A glycogenin homolog controls *Toxoplasma gondii* growth via glycosylation of an E3 ubiquitin ligase

Msano Mandalasi[1,2,*], Hyun W. Kim[1,*], Kazi Rahman[1,4], M. Osman Sheikh[3,5], David Thieker[3,6], Elisabet Gas-Pascual[1,2], Peng Zhao[3], Nitin G. Daniel[1], Hanke van der Wel[1], Travis H. Ichikawa[1,7], John N. Glushka[3], Lance Wells[1,3], Zachary A. Wood[1], Christopher M. West[1,2,3,8]

[1]Department of Biochemistry & Molecular Biology, [2]Center for Tropical and Emerging Global Diseases, [3]Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602 USA

[4] Current address: HIV Dynamics and Replication Program, Center for Cancer Research, National Cancer Institute, Frederick, Maryland, 21702

[5] Current address: Amicus Therapeutics, Philadelphia, PA

[6] Current address: Department of Biochemistry and Biophysics, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

[7] Current address: New Materials Institute, University of Georgia, Athens, GA 30602 USA

[8] To whom correspondence should be addressed: 120 E. Green St., Davison Life Sciences A310, Athens, GA 30602. Tel.: 706-542-4259; E-mail: westcm@uga.edu

* These authors contributed equally to this work.

Running title: A glycogenin predecessor glycosylates Skp1 in protists

Key words: crystallography, NMR, molecular dynamics, Skp1, SCF, glycogenin, cytoplasmic glycosylation, *Toxoplasma*, *Pythium*

ABSTRACT

Skp1, a subunit of E3-Skp1/Cullin-1/F-box protein ubiquitin ligases, is uniquely modified in protists by an O_2-dependent prolyl hydroxylase, which forms the attachment site for a novel pentasaccharide. Mutational studies demonstrate the importance of the core glycan for growth of the parasite *Toxoplasma gondii* in fibroblasts, but the significance of the non-reducing terminal sugar is unknown. Here we investigated a homolog of glycogenin, an enzyme that can initiate and prime glycogen synthesis in yeast and animals. Gat1 is required for pentasaccharide assembly in cells and catalyzes the addition of an α-galactose in 3-linkage to the subterminal α3-linked glucose residue *in vitro*. A strong selectivity of Gat1 for Skp1 in extracts is consistent with evidence that Skp1 is the sole target of the glycosyltransferase pathway. *gat1*-disruption resulted in slow growth indicating the importance of the complete glycan. Molecular dynamics simulations suggested that the full glycan helps organize Skp1 as previously described in the amoebozoan *Dictyostelium* where a distinct glycosyltransferase assembles a different terminal disaccharide. The crystal structure of Gat1 from the plant pathogen *Pythium ultimum* confirmed the striking similarity to glycogenin, with differences in the active sites providing an explanation for its distinct substrate preference and regiospecificity. Gat1 also exhibited low α-glucosyltransferase activity like glycogenin, but autoglycosylation was not detected and *gat1*-disruption revealed no effect on starch accumulation in *Toxoplasma*. A phylogenetic analysis suggested that Gat1 was a progenitor of glycogenin, and acquired its role in glycogen formation following the ancestral disappearance of the underlying Skp1 glycosyltransferase Glt1 prior to amoebozoan evolution.

INTRODUCTION

A novel glycosylation pathway whose glycosyltransferases (GTs) reside in the cytoplasm evolved early in eukaryotic evolution and persisted in most branches of protist radiation before disappearing in advanced fungi, land plants, and animals. Studies of the pathway in the amoebozoan *Dictyostelium discoideum* and the apicomplexan *Toxoplasma gondii* show a role in O_2-sensing, owing to the requirement for an O_2-dependent prolyl 4-hydroxylase, PhyA, that modifies a critical prolyl residue of Skp1, a subunit of the SCF class of E3 ubiquitin (Ub) ligases (West and Blader, 2015). PhyA action generates a 4(*trans*)-hydroxyproline (Hyp) residue, which serves as the attachment point of the pentasaccharide that is assembled by the sequential action of five GT activities. Thus Skp1 is a rare example of a glycoprotein that is not only glycosylated in a complex manner in the cytoplasm, but also remains to function in this compartment and the nucleus, rather than being exported (West and Hart, 2017). Skp1 is the only protein that is detectably glycosylated by any of the five GT activities in *D. discoideum* (West and Blader, 2015), which is at a variance from most GTs, which typically modify a range of proteins. Genetic manipulation of Skp1 expression, substitution of the target Pro, and genetic interactions with a Skp1 GT, affect *D. discoideum* O_2-sensing in ways that are consistent with Skp1 being the functional target of prolyl hydroxylation in O_2-sensing (Wang et al., 2011). Hydroxylation/glycosylation of Skp1 promotes its interaction with three different F-box proteins *in vitro* and *in vivo* (Sheikh et al., 2014, 2015), implicating an effect on their respective E3(SCF)Ub ligases. The mechanism by which the Hyp-linked pentasaccharide controls the activity of Skp1 is unclear, but structural studies suggest an effect on Skp1 conformation that is consistent with receptivity to binding FBPs (Sheikh et al., 2017; Xu et al., 2018; West and Kim, 2019).

Studies of the GTs that catalyze assembly of the glycan in *T. gondii* revealed that, while a linear pentasaccharide is assembled at a conserved prolyl residue as in *D. discoideum*, the glycan has a distinct structure (Fig. 1A). The core trisaccharides are evidently identical, and the enzymes that assemble them are orthologous, but the GT that assembles the fourth and fifth sugars in *D. discoideum* is not detected in the *T. gondii* genome. Further studies revealed that the fourth sugar is an α3-linked Glc rather than α3-linked Gal. This distinction led to the discovery of a novel enzyme, Glt1, that catalyzes the formation of this linkage (Rahman et al., 2017). The question of the identity of the fifth sugar, known to be a hexose from MS studies, and its mechanism of addition, remains unanswered. Genomics studies predict the existence of four cytoplasmically localized GTs whose functions are not assigned (Gas-Pascual et al., 2019), and one of these, referred to as Gat1, was predicted to be the missing Skp1 GT on account of the apparent co-distribution of its gene in protists that possess Glt1-like genes (Rahman et al., 2017). However, this gene is highly similar to glycogenin, a dimeric α4-glucosyl transferase that can prime the synthesis of glycogen in the cytoplasm of yeast and animals by a mechanism that involves auto-glucosylation. The cyst-forming stage of *T. gondii* accumulates crystalline amylopectin (Coppin et al., 2003), an α1,4Glc polymer with α1,6-linked branches, in its cytoplasmic compartment. *T. gondii* amylopectin is assembled by a UDP-Glc based metabolism that is related to the floridean starch of the red alga *Cyanidioschyzon merolae* and, to a lesser extent, to those of glycogen storing animals or fungi. Homologs of glycogen synthase (Coppin et al., 2003) and glycogen phosphorylase (Sugi et al., 2017) regulate the accumulation of amylopectin in this parasite, and TgGat1 has been annotated as a glycogenin (Coppin et al., 2005; Sugi et al., 2017). Related genes have been implicated in promoting starch formation in red algae (Pancha et al., 2019a, 2019b).

The identity of the missing Skp1 GT is important for determining the terminal sugar and linkage of the glycan of *T. gondii* Skp1, because recovering enough Skp1 from the parasite for direct determination is not possible. Given that GTs commonly have multiple targets, access to the enzyme is also critical to assessing its acceptor and donor specificity to understand how it contributes to cell regulation. Finally, the complete description of the glycan with its linkages is critical for testing the *cis*-acting structure model for how the glycan controls Skp1 activity as proposed from the *D. discoideum* studies, which in turn bears on the question of the target specificity of the GT from a separate direction. Our studies reported here show that Gat1 is a conserved, cytoplasmic, retaining α3-galactosyl transferase, not an α4-glucosyl transferase as for glycogenin, that modifies the Glc terminus of Glcα1,3Fucα1,2Galβ1,3GlcNAcα1-Skp1 in support

of optimal growth on fibroblast monolayers. We also incorporated into our approach the predicted ortholog from the oomycete plant pathogen *Pythium ultimum* (Pu), the agent for root rot disease in agriculturally important crops (Lévesque et al., 2010; Kamoun et al., 2015). This parasite is related to *Pythium insidiosum*, the agent for debilitating pythiosus in humans and other mammals (Gaastra et al., 2010). We infer that the glycogenin lineage originally evolved as a modifier of Skp1 that contributed to a novel mechanism of O_2-sensing in a range of aerobic unicellular and pathogenic eukaryotes, before changing specificity to become the glycogenin enzymes that emerged during Opisthokont evolution.

RESULTS

The fifth Skp1 sugar is an α-linked Gal and depends on TgGat1

Previous studies described the mechanism of assembly of the first four sugars on TgSkp1-Pro154 (Fig. 1A), but the left the identity of the final sugar (other than its being a hexose) unresolved (Rahman et al., 2016, 2017). The corresponding sugar in *Dictyostelium* (Fig. 1B) is a 3-linked αGal that is susceptible to removal with green coffee bean α-galactosidase. Similar treatment of tryptic peptides from TgSkp1, isolated from the standard type 1 strain RH by immunoprecipitation, resulted in the appearance of the tetrasaccharide form of the glycopeptide at the expense of the pentasaccharide form, indicating that the terminal Hex is an αGal residue (Fig. 1C).

According to a recent genomic analysis (Rahman et al., 2017), TgGat1 is a candidate for the unknown glycosyltransferase that catalyzes the addition of the fifth and final sugar to the linear glycan of Skp1 of *T. gondii* (Fig. 1A). To test the dependence of Skp1 glycosylation on Tg*gat1*, the gene was disrupted using CRISPR/Cas9 in the RH strain, yielding *gat1*Δ as described in Fig. S4A. PCR studies confirmed replacement with an *dhfr* cassette, and enzyme assays (see below) showed a loss of enzyme activity. The resulting clone produced a version of Skp1 in which the 5-sugar glycopeptide was no longer detectable, but ions corresponding to the 4-sugar glycopeptide were detected at similar abundance (Fig. 1C). Similar results were obtained in the type 2 ME49 strain (Fig. S4A, Fig. 1C), and in strain RHΔΔ in which *gat1* was disrupted by homologous recombination using a different selection marker, referred to as Δ*gat1-1* (Table 1; Fig. S3B; not shown). The similar results obtained by different genetic methods in distinct genetic backgrounds indicate that Tg*gat1* is required for addition of the terminal sugar, but whether this was a direct or indirect effect was unclear.

Dependence of parasite growth on *gat1*

Parasites require contact with and invasion of mammalian host cells to establish a niche within an intracellular parasitophorous vacuole in order to proliferate. In the context of two-dimensional monolayers of fibroblasts, parasites divide and eventually lyse out to invade neighboring cells to repeat the cycle. The area of the resulting plaque provides a measure of efficiency of a number of cellular processes. Past studies showed that plaque size growth is compromised by mutational blockade of Skp1 hydroxylation and earlier steps of the glycosylation pathway (Xu et al., 2012; Rahman et al., 2016, 2017). Here we find that disruption of *gat1* resulted in modestly smaller plaques, on average, after 5 d of growth, in either the RH or RHΔΔ background (Fig. 2A,B). Furthermore, plaque sizes were similarly reduced in an independent *gat1*Δ strain prepared in RHΔΔ using homologous recombination without CRISPR-Cas9 (Fig. 2C). To determine whether the effects were specific to the genetic lesion at the *gat1* locus, the RH and RHΔΔ KO strains generated using CRISPR/Cas9 were modified again by CRISPR/Cas9 to introduce single copies of epitope tagged versions of the *gat1* coding locus, downstream of an endogenous *gat1* promoter cassette or a tubulin promoter cassette, respectively, into the *uprt* locus. The expected insertions were confirmed using PCR (Figs. S3C, S4B). As a result, TgGat1-3xHA could be detected by Western blotting of tachyzoites in the RHΔΔ background (Fig. S4D) and, as discussed below, the complemented RHΔΔ strain restored Skp1 glycosylation according to a biochemical complementation test. Although TgGat1-Ty expressed under its own promoter cassette in the RH background was not detected (not shown), enzyme activity was partially restored in the TgGat1-Ty strain (Fig. S4C). Both strains exhibited larger plaque sizes than their respective KO parents (Fig. 2A, B).

Evolution of the Gat1 sequence

An evolutionary analysis was conducted to gain insight into the function of the TgGat1 gene product. Based on searches of genomic databases using BLASTP, TgGat1 is most closely related to CAZy GT8 sequences. The top-scoring hits, with Expect values of $<10^{-32}$, were found only in protists that contain *Toxoplasma* PgtA-like and Glt1-like sequences (Fig. S7) and lack *Dictyostelium* AgtA-like sequences, suggesting a common function. The most similar sequences, in searches seeded with the putative catalytic domain, belong to glycogenin, with Expect values of $\geq E^{-27}$. All other homologous sequences had Expect values of $\geq 10^{-22}$. Glycogenin is a self-modifying α4-glucosyltransferase that is a primer for glycogen synthesis by glycogen synthase and branching enzyme. *T. gondii* produces amylopectin, whose sequence is similar to that of glycogen, raising the possibility that Gat1 has a similar function in *T. gondii*.

