

# Characterization of a cytoplasmic glucosyltransferase that extends the core trisaccharide of the *Toxoplasma* Skp1 E3 ubiquitin ligase subunit

Received for publication, July 28, 2017, and in revised form, September 12, 2017. Published, Papers in Press, September 19, 2017, DOI 10.1074/jbc.M117.809301

Kazi Rahman<sup>‡§1</sup>, Msano Mandalasi<sup>‡¶</sup>, Peng Zhao<sup>||</sup>, M. Osman Sheikh<sup>||</sup>, Rahil Taujale<sup>||\*\*2</sup>, Hyun W. Kim<sup>‡</sup>, Hanke van der Wel<sup>‡</sup>, Khushi Matta<sup>‡‡</sup>, Natarajan Kannan<sup>‡\*\*</sup>, John N. Glushka<sup>||</sup>, Lance Wells<sup>‡||</sup>, and Christopher M. West<sup>‡¶13</sup>

From the <sup>‡</sup>Department of Biochemistry and Molecular Biology, the <sup>||</sup>Complex Carbohydrate Research Center, and the <sup>\*\*</sup>Institute of Bioinformatics, University of Georgia, Athens, Georgia 30602, the Departments of <sup>§</sup>Microbiology and Immunology and <sup>¶</sup>Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104, and the <sup>‡‡</sup>Department of Chemical and Biological Engineering, State University of New York, Buffalo, New York 14260

Edited by Gerald Hart

Skp1 is a subunit of the SCF (Skp1/Cullin 1/F-box protein) class of E3 ubiquitin ligases that are important for eukaryotic protein degradation. Unlike its animal counterparts, Skp1 from *Toxoplasma gondii* is hydroxylated by an O<sub>2</sub>-dependent prolyl-4-hydroxylase (PhyA), and the resulting hydroxyproline can subsequently be modified by a five-sugar chain. A similar modification is found in the social amoeba *Dictyostelium*, where it regulates SCF assembly and O<sub>2</sub>-dependent development. Homologous glucosyltransferases assemble a similar core trisaccharide in both organisms, and a bifunctional  $\alpha$ -galactosyltransferase from CAZy family GT77 mediates the addition of the final two sugars in *Dictyostelium*, generating Gal $\alpha$ 1,3Gal $\alpha$ 1,3Fuc $\alpha$ 1,2Gal $\beta$ 1,3GlcNAc $\alpha$ 1-. Here, we found that *Toxoplasma* utilizes a cytoplasmic glucosyltransferase from an ancient clade of CAZy family GT32 to catalyze transfer of the fourth sugar. Catalytically active Glt1 was required for the addition of the terminal disaccharide in cells, and cytosolic extracts catalyzed transfer of [<sup>3</sup>H]glucose from UDP-[<sup>3</sup>H]glucose to the trisaccharide form of Skp1 in a *glt1*-dependent fashion. Recombinant Glt1 catalyzed the same reaction, confirming that it directly mediates Skp1 glucosylation, and NMR demonstrated formation of a Glc $\alpha$ 1,3Fuc linkage. Recombinant Glt1 strongly preferred the full core trisaccharide attached to Skp1 and labeled only Skp1 in *glt1* $\Delta$  extracts, suggesting specificity for Skp1. *glt1*-knock-out parasites exhibited a growth defect not rescued by catalytically inactive Glt1, indicating that the glycan acts in concert with the first enzyme in the pathway, PhyA, in

cells. A genomic bioinformatics survey suggested that Glt1 belongs to the ancestral Skp1 glycosylation pathway in protists and evolved separately from related Golgi-resident GT32 glucosyltransferases.

Skp1 is an adaptor subunit of the Skp1/Cullin-1/F-box protein (SCF)<sup>4</sup> class of E3 ubiquitin ligases that target proteins for polyubiquitination and degradation via the 26S proteasome (1). In the agent for toxoplasmosis, *Toxoplasma gondii*, Skp1 is hydroxylated by the cytoplasmic prolyl 4-hydroxylase PhyA at Pro-154 (2) and subsequently modified by a linear pentasaccharide (3). In the social amoeba *Dictyostelium*, Skp1 is also modified by hydroxylation and a pentasaccharide, which represents a novel form of SCF regulation (4). Biochemical and interaction studies indicate that full glycosylation of Skp1 promotes association with three different F-box proteins (FBPs) (5, 6) which, in a developmentally regulated manner, is associated with their reduced steady-state levels in cells.<sup>5</sup> Many FBPs are substrate receptors for ligands whose polyubiquitination controls their abundance, whereas others are considered to have enzymatic or other functions (2). FBPs possess a 40-amino acid F-box domain that binds to the C-terminal region of Skp1 (7, 8). Recent studies show that glycosylation influences the organization and range of motions of this region of Skp1, in part by hydrogen bonding along the polypeptide in *cis* (9). Glycosylation is regulated by PhyA, whose action on Skp1 is rate-limited by O<sub>2</sub> availability in the cell (6, 10). This biochemical mechanism underlies cellular O<sub>2</sub> sensing, which controls the slug-to-fruit switch and sporulation during starvation-induced development (4). O<sub>2</sub> signifies positional information in the native soil environment of *Dictyostelium*, and sensing O<sub>2</sub> is key for the

This work was supported in part by Grant 14-140 from the Mizutani Foundation for Glycoscience (to C. M. W.) and National Institutes of Health Grants R01-GM084383 (to C. M. W.), P41-GM103490 (to L. W., senior investigator), and 8P41-GM103390 (Resource for Integrated Glycotechnology). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This article contains supplemental Table S1 and Figs. S1 and S2.

<sup>1</sup> Present address: HIV Dynamics and Replication Program, Center for Cancer Research, NCI-Frederick, Frederick, MD 21702.

<sup>2</sup> Supported by National Institutes of Health Grant T32-GM107004.

<sup>3</sup> To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Biology, University of Georgia, 120 E. Green St., Athens, GA 30602. Tel.: 706-542-4259; E-mail: westcm@uga.edu.

<sup>4</sup> The abbreviations used are: SCF, Skp1/Cullin-1/F-box protein; FBP, F-box protein; Fuc, fucose; Hex, hexose; HexNAc, N-acetylhexosamine; GalNAc, N-acetylgalactosamine; GlcNAc, N-acetylglucosamine; Man, mannose; TOCSY, total correlation spectroscopy; HMBC, heteronuclear multiple bond correlation; HSQC, heteronuclear single quantum correlation; HMM, hidden Markov model; FGn-, Fuc $\alpha$ 1,2Gal $\beta$ 1,3GlcNAc $\alpha$ 1-; FG-, Fuc $\alpha$ 1,2Gal $\beta$ 1-; pNP, p-nitrophenol; CID, collision-induced dissociation.

<sup>5</sup> M. O. Sheikh, B. Nottingham, A. Boland, and C. M. West, unpublished data.

ability of developing cells to navigate to the soil surface for fruiting body formation (11).

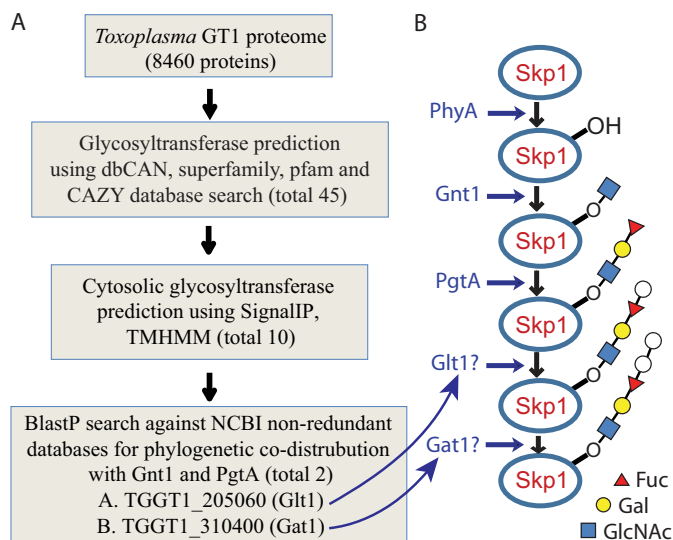
PhyA is important for *Toxoplasma* tachyzoite proliferation on cultured human fibroblasts (12), which involves successive cycles of invasion, intracellular replication, egress, and reinvasion to form cell-free plaques. Glycosylation of *Toxoplasma* Skp1 is important too, because disruption of genes that mediate the addition of the first three monosaccharides also results in reduced parasite growth (3). However, the significance of the full pentasaccharide has not been examined, because the gene that mediates the addition of the final two sugars on the *Dictyostelium* pentasaccharide, *agtA* (13), is evidently absent from the *Toxoplasma* genome (14).

To investigate the function of the terminal disaccharide of the *Toxoplasma* Skp1 glycan, we searched for glycosyltransferase-like genes in the parasite's genome whose protein products are predicted to reside in the cytoplasm or nucleus and whose phylogenetic distribution correlates with the presence of the *pgtA* gene that mediates the addition of the second and third monosaccharides (15). This search netted two genes, not found in the amoebozoia, including *Dictyostelium*. Characterization of one of these (TGGT1\_205060), which we named *glt1*, revealed a previously non-annotated gene that encodes a cytoplasmic glucosyltransferase from CAZy glycosyltransferase family 32 that modifies the Fuc terminus of the Skp1 trisaccharide. Gene disruption and complementation studies show that the fourth sugar contributes to Skp1 functionality in parasite growth. Thus, these unrelated protists employ different mechanisms to assemble a related but distinct pentasaccharide to regulate, presumably, the critical process of protein turnover in cells. A deeper understanding of the enzymes involved may offer unique strategies to control toxoplasmosis, which affects a large fraction of the world's human population and for which control of acute and chronic phases is lacking (16).

## Results

### Prediction of candidate TgSkp1-modifying glycosyltransferases

To identify candidate proteins for catalyzing the addition of the fourth and fifth Skp1 monosaccharides, genome sequence databases were searched for predicted glycosyltransferase domains that reside in the cytoplasm and exist only in protists whose genomes harbor other predicted Skp1 modification pathway enzymes, *phyA*, *gnt1*, and *pgtA*, but not *agtA*, as outlined in Fig. 1. Briefly, the predicted proteome (8460 proteins) of the *T. gondii* Type I GT1 strain was searched using (i) the SUPERFAMILY server, which assigns protein domains at the SCOP "superfamily" level using hidden Markov models; (ii) dbCAN, an automated carbohydrate-active enzyme annotation database, which utilizes a CAZyme signature domain-based annotation based on a CDD (conserved domain database) search, literature curation, and a hidden Markov model; and (iii) the Pfam database. The resulting 45 sequences were scanned by SignalIP version 4.1 and TMHMM servers for signal sequences or transmembrane domains, which yielded 10 candidate cytoplasmic glycosyltransferases. Among those, TgGnt1 and TgPgtA were already identified as TgSkp1-modifying gly-

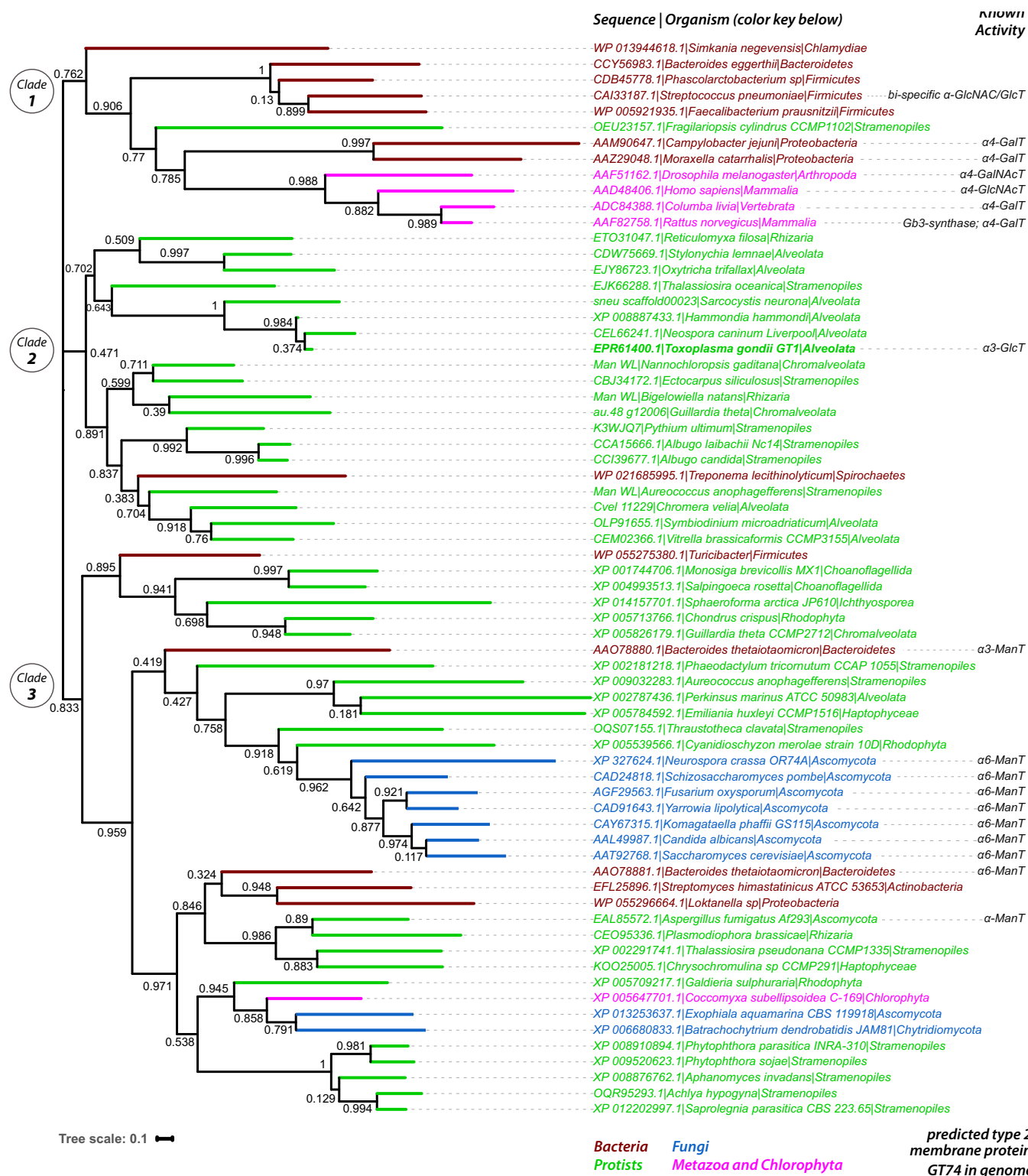


**Figure 1. Bioinformatics strategy used to identify potential TgSkp1-modifying glycosyltransferases.** A, steps utilized in bioinformatic analysis to identify potential terminal glycosyltransferase candidates. B, the Skp1 glycosylation pathway in *Toxoplasma*. Arrows from left to right show at which step of the pathway putative glycosyltransferases are predicted to function.

cosyltransferases (3). The remaining eight sequences were subjected to BlastP analysis against the NCBI non-redundant database to search for their phylogenetic co-distribution with *gnt1* and *pgtA*. This yielded two candidates, one from CAZy family GT32, TGGT1\_205060, and another from CAZy family GT8. The GT32 sequence, here referred to as Glt1, is the subject of this study.

