

A terminal α 3-galactose modification regulates an E3 ubiquitin ligase subunit in *Toxoplasma gondii*

Msano Mandalasi^{1,2,*}, Hyun W. Kim^{1,*}, David Thieker^{3,4}, M. Osman Sheikh^{3,5}, Elisabet Gas-Pascual^{1,2}, Kazi Rahman^{1,6}, Peng Zhao³, Nitin G. Daniel¹, Hanke van der Wel¹, H. Travis Ichikawa^{1,7}, John N. Glushka³, Lance Wells^{1,3}, Robert J. Woods^{1,3}, Zachary A. Wood¹, Christopher M. West^{1,2,3,8}

¹Department of Biochemistry & Molecular Biology, ²Center for Tropical and Emerging Global Diseases,

³Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602 USA

⁴ Current address: Department of Biochemistry and Biophysics, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

⁵ Current address: Amicus Therapeutics, Philadelphia, PA

⁶ Current address: HIV Dynamics and Replication Program, Center for Cancer Research, National Cancer Institute, Frederick, Maryland, 21702

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 homolog glycosylates Skp1 in protists

Key words: X-ray crystallography, NMR, molecular dynamics simulation, Skp1, SCF, glycogenin, cytoplasmic glycosylation, posttranslational modification, *Toxoplasma*, *Pythium*

Abstract

Skp1, a subunit of E3 Skp1/Cullin-1/F-box protein ubiquitin ligases, is modified by a prolyl hydroxylase that mediates O₂-regulation of the social amoeba *Dictyostelium* and the parasite *Toxoplasma gondii*. The full effect of hydroxylation requires modification of the hydroxyproline by a pentasaccharide that, in *Dictyostelium*, influences Skp1 structure to favor assembly of Skp1/F-box protein subcomplexes. In *Toxoplasma*, the presence of a contrasting penultimate sugar assembled by a different glycosyltransferase enables testing of the conformational control model. To define the final sugar and its linkage, here we identified the glycosyltransferase that completes the glycan and found that it is closely related to glycogenin, an enzyme that may prime glycogen synthesis in yeast and animals. However, the *Toxoplasma* enzyme catalyzes formation of a Gal α 1,3Glc α - rather than the Glc α 1,4Glc α - linkage formed by glycogenin.

Kinetic and crystallographic experiments showed that the glycosyltransferase Gat1 is specific for Skp1 in *Toxoplasma* and also in another protist, the crop pathogen *Pythium ultimum*. The fifth sugar is important for glycan function as indicated by the slow-growth phenotype of *gat1* Δ parasites. Computational analyses indicated that, despite the sequence difference, the *Toxoplasma* glycan still assumes an ordered conformation that controls Skp1 structure and revealed the importance of non-polar packing interactions of the fifth sugar. The substitution of glycosyltransferases in *Toxoplasma* and *Pythium* by an unrelated bifunctional enzyme that assembles a distinct but structurally compatible glycan in *Dictyostelium* is a remarkable case of convergent evolution, that emphasizes the importance of the terminal α -galactose and establishes the phylogenetic breadth of Skp1 glycoregulation.

Introduction

A prominent mechanism of O₂-sensing in metazoa involves an O₂-dependent prolyl 4-hydroxylase (PHD2) that generates degrons on Hypoxia Inducible Factor-α (HIFα) that are recognized by the von Hippel Lindau subunit of the E3(VBC) ubiquitin ligase, leading to its polyubiquitination and degradation in the 26S-proteasome (1). Thus low O₂ inhibits PHD2 and stabilizes HIFα to dimerize with HIFβ and transcriptionally activate genes appropriate to respond to low O₂ (2). Protists also have a PHD2-like gene that encodes the evolutionary predecessor and likely ortholog of PHD2 (3). However, protists lack HIFα, and protist PhyA enzymes instead hydroxylate Skp1, a subunit of the SCF (Skp1/cullin-1/F-box protein)¹ family of E3 ubiquitin (Ub) ligases (4) that are related to the E3(VBC) Ub ligases. In the social amoeba *Dictyostelium discoideum*, Skp1 prolyl hydroxylation is involved in mediating the O₂-checkpoint for fruiting body formation (5). Thus, both prolyl hydroxylases contribute to regulation of the proteome, but transcriptionally in animals and likely via degradation in protists.

The mechanism by which Skp1 hydroxylation contributes to O₂-sensing has been examined in *D. discoideum* (6). Remarkably, the hydroxyproline (Hyp) residue is sequentially modified by a series of 5 glycosyltransferase (GT) reactions leading to the assembly of a linear pentasaccharide of recently defined structure (7). Genetic studies reveal a strong Skp1-dependent role for the GTs in O₂-sensing (8), and radiotracer and biochemical complementation studies indicate that Skp1 is the sole target of these GTs. Skp1 serves as an adaptor that links the F-box protein (FBP) and cullin-1 subunits of the SCF complex through mostly independent binding events. Interactome studies indicate that glycosylation promotes accumulation of Skp1/FBP subcomplexes *in vitro* and *in vivo* (9, 10), thereby potentially activating the respective SCF complexes and contributing to the degradation of all the substrates that are recognized by the cellular repertoire of FBPs. Current evidence indicates that the glycan assumes a relatively rigid structure that organizes the intrinsically disordered subsite-2 region involved in FBP recognition. By increasing the fraction of time that the C-terminal region is folded into α-helix-8 and extended to form subsite-2, the glycan is thought to promote the

interaction of Skp1 with FBPs in the cell (7). However, the importance of the exact glycan sequence and the significance of the terminal (fifth) sugar deserve further attention because there is limited precedent for glycan-imposed order on local protein structure (6), and because the information will provide needed new approaches to probe its role in cellular O₂-sensing.

Toxoplasma gondii is an apicomplexan parasite that latently infects a sizeable fraction of many human populations (11). Although chronic infections are usually clinically benign, reactivation, as can occur in immunosuppressed individuals, can lead to serious diseases of the central nervous system and other organs. Moreover, human fetuses are subject to serious birth defects in the case of acute maternal infection. Previous studies documented that the *T. gondii* prolyl hydroxylase (*phyA*) can complement a disruption of *phyA* in *D. discoideum*, and contribute to O₂-dependent growth of *T. gondii* in a fibroblast monolayer growth assay (12). Furthermore, coding sequences for the first three GT activities are present though, surprisingly, *T. gondii* still assembles a pentasaccharide glycan on its Skp1 (13). This led to the discovery of a novel enzyme that assembles the fourth sugar (14) and, as reported here, another enzyme that mediates addition of the final sugar. Analytical studies revealed that the fourth sugar is different from that of *D. discoideum*, but the fifth has remained unknown. These findings have raised important questions, including why are the GTs and glycans different between these two protists, what constrains their evolution, and are they compatible with the conformational regulation model posited for *D. discoideum*?

Pythium ultimum is an oomycete plant pathogen that resides in the stramenopile branch of the larger TSAR group of protozoans to which the apicomplexans also belong. *P. ultimum* is the agent for root rot disease in agriculturally important crops (15, 16), and is related to *P. insidiosum*, the agent for debilitating pythiosis in humans and other mammals (17). We recently showed *P. ultimum* PhyA and the first GT Gnt1 constitute a bifunctional protein that is active toward Skp1A but not the second Skp1B encoded by its genome (18). Thus O₂-dependent posttranslational regulation of Skp1 appears to be widespread, and *P. ultimum*, and protists from other lineages, have evidently evolved a second Skp1 that avoids O₂-dependent

glycoregulation. We show here that the terminal glycosylation of *P. ultimum* Skp1 is like that of *T. gondii* rather than that of *D. discoideum*.

By identifying the Gat1 GT that mediates addition of the final sugar to the *T. gondii*, we have been able to determine the complete glycan sequence, and to show that Gat1 is specific for Skp1, regulates Skp1, and contributes to optimal growth in cell culture. Despite a distinct structure, the glycan mediates a conserved conformational effect on Skp1. We propose that the selection pressure for a 5-sugar chain was so strong that, in lineages where the Gat1 progenitor was evolving to acquire a new function, the later-evolving amoebozoan acquired a novel GT mechanism to ensure assembly of a glycan that was structurally similar enough to still conformationally control Skp1.

Results

The fifth and final 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 (13, 14). 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 complete conversion of the pentasaccharide form of the glycopeptide to the tetrasaccharide form, indicating that the terminal Hex is an αGal residue (Fig. 1C).

Genomics studies predict the existence of four cytoplasmically localized GTs whose functions are not assigned (20), and one of these, referred to as Gat1, was predicted to be the missing Skp1 GT on account of the phylogenetic co-distribution of its gene in protists that possess Glt1-like genes (14). To test the dependence of Skp1 glycosylation on Tggat1, the gene was disrupted using CRISPR/Cas9 in the RH strain, yielding *gat1Δ* as described in Fig. S4A. PCR studies confirmed replacement with a *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 Tggat1 is required for addition of the terminal sugar, but whether this was a direct or indirect effect was unclear.

Toxoplasma growth depends partially on *gat1*

Parasites require invasion of mammalian host cells to establish a niche within an intracellular parasitophorous vacuole in order to proliferate (24, 25). In two-dimensional fibroblast monolayers, parasites lyse out and invade neighboring cells to repeat the cycle, resulting in plaques whose area is a measure of efficiency of these cellular processes. Past studies showed that plaque size growth is compromised by mutational blockade of Skp1 hydroxylation and earlier steps of the glycosylation pathway (12-14). Similarly, disruption of *gat1* also resulted in modestly smaller plaques in both RH or RHΔΔ backgrounds (Fig. 2A,B), and in RHΔΔ in which *gat1* was disrupted 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), confirming that the effect on growth was due to the original loss of Gat1.

Gat1 is closely related to glycogenin

An evolutionary analysis was conducted to gain further insight into the function of TgGat1. Based on searches of genomic databases using BLASTP, TgGat1 belongs to the CAZy GT8 family. 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 *D. discoideum* 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 dimeric α 4-glucosyltransferase that may prime the synthesis of glycogen in the cytoplasm of yeast and animals by a mechanism that involves auto-glycosylation. Glycogenin appears not to be involved in starch formation and is an evolutionarily recent addition to the glycogen biosynthesis pathways that occur in the absence of glycogenin activity in bacteria and many unicellular eukaryotes (26). The cyst-forming stage of *T. gondii* accumulates crystalline amylopectin (27), an α 1,4Glc polymer with α 1,6-linked branches that resembles glycogen, 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 that of glycogen storing animals and fungi. Homologs of glycogen synthase (27) and glycogen phosphorylase (28) regulate the accumulation of amylopectin in this parasite, and TgGat1 has been annotated as a glycogenin (28, 29). Related genes have been implicated in promoting starch formation in red algae (30,31). Thus Gat1 might have a function in regulating *T. gondii* amylopectin formation.

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). This region is likely unstructured based on secondary structure prediction by the XtalPred server (32).

The Gat1-like sequence from *P. ultimum* (Uniprot K3WC47) lacks this sequence, so was analyzed for comparison. 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 further evaluate the evolutionary relationship of these putative GTs, their catalytic domains and 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 potentially the predecessor to glycogenin owing to its presence in other primitive unicellular eukaryotes which bear no evidence of glycogenin-like sequences, and because there is currently no evidence for the existence of close homologs of Gat1 and glycogenin within the same clade that would suggest that they evolved as paralogs of a gene duplication. However, the possibility that either one or the other product of an ancestral gene duplication was always lost in every extant derivative cannot be excluded. Gat1 and glycogenin each possess unique conserved sequence motifs (Fig. S6) that potentially support functional differences.

Gat1 is a terminal Skp1 α -galactosyltransferase

To determine if TgGat1 can directly modify Skp1, the predicted full-length protein was expressed as a His₆-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 presumptive ortholog 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 analog for glycogenin, Glc α 1,4Glc α 1-pNP (maltose-pNP), which mimics the terminal disaccharide of glycogen and starch, and has a terminal α Glc as found on the Skp1 tetrasaccharide.

Although Gat1 from either *T. gondii* or *P. ultimum* could modify maltose-pNP using either sugar nucleotide (Fig. 4C), the enzymes strongly preferred UDP-Gal (Fig. 4D). Furthermore, TgGat1 activity was not saturated by UDP-Glc at 0.5 mM (Fig. 4D), whereas TgGat1 and PuGat1 exhibited apparent K_m values for UDP-Gal in the range of 30–15 μ M (Fig. S9F). These values were greater than the 2–4 μ M values reported for rabbit and yeast glycogenins for UDP-Glc (33, 34). TgGat1 and PuGat1 exhibited higher apparent 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 (35).

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 *P. ultimum* and *T. gondii* enzymes. The enzymes were specific for terminal Glc acceptors as activity towards Gal-pNP was not detected. 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 α 4-linked oligomers Glc.

The importance of Skp1 as context for the acceptor glycan was examined 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 comparison of acceptor concentration dependence, TgGat1 was about 33x more active toward GlFGaGn-Skp1 than free 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 μ M. Thus the TgGat1 reaction was much more efficient when the tetrasaccharide was associated with its native substrate Skp1, and thus consistent with Gat1 being 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 α Glc, and then to repeatedly modify the 4-position of the Glc with another α Glc, and repeat the process up to 8–12 sugars. Thus, when isolated as a recombinant protein expressed in UDP-Glc positive *E. coli*, glycogenin is partly glucosylated (36). However, neither TgGat1 nor PuGat1 prepared in a similar manner were found to be glycosylated, based on an exact mass measurement using nLC/MS (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 substrate of Gat1 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 1D 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 incorporated 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 RH $\Delta\Delta$ and Me49 backgrounds (Fig. S9H, I). Finally, complementation of the *gat1* Δ clone in RH $\Delta\Delta$ with Gat1 expressed under the tubulin promoter resulted in absence of measurable incorporation into Skp1, confirming specificity for Gat1 expression *per se*.

Although Gat1 was unable to serve as its own GT acceptor in the manner of glycogenin, its ability to modify α 4Glc oligomers *in vitro*, albeit with low efficiency, raised the possibility that it may affect amylopectin (starch) biosynthesis in cells by applying α Gal residues to its non-reducing termini. Starch normally 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 does not appear to affect starch synthesis, which is consistent with the absence of terminal Gal residues in a sugar composition analyses of *T. gondii* starch (29, 37), and the finding that *gat1* is expressed equally in starch-poor tachyzoites and starch-rich bradyzoites based on transcript analysis (29).

Gat1 generates a Galα1,3Glc-linkage

To determine the glycosidic linkage of the αGal residue transferred by TgGat1, the previously prepared ($^{13}\text{C}_6$)GlFGaGn-pNP (14) was modified with TgGat1 using UDP-[U- ^{13}C]Gal as the donor substrate. The pentasaccharide reaction product (approximately 30% conversion) was analyzed together with the tetrasaccharide starting material by NMR. The previous assignment of the chemical shifts of the tetrasaccharide (14) facilitated provisional assignment of the additional terminal Gal chemical shifts using the CASPER program (38) and confirmed by analysis of the 2D COSY, TOCSY and HMBC spectra (Fig. 6). One-dimensional ^1H -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. 6A). The HMBC ^1H – ^{13}C correlation spectrum shows the (Fig. 6B, center panel) connection from Gal-H1 to Glc-H3, consistent with the downfield peak in the ^1H – ^{13}C -HSQC spectrum (Fig. 6B, top panel), and the proton resonances in the HSQC-TOCSY (Fig. 6B, bottom panel), establishing the glycosidic linkage between the terminal αGal and underlying αGlc as 1 \rightarrow 3. Finally, a ^1H – ^1H -COSY spectrum confirms the assignments of the underlying Glc-H1, H2 and H3 (Fig. 6C). Consistent 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:

Galα1,3Glcα1,3Fucα1,2Galβ1,3GlcNAca-, indicating that 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 αGlc residue.

Comparison of crystal structures explains the catalytic differences between Gat1 and 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 co-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 (39) consisting of eight α-helices and eight β-strands. The N-terminus (residues 1-8) and two 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 (40) for the Oc-glycogenin-1 structure (Fig. 7B). According to PISA (41) 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 dimeric even at 0.3 μM (Fig. S12), suggesting that it forms a dimer with an affinity >2-fold higher than that of Oc-glycogenin-1 (which has a reported K_d of 0.85 μM) (42). Based on gel filtration and preliminary sedimentation velocity experiments (not shown), TgGat1 is also a dimer.

The PuGat1 active site shows that the conserved DxD motif (43) 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 Mn²⁺ 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). Other changes in Gat1 active site include a Leu to Ser substitution at residue 233, which can potentially remove a packing interaction with the donor 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 and forms a hydrogen bond with Asn149.

Comparing the crystal structure of UDP-Glc bound Oc-glycogenin-1 with PuGat1 immediately suggests a reason for why these enzymes have different donor specificities (Fig. 8). Superimposing the Oc-glycogenin-1:UDP-Glc structure onto PuGat1 shows that UDP-Glc would displace water499 coordinated by Thr180 (Fig. 8A). This would leave the Thr180 hydroxyl group unsatisfied, and the unfavorable burial of a polar group likely explains why UDP-Glc is a poor donor (Fig. 8A). In contrast, we modelled in UDP-Gal by flipping the stereochemistry at C4 position. The O4 atom of Gal would be ideally positioned to satisfy the Thr180 hydroxyl group. Concurrently, Asp176, whose C α atom underwent a 2.3 Å shift relative to its location in glycogenin, would be in position to receive a hydrogen bond from the O4 atom of the Gal (Fig. 8B). Altogether, Gat1's sugar donor preference for UDP-Gal is likely due to formation of favorable hydrogen bonds with O4 atom of Gal in contrast to the burial of Thr180 hydroxyl group when binding UDP-Glc.

Computational modeling predicts the specificity of Gat1 towards the Skp1 tetrasaccharide

To address the basis of Gat1's preference for the GlFGaGn- glycan, the lowest energy conformation of the reducing form of the glycan generated by GLYCAM-web (www.glycam.org) 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 adopted an alignment in a groove formed by Gat1 dimerization (Figs. 9A, B). The glycan is stabilized by hydrogen bond contributions from the sidechains or peptide backbones of residues D141, F143, and S234 from subunit A, and residues L212, K216, N217, and Y220 from subunit (Fig. 9C). In addition, non-polar interactions against the faces of the sugar moieties are 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 packing interactions from conserved hydrophobic residues are likely the major contributors in terms of binding energy. The extensive electrostatic and packing complementarity, which could not be achieved using the same approach using the α 4Glc-terramer recognized by glycogenin, can explain the distinct preference of Gat1 for the Skp1 tetrasaccharide acceptor substrate.

The *Toxoplasma* glycan influences Skp1 helix-8 extension via sugar-protein contacts

The glycan-protein contacts described previously for *D. discoideum* Skp1 correlated with the extension of helix-8, which was interpreted to provide better access by FBPs. To address whether this mechanism is conserved for *T. gondii*, despite the difference of the fourth sugar, the computational studies were repeated on glycosylated TgSkp1. Energy-minimized structures of the glycans from DdSkp1 and TgSkp1 revealed only a difference in the position of O4 of the fourth sugar (due to the Glc/Gal configurational inversion) (Fig. S14A). Six all-atom MD simulations were performed without coordinate constraints for 250 ns each to allow greater sampling of conformational space. Three of the simulations began with a 50-ns pre-equilibration of the glycan with C α atom constraints on the polypeptide while the others proceeded directly. As before, the simulations did not converge on a common structure, so a combined time-resolved linear regression analysis of the six simulations was conducted to identify correlations between helix extension and calculated polar and nonpolar interaction energies with each moiety of the glycan, and the results are shown in Table 2.

A representative frame from a simulation with

a strong correlation between the extension of helix-8 (dotted green line) and glycan contacts with the polypeptide chain is shown in Figs. 10A-C). Glycan-protein contacts involve both polar hydrogen bonds and non-polar van der Waals interactions between sugar and amino acid residues. The strongest correlated polar interactions (Table 2) correspond to three hydrogen bonds between sugars and amino acids within the unstructured peptide chain located between helix 7 and 8 (residues 147-152), as depicted in Figs. 10B, 9C. Notably, the 4-OH of α Glc (the fourth sugar), which is epimeric to the *D. discoideum* fourth sugar, is oriented toward the solvent thus not affecting protein contacts. Nonpolar interactions also occur within this region, with the terminal α Gal moiety showing the highest correlation (Table 2). The hydrophobic face of the terminal α Gal can pack against a nonpolar pocket consisting of planar faces of peptide backbone (Figs. 10D, 9E). Specifically, its C3-C5 hydrogens can pack with the backbone of residues 147-149, and a hydrogen on C6 can be buried within residues 143 and 144 (Fig. 10D). Both the polar and nonpolar interactions identified here involve amino acids that are conserved across both organisms, with the exception of V149 which is Lys in DdSkp1 (Fig. S14D). This substitution would not be expected to affect Gal burial since the majority of interactions are with the protein backbone at this position (not shown). The previously noted hydrogen bonds-1 and -3 to helix-7, involving E140 (E129 in DdSkp1), were also observed in this study (Fig. S14C), but were poorly correlated ($R^2=0.0$, vs. 0.63 average for the other three) with helix-8 extension (Fig. S14E). The energetic analyses applied to the *T. gondii* variant confirm its ability to form an organized structure that can still similarly influence local Skp1 polypeptide organization. The new studies emphasize the interaction of the glycan with the interhelix region rather than helix-7 and contributions of packing of the outer sugars including the terminal α Gal. As previously hypothesized, these effects on the conformational ensemble of TgSkp1 have the potential to make it more receptive to FBP binding.

Discussion

Glycosylation mediates much of the effect of O₂-dependent prolyl hydroxylation of Skp1 in *D. discoideum* and *T. gondii*, and the present work emphasizes the importance of the fifth and final

sugar of the glycan. The *T. gondii* glycan adopts a similar constrained conformation as described for *D. discoideum* despite the sequence different (Figs. S14A, B), and new all-atoms molecular dynamics simulations showed a conserved *cis*-interaction with the back side of the intrinsically disordered region of Skp1 that comprises subsite-2 of the FBP binding pocket (Fig. S14C). A deeper time-resolved energetics analysis of polar and non-polar contacts showed a strong dynamic correlation between burial of the peripheral (especially the fifth) sugars in the turn from helix-7 to the loop connecting helix-8, and the extension of helix-8, which exposes subsite-2 for FBP docking (Fig. 10; Table 2). This analysis de-emphasized the previous correlation of hydrogen bonds with helix-7 in favor of contacts with loop residues between helix-7 and helix-8 (Fig. S14E). This quantitative analysis of the trajectories extends critical features of the prior study in *D. discoideum* (7) by emphasizing the role of non-polar packing of the glycan terminus in helix-8 extension.

The fifth sugar is attached by Gat1, an unusual GT that resides in the cytoplasmic compartment rather than the secretory pathway of the cell and appears to be dedicated to only Skp1. While confined to the protist kingdom, it is nevertheless descended from a widely distributed lineage of sugar nucleotide-dependent CAZy GT8 GTs, and may be the evolutionary predecessor of glycogenin, a GT that modulates glycogen formation in the cytoplasm of yeast, fungi and animals. Though Gat1s from *T. gondii* and *P. ultimum* catalyze the transfer of α Gal from UDP-Gal to non-reducing terminal Glc acceptors (Figs. 4, S9), their ability to utilize UDP-Glc at low efficiency could be rationalized by the organization of the active site in the crystal structure (Fig. 8) and anticipates this evolutionary transition. The ability of UDP-Glc to inhibit Gat1's galactosyltransferase activity (Fig. 4G) might be regulatory in cells. The recently reported capability of glycogenin to transfer Gal at certain steps (44) lends further support to the evolutionary relationship. Though Gat1 is able to modify Glc in multiple contexts, it prefers Glc at the terminus of the native Skp1 tetrasaccharide over Glc in a native starch- or glycogen-like glycan (Figs. 4, S9). Furthermore, Gat1 was substantially more reactive when the glycan was attached to Skp1 (Fig. 4H, I), indicating that the apoprotein contributes to increased activity. The increase in

activity may be due to the recognition of Skp1 protein by Gat1, but this hypothesis remains to be tested.

Computational docking showed how the Skp1 glycan alone in its calculated lowest energy state can naturally fit within a groove along the dimer interface of PuGat1 (Fig. 9A), in a manner that cannot be achieved with an α 4-glucan as in starch or glycogen in its lowest energy state (not shown). The orientation of the acceptor Glc of the Skp1 glycan in this docking mode also supports the formation of the α 1,3-linkage that was determined by NMR analysis of the product of the reaction of both PuGat1 and TgGat1 with a synthetic version of the Skp1 tetrasaccharide as an acceptor (Fig. 6). Furthermore, a high degree of conservation between the two enzymes is supported by their very similar activity and specificity characteristics (Figs. 4, S9). The described binding mode is a possible explanation for the greater activity of the free tetrasaccharide as an acceptor substrate compared to the α 4-linked glucans. Overall, the data suggest that recognition of Skp1 as a substrate includes both active site preference for the specific glycan and separate determinants on the polypeptide. Owing to the limited computational scope of this study, however, further analysis is needed to fully define the basis for Skp1 recognition.

The significance of the terminal sugar is echoed by the slow growth phenotype of *gat1* Δ parasites in the monolayer plaque assay. This was attributed 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 no other Gat1 targets were detected by biochemical complementation of *gat1* Δ extracts (Figs. 4J, S9H, I), and no effect on starch formation was observed (Fig. 5), 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 earlier GTs in the pathway and their selectivity for Skp1 *in vitro* (13, 14). Furthermore, in *D. discoideum*, genetic manipulations of Skp1 levels interact with its glycosylation with respect to O₂-sensing (8). Skp1 interactome studies in *D. discoideum* indicated that 5 sugars were clearly more effective in promoting binding of FBPs than 3 sugars (10), but were however unable to resolve the roles of the fourth

and fifth sugars owing to their being added by the same GT (45). Though we cannot completely exclude some other function of Gat1, we note that at 345 amino acids, nearly all of its sequence appears devoted to the enzymatic domain (Fig. 3A). Studies are underway to identify SCF substrates and how they are affected by Skp1 modification and influence parasite growth and fitness.

The importance of the fifth sugar is reinforced by the observation that the amoebozoan, which lack both Glt1 and Gat1 (Figs. 1A, B), evolved a new unrelated enzyme, AgtA, to fulfill the role. AgtA is a dual function GT that applies both the 4th and 5th sugars: each an α 3-linked Gal. This raises the interesting possibility that AgtA evolved to compensate for the unexplained loss of Glt1 and Gat1, with recursive addition of the same sugar being an accessible evolutionary pathway to recover the pentasaccharide. Gat1 sequences are found in alveolates (includes *T. gondii*), stramenopiles (includes *P. ultimum*), and rhizaria (together with telonemids are known as the TSAR group), and archaeoplastids (together with the TSAR group are known as bikonts), but not in the unikonts that include the amoebozoan, fungi and animals (46, 47). While the fate of Glt1 in the unicots is unknown, its disappearance potentially freed Gat1 to evolve into glycogenin, an α 4-glucosyltransferase that, *in vitro*, can autoglucosylate itself to prime glycogen synthesis in fungi and metazoa. The enzymes share catalytic properties (see above) and perfectly conserve structural similarities including the same dimer interface (Figs. 7, 9), though further studies are needed to support the model of lineal descent. Thus it is interesting to speculate whether the proposed transition conserved a cellular function. Though glycogenin acquired a C-terminal glycogen synthase binding domain (Fig. 3A; 48) that seems consistent with its *in vitro* capacity to initiate glycogen assembly, data now exist that glycogenin is not required for glycogen synthesis in yeast and mice (49, 50), though glycogen levels are affected by an unknown mechanism. The *trans* glycosyltransferase activity for this enzyme lineage in bikonts protists raises the possibility of an unanticipated activity in unicots that could help explain the glycogenin-KO findings in yeast and mice.

Experimental Procedures

Maintenance of host cells and parasite manipulations

Cultures of human foreskin fibroblast (HFF, ATCC SCRC-1041) or hTERT HFF (BJ-5ta, ATCC CRL-4001) were maintained Dulbecco's modified Eagle's Medium supplemented with 10% (v/v) fetal bovine serum, 2 mM L-glutamine and 100 units/ml penicillin/streptomycin (Corning) at 37°C in a humidified CO₂ (5%) incubator. Type 1 RH (21), RHΔKu80ΔHXGPRT (RHΔΔ) (22) and type 2 ME49-RFP (23) strains of *Toxoplasma gondii* were cultured on HFF or hTERT HFF monolayers in the same medium as described for host cells except that 1% (v/v) fetal bovine serum was used where stated, and cloned by limiting dilution in 96-well plates. Parasites were kept in media without drug for plaque assays, which were performed as before (20). Transfections were conducted by electroporation using a BioRad Gene Pulser Xcell at 1.5 kV and 25 μF with 2 mm electroporation cuvettes (VWR international) in Cytomix buffer (10 mM KH₂PO₄/K₂HPO₄ (pH 7.6), 120 mM KCl, 0.15 mM CaCl₂, 5 mM MgCl₂, 25 mM HEPES, 2 mM EDTA).

Disruption and complementation of *gat1*

TGGT1_310400 (Toxodb.org), referred to as *gat1* (Fig. S1), was disrupted by two independent approaches. In the first method, a disruption DNA was prepared from the vector pmini-GFP.ht (a gift from Dr. Gustavo Arrizabalaga) in which the *hxgpprt* gene is flanked by multiple cloning sites as described (13). 5'-flank and 3'-flank targeting sequences were PCR amplified from strain RHΔΔ with primer pairs Fa and Ra and pairs Fb and Rb, respectively (Table S1). The 5'-fragment was digested with ApaI and XhoI and cloned into similarly digested pminiGFP.ht. The resulting plasmid was digested with XbaI and NotI and ligated to the similarly digested 3'-flank DNA. The resulting vector was linearized with PacI and electroporated into strain RHΔΔ, selected under 25 μg/ml mycophenolic acid and 25 μg/ml xanthine, and cloned by limiting dilution. Genomic DNA was screened by PCR to identify Tggat1 disruption clones (Fig. S3A), using primers listed in Table S1.

The second approach was based on a double-CRISPR/Cas9 method as previously detailed (20), with minor modifications. To generate the dual guide (DG) plasmid, a fragment of p2 containing

the guide RNA *gat1*-63 expression cassette was PCR amplified using primers plasmid 3 FOR and plasmid 3 REV (Table S1), digested with NsiI, and ligated into the NsiI site of a dephosphorylated p3 containing guide RNA *gat1*-968. The type 1 RH and type 2 ME49 strains were co-transfected with pDG-Gat1 (10 μg) and a dihydroxyfolate reductase (*dhfr*) amplicon (1 μg) by electroporation (Fig. S4A). CRISPR/Cas9-mediated disruption in RHΔΔ parasites was done similarly except that pDG-Gat1 was co transfected with a *dhfr* amplicon containing 45 bp homology arms targeting *gat1* (Fig. S3B). *Gat1*Δ parasites were subsequently selected in 1 μM pyrimethamine (Sigma). The expected replacement of nt 63-968 (relative to A of ATG start codon) with the *dhfr* cassette was confirmed by PCR using primers listed in Table S1.

To complement *gat1*Δ RH parasites, a Tggat1 DNA fragment consisting of the coding sequence of *gat1* plus approximately 1.2 kb each of 5'- and 3'- flanking DNA sequences was generated by PCR from RH genomic DNA using primers Fa and Rb, which contained ApaI and NotI restriction sites respectively. After treatment with ApaI and NotI, the PCR product was cloned into similarly digested pmini-GFP-ht plasmid in place of its *hxgpprt* cassette, to generate pmini-Tggat1. The plasmid was transformed into *E. coli* Top10 cells and purified by using a Monarch miniprep kit (NEB). A Ty tag DNA sequence was inserted at the 3'-end of the *gat1* coding sequence using a Q5 site directed mutagenesis kit (NEB) and Fn and Rn primers, yielding pmini-Tggat1-Ty. The sequence was confirmed using primers Fl and Rl (Table S1). RH *gat1*Δ clone 8 was complemented by co-electroporation with a PCR amplicon from pmini-Tggat1-Ty (1 μg) and a sgUPRT CRISPR/Cas9 plasmid (10 μg) targeting the *uprt* locus using the guide DNA sequence 5'-ggcgctcgattgtgagac (51) (Fig. S4B). Transformants were selected with 10 μM fluorodeoxyuridine (FUdR, Sigma), and drug resistant clones were screened by PCR with primers listed in Table S1.

To complement *gat1* in RHΔΔ parasites, the UPRT Vha1 cDNA shuttle vector containing TgVha1 cDNA (52) was modified to generate a Gat1-HA complementation plasmid using NEB HiFi Builder method. The vector backbone, containing 5'-flank and 3'-flank *uprt* targeting sequences and a Tg-tubulin promoter and 3×HA sequence, was PCR amplified from the shuttle

vector using primers Ft and Rt. The coding sequence of *TgGat1* was PCR amplified from pmini-*Tggat1* plasmid using primers Fu and Ru, which had 18-21 nts complementary to the terminal ends of the vector (Fig. S3C). The gel-purified PCR fragments were incubated with HiFi DNA assembly enzyme mix (NEB) and transformed into *E. coli* Top10 cells, yielding pUPRT*gat1*3xHA. The *Tggat1* sequence was confirmed using primers Fl and Rl. Complementation in a *gat1* Δ clone derived from RH $\Delta\Delta$ was done similarly except that a *gat1*-3xHA PCR amplicon with 5' and 3' *uprt* homology arms was used in the transfection.