Glycogenin consists of a CAZy GT8 family catalytic domain plus a C-terminal glycogen synthase binding domain separated by a linker, whereas Gat1 consists only of a single catalytic domain (Fig. 3A). TgGat1 is predicted to be a 345-amino acid protein encoded by a single exon gene in the Type I GT1 strain (TGGT1_310400). It is 34% identical to rabbit (*Oryctolagus cuniculus*) glycogenin over the catalytic domain, but includes a poorly conserved 90-amino acid sequence that interrupts the catalytic domain (Fig. 3A). A homologous Gat1-like sequence found in another, distantly related, protist, *Pythium ultimum*, (Uniprot K3WC47) was analyzed because it lacks this sequence. PuGat1 is predicted to be a 266-amino acid protein encoded by a 2-exon gene, annotated as PYU1_G002535-201 (Transcript ID PYU1_T002538).

To evaluate the evolutionary relationship of these putative glycosyltransferases, their catalytic domains those of the most closely related or known sequences from the CAZy GT8 family (Fig. S7) were aligned (Fig. S8) and analyzed by a Maximum Likelihood method (Fig. 3B). The results suggest that Gat1 and glycogenin evolved separately from a common ancestor. Though the last common ancestor was not resolved, Gat1 was presumably the predecessor to glycogenin owing to its presence in more primitive unicellular eukaryotes which bear no evidence of glycogenin-like sequences, and putative orthologs of Gat1 and glycogenin have not been observed in the same clade. Gat1 and glycogenin each possess unique conserved sequence motifs (Fig. S6) that likely support functional differences.

Enzymatic characteristics of Gat1

To address whether TgGat1 has the potential to directly modify Skp1, the predicted full-length protein was expressed as a His_6-tagged conjugate in *E. coli*, purified on a TALON resin, and treated with TEV protease leaving an N-terminal GlyHis-dipeptide stub before the start Met (Fig. 4A). The homolog from *Pythium ultimum* was prepared similarly. A screen for UDP-sugar hydrolysis activity of TgGat1 yielded, after extended reaction times, only UDP-Gal and UDP-Glc as candidate substrates from a panel of six common UDP-sugars (Fig. S9A). A quantitative comparison showed approximately 7-fold greater activity toward UDP-Gal than UDP-Glc (Fig. 4B).

The ability of Gat1 to transfer Gal or Glc to another sugar, rather than water, was tested using a substrate for glycogenin, Glcα1,4Glcα1-pNP (maltose-pNP), which mimics the terminal disaccharide of glycogen and starch and has a terminal αGlc as found on the Skp1 tetrasaccharide. Although Gat1 from either *T. gondii* or *P. ultimum* could modify maltose-pNP using either sugar nucleotide (Fig. 4D), the enzymes strongly preferred UDP-Gal. TgGat1 exhibited a K_m value for UDP-Gal of 30-72 μM but was not saturated with UDP-Glc at 0.5 mM. PuGat1 had a slightly lower apparent K_m for UDP-Gal (Fig. S9F), and both values were greater than the 2-4 μM values reported for rabbit and yeast glycogenins for UDP-Glc (Lomako et al., 1988, de Paula et al., 2005). TgGat1 and PuGat1 exhibited higher K_m values for maltose-pNP in the range of 16-43 mM (Fig. 4D), which were greater than the 4 mM value reported for rabbit glycogenin (Cao et al., 1995).

Extending the acceptor to 3 sugars or decreasing it to one resulted in less activity, but either anomer of Glc-pNP was acceptable (Fig. 4F). A similar pattern was observed for the *Pythium* and *Toxoplasma* enzymes. The enzymes were specific for Glc as Gal-pNP was inactive. In comparison, GlFGaGn-pNP, which mimics the natural acceptor on Skp1, was a superior acceptor substrate (Fig. 4F) with a K_m of 1.5

mM (Fig. S9G), and the truncated trisaccharide FGaGn-pNP was inactive indicating that Glc was the position of attachment. Thus the Gat1 enzymes preferred their native tetrasaccharide acceptor substrate and UDP-Gal as a donor, but tolerated, with low efficiency, the preferred substrates of glycogenin, UDP-Glc and α4-linked oligomers Glc.

The importance of acceptor glycan context was analyzed using GlFGaGn-Skp1, which was prepared by reaction of FGaGn-Skp1 with UDP-Glc and Glt1 resulting in loss of the trisaccharide epitope (Fig. 4H inset). In a direct comparison of acceptor concentration dependence, TgGat1 was about 33x more active toward GlFGaGn-Skp1 than GlFGaGn-pNP (Fig. 4H,I), based on ~33x less activity (dpm incorporated) at 0.001x the substrate concentration (substrate concentrations were both in the linear response range). The reaction with GlFGaGn-Skp1 did not approach saturation at the highest concentration tested, 6 μM. Thus the TgGat1 reaction was much more efficient when the tetrasaccharide was associated with its native substrate Skp1, and indicates that Gat1 is directly responsible for modifying Skp1 in the cell.

A characteristic of glycogenin is its ability to modify the HO-group of a Tyr side chain near its active site with αGlc, and then to modify the 4-position of the Glc with another αGlc, and repeat the process up to 8-12 sugars. When isolated in their recombinant forms from *E. coli*, neither TgGat1 nor PuGat1 were found to be unmodified, based on nLC/MS exact measurements (Fig. S10A-E). Furthermore, following incubation with either UDP-Gal or UDP-Glc, no change in SDS-PAGE mobility (Fig. S10D) or exact mass were observed (Fig. S10E). Thus, no evidence for autoglycosylation activity of Gat1 from either species could be detected.

Skp1 is the only detectable protein substrate in parasite extracts.

GlFGaGn-Skp1 is a substrate for Gat1, but are there others? This was addressed by complementing extracts of *gat1*Δ parasites with recombinant TgGat1 in the presence of UDP-[^{3}H]Gal, and measuring incorporation of ^{3}H after display of the proteome on a 1-D SDS-PAGE gel. A high level of incorporation of ^{3}H that depended on the addition of enzyme was observed at the position of Skp1 (Fig. 4J), as expected, but negligible dpm were detected elsewhere in the gel. Furthermore, negligible dpm were into Skp1 in RH parental cells, indicating that little GlFGaGn-Skp1 accumulates in wild-type cells. Similar results were observed in studies of *gat1*Δ clones in the RHΔΔ and Me49 backgrounds (Fig. S9H, I). and complementation of the *gat1*Δ clone in RHΔΔ with Gat1 expressed under the tubulin promoter resulted in absence of measurable incorporation into Skp1, confirming specificity for Gat1 expression *per se*.

Starch accumulates in Tg*gat1*Δ parasites

Although Gat1 was unable to serve as its own GT acceptor in the manner of glycogenin, its ability to modify α4Glc oligomers *in vitro*, albeit with low efficiency, raised the possibility that it may affect amylopectin (starch) biosynthesis in cells by applying αGal residues to its non-reducing termini. Starch accumulates to substantial levels in bradyzoites, a slow growing form of the parasite that is induced by stress. Since induction of bradyzoite differentiation in cell culture is more efficient in type 2 strains, Me49 and *gat1*Δ/Me49 cells were induced by pH up-shift and examined for differentiation by labeling of bradyzoite cyst walls with FITC-DBA-lectin, and for starch with the Periodic acid/Schiff's base reagent. As shown in Fig. 5, ME49 bradyzoites accumulated substantial levels of starch relative to tachyzoites, and no difference in the pattern or level was ascertained in *gat1*Δ cells using this qualitative assessment. Thus Gat1 is not required for starch synthesis, which is consistent with the absence of terminal Gal residues in sugar composition analyses in *T. gondii* starch (Coppin et al., 2005; Guerardel et al., 2005), and does not dramatically control starch abundance.

Gat1 generates a Galα1,3Glc- linkage

To determine the glycosidic linkage of the αGal residue transferred by TgGat1, the previously prepared ($^{13}C_6$)GlFGaGn-pNP (Rahman et al., 2017) was modified with TgGat1 using UDP-[U-^{13}C]Gal as the donor substrate. The pentasaccharide reaction product (approximately 25% conversion) was analyzed together with the tetrasaccharide starting material by NMR. The previous assignment of the chemical shifts of the tetrasaccharide (Rahman et al., 2017) facilitated provisional assignment of the

additional terminal Gal chemical shifts using the *CASPER* program (Jansson et al., 2006) and confirmed by analysis of the 2D COSY, TOCSY and HMBC spectra (Fig. S11). One-dimensional ^{1}H-NMR spectra reveals the presence of the mixture of tetra- and pentasaccharides, followed by a downfield shift in the ^{13}C-Glc-H1 peaks upon linkage to the terminal ^{13}C-Gal (Fig. S11A). The HMBC ^{1}H –^{13}C correlation spectrum shows the (Fig. S11B, center panel) connection from Gal-H1 to Glc-H3, consistent with the downfield peak in the ^{1}H –^{13}C-HSQC spectrum (Fig. S11B, top panel), and the proton resonances in the HSQC-TOCSY (Fig. S11B, bottom panel), establishing the glycosidic linkage between the terminal αGal and underlying αGlc as 1→3. Finally, a ^{1}H –^{1}H-COSY spectrum confirms the assignments of the underlying Glc-H1, H2 and H3 (Fig. S11C). Similar results were obtained when the tetrasaccharide was modified in the presence of PuGat1 (data not shown). Taken together, our NMR analyses are most consistent with the glycan structure: Galα1,3Glcα1,3Fucα1,2Galβ1,3GlcNAcα-. Thus, TgGat1 is a retaining UDP-Gal:glucoside α1,3-galactosyltransferase. Not only does Gat1 transfer the α-anomer of a different sugar compared to glycogenin, it attaches it to a different position (4- not 3-) of the acceptor α Glc residue. Significantly, energy minimization of the TgSkp1 glycan attached to Hyp, GaGlFGaGn-Hyp, yielded a very similar conformation to the DdSkp1 glycan, GaGaFGaGn-Hyp, using the GLYCAM program (Woods, 2014).

Skp1 glycan conformation

Previous energy minimization, NMR, and molecular dynamics studies suggested that the glycan of DdSkp1 forms a relatively stable conformation around which the local region of the polypeptide partially conforms. The association of the glycan with the polypeptide, which was consistent with NMR-relaxation measurements on individual sugar residues, was correlated with a dynamic opening of the neighboring F-box combining region of Skp1. To assess whether the glycan of TgSkp1, with its fourth sugar being the 4-epimer of the αGal found in *Dictyostelium*, conserves this conformation and correlation, the all-atoms molecular dynamics simulations applied to analyze glycosylated DdSkp1 were repeated on TgSkp1. As before, six 250 ns simulations were performed, with three including a 50-ns pre-equilibration of the glycan with a restrained protein backbone (Equil-1,-2,-3). One of the three demonstrated enhanced extension of the C-terminal helix-8 for the entire trajectory compared to the other two (Fig. S12), and exhibited a high frequency of hydrogen bonds 1-5 (Fig. 6A, B). The orientation of the glycan is superimposable upon the structure inferred for DdSkp1 (Fig. 6C), whose amino acid contacts are highly conserved (Fig. 6D). In Equil-1 and Equil-2, helix-8 folded back down part way through the trajectory, and this correlated well with reduced frequencies of hydrogen bonds 1, 2, 4, and 5, as illustrated in Fig. 6E for Equil-2. In the other three simulations (Sim-1-3), helix-8 was rarely extended, and hydrogen bonds 2, 4, and 5 were infrequent (Fig. S12). Thus high frequency occupancy of hydrogen bonds 1, 2, 4, and 5 were correlated with high frequency of extension of helix-8, which forms the binding pocket for the F-box domain. Although subject to the same caveats as the model developed for DdSkp1, the analysis indicates that the *T. gondii* glycan, with its different 4th sugar, has the same potential to influence the conformational ensemble of TgSkp1 to be more receptive to F-box protein binding.

Structural relationship of Gat1 to glycogenin

To further probe the relationship between Gat1 and glycogenin, we compared their structures by X-ray crystallography. Attempts to crystallize TgGat1 were unsuccessful, even after deletion of its unconserved insert (Fig. 3A). However, PuGat1, which lacks this insert, was crystallized in the presence of Mn^{2+} and UDP.

The crystal structure of PuGat1 in complex with Mn^{2+} and UDP was solved using single-wavelength anomalous dispersion phasing of a Pt^{2+} derivative, and the resolution was extended to 1.76 Å using a native data set (Table S3). The asymmetric unit contains a single chain of PuGat1 with unambiguous electron density for the nucleotide and Mn^{2+} ion (Fig. 7A). The overall structure of PuGat1 reveals a canonical GT-A fold (Bourne and Henrissat, 2001) consisting of eight α-helices and eight β-sheets. The N-terminus (residues 1-8) and two short loops (residues 80-96 and 242-244) are disordered and were not modeled. The structure is similar to glycogenin-1 from *Oryctolagus cuniculus* (Oc-glycogenin-1), which

superimposes 213 corresponding Cα atoms with an RMSD of 3.3 Å despite a sequence identity of only 34% (Fig. 7B). The application of crystallographic symmetry shows that PuGat1 forms the same dimer described (Gibbons et al., 2002) for the Oc-glycogenin-1 structure (Fig. 7B). According to PISA (Krissenel and Henrick, 2007) analysis, the PuGat1 dimer interface buries 1090 $Å^2$ with a favorable P-value of 0.107, which suggests that the dimer contact is stable. Sedimentation velocity analysis of 3.5 μM PuGat1 reveals a *c(s)* distribution consisting of single species at 4.0 S, which corresponds to the predicted value of 4.2 S for a dimer (Fig. 7C). The slightly slower sedimentation indicates that the enzyme in solution is less compact than that observed in the crystal structure. PuGat1 was substantially dimeric even at 0.3 μM (Fig. S14), suggesting that it forms a more stable dimer than Oc-glycogenin-1 (which has a reported K_d of 0.85 μM) (Bazan et al., 2008). Based on gel filtration and preliminary sedimentation velocity experiments (not shown), TgGat1 is also a high affinity dimer.