### Glt1 sequence characteristics and phylogeny

In the *Toxoplasma* database ToxoDB V29, TGGT1\_205060 (type 1 GT1 strain) is annotated as a 605-amino acid protein encoded by five exons that is conserved in ME49 (type II) and VEG (type III) strains with 99% nucleotide sequence identity. The protein contains a DXD motif typical of and essential for CAZy GT32 family and superfamily A glycosyltransferases (17). Proteomics analyses of *Toxoplasma*, as reported on ToxoDB, identified five Glt1 peptides, one of which was phosphorylated at Ser-601 (supplemental Fig. S1), confirming that the gene is expressed in tachyzoites grown in fibroblasts. To investigate the evolutionary origin of Glt1, 68 related and representative sequences from CAZy family GT32 were aligned according to their catalytic domains and subjected to a phylogenetic analysis using a maximum likelihood algorithm (Fig. 2). Glt1 was found in clade 2, which was almost uniquely populated by sequences from other protists that possess a CAZy family GT74-like sequence related to PgtA, which mediates the addition of the third Skp1 monosaccharide, upon which Glt1 is postulated to act. (Instances of GT74 association with sequences in other clades merely represent a second GT32-like sequence in a clade 2 organism.) The sole exception was a bacterial sequence that lacks the Skp1 glycosylation pathway. All sequences in this clade are predicted to encode cytoplasmic proteins based on the absence of signal peptide or anchor sequences (Fig. 2). Thus, they potentially modify Skp1, which also resides in this compartment and possesses the equivalent of Pro-154 in at least one

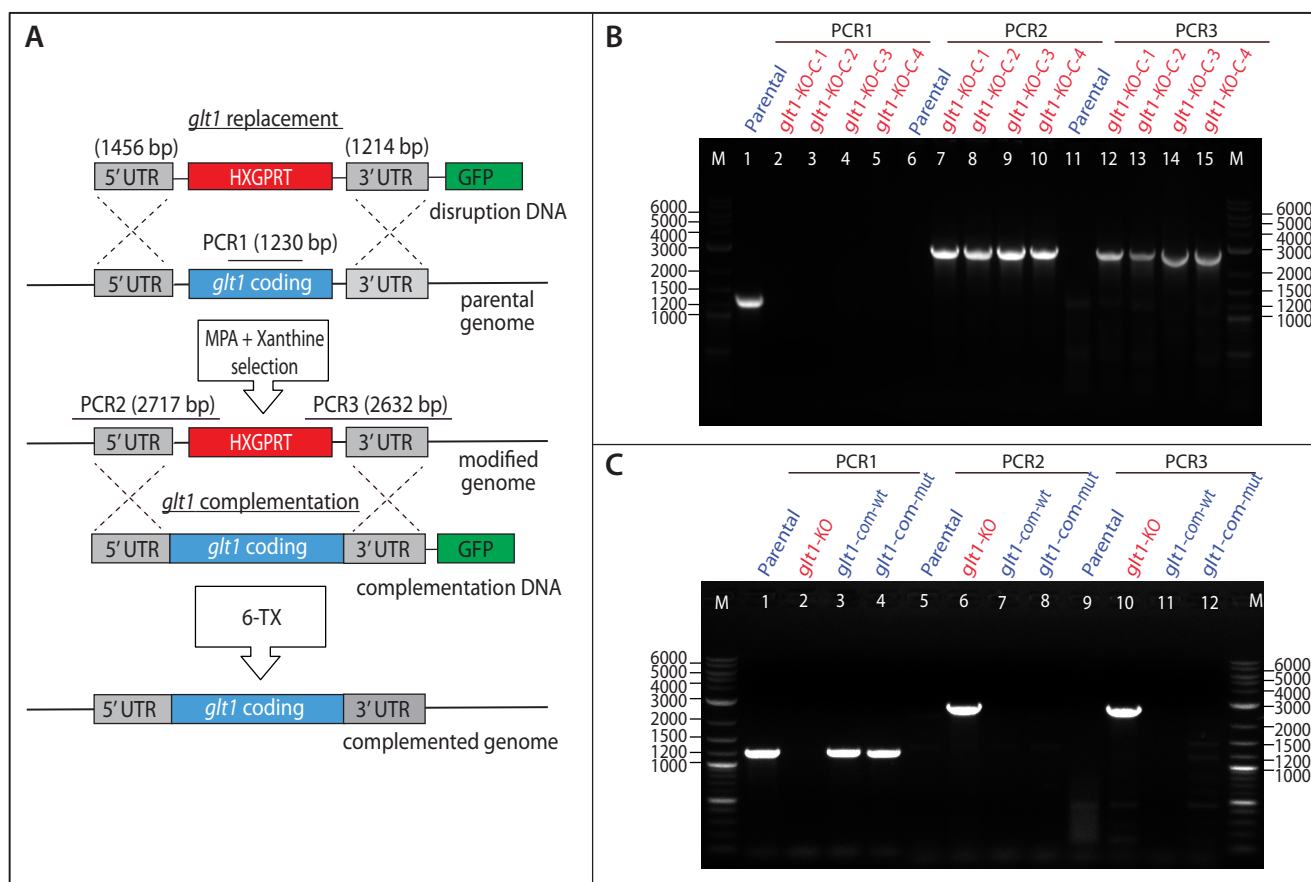


**Figure 2. Phylogenetic distribution of Glt1-like sequences.** Phylogenetic tree showing three distinct clades for the CAZY GT32 family sequences. Different colored branches and leaves indicate different taxonomic groups as labeled. Local support values for the branch splits are shown. Known enzymatic activity for experimentally characterized sequences is provided in black after the leaf name as per the CAZY database. Teal-colored circles indicate the presence of a GT74 fucosyltransferase-related protein in the species; empty circles indicate its absence. Orange coloration indicates evidence for a signal peptide or signal anchor sequence expected to direct the protein to the secretory pathway.

of its Skp1 genes (in cases where there are more than one). An alignment of Glt1-like amino acid sequences from diverse organisms identified Pro-234, Asp-348, Arg-351, Asp-363, Asp-365, Gly-470, Pro-471, and His-586 of Glt1 as generally

characteristic of all GT32 family members (supplemental Fig. S2). In contrast, Gly-243, Glu-393, and Asn-561 are unique for the Glt1-like clade 2, and Trp-241, Trp-318, Gly-385, Asn-396, His-405, Pro-552, and Trp-592 are characteristic of but not





**Figure 3. Disruption and complementation of *glt1*.** A, outline of strategy for double crossover gene replacements and PCRs to monitor outcomes. B, PCR analysis of *hxgprt*-positive clones that were resistant to mycophenolic acid and xanthine and GFP-negative. Desired replacement clones were characterized by the absence of products of PCR1 and the presence of PCR2 and PCR3. C, PCR analysis of *glt1* replacement clone-1 with gDNA encoding wild-type Glt1 or a triple mutant form encoding Glt1(D363A/D365A/D348N).

exclusive to this clade, which support the uniqueness of the Glt1 clade. The other two major clades contain sequences from all kingdoms, including bacteria, whose members are usually cytoplasmic as is common for bacterial glycan biogenesis, and protists/fungi/metazoan/chlorophytes, whose members are usually localized to the secretory pathway. Known functions of clade 1 of Fig. 2 include  $\alpha$ 4Gal- and  $\alpha$ 4GalNAc-transferases, whereas known functions in clade 3 include  $\alpha$ 6-Man-transferases. This analysis suggests that clade 2 with predominantly Glt1-like sequences diverged from the other two clades during the earliest radiation of the GT32 family and might have evolved to uniquely modify Skp1. Glt1 is likely to be a retaining enzyme that attaches a D-sugar in  $\alpha$ -linkage, but the identity of the sugar cannot be predicted.

### Glt1 is required for *Tg*Skp1 glycosylation

To determine whether *glt1* is involved in Skp1 glycosylation, the gene was replaced by double-crossover homologous recombination in the type 1 RH $\Delta\Delta$  strain, as illustrated in Fig. 3A. *glt1* deletion mutants were confirmed by negative PCRs for *glt1*-coding DNA and positive PCR products for the insertion of the selection marker *hxgprt* between *glt1* flanking sequences (Fig. 3B). To control for off-target genetic modifications, a complementation construct containing a version of the original genomic DNA was used to replace the *hxgprt* locus in *glt1* dis-

ruption clone-1 using counterselection for loss of *hxgprt* (Fig. 3A). The same set of PCRs were used to confirm the desired gene restoration in clonal isolates (Fig. 3C). The same clone was also transformed with a mutant version in which three conserved Asp residues, including the DXD motif at positions 363 and 365 and another at codon 348, were changed to generate a potentially inactive mutant Glt1(D363A/D365A/D348N). Strains are listed in Table 1.

Western blot analysis of whole cells suggested that Skp1 from parasites lacking *glt1* (Fig. 4A, lane 3) migrates slightly more rapidly than wild-type Skp1 (lanes 1 and 6) but slower than Skp1 from a *phyA* $\Delta$  strain (lane 2). Complementation with wild-type *glt1* (lane 4), but not the triple mutant (lane 5), restored normal mobility. This confirmed the specificity of the genetic disruption to *glt1* and further implicated Glt1 as a Skp1 glucosyltransferase. To confirm an effect on Skp1 glycosylation and to pinpoint its location, Skp1 was immunoprecipitated, converted to peptides using trypsin, and analyzed by nanoLC/MS-MS in an Orbitrap Fusion mass spectrometer. As shown in the extracted ion chromatograms from a sample from the *glt1*-complemented strain (Fig. 4B, left, dashed box 1), doubly and triply charged ions with exact mass matches to the peptide ( $^{145}$ IFNIVNDFTPEEEAQVR $^{161}$ ) that bears the known modification site, Pro-154, were readily detected. As expected, the

**Table 1**
*Toxoplasma* strains used in this study

Strain	Parental strain	Genotype	Gene targeted	Selection marker	Selection drug	Source/Reference
RHΔΔ	RH(1)	Δhxgprt;Δku80				Ref. 24
<i>phyA</i> Δ	RHΔΔ	<i>phyA</i> Δ;Δku80	<i>phyA</i> -exon 1	Hxgprt	MPA, <sup>a</sup> xanthine	Ref. 12
<i>glt1</i> Δ	RHΔΔ	<i>glt1</i> Δ;Δku80	<i>glt1</i> -exons 1–5 (all)	Hxgprt	MPA, xanthine	This work
<i>glt1</i> Δ; <i>glt1</i> <sup>+</sup>	<i>glt1</i> Δ-clone 1	Δhxgprt;Δku80		ΔHxgprt	6-Thioxanthine	This work
<i>glt1</i> Δ; <i>glt1</i> (m) <sup>+</sup> <sup>b</sup>	<i>glt1</i> Δ-clone 1	Δhxgprt;Δku80; <i>glt1</i> (D363A/D365A/D348A)		ΔHxgprt	6-Thioxanthine	This work

<sup>a</sup> MPA, mycophenolic acid.

<sup>b</sup> (m), the triple mutant form encoding *Glt1*(D363A/D365A/D348N).

pentasaccharide form was also readily detected at an earlier elution time (Fig. 4E, left, dashed box 2), as reported previously for the normal parental strain (3) and shown in Table 2. In addition, a very low level of the disaccharide form was apparent in Fig. 4C, but no trisaccharide form was detected (Fig. 4D). In contrast, Skp1 from a *glt1*Δ clone yielded no detectable pentasaccharide but abundant trisaccharide (box 3), as well as the unmodified peptide and a minor level of the disaccharide form. The doubly charged version of this putative dHex-Hex-HexNAc-peptide (Fig. 5A) was subjected to further CID analysis, yielding a series of cleavage products confirming the order of sugar residues (Fig. 5B and Table 2) and the sequence of the peptide and its attachment, as expected, at a hydroxylated form of Pro-154 (Fig. 5C). In a clone in which *glt1*Δ was complemented with mutated (D363A/D365A/D348N) *glt1* genomic DNA, the profiles resembled those of the *glt1*Δ strain (Table 2), demonstrating the importance of its predicted glucosyltransferase activity. These findings implicate *Glt1* as the glucosyltransferase that catalyzes the addition of the fourth sugar and indicate that the addition of the fifth sugar depends on the fourth.

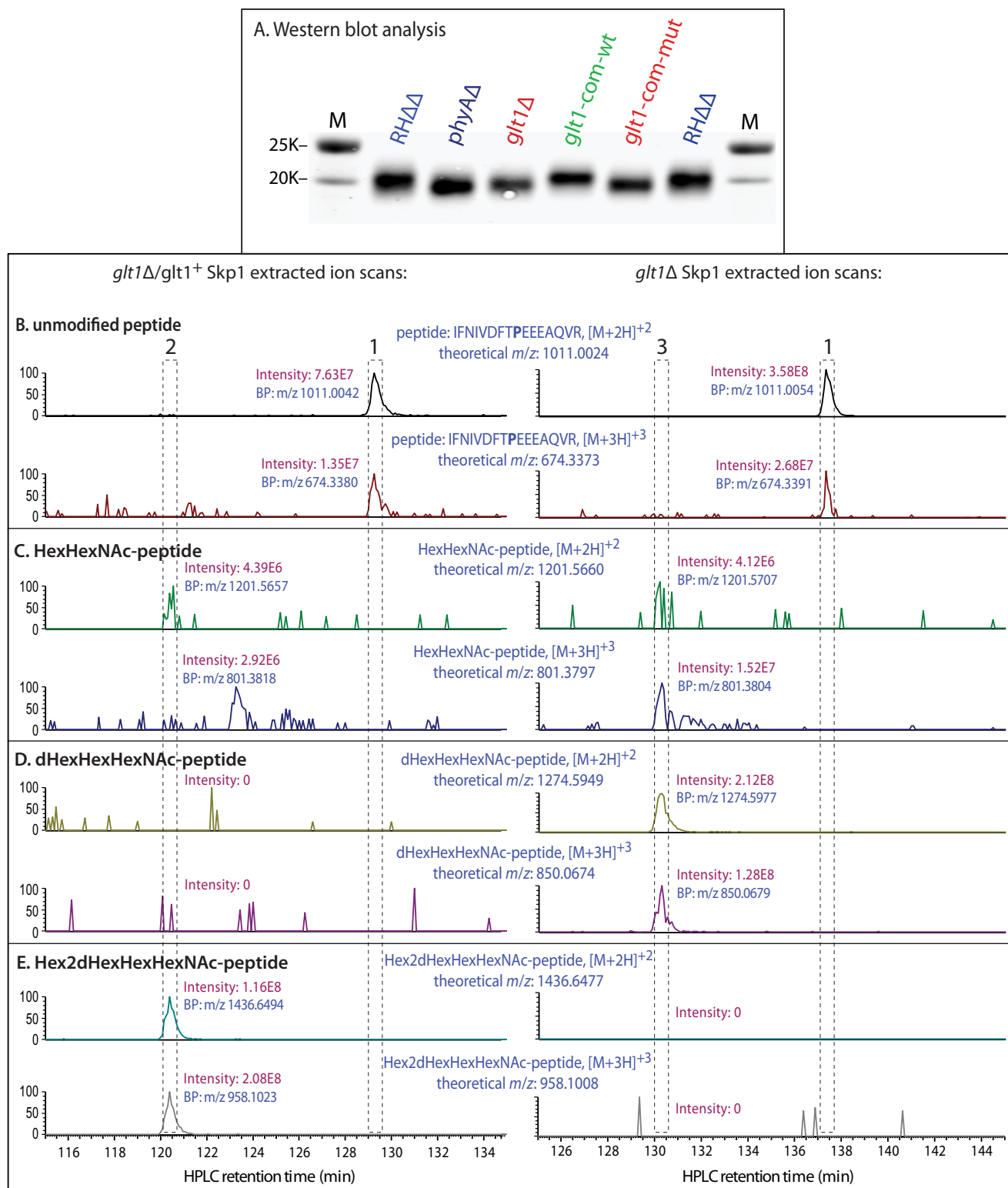
### *Toxoplasma* expresses a *glt1*-dependent Skp1 glucosyltransferase activity

To characterize the implied activity of *Glt1*, exogenous radioactive sugar nucleotide donors and Skp1 glycoforms were introduced into soluble parasite extracts in an attempt to recapitulate reactions in the cell. Because *Toxoplasma* and *Dictyostelium* apparently share the same core trisaccharide on their Skp1s, and their Skp1 sequences are highly conserved and serve as substrates for each other's *PhyA*, *Gnt1*, and *PgtA* enzymes (3, 12), we used recombinantly expressed *Dictyostelium* FGn-Skp1 (13) as a surrogate acceptor substrate. *In vitro* reactions were performed with cytosolic extracts in the presence of UDP-[<sup>3</sup>H]Gal, based on the occurrence of Gal in *Dictyostelium* Skp1 (Fig. 6A). Substantial activity was detected, which was time- and Skp1-dependent and absent from *glt1*Δ extracts. To confirm the chemical identity of the transferred <sup>3</sup>H, the radioactive Skp1 band was excised after electroblot transfer to a PVDF membrane and subjected to acid hydrolysis and high-pH anion exchange chromatography. Surprisingly, radioactivity co-chromatographed with D-Glc (Fig. 6B). New reactions using UDP-[<sup>3</sup>H]Glc revealed more efficient incorporation (Fig. 6A), whose product also co-eluted with the Glc standard (data not shown). The incorporation of <sup>3</sup>H from UDP-[<sup>3</sup>H]Gal was probably the result of the action of TGGT1\_225880, a putative UDP-Glc 4-epimerase in the extract, which would explain the lower incorporation of <sup>3</sup>H from this substrate. Therefore, *Glt1* is inferred to utilize UDP-Glc as its substrate.

### Activity of recombinant *Glt1*

To test whether *Glt1* is capable of directly glycosylating FGn-Skp1, a codon-optimized *Toxoplasma glt1* cDNA (supplemental Fig. S1) was expressed in and purified from *Escherichia coli* as an N-terminally His<sub>6</sub>-tagged protein. After enrichment over a Co<sup>2+</sup>-Talon column (Fig. 7A), Western blot analysis of fractions from a Q-anion exchange column revealed an anti-His<sub>6</sub>-reactive band at the expected *M<sub>r</sub>* value of 69,500 (Fig. 7C). To simplify detection of enzymatic activity, a synthetic small molecule, FGn-pNP, was used as a surrogate for FGn-Skp1, and its validity is described below. A glucosyltransferase activity catalyzing the transfer of <sup>3</sup>H from UDP-[<sup>3</sup>H]Glc to FGn-pNP coeluted with the 69,500 band (Fig. 7B). Inspection of a parallel gel stained for total protein with Coomassie Blue indicated a prominent band that co-migrated with the His<sub>6</sub> band (Fig. 7D), suggesting that the *Glt1* protein was major component of the preparation. The great majority of *Glt1* was full-length (not shown), and densitometry indicated that *Glt1* represented 14% of the protein in fraction 21 (not shown). Pilot studies showed that the transferase activity was stable on ice for days and freeze-thawing, most active in the presence of bovine serum albumin, unaffected by NaCl concentration over the range of 10–400 mM, and more active at increasing pH values up to the highest value tested (pH 8.5). Activity was blocked by the addition of EDTA, consistent with the importance of the DXD sequence for divalent cation coordination. A recombinant version of the triple (D348N/D363A/D365A) mutant described above, as well as each of the individual point mutants, expressed well but were inactive (data not shown). MnCl<sub>2</sub>, but not MgCl<sub>2</sub>, supported activity, and 2 mM was sufficient for maximal activation. These data guided the design of the standard reaction to examine the substrate specificities of the enzyme.

