University of Alberta

Chemical Studies Exploring the Specificity of Bacterial Furanoside Biosynthesis

by

Myles Bradley Poulin

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

Fall 2012
Edmonton, Alberta

Dedicated to All My Friends and Family

Abstract

Mammalian glycoconjugates are composed exclusively of sugar in the thermodynamically favorable pyranose ring form; on the other hand, sugars in the five-membered furanose ring form are widespread in many bacteria, fungi, and protozoan pathogens. This makes the enzymes involved in furanoside biosynthesis a potential selective drug target for anti-microbial therapeutics. However, many questions regarding the binding interactions that occur in the active sites of these enzymes still remain unanswered. In particular, protein structural motifs involved in substrate discrimination remain poorly understood. Galactofuranose, the most common hexofuranose sugar, is biosynthesized by an enzyme known as UDP-galactopyranose mutase (UGM). In contrast, the role of similar enzymes involved in the biosynthesis of various other furanose sugars has not been established.

Of particular interest to this thesis are the furanose sugars found in the capsular polysaccharides of *Campylobacter jejuni*; specifically the 2-acetamido-2-deoxy-D-galactofuranose residue produced by *C. jejuni* serotype HS:2 and the D-fucofuranose, 6-deoxy-L-altrofuranose, L-arabinofuranose, and 6-deoxy-D-*altro*-heptofuranose produced by *C. jejuni* serotype HS:41 whose biosynthesis remain unexplored. Herein, we examine the activity and specificity of the pyranose–furanose mutase enzymes responsible for the biosynthesis of these sugars. Using synthetic substrate analogs and molecular biology techniques, we have also evaluated specific binding interactions responsible for the substrate specificity of these enzymes.

In addition to the pyranose–furanose mutase, furanosyltransferase enzymes are also involved in furanoside biosynthesis. Specifically, the mycobacterial cell wall galactan, composed of alternating β-(1→5) and β-(1→6) galactofuranosyl residues, is assembled by the action of two bifunctional galactofuranosyltransferases, GlfT1 and GlfT2. The second, GlfT2, adds the third and subsequent Gal*f* residues using a single active site to carry out both the formation of β-D-Gal*f*-(1→5)-β-D-Gal*f* and β-D-Gal*f*-(1→6)-β-D-Gal*f* linkages. Previous work has largely focused on the specificity of the acceptor species with little known with regards to the binding and specificity of the UDP-D-Gal*f* donor. Herein, we have used a range of synthetic UDP-D-Gal*f* analogs to probe the specificity of GlfT2, and site-directed mutagenesis to explore the mechanism of alternating β-(1→5) and β-(1→6)-GlfT activity. Together, these observations provide insight into specific protein–carbohydrate interactions in GlfT2 and may facilitate the design of future inhibitors.

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Todd Lowary, for his help, guidance, and support throughout the last five years of my studies. Todd has provided an exceptional environment in which to work and learn, and I am thankful for all of the help and opportunities he has given me while I have been here at the University of Alberta.

Next, I would like to show my appreciation to Dr. Harald Nothaft, Bernadette Beadle, Dr. Isabelle Hug, Dr. Sean Dalrymple, Carla Protsko, Professor Christine Szymanski, Professor Mario Feldman, and Professor David Sanders for their collaborations throughout the course of my Ph.D. studies.

I have been fortunate to have the opportunity to work with a great group of people during the course of this dissertation and I would like to thank all the members of the Lowary group, past and present, for their help and support during this time. Specifically, I would like to thank Jean Pearcey and Ruixiang Blake Zheng for their training at the outset of my studies, Ruokun Zhou for his initial contributions to the synthesis of the substrate analogs discussed in this thesis, Dr. Maju Joe and Dr. Peter Meloncelli for their helpful discussions and suggestions, and Li Xia and Lu Zou for their help and support over the past years.

I would also like to thank the Alberta Glycomics Centre, the University of Alberta and the Natural Science, and Engineering Research Council of Canada for supporting my thesis research, and to Alberta Innovates Technology

futures and NSERC for providing me with financial support during the course of this dissertation.

Last but not least, I would like to thank my parents and my brothers for their constant support and encouragement, not only during the course of my studies, but throughout my life. And finally to my fiancé Lily for her constant help and support. I would not be here today if not for all of them.

Table of Contents

List of Tables

List of Figures

List of Schemes

List of Abbreviations

1P	1-phosphate
5P-DRP	5′-phospho-decaprenylphosphoryl-β-D-ribofuranose
5P-RPP	5-phospho-α-D-ribose 1-pyrophophate
6d-L-Alt*f*	6-deoxy-L-altrofuranose
6d-L-Alt*p*	6-deoxy-L-altropyranose
6d-D-*altro*-Hep*f*	6-deoxy-D-*altro*-heptofuranose
6d-D-*altro*-Hep*p*	6-deoxy-D-*altro*-heptopyranose
*A*600	Absorbance at 600 nm
Ac	Acetyl
ADP	Adenine 5′-diphosphate
AfT	Arabinofuranosyltransferase
Alt*f*	Altrofuranose
AP	Alkaline phosphatase
Ara*f*	Arabinofuranose
Ara*p*	Arabinopyranose
BCA	bicinchoninic acid assay
BLAST	Basic local alignment search tool
Bn	Benzyl
Bz	Benzoyl
cat	Chloramphenicol
CDP	Citydine 5′-diphosphate
CE	Capilary electrophoresis
cj	*Campylobacter jejuni*

CMP	Citydine 5′-monophosphate
CPS	Capsular polysaccharide
Cps2L	α-D-glucopyranosyl-1-phosphate thymidylyltransferase
ΔA_{500}	Change in absorbance at 500 nm
DMAP	*N,N*-Dimethyl-4-amminopyridine
DMF	*N,N*-Dimethyl formamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DPA	decaprenylphosphoryl-β-D-arabinofuranose
DPR	decaprenylphosphoryl-β-D-ribofuranose
DPX	decaprenylphosphory-β-D-2′-keto-*erythro*-pentofuranose
dr	*Deinococcus radiodurans*
TDP	Deoxy thymidine 5′-diphosphate
dTTP	Deoxy thymidine 5′-triphosphate
ec	*Escherichia coli*
Emb	Ethambutol
ESI	Electrospray ionization
Et	Ethyl
FAD	Flavin adenine dinucleotide (oxidized)
$FADH^-$	Flavin adenine dinucleotide (reduced)
FPA	Farnesylphosphoryl-β-D-arabinofuranose
FPR	Farnesylphosphoryl-β-D-ribofuranose
Fuc*f*	6-Deoxy-galactofuranose, fucofuranose
Fuc*p*	6-Deoxy-galactopyranose, fucopyranose

GaHM	GDP-*altro*-heptopyranose mutase
GalE	galactopyranose epimerase
Gal*f*	Galactofuranose
Gal*f*NAc	2-Acetamido-2-deoxy-galactofuranose
Gal*p*	Galactopyranose
Gal*p*NAc	2-Acetamido-2-deoxy-galactopyranose
GalPUT	Galactose-1-phosphate uridylyltransferase
GalU	glucose-1-phosphate uridylyltransferase
GBS	Guillain–Barré syndrome
GDP	Guanosine 5′-diphosphate
GlcA	Glucuronic acid
Glc*p*	Glucopyranose
Glc*p*NAc	2-Acetamido-2-deoxy-glucopyranose
GlfT1	Mycobacterial galactofuranosyltransferase 1
GlfT2	Mycobacterial galactofuranosyltransferase 2
GMP	Guanosine 5′-monophosphate
GTA	Glycosyltransferase A
GTB	Glycosyltransferase B
His6	Hexa histidine tag
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
IC_{50}	50% inhibitory concentration
IPP	Inorganic pyrophosphorylase, pyrophosphatase
IPTG	Isopropyl 1-thio-β-D-galactopyranoside
kan	Kanamycin

k_{cat}	Apparent unimolecular rate constant
K_M	Michaelis constant
kp	*Klebsiella pneumoniae*
LAM	Lipoarabinomannan
LOS	Lipooligosaccharide
LPS	Lipopolysacharide
mAG	Mycolyl-arabinogalactan
MALDI	Matrix assisted laser desorption ionization
Me	Methyl
MS	Mass spectrometry
mt	*Mycobacterium tuberculosis*
NAD+	Nicotinamide adenine dinucleotide (oxidized)
NADH	Nicotinamide adenine dinucleotide (reduced)
Ni-NTA	Nickel nitrilotriacetic acid
NMR	Nuclear magnetic resonance
OD	Optical density
p-TSA	para-Toluenesulfonic acid
PCR	Polymerase chain reaction
PDB	Protein Data Bank
P*i*	Inorganic phosphate
PP*i*	Inorganic pyrophosphate
PIX	Positional isotope exchange
Py	Pyridine
R*f*	Retention factor
RGP	Reversibly glycosylated protein

Rha*p*	Rhamnopyranose
Rib*f*	Ribofuranose
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SET	Single electron transfer
STD-NMR	Saturation transfer difference nuclear magnetic resonance
STDD	Saturation transfer double difference
TB	Tuberculosis
TBAF	Tetrabutylammonium fluoride
TBDPS	Tertbutyldiphenylsilyl
THF	Tetrahydrofurane
TLC	Thin layer chrpmatography
Tr	Trityl
UAM	UDP-arabinopyranose mutase
UDP	Uridine 5′-diphosphate
UGM	UDP-galactopyranose mutase
UMP	Uridine 5′-monophosphate
UNGM	UDP-N-acetyl-galactopyranose mutase
UTP	Uridine 5′-triphosphate
UV	Ultravilolet
vis	Visible
wt	Wild-type
Xyl	Xylose

Chapter 1

Activated Furanoside Donors. Structural and Mechanistic Insight Into Furanoside Biosynthesis

A portion of this Chapter has been previously published:

Poulin, M. B.; Lowary, T. L., *Methods in Enzymology*. **2010**, *478*, 389-411.

1.1 Furanose Sugars in Nature

Carbohydrate residues in nature typically exist in the thermodynamically-favored six membered pyranose ring form though monosaccharides in the disfavored five membered furanose ring form are also widespread. Furanose sugars have been found in the glycoconjugates of bacteria, fungi, plants, and protozoa including many human and animal pathogens;[1, 2] however, they are not found in mammalian glycoconjugates. For many pathogens, including mycobacterial species such as *Mycobacterium tuberculosis*, fungal pathogens such as *Aspergillus* species, and protozoan pathogens including *Leishmania* species, the furanoside glycoconjugates are essential for the proper growth or virulence of the organism.[1-3] Consequently, there has been a significant interest in studying the enzymes involved in the biosynthesis of furanose sugars as potential targets for anti-microbial chemotherapeutics.[4]

Natural glycans, including furanosides, are typically assembled by the action of glycosyltransferase enzymes, which use nucleotide-diphosphate or lipid-(di)phosphate activated sugar donors as the source of the carbohydrate residues added to the glycan chain. For example, the UDP-D-galactofuranose (UDP-D-Gal*f*) serves as the source of D-galactofuranose (D-Gal*f*) in the lipopolysaccharide (LPS) galactan I domains of *Klebsiella pneumoniae*,[5, 6] and the cell wall galactan of *Mycobacterium tuberculosis* (Figure 1-1), as will be discussed in more detail later in this chapter.

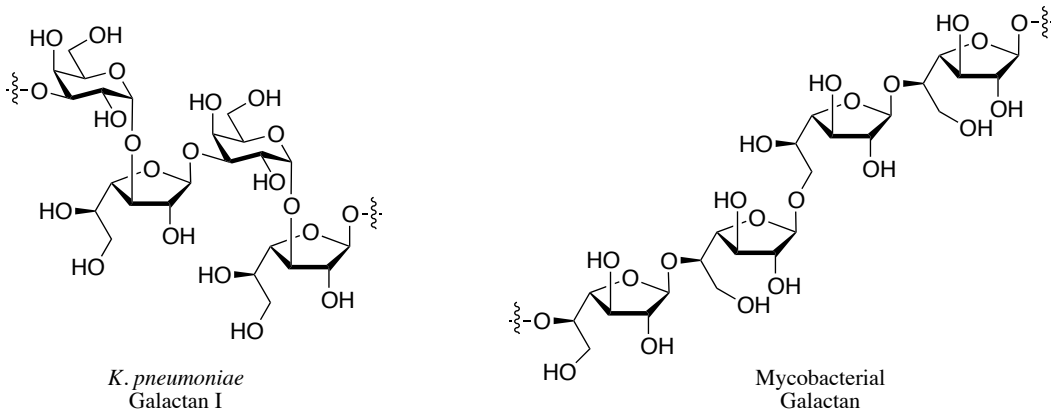


Figure 1-1. Structures of D-Gal*f* containing glycoconjugates. A tetrasaccharide portion of the *K. pneumoniae* galactan I structure (*left*) and a tetrasaccharide potion of the cell wall galactan of mycobacterial species (*right*). In both structures, UDP-D-Gal*f* is the precursor of D-Gal*f* residues.

Glycans containing D-Gal*f*, D-arabinofuranose (D-Ara*f*), and L-arabinofuranose (L-Ara*f*) are the most common furanoside structures indentified in microbial and plant glycans, and, as such, there has been a substantial amount of work dedicated over the past two decades to elucidate the biosynthesis of these furanose sugars. In each case, an activated furanose donor serves as the source of the D-Gal*f*, D-Ara*f*, or L-Ara*f* residue. The aim of this introduction is not to provide a comprehensive overview of the structure and function of furanoside containing glycoconjugates, many of which have been described in recent reviews.[1, 2, 9-12] Instead, this chapter will introduce the key enzymes involved in the biosynthesis of these furanose sugars and, in particular, recent structural and mechanistic insights into the methods by which the enzymes prepare these thermodynamically disfavored sugar donors.

1.2 Enzymes involved in D-Arabinofuranose Biosynthesis

Of the three monosaccharides introduced above, D-Ara*f* is a relatively rare sugar found predominantly in the cell wall glycoconjugates of mycobacterial species and related *Actinomycetes*.[7-10] In contrast to D-Ara*p*, which has been found in the glycans of several eukaryotic species including trypanosomatids[11] and *Leishmania* species,[12, 13] D-Ara*f* has only been found in bacterial species.[14] In *Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB), and other mycobacterial species D-Ara*f* is found in the lipoarabinomannan (LAM) glycolipid as well as the mycolylarabinogalactan (mAG) complex of the bacterial cell wall (Figure 1-2).[7] Biosynthesis of the mAG arabinan is essential for mycobacterial viability and has been the target of anti-mycobacterial chemotherapeutics. Of particular relevance is ethambutol, a current front line antibiotic used in the treatment of mycobacterial infections, which targets arabinofuranosyltransferases (AfTs) involved in the assembly of mycobacterial arabinan domains.[15, 16] Thus, there has been an increased interest in studying the enzymes involved in L-Ara*f* metabolism as putative targets for thereputics against TB and other mycobacterial infections.