The PuGat1 active site shows that the conserved DxD motif (Bourne and Henrissat, 2001) and a conserved His residue coordinate the Mn^{2+} ion using the Oδ2 atom of D117, both Oδ1 and Oδ2 atoms of D119, and NE2 atom of His231 (Fig. 7D). The Mn^{2+} ion is also coordinated by the oxygen atoms from the α and β phosphates of UDP. Comparing the PuGat1 and Oc-glycogenin-1 active sites shows that all of the interactions with the nucleotide are conserved, with the exception of the interactions with N3 and O4 of the uracil ring (Fig. S13).

Comparing the crystal structures of the UDP-Glc:Oc-glycogenin-1 complex with PuGat1 immediately suggests a reason for why these enzymes have different donor specificities (Fig. 8). In the active site of the PuGat1 crystal structure, the Oγ atom of Thr180 forms a hydrogen bond with water499. Superimposing the Oc-glycogenin-1 structure onto PuGat1 shows that UDP-Glc would displace the water499 and bury the Thr180 with an unsatisfied hydroxyl (Fig. 8A). Since Thr180 is conserved in TgGat1, the unfavorable burial of a polar group likely explains why UDP-Glc is a poor donor (Fig. 8A). In contrast, the O4 atom of Gal would be ideally positioned to satisfy the buried Thr hydroxyl, offering a plausible explanation for the donor preference (Fig. 8B). Asp176, whose Cα atom underwent a 2.3 Å shift relative to its location in glycogenin, is in position to receive a hydrogen bond from the O4 atom of the modelled Gal (Fig. 8B). Other significant changes in the active site include a Leu to Ser substitution at residue 233, which removes a packing interaction with the sugar in PuGat1 (Fig. 8). The hydroxyl of the substituted Ser233 forms a hydrogen bond with the adjacent side chain of Gln206. A water molecule (W509) replaces the Leu side chain is occupied by water molecule and forms a hydrogen bond with Asn149.

To address the basis of Gat1's preference for the GlFGaGn- glycan, the lowest energy conformation of the alcohol form of the glycan generated by GLYCAM was docked using AutoDock Vina. A plausible docking mode was selected based on the requirement that the C'3-hydroxyl group must be oriented towards the anomeric carbon of the donor sugar to serve as the nucleophile for addition of the Gal, and that the glycan does not clash with the other subunit of the dimer. Out of 100 docking simulations, only the top scoring pose with a binding energy score of -5.7 kcal/mol satisfied the selection requirement. In this pose, the glycan was docked into a groove formed by Gat1 dimerization (Figs. 9A, B). The glycan is stabilized by hydrogen bonds contributed by residues D141, F143, and S234 from subunit A, and residues L212, K216, N217, and Y220 from subunit B with either the sidechain or the peptide backbone (Fig. 9C). In addition, hydrophobic packing against the faces of the sugar moieties is provided by residues T208, L212, F143, and W221 from subunit A, and residue Y220 from subunit B (Fig. 9D). F143, Y220, W221, and S234 are uniquely conserved in Gat1 proteins relative to glycogenins (Fig. S6). The conserved hydrophobic residues are probably the major contributors in terms of binding energy. The extensive shape complementarity, which could not be achieved using the same approach using the α4Glc-tertramer recognized by glycogenin, can explain the distinct preference of Gat1 for the Skp1 tetrasaccharide acceptor substrate.

DISCUSSION

Gat1 is unusual for glycosyltransferases. First, it appears to be dedicated to the glycosylation of a single target protein, Skp1. Second, it resides in the cytoplasmic compartment, rather than the secretory

pathway, of the cell. Third, it is widespread among protists, but not found outside of this kingdom. Fourth, it substitutes for another unrelated glycosyltransferase that modifies Skp1 in amoebozoa, but appears to be contribute a related regulatory function. Nevertheless, it is descended from a widely distributed lineage of sugar nucleotide-dependent glycosyltransferases. It is most closely related to glycogenin, a sublineage of retaining glycosyltransferases that potentially evolved from Gat1 to modulate glycogen formation outside of the protist kingdom, in yeast, fungi and animals.

Recombinant Gat1 from either *T. gondii* or *P. ultimum* is a CAZy GT8 family enzyme that, *in vitro*, can catalyze the transfer of αGal from UDP-Gal to non-reducing terminal Glc acceptors in an α1,3-linkage, as determined using NMR analysis of a synthetic version of the TgSkp1 glycan as an acceptor (Fig. S11). The enzyme was highly selective for UDP-Gal as the donor, but was able to transfer Glc from UDP-Glc at low efficiency (Figs. 4, S9). The enzyme also preferred the TgSkp1 tetrasaccharide relative to other mono- and oligo-saccharides, but is able to modify any non-reducing terminal Glc residue, in α- or β-configuration. These reactions are characterized by K_m values in the range of 70 μM for UDP-Gal and 1.5 mM for the TgSkp1 acceptor glycan, and exhibit pH and ionic strength optima that are consistent with action in the cytoplasmic compartment. In side-by-side comparisons, no substantial differences were found for Gat1's from *T. gondii* and *P. ultimum*, indicating a high degree of conservation despite the wide phylogenetic distance between the species.

In *T. gondii*, the *gat1* gene is required for addition of the terminal hexose on TgSkp1. Since the terminal hexose is αGal, based on its removal with green coffee bean α-galactosidase (Fig. 1C), and because Gat1 can efficiently and selectively catalyze the corresponding reaction *in vitro* (Fig. 4), it is concluded that Gat1 is directly responsible for this modification *in vivo*. To address whether other proteins might be modified by Gat1 *in vivo*, cytosolic extracts of tachyzoites were prepared from *gat1*Δ cells and incubated with recombinant Gat1 and UDP-[^{3}H]Gal. Significant incorporation was only detected in Skp1 (Figs. 4J, S9H, I). Because Gat1 is related to glycogenin, which is involved in glycogen formation in yeast and animals, amylopectin (starch) formation was examined in bradyzoites. However, there was no evidence for an effect on its accumulation (Fig. 5). These findings are consistent with Gat1's preference for Skp1-related acceptors *in vitro* (Fig. 4F). Furthermore, Gat1 was even more reactive with the Skp1 glycan as part of the Skp1 glycoprotein (Fig. 4H, I), indicating that the apoprotein contributes to enzyme recognition. These results are consistent with evidence that the underlying glycosyltransferases are also specific for Skp1 in *T. gondii* (Rahman et al., 2016; Rahman et al., 2017). The evidence that Gat1 is dedicated to TgSkp1 parallels apparent similar specificity for Skp1 in *D. discoideum* (West and Blader, 2015). In *D. discoideum*, the conclusion of specificity is strengthened by findings that genetic manipulation of Skp1 levels interacts genetically with its glycosylation (Wang et al., 2011).

The slow growth phenotype in the monolayer plaque assay is likely attributable to the absence of Gat1 because similar results were obtained in independent knockouts in three different genetic backgrounds, and the defect was corrected by genetic complementation under its own promoter or a strong tubulin promoter in the uprt locus (Fig. 2). Since the only substrate for Gat1's enzymatic activity that we detected was GlFGaGn-Skp1, it is likely that failure to fully glycosylate Skp1 is responsible for the growth defect. This interpretation is consistent with similar effects of knocking out early GTs in the pathway, but is milder than the *phyA*Δ phenotype (Rahman et al., 2016; Rahman et al., 2017; Xu et al., 2012). However, we cannot rule out some other function of Gat1, though we note that at 345 amino acids, nearly all of its sequence appears devoted to comprising the enzymatic domain (Fig. 3A). Future studies will be needed to address whether Gat1 contributes to fibroblast adhesion, invasion, proliferation within the parasitophorous vacuole, or egression from fibroblasts.

The mechanism by which full glycosylation affects Skp1 is likely to be similar to the model proposed for *D. discoideum*. According to the evidence from NMR relaxation studies and molecular dynamics simulations, the glycan was found to adopt a rather rigid conformation around which an intrinsically disordered region of the Skp1 polypeptide tended to fold, resulting in an opening of subsite 2 of the F-box binding region (Sheikh et al., 2017). Lowest energy calculations and MD simulations indicated that the TgSkp1 glycan, despite its placement of Glc rather than Gal at the fourth position, adopts a similar conformation (Fig. 6). Furthermore, the simulations of TgSkp1 suggest the occurrence of a hydrogen

bond involving the fifth sugar to the side chain of N147 (Fig. 6C) that was not described for DdSkp1, and reveal a strong temporal correlation between the presence of hydrogen bonds 1, 2, 4 and 5 and extension of the Skp1 C-terminus (Fig. S12). The importance of generating a pentasaccharide is reinforced by the observation that in the amoebozoa, which lack the prior enzyme Glt1 (Figs. 1A, B) that generates the substrate for Gat1, Glt1 (which evidently disappeared) and Gat1 were replaced by a new, unrelated enzyme, AgtA. AgtA is a dual function GT in the sense that it applies both the 4th and 5th sugars: each an α3linked Gal. This raises the interesting possibility that AgtA evolved to substitute an unexplained disappearance of Glt1 (and therefore Gat1) with recursive addition of the same sugar being an accessible evolutionary pathway to recover the pentasaccharide.

The most closely related protein that can be found in genomic sequence databases, glycogenin (Fig. 3), has the peculiar property of appending an αGlc residue to a particular Tyr within its own sequence and, in addition, is able to continue adding αGlc residues to the 4-position of the underlying Glc to form a linear oligomer of up to 12 residues. This activity has been proposed to prime glycogen synthesis where the glycogenin sequence is found throughout fungi and metazoa. The Gat1 sequence is observed in a range of alveolates (includes *T. gondii*), stramenopiles (includes *P. ultimum*), and rhizaria (altogether known as the SAR group), and archaeplastids (with the SAR group known as bikonts), but not in the more recently emerging unikonts, that include the amoebozoa, fungi and animals (Burki, 2014; Brown et al., 2018). Since glycogenin has the complementary distribution being found only in fungi and animals, it is tempting to speculate that Gat1 is the evolutionary precursor of glycogenin. This concept is supported by our failure to find both sequences in the same organisms or group of organisms in extant databases (Fig. 3). The aforementioned loss of Glt1 might have created the opportunity for Gat1 to evolve to a new biochemical role, and it is interesting to speculate whether this conserved some common shared cellular function.

To explore the hypothesis that Gat1 is the evolutionary precursor of glycogenin, the crystal structure of PuGat1 was determined and compared with that of glycogenin-1 from rabbit. The dimer interface of Gat1 precisely overlaps that of glycogenin (Figs. 7, S14). The UDP moiety of the sugar nucleotide donor is coordinated in the same manner in the active site (Fig. S13). Residues that recognize the Glc-moiety of UDP-Glc in glycogenin-1 are distinct in Gat1 in a way that can explain the preference for UDP-Gal (Fig. 8) while allowing inefficient processing of UDP-Glc as observed in the *in vitro* reactions (Fig. 4B-D). Importantly, UDP-Glc is able to inhibit the αGalT activity of Gat1 at concentrations that may be achieved *in vivo* (Fig. 4G), potentially explaining its ability to bind and process UDP-Glc albeit at low efficiency. Interestingly, this duality presages the recently discovered ability of glycogenin to utilize UDP-Gal during intermediate phases of glucan extension (Bilyard et al., 2018). On the other hand, presentation of the acceptor oligosaccharide is distinct. Computational modeling indicates a unique presentation of the Skp1 tetrasaccharide that involves contacts with both monomers at the dimer interface (Fig. 9), that is incompatible with docking of an α4-glucan. The 3-OH of the acceptor Glc is positioned appropriately to act as the nucleophile. If glycogenin evolved from Gat1, substantial changes occurred to allow both auto-glucosylation and extension of the initial Tyr-linked αGlc with additional α4-linked Glc residues in a manner that evidently depends on *trans*-interactions across a pre-existing dimer interface (Issoglio et al., 2012). Although the target Tyr-residue is present in a similar position in some Gat1's, this immediate region is poorly conserved among Gat1's (Fig. S6). Possibly, these features provided an evolutionary opportunity for the novel mechanism adopted by glycogenin once it was free to evolve new roles following the evident disappearance of Glt1. In addition, the protein acquired a C-terminal glycogen synthase binding domain of unknown function (Fig. 3A; Zequiraj et al., 2014)

Past and recent studies indicate that, despite the compelling *in vitro* data, glycogenin is not required for glycogen synthesis in yeast and mice (Torija et al., 2005; Testoni et al., 2017). Nevertheless, glycogen levels are affected in a manner that is currently unexplained. This role was evidently acquired later in evolution because Gat1 does not appear to have a significant effect on amylopectin accumulation in *T. gondii*, and *gat1* is expressed equally in tachyzoites and bradyzoites based on transcript analysis (Coppin et al., 2005), whereas starch accumulates predominantly in bradyzoites. Although the role of this enzyme lineage in Skp1 modification was lost by the time amoebozoa evolved *en route* to the appearance of

fungi/yeast and animals, this former activity indicates the potential for a new enzyme activity distinct from the glycogen priming activity that can be detected for glycogenin *in vitro*. Further studies to investigate additional potential targets of this protein in later-evolving eukaryotes are warranted.