Most glucosyltransferases are capable of specifically hydrolyzing their donor substrates by transferring the sugar to water in the absence of an appropriate acceptor (18). A screen for the ability of purified His<sub>6</sub>-*Glt1* to hydrolyze six different UDP-sugar donors, based on generation of UDP, revealed strong selectivity for UDP-Glc (Fig. 8A), which was consistent with the transferase activity in cytosolic extracts (Fig. 6). GDP-Man was not a substrate in a glucosyltransferase assay utilizing FG-Bn, which mimics the non-reducing terminal disaccharide of the Skp1 glycan, using a variation of the UDP-Glc assay that measures the generation of GDP or UDP (Fig. 8B). Kinetic analysis of UDP generation yielded a *K<sub>m</sub>* for UDP-Glc of 7.5 ± 2.3 μM. Transferase activity toward FG-Bn was also monitored as the incorporation of <sup>3</sup>H from UDP-[<sup>3</sup>H]Glc into FG-Bn. These data confirmed the expected inverse hyperbolic dependence on



**Figure 4. *glt1Δ* parasites accumulate the trisaccharide form of Skp1.** A, Western blot analysis of Skp1 from equivalent numbers ( $3 \times 10^6$  cells) of parental RH $\Delta\Delta$ , *phyAΔ*, *glt1Δ*, or complemented cells. Soluble S16 fractions were separated on a 4–12% SDS-polyacrylamide gel, electroblotted, and probed with anti-TgSkp1 (UOK75) antiserum. B–E, analysis of the Skp1 glycosylation by mass spectrometry, from RH $\Delta\Delta$  (left half) and *glt1Δ* (right half) cells. Skp1 was enriched by immunoprecipitation and converted to peptides with trypsin. Total peptides were analyzed by nanoLC/MS-MS (note that retention times varied between the left- and right-hand trials, which were conducted at different times). The vertical dashed boxes show approximate elution intervals for glycoforms of the Pro-154-containing peptides from Skp1. Spectral counts corresponding to exact m/z matches that appear as both doubly and triply charged forms are annotated at the side. B, extracted ion scans from full-ion mode for the unmodified peptide containing Pro-154, showing scans for [M + 2H]<sup>2+</sup> and [M + 3H]<sup>3+</sup> ions. C, extracted ion scans for corresponding to the HexHexNAC- derivative of the Pro-hydroxylated peptide. D, same for the dHexHexHexNAC- derivative. E, same for the fully glycosylated Hex2dHexHexHexNAC- derivative. No other glycoforms were detected.



**Table 2**
**MS detection of Skp1 glycopeptides in strains**

Isoforms of the Skp1 peptide <sup>145</sup>IFNIVNDFTPEEEAQVR were detected as described in Fig. 4. The distribution of raw ion counts among the detected isoforms is shown for the strains analyzed.

Strain	unmodified peptide			Hex-HexNAc-peptide			dHex-Hex-HexNAc-peptide			Hex-Hex-dHex-Hex-HexNAc-peptide		
	fraction <sup>a</sup>	[M+2H] <sup>2+</sup>	ΔM <sup>b</sup>	fraction	[M+2H] <sup>2+</sup>	ΔM	fraction	[M+2H] <sup>2+</sup>	ΔM	fraction	[M+2H] <sup>2+</sup>	ΔM
		[M+3H] <sup>3+</sup>	ΔM		[M+3H] <sup>3+</sup>	ΔM		[M+3H] <sup>3+</sup>	ΔM		[M+3H] <sup>3+</sup>	ΔM
RHΔΔ <sup>c</sup>	0.68	1011.0002 n.r. <sup>d</sup>	-2.18	0	n.d. <sup>e</sup>		0	n.d.		0.31	1436.6464 n.r.	-0.92
<i>glt1</i> Δ	0.45	1011.0042 674.3391	1.78 2.67	0.03	1201.5630 801.3804	-2.47 0.90	0.51	1274.5964 850.0674	1.16 2.06	0	n.d. <sup>e</sup>	
<i>glt1</i> Δ; <i>glt1</i> <sup>+</sup>	0.27	1011.0042 674.3380	1.78 1.04	0.01	1201.5657 801.3818	-0.22 2.65	0	n.d.		0.71	1436.6494 958.1023	1.17 1.52
<i>glt1</i> Δ; <i>glt1</i> (m) <sup>+</sup> <sub>f</sub>	0.36	1011.0020 674.3370	-0.40 -0.44	0.05	1201.5651 801.3817	-0.72 2.52	0.59	1274.5948 850.0659	-0.09 0.30	0	n.d.	

<sup>a</sup> Apparent fractional occupancy, averaged from doubly and triply charged ion raw spectral counts. <sup>b</sup> Δm/z in ppm. <sup>c</sup> From Ref. 3. <sup>d</sup> Not reported. <sup>e</sup> Not detected. <sup>f</sup> (m), the triple mutant form encoding Glt1(D363A/D365A/D348N).

UDP-Glc concentration and yielded an apparent  $K_m$  of  $6.0 \pm 0.77 \mu\text{M}$  (Fig. 8C), consistent with the value from the other assay. UDP was a more potent inhibitor than GDP of this transferase reaction (data not shown), confirming that GDP-Man is not a substrate.

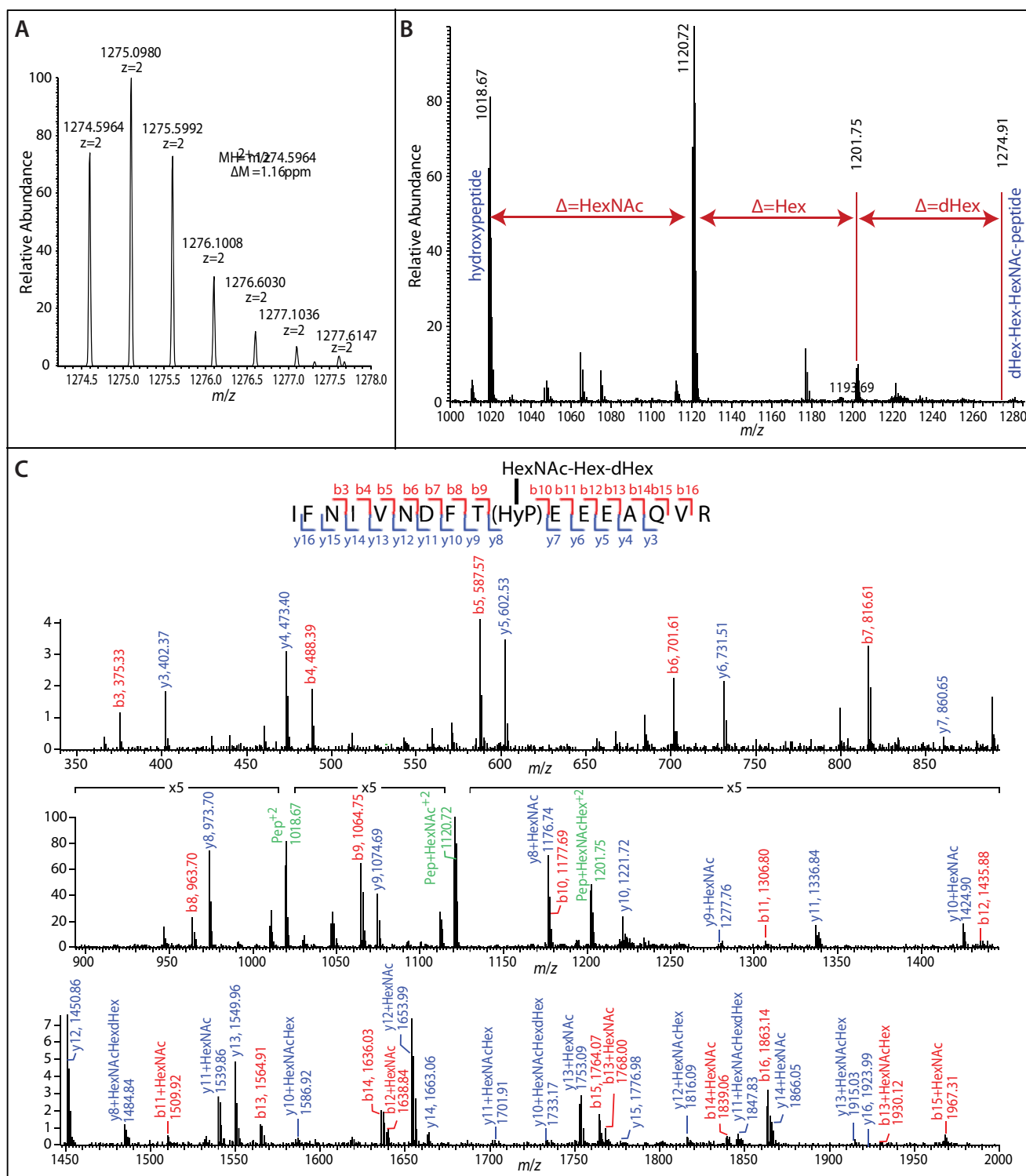
A comparison of FG-Bn and FGGn-pNP acceptors showed markedly improved time-dependent activity toward the full trisaccharide (Fig. 8D), although an effect of the different aglycon moieties cannot be excluded. Fuc-pNP alone was inactive as an acceptor, as were various other mono- and disaccharides representing facets of the Skp1 or related glycans (Fig. 8E). Glt1 transferase activity toward varied concentrations of FGGn-pNP yielded a  $K_m$  of 13 mM and a  $V_{\max}$  of 330 nmol/h/μg, or about 6 nmol of Glc/nmol of Glt1/s (Fig. 8F). To compare with the native substrate, *Toxoplasma* Skp1 was co-expressed in *E. coli* with *Dictyostelium* PhyA and Gnt1, purified to near homogeneity under non-denaturing conditions, and modified to completion *ex vivo* using *Dictyostelium* PgtA. Activity toward *Toxoplasma* FGGn-Skp1 yielded an apparent  $K_m$  over 3 orders of magnitude lower than toward FGGn-pNP, 4.2 μM, and an apparent  $V_{\max}$  that was reduced by about 200-fold (Fig. 8G). Thus, at the low micromolar concentrations expected in the cell, Glt1 is calculated to exhibit a strong preference for the FGGn-trisaccharide when attached to Skp1 compared with unknown potential aglycons.

To address the question of whether Glt1 has cellular targets other than Skp1, we searched for substrates that might accumulate in *glt1*Δ parasites and be susceptible to glucosylation in extracts. This was tested by incubation of desalted wild-type or *glt1*Δ parasite extracts in the presence of His<sub>6</sub>-Glt1 and UDP-[<sup>3</sup>H]Glc, followed by separation of the entire reaction mixture by SDS-PAGE and scintillation counting of gel slices. As shown

in Fig. 8H, a high level of incorporation of <sup>3</sup>H was observed at the position of Skp1 in the *glt1*Δ extract. Incorporation was dependent on the addition of His<sub>6</sub>-Glt1 and was not observed in parental (RHΔΔ) extracts consistent with Skp1 being already modified before cell extraction. Minimal incorporation was detected at higher  $M_r$  positions, but it was also observed in the controls. Thus, within the sensitivity of the method and assuming that potential other Glt1 substrates are accessible and not alternatively processed in its absence, Glt1's main or only substrate is Skp1.

### Linkage position of Glc

To characterize the glycosidic linkage of the Glc residue, FGGn-pNP was used as the acceptor substrate in a scaled up reaction with Glt1, and the glycan product was recovered by solid phase extraction on a C<sub>18</sub> Sep-Pak cartridge and analyzed by NMR. A 1D <sup>1</sup>H NMR spectrum revealed that the sample was of high purity but also contained many overlapping peaks in the proton dimension (Fig. 9A, between ~3.5 and 5.5 ppm). Much of the peak overlap in the proton dimension was resolved in the two-dimensional <sup>1</sup>H-<sup>13</sup>C HSQC spectrum, which was provisionally assigned using the CASPER program (19) (Fig. 9B). Assignments were confirmed by analysis of the 2D COSY, TOCSY, and HMBC spectra (data not shown). The HMBC spectra of the Glt1 reaction product also revealed through-bond connectivities between the anomeric carbons to the ring protons (Fig. 9C) and between the ring carbons and anomeric protons (Fig. 9D), clearly demonstrating the glycosidic linkage between the terminal αGlc and underlying αFuc as 1→3. Taken together, the NMR analyses are most consistent with the Skp1 glycan structure: Glcα1,3Fucα1,2Galβ1,3GlcNAcα1-. Thus, Glt1 is a UDP-Glc: fucoside α1,3-glucosyltransferase.



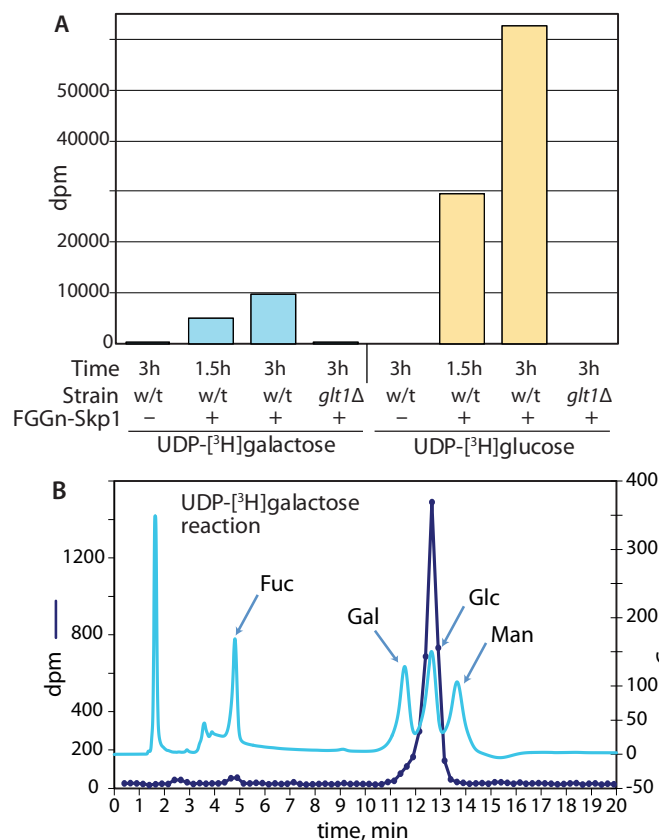
**Figure 5. MS-MS confirmation of the trisaccharide-glycopeptide.** A, the doubly charged ion from Fig. 4, whose  $m/z$  (1274.5954) matched that expected of IFNIVFTP(HexNAc-dHex-Hex)EEEAQVR. B, CID fragmentation of the doubly charged precursor ion yields a sequential loss of monosaccharide residues corresponding to dHex, Hex, and HexNAc, indicating the presence of a linear trisaccharide. C, the full CID fragmentation spectrum showing b (blue annotations) and y (red annotations) ion series that match the predicted peptide sequence, as illustrated in the inset. The glycan is linked via a hydroxylated derivative of Pro-154. Peptides with residual sugars are annotated in green.

### *Glt1* is important for *Toxoplasma* proliferation

The ability of the parasite to infect and proliferate on a monolayer culture of fibroblasts, as measured from the area of clearance (plaques) of cells, is a model for potential virulence in animals. Confluent monolayers were infected with parental,

mutant, and complemented parasites, and plaque areas were analyzed 5.5 days later. As shown in Fig. 10, the average plaque area generated by *glt1*Δ clones is reduced compared with that of parental (RHΔΔ) cells, but larger than that of *phyA*Δ strains that lack the entire modification on Skp1. There was no evi-



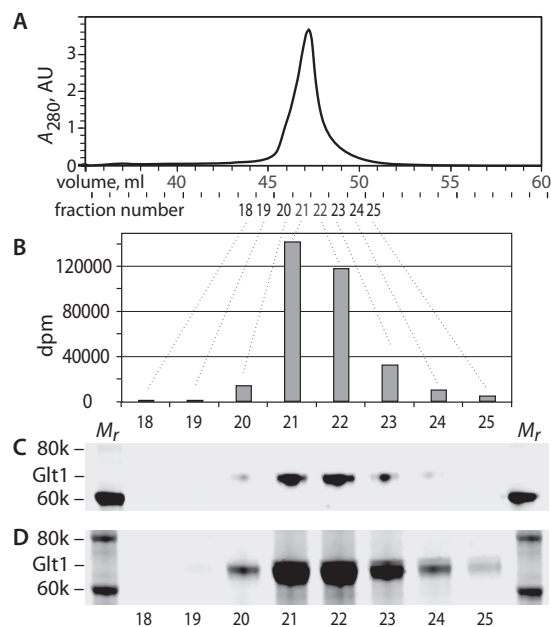


**Figure 6. *Toxoplasma* extracts express *glt1*-dependent Skp1 glucosyltransferase activity.** A, desalted cytosolic extracts were assayed for the ability to transfer <sup>3</sup>H from UDP-[<sup>3</sup>H]Gal or UDP-[<sup>3</sup>H]Glc to FGGn-Skp1, based on incorporation into the Skp1 band after SDS-PAGE. Normal (RHΔΔ) and *glt1*Δ mutant extracts were assayed for 1.5 or 3 h as indicated. Negligible incorporation was detected in the rest of the SDS-PAGE lane (data not shown). These data, which were collected in parallel on the same extracts, are representative of other trials conducted under varied conditions (data not shown). B, after electrophoresis from the SDS-polyacrylamide gel to a polyvinylidene difluoride membrane and staining with Ponceau, the Skp1 band and surrounding regions were separately acid hydrolyzed and subjected to high pH anion exchange chromatography on a Dionex PA-1 column in the presence of internal standards. Monosaccharide elution was monitored by pulsed amperometric detection (nC), and fractions were collected for scintillation counting for <sup>3</sup>H. Similar results were obtained for UDP-[<sup>3</sup>H]Gal (shown) and UDP-[<sup>3</sup>H]Glc reaction products (data not shown).

dence for reduced plating efficiency (data not shown). Genetic complementation with the original *glt1* sequence at the same locus restored normal growth, verifying that the growth defect in the original disruption strain was due to disruption of *glt1*. In contrast, complementation with the enzymatically inactive mutant version of *glt1* failed to rescue the growth defect of *glt1*Δ cells. Therefore, the slow growth of the strain can be attributed to loss of the enzymatic activity itself rather than another potential function of Glt1.