Bradyzoite induction

ME49-RFP tachyzoites were differentiated to bradyzoites using alkaline pH (53). HFF monolayers were pre-incubated with sodium bicarbonate free RPMI (Corning) containing 50 mM HEPES-NaOH (pH 8.1) for 24 h, and then infected with tachyzoites and maintained at ambient atmosphere at 37°C with medium replacement every 24 h. Differentiation was monitored by labeling with *Dolichos biflorus* agglutinin (DBA) (54). Infected HFF monolayers formed on 25 mm coverslips were washed with PBS (Corning), fixed with 4% paraformaldehyde in PBS for 10 min, washed with PBS, and permeabilized with 1% Triton X-100 (BioRad) in PBS for 10 min, all at room temperature. Samples were blocked with 3% (w/v) bovine serum albumin (BSA) in PBS for 1 h, incubated for 2 h with 5 μ g/ml FITC-DBA lectin (Vector Laboratories Inc.) in 1% BSA in PBS, and washed with PBS. The coverslips were mounted with ProLong Gold antifade reagent (Invitrogen) on glass slides. The slides were imaged by phase contrast and fluorescence microscopy on a Zeiss Axioskop 2 Mot plus.

Periodic acid staining

To assess amylopectin levels, parasite-infected HFF monolayers on 25 mm glass coverslips were washed with plain PBS, fixed with ice-cold MeOH for 5 min, washed with PBS, and incubated with 1% periodic acid in deionized H₂O in the dark for 10 min (55). The coverslips were washed with deionized H₂O, incubated with Schiff reagent for 15 min, washed once with deionized H₂O, and finally rinsed with running tap water for 10 min. The stained coverslips were dehydrated by sequential immersion in 70% (v/v), 80%, 90% and 100%

EtOH, mounted on glass slides using Permount mounting media (Fisher Scientific), and imaged on an EVOS XL Core microscope (Invitrogen).

Expression and purification of recombinant *TgGat1* and *PuGat1*

The single exon coding sequence of *Tggat1* cDNA was amplified by PCR from RH genomic DNA using primers, Gat1 Fw and Gat1 Rv (Table S1), cloned into PCR4-TOPO TA (Invitrogen), and transformed into *E. coli* Top 10 cells. The plasmid was double digested with BamH1 and Nhe1 to yield the *gat1* fragment that was cloned into similarly digested pET15-TEVi plasmid (Invitrogen), resulting in the original 346 amino acid coding sequence of Gat1 extended at its N-terminus with a His₆-tag and TEV protease cleavage site (MGSSHHHHHSSSGRENLYFQGH-). A similar N-terminal modification of rabbit glycogenin did not significantly alter its enzymatic activity (35). The predicted coding sequence for *Pugat1* was inferred from PYU1_G002535-201 at protists.ensembl.org/*Pythium ultimum*. The coding sequence was codon optimized for *E. coli* expression, chemically synthesized by Norclone Biotech (Ontario, Canada), and inserted in the NdeI and XhoI sites of pET15b-TEV. The expressed protein was extended at its N-terminus with MGSSHHHHHSSGENLYFQGH-.

TgGat1 or *PuGat1* were expressed in and purified from *E. coli* BL21-Gold cells as previously described for *TgGlt1* (14), through the purification on a 5-ml Co⁺² TALON resin column. The eluted protein was dialyzed in 50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 1 mM EDTA and 2 mM β -mercaptoethanol, followed by 25 mM Tris-HCl (pH 8.0), 200 mM NaCl, 1 mM β -mercaptoethanol and 5 mM MnCl₂. The sample was treated with 2 μ M His₆-TEV protease, 5 μ M TCEP in the same buffer overnight at 20°C, and reapplied to another Co⁺² TALON column. The flow-through fraction was dialyzed in 25 mM Tris-HCl (pH 8.0), 50 mM NaCl, 1 mM β -mercaptoethanol, 2 mM MnCl₂. The sample was concentrated by centrifugal ultrafiltration and aliquots were stored at -80°C. As indicated, the preparations were further purified by gel filtration on a Superdex200 column in the same buffer.

SDS-PAGE and Western blotting

Samples were suspended in diluted with

Laemmli sample buffer and typically electrophoresed on a 4–12% gradient SDS-polyacrylamide gel (NuPAGE Novex, Invitrogen). Gels were either stained with Coomassie blue or transferred to a nitrocellulose membrane using an iBlot system (Invitrogen). Blots were typically blocked in 5% non-fat dry milk in Tris-buffered saline and probed with a 1:1000-fold dilution of the antibody of interest in the milk solution, followed by secondary probing with a 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 and analyzed in Adobe Photoshop with no contrast enhancement. For measuring incorporation of radioactivity, 1-mm thick 7–20% acrylamide gels were prepared manually, as detailed (13).

Preparation of Skp1 peptides

To monitor Skp1 glycosylation status, endogenous TgSkp1 was purified from parasite extracts essentially as described (13). Briefly, frozen pelleted tachyzoites (1×10^8) were resuspended in 8 M urea in 50 mM HEPES-NaOH (pH 7.4), incubated on ice for 1 h and at 50°C for 5 mins, and diluted 8-fold in IP buffer (0.2% Nonidet P-40 in 50 mM HEPES-NaOH, pH 7.4). The lysates were centrifuged at 21,000 × g for 20 min at 4°C, and 100 µl of the supernatant (5×10^7 cells) were incubated with 5 µl UOK75 rabbit antibody (affinity-purified anti-TgSkp1) coupled to protein A/G magnetic agarose beads (Pierce, 78609) for 1 h at 4°C. Beads were captured in a DynaMag-2 magnet (Life Technologies) according to the manufacturer's directions, and washed 3× with 50 mM HEPES-NaOH (pH 7.4), 3× with 10 mM Tris-HCl (pH 7.4), 50 mM NaCl, and once with 50 mM NaCl. Bound Skp1 was eluted twice with 60 µl 133 mM triethylamine (TEA, Sequencing Grade, Pierce, 25108) by incubating for 15 mins at RT, and immediately neutralized with 40 µl of 0.2 M acetic acid. The eluted material was pooled and dried under vacuum, reconstituted in 8 M urea, 10 mM Tris HCl (pH 7.4), reduced in 10 mM DTT for 40 min at RT, and alkylated in 50 mM chloroacetamide for 30 min at RT. Samples were then diluted to 2 M urea with 10 mM Tris HCl (pH 7.4), and digested in 10 µg/ml Trypsin Gold (Mass Spectrometry Grade, Promega, V5280) overnight at RT. Excess trypsin was quenched by the addition of 1% (v/v) trifluoroacetic acid (TFA, Pierce, 28904)

on ice for 15 min and centrifuged at 1,800 × g for 15 min at 4°C. The supernatants were adsorbed to C18 pipette tips (Bond Elut OMIX C18, Agilent, A7003100), and eluted in 50% acetonitrile (ACN, Optima™ LC/MS Grade, Fisher Chemical, A955-4), 0.1% formic acid (FA, LC-MS Grade, Pierce, 28905). Eluted peptides were vacuum dried, reconstituted in 40 µl 5% ACN, 0.05% TFA, and 4 µl were analyzed by nLC-MS/MS.

Treatment of TgSkp1 peptides with α-galactosidase

Trypsinates from above were centrifuged at 1800 × g for 15 min at 4°C. The supernatants were dried under vacuum, resuspended in 100 mM sodium citrate phosphate buffer (pH 6.0), and treated with 3.6 mU of green coffee-bean α-galactosidase (CalBiochem) for 18 h at 37°C. An additional 3.6 mU of α-galactosidase was added for an 8 h. After treatment, peptides were processed as above.

Mass spectrometry of TgSkp1 peptides

Reconstituted peptides were loaded onto an Acclaim PepMap C18 trap column (300 µm, 100 Å) in 2% (v/v) ACN, 0.05% (v/v) TFA at 5 µl/min, eluted onto and from an Acclaim PepMap RSLC C18 column (75 µm × 150 mm, 2 µm, 100 Å) with a linear gradient consisting of 4–90% solvent B (solvent A: 0.1% FA; solvent B: 90% ACN, 0.08% (v/v) FA) over 180 min, at a flow rate of 300 nl/min with an Ultimate 3000 RSLC nano UHPLC system, into the ion source of an Orbitrap QE+ mass spectrometer (Thermo Fisher Scientific). The spray voltage was set to 1.9 kV and the heated capillary temperature was set to 280°C. Full MS scans were acquired from *m/z* 350 to 2000 at 70k resolution, and MS² 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 MS² 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. All known glycoforms for TgSkp1 specific glycopeptides were manually searched for and verified. Searches were performed against the *T. gondii* (strain ATCC 50853/GT1) proteome (Uniprot proteome ID

UP000005641, downloaded May 18, 2018; 8,450 entries) plus a list of common keratin, immunoglobulin-derived and trypsin contaminants. Carbamidomethylation of Cys was set as a fixed/static modification and oxidation of Met, deamidation of Asn and Gln residues, and acetylation of protein N-termini were set as variable/dynamic modifications. Searches were performed with trypsin cleavage specificity, allowing two missed cleavage events and a minimum peptide length of 6 residues. All known glycoforms for TgSkp1 specific glycopeptides were manually searched for and verified. The raw files were uploaded to the Figshare server at <https://figshare.com/> with ID 10.6084/m9.figshare.12272882 (Skp1 glycopeptides raw data Fig. S5).

Enzyme assays

Sugar nucleotide hydrolysis: Recombinant TgGat1 (0.625–2.5 μM) was incubated with a given UDP-sugar (50 μM) (Promega) in 50 mM HEPES-NaOH (pH 7.4), 2 mM MnCl₂, 5 mM DTT, in a final volume of 20 μl, for 1–16 h at 37°C. The UDP generated was quantitated using the UDP-Glo assay (Promega) as described (56).

Glycosyltransferase activity toward small glycosides: In the standard reaction, TgGat1 or PgGat1 was incubated with 2 mM synthetic glycosides [pNP-α-galactoside (pNP-αGal), pNP-β-galactoside (pNP-β-Gal), pNP-α-glucoside (pNP-αGlc), pNP-β-glucoside (pNP-βGlc), pNP-α-maltoside, (pNP-malt), chloro-4-nitrophenyl-α-maltotrioside (pNCIP-trimalt), pNP-Skp1 trisaccharide (FGaGn-pNP), pNP-Skp1 tetrasaccharide (GlFGaGn-pNP)], 8 μM UDP-Gal (unlabeled), 0.17 μM UDP-[³H]Gal (15.6 μCi/nmol, American Radiolabeled Chemicals), 50 mM HEPES-NaOH (pH 7.0), 2 mM MnCl₂, 5 mM DTT, in a final volume of 30 μl, for 1 h at 37°C. Pilot studies indicated a pH optimum of 7.0, with 50% activity at pH 8.0 and 75% activity at pH 6.0 for TgGat1. Salt dependence studies showed maximal activity with no added NaCl or KCl, and 35% activity at 800 mM of either salt. Activity showed a ~6-fold preference for MnCl₂ over MgCl₂, with activity maximal at 2 mM MnCl₂. The enzyme was essentially inactive in NiCl₂, CoCl₂ and CaCl₂ (Fig. S9A–D). For kinetic studies, concentrations and times were varied as indicated, and kinetic parameters were analyzed according to

the Michaelis-Menten model based on the least squares fitting method in Graph pad Prism software. Reactions were stopped by addition of 1 ml 1 mM ice-cold Na-EDTA (pH 8.0), and incorporation of radioactivity into pNP-glycosides was analyzed by capture and release from a Sep-Pak C18 cartridge and scintillation counting (14).

Glycosyltransferase activity toward GlFGaGn-Skp1: To prepare Tg-GlFGaGn-Skp1, 2 nmols (40 μg) of recombinant TgSkp1 FGaGn-Skp1 (14) was incubated with 1.3 nmols (88 μg) of Tg-His₆-Glt1, 4 nmols UDP-Glc, 1.2 units alkaline phosphatase (Promega), 50 mM HEPES-NaOH (pH 8.0), 5 mM DTT, 2 mM MnCl₂, 2 mM MgCl₂, 2 mg/ml BSA in a final volume of 121 μl for 3.5 h at 37°C. The reaction was initiated by addition of UDP-Glc and terminated by freezing at -80°C. Reaction progress was monitored by Western blotting with pAb UOK104, which is specific for FGaGn-Skp1 from either *D. discoideum* or *T. gondii* (57), followed by probing with pAb UOK75, which is pan-specific for all TgSkp1 isoforms, for normalization. Approximately 85% of total Skp1 was modified. GlFGaGn-Skp1 was purified from Glt1 on a mini-QAE-column using a Pharmacia Biotech SMART System. ~19 μg of GlFGaGn-Skp1 was applied to a mini QAE column pre-equilibrated with 50 mM Tris-HCl (pH 7.8), 5 mM MgCl₂, 0.1 mM EDTA (buffer A) and eluted with a gradient from 0% A to 100% buffer B (50 mM Tris-HCl (pH 7.8), 5 mM MgCl₂, 0.1 mM EDTA, 300 mM NaCl) in 40 min at a flow rate of 240 μl/min. GlFGaGn-Skp1 fractions were identified based on A₂₈₀ and Western blot probing with pAb UOK75.

TgGat1 (0.127 μM) was incubated with GlFGaGn-Skp1 or FGaGn-Skp1 (0.9–3.6 μM), 3.2 μM UDP-[³H]Gal (15.6 μCi/nmol), 40 μM UDP-Gal (unlabeled), 0.2% (v/v) Tween-20, 50 mM HEPES-NaOH (pH 7.0), 2 mM MnCl₂, 5 mM DTT, in a final volume of 20 μl, for 1 h at 37°C. The reaction was stopped by addition of 4×-Laemmli sample buffer, 1 M DTT (50 mM final concentration), and 2 μg soybean trypsin inhibitor, and boiled for 3 min. SDS-PAGE and incorporation of radioactivity into Skp1 was performed as described above.

To detect Gat1 activity in cells, a cytosolic extract was prepared by hypotonic lysis, ultracentrifugation at 100,000 g × 1 h, and desalted as previously described (13). 25 μg of desalted S100 protein was incubated with 10–50 nmol Tg-

GIFGaGn-Skp1 and 1.0 μ Ci UDP-[3 H]Gal (15.6 μ Ci/nmol) for 5 h, and incorporation of radioactivity into protein was assayed by SDS-PAGE and scintillation counting as described above.

Glycosyltransferase activity toward parasite extracts: To search for Gat1 substrates, cytosolic S100 fractions (180 μ g protein) were incubated with 0.13 μ M TgGat1, 2.0 μ Ci UDP-[3 H]Gal (15.6 μ Ci/nmol) in a final volume of 60 μ l containing 50 mM HEPES-NaOH (pH 7.0), 2 mM MnCl₂, 5 mM DTT, 30 mM NaF, 0.2% Tween-20, at 37°C for 1 h, supplemented with 1.7 μ g Tg-GIFGaGN-Skp1 as indicated. Incorporation of radioactivity was monitored by the SDS-PAGE assay as described above.

Mass spectrometry of Skp1

To evaluate its glycosylation status, recombinant PuGat1 or TgGat1 (purified by gel filtration) were incubated with or without UDP-Gal or UDP-Glc in the absence of added acceptor substrate, and diluted to 50 ng/ μ l Skp1 with 2% acetonitrile, 0.05% (v/v) trifluoroacetic acid. 250–500 ng of protein (5–10 μ l) was injected into an Acclaim PepMap C4 trap cartridge (300 μ m \times 5 mm) equilibrated with 0.05% trifluoroacetic acid, 2% acetonitrile, ramped up with an increasing gradient to 0.1% formic acid, 25% acetonitrile, and introduced into an Acclaim PepMap analytical C4 column (75 μ m \times 15 cm, 5 μ m pore size) maintained at 35°C in an Ultimate 3000 RSLC system coupled to a QE+ Orbitrap mass spectrometer (Thermo Scientific). After equilibrating the analytical column in 98% LC-MS Buffer A (water, 0.1% formic acid) for 10 min and 6-min ramp up to 27% LC-MS Buffer B [90% (v/v) acetonitrile, 0.1% formic acid], separation was achieved using a linear gradient from 27% to 98% Buffer B over 20 min at a flow rate of 300 nl/min. The column was regenerated after each run by maintaining it at 98% Buffer B for 5 min. The effluent was introduced into the mass spectrometer by nanospray ionization in positive ion mode via a stainless-steel emitter with spray voltage set to 1.9 k, capillary temperature set at 250°C and probe heater temperature set at 350°C. The MS method consisted of collecting Full ITMS (MS¹) scans (400–2000 *m/z*) at 140,000 resolution in intact protein mode (default gas P set to 0.2). PuGat1 species eluting between 17.5 and 21.5 min and

TgGat1 species eluting between 18.5 and 22.5 min (~60 to 80% acetonitrile) were processed with Xcalibur Xtract deconvolution software to generate monoisotopic masses from the multicharged, protonated ion series. Since TgGat1 MS spectra were not isotopically resolved, masses were extracted after MS spectra deconvolution using the ReSpect algorithm in the BioPharma Finder suite (Thermo Scientific), with a 20 ppm deconvolution mass tolerance and 25 ppm protein sequence matching mass tolerance. For consistency, PuGat1 was also deconvoluted and re-extracted using the ReSpect algorithm and the same conditions. The data raw files were uploaded to the Figshare server at <https://figshare.com/> with ID 10.6084/m9.figshare.12272909 (Recombinant Gat1 intact protein raw files Fig. S10).

Structure determination of PuGat1

A PuGat1:UDP:Mn²⁺ complex in 50 mM HEPES-NaOH, pH 7.4, 75 mM NaCl, 2 mM DTT, 5 mM UDP, and 5 mM MnCl₂ was crystallized at 20°C 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. Crystals were obtained overnight and were transferred to a reservoir solution containing 15% (v/v) of a cryoprotectant mixture (1:1:1 ethylene glycol:dimethyl sulfoxide:glycerol), and flash cooled with liquid N₂. The complex crystallized in space group P4₂12 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 (58), with 5% of the data omitted for cross validation.

PuGat1:UDP:Mn²⁺ crystals were soaked with platinum cyanide for heavy-atom phasing, cryoprotected, and frozen as above. PuGat1:UDP:Pt²⁺ crystals were isomorphous to PuGat1:UDP:Mn²⁺ crystals and diffracted to 2.1 Å (Table S3). PuGat1:UDP:Mn²⁺ crystals were alternatively soaked in UO₂ (not shown). The crystals, which diffracted to 2.4 Å in the same space group, lacked UDP density, suggesting displacement by UO₂.

The crystal structure of PuGat1:Pt²⁺ was solved using single-wavelength anomalous dispersion (SAD). The data was obtained at a wavelength of 1.85 Å for maximum anomalous signal. A single Pt²⁺ site was located using PHENIX (59), and the

resulting phases had an acceptable figure of merit of 0.31. The model was subjected to iterative cycles of refinement and yielded a final model with $R_{\text{work}}/R_{\text{free}}$ of 0.21/0.24 (Table S3). The structure of PuGat1:UDP:Mn²⁺ was solved using rigid body refinement of PuGat1:Pt²⁺. The resulting model was subjected to iterative cycles of refinement and yielded a final model with $R_{\text{work}}/R_{\text{free}}$ of 0.18/0.21 (Table S3). Images were rendered in PyMol (60), and secondary structures were assigned based on DSSP (61, 62).

Glycan docking

The lowest energy conformation of the TgSkp1 tetrasaccharide ($\text{Glc}\alpha 1,3\text{Fuc}\alpha 1,2\text{Gal}\beta 1,3\text{GlcNAc}1\text{-OH}$) was generated via GLYCAM (63). Hydrogen atoms were added, and the electrostatic surface was generated using AutoDockTools (64). A grid box with dimensions $26 \text{ \AA} \times 26 \text{ \AA} \times 34 \text{ \AA}$ was placed over the ligand binding site based on where the acceptor is bound on glycogenin/glucan complex. The ligand was kept rigid, since AutoDock Vina is not parameterized specific to glycan torsion angles. 100 binding modes were calculated with the lowest binding energy scored at -4.4 kcal/mol and the highest binding energy scored at -5.7 kcal/mol.

Sedimentation velocity studies

PuGat1 was further purified on a Superdex S200 gel filtration column (GE Healthcare) equilibrated with 20 mM potassium phosphate (pH 7.4), 50 mM KCl, 0.5 mM TCEP. Protein concentration was calculated from A_{280} measured in an Agilent 8453 UV/Vis spectrophotometer, based on a molar absorptivity (ϵ_{280}) of 60390 M⁻¹cm⁻¹, which was calculated from the PuGat1 sequence using ProtParam (65). Samples were diluted to 0.3–11 μM, loaded into 12 mm double-sector Epon centerpieces equipped with quartz windows, and equilibrated for 2 h at 20°C in an An60 Ti rotor. Sedimentation velocity data were collected using an Optima XLA analytical ultracentrifuge (Beckman Coulter) at a rotor speed of 50000 RPM at 20°C. Data were recorded at 280 nm for protein samples at 3.5–11 μM, and at 230/220 nm for samples at 0.3–1.5 μM, in radial step sizes of 0.003 cm. SEDNTERP (66) was used to model the partial specific volume of PuGat1 (0.73818 mL/g), and the density (1.0034 g/ml) and viscosity (0.0100757 P) of the buffer. Using SEDFIT (67), data were

modeled as continuous $c(s)$ distributions and were fit using baseline, meniscus, frictional coefficient, and systematic time-invariant and radial-invariant noise. Predicted sedimentation coefficient (s) values for the PuGat1 monomer (2.8 S) and dimer (4.2 S) were calculated using HYDROPRO (68). Data fit and $c(s)$ plots were generated using GUSSI (69).

Molecular dynamics simulations

The model for TgSkp1 was built as described previously for DdSkp1 (7). Briefly, a homology model of TgSkp1 was generated with the SWISS-MODEL web server (70) based on the human Skp1 template from PDB ID: 2ASS (71), and missing residues were appended with UCSF Chimera (72). Molecular dynamics simulations were performed as described previously. Briefly, MD simulations were performed with the pmemd.cuda version of AMBER14 (73). The amino acid and carbohydrate residues were parameterized with the FF12SB and GLYCAM06 (J-1) force fields, respectively (74, 75). The systems were neutralized with Na⁺ ions and solvated using the TIP3P water model (76) in a truncated octahedral box with 15-Å distance from the solute to the end of the unit cell. Electrostatic interactions were treated with the particle mesh-Ewald algorithm, and a cut-off for non-bonded interactions was set to 8 Å (77). SHAKE was employed to constrain hydrogen-containing bonds, enabling an integration time step of 2 fs. Restraints were imposed in specific situations and were enforced with a 10-kcal/mol Å² energy barrier in each case. Each minimization step consisted of 1000 cycles of the steepest descent method (1000 cycles), followed by 24,000 cycles using the conjugate gradient approach. The systems were heated to 300 °K under NVT conditions over 60 ps, employing the Berendsen thermostat with a coupling time constant of 1 ps. The subsequent simulations were performed under NPT conditions. A torsion term that corrects 4(*trans*)-hydroxyproline residue (Hyp) ring puckering was included in simulations of the O-linked residue type (OLP) based on previous studies that indicate that the ring is primarily exo when glycosylated (78, 79). This torsion term has been adopted in GLYCAM06 (version K).

A 50-ns simulation of the protein was performed with Cα cartesian constraints on all amino acids except those generated by Chimera.

The fully glycosylated isoform was created by adding the TgSkp1 pentasaccharide to the exo-pucker conformation of hydroxyproline (residue 154). Six independent simulations were performed. Three ran for 250 ns directly, while the other three began with an additional 50 ns in which the protein was restrained to allow the glycan time to adapt to the protein conformation.

Computational Analysis

Structural images were created with Visual Molecular Dynamics (80) and the 3D-SNFG plugin (81). The structure depicted in Fig. 7 was created by identifying the frame from equil-1 that consisted of the lowest RMSD to the average structure as calculated by cpptraj (82). The cpptraj program was also used to distribute the latter 200 ns of the six simulations into 48 bins containing 250 frames each for analysis. Per-residue MMGBSA energies were calculated with MMPBSA.py.MPI with $\text{igb}=2$ and $\text{idecomp}=3$. A bash script was used to calculate the correlation coefficients (83).

Phylogenetic analysis of enzyme sequences

Proteins related to TgGat1 were searched for using a BLASTP (**V 2.4.0**) search seeded with the full-length TgGat1 protein sequence against the NCBI non-redundant database (December 2016). The evolutionary relationship of Gat1-like sequences was investigated by using a Maximum Likelihood method (84) and conducted in MEGA7 (85). Catalytic domains from 43 CAZY GT8 sequences selected based on their relatedness to Gat1, glycogenin, or known function, and consisted of 196 positions. 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).

Data availability

All atomic coordinates and structure factors were deposited in the RCSB Protein Data Bank (PDB) under accession codes 6MW5 (PuGat1 Pt²⁺ derivative) and 6MW8 (PuGat1 native).

The raw data files for mass spectrometry were uploaded to the Figshare server at <https://figshare.com/> with ID 10.6084/m9.figshare.12272882 (Skp1 glycopeptides raw data Fig. S5) and ID 10.6084/m9.figshare.12272909 (Recombinant Gat1 intact protein raw files Fig. S10).

Acknowledgments

We thank Kentuan Hicks, Nathan Beattie, and Safal Shrestha for their assistance.

Funding and Additional Information

MM was partially supported by NIH T32-AI060546; HWK was partially supported by NIH T32-GM107004; and KH was supported by 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 JHP), NIH R01-GM114298 to ZAW, and NIH S10-OD021762 to John Rose. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing Interests

The authors declare that they have no conflicts of interest with the contents of this article.

Supporting Information

This section contains Tables S1-S3 and Figures S1-S14 as a single pdf document.

References

1. Islam, M. S., Leissing, T. M. , Chowdhury, R., Hopkinson, R. J., and Schofield, C. J. (2018) 2-oxoglutarate-dependent oxygenases. *Annu. Rev. Biochem.* **87**, 585-620
2. Pugh, C. W., and Ratcliffe, P. J. (2017) New horizons in hypoxia signaling pathways. *Exp. Cell Res.* **356**, 116-121
3. West, C. M., and Blader, I. J. (2015) Oxygen sensing by protozoans: how they catch their breath. *Curr. Opin. Microbiol.* **26**, 41-47
4. Willems, A. R., Schwab, M., and Tyers, M. (2004) A hitchhiker's guide to the cullin ubiquitin ligases: SCF and its kin. *Biochim. Biophys. Acta* **1695**, 133–170
5. West, C. M., van der Wel, H., and Wang, Z. A. (2007) Prolyl 4-hydroxylase-1 mediates O₂ signaling during development of *Dictyostelium*. *Development* **134**, 3349–3358
6. West, C. M., and Kim, H. W. (2019) Nucleocytoplasmic O-glycosylation in protists. *Curr. Opin. Struct. Biol.* **56**, 204-212
7. Sheikh, M. O., Thieker, D., Chalmers, G., Schafer, C. M., Ishihara, M., Azadi, P., Woods, R. J., Glushka, J. N., Bendiak, B., Prestegard, J. H., and West, C. M. (2017) O₂ sensing-associated glycosylation exposes the F-box-combining site of the *Dictyostelium* Skp1 subunit in E3 ubiquitin ligases. *J. Biol. Chem.* **292**, 18897-18915
8. Wang, Z. A., Singh, D., van der Wel, H., and West, C. M. (2011) Prolyl hydroxylation- and glycosylation-dependent functions of Skp1 in O₂-regulated development of *Dictyostelium*. *Dev. Biol.* **349**, 283-295
9. Sheikh, M. O., Schafer, C. M., Powell, J. T., Rodgers, K. K., Mooers, B. H., and West, C. M. (2014) Glycosylation of Skp1 affects its conformation and promotes binding to a model F-box protein. *Biochemistry* **53**, 1657-1669
10. Sheikh, M. O., Xu, Y., van der Wel, H., Walden, P., Hartson, S. D., and West, C. M. (2015) Glycosylation of Skp1 promotes formation of Skp1/cullin-1/F-box protein complexes in *Dictyostelium*. *Mol. Cell. Proteomics* **14**, 66-80
11. Halonen, S. K., and Weiss, L. M. (2013) Toxoplasmosis. *Handb. Clin. Neurol.* **114**, 125-145
12. Xu, Y., Brown, K. M., Wang, Z. A., van der Wel, H., Teygong, C., Zhang, D., Blader, I. J., and West, C. M. (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
13. Rahman, K., Zhao, P., Mandalasi, M., van der Wel, H., Wells, L., Blader, I. J., and West, C. M. (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
14. Rahman, K., Mandalasi, M., Zhao, P., Sheikh, M. O., Taujale, R., Kim, H. W., van der Wel, H., Matta, K., Kannan, N., Glushka, J. N., Wells, L., and West, C. M. (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

15. Lévesque, C. A., Brouwer, H., Cano, L., Hamilton, J. P., Holt, C., Huitema, E., Raffaele, S., Robideau, G. P., Thines, M., Win, J., Zerillo, M. M., Beakes, G. W., Boore, J. L., Busam, D., Dumas, B., Ferriera, S., Fuerstenberg, S. I., Gachon, C. M., Gaulin, E., Govers, F., Grenville-Briggs, L., Horner, N., Hostetler, J., Jiang, R. H., Johnson, J., Krajaejun, T., Lin, H., Meijer, H. J., Moore, B., Morris, P., Phuntmart, V., Puiu, D., Shetty, J., Stajich, J. E., Tripathy, S., Wawra, S., van West, P., Whitty, B. R., Coutinho, P. M., Henrissat, B., Martin, F., Thomas, P. D., Tyler, B. M., De Vries, R. P., Kamoun, S., Yandell, M., Tisserat, N., and Buell, C. R. (2010) Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biol.* **11**, 1-22
16. Kamoun, S., Furzer, O., Jones, J. D., Judelson, H. S., Ali, G. S., Dalio, R. J., Roy, S. G., Schena, L., Zambounis, A., Panabières, F., Cahill, D., Ruocco, M., Figueiredo, A., Chen, X. R., Hulvey, J., Stam, R., Lamour, K., Gijzen, M., Tyler, B. M., Grünwald, N. J., Mukhtar, M. S., Tomé, D. F., Tör, M., Van Den Ackerveken, G., McDowell, J., Daayf, F., Fry, W. E., Lindqvist-Kreuze, H., Meijer, H. J., Petre, B., Ristaino, J., Yoshida, K., Birch, P. R., and Govers, F. (2015) The Top 10 oomycete pathogens in molecular plant pathology. *Mol. Plant Pathol.* **16**, 413-434
17. Gaastra, W., Lipman, L. J., De Cock, A. W., Exel, T. K., Pegge, R. B., Scheurwater, J., Vilela, R., and Mendoza, L. (2010) *Pythium insidiosum*: an overview. *Vet. Microbiol.* **146**, 1-16
18. van der Wel, H., Gas-Pascual, E., and West, C. M. (2019) Skp1 isoforms are differentially modified by a dual function prolyl 4-hydroxylase/N-acetylglucosaminyltransferase in a plant pathogen. *Glycobiology* **29**, 705-714
19. Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J. J., Schnaar, R. L., Freeze, H. H., Marth, J. D., Bertozzi, C. R., Etzler, M. E., Frank, M., Vliegenthart, J. F., Lutteke, T., Perez, S., Bolton, E., Rudd, P., Paulson, J., Kanehisa, M., Toukach, P., Aoki-Kinoshita, K. F., Dell, A., Narimatsu, H., York, W., Taniguchi, N., and Kornfeld, S. (2015) Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25**, 1323-1324
20. Gas-Pascual, E., Ichikawa, H. T., Sheikh, M. O., Serji, M. I., Deng, B., Mandalasi, M., Bandini, G., Samuelson, J., Wells, L., and West, C. M. (2019) CRISPR/Cas9 and glycomics tools for *Toxoplasma* glycobiology. *J. Biol. Chem.* **294**, 1104-1125
21. Sabin, A. B. (1941) Toxoplasmic encephalitis in children. *J. Am. Med. Assoc.* **116**, 801-807
22. Fox, B. A., Ristuccia, J. G., Gigley, J. P., and Bzik, D. J. (2009) Efficient gene replacements in *Toxoplasma gondii* strains deficient for nonhomologous end joining. *Eukaryot. Cell* **8**, 520-529
23. Grainger, J. R., Wohlfert, E. A., Fuss, I. J., Bouladoux, N., Askenase, M. H., Legrand, F., Koo, L. Y., Brenchley, J. M., Fraser, I. D., and Belkaid, Y. (2013) Inflammatory monocytes regulate pathologic responses to commensals during acute gastrointestinal infection. *Nat. Med.* **19**, 713-721
24. Blader, I. J., Coleman, B. I., Chen, C. T., and Gubbels, M. J. (2015) Lytic cycle of *Toxoplasma gondii*: 15 years later. *Annu. Rev. Microbiol.* **69**, 463-485
25. Bisio, H., and Soldati-Favre, D. (2019) Signaling cascades governing entry into and exit from host cells by *Toxoplasma gondii*. *Annu. Rev. Microbiol.* **73**, 579-599
26. Ball, S., Colleoni, C., Cenci, U., Raj, J. N., and Tirtiaux, C. (2011) The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis.