ABBREVIATED EXPERIMENTAL PROCEDURES (See Supplement for detailed Procedures)

Maintenance of host cells and parasite manipulations

RH, RHΔΔ, ME49-RFP, and their genetically modified derivatives were maintained on HFF or hTERT cells in Dulbecco's modified Eagle's Medium supplemented with 10% (v/v) fetal bovine serum, and cloned by limiting dilution in 96-well plates. Parasites were transfected by electroporation.

Disruption and complementation of *gat1*

TGGT1_310400 (Toxodb.org), referred to as *gat1*, was disrupted by homologous recombination or CRISPR/Cas9, and complemented in the uprt locus using a tubulin or native gat1 promoter using CRISPR/Cas9. Desired modifications were confirmed by PCR, enzyme expression analysis, DNA sequencing, and protein expression.

Bradyzoite induction

ME49-RFP tachyzoites were differentiated to bradyzoites using alkaline pH, and differentiation was monitors by immunolabeling with FITC-DBA lectin.

Periodic acid staining

Amylopectin expression was assayed immunocytologically using a 1% periodic acid and Schiff reagent.

Expression and purification of recombinant TgGat1 and PuGat1

Recombinant forms of TgGat1 and PuGat1 were expressed from synthetic genes with N-terminal His_6-tags in E. coli, purified on Co^{+2} TALON resin, treated with TEV protease to remove the His_6-tag, and as indicated, purified further by gel filtration on a Superdex 200 column.

SDS-PAGE and Western blotting

Samples were separated by SDS-PAGE, Western blotted onto nitrocellulose, and probed with the indicated primary and fluorescent secondary antibodies. Fluorescence was recorded on a Li-Cor Odyssey infrared scanner.

Preparation of Skp1 peptides

To monitor Skp1 glycosylation status, Skp1 was immunoprecipitated from urea solubilized parasites, eluted with triethanolamine, digested with trypsin, and recovered on C18-ZipTips.

Treatment of TgSkp1 peptides with α-galactosidase

Trypsinates from above were treated with 3.6 mU of green coffee-bean α-galactosidase, and processed as above.

Mass spectrometry of TgSkp1 peptides

Peptides were separated on an Acclaim PepMap RSLC C18 column using a gradient of acetonitrile with 0.1% formic acid on an Ultimate 3000 RSLCnano UHPLC system, and introduced into the ion source of an Orbitrap QE+ mass spectrometer (Thermo Fisher Scientific). Full MS scans were acquired from *m/z* 350-2000 at 70k resolution, and MS^2 scans following higher energy collision-induced dissociation (HCD, 30) were collected for the Top10 most intense ions, with a 30-sec dynamic exclusion. The acquired raw spectra were analyzed using Sequest HT (Proteome Discoverer 2.2, Thermo Fisher Scientific) with a full MS peptide tolerance of 10 ppm and MS^2 peptide fragment tolerance of 0.02 Da,

and filtered to generate a 1% target decoy peptide-spectrum match (PSM) false discovery rate for protein assignments.

Enzyme assays

Sugar nucleotide hydrolysis: Recombinant TgGat1 (0.625-2.5 µM) was incubated with the indicated UDP-sugar, and UDP generated was quantitated using a UDP-Glo assay (Promega).

Glycosyltransferase activity toward small glycosides: In the standard reaction, TgGat1 or PuGat1 was incubated with 2 mM synthetic glycosides, UDP-[^{3}H]Gal, 50 mM HEPES-NaOH (pH 7.0), 2 mM $MnCl_2$, 5 mM DTT, in a final volume of 30 µl, for 1 h at 37°C. For kinetic studies, concentrations and times were varied as indicated, and kinetic parameters were analyzed according to the Michaelis-Menten model using Graph pad Prism software. Incorporation of radioactivity into pNP-glycosides was analyzed by capture and release from a Sep-Pak C18 cartridge and scintillation counting (Rahman et al., 2017).

Glycosyltransferase activity toward GlFGaGn-Skp1: To prepare Tg-GlFGaGn-Skp1, FGaGn-Skp1 (Rahman et. al 2017) was modified using TgGlt1 and UDP-Glc. Reaction progress was monitored by Western blotting with pAb UOK104, which is specific for FGGn-Skp1; approximately 85% of total Skp1 was modified. GlFGaGn-Skp1 was purified from Glt1 on a mini-QAE-column using a Pharmacia Biotech SMART System. TgGat1 was incubated with GlFGaGn-Skp1, UDP-[^{3}H]Gal, 0.2% (v/v) Tween-20, 50 mM HEPES-NaOH (pH 7.0), 2 mM $MnCl_2$, 5 mM DTT, in a final volume of 20 µl, for 1 h at 37°C. Incorporation into Skp1 was assayed by displaying the proteins on an SDS-PAGE gel, slicing the gel lanes into ~1-mm long pieces, and counting for radioactivity in a liquid scintillation counter.

To detect cellular Gat1 activity, a cytosolic S100 extract was prepared by hypotonic lysis, ultracentrifugation at 100,000 *g* × 1 h, and desalting (Rahman et al., 2016). S100 protein was incubated with Tg-GlFGaGn-Skp1 and UDP-[^{3}H]Gal, and incorporation of radioactivity into protein was assayed by SDS-PAGE and scintillation counting.

Glycosyltransferase activity toward parasite extracts: Cytosolic S100 fractions were incubated with TgGat1 and UDP-[^{3}H]Gal in a final volume of 60 µl containing 50 mM HEPES-NaOH (pH 7.0), 2 mM $MnCl_2$, 5 mM DTT, 30 mM NaF, 0.2% Tween-20, at 37°C for 1 h. Incorporation of radioactivity into protein was monitored by the SDS-PAGE assay.

Mass spectrometry of Skp1

To evaluate its glycosylation status, recombinant PuGat1 or TgGat1 was incubated with or without UDP-Gal or UDP-Glc, and introduced into an Acclaim PepMap analytical C4 column on an Ultimate 3000 RSLC system coupled to a QE+ Orbitrap mass spectrometer (Thermo Scientific). Protein was eluted with a gradient pf acetonitrile in formic acid into a Thermo QE+ Orbitrap mass spectrometer by nanospray ionization in positive ion mode. The MS method consisted of collecting Full ITMS (MS^1) scans (400-2000 *m*/*z*) at 140,000 resolution in intact protein mode. Data were processed with Xcalibur Xtract deconvolution software to generate monoisotopic masses from the multicharged, protonated ion series, or masses were extracted after MS spectra deconvolution using the ReSpect algorithm in the BioPharma Finder suite (Thermo Scientific).

Crystallography and modeling of PuGat1

A *Pu*Gat1:UDP:Mn^{2+} complex was crystallized using a hanging drop vapor diffusion method over a reservoir containing 8-12% (w/v) PEG4000, 0.4 M ammonium sulfate, and 0.1 M sodium acetate at pH 4.0. The complex crystallized in space group $P4_22_12$ and diffracted to 1.76 Å (Table S3). X-ray data were collected remotely at the SER-CAT 22-BM beamline at the Argonne National Laboratory using a Fast Rayonix 300HS detector, and processed using XDS (Kabsch 2010), with 5% of the data omitted for cross validation. Crystals were soaked with platinum cyanide for heavy-atom phasing. *Pu*Gat1:UDP:Pt^{2+} crystals were isomorphous and diffracted to 2.1 Å. The crystal structure of *Pu*Gat1:Pt^{2+} was solved using single-wavelength anomalous dispersion (SAD). The resulting protein phases had an acceptable figure of merit of 0.31 and the model was subjected to iterative cycles of refinement. The structure of *Pu*Gat1:UDP:Mn^{2+} was solved using rigid body refinement of *Pu*Gat1:Pt^{2+}. The resulting model was

subjected to iterative cycles of refinement and yielded a final model with Rwork/Rfree of 0.18/0.21 (Table S3). Images were rendered in PyMol (Delano 2002).

Glycan docking

The lowest energy conformation of the TgSkp1 tetrasaccharide (Glcα1,3Fucα1,2Galβ1,3GlcNAcα1-OH) was generated via GLYCAM (Woods, 2014), hydrogen atoms were added, and the electrostatic surface was assigned using AutoDockTools (Morris et al., 2009). The ligand was kept rigid and docked to PuGat1 using AutoDock Vina (Trott and Olson, 2010). The binding site described was selected from 100 docking modes examined.

Sedimentation velocity

Protein concentration was calculated from A_{280} , and loaded at 0.3-11 µM concentration into 12 mm double-sector Epon centerpieces equipped with quartz windows in an An60 Ti rotor. Sedimentation velocity data were collected using an Optima XLA analytical ultracentrifuge (Beckman Coulter) and data were recorded by absorbance in radial step sizes of 0.003 cm. SEDNTERP (Laue et al., 1992) was used to model the partial specific volume of *Pu*Gat1, and the density and viscosity of the buffer. Data were modeled as continuous *c*(*s*) distributions using SEDFIT (Schuck, 2000), and were fit using baseline, meniscus, frictional coefficient, and systematic time-invariant and radial-invariant noise. Predicted sedimentation coefficient (s) values for the monomer (2.8 S) and dimer (4.2 S) were calculated using HYDROPRO (Ortega et al., 2011). Data fit and *c*(*s*) plots were generated using GUSSI (Brautigam, 2015).

Molecular dynamics simulations

Simulations were performed basically as described before (Sheikh et al., 2017). Three independent 250-ns simulations were performed, and another three were run for 300 ns with the protein restrained for the first 50 ns to allow the glycan to relax.

Phylogenetic analysis of enzyme sequences

Proteins related to TgGat1 were searched for using a BLASTP against the NCBI non-redundant database. Their evolutionary relationships were investigated by using a Maximum Likelihood method (Le and Gascuel, 2008) and conducted in MEGA7 (Kumar et al., 2016). Sequence alignments were manually-curated in BioEdit (v 7.2.5). 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*], 1.02% sites).

ACKNOWLEDGMENTS

We thank Kentuan Hicks and Nathan Beattie for their assistance. KH was supported by an NSF REU grant: DBI-1426834. This project was supported in part by NIH RO1-GM084383 to CMW and Ira Blader, Grant #14-140 from the Mizutani Foundation for Glycoscience to CMW and Ira Blader, NIH P41-GM103490 (to LW, senior investigator), NIH 8P41-GM103390 (Resource for Integrated Glycotechnology to J. Prestegard), NIH R01-GM114298 to ZAW, and NIH P01-GM107012 to John Rose. MM was partially supported by NIH T32 training grant AI060546 to the Center for Tropical and Emerging Global Diseases.

ABBREVIATIONS

DBA, *Dolichos biflorus* lectin; Dd, *Dictyostelium discoideum*; FBP, F-box protein; GaGlFGaGn–, Galα1,3Glcα1,3Fucα1,2Galβ1,3GlcNAcα1–; GalT, galactosyl transferase; GT, glycosyltransferase; Hyp, (*2S*,*4R*)-4-hydroxy-l-proline; mAb, monoclonal antibody; P4H, prolyl 4-hydroxylase; pAb, polyclonal

antibody; pNP, para-nitrophenol; Pu, *Pythium ultimum*; RHΔΔ, RHΔ*ku80*Δ*hxgprt*; SCF, Skp1/Cullin-1/F-box protein subcomplex of E3 Cullin-RING-1 ubiquitin ligases; Tg, *Toxoplasma gondii*; Ub, ubiquitin

COMPETING INTERESTS

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

Msano Mandalasi, Investigation, Methodology, Writing–review and editing; Hyun W. Kim, Investigation, Methodology, Data curation, Writing–review and editing; Kazi Rahman, Investigation, Methodology; M. O. Sheikh, Methodology, Writing–review and editing; David Thieker, Methodology, Writing–review and editing; Elisabet Gas-Pascual, Methodology, Data curation; Peng Zhao, Methodology, Data curation; Nitin Daniel, Methodology, Data curation; Hanke van der Wel, Methodology; Travis H. Ichikawa, Methodology; John N. Glushka, Methodology, Data Curation; Lance Wells, Investigation, Methodology; Zachary A. Wood, Conceptualization, Supervision, Writing–review and editing; Christopher M. West, Conceptualization, Supervision, Funding acquisition, Writing–original draft, Project administration

SUPPLEMENT

The Supplement contains extended Experimental Methods, Figures S1-S14, and Tables S1-S4.

REFERENCES

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS,Terwilliger TC, PH Zwart (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst D* **66**:213-221

Bazan S, Issoglio FM, Carrizo ME, Curtino JA (2008) The intramolecular autoglucosylation of monomeric glycogenin. *Biochem Biophys Res Comm* **371**:328-332.

Bilyard MK, Bailey HJ, Raich L, Gafitescu MA, Machida T, Iglésias-Fernández J, Lee SS, Spicer CD, Rovira C, Yue WW, Davis BG (2018) Palladium-mediated enzyme activation suggests multiphase initiation of glycogenesis. *Nature* **563**:235-240.

Brautigam CA (2015) Calculations and publication-quality illustrations for analytical ultracentrifugation data. *Methods Enzymology* **562**:109-133.

Brown MW, Heiss AA, Kamikawa R, Inagaki Y, Yabuki A, Tice AK, Shiratori T, Ishida KI, Hashimoto T, Simpson AGB, Roger AJ (2018) Phylogenomics places orphan protistan lineages in a novel eukaryotic super-group. *Genome Biol Evol* **10**:427-433.

Burki F (2014) The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb Perspect Biol* **6**:a016147.

Cao Y, Steinrauf LK, Roach PJ (1995) Mechanism of glycogenin self-glucosylation. *Arch Biochem Biophys* **319**:293-298.

Chan MC, Holt-Martyn JP, Schofield CJ, Ratcliffe PJ (2016) Pharmacological targeting of the HIF hydroxylases--A new field in medicine development. *Mol Aspects Med* **47-48**:54-75.

