Structural Insights into Mycobacteriales Galactan Biosynthesis

By
Alan Wylde Carter

B.S. Chemistry
University of Florida 2018

M.S. Biochemistry and Molecular Biology
University of Florida 2019

Submitted to the Department of Chemistry
in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY IN CHEMISTRY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2025

Authored By: _____
Alan W. Carter
Department of Chemistry
May 9th, 2025

Certified By: _____
Laura L. Kiessling
Novartis Professor of Chemistry
Thesis Supervisor

Accepted By: _____
Adam P. Willard
Professor of Chemistry
Graduate Officer
Chair, Department Committee on Graduate Students

This doctoral thesis has been examined by a committee of professors
from the Department of Chemistry as follows:

Matthew D. Shoulders _____
Thesis Chair
Class of 1942 Professor of Chemistry

Laura L. Kiessling _____
Thesis Supervisor
Novartis Professor of Chemistry

Catherine L. Drennan _____
Thesis Committee Member
John and Dorothy Wilson Professor of Chemistry

Structural Insights into Mycobacteriales Galactan Biosynthesis

By
Alan Wylde Carter

Submitted to the Department of Chemistry
on May 9th, 2025 in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Chemistry

Abstract

The order Mycobacteriales includes a number of severe human pathogens, including *Mycobacterium tuberculosis*, the causative agent of tuberculosis and a leading cause of infectious disease-related mortality worldwide. The unique cell wall structure of these bacteria is essential for their viability, and has been studied as a potential target for novel therapeutics development. A key component of the mycobacterial cell wall is the galactan, a 30-40 residue linear polysaccharide of galactofuranose (Gal*f*) with an alternating β(1,5) and β(1,6) linkage pattern, synthesized by the polymerase Galactofuranosyl Transferase 2 (GlfT2). While GlfT2 has been established as a processive polymerase with intrinsic sequence control, the mechanism underlying this activity remains unclear. In the studies presented here, we provide structural insights into *Nocardia brasiliensis* GlfT2 (*Nbr*GlfT2) using X-ray crystallography and cryo-electron microscopy. We characterize both the acceptor-bound and membrane-embedded structures of NbrGlfT2 and propose three models for its catalysis: Processive Galactan Sliding, Feedback-Regulated Sequence Control, and Membrane Curvature-Mediated Polymerization. Furthermore, we structurally characterize a previously undescribed GlfT2 paralog from *Rhodococcus equi*, which we term *Req*GlfT3. We confirm its galactofuranosyl transferase activity and identify the production of β(1,3) and β(1,5) linkages. These findings offer new insights into GlfT2 and related polymerizing glycosyltransferases, which will provide insights into enzymatic regioselectivity mechanisms and polysaccharide biosynthesis across the bacterial kingdom.

Acknowledgements

This Ph.D. would not have been possible without the help and support of so many incredible people. I would first like to acknowledge the help and support of Prof. Laura L. Kiessling. Laura has allowed me to pursue my passions and interests at every turn during my five years in the Kiessling Lab, giving me a project that I have enjoyed becoming immersed in. Her trust and belief in me these past few years has been a huge driving force in completing this thesis work.

I would like to fully acknowledge the immense love and support I've received from my family. Though my family is large, every single person has been behind me, supporting me and making sure I survived these last five years; being one of the youngest in the family has its benefits. All of my parents, Howard, Wendy, Jackie, Joerg, have all been my biggest supporters, cheering me on no matter my decisions, making sure I known that I've got undying love and support behind me at all times. My siblings, Anna, Shelena, Kristine, Bruce, and Travis, have all been an amazing support group for me as well, cheering me on but also making sure to keep me humble when I need it. The trips and get-togethers I've been able to enjoy with them have been indispensable. Also, big shout out to Brooklyn and MacDougal St. for being a get-away when I needed it.

During my undergraduate and graduate studies, I've made a number of amazing friends that have been with me through it all. At UF, I met my old roommates and great friends Uday and Alan D. Here at MIT, I've been able to meet many amazing people. My first and second year was defined by the amazing memories I made with great friends at Sciarappa St. While in lab, I became great friends with Melanie and Mike, who have continued to keep me constantly entertained, supported, and slightly inebriated throughout my Ph.D. I would also like to give a huge acknowledgement to the Griggs House, where I've had upwards of 20 people live with me during my time at MIT. Griggs has never ceased to be an amazing place to live, and I am positive I would not have been as successful or happy without the amazing crew of people I have been able to live with.

To every mentor I have had in the labs I've worked in. Prof. Matthew Merritt and Dr. Mukundan Ragavan were amazing mentors and huge supporters of me, and if I did not have them early on in my career, I may not have gotten as far as I have. Dr. Lyle McPherson helped me get started in

my Ph.D. work, and was a big help in my raising my confidence within the Kiessling Lab. Dr. Greg Dodge was absolutely instrumental in getting me as far as I have in structural biology, and has never ceased to be an amazing mentor and supporter, even after he's left MIT.

Big shout out to Flatline Charters in Plymouth for making sure my fishing in Massachusetts has been as good as it has been, and making the transition from Florida a little bit easier.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANL	Argonne National Labs
APS	Advanced Photon Source
Ara*f*	Arabinofuranose
BLAST	Basic Local Alignment Search Tool
BNL	Brookhaven National Labs
CBM	Carbohydrate-Binding Motif
CBMS	Center for BioMolecular Structure
CCP4	Collaborative Computational Project Number 4
CCRC	Complex Carbohydrate Research Center
COSY	Correlation Spectroscopy
Cryo-EM	Cryo-Electron Microscopy
CryoSPARC	Cryo-EM Single Particle Ab Initio Reconstruction and Classification
CTF	Contrast Transfer Function
DOE	Department of Energy
DPPC	Dipalmitoylphosphatidylcholine
eLBOW	electronic Ligand Builder and Optimisation Workbench
HMBC	Heteronuclear Multiple Bond Correlation
HSQC	Heteronuclear Single Quantum Coherence
Gal*f*	Galactofuranose
GlcNAc	N-Acetylglucosamine
GlcNAc-1-P	N-Acetylglucosamine-1-Phosphate
GlfT1	Galactofuranosyl Transferase 1
GlfT2	Galactofuranosyl Transferase 2
GlfT3	Galactofuranosyl Transferase 3
GSFSC	Gold Standard Fourier Shell Correlation
GT	Glycosyltransferase
iTol	Interactive Tree of Life

KEGG	Kyoto Encyclopedia of Genes and Genomes
LCMS	Liquid Chromatography Mass Spectrometry
LPS	Lipopolysaccharide
mAGP	Mycolyl-Arabinogalactan-Peptidoglycan
m/z	Mass-to-Charge Ratio
MALDI	Matrix-Assisted Laser Desorption Ionization
MD	Molecular Dynamics
MIT	Massachusetts Institute of Technology
MS	Mass Photometry
MraY	Phospho-MurNAc-Pentapeptide Translocase
Msm	*Mycobacterium smegmatis*
Mtb	*Mycobacterium tuberculosis*
Nbr	*Nocardia brasiliensis*
NIGMS	National Institute of General Medical Sciences
NSLS-II	National Synchrotron Light Source II
OPM	Orientation of Protein in Membrane
PBS	Phosphate Buffer Saline
PDB	Protein Data Bank
PGT	Phosphoglycosyltransferase
PHENIX	Python-based Hierarchical Environment for Integrated Xtallography
PP	Pyrophosphate
ppm	Parts Per Million
PPM	Position of Protein in Membrane
PNPT	Polyisoprenyl-Phosphate N-Acetylaminosugar-1-Phosphate Transferases
Req	*Rhodococcus equi*
Rha	Rhamnose
SMALP	Styrene Maleic Anhydride Liponanoparticle
TB	Tuberculosis

TFA	Trifluoroacetic Acid
TM	Transmembrane
TOCSY	Total Correlation Spectroscopy
TOF	Time of Flight
UDP	Uridine Diphosphate
UF	University of Florida
UMP	Uridine Monophosphate
UGM	UDP-Galactopyranose Mutase
UV-TPEF	Ultraviolet-Two Photon Excited Fluorescence

Chapter 1:

Structural and Functional Insights into Mycobacterial Galactan Biosynthesis

1.1 Summary

The genus *Mycobacterium* includes pathogens that represent critical global health threats, such as the causative agent of tuberculosis, *Mycobacterium tuberculosis*. Efforts to reduce the burden of these pathogens and improve treatment outcomes have in part focused on identifying novel cellular targets for developing new antibiotics. The mycobacterial cell wall, which is essential to their viability and forms their outermost boundary, is an attractive target. This structure consists of a unique arabinogalactan polysaccharide that forms the central covalent adduct between the peptidoglycan and outer cell membrane. The galactan component of the arabinogalactan is essential in mycobacteria, and disruption in its biosynthesis and length has been shown to reduce their viability.

Here, we review the biosynthetic enzymes involved in mycobacterial galactan biosynthesis, focusing on the enzyme Galactofuranosyl Transferase 2 (GlfT2), which is responsible for synthesizing the bulk of the galactan polysaccharide. GlfT2 catalyzes the polymerization of galactofuranose sugars (Gal*f*), producing a 30-40 unit-long galactan polymer. Two decades of research have revealed GlfT2 to function as a processive polymerase with intrinsic sequence control, mediating the formation of an alternative pattern of β(1,5) and β(1,6) linkages in the absence of a template. Biochemical and structural characterization of GlfT2 has informed several proposed models for GlfT2 catalysis and galactan length control: the cavity-filling model, the tethering model, and the transporter-mediated termination model.

Recently, GlfT2 orthologs have been identified outside of mycobacteria, with predicted structures that are highly homologous to GlfT2. These orthologs are found in species ranging from gram-positive to gram-negative bacteria and appear to produce diverse Gal*f* linkages. The Galf-containing polysaccharides of these organisms suggest these GlfT2 orthologs produce

diverse polysaccharides and likely mediate distinct functions from that of GlfT2. Understanding the mechanism and activity of GlfT2 from Mtb and related organisms will provide far-reaching insights into polysaccharide biosynthesis and Galf incorporation across the bacterial kingdom.

1.2 Introduction

1.2.1 Mycobacterial Pathogens

The genus *Mycobacterium* is notorious for containing several critical human pathogens. Among these, the most prominent is *Mycobacterium tuberculosis* (Mtb), the causative agent of tuberculosis (TB). Prior to the COVID-19 pandemic, TB was the leading cause of infectious disease-related deaths worldwide and is expected to regain this position in the near future.[1] In 2021 alone, TB accounted for 10.6 million new infections and 1.6 million deaths.[2] These numbers underscore the global health significance of TB and the urgent need for research to develop novel strategies to treat and eventually eradicate this devastating disease.

Beyond Mtb, related mycobacterial pathogens also significantly impact human health. For instance, *Mycobacterium leprae* causes leprosy, a chronic disease affecting the skin and nerves, while *Mycobacterium avium* is a major pathogen in immunocompromised individuals, leading to non-tuberculous lung disease.[3,4] Related genera also display pathogenic potential, as seen with *Corynebacterium diphtheriae*, the cause of diphtheria.[5] The prevalence of human pathogens within *Mycobacterium* and its close relatives highlights the importance of investigating shared biological features to uncover novel therapeutic targets and advance treatment strategies.

A key focus of TB treatment is the disruption of mycobacterial cell wall biosynthesis, a validated target for frontline antibiotics.[6] Drugs like isoniazid and ethambutol exploit this vulnerability by inhibiting the biosynthesis of mycolic acids and arabinogalactan, essential components of the mycobacterial cell wall.[6] Given the central role of the cell wall in protecting Mtb against environmental stresses and enabling its survival, understanding the molecular mechanisms

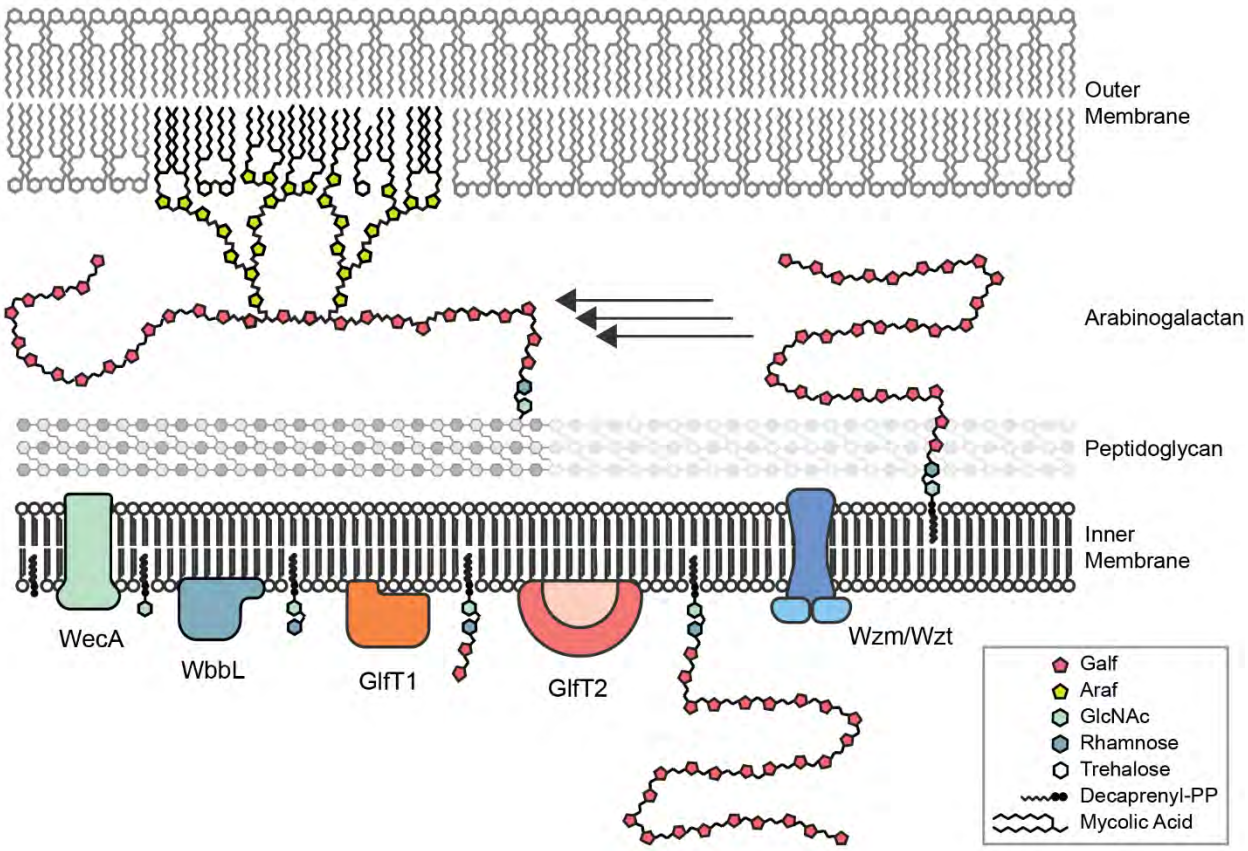

Figure 1.1: Model of Mycobacterial Galactan Biosynthesis and Incorporation. The enzymes involved in intracellular galactan biosynthesis and their individual activities are represented. Galactan transport by the Wzm-Wzt complex and its incorporation into the final arabinogalactan structure are shown. Legend: galactofuranose – red; arabinofuranose – yellow; N-acetyl-glucosamine – green; rhamnose – blue; trehalose – white.

underlying cell wall biosynthesis holds great promise for informing the development of new treatments for mycobacterial infections.

1.2.2 Mycobacterial Cell Wall Structure

Mycobacteria are gram-positive bacteria that have evolved a unique cell wall. This cell wall is characterized by a bilayer of mycolic acids—long-chain fatty acids up to 90 carbons—forming an exceptionally thick and hydrophobic membrane.[7] This structure serves as the primary barrier between the bacterium and its environment, playing a critical role in its survival and resistance to host defenses. Between the inner membrane and the mycolic acid bilayer lies the periplasmic

space, which houses the peptidoglycan layer and the arabinogalactan polysaccharide (Figure 1). The arabinogalactan is unique to mycobacteria and closely related genera, forming a covalent bridge between the peptidoglycan and the mycolic acid bilayer. Together, these components create the mycolyl-arabinogalactan-peptidoglycan (mAGP) complex, an essential structure for the viability of mycobacteria.[7]

The arabinogalactan itself consists of two distinct polysaccharides: the galactan and the arabinan (Figure 2). The galactan is a linear polymer of galactofuranose (Gal*f*) units synthesized on the cytoplasmic side of the inner membrane.[8] Initially, the galactan exists as a glycolipid precursor, decaprenylphosphate-GlcNAc-Rha-Galf_n, where *n* can range from 30 to 40 Gal*f* residues.[9] Interestingly, the galactan contains a defined sequence, displaying an alternating pattern of β(1,5) and β(1,6) linkages (Figure 2a).[10] Its biosynthesis requires the coordinated activity of four transferases and a membrane transport complex (Figure 1). Once translocated into the periplasmic space, the galactan becomes covalently linked to the peptidoglycan, forming the scaffold for arabinan attachment.[7,9]

The arabinan is a highly branched polysaccharide composed of arabinofuranose (Ara*f*) units, synthesized in the periplasmic space.[7,9] It features a linear α(1,5)-linked backbone with α(1,3)-linked branch points along the chain (Figure 2b). Arabinan polymerization is initiated by a branching modification to the galactan, transferring an Ara*f* to the 5-OH position of a β-D-Gal*f*-(1,6) unit. The arabinan terminates with a unique β(1,2) linkage, forming the β-D-Ara*f*-(1,2)-α-D-Ara*f* motif. This terminal motif forms the attachment site for mycolic acids, anchoring the lipids to the 5-OH position of both monosaccharides. These arabinan-mycolic acid displays form the periplasmic leaflet of the mycobacterial outer membrane (Figure 1).[7]

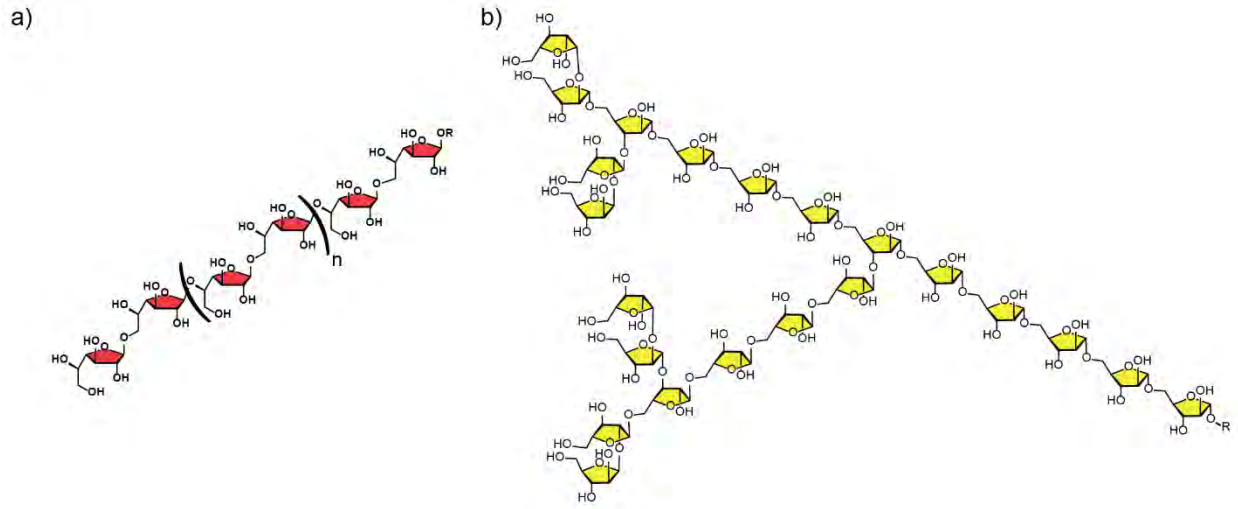

Figure 1.2: Chemical Structures of the Mycobacterial Galactan and Arabinan. a) The linear galactan polysaccharide consists of a chain of Gal*f* sugars with a repeating pattern of alternating β(1,5) and β(1,6) linkages. b) The arabinan is a branched polysaccharide of Ara*f* sugars, with a linear α(1,5) backbone, α(1,3) branching points, and a β(1,2) termination linkage.

While the structural composition of the mycobacterial cell wall is well-characterized, the processes underlying its biosynthesis remain a key area of research. This includes the synthesis of the galactan, which is integral to the formation and function of the cell wall.

1.3 Galactan Precursor Biosynthesis

The galactan is synthesized through a biosynthetic pathway involving four key transferases: WecA, WbbL, GlfT1, and GlfT2 (Figure 1).[11–16] The process begins with the phosphoglycosyltransferase activity of WecA, which transfers an N-acetylglucosamine-1-phosphate (GlcNAc-1-P) to a decaprenylphosphate acceptor.[15] This step represents the first committed step in galactan biosynthesis, producing the decaprenylpyrophosphate-GlcNAc intermediate. Next, the glycosyltransferase (GT) WbbL catalyzes the addition of a rhamnose (Rha) monosaccharide to the WecA product.[17,18] This rhamnose allows for acceptor recognition by Galactofuranosyl Transferase 1 (GlfT1), which subsequently adds the first two galactofuranose (Galf) units.[13] This di-Gal*f* moiety is believed to allow recognition by

Galactofuranosyl Transferase 2 (GlfT2), which catalyzes the full polymerization of the galactan.[11,12,19,20]

1.3.1 Insights from the Predicted Structure of WecA

WecA (Rv1302) is a phosphoglycosyltransferase (PGT) responsible for catalyzing the initial committed step of galactan biosynthesis in Mtb.[15] It transfers N-acetylglycosamine-1-phosphate (GlcNAc-1-P) from a UDP-GlcNAc donor to a decaprenylphosphate acceptor, forming UMP and decaprenyl-PP-GlcNAc as products. WecA was identified due to its homology to *E. coli* WecA, which performs an analogous GlcNAc-1-P transfer to an undecaprenylphosphate substrate in the gram-negative bacterium.[15,21]

Knockout and silencing studies on WecA in mycobacteria have found its activity to be essential to the viability of both Mtb and *Mycobacterium smegmatis* (Msm).[21–24] As members of the polyprenyl-phosphate N-acetyl hexosamine-1-phosphate transferase (PNPT) family, WecA enzymes represent paralogs of the tunicamycin-targeted phospho-MurNAc-pentapeptide translocase (MraY).[25,26] Silencing of WecA in Mtb indeed conferred tunicamycin-hypersensitivity, highlighting its potential as a therapeutic target.[15] Insights into WecA and related enzymes are likely to provide new opportunities for targeting galactan biosynthesis.

The predicted structure of Mtb WecA consists of 11 transmembrane (TM) helices, with a highly similar structure to the 10 TM helix MraY (PDB: 4J72, Figure 3a).[27–29] Both proteins appear to include an N-terminal interfacial helix that resides within the periplasmic space, while Mtb WecA contains an extra TM helix at its C-terminus. The aspartate triad essential for facilitating catalysis in MraY and other PNPT members is conserved in Mtb WecA (D126, D127, D258).[27] The tunicamycin-binding site of MraY (PDB: 5JNQ) is highly similar to the homologous site of

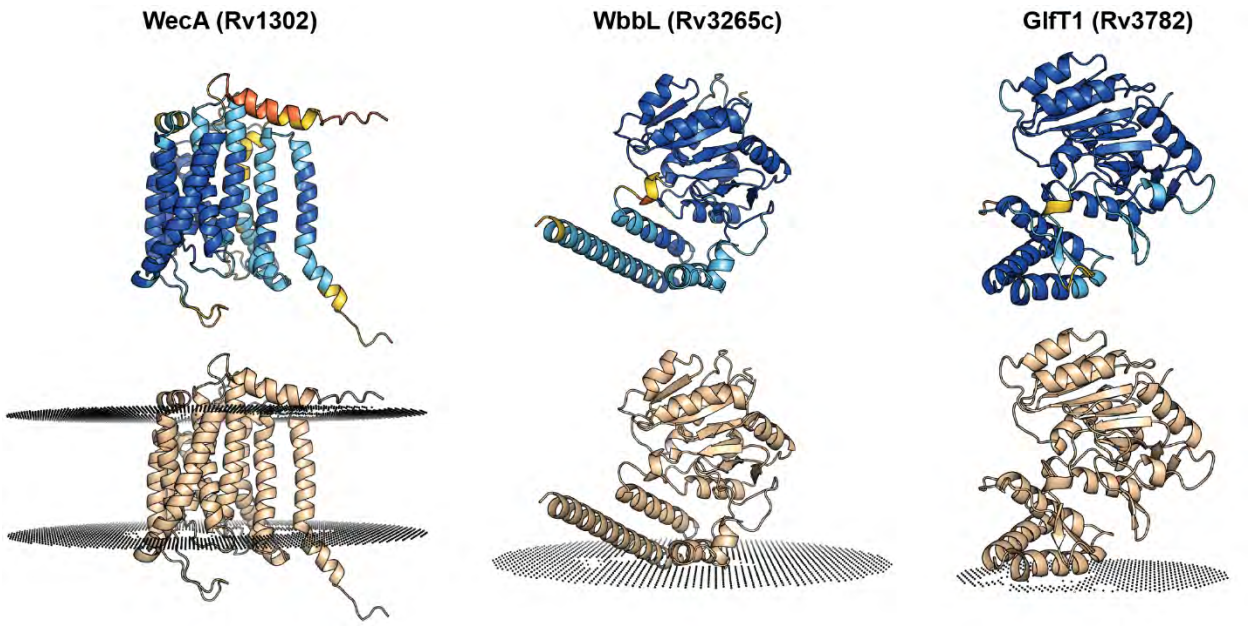

Figure 1.3: Structure Predictions of WecA, WbbL, and GlfT1. Each structure is colored by per-residue confidence, represented by their predicted local-distance difference test (pLDDT) value; blue to orange: high to low confidence. The PPM Server-predicted membrane boundaries are depicted for each enzyme, represented as small black spheres.

Mtb WecA, suggesting similar interactions and inhibition that produced tunicamycin-hypersensitivity in WecA-depleted Mtb.[15,30] These structural insights are expected to advance efforts towards novel inhibitor development targeting WecA activity.

Following the initial commitment to galactan biosynthesis by WecA, the next step involves the addition of rhamnose by WbbL, priming the substrate for subsequent galactofuranose incorporation.

1.3.2 Insights from the Predicted Structure of WbbL

WbbL (Rv3265c) is a GT responsible for adding rhamnose to the WecA product, priming the lipid-linked disaccharide for recognition by GlfT1.[17,18] WbbL catalyzes the transfer of rhamnose from a TDP-β-Rha donor, forming an α-L-Rha-(1,3) linkage onto the acceptor. Like WecA, WbbL was identified through its homology to the WbbL rhamnosyltransferase in *E. coli*.[18] WbbL

activity was shown to be essential in Mtb, and was indicated to be vulnerable to disruption.[22] Indeed, a strain of Msm with a single mutation in its WbbL led to diminished activity, introducing a temperature-sensitivity phenotype and loss of viability at elevated growth temperatures.[18]

The predicted structure of Mtb WbbL reveals a GT-2 family enzyme with a central catalytic GT-A domain and a C-terminal α-helical domain formed from a helix-turn-helix motif with two antiparallel α-helices (Figure 3b).[28,29,31] A loop predicted to contain a short helix-turn-helix is predicted to extend from β5 of the central GT-A β-sheet, in proximity to the expected acceptor binding site. The α2 of the α-helical domain and the extended loop are predicted to display hydrophobic amino acids that can potentially mediate WbbL membrane association (Figure 3b).[32] WbbL includes the conserved GT-A motifs essential for catalysis, with the DXD metal-binding motif being represented by a less common NXD motif with a noteworthy proline in the middle of the motif ($_{99}$NPD$_{101}$) and the DDX catalytic motif becoming an EDX motif instead ($_{210}$EDV$_{212}$).[31] The distance between the catalytic residue D211 and an estimated membrane boundary, calculated by the PPM-server, is nearly 18 Å, suggesting the GlcNAc-PP moiety of the acceptor must span this distance to facilitate transferase activity.[33]

Once the disaccharide precursor is formed by WbbL, GlfT1 facilitates the initial incorporation of galactofuranose residues. This final acceptor-priming step sets the stage for full polymerization by GlfT2.

1.3.3 Insights from the Predicted Structure of GlfT1

GlfT1 (Rv3782) is a mycobacterial GT that mediates the addition of the first two Gal*f* residues to the WbbL product, priming the acceptor for galactan biosynthesis by GlfT2.[13] Using UDP-Gal*f*

as a donor, GlfT1 catalyzes the addition of a β-D-Gal*f*-(1,5)-β-D-Gal*f* disaccharide to the 4-OH of rhamnose.[16] GlfT1 was shown to be essential to mycobacteria viability and highly vulnerable to interference.[18,22] Silencing of GlfT1 expression was observed to induce ethambutol susceptibility and reduce biofilm formation in Mtb.[34] Synthetic acceptor analogs of the endogenous substrate were developed to study the activity of GlfT1 *in vitro*.[16] These analogs included shortened isoprenoid lipid tails, which reestablished GlfT1 activity for the addition of 1-2 Gal*f* residues. These assays confirmed the Rha-specificity of GlfT1 and an improved *in vitro* reactivity with increased isoprenoid-lipid length.

The predicted structure of Mtb GlfT1 reveals a GT-2 family enzyme with a central catalytic GT-A domain, similar to WbbL (Figure 3c).[28,29,31] GlfT1 likewise has a C-terminal α-helical domain, formed instead by two helix-turn-helix motifs containing three antiparallel helices. GlfT1 contains the conserved GT-A DXD metal-binding motif as $_{93}$DDD$_{95}$ and the catalytic DDX motif as $_{190}$DEV$_{192}$, which interestingly suggests residue E191 facilitates catalysis.[31]

Two loops of interest are present in GlfT1, one containing an apparent β-hairpin that extends from β5 of the central GT-A β-sheet, and another β-hairpin that spans the GT-A and α-helical domains. The spanning loop, and helices α2 and α3 of the α-helical domain are predicted to display hydrophobic residues and likely mediate GlfT1 membrane association. Interestingly, the β5-extended loop is enriched for hydrophobic residues, suggesting it may also aid in membrane association as with WbbL or some unknown protein-protein interaction. Both extended loops are in proximity to the predicted catalytic glutamate E191, and potentially aid in acceptor substrate binding. The distance between the predicted catalytic residue and the membrane boundary estimated by the PPM-server is 22 Å, suggesting the Rha-GlcNAc-PP moiety spans this distance

for catalysis. The 4 Å distance increase compared to WbbL suggests the added Rha occupies this distance, and these differences may function to aid glycolipid differentiation at the membrane.

1.4 Galactan Polymerase GlfT2

GlfT2 (Rv3808c) is a mycobacterial GT that functions as a polymerase, catalyzing the bulk of galactan extension.[11,19,35] Using UDP-Gal*f* as a donor, GlfT2 recognizes the β-D-Gal*f*-(1,5)-β-D-Gal*f* disaccharide product of GlfT1 and catalyzes the stepwise addition of the remaining Gal*f* to a polysaccharide length of 30-40 units (Figure 4a).[19,35] The activity of GlfT2 is essential to the viability of mycobacteria.[22,36] Interestingly, Msm strains engineered with GlfT2 variants that produce shorter galactan products were observed to have reduced periplasmic thickness and increased antibiotic susceptibility.[37] These observations suggest that the GlfT2 galactan product may play an unrecognized role in structuring the periplasm, in addition to its role as a scaffold for the arabinan. GlfT2 activity is an attractive target for antibiotic development, and its targeting may synergize with antibiotics currently used to treat Mtb.[37]

1.4.1 The Galactofuranosyl Transferase Activity of GlfT2

Extensive work has been done to define the activity of GlfT2. Initial insights into its galactofuranosyl transferase activity were informed using simple glycolipid acceptor-surrogates that reproduced functionality in a cell-free assay with UDP-[^{14}C]-Gal.[38] Early GlfT2 acceptor surrogates were di-Gal*f*-octyl with either a β(1,5) or β(1,6) glycosidic linkage, displaying modest extension to a tetrasaccharide product. Recombinant expression and purification of *Mtb*GlfT2 was achieved soon after, enabling the isolated *in vitro* characterization of GlfT2 activity.[39] GlfT2 *in vitro* activity required the inclusion of the dication Mg^{2+} and a UDP-Galactose Mutase (UGM) to form the donor UDP-Gal*f* prior to the development of a synthetic route for the donor. Using

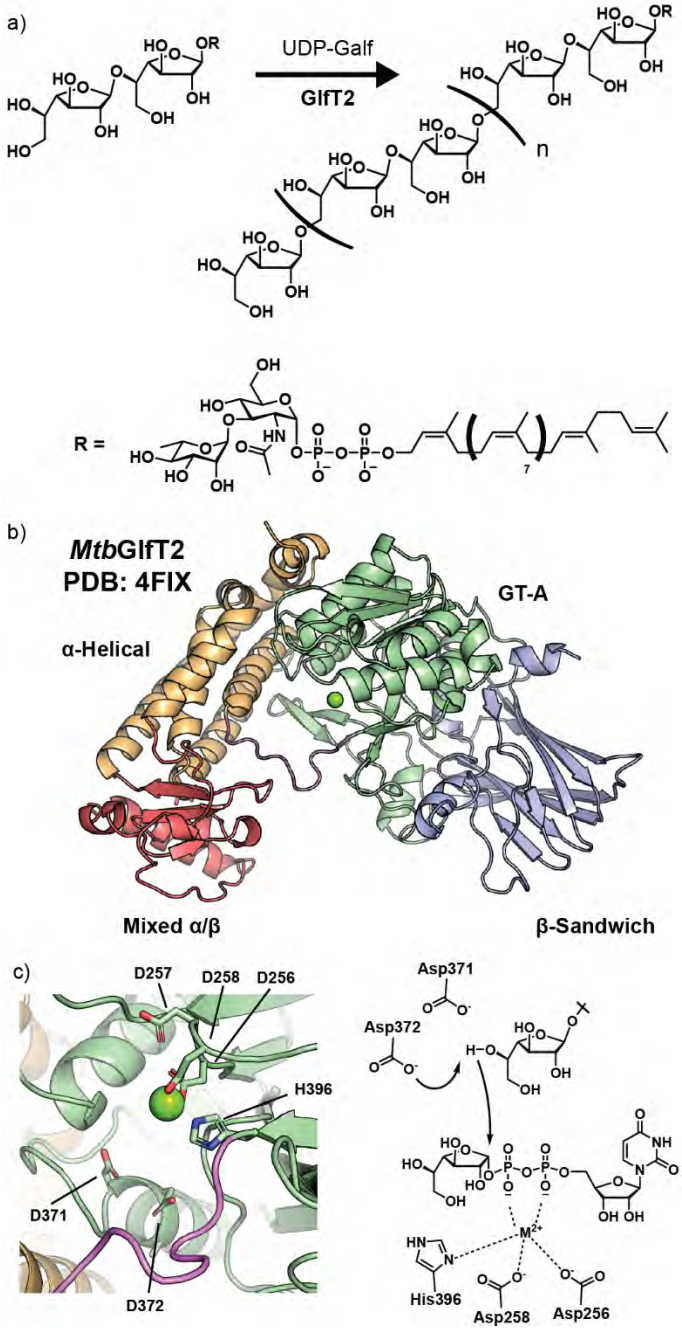

Figure 1.4: Activity and Structure of GlfT2. a) Reaction catalyzed by GlfT2, recognizing a Gal*f* disaccharide to produce the galactan polysaccharide. b) Monomer structure of *Mtb*GlfT2 (PDB: 4FIX), with domains highlighted: β-Sandwich – blue; GT-A – green; α-Helical – yellow; Mixed α/β – red. c) Structure of the essential GT-A DXD and DDX motifs. The proposed catalytic mechanism for GlfT2 is shown.

the same synthetic acceptor, product extensions could now be measured up to a hexasaccharide product in this *in vitro* system.

Although these *in vitro* results confirmed the galactofuranosyl transferase activity of GlfT2, the product lengths were significantly shorter than known endogenous galactan lengths. To evaluate whether an unidentified protein partner was required for galactan length control or if GlfT2 could have this property on its own, the *in vitro* system was optimized. Synthetic UDP-Gal*f* was used to remove UGM from the reaction mixture, isolating GlfT2 from all other proteins.[40] A stepwise optimization approach was applied to the lipid moiety of the synthetic acceptor surrogate, and larger lipids were indeed observed to increase polymerization extent, with extensions up to 48 additional Gal*f* residues.[19] These results revealed that GlfT2 could extend

biologically relevant lengths of galactan in isolation, suggesting that GlfT2 could intrinsically control galactan length.

1.4.2 Structural Insights into GlfT2 Activity

The structure of Mtb GlfT2 was determined by Xray crystallography for both the apo and UDP-bound structures (PDB: 4FIX, 4FIY).[35] The single-chain structure for *Mtb*GlfT2 reveals a GT-2 family enzyme with four domains assembled into a kidney-shaped structure (Figure 4b). A catalytic GT-A domain resides in the center of the chain. Towards the N-terminus, a β-sandwich domain is formed. The β-sandwich domain has been shown to have homology to carbohydrate-binding modules (CBM) present in some glycosylhydrolases.[35,41] The C-terminus contains a five-helix bundle of antiparallel α-helices that form an α-helical domain, very similar to the α-helical domains of WbbL and GlfT1. The sequence of the α-helical domain is split by a mixed α/β domain, in which a central three-strand β-sheet is surrounded by three α-helices. A 16-residue long helix is extended from this domain. The GT-A and α-helical domains are connected by a disordered loop, commonly found in GT-A enzymes.[31]

The central GT-A domain contains the conserved motifs necessary for catalysis (Figure 4c). The DXD metal-binding motif is represented by $_{256}$DDD$_{258}$, forming a tri-aspartate sequence similar to GlfT1. The DDX catalytic motif is represented by $_{371}$DDA$_{373}$. Single-point substitutions of the first aspartate within the DDD motif to glutamate or alanine led to severe reductions in polymerization efficiency and product lengths.[37] Substitution of either aspartate within the DDA motif led to a complete loss of catalytic activity.[35] Recombinant expression of the inefficient glutamate variant in Msm likewise showed reduced galactan lengths *in vivo*.[37] These observations reveal the GT-A domain to be the sole catalytic domain responsible for GlfT2 activity, facilitating galactan polysaccharide polymerization in mycobacteria.