27. Coppin, A., Dzierszinski, F., Legrand, S., Mortuaire, M., Ferguson, D., and Tomavo, S. (2003) Developmentally regulated biosynthesis of carbohydrate and storage polysaccharide during differentiation and tissue cyst formation in *Toxoplasma gondii*. *Biochimie* **85**, 353-361
28. Sugi, T., Tu, V., Ma, Y., Tomita, T., and Weiss, L. M. (2017) *Toxoplasma gondii* requires glycogen phosphorylase for balancing amylopectin storage and for efficient production of brain cysts. *MBio* **8**(4)
29. Coppin, A., Varre, J. S., Lienard, L., Dauvillee, D., Guerardel, Y., Soyer-Gobillard, M. O., Buleon, A., Ball, S., and 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
30. Pancha, I., Shima, H., Higashitani, N., Igarashi, K., Higashitani, A., Tanaka, K., and Imamura, S. (2019) Target of rapamycin-signaling modulates starch accumulation via glycogenin phosphorylation status in the unicellular red alga *Cyanidioschyzon merolae*. *Plant J.* **97**, 485-499
31. Pancha, I., Tanaka, K., and Imamura, S. (2019) Overexpression of a glycogenin, CmGLG2, enhances floridean starch accumulation in the red alga *Cyanidioschyzon merolae*. *Plant. Signal. Behav.* **14**, 1596718
32. Slabinski, L., Jaroszewski, L., Rychlewski, L., Wilson, I. A., Lesley S. A., Godzik, A. (2007) XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **23**, 3403-3405
33. Lomako, J., Lomako, W. M., and Whelan, W. J. (1988) A self-glucosylating protein is the primer for rabbit muscle glycogen biosynthesis. *FASEB J.* **2**, 3097-3103
34. de Paula, R. M., Wilson, W. A., Roach, P. J., Terenzi, H. F., and Bertolini, M. C. (2005) Biochemical characterization of *Neurospora crassa* glycogenin (GNN), the self-glucosylating initiator of glycogen synthesis. *FEBS Lett.* **579**, 2208-2214
35. Cao, Y., Steinrauf, L. K., and Roach, P. J. (1995) Mechanism of glycogenin self-glucosylation. *Arch. Biochem. Biophys.* **319**, 293-298
36. Alonso, M. D., Lomako, J., Lomako, W. M., Whelan, W. J., and Preiss, J. (1994) Properties of carbohydrate-free recombinant glycogenin expressed in an *Escherichia coli* mutant lacking UDP-glucose pyrophosphorylase activity. *FEBS Lett.* **352**, 222-226
37. Guérardel, Y., Leleu, D., Coppin, A., Liénard, L., Slomianny, C., Strecker, G., Ball, S., and 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
38. Jansson, P. E., Stenutz, R., and 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
39. Moremen, K. W., and Haltiwanger, R. S. (2019) Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. *Nat. Chem. Biol.* **15**, 853-864
40. Gibbons, B. J., Roach, P. J., and Hurley, T. D. (2002) Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *J. Mol. Biol.* **319**, 463-477

41. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774-797
42. Bazan, S., Issoglio, F. M., Carrizo, M. E., and Curtino, J. A. (2008) The intramolecular autoglucosylation of monomeric glycogenin. *Biochem. Biophys. Res. Comm.* **371**, 328-332
43. Bourne, Y., and Henrissat, B. (2001) Glycoside hydrolases and glycosyltransferases: families and functional modules. *Curr. Opin. Struct. Biol.* **11**, 593-600
44. Bilyard, M. K., Bailey, H. J., Raich, L., Gafitescu, M. A., Machida, T., Iglesias-Fernández, J., Lee, S. S., Spicer, C. D., Rovira, C., Yue, W. W., and Davis, B. G. (2018) Palladium-mediated enzyme activation suggests multiphase initiation of glycogenesis. *Nature* **563**, 235-240
45. Schafer, C. M., Sheikh, M. O., Zhang, D., and West, C. M. (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
46. Burki, F., Roger, A. J., Brown, M. W., and Simpson, A. G. B. (2020) The new tree of eukaryotes. *Trends Ecol. Evol.* **35**, 43-55
47. Brown, M. W., Heiss, A. A., Kamikawa, R., Inagaki, Y., Yabuki, A., Tice, A. K., Shiratori, T., Ishida, K. I., Hashimoto, T., Simpson, A. G. B., and Roger, A. J. (2018) Phylogenomics places orphan protistan lineages in a novel eukaryotic super-group. *Genome Biol. Evol.* **10**, 427-433
48. Zeqiraj, E., Tang, X., Hunter R. W., García-Rocha, M., Judd, A., Deak, M., von Wilamowitz-Moellendorff, A., Kurinov, I., Guinovart, J. J., Tyers, M., Sakamoto, K., and Sicheri, F. (2014) Structural basis for the recruitment of glycogen synthase by glycogenin. *Proc. Natl. Acad. Sci. USA* **111**, E2831-2840
49. Torija, M. J., Novo, M., Lemassu, A., Wilson, W., Roach, P. J., François, J., and Parrou, J. L. (2005) Glycogen synthesis in the absence of glycogenin in the yeast *Saccharomyces cerevisiae*. *FEBS Lett.* **579**, 3999-4004
50. Testoni, G., Duran, J., García-Rocha, M., Vilaplana, F., Serrano, A. L., Sebastián, D., López-Soldado, I., Sullivan, M. A., Slebe, F., Vilaseca, M., Muñoz-Cánoves, P., and Guinovart, J. J. (2017) Lack of glycogenin causes glycogen accumulation and muscle function impairment. *Cell Metab.* **26**, 256-266
51. Shen, B., Brown, K., Long, S., and Sibley, L. D. (2017) Development of CRISPR/Cas9 for efficient genome editing in *Toxoplasma gondii*. *Methods Mol. Biol.* **1498**, 79-103
52. Stasic, A. J., Chasen, N. M., Dykes, E. J., Vella, S. A., Asady, B., Starai, V. J., and Moreno, S. N. J. (2019) The *Toxoplasma* vacuolar H⁺-ATPase regulates intracellular pH and impacts the maturation of essential secretory proteins. *Cell Reports* **27**, 2132-2146
53. Soête, M., Camus, D., and Dubremetz, J. F. (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
54. Zhang, Y. W., Halonen, S. K., Ma, Y. F., Wittner, M., and Weiss, L. M. (2001) Initial characterization of CST1, a *Toxoplasma gondii* cyst wall glycoprotein. *Infect. Immun.* **69**, 501-507
55. Dubey, J. P., Lindsay, D. S., and Speer, C. A. (1998) Structures of *Toxoplasma gondii* tachyzoites,

- bradyzoites, and sporozoites and biology and development of tissue cysts. *Clin. Microbiol. Rev.* **11**, 267–299
56. Sheikh, M. O., Halmo, S. M., Patel, S., Middleton, D., Takeuchi, H., Schafer, C. M., West, C. M., Haltiwanger, R. S., Avci, F. Y., Moremen, K. W., and Wells, L. (2017) Rapid screening of sugar-nucleotide donor specificities of putative glycosyltransferases. *Glycobiology* **27**, 206-212
57. Chinoy ZS, Schafer CM, West CM, and Boons GJ (2015) Chemical synthesis of a glycopeptide derived from Skp1 for probing protein specific glycosylation. *Chemistry* **21**, 11779-11787
58. Kabsch, W. (2010) XDS. *Acta Cryst. D* **66**, 125-132
59. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D* **66**, 213-221
60. DeLano, W. L. (2002) PyMOL. DeLano Scientific, San Carlos, CA, 700
61. Touw, W. G., Baakman, C., Black, J., te Beek, T. A. H., Krieger, E., Joosten, R. P., and Vriend, G. (2015) A series of PDB related databases for everyday needs. *Nuc. Ac. Res.* **43**, D364-D368
62. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637
63. Woods, R. J. (2014) GLYCAM Web, Complex Carbohydrate Research Center, University of Georgia, Athens, GA
64. 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 flexibility. *J. Comput. Chem.* **16**, 2785-2791
65. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and 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
66. Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (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
67. Schuck, P. (2000) Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606-1619
68. Ortega, A., Amorós, D., and García de la Torre, J. (2011) .Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* **101**, 892-898
69. Brautigam, C. A. (2015) Calculations and publication-quality illustrations for analytical ultracentrifugation data. *Meths. Enzymol.* **562**, 109-133
70. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: a web-

based environment for protein structure homology modelling. *Bioinformatics* **22**, 195-201

71. Hao, B., Zheng, N., Schulman, B. A., Wu, G., Miller, J. J., Pagano, M., and Pavletich, N. P. (2005) Structural basis of the Cks1-dependent recognition of p27Kip1 by the SCFSkp2 ubiquitin ligase. *Molec. Cell* **20**, 9-19
72. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612
73. Götz, A.W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **8**, 1542–1555
74. Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr, Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688
75. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C.R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **29**, 622–655
76. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935
77. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Physics* **98**, 10089–10092
78. Park, S., Radmer, R. J., Klein, T. E., and Pande, V. S. (2005) A new set of molecular mechanics parameters for hydroxyproline and its use in molecular dynamics simulations of collagen-like peptides. *J. Comput. Chem.* **26**, 1612–1616
79. Owens, N. W., Braun, C., O'Neil, J. D., Marat, K., and Schweizer, F. (2007) Effects of glycosylation of (2 S, 4 R)-4-hydroxyproline on the conformation, kinetics, and thermodynamics of prolyl amide isomerization. *J. Am. Chem. Soc.* **129**, 11670–11671
80. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Molecular Graphics* **14**, 33-38
81. Thieker, D. F., Hadden, J. A., Schulten, K., and Woods, R. J. (2016) 3D implementation of the symbol nomenclature for graphical representation of glycans. *Glycobiology* **26**, 786-787
82. Roe, D. R., and Cheatham, III, T. E. (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084-3095
83. Miller, III, B. R., McGee, Jr., T. D., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012) MMPBSA.py: an efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* **8**, 3314-3321
84. Le, S. Q., and Gascuel, O. (2008) An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307-1320

85. Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870-1874
86. Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality. *Science* **336**, 1030-1033
87. Laskowski, R. A., and Swindells, M. B. (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model* **51**, 2778-2786
88. West, C. M., van der Wel, H., and Gaucher, E. A. (2002) Complex glycosylation of Skp1 in *Dictyostelium*: implications for the modification of other eukaryotic cytoplasmic and nuclear proteins. *Glycobiology* **12**, 17R-27R

Footnotes

¹ The abbreviations used are: DBA, *Dolichos biflorus* lectin; Dd, *Dictyostelium discoideum*; FBP, F-box protein; GaGIFGaGn-, Gal α 1,3Glc α 1,3Fuca1,2Gal β 1,3GlcNAc α 1-; 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*; RH $\Delta\Delta$, RH Δ ku80 Δ hxgprt; SCF, Skp1/Cullin-1/F-box protein subcomplex of E3 Cullin-RING-1 ubiquitin ligases; Tg, *Toxoplasma gondii*; Ub, ubiquitin

Table 1. Strains employed in this study

Description	Name	Parental Strain	Genotype	Gene Targeted	Selection marker	Reference
RH (type 1)			WT			21
$\Delta gat1$ /RH	MM12, cl.A8	RH	$\Delta gat1$	$gat1$	DHFR	this report
$gat1^+/\Delta gat1$ /RH	MM21, cl.E12	MM21	$\Delta uprt/gat1::gat1$ -Ty	$uprt$	$\Delta uprt$	this report
RH $\Delta\Delta$		RH	$\Delta ku80/\Delta hxgppt$			22
$\Delta gat1$ -1/RH $\Delta\Delta$	KR10, cl.A1	RH $\Delta\Delta$	$\Delta ku80/\Delta gat1$	$gat1$	HXGPRT	this report
$\Delta gat1$ -2/RH $\Delta\Delta$	MM16, cl.B7	RH $\Delta\Delta$	$\Delta ku80/\Delta hxgppt/\Delta gat1$	$gat1$	DHFR	this report
$gat1^+/\Delta gat1$ -2/RH $\Delta\Delta$	MM24, cl.G10	MM16	$\Delta ku80/\Delta hxgppt/\Delta uprt/tub::gat1$ -3xHA	$uprt$	$\Delta uprt$	this report
Me49-RFP (type 2)	MM8, cl.A10*	Me49	rfp+			23
$\Delta gat1$ /Me49-RFP	MM14, cl.B5	Me49-RFP	$\Delta gat1$	$gat1$	DHFR	this report

*high RFP expression level clone from original source

Table 2. Table of MMGBSA-derived per-residue energies between the protein and glycan that exhibit a strong correlation with helix extension (distance in Fig. 7A) according to a linear regression analysis of 48 bins from the six MD simulations (Fig. S14E). Polar energies represent a sum of the electrostatic and polar solvation energies while the nonpolar energies are composed of the van der Waals and nonpolar solvation energies. Only interactions with an average polar/non-polar energy less than -0.5 were considered.

Glycan	Protein	Avg. Energy*	R ²
Polar	α Gal	N147	-0.6
	Glc	N150	-0.6
	Fuc	N150	-1.2
	α Gal	N147	-0.81
Non-Polar	GlcNAc	F152	-1.5
	β Gal	F152	-1.1
	GlcNAc	N150	-0.65
	α Gal	I148	-0.67

*kcal/mol

Figure Legends

Fig. 1. Gat1 is required for terminal α -galactosylation of Skp1 in parasites. *A, B*, Schematic of Skp1 glycosylation pathway in *Toxoplasma* (14, and herein) and *Dictyostelium* (3), which modifies a single hydroxyproline associated with its F-box binding region, is depicted using CFG glycan symbols (19). The identity of Gat1 as the final enzyme in the *Toxoplasma* pathway and the nature of the final sugar are reported here. *C*, Dependence of the terminal sugar on Gat1, and its characterization. Skp1 was immunoprecipitated from type 1 RH and type 2 ME49 strains and their genetic derivatives, and peptides generated by trypsinization were analyzed by nLC-MS. In addition, Skp1 from RH was treated with green coffee bean α -galactosidase after trypsinization. The values represent the levels of pentasaccharide-peptide and tetrasaccharide-peptide levels detected, after normalization to all detected modification states of the peptide; note that the values have only partial relative meaning because of the low and varied detection efficiency of glycopeptides. As indicated, the detection threshold was <0.0005. The open circle for the terminal sugar of the pentasaccharide indicates that it was only known as a hexose at the time of the experiment. See Table

S2 and Fig. S5 for primary data. Similar results were obtained in independent samples from RH and $\Delta gat1$ /RH (not shown).

Fig. 2. Parasite growth depends 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 RH $\Delta\Delta$, *gat1*-2 Δ /RH $\Delta\Delta$, and the latter complemented with *gat1* under control of a tubulin promoter in the uprt locus. *C*, Comparison of *gat1*-1 Δ /RH $\Delta\Delta$, prepared by homologous recombination. Significance of differences in plaque areas between parasite strains was assessed by Student's *t*-test.

Fig. 3. Gat1 is closely related to glycogenin 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. Gat1 preferentially galactosylates the Skp1 glycan *in vitro*. *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 μ M UDP-Gal or UDP-Glc toward 20 mM Glc α 1,4Glc α -pNP (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 μ M 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 GlcGaGn-Skp1, in the presence of 40 μ M UDP-Gal (1 μ Ci) 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 GlcGaGn-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 appears unaffected 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 μ m. Two independent trials yielded similar results.

Fig. 6. Gat1 assembles a Gal α 1,3Glc linkage on the Skp1 tetrasaccharide. NMR Analysis of the TgSkp1 pentasaccharide. [1- 13 C]Glc α 1,3Fuc α 1,2Gal- β 1,3GlcNAc α 1-pNP was partially (30%) modified by TgGat1 or PuGat1 in the presence of UDP-[U- 13 C]Gal. *A*, 1D 600 MHz 1 H-NMR spectrum. Magnification shows the region of 13 C-anomeric carbons in the mixture of modified and unmodified tetrasaccharide. A cartoon diagram of the TgSkp1 pentasaccharide attached to pNP (Gal α 1,3Glc α 1,3Fuc α 1,2Gal- β 1,3GlcNAc α 1-pNP) is shown at the top using CFG glycan symbols (19). *B*, 1 H- 13 C-HSQC, -HMBC, and -HSQC TOCSY spectra demonstrating anomeric carbon to ring proton correlations. The Gal-H1/C1 doublet peaks were too weak to observe and are indicated by boxes. *C*, 1 H- 1 H-COSY and 1 H- 13 C-HMBC spectra. Identical results were obtained using [1- 13 C]GIFGaGn-pNP modified by PuGat1 (not shown). See Fig. S11 for a listing of chemical shift values.

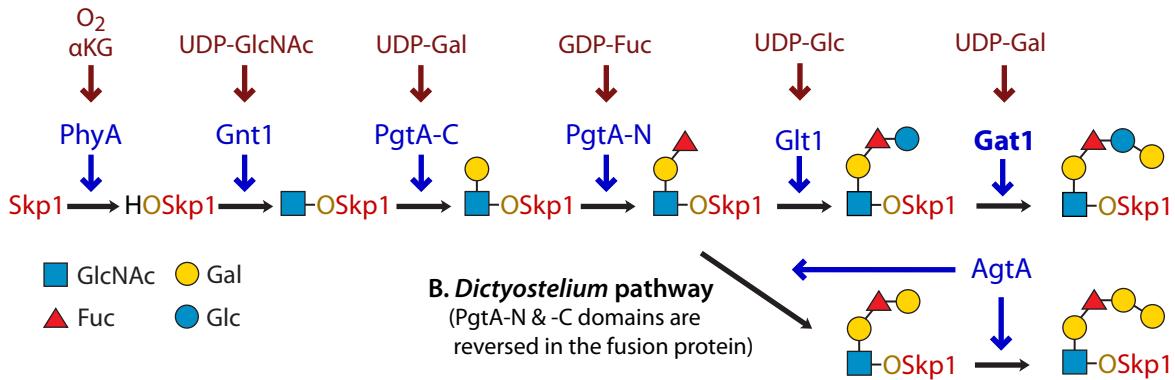
Fig. 7. Gat1 is structurally related to glycogenin. *A*, The asymmetric unit of crystallized Gat1 from *P. ultimum* in complex with UDP and Mn $^{2+}$. α -helices are in red, β -strands in yellow, loops in green, and the active site with bound ligand is boxed. *B*, PuGat1 forms a homodimer that closely superimposes on Oc-glycogenin-1 (PDB entry 1LL2). The cylinders represent α -helices, the arrows represent β -sheets, and the red ellipse marks the two-fold symmetry axis perpendicular to the page. *C*, Sedimentation velocity data modeled as a continuous *c(s)* distribution (normalized to 1.0) yields an *S*-value for 3.5 μ M PuGat1 that is close to the predicted value for a stable dimer in solution. Fig. S12 shows that the dimer is stable down to at least 0.3 μ M. *D*, UDP is coordinated in near identical fashion to that of glycogenin-1, based on the difference density map (F_o-F_c) that was contoured at 5 σ , calculated after omitting UDP and Mn $^{2+}$ and subjecting the structure to simulated annealing. Octahedral coordination of Mn $^{2+}$ is satisfied by the DxD motif, His231, and UDP. The comparison with glycogenin-1 is illustrated in Fig. S13.

Fig. 8. Active site geometry explains Gat1's preference for UDP-Gal rather than UDP-Glc. Comparison of the sugar binding pockets of PuGat1 and Oc-glycogenin-1 displayed as wall-eyed stereoview. *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, the yellow dashes represent hydrogen bonds, and water molecules are represented by blue spheres.

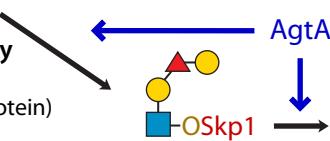
Fig. 9. Computational docking explains specificity of Gat1 for the Skp1-tetrasaccharide. *A*, The top docking pose of the Skp1-tetrasaccharide in the PuGat1 active site and a groove formed by the dimer. *B*, 90° turn of the image shown in panel A. *C*, Hydrogen bonding interactions of the glycan. *D*, Hydrophobic packing of the sugar faces and fucose-methyl (in sticks) with non-polar surfaces (sticks/dots) of Gat1 subunit A (green) and B (cyan).

Fig. 10. Packing of the glycan with Skp1 can explain F-box binding site conformation. *T. gondii* GaGIFGaGn-Skp1 was subjected to six 250-ns all-atoms molecular dynamics simulations. *A*, A frame representative of the glycan-protein interaction and associated helix-8 extension, from a simulation (Equil-1, see Fig. S14E) in which the glycan was pre-equilibrated for 50 ns prior to the start of the simulation. The dotted green line refers to the distance from C-terminus to the center of mass of residues 1-136, and ranged from 18 to 61 Å. *B*, Zoom-in of panel A depicting the glycan (C-atoms in green) and amino acids (C-atoms

in gray) described in Table 2. Dotted black lines depict H-bonds contributing to the polar energies described in Table 2. *C*, The back side of panel B. *D, E*, Packing of terminal sugars against the polypeptide. C-atoms of the peptide are in orange. *D*, Glycan and peptide represented as sticks. *E*, As in *D*, with glycan and peptide represented by spheres.

Figure 1**A. *Toxoplasma* pathway****B. *Dictyostelium* pathway**

(*PgtA-N* & -C domains are reversed in the fusion protein)



Toxoplasma glycan: $\text{Gal}\alpha 1\text{-}3\text{Glc}\alpha 1\text{-}3\text{Fuca1}\text{-}2\text{Gal}\beta 1\text{-}3\text{GlcNAca1}\text{-}4\text{O(trans)-Pro154-Skp1}$

Dictyostelium glycan: $\text{Gal}\alpha 1\text{-}3\text{Gal}\alpha 1\text{-}3\text{Fuca1}\text{-}2\text{Gal}\beta 1\text{-}3\text{GlcNAca1}\text{-}4\text{O(trans)-Pro143-Skp1}$

C. Glycosylation status of Skp1

relative abundance

	Source of peptides					$\Delta\text{Gat1}/\text{Me49}$
	RH	RH + α Galase	$\Delta\text{gat1}/\text{RH}$	$\text{gat1::gat1}/\Delta\text{gat1/RH}$	Me49	
peptide -	0.12	0	0	0.08	0.08	0
peptide -	0*	0.14	0.09	0	0	0.09

* <0.0005

Figure 2

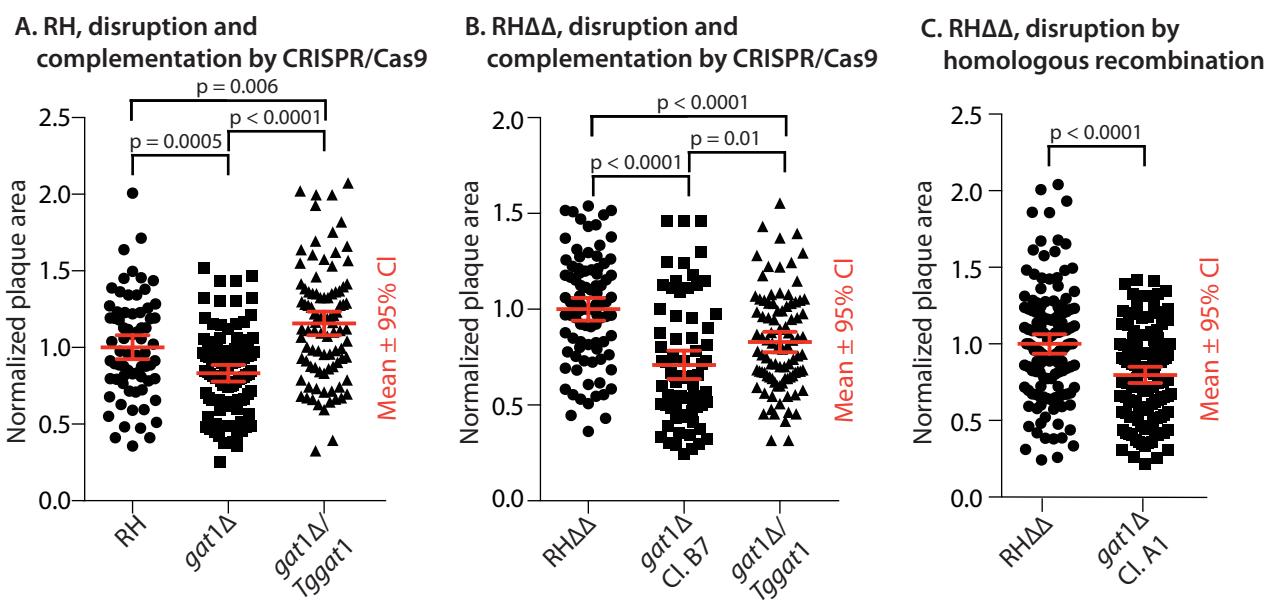


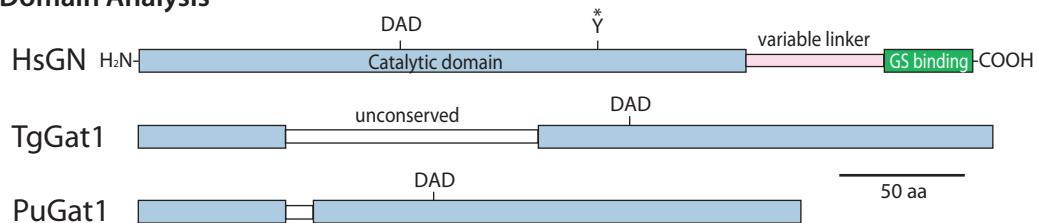
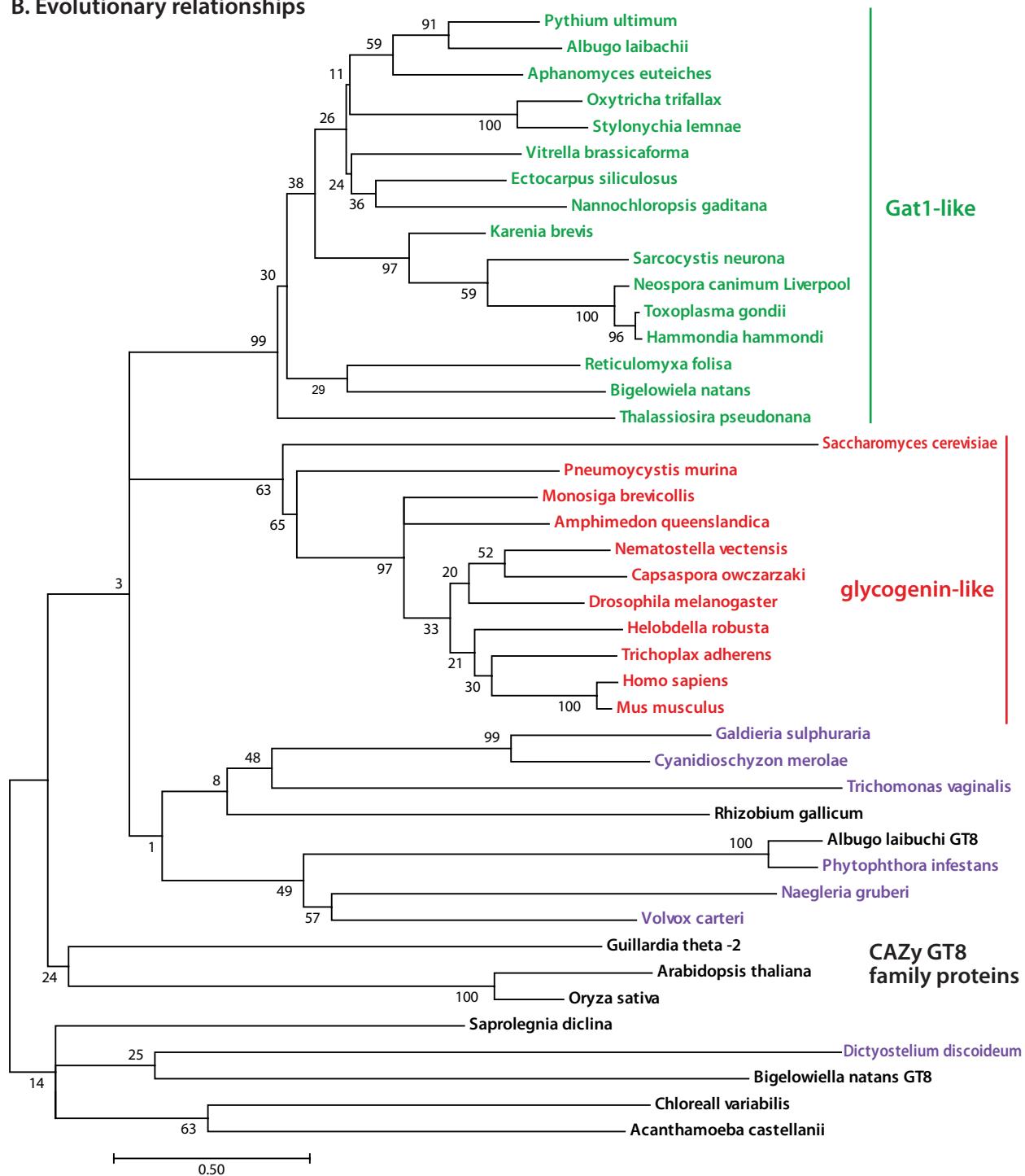
Figure 3**A. Domain Analysis****B. Evolutionary relationships**

