STRUCTURAL INSIGHTS INTO HOW GLYCOSYLATION CONTROLS SKP1 FUNCTION IN PROTISTS

by

HYUN WOO KIM

(Under the Direction of Christopher M West)

ABSTRACT

Skp1 is a homodimeric protein that recruits substrate receptor F-box proteins (FBP) to Cullin-1 as part of the **S**kp1/**C**ullin-1/**F**BP (SCF) complex. The SCF is a class of E3 ligases that conjugates substrates with ubiquitins and signals them for proteasomal degradation. E3 ligases therefore serve a role in controlling the proteome. In protists, the intrinsically disordered C-terminal domain of Skp1 undergoes an O_2-sensing process resulting in the formation of an O-linked pentasaccharide. First characterized using *Dictyostelium* Skp1, NMR-supervised molecular dynamics (MD) simulations indicated that the glycan alters the surrounding ensemble of local conformers to be more conducive to FBP binding, in concordance with Skp1 interactome studies in cell extracts. Analysis of the Skp1 glycan from the unrelated parasite *Toxoplasma* led to the finding that the terminal disaccharide is distinctive and assembled by two glycosyltransferases (GTs) unrelated to the one from *Dictyostelium*. X-ray crystallographic and modeling studies of the terminal GT ortholog from another species, *Pythium ultimum* Gat1, rationalized how the *Toxoplasma* GT was a UDP-Gal:α-glucoside α3-galactosyltransferase specific for Skp1, rather than a UDP-Glc:α-glucoside α4-glucosyltransferase in glycogen synthesis as previously annotated. Despite

the substitution of a glucose for a galactose in the fourth position, MD analysis indicated that both pentasaccharides were structurally compatible and alter Skp1 conformations in a similar manner. The convergent evolution represented by the GT switch supports the significance and specificity of the pentasaccharide structure. A previously described second function of the glycan, that it inhibits Skp1 homodimerization, was explored. Truncation studies of *Dictyostelium* Skp1 showed that the intrinsically disordered C-terminal domain is not required for dimerization. NMR analysis mapped the dimer interface onto FBP-binding subsite-1, which was confirmed by Rosetta-informed mutagenesis. The dimer configuration would place the C-terminal domains and their glycans within close proximity which suggests the potential for physical interactions. To address this possibility, NMR analysis of glycosylated Skp1 identified chemical shift differences of residues inferred to reside in the intrinsically disordered C-terminal domain. These findings reinforce dual roles for the Skp1 glycan in modulating O_2-sensing by regulating the availability of monomeric Skp1 for FBP recruitment and Skp1 conformations that are receptive to bind FBPs.

INDEX WORDS: Skp1, F-box protein, Cullin-1, SCF, Ubiquitin ligase, O_2 sensing, Protist, Glycosylation, NMR, Molecular dynamics, Protein-protein interaction, Protein-glycan interaction, Intrinsic disorder, *Dictyostelium*, *Toxoplasma.*

STRUCTURAL INSIGHTS INTO HOW GLYCOSYLATION CONTROLS SKP1 FUNCTION

by

HYUN WOO KIM

BS, University of Georgia, 2016

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2022

STRUCTURAL INSIGHTS INTO HOW GLYCOSYLATION CONTROLS SKP1 FUNCTION

by

HYUN WOO KIM

Major Professor:	Christopher M West
Committee:	Zachary A Wood
	James H Prestegard
	Robert S Haltiwanger

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2022

DEDICATION

To Mom and Dad

ACKNOWLEDGEMENTS

I would like to take this moment to deliver special thanks to those who supported my journey into the world of scientific research. My academic path was filled with amazing mentors, colleagues, and friends, and only with their encouragement and support I got through graduate school--relatively unscathed. First, I would like to express my deepest appreciation for my advisor Dr. Chris West for taking me into the lab especially at a moment in time where I was lost and did not know what to do with my life. He provided a fruitful environment to think and to conduct research, and his exceptional mentorship kept me sane through difficult experiments. I will cherish our discussions, and your brilliance, composure, and passion for solving problems will always serve as a model for my career as a scientist.

I would like to acknowledge my dissertation committee members, both past and present. Dr. Zac Wood, thank you for holding me to the highest standards since I held my first pipette as an undergraduate researcher in your lab. Your continued support and mentorship throughout the years really provided the fuel to keep the fire burning. Dr. Jim Prestegard, thank you for providing me with lab space to facilitate research. The world-class expertise in protein NMR in your lab made it possible to model and solve Skp1 problems. Dr. Bob Haltiwanger, thank you for demonstrating the highest level of professionalism, and keeping my thoughts about both graduation and post-graduation in check. Dr. Natarajan Kannan and Dr. Rob Woods, thank you for initially serving on my

committee when the project was fresh and at conception. Your perspectives led to insightful discussions and potential experiments.

I would like to acknowledge my collaborators. Dr. Alex Eletsky, thank you for your knowledge and expertise in structural NMR. I started working with you when I had absolute zero experience with NMR, so thank you for being patient and walking through the process step-by-step. Dr. Eva Strauch, thank you and your graduate student Karen Gonzalez for the contribution with computational work on Skp1 dimer interface. I'm glad our short chat at a student recruitment dinner culminated into a fruitful collaboration.

I would like to acknowledge the West lab members. First, Hanke, thank you for being the pillar of the lab as the manager and a research professional. You know the ins and outs of the West lab operation, and your technical support was second to none. To our research scientists Dr. Elisabet Gas-Pascual and Dr. Msano Mandalasi, thank you for working with a scrub graduate student like me. Your professionalism and grit set an example to follow, and I am excited to have collaborated in many projects with you. To my homie graduate students Drew, Bowen, Donovan, and Megna, I just want to say we were a dynamic and a diverse bunch of humans with an extraordinary sense of humor. Thank you for all the crazy lab jokes and shenanigans (and of course, scientific discussions).

Finally, I would like to express my top-shelf gratitude to my dear best friend and wife June, for your love and support through our academic journey together. It has been my greatest joy to share knowledge and "AHA" moments with you, and you challenge and test me every bit of the way. I think we've come a long way since general chemistry 101,

so the test of time has been on our side. What an honor getting through grad school together with my wife! Love you!

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

ABBREVIATIONS

AUC	Analytical ultracentrifugation
BTB	BR-C, ttk, bab
CD	Circular dichroism
CRL	Cullin-RING ligase
Cullin	Cul
Dd	*Dictyostelium discoideum*
FBP	F-box protein
Fuc	Fucose
Gal	Galactose
Glc	Glucose
GlcNAc	N-acetyl glucosamine
GT	Glycosyltransferase
HIF-α	Hypoxia inducible factor-α
HSQC	Heteronuclear single quantum coherence spectroscopy
MD	Molecular dynamics
NMR	Nuclear magnetic resonance
PHD	prolyl-4-hydroxylase domain
Rbx1	RING-box protein
SAXS	Small-angle X-ray scattering
SCF	Skp1-Cullin-1-F-box protein
RCI	Random coil index
SOCS	Suppressor of cytokine signaling

SR	Substrate receptor
Tg	*Toxoplasma gondii*
UPS	Ubiquitin-proteasome system
VHL	Von Hippel Lindau
Tg	*Toxoplasma gondii*

CHAPTER 1

LITERATURE REVIEW

1.1 Protein degradation: autophagy and ubiquitin-proteasome system

All living organisms experience growth, change and adaptation according to their environment that poses a challenge with a wide range of stress factors including starvation, hypoxia, and heat. Eukaryotic cells can respond to stress by regulating their proteome to control the levels of key proteins that are beneficial for survival.[1,2] Protein levels can be controlled globally at the transcriptional stage where cells, in response to stimuli or stress, can regulate gene expression. This determines the rate of protein synthesis which consequently shapes the proteome landscape by upregulating or downregulating expression of specific proteins.

Together with transcriptional control, the proteome is also regulated by intracellular protein degradation which allows turnover of proteins that are misfolded, damaged, aggregated or functional but simply no longer needed. This can be can be achieved by two major intracellular pathways - autophagy and the ubiquitin-proteasome system (UPS).[3,4] Autophagy is a recycling process that begins by sequestering cargos including proteins and whole organelles into vesicles.[3-6] These packaged vesicles are then transported and fused with late endosomes and lysosomes, eventually forming autolysosomes. In these new compartments, lysosomal hydrolases degrade the cargos into building blocks which are released back into cytosol for reuse. Major targets of autophagy include long-lived proteins, insoluble protein aggregates, and degenerated organelles. Because of the large

size of vesicles that can form, higher order macromolecular structures including multi-protein complexes can be efficiently degraded in bulk.

The other major process for protein turnover is ubiquitin-proteasome system, where specific proteins are tagged with poly-ubiquitin chains that are recognized as a signal for degradation by the 26S proteasome.[3,4,7,8] Ubiquitin is a highly conserved 8.5 kDa protein that is covalently attached to target proteins via the concerted action of E1, E2, and E3 subunits of ubiquitin ligase complex. E1 activates ubiquitin by linking a ubiquitin molecule onto itself via a thiolester linkage in an ATP dependent manner, and then the E1 transfers the ubiquitin to E2. E1 and E2 pair is the common conjugating machinery used for all ubiquitin ligases. E3 is a scaffold protein that binds both the substrate receptor (SR) and the charged E2 enzyme which places the two in close proximity. Subsequently, the E2 conjugates ubiquitin onto the target recruited by the SR. The substrate specificity is provided by the types of E3 ligases and the myriad of SRs that exists in an organism. Repeated action of ubiquitination results in a polyubiquitin chain conjugated on a target protein which then is recognized by the 26S proteasome for degradation.[9] The 26S proteasome in an ATP dependent manner unfolds and translocates target proteins into a barrel-like compartment where they are proteolytically degraded into short peptides and amino acids.

These two pathways cover a substantial range of targets and provide a mechanism for cells to modulate the intracellular proteins in a balanced manner. Autophagic engulfment is generally viewed as a non-selective process for regular maintenance of components too big to be degraded by the proteasome.[3,4] These include entire portions of cytoplasm and endoplasmic reticulum, ribosomes, large aggregated proteins and even

pathogenic intracellular invaders. More recent findings suggest selective targeting can also occur in a cross-talk fashion with UPS, as ubiquitination of proteins is a key step in the events leading up to their autophagic removal. Contrastingly, UPS degrades proteins in a precise manner with the specificities provided by the abundance of unique E3 ubiquitin ligases in a cell. They exist in dozens to hundreds in a given organism depending on the type of ligase complexes and the SRs that are expressed.[7,8]

1.2 Cullin-Ring E3 ligases (CRLs)

The E3 ligases consists of four families – HECT, RING-finger, U-box, and PHD-finger, and of these, the largest and most studied is the RING-finger ligase family.[7] Also known as Cullin-RING ligases (CRLs), this class of E3 complex is centered around a number of Cullins (Culs) whose scaffold-like role consists of bridging E2 to the substrate receptors ***(Figure 1.1)***. The structure of Cullin consists of elongated α-helix bundles with protein-protein interaction sites near the opposite ends of the termini.[10] RING-box protein (Rbx1) binds to the C-terminus and serves the critical role in recruiting E2 to the complex, and an adaptor recruiting SRs or a substrate itself binds to the N-terminus.[7,8,10] The plethora of SRs that encoded in a species represent the number of unique CRLs that can form within a cell.

Up to 7 Cullins that form CRLs have been identified, and each harbor a distinct set of adaptors and SRs that make up each class.[7,8,11] Diagrams of the most studied CRL1-5 are shown in Figure 1.1. Cul-1 binds the adaptor protein Skp1, which recruits F-box motif containing SRs called F-box proteins (FBPs)[7,8,11] ***(Figure 1.1A).*** Cul-2 and Cul-5 are similar in the way they utilize the heterodimer adaptor elongin C and elongin B, but differ

in the type of recruited SRs. Elongin C binds SRs that contain either a Von Hippel Lindau (VHL)-box or a suppressor of cytokines signaling (SOCS)-box, and each contain Cul-2 or Cul-5 binding subdomains which recognize and provides specificity for their respective Cullin partner[12-14] ***(Figures 1.1B, 1.1E)***. CRL-3 differs from the other classes and does not have an adaptor protein. Instead, CRL-3 SRs contain a BR-C, ttk and bab (BTB) domain that can directly dock onto the N-terminus of Cul-3 ***(Figure 1.1C)***.[15-16] CRL-4 utilizes DDB1 as an adaptor to bind to SRs that contain a WDXR domain.[17]

CRLs are subjected to global regulation by multiple dynamic processes ***(Figure 1.2)***. Cand1 has been shown to inhibit ubiquitin ligase activity by protein-protein interactions.[18] It wraps around the N-terminus of Cullin and renders its binding site inaccessible to interact with adaptor proteins (Skp1, Elongin C, BTB protein, etc.).[19] This action prevents the recruitment of SRs and substrates and ultimately the formation of a catalytic CRL complex. Neddylation of Cullin prevents Cand1 binding and promotes functional CRL complex assembly and ubiquitin ligase activity.[21] Nedd8 is conjugated near the C-terminal region of Cullin, and studies have shown that Nedd8 promotes conformational changes of Rbx1 and E2 in a manner that optimizes and enhances CRL catalytic activity.[22,23] The presence of Nedd8 on Cullin blocks Cand1 from binding and frees Cullin to interact and recruit adaptors/SRs.[21] Removal of Nedd8 is achieved by COP9 signalosome, and the regulatory cycle by which free Cullin is subjected to either inhibition by Cand-1 binding or promotion by neddylation is reestablished.[24]

1.3 The SCF ubiquitin ligase (CRL-1)

Of the RING-finger ligase family, the best characterized and the main focus of this dissertation is CRL-1, also known as the SCF (Skp1/Cullin-1/F-box protein) complex ***(Figure 1.3)***.[7,8,10] The key element in the SCF complex is the adaptor protein Skp1 whose major role is recruiting FBPs to Cul-1 ***(Figure 1.4)***. Skp1 is classified as a BTB protein along with Elongin C from CRL-2/5 and CRL-3 SRs and much like other CRL adaptors, docks at the N-terminus of Cul-1 ***(Figure 1.4A)***.[12,14,25] The N-terminal part of Skp1 binds to Cul-1 and the hydrophobic surface of the remaining helices 5 and 6 of the BTB domain make up the FBP-binding subsite 1 ***(Figure 1.4B)***.[10,25] The feature unique to Skp1 is the C-terminal extension following the BTB domain that contributes to FBP binding as the subsite-2 ***(Figure 1.4A)***.[10,25] This region while disordered and highly mobile in the free state, is ordered into a helix-loop-helix (Helix-7 and -8) when bound to an FBP as demonstrated by numerous crystal structures ***(Fig 1.4B)***.[7,10,26,27] The combination of FBP-binding subsite 1 and 2 acts as a molecular clamp and latches onto the F-box domain of FBPs.

The target recognition and specificity of the SCF complex is determined through the substrate binding regions of FBPs.[7,10,28,29] FBPs are composed of an F-box motif and a substrate recognition domain. An F-box consists of a 3-helix bundle which serves as interface to bind Skp1 and recruit an FBP to the Cul-1 scaffold ***(Fig 1.4B)***. The substrate recognition domain of FBPs exists in a variety of protein-protein binding motifs including leucine-rich repeats and WD40 repeats.[29,29] The specificity and the hydrophobic nature of these protein binding motifs allow recruitment of target proteins to the complex. On the contrary, some FBPs contain an enzymatic or a ligand binding domain instead of a substrate

recognition domain, in which case, FBPs themselves can be directly ubiquitinated for downstream degradation. The number of FBPs range from dozens to hundreds depending on the species, therefore those numbers can represent unique SCF complexes existing in an organism.

Besides the modes of CRL regulation describe above, there are several examples of regulatory mechanisms specific for the SCF complex. Characterized in yeasts, Met30 is an F-box protein involved in responding to heavy metal stress.[30] Upon cadmium exposure, SCF-Met30 is inactivated and its basal substrates including Met4 and Met34 are no longer ubiquitinated or degraded.[30-32] Inactivation of SCF-Met30 complex is achieved by the removal of Met30 from Skp1. Cdc48/Shp1 complex associates with Met30, and facilitates the dissociation of Met30 from the SCF using ATP hydrolysis.[30-32] Following, Met4 as a transcription factor induces gene expression associated with heavy metal detoxification, and Met32 inhibits cell cycle progression until the stress is mitigated. [30-32]

In a large-scale mass-spectrometry data analysis, Skp1 was found to be a phosphorylated protein.[33] Phosphorylation sites were detected at the Ser or Thr residue preceding the helix-7 and helix-8 of FBP subsite-2, which led to the hypothesis that phosphorylation may modulate Skp1/FBP interaction. A cell-based assay in yeasts using Skp1 with phosphomimetic substitution from Ser to Glu demonstrated less association with an FBP indicating that Skp1 phosphorylation may reduce binding.[33]

1.4 A novel mode of SCF regulation: O_2-dependent Skp1 glycosylation

West discovered a unique fucose containing glycan in the cytosol of the social amoeba *Dictyostelium discoideum* (*Dd*), which turned out to be Skp1 that is post-

translationally modified by a complex O-linked glycan ***(Figure 1.5)***.[34,35] Pro143 of Skp1 at the FBP-binding subsite-2 ***(Figure 1.4A)*** is modified by a prolyl-4-hydroxylase (P4H) named PhyA in an O_2-dependent manner, and the resulting 4-*trans*-hydroxyproline (Hyp) residue is subjected to a multi-step glycosylation event by set of glycosyltransferases (GT) specific for Skp1[36-39] ***(Figure 1.5A)***. A GlcNAc transferase, Gnt1, catalyzes the first sugar moiety onto the Hyp, which is followed by PgtA with its two GT domains transferring a Galactose (Gal) and a Fucose (Fuc) residue in a processive manner.[38,39] AgtA then extends the core trisaccharide into the final form with the addition of Gal-Gal terminal disaccharide resulting in a glycan with the following sequence and linkages on Skp1: Galα1-3Galα1-3Fucα1-2Galβ1-2GlcNAcα1-4(*trans*)Hyp143 ***(Figure 1.5C)***.[34,35,40] The terminal disaccharide can vary depending on the species as demonstrated by recent studies of Skp1 from the human pathogen *Toxoplasma gondii* (*Tg*), which revealed that rather than AgtA, a set of evolutionarily distinct GTs Glt1 and Gat1 catalyzes the 4th and the 5th sugars ***(Figure 1.5B).***[41-43]

Substantial studies in *Dictyostelium* and *Toxoplasma* support the model by which the prolyl hydroxylation together with glycosylation sense O_2.[34-36,41,44,45-47] *Dictyostelium* lives in the soil where the local O_2 levels can vary owing to the infusion of gas from the atmosphere, as well as constant aerobic metabolism from the soil microbiome. The amoeba feed on bacteria as a nutrient source, and upon starvation proceed to undergo various morphological changes into dormant spores. During the process, the amoeba gathers into multicellular aggregates and rearranges into slugs that travel to the surface of the soil. At the surface, these slugs transform into stalks with fruiting bodies containing dormant spores which are activated by a nutrient rich environment. O_2 sensing is important for

Dictyostelium to assess their environment whether they are in the soil or above ground in an area suitable for fruiting body formation. Therefore, culmination is dependent on O_2 levels, the threshold of O_2 for culmination can be controlled by PhyA[46-47]. For example, over-expression of PhyA requires higher O_2 levels for culmination while under-expression requires lower O_2 levels. Underexpressing Skp1 has the same effect on the O_2 levels required as PhyA over-expression, and the converse is true as well where Skp1 overexpression requires higher O_2 levels. This suggests that environment O_2 sensing in *Dictyostelium* is mediated through PhyA's activity on Skp1.

The importance of subsequent glycosylation can be demonstrated by genetic manipulation of each Skp1 GTs[37,38,39]. *Gnt1* disruption yields Hyp143 on Skp1, and its culmination O_2 level lies in between that of *PhyA*-KO and wild-type[37]. Additionally, the fruiting bodies form in a morphologically defective manner. Knocking out the second GT leaves Skp1 with a single GlcNAc modification which results in an O_2 requirement that is comparable to that of *gnt1*-KO, but with proper fruiting body morphology. The knock out of the last GT AgtA leaves Skp1 with a trisaccharide and interestingly with an O_2 dependence similar to that of *phyA*-KO, which can be attributed to AgtA's WD40 domain serving an addition function of binding Skp1[39]. Altogether the genetic data suggests that the full capability of O_2 sensing in *Dd* requires Skp1 to be modified by the full pentasaccharide.

Studies of Skp1 glycosylation system in an unrelated parasite *Toxoplasma gondii* (*Tg*) demonstrated its prominence in protists.[34,35,41-44] *Toxoplasma* infects most mammals, and the parasites are passed between hosts by ingestion of oocysts in cat fecal matter, or tissues cysts found in undercooked meat from other infected mammals. Once ingested the

parasites from oocysts or tissue cysts can be released within the acidic environment of the gut. The released parasites can navigate through the anaerobic environment in the gut to infect intestinal epithelial cells, where they can transform into replicative tachyzoites and disseminate through other tissues throughout the body using monocytes and dendritic cells. This way *Toxoplasma* is similar to *Dictyosteium* where it must navigate diverse O_2 tensions during its life cycle. Discovery of homologous genes PhyA and GTs in *Toxoplasma* gave clues as to how the parasite might sense O_2[34,35]. *PhyA* disruption of the parasite results in moderately slow growth phenotype at atmospheric O_2 levels, however, the growth defect is exaggerated at a low O_2 level where the parasites are likely to experience at physiological level in mammalian cell and tissues.[48] Like *Dictyostelium,* GT disruption resulted in milder growth defects that lies between *PhyA*-KO and w/t, suggesting that the full glycan is required on Skp1 to mediate the effect of PhyA.

The discovery of Skp1 glycan in protists presents itself as a unique O_2 sensing and SCF regulation in eukaryotes[2,10]. The O_2 sensing mechanism through Skp1 modification regulates Skp1/FBP binding to elicit a response by modulating protein degradation, but animals, fungi, and plants respond to low-oxygen stress (hypoxia) through transcriptional regulation.[2,10,49] The well-studied example is the animal hypoxia pathway modulated by the hypoxia inducible factor-α (HIF-α) ***(Figure 1.6)***. Under normoxic conditions, a P4H homologue PHD hydroxylates HIF-α at 2 Pro residues, which renders it a target substrate by a CRL-2 ubiquitin ligase and is subsequently degraded[49-53]. Under hypoxia, owing to the lack of O_2 substrates for PHD, HIF-α is not hydroxylated or degraded, therefore available to form a functional transcription factor by oligomerizing with HIF-β[50-53]. HIF-α/ HIF-β heterodimer complex then promotes the expression of genes associated with

hypoxia response.[50-53] Other pathways have been described in fungi and plants well, and much like PDH/HIF-α system, the downstream response to hypoxia is a transcriptional response.[49] In fungi, Sre1N is the transcriptional factor prolyl-hydoxylated by Ofd1 under normoxia, and it is ubiquitinated by UBR1 then degraded.[54] Rather than prolyl hydroxylation, examples of cysteine oxidases such as ADO from animals and PCO from plants also have been characterized, where hypoxia transcriptional factors are modified with cysteine sulfinic acids instead, but still are substrates to a ubiquitin ligase.[49]

1.5 The functional and structural effect of glycan on Skp1 structure and Skp1/FBP complex assembly

While in 1:1 molar ratio in a complex with an FBP or Cullin-1, the native state of Skp1 alone exists as a homodimer with an affinity that can range in the low micromolar K_D based on analytical ultracentrifugation (AUC) analysis.[55,56] Substantial *in vitro* studies with *Dd*Skp1 suggest that the glycan induces conformational changes, affecting its propensity for homodimerization.[57] Superdex 200 (S200) size-exclusion analysis on Skp1 glycoforms showed that glycosylated Skp1 elutes later than unmodified Skp1***(Figure 1.7)***. This indicated that the glycan influences Skp1 to behave as a smaller molecular weight species implying that Skp1 exists less as a dimer when it is modified with the glycan. DSS cross-linking analysis demonstrated consistent observation showing that unmodified Skp1 is more prone to dimer cross-linking than glycosylated Skp1. Probing Skp1 with small angle X-ray scattering (SAXS) and circular dichroism (CD) revealed that glycosylated Skp1 exhibits altered structural behavior. The SAXS envelope of Skp1 appeared to be more elongated in shape when modified showing that the glycan exerts a degree of structural

rearrangement of Skp1 ***(Figure 1.8)***. In support of this model, secondary structure composition analysis by CD indicated a higher level of α-helical content even with a single sugar modification ***(Figure 1.9)***. Altogether, these analyses suggest that the glycan exerts changes to the intrinsic property of Skp1 structure thereby inhibiting homodimerization.

The effect of glycan extends to the functional properties of Skp1 binding to FBPs[57]. Pull-down assays using a soluble FBP model (Fbs1 from guinea pig) as bait and Skp1 as targets demonstrated that all glycoforms are binding competent ***(Figure 1.7A)***. However, when subjected to a binding competition with a pool of Skp1 isoforms present, higher level of glycosylated Skp1 content was shown to elute with FBP based on S200 analysis, which demonstrated that glycosylated Skp1 has a higher propensity to interact FBP ***(Figure 1.7)***. This model is consistent with Skp1 co-immunoprecipitation studies performed with PhyA or GT knockout *Dictyostelium* strains representing Skp1 at various glycosylation stages.[58] The results showed the greatest association with an FBP with fully modified Skp1 and the least with unmodified Skp1, while the intermediate glycosylation stages revealed a general trend of step-wise increase with additional sugar moieties ***(Figure 1.10)***.

Owing to the flexibility and the disordered nature of the glycan and the C-terminal tail of Skp1, NMR spectroscopy and molecular dynamics (MD) simulation were employed to provide information regarding motions and the conformation of this region.[40] MD trajectories of glycosylated Skp1 where the extracted glycan relaxation kinetics and rotational correlation times agree with NMR experiments demonstrated that the glycan interacts with Skp1. The glycan folded back onto the loop region between helix 7 and 8, demonstrating hydrophobic packing as well as hydrogen bonding between the glycan hydroxyl group with Skp1 side chains and backbone amides ***(Figure 1.11)***. Consequently,

the C-terminal tail region adopted an open conformation and widely exposing the FBP binding site while closed and collapsed conformations were shown by the trajectories of unmodified and intermediate glycoforms. These observations suggest that the glycan promotes the C-terminal region into ensembles more likely to be accessible to FBP binding. Further residue-specific NMR analysis showed an abundance of chemical shift differences at the helix-8 when comparing various glycosylated Skp1 and unglycosylated Skp1 which suggests a change in the local structure ***(Figure 1.12A)***.[26] Even with a single sugar addition, Skp1 showed a higher $^{15}N\{^{1}H\}$-HNOE signal and a lower random coil index (RCI) which suggest an occurance or local ordering of disordered coil to a helical structure.[26]

1.6 Foundation

These experiment results point to two mechanism by which the glycan effects Skp1 to mediate the interaction with FBPs. First, the glycan promotes the dissociation of homodimers which increases available monomeric Skp1 for FBP interaction.[57] Second, the glycan promotes the extended and open conformation of Skp1 exposing FBP binding site, and the glycan induces a change in the secondary structure of the C-terminal disordered region into a helical structure perhaps with some resemblance to the one that is in a complex with FBP.[7,10,26,40] These models together suggest that the glycan tailors Skp1 structure in a manner that promotes FBP interaction.

Phylogenetic analysis suggests the widespread of Skp1 glycosylation systems in protists as exemplified the prominence of Skp1 modification genes in other organisms. The discovery of PhyA and Skp1 GTs in *Toxoplasma gondii* allowed characterization of a Skp1 glycan with a different sequence owing to a set of GTs (Glt1 and Gat1) with distinct

evolutionary origins that modify the terminal disaccharide.[34,35,42-44] This raises questions as to whether this mode of Skp1 regulation is specific for *Dictyostelium* or conserved throughout the other aerobic protists harboring PhyA and Skp1 GTs. In Chapter 2, the last step of *Tg*Skp1 glycan biosynthesis by Gat1 and the resulting glycan structure will be described in detail. Gat1 is a homolog of glycogenin in animals and yeasts, and whilst annotated as one in *Toxoplama,* it was demonstrated that Gat1 is an α1,3-galactosyltransferase specific for Skp1 and is not involved in starch synthesis. The resulting *Tg*Skp1 pentasaccharide sequence and linkage analysis showed that in comparison with the *Dd*Skp1, the only difference occurred at the 4th sugar position with Glc instead of Gal where the C'4 hydroxyl group is placed at an equatorial position. Further modeling by MD simulation of glycosylated *Tg*Skp1 showed a similar mode of action to that of *Dd*, where the glycan interacts with the Skp1 polypeptide and promoting an extended conformation. The C'4 hydroxyl group at this position resides in the bulk solvent, whether it is Glc or Gal, therefore did not alter its interaction with Skp1. Despite having synthesized by different sets of GTs, both *Dd* and *Tg* glycans demonstrated to be structurally compatible regarding the effect on Skp1 conformation.

Although the inhibitory effect of the glycan on Skp1 homodimerization has been well documented, the mechanism of action has been elusive owing to the gap in knowledge on how the dimer is arranged and assembled. Previous SAXS experiments yielded molecular envelopes that can accommodate two subunits of Skp1, but the resolution was not sufficient enough to confidently model the dimer interface nor the subunit orientations.[57] To this end in Chapter 3, efforts to characterize and solve the Skp1 dimer interface by Analytical ultracentrifugation (AUC) and NMR spectroscopy will be

described.[56] AUC was used to determine the binding affinity of *Dd*Skp1. Then, a truncated Skp1 construct lacking the disordered C-terminus and the internal loop was generated, effectively reducing the molecular weight while retaining its dimerization properties. This construct was amenable to high resolution NMR analysis, and the dimer structure was solved at an atomic resolution. The resulting structure revealed a two-fold symmetry arrangement with a predominantly hydrophobic interface. The dimer interface overlaps with FBP binding subsite 1, which explains why only the Skp1 monomer binds FBPs. The interface was tested with an amino acid substitution predicted to disrupt dimerization, and this construct behaved as a stable monomer and actively bound to an FBP, which suggested dimerization is not required for Skp1/FBP binding. The semi-parallel orientation of the dimer, which would presumably place each subunit's C-terminal region in an overlapping space with a potential for steric clash, and likely the respective glycans would reside in close proximity to each other as well as further crowding the space.

With the new and accurate Skp1 dimer structure model, how the glycan inhibits dimerization and glycan organizes the C-terminal region of Skp1 can be experimentally tested and observed using NMR. Chapter 4 will present the preparation of uniformly [^{13}C,^{15}N]-labeled glycosylated Skp1 with ^{13}C-labeled sugars. A list of NMR spectra was collected, and the analysis based on 2D HSQC spectra indicated that the sample is well behaved. Although the molecular weight of the glycosylated Skp1 dimer is greater than 40 kDa, NMR data quality is expected to be sufficient as the focus will be at the mobile and flexible region of Skp1.

1.7 Figures

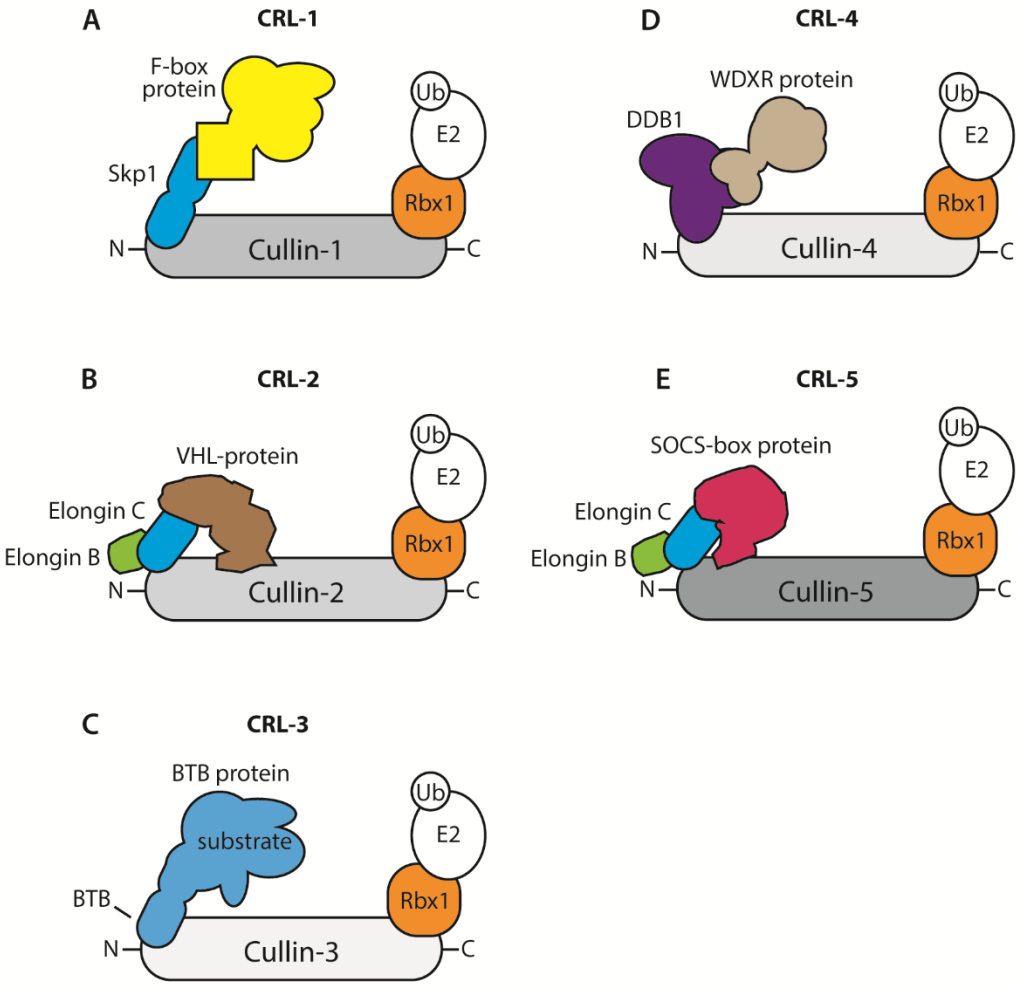


Figure 1.1. Diagrams of Cullin-RING ligase 1-5. Each CRL class consists of a Cullin with a distinct set of adaptors and substrate receptors. Rbx1 and the E2 enzyme are common to all CRLs. (A) CRL-1: Cullin-1 interacts with the adaptor protein Skp1 that recruits F-box motif containing substrate receptors. (B) CRL-2: Cullin-2 binds to the adaptor Elongin C of Elongin B/Elongin C heterodimer. Elongin-C recruits VHL-box containing substrate receptors (C) CRL-3: this class does not does not have an adaptor. Instead BTB domain containing substrates dock directly to Cullin-3 (D) CRL-4: Cullin-4 utilizes DDB1 adaptor to recruit WDXR domain containing receptors (E) CRL-5: Like CRL-2, Cullin-5 binds to Elongin B/Elongin C heterodimer adaptor, but recruits SOCS-box containing substrate receptors. The figures are adapted from Bosu, et al.[8]

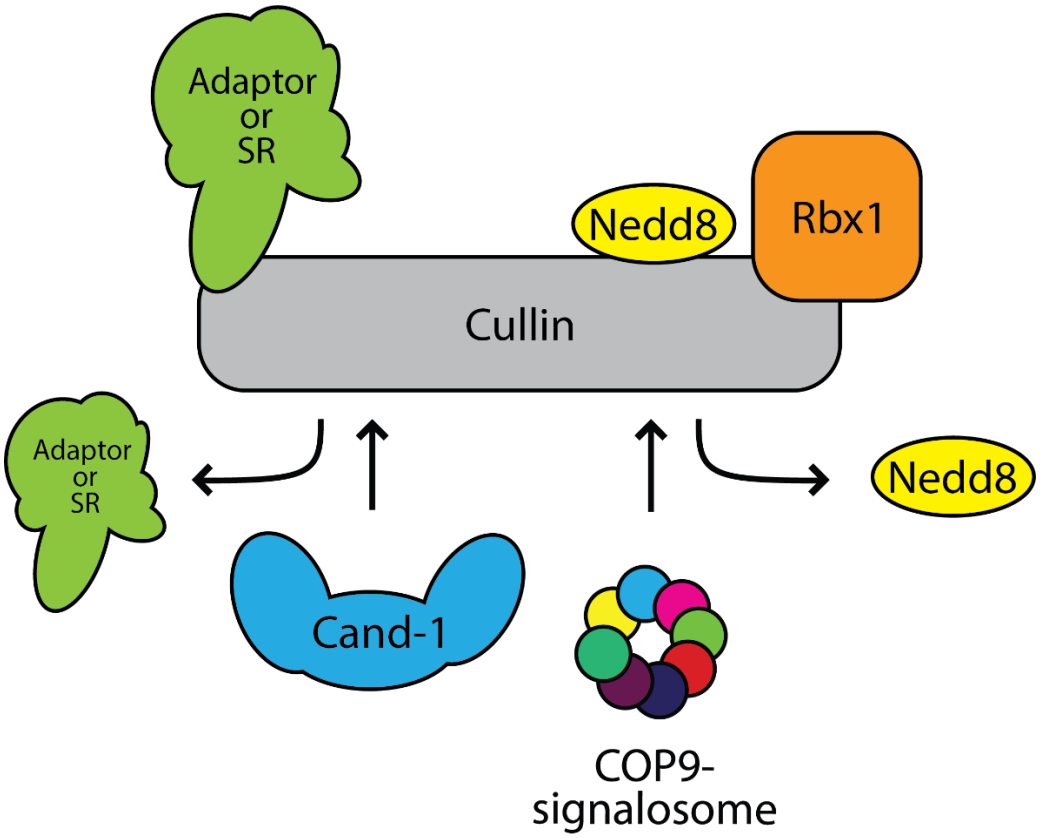


Figure 1.2. Regulatory elements of CRLs. A diagram representing the common regulatory mechanisms of CRLs. Nedd8 conjugation promotes the ligase activity by acting near Rbx1, and Nedd8 is removed by COP9 signalosome. Cand-1 binding dissociates the adaptor or substrate receptors.

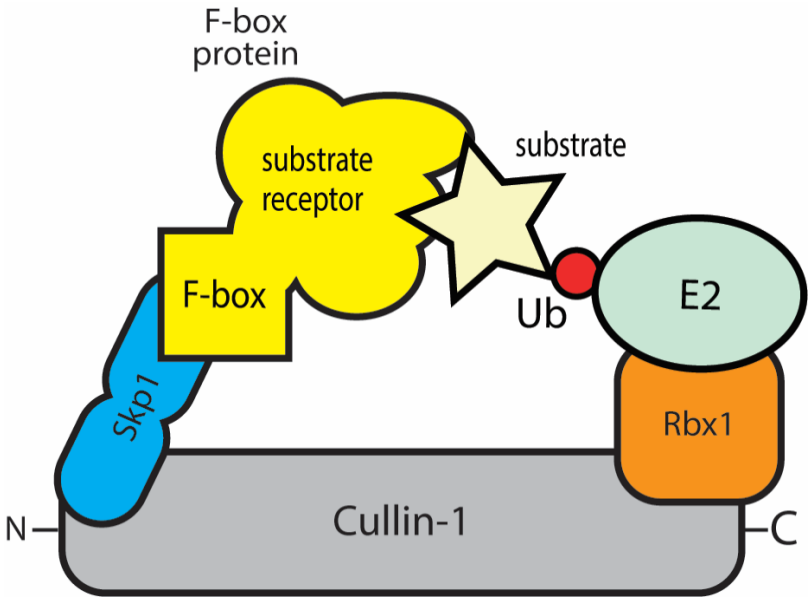


Figure 1.3. A diagram of the E3 SCF ubiquitin ligase (CRL-1). The SCF complex is composed Cullin-1 using Skp1 as an adaptor. In most organisms, a single Skp1 binds many dozens of F-box proteins which can recruit substrates to ubiquitinated and marked for proteasomal degradation.

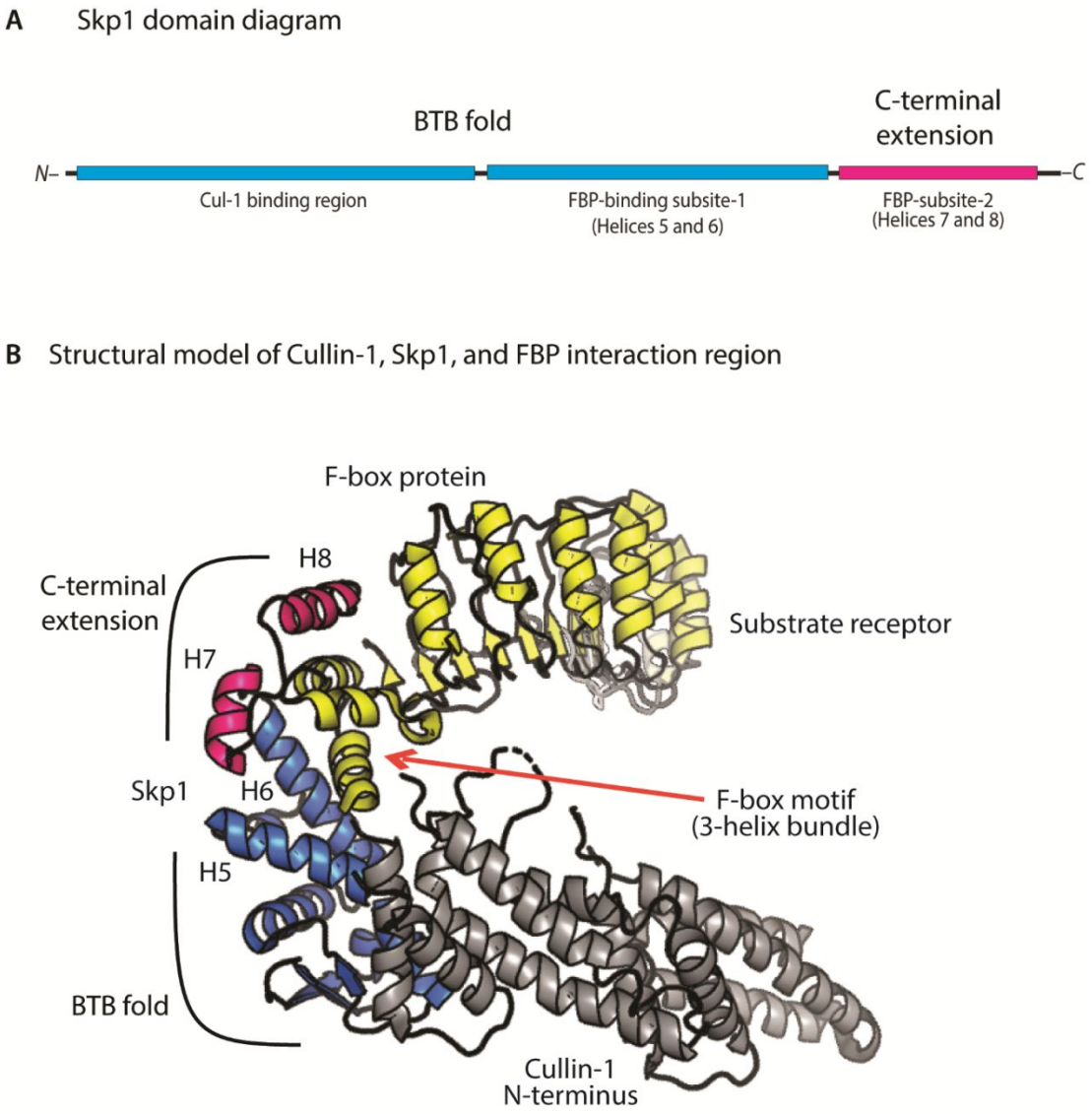
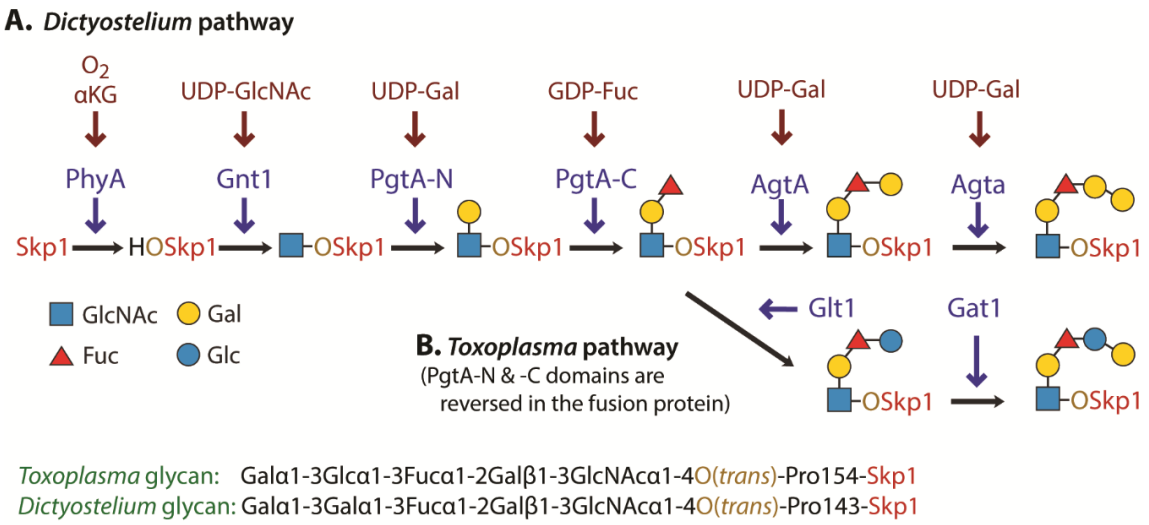


Figure 1.4. Skp1 is a BTB protein with a C-terminal extension. (A) A domain diagram depicting the regions of Skp1 associated with Cullin-1 and F-box protein interactions. Skp1 consists of a BTB fold with a C-terminal extension. (B) A structural model of Skp1/Cullin-1/F-box protein binding regions centered at Skp1. The BTB fold of Skp1 (blue) interacts with the N-terminus of Cullin-1 (gray). Skp1 recognizes the F-box motif (yellow) via a clasp-like structure formed between helices 5/6 of the BTB domain (blue) and helices 7/8 of the C-terminal extension (magenta). The model was assembled by aligning human Skp1's from Cullin-1/Skp1 and Skp1/FBP complex crystal structures.[10] (PDB: 1LDK, 1FQV)



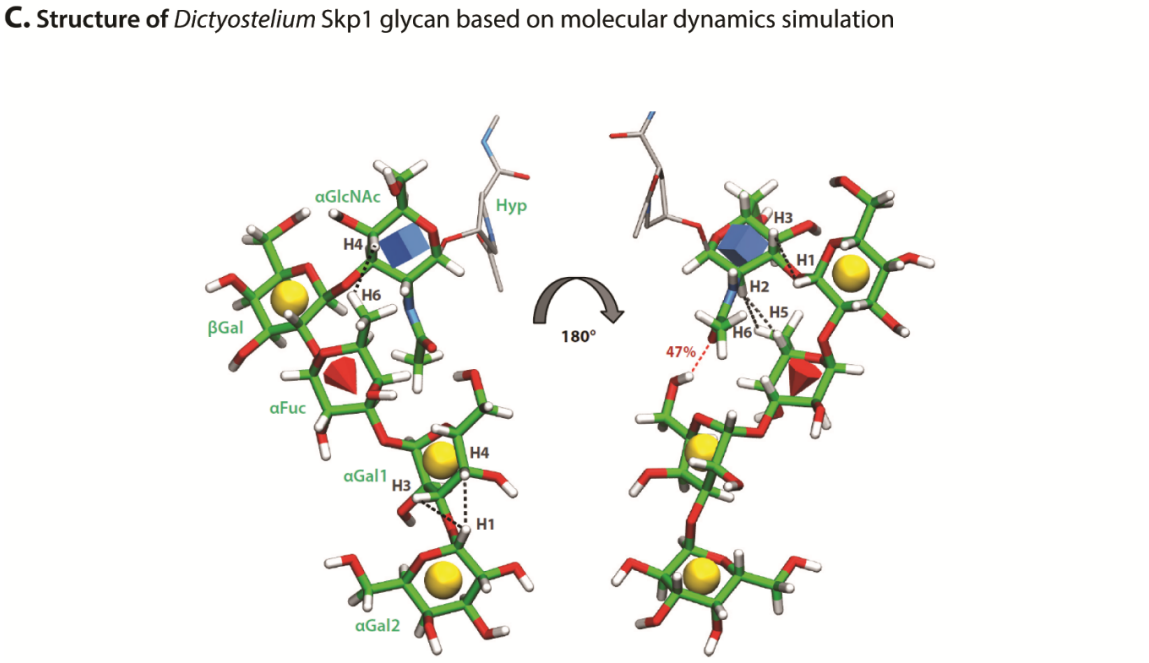


Figure 1.5. Skp1 glycosylation pathways. (A and B) A schematic diagram showing the enzymes associated with Skp1 glycosylation in *Dictyostelium* and *Toxoplasma*. In both species, PhyA hydroxylates Pro143 (position Pro156 in *Toxoplasma*), then Gnt1 and PgtA assembles the core trisaccharide. *Dictyostelium* has the Gal transferase AgtA that catalyzes the terminal disaccharide –Gal-Gal. In *Toxoplasma*, two glycosyltransferases Glt1 and Gat1 catalyzes assembly of –Glc-Gal. Sequences of the two glycans are shown. With the exception of Gal vs Glc at the 4th sugar position, both glycans are linked in an identical manner. (C) A representative structure of Hyp-glycan from a molecular dynamics simulation trajectory and energy minimization of the *Dictyostelium* version. Black dashed lines represent proton-proton contacts measurable by NMR NOESY experiments. The red

dash is the only hydrogen bond maintained >20% during the simulation. Figure C was adapted from Sheikh et al.[40]

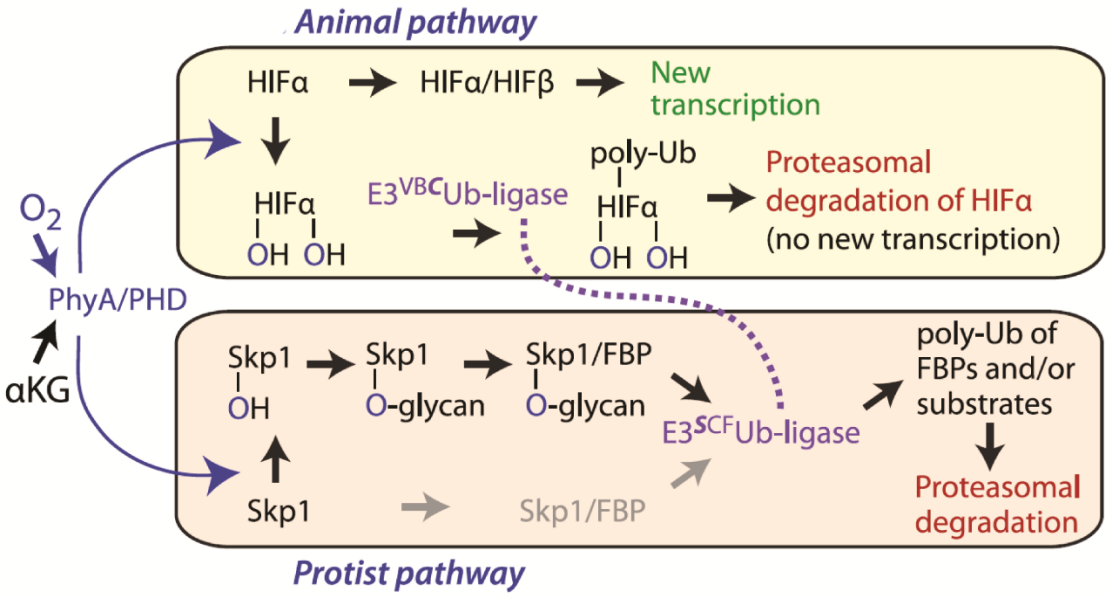


Figure 1.6. A contrast between the O_2 sensing modes in animals and protists. Shown is a schematic diagram of oxygen sensing processes in animals or protists. PhyA and PHD are prolyl-4-hydroxylases from Fe (II) and 2-alphaketoglutarate dependent dioxygenase family of enzymes, and are thought to be evolutionary orthologs despite their different substrates. The diagram was adapted from Xu et al.[50]

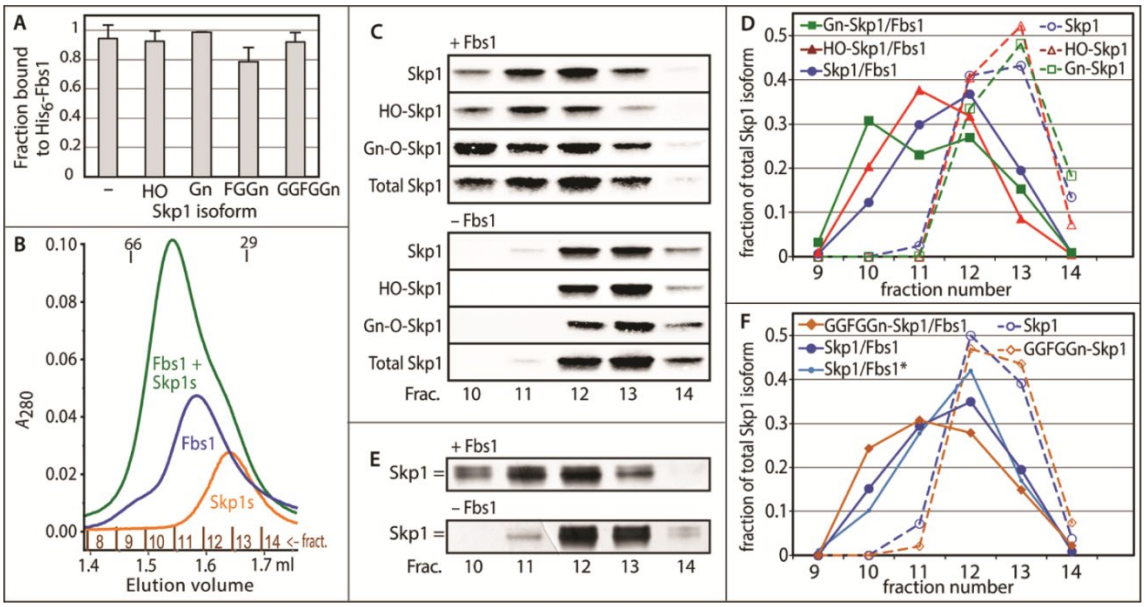


Figure 1.7. *in vitro* analysis of Skp1 glycoforms and their interaction with a model F-box protein. (A) Ni^{2+}-Sepharose pull-down assay using His_6Fbs1 from guinea pig and purified *Dictyostelium* Skp1 isoforms. Data shown is derived from densitrometric analysis of Western Blots. (B) Superdex 200 column elution profiles of Skp1s (Orange-Skp1, HO-Skp1, and Gn-Skp1), Fbs1 (blue), and Skp1s preincubated with Fbs1. (C) Western blots of the fractions from panel B detected using isoform specific antibodies. (D) Densitometric analysis of Western blots from panel C. (E) Similar S200 experiment with Fbs1 and a mix of Skp1 and GGFGGn-Skp1 monitored by SDS-PAGE and silver staining. The signal shows two bands consistent with molecular weight differences between modified and unmodified Skp1. (F) Densitometry analysis of the fractions shown in panel E. The figures were adapted from Sheikh et al.[40]

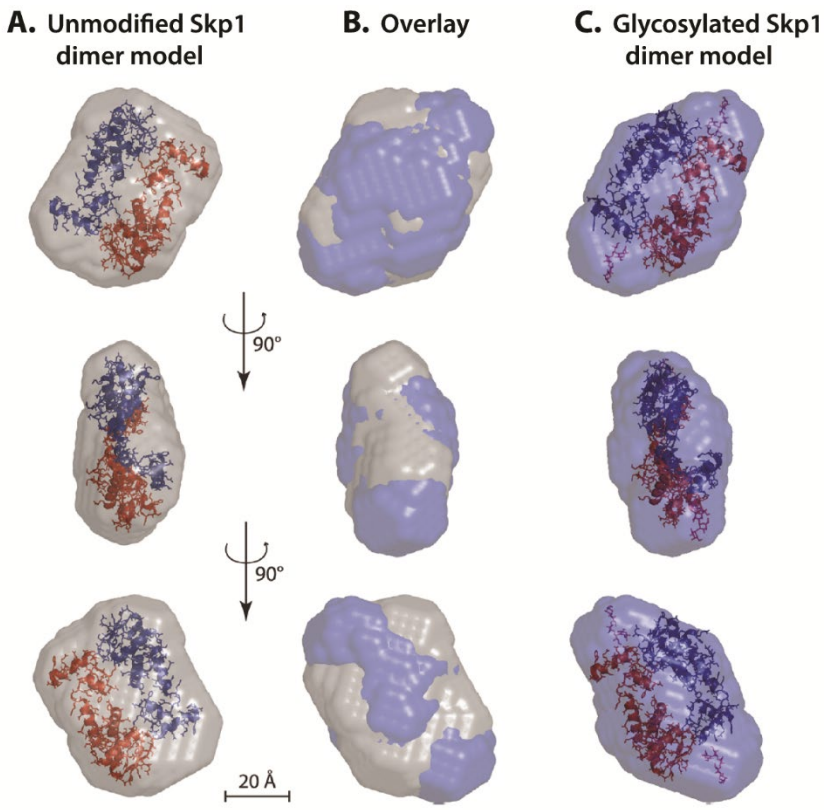


Figure 1.8. SAXS envelopes of Skp1 dimers. Envelopes were calculated from unmodified Skp1 (A) and GGFGGn-Skp1 (B). (C) Overlay of the envelopes shown in panels A and C. The figure was adapted from Sheikh et al.[40]