In addition to the *Actinomycetes*, D-Ara*f* has also been found in the pilin O-glycan of *Pseudomona aeruginosa*,[17] a Gram-negative opportunistic pathogen, and in the Nod factor glycans of *Azorhizobium caulinodans*,[18] a symbiotic organism in leguminous plants. It was also shown that the biosynthesis of D-Ara*f* is conserved between *P. aeruginosa*,[19] mycobacterial species and *A. caulinodans*,[18] as will be discussed below. Most recently, D-Ara*f* has been found

in an arabinomannan isolated from the cell wall of the chlorococcal algae *Chlorella vulgaris*. The structure of these glycan was shown to be related to LAM in *M. tuberculosis*;[20] however, the biosynthesis of this polysaccharide is still not known.

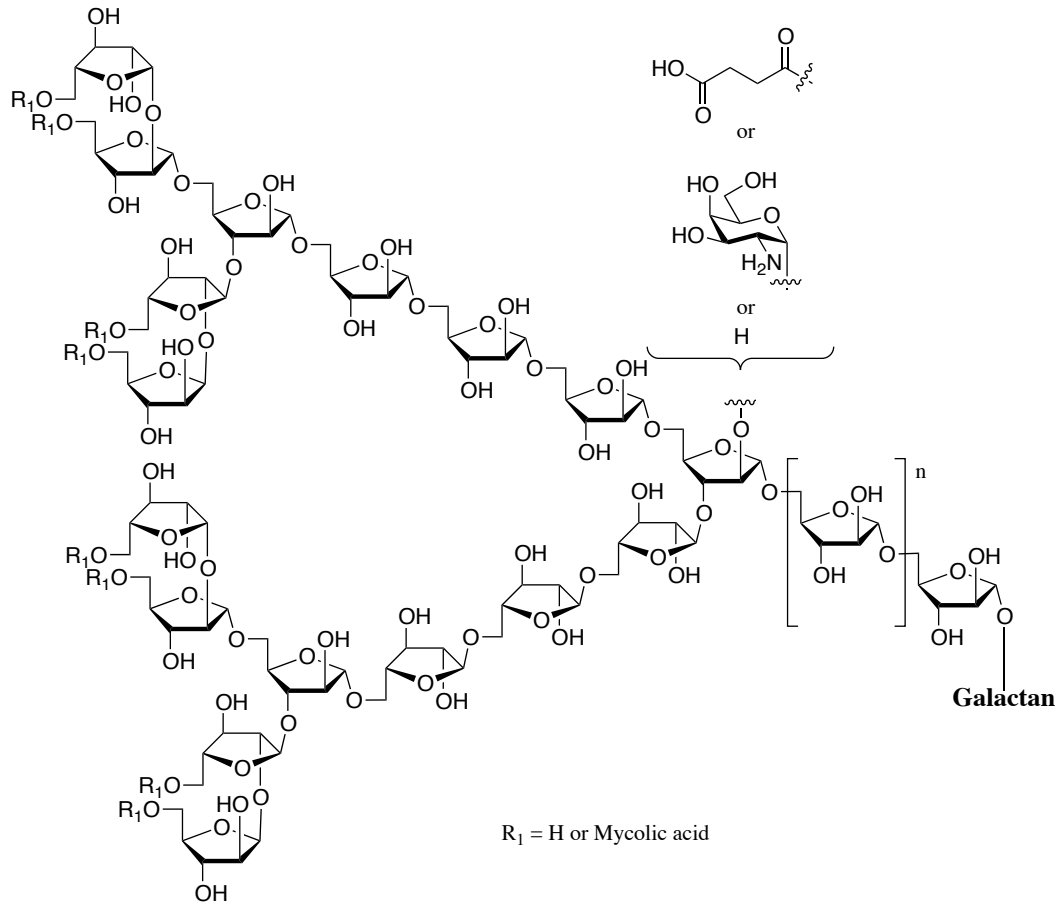


Figure 1-2. Structure of mAG arabinan domains. Three arabinan domains are attached at O-5 to the 8th, 10th, and 12th D-Gal*f* residue of the galactan. Two thirds of the arabinan domains are further esterified with long chain mycolic acid lipids at O-5 of the non-reducing terminal L-Ara*f* residues. A fraction of the internal 1,3,5-substituted L-Ara*f* residues are also further substituted with succinyl or α-D-galactosamine residues.[8]

1.2.1 Biosynthesis of D-Ara*f*

Guanosine 5′-diphosphate-α-D-arabinopyranose (GDP-D-Ara*p*) serves as the precursor for D-Ara*p* residues found in the lipophosphoglycan of protozoan species,[21, 22] where it is derived from glucose by C1 truncation (Figure 1-3). However, no GDP-D-Ara*f* species has been identified in the biosynthesis of bacterial D-Ara*f*-containing glycans. The biosynthesis of D-Ara*f* in the LAM and mAG of mycobacterial species remained unclear until 1994 when a family of polyprenylphosphate monosaccharides were isolated from *Mycobacterium smegmatis* treated with ethambutol.[15] These include two pentose polyprenylphosphates, decaprenylphosphoryl-β-D-ribofuranose (DPR) and decaprenylphosphoryl-β-D-arabinofuranose (DPA), the later of which serves as the immediate precursor to L-Ara*f* in mycobacterial arabinan. A water-soluble uridine 5′-diphosphate (UDP) derivative of L-Ara*f* has also been identified in the extracts of *M. smegmatis*;[23] however, there have been no subsequent reports to demonstrate a possible role for UDP-D-Ara*f* in mycobacterial glycan biosynthesis.

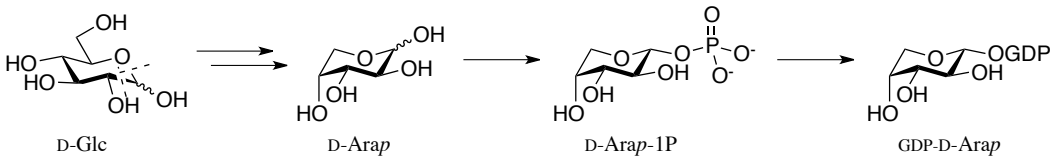


Figure 1-3. GDP-D-Ara*p* biosynthesis in protozoan species.

Isotope labeling experiments in which M. *smegmatis* cells were incubated with specifically ^{14}C- or ^{13}C-labeled glucose showed that neither the

pentose phosphate pathway (Figure 1-4A), which would result in the loss of C1 of glucose, or the uronic acid pathway (Figure 1-4B), which would result in the loss of C6 of glucose, are involved in D-Ara*f* biosynthesis.[24] Instead, the majority of the C1 and C6 isotope labels were retained in D-Ara*f*. The same isotope labeling patterns were also observed for D-ribose, isolated in the same experiments, suggesting it is a precursor to the D-Ara*f* via the non-oxidative pentose phosphate pathway (Figure 1-4C).

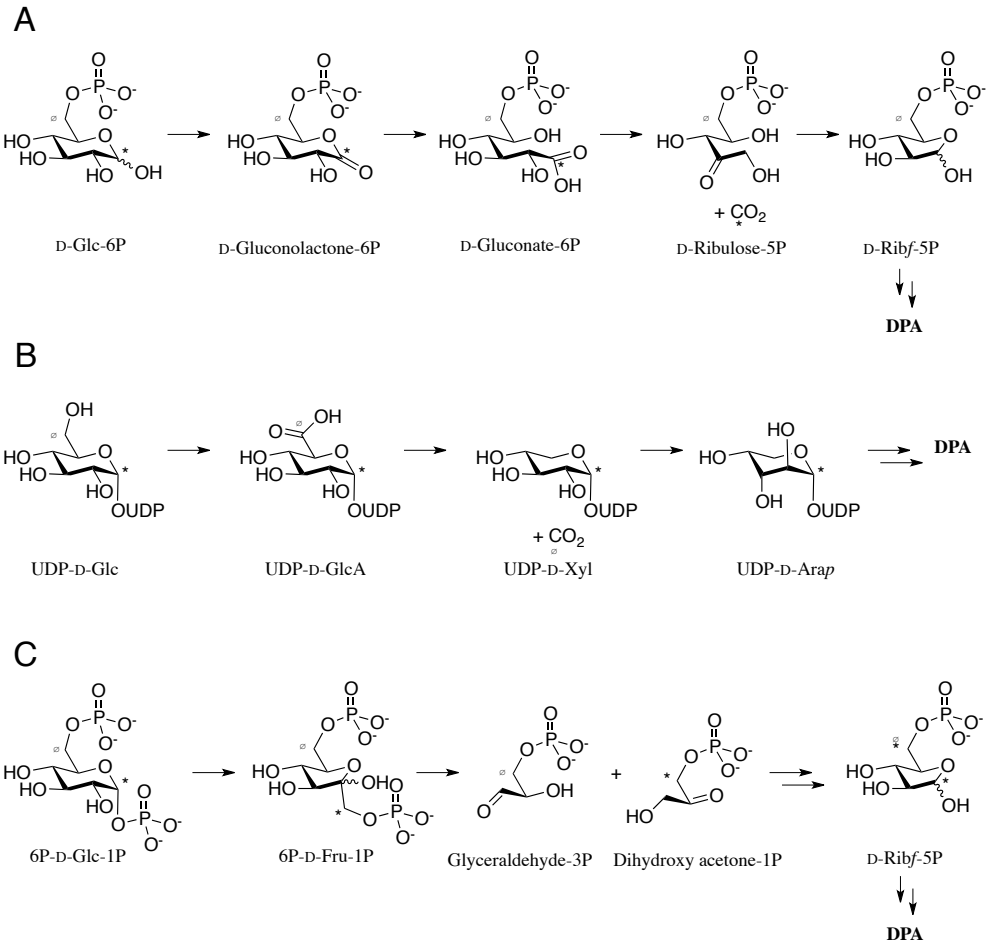


Figure 1-4. Oxidative pentose phosphate (A), uronic acid (B), and non-oxidative pentose phosphate (C) pathways for the biosynthesis of DPA. The position of the glucose ^{13}C1 (*) and ^{13}C6 (^{Ø}) labels are tracked through each pathway. Only the non-oxidative pentose phosphate pathway (C) accounts for the retention of both the ^{13}C1 and ^{13}C6 labels into D-Ara*f*.

Other isotope labeling experiments with ^{14}C-labeled 5-phospho ribofuranose-1-pyrophosphate (5P-RPP) implicate this species as an intermediate in the biosynthesis of both DPA and DPR (Figure 1-5).[25] The subsequent identification of a DPR 5′-phosphate synthetase (DprS)[26] and a two protein DPR epimerase complex (DprE1/DprE2)[27] in 2005, based on similarity to the Nod factor biosynthetic gene locus of *A. caulinodans*,[18] lead to the currently accepted pathway for D-Ara*f* biosynthesis (Figure 1-5). These same genes were also found in the D-Ara*f* biosynthetic locus of *P. aeruginosa*[19] and are required for glycosylation of the pilin.

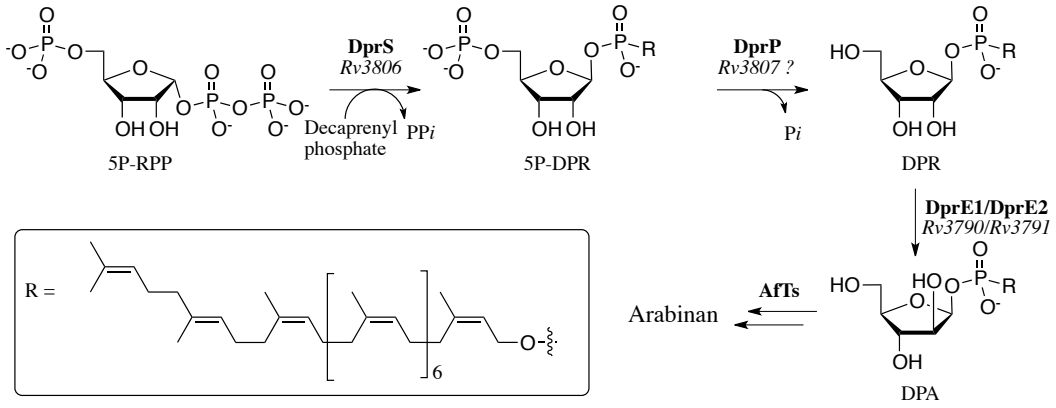


Figure 1-5. Biosynthesis of the mycobacterial arabinan uses a decaprenylphosphoryl-β-D-arabinofuranose (DPA) donor. DPA is derived from 5-phospho-α-D-ribofuranose 1-pyrophophate (5P-RPP) by the action of three enzyme, DprS, DprP, and DprE1/DprE2.

No sugar nucleotide intermediate is used in this pathway; instead, 5P-RPP serves directly as the substrate for DprS to form the lipid linked 5′-phospho decaprenylphosphoryl-β-D-ribofuranose (5P-RPP) precursor to DPA.[25] DprS in *M. tuberculosis* is encoded by the *Rv3806* gene and is predicted to contain nine

trans-membrane domains. Consistent with this hypothesis, recombinant DprS is isolated from the membrane fractions *E. coli*, and catalyzes the reversible displacement of pyrophosphate, from 5P-RPP, by decaprenyl phosphate.[26] A phosphatase, encoded by *Rv3807* in *M. tuberculosis*, is predicted to cleave the 5′-phosphate of 5P-DPR to give DPR. However, *Rv3807* is apparently not essential in mycobacteria,[28] whereas all other D-Ara*f* biosynthetic genes are. Thus, it is still not clear what role *Rv3807* plays in D-Ara*f* biosynthesis, or whether other phosphatases are involved. The final step in DPA biosynthesis involves an epimerization of the 2′-OH of DPR. As this is the key step in D-Ara*f* biosynthesis, it will be described in more detail below.