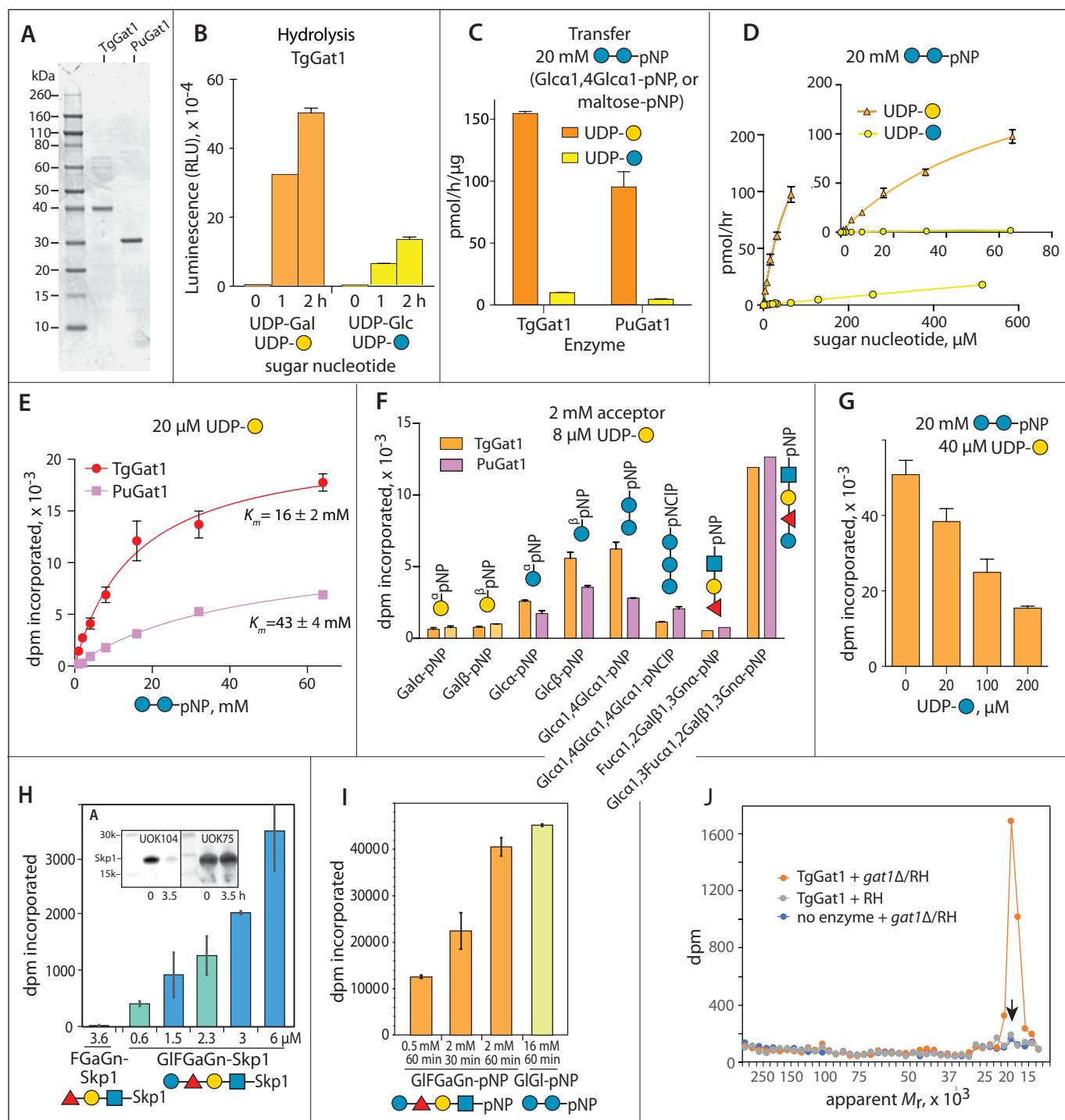
Figure 4

Figure 5

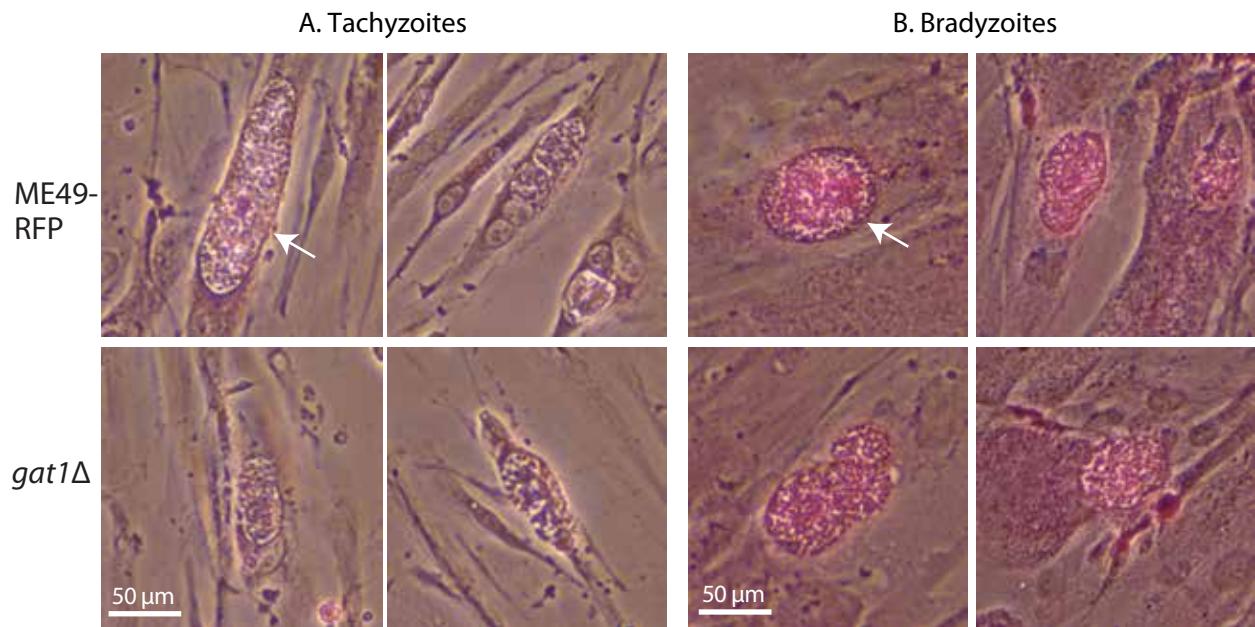


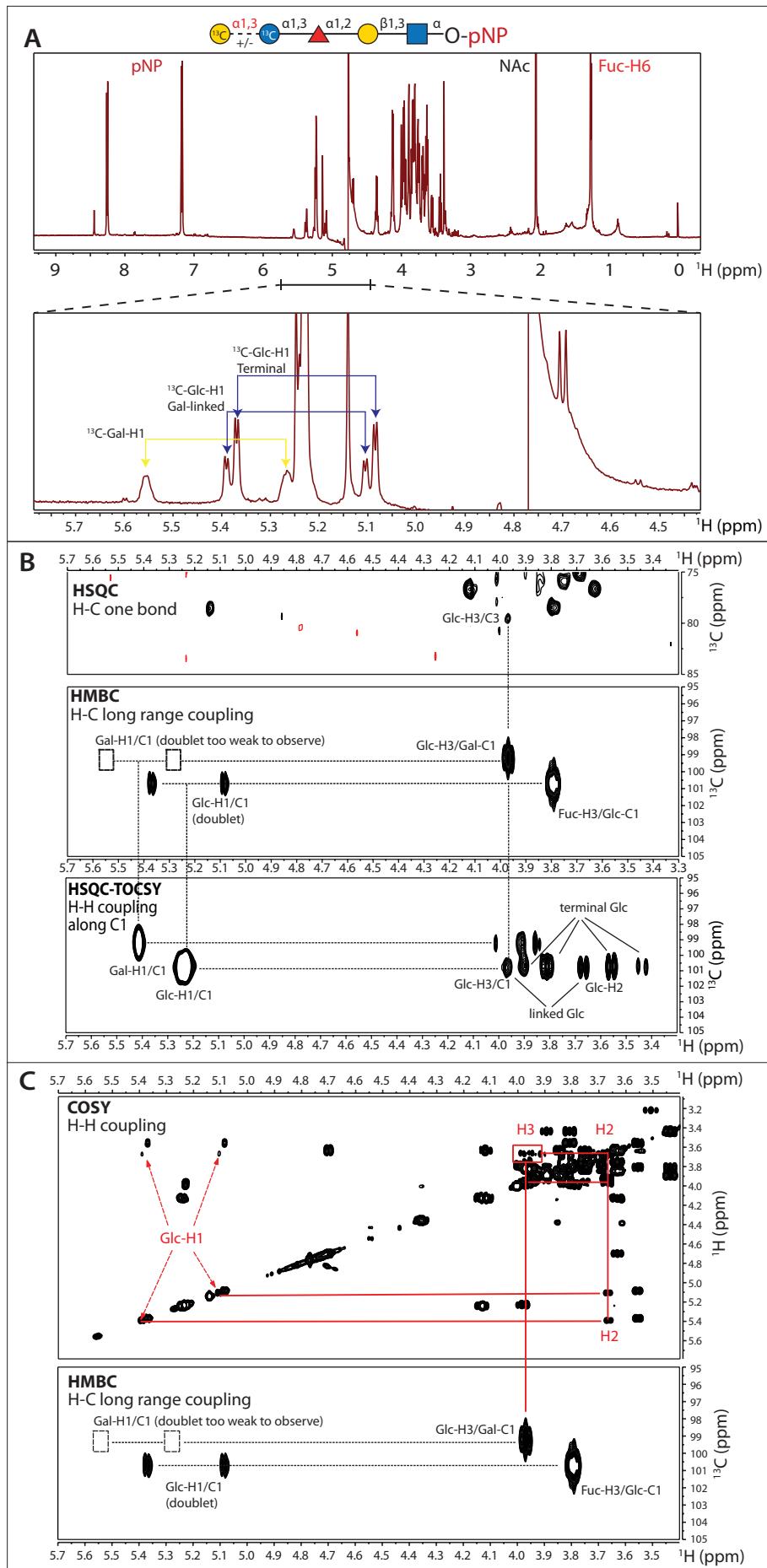
Figure 6

Figure 7

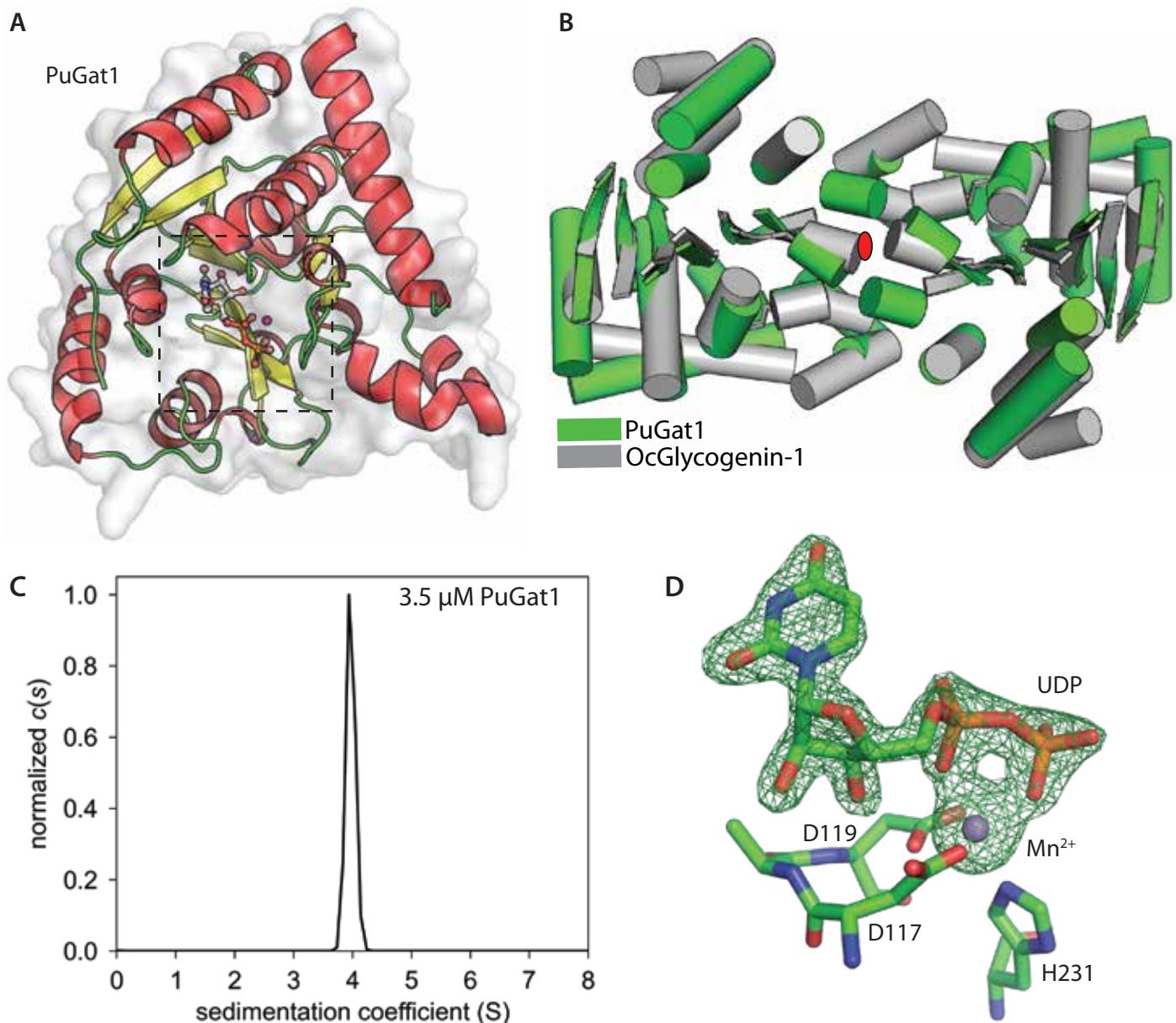


Figure 8

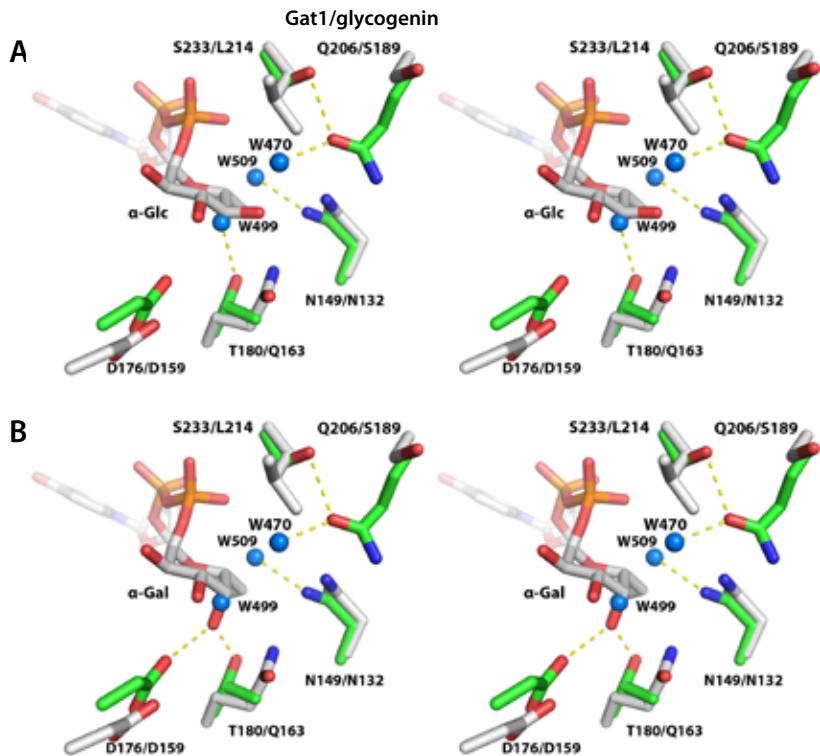


Figure 9

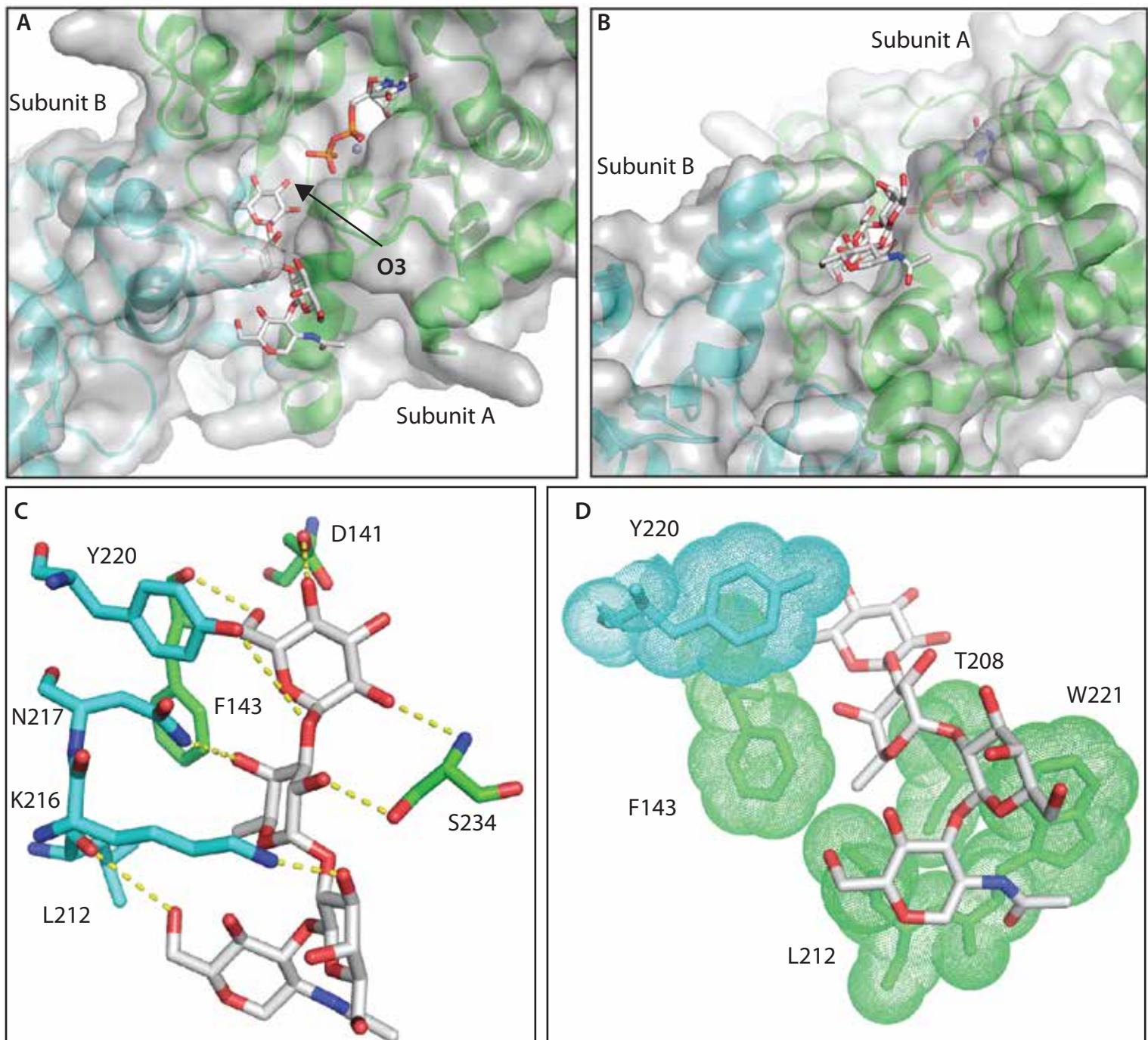
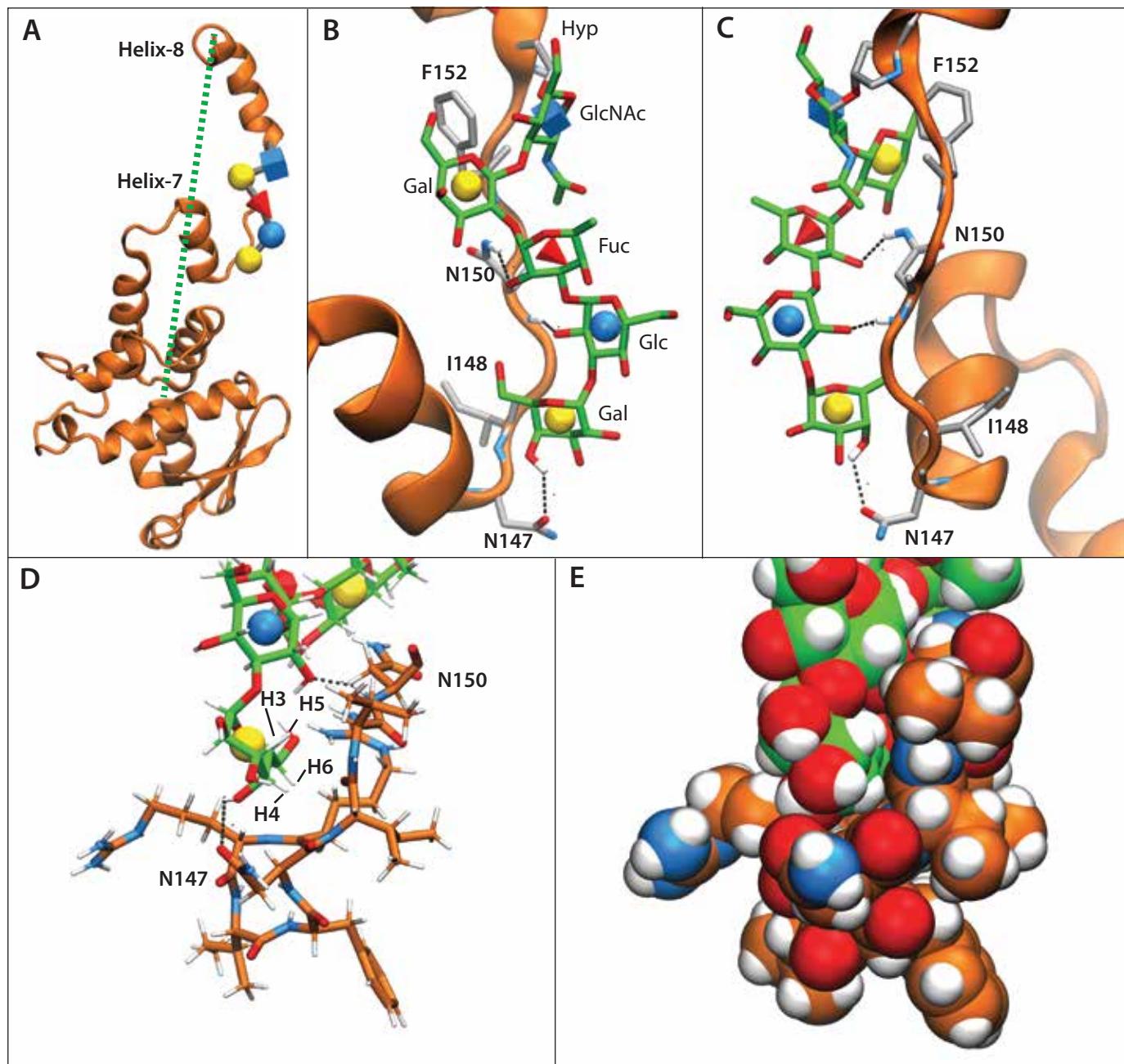


Figure 10



**A terminal α 3-galactose modification regulates an E3 ubiquitin ligase subunit in
*Toxoplasma gondii***

Msano Mandalasi, Hyun W. Kim, David Thieker, M. Osman Sheikh, Elisabet Gas-Pascual, Kazi Rahman, Peng Zhao, Nitin G. Daniel, Hanke van der Wel, H. Travis Ichikawa, John N. Glushka, Lance Wells, Robert J. Woods, Zachary A. Wood and Christopher M. West

J. Biol. Chem. published online May 15, 2020

Access the most updated version of this article at doi: [10.1074/jbc.RA120.013792](https://doi.org/10.1074/jbc.RA120.013792)

Alerts:

- When this article is cited
- When a correction for this article is posted

[Click here](#) to choose from all of JBC's e-mail alerts

SUPPORTING INFORMATION

A terminal α 3-galactose modification regulates an E3 ubiquitin ligase subunit in *Toxoplasma gondii*

Msano Mandalasi*, Hyun W. Kim*, David Thieker, M. Osman Sheikh, Elisabet Gas-Pascual, Kazi Rahman, Peng Zhao, Nitin G. Daniel, Hanke van der Wel, H. Travis Ichikawa, John N. Glushka, Lance Wells, Robert J. Woods, Zachary A. Wood, Christopher M. West

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

* These authors contributed equally to this work.

Table of Contents

Table S1. Oligonucleotides employed

Table S2. Skp1 glycopeptide mass measurements (supports Fig. 1C)

Table S3. PuGat1 crystal parameters

Fig. S1. Nucleotide and amino acid sequences of TgGat1

Fig. S2. Nucleotide and amino acid sequences of PuGat1

Fig. S3 A-C. Disruption and complementation of *Tggat1* in RH $\Delta\Delta$

Fig. S4 A-B. Disruption and complementation of *Tggat1* in Ku80+ strains

Fig. S5 A-G. nLC/MS of Skp1 glycopeptides (supports Fig. 1C, Table S2)

Fig. S6. Comparison of Gat1-like and glycogenin sequences (basis for Fig. 3A)

Fig. S7. Summary of Gat1-related sequences selected for phylogenetic analysis (supports Fig. S8)

Fig. S8. Alignment of Gat1-like, glycogenin-like, and other CAZy GT8 sequences (supports phylogenetic tree in Fig. 2)

Fig. S9 A-I. Characterization of the α GalT activity of Gat1 and biochemical complementation of *Toxoplasma* extracts (supports Fig. 4)

Fig. S10 A-E. Absence of Gat1 autoglycosylation

Fig. S11. Chemical shifts of Gat1 substrate and reaction product (supports Fig. 6)

Fig. S12 A, B. Sedimentation velocity analyses of PuGat1 (related to Fig. 7C)

Fig. S13. PuGat1 and Oc-glycogenin-1 ligand interactions (related to Fig. 8)

Fig. S14 A-E. Computational comparison of the Skp1 glycans from *T. gondii* and *D. discoideum* (related to Fig. 10)

Table S1. List of primers

Purpose	Code name	Primer name	Primer sequence	Location
gat1 disruption in RHΔΔ and complementation in RH	Fa	a) Gat1F1 5'-flank 5'-end 5'	GGGGGCCAACAGCGGATCTTCTGAAC (ApaI)	Tggat1 homologous recombination disruption & complementation plasmids
	Ra	a') Gat1R1 5'-flank 3'-end 5'	GGCTCGAGACCGCGTTGAGCGATTGA (Xho1)	
	Fb	b) Gat1F2 3'-flank 5'-end 5'	GCTCTAGAGAGGGAGAACCAAAGTGATGAT (XbaI)	
	Rb	b') Gat1R2 3'-flank 3'-end 5'	CGCGGCCGCTCGTAGAACACAAGGAGAAC (NotI)	
PCR confirmation for gat1 disruption in RHΔΔ				
PCR1	Fc	Forward	TACCCTGTTGACCGACAATT	Tggat1 genomic sequence
	Rc	Reverse	CTTGCTGGTTGTTCCCAAG	
PCR2	Fd	Forward	GAACCGAAATGACAACGCATTAC	HXGPRT sequence
	Rd	Reverse	AGTCGCGGAACATCTCGTTGAAGT	
PCR3	Fe	Forward	ATTGCATCCTGAAAGGCTCTCGC	Tggat1 genomic sequence
	Re	Reverse	TCTGAAATGGAGTCGCCTTG	
Dual guide CRISPR plasmid for gat1 disruption in RH, Me49-RFP				
NsiI PCR	Ff	Plasmid 3 FOR	CGTGGGGATGCATTACCGCGGCCACATGTIG	Dual guide gat1 CRISPR disruption plasmid
	Rf	Plasmid 3 REV	GCGATGAGCGCAAGCCGTCTGAGTTACG	
dg plasmid sequencing	Fg	gRNA FOR	CAAAGTGCAGCAGTTGAAATCG	
	Rg	gRNA REV	GAGACGATGATTCTGATCACTCCG	
PCR confirmation for gat1 disruption in RH and Me49-RFP	Fh	Gat1 63 seq Fw (P1)	CGTACGCTACCCCTGTTGACG	Tggat1 genomic sequence
	Rh	Gat1 968 Seq Rv (P2)	AGAACATCAGTTGGCACAGTGCC	
	Fi	DHFR F/R Fw (P4)	CCATTGCGGTGTCGTGGATT	
	Ri	DHFR RO Rv (P3)	CCCCTGTGTCCTTATCGAAG	
Complementation plasmid sequencing	Fl	TgGat1 seq Fw	GGACTGTTCACCAAACTAGCGTGTGT	Tggat1 genomic sequence
	Rl	Gat1 3'UTR Rv	CTAGTCAGTCCCTAACGGCTAGT	
Ty tag insertion on complementation plasmids				
Tggat1 Ty Tag insertion	Fn	TgGat1-Ty Fw	GAAGTACACACAAACCAAGACCCACTAGACTAGTGGAGGGAGA	Tggat1 genomic sequence & Ty tag
	Rn	TgGat1-Ty Rv	GTTTGTGTGTACTTCCACGATATCAGAACATCAGTGGC	
PCR confirmation for gat1 complementation in RH	Fp	UPRT Fw	GTCCTAACGTCGCAAGTAA	UPRT genomic sequence
	Rp	UPRT Rv	ATGCGGACTTCCGGTATTTC	
	Fq	Gat1 check Fw	TGGGAACACCAGCAAAGA	
	Rq	Gat1 FO Rv	GGGGTTGCAGCCTATGG	
TgGat1 <i>E. coli</i> expression plasmid	Fr	Gat1 Fw	AAGCTAGCATGTCTCTCGGTACGCGTACGCT	Tggat1 genomic sequence & pET15b expression plasmid
	Rr	Gat1 Rv	AAGGATCCCTACACGATATCAGAACATCAGTTGGCACAG	
DHFR amplicon with 45 bp gat1 arms for CRISPR disruption in RHΔΔ	Fs	63 Fw_dhfr Fw	CGGACAATTCTCTACTATGGTGTGAGGGACTGCTCAAGTCAC	Tggat1 and DHFR
	Rs	968 Rv_dhfr Rv	AAGCTTCGCCAGGCTGTAAAT	
Gat1-HA complementation plasmid in RHΔΔ	Ft	3HA Fw	ATCAGAACATCAGTTGGCACAGTCCCCGTAAGGAAGACTTTCCACCA	pUPRT a1 WT cDNA shuttle Vector-Tub1-3xHA
	Rt	Tub-5'UTR Rv	CATCCTGCAAAGTGCATAGAAG	
	Fu	Tub-5'UTR-Gat1 Fw	CTTTTCGACACGGCGCGCCATGTCCTCGGTACGCG	
	Ru	HA-Gat1 Rv	ACGTCGTACGGTAGGTACCCACGATATCAGAACATCAGTTGGC	
PCR confirmation for Gat1 complementation in RH	Fv	UPRT 5'Arm Fw	GCTGTGCCTAGTATCGAAAGCTGTA	UPRT genomic sequence
	Rv	Gat1 at STOP Rv	CTACACGATATCAGAACATCAGTTGGCACA	
	Fl	TgGat1 seq Fw	GGACTGTTCACCAAACTAGCGTGTGTG	
	Rw	UPRT 3'Arm Rv	CGACGTCACTGTACGACATCC	

Table S2. Skp1 glycopeptide mass measurements (supports Fig. 1C)

Isoforms of the Skp1 peptide 145-IFNIVNDFTPEEEAQVR were detected and quantified as described in Materials and Methods. See Fig. S5 for interpretation of raw data.

The abundances of raw ion counts for the detected isoforms are shown for all the strains analyzed. Hydroxylated, mono, di and trisaccharide glycopeptides were not detected.