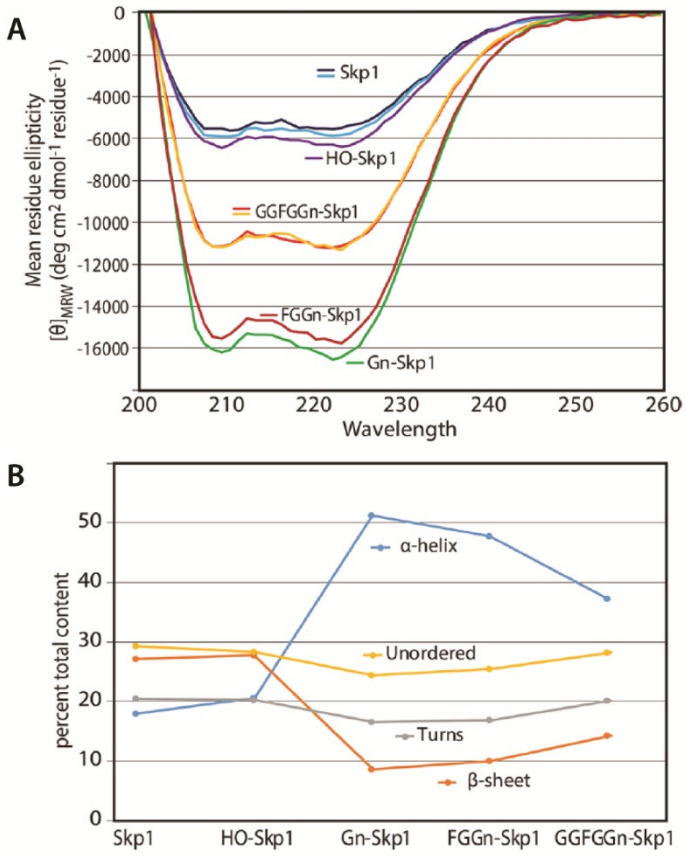


Figure 1.9. Circular dichroism spectroscopy of Skp1 isoforms. (A) CD spectra of Skp1 isoforms including replicates of sample subjected to freeze-thaw cycles. (B) A comparison of secondary structure composition of the isoforms derived from the spectra. The figure was adapted from Sheikh et al.[40]

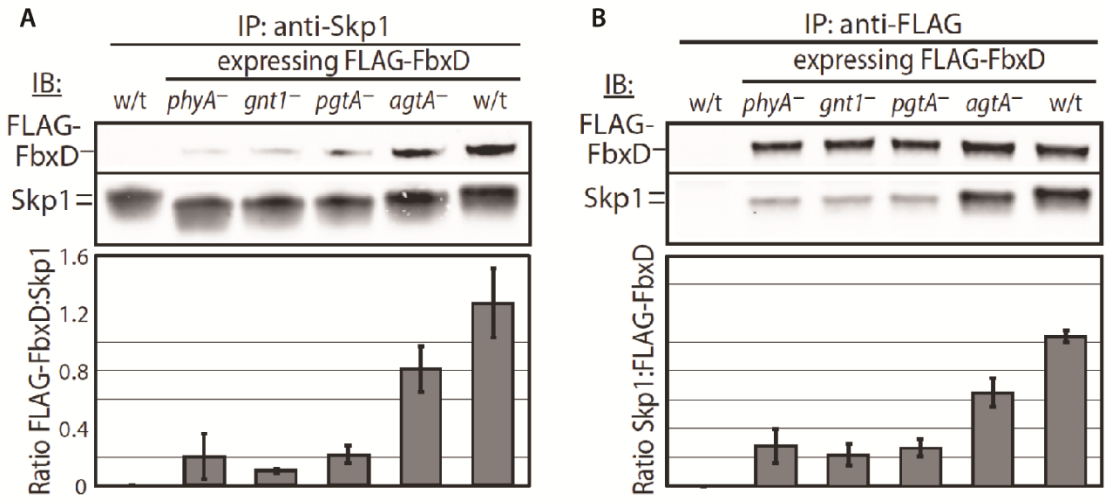


Figure 1.10. Glycosylation dependence assay using glycosylation pathway knockouts. *Dictyostelium* slug extracts overexpressing FLAG-FbxD in pathway knockout backgrounds were subjected to co-immunoprecipitation using anti-Skp1 (A) or anti-FLAG (B) antibodies. The bar graphs show the ratios of Western blot signal intensities based on densitometry. The figure was adapted from Sheikh et al.[58]

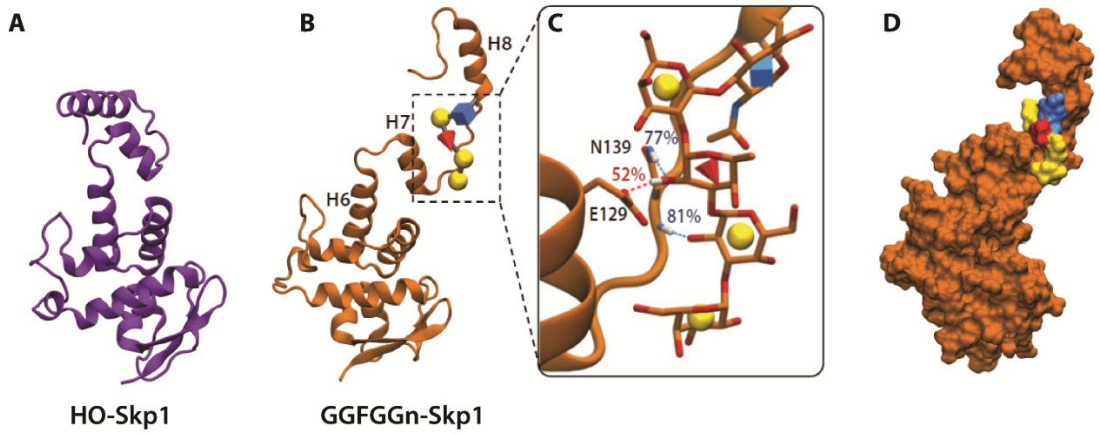


Figure 1.11. Molecular dynamics analysis of HO-Skp1 and GGFGGn-Skp1. (A) Average structure of HO-Skp. (B) The average structure GGFGGn-Skp1. (C) Zoomed-in and detailed view of panel B, showing the glycan interacting with Skp1. The dashes between the glycan and the protein are hydrogen bonds that are maintained over 40% throughout the course of the simulation. The colored dashes are according to which atom the hydrogen is connected to (Oxygen = red, Nitrogen=blue). (D) Space filling model of the ribbon diagram from panel B shows the packing of the glycan along the surface of the loop region between helix-7 and 8. The figure was adapted from Sheikh et al.[40]

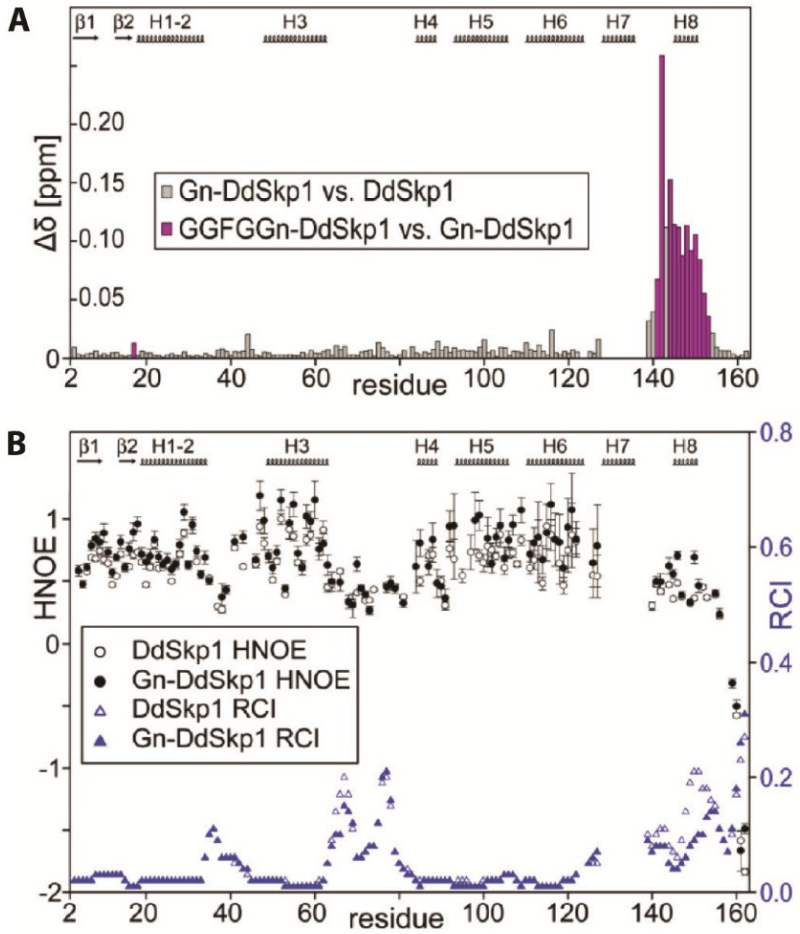


Figure 1.12. The effect of glycan on Skp1 monitored by NMR. (A) Weighted root mean square of $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$ and ^{13}CO chemical shift differences between Skp1 and Gn-Skp1 in a residue specific manner (gray). Residue whose $^{1}H^{N}$ ^{15}N chemical shifts are significantly perturbed in GGFGGn-Skp1 relative to Gn-Skp1 are in magenta. (B) Comparison of $^{15}N\{^{1}H\}$-HNOE and random coil index (RCI) between Skp1 and Gn-Skp1. Above each chart is a secondary structure element outlined for reference. The figure and the legend are adapted from Xu and Eletsky.[26]

CHAPTER 2

A TERMINAL α3-GALACTOSE MODIFICATION REGULATES AN E3 UBIQUITIN LIGASE SUBUNIT IN *TOXOPLASMA GONDII* [1]

[1] Mandalasi M‡, **Kim HW**‡, Thieker D, Sheikh MO, Gas-Pascual E, Rahman K, Zhao P, Daniel NG, van der Wel H, Ichikawa TH, Glushka JN, Wells L, Woods RJ, Wood ZA, West CM. 2020. *Journal of Biological Chemistry.* **295**:9223-9243. Reprinted here with permission of the publisher. ‡ denotes co-first authorship.

2.1 Abstract

Skp1, a subunit of E3 Skp1/Cullin-1/F-box protein ubiquitin ligases, is modified by a prolyl hydroxylase that mediates O2-regulation of the social amoeba *Dictyostelium* and the parasite *Toxoplasma gondii*. The full effect of hydroxylation requires modification of the hydroxyproline by a pentasaccharide that, in *Dictyostelium*, influences Skp1 structure to favor assembly of Skp1/F-box protein subcomplexes. In *Toxoplasma*, the presence of a contrasting penultimate sugar assembled by a different glycosyltransferase enables testing of the conformational control model. To define the final sugar and its linkage, here we identified the glycosyltransferase that completes the glycan and found that it is closely related to glycogenin, an enzyme that may prime glycogen synthesis in yeast and animals. However, the Toxoplasma enzyme catalyzes formation of a Galα1,3Glcα– rather than the Glcα1,4Glcα– linkage formed by glycogenin. Kinetic and crystallographic experiments showed that the glycosyltransferase Gat1 is specific for Skp1 in Toxoplasma and also in another protist, the crop pathogen *Pythium ultimum*. The fifth sugar is important for glycan function as indicated by the slow-growth phenotype of gat1Δ parasites. Computational analyses indicated that, despite the sequence difference, the Toxoplasma glycan still assumes an ordered conformation that controls Skp1 structure and revealed the importance of non-polar packing interactions of the fifth sugar. The substitution of glycosyltransferases in Toxoplasma and Pythium by an unrelated bifunctional enzyme that assembles a distinct but structurally compatible glycan in *Dictyostelium* is a remarkable case of convergent evolution, that emphasizes the importance of the terminal α-galactose and establishes the phylogenetic breadth of Skp1 glycoregulation.

2.2 Introduction

A prominent mechanism of O_2-sensing in metazoa involves an O_2-dependent prolyl 4-hydroxylase (PHD2) that generates degrons on Hypoxia Inducible Factor-α (HIFα) that are recognized by the von Hippel Lindau subunit of the E3(VBC) ubiquitin ligase, leading to its polyubiquitination and degradation in the 26S-proteasome.[1] Thus low O_2 inhibits PHD2 and stabilizes HIFα to dimerize with HIFβ and transcriptionally activate genes appropriate to respond to low O_2.[2] Protists also have a PHD2-like gene that encodes the evolutionary predecessor and likely ortholog of PHD2.[3] However, protists lack HIFα, and protist PhyA enzymes instead hydroxylate Skp1, a subunit of the SCF (Skp1/cullin-1/F-box protein)[1] family of E3 ubiquitin (Ub) ligases[4] that are related to the E3(VBC) Ub ligases. In the social amoeba *Dictyostelium discoideum*, Skp1 prolyl hydroxylation is involved in mediating the O_2-checkpoint for fruiting body formation.[5] Thus, both prolyl hydroxylases contribute to regulation of the proteome, but transcriptionally in animals and likely via degradation in protists.

The mechanism by which Skp1 hydroxylation contributes to O_2-sensing has been examined in *D. discoideum.*[6] Remarkably, the hydroxyproline (Hyp) residue is sequentially modified by a series of 5 glycosyltransferase (GT) reactions leading to the assembly of a linear pentasaccharide of recently defined structure.[7] Genetic studies reveal a strong Skp1-dependent role for the GTs in O_2-sensing [8], and radiotracer and biochemical complementation studies indicate that Skp1 is the sole target of these GTs. Skp1 serves as an adaptor that links the F-box protein (FBP) and cullin-1 subunits of the SCF complex through mostly independent binding events. Interactome studies indicate that glycosylation promotes accumulation of Skp1/FBP subcomplexes *in vitro* and *in vivo*[9,10], thereby

potentially activating the respective SCF complexes and contributing to the degradation of all the substrates that are recognized by the cellular repertoire of FBPs. Current evidence indicates that the glycan assumes a relatively rigid structure that organizes the intrinsically disordered subsite-2 region involved in FBP recognition. By increasing the fraction of time that the C-terminal region is folded into α-helix-8 and extended to form subsite-2, the glycan is thought to promote the interaction of Skp1 with FBPs in the cell.[7] However, the importance of the exact glycan sequence and the significance of the terminal (fifth) sugar deserve further attention because there is limited precedent for glycan-imposed order on local protein structure[6], and because the information will provide needed new approaches to probe its role in cellular O_2-sensing.

Toxoplasma gondii is an apicomplexan parasite that latently infects a sizeable fraction of many human populations[11]. Although chronic infections are usually clinically benign, reactivation, as can occur in immunosuppressed individuals, can lead to serious diseases of the central nervous system and other organs. Moreover, human fetuses are subject to serious birth defects in the case of acute maternal infection. Previous studies documented that the *T. gondii* prolyl hydroxylase (*phyA*) can complement a disruption of *phyA* in *D. discoideum*, and contribute to O_2-dependent growth of *T. gondii* in a fibroblast monolayer growth assay.[12] Furthermore, coding sequences for the first three GT activities are present though, surprisingly, *T. gondii* still assembles a pentasaccharide glycan on its Skp1[13]. This led to the discovery of a novel enzyme that assembles the fourth sugar[14] and, as reported here, another enzyme that mediates addition of the final sugar. Analytical studies revealed that the fourth sugar is different from that of *D. discoideum*, but the fifth has remained unknown. These findings have raised important questions, including why are

the GTs and glycans different between these two protists, what constrains their evolution, and are they compatible with the conformational regulation model posited for *D. discoideum*?

Pythium ultimum is an oomycete plant pathogen that resides in the stramenopile branch of the larger TSAR group of protozoans to which the apicomplexans also belong. *P. ultimum* is the agent for root rot disease in agriculturally important crops[15,16], and is related to *P. insidiosum*, the agent for debilitating pythiosus in humans and other mammals.[17] We recently showed *P. ultimum* PhyA and the first GT Gnt1 constitute a bifunctional protein that is active toward Skp1A but not the second Skp1B encoded by its genome.[18] Thus O_2-dependent posttranslational regulation of Skp1 appears to be widespread, and *P. ultimum*, and protists from other lineages, have evidently evolved a second Skp1 that avoids O_2-dependent glycoregulation. We show here that the terminal glycosylation of *P. ultimum* Skp1 is like that of *T. gondii* rather than that of *D. discoideum*.

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

2.3 Materials and Methods

Maintenance of host cells and parasite manipulations

Cultures of human foreskin fibroblast (HFF, ATCC SCRC-1041) or hTERT HFF (BJ-5ta, ATCC CRL-4001) were maintained Dulbecco's modified Eagle's Medium supplemented with 10% (v/v) fetal bovine serum, 2 mM L-glutamine and 100 units/ml penicillin/streptomycin (Corning) at 37°C in a humidified CO_2 (5%) incubator. Type 1 RH (21), RHΔ*Ku80*Δ*HXGPRT* (RHΔΔ)[22]and type 2 ME49-RFP[23] strains of *Toxoplasma gondii* were cultured on HFF or hTERT HFF monolayers in the same medium as described for host cells except that 1% (v/v) fetal bovine serum was used where stated, and cloned by limiting dilution in 96-well plates. Parasites were kept in media without drug for plaque assays, which were performed as before.[20] Transfections were conducted by electroporation using a BioRad Gene Pulser Xcell at 1.5 kV and 25 µF with 2 mm electroporation cuvettes (VWR international) in Cytomix buffer (10 mM KH_2PO_4/K_2HPO_4 (pH 7.6), 120 mM KCl, 0.15 mM $CaCl_2$, 5 mM $MgCl_2$, 25 mM HEPES, 2 mM EDTA).

Disruption and complementation of gat1

TGGT1_310400 (Toxodb.org), referred to as *gat1* ***(Figure S2.1)***, was disrupted by two independent approaches. In the first method, a disruption DNA was prepared from the vector pmini-GFP.*ht* (a gift from Dr. Gustavo Arrizabalaga) in which the *hxgprt* gene is flanked by multiple cloning sites as described[13]. 5'-flank and 3'-flank targeting sequences were PCR amplified from strain RHΔΔ with primer pairs Fa and Ra and pairs Fb and Rb, respectively ***(Table S2.1)***. The 5'-fragment was digested with ApaI and XhoI and cloned into similarly digested pminiGFP.ht. The resulting plasmid was digested with XbaI and NotI and ligated to the similarly digested 3'-flank DNA. The resulting vector was

linearized with PacI and electroporated into strain RHΔΔ, selected under 25 µg/ml mycophenolic acid and 25 µg/ml xanthine, and cloned by limiting dilution. Genomic DNA was screened by PCR to identify Tg*gat1* disruption clones ***(Fig. S2.3A)***, using primers listed in Table S1.

The second approach was based on a double-CRISPR/Cas9 method as previously detailed[20], with minor modifications. To generate the dual guide (DG) plasmid, a fragment of p2 containing the guide RNA gat1-63 expression cassette was PCR amplified using primers plasmid 3 FOR and plasmid 3 REV ***(Table S2.1)***, digested with NsiI, and ligated into the NsiI site of a dephosphorylated p3 containing guide RNA gat1-968. The type 1 RH and type 2 ME49 strains were co-transfected with pDG-Gat1 (10 µg) and a dihydroxyfolate reductase (*dhfr*) amplicon (1 µg) by electroporation ***(Figure S2.4A)***. CRISPR/Cas9-mediated disruption in RHΔΔ parasites was done similarly except that pDG-Gat1 was co transfected with a *dhfr* amplicon containing 45 bp homology arms targeting *gat1* ***(Figure S2.3B)***. *Gat*1Δ parasites were subsequently selected in 1 µM pyrimethamine (Sigma). The expected replacement of nt 63-968 (relative to A of ATG start codon) with the *dhfr* cassette was confirmed by PCR using primers listed in Table S2.1.

To complement *gat1*Δ RH parasites, a Tg*gat1* DNA fragment consisting of the coding sequence of *gat1* plus approximately 1.2 kb each of 5'- and 3'- flanking DNA sequences was generated by PCR from RH genomic DNA using primers Fa and Rb, which contained Apa1 and Not1 restriction sites respectively. After treatment with Apa1 and Not1, the PCR product was cloned into similarly digested pmini-GFP-*ht* plasmid in place of its *hxgprt* cassette, to generate pmini-Tg*gat1*. The plasmid was transformed into *E. coli* Top10 cells and purified by using a Monarch miniprep kit (NEB). A Ty tag DNA sequence was inserted

at the 3'-end of the *gat1* coding sequence using a Q5 site directed mutagenesis kit (NEB) and Fn and Rn primers, yielding pmini-Tg*gat1*-Ty. The sequence was confirmed using primers Fl and Rl ***(Table S2.1)***. RH *gat1*Δ clone 8 was complemented by co-electroporation with a PCR amplicon from pmini-Tg*gat1-Ty* (1 µg) and a sgUPRT CRISPR/Cas9 plasmid (10 µg) targeting the *uprt* locus using the guide DNA sequence 5'-ggcgtctcgattgtgagagc ***(Figure S2.4B)***.[51] Transformants were selected with 10 µM fluorodeoxyuridine (FUdR, Sigma), and drug resistant clones were screened by PCR with primers listed in Table S2.1.

To complement *gat1* in RHΔΔ parasites, the UPRT Vha1 cDNA shuttle vector containing TgVhaI cDNA [52] was modified to generate a Gat1-HA complementation plasmid using NEB HiFi Builder method. The vector backbone, containing 5'-flank and 3'-flank *uprt* targeting sequences and a Tg-tubulin promoter and 3×HA sequence, was PCR amplified from the shuttle vector using primers Ft and Rt. The coding sequence of *Tg*Gat1 was PCR amplified from pmini-Tg*gat1* plasmid using primers Fu and Ru, which had 18-21 nts complementary to the terminal ends of the vector ***(Figure S2.3C)***. The gel-purified PCR fragments were incubated with HiFi DNA assembly enzyme mix (NEB) and transformed into *E. coli* Top10 cells, yielding pUPRTg*at1*3xHA. The Tg*gat1* sequence was confirmed using primers Fl and Rl. Complementation in a *gat1*Δ clone derived from RHΔΔ was done similarly except that a *gat1*-3xHA PCR amplicon with 5'and 3' *uprt* homology arms was used in the transfection.

Bradyzoite induction

ME49-RFP tachyzoites were differentiated to bradyzoites using alkaline pH[53]. HFF monolayers were pre-incubated with sodium bicarbonate free RPMI (Corning) containing

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

Periodic acid staining

To assess amylopectin levels, parasite-infected HFF monolayers on 25 mm glass coverslips were washed with plain PBS, fixed with ice-cold MeOH for 5 min, washed with PBS, and incubated with 1% periodic acid in deionized H_2O in the dark for 10 min.[55] The coverslips were washed with deionized H_2O, incubated with Schiff reagent for 15 min, washed once with deionized H_2O, and finally rinsed with running tap water for 10 min. The stained coverslips were dehydrated by sequential immersion in 70% (v/v), 80%, 90% and 100% EtOH, mounted on glass slides using Permount mounting media (Fisher Scientific), and imaged on an EVOS XL Core microscope (Invitrogen).

Expression and purification of recombinant TgGat1 and PuGat1

The single exon coding sequence of Tg*gat1* cDNA was amplified by PCR from RH genomic DNA using primers, Gat1 Fw and Gat1 Rv (Table S2.1), cloned into PCR4-TOPO

TA (Invitrogen), and transformed into *E. coli* Top 10 cells. The plasmid was double digested with BamH1 and Nhe1 to yield the *gat1* fragment that was cloned into similarly digested pET15-TEVi plasmid (Invitrogen), resulting in the original 346 amino acid coding sequence of Gat1 extended at its N-terminus with a His_6-tag and TEV protease cleavage site (MGSSHHHHHHSSGRENLYFQGH-). A similar N-terminal modification of rabbit glycogenin did not significantly alter its enzymatic activity.[35] The predicted coding sequence for Pu*gat1* was inferred from PYU1_G002535-201 at protists.ensembl.org/Pythium_ultimum. The coding sequence was codon optimized for *E. coli* expression, chemically synthesized by Norclone Biotech (Ontario, Canada), and inserted in the NdeI and XhoI sites of pET15b-TEV. The expressed protein was extended at its N-terminus with MGSSHHHHHHSSGENLYFQGH-.

TgGat1 or PuGat1 were expressed in and purified from *E. coli* BL21-Gold cells as previously described for TgGlt1[14], through the purification on a 5-ml Co^{+2} TALON resin column. The eluted protein was dialyzed in 50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 1 mM EDTA and 2 mM β-mercaptoethanol, followed by 25 mM Tris-HCl (pH 8.0), 200 mM NaCl, 1 mM β-mercaptoethanol and 5 mM $MnCl_2$. The sample was treated with 2 μM His_6-TEV protease, 5 μM TCEP in the same buffer overnight at 20°C, and reapplied to another Co^{+2} TALON column. The flow-through fraction was dialyzed in 25 mM Tris-HCl (pH 8.0), 50 mM NaCl, 1 mM β-mercaptoethanol, 2 mM $MnCl_2$. The sample was concentrated by centrifugal ultrafiltration and aliquots were stored at -80°C. As indicated, the preparations were further purified by gel filtration on a Superdex200 column in the same buffer.

SDS-PAGE and Western blotting

Samples were suspended in diluted with Laemmli sample buffer and typically electrophoresed on a 4–12% gradient SDS-polyacrylamide gel (NuPAGE Novex, Invitrogen). Gels were either stained with Coomassie blue or transferred to a nitrocellulose membrane using an iBlot system (Invitrogen). Blots were typically blocked in 5% non-fat dry milk in Tris-buffered saline and probed with a 1:1000-fold dilution of the antibody of interest in the milk solution, followed by secondary probing with a 1:10,000-fold dilution of Alexa-680-labeled goat anti-rabbit IgG secondary antibody (Invitrogen). Blots were imaged on a Li-Cor Odyssey infrared scanner and analyzed in Adobe Photoshop with no contrast enhancement. For measuring incorporation of radioactivity, 1-mm thick 7-20% acrylamide gels were prepared manually, as detailed[13].

Preparation of Skp1 peptides

To monitor Skp1 glycosylation status, endogenous TgSkp1 was purified from parasite extracts essentially as described[13]. Briefly, frozen pelleted tachyzoites (1×10^8) were resuspended in 8 M urea in 50 mM HEPES-NaOH (pH 7.4), incubated on ice for 1 h and at 50°C for 5 mins, and diluted 8-fold in IP buffer (0.2% Nonidet P-40 in 50 mM HEPES-NaOH, pH 7.4). The lysates were centrifuged at 21,000 × *g* for 20 min at 4°C, and 100 µl of the supernatant (5×10^7 cells) were incubated with 5 µl UOK75 rabbit antibody (affinity-purified anti-TgSkp1) coupled to protein A/G magnetic agarose beads (Pierce, 78609) for 1 h at 4°C. Beads were captured in a DynaMag-2 magnet (Life Technologies) according to the manufacturer's directions, and washed 3× with 50 mM HEPES-NaOH (pH 7.4), 3× with 10 mM Tris-HCl (pH 7.4), 50 mM NaCl, and once with 50 mM NaCl. Bound Skp1 was eluted twice with 60 µl 133 mM triethylamine (TEA, Sequencing Grade,

Pierce, 25108) by incubating for 15 mins at RT, and immediately neutralized with 40 µl of 0.2 M acetic acid. The eluted material was pooled and dried under vacuum, reconstituted in 8 M urea, 10 mM Tris HCl (pH 7.4), reduced in 10 mM DTT for 40 min at RT, and alkylated in 50 mM chloroacetamide for 30 min at RT. Samples were then diluted to 2 M urea with 10 mM Tris HCl (pH 7.4), and digested in 10 µg/ml Trypsin Gold (Mass Spectrometry Grade, Promega, V5280) overnight at RT. Excess trypsin was quenched by the addition of 1% (v/v) trifluoroacetic acid (TFA, Pierce, 28904) on ice for 15 min and centrifuged at 1,800 × *g* for 15 min at 4°C. The supernatants were adsorbed to C18 pipette tips (Bond Elut OMIX C18, Agilent, A7003100), and eluted in 50% acetonitrile (ACN, Optima^{TM} LC/MS Grade, Fisher Chemical, A955-4), 0.1% formic acid (FA, LC-MS Grade, Pierce, 28905). Eluted peptides were vacuum dried, reconstituted in 40 µl 5% ACN, 0.05% TFA, and 4 µl were analyzed by nLC-MS/MS.

Treatment of TgSkp1 peptides with α-galactosidase

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

Mass spectrometry of TgSkp1 peptides

Reconstituted peptides were loaded onto an Acclaim PepMap C18 trap column (300 µm, 100 Å) in 2% (v/v) ACN, 0.05% (v/v) TFA at 5 µl/min, eluted onto and from an Acclaim PepMap RSLC C18 column (75 µm × 150 mm, 2 µm, 100 Å) with a linear gradient consisting of 4-90% solvent B (solvent A: 0.1% FA; solvent B: 90% ACN, 0.08% (v/v)

FA) over 180 min, at a flow rate of 300 nl/min with an Ultimate 3000 RSLCnano UHPLC system, into the ion source of an Orbitrap QE+ mass spectrometer (Thermo Fisher Scientific). The spray voltage was set to 1.9 kV and the heated capillary temperature was set to 280°C. Full MS scans were acquired from *m/z* 350 to 2000 at 70k resolution, and MS^2 scans following higher energy collision-induced dissociation (HCD, 30) were collected for the Top10 most intense ions, with a 30-sec dynamic exclusion. The acquired raw spectra were analyzed using Sequest HT (Proteome Discoverer 2.2, Thermo Fisher Scientific) with a full MS peptide tolerance of 10 ppm and MS^2 peptide fragment tolerance of 0.02 Da, and filtered to generate a 1% target decoy peptide-spectrum match (PSM) false discovery rate for protein assignments. All known glycoforms for TgSkp1 specific glycopeptides were manually searched for and verified. Searches were performed against the *T. gondii* (strain ATCC 50853/GT1) proteome (Uniprot proteome ID UP000005641, downloaded May 18, 2018; 8,450 entries) plus a list of common keratin, immunoglobulin-derived and trypsin contaminants. Carbamidomethylation of Cys was set as a fixed/static modification and oxidation of Met, deamidation of Asn and Gln residues, and acetylation of protein N-termini were set as variable/dynamic modifications. Searches were performed with trypsin cleavage specificity, allowing two missed cleavage events and a minimum peptide length of 6 residues. All known glycoforms for TgSkp1 specific glycopeptides were manually searched for and verified. The raw files were uploaded to the Figshare server at https://figshare.com/ with ID 10.6084/m9.figshare.12272882 (Skp1 glycopeptides raw data Fig. S2.5).

Enzyme assays

Sugar nucleotide hydrolysis: Recombinant TgGat1 (0.625-2.5 μM) was incubated with a given UDP-sugar (50 μM) (Promega) in 50 mM HEPES-NaOH (pH 7.4), 2 mM $MnCl_2$, 5 mM DTT, in a final volume of 20 μl, for 1-16 h at 37°C. The UDP generated was quantitated using the UDP-Glo assay (Promega) as described.[56]

Glycosyltransferase activity toward small glycosides: In the standard reaction, TgGat1 or PuGat1 was incubated with 2 mM synthetic glycosides [pNP-α-galactoside (pNP-αGal), pNP-β-galactoside (pNP-β-Gal), pNP-α-glucoside (pNP-αGlc), pNP-β-glucoside (pNP-βGlc), pNP-α-maltoside, (pNP-malt), chloro-4-nitrophenyl-α-maltotrioside (pNClP-trimalt), pNP-Skp1 trisaccharide (FGaGn-pNP), pNP-Skp1 tetrasaccharide (GlFGaGn-pNP)], 8 μM UDP-Gal (unlabeled), 0.17 μM UDP-[^{3}H]Gal (15.6 μCi/nmol, American Radiolabeled Chemicals), 50 mM HEPES-NaOH (pH 7.0), 2 mM $MnCl_2$, 5 mM DTT, in a final volume of 30 μl, for 1 h at 37°C. Pilot studies indicated a pH optimum of 7.0, with 50% activity at pH 8.0 and 75% activity at pH 6.0 for TgGat1. Salt dependence studies showed maximal activity with no added NaCl or KCl, and 35% activity at 800 mM of either salt. Activity showed a ~6-fold preference for $MnCl_2$ over $MgCl_2$, with activity maximal at 2 mM $MnCl_2$. The enzyme was essentially inactive in $NiCl_2$, $CoCl_2$ and $CaCl_2$ (Fig. S2.9A-D). For kinetic studies, concentrations and times were varied as indicated, and kinetic parameters were analyzed according to the Michaelis-Menten model based on the least squares fitting method in Graph pad Prism software. Reactions were stopped by addition of 1 ml 1 mM ice-cold Na-EDTA (pH 8.0), and incorporation of radioactivity into pNP-glycosides was analyzed by capture and release from a Sep-Pak C18 cartridge and scintillation counting.[14]

Glycosyltransferase activity toward GlFGaGn-Skp1: To prepare Tg-GlFGaGn-Skp1, 2 nmols (40 µg) of recombinant TgSkp1 FGaGn-Skp1[14] was incubated with 1.3 nmols (88 µg) of Tg-His_6-Glt1, 4 nmols UDP-Glc, 1.2 units alkaline phosphatase (Promega), 50 mM HEPES-NaOH (pH 8.0), 5 mM DTT, 2 mM $MnCl_2$, 2 mM $MgCl_2$, 2 mg/ml BSA in a final volume of 121 µl for 3.5 h at 37°C. The reaction was initiated by addition of UDP-Glc and terminated by freezing at -80°C. Reaction progress was monitored by Western blotting with pAb UOK104, which is specific for FGGn-Skp1 from either *D. discoideum* or *T. gondii*[57], followed by probing with pAb UOK75, which is pan-specific for all TgSkp1 isoforms, for normalization. Approximately 85% of total Skp1 was modified. GlFGaGn-Skp1 was purified from Glt1 on a mini-QAE-column using a Pharmacia Biotech SMART System. ~19µg of GlFGaGn-Skp1 was applied to a mini QAE column pre-equilibrated with 50 mM Tris-HCl (pH 7.8), 5mM $MgCl_2$, 0.1 mM EDTA (buffer A) and eluted with a gradient from 0% A to 100% buffer B (50 mM Tris-HCl (pH 7.8), 5 mM $MgCl_2$, 0.1mM EDTA, 300 mM NaCl) in 40 min at a flow rate of 240 µl/min. GlFGaGn-Skp1fractions were identified based on A280 and Western blot probing with pAb UOK75.

TgGat1 (0.127 µM) was incubated with GlFGaGn-Skp1 or FGaGn-Skp1 (0.9-3.6 µM), 3.2 µM UDP-[^{3}H]Gal (15.6 µCi/nmol), 40 µM UDP-Gal (unlabeled), 0.2% (v/v) Tween-20, 50 mM HEPES-NaOH (pH 7.0), 2 mM $MnCl_2$, 5 mM DTT, in a final volume of 20 µl, for 1 h at 37°C. The reaction was stopped by addition of 4×-Laemmli sample buffer, 1 M DTT (50 mM final concentration), and 2 µg soybean trypsin inhibitor, and boiled for 3 min. SDS-PAGE and incorporation of radioactivity into Skp1 was performed as described above.

To detect Gat1 activity in cells, a cytosolic extract was prepared by hypotonic lysis, ultracentrifugation at 100,000 *g* × 1 h, and desalted as previously described[13]. 25 µg of desalted S100 protein was incubated with 10-50 nmol Tg-GlFGaGn-Skp1 and 1.0 µCi UDP-[^{3}H]Gal (15.6 µCi/nmol) for 5 h, and incorporation of radioactivity into protein was assayed by SDS-PAGE and scintillation counting as described above.

Glycosyltransferase activity toward parasite extracts: To search for Gat1 substrates, cytosolic S100 fractions (180 µg protein) were incubated with 0.13 µM TgGat1, 2.0 µCi UDP-[^{3}H]Gal (15.6 µCi/nmol) in a final volume of 60 µl containing 50 mM HEPES-NaOH (pH 7.0), 2 mM $MnCl_2$, 5 mM DTT, 30 mM NaF, 0.2% Tween-20, at 37°C for 1 h, supplemented with 1.7 µg Tg-GlFGaGN-Skp1 as indicated. Incorporation of radioactivity was monitored by the SDS-PAGE assay as described above.

Mass spectrometry of Skp1

To evaluate its glycosylation status, recombinant PuGat1 or TgGat1 (purified by gel filtration) were incubated with or without UDP-Gal or UDP-Glc in the absence of added acceptor substrate, and diluted to 50 ng/µl Skp1 with 2% acetonitrile, 0.05% (v/v) trifluoroacetic acid. 250-500 ng of protein (5-10 µl) was injected into an Acclaim PepMap C4 trap cartridge (300 µm × 5 mm) equilibrated with 0.05% trifluoroacetic acid, 2% acetonitrile, ramped up with an increasing gradient to 0.1% formic acid, 25% acetonitrile, and introduced into an Acclaim PepMap analytical C4 column (75 µm × 15 cm, 5 µm pore size) maintained at 35°C in an Ultimate 3000 RSLC system coupled to a QE+ Orbitrap mass spectrometer (Thermo Scientific). After equilibrating the analytical column in 98% LC-MS Buffer A (water, 0.1% formic acid) for 10 min and 6-min ramp up to 27% LC-MS Buffer B [90% (v/v) acetonitrile, 0.1% formic acid], separation was achieved using a linear

gradient from 27% to 98% Buffer B over 20 min at a flow rate of 300 nl/min. The column was regenerated after each run by maintaining it at 98% Buffer B for 5 min. The effluent was introduced into the mass spectrometer by nanospray ionization in positive ion mode via a stainless-steel emitter with spray voltage set to 1.9 k, capillary temperature set at 250°C and probe heater temperature set at 350°C. The MS method consisted of collecting Full ITMS (MS^1) scans (400-2000 *m*/*z*) at 140,000 resolution in intact protein mode (default gas P set to 0.2). PuGat1 species eluting between 17.5 and 21.5 min and TgGat1 species eluting between 18.5 and 22.5 min (~60 to 80% acetonitrile) were processed with Xcalibur Xtract deconvolution software to generate monoisotopic masses from the multicharged, protonated ion series. Since TgGat1 MS spectra were not isotopically resolved, masses were extracted after MS spectra deconvolution using the ReSpect algorithm in the BioPharma Finder suite (Thermo Scientific), with a 20 ppm deconvolution mass tolerance and 25 ppm protein sequence matching mass tolerance. For consistency, PuGat1 was also deconvoluted and re-extracted using the ReSpect algorithm and the same conditions. The data raw files were uploaded to the Figshare server at https://figshare.com/ with ID 10.6084/m9.figshare.12272909 (Recombinant Gat1 intact protein raw files Fig. S2.10).

Structure determination of PuGat1

A PuGat1:UDP:Mn^{2+} complex in 50 mM HEPES-NaOH, pH 7.4, 75 mM NaCl, 2 mM DTT, 5 mM UDP, and 5 mM $MnCl_2$ was crystallized at 20°C using a hanging drop vapor diffusion method over a reservoir containing 8-12% (w/v) PEG4000, 0.4 M ammonium sulfate, and 0.1 M sodium acetate at pH 4.0. Crystals were obtained overnight and were transferred to a reservoir solution containing 15% (v/v) of a cryoprotectant

mixture (1:1:1 ethylene glycol:dimethyl sulfoxide:glycerol), and flash cooled with liquid N_2. The complex crystallized in space group $P4_22_12$ and diffracted to 1.76 Å (Table S2.3). X-ray data were collected remotely at the SER-CAT 22-BM beamline at the Argonne National Laboratory using a Fast Rayonix 300HS detector, and processed using XDS[58], with 5% of the data omitted for cross validation.

PuGat1:UDP:Mn^{2+} crystals were soaked with platinum cyanide for heavy-atom phasing, cryoprotected, and frozen as above. PuGat1:UDP:Pt^{2+} crystals were isomorphous to PuGat1:UDP:Mn^{2+} crystals and diffracted to 2.1 Å (Table S2.3). PuGat1:UDP:Mn^{2+} crystals were alternatively soaked in UO_2 (not shown). The crystals, which diffracted to 2.4 Å in the same space group, lacked UDP density, suggesting displacement by UO_2.

The crystal structure of PuGat1:Pt^{2+} was solved using single-wavelength anomalous dispersion (SAD). The data was obtained at a wavelength of 1.85 Å for maximum anomalous signal. A single Pt^{2+} site was located using PHENIX[59], and the resulting phases had an acceptable figure of merit of 0.31. The model was subjected to iterative cycles of refinement and yielded a final model with R_{work}/R_{free} of 0.21/0.24 (Table S2.3). The structure of PuGat1:UDP:Mn^{2+} was solved using rigid body refinement of PuGat1:Pt^{2+}. The resulting model was subjected to iterative cycles of refinement and yielded a final model with R_{work}/R_{free} of 0.18/0.21 ***(Table S2.3)***. Images were rendered in PyMol[60], and secondary structures were assigned based on DSSP.[61,62]

<u>Glycan docking</u>

The lowest energy conformation of the TgSkp1 tetrasaccharide (Glcα1,3Fucα1,2Galβ1,3GlcNAcα1-OH) was generated via GLYCAM5[63]. Hydrogen atoms were added, and the electrostatic surface was generated using AutoDockTools.[64] A

grid box with dimensions 26 Å × 26 Å × 34 Å was placed over the ligand binding site based on where the acceptor is bound on glycogenin/glucan complex. The ligand was kept rigid, since AutoDock Vina is not parameterized specific to glycan torsion angles. 100 binding modes were calculated with the lowest binding energy scored at -4.4 kcal/mol and the highest binding energy scored at -5.7 kcal/mol.

Sedimentation velocity studies

PuGat1 was further purified on a Superdex S200 gel filtration column (GE Healthcare) equilibrated with 20 mM potassium phosphate (pH 7.4), 50 mM KCl, 0.5 mM TCEP. Protein concentration was calculated from A_{280} measured in an Agilent 8453 UV/Vis spectrophotometer, based on a molar absorptivity (ε_{280}) of 60390 $M^{-1}cm^{-1}$, which was calculated from the PuGat1 sequence using ProtParam.[65] Samples were diluted to 0.3-11 μM, loaded into 12 mm double-sector Epon centerpieces equipped with quartz windows, and equilibrated for 2 h at 20°C in an An60 Ti rotor. Sedimentation velocity data were collected using an Optima XLA analytical ultracentrifuge (Beckman Coulter) at a rotor speed of 50000 RPM at 20°C. Data were recorded at 280 nm for protein samples at 3.5-11 μM, and at 230/220 nm for samples at 0.3-1.5 μM, in radial step sizes of 0.003 cm. SEDNTERP[66] was used to model the partial specific volume of PuGat1 (0.73818 mL/g), and the density (1.0034 g/ml) and viscosity (0.0100757 P) of the buffer. Using SEDFIT[67], data were modeled as continuous *c*(*s*) distributions and were fit using baseline, meniscus, frictional coefficient, and systematic time-invariant and radial-invariant noise. Predicted sedimentation coefficient (s) values for the PuGat1 monomer (2.8 S) and dimer (4.2 S) were calculated using HYDROPRO[68]. Data fit and *c*(*s*) plots were generated using GUSSI.[69]

Molecular dynamics simulations

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

A 50-ns simulation of the protein was performed with Cα cartesian constraints on

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

Computational Analysis

Structural images were created with Visual Molecular Dynamics[80] and the 3D-SNFG plugin.[81] The structure depicted in Fig. 2.7 was created by identifying the frame from equil-1 that consisted of the lowest RMSD to the average structure as calculated by cpptraj[82]. The cpptraj program was also used to distribute the latter 200 ns of the six simulations into 48 bins containing 250 frames each for analysis. Per-residue MMGBSA energies were calculated with MMPBSA.py.MPI with igb=2 and idecomp=3. A bash script was used to calculate the correlation coefficients.[83]

Phylogenetic analysis of enzyme sequences

Proteins related to TgGat1 were searched for using a BLASTP **(V 2.4.0)** search seeded with the full-length TgGat1 protein sequence against the NCBI non-redundant database (December 2016). The evolutionary relationship of Gat1-like sequences was investigated by using a Maximum Likelihood method[84] and conducted in MEGA7[85]. Catalytic domains from 43 CAZy GT8 sequences selected based on their relatedness to Gat1, glycogenin, or known function, and consisted of 196 positions. Sequence alignments were manually-curated in BioEdit (v 7.2.5). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with

superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 1.1608)). The rate variation model allowed for some sites to be evolutionarily invariable ([+*I*], 1.02% sites).

2.4 Results

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

Previous studies described the mechanism of assembly of the first four sugars on TgSkp1-Pro154 ***(Figure 2.1A)***, but the left the identity of the final sugar (other than its being a hexose) unresolved[13,14]. The corresponding sugar in *Dictyostelium* ***(Figure 2.1B)*** is a 3-linked αGal that is susceptible to removal with green coffee bean α-galactosidase. Similar treatment of tryptic peptides from TgSkp1, isolated from the standard type 1 strain RH by immunoprecipitation, resulted in complete conversion of the pentasaccharide form of the glycopeptide to the tetrasaccharide form, indicating that the terminal Hex is an αGal residue ***(Figure 2.1C)***.

Genomics studies predict the existence of four cytoplasmically localized GTs whose functions are not assigned[20], and one of these, referred to as Gat1, was predicted to be the missing Skp1 GT on account of the phylogenetic co-distribution of its gene in protists that possess Glt1-like genes.[14] To test the dependence of Skp1 glycosylation on Tg*gat1*, the gene was disrupted using CRISPR/Cas9 in the RH strain, yielding *gat1*Δ as described in Fig. S2.4A. PCR studies confirmed replacement with a *dhfr* cassette, and enzyme assays (see below) showed a loss of enzyme activity. The resulting clone produced a version of Skp1 in which the 5-sugar glycopeptide was no longer detectable, but ions corresponding to the 4-sugar glycopeptide were detected at similar abundance ***(Figure***

2.1C). Similar results were obtained in the type 2 ME49 strain ***(Figure S2.4A, 2.1C)***, and in strain RHΔΔ in which *gat1* was disrupted by homologous recombination using a different selection marker, referred to as Δ*gat1-1* ***(Table 2.1; Figure S2.3B; not shown)***. The similar results obtained by different genetic methods in distinct genetic backgrounds indicate that Tg*gat1* is required for addition of the terminal sugar, but whether this was a direct or indirect effect was unclear.

Toxoplasma growth depends partially on gat1

Parasites require invasion of mammalian host cells to establish a niche within an intracellular parasitophorous vacuole in order to proliferate.[24,25] In two-dimensional fibroblast monolayers, parasites lyse out and invade neighboring cells to repeat the cycle, resulting in plaques whose area is a measure of efficiency of these cellular processes. Past studies showed that plaque size growth is compromised by mutational blockade of Skp1 hydroxylation and earlier steps of the glycosylation pathway.[12-14] Similarly, disruption of *gat1* also resulted in modestly smaller plaques in both RH or RHΔΔ backgrounds (Fig. 2.2A,B), and in RHΔΔ in which *gat1* was disrupted using homologous recombination without CRISPR-Cas9 ***(Figure 2.2C)***. To determine whether the effects were specific to the genetic lesion at the *gat1* locus, the RH and RHΔΔ KO strains generated using CRISPR/Cas9 were modified again by CRISPR/Cas9 to introduce single copies of epitope tagged versions of the *gat1* coding locus, downstream of an endogenous *gat1* promoter cassette or a tubulin promoter cassette, respectively, into the *uprt* locus. The expected insertions were confirmed using PCR ***(Figures S2.3C, S2.4B)***. As a result, TgGat1-3xHA could be detected by Western blotting of tachyzoites in the RHΔΔ background ***(Figure S2.4D)*** and, as discussed below, the complemented RHΔΔ strain restored Skp1

glycosylation according to a biochemical complementation test. Although TgGat1-Ty expressed under its own promoter cassette in the RH background was not detected (not shown), enzyme activity was partially restored in the TgGat1-Ty strain ***(Figure S2.4C)***. Both strains exhibited larger plaque sizes than their respective KO parents ***(Figures 2.2A, B)***, confirming that the effect on growth was due to the original loss of Gat1.

Gat1 is closely related to glycogenin

An evolutionary analysis was conducted to gain further insight into the function of TgGat1. Based on searches of genomic databases using BLASTP, TgGat1 belongs to the CAZy GT8 family. The top-scoring hits, with Expect values of $<10^{-32}$, were found only in protists that contain *Toxoplasma* PgtA-like and Glt1-like sequences ***(Figure S2.7)*** and lack *D. discoideum* AgtA-like sequences, suggesting a common function. The most similar sequences, in searches seeded with the putative catalytic domain, belong to glycogenin, with Expect values of $\geq E^{-27}$. All other homologous sequences had Expect values of $\geq 10^{-22}$. Glycogenin is a dimeric α4-glucosyltransferase that may prime the synthesis of glycogen in the cytoplasm of yeast and animals by a mechanism that involves auto-glucosylation. Glycogenin appears not to be involved in starch formation and is an evolutionarily recent addition to the glycogen biosynthesis pathways that occur in the absence of glycogenin activity in bacteria and many unicellular eukaryotes.[26] The cyst-forming stage of *T. gondii* accumulates crystalline amylopectin[27], an α1,4Glc polymer with α1,6-linked branches that resembles glycogen, in its cytoplasmic compartment. *T. gondii* amylopectin is assembled by a UDP-Glc based metabolism that is related to the floridean starch of the red alga *Cyanidioschyzon merolae* and, to a lesser extent, to that of glycogen storing animals and fungi. Homologs of glycogen synthase[27] and glycogen phosphorylase[28]

regulate the accumulation of amylopectin in this parasite, and TgGat1 has been annotated as a glycogenin.[28,29] Related genes have been implicated in promoting starch formation in red algae.[30,31] Thus Gat1 might have a function in regulating *T. gondii* amylopectin formation.

Glycogenin consists of a CAZy GT8 family catalytic domain plus a C-terminal glycogen synthase binding domain separated by a linker, whereas Gat1 consists only of a single catalytic domain ***(Figure 2.3A)***. TgGat1 is predicted to be a 345-amino acid protein encoded by a single exon gene in the Type I GT1 strain (TGGT1_310400). It is 34% identical to rabbit (*Oryctolagus cuniculus*) glycogenin over the catalytic domain, but includes a poorly conserved 90-amino acid sequence that interrupts the catalytic domain ***(Figure 2.3A).*** This region is likely unstructured based on secondary structure prediction by the XtalPred server[32]. The Gat1-like sequence from *P. ultimum* (Uniprot K3WC47) lacks this sequence, so was analyzed for comparison. PuGat1 is predicted to be a 266-amino acid protein encoded by a 2-exon gene, annotated as PYU1_G002535-201 (Transcript ID PYU1_T002538).

To further evaluate the evolutionary relationship of these putative GTs, their catalytic domains and those of the most closely related or known sequences from the CAZy GT8 family ***(Figure S7)*** were aligned ***(Figure S2.8)*** and analyzed by a Maximum Likelihood method ***(Figure 2.3B)***. The results suggest that Gat1 and glycogenin evolved separately from a common ancestor. Though the last common ancestor was not resolved, Gat1 was potentially the predecessor to glycogenin owing to its presence in other primitive unicellular eukaryotes which bear no evidence of glycogenin-like sequences, and because there is currently no evidence for the existence of close homologs of Gat1 and glycogenin

within the same clade that would suggest that they evolved as paralogs of a gene duplication. However, the possibility that either one or the other product of an ancestral gene duplication was always lost in every extant derivative cannot be excluded. Gat1 and glycogenin each possess unique conserved sequence motifs ***(Figure S2.6)*** that potentially support functional differences.

Gat1 is a terminal Skp1 α-galactosyltransferase

To determine if TgGat1 can directly modify Skp1, the predicted full-length protein was expressed as a His_6-tagged conjugate in *E. coli*, purified on a TALON resin, and treated with TEV protease leaving an N-terminal GlyHis-dipeptide stub before the start Met ***(Figure 2.4A)***. The presumptive ortholog from *Pythium ultimum* was prepared similarly. A screen for UDP-sugar hydrolysis activity of TgGat1 yielded, after extended reaction times, only UDP-Gal and UDP-Glc as candidate substrates from a panel of six common UDP-sugars ***(Figure S2.9A)***. A quantitative comparison showed approximately 7-fold greater activity toward UDP-Gal than UDP-Glc ***(Figure 2.4B)***.

The ability of Gat1 to transfer Gal or Glc to another sugar, rather than water, was tested using a substrate analog for glycogenin, Glcα1,4Glcα1-pNP (maltose-pNP), which mimics the terminal disaccharide of glycogen and starch, and has a terminal αGlc as found on the Skp1 tetrasaccharide. Although Gat1 from either *T. gondii* or *P. ultimum* could modify maltose-pNP using either sugar nucleotide ***(Figure 2.4C),*** the enzymes strongly preferred UDP-Gal ***(Figure 2.4D)***. Furthermore, TgGat1 activity was not saturated by UDP-Glc at 0.5 mM ***(Figure 2.4D)***, whereas TgGat1 and PuGat1 exhibited apparent K_m values for UDP-Gal in the range of 30-15 μM ***(Figure S2.9F)*** These values were greater than the 2-4 μM values reported for rabbit and yeast glycogenins for UDP-Glc.[33,34] TgGat1

and PuGat1 exhibited higher apparent K_m values for maltose-pNP in the range of 16-43 mM ***(Figure 2.4D)***, which were greater than the 4 mM value reported for rabbit glycogenin[35].

Extending the acceptor to 3 sugars or decreasing it to one resulted in less activity, but either anomer of Glc-pNP was acceptable ***(Figure 2.4F)***. A similar pattern was observed for the *P. ultimum* and *T. gondii* enzymes. The enzymes were specific for terminal Glc acceptors as activity towards Gal-pNP was not detected. In comparison, GlFGaGn-pNP, which mimics the natural acceptor on Skp1, was a superior acceptor substrate ***(Figure 2.4F)*** with a K_m of 1.5 mM ***(Figure S2.9G)***, and the truncated trisaccharide FGaGn-pNP was inactive indicating that Glc was the position of attachment. Thus the Gat1 enzymes preferred their native tetrasaccharide acceptor substrate and UDP-Gal as a donor, but tolerated, with low efficiency, the preferred substrates of glycogenin, UDP-Glc and α4-linked oligomers Glc.

The importance of Skp1 as context for the acceptor glycan was examined using GlFGaGn-Skp1, which was prepared by reaction of FGaGn-Skp1 with UDP-Glc and Glt1 resulting in loss of the trisaccharide epitope ***(Figure 2.4H inset)***. In a comparison of acceptor concentration dependence, TgGat1 was about 33x more active toward GlFGaGn-Skp1 than free GlFGaGn-pNP ***(Figures 2.4H,I)***, based on ~33x less activity (dpm incorporated) at 0.001x the substrate concentration (substrate concentrations were both in the linear response range). The reaction with GlFGaGn-Skp1 did not approach saturation at the highest concentration tested, 6 μM. Thus the TgGat1 reaction was much more efficient when the tetrasaccharide was associated with its native substrate Skp1, and thus consistent with Gat1 being directly responsible for modifying Skp1 in the cell.

A characteristic of glycogenin is its ability to modify the HO-group of a Tyr side chain near its active site with αGlc, and then to repeatedly modify the 4-position of the Glc with another αGlc, and repeat the process up to 8-12 sugars. Thus, when isolated as a recombinant protein expressed in UDP-Glc positive *E. coli*, glycogenin is partly glucosylated[36]. However, neither TgGat1 nor PuGat1 prepared in a similar manner were found to be glycosylated, based on an exact mass measurement using nLC/MS ***(Figures S2.10A-E)***. Furthermore, following incubation with either UDP-Gal or UDP-Glc, no change in SDS-PAGE mobility ***(Figure S2.10D)*** or exact mass were observed ***(Figure S2.10E)***. Thus, no evidence for autoglycosylation activity of Gat1 from either species could be detected.