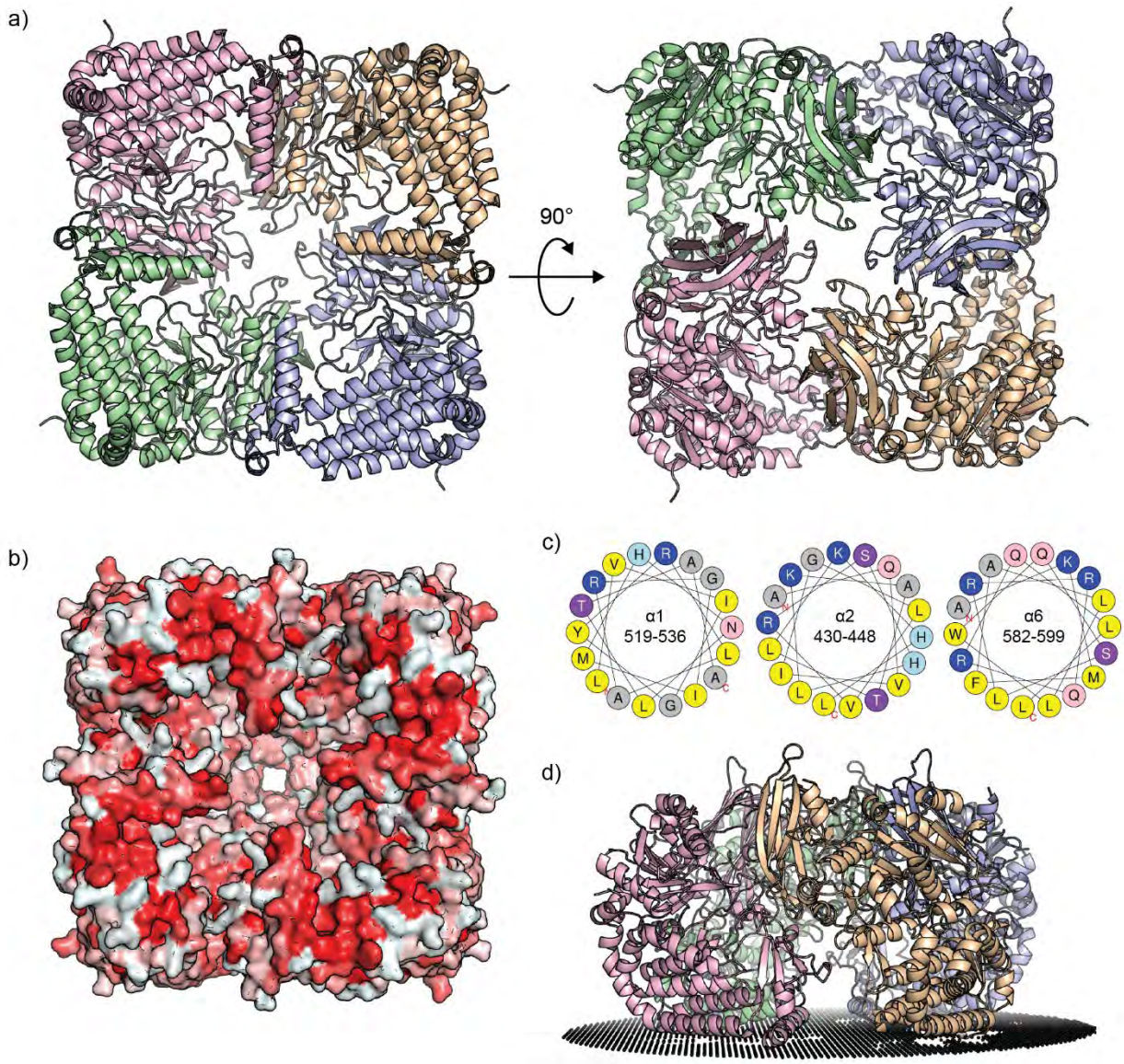

Figure 1.5: Complex Formation and Membrane Association by *Mtb*GlfT2. a) The tetrameric structure of *Mtb*GlfT2 (PDB: 4FIX). The individual protomers are differentiated by color. b) Hydrophobic amino acids are enriched on one side of the tetramer. Hydrophobic residues – red; hydrophilic residues – white. c) Helical wheels for the three amphipathic helices proposed to mediate *Mtb*GlfT2 membrane binding. d) PPM-Server predicted membrane association by the *Mtb*GlfT2 tetramer. Membrane boundary represented by small black spheres.

1.4.3 GlfT2 Oligomerization and Membrane Association

The crystal structure of *Mtb*GlfT2 revealed the formation of a homotetrameric complex (Figure 5a).[35] GlfT2 complexation is mediated by contacts between the β-sandwich and mixed α/β domains (Figure 4b). The tetramer forms with C4-symmetry, forming a large cavity in the center

of the complex. Three helices from the α-helical and mixed α/β domains of each protomer surround the opening of the cavity: α1 of the mixed α/β domain and α2 and α6 of the α-helical domain. These helices are amphipathic, with their solvent-exposed regions enriched for hydrophobic amino acids and the opposite regions enriched for basic amino acids (Figure 5b, c). These helices were hypothesized to mediate GlfT2 membrane association, localizing GlfT2 in proximity to its glycolipid substrate (Figure 5d).[35]

The estimated distance between the catalytic aspartate D372 and the predicted membrane interface is 30 Å.[33] This requires the tetrasaccharide-pyrophosphate moiety of the GlfT1 product to span this distance. This suggests the two Gal*f* residues are likely required to span the extra 8 Å distance compared to GlfT1, providing a mechanism for glycolipid substrate selectivity by GlfT2. The incremental distance increase between WbbL, GlfT1, and GlfT2 indeed suggests the use of membrane-to-active site distance as a mechanism for acceptor specificity by these GT-2 enzymes.

1.4.4 Galactan Sequence-Control by GlfT2

Linkage analysis of the galactan linkage sequence within mycobacteria revealed the presence of β(1,5) and β(1,6) linkages in an alternating pattern.[42] Production of this linkage pattern was confirmed to be reproduced by GlfT2 on the synthetic acceptor surrogates in membrane isolates and *in vitro*.[12,39] The propensity for GlfT2 to control this linkage sequence was tested using donor analogs substituted at either the 5 or 6 position of Gal*f* (Figure 6).[43] In the presence of a β(1,6) di-Gal*f* acceptor, the 5- and 6-fluoro donors only produced +2 and +1 products, respectively, indicating GlfT2 has strong fidelity for producing the alternating pattern. As this fidelity was reproduced in an *in vitro* system with GlfT2 in isolation, these observations suggest that GlfT2

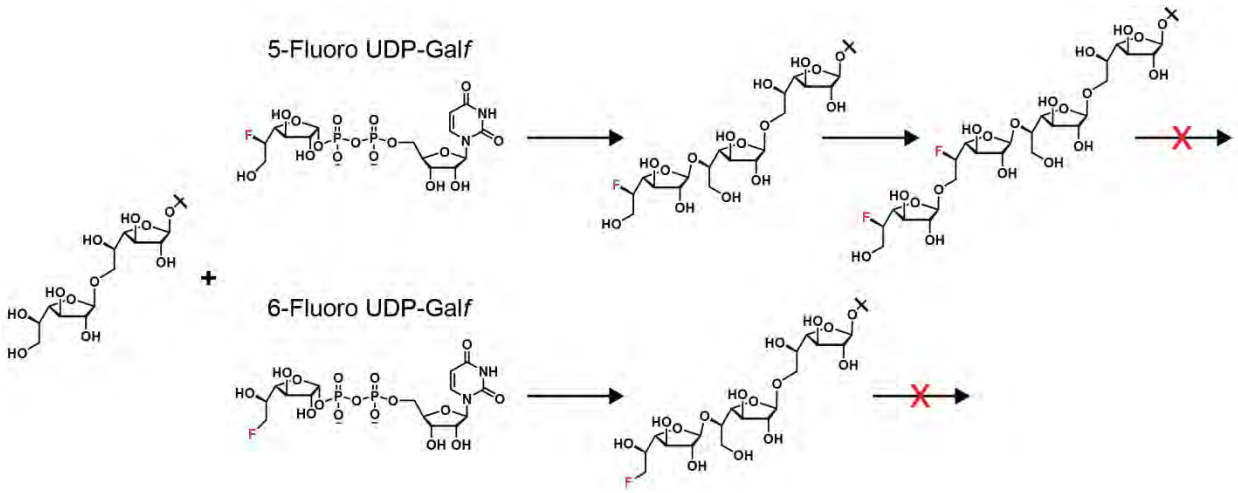

Figure 1.6: Evaluating GlfT2 Fidelity using Fluorosugar Donors. 5- and 6- fluoro UDP-Gal*f* donor analogs were used to determine the fidelity of GlfT2 for the alternating β(1,5) and β(1,6) linkage pattern. High pattern fidelity led to the production of +2 and +1 products, respectfully.

has intrinsic sequence control without needing a template or cofactor. However, the function of this linkage pattern within the periplasm remains undefined.

1.4.5 GlfT2 is a Processive Polymerase

GlfT2 functions as a polymerase, producing the galactan by stepwise Gal*f* additions to its own product. Polymerizing glycosyltransferases are often classified by two main mechanisms: distributive and processive polymerization.[44] Distributive polymerization involves a single catalytic turnover for each acceptor binding event, producing a distribution of product lengths. Processive polymerization involves higher affinity binding of the acceptor substrate, retaining the polymer after each catalytic turnover, and releasing the final, full-length product (Figure 7). A mechanism in between these two can also occur, in which a lower degree of processivity allows the addition of multiple sugars during each binding event. Determining the type of polymerization GlfT2 performs would inform its overall mechanism and provide insights into targeting its activity.

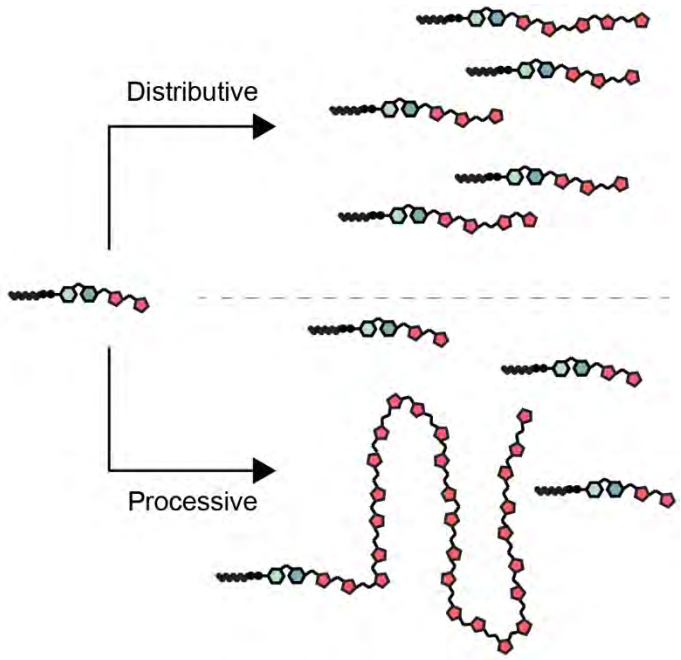

Figure 1.7: Distributive vs. Processive Glycosyltranferase Products. Representative product distributions are shown for the distributive and processive polymerization of oligo- and polysaccharides.

To determine if GlfT2 displays an apparent processivity when producing its galactan product, a competition assay between an unlabeled and an isotopically labeled acceptor was performed.[45] In this assay, the unlabeled acceptor is allowed to initiate polymerization, and the labeled acceptor is introduced to "distract" GlfT2 from the unlabeled acceptor. Indeed, enrichment for longer polymer lengths on the unlabeled acceptor was observed, suggesting GlfT2 functions processively. GlfT2 also appeared to increase its polymerization efficiency as its acceptor grew, a feature seen in other processive polymerases.[44,45]

Processive polymerization by GlfT2 is likely advantageous to galactan biosynthesis and length control. Production of the donor UDP-Gal*f* is thermodynamically disfavored, with the equilibrium favoring the pyranose-form UDP-Gal*p* with an 11:1 population ratio.[46,47] Processively polymerizing the galactan is expected to utilize the lower abundance UDP-Gal*f* to produce functional-length products more efficiently. Retention of the growing galactan also prevents the incorporation of short- and intermediate-length polymers into the mAGP. As shortened galactan was shown to reduce cell viability and increase antibiotic susceptibly in Msm,

controlling product release through this processive mechanism is likely critical to mycobacterial viability.[37]

1.4.6 Proposed Models for Galactan Length Control by GlfT2

Informed by the structure of *Mtb*GlfT2 and its activities *in vitro* and *in vivo*, a number of models for the mechanism of GlfT2 have been proposed (Figure 8a). The tetrameric complex of GlfT2 was hypothesized to form a large cavity above the membrane interface, and the acceptor-binding site for GlfT2 is present within this cavity.[35] The functional purpose of this cavity is not known. Still, it was suggested to mediate acceptor retention during galactan growth, with the growing chains filling the cavity during polymerization and the cavity's volume determining the galactan's final length. This cavity-filling model proposes that the cavity mediates processive galactan growth and length control (Figure 8a). However, the

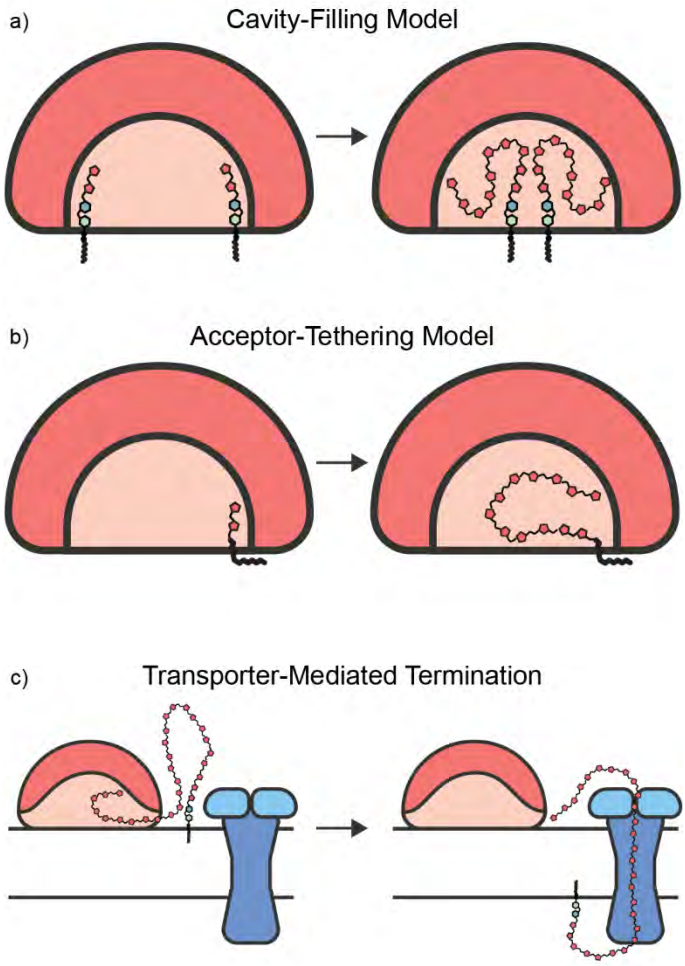

Figure 1.8: Proposed Models for GlfT2 Polymerization. a) A Cavity-Filling Model suggests the cavity formed by the GlfT2 tetramer encloses the growing galactan, and the volume of the cavity dictates product lengths. b) An Acceptor-Tethering Model proposes that the lipid of the acceptor surrogates interacts with an allosteric site on GlfT2, and facilitates retention and polymerization *in vitro*. c) The Transporter-Mediated Termination Model suggests that the transporter complex Wzm-Wzt can interact with the growing galactan outside of the GlfT2 cavity, prematurely terminating polymerization and influencing galactan length control.

model as proposed would require simulatenous initiation of polymerization by each protomeric active site, but cooperativity does not appear to be present in GlfT2's activity.[45]

A second model was proposed for GlfT2 polymerization, informed by the observation that the overall size of the synthetic acceptor lipid influences the length of the galactan product *in vitro*.[19] These data suggest that the lipid interacts directly with GlfT2 and binds to an allosteric site to influence acceptor retention and polymerization extent. The acceptor-tethering model proposed that galactan processive polymerization is driven by acceptor affinity for GlfT2, and enhanced retention of the acceptor by the lipid permits the polymerization of the biologically relevant length of galactan.

A recently proposed model for galactan polymerization at the membrane interface was informed by *in vivo* observations, and suggests the galactan transporter may interact with the growing galactan chain. The Wzm-Wzt complex (Rv3781, Rv 3783) is predicted to mediate galactan transport to the periplasm after GlfT2 polymerization (Figure 1).[48] Knockdown of Wzm, the transmembrane unit of this complex unexpectedly led to an accumulation of aberrantly long lipid-linked galactan within the cell.[14] With GlfT2 catalyzing processive polymerization, this observation suggests that the transporter complex may mediate premature termination of galactan extension, producing shorter product lengths. A transporter-mediated model was proposed in which GlfT2 and the Wzm-Wzt complex are in proximity to each other, and during polymerization, the reducing end of the growing galactan extends outside the tetrameric cavity, allowing the transporter complex to recognize the reducing end of the galactan. The transporter would then translocate the galactan to the periplasm, potentially leading to premature termination of galactan elongation and influencing galactan length control.

Although GlfT2 has been thoroughly examined in *Mycobacterium tuberculosis*, the recent identification of GlfT2-like enzymes in various other bacterial species underscores the variety of Gal*f*-utilizing pathways across the bacterial kingdom. These orthologs offer valuable insights into conserved and unique mechanisms, enhancing our comprehension of polysaccharide biosynthesis.

1.5 Insights into Galactofuranose Polymerases Outside Mycobacteria

1.5.1 Diverse Linkages Throughout the Mycobacteriales Order

The linkage pattern of the galactan in mycobacteria has been documented, and the role of GlfT2 in producing the alternating β(1,5) and β(1,6) linkage sequence is well known.[12,39,42,43] Mycobacteria are a part of the Mycobacteriales order, which contains a number of other interesting families, including Corynebacteriaceae and Nocardiaceae. In studying the galactan linkage patterns of members of Nocardiaceae, a number of alternative linkage patterns were identified in the galactan.[42] These included the apparent loss of β(1,6) linkages in *Nocardia* species and the introduction of either β(1,2) or β(1,3) linkages in *Rhodococcus*. These observations suggest that non-mycobacterial species present within the Mycobacteriales order may have different activities for their GlfT2 orthologs or incorporate novel modifications on their arabinogalactan products. Investigation of these outstanding questions may provide additional insight into the role of the galactan and its linkage pattern.

1.5.2 GlfT2-Like Enzymes Outside of Mycobacteria

A recent investigation of the T1 antigen biosynthetic pathway in *Salmonella enterica* revealed the gene *STM0724* encodes for an enzyme that produces the O-polysaccharide for its surface lipopolysaccharide (LPS). The structure of *STM0724* was predicted to fold with a three-

dimensional structure highly similar to GlfT2.[28,29,49] Indeed, *STM0724* was shown to function as a galactofuranosyl transferase, producing an oligosaccharide product with β(1,3) and β(1,6) linkages. Though the polysaccharide products were not polymerized to the lengths that *Mtb*GlfT2 produces the galactan, the high structural homology suggests *STM0724* likely functions similarly to GlfT2.

Salmonella species are gram-negative bacteria and possess a distinct cell wall structure compared to gram-positive bacteria like mycobacteria. A number of other gram-negative bacteria have been suggested to produce GlfT2 orthologs, including *E.coli* serotype O62 (ARJ35764.1) and *Kingella kingae* (CRZ20666.1, AEK99054.1), that are potentially involved in Gal*f* incorporation into their LPS structures.[49] Within the gram-positive species *Bifidobacterium bifidum*, it has been suggested that Gal*f* incorporation into a surface-exposed exopolysaccharide is facilitated by its own GlfT2 ortholog. These observations suggest the galactofuranosyl transferase activity of GlfT2 orthologs outside of mycobacteria may facilitate the production of Gal*f*-containing polysaccharides with diverse functional roles and linkage patterns.

Exploring GlfT2 orthologs underscores the versatility of galactofuranose polymerases and their roles in diverse bacterial contexts. These findings reinforce GlfT2's significance as a model enzyme for studying polysaccharide biosynthesis, which has implications for understanding bacterial survival and pathogenicity.

1.6 Concluding Remarks

The genus *Mycobacterium* contains multiple pathogens that represent critical threats to human health globally. Efforts to reduce the burden of these pathogens and improve treatment outcomes have, in part, focused on identifying novel cellular targets and elucidating their functions to

potentially inform the development of new antibiotics. The mycobacterial cell wall is a unique feature of these species, and its essential function makes it a promising candidate for disrupting the viability and fitness of these pathogens.

The mycobacterial cell wall consists of an arabinogalactan polysaccharide that forms the central covalent adduct between the peptidoglycan and outer cell membrane. The galactan component of the arabinogalactan is essential in mycobacteria, and disruption in its biosynthesis and length has been shown to reduce their viability.[22,36,37] Four biosynthetic transferases have been shown to be involved in galactan biosynthesis, each performing an essential function within mycobacteria.

GlfT2 is the final enzyme in the galactan biosynthetic pathway and is responsible for synthesizing the bulk of the galactan polysaccharide. Two decades of research have revealed GlfT2 to function as a processive polymerase, controlling the formation of an alternating pattern of β(1,5) and β(1,6) linkages.[10,19,45] Structural characterization of GlfT2 revealed that the protein forms a homotetrameric structure that associates with the membrane interface, forming a large central cavity. Several models have been proposed for GlfT2 catalysis and are likely to inform mechanistic studies into GlfT2 in the future.

GlfT2 orthologs have been identified outside of mycobacteria, and their predicted structures suggest they may function similarly to GlfT2. These orthologs range from gram-positive to gram-negative bacteria and appear to produce a number of different Gal*f* linkages. The varied Gal*f*-containing polysaccharides of these organisms suggest these GlfT2 orthologs function to produce diverse polysaccharides that likely perform functions different from the mycobacterial galactan. Understanding the mechanism and activity of GlfT2 from Mtb and related organisms will likely produce far-reaching insights into enzymatic regioselectivity mechanisms, polysaccharide biosynthesis and Gal*f* incorporation across the bacterial kingdom.

1.7 Contributions of this Thesis Research

My thesis research has focused on structural characterization and mechanistic elucidation of GlfT2 and related enzymes. This work has resulted in three chapters. **Chapter 2** describes the functional characterization of GlfT2 from *Nocardia brasiliensis* (Nbr), a bacterium from the family Nocardiaceae within the order Mycobacteriales. This organism is distantly related to mycobacteria and was previously suggested to produce a galactan of only β(1,5) linkages.[42] Biochemical studies have revealed that *Nbr*GlfT2 produces the same alternating pattern of β(1,5) and β(1,6) linkages seen in Mtb. This suggests the presence of additional *in vivo* modification to the Nbr galactan polysaccharide post-polymerization.

Structural characterization of *Nbr*GlfT2 at 3.1 Å by Xray Crystallography reveals a homotetrameric complex homologous to *Mtb*GlfT2.[35] The structures of *Nbr*GlfT2 bound to synthetic surrogates representing a tri- and tetra-Gal*f* acceptor reveal the GlfT2 acceptor binding site and the residues involved in galactan polymerization. From these structures, three sugar-binding subsites were identified. These subsites appear to mediate galactan polymer sliding during catalysis, likely facilitating processive polymerization. Characterization of both a β(1,5)- and β(1,6)-terminal acceptor reveals differential active-site poses for the non-reducing end Gal*f*. From these findings, we propose a model for the mechanism of GlfT2 polymerization, in which galactan sliding mediates processive polymerization, and the length difference between the β(1,5) and β(1,6) linkages within the acceptor site facilitates feedback-regulated sequence control to produce the alternating pattern.

Chapter 3 describes the structural characterization of a nanodisc-bound *Nbr*GlfT2 by Cryo-Electron Microscopy (Cryo-EM). This work utilized styrene-maleic anhydride liponanoparticles

(SMALP) to produce GlfT2-bound nanodiscs containing native *E. coli* lipids. The work confirmed the tetrameric assembly of *Nbr*GlfT2 at the membrane interface and GlfT2 membrane association mediated by the three hypothesized amphipathic helices.[35] The membrane interface of nanodisc-bound *Nbr*GlfT2 revealed local curvature within the tetramer cavity, with the interface curving into the acceptor-binding site and away near the exit of the tetramer. These observations informed a proposed model in which GlfT2-induced membrane curvature facilitates polymer translocation and cavity exit during elongation.

Chapter 4 describes the identification of an uncharacterized group of enzymes within the order Mycobacteriales that represent paralogs of GlfT2. These paralogs, which we term GlfT3, are not present within the genera *Mycobacterium* and *Corynebacterium* and potentially influence the sequences of galactan within species expressing these genes. Sequence analysis reveals these GlfT3 enzymes to have higher sequence similarity between each other than their GlfT2 paralogs, suggesting they may perform distinct activities.

Biochemical characterization of *Rhodococcus equi* GlfT3 confirms its galactofuranosyl transferase activity and suggests that it performs distributive polymerization. Linkage analysis of its products suggests it produces β(1,3) and β(1,5) linkages, similar to previously observed linkages in the Req galactan.[42] Structural characterization of *Req*GlfT3 to 3.3 Å reveals a monomeric structure highly similar to GlfT2, as well as similar tetrameric complexation. The *Req*GlfT3 structure reveals an acceptor-binding site distinct to GlfT2 and enrichment of solvent-exposed hydrophobic amino acids that likely mediate membrane association.

Overall, the findings described here are expected to provide important information regarding the mechanism of GlfT2 polymerization. The structural insights provided by the acceptor-bound and membrane-bound structures are likely to inform future studies into GlfT2 catalysis and the

mechanisms of other polymerizing GTs. Insights from *Req*GlfT3 are expected to inform arabinogalactan biosynthesis in other Mycobacteriales pathogens and provide insights into the mechanisms mediating GT regioselectivity and linkage production.

Chapter 2:

Mechanism of the Processive, Sequence-Controlled Polysaccharide Biosynthesis Mediated by the Glycosyltransferase GlfT2

Contributions: Conceptualization by Alan W. Carter and Laura L. Kiessling. Cloning, protein purification, and biochemical characterization was performed by Alan W. Carter, Alexander M. Justen, and Katherine I. Taylor. Protein crystallization and X-ray diffraction data collection was performed by Alan W. Carter and Greg J. Dodge. Diffraction data processing by Alan W. Carter, Greg J. Dodge, and Catherine L. Drennan. Protein structure modeling and density refinement was performed by Alan W. Carter. Molecular dynamics simulations were performed by Allison Keys. The chapter was written by A.W.C. with input from authors.

2.1 Abstract

Polymerizing glycosyltransferases (GTs) are involved in the biosynthesis of many important polysaccharides. The roles of their polysaccharide products include cell-cell communication, immune recognition and evasion, biofilm formation, energy storage, and cell wall integrity. In many of these roles, polysaccharide length and composition are critical to their function; polymerizing GTs often control these properties through processive elongation and sequence-control. Galactofuranosyl transferase 2 (GlfT2) performs both of these functions using a single active site to produce the mycobacterial galactan, a polysaccharide of galactofuranose (Gal*f*) 30-40 residues long with a defined pattern of alternating β(1,5)-Gal*f* and β(1,6)-Gal*f* linkages. The mechanism facilitating these activities was previously unknown. Herein, we used X-ray Crystallography to determine the structure of GlfT2 from *Nocardia brasiliensis* bound to synthetic acceptor-substrate surrogates. These structures reveal a multi-residue binding site with alternate active site poses for β(1,5)-Gal*f* and β(1,6)-Gal*f* linkage formation. These findings suggest a model for galactan polymerization by GlfT2 in which acceptor sliding and feedback-regulated regioselectivity mediate processive, sequence-controlled polymerization. Broadly, these structures will provide insight into the activities of galactofuranosyl transferases enzymes in critical human pathogens and broadly inform on sequence-controlled glycosyltransferase polymerization mechanisms.

2.2 Introduction

Polymerizing glycosyltransferases (GT) are involved in the biosynthesis of polysaccharides, sugar biopolymers that play diverse and important roles across all kingdoms of life. These roles include mediating cell-cell communication, immune recognition, pathogen evasion, bacterial biofilm formation, energy storage, and cell wall integrity. The specific role played by a particular polysaccharide is dictated by its composition and length. To facilitate control of these properties, many polymerizing GTs perform their product elongation with an apparent processivity, in which association to the growing polysaccharide is maintained during elongation, permitting sequence and length control of the polysaccharide product.[44] Product release then produces the final full-length polysaccharide.

Galactofuranosyl transferase 2 (GlfT2) is a polymerizing glycosyltransferase that produces the mycobacterial galactan, a polysaccharide of galactofuranose (Gal*f*) that forms an essential component of the mycolyl-arabinogalactan-peptidoglycan structure of the mycobacterial cell wall.[6] The galactan is a 30-40 Gal*f* unit-long product composed of a defined sequence of alternating β(1,5)-Gal*f* and β(1,6)-Gal*f* glycosidic linkages.[19,42,50] The specific role of the galactan's length and sequence within the cell wall remains unknown, but we previously observed shortened galactan lengths to have a deleterious effect on cell viability and antibiotic resistance in *Mycobacterium smegmatis* (Msm).[37] Understanding the biosynthesis of the galactan by GlfT2 would inform efforts towards targeting this vulnerability and developing novel antibiotics against this essential polysaccharide.[51]

GlfT2 is a membrane-resident GT-2 family enzyme that uses the nucleotide-sugar donor UDP-Gal*f* to produce most of the mass of the galactan.[12,35,52] Previous knockout and knockdown

studies in *Mycobacterium tuberculosis* (Mtb) and Msm involving GlfT2 revealed its production to be essential to pathogen survival and highly vulnerable to perturbation.[22,36] GlfT2 associates to the cytoplasmic-face of the inner membrane, recognizing a primed acceptor of (Gal*f*)$_{2\text{-}3}$-GlcNAc-Rha linked to a decaprenyl lipid.[8] GlfT2 has been observed to extend its acceptor substrate processively, retaining the acceptor and extending it over rounds of UDP-Gal*f* turnover.[19,45,53,54] This retention and processive elongation would therefore provide GlfT2 with a mechanism for galactan sequence and length control, the latter of which is known to be critical for mycobacteria viability.[37]

Structural studies of GlfT2 from Mtb revealed a protomer containing a central catalytic GT-A domain, which assembles into a C4-symmetric tetrameric complex.[35,55] This complex forms a cup-shaped structure with a central cavity surrounded by patches of surface-exposed hydrophobic residues. These hydrophobic patches were predicted to mediate membrane association by GlfT2, leading to partial enclosure of the central cavity by the membrane interface. The UDP-bound complex of GlfT2 (PDB: 4FIY) revealed a surface-accessible donor-binding site, which informed a potential acceptor-binding site within the enclosed cavity. Informed by these observations, two models for GlfT2 galactan polymerization have been proposed: a cavity-based enclosure model and an acceptor-tethering model.[19,35] The former suggests the tetrameric cavity restricts acceptor diffusion during elongation to influence acceptor retention and polymerization extent, while the latter suggests that in an *in vitro* environment the acceptor lipid moiety influences substrate retention to drive processivity and length control.

Herein, we used X-ray Crystallography to determine the structures of GlfT2 from *Nocardia brasiliensis*, a Mycobacteriales pathogen involved in cutaneous nocardiosis.[56–58] These

structures were determined to 3.1 Å resolution and include an apo structure and two holo structures bound to acceptor-substrate surrogates (PDB: 9DTP, 9NJ0, 9NJ1). These surrogates include a phenoxy-dodec-2-enyl lipid linked to either a β(1,5)-terminal trisaccharide or a β(1,6)-terminal tetrasaccharide.[10] The acceptor-bound structures revealed three main observations. First, a common binding site that accommodates up to three Gal*f* residues regardless of acceptor length or sequence. A disordered-to-ordered transition of an active site loop appears to assist in binding. Second, two different active site binding poses are observed, where the previous β(1,5)-Gal*f* or β(1,6)-Gal*f* linkage dictates presentation of either the 6-OH or 5-OH, respectively, of Gal*f* for catalysis. Finally, partial density for the lipid moiety of the acceptor resides in proximity to the membrane-associating face of GlfT2. Molecular dynamics simulations suggest lipid binding to this region drives acceptor affinity *in vitro*.

From these observations, we propose a sliding model for processive and sequence-controlled polymerization by GlfT2. The multi-sugar binding site forms three subsites that retain the galactan during elongation regardless of linkage pattern, sliding the growing product between these subsites to reform the productive complex after catalysis. Differential positioning of the final Gal*f*, presenting either the 5-OH or 6-OH for catalysis, dictates reaction regioselectivity and is influenced by linkage length differences to the previous sugar. The influence of the previously formed linkage in this way suggests a feedback-regulated sequence-control mechanism during catalysis. We believe these finding will help inform future efforts towards targeting GlfT2 in critical human pathogens, and broadly inform the mechanisms underlying processive and sequenced-controlling glycosyltransferases.

2.3 Results

2.3.1 *Nbr*GlfT2 Functions as a GT-A Polymerase of the GT-2 Family

In order to characterize the activity and structure of *Nbr*GlfT2 (67.1% shared sequence with *Mtb*GlfT2), we produced a His_6-*Nbr*GlfT2 plasmid construct for recombinant expression in *E. coli* (Figure 2.1). The *glft2* gene was identified within a gene cluster following *ugm*, which codes for UDP-Galactopyranose Mutase (UGM).[59,60] UGM produces the UDP-Gal*f* donor for GlfT2 catalysis, and similarly precedes *glft2* in Mtb and Msm.[51,61] Purified His_6-*Nbr*GlfT2 produced galactan in the presence of the synthetic disaccharide acceptor **Compound 1** and UDP-Gal*f in vitro* (Figure 2.2a). Product distributions were measured by Matrix-Assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF) mass spectrometry, revealing the highest measured product mass corresponding to the addition of 19 Gal*f* units, with the highest abundance long-extension peak at 15 additional Gal*f* (Figure 2.2a). Compared to our previous measurements of *Mtb*GlfT2 with **Compound 1**, *Nbr*GlfT2 produces shorter *in vitro* product lengths.[19] This product length reduction correlates to our previous *in cellulo* galactan length analysis from Msm Δ*glft2* complemented with *Nbr*GlfT2.[37] This data suggests *Nbr*GlfT2 functions as a polymerase, controlling product length similarly in isolation and within a cellular environment.

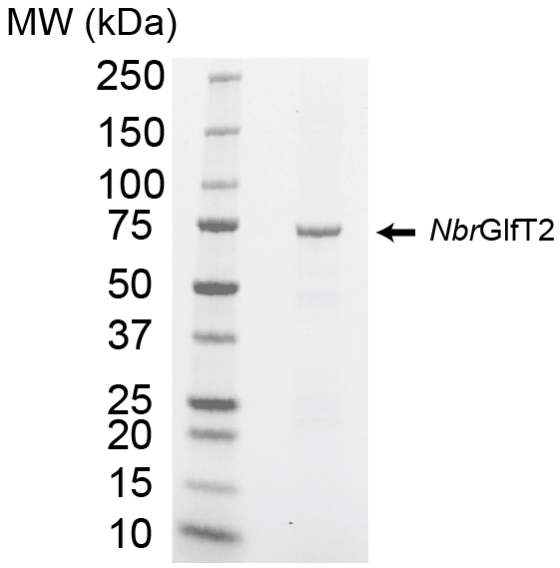

Figure 2.1: Gel of Purified *Nbr*GlfT2. Reducing SDS-PAGE gel of *Nbr*Glft2. The molecular weight corresponds to a single protomer of *Nbr*GlfT2.

*Nbr*GlfT2 is a GT-2 family enzyme, containing a central GT-A motif responsible for facilitating

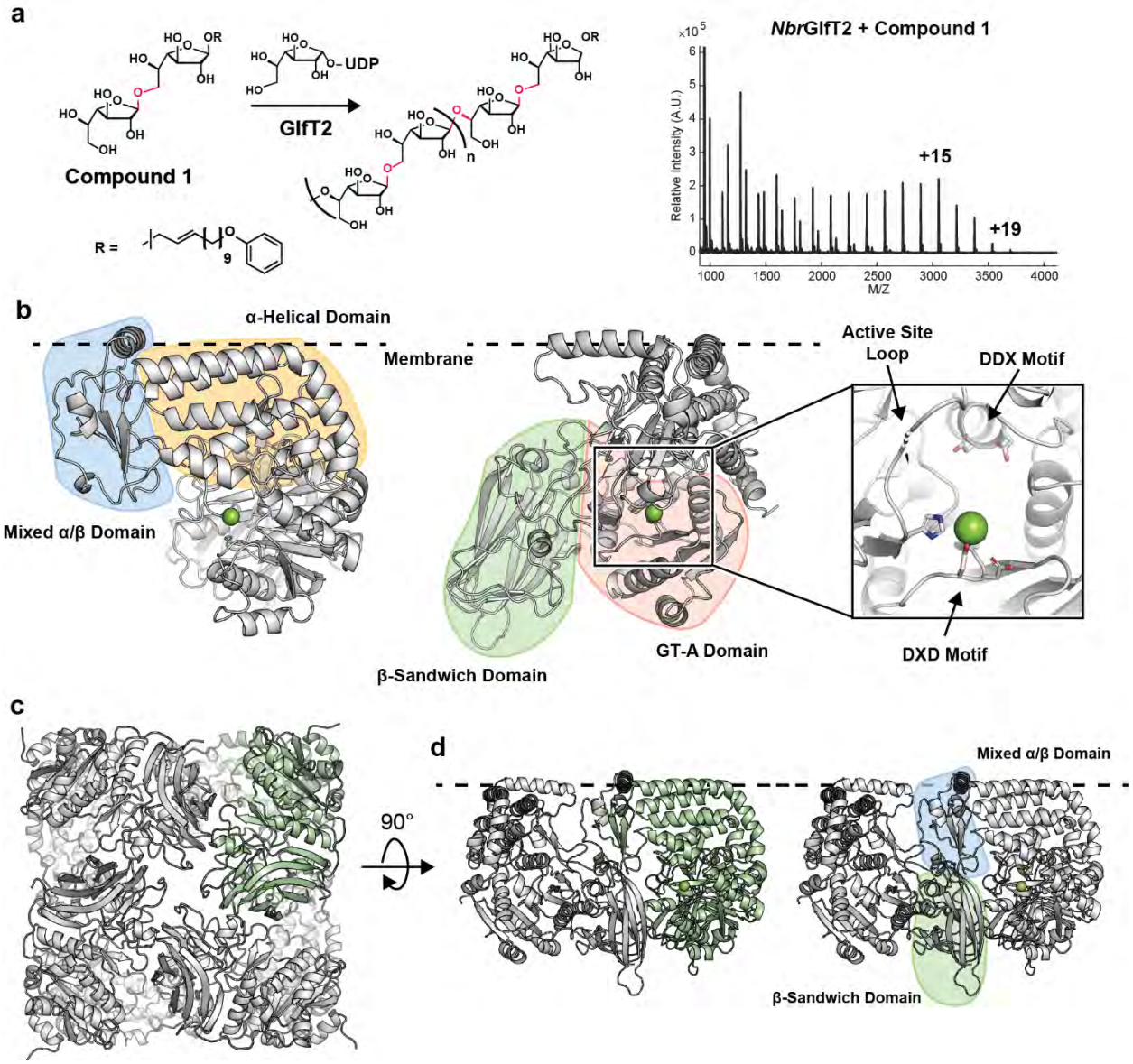

Figure 2.2: *Nbr*GlfT2 Polymerization Reaction and Structural Model. a) Reaction schematic of galactan polymerization by *Nbr*GlfT2 in the presence of synthetic disaccharide acceptor **Compound 1** and UDP-Gal*f*. The MALDI-TOF MS spectrum of reaction products for wild-type *Nbr*GlfT2 is shown. b) Structural model of the *Nbr*GlfT2 protomer bound to membrane, with the four domains highlighted. The active site of *Nbr*GlfT2 is shown, labeled with the donor-binding DXD motif coordinated to a Mg^{2+} ion, the catalytic DDX motif, and the active site loop. c) Structure of the C4-symmetric tetrameric assembly of *Nbr*GlfT2, colored for individual protomer chain. d) Membrane association of the tetrameric assembly, with the oligomerizing domains highlighted.

glycosyltransferase activity.[31,62] Two essential motifs are generally required for GT-A activity:

the metal-binding DXD motif ($_{268}$DDD$_{270}$) and catalytic DDX motif ($_{382}$DDV$_{384}$; Figure

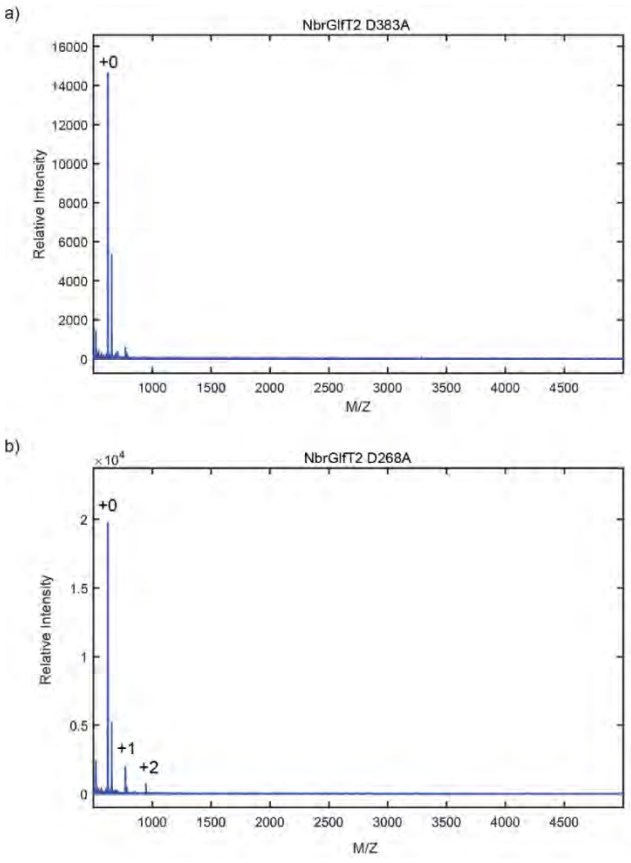

Figure 2.3: Substitution of Active Site Residues in *Nbr*GlfT2. a) Substitution of the catalytic aspartate base D377 to alanine shows complete loss of *Nbr*GlfT2 activity. b) Substitution of metal-ion binding aspartate D268 to alanine shows substantial reduction of GlfT2 activity.