Strain ^a	unmodified peptide			H-dH-H-HN-O-peptide ^c			H-H-dH-H-HN-O-peptide			All peptides
	Abundance ^b	[M+2H] ²⁺	Δm/z ^d	abundance	[M+2H] ²⁺	Δm/z	abundance	[M+2H] ²⁺	Δm/z	total abundance
		[M+3H] ³⁺	Δm/z		[M+3H] ³⁺	Δm/z		[M+3H] ³⁺	Δm/z	
RH, wt	1.12E+07 8.59E+06	1011.002 674.337	0.40 0.56	nd ^e			3.96E+04 1.40E+06	1436.650 958.103	0.28 -0.30	2.13E+07
RH + αGalase	4.71E+06 1.34E+06	1011.000 674.335	2.37 3.53	1.07E+05 8.41E+05	1355.62 904.081	3.61 4.47	nd			7.00E+06
Δgat1/RH MM12.A8	9.32E+06 5.67E+06	1011.001 674.336	1.38 2.05	9.99E+04 1.46E+06	1355.62 904.083	1.40 2.26	nd			1.65E+07
gat1::gat1-ty/ gat1Δ/RH MM21.E12	1.42E+07 6.08E+06	1011.000 674.336	2.37 2.05	nd			5.77E+04 1.79E+06	1436.65 958.101	2.37 1.78	2.21E+07
Me49-RFP MM8.A10	1.81E+07 1.20E+07	1011.000 674.336	2.37 2.05	nd			3.88E+04 2.58E+06	1436.65 958.100	2.37 2.83	3.26E+07
Δgat1/ME49 MM14.B5	8.27E+06 3.64E+06	1011.000 674.336	2.37 2.05	4.57E+04 1.08E+06	1355.61 904.081	10.99 4.47	nd			1.30E+07

Notes:

^a see Table 1 for detailed descriptions

^b abundance from ion raw spectral counts

^c H=Hex; dH=deoxyHex; HN=HexNAc

^d Δm/z in ppm, see below for expected m/z values

^e nd: not detected (<5000)

Expected masses for each glycoform are as follows:

unmodified peptide	tetrosaccharide-peptide	pentasaccharide-peptide
[M+2H] ²⁺ 1011.002	[M+2H] ²⁺ 1355.624	[M+2H] ²⁺ 1436.650
[M+3H] ³⁺ 674.337	[M+3H] ³⁺ 904.085	[M+3H] ³⁺ 958.103

Table S3. Crystallographic data

Data collection	PuGat1:UDP:Pt ²⁺ (PDB_6MW5)	PuGat1:UDP:Mn ²⁺ (PDB_6MW8)
Wavelength (Å)	1.85	1.0
Space group	P4 ₂ 2 ₁ 2	P4 ₂ 2 ₁ 2
Unit cell dimensions (a, b, c)	83.78, 83.78, 75.84, 90.00, 90.00, 90.00	84.06, 84.06, 76.08, 90.00, 90.00, 90.00
Completeness (%)	97.4 (94.8) ^a	99.9 (99.8) ^a
Total number of reflections	396057 (13045)	800853 (52625)
Unique reflections	29424 (2114)	28125 (2042)
Redundancy	13.5 (6.2)	28.4 (25.8)
I/σ(I)	28.6 (1.21)	36.94 (1.85)
R _{meas} ^b (%)	6.1 (145.4)	6.2 (203.9)
CC _{1/2} ^c (%)	100.0 (49.0)	100.0 (65.5)
Refinement		
Resolution (Å)	2.1	1.76
R _{work} /R _{free}	0.196/0.242	0.181/0.210
No. of atoms Protein/ Ligand / Water	1957/30/54	1949/42/122
Wilson B-factor (Å ²)	45.2	39.9
B-factors (Å ²) Protein/ Ligands and Water	45.2/45.66	39.4/44.2
Stereochemical Ideality		
Bond lengths (Å)	0.006	0.006
Bond angles (°)	0.786	0.784
φ, ψ Most favored (%)	97	99
φ, ψ Additionally allowed (%)	3	1
SAD Phasing statistics		
Heavy atom sites	1	
Figure of merit	0.31	

^a Values in parentheses are for highest-resolution shell

^b R_{meas} is the redundancy independent merging R-factor of Karplus and Diederichs (86)

^c CC_{1/2} is the percentage of correlation between intensities from random half-data sets

Figure S1. Genomic sequence surrounding the open reading frame of Gat1 (TGME49_310400 model from Toxodb.org). Numbering begins at the A of the start codon ATG. Coding sequences, including those upstream and downstream of Gat1, are capitalized; non-coding sequences are lower case. Amino acid sequence of Gat1 is above its coding sequence. Sequences of oligonucleotides from Table S1 are shown and mapped. For forward PCR primers, cognate sequences are colored purple; cognate sequences of reverse PCR primers are in red; guide DNA sequences are in blue. nt differences observed in the type 1 RH strain are indicated.

```

>TGME49_chrXI:1317070..1322069
ATCGTCGCTATACATGAGGAAGCGAACGTCAGACAAACCCAGGGCAAACGAAATTGGAAGA
AAGACACTCCTAAACTCGTCGGCGGGGGTGCCAGTGAATCGATTGCGATTGACTTGC
TCTTGAGTTGGATACTCCGGCCGGCCAATTGCAACATTGTCATTATCCAGTCCATTGCT
AGGTAGAGTGTGGAGAAGAGAGAGAAAACGCAATGCAAATCGAACCCAAAAGGCTGCC
TGTCTACTCCTGGCACGCCATgatttccggaggaactgaggacatctccttttgc TGGT1_310390 reverse
gtagataactttgccctcgctcgctccctttgttccttctggccaccccttgacacctct
tcttgcttaattgaatctcgccgtctctttcgggtcttctccgtcgtaaacgggtcctt
cccgtcactttgtccgtctctttcgggtcttctccgtcgtaaacgggtccttgc
tttcatcagccgcgacagtctgccttcgtgtttccccctctgttctgtgtttcg
tggtgtctcaactcgctgtgtactcctccctccgtccgtttctgttctacttaggt
tccctccctctggctatccaaaaaaggatctaattgacagcgaacatcccacaact
ttttctcagctaaaatcacaactatgccaacacacaaggcgcgtaccataattccgt
ataggttccctcctacacggcaacacttaaacaggcgcgttctaaactattacgaagtc
gaggcgagaacccccacaacttcctatccagtgccacaagaaccgaatgacaacgcattact
5'-gaaccgaatgacaacgcattac Fd

tttacacaagtgcctagttgaacacactagatacacaattttgttgcgttatggag
tccaccagtgacaccacatcgaaaggcgttgcgtactttgacccgttgcgttgcctcg
5'-ggg Fa

cgaaaaaccagggatcttctgaacacgtttctgcgaagaagaaggtaagctaccact
ggcccaaccagggatcttctgaac Fa

tttcagtgacgcttattgaaagaacaaatgaaggtgacagaatcaaaaggaaacacagga
ccgcaaagccaaacactcttccacatttcagcggaaatttactcacgtaccatgcctgac
a
tgaggccgtaggctgcaaccccccctccaccgcctccgtttctcagcctgacggcac RH
ggtatccgacgttgggg-5' Rq

tcagacggcgagttcgggaaggcagcctgcctttcaaacatatataatcaatctactgcgtt
cccgattatttgcgcagaacgcgaaaaagaagtcttcacatctacatccccccacg
gtgtcgcttggaaaactgtcttaatgccaatgtatggcattagtgcacacatgaaagc
cgattttctaagcgaataataccacgggaaactccgggtgtcttggaaagacgaaaa
atccaatgcatgactcgatgtaaatggatggcttgcgttgcgttgcgttgcgttgcgtt
cgcaattctggcttgcgttgcacttccccccctccggccccccgcctccagcgaagcga
tgtgtcgaggcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
tcttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
tgaagtgtcaggcattatttcactgggtggacgtggcggtgttgcgttgcgttgcgtt
atgtgtgtctcaaattccctttctcgcggcacttcgaggcggacgggggggggggggg
cacacgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
cttcaactgcacatgcacatgcacatgcacatgcacatgcacatgcacatgcacatgcacat
ttgtatgtgtggcatcattcactgggtggacgtggcggtgttgcgttgcgttgcgttgcgtt
caggaagttgcactaaatggaaaattggcttgcgttgcgttgcgttgcgttgcgttgcgtt
ctttgtctcccttcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt

t RH
gcattaagcggcacttcagtctgaatcaatcgctcaacgcgtcatttccttcctttt
agttagcgagttgcgcagagctcg-5' Ra
gcagtaaaagaaaaggaaaaaa Rk
5'-cttttcgacac Fu
M S P R Y A Y A T L L T D N S F Y 17
gctaaggagATGTCTCCTCGGTACGCGTACGCTACCCCTGTTGACGGACAATTCTTCTAC 51
cgattcctctactgcag-5' Rk
aagctagcatgtctcctcggtacgcgtacgct-3' Fr

```

ggccgccc **atgtctcctcggtacgcg-3'** Fu
 5'-**cgtacgc**taccctgttgacg Fh
 5'-taccctgttgacggacaatt Fc
 5'-CGGACAATTCTTCTAC Fs

Y G V E A L L K S L E A T K T P Y P V L 37
 TATGGTGTGAGGCACTGCTCAAGTCACTGGAGGCTACGAAGACGCCCTAACCCGTGCTT 111
 5'-**ggcactgctcaagtactgg** gDNA-63
 TATGGTGTGAGGCACTGCTCAAGTCACAAAGCTCGCCAGGCTGTAAAT Fs

L L H T S D V S Q S T I K A L V Y Q R R 57
 CTTTGACACATCTGATGTTCTCAGAGTACAATAAAGCGTTGGTTATCAGCGTCGA 171

K A P A S E D A G T T G K E M K T G Q E 77
 AAAGCCCCGGCGAGTGAGGATGCGGGAACTACAGGAAGGAAATGAAAACAGGGCAGGAA 231

V I P S S Q C P E H T P G R N L H S P I 97
 GTCATCCCAGTTCACAGTGTCCAGAACACACCCCCAGGTAGAAACTTGCACCCCCCATT 291

G R K G V N P V S C S V T Q D E T R V R 117
 GGCAGGAAAGGGTAAACCCCTGTGAGTTGCTCCGTACACAAGACGAGACTAGGGTCGT 351

T D S D R I E E A E R R A S E R T S E R 137
 ACTGATTCAAGATCGTATAGAAGAACAGCGTCGAGCCTCAGAGAGAACCTCGGAGCGA 411

A R A G E T E E Q G I C V I P R L V G S 157
 GCGAGAGCTGGGAAACGGAGGAACAGGGCATTTGCGTTATTCCCCGACTCGTTGGTTCT 471

V A Y P K A E R D T C P V E G W K D C F 177
 GTCGCGTACCCCTAAAGCGGAACGGGACACGTGCCCTGTTGAAGGGTGGAAGGACTGTTTC 531
 5'-**ggactgtttc** F1

T K L R V W E Q V D F D V I V Y V D A D 197
 ACCAAACTCGCGTGTGGGAGCAGGTTGACTTCGATGTGATTGTATGTCGACGCCGAC 591
accaaactcgctgtgtg F1

C I V L R P V D E L F L R Q P L P A F A 217
 TGTATAGTTTGGCCGGTAGACGAGCTTTCTTAGGCAGCCACTACCCGCCTTGCA 651

P D I F P P D K F N A G V A V L K P D L 237
 CCAGATATCTCCCTCCCGATAATTAAACGCGGGAGTCGAGTGTGAAGCCCCGACCTC 711

G E Y G N M V A A V E R L P S Y D G G D 257
 GGCGAATACGGAAATATGGTAGCCGCGTCGAGCGTTACCTCATATGACGGAGGCGAC 771

T G F L N A Y F S S W Y E N A A G A R L 277
 ACAGGGTTTTGAACCGTATTCTCATCGTGGTATGAAAACGAGCTGGCGCCGTTTG 831

P F R Y N A L R T L Y H M T Y S S R K G 297
 CCCTTCGGTACAATGCTCTGCGCACACTGTATCACATGACGTACTCCAGTCGAAAAGGA 891

Y W N A V K P I K I L H F C S S P K P W 317
 TACTGGAATGCCGTCAAGCCGATCAAATCTGCACTTCTGCTCCTCCCCGAAGCCTTGG 951
gaacc Rc
5'-tgg Fq

E Q P A K T D L E E L W W K V F L T G T 337
 GAACAAACCAGCAAAGACCGACCTCGAGGAACATGGTGGAAAGTCTCCTACGGGCACT 1011
gaacaaccagcaaaga Fq
cttgttggtcgtttc-5' Rc
 5'-**ccgacctcgaggaactatgg** gDNA-968
ccgtga Rh
 5'-**gaagta** Fn

ct Rr
gaagatacgtaaacgtcctacaccacccatcagaaggaaatgcccgtga Rs

V P T D S D I V *	345
GTGCCAACTGATTCTGATATCGTAGtgaggagaaccaaagtgtatgaaatg	1071
cacggttactaagactatagcacc-5'	a Rv
cacggttactaaga-5'	Rv
cacacaaaccaagacccactagactgtggagggaga	Rh
cacggttactaagactatagcacatcgatcctt-5'	Fn
gcctgtactaagactatagcacatgtgtgtttg-5'	Rr
cgttgtactaagactatagcacatggatggcatgtgc-5'	Rn
5'-cttggatagtggagggagaaccaaagtgtatg	Ru
5'-gctctaaggagaaccaaagtgtatg	Fk
cacggttactaagacta-5'	Fb
cacggttactaagacta-5'	Rs

accgacttccaaaagaacggaaacgcggacagctgcctcggtacctggaaaagag
cgggacgtgtgaatcctgtcaactatctttctgtgtcacctgtggacgaattgtaaa
tctttaaagtacaaacggagtagcagcttaatctgttaatttcttcgaaggacgc

a	RH
cagtgcgcccaagcgtctagtggctgtcaaagactagccttagggactgactgggttcgc	R1
tgatcggaatccctgactgatc-5'	R1

gtacgcaaccatcacgcacaaggatgtttatggccactgggtgtcactcagctagacg
cgtcatgttatgtatacgtacgtttcacagccctctcagagacatccgcacaacgcac
aaccgctgcaaccagaataactgaccgtcagcggttcgctgtttaaatcggttttt
tggaaaaactcaaaggtagtcgttctgtacatctccctaagttaaatcggttttt
acgagcatacattgacaataagacggttctcacaatgaacatccaaagaggcactaga
ccaaacaaaggagctaaagacacgagcaagatgaagataaaacgcacccatcgacgg
catataacaaagtggatcttcacagtagtattctgtgtccgtaccagtgcgtgcacaa
gaagacgcattgtgaacgggtCTACTCTGGATTGAAAGCATTCAATACGCCAGAGCTG TGGT1_310410 reverse
CCCACTACGCACACCGAACACCGCCAGGAACCGGTTTCGTCAGATGAGCAGCATCC
GCAACAGACGTTTGATCGGTCCACACGCTCCACCGAAAAAATTGTTGTCGATG
GTTGTAAGCTGCCCAATATGCCGCTGCTGGCATTACCGCAGCGCATTAAACCTC
TGGCGGTAAGCGATGGTTCTTGTGGCACATGAACGCTGGATAGGCTCCGCTCTTC
ACCAGTCGAGATGAACTATTTCTGTTCCCATCGCATCCACAGAGGCCACACCGGACGA
ATAGTCCACACCTGCACCCATCGATCCCAGCCTACAGTAAACAGAAGGGGGAAAGAGAAGC
GAGCAACTCGTGTGACGACGCATGAGCGTTTACTCTACAGGCGGATATCAACCTA

T	RH
AGGGAGAGAGACCGGTTGTTCTCTCGTGTCTACGCAAATAAAACTGGACCGTTACTGAC	Rb
caagaggaacacaagatgcgtcgccggcgc-5'	Rb

TGGTCATCACCGCACGGGACCGAGCAGTAGCAAGGGACTCCATTCTAGACTCTGGAGG
gttccgctgaggtaaagtct-5' Re

CGGGACGAACCCAGCCCTGCAGGAGATGCGACCGACGCCGCTTCAACCTTGTCCCC
AAAGCGTTTGGTTGGCCACTTCAAGTCATTACGCCACACTGGTGTGCGTGTAGA
GTGCCCAATCGCTCCAGTCGAGGCACCGCAGCGCTCGTGTCCCCCAGGAACACCGCGATT
TCTCCGGTGGTCATTCCGACCAGTACCAAGACCCCATGTGTGTTCTCGACGGTTCC
TCACTCCGAACCTGTCTGGAGCTAGCAAAGGAGCAGCCTCCAGCATCGTTGCATGC
GTGTCTAGACAAACGACACGAACGAGAGAGCCGCGCCTCCAGCGGGCATGATGCTTT
TTTCTAAGCGACTGTTCCGGCTCCGGTACCATGAGCCAGCTGCCGAATTCTGAGAC
CTCGTCAGCACAGTGTGCTTCCACATACCGGACTTCGGCCAGCAGCAGCCGTCCAC
ATGTGCACAGTCAGAGACTC

Fig. S2. *Pythium ultimum* Gat1 sequences

Bold: amino acid sequence

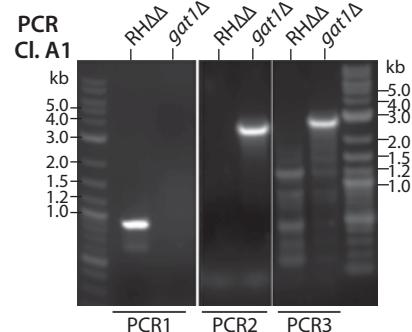
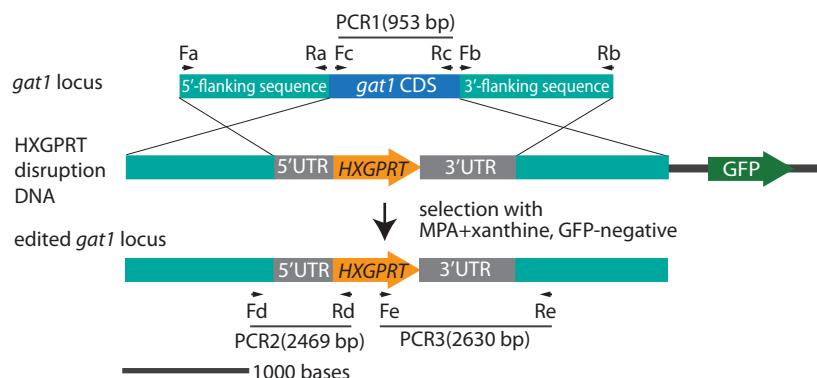
Black: native genomic coding sequence, from PYU1_G002535-201 (UniProtK3WCV7)
Red: Synthetic codon optimized sequence

M T V G T R R A A Y A T L I T S D A Y V		
atgaccgtcgacgcgcaggcgccgtacgcaacactgatcacgtccatgcgtacgtc ATGACTGTGGAACACGTCGTGCCATTGATCACCAGCGATGCGTACGTT	60	
M G V E A L V Y S L F K A R V A F P L V		
atggcgctcgagggcgctcgactcgcttcaaggcgccgttagcggtttcccgctcgta ATGGCGTCGAGGCATTAGTGTATAAGCTTAAAGCGCGTGTGCCTCCCACTTGTG	120	
V L H S S Q V T Q P T V A K L T R F C A		
gtgctgcattcgtcgcagggtacgcgcggccacgggtggccaaactcacgcgttcgcgc GTGTTACACAGCAGCCAGGTTACTCAGCCAACGGTGGCGAAGCTTACCGTTCTGCGCC	180	
P F Q S S T W R I S F R S V P D I G I P		
ccattccaggcatcaacgtggcgcattcgttccgtctgtcccgatatacgccatcccc CCCTTCAAAGCAGCACATGGCGTATTAGCTTCGTAGCGTTCTGATATCGGTATCCCA	240	
D E V T D R S T V H V P G W V N S G Y T		
gacgaagtcaactgataggacacgggtcacgtgcctggatgggtcaactcggggtacacc GACGAGGTAACTGATCGTAGTACCGTCCATGTGCCGGATGGGTTAATTCAAGGTACACA	300	
K L H I F A M D D F E Q I V Y I D A D A		
aagctccacatcttcgcccattggacgactttgagcagatcgatcattgacgcgcgc AAGCTTCATATCTCGCTATGGACGACTTCGAGCAAATCGTCTATATTGACGCCGATGCC	360	
I V L Q N V D E L F D R S T S F A A A P		
atcgcttacagaacgtcgacgagctttcgatcgctcaacgagactttgcggctgcgc ATTGTTCTCAAACGTAGACGAGTTGTTGACCGTAGTACAGCTTCGCGGGCGCCT	420	
D V F P P D R F N A G V L V I R P N K Q		
gacgtgtttccaccgcaccgcttcaacgcggcgctgtgtatccgtccgaacaaggc GACGTATTTCACCAACGACCGTTAACGCAGGGTGCTTGTCAATTGCTTAACAAACAA	480	
L F A D L L A K A K E L K S Y D G G D T		
ctctttcgagacttactggcggaaaggccaaaggagctcaagtcgtacgtggcgacac CTTTTCGCCGACTTGTAGCGAAGGCCAAGGAATTGAAAAGCTATGATGGGGCGATACA	540	
G F L N A F F P K W F E S D A A S R L P		
ggcttcctcaatgcgtttcccaagtggctgaatcgacgcgcgcgtcgagactgcgc GGATTCTAACGCTTTTCCCAAGTGGTCGAGTCCGATGCCGCTCACGTTGCCT	600	
F G Y N A Q R T M Y W L V N G K N P G Y		
tttggatacaacgcgcaggcgacgtactggctcgtaacggcaagaacccgggtac TTTGGTTACAATGCTCAGCGTACGATGTACTGGCTTGTGAAACGGGAAGAACCTGGGTAC	660	
W N A V Q P L K I L H Y S S N P K P W E		
tggAACGCCGTCAGCCTTGAAGATTCTTCACTATTCAATCCTAACCCCTGGGAG TGGAACGCCGTCAGCCTTGAAGATTCTTCACTATTCAATCCTAACCCCTGGGAG	720	
D P S R K G D L E I L W W Q M Y T E S R		
gatccgagtcgcagggtgacctggagatcctgtggcaaatgtacacggaaatccaga GACCCAAGTCGTAAGGGTACTGGAGATTCTTGGTGGCAAATGTATACCGAAAGTCGT	780	
C M S F L G *		
tgcatgagcttctggggtag TGTATGAGCTTCTGGATAG	801	

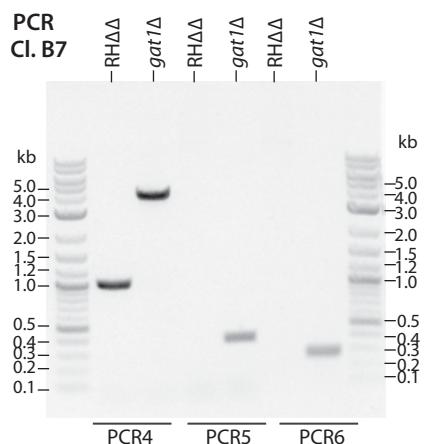
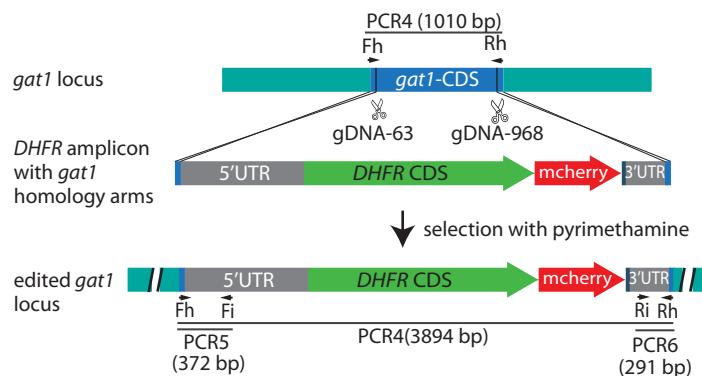
Fig. S3. Disruption and complementation of Tggat1 in the RHΔΔ type 1 strain. *A*, Disruption of *gat1* by double cross-over homologous recombination in RHΔΔ. The disruption DNA consisted of an HXGPRT cassette flanked by a 1177-nt 5'-upstream DNA (prepared by PCR using primers Fa and Ra) and a 1209-nt 3'-downstream DNA (prepared using primers Fb and Rb) of the *gat1* coding sequence (CDS), and an adjacent GFP-expression cassette. GFP-negative clones that grew in mycophenolic acid (MPA) and xanthine showed evidence of a gel shift of Skp1 relative to parental and *phyA*Δ cells based on Western blotting using pAb UOK75. Clone A1 was confirmed to have the desired integration by PCR reaction #1 (primers Fc and Rc), which showed loss of the *gat1* CDS, and PCR2 (Fd and Rd) and PCR3 (Fe and Re), which showed integration of the HXGPRT within the *gat1* locus. *B*, Disruption of *gat1* using a double CRISPR/Cas9 strategy. RHΔΔ parasites were transiently transfected with a plasmid encoding gDNA-63 and gDNA-968 guide DNA's and Cas9, and a PCR amplicon expressing the DHFR resistance cassette flanked by 45-bp *gat1* homology arms. Pyrimethamine-resistant clones that replaced the *gat1* CDS with the DHFR cassette were confirmed using PCR4 (Fh and Rh), which showed loss of *gat1* CDS, and PCR2 and PCR3, which showed the integration of DHFR in the forward orientation. *C*, Complementation of clone B7 from panel B by replacement of the *uprt* locus with a *gat1* expression cassette consisting of a tubulin promoter, Tggat1 CDS modified with DNA encoding a C-terminal 3×HA tag, and *uprt* targeting sequences. Correct insertion of the *gat1* expression cassette was assessed by PCR reactions PCR7, PCR8 and PCR9, and confirmed by Western blot analysis for a predicted M_r 45,000 protein band that could be detected with mAb 12CA5 that recognizes the 3xHA epitope. A parallel gel containing samples without reducing reagent was Western blotted to detect Sag1 as a loading control.

Figure S3. Disruption and complementation of Tggat1 in the RHΔΔ type 1 strain

A. Double cross over homologous recombination in RHΔΔ



B. CRISPR/Cas9 mediated replacement in RHΔΔ



C. CRISPR/Cas9 mediated complementation at the *uprt* locus in *gat1Δ/RHΔΔ*

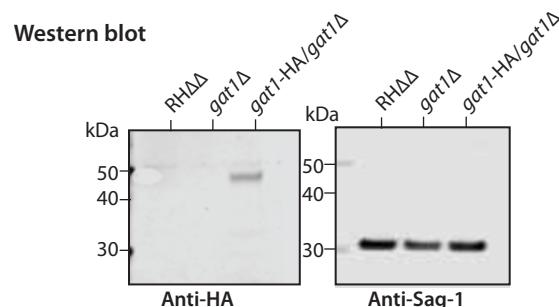
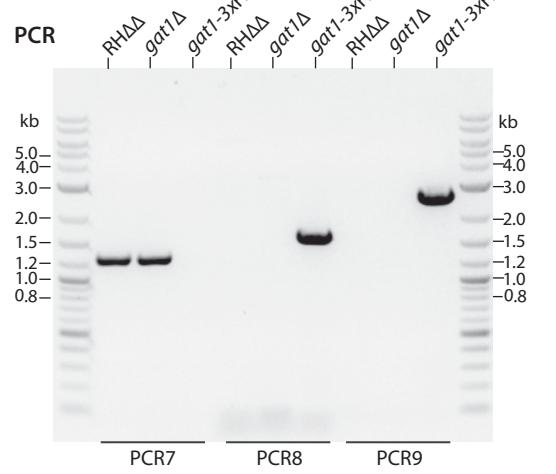
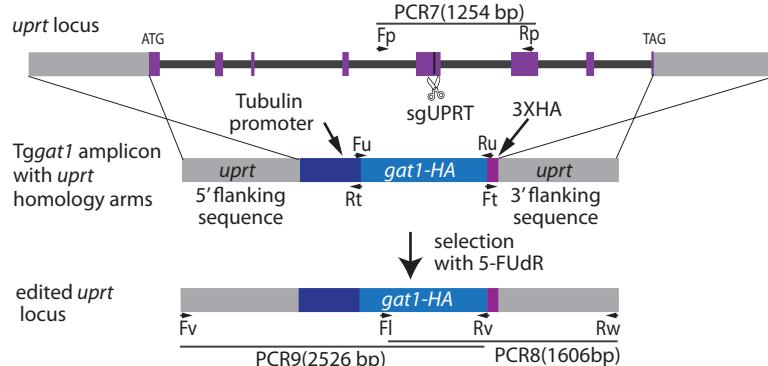
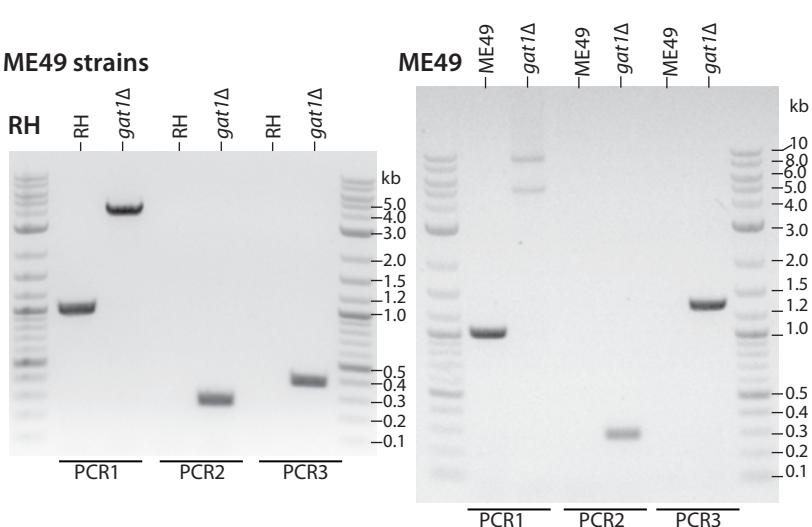
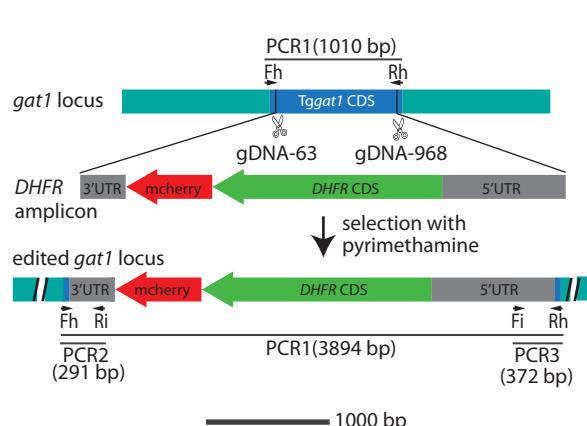


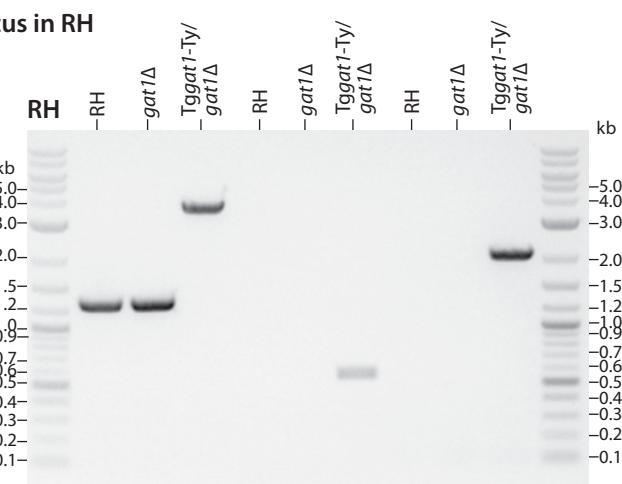
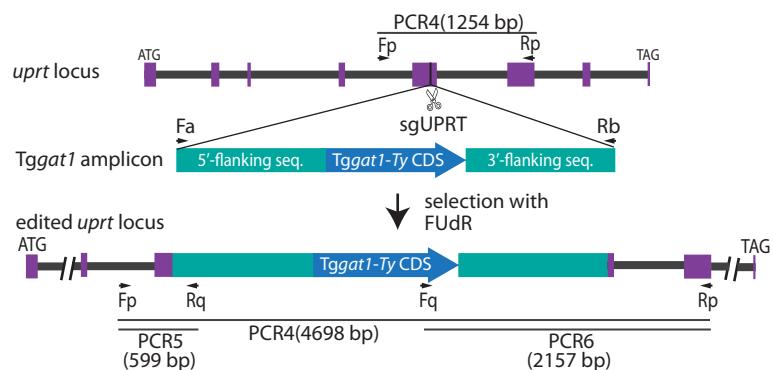
Fig. S4. Disruption and complementation of *Tggat1* in Ku80+ type 1 and type 2 strains. *A*, Disruption of *gat1* was achieved using the double CRISPR/Cas9 strategy described in Fig. S3B, except that the DHFR amplicon lacked *gat1* homology arms owing to the presence of non-homologous end joining activity. Successful replacement was evaluated for strains RH and ME49, by PCR as in Fig. S3B. *B*, The RH *gat1* Δ strain was complemented by insertion of a genomic fragment of *Tggat1* including its CDS, DNA encoding a C-terminal Ty-tag, and >1 kb of flanking DNA from both directions. Successful integration was verified using *uprt*-specific primers (Fp and Rp, Table S1) flanking the CRISPR/Cas9 cut site in PCR reaction #4. The identity of the integrated DNA was verified using primer pairs Fp and Rq, and Fq and Rp, in which Rq and Fq were specific to *Tggat1* DNA. *C*. Extracts of *gat1* Δ and complemented clones from panels A and B were analyzed for Skp1 α GalT activity. Desalted S100 extracts were prepared by hypotonic lysis and gel filtration, and incubated in the presence of GlFGaGn-Skp1 and UDP-[³H]Gal. The reactions were separated on SDS-PAGE gels, the Skp1 band was excised after Coomassie blue staining, and radioactivity determined by liquid scintillation counting. Error bars represent S.D. of two technical replicates of the same samples.

Figure S4. Disruption and complementation of *Tggat1* in *Ku80⁺* type 1 and type 2 strains

A. CRISPR/Cas9 mediated *gat1* replacement in RH or ME49 strains



B. CRISPR/Cas9 mediated *Tggat1-Ty* complementation at the *uprt* locus in RH



C. Enzyme assay for TgGat1 αGalT activity in parasite extracts

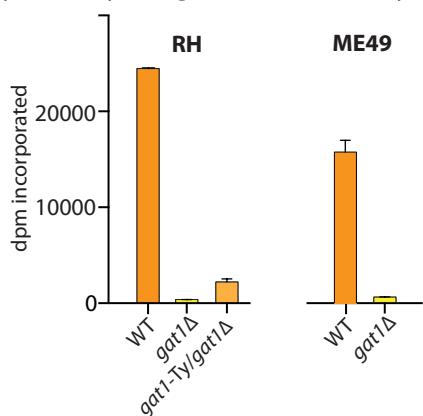


Fig. S5. nLC/MS of Skp1 glycopeptides (supports Fig. 1C, Table S2)

TgSkp1 isolated by immunoprecipitation from tachyzoite extracts were reduced and alkylated, trypsinized treated with green coffee bean α -galactosidase as indicated, and analyzed by a standard proteomics workflow consisting of separation on a C18 nLC column and analysis in an QE-Plus Orbitrap mass spectrometer.

The raw data files listed below are deposited in a data repository at <https://figshare.com/>

Figshare ID 10.6084/m9.figshare.12272882 (Skp1 glycopeptides raw data Fig. S5). The data are analyzed in Panels A-G.

Samples analyzed (as in Fig. 1C): Original datafile name:

RH (type 1 parental)	RH.raw
<i>gat1</i> Δ /RH	<i>gat1</i> _delta_RH.raw
Tggat1/ <i>gat1</i> Δ /RH (complemented)	Tggat1_$gat1$_delta_RH.raw
RH, incubated with α -galactosidase	RH_aGal_ase.raw
ME49 (type 2 parental)	ME49.raw
<i>gat1</i> Δ /ME49	<i>gat1</i> _delta_ME49.raw

A, Stacked extracted ion chromatograms for all isoforms of peptide(134-150), which contains the modifiable Pro143, that were detected in the RH and ME49 backgrounds.

B, Selected extracted ion chromatograms for unmodified peptide(134-150), and an example of an MS¹ spectrum, from the RH sample, highlighting the parent ion for the unmodified peptide.

C, Selected extracted ion chromatograms for pentasaccharide-modified peptide(134-150), and an example of an MS¹ spectrum, from the TgGat1 complemented sample, highlighting the parent ion for the pentasaccharide-modified peptide.

D, Selected extracted ion chromatograms for tetrasaccharide-modified peptide(134-150), and an example of an MS¹ spectrum, from the *gat1* Δ /RH sample, highlighting the parent ion for the tetrasaccharide-modified peptide.