1.2.2 DPR epimerase activity of DprE1/DprE2

The DPR 2′-epimerase responsible for the final step in DPA biosynthesis is a heterodimeric enzyme composed of two polypeptide chains, DprE1 and DprE2, which are encoded by the *Rv3790* and *Rv3791* genes, respectively.[14, 27] DprE1 shares sequence similarity to oxidoreductase enzymes and binds a flavin adenine dinucleotide (FAD) cofactor in the enzyme active site. The second protein component, DprE2, belongs to a family of short-chain dehydrogenase enzymes containing a nicotinamide adenine dinucleotide (NAD) binding Rossman fold. The epimerization reaction proceeds via a decaprenyl-phospho-D-2′-keto-*erythro*-pentofuranose (DPX) intermediate (Figure 1-6A), which has been observed during the *in vitro* interconversion of DPR and DPA but likely remains bound to the DprE1/DprE2 complex during the epimerization.[27] It was

originally reported that both DprE1 and DprE2 components and NAD(P)H were required for activity.[14] However, more recent reports show that, in the absence of DprE2 and NAD(P)H, DprE1 can catalyze the formation of DPX.[29] Reduction of DPX is then dependent on DprE2 and the presence of NAD(P)H. In addition to DPR, the more synthetically tractable farnesylphosphoryl-β-D-ribofuranose (FPR), the synthesis of which will be discussed in section 1.5.3, also serves as a substrate for DprE1/DprE2 and is turned over via an analogous mechanism.[29]

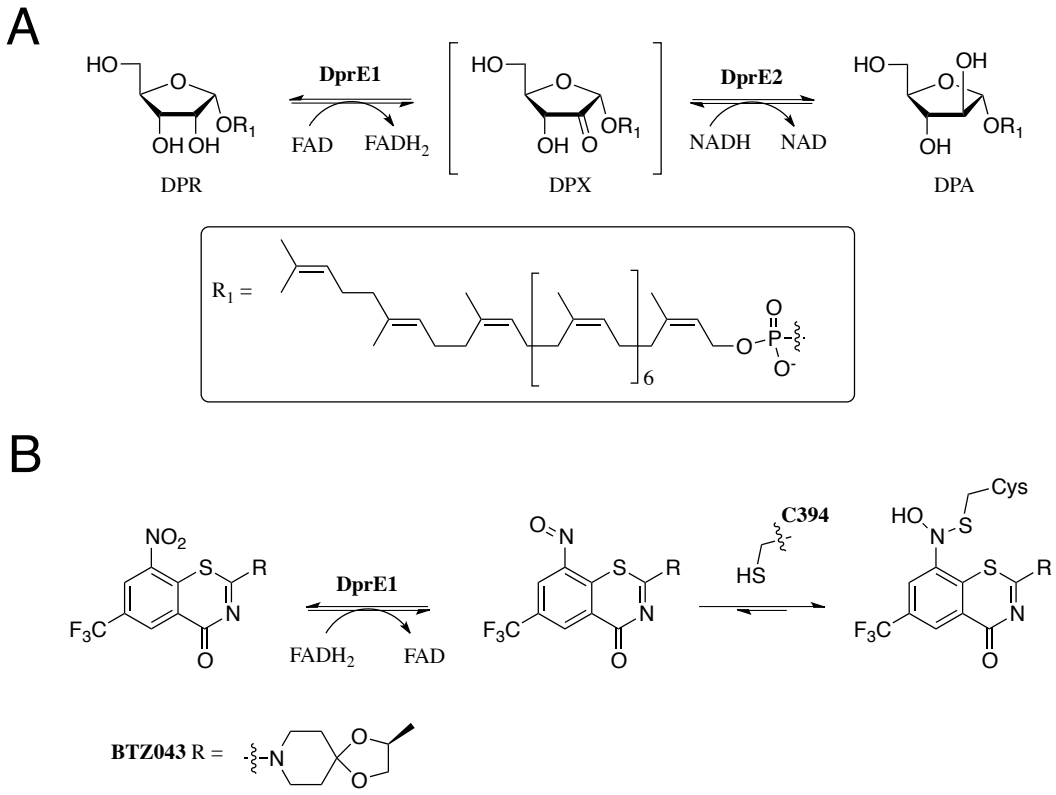


Figure 1-6. Isomerization of DPR to DPA by DprE1/DprE2 proceeds through a 2-keto (DPX) intermediate (A). DprE1 is the target of benzathiazinones such as BTZ043 (B), which function as pro-drugs that act as suicide inhibitors of arabinan biosythesis.

Although the exact molecular interactions required for epimerase activity are still not known, DprE1 and DprE2 have received significant attention in recent years after they were demonstrated to be the target of the recently discovered benzothiazinone anti-mycobacterial agents.[30, 31] Specifically, these benzothiazinone compounds, which contain a nitro functionality (i.e., BTZ043, Figure 1-6 B), are pro-drugs that covalently modify DprE1.[32] The mechanism of inhibition involves reduction of the nitro-functionality to a nitroso derivative by the reduced $FADH_2$ cofactor of DprE1 in turn producing a nascent electrophile that reacts with an active site cysteine forming a covalent complex (Figure 1-6 B).[29] Besra and coworkers recently reported structural evidence for such a covalent complex seen in a crystal structure of DprE1 with a benzothiazinone derivative.[33] As predicted, a covalent bond was observed between the derivative and the conserved active site cysteine of DprE1. Indeed, DprE1 appears to be a promising target for new anti-mycobacterial therapeutics,[31] although the structure of the DprE1/DprE2 complex, and many of the mechanistic details of the DPR to DPA conversion remain to be determined.

1.2.3 AfTs in mycobacterial arabinan assembly

The number of different glycosidic linkages found in mycobacterial LAM and mAG suggest that AfT enzymes with a range of specificities are required for their assembly. To date, seven AfT enzymes involved in the mycobacterial arabinan biosynthesis have been identified.[34, 35] However, the exact role of many of these proteins in mAG and LAM assembly remain to be

determined, and it is likely that other AfTs are required. All these AfTs appear to be integral membrane proteins, containing a number of transmembrane domains, and use DPA as the sole D-Ara*f* donor. These factors contribute to the difficulty in studying arabinan biosynthesis.[35, 36]

The AfTs identified to date include AftA, discovered in 2006,[37] which is the priming enzyme that adds the first D-Ara*f* residue to the galactan in mAG biosynthesis (Figure 1-7 A); however, the enzyme catalyzing the same transformation in LAM biosynthesis remains to be discovered. The AftB enzyme functions as a β-(1→2)-AfT adding the capping β-D-Ara*f* moiety to both mAG and LAM arabinan.[38] On the other hand, both AftB and AftC have been shown to exhibit α-(1→3)-AfT activity[45, 46] and likely have a role in introducing the (1→3)-D-Ara*f* branching in the mAG and LAM structures. Recombinant AftC, reconstituted in proteoliposomes, was recently evaluated *in vitro* with a series of arabinan fragment analogs.[39] Linear α-(1→5) linked arabinan analogs, containing a minimum of three D-Ara*f* residues, served as acceptors for AftC whereas shorter analogs or those containing a α-(1→3)-D-Ara*f* branches did not. Thus, the *in vitro* results support the role of this enzyme in introducing α-(1→3)-D-Ara*f* branches.

The Emb proteins, named such as they were found to be the targets of the anti-mycobacterial drug ethambutol, also function in mAG (EmbA, EmbB) and LAM (EmbC) biosynthesis (Figure 1-7). Although the exact function of EmbA and EmbB are unknown, they appear to play a role in the formation of the terminal hexasaccharide motif of the mAG arabinan.[40] EmbC plays no role in

mAG biosynthesis, but is required for LAM biosynthesis.[41] In the Emb enzymes, the N-terminal integral membrane domain is hypothesized to function in the recognition and potentially in flipping the DPA donor of lipid-linked acceptor

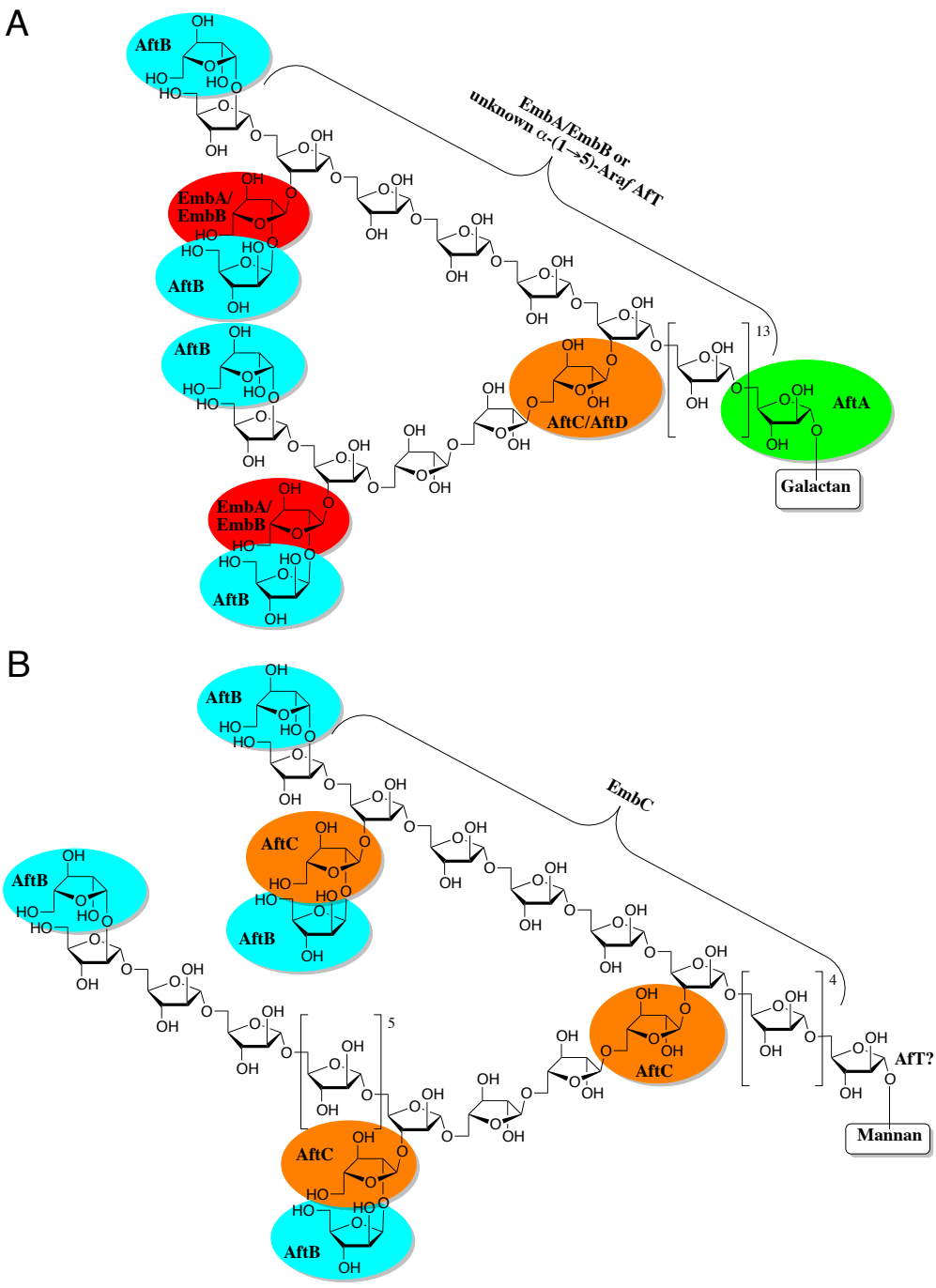


Figure 1-7. Structure of mAG arabinan (A) and LAM arabinan (B). Where known, the AfTs involved in adding a particular D-Ara*f* residue are shown.

substrates across the plasma membrane.[41] The C-terminal domain then functions in the extension of the arabinan chain.[42] A crystal structure of the C-terminal domain for EmbC, which contains a lectin-like carbohydrate binding module (CBM), was recently reported by Besra and co-workers (Figure 1-8).[43] This CBM appears to function in the recognition of the arabinan acceptor substrate. However, further studies are required to determine the exact role this domain plays in EmbC activity. Further work is also needed to establish the structure and function to the additional domains of EmbC, as well as the other AfTs involved in arabinan biosynthesis. In particular, interactions of these enzymes with the DPA donor remain to be elucidated.

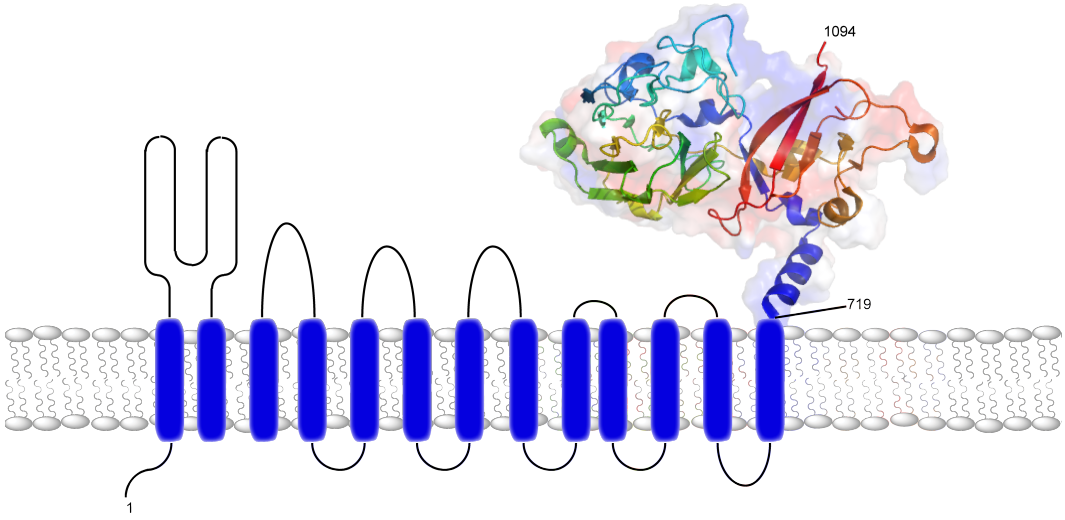


Figure 1-8. The structure of EmbC contains 13 trans-membrane domains. The crystal structure of the C-terminal amino acids 719–1094 (PDB id: 3PTY) is shown.[43] This region contains a lectin-like domain involved in acceptor substrate recognition.

1.3 Biosynthesis of D-Galactofuranose

D-Gal*f* is by far the most abundant and widespread hexose sugar existing in the furanose ring form, and is present in the glycoconjugates of bacteria, fungi, algae and protozoa including a number of human and animal pathogens.[1, 12, 52] One of the most impressive displays of D-Gal*f* is found in the cell wall galactan of mycobacterial species, which is composed entirely of galactose in the furanose ring form,[7, 9] and which is essential for the viability of the bacterium.[44] D-Gal*f* is also found in the lipopolysaccharide (LPS) of many Gram-negative bacteria, where it serves as a virulence factor. These include the O-antigen repeating unit of *Escherichia coli* strains,[45-48] *Actinobacillus pleuropneumoniae*,[49] and the *Klebsiella pneumoniae* galactan I (Figure 1-9).[6] In eukaryotic pathogens, D-Gal*f* residues have been found in a number of glycan structures, including the main antigenic determinant galactomannan of *Aspergillus* species and the lipid linked glycan of *Trypanosoma cruzi* and *Leishmania major*.[50] However, D-Gal*f* is not found in mammalian glycans.