Chinoy ZS, Schafer CM, West CM, Boons GJ (2015) Chemical synthesis of a glycopeptide derived from Skp1 for probing protein specific glycosylation. *Chemistry* **21**:11779-11787.

Coppin A, Dzierszinski F, Legrand S, Mortuaire M, Ferguson D, Tomavo S (2003) Developmentally regulated biosynthesis of carbohydrate and storage polysaccharide during differentiation and tissue cyst formation in *Toxoplasma gondii*. *Biochimie* **85**:353-361.

Coppin A, Varre JS, Lienard L, Dauvillee D, Guerardel Y, Soyer-Gobillard MO, Buleon A, Ball S, Tomavo S (2005) Evolution of plant-like crystalline storage polysaccharide in the protozoan parasite Toxoplasma gondii argues for a red alga ancestry. *J Mol Evol* **60**:257-267.

DeLano, W. L. (2002) PyMOL. DeLano Scientific, San Carlos, CA, 700

de Paula RM, Wilson WA, Roach PJ, Terenzi HF, Bertolini MC (2005) Biochemical characterization of *Neurospora crassa* glycogenin (GNN), the self-glucosylating initiator of glycogen synthesis. *FEBS Lett* **579**:2208-2214.

Dubey JP, Lindsay DS, Speer CA (1998) Structures of *Toxoplasma gondii* tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts. Clin Microbiol Rev 11:267–299.

Fox BA, Ristuccia JG, Gigley JP, Bzik DJ (2009) Efficient gene replacements in *Toxoplasma gondi*i strains deficient for nonhomologous end joining. *Eukaryot Cell* **8**:520-529.

Gaastra W, Lipman LJ, De Cock AW, Exel TK, Pegge RB, Scheurwater J, Vilela R, Mendoza L (2010) *Pythium insidiosum*: an overview. *Vet Microbiol* **146**:1-16.

Gas-Pascual E, Ichikawa HT, Sheikh MO, Serji MI, Deng B, Mandalasi M, Bandini G, Samuelson J, Wells L, West CM (2019) CRISPR/Cas9 and glycomics tools for *Toxoplasma* glycobiology. *J Biol Chem* **294**:1104-1125.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press. pp. 571-607.

Gibbons BJ, Roach PJ, Hurley TD (2002) Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *J Mol Biol* **319**:463-477.

Grainger JR, Wohlfert EA, Fuss IJ, Bouladoux N, Askenase MH, Legrand F, Koo, LY, Brenchley JM, Fraser ID, Belkaid Y (2013) Inflammatory monocytes regulate pathologic responses to commensals during acute gastrointestinal infection. *Nat Med* **19**:713-721.

Guérardel Y, Leleu D, Coppin A, Liénard L, Slomianny C, Strecker G, Ball S, Tomavo S (2005) Amylopectin biogenesis and characterization in the protozoan parasite *Toxoplasma gondii*, the intracellular development of which is restricted in the HepG2 cell line. *Microbes Infect* **7**:41-48.

Hurley TD, Stout S, Miner E, Zhou J, Roach PJ (2005) Requirements for catalysis in mammalian glycogenin. *J Biol Chem* **280**:23892-23899.

Islam MS, Leissing TM, Chowdhury R, Hopkinson RJ, Schofield CJ (2018) 2-oxoglutarate-dependent oxygenases. *Annu Rev Biochem* **87**:585-620.

Issoglio FM, Carrizo ME, Romero JM, Curtino JA (2012) Mechanisms of monomeric and dimeric glycogenin autoglucosylation. *J Biol Chem* **287**:1955-1961.

Jansson PE, Stenutz R, Widmalm G (2006) Sequence determination of oligosaccharides and regular polysaccharides using NMR spectroscopy and a novel Web-based version of the computer program CASPER. *Carbohydr Res* **341**:1003-1010.

Kabsch W (2010a) XDS. *Acta Cryst D* **66**:125-132

Kamoun S, Furzer O, Jones JD, Judelson HS, Ali GS, Dalio RJ, Roy SG, Schena L, Zambounis A, Panabières F, Cahill D, Ruocco M, Figueiredo A, Chen XR, Hulvey J, Stam R, Lamour K, Gijzen M, Tyler BM, Grünwald NJ, Mukhtar MS, Tomé DF, Tör M, Van Den Ackerveken G, McDowell J, Daayf F, Fry WE, Lindqvist-Kreuze H, Meijer HJ, Petre B, Ristaino J, Yoshida K, Birch PR, Govers F (2015) The Top 10 oomycete pathogens in molecular plant pathology. *Mol Plant Pathol* **16**:413-434.

Karplus PA, Diederichs K (2012) Linking crystallographic model and data quality. *Science* **336:**1030-1033

Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**:774-797.

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* **33**:1870-1874.

Laskowski RA, Swindells MB (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* **51**:2778-2786.

Laue TM, Shah BD, Ridgeway TM, Pelletier SL (1992) Computer-aided interpretation of analytical sedimentation data for proteins. In: Harding SE, Rowe AJ, Horton J, editors. Analytical Ultracentrifugation in Biochemistry and Polymer Science. Cambridge: The Royal Society of Chemistry. pp. 90–125.

Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* **25**:1307-1320.

Lévesque CA, Brouwer H, Cano L, Hamilton JP, Holt C, Huitema E, Raffaele S, Robideau GP, Thines M, Win J, Zerillo MM, Beakes GW, Boore JL, Busam D, Dumas B, Ferriera S, Fuerstenberg SI, Gachon CM, Gaulin E, Govers F, Grenville-Briggs L, Horner N, Hostetler J, Jiang RH, Johnson J, Krajaejun T, Lin H, Meijer HJ, Moore B, Morris P, Phuntmart V, Puiu D, Shetty J, Stajich JE, Tripathy S, Wawra S, van West P, Whitty BR, Coutinho PM, Henrissat B, Martin F, Thomas PD, Tyler BM, De Vries RP, Kamoun S, Yandell M, Tisserat N, Buell CR (2010) Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biol* **11**:1-22.

Lomako J, Lomako WM, Whelan WJ (1988) A self-glucosylating protein is the primer for rabbit muscle glycogen biosynthesis. *FASEB J* **2**:3097-3103.

Lomako J, Lomako WM, Whelan WJ (1990) The biogenesis of muscle glycogen: regulation of the activity of the autocatalytic primer protein. *BioFactors* **2**:251-254.

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

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. *J Computational Chemistry* **16**:2785-2791.

Ortega A, D. Amorós, J. Garc1a de la Torre (2011) .Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys J* **101**:892-898.

Pancha I, Shima H, Higashitani N, Igarashi K, Higashitani A, Tanaka K, Imamura S (2019a) Target of rapamycin-signaling modulates starch accumulation via glycogenin phosphorylation status in the unicellular red alga *Cyanidioschyzon merolae*. Plant J. **97**:485-499.

Pancha I, Tanaka K, Imamura S (2019b) Overexpression of a glycogenin, CmGLG2, enhances floridean starch accumulation in the red alga *Cyanidioschyzon merolae Plant Signal Behav* **14**:1596718.

Rahman K, Zhao P, Mandalasi M, van der Wel H, Wells L, Blader IJ, West CM (2016) The E3 ubiquitin ligase adaptor protein Skp1 Is glycosylated by an evolutionarily conserved pathway that regulates protist growth and development. *J Biol Chem* **291**:4268-4280.

Rahman K, Mandalasi M, Zhao P, Sheikh MO, Taujale R, Kim HW, van der Wel H, Matta K, Kannan N, Glushka JN, Wells L, West CM (2017) Characterization of a cytoplasmic glucosyltransferase that extends the core trisaccharide of the *Toxoplasma* Skp1 E3 ubiquitin ligase subunit. *J Biol Chem* **292**:18644-18659.

Roach PJ, Skurat AV (1997) Self-glucosylating initiator proteins and their role in glycogen biosynthesis. *Prog Nucleic Acid Res Mol Biol* **57**:289-316.

Sabin AB (1941) Toxoplasmic encephalitis in children. J. Am. Med. Assoc 116:801–807.

Schafer CM, Sheikh MO, Zhang D, West CM (2014) Novel regulation of Skp1 by the *Dictyostelium* AgtA α-galactosyltransferase involves the Skp1-binding activity of its WD40 repeat domain. J Biol Chem 289:9076-9088.

Schuck P (2000) Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophysical J* **78**:1606-1619.

Sheikh MO, Schafer CM, Powell JT, Rodgers KK, Mooers BH, West CM (2014) Glycosylation of Skp1 affects its conformation and promotes binding to a model F-box protein. *Biochemistry* **53**:1657-1669.

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

Sheikh MO, Halmo SM, Patel S, Middleton D, Takeuchi H, Schafer CM, West CM, Haltiwanger RS, Avci FY, Moremen KW, Wells L (2017) Rapid screening of sugar-nucleotide donor specificities of putative glycosyltransferases. Glycobiology **27**:206-212.

Sheikh MO, Thieker D, Chalmers G, Schafer CM, Ishihara M, Azadi P, Woods RJ, Glushka JN, Bendiak B, Prestegard JH, West CM (2017) O_2 sensing-associated glycosylation exposes the F-box-combining site of the *Dictyostelium* Skp1 subunit in E3 ubiquitin ligases. *J Biol Chem* **292**:18897-18915.

Shen B, Brown K, Long S, Sibley LD (2017) Development of CRISPR/Cas9 for efficient genome editing

in *Toxoplasma gondii*. *Methods Mol Biol* **1498**:79-103.

Soête M, Camus D, Dubremetz JF (1994) Experimental induction of bradyzoite-specific antigen expression and cyst formation by the RH strain of *Toxoplasma gondii in vitro*. *Exp Parasitol* **78**:361-370.

Stasic AJ, Chasen NM, Dykes EJ, Vella SA, Asady B, Starai VJ, Moreno SNJ (2019) The *Toxoplasma* vacuolar H^+-ATPase regulates intracellular pH and impacts the maturation of essential secretory proteins. *Cell Reports* **27**:2132-2146.

Striepen B, Soldati D (2007) Genetic Manipulation of *Toxoplasma gondii*. In: Weiss LM and Kim K. (Eds.), *Toxoplasma gondii*: the model apicomplexan. Perspectives and Methods. Elsevier Ltd.

Sugi T, Tu V, Ma Y, Tomita T, Weiss LM (2017) *Toxoplasma gondii* requires glycogen phosphorylase for balancing amylopectin storage and for efficient production of brain cysts. *MBio* **8**(4).

Testoni G, Duran J, García-Rocha M, Vilaplana F, Serrano AL, Sebastián D, López-Soldado I, Sullivan MA, Slebe F, Vilaseca M, Muñoz-Cánoves P, Guinovart JJ (2017) Lack of glycogenin causes glycogen accumulation and muscle function impairment. *Cell Metab* **26**:256-266.

Torija MJ, Novo M, Lemassu A, Wilson W, Roach PJ, François J, Parrou JL (2005) Glycogen synthesis in the absence of glycogenin in the yeast *Saccharomyces cerevisiae*. *FEBS Lett* **579**:3999-4004.

Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Computational Chemistry* **31**:455-461.

van der Wel H, Gas-Pascual E, West CM (2019) Skp1 isoforms are differentially modified by a dual function prolyl 4-hydroxylase/N-acetylglucosaminyltransferase in a plant pathogen. *Glycobiology*, in press.

Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, Stanley P, Hart G, Darvill A, Kinoshita T, Prestegard JJ, Schnaar RL, Freeze HH, Marth JD, Bertozzi CR, Etzler ME, Frank M, Vliegenthart JF, Lutteke T, Perez S, Bolton E, Rudd P, Paulson J, Kanehisa M, Toukach P, Aoki-Kinoshita KF, Dell A, Narimatsu H, York W, Taniguchi N, Kornfeld S (2015) Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25**:1323-1324.

West CM, Blader IJ (2015) Oxygen sensing by protozoans: how they catch their breath. *Curr Opin Microbiol* **26**:41-47.

West CM, Hart GW (2017) Nucleocytoplasmic glycosylation. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, Seeberger PH, editors. Essentials of Glycobiology [Internet]. 3rd edition. Cold Spring Harbor Laboratory Press; Chapter 18

West CM, Kim HW (2019) Nucleocytoplasmic O-glycosylation in protists. *Curr Opin Struct Biol* **56**:204-212.

West CM, Wang ZA, van der Wel H (2010) A cytoplasmic prolyl hydroxylation and glycosylation pathway modifies Skp1 and regulates O_2-dependent development in *Dictyostelium*. *Biochim Biophys Acta* **1800**:160-171.

Woods RJ (2014) GLYCAM Web, Complex Carbohydrate Research Center, University of Georgia,

Athens, GA

Xu X, Eletsky A, Sheikh MO, Prestegard JH, West CM (2018) Glycosylation promotes the random coil to helix transition in a region of a protist Skp1 associated with F-Box binding. *Biochemistry* **57**:511-515.

Xu Y, Brown KM, Wang ZA, van der Wel H, Teygong C, Zhang D, Blader IJ, West CM (2012) 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* **287**:25098-25110.

Zhang YW, Halonen SK, Ma YF, Wittner M, Weiss LM (2001) Initial characterization of CST1, a *Toxoplasma gondii* cyst wall glycoprotein. *Infect Immun* **69**:501-507.

Zeqiraj E, Tang X, Hunter RW, García-Rocha M, Judd A, Deak M, von Wilamowitz-Moellendorff A, Kurinov I, Guinovart JJ, Tyers M, Sakamoto K, Sicheri F (2014) Structural basis for the recruitment of glycogen synthase by glycogenin. *Proc Natl Acad Sci USA* **111**:E2831-2840.

Table 1. Strains employed in this study

Description	Name	Parental Strain	Genotype	Gene Targeted	Selection marker	Reference
RH (type 1)			WT			Sabin et al.
Δ*gat1*/RH	MM12, cl.8	RH	Δ*gat1*	*gat1*	DHFR	this report
gat1+/Δ*gat1*/RH	MM21, cl.E12	MM21	Δ*uprt*/tub::*gat1*-Ty	*uprt*	Δ*uprt*	this report
RHΔΔ		RH	Δku80/Δ*hxgprt*			Fox et al.
Δ*gat1*-1/RHΔΔ	KR10, cl.A1	RHΔΔ	Δku80/Δ*gat1*	*gat1*	HXGPRT	KR
Δ*gat1*-2/RHΔΔ	MM16, cl.B7	RHΔΔ	Δku80/Δ*hxgprt*/Δ*gat1*	*gat1*	DHFR	this report
gat1+/Δ*gat1*-2/ RHΔΔ	MM24, cl.G10	MM16	Δku80/Δ*hxgprt*/Δ*uprt*/ *gat1*::*gat1*-3xHA	*uprt*	Δ*uprt*	this report
Me49-RFP (type 2)	MM8, cl.A10	Me49	rfp+			Grainger et al.
Δ*gat1*/Me49-RFP	MM14, cl.B5	Me49-RFP	Δ*gat1*	*gat1*	DHFR	this report

Figure Legends

Fig. 1. Gat1 is required for terminal α-galactosylation of Skp1. (**A**, **B**) Schematic of Skp1 glycosylation pathway in *Toxoplasma* (Rahman et al., 2017, and herein) and *Dictyostelium* (West and Blader 2015). Gat1, the final enzyme in the *Toxoplasma* pathway, is in bold. (**C**) Glycopeptide analysis of Skp1 immunoprecipitated from type 1 RH and type 2 ME49 strains and their genetic derivatives, after trypsinization and mass analysis by nLC-MS. Pentasaccharide levels are shown above the x-axis, and tetrasaccharide levels are below, after normalization to total Skp1 peptides detected. In addition, tryptic peptides from RH were treated with green coffee bean α-galactosidase prior to nLC-MS analysis. Similar results were obtained in an independent trial from RH and Δ*gat1*/RH (not shown). See Table S2 and Fig. S5 for primary data.

Fig. 2. Dependence of parasite growth on Gat1. Parasites were plated at clonal density on two-dimensional monolayers of human foreskin fibroblasts (HFFs), and allowed to invade, proliferate, lyse out, and reinfect neighboring fibroblasts. After 5.5 d, cultures were fixed, stained, and analyzed for the areas occupied by lysed fibroblasts (plaques). Data from each of 3 independent trials, which were each normalized to the parental strain, were merged for presentation. (**A**) Comparison of the type 1 RH strain before and after *gat1* replacement using CRISPR/Cas9, and complementation with *gat1* under control of its own promoter cassette in the uprt locus. (**B**) Comparison of RHΔΔ, *gat1-2*Δ/RHΔΔ, and the latter complemented with gat1 under control of a tubulin promoter in the uprt locus. (**C**) Comparison of *gat1-1*Δ/RHΔΔ, prepared by homologous recombination.

Fig. 3. Phylogenetic analysis of Gat1- and glycogenin-like sequences. (**A**) Domain analysis of Gat1 from *T. gondii* and *P. ultimum* in comparison with human glycogenin-1. (**B**) The evolutionary history of the sequence of the Gat1 catalytic domain was inferred by using a Maximum Likelihood method. The tree with the highest log likelihood (-13279.56) is shown. Gat1 and Gat1-like sequences are colored green, glycogenin and glycogenin-like sequences are in red, and characterized and other selected other CAZy GT8 sequences are in black, or purple if predicted to reside in the secretory pathway rather than the cytoplasm. The percentage of trees in which the associated taxa clustered together is shown at each branch. Branch lengths are measured by the number of substitutions per site. See Figs. S6-S8 for alignments.

Fig. 4. Substrate specificity of Gat1. (**A**) Recombinantly expressed and purified preparations of TgGat1

and PuGat1 were analyzed by SDS-PAGE and staining with Coomassie blue. (**B**) Temporal dependence of UDP-Gal and UDP-Glc hydrolysis. The averages and standard deviations of 3 technical replicates are shown. A similar profile was observed with a different enzyme concentration. See Fig. S9E for a trial with higher enzyme concentrations. (**C**) Transferase activity utilizing 8 µM UDP-Gal or UDP-Glc toward 20 mM maltose-pNP for TgGat1 and PuGat1. The averages and standard deviations of two technical replicates are shown; similar profiles were in 2 independent assays with a different TgGat1 preparation. (**D**) UDP-Gal and UDP-Glc concentration dependence of TgGat1 transferase activity toward 20 mM maltose-pNP. The averages and standard deviations of two technical replicates are shown, and an independent trial with TgGat1 and PuGat1 against UDP-Gal is shown in Fig S9F. (**E**) Maltose-pNP concentration dependence of TgGat1 and PuGat1 transferase activity from 20 µM UDP-Gal. The averages and standard deviations of two technical replicates are shown. (**F**) Relative Gal-transferase activity of TgGat1 and PuGat1 toward different acceptors. The averages and standard deviations of three technical replicates are shown. Similar results were obtained in three independent trials. (**G**) Effect of UDP-Glc concentration on the Gal-transferase activity of TgGat1. Reactions were incubated for 1 h. The averages and standard deviations of two technical replicates are shown. **(H**) Gal-transferase activity of TgGaT1 toward varied concentrations of GlFGaGn-Skp1, in the presence of 40 µM UDP-Gal (1 µCi) after 1 h incubation. Data from independent preparations of TgSkp1 are colored in different shades. FGaGn-Skp1 is included for comparison. Error bars represent S.D. of duplicate measurements. Inset shows Western blots of the Skp1 preparations used, where FGaGn-Skp1, which is recognized specifically by pAb UOK104, is largely converted in a 3.5-h reaction using Glt1 and UDP-Glc to GlFGaGn-Skp1, which is recognized only by the pan-specific pAb UOK75. (**I**) Reactions with synthetic oligosaccharides conjugated to pNP were conducted in parallel using the same conditions. (**J**) Biochemical complementation to detect Gat1 substrates. Desalted S100 extracts of RH and *gat1*Δ/RH were reacted with recombinant Gat1 in the presence of UDP-[^{3}H]Gal, and the product of the reaction was separated on an SDS-PAGE gel which was sliced into 40 bands for liquid scintillation counting. The migration position of Skp1 is marked with an arrow. See Figs. S9H and S9I for trials using different strains.

Fig. 5. Starch accumulates normally in *gat1*Δ parasites. To promote normal starch accumulation, rapidly proliferating tachyzoites (panel **A**) of the type II strain Me49 (RFP expressing) and its *gat1*Δ derivative were induced to differentiate as slow-proliferating bradyzoite cysts (panel **B**) in human foreskin fibroblasts. Cultures were fixed and stained with Periodic acid/Schiff's base to reveal starch as a purple adduct. Arrow indicates a parasitophorous vacuole containing dozens to hundreds of tachyzoites within a fibroblast. Arrowhead indicates a cyst containing dozens of slow-growing bradyzoites, as confirmed by labeling of the cyst wall with FITC-DBA lectin (not shown). Scale bar = 50 µm. Two independent trials yielded similar results.

Fig. 6. Conformational analysis of glycosylated TgSkp1. Skp1, with its appended glycan whose terminal sugar is linked as determined in Fig. S11, was subjected to six 250-ns all-atoms molecular dynamics simulations. (**A**) The average ensemble of conformations of the full glycoprotein from a trajectory (Equil-3) in which the glycan was allowed to equilibrate during the first 50 ns. (**B**) Zoom-in of panel A to show hydrogen bonds present at >25% occupancy. (**C**) Superposition of glycans from TgSkp1 and DdSkp1 (from Sheikh et al., 2017). (**D**) Comparison of amino acid sequences of TgSkp1 and DdSkp1 over the region depicted. Red asterisks indicate residues involved in hydrogen bonding, and the black asterisk indicates the Hyp attachment residue. (**E**) Time course of a second trajectory (Equil-2), comparing the distance from C156 to the center of mass of residues of 1-136 (dashed green line in panel A) with occupancy of 5 hydrogen bonds (labeled in panel B). Data for this and the other 5 trajectories are shown in Fig. S12 and listed in Table S4.

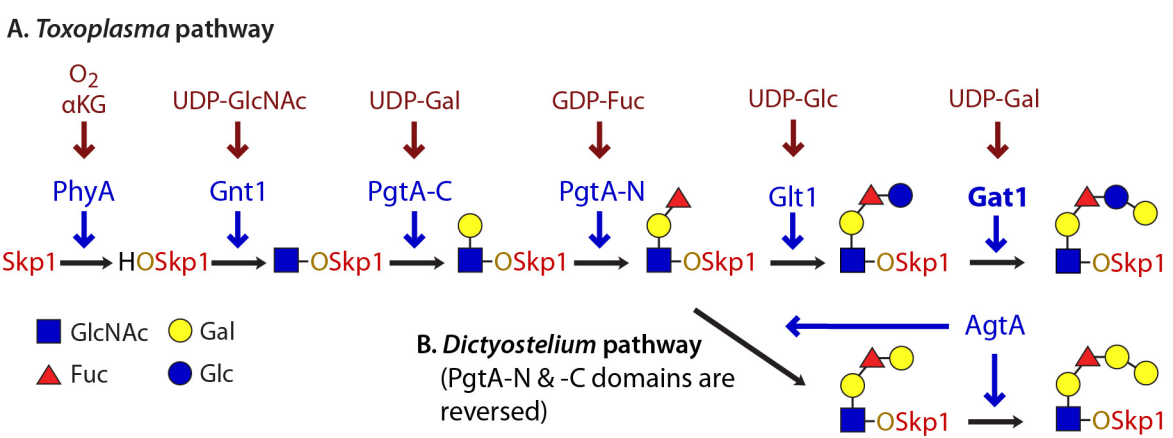
Fig. 7. Crystal structure of PuGat1 and its oligomeric state in solution. (**A**) The asymmetric unit of PuGat1:UDP:Mn^{2+} is shown with α-helices in red, β-strands in yellow, and loops in green. Secondary structures are assigned based on DSSP. The box highlights the active site with bound ligands. (**B**) PuGat1

and Oc-glycogenin-1 (PDB entry 1LL2) dimers are superimposed. The cylinders represent α-helices, and the arrows represent β-sheets. The red ellipse is the two-fold symmetry perpendicular to the page. (**C**) Sedimentation velocity experiment of 3.5 μM PuGat1 is displayed as a continuous *c*(*s*) distribution (normalized to 1.0). The black dashed represent the predicted dimer S-values based on the crystal structure. See Fig. S14 for additional data. (**D**) The difference density map (F_o-F_c) contoured at 5σ was calculated after omitting UDP and Mn^{2+} and subjecting the structure to simulated annealing Octahedral coordination of Mn^{2+} is satisfied by Asp117 and Asp119 of DxD motif, His231, and the α and β phosphates of UDP. See Fig. S13 for additional information.

Fig. 8. Active site comparison of PuGat1 and Oc-glycogenin-1. Comparison of the sugar binding pockets of PuGat1 and Oc-glycogenin-1. (**A)** The Glc moiety is modeled based on the Oc-glycogenin-1 crystal structure with its intact sugar nucleotide. **(B)** The Gal moiety is modeled by flipping the stereochemistry of Glc at the C4' position. PuGat1 and glycogenin-1 side chains are represented by green and gray sticks, respectively, and the yellow dashes represent hydrogen bonds. Water molecules are represented by blue spheres.

Fig. 9. Top docking pose reveals Gat1/Skp1-tetrasaccharide complementarity. (**A**) Skp1-tetrasaccharide binds in the PuGat1 active site and is accommodated by a groove formed by the dimer. (**B**) 90° turn of the image shown in panel A. (**C**) Hydrogen bonding interaction with the glycan is shown with the residues/ligand in sticks. (**D**) Hydrophobic packing with the faces of the sugar moieties and the methyl moiety of fucose is shown with residues in sticks/dots and ligand in sticks. Gat1 subunit A is represented in green, and subunit B in cyan.

Figure 1



Toxoplasma glycan: Galα1-3Glcα1-3Fucα1-2Galβ1-3GlcNAcα1-4O(*trans*)-Pro154-Skp1
Dictyostelium glycan: Galα1-3Galα1-3Fucα1-2Galβ1-3GlcNAcα1-4O(*trans*)-Pro143-Skp1

C. Glycosylation status of Skp1

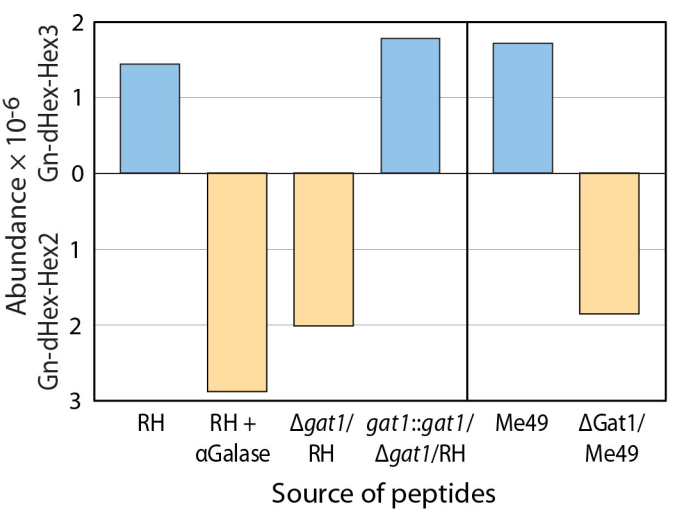


Figure 2

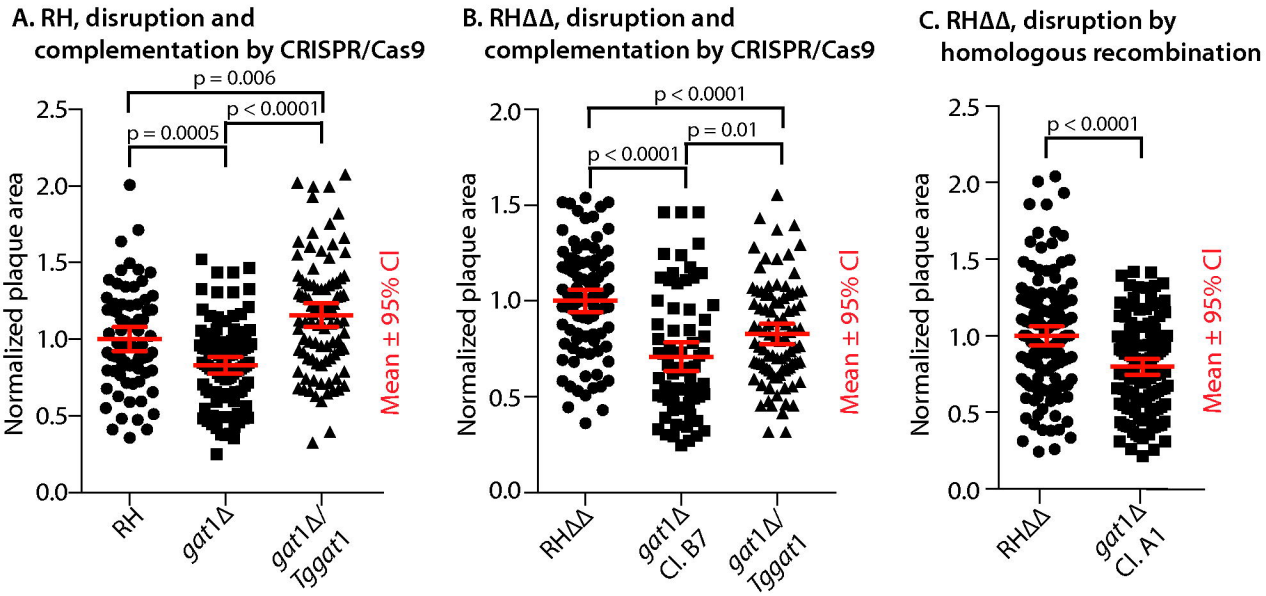
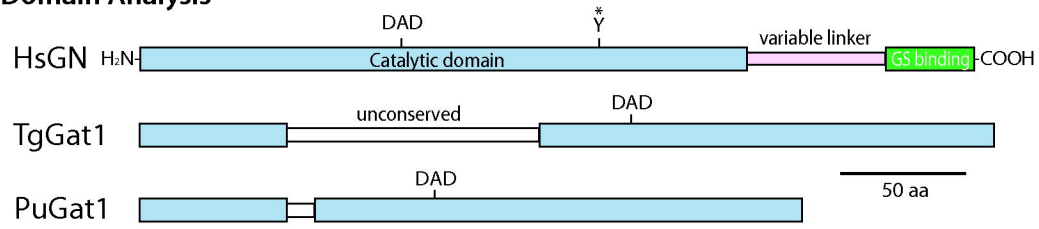