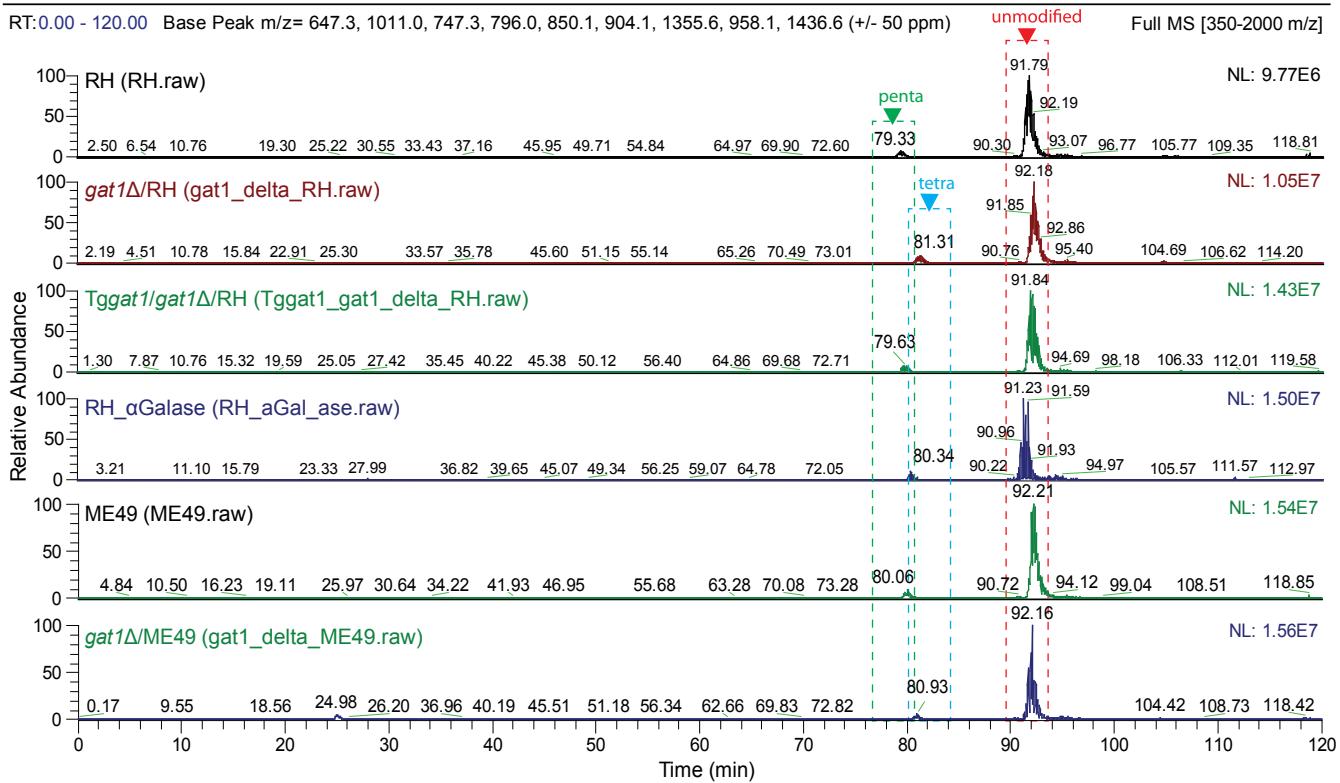
E, MS² of unmodified peptide(134-150) from RH, with associated extracted ion chromatogram and MS¹. Detected b and y fragment ions that define the peptide sequence are in bold in the list of predicted fragments ions at the bottom.

F, MS² of pentasaccharide peptide(134-150) from RH, with associated extracted ion chromatogram and MS¹. An expanded table of predicted b and y fragment ions, calculated to include the full pentasaccharide or a GlcNAc stub, is at the bottom. MS² fragmentation resulted in either loss of the full glycan leaving Hyp, or retention of a GlcNAc stub (encircled with a dashed green line).

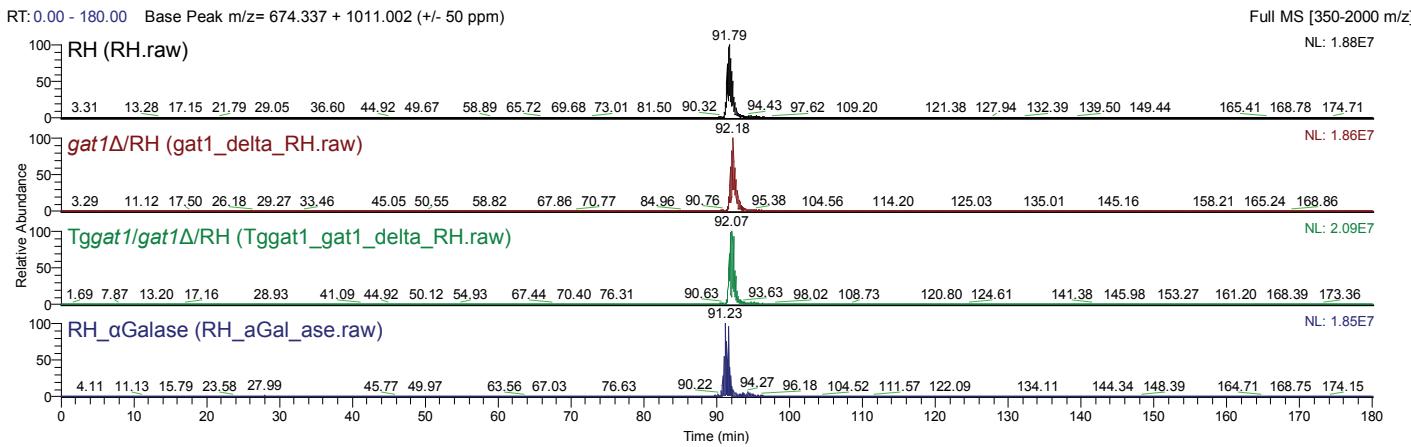
G, MS² of tetrasaccharide peptide(134-150) from *gat1* Δ /RH, with associated extracted ion chromatogram and MS¹. An expanded table of predicted b and y fragment ions, calculated to include the full tetrasaccharide or a GlcNAc stub, is at the bottom. MS² fragmentation resulted in either loss of the full glycan leaving Hyp, or retention of a GlcNAc stub (encircled with a dashed green line).

Figure S5A,B

A. Extracted ion chromatogram Summary: IFNIVNDFT(HyP)EEEAQVR (all glycoforms)



B. Extracted ion chromatograms, MS1: IFNIVNDFT(HyP)EEEAQVR (unmodified)



RH (RH.raw)

RH.raw # 29612 RT: 91.54 AV:1 NL: 4.22E6
T: FTMS + p NSI Full ms [350.0000-2000.0000]

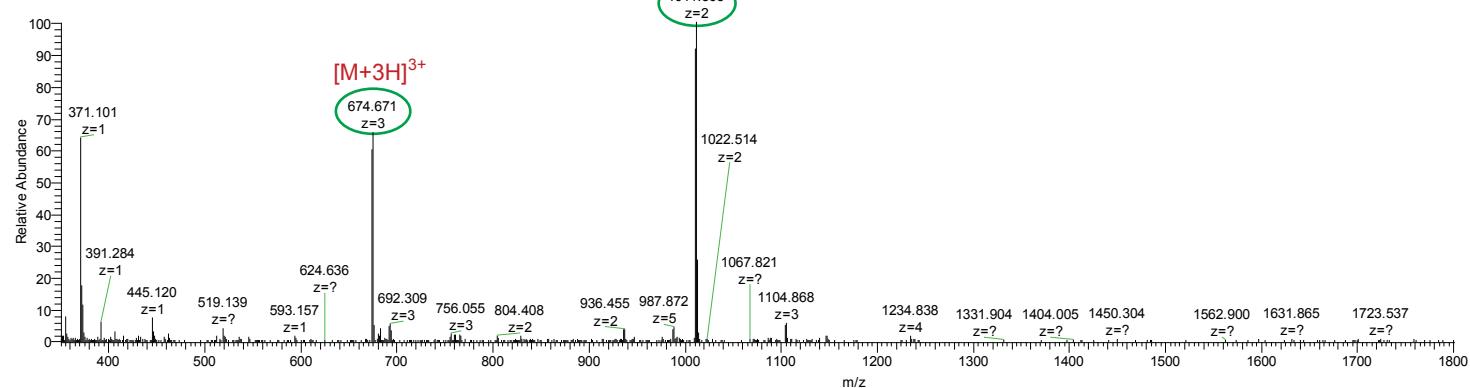
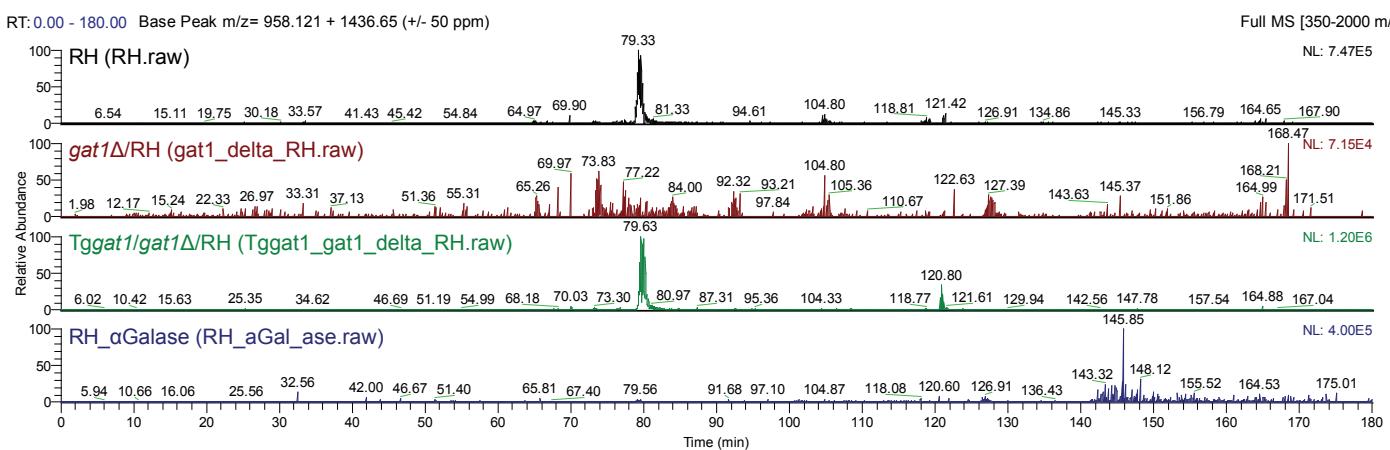


Figure S5C,D

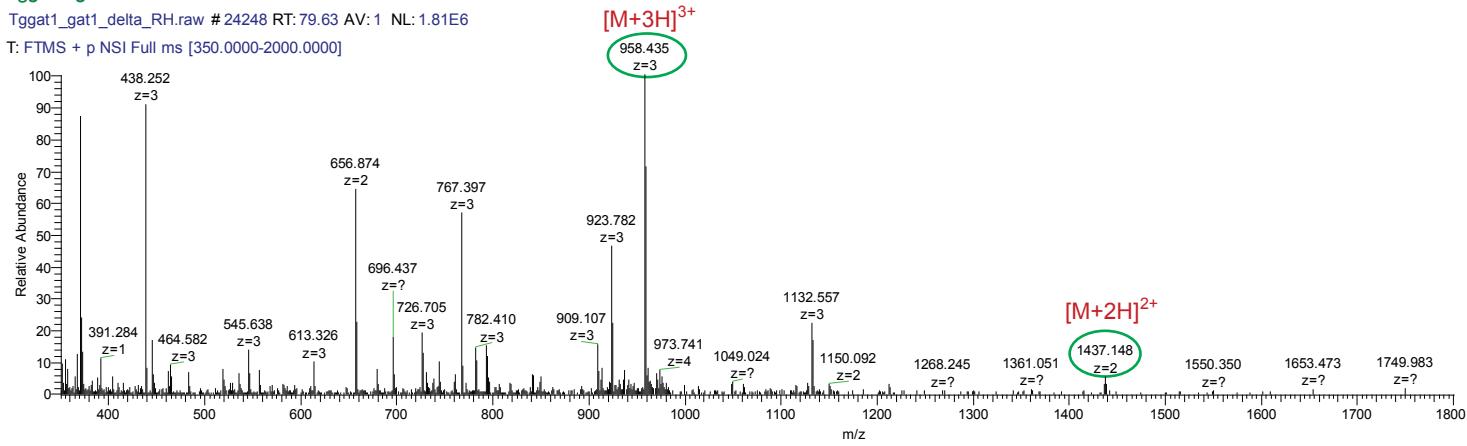
C. IFNIVNDFT(HyP)EEEAQVR + pentasaccharide



Tggat1/gat1Δ/RH

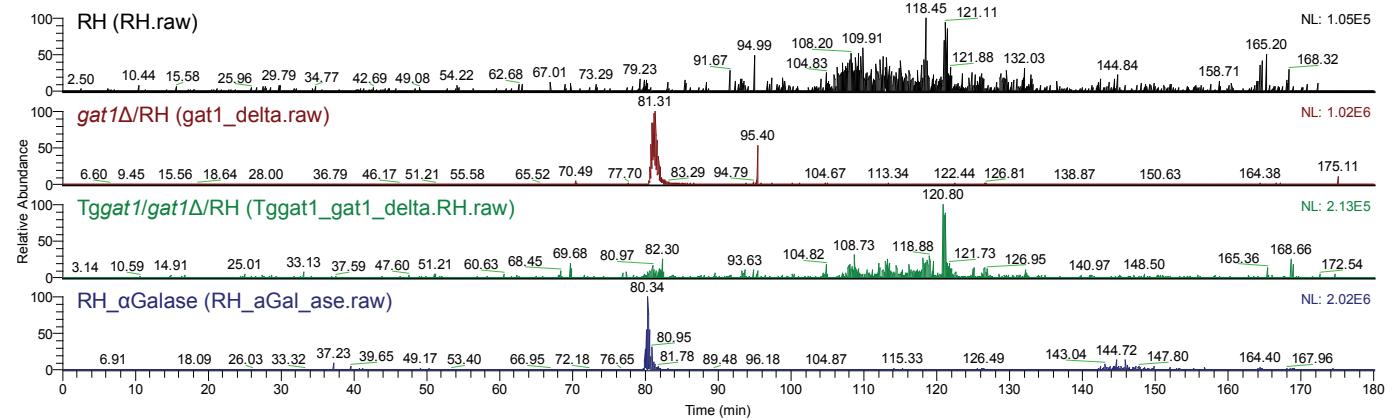
Tggat1_gat1_delta_RH.raw # 24248 RT: 79.63 AV: 1 NL: 1.81E6

T: FTMS + p NSI Full ms [350.0000-2000.0000]



D. IFNIVNDFT(HyP)EEEAQVR + tetrasaccharide

RT: 0.00 - 180.00 Base Peak m/z = 904.085 + 1355.624 (+/- 50 ppm)



gat1Δ/RH (gat1_delta.raw)

gat1_delta_RH.raw # 23551 RT: 81.13 AV: 1 NL: 1.63E6

T: FTMS + p NSI Full ms [350.0000-2000.0000]

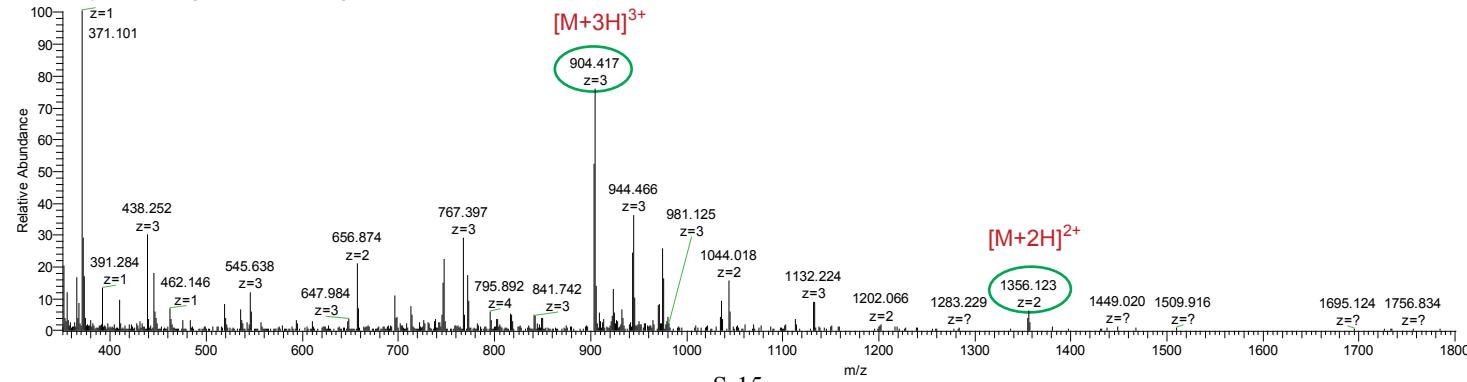
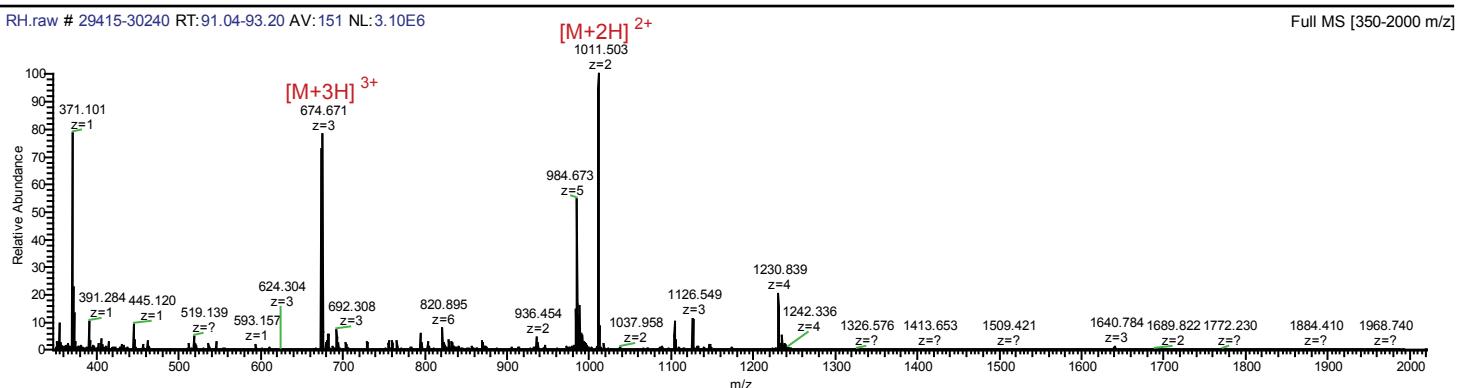
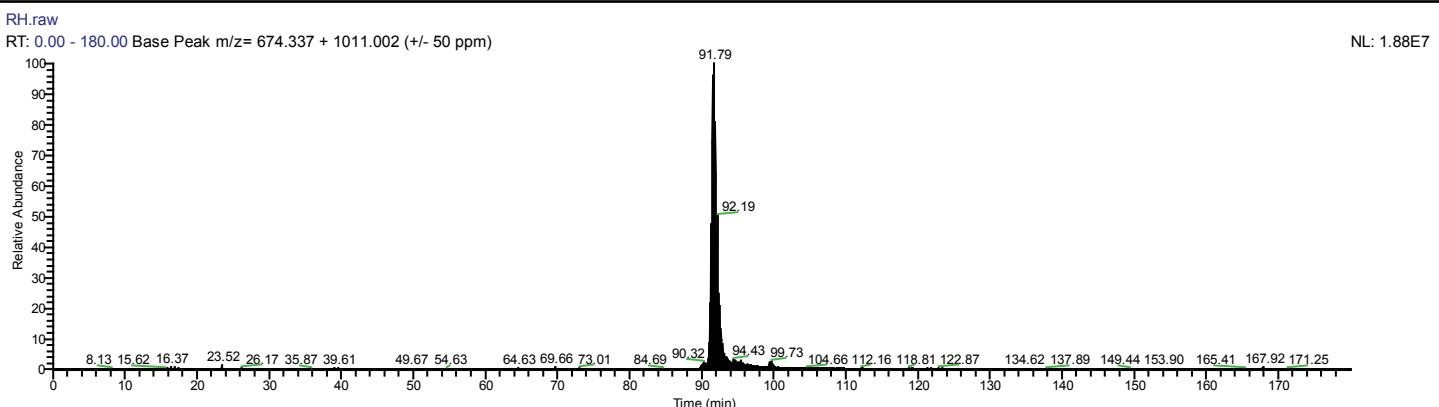


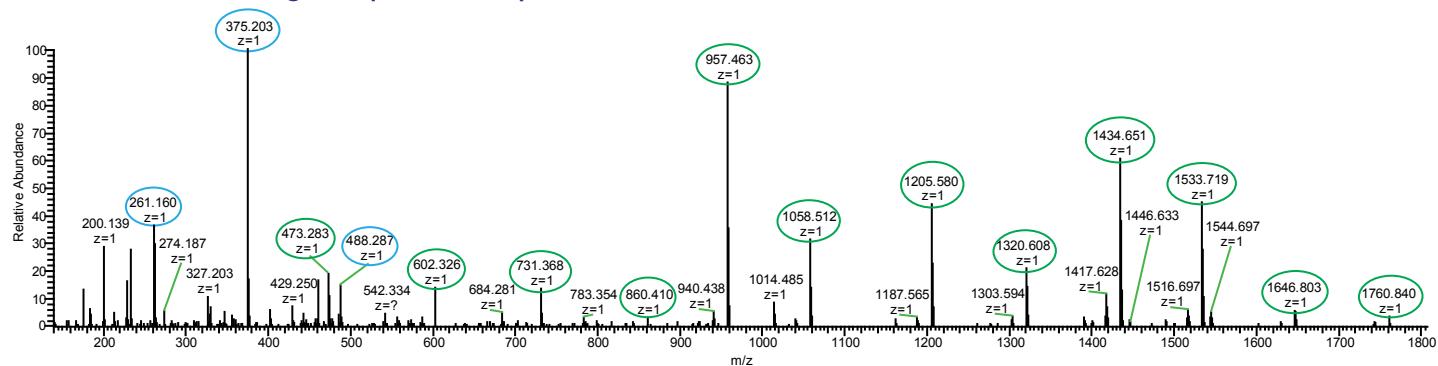
Figure S5E. MS2 of unmodified peptide from RH

IFNIVNDFT(Pro)EEEAQVR (unmodified)

RH (RH.raw)



RH.raw # 29603 RT:91.52 AV:1 NL:1.76E6
T: FTMS + c NSI d Full ms2 1011.5025@hcd30.00 [139.0000-2085.0000]



IFNIVNDFT(Pro)EEEAQVR

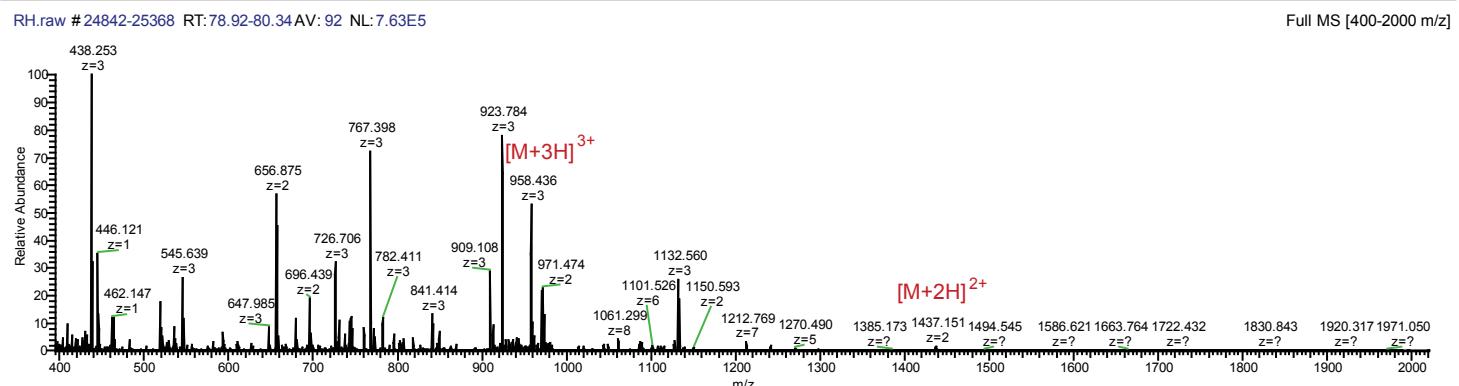
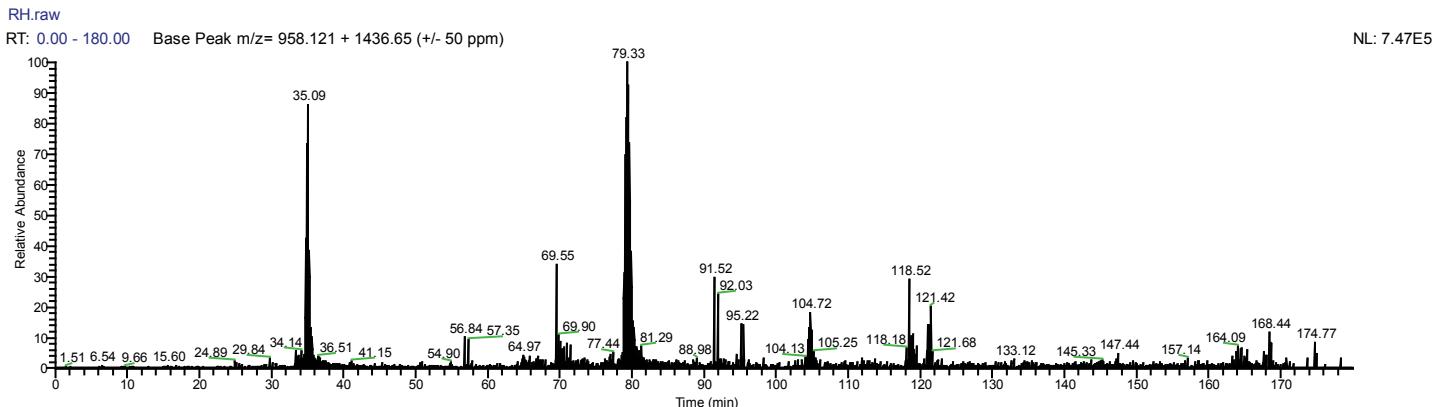
unmodified		
	b	y
I 1	114.0914	2020.997
F 2	261.1598	1907.913
N 3	375.2027	1760.845
I 4	488.2868	1646.802
V 5	587.3552	1533.718
N 6	701.3981	1434.65
D 7	816.4251	1320.607
F 8	963.4935	1205.58
T 9	1064.541	1058.511
P 10	1161.594	957.4636
E 11	1290.637	860.4109
E 12	1419.679	731.3683
E 13	1548.722	602.3257
A 14	1619.759	473.2831
Q 15	1747.817	402.246
V 16	1846.886	274.1874
R 17	2002.987	175.119

b fragments circled in blue, y fragments circled in green; detected fragments bold.

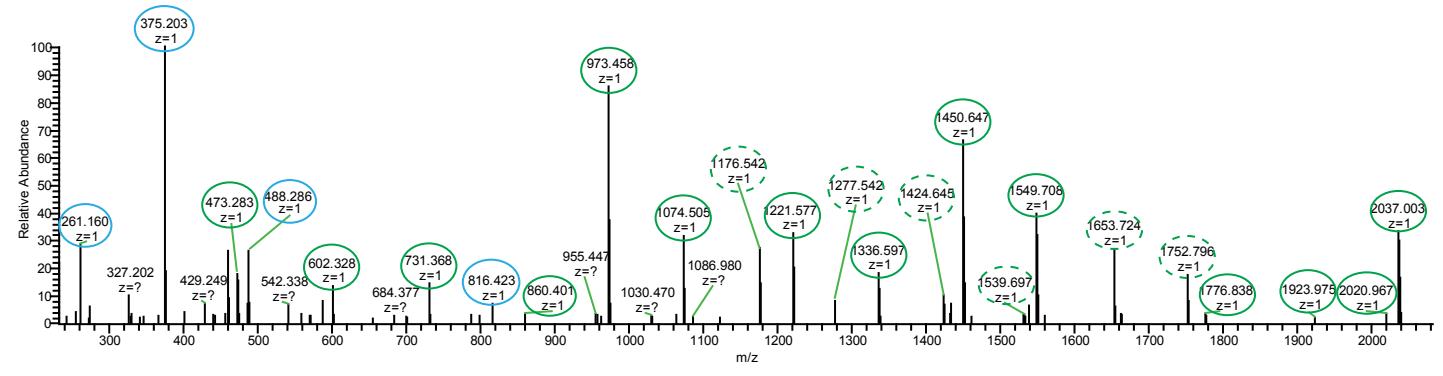
Figure S5F. MS2 of pentasaccharide peptide from RH

IFNIVNDFT(HyP+HexNAc+Fuc+Hex3)EEEAQVR

RH (RH.raw)



RH.raw # 24953 RT: 79.22 AV: 1 NL: 5.48E4
T: FTMS + c NSI d Full ms2 958.4366@hcd30.00 [197.0000-2955.0000]



IFNIVNDFT(Pro)EEEAQVR

	unmodified		Hyp		Hyp+HexNAc		Hyp+penta		
	b	y	b	y	b	y	b	y	
I 1	114.0914	2020.997	114.0914	2036.997	114.0914	2239.997	114.0914	2872.293	17
F 2	261.1598	1907.913	261.1598	1923.913	261.1598	2126.913	261.1598	2759.209	16
N 3	375.2027	1760.845	375.2027	1776.845	375.2027	1979.845	375.2027	2612.141	15
I 4	488.2868	1646.802	488.2868	1662.802	488.2868	1865.802	488.2868	2498.098	14
V 5	587.3552	1533.718	587.3552	1549.718	587.3552	1752.718	587.3552	2385.014	13
N 6	701.3981	1434.65	701.3981	1450.65	701.3981	1653.65	701.3981	2285.946	12
D 7	816.4251	1320.607	816.4251	1336.607	816.4251	1539.607	816.4251	2171.903	11
F 8	963.4935	1205.58	963.4935	1221.58	963.4935	1424.58	963.4935	2056.876	10
T 9	1064.541	1058.511	1064.541	1074.511	1064.541	1277.511	1064.541	1909.807	9
P 10	1161.594	957.4636	1177.594	973.4636	1380.594	1176.464	2012.89	1808.76	8
E 11	1290.637	860.4109	1306.637	860.4109	1509.637	860.4109	2141.933	860.4109	7
E 12	1419.679	731.3683	1435.679	731.3683	1638.679	731.3683	2270.975	731.3683	6
E 13	1548.722	602.3257	1564.722	602.3257	1767.722	602.3257	2400.018	602.3257	5
A 14	1619.759	473.2831	1635.759	473.2831	1838.759	473.2831	2471.055	473.2831	4
Q 15	1747.817	402.246	1763.817	402.246	1966.817	402.246	2599.113	402.246	3
V 16	1846.886	274.1874	1862.886	274.1874	2065.886	274.1874	2698.182	274.1874	2
R 17	2002.987	175.119	2018.987	175.119	2221.987	175.119	2854.283	175.119	1

b fragments circled in blue, y fragments circled in green; detected fragments in bold.
Specific HexNAc fragments dashed in green.
No specific pentasaccharide fragments detected.