Skp1 is the only detectable substrate of Gat1 in parasite extracts

GlFGaGn-Skp1 is a substrate for Gat1, but are there others? This was addressed by complementing extracts of *gat1*Δ parasites with recombinant TgGat1 in the presence of UDP-[^{3}H]Gal, and measuring incorporation of ^{3}H after display of the proteome on a 1D SDS-PAGE gel. A high level of incorporation of ^{3}H that depended on the addition of enzyme was observed at the position of Skp1 ***(Figure 2.4J)***, as expected, but negligible dpm were detected elsewhere in the gel. Furthermore, negligible dpm were incorporated into Skp1 in RH parental cells, indicating that little GlFGaGn-Skp1 accumulates in wild-type cells. Similar results were observed in studies of *gat1*Δ clones in the RHΔΔ and Me49 backgrounds ***(Figures S2.9H, I)***. Finally, complementation of the *gat1*Δ clone in RHΔΔ with Gat1 expressed under the tubulin promoter resulted in absence of measurable incorporation into Skp1, confirming specificity for Gat1 expression *per se*.

Although Gat1 was unable to serve as its own GT acceptor in the manner of

glycogenin, its ability to modify α4Glc oligomers *in vitro*, albeit with low efficiency, raised the possibility that it may affect amylopectin (starch) biosynthesis in cells by applying αGal residues to its non-reducing termini. Starch normally accumulates to substantial levels in bradyzoites, a slow growing form of the parasite that is induced by stress. Since induction of bradyzoite differentiation in cell culture is more efficient in type 2 strains, Me49 and *gat1*Δ/Me49 cells were induced by pH up-shift and examined for differentiation by labeling of bradyzoite cyst walls with FITC-DBA-lectin, and for starch with the Periodic acid/Schiff's base reagent. As shown in Fig. 2.5, ME49 bradyzoites accumulated substantial levels of starch relative to tachyzoites, and no difference in the pattern or level was ascertained in *gat1*Δ cells using this qualitative assessment. Thus Gat1 does not appear to affect starch synthesis, which is consistent with the absence of terminal Gal residues in a sugar composition analyses of *T. gondii* starch[29,37], and the finding that *gat1* is expressed equally in starch-poor tachyzoites and starch-rich bradyzoites based on transcript analysis[29].

Gat1 generates a Galα1,3Glc- linkage

To determine the glycosidic linkage of the αGal residue transferred by TgGat1, the previously prepared ($^{13}C_6$)GlFGaGn-pNP[14] was modified with TgGat1 using UDP-[U-^{13}C]Gal as the donor substrate. The pentasaccharide reaction product (approximately 30% conversion) was analyzed together with the tetrasaccharide starting material by NMR. The previous assignment of the chemical shifts of the tetrasaccharide[14] facilitated provisional assignment of the additional terminal Gal chemical shifts using the *CASPER* program[38] and confirmed by analysis of the 2D COSY, TOCSY and HMBC spectra ***(Figure 2.6)***. One-dimensional ^{1}H-NMR spectra reveals the presence of the mixture of tetra- and

pentasaccharides, followed by a downfield shift in the ^{13}C-Glc-H1 peaks upon linkage to the terminal ^{13}C-Gal ***(Figure 2.6A).*** The HMBC ^{1}H –^{13}C correlation spectrum shows the ***(Figure 2.6B, center panel)*** connection from Gal-H1 to Glc-H3, consistent with the downfield peak in the ^{1}H –^{13}C-HSQC spectrum ***(Figure 2.6B, top panel)***, and the proton resonances in the HSQC-TOCSY ***(Figure 2.6B, bottom panel)***, establishing the glycosidic linkage between the terminal αGal and underlying αGlc as 1→3. Finally, a ^{1}H –^{1}H-COSY spectrum confirms the assignments of the underlying Glc-H1, H2 and H3 ***(Figure 2.6C)***. Consistent results were obtained when the tetrasaccharide was modified in the presence of PuGat1 (data not shown). Taken together, our NMR analyses are most consistent with the glycan structure: Galα1,3Glcα1,3Fucα1,2Galβ1,3GlcNAcα-, indicating that TgGat1 is a retaining UDP-Gal:glucoside α1,3-galactosyltransferase. Not only does Gat1 transfer the α-anomer of a different sugar compared to glycogenin, it attaches it to a different position (4- not 3-) of the acceptor αGlc residue.

Comparison of crystal structures explains the catalytic differences between Gat1 and glycogenin

To further probe the relationship between Gat1 and glycogenin, we compared their structures by X-ray crystallography. Attempts to crystallize TgGat1 were unsuccessful, even after deletion of its unconserved insert ***(Figure 2.3A)***. However, PuGat1, which lacks this insert, was co-crystallized in the presence of Mn^{2+} and UDP.

The crystal structure of PuGat1 in complex with Mn^{2+} and UDP was solved using single-wavelength anomalous dispersion phasing of a Pt^{2+} derivative, and the resolution was extended to 1.76 Å using a native data set ***(Table S2.3)***. The asymmetric unit contains a single chain of PuGat1 with unambiguous electron density for the nucleotide and Mn^{2+}

ion ***(Figure 2.7A)***. The overall structure of PuGat1 reveals a canonical GT-A fold[39] consisting of eight α-helices and eight β-strands. The N-terminus (residues 1-8) and two loops (residues 80-96 and 242-244) are disordered and were not modeled. The structure is similar to glycogenin-1 from *Oryctolagus cuniculus* (Oc-glycogenin-1), which superimposes 213 corresponding Cα atoms with an RMSD of 3.3 Å despite a sequence identity of only 34% ***(Figure 2.7B)***. The application of crystallographic symmetry shows that PuGat1 forms the same dimer described40 for the Oc-glycogenin-1 structure ***(Figure 2.7B)***. According to PISA[41] analysis, the PuGat1 dimer interface buries 1090 Å^2 with a favorable P-value of 0.107, which suggests that the dimer contact is stable. Sedimentation velocity analysis of 3.5 μM PuGat1 reveals a *c(s)* distribution consisting of single species at 4.0 S, which corresponds to the predicted value of 4.2 S for a dimer ***(Figure 2.7C)***. The slightly slower sedimentation indicates that the enzyme in solution is less compact than that observed in the crystal structure. PuGat1 was dimeric even at 0.3 μM ***(Figure S2.12)***, suggesting that it forms a dimer with an affinity >2-fold higher than that of Oc-glycogenin-1 (which has a reported K_d of 0.85 μM)[42]. Based on gel filtration and preliminary sedimentation velocity experiments (not shown), TgGat1 is also a dimer.

The PuGat1 active site shows that the conserved DxD motif[43] and a conserved His residue coordinate the Mn^{2+} ion using the Oδ2 atom of D117, both Oδ1 and Oδ2 atoms of D119, and Nε2 atom of His231 ***(Figure 2.7D)***. The Mn^{2+} ion is also coordinated by the oxygen atoms from the α and β phosphates of UDP. Comparing the PuGat1 and Oc-glycogenin-1 active sites shows that all of the interactions with the nucleotide are conserved, with the exception of the interactions with N3 and O4 of the uracil ring ***(Figure S2.13)***. Other changes in Gat1 active site include a Leu to Ser substitution at residue 233,

which can potentially remove a packing interaction with the donor sugar in PuGat1 ***(Figure 2.8)***. The hydroxyl of the substituted Ser233 forms a hydrogen bond with the adjacent side chain of Gln206. A water molecule (W509) replaces the Leu side chain and forms a hydrogen bond with Asn149.

Comparing the crystal structure of UDP-Glc bound Oc-glycogenin-1 with PuGat1 immediately suggests a reason for why these enzymes have different donor specificities (Fig. 2.8). Superimposing the Oc-glycogenin-1:UDP-Glc structure onto PuGat1 shows that UDP-Glc would displace water499 coordinated by Thr180 ***(Figure 2.8A)***. This would leave the Thr180 hydroxyl group unsatisfied, and the unfavorable burial of a polar group likely explains why UDP-Glc is a poor donor ***(Figure 2.8A)***. In contrast, we modelled in UDP-Gal by flipping the stereochemistry at C4 position. The O4 atom of Gal would be ideally positioned to satisfy the Thr180 hydroxyl group. Concurrently, Asp176, whose Cα atom underwent a 2.3 Å shift relative to its location in glycogenin, would be in position to receive a hydrogen bond from the O4 atom of the Gal ***(Figure 2.8B)***. Altogether, Gat1's sugar donor preference for UDP-Gal is likely due to formation of favorable hydrogen bonds with O4 atom of Gal in contrast to the burial of Thr180 hydroxyl group when binding UDP-Glc.

Computational modeling predicts the specificity of Gat1 towards the Skp1 tetrasaccharide

To address the basis of Gat1's preference for the GlFGaGn- glycan, the lowest energy conformation of the reducing form of the glycan generated by GLYCAM-web (www.glycam.org) was docked using AutoDock Vina. A plausible docking mode was selected based on the requirement that the C'3-hydroxyl group must be oriented towards

the anomeric carbon of the donor sugar to serve as the nucleophile for addition of the Gal, and that the glycan does not clash with the other subunit of the dimer. Out of 100 docking simulations, only the top scoring pose with a binding energy score of -5.7 kcal/mol satisfied the selection requirement. In this pose, the glycan adopted an alignment in a groove formed by Gat1 dimerization ***(Figsures 2.9A, B)***. The glycan is stabilized by hydrogen bond contributions from the sidechains or peptide backbones of residues D141, F143, and S234 from subunit A, and residues L212, K216, N217, and Y220 from subunit B ***(Figure 2.9C)***. In addition, non-polar interactions against the faces of the sugar moieties are provided by residues T208, L212, F143, and W221 from subunit A, and residue Y220 from subunit B ***(Figure 2.9D)***. F143, Y220, W221, and S234 are uniquely conserved in Gat1 proteins relative to glycogenins ***(Figure S2.6)***. The packing interactions from conserved hydrophobic residues are likely the major contributors in terms of binding energy. The extensive electrostatic and packing complementarity, which could not be achieved using the same approach using the α4Glc-tertramer recognized by glycogenin, can explain the distinct preference of Gat1 for the Skp1 tetrasaccharide acceptor substrate.

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

The glycan-protein contacts described previously for *D. discoideum* Skp1 correlated with the extension of helix-8, which was interpreted to provide better access by FBPs. To address whether this mechanism is conserved for *T. gondii*, despite the difference of the fourth sugar, the computational studies were repeated on glycosylated TgSkp1. Energy-minimized structures of the glycans from DdSkp1 and TgSkp1 revealed only a difference in the position of O4 of the fourth sugar (due to the Glc/Gal configurational inversion) ***(Figure S2.14A)***. Six all-atom MD simulations were performed without

coordinate constraints for 250 ns each to allow greater sampling of conformational space. Three of the simulations began with a 50-ns pre-equilibration of the glycan with Cα atom constraints on the polypeptide while the others proceeded directly. As before, the simulations did not converge on a common structure, so a combined time-resolved linear regression analysis of the six simulations was conducted to identify correlations between helix extension and calculated polar and nonpolar interaction energies with each moiety of the glycan, and the results are shown in Table 2.2.

A representative frame from a simulation with a strong correlation between the extension of helix-8 (dotted green line) and glycan contacts with the polypeptide chain is shown in Figs. 2.10A-C. Glycan-protein contacts involve both polar hydrogen bonds and non-polar van der Waals interactions between sugar and amino acid residues. The strongest correlated polar interactions ***(Table 2.2)*** correspond to three hydrogen bonds between sugars and amino acids within the unstructured peptide chain located between helix 7 and 8 (residues 147-152), as depicted in Figs. 10B, 9C. Notably, the 4-OH of αGlc (the fourth sugar), which is epimeric to the *D. discoideum* fourth sugar, is oriented toward the solvent thus not affecting protein contacts. Nonpolar interactions also occur within this region, with the terminal αGal moiety showing the highest correlation ***(Table 2.2)***. The hydrophobic face of the terminal αGal can pack against a nonpolar pocket consisting of planar faces of peptide backbone ***(Figsure 2.10D, 2.9E)***. Specifically, its C3-C5 hydrogens can pack with the backbone of residues 147-149, and a hydrogen on C6 can be buried within residues 143 and 144 ***(Figure 2.10D)***. Both the polar and nonpolar interactions identified here involve amino acids that are conserved across both organisms, with the exception of V149 which is Lys in DdSkp1 ***(Figure S2.14D)***. This substitution would not be expected to affect Gal

burial since the majority of interactions are with the protein backbone at this position (not shown). The previously noted hydrogen bonds-1 and -3 to helix-7, involving E140 (E129 in DdSkp1), were also observed in this study ***(Figure S2.14C)***, but were poorly correlated (R^2=0.0, vs. 0.63 average for the other three) with helix-8 extension ***(Figure S2.14E)***. The energetic analyses applied to the *T. gondii* variant confirm its ability to form an organized structure that can still similarly influence local Skp1 polypeptide organization. The new studies emphasize the interaction of the glycan with the interhelix region rather than helix-7 and contributions of packing of the outer sugars including the terminal αGal. As previously hypothesized, these effects on the conformational ensemble of TgSkp1 have the potential to make it more receptive to FBP binding.

2.5 Discussion

Glycosylation mediates much of the effect of O_2-dependent prolyl hydroxylation of Skp1 in *D. discoideum* and *T. gondii*, and the present work emphasizes the importance of the fifth and final sugar of the glycan. The *T. gondii* glycan adopts a similar constrained conformation as described for *D. discoideum* despite the sequence differences ***(Figures S2.14A, B)***, and new all-atoms molecular dynamics simulations showed a conserved *cis*-interaction with the back side of the intrinsically disordered region of Skp1 that comprises subsite-2 of the FBP binding pocket ***(Figure S2.14C)***. A deeper time-resolved energetics analysis of polar and non-polar contacts showed a strong dynamic correlation between burial of the peripheral (especially the fifth) sugars in the turn from helix-7 to the loop connecting helix-8, and the extension of helix-8, which exposes subsite-2 for FBP docking ***(Figure 2.10; Table 2.2)***. This analysis de-emphasized the previous correlation of

hydrogen bonds with helix-7 in favor of contacts with loop residues between helix-7 and helix-8 ***(Figure S2.14E)***. This quantitative analysis of the trajectories extends critical features of the prior study in *D. discoideum*[7] by emphasizing the role of non-polar packing of the glycan terminus in helix-8 extension.

The fifth sugar is attached by Gat1, an unusual GT that resides in the cytoplasmic compartment rather than the secretory pathway of the cell and appears to be dedicated to only Skp1. While confined to the protist kingdom, it is nevertheless descended from a widely distributed lineage of sugar nucleotide-dependent CAZy GT8 GTs, and may be the evolutionary predecessor of glycogenin, a GT that modulates glycogen formation in the cytoplasm of yeast, fungi and animals. Though Gat1s from *T. gondii* and *P. ultimum* catalyze the transfer of αGal from UDP-Gal to non-reducing terminal Glc acceptors ***(Figures 2.4, S2.9)***, their ability to utilize UDP-Glc at low efficiency could be rationalized by the organization of the active site in the crystal structure ***(Figure 2.8)*** and anticipates this evolutionary transition. The ability of UDP-Glc to inhibit Gat1's galactosyltransferase activity ***(Figure 2.4G)*** might be regulatory in cells. The recently reported capability of glycogenin to transfer Gal at certain steps[44] lends further support to the evolutionary relationship. Though Gat1 is able to modify Glc in multiple contexts, it prefers Glc at the terminus of the native Skp1 tetrasaccharide over Glc in a native starch- or glycogen-like glycan ***(Figures 2.4, S2.9)***. Furthermore, Gat1 was substantially more reactive when the glycan was attached to Skp1 ***(Figures 2.4H, I)***, indicating that the apoprotein contributes to increased activity. The increase in activity may be due to the recognition of Skp1 protein by Gat1, but this hypothesis remains to be tested.

Computational docking showed how the Skp1 glycan alone in its calculated lowest

energy state can naturally fit within a groove along the dimer interface of PuGat1 ***(Figure 2.9A)***, in a manner that cannot be achieved with an α4-glucan as in starch or glycogen in its lowest energy state (not shown). The orientation of the acceptor Glc of the Skp1 glycan in this docking mode also supports the formation of the α1,3-linkage that was determined by NMR analysis of the product of the reaction of both PuGat1 and TgGat1 with a synthetic version of the Skp1 tetrasaccharide as an acceptor ***(Figure 2.6)***. Furthermore, a high degree of conservation between the two enzymes is supported by their very similar activity and specificity characteristics ***(Figures 2.4, S2.9)***. The described binding mode is a possible explanation for the greater activity of the free tetrasaccharide as an acceptor substrate compared to the α4-linked glucans. Overall, the data suggest that recognition of Skp1 as a substrate includes both active site preference for the specific glycan and separate determinants on the polypeptide. Owing to the limited computational scope of this study, however, further analysis is needed to fully define the basis for Skp1 recognition.

The significance of the terminal sugar is echoed by the slow growth phenotype of *gat1*Δ parasites in the monolayer plaque assay. This was attributed to the absence of Gat1 because similar results were obtained in independent knockouts in three different genetic backgrounds, and the defect was corrected by genetic complementation under its own promoter or a strong tubulin promoter in the *uprt* locus ***(Figure 2.2)***. Since no other Gat1 targets were detected by biochemical complementation of *gat1*Δ extracts ***(Figures 2.4J, S2.9H, I)***, and no effect on starch formation was observed ***(Figure 2.5)***, it is likely that failure to fully glycosylate Skp1 is responsible for the growth defect. This interpretation is consistent with similar effects of knocking out earlier GTs in the pathway and their selectivity for Skp1 *in vitro.*[13,14] Furthermore, in *D. discoideum*, genetic manipulations of

Skp1 levels interact with its glycosylation with respect to O_2-sensing.[8] Skp1 interactome studies in *D. discoideum* indicated that 5 sugars were clearly more effective in promoting binding of FBPs than 3 sugars[10], but were however unable to resolve the roles of the fourth and fifth sugars owing to their being added by the same GT.[45] Though we cannot completely exclude some other function of Gat1, we note that at 345 amino acids, nearly all of its sequence appears devoted to the enzymatic domain ***(Figure 2.3A)***. Studies are underway to identify SCF substrates and how they are affected by Skp1 modification and influence parasite growth and fitness.

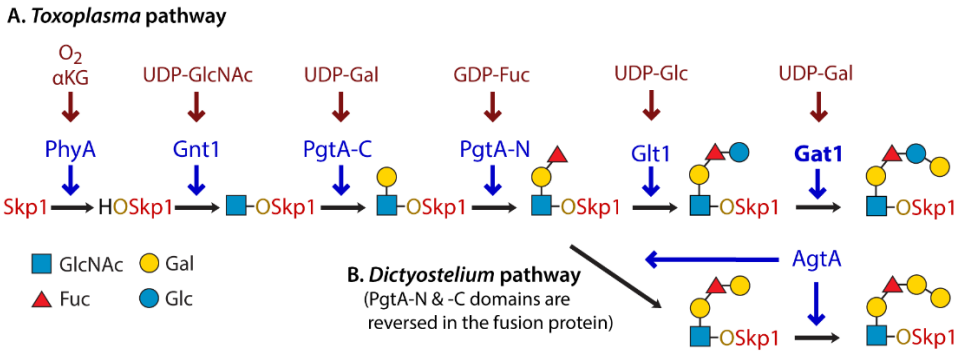
The importance of the fifth sugar is reinforced by the observation that the amoebozoa, which lack both Glt1 and Gat1 ***(Figure 2.1A, B)***, evolved a new unrelated enzyme, AgtA, to fulfill the role. AgtA is a dual function GT that applies both the 4th and 5th sugars: each an α3-linked Gal. This raises the interesting possibility that AgtA evolved to compensate for the unexplained loss of Glt1 and Gat1, with recursive addition of the same sugar being an accessible evolutionary pathway to recover the pentasaccharide. Gat1 sequences are found in alveolates (includes *T. gondii*), stramenopiles (includes *P. ultimum*), and rhizaria (together with telonemids are known as the TSAR group), and archaeplastids (together with the TSAR group are known as bikonts), but not in the unikonts that include the amoebozoa, fungi and animals.[46,47] While the fate of Glt1 in the unicots is unknown, its disappearance potentially freed Gat1 to evolve into glycogenin, an α4-glucosyltransferase that, *in vitro*, can autoglucosylate itself to prime glycogen synthesis in fungi and metazoa. The enzymes share catalytic properties (see above) and perfectly conserve structural similarities including the same dimer interface ***(Figures 2.7, 2.9)***, though further studies are needed to support the model of lineal descent. Thus it is

interesting to speculate whether the proposed transition conserved a cellular function. Though glycogenin acquired a C-terminal glycogen synthase binding domain[48] ***(Figure 2.3A)*** that seems consistent with its *in vitro* capacity to initiate glycogen assembly, data now exist that glycogenin is not required for glycogen synthesis in yeast and mice[49,50], though glycogen levels are affected by an unknown mechanism. The *trans* glycosyltransferase activity for this enzyme lineage in bikonts protists raises the possibility of an unanticipated activity in unicots that could help explain the glycogenin-KO findings in yeast and mice.

2.6 Acknowledgements

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

2.7 Figures and Tables



C. Glycosylation status of Skp1

relative abundance

	Source of peptides					
	RH	RH + αGalase	Δ*gat1*/ RH	*gat1*::*gat1*/ Δ*gat1*/RH	Me49	ΔGat1/ Me49
peptide–Hex5	0.12	0	0	0.08	0.08	0
peptide–Hex4	0*	0.14	0.09	0	0	0.09

* <0.0005

Figure 2.1. Gat1 is required for terminal α-galactosylation of Skp1 in parasites. (A, B) Schematic of Skp1 glycosylation pathway in *Toxoplasma*[14] and *Dictyostelium*[3], which modifies a single hydroxyproline associated with its F-box binding region, is depicted using CFG glycan symbols[19].The identity of Gat1 as the final enzyme in the *Toxoplasma* pathway and the nature of the final sugar are reported here. (C) Dependence of the terminal sugar on Gat1, and its characterization. Skp1 was immunoprecipitated from type 1 RH and type 2 ME49 strains and their genetic derivatives, and peptides generated by trypsinization were analyzed by nLC-MS. In addition, Skp1 from RH was treated with green coffee bean α-galactosidase after trypsinization. The values represent the levels of pentasaccharide-peptide and tetrasaccharide-peptide levels detected, after normalization to all detected modification states of the peptide; note that the values have only partial relative meaning because of the low and varied detection efficiency of glycopeptides. As indicated, the detection threshold was <0.0005. The open circle for the terminal sugar of the pentasaccharide indicates that it was only known as a hexose at the time of the experiment. See Table S2.2 and Fig. S2.5 for primary data. Similar results were obtained in independent samples from RH and Δ*gat1*/RH (not shown).

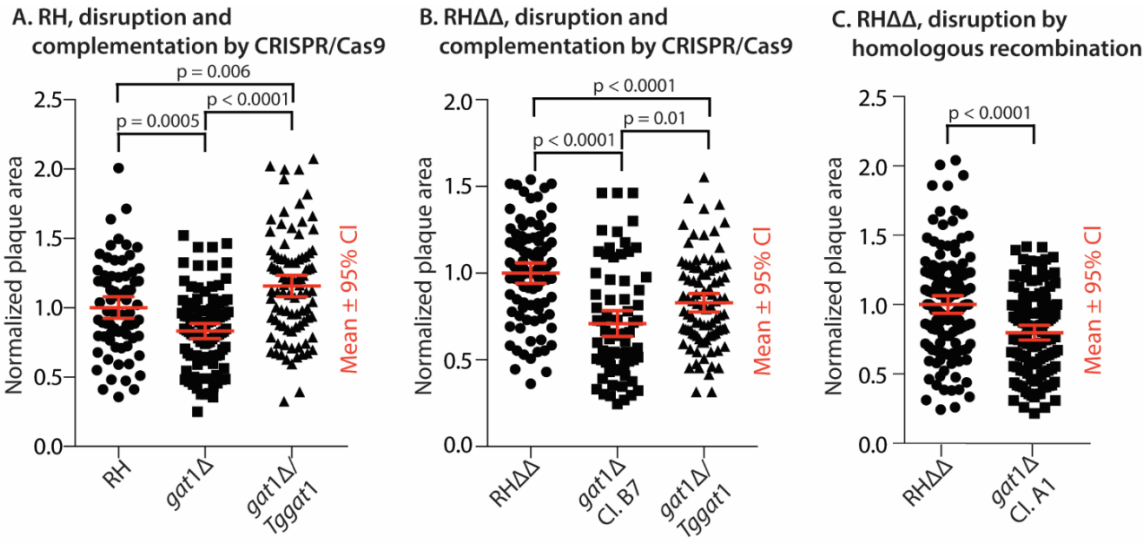


Figure 2.2. Parasite growth depends on Gat1. Parasites were plated at clonal density on two-dimensional monolayers of human foreskin fibroblasts (HFFs), and allowed to invade, proliferate, lyse out, and reinfect neighboring fibroblasts. After 5.5 d, cultures were fixed, stained, and analyzed for the areas occupied by lysed fibroblasts (plaques). Data from each of 3 independent trials, which were each normalized to the parental strain, were merged for presentation. (A) Comparison of the type 1 RH strain before and after *gat1* replacement using CRISPR/Cas9, and complementation with *gat1* under control of its own promoter cassette in the uprt locus. (B) Comparison of RHΔΔ, *gat1-2*Δ/RHΔΔ, and the latter complemented with gat1 under control of a tubulin promoter in the uprt locus. (C) Comparison of *gat1-1*Δ/RHΔΔ, prepared by homologous recombination. Significance of differences in plaque areas between parasite strains was assessed by Student's *t*-test.

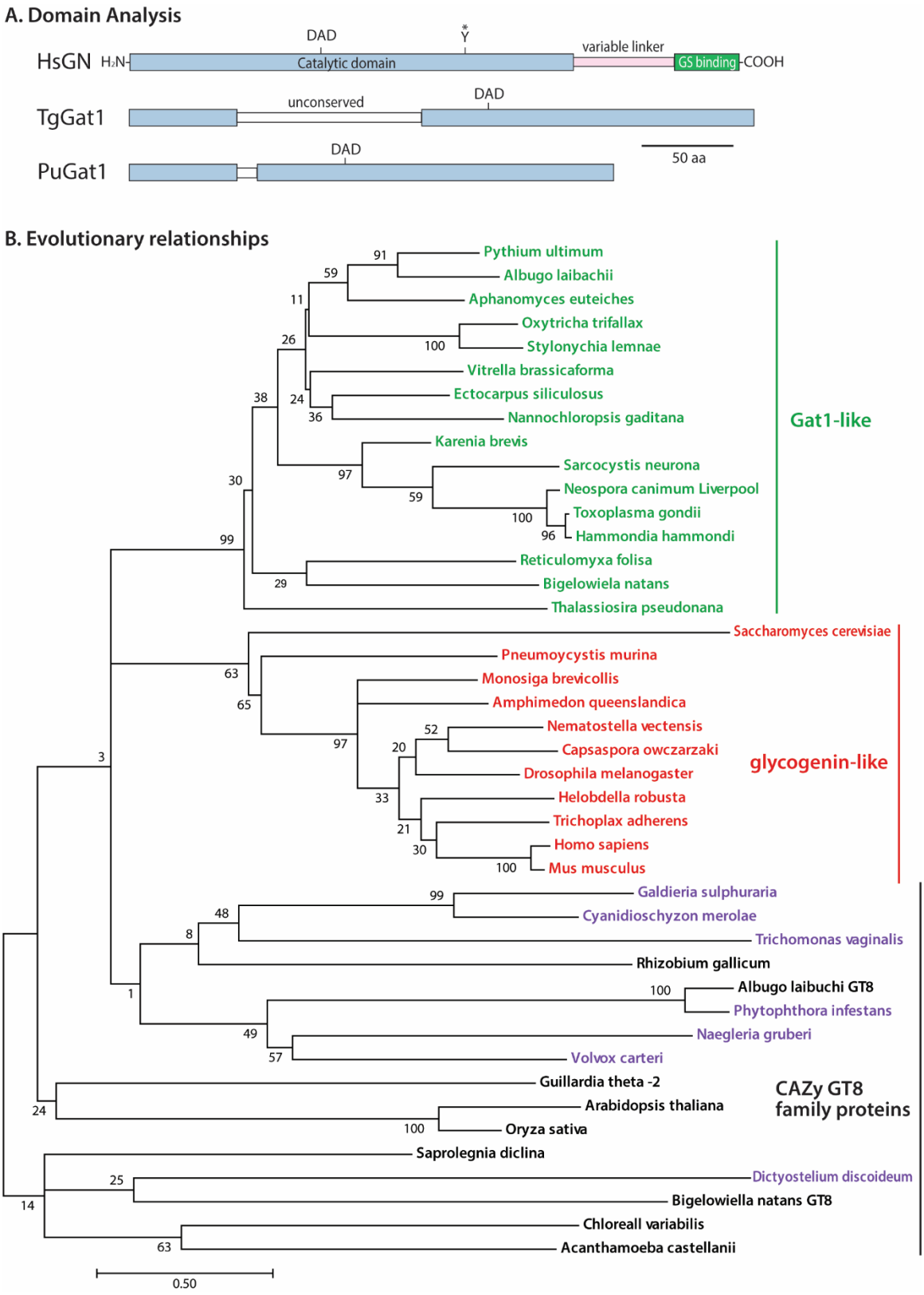


Figure 2.3. Gat1 is closely related to glycogenin sequences. (A) Domain analysis of Gat1 from *T. gondii* and *P. ultimum* in comparison with human glycogenin-1. (B) The

evolutionary history of the sequence of the Gat1 catalytic domain was inferred by using a Maximum Likelihood method. The tree with the highest log likelihood (-13279.56) is shown. Gat1 and Gat1-like sequences are colored green, glycogenin and glycogenin-like sequences are in red, and characterized and other selected other CAZy GT8 sequences are in black, or purple if predicted to reside in the secretory pathway rather than the cytoplasm. The percentage of trees in which the associated taxa clustered together is shown at each branch. Branch lengths are measured by the number of substitutions per site. See Figs. S2.6-S2.8 for alignments.

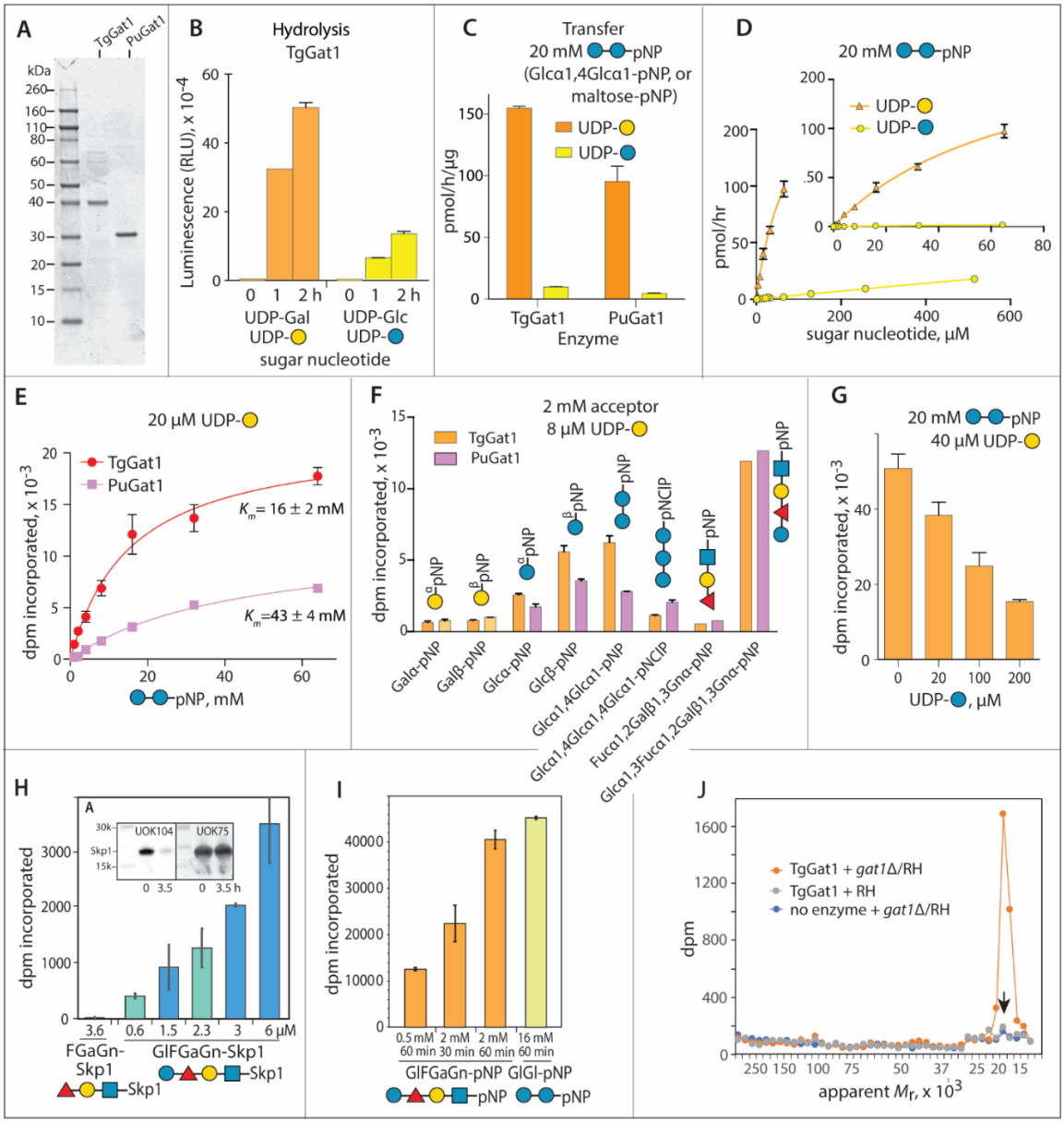


Figure 2.4. Gat1 preferentially galactosylates the Skp1 glycan *in vitro*. (A) Recombinantly expressed and purified preparations of TgGat1 and PuGat1 were analyzed by SDS-PAGE and staining with Coomassie blue. (B) Temporal dependence of UDP-Gal and UDP-Glc hydrolysis. The averages and standard deviations of 3 technical replicates are shown. A similar profile was observed with a different enzyme concentration. See Fig. S9E for a trial with higher enzyme concentrations. (C) Transferase activity utilizing 8 µM UDP-Gal or UDP-Glc toward 20 mM Glcα1,4Glcα-pNP (maltose-pNP) for TgGat1 and PuGat1. The averages and standard deviations of two technical replicates are shown; similar profiles were in 2 independent assays with a different TgGat1 preparation. (D) UDP-Gal and UDP-Glc concentration dependence of TgGat1 transferase activity toward 20 mM maltose-pNP. The averages and standard deviations of two technical replicates are shown, and an independent trial with TgGat1 and PuGat1 against UDP-Gal is shown in Fig S2.9F. (E) Maltose-pNP concentration dependence of TgGat1 and PuGat1 transferase

activity from 20 μM UDP-Gal. The averages and standard deviations of two technical replicates are shown. (F) Relative Gal-transferase activity of TgGat1 and PuGat1 toward different acceptors. The averages and standard deviations of three technical replicates are shown. Similar results were obtained in three independent trials. (G) Effect of UDP-Glc concentration on the Gal-transferase activity of TgGat1. Reactions were incubated for 1 h. The averages and standard deviations of two technical replicates are shown. (H) Gal-transferase activity of TgGaT1 toward varied concentrations of GlFGaGn-Skp1, in the presence of 40 μM UDP-Gal (1 μCi) after 1 h incubation. Data from independent preparations of TgSkp1 are colored in different shades. FGaGn-Skp1 is included for comparison. Error bars represent S.D. of duplicate measurements. Inset shows Western blots of the Skp1 preparations used, where FGaGn-Skp1, which is recognized specifically by pAb UOK104, is largely converted in a 3.5-h reaction using Glt1 and UDP-Glc to GlFGaGn-Skp1, which is recognized only by the pan-specific pAb UOK75. (I) Reactions with synthetic oligosaccharides conjugated to pNP were conducted in parallel using the same conditions. (J) Biochemical complementation to detect Gat1 substrates. Desalted S100 extracts of RH and *gat1*Δ/RH were reacted with recombinant Gat1 in the presence of UDP-[^{3}H]Gal, and the product of the reaction was separated on an SDS-PAGE gel which was sliced into 40 bands for liquid scintillation counting. The migration position of Skp1 is marked with an arrow. See Figs. S2.9H and S2.9I for trials using different strains.

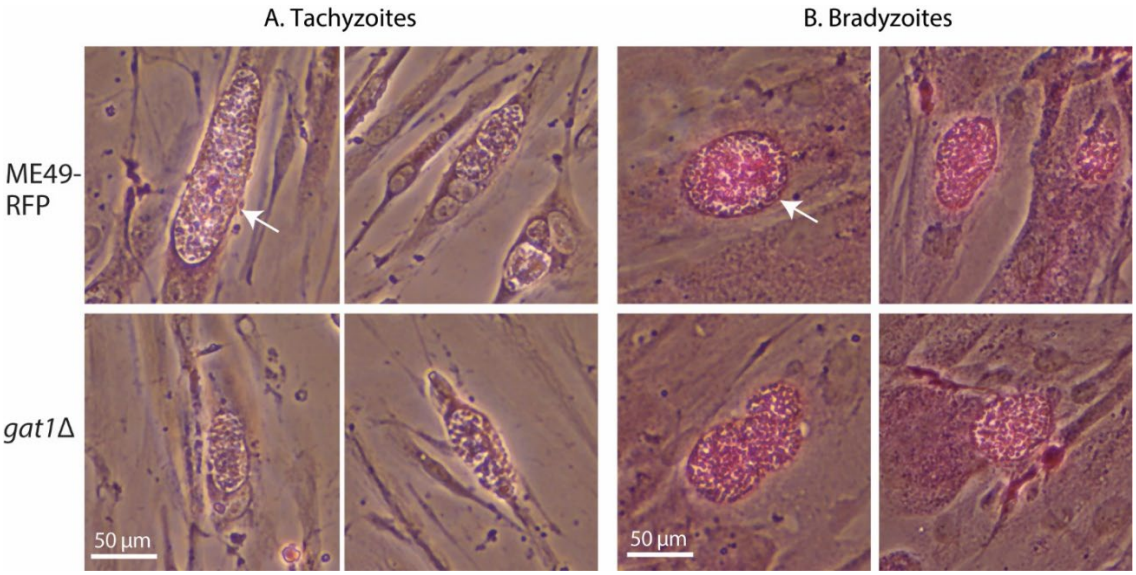


Figure 2.5. Starch appears unaffected in *gat1*Δ parasites. To promote normal starch accumulation, rapidly proliferating tachyzoites (panel A) of the type II strain Me49 (RFP expressing) and its *gat1*Δ derivative were induced to differentiate as slow-proliferating bradyzoite cysts (panel B) in human foreskin fibroblasts. Cultures were fixed and stained with Periodic acid/Schiff's base to reveal starch as a purple adduct. Arrow indicates a parasitophorous vacuole containing dozens to hundreds of tachyzoites within a fibroblast. White arrowhead indicates a cyst containing dozens of slow-growing bradyzoites, as confirmed by labeling of the cyst wall with FITC-DBA lectin (not shown). Scale bar = 50 µm. Two independent trials yielded similar results.

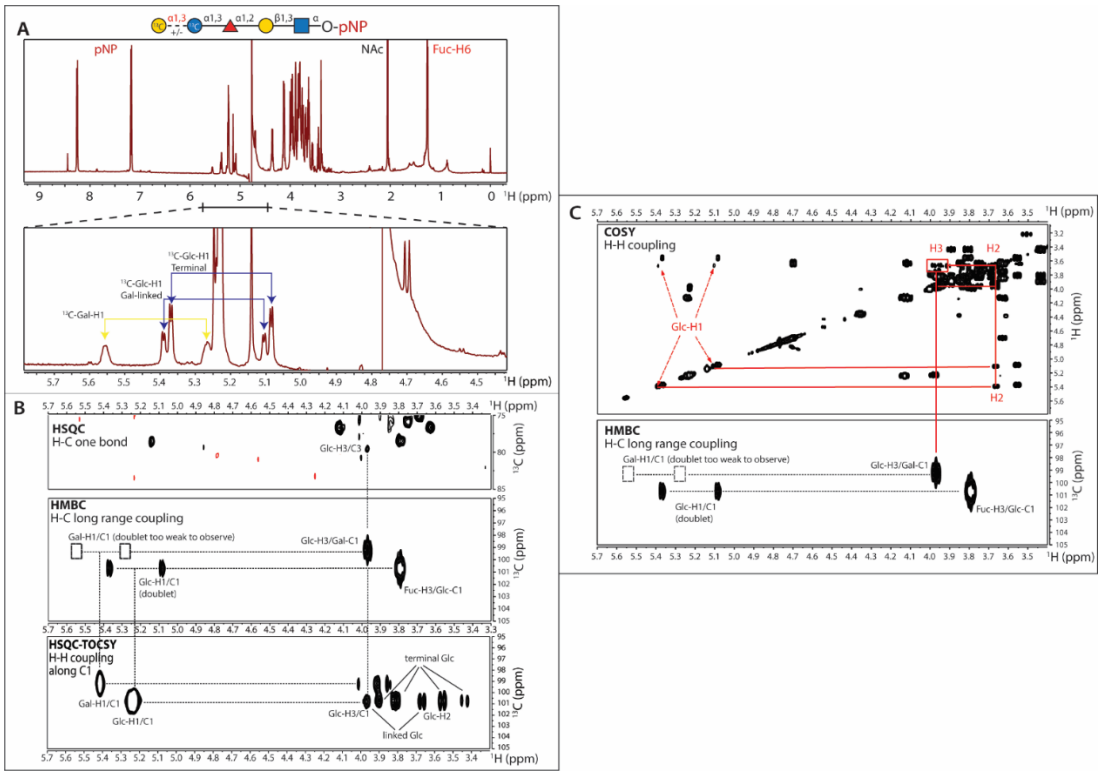


Figure 2.6. Gat1 assembles a Galα1,3Glc linkage on the Skp1 tetrasaccharide. NMR Analysis of the TgSkp1 pentasaccharide. [1-^{13}C]Glcα1,3Fucα1,2Gal-β1,3GlcNAcα1-pNP was partially (30%) modified by TgGat1 or PuGat1 in the presence of UDP-[U-^{13}C]Gal. (A) 1D 600 MHz ^{1}H-NMR spectrum. Magnification shows the region of ^{13}C-anomeric carbons in the mixture of modified and unmodified tetrasaccharide. A cartoon diagram of the TgSkp1 pentasaccharide attached to pNP (Galα1,3Glcα1,3Fucα1,2Gal-β1,3GlcNAcα1-pNP) is shown at the top using CFG glycan symbols (19). (B) ^{1}H –^{13}C-HSQC, -HMBC, and –HSQC TOCSY spectra demonstrating anomeric carbon to ring proton correlations. The Gal-H1/C1 doublet peaks were too weak to observe and are indicated by boxes. (C) ^{1}H –^{1}H –COSY and ^{1}H –^{13}C-HMBC spectra. Identical results were obtained using [1-^{13}C]GlFGaGn-pNP modified by PuGat1 (not shown). See Fig. S2.11 for a listing of chemical shift values.

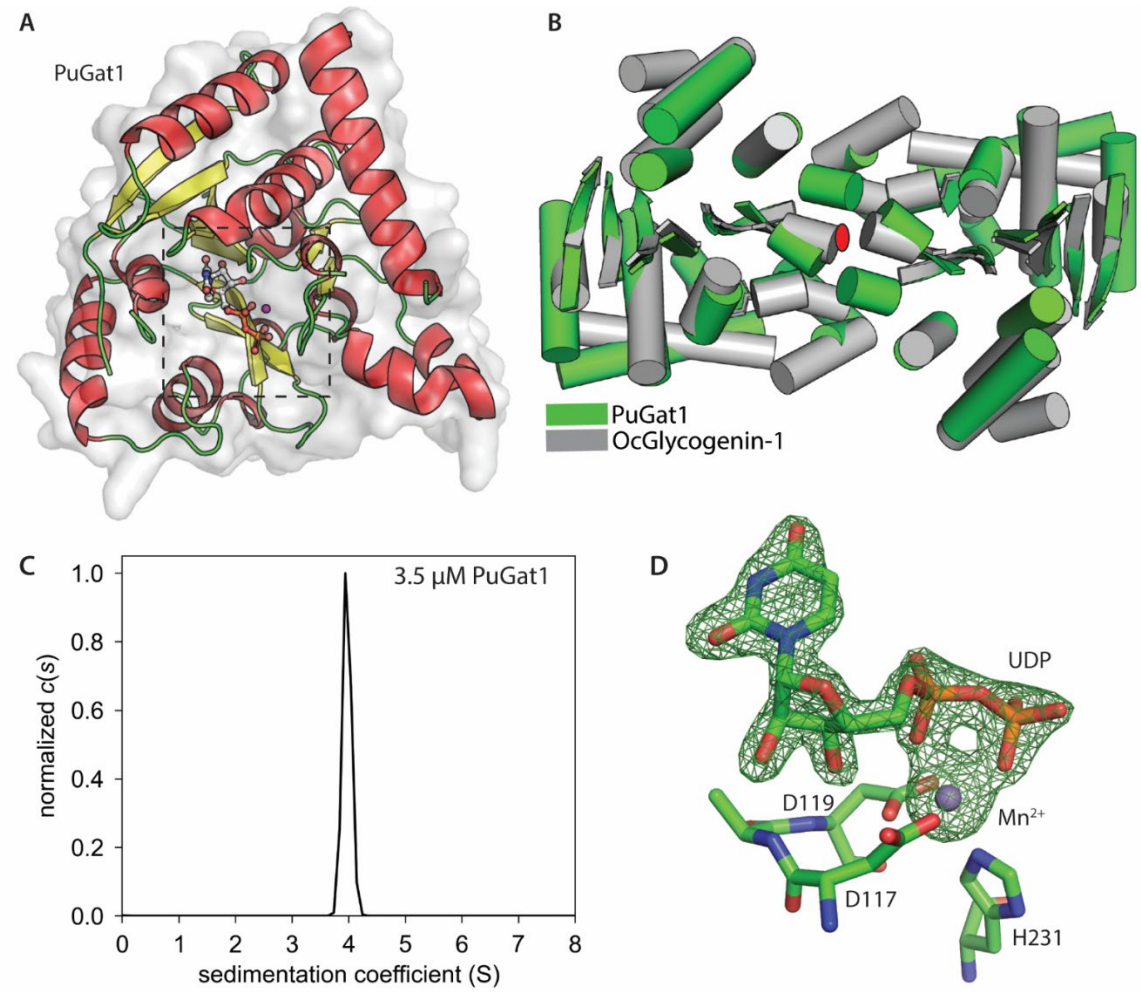


Figure 2.7. Gat1 is structurally related to glycogenin. (A) The asymmetric unit of crystallized Gat1 from *P. ultimum* in complex with UDP and Mn^{2+}. α-helices are in red, β-strands in yellow, loops in green, and the active site with bound ligand is boxed. (B) PuGat1 forms a homodimer that closely superimposes on Oc-glycogenin-1 (PDB entry 1LL2). The cylinders represent α-helices, the arrows represent β-sheets, and the red ellipse marks the two-fold symmetry axis perpendicular to the page. (C) Sedimentation velocity data modeled as a continuous $c(s)$ distribution (normalized to 1.0) yields an S-value for 3.5 µM PuGat1 that is close to the predicted value for a stable dimer in solution. Fig. S2.12 shows that the dimer is stable down to at least 0.3 µM. (D) UDP is coordinated in near identical fashion to that of glycogenin-1, based on the difference density map (F_o-F_c) that was contoured at 5σ, calculated after omitting UDP and Mn^{2+} and subjecting the structure to simulated annealing. Octahedral coordination of Mn^{2+} is satisfied by the DxD motif, His231, and UDP. The comparison with glycogenin-1 is illustrated in Fig. S2.13.

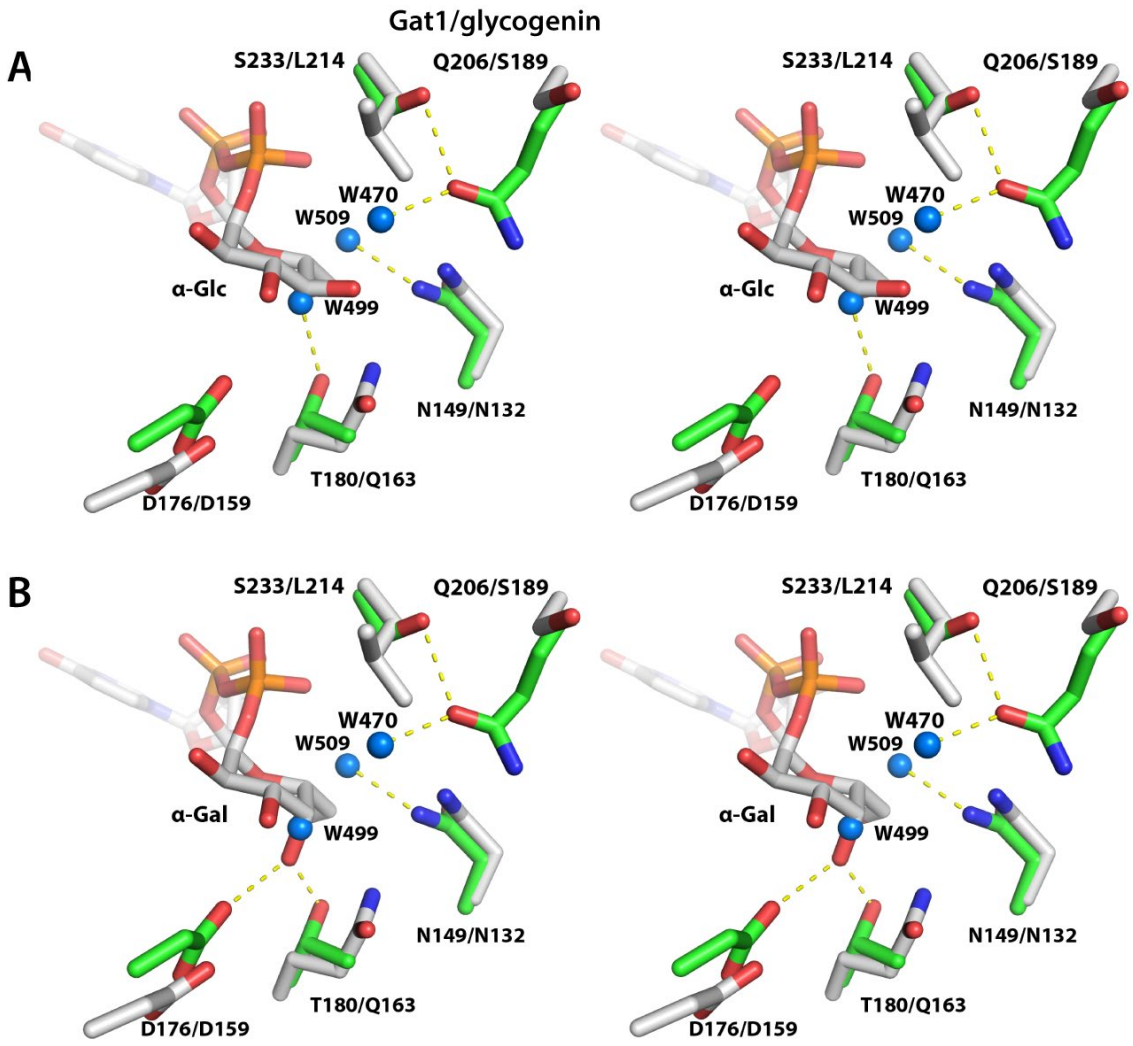


Figure 2.8. Active site geometry explains Gat1's preference for UDP-Gal rather than UDP-Glc. Comparison of the sugar binding pockets of PuGat1 and Oc-glycogenin-1 displayed as wall-eyed stereoview. (A) The Glc moiety is modeled based on the Oc-glycogenin-1 crystal structure with its intact sugar nucleotide. (B) The Gal moiety is modeled by flipping the stereochemistry of Glc at the C4' position. PuGat1 and glycogenin-1 side chains are represented by green and gray sticks, respectively, the yellow dashes represent hydrogen bonds, and water molecules are represented by blue spheres.

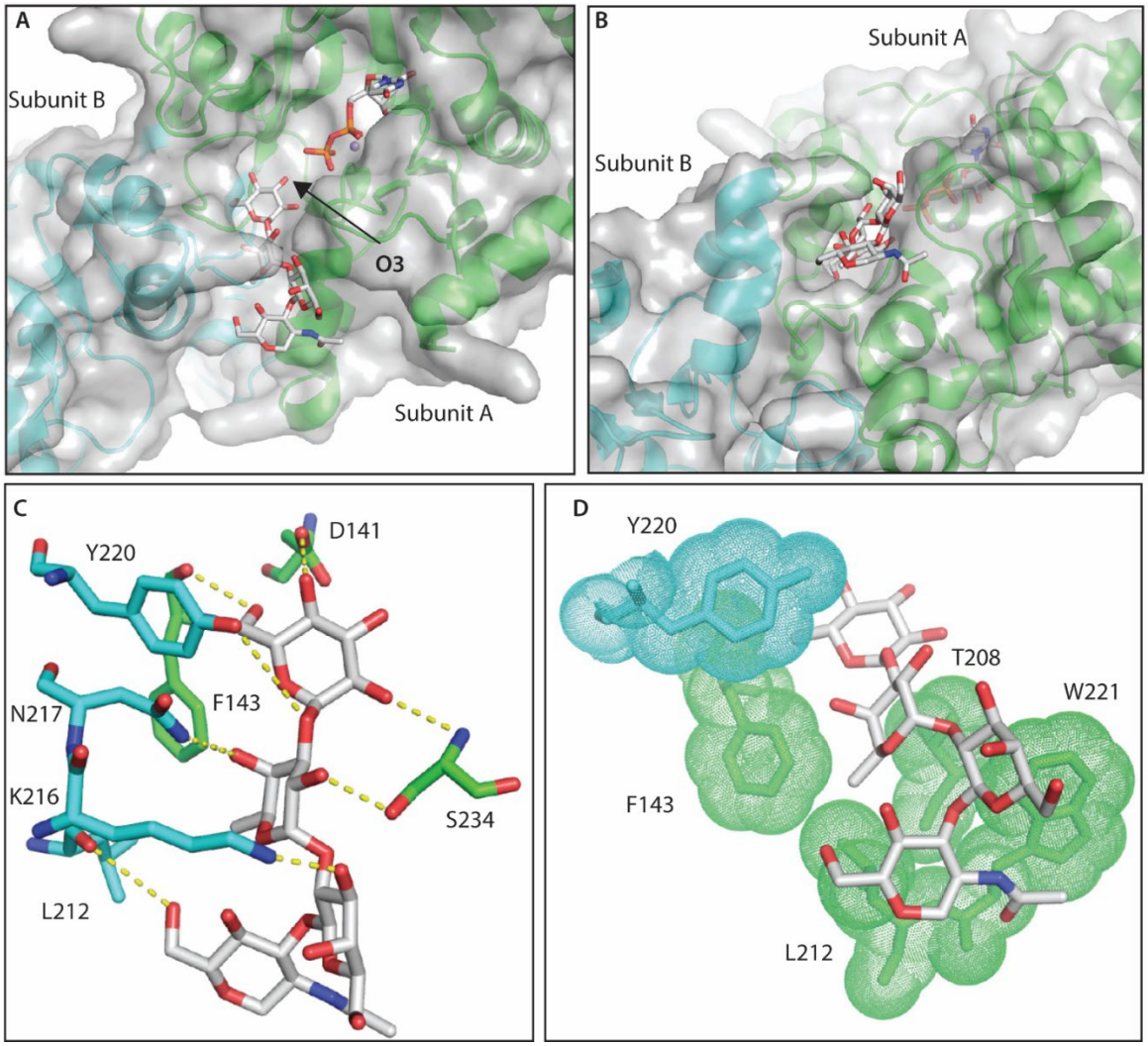


Figure 2.9. Computational docking explains specificity of Gat1 for the Skp1-tetrasaccharide. (A) The top docking pose of the Skp1-tetrasaccharide in the PuGat1 active site and a groove formed by the dimer. (B) 90° turn of the image shown in panel A. (C) Hydrogen bonding interactions of the glycan. (D) Hydrophobic packing of the sugar faces and fucose-methyl (in sticks) with non-polar surfaces (sticks/dots) of Gat1 subunit A (green) and B (cyan).