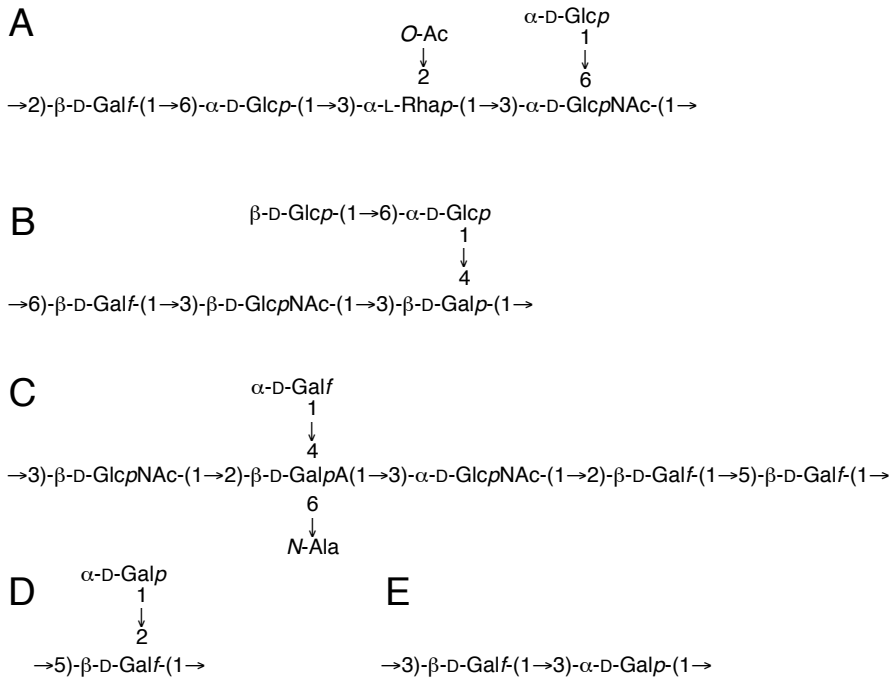


Figure 1-9. Representative D-Gal*f* containing glycans of gram-negative bacteria. (A) The LPS O-antigen repeating units of *E. coli* K-12, (B) O164, (C) O167, (D) *A. pleuropneumoniae* serotype 14 and (E) the galactan I repeat unit of *K. pneumoniae*.

Both eukaryotes and prokaryotes use the same sugar nucleotide, UDP-D-Gal*f,* as the immediate precursor to D-Gal*f* found in their various glycoconjugates.[50, 51] The biosynthesis of UDP-D-Gal*f* from UDP-D-Glc*p* involves the action of two enzymes: UDP-galactose 4-epimerase (GalE) facilitates the epimerization of UDP-D-Glc*p* to UDP-D-Gal*p*,[52] and UDP-galactopyranose mutase (UGM) catalyzes the pyranose-to-furanose ring contraction of UDP-D-Gal*p* (Figure 1-10).[53] Only the latter activity will be discussed here. The first UGM enzyme was identified in *E. coli* K-12 where it is encoded by the *orf6* gene, later renamed *glf*.[53] Homologs of *glf* have since been identified in *K. pneumoniae*,[5] mycobacterial species,[54] and various eukaryotic

pathogens,[55, 56] in addition to numerous other bacterial and eukaryotic organisms where the activity and function of *glf* has not been experimentally determined (See section 1.3.3). In addition to UGM, galactofuranosyltransferase (GlfT) enzymes catalyze the transfer of D-Gal*f* from UDP-D-Gal*f* onto the glycoconjugate. Here, we will discuss recent studies to elucidate the structure, mechanism, and specificity of these two groups of enzymes.

Figure 1-10. Reaction catalyzed by UGM. The equilibrium favors the pyranose sugar in a ratio of ~9:1.

1.3.1 Mechanism of UGM

Since the first UGM, from *E. coli,* was reported,[53] there has been a significant effort invested to elucidate both the structure and mechanism of this enzyme, which catalyzes the ring contraction of UDP-D-Gal*p* to form UDP-D-Gal*f*. UGMs catalyze an equilibrium process in which the pyranose ring form is favored over the furanose ring form, in a ratio of approximately 9:1. In early studies, Blanchard and coworkers, using positional isotope exchange (PIX) experiments, showed that the C–O bond connecting the galactose anomeric carbon to UDP is cleaved during the isomerization from UDP-D-Gal*p* to UDP-D-Gal*f*.[57] On the basis of the PIX results, they proposed a reaction mechanism

proceeding through a 1,4-anhydro-galactopyranose intermediate (**1.1**, Figure 1-11). Vincent and coworkers later synthesized the proposed intermediate **1.1** and evaluated it as a substrate for UGM.[58] Incubating **1.1** with UGM failed to produce either UDP-D-Gal*f* or UDP-D-Gal*p* at any of the concentrations tested (0.5–5 mM), largely ruling out **1.1** as an intermediate in the UGM reaction.

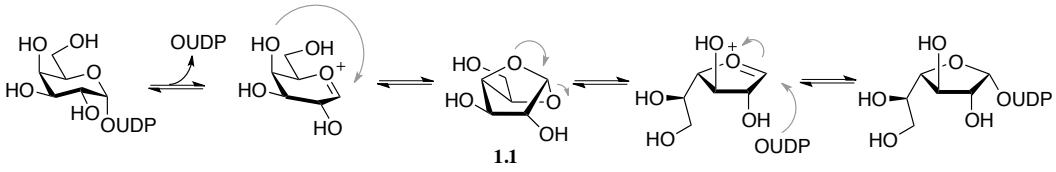


Figure 1-11. A proposed UGM mechanism involving formation of the bicyclic intermidiate **1.1**.

A number of other substrate analogs have been designed specifically to probe the mechanistic details of the UGM reaction. Both Barlow and Blanchard, and Zhang and Liu independently reported that 2′′- and 3′′-deoxyfluoro UDP-D-Gal derivatives in both the pyranose[59, 60] and furanose[61] ring forms, respectively, can act as substrates for UGM. These results establish that the UGM mechanism does not involve either oxidation or dehydration at C-2′′ or C-3′′, or the transfer of UDP from C-1′′ to either of these positions during the reaction. These studies were later expanded to include the 6′′-deoxyfluoro UDP-D-Gal*p* and UDP-D-Gal*f* derivatives,[69, 70] both of which also serve as UGM substrates. With the exception of a 4′′-deoxyfluoro UDP-D-Gal*p* derivative reported by Boons and coworkers,[60] all of the fluorinated derivatives tested to date function as UGM substrates. However, the rate of the reaction (k_{cat}/K_M) is dependent on the position of fluorination in the substrate.[59, 62] The lowest rates were observed for

the derivative with fluorine at C-2′′, which suggest an oxocarbenium ion, or another positively-charged species is formed during the UGM reaction mechanism.

A peculiar feature of UGM is that it is a flavo-enzyme[53]—possessing one non-covalently bound FAD cofactor per monomer unit of the enzyme[63]—even though the reaction catalyzed by the enzyme is not a redox process (i.e., the starting material and product exist in the same oxidation state). In addition, the UGM reaction requires a reduced flavin and, in the presence of an external oxidant (i.e., $K_3Fe_3(CN)_6$, the enzyme shows no activity.[64] Modeling studies,[64] which will be discussed more in the next section, suggest the UDP-D-Gal substrate binds adjacent to the isoalloxazine ring of the FAD cofactor. These results, taken together with prior chemical studies, support a direct role for the FAD cofactor in catalysis.

Potentiometric titration of the UGM enzyme FAD cofactor suggests that the fully reduced flavin species exists as the anionic $FADH^-$ rather then as the more common neutral $FADH_2$ species.[65] In addition, the neutral semiquinone radical species ($FADH^{\bullet}$) is stabilized in the presence UDP-D-Gal*p*. Although this stabilization may have no direct role in the UGM mechanism, it was suggested these results support a mechanism involving single electron transfer to UDP-D-Gal. Additional studies, using UGM reconstituted with 1-deaza-FAD or 5-deaza-FAD, also support this role for the FAD cofactor.[66] FAD and 1-deaza-FAD are capable of both one and two electron processes, whereas the 5-deaza-FAD is restricted to net two electron reactions. Only UGM reconstituted with 1-deaza-

FAD showed activity, consistent with a radical mechanism (i.e. route C in Figure 1-13).

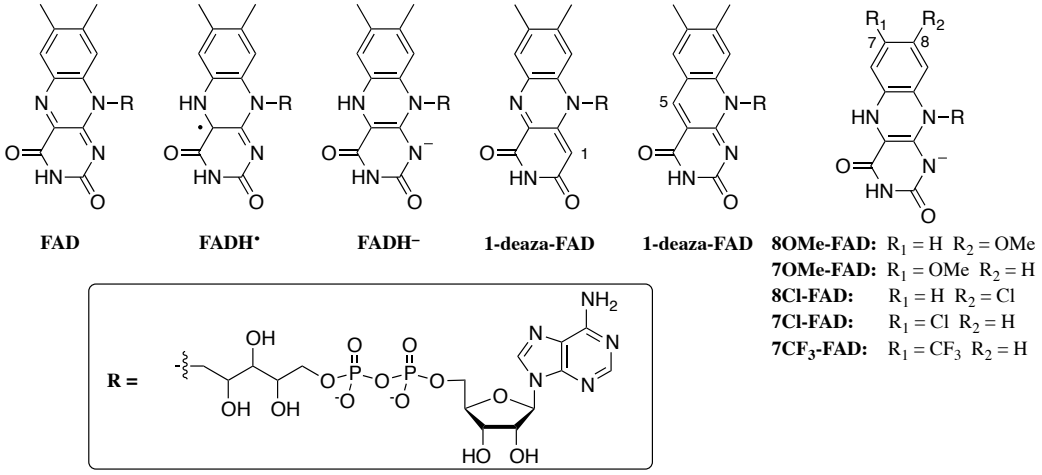


Figure 1-12. Different oxidation states and derivatives of FAD discussed in this chapter.

Kiessling and coworkers suggested an alternative role for the FAD cofactor, one that does not involve single electron transfer. Instead, they suggest a novel mechanism in which the N5 of the reduced $FADH^-$ species functions as a nucleophile to attack the anomeric position of galactose ring displacing UDP to form a covalent adduct (i.e., **1.2** and **1.4**, Figure 1-13).[67] The isomerization then proceeds through an iminium ion intermediate **1.3**, evidence of which they could detect spectrophotometrically. They were also able to indirectly detect **1.3** by treating the UGM reaction mixture with $NaCNBH_3$ to give **1.5**, the structure of which was confirmed by mass spectrometry[67] and, more recently, NMR spectroscopy.[68] There are, however, multiple mechanisms that could lead to the formation of **1.2**/**1.4**: One is through direct S_N2 attack (*Route A*), another

involves an oxocabenium ion and an SN1 mechanism (*Route B*), the third requires a single electron transfer followed by radical coupling (*Route C*).

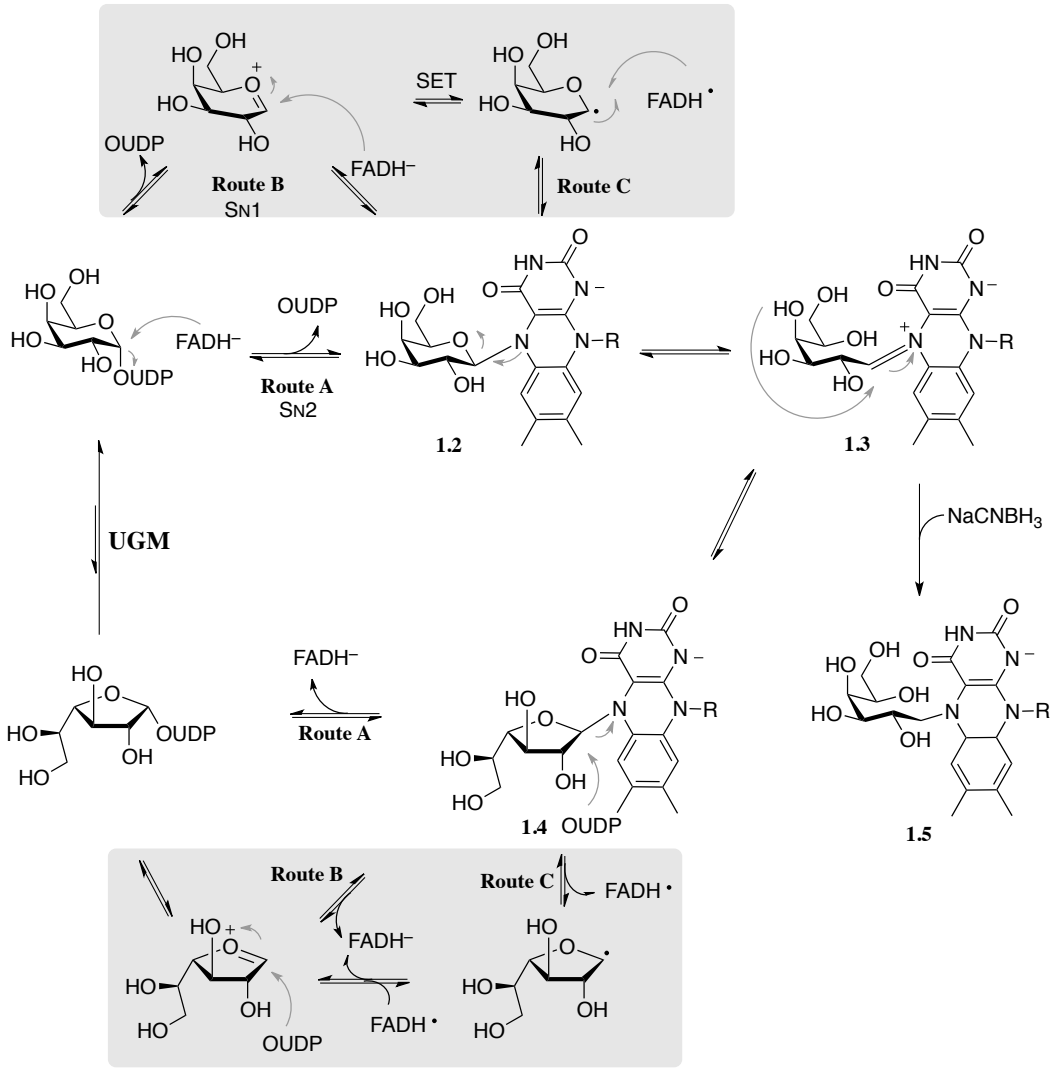


Figure 1-13. Proposed UGM reaction mechanisms. The mechanism involving formation of **1.2** via SN2 attack (route A) agrees the best with the currently available data.[69] The putative flavin-derived iminium species **1.3** has been observed spetrophotometrically and indirectly after reduction with $NaCNBH_3$ to give **1.5**.[67]

To distinguish between these putative mechanisms, Liu and coworkers earlier this year reconstituted UGM with FAD derivatives substituted at C7 and

C8 with various electron-donating or electron-withdrawing groups.[69] By measuring linear free energy relationships between the log(k_{cat}) of the reaction by UGM versus the *para*- and *meta*-substituent effects of each substituted FAD derivative, they showed that the rate of steady state turnover is, at least in part, affected by the electronegativity of the flavin N5. In addition, PIX experiments with UGM reconstituted with 5-deaza-FAD, which lacks the flavin N5, show that little to no cleavage of the anomeric C–O bond occurs, largely ruling out an S_N1 mechanism (Figure 1-13, *Route B*). Taken together, these experiments support an S_N2 mechanism (Figure 1-13, *Route A*) where formation of the covalent (**1.2** or **1.4**) or iminium ion intermediate (**1.3**) is at least partially rate limiting.