## Discussion

*glt1*Δ parasites exhibit a growth defect in fibroblast cultures. The effect was specific for *glt1*, as complementation with *glt1* restored normal growth. Failure to complement with a catalytically inactive sequence demonstrated the importance of the catalytic activity of Glt1. The deficit was not as strong as that of disrupting *phyA*, an earlier gene in the Skp1 modification pathway. Key to understanding how *glt1* contributes to parasite

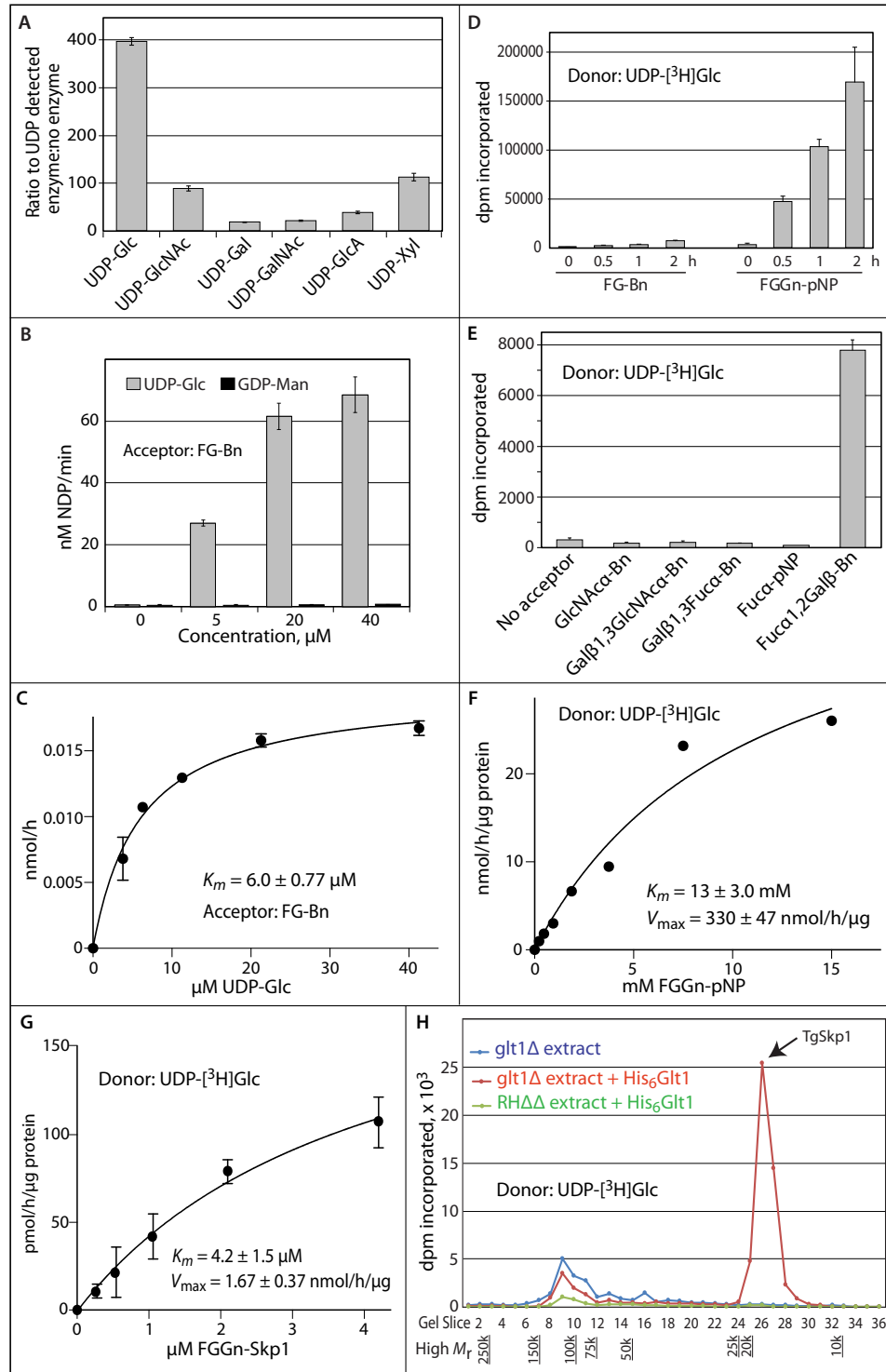


**Figure 7. Purification of recombinant Glt1 and *Toxoplasma* FGGn-Skp1.** A, recombinant His<sub>6</sub>-Glt1 was purified from *E. coli* over Co<sup>2+</sup>-Talon and Q-anion exchange columns. Protein was eluted from the Q-column using a gradient of NaCl and monitored at A<sub>280</sub>. AU, absorbance units. B, fractions were assayed for glucosyltransferase activity using FGGn-pNP as the acceptor substrate. C, fractions were analyzed for Glt1 protein by Western blotting with an anti-His<sub>6</sub> tag antibody. The region of the gel showing reactivity with the antibody is shown. D, fractions were analyzed for total protein using SDS-PAGE and Coomassie Blue staining. Only the region of the gel corresponding to C is shown. Equal volumes of fractions were analyzed in B–D.

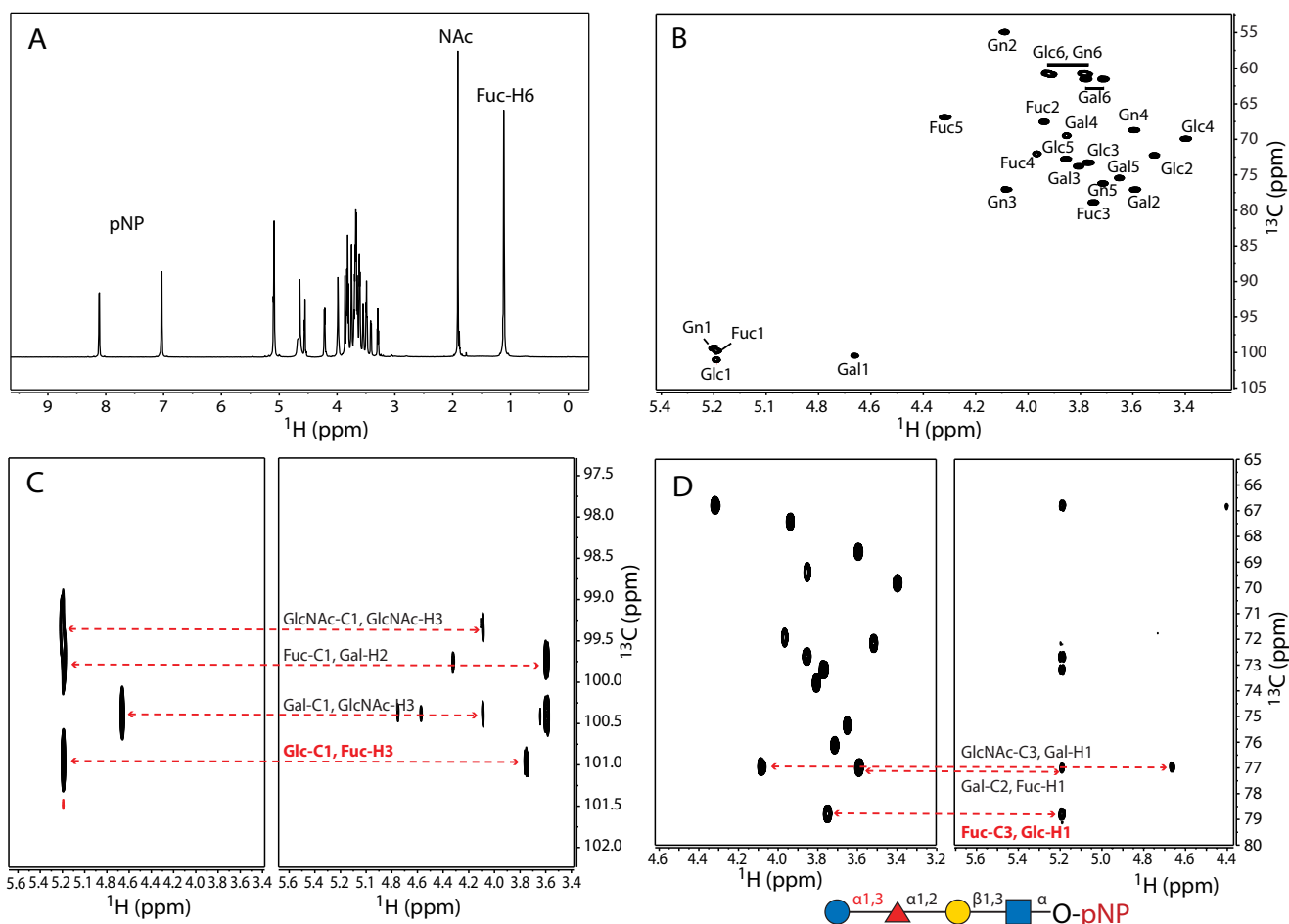
growth, and interpreting the difference between disrupting *phyA* and *glt1*, is characterizing the biochemical contributions of the *glt1* gene product.

## Glt1 is a novel α3-glucosyltransferase

Here we find that Glt1, an enzyme from the CAZy GT32 family of retaining glycosyltransferases, mediates the addition of the fourth sugar on *Toxoplasma* Skp1. This enabled us, in turn, to infer the sugar to be the pyranose form of D-Glc in α-linkage to the 3-position of the underlying Fuc. This conclusion is based on (i) the mass spectrometric analysis of Skp1 that shows that disruption of *glt1* leads to accumulation of the truncated trisaccharide form of Skp1 (FGGn-Skp1), suggesting an inability to transfer the fourth sugar, a residue of hexose (Figs. 4 and 5 and Table 2); (ii) the ability of parasite cell extracts to transfer Glc from UDP-α-D-Glc<sub>4</sub>p to FGGn-Skp1 by a mechanism that depends upon *glt1* (Fig. 6); and (iii) the ability of recombinant His<sub>6</sub>-Glt1 (Fig. 7) to directly, specifically, and efficiently catalyze the addition of Glc to the Skp1 trisaccharide (FGGn-Skp1) and synthetic glycan models (Fig. 8). UDP-Glc was the only UDP-sugar, of the six tested, to be efficiently hydrolyzed by purified Glt1 (Fig. 8A). GDP-Man was not a substrate (Fig. 8B). Furthermore, UDP-Glc is likely to be the native substrate of Glt1 due to its ability to efficiently transfer Glc to the synthetic acceptor FGGn-pNP (Fig. 8F) at a calculated turnover number of about 6/s at maximal velocity, which is rapid for a glycosyltransferase. The Glt1 protein sequence is related to the CAZy GT32 family, and the loss of enzymatic activity of point mutants that inactivate other members of this family sup-



**Figure 8. Substrate specificities of His<sub>6</sub>-Glt1.** A, comparison of glycosyl donor specificities, based on UDP-sugar hydrolysis. Recombinant Glt1 was incubated with each UDP-sugar at 50  $\mu$ M, and released UDP was detected using the UDP-Glo™ assay. Data are representative of three trials, and error bars represent S.D. for one reaction conducted in triplicate. B, comparison of utilization of UDP-Glc or GDP-Man as donors in a transferase assay toward 1.5 mM FG-Bn, which represents the non-reducing terminal disaccharide of Skp1, using a UDP or GDP (NDP) detection assay. S.D. values of three technical replicates are shown. C, dependence of transferase activity on UDP-Glc concentration, using 2 mM FG-Bn as the acceptor substrate in reactions monitored by incorporation of  $^3$ H from UDP-[ $^3$ H]Glc. Data represent the mean of two technical replicates  $\pm$  S.D. The  $K_m$  calculation is based on the Michaelis–Menten model. D, time dependence of reactions comparing FG-pNP with FGn-pNP, which models the full Skp1 trisaccharide, assayed as in C. The acceptor and donor substrates were 2 mM and 2  $\mu$ M, respectively. Error bars, S.D. between duplicate samples. E, comparison of different potential acceptor substrates, as in D, in reactions incubated for 2 h. F, dependence of transferase activity on FGn-pNP concentration, in the presence of 40  $\mu$ M UDP-Glc. Data are representative of two independent experiments, and the kinetic calculations are based on the Michaelis–Menten model. G, dependence on *Toxoplasma* FGn-Skp1 concentration, as in F. H, biochemical complementation of parasite extracts with His<sub>6</sub>-Glt1. Transferase reactions were conducted using UDP-[ $^3$ H]Glc as the donor and desalted S100 extracts of parental (RH $\Delta$ ) or glt1 $\Delta$  extracts as the acceptor in the presence or absence of His<sub>6</sub>-Glt1. Samples were separated by SDS-PAGE, and, after fixation, gel lanes were sliced from top to bottom. Incorporation was determined by scintillation counting.



**Figure 9. Linkage analysis of the His<sub>6</sub>-Glt1 reaction product.** FGn-pNP was incubated with UDP-Glc in the presence of His<sub>6</sub>-Glt1. The reaction product was recovered on a C<sub>18</sub> Sep-Pak cartridge and subjected to NMR analysis. A, 1D 900-MHz <sup>1</sup>H NMR spectrum of Glc- $\alpha$ 1,3-Fuc- $\alpha$ 1,2-Gal- $\beta$ 1,3-GlcNAc- $\alpha$ -pNP. B, 2D <sup>1</sup>H-<sup>13</sup>C HSQC spectrum. All expected peaks are observed and annotated in this region except for Fuc-6 and GlcNAc methyl groups, which are annotated in A. Numbers indicate ring position. C, 2D <sup>1</sup>H-<sup>13</sup>C HMBC spectrum region demonstrating anomeric carbon to ring proton correlations (red dashed lines). D, 2D <sup>1</sup>H-<sup>13</sup>C HMBC spectrum region demonstrating ring carbon to anomeric proton correlations (red dashed lines) and schematic diagram of the Tg Skp1 tetrasaccharide attached to pNP with linkages indicated. Glycan symbols used are adopted from Ref. 20. In C and D, interglycosidic correlations are annotated in red.

ports this association (data not shown). The CAZy GT32 family includes several characterized retaining glycosyltransferases that use either UDP-Gal or GDP-Man, although none are known to utilize UDP-Glc, and catalyze formation of  $\alpha$ 1,3,  $\alpha$ 1,4, or  $\alpha$ 1,6 linkages. The NMR-HMBC analysis confirms the  $\alpha$ -linkage and furthermore establishes glycosidic attachment to the 3-OH of the underlying Fuc (Fig. 9).

### Skp1 is probably the natural substrate of Glt1

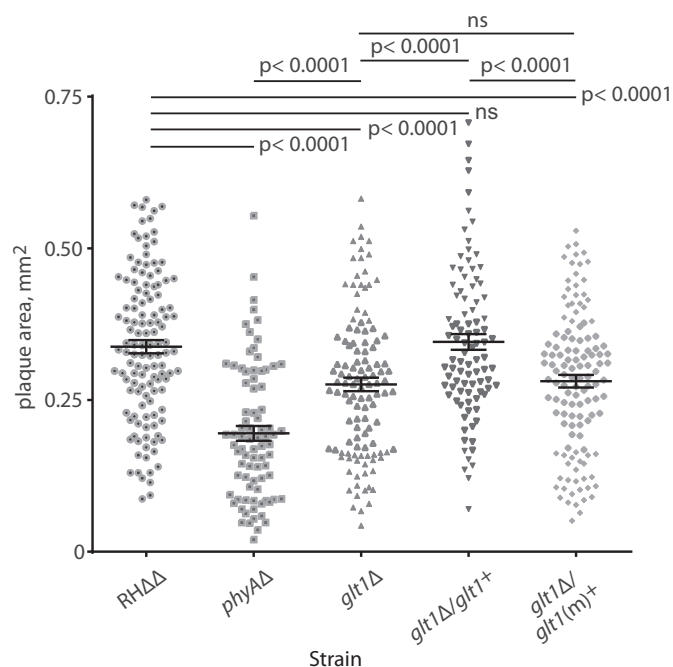
Glt1 and Skp1 both reside in cytoplasmic and possibly nuclear compartments based on the absence of apparent signal peptides or transmembrane domains, consistent with their recovery from cytosolic *Toxoplasma* extracts. Glt1 has a strong preference for non-reducing terminal Fuc in the context of the native Skp1 trisaccharide, relative to the non-reducing disaccharide, and did not modify GlcNAc or Gal in different contexts (Fig. 8, D and E). Furthermore, it did not modify Fuc alone, indicating that it will not target the abundant O-Fuc modifications in the nucleus of *Toxoplasma* (21). FGn-Skp1 is also an excellent substrate, with an apparent  $K_m$  of 4  $\mu$ M, dramatically lower than that of FGn-pNP and consistent with favorable

selectivity for the trisaccharide in the context of Skp1. What FGn-Skp1 gains in  $K_m$  is, however, sacrificed in  $V_{max}$ . Nevertheless, at 1  $\mu$ M concentrations, FGn-Skp1 is still calculated to enjoy a 15-fold catalytic advantage over FGn-pNP. Evidence that Skp1 is a primary target of Glt1 comes from a biochemical complementation experiment, in which Skp1 is the only acceptor substrate detected after incubation of *glt1* $\Delta$  extracts with His<sub>6</sub>-Glt1 and radioactive UDP-Glc (Fig. 8H). Although other substrates could have been missed because they were excluded from or not accessible in the extract, or alternatively modified in the absence of Glt1, the interpretation that Glt1 is dedicated to Skp1 is consistent with evidence that the two earlier glycosyltransferases in the pathway are specific for Skp1 (3) and with more extensive evidence that the Skp1 modification pathway is specific for Skp1 in *Dictyostelium* (4, 14).