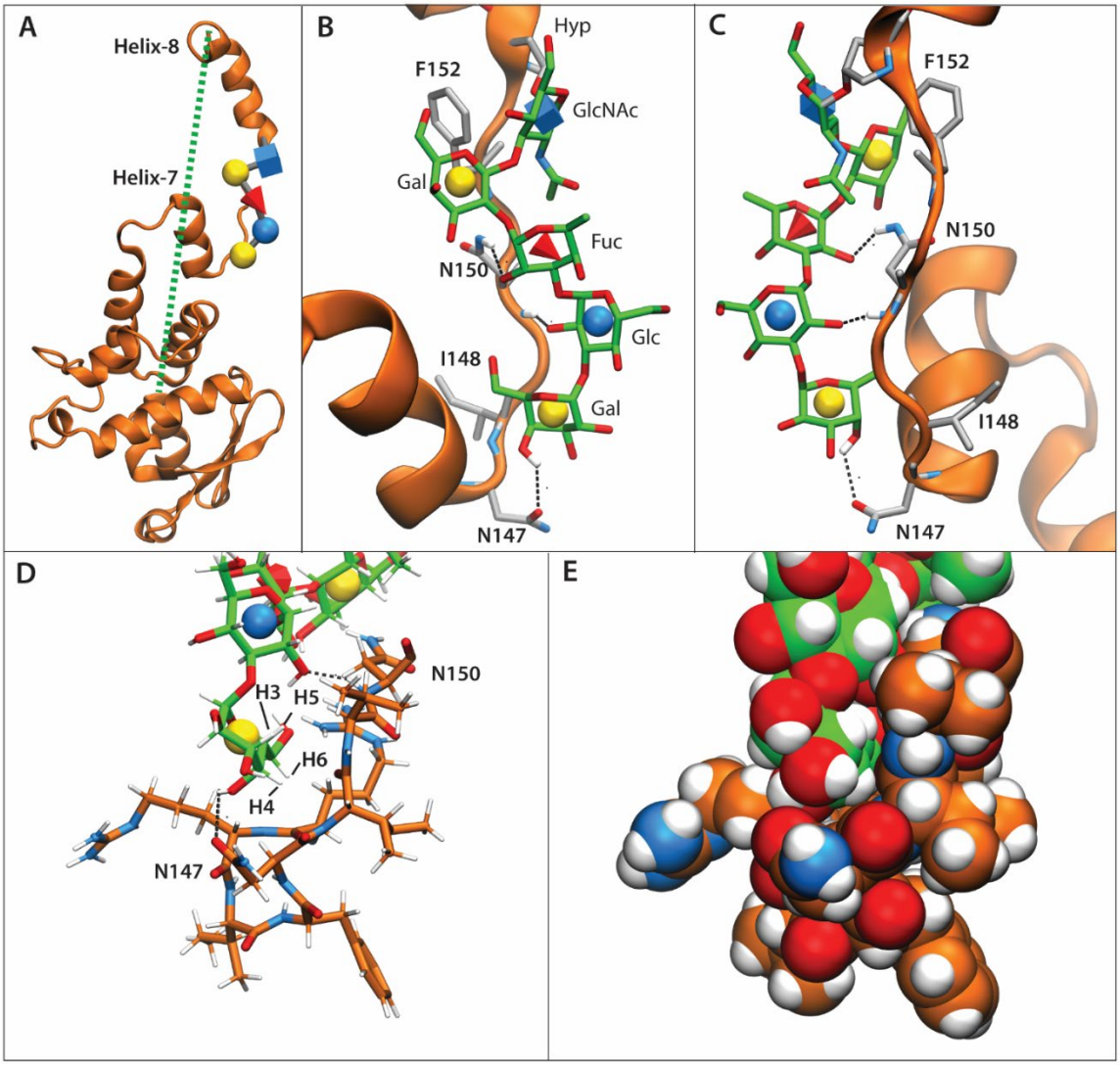


Figure 2.10. Packing of the glycan with Skp1 can explain F-box binding site conformation. *T. gondii* GaGlFGaGn-Skp1 was subjected to six 250-ns all-atoms molecular dynamics simulations. (A) A frame representative of the glycan-protein interaction and associated helix-8 extension, from a simulation (Equil-1, see Fig. S2.14E) in which the glycan was pre-equilibrated for 50 ns prior to the start of the simulation. The dotted green line refers to the distance from C-terminus to the center of mass of residues 1-136, and ranged from 18 to 61 Å. (B) Zoom-in of panel A depicting the glycan (C-atoms in green) and amino acids (C-atoms in gray) described in Table 2.2. Dotted black lines depict H-bonds contributing to the polar energies described in Table 2.2. (C) The back side of panel B. *D*, *E*, Packing of terminal sugars against the polypeptide. C-atoms of the peptide are in orange. (D) Glycan and peptide represented as sticks. (E) As in D, with glycan and peptide represented by spheres.

Description	Name	Parental Strain	Genotype	Gene Targeted	Selection marker	Reference
RH (type 1)			WT			21
Δ*gat1*/RH	MM12, cl.A8	RH	Δ*gat1*	*gat1*	DHFR	this report
gat1+/Δ*gat1*/RH	MM21, cl.E12	MM21	Δ*uprt*/*gat1*::*gat1*-Ty	*uprt*	Δ*uprt*	this report
RHΔΔ		RH	Δku80/Δ*hxgprt*			22
Δ*gat1*-1/RHΔΔ	KR10, cl.A1	RHΔΔ	Δku80/Δ*gat1*	*gat1*	HXGPRT	this report
Δ*gat1*-2/RHΔΔ	MM16, cl.B7	RHΔΔ	Δku80/Δ*hxgprt*/Δ*gat1*	*gat1*	DHFR	this report
gat1+/Δ*gat1*-2/ RHΔΔ	MM24, cl.G10	MM16	Δku80/Δ*hxgprt*/Δ*uprt*/ *tub*::*gat1*-3xHA	*uprt*	Δ*uprt*	this report
Me49-RFP (type 2)	MM8, cl.A10*	Me49	rfp+			23
Δ*gat1*/Me49-RFP	MM14, cl.B5	Me49-RFP	Δ*gat1*	*gat1*	DHFR	this report

*high RFP expression level clone from original source

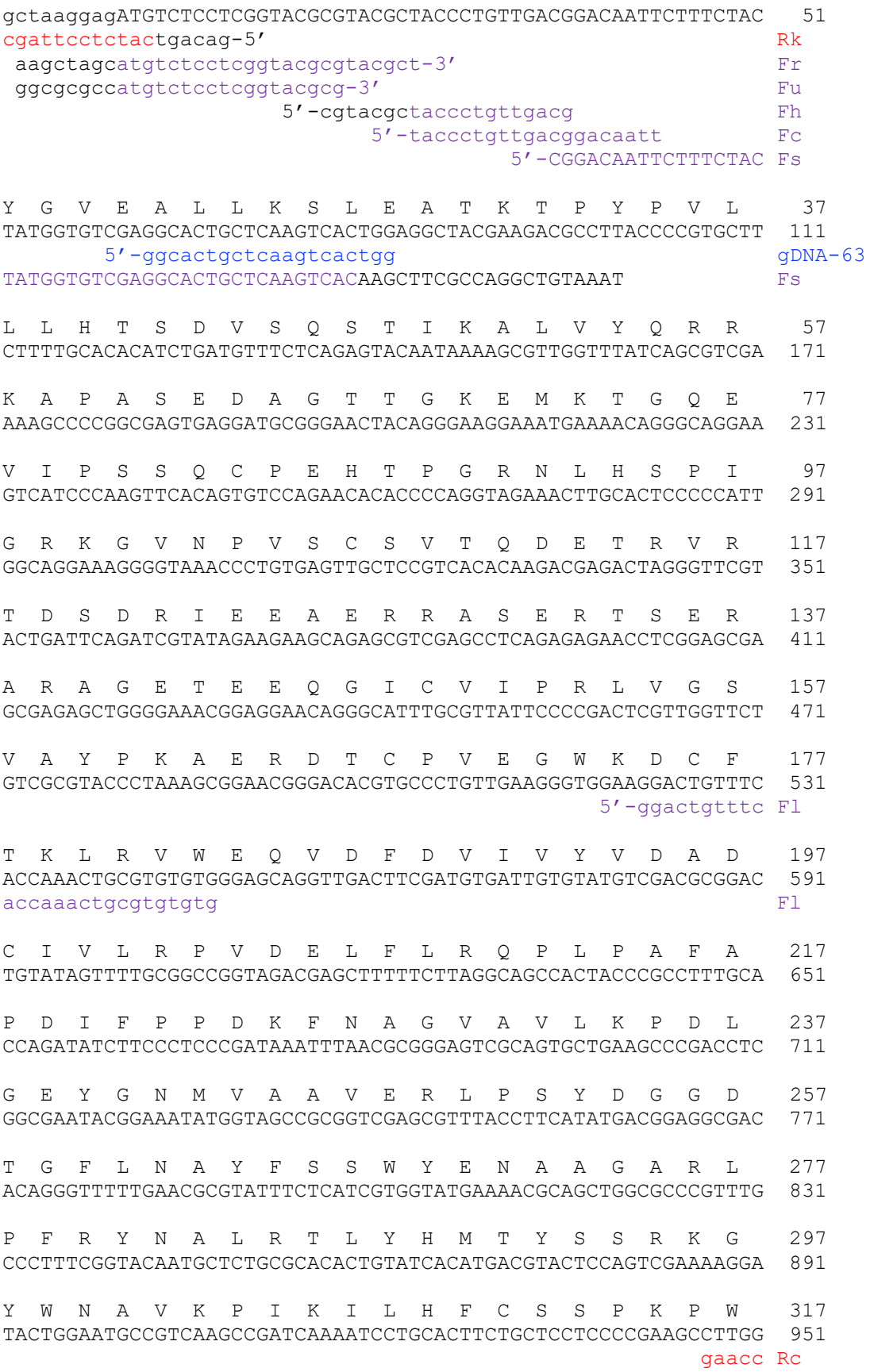
Table 2.1. Strains employed in this study

	Glycan	Protein	Avg. Energy*	R^2
	αGal	N147	-0.6	0.73
	Glc	N150	-0.6	0.60
Polar	Fuc	N150	-1.2	0.50
	αGal	N147	-0.81	0.73
	GlcNAc	F152	-1.5	0.66
	βGal	F152	-1.1	0.64
	GlcNAc	N150	-0.65	0.54
Non-Polar	αGal	I148	-0.67	0.52

*kcal/mol

Table 2.2. Table of MMGBSA-derived per-residue energies between the protein and glycan that exhibit a strong correlation with helix extension (distance in Fig. 2.7A) according to a linear regression analysis of 48 bins from the six MD simulations (Fig. S2.14E). Polar energies represent a sum of the electrostatic and polar solvation energies while the nonpolar energies are composed of the van der Waals and nonpolar solvation energies. Only interactions with an average polar/non-polar energy less than -0.5 were considered.

2.8 Supplementary Figures and Tables


gctaaggagATGTCTCCTCGGTACGCGTACGCTACCCTGTTGACGGACAATTCTTTCTAC 51
cgattcctctactgacag-5′ Rk
aagctagcatgtctcctcggtacgcgtacgct-3′ Fr
ggcgcgccatgtctcctcggtacgcg-3′ Fu
5′-cgtacgctaccctgttgacg Fh
5′-taccctgttgacggacaatt Fc
5′-CGGACAATTCTTTCTAC Fs
Y G V E A L L K S L E A T K T P Y P V L 37
TATGGTGTCGAGGCACTGCTCAAGTCACTGGAGGCTACGAAGACGCCTTACCCCGTGCTT 111
5′-ggcactgctcaagtcactgg gDNA-63
TATGGTGTCGAGGCACTGCTCAAGTCACAAGCTTCGCCAGGCTGTAAAT Fs
L L H T S D V S Q S T I K A L V Y Q R R 57
CTTTTGCACACATCTGATGTTTCTCAGAGTACAATAAAAGCGTTGGTTTATCAGCGTCGA 171
K A P A S E D A G T T G K E M K T G Q E 77
AAAGCCCCGGCGAGTGAGGATGCGGGAACTACAGGGAAGGAAATGAAAACAGGGCAGGAA 231
V I P S S Q C P E H T P G R N L H S P I 97
GTCATCCCAAGTTCACAGTGTCCAGAACACACCCCAGGTAGAAACTTGCACTCCCCCATT 291
G R K G V N P V S C S V T Q D E T R V R 117
GGCAGGAAAGGGGTAAACCCTGTGAGTTGCTCCGTCACACAAGACGAGACTAGGGTTCGT 351
T D S D R I E E A E R R A S E R T S E R 137
ACTGATTCAGATCGTATAGAAGAAGCAGAGCGTCGAGCCTCAGAGAGAACCTCGGAGCGA 411
A R A G E T E E Q G I C V I P R L V G S 157
GCGAGAGCTGGGGAAACGGAGGAACAGGGCATTTGCGTTATTCCCCGACTCGTTGGTTCT 471
V A Y P K A E R D T C P V E G W K D C F 177
GTCGCGTACCCTAAAGCGGAACGGGACACGTGCCCTGTTGAAGGGTGGAAGGACTGTTTC 531
5′-ggactgtttc Fl
T K L R V W E Q V D F D V I V Y V D A D 197
ACCAAACTGCGTGTGTGGGAGCAGGTTGACTTCGATGTGATTGTGTATGTCGACGCGGAC 591
accaaactgcgtgtgtg Fl
C I V L R P V D E L F L R Q P L P A F A 217
TGTATAGTTTTGCGGCCGGTAGACGAGCTTTTTCTTAGGCAGCCACTACCCGCCTTTGCA 651
P D I F P P D K F N A G V A V L K P D L 237
CCAGATATCTTCCCTCCCGATAAATTTAACGCGGGAGTCGCAGTGCTGAAGCCCGACCTC 711
G E Y G N M V A A V E R L P S Y D G G D 257
GGCGAATACGGAAATATGGTAGCCGCGGTCGAGCGTTTACCTTCATATGACGGAGGCGAC 771
T G F L N A Y F S S W Y E N A A G A R L 277
ACAGGGTTTTTGAACGCGTATTTCTCATCGTGGTATGAAAACGCAGCTGGCGCCCGTTTG 831
P F R Y N A L R T L Y H M T Y S S R K G 297
CCCTTTCGGTACAATGCTCTGCGCACACTGTATCACATGACGTACTCCAGTCGAAAAGGA 891
Y W N A V K P I K I L H F C S S P K P W 317
TACTGGAATGCCGTCAAGCCGATCAAAATCCTGCACTTCTGCTCCTCCCCGAAGCCTTGG 951
gaacc Rc


5′-tgg Fq
E Q P A K T D L E E L W W K V F L T G T 337
GAACAACCAGCAAAGACCGACCTCGAGGAACTATGGTGGAAAGTCTTCCTTACGGGCACT 1011
gaacaaccagcaaaga Fq
cttgttggtcgtttc-5′ Rc
5′-ccgacctcgaggaactatgg gDNA-968
ccgtga Rh
5′-gaagta Fn
ct Rr
gaagatacgtgaaacgtcctacaccacctttcagaaggaatgcccgtga Rs
V P T D S D I V * 345
GTGCCAACTGATTCTGATATCGTGTAGtggagggagaaccaaagtgatgatgaaagaatg 1071
a Rv
cacggttgactaagactatagcacc-5′ Rv
cacggttgactaaga-5′ Rh
cacacaaaccaagacccactagactagtggagggaga Fn
cacggttgactaagactatagcacatcggatcctt-5′ Rr
gccttgactaagactatagcaccttcatgtgtgtttg-5′ Rn
cggttgactaagactatagcaccatggatgggcatgctgca-5′ Ru
5′-cttggatagtggagggagaaccaaagtgatga Fk
5′-gctctagagggagaaccaaagtgatgat Fb
cacggttgactaagacta-5′ Rs
accgacttccaaaagaacggaaacgccggacagctgcctcgcggtaccttgggaaaagag
cgggacgtgtggaatcctgtcaactatctctttctgtgtcacctgtggacgaattgtaaa
tcttgtaaagtacaaacggagtacagcttaatcttgtaattttctttcttcgaaggacgc
a RH
cagtgcgccgcaagcgtctagtggcctgcaaagactagccttagggactgactggttcgc
tgatcggaatccctgactgatc-5′ Rl
gtacgcaaccatcacgcacaagcatgtttatttgccactggtgtgtcactcagctagacg
cgtcatgtttatgtatacgtacgtttcacagcctctcagagacatcccgcacaacgcatg
aaccgctgcaaccagaatactgaccgtcagcggtttcgcttgctttaaatcgggttgttt
tggaaaaactcaaaggtagtcgttctgtacatctccctaagttaagcggtaagttactcg
acgagcatacattgacaataagacggttctcacaatgaacatcccaaagagggcactaga
ccaaacaaaggagctaaaagacacgagcaagatgaagataaaacgcaccttagcgaaggc
catataacaaagtggatcttcacagtatcattctgtgtccgtaccagtcgctgcgaacaa
gaagacgcatgtgaacggttCTACTCTCGGAATTGAAAGCATTCAATACCGCCAGAGCTG
TGGT1_310410 reverse
CCCACTACGCACACCCGAACACCGCCAGGAACCGGTTTTCCGTCCAGATGAGCAGCATCC
GCAACAGACGTTTTGATCGCGTCCACACGCTCCACCGAAAAAATTTGTGTCTCGTCGATG
GTTGTAGACTCGCCCCAATATCGCCGCTGCTGGCATTCACGCGACCGCGCATTTAACCTC
TGGGCGGTAAGCGATGGTTTTCTTGTGGCACATGAACGCTGGATAGGCTCCGCTCCTTTC
ACCAGTCGAGATGAACTATTTTCGTTCCCATCGCATCCACCAGAGGCCACACCCGGACGA
ATAGTCCACACCTGCACCCATCGATCCCAGCCTACAGTAAACAGAAGGGGGAAGAGAAGC
GAGCAACTTCGTGCTGACGACGCATGAGCGTTTTCTACTCTACAGGCGGATATCAACCTA
T RH
AGGGAGAGAGACGCGTTGTTCTCCTCGTGTTCTACGCAAATAAACTGGACCGTTACTGAC
caagaggaacacaagatgcgtcgccggcgc-5′ Rb
TGGTCATCACCGCACGAGGCGACGCAGTAGCAAGGCGACTCCATTTCAGACTCTGCGAGG
gttccgctgaggtaaagtct-5′ Re

GCGGACGAACCCAGCCCTGCAGGAGATGCGACCGACGCGCCGCTTTCAACCTTTGTCCCC
AAAGCGTTTTGCGTTTCGGCCACTTCCAAGTCATTTACGCCACACTGGTGTGCGTGTAGA
GTGCCCAATCGCTCCAGTCGAGGCACCGCAGCGCTCGTGTCCCCCAGGAACACCGCGATT
TCTCCGGTGGTCATTCCGACCAGTACCAAGACCCCATGTGTGTTCGTTCTCGACGGTTCC
TCACTCCCGAACTCTGTCTGGGAGCTAGCAAAGGAGCAGCCTCCAGCATCGTTTGCATGC
GTGTCTAGACAAACGACACGAACGCAGAGAGCCGCGCCTTCCAGCGGGCATGATGCTCTT
TTTTCTAAGCGACTGTTCCGGCTCCGGTCACCATGAGCCAGCTGCCCGAATTCCTGAGAC
CTCGTCAGCACAAGTGTCGCTTCCACATACGCGACTTCGGCCGACGACGCAGCCGTCCAC
ATGTGCACAGTCAGAGACTC

Figure S2.1. Genomic sequence surrounding the open reading frame of Gat1 (TGME49_310400 model from Toxodb.org). Numbering begins at the A of the start codon ATG. Coding sequences, including those upstream and downstream of Gat1, are capitalized; non-coding sequences are lower case. Amino acid sequence of Gat1 is above its coding sequence. Sequences of oligonucleotides from Table S2.1 are shown and mapped. For forward PCR primers, cognate sequences are colored purple; cognate sequences of reverse PCR primers are in red; guide DNA sequences are in blue. nt differences observed in the type 1 RH strain are indicated.

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

```
M  T  V  G  T  R  R  A  A  Y  A  T  L  I  T  S  D  A  Y  V
atgaccgtcggcacgcgcagggcggcgtacgcaacactgatcacgtccgatgcgtacgtc
ATGACTGTCGGAACACGTCGTGCGGCTTATGCCACTTTGATCACCAGCGATGCGTACGTT        60

M  G  V  E  A  L  V  Y  S  L  F  K  A  R  V  A  F  P  L  V
atgggcgtcgaggcgctcgtctactcgctcttcaaggcgcgcgtagcgtttccgctcgtg
ATGGGCGTCGAGGCATTAGTGTATAGCTTGTTTAAGGCGCGTGTTGCCTTCCCACTTGTG        120

V  L  H  S  S  Q  V  T  Q  P  T  V  A  K  L  T  R  F  C  A
gtgctgcattcgtcgcaggtgacgcagcccacggtggccaaactcacgcgcttctgcgcg
GTGTTACACAGCAGCCAGGTTACTCAGCCAACGGTGGCGAAGCTTACCCGTTTCTGCGCC        180

P  F  Q  S  S  T  W  R  I  S  F  R  S  V  P  D  I  G  I  P
ccattccagtcatcaacgtggcgcatttcgttccgctctgtcccagatatcggcatccca
CCCTTCCAAAGCAGCACATGGCGTATTAGCTTTCGTAGCGTTCCTGATATCGGTATCCCA        240

D  E  V  T  D  R  S  T  V  H  V  P  G  W  V  N  S  G  Y  T
gacgaagtcactgataggagcacggtgcacgtgcctggatgggtcaactcggggtacacc
GACGAGGTAACTGATCGTAGTACCGTCCATGTGCCGGGATGGGTTAATTCAGGTTACACA        300

K  L  H  I  F  A  M  D  D  F  E  Q  I  V  Y  I  D  A  D  A
aagctccacatcttcgccatggacgactttgagcagatcgtgtacattgacgccgacgcc
AAGCTTCATATCTTCGCTATGGACGACTTCGAGCAAATCGTCTATATTGACGCCGATGCC        360

I  V  L  Q  N  V  D  E  L  F  D  R  S  T  S  F  A  A  A  P
atcgtcctacagaacgtcgacgagcttttcgatcgctcaacgagctttgcggctgcgccc
ATTGTTCTTCAAAACGTAGACGAGTTGTTCGACCGTAGTACCAGCTTCGCGGCGGCGCCT        420

D  V  F  P  P  D  R  F  N  A  G  V  L  V  I  R  P  N  K  Q
gacgtgtttccacccgaccgcttcaacgccggcgtgctcgtgatccgtccgaacaagcag
GACGTATTTCCACCAGACCGTTTTAACGCGGGGGTGCTTGTCATTCGTCCTAACAAACAA        480

L  F  A  D  L  L  A  K  A  K  E  L  K  S  Y  D  G  G  D  T
ctctttgcagacttactggcgaaagccaaggagctcaagtcgtacgatggcggcgacacg
CTTTTCGCCGACTTGTTAGCGAAGGCCAAGGAATTGAAAAGCTATGATGGGGGCGATACA        540

G  F  L  N  A  F  F  P  K  W  F  E  S  D  A  A  S  R  L  P
ggcttcctcaatgcgtttttccccaagtggttcgaatcggacgccgcgtcgagactgccg
GGATTCTTAAACGCTTTTTTCCCCAAGTGGTTCGAGTCCGATGCCGCCTCACGTTTGCCT        600

F  G  Y  N  A  Q  R  T  M  Y  W  L  V  N  G  K  N  P  G  Y
tttggatacaacgcgcagcgcacgatgtactggctcgtgaacggcaagaaccccgggtac
TTTGGTTACAATGCTCAGCGTACGATGTACTGGCTTGTGAACGGGAAGAACCCTGGGTAC        660

W  N  A  V  Q  P  L  K  I  L  H  Y  S  S  N  P  K  P  W  E
tggaacgccgtccagccgctcaagatcctgcactactcgtcgaatccaaagccctgggag
TGGAACGCGGTCCAGCCTTTGAAGATTCTTCACTATTCATCCAATCCTAAACCCTGGGAG        720

D  P  S  R  K  G  D  L  E  I  L  W  W  Q  M  Y  T  E  S  R
gatccgagtcgcaagggtgacctggagatcctgtggtggcaaatgtacacggaatccaga
GACCCAAGTCGTAAGGGTGACTTGGAGATTCTTTGGTGGCAAATGTATACCGAAAGTCGT        780

C  M  S  F  L  G  *
tgcatgagctttctggggtag
TGTATGAGCTTTCTTGGATAG
                                801
```

Figure S2.2 *Pythium ultimum* Gat1 sequences

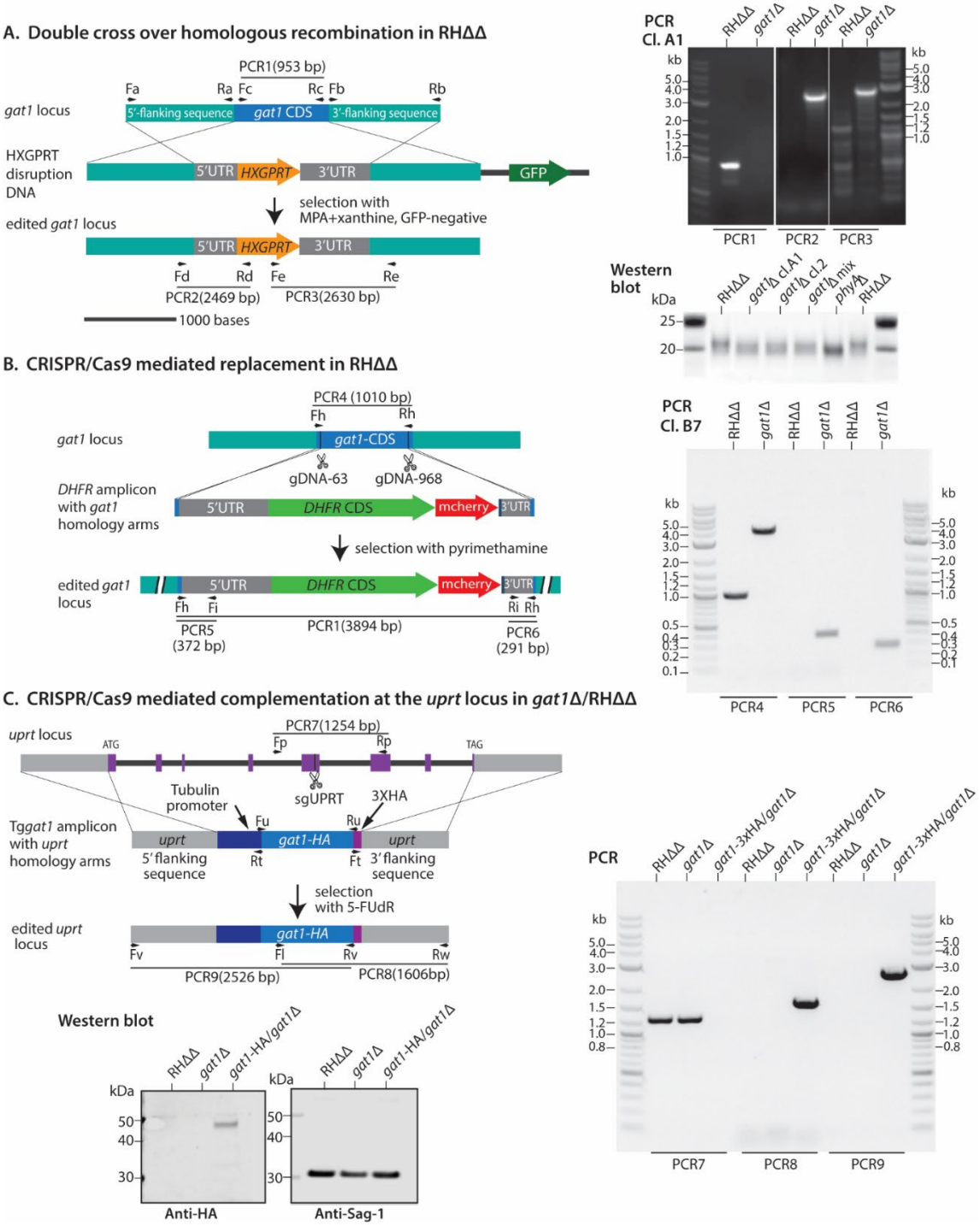


Figure S2.3. Disruption and complementation of Tg*gat1* in the RHΔΔ type 1 strain. (A) Disruption of *gat1* by double cross-over homologous recombination in RHΔΔ. The disruption DNA consisted of an HXGPRT cassette flanked by a 1177-nt 5'-upstream DNA (prepared by PCR using primers Fa and Ra) and a 1209-nt 3'-downstream DNA (prepared using primers Fb and Rb) of the *gat1* coding sequence (CDS), and an adjacent GFP-expression cassette. GFP-negative clones that grew in mycophenolic acid (MPA) and xanthine showed evidence of a gel shift of Skp1 relative to parental and *phyA*Δ cells based

on Western blotting using pAb UOK75. Clone A1 was confirmed to have the desired integration by PCR reaction #1 (primers Fc and Rc), which showed loss of the *gat1* CDS, and PCR2 (Fd and Rd) and PCR3 (Fe and Re), which showed integration of the HXGPRT within the gat1 locus. (B) Disruption of *gat1* using a double CRISPR/Cas9 strategy. RHΔΔ parasites were transiently transfected with a plasmid encoding gDNA-63 and gDNA-968 guide DNA's and Cas9, and a PCR amplicon expressing the DHFR resistance cassette flanked by 45-bp *gat1* homology arms. Pyrimethamine-resistant clones that replaced the *gat1* CDS with the DHFR cassette were confirmed using PCR4 (Fh and Rh), which showed loss of *gat1* CDS, and PCR2 and PCR3, which showed the integration of DHFR in the forward orientation. (C) Complementation of clone B7 from panel B by replacement of the *uprt* locus with a *gat1* expression cassette consisting of a tubulin promoter, Tg*gat1* CDS modified with DNA encoding a C-terminal 3×HA tag, and *uprt* targeting sequences. Correct insertion of the gat1 expression cassette was assessed by PCR reactions PCR7, PCR8 and PCR9, and confirmed by Western blot analysis for a predicted M_r 45,000 protein band that could be detected with mAb 12CA5 that recognizes the 3xHA epitope. A parallel gel containing samples without reducing reagent was Western blotted to detect Sag1 as a loading control.

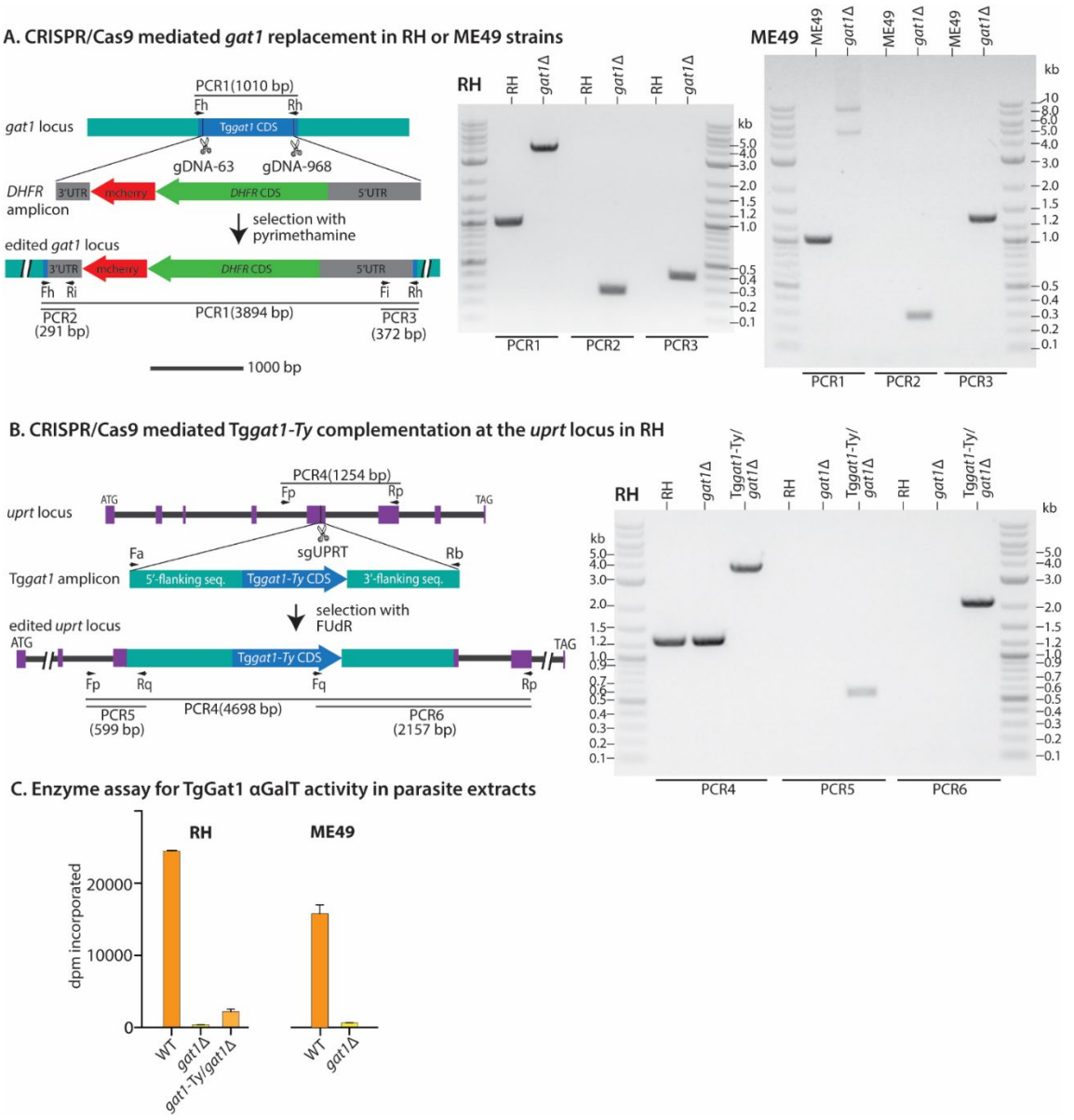
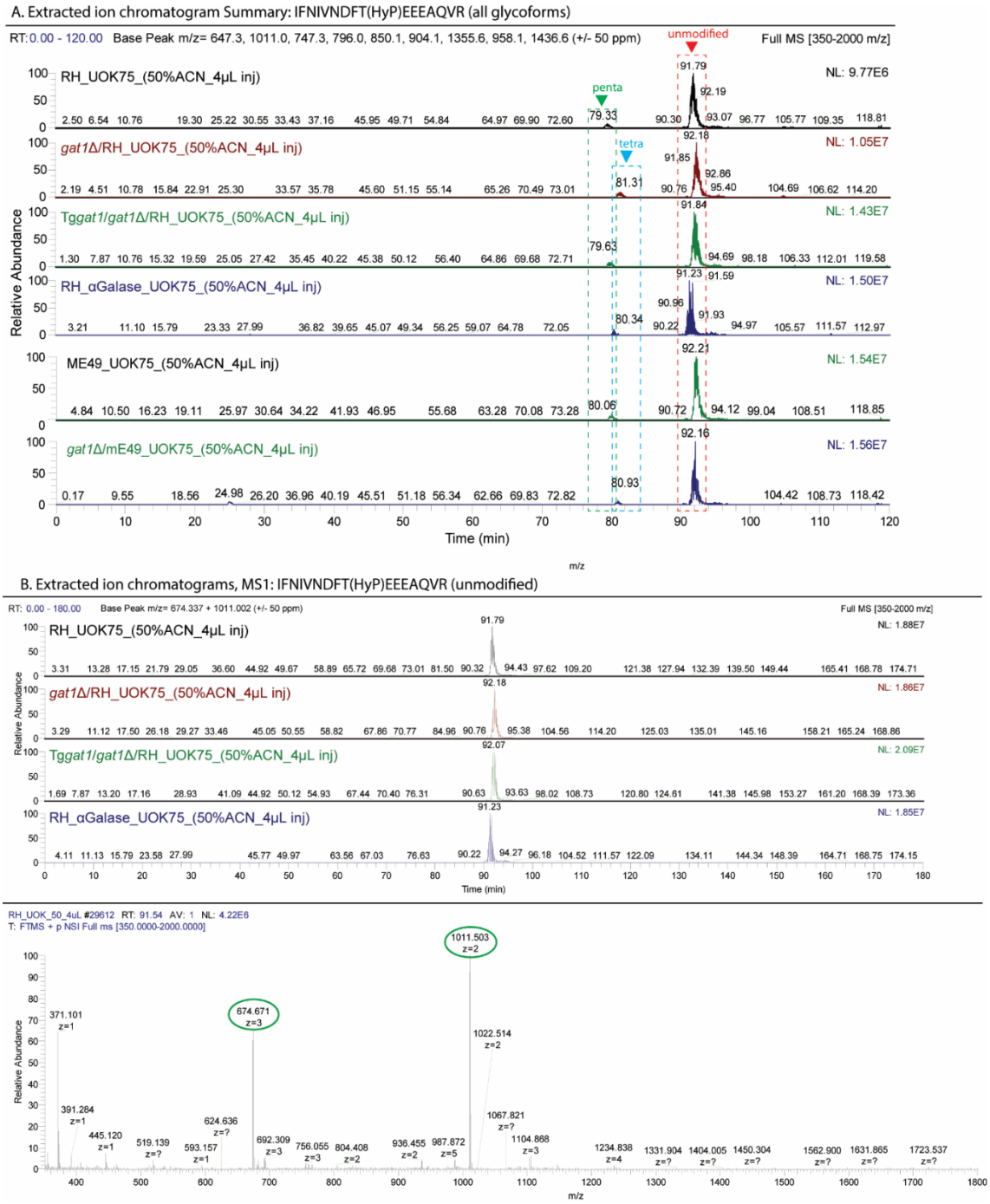
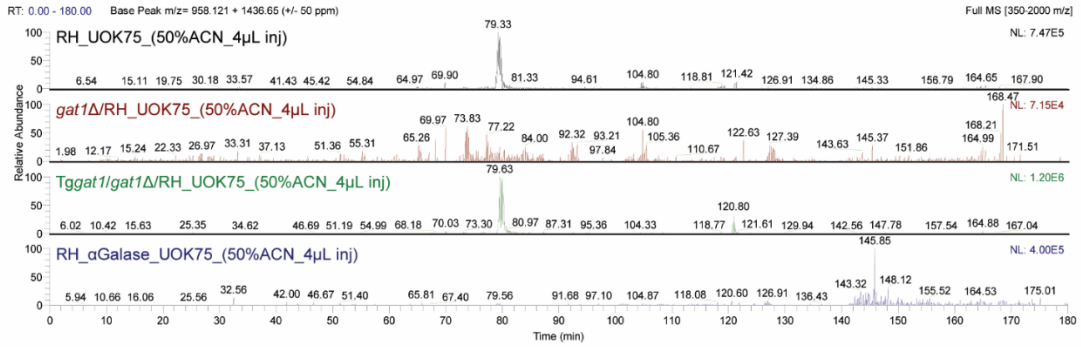


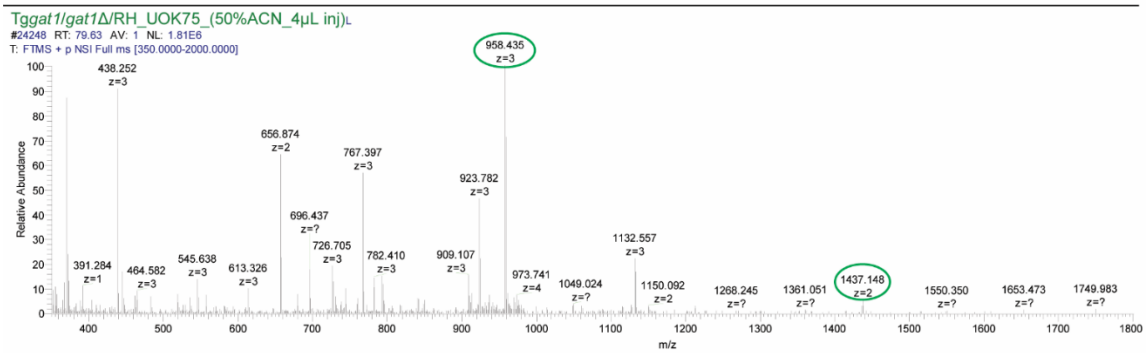
Figure S2.4. Disruption and complementation of Tg*gat1* in Ku80+ type 1 and type 2 strains. (A) Disruption of *gat1* was achieved using the double CRISPR/Cas9 strategy described in Fig. S2.3B, except that the DHFR amplicon lacked *gat1* homology arms owing to the presence of non-homologous end joining activity. Successful replacement was evaluated for strains RH and ME49, by PCR as in Fig. S2.3B. (B) The RH *gat1*Δ strain was complemented by insertion of a genomic fragment of Tg*gat1* including its CDS, DNA encoding a C-terminal Ty-tag, and >1 kb of flanking DNA from both directions. Successful integration was verified using *uprt*-specific primers (Fp and Rp, Table S1) flanking the CRISPR/Cas9 cut site in PCR reaction #4. The identity of the integrated DNA was verified using primer pairs Fp and Rq, and Fq and Rp, in which Rq and Fq were specific to Tg*gat1* DNA. (C) Extracts of *gat1*Δ and complemented clones from panels A and B were analyzed for Skp1 αGalT activity. Desalted S100 extracts were prepared by hypotonic lysis and gel filtration, and incubated in the presence of GlFGaGn-Skp1 and UDP-[^{3}H]Gal. The

reactions were separated on SDS-PAGE gels, the Skp1 band was excised after Coomassie blue staining, and radioactivity determined by liquid scintillation counting. Error bars represent S.D. of two technical replicates of the same samples.

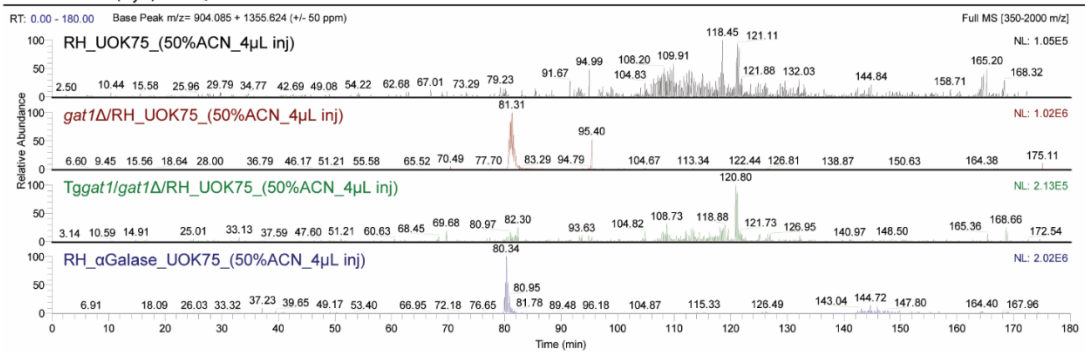

A. Extracted ion chromatogram Summary: IFNIVNDFT(HyP)EEEAQVR (all glycoforms)
RT: 0.00 - 120.00 Base Peak m/z= 647.3, 1011.0, 747.3, 796.0, 850.1, 904.1, 1355.6, 958.1, 1436.6 (+/- 50 ppm)
Full MS [350-2000 m/z]
unmodified
penta
tetra
RH_UOK75_(50%ACN_4µL inj)
NL: 9.77E6
gat1Δ/RH_UOK75_(50%ACN_4µL inj)
NL: 1.05E7
Tggat1/gat1Δ/RH_UOK75_(50%ACN_4µL inj)
NL: 1.43E7
RH_αGalase_UOK75_(50%ACN_4µL inj)
NL: 1.50E7
ME49_UOK75_(50%ACN_4µL inj)
NL: 1.54E7
gat1Δ/mE49_UOK75_(50%ACN_4µL inj)
NL: 1.56E7
Relative Abundance
Time (min)
m/z
B. Extracted ion chromatograms, MS1: IFNIVNDFT(HyP)EEEAQVR (unmodified)
RT: 0.00 - 180.00 Base Peak m/z= 674.337 + 1011.002 (+/- 50 ppm)
Full MS [350-2000 m/z]
RH_UOK75_(50%ACN_4µL inj)
NL: 1.88E7
gat1Δ/RH_UOK75_(50%ACN_4µL inj)
NL: 1.86E7
Tggat1/gat1Δ/RH_UOK75_(50%ACN_4µL inj)
NL: 2.09E7
RH_αGalase_UOK75_(50%ACN_4µL inj)
NL: 1.85E7
Relative Abundance
Time (min)
RH_UOK_50_4uL #29612 RT: 91.54 AV: 1 NL: 4.22E6
T: FTMS + p NSI Full ms [350.0000-2000.0000]
1011.503 z=2
674.671 z=3
371.101 z=1
1022.514 z=2
391.284 z=1
445.120 z=1
519.139 z=?
624.636 z=?
593.157 z=1
692.309 z=3
756.055 z=3
804.408 z=2
936.455 z=2
987.872 z=5
1067.821 z=?
1104.868 z=3
1234.838 z=4
1331.904 z=?
1404.005 z=?
1450.304 z=?
1562.900 z=?
1631.865 z=?
1723.537 z=?
Relative Abundance
m/z

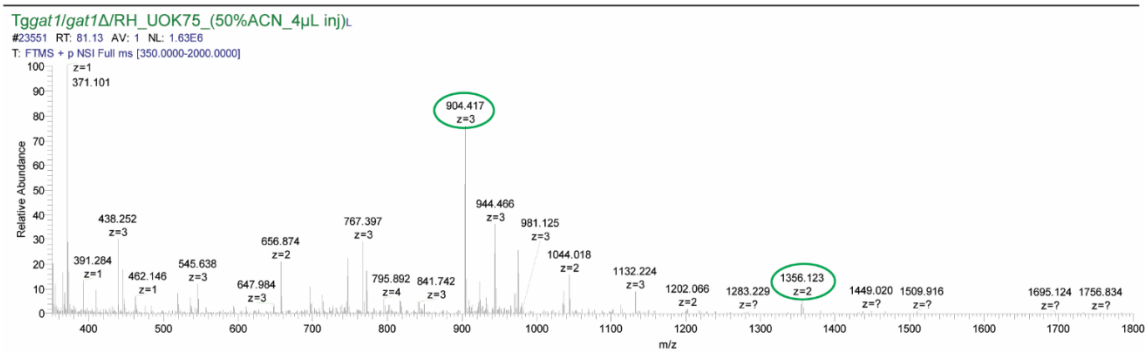
C. IFNIVNDFT(HyP)EEEAQVR + pentasaccharide





D. IFNIVNDFT(HyP)EEEAQVR + tetrasaccharide

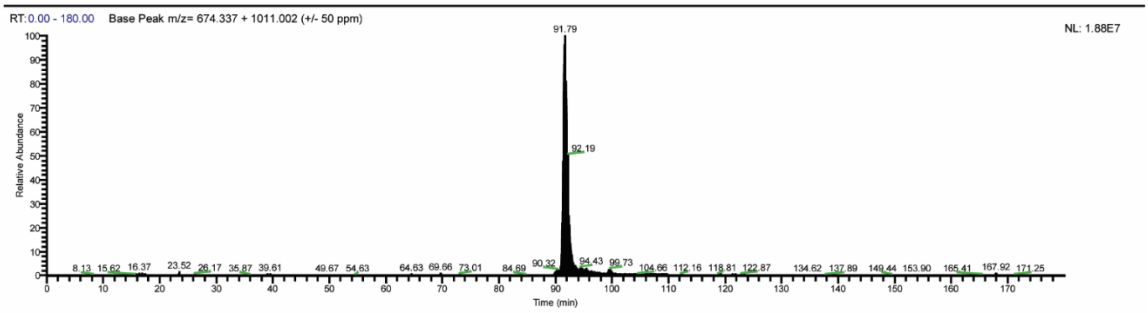


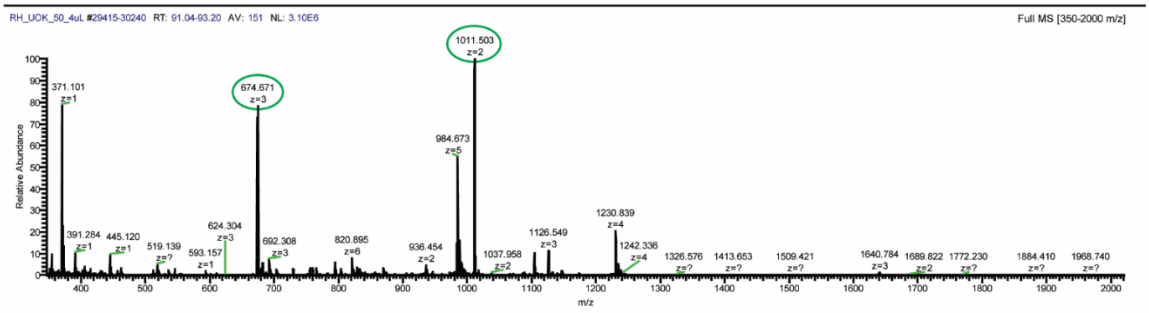


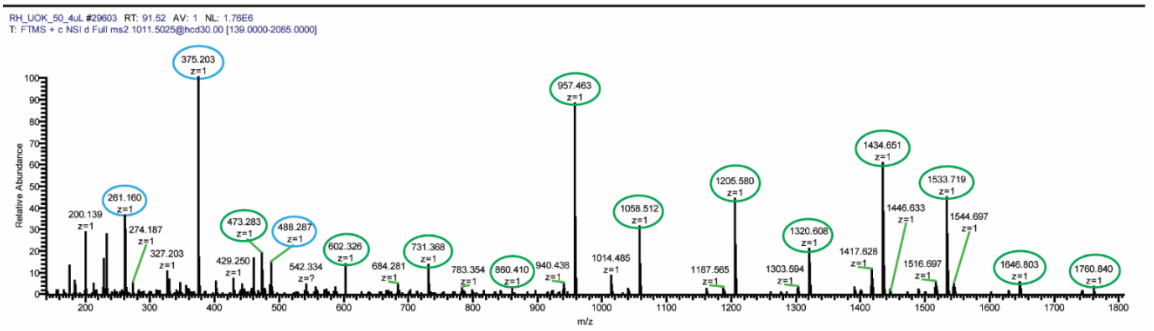
E. MS2 of unmodified peptide from RH

IFNIVNDFT(Pro)EEEAQVR (unmodified)

RH_UOK75_(50%ACN_4µL inj)







IFNIVNDFT(Pro)EEEAQVR

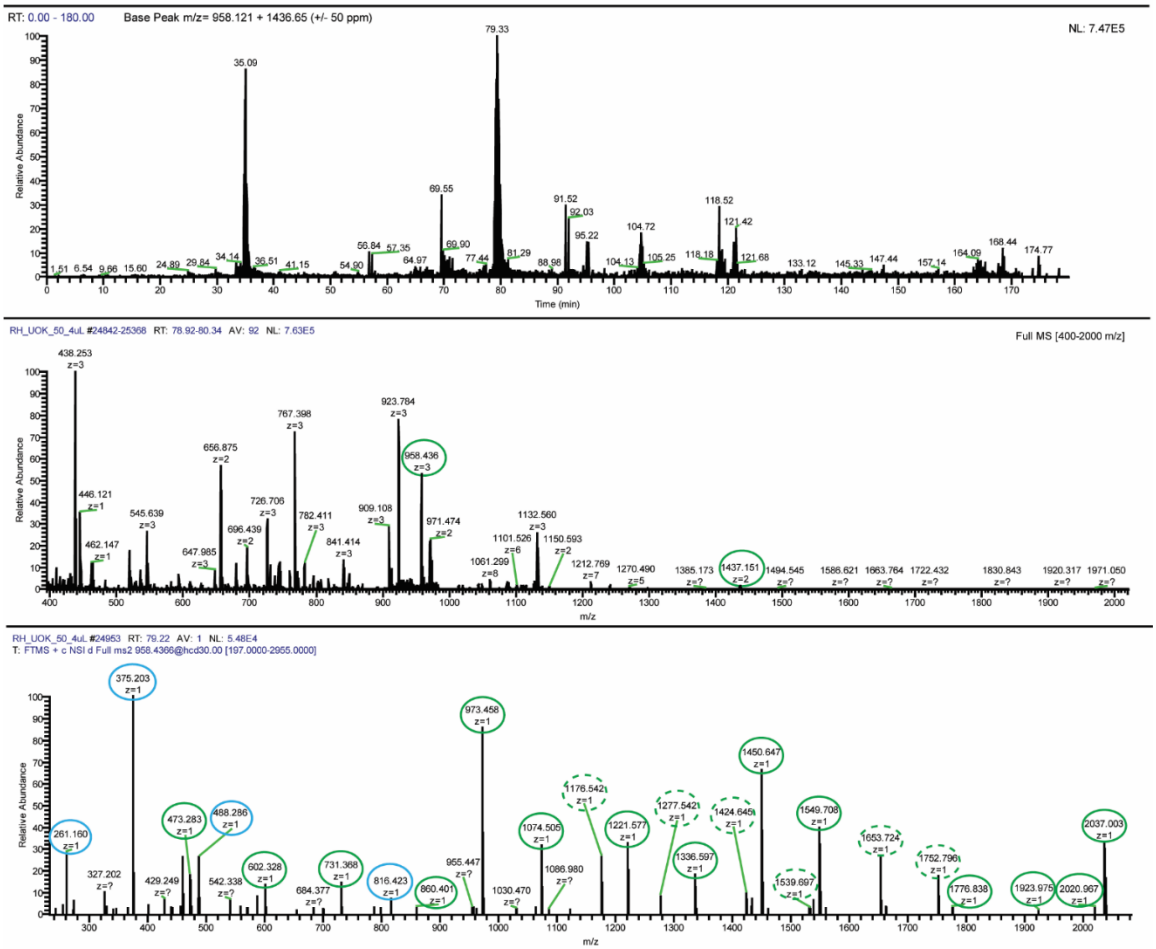
		unmodified		
		b	y	
I	1	114.0914	2020.997	1
F	2	**261.1598**	1907.913	2
N	3	**375.2027**	**1760.845**	3
I	4	**488.2868**	**1646.802**	4
V	5	587.3552	**1533.718**	5
N	6	701.3981	**1434.65**	6
D	7	816.4251	**1320.607**	7
F	8	963.4935	**1205.58**	8
T	9	1064.541	**1058.511**	9
P	10	1161.594	**957.4636**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

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

F. MS2 of pentasaccharide peptide from RH

IFNIVNDFT(HyP+HexNAc+Fuc+Hex3)EEEAQVR

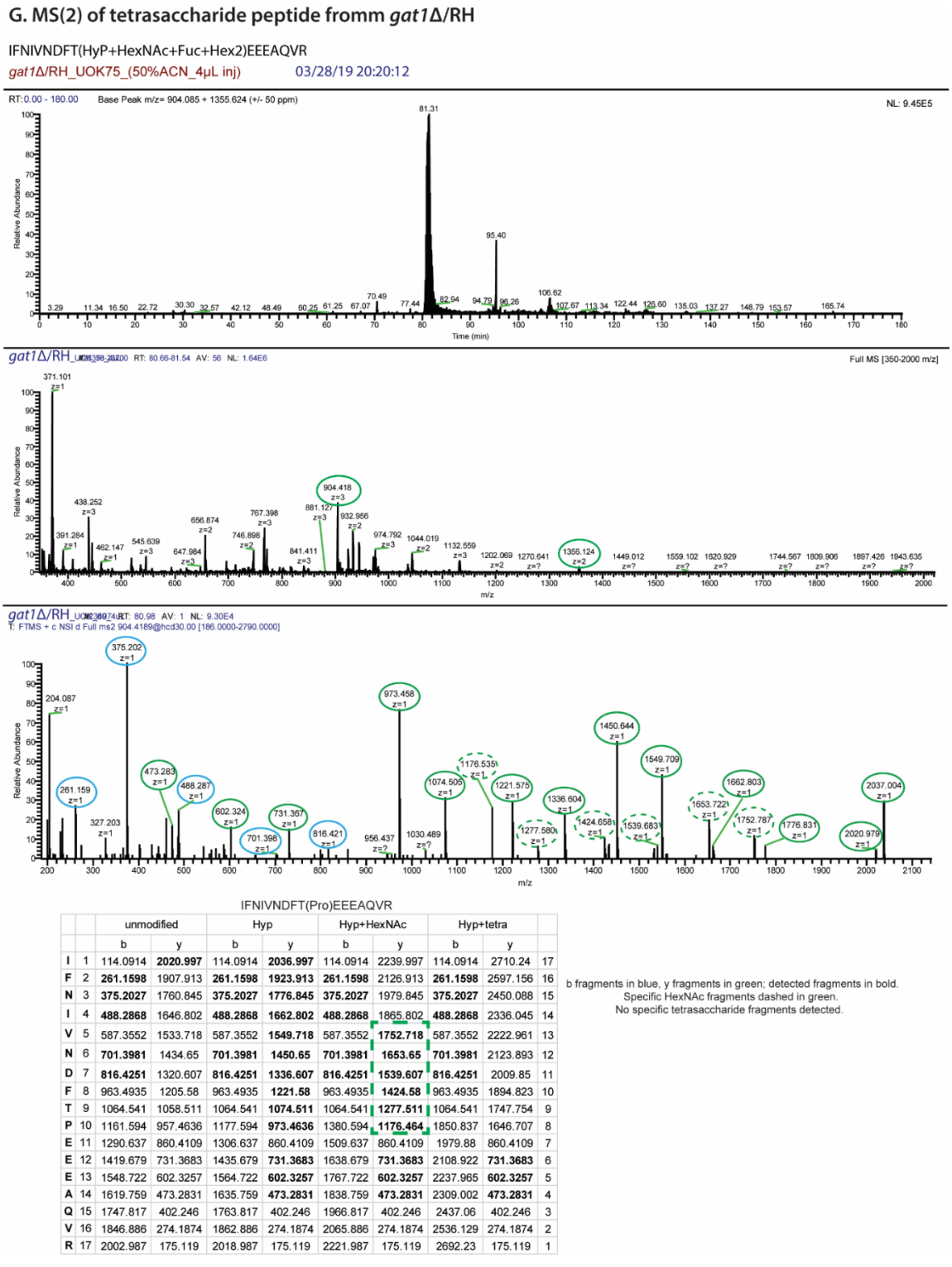
RH_UOK75_(50%ACN_4µL inj)



IFNIVNDFT(Pro)EEEAQVR

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

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



IFNIVNDFT(Pro)EEEAQVR

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

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

Figure S2.5. nLC/MS of Skp1 glycopeptides (supports Figure 2.1C, Table S2.2)
TgSkp1 isolated by immunoprecipitation from tachyzoite extracts were reduced and alkylated, trypsinized treated with green coffee bean α-galactosidase as indicated, and analyzed by a standard proteomics workflow consisting of separation on a C18 nLC column and analysis in an QE-Plus Orbitrap mass spectrometer. (A) Stacked extracted ion

chromatograms for all isoforms of peptide(134-150), which contains the modifiable Pro143, that were detected in the RH and ME49 backgrounds. (B) Selected extracted ion chromatograms for unmodified peptide(134-150), and an example of an MS^1 spectrum, from the RH sample, highlighting the parent ion for the unmodified peptide. (C) Selected extracted ion chromatograms for pentasaccharide-modified peptide(134-150), and an example of an MS^1 spectrum, from the TgGat1 complemented sample, highlighting the parent ion for the pentasaccharide-modified peptide. (D) Selected extracted ion chromatograms for tetrasaccharide-modified peptide(134-150), and an example of an MS^1 spectrum, from the *gat1Δ*/RH sample, highlighting the parent ion for the tetrasaccharide-modified peptide. (E) MS^2 of unmodified peptide(134-150) from RH, with associated extracted ion chromatogram and MS^1. Detected b and y fragment ions that define the peptide sequence are in bold in the list of predicted fragments ions at the bottom. (F) MS^2 of pentasaccharide peptide(134-150) from RH, with associated extracted ion chromatogram and MS^1. An expanded table of predicted b and y fragment ions, calculated to include the full pentasaccharide or a GlcNAc stub, is at the bottom. MS^2 fragmentation resulted in either loss of the full glycan leaving Hyp, or retention of a GlcNAc stub (encircled with a dashed green line). (G) MS^2 of tetrasaccharide peptide(134-150) from *gat1*Δ/RH, with associated extracted ion chromatogram and MS^1. An expanded table of predicted b and y fragment ions, calculated to include the full tetrasaccharide or a GlcNAc stub, is at the bottom. MS^2 fragmentation resulted in either loss of the full glycan leaving Hyp, or retention of a GlcNAc stub (encircled with a dashed green line).

Figure S2.6. Alignment catalytic domains of Gat1-like sequences and glycogenins (supports Fig. S8). To facilitate visualization of relatedness, acidic residues are in blue, basic in dark red, Gly and Pro in red, and hydrophobic in green, as previously described[88]. Positions possessing a consensus chemical characteristic are highlighted in yellow (hydrophobic), gray (acidic), dark grey (basic), or teal (small). Positions of near perfect conservation are bold.

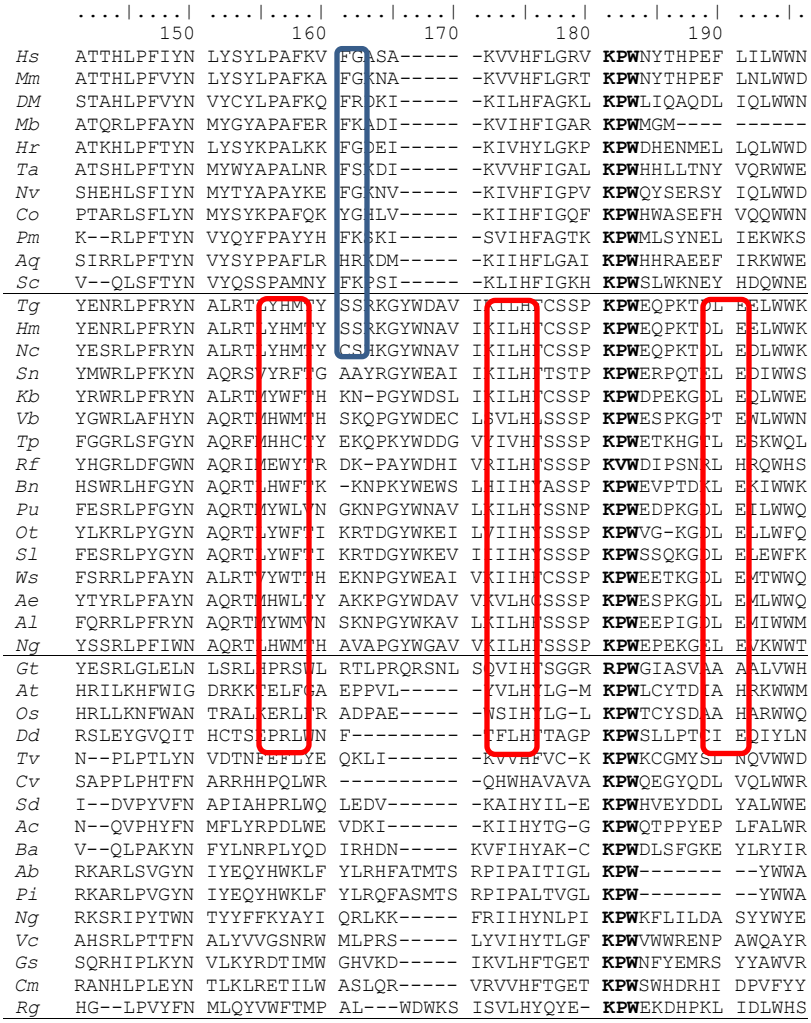
A. Gat1-like sequences from PgtA containing Protists	**B. Glycogenin-like sequences**
Toxoplasma gondii EPR60889.1	*Trichoplax adhaerens (*E^{-27}*)* XP_002116183.1 *Amphimedon queenslandica* (E^{-30}) XP_003383748.1 *Nematostella vectensis* (E^{-24}) XP_001625718.1 (Simplest animals)
Hammondia hammondii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

Fig. S2.7. Summary of Gat1-related sequences selected for phylogenetic analysis (basis for Fig. 3B). The best scoring hits (based on BLAST) from different categories of Gat1-like sequences were selected for manual alignment and phylogenetic analysis. (A) Predicted Gat1 sequences, from protists that have PgtA-like sequences but not AgtA-like sequences. (B) Glycogenin and glycogenin-like sequences. (C) Closest CAZy GT8 sequences from vascular plants. (D) Closest CAZy GT8 sequences from organisms

(protists) that possess Gnt1-like but not PgtA-like sequences. (E) Closest CAZy GT8 from protists that possess PgtA-like sequences but lack apparent Gat1. (F) Closest CAZy GT8 sequences from prokaryotes. Expect values, gene IDs, and known functions are indicated.

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

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



Species names. Sequence IDs are in Fig. S7.

Glycogenin-like:

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

Gat1-like

Tg: Toxoplasma gondii
Hm: Hammondia hammondi
Nc: Neospora canimum
Sn: Sarcocystis neurona
Kb: Karenia brevis
Vb: Vitrella brassicaforma
Tp: Thalassiosira pseudonana

Rf: ReticulomYxa folisa
Bn: Bigelowiela natans
Pu: Pythium ultimum
Ot: Oxytricha trifallax
Sl: StylonYchia lemnae
Ws: Ectocarpus siliculosus
Ae: Aphanomyces euteiches
Al: Albugo laibachii
Ng: Nannochloropsis gaditana
Other CAZy GT8 family
Gt: Guillardia theta
At: Arabidopsis thaliana
Os: Oryza sativa
Dd: Dictyostelium discoideum
Tv: Trichomonas vaginalis
Cv: Chlorella variabilis
Sd: Saprolegnia diclina
Ac: Acanthamoeba castellanii
Ba: Bigelowiella natans GT8
Ab: Albugo laibuchi GT8
Pi: Phytophthora infestans
Ng: Naegleria gruberi
Vc: Volvox carteri
Gs: Galdieria sulphuraria
Cm: Cyanidioschyzon merolae
Rg: Rhizobium gallicum

Figure S2.8. Alignment of glycogenin-like, Gat1-like, and other CAZy GT8 sequences used to construct the phylogenetic tree in Fig. 2.2. The amino acid sequence of Gat1-like proteins described in Fig. S2.7 (middle panel) were aligned with the amino acid sequences of representative known and predicted glycogenins (top panel) or CAZy GT8 sequences (bottom panel) as described in "Experimental Procedures". Species names are spelled out at the bottom, and sequence sources are listed in Fig. S2.7. Amino acids are color-coded with respect to chemical similarities that guided the alignments, giving preference to the registration of hydrophobic residues: green, hydrophobic; blue, acidic; dark red, basic; black, polar; bright red, secondary structure breaking (P or G). Positions occupied by identical amino acids across all the organisms are bolded. Unique motifs that are specific for glycogenins are boxed in blue color, and Gat1-specific motifs are boxed in red.

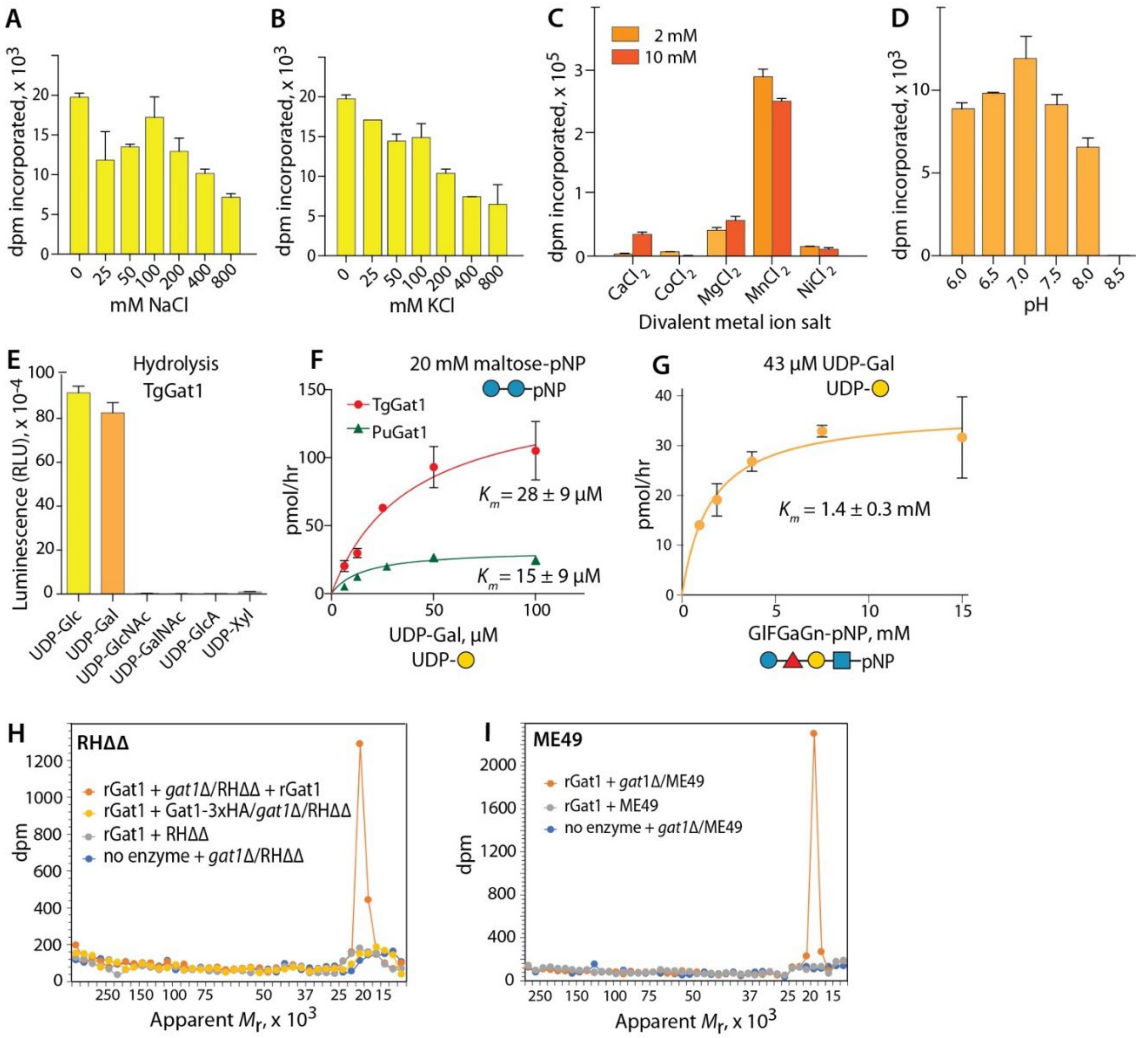
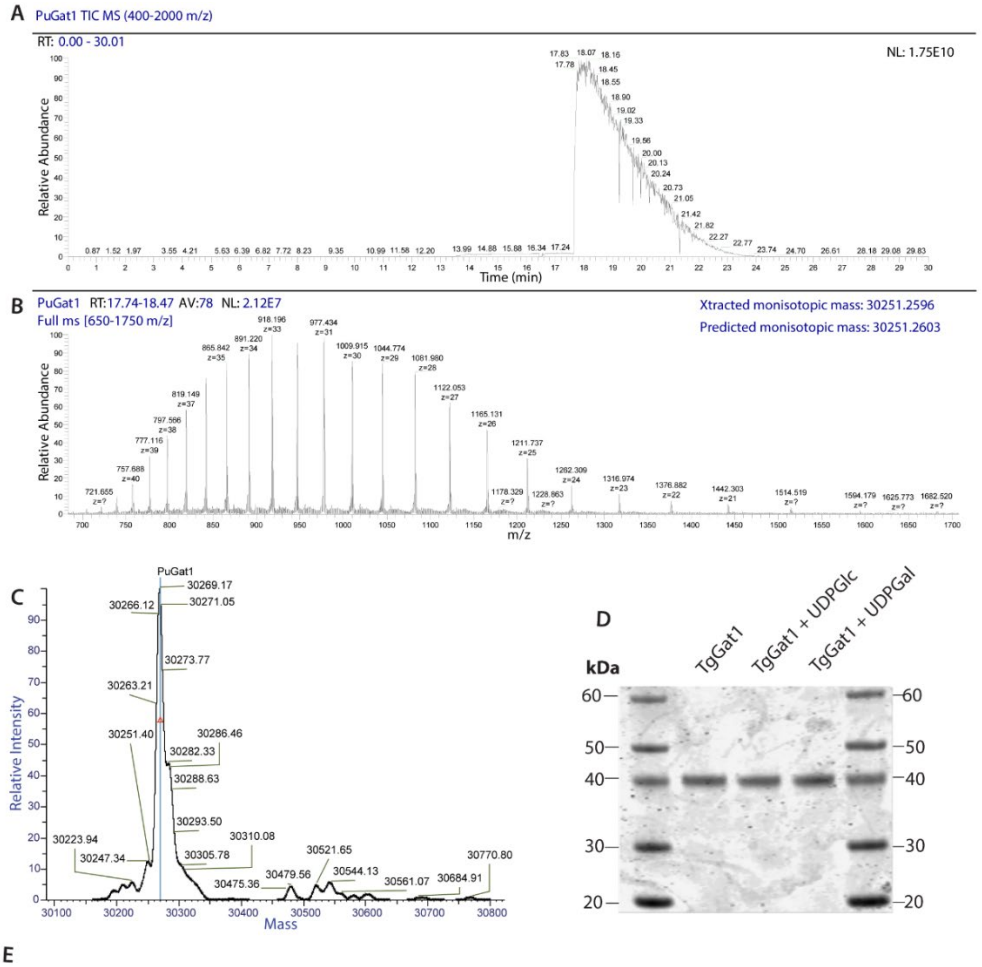


Figure S2.9. Characterization of Gat1 enzyme activity and biochemical complementation of *T. gondii* extracts (supports Fig. 2.4). TgGat1 glycosyltransferase activity was assayed using 20 mM maltose-pNP as an acceptor in the presence of 4 µM UDP-Glc, 2 mM $MnCl_2$, pH 7.0, and varying concentrations of NaCl or KCl (A, B). (C) TgGat1 was assayed using 20 mM maltose-pNP in the presence of 5.2 µM UDP-Gal, no added salt, 2 mM $MnCl_2$, pH 7.0, and the indicated divalent metal ions. (D) TgGat1 was assayed using 20 mM maltose-pNP, 8 µM UDP-Gal, no added salt, 2 mM $MnCl_2$, at different pH values. (E) Donor specificity of TgGat1, based on the UDP-Glo assay in an overnight reaction that consumed all UDP-Gal. (F) UDP-Gal concentration dependence of TgGat1 and PuGat1 Gal-transferase activity toward 20 mM maltose-pNP. Symbols are after Fig. 1. (G) Concentration dependence of TgGat1 Gal-transferase activity on GlFGaGn-pNP concentration. Error bars represent ±S.D. of 3 technical replicates of the same reaction. (H, I) Biochemical complementation to detect Gat1 substrates. Desalted S100 extracts from strains RHΔΔ, *gat1*Δ/RHΔΔ, ME49 and *gat1*Δ/ME49 were reacted with recombinant Gat1 (rGat1) in the presence of UDP-[^{3}H]Gal, and the product of the reaction was separated on an SDS-PAGE gel which was divided into 40 slices for liquid scintillation counting.



E

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

Figure S2.10. TgGat1 lacks auto-glycosylation activity. PuGat1 and TgGat1 were prepared in *E. coli*, purified to near homogeneity (Fig. 2.4A), and analyzed by nLC/MS analysis. (A) Total ion current for elution of PuGat1 in a gradient of acetonitrile from a C4 column. (B) Mass spectrum showing multiply protonated species. Xtract deconvolution yielded virtually only one species with an M_r 30251.2596, which closely matched the predicted theoretical monoisotopic mass of M_r 30251.2603 (error= 0.02 ppm). (C) Deconvolution of data in panel B using the ReSpect algorithm in BioPharma to yield a measurement of the average mass. (D) SDS-PAGE and Coomassie blue staining of TgGat1, before and after incubation with UDP-Glc or UDP-Gal for 30 min. (E) Summary of average mass measurements of TgGat1 and PuGat1 based on ReSpect deconvolution. After isolation from *E. coli*, both TgGat1 and PuGat1 yielded predominantly only the unmodified versions of the recombinant proteins, with M_r 39051.9687 for TgGat1 (theoretical average mass: 39051.9161, error= 1.3 ppm) and M_r 30269.1738 for PuGat1 (theoretical average mass: 30269.2901, error= 3.8 ppm). Their masses were essentially

unaffected by *ex vivo* reaction in the presence of UDP-Gal or UDP-Glc.

Chemical shifts* of starting tetra-saccharide:
αD-Glc(1-3)αL-Fuc(1-2)βD-Gal(1-3)αD-GlcNAc-O-pNP

Proton	H1	H2	H3	H4	H5	H6,6'	other
αGlc	5.23	3.56	3.81	3.44	3.89	3.82,3.96	
-3)αFuc	5.23	3.97	3.79	4.00	4.36	1.26	
-2)βGal	4.70	3.63	3.85	3.89	3.56	3.75,3.82	
-3)αGlcNAc	5.24	4.13	4.12	3.63	3.75	3.82,3.95	Ac 2.05

Carbon	C1	C2	C3	C4	C5	C6	other
αGlc	103.4	71.6	75.6	72.3	75.1	63.2	
-3)αFuc	102.1	69.9	81.2	74.4	69.2	18.1	
-2)βGal	102.8	79.4	76.1	71.8	74.6	63.9	
-3)αGlcNAc	101.8	57.2	79.4	71.0	78.6	63.1	Ac 25.1

Chemical shifts** of penta-saccharide:
αD-Gal(1-3)αD-Glc(1-3)αL-Fuc(1-2)βD-Gal(1-3)αD-GlcNAc-O-pNP

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

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

*Proton shifts in ppm referenced to DSS in the pentasaccharide sample. Carbon shifts in ppm derived from indirect referencing. Values for the tetrasaccharide were based on identical peaks in both samples.
** The additional chemical shifts for the terminal α-galactosyl residue and some of the penultimate 3-linked α-glucosyl are entered. Other peaks could either not be clearly distinguished (nd) or were identical (s) to the tetrasaccharide in the mixture.

Figure S2.11. Chemical shifts of Gat1 substrate and reaction product (supports Fig. 2.6)

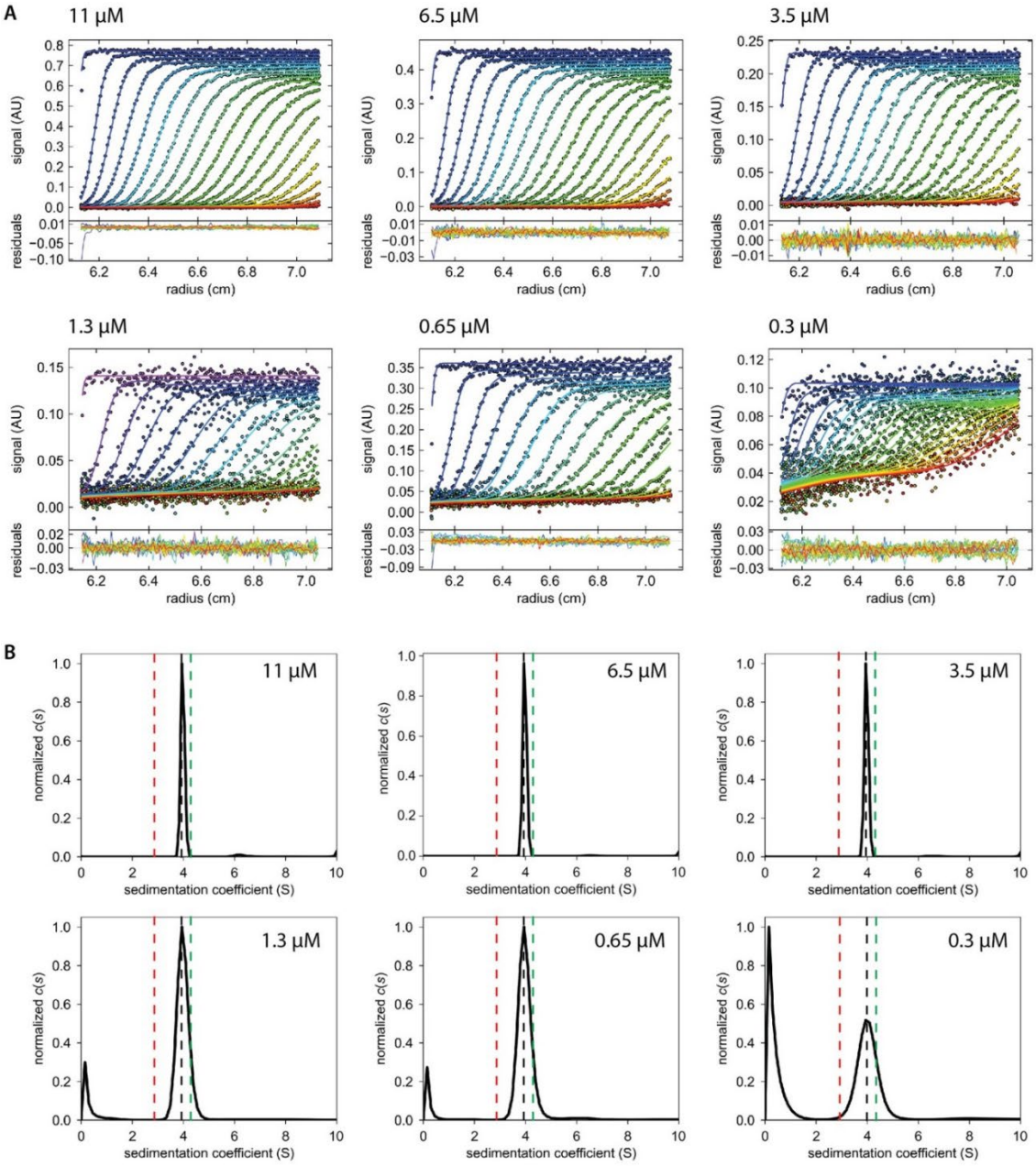


Figure S2.12. Gat1 is a dimer at all concentrations tested (related to Fig. 2.7C). (A) Sedimentation velocity profiles of different concentrations of PuGat1 are displayed with fit data and residuals. 11 μM, 6.5 μM, and 3.5 μM concentrations were detected at 280 nm, 1.3 μM data were collected at 230 nm, and 0.65 μM and 0.3 μM data were collected at 220 nm. (B) Data modeled as continuous *c*(*s*) distributions are shown (normalized to a value of 1 for the tallest peak). Black dashed line represents the determined S-value, and the red and green dashed lines respectively represent the predicted monomer and dimer S-values. The peak appearing at a near-zero S-value at the lower concentrations may be due to a buffer mismatch that became apparent at lower wavelengths.

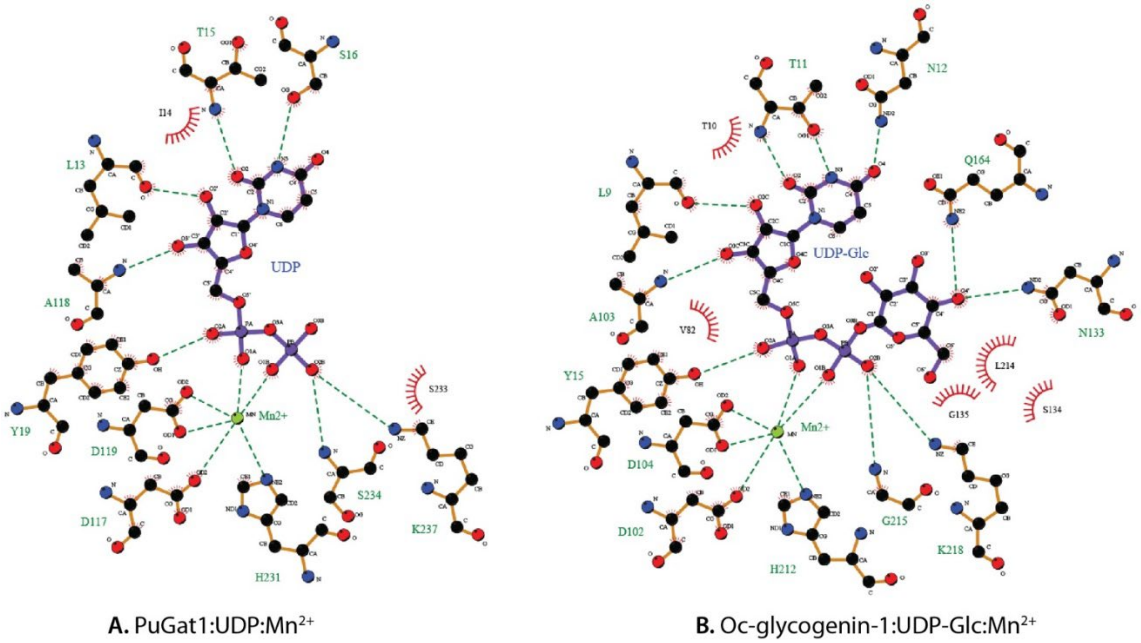


Figure S2.13. Gat1 and glycogenin coordinate UDP and Mn^{2+} in similar fashion (related to Fig. 2.8). PuGat1:UDP:Mn^{2+} (A) and Oc-glycogenin-1:UDP (PDB 1LL2) (B) are displayed as Ligplots[87]. Green dotted lines represent the interactions between the protein and the ligand, and red arcs represent packing interactions.

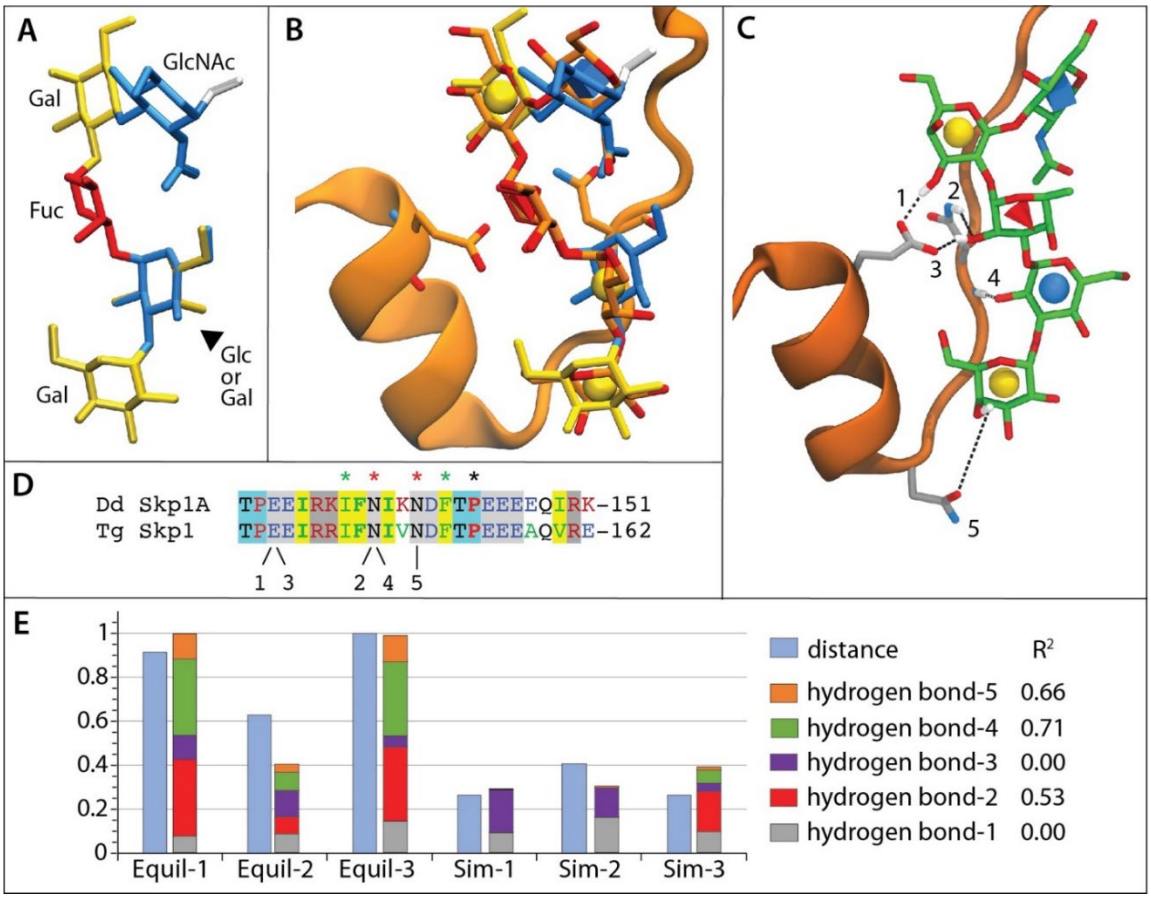


Figure S2.14. The *T. gondi* glycan/Skp1 relationship is reminiscent of *D. discoideum* (related to Fig. 2.10). (A) Superimposition of the two energy-minimized glycan structures produced by the Glycam webserver (63). Residues are colored according to the SNFG system. The differing Glc (blue) and Gal (yellow) residues (arrowhead) mark the difference between *Toxoplasma* and *Dictyostelium* glycans, respectively. (B) Superposition of the glycans in the context of Skp1 (orange ribbon); note that the linkage to Hyp is not shown. (C) Illustration of hydrogen bonds present at >25% occupancy over all simulations (1.5 µs) in *Toxoplasma* Skp1. (D) Comparison of amino acid sequences of TgSkp1 and DdSkp1 over the region depicted. Red asterisks indicate residues involved in hydrogen bonds that correlate best with extension of helix-8 (see panel E), green asterisks indicated residues that contribute most to non-polar packing interactions (see Table 2), and the black asterisk indicates the attachment site after hydroxylation. Residues are labeled from below according to the hydrogen bond with which they are associated. (E) Summary of the six 250-ns trajectories (3 pre-equilibrated; 3 were not). Left bars of each pair summarize the average distances for each trajectory between C156, near the C-terminus, to the center of mass of residues 1-136 (dashed green line in Fig. 10A), scaled to the highest average distance (Equil-3, in which the average value was >50 Å in the observed range of 18-61 Å for at least 85% of the time sampled at 0.1 ns increments). Right bars summarize the 5 most frequent hydrogen bonds between the glycan and Skp1, normalized to the highest level of hydrogen bonds observed in a single trajectory (Equil-1, in which at least one of the hydrogen bonds was occupied >99.7% of the time sampled in 0.1 ns increments). At the right is shown a time-resolved analysis of the correlation of helix-8 extension with the

occupancy of each hydrogen bond over the entire 1.5 μs of simulation time, based on the Pearson's correlation coefficient (linear regression R^2).

Purpose	Code name	Primer name	Primer sequence	Location
gat1 disruption in RHΔΔ and complementation in RH	Fa	a) Gat1F1 5'-flank 5'-end 5'	GGGGGCCCAACCAGCGGATCTTCTGAAC (ApaI)	Tg*gat1* homologous recombination disruption & complementation plasmids
	Ra	a') Gat1R1 5'-flank 3'-end 5'	GGCTCGAGACGCGTTGAGCGATTGA (Xho1)	
	Fb	b) Gat1F2 3'-flank 5'-end 5'	GCTCTAGAGAGGGAGAACCAAAGTGATGAT (XbaI)	
	Rb	b') Gat1R2 3'-flank 3'-end 5'	CGCGGCCGCTGCGTAGAACACAAGGAGAAC (NotI)	
PCR confirmation for *gat1* disruption in RHΔΔ				
PCR1	Fc	Forward	TACCCTGTTGACGGACAATT	Tg*gat1* genomic sequence
	Rc	Reverse	CTTTGCTGGTTGTTCCCAAG	
PCR2	Fd	Forward	GAACCGAATGACAACGCATTAC	
	Rd	Reverse	AGTCGCGGAACATCTCGTTGAAGT	HXGPRT sequence
PCR3	Fe	Forward	ATTTGCATCCTGAAAGGCTCTCGC	
	Re	Reverse	TCTGAAATGGAGTCGCCTTG	Tg*gat1* genomic sequence
Dual guide CRISPR plasmid for *gat1* disruption in RH, Me49-RFP				
NsiI PCR	Ff	Plasmid 3 FOR	CGTGGGGATGCATTCACCGCGGCCACATGTTG	Dual guide *gat1* CRISPR disruption plasmid
	Rf	Plasmid 3 REV	GCGATGAGCGCAAGCCGTCCTGAGTTACG	
dg plasmid sequencing	Fg	gRNA FOR	CAAAGTGCGCGAGTTGAAATCG	
	Rg	gRNA REV	GAGACGATGATTCCTGATCACTCCG	
PCR confirmation for *gat1* disruption in RH and Me49-RFP	Fh	Gat1 63 seq Fw (P1)	CGTACGCTACCCTGTTGACG	Tg*gat1* genomic sequence
	Rh	Gat1 968 Seq Rv (P2)	AGAATCAGTTGGCACAGTGCC	
	Fi	DHFR F/R Fw (P4)	CCATTGCGGTGTCGTGGATT	DHFR sequence
	Ri	DHFR RO Rv (P3)	CCCCTGTGTCCTTTATCGAAG	
Complementation plasmid sequencing	Fl	TgGat1 seq Fw	GGACTGTTTCACCAAACTGCGTGTGTG	Tg*gat1* genomic sequence
	Rl	Gat1 3'UTR Rv	CTAGTCAGTCCCTAAGGCTAGT	
Ty tag insertion on complementation plasmids				
Tg*gat1* Ty Tag insertion	Fn	TgGat1-Ty Fw	GAAGTACACACAAACCAAGACCCACTAGACTAGTGGAGGGA	Tg*gat1* genomic sequence & Ty tag
	Rn	TgGat1-Ty Rv	GTTTGTGTGTACTTCCACGATATCAGAATCAGTTGGC	
PCR confirmation for *gat1* complementation in RH	Fp	UPRT Fw	GTCCCAACGTCGCAAGTAA	UPRT genomic sequence
	Rp	UPRT Rv	ATGCGGACTTTCGGGTATTC	
	Fq	Gat1 cheek Fw	TGGGAACAACCAGCAAAGA	Tg*gat1* genomic sequence
	Rq	Gat1 FO Rv	GGGGTTGCAGCCTATGG	
TgGat1 *E. coli* expression plasmid	Fr	Gat1 Fw	AAGCTAGCATGTCTCCTCGGTACGCGTACGCT	Tg*gat1* genomic sequence & pET15b expression plasmid
	Rr	Gat1 Rv	AAGGATCCCTACACGATATCAGAATCAGTTGGCACAG	
DHFR amplicon with 45 bp *gat1* arms for CRISPR disruption in RHΔΔ	Fs	63 Fw_dhfr Fw	CGGACAATTCTTTCTACTATGGTGTCGAGGCACTGCTCAAG TCACAAGCTTCGCCAGGCTGTAAAT	Tg*gat1* and DHFR
	Rs	968 Rv_dhfr Rv	ATCAGAATCAGTTGGCACAGTGCCCGTAAGGAAGACTTTCC ACCACATCCTGCAAAGTGCATAGAAG	
Gat1-HA complementation plasmid in RHΔΔ	Ft	3HA Fw	GGTACCTACCCGTACGACGTC	pUPRT a1 WT cDNA shuttle Vector-Tub1-3xHA
	Rt	Tub-5'UTR Rv	GGCGCGCCGTGTCGAAAA	
	Fu	Tub-5'UTR-Gat1 Fw	CTTTTTCGACACGGCGCGCCATGTCTCCTCGGTACGCG	pmini-Tg*gat1* plasmid
	Ru	HA-Gat1 Rv	ACGTCGTACGGGTAGGTACCCACGATATCAGAATCAGTTGG	
PCR confirmation for Gat1 complementation in RH	Fv	UPRT 5'Arm Fw	GCTGTGCCTAGTATCGAAAGCTGTA	UPRT genomic sequence
	Rv	Gat1 at STOP Rv	CTACACGATATCAGAATCAGTTGGCACA	Tg*gat1* genomic sequence
	Fl	TgGat1 seq Fw	GGACTGTTTCACCAAACTGCGTGTGTG	
	Rw	UPRT 3'Arm Rv	CGACGTCACTGTACGACATCC	UPRT genomic sequence

Table S2.1. List of primers

Data collection	PuGat1:UDP:Pt^{2+} (PDB_6MW5)	PuGat1:UDP:Mn^{2+} (PDB_6MW8)
Wavelength (Å)	1.85	1.0
Space group	$P4_2\,2_1\,2$	$P4_2\,2_1\,2$
Unit cell dimensions (a, b, c)	83.78, 83.78, 75.84, 90.00, 90.00, 90.00	84.06, 84.06, 76.08, 90.00, 90.00, 90.00
Completeness (%)	97.4 (94.8)[a]	99.9 (99.8)[a]
Total number of reflections	396057 (13045)	800853 (52625)
Unique reflections	29424 (2114)	28125 (2042)
Redundancy	13.5 (6.2)	28.4 (25.8)
I/σ(I)	28.6 (1.21)	36.94 (1.85)
R_{meas}[b] (%)	6.1 (145.4)	6.2 (203.9)
CC1/2[c] (%)	100.0 (49.0)	100.0 (65.5)
Refinement		
Resolution (Å)	2.1	1.76
R_{work}/R_{free}	0.196/0.242	0.181/0.210
No. of atoms Protein/ Ligand / Water	1957/30/54	1949/42/122
Wilson B-factor ($Å^2$)	45.2	39.9
B-factors ($Å^2$) Protein/ Ligands and Water	45.2/45.66	39.4/44.2
Stereochemical Ideality		
Bond lengths (Å)	0.006	0.006
Bond angles (°)	0.786	0.784
φ, ψ Most favored (%)	97	99
φ, ψ Additionally allowed (%)	3	1
SAD Phasing statistics		
Heavy atom sites	1	
Figure of merit	0.31	

[a] Values in parentheses are for highest-resolution shell
[b] R_{meas} is the redundancy independent merging R-factor of Karplus and Diederichs (86)
[c] $CC_{1/2}$ is the percentage of correlation between intensities from random half-data sets

Table S2.3. Crystallographic data

CHAPTER 3

SKP1 DIMERIZATION CONCEALS ITS F-BOX PROTEIN BINDING SITE [2]

[2] **Kim HW**‡, Eletsky A‡, Gonzalez KJ van der Wel H, Strauch EM, Prestegard JH, West CM. 2020. *Biochemistry*. **59**:1527-1536. Reprinted here with permission of the publisher. ‡ denotes co-first authorship.

3.1 Abstract

Skp1 is an adapter that links F-box proteins to cullin-1 in the Skp1/cullin-1/F-box (SCF) protein family of E3 ubiquitin ligases that targets specific proteins for polyubiquitination and subsequent protein degradation. Skp1 from the amoebozoan *Dictyostelium* forms a stable homodimer *in vitro* with a K_d of 2.5 μM as determined by sedimentation velocity studies, yet is monomeric in crystal complexes with F-box proteins. To investigate the molecular basis for the difference, we determined the solution NMR structure of a doubly truncated Skp1 homodimer (Skp1ΔΔ). The solution structure of the Skp1ΔΔ dimer reveals a 2-fold symmetry with an interface that buries ~750 $Å^2$ of predominantly hydrophobic surface. The dimer interface overlaps with subsite-1 of the F-box interaction area, explaining why only the Skp1 monomer binds F-box proteins (FBPs). To confirm the model, Rosetta was used to predict amino acid substitutions that might disrupt the dimer interface, and the F97E substitution was chosen to potentially minimize interference with F-box interactions. A nearly full-length version of Skp1 with this substitution (Skp1ΔF97E) behaved as a stable monomer at concentrations up to 500 μM and actively bound a model FBP, mammalian Fbs1, which suggests that the dimeric state is not required for Skp1 to carry out a basic biochemical function. Finally, Skp1ΔF97E is expected to serve as a monomer model for high-resolution NMR studies previously hindered by dimerization.

3.2 Introduction

The Skp1/Cullin-1/F-box (SCF) protein family of E3 ubiquitin ligases is an important mediator of protein turnover in yeast/fungi, higher plants, and animals, owing to the role of polyubiquitination in serving as a signal for recognition and degradation in the 26S-proteasome. Evidence supports the importance of the SCF complex in the protist kingdom as well,[1] where a novel posttranslational modification has been discovered in Skp1 orthologs from groups as diverse as amoebozoa (*Dictyostelium discoideum*), apicomplexans (*Toxoplasma gondii*), and oomycetes (*Pythium ultimum*).[2] Protist Skp1 is subject, in the presence of sufficient O_2 and α-ketoglutarate, to hydroxylation of a Pro-residue that lies on the backside of subsite-2 of the F-box binding domain of Skp1. Once Skp1 is hydroxylated, the hydroxyproline (Hyp) residue is recognized and glycosylated by a series of glycosyltransferases, resulting in the assembly of a canonical pentasaccharide. Mutational studies show that both hydroxylation and full glycosylation are required for optimal O_2-sensing in *Dictyostelium* and *Toxoplasma*.[1,3] Biophysical and computational studies have generated a model by which the relatively organized structure of the pentasaccharide organizes the surrounding intrinsically disordered region of *Dictyostelium* Skp1 in such a way as to be more conducive to binding the F-box domain of FBPs,[4] and recent studies indicate that this model is also relevant to Skp1 from *Toxoplasma*.[3] Confirmatory biophysical studies using NMR are hampered by the dimeric state of glycosylated Skp1 (GGFGGn-Skp1) because of its relatively large size, 324 amino acids.

Recombinant guinea pig Skp1, whose sequence is identical across mammals, was previously reported to dimerize with a K_d of 1.1 μM.[5] This value is significantly below the estimated concentration of Skp1 (a.k.a. OCP2) in the inner ear tissues (2 mM) where

Skp1/OCP-2 was initially characterized, suggesting that dimerization might influence Skp1 activity in cells. Studies of *Dictyostelium* Skp1, where small angle X-ray scattering, gel filtration, and NMR studies confirmed its dimeric status at higher concentrations, indicate that glycosylation modestly inhibits dimerization.[6] We sought to investigate the significance of Skp1 dimerization by mapping its dimer interface. The structure of Skp1 from mammals, yeast and higher plants is known when it is bound to F-box proteins.[7,8] In contrast to the extensive sequence variations of F-box domains, the sequence and structure of Skp1 in these complexes is highly conserved. However, free Skp1 has defied structural characterization, potentially because of intrinsically disordered regions (including its C-terminal region that contributes to F-box domain recognition) that interfere with the formation of crystals for X-ray crystallography. At the same time, the Skp1 homodimer is too large for solution NMR studies without resorting to ^{2}H-isotope labeling. We have found that *Dictyostelium* Skp1A remains a dimer in the absence of both an internal disordered region that was originally removed to allow crystallization with F-box proteins,[7] and the predominantly disordered region that comprises the C-terminal subsite-2 of the F-box binding region. The doubly truncated Skp1ΔΔ dimer variant (2 × 118 aa) was sufficiently small to pursue high-resolution solution NMR structure determination using uniform ^{15}N- and ^{13}C-isotope labeling. We found that the homodimer interface overlaps with the F-box binding interface and confirmed the finding by Rosetta-guided mutagenesis, and discuss the implications for Skp1 function and future studies on its posttranslational regulation.

3.3 Materials and Methods

Expression Plasmids

The *E. coli* expression plasmid pET19b-Skp1AΔΔ was derived from pET19b-Skp1A[9] by site-directed mutagenesis, in which primers designed to bridge the deleted sequence as described in Figure S1C were used in a PCR reaction with Q5® High-Fidelity DNA Polymerase (New England Biolabs) to amplify the modified vector. After treatment with DpnI to destroy the original vector, the sample was cloned into *E. coli* strain BL21-Gold (DE3).

For improved recovery and purification of Skp1, the *Dictyostelium* Skp1A coding sequence was codon optimized for expression in E. *coli*, and appended with an N-terminal His_6-tag which, when excised by treatment with TEV protease, yielded the native sequence with an N-terminus of SMSL-, compared to the N-terminal SL- that occurs natively after removal of the start Met.[10] The cDNA ***(Figure S1A)*** was synthesized and provided in pUC57 by GenScript, excised using NcoI and BamHI, and ligated into the NcoI and BamHI restriction sites of pET19b, yielding pET19b-His_6DdSkp1A-optim. A second cDNA in which 12 internal amino acids (SPQGDDKKDEKR) were replaced with GGSG ***(Figure S1B)*** was synthesized and similarly ligated into pET19B yielding pET19b-His_6DdSkp1AΔLoop-optim. This plasmid was modified to generate an F97E point mutation by site-directed mutagenesis, in which the indicated primers were used in a PCR reaction to amplify the modified vector as above.

Expression and Purification of Skp1 constructs

Skp1 and Skp1ΔΔ, which each lacked an affinity tag to ensure native-like behavior, were purified from *E. coli* to near homogeneity under non-denaturing conditions (DEAE,

phenyl, Q and S200 Superdex columns) as described previously.[6] Sample purity and integrity was assessed by SDS-PAGE and Coomassie blue staining to be >90%.

E. coli cells expressing His_6Skp1Δ or His_6Skp1ΔF97E were incubated at 37 ºC in 2 × 1 L of Terrific Broth medium in the presence of 100 µg/ml ampicillin. At an OD_{600} of 0.6, protein expression was induced by addition of 125 µM isopropyl 1-thio-β-D-galactopyranoside (IPTG) at 20 ºC. After 12-16 h, bacteria were collected by centrifugation at 5000 × *g* for 10 min and resuspended in 50 mM Na^+/K^+ phosphate (pH 7.8), 300 mM NaCl, 5 µg/ml aprotinin, 5 µg/ml leupeptin at 4 ºC. Cells were lysed using a probe sonicator (model 500, Thermo Fisher Scientific) for a total sonication time of 5 min. The lysate was centrifuged at 25,000 × *g* for 45 min at 4 ºC, and the supernatant was immediately applied to a 1.5-ml column of Co^{2+} Talon resin (Clontech) pre-equilibrated at 4 ºC in the buffer described above. The column was washed successively with the same buffer supplemented with either 1 M NaCl, 10% glycerol, or 5 mM imidazole. Protein was eluted with buffer containing 300 mM imidazole, and dialyzed against 50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 1 mM EDTA, 2 mM β-mercaptoethanol. The sample was incubated overnight at 22 °C with His_6TEV protease to cleave the His_6-tag from Skp1, and the sample was re-applied to the Talon resin. The flow-through was concentrated to 1.5 ml using a spin concentrator (Amicon) with a 3 kDa molecular weight cut-off. The concentrated sample was further purified over a Superdex 200 Hi-load 16/60 gel filtration column (GE Healthcare) pre-equilibrated with 20 mM potassium phosphate (pH 7.4), 50 mM KCl, and 0.5 mM TCEP. The sample was estimated to be >95% pure by SDS-PAGE and staining with Coomassie blue.

Stable isotope labeled Skp1ΔΔ and His_6Skp1ΔF97E were prepared by expression in *E.*

coli in the presence of isotope enriched minimal media as previously described.[9] His_6Skp1ΔF97E was uniformly enriched with ^{15}N, and Skp1ΔΔ was labeled with ^{15}N and ^{13}C. The final Skp1ΔΔ NMR sample (105 µl in a 3-mm Shigemi tube) contained a 1:1 mixture of ^{15}N,^{13}C-Skp1ΔΔ and natural abundance Skp1ΔΔ at ~1.0 mM concentration in 20 mM MES-NaOH (pH 6.0), 50 mM NaCl, 5 mM dithiothreitol, 0.05% NaN_3. ^{15}N-labeled Skp1ΔF97E NMR samples (300 µl in 5-mm Shigemi tubes) were prepared at concentrations of 100 µM and 500 µM in the same buffer. All NMR samples contained 10% D_2O for spectrometer lock.

Analytical Ultracentrifugation

Protein was quantified based on molar absorptivity calculated from the protein sequence using ProtParam.[11] Samples were loaded into 12-mm double-sector Epon centerpieces equipped with quartz windows and equilibrated for 2 h at 20 °C in an An60 Ti rotor. Sedimentation velocity data were collected using an Optima XLA analytical ultracentrifuge (Beckman Coulter) at 50,000 rpm at 20 °C. Data were recorded with absorbance optics at 280 nm, 230 nm or 215 nm in radial step sizes of 0.003 cm. SEDNTERP[12] was used to model the partial specific volume as well as the density and viscosity of the buffer. SEDFIT[13] was used to analyze sedimentation data. All data were modeled as continuous *c(s)* distributions and were fit using baseline, meniscus, frictional coefficient, and systematic time-invariant and radial-invariant noise. Predicted sedimentation coefficient (s) values for Skp1 monomer and dimer states were calculated using HYDROPRO[14] with a homology model generated on the ROBETTA server.[15] Data fit and *c(s)* plots were generated using GUSSI.[16] Weight-averaged S values (S_w) at each concentration were determined by integrating *c(s)* distributions. Constructed S_w isotherms

were fitted with a A+A⇌AA self-association model using SEDPHAT to calculate the dissociation constant.[17]

NMR Spectroscopy and Structure Determination of Skp1ΔΔ

NMR spectra for Skp1ΔΔ were acquired at 35 °C using a Bruker AVANCE NEO 800 MHz spectrometer equipped with a 5-mm cryogenic TCI $^{1}H\{^{13}C,^{15}N\}$ probe, and an Agilent VNMRS 600 MHz spectrometer equipped with a 3-mm cryogenic $^{1}H\{^{13}C,^{15}N\}$ probe. NMR spectra for 100 μM and 500 μM Skp1ΔF97E samples were acquired using a Bruker AVANCE NEO 900 MHz spectrometer equipped with a 5-mm cryogenic TXO $^{13}C,^{15}N\{^{1}H\}$ probe, and the 600 MHz spectrometer equipped with a 5-mm cryogenic $^{1}H\{^{13}C,^{15}N\}$ probe. The acquired NMR spectra are summarized in Table S1. NOE mixing times were 70 ms for $^{13}C/^{15}N$-edited $[^{1}H,^{1}H]$-NOESY and 120 ms for $^{13}C/^{15}N$-filtered $^{13}C/^{15}N$-edited $[^{1}H,^{1}H]$-NOESY experiments. Fourier transform was performed with TopSpin (Bruker BioSpin) for Bruker NMR data, and NMRPipe[18] for Varian NMR data. ^{1}H chemical shifts were referenced relative to 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS), and ^{13}C and ^{15}N chemical shifts were referenced indirectly via gyromagnetic ratios. 2D and 3D NMR spectra were analyzed using CARA.[19]

Relaxation delays were 0.1, 0.2, 0.3, 0.4, 0.7, 1.0, 1.5, and 2.0 seconds in 1D proton-detected ^{15}N T_1 experiments, and 10, 30, 50, 70, 90, 110, 130, 150 and 170 ms in 1D proton-detected ^{15}N T_2 experiments acquired for the 500 μM Skp1ΔF97E sample. 1D ^{15}N T_1/T_2 relaxation spectra were processed and analyzed with VnmrJ v4.2 (Agilent Inc). Macro "tc" (wiki.nesg.org) was used to integrate the regions between ^{1}H chemical shifts of 8.8 and 9.9 ppm, determine average ^{15}N T_1 and T_2 relaxation times via exponential fitting, and calculate global rotational correlation time τ_c.

Sequence-specific backbone and side-chain resonance assignments for Skp1ΔΔ were derived using CARA based on existing resonance assignments of full-length Skp1.[9] Structure calculation of the Skp1ΔΔ homodimer was performed using CYANA[20] based on ^{1}H-^{1}H upper distance constraints derived from ^{13}C/^{15}N-edited [^{1}H,^{1}H]-NOESY, as well as backbone φ and ψ and side-chain χ_1 dihedral angle restraints from TALOS-N.[21] Automated NOESY peak assignment was performed initially with CYANA, with 22 manually assigned intermolecular 5 Å ^{1}H-^{1}H upper distance constraints (Table S2) applied after cycle 1 of simulated annealing. These intermolecular ^{1}H-^{1}H upper distance constraints were derived from selected strong peaks in a ^{13}C/^{15}N-filtered ^{13}C/^{15}N-edited [^{1}H,^{1}H]-NOESY spectrum. Only those peaks that could be unambiguously assigned and could not be explained by intramolecular contacts were chosen. After several rounds of iterative refinement of NOE peak assignments and calibration of distance constraints, the final structure calculation was performed with CYANA. Stereospecific resonance assignment of methylene proton spins and methyl groups of Leu and Val residues were obtained with the GLOMSA module of CYANA. Out of 100 calculated conformers, 20 conformers with the lowest target function values were selected for subsequent refinement in explicit water bath using the program CNS[22] with upper distance constraints relaxed by 5%. The structure statistics are outlined in Table 1.

Predicted rotational correlation time τ_c was calculated for Skp1ΔΔ and Skp1ΔF97E using HYDRONMR,[23] assuming water viscosity as 0.0072 cP at 35 °C. The lowest energy conformer of Skp1ΔΔ NMR ensemble was used, and only the first chain was used to calculate τ_c of a hypothetical Skp1ΔΔ. Homology model of Skp1ΔF97E monomer for τ_c calculation was generated with SWISS-MODEL based on X-ray structure of human Skp1

(PDB ID 3L2O) as a template. A hypothetical dimer model of Skp1ΔF97E was then produced by structural alignment of individual subunits to those of Skp1ΔΔ dimer in Chimera (http://www.rbvi.ucsf.edu/chimera),[24] followed by adjustment of backbone dihedral angles to eliminate clashes in the C-terminal region and energy minimization.

Analysis of the Dimer Interface

Computational investigation of the dimer interface was performed using the conformer with the lowest CYANA target function of the initial NMR structures of Skp1ΔΔ. To prevent inaccurate predictions due to small clashes in the structure, the protein was prepared using the standard Rosetta optimization protocol, "FastRelax".[25,26] Briefly, five cycles of rotamer packing and minimization were carried out, ramping up the repulsive weight in the scoring function within each cycle. After three rounds of symmetrical FastRelax with atom-atom pair distance constraints, the quality of the generated models was validated with the Molprobity web service.[27] The lowest scoring structure based on the Rosetta energy score and the Molprobity score was selected for mutational analysis.

To identify mutations that could disrupt the Skp1 dimerization, an all amino acids-scanning mutagenesis was carried out across the homodimer interface *in silico*. The "flex ddG" protocol implemented in RosettaScripts[28,29] was used to model and predict the effect of the mutations on the binding free energy of the complex. All parameters in the protocol were set up according to the default values described previously.[28] Overall, the flex ddG method takes advantage of the Rosetta backrub approach[30] to sample side-chain and backbone conformational changes around the mutated position. Once the backrub ensembles are generated, the structures are optimized by sidechain repacking and torsion minimization. The interface ΔΔG score corresponds to the average difference in binding

free energy between the mutant structure and the wild-type complex. Stabilizing mutations are defined as those with interface ΔΔG scores < −1.0 Rosetta energy units (REU), while destabilizing mutations are assigned to interface ΔΔG scores > 1.0.[28]

To determine which dimer-destabilizing substitutions do not perturb the stability of the individual monomers, the change in the total free energy of the monomer was estimated for all mutations. This analysis was performed with the current state-of-the-art Rosetta ΔΔG protocol, "cartesian_ddg".[31] As a preparation step, the wild-type monomer was relaxed in Cartesian space, using backbone and sidechain coordinates constraints. Since the input structure was previously refined with FastRelax, the restrained Cartesian-refinement was intended to adjust the (x,y,z) position of each atom without large deviations from the optimized structure. In this regard, coordinate constraints corresponding to harmonic restraints (Supplementary information- Scripts) were used to penalize large movements[32]. The model with the lowest Rosetta score was then used as input for the cartesian_ddg protocol. Within the cartesian_ddg application, the protein was relaxed again in the Cartesian space, allowing movement of only the backbone and sidechains around the mutated position.[31] All parameters in the cartesian_ddg protocol were configured as previously described.[31] The total ΔΔG score was finally considered as the difference in the total ΔΔG between the mutant and the wild-type monomer, multiplied by an energy scaling factor of 1.0/2.94. As above, stabilizing mutations correspond to total ΔΔG scores < -1.0, and destabilizing mutations refer to total ΔΔG scores > 1.0.

All Rosetta commands for this report were run with the same Rosetta static executable (RosettaCommons/main.git2019-03-07, version 4ab48a76160c888257155619edb9817845bd8a67). The protocols previously described can

be found at Supplementary information- Scripts.

Analytical Gel Filtration

Skp1Δ and Skp1ΔF97E with or without Fbs1 at a limiting concentration were subjected to Superdex 200 PC 3.2/30 gel filtration analysis using a Pharmacia SMARTSystem HPLC as previously described.[6]

3.4 Results

Characterization of the Skp1 Dimer

Sedimentation velocity experiments were conducted on *Dictyostelium* Skp1A (Skp1) that was recombinantly expressed without a peptide tag in *E. coli*. Over a concentration range of 0.5 – 45 μM , the samples yielded peaks at 1.8 S and 2.7 S ***(Figure 3.1A)***, values which are slightly less than the predicted S-values for monomer and dimer forms, 1.9 S and 2.8 S. The homology model used for predicting S-values assumed that the C-terminal region of Skp1 is organized as α-helices as occurs in complexes with F-box proteins. However, the C-terminal region of free Skp1 is predominantly disordered based on previous NMR studies,[9] which is expected to cause Skp1 to sediment more slowly and would explain the slight discrepancy between the observed and predicted S-values. The separate peaks in this non-equilibrium analysis indicate that interconversion between monomer and dimer states is slow relative to the time scale of sedimentation. A Skp1 dimer binding isotherm constructed using weighted S-values from across the concentration range yielded a dissociation constant for the Skp1 dimer of 2.5 μM ***(Figure 3.1B)***.

Skp1 has resisted crystallization and the large size of the full-length homodimer (324 amino acids) inhibited structure determination by NMR. To initiate mapping of the

dimer interface, we examined a truncated Skp1 variant, Skp1ΔΔ, which lacks the mainly disordered C-terminal F-box binding domain and an internal disordered loop that is frequently removed for Skp1/FBP crystallization ***(Figure 3.2A)***. We demonstrated that Skp1ΔΔ (118 × 2 amino acids) still forms a stable homodimer based on sedimentation velocity studies, which yielded an S-value of 1.9, in agreement with the 1.9 S-value predicted by HYDROPRO ***(Figure S3.2)***. Also, a 2D [^{15}N,^{1}H] HSQC spectrum of Skp1ΔΔ correlated well with 2D [^{15}N,^{1}H] TROSY of full-length Skp1 (data not shown), indicating that truncations in Skp1ΔΔ do not perturb the overall structure.

Solution NMR Structure of Skp1ΔΔ

Using a suite of standard NMR experiments ***(Table S2.1)***, we obtained complete sequence-specific assignments of backbone and side-chain ^{1}H, ^{15}N and ^{13}C resonances of Skp1ΔΔ ***(Table 3.1; Figure S3.3)***. These resonance assignments allowed us to pursue structure calculations based on ^{1}H-^{1}H distance constraints derived from 3D ^{15}N/^{13}C-edited [^{1}H,^{1}H] NOESY spectrum. To ensure proper modeling of the subunit interaction, we applied weak intermolecular distance ^{1}H-^{1}H constraints ***(Table S3.2)*** derived from NOE peaks identified in a separate ^{13}C/^{15}N-filtered and ^{13}C-edited [^{1}H,^{1}H] NOESY spectrum recorded with a sample of mixed U-^{15}N,^{13}C-labeled and natural abundance Skp1ΔΔ. Representative fragments of NOESY spectra are shown in Figure S3.4. Full attenuation of intra-chain NOE cross-peaks was not achieved in the ^{13}C/^{15}N-filtered spectrum ***(right strip in Figure S3.4)*** because of incomplete isotope incorporation (~85%). However, comparison with the corresponding strip from the NOESY spectrum without isotope filtering (strip on the left) allowed us to distinguish between peaks of comparable intensity (inter-chain) and peaks with significantly reduced intensity in the filtered set (intra-chain).

We obtained a high-quality solution NMR structure of Skp1ΔΔ ***(Table 3.1; Figure 3.2)***. A ribbon diagram of the best scoring structure of the top 20 conformers ***(Figure 3.2B)*** is shown in Figure 3.2C. The structure features a semi-parallel orientation of subunits with respect to their N- and C-termini, in contrast to the previously hypothesized dimerization model from SAXS analysis.[6] The dimer interface is organized as a four-helix bundle, with symmetrical packing contributions from residues of helices 5 & 6 of each chain ***(Figure S3.5)*** as identified using PISA.[37] The interface buries ~743 Å^2 of predominantly hydrophobic surface whose participating amino acids are labeled in Figure 3.2D. Superimposition of the corresponding Cα atoms with human Skp1 from a crystal structure in complex with the human FBP βTRCP revealed an RMSD of 1.1 Å ***(Figure S3.6)***, indicating that the structure of amino acids 1-125 of free Skp1 changes little when complexed as a monomer with FBPs. Furthermore, the homodimer interface involves the previously described subsite-1 of the binding site for F-box domains.[7] The physical overlap ***(compare Figures 3.2E, F)*** was quantified by modeling the F-box domain of human FBXW7 from a crystal structure with human Skp1, with Skp1ΔΔ. This revealed a ~650 Å^2 overlap of the homodimer and heterodimer interfaces, in the region of helices 5 and 6, and explains why only the Skp1 monomer is found in complexes with FBPs.

Conservation of the Dimer Interface

The two α-helices contributing to the homodimer interface, extending from L96 to I123, are highly conserved throughout phylogeny. Furthermore, each dimer contact residue except K117 is almost perfectly conserved from stramenopiles to humans ***(Figure S3.7)***, and the dimer interface region is immediately surrounded by Gly or Pro residues and length variations (indels), indicating that this region represents a functional unit under selective

pressure to remain intact. This is consistent with the finding that human Skp1 also dimerizes in this concentration range.[5]

Computationally-Guided Selection of a Skp1 Monomer Mutant

To test the dimer structure model, we searched for point mutations of interface amino acids that might destabilize the dimer. An initial alanine-scanning mutagenesis calculation suggested Phe97 Leu101, and Ile123 as potential destabilizing positions based on a weakening of the binding free energy of the complex (interface ΔΔG score > 2) ***(Figures 3.3A,B).*** All 20 amino acids were then substituted at each position to predict the effect of different side chains. Indeed, most substitutions at these 3 positions were destabilizing, and 85% of the variations at position 97 yielded interface ΔΔG scores greater than 1.5. The mutations F97G, F97D and F97E showed the highest dimer-destabilizing effect, with interface ΔΔG scores of 5.41, 5.23 and 4.87, respectively ***(Figure 3.3C)***.

To determine whether the dimer-destabilizing mutations might affect monomer stability, the total monomer ΔΔG was estimated for each substitution. Remarkably, most of the mutations at positions 97, 101, and 123 presented total ΔΔG scores between 1 and -1 ***(Figure S3.8)***. In particular, the highly dimer destabilizing F97E substitution is predicted to have a neutral effect on monomer stability with its ΔΔG score at 0.15.

Based on the above analysis, the conserved Phe97 was changed to Glu in the nearly full length Skp1Δ isoform. Glu was chosen over Asp to allow extension of the carboxyl group to provide better accessibility to solvent rather than clashing when binding an FBP. As predicted by the dimer interface model, Skp1ΔF97E eluted as a monomer based on gel filtration and was predominantly a monomer by AUC at 100 μM ***(Figure 3.4A)***. A 2D [^{1}H,^{15}N]-HSQC spectrum recorded for a ^{15}N-labeled Skp1ΔF97E sample exhibited peak

dispersion consistent with a well-folded protein ***(Figure 3.4B)***. The peak pattern was comparable to that of wild-type Skp1 considering that there are a number of differences including an N-terminal SM- extension, replacement of 12 amino acids with 4 different amino acids in the internal loop, and truncation after amino acid 125 (not shown). Based on average ^{15}N T_1 and T_2 relaxation times measured for all ^{1}H amide resonances between 8.8 and 9.9 ppm in a 500 μM sample of Skp1ΔF97E, the effective rotational correlation time (τ_c) was calculated at 9.4 ± 0.5 ns. This is close to the 10.9 ns value predicted for the monomer ***(Table 3.2)***, and contrasts with the previously reported 19.9 ± 2.2 ns value of the full-length protein.[4] Thus, Skp1ΔF97E remains predominantly monomeric even at high concentrations typical for solution NMR.

Skp1ΔF97E is Binding Competent with a Model FBP

Phe97 is conserved as a Phe or Tyr in known Skp1s ***(Figure S3.7)***. Analysis of Skp1 in crystal structures of complexes with 3 different FBPs (Tir1, βTRCP1, Fbs1) shows that, compared to the homodimer, Phe (or Tyr) resides in a different rotamer state with solvent exposure (not shown). This suggests that the F97E replacement will, though removing a favorable hydrophobic contact, not directly disrupt the FBP interaction. To test the functionality of Skp1ΔF97E, we used a Superdex200 gel filtration column to examine the elution profile of Skp1 in the absence and presence of a heterologous FBP, guinea pig Fbs1. As previously described[6] and replicated in Figure 4C, a mixture of Fbs1 and Skp1 eluted prior to the elution positions of either protein alone. As shown in Figure 4D, Skp1ΔF97E exhibited similar behavior. In addition, Skp1ΔF97E clearly eluted later than native Skp1, consistent with its monomeric character as described by AUC and its NMR-derived rotational correlation time. Thus Skp1(F97E) retained its FBP binding function,

though a possible reduction in affinity is not excluded by this analysis.

2.5 Discussion

Our findings confirm that dimerization is a highly conserved property of Skp1, based on similar dissociation constant values from organisms as phylogenetically distant as *Dictyostelium* and humans. Their measured K_d values range from 1.1 to 2.5 μM under *in vitro* conditions, though actual affinities in the cell may vary. These values are similar to the predicted total monomer Skp1 concentration in a mammalian cell line, ~2 μM.[38] The significance of dimerization is suggested by the exceptionally high degree of conservation of the contact residues ***(labeled D in Figure S3.7)***, and its ability to mask the hydrophobic character of this solvent exposed surface.

The Skp1 dimer interface occupies ~740 Å^2 of predominantly hydrophobic surface, a substantial area that overlaps subsite-1 of its FBP binding site ***(Figure 3.2E)***. In human Skp1, there is a lack of NMR resonance assignments in the corresponding interface region ***(residues 99-130 in human vs. residues 93-124 in Dictyostelium Skp1, Figure S7)***[39], potentially due to effects of unanticipated dimerization of human Skp1. The homodimerization model was supported by the predicted effect of an amino acid substitution within the interface, F97E, to destabilize the interaction ***(Figure 3.4A)*** without unfolding the protein ***(Figure 3.4B)***. Indeed, monomeric Skp1 maintained its ability to bind a model FBP, Fbs1 ***(Figure 3.4C, D)***, and was also competent to be enzymatically hydroxylated and fully glycosylated *in vitro* (unpublished data). A further contribution to the dimer interface from beyond the truncation site at residue 125 seems unlikely, because previous studies indicated substantial disorder for this region in the free dimer.[9]

Based on the new structure, the dimer interface also contributes to subsite-1 of the F-box binding region of Skp1, which explains why Skp1 is a monomer in complexes with FBPs.[8] Interestingly, 8 of the 13 alleles of Skp1 known to affect its function in budding and fission yeast have point mutations located on this region,[40-44] raising the question of whether these mutations affect dimerization, FBP binding, or both. Since the Skp1/FBP interaction is very strong (K_d for binding to the FBP Fbs1/OCP1 is ca. two orders of magnitude smaller than the homodimerization K_d)[45], and quantitative mass spectrometry indicates that Skp1 is not in great excess over FBPs in cells,[38] the average concentration of Skp1 does not appear to be high enough to generate a substantial homodimer pool. However, if higher local concentrations occur in the cell, homodimerization might occur to protect Skp1 from interacting promiscuously with other macromolecules. Our ability to selectively perturb dimerization relative to Fbs1 binding using the F97E mutation might allow an investigation of this question *in vivo*; however, we cannot exclude the possibility that interaction with FBPs is quantitatively affected.

Chemical shift index analysis of assigned residues of free human[39] and *Dictyostelium*[9] Skp1s indicated that the overall secondary structure elements of the dimers were similar to one another and to human Skp1 in complexes with FBPs, except for the C-terminal F-box subsite-2 region ***(residues 126-162)*** which was predominantly disordered in free Skp1's. The current study of residues 1-125 extends to show that free Skp1 (dimer) assumes essentially the same structure as for human Skp1 bound to the FBP βTRCP, with an RMSD for the corresponding Cα atoms of 1.1 Å ***(Figure S3.6)***. Thus interactions with proteins including Cul1[8] and Sgt1,[46] whose crystallographically defined binding interfaces lie within this region but N-terminal to the dimer interface (ca. residues 1-90), are likely to

be unaffected by the dimer status of Skp1.

Owing to the semi-parallel arrangement of the monomers, the two C-termini of the Skp1ΔΔ dimer are close enough to one another that the missing C-terminal regions have the potential to influence one another in the native protein. The availability of a stable monomeric form of Skp1 will now allow a direct NMR analysis of full length Skp1 and the consequences of its glycosylation, which has been postulated to influence the organization of F-box binding subsite-2[4,47] – the C-terminal region that was truncated to enable the dimer structure reported here.

Protein homo-dimerization is a common phenomenon with highly varied roles,[48] and was originally described for Skp1 over two decades ago.[5] Understanding the significance of Skp1 dimerization has been impeded by its recalcitrance to crystallization, likely owing to its intrinsically disordered regions, and the limitations of NMR to study larger protein structures. Here we assigned a K_d of 2.5 μM for native Skp1 from *Dictyostelium*, a protist that regulates its Skp1 via O_2-dependent glycosylation. By preparing a truncated version that remains dimeric, we solved its solution structure using NMR of isotopically labeled samples. The structure revealed that the dimer interface is contributed by a pair of α-helices from each subunit and buries 740 $Å^2$ of predominantly hydrophobic surface, with the monomer units assuming a semi-parallel orientation around a 2-fold rotational axis of symmetry. The portion of Skp1 remaining after truncation adopts a structure equivalent to that to the monomer unit found in crystal structures of complexes with F-box proteins. The dimer interface substantially overlaps with subsite-1 of the binding interface with the F-box domain of F-box proteins, but is not expected to compete with the other main binding function of Skp1 with cullin-1. The interface was verified by

finding that a Rosetta-predicted point mutation interferes with homodimerization of full-length Skp1 but not with binding a model F-box protein. The high degree of conservation of the dimerization patch throughout phylogeny suggests a potential role in buffering Skp1 function in cells, and the availability of a monomeric isoform of Skp1 will enable new structural studies on the mechanism of its control by glycosylation.

3.6 Acknowledgements

The authors thank John Glushka for NMR technical support.

3.7 Figures and Tables

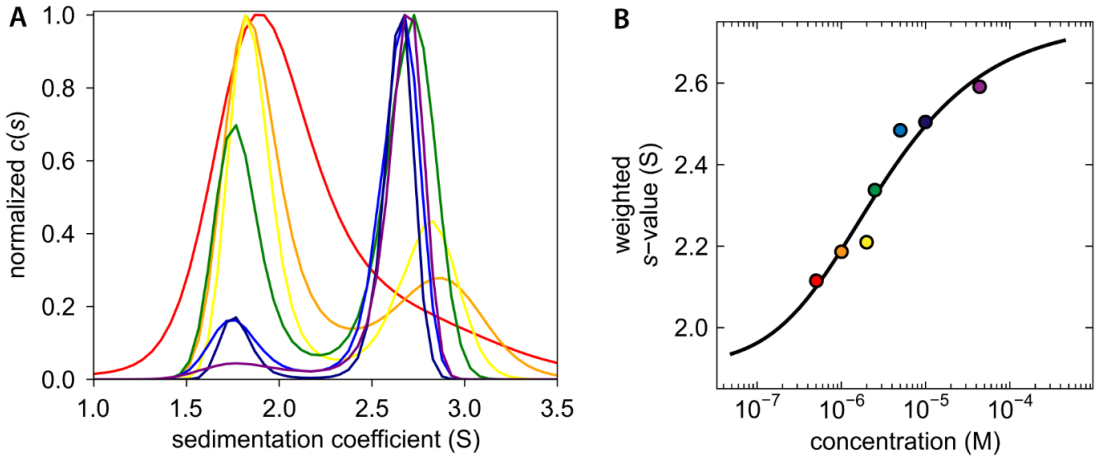


Figure 3.1. Sedimentation velocity analysis of *Dictyostelium* Skp1. (A) *c(s)* distribution reveals concentration dependence of dimerization. The concentration range is depicted by a rainbow spectrum with the lowest concentration in red and the highest in purple. (B) An isotherm was constructed with weighted s-values (S_w); the fitted model indicates a K_d of 2.5 μM. The color of each data point corresponds to the respective *c(s)* distribution in panel A.

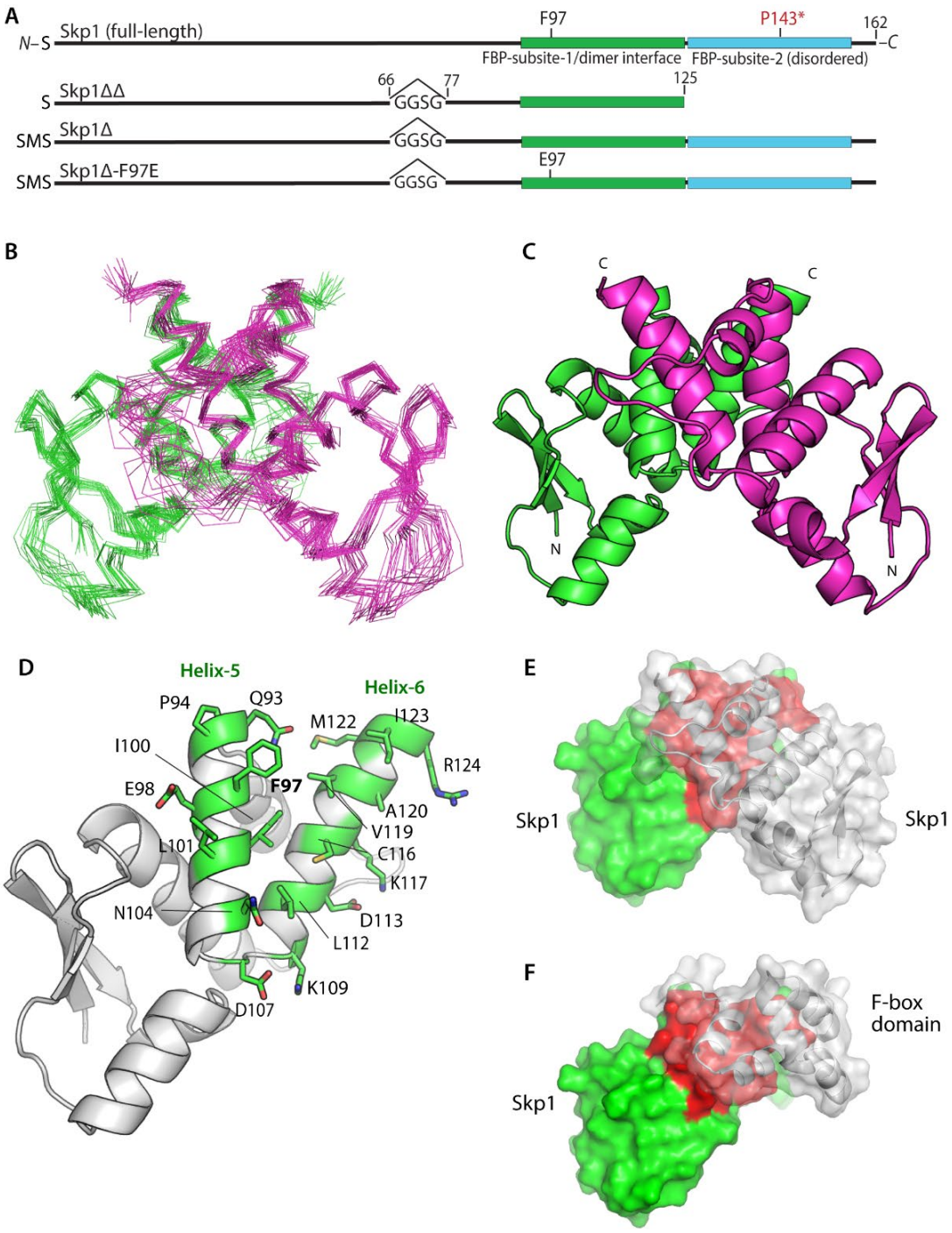


Figure 3.2. Structure of the Skp1 dimer. (A) Domain diagrams of the constructs examined. Note that versions derived from His_6Skp1 have a SerMet-extension beyond the native Ser- resulting from Met removal. See Figure S1 for details. (B) Superimposition of Cα-traces of 20 calculated conformers of Skp1ΔΔ. (C) Ribbon representation of the lowest energy Skp1ΔΔ conformer (PDB ID 6V88). Dimer subunits are colored in green or magenta. A 2-fold axis of rotational symmetry lies vertically between the subunits. (D)

Ribbon representation of a single Skp1ΔΔ, with the residues contributing to intermolecular contacts (<5 Å) shown in green with stick representations of their side chains. (E) Surface representation of the Skp1ΔΔ dimer is shown with the rear subunit colored in green and red, and the front in transparent gray. Red shading represents the homodimer contact region. (F) Surface representation of a hypothetical Skp1ΔΔ/F-box heterodimer model, generated by substitution of a single Skp1ΔΔ subunit for Skp1 in a human Skp1/FBXW7 complex (PDB ID 5V4B). Coloration is as in E, with FBXW7 residues 2263-2355 in gray.

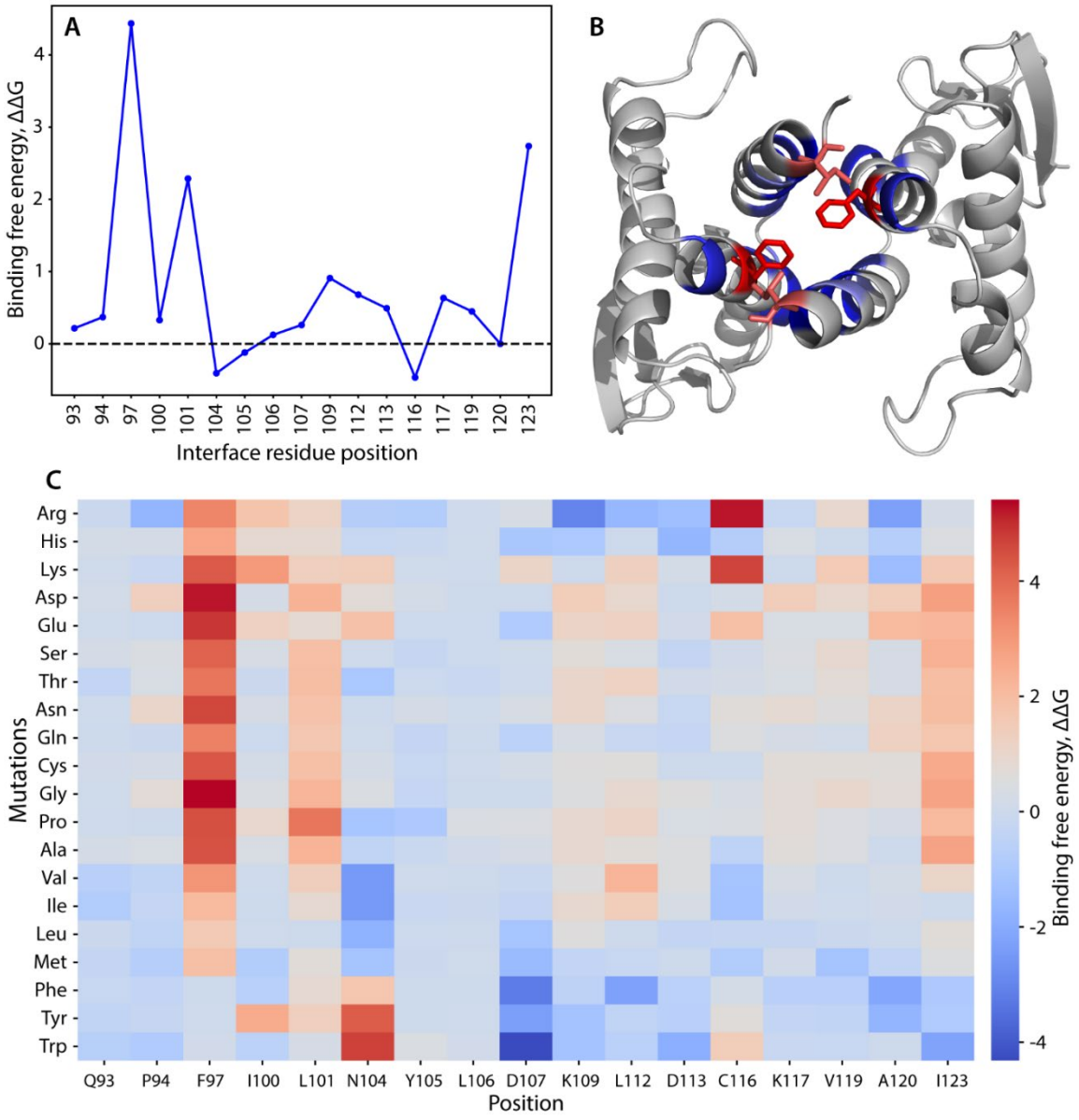


Figure 3.3. Computational scanning mutagenesis of the Skp1ΔΔ homodimer interface. (A) Alanine-scanning mutagenesis using Rosetta. Changes in binding free energy upon replacement with alanine are shown according to the interface residue positions. (B) Skp1 dimer protein-protein interface. Residues with the highest binding free energy change (Phe97 and Ile123) are emphasized in stick representation and in red; other mutated residues are in blue. See panel C for color code explanation. (C) Heatmap of the changes in binding free energy upon all amino acid substitutions. Effects of amino acid replacements are shown for each interface position. The colors represent the changes in the binding free energy of the dimer (interface ΔΔG score). Values greater than one (warmer colors) indicate destabilizing mutations, and values less than one (colder colors) imply stabilizing mutation.[28] Compare with effects on the monomer state (Fig. S3.8).

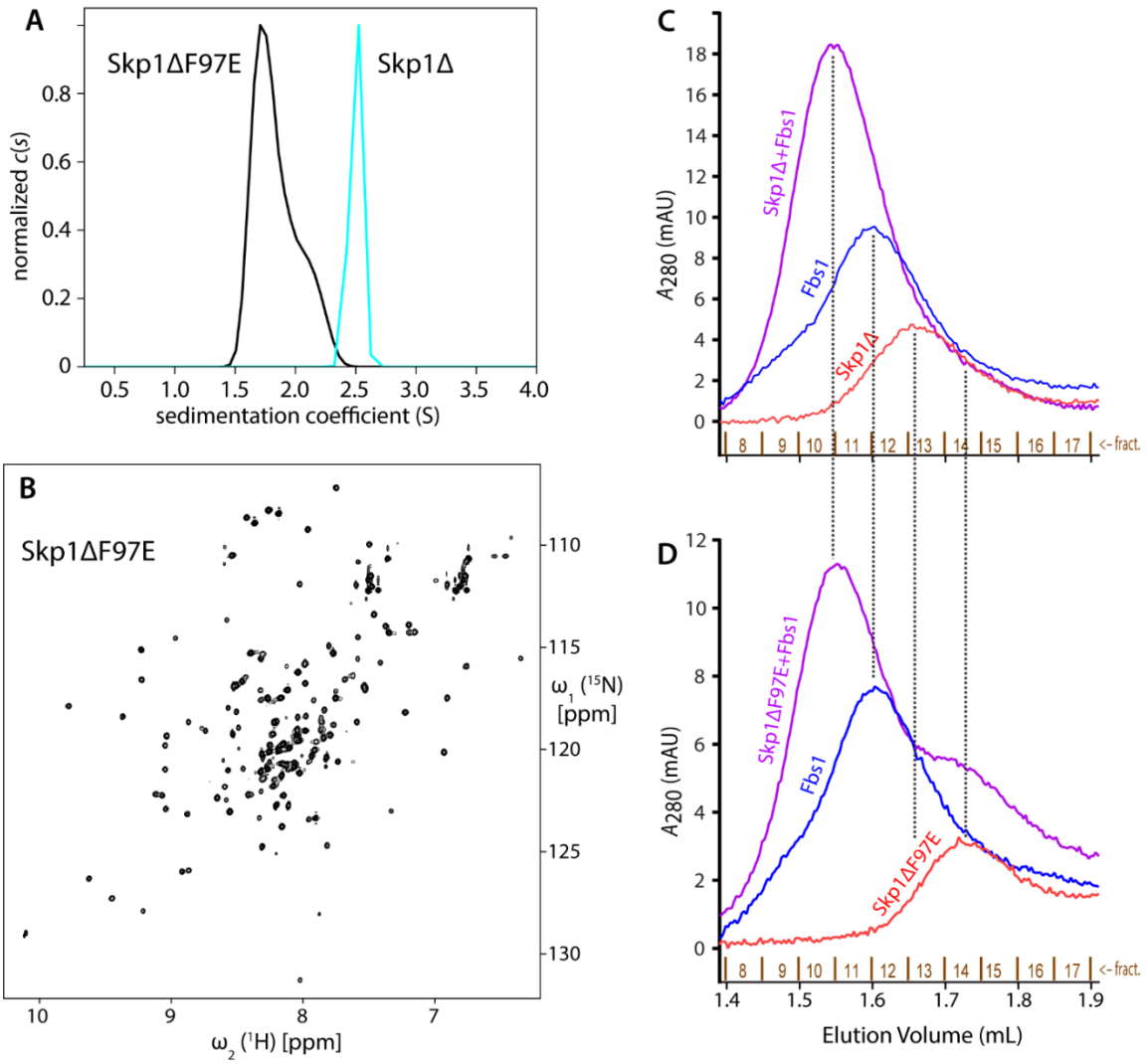


Figure 3.4. Skp1ΔF97E is a stable and functional monomer in solution. (A) *c(s)* distributions of 100 µM Skp1Δ or Skp1ΔF97E are shown in cyan and black, respectively. (B) $^{1}H/^{15}N$-HSQC of 100 µM Skp1ΔF97E at 900 MHz and 35°C, with a 4 h collection time. The 500 µM spectrum (not shown) was indistinguishable. (C, D) Skp1ΔF97E binds the model F-box protein Fbs1. His_6Fbs1 (1.5 µM) and an estimated 2.25 µM Skp1Δ (C) or Skp1Δ(F97E) (D) were analyzed on a Superdex 200 gel filtration column. Elution was monitored by A_{280}, which favors detection of Fbs1 relative to Skp1 because of its higher extinction coefficient.

Completeness of resonance assignments[a] [%]	
Backbone/Side-chain	100.0/100.0
Conformation-restricting distance constraints[b]	
Intra-residue [i = j]	696
Sequential [|i - j| = 1]	1256
Medium range [1 < |i - j| < 5]	1360
Long range [| i - j | ≥ 5]	1470
Total	4782
Intermolecular NOE constraints (included in above)	182
Dihedral angle constraints	406
NOE constraints per restrained residue (of those, long range)	22.6(6.4)
CYANA target function [Å^2]	2.33
Average number of distance constraint violations per conformer	
0.1 - 0.2 Å	9.25
0.2 – 0.5 Å	2.2
>0.5 Å	0
Average number of dihedral angle constraint violations per conformer	
>10°	0.1
Average RMSD from mean coordinates [Å]	
backbone atoms[c] (all)	0.8 (1.2)
heavy atoms[c] (all)	1.1 (1.4)
Global quality scores[c] (raw / Z-score)	
PROCHECK[2] G-factor (phi-psi)	0.16/0.94
PROCHECK[2] G-factor (all)	0.14/0.83
Molprobity[3] clash score	2.55/1.09
ProsaII[4]	0.74/0.37
Molprobity[3] Ramachandran summary [%]	
Most favored regions	98.7
Additionally allowed regions	1.3
Disallowed regions	0.0

[a] Commonly observed protein NMR resonances. Excludes amino group of N-terminal Ser, side-chain amino groups of Lys, side-chain guanidinium groups of Arg, carboxyl groups of Asp and Glu, thiol and hydroxyl ^{1}H of Cys, Ser, Thr and Tyr, and non-protonated aromatic ^{13}C.
[b] Calculated with Protein Structure Validation Software (PSVS 1.5; http://psvs.nesg.org/)
[c] Ordered residue ranges: 3-34, 38-44, 46-64, 74-115
[1] Ref. 33
[2] Ref. 34
[3] Ref. 35
[4] Ref. 36

Table 3.1. Skp1ΔΔ dimer NMR structure statistics (PDB ID: 6V88, BMRB ID:30696)

	Predicted[a] τ_c (ns), monomer/dimer	Experimental τ_c (ns)
Skp1ΔF97E, 500 μM	10.9/19.6	9.4 ± 0.5
Skp1(native), 850 μM	Similar to above	19.5 ± 2.2[b]

[a] Predicted using HYDRONMR[23]
[b] from Ref. 7

Table 3.2. Predicted and Experimental τ_c for Skp1 isoforms at 35°C

3.8 Supplementary Figures and Tables

(A) His_6Skp1A

```
                                                                M  G  H   -20
Syn                                                          CCATGGgccac -60
NcoI

    H  H  H  H  H  S  S  G  V  D  L  G  T  E  N  L  Y  F  Q  S     0
Syn catcatcatcatcattcttctggtgtagatctgggtaccgagaacctgtacttccaatcc   0

    M  S  L  V  K  L  E  S  S  D  E  K  V  F  E  I  E  K  E  I    20
Dd  ATGTCTTTAGTTAAATTAGAATCTTCAGATGAAAAAGTCTTTGAAATTGAAAAAGAAATC  60
Syn ATGAGCCTGGTGAAACTGGAAGCAGCGACGAGAAGGTGTTCGAGATTGAGAAAGAGATT

    A  C  M  S  V  T  I  K  N  M  I  E  D  I  G  E  S  D  S  P    40
Dd  GCTTGTATGTCAGTTACAATCAAGAATATGATTGAAGATATTGGTGAATCAGATAGTCCA 120
Syn GCGTGCATGAGCGTGACCATTAAAAACATGATCGAGGACATTGGTGAAAGCGATAGCCCG

    I  P  L  P  N  V  T  S  T  I  L  E  K  V  L  D  Y  C  R  H    60
Dd  ATTCCATTACCAAATGTTACTAGCACTATTTTAGAGAAAGTTCTTGACTATTGCAGACAT 180
Syn ATTCCGCTGCCGAACGTGACCAGCACCATCCTGGAGAAGGTTCTGGACTACTGCCGTCAC

    H  H  Q  H  P  S  P  Q  G  D  D  K  K  D  E  K  R  L  D  D    80
Dd  CACCATCAACATCCATCACCACAAGGTGACGATAAAAAGGATGAAAAGAGATTAGATGAT 240
Syn CATCACCAGCACCCGAGCCCGCAAGGTGACGATAAGAAAGATGAAAAGCGTCTGGACGAT

    I  P  P  Y  D  R  D  F  C  K  V  D  Q  P  T  L  F  E  L  I   100
Dd  ATCCCACCATATGATAGAGATTTCTGTAAAGTCGATCAACCAACCTTATTCGAATTAATC 300
Syn ATTCCGCCGTACGACCGTGATTTCTGCAAAGTGGACCAGCCGACCCTGTTTGAGCTGATT

    L  A  A  N  Y  L  D  I  K  P  L  L  D  V  T  C  K  T  V  A   120
Dd  TTGGCAGCCAATTATTTGGATATCAAACCATTATTAGATGTTACCTGTAAAACTGTTGCC 360
Syn CTGGCGGCGAACTATCTGGACATCAAGCCGCTGCTGGATGTGACCTGCAAAACCGTTGCG

    N  M  I  R  G  K  T  P  E  E  I  R  K  I  F  N  I  K  N  D   140
Dd  AATATGATCAGAGGTAAAACCCCAGAAGAAATCAGAAAAATCTTCAACATCAAGAACGAC 420
Syn AACATGATCCGTGGTAAAACCCCGGAGGAAATCCGTAAGATTTTCAACATCAAGAACGAT

    F  T  P  E  E  E  E  Q  I  R  K  E  N  E  W  C  E  D  K  G   160
Dd  TTTACTCCAGAAGAAGAAGAACAAATCAGAAAAGAAAATGAATGGTGTGAAGATAAAGGT 480
Syn TTCACCCCGGAAGAAGAGGAACAAATCCGTAAGGAGAACGAATGGTGCGAGGACAAGGGT

    G  N  *                                                      162
Dd  GGAAACTAA                                                    489
Syn GGTAACTAAGGATCC                                              BamHI
```

(B) His$_6$Skp1AΔloop

```
                                                   M  G  H    -20
Syn                                              CCATGGgccac -60
NcoI

    H  H  H  H  H  S  S  G  V  D  L  G  T  E  N  L  Y  F  Q  S      0
Syn catcatcatcatcattcttctggtgtagatctgggtaccgagaacctgtacttccaatcc    0

    M  S  L  V  K  L  E  S  S  D  E  K  V  F  E  I  E  K  E  I     20
Dd  ATGTCTTTAGTTAAATTAGAATCTTCAGATGAAAAAGTCTTTGAAATTGAAAAAGAAATC   60
Syn ATGAGCCTGGTGAAACTGGAAAGCAGCGACGAGAAGGTGTTCGAGATTGAGAAAGAGATT

    A  C  M  S  V  T  I  K  N  M  I  E  D  I  G  E  S  D  S  P     40
Dd  GCTTGTATGTCAGTTACAATCAAGAATATGATTGAAGATATTGGTGAATCAGATAGTCCA  120
Syn GCGTGCATGAGCGTGACCATTAAAAACATGATCGAGGACATTGGTGAAAGCGATAGCCCG

    I  P  L  P  N  V  T  S  T  I  L  E  K  V  L  D  Y  C  R  H     60
Dd  ATTCCATTACCAAATGTTACTAGCACTATTTTAGAGAAAGTTCTTGACTATTGCAGACAT  180
Syn ATTCCGCTGCCGAACGTGACCAGCACCATCCTGGAGAAGGTTCTGGACTACTGCCGTCAC

    H  H  Q  H  P  G  G  S  G                          L  D  D     80
Dd  CACCATCAACATCCAGGTGGTTCCGGA------------------------TTAGATGAT  240
Syn CATCACCAGCACCCGGGTGGTTCCGGA------------------------CTGGACGAT

    I  P  P  Y  D  R  D  F  C  K  V  D  Q  P  T  L  F  E  L  I    100
Dd  ATCCCACCATATGATAGAGATTTCTGTAAAGTCGATCAACCAACCTTATTCGAATTAATC  300
Syn ATTCCGCCGTACGACCGTGATTTCTGCAAAGTGGACCAGCCGACCCTGTTTGAGCTGATT   F97E-S
 AS TAAGGCGGCATGCTGGCACTAAAGACGTTTCACCTGGTCGGCTGGGACAAACTCGACTAA   F97E-
AS

    L  A  A  N  Y  L  D  I  K  P  L  L  D  V  T  C  K  T  V  A    120
Dd  TTGGCAGCCAATTATTTGGATATCAAACCATTATTAGATGTTACCTGTAAAACTGTTGCC  360
Syn CTGGCGGCGAACTATCTGGACATCAAGCCGCTGCTGGATGTGACCTGCAAAACCGTTGCG   F97E-S
 AS GACCGCCGCTTGATAGACCTGTAGTTCGGCGACGACCTACACTGGACGTTTTGGCAACGC

    N  M  I  R  G  K  T  P  E  E  I  R  K  I  F  N  I  K  N  D    140
Dd  AATATGATCAGAGGTAAAACCCCAGAAGAAATCAGAAAAATCTTCAACATCAAGAACGAC  420
Syn AACATGATCCGTGGTAAAACCCCGGAGGAAATCCGTAAGATTTTCAACATCAAGAACGAT

    F  T  P  E  E  E  E  Q  I  R  K  E  N  E  W  C  E  D  K  G    160
Dd  TTTACTCCAGAAGAAGAAGAACAAATCAGAAAAGAAAATGAATGGTGTGAAGATAAAGGT  480
Syn TTCACCCCGGAAGAAGAGGAACAAATCCGTAAGGAGAACGAATGGTGCGAGGACAAGGGT

    G  N  *                                                       162
Dd  GGAAACTAA                                                     489
Syn GGTAACTAAGGATCC                                               BamHI
```

F97E mutation:

Dd-Skp1A-optim-F97E-S: 5'-GACCCTCGAGGAGCTGATTCTGGCGGCG

Dd-Skp1A-optim-F97E-AS: 5'-CAGCTCCTCGAGGGTCGGCTGGTCCAC

(C) Skp1AΔloopΔCterm (Skp1ΔΔ)

```
M  S  L  V  K  L  E  S  S  D  E  K  V  F  E  I  E  K  E  I
ATGTCTTTAGTTAAATTAGAATCTTCAGATGAAAAAGTCTTTGAAATTGAAAAAGAAATC
```

Figure S3.1. Synthetic *Dictyostelium* Skp1A cDNAs. (A) The native (Dd) and synthetic (Syn) coding sequences of Skp1A are shown together with their identical translated amino acid sequences. Amino acid (bold) and nucleotide numbering starts at the original start Met. The start codons are highlighted in green, the new N-terminus (Ser) generated by TEV protease cleavage and Phe97 are highlighted in yellow, nucleotide replacements in the synthetic DNA are in red, and nucleotides 196-232 replaced in Skp1Δloop (panels B, C) are underlined. (B) The native (Dd) and synthetic (Syn) cDNA sequences for His_6Skp1Δloop, in which amino acids SPQGDDKKDEKR are replaced by GGSG (in blue). Original numbering for downstream positions is retained. Oligonucleotide sequences utilized for site-directed mutagenesis to generate F97E are indicated below the sequences, including the lower strand (AS) where needed to locate antisense primers. (C) Deletions were generated in the native sequence of Skp1A in pET19b-Skp1 by a variation of site-directed mutagenesis using the indicated oligonucleotides, with matching sequences highlighted in yellow or gray. Truncated C-terminal sequences are in light gray.

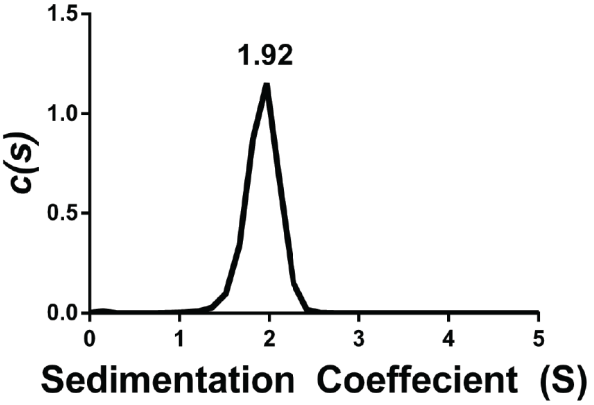


Figure S3.2. Sedimentation velocity analysis of Skp1ΔΔ. Analytical ultracentrifugation of 100 μM Skp1ΔΔ, as described in Figure 3.1. DdSkp1ΔΔ sediments at 1.92 S, which matches the predicted 1.92 value for the dimer using HYDROPRO. In contrast, the sedimentation coefficient for a monomeric Skp1ΔΔ was predicted be 1.28 S. The breadth of the peak width suggests that a small amount of monomeric Skp1ΔΔ may exist in rapid equilibrium with the dimer.

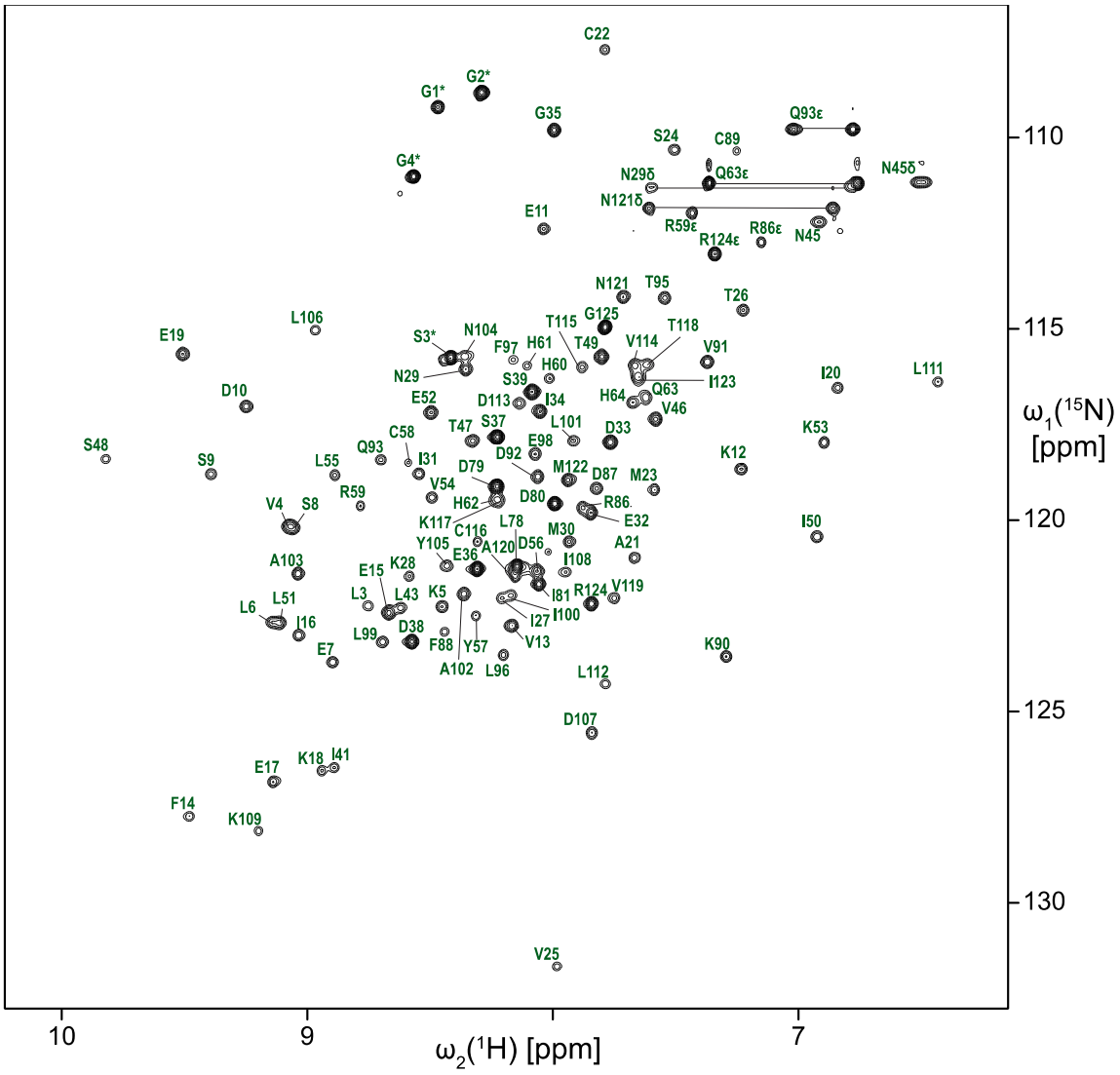


Figure S3.3. 2D [^{15}N, ^{1}H] HSQC spectrum of a 1:1 mixed sample of U-^{15}N,^{13}C labeled and natural abundance Skp1ΔΔ acquired at 800 MHz field strength. Peak assignments are indicated in green with residue numbers corresponding to the wild-type sequence. Assignments associated with the non-native tetrapeptide insertion (GGSG) that replaced the internal loop (residues 66-77) are denoted with asterisks.

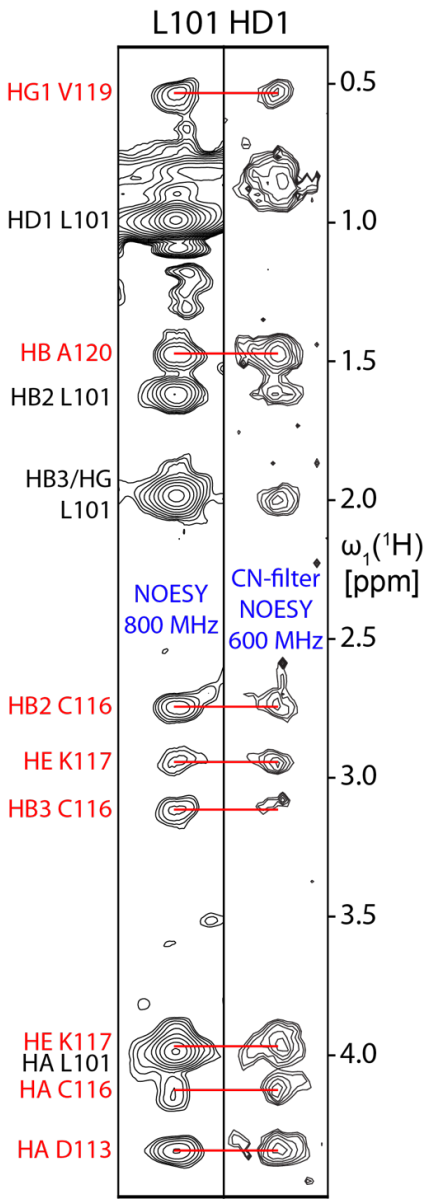


Figure S3.4. Intermolecular ^{1}H-^{1}H NOE contacts between Skp1ΔΔ dimer subunits. Representative strips from 3D ^{13}C/^{15}N-edited [^{1}H,^{1}H] NOESY and 3D ^{13}C/^{15}N-filtered, ^{13}C-edited [^{1}H,^{1}H] NOESY spectra showing contacts for HD1 of L101 recorded on a mixture of natural abundance and ^{13}C,^{15}N-enriched Skp1ΔΔ. Spectra were recorded at 800 and 600 MHz magnetic field strengths, with mixing times of 70 and 120 ms, respectively. Intermolecular and intramolecular NOE correlations are labeled in red and black, respectively

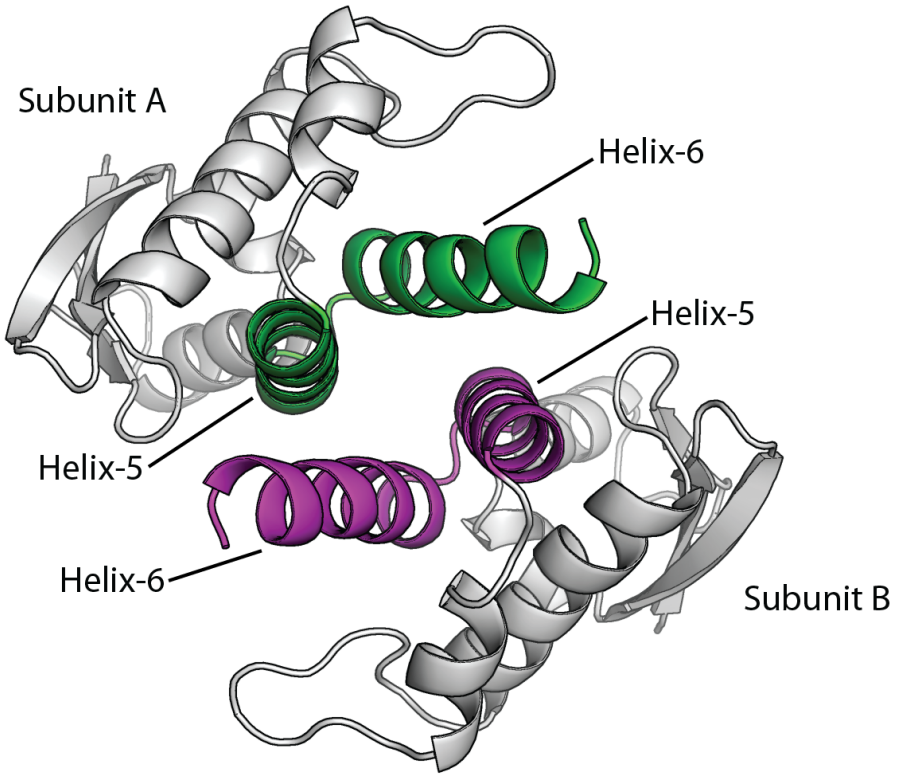


Figure S3.5. The dimer interface is composed of a 4-helix bundle, or a pair of pairs that are related by two-fold rotational symmetry. Green and magenta represent the helical pairs contributed by each Skp1ΔΔ subunit.