Figure S5G. MS(2) of tetrasaccharide peptide fromm *gat1Δ/RH*

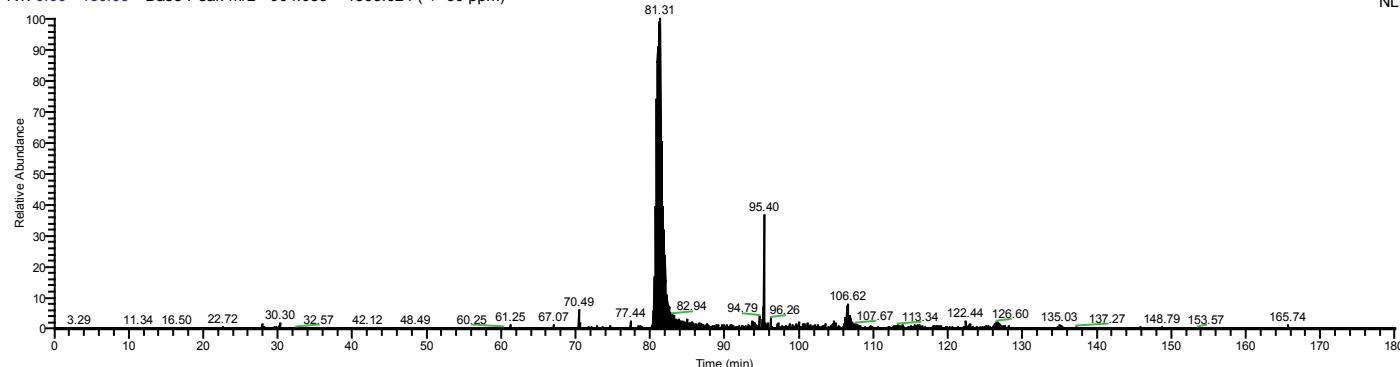
IFNIVNDFT(HyP+HexNAc+Fuc+Hex2)EEEQVR

gat1Δ/RH (*gat1_delta_RH.raw*)

gat1_delta_RH.raw

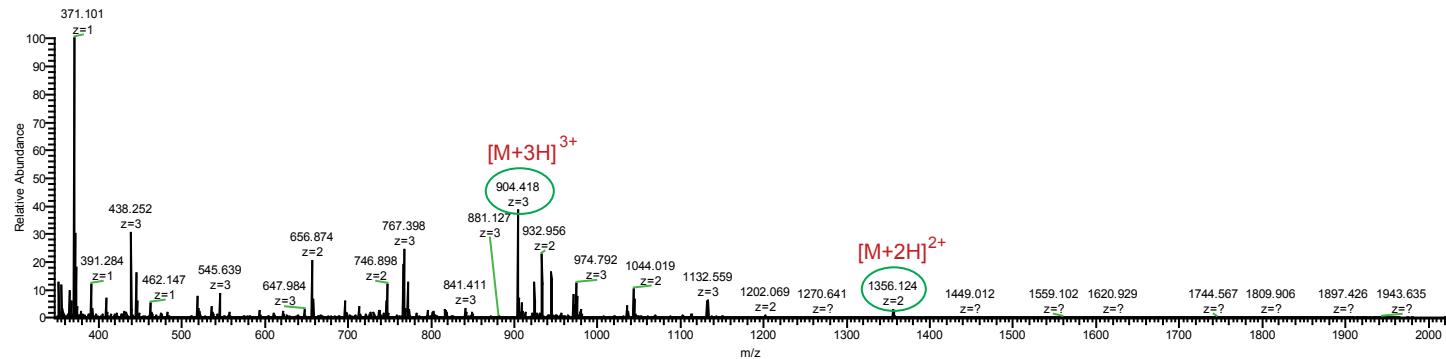
RT: 0.00 - 180.00 Base Peak m/z = 904.085 + 1355.624 (+/- 50 ppm)

NL: 9.45E5



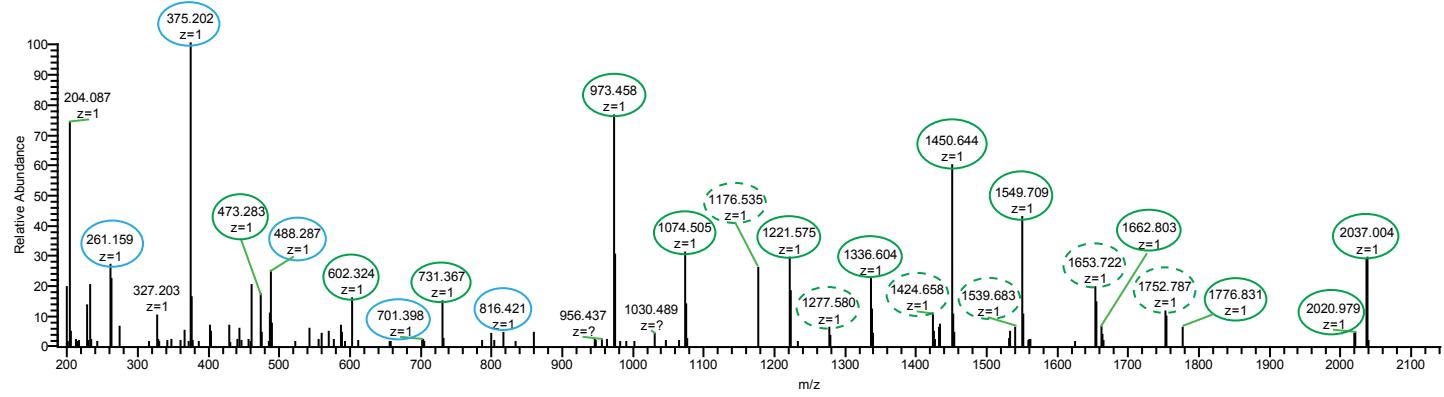
gat1_delta_RH.raw # 23378-23700 RT: 80.66-81.54 AV: 56 NL: 1.64E6

Full MS [350-2000 m/z]



gat1_delta_RH.raw # 23497 RT: 80.98 AV: 1 NL: 9.30E4

T: FTMS + c NSI d Full ms2 904.4189@hcd30.00 [186.0000-2790.0000]



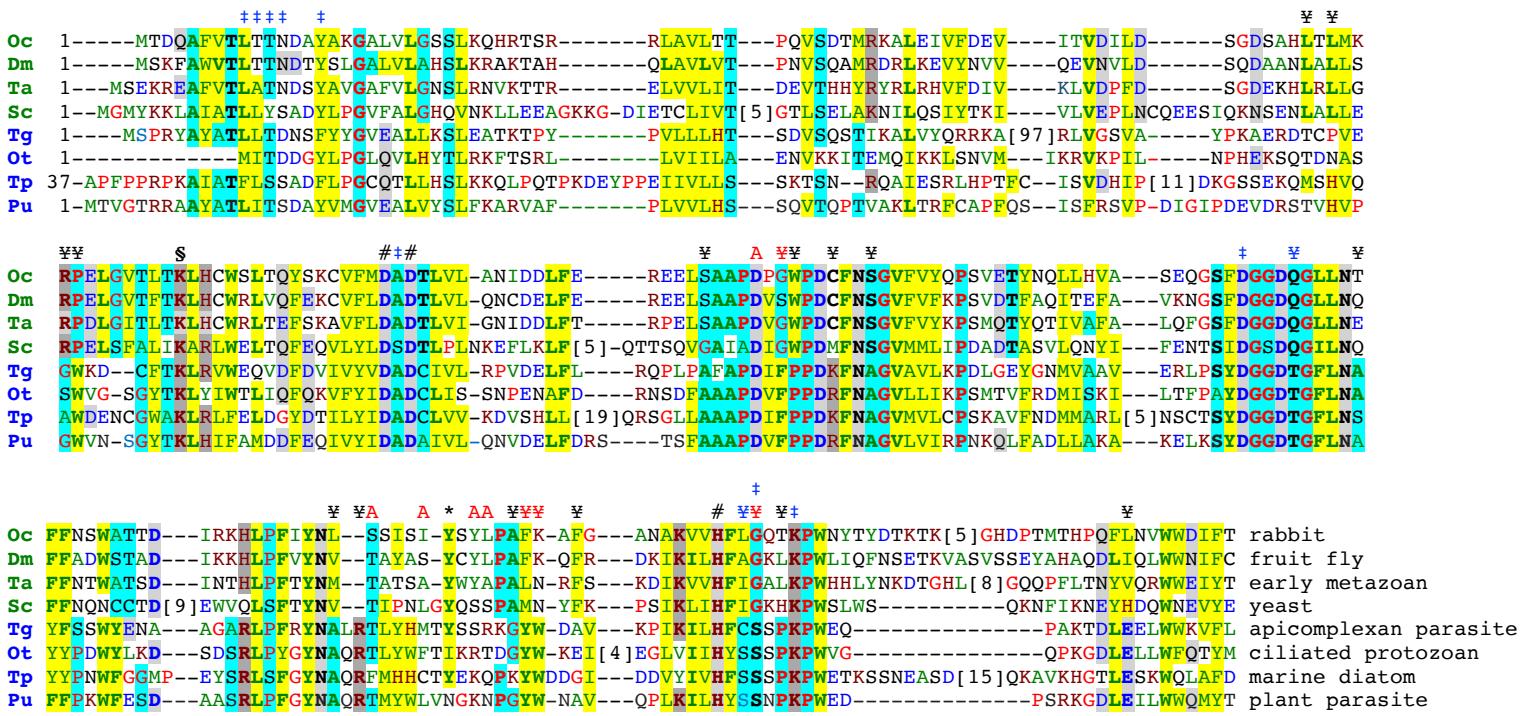
IFNIVNDFT(Pro)EEEQVR

	unmodified		Hyp		Hyp+HexNAc		Hyp+tetra		
	b	y	b	y	b	y	b	y	
I 1	114.0914	2020.997	114.0914	2036.997	114.0914	2239.997	114.0914	2710.24	17
F 2	261.1598	1907.913	261.1598	1923.913	261.1598	2126.913	261.1598	2597.156	16
N 3	375.2027	1760.845	375.2027	1776.845	375.2027	1979.845	375.2027	2450.088	15
I 4	488.2868	1646.802	488.2868	1662.802	488.2868	1865.802	488.2868	2336.045	14
V 5	587.3552	1533.718	587.3552	1549.718	587.3552	1752.718	587.3552	2222.961	13
N 6	701.3981	1434.65	701.3981	1450.65	701.3981	1653.65	701.3981	2123.893	12
D 7	816.4251	1320.607	816.4251	1336.607	816.4251	1539.607	816.4251	2009.85	11
F 8	963.4935	1205.58	963.4935	1221.58	963.4935	1424.58	963.4935	1894.823	10
T 9	1064.541	1058.511	1064.541	1074.511	1064.541	1277.511	1064.541	1747.754	9
P 10	1161.594	957.4636	1177.594	973.4636	1380.594	1176.464	1850.837	1646.707	8
E 11	1290.637	860.4109	1306.637	860.4109	1509.637	860.4109	1979.88	860.4109	7
E 12	1419.679	731.3683	1435.679	731.3683	1638.679	731.3683	2108.922	731.3683	6
E 13	1548.722	602.3257	1564.722	602.3257	1767.722	602.3257	2237.965	602.3257	5
A 14	1619.759	473.2831	1635.759	473.2831	1838.759	473.2831	2309.002	473.2831	4
Q 15	1747.817	402.246	1763.817	402.246	1966.817	402.246	2437.06	402.246	3
V 16	1846.886	274.1874	1862.886	274.1874	2065.886	274.1874	2536.129	274.1874	2
R 17	2002.987	175.119	2018.987	175.119	2221.987	175.119	2692.23	175.119	1

b fragments in blue, y fragments in green; detected fragments in bold.
Specific HexNAc fragments dashed in green.
No specific tetrasaccharide fragments detected.

Figure S6. Alignment catalytic domains of Gat1-like sequences and glycogenins (supports Fig. S8). To facilitate visualization of relatedness, acidic residues are in blue, basic in dark red, Gly and Pro in red, and hydrophobic in green, as previously described (88). Positions possessing a consensus chemical characteristic are highlighted in yellow (hydrophobic), gray (acidic), dark grey (basic), or teal (small). Positions of near perfect conservation are bolded.

green names- glycogenin like
blue names- Gat1-like



metal binding
\$ present in all GT8 sequences and catalytically essential
‡ & § sugar nucleotide binding (either nucleotide or sugar)
¥ group-specific
A & ¥ hydrogen bond and hydrophobic packing contacts with GlcGaGn- acceptor in PuGat1
* autoglycosylation site in glycogenin

Oc: *Oryctolagus cuniculus*
Dm: *Drosophila melanogaster*
Ta: *Trichoplax adhaerens*
Sc: *Saccharomyces cerevisiae*

Tg: *Toxoplasma gondii*
Ot: *Oxytricha trifallax*
Tp: *Thalassiosira pseudonana*
Pu: *Pythium ultimum*

Fig. S7. Summary of Gat1-related sequences selected for phylogenetic analysis (basis for Fig. 3B). The best scoring hits (based on BLAST) from different categories of Gat1-like sequences were selected for manual alignment and phylogenetic analysis. *A*, Predicted Gat1 sequences, from protists that have PgtA-like sequences but not AgtA-like sequences. *B*, Glycogenin and glycogenin-like sequences. *C*, Closest CAZy GT8 sequences from vascular plants. *D*, Closest CAZy GT8 sequences from organisms (protists) that possess Gnt1-like but not PgtA-like sequences. *E*, Closest CAZy GT8 from protists that possess PgtA-like sequences but lack apparent Gat1. *F*, Closest CAZy GT8 sequences from prokaryotes. Expect values, gene IDs, and known functions are indicated.

A. Gat1-like sequences from PgtA containing Protists	B. Glycogenin-like sequences
<i>Toxoplasma gondii</i> EPR60889.1	<i>Trichoplax adhaerens</i> (E ⁻²⁷) XP_002116183.1
<i>Hammondia hammondii</i> XP_008886569.1	<i>Amphimedon queenslandica</i> (E ⁻³⁰) XP_003383748.1
<i>Neospora caninum Liverpool</i> XP_003885051.1	<i>Nematostella vectensis</i> (E ⁻²⁴) XP_001625718.1 (Simplest animals)
<i>Ectocarpus siliculosus</i> CBJ26265.1	<i>Saccharomyces cerevisiae</i> (E ⁻¹³) (yeast) E7QGE5 (known function: primes glycogen synthesis)
<i>Albugo laibachii</i> CCA19642.1	<i>Monosiga brevicollis</i> (E ⁻²⁷) (choanoflagellate) XP_001744585.1
<i>Vitrella brassicaformis</i> CEM34465.1	<i>Capsaspora owczarzaki</i> (Filesterea) XP_004349815.2
<i>Nannochloropsis gaditana</i> EWM28655.1	<i>Helobdela robusta</i> (E ⁻²⁷) (anneleid) XP_009013909.1
<i>Oxytricha trifallax</i> EJY67427.1	<i>Drosophila melanogaster</i> (E ⁻²⁶) (fruit fly) NP_001163232.2 (known function: primes glycogen synthesis)
<i>Styloynchia lemnae</i> CDW86810.1	<i>Mus Musculus</i> (E ⁻²³) (animal) NP_038783.1
<i>Thalassiosira pseudonana</i> XP_002291959.1	<i>Homo sapiens</i> (E ⁻²³) (animal) AAH31096.2 (known function: primes glycogen synthesis)
<i>Styloynchia lemnae</i> CDW86810.1	D. Gat1-like sequences (E value <10⁻⁵) from the protists that have Gnt1 but not PgtA
<i>Reticulomyxa filosa</i> X6P0J2	<i>Acanthamoeba castellanii</i> (E ⁻¹⁰) XP_004352787.1
<i>Bigowiella natans</i> JGI: aug1.92_g19606	<i>Cyanidioschyzon merolae</i> (E ⁻²²) (red Alga) XP_005535960.1
<i>Sarcocystis neurona</i> SN3_01500095	<i>Galdieria sulphuraria</i> (E ⁻²¹) (red Alga) XP_005708321.1
<i>Karenia brevis</i> EX959504.1	<i>Volvox carteri</i> (E ⁻¹⁵) (green algae) XP_002954821.1
<i>Pythium ultimum</i> K3WC47	<i>Phytophthora infestans</i> (E ⁻¹³) XP_002997946.1
<i>Aphanomyces euteiches</i> (Aphanodb2: Ae201684 9096.1)	<i>Naegleria gruberi</i> (E ⁻¹³) XP_002672734.1
C. Closest Gat1-like sequences from plants	<i>Saprolegnia diclina</i> (E ⁻⁷) XP_008603979.1
<i>Arabidopsis thaliana</i> (E ⁻¹⁶) NP_175891.1	<i>Chlorella variabilis</i> (E ⁻¹⁰) XP_005850943.1
<i>Oryza sativa</i> (E ⁻¹⁷) A2XDA4	<i>Trichomonas vaginalis</i> (E ⁻¹⁰) XP_001309036.1
E. Gat1-like GT8 sequences from organisms that have PgtA but not Gat1	F. Gat1 like sequence from bacteria
<i>Dictyostelium discoideum</i> (E ⁻⁶) Q54L24	<i>Rhizobium meliloti</i> (E ⁻¹⁹) WP_029616784.1
<i>Albugo laibachii</i> (E ⁻¹⁴) F0W520	
<i>Bigowiella natans</i> (E ⁻⁵)	
<i>Guillardia theta</i> CCMP2712 (E ⁻¹⁵) L1J9Y4	

Figure S8. Alignment of glycogenin-like, Gat1-like, and other CAZy GT8 sequences used to construct the phylogenetic tree in Fig. 2. The amino acid sequence of Gat1-like proteins described in Fig. S7 (middle panel) were aligned with the amino acid sequences of representative known and predicted glycogenins (top panel) or CAZy GT8 sequences (bottom panel) as described in “Experimental Procedures”. Species names are spelled out at the bottom, and sequence sources are listed in Fig. S7. 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). Positions occupied by identical amino acids across all the organisms are bolded. Unique motifs that are specific for glycogenins are boxed in blue color, and Gat1-specific motifs are boxed in red.

	10 20 30 40 50 60 70
<i>Hs</i>	QAFVTLTTND AYAKGALVG SSLKQHRTTR RLVVLATLTL MKRPELGVTI TKLHCWSLTQ YSKCVFMDAD
<i>Mm</i>	QAFVTLTTND AYAKGALVG SSLKQHRTTR RMVVLTSLTL MKRPELGITL TKLHCWSLTQ YSKCVFMDAD
<i>DM</i>	FAFWVTLTTND TYSLGALVLA HSLKRKTAH QLAVLVTLLAL LSRPELGVTF TKLHCWRLVQ FEKCVFLDAD
<i>Mb</i>	QAYVTLCTND AYVVGAMLLA HSLRRTGTRR QIVCMITLGL LQRPELGVTI TKLHAWKLTH YDNCVFLDAD
<i>Hr</i>	-AYVTMATND VYAVGALVLA ETLRQTNQQ DLVIMITLSL LQRSELGVTF TKIQAWRLVE YRKCVFMDAD
<i>Ta</i>	EAFVTLATND SYAVGAFVLG NSLRNVKTTR ELVVLITLRL LGRPDLGITL TKLHCWRLTE FSKAVFLDAD
<i>Nv</i>	EAFVSLVTND NYANGALVG YSLRRVNNTTR KLALLVTLAL LSRPELGITF TKIRCWNLTH YQKCVFMDAD
<i>Co</i>	EAFVTLVTND GYALGALVLA KSLRDVNNTTR KIAVLITLAL LGRPELGVTI TKIYAWKLQ FTKCVFLDAD
<i>Pm</i>	ETYMTLVLTD SYLIGSQVLA WSLRDGSKK HLTALVTLYL LGRPDLRSSF TKIHIWAQEKF FKIIYLDAD
<i>Aq</i>	EAYVSLATNN DYCCHGAIALA CSLRLNTNSR KLCLLISLAL IKRPELGVTI SKLHIWRLVH YSKCVFLDAD
<i>Sc</i>	LAIATLLYSA DYLPGVFALG HOVNKLKGDI ETCLIVTLAL LERPELSFAL IKARLWEI TQ FEQVLYLDSD
<i>Tg</i>	YAYATLLTDN SFYYGVEALL KSLEATKTPY PVLLLHTVGS VAYPKAEDCF TKLRRWEQVD FDVIVYV DAD
<i>Hm</i>	YAYATLLTDN SFYYGVEALL KSLEATKTPY PVLLLHTVGS VAYPKAEDCF TKLRRWEQVD FDVIVYV DAD
<i>Nc</i>	YAYATLLTDN SFYYGVEALL KSLEATKTPY PVLLLYTVGS IAYPEKENCF TKLRAWEQVD FDVIVYI DAD
<i>Sn</i>	KAYATLLDD SFFYGVAAALI RSLAKTRTRY PLLLLHTVVE VRGPAKARLY TKLRLWEQED FDLLVYI DAD
<i>Kb</i>	EAYVSLLTSD SFLMAVQALI ASLKATGTAR RLLLLHTVAA IPNPHQTSGF TKLRRWEQVD FDKLVYI DAD
<i>Vb</i>	CAYITLLTD SFAIGVETLA FSLRKTGTPH PFIVLGVGD IANPNAESGF TKLHVWSLTE FQRVYI DAD
<i>Tp</i>	KAIATFLSSA DFLPGCQTLL HSLKKQLPQT PIIVLLSDNN NSDNNDKCGW AKLRLFELDG YDTILYI DAD
<i>Rf</i>	YAVVSLVTSE SYVVGAVQLI HSLHRNGGLK GSNVLVTVSE IPNPLEKSGY TKLrifemvq LKKLFYI DAD
<i>Bn</i>	YGYVSLLTND GFLPGAIVLA KSSLKVEARY PGAVMVTIPI EPLPCPNVGL TKLRRWQLGD FAKVYI DAD
<i>Pu</i>	AAyatlitSD AYVMGVEALV YSLFKARVAF PLVVLHSVPD IGIPDEVSGY TKLHIFAMDD FEQIVYI DAD
<i>Ot</i>	-----MITDD GYLPGQLQVLH YTLRKF-TSR LLVIIILAVKP ILNPHEKSGY TKLYIWTLQ FQKVFYI DAD
<i>S1</i>	-----MITED SYLPGQLQVMH YSLRKF-TQR TLVVIMTVKP IGNPNEKSGY TKFYIWSLTQ YKRIFYI DAD
<i>Ws</i>	-----MVTSD DFVIGAEVML HSLREHSTRR PLVVMVTVEP IAMPMKRVGY TKLRRWGLI Q FRCVYI DAD
<i>Ae</i>	KTFATLVTSD DFVIGVQVLA YSLRKHGAKY PLIVLYTVEA LPNPNVHSGY TKLHVFNLVE FSTVFYI DSD
<i>A1</i>	QAYATMITSDFQMGEVALL YSWSCTHSSI NFLILYTVDS IPIPASSSSAY TKNLIFGLEE YQKIVYI DAD
<i>Ng</i>	HAFVTLTGP GAQVLLHSLR TSISAKVAIR PVVVLVTVEP IANPYAESGF TKLQIWGL TQ FERVVYLD AD
<i>Gt</i>	EAYATLITTK EYIQAIVLS RIVKSTDEER PFIALVLVPR VKRPTGATTY SKLFVWNLTA YRLVLYLD AD
<i>At</i>	EAYATILHAH VYVCGAIAAA QSIRQSGSTR DLVILVDNPK AEKDAYNWNY SKFRLWQLTD YDKIIF IAD
<i>Os</i>	EAYATVLHSD TYLCGAIVLA QSIRRAGSTR DLVLLHDNPR AERGTYNYY SKFRLWQLTD YDRVVF DAD
<i>Dd</i>	NVYVTFADNA EYLKGIVALR MSMINTKCNY GLIVFTIEM VDIPKEVPAF TKFRAWQLVE YERVIWL DSD
<i>Tv</i>	YAFATVT-TP AFCMGAVVLG YTLRKYGNDY SYLCLVTVND A-KPYLWRSW IKLELWTFTYEKIVY LTD
<i>Cv</i>	MARRGSTWPD SYLMGVQALA RSLLAAQAQH PLLVMYTVER YV-PAGHECW NKLRIWELEE YERLAY IAD
<i>Sd</i>	RAYATLVCTD AYAIGAQVLR ASLHRVGSTL PLVVLVTYDV APIPLRSHAW AKLRFVFELEM FDTIVFL DAD
<i>Ac</i>	EAFVTLSSR SYYPGVVALA RSLRQFS-A-R ELLVLTTPV ERVPPPDCF TKFRMFELKN YTKFVY LDAD
<i>Ba</i>	EAYVTHLTND QYIKGAQVLA ESLREAGATR PPLAMITVPE FGDGRKDGF TKLEAWRLPC -TRVIY LTD
<i>Ab</i>	FAYVTVHYDQ EYVLGIQVLM QSIKLSGTRH DLVVLVSVD ITNPFLNHTL NKLHVWNLL EYDRVVY LDAD
<i>Pi</i>	FAYVTVHYDA EYVLGVQVMM HS1KLTGSPY DLVVLASVTN IDNPFVGYTL NKLHVWNMLE YERVVY LDAD
<i>Ng</i>	YAYATLVSSE GYLSGALAMY KSIIARGGKY DLVLLVTASY IDNPNAKDTY NKLH1WKL DQ YKRLVF VDSD
<i>Vc</i>	EAYATLVYGE DFVLAARVLG QSLRESGTTR DMVALTTVAP VKNPGTYVY TKLYIFQMTE YKKIVFL DAD
<i>Gs</i>	YAYATLLCDD VMLPATRAWL QSLKMTNTSF PIVVVLVTP LEYPFTLCRY SKLHLWNLLN YDKVVY MDSD
<i>Cm</i>	YAYATLLCDE RMLRAVAALV HSLRVRNTSY PILVLTREP LPYPFALCRY AKLHLWSLT T YEKIVFL DGD
<i>Rg</i>	YAYITLVTNA DYAKGATALV RSLRLTKAA NIVVLTIAL APLADLGCFN CKLRLWQLTE YERIVF IAD

						
	80	90	100	110	120	130	140
<i>Hs</i>	TLVLANIDDL	FDREELSAAP	D P GWPDCFNS	G VFVYQPSVE	TYNQLLHLAS	EQGSFDGGDQ	GILNTFFSSW
<i>Mm</i>	TLVLSNIDDL	FEREELSAAP	D P GWPDCFNS	G VFVYQPSIE	TYNQLLHLAS	EQGSFDGGDQ	GLLNNTYFSGW
<i>DM</i>	TLVLQNCDDEL	FEREELSAAP	D V SWPDCFNS	G VVFVKPSVD	TFAQITEFAV	KNGSF DGGDQ	GLLNQFFADW
<i>Mb</i>	TLVLTNIDEL	FERNCFAAAP	D I GWPDCFNS	G VVFQPSA	KFEDLVRLLA	STGSFDGGDQ	GLLNNEYFADW
<i>Hr</i>	TLVLQNVDDL	FSRDPFAAAP	DAGWPDCFNS	G IFLYQPSFE	MYGDLLQFAL	KIGSF DGGDQ	GLLNLF FSDW
<i>Ta</i>	TLVIGNIDDL	FTRPELSEAAP	DVGWPDCFNS	G VFVYKPSMQ	TYQTIVAFAL	QFGSF DGGDQ	GLLNNEFFNTW
<i>Nv</i>	MLVLQNCDDEL	FDRCELSAVP	DIGWPDCFNS	G MVFEP SRA	THEALLKYAI	DHGSFDGGDQ	GLLN SFFS QW
<i>Co</i>	TLVVQNVDEL	FDRPEI AAAP	DVGWPDCFNS	G VFVFVPSAA	TFEKLAEHAV	STGSFDGGDQ	GLLNNTFFDYW
<i>Pm</i>	AFCLKNIDDEL	FDLDTFAAVP	DVGWPDI FNS	G VFITKPNIS	VYNSSLNLAK	NSISFDGGDQ	GLLN FYFSNW
<i>Aq</i>	TLVLTNVDEL	FEREEMSAAP	DIGWPDLFNS	G VFVFRPSLE	TFASLLELAD	KEGSYDGGDQ	GLLNLYWRDW
<i>Sc</i>	TLPLNKEFLL	FDIMSVGAIA	D I GWPDMFNS	G VMM LIPDAD	TASVLQNYIF	ENTSIDGSDQ	GILNOFFREW
<i>Tg</i>	CIVLRPVDEL	FLRQP PAFAP	D I FPPDKFNA	G VAVLKP DLG	EYGNMVA AVE	R LP SYDGGDT	GFLNAYFSSW
<i>Hm</i>	CIVLRPIDDL	FLRQP PAFAP	D I FPPDKFNA	G VAVLKP DLD	EYGMVA AVE	R LP SYDGGDT	GFLNAYFSSW
<i>Nc</i>	CIVLGPVDEL	FLRKPPAFAP	D I FPPDKFNA	G VVVLKP DLG	EYGMIA AIE	R LP SYDGGDT	GFLNAYFSSW
<i>Sn</i>	CVVLQNVDEL	FERLSPAF AA	D V FPPDRFNA	G VIVLQPNVE	LFSRMLRAAG	LLPAADGGDT	GFLNSFFSDW
<i>Kb</i>	CVVLERVDEL	FERPSPAF CP	D V FPPDKFNA	G VIVLSPSRE	LFEKMQERIA	ELPSHDGGDT	GFLNAFFPDW
<i>Vb</i>	CIVMRKIDCL	FDPAAPAFAP	D V FPPDRFNA	G VMVIEPSLA	VYEDLLAKRT	VLRSYD RGD T	GFLNAYFSGW
<i>Tp</i>	CLVVKDVS HL	LRVDSLAAAP	D I FPPDKFNA	G VMVLCPSKA	VFN DMMARLN	SCTS YDGGDT	GFLNSYYPNW
<i>Rf</i>	CIVVRDISDI	FKLPDFAAAP	D I LCPPDHFNA	G VLFIQPNVQ	TFQQLLRNVA	YVNSYDGGDT	GFLNSYFNDW
<i>Bn</i>	AI VVRNV DHL	FKMIPFAAAP	D I FPPDKFNA	G VVLVQPN SV	MFAYILRLAY	GLGSYDGGDT	GFLNRIFPRW
<i>Pu</i>	AI VLQNVDEL	FDRSTFAAAP	D V FPPDKFNA	G VVLVIRPNQ	LFADLLAKAK	ELKS YDGGDT	GFLNAFFPKW
<i>Ot</i>	CLISSNPENA	FDRNSFAAAP	D V FPPDKFNA	G VLLIKPSMT	VFRDMISKIL	TFPAYDGGDT	GFLNAYYPDW
<i>S1</i>	CLIMQN PENI	FLRDTFAAAP	D V FPPDKFNA	G VLYIEPSMK	IFTDLISKI Q	ILSTYDGGDT	GFLNAYFPNW
<i>Ws</i>	ALVMEDLDEL	FDREVFAAAP	D V FPPDKFNA	G VMV VVPSLI	VLEDMMSKVE	ELPSYDGGDT	GFLNAYFADW
<i>Ae</i>	AFVLANDEV	LERDIFAAAP	D I FPPDKFNA	G VLLIKPSNAE	LFQRLVQSQA	QFQS YDGGDT	GYLNAVFPDW
<i>A1</i>	ALILTNIDDEL	FEMDTFAAAP	D I FPPDKFNA	G VLVIKPGKD	VENLLAKAK	TIKS YDGGDT	GFLNLVFS DW
<i>Ng</i>	CLVVEDIQEL	FSADVFAAAP	D I FPPDKFNA	G VMLVRPNLD	VYEDMLRAVG	ALPSYDGGDT	GFLNAFFPKW
<i>Gt</i>	LLPLSSLAPL	FDRDVVAAP	D I SLFDHFNS	A LVL RP NLL	HLQRLLALSS	SLEPYDGGDT	GLLNNEFFNAW
<i>At</i>	LLILRNIDFL	FSMPEISATG	NNGTL--FNS	G VMVIEPCNC	TFQLLMEHIN	EIES YNGGDQ	GYLNEVFTWW
<i>Os</i>	ILVLRDLDAL	FGFPQLTAVG	NDGSL--FNS	G VMVIEPSQC	TFQSLIRQRR	TIRS YNGGDQ	GFLNEVFWWW
<i>Dd</i>	MLLKSLDHL	FDLVDLYAAI	DADANSCINS	G IMLLSPSID	VYNLLIDGMK	LPNQSTVNDQ	DVINTTLPHW
<i>Tv</i>	TLPTQRIDE L	FNHSELSCVS	DPMPPQICNT	G LLVLEPNLT	TFKHMKKLSD	LYANNPPGDQ	GFINFFFQF
<i>Cv</i>	MLVLRNIDHL	FALPPFYAAP	DCTAGRQFNA	G FFLVT P SRA	ELARFQSLV	RIGGY--AEQ	DLLNEV LHEF
<i>Sd</i>	MLCVRNMDDL	FDAIAAASRA	CTCNPQRFNS	G MLVLPSCA	TLESLLAKLR	SVERFVFSDQ	CFLNEA FPDF
<i>Ac</i>	MLVVGVDDEL	FSYPSFAAAP	NFQLKKSFNA	G LFVVRDRDEG	LHRQFLDH YH	YDKAWSWADQ	SLLNDFFKKW
<i>Ba</i>	ILAVGNPDVL	FELAQFAVQD	SQPHMQGPNT	G VMLKP DIR	VYARIVETLT	PLHEMPFYEQ	GFIGKFFAKW
<i>Ab</i>	NIVLRNADEL	FMCGPFCAVF	MNPCH--FHT	G LLVVT PDK	EYQRLLHQLE	YQSSFDGADQ	GFLSSVYSEL
<i>Pi</i>	NVLIRNSDEL	FLCGEFCAVF	MNPCH--FHT	G LLVVT PSSA	EYQRLLS ALG	HLESFDGADQ	GFLSSMYSML
<i>Ng</i>	CIIFKNVDLL	FNCVGVC SGS	DMGNTEFFNG	G IMVLEP STK	TYDDMMDKMP	AYKS YDGGEQ	GFINLYFDFH
<i>Vc</i>	VLVIRNM DVI	FKCPGFCAAL	RHSER--FNT	G VMSL VPSLE	MYDDMMAKMR	SMPS YTG GDQ	GFLNSYFPSF
<i>Gs</i>	MLVMQNIDNL	FVEFDLSACA	DLYPDT-FNS	G IMVIQPNET	TFRNM KAVYK	NVSS YNVGDQ	GFLNWFFGEW
<i>Cm</i>	TLVLAPIDDL	FEKYDLAAAP	DLYPET-FNS	G VMVLEPRHD	VYASMLARYR	ETPSY NLGDQ	GFLNSFFGQW
<i>Rg</i>	AI ILKNIDKL	FAYPEFSAAP	NVYETRRMNS	G VFVARPSEE	TFGRMLAMLD	QPDA FRRTDQ	TFLEAFFPDW

	150 160 170 180 190
<i>Hs</i>	ATTHLPFIYN LYSYLPAFKV FGASA----- -KVVHFLGRV KPWNYTHPEF LILWWN
<i>Mm</i>	ATTHLPFVYN LYSYLPAFKA FGKNA----- -KVVHFLGRT KPWNYTHPEF LNLWWD
<i>DM</i>	STAHLPFVYN VYCYLPAFKQ FRDKI----- -KILHFAGKL KPWLIQAQDL IQLWWN
<i>Mb</i>	ATQRLPFAYN MYGYAPAFER FKADI----- -KVIHFIGAR KPWMGM -----
<i>Hr</i>	ATKHLPTFTYN LYSYKPALKK FGDEI----- -KIVHYLGKP KPWDHENMEL LQLWWD
<i>Ta</i>	ATSHLPFTYN MYWYAPALNR FSKDI----- -KVVHFIGAL KPWHHLLTNY VQRWWE
<i>Nv</i>	SHEHLSFIYN MYTYAPAYKE FGKNV----- -KIVHFIGPV KPWQYSERSY IQLWWD
<i>Co</i>	PTARLSPFLYN MYSYKPAFKQ YGHLV----- -KIIHFIGQF KPWHWASEFH VQQWN
<i>Pm</i>	K--RLPFTYN VYQYFPAYYH FKSKI----- -SVIHFAGTK KPWMLSYNEL IEKWKS
<i>Aq</i>	SIRRLPFTYN VYSYPPAFLR HRKDM----- -KIIHFIGAI KPWHHRAEEF IRKWW
<i>Sc</i>	V--QLSFTYN VYQSSPAMNY FKPSI----- -KLIHFIGKH KPWSLWKNEY HDQWNE
<i>Tg</i>	YENRLPFRYN ALRFLYHMTY SSRKGYWDAV IKILHFCSSP KPWEQPKTDL EELWWK
<i>Hm</i>	YENRLPFRYN ALRFLYHMTY SSRKGYWNAV IKILHFCSSP KPWEQPKTDL EELWWK
<i>Nc</i>	YESRLPFRYN ALRFLYHMTY CSHKGYWNAV IKILHFCSSP KPWEQPKTDL EDLWWK
<i>Sn</i>	YMWRLPFRYN AQRSVYRFTG AAYRGYWEAI IKILHFTSTP KPWERPQTEL EDIWW
<i>Kb</i>	YRWRLPFRYN ALRFLMYWFTH KN-EGYWDSDL IKILHFCSSP KPWDPEKGDL EQLWW
<i>Vb</i>	YGWRLAFAHYN AQRFLMHWMTH SKQHGYWDEC LSVLHLSSSP KPWESPKGPT EWLWW
<i>Tp</i>	FGGRLSFGYN AQRFMHHCTY EKQHKGWDDG VYIVHFSSSP KPWETKHGTL ESKWQL
<i>Rf</i>	YHGRLDGFWN AQRIMEWYTR DK-HAYWDHI VRILHFSSSP KVWDIPSNRL HRQWHS
<i>Bn</i>	HSWRLHFGYN AQRFLHWFTK -KNEHKGWEWS LHIHYASSP KPWEVPTDKL EKIWWK
<i>Pu</i>	FESRLPFGYN AQRFLMYWLNV GKNHGYWNAV LKILHYSSNP KPWEDPKGDL EILWWQ
<i>Ot</i>	YLKRLPYGYN AQRFLYWFTH KRTDGWKEI LVIHYSSSP KPWVG-KGDL ELLWFQ
<i>S1</i>	FESRLPFGYN AQRFLYWFTH KRTDGWKEV IIIHYSSSP KPWSSQKGDL ELEWFK
<i>Ws</i>	FSRRLPFAYN ALRFLVYWTTH EKNHGYWEAI VKIIFCSSP KPWEETKGDL EMTWWQ
<i>Ae</i>	YTYRLPFAYN AQRFLMHWLTY AKKEGYWDAV VKVLHCSSSP KPWESPKGDL EMLWWQ
<i>A1</i>	FQRRLPFRYN AQRFLMYWMVN SKNHGYWKAV LKILHFSSSP KPWEPIGDL EMIWWM
<i>Ng</i>	YSSRLPFIWN AQRFLHWMTY AVAEGYWGAV VKILHFSSSP KPWEPEKGEL EVKWWT
<i>Gt</i>	YESRLGLELN LSRLHPRSWL RTLPRQRSNL SQVIHFSGGR RPWGIASVAA AALVWH
<i>At</i>	HRILKHFWIG DRKKTELFGA EPPVL----- -YVLHYLG-M KPWLCYTDIA HRKWWM
<i>Os</i>	HRLLKNFWAN TRALKERLFR ADPAE----- -WSIHYLG-L KPWTCYSDAA HARWWQ
<i>Dd</i>	RSLEYGVQIT HCTSEPRLWN F----- -TFLHFTAGP KPWSLLEPTCI EQIYLN
<i>Tv</i>	N--PLPTLYN VDTNFEFFLYE QKLI----- -KVVHFVC-K KPWKGCGMYSL NQVWWD
<i>Cv</i>	SAPPLPHTFN ARRHPQLWR ----- -OHWHAVAVA KPWQEGLYQDL VOLWWR
<i>Sd</i>	I--DVPYVFN APIAHPRLWQ LEDV----- -KAIHYIL-E KPHVHEYDDL YALWWE
<i>Ac</i>	N--QVPHYFN MFLYRPDLWE VDKI----- -KIIHYTG-G KPWQTPPYEP LFALWR
<i>Ba</i>	V--QLPAKYN FYLNRPLYQD IRHDN----- KVFIHYAK-C KPWDLSEFGKE YLRYIR
<i>Ab</i>	RKARLSPVGYN IYEQYHWKLF YLRHFATMTS RPIPAITIGL KPW ----- --YWWA
<i>Pi</i>	RKARLPVGYN IYEQYHWKLF YLRQFASMTS RPIPALTVGL KPW ----- --YWWA
<i>Ng</i>	RKSRIPYTWN TYYFFKYAYI QRLKK----- FRIIHYNLPI KPWKFLLIDA SYYWYE
<i>Vc</i>	AHSRLPTTFN ALYVVGSNRW MLPRS----- LYVIHYTLGF KPWVWWREN AWQAYR
<i>Gs</i>	SQRHIPLKYN VLKYRDTIMW GHVKD----- IKVLHFTGET KPWNFYEMRS YYAWVR
<i>Cm</i>	RANHLPLEYN TLKLRETIW ASLQR----- VRVVHFTGET KPWSWHDRI DPVFYY
<i>Rg</i>	HG--LPVYFN MLQYVWFTMP AL---WDWKS ISVLIHYQYE- KPWEKDHPKL IDLWHS

Species names. Sequence IDs are in Fig. S7.

Glycogenin-like:

Hs: *Homo sapiens*
Mm: *Mus musculus*
DM: *Drosophila melanogaster*
Mb: *Monosiga brevicollis*
Hr: *Helobdella robusta*
Ta: *Trichoplax adherens*
Nv: *Nematostella vectensis*
Co: *Capsaspora owczarzaki*
Pm: *Pneumocystis murina*
Aq: *Amphimedon queenslandica*
Sc: *Saccharomyces cerevisiae*

Gat1-like

Tg: *Toxoplasma gondii*
Hm: *Hammondia hammondi*
Nc: *Neospora caninum*
Sn: *Sarcocystis neurona*
Kb: *Karenia brevis*
Vb: *Vitrella brassicaform*
Tp: *Thalassiosira pseudonana*
Rf: *Reticulomyxa folisa*
Bn: *Bigelowiella natans*
Pu: *Pythium ultimum*
Ot: *Oxytricha trifallax*
Sl: *Styloynchia lemnae*
Ws: *Ectocarpus siliculosus*
Ae: *Aphanomyces euteiches*
Al: *Albugo laibachii*
Nq: *Nannochloropsis gaditana*

Other CAZy GT8 family

Gt: *Guillardia theta*
At: *Arabidopsis thaliana*
Os: *Oryza sativa*
Dd: *Dictyostelium discoideum*
Tv: *Trichomonas vaginalis*
Cv: *Chlorella variabilis*
Sd: *Saprolegnia diclina*
Ac: *Acanthamoeba castellanii*
Ba: *Bigelowiella natans GT8*
Ab: *Albugo laibuchi GT8*
Pi: *Phytophthora infestans*
Ng: *Naegleria gruberi*
Vc: *Volvox carteri*
Gs: *Galdieria sulphuraria*
Cm: *Cyanidioschyzon merolae*
Rg: *Rhizobium gallicum*

Fig. S9. Characterization of Gat1 enzyme activity and biochemical complementation of *T. gondii* extracts (supports Fig. 4). TgGat1 glycosyltransferase activity was assayed using 20 mM maltose-pNP as an acceptor in the presence of 4 μ M UDP-Glc, 2 mM MnCl₂, pH 7.0, and varying concentrations of NaCl or KCl (A, B). C, TgGat1 was assayed using 20 mM maltose-pNP in the presence of 5.2 μ M UDP-Gal, no added salt, 2 mM MnCl₂, pH 7.0, and the indicated divalent metal ions. D, TgGat1 was assayed using 20 mM maltose-pNP, 8 μ M UDP-Gal, no added salt, 2 mM MnCl₂, at different pH values. E, Donor specificity of TgGat1, based on the UDP-Glo assay in an overnight reaction that consumed all UDP-Gal. F, UDP-Gal concentration dependence of TgGat1 and PuGat1 Gal-transferase activity toward 20 mM maltose-pNP. Symbols are after Fig. 1. G, Concentration dependence of TgGat1 Gal-transferase activity on GlFGaGn-pNP concentration. Error bars represent \pm S.D. of 3 technical replicates of the same reaction. H, I, Biochemical complementation to detect Gat1 substrates. Desalted S100 extracts from strains RH $\Delta\Delta$, *gat1* Δ /RH $\Delta\Delta$, ME49 and *gat1* Δ /ME49 were reacted with recombinant Gat1 (rGat1) in the presence of UDP-[³H]Gal, and the product of the reaction was separated on an SDS-PAGE gel which was divided into 40 slices for liquid scintillation counting.

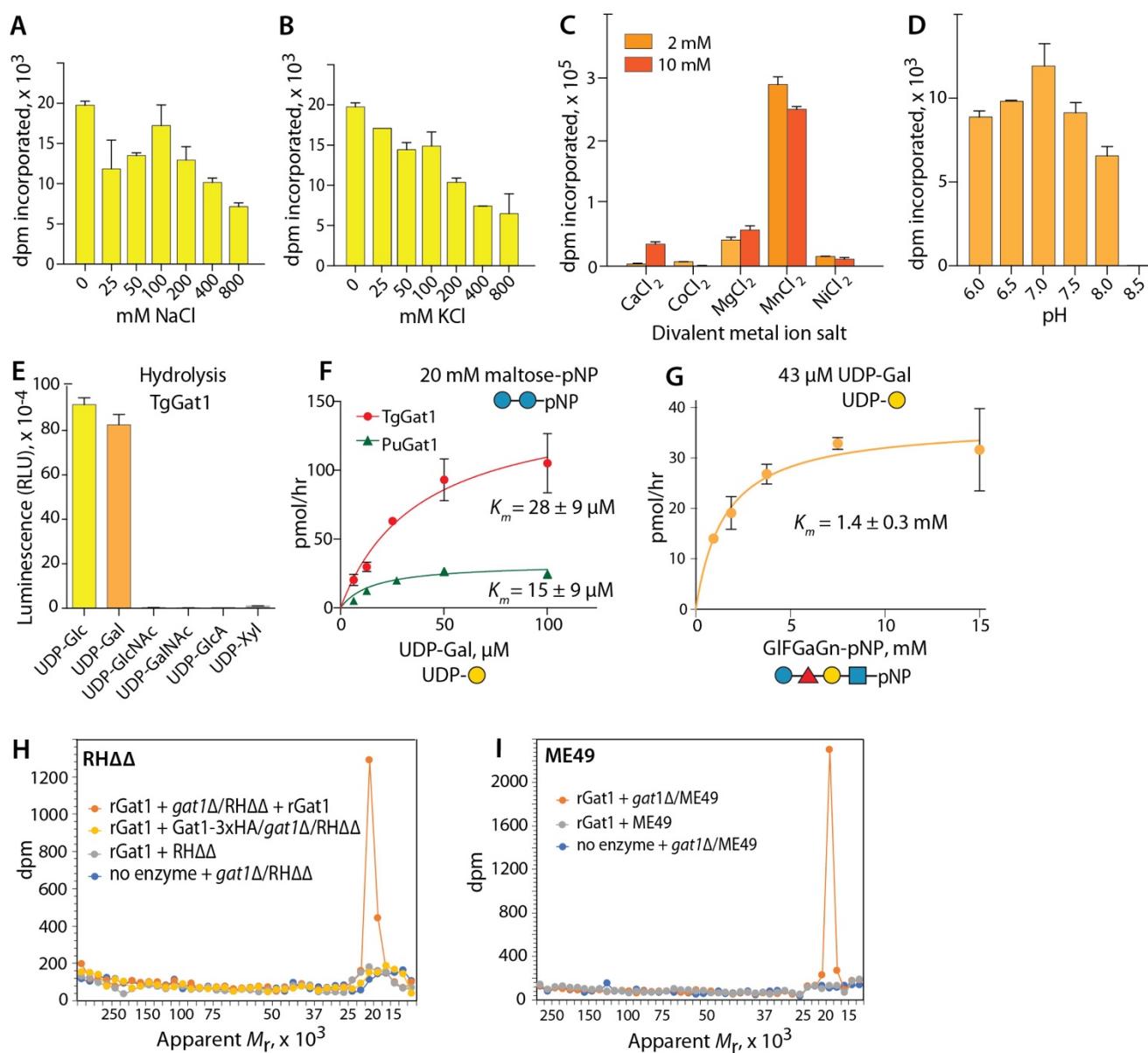


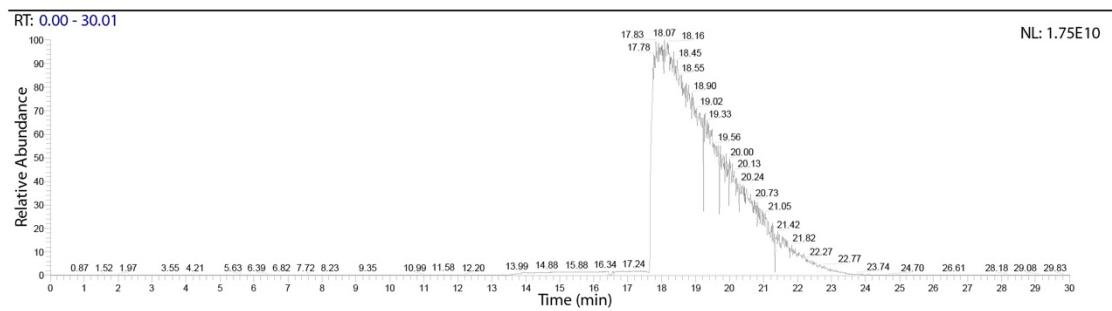
Fig. S10. TgGat1 lacks auto-glycosylation activity. PuGat1 and TgGat1 were prepared in *E. coli*, purified to near homogeneity (Fig. 4A), and analyzed by nLC/MS analysis. *A*, Total ion current for elution of PuGat1 in a gradient of acetonitrile from a C4 column. *B*, Mass spectrum showing multiply protonated species. Xtract deconvolution yielded virtually only one species with an M_r 30251.2596, which closely matched the predicted theoretical monoisotopic mass of M_r 30251.2603 (error= 0.02 ppm). *C*, Deconvolution of data in panel B using the ReSpect algorithm in BioPharma to yield a measurement of the average mass. *D*, SDS-PAGE and Coomassie blue staining of TgGat1, before and after incubation with UDP-Glc or UDP-Gal for 30 min. *E*, Summary of average mass measurements of TgGat1 and PuGat1 based on ReSpect deconvolution. After isolation from *E. coli*, both TgGat1 and PuGat1 yielded predominantly only the unmodified versions of the recombinant proteins, with M_r 39051.9687 for TgGat1 (theoretical average mass: 39051.9161, error= 1.3 ppm) and M_r 30269.1738 for PuGat1 (theoretical average mass: 30269.2901, error= 3.8 ppm). Their masses were essentially unaffected by *ex vivo* reaction in the presence of UDP-Gal or UDP-Glc.

The raw data files listed below are deposited in a data repository at <https://figshare.com/> Figshare ID: 10.6084/m9.figshare.12272909 (Recombinant Gat1 intact protein raw files Fig. S10). The data are analyzed in Panels A-C, E.

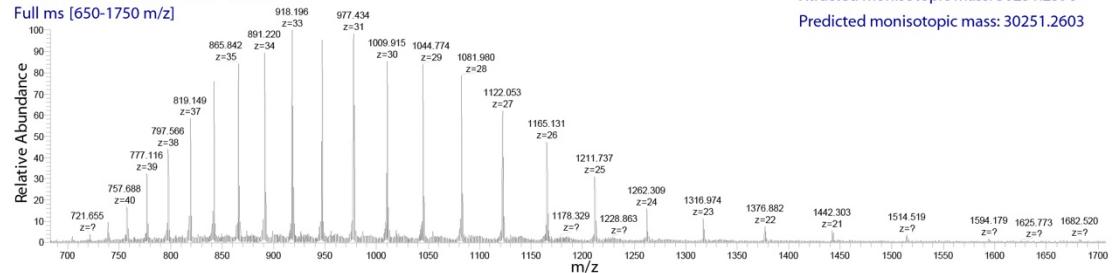
Samples analyzed: Original datafile:

TgGat1+Gal	TgGat1+Gal.raw
TgGat1+Glc	TgGat1+Glc.raw
TgGat1	TgGat1.raw
PuGat1	PuGat1.raw

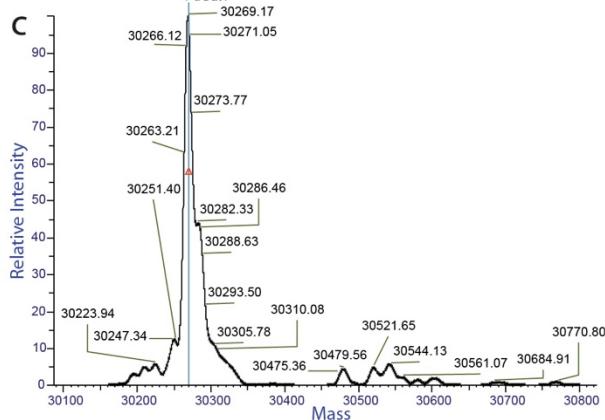
A PuGat1 TIC MS (400-2000 m/z)



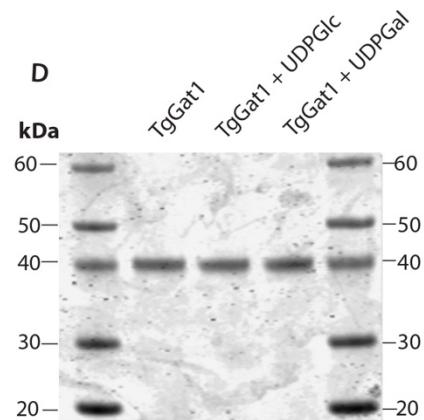
B PuGat1 RT:17.74-18.47 AV:78 NL: 2.12E7



C PuGat1



D



E

Sample Name	Protein Name	Average Mass	Theoretical Mass (Da)	ΔMass (ppm)	Intensity	Fractional Abundance	Score	Number of Charge States	Charge State Distribution	Mass Std Dev	PPM Std Dev
TgGat1	TgGat1	39051.9688	39051.9161	1.3	1.51E+08	5.44	136.39	34	21 - 54	1.68	42.93
TgGat1+Glc	TgGat1	39051.3359	39051.9161	14.9	4.73E+07	2.25	103.70	28	26 - 53	1.92	49.12
TgGat1+Gal	TgGat1	39052.7539	39051.9161	21.5	7.38E+07	4.73	83.04	31	22 - 52	2.39	61.19
PuGat1	PuGat1	30269.1738	30269.2901	3.8	1.86E+09	20.86	92.16	26	17 - 42	0.98	32.34

Fig. S11. Chemical shifts of Gat1 substrate and reaction product (supports Fig. 6)

Chemical shifts* of starting tetra-saccharide:
 $\alpha\text{D-Glc}(1\text{-}3)\alpha\text{L-Fuc}(1\text{-}2)\beta\text{D-Gal}(1\text{-}3)\alpha\text{D-GlcNAc-O-pNP}$

Proton	H1	H2	H3	H4	H5	H6,6'	other
αGlc	5.23	3.56	3.81	3.44	3.89	3.82,3.96	
-3) αFuc	5.23	3.97	3.79	4.00	4.36	1.26	
-2) βGal	4.70	3.63	3.85	3.89	3.56	3.75,3.82	
-3) αGlcNAc	5.24	4.13	4.12	3.63	3.75	3.82,3.95	Ac 2.05
Carbon	C1	C2	C3	C4	C5	C6	other
αGlc	103.4	71.6	75.6	72.3	75.1	63.2	
-3) αFuc	102.1	69.9	81.2	74.4	69.2	18.1	
-2) βGal	102.8	79.4	76.1	71.8	74.6	63.9	
-3) αGlcNAc	101.8	57.2	79.4	71.0	78.6	63.1	Ac 25.1

Chemical shifts** of penta-saccharide:
 $\alpha\text{D-Gal}(1\text{-}3)\alpha\text{D-Glc}(1\text{-}3)\alpha\text{L-Fuc}(1\text{-}2)\beta\text{D-Gal}(1\text{-}3)\alpha\text{D-GlcNAc-O-pNP}$

Proton	H1	H2	H3	H4	H5	H6,6'	other
αGal	5.41	3.84	3.91	4.01	4.27	3.74	
-3) αGlc	5.25	3.67	3.97	nd/s	nd/s	nd/s	
-3) αFuc	s	s	s	s	s	s	
-2) βGal	s	s	s	s	s	s	
-3) αGlcNAc	s	s	s	s	s	s	s

Carbon	C1	C2	C3	C4	C5	C6	other
αGal	101.9	71.4	71.7	71.8	73.3	63.5	
-3) αGlc	103.5	72.8	82.4	nd	nd	nd	
-3) αFuc	s	s	s	s	s	s	
-2) βGal	s	s	s	s	s	s	
-3) αGlcNAc	s	s	s	s	s	s	s

*Proton shifts in ppm referenced to DSS in the pentasaccharide sample. Carbon shifts in ppm derived from indirect referencing. Values for the tetrasaccharide were based on identical peaks in both samples.

** The additional chemical shifts for the terminal α -galactosyl residue and some of the penultimate 3-linked α -glucosyl are entered. Other peaks could either not be clearly distinguished (nd) or were identical (s) to the tetrasaccharide in the mixture.

Fig. S12. Gat1 is a dimer at all concentrations tested (related to Fig. 7C). A, Sedimentation velocity profiles of different concentrations of PuGat1 are displayed with fit data and residuals. 11 μM , 6.5 μM , and 3.5 μM concentrations were detected at 280 nm, 1.3 μM data were collected at 230 nm, and 0.65 μM and 0.3 μM data were collected at 220 nm. B, Data modeled as continuous $c(s)$ distributions are shown (normalized to a value of 1 for the tallest peak). Black dashed line represents the determined S-value, and the red and green dashed lines respectively represent the predicted monomer and dimer S-values. The peak appearing at a near-zero S-value at the lower concentrations may be due to a buffer mismatch that became apparent at lower wavelengths.

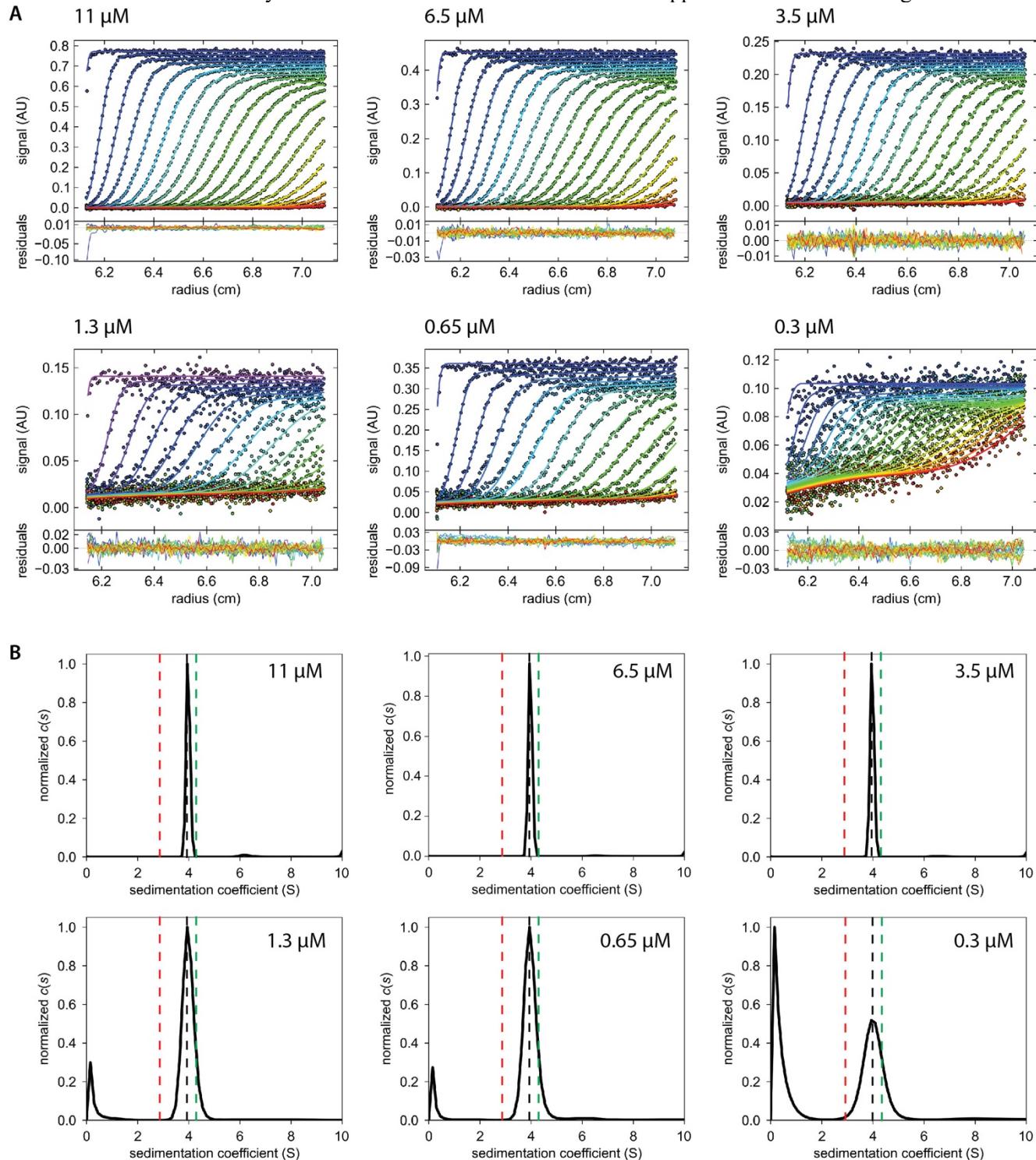


Fig. S13. Gat1 and glycogenin coordinate UDP and Mn²⁺ in similar fashion (related to Fig. 8).

PuGat1:UDP:Mn²⁺ (A) and Oc-glycogenin-1:UDP (PDB 1LL2) (B) are displayed as Ligplots (87). Green dotted lines represent the interactions between the protein and the ligand, and red arcs represent packing interactions.

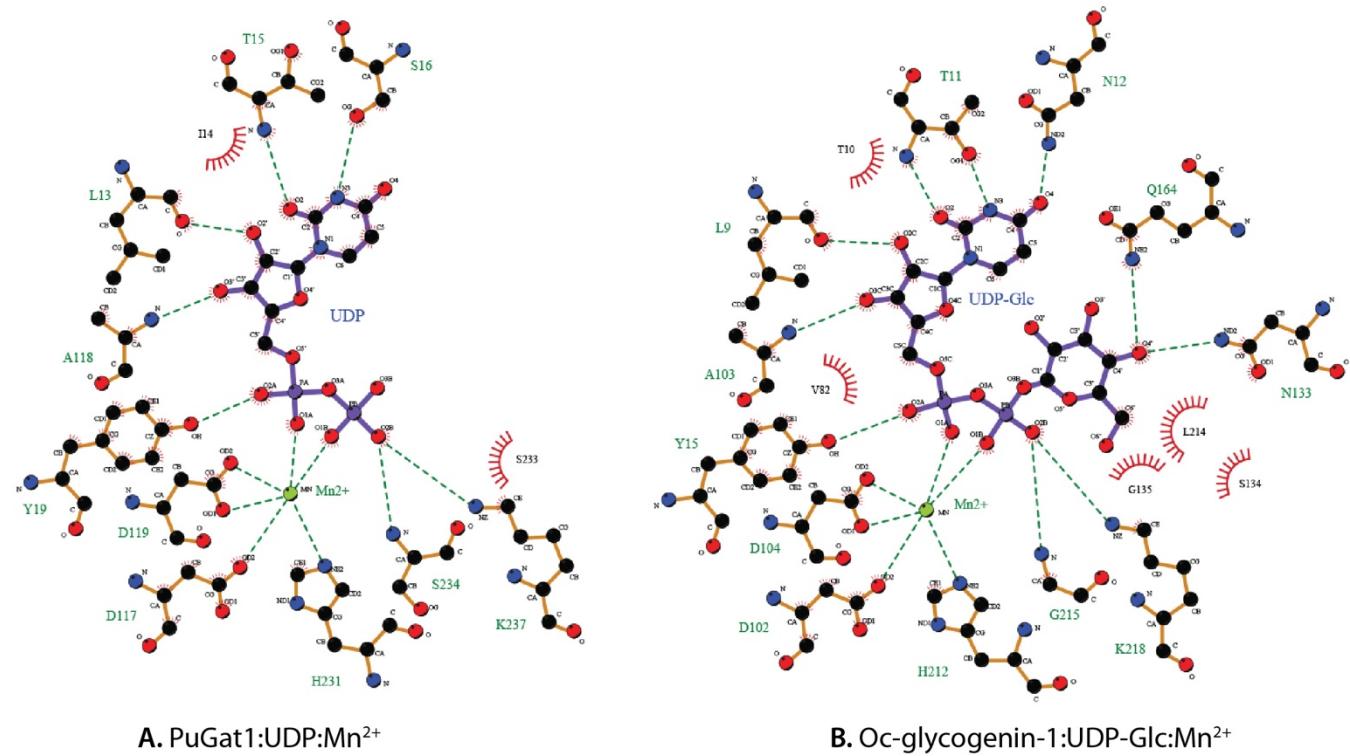


Fig. S14. The *T. gondii* glycan/Skp1 relationship is reminiscent of *D. discoideum* (related to Fig. 10). *A*, Superimposition of the two energy-minimized glycan structures produced by the Glycam webserver (63). Residues are colored according to the SNFG system. The differing Glc (blue) and Gal (yellow) residues (arrowhead) mark the difference between *Toxoplasma* and *Dictyostelium* glycans, respectively. *B*, Superposition of the glycans in the context of Skp1 (orange ribbon); note that the linkage to Hyp is not shown. *C*, Illustration of hydrogen bonds present at >25% occupancy over all simulations (1.5 μ s) in *Toxoplasma* Skp1. *D*, Comparison of amino acid sequences of TgSkp1 and DdSkp1 over the region depicted. Red asterisks indicate residues involved in hydrogen bonds that correlate best with extension of helix-8 (see panel E), green asterisks indicated residues that contribute most to non-polar packing interactions (see Table 2), and the black asterisk indicates the attachment site after hydroxylation. Residues are labeled from below according to the hydrogen bond with which they are associated. *E*, Summary of the six 250-ns trajectories (3 pre-equilibrated; 3 were not). Left bars of each pair summarize the average distances for each trajectory between C156, near the C-terminus, to the center of mass of residues 1-136 (dashed green line in Fig. 10A), scaled to the highest average distance (Equil-3, in which the average value was >50 Å in the observed range of 18-61 Å for at least 85% of the time sampled at 0.1 ns increments). Right bars summarize the 5 most frequent hydrogen bonds between the glycan and Skp1, normalized to the highest level of hydrogen bonds observed in a single trajectory (Equil-1, in which at least one of the hydrogen bonds was occupied >99.7% of the time sampled in 0.1 ns increments). At the right is shown a time-resolved analysis of the correlation of helix-8 extension with the occupancy of each hydrogen bond over the entire 1.5 μ s of simulation time, based on the Pearson's correlation coefficient (linear regression R^2).