### Functional variations between *Toxoplasma* Glt1 and *Dictyostelium* AgtA

Skp1 from *Dictyostelium* also possesses a pentasaccharide on the corresponding 4-hydroxyproline residue, and evidence suggests that the core trisaccharide of *Toxoplasma* Skp1 is identi-





**Figure 10. Role of Glt1 in parasite growth.** Human foreskin fibroblast monolayers were inoculated with the indicated strains in parallel, at a multiplicity of infection of 0.002. After 5.5 days, monolayers were stained with crystal violet. The images were digitized, and areas of manually encircled plaques, representing cleared areas of the monolayer due to cell lysis, were measured and analyzed using GraphPad Prism. Average plaque areas  $\pm$  S.E. (error bars) from a representative experiment are shown. ns, not significant.

cal to that of the recently confirmed (9) core trisaccharide of *Dictyostelium* Skp1,  $\text{Fuca}1,2\text{Gal}\beta1,3\text{GlcNAc}\alpha1-$ . The evidence for identity is based primarily on sequence homologies of their Gnt1 and PgtA enzymes (3) and confirmation of their donor substrates as UDP-GlcNAc and as UDP-Gal and GDP-Fuc, respectively. Identity of the trisaccharide cores is also supported by the robust activity of  $\text{His}_6$ -Glt1 toward endogenous *Toxoplasma* Skp1 in *glt1* $\Delta$  extracts (Fig. 8H). More direct confirmation of the linkages of the core trisaccharide is confounded by the small amounts of the intracellular pathogen that are available and difficulties encountered in expressing soluble protein from *Toxoplasma* *pgtA* cDNA in *E. coli* or *Dictyostelium*.<sup>6</sup>

The fourth sugar in the *Dictyostelium* Skp1 glycan, an  $\alpha$ Gal, differs from the  $\alpha$ Glc in *Toxoplasma* but conserves the 3-linkage to Fuc. The enzyme catalyzing its addition, AgtA, belongs to CAZy family GT77 and is evolutionarily unrelated to Glt1. AgtA also catalyzes the addition of the fifth sugar, an  $\alpha$ Gal that is 3-linked to the fourth sugar to form a linear chain. In contrast, in *Toxoplasma*, the fifth sugar is added by a separate enzyme.<sup>7</sup> The mechanism of recognition of FGn-Skp1 is distinct for the two enzymes. Whereas AgtA strongly prefers FG- relative to Fuc- or FGn- (as conjugates to small aglycons) as an acceptor (22), Glt1 strongly favors the full trisaccharide and is unable to modify Fuc alone (Fig. 8, D and E). Genetic and biochemical studies reveal that *Dictyostelium* AgtA has an independent function in modulating Skp1 activity that probably involves a

physical interaction via a separate WD40 repeat domain (13). Unlike AgtA, Glt1 lacks an identifiable domain separate from the catalytic domain, although it does possess short sequences at various positions within the conserved catalytic domain (supplemental Fig. S2). However, these insert sequences, variations of which are commonly observed in *Toxoplasma* proteins (23), are not conserved at the sequence level in another apicomplexan expected to harbor the Skp1 modification pathway, *Sarcocystis neurona*. Furthermore, the insert sequences tend to be absent from other protists that possess *glt1*-like genes. Therefore, Glt1 lacks a structural basis for the second function possessed by AgtA.

Structural studies of *Dictyostelium* Skp1 suggest that the full-length glycan encourages an ensemble of conformations that promote interactions with at least certain FBPs (9). The trisaccharide form of *Dictyostelium* Skp1 exhibits intermediate interaction with two of the FBPs (6), and, if an analogous mechanism operates in *Toxoplasma*, this might explain why the *glt1* deletion results in a growth phenotype intermediate between complete absence of the glycan and its full assembly. Further studies, to be conducted when the final linkage of the *Toxoplasma* glycan is ascertained, will address this possible explanation.

#### Glt1 is ancestral for Skp1 glycosylation in protists

A comprehensive reconstruction of the evolution of CAZy GT32-related sequences indicates that Glt1-like sequences form a distinct clade within the GT32 group and have evolutionarily diverged from other members through variations in the GT domain. This interpretation is supported by conservation of amino acids at 10 positions throughout the protein, including three that are exclusive to the Glt1 clade (supplemental Fig. S2). Glt1-like sequences are found only in protists (Fig. 2), with one exception in a bacterium, which might reflect horizontal transfer. But the broad distribution within this diverse phylogeny suggests that Glt1 was present in the last common protistan ancestor and selectively lost or modified where it did not afford a selective advantage in, for example,  $\text{O}_2$  sensing. Furthermore, the Glt1 clade is populated only by protists that also possess CAZy GT74-like sequences predicted to encode the Skp1 PgtA-like enzyme that assembles the Fuc residue on which Glt1 acts in *Toxoplasma*. This evidence of co-evolution implicates all members of this clade in the assembly of the Skp1 glycan. However, Glt1-like sequences are absent from the amoebozoans (using BlastP search at  $E < 10^{-1}$ ), which include *Dictyostelium*, where the CAZy family GT77 member AgtA performs a related biochemical function. The simplest explanation is that Glt1 is the ancestral Skp1 glycosyltransferase whose function was replaced in amoebozoans, which might have occurred to compensate for a loss of Glt1 or the final glycosyltransferase (Gat1),<sup>7</sup> or because of the selective advantage afforded by AgtA's additional function in Skp1 suppression (13).

#### Experimental procedures

##### Parasite manipulations

Strain *RH* $\Delta$ *ku80* $\Delta$ *hxgprt* (*RH* $\Delta$  $\Delta$ ) of *T. gondii* was cultured on preformed monolayers of human foreskin fibroblasts (ATCC catalog no. SCRC-1041), or BJ fibroblasts (ATCC cata-

<sup>6</sup> H. van der Wel, and C. M. West, unpublished data.

<sup>7</sup> M. Mandalasi, K. Rahman, and H. Kim, unpublished data.

log no. CRL-2522) if indicated, in the presence of Complete Medium, which consisted of DMEM (Corning Inc.) supplemented with 10% (v/v) FBS, 2 mM L-glutamine, and 100 units/ml penicillin/streptomycin (Corning) at 37 °C in a humidified CO<sub>2</sub> (5%) incubator. RHΔ*ku80Δglt1* (*glt1Δ*) strain was selected and maintained in the same medium with 25 μg/ml mycophenolic acid (Sigma) and 25 μg/ml xanthine (Sigma). The medium of complemented *glt1Δ* strains was supplemented with 250 μg/ml 6-thioxanthine (Matrix Scientific). Strains were cloned by limiting dilution on BJ fibroblasts in 96-well plates. Plaque assays were conducted in media without selection drug (3). Cytosolic S100 extracts were prepared from tachyzoites isolated from infected fibroblasts and desalted, and soluble S16 urea extracts were prepared as before (3).

## *glt1* gene replacement and complementation

DNA for the gene replacement was generated from pminiGFP.ht, in which the *hxpprt* gene is flanked by multiple cloning sites, as described (3). Briefly, 5'-flanking and 3'-flanking targeting sequences of *glt1* from RHΔΔ were PCR-amplified with primer pairs a and a' and pairs b and b', respectively (supplemental Table S1). The 5'-fragment was released by digestion with ApaI and HindIII and cloned into similarly digested pminiGFP.ht. The 3'-fragment was similarly inserted using XbaI and NotI. The resulting vector was linearized with ApaI and electroporated into strain RHΔΔ. Transformants were selected, and GFP-negative parasites were cloned by limiting dilution. Genomic DNA was prepared and screened by PCR (3).

To complement a *glt1Δ* clone, the *hxpprt* cassette of pminiGFP.ht was replaced with a ~5-kb gDNA fragment containing the *glt1* coding region from RHΔΔ (2860 nucleotides) plus ~1 kb each of 5'-flanking and 3'-flanking DNA, using the complementary annealing mediated by exonuclease cloning method (25). Briefly, the vector and insert were PCR-amplified separately for 20 cycles by Q5 high-fidelity DNA polymerase (primers in supplemental Table S1). The amplified insert contained 15-base overhangs matching the termini of the amplified vector. The gel-purified amplicons were mixed at a 1:3 molar vector/insert ratio (total ~100 ng), incubated with T4 DNA polymerase (Novagen) at 22 °C for 2.5 min to generate 5'-overhangs, incubated at 75 °C for 20 min to inactivate the polymerase, and annealed at 50 °C for 30 min. 2 μl was transformed into *E. coli* Top10-competent cells. The recovered plasmid was electroporated into *glt1Δ* parasites, which were subjected to selection (3). A triple point mutant (D348N/D363A/D365A) of the complementation plasmid was generated by site-directed mutagenesis (primers in supplemental Table S1) as described (12) and separately electroporated.

## Mass spectrometry of Skp1 peptides

Skp1 was immunoprecipitated from urea-solubilized parasite extracts as described (3), except that the sample was first precleared by incubation with 50 μl of anti-rabbit antibody bound beads for 1 h at 4 °C. The dried, purified Skp1 samples were dissolved in 100 μl of 10 mM dithiothreitol in 50 mM NH<sub>4</sub>HCO<sub>3</sub>, incubated at 56 °C for 1 h, alkylated with 22.5 mM iodoacetamide for 45 min in the dark, and digested with trypsin (Promega) at 37 °C overnight. The resulting peptides were

recovered by addition to a C18 spin column (MicroSpin<sup>TM</sup> column, The Nest Group), elution with 0.1% formic acid in 80% acetonitrile, and drying under vacuum. Peptides were reconstituted in 19.5 μl of solvent A (0.1% formic acid) and 0.5 μl of solvent B (0.1% formic acid in 80% acetonitrile), separated on an Acclaim PepMap RSLC C18 column (75 μm × 15 cm), and eluted into the ion source of an Orbitrap Fusion Lumos Tribrid<sup>TM</sup> mass spectrometer (Thermo Fisher Scientific) with a linear gradient consisting of 0.5–100% solvent B over 150 min at a flow rate of 200 nl/min. The spray voltage was set to 2.2 kV, and the temperature of the heated capillary was set to 280 °C. Full MS scans were acquired from *m/z* 300 to 2000 at 120,000 resolution, and MS<sup>2</sup> scans following collision-induced fragmentation were collected in the ion trap for the most intense ions in the Top-Speed mode within a 3-s cycle using Fusion instrument software (version 2.0, Thermo Fisher Scientific). The acquired raw spectra were analyzed using SEQUEST (Proteome Discoverer version 1.4, Thermo Fisher Scientific) with a full MS peptide tolerance of 20 ppm and MS<sup>2</sup> peptide fragment tolerance of 0.5 Da and filtered to generate a 1% target decoy peptide-spectrum-match false discovery rate for protein assignments. Spectra assigned as glycosylated TgSkp1 peptides were manually validated.

## Preparation of Toxoplasma FGn-Skp1

The *E. coli* dual expression plasmid for *Dictyostelium* Skp1A (5) and *Dictyostelium* PhyA was modified by substitution with synthetic cDNA encoding *Toxoplasma* Skp1 and introduced into *E. coli* together with the plasmid encoding *Dictyostelium* DdDp-Gnt1. Gn-Skp1 was purified as for its *Dictyostelium* counterpart and modified *in vitro* using *Dictyostelium* FLAG-PgtA. The reaction consisted of 6.25 μM *Toxoplasma* Gn-Skp1, 112 nM *Dictyostelium* FLAG-PgtA, 25 μM UDP-Gal, 50 μM GDP-Fuc, 120 mM NaCl in 50 mM Tris-HCl (pH 7.5). The reaction was monitored by dot blotting onto nitrocellulose filters and probing with mAb 1C9, which is specific for *Dictyostelium* Gn-Skp1 relative to other modified glycoforms (15), and a 1:1000 dilution of polyclonal antibody UOK104, which is specific for *Dictyostelium* FGn-Skp1 relative to other glycoforms (26). Alexa680-coupled secondary antibodies were applied and detected in an Odyssey infrared scanner (LI-COR). Similar specificity was observed for the *Dictyostelium* and *Toxoplasma* glycoforms, and the glycosylation reaction was taken to completion within the sensitivity of the method (data not shown).

## Expression and purification of recombinant His<sub>6</sub>-Glt1

The predicted coding sequence of Glt1 (TGGT1\_205060) was codon-optimized for expression in *E. coli*, chemically synthesized by GenScript (Piscataway, NJ), and inserted into the pUC57 vector between its NdeI and BamHI sites (supplemental Fig. S1). After treatment with NdeI and BamHI, the released coding fragment was inserted into similarly digested pET15b (Invitrogen), which resulted in the full-length 605-amino acid coding sequence preceded by an N-terminal His<sub>6</sub> tag and tobacco etch virus protease cleavage site (MGSSHHHHHSS-GRENLYFQGH-). *E. coli* Gold cells expressing His<sub>6</sub>-Glt1 were incubated for 24 h in 8 × 1 liter of Terrific Broth medium in the presence of 100 μg/ml ampicillin, 2 g/liter lactose, and 125 μM

isopropyl 1-thio- $\beta$ -D-galactopyranoside at 20 °C. Pilot studies showed that the autoinduction during bacterial growth (27) was much superior to standard induction at high density for expression of soluble Glt1 (data not shown). After 24 h, cells were collected by centrifugation at  $2000 \times g$  for 10 min and resuspended in 50 mM  $\text{Na}^+/\text{K}^+$  phosphate (pH 7.8), 300 mM NaCl, 2 mM benzamidine, 0.5  $\mu\text{g}/\text{ml}$  pepstatin A, 5  $\mu\text{g}/\text{ml}$  aprotinin, 5  $\mu\text{g}/\text{ml}$  leupeptin, and 0.5 mM phenylmethylsulfonyl fluoride at 4 °C;  $\text{Na}^+/\text{K}^+$  phosphate buffer was prepared by titrating monosodium phosphate into dipotassium phosphate of equal molarity. Bacteria were lysed using a probe sonicator (model 500, Thermo Fisher Scientific). The lysate was immediately centrifuged at  $21,000 \times g$  for 30 min at 4 °C, and the supernatant (S21) was applied to a column containing 1.5 ml of  $\text{Co}^{2+}$  TALON resin (Clontech) pre-equilibrated at 4 °C in the buffer described above. Protein was eluted with 300 mM imidazole in the same buffer. The major  $A_{280}$  peak was dialyzed against 40 mM Tris-HCl (pH 8.0) overnight, and applied to a 1-ml Hi-Trap Q-Sepharose column (GE Healthcare) pre-equilibrated at 4 °C in 40 mM Tris-HCl (pH 8.0), 2 mM DTT, 2 mM  $\text{MgCl}_2$ , 10% (v/v) glycerol, aprotinin, and leupeptin (as above). The column was eluted with a 0–1 M gradient of NaCl in the same buffer. Fractions with the highest enzymatic activity were confirmed for the presence of His<sub>6</sub>-Glt1 by Western blotting with anti-His<sub>6</sub> monoclonal antibody (Novagen catalog no. 70796-3), pooled, and frozen as aliquots at –80 °C.

### Western blot analysis

SDS-PAGE was performed on 4–12% preformed gels (NuPAGE Novex, Invitrogen), Western blotted, probed with primary and fluorescent secondary antibodies, and scanned as described previously (3).

### Glt1 enzyme activity assays

Hydrolysis of UDP-sugars was conducted using the UDP-Glo assay (Promega) as described (18). Briefly, His<sub>6</sub>-Glt1 (after Q-column purification) was incubated in the presence of 50  $\mu\text{M}$  sugar nucleotides in 20- $\mu\text{l}$  reactions containing 50 mM HEPES-NaOH (pH 7.4), 2 mM  $\text{MnCl}_2$ , 5 mM DTT at 37 °C for 16 h, and activity was quantitated based on conversion of the UDP reaction product to ATP.

Glt1 glucosyltransferase activity was assayed as the transfer of  $^3\text{H}$  from UDP-[6- $^3\text{H}$ ]Glc to synthetic acceptors or FGn-Skp1. Synthetic acceptors were Fuca1-*para*-nitrophenol (Sigma), Fuca1,2Gal $\beta$ 1-Bn (FG-Bn), Fuca1,2Gal $\beta$ 1,3GlcNAc $\alpha$ 1-pNP (FGn-pNP), Gal $\beta$ 1,3GlcNAc $\alpha$ 1-Bn, GlcNAc $\alpha$ 1-Bn, and Gal $\beta$ 1,3Fuca1-Bn (22, 28). The standard reaction consisted of Q-column-purified His<sub>6</sub>-Glt1, and 2 mM synthetic acceptor or 1  $\mu\text{M}$  FGn-TgSkp1, 50 mM HEPES-NaOH (pH 7.4), 2  $\mu\text{M}$  UDP-Glc (Sigma-Aldrich), including 0.5  $\mu\text{Ci}$  of UDP-[6- $^3\text{H}$ ]Glc (38.2  $\mu\text{Ci}/\text{nmol}$ ; PerkinElmer Life Sciences), 2 mM  $\text{MnCl}_2$ , 50 mM NaCl, 5 mM DTT, and 2 mg/ml bovine serum albumin (Sigma), in a final reaction volume of 20  $\mu\text{l}$  that was incubated for 1 h at 37 °C. Concentrations, pH (using a mixture of MES and HEPES), and reaction times were varied as indicated and were confirmed to be linearly dependent on enzyme concentration and reaction time. Reactions using synthetic acceptor substrates were terminated by the addition of 1 ml of ice-cold 2 mM

sodium EDTA (pH 8.0) and applied to and eluted from a Sep-Pak C<sub>18</sub> cartridge (360 mg, Waters Corp.) and counted for radioactivity as described (3). Activity from a reaction lacking an acceptor substrate was subtracted as nonspecific background. Reactions using Skp1 as an acceptor substrate were stopped by the sequential addition of 863  $\mu\text{l}$  of 2 mM Na-EDTA (pH 8.0), 20  $\mu\text{l}$  of 20 mg/ml BSA, and 85% TCA to a final concentration of 10% and incubation on ice. The precipitate was recovered by filtration on a 0.45- $\mu\text{m}$  glass filter, washed with ice-cold 10% TCA and ice-cold acetone, and transferred into Biosafe NA (RPI) scintillation mixture and counted as above.

Alternatively, Glt1 transferase activity was assayed based on release of GDP or UDP from unlabeled GDP-Man or UDP-Glc, using GDP-Glo or UDP-Glo assays (Promega) according to the manufacturer's protocol. Conditions were as described above.

Kinetic parameters were determined assuming Michaelis–Menten kinetics using GraphPad Prism software. Acceptor substrate kinetics were determined at 40  $\mu\text{M}$  UDP-Glc, and donor substrate kinetics were determined at 2 mM FG-Bn.