In more recent studies of the *T. cruzi* UGM, Oppenheimer *et al.* were able to measure specific rate constants for individual steps of the UGM mechanism using rapid reaction kinetics experiments.[70] Substrate dissociation is not rate limiting, as carrying out the reaction in different viscosity solvents (i.e., in water with increasing concentrations of glycerol) showed no effect on the reaction rate. Instead, the kinetic measurements in this study also support an S_N2 mechanism for the UGM reaction, where formation of the iminium ion intermediate is at least partially rate limiting.[70]

1.3.2 Structure of bacterial UGMs

In addition to the mechanistic studies described above, there has been a considerable amount of effort directed at elucidating the structure of UGMs.

Naismith and coworkers solved the first crystal structure of UGM, from *E. coli*, in 2001.[64] The protein exists as a homodimer with each monomer unit binding a single FAD molecule. In the structure, the FAD is bound predominantly in a domain composed of both the C- and N-terminus of the polypeptide. Adjacent to the isoalloxazine ring is a cleft proposed to bind the UDP-D-Gal substrate. Although this 2001 structure contained no bound ligand, Naismith and coworkers were able to propose a number of putative binding interactions based on site-directed mutagenesis and comparison to other UDP-D-Gal binding proteins.[64] Specifically, they proposed that a tryptophan 156 stacks against the uridine ring, and that four conserved tyrosine residues hydrogen bond to the Gal hydroxyl groups although these interactions would later prove incorrect (see section 1.3.2.2 for further discussion). Mutation or any of these residues resulted in a substantial decrease in UGM activity.[64]

Naismith and coworkers subsequently reported the structures of two additional UGM structures in 2005: those of *M. tuberculosis* with an oxidized (inactive) FAD moiety, and *K. pneumoniae* with both oxidized and reduced (active) FAD moieties.[71] The overall three-dimensional structure and domain organization is highly conserved in all of these structures (Figure 1-14), despite showing only moderate sequence homology (38–48% identity) between the three UGMs.

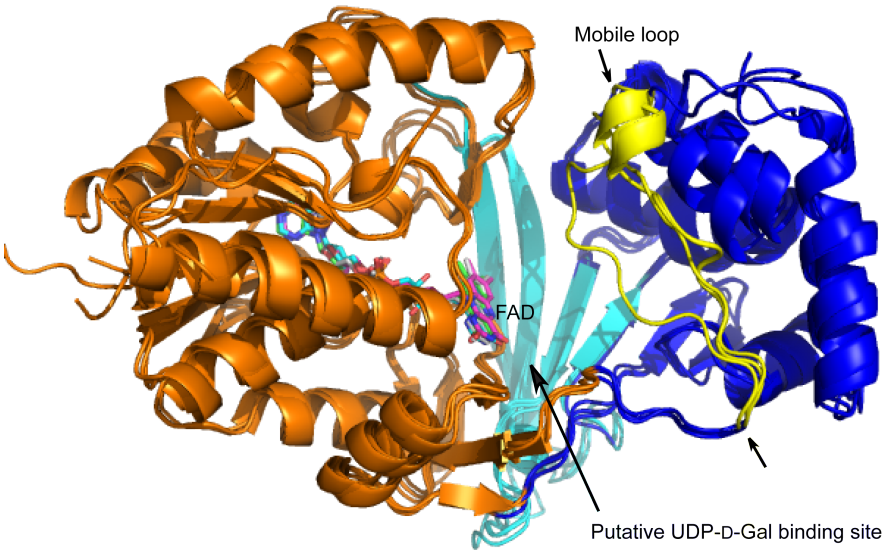


Figure 1-14. Overlay of the *E. coli* (PDB id# 1I8T chain A), *M. tuberculosis* (PDB id# 1V0J chain A), and *K. pneumoniae* (oxidized PDB id# 2B17, reduced PDB id# 2B18) UGM structures. The greatest divergence in the structure occurs in the mobile loop region (shown between arrows).

1.3.2.1 Mobile loop movement

The structure of the UGM described above varied predominantly in a loop region consisting of amino acid residues adjacent to the putative UDP-D-Gal binding site. This loop adopts either a "closed" conformation (*E. coli* UGM),[64] lying over the proposed substrate binding pocket, or an "open" solvent-exposed conformation (*K. pneumoniae* and *M. tuberculosis* UGM, Figure 1-14).[71] Further site-directed mutagenesis experiments, by Sanders and coworkers, identified an arginine in this mobile loop (R174 in *K. pneumoniae*, R170 in *E. coli*, and R180 in *M. tuberculosis*) that plays an essential role in UGM activity.[72]

Mutating this amino acid to alanine resulted in a complete loss of activity. Molecular modeling, based on saturation transfer difference (STD) NMR spectroscopic data, suggests this arginine residue interacts with the β-phosphate of UDP-D-Gal to stabilize the negatively charged phosphates.[73, 74] Yao *et al.* subsequently showed—through a combination of tryptophan fluorescence measurements, site-directed mutagenesis, STD-NMR spectroscopic studies, and molecular dynamics simulations—that binding of the substrate to UGM induces closure of this mobile loop, which is required for the enzyme activity.[75]

In 2009, Kiessling and coworkers reported the first structure of a UGM co-crystalized with a ligand: UDP-D-Glc*p*, a species that is not a substrate for the enzyme.[76] The UDP-D-Glc*p* serves as a moderate ligand for UGM, binding only 2-fold weaker than the native UDP-D-Gal*p* substrate, but the equatorial 4″-OH group precludes its turnover by the enzyme to the furanose ring form. Binding to UDP-D-Glc*p* induces a conformational change in the enzyme bringing the arginine 174 of the mobile loop within 4.6 Å of the pyrophosphate of the ligand,[76] similar to what was predicted in molecular dynamics simulations.[73, 75] The same year, Keissling and coworkers and Sanders and coworkers subsequently reported structures of the *K. pneumoniae* UGM,[68] and a *Deinococcus radiodurans* UGM,[77, 78] in complex with the native UDP-D-Gal*p* substrate. In both of these structures, the mobile loop is closed over the substrate-binding pocket placing the mobile loop arginine 174 (R196 in *D. radiodurans*) within 2.5–2.8 Å from the pyrophosphate group of the substrate (Figure 1-15). This mobile loop movement is essential for UGM activity. Thus,

developing small molecules that interfere with this movement could provide a possible method to inhibit UGM.

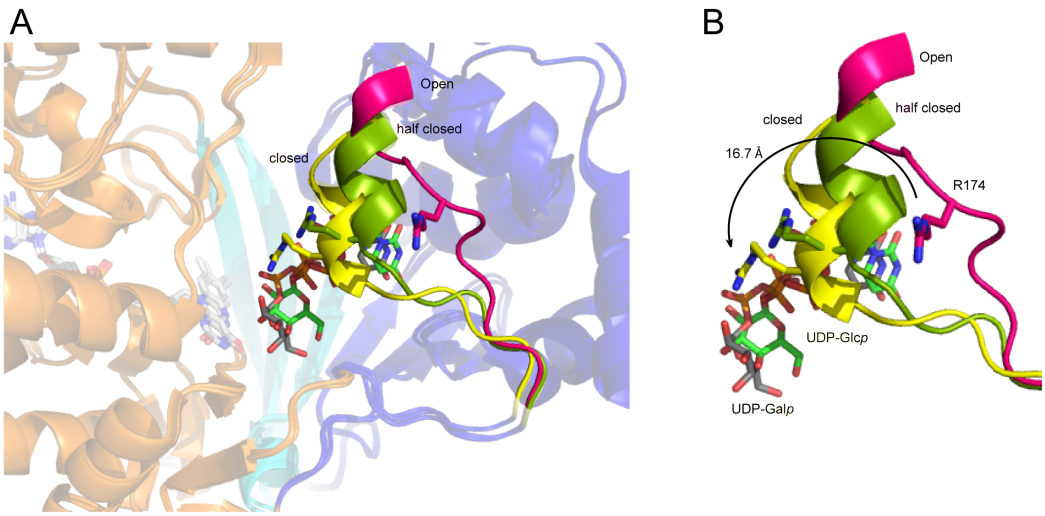


Figure 1-15. Mobile loop movement in *K. pneumoniae* UGM. Shown is an overlay of the reduced enzyme in the inactive "open" state (PDB id# 2B18), the structure with UDP-D-Glc*p* bound (PDB id# 3GF4) in the "half closed" state, and the structure with UDP-D-Gal*p* bound (PDB id# 3INT) in the "closed" state (A). The mobile loop movement is highlighted (B). R174 in the mobile loop region moves ~16 Å from the "open" to the "closed" state upon substrate binding and interacts with the β-phosphate through an electrostatic interaction,[68] consistent with molecular dynamics predictions.[73]

1.3.2.2 Ligand binding interactions and substrate specificity

The structures of UGM in complex with both UDP-D-Glc*p*[76] and UDP-D-Gal*p*[68, 78] aided the identification of a number of binding interactions occurring in the active site of these enzymes (Figure 1-16 and Figure 1-17). The uridine base stacks against a conserved tyrosine residue in the active site (Y155 in *K. pneumoniae*, Y179 in *D. radiodurans*) and not the tryptophan originally

proposed by Naismith and coworkers in 2001.[64] Instead, this tryptophan forms hydrogen bonds to the 2′- and 3′-hydroxyl groups of the ribose moiety. All of the galactose hydroxyl groups of the UDP-D-Gal*p* substrate form at least one direct hydrogen bond, or water-mediated hydrogen bond, to the surrounding active site amino acid side chains. One of the most notable features of these structures is that the D-Gal*p* binds in the active site with the anomeric carbon aligned with N5 of the FAD poised for nucleophilic attack (Figure 1-16 A).[68] In fact, the *D. radiodurans* structure shows partial electron density forming between N-5 and the anomeric carbon of D-Gal*p* (Figure 1-17 A).[78] In complex with UDP-D-Glc*p*, which does not serve a UGM substrate, the anomeric carbon is pointing away from FAD, preventing any possible reaction from taking place (Figure 1-16 B).[76] This difference in the binding modes may explain, at least in part, how UGM discriminates between these two structurally similar ligands (differing only at one stereocenter).

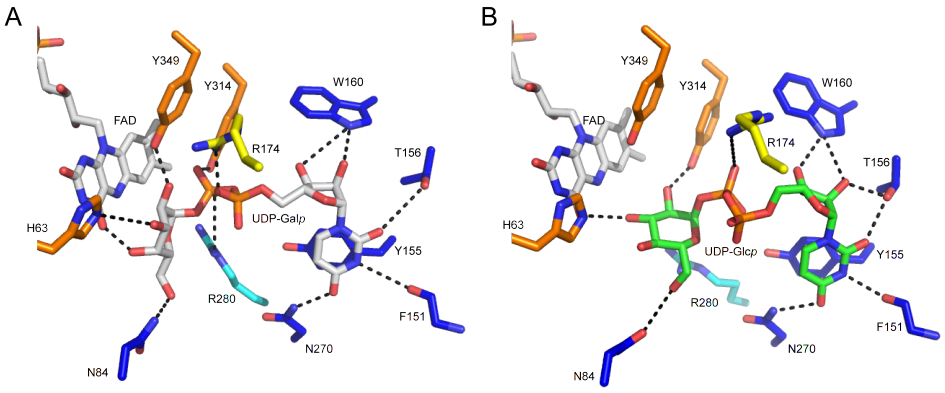


Figure 1-16. Active site interaction for the *K. pneumoniae* UGM in complex with UDP-D-Gal*p* (PDB id# 3INT) (A), and UDP-D-Glc*p* (PDB id# 3GF4) (B). A number of the interactions are shown as dashed lines. The uridine base of both ligands stacks against Y155. W160 form hydrogen bonds to the 2′-OH and 3′-OH of ribose.

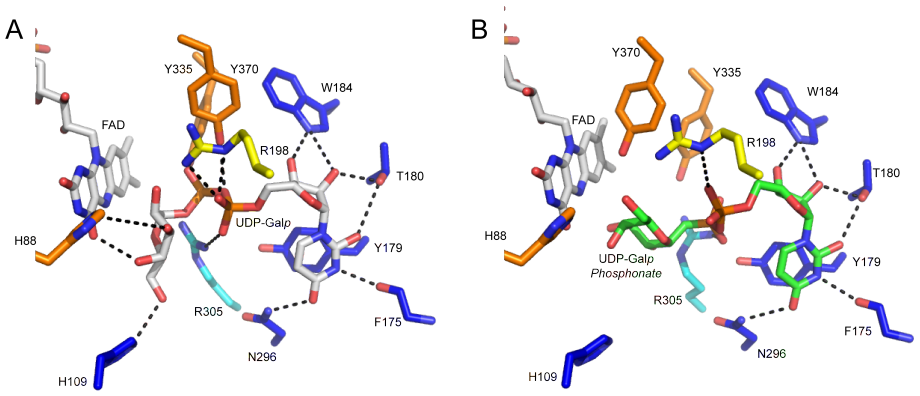


Figure 1-17. Comparison of the binding interactions of UDP-D-Gal*p* (A) and a C-phosphonate derivative of UDP-D-Gal*p* (B) to the *D. radiodurans* UGM (PDB id# 3HDY and 3MJ4, respectively).

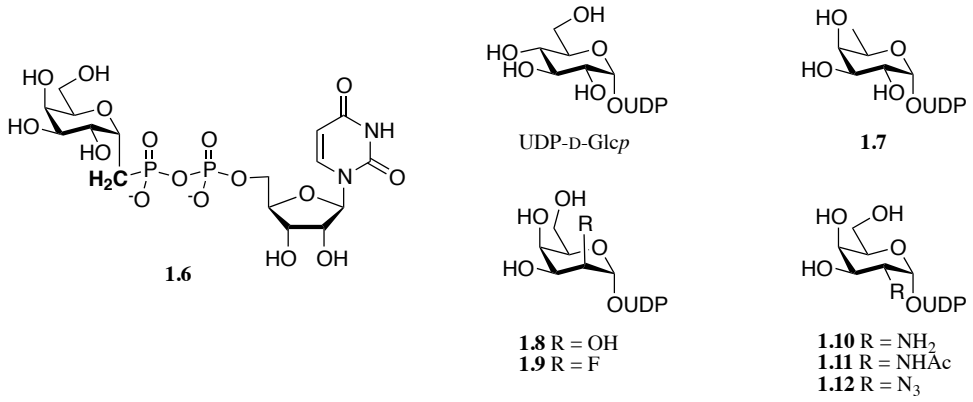


Figure 1-18. Analogs of UDP-D-Gal*p* used to study UGM specificity.

Jakeman and coworkers have also prepared C-phosphonate **1.6** (Figure 1-18), designed as a non-hydrolysable isostere of UDP-D-Gal*p*, that was then co-crystallized with the *D. radiodurans* UGM by Sanders and coworkers.[79] This derivative was designed to mimic the native substrate binding interaction without being turned over by the enzyme, thus making it a potent UGM

inhibitor. However, the observed binding to **1.6** does not match the native binding mode of UDP-D-Gal*p* to UGM (Figure 1-17 B). Instead the D-Gal*p* is flipped away from the FAD cofactor, in turn disrupting the normal binding interactions and making **1.6** a poor ligand for UGM.

In addition to the deoxyfluoro analogs discussed in section 1.3.1, a number of other C-2′′ and C-6′′ modified UDP-D-Gal*p* derivatives (Figure 1-18) have also been evaluated as UGM substrate to examine the enzyme's specificity. Field and coworkers were unable to detect any UGM activity using the 6′′-deoxy derivative **1.7**,[62] indicating that the hydrogen bond to the 6′′-hydroxyl group seen in UGM crystal structures is important for activity. They also observed no activity with 2′′-epimeric derivatives **1.8** and **1.9**. A 2′′-amino derivative **1.10** did show turnover at a rate about 20-fold lower then the native substrate.[62] It is likely that this analog still maintains the same hydrogen boding interactions at C-2′′ as the native UDP-D-Gal*p* substrate. 2′′-Acetamido and azido derizatives **1.11** and **1.12**, however, showed no turnover. It was hypothesized that the extra steric bulk of the acetamido group and the loss of hydrogen bonding for the azido derivative prevents proper binding of these analogs.

1.3.3 Structure and Activity of Eukaryotic UGM

The first example of a eukaryotic UGM was reported by McNeil and coworkers in 2005.[55] Here they used to a combinatorial bioinformatics approach to identify homologs of the *glf* gene in a number of eukaryotic species—including protozoan pathogens *Leishmania major* and *T. cruzi*, and the fugal

pathogen *Cryptococcus neoformans*—that when expressed in *E. coli* showed UGM activity both *in vitro* and *in vivo*. This investigation was followed by an independent report from Bakker *et al.* identifying UGM in *Aspergillus fumigatus* and again in *L. major*.[56] These eukaryotic enzymes share less then 20% sequence identity with their bacterial counterparts; however, conservation is mostly seen in the amino acid residues involved in cofactor and substrate binding.[56] A number of these enzymes have subsequently been expressed recombianantly and biochemically characterized.[70, 80, 81] All are flavo-proteins and display a similar activity[80, 81] and mechanism[70] to previously reported bacterial UGMs.

Tanner and coworkers and Sanders and coworkers reported the first structures of a eukaryotic UGM, from *A. fumigatus*, independently, in 2012.[82-84] This enzyme exists as a homotetramer, with each monomer unit binding a molecule of FAD (Figure 1-19 C). Like the bacterial UGMs, each *A. fumigatus* UGM monomer is composed of three domains, with an extended α-helix in domain 1 and domain 2 that are responsible for the majority of the protein–protein contacts between the monomer units (Figure 1-19 D). The crystal structure of *T. cruzi* UGM was also reported in 2012.[85] In this case, the enzyme exists as a monomer, and it is not clear what effect, if any, the quaternary structure seen in the *A. fumigatus* enzyme has on its activity. Apart from the quaternary structure, the bacterial, fungal and protozoan UGM are well conserved in the overall three-dimensional architecture and domain organization (Figure 1-19).

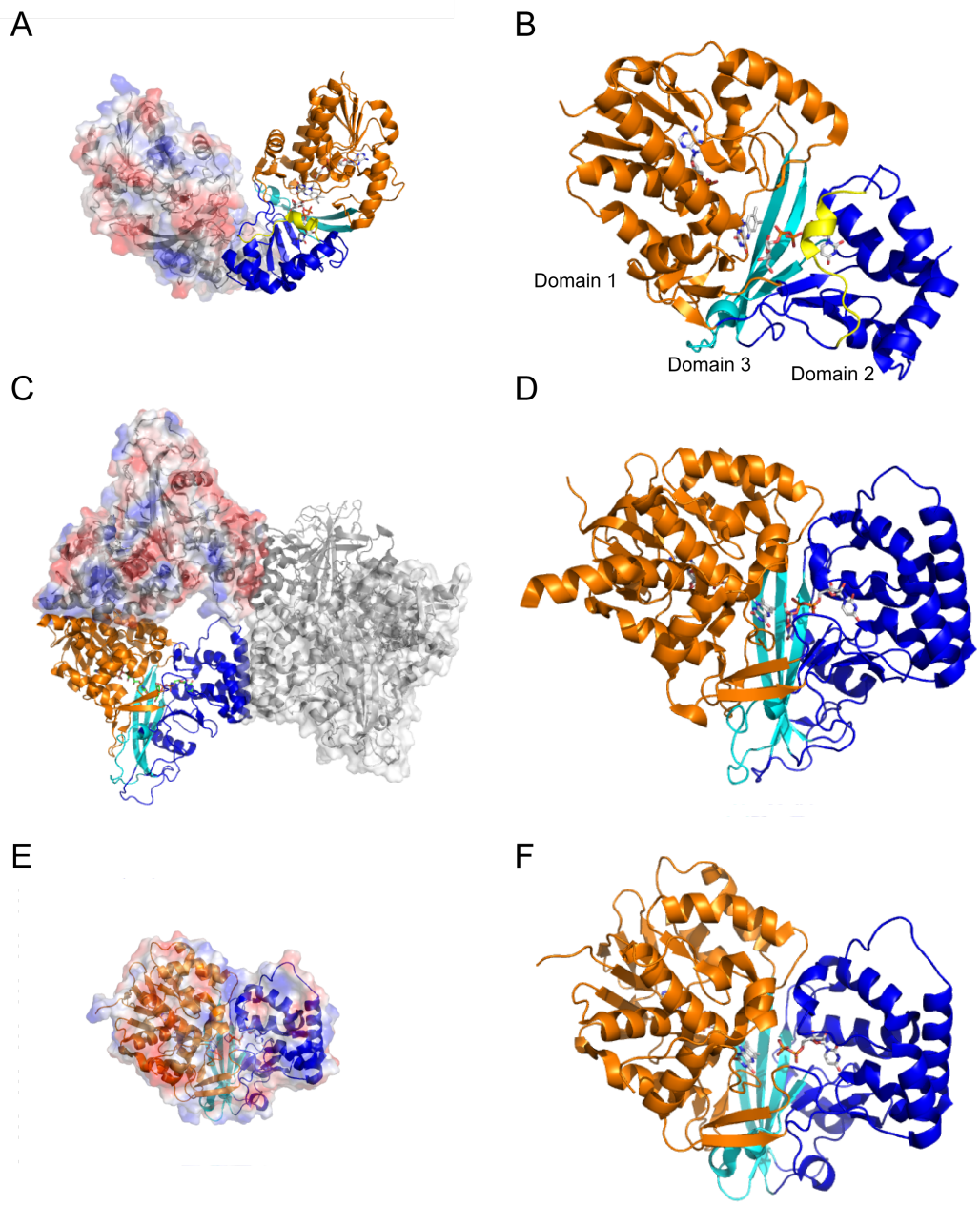


Figure 1-19. Comparison of the structure and domain organization of UGM from bacteria (*K. pneumoniae*, A, B), fungi (*A. fumigatus*, C, D), and protazoa (*T. cruzi*, E, F). Bacterial UGM studied to date are present as dimers (A), with three domains per monomer (numbered in B) and a mobile loop region (B). The *A. fumigatus* UGM is present as a homotetramer (C) and each monomer again consists of three domains and a mobile loop (D). Other identified eukaryotic UGM, including the *T. cruzi* UGM, are present as monomers (E) again consisting of three domains (F). All use a reduced $FADH^-$ co-factor.

1.3.3.1 Mutations of A. fumigatus *UGM disrupt proper substrate binding*

A number of the substrate binding interactions are conserved between the *A. fumigatus* UGM, *T. cruzi* UGM and their bacterial equivalents. Notably, the essential mobile loop arginine (R182 in *A. fumigatus*, R176 in *T. cruzi*) and a second essential active site arginine residue (R327 in *A. fumigatus*, R327 in *T. cruzi*), which interact with the pyrophosphate backbone of UDP-D-Gal in the bacterial enzymes, are also present in these eukaryotic UGM structures (Figure 1-20 A).[82, 84, 85] Site directed mutagenesis of either arginine to lysine resulted in a ten to hundred-fold decrease in activity for the *A. fumigatus* UGM, whereas, mutagenesis to alanine resulted in over a 10,000-fold loss of activity.[84]

Sanders and coworkers were able to obtain crystal structures of a number of these mutants in complex with either UDP-D-Gal*p* or UDP that help explain the observed mutagenesis results.[84] In both the R182K (Figure 1-20 B) and R327K mutants (Figure 1-20 C) the UDP or UDP-D-Gal*p* appear to bind in approximately the same orientation as seen in the native *A. fumigatus* UGM structure with the positively charged lysine still able to stabilize the ligand's negatively charged pyrophosphate group. However, in the R327A mutant (Figure 1-20 D) the alanine cannot stabilize the pyrophosphate backbone resulting in a UDP-D-Gal*p* binding mode with the D-Gal*p* anomeric carbon pointing away from the N5 of FAD, in turn preventing any reaction with the cofactor. Thus, it appears that these two positively charged arginines not only stabilize the negative charge of the UDP-D-Gal*p* pyrophosphate backbone, but

also orient the substrate for reaction with the FAD. These structures also explain previous mutagenesis results in the bacterial UGM enzymes.[72]

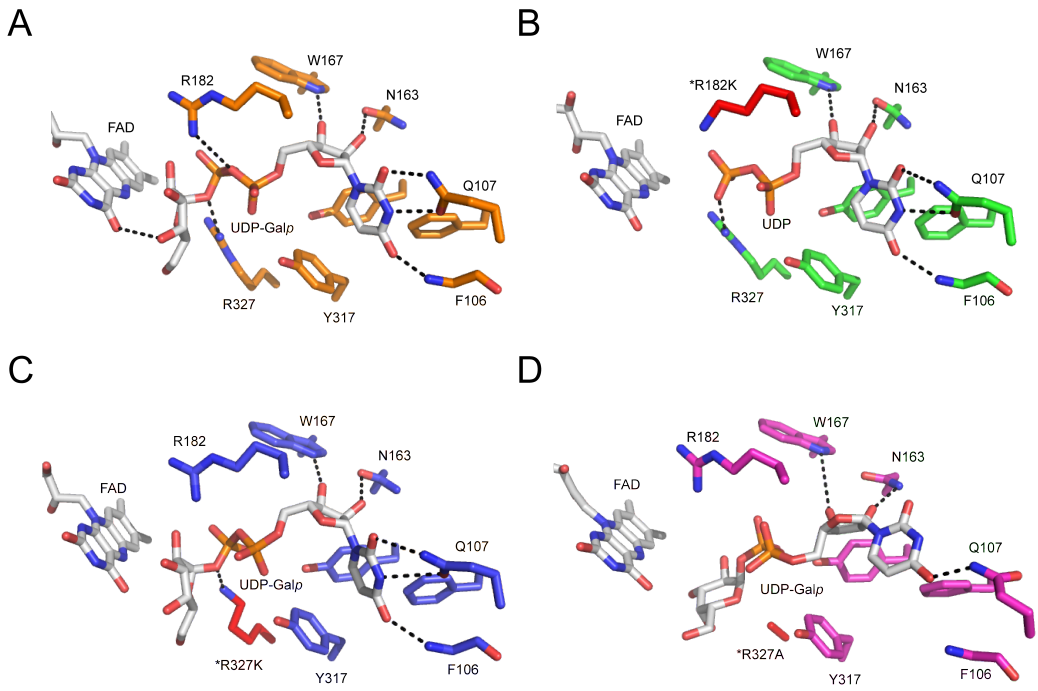


Figure 1-20. *A. fumigatus* arginine mutants show disrupted substrate binding. Many of the binding interactions are conserved between the *A. fumigatus* UGM and those of bacteria (A). R182K (B) and R327K (C) mutants appear to maintain the native substrate-binding mode. The R327A mutant (D) binds so that the anomeric carbon of UDP-D-Gal*p* is pointed away from the FAD cofactor, preventing the UGM reaction. The mutated amino acids are indicated with an asterisks.

1.3.4 Putative *glf* gene homologs in other organisms

In addition to the UGMs described above, rapid genome sequencing has allowed for the identification of a large number of *glf* homologs in numerous bacterial and eukaryotic organisms. A search of the National Center for

Biotechnology Information (NCBI) databases reveals over three thousand putative *glf* sequences although it is likely that some of these are duplicates of the same sequence. However, few of these have had their function experimentally characterized.

Similar pyranose–furanose mutase enzymes play a role in the biosynthesis of other furanose sugar nucleotides. For example, in 2009 Wang *et al.* discovered a UGM homolog in the genome of *E. coli* O52, involved in deoxy thymidine 5′-diphosphate-α-D-6-deoxy-galactofuranose (TDP-D-Fuc*f*) biosynthesis.[86] The enzyme, encoded by the *fcf2* gene, shares 60% sequence identity with the UGM from *K. pneumoniae* and falls into the same protein family. Wang *et al.* demonstrated that this enzyme recognizes TDP-D-Fucopyranose and converts it to TDP-D-Fuc*f*, which serves as the source of D-Fuc*f* in the O52 LPS O-antigen repeating unit (Figure 1-21).[87] Like the UGM enzymes described above, the equilibrium of Fcf2 favors the pyranose ring form in a ratio of ~9:1. The substrate specificity of this enzyme, and specific binding interactions leading to recognition of the TDP-D-Fuc substrate, have not been reported.

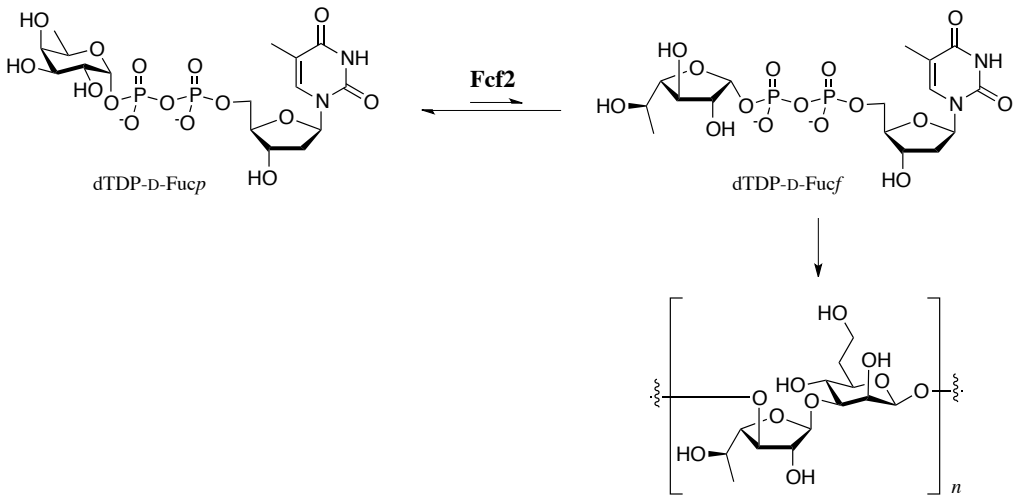


Figure 1-21. Fcf2 serves as a pyranose–furanose mutase enzyme in the biosynthesis of D-Fuc*f* found in the LPS O-antigen heteropolysaccharide from *E. coli* O52. The equilibrium of Fcf2 favours TDP-D-Fuc*p* in a ratio of ~9:1 consistent with UGM enzymes.

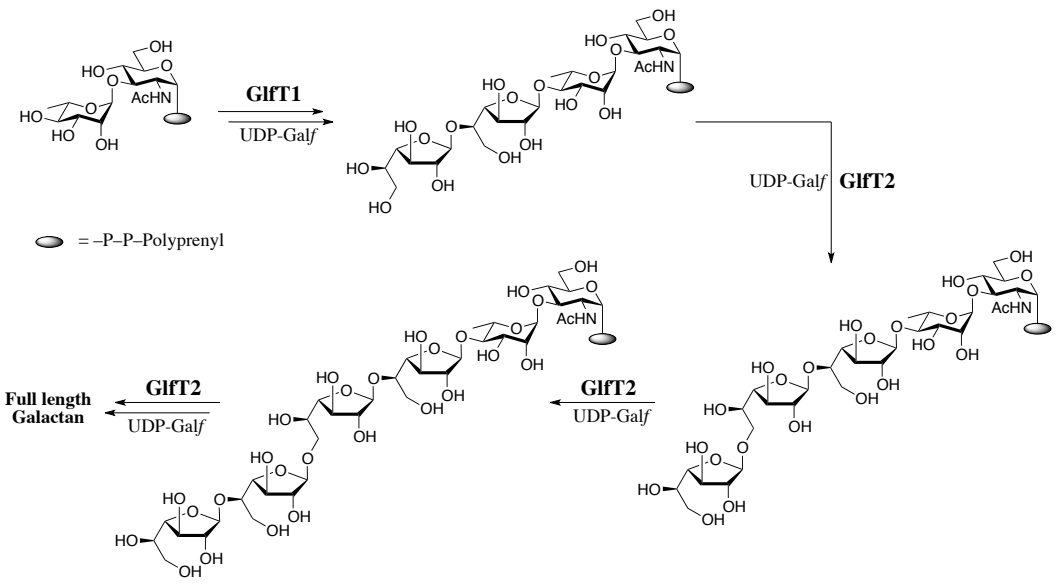
Two other UGM homologs have also been predicted to function in the biosynthesis of 3,6-dideoxy-D-glucofuranose (paratofuranose) and 6-deoxy-L-altrofuranose found in the LPS of *Yersinia pseudotuberculosis* serotype O11 and O1b strains, respectively.[88] However, biochemical or functional characterization of either of these enzymes is still required.

1.3.5 Galactofuranosyltransferases

A number of galactofuransosyltransferase (GlfT) enzymes have been identified in the past decade; however, compared to the UGM enzymes discussed above, their activities remain poorly characterized. In 1996 it was hypothesized that the *orf8* gene of *E. coli* K-12 functioned as a GlfT,[53] adding the non-reducing β-D-Gal*f* residue to the 6-OH of glucose in the LPS O-antigen

repeating unit. The WbbI protein encoded by this gene was later shown, *in vitro*, to transfer D-Gal*f* from UDP-D-Gal*f* onto an octyl linked α-D-Glc*p* acceptor.[89] This α-D-Glc*p* stereochemistry, which is the same as that found in the native acceptor, appears to be required for the enzyme, as it was unable to catalyze D-Gal*f* transfer to an octyl β-D-Glc*p* acceptor. Another example, the WbbO enzyme, from *K. pneumoniae*, functions as a β-(1→3)-GlfT in the biosynthesis of galactan I.[90] Whitfield and coworkers showed in 2001 that this enzyme is bifunctional and adds both the first α-(1→3)-D-Gal*p* residue to initiate galactan I biosynthesis, as well as all of the subsequent β-(1→3)-D-Gal*f* residues. Thus, the enzyme has dual pyranosyltransferase and furanosyltransferase activities.

There are also two bifunctional GlfT enzymes identified in the biosynthesis of the mAG galactan of mycobacterial species, which will be discussed in greater detail in Chapter 5. The first of these, GlfT1, is encoded by the *Rv3782* gene of *M. tuberculosis*,[91] and has both β-Gal*f*-(1→4)-α-Rha*p* and β-Gal*f*-(1→5)-β-Gal*f* GlfT activity. This enzyme adds the first and second D-Gal*f* residues to a lipid linked pyrophosphate-disaccharide acceptor to initial galactan biosynthesis (Scheme 1-1).[92]



Scheme 1-1. Biosynthesis of the mycobacterial mAG complex requires two bifunctional galactofuranosyltransferase enzymes, GlfT1 and GlfT2.

Brennan and coworkers discovered the second enzyme, GlfT2, encoded by the *Rv3808c* gene of *M. tuberculosis,* in 2000.[93] The enzyme has subsequently been expressed and purified in *E. coli*.[94, 95] GlfT2 is a bifunctional galactofuranosyltransferase, which adds the third and all subsequent D-Gal*f* residues to the growing galactan via alternating β-(1→5)- and β-(1→6)-glycosidic linkages. A number of synthetic galactan derivatives have been evaluated as substrates for GlfT2, which showed that a D-Gal*f*-(1→5)-β-D-Gal*f* or D-Gal*f*-(1→6)-β-D-Gal*f* disaccharide is the minimal structural motif required for GlfT2 activity.[102, 103] GlfT2 is a polymerase with only a single active site capable of catalyzing both the enzymes β-(1→5)- and β-(1→6)-transferase activities.[96, 97] The enzyme is also processive, catalyzing the addition of multiple

D-Gal*f* residues before the acceptor substrate dissociates from the enzyme active site.[98]

Kiessling and coworkers have also shown that GlfT2 is capable of forming full-length galactan polymers, containing 30–35 D-Gal*f* residues, without the need for additional GlfT enzymes.[99] The authors suggest that a distal lipid-binding site in the enzyme dictates chain length, by controlling the dissociation of the acceptor substrate. Using acceptor analogs with different lipid chains it was shown that the amount of polymerization catalyzed by GlfT2 is dependent on the lipid length.[99] However, acceptor substrates with the native lipid structure were never tested to determine if the same levels of polymerization are observed. It is also worth noting that mycobacterial galactan biosynthesis occurs at the plasma membrane, and thus the lipid linked acceptor substrate would likely be embedded in the membrane. Ng, Lowary and coworkers recently reported the first X-ray structure of GlfT2 and the structure does not provide evidence for a lipid-binding site in the enzyme.[100] Thus, more work is still required to determine how GlfT2 controls product length. Although there has been substantial progress in studying GlfT enzymes over the last decade, there is still much work to be done, specifically to identify binding interactions involved in substrate recognition.

1.4 UDP-L-Arabinofuranose and UAM Enzymes

Glycans containing L-Ara*f* are more common than those containing D-Ara*f*, and make up a large and important component of plant cell wall polysaccharides including the hemicellulose arabinoxylan, pectic

rhamnogalacturonan I and II (RGI and RGII), and arabinogalactan glycoproteins.[101] L-Ara*f* is the most abundant side chain in xylan hemicellulose fibers (β-D-Xyl*p*-(1→4)-β-D-Xyl*p*).[102, 103] Single L-Ara*f* residues are attached to the D-Xyl*p* backbone, typically at O-3, and the number of L-Ara*f* residues influences the solubility and cellulose-binding properties of the xylan fibers. Polymers containing high degrees of L-Ara*f* substitution tend to be more soluble and bind poorly to cellulose, compared to polymers with lower levels of L-Ara*f*.[104] Arabinan domains, consisting of an α-(1→5)-linked L-Ara*f* polysaccharides with single α-L-Ara*f* branch points at O-2 or O-3, are found as side chains in the pectin RGI polysaccharides (Figure 1-22). Despite the abundance, and importance, of L-Ara*f* in plant cell wall glycans, it was only recently that the first enzymes involved in L-Ara*f* biosynthesis were identified and experimentally characterized.[105, 106]

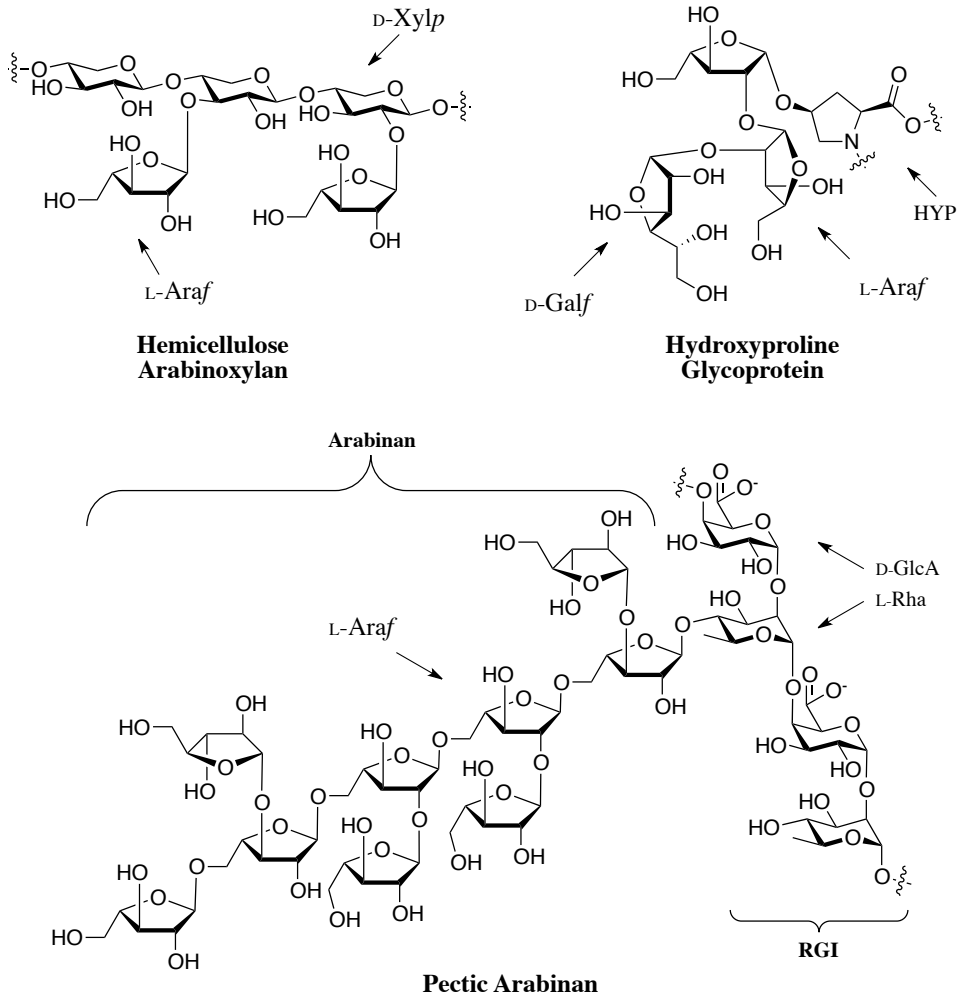


Figure 1-22. Representative examples of plant L-Ara*f*-containing glycans. L-Ara*f* residues are α-(1→3)- or α-(1→2)-linked to D-Xyl*p* in the hemicellulose arabinoxylan. L-Ara*f* are found in hydroxyproline (HYP) linked glycoprotein in *Chamydomonas* and land plants.[107] Arabinan domains, consisting of α-(1→5)-linked L-Ara*f* with α-(1→2)- and α-(1→3) branched L-Ara*f* domains are attached to the 4-hydroxyl group of 20–80% of the L-Rha residues in the backbone of pectin RGI polysaccharides.

1.4.1 Biosynthesis of UDP-L-Ara*f*

As with D-Gal*f* biosynthesis, plant L-Ara*f* is believed to be incorporated into pectin arabinan domains, arabinoxylans, and glycoproteins from a UDP-L-

arabinose sugar nucleotide donor. Plant UDP-L-Ara*p* has been shown to be derived either from UDP-D-Xyl*p* through the activity of a C4-epimerase,[108] or via the arabinose salvage pathway (Figure 1-23A).[109]

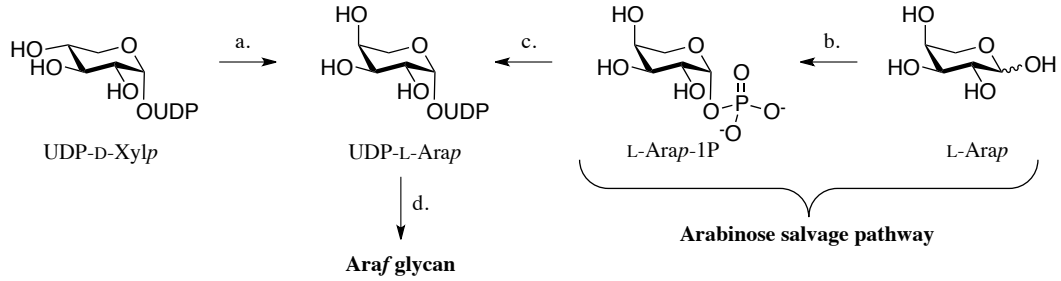


Figure 1-23. Putative biosynthesis of L-Ara*f* glycans. *a.* UDP-xylose epimerase catalyses the epimerization of UDP-D-Xyl*p* to UDP-L-Ara*p*. *b*. Arabinose kinase (AraK) converts free L-Ara*p* to L-Ara*p*-1P and *c*. UDP-L-arabinose pyrophosphorylase catalyzes formation of UDP-L-Ara*p*. *d.* It was originally proposed that AfT enzymes catalyzed the isomerization of L-Ara*p* to L-Ara*f* during the transfer from UDP-L-Ara*p* to L-Ara*f* glycans; however, this has shown to be incorrect (see discussion below).

In vitro experiments, with enzyme preparations isolated from the stellar and callus tissues of the common bean *Phaseolus vulgaris*, showed the specific incorporation of radiolabeled [1-^{3}H]-arabinose, from UDP-L-[1′′-^{3}H]-Ara*p*, into pectic arabinan polymers.[110] No lipid-linked intermediates were observed, which demonstrated that L-Ara*f* could be derived directly from a sugar nucleotide precursor. However, at that time, it was not known at what step the L-Ara*p* was isomerized to the furanose ring form, which is the predominant form incorporated into plant arabinans.

Feeding [^{3}H]-arabinose to *Spinacia oleracea* cell lines also showed incorporation of radiolabeled pentose sugars into arabinoxylan polymers.[111] The

[^{3}H]-pentose sugars identified in these polymers were predominantly found in the L-Ara*f* ring form, with a small amount L-Ara*p* and D-Xyl*p* also identified. In these experiments, the only pentose sugar nucleotide precursors identified from the cell extracts were UDP-L-Ara*p* and UDP-D-Xyl*p*, along with their corresponding L-Ara*p* and D-Xyl*p* 1-phosphate precursors (L-Ara*p*-1P, D-Xyl*p*-1P), and no UDP-L-Ara*f* was observed.[111] This work lead to the hypothesis that arabinan synthase (arabinosyltransferase, AfT) enzymes, which transfer L-Ara*f* residues to L-Ara*f* containing glycans, also catalyze the isomerization of L-Ara*p* to L-Ara*f* concurrent with glycosylation.

AfT activity has also been reported in Golgi membranes of mung bean (*Vigna radiata*) in the presence of UDP-L-[^{14}C]-Ara*p* using both endogenous and exogenous acceptor substrates.[112] Only a non-specific arabinofuranosidase was able to partially degrade (~25%) the radioactive polysaccharide products isolated from these incubations. Use of a specific *endo*-arabinanase, which cleaves only linear (1→5)-linked α-L-Ara*f*-chains, did not release any radiolabeled products. This suggests the product arabinan is highly branched, but the exact structure was not reported.

When a partially solubilized membrane and an exogenous arabinan acceptor substrate (**1.13**, Figure 1-24) was instead used, L-Ara*p* and not L-Ara*f* was transferred to O-3 of the non-reducing terminal L-Ara*f* residue,[112, 113] and no incorporation of L-Ara*f* was observed. This shows the L-AfT enzyme does not catalyze isomerization to L-Ara*f* during transfer as originally hypothesized. Using UDP-L-Ara*f* as the donor, (1→5)-linked α-L-Ara*f* residues were added[114]

with a degree of polymerization (DP) similar to that seen in the native mung bean arabinan. These studies showed that UDP-L-Ara*f*, and not UDP-L-Ara*p*, is the donor involved in pectin arabinan biosynthesis, and likely the biosynthesis of other plant L-Ara*f* glycans.

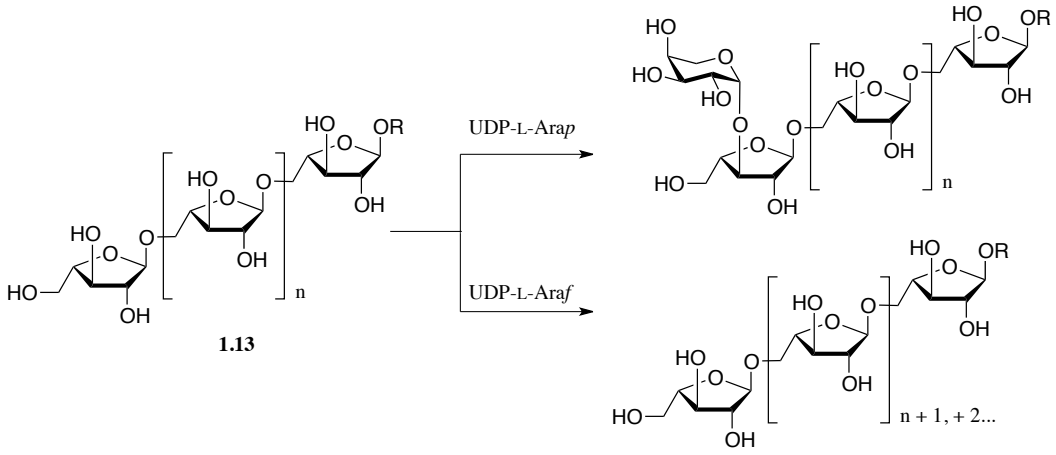


Figure 1-24. Glycans formed by incubating acceptor **1.13** with UDP-L-Ara*p* or UDP-L-Ara*f* in the presence of a Golgi membrane preparation from *Vigna radiata*. L-Ara*f* was only introduced when UDP-L-Ara*f* was used as a donor.

Further evidence for a UDP-L-Ara*f* donor in plants was provided when the first enzymes with UDP-arabinopyranose mutase (UAM) activity, which catalyze the reversible isomerization of UDP-L-Ara*p* and UDP-L-Ara*f*, were identified in rice (*Oryza sativa*).[105] Three UAMs (osUAM1, osUAM2, and osUAM3) were identified in this species; however, only osUAM1 and osUAM3 were able to catalyze the interconversion of UDP-L-Ara*p* and UDP-L-Ara*f*.

1.4.2 Plant UAMs

Plant UAMs belong to the reversibly glycosylated polypeptides (RGP) family of proteins,[105] which are specific to plants and are found in the *trans*-

Golgi. These proteins self-glycosylate using a variety of sugar nucleotide donors including UDP-D-Glc*p*, UDP-D-Xyl*p*, and UDP-D-Gal*p*.[115] In addition to catalyzing the isomerization of UDP-L-Ara (Figure 1-25), rice osUAM1 and osUAM3 also undergo self-glycosylation in the presence of UDP-D-[^{14}C]-Glc*p*, which could then be placed with UDP-D-Xyl*p*, UDP-D-Gal*p*, UDP-L-Ara*p*, or UDP-L-Ara*f*,[105] which will be discussed more in the following section.

The sequence identity between UAMs and previously reported RGPs associated with the plant Golgi, suggest these proteins likely share the same functions. This is supported by knock-out experiments that suggest RGPs play a role associated with plant cell wall biosynthesis, and in particular with pectin and xyloarabinan biosynthesis, [116, 117] consistent with functioning as UAMs. Thus, it is likely that all of the RGP proteins identified to date play a role in furanose sugar biosynthesis *in vivo*; however, this remains to be evaluated experimentally.

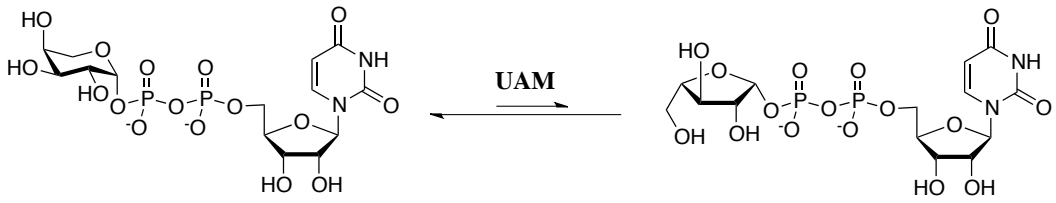


Figure 1-25. The reaction catalyzed by plant UAM enzymes. The equilibrium favors the pyranose ring form in an ~9:1 ratio.

1.4.3 Mechanism of plant UAMs

The exact mechanism for the UAM catalyzed interconversion of UDP-L-Ara*p* and UDP-L-Ara*f* is not known, but it is likely related to the RGP activity. Both osUAM1 and osUAM3 contain a DXD motif (D110 D111 D112 in

osUAM1, D108 D109 D110 in osUAM3), characteristic of GTA family glycosyltransferase enzymes, which is hypothesized to bind Mn^{2+} and stabilize the sugar nucleotide substrate (Figure 1-26). In UAMs, this DXD motif is essential for mutase activity and mutation of D112N in osUAM1 results in a complete loss of function.[106] The nonfunctional rice osUAM2 lacks a DXD motif and is replaced by DDN. Converting the DDN to DDD, however, did not restore UAM activity in osUAM2; therefore, the DXD motif is not the only factor required for a functional protein. In addition, the arginine residue (R158 in osUAM1 and R156 in osUAM3), which is found to be reversibly glycosylated with UDP-D-Glc*p* via the RPG activity of osUAMs, is also essential for the mutase activity of the enzyme.[106] Glycosylation of this arginine residue was also observed with UDP-L-Ara*f*, where the arginine may play a similar role as fulfilled by $FADH^{-}$ in the UGM mechanism (Figure 1-27).

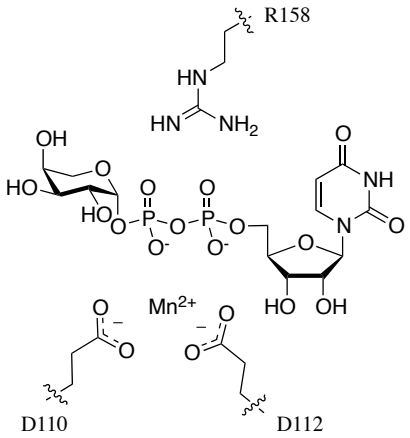


Figure 1-26. Active site amino acids essential for activity in osUAM1 and osUAM3 (numbering shown for UAM1) from *O. sativa*. UAM2, which lacks the DXD motif (DDN in UAM2), is inactive.

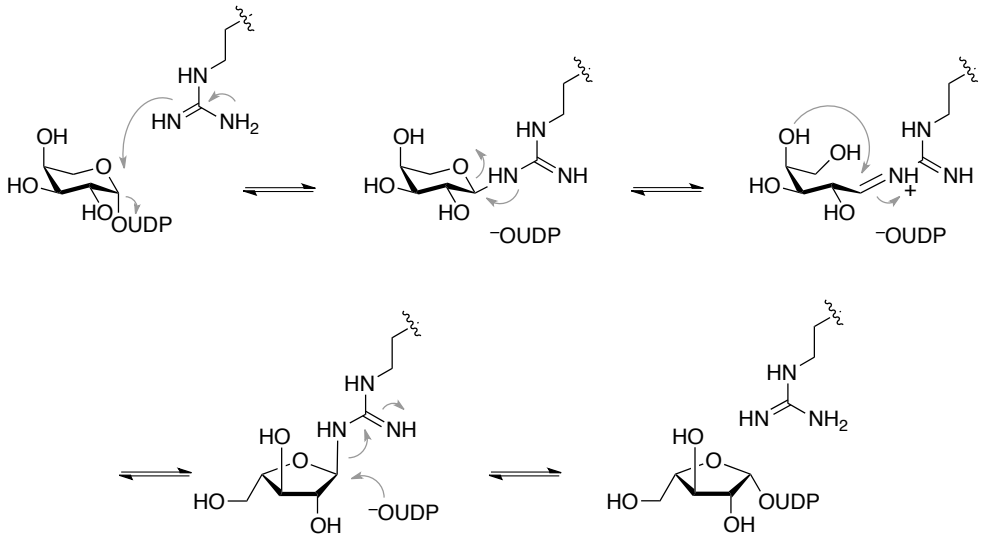


Figure 1-27. A putative mechanism for the plant UAM isomerization of UDP-L-Ara*p* to UDP-L-Ara*f*, by analogy to the microbial UGM mechanism.

1.4.4 L-Ara*f* in other organisms

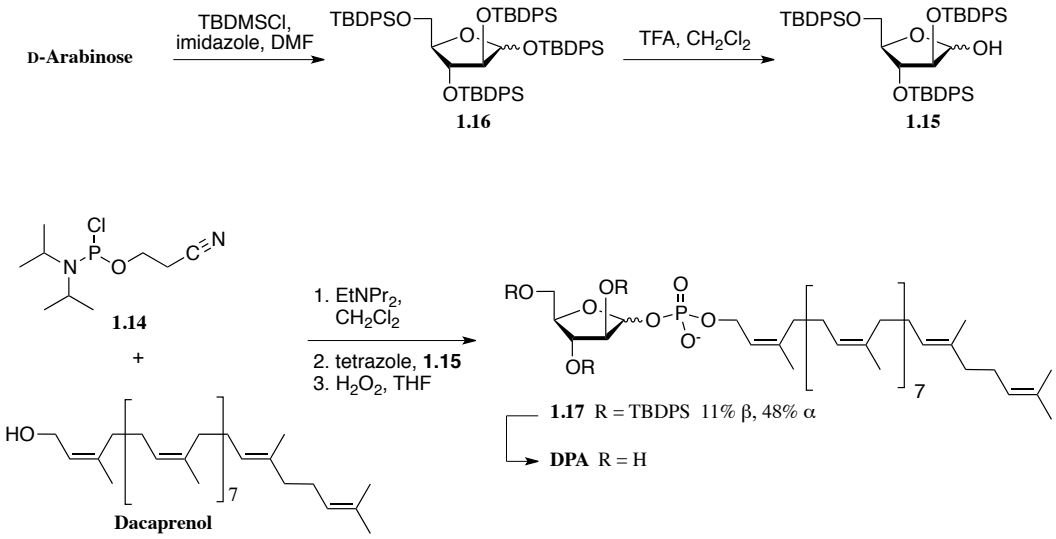
L-Ara*f*, although most widespread in plant species, has also been found in the glycans of the gram-negative bacterium *Campylobacter jejuni*. In this organism, L-Ara*f* residues are found in the CPS repeating units of the HS:15[118] and HS:41[19] serotype strains. The sequence of the CPS gene locus for both serotypes have been reported,[119, 120] however, neither contains a homolog of the plant UAM protein. RGPs, including UAM, appear to be plant specific enzymes; suggesting an alternative pathway is present for the biosynthesis of UDP-L-Ara*f* in bacteria.

1.5 Synthesis of Activated Furanose Donors

One factor that often limits the characterization of biosynthetic enzymes is the ability to access sufficient quantities of the enzyme substrates. Many of these substrates cannot be easily prepared or isolated from biological systems. Thus, chemical synthesis of the substrate, or at least partial synthesis, is required to access sufficient quantities. This is particularly true for the activated donor substrates in furanoside biosynthesis, such as the UDP-D-Gal*f*, DPA, DPR, or UDP-L-Ara*f* species discussed above. This section will discuss the chemical and chemo-enzymatic methods developed to access these donor molecules and derivatives thereof.

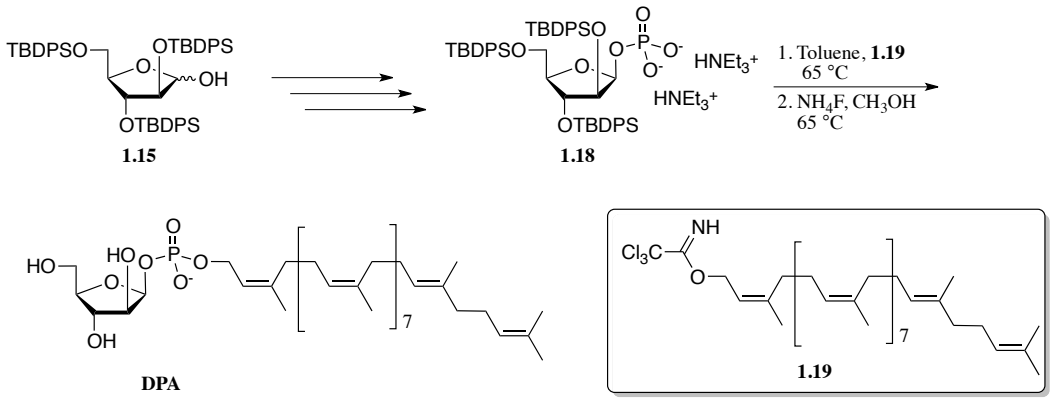
1.5.1 Synthesis of DPR and DPA analogs

Lee *et al.* reported the first chemical synthesis of DPA in 1995,[16] when they prepared ^{14}C-labeled DPA to be used as a substrate to evaluