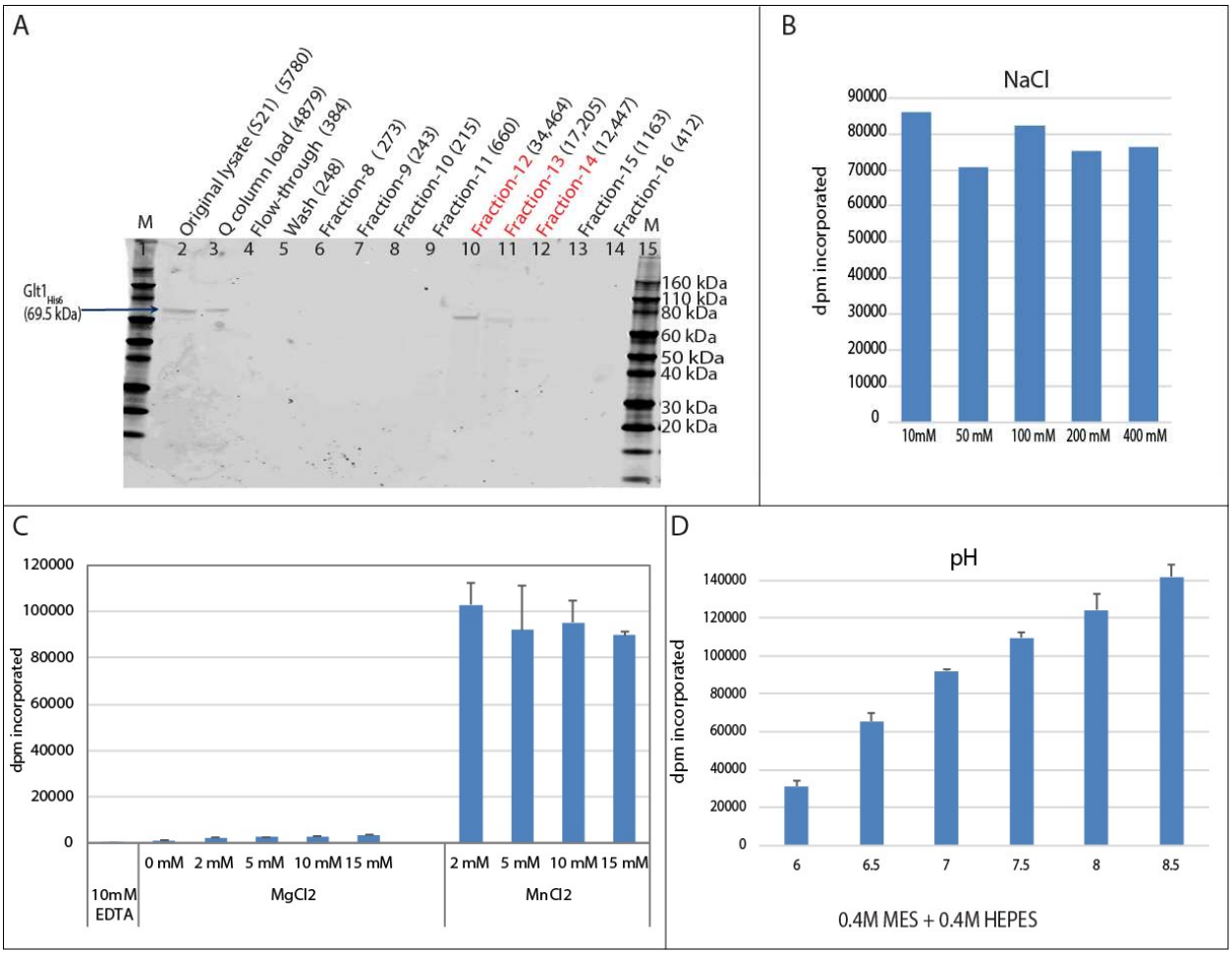


Figure 3.8. Recombinant Glt1 purification and characterization. (A) His_6-tagged TgGlt1 was expressed in *E. coli* and purified over the Ni^{2+} column using a gradient of 0.005-1.0 M imidazole. Fractions were assayed for GlcT activity with 1 mM FGGn-pNP as acceptor. Active fractions were pooled, dialyzed and applied to a Q-Sepharose column, which was eluted with a 0-1 M gradient of NaCl. Fractions were assayed for activity and Western blotted for the presence of His_6-Glt1, whose expected M_r is 69,500. The panel shows the expected Glt1 band in the original *E. coli* lysate (lane-2), the Q column load (lane -3), and in Q column fractions 12 -14. Enzymatic activity is shown in dpm above the lanes inside brackets. (**B-D**) 1 µl of fraction 12 was assayed for glucosyltransferase activity using 2 mM FGGn-pNP for 1 h at varying concentrations of NaCl in the presence of 2 mM $MnCl_2$ at pH 8 (panel B), varying concentrations of divalent cations in the presence of 50 mM NaCl at pH 8 (panel C), and varied pH values in the presence of 2 mM $MnCl_2$ and 50 mM NaCl (panel D). Error bars indicates standard deviation of the mean between duplicate samples.

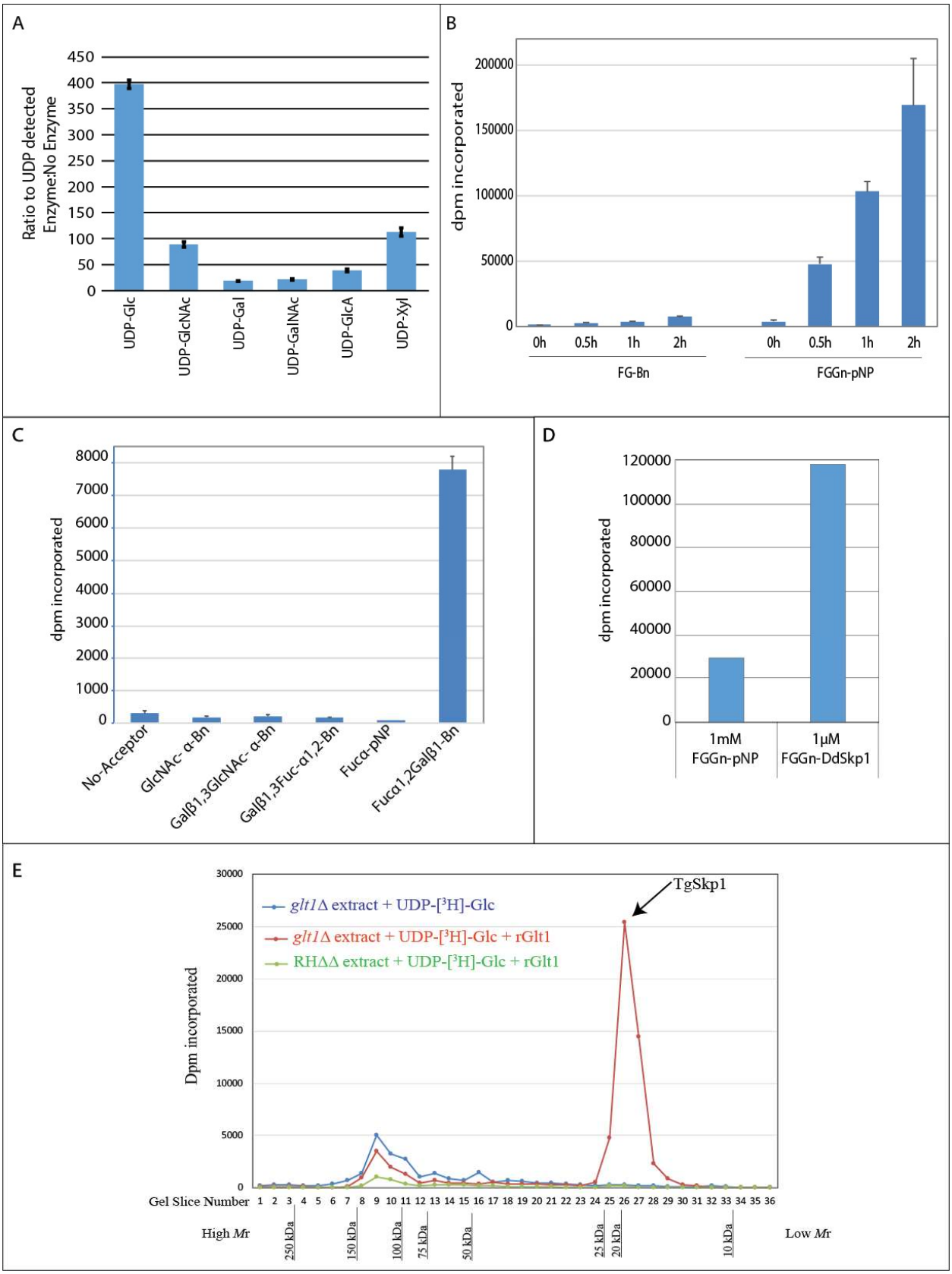

A
Ratio to UDP detected Enzyme:No Enzyme
450
400
350
300
250
200
150
100
50
0
UDP-Glc
UDP-GlcNAc
UDP-Gal
UDP-GalNAc
UDP-GlcA
UDP-Xyl
B
dpm incorporated
200000
150000
100000
50000
0
0h
0.5h
1h
2h
FG-Bn
FGGn-pNP
C
dpm incorporated
8000
7000
6000
5000
4000
3000
2000
1000
0
No-Acceptor
GlcNAc-α-Bn
Galβ1,3GlcNAc-α-Bn
Galβ1,3Fuc-α1,2-Bn
Fucα-pNP
Fucα1,2Galβ1-Bn
D
dpm incorporated
120000
100000
80000
60000
40000
20000
0
1mM FGGn-pNP
1μM FGGn-DdSkp1
E
Dpm incorporated
30000
25000
20000
15000
10000
5000
0
glt1Δ extract + UDP-[3H]-Glc
glt1Δ extract + UDP-[3H]-Glc + rGlt1
RHΔΔ extract + UDP-[3H]-Glc + rGlt1
TgSkp1
Gel Slice Number 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
High Mr
250 kDa
150 kDa
100 kDa
75 kDa
50 kDa
25 kDa
20 kDa
10 kDa
Low Mr

Figure 3.9. Activity of His_6-Glt1 toward UDP-sugars and synthetic or endogenous acceptors. **(A)** Comparison of glycosyl-donor specificities of His_6-Glt1. Recombinant Glt1 was incubated with 50 μM of each UDP-sugar (shown at the bottom) at 37°C for 16 h, and released UDP was detected by using UDP-Glo^{TM} assay as described in the "Experimental Procedures". Data are represented as the ratio of the UDP detected from the Glt-containing reactions to the respective negative controls, which have all reaction components except the enzyme. Data are representative from three trials, and error bars represent standard deviation of the mean (Osman Sheikh, Msano Mandalasi, and Lance Wells are thanked for their assistance in performing this assay). **(B-D)** Recombinantly expressed His_6-Glt1 was incubated in 2.0 mM synthetic glycoside for 1 h in the presence of 2 μM UDP-[^{3}H]Glc. Transfer of [^{3}H] to the glycoside was assayed by absorption to a C_{18} Sep-Pak cartridge and elution with methanol. **(B)** Time-dependent incorporation of [^{3}H] in di- or tri-saccharide conjugates. **(C)** A series of Bn conjugates. **(D)** His_6-Glt1 was incubated with the indicated concentrations of FGGn-pNP or FGGn- DdSkp1 in the presence of 2 μM UDP-[^{3}H]Glc, and ^{3}H incorporation was counted after recovery from a C18 Sep-Pak cartridge for FGGn-pNP, or after TCA-precipitation for FGGn-*Dd*Skp1, as described in "Experimental Procedures". **(E)** Biochemical complementation of parasite extracts with His_6-Glt1. 2μl of His_6-Glt1 was added in the reactions containing 2 μCi of UDP-^{3}H-Glc, 170 μl of desalted RHΔΔ or *glt1*Δ parasite extracts and incubated for 3h at 37°C. Reactions without His_6-Glt1 was used as the negative control. Concentrated samples were then loaded onto SDS-PAGE gel and top to bottom of each lane was sliced into multiple pieces and incorporation of radioactivity was measured as described in the "Experimental procedures". Incorporation in the Skp1 band position was observed only for the *glt1*Δ parasite extracts incubated with exogenous His_6-Glt1 (red line).

that Glt1 directly catalyzes the addition of αGlc to the non-reducing terminus of the Skp1 trisaccharide in cells. To test if Glt1 has other cellular targets beyond Skp1, His_6-Glt1 was added in the reactions containing UDP-^{3}H-Glc and desalted wild-type or *glt1*Δ parasites extracts and ^{3}H incorporation into any other endogenous proteins was investigated by analysis of the entire SDS-PAGE gel as described in chapter 2. As shown in **Fig. 3.9E**, similar incorporations into several high M_r band positions were observed in the presence or absence of His_6-Glt1 suggesting that incorporation into those positions are Glt1-independent. Interestingly, in the presence of His_6-Glt1, a large incorporation was observed in the Skp1 band position only for the *glt1*Δ parasite extracts but not for the wild-type parasite extracts suggesting that the endogenous FGGn-Skp1 in the *glt1*Δ parasite extract was modified by the exogenous His_6-Glt1, which was already modified in the wild-type parasite extracts by the endogenous Glt1 and thereby no incorporation was observed. This data suggest that Skp1 might be the sole target of Glt1.

Glt1 is Important for Toxoplasma Proliferation- Previous studies revealed that disruption of *phyA, gnt1* and *pgtA* result in a parasite growth defect, which could be detected as reduced plaque areas after 5 d of replication on a fibroblast monolayer. To examine the role *glt1*, the plaque-forming abilities of the disruption strains described above were analyzed. As shown in **Fig. 3.10**, the average plaque areas of *glt1*Δ cells are reduced in comparison to that of parental (RHΔΔ) cells, but larger than that of *phyA*Δ strains that lack the entire modification on TgSkp1. Both differences were statistically significant at $p<0.015$. Genetic complementation with the original *glt1* sequence at the same locus restored normal growth, verifying that the growth defect in the original disruption strain was due to the disruption of *glt1*. Furthermore, complementation with a mutated version of *glt1*, consisting of 3 point mutations that were shown to inactivate the glucosyltransferase activity

of His_6-Glt1 expressed in *E. coli* (data not shown), failed to rescue the growth defect of *glt1*Δ cells. Therefore, the slow growth of the strain can be attributed to loss of the enzymatic activity itself rather than another potential function of Glt1.

Discussion

Glt1 is a Novel TgSkp1 Modifying Glucosyltransferase- Here we have shown that Glt1, an enzyme from CAZy GT32 family of retaining glycosyltransferases, mediates addition of the fourth sugar on TgSkp1. This interpretation is based on (a) the mass spectrometric analysis of Skp1 that shows that *glt1* knock out and its catalytically dead mutants accumulate trisaccharide while the parental and Glt1 complemented strains accumulate pentasaccharide (**Fig 3.5, 3.6**) (**Table 3.3**), (b) enzymatic assays with parasite cell extracts, which show transfer of [^{3}H]Glc from UDP-[^{3}H]Glc to exogenous FGGn-DdSkp1 in a Glt1-dependent fashion, and (c) the ability of recombinant His_6-Glt1 to efficiently catalyze the addition of the fourth sugar to the Skp1 trisaccharide (FGGn-Skp1) and synthetic glycan models. Glt1 is concluded to be a glucosyltransferase based on (a) its preference for UDP-Glc, relative to all other UDP-sugars tested, in a sugar nucleotide hydrolysis assay, and b) its efficient utilization of UDP-Glc as a donor substrate for various synthetic acceptors as well as FGGn-Skp1 using either the cell extract or the

recombinant enzyme. This conclusion explains the Hex-dHex-Hex-HexNAc- found by MS analysis (160), and is consistent with the inability of green coffee bean α-galactosidase to mediate its removal (Chapter 4). The Glt1 protein sequence is related to the CAZy GT32 family, and the loss of enzymatic activity of point mutants that inactivate other members of this family supports this association. This family includes a number of characterized retaining glycosyltransferases,

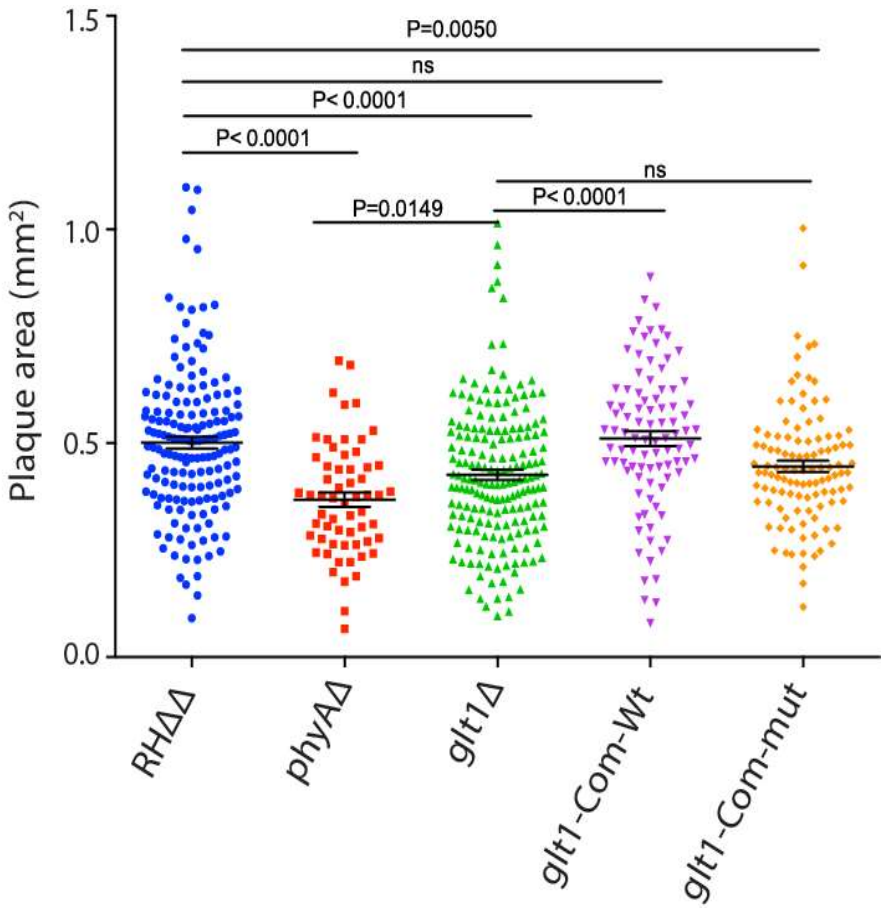


Figure 3.10. Role of TgGlt1 in parasite proliferation. BJ HFF monolayers were inoculated with parental, Tg*phyAΔ*, Tg*glt1Δ* and complemented strains at an MOI of 0.002. After 6 days, monolayers were stained with crystal violet. **(A)** The images were digitized and areas of manually encircled plaques (~80 plaques) formed by the individual strains were measured by the imageJ software. Average plaque areas ± SEM from a representative experiment are reported here.

that use varied sugar nucleotide donors including UDP-Gal and GDP-Man, and catalyze α1,3, α1,4, or α1,6 linkages, though none are known to us to utilize UDP-Glc. This family association indicates that the Glc is α-linked to the Fuc, which to our knowledge is an unprecedented linkage in eukaryotic proteins. The corresponding structure in *Dictyostelium* Skp1 consists of Galα1,3Fuc-, which is mediated by the unrelated AgtA galactosyltransferase from the CAZy GT77 family. Further studies will be required to determine the position on Fuc to which the Glc residue is linked.

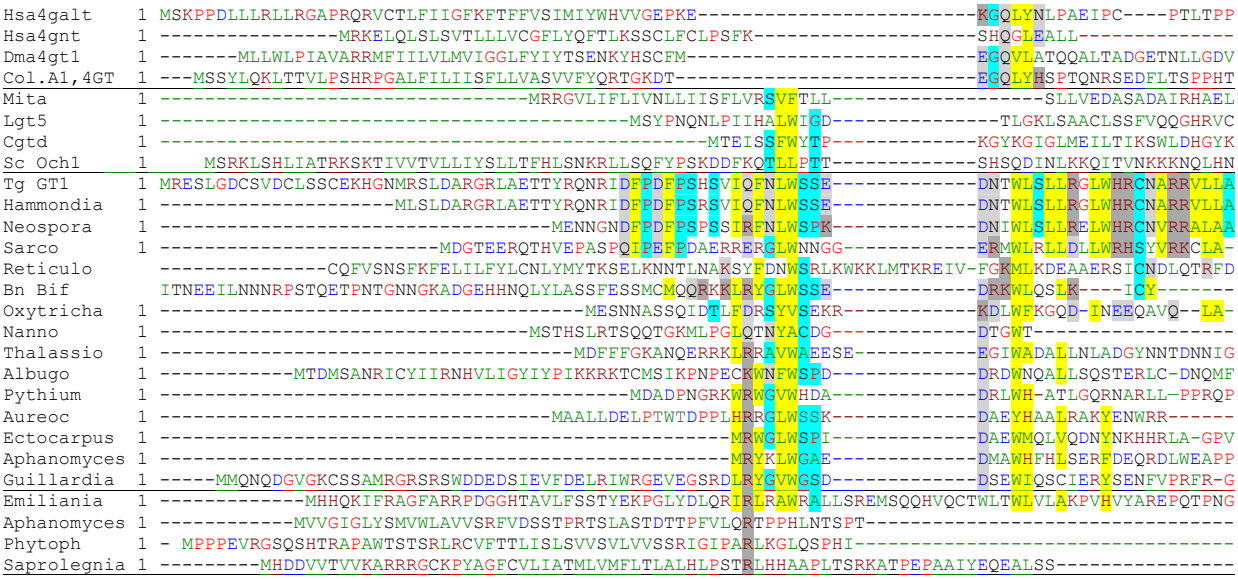
Skp1 is Natural Substrate of Glt1- Skp1 appears to be a natural substrate of Glt1. The enzyme is colocalized with Skp1 in the cytoplasmic compartment based on the absence of a detectable signal sequence or transmembrane domain (**Fig. 3.1**, **Supplemental Fig. S3.1** and **3.2**), and the presence of its enzymatic activity in the cytosolic fraction of the parasites prepared by hypotonic lysis (**Fig. 3.7A**). FGGn-Skp1 is an excellent substrate with a catalytic efficiency that is ~ 4,000-fold better than that of synthetic acceptor FGGn-pNP (**Fig. 3.9D**), which itself is a dramatically better acceptor than FG-Bn or F-Bn. The interpretation that the Glc is linked to Fuc is consistent with the observation that synthetic glycans that represent other motifs in the FGGn-trisaccharide, including GlcNAcα-Bn, Galβ1,3Fucα1,2-Bn and Galβ1,3GlcNAcα1-Bn, are not acceptors. Substantial biochemical and genetic evidence indicate that, in *Dictyostelium*, FGGn-DdSkp1 is the sole target of the AgtA enzyme that mediates the addition of its fourth sugar, a conclusion that has been reached for each of the other enzymes in the pathway as well. Glt1 also appears to solely target Skp1 (**Fig. 3.9E**), suggesting that Glt1, along with the other enzymes, are dedicated to Skp1 as in *Dictyostelium*.

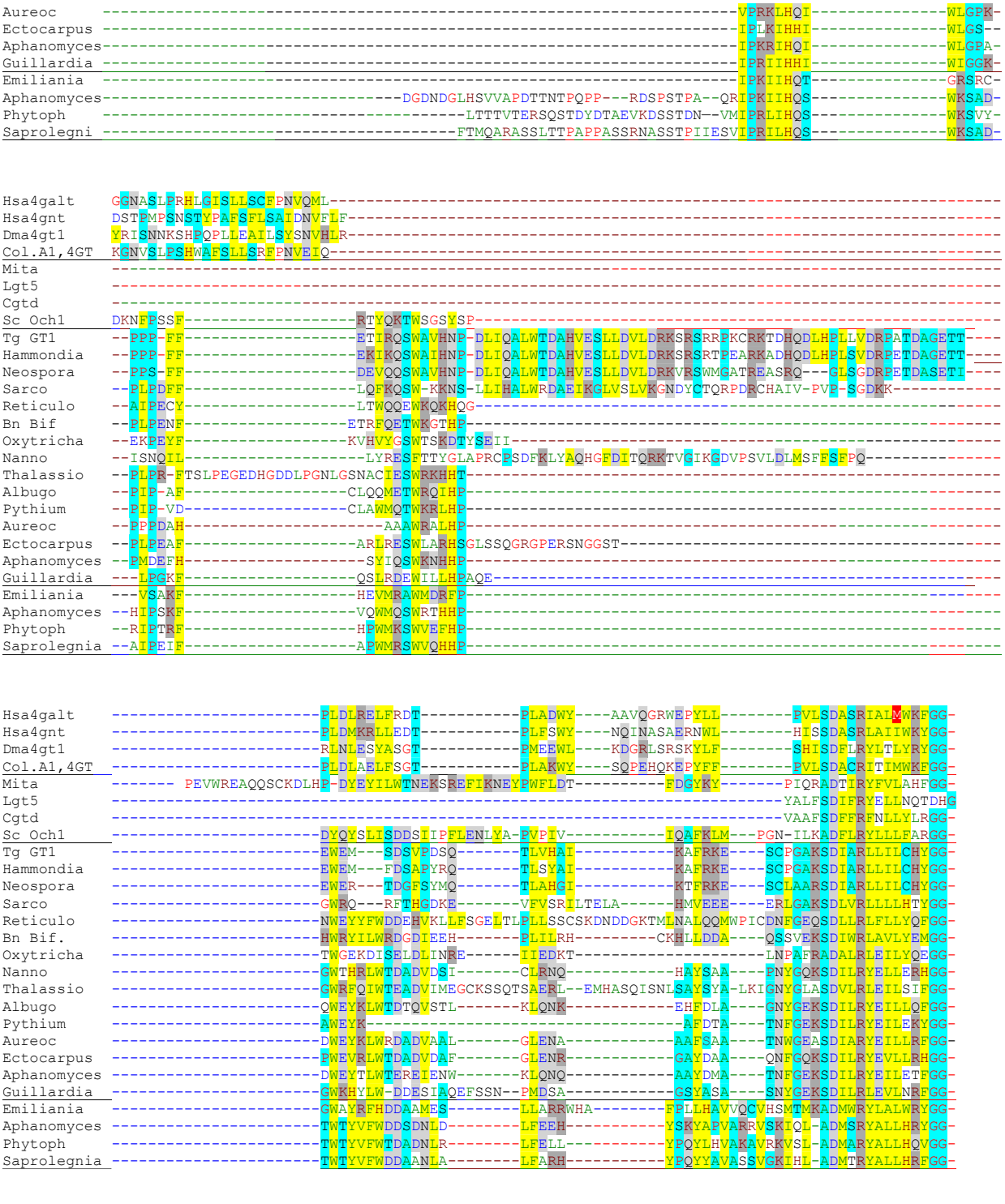
Glt1 is Selectively Distributed in the Wide Range of Protist Groups - The broad phylogenetic distribution of Glt1 in the organisms that have *Toxoplasma-like* Skp1 modification pathway **(Fig. 3.3)** suggests its presence in the common eukaryotic ancestor and selective loss or modifications in the successors where the pathway deemed unnecessary. BlastP search in *Dictyostelium* Database (Dictybase) shows no evidence of any putative homologues even at $E<10^{-1}$ in the sequenced amoebazoans suggesting that the gene might have completely been lost from this group to avoid the potential competition between AgtA and Glt1 for the same substrate (i.e., FGGn-Skp1). Related GT32A sequences were found in protists *Eimiliania huxlei* and *Phytophthora parasitica* that have only *phyA and gnt1* genes of the pathway and were grouped closely but far from the putative Glt1 clade in the phylogenetic tree. Interestingly, all these sequences predicted to have signal sequences or transmembrane domains whereas all the Glt1-like sequences are predicted to be cytoplasmic. In addition, a scrutiny of the sequence alignment (**Fig 3.2** and **Supplemental Fig. S3.1**) shows that all these sequences lack the glutamate at position at 393 which is conserved only in the Glt1-like sequences and in a bacterium, *Treponema lecithinolyticum.* Although the glutamate is conserved in the bacterium *Treponema lecithinolyticum*, its position in the out of Glt1-specific clade as well as the absence of any other Skp1-modification pathway genes in this bacterium indicates that it might be obtained via horizontal gene transfer and have other functional target(s). A potential role of this conserved glutamate in Glt1-specific function needs further investigation.

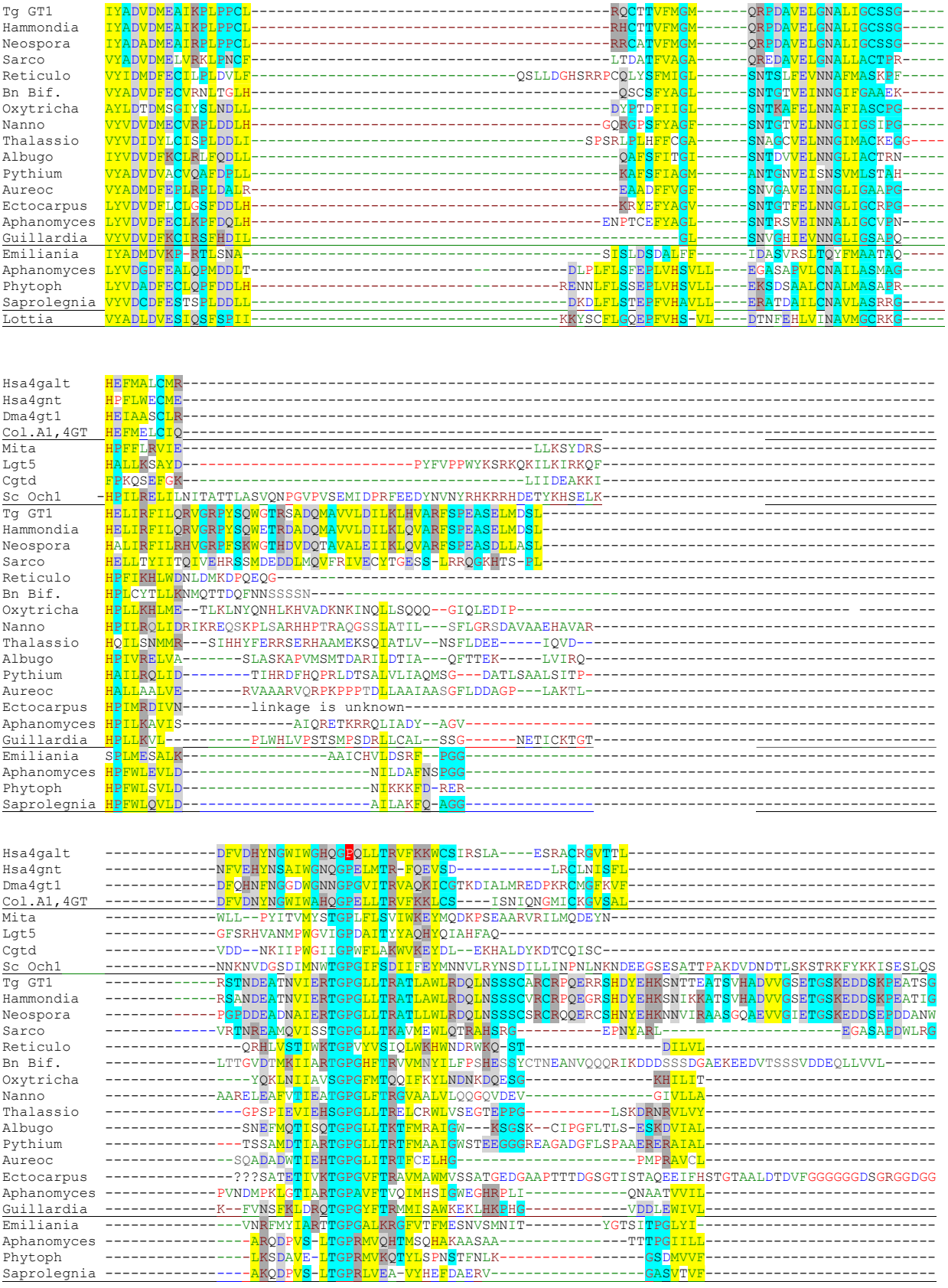
*Glt1 has Important Roles in Parasite Proliferation- glt1*Δ parasites have a growth defect that is intermediate between that of *phyA*Δ and wild-type parasites. Thus full glycosylation of Skp1 may be important to mediate the growth advantage afforded by the action of phyA, as described

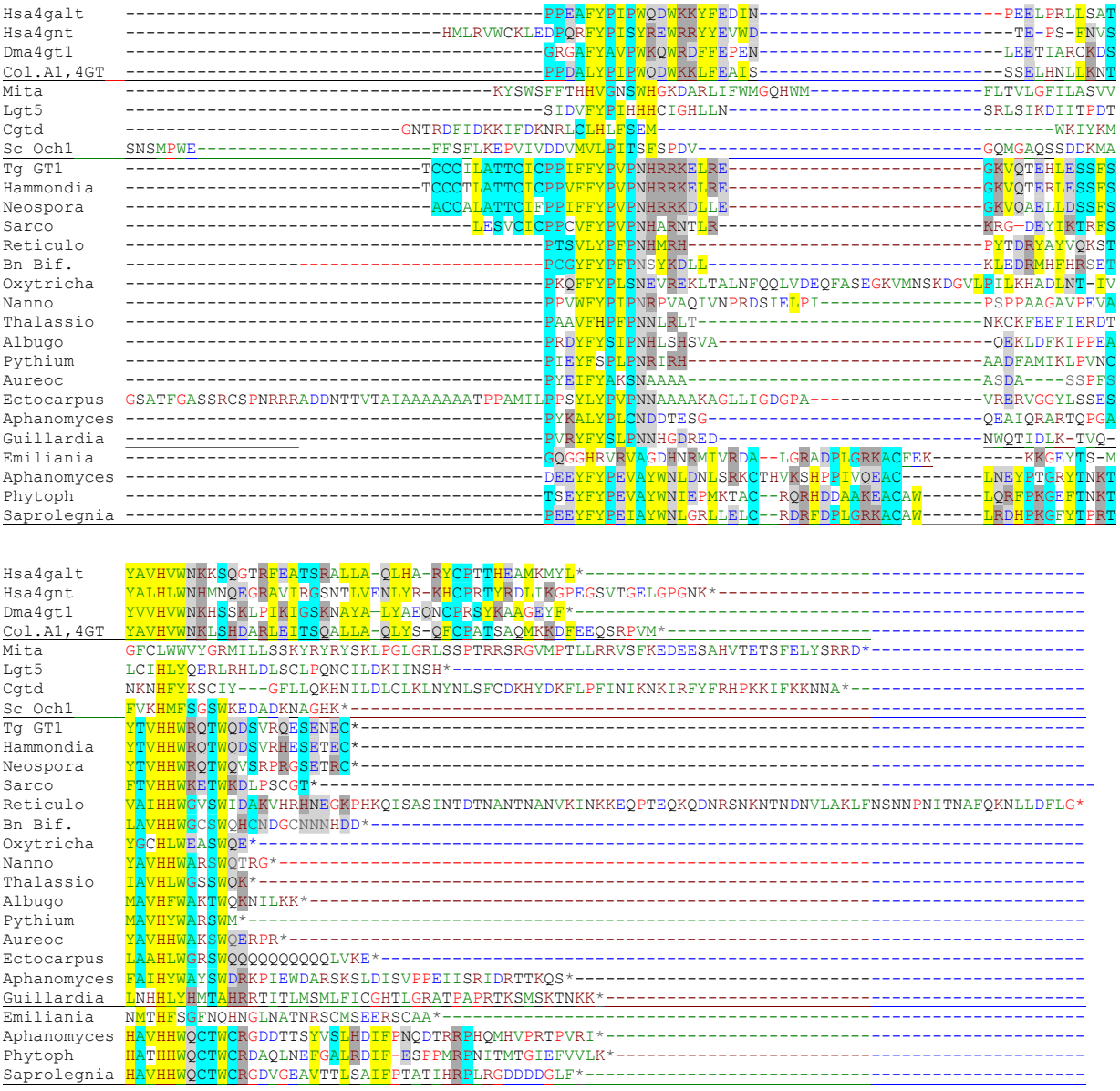
in *Dictsyostelium* AgtA-KO, which is also important for *Dictyostelium* culmination into fruiting bodies (169). The mechanism of this effect is under further investigation.

Functional Variations Between Toxoplasma Glt1 and Dictyostelium AgtA- Although Glt1 and AgtA both show superior preferences for FGGn-Skp1 over the synthetic free glycans, interestingly, the preference for the free glycans of DdAgtA is opposite of Glt1 and that is: F-pNP>FG-Octyl>FGGn-pNP (113). Whereas the C-terminal WD-40 domain of AgtA preferentially binds with the trisaccharide modified Skp1 for dampening the functions of unmodified Skp1 and act as an F-box exchange factor (113), Glt1 apparently lacks any additional domain for similar functions. Therefore, it is not clear if similar roles are important in *Toxoplasma* or the parasite is regulating similar functions through other unknown mechanisms. It might be possible that unlike *Dictyostelium,* instead of maintaining another round of regulation on glycosylation at this final stage, *Toxoplasma* might regulate this mechanism in the very earlier stage. Significant levels of either modified or unmodified Skp1 peptide in all the strains support this hypothesis, as PhyA activity may be the rate limiting factor for sugar extension. Minor abundance of disaccharide in all the strains does not oppose the hypothesis as it might be due to a reduced rate of fucosylation resulting from inefficient transfer from Gal-Tase domain to the Fuc-Tase domain of TgPgtA for the same Skp1 docking site as observed before for DdPgtA (109). Although there is no supporting data yet, it might also be possible that presence of glucose instead of galactose at the 4^{th} position of *Toxoplasma* glycan may regulate its function too. For detailed understanding on the role and regulation of distinct glycans in Skp1 function, a gene swapping experiment is warranted through expressing Glt1 in AgtA-KO *Dictyostelium* and in vice versa expressing AgtA in Glt1-KO parasites.


Hsa4galt 1 MSKPPDLLLRLLRGAPRQRVCTLFIIGFKFTFFVSIMIYWHVVGEPKE
Hsa4gnt 1 MRKELQLSLSVTLLLVCGFLYQFTLKSSCLFCLPSFK
Dma4gt1 1 MLLWLPIAVARRMFIILVLMVIGGLFYIYTSENKYHSCFM
Col.A1,4GT 1 MSSYLQKLTTVLPSHRPGALFILIISFLLVASVVFYQRTGKDT
Mita 1 MRRGVLIFLIVNLLIISFLVRSVFTLL
Lgt5 1 MSYPNQNLPIIHALWIGD
Cgtd 1 MTEISSFWYTP
Sc Och1 1 MSRKLSHLIATRKSKTIVVTVLLIYSLLTFHLSNKRLLSQFYPSKDDFKQTLLPTT
Tg GT1 1 MRESLGDCSVDCLSSCEKHGNMRSLDARGRLAETTYRQNRIDFPDFPSHSVIQFNLWSSE
Hammondia 1 MLSLDARGRLAETTYRQNRIDFPDFPSRSVIQFNLWSSE
Neospora 1 MENNGNDFPDFPSPSSIRFNLWSPK
Sarco 1 MDGTEERQTHVEPASPQIPEFPDAERRERGLWNNGG
Reticulo CQFVSNSFKFELILFYLCNLYMYTKSELKNNTLNAKSYFDNWSRLKWKKLMTKREIV-FGKMLKDEAAERSICNDLQTRFD
Bn Bif ITNEEILNNNRPSTQETPNTGNNGKADGEHHNQLYLASSFESSMCMQQRKKLRYGLWSSE
Oxytricha 1 MESNNASSQIDTLFDRSYVSEKR
Nanno 1 MSTHSLRTSQQTGKMLPGLQTNYACDG
Thalassio 1 MDFFFGKANQERRKLRRAVWAEESE
Albugo 1 MTDMSANRICYIIRNHVLIGYIYPIKKRKTCMSIKPNPECKWNFWSPD
Pythium 1 MDADPNGRKWRWGVWHDA
Aureoc 1 MAALLDELPTWTDPPLHRRGLWSSK
Ectocarpus 1 MRWGLWSPI
Aphanomyces 1 MRYKLWGAE
Guillardia 1 MMQNQDGVGKCSSAMRGRSRSWDDEDSIEVFDELRIWRGEVEGSRDLRYGVWGSD
Emiliania 1 MHHQKIFRAGFARRPDGGHTAVLFSSTYEKPGLYDLQRIRLRAWRALLSREMSQQHVQCTWLTWLVLAKPVHVYAREPQTPNG
Aphanomyces 1 MVVGIGLYSMVWLAVVSRFVDSSTPRTSLASTDTTPFVLQRTPPHLNTSPT
Phytoph 1 - MPPPEVRGSQSHTRAPAWTSTSRLRCVFTTLISLSVVSVLVVSSRIGIPARLKGLQSPHI
Saprolegnia 1 MHDDVVTVVKARRRGCKPYAGFCVLIATMLVMFLTLALHLPSTRLHHAAPLTSRKATPEPAAIYEQEALSS
Hsa4galt TPPSHGPTPGN-IFFLETSDRT
Hsa4gnt SHRRGIVFLETSERM
Dma4gt1 LQADPKPSPGNSIFFHETSCRLSENRQLETLKVTARQACAIESAAMH
Col.A1,4GT IAGGSPPSPGD-VFFVETSERI
Mita PSPN-SSLIEQRPQI
Lgt5 LHAYTPISDVPSGIDICDANLI
Cgtd FHLYTYNLEDKIFLKFQELFDNFILKDANEI
Sc Och1 LRDQLSFAFPYDSQAP
Tg GT1 AS-GDPREPAHLSEEVTSKRHRLPGVATPRLNCDSELLKCYKNGGDLKGVDHHCPQPA PADSRGDRDGEAFCADRALSAS
Hammondia AS-GDFREQAHLSVEATSTRHMLPGVATPRLNCDSELLMWYENGGDLKGVDHHRPQQA PADSRGDRGGEAFCADRQLSAS
Neospora ASKAATSKAAASKAAACGDFPHDAQLSVNATMKSHRMSIAATPWFRYDTDFPACHNGGG-MKG-EHQRRRRA TADVRDDSAREAVCADRALNAS
Sarco RDALLGHVRTAGFVRTTHSLEHGSAPG LPVV QEHPAHQRDNRMQTADADALYLRPLMRA-IASAAS
Reticulo EQFKSRVDDDKHVMNNTINHNDDDDDDDG-KSWTD
Bn Bif DKWCCHQRRVLIEKDDDRTMESTKRMI
Oxytricha KDQFKNIVIDHLSSQSR
Nanno
Thalassio ADGSYLAAETKVKTDQQGAMMVGTA
Albugo PTVDNAVR
Pythium TGERGGSSE
Aureoc GASGSAN
Ectocarpus REDGEAL
Aphanomyces SPGPPV
Guillardia TGNKSEGCEER
Emiliania RHYGRETQSKQDTSVAGGQASQCPPGLAFVSDVSPDPSSRVNEPPS-R
Aphanomyces PPSLIRAIPDDQAMHDVRGFGLATKSSPTSNPNLSLVDGVKASRTNDKEGPKPFLL
Phytoph
Saprolegnia TLEPIS
Hsa4galt YPEW-PVVF FMKGLT
Hsa4gnt NPNF-QVFV LFAGPT
Dma4gt1 HPGT-RVVV LMKGLA
Col.A1,4GT IPKIIHQT YKNETI
Mita IPKEK-IIK HKATGS
Lgt5 IPFEE-YFS DDRCAG
Cgtd IPNIIHQI WIQGYE
Sc Och1 IPQRV-WQT WKVGAD-
Tg GT1 LAQAIREAKLYCGHPERTVVDCRSSCCLEETGRHTPPELPLVDGGHSSAGCKNFSSAFMKIPSEFVSGKAIPPLLHFV WLGGH-
Hammondia LAKAINEAKLYCGHPQRTVVDCRCPCSPEETGGHTPPDLPVVDGGHSSTGCKNFSSPFTQIPSEFVCGKAIPPLLHFV WLGGH-
Neospora LAGAIREAQLHCRHPERTAEDSGSARSPETSGQLTLPVVPVTPAGPSVTGDEHLLSPVVQSPHAWDRGQAIPPLLHFV WLGGHR
Sarco LECAGRERNTQATVCALSGRPTAAEEHPE PRADSGAFNTTENFLLKSLRN-TSWH IPPLLHFV WCGKE-
Reticulo IPKRIHQI WLGDK-
Bn Bif IPKIIHQI WFGSN-
Oxytricha KSIPKKIHFI WLGEK-
Nanno LPKKKHFYAKLLLISKISTEA WKGRH-
Thalassio IPKILHFI WLGGN-
Albugo IPFIIHQI WLGPH-
Pythium IPHVIHQI WLGPH-


Aureoc
Ectocarpus
Aphanomyces
Guillardia
Emiliania
Aphanomyces
Phytoph
Saprolegni
Hsa4galt
Hsa4gnt
Dma4gt1
Col.A1,4GT
Mita
Lgt5
Cgtd
Sc Och1
Tg GT1
Hammondia
Neospora
Sarco
Reticulo
Bn Bif
Oxytricha
Nanno
Thalassio
Albugo
Pythium
Aureoc
Ectocarpus
Aphanomyces
Guillardia
Emiliania
Aphanomyces
Phytoph
Saprolegnia
Hsa4galt
Hsa4gnt
Dma4gt1
Col.A1,4GT
Mita
Lgt5
Cgtd
Sc Och1
Tg GT1
Hammondia
Neospora
Sarco
Reticulo
Bn Bif.
Oxytricha
Nanno
Thalassio
Albugo
Pythium
Aureoc
Ectocarpus
Aphanomyces
Guillardia
Emiliania
Aphanomyces
Phytoph
Saprolegnia
Hsa4galt
Hsa4gnt
Dma4gt1
Col.A1,4GT
Mita
Lgt5
Cgtd
Sc Och1


Tg GT1
Hammondia
Neospora
Sarco
Reticulo
Bn Bif.
Oxytricha
Nanno
Thalassio
Albugo
Pythium
Aureoc
Ectocarpus
Aphanomyces
Guillardia
Emiliania
Aphanomyces
Phytoph
Saprolegnia
Lottia
Hsa4galt
Hsa4gnt
Dma4gt1
Col.A1,4GT
Mita
Lgt5
Cgtd
Sc Och1
Tg GT1
Hammondia
Neospora
Sarco
Reticulo
Bn Bif.
Oxytricha
Nanno
Thalassio
Albugo
Pythium
Aureoc
Ectocarpus
Aphanomyces
Guillardia
Emiliania
Aphanomyces
Phytoph
Saprolegnia
Hsa4galt
Hsa4gnt
Dma4gt1
Col.A1,4GT
Mita
Lgt5
Cgtd
Sc Och1
Tg GT1
Hammondia
Neospora
Sarco
Reticulo
Bn Bif.
Oxytricha
Nanno
Thalassio
Albugo
Pythium
Aureoc
Ectocarpus
Aphanomyces
Guillardia
Emiliania
Aphanomyces
Phytoph
Saprolegnia



Sequence Ids:

Hsa4galt: Lactosylceramide 4-alpha-galactosyltransferase [*Homo sapiens*] (BAA95915.1)
Hsa4gnt: alpha-1,4-N-acetylglucosaminyltransferase [*Homo sapiens*] (AAD48406.1)
Dma4gt1: alpha4GT1 [*Drosophila melanogaster*] (AAF51162.1)
Col.A1,4GT : UDP-galactose:beta-D-galactoside alpha-1,4-galactosyltransferase [*Columba livia*] (ADC84388.1)
Mita: MIPC synthase subunit (SurA), [*Aspergillus fumigatus* Af293] (EAL85572.1)
Lgt5: alpha-1,4-galactosyltransferase [*Moraxella catarrhalis* 2951] (AAZ29048.1)
Cgtd: alpha-1,4-galactosyltransferase [*Campylobacter jejuni*] (AAM90647.1)

Supplemental Fig. S3.1. Alignment of Glt1-like sequences with other CAZy GT32 family proteins. Complete alignment of Glt1-like sequences (third-panel) with representative characterized CAZy GT32 family proteins from eukaryotes (top-panel) and bacteria (bottom-panel) or with related uncharacterized CAZy GT32 sequences from protists that have partial or full TgSkp1-like Skp1 modification pathway. All the sequences were manually aligned and color coded as described in **Fig 3.2** except that similar amino acids are highlighted when they have majority representation at a given position. Additional sequence ids used in this alignment only (not present in **Fig 3.2**) are given at the bottom.

```
Atgggcagcagccatcatcatcatcacagcagcggcagagaaaacttgtatttccagggccat
 M  G  S  S  H  H  H  H  H  H  S  S  G  R  E  N  L  Y  F  Q  G  H pET15b(TEV)

atgcgtgaaagcctgggtgactgtagcgtggattgtctgtcctcctgcgaaaaacatggt   60
 M  R  E  S  L  G  D  C  S  V  D  C  L  S  S  C  E  K  H  G    20

aacatgcgttccctggatgcccgtggtcgtctggcggaaaccacgtatcgtcaaaaccgc  120
 N  M  R  S  L  D  A  R  G  R  L  A  E  T  T  Y  R  Q  N  R    40

attgattttccggacttcccgtcccattcagtcatccagtttaacctgtggagctctgaa  180
 I  D  F  P  D  F  P  S  H  S  V  I  Q  F  N  L  W  S  S  E    60

gataatacctggctgagtctgctgcgtggcctgtggcaccgctgtaatgcacgtcgtgtg  240
 D  N  T  W  L  S  L  L  R  G  L  W  H  R  C  N  A  R  R  V    80

ctgctggcagcatccggtgatccgcgtgaaccggcacatctgtctgaagaagttaccagt  300
 L  L  A  A  S  G  D  P  R  E  P  A  H  L  S  E  E  V  T  S   100

aaacgtcaccgtctgccgggtgtcgcaacgccgcgtctgaactgtgattctgaactgctg  360
 K  R  H  R  L  P  G  V  A  T  P  R  L  N  C  D  S  E  L  L   120

aaatgctataaaaatggcggtgatctgaaaggcgtggaccatcactgtccgcaaccggca  420
 K  C  Y  K  N  G  G  D  L  K  G  V  D  H  H  C  P  Q  P  A   140

ccggctgatagccgtggtgatcgtgacggtgaagcattctgcgctgaccgtgcgctgtcg  480
 P  A  D  S  R  G  D  R  D  G  E  A  F  C  A  D  R  A  L  S   160

gcaagcctggcacaggcaattcgtgaagccaaactgtactgtggtcatccggaacgtacc  540
 A  S  L  A  Q  A  I  R  E  A  K  L  Y  C  G  H  P  E  R  T   180

gtggttgattgccgcagttcctgctgtctggaagaaaccggtcgtcatacgccgccggaa  600
 V  V  D  C  R  S  S  C  C  L  E  E  T  G  R  H  T  P  P  E   200

ctgccgctggttgatggtggtcactcatcggcaggttgtaaaaacttcagctctgctttc  660
 L  P  L  V  D  G  G  H  S  S  A  G  C  K  N  F  S  S  A  F   220

atgaaaattccgtctgaatttgtcagtggcaaagcgatcccgccgctgctgcatttcgtg  720
 M  K  I  P  S  E  F  V  S  G  K  A  I  P  P  L  L  H  F  V   240

tggctgggtggtcacccgccgccgtttttcgaaaccattcgccaaagctgggccgttcat  780
 W  L  G  G  H  P  P  P  F  F  E  T  I  R  Q  S  W  A  V  H   260

aatccggatctgatccaggcactgtggaccgacgctcacgtggaaagcctgctggatgtt  840
 N  P  D  L  I  Q  A  L  W  T  D  A  H  V  E  S  L  L  D  V   280

ctggaccgtaaatctcgcagtcgtcgcccgaaatgccgtaaaaccgatcatcaggacctg  900
 L  D  R  K  S  R  S  R  R  P  K  C  R  K  T  D  H  Q  D  L   300

cacccgctgctggtcgatcgtccggcaacggacgctggtgaaaccacggaatgggaaatg  960
 H  P  L  L  V  D  R  P  A  T  D  A  G  E  T  T  E  W  E  M   320

tccgattcagtcccggacagccagaccctggtgcatgcgattaaagcctttcgtaaagaa 1020
 S  D  S  V  P  D  S  Q  T  L  V  H  A  I  K  A  F  R  K  E   340
tcctgtccgggcgcgaaatcagatattgcccgcctgctgatcctgtgccactatggcggt 1080
 S  C  P  G  A  K  S  D  I  A  R  L  L  I  L  C  H  Y  G  G   360
```

```
atttacgcggatgtggacatggaagccatcaaaccgctgccgccgtgtctgcgtcaatgc 1140
 I  Y  A  D  V  D  M  E  A  I  K  P  L  P  P  C  L  R  Q  C   380

accacggtgtttatgggcatgcagcgcccggatgcagttgaactgggtaacgctctgatc 1200
 T  T  V  F  M  G  M  Q  R  P  D  A  V  E  L  G  N  A  L  I   400

ggctgcagttccggtcatgaactgattcgtttcatcctgcagcgtgttggccgcccgtat 1260
 G  C  S  S  G  H  E  L  I  R  F  I  L  Q  R  V  G  R  P  Y   420

tcgcaatggggtacccgtagcgcagatcagatggcagtcgtgctggacattctgaaactg 1320
 S  Q  W  G  T  R  S  A  D  Q  M  A  V  V  L  D  I  L  K  L   440

cacgtggcgcgtttctccccggaagcctcagaactgatggattcgctgcgtagcaccaac 1380
 H  V  A  R  F  S  P  E  A  S  E  L  M  D  S  L  R  S  T  N   460

gacgaagcgacgaatgttatcgaacgtaccggcccgggtctgctgacccgtgcaacgctg 1440
 D  E  A  T  N  V  I  E  R  T  G  P  G  L  L  T  R  A  T  L   480

gcttggctgcgcgatcaactgaactcatcgagctgtgcacgttgccgtccgcaggaacgt 1500
 A  W  L  R  D  Q  L  N  S  S  S  C  A  R  C  R  P  Q  E  R   500

cgttcccatgattacgaacacaaatcaaataccacggaagcaacctcggtgcatgctgac 1560
 R  S  H  D  Y  E  H  K  S  N  T  T  E  A  T  S  V  H  A  D   520

gttgtcggctcggaaacgggtagcaaagaagatgactctaaaccggaagcaaccagtggc 1620
 V  V  G  S  E  T  G  S  K  E  D  D  S  K  P  E  A  T  S  G   540

acgtgctgttgcattctggcgaccacgtgtatttgcccgccgatctttttctatccggtt 1680
 T  C  C  C  I  L  A  T  T  C  I  C  P  P  I  F  F  Y  P  V   560

ccgaatcatcgtcgcaaagaactgcgcgaaggcaaagtccagaccgaacacctggaatct 1740
 P  N  H  R  R  K  E  L  R  E  G  K  V  Q  T  E  H  L  E  S   580

agtttttcttacacggtccatcactggcgtcagacctggcaagacagcgtccgtcaagaa 1800
 S  F  S  Y  T  V  H  H  W  R  Q  T  W  Q  D  S  V  R  Q  E   600

agcgaaaatgaatgttaa 1818
 S  E  N  E  C  -   605
```

Supplemental Fig. S3.2. Sequence of TgGlt1 cDNA. Codon optimized synthetic nt sequences of TgGlt1 cDNA is shown in black. N-terminal His6-tag contributed by pET15b(TEV), is shown in purple. The translated amino acid sequence is shown below.

CHAPTER 4

MAMMALIAN GLYCOGENIN-LIKE GLYCOSYLTRANSFERASE ADDS THE FIFTH SUGAR ON *TOXOPLASMA* SKP1 AND IS IMPORTANT FOR PARASITE GROWTH[1]

[1]Kazi Rahman, Msano Mandalasi, Peng Zhao, Hanke van der Wel, Lance Wells, Christopher M West. To be submitted to the *Journal of Biological Chemistry*.

Abstract

Skp1 is an adaptor subunit of the SCF (Skp1/Cullin-1/F-box protein) class of E3 ubiquitin ligases that are important for cell cycle and developmental regulation. Unlike its human counterpart, Skp1 from *Toxoplasma gondii* (Tg) is hydroxylated at Pro154 by an O_2-dependent prolyl-4-hydroxylase, and the resulting hydroxyproline is sequentially modified by a chain of 5 sugars. We previously reported the enzymes that assemble the core tetrasaccharide, but the mechanism of addition of the terminal sugar on TgSkp1 remained unknown. Using bioinformatics approaches, we predicted a cytoplasmic glycosyltransferase from CAZy GT family 8, named Gat1, which is highly conserved across a broad range of protozoa that harbor homologs of earlier genes of the Skp1 glycosylation pathway. Disruption of *gat1* in type I RHΔΔ parasites interrupted addition of the 5th sugar on TgSkp1 based on mass spectrometry, and reduced parasite growth based on plaque assay. The missing fifth sugar was assessed to be an α-linked Gal based on sensitivity to green coffee bean a-galactosidase. Gat1 has very high sequence similarity to the catalytic domain of glycogenin, an αGlc-transferase that primes glycogen synthesis in animals and yeasts. A phylogenetic analysis of glycogenin, Gat1 and selected other CAZy GT8 sequences reveal that Gat1s and glycogenins form separate clades with a common ancestor, and their exclusive phylogenetic distributions suggest that glycogenins derived from an ancestral Gat1 concomitant with its replacement by AgtA for Skp1 glycosylation in amoebazoa. We speculate that these dramatic changes in biochemical gene function nevertheless conserved a cellular function in nutrition regulation, consistent with a role for Skp1 glycosylation in growth regulation.

Introduction

Skp1 is an adaptor subunit of SCF (Skp1/Cullin-1/F-box protein)-class E3 ubiquitin ligases that target proteins for polyubiquitination and degradation via 26S proteasome (159). In the agent for human toxoplasmosis, *Toxoplasma gondii*, Skp1 is hydroxylated by the cytoplasmic prolyl 4-hydroxylase (*phyA*) at Pro154, and subsequently *modified* by a linear pentasaccharide that regulates parasite growth in fibroblasts (chapter 2) (16). In the social amoeba *Dictyostelium*, Skp1 is also modified by a pentasaccharide, which regulates SCF assembly and O_2-dependent development (10).

Both in *Toxoplasma* and *Dictyostelium*, the first 3 sugars are added sequentially to Hyp of Skp1 by Gnt1, a polypeptide αGlcNAcT, and PgtA, a processive bifunctional enzyme with β1,3-GalT and α1,2-FucT activities (10, 16). These sugar nucleotide-dependent enzymes are soluble in the cytoplasm and lack a rough endoplasmic reticulum targeting sequence or membrane anchor motifs. The Fuc terminus of the core trisaccharide of *Dictyostelium* Skp1 is further modified by two α,1,3-linked Gal residues by another cytosolic glycosyltransferase AgtA. Interestingly, *Toxoplasma* lacks a homologue of AgtA suggesting that addition of the terminal disaccharide is catalyzed by other enzymes. Previous studies have shown that the fourth sugar is located on the Fuc terminus of the *Toxoplasma* Skp1 trisaccharide, and its addition is mediated by a cytoplasmic glycosyltransferase from CAZy GT family 32, named Glt1, But the enzyme that is responsible for the addition of the fifth sugar is unknown. Bioinformatics analysis suggest that the presence of AgtA is strictly restricted to *Dictyostelium* and other amoebazoans. *Toxoplasma* and a wide group of protists lack *agtA* but harbor *glt1* and earlier genes of the Skp1 modification pathway (i.e. *phyA*, *gnt1* and *pgtA*).

A search of predicted proteins encoded by the *Toxoplasma* genome identified several putative glycosyltransferases predicted to reside in the cytoplasm or nucleus, which are therefore candidates for completing the Skp1 pentasaccharide in *Toxoplasma*. One of these, named Gat1, is annotated in some databases as glycogenin, an enzyme involved in glycogen biosynthesis. Glycogen is an important nutrient storage polymer consisting of α-1,4-linked glucose subunits with α-1,6-linked glucose at the branching points. It is present in many organisms, from bacteria to humans. *De novo* synthesis of glycogen in yeasts and mammals requires a priming mechanism by an autocatalytic self-glucosylating protein named glycogenin (170). A glycogenin dimer autoglucosylates a conserved tyrosine residue (Tyr230 in yeast and Tyr195 in human) leading to an α1,4-linked chain of 8–12 glucose units. Glucosylated glycogenin serves as the substrate for glycogen synthase that in combination with glycogen branching enzyme forms the complete glycogen particle comprised of up to ~55,000 glucose residues (171). Glycogenin is a member of the CAZy GT8 family of glycosyltransferases containing an N-terminal catalytic domain and C-terminal extension of variable length. The last 35 amino acids of the C-terminus tail in yeast and human contain a conserved motif that is sufficient for binding to glycogen synthase (172).

Toxoplasma and a number of other apicomplexans including *Eimeria* (173) and *Sarcocystis* (174) lack glycogen but accumulate a related polysaccharide named amylopectin in the cytoplasm of the bradyzoite and the sporozoites (175). Toxoplasma tachyzoites can also make amylopectin in stress conditions such as acidic pH (176). Amylopectin is a major component of starch, a storage material in the plastid of photosynthetic eukaryotes or in the cytoplasm of red algae (177), dinoflagellates (178) and other unicellular protists that acquired a plastid-like organelle by secondary endosymbiosis. Like glycogen, amylopectin also consists of α-1,4 linked glucose chains that are branched through α-1,6 linkages, but the branches are less frequent (176). The *Toxoplasma*

genome encodes a glycosyltransferase 8 family protein, which we named as Gat1 shows 34.21%/46.31% identity/similarity over 190 amino acid long catalytic domain of human glycogenin (**Table 4.1**), and was hypothesized to be involved in the priming of *Toxoplasma* amylopectin synthesis (175). However, the absence of the glycogen synthase binding motif in the *Toxoplasma* Gat1, and the absence of a glycogenin-like protein in many other amylopectin harboring protists, raises the question of whether *Toxoplasma* Gat1 is involved in the priming of *Toxoplasma* amylopectin synthesis, or has any other function.

Here, using genomic studies, and reverse genetic and mass spectrometry approaches, we have shown that *Toxoplasma* Gat1 is involved in the transfer of a galactose to the 5th position of the Skp1 glycan. Phylogenetically Gat1 is related to glycogenin and is required for optimum parasite proliferation.

Experimental Procedures

Parasites, Cell Culture and Plaque Assays– Toxoplasma strain RHΔ*ku80*Δ*hxgprt* (RHΔΔ) was cultured in human foreskin fibroblasts (HFFs) using Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% (v/v) fetal bovine serum, 2 mM L-glutamine and 100 units/ml penicillin/streptomycin at 37°C in a humidified CO_2 (5%) incubator. RHΔ*ku80*Δ*gat1* (RHΔ*gat1*Δ) strains were cultured in the same medium supplemented with 25 μg/ml mycophenolic acid (Sigma) and 25 μg/ml xanthine (Sigma). Strains were cloned by limiting dilution in 96-well plates.

Table 4.1. Percent identity and similarity of mammalian glycogenin and *Toxoplasma* Gat1.

Organisms	Region compared	Total number of amino acids	Identity	Similarity
Homo sapiens	4-251	190	34.2%	46.3%
Toxoplasma gondii	5-331	190		

Tggat1 Disruption Strain– DNAs for gene disruptions were generated from the vector pminiGFP.ht, in which the *hxgprt* gene is flanked by multiple cloning sites as described (160). Briefly, the 5'-flank and 3'-flank targeting sequences of *Tggat1* from RHΔΔ were PCR amplified with primer pairs a and a' and pairs b and b', respectively (**Table 4.2**). The 5'-fragment was liberated with ApaI and XhoI and inserted into similarly digested pminiGFP.ht. The resulting plasmid was digested with XbaI and NotI and ligated to the XbaI and NotI digested 3'-flank. The resulting vector was linearized with PacI and electroporated into RHΔΔ strain, selected under 25 µg/ml MPA and 25 µg/ml xanthine, and cloned by limiting dilution. Genomic DNA from several clones were screened by PCR to identify the *Tggat1* disruption strain.

Immunoprecipitation and Mass Spectrometry of TgSkp1 peptides– TgSkp1 was immunoprecipitated and Orbitrap LC-MS/MS analysis was as described before in Chapter 2 (160).

Exoglycosidase Treatment of TgSkp1 peptides– For exoglycosidase treatment, TgSkp1 was immunoprecipitated from 6×10^8 RHΔΔ parasites and alkylated, reduced and trypsinized as described before in chapter 2. Trypsinized TgSkp1 glycopeptides were then purified by adsorption to a C18 Toptip column (PolyLC InC. Item # TT200MC18) pre-equilibrated with 0.05% trifluroacetic acid (v/v) and eluted with 60% acetonitrile in 0.05% trifluroacetic acid. The sample was vacuum centrifuged, dissolved in 10 µl water, boiled for 10 mins to heat inactivate the trypsin, vacuum centrifuged again and re-dissolved in the exoglycosidase preparation to a final volume of 10 µl 25mM NH_4AC (pH 5.5) containing 20 ng of trypsin inhibitor (aprotinin) and 10 milliunits of green coffee bean alpha galactosidase (Boehringer Mannheim) (179). The sample was

Table 4.2. Oligonucleotide sequences employed

Targeting sequence amplification:

T*gat1*-disruption

a) Gat1F1 5'-flank 5'-end 5'- GGGGGCCCAACCAGCGGATCTTCTGAAC (ApaI)
a') Gat1R1 5'-flank 3'-end 5'- GGCTCGAGACGCGTTGAGCGATTGA (Xho1)
b) Gat1F2 3'-flank 5'-end 5'- GCTCTAGAGAGGGAGAACCAAAGTGATGAT (XbaI)
b') Gat1R2 3'-flank 3'-end 5'- CGCGGCCGCTGCGTAGAACACAAGGAGAAC (NotI)

PCR for confirmation:

PCR1

Forward: TACCCTGTTGACGGACAATT
Reverse: CTTTGCTGGTTGTTCCCAAG

PCR2

Forward: GAACCGAATGACAACGCATTAC
Reverse: AGTCGCGGAACATCTCGTTGAAGT

PCR3

Forward: ATTTGCATCCTGAAAGGCTCTCGC
Reverse: TCTGAAATGGAGTCGCCTTG

incubated at 37ºC for 3 h and subsequently analyzed by Orbitrap mass spectrometry as described (chapter 2) (160).

Phylogenetic Analysis– For phylogenetic inferences of Gat1, 43 homologous protein sequences from different organisms were at first manually aligned, uploaded in BioEdit sequence alignment editor V 7.2.5 and again manually refined to remove most of the gaps and ambiguous sites, which resulted in a sequence alignment based on 196 amino acids of *Toxoplasma* Gat1**.** For phylogenetic inferences, MEGA6 software (164) was used to estimate the best substitution model as well as estimate for the Maximum Likelihood (ML) trees. The phylogenetic tree was inferred based on a LG amino acid substitution model with a gamma distribution (165). To determine the reliability of each node of the tree, a total of 10000 repetitions were performed using the bootstrapping method (166).

Plaque assays, Western Blotting– Procedures were performed as described (160).

Results

Prediction of Gat1 as a Candidate Skp1 Modifying Glycosyltransferase– Previously we showed that assembly of the core trisaccharide of TgSkp1 is mediated by the action of *Dictyostelium* homologs of Gnt1 and PgtA, and that a novel enzyme Glt1 is responsible for addition of the fourth sugar. Glt1-related sequences are present only in select protists ($E<10^{-5}$ in NCBI non-redundant database) that harbor sequences related to *phyA*, *gnt1*, and *pgtA*, but lack *Dictyostelium agtA*-like sequences.

The protists might like *Toxoplasma*, assemble a pentasaccharide on the their Skp1 proteins. In order to identify terminal Skp1 modifying glycosyltransferase in *Toxoplasma* and related protozoans, we adopted an evolutionary and comparative genomics approach. We hypothesized that the terminal Skp1-modifying GT candidate should have the following characteristics: Firstly, it should be cytoplasmic as the TgSkp1 modification is predicted to occur in the cytoplasm. Secondly, it should be conserved only in protists that have *glt1*-like sequences and earlier genes of the pathway but lack *agtA*. Therefore, from a group of previously predicted cytoplasmic glycosyltransferases (described in chapter 3) we searched for genes that satisfied these criteria. A BLASTP search was performed individually for all of these genes against NCBI non-redundant database. This yielded a putative glycosyltransferase from the CAZy GT8 family protein, which we named Gat1.

Gat1-disrupted Parasites Lack Terminal Sugar on TgSkp1– To test the contribution of *gat1* to *Toxoplasma* Skp1-modification, we disrupted the gene in the RHΔΔ strain by a double-crossover homologous recombination strategy, as described in "Experimental Procedures" and depicted in **Fig. 4.1A**. As described in **Fig. 4.1B**, the chosen RHΔ*gat1*Δ clone shows a negative PCR reaction for *gat1* coding region (PCR1), and a positive PCR reaction for insertion of the selection marker *hxgprt* (PCR2 and PCR3), using PCR primer pairs matching *hxgprt* and sequences in external *gat1* flanking regions suggesting the disruption of the gene.

To determine if Gat1 disruption affects Skp1 modification, TgSkp1 was immunoprecipitated from parental (RHΔΔ) and RHΔ*gat1*Δ knock-out tachyzoite extracts and its tryptic peptides were analyzed by conventional nanoLC/MS in an LTQ-XL Orbitrap mass spectrometer for the presence of Skp1-glycopeptides.

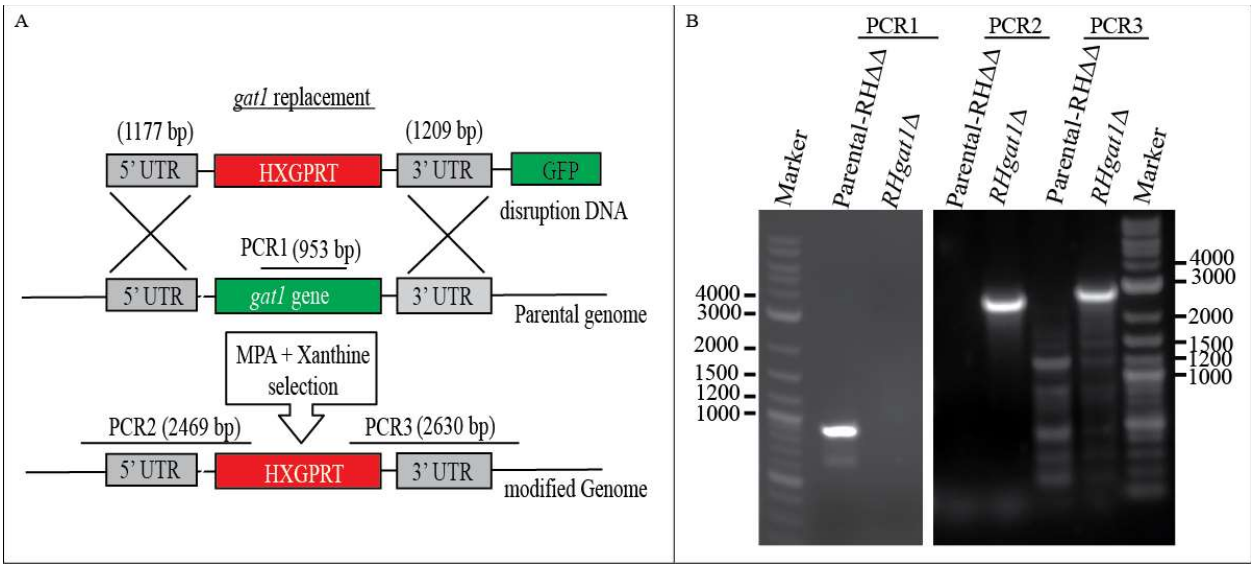


Figure 4.1. Knockout of *Tggat1*. Strategy for deletion and validation of the knockout status of endogenous Tg*gat1* locus in RHΔΔ. **(A)** Strategy for deletion of Tg*gat1*. To knock out Tg*gat1*, a *hxgprt* cassette, which is flanked by the 5'- and 3'- untranslated regions of the Tg*gat1* locus, was transfected into RHΔΔ strain. GFP negative parasites that are expected to gain Tg*gat1* locus replacement by the *hxgprt* cassette upon a double cross crossover event, were selected under MPA/Xanthine and serially diluted. A GFP-negative clone was examined by three sets of PCR reactions. Approximate locations and expected lengths of PCR products to validate the genotypes are depicted by lines. **(B)** PCR validation of the Tg*gat1* knock-out. The parental RHΔΔ was positive for PCR1 while the Tg*gat1*Δ clone was negative. On the other hand, Tg*gat1*Δ clone, but not the RHΔΔ parent, yielded PCR2 and PCR3 products of the expected length, indicating site specific integration of *hxgprt*.

As in our previous studies (160) of the RHΔΔ strain, we detected an unmodified form of Skp1 peptide, ^{145}IFNIVNDFT**P**EEEAQVR^{161} that includes Pro-154, as well as its hydroxy-pentasaccharide bearing form, but no other intermediate glycoforms of the peptide (**Fig. 4.2A**). In contrast, from RHΔ*gat1*Δ parasites we the detected hydroxy-tetrasaccharide (Hex_2dHexHexNAc) form of this peptide, ($[M+2H]^{2+}$, 1355.61; $[M+3H]^{3+}$, 904.08), in addition to its completely unmodified form (**Fig. 4.2B**). No pentasaccharide form of the peptide was observed. The putative tetrasaccharide-bearing doubly charged glycopeptide ion (($[M+2H]^{2+}$, 1355.61) was subjected to MS/MS analysis to confirm its composition and characterize its organization. CID fragmentation of the doubly charged ion yielded a series of ions resulting from sequential loss of a hex, dHex, Hex and HexNAc, representing the loss of core Glc, Fuc, Gal and GlcNAc from Skp1 as expected and ultimately resulted in the hydroxypeptide ion (**Fig. 4.3A**). A series of b- and y-ions also confirmed the identity of the glycopeptide ion (**Fig. 4.3B**). These data indicate that *gat1* knock-out parasites lack the fifth sugar on Skp1 and, therefore, Gat1 might directly mediate transfer of the last sugar of the TgSkp1 pentasaccharide.

Native Toxoplasma Skp1 Contains a Terminal α-Gal Residue– In *Dictyostelium* the terminal disaccharide on Skp1 is Galα1,3Galα1-. In *Toxoplasma* the fourth sugar is a glucose but the identity of the fifth sugar is unknown. To test if the terminal sugar on TgSkp1 is also a galactose, we treated trypsinized wild-type Skp1 with green coffee bean alpha-galactosidase and analyzed the peptides in an LTQ-XL Orbitrap mass spectrometer. The analysis showed that while the untreated sample contained unmodified and pentasaccharide modified TgSkp1 glycopeptide

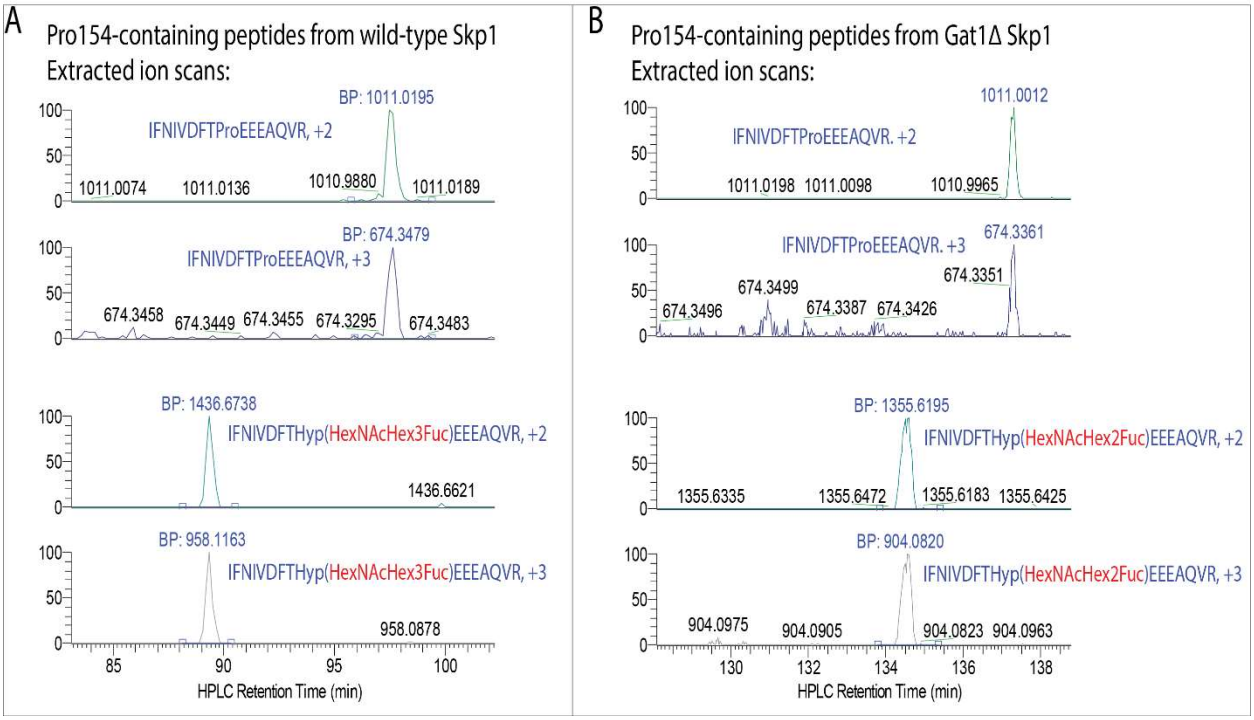


Figure 4.2. Gat1 is required to add the 5th sugar on *Toxoplasma* Skp1. TgSkp1 purified from extracts of RHΔΔ and RHΔ*gat1*Δ tachyzoites by immunoprecipitation was trypsinized and analyzed by reverse phase HPLC on an LTQ-XL Orbitrap MS. (**A, B**) The extracted ion scans showed *m*/*z*-values consistent with unmodified Skp1 peptide with double and triple charged ions in both samples. In addition, (**A**) the wildtype sample showed doubly and triply charged ions eluting at an earlier position for the fully modified Skp1 glycopeptide with a HexNAcHex3Fuc pentasaccharide but no mono-, di-, tri- or tetrasaccharide bearing glycoforms were observed. (**B**) Analysis of Gat1Δ purified extracts showed doubly and triply charged ions corresponding to a TgSkp1 glycopeptide containing a tetrasaccharide with the composition HexNAcHex2Fuc but no other mono-, di-, tri- or pentasaccharide bearing glycoforms. Linda Peng Zhao and Lance Wells are acknowledged for performing this experiment.

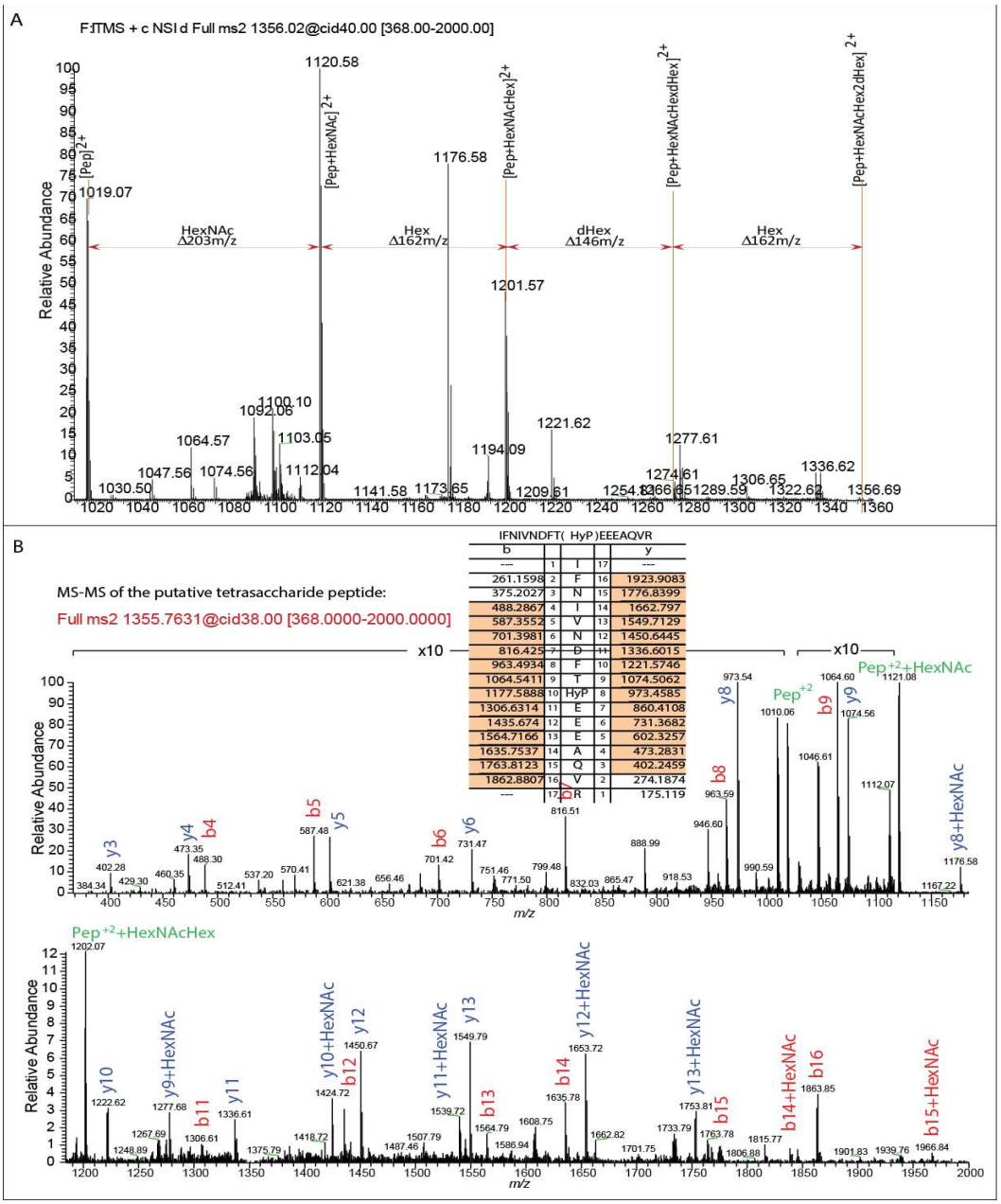


Figure. 4.3. MS-MS of doubly charged TgSkp1 glycopeptide from RHΔ*gat1*Δ strains. (**A**) CID fragmentation of the doubly-charged precursor ion yields a sequential loss of monosaccharide residues corresponding to Hex, dHex, Hex and HexNAc, indicating the presence of a linear tetrasacchararide. (**B)** Inspection of the full CID fragmentation spectrum shows b- (blue annotations) and y- (red annotations) ion series that match the predicted peptide sequence, as illustrated in the inset, and demonstrate that the glycan is linked via a hydroxylated derivative of Pro154. Peptides with residual sugars are annotated in green. Linda Peng Zhao and Lance Wells are acknowledged for performing this experiment.

(data not shown), the α-galactosidase treated sample contained a glycopeptide containing a HexNAcHex2dHex tetrasaccharide **(Fig. 4.4C-D)**. The original pentasaccharide-peptide **(Fig. 4.4E-F)** and unmodified peptide **(Fig. 4.4A-B)** were also present. The appearance of the tetrasaccharide after α-galactosidase treatment suggests the presence of a terminal α-linked Gal on the *Toxoplasma* Skp1 glycan. The presence of pentasaccharide may be the result of incomplete digestion, but the possibility of heterogeneity is not excluded.

Gat1 is Phylogenetically Related to Glycogenin– The 345-amino acid long Gat1 protein is encoded by a single exon gene model in Type I GT1 strain (TGGT1_310400), which is 99% identical to that of the type II ME49 and Type III VEG strain sequences. The gene is expressed equally both in tachyzoites and bradyzoites based on transcript analysis (175). To examine the evolution of Gat1, a BLASTP **(V 2.4.0)** search using the full-length protein sequence as the query was performed against the NCBI non-redundant database (December 2016). The top-scoring sequences, with Expect values of $<10^{-32}$, were found in other protists that contain *Toxoplasma* PgtA-like sequences and lack *Dictyostelium* AgtA-like sequences. All sequences with larger Expect values came from organisms that lack PgtA-like sequences. The most similar sequence of known function is a glycogenin, with an Expect value of E^{-26}. Organisms that possessed Gat1-like sequences lacked related sequences with an Expect value less than $<10^{-16}$. To evaluate the evolutionary relationship of Gat1 with glycogenin and other glycosyltransferases, a phylogenetic analysis of a manually-aligned sequences from the following representatives was performed (**Table 4.3** and **Fig. 4.5**): **(A)** Gat1-like sequences from organisms that contain PgtA-like but lack AgtA-like sequences, **(B)** glycogenin-like sequences from simplest to complex animals **(C)** closest Gat1-like sequences from plants, **(D)** closest Gat1-like GT8 sequences from organisms

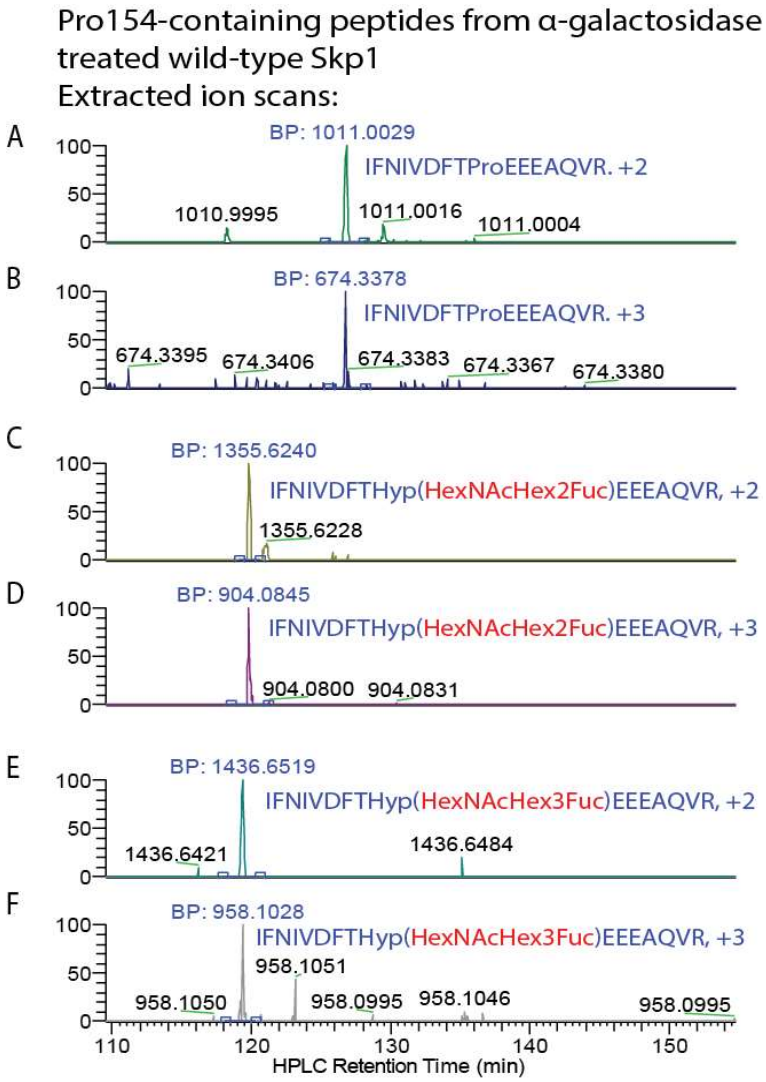
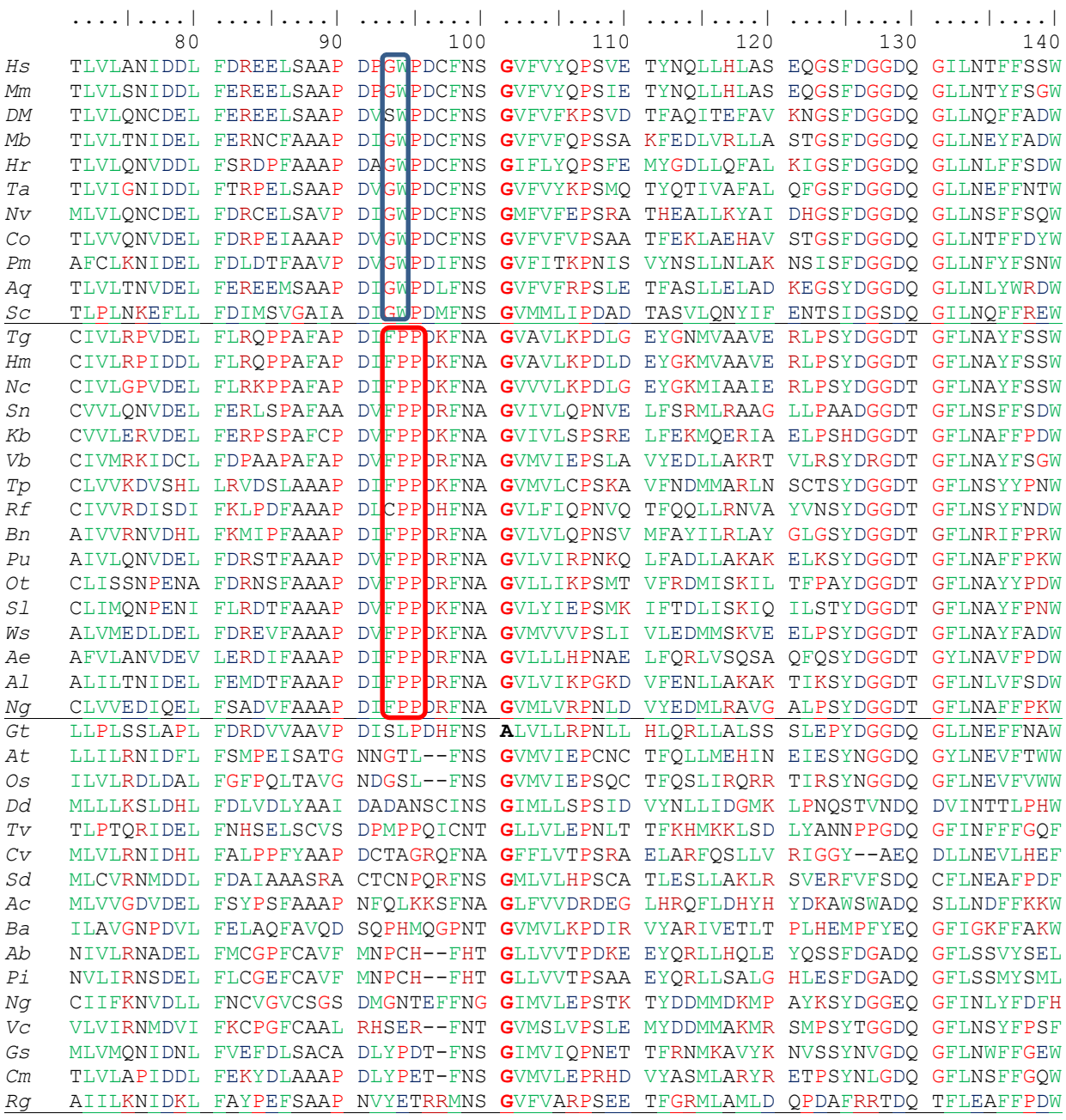


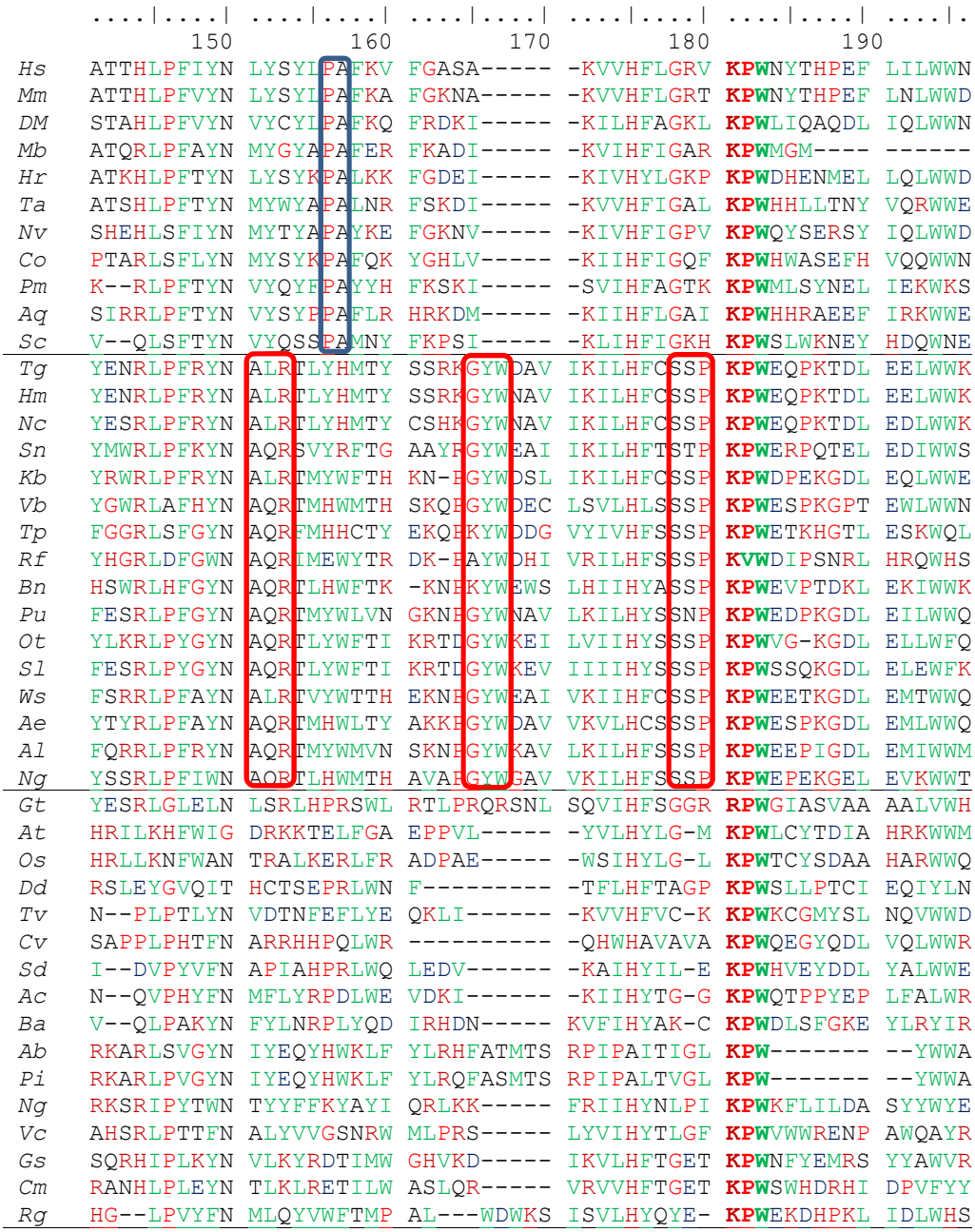
Figure 4.4. Native *Toxoplasma* Skp1 contains a terminal α-Gal residue. TgSkp1 immunoprecipitated from parasite extracts was trypsinized, purified on C18, treated with green coffee bean α-galactosidase, and analyzed by RP HPLC in an LTQ-XL Orbitrap MS. (**A-F**) The analysis shows doubly and triply charged ions eluting at 118.5 mins corresponding to the glycopeptide containing a HexNAcHex2Fuc tetrasaccharide, which was confirmed by MS-MS (data not shown). The original pentasaccharide-peptide was also present. Linda Peng Zhao and Lance Wells are acknowledged for performing this experiment.

Table 4.3. List of sequences/organisms utilized in the phylogenetic tree generation. The best scoring hits in each organism based on BLAST analysis are reported. Known function of any gene is reported at the right inside bracket.

A. Gat1-like sequences from PgtA containing Protists	**B. Glycogenin-like sequences from**
Toxoplasma gondii EPR60889.1	*Trichoplax adeherans (*E^{-27}*)* XP_002116183.1 *Amphimedon queenslandica (*E^{-30}*)* XP_003383748.1 *Nematostella vectensis (E-24)* XP_001625718.1 ***(Simplest animals)***
Hammondia hammondi XP_008886569.1	
Neospora caninum Liverpool XP_003885051.1	
Ectocarpus siliculosus CBJ26265.1	*Saccharomyces cerevisiae* (E^{-13}) **(Yeast)** E7QGE5 (known function: primes glycogen synthesis)
Albugo laibachii CCA19642.1	*Monosiga brevicollis* (E^{-27}) **(Choanoflagellate)** XP_001744585.1
Vitrella brassicaformis CEM34465.1	*Capsaspora owczarzaki* **(Filesterea)** XP_004349815.2
Nannochloropsis gaditana EWM28655.1	*Helobdela robusta* (E^{-27})**((Anneleid)** XP_009013909.1
Oxytricha trifallaxs EJY67427.1	*Drosophila melanogaster* (E^{-26}) **(Fruit fly)** NP_001163232.2 (known function: primes glycogen synthesis)
Stylonychia lemnae CDW86810.1	*Mus Musculus* (E^{-23}) (**Animal**) NP_038783.1
Thalassiosira pseudonana XP_002291959.1	*Homo sapiens* (E^{-23}) (**Animal**) AAH31096.2 (known function: primes glycogen synthesis)
Stylonychia lemnae CDW86810.1	**D. Gat1-like sequences (E value <10^{-5}) from the protists that have Gnt1 but not the PgtA**
Reticulomyxa filosa X6P0J2	*Acanthamoeba castellanii (*E^{-10}*)* XP_004352787.1
Bigowiella natans JGI: aug1.92_g19606	*Cyanidioschyzon merolae ((*E^{-22}*) (**Red Alga**)* XP_005535960.1
Sarcocystis neurona SN3_01500095	*Galdieria sulphuraria (*E^{-21}*) (**Red Alga**)* XP_005708321.1
Karenia brevis EX959504.1	*Volvox carteri (*E^{-15}*) (**Green algae**)* XP_002954821.1
Pythium ultimum K3WC47	*Phytophthora infestans (*E^{-13}*)* XP_002997946.1
Aphanomyces euteiches (Aphanodb2: Ae201684_9096.1)	*Naegleria gruberi (*E^{-13}*)* XP_002672734.1
C. Closest Gat1-like sequences from plants	*Saprolegnia diclina (*E^{-7}*)* XP_008603979.1
Arabidopsis thaliana (E^{-16}*)* NP_175891.1	*Chlorella variabilis (*E^{-10}*)* XP_005850943.1
Oryza sativa (E^{-17}*)* A2XDA4	*Trichomonas vaginalis (*E^{-10}*)* XP_001309036.1
E. The closest Gat1- like GT8 sequences from organisms that have pgtA	**F. Gat1 like sequence from Bacteria**
Dictyostelium discoideum (E^{-6}*)* Q54L24	*Rhizobium meliloti (*E^{-19}*)* WP_029616784.1
Albugo laibachii (E^{-14}*)* F0W520	
Bigowiella natans (E^{-5}*)*	
Guillardia theta CCMP2712 (E^{-15}*)* L1J9Y4	

```
     ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
             10         20         30         40         50         60         70
Hs   QAFVTLTTND AYAKGALVLG SSLKQHRTTR RLVVLATLTL MKRPELGVTL TKLHCWSLTQ YSKCVFMDAD
Mm   QAFVTLTTND AYAKGALVLG SSLKQHRTTR RMVVLTSLTL MKRPELGITL TKLHCWSLTQ YSKCVFMDAD
DM   FAWVTLTTND TYSLGALVLA HSLKRAKTAH QLAVLVTLAL LSRPELGVTF TKLHCWRLVQ FEKCVFLDAD
Mb   QAYVTLCTND AYVVGAMLLA HSLRRTGTRR QIVCMITLGL LQRPELGVTL TKLHAWKLTH YDNCVFLDAD
Hr   -AYVTMATND VYAVGALVLA ETLRQTNTQQ DLVIMITLSL LQRSELGVTF TKIQAWRLVE YRKCVFMDAD
Ta   EAFVTLATND SYAVGAFVLG NSLRNVKTTR ELVVLITLRL LGRPDLGITL TKLHCWRLTE FSKAVFLDAD
Nv   EAFVSLVTND NYANGALVLG YSLRRVNTTR KLALLVTLAL LSRPELGITF TKIRCWNLTH YQKCVFMDAD
Co   EAFVTLVTND GYALGALVLA KSLRDVNTTR KIAVLITLAL LGRPELGVTL TKIYAWKLTQ FTKCVFLDAD
Pm   ETYMTLVLTD SYLIGSQVLA WSLRDSGSKK HLTALVTLYL LGRPDLRSSF TKIHIWAQEK FKKIIYLDAD
Aq   EAYVSLATNN DYCHGAIALA CSLRLTNTSR KLCLLISLAL IKRPELGVTF SKLHIWRLVH YSKCVFLDAD
Sc   LAIATLLYSA DYLPGVFALG HQVNKLKGDI ETCLIVTLAL LERPELSFAL IKARLWELTQ FEQVLYLDSD
Tg   YAYATLLTDN SFYYGVEALL KSLEATKTPY PVLLLHTVGS VAYPKAEDCF TKLRVWEQVD FDVIVYVDAD
Hm   YAYATLLTDN SFYYGVEALL KSLEATKTPY PVLLLHTVGS VAYPKAEDCF TKLRVWEQVD FDVIVYVDAD
Nc   YAYATLLTDN SFYYGVEALL KSLEATKTPY PVLLLYTVGS IAYPEKENCF TKLRAWEQVD FDVIVYIDAD
Sn   KAYATLLLDD SFFYGVAALI RSLAKTRTRY PLLLLHTVEE VRGPAKARLY TKLRLWEQED FDLLVYIDAD
Kb   EAYVSLLTSD SFLMAVQALI ASLKATGTAR RLLLLHTVAA IPNPHQTSGF TKLRVWEQVD FDKLVYIDAD
Vb   CAYITLLTSD SFAIGVETLA FSLRKTGTPH PFIVLVGVDG IANPHAESGF TKLHVWSLTE FQRVVYIDAD
Tp   KAIATFLSSA DFLPGCQTLL HSLKKQLPQT PIIVLLSDNN NSDNNDKCGW AKLRLFELDG YDTILYIDAD
Rf   YAVVSLVTSE SYVVGAQVLI HSLHRNGGLK GSNVLVTVSE IPNPLEKSGY TKLRIFEMVQ LKKLFYIDAD
Bn   YGYVSLLTND GFLPGAIVLA KSLLKVEARY PGAVMVTIPI EPLPCPNVGL TKLRVWQLGD FAKVVYIDAD
Pu   AAYATLITSD AYVMGVEALV YSLFKARVAF PLVVLHSVPD IGIPDEVSGY TKLHIFAMDD FEQIVYIDAD
Ot   -----MITDD GYLPGLQVLH YTLRKF-TSR LLVIILAVKP ILNPHEKSGY TKLYIWTLIQ FQKVFYIDAD
Sl   -----MITED SYLPGLQVMH YSLRKF-TQR TLVVIMTVKP IGNPNEKSGY TKFYIWSLTQ YKRIFYIDAD
Ws   -----MVTSD DFVIGAEVML HSLREHSTRR PLVVMVTVEP IAMPMKRVGY TKLRVWGLIQ FRCVVYIDAD
Ae   KTFATLVTSD DFVIGVQVLA YSLRKHGAKY PLIVLYTVEA LPNPNVHSGY TKLHVFNLVE FSTVFYIDSD
Al   QAYATMITSD DFQMGVEALL YSWSCTHSSI NFLILYTVDS IPIPASSSAY TKLNIFGLEE YQKIVYIDAD
Ng   HAFVTLLTGP GAQVLLHSLR TSISAKVAIR PVVVLVTVEP IANPYAESGF TKLQIWGLTQ FERVVYLDAD
Gt   EAYATLITTK EYIQGAIVLS RIVKSTDEER PFIALVLVPR VKRPTGATTY SKLFVWNLTA YRLVLYLDAD
At   EAYATILHAH VYVCGAIAAA QSIRQSGSTR DLVILVDNPK AEKDAYNWNY SKFRLWQLTD YDKIIFIDAD
Os   EAYATVLHSD TYLCGAIVLA QSIRRAGSTR DLVLLHDNPR AERGTYNYNY SKFRLWQLTD YDRVVFVDAD
Dd   NVYVTFADNA EYLKGIVALR MSMINTKCNY GLIVFVTIEM VDIPKEVPAF TKFRAWQLVE YERVIWLDSD
Tv   YAFATVT-TP AFCMGAVVLG YTLRKYGNDY SYLCLVTVND A-KPYLWRSW IKLELWTFTE YEKIVYLDTD
Cv   MARRGSTWPD SYLMGVQALA RSLLAAQAQH PLLVMYTVER YV-PAGHECW NKLRIWELEE YERLAYLDAD
Sd   RAYATLVCTD AYAIGAQVLR ASLHRVGSTL PLVVLVTYDV APIPLRSHAW AKLRVFELEM FDTIVFLDAD
Ac   EAFVTLLSSR SYYPGVVALA RSLRQFSA-R ELLVLTTVPV ERVPPPEDCF TKFRMFELKN YTKFVYLDAD
Ba   EAYVTHLTND QYIKGAQVLA ESLREAGATR PPLAMITVPE FGDGRKDGFF TKLEAWRLPC -TRVIYLDTD
Ab   FAYVTVHYDQ EYVLGIQVLM QSIKLSGTRH DLVVLVSVVD ITNPFLNHTL NKLHVWNLLE YDRVVYLDAD
Pi   FAYVTVHYDA EYVLGVQVMM HSIKLTGSPY DLVVLASVTN IDNPFVGYTL NKLHVWNMLE YERVVYLDAD
Ng   YAYATLVSSE GYLSGALAMY KSIIARGGKY DLVLVVTASY IDNPNAKDTY NKLHIWKLDQ YKRLVFVDSD
Vc   EAYATLVYGE DFVLAARVLG QSLRESGTTR DMVALTTVAP VKNPGTGYVY TKLYIFQMTE YKKIVFLDAD
Gs   YAYATLLCDD VMLPATRAWL QSLKMTNTSF PIVVLVLVTP LEYPFTLCRY SKLHLWNLLN YDKVVYMDSD
Cm   YAYATLLCDE RMLRAVAALV HSLRVRNTSY PILVLTTREP LPYPFALCRY AKLHLWSLTT YEKIVFLDGD
Rg   YAYITLVTNA DYAKGATALV RSLRLTKTAA NIVVLHTIAL APLADLGCNF CKLRLWQLTE YERIVFIDAD
```


80
90
100
110
120
130
140
Hs TLVLANIDDL FDREELSAAP DPGWPDCFNS GVFVYQPSVE TYNQLLHLAS EQGSFDGGDQ GILNTFFSSW
Mm TLVLSNIDDL FEREELSAAP DPGWPDCFNS GVFVYQPSIE TYNQLLHLAS EQGSFDGGDQ GLLNTYFSGW
DM TLVLQNCDEL FEREELSAAP DVSWPDCFNS GVFVFKPSVD TFAQITEFAV KNGSFDGGDQ GLLNQFFADW
Mb TLVLTNIDEL FERNCFAAAP DIGWPDCFNS GVFVFQPSSA KFEDLVRLLA STGSFDGGDQ GLLNEYFADW
Hr TLVLQNVDDL FSRDPFAAAP DAGWPDCFNS GIFLYQPSFE MYGDLLQFAL KIGSFDGGDQ GLLNLFFSDW
Ta TLVIGNIDDL FTRPELSAAP DVGWPDCFNS GVFVYKPSMQ TYQTIVAFAL QFGSFDGGDQ GLLNEFFNTW
Nv MLVLQNCDEL FDRCELSAVP DIGWPDCFNS GMFVFEPSRA THEALLKYAI DHGSFDGGDQ GLLNSFFSQW
Co TLVVQNVDEL FDRPEIAAAP DVGWPDCFNS GVFVFVPSAA TFEKLAEHAV STGSFDGGDQ GLLNTFFDYW
Pm AFCLKNIDEL FDLDTFAAVP DVGWPDIFNS GVFITKPNIS VYNSLLNLAK NSISFDGGDQ GLLNFYFSNW
Aq TLVLTNVDEL FEREEMSAAP DIGWPDLFNS GVFVFRPSLE TFASLLELAD KEGSYDGGDQ GLLNLYWRDW
Sc TLPLNKEFLL FDIMSVGAIA DIGWPDMFNS GVMMLIPDAD TASVLQNYIF ENTSIDGSDQ GILNQFFREW
Tg CIVLRPVDEL FLRQPPAFAP DIFPPDKFNA GVAVLKPDLG EYGNMVAAVE RLPSYDGGDT GFLNAYFSSW
Hm CIVLRPIDDL FLRQPPAFAP DIFPPDKFNA GVAVLKPDLD EYGKMVAAVE RLPSYDGGDT GFLNAYFSSW
Nc CIVLGPVDEL FLRKPPAFAP DIFPPDKFNA GVVVLKPDLG EYGKMIAAIE RLPSYDGGDT GFLNAYFSSW
Sn CVVLQNVDEL FERLSPAFAA DVFPPDRFNA GVIVLQPNVE LFSRMLRAAG LLPAADGGDT GFLNSFFSDW
Kb CVVLERVDEL FERPSPAFCP DVFPPDKFNA GVIVLSPSRE LFEKMQERIA ELPSHDGGDT GFLNAFFPDW
Vb CIVMRKIDCL FDPAAPAFAP DVFPPDRFNA GVMVIEPSLA VYEDLLAKRT VLRSYDRGDT GFLNAYFSGW
Tp CLVVKDVSHL LRVDSLAAAP DIFPPDKFNA GVMVLCPSKA VFNDMMARLN SCTSYDGGDT GFLNSYYPNW
Rf CIVVRDISDI FKLPDFAAAP DLCPPDHFNA GVLFIQPNVQ TFQQLLRNVA YVNSYDGGDT GFLNSYFNDW
Bn AIVVRNVDHL FKMIPFAAAP DIFPPDRFNA GVLVLQPNSV MFAYILRLAY GLGSYDGGDT GFLNRIFPRW
Pu AIVLQNVDEL FDRSTFAAAP DVFPPDRFNA GVLVIRPNKQ LFADLLAKAK ELKSYDGGDT GFLNAFFPKW
Ot CLISSNPENA FDRNSFAAAP DVFPPDRFNA GVLLIKPSMT VFRDMISKIL TFPAYDGGDT GFLNAYYPDW
Sl CLIMQNPENI FLRDTFAAAP DVFPPDKFNA GVLYIEPSMK IFTDLISKIQ ILSTYDGGDT GFLNAYFPNW
Ws ALVMEDLDEL FDREVFAAAP DVFPPDKFNA GVMVVVPSLI VLEDMMSKVE ELPSYDGGDT GFLNAYFADW
Ae AFVLANVDEV LERDIFAAAP DIFPPDRFNA GVLLLHPNAE LFQRLVSQSA QFQSYDGGDT GYLNAVFPDW
Al ALILTNIDEL FEMDTFAAAP DIFPPDRFNA GVLVIKPGKD VFENLLAKAK TIKSYDGGDT GFLNLVFSDW
Ng CLVVEDIQEL FSADVFAAAP DIFPPDRFNA GVMLVRPNLD VYEDMLRAVG ALPSYDGGDT GFLNAFFPKW
Gt LLPLSSLAPL FDRDVVAAVP DISLPDHFNS ALVLLRPNLL HLQRLLALSS SLEPYDGGDQ GLLNEFFNAW
At LLILRNIDFL FSMPEISATG NNGTL--FNS GVMVIEPCNC TFQLLMEHIN EIESYNGGDQ GYLNEVFTWW
Os ILVLRDLDAL FGFPQLTAVG NDGSL--FNS GVMVIEPSQC TFQSLIRQRR TIRSYNGGDQ GFLNEVFVWW
Dd MLLLKSLDHL FDLVDLYAAI DADANSCINS GIMLLSPSID VYNLLIDGMK LPNQSTVNDQ DVINTTLPHW
Tv TLPTQRIDEL FNHSELSCVS DPMPPQICNT GLLVLEPNLT TFKHMKKLSD LYANNPPGDQ GFINFFFGQF
Cv MLVLRNIDHL FALPPFYAAP DCTAGRQFNA GFFLVTPSRA ELARFQSLLV RIGGY--AEQ DLLNEVLHEF
Sd MLCVRNMDDL FDAIAAASRA CTCNPQRFNS GMLVLHPSCA TLESLLAKLR SVERFVFSDQ CFLNEAFPDF
Ac MLVVGDVDEL FSYPSFAAAP NFQLKKSFNA GLFVVDRDEG LHRQFLDHYH YDKAWSWADQ SLLNDFFKKW
Ba ILAVGNPDVL FELAQFAVQD SQPHMQGPNT GVMVLKPDIR VYARIVETLT PLHEMPFYEQ GFIGKFFAKW
Ab NIVLRNADEL FMCGPFCAVF MNPCH--FHT GLLVVTPDKE EYQRLLHQLE YQSSFDGADQ GFLSSVYSEL
Pi NVLIRNSDEL FLCGEFCAVF MNPCH--FHT GLLVVTPSAA EYQRLLSALG HLESFDGADQ GFLSSMYSML
Ng CIIFKNVDLL FNCVGVCSGS DMGNTEFFNG GIMVLEPSTK TYDDMMDKMP AYKSYDGGEQ GFINLYFDFH
Vc VLVIRNMDVI FKCPGFCAAL RHSER--FNT GVMSLVPSLE MYDDMMAKMR SMPSYTGGDQ GFLNSYFPSF
Gs MLVMQNIDNL FVEFDLSACA DLYPDT-FNS GIMVIQPNET TFRNMKAVYK NVSSYNVGDQ GFLNWFFGEW
Cm TLVLAPIDDL FEKYDLAAAP DLYPET-FNS GVMVLEPRHD VYASMLARYR ETPSYNLGDQ GFLNSFFGQW
Rg AIILKNIDKL FAYPEFSAAP NVYETRRMNS GVFVARPSEE TFGRMLAMLD QPDAFRRTDQ TFLEAFFPDW



Sequence Ids are in Table 4.3

Hs: Homo sapiens
Mm: Mus musculus
DM: Drosophila melanogaster
Mb: Monosiga brevicollis
Hr: Helobdella robusta

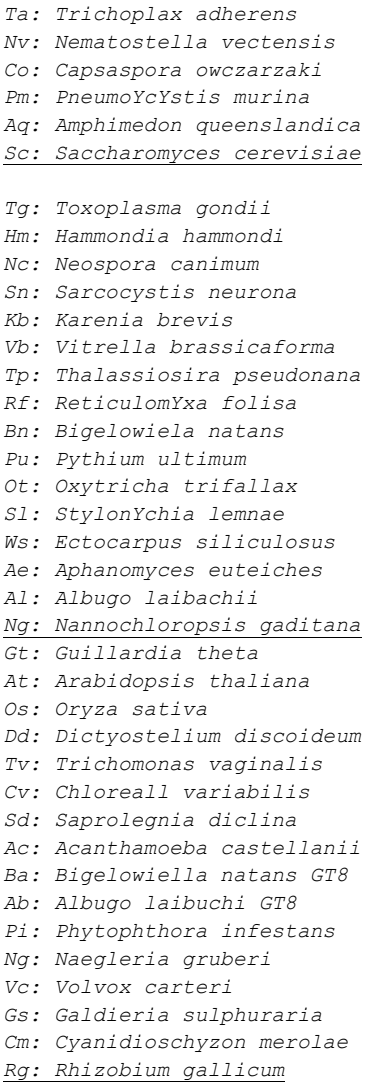


Figure 4.5. Alignment of glycogenins, Gat1, and GT8 sequences. The amino acid sequence of Gat1 (middle panel) were aligned with the amino acid sequences of representative known and predicted Glycogenins (top panel) or GT8 sequences (bottom panel) as described in "Experimental Procedures" and listed in **Table 4.3**. Amino acids are color-coded with respect to chemical similarities that guided the alignments, giving preference to the registration of hydrophobic residues: green, hydrophobic; blue, acidic; dark red, basic; black, polar; bright red, secondary structure breaking (P or G). To facilitate the aligning, positions occupied by identical amino acids across all the organisms are bolded. Unique motifs that are specific for glycogenins are shown in blue color or Gat1-specific motifs are shown in red color.

that contain Gnt1-like but lack PgtA/AgtA-like sequences, **(E)** closest Gat1-like GT8 family protein sequences from organisms that contain PgtA-like but lack AgtA-like sequences, and **(F)** closest Gat1-like sequence from bacteria. The aligned sequences were manually-curated in BioEdit (v 7.2.5) and a phylogenetic tree was made by Mega6 as described in the 'Experimental Procedures'. The result shows that both Gat1 and glycogenin form separate clades with a common ancestor (**Fig. 4.6**). Though the last common ancestor was not resolved, it can suggested that Gat1 was predecessor to glycogenin owing to its presence in more phylogenetically primitive eukaryotes, and the lack of both genes in the same species. Gat1 and glycogenins each have their own unique motifs **(Fig 4.5),** which might play role in their functional differences.

Gat1 May be Important for Type I Toxoplasma Growth– Previous studies show that *Toxoplasma* PhyA and earlier glycosyltransferases of the Skp1 modification pathway are important for parasite growth. To determine if Gat1 is also important, RHΔ*gat1*Δ parasites were inoculated on human foreskin fibroblast monolayers in parallel with RHΔΔ or RHΔ*phyA*Δ cells, and cultured for 5 days. Comparison of average plaque areas among these groups show that RHΔ*gat1*Δ parasites form plaque areas which are on average smaller than the RHΔΔ but larger than RHΔ*phyA*Δ **(Fig 4.7)**, suggesting its importance in optimum parasite growth. However, since this study was performed on a single clone, additional studies such as genetic complementation of the *gat1* locus are required to confirm that slow growth was the result of *gat1*-disruption.

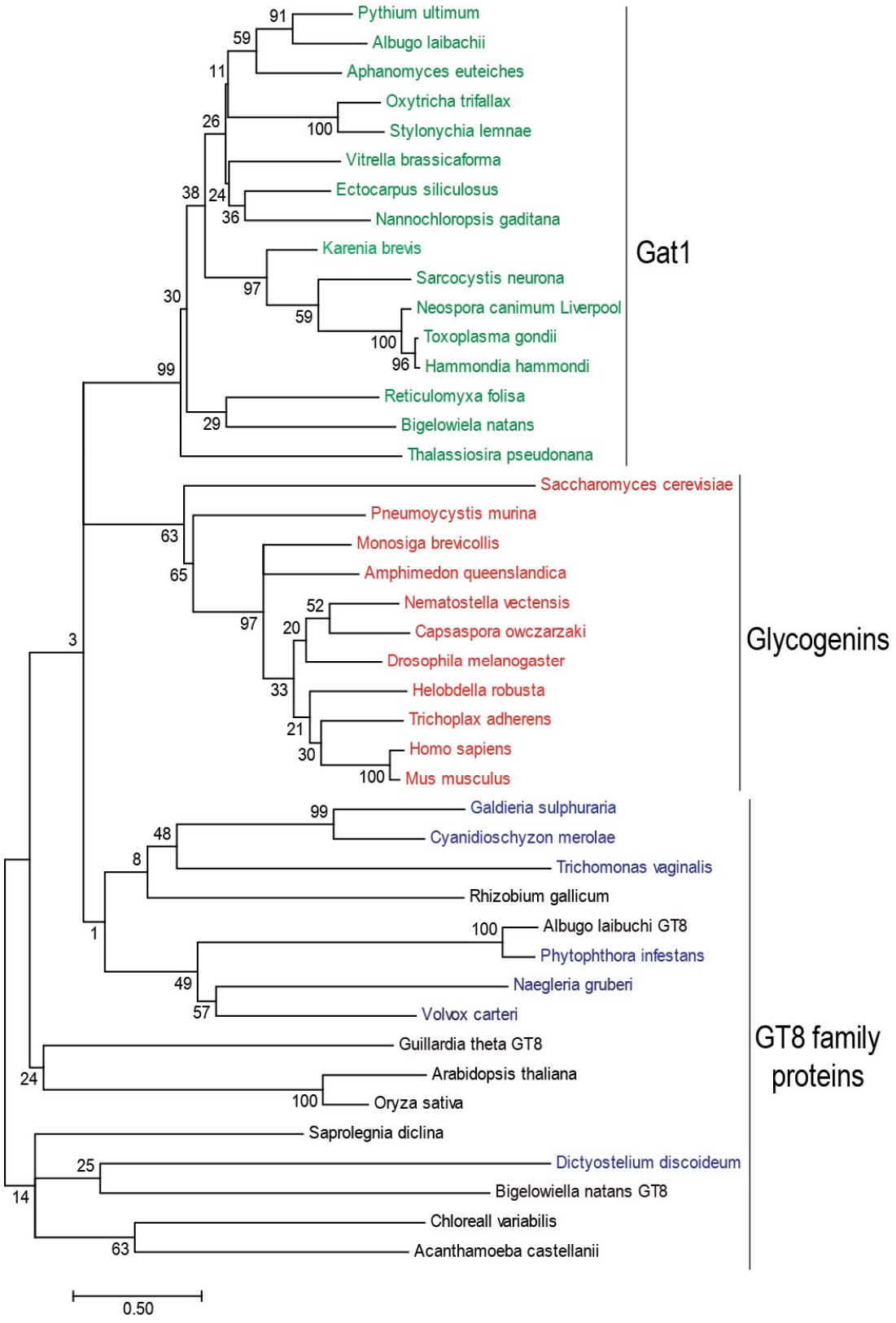

Pythium ultimum
Albugo laibachii
Aphanomyces euteiches
Oxytricha trifallax
Stylonychia lemnae
Vitrella brassicaforma
Ectocarpus siliculosus
Nannochloropsis gaditana
Karenia brevis
Sarcocystis neurona
Neospora canimum Liverpool
Toxoplasma gondii
Hammondia hammondi
Reticulomyxa folisa
Bigelowiela natans
Thalassiosira pseudonana
Gat1
Saccharomyces cerevisiae
Pneumoycystis murina
Monosiga brevicollis
Amphimedon queenslandica
Nematostella vectensis
Capsaspora owczarzaki
Drosophila melanogaster
Helobdella robusta
Trichoplax adherens
Homo sapiens
Mus musculus
Glycogenins
Galdieria sulphuraria
Cyanidioschyzon merolae
Trichomonas vaginalis
Rhizobium gallicum
Albugo laibuchi GT8
Phytophthora infestans
Naegleria gruberi
Volvox carteri
Guillardia theta GT8
Arabidopsis thaliana
Oryza sativa
Saprolegnia diclina
Dictyostelium discoideum
Bigelowiella natans GT8
Chloreall variabilis
Acanthamoeba castellanii
GT8 family proteins
0.50

Figure 4.6. Gat1 is phylogenetically related to mammalian glycogenin. An unrooted tree showing the phylogenetic position of Gat1 (green color), glycogenins (red color) and selected other CAZy GT8 sequences. Sequences having the predicted signal sequence or transmembrane domain are colored in blue. The evolutionary history was inferred by using the Maximum Likelihood method based on the Le_Gascuel_2008 model. The tree with the highest log likelihood (-13279.5590) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.1608)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 1.0204% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 43 amino acid sequences. There were a total of 196 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.

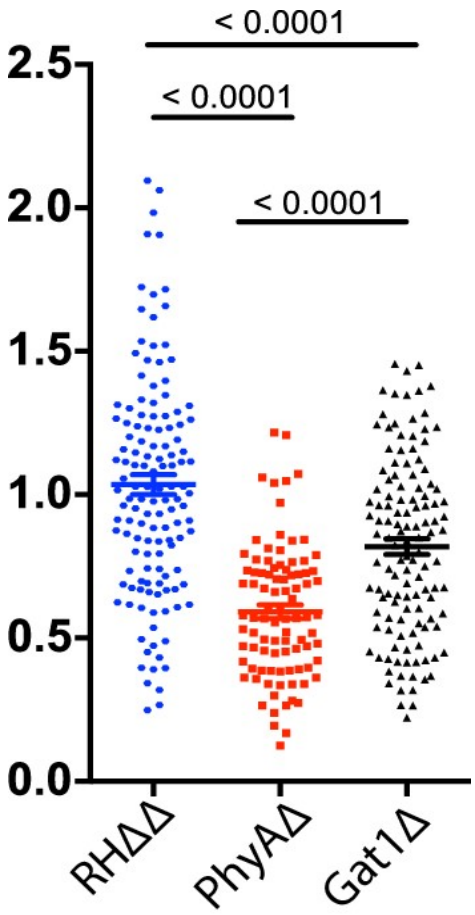


Figure 4.7. Cellular role of Gat1. HFF monolayers were inoculated with freshly isolated tachyzoite stage RHΔΔ, *phyA*Δ and *gat1*Δ parasites at a multiplicity of infection of 0.002. After 5.5 days, monolayers were fixed with methanol and stained with crystal violet. The images were digitized and plaque areas were measured manually by ImageJ software. The dot plots show the area distributions and average values ± S.E.M. from a representative experiment. P values for statistical significance of the differences, based on a one-way ANOVA test, are shown above.

Discussion

Here we have shown that Gat1, an enzyme from CAZy GT8 family, is required for addition of the 5th sugar of the *Toxoplasma* Skp1 pentasaccharide. TgSkp1 isolated from the parasite is at least partially susceptible to loss of a single Hex in the presence of green coffee bean α-galactosidase. Thus the terminal sugar is probably Gal but the possibility that a fraction of Skp1 has a terminal Glc or any other sugar needs further investigation. It is also very likely to be α-linked, which is consistent with the retaining characteristic of the transferase activities of GT8 enzymes. Nevertheless, it is not clear yet from this analysis if Gat1 directly transfers the sugar on TgSkp1 or is involved in an indirect manner. The position to which the αGal residue is attached is also not revealed by these studies, which might be any of several –OH moieties on, according to prior MS fragmentation data (14), the 3rd (αFuc) or 4th (αGlc) sugars. Relevant to this issue, other studies in the lab (Mandalasi, M., van der Wel, H., West, C.M., unpublished data) show that a recombinant version of Gat1 has GalT activity toward Glc acceptors *in vitro*, consistent with its hypothesized direct role in modification of Glc on Skp1 with Gal in cells.

RHΔ*gat1*Δ clone 1 parasites have a growth defect that is intermediate between that of *phyA*Δ and wild-type parasites. Thus full glycosylation of Skp1 may be important to mediate the growth advantage afforded by the action of *phyA*, as described in *Dictsyostelium* (147). However, two caveats exist for this interpretation. First, genetic complementation of the gat1 locus and/or analysis of independent clones are needed to rule out the involvement of other unknown genetic lesions in clone 1. In this vein, *gat1* was recently disrupted by an independent method utilizing CRISPR/Cas9 targeting, and a provisional slow growth phenotype suggests that *gat1* does contribute to optimal growth in fibroblast monolayers (Ichikawa, H., Mandalasi, M., West, C.M.,

unpublished data). However, slow growth was not predicted by a fitness value of +2.45 (139), so further quantitative studies on genetically defined strains will be important.

The second caveat is that if Gat1 has a second function in cells, such as modifying other acceptors, any growth phenotype observed in the disruption strain might be due to abrogation of the function. The high sequence similarity between Gat1 and glycogenin, an αGlc-transferase that primes glycogen synthesis in animals and yeasts, raises the possibility that starch biosynthesis is affected in the mutant. Preliminary studies on recombinant Gat1 suggests glycogenin-like autoglucosylation activity (Rahman, K., van der Wel, H., West, C.M., unpublished data), and autogalactosylation activity, consistent with this possibility. Alternatively, since Gat1- and glycogenin-like genes have not been observed in the same genome, Skp1 glycosylation may have been the ancestral function of this gene lineage and that the autoglycosylation activity is vestigial. It is intriguing to speculate that this switch conserved a cellular function in nutritional regulation. The broad phylogenetic distribution of Glt1 and Gat1 in protists contrasts with the narrow distribution of their replacement homolog in amoebazoa, AgtA. This suggests that the Glt1/Gat1 is ancestral, and the switch that is evident now only in amoebazoa may have resulted from a loss of Glt1, which liberated Gat1 to undergo the hypothesized repurposing to a direct role in priming glycogen synthesis.

CHAPTER V

SUMMARY AND FUTURE DIRECTIONS

Summary

Toxoplasmosis is one of the most common parasitic infections in the warm-blooded animals, which can be devastating to immunocompromised human patients such as HIV-infected or bone-marrow transplant patients. Due to its intracellular nature, finding a *Toxoplasma*-specific drug target is challenging. Skp1-glycosylation in *Toxoplasma* is a parasite-specific pathway that is absent from its mammalian host. Gaining fundamental knowledge on the biochemical and cellular roles of this pathway might help exploit it against the parasite and other pathogens having the common target. In this dissertation, I attempted to understand the basic nature of Skp1-modification in *Toxoplasma*, and characterize the glycosylation pathway and its role in parasite proliferation in fibroblasts. I also investigated its phylogenetic distribution to consider applicability of my findings to other organisms. Although *Toxoplasma* and *Dictyostelium* are unrelated protists, studies from *Dictyostelium* Skp1 modification pathway, where Skp1 glycosylation is fundamentally important for its O_2-dependent development, has guided us throughout the studies. Using mass spectrometry, reverse genetics approaches and growth assays, we found that *Toxoplasma* has a unique but partially overlapping Skp1-glycosylation pathway in comparison to *Dictyostelium*, which is important for parasite growth. Our findings in *Toxoplasma* could be summarized as follows:

A. Like *Dictyostelium,* Skp1 in *Toxoplasma* is modified at an equivalent Pro residue by a hydroxylation-dependent pentasaccharide, which is evidently linear. The glycan chain

consists of Galα-Glcα-Fucα-Galβ-GlcNAcα-. The core trisaccharide is thought to be the same as that of *Dictyostelium* Skp1, Fucα1-2Galβ1-3GlcNAcα-, and linked to 4(*trans*)Hyp, based on enzyme homology and preliminary studies using β-galactosidases. The outer disaccharide differs from that of the *Dictyostelium* glycan structure, by having an αGlc in place of an αGal at the 4th position. The regiospecificity of the outer two *Toxoplasma* sugar linkages is not known.

B. *Dictyostelium* homologues of Gnt1 and PgtA in *Toxoplasma* form the similar core-trisaccharide, though the physical order of the GT domains of PgtA are reversed in *Toxoplasma*. Two novel glycosyltransferases, Glt1 and Gat1, replace the function of AgtA in *Toxoplasma* and mediate the fourth and fifth sugar additions on TgSkp1, respectively.

C. The *Toxoplasma* pathway is evidently ancestral. The last two enzymes are absent from *Dictyostelium* but widely distributed in select group of protists (**Fig 5.1**) that have potential *Toxoplasma*-like Skp1 modification pathway genes (*phyA, gnt1, pgtA, glt1* and *gat1*, but lack *agtA*) (**Table 5.1**). The genes seem to be lost or switched their targets in organisms where the pathway deemed unnecessary. Gat1 appears to be the ancestor of glycogenin, the enzyme important for initiation of mammalian and yeast glycogen synthesis.

D. All the glycosyltransferases are important for parasite proliferation in fibroblasts suggesting that PhyA mediates its function through glycosylation. Among them Gnt1 is the most important for the parasite growth as *gnt1* disrupted parasites show almost similar growth defect as *phyA*Δ parasites (Chapter 2). Other glycosyltransferase mutants show intermediate growth rates between wild-type and *phyA* disrupted parasites based on 5-day plaque assays. Therefore, it appears that full glycosylation is important for the PhyA-mediated function.

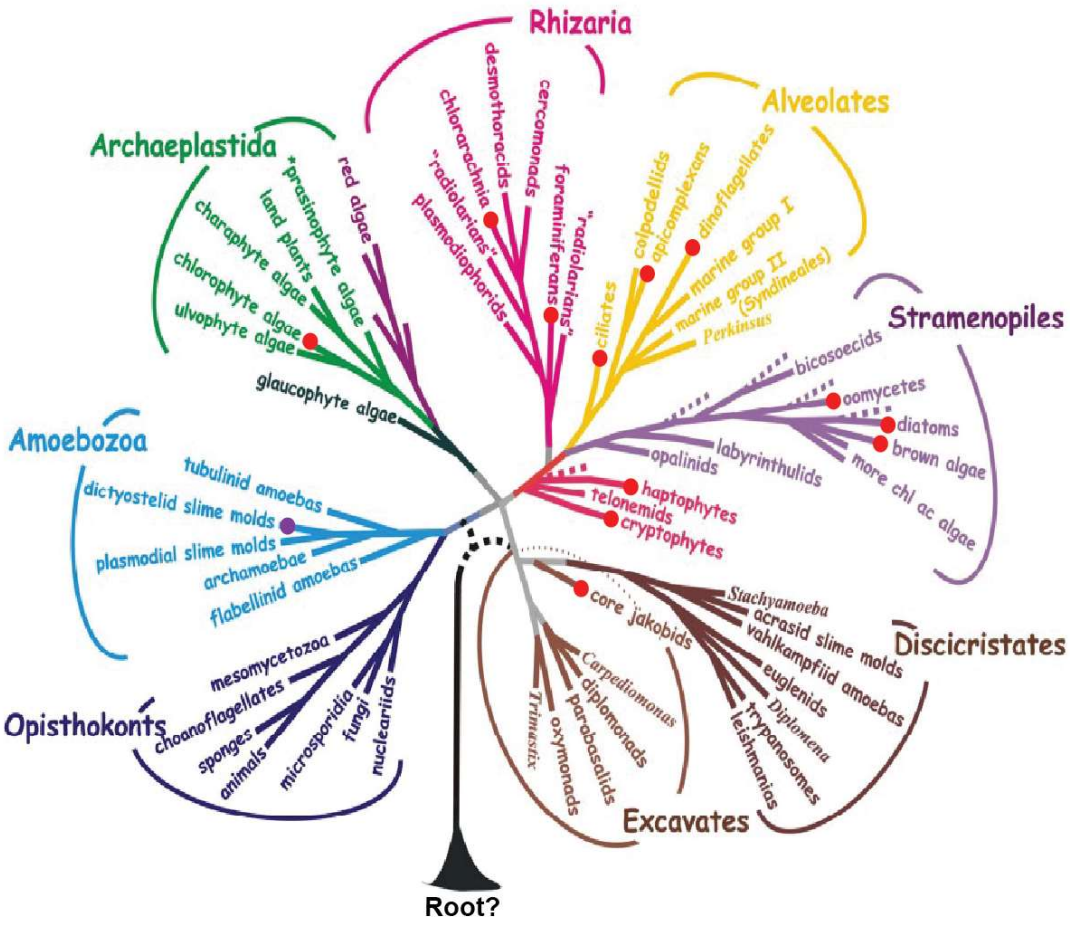


Figure 5.1. *Toxoplasma*-like Skp1 modification pathway is more widely distributed among unicellular protists than *Dictyostelium* according to genomic analyses. An illustration from Baldauf (180) that presents a consensus phylogeny of the eukaryotes. Groups or organisms where *Toxoplasma*-like Skp1 pathway gene candidates (i.e. presence of *glt1* and *gat1* in addition to *phyA*, *gnt1* and *pgtA*) are identified are marked with red dots. AgtA is restricted to *Dictyostelium* and other amoebazoans, which also lack *glt1* and *gat1* genes (marked with purple dot).

Table 5.1. Phylogenetic distribution of Skp1 modification pathway-like genes. An updated version of Table 1.2 (chapter 1) that additionally includes phylogenetic distribution of our newly discovered GT genes from *Toxoplasma.* Before, BLASTP and TBLASTN searches were performed for sequences corresponding to *Dictyostelium* Skp1 and its Hyp-modification pathway related genes (*phyA*, *gnt1*, *pgtA* and *agtA*) and their phylogenetic distribution was listed in Table 1.2. The presence (+), dual functional (---------) and reverse orientation of the domains (R) relative to *Dictyostelium* enzyme is shown. A similar search was performed for Glt1 and Gat1 that mediates addition of the final 2 sugars on *Toxoplasma* Skp1 as described in chapter 2 and 3 and the organisms having the sequences are added with Table 1.2 to generate this table. The *Dictyostelium* Skp1 modification genes are on the top row. Bioinformatics studies predicted that DdAgtA (highlighted in green) is substituted by these two new genes (highlighted in purple), which coexist in numerous other protists that also possess the underlying GT, PgtA.

		Skp1-	P4H1	Gnt1	PgtA		AgtA		Glt1	Gat1
Group	**Organism**	**Pro**	**PhyA**	**α-GlcNAc**	**β1,3Gal**	**α1,2Fuc**	**α1,3Gal-1**	**α1,3Gal-2**	**α-Glc**	**α-Gal**
Amoebazoa	*Dictyostelium discoideum* (cellular slime mold)	+(2)	+	+	+-------	-------+	+-------	-------+		
	D. purpureum (cellular slime mold)	+	+	+	+-------	-------+	+-------	-------+		
	D. fasciculatum (cellular slime mold)	++(2)	+	+	+-------	-------+	+-------	-------+		
	Polysphondylium pallidum (cellular slime mold)	++(2)	+	+	+-------	-------+	+-------	-------+		
	Acytostelium subglobosum (cellular slime mold)	++(2)	+	+	+-------	-------+	+-------	-------+		
	Physarum polycephalum (acellular slime mold)	+	+	+	+-------	-------+	+-------	-------+		
	Acanthamoeba castellani (Neff)(soil protozoa)	+	+	+						
Rhizaria	*Reticulomyxa filosa* (fresh water foraminifera)	+(4)	+	+	+	+			+-------	-------+R
	Bigelowiella natans CCMP2755 (chlorarachnea)	+(?more)	+	+		+			+-------	-------+R
Alveolates	***Toxoplasma gondii*** (apicomplexan)	+(2)	+	+	+-------	-------+R			+	+
	Hammondia hammondi (apicomplexan)	+	+	+	+-------	-------+R			+	+
	Neospora caninum (apicomplexan)	+	+	+	+-------	-------+R			+	+
	Sarcocystis neurona (apicomplexan)	+	+	+	+-------	-------+R			+	+
	Chromera velia (chromerida)	+	+	+	+-------	-------+R			+	+
	Vitrella brassicaformis (chromerida)	+	+	+	+-------	-------+R			+	+
	Stylonychia lemnae (ciliated)	+	+-------	-------+	+-------	-------+R			+	+
	Oxytricha trifallax (ciliated)	+	+-------	-------+	+-------	-------+R			+	+
	Simbiodinium microadriaticum (dinoflagellate)	+	+	+	+-------	-------+R			+	+
	Karenia brevis (marine unicellar dinoflagellate; red tide)	+	+	+	+-------	-------+R				+
Chromalveolate	*Nannochloropsis gaditana* (marine microalga)	+	+-------	-------+	+-------	-------+R			+	+
	Guillardia theta CCMP 2712 (cryptophyte)	+	+-------	-------+		+			+	+
Heterokonts/	*Thalassiosira pseudonana* (diatom)	+	+-------	-------+	+-------	-------+R			+	+
Stramenopiles	*Albugo laibachii* (oomycete)	+	+-------	-------+	+-------	-------+R			+	+
	Pythium ultimum (oomycete)	+	+-------	-------+	+-------	-------+R			+	+
	Aureococcus anophagefferens (non-motile pelagophyte)	+		+	+-------	-------+R			+	+
	Ectocarpus siliculosis (brown alga)	+(2)	+-------	-------+	+-------	-------+R			+	+
	Aphanomyces euteiches, strain: 109 (oomycete)	+(2)	+-------	-------+	+-------	-------+R			+	+
	Phytophtera infestans (oomycete)	+	+-------	-------+						
	P. ramorum (oomycete)	+	+-------	-------+						
	P. sojae (oomycete)	+	+-------	-------+						
	Phaeodactylum tricornutum (marine pennate diatom)	+	**+(2)**	+						
	Fragilariopsis cylindrus (coldwater diatom)	+	+-------	-------+						
	Emiliania huxleyi (marine haptophyte alga)	+	+-------	-------+						
	Chlamydomonas reinhardtii (chlorophyte alga)	+	+	+						
	Volvox carteri (chlorophyte alga)	+	+-------	-------+						
	Coccomyxa subellipsoidea (chlorophyte alga)	+	+-------	-------+						
	Chlorella variabilis (chlorophyte single-cell green alga)	+	+-------	-------+						
Excavates	*Malawimonas californiana*	+	+	+						
	Naegleria gruberi (amoebe flagellate, heterolobosea)	+	+-------	-------+						

Although our studies did not investigate the significance of partially distinct glycan structure on TgSkp1 function, by analogy to the *Dictyostelium* it can be proposed that Skp1 glycosylation might have impact on the structure of Skp1 (11) and affect SCF-complex assembly (12). In fact, a recent analysis of the TgSkp1 interactome (**Fig. 5.2**) detected a cullin-like protein, predicted to be a Cul1 homolog, and a predicted F-box protein (TGGT1_310930), suggesting the existence of SCF complex. Several proteins were differentially enriched in normal or *phyA*$^{-}$ cells. Differences shown were validated in an independent trial using TgSkp1-TAP tagged strains (160). Interestingly, most proteins preferentially associated with unmodified Skp1 and only a single hypothetical protein TGGT1_259880 preferentially binds the modified Skp1. Recent studies by Ashely Dittmer in our collaborator Ira Blader's lab has confirmed that TGGT1_310930 is an F-box protein (named as FbxO1) that binds directly and preferentially with unmodified Skp1, based on Western blotting (unpublished data). The significance of this preferential binding with unmodified Skp1 as well as its potential role in parasite growth is currently under investigation. On the other hand, the role of TGGT1_259880, the only protein that preferentially binds glycosylated Skp1, is also unknown due to the absence of any known functional protein domains. Preliminary BLASTP analysis against the NCBI non-redundant database suggests that the sequence is conserved only in apicomplexans and absent from other sequenced organisms at $E<10^{-1}$. Further studies are needed to decipher any role of this protein in Skp1 glycosylation regulated pathway.

While in *Dictyostelium*, glycosylation pathway mutants affect development of the slug to a fruiting body, in *Toxoplasma* the effects are on generalized parasite growth as evident from the smaller plaque areas of PhyA and GT mutants in comparison to the wild-type parasites. The observed growth defects might not result from a defect in parasite invasion ability since the plaque

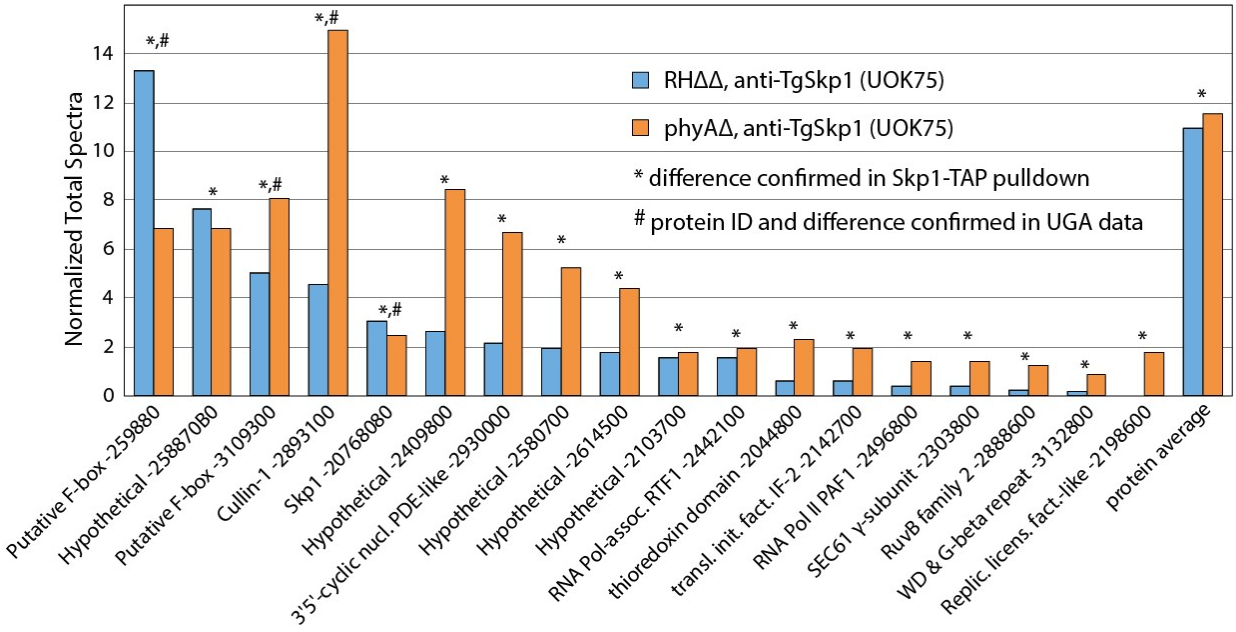


Figure 5.2. TgSkp1 interactome is affected by Skp1's modification. 6 x10^8 RHΔΔ or *phyA*Δ parasites were lysed in 1% NP-40 for 10 mins, precleared with anti-rabbit IgG beads for 1h at 4ºC and Skp1 was immunoprecipitated with anti-Skp1 (pAb UOK75) bound beads for 1 hour as described (160). The samples were analyzed on an Orbitrap MS at Ok. State Univ as described (12). Normalized spectral counts are shown for protein IDs (1% false discovery rate) that were not found in parallel control co-IPs with native rabbit IgG. Differences between strains that were confirmed in similar co-IPs with anti-FLAG (M2) performed in Skp1-TAP strains (149) are denoted by an asterisk. Trends confirmed in a separate (UGa) dataset are denoted by # (and no contradictions were seen for other proteins).

formation efficiency was similar for both wild-type and pathway mutants (unpublished data). The critical role of SCF complexes in regulation of mammalian and yeast cell cycle suggests that the observed growth defect might result from the defects in parasite proliferation inside the PV. Skp1 was originally discovered as a connector of cell cycle regulators to the SCF complex through the F-box proteins for their timely degradation by the proteasome (181). As described in chapter 1, Skp1 mediates the degradation of a number of cyclin-dependent kinases (CDKs) and CDK inhibitors for the G1/S transition as well as mitotic entry. Although it is too early to predict how TgSkp1 modification affects parasite growth, however, an indirect glycosylation-dependent regulation of the cell cycle regulators could be anticipated. Based on our findings in *Toxoplasma* so far and by analogy to *Dictyostelium* studies, we propose a model for the potential role of TgSkp1 modification (**Fig 5.3**) on cell cycle regulation. Reduced pull down of FbxO1 with modified Skp1 supports two hypothesis: (a) glycosylation causes such a conformational change in TgSkp1 that decreases its affinity for binding to certain F-box proteins that result in reduced degradation of cell cycle specific activators to promote cell cycle progression, or (b) glycosylation-dependent conformational changes in Skp1 increase the affinity for binding to certain F-box proteins that result in their autoubiquitination and degradation by the 26S proteasome resulting in upregulation of cell cycle activators for timely cell division. This is also causes less F-box protein for co-IP with modified Skp1. However, increased binding of TGGT1_259880, a potentially non-F-box protein, with modified Skp1 instead suggests another novel mechanism where glycosylation on TgSkp1 increases its affinity for certain proteins that act as inhibitor for binding to certain F-box proteins, thereby resulting in increased abundance of cell cycle activators for timely cell cycle progression. Further studies are needed to investigate if Skp1 glycosylation regulates parasite proliferation using any of these or other novel ways.

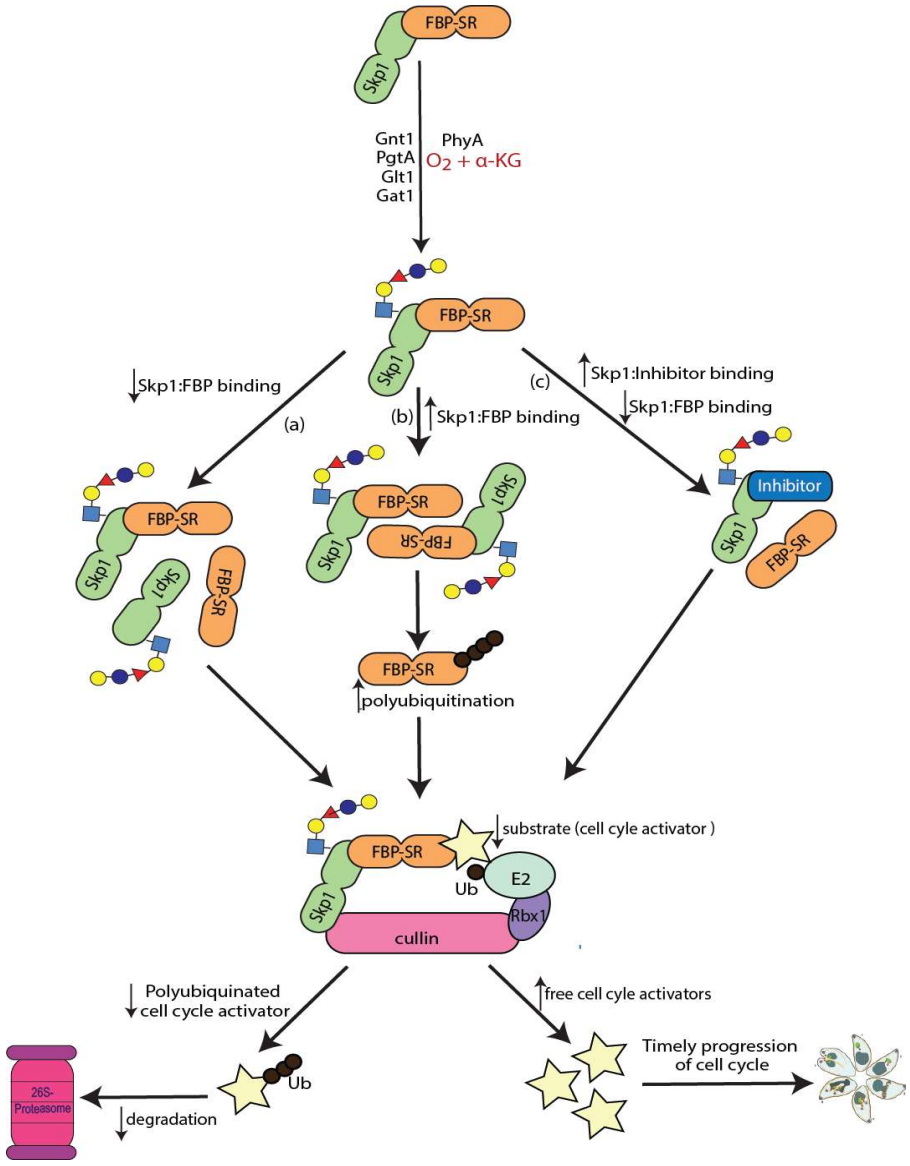


Figure 5.3. Model for the role of glycosylation of Skp1 in parasite growth. Three hypotheses on how TgSkp1 modification might affect parasite proliferation. (a) Hydroxylation of Pro154 of TgSkp1 and its subsequent glycosylation might decrease its association with specific Fbox proteins, or (b) increase its association with certain F-box proteins potentially leading to the autoubiquitination and degradation of the F-box proteins, or (c) inhibits the binding of certain F-box proteins through glycosylation-dependent binding of certain inhibitor proteins. Any of these mechanisms or their combinations might allow for the reduced turnover of substrates/cell-cycle-activators specific to the F-box proteins that result in timely cell cycle progression. Accumulation of substrates/activators might create a negative feedback loop that ultimately results in their degradation by the proteasome (not shown).

Future Directions

Further Biochemical and Functional Characterization of TgSkp1 Glycosylation and Pathway Enzymes- While the basic nature of Skp1 glycosylation has been elucidated in this study, still questions remain to be answered for better understanding the role of this pathway that include:

Linkage Analysis- Defining the glycosidic linkages and anomeric configuration is important to elucidate the complete structure of Skp1 glycans and its impact on Skp1 structure/function. Since, homologous enzymes mediate the core trisaccharide in *Dictyostelium* and *Toxoplasma*, we expect same core trisaccharide with the following sequence Fucα1,2Galβ1,3GlcNAcα1-. To address whether Gal has been linked with GlcNAc in a β1,3 linkage, an *in vitro* reaction was performed where exogenous GlcNAc-DdSkp1 was incubated with UDP-[^{3}H]-Gal and desalted wild-type parasite cytosolic extracts as described (160). The purified reaction product containing [^{3}H]-Gal-GlcNAc-DdSkp1 was trypsinized and incubated with β1,3 or β1,4/6 galactosidases to release the [^{3}H]-Gal as described in **Fig. 5.4**, which showed selective sensitivity to β1,3 galactosidase as seen for *Dictyostelium* Skp1 (109). To check the linkage between Fuc and Gal we have done a similar reaction but instead of UDP-[^{3}H]-Gal, cold UDP-Gal was used with GDP-[^{3}H]-Fuc and desalted *glt1Δ* parasite cell extracts. The radioactive fraction containing the [^{3}H]-Fuc-Gal-GlcNAc-DdSkp1 product was purified over the Q-column and the trypsinized product will be treated with α1-2, or α1-3/4 fucosidases. However, we have no precedent for the linkage between αGlc and αFuc (fourth and third sugars). Due to the assignment of Glt1 in CAZy family 32 we expect that it mediates α-linkage. To determine the nature of this

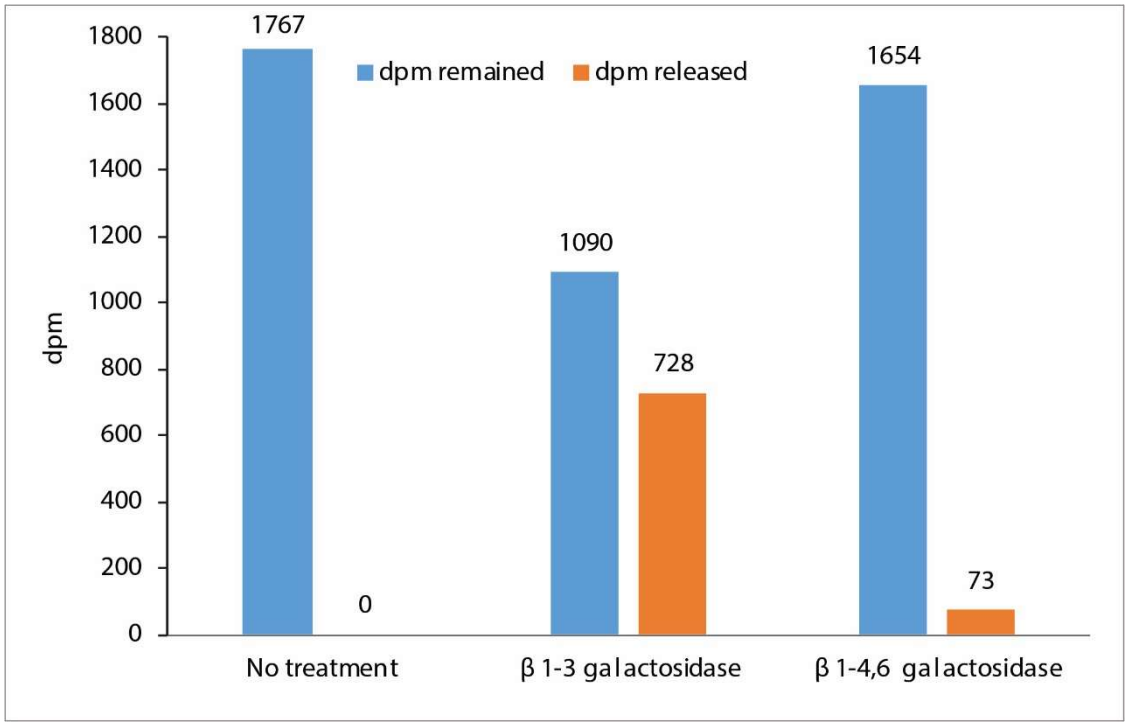


Figure 5.4. Galactosyltransferase assays in RHΔΔ parasite extracts. 100 µl of desalted S100 extracts of RHΔΔ containing 1 mg protein/ml, were incubated with 1 µCi UDP- [^{3}H]Gal and 100 pmole of Dd Gn-kp1 for 3h as described (160). The reaction product was applied to a 1-ml Hi-Trap Q-Sepharose column (GE-Healthcare) pre-equilibrated at 4°C in 50mM Tris pH 7.4, 5mM $MgCl_2$ and 0.5mM EDTA to separate the proteins from UDP-[^{3}H]Gal. Protein was eluted with a gradient of NaCl from 0 to 1M prepared in the same buffer supplemented with 1 µg/ml each of the protease inhibitors aprotinin and leupeptin. The radioactive protein fractions were trypsinized to generate low M_r peptides, which were then purified by adsorption to a C_{18} Toptip catridge (PolyLC InC. Item # TT200MC18) pre-equilibriated with 0.05% (v/v) trifluroacetic acid and eluted with 60% acetonitrile in 0.05% trifluroacetic acid. The sample was vacuum centrifuged, dissolved in 10µl water, boiled for 10 mins to heat inactivate the trypsin, vacuum centrifuged again and re-dissolved in exoglycosidase preparation to a final volume of 10 µl 25mM NH_4Ac (pH 5.5) containing 20 ng of aprotinin (trypsin inhibitor) and 10 milliunits of either β1,3-galactosidase (NEB), β1,4/6-galactosidase (Glyko) or buffer only (untreated) and incubated for 5 h at 37°C. Released radioactivity was quantitated by liquid scintillation counter after passage through the C_{18} Toptip column as described above.

linkage a glucosyltransferase reaction will be performed in the presence of FGGn-pNP, UDP-[1-^{13}C]-Glc and recombinantly expressed His_6-Glt1 as described in chapter 3 and NMR spectroscopy will be performed to assign the linkage. Previous studies have shown that a non-specific α-galactosidase treatment partially released terminal galactose (chapter 4) from the endogenous Skp1 pentasaccharide suggesting that the terminal galactose is α-linked to the glucose. To confirm the specificity of the linkage, we will initially generate the Skp1-tetrasaccharide glycans, GFGGn-pNP by incubating FGGn-pNP, cold UDP-Glc and rGlt1 and which will be purified by C18 Sep Pack column as described in chapter 3. The resulting product will be incubated with UDP-[1-^{13}C]-Gal and rGat1 to make the pentasaccharide which will be analyzed by NMR spectroscopy to assign the linkage.

Investigate for Additional Endogenous Target of Glt1/Gat1- Previous studies have suggested that Skp1 is the sole target of *Dictyostelium* enzymes (10), which is the basis for predicting that *Toxoplasma* Gnt1 and PgtA target only Skp1 due to their homologous characteristics. Our previous data indicate that Skp1 might be the only target of *Toxoplasma* Glt1 as well (chapter 3). However, it is unknown if Gat1 has any other target(s) or role(s) inside the cell in addition to Skp1 glycosylation. To investigate if Gat1 has any other cellular target(s), an experiment could be performed where recombinantly expressed His_6-Gat1 could be added in a reaction containing UDP-[^{3}H]-Gal and desalted wild-type or *gat1*Δ parasite cell extracts. The reaction products could be loaded into the SDS-PAGE gel, sliced each lane into multiple pieces to monitor the incorporation of radioactivity as a function of migration. If Gat1 has any other cellular target(s), we expect that upon addition of His_6-Gat1, *^{3}H*-incorporation could be detected at the appropriate gel slice from *gat1*Δ reaction products, but not from the wild-type reaction product as in the former the target(s) would be modified by rGat1 with radioactive ^{3}H-Gal, but in the latter

the target(s) are expected to be already modified by endogenous Gat1 and therefore no incorporation is expected.

Determine if Gat1 has Any Role in Parasite Amylopectin Synthesis- As described in chapter 4, Gat1 has very high sequence similarity with glycogenin, a glucosyltransferase responsible for mammalian and yeast glycogen synthesis. In fact, X-ray crystallography of a Gat1 homolog from *Pythium ultimum* (John Kim, Zachary Wood, Chris West; unpublished data) shows extensive structural conservation between Gat1 and glycogenin. *Toxoplasma* produces amylopectin which is glycogen-like but has fewer α1,6 branches. To test if Gat1 is responsible for the amylopectin synthesis in addition to TgSkp1 modification in *Toxoplasma*, an experiment could be performed where amylopectin synthesis could be induced in the type I RH parasites by stress as described earlier (176). The amount of amylopectin generated in wild-type and *gat1Δ* parasites could be compared by analysis of extracts for Glc after acid hydrolysis by high performance Chromatography (Dionex), or via monitoring the amylopectin granules by electron microscopy. If Gat1 is responsible for initiating amylopectin synthesis in *Toxoplasma*, then we expect that *gat1*Δ parasites would be deficient of that. On the other hand, *Toxoplasma* bradyzoites accumulate significantly more amylopection than tachyzoites (25) and a number of studies suggest that bradyzoite metabolism is dependent on the catabolism of amylopection to lactate (182, 183). Therefore, an alternative way to test the importance of Gat1 in amylopectin synthesis is through disrupting the *gat1* gene from the Type II Me49 strain that *in vitro* converts to bradyzoites from tachyzoites under alkaline pH, and the amount of amylopectin in bradyzoites could be compared between wild-type and *gat1Δ* parasites as described above.

Study the Role of Skp1 Glycosylation in Parasite Growth and Virulence at Low O_2- In *Dictyostelium*, PhyA is required for culmination to fruiting bodies (111) and *phyA* mutants have a higher O_2 requirement for culmination, suggesting its role as O_2 sensor to drive the process (131). Each of the GTs also have variable roles in this developmental process (10). As described in chapter 1, after digestion of *Toxoplasma* cysts, the released parasites infect the intestine, which is one of the most O_2 poor regions of the body, and then disseminate to the brain, retina, and other parts of the body. Previous studies have shown that *phyAΔ* parasites that lack the complete modification on TgSkp1 show a more severe growth defect at 0.5% O_2 than atmospheric O_2 level suggesting the importance of PhyA in parasite O_2 sensing and the importance of complete modification in this process (16). Although we have shown that all the GTs are important for parasite growth at atmospheric O_2 level, but their importance in low O_2 has not been tested yet. To evaluate the importance of GTs in parasite growth at low O_2 level, plaque areas could be compared among wild-type and individual GT KO parasites at 0.5% O_2 similarly as tested for in the atmospheric O_2 levels (chapters 2-4). In addition, to study the importance of overall glycosylation in parasite virulence, mice could be gavage feed with wild-type and *phyAΔ* tissue cysts containing bradyzoites, and after a certain period tissue cysts could be harvested from the brains to compare the parasite load as described before (184). Dr. Msano Mandalasi in our lab has generated a *phyA* mutant type II Me49 strain in that purpose, which should be suitable for mice studies due to less virulence properties of type II strains compared to the type I RH strain.

Investigate Specificity of the Glycan Sequence- The unexpected finding that the TgSkp1 glycan partially varies from the DdSkp1 glycan would permit examination of the importance of the precise structure by gene swapping between *Toxoplasma* and *Dictyostelium*. To investigate if *Dictyostelium* AgtA complementation could rescue the defect in Skp1 glycosylation as well as the growth defect of Glt1-

disrupted parasites, we have generated plasmids to overexpress the codon optimized *Dicytostelium purpureum* catalytic-only (AgtA-CAT) or full-length AgtA (AgtA-FL) (113) as well as plasmids to express them from the endogenous Glt1 locus. In addition, to assess the importance of WD-40 domain of AgtA, AgtA-FL overexpression or endogenous constructs were generated with enzymatically dead catalytic site. All the plasmids will be transfected into the *glt1*Δ parasites, and Skp1 from the complemented clones will be immunoprecipitated and analyzed by mass spectrometry to determine the nature of TgSkp1 sugars and the growth rate of the parasites will be compared with the wild-type parasites as described in chapters 2-4. In addition, Angela Park in our lab will express codon optimized Glt1 and Gat1 from the endogenous locus of AgtA, or overexpress them in *agtA*-null background *Dictyostelium*. The Skp1 from the complemented mutants will be analyzed by mass spectrometry, and its effect on *Dictyostelium* development will be investigated as described before (113).

Functional Validation of TgSkp1 Interactome and Identify Additional Components of TgSkp1 Signaling Pathway- As described above and shown in **Fig 5.2**, we observed a glycosylation dependent variation in several putative TgSkp1 binding partners. These preliminary studies need additional trials for confirmation. Among them, TGGT1_310900 (Fbox1) has been validated as a F-box protein as described above and validation of the others remains to be done. Tagging of additional genes, examining the effect of Skp1 modification on the level and possibly localization of the encoded protein, and reciprocal co-immunoprecipitation, is a useful strategy to confirm their interaction differences. To decipher the TgSkp1 signaling cascade any of the interesting genes that shows glycosylation-dependent binding affinities could be tagged and co-immunoprecipitated to find binding partner candidates. Any putative cell cycle regulators could be investigated further by reverse genetics approaches for their involvement in cell cycle regulation.

Development of Inhibitors Against TgPhyA and Pathway Glycosyltransferases- Prolyl 4-hydroxylases (PHDs) and the larger family of αKG dioxygenases are now being targeted for therapeutic intervention for diseases including anemia, inflammation and cancer (185, 186). The importance of TgPhyA in parasite growth suggests its potential as an important drug target. Our collaborator Christopher Schofield at the University of Oxford has an extensive collection of >1000 compounds with the potential to inhibit the PHD-like oxygenases and developed a quantitative proteomics assay suitable for screening the compound library (187), which could be tested against TgPhyA due to close enzymatic and structural similarities with other PHDs (104). In fact, a recent screen revealed several compounds with sub-μM IC50 values toward TgPhyA and DdPhyA (unpublished data). The profile varied from the inhibitory profile toward host PHD2, and studies done by Ashley Dittmar in our collaborator Ira Blader's lab indicated that two of those inhibitors inhibit *Toxoplasma* growth in the cell culture. Future studies will focus on improvement of compounds that selectively inhibit TgPhyA, and address their actual targets in cells.

On the other hand, the importance of Skp1 modifying glycosyltransferases in parasite growth also suggests their potential as drug targets. Among them Gnt1 disrupted parasites have shown highest growth defect in comparison to other glycosyltransferases (chapter 2), which could be targeted for potential inhibitor development. Rational design of glycosyltransferases still remains a difficult task due to intrinsic features of glycosyltransferases such as complex four-partner transition state (sugar donor, acceptor, metal, nucleotide), limited structural data and weak binding of the enzyme with their natural substrates (usual K_m values are in the mM range). Evidence suggests, however, that Skp1 glycosyltransferases have atypically high affinities toward their substrates. Other groups have developed approaches to identify small molecule inhibitor scaffolds for glycosyltransferases, including O-GlcNAc transferase, which could be applied against the TgGnt1 or other glycosyltransferases of the pathway (188-191). In addition, John Kim

in our lab has solved the structure of Gat1 (unpublished data), an approach that could inform complementary structure based approaches for rational designing of inhibitors against this glycosyltransferase.

References

1. Blader IJ, Coleman BI, Chen CT, Gubbels MJ. Lytic Cycle of Toxoplasma gondii: 15 Years Later. Annu Rev Microbiol. 2015;69:463-85.

2. Montoya JG, Liesenfeld O. Toxoplasmosis. Lancet. 2004;363(9425):1965-76.

3. Baatz H, Mirshahi A, Puchta J, Gumbel H, Hattenbach LO. Reactivation of toxoplasma retinochoroiditis under atovaquone therapy in an immunocompetent patient. Ocul Immunol Inflamm. 2006;14(3):185-7.

4. Aspinall TV, Joynson DH, Guy E, Hyde JE, Sims PF. The molecular basis of sulfonamide resistance in Toxoplasma gondii and implications for the clinical management of toxoplasmosis. J Infect Dis. 2002;185(11):1637-43.

5. Dannemann B, McCutchan JA, Israelski D, Antoniskis D, Leport C, Luft B, Nussbaum J, Clumeck N, Morlat P, Chiu J, et al. Treatment of toxoplasmic encephalitis in patients with AIDS. A randomized trial comparing pyrimethamine plus clindamycin to pyrimethamine plus sulfadiazine. The California Collaborative Treatment Group. Ann Intern Med. 1992;116(1):33-43.

6. Cardozo T, Pagano M. The SCF ubiquitin ligase: insights into a molecular machine. Nat Rev Mol Cell Biol. 2004;5(9):739-51.

7. Spruck CH, Strohmaier HM. Seek and destroy: SCF ubiquitin ligases in mammalian cell cycle control. Cell Cycle. 2002;1(4):250-4.

8. Santner A, Estelle M. The ubiquitin-proteasome system regulates plant hormone signaling. Plant J. 2010;61(6):1029-40.

9. Petroski MD, Deshaies RJ. Function and regulation of cullin-RING ubiquitin ligases. Nat Rev Mol Cell Biol. 2005;6(1):9-20.

10. West CM, Wang ZA, van der Wel H. A cytoplasmic prolyl hydroxylation and glycosylation pathway modifies Skp1 and regulates O2-dependent development in Dictyostelium. Biochim Biophys Acta. 2010;1800(2):160-71.

11. Sheikh MO, Schafer CM, Powell JT, Rodgers KK, Mooers BH, West CM. Glycosylation of Skp1 affects its conformation and promotes binding to a model f-box protein. Biochemistry. 2014;53(10):1657-69.

12. Sheikh MO, Xu Y, van der Wel H, Walden P, Hartson SD, West CM. Glycosylation of Skp1 promotes formation of Skp1-cullin-1-F-box protein complexes in dictyostelium. Mol Cell Proteomics. 2015;14(1):66-80.

13. Paugam A, Creuzet C, Dupouy-Camet J, Roisin MP. Evidence for the existence of a proteasome in Toxoplasma gondii: intracellular localization and specific peptidase activities. Parasite. 2001;8(4):267-73.

14. Ponts N, Yang J, Chung DW, Prudhomme J, Girke T, Horrocks P, Le Roch KG. Deciphering the ubiquitin-mediated pathway in apicomplexan parasites: a potential strategy to interfere with parasite virulence. PLoS One. 2008;3(6):e2386.
15. Silmon de Monerri NC, Yakubu RR, Chen AL, Bradley PJ, Nieves E, Weiss LM, Kim K. The Ubiquitin Proteome of Toxoplasma gondii Reveals Roles for Protein Ubiquitination in Cell-Cycle Transitions. Cell Host Microbe. 2015;18(5):621-33.

16. Xu Y, Brown KM, Wang ZA, van der Wel H, Teygong C, Zhang D, Blader IJ, West CM. The Skp1 protein from Toxoplasma is modified by a cytoplasmic prolyl 4-hydroxylase associated with oxygen sensing in the social amoeba Dictyostelium. J Biol Chem. 2012;287(30):25098-110.

17. Weiss LM, Dubey JP. Toxoplasmosis: A history of clinical observations. Int J Parasitol. 2009;39(8):895-901.

18. Dubey JP, Miller NL, Frenkel JK. Characterization of the new fecal form of Toxoplasma gondii. J Parasitol. 1970;56(3):447-56.

19. Portman N, Slapeta J. The flagellar contribution to the apical complex: a new tool for the eukaryotic Swiss Army knife? Trends Parasitol. 2014;30(2):58-64.

20. Lim L, McFadden GI. The evolution, metabolism and functions of the apicoplast. Philos Trans R Soc Lond B Biol Sci. 2010;365(1541):749-63.

21. Morrissette NS, Sibley LD. Cytoskeleton of apicomplexan parasites. Microbiol Mol Biol Rev. 2002;66(1):21-38; table of contents.

22. Dubey JP, Hill DE, Lindsay DS, Jenkins MC, Uggla A, Speer CA. Neospora caninum and Hammondia heydorni are separate species/organisms. Trends Parasitol. 2002;18(2):66-9.

23. Kopecna J, Jirku M, Obornik M, Tokarev YS, Lukes J, Modry D. Phylogenetic analysis of coccidian parasites from invertebrates: search for missing links. Protist. 2006;157(2):173-83.

24. Francia ME, Striepen B. Cell division in apicomplexan parasites. Nat Rev Microbiol. 2014;12(2):125-36.

25. Dubey JP, Lindsay DS, Speer CA. Structures of Toxoplasma gondii tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts. Clin Microbiol Rev. 1998;11(2):267-99.

26. Dubey JP. Toxoplasmosis - a waterborne zoonosis. Vet Parasitol. 2004;126(1-2):57-72.

27. Muller J, Hemphill A. In vitro culture systems for the study of apicomplexan parasites in farm animals. Int J Parasitol. 2013;43(2):115-24.

28. Moura Mde A, Amendoeira MR, Barbosa HS. Primary culture of intestinal epithelial cells as a potential model for Toxoplasma gondii enteric cycle studies. Mem Inst Oswaldo Cruz. 2009;104(6):862-4.

29. Scheidegger A, Vonlaufen N, Naguleswaran A, Gianinazzi C, Muller N, Leib SL, Hemphill A. Differential effects of interferon-gamma and tumor necrosis factor-alpha on Toxoplasma gondii proliferation in organotypic rat brain slice cultures. J Parasitol. 2005;91(2):307-15.

30. Ferreira da Silva Mda F, Barbosa HS, Gross U, Luder CG. Stress-related and spontaneous stage differentiation of Toxoplasma gondii. Mol Biosyst. 2008;4(8):824-34.

31. Luder CG, Giraldo-Velasquez M, Sendtner M, Gross U. Toxoplasma gondii in primary rat CNS cells: differential contribution of neurons, astrocytes, and microglial cells for the intracerebral development and stage differentiation. Exp Parasitol. 1999;93(1):23-32.

32. Ferreira-da-Silva Mda F, Takacs AC, Barbosa HS, Gross U, Luder CG. Primary skeletal muscle cells trigger spontaneous Toxoplasma gondii tachyzoite-to-bradyzoite conversion at higher rates than fibroblasts. Int J Med Microbiol. 2009;299(5):381-8.

33. Ferreira-da-Silva Mda F, Rodrigues RM, Andrade EF, Carvalho L, Gross U, Luder CG, Barbosa HS. Spontaneous stage differentiation of mouse-virulent Toxoplasma gondii RH parasites in skeletal muscle cells: an ultrastructural evaluation. Mem Inst Oswaldo Cruz. 2009;104(2):196-200.

34. Skariah S, McIntyre MK, Mordue DG. Toxoplasma gondii: determinants of tachyzoite to bradyzoite conversion. Parasitol Res. 2010;107(2):253-60.

35. Carruthers VB, Sibley LD. Sequential protein secretion from three distinct organelles of Toxoplasma gondii accompanies invasion of human fibroblasts. Eur J Cell Biol. 1997;73(2):114-23.

36. Dobrowolski JM, Sibley LD. Toxoplasma invasion of mammalian cells is powered by the actin cytoskeleton of the parasite. Cell. 1996;84(6):933-9.

37. Carruthers V, Boothroyd JC. Pulling together: an integrated model of Toxoplasma cell invasion. Curr Opin Microbiol. 2007;10(1):83-9.

38. Charron AJ, Sibley LD. Molecular partitioning during host cell penetration by Toxoplasma gondii. Traffic. 2004;5(11):855-67.

39. Coppens I, Dunn JD, Romano JD, Pypaert M, Zhang H, Boothroyd JC, Joiner KA. Toxoplasma gondii sequesters lysosomes from mammalian hosts in the vacuolar space. Cell. 2006;125(2):261-74.

40. Coppens I, Sinai AP, Joiner KA. Toxoplasma gondii exploits host low-density lipoprotein receptor-mediated endocytosis for cholesterol acquisition. J Cell Biol. 2000;149(1):167-80.

41. Sibley LD, Niesman IR, Parmley SF, Cesbron-Delauw MF. Regulated secretion of multi-lamellar vesicles leads to formation of a tubulo-vesicular network in host-cell vacuoles occupied by Toxoplasma gondii. J Cell Sci. 1995;108 (Pt 4):1669-77.

42. Sinai AP, Joiner KA. The Toxoplasma gondii protein ROP2 mediates host organelle association with the parasitophorous vacuole membrane. J Cell Biol. 2001;154(1):95-108.

43. Melo EJ, Carvalho TM, De Souza W. Behaviour of microtubules in cells infected with Toxoplasma gondii. Biocell. 2001;25(1):53-9.

44. Sinai AP, Webster P, Joiner KA. Association of host cell endoplasmic reticulum and mitochondria with the Toxoplasma gondii parasitophorous vacuole membrane: a high affinity interaction. J Cell Sci. 1997;110 (Pt 17):2117-28.

45. Chaudhary K, Darling JA, Fohl LM, Sullivan WJ, Jr., Donald RG, Pfefferkorn ER, Ullman B, Roos DS. Purine salvage pathways in the apicomplexan parasite Toxoplasma gondii. J Biol Chem. 2004;279(30):31221-7.

46. Hoff EF, Carruthers VB. Is Toxoplasma egress the first step in invasion? Trends Parasitol. 2002;18(6):251-5.

47. Lyons RE, McLeod R, Roberts CW. Toxoplasma gondii tachyzoite-bradyzoite interconversion. Trends Parasitol. 2002;18(5):198-201.

48. Hershko A, Ciechanover A. The ubiquitin system. Annu Rev Biochem. 1998;67:425-79.

49. Komander D, Rape M. The ubiquitin code. Annu Rev Biochem. 2012;81:203-29.

50. Pickart CM. Mechanisms underlying ubiquitination. Annu Rev Biochem. 2001;70:503-33.

51. Sarikas A, Hartmann T, Pan ZQ. The cullin protein family. Genome Biol. 2011;12(4):220.

52. Zheng N, Schulman BA, Song L, Miller JJ, Jeffrey PD, Wang P, Chu C, Koepp DM, Elledge SJ, Pagano M, Conaway RC, Conaway JW, Harper JW, Pavletich NP. Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. Nature. 2002;416(6882):703-9.

53. Xu G, Ma H, Nei M, Kong H. Evolution of F-box genes in plants: different modes of sequence divergence and their relationships with functional diversification. Proc Natl Acad Sci U S A. 2009;106(3):835-40.

54. Lydeard JR, Schulman BA, Harper JW. Building and remodelling Cullin-RING E3 ubiquitin ligases. EMBO Rep. 2013;14(12):1050-61.

55. Duda DM, Borg LA, Scott DC, Hunt HW, Hammel M, Schulman BA. Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. Cell. 2008;134(6):995-1006.

56. Saha A, Deshaies RJ. Multimodal activation of the ubiquitin ligase SCF by Nedd8 conjugation. Mol Cell. 2008;32(1):21-31.

57. Handeli S, Weintraub H. The ts41 mutation in Chinese hamster cells leads to successive S phases in the absence of intervening G2, M, and G1. Cell. 1992;71(4):599-611.

58. Jones D, Crowe E, Stevens TA, Candido EP. Functional and phylogenetic analysis of the ubiquitylation system in Caenorhabditis elegans: ubiquitin-conjugating enzymes, ubiquitin-activating enzymes, and ubiquitin-like proteins. Genome Biol. 2002;3(1):RESEARCH0002.

59. Bostick M, Lochhead SR, Honda A, Palmer S, Callis J. Related to ubiquitin 1 and 2 are redundant and essential and regulate vegetative growth, auxin signaling, and ethylene production in Arabidopsis. Plant Cell. 2004;16(9):2418-32.

60. Osaka F, Saeki M, Katayama S, Aida N, Toh EA, Kominami K, Toda T, Suzuki T, Chiba T, Tanaka K, Kato S. Covalent modifier NEDD8 is essential for SCF ubiquitin-ligase in fission yeast. EMBO J. 2000;19(13):3475-84.

61. Lammer D, Mathias N, Laplaza JM, Jiang W, Liu Y, Callis J, Goebl M, Estelle M. Modification of yeast Cdc53p by the ubiquitin-related protein rub1p affects function of the SCFCdc4 complex. Genes Dev. 1998;12(7):914-26.

62. Liakopoulos D, Doenges G, Matuschewski K, Jentsch S. A novel protein modification pathway related to the ubiquitin system. EMBO J. 1998;17(8):2208-14.

63. Kurz T, Ozlu N, Rudolf F, O'Rourke SM, Luke B, Hofmann K, Hyman AA, Bowerman B, Peter M. The conserved protein DCN-1/Dcn1p is required for cullin neddylation in C. elegans and S. cerevisiae. Nature. 2005;435(7046):1257-61.

64. Schwechheimer C, Serino G, Callis J, Crosby WL, Lyapina S, Deshaies RJ, Gray WM, Estelle M, Deng XW. Interactions of the COP9 signalosome with the E3 ubiquitin ligase SCFTIRI in mediating auxin response. Science. 2001;292(5520):1379-82.

65. Cavadini S, Fischer ES, Bunker RD, Potenza A, Lingaraju GM, Goldie KN, Mohamed WI, Faty M, Petzold G, Beckwith RE, Tichkule RB, Hassiepen U, Abdulrahman W, Pantelic RS, Matsumoto S, Sugasawa K, Stahlberg H, Thoma NH. Cullin-RING ubiquitin E3 ligase regulation by the COP9 signalosome. Nature. 2016;531(7596):598-603.

66. Fischer ES, Scrima A, Bohm K, Matsumoto S, Lingaraju GM, Faty M, Yasuda T, Cavadini S, Wakasugi M, Hanaoka F, Iwai S, Gut H, Sugasawa K, Thoma NH. The molecular basis of CRL4DDB2/CSA ubiquitin ligase architecture, targeting, and activation. Cell. 2011;147(5):1024-39.

67. Emberley ED, Mosadeghi R, Deshaies RJ. Deconjugation of Nedd8 from Cul1 is directly regulated by Skp1-F-box and substrate, and the COP9 signalosome inhibits deneddylated SCF by a noncatalytic mechanism. J Biol Chem. 2012;287(35):29679-89.

68. Enchev RI, Scott DC, da Fonseca PC, Schreiber A, Monda JK, Schulman BA, Peter M, Morris EP. Structural basis for a reciprocal regulation between SCF and CSN. Cell Rep. 2012;2(3):616-27.

69. Groisman R, Polanowska J, Kuraoka I, Sawada J, Saijo M, Drapkin R, Kisselev AF, Tanaka K, Nakatani Y. The ubiquitin ligase activity in the DDB2 and CSA complexes is

differentially regulated by the COP9 signalosome in response to DNA damage. Cell. 2003;113(3):357-67.

70. Fuzesi-Levi MG, Ben-Nissan G, Bianchi E, Zhou H, Deery MJ, Lilley KS, Levin Y, Sharon M. Dynamic regulation of the COP9 signalosome in response to DNA damage. Mol Cell Biol. 2014;34(6):1066-76.

71. Bornstein G, Ganoth D, Hershko A. Regulation of neddylation and deneddylation of cullin1 in SCFSkp2 ubiquitin ligase by F-box protein and substrate. Proc Natl Acad Sci U S A. 2006;103(31):11515-20.

72. Liu J, Furukawa M, Matsumoto T, Xiong Y. NEDD8 modification of CUL1 dissociates p120(CAND1), an inhibitor of CUL1-SKP1 binding and SCF ligases. Mol Cell. 2002;10(6):1511-8.

73. Goldenberg SJ, Cascio TC, Shumway SD, Garbutt KC, Liu J, Xiong Y, Zheng N. Structure of the Cand1-Cul1-Roc1 complex reveals regulatory mechanisms for the assembly of the multisubunit cullin-dependent ubiquitin ligases. Cell. 2004;119(4):517-28.

74. Deshaies RJ, Joazeiro CA. RING domain E3 ubiquitin ligases. Annu Rev Biochem. 2009;78:399-434.

75. Cheng Y, Dai X, Zhao Y. AtCAND1, a HEAT-repeat protein that participates in auxin signaling in Arabidopsis. Plant Physiol. 2004;135(2):1020-6.

76. Bosu DR, Feng H, Min K, Kim Y, Wallenfang MR, Kipreos ET. C. elegans CAND-1 regulates cullin neddylation, cell proliferation and morphogenesis in specific tissues. Dev Biol. 2010;346(1):113-26.

77. Sela N, Atir-Lande A, Kornitzer D. Neddylation and CAND1 independently stimulate SCF ubiquitin ligase activity in Candida albicans. Eukaryot Cell. 2012;11(1):42-52.