2.2b).[31,35,53] Substitutions within each motif either disrupted or abolished activity, with the D268A construct catalyzing up to 2 Gal*f* additions and the D383A construct producing no catalysis (Figure 2.3a, b). Therefore, the GT-A domain of *Nbr*GlfT2 functions as the single catalytic domain responsible for galactan biosynthesis.

2.3.2 *Nbr*GlfT2 Produces Alternating Pattern of β(1,5) and β(1,6) Glycosidic Linkages

To evaluate the glycosidic linkage pattern of the Nbr galactan produced by *Nbr*GlfT2, we analyzed the galactan product from *in vitro* catalysis with **Compound 1** and purified cell wall from a *Nbr*GlfT2-complemented Msm Δ*glft2* strain.[37] The presence of 5-linked Gal*f* and 6-linked Gal*f* products were identified in both *in vitro* and *in cellulo* galactan products, with an additional 5,6-linked Gal*f* product *in cellulo* from arabinan adduct formation (Figure 2.4 and 2.5). This dual linkage composition is in contrast to previous analysis of endogenous galactan in *N. brasiliensis* and *N. asteroides*, which were enriched for 5-Gal*f* with only trace amounts of 6-Gal*f* linkages.[42] This discrepancy could arise from extensive modification of the endogenous

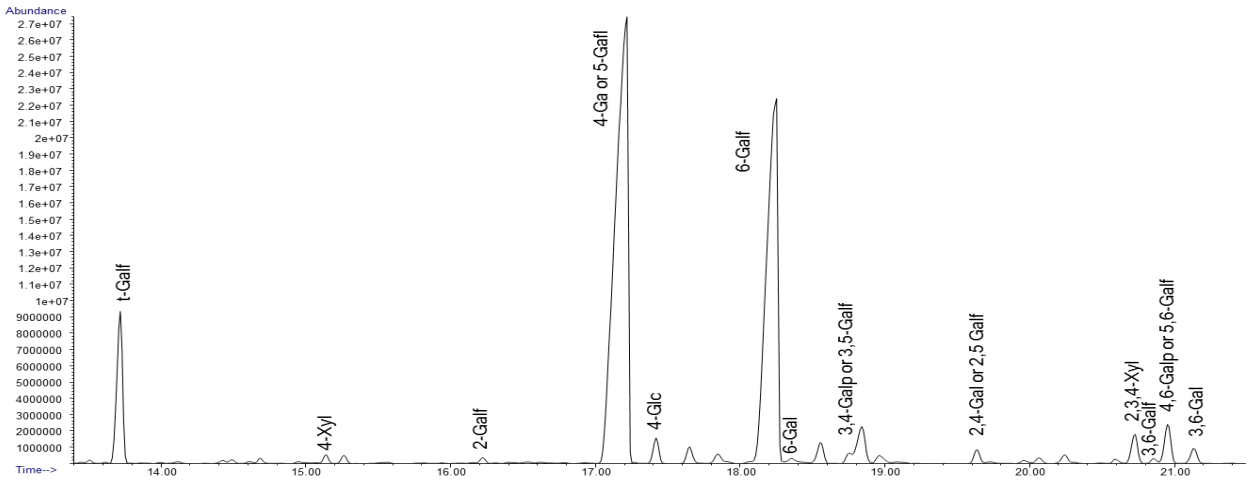

Figure 2.4: *In Vitro* Galactan Product Analysis. Linkage analysis of galactan produced from *Nbr*GlfT2 and Compound 1 by GC-MS analysis of partially methylated alditol acetate products.

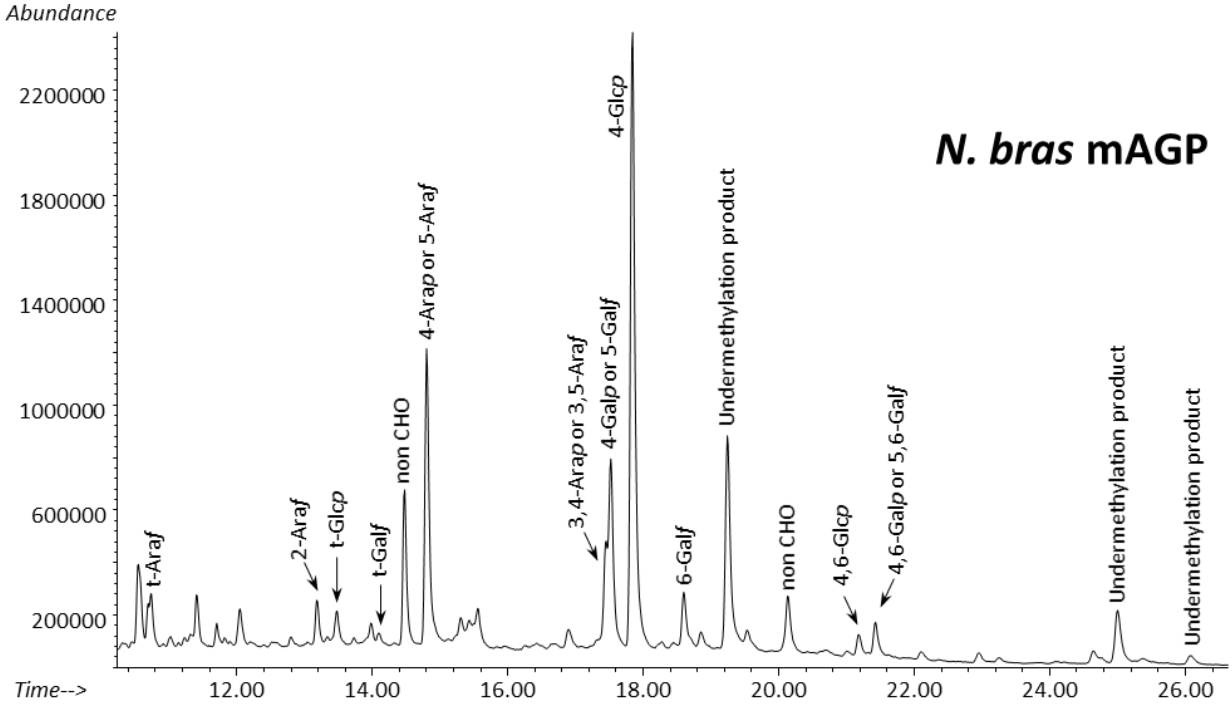

Figure 2.5: Linkage Analysis of Isolated mAGP. Linkage analysis of arabinogalactan from *M. smegmatis* Δ*glfT2* complemented with *Nbr*GlfT2 by GC-MS of partially methylated alditol acetate products. The major peak 4-Glc*p* constitutes a contaminant in these analyses.

Nocardia galactan, increasing 5,6-linked products while reducing 6-linked products to traces levels. These results suggest additional investigation is needed to fully characterize the *Nocardia*

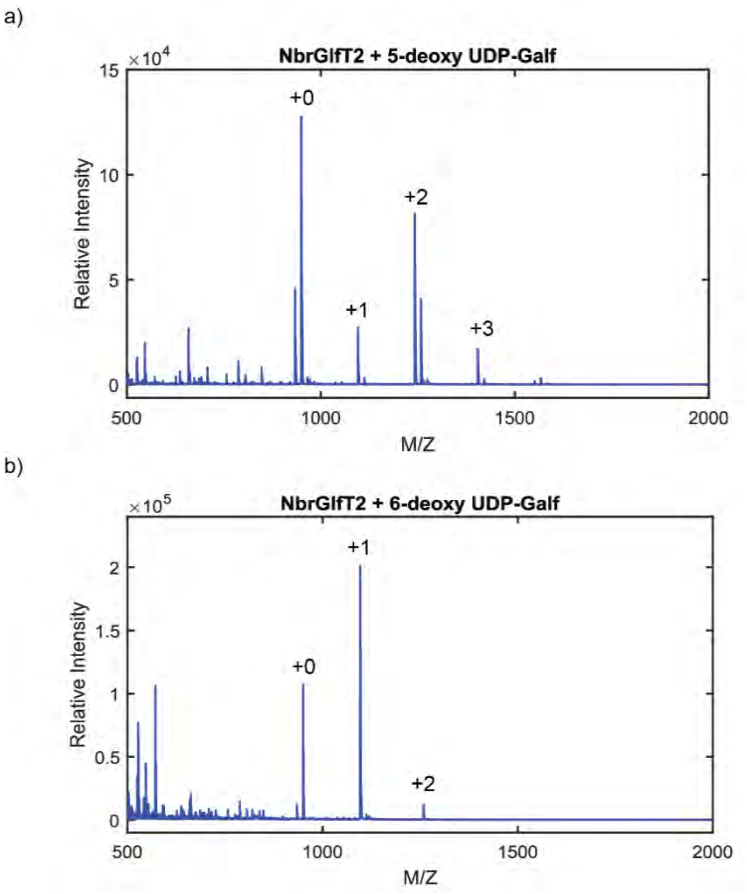

Figure 2.6: Deoxy Donors Reveal Conserved Fidelity for Alternating Pattern. *In vitro* NbrGlfT2 galactan extension reactions using tetrasaccharide acceptor, with (a) 5-deoxy UDP-Gal*f* and (b) 6-deoxy UDP-Gal*f*. Both reactions led to truncation of the galactan, with extent of truncation matching what is expected for the GlfT2 product alternating pattern.

arabinogalactan and characterize potential modifying enzymes.

To determine the pattern of β(1,5) and β(1,6) linkage distributions produced by *Nbr*GlfT2, we incubated *Nbr*GlfT2 and **Compound 1** in the presence of either 5-deoxy or 6-deoxy UDP-Gal*f*. Donor-analogs modified at the 5- and 6-positions were previously used to confirm the fidelity of *Mtb*GlfT2 for producing an alternating β(1,5) and β(1,6) linkage pattern.[43,63] Catalysis with 5-deoxy and 6-deoxy UDP-Gal*f* led to polymerization termination, producing +2 and +1 Gal*f* extensions as the major products, respectively (Figure 2.6a, b). We therefore conclude that *Nbr*GlfT2 functions as a polymerase similar to *Mtb*GlfT2, producing both β(1,5)-Gal*f* and β(1,6)-Gal*f* products through bifunctional catalysis, alternating the linkage pattern throughout the galactan product.

2.3.3 *Nbr*GlfT2 Structure Reveals Conservation of Essential Domains

X-ray Crystallography was performed to elucidate the structure of *Nbr*GlfT2. The apo and acceptor-bound structures of *Nbr*GlfT2 were solved to 3.0 Å (PDB: 9DTP, 9NJ0, 9NJ1; Table

Table 2.1: *Nbr*GlfT2 Data Collection and Refinement Statistics (Molecular Replacement)

	NbrGlfT2 (Apo) 9DTP	NbrGlfT2 (Compound 2) 9NJ0	NbrGlfT2 (Compound 3) 9NJ1
Data collection			
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions:			
a, *b*, *c* (Å)	107.0, 205.0, 295.3	106.8, 205.0, 294.1	106.2, 203.9, 292.5
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	1.03	0.92	0.92
Resolution (Å)	20.05-3.11 (3.22-3.11)	34.61-3.20 (3.31-3.20)	34.88-3.15 (3.26-3.15)
R_{pim}	0.14 (0.72)	0.14 (0.44)	0.14 (0.60)
I / σI	4.9 (1.0)	5.5 (1.6)	5.1 (1.2)
Completeness (%)	96.0 (99.0)	94.9 (94.7)	98.6 (98.5)
Redundancy	4.4 (4.5)	5.2 (5.3)	6.9 (7.0)
Refinement			
No. reflections	111,965 (11402)	101,657 (10,021)	128,630 (12,730)
R_{work} / R_{free}	0.221 / 0.258	0.200 / 0.248	0.238 / 0.278
No. atoms			
Protein	39,068	39,217	39,091
Ligand/ion	262	474	547
Water	0	0	0
B-factors			
Protein	64.7	51.8	53.9
Ligand/ion	76.4	77.5	79.2
R.m.s. deviations			
Bond lengths (Å)	0.002	0.003	0.011
Bond angles (°)	0.43	0.61	0.61

* Values in parentheses are for highest-resolution shell.

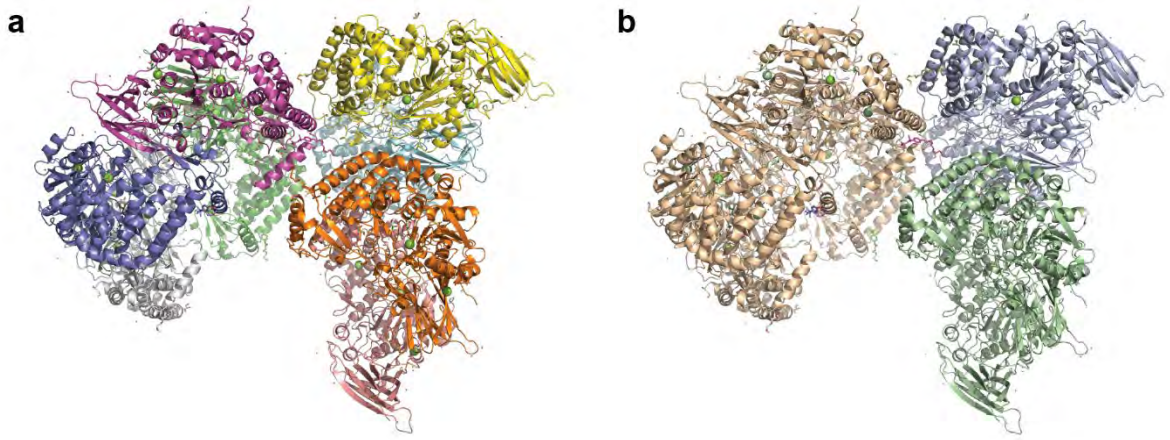

Figure 2.7: Asymmetric Unit of Apo *Nbr*GlfT2. The structure of Apo *Nbr*GlfT2 was solved to 3.0 Å in the $P2_12_12_1$ space group. The structure consists of 8 separate chains. Asymmetric unit colored by chain. b) Asymmetric unit colored to highlight different tetrameric complexes or parts of tetrameric complexes.

2.1), revealing an asymmetric unit containing a C_4-symmetric tetrameric complex (Figure 2.2c; Figure 2.6). The tetrameric assembly of *Nbr*GlfT2 agrees with the biological-assembly proposed for *Mtb*GlfT2 (PDB: 4FIX, 4FIY), with high structural similarity ($RMSD_{tetramer}$ = 33.4 Å, Figure 2.8a).[35]

*Nbr*GlfT2 consists of four domains: a core catalytic GT-A domain surrounded by a β-sandwich domain at the N-terminus, and an α-helical and an α/β domain at the C-terminus (Figure 2.2b). This is homologous to *Mtb*GlfT2 ($RMSD_{monomer}$ = 0.58 Å, Figure 2.9).[35] The active site channel of *Nbr*GlfT2 is formed by the GT-A and α-helical domains, and is highly conserved compared to *Mtb*GlfT2 (Figure 2.10). Despite the overall similarity between both orthologs, distinct differences are present at the β5-β6 hairpin of the β-sandwich domain and the joining loop between the α/β and β-helical domains (Figure 2.9). *Nbr*GlfT2 oligomerizes at the interfaces of the β-sandwich and mixed α/β domains (Figure 2.2d), with the surface area for the primary oligomerizing interface of nearly 1200 $Å^2$.[64] The high structural similarity between *Nbr*GlfT2

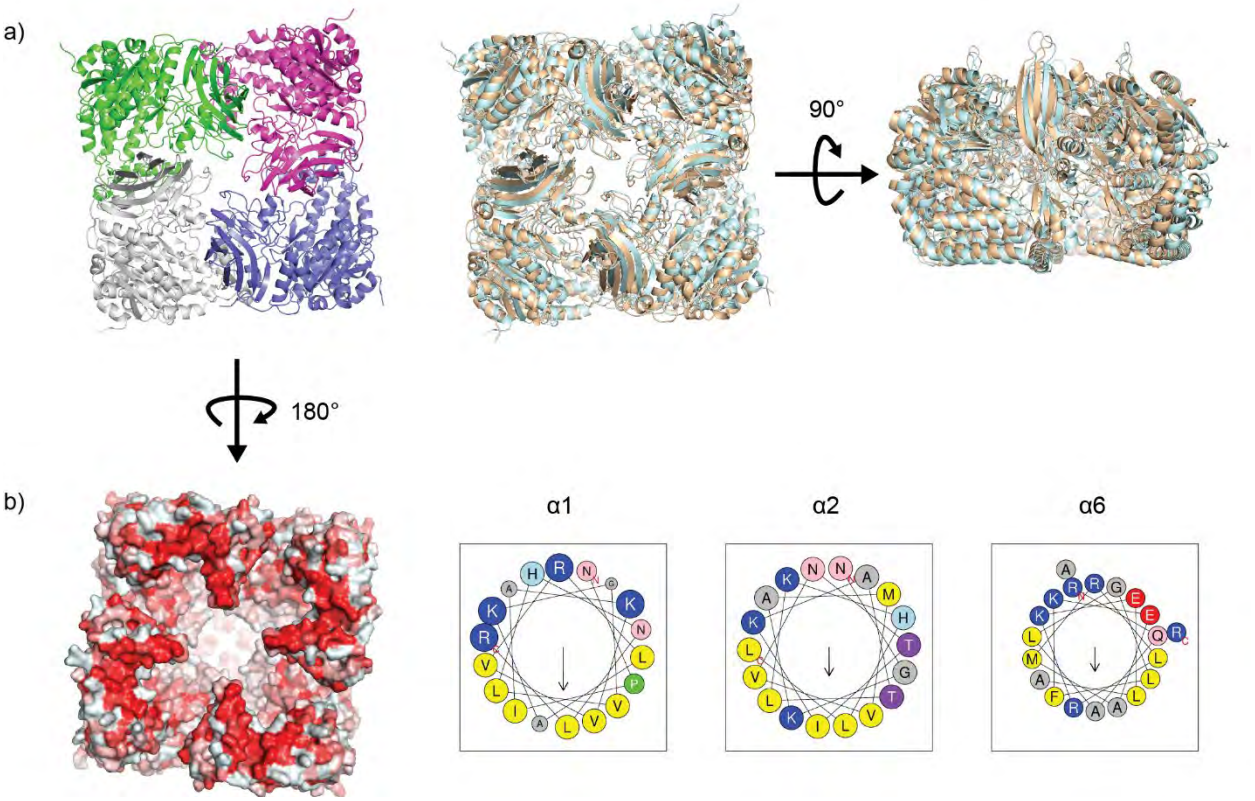

Figure 2.8: Biological Assembly of *Nbr*GlfT2 and *Mtb*GlfT2. a) NbrGlfT2 forms a C_4-symetric tetramer. Superimposition of NbrGlfT2 (cyan) and MtbGlfT2 (wheat, PDB:4FIX) biological assemblies shows high conservation of overall quaternary structure. b) NbrGlfT2 forms a highly hydrophobic surface on one side of its tetramer, very similar to MtbGlfT2. This surface is formed by amphipathic helices with the hydrophobic residues oriented outside of the tetramer. Surface colored from most hydrophobic (red) to most hydrophilic (white) residues (Eisenberg, et. al. 1984). Helix wheels reveal polarization of the hydrophobic residues towards one face of each helix, with hydrophobic moment represented by the arrow in the center of the wheel.

and *Mtb*GlfT2 highlights the evolutionary conservation of the tertiary and quaternary structure of GlfT2.

In *Nbr*GlfT2, residues 402-412 form a flexible loop between the GT-A and α-helical domains to span the active site (Figure 2.2b). This flexible active-site loop is present in other GT-A enzymes and is suggested to be involved in substrate binding and catalysis.[31] In apo *Nbr*GlfT2, the active-site loop resides in region of discontinuous electron density that precluded complete loop

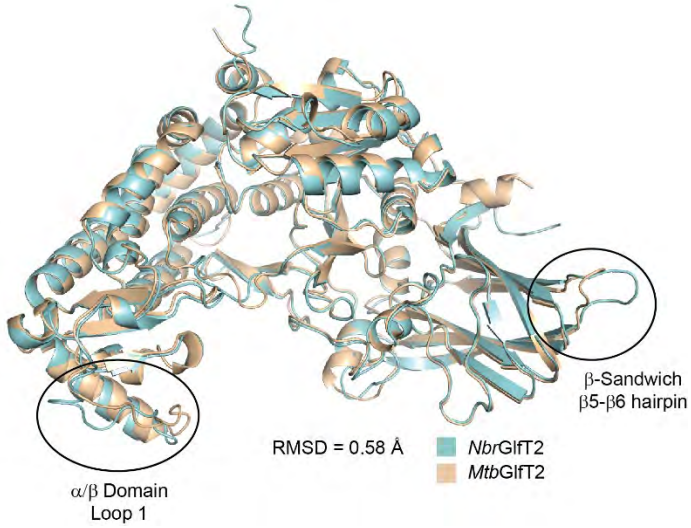

Figure 2.9: Comparison of Monomeric *Nbr*GlfT2 and *Mtb*GlfT2. Superimposition of the protomers of *Nbr*GlfT2 (cyan) and *Mtb*GlfT2 (wheat, PDB: 4FIX), highlighting their high structural similarity (RMSD = 0.58 Å). Distinct differences are apparent at the β5-β6 hairpin loop of the β-sandwich domain, and loop 1 of the α/β domain that joins it to the α-helical domain.

modeling. The active-site loop sequence is fully conserved between *Nbr*GlfT2 and *Mtb*GlfT2, and high flexibility was observed for both apo and UDP-bound *Mtb*GlfT2.[35]

Evaluation of unmodeled density in apo *Nbr*GlfT2 revealed occupation of Mg^{2+} (Figure 2.2b) and glycerol in the active site. Mg^{2+} coordinates to the DXD metal-binding motif of the donor binding site. Glycerol is modeled in density at the catalytic site, forming interactions with the DDX catalytic motif (Figure 2.11). Glycerol is present in both *Nbr*GlfT2 and *Mtb*GlfT2 crystallization conditions, and density larger-than-expected for water was present in both

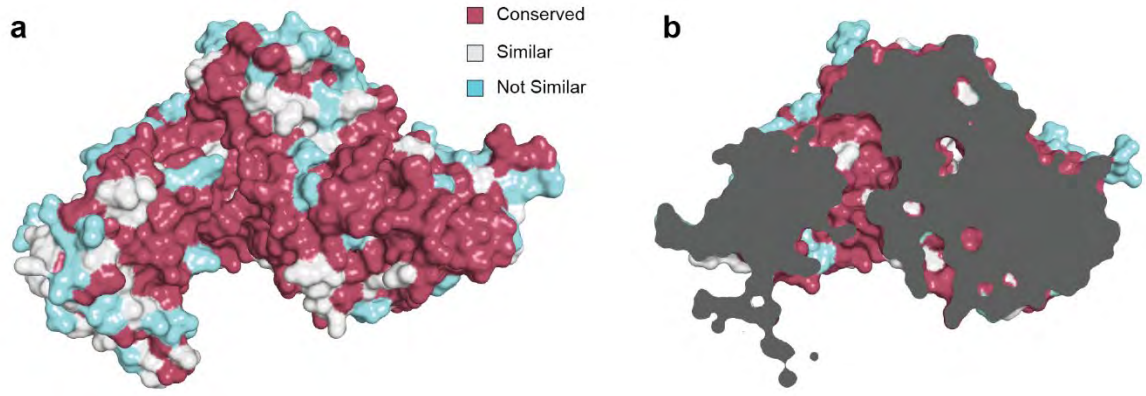

Figure 2.10: Conservation of Residues between *Nbr*GlfT2 and *Mtb*GlfT2. a, Surface representation of *Nbr*GlfT2, colored for surface-residue conservation. Maroon – conserved; white – similar; cyan – not similar. **b,** Surface representation clipped to highlight the active site channel, containing the donor- and acceptor-binding sites. A highly conserved channel is illustrated, with three residues substituted to similar amino acids.

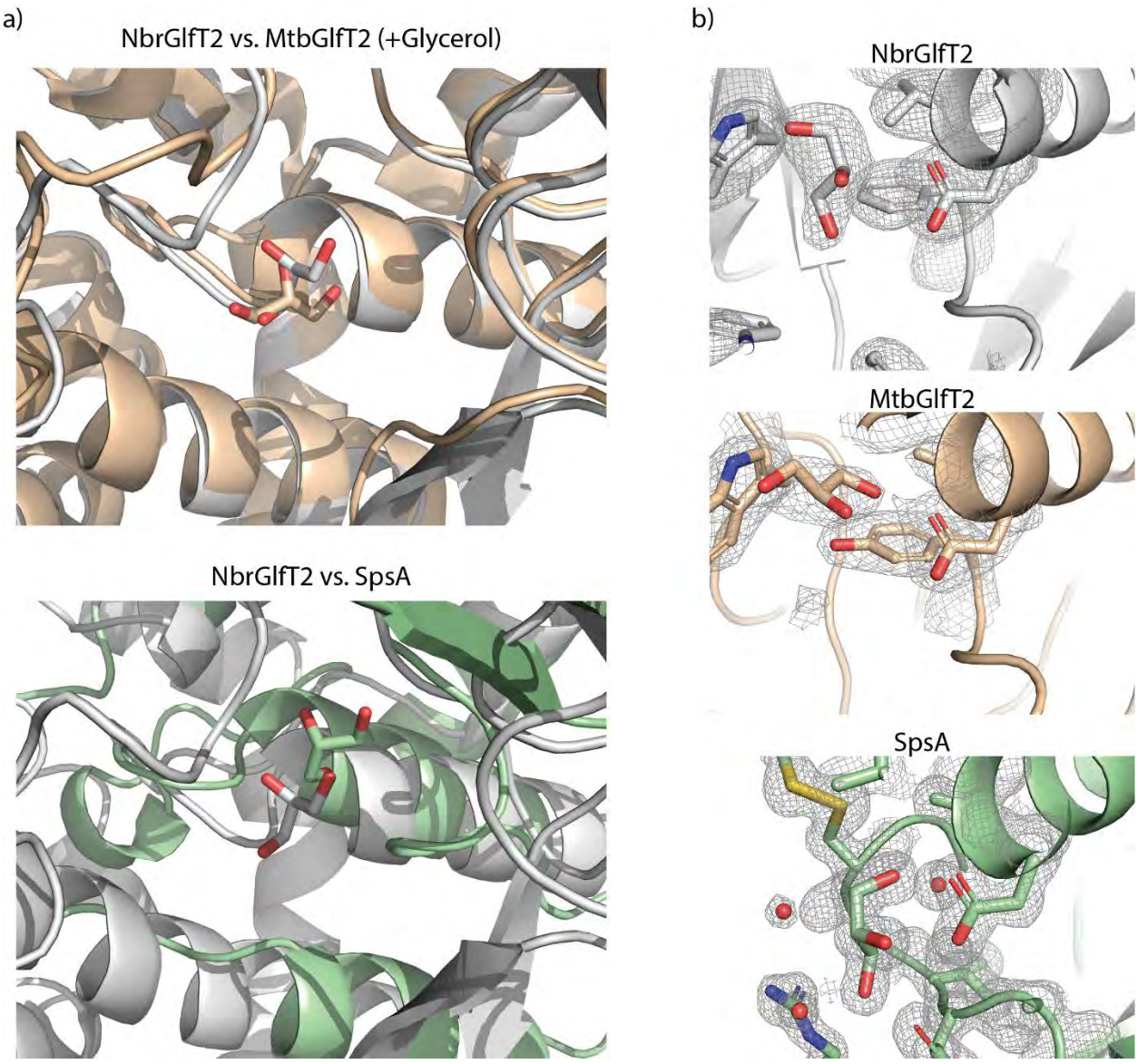

Figure 2.11: Glycerol Occupation of Active Site in GlfT2 and SpsA. a) Structure superimpositions of MtbGlfT2 (PDB: 4FIX) refined with glycerol included and published model of SpsA (PDB: 1QG8) with NbrGlfT2. b) Electron density of active site glycerol in NbrGlfT2, MtbGlfT2, and SpsA highlighting close interactions with catalytic aspartates of each enzyme (2mFo-DFc electron density map in grey, contoured to 1.0 σ).

electron density maps near the DDX motif; this density was previously modeled as a chain of water molecules in *Mtb*GlfT2.[35] Similar glycerol occupation was seen for the GT-2 family enzyme SpsA (PDB: 1QG8, Figure 2.11a, b). Interactions between the glycerol and the catalytic

base D383 suggests glycerol may mimic the association of the exocyclic diol of the final acceptor Gal*f*.

2.3.4 Membrane Association and Cavity Formation by Tetrameric *Nbr*GlfT2

The structure of the *Nbr*GlfT2 tetrameric complex reveals enrichment of surface exposed hydrophobic residues at a single face (Figure 2.8). This region consists of three amphipathic helices per protomer enriched for hydrophobic and basic amino acids: α2 (residues 432-450) and α6 (residues 585-604) of the α-helical domain, and α1 (residues 520-537) of the α/β domain.[65] These helices are analogous to the proposed membrane-associating helices of *Mtb*GlfT2, suggesting a similar role in localizing *Nbr*GlfT2 to its site of action (Figure 2.2d).[35]

We sought to inform whether structural differences in *Nbr*GlfT2 significantly influenced its reduced galactan product length distribution compared to the *Mtb* ortholog by comparing the calculated cavity volume to measured product lengths. The previous cavity-filling model for GlfT2 length control proposed a positive correlation between tetramer cavity volume and galactan product length.[35] We determined the volumes of cavities formed by two stacked tetramers in order to reduce overestimation by inclusion of volumes below the tetramer (Figure 2.12). Calculated volumes for apo *Nbr*GlfT2 and *Mtb*GlfT2 (PDB 4FIX) were nearly 62,000 Å^3 and 52,000 Å^3, respectively, for a single tetramer.[66] This almost 20% increase in cavity volume for *Nbr*GlfT2 is in contrast to it shorter galactan product lengths. This disagreement with the proposed cavity model for GlfT2 length control suggests a need to reevaluate the role the GlfT2 cavity performs during catalysis.

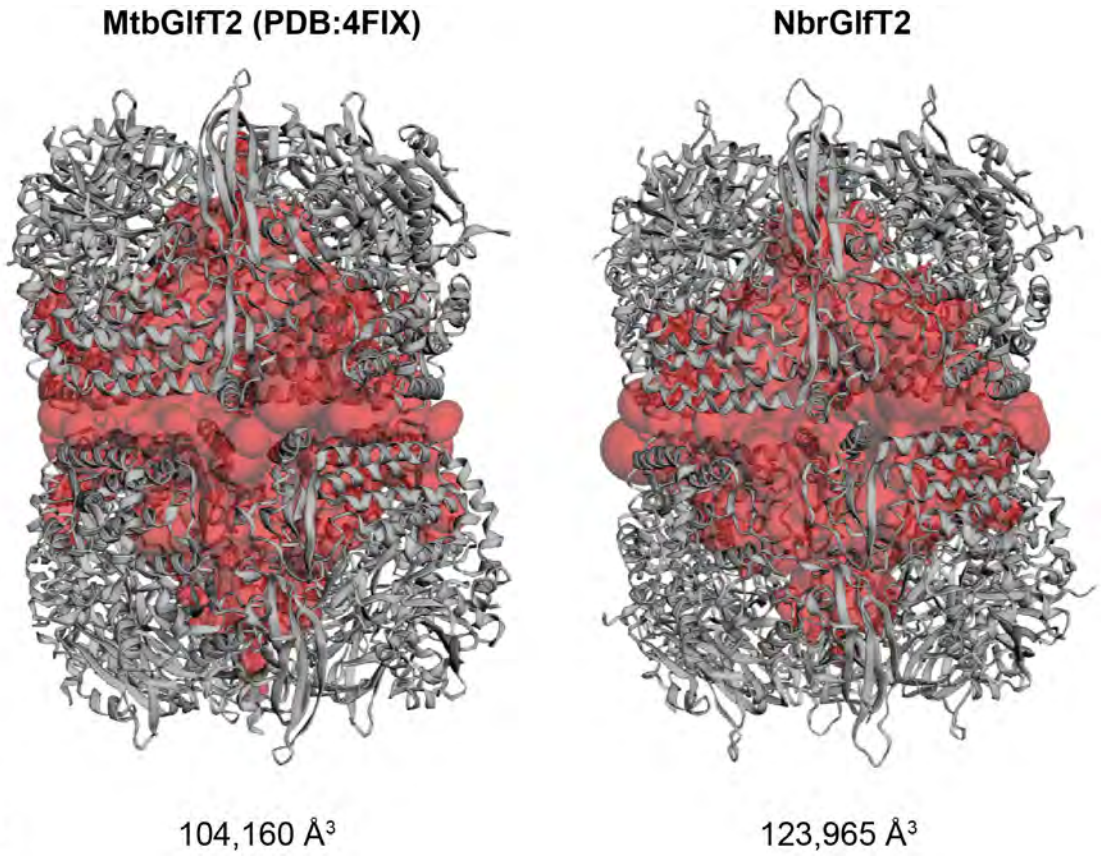

Figure 2.12: Volume Comparison of *Mtb*GlfT2 and *Nbr*GlfT2. The CastP server was used to calculate the volumes for the MtbGlfT2 (PDB: 4FIX) and NbrGlfT2 cavities. Two tetramers were stacking with the membrane-associating regions pointed toward each other to reduce inclusion of empty space below the tetramer into volume calculation. Tetramer cavity volumes are assumed to be estimated from the stacked tetramer volume by $V_{cavity} = V_{stacked} / 2$ (V_{cavity}: single tetramer cavity volume; $V_{stacked}$: CastP-calculated volume of two stacked tetramers).

2.3.5 Multi-Sugar Galactan-Binding Mediates Processive Polymerization

To inform the mechanism of galactan polymerization and sequence control by GlfT2, we determined the structure of *Nbr*GlfT2 with the trisaccharide acceptor **Compound 2** and tetrasaccharide acceptor **Compound 3** (Figure 2.13a, b; PDB 9NJ0 and 9NJ1). The *Nbr*GlfT2 + **Compound 3** complex represents the Gal*f*-β(1,6) addition to *Nbr*GlfT2 + **Compound 2**, informing both acceptor elongation and linkage production.

To facilitate processive polymerization, GlfT2 was proposed to retain its acceptor at a single active site during elongation.[19,45,53,54,67] The acceptor-binding site of both complexes reveals

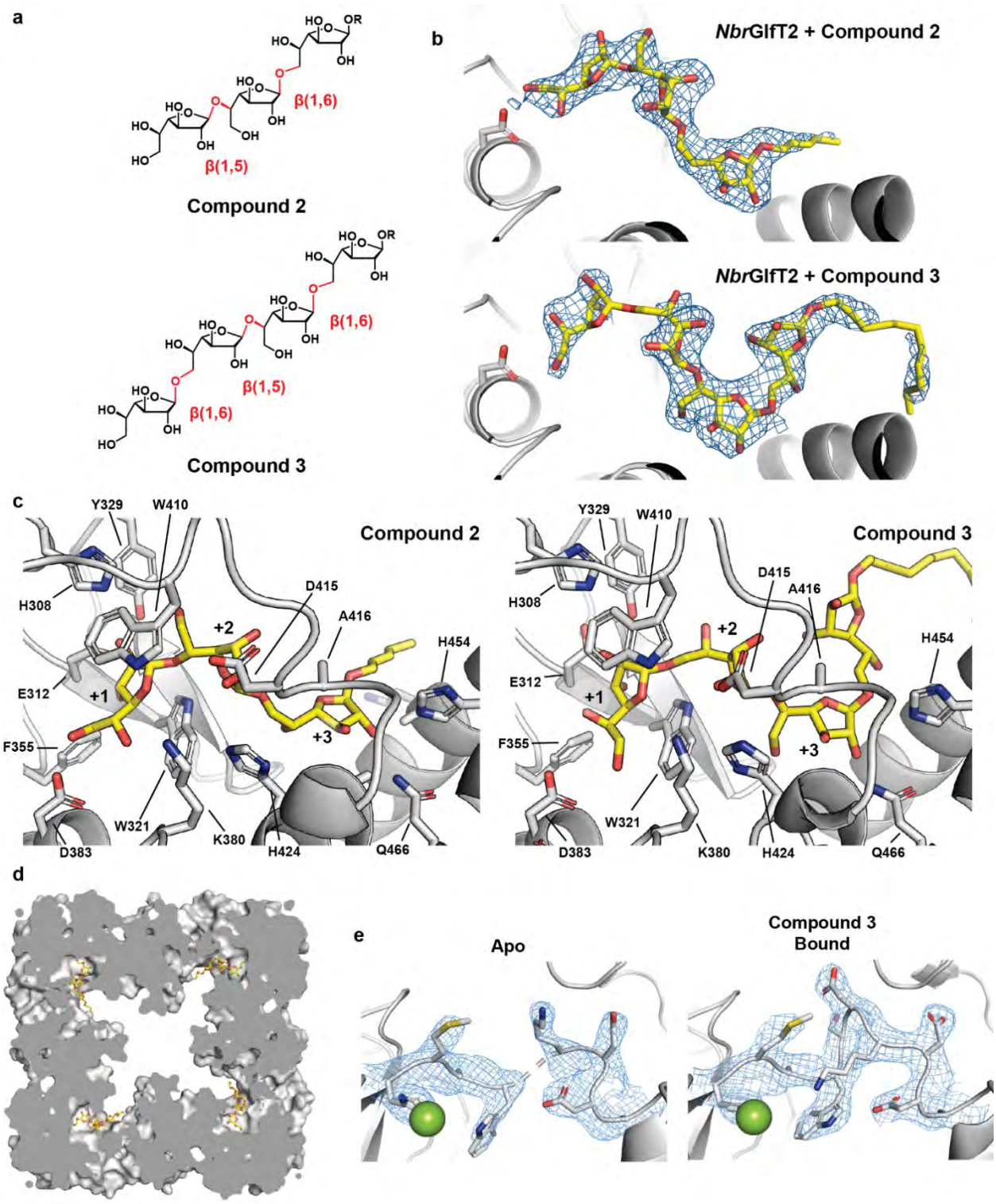

Figure 2.13: NbrGlfT2 Acceptor-Bound Structures and Subsite Interactions. a, Chemical structures of the synthetic trisaccharide acceptor **Compound 2** and tetrasaccharide acceptor **Compound 3**. Individual glycosidic linkages are labeled and highlighted in red. **b,** Electron density maps around for **Compound 2** and **Compound 3** in their bound complexes with *Nbr*GlfT2. The catalytic aspartate D383 is shown. 2Fo-Fc density map contoured to 1.0 σ. **c,**

Residues making significant interactions with **Compound 2** and **Compound 3** in their bound complexes. Three major binding subsites are labeled. **d,** The *Nbr*GlfT2 + **Compound 2** complex clipped from above the membrane reveals the acceptor-binding site within the cavity**. e,** The active site loop undergoes a ordering transition upon acceptor binding, illustrated by improved electron density and side chain placement (illustrated with **Compound 3**). 2Fo-Fc density map contoured to 1.0 σ.

three major subsites spanning nearly 20 Å, stabilizing three Gal*f* residues at the non-reducing end of the acceptor (Figure 2.13c). Galactan binding occurs within the cavity of *Nbr*GlfT2, suggesting the elongated galactan would occupy this space above the membrane interface (Figure 2.13d). Similar acceptor binding in both complexes reveals the acceptors are accommodated by the three subsites regardless of linkage pattern. This promiscuous binding would accommodate the acceptor during elongation, while the subsites would provide sufficient binding interactions to retain the acceptor. Acceptor retention across a multisite binding region suggests processive polymerization of the galactan by GlfT2 occurs through a sliding mechanism.[67]

Subsite +1 forms the acceptor active site, containing the catalytic DDX motif. The final Gal*f* of the acceptor binds to subsite +1 through interactions with the GT-A domain and active site loop at residues H308, E312, W321, F355, W359, K380, D383, and W410 (Figure 2.13c). Previous substitution of the corresponding residues in *Mtb*GlfT2 for E312 and W410 revealed disrupted activity.[35] Interestingly, W321 appears to undergo a nearly 180° flip upon acceptor binding. This subsite reveals a pocket highly enriched for aromatic residues. Carbohydrate-binding pockets are often enriched for aromatics, with association driven by entropic contributions and CH-π interactions with the face of aromatic residues.[68,69] We hypothesize this highly aromatic pocket provides necessary interactions to stabilize the reacting sugar while permitting the contrasting orientations needed for Gal*f*-β(1,5) and Gal*f*-β(1,6) formation.