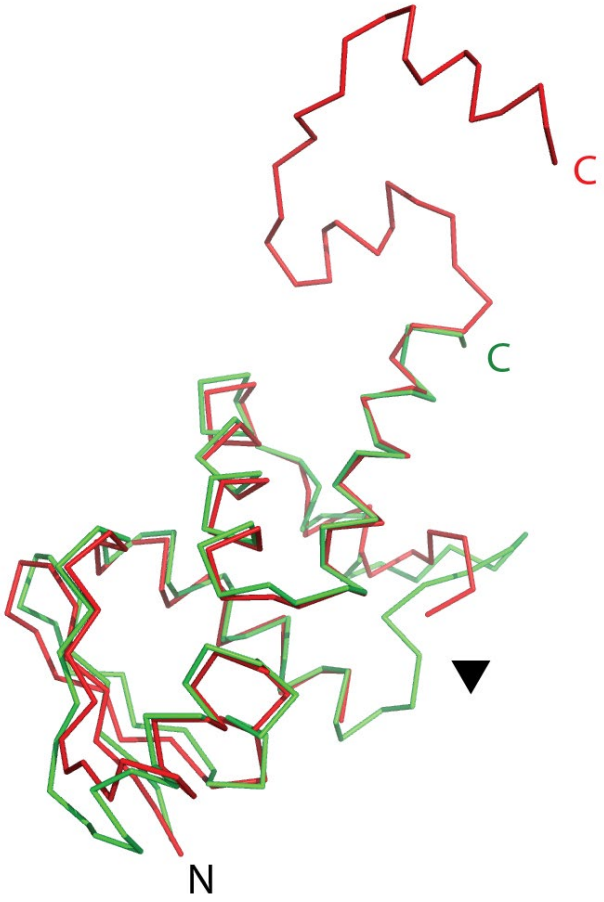
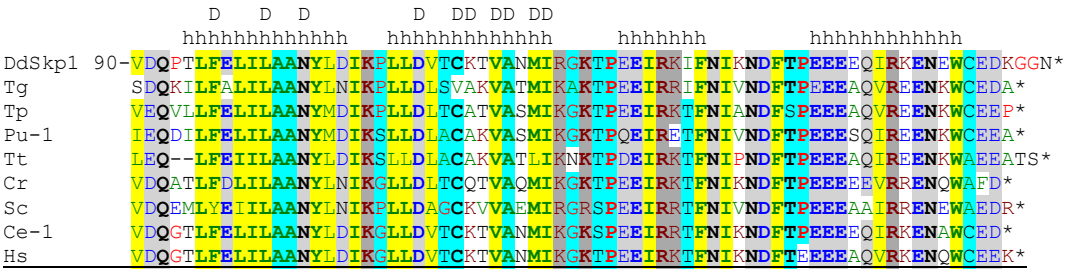


Figure S3.6. Overlay of Cα traces of a single subunit from the *Dictyostelium* Skp1ΔΔ dimer (green) and human Skp1 (red). HsSkp1 was excerpted from a crystal structure of its complex with βTRCP, a human FBP (PDB ID 6M90). The N- and C-termini are indicated. The corresponding 97 Cα atoms of DdSkp1 (amino acids 2-33, 40-62, 83-123) and HsSkp1 (amino acids 2-33, 38-60, 88-128) align with an RMSD of 1.1 Å. An internal loop in HsSkp1, which corresponds to the segment substituted with a GGSG tetrapeptide in DdSkp1ΔΔ, is absent from its Cα trace at the black triangle.



Dd, *Dictyostelium discoideum* (cellular slime mold), NCBI: AAA67888.1
Tg, *Toxoplasma gondii* (apicomplexa), ToxoDB: TGARI_207680
Tp, *Thalassiosira pseudonana* 38460 (diatom/stramenopile), NCBI, XP_002294707.1
Pu, *Pythium ultimum*, oomycete plant pathogen, PYU1_T013675, www.pythiumdb.org
Tt, *tetrahymena thermophila* (ciliated protozoan), TTHERM_00426320
Cr, *Chlamydomonas reinhardtii* (green algae/chlorophyta), XP_001690964.1
Sc, *Saccharomyces cerevisiae* (budding yeast), NCBI, NP_010615.3
Ce, *Caenorhabditis elegans* (nematoda) NP_492513.1
Hs, *Homo sapiens* (vertebrata), NCBI, NP_008861.2

Figure S3.7. Alignment of Skp1 sequences from across eukaryotic phylogeny. The D letters in the top row indicate positions of intermolecular contacts observed for *Dictyostelium* Skp1ΔΔ. Lower case h letters depict the positions of helices-5, -6, -7, and -8 (from left to right) in crystal structures with FBPs. To facilitate visualization of relatedness, acidic residues are in blue, basic in dark red, Gly and Pro in red, and hydrophobic in green, as previously described (West, C. M.; van der Wel, H.; Gaucher, E. A. (2002) Complex Glycosylation of Skp1 in *Dictyostelium*: Implications for the Modification of Other Eukaryotic Cytoplasmic and Nuclear Proteins. Glycobiology 12, 17R-27R). Positions possessing a consensus chemical characteristic are highlighted, in yellow (hydrophobic), gray (acidic), dark grey (basic), or teal (small). Positions of near perfect conservation are bolded. Asterisk denotes the C-terminus.

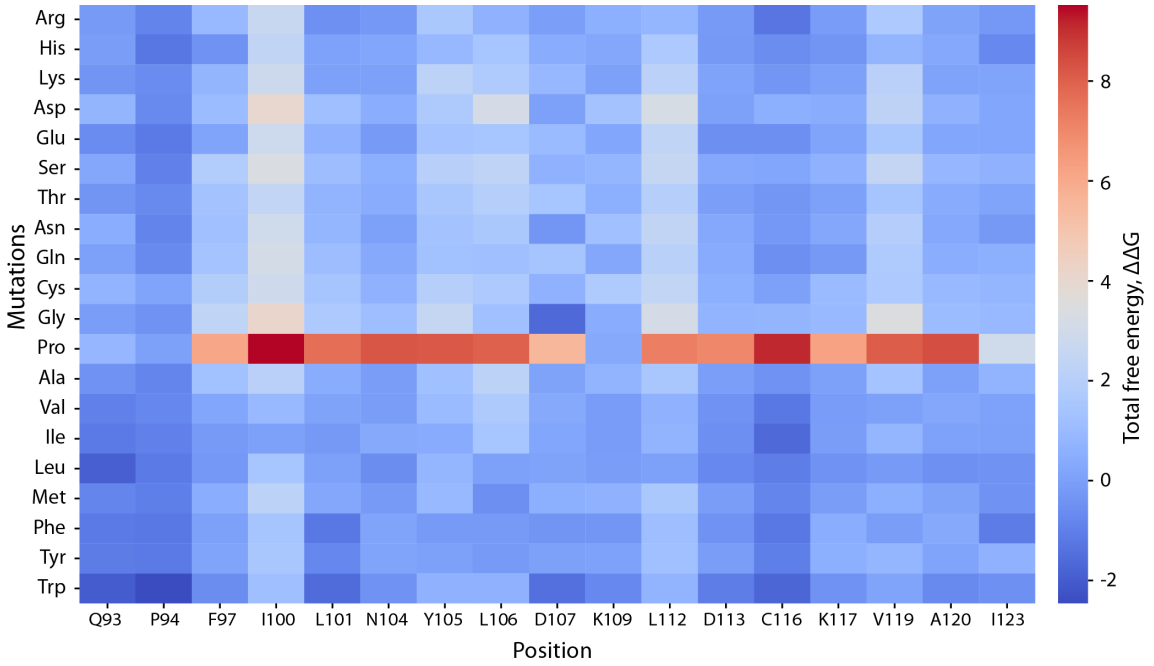


Figure S3.8. Heatmap of total *in silico* free energy changes in the Skp1 monomer for each amino acid substitution at the dimer interface. The effects of amino acid replacements are plotted on the y-axis as a for each interface position (x-axis). Colors represent calculated changes in the total free energy of the monomer (total ΔΔG score). Values greater than one (warmer colors) indicate destabilizing mutations, and values less than one (colder colors) imply stabilizing mutations.[31] Compare with effects on the dimer state (Fig. 3.3C).

Sample	Experiment*	Magnetic field (MHz)
Skp1ΔΔ	2D [^{15}N, ^{1}H]-HSQC	800
	2D [^{13}C, ^{1}H] CT-HSQC (aliphatic)	800
	2D [^{13}C, ^{1}H] CT-HSQC (aromatic)	800
	3D ^{13}C/^{15}N-edited [^{1}H,^{1}H] NOESY	800
	3D TROSY-HNCACB	800
	3D (H)CCH-COSY aromatic	800
	3D (H)CCH-COSY aliphatic	800
	3D (H)CCH-TOCSY aliphatic	800
	3D ^{13}C/^{15}N-filtered ^{13}C/^{15}N-edited [^{1}H,^{1}H] NOESY	600
	3D CBCA(CO)NH	600
	3D HBHA(CO)NH	600
	3D HNCO	600
	3D (HCA)CONH	600
	2D long-range [^{15}N, ^{1}H]-HSQC	600
100 µM Skp1ΔF97E	2D [^{15}N, ^{1}H]-HSQC	900
	1D ^{15}N T_1	600
	1D ^{15}N T_2	600
500 µM Skp1ΔF97E	2D [^{15}N, ^{1}H]-HSQC	900
	1D ^{15}N T_1	600
	1D ^{15}N T_2	600

***** wiki.nesg.org

Table S3.1 List of NMR experiments.

CHAPTER 4

NMR ANALYSIS OF GLYCOSYLATED SKP1Δ

4.1 Introduction

Skp1 is protein-protein interactor that recruits substrate receptor F-box proteins (FBP) to cullin-1 in the Skp1/cullin-1/FBP (SCF) family of E3 ubiquitin ligases which is involved in targeting proteins for degradation.[3,7,8,10,] In many protists, a proline residue within the C-terminal region of Skp1 undergoes an O_2-dependent post-translational modification that results in the formation of a 4(*trans*)-hydroxyproline with an O-linked pentasaccharide ***(Figure 1.5)***.[34,35] Investigation of the Skp1 glycosylation system in *Dictyostelium* and *Toxoplasma* showed that full glycosylation is required for O_2 sensing and optimal growth, and the enzymes in this pathway have been characterized to be specific for Skp1 suggesting that the O_2 sensing is mediated through Skp1.[36-39,42-48]

Skp1 exists as a homodimer in solution, but as a monomer in Skp1/FBP or the full SCF complex as demonstrated by numerous crystal structures ***(Figures 3.2C-F, Figure 1.3C)***.[10,18,55] These studies showed that the BTB domain of Skp1 docks onto Cullin-1 and makes up a portion of FBP binding interface (subsite-1) ***(Figure 1.3)***. The C-terminal extension of Skp1 forms the remaining FBP interface (subsite-2) with helices-7 and 8 that latches around the F-box motif. However, pioneering studies of Skp1 alone by NMR showed that the C-terminal extension is highly disordered and mobile in its free state, indicating that the region undergoes a disorder-to-helix transition when binding to an

FBP.[26] The glycan is located at the proline position preceding helix-8, and it was shown to affect the structure of the surrounding C-terminal region. The addition of the α-GlcNAc promoted α-helicity within this region showing that the glycan directly affects Skp1 structure ***(Figure 1.9)***.

Clues as to how Skp1 is affected by the glycan have been informed by substantial biochemical, biophysical, and computational experiments that point to two modes of action. A binding competition assay with a model FBP indicated preferential binding to glycosylated forms of Skp1 ***(Figure 1.7)***.[57] Cross-linking assays and size-exclusion analysis indicate that glycan inhibits Skp1 homodimerization, potentially increasing available monomeric pool of Skp1 to associate with FBPs. A structural model of how this preference is modulated by the glycan was indicated by NMR-supervised molecular dynamic (MD) simulations ***(Figure 1.11)***.[40] In the model, the glycan interact with Skp1 through hydrogen bonds and hydrophobic interactions along the loop connecting helices-7 and 8 ***(Figure 1.11C)***. Consequentially, the C-terminal region adopts an extended conformational and exposes the FBP binding site whereas the unglycosylated C-terminus forms a collapsed conformation. These results together indicate that the glycan inhibits homodimerization and alters the C-terminal structure via molecular contacts to enhance FBP binding. Owing to the relatively large size of the Skp1 dimer (>40 kDa), NMR analysis of Skp1 required perdeuteration for resonance assignments. Therefore, experimental investigation of the contacts between the C-terminal region and the glycan has been elusive, because protonated sample is required to measure NOEs.

In search of structural data to test our model, we first determined the Skp1 structure by NMR in order to assess how the dimer is assembled ***(Figure 3.2B)***.[56] A doubly-

truncated construct Skp1ΔΔ (~26 kDa) missing the C-terminal region and an internal loop that still retains its dimeric properties was used in order to maximize data signal and to reduce spectral complexity ***(Fig 4.1)***. Skp1ΔΔ was expressed in a protonated background in order to measure ^{1}H-^{1}H contacts that are required for structure calculation. The sample yielded enough quality data for complete resonance assignment as well as sufficient NOEs to calculate a high-resolution structure. The structure revealed that the dimer interface is composed of a 4-helix bundle with predominantly hydrophobic contributions from helices 5 and 6 ***(Figure 3.2D)***. This interface overlaps with FBP-subsite 1 providing an explanation why Skp1 is found as a monomer in SCF complexes ***(Figures 3.2E,F)***. The structure allows visualization how the glycosylated regions of Skp1 are placed relative to each other and to the dimer interface. This arrangement revealed that each subunit's C-terminal region would presumably reside in close proximity in space with the potential for interactions between the C-terminal regions, the glycans, and the dimer interface.

Building on the newly determined dimer structure, this chapter focuses on preparation of a glycosylated sample to be examined by NMR for structural evidence regarding the predicted MD Skp1-glycan contacts and for clues as to how the glycan might act to inhibit dimerization. Skp1Δ is a truncated construct where a 12-amino acid internal loop is replaced by a GGSG linker which is expected to reduce spectral complexity in the flexible region of NMR spectra. Skp1Δ was co-expressed in *E.* coli together with PhyA and Gnt1 in minimimal media containing ^{15}N and ^{13}C isotopes in a protonated background. The uniformly labeled Gn-Skp1Δ enriched with^{15}N and ^{13}C was isolated and the remaining 4 sugars were attached *in vitro* using purified Skp1 glycosyltransferase PgtA and AgtA and ^{13}C-labeled sugar nucleotides. The Skp1 glycoprotein was purified to homogeneity and the

list of NMR spectra recorded is outlined in Table 4.1. The data analysis at early stages demonstrated that the sample behaved stably based on ^{15}N-HSQC, and all five anomeric sugar signals were detected by ^{13}C-HSQC spectrum.

4.2 Materials and Methods

Expression and purification of uniformly labeled ^{15}N,^{13}C-His$_6$GnSkp1Δ and ^{15}N,^{13}C-GnSkp1Δ

E. coli strain ER2566 containing pET19b-His$_6$Skp1Δ (Amp^R) and pET28a-DdPhyA+gnt1chimera (Kan^R) were used as a source for His$_6$GnSkp1Δ. The strain was grown in 8 X 1 liter of LB media containing 100 µg/ml carbenicillin and 100 µg/ml kanamycin in Fernbach flasks. The media was inoculated with 2% overnight starter culture and incubated at 37°C and 200 RPM. The culture was harvested at OD_{600} of 0.6-0.8 by centrifugation at 5000 X g for 10 minutes at 4°C. The *E. coli* pellet was resuspended with 2 x 1 liter of M9 minimal media containing 100 µg/ml carbenicillin and 100 µg/ml, kanamycin, 1g/L $^{15}NH_4Cl$ (Cambridge Isotopes), 0.4% (w/v) ^{13}C-glucose (Cambridge Isotopes), 0.1 mM $CaCl_2$, 0.5 mM $MgSO_4$, and 1% vitamin solution. The vitamin solution was prepared by dissolving 1 multi-vitamin tablet (Centrum) into 30 ml of water, followed by centrifugation for 30 min at 30,000 x g; The vitamin solution was sterile filtered prior to use. The culture resuspended in M9 media was incubated at 37°C for 20 min at 200 RPM, and then protein expression was induced with 1mM IPTG at 22°C 18-20 hours. The culture was harvested at 5000 x g and 4°C, and the *E. coli* pellet was stored at -80°C until preparation.

The *E. coli* pellet was resuspended in 50 mM Na^+/K^+ phosphate (pH 7.4), 300 mM NaCl, 10 µg/ml aprotinin, 10 µg/ml leupeptin, and 1 mM PMSF at 4 ºC. Cells were lysed using a probe sonicator (model 500, Thermo Fisher Scientific) for a total sonication time of 5 min. The lysate was centrifuged at 27,000 x g for 45 min at 4 ºC, and the supernatant was applied to a 5 ml Talon cobalt resin (Takara) pre-equilibrated with the lysis buffer described above. The column was washed with the same buffer supplemented with either 1M NaCl, 10% glycerol, or 10 mM imidazole until OD_{280} reached a baseline. The target protein was eluted with buffer containing 300 mM imidazole, and the peak fractions were dialyzed using a 10 kDa cut-off membrane overnight against 20 mM Tris-HCl (pH 7.5), 1 mM EDTA, 5 mM $MgCl_2$ and 1 mM DTT. 10% of the total sample was treated with His_6-Tev protease to obtain $^{15}N,^{13}C$-GnSkp1Δ. The rest of the protein sample was loaded onto a 5 ml Q-sepharose HiTrap column (GE Healthcare) pre-equilibrated with the same Tris buffer described above. The column was washed until baseline, and the protein was eluted using a 0-300 mM NaCl gradient. The eluted Skp1 sample fractions were confirmed by SDS-PAGE with Comassie blue staining and anti-His_6 western blotting. MALDI-TOF-MS was used to evaluate isotopic incorporation.

Expression and purification of Flag-DdPgtA and DdAgtA

18 liters of *Dd* Ax3 strain was transiently transfected with a pDM320-Flag-DdPgtA and was allowed to express until *Dd* reached a density of 1 x 10^7 cells/ml. The cells were lysed using 5 micron syringe filters, and the lysate was centrifuged at 27,000 x g for 45 minutes followed by ultracentrifugation at 100,000 x g for 45 minutes. The clarified lysate was applied over a 90 ml DEAE-sepharose column and eluted using a 0-300 mM NaCl gradient in buffer composed of 50 mM Tris-HCl (pH 7.4), 5 mM $MgCl_2$, 0.1 mM EDTA,

and 1 mM DTT. Fractions containing Flag-DdPgtA were identified using a dot blot on nitrocellulose that was probed M2 primary antibody (αFlag) and anti-mouse secondary antibody. Flag-DdPgtA containing fractions were loaded onto a 5 ml M2-agarose column (Sigma), washed, and eluted with 0.1 mg/ml 3x Flag peptide (GLPBIO). Fractions containing Flag-DdPgtA were concentrated and applied over a Superdex-200 size-exclusion column (HiLoad 16/600, Cytiva) preequilibrated with 50 mM HEPES-NaOH (pH 7.4), 150 mM NaCl, and 1 mM DTT. Protein elution was detected based on absorbance at 280nm. The purity of the eluted peak was estimated to be greater than 95% based on SDS-PAGE, and the elution fractions were flash frozen with liquid N_2 and stored in -80°C for later use.

E. coli Gold strain containing pET-15b-His-DdAgtA was used as a source for *Dd*AgtA. The strain was grown in 4 x 1 liter of LB media supplemented with 100 µg/ml ampicillin. The culture was incubated at 37°C and 200 RPM until OD_{600} (1 cm pathlength) reached ~1.0, and the expression was induced with 0.5 mM IPTG for 20 hours at 18°C. The cell pellet was harvested at 5000 x g and 4°C. His-*Dd*AgtA was purified with a 5ml Talon cobalt column as described above, and the His_6-tag was removed using Tev protease overnight. *Dd*AgtA was retrieved over a Talon column, then the sample was concentrated and applied over a Superdex-200 size-exclusion column (HiLoad 16/600, Cytiva) preequilibrated with the buffer described above. The purity of the eluted fraction was estimated to be greater than 95% based on Coomassie blue stained SDS-PAGE gels.

Enzymatic synthesis and purification of $^{15}N,^{13}C$-GGFGGnSkp1Δ

To prepare $^{15}N,^{13}C$-His_6FGGnSkp1Δ, 100 µM $^{15}N,^{13}C$-His_6GnSkp1Δ was reacted with 0.4 µM FlagDdPgtA (250:1 molar ratio). The reaction vessel contained 75 mM

HEPES-NaOH (pH 7.4), 75 mM NaCl, 5 mM $MgCl_2$, 2 mM $MnCl_2$, 0.1% Tween-20, 200 µM UDP-[U-$^{13}C_6$]-Galactose (Cambridge Isotopes), and 200 µM GDP-[U-$^{13}C_6$]-Fucose. The PgtA reaction was supplemented with 0.01 U/µl calf intestine alkaline phosphatase (CIAP, New England Biolabs) to reduce product inhibition from UDP and GDP, and 10 µg/ml Leupeptin, 10 µg/ml aprotinin, and 5 mM benzamidine for reducing potential proteolysis which was observed in their absence. GDP-[U-$^{13}C_6$]-Fucose, in brief, was prepared enzymatically using L-fucokinase/GDP-fucose pyrophosphorylase (FKP)[59] and [U-$^{13}C_6$]-Fucose (Cambridge Isotopes), and the final product was purified over a P2 column and estimated to be ~85% purity based on LC-ESI mass spectrometry analysis (not shown). The contamination is likely composed of reaction components and byproducts such as nucleotides and phosphatases that are not expected to intefere with Skp1 glycosylation reactions. The PgtA reaction proceeded overnight at room temperature, and the progress was monitored by MALDI-TOF-MS. The resulting ^{15}N,^{13}C-His-FGGnSkp1Δ was purified using a 5 ml Q-column as described above.

Q-column purified fractions containing ^{15}N,^{13}C-His_6FGGnSkp1Δ were modified with terminal disaccharide -Gal-Gal using AgtA. The reaction condition was carried out in a 10 kDa cut-off centrifugal ultrafiltration cartridge and consisted of 150 µM ^{15}N,^{13}C-His_6FGGnSkp1Δ, 3 µM AgtA (50:1 molar ratio), 75 mM HEPES-NaOH (pH 7.4), 75 mM NaCl, 600 µM UDP-[U-$^{13}C_6$]-Gal, and 0.01 U/µl CIAP. This reaction was supplemented with 1 mM PMSF in addition to the other protease inhibitors mentioned above. The reaction proceeded overnight at room temperature, and its progress was monitored by MALDI-TOF-MS. The resulting U-^{15}N,^{13}C-His_6GGFGGnSkp1Δ reaction was exchanged into a buffer containing 50 mM Na^+/K^+ phosphate (pH 7.4) and 300 mM NaCl, and then

applied over a Talon column to isolate Skp1 as described above. 300 mM imidazole (in the buffer described above) elution fractions containing $^{15}N,^{13}C$-His_6GGFGGnSkp1Δ was dialyzed using a 10 kDa cut-off membrane against 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM TCEP, then supplemented with Tev protease overnight. His_6 cleaved U-$^{15}N,^{13}C$-GGFGGnSkp1Δ was retrieved over a Talon column. The sample was concentrated and applied over a Superdex-75 size-exclusion column (HiLoad 16/600, Cytiva) preequilibrated with the NMR sample buffer containing 20 mM MES-NaOH (pH 6.0), 50 mM NaCl, 5 mM DTT, 0.05% NaN_3, 10 µg/ml Leupeptin, and 10 µg/ml aprotinin. The purity and the reaction efficiency were assessed via SDS-PAGE and MALDI-TOF-MS, and the content was estimated to be 90% fully modified with GGFGGn-. The eluted samples were immediately prepared to be loaded into NMR tubes for data aquisition.

NMR sample preparation and data aquisition

$^{15}N,^{13}C$-GGFGGnSkp1Δ fractions from a Superdex-75 purification were immediately concentrated to 2 mM concentration using a 10 kDa cut-off centrifugal ultrafiltration cartridge, and the concentration was calculated using OD_{280} value determined using a NanoDrop UV-vis (Thermo) and the extinction coefficient of Skp1Δ (9970 M^{-1} cm^{-1}) which was calculated using ProtParam[61]. The concentrated sample were supplemented with 3% (v/v) D_2O for spectrometer lock and 30 µM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) for spectral reference. Additional samples including 300 µM $^{15}N,^{13}C$-GGFGGnSkp1Δ and 300 µM $^{15}N,^{13}C$-GnSkp1Δ were prepared in a similar manner to the 2mM sample. 35 µl of each sample was loaded into 1.7 mm capillary tubes, and NMR spectra were recorded at 35°C using a Bruker NEO 800 MHz spectrometer equipped with a 1.7 mm TXI H{CN} cryoprobe. The recorded NMR spectra are summarized in Table

4.1. Fourier transform and phase adjustments were performed with NMRPipe, and data collected with non-uniform sampling were reconstructed with SMILE module within NMRPipe. Processed NMR spectra were visualized using CARA. These softwares were accessed virtually via NMRbox server[60].

MALDI-TOF mass spectrometry analysis

To assess the degree of ^{15}N and ^{13}C isotopic incorporation in Skp1 samples and the efficiency of Skp1 glycosylation reactions, the masses of Skp1 samples were monitored using MALDI-TOF mass spectrometry. In brief, Skp1 samples were desalted using reverse phase C-4 columns (TopTips, PolyLC INC.). Then, the samples were spotted onto a steel target plate with an equal volume of saturated sinapinic acid dissolved in 50% acetonitrile and 0.1% trifluoroacetic acid. The samples were air-dried at room temperature. The data were collected using AB SCIEX TOF/TOF 5800 MALDI mass spectrometer operated in positive ion and linear mode settings. The *m/z* peaks were modeled using gaussian distribution fitting were analyzed using Data Explorer (Applied Biosystems).

4.3 Results and Conclusion

RESULTS

Assessing the degree of isotopic incorporation and glycosylation of U-^{15}N,^{13}C-GGFGGnSkp1Δ sample

MALDI-TOF mass spectrometry was used to first assess the incorporation of ^{15}N and ^{13}C isotopes into *Dictyostelium* U-^{15}N,^{13}C-His-GnSkp1Δ by comparing its mass against that of natural abundance His-GnSkp1Δ ***(Figure 4.2A).*** Doubly-charged *m/z* species were analyzed, because they were sharper and well-defined when compared to the

singly-charged species that showed broad peaks (not shown). The spectra show multiple peaks, and the major peak of each sample was interpreted to be either natural abundance His-GnSkp1Δ and U-^{15}N,^{13}C-His-GnSkp1Δ ***(Figure 4.2A).*** The doubly-charged *m/z* species of natural abundance His-GnSkp1Δ was observed at 10137.1 which agrees with the expected *m/z* of 10138.5. The minor discrepancy of observed and expected values were within the noise of the mass spectrometer. The expected mass difference between U-^{15}N,^{13}C-His-GGFGGnSkp1Δ and His-GnSkp1Δ, based on average masses, was calculated to be 564 *m/z*, and the observed mass difference was 528.7 *m/z*, which indicated ^{15}N and ^{13}C labeling efficiency of 93.7%. Complete isotopic labeling was not achieved likely due to the residual natural abundance nutrients carried over from LB media into M9 labeling media.

This U-^{15}N,^{13}C-His-GnSkp1Δ sample was further glycosylated *in-vitro* using the recombinantly expressed and purified *Dictyostelium* Skp1 glycosyltransferases, PgtA and AgtA, and sugar nucleotides containing U-^{13}C-labeled sugar. The resulting U-^{15}N,^{13}C-His-GGFGGnSkp1Δ was subjected to TEV digestion and a series of purification steps ending with a Superdex 75 size-exclusion purification, yielding a highly purified U-^{15}N,^{13}C-GGFGGnSkp1Δ sample ***(Figure 4.2B)***. An aliquot of U-^{15}N,^{13}C-His-GnSkp1Δ was processed in a similar manner excluding the glycosylation steps, yielding U-^{15}N,^{13}C-GnSkp1Δ ***(Figure 4.2B)***.

The Superdex-75 purified ^{15}N,^{13}C-His-GnSkp1Δ and U-^{15}N,^{13}C-GGFGGnSkp1Δ were compared by MALDI-TOF MS to evaluate the degree of glycosylation ***(Figure 4.2C)***. The expected mass difference between the doubly-charged ions representing Gn- and GGFGGn-Skp1Δ was calculated to be 328.1 *m/z*, and the observed difference was 329.4

m/z, which indicated that the sample is consistent with a high degree of pentasaccharide modification ***(Figure 4.2C)***. The other minor peaks were observed which are consistent with a potential unmodified or hydroxylated form of Skp1Δ and GGn-Skp1Δ ***(Red arrows, Figure 4.2C).*** These observations indicate incomplete *in vivo* processing of Skp1Δ into GnSkp1Δ while expressing in *E. coli.* and incomplete *in vitro* addition of the αFuc onto GGn-Skp1Δ by PgtA, respectively. Integration of these peaks indicated that 91% of Skp1Δ is fully modified with pentasaccharide and the remaining 9% of Skp1Δ is either unmodified, hydroxylated, or GGn- Skp1Δ.

Comparison of GGFGGnSkp1Δ and previous Skp1ΔΔ spectra

To evaluate the quality and integrity of GGFGGnSkp1Δ sample, its ^{15}N-TROSY-HSQC spectra were overlaid and compared with the ^{15}N-HSQC spectrum of Skp1ΔΔ which lacks the C-terminal region ***(Fig 4.3A)***. For the more peripheral regions displaying the well-folded region of the protein, both spectra demonstrated high level of correspondence indicating that GGFGGnSkp1Δ is properly folded. Additional peaks for GGFGGnSkp1Δ are detected towards the middle region of the spectra where signals from flexible and mobile regions appear ***(Fig 4.3B)***. These peaks correspond to the residues from the flexible C-terminal region of Skp1 which are absent in the Skp1ΔΔ ***(Fig 4.1).***

Glycan extension induces chemical shifts

GGFGGnSkp1Δ and GnSkp1Δ spectra were compared to assess whether there are chemical shifts associated with having the glycan extended from monosaccharide to the full pentasaccharide. ^{15}N-TROSY-HSQC spectra for both glycoforms at equal concentrations (300 μM) were recorded and the spectra are shown in Figure 4.3. The signals overlay to a high degree in the peripheral region, indicating that the folded core of

the protein is not perturbed when the glycan extended. Interestingly, when the middle of the spectra is examined, there are a number of shifted peaks ***(Figure 4.4)***, which indicates that the glycan is perturbing these residue's chemical environment.

All sugar anomeric carbon signals are detected

To evaluate the quality of carbon signals from the glycan, ^{13}C-HSQC spectrum was recorded ***(Figure 4.5A)***, and the region associated with anomeric carbons (H^1= 4.5-6 ppm, C^{13}=90-110 ppm) was examined. Five strong cross peaks corresponding to the anomeric carbons of the Skp1 pentasaccharide are well resolved, indicating high integrity of the glycosylated Skp1 sample ***(Figure 4.5B)***.

Conclusion

Here we demonstrate that our new glycosylated sample ^{15}N,^{13}C- GGFGGn-Skp1Δ is stable and appears to be amenable to in-depth NMR analysis. Previous NMR studies on glycosylated Skp1 required perdeuteration of the sample to assign chemical shifts and were conducted at pH 7.0 in an effort to be closer to a physiological pH[26]. Now with a protonated sample we have access to the full collection of NOESY based spectra at pH 6.0 and 35°C that can detect molecular interactions from proton-proton contacts. These experiments would be particularly useful in testing our model that the glycan makes hydrogen bonds and van der Waal contacts with the neighboring regions of the Skp1.

A preliminary review of several NMR spectra indicated a promising start to the analysis. When GGFGGn-Skp1Δ and Skp1ΔΔ ^{15}N-HSQC spectra were compared ***(Figure 4.3A)***, most peaks demonstrated identical chemical shift pattern which showed that the core protein fold is the same since Skp1ΔΔ is only the folded part of Skp1. Therefore, the new peaks that arise are expected to be resonances associated with the C-terminal region of the

protein ***(Figure 4.3B)***. GGFGGn-Skp1Δ was compared with GnSkp1Δ sample that was withheld from further glycosylation processing ***(Figure 4.4)***. At equal concentrations, the peaks generally aligned identically, with the exception of several that likely are due to the effect of glycan extension. The GGFGGn-Skp1Δ samples were also uniformly labeled with ^{13}C sugars and each anomeric carbon from the five sugar moieties were clearly visible attesting to the integrity of the glycan on Skp1 ***(Figure 4.5).***

Many questions remain to be answered with information which we believe our new set of data can provide. Is the glycan making interactions with Skp1 as modeled by MD simulations, therefore does it show any NOEs between the glycan and the protein? Are there any clues regarding contacts between the C-termini and the respective glycans that might affect Skp1 dimerization? Through this ongoing analysis we hope to refine our model on how the glycan impacts Skp1 structure.

4.4 Acknowledgements

We thank Dr. Anthony Prudden and Dr. Geert-Jan Boons for the enzymatically synthesized GDP-[U-$^{13}C_6$]-Fucose for PgtA reactions. We thank Dr. Alex Eletsky for his advice and assistance with NMR data acquisition and analysis.

4.4 Figures and Tables

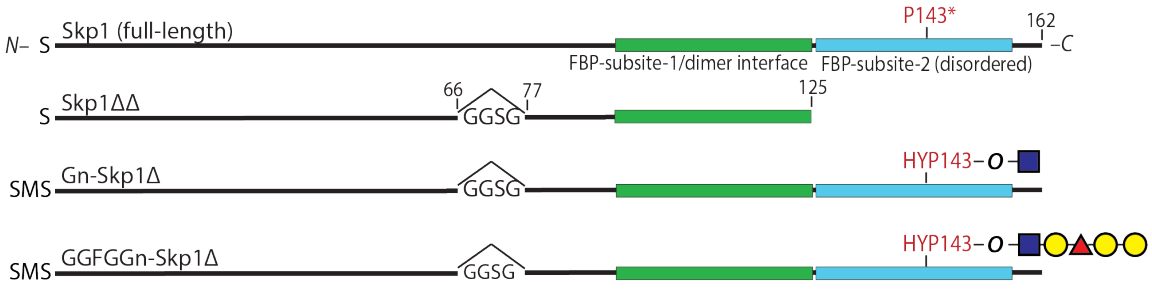
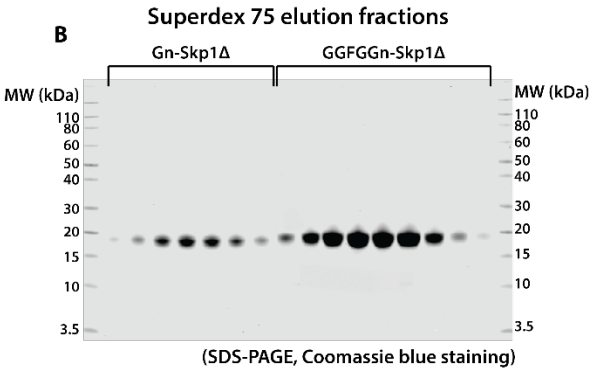


Figure 4.1. Skp1 constructs used in this study. A diagram showing the domains of Skp1 recombinantly expressed Skp1. Asterisk denotes Skp1 Proline-143 position that is hydroxylated and glycosylated in *Dictyostelium*. Note that a Ser-Met extension occurs at the N-terminus upon His_6-tag removal, compared to the Ser- N-terminus that occurs after normal proceesing in *E. coli*. The Skp1 construct was coexpressed with PhyA and Gnt1 to yield Gn-Skp1. Subsequent glycosylation reactions were carried out *in vitro* using purified PgtA and AgtA.

A

Sample	Expected *m/z*	Observed *m/z*
Natural abundance His-GnSkp1Δ	10138.5	10137.1
U-^{15}N,^{13}C His-GnSkp1Δ	10702.5	10665.8

Expected Δ = 564
Observed Δ = 528.7
Labeling efficiency = 93.7 %



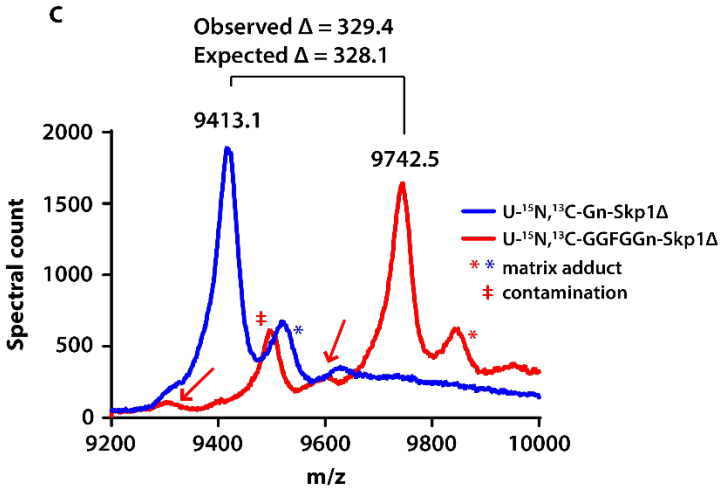


Figure 4.2. Mass analysis of the glycosylated Skp1Δ sample used for NMR data acquisition. (A) A table of expected and observed doubly-charged *m/z* species (average masses) of natural abundance and uniformly ^{15}N and ^{13}C labeled His-GnSkp1Δ. Natural abundance His-GnSkp1 and U-^{15}N,^{13}C-His-GnSkp1Δ were expressed recombinantly in *E. coli* grown in LB media (natural isotopic abundance) or M9 media with ^{13}C-glucose and ^{15}N-NH_4Cl nutrient source. (B) A coomassie blue stained SDS-PAGE gel loaded with Superdex-75 elution fractions of fully processed (further glycosylation and/or Tev treatment) and purified U-^{15}N,^{13}C- GnSkp1 and U-^{15}N,^{13}C- GGFGGnSkp1. These samples were pooled for NMR analysis. (C) Mass spectra comparing the doubly-charged *m/z* species of U-^{15}N,^{13}C-GnSkp1 and U-^{15}N,^{13}C-GGFGGnSkp1 shown in panel B. Blue and red asterisks denote matrix adduct peaks that are commonly present when Skp1 is evaluated by MALDI-TOF-MS. The peak around 9500 *m/z* marked with red double dagger in the U-^{15}N,^{13}C-GGFGGnSkp1 is a contaminant from PgtA or AgtA stock solution that co-purifies with Skp1 and would not be visible by NMR since they are not isotopically enriched. This mass does not appear in ^{15}N,^{13}C-GnSkp1 sample, because it was not treated with PgtA and/or AgtA. The red arrows point to *m/z* values that correspond to HO-Skp1Δ and GGn-Skp1Δ which appear as a result of incomplete GlcNAcylation during expression in *E. coli* and incomplete PgtA reaction *in vitro*, respectively.

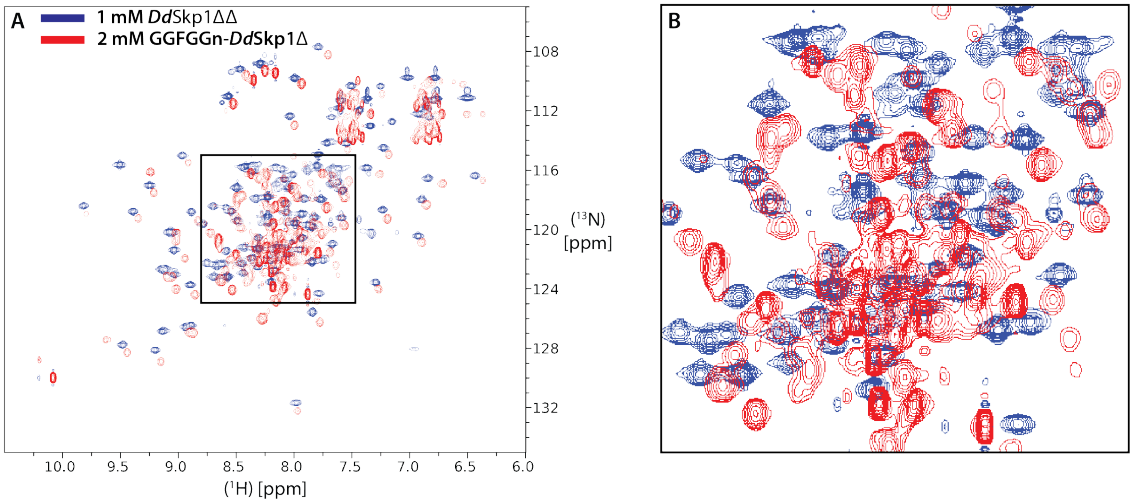


Figure 4.3. GGFGGn-Skp1Δ ^{15}N-TROSY HSQC and Skp1ΔΔ ^{15}N-HSQC spectra overlay. The diagonal shifts in corresponding peaks between the two spectra occur due to effect of TROSY HSQC data acquisition . (A) 2 mM GGFGGn-Skp1Δ (blue) peaks show a high degree of similarity in the well-ordered regions of the spectra when compared with Skp1ΔΔ peaks (red; see Figure 4.1 for protein construct details). (B) Zoomed-in image of the boxed area in panel A shows peaks associated with mobile and dynamic region of the Skp1. Extra peaks shown in GGFGGn-Skp1Δ likely correspond to the glycosylated C-terminal region.

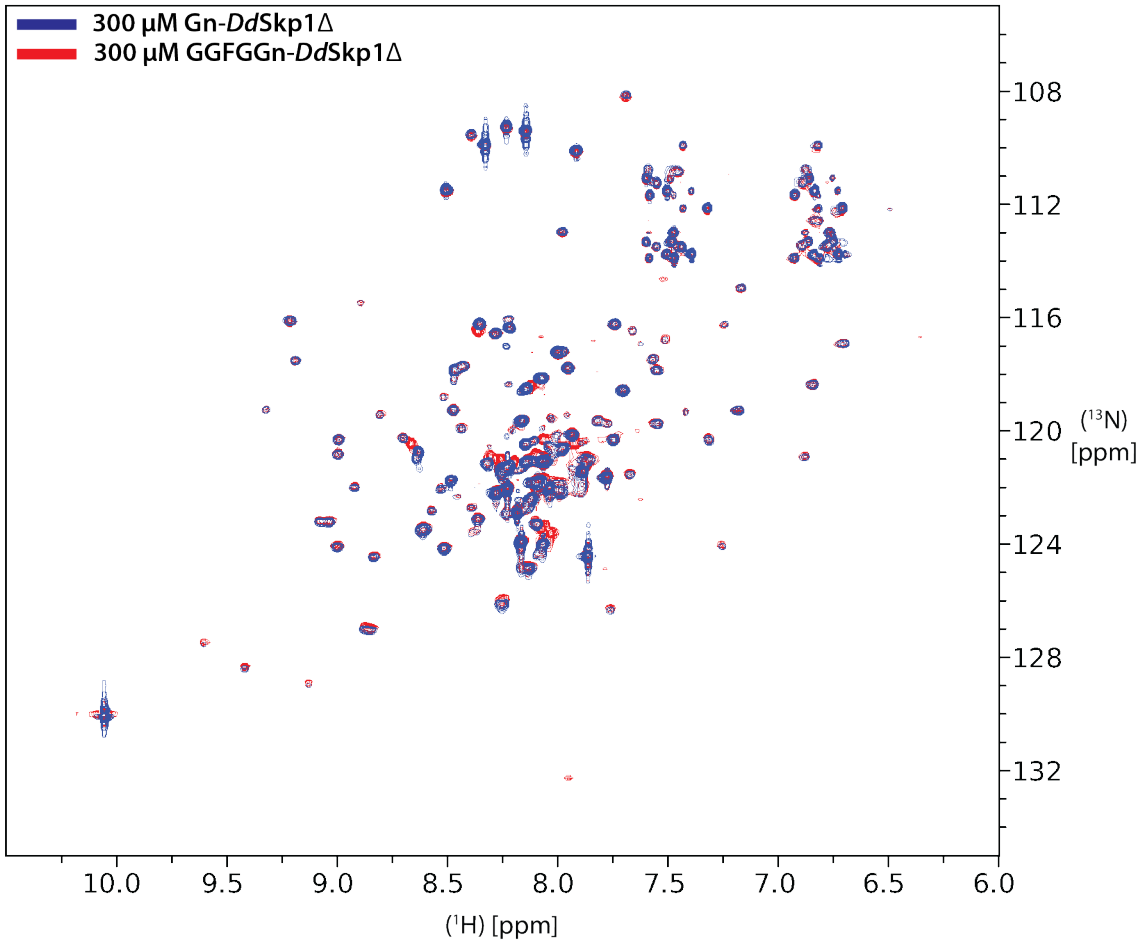


Figure 4.4. Extension of the glycan induces chemical shifts. Shown is an overlay of GGFGGn-Skp1Δ (red) and GnSkp1Δ (blue) ^{15}N-TROSY HSQC spectra. Fully glycosylated Skp1 demonstrates peak shifts near the flexible and mobile peptide region (middle) of the spectra when compared with Skp1 with a single sugar modification. Both spectra were recorded at 300 µM protein concentration for equal signal-to-noise and controlled comparison.

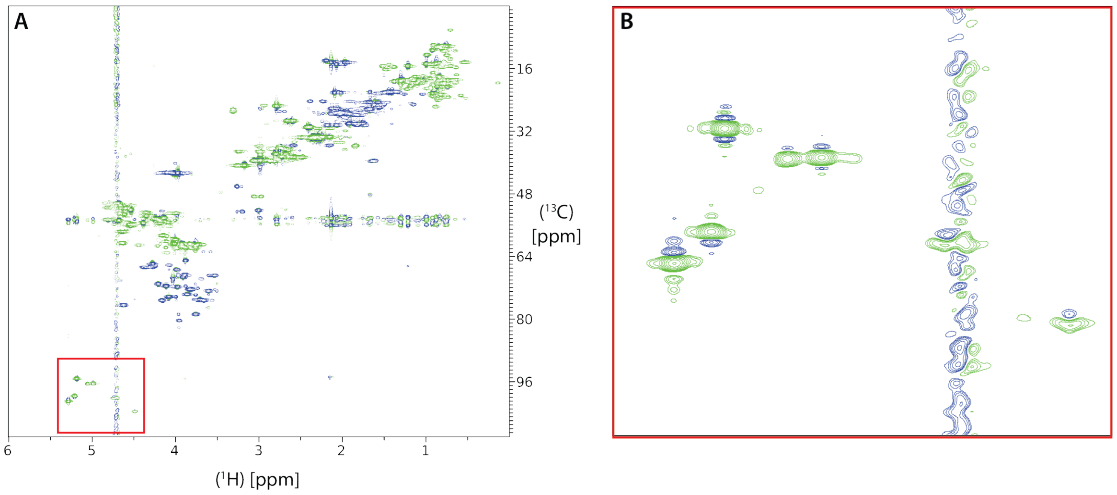


Figure 4.5. GGFGGnSkp1Δ ^{13}C-HSQC spectrum and its sugar anomeric carbon region. (A) A ^{13}C-HSQC spectrum of 2 mM glycosylated Skp1Δ showing signals from the protein sidechains and the sugars. (B) Zoomed-in image of the red-boxed area in panel A shows anomeric carbon signals. Five strong peaks correspond to each sugar of the Skp1 glycan. The weak, disformed signal located upfield of the water line is likely an artifact.

Sample	Experiment	Magnetic field (MHz)
300 µM GnSkp1Δ	2D [^{15}N, ^{1}H]-TROSY-HSQC	800
300 µM GGFGGnSkp1Δ	2D [^{15}N, ^{1}H]-TROSY-HSQC	800
2 mM GGFGGnSkp1Δ	2D [^{15}N, ^{1}H]-TROSY-HSQC	800
	2D H-N plane of 3D HNCO	800
	2D H-N plane of 3D TROSY-HNCO	800
	2D H-N plane of 3D CBCA(CO)NH	800
	2D H-N plane of 3D CBCA(CO)NH-TROSY	800
	3D CBCA(CO)NH-TROSY	800
	2D 13C CT-HSQC aliphatic	800
	2D 13C CT-HSQC aromatic	800
	2D 13C TROSY aromatic	800
	2D HN-C plane of CBCA(CO)NH	800
	2D HN-C plane of HNCACB	800
	2D H-N/C plane of simNOESY	800
	3D simNOESY	800
	3D TROSY-HNCACB (25% NUS)	800
	3D TROSY-HNCO (25% NUS)	800
	3D CBCA(CO)NH (25% NUS)	800
	3D HBHA(CO)NH (12.5% NUS)	800
	3D (H)CCH-COSY ali, (20% NUS)	800
	3D (H)CCH-TOCSY ali, (20% NUS)	800
	3D CBCA(CO)NH (25% NUS)	800
	3D HBHA(CO)NH (12.5% NUS)	800
	3D (H)CCH-COSY aromatic	800
	2D H-CO plane of 3D HN(CA)CO	800
	3D HN(CA)CO (25% NUS)	800
	3D (H)CC(CO)NH-TOCSY (12.5% NUS)	800
	3D H(CCCO)NH-TOCSY (25% NUS)	800
	2D H-CA plane of 3D TROSY-HNCA	800
	3D TROSY-HNCA (31.25% NUS)	800
	2D H-N plane of 3D TROSY-HNCO	800

Table 4.1. List of NMR experiments for U-^{15}N,^{13}C-GGFGGnSkp1Δ and U-^{15}N,^{13}C-GnSkp1Δ

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Conclusions

Conservation of Skp1 glycans from unrelated protists

The O_2-sensing pathway through Skp1 pentasaccharide modification was initially found and characterized in *Dictyostelium*.[34-39] The question of whether it might be an abundant process in other protists was addressed with the discovery of PhyA and Skp1 GT homologues in the unrelated parasite *Toxoplasma gondii.*[42] Although the enzymes for synthesizing the first three sugars (Gnt1 and PgtA) were conserved, the GTs that catalyze the terminal disaccharides were different.[39,43,44] Instead of having AgtA catalyze addition of the last two sugars as in *Dictyostelium*, *Toxoplasma* was found to have two unrelated GTs, Glt1 and Gat1, that modify the 3-position of the third sugar, αFuc, with Galα1,3-Glcα1- instead of –Galα1,3Galα-.[43,44] Even with the different sequence, the two glycans were highly similar with the only difference occurring at the stereochemistry at the C'4 hydroxyl position (equatorial for Glc, axial for gal) ***(Figure S2.14A)***.[44]

At the time when a CAZy GT-8 family enzyme was suspected to be Gat1, it was annotated as a glycogenin, a well-characterized enzyme involved in priming and polymerization of –Glcα1,4-Glcα1,4- linkages in the formation glycogen from yeasts and animals. This led to the investigation of whether or not Gat1 might be involved in starch, which contains glucose in the same linkages as in glycogen. Interstingly, Gat1 knock-out

Toxoplasma did not demonstrate starch deficiency, therefore not Gat1, but another enzyme might be involved in priming starch formation, if priming is even needed by starch synthase ***(Figure 2.5)***. Examination of recombinant Gat1 showed preferential specificity for hydrolysis of UDP-Gal instead of UDP-Glc, and no evidence for auto-glycosylation which suggested that Gat1 is a galactosyltransferase that targets acceptors on a different protein(s) ***(Figure 2.4B)***. Acceptor assays with a variety of potential sugar acceptors indicated that Gat1 is selective for a Skp1 tetrasaccharide with a terminal α-glucoside, compared to starch like non-reducing terminal oligosaccharides, and a higher level of transfer occurs when it is attached to Skp1. Radiotracer analysis of *gat1*-KO extracts supplemented with recombinant Gat1 and ^{3}H-labeled sugar nucleotides showed that only incorporation was achieved at 20 kDa (Skp1 MW ~20 kDa), supporting high specificity of Skp1 as the sole substrate of Gat1 ***(Figure 2.4J)***. To further investigate the substrate specificity, the crystal structure of the ortholog of Gat1 from *Pythium ultimum* was determined, and computational modeling of the substrates indicated that the Gat1 active site accommodates binding to UDP-Gal and Skp1-tetrasaccharides in a complementary manner that cannot be achieved with glycogenin substrates UDP-Glc and a starch or glycogen polymer ***(Figures 2.8,2.9)***. Interestingly, Gat1 forms a homodimer whose dimer interface almost perfectly matches that of glycogenin homodimer, and the interface also forms a groove that hosts the acceptor glycan. Glycogenin, rather, requires conformational flexibility for priming a glucose moiety onto its Tyr-195 due to its distance from active site, and subsequent glucosylation (up to 10-12 moieties) is dependent on recognition of 3 terminal glucoses by a single subunit[62]. This work altogether supported that Gat1 is a galactosyltransferase specific to

the modification of Skp1, and was potentially repurposed as glycogenin when the Skp1 modification pathway was lost in evolution.

The first structural model of glycosylated Skp1 was provided by NMR-guided MD simulation which showed that the glycan interacts with Skp1 polypeptide and affects Skp1 conformation in a manner that is more conducive to interact with F-box proteins. The structure of glycosylated *Toxoplasma* Skp1 was examined to address whether the different sequence of its glycans was compatible with the current model of its structural relationship with Skp1 in *Dictyostelium.* MD simulations on glycosylated *Toxoplasma* Skp1 showed that the glycan behaved in a similar manner to previous simulations of *Dictyostelium* Skp1 ***(Figure 2.10)***. The simulations showed that the helix-8 was oriented outward as expected, and interactions including hydrogen bonds and glycan-protein packing were similarly observed especially for the fifth sugar that was added by Gat1. Examination of the C'4 hydroxyl of the fourth sugar showed it to be exposed to the solvent and not involved in glycan-protein contacts throughout the course of the simulation. This indicates that the 4^{th} sugar is amenable to either Glc or Gal, and is consistent with the idea that the glycan serves a structural role within Skp1 rather than as an interface with another protein, unless that protein was conserved and had evolved to accommodate the distinct stereochemistry of the different glycan. Nevertheless, these simulations require confirmation by additional experimentation.

Orientation of the Skp1 dimer and the relative positioning of the C-terminus and the glycan

Examining how glycan impacts the dimer of Skp1 has been elusive owing to the lack of knowledge on how the dimer is assembled. First we tried crystallizing Skp1 using a trimmed version that is missing the internal loop and the disordered C-terminal tail in order to ensure crystal packing is not disturbed by these regions ***(Figure 3.2A)***. However, only 2D crystals formed, and we were unable to optimize crystal packing in the third dimension to obtain crystals sufficient for data collection. Since others have also been unable to crystallize free Skp1 from other species, we employed NMR spectroscopy as an alternative for structural determination. The reduced size of the truncated Skp1 allowed for sufficient signal-to-noise and spectral resolution to calculate a high-resolution structure.

The dimer structure was composed of a 4-helix bundle with an interface composed of predominantly hydrophobic contributions from the pair of two helices (5 and 6) from each subunit ***(Figures 3.2B,C,D)***. Comparison of the Skp1 dimer structure with the monomer in complex with FBP showed that the dimer interface overlaps substantially with FBP binding subsite-1 ***(Figure 3.2E,F)***. This provided an explanation as to how Skp1 is a dimer in its free state, but a monomer in the Skp1/FBP complex.[10] Furthermore, the overlapping of the interfaces suggests that Skp1 binding to FBP can be competed by Skp1 dimerization albeit the difference of ~100 fold in binding affinity, atleast for *Dictyostelium* Skp1 *in vitro*. Interestingly, preliminary evidence suggests that the K_D for *Toxoplasma* Skp1 homodimerization may be similar to that of the FBP interaction.

The dimer structure allows speculation on the placement of the C-terminal tail harboring the glycan. With the dimer configuration, the interface helices are arranged in a

semi-parallel fashion where the C-termini would in close proximity with each other and to the dimer interface. This suggests that the relatively disordered C-terminus including the glycan would have the potential for inter-subunit interactions as well as with the central dimer interface and potentially the internal disordered loop.

A promising start to NMR analysis of a glycosylated Skp1

The dimer structure allowed us to orient the glycosylated region with respect to each subunit. With the ultimate goal of understanding how the glycan impacts Skp1 structure, we reintroduced the C-terminus and the glycan, and everything including the protein and the sugars was uniformly incorporated with ^{13}N and ^{13}C isotopes in a protonated background. This particular sample ***(GGFGGn-Skp1Δ)*** is advantageous when considering that previous glycosylated Skp1 studies were done with perdeuterated sample due to the size of the dimer, therefore was not amendable to the full library of ^{1}H-^{1}H NOESY based experiments to probe protein-glycan interactions. Also, pH of the NMR sample was lowered to 6.0 (previously 7.0) in order to reduce the rate of amide proton exchanging with the solvent.

With the new set of data including NOESY we seek to assess Skp1-glycan interactions, both those predicted as described above as well as new possibilities. The early analysis on this set of collected NMR spectra ***(Table 4.1)*** indicated that the sample is soluble, stable, nearly fully ^{15}N and ^{13}C labeled, and nearly fully glycosylated. When GGFGGn-Skp1Δ and Skp1ΔΔ ^{15}N-HSQC spectra were compared ***(Figure 4.2A)***, the peaks showed identical chemical shift pattern which indicated that the core protein fold is the same since Skp1ΔΔ includes only the stably folded parts of Skp1 based on our NMR studies. Therefore, the new cross-peaks unique to GGFGGn-Skp1Δ are expected to be

resonances associated with the C-terminal region of the protein ***(Figure 4.2B)***. GGFGGn-Skp1Δ was compared with GnSkp1Δ sample that was withheld from further glycosylation processing ***(Figure 4.3)***. At equal concentrations, the peaks generally aligned identically, with the exception of several that likely are due to the effect of glycan extension. The GGFGGn-Skp1Δ samples were also uniformly labeled with ^{13}C sugars and each anomeric carbon from the five sugar moieties were clearly visible attesting to the integrity of the glycan on Skp1 ***(Figure 4.4)***.

5.2 Future directions

My dissertation work has expanded our knowledge on Skp1 glycosylation system in protists and tuned the structural model of Skp1. Here, I propose future directions and experiments to further refine our understanding on the structural mechanism of Skp1 glycan and how it affects homodimerization and modulates SCF complex formation.