For assay of transferase activity in parasite extracts, the typical 50- $\mu\text{l}$  reaction volume containing 30  $\mu\text{l}$  of S100 fraction (1 mg of protein/ml), 50 pmol of FGn-DdSkp1 (5), 1  $\mu\text{Ci}$  of UDP-[ $^3\text{H}$ ]Glc or UDP-[ $^3\text{H}$ ]Gal (13), in 50 mM HEPES-NaOH (pH 7.4), 10 mM  $\text{MgCl}_2$ , 2 mM  $\text{MnCl}_2$ , 2 mM DTT, 3 mM NaF, and protease inhibitors, was incubated at 37 °C for 1.5 or 3 h. Reactions were stopped by the addition of 4 $\times$  Laemmli electrophoresis sample buffer, and incorporation was after SDS-PAGE as described (3). The chemical form of incorporated radioactivity was determined after electroblot transfer to a polyvinylidene difluoride membrane, acid hydrolysis, and high-pH anionic exchange chromatography, using 1.5 nmol each of L-Fuc, D-Glc, D-Gal, and D-Man as internal standards (3).

For assay of acceptor substrate activity in extracts, reactions were modified to contain 2  $\mu\text{Ci}$  of UDP-[ $^3\text{H}$ ]Glc, 170  $\mu\text{l}$  of desalted S100 fraction (1 mg/ml), His<sub>6</sub>-Glt1 in a final volume of 200  $\mu\text{l}$ . After incubation for 3 h at 37 °C, the reaction was concentrated by centrifugal ultrafiltration using a Nanosep 3K concentrator (Pall Corp.). The samples were resolved by SDS-PAGE, and each lane was sliced into 36 equal pieces, which were counted for radioactivity as above.

### Glc-Fuc linkage analysis

To determine the linkage of the Glt1 product, 4  $\mu\text{mol}$  of FGn-pNP and 8  $\mu\text{mol}$  of UDP-Glc were incubated with 34  $\mu\text{g}$  (~0.65 nmol) of His<sub>6</sub>-Glt1 and 40 units of calf intestinal alkaline phosphatase (Promega) in 2 ml of reaction buffer (50 mM HEPES-NaOH (pH 8.0), 5 mM DTT, 8 mM  $\text{MnCl}_2$ , and 10 mM  $\text{MgCl}_2$ ) at 37 °C for 5 h. The reaction was terminated with 8 ml of 5 mM Na-EDTA, pH 8.0. The product was recovered using a Sep-Pak C<sub>18</sub> cartridge as above, dried under  $\text{N}_2$ , and reconstituted in 50  $\mu\text{l}$  of  $\text{H}_2\text{O}$ . Quantitative conversion to the tetrasaccharide was confirmed using MALDI-TOF mass spectrometry (ABSciEx 5800, Applied Biosystems), and the concentration was determined spectrophotometrically using an extinction coefficient of  $1.15 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1}$  at 300 nm (22). 3  $\mu\text{mol}$  of dried FGn-Skp1 was resuspended in 99.96%  $\text{D}_2\text{O}$  and analyzed on an Agilent 900-MHz DD2 spectrometer equipped with a 5-mm cryogenically cooled probe. NMR experiments were



performed at 25 °C after stabilization and shimming and used standard pulse sequences (PRESAT, gCOSY, zTOCSY, HSQCAD, and gHMBCAD) from the Agilent library. Two-dimensional data were collected with default values except for increased digital resolution in both dimensions. Spectra were processed with MestReNova software (Mestrelab Research S.L.). Peaks of the <sup>1</sup>H–<sup>13</sup>C HSQC spectra were provisionally assigned by comparison with predicted chemical shifts calculated by the CASPER program, and all signals and residue linkages were confirmed from analysis of the 2D data.

## Phylogenetic analysis

Full-length sequences for all of the experimentally characterized GT32 family proteins listed in the CAZy database (accessed May 10, 2017, <http://www.cazy.org/>)<sup>8</sup> (29) were collected. To identify Glt1-related sequences, *T. gondii* (EPR61400.1) and *Vitrella brassicaformis* (CEM02366.1) Glt1 protein sequences were used as BLAST queries to search against the NCBI nr database and available protist proteome databases (Ensembl protist and JGI) (<http://protists.ensembl.org/index.html> and <http://genome.jgi.doe.gov/>).<sup>8</sup> The putative ortholog from the chromerid *V. brassicaformis*, a close relative of apicomplexa that is predicted to express the *Toxoplasma*-like Skp1 modification pathway, was included due to its lack of apicomplexan-specific insertions within its proteins (30). For the GT74 family, the only characterized sequence from *D. discoideum* (AAF82378.1) was used as a BLAST query. For both families (GT32 and GT74), the collected BLAST best-hit sequences were then aligned using the MAFFT L-INS-I strategy (31), and the alignment was used to generate a hidden Markov model (HMM) (<http://hmmer.org/>)<sup>8</sup> (32). The built HMM profile was used to identify additional related sequences in NCBI nr from diverse taxonomic groups using an *e*-value cut-off of 1e–5. A total of 68 GT32-related sequences from bacteria, protists, fungi, chlorophyte, and metazoans were collected and used for the phylogenetic tree construction. Hits collected for the GT74 family were used to identify the presence of a GT74-related sequence in the species.

The 68 full-length sequences representing the GT32 family members were aligned using an MAFFT L-INS-I strategy. The boundaries of the GT32 domain were marked by identifying conserved regions in the alignment. The sequences were initially trimmed to extract only the GT32 domain and realigned using the same strategy. This domain alignment was further refined by removing large insert segments and poorly aligned regions. The trimmed alignment was used to generate the final phylogenetic tree. FastTree (33) was used to build the tree with default parameters using the following options: -wag for the WAG (34) model of amino acid evolution, -gamma for the rescaling of the branch lengths and the computation of Gamma20-based likelihood scores, -pseudo to add pseudo counts for highly gapped segments in the alignment. Local support values for the internal nodes were computed by FastTree using the Shimodaira–Hasegawa test and are displayed in the figure. Trees generated using full-length GT domain alignments also

resulted in similar tree topologies. Runs using RAxML also generated similar topologies (not shown). Functional annotations were collected from the CAZy database for the characterized sequences. SignalP3.0 ([www.cbs.dtu.dk/services/SignalP-3.0/](http://www.cbs.dtu.dk/services/SignalP-3.0/))<sup>8</sup> was used to predict signal peptide sequences, using separate runs for eukaryotes, Gram-negative bacteria, and Gram-positive bacteria, and TMHMM version 2.0 ([www.cbs.dtu.dk/services/TMHMM/](http://www.cbs.dtu.dk/services/TMHMM/))<sup>8</sup> was used to predict the transmembrane regions.

**Author contributions**—K. R. conducted most of the experiments, analyzed the results, assembled the figures, and wrote most of the first draft. P. Z. conducted the mass spectrometry and interpreted the data together with L. W. M. M. and H. V. D. W. contributed to the enzymatic assays and high-performance anion exchange chromatography analyses. M. O. S. performed the UDP-sugar hydrolysis assays, and M. O. S. and J. N. G. performed the linkage analyses. H. W. K. prepared and purified the Glt1 enzyme. K. M. synthesized the glycan substrates. R. T. conducted the phylogenetic analyses which were edited by N. K. and C. M. W. C. M. W. conceived most ideas and experimental strategies for the project and compiled the final draft, which was approved by all authors.

**Acknowledgments**—We thank Dr. Hicham Zegzouti (Promega R&D) for providing the GDP-Glo assay kit, and we are grateful to Ira Blader (State University of New York, Buffalo, NY) for reviewing the manuscript.

## References

- 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
- Skaar, J. R., Pagan, J. K., and Pagano, M. (2013) Mechanisms and function of substrate recruitment by F-box proteins. *Nat. Rev. Mol. Cell Biol.* **14**, 369–381
- 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
- West, C. M., and Blader, I. J. (2015) Oxygen sensing by protozoans: how they catch their breath. *Curr. Opin. Microbiol.* **26**, 41–47
- 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
- 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 SCF complexes in *Dictyostelium*. *Mol. Cell. Proteomics* **14**, 66–80
- Schulman, B. A., Carrano, A. C., Jeffrey, P. D., Bowen, Z., Kinnucan, E. R. E., Finnin, M. S., Elledge, S. J., Harper, J. W., Pagano, M., and Pavletich, N. P. (2000) Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. *Nature* **408**, 381–386
- Zimmerman, E. S., Schulman, B. A., and Zheng, N. (2010) Structural assembly of cullin-RING ubiquitin ligase complexes. *Curr. Opin. Struct. Biol.* **20**, 714–721
- 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) O<sub>2</sub> sensing-associated glycosylation exposes the F-box—combining site of the *Dictyostelium* Skp1 subunit in E3 ubiquitin ligases. *J. Biol. Chem.* **292**, 18897–18915
- Xu, Y., Wang, Z. A., Green, R. S., and West, C. M. (2012) Role of the Skp1 prolyl-hydroxylation/glycosylation pathway in oxygen dependent submerged development of *Dictyostelium*. *BMC Dev. Biol.* **12**, 31
- Bonner, J. T., and Lamont, D. S. (2005) Behavior of cellular slime molds in the soil. *Mycologia* **97**, 178–184

<sup>8</sup> Please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third party hosted site.

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. Schafer, C. M., Sheikh, M. O., Zhang, D., and West, C. M. (2014) Novel regulation of Skp1 by the *Dictyostelium* AgtA  $\alpha$ -galactosyltransferase involves the Skp1-binding activity of its WD40 repeat domain. *J. Biol. Chem.* **289**, 9076–9088
14. West, C. M., Wang, Z. A., and van der Wel, H. (2010) A cytoplasmic prolyl hydroxylation and glycosylation pathway modifies Skp1 and regulates O<sub>2</sub>-dependent development in *Dictyostelium*. *Biochim. Biophys. Acta* **1800**, 160–171
15. Wang, Z. A., van der Wel, H., Vohra, Y., Buskas, T., Boons, G. J., and West, C. M. (2009) Role of a cytoplasmic dual-function glycosyltransferase in O<sub>2</sub> regulation of development in *Dictyostelium*. *J. Biol. Chem.* **284**, 28896–28904
16. Blader, I. J., and Koshy, A. A. (2014) *Toxoplasma gondii* development of its replicative niche: in its host cell and beyond. *Eukaryot. Cell* **13**, 965–976
17. Wiggins, C. A., and Munro, S. (1998) Activity of the yeast MNN1  $\alpha$ -1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7945–7950
18. Sheikh, M. O., Halm, 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
19. Rönnols, J., Pendrill, R., Fontana, C., Hamark, C., d'Ortoli, T. A., Engström, O., Ståhle, J., Zaccheus, M. V., Sävén, E., Hahn, L. E., Iqbal, S., and Widmalm, G. (2013) Complete <sup>1</sup>H and <sup>13</sup>C NMR chemical shift assignments of mono- to tetrasaccharides as basis for NMR chemical shift predictions of oligosaccharides using the computer program CASPER. *Carbohydr. Res.* **380**, 156–166
20. 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., et al. (2015) Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25**, 1323–1324
21. Bandini, G., Haserick, J. R., Motari, E., Ouologuem, D. T., Lourido, S., Roos, D. S., Costello, C. E., Robbins, P. W., and Samuelson, J. (2016) O-Fucosylated glycoproteins form assemblies in close proximity to the nuclear pore complexes of *Toxoplasma gondii*. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11567–11572
22. Ketcham, C., Wang, F., Fisher, S. Z., Ercan, A., van der Wel, H., Locke, R. D., Sirajud-Doulah, K., Matta, K. L., and West, C. M. (2004) Specificity of a soluble UDP-galactose:fucoside  $\alpha$ 1,3-galactosyltransferase that modifies the cytoplasmic glycoprotein Skp1 in *Dictyostelium*. *J. Biol. Chem.* **279**, 29050–29059
23. Narasimhan, J., Joyce, B. R., Naguleswaran, A., Smith, A. T., Livingston, M. R., Dixon, S. E., Coppens, I., Wek, R. C., and Sullivan, W. J., Jr. (2008) Translation regulation by eukaryotic initiation factor-2 kinases in the development of latent cysts in *Toxoplasma gondii*. *J. Biol. Chem.* **283**, 16591–16601
24. 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
25. Sun, S., Huang, H., Qi, Y. B., Qiu, M., and Dai, Z. M. (2015) Complementary annealing mediated by exonuclease: a method for seamless cloning and conditioning site-directed mutagenesis. *Biotechnol. Biotechnol. Equip.* **29**, 105–110
26. Chinoy, Z. S., Schafer, C. M., West, C. M., and Boons, G. J. (2015) Chemical synthesis of a glycopeptide derived from Skp1 for probing protein specific glycosylation. *Chemistry* **21**, 11779–11787
27. Li, Z., Kessler, W., van den Heuvel, J., and Rinas, U. (2011) Simple defined autoinduction medium for high-level recombinant protein production using T7-based *Escherichia coli* expression systems. *Appl. Microbiol. Biotechnol.* **91**, 1203–1213
28. Rana, S. S., Barlow, J. J., and Matta, K. L. (1981) The chemical synthesis of O- $\alpha$ -L-fucopyranosyl-(1 $\rightarrow$ 2)-O- $\beta$ -D-galactopyranosyl-(1 $\rightarrow$ 3)-O-[ $\alpha$ -L-fucopyranosyl-(1 $\rightarrow$ 4)]-2-acetamido-2-deoxy-D-glucopyranose, the Lewis b blood-group antigenic determinant. *Carbohydr. Res.* **96**, 231–239
29. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495
30. Fuss, Z., Masařová, P., Kručinská, J., Esson, H. J., and Oborník, M. (2017) Budding of the alveolate alga *Vitrella brassicaformis* resembles sexual and asexual processes in apicomplexan parasites. *Protist* **168**, 80–91
31. Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066
32. Eddy, S. R. (2011) Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195
33. Price, M. N., Dehal, P. S., and Arkin, A. P. (2010) FastTree 2: approximately maximum-likelihood trees for large alignments. *PLOS One* **5**, e9490
34. Whelan, S., and Goldman, N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691–699

**Characterization of a cytoplasmic glucosyltransferase that extends the core trisaccharide of the *Toxoplasma* Skp1 E3 ubiquitin ligase subunit**

Kazi Rahman, Msano Mandalasi, Peng Zhao, M. Osman Sheikh, Rahil Taujale, Hyun W. Kim, Hanke van der Wel, Khushi Matta, Natarajan Kannan, John N. Glushka, Lance Wells and Christopher M. West

*J. Biol. Chem.* 2017, 292:18644-18659.

doi: 10.1074/jbc.M117.809301 originally published online September 19, 2017

---

Access the most updated version of this article at doi: [10.1074/jbc.M117.809301](https://doi.org/10.1074/jbc.M117.809301)

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

Supplemental material:

<http://www.jbc.org/content/suppl/2017/09/19/M117.809301.DC1>

This article cites 34 references, 12 of which can be accessed free at

<http://www.jbc.org/content/292/45/18644.full.html#ref-list-1>



## SUPPLEMENT

*Characterization of a cytoplasmic glucosyltransferase that extends the core trisaccharide of the Toxoplasma  
Skp1 E3 ubiquitin ligase subunit*

**Kazi Rahman<sup>1,4,8</sup>, Msano Mandalasi<sup>1,5</sup>, Peng Zhao<sup>2</sup>, M. Osman Sheikh<sup>2</sup>, Rahil Taujale<sup>2,3</sup>, Hyun W.  
Kim<sup>1</sup>, Hanke van der Wel<sup>1</sup>, Khushi Matta<sup>6</sup>, Natarajan Kannan<sup>1,3</sup>, John N. Glushka<sup>2</sup>, Lance Wells<sup>1,2</sup>,  
Christopher M. West<sup>1,5,7</sup>**

<sup>1</sup>Department of Biochemistry & Molecular Biology, <sup>2</sup>Complex Carbohydrate Research Center, <sup>3</sup>Institute of Bioinformatics, University of Georgia, Athens, GA 30602 USA; <sup>4</sup>Departments of Microbiology & Immunology, and <sup>5</sup>Biochemistry & Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104 USA; <sup>6</sup>Chemical and Biological Engineering, State University of New York, Buffalo, NY 14260 USA

### Contents:

**SUPPLEMENTAL TABLE S1. Oligonucleotide sequences employed.**

**SUPPLEMENTAL FIGURE S1. Sequence of Glt1 cDNA.**

**SUPPLEMENTAL FIGURE S2. Alignment of Glt1-like sequences with other CAZy GT32 family proteins.**

## SUPPLEMENTAL TABLE S1. Oligonucleotide sequences employed.

### *Targeting sequence amplification:*

#### **Tggl1-disruption**

a)	Glt1F1	5'-flank	5'-end	5'-CCGGGCCCCAAGCAAACCCTGTCTCTCTAT	(ApaI)
a')	Glt1R1	5'-flank	3'-end	5'-GGAAGCTTCAATTCCACACTCAAGCGAATG	(HindIII)
b)	Glt1F2	3'-flank	5'-end	5'-GCTCTAGAGAAGATGCGCAGGATTCATCTA	(XbaI)
b')	Glt1R2	3'-flank	3'-end	5'-GCGCGCCGCTGTTAGGCGATGCTGTCTT	(NotI)

#### **Tgggl1 complementation** (*italics* denote complementary regions)

Vector (forward): 5'-GCAGCCACGATTTTCCAAG

Vector (Reverse): 5'-TATTGATGAACCCGTTGTTCC

Insert (Forward): 5'-ACGGGTTCATCAATAGCTGGCTGACGATCTTTAAGT

Insert (Reverse): 5'-GAAAATCGTGGCTGCCACATAACTCCTGCTTACAGAC

#### **Mutation of Tgggl1 complementation construct:**

<sup>363</sup>DVD<sup>365</sup> to <sup>363</sup>AVA<sup>365</sup>:

Forward: 5'-ggttttatagcctccatggccacagctgcgtatattcctgaat-3'

Reverse: 5'-attcaggaatatacgcagctgtggccatggaggctataaaacc-3'

#### **D348N:**

Forward: 5'-gcaggcgggcaatattcgattttgcacccgg-3'

Reverse: 5'-ccgggtgcaaaatcgaatattgccgcctgc-3'