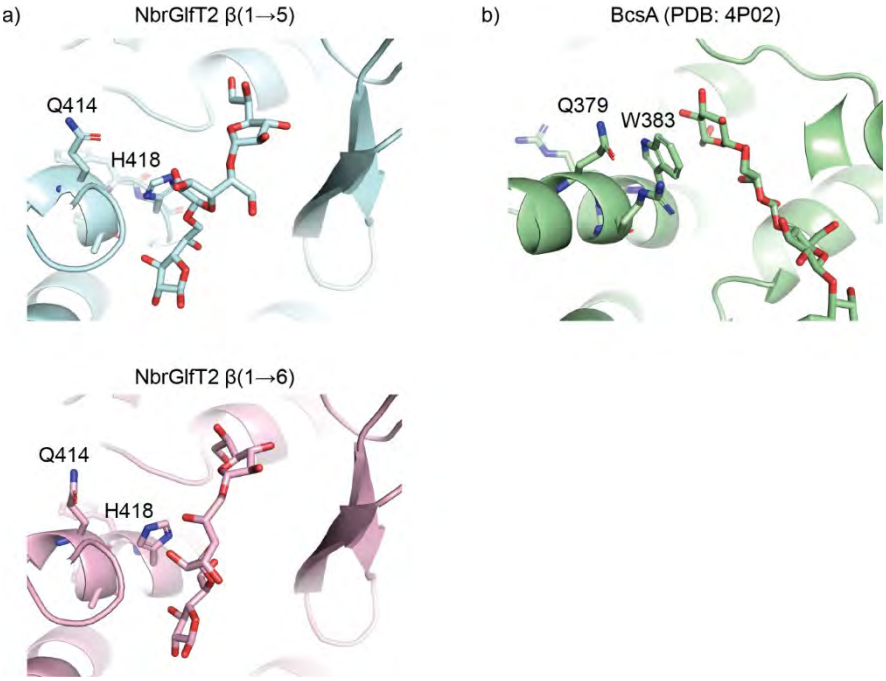

Figure 2.14: QXXRW-like Motif in GlfT2 Compared to Processive Cellulose Synthase BcsA. a) Structures of **Compound 2** and **3** bound to *Nbr*GlfT2. b) Interaction of QXXRW motif and cellulose acceptor in cellulose synthase BcsA (PDB: 4P02), revealing significant interactions between W383 and the non-reducing end of the cellulose acceptor.

Subsite +2 and +3 bind the next two Gal*f* residues of the acceptor outside of the active site (Figure 2.13c). Both subsites include interactions with the active site loop and α-helical domain, suggesting both structural features are essential for GlfT2 processivity. Subsite +2 forms binding interactions with W321, D415, and H424. Interestingly, H424 replaces the tryptophan of the QXXRW motif, which was implicated in processive polymerization for other GT-2 enzymes (Figure 2.14).[70] Reactivity of the D415A and H424A substitutions revealed retention of polymerase activity, but to shorter product lengths (Figure 2.15), highlighting their importance to the processive mechanism of GlfT2. Subsite +3 stabilizes the third binding Gal*f* through interactions with A416, H454, and Q466. The interactions at subsites +2 and +3 likely help stabilize the final sugar in the highly aromatic active site.

2.3.6 Synthetic Acceptor Lipid Aglycon and Active Site Loop Promote Acceptor Association

Previous optimization of the synthetic acceptor in the presence of *Mtb*GlfT2 revealed the influence of the lipid aglycon on *in vitro* reactivity, observing a correlation between galactan

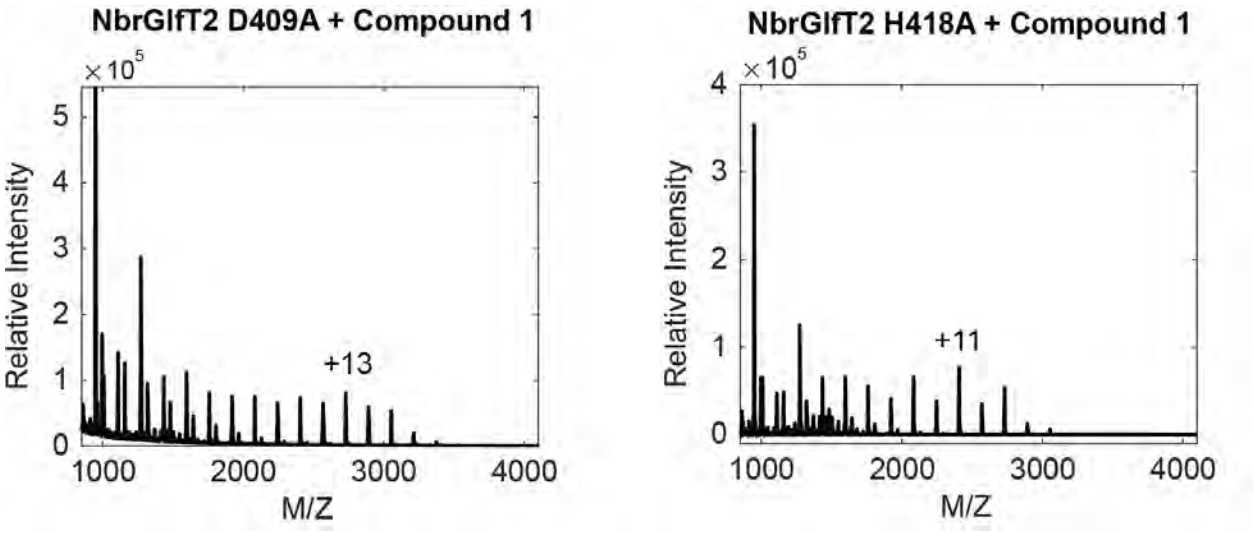

Figure 2.15: *In Vitro* Activity of D409A and H418A Reveal Reduced Polymerase Activity. MALDI-TOF spectra for *in vitro* reaction products of **Compound 1** extension in the presence of *Nbr*GlfT2 and UDP-Gal*f*. Product lengths reveal shorter product lengths and lower product abundance for both the D409A and H418A constructs.

product length and lipid length.[19] Both acceptor-complexes of *Nbr*GlfT2 reveal electron density for a portion of the lipid aglycon (Figure 2.13). This density suggests localization of the lipid near the membrane-associating hydrophobic patch formed by the α-helical and mixed α/β domains (Figure 2.16). The electron density of the lipid aglycon suggests its influence on *in vitro* reactivity arises from its interactions with the membrane-associating region of GlfT2. Across all three subsites, interactions occur between the bound acceptor and the flexible active-site loop. These include interactions with W410, D415 and A416. Interactions between these residues and the bound acceptor were observed to induce an ordering transition of the flexible loop (Figure 2.13e). The acceptor-bound structures suggest that the loop may either perform a sensing function to control catalysis, or assist in retaining the acceptor during elongation.

2.3.7 Common Binding to Acceptor-Binding Subsite Promotes Regioselective Catalysis

To inform the mechanism of galactan sequence control by GlfT2, we compared the active site binding of **Compound 2** and **Compound 3** (Figure 2.17a). Both complexes show similar

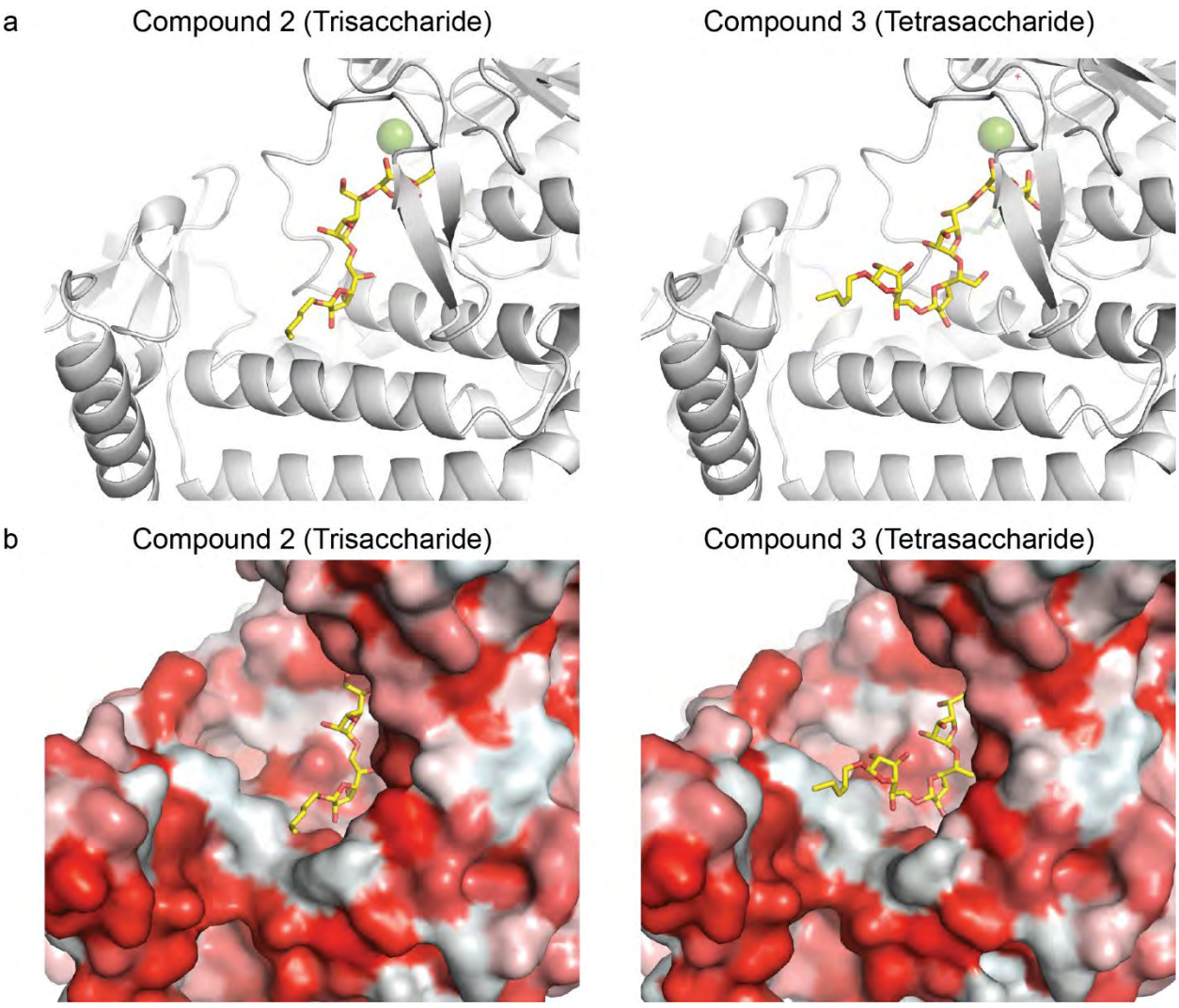

Figure 2.16: Proximity of Synthetic Acceptor Lipid Aglycon to Membrane-Associating Region. a, Cartoon representation a single protomer of *Nbr*GlfT2 in complex with either **Compound 2** or **Compound 3** (both shown in yellow). The lipid aglycon is shown with only five carbons, and localization is shown near α2 (residues 432-450) and α6 (residues 585-604) of the α-helical domain and α1 (residues 520-537) of the α/β domain. **b,** Surface representation, colored to show surface residue hydrophobicity with red indicating highly hydrophobic residues. The lipid tail of both acceptors is in close proximity to the highly hydrophobic region of GlfT2, suggesting significant binding interactions between the acceptor aglycon and the GlfT2 membrane associating region.

binding to subsites +2 and +3, regardless of linkage pattern. In subsite +1, *Nbr*GlfT2 +

Compound 2 reveals electron density for the reactive Gal*f* and the previous β(1,5) linkage, while

*Nbr*GlfT2 + **Compound 3** was absent of density for the β(1,6) linkage, likely due to the

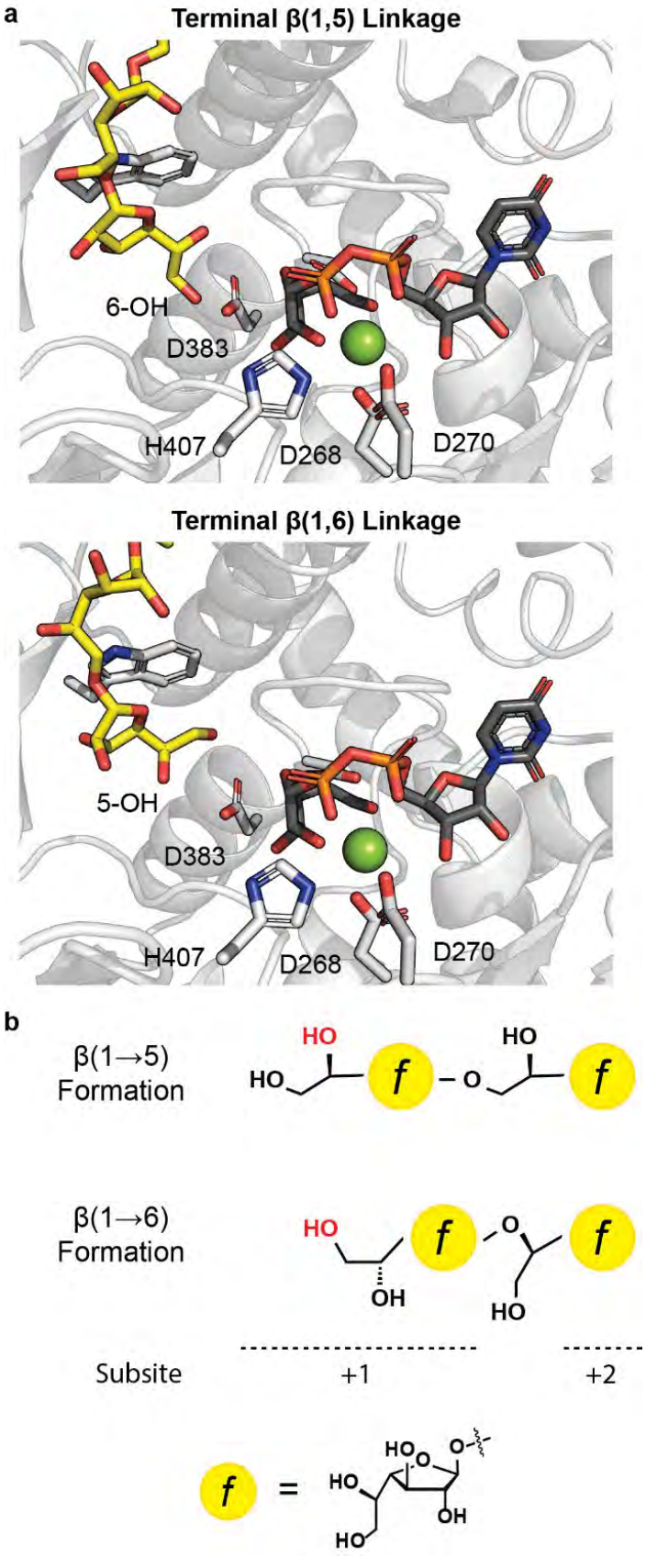

Figure 2.17: Linkage-Dependent Regioselectivity. a, Proximity of trisaccharide and tetrasaccharide galactan acceptors (yellow) to a modeled UDP-Gal*f* donor (dark gray); donor placement was guided by PDB 4FIY.[9] The reacting acceptor hydroxyl is labeled, along with the catalytic aspartate D383 and the DDX motif. **b,** Model for regioselectivity and bifunctional glycosyltransferase activity by GlfT2, in which common binding to subsite +2 mediates linkage control through a linkage-length mechanism. The hydroxyl-group involved in catalysis is shown in red.

flexibility of the longer linkage.

Comparing the coordination of the exocyclic diol of the reacting Gal*f*, **Compound 2** coordinates the 6-OH to the catalytic residue D383, while **Compound 3** presents the 5-OH. Modeling the donor UDP-Gal*f* in the active site reveals proximity of the reacting hydroxyls to the anomeric position of the donor (Figure 2.17a). This differential hydroxyl coordination agrees with the alternating pattern production expected.

We hypothesize that common binding to subsite +2 influences the regioselectivity

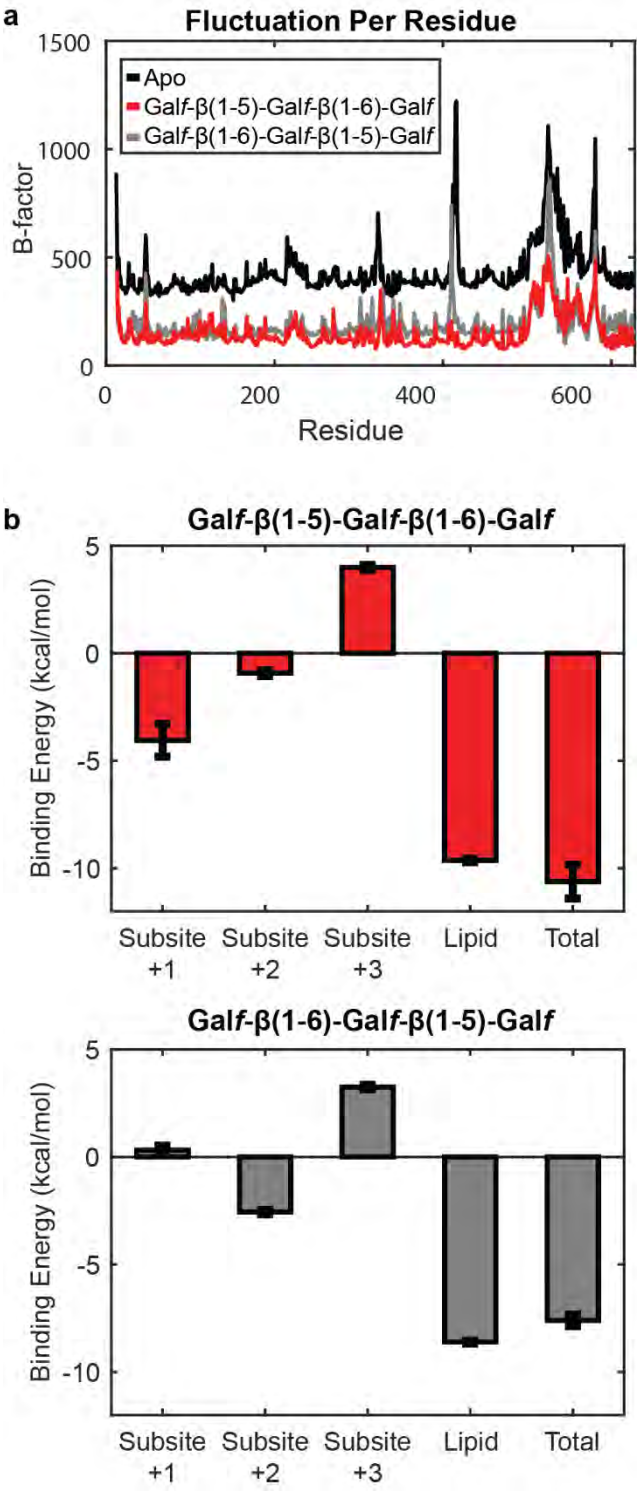

Figure 2.18: Molecular Dynamics Simulation of Acceptor-Bound Complexes. Simulations were performed on monomeric *Nbr*GlfT2 as the apo enzyme or in complex with a Gal*f*-β(1,5)-Gal*f*-β(1,6)-Gal*f*-R or Gal*f*-β(1,5)-Gal*f*-β(1,6)-Gal*f*-R trisaccharide. a) Per-residue fluctuation over the course of the simulations, represented as residue B-factor. b) Energy-per-subsite for each complex, defined by the interacting sugar (kcal mol^{-1}). Apo – black; Gal*f*-β(1,5)-Gal*f*-β(1,6)-Gal*f* – red; Gal*f*-β(1,6)-Gal*f*-β(1,5)-Gal*f* – grey; Lipid - dodec-2-enyl.

of GlfT2 catalysis. Similar positioning within subsite +2 would permit length differences from the previous linkage to dictate hydroxyl presentation. This model suggests a feedback-regulated sequence control mechanism for GlfT2 regioselective and linkage pattern fidelity (Figure 2.17b).

2.3.8 Molecular Dynamics of Acceptor-Bound *Nbr*GlfT2

To evaluate the contributions of binding site residues to acceptor stabilization, molecular dynamics (MD) simulations were performed on monomeric *Nbr*GlfT2, including the *apo* enzyme and two acceptor-bound *holo* enzymes. The acceptor-bound complexes contained either a Gal*f*-β(1,5)-Gal*f*-β(1,6)-Gal*f* (6-OH reactive) or Gal*f*-β(1,6)-Gal*f*-β(1,5)-Gal*f* (5-OH reactive)

glycan component, and a dodecyl lipid aglycon. Values for all-atom, per-residue B-factors reveal comparable stabilization of the *Nbr*GlfT2 monomer when bound to either acceptor (Figure 2.18a). Notably, the active site loop (residues 402-412) becomes highly stabilized when bound to the acceptor, except for W410 which remained highly flexible when complexed with the 5-OH reactive acceptor. This is similar to the observed flexibility in the final β(1,6) linkage in the experimental structure of *Nbr*GlfT2 + **Compound 3**, suggesting the bound conformation of the final Gal*f* of **Compound 3** in subsite +1 does not become fully stabilized in the active site.

Energy decomposition analysis (MM/GBSA, see Methods) for both simulated acceptor complexes reveal an overall stabilizing energy for binding, driven significantly by the lipid aglycon (Figure 2.18b, c). Binding enthalpies for the active site Gal*f* in the Gal*f*-β(1,5)-Gal*f*-β(1,6)-Gal*f* complex contributed to glycan stabilization, with additional contribution from the subsite +2 Gal*f* residue. A similar trend was determined for the Gal*f*-β(1,6)-Gal*f*-β(1,5)-Gal*f* simulation, except the active site Gal*f* was slightly destabilizing, and the subsite +2 Gal*f* had a higher contribution to binding stabilization.

Subsite +1 stabilization in both complexes includes residues E312, W321, K380, D383, and W410 (Table 2.2). A shift in relative contributions within subsite +1 is seen between both complexes, where the Gal*f*-β(1,5)-Gal*f*-β(1,6)-Gal*f* binding energy is contributed to by E312 much more than D383, while the opposite is seen for Gal*f*-β(1,6)-Gal*f*-β(1,5)-Gal*f*, likely arising from the difference in insertion depth and positioning of the final Gal*f* in the active site. Binding energy contribution from van der Waals interactions with W321 and W410 are substantial in both complexes, with significant difference in W410, likely arising from the higher flexibility of the Gal*f*-β(1,6) active site complex. Tryptophan contributions in these complexes agree with previous observations of aromatic amino acid enrichment in sugar binding sites of proteins.[68]

Table 2.2: Residue-Ligand Interaction Energies from MD Simulations (<-0.5 kcal mol^{-1})

Gal*f*-β(1→5)-Gal*f*-β(1→6)-Gal*f*

Ligand	Residue	van der Waals (kcal mol^{-1})	Electrostatic (kcal mol^{-1})	Polar Solvation (kcal mol^{-1})	Non-polar Solvation (kcal mol^{-1})	Total (kcal mol^{-1})	Standard Error (kcal mol^{-1})
Galf +1	GLU 312	0.02	-18.29	11.59	-0.90	-7.58	0.11
	LYS 380	-0.26	-4.20	-1.11	-0.85	-6.43	0.13
	TRP 321	-3.02	-0.38	-0.28	-1.89	-5.57	0.04
	TRP 410	-2.51	-1.08	0.09	-1.74	-5.23	0.03
	ASP 383	0.13	-6.30	4.51	-0.27	-1.93	0.07
	TRP 359	-0.53	0.02	0.03	-0.45	-0.92	0.01
Galf +2	ASP 415	-0.02	-15.92	8.75	-0.82	-8.00	0.09
	TRP 321	-1.54	-1.20	0.19	-1.13	-3.68	0.04
	TRP 410	-1.64	0.71	-0.36	-1.17	-2.46	0.03
	ALA 416	-0.88	0.05	-0.10	-0.70	-1.63	0.01
	HID 424	-0.83	1.39	-0.93	-0.53	-0.90	0.02
Galf +3	ALA 416	-0.64	-4.53	1.71	-1.07	-4.54	0.74
	GLN 466	-0.35	-3.75	1.29	-0.66	-3.47	1.62
	HID 424	-1.19	0.02	-0.38	-0.83	-2.38	0.64
	HIE 454	-0.51	-0.80	-0.32	-0.65	-2.28	0.70
	TRP 321	-0.42	-0.57	0.33	-0.22	-0.87	0.32
	ALA 450	-0.44	0.16	-0.03	-0.43	-0.74	0.26
Lipid Aglycon	LYS 453	-1.71	-0.46	0.31	-1.64	-3.50	0.72
	LYS 449	-1.67	0.24	-0.25	-1.43	-3.11	0.88
	TRP 564	-1.37	0.06	-0.08	-1.14	-2.53	0.97
	ALA 450	-1.03	-0.04	-0.06	-0.84	-1.96	0.46
	LEU 534	-0.66	0.01	-0.01	-0.84	-1.50	0.53
	LEU 452	-0.74	0.00	0.00	-0.63	-1.37	0.69
	HIE 454	-0.34	-0.45	0.28	-0.13	-0.64	0.39
	LEU 456	-0.30	-0.02	0.02	-0.30	-0.60	0.40

Gal*f*-β(1→6)-Gal*f*-β(1→5)-Gal*f*

Ligand	Residue	van der Waals (kcal mol^{-1})	Electrostatic (kcal mol^{-1})	Polar Solvation (kcal mol^{-1})	Non-polar Solvation (kcal mol^{-1})	Total (kcal mol^{-1})	Standard Error (kcal mol^{-1})
Galf +1	ASP 371	0.41	-15.12	10.02	-0.66	-5.34	0.12
	TRP 309	-2.48	-0.35	-0.24	-1.71	-4.79	0.03
	GLU 300	-0.49	-10.49	8.02	-0.77	-3.72	0.09
	TRP 398	-1.28	-0.73	-0.02	-0.94	-2.96	0.04
	LYS 368	-0.77	1.69	-3.14	-0.67	-2.88	0.09
	TRP 347	-0.41	-0.01	-0.02	-0.39	-0.83	0.01
	HIE 296	-0.14	0.14	-0.48	-0.19	-0.66	0.04
Galf +2	ASP 403	-0.63	-13.14	8.64	-0.79	-5.91	0.06
	TRP 398	-2.30	0.38	0.00	-1.61	-3.54	0.04
	TRP 309	-1.46	-0.63	0.08	-1.06	-3.06	0.02
	ALA 404	-0.64	0.03	-0.08	-0.52	-1.20	0.01
	LYS 368	-0.29	0.36	-0.86	-0.31	-1.09	0.04
	HIE 412	-0.37	0.08	-0.31	-0.21	-0.80	0.03
Galf +3	ALA 404	-1.13	-0.51	0.31	-0.97	-2.31	0.02
	HIE 412	-0.91	-0.55	0.28	-0.86	-2.04	0.04
	TRP 309	-0.77	-0.50	0.39	-0.47	-1.34	0.03
	LYS 368	-0.13	-2.63	2.12	-0.08	-0.72	0.04
	HID 442	-0.19	-0.37	0.12	-0.13	-0.58	0.04

Lipid Aglycon	LYS 437	-1.75	0.13	-0.10	-1.53	-3.25	0.03
	LYS 441	-1.68	0.95	-0.80	-1.63	-3.16	0.03
	TRP 552	-1.32	0.05	-0.07	-1.20	-2.53	0.04
	LEU 522	-0.82	0.00	0.00	-0.92	-1.74	0.02
	ALA 438	-0.91	-0.05	0.03	-0.75	-1.68	0.02
	LEU 440	-0.71	0.01	0.01	-0.53	-1.22	0.03

Subsite +2 stabilization included D415, A416, W321, and W410. Interactions between D415 and the cyclic diol of the penultimate Gal*f* residue in both complexes was the major contributor within this subsite. Both W321 and W410 provided similar stabilization energies between both complexes. Subsite +3 binding was driven in both simulations by A416, with an additional contribution by H424. Gal*f* stabilization within subsite +3 was low throughout the simulation for both complexes.

2.4 Discussion

Based on the structures of NbrGlfT2 bound to its galactan acceptor, we propose a model for processive sequence-controlled polymerization by GlfT2 (Figure 2.19). GlfT2 uses three major subsites within its acceptor-binding cleft to bind and retain the galactan acceptor during elongation. Upon binding the acceptor, the flexible active-site loop undergoes a disordered-to-ordered transition to promote glycan retention. The acceptor-binding subsites accommodate the galactan regardless of glycosidic linkage pattern; this promiscuity allows the acceptor to slide between these subsites during elongation. Acceptor sliding facilitates GlfT2 processivity. During sliding, subsite +2 stabilizes the furanose ring of the penultimate Gal*f* residue, allowing subsite +2 to influence the orientation of the final Gal*f* in the active site through length differences between the β(1,5) and β(1,6) linkages. The influence of subsite +2 produces a feedback-

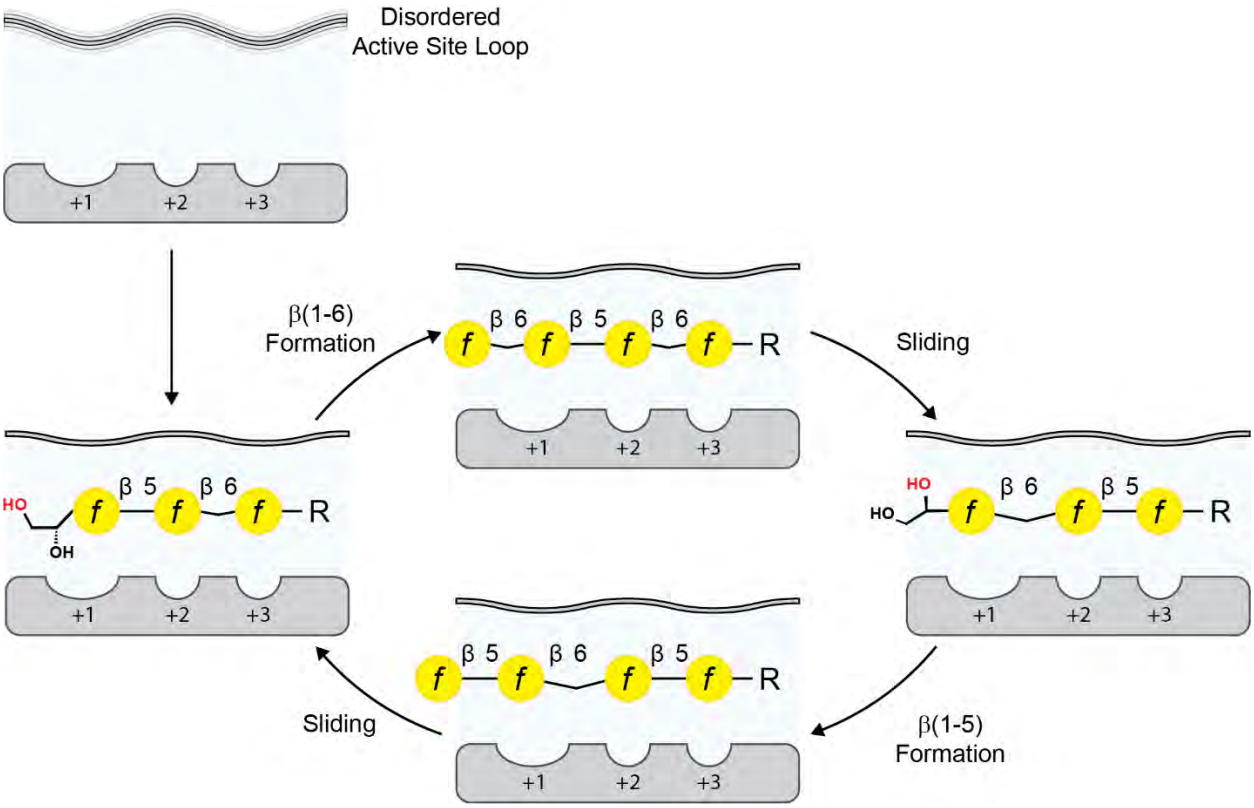

Figure 2.19: Model for Processive Bifunctional Polymerization by *Nbr*GlfT2. A schematic depicting subsite-mediated processive sliding of the galactan during polymerization. The flexible active-site loop is disordered in the apo-state, and transitions to an ordered structure upon acceptor binding. During polymerization, the acceptor remains bound to GlfT2, and slides between each subsite after each Gal*f* addition. Subsite +2 dictates presentation of the non-reducing end 5-OH or 6-OH for regioselective bifunctional catalysis.

regulated sequence control, permitting GlfT2 to produce an alternating linkage pattern while retaining the elongating galactan during processive polymerization.

Energy decomposition analysis from MD simulations of bound acceptors suggest the *in vitro* association of the acceptor to GlfT2 is driven by the lipid moeity at the reducing end of the glycan. This is in agreement with a previously observed correlation between lipid aglycon sioze to galactan polymerization extent.[19] Energies for the acceptor glycan suggest subsite +3 to be destabilizing, and may influence the ability for the acceptor to unbind at this site and slide during polymerization. The 5-OH reactive acceptor was determined to be less stabilizing compared to

the 6-OH reactive acceptor, potentially in agreement with product distributions from both *Nbr*GlfT2 and *Mtb*GlfT2, which were observed to enrich slightly for Gal*f*-β(1,5) products (Figure 2.2b, 2.15).[10,19,45,53]

The proposed processive sliding model agrees with previous observations of GlfT2 activity. Galactan acceptor retention during elongation was observed when products were measured after short reaction times and under single-hit conditions, suggesting galactan sliding permits this retention during polymerization.[19,45] GlfT2 achieves this processivity by sliding the galactan between three subsites. These subsites agree with an observed initial lag phase for the disaccharide acceptor (**Compound 1**), suggesting the acceptor either preferentially binds subsites +2 and +3, or does not have sufficient association strength to initiate polymerization.[10] Electron density for the lipid moeity of the synthetic acceptor near the hydrophobic membrane-associating motif suggests a potential two-site association *in vitro*. This is in agreement with our previous *in vitro* tethering model for GlfT2 activity, in which lipid binding influences galactan polymerization extent.[19]

GlfT2 was shown to produce its alternating pattern with high fidelity, recognizing only the final glycosidic linkage, while having promiscuity for other linkages throughout the acceptor.[10] This agrees with the proposed feedback-regulated sequence control model, where subsite +2 dictates regioselectivity, irrespective of the linkage between subsites +2 and +3. The catalytic importance of recognition by subsite +2 is highlighted by the requirement of two Gal*f* residues on the acceptor to initiate catalysis, although a monosaccharide acceptor has been show to function as a substrate.[19,71] *In vivo*, this requirement is fulfilled by GalactofuranosylTransferase 1 (GlfT1), which primes the Rha-GlcNAc-decaprenylpyrophosphate acceptor with 2-3 Galf residues to

initiate GlfT2 polymerization.[16] This provides an additional checkpoint of control within the galactan biosynthetic pathway.

Acceptor binding to *Nbr*GlfT2 revealed an ordering transition for the GlfT2 active site loop. This loop interacts with the acceptor glycan at all three subsites. Flexibility of the active site loop may be involved in promoting processive polymerization, maintaining partial interactions with the sliding acceptor, retaining the bound complex while reducing the energetic barrier to translation. Previous investigations of GT-A enzymes suggested that the active site loop could additionally be involved in ordered binding of the donor and acceptor substrates.[31] Additional structural and biochemical studies are needed to inform the role of the active site loop in GlfT2 catalysis.

GlfT2 was previously hypothesized to mediate its galactan length control by filling a cavity formed above the membrane interface with the full length galactan. The structures reported here indicate the growing galactan enters the cavity inside the tetramer complex. Comparison of cavity volumes between *Nbr*GlfT2 and *Mtb*GlfT2 suggests a larger cavity for *Nbr*GlfT2, predicting a longer galactan product. This disagrees with the galactan lengths seen *in vitro* and *in cellulo* for *Nbr*GlfT2. However, a recently solved structure of the wall teichoic acid polymerase TarL revealed a membrane-associating tetramer with C_4-symmetry that forms a cavity similar to GlfT2 above the membrane (EMD-43084).[72] TarL uses a catalytic GT-B domain, structurally distinct from the GT-A domain of GlfT2. This suggests the evolutionary convergence of these two enzymes to a common quarternary structure, indicating a possibly significant and poorly understood role for the cavity in GT polymerase activity.

Informed by AlphaFold structure predictions and FoldSeek structural alignment, a number of protein sequences were predicted to have structures highly similar to *Nbr*GlfT2.[28,29,73] Accession O3I_000895 within *N. brasiliensis* was predicted to fold similarly to *Nbr*GlfT2, with a distinct

acceptor-binding region. This protein likely constitutes a paralog with distinct activity and linkage production compared to *Nbr*GlfT2, possibly influencing previous linkage analysis of the *N. brasiliensis* galactan.[42] This paralog does not have an ortholog in *M. tuberculosis* or other mycobacteria, but appears to be present throughout the Mycobacteriales families Nocardiaceae and Gordoniaceae, and may influence unique glycan structures in non-mycobacterial organisms throughout Mycobacteriales.

Predicted GlfT2-like proteins appear to be present outside of Mycobacteriales. A recent publication described the biosynthetic gene cluster responsible from T1-antigen production in *Salmonella enterica* (strain SL3770). The polysaccharide product of this gene cluster is formed with a core repeat-unit containing alternating Gal*f*-β(1,6) and Gal*f*-β(1,3) glycosidic linkages. The enzyme responsible, STM0724, was described to be an ortholog of GlfT2 from mycobacteria. The predicted structure is highly similar to *Nbr*GlfT2, and suggests a similar activity. Future investigations into the activities and regioselectivities of STM0724 and other galactofuranosyl transferases outside of Mycobacteriales will certainly be informed by the structures and mechanistic models described for herein for *Nbr*GlfT2.

In conclusion, we describe here the structures of galactan acceptor-bound GlfT2. These results reveal a binding site spanning three Gal*f* units that facilitates polymerization and sequence control. We proposed two models for GlfT2 activity: processive polymerization facilitated by acceptor sliding, and linkage pattern control mediated by linkage-length feedback regulation. Insights from these structures are expected to inform efforts towards GlfT2 inhibitor design, an attractive target for drug development against pathogenic members of Mycobacteriales. Broadly, this work and the models proposed are expected to further our understanding of processive

polymerization performed by glycosyltransferases, and guide future studies into their mechanisms of action.

2.5 Methods

2.5.1 Cloning and Mutagenesis

The *Nbr*GlfT2 (O3I_000690 in strain HUJEG-1) expression vector was produced using a gene block of the *Nbr*GlfT2 gene (IDT DNA), codon optimized for expression in *E. coli*. Isothermal assembly was performed to insert the gene block into a Nde1 (New England Biolabs) linearized pET28a backbone. Phosphorylation by T4 Polynucleotide Kinase (New England Biolabs) and ligation by Fermentas T4 DNA Ligase (ThermoScientific) joined the two linear fragments. Plasmid amplification was performed in DH5α *E. coli* growing with LB broth in the presence of kanamycin (50 μg/mL). Plasmid was purified using the PureYield Miniprep Kit (Promega).

Table 2.3: Primers Used for *Nbr*GlfT2 Mutagenesis

Primer Name	Primer Sequence
D377A *Nbr*GlfT2 F	CTGTGGAATACGGCCTGCG
D377A *Nbr*GlfT2 R	CATCCCACTTCAGAAACAAAGGCAG
D262A *Nbr*GlfT2 F	GCTGATGATATCGAGATCGAGCC
D262A *Nbr*GlfT2 R	CATATACACAATATATTCAGCGTCCGTATTC
D409A *Nbr*GlfT2 F	CCTGGAGCGATAAGGACGCCGCTATTGATTGGCAGGCGTATTTCC
D409A *Nbr*GlfT2 R	GTCCTTATCGCTCCAGGCCATATGCCAAACCGCCGC
H418A *Nbr*GlfT2 F	TGATTGGCAGGCGTATTTCGCTCTGCGCAACCGTCTGGTGG
H418A *Nbr*GlfT2 R	GAAATACGCCTGCCAATCAATAGCGTCGTCCTTATCGCTCC

Plasmid mutations were introduced using the *In Vivo* Assembly method.[74] Primers were designed to include the point mutation of interest and a 5' homologous linear region (IDT DNA, Table

2.3). PCR was performed with the Q5 High-Fidelity Polymerase (New England Biolabs) and the linear products were chemically transformed into DH5α *E. coli* for product ligation and amplification. Mutation success was confirmed by sanger sequencing (Quintara Biosciences).

2.5.2 *Nbr*GlfT2 Expression and Purification

Protein production was performed using DE3 *E. coli* with chemical transformation of the expression plasmid. Protein was expressed using an autoinduction protocol, using terrific broth supplemented with 150 ug/mL of kanamycin as base media.[75,76] Cells were grown on a shaker in 0.5 L Terrific Broth and 150 μg/mL of kanamycin to an optical density (600 nm) of 0.8 at 37 ˚C, and cooled to 18 ˚C for ~20 hour protein expression. Cells were pelleted by centrifugation at 4,000 xg for 10 minutes at 4 ˚C, and the pellets stored at -80 ˚C.

Cells were resuspended in a 1:1 dilution of 2x Bacterial Protein Extraction Reagent (Thermo Fisher Scientific) and 2x Buffer A (50 mM HEPES pH 7.8, 300 mM NaCl, 20 mM imidazole) at 5 mL/g, along with 1 mg/mL Egg White Lysozyme (Gold Biotechnology), 0.1% Triton X-100 (Sigma Aldrich), and 5 units of Benzonase Nuclease (EMD Millipore). The suspension was incubated for 30 minutes at room temperature with gently mixing, and centrifuged at 24,000 xg for 1 hour at 4 ˚C. The supernatant was filtered with a Millex-GP 0.22 μm PES Syringe Filter (MilliporeSigma), and injected onto a HisTrap HP column (Cytiva). Purification was accomplished using a NGC FPLC (Biorad) at 4 ˚C with a 0% to 100% gradient of Buffer A and Buffer B (50 mM HEPES pH 7.8, 300 mM NaCl, 500 mM imidazole). Purified protein was pooled and dialyzed against Buffer C (50 mM HEPES pH 7.4, 100 mM NaCl) with 5 mM EDTA and 5mM BME using a 20k Da molecular weight cutoff Slide-A-Lyzer Dialysis Cassette (Thermo Fisher Scientific). Dialysis was then moved to Buffer C + 5% glycerol. Protein was flash frozen in liquid nitrogen and stored at -80 ˚C.

2.5.3 GlfT2 Activity Assay and Product Length Analysis by MALDI-TOF MS

Elongation reactions were performed in 10 µL volumes containing 0.2 µM GlfT2, 1.25 mM UDP-Gal*f*, and 200 µM acceptor substrate in 50 mM HEPES pH 7.4 and 100mM NaCl. Reactions proceeded overnight at room temperature. The reaction mixtures were mixed to contain 0.05% trifluoroacetic acid (TFA) and loaded onto Hypersep Hypercarb Tips (10 µL, Thermo Fisher Scientific). The tips were washed with 0.05% TFA in 1:3 acetonitrile:water, and then eluted in 70% methanol saturated with α-cyano-4-hydroxycinnamic acid. Mass spectra of galactan products were acquired using a Bruker Autoflex LRF Speed MALDI-TOF in positive reflectron mode.

2.5.4 *In Cellulo* Extraction of Arabinogalactan from Msm Δ*glft2* Strain

Arabinogalactan was extracted from *Mycobacterium smegmatis* as previously described.[37] Briefly, Msm mc^{2}155 strains complemented with *Nbr*GlfT2 in pAMJ7 were grown in 7H9 media (Remel) supplemented with 0.2% glycerol, 0.2% glucose, 0.5% bovine serum albumin (US Biological), 4 mg/L catalase, 15 mM sodium chloride, and 0.05% Tween-80. Cells were grown to mid-log phase in 50 mL cultures, and harvested by centrifugation (4000 RCF for 10 mins). Cell pellets were decanted and stored at -80 ˚C.

After thawing on ice, cell pellets were resuspended in 50 mL phosphate buffered saline (PBS) supplemented with 2% Triton-X 100. Cell lysis was performed at 22,000 PSI with a benchtop cell disruptor from Constant Systems (Kennesaw, GA), and the lysate was clarified by centrifugation (20,000 RCF for 1 hour). Supernatant was discarded and insoluble pellet was extracted three times with 2% SDS in PBS for one hour at 95 ˚C. Following the third extraction,

mAGP material was washed with milli-Q water, 80% acetone in water, and 100% acetone. Residual acetone was allowed to evaporate overnight.

2.5.5 Galactan Linkage Analysis

Linkage analysis of both desalted *in vitro* extension products of **Compound 1** and extracted *in cellulo* arabinogalactan from *Nbr*GlfT2-complemented Msm Δ*glft2* were performed as previously described.[77] Briefly, the samples were permethylated by two treatments with sodium hydroxide (15 minutes) and methyl iodide (45 minutes). The permethylated material was hydrolyzed using 2 M trifluoroacetic acid (TFA, 400 µL, 2 hours at 121 ˚C in a sealed tube). The hydrolyzed material was reduced with $NaBD_4$ and acetylated with a 52:48 mix of acetic anhydride:TFA (25 minutes at 45 ˚C). The resulting products were analyzed with an Agilent 7890A Gas Chromatograph interfaced with a 5975C Mass Selective Detector (electron impact ionization mode). Separatation was performed on a 30 m Supelco SP-2331 bonded phase fused silica capillary column.

2.5.6 *Nbr*GlfT2 Crystallography

Protein samples were diluted with Buffer C1 with Triton X-100 to final concentrations of 6-7 mg/mL protein and 0.02% Triton X-100. Samples used for ligand soaking were additionally dialyzed for 20 minutes against Buffer C + 0.02% Triton X-100 to remove glycerol. The crystallization condition composition was 0.06-0.1 M imidazole, 2 mM MgCl2, and 7.5-8.75% each of PEG 3350 (w/v), PEG 1000 (w/v), and MPD (v/v; Molecular Dimensions). Crystallization drops were set with the hanging drop method as 2 µL of 1:1 protein:mother liquor. Crystals formed within 1-7 days with a plate morphology, and were harvested at most a month after forming. Crystals were harvested and flash frozen in liquid nitrogen. For ligand

soaking, crystals were transferred to drops made of a 1:1 mixture of 20 mM Compound 2 or Compound 3 and mother liquor with 10% more precipitant mixture, and incubated overnight prior to harvesting and flash freezing.

2.5.7 Crystal Diffraction, Data Phasing and Modeling

Diffraction data for *Nbr*GlfT2 (Apo) was collected at the Advanced Photon Source (APS, Argonne National Laboratory) using the GM/CA beamline 23-ID-B at a wavelength of 1.033 Å. The acceptor bound complexes *Nbr*GlfT2 (**Compound 2**) and *Nbr*GlfT2 (**Compound 3**) were at the NSLSII (National Synchrotron Light Source II, Brookhaven National Laboratory) using the AMX beamline 17-ID-1 at a wavelength of 0.920 Å. The protein crystals diffracted to 3.0 Å and were in the space group $P2_12_12_1$, containing eight protomer chains of *Nbr*GlfT2. The structure was solved by molecular replacement in Phaser using an Alphafold-prediction of *Nbr*GlfT2 for apo *Nbr*GlfT2, and then the experimental structure for all liganded datasets.[28,78] The initial atomic model was refined in PHENIX.[79] Manual real-space refinement was performed in Coot 0.9.8.95 (glycan refinement required Coot 0.9.6). The refined protein model was validated using MolProbity.[80] Protein structure visualization performed in PyMOL.[81] Protein oligomer interfaces were calculated with the PISA server.[64]

Galactofuranose ligand constraints were optimized in the electron Ligand Builder and Optimization Workbench.[82] Linkage constraints for the β(1,5) and β(1,6) glycosidic linkages were generated in CCP4 AceDRG.[83]

Ramachandran statistics for the protein models were as follows (Favored, Allowed, Outliers): Apo – 94.9%, 4.2%, 0.9%; **Compound 2** – 94.9%, 4.7%, 0.4%; **Compound 3** – 95.3%, 4.2%, 0.5%.

Table 2.4: Residues Frozen for MD Simulation. Residues in the frozen region that are restrained during the minimization and equilibration steps. The atoms frozen in each residue are specified. Residues 629 and 630 have different names based on the glycosidic linkages in the trisaccharide. The 6-OH reactive trisaccharide has residues 6LU 629 and 5LU 630, and the 5-OH reactive trisaccharide has residues 5LU 629 and 6LU 630.

Macromolecule	Residue Name	Residue Number	Atoms Frozen
Protein	MET	320	Backbone atoms (C, Ca, O, N, H)
Protein	TRP	321	Backbone atoms
Protein	TRP	410	Backbone atoms
Protein	ASP	415	Backbone atoms
Protein	ALA	416	Backbone atoms
Protein	ILE	417	Backbone atoms
Protein	HID (His)	424	Backbone atoms
Protein	ARG	428	Backbone atoms
Protein	HIE (His)	454	Backbone atoms
Protein	THR	462	Backbone atoms
Protein	GLN	466	Backbone atoms
Alkane tail	F01	640	All C atoms
Carbohydrate	5LU/6LU	641	All C and O atoms
Carbohydrate	6LU/5LU	642	All C and O atoms
Carbohydrate	0LU	643	All C and O atoms

2.5.8 Molecular Dynamics Simulations

We used the X-ray crystal structures as the starting point for simulations of apo GlfT2 and GlfT2 bound to one of two trisaccharide ligands. Protonation states were assigned using the H++ web server with an internal dielectric constant of 10.0, a pH of 7.0, and all other default settings

Table 2.5: Residues for which a Pairwise Energy Decomposition was Computed. Residues 629 and 630 have different names based on the glycosidic linkages in the trisaccharide. The 6-OH reactive trisaccharide has residues 6LU 629 and 5LU 630, and the 5-OH reactive trisaccharide has residues 5LU 629 and 6LU 630.

Macromolecule	Residue Name	Residue Number
Protein	HIE (His)	308
Protein	MET	312
Protein	TRP	321
Protein	TYR	329
Protein	TRP	359
Protein	LYS	380
Protein	ASP	383
Protein	TRP	410
Protein	ASP	415
Protein	ALA	416
Protein	HID (His)	424
Protein	LYS	449
Protein	ALA	450
Protein	LEU	452
Protein	LYS	453
Protein	HIE (His)	454
Protein	LEU	456
Protein	GLN	466
Protein	LEU	534
Protein	TRP	564
Alkane tail	F01	640
Carbohydrate	5LU/6LU (Gal*f*)	641
Carbohydrate	6LU/5LU (Gal*f*)	642
Carbohydrate	0LU (Gal*f*)	643

applied.[84] The residue labels for the ligands were manually updated to match the GLYCAM_06 force field naming conventions.[85] The AMBER LEaP program was used to solvate the protein in

a water box consisting of at least 10 Å initial buffer around all dimensions of the protein.[86] LEaP

was also used to neutralize the system using 1 sodium ion (holo systems) or 1 chloride ion (apo system) and generate final topology and coordinate files. We used forcefields ff14SB for the protein and TIP3P for water.[87,88] We added galactofuranose ligand parameters prepared according to GLYCAM_06j-1 forcefield conventions and we generated forcefield parameters for the alkane tail of the ligand using the RESP charges from the REDS server according to the procedure for creating a GLYCAM 2006 type forcefield (see

Supporting Information data file).[89–91] The final system sizes were 75453, 75968, and 74672 atoms for apo GlfT2, GlfT2 bound to the 5-OH reactive trisaccharide ligand, and GlfT2 bound to the 6-OH reactive trisaccharide ligand, respectively.

Molecular dynamics (MD) simulations were performed with AMBER18 using the GPU-accelerated particle mesh Ewald (PME) molecular dynamics (PMEMD) module.[86,92,93] Each system was equilibrated using the following procedure. We ran a 2000 cycle minimization of the water molecules and hydrogen atoms with a restraint applied to protein and ligand heavy atoms and a 2000 cycle minimization of all atoms with a restraint applied to the frozen region for the holo systems. All restraints have a 200 kcal $mol^{-1}Å^{-2}$ weight and the 'frozen region' includes ligand heavy atoms and the backbone atoms of nearby protein residues (Supporting Information Table 2.3). Next, we performed a 10 ps controlled NVT equilibration to heat the system from 100K to 300K and a 1 ns NpT equilibration using the Berendsen barostat with a pressure relaxation time of 1 ps. Both steps used a Langevin thermostat for temperature control with a collision frequency of 5.0 ps^{-1}, a 2 fs timestep, periodic boundary conditions, the SHAKE algorithm for constraining hydrogen atom bond lengths, and restraints on the frozen region of the holo systems.[94] Finally, a 200 ns NpT production run was performed with no restraints to observe the dynamics of each system and all other settings the same as the NpT equilibration. This full MD procedure was carried out three times for each system.

The three production runs of each system were combined using the AMBER CPPTRAJ trajectory processing program.[95] Each concatenated trajectory was clustered using the k-means algorithm implemented in CPPTRAJ, with k=5, to identify the major structural conformations of the equilibrated systems. The resulting trajectory of the predominant cluster for each holo system

was used as input for molecular mechanics with generalized born and surface area solvation (MM-GBSA) evaluation of binding interaction energetics.[96] The MM-GBSA calculations were performed using the AMBER MMPBSA.py script with the OBC GB model, salt concentration of 0.1.[97,98] Total interaction energies are decomposed into pairwise per-residue energies for relevant residues with 1-4 terms added to internal potential terms (Table 2.5). Trajectory snapshots were spaced at least 400 ps apart, totaling 657 frames for the system bound to the 5-OH reactive trisaccharide ligand and 563 frames for the system bound to the 6-OH reactive trisaccharide ligand.

2.6 Acknowledgements

We thank the Structural Biology Core Facility at MIT for their help in initial protein crystallization and the Department of Chemistry Instrumentation Facility for their assistance in collection of MALDI-TOF data. We acknowledge the beamlines at APS and NSLS-II for their assistance in X-ray diffraction collection. GM/CA@APS has been funded by the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006, P30GM138396). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. The Eiger 16M detector at GM/CA-XSD was funded by NIH grant S10 OD012289. This research used resources of the National Synchrotron Light Source II, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under Contract No. DE-SC0012704. The Center for BioMolecular Structure (CBMS) is primarily supported by the National Institutes of Health, National Institute of General Medical Sciences (NIGMS) through a Center Core P30 Grant (P30GM133893), and by the DOE Office

of Biological and Environmental Research (KP1605010). The Complex Carbohydrate Research Center (CCRC) is supported by U.S. Department of Energy grant (DE-SC0015662) to Parastoo Azadi. We thank Sara Porfirio, Ian Black, and Parastoo Azadi at the CCRC for technical assistance

Funding: Research reported in this publication was supported by the National Institute of General Medical Sciences and the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number F31GM148069 (to A.W.C.) and R01AI126592 (to A.W.C., A.K, A.M.J., K.I.T., L.L.K.).

2.7 Author Information

Alan W. Carter[1], Allison M. Keys[1,2,3], Alexander M. Justen[1], Katherine I. Taylor[1], Greg J. Dodge[4], Barbara Imperiali[1,4], Heather J. Kulik[1,3], Laura L. Kiessling[1,4]

Affiliations:

[1]Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA – A.W.C., A.M.J., A.M.K., K.I.T., C.L.D, B.I., H.J.K, L.L.K.

[2]Program in Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA – A.M.K, L.L.K, H.KJ.K

[3]Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA – A.M.K., H.J.K.

[4]Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA – G.J.D., B.I., C.L.D

Chapter 3

Structural Characterization of GlfT2 at the Membrane Interface

Contributions: Conceptualization by Alan W. Carter and Laura L. Kiessling. Protein expression and purification was performed by Alan W. Carter. Micrograph collection was performed by Alan W. Carter. Image processing, volume reconstruction, and model refinement was performed by Greg J. Dodge and Alan W. Carter. Molecular dynamics simulations were performed by Alan W. Carter.

3.1 Abstract

Polysaccharides play crucial roles, including cell-cell communication, immune recognition and evasion, energy storage, and cell wall integrity. The biosynthesis of many essential polysaccharides is controlled by polymerizing glycosyltransferases (GTs). Galactofuranosyl transferase 2 (GlfT2) is a polymerizing GT responsible for producing the galactan polysaccharide. The galactan is a vital component of the cell wall structure of the Mycobacteriales order, a group containing key human pathogens such as *Mycobacterium tuberculosis*. GlfT2 is a membrane-associated enzyme that was previously shown to perform processive polymerization to synthesize the galactan. Recent studies have identified mechanisms governing processivity and linkage control, but polymer length regulation remains poorly understood. Here, we describe the Cryo-EM structure of GlfT2 bound to a membrane nanodisc, revealing three amphipathic helices that mediate membrane association and induce membrane interface curvature. Based on these findings, we propose a model in which GlfT2-induced membrane curvature facilitates polymer translocation during elongation. We believe this model will guide future investigations into the polymerization mechanisms of GlfT2 and other critical polymerizing GTs.

3.2 Introduction

Polysaccharides are sugar polymers that play a variety of critical roles across biology. These roles are diverse, and include cell-cell communication, immune recognition and evasion, energy storage, and cell wall integrity. The specific roles performed by polysaccharides are dictated by their monosaccharide constituents, composition pattern, and polymer length. These properties are controlled by the enzymes responsible for their biosynthesis. Therefore, maintaining proper function of these enzymatic activities is critical to many essential processes throughout cellular life.

To produce polysaccharides with distinct polymeric properties, many glycosyltransferases (GTs) polymerize their polysaccharides processively, retaining the substrate throughout its elongation, releasing the final full-length product.[44] This processivity permits polymerases to dictate the length and composition of the final product upon release. Various structural features have been proposed to facilitate processive polymerization, including polymer sliding and substrate enclosure.[67] Understanding the mechanisms that underlie the functions of processively polymerizing GTs will provide critical insights into how polysaccharide composition is controlled.

Galactofuranosyl transferase 2 (GlfT2) is a polymerizing GT that is responsible for the production of the bulk of the Mycobacteriales galactan.[19,35] The galactan is a polysaccharide of galactofuranose (Gal*f*) 20-40 residues long that forms an essential component of the cell walls of bacteria from the order Mycobacteriales.[6,19,42,50] The activity of GlfT2 in mycobacteria was shown to be essential and highly vulnerable to disruption, with galactan length reduction significantly affecting cell wall physiology and cellular antibiotic resistance.[22,36,37] With the

Mycobacteriales order consisting of a number of critical human pathogens, include *Mycobacterium tuberculosis* and *Corynebacterium diphtheria*, GlfT2 and its activity represent a promising target for novel antibiotic development.[6,20,36,51]

GlfT2 is a membrane-associated GT-2 family enzyme, binding to the cytoplasmic face of the inner membrane.[52] At the membrane interface, GlfT2 forms a homotetrameric complex with a membrane-enclosed central cavity (Chapter 2).[35] It forms the galactan within this cavity, catalyzing repeated transfers of Gal*f* from the donor UDP-Gal*f* onto its acceptor substrate (Gal*f*)$_{2-3}$-GlcNAc-Rha-decaprenylpyrophosphate.[8] Previous studies into the activity of GlfT2 from *Mycobacterium tuberculosis* have observed this enzyme to polymerize the galactan processively.[19,45,53] Structural characterization of acceptor-bound GlfT2 confirmed acceptor binding within the central cavity, utilizing three subsites to mediate both glycan retention and linkage pattern control during polymerization (Chapter 2). We hypothesized that glycan-sliding between these subsites permits processive polymerization of the galactan product. An additional model previously proposed galactan length control by GlfT2 was mediated by its central cavity.[35] The role of the GlfT2 cavity and the membrane interface on galactan polymerization still remains poorly understood.

To inform the membrane-association mechanism of GlfT2 and the topology of its cavity near the membrane interface, structural characterization of GlfT2 at the membrane is needed. Herein, we describe the cryo-electron microscopy (Cryo-EM) structure of membrane-embedded *Nocardia brasiliensis* GlfT2, embedded into a styrene-maleic anhydride liponanoparticle (SMALP) solubilized from *E. coli*. The experimental map reveals a nanodisc-stabilized membrane, with GlfT2 associated through interactions with its proposed amphipathic helices (Chapter 2).[35] These helices appear to induce membrane curvature within the tetrameric cavity, influenced by the

arrangement of the membrane-binding helices. Coarse-grained molecular dynamics simulations of GlfT2 on a model membrane supports its propensity to induce membrane curvature.[99]

From these observations, we propose a membrane curvature-assisted model for galactan polymerization by GlfT2. Membrane curvature induced by GlfT2 association appears to reduce its cavity volume, potentially constraining the acceptor to promote retention. During galactan polymerization, the acceptor lipid moiety would preferentially move along the membrane-curvature gradient, promoting acceptor exit during polymer growth, potentially supporting longer product lengths. We believe this model has the potential to inform future investigations into GlfT2 catalysis and inhibitor design, and broadly inform mechanisms of other critical polymerizing glycosyltransferases.

3.3 Results

3.3.1 Solubilization of *Nbr*GlfT2 in SMALP Nanodisc

Strep-Tactin purification of SMALP solubilized proteins with a StrepII-tag has previously been reported to improve the purity and yield of the final liponanoparticle-bound protein.[100] StrepII-tag purification is suggested to avoid non-specific charged interactions between the SMALP and the Ni-NTA resin often used for His-tag purification, while increasing the resin- tag binding affinity compared to a Hig-tag. Therefore, the affinity tag used for *Nbr*GlfT2 was switched from 6x His (pAMJ24, Chapter 2) to a Strep-II tag (WSHPQFEK, pAWC25) by PCR amplification.[74,100]

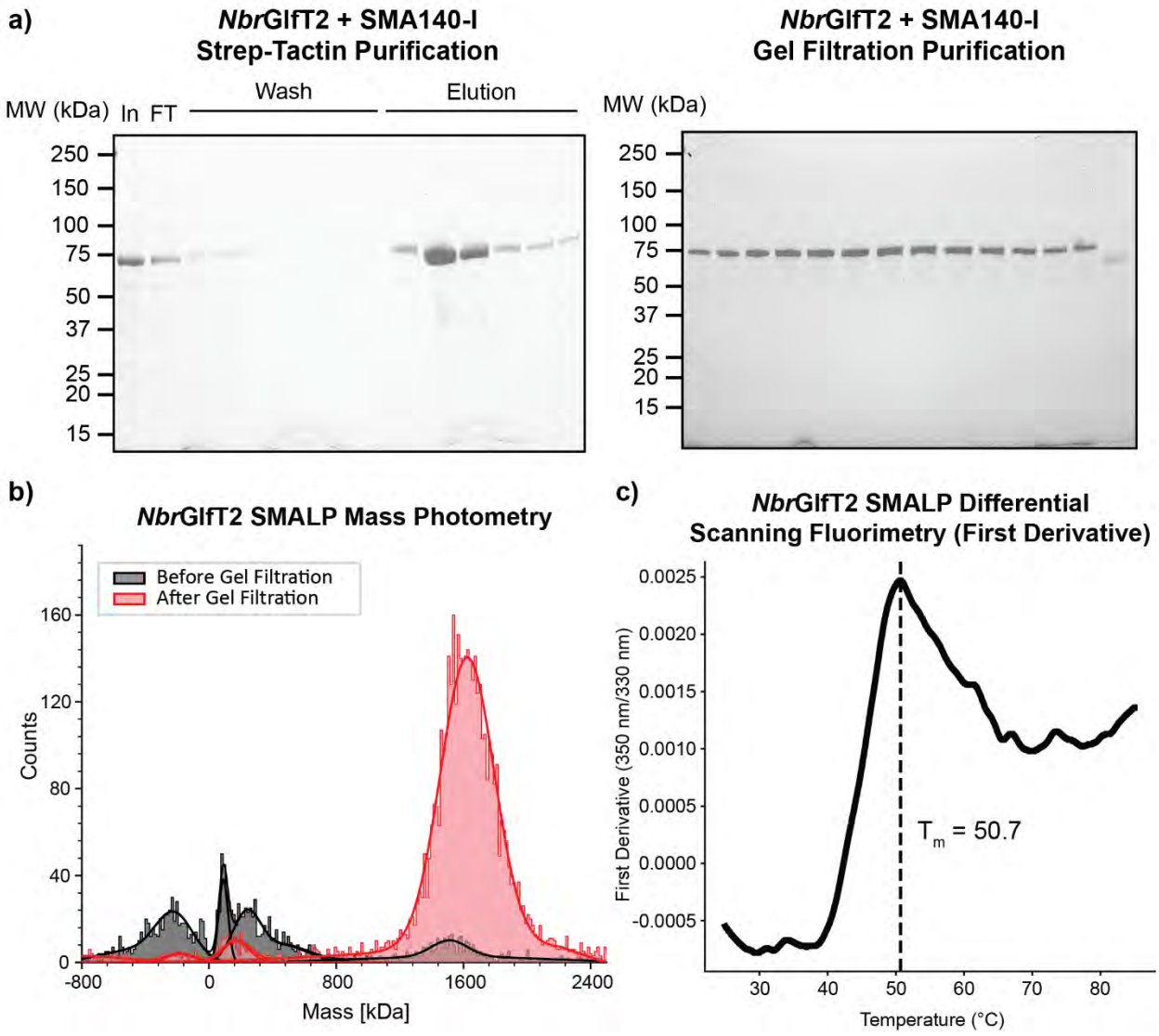

Figure 3.1: Purification and Characterization of SMALP *Nbr*GlfT2. a) Purification and biophysical characterization of *Nbr*GlfT2. Membrane fraction preparations of StrepII-*Nbr*GlfT2 were solubilized with SMA 140I and purified using StrepTactin resin, and eluted using biotin. Additional purification was performed using gel filtration. b) Size estimation and purity determination by Mass Photometry, where population dispersity was compared between samples before and after gel filtration. c) Melting temperature determination of SMALP *Nbr*GlfT2, determined to be 50.7 ˚C.

To determine the optimal liponanoparticle solubilization conditions for StrepII-*Nbr*GlfT2, we performed a small-scale screen of solubilization and purification across various synthetic nanodisc-forming polymers.[100] The best purity and yield was achieved with SMA-140I, a 1.4:1 ratio styrene-maleic anhydride copolymer functionalized with dimethylaminopropylamine. Gel

filtration was performed to improve purity for single particle imaging and analysis (Figure 3.1a). We confirm the monodispersity of the final purified product by Mass Photometry (MP, Figure 3.1b). Mass estimation by MP observes a particle size distribution centered near 1.6 MDa, which would represent the GlfT2 tetramer embedded into to a SMALP. Retention of properly folded *Nbr*GlfT2 within the SMALP nanodiscs was confirmed by differential scanning fluorimetry (Figure 3.1c). The melting temperature of the nanodisc-embedded protein was determined to be 50.7 ˚C.

3.3.2 Cryo-EM of Nanodisc-Embedded *Nbr*GlfT2

Single-particle Cryo-EM imaging was performed on StrepII-*Nbr*GlfT2 purified in a SMA-140I liponanoparticle. Initial imaging produced 8,583 micrographs with 1,552,517 automatically selected particles (Figure 3.2a, b). 2D classification and templated particle picking were performed to improve particle quality. Vitrification conditions used during grid freezing for *Nbr*GlfT2 led to an apparent preferential particle orientation during alignment and 2D classification (Figure 3.3). To mitigate map streaking along the primary axis, a Topaz neural network was trained and executed for particle picking of the underrepresented side view.[101] Orientation distributions were rebalanced to reduce overrepresented orientations, resulting in 356,815 remaining particles.[102] 3D classification further selected for particle quality, reducing the final number of particles to 182,122 particles.

Three-dimensional reconstruction of the 182,122 particles reveals a map at 3.48 Å resolution (Figure 3.2c, EMDB 70290). The C1 refinement map contains a structured volume representing the *Nbr*GlfT2 tetramer (Figure 3.2d), confirming the formation of tetrameric GlfT2 in a non-crystallographic context. The map reveals the protein complex associated with a large blob feature below the tetramer, likely representing the membrane nanodisc. C4 symmetry-expanded

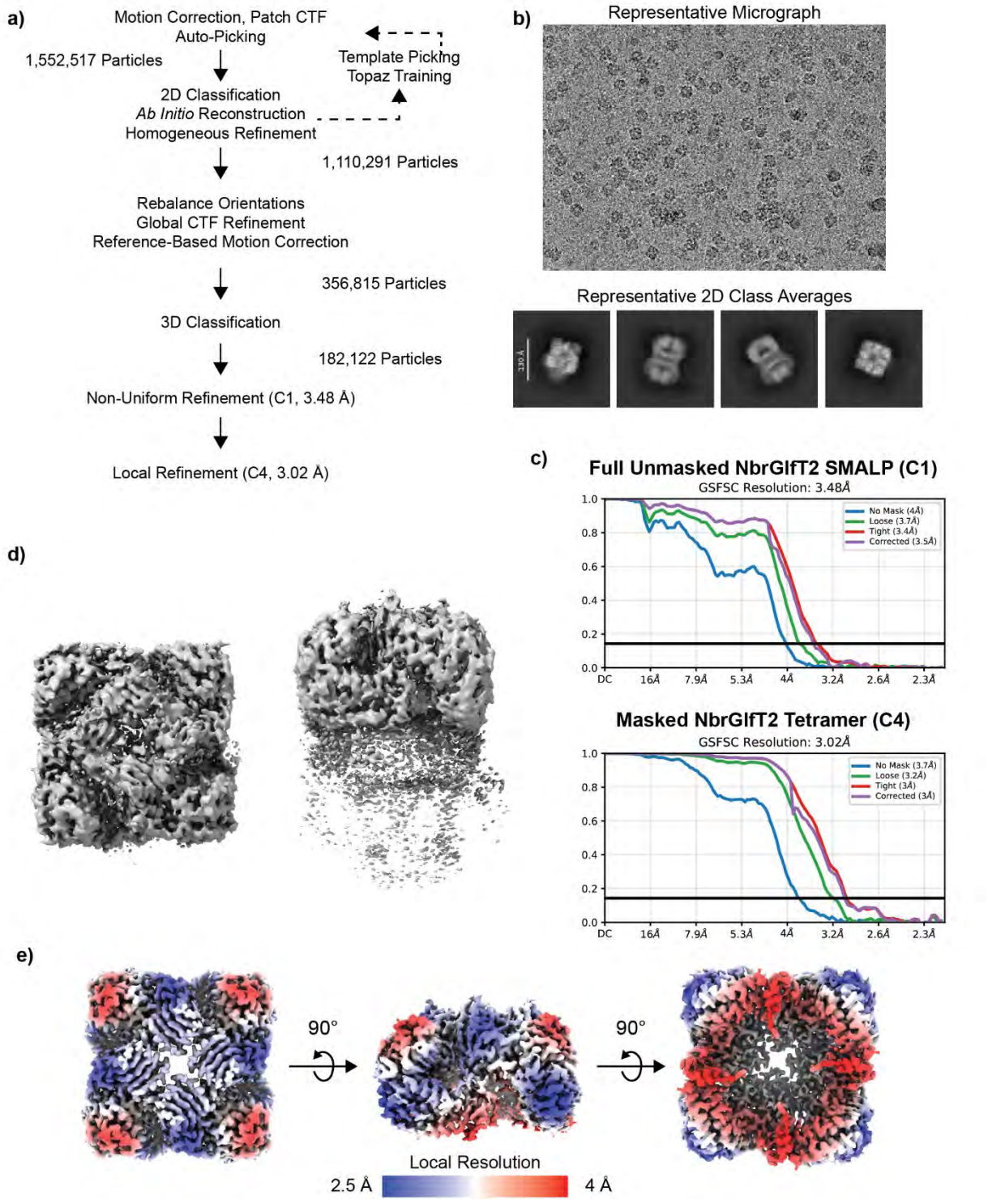

Figure 3.2: Single-Particle Reconstruction of *Nbr*GlfT2 by Cryo-EM. a) Workflow for micrograph processing, particle picking, volume reconstruction and refinement in CryoSPARC. b) Representative micrograph of particles suspended within the grid hole. Representative 2D classes of aligned particles. c) The Gold-Standard Fourier Shell Correlation for resolution determination of unmasked C1 refinement and masked tetramer C4 refinement. d) Map of C1 refinement of unmasked refinement. e) Symmetry-imposed refinement of GlfT2 tetramer map.

Table 3.1: *Nbr*GlfT2 SMA140I Cryo-EM Data Collection, Refinement and Validation Statistics

	C1, EMDB-70290	C4, EMDB-70273, PDB 9OA3
Data collection and processing		
Magnification	81,000	81,000
Voltage (kV)	300	300
Electron exposure (e–/Å^2)	50.682	50.682
Defocus range (μm)	-.25 to -2.25	-.25 to -2.25
Pixel size (Å)	1.06	1.06
Symmetry imposed	C1	C4
Initial particle images (no.)	1,552,517	1,552,517
Final particle images (no.)	182,122	182,122
Map resolution (Å)	3.7	3.0
FSC threshold	0.143	0.143
Map resolution range (Å)		2.2-4.6
Refinement		
Initial model used (PDB code)		9DTP
Model resolution (Å)		3.0
FSC threshold		0.143
Model resolution range (Å)		
Map sharpening *B* factor (Å^2)		108.4
Model composition		
Non-hydrogen atoms		19,180
Protein residues		2,448
Ligands		4
B factors (Å^2)		
Protein		87.44
Ligand		76.04
R.m.s. deviations		
Bond lengths (Å)		0.003
Bond angles (°)		0.579
Validation		
MolProbity score		1.68
Clashscore		5.36
Poor rotamers (%)		1.80
Ramachandran plot		
Favored (%)		96.76
Allowed (%)		3.11
Disallowed (%)		0.12

local refinement of the structured portion of the map produced a final refined 3.02 Å map highly similar to the *Nbr*GlfT2 model determined by X-ray crystallography (EMDB 70273 PDB 9OA3, Figure 3.2c, e, Table 3.1).

Distance comparisons between the embedded *Nbr*GlfT2 and nanodisc maps reveal a protein complex 60 Å in length perpendicular to the protein-membrane interface, with a nanodisc blob nearly 90 Å in length (Figure 3.4a). This blob potentially represents a large membrane bilayer, or a bilayer with an additional unassigned feature. The tetrameric volume observed by local refinement reveals four main domains, as seen with the model solved by X-ray crystallography (Chapter 2). An N-terminal β-sandwich domain similarly forms

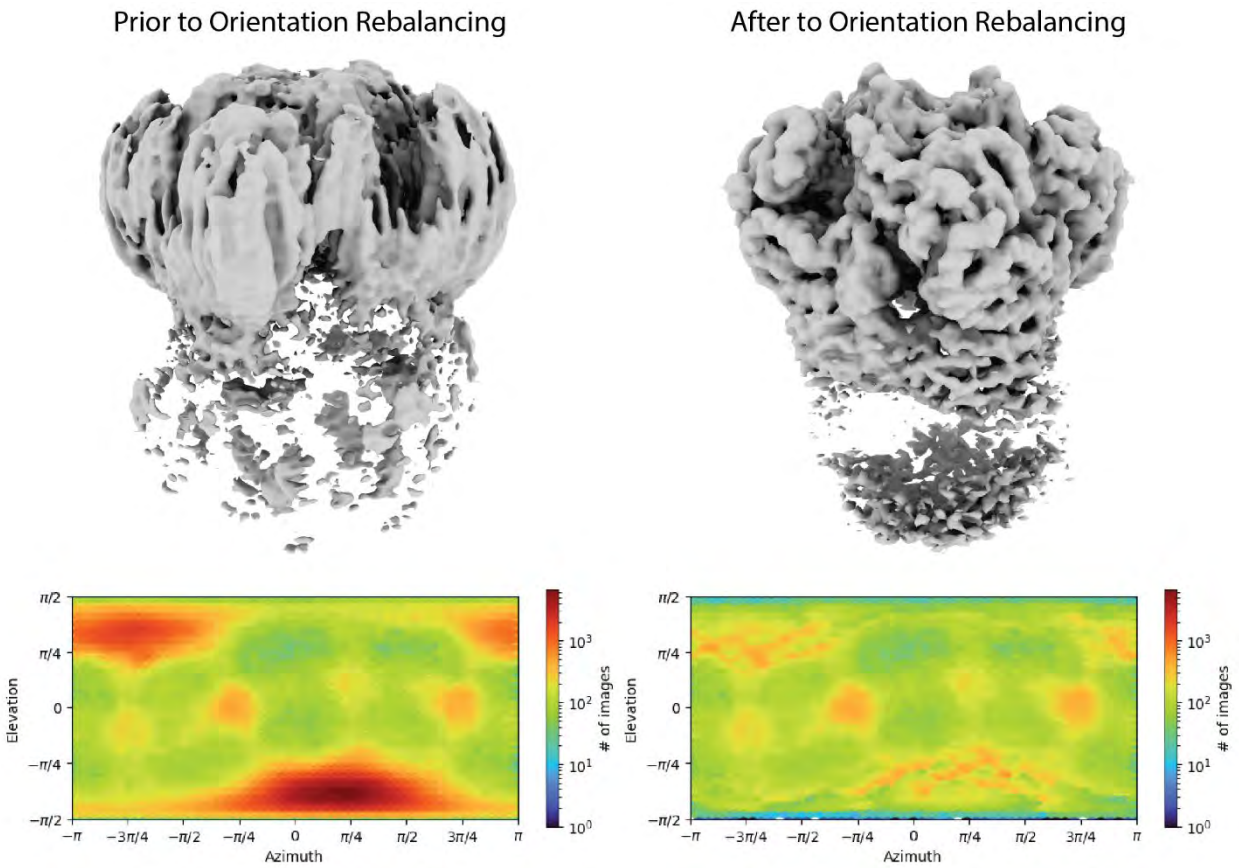

Figure 3.3: Rebalancing Preferred *Nbr*GlfT2 Particle Orientation. Rebalancing of orientation distribution reduces streaking causing by preferred particle orientiation. Above are the volumes before (1,100k particles) and after (360k particles) orientation rebalancing. Below are the distribution maps for orientation representation.

oligomerizing contacts with the mixed α/β domain. The GT-A domain extends toward the corners of the complex, and is known to mediate catalysis in GlfT2 (Chapter 1, 2).[31,35,62] An

α-helical domain is split by a mixed α/β domain, both of which were previously hypothesized to mediate membrane association using sovlent-exposed amphipathic helices.[35] Absences in the refined map are present along the β5-β6 hairpin of the β-sandwich domain, the loop following β7 of the GT-A domain, and the loop linking the N-terminal portion of the α-helical domain and the α/β domain.

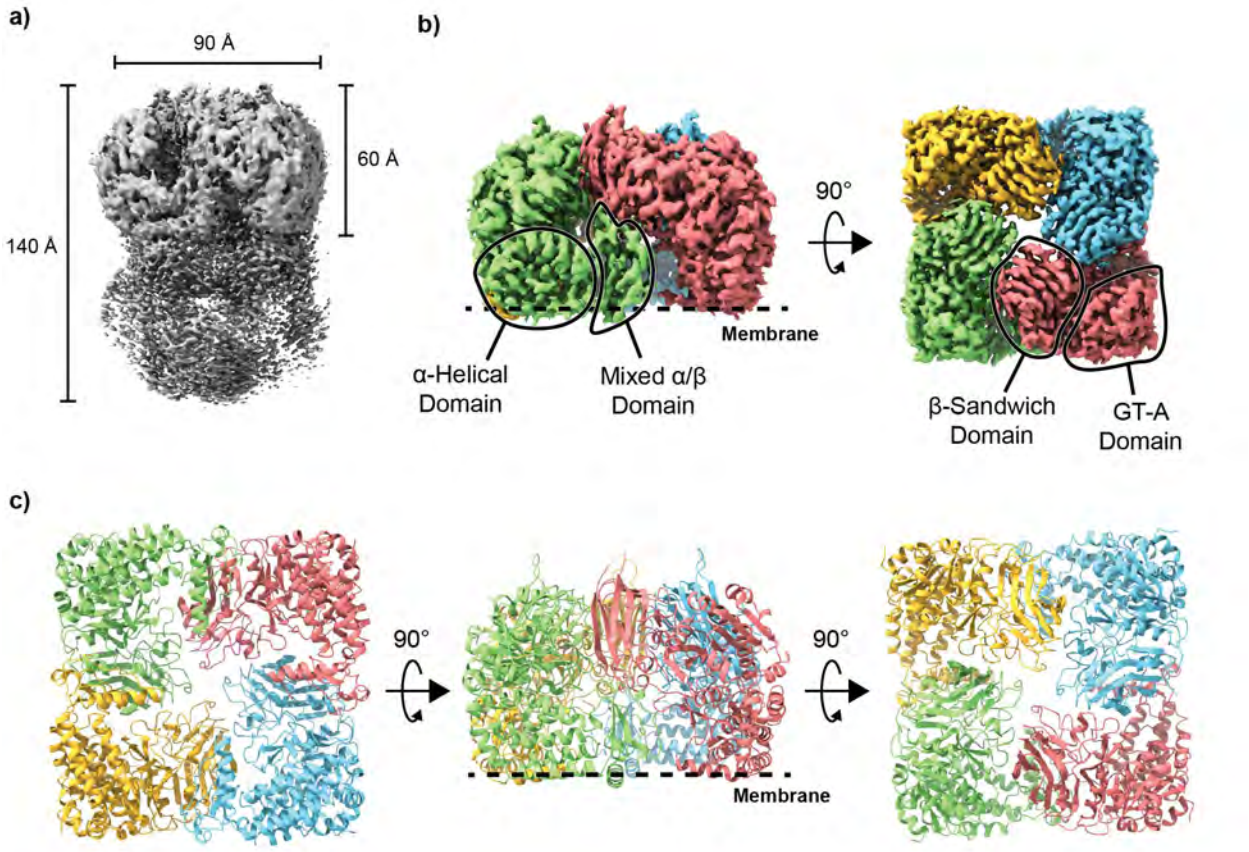

Figure 3.4: Model of Tetrameric *Nbr*GlfT2 in Cryo-EM Map. a) Distance measurements for the dimensions of the nanodisc-associated GlfT2. b) Protein domains visible in the masked volume reconstruction. A depiction of membrane interface is present for the side view of the tetramer. Volume colored by proximity for each protomer docked. c) Protein model built from the masked volume reconstruction.

3.3.3 Membrane Association by *Nbr*GlfT2

GlfT2 was previously hypothesized to associate to the cytoplasmic face of the inner membrane, mediating this binding through three amphipathic helices: α1 of the mixed α/β domain, and α2 and α6 of the α-helical domain (Figure 3.4, 3.5).[35] The refined map of the *Nbr*GlfT2-embedded nanodisc confirms the interaction of these three amphipathic helices with the membrane interface. The amphipathic helices are organized with hydrophobic residues enriched on their solvent-exposed face, orienting these residues away from the tetrameric cavity (Figure 3.5a, b).[35] The opposite faces of these helices are enriched for basic residues that are predicted to form

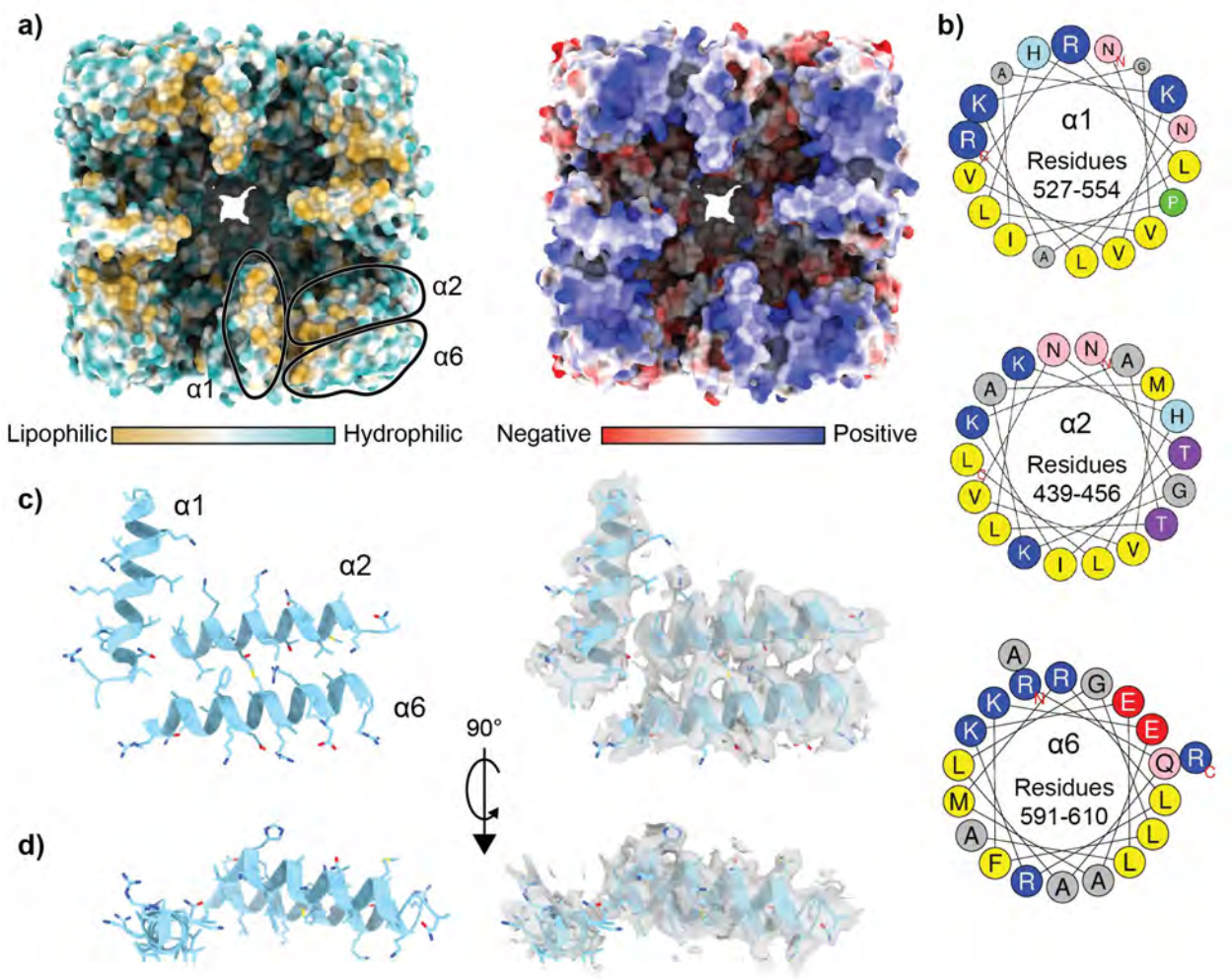

Figure 3.5: Membrane-Associating Amphipathic Helices of *Nbr*GlfT2. a) Surface representation colored for both lipophilicity (cyan: hydrophilic, yellow: lipophilic) and electrostatic potential (blue: positive, red: negative). b) Helical wheel for α1 in the α/β domain, and α2 and α6 in the α-helical domain. c) Depiction of membrane-binding amphipathic helices, with the volume highlighted. d) Different orientation of membrane-binding amphipathic helices.

coulombic interactions with the negatively charged head-groups of phospholipids at the membrane interface (Figure 3.5c, d).

The GlfT2-embedded blob is nearly 90 Å in length, and potentially represents a membrane bilayer with a bifurcation in the center between the membrane leaflets, or an uncharacterized particle aligned with GlfT2 at the membrane. This bifurcation forms a plane parallel to the

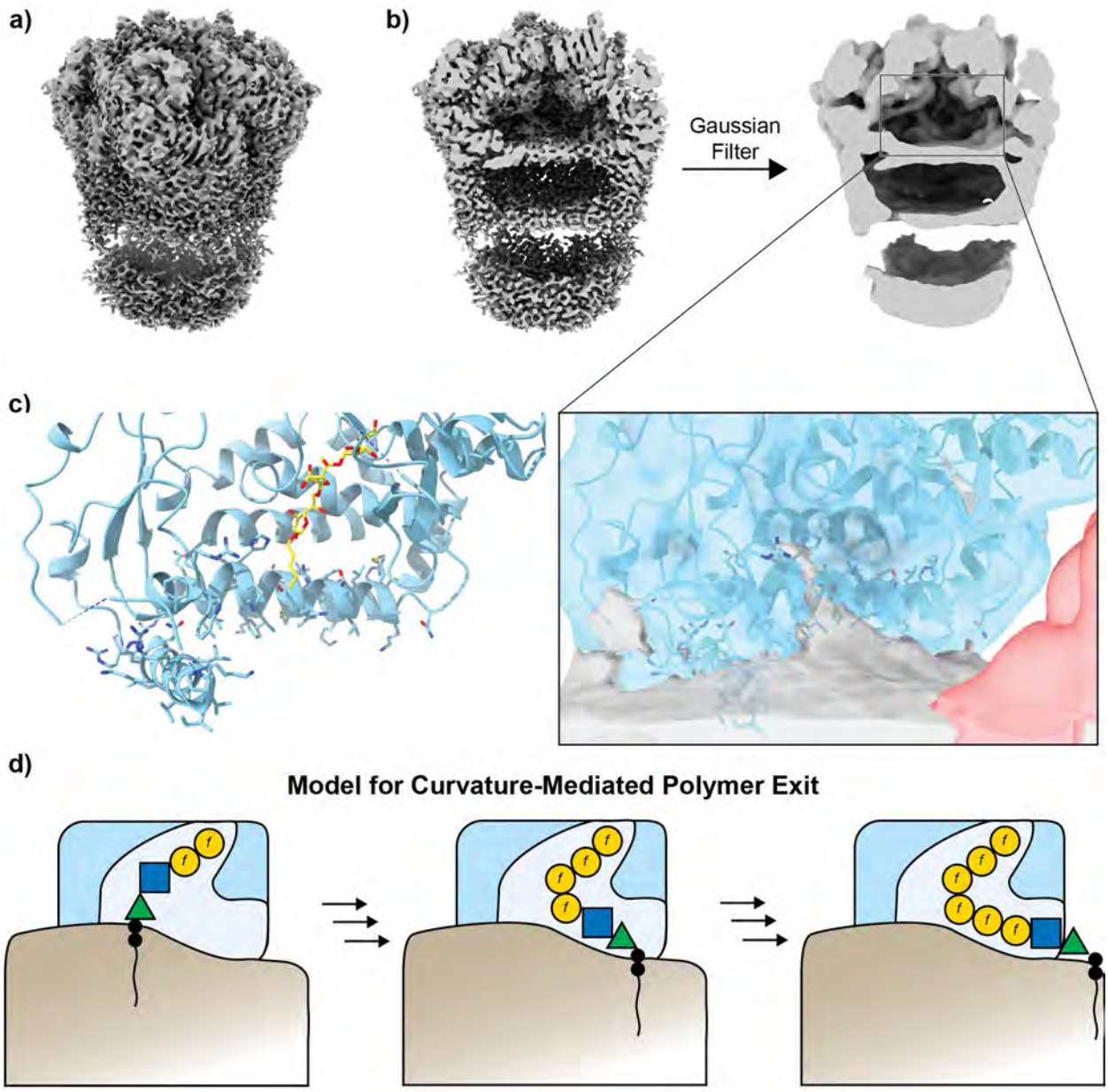

Figure 3.6: GlfT2-Induced Membrane Curvature. a) Map of C4 symmetry expanded volume masked map. b) Map clipped to highlight the membrane boundary. Gaussian filtering to 2σ was used to smooth volume for interpretation. c) Structural model of NbrGlfT2 modeled with a tri-Gal*f* acceptor, with residues shown along the membrane-binding amphipathic helices. The Gaussian-filtered volume reveals apparent membrane curvature surrounding the amphipathic helices, and a rising membrane towards the acceptor. d) Model proposed for curvature-assisted acceptor translocation

membrane-associating region of the GlfT2 complex. The overall map agrees with the expected membrane association previously proposed for GlfT2.[35]

3.3.4 Membrane Curvature Induced by GlfT2 Binding

Inspection of the membrane interface within the cavity of GlfT2, smoothed by Gaussian filtering of the refined map (Figure 3.6a, b), suggests the presence of membrane curvature induced by GlfT2 association, leading into the acceptor-binding site (Figure 3.6c). Localization of the three amphipathic helices with the membrane interface suggested α1 of the mixed α/β domain potentially organizes its hydrophobic and basic residues lower than those of the two amphipathic helices of the α-helical domain (Figure 3.5d). In particular, α2 and α6 of the α-helical domain appear to have solvent-exposed hydrophobic residues above α1 (Figure 3.6c). Differential exposure of these residues suggests the membrane interface may undergo localized reshaping upon GlfT2 association to accommodate these residues.

Visualization of the membrane interface around these helices indeed reveals membrane curvature lower and away from the central cavity near α1, and up and into the GlfT2 protomer near α2 (Figure 3.6c). Interestingly, the apparent interface curvature into the protomer is at the previously characterized GlfT2 acceptor binding site (Chapter 2). This induced-curvature may play a role in assisting the primed acceptor in binding to the active site of GlfT2.

Informed by the apparent membrane-interface curvature induced by GlfT2, we propose a model in which this curvature influences acceptor extension and exit from the central cavity (Figure 3.6d). In this model, acceptor binding to the GlfT2 active site occurs above the membrane interface, potentially assisted by the observed curvature induced by the GlfT2 α1 and α2 helices. Upon initiation of galactan polymerization, the phospholipid portion of the galactan acceptor is permitted to move away from the active site while catalysis proceeds. The acceptor-site subsites observed for galactan binding (Chapter 2) retain the glycan during elongation, allowing processive elongation to continue. As elongation proceeds, the growing galactan product within

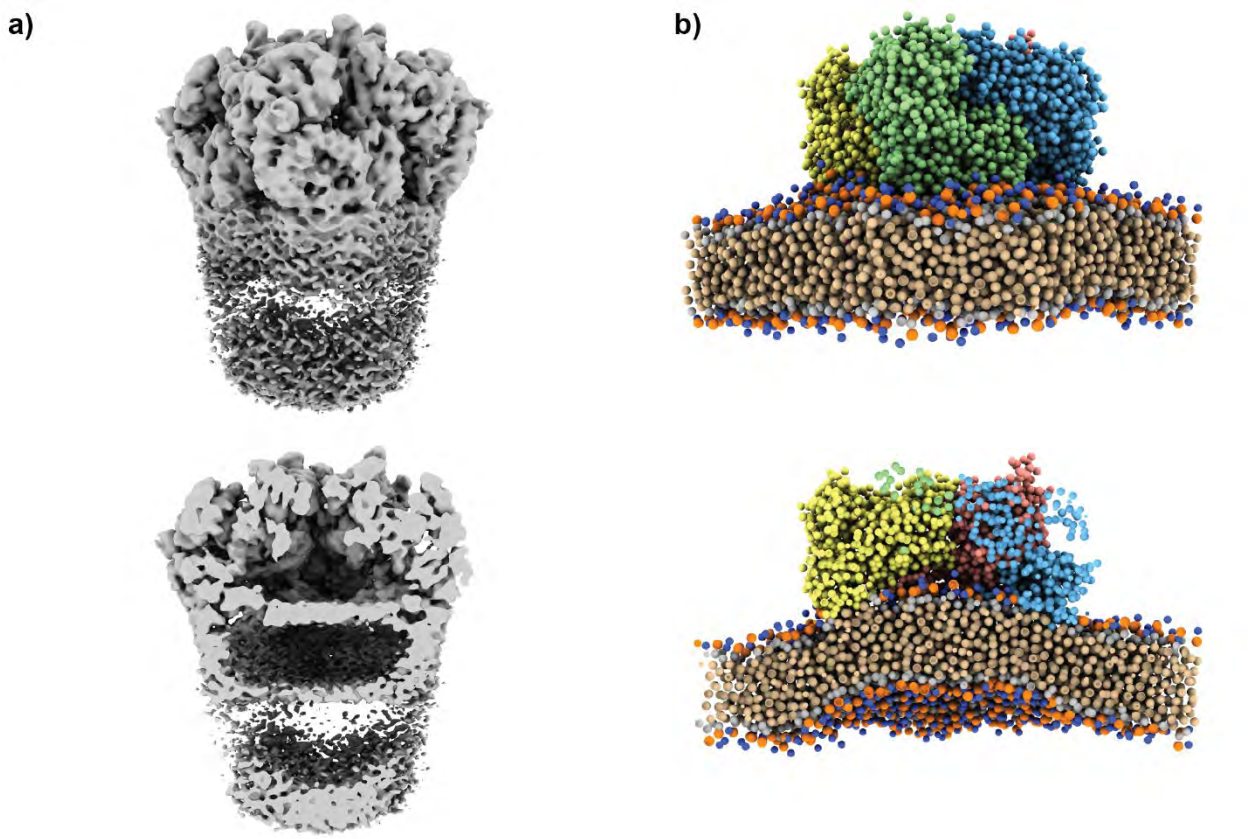

Figure 3.7: Coarse-Grained Simulation of GlfT2 Membrane Association. a) Map of masked C4-symmetry expanded *Nbr*GlfT2. Below is clipped halfway into tetrameric cavity. b) Coarse-grained simulation of *Nbr*GlfT2 in a dipalmitoylphosphatidylcholine bilayer. (blue: NH^{3+}, orange: PO_4^{2-}, grey: glycerol group, beige: lipid group)

the cavity is predicted to translate the phospholipid portion of the acceptor further down the gradient of the curved membrane interface, leading outside the cavity. This model would suggest a curvature-assisted polymerization mediated by the membrane.

3.3.5 Coarse-Grained Molecular Dynamics of *Nbr*GlfT2 Membrane Binding

Molecular dynamics simulation of a coarse-grained *Nbr*GlfT2 embedded into a model dipalmitoylphosphatidylcholine (DPPC) membrane bilayer was performed to further validate the propensity for GlfT2 to induce membrane curvature (Figure 3.7). GlfT2 orientation and proximity to the membrane was initially estimated with the PPM server.[32] Conversion of the full

atom system to coarse-grained groups was defined through Maritini in order to simplify the simulation of the membrane-associated system, and promote system convergence to a representative final state.[99]

During the course of the simulation, the coarse-grained membrane forms an apparent puckered shape in the cavity of GlfT2 (Figure 3.7). The membrane-binding amphipathic helices remain associated with the membrane interface.[35] Similar to the curvature observed for the experimental map, coarse-grained GlfT2 appears to induce membrane curvature in a simplified membrane system. Using the phosphate and glycerol groups as representative of the membrane interface observed by Cryo-EM, a similar curvature is seen between the simulated and experimental structures (Figure 3.7). These observations suggest GlfT2 can intrinsically induce membrane curvature, and the orientations of residues along the amphipathic helices may further alter the specific extent of curvature within the GlfT2 cavity.

3.4 Discussion

Based on the Cryo-EM volume reconstruction of *Nbr*GlfT2 embedded into a SMA-140I solubilized liponanoparticle, we propose a model in which membrane curvature assists in galactan polysaccharide elongation. In this model, three amphipathic helices mediate membrane-association by GlfT2. These helices induce an apparent curvature of the membrane interface within the central cavity (Chapter 2).[35] The membrane interface from the refined map appears to curve into the cavity in a convex manner, and curves down towards the edge of the tetrameric complex. Additional apparent curvature is seen near the acceptor-binding site, where the membrane slightly rises into the pocket between the amphipathic helices (Figure 3.5c). This continual membrane curvature from the acceptor site to the exit of the GlfT2 cavity suggests a

role in mediating galactan translocation during polymerization. In our model, as the galactan elongates, crowding of the growing polymer promotes acceptor translation along the curving membrane gradient, assisting galactan exit from the cavity. This mechanism could potentially reduce polymerization stalling due to cavity overcrowding during simultaneous elongation of the four protomers within the tetramer.

The use of a membrane interface in this model is similar to enclosure-mediated processive polymerization.[67] In enclosure-mediated polymerization, a long channel or enclosed cleft promotes retention of a growing polymer product by reducing the degrees of freedom of the acceptor substrate. A similar mechanism appears to potentially be involved in the bacterial cellulose synthase A (BscA), also a GT-2 family polymerase similar to GlfT2, which forms a long acceptor-binding channel.[103] In addition to the previously defined sliding mechanism for the GlfT2 acceptor (Chapter 2), this mechanism may help to influence the length of galactan product produced by GlfT2. Here, the membrane interface would potentially form a barrier that promotes enclosure, rather than a protein channel.

A previous model for GlfT2 activity suggested that the central GlfT2 cavity would potentially control polymerization and galactan length through a cavity-filling mechanism (Chapter 1).[35] The observations described here suggest the membrane interface may potentially reduce the overall volume of the cavity and instead promote galactan elongation outside of the cavity. This model is in agreement with a recent study that observed longer galactan products after silencing the galactan ABC transporter complex Wzm/Wzt.[14] This suggests that the transporter complex may influence galactan length by prematurely terminating extension. Binding of the still-growing galactan to the ABC transporter could only occur for a processive GlfT2 if the lipid moiety were

allowed to exit the cavity during elongation, rather than remaining in the cavity as previously suggested.[35]

The proposed curvature-assisted polymerization model will potentially inform mechanisms performed by other membrane associated glycosyltransferases. A recent structural investigation of TarL from *Staphylococcus aureus* revealed a micelle-embedded tetrameric structure.[72] This GT-B enzyme appears to form a C4-symmetric complex similar to GlfT2, with membrane association mediated by two amphipathic helices. As ribitol-phosphate polymerization occurs within TarL's tetrameric cavity, its activity may likewise be similarly assisted by membrane enclosure and curvature-assisted cavity exit. The Cryo-EM map of micelle-embedded TarL indeed reveals slight curvature from the micelle interface, but its propensity to induce curvature at a native membrane interface remains unclear. Interestingly, the evolutionary convergence of a cavity-forming tetrameric structure for the evolutionarily-distinct GT-A enzyme GlfT2 and GT-B enzyme TarL suggests an important role for the cavity in their polymerizing glycosyltransferase activity.

In conclusion, we describe here the structure of *Nbr*GlfT2 embedded to a liponanoparticle. This structure informs the membrane-associated structure of GlfT2, and confirms the role of three amphipathic helices in mediating membrane association. These amphipathic helices appear to induce curvature of the membrane interface within the cavity formed by GlfT2. We propose a model in which this curvature may assist in galactan polysaccharide translation outside of the cavity during elongation. This structure and model are expected to inform future mechanistic investigations into GlfT2 and other critical polymerizing glycosyltransferases.

3.5 Methods

3.5.1 Plasmid Preparation

Plasmid preparation for StrepII-*Nbr*GlfT2 (pAWC25) was accomplished by switching out the tag from His_6-*Nbr*GlfT2 (pAMJ24) using PCR amplification of primers designed with 5'-end StrepII tag sequence and a 3'-end template binding region (Table 3.2).[74] Plasmid assembly was performed using an In Vivo Assembly protocol in chemically-competent *E. coli* DH5α cells. Plasmid purification was performed with a Miniprep Kit (Promega). Successful tag switching was confirmed by Sanger Sequencing (Quintara Biosciences).

Table 3.2: Primers Used for Plasmid Preparation

Primer Name	Primer Sequence
Nbr_Strep-Add_F	GATATACCATGGGCAGCAGCTGGAGCCATCCGCAGTTTGAAAAG
Nbr_Strep-Add_R	CGCGGCACCAGGCCGCTGCTCTTTTCAAACTGCGGATGGCTCCA

3.5.2 Protein Expression and Purification

Protein production was performed using *E. coli* DE3 chemically transformed with the expression vector pAWC25. Protein was expressed using an autoinduction protocol, using Terrific Broth (TB) supplemented with 150 ug/mL of kanamycin as base media.[75,76] Cells were grown on an orbital shaker in 0.5 L TB and 150 μg/mL of kanamycin to an optical density (600 nm) of 0.8 at 37 ˚C, then cooled to 18 ˚C and shaken for ~20 hour of protein expression. Cells were pelleted by centrifugation at 4,000 xg for 10 minutes at 4 ˚C, and the pellets stored at -80 ˚C.

Cells pellets were resuspended in Buffer A (50 mM HEPES, 300 mM NaCl, pH 7.4) at 5 mL/g, along with 1 mg/mL Egg White Lysozyme (Gold Biotechnology) and 5 units of Benzonase

Nuclease (EMD Millipore). The suspension was lysed by mechanical disruption (Constant Systems). Cell lysate was initially centrifuged at 9,000 xg for 30 minutes at 4 ˚C, and the supernatant was additionally spun at 100,000 xg for 1 hour at 4 ˚C to separate the membrane fraction. The pellet was resuspended in Buffer A, and flash frozen in liquid nitrogen for storage at -80 ˚C.

The membrane fraction was thawed and mixed with SMA-140I to 1% (w/v) copolymer. This mixture was incubated at room temperature for 30 minutes, then diluted to a final copolymer concentration of 0.25%, and spun down at 100,000 xg for 1 hour at 4 ˚C. The supernatant was passed through Strep-Tactin resin to bind the StrepII tagged protein by gravity flow. The resin was washed 5 times with 1 column volume of Buffer A. The protein was eluted from the resin 6 times with 0.5 column volumes of Buffer A with 5 mM desthiobiotin. Further purification and desthiobiotin removal was performed on a Superdex 200 prep grade filtration column (Cytiva). The purified sample was concentrated using an Amicon Ultra-4 Centrifugal Filter (MilliporeSigma) to a final concentration of 1 mg/mL as measured by UV-Vis. Concentrated protein was aliquoted, flash frozen in liquid nitrogen, and stored at -80 ˚C.

3.5.3 Biophysical Characterization of *Nbr*GlfT2 Nanodisc

Protein denaturing and melting temperature determination were performed using a Prometheus Panta nanoDSF (Nanotemper). Protein unfolding was monitored from 25 ˚C to 85 ˚C with a 0.5 ˚C/min ramping rate. A Refeyn Two Mass Photometer was used for particle size distribution determination. Buffer A was prepared for sample dilution by filtration with a 0.22 μm polyethersulfone syringe filter (Millipore Sigma). Background measurement was performed on 18 μL of Buffer A, and sample measurement was performed immediately after mixing in 2 μL of 1 mg/mL of protein sample.

3.5.4 Cryo-TEM Grid Preparation and Particle Imaging

Grid preparation of *Nbr*GlfT2 solubilized in SMA-140I nanodiscs was performed using CFlat Holey Carbon Grid 1.2 μm/1.3 μm, 300 mesh copper grids. Samples were spotted on the grid as 3 μL drops at a 1.3 mg/mL protein concentration, and blotting was performed using the Vitrobot Mark IV (Thermo Fisher Scientific), and plunge frozen in liquid ethane. Grids were screened on a Talos Arctica G2 Cryo-TEM at 200 kV with a Falcon 3EC detector, and final images were captured using a Titan Krios G3i Cryo-TEM at 300 kV with a K3 Gatan detector at 81 kX.

3.5.5 Cryo-EM Image Processing and Volume Reconstruction

Motion correction, Contrast Transfer Function (CTF) estimation, *ab initio* reconstruction, and refinement steps were performed using CryoSPARC (Cryo-EM Single Particle Ab Initio Reconstruction and Classification) v4.6.0.[102] PHENIX 1.21.2-5419, Coot 0.9.8.95, and Isolde 1.8 in ChimeraX 1.8 were used for model docking and real-space refinement of the *Nbr*GlfT2 crystal structure (PDB: 9DTP) into the refined map.[79,104,105] ChimeraX 1.8 was used for volume and structural model visualization, and map Gaussian filtering.[106]

3.5.6 Coarse-Grained Molecular Dynamics of Membrane-Embedded GlfT2

The crystal structure of *Nbr*GlfT2 (PDB: 9DTP) was used as a starting point. The relative membrane boundary to the *Nbr*GlfT2 tetramer was estimated using the Positioning of Proteins in Membrane (PPM) server.[32] An all-atom *Nbr*GlfT2 model with an associated membrane was produced using CharmmGUI, with a 16 nm x 16 nm x 15 nm periodic boundary.[107,108] The membrane used was a bilayer of dipalmitoylphosphatidylcholine (DPPC). Counter ions were added to a concentration of 100 mM NaCl. The atomic structure of the *Nbr*GlfT2-DPPC

membrane complex was converted to a coarse-grained system using the Martini3 definition.[99,109] Three groups were defined for the protein chains, membrane bilayer, and solute.

All simulation steps were defined by CharmmGUI for GROMACS, using the Martini 3 force field.[99,109,110] The systems were simulated at 303.15 K using Reaction-Field electrostatics with a coulomb cutoff of 1.1. A Verlet cutoff scheme was enforced with a tolerance of 0.005. Two minimizations were run for 5 ps each, with a single initial soft-core minimization step. Four separate equilibration steps were run with increasing durations of 2 to 15 fs. A single production step was run for 30 ns on the equilibrated system.

3.6 Acknowledgements

We would like to acknowledge the help of the MIT.Nano staff in preparation and collection of Cryo-EM images, including the assistance of Dr. Nicholas Menousos and Dr. Dana Westmoreland in grid preparation, and Jenn Podgorski and Sarah Sterling in particle imaging. The MIT Biophysical Instrumentation Facility and Dr. Ky Lowenhaupt were instrumental in performing the DSF and MS experiments. Special thanks to Dr. Greg Dodge for his indispensable assistance and advice concerning image processing, particle reconstruction and map refinement.

Funding: Research reported in this publication was supported by the National Institute of General Medical Sciences and the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number F31GM148069 and R01AI126592.

Chapter 4:

Structural Characterization of a Novel Galactofuranosyl Transferase in *Rhodococcus equi*

Contributions: Conceptualization by Alan W. Carter and Laura L. Kiessling. Initial identification of GlfT3 was informed by Alexander M. Justen and Katherine Taylor. Protein expression and purification was performed by Alan W. Carter. *In vitro* activity assays were performed by Alan W. Carter and Hallie O. Trial. The one-pot activity assay was developed and optimized by Hallie Trial and Katherine I. Taylor. Linkage analysis was performed by Hallie O. Trial and Alan W. Carter. Research designed by Alan W. Carter, Hallie O. Trial, Katherine I. Taylor, Alexander M. Justen, and Laura L. Kiessling.

4.1 Abstract

Polysaccharides serve vital functions across biology, including roles in cell-cell communication, immune recognition and evasion, energy storage, protein quality control, and cell wall structures. Glycosyltransferases (GTs) are key enzymes responsible for synthesizing polysaccharides, often working in concert within pathways to create polymers with defined linkage patterns and compositions. In mycobacteria, the galactan polysaccharide—a crucial component of the cell wall—is synthesized by four GTs, including the processive polymerase Galactofuranosyl Transferase 2 (GlfT2). This galactan consists of a linear polymer with an alternating β(1,5) and β(1,6) linkage pattern. In non-mycobacterial species, variations in this sequence have been observed, such as the incorporation of an additional β(1,3) linkage in *Rhodococcus equi* (Req). Here, we characterize *Req*GlfT3, a paralog of GlfT2 which produces galactofuranosyl oligosaccharides. The crystal structure of *Req*GlfT3, solved at 3.3 Å resolution, reveals a tetrameric organization with high structural similarity to GlfT2 but distinct features in the acceptor-binding site. These findings shed light on the functional implications of galactan sequence diversity and offer insights into the broader roles of GlfT enzymes across bacterial species.

4.2 Introduction

Glycosyltransferases (GTs) are enzymes that catalyze the transfer of sugar moieties onto acceptor substrates, often leading to the formation of oligosaccharides and polysaccharides. These carbohydrates play diverse and essential roles in biological systems, including cell-cell communication, immune system recognition and evasion, energy storage, protein quality control, and cell wall structures. GTs often operate within biosynthetic pathways, collaborating with other GTs of varying specificities to add sugars stepwise, thereby determining the sequence and length of the resulting polymer. Determining the mechanisms of these GTs may inform the development of inhibitors targeting pathogenic polysaccharide biosynthesis, and engineering enzymes for industrial polysaccharide production.[6,111,112]

The galactan polysaccharide of Mycobacteriales species is a critical component of their cell wall, forming a portion of the mycolyl-arabinogalactan-peptidoglycan (mAGP) structure.[6] The order Mycobacteriales includes several significant human pathogens, making the galactan an attractive target for novel antibiotic development. The GTs of the galactan biosynthetic pathway have been shown to be essential for mycobacteria survival and highly vulnerable to disruption.[6,22] Perturbations that lead to shortened galactan were shown to impair mycobacteria fitness, decreasing periplasm periplasmic thickness and increasing antibiotic susceptibility in *Mycobacterium smegmatis*.[37] As mycobacterial galactan biosynthesis remains untargeted by any known antibiotics, these findings underscore the potential of the galactan as a viable target for novel therapeutic development.[6,37]

The galactan is a polymer of Gal*f* units, and its formation is catalyzed by the GT Galactofuranosyl Transferase 2 (GlfT2).[19,35] GlfT2 is responsible for generating a galactan

sequence containing alternating β(1,5) and β(1,6) linkages.[19,42] The purpose of the mycobacterial galactan linkage pattern is unknown, but its production by GlfT2 is executed with high fidelity.[10] Interestingly, linkage analysis of galactan from non-mycobacterial species throughout the Mycobacteriales order suggested the presence of different Gal*f* linkage patterns. Investigation of GlfT2 from the non-mycobacterial species *Nocardia brasiliensis* (Nbr), previously reported to have a β(1,5)-repeating galactan, found its function to be the same as *Mycobacterium tuberculosis* GlfT2, producing both β(1,5) and β(1,6) linkages (Chapter 2).[42] This suggests GlfT2 enzymes across Mycobacteriales may similarly function to produce β(1,5) and β(1,6) linkages. Therefore, other unidentified enzymes may be involved in either galactan biosynthesis or processing within these organisms to generate the observed diverse Gal*f* linkage patterns.

Here, we identify an uncharacterized class of GTs present throughout Mycobacteriales that are absent in the families Mycobacteriaceae and Corynebacteriaceae. Sequence alignments reveal this class to represent paralogs of GlfT2, which we designate as GlfT3. Biochemical characterization of GlfT3 from Nbr and *Rhodococcus equi* (Req) confirm their Gal*f* transferase activities, and their ability to polymerize oligosaccharides in a distributive manner.[44] Structural characterization of *Req*GlfT3 reveals three-dimensional homology to GlfT2, including a membrane-associating region, but also identifies substitutions of key residues known to be critical to galactan polymerization.[35] These findings suggest that the GlfT3 enzyme may produce alternative galactan linkages, expanding the galactan biosynthetic pathway and potentially altering the function of the galactan in GlfT3-containing organisms.

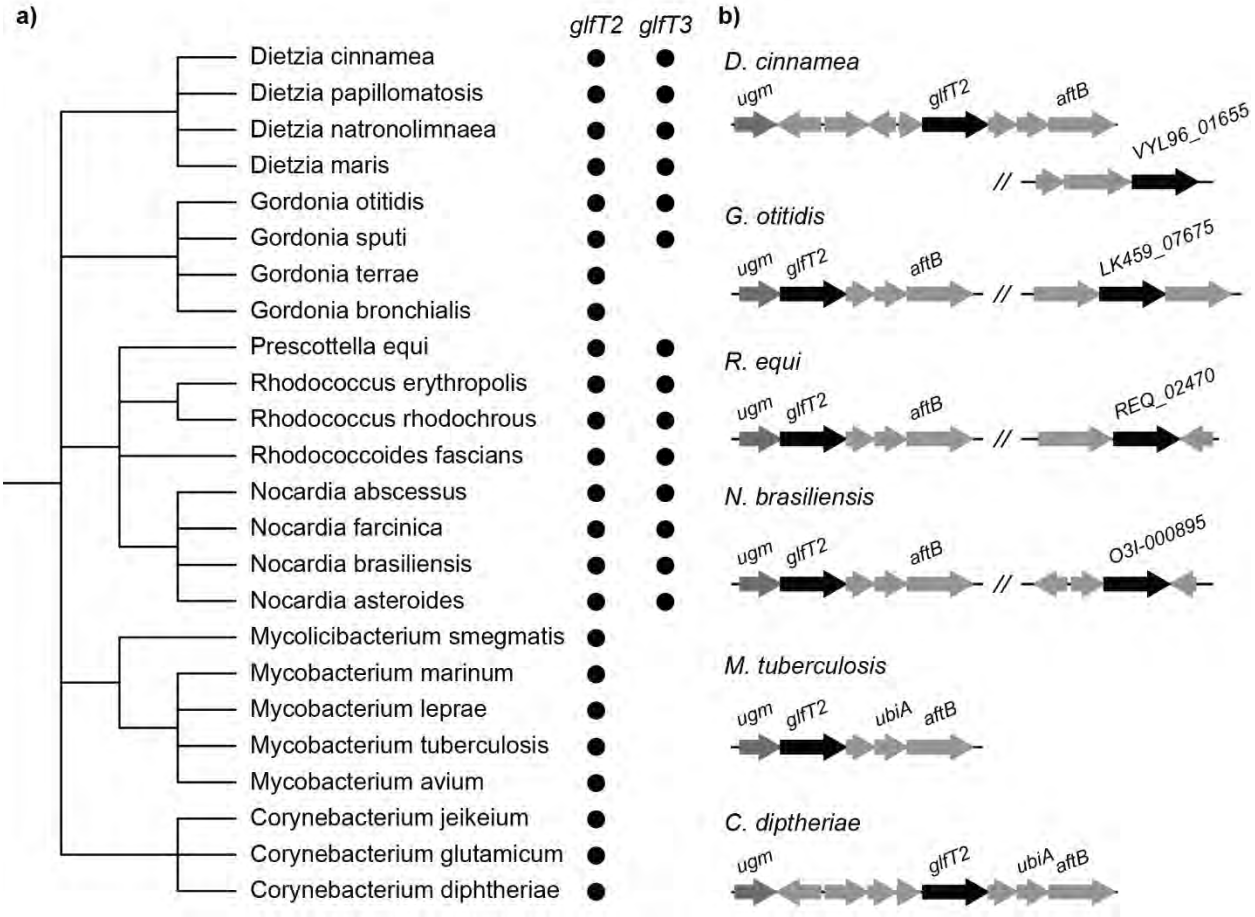

Figure 4.1: Identification of Uncharacterized GlfT2-Paralogs in Mycobacteriales. a) Phylogenetic tree of select species from within the Mycobacteriales order. The presence of an identified *glfT2* and *glfT3* gene are indicated. b) The local gene neighborhoods surround the *glfT2* and *glfT3* genes from select species within each indicated genus.

4.3 Results

4.3.1 Novel GlfT Paralog Present in Mycobacteriales

To inform the potential differences in galactan linkage pattern, representative species from within the order Mycobacteriales were chosen (Figure 4.1a).[42] The genomes of these species were searched for sequences similar to that of *Nbr*GlfT2.[113] Possible genes representing paralogs of similar length to GlfT2 (> 600 amino acids) were labeled as *glfT3*, and their overall three-dimensional structure was predicted and confirmed to have homology to known structures of

*Mtb*GlfT2 (PDB: 4FIX) and *Nbr*GlfT2 (PDB: 9DTP).[28,29,114] Across these species, the families Dietziaceae, Gordoniaceae, and Nocardiaceae contained genes that represented possible *glfT3* genes, with exception of *Gordonia terrae* and *Gordonia bronchialis*. Interesting, species within the genera *Mycobacterium* and *Corynebacterium* lacked this GlfT2 paralog.

The local genomic neighborhood surrounding *glfT2* is similar across each selected species (Figure 4.1b). Preceding GlfT2, this gene cluster commonly consists of a *ugm* gene to encode for the Gal*f*-forming enzyme UDP-Galactopyranose Mutase (UGM) critical for galactan biosynthesis. Following GlfT2, the gene cluster commonly contains enzymes involved in arabinan biosynthesis, including a putative decaprenylphophoryl-5-phosphoribose phosphatase and a *ubiA* gene implicated in decaprenylphosphate-arabinose (DPA) production, and an *aftB* gene involved in arabinan capping.[115–117] The conserved proximity of these genes highlights the essential interplay of these pathways in arabinogalactan biosynthesis.

The genomic neighborhoods for *glfT3* were much more diverse than that seen for *glfT2* (Figure 4.1b). In *Dietzia cinnamea* (Dcd), two genes potentially co-transcribed with *glfT3* are predicted to encode for a GtrA-family enzyme, flippases often implicated in glycolipid transport, and a GT-C family glycosyltransferase structurally similar to Rv3779.[118–121] *Gordonia otitidis glfT3* is preceded by a potential mannosyltransferase predicted to be similar to that involved in ManLAM biosynthesis in Mtb.[122,123] The *glfT3* of *Rhodococcus equi* (Req) is potentially co-transcribed with an enzyme predicted to be structurally similar to the proposed lipotechoic acid biosynthetic enzyme YfhO.[124,125] *Nocardia brasiliensis* (Nbr) is also preceded by a hypothesized GtrA-family protein.[59] Interestingly, each *glfT3* appears to populate a unique genomic region across the various families highlighted, but a common trend of proximity to genes encoding proteins involved in glycosylation pathways is apparent.

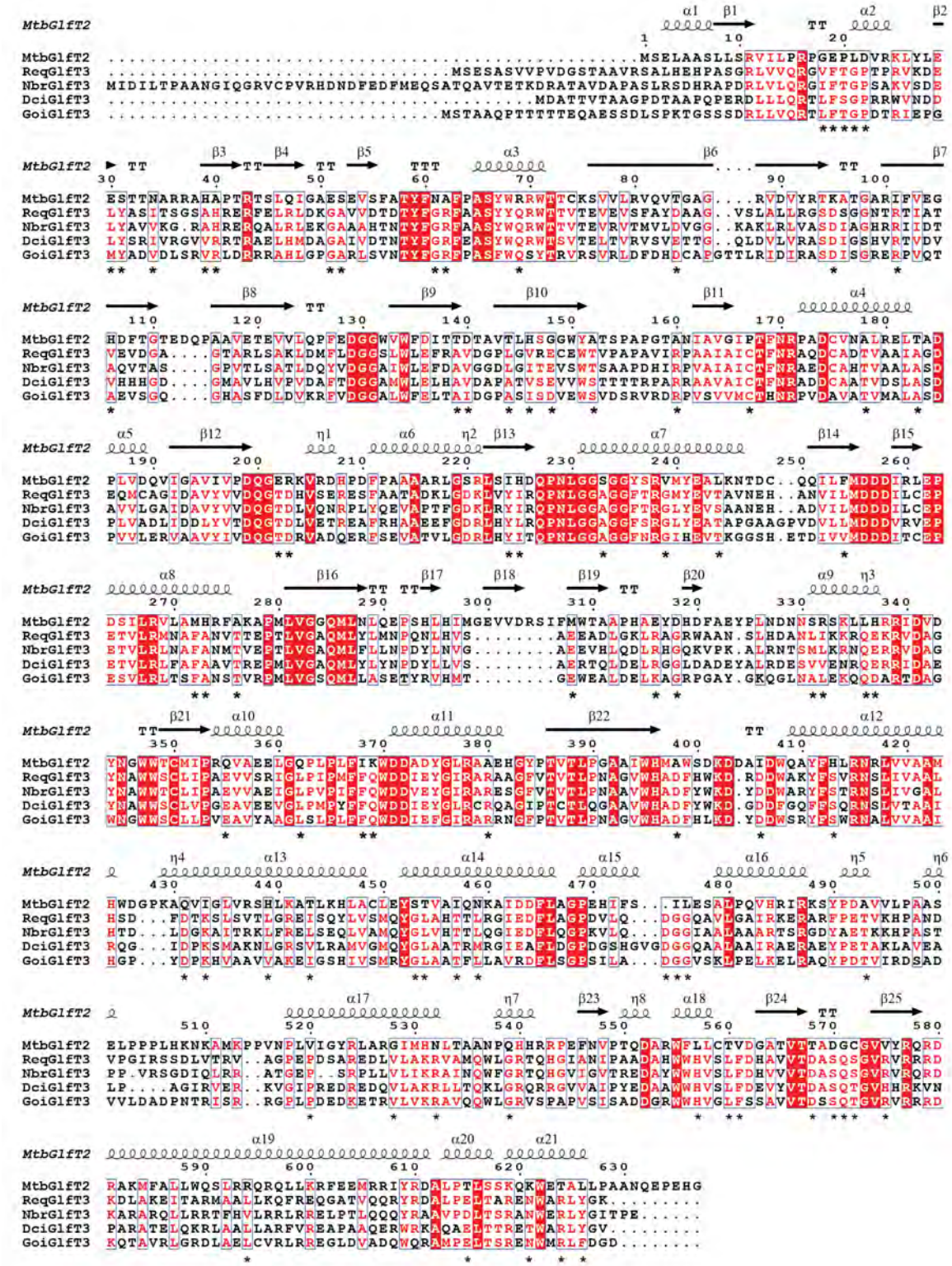

Figure 4.2: Sequence Alignment of GlfT2 and GlfT3. Sequence alignment of *Mtb*GlfT2 with GlfT3 from select species. Regions with GlfT3-specific similarities are indicated with an asterisk. Identical residues are show in red boxes, similar regions of similarity boxed and indicated with red text.

Sequence identities for *glfT3* genes had higher similarity between themselves than to their *glfT2* paralog (Figure 4.2). This is seen for the protein encoded by *glfT3* from Req (REQ_2470), which shares 62.3% sequence identity to its ortholog in Nbr (O3I_000895), compared to only 32.9% shared identity with the *glfT2* gene within its own genome. This is similarly seen for *glfT3* from Dcd (VYL96_01655), a more distantly-related Mycobacteriales species that likewise has 53.7% sequence identity to *glfT3* from Req.[126] This suggests that the enzyme encoded by *glfT3* likely performs a role or function distinct from that of GlfT2 and is under evolutionary pressure to conserve this function and the majority of its amino acid sequence.

4.3.2 *Req*GlfT3 and *Nbr*GlfT3 Display Distributive Gal*f* Transferase Activity

The *glfT3* sequences from Nbr and Req were added to vectors pAWC21 and pAWC22, respectively, for recombinant expression in *E. coli*. Each enzyme was expressed with an N-terminal His_6-tag for affinity purification (Figure 4.3a). Purification of both *Nbr*GlfT3 and *Req*GlfT3 was successful, with the best purity and yield seen for *Req*GlfT3.

Due to the overall sequence similarity and predicted structural homology of GlfT3 to GlfT2, we hypothesized that it would function as a galactofuranosyl transferase. This was tested with an *in vitro* reaction mixture containing the GlfT2 donor UDP-Gal*f* and two synthetic acceptor surrogates previously show to produce activity in GlfT2, the tri-Gal*f* **Compound 1** and the tetra-Gal*f* **Compound 2**. Both *Req*GlfT3 and *Nbr*GlfT3 produced extension products by Matrix-Assisted Laser Desorption Ionization Mass Spectrometry (MALDI-MS), revealing both enzymes to indeed function as galactofuranosyl transferases.

The product distributions for both enzymes showed extension products out to +4 Gal*f* units, with *Req*GlfT3 performing the most robustly. The higher effective concentration of the purer

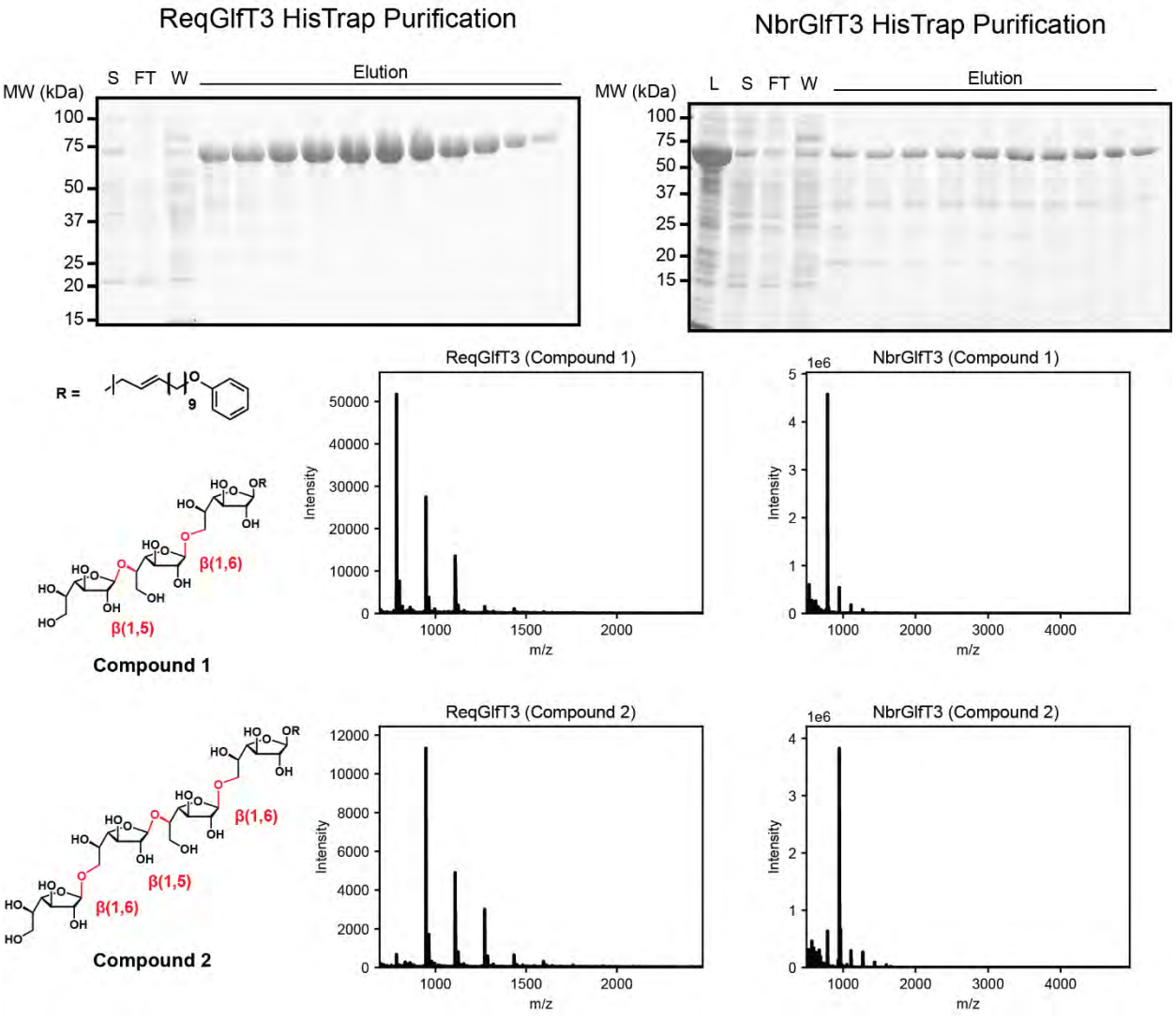

Figure 4.3: Purification and *In Vitro* Reactivity of GlfT3 from Req and Nbr. GlfT3 orthologs from Req and Nbr were purified by nickel affinity chromatography. Galactofuranosyl transferase activity was confirmed *in vitro* using the tri- and tetrasaccharide acceptor surrogates **Compound 1** and **2**. The product distributions were assessed by MALDI-TOF.

*Req*GlfT3 likely influenced the product abundance differences between both enzymes. The high abundance of short oligosaccharide product distributions indicate that both *Req*GlfT3 and *Nbr*GlfT3 likely function as distributive polymerases rather than processive polymerases like GlfT2.[19,45,53] This suggests that GlfT3 may function to add a single Gal*f* unit at a time to its oligosaccharide acceptor, rather than elongating the full product prior to unbinding.[44]

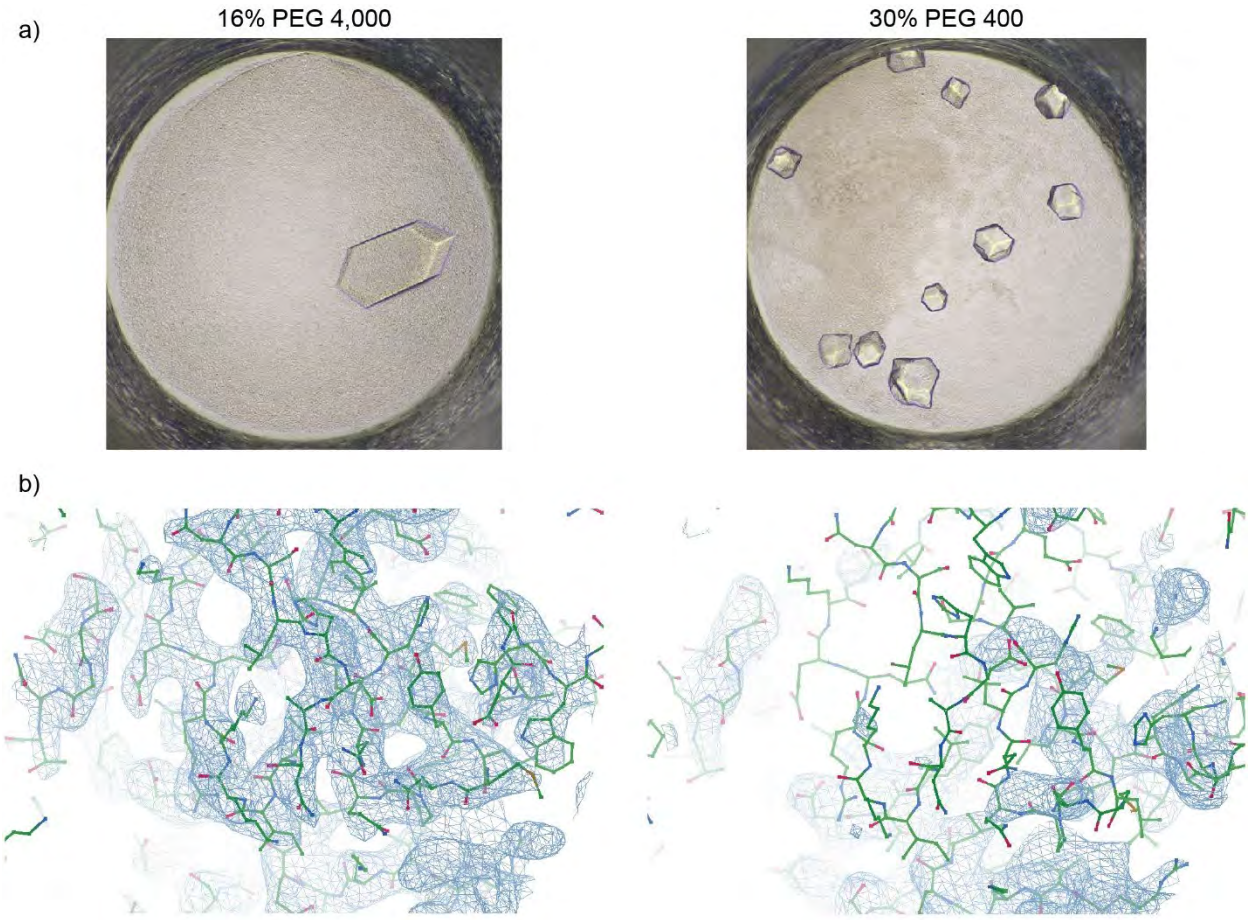

Figure 4.4: Systematic Absences Present in PEG 400 Crystallization Conditions. a) Representative images of crystal formations form screening conditions including PEG 4,000 and PEG 400. b) Electron density for PEG 4,000 and PEG 400 crystals after molecular replacement phasing and refinement. 2Fo-Fc density map contoured to 1.0 σ.

4.3.3 Crystallization Condition Optimization

Both *Req*GlfT3 and *Nbr*GlfT3 were screened for conditions conducive to crystal formation. Positive hits were seen for *Req*GlfT3, with robust crystal formation in conditions containing PEG 400 and PEG 4,000 as the precipitant (Figure 4.3a). These conditions were optimized and successfully led to crystal growth, with a three-dimension hexagonal prism and a cube morphology for PEG 400 and PEG 4,000, respectively.

Initial diffraction experiments revealed the concentration of PEG 4,000 used for crystallization was not cryoprotective, leading to loss of high-resolution diffraction.[127] PEG 400 successfully

Table 4.1: *Req*GlfT3 Crystallization Conditions and Crystal Lattice Space Groups

Precipitant Condition	Space Group
PEG 4K	$P3_2$
PEG 400	$P2_12_12$
PEG 4K + Ethylene Glycol	$C2$

cryoprotected crystals formed in its presence, but phasing and refinement revealed significant systematic absences in electron density for residues 305-352 (Figure 4.4). Cryoprotection of the PEG 4,000 condition was tested by additive soaking and additive cocrystallization. Short-duration additive soaking with PEG 400, glycerol, 2-methyl-2,4,pentanediol, ethylene glycol, or additional PEG 4,000 were screened. Most led to low-resolution, streaking diffraction spots, with additional PEG 4,000 resulting in diffraction with a resolution of 3.6 Å. Additive cocrystallization of ethylene glycol successfully led diffraction of 3.3 Å and was successfully used for structure determination. Successful phasing of some of these conditions reveals a change in lattice space group between PEG 400 and PEG 4,000, and in the presence of ethylene glycol, suggesting the losses in high-resolution diffraction may have come from incomplete space group rearrangement (Table 4.1).

4.3.4 Structural Characterization of *Req*GlfT3

Structural determination of *Req*GlfT3 revealed a kidney-shaped structure containing four domains highly similar to those shown for GlfT2 (Table 4.2).[35] The single-chain structure of *Req*GlfT3 allowed placement of 618 residues from residue 11 to 629. The protomeric structure consists of a central catalytic GT-A domain (Figure 4.5a), consisting of the metal binding DXD motif ($_{272}$DDD$_{274}$) required for UDP-sugar binding, and the catalytic DDX motif ($_{378}$DDI$_{380}$) require for sugar transfer to the acceptor substrate.[31,35] At the N-terminus, there is β-sandwich domain with a central 10-strand jelly roll topology. Following the GT-A domain, the α-helical

Table 4.2: *Req*GlfT3 Data Collection and Refinement Statistics (Molecular Replacement)

	*Req*GlfT3 (PEG 4K + EG)
Data collection	
Space group	*C*2
Cell dimensions	
a, *b*, *c* (Å)	413.725 112.006 144.864
α, β, γ (°)	90 89.6 90
Wavelength (Å)	0.92
Resolution (Å)	34.6-3.30 (3.42-3.30)
R_{sym} or R_{merge}	0.21 (2.33)
$I / \sigma I$	6.3 (0.7)
Completeness (%)	99.9 (99.8)
Redundancy	5.3 (5.5)
Refinement	
No. reflections	98,615 (9,753)
R_{work} / R_{free}	0.235 / 0.275
No. atoms	
Protein	38,127
Ligand/ion	0
Water	0
B-factors	
Protein	122.09
R.m.s. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.439

domain is split by a mixed α/β domain. The first section of the α-helical domain forms a four-helix bundle of antiparallel helices, with a short helix connecting the third and fourth helix. The second section of the α-helical domain folds onto the bundle with a bent helix. The mixed α/β domain is formed by two sequential β-α-β motifs and a single β-hairpin that interact to form a four-strand twisted β-sheet.

Contacts within the solved structure suggest the formation of a homotetrameric complex with C4 symmetry (Figure 4.5b). This is similar to the tetramer structure of GlfT2 and likewise is mediated by interactions between the β-sandwich domain and the mixed α/β domain. The tetramer has a large cavity within its tetrameric complex, with a large opening below the mixed α/β domains, and a smaller hole between the β-sandwich domains (Figure 4.5c). Additional openings into the cavity are present between the α-helical and mixed α/β domains.

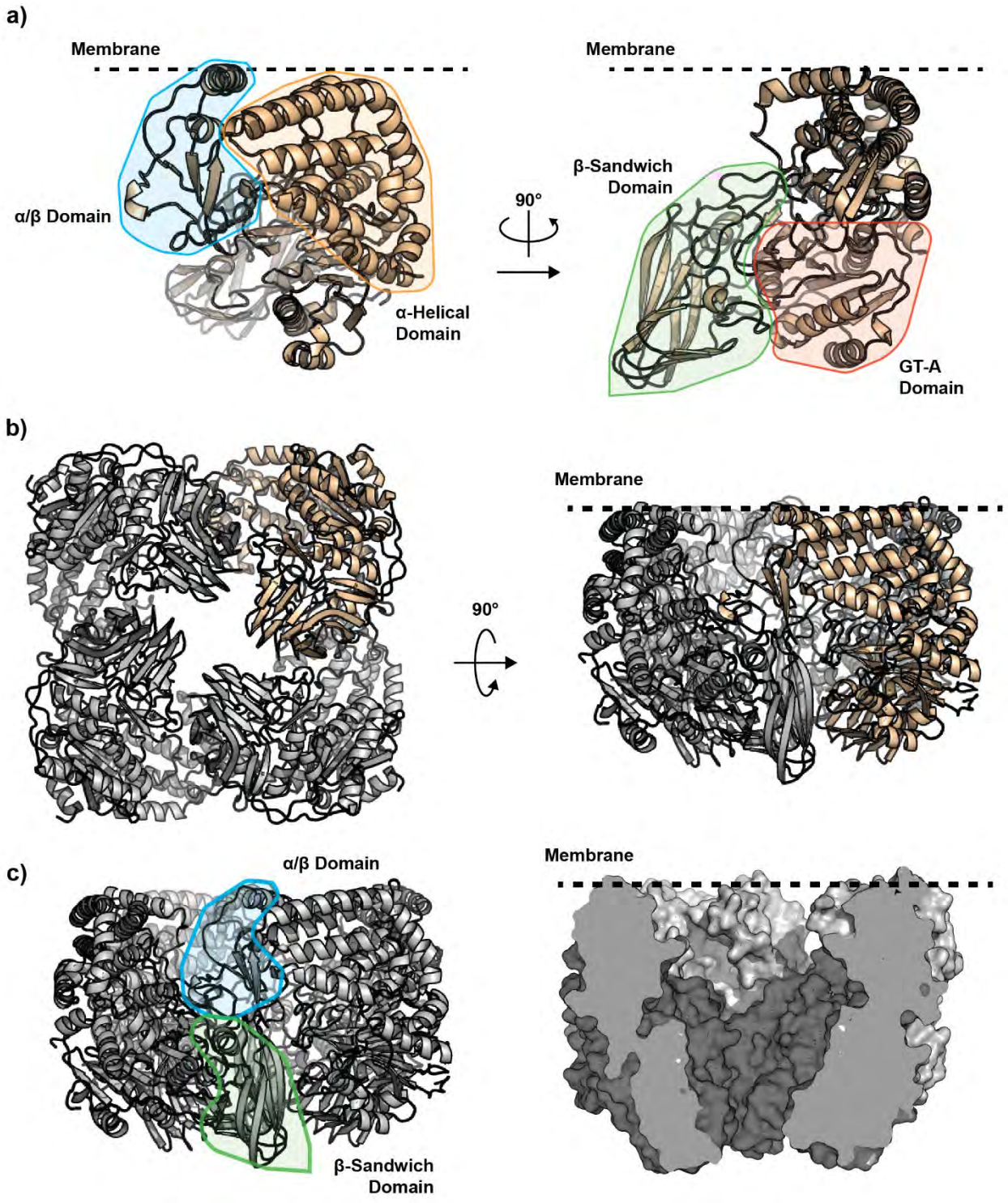

Figure 4.5: Structural Characterization of *Rhodococcus equi* GlfT3. a) Monomeric structure of *Req*GlfT3 with a predicted membrane boundary drawn. The four domains are highlighted: blue – mixed α/β domain; orange – α-helical domain; green – β-sandwich domain; red – GT-A domain. b) The tetrameric structure of GlfT3, with a single protomer highlighted. A predicted membrane boundary is drawn. c) The tetramer-forming interactions between the α/β domain and the β-sandwich domain are highlighted. The cavity formed by the tetramer is highlighted.

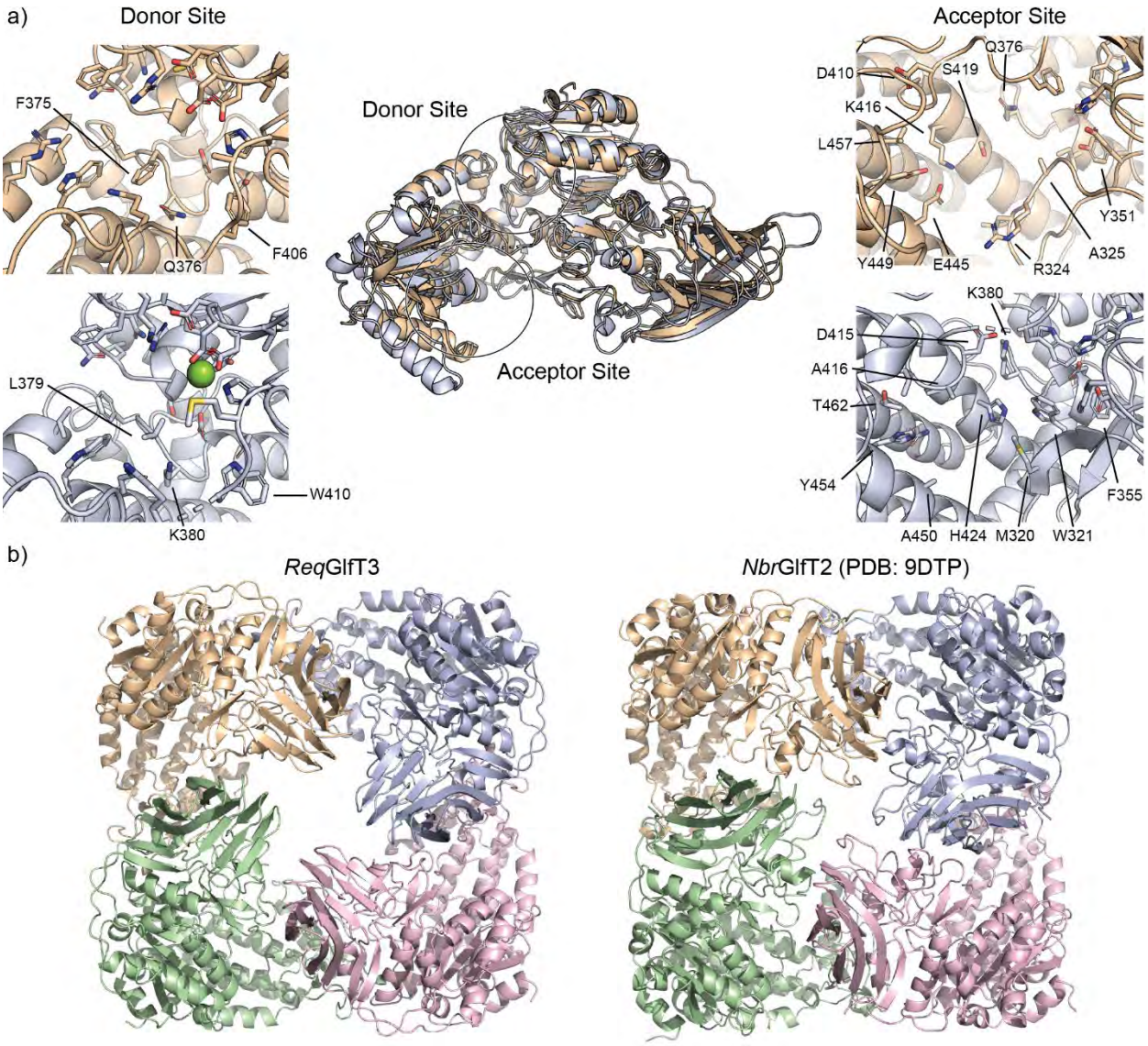

Figure 4.6: Residue and Structural Differences between *Req*GlfT3 and *Nbr*GlfT2. a) Superimposition of the *Req*GlfT3 and *Nbr*GlfT2 (PDB: 9DTP) protomers. Residue differences between both structures are highlighted within their respective donor and acceptor binding sites. b) Comparison of the tetrameric complexes for ReqGlfT3 and NbrGlfT2.

The donor site of *Req*GlfT3 was compared to that of *Nbr*GlfT2 (Chapter 2, 3) to confirm the use of similar mechanisms for their transferase activity (Figure 4.6a). This site consists of the DDD metal-binding motif, shown bound to magnesium in *Nbr*GlfT2 (PDB: 9DTP). Using the UDP-bound structure of *Mtb*GlfT2 (PDB: 4FIY) as a guide, residues within interacting distance to the donor were represented. Of the 22 residues within proximity to UDP, only three homologous

substitutions were identified: L379 to F375, K380 to Q376, and W410 to F406. The overall conservation of this site and the similarity of the substituted residues highlights the conserved Gal*f* transferase activity in GlfT3. To inform the potential reactivity of GlfT3 compared to GlfT2, the acceptor site of *Req*GlfT3 was compared to *Nbr*GlfT2. The residues within interacting distance to the *Nbr*GlfT2 acceptor were highlighted (Figure 4.6a), and significant homologous substitutions were identified. The *Nbr*GlfT2 residues W321 and H424 were shown to interact with the first and second Gal*f* of the galactan acceptor, and in *Req*GlfT3, these residues are substituted to A325 and S419, respectively. These residues were shown to be critical to GlfT2 processive polymerization and linkage control (Chapter 2).[35] These substitutions are expected to be involved in mediating the distributive mechanism of GlfT3 and its linkage production.

Comparison of the tetrameric structures of *Req*GlfT3 and *Nbr*GlfT2 reveal very similar C4-symmetric oligomerization (Figure 4.6b). The cavity of GlfT2 was previously predicted to mediate either galactan length or galactan sliding and polymerization (Chapter 3).[35] Unlike GlfT2, GlfT3 appears to produce short oligosaccharides through distributive polymerization. The purpose of conserving the tetrameric complex in GlfT3 is therefore unknown, and understanding its role in oligosaccharide biosynthesis by GlfT3 may inform its function in GlfT2.

4.3.5 *Req*GlfT3 is a Membrane-Associating Enzyme

Analysis of the biophysical properties of the solvent-exposed residues of the *Req*GlfT3 tetramer reveals enrichment for hydrophobic and basic amino acids along the larger cavity opening (Figure 4.7a). These solvent-exposed hydrophobic residues are displayed along three amphipathic helices, α1 of the first β-α-β motif of the mixed α/β domain, and α2 and α6 of the α-helical domain (Figure 4.7b). This is also seen in GlfT2 and was confirmed to mediate membrane association (Chapter 3). These amphipathic helices likely mediate similar membrane association

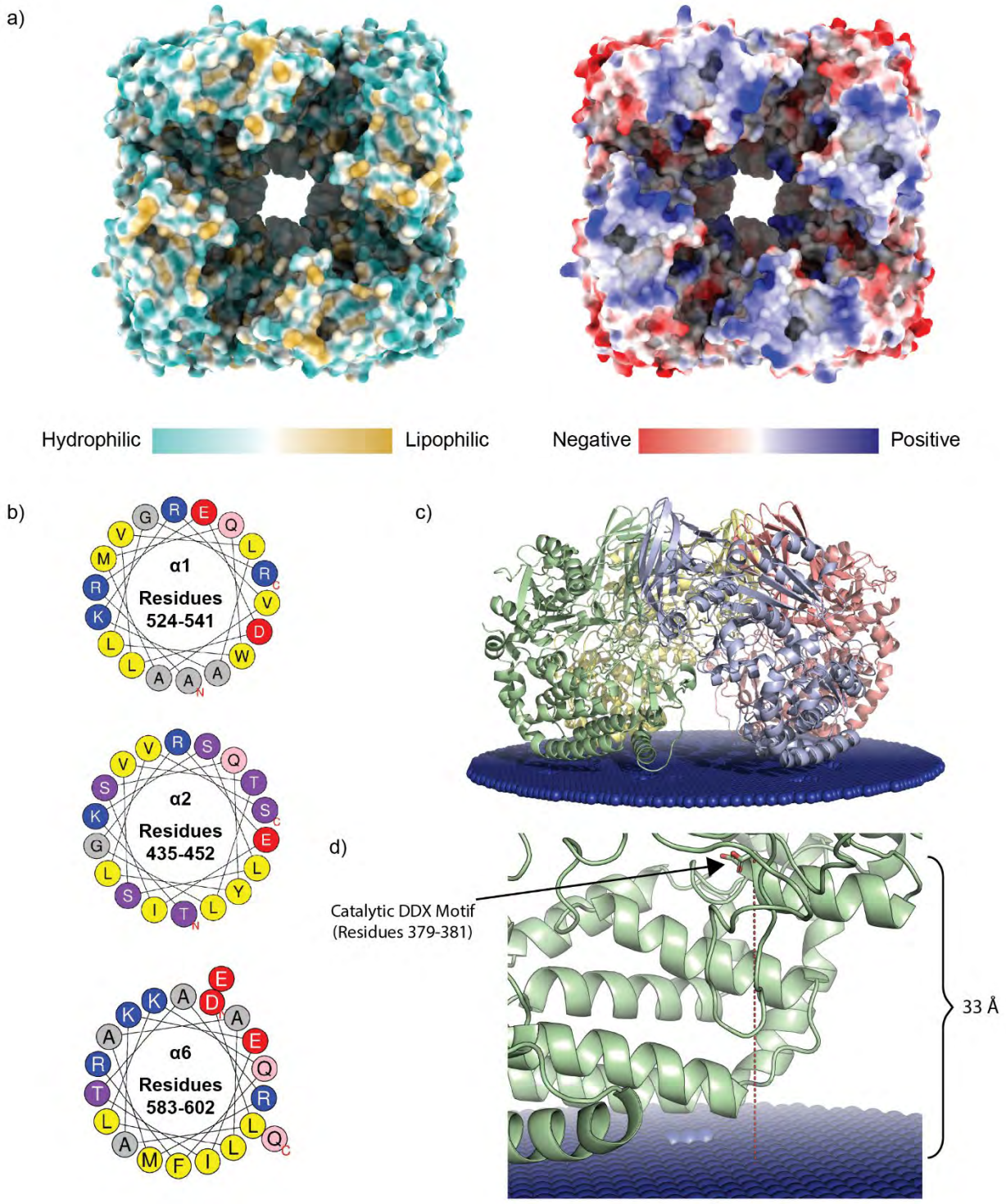

Figure 4.7: Predicted Membrane Association by *Req*GlfT3. a) Apparent enrichment of hydrophobic and basic amino acids along the cavity opening of *Req*GlfT3 are highlighted. Residues are colored as follows: cyan – hydrophilic; beige – lipophilic; red – negatively charged; blue – positively charged. b) Helical wheels for the predicted membrane-binding amphipathic helices. c) Predicted ReqGlfT3 membrane binding predicted by the PPM Server. d) Distance measure between the catalytic aspartate D380 and the membrane boundary.

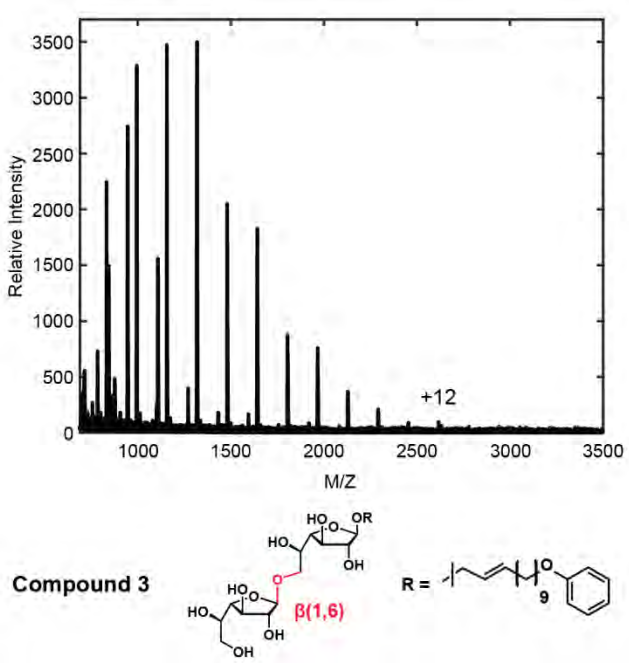

Figure 4.8: Product Distribution of *Req*GlfT3 One-Pot Reaction. Extension products of *Req*GlfT3 in a one-pot reaction with **Compound 3**. Products were measured up to 12 additional Gal*f* residues, with a +48 m/z modification.

for GlfT3, localizing the enzyme to the membrane interface. This suggests *Req*GlfT3 likely recognizes a glycolipid acceptor substrate, as with GlfT2.[35] Orientation optimization of static *Req*GlfT3 in the presence of an aqueous-organic boundary representing a membrane interface indeed predicts its membrane association (Figure 4.7c).[32] The distance from this boundary to the catalytic residue D380 was nearly 33 Å (Figure 4.7d), suggesting GlfT3 may recognize an oligosaccharide acceptor already primed with multiple sugars in order to reach its active site. This distance is similar to GlfT2 and may suggest both GlfT2 and GlfT3 can recognize the same acceptor substrate.

4.3.6 Linkage Analysis of *Req*GlfT3 Product Formation

To inform the Gal*f* linkages produced by *Req*GlfT3, an *in vitro* one-pot reaction system was developed and optimized, producing linkages up to an estimated 12 additional Gal*f* residues (Figure 4.8). **Compound 3**, a synthetic disaccharide acceptor, was used to reduce the background signal of acceptor-substrate sugars.[19] Due to the length of the reaction incubation, a modification appears to have shifted the product masses by +48 m/z (Figure 4.8). The optimized extension distribution exhibits a Poisson distribution for products, suggesting a distributive polymerization mechanism.[44]

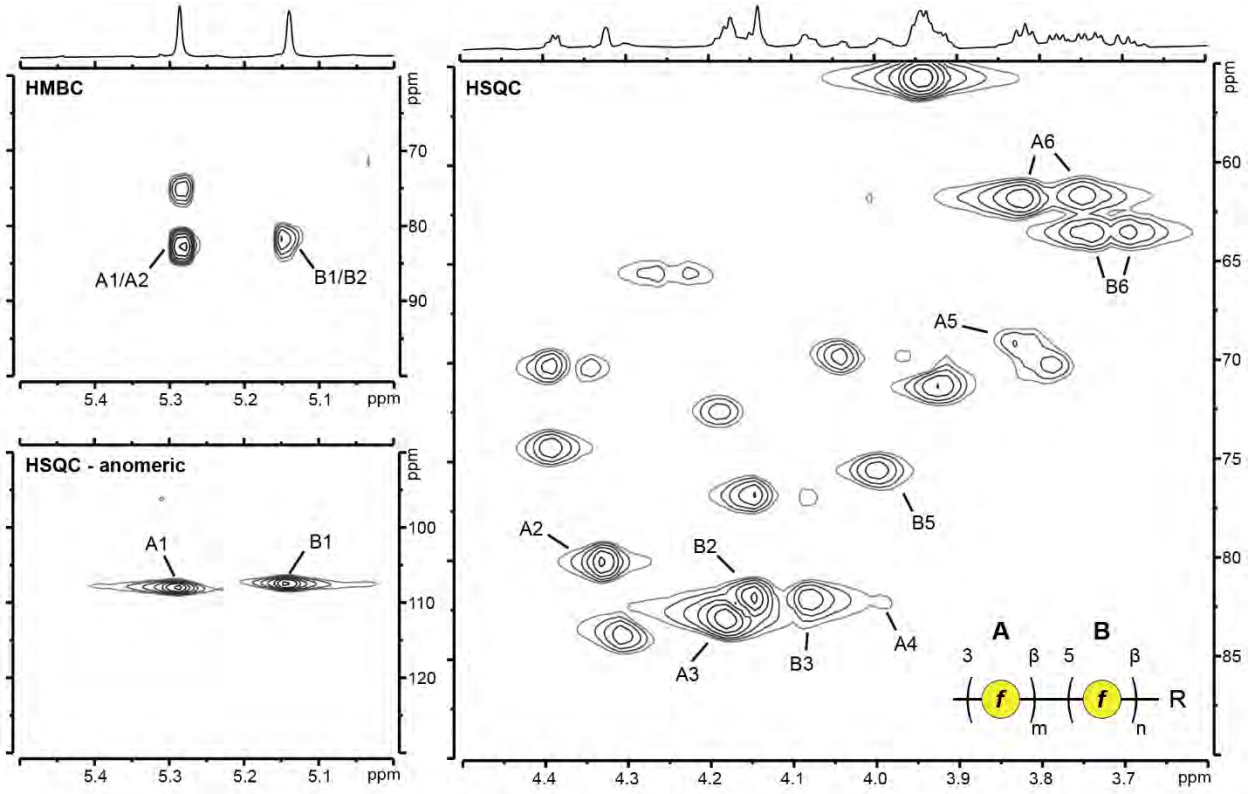

Figure 4.9: *Req*GlfT3 Product Linkage Analysis by HSQC and HMBC. HSQC and HMBC spectra show the tentative assignments for Gal*f* products produced by *Req*GlfT3. The assignments are numbered for each carbon position in Gal*f*. Unit **A** represents a 3-modified Gal*f*, and unit **B** represents a 5-modified Gal*f*.

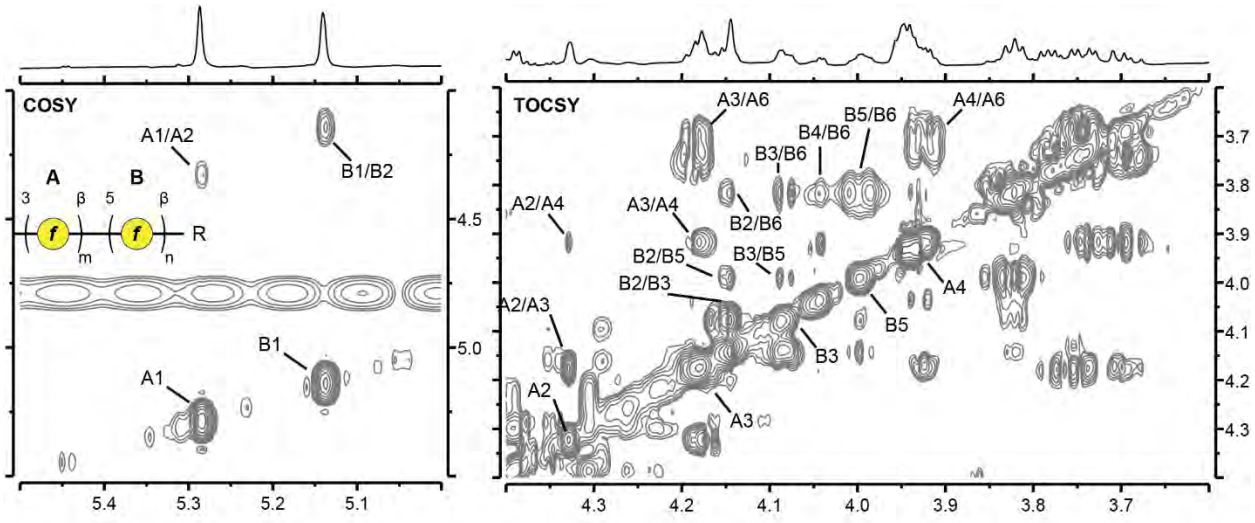

Figure 4.10: *Req*GlfT3 Product Linkage Analysis by COSY and TOCSY. COSY and TOCSY spectra with tentative assignments for the Gal*f* linkages produced by *Req*GlfT3. Cross-peaks are labeled for intra-sugar correlations. Unit **A** represents a 3-modified Gal*f*, and unit **B** represents a 5-modified Gal*f*.

Table 4.3: NMR Resonance Assignments for Product Linkage Analysis

Residue	Linkage		1	2	3	4	5	6
A	→3-Galf-1→	^{1}H	5.29	4.33	4.17	3.92	3.82	3.73, 3.69
		^{13}C	108.1	80.2	83.2	82.5	69.2	63.6
B	→5-Galf-1→	^{1}H	5.14	4.14	4.09	4.00	3.99	3.83, 3.75
		^{13}C	107.4	82.0	82.2	82.2	75.6	61.6

Nuclear magnetic resonance (NMR) experiments were performed to determine the Gal*f* linkages within the *Req*GlfT3 products. ^{1}H resonances for the anomeric positions of sugars are often used for monosaccharide unit identification, and ^{1}H peaks at 5.29 and 5.14 ppm were within the expected resonance range.[19,42,49] A Heteronuclear Single Quantum Coherence (HSQC) experiment identified peaks anomeric ^{13}C resonances at 108.1 and 107.4 ppm (Figure 4.9). A Heteronuclear Multiple Bond Correlation (HMBC) experiment identified the cross peaks for the anomeric positions and the intra-saccharide C2 correlation. Using these initial identifications, Correlated Spectroscopy (COSY) and Total Correlation Spectroscopy (TOCSY) experiments identified connected resonances (Figure 4.10), which had their ^{13}C chemical shifts assigned by HSQC. Resonance for C4 of the 5-modified Gal*f* was not assigned. The resonances are listed in Table 4.3, and suggest the production of β(1,3) and β(1,5) linkages.

4.4 Discussion

Here we described the identification of a novel GlfT2 paralog, termed GlfT3, that may influence galactan linkage patterns in Mycobacteriales species outside the families Mycobacteriaceae and Corynebacteriaceae. Two GlfT3 orthologs from Req and Nbr were expressed, purified, and confirmed to have galactofuranosyl transferase activities *in vitro*, appearing to perform

distributive polymerization.[44] Structural characterization of *Req*GlfT3 revealed an enzyme with structural homology to GlfT2 and a highly similar UDP-Gal*f* binding site.[35] The acceptor-binding site of *Req*GlfT3 reveals substitutions at positions known to be critical for GlfT2 activity, and may suggest different reactivities (Chapter 2).[35] Distance measurement from the active site to the predicted membrane suggest GlfT3 recognizes a oligosaccharide, and its similarity to GlfT2 suggests it may bind the same acceptor used by GlfT2 in the galactan biosynthetic pathway. *In vitro* product extension and linkage analysis indicates the formation of β(1,3) and β(1,5) linkages by *Req*GlfT3. From these insights, we propose that GlfT3 may be a new enzyme involved in the galactan biosynthetic pathway, likely catalyzing the addition of new Gal*f* linkages onto the galactan polysaccharide.

Additional work is required to determine the role GlfT3 plays within the Mycobacteriales species identified (Figure 4.1), either as an essential enzyme critical to cell wall biosynthesis or as a nonessential enzyme used to modify and modulate the galactan. If it plays a non-essential role, GlfT3 may have undergone evolutionary pressure that may allow for the formation of new linkages across different bacterial species. Understanding its expression patterns under different stress conditions and the effects the new linkages produce may further inform the role the galactan linkage pattern plays within the cell wall.

The structure of *Req*GlfT3 provides additional insight into the activities of GlfT2-like enzymes across the bacterial kingdom. Recent investigation of the gram-negative bacterium *Salmonella enterica* revealed its T1 antigen polysaccharide to be biosynthesized in part by a GlfT2 ortholog, *STM0724*.[49] This enzyme was shown to catalyze Gal*f* transfer using UDP-Gal*f*, producing both β(1,3) and β(1,6) linkages. A number of other non-mycobacterial species were shown to include GlfT2 orthologs, all with variable Gal*f* linkage compositions.[49] These results indicate that the

GlfT2 architecture may be present across diverse bacteria and function within a variety of pathways to produce different polysaccharide products. Comparison of these enzymes to the structures and activities of GlfT2 and GlfT3 is expected to inform their mechanisms and regioselectivity.

Future research into *Req*GlfT3 will inform the mechanism of its galactan linkage production. Here, we show the production of β(1,3) and β(1,5) linkages by *Req*GlfT3. Previous arabinogalactan linkage analysis in Req indicated the production of →5-Gal*f*-1→5-Gal*f*-1→3, →5-Gal*f*-1→3-Gal*f*-1→5, and →5-Gal*f*-1→3 motifs. These suggest a potentially random linkage production, rather than the high-fidelity linkage sequence produced by GlfT2.[10] Structural characterization of *Req*GlfT3 with its acceptor-substrates bound may provide insight into its mechanism of regioselectivity. Comparison of these structures to the acceptor-bound structures of *Nbr*GlfT2 will inform the evolutionary changes that dictate Gal*f* linkage production by this class of enzyme.

In summary, we provide the first experimental insight into the structure and activity of GlfT3 from the organism *Rhodococcus equi*. GlfT3 from Req and Nbr were confirmed to be Gal*f* transferases, recognizing a Gal*f*-containing acceptor and functioning to produce an oligosaccharide product through distributive polymerization. Linkage analysis indicates the production of β(1,3) and β(1,5) linkages by *Req*GlfT3. The structure of *Req*GlfT3 reveals an enzyme with homology to GlfT2, forming a tetrameric structure that likely associates to the membrane interface. We believe the structure and activity of *Req*GlfT3 described here will inform future work into GlfT's across the bacterial kingdom and provide insight into their functions and polysaccharide products.

4.5 Methods

4.5.1 Plasmid Preparation and Mutagenesis

DNA sequences for *Rhodococcus equi* GlfT3 (REQ_02470) and *Nocardia brasiliensis* GlfT3 (O3I_000895) were codon optimized for expression in *E. coli* and purchased from Integrated DNA Technologies. Gene incorporation into the pET28a vector was prepared using In Vitro Assembly by mixing the PCR amplified vector backbone (primers from Table 4.1) and the DNA gblock, using chemically-competent *E. coli* DH5α (New England Biolabs). Proper incorporation was confirmed using Sanger Sequencing. Mutagenesis of the produced plasmids were produced using In Vitro Assembly with primers from Table 4.1 designed to have nucleotide substitutions in the homologous-sequence region.

4.5.2 Protein Expression and Purification

Protein was expressed using an autoinduction protocol, using terrific broth (TB) supplemented with 150 ug/mL of kanamycin as base media.[75,76] Cells were grown on a shaker in 0.5 L TB and 150 μg/mL of kanamycin to an optical density (600 nm) of 0.8 at 37 ˚C, and cooled to 18 ˚C for ~20 hour protein expression. Cells were pelleted by centrifugation at 4,000 xg for 10 minutes at 4 ˚C, and the pellets stored at -80 ˚C.

Cells were resuspended in Bacterial Protein Extraction Reagent (Thermo Fisher Scientific) at 5 mL/g, along with 1 mg/mL Egg White Lysozyme (Gold Biotechnology), 0.1% Triton X-100 (Sigma Aldrich), and 5 units of Benzonase Nuclease (EMD Millipore). The suspension was incubated for 30 minutes at room temperature with gently mixing, and centrifuged at 24,000 xg for 1 hour at 4 ˚C. The supernatant was filtered with a Millex-GP 0.22 μm PES Syringe Filter

(MilliporeSigma), and injected onto a HisTrap HP column (Cytiva). Purification was accomplished using a NGC FPLC (Biorad) at 4 ˚C with a 0% to 100% gradient of Buffer A (50

Table 4.4: Gene Accessions Used in Phylogenetic Analysis

Organism	***glfT2***	***glfT3***
Dietzia cinnamea	VYL96_01550	VYL96_01655
Dietzia papillomatosis	WP_061916275.1	WP_061916240.1
Dietzia natronolimnaea	WP_095717071.1	WP_182645776.1
Dietzia maris	KZO58818.1	KZO58835.1
Gordonia otitidis	LK459_07830	LK459_07675
Gordonia sputi	MFF0708407.1	MFF0708440.1
Gordonia terrae	BCM27_01520	
Gordonia bronchialis	Gbro_0205	
Rhodococcus equi	REQ_01940	A0A3S5Y1I6
Rhodococcus erythropolis	RER_02120	RER_02550
Rhodococcus rhodochrous	SNV08814.1	OOL32513.1
Rhodococcoides fascians	MBY4404039.1	MBY4205797.1
Nocardia abscessus	WP_043684864.1	WP_043685134.1
Nocardia farcinica	WP_280505208.1	WP_280502878.1
Nocardia brasiliensis	O3I_000690	O3I_000895
Nocardia asteroides	WP_233015472.1	WP_233015504.1
Mycolicibacterium smegmatis	MSMEG_6403	
Mycobacterium marinum	MMAR_5372	
Mycobacterium leprae	ML0093	

Mycobacterium tuberculosis	Rv3808c
Mycobacterium avium	MAV_0209
Corynebacterium jeikeium	Jk0130
Corynebacterium glutamicum	Cgl2882
Corynebacterium diphtheriae	DIP2198

mM HEPES pH 7.8, 300 mM NaCl, 20 mM imidazole) and Buffer B (50 mM HEPES pH 7.8, 300 mM NaCl, 500 mM imidazole). Purified protein was pooled and dialyzed against Buffer C (50 mM HEPES pH 7.4, 100 mM NaCl) with 5 mM EDTA and 5mM BME using a 20k Da molecular weight cutoff Slide-A-Lyzer Dialysis Cassette (Thermo Fisher Scientific). Dialysis was then moved to Buffer C + 5% glycerol. Protein was flash frozen in liquid nitrogen and stored at -80 ˚C.

4.5.3 Phylogenetic Analysis

Representative organisms from within Mycobacteriales were selected for phylogenetic analysis. The amino acid sequence of *Nbr*GlfT2 was used in the Basic Local Alignment Search Tool (BLAST) for sequence similarity searches for each organism. Confirmed GlfT2-paralog sequences were identified as GlfT3, and their three-dimensional structural similarity was confirmed by Alphafold2 structural prediction (Table 4.2).[28,29,114] Primary sequence alignment was performed in SnapGene 7.2.1. The NCBI Taxonomy database was used to phylogenetic grouping of the selected organisms.[128,129] The phylogenetic tree was visualized in Interactive Tree of Life (iTOL) online tool.[130] Genome neighborhoods and gene annotations were adapted from the Kyoto Encyclopedia of Genes and Genomes (KEGG) knowledge base.[131]

4.5.4 *In Vitro* Activity Assay by MALDI-TOF

Elongation reactions were performed in 10 µL volumes containing 0.2 µM GlfT3, 1.25 mM UDP-Gal*f*, and 200 µM acceptor substrate in 50 mM HEPES pH 7.4 and 100mM NaCl. Reactions proceeded overnight at room temperature. The reaction mixtures were mixed to contain 0.05% trifluoroacetic acid (TFA) and loaded onto Hypersep Hypercarb Tips (10 µL, Thermo Fisher Scientific). The tips were washed with 0.05% TFA in 1:3 acetonitrile:water, and then eluted in 70% methanol saturated with α-cyano-4-hydroxycinnamic acid. Mass spectra of galactan products were acquired using a Bruker Autoflex LRF Speed MALDI-TOF in positive reflectron mode.

4.5.5 Protein Crystallography

Protein crystallization conditions were screened with the National Crystallization Center at the Hauptman-Woodward Medical Research Institute. The protein stock of *Req*GlfT3 was concentrated to 5 mg/mL and used in a high-throughput screened across conditions designed for membrane proteins. Protein crystallization was monitored for 6 weeks, and protein crystallization was confirmed by Ultraviolet-Two Photon Excited Fluorescence (UV-TPEF) imaging of intrinsic protein fluorescence.

Protein samples were diluted with Buffer C + 0.02% Triton X-100 to final concentrations of 6-7 mg/mL protein. The sample was dialyzed in a 20 kDa cutoff Slide-A-Lyzer MINI (Invitrogen) for 20 minutes against 500 mL Buffer C + 0.02% Triton X-100 to remove glycerol. Crystallization was achieved using the hanging drop method. Proteins used for diffraction collection were crystallized with a mother liquor consisting of 0.1 M Tris pH 8.0, 0.4 M NaCl,

15-17.5% PEG 4,000, and 6-10.5% ethylene glycol. Crystallization drops made as a 2 µL mixture of 1:1 protein:mother liquor, and stored at 20 °C. Crystals formed within 1-7 days with a three-dimensional cube-like morphology. Crystals were harvested and flash frozen in liquid nitrogen.

4.5.6 Crystal Diffraction, Data Phasing, Refinement, and Modeling

Diffraction data for *Req*GlfT3 was collected at the NSLSII (National Synchrotron Light Source II, Brookhaven National Laboratory) using the AMX beamline 17-ID-1 at a wavelength of 0.920 Å. The protein crystals diffracted to 3.3 Å with the space group *C*2, containing eight protomer chains of *Req*GlfT3. The structure was solved by molecular replacement in Phaser using an Alphafold2-prediction of *Req*GlfT3 processed with PHENIX 1.21.2-5419.[28,78] The initial atomic model was refined in PHENIX Refine.[79] Manual real-space refinement was performed in Coot 0.9.8.95. The refined protein model was validated using MolProbity.[80] Protein structure visualization performed in PyMOL.[81] Protein oligomer interfaces were calculated with the PISA server.[64]

Ramachandran statistics for the protein models were as follows (Favored, Allowed, Outliers):
Apo – 92.3%, 7.21%, 0.51%

4.5.7 One-Pot *Req*GlfT3 Reaction and Linkage Analysis

One-pot reactions with *Req*GlfT3 were performed in Buffer C (50 mM HEPES, 100 mM NaCl, pH 7.4), with final concentrations of 25 mM $MgCl_2$, 27 mM UDP-Gal*p*, 20 mM dithionite, 1 µM UGM, 0.2 mM **Compound 3**, and 0.2 µM *Req*GlfT3. UGM and *Req*GlfT3 concentrations were determined by UV absorbance at 450 nm and 280 nm, respectively. Dithionite was prepared fresh. All aqueous solutions other than dithionite were degassed by sonication. The reaction

mixture was prepared with 10x Buffer C, $MgCl_2$, and UDP-Gal*p* mixed to a final volume of 3 mL. UGM and dithionite were added, placed under an argon blanket, and incubated for 1 hour. **Compound 3** and *Req*GlfT3 were added, the argon blanket was reestablished, sealed, and incubated at room temperature for 5 days.

The reaction products were purified by Liquid Chromatography Mass Spectrometry (LCMS) using a C18 column. After injection, 15% acetonitrile in water with 0.1% TFA was run for 10 minutes, then ramped to 85% acetonitrile with 0.1% TFA over a 35-minute period. Final column flushing was performed with acetonitrile + 0.1% TFA.

NMR acquisition was performed on a Bruker Avance Neo 600 MHz spectrometer at 298.1 K. HSQC was measured on within an acquisition time of 0.143 seconds with a spectral with of 11.9 ppm along the ^{1}H and 169.9 ppm along ^{13}C, over 64 scans. The HMBC was run with a 0.143 second acquisition time, with a 5.86 ppm and 220.8 ppm spectral width along ^{1}H and ^{13}C, respectively, over 64 scans. COSY was measured over 16 scans, with a 1.85 second delay and 0.29 second acquisition time over a 5.8 ppm spectral width along ^{1}H. TOCSY was run over 32 scans, with a relaxation delay of 1.79 seconds and a 0.38 second acquisition time, over 32 scans.

4.6 Acknowledgements

Crystallization screening at the National Crystallization Center at HWI was supported through NIH grant R24GM141256. The Department of Chemistry Instrumentation Facility assisted in collection of MALDI-TOF data. We acknowledge the beamline at NSLS-II for their assistance in X-ray diffraction collection. This research used resources of the National Synchrotron Light Source II, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under Contract No. DE-SC0012704.

The Center for BioMolecular Structure (CBMS) is primarily supported by the National Institutes of Health, National Institute of General Medical Sciences (NIGMS) through a Center Core P30 Grant (P30GM133893), and by the DOE Office of Biological and Environmental Research (KP1605010).

References

1. Boshoff, H. I. M., Warner, D. F. & Gold, B. Editorial: Drug-resistant Mycobacterium tuberculosis. *Front. Cell. Infect. Microbiol.* **13**, 1215294 (2023).
2. Bagcchi, S. WHO's Global Tuberculosis Report 2022. *The Lancet Microbe* **4**, e20 (2023).
3. Bhat, R. M. & Prakash, C. Leprosy: An Overview of Pathophysiology. *Interdisciplinary Perspectives on Infectious Diseases* **2012**, 1–6 (2012).
4. Kiehn, T. E. *et al.* Infections caused by Mycobacterium avium complex in immunocompromised patients: diagnosis by blood culture and fecal examination, antimicrobial susceptibility tests, and morphological and seroagglutination characteristics. *J Clin Microbiol* **21**, 168–173 (1985).
5. Hadfield, T. L., McEvoy, P., Polotsky, Y., Tzinserling, V. A. & Yakovlev, A. A. The Pathology of Diphtheria. *J INFECT DIS* **181**, S116–S120 (2000).
6. Abrahams, K. A. & Besra, G. S. Mycobacterial cell wall biosynthesis: a multifaceted antibiotic target. *Parasitology* **145**, 116–133 (2018).
7. Jankute, M., Cox, J. A. G., Harrison, J. & Besra, G. S. Assembly of the Mycobacterial Cell Wall. *Annu. Rev. Microbiol.* **69**, 405–423 (2015).
8. Mikušová, K. *et al.* Biosynthesis of the Galactan Component of the Mycobacterial Cell Wall. *Journal of Biological Chemistry* **275**, 33890–33897 (2000).
9. Alderwick, L. J., Harrison, J., Lloyd, G. S. & Birch, H. L. The Mycobacterial Cell Wall—Peptidoglycan and Arabinogalactan. *Cold Spring Harb Perspect Med* **5**, a021113 (2015).
10. Yamatsugu, K., Splain, R. A. & Kiessling, L. L. Fidelity and Promiscuity of a Mycobacterial Glycosyltransferase. *J. Am. Chem. Soc.* **138**, 9205–9211 (2016).
11. Beláňová, M. *et al.* Galactosyl Transferases in Mycobacterial Cell Wall Synthesis. *J Bacteriol* **190**, 1141–1145 (2008).

12. Kremer, L. *et al.* Galactan Biosynthesis in Mycobacterium tuberculosis. *Journal of Biological Chemistry* **276**, 26430–26440 (2001).

13. Mikušová, K. *et al.* Identification of a Novel Galactosyl Transferase Involved in Biosynthesis of the Mycobacterial Cell Wall. *J Bacteriol* **188**, 6592–6598 (2006).

14. Savková, K. *et al.* An ABC transporter Wzm–Wzt catalyzes translocation of lipid-linked galactan across the plasma membrane in mycobacteria. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023663118 (2021).

15. Huszár, S. *et al. N* -Acetylglucosamine-1-Phosphate Transferase, WecA, as a Validated Drug Target in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* **61**, e01310-17 (2017).

16. Martinez Farias, M. A., Kincaid, V. A., Annamalai, V. R. & Kiessling, L. L. Isoprenoid Phosphonophosphates as Glycosyltransferase Acceptor Substrates. *J. Am. Chem. Soc.* **136**, 8492–8495 (2014).

17. Grzegorzewicz, A. E. *et al.* Development of a microtitre plate-based assay for lipid-linked glycosyltransferase products using the mycobacterial cell wall rhamnosyltransferase WbbL. *Microbiology* **154**, 3724–3730 (2008).

18. Mills, J. A. *et al.* Inactivation of the Mycobacterial Rhamnosyltransferase, Which Is Needed for the Formation of the Arabinogalactan-Peptidoglycan Linker, Leads to Irreversible Loss of Viability. *Journal of Biological Chemistry* **279**, 43540–43546 (2004).

19. May, J. F., Splain, R. A., Brotschi, C. & Kiessling, L. L. A tethering mechanism for length control in a processive carbohydrate polymerization. *Proceedings of the National Academy of Sciences* **106**, 11851–11856 (2009).

20. Dulberger, C. L., Rubin, E. J. & Boutte, C. C. The mycobacterial cell envelope — a moving target. *Nat Rev Microbiol* **18**, 47–59 (2020).

21. Jin, Y., Xin, Y., Zhang, W. & Ma, Y. Mycobacterium tuberculosis Rv1302 and Mycobacterium smegmatis MSMEG___4947 have WecA function and MSMEG__4947 is required for the growth of M. smegmatis: Mycobacterial WecA. *FEMS Microbiology Letters* **310**, 54–61 (2010).

22. Bosch, B. *et al.* Genome-wide gene expression tuning reveals diverse vulnerabilities of *M. tuberculosis*. *Cell* **184**, 4579-4592.e24 (2021).

23. Griffin, J. E. *et al.* High-Resolution Phenotypic Profiling Defines Genes Essential for Mycobacterial Growth and Cholesterol Catabolism. *PLoS Pathog* **7**, e1002251 (2011).

24. DeJesus, M. A. *et al.* Comprehensive Essentiality Analysis of the *Mycobacterium tuberculosis* Genome via Saturating Transposon Mutagenesis. *mBio* **8**, e02133-16 (2017).

25. Lehrer, J., Vigeant, K. A., Tatar, L. D. & Valvano, M. A. Functional Characterization and Membrane Topology of *Escherichia coli* WecA, a Sugar-Phosphate Transferase Initiating the Biosynthesis of Enterobacterial Common Antigen and O-Antigen Lipopolysaccharide. *J Bacteriol* **189**, 2618–2628 (2007).

26. Lehrman, M. A. Commentary: A family of UDP-GlcNAc/MurNAc: polyisoprenol-P GlcNAc/MurNAc-1-P transferases. *Glycobiology* **4**, 768–771 (1994).

27. Chung, B. C. *et al.* Crystal Structure of MraY, an Essential Membrane Enzyme for Bacterial Cell Wall Synthesis. *Science* **341**, 1012–1016 (2013).

28. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

29. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* **50**, D439–D444 (2022).

30. Hakulinen, J. K. *et al.* MraY–antibiotic complex reveals details of tunicamycin mode of action. *Nat Chem Biol* **13**, 265–267 (2017).

31. Breton, C., Šnajdrová, L., Jeanneau, C., Koča, J. & Imberty, A. Structures and mechanisms of glycosyltransferases. *Glycobiology* **16**, 29R-37R (2006).

32. Lomize, A. L., Todd, S. C. & Pogozheva, I. D. Spatial arrangement of proteins in planar and curved membranes by PPM 3.0. *Protein Science* **31**, 209–220 (2022).

33. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Research* **40**, D370–D376 (2012).

34. Chauhan, A. *et al.* GlfT1 down-regulation affects Mycobacterium tuberculosis biofilm formation and its in-vitro and in-vivo survival. *Tuberculosis* **141**, 102352 (2023).

35. Wheatley, R. W., Zheng, R. B., Richards, M. R., Lowary, T. L. & Ng, K. K. S. Tetrameric Structure of the GlfT2 Galactofuranosyltransferase Reveals a Scaffold for the Assembly of Mycobacterial Arabinogalactan. *J. Biol. Chem.* **287**, 28132–28143 (2012).

36. Pan, F., Jackson, M., Ma, Y. & Mcneil, M. Cell Wall Core Galactofuran Synthesis Is Essential for Growth of Mycobacteria. *J. BACTERIOL.* **183**, 9 (2001).

37. Justen, A. M. *et al.* Polysaccharide length affects mycobacterial cell shape and antibiotic susceptibility. *Sci Adv* **6**, eaba4015 (2020).

38. Pathak, A. K. *et al.* Studies on (β,1→5) and (β,1→6) linked octyl Galf disaccharides as substrates for mycobacterial galactosyltransferase activity. *Bioorganic & Medicinal Chemistry* **9**, 3129–3143 (2001).

39. Rose, N. L. *et al.* Expression, Purification, and Characterization of a Galactofuranosyltransferase Involved in *Mycobacterium t uberculosis* Arabinogalactan Biosynthesis. *J. Am. Chem. Soc.* **128**, 6721–6729 (2006).

40. Marlow, A. L. & Kiessling, L. L. Improved Chemical Synthesis of UDP-Galactofuranose. *Org. Lett.* **3**, 2517–2519 (2001).

41. Bae, B. *et al.* Molecular Basis for the Selectivity and Specificity of Ligand Recognition by the Family 16 Carbohydrate-binding Modules from Thermoanaerobacterium polysaccharolyticum ManA. *Journal of Biological Chemistry* **283**, 12415–12425 (2008).

42. Daffe, M., McNeil, M. & Brennan, P. J. Major structural features of the cell wall arabinogalactans of Mycobacterium, Rhodococcus, and Nocardia spp. *Carbohydrate Research* **249**, 383–398 (1993).

43. Brown, C. D., Rusek, M. S. & Kiessling, L. L. Fluorosugar Chain Termination Agents as Probes of the Sequence Specificity of a Carbohydrate Polymerase. *J. Am. Chem. Soc.* 4 (2012).

44. Yakovlieva, L. & Walvoort, M. T. C. Processivity in Bacterial Glycosyltransferases. *ACS Chem. Biol.* **15**, 3–16 (2020).

45. Levengood, M. R., Splain, R. A. & Kiessling, L. L. Monitoring Processivity and Length Control of a Carbohydrate Polymerase. *J. Am. Chem. Soc.* **133**, 12758–12766 (2011).

46. Nassau, P. M. *et al.* Galactofuranose biosynthesis in Escherichia coli K-12: identification and cloning of UDP-galactopyranose mutase. *J Bacteriol* **178**, 1047–1052 (1996).

47. Tanner, J. J., Boechi, L., Andrew McCammon, J. & Sobrado, P. Structure, mechanism, and dynamics of UDP-galactopyranose mutase. *Archives of Biochemistry and Biophysics* **544**, 128–141 (2014).

48. Dianišková, P. *et al.* Investigation of ABC transporter from mycobacterial arabinogalactan biosynthetic cluster. *gpb* **30**, 239–250 (2011).

49. Kelly, S. D., Allas, M. J., Goodridge, L. D., Lowary, T. L. & Whitfield, C. Structure, biosynthesis and regulation of the T1 antigen, a phase-variable surface polysaccharide conserved in many Salmonella serovars. *Nat Commun* **15**, 6504 (2024).

50. Janoš, P., Kozmon, S., Tvaroška, I. & Koča, J. How *Mycobacterium tuberculosis* Galactofuranosyl Transferase 2 (GlfT2) Generates Alternating β-(1-6) and β-(1-5) Linkages: A QM/MM Molecular Dynamics Study of the Chemical Steps. *Chem. Eur. J.* **24**, 7051–7059 (2018).

51. Konyariková, Z., Savková, K., Kozmon, S. & Mikušová, K. Biosynthesis of Galactan in *Mycobacterium tuberculosis* as a Viable TB Drug Target? *Antibiotics* **9**, 20 (2020).

52. Hayashi, J. M. *et al.* Spatially distinct and metabolically active membrane domain in mycobacteria. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5400–5405 (2016).

53. May, J. F., Levengood, M. R., Splain, R. A., Brown, C. D. & Kiessling, L. L. A Processive Carbohydrate Polymerase That Mediates Bifunctional Catalysis Using a Single Active Site. *Biochemistry* **51**, 1148–1159 (2012).

54. Janoš, P., Tvaroška, I., Dellago, C. & Koča, J. Catalytic Mechanism of Processive GlfT2: Transition Path Sampling Investigation of Substrate Translocation. *ACS Omega* **5**, 21374–21384 (2020).

55. Poulin, M. B. & Lowary, T. L. Chemical Insight into the Mechanism and Specificity of GlfT2, a Bifunctional Galactofuranosyltransferase from Mycobacteria. *J. Org. Chem.* **81**, 8123–8130 (2016).

56. Smego, R. A. & Gallis, H. A. The Clinical Spectrum of *Nocardia brasiliensis* Infection in the United States. *Clinical Infectious Diseases* **6**, 164–180 (1984).

57. Fukuda, H., Saotome, A., Usami, N., Urushibata, O. & Mukai, H. Lymphocutaneous type of nocardiosis caused by *Nocardia brasiliensis* : A case report and review of primary cutaneous nocardiosis caused by *N. brasiliensis* reported in Japan. *The Journal of Dermatology* **35**, 346–353 (2008).

58. Ramos-e-Silva, M., Lopes, R. S. & Trope, B. M. Cutaneous nocardiosis: A great imitator. *Clinics in Dermatology* **38**, 152–159 (2020).

59. Vera-Cabrera, L., Ortiz-Lopez, R., Elizondo-Gonzalez, R., Perez-Maya, A. A. & Ocampo-Candiani, J. Complete Genome Sequence of *Nocardia brasiliensis* HUJEG-1. *Journal of Bacteriology* **194**, 2761–2762 (2012).

60. Soltero-Higgin, M., Carlson, E. E., Gruber, T. D. & Kiessling, L. L. A unique catalytic mechanism for UDP-galactopyranose mutase. *Nat Struct Mol Biol* **11**, 539–543 (2004).

61. Cole, S. T. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544 (1998).

62. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* **77**, 521–555 (2008).

63. Poulin, M. B., Zhou, R. & Lowary, T. L. Synthetic UDP-galactofuranose analogs reveal critical enzyme–substrate interactions in GlfT2-catalyzed mycobacterial galactan assembly. *Org. Biomol. Chem.* **10**, 4074 (2012).

64. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology* **372**, 774–797 (2007).

65. Gautier, R., Douguet, D., Antonny, B. & Drin, G. HELIQUEST: a web server to screen sequences with specific α-helical properties. *Bioinformatics* **24**, 2101–2102 (2008).

66. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Research* **46**, W363–W367 (2018).

67. Breyer, W. A. & Matthews, B. W. A structural basis for processivity. *Protein Sci.* **10**, 1699–1711 (2001).

68. Hudson, K. L. *et al.* Carbohydrate–Aromatic Interactions in Proteins. *J. Am. Chem. Soc.* **137**, 15152–15160 (2015).

69. Kiessling, L. L. & Diehl, R. C. CH−π Interactions in Glycan Recognition. *ACS Chem. Biol.* **16**, 1884–1893 (2021).

70. Morgan, J. L. W., McNamara, J. T. & Zimmer, J. Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. *Nat Struct Mol Biol* **21**, 489–496 (2014).

71. Splain, R. A. & Kiessling, L. L. Synthesis of galactofuranose-based acceptor substrates for the study of the carbohydrate polymerase GlfT2. *Bioorganic & Medicinal Chemistry* **18**, 3753–3759 (2010).

72. Li, F. K. K., Worrall, L. J., Gale, R. T., Brown, E. D. & Strynadka, N. C. J. Cryo-EM analysis of *S. aureus* TarL, a polymerase in wall teichoic acid biogenesis central to virulence and antibiotic resistance. *Sci. Adv.* **10**, eadj3864 (2024).

73. Van Kempen, M. *et al.* Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* (2023) doi:10.1038/s41587-023-01773-0.

74. García-Nafría, J., Watson, J. F. & Greger, I. H. IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. *Sci Rep* **6**, 27459 (2016).

75. Tartof, K. Improved media for growing plasmid and cosmid clones. *Bethesda Res Lab Focus* **9**, 12 (1987).

76. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **41**, 207–234 (2005).

77. Heiss, C., Stacey Klutts, J., Wang, Z., Doering, T. L. & Azadi, P. The structure of *Cryptococcus neoformans* galactoxylomannan contains β-d-glucuronic acid. *Carbohydrate Research* **344**, 915–920 (2009).

78. McCoy, A. J. *et al. Phaser* crystallographic software. *J Appl Crystallogr* **40**, 658–674 (2007).

79. Adams, P. D. *et al. PHENIX* : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221 (2010).

80. Chen, V. B. *et al. MolProbity* : all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12–21 (2010).

81. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).

82. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. *electronic Ligand Builder and Optimization Workbench* (*eLBOW*): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr* **65**, 1074–1080 (2009).

83. Long, F. *et al. AceDRG* : a stereochemical description generator for ligands. *Acta Crystallogr D Struct Biol* **73**, 112–122 (2017).

84. Gordon, J. C. *et al.* H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Research* **33**, W368–W371 (2005).

85. Kirschner, K. N. *et al.* GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J Comput Chem* **29**, 622–655 (2008).

86. Pearlman, D. A. *et al.* AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **91**, 1–41 (1995).

87. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).

88. Price, D. J. & Brooks, C. L. A modified TIP3P water potential for simulation with Ewald summation. *The Journal of Chemical Physics* **121**, 10096–10103 (2004).

89. Dupradeau, F.-Y. *et al.* The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **12**, 7821 (2010).

90. Vanquelef, E. *et al.* R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res* **39**, W511-7 (2011).

91. Wang, F., Becker, J. P., Cieplak, P. & Dupradeau, F. Y. Ed. Online (2014).

92. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089–10092 (1993).

93. Götz, A. W. *et al.* Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **8**, 1542–1555 (2012).

94. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327–341 (1977).

95. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).

96. Massova, I. & Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspectives in Drug Discovery and Design* **18**, 113–135 (2000).

97. Miller, B. R. *et al. MMPBSA.py* : An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **8**, 3314–3321 (2012).

98. Onufriev, A., Bashford, D. & Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **55**, 383–394 (2004).

99. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & De Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).

100. Dodge, G. J., Bernstein, H. M. & Imperiali, B. A generalizable protocol for expression and purification of membrane-bound bacterial phosphoglycosyl transferases in liponanoparticles. *Protein Expression and Purification* **207**, 106273 (2023).

101. Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* **16**, 1153–1160 (2019).

102. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290–296 (2017).

103. Morgan, J. L. W., Strumillo, J. & Zimmer, J. Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature* **493**, 181–186 (2013).

104. Emsley, P. & Cowtan, K. *Coot* : model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132 (2004).

105. Croll, T. I. *ISOLDE* : a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519–530 (2018).

106. Pettersen, E. F. *et al.* UCSF CHIMERAX : Structure visualization for researchers, educators, and developers. *Protein Science* **30**, 70–82 (2021).

107. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J Comput Chem* **29**, 1859–1865 (2008).

108. Wu, E. L. *et al.* CHARMM-GUI *Membrane Builder* toward realistic biological membrane simulations. *J. Comput. Chem.* **35**, 1997–2004 (2014).

109. Qi, Y. *et al.* CHARMM-GUI Martini Maker for Coarse-Grained Simulations with the Martini Force Field. *J. Chem. Theory Comput.* **11**, 4486–4494 (2015).

110. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

111. Weijers, C. A. G. M., Franssen, M. C. R. & Visser, G. M. Glycosyltransferase-catalyzed synthesis of bioactive oligosaccharides. *Biotechnology Advances* **26**, 436–456 (2008).

112. Liang, D.-M. *et al.* Glycosyltransferases: mechanisms and applications in natural product development. *Chem. Soc. Rev.* **44**, 8350–8374 (2015).

113. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

114. Varadi, M. *et al.* AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research* **52**, D368–D375 (2024).

115. Seidel, M. *et al.* Identification of a Novel Arabinofuranosyltransferase AftB Involved in a Terminal Step of Cell Wall Arabinan Biosynthesis in Corynebacterianeae, such as Corynebacterium glutamicum and Mycobacterium tuberculosis. *Journal of Biological Chemistry* **282**, 14729–14740 (2007).

116. Gao, S. *et al.* Structural analysis of phosphoribosyltransferase-mediated cell wall precursor synthesis in Mycobacterium tuberculosis. *Nat Microbiol* **9**, 976–987 (2024).

117. Cai, L. *et al.* Prokaryotic Expression, Identification and Bioinformatics Analysis of the Mycobacterium tuberculosis Rv3807c Gene Encoding the Putative Enzyme Committed to Decaprenylphosphoryl-d-arabinose Synthesis. *Indian J Microbiol* **54**, 46–51 (2014).

118. Korres, H., Mavris, M., Morona, R., Manning, P. A. & Verma, N. K. Topological analysis of GtrA and GtrB proteins encoded by the serotype-converting cassette of Shigella flexneri. *Biochemical and Biophysical Research Communications* **328**, 1252–1260 (2005).

119. Kolly, G. S. *et al.* GtrA Protein Rv3789 Is Required for Arabinosylation of Arabinogalactan in Mycobacterium tuberculosis. *J Bacteriol* **197**, 3686–3697 (2015).

120. Scherman, H. *et al.* Identification of a Polyprenylphosphomannosyl Synthase Involved in the Synthesis of Mycobacterial Mannosides. *J Bacteriol* **191**, 6769–6772 (2009).

121. Škovierová, H. *et al.* Biosynthetic Origin of the Galactosamine Substituent of Arabinogalactan in Mycobacterium tuberculosis. *Journal of Biological Chemistry* **285**, 41348–41355 (2010).

122. Sichtig, H. *et al.* FDA-ARGOS is a database with public quality-controlled reference genomes for diagnostic use and regulatory science. *Nat Commun* **10**, 3313 (2019).

123. Kaur, D. *et al.* Lipoarabinomannan of *Mycobacterium* : Mannose capping by a multifunctional terminal mannosyltransferase. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17973–17977 (2008).

124. Letek, M. *et al.* The Genome of a Pathogenic Rhodococcus: Cooptive Virulence Underpinned by Key Gene Acquisitions. *PLoS Genet* **6**, e1001145 (2010).

125. Rismondo, J., Percy, M. G. & Gründling, A. Discovery of genes required for lipoteichoic acid glycosylation predicts two distinct mechanisms for wall teichoic acid glycosylation. *Journal of Biological Chemistry* **293**, 3293–3306 (2018).

126. Hirsch, A. M. *et al.* Genome of *Dietzia cinnamea* 55, a desert-isolated microbe with plant growth-promoting properties for grain crops. *Microbiol Resour Announc* **13**, e00257-24 (2024).

127. Kempkes, R., Stofko, E., Lam, K. & Snell, E. H. Glycerol concentrations required for the successful vitrification of cocktail conditions in a high-throughput crystallization screen. *Acta Crystallogr D Biol Crystallogr* **64**, 287–301 (2008).

128. Sayers, E. W. *et al.* GenBank. *Nucleic Acids Research* **47**, D94–D99 (2019).

129. Schoch, C. L. *et al.* NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database* **2020**, baaa062 (2020).

130. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research* **52**, W78–W82 (2024).

131. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27–30 (2000).