Determination of glycosylated Skp1 dimer affinity

While the effect of glycosylation on Skp1 dimerization is documented quantitatively[57], precise quantitative measurements have yet to be performed in order compare the affinity of the Skp1 dimers under various glycosylation states. Unmodified Skp1 dimer affinity has been determined to be 2.5 μM K_D based on AUC studies of *Dictyostelium* Skp1[56] ***(Chapter 3)***. More recently, AUC analysis on *Toxoplasma* Skp1 showed that the dimer affinity is estimated to be low nanomolar K_D (a difference of ~2 orders of magnitude), which suggests that different species might have evolved varied Skp1 homodimer affinities to tune the availability of monomeric Skp1 for FBP binding according to their own needs (Cantrell et al., unpublished).

Skp1 expression construct, GT enzymes, and other reagents are available in the West lab to prepare glycosylated *Dictoystelium,* and *Toxoplasma* samples which we can test using the same AUC methods employed for unglycosylated Skp1. Preliminary evidence already exists for over an order of magnitude decrease in affinity for *Toxoplasma* Skp1 owing to glycosylation (Cantrell et al., unpublished)

We expect to see that the glycosylated Skp1 will demonstrate K_D values greater than unmodified Skp1 dimer K_D based on the unglycosylated sample affinity and biochemical results indicating the glycan inhibits dimerization. The difference in the value of K_D can be used to derive thermondynamics contribution of the glycan to the dimerization system and perhaps can be correlated to newly emerging NMR data on the glycan-protein contacts.

Assessing the contribution of Skp1 internal loop on Skp1 dimer affinity

Skp1 dimerization so far has been investigated from the glycan and the disordered C-terminus region point of view. However, the disordered internal loop that is present in Skp1 has yet to be thoroughly evaluated beyond finding that MD studies do not suggest meaningful conformations. Based on our dimer structure, the loop may be residing in locations where it can potentially interact with either the dimer interface or the disordered C-terminal region of Skp1.

Preliminary AUC analysis of internal loop truncated *Dictyostelium* Skp1 (Skp1Δ) showed an absence of monomers at 5 μM concentration where the full-length Skp1 showed a substantial degree of monomers, which indicated that Skp1Δ dimer displays a higher apparent affinity (unpublished). Based on this information, a model of how the internal

loop can affect Skp1 dimerization can be tested. Perhaps Skp1Δ is more stable and consequentially forms a more stable dimer.

A contrasting observation was made when assessing the dimer interface substitution (F97E, see chapter 3) in both full-length Skp1 and Skp1Δ background. AUC analysis at 100 μM protein concentration showed that despite having a mutation in the dimer interface, full-length Skp1(F97E) formed predominantly dimers indicating the mutation did not impact the dimer affinity to a degree to dissociate at the tested concentration (unpublished). Surprisingly, Skp1Δ(F97E) demonstrated to be monomers at 100 μM, which indicated that absence of the internal loop led to a more stable monomer. This suggests that the loop had a positive effect in stabilizing the dimer in the F97E substitution background. Perhaps the glutamic acids (E97) are stabilized by electrostatic interactions from the loop side chains, where in absence would cause electrostatic repulsion between the dimers.

We have *Toxoplasma* and *Dictyostelium* Skp1 constructs with various combination of internal loop and/or disordered C-terminal region truncations to assess the dimer dependence on the internal loop. This can also be tested in combination of the disordered C-terminal region and glycosylation, which may bring insights into the cross-talk between the internal loop, the disordered C-terminal region, and the glycan. The specificity of the internal loop can be tested by replacing the loop with different amino acid compositions and various loop lengths. This approach could conveivably be applied to the C-terminal region as well. The proposed constructs can first be examined by AUC to assess to quantify the effect of the loop on dimerization, and further studies can be performed using NMR to

assess whether its effect on dimerization is a consequence of intermolecular contacts or motions of the internal loop.

Resonance assignment of GGFGGn-Skp1Δ

A preliminary analysis of GGFGGn-Skp1Δ indicated that the sample is amenable for futher NMR analysis based on HSQC spectra, and all of the five sugar signals were accounted for. Moving on, the amide backbone, the sidechain and, the glycan resonances need to be assigned in order for further in-depth analysis. The BTB core of the GGFGGn-Skp1Δ and Skp1ΔΔ is essentially identical, based on resonance assignments for Skp1ΔΔ as indicated by spectra overlay. Therefore, instead of *de novo* assignments, the information from Skp1ΔΔ can be used to expedite the process. However, assigning resonances of the C-terminal region will require careful assessment, as the peaks from the flexible regions appear closely together often overlapping showing ambiguous signals. Once the resonance assignment is complete, NOESY spectra can be examined for potential contacts between the glycan and Skp1. First, the sugar moieties and the parts of the protein that were predicted to interact based on MD should be prioritized.

Analysis of Skp1 with glycan contact residue substitutions

Once the glycan contacts are proposed, it would be informative to substitute amino acids at the residue positions involved with glycan contact to probe the specifity of the interaction. The substitution can be selected based on sequences of Skp1 glycan from species that do not have the Skp1 glycosylation pathway, or from species with a second Skp1 that is not subject to modification, and therefore would not have been selected for glycan contact residues.

Other potential NMR approaches for investigating the motions of Skp1 and the glycan

The recent advancements in examining glycosylated Skp1 by NMR has been focused at the molecular interactions between Skp1 and the glycan predicted by MD. However, the other prediction by MD suggesting that the glycan promotes the extension of helix-8 remains to be investigated.

First, conventional relaxation studies are warrented, especially with the newly emerging chemical shift assignments which would allow probing residue specific motions of the Skp1 disordered C-terminal region and the glycan. The current NMR methods on dynamics can reveal relative motions of bonds, collective motions of the loop, and relative motion relative to stably folded Skp1 domain. This would require perdeuteration of glycosylated Skp1 to accurately measure relaxation values (R_1 and R_2).

Further spacial orientation of these regions can be probed using long-range pseudo contact shifts (PCSs) and paramagnetic relaxation enhancements (PREs) which can be used to detect ranges of motions up to 40 Å, unlike the traditional NOESY approaches which are limited to much shorter distances less than 5 Å. PCSs would require insertion of a lanthanide binding loop by modifying the internal loop of Skp1, and PREs would require the modification of a cysteine residue with a paramagnetic probe. Both are viable approaches, however are highly dependent on the stability of Skp1 when either modification is introduced.

Analysis of Skp1 with altered glycan sequence and linkages

Alternatively, perhaps the glycan could be altered in a way that could form a pentasaccharide with the combination of different sequences and linkages to probe whether the mode of Skp1 glycan mechanism is sequence specific. This might be feasible as GTs

are highly diverse and at times found to be promiscuous towards acceptor substrates. However, in practice, this might not be achievable due to the unpredictability of GT activity on Skp1. Therefore, unless a high-throughput equipment is employed to express, purify, and test GT activity on Skp1 or activity in the cytosolic environment, this might not be achievable in a timely fashion.

NMR-guided MD simulations

Previously, the MD simulations were performed with a monomeric Skp1 with the C-terminal region modeled as α-helices as it is in Skp1/FBP crystal structures. Therefore, with future simulations, the C-terminal region will be modeled as an unstructured peptide that is consistent with backbone geometries derived from chemical shifts. The F97E substitution was in part engineered to experimentally examine how the glycan and Skp1 is behaving as a monomer, however, this construct was prone to aggregation during the course of NMR data acquisition. Therefore, other substitutions predicted to disrupt the dimer interface as predicted previously by Rosetta can be explored for a more stable monomeric Skp1 construct.

If we are successful in gathering structural information from GGFGGn-Skp1Δ (dimer) by NMR, these data can be useful in providing a visual model by MD simulations. Any contacts that we observe can be used as constraints for simulations, thereby yielding structures that are consistent with NMR data. Plausible monomer and dimer structural models of glycosylated Skp1 would lead to further understanding on how the glycan serves the dual role by first inhibiting homodimer formation and then promoting monomer conformations to be more receptive to FBP binding.

LIST OF REFERENCES

CHAPTERS 1, 4, & 5

1. Richter, K., Haslbeck, M., Buchner, J. (2010) The heat shock response: life on the verge of death. *Mol Cell.* **40**, 253-266.

2. Casamassimi, A., Ciccodicola, A. (2019) Transcriptional regulation: molecules, involved mechanisms, and misregulation. *Int J Mol Sci.* **20**, 1281.

3. Kocaturk, N. M., Gozuaci, D. (2018) Crosstalk between mammalian autophagy and the ubiquitin-proteasome system. *Front. Cell Dev.* **6**, 00128.

4. Pohl, C., Dikic, I. (2019) Cellular quality control by the ubiquitin-proteasome system and autophagy. *Science.* **366**, 818-822.

5. Klionsky, D. J. (2007). Autophagy: from phenomenology to molecular understanding in less than a decade. *Nat. Rev. Mol. Cell Biol*. **8**, 931–937.

6. Lamb, C. A., Yoshimori, T., and Tooze, S. A. (2013). The autophagosome: origins unknown, biogenesis complex. *Nat. Rev. Mol. Cell Biol.* **14**, 759–774.

7. Zheng, N., Shabek, N. (2017) Ubiquitin Ligases: Structure, Function, and Regulation. *Annu. Rev. Biochem.* **86**, 129-157.

8. Bosu, D. R., Kipreos, E. T. (2008) Cullin-RING ubiquitin ligases: global regulation and activation cycles. *Cell Division. **3**,* 7.

9. Marshall, R. S., Vierstra, R. D. (2019) Dynamic regulation of the 26S proteasome: From synthesis to degradation. *Front. Mol. Biosci.* **6**, 40.

10. Zheng, N., Schulman, B. A., Song, L., Miller, J. J., Jeffrey, P. D., Wang, P., Chu, C., Koepp, D. M., Elledge, S. J., Pagano, M., Conaway, R. C., Conaway, J. W., Harper, J. W., Pavletich, N. P. (2002) Structure of the Cul-1-Rbx1-Skp1-F-boxSkp2 SCF ubiquitin ligase complex. *Nature*. **416**, 703-709.

11. Sarikas, A., Harmann, T., Pan, Z. Q. (2011) The cullin protein family. *Genome Biol.* **12**, 220.

12. Kamura, T., Koepp, D. M. , Conrad, M. N. , Skowyra, D. , Moreland, R. J., Iliopoulos, O., Lane, W. S., Kaelin, W. G., Elledge ,S. J., Conaway, R. C., Harper, J. W.,

Conaway, J. W. (1999) Rbx1, a component of the VHL tumor suppressor complex and SCF ubiquitin ligase. *Science*. **284**, 657-661.

13. Pause, A., Peterson, B., Schaffar, G., Stearman, R., Klausner, R. D. (1999) Studying interactions of four proteins in the yeast two-hybrid system:structural resemblance of the pVHL/elongin BC/hCUL-2 complex with the ubiquitin ligase complex SKP1/cullin/F-box protein. *PNAS.* **96**, 9533-9538.

14. Kamura, T., Maenaka, K., Kotoshiba, S., Matsumoto, M,. Kohda, D., Conaway, R. C., Conaway, J. W., Nakayama, K.I. (2004) VHL-box and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases. *Genes Dev.* **18**, 3055-3065.

15. Furukawa, M., He, Y. J., Borchers, C, Xiong, Y. (2003) Targeting of protein ubiquitination by BTB-Cullin 3-Roc1 ubiquitin ligases. *Nat. Cell.Biol.* **5,** 1001-1007.

16. Pintard, L., Willis, J. H., Willems, A., Johnson, J. L., Srayko, M., Kurz, T., Glaser, S., Mains, P. E., Tyers M., Bowerman, B., Peter ,M. (2003) The BTB protein MEL-26 is a substrate-specific adaptor of the CUL-3 ubiquitin-ligase. *Nature*. **425,** 311-316.

17. Angers, S., Li, T., Y,i X., MacCoss, M.J., Moon, R.T., Zheng, N. (2006) Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature* **443**, 590-593.

18. Pierce N. W., Lee J. E., Liu X., Sweredoski M. J., Graham R. L., Larimore E. A., Rome M., Zheng N., Clurman B. E., Hess S., Shan S. O., Deshaies R. J. (2013) Cand1 promotes assembly of new SCF complexes through dynamic exchange of F box proteins. *Cell*. 153, 206-215.

19. Goldenberg, S. J., Cascio, T. C., Shumway, S. D., Garbutt, K. C., Liu J., Xion, Y., Zheng, N. (2004) Structure of the Cand1-Cul1-Roc1 complex reveals regulatory mechanisms for the assembly of the multisubunit Cullin-dependent ubiquitin ligases. *Cell.* **119**, 517-528.

20. Huang, D. T., Ayrault, O., Hunt, H. W., Taherbhoy, A. M., Duda, D. M., Scott, D. C., Borg, L. A., Neale, G., Murray, P. J., Roussel, M. F., Schulman, B. A. (2009) E2-RING expansion of the NEDD8 cascade confers specificity to cullin modification. *Mol. Cell. **33**, 483-495.*

21. Siergiejuk, E., Scott, D. C., Schulman, B. A., Hofmann, K., Kurz, T., Peter, M. (2009) Cullin neddylation and substrate-adaptors counteract SCF inhibition by the CAND-like protein Lag2 in Saccharomyces cerevisiae. *EMBO J.* **28**, 3845-3856.

22. Liu, J., Nussinov, R. (2010) Rbx1 Flexible linker facilitates Cullin-RING ligase function before neddylation and after neddylation. *Biophys J.* **99**, 736-744.

23. Kurz T., Ozlu N., Rudolf F., O'Rourke S. M., Luke B., Hofmann K., Hyman A. A., Bowerman B., Peter M. (2005) The conserved protein DCN-1/Dcn1p is required for cullin neddylation in C. elegans and S. cerevisiae. *Nature.* 435, 1257–1261.

24. Lyapina, S., Cope, G., Shevchenko, A., Serino, G., Tsuge, T., Zhou, C., Wof, D. A., Wei, N., Shevchenko, A., Deshaie, R. J. (2001) Promotion of NEDD8-CUL1 Conjugate Cleavage by COP9 Signalosome. *Science.* 292, 1382-1385.

25. Stogios, P. J., Downs, G. S., Jauhal, J. J. S., Nandra. S. K., Prive, G. G. (2005) Sequence and structural analysis of BTB domain proteins. *Genome Biol.* **6**, R82.

26. Xu, X., Eletsky, A., Sheikh, M. O., Prestegard, J. H., West, C. M. (2018) Glycosylation promotes random coil to helix transition in a region of a protist Skp1 associated with F-box binding. *Biochemistry.* 57, 511-515.

27. Kachariya, N. N., Dantu, S. C., Kumar, A. (2016) Backbone and side chain assignments of human cell cycle regulatory protein S-phase kinase-associated protein 1. *Biol. NMR Assign.* **10**, 351-355.

28. Kipreos, E.T., and Pagano, M. (2000) The F-box protein family. *Genome biology*. 5, 3002.1-3002.7

29. Skaa, J. R., Pagan, J. K., Pagano, M. (2013) Mechanisms and function of substrate recruitment by F-box proteins. *Nat. Rev. Mo. Cell Biol.* **14**, 369-381.

30. Lauinger, L., Flick, K., Kaiser, P. (2021) Cdc48/Shp1 participates in dissociation of protein complex to regulate their activity. *Curr. Genet.* **67**, 263-265.

31. Barbey, R., Baudouin-Cornu, P., Lee, T. A., Rouillon, A., Zarzov, P., Tyers, M., Thomas, D. (2005) Inducible dissociation of SCFMet30 ubiquitin ligase mediates a rapid transcriptional response to cadmium. *EMBO J.* **24**, 521-532.

32. Yen, J. L. L., Flick, K., Papagiannis, C. V. V., Mathur, R., Tyrrel, A., Ouni, I., Kaake, R. M., Huang, L., Kaiser, P. (2012) Signal-induced disassembly of the SCF ubiquitin ligase complex by Cdc48/p97. *Mol. Cell*. **48**, 288-297.

33. Beltrao, P., Albanese, V., Kenner, L. R., Swaney, D. L., Burlingame, A., Villen, J., Lim, W. A., Fraser, J. S., Frydman, J., Krogan, N. J. (2012) Systematic functional prioritization of protein posttranslational modifications. *Cell.* **150**, 413-425.

34. West, C. M., and Blader, I. J. (2015) Oxygen Sensing by Protozoans: How They Catch Their Breath. *Curr. Opin. Microbiol.* 26, 41−47.

35. West, C. M., and Kim, H. W. (2019) Nucleocytoplasmic O-glycosylation in protists. *Curr. Opin. Struc. Biol.* 56, 204-212.

36. van der Wel, H., Ercan, A., West, C. M. (2005) The Skp1 prolyl hydroxylase from Dictyostelium is related to the hypoxia-inducible factor-alpha class of animal prolyl 4-hydroxylases. *J Biol Chem.* 280, 14645-14655.

37. van der Wel, H., Johnson, J. M., Xu, Y., Karunaratne, C. V., Wilson, K. D., Vohra, Y., Boons, G. J., Taylor, C. M., Bendiak, B., West, C. M. (2011) Requirements for Skp1 processing by cytosolic prolyl 4(trans)-hydroxylase and α-N-acetylglucosaminyltransferase enzymes involved in O_2 signaling in dictyostelium. *Biochemistry.* 50, 1700-1713.

38. van der Wel, H., Fisher, S. Z., West, C. M. (2002). A bifunctional diglycosyltransferase forms the Fucalpha1,2Galbeta1,3-disaccharide on Skp1 in the cytoplasm of Dictyostelium. *J Biol Chem*. 277, 46527–46534.

39. Schafer, C. M., Sheikh, M. O., Zhang, D., West, C. M. (2014) Novel Regulation of Skp1 by the *Dictyostelium* AgtA α-Galactosyltransferase Involves the Skp1-binding Activity of Its WD40 Repeat Domain. *J Biol Chem.* 289, 9076-9088.

40. Sheikh, M. O., Thieker, D., Chalmers, G., Schafer, C. M., Ishihara, M., Azadi, P., Woods, R. J., Glushka, J. N., Bendiak, B., Prestegard, J. H., and West, C. M. (2017) O2 Sensing−Associated Glycosylation Exposes the F-Box−Combining Site of the Dictyostelium Skp1 Subunit in E3 Ubiquitin Ligases. *J Biol Chem*. 292, 18897−18915.

41. Florimond, C., Coronnier, C., Taujale, R., van der Wel, H., Kannan, N., West, C. M., Blader, I. J. (2019) A *Toxoplasma* prolyl hydroxylase mediates oxygen stress responses by regulation translation elongation. *mBio.* **10**, e00234-19.

42. Rahman K., Zhao, P., Mandalasi, M., van der Wel, H., Wells, L., Blader, I. J., 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.

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

44. Mandalasi, M., Kim, H. W., Thieker, D., Sheikh, M. O., Gas-Pascual, E., Rahman, K., Zhao, P., Daniel, N. G., van der Wel, H., Ichikawa, T. H., Glushka, J. N., Wells, L., Woods, R. J., Wood, Z. A, West, C. M. (2020) A terminal α-galactose modification regulates an E3 ubiquitin ligase subunit in Toxoplasma gondii. *J Biol Chem.* 295, 9223-9243.

45. Zhang, D., van der Wel, H., Johnson, J. M., West, C. M. (2012) Skp1 prolyl 4-hydroxylase of Dictyostelium mediates glycosylation-independent and -dependent responses to O2 without affecting Skp1 stability. *J Biol Chem.* 287, 2006-2016.

46. West, C. M., van der Wel, H., Wang, Z. A. (2007) Prolyl 4-hydroxylase-1 mediates O2 signaling during development of Dictyostelium. *Development*. 134, 3349-3358.

47. Wang, Z.A., Singh, D., van der Wel, H., West, C. M. (2011) Prolyl hydroxylation- and glycosylation-dependent functions of Skp1 in O2-regulated development of Dictyostelium. *Dev Biol*. 349, 283-295.

48. Xu, Y., Brown, K. M., Wang, Z. A., van der Wel, H., Teygong, C., Zhang, D., Blader, I. J., 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.

49. Hammarlund, E. U., Flashman, E., Mohlin, S., Licausi, F. (2020) Oxygen-sensing mechanisms across eukaryotic kingdoms and their roles in complex multicellularity. *Science.* **370**, 421.

50. Xu, Y., Wang, Z. A., Green, R. S., West, C. M. (2012) Role of the Skp1 prolyl-hydroxylation/glycosylation pathway in oxygen dependent submerged development of *Dictyostelium. BMC Dev. Biol.* **12**, 31.

51. Kaelin Jr., W. G. J., Ratcliffe, P. J. (2008) Oxygen sensing by metazoans: The central role of the HIF hydroxylase pathway. *Mol. Cell.* **30**, 393–402.

52. Epstein, A. C., Gleadle, J. M., McNeill, L. A., Hewitson, K. S., O'Rourke, J., Mole, D. R., Mukherji, M., Metzen, E., Wilson, M. I., Dhanda, A., Tian, Y. M.,Masson, N., Hamilton, D. L., Jaakkola, P., Barstead, R., Hodgkin, J., Maxwell, P. H., Pugh, C.W., Schofield, C. J., Ratcliffe, P. J. (2001) C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell* **107**, 43–54

53. Hirsilä, M., Koivunen, P., Günzler, V., Kivirikko, K. I., Myllyharju, J. (2003) Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. *J. Biol. Chem.* **278**, 30772–30780.

54. Hughes, B. T., Espenshade, P. J. (2008) Oxygen-regulated degradation of fission yeast SREBP by Ofd1, a prolyl hydroxylase family member. *EMBO J.* **27**, 1491–1501.

55. Henzl, M. T., Thalmann, I., Thalman, R. (1998) OCP2 exists as a dimer in the organ of Corti. *Hearing Research.* **126**, 37-46.

56. Kim, H. W., Eletsky, A., Gonzalez, K. J., van der Wel, H., Strauch, E. M., Prestegard, J. H., West, C. M. (2020) Skp1 Dimerization Conceals its F-box Protein Binding Site. *Biochemistry.* **59**, 1527-1536.

57. Sheikh, M. O., Schafer, C. M., Powell, J. T., Rodgers, K. K., Mooers, B. H. M., West, C. M. (2014) Glycosylation of Skp1 Affects its Conformation and Promotes Binding to a Model F-box Protein. *Biochemistry.* **53**, 1657-1669.

58. Sheikh, M. O., Xu, Y., van der Wel, H., Walden, P., Hartson, S. D., West, C. M. (2015) Glycosylation of Skp1 Promotes Formation of Skp1-Cullin-1-F-box Protein Complexes in *Dictyostelium*. *MCP*. **14**, 66-80.

59. Zhao, G., Guan, W., Cai, L., Wang, P. G. (2010) Enzymatic route to preparative-scale synthesis of UDP-GlcNAc/GalNAc, their analogues and GDP-fucose. *Nat. Protoc.* **5**, 636-646.

60. Maciejewski, M. W., Schuyler, A. D., Gryk, M. R., Moraru, I. I., Romero, P. R., Ulrich, E. L., Eghbalnia, H. R., Livny, M., Delaglio, F., and Hoch, J. C. (2017) NMRbox: A Resource for Biomolecular NMR Computation, *Biophys J,* **112,** 1529-1534.

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

62. Chaikuad, A., Froese, D. S., Berridge, G., von Delft, F., Oppermann, U., Yue, W. W. (2011) Conformational plasticity of glycogen and its maltosaccharide substrate during glycogen biogenesis. *PNAS.* **108**, 21028-21033.

CHAPTER 2

1. Islam, M. S., Leissing, T. M. , Chowdhury, R., Hopkinson, R. J., and Schofield, C. J. (2018) 2-oxoglutarate-dependent oxygenases. *Annu. Rev. Biochem.* **87**, 585-620

2. Pugh, C. W., and Ratcliffe, P. J. (2017) New horizons in hypoxia signaling pathways. *Exp. Cell Res.* **356**, 116-121

3. West, C. M., and Blader, I. J. (2015) Oxygen sensing by protozoans: how they catch their breath. *Curr. Opin. Microbiol.* **26**, 41-47

4. Willems, A. R., Schwab, M., and Tyers, M. (2004) A hitchhiker's guide to the cullin ubiquitin ligases: SCF and its kin. *Biochim. Biophys. Acta* **1695**, 133–170

5. West, C. M., van der Wel, H., and Wang, Z. A. (2007) Prolyl 4-hydroxylase-1 mediates O_2 signaling during development of *Dictyostelium*. *Development* **134**, 3349–3358

6. West, C. M., and Kim, H. W. (2019) Nucleocytoplasmic O-glycosylation in protists. *Curr. Opin. Struct. Biol.* **56**, 204-212

7. Sheikh, M. O., Thieker, D., Chalmers, G., Schafer, C. M., Ishihara, M., Azadi, P., Woods, R. J., Glushka, J. N., Bendiak, B., Prestegard, J. H., and West, C. M. (2017) O_2 sensing-associated glycosylation exposes the F-box-combining site of the *Dictyostelium* Skp1 subunit in E3 ubiquitin ligases. *J. Biol. Chem.* **292**, 18897-18915

8. Wang, Z. A., Singh, D., van der Wel, H., and West, C. M. (2011) Prolyl hydroxylation- and glycosylation-dependent functions of Skp1 in O_2-regulated development of *Dictyostelium*. *Dev. Biol.* **349**, 283-295

9. Sheikh, M. O., Schafer, C. M., Powell, J. T., Rodgers, K. K., Mooers, B. H., and West, C. M. (2014) Glycosylation of Skp1 affects its conformation and promotes binding to a model F-box protein. *Biochemistry* **53**, 1657-1669

10. Sheikh, M. O., Xu, Y., van der Wel, H., Walden, P., Hartson, S. D., and West, C. M. (2015) Glycosylation of Skp1 promotes formation of Skp1/cullin-1/F-box protein complexes in *Dictyostelium*. *Mol. Cell. Proteomics* **14**, 66-80

11. Halonen, S. K., and Weiss, L. M. (2013) Toxoplasmosis. *Handb. Clin. Neurol.* **114**, 125-145

12. Xu, Y., Brown, K. M., Wang, Z. A., van der Wel, H., Teygong, C., Zhang, D., Blader, I. J., and West, C. M. (2012) The Skp1 protein from *Toxoplasma* is modified by a cytoplasmic prolyl 4-hydroxylase associated with oxygen sensing in the social amoeba *Dictyostelium*. *J. Biol. Chem.* **287**, 25098-25110

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

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

15. Lévesque, C. A., Brouwer, H., Cano, L., Hamilton, J. P., Holt, C., Huitema, E., Raffaele, S., Robideau, G. P., Thines, M., Win, J., Zerillo, M. M., Beakes, G. W., Boore, J. L., Busam, D., Dumas, B., Ferriera, S., Fuerstenberg, S. I., Gachon, C. M., Gaulin, E., Govers, F., Grenville-Briggs, L., Horner, N., Hostetler, J., Jiang, R. H., Johnson, J., Krajaejun, T., Lin, H., Meijer, H. J., Moore, B., Morris, P., Phuntmart, V., Puiu, D., Shetty, J., Stajich, J. E., Tripathy, S., Wawra, S., van West, P., Whitty, B. R., Coutinho, P. M., Henrissat, B., Martin, F., Thomas, P. D., Tyler, B. M., De Vries, R. P., Kamoun, S., Yandell, M., Tisserat, N., and Buell, C. R. (2010) Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biol.* **11**, 1-22

16. Kamoun, S., Furzer, O., Jones, J. D., Judelson, H. S., Ali, G. S., Dalio, R. J., Roy, S. G., Schena, L., Zambounis, A., Panabières, F., Cahill, D., Ruocco, M., Figueiredo, A., Chen, X. R., Hulvey, J., Stam, R., Lamour, K., Gijzen, M., Tyler, B. M., Grünwald, N. J., Mukhtar, M. S., Tomé, D. F., Tör, M., Van Den Ackerveken, G., McDowell, J., Daayf, F., Fry, W. E., Lindqvist-Kreuze, H., Meijer, H. J., Petre, B., Ristaino, J., Yoshida, K., Birch, P. R., and Govers, F. (2015) The Top 10 oomycete pathogens in molecular plant pathology. *Mol. Plant Pathol.* **16**, 413-434

17. Gaastra, W., Lipman, L. J., De Cock, A. W., Exel, T. K., Pegge, R. B., Scheurwater, J., Vilela, R., and Mendoza, L. (2010) *Pythium insidiosum*: an overview. *Vet. Microbiol.* **146**, 1-16

18. van der Wel, H., Gas-Pascual, E., and West, C. M. (2019) Skp1 isoforms are differentially modified by a dual function prolyl 4-hydroxylase/N-acetylglucosaminyltransferase in a plant pathogen. *Glycobiology* **29**, 705-714

19. Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J. J., Schnaar, R. L., Freeze, H. H., Marth, J. D., Bertozzi, C. R., Etzler, M. E., Frank, M., Vliegenthart, J. F., Lutteke, T., Perez, S., Bolton, E., Rudd, P., Paulson, J., Kanehisa, M., Toukach, P., Aoki-Kinoshita, K. F., Dell, A., Narimatsu, H., York, W., Taniguchi, N., and Kornfeld, S. (2015) Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25**, 1323-1324

20. Gas-Pascual, E., Ichikawa, H. T., Sheikh, M. O., Serji, M. I., Deng, B., Mandalasi, M., Bandini, G., Samuelson, J., Wells, L., and West, C. M. (2019) CRISPR/Cas9 and glycomics tools for *Toxoplasma* glycobiology. *J. Biol. Chem.* **294**, 1104-1125

21. Sabin, A. B. (1941) Toxoplasmic encephalitis in children. *J. Am. Med. Assoc.* **116**, 801–807

22. Fox, B. A., Ristuccia, J. G., Gigley, J. P., and Bzik, D. J. (2009) Efficient gene replacements in *Toxoplasma gondi*i strains deficient for nonhomologous end joining. *Eukaryot. Cell* **8**, 520-529

23. Grainger, J. R., Wohlfert, E. A., Fuss, I. J., Bouladoux, N., Askenase, M. H., Legrand, F., Koo, L. Y., Brenchley, J. M., Fraser, I. D., and Belkaid, Y. (2013) Inflammatory monocytes regulate pathologic responses to commensals during acute gastrointestinal infection. *Nat. Med.* **19**, 713-721

24. Blader, I. J., Coleman, B. I., Chen, C. T., and Gubbels, M. J. (2015) Lytic cycle of *Toxoplasma gondii*: 15 years later. *Annu. Rev. Microbiol.* **69**, 463-485

25. Bisio, H., and Soldati-Favre, D. (2019) Signaling cascades governing entry into and exit from host cells by *Toxoplasma gondii*. *Annu. Rev. Microbiol.* **73**, 579-599

26. Ball, S., Colleoni, C., Cenci, U., Raj, J. N., and Tirtiaux, C. (2011) The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. *J. Exp. Bot.* **62**, 1775-1801

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

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

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

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

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

32. Slabinski, L., Jaroszewski, L., Rychlewski, L., Wilson, I. A., Lesley S. A., Godzik, A. (2007) XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **23**, 3403-3405

33. Lomako, J., Lomako, W. M., and Whelan, W. J. (1988) A self-glucosylating protein is the primer for rabbit muscle glycogen biosynthesis. *FASEB J.* **2**, 3097-3103

34. de Paula, R. M., Wilson, W. A., Roach, P. J., Terenzi, H. F., and Bertolini, M. C. (2005) Biochemical characterization of *Neurospora crassa* glycogenin (GNN), the self-glucosylating initiator of glycogen synthesis. *FEBS Lett.* **579**, 2208-2214

35. Cao, Y., Steinrauf, L. K., and Roach, P. J. (1995) Mechanism of glycogenin self-glucosylation. *Arch. Biochem. Biophys.* **319**, 293-298

36. Alonso, M. D., Lomako, J., Lomako, W. M., Whelan, W. J., and Preiss, J. (1994) Properties of carbohydrate-free recombinant glycogenin expressed in an *Escherichia coli* mutant lacking UDP-glucose pyrophosphorylase activity. *FEBS Lett.* **352**, 222-226

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

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

39. Moremen, K. W., and Haltiwanger, R. S. (2019) Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. *Nat. Chem. Biol.* **15**, 853-864

40. Gibbons, B. J., Roach, P. J., and Hurley, T. D. (2002) Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *J. Mol. Biol.* **319**, 463-477

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

42. Bazan, S., Issoglio, F. M., Carrizo, M. E., and Curtino, J. A. (2008) The intramolecular autoglucosylation of monomeric glycogenin. *Biochem. Biophys. Res. Comm.* **371**, 328-332

43. Bourne, Y., and Henrissat, B. (2001) Glycoside hydrolases and glycosyltransferases: families and functional modules. *Curr. Opin. Struct. Biol.* **11**, 593-600

44. Bilyard, M. K., Bailey, H. J., Raich, L., Gafitescu, M. A., Machida, T., Iglésias-Fernández, J., Lee, S. S., Spicer, C. D., Rovira, C., Yue, W. W., and Davis, B. G. (2018) Palladium-mediated enzyme activation suggests multiphase initiation of glycogenesis. *Nature* **563**, 235-240

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

46. Burki, F., Roger, A. J., Brown, M. W., and Simpson, A. G. B. (2020) The new tree of eukaryotes. *Trends Ecol. Evol.* **35**, 43-55

47. Brown, M. W., Heiss, A. A., Kamikawa, R., Inagaki, Y., Yabuki, A., Tice, A. K., Shiratori, T., Ishida, K. I., Hashimoto, T., Simpson, A. G. B., and Roger, A. J. (2018) Phylogenomics places orphan protistan lineages in a novel eukaryotic super-group. *Genome Biol. Evol.* **10**, 427-433

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

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

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

51. Shen, B., Brown, K., Long, S., and Sibley, L. D. (2017) Development of CRISPR/Cas9 for efficient genome editing in *Toxoplasma gondii*. *Methods Mol. Biol.* **1498**, 79-103

52. Stasic, A. J., Chasen, N. M., Dykes, E. J., Vella, S. A., Asady, B., Starai, V. J., and Moreno, S. N. J. (2019) The *Toxoplasma* vacuolar H^+-ATPase regulates intracellular pH and impacts the maturation of essential secretory proteins. *Cell Reports* **27**, 2132-2146

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

54. Zhang, Y. W., Halonen, S. K., Ma, Y. F., Wittner, M., and Weiss, L. M. (2001) Initial characterization of CST1, a *Toxoplasma gondii* cyst wall glycoprotein. *Infect. Immun.* **69**, 501-507

55. Dubey, J. P., Lindsay, D. S., and Speer, C. A. (1998) Structures of *Toxoplasma gondii* tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts. *Clin. Microbiol. Rev.* **11**, 267–299

56. Sheikh, M. O., Halmo, S. M., Patel, S., Middleton, D., Takeuchi, H., Schafer, C. M., West, C. M., Haltiwanger, R. S., Avci, F. Y., Moremen, K. W., and Wells, L. (2017) Rapid screening of sugar-nucleotide donor specificities of putative glycosyltransferases. *Glycobiology* **27**, 206-212

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

58. Kabsch, W. (2010) XDS. *Acta Cryst, D* **66**, 125-132

59. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D* **66**, 213-221

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

61. Touw, W. G., Baakman, C., Black, J., te Beek, T. A. H., Krieger, E., Joosten, R. P., and Vriend, G. (2015) A series of PDB related databases for everyday needs. *Nuc. Ac. Res.* **43**, D364-D368

62. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637

63. Woods, R. J. (2014) GLYCAM Web, Complex Carbohydrate Research Center, University of Georgia, Athens, GA

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

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

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

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

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

69. Brautigam, C. A. (2015) Calculations and publication-quality illustrations for analytical ultracentrifugation data. *Meths. Enzymol.* **562**, 109-133

70. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195-201

71. Hao, B., Zheng, N., Schulman, B. A., Wu, G., Miller, J. J., Pagano, M., and Pavletich, N. P. (2005) Structural basis of the Cks1-dependent recognition of p27Kip1 by the SCFSkp2 ubiquitin ligase. *Molec. Cell* **20**, 9-19

72. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612

73. Götz, A.W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **8**, 1542–1555

74. Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr, Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688

75. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C.R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **29**, 622–655

76. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935

77. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N· log (N)

method for Ewald sums in large systems. *J. Chem. Physics* **98**, 10089–10092

78. Park, S., Radmer, R. J., Klein, T. E., and Pande, V. S. (2005) A new set of molecular mechanics parameters for hydroxyproline and its use in molecular dynamics simulations of collagen-like peptides. *J. Comput. Chem.* **26**, 1612–1616

79. Owens, N. W., Braun, C., O'Neil, J. D., Marat, K., and Schweizer, F. (2007) Effects of glycosylation of (2 S, 4 R)-4-hydroxyproline on the conformation, kinetics, and thermodynamics of prolyl amide isomerization. *J. Am. Chem. Soc.* **129**, 11670–11671

80. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Molecular Graphics* **14**, 33-38

81. Thieker, D. F., Hadden, J. A., Schulten, K., and Woods, R. J. (2016) 3D implementation of the symbol nomenclature for graphical representation of glycans. *Glycobiology* **26**, 786-787

82. Roe, D. R., and Cheatham, III, T. E. (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084-3095

83. Miller, III, B. R., McGee, Jr., T. D., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012) MMPBSA.py: an efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* **8**, 3314-3321

84. Le, S. Q., and Gascuel, O. (2008) An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307-1320

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

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

87. Laskowski, R. A., and Swindells, M. B. (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model* **51**, 2778-2786

88. West, C. M., van der Wel, H., and Gaucher, E. A. (2002) Complex glycosylation of Skp1 in *Dictyostelium*: implications for the modification of other eukaryotic cytoplasmic and nuclear proteins. *Glycobiology* **12**, 17R−27R

CHAPTER 3

1. West, C. M., Blader, I. J. (2015) Oxygen Sensing by Protozoans: How They Catch Their Breath. *Curr. Opin. Microbiol*. **26**, 41–47.

2. van der Wel, H.,Gas-Pascual, E., West, C. M. (2019) Skp1 Isoforms are Differentially Modified by a Dual Function Prolyl 4-Hydroxylase/N-Acetylglucosaminyltransferase in a Plant Pathogen. *Glycobiology* **29**, 705-714.

3. Mandalasi, M., Kim, H. W., Thieker, D., Sheikh, M. O., Gas-Pascual, E., Rahman, K., Zhao, P., Daniel, N., van der Wel, H., Ichikawa, T. H., Glushka, J. N., Wells, L.,Woods, R. J., Wood, Z. A., West, C. M. (2020) A Glycogenin Homolog Controls *Toxoplasma gondii* Growth Via Glycosylation of an E3 Ubiquitin Ligase. *bioRxiv*.

4. 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., West, C. M. (2017) O_2 Sensing–Associated Glycosylation Exposes the F-Box–Combining Site of the *Dictyostelium* Skp1 Subunit in E3 Ubiquitin Ligases. *J. Biol. Chem.* **292**, 18897-18915.

5. Henzl, M. T., Thalmann, I., Thalmann, R. (1998) OCP2 Exists as a Dimer in the Organ of Corti. *Hear. Res*. **126**, 37-46.

6. Sheikh, M. O., Schafer, C. M., Powell, J. T., Rodgers, K. K., Mooers, B. H. M., West, C. M. (2014) Glycosylation of Skp1 Affects its Conformation and Promotes Binding to a Model F-box Protein. *Biochemistry.***53**, 1657-1669.

7. 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., Pavietich, N. P. (2000) Insights into SCF Ubiquitin Ligases from the Structure of the Skp1-Skp2 Complex. *Nature*. **408**, 381–386.

8. Zimmerman, E. S., Schulman, B. A., Zheng, N. (2010) Structural Assembly of Cullin-RING Ubiquitin Ligase Complexes. *Curr. Opin. Struct. Biol*. **20**, 714–721.

9. Xu, X., Eletsky, A., Sheikh, M. O., Prestegard, J. H., West, C. M. (2018) Glycosylation Promotes the Random Coil to Helix Transition in a Region of a Protist Skp1 Associated with F-box Binding. *Biochemistry* **57**, 511-515.

10. West, C. M., Kozarov, E., Teng-umnuay, P. (1997) The Cytosolic Glycoprotein FP21 of *Dictyostelium discoideum* is Encoded by Two Genes Resulting in a Polymorphism at a Single Amino Acid Position. *Gen*e **200**, 1-10.
11. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., Bairoch, A. (2005) Protein Identification and Analysis Tools on the Expasy Server. In: Walker JM (ed): The Proteomics Protocols Handbook, Humana Press. pp. 571-607.

12. Laue, T. M., Shah, B. D., Ridgeway, T. M., Pelletier, S. L. (1992) Computer-Aided Interpretation of Analytical Sedimentation Data for Proteins. In: Harding SE, Rowe AJ, Horton J, editors. Analytical Ultracentrifugation in Biochemistry and Polymer Science. Cambridge: The Royal Society of Chemistry. pp. 90–125.

13. Schuck, P. (2000) Size Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys. J.* **78**, 1606-1619.

14. Ortega, A., Amorós, D., Garcia de la Torre, J. (2011) Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophys. J.* **101**, 892-898.

15. Kim, D. E., Chivian, D., Baker, D. (2004) Protein Structure Prediction and Analysis Using the Robetta Server. *Nucleic Acid Res*. **32**(Web Server issue), W526-531.

16. Brautigam, C. A. (2015) Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. *Meth. Enzymol*. **562**,109-133.

17. Schuck, P. (2003) On the Analysis of Protein Self-Association by Sedimentation Velocity Analytical Ultracentrifugation. *Anal. Biochem*. 320, 104-124.

18. Delaglio, F., Grzesiek, S., Vuistere, G. W., Zhu, G., Pfeifer, J., Bax, A. (1995) NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* 6, 277-293.

19. Keller, R. (2004) The Computer Aided Resonance Assignment Tutorial, CANTINA Verlag, Goldau.

20. Güntert, P., Buchner, L. (2015) Combined Automated NOE Assignment and Structure Calculation with CYANA. *J. Biomol. NMR* 62, 453-471.

21. Shen, Y., Bax, A. (2013) Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks. *J. Biomol. NMR* 56, 227-241.

22. Brunger, A. T., Adams, P. D., Clore, G. M., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., Warren, G. L. (1998) Crystallography & NMR system (CNS), A New Software Suite for Macromolecular Structure Determination. *Acta Cryst*. 54, 905-921.

23. García de la Torre, J., Huertas, M. L., Carrasco, B. (2000) HYDRONMR: Prediction of NMR Relaxation of Globular Proteins from Atomic-Level Structures and Hydrodynamic Calculations. *J. Magnet. Resonance* 147, 138-146.

24. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., Ferrin, T. E. (2004) UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem*. 25, 1605-1612.

25. Tyka, M. D., Keedy, D. A., Andre, I., Dimaio, F., Song, Y., Richardson, D. C., Richardson, J. S., Baker, D. (2011) Alternate States of Proteins Revealed by Detailed Energy Landscape Mapping. *J. Mol. Biol*. 405, 607-618.

26. Das, R., Baker, D. (2008) Macromolecular Modeling with Rosetta. *Annu. Rev. Biochem*. 77, 363-382.

27. Chen, V. B., Arendall, W. B. 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., Richardson, D. C. (2010) MolProbity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr*. Sec D *Biol. Crystallogr*. **66**, 12-21.

28. Barlow, K. A., Ó Conchuir, S., Thompson, S., Suresh, P., Lucas, J. E., Heinonen, M., Kortemme, T. (2018) Flex ddG: Rosetta Ensemble-Based Estimation of Changes in Protein-Protein Binding Affinity upon Mutation. *J. Phys. Chem*. **122**, 5389-5399.

29. Fleishman, S. J., Leaver-Fay, A., Corn, J. E., Strauch, E. M., Khare, S. D., Koga, N., Ashworth, J., Murphy, P., Richter, F., Lemmon, G., Meiler, J., Baker, D. (2011) RosettaScripts: a Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. *PLoS One* **6**, e20161 (10 pages).

30. Smith, C. A., Kortemme, T. (2008) Backrub-Like Backbone Simulation Recapitulates Natural Protein Conformational Variability and Improves Mutant Side-Chain Prediction. *J. Mol. Biol*. **380**, 742-756.

31. Park, H., Bradley, P., Greisen, P., Liu, Y., Mulligan, V. K., Kim, D. E., Baker, D., DiMaio, F. (2016) Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theory Comput*. **12**, 6201-6212.

32. Nivon, L. G., Moretti, R., Baker, D. (2013) A Pareto-Optimal Refinement Method for Protein Design Scaffolds. *PLoS One* **8**, e59004 (5 pages).

33. Bhattacharya, A., Tejero, R., Montelione, G. T. (2007) Evaluating Protein Structures Determined by Structural Genomics Consortia. *Proteins* **50**, 778-795.

34. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein R., Thornton, J. M. (1996) AQUA and PROCHECK-NMR: Programs for Checking the Quality of Protein Structures Solved by NMR. *J. Biomol. NMR* **8,** 477-486.

35. Lovell, S. C., Davis, I. W., Arendall, W. B. III, de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., Richardson, D. C. (2003) Structure Validation by C-alpha

Geometry: phi,psi and Cbeta deviation. *Protei*ns **50**, 437–450.

36. Sippl, M. J. (1993) Recognition of Errors in Three-Dimensional Structures of Proteins. *Proteins* **17**, 355-362.

37. Krissinel, E. (2015) Stock-Based Detection of Protein Oligomeric States in jsPISA. *Nuc. Ac.* Res. **43**, W314-319.

38. Reitsma, J. M., Liu,, Reichermeier, K. M., Moradian, A., Sweredoski, M. J., Hess, S., Deshaies, R. J. (2017) Composition and Regulation of the Cellular Repertoire of SCF Ubiquitin Ligases. *Cell* **171**, 1326-1339.

39. Kachariy, N. N., Dantu, S. C., Kumar, A. (2016) Backbone and Side Chain Assignments of Human Cell Cycle Regulatory Protein S-Phase Kinase-Associated Protein 1. *Biomol. NMR Assign*. **10**, 351-355.

40. Bai, C., Sen, P., Hofmann, K., Ma, L., Goebl, M., Harper, J. W., Elledge, S. J. (1996) SKP1 Connects Cell Cycle Regulators to the Ubiquitin Proteolysis Machinery Through a Novel Motif, the F-box. *Cell* **86**, 263-274.

41. Connelly, C., Hieter, P. (1996) Budding Yeast SKP1 Encodes an Evolutionarily Conserved Kinetochore Protein Required for Cell Cycle Progression. *Cell* **86**, 275-285.

42. Beltrao, P., Albanèse, V., Kenner, L. R., Swaney, D. L., Burlingame, A., Villén, J., Lim, W. A., Fraser, J. S., Frydman, J., Krogan, N. J. (2012) Systematic Functional Prioritization of Protein Posttranslational Modifications. *Cell* **150**, 413-425.

43. Lehmann, A., Katayama, S., Harrison, C., Dhut, S., Kitamura, K., McDonald, N., Toda, T. (2004) Molecular Interactions of Fission Yeast Skp1 and its Role in the DNA Damage Checkpoint. *Genes Cells* **9**, 367-382.

44. Leber, V., Nans, A., Singleton, M. R. (2018) Structural Basis for Assembly of the CBF3 Kinetochore Complex. *EMBO J.* **37**, 269-281.

45. Tan, A., Tanner, J. J., Henzl, M. T. (2008) Energetics of OCP1-OCP2 Complex Formation. *Biophys. Chem*. **134**, 64-71.

46. Willhoft, O., Kerr, R., Patel, D., Zhang, W., Al-Jassar, C., Daviter, T., Millson, S. H., Thalassinos, K., Vaughan, C. K. (2017) The Crystal Structure of the Sgt1-Skp1 Complex: the Link between Hsp90 and Both SCF E3 Ubiquitin Ligases and Kinetochores. *Sci. Rep*. **7**, 41626 (13 pages).

47. West, C. M., Kim, H. W. (2019) Nucleocytoplasmic O-Glycosylation in Protists. *Curr. Opin. Struct. Biol.* **56**, 204-212.

48. Marianayagam, N. J., Sunde, M., Matthews, J. M. (2004) The Power of Two: Protein Dimerization in Biology. Tr. *Biochem. Sci*. **29**, 618-625.

49. West, C. M., van der Wel, H., Gaucher, E. A. (2002) Complex Glycosylation of Skp1 in *Dictyostelium*: Implications for the Modification of Other Eukaryotic Cytoplasmic and Nuclear Proteins. *Glycobiology* **12**, 17R-27R.

APPENDICES

Appendix A

Spindly is a nucleocytosolic O-fucosyltransferase in *Dictyostelium* and related proteins are widespread in protists and bacteria

Van der Wel H, Garcia AM Gas-Pascua E, Willis MM, **Kim HW**, Bandini G, Gaye MM, Costello CE, Samuelson J, West CM (2022) *Spindly* is a nucleocytosolic O-fucosyltransferase in *Dictyostelium* and related proteins are widespread in protists and bacteria. *Glycobiology*

Abstract
O-GlcNAcylation is a prominent modification of nuclear and cytoplasmic proteins in animals and plants, and is mediated by a single O-GlcNAc transferase (OGT). Spindly (Spy), a paralog of OGT first discovered in higher plants, has an ortholog in the apicomplexan parasite Toxoplasma gondii, and both enzymes are now recognized as O-fucosyltransferases (OFTs). Here we investigate the evolution of spy-like genes and experimentally confirm OFT activity in the social amoeba Dictyostelium – a protist that is more related to fungi and metazoa. Immunofluorescence probing with the fucose-specific Aleuria aurantia lectin (AAL) and biochemical cell fractionation combined with western blotting suggested the occurrence of nucleocytoplasmic fucosylation. The absence of reactivity in mutants deleted in spy or gmd (unable to synthesize GDP-Fuc) suggested monofucosylation mediated by Spy. Genetic ablation of the modE locus, previously predicted to encode a GDP-fucose transporter, confirmed its necessity for fucosylation in the secretory pathway but not for the nucleocytoplasmic proteins. Affinity capture of these proteins combined with mass spectrometry confirmed monofucosylation of Ser and Thr residues of several known nucleocytoplasmic proteins. As in Toxoplasma, the Spy OFT was required for optimal proliferation of Dictyostelium under laboratory conditions. These findings support a new phylogenetic analysis of OGT and OFT evolution that indicates their occurrence in the last eukaryotic common ancestor but mostly complementary presence in its eukaryotic descendants with the notable exception that both occur in red algae and plants. Their generally exclusive expression, high degree of conservation and shared monoglycosylation targets suggest overlapping roles in physiological regulation.

Appendix B

Human poly-N-acetyl-lactosamine synthase structure demonstrates a modular assembly of catalytic subsites for GT-A glycosyltransferase

Kadirvelraj R, Yang JY, **Kim HW,** Sanders JS., Moremen KW, Wood ZA (2021) Human poly-N-acetyl-lactosamine synthase structure demonstrates a modular assembly of catalytic subsites for GT-A glycosyltransferases. *J. Biol. Chem.* **296**:100110

Abstract

Poly-N-acetyl-lactosamine (poly-LacNAc) structures are composed of repeating [-Galβ(1,4)-GlcNAcβ(1,3)-]$_n$ glycan extensions. They are found on both N- and O-glycoproteins and glycolipids and play an important role in development, immune function, and human disease. The majority of mammalian poly-LacNAc is synthesized by the alternating iterative action of β1,3-N-acetylglucosaminyltransferase 2 (B3GNT2) and β1,4-galactosyltransferases. B3GNT2 is in the largest mammalian glycosyltransferase family, GT31, but little is known about the structure, substrate recognition, or catalysis by family members. Here we report the structures of human B3GNT2 in complex with UDP:Mg^{2+} and in complex with both UDP:Mg^{2+} and a glycan acceptor, lacto-N-neotetraose. The B3GNT2 structure conserves the GT-A fold and the DxD motif that coordinates a Mg^{2+} ion for binding the UDP-GlcNAc sugar donor. The acceptor complex shows interactions with only the terminal Galβ(1,4)-GlcNAcβ(1,3)- disaccharide unit, which likely explains the specificity for both N- and O-glycan acceptors. Modeling of the UDP-GlcNAc donor supports a direct displacement inverting catalytic mechanism. Comparative structural analysis indicates that nucleotide sugar donors for GT-A fold glycosyltransferases bind in similar positions and conformations without conserving interacting residues, even for enzymes that use the same donor substrate. In contrast, the B3GNT2 acceptor binding site is consistent with prior models suggesting that the evolution of acceptor specificity involves loops inserted into the stable GT-A fold. These observations support the hypothesis that GT-A fold glycosyltransferases employ coevolving donor, acceptor, and catalytic subsite modules as templates to achieve the complex diversity of glycan linkages in biological systems.

Appendix C

The nucleocytoplasmic *O*-fucosyltransferase *Spindly* affects protein expression and virulence in *Toxoplasma gondii*

Bandini G, Agop-Nersesian C, van der Wel H, Mandalasi M, **Kim HW**, West CM, Samuelson J (2021) The nucleocytoplasmic *O*-fucosyltransferase *Spindly* affects protein expression and virulence in *Toxoplasma gondii*. *J. Biol. Chem.* **296**:100039

Abstract

Once considered unusual, nucleocytoplasmic glycosylation is now recognized as a conserved feature of eukaryotes. While in animals, *O*-GlcNAc transferase (OGT) modifies thousands of intracellular proteins, the human pathogen *Toxoplasma gondii* transfers a different sugar, fucose, to proteins involved in transcription, mRNA processing, and signaling. Knockout experiments showed that *Tg*SPY, an ortholog of plant SPINDLY and paralog of host OGT, is required for nuclear *O*-fucosylation. Here we verify that *Tg*SPY is the nucleocytoplasmic *O*-fucosyltransferase (OFT) by 1) complementation with *Tg*SPY-MYC_3, 2) its functional dependence on amino acids critical for OGT activity, and 3) its ability to *O*-fucosylate itself and a model substrate and to specifically hydrolyze GDP-Fuc. While many of the endogenous proteins modified by *O*-Fuc are important for tachyzoite fitness, *O*-fucosylation by *Tg*SPY is not essential. Growth of Δ*spy* tachyzoites in fibroblasts is modestly affected, despite marked reductions in the levels of ectopically expressed proteins normally modified with *O*-fucose. Intact *Tg*SPY-MYC_3 localizes to the nucleus and cytoplasm, whereas catalytic mutants often displayed reduced abundance. Δ*spy* tachyzoites of a luciferase-expressing type II strain exhibited infection kinetics in mice similar to wild-type but increased persistence in the chronic brain phase, potentially due to an imbalance of regulatory protein levels. The modest changes in parasite fitness *in vitro* and in mice, despite profound effects on reporter protein accumulation, and the characteristic punctate localization of *O*-fucosylated proteins suggest that *Tg*SPY controls the levels of proteins to be held in reserve for response to novel stresses.

Appendix D

A redox-active switch in fructosamine-3-kinases expands the regulatory repertoire of the protein kinase super-family

Shrestha S, Katiyar S, Sans-Rodriguez CE, Kemppinen NR, **Kim HW**, Kadirvelraj R, Panagos C, Keyhaninejad N, Chopra P, Byrne DP, Boons GJ, Knapp EV, Eyers PA, Edison AS, Wood ZA, Kannan N (2020) A redox-active switch in fructosamine-3-kinases expands the regulatory repertoire of the protein kinase super-family. *Sci. Signal.* **13**:639

Abstract

Aberrant regulation of metabolic kinases by altered redox homeostasis substantially contributes to aging and various diseases, such as diabetes. We found that the catalytic activity of a conserved family of fructosamine-3-kinases (FN3Ks), which are evolutionarily related to eukaryotic protein kinases, is regulated by redox-sensitive cysteine residues in the kinase domain. The crystal structure of the FN3K homolog from *Arabidopsis thaliana* revealed that it forms an unexpected strand-exchange dimer in which the ATP-binding P-loop and adjoining β strands are swapped between two chains in the dimer. This dimeric configuration is characterized by strained interchain disulfide bonds that stabilize the P-loop in an extended conformation. Mutational analysis and solution studies confirmed that the strained disulfides function as redox "switches" to reversibly regulate the activity and dimerization of FN3K. Human FN3K, which contains an equivalent P-loop Cys, was also redox sensitive, whereas ancestral bacterial FN3K homologs, which lack a P-loop Cys, were not. Furthermore, CRISPR-mediated knockout of FN3K in human liver cancer cells altered the abundance of redox metabolites, including an increase in glutathione. We propose that redox regulation evolved in FN3K homologs in response to changing cellular redox conditions. Our findings provide insights into the origin and evolution of redox regulation in the protein kinase superfamily and may open new avenues for targeting human FN3K in diabetic complications.