#### **PCR confirmation of Tggl1 disruption and complementation**

##### **PCR 1**

Forward: 5'-CTGTGACTCAGAACTCCTCAAG

Reverse: 5'-CCCGAAGAACATCCGATTAGAG

##### **PCR 2**

Forward: 5'-GCTAGTGTACAGTTCCCACTAAG

Reverse: 5'-AGTCGCGGAACATCTCGTTGAAGT

##### **PCR 3**

Forward: 5'-ATTTGCATCCTGAAAGGCTCTCGC

Reverse: 5'-TCGAGCAAGTATCGCCTAATG

SUPPLEMENTAL FIGURE S1. **Sequence of Glt1 cDNA.** Codon optimized synthetic nt sequences of *Toxoplasma* Glt1 cDNA is in black lower case. The translated amino acid sequence is below in black upper case, and matches TGGT1\_205060. The N-terminal His<sub>6</sub>tag contributed by pET15b(TEV) is in purple. Numbering refers to the sequence of the native cDNA and protein. Peptides from Glt1 detected in strains GT1 and ME49 as recorded at ToxoDB are in red, with the first amino acid underlined. A detected phosphosite is asterisked in red.

```

Atgggcagcagccatcatcatcatcacagcagcggcagagaaaacttgatttccagggccat
M G S S H H H H H S S G R E N L Y F Q G H pET15b (TEV)
atgcgtagaaagcctgggtgactgtagcgtggattgtctgtcctcctgcaaaaacatggt 60
M R E S L G D C S V D C L S S C E K H G 20
aacatgcgttccctggatgccgtgggtcgtctggcggaaaccacgtatcgtcaaaaccgc 120
N M R S L D A R G R L A E T T Y R Q N R 40
attgattttccggacttcccgtcccatcagtcattcagtttaacctgtggagctctgaa 180
I D F P D F P S H S V I Q F N L W S S E 60
gataatacctggctgagctctgctgctggcctgtggcaccgctgtaatgcacgtcgtgtg 240
D N T W L S L L R G L W H R C N A R R V 80 peptide 1
ctgctggcagcatccgggtgatccggtgaaccggcacatctgtctgaagaagttaccagt 300
L L A A S G D P R E P A H L S E E V T S 100
aaacgtcacctctgcccgggtgtcgcaacgccgctctgaactgtgattctgaactgctg 360
K R H R L P G V A T P R L N C D S E L L 120
aaatgctataaaaatggcgggtgatctgaaaggcgtggaccatcactgtccgcaaccggca 420
K C Y K N G G D L K G V D H H C P Q P A 140
ccggctgatagccgtgggtgatcgtgacgggtgaagcattctgcgctgaccgtgcgctgtcg 480
P A D S R G D R D G E A F C A D R A L S 160 peptide 2
gcaagcctggcacaggcaattcgtgaagccaaactgtactgtggtcatccggaacgtacc 540
A S L A Q A I R E A K L Y C G H P E R T 180 peptide 3
gtggttgattgccgcagttcctgctgtctggaagaaaccgggtcgtcatacgccgccggaa 600
V V D C R S S C C L E E T G R H T P P E 200
ctgccgctgggttgatgggtgggtcactcatcggcaggttgtaaaaacttcagctctgctttc 660
L P L V D G G H S S A G C K N F S S A F 220
atgaaaattccgtctgaattttgtcagtggaagcgatcccgccgctgctgcatttcgtg 720
M K I P S E F V S G K A I P P L L H F V 240
tggtcgggtgggtcacccgccgcccgtttttcgaaccattcgccaaagctgggcccgttcac 780
W L G G H P P P F F E T I R Q S W A V H 260
aatccggatctgatccaggcactgtggaccgacgctcacgtggaaagcctgctggatgtt 840
N P D L I Q A L W T D A H V E S L L D V 280
ctggaccgtaaatctcgagtcgtcgcccgaaatgccgtaaaaccgatcatcaggacctg 900
L D R K S R S R R P K C R K T D H Q D L 300
caccgctgctgggtcgatcgtccggcaacggacgctgggtgaaaccacggaatgggaaatg 960
H P L L V D R P A T D A G E T T E W E M 320
tccgattcagtcgccgacagccagaccctgggtgcatgcatgaaagccttttcgtaaagaa 1020
S D S V P D S Q T L V H A I K A F R K E 340
tcctgtccggggcgcaaatcagatattgcccgctgctgatcctgtgccactatggcggt 1080
S C P G A K S D I A R L L I L C H Y G G 360

```



```

atttacgcggatgtggacatggaagccatcaaaccgctgccgccgtgtctgcgtcaatgc 1140
I Y A D V D M E A I K P L P P C L R Q C 380
accacggtgtttatgggcatgcagcgcccgatgcagttgaactgggtaacgctctgatc 1200
T T V F M G M Q R P D A V E L G N A L I 400
ggctgcagttccggtcatgaactgattcgtttcattcctgcagcggtgttgccgcccgatat 1260
G C S S G H E L I R F I L Q R V G R P Y 420
tcgcaatggggtaccgtagcgcagatcagatggcagtcgtgctggacattctgaaactg 1320
S Q W G T R S A D Q M A V V L D I L K L 440 peptide 4
cacgtggcgcggtttctccccggaagcctcagaactgatggattcgctgcgtagcaccaac 1380
H V A R F S P E A S E L M D S L R S T N 460
gacgaagcgacgaatgttatcgaaactaccggcccggtctgctgaccgctgcaacgctg 1440
D E A T N V I E R T G P G L L T R A T L 480
gcttggtgctgcgcgatcaactgaactcatcgagctgtgcacgttgccgtccgcaggaacgt 1500
A W L R D Q L N S S S C A R C R P Q E R 500
cgttcccatgattacgaacacaaatcaaataccacggaagcaacctcggtgcatgctgac 1560
R S H D Y E H K S N T T E A T S V H A D 520
gttgctcggtcggaacgggttagcaaagaagatgactctaaaccggaagcaaccagtggc 1620
V V G S E T G S K E D D S K P E A T S G 540
acgtgctgttgcatctctggcgaccacgtgtatttgcccgcgatctttttctatccggtt 1680
T C C C I L A T T C I C P P I F F Y P V 560
ccgaatcatcgctcgaaagaactgcgcgaaggcaaagtccagaccgaacacctggaatct 1740
P N H R R K E L R E G K V Q T E H L E S 580
agtttttcttacacggtccatcactggcgctcagacctggcaagacagcggtccgtcaagaa 1800
S F S Y T V H H W R Q T W Q D S V R Q E 600 peptide 5
agcgaaaatgaatgttaa 1818
S* E N E C - 605 *phosphosite 1

```

W64

*Hs1* --TPPSHGPTPGN**IFFLE**TS**DRT**-----NP**NFL**FM**CS**VE**SA**ART-  
*Hs2* -----SHRR**G**IV**FL**ET**S**ERM-----EP**PH**LV**SC**SV**ES**AA**KI**-  
*Dm* LQAD**PK**PS**PG**NS**IFF**H**ET**SC**RL**SE**NR**Q**LE**TL**KV**TARQ**CA**IESA**AM**H-  
*C1* -IAGGS**PP**SP**GD**V**FF**VE**T**SE**RI**-----K**PS**Y**LF**T**CS**VE**SA**ARA-  
*Af* -----  
*Mc* -----  
*Cj* -----  
*Sc* QAP-----  
*Tg* LLAAS-----GDP**RE**PA**HL**SE**EV**TSKR**HR**LP**GV**AT**PR**LN**CD**SE**LL**K**CY**KN**GG**DL**KG**V**DH**HC**PQ**AP**AD**SR**GD**R**DG**E**AF**CA**DR**ALS-  
*Hh* LLAAS-----GDP**RE**QA**HL**S**VE**AT**STR**H**ML**P**GV**AT**PR**LN**CD**SE**LL**M**WY**EN**GG**DL**KG**V**DH**HR**PQ**Q**AP**AD**SR**GD**R**GG**E**AF**CA**DR**QL**S-  
*Nc* LAAAS**KA**AT**SK**AAAS**KAA**AC**GD**PP**HA**Q**LS**V**N**AT**MK**SR**MS**IA**AT**P**WF**RY**DT**D**FP**ACH**NG**GGG-M**KG**-E**H**QRRRR**AT**AD**VR**DD**S**ARE**AV**CA**DR**ALN-  
*Lar* -----DALL**GH**VRT**AG**F**VRT**TH**SL**EH**GS**APG-----LP**VV**-----QE**HP**-A**H**Q**R**DN**R**MQ**T**AD**AD**AL**YL**R**PL**M**RA**IASA-  
*Rf* R**F**DE**Q**F**K**SR**V**DD**D**K**H**VM**N**T**IN**H**N**DD**DD**DD**D**G**K**SW**D**H-  
*Bn* R**VL**IE**K**DD**D**RT**M**EST**K**R**M**I-----  
*Ot* D**Q**F**K**N**I**V**ID**H**L**SS**Q**SR-----  
*Ng* -----  
*To* NIGAD**GS**Y**L**AA**ET**K**V**K**TD**Q**Q**G**AM**M**V**G**TA**-----  
*Ac* MF**P**T**V**D**NA**VR-----  
*Pu* P**T**GER**GS**SE-----  
*Aa* G**S**AN-----  
*Es* P**V**RED**G**EAL-----  
*Ap1* A**PP**SP**G**PP**V**-----  
*Gt* RGT**G**N**K**SE**G**CEER-----  
*Eh* PQT**PN**GR**HY**GRE**T**Q**SK**Q**DT**SV**AG**G**Q**AS**QC**PP**GL**AF**VS**D**V**SP**D**SS**RV**NE**PP**SR-----  
*Pp* -----LTT**T**VT**TE**RS**Q**ST**DY**DT**AE**V**K**DS**ST**D**N**VM-----  
*Ap2* -----P**PS**L**IR**AI**P**DD**Q**AM**H**D**VR**GF**GL**AT**K**SS**P**TS**NP**LS**IV**D**GV**KAS**RT**N**D**KE**G**PK**P**FL**D**GD**ND**GL**HS**V**V**AP**DT**T**NT**P**Q**PP**RD**SP**ST**PA**QR**-----  
*Sp* -----TL**EP**IS**FT**MOAR**ASS**LTT**P**APP**ASS**R**N**AS**ST**PI**IE**SV-----

		P234		W241
		*		*
<i>Hs1</i>	-----	YPEWVPVF	-----	FM
<i>Hs2</i>	-----	NPNFQVFV	-----	LF
<i>Dm</i>	-----	HPGTRVVV	-----	LM
<i>Cl</i>	-----	IPKIIHQT	-----	YK
<i>Af</i>	-----	IPKEKIIK	-----	HK
<i>Mc</i>	-----	IPFEEYFS	-----	DD
<i>Cj</i>	-----	IPNIIHQI	-----	WI
<i>Sc</i>	-----	LRDQLSFAFPYDSQAPI	-----	WK
<i>Tg</i>	ASLAQAIREAKLYCGHPERTVVDCRSSCCLEETGRHTPPLEPLVDGGHSSAGCKNFSSAFMKIPSEFVSGKAI	PPLLHFV	-----	WL
<i>Hh</i>	ASLAKAINEAKLYCGHPQRTVVDCRCPCSPREETGGHTPPDLPVVDGGHSSSTGCKNFSSPFTQIPSEFVCGKAI	PPLLHFV	-----	WL
<i>Nc</i>	ASLAGAIREAQLHCRHPERTAEDSGSARSFETSQGLTLPVVPVTPAGPSVTGDEHLLSPVVQSPHAWDRGQAI	PPLLHFV	-----	WL
<i>Sn</i>	ASLECAGRERNTOATVCALSGRPTAAEEHPE-----PRADSGAFNTTENFLLKSLRNTSWH-----	IPLLHFV	-----	WC
<i>Rf</i>	-----	IPKRIHQI	-----	WL
<i>Bn</i>	-----	IPKIIHQI	-----	WF
<i>Ot</i>	-----	KSIPKIIHFI	-----	WL
<i>Ng</i>	-----	LPKKKHFFYAKLLLSKISTEAWK	-----	WK
<i>To</i>	-----	IPKILHFI	-----	WL
<i>Ac</i>	-----	IPFIIHQI	-----	WL
<i>Pu</i>	-----	IPHVIHQI	-----	WL
<i>Aa</i>	-----	VPRKLHQI	-----	WL
<i>Es</i>	-----	IPKIIHHI	-----	WL
<i>Ap1</i>	-----	IPKRIHQI	-----	WL
<i>Gt</i>	-----	IPRIIHHI	-----	WI
<i>Eh</i>	-----	IPKIIHQT	-----	GR
<i>Pp</i>	-----	IPRLIHQS	-----	WK
<i>Ap2</i>	-----	IPKIIHQS	-----	WK
<i>Sp</i>	-----	IPRLIHQS	-----	WK

	G243	
	*	
<i>Hs1</i>	KGLT--GGNASLPRHLGTSLLSCTFPNVQML	-----
<i>Hs2</i>	AGPTDSTMPNSNTYPAFSFLSAIDNVFLF	-----
<i>Dm</i>	KGLAYRISNNKSHPPQLLEAIIYSNVHLR	-----
<i>Cl</i>	NETI--KGNVSLPSHWAFSLLSREFPNVEIQ	-----
<i>Af</i>	ATGS--	-----
<i>Mc</i>	RGAG--	-----
<i>Cj</i>	QGYE--	-----
<i>Sc</i>	VGADDKNFFSSF	RTYQKTWSGSYSP-----
<i>Tg</i>	GGH---PPP-FF	ETIRQSWAVHNPDLIQALWTDHAVESLLDVLDKRSRRRPKCRKTDHQLHPLLVDPRPATDAGET
<i>Hh</i>	GGH---PPP-FF	EKIKQSWAIHNPDLIQALWTDHAVESLLDVLDKRSRSTPEARKADHQLHPLSVDRPETDAGET
<i>Nc</i>	GGHR---PPS-FF	DEVQQSWAVHNPDLIQALWTDHAVESLLDVLDKRVRSWMGATREASRQ---GLSGDRPETDASET
<i>Sn</i>	GKE---PLDFFF	LOFKQSW-KKNSLLIHALWRDAEIKGLVSLVKGNDYCTQRPDRCHAV-PVP-SGDKK-----
<i>Rf</i>	GDK---AIPECY	LTWQOEWKQKHOG-----
<i>Bn</i>	GSN---PLPENF	ETRFQETWKGTHT-----
<i>Ot</i>	GEK---EKPEYF	KVHVYGSWTSKDTYSEII-----
<i>Ng</i>	GRH---ISNQIL	LYRESFTTYGLAPRCPSDFKLYAQHGFDTQRKTVGIKGDVPSVLDLMSFFSFPO-----
<i>To</i>	GGN---PLPR-F	TSLPEGEDHGDPLPGNLGSNACIESWRKHHT-----
<i>Ac</i>	GPH---PIP-AF	CLQQMETWRQIHP-----
<i>Pu</i>	GPH---PIP-VD	CLAWMQTWKRLHP-----
<i>Aa</i>	GPK---PPDDAH	AAAWRALHP-----
<i>Es</i>	GS---PLPEAF	ARLRSWLARHSGLSSQGRGPERSNNGGST-----
<i>Ap1</i>	GPA---PMDEFH	SYIQSWKNHHP-----
<i>Gt</i>	GGK---LPGKF	QSLRDEWILLHPAQE-----
<i>Eh</i>	SRC---VSAKF	HEVMRAWMDRFP-----
<i>Pp</i>	SVY---RIPTRF	HPWMKSWVEFHP-----
<i>Ap2</i>	SAD---HIPSKF	VQWMQSWRTHHP-----
<i>Sp</i>	SAD---AIPEIF	APWMRSWVQHHP-----

	W318		D348	R351	D363	D365
	*		*	*	*	*
Hs1	PLDLRELEFRDT	PLADWY	AAVQGRWEPYLL	PVLSASRIALMWKFGG	IYLDTFDIVL	
Hs2	PLDMKRLLED	PLFSWY	NQINASAERNWL	HTSSDASRLAIWKYGG	IYMDTDVLSI	
Dm	RLNLESYASGT	PMEEWL	KDGRLSRSKYLE	SHISDFLRYLTLYRYGG	LYLDMDVVVL	
C1	PLDLAELEFGT	PLAKWY	SQPEHQKPYFF	PVLSDACRITIMWKFGG	IYLDTFDIVL	
Af	PEVWREAQQSCKDLHP	DYEYILWTNEKSREFIKNEY	PWFLDT	PIQRADTIRYFVLAHFGG	TYIDLDDGCN	
Mc				YALFSDIFRYELLNQT	DHGIYVDCDVYCL	
Cj				VAAFSDFFRFNLLYL	RGGVWVDLDMVCL	
Sc	DYOYSLISDSSIIFLENLYA	PVPIV	IQAFKLM	PGNILKADFLRYLLLFARGG	IYSDMDTMLL	
Tg	T-----EWEM-----SDSVPSQ	TLVHAI	KAFRKE	SCPGAASDIARLLILCHYGG	IYADVDMETAI	
Hh	T-----EWEM-----FDSAPYRQ	TLSYAI	KAFRKE	SCPGAASDIARLLILCHYGG	IYADVDMETAI	
Nc	I-----EWEK-----TDGFSYMQ	TLAHGI	KTFRKE	SCLAARSDIARLLILCHYGG	IYADADMETAI	
Sn	-----GWRQ-----RFTHGDK	VEVSRILTELA	HMVEE	ERLGAASDLVRLLLLHTYGG	VYADVDMELV	
Rf	NWEYFWDDEHVKLLFS	GELTLP	LLSSCSKDNDDGKTMLNAL	QQMWPICDNFGEQSDLLRLFLLYQFGG	VYIDMDDFECI	
Bn	HWRYILWRDGDIEEH	PLILRH	CKHLLDDA	QSSVEKSDIWRRLAVLYEMGG	VYADVDFECV	
Ot	-----TWGEKDISELDLINE	-----TIEDKT		LNPAFRADALRLIILYQEGG	AYLDTDMSGI	
Ng	-----GWTTRLWTDADVDSI	-----CLRNO		PNYQKSDILRYELLERHGG	VYVDVDMCEV	
To	GWRFQIWTEADVIMEGCKSSQTS	EAERL	EMHASQISNL	SAYSYA-LKIGNYGLASDVLRLEILSIFGG	VYVDIDYLCI	
Ac	QWEYKLWTDQTQVSTL	-----KLQNK	-----EHFDLA	GNYEKSDILRYELLQFGG	IYVDVDFECL	
Pu	AWEYKAFDTA	-----	-----	TNFEKSDILRYELLKYGG	VYADVDAVACV	
Aa	DWEYKLWRDADVAAL	-----GLENA	-----AAFSA	TNWGEASDIARYELLRFGG	VYADMDFEPL	
Es	PWEVRLWTDADVDAF	-----GLENR	-----GAYDA	QNFQKSDILRYELLRHGG	LYVDVDFECL	
Ap1	DWEYTLWTEREINW	-----KLQNK	-----AAYDMA	TNFEKSDILRYELLTFGG	LYVDVDFECL	
Gt	GWKHLYLW-DDESIAQEFSSN	-----PMDSA	-----GSYASA	SNYEKSDILRLVLRNRFGG	VYVDVDFKCI	
Eh	GWAYRFHDDAAMES	-----LLARRHA	-----FPLHLHAVVQCVHSM	TMKADMWRYLLALWRYGG	IYADMDVKPF	
Pp	TWTYVFWTDADNLR	-----LFELL	-----YPOYLHVAKAVRKVSL	ADMARYALLHQVGG	LYVDADFECV	
Ap2	TWTYVFWDDSDNLD	-----LFEEH	-----YSKYAPVARRVSKIQL	ADMSRYALLHRYGG	LYVDGDFEAL	
Sp	TWTYVFWDDAANLA	-----LFARH	-----YPOYAVASSVGKIHL	ADMTRYALLHRRFGG	VYVDCEFEET	