78. Wu JT, Lin HC, Hu YC, Chien CT. Neddylation and deneddylation regulate Cul1 and Cul3 protein accumulation. Nat Cell Biol. 2005;7(10):1014-20.

79. Cope GA, Deshaies RJ. Targeted silencing of Jab1/Csn5 in human cells downregulates SCF activity through reduction of F-box protein levels. BMC Biochem. 2006;7:1.

80. He Q, Cheng P, He Q, Liu Y. The COP9 signalosome regulates the Neurospora circadian clock by controlling the stability of the SCFFWD-1 complex. Genes Dev. 2005;19(13):1518-31.

81. Wee S, Geyer RK, Toda T, Wolf DA. CSN facilitates Cullin-RING ubiquitin ligase function by counteracting autocatalytic adapter instability. Nat Cell Biol. 2005;7(4):387-91.

82. Bosu DR, Kipreos ET. Cullin-RING ubiquitin ligases: global regulation and activation cycles. Cell Div. 2008;3:7.

83. Schmidt MW, McQuary PR, Wee S, Hofmann K, Wolf DA. F-box-directed CRL complex assembly and regulation by the CSN and CAND1. Mol Cell. 2009;35(5):586-97.

84. Chew EH, Poobalasingam T, Hawkey CJ, Hagen T. Characterization of cullin-based E3 ubiquitin ligases in intact mammalian cells--evidence for cullin dimerization. Cell Signal. 2007;19(5):1071-80.

85. Tang X, Orlicky S, Lin Z, Willems A, Neculai D, Ceccarelli D, Mercurio F, Shilton BH, Sicheri F, Tyers M. Suprafacial orientation of the SCFCdc4 dimer accommodates multiple geometries for substrate ubiquitination. Cell. 2007;129(6):1165-76.

86. Suzuki H, Chiba T, Suzuki T, Fujita T, Ikenoue T, Omata M, Furuichi K, Shikama H, Tanaka K. Homodimer of two F-box proteins betaTrCP1 or betaTrCP2 binds to IkappaBalpha for signal-dependent ubiquitination. J Biol Chem. 2000;275(4):2877-84.

87. Dixon C, Brunson LE, Roy MM, Smothers D, Sehorn MG, Mathias N. Overproduction of polypeptides corresponding to the amino terminus of the F-box proteins Cdc4p and Met30p inhibits ubiquitin ligase activities of their SCF complexes. Eukaryot Cell. 2003;2(1):123-33.

88. Kominami K, Ochotorena I, Toda T. Two F-box/WD-repeat proteins Pop1 and Pop2 form hetero- and homo-complexes together with cullin-1 in the fission yeast SCF (Skp1-Cullin-1-F-box) ubiquitin ligase. Genes Cells. 1998;3(11):721-35.

89. Welcker M, Clurman BE. Fbw7/hCDC4 dimerization regulates its substrate interactions. Cell Div. 2007;2:7.

90. Zhang W, Koepp DM. Fbw7 isoform interaction contributes to cyclin E proteolysis. Mol Cancer Res. 2006;4(12):935-43.

91. Barbash O, Zamfirova P, Lin DI, Chen X, Yang K, Nakagawa H, Lu F, Rustgi AK, Diehl JA. Mutations in Fbx4 inhibit dimerization of the SCF(Fbx4) ligase and contribute to cyclin D1 overexpression in human cancer. Cancer Cell. 2008;14(1):68-78.

92. Zhou P, Howley PM. Ubiquitination and degradation of the substrate recognition subunits of SCF ubiquitin-protein ligases. Mol Cell. 1998;2(5):571-80.

93. Galan JM, Peter M. Ubiquitin-dependent degradation of multiple F-box proteins by an autocatalytic mechanism. Proc Natl Acad Sci U S A. 1999;96(16):9124-9.

94. Smothers DB, Kozubowski L, Dixon C, Goebl MG, Mathias N. The abundance of Met30p limits SCF(Met30p) complex activity and is regulated by methionine availability. Mol Cell Biol. 2000;20(21):7845-52.

95. Wirbelauer C, Sutterluty H, Blondel M, Gstaiger M, Peter M, Reymond F, Krek W. The F-box protein Skp2 is a ubiquitylation target of a Cul1-based core ubiquitin ligase complex: evidence for a role of Cul1 in the suppression of Skp2 expression in quiescent fibroblasts. EMBO J. 2000;19(20):5362-75.

96. Li Y, Gazdoiu S, Pan ZQ, Fuchs SY. Stability of homologue of Slimb F-box protein is regulated by availability of its substrate. J Biol Chem. 2004;279(12):11074-80.

97. Winston JT, Chu C, Harper JW. Culprits in the degradation of cyclin E apprehended. Genes Dev. 1999;13(21):2751-7.

98. Nash P, Tang X, Orlicky S, Chen Q, Gertler FB, Mendenhall MD, Sicheri F, Pawson T, Tyers M. Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. Nature. 2001;414(6863):514-21.

99. Coccetti P, Tripodi F, Tedeschi G, Nonnis S, Marin O, Fantinato S, Cirulli C, Vanoni M, Alberghina L. The CK2 phosphorylation of catalytic domain of Cdc34 modulates its activity at the G1 to S transition in Saccharomyces cerevisiae. Cell Cycle. 2008;7(10):1391-401.

100. Sadowski M, Mawson A, Baker R, Sarcevic B. Cdc34 C-terminal tail phosphorylation regulates Skp1/cullin/F-box (SCF)-mediated ubiquitination and cell cycle progression. Biochem J. 2007;405(3):569-81.

101. Mizushima T, Yoshida Y, Kumanomidou T, Hasegawa Y, Suzuki A, Yamane T, Tanaka K. Structural basis for the selection of glycosylated substrates by SCF(Fbs1) ubiquitin ligase. Proc Natl Acad Sci U S A. 2007;104(14):5777-81.

102. Kozarov E, van der Wel H, Field M, Gritzali M, Brown RD, Jr., West CM. Characterization of FP21, a cytosolic glycoprotein from Dictyostelium. J Biol Chem. 1995;270(7):3022-30.

103. West CM, Van Der Wel H, Sassi S, Gaucher EA. Cytoplasmic glycosylation of protein-hydroxyproline and its relationship to other glycosylation pathways. Biochim Biophys Acta. 2004;1673(1-2):29-44.

104. van der Wel H, Ercan A, West CM. The Skp1 prolyl hydroxylase from Dictyostelium is related to the hypoxia-inducible factor-alpha class of animal prolyl 4-hydroxylases. J Biol Chem. 2005;280(15):14645-55.

105. Kaelin WG, Jr., Ratcliffe PJ. Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. Mol Cell. 2008;30(4):393-402.

106. Teng-Umnuay P, van der Wel H, West CM. Identification of a UDP-GlcNAc:Skp1-hydroxyproline GlcNAc-transferase in the cytoplasm of Dictyostelium. J Biol Chem. 1999;274(51):36392-402.

107. Van Der Wel H, Morris HR, Panico M, Paxton T, Dell A, Kaplan L, West CM. Molecular cloning and expression of a UDP-N-acetylglucosamine (GlcNAc):hydroxyproline polypeptide GlcNAc-transferase that modifies Skp1 in the cytoplasm of dictyostelium. J Biol Chem. 2002;277(48):46328-37.

108. van Der Wel H, Morris HR, Panico M, Paxton T, North SJ, Dell A, Thomson JM, West CM. A non-Golgi alpha 1,2-fucosyltransferase that modifies Skp1 in the cytoplasm of Dictyostelium. J Biol Chem. 2001;276(36):33952-63.

109. Van Der Wel H, Fisher SZ, West CM. A bifunctional diglycosyltransferase forms the Fucalpha1,2Galbeta1,3-disaccharide on Skp1 in the cytoplasm of dictyostelium. J Biol Chem. 2002;277(48):46527-34.

110. West CM, Scott-Ward T, Teng-umnuay P, van der Wel H, Kozarov E, Huynh A. Purification and characterization of an alpha1,2,-L-fucosyltransferase, which modifies the cytosolic protein FP21,from the cytosol of Dictyostelium. J Biol Chem. 1996;271(20):12024-35.

111. Ercan A, Panico M, Sutton-Smith M, Dell A, Morris HR, Matta KL, Gay DF, West CM. Molecular characterization of a novel UDP-galactose:fucoside alpha3-galactosyltransferase that modifies Skp1 in the cytoplasm of Dictyostelium. J Biol Chem. 2006;281(18):12713-21.

112. Ketcham C, Wang F, Fisher SZ, Ercan A, van der Wel H, Locke RD, Sirajud-Doulah K, Matta KL, West CM. Specificity of a soluble UDP-galactose: fucoside alpha1,3-galactosyltransferase that modifies the cytoplasmic glycoprotein Skp1 in Dictyostelium. J Biol Chem. 2004;279(28):29050-9.

113. Schafer CM, Sheikh MO, Zhang D, West CM. Novel regulation of Skp1 by the Dictyostelium AgtA alpha-galactosyltransferase involves the Skp1-binding activity of its WD40 repeat domain. J Biol Chem. 2014;289(13):9076-88.

114. Sassi S, Sweetinburgh M, Erogul J, Zhang P, Teng-Umnuay P, West CM. Analysis of Skp1 glycosylation and nuclear enrichment in Dictyostelium. Glycobiology. 2001;11(4):283-95.

115. Kaplan KB, Hyman AA, Sorger PK. Regulating the yeast kinetochore by ubiquitin-dependent degradation and Skp1p-mediated phosphorylation. Cell. 1997;91(4):491-500.

116. Beltrao P, Albanese V, Kenner LR, Swaney DL, Burlingame A, Villen J, Lim WA, Fraser JS, Frydman J, Krogan NJ. Systematic functional prioritization of protein posttranslational modifications. Cell. 2012;150(2):413-25.

117. Soucy TA, Smith PG, Milhollen MA, Berger AJ, Gavin JM, Adhikari S, Brownell JE, Burke KE, Cardin DP, Critchley S, Cullis CA, Doucette A, Garnsey JJ, Gaulin JL, Gershman RE, Lublinsky AR, McDonald A, Mizutani H, Narayanan U, Olhava EJ, Peluso S, Rezaei M, Sintchak MD, Talreja T, Thomas MP, Traore T, Vyskocil S, Weatherhead GS, Yu J, Zhang J, Dick LR, Claiborne CF, Rolfe M, Bolen JB, Langston SP. An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. Nature. 2009;458(7239):732-6.

118. Marin I. Diversification of the cullin family. BMC Evol Biol. 2009;9:267.

119. Stone SL, Hauksdottir H, Troy A, Herschleb J, Kraft E, Callis J. Functional analysis of the RING-type ubiquitin ligase family of Arabidopsis. Plant Physiol. 2005;137(1):13-30.

120. Sadanandom A, Bailey M, Ewan R, Lee J, Nelis S. The ubiquitin-proteasome system: central modifier of plant signalling. New Phytol. 2012;196(1):13-28.

121. Jin J, Cardozo T, Lovering RC, Elledge SJ, Pagano M, Harper JW. Systematic analysis and nomenclature of mammalian F-box proteins. Genes Dev. 2004;18(21):2573-80.

122. Teixeira LK, Reed SI. Ubiquitin ligases and cell cycle control. Annu Rev Biochem. 2013;82:387-414.

123. Frescas D, Pagano M. Deregulated proteolysis by the F-box proteins SKP2 and beta-TrCP: tipping the scales of cancer. Nat Rev Cancer. 2008;8(6):438-49.

124. Welcker M, Clurman BE. FBW7 ubiquitin ligase: a tumour suppressor at the crossroads of cell division, growth and differentiation. Nat Rev Cancer. 2008;8(2):83-93.

125. Crusio KM, King B, Reavie LB, Aifantis I. The ubiquitous nature of cancer: the role of the SCF(Fbw7) complex in development and transformation. Oncogene. 2010;29(35):4865-73.

126. Nakayama KI, Nakayama K. Ubiquitin ligases: cell-cycle control and cancer. Nat Rev Cancer. 2006;6(5):369-81.

127. Dharmasiri N, Dharmasiri S, Weijers D, Lechner E, Yamada M, Hobbie L, Ehrismann JS, Jurgens G, Estelle M. Plant development is regulated by a family of auxin receptor F box proteins. Dev Cell. 2005;9(1):109-19.

128. Greenham K, Santner A, Castillejo C, Mooney S, Sairanen I, Ljung K, Estelle M. The AFB4 auxin receptor is a negative regulator of auxin signaling in seedlings. Curr Biol. 2011;21(6):520-5.

129. Sheard LB, Tan X, Mao H, Withers J, Ben-Nissan G, Hinds TR, Kobayashi Y, Hsu FF, Sharon M, Browse J, He SY, Rizo J, Howe GA, Zheng N. Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor. Nature. 2010;468(7322):400-5.

130. Xu Y, Wang ZA, Green RS, West CM. Role of the Skp1 prolyl-hydroxylation/glycosylation pathway in oxygen dependent submerged development of Dictyostelium. BMC Dev Biol. 2012;12:31.

131. West CM, van der Wel H, Wang ZA. Prolyl 4-hydroxylase-1 mediates O2 signaling during development of Dictyostelium. Development. 2007;134(18):3349-58.

132. Nishi M, Hu K, Murray JM, Roos DS. Organellar dynamics during the cell cycle of Toxoplasma gondii. J Cell Sci. 2008;121(Pt 9):1559-68.

133. Sinai AP, Roepe PD. Autophagy in Apicomplexa: a life sustaining death mechanism? Trends Parasitol. 2012;28(9):358-64.

134. Gonzalez J, Ramalho-Pinto FJ, Frevert U, Ghiso J, Tomlinson S, Scharfstein J, Corey EJ, Nussenzweig V. Proteasome activity is required for the stage-specific transformation of a protozoan parasite. J Exp Med. 1996;184(5):1909-18.

135. Robertson CD. The Leishmania mexicana proteasome. Mol Biochem Parasitol. 1999;103(1):49-60.

136. Gantt SM, Myung JM, Briones MR, Li WD, Corey EJ, Omura S, Nussenzweig V, Sinnis P. Proteasome inhibitors block development of Plasmodium spp. Antimicrob Agents Chemother. 1998;42(10):2731-8.

137. Agrawal S, Chung DW, Ponts N, van Dooren GG, Prudhomme J, Brooks CF, Rodrigues EM, Tan JC, Ferdig MT, Striepen B, Le Roch KG. An apicoplast localized ubiquitylation system is required for the import of nuclear-encoded plastid proteins. PLoS Pathog. 2013;9(6):e1003426.

138. Shaw MK, He CY, Roos DS, Tilney LG. Proteasome inhibitors block intracellular growth and replication of Toxoplasma gondii. Parasitology. 2000;121 (Pt 1):35-47.

139. Sidik SM, Huet D, Ganesan SM, Huynh MH, Wang T, Nasamu AS, Thiru P, Saeij JP, Carruthers VB, Niles JC, Lourido S. A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes. Cell. 2016;166(6):1423-35 e12.

140. Dhara A, Sinai AP. A Cell Cycle-Regulated Toxoplasma Deubiquitinase, TgOTUD3A, Targets Polyubiquitins with Specific Lysine Linkages. mSphere. 2016;1(3).

141. West CM, van der Wel H, Blader IJ. Detection of cytoplasmic glycosylation associated with hydroxyproline. Methods Enzymol. 2006;417:389-404.

142. Kim K, Weiss LM. Toxoplasma gondii: the model apicomplexan. Int J Parasitol. 2004;34(3):423-32.

143. Pereira-Chioccola VL, Vidal JE, Su C. Toxoplasma gondii infection and cerebral toxoplasmosis in HIV-infected patients. Future Microbiol. 2009;4(10):1363-79.

144. Luft BJ, Remington JS. Toxoplasmic encephalitis in AIDS. Clin Infect Dis. 1992;15(2):211-22.

145. Lambert H, Barragan A. Modelling parasite dissemination: host cell subversion and immune evasion by Toxoplasma gondii. Cell Microbiol. 2010;12(3):292-300.

146. Zhang D, van der Wel H, Johnson JM, West CM. Skp1 prolyl 4-hydroxylase of dictyostelium mediates glycosylation-independent and -dependent responses to O2 without affecting Skp1 stability. J Biol Chem. 2012;287(3):2006-16.

147. West CM, Blader IJ. Oxygen sensing by protozoans: how they catch their breath. Curr Opin Microbiol. 2015;26:41-7.

148. Vinayak S, Brooks CF, Naumov A, Suvorova ES, White MW, Striepen B. Genetic manipulation of the Toxoplasma gondii genome by fosmid recombineering. MBio. 2014;5(6):e02021.

149. Huynh MH, Carruthers VB. Tagging of endogenous genes in a Toxoplasma gondii strain lacking Ku80. Eukaryot Cell. 2009;8(4):530-9.

150. Fox BA, Ristuccia JG, Gigley JP, Bzik DJ. Efficient gene replacements in Toxoplasma gondii strains deficient for nonhomologous end joining. Eukaryot Cell. 2009;8(4):520-9.

151. Figueras MJ, Martin OA, Echeverria PC, de Miguel N, Naguleswaran A, Sullivan WJ, Jr., Corvi MM, Angel SO. Toxoplasma gondii Sis1-like J-domain protein is a cytosolic chaperone associated to HSP90/HSP70 complex. Int J Biol Macromol. 2012;50(3):725-33.

152. Kimmel J, Smith TK, Azzouz N, Gerold P, Seeber F, Lingelbach K, Dubremetz JF, Schwarz RT. Membrane topology and transient acylation of Toxoplasma gondii glycosylphosphatidylinositols. Eukaryot Cell. 2006;5(8):1420-9.

153. Narasimhan J, Joyce BR, Naguleswaran A, Smith AT, Livingston MR, Dixon SE, Coppens I, Wek RC, Sullivan WJ, Jr. Translation regulation by eukaryotic initiation factor-2 kinases in the development of latent cysts in Toxoplasma gondii. J Biol Chem. 2008;283(24):16591-601.

154. Moore RB, Obornik M, Janouskovec J, Chrudimsky T, Vancova M, Green DH, Wright SW, Davies NW, Bolch CJ, Heimann K, Slapeta J, Hoegh-Guldberg O, Logsdon JM, Carter DA. A photosynthetic alveolate closely related to apicomplexan parasites. Nature. 2008;451(7181):959-63.

155. Wojczyk BS, Stwora-Wojczyk MM, Hagen FK, Striepen B, Hang HC, Bertozzi CR, Roos DS, Spitalnik SL. cDNA cloning and expression of UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase T1 from Toxoplasma gondii. Mol Biochem Parasitol. 2003;131(2):93-107.

156. Wang ZA, Singh D, van der Wel H, West CM. Prolyl hydroxylation- and glycosylation-dependent functions of Skp1 in O2-regulated development of Dictyostelium. Dev Biol. 2011;349(2):283-95.

157. Willems AR, Schwab M, Tyers M. A hitchhiker's guide to the cullin ubiquitin ligases: SCF and its kin. Biochim Biophys Acta. 2004;1695(1-3):133-70.

158. Skaar JR, Pagan JK, Pagano M. Mechanisms and function of substrate recruitment by F-box proteins. Nat Rev Mol Cell Biol. 2013;14(6):369-81.

159. Deshaies RJ. SCF and Cullin/Ring H2-based ubiquitin ligases. Annu Rev Cell Dev Biol. 1999;15:435-67.

160. Rahman K, Zhao P, Mandalasi M, van der Wel H, Wells L, Blader IJ, West CM. The E3 Ubiquitin Ligase Adaptor Protein Skp1 Is Glycosylated by an Evolutionarily Conserved Pathway That Regulates Protist Growth and Development. J Biol Chem. 2016;291(9):4268-80.

161. Sun S, Huang H, Qi YB, Qiu M, Dai ZM. Complementary annealing mediated by exonuclease: a method for seamless cloning and conditioning site-directed mutagenesis. Biotechnology, biotechnological equipment. 2015;29(1):105-10.

162. Sheikh MO, Halmo SM, Patel S, Middleton D, Takeuchi H, Schafer CM, West CM, Haltiwanger RS, Avci FY, Moremen KW, Wells L. Rapid screening of sugar-nucleotide donor specificities of putative glycosyltransferases. Glycobiology. 2017;27(3):206-12.

163. Rana SS, Barlow JJ, Matta KL. The chemical synthesis of O-alpha-L-fucopyranosyl-(1 goes to 2)-O-beta-D-galactopyranosyl-(1 goes to 3)-O-[alpha-L-fucopyranosyl-(1 goes to 4)]-2-acetamido-2-deoxy-D-glucopyranose, the Lewis b blood-group antigenic determinant. Carbohydr Res. 1981;96(2):231-9.

164. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33(7):1870-4.

165. Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol. 2008;25(7):1307-20.

166. McCarthy PJ, Snowden CB. The bootstrap and finite population sampling. Vital Health Stat 2. 1985(95):1-23.

167. Wiggins CA, Munro S. Activity of the yeast MNN1 alpha-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. Proc Natl Acad Sci U S A. 1998;95(14):7945-50.

168. Fussy Z, Masarova P, Krucinska J, Esson HJ, Obornik M. Budding of the Alveolate Alga Vitrella brassicaformis Resembles Sexual and Asexual Processes in Apicomplexan Parasites. Protist. 2016;168(1):80-91.

169. Wang ZA, van der Wel H, Vohra Y, Buskas T, Boons GJ, West CM. Role of a cytoplasmic dual-function glycosyltransferase in O2 regulation of development in Dictyostelium. J Biol Chem. 2009;284(42):28896-904.

170. Cheng C, Mu J, Farkas I, Huang D, Goebl MG, Roach PJ. Requirement of the self-glucosylating initiator proteins Glg1p and Glg2p for glycogen accumulation in Saccharomyces cerevisiae. Mol Cell Biol. 1995;15(12):6632-40.

171. Roach PJ, Depaoli-Roach AA, Hurley TD, Tagliabracci VS. Glycogen and its metabolism: some new developments and old themes. Biochem J. 2012;441(3):763-87.

172. Skurat AV, Dietrich AD, Roach PJ. Interaction between glycogenin and glycogen synthase. Arch Biochem Biophys. 2006;456(1):93-7.

173. Ryley JF, Bentley M, Manners DJ, Stark JR. Amylopectin, the storage polysaccharide of the Coccidia Eimeria brunetti and E. tenella. J Parasitol. 1969;55(4):839-45.

174. Sheffield HG, Frenkel JK, Ruiz A. Ultrastructure of the cyst of Sarcocystis muris. J Parasitol. 1977;63(4):629-41.

175. Coppin A, Varre JS, Lienard L, Dauvillee D, Guerardel Y, Soyer-Gobillard MO, Buleon A, Ball S, Tomavo S. Evolution of plant-like crystalline storage polysaccharide in the protozoan parasite Toxoplasma gondii argues for a red alga ancestry. J Mol Evol. 2005;60(2):257-67.

176. Guerardel Y, Leleu D, Coppin A, Lienard L, Slomianny C, Strecker G, Ball S, Tomavo S. Amylopectin biogenesis and characterization in the protozoan parasite Toxoplasma gondii, the intracellular development of which is restricted in the HepG2 cell line. Microbes Infect. 2005;7(1):41-8.

177. Viola R, Nyvall P, Pedersen M. The unique features of starch metabolism in red algae. Proc Biol Sci. 2001;268(1474):1417-22.

178. Dauvillee D, Deschamps P, Ral JP, Plancke C, Putaux JL, Devassine J, Durand-Terrasson A, Devin A, Ball SG. Genetic dissection of floridean starch synthesis in the cytosol of the model dinoflagellate Crypthecodinium cohnii. Proc Natl Acad Sci U S A. 2009;106(50):21126-30.

179. Teng-umnuay P, Morris HR, Dell A, Panico M, Paxton T, West CM. The cytoplasmic F-box binding protein SKP1 contains a novel pentasaccharide linked to hydroxyproline in Dictyostelium. J Biol Chem. 1998;273(29):18242-9.

180. Baldauf SL. The deep roots of eukaryotes. Science. 2003;300(5626):1703-6.

181. Bai C, Sen P, Hofmann K, Ma L, Goebl M, Harper JW, Elledge SJ. SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. Cell. 1996;86(2):263-74.

182. Sullivan WJ, Jr., Jeffers V. Mechanisms of Toxoplasma gondii persistence and latency. FEMS Microbiol Rev. 2012;36(3):717-33.

183. Weiss LM, Kim K. The development and biology of bradyzoites of Toxoplasma gondii. Front Biosci. 2000;5:D391-405.

184. Charles E, Callegan MC, Blader IJ. The SAG1 Toxoplasma gondii surface protein is not required for acute ocular toxoplasmosis in mice. Infect Immun. 2007;75(4):2079-83.

185. Yan L, Colandrea VJ, Hale JJ. Prolyl hydroxylase domain-containing protein inhibitors as stabilizers of hypoxia-inducible factor: small molecule-based therapeutics for anemia. Expert Opin Ther Pat. 2010;20(9):1219-45.

186. Rose NR, McDonough MA, King ON, Kawamura A, Schofield CJ. Inhibition of 2-oxoglutarate dependent oxygenases. Chem Soc Rev. 2011;40(8):4364-97.

187. Ge W, Wolf A, Feng T, Ho CH, Sekirnik R, Zayer A, Granatino N, Cockman ME, Loenarz C, Loik ND, Hardy AP, Claridge TD, Hamed RB, Chowdhury R, Gong L, Robinson CV, Trudgian DC, Jiang M, Mackeen MM, McCullagh JS, Gordiyenko Y, Thalhammer A, Yamamoto A, Yang M, Liu-Yi P, Zhang Z, Schmidt-Zachmann M, Kessler BM, Ratcliffe PJ, Preston GM, Coleman ML, Schofield CJ. Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. Nat Chem Biol. 2012;8(12):960-2.

188. Wang Y, Zhu J, Zhang L. Discovery of Cell-Permeable O-GlcNAc Transferase Inhibitors via Tethering in Situ Click Chemistry. J Med Chem. 2017;60(1):263-72.

189. Gross BJ, Kraybill BC, Walker S. Discovery of O-GlcNAc transferase inhibitors. J Am Chem Soc. 2005;127(42):14588-9.

190. Gross BJ, Swoboda JG, Walker S. A strategy to discover inhibitors of O-linked glycosylation. J Am Chem Soc. 2008;130(2):440-1.

191. Hu Y, Helm JS, Chen L, Ginsberg C, Gross B, Kraybill B, Tiyanont K, Fang X, Wu T, Walker S. Identification of selective inhibitors for the glycosyltransferase MurG via high-throughput screening. Chem Biol. 2004;11(5):703-11.