Figure 3

A. Domain Analysis



B. Evolutionary relationships

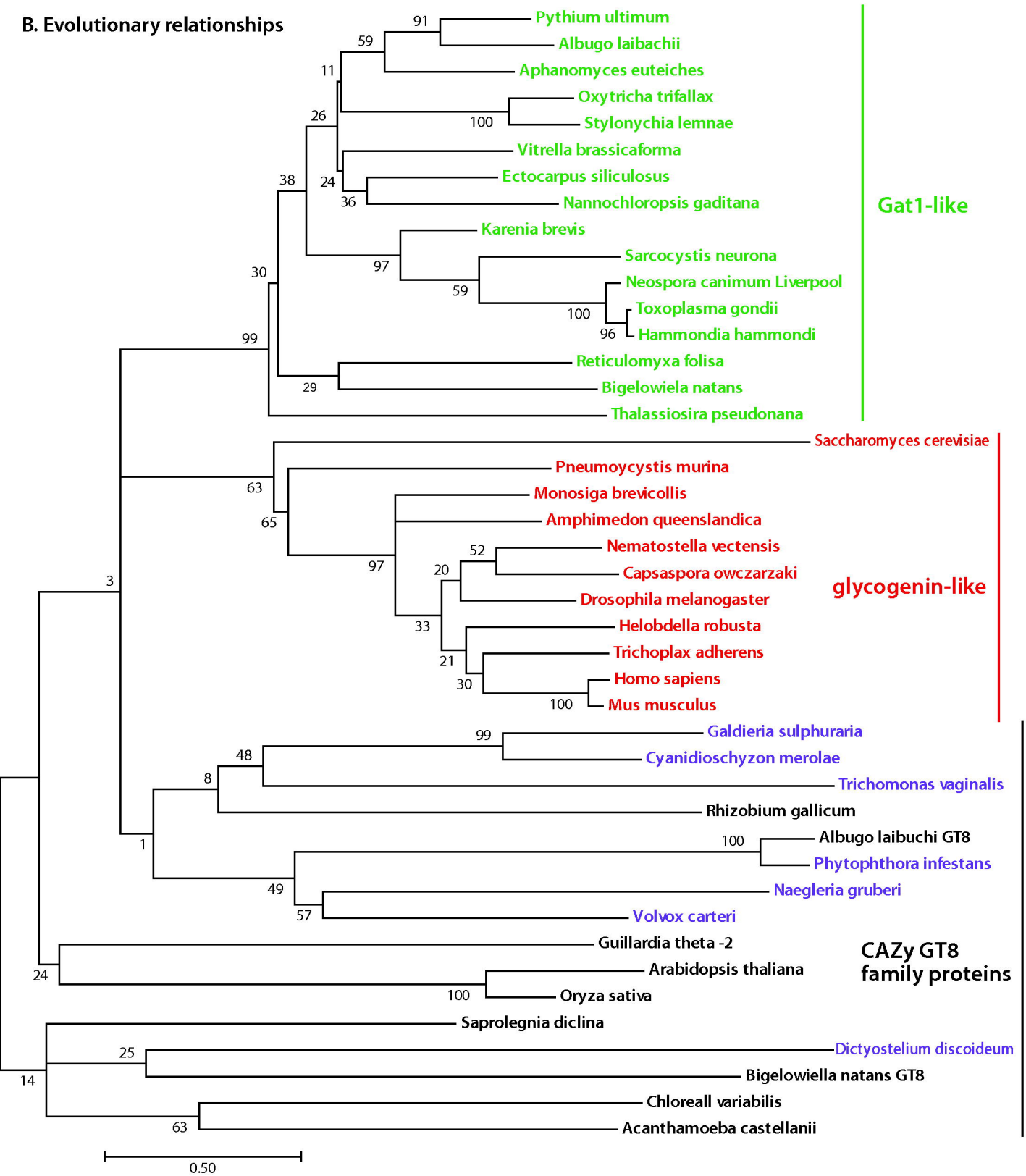


Figure 4

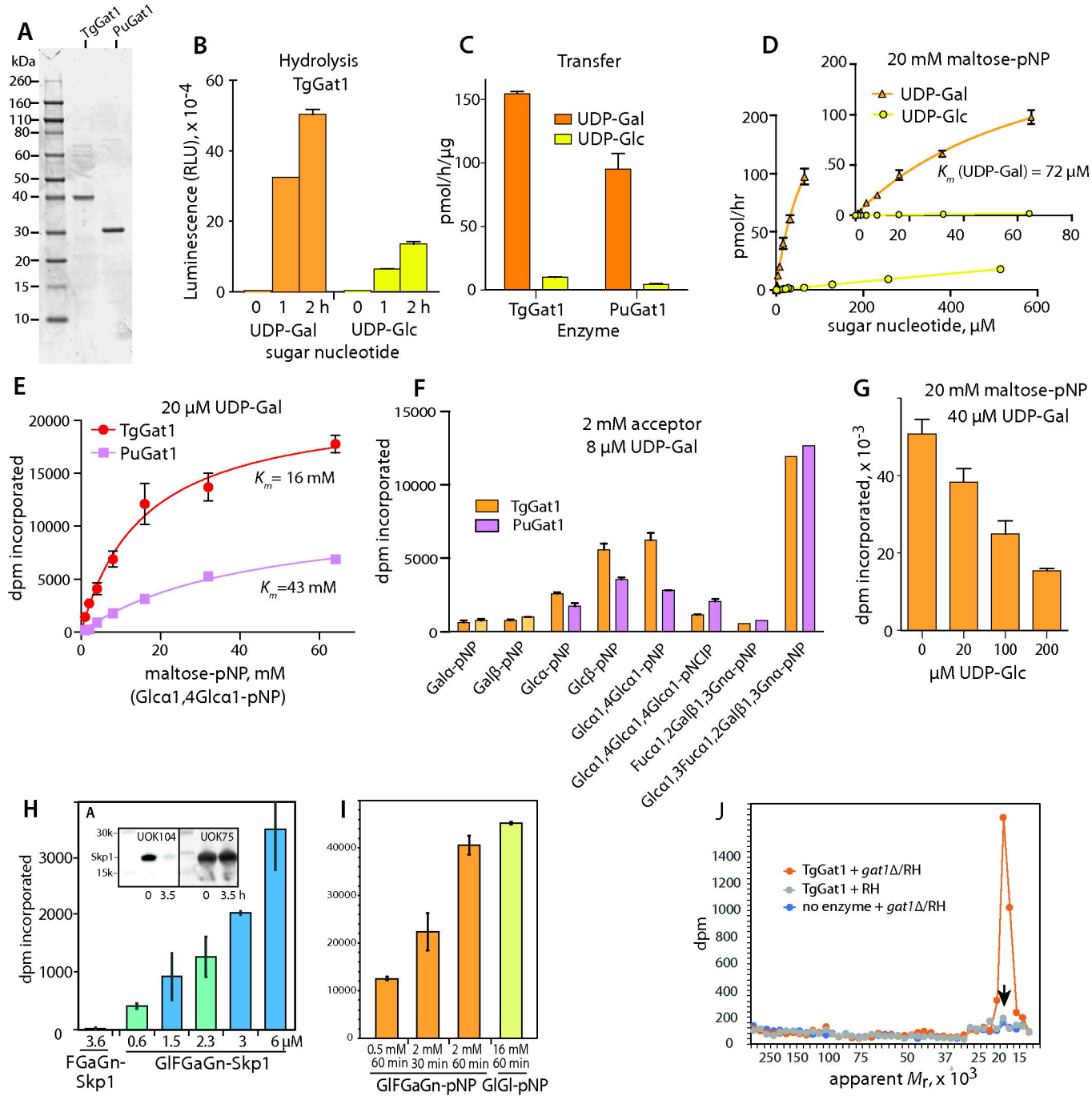


Figure 5

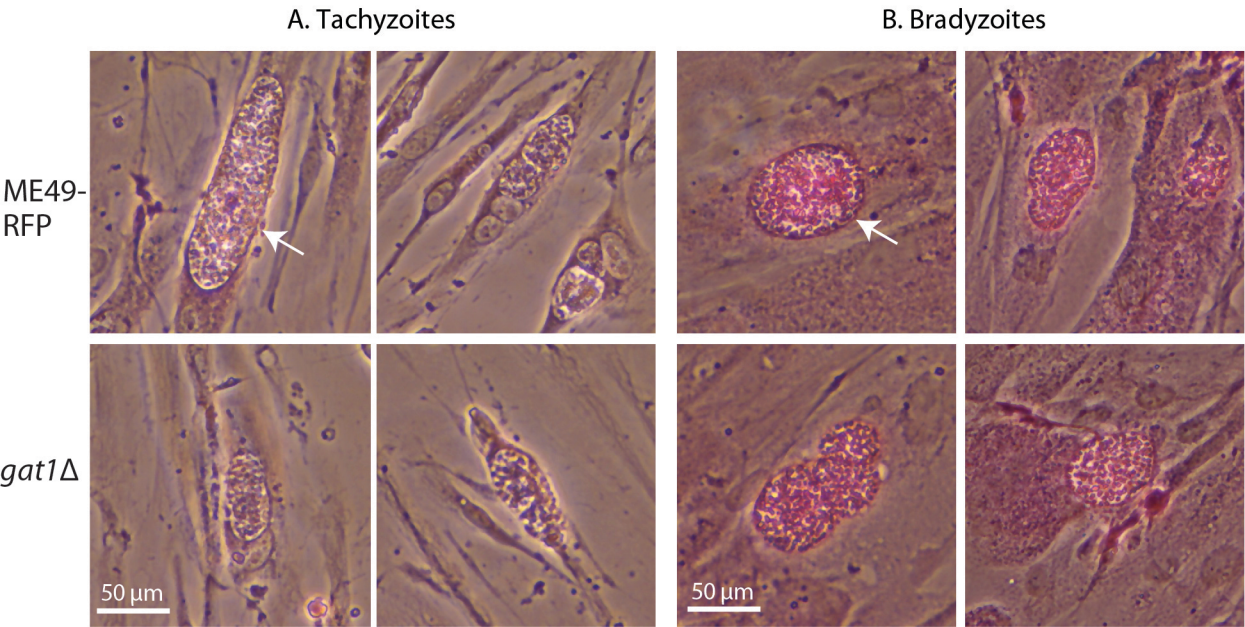


Figure 6

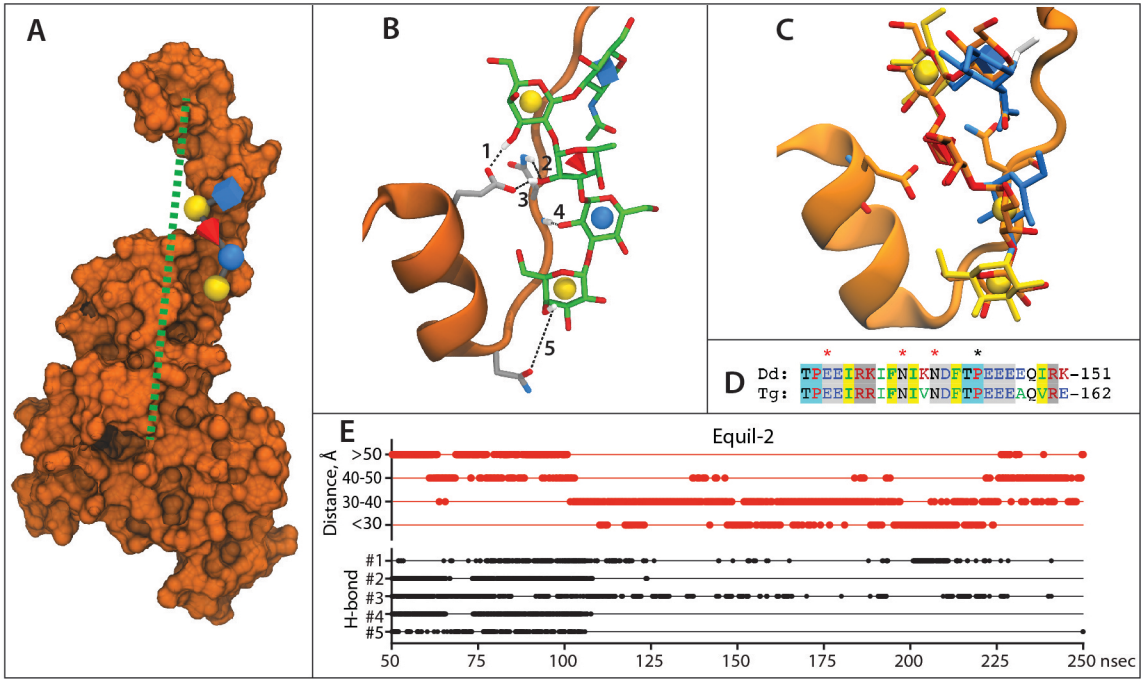

A
B
1
2
3
4
5
C
D
Dd: TPEEIRKIFNIKNDFTPEEEEQIRK-151
Tg: TPEEIRRIFNIVNDFTPEEEAQVRE-162
E
Equil-2
Distance, Å
>50
40-50
30-40
<30
H-bond
#1
#2
#3
#4
#5
50
75
100
125
150
175
200
225
250 nsec

Figure 7

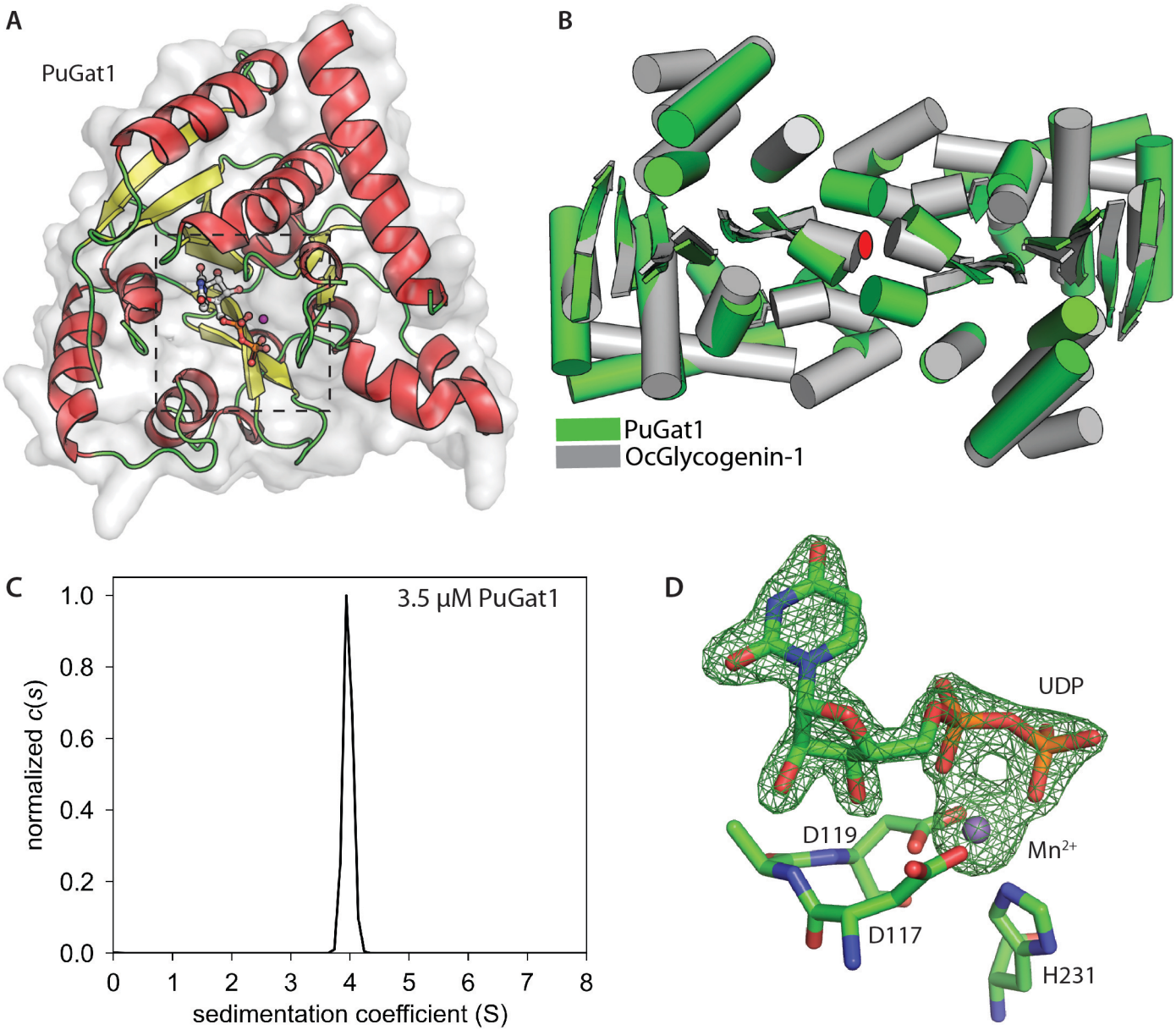

A
PuGat1
B
PuGat1
OcGlycogenin-1
C
3.5 μM PuGat1
normalized c(s)
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
7
8
sedimentation coefficient (S)
D
UDP
D119
Mn2+
D117
H231

Figure 8

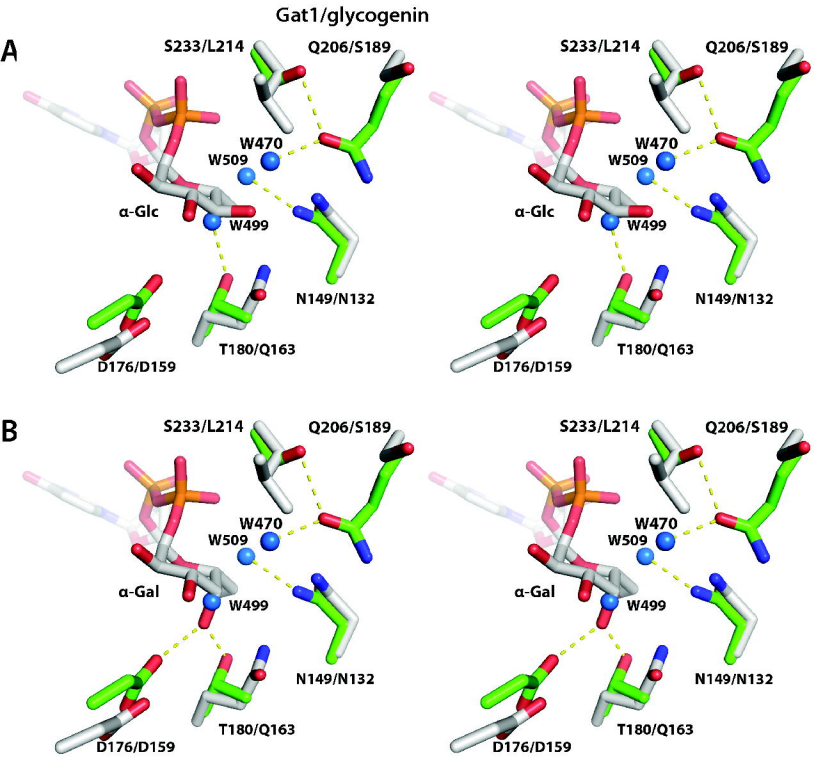
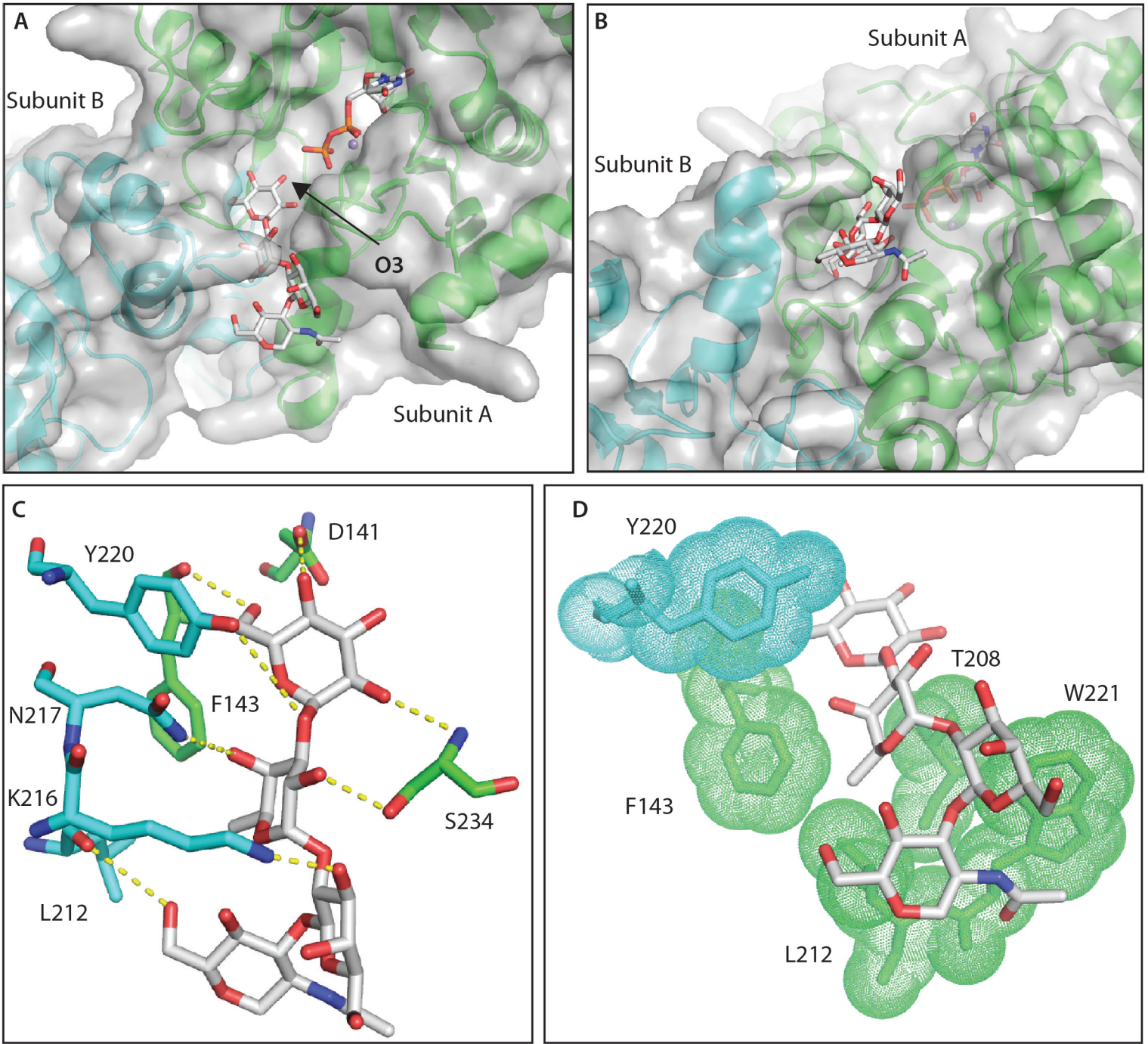


Figure 9


A
Subunit B
O3
Subunit A
B
Subunit A
Subunit B
C
D141
Y220
N217
F143
K216
S234
L212
D
Y220
T208
W221
F143
L212