	G385	E393	N396	H405
	*	*	*	*
Hs1	KNLRNLT	NVLGTQS	RYVLNGAFLAFERR	HEFMAICMR
Hs2	RPIPEE	NFLAAQA	SRYSSNGIFGFLPH	HPFLWECME
Dm	RNMEKVP	PNYTGAE	NTHLAAGVMNLAATFG	HEIAASCLR
C1	KNLRNLT	NALGLQS	QDVNLGAFLSFKPK	HEFMELCIQ
Af	RRLDPLL	SYPAWRR	TVPTGISNDAMGSVPO	HPFFLRVIELLSYDRS
Mc	KPITMPN	HGYLLGY	EDDTKINGAILALPKES	ELLHALLKSAYDPYFVPPWYK
Cj	NHYDYDK	KEYIFSK	EIDNDLSKARITTSLLK	FPKQSEFGKLIIDEAKKI
Sc	KPIDSWPSQNKSWLNNIIDLNKPIPYKNSKPSLLSSDEISHQPG	LVIGIEADPDRDDWSEWYARRIQFCQWTIQAKPGH	FILRELILNITATTLAS	
Tg	KPLPPCL	RQCTVFMGM	QRPDAVELCNALICSSG	HELIRFILQVRGRPYSQW
Hh	KPLPPCL	RHCTVFMGM	QRPDAVELCNALICSSG	HELIRFILQVRGRPYSQW
Nc	RPLPPCL	RRCATVFMGM	QRPDAVELCNALICSSG	HALIRFILRHVGRPFSSK
Rf	RKLPPCF	LTDAFVAGA	QREDAVELCNALLACTPR	HELLTYIITQIVHRSSM
Sn	LPLDVLF	SNTSLFEVNNAFMASKPF	HPFIKHLWDNLDMDKDPQE	
Bn	RNLTLGLH	QSCSFYAGL	SNTCTVEINNGIFCAA	EEKHPLCYTLLKNMQTTDQFN
Ot	YSLNDLL	DYPTDFIIGL	SNTKAFELNNAFIASCPG	HPLLKHLMETLKLNYQNH
Ng	RPLDDLH	GQRGPSFYAGF	SNTCTVELNNGIIGSIPG	HPILRQLIDRIKREQSKP
To	SPLDDLH	SPSRLPLHFFCGA	SNAGCVELNNGIMACKEGG	HQILSNMMSRIHHYFERR
Ac	RLFQDLL	QAFSFIITGI	SNTDVVELNNGLIAC	TRNHPIVRELVASLASKAPVM
Pu	QAFDPLL	KAFSFIAGM	ANTGNVEISNSVMLSTAH	HAILRQLIDTIHRDFHQP
Aa	RPLDALR	EAADFFVGF	SNVGAVEINNGLIGA	APGHALLAALVERVAARVQR
Es	GSFDDLH	KRYEFYAGV	SNTGTVELNNGLICRPG	HPIMRDI
Ap1	KPFQDLH	ENPTCFEYAGL	SNTRSVEINNALIGCVPN	HPILKAVISAIQRETKRR
Gt	RSFHDIL	GL	SNVGHIEVNNGLIGSAPO	HPLLKVLPLWLHVPSTSM
Eh	RTLNSAS	ISLSDSALFF	IDASVRSLTQYFMAATAQ	SPLMESALKAAACHVLDS
Pp	QPFDDLH	RENNFLSSEPLVH	SVLL-EKSDSAAALCNALMASAPR	HPFWLSVLDNIKKKFD-R
Ap2	QPMDDLH	DLPLFLSFEPVH	SVLL-EGASAPVLCNAILAS	MAGHPFWLEVLDNILDAFNSP
Sp	SPLDDLH	DKDLFLSTEPFVH	SVLL-ERATDAILCNAVLASRRG	HPFWLQVLDAILAKFQ-A



G470 P471

\*\*

```

Hs1 -----DFVDHYNGWIWGHQGPQLTRVFKKWCSIRSLAESRACRGVTTL-----
Hs2 -----NFVEHYNSAIWGNQGPPELMTR-FQEVSDLRCLNISFL-----
Dm -----DFQHNFGGWDGNNNGPGVITRVAQKICGTKDIALMREDPKRCMGFKVF-----
Cl -----DFVDNYNGWIWAHQGPPELLTRVFKKLCISINIQNGMICKGVSALE-----
Af -----WLLPYITVMYSTGPPLFLSVIWKKEYMQDKPSEAAVRILMQDEYN-----
Mc SRKQKILKIRKQF-----GFSRHVANMPWGVIGPDAITYYAQHYQIAHFAQ-----
Cj -----VDNKKIIPWGIIGPWFLAKWVKEYDLEKHALDYKDTQCISC-----
Sc VQNPQGVPSVSEMIIDPRFEEDYNVNYRHKRRHDETYKHSELKNNKNVDGSDIMNWTGPGIFSDIIFEYMNVLRYNSDILLINPNLNKNDEEGSESAT
Tg GTRSADQMAVVDLILKLHVARFSPASELMDSL-----RSTNDEATNVIERTGPGLLTRATLAWLRDQLNSSSCARCRPQERRSHDYEHKSNTT
Hh ETRDADQMAVVDLILKLQVARFSPASELMDSL-----RSANDEATNVIERTGPGLLTRATLAWLRDQLNSSSCVRCRPQEGRSHDYEHKSNIK
Nc GTHDVDQTAVALIILKLQVARFSPASDILLASL-----PGPDDEADNAIERTGPGLLTRATLLWLRDQLNSSSCSRCQQERCCHNYEHKNVNI
Sn DEDDLMQVFRIVECYTGESS-LRRQGKHTS-PL-----VRTNREAMQVISSTGPGLLTKAVMEWLQTRAHSRG-----EPNYARL-----
Rf QG-----QRHLVSTIWKTPGVVYSIQLWKHWDRWKQ-ST-----DILV
Bn NSSSSN-----LTTGVDTMKIARTGPGHFTRVVMNYILFPSHESYCTNEANVQQORIKDDSSSD
Ot LKHVADKNKINQLLSQQQGIQLEDIP-----YQKLNIIAVSGPGFMTQQIFKYLNNDKQDESG-----KHILI
Ng LSARHHPTRAQGSSSLATILSFLGRSDAVAAEHAVAR-----AARELEAFVTIEATGPGLTRGVAALVLQGGQVDEV-----GIVLL
To SERHAAMEKSQIATLVNSFLDEEIQVD-----GPSPIEVIEHSGPGLLTRELCRWLVSSEGTPEPPG-----LSKDRNRVLV
Ac SMTDARILDITIAQFTTEKLVIRQ-----SNEFMQTISQTGPGLLTRTFMRAIGW-----KSGSK--CIPGFLTSL-ESKDVIA
Pu RLDTSALVLIAQMSGDATLSAALSITP-----TSSAMDTIARTGPGLLTRTFMAAIGWSTEEGGGREAGADGFLSPAERERAI
Aa PKPPPTDLLAAIAASGFLDDAGPLAKTL-----SQADADWTIEHTGPGLLTRTFCELHG-----PMPRAVC
Es -----linkage is unknown-----??SATEIVVKTGPGVFTRAVMAWVSSATGEDGAAPTTTDSGSGTISTAQEEIF
Ap1 QLIADYAGV-----PVNDMPKLGTIARTGPAVFTVQIMHSIGWEGHRLPI-----QNAATVVI
Gt PSDRLLCALSSGNETICKTGT-----K--FVNSFKLDROTGPGYFTRMMISAWKEKLHKPHG-----VDDLEWIV
Eh RFPGG-----VNRFMYIARTGPGLKRGFTFMESNVSMNIT-----YGTSTPGLYI
Pp ER-----LKSDAVELTGPRMVKQTYLSPNSTFNLK-----GSDMVV
Ap2 GG-----ARQDPVSLTGPRMVQHTMSQHAKAASAA-----TTTGGIIL
Sp GG-----AKQDPVSLTGPRLVLEA-VYHEFDAERV-----GASVTV

```

P552      N561  
\*                   \*

```

Hs1 -----PPEAFYPIPWQDWKKYFEDIN-----
Hs2 -----HMLRVWCKLEDPPQRFYPISYREWRRYYEVWD-----
Dm -----GRGAFYAVPWKQWRDFFEPEN-----
Cl -----PPDALYPIPWQDWKKLFEAIS-----
Af -----KYSWS-----
Mc -----SIDVFYPIHHHCIGHLLN-----
Cj -----GNTRDFIDKKIFDKNRCLHLFSEM-----
Sc TPAKDVDNDTLKSTRKFYKKISLSQSSNSMPWE-----FFSFLKEPVIIVDDVMVLPTSFSFSPDV-----
Tg EATSVHADVVGSETGSKEDDSKPEATSG-----TCCCLATTCICPPIFFYPVPHNRRKELRE-----
Hh KATSVHADVVGSETGSKEDDSKPEATIG-----TCCCLATTCICPPIFFYPVPHNRRKELRE-----
Nc RAASGQAEVVGIEGTGSKEDDSEPDANW-----ACCALATTCICPPIFFYPVPHNRRKDLLE-----
Sn -----EGASAPDWLRG-----LESVICICPCVFYPVPHNRRKELR-----
Rf L-----PISVLYPFPNMRH-----
Bn GAEKEEDVTSSSVDEQLLVVL-----PCGYFYPFPNSYKDLL-----
Ot T-----PKQFFYPPLSNVREKLTALNFQQLVDE-----
Ng A-----PPVWFYPIPNRPVQAQIVNPRDSIELPI-----
To Y-----PAAVFHPPFNRLRLT-----
Ac L-----PRDYFYSIPNHLSSHVA-----
Pu L-----PIEYFSPLPNRIH-----
Aa L-----PYEIFYAKSNAAAA-----
Es HSTGTAALDTDVFGGGGGGDSGRGGGGGSATFGASSRCSNRRRADDNTTVTAIAAAAAAATPPAMILPSPSYLYPVPNAAAAKAGLLIGDGA-
Ap1 L-----PYKALYPLCNDDESG-----
Gt L-----PVRYFYSLPNHGDRED-----
Eh -----GQGGHVRVAGDHNRMIVRDA--LGRA
Pp F-----TSEYFYPEVAYWNIPEPMKTAC--RQRH
Ap2 L-----DEEYFYPEVAYWNLNLSRKCTHVKSH
Sp F-----PEEYFYPEIAYWNLGRLLLELC--RDRF

```

H586      W592  
                 \*      \*

```

Hs1 -----PEELPRLLSATYAVHVNKKSQGTRFEATSRAALLA-QLHARYCPTTHEAMKMYL*
Hs2 -----TE-PS-FNVSYALHLWNHMQEGRAVIRGSNTLVENLYRKHCPRITYRDLIKGPEGSVTGELGPGNK*
Dm -----LEETIARCKDSYVVHVNKHSKLPKIGSKNAYA-LYAEQNCPRSYKAAGEYF*
Cl -----SSELHNLKNTYAVHVNKLSHDARLEITSQALLA-QLYSQFCPATSAQMKKDFEEQSRPVM*
Af -----FFTTHVGNSTWHGKDARLIFWMGQHWMLTVLGFILASVVGFCLEWVWYGRMILLSSKYRRY [ 50 ] *
Mc -----SRLSIKDIITPDTLCIHLVQERLRHLDLSCLPQNCILDKIINSH*
Cj -----WKIYKMNKNHFKSCYIGFLLQKHNLDLCLKLNYNLSFCDKHYDKFLPFINIKNKIRFYFRHPKKIF [ 5 ] *
Sc -----GOMGAQSSDDKMAFVKHMFSGSWKEDADKNAGHK*
Tg -----GKVQTEHLESSFSYTVHHWRQTWQDSVRQSENEC*
Hh -----GKVQTERLESSFSYTVHHWRQTWQDSVRHESETEC*
Nc -----GKVQAELDSSFSYTVHHWRQTWQVSRPRGSETRC*
Sn -----KRGDEYIKTRFSFTVHHWKETWKDLPSCGT*
Rf -----PYTDRIAYVQKSTVAIHVHWGVSWIDAKVHRHNEGKPKHQISASINTDTNANTNANVKINKKEQPTTEQKQDNRSN [ 32 ] *
Bn -----KLEDRMHFHRSETLAVHHWGCSWOHCNDGCNNNHDD*
Ot QFASLEGKVMNSKDGVLPLILKHADLNTIVYGCHLWEASWQE*
Ng -----PSPPAAGAVPEVAYAVHWWARSWQTRG*
To -----NKCKFEFIERDTIAVHLWGSSWQK*
Ac -----QEKLDKFIPEPEAMAVHFWAKTWOKNILKK*
Pu -----AADFAMIKLPVNCMAVHYWARSWM*
Aa -----ASDASSPFSYAVHWWAKSWOERPR*
Es -----VREVRGGYLSSESLAAHLWGRSWQQQQQQQQQLVKE*
Ap1 -----QEAIQARTQPGAFIHWYASWDRKPIEWDARSKSLDISVPEIISRIDRTTKQS*
Gt -----NWQTIDLKTVQLNHHLHMTAHRRTITLMSMLFICGHTLGRATPAPRTKSMSTNKK*
Eh DPLGRKACFEK-----KKGEYTS-MNMTTFSGFNQHNGLNATNRSCMSEERSCAA*
Pp DDAAKEACAW-----LQRFPKGEFTNKTHATHHWQCTWCRDAQLNEFGALRDIF-ESPPMRPNITMTGIEFVVLK*
Ap2 PPIVQEAC-----LNEYPTGRYTNKTHAVHHWQCTWCRGDDTTSYVSLHDIFPNQDTRRPHQMHPVPRTPVRI*
Sp DPLGRKACAW-----LRDHPKGFYTPRTHAVHHWQCTWCRGDVGEAVTTLSAIFPTATIHRLRGDDDDGLF

```

#### Sequence IDs:

**Hs1** *Homo sapiens* (BAA95915.1): Hsa4galt, lactosylceramide  $\alpha$ - galactosyltransferase  
**Hs2** *Homo sapiens* (AAD48406.1): Hsa4gnt,  $\alpha$ 4-N-acetylglucosaminyltransferase  
**Dm** *Drosophila melanogaster* (AAF51162.1): Dma4gt1:  $\alpha$ 4GT1  
**Cl** *Columba livia* (ADC84388.1): collagen GT,  $\alpha$ 4-galactosyltransferase  
**Af** *Aspergillus fumigatus* Af293 (EAL85572.1): Mita, MIPC synthase subunit (SurA),  $\alpha$ -mannosyltransferase  
**Mc** *Moraxella catarrhalis* 2951 (AAZ29048.1): Lgt5,  $\alpha$ 1,4-galactosyltransferase  
**Cj** *Campylobacter jejuni* (AAM90647.1): Cgtd,  $\alpha$ 1,4-galactosyltransferase  
**Sc** *Saccharomyces cerevisiae* (P31755): Och1,  $\alpha$ 1,6-mannosyltransferase  
**Tg** *Toxoplasma gondii* ([EPR61400.1](#))  
**Hh** *Hammondia hammondi* ([XP\\_008887433.1](#))  
**Nc** *Neospora caninum* ([CEL66241.1](#))  
**Sn** *Sarcocystis neurona* (EupathDB:SRCN\_1642)  
**Rf** *Reticulomyxa filosa* (X6P0J2)  
**Bn** *Bigelowiella natans* (JGI:aug1.98\_g20090)  
**Ot** *Oxytricha trifallax* (J9J850)  
**Ng** *Nannochloropsis gaditana* (W7UA89)  
**To** *Thalassiosira oceanica* ([EJK66288.1](#))  
**Ac** *Albugo candida* ([CCI39677.1](#))  
**Pu** *Pythium ultimum* (K3WJQ7)  
**Aa** *Aureococcus anophagefferens*  
**Es** *Ectocarpus siliculosus* CBJ34172.1  
**Ap** *Aphanomyces astaci* ([XP\\_009836009.1](#))  
**Gt** *Guillardia theta* XP\_005836918.1  
**Eh** *Emiliania huxleyi* (R1FXS0)  
**Pp** *Phytophthora parasitica* (V9EJ41)  
**Ap** *Aphanomyces*  
**Sp** *Saprolegnia parasitica* (A0A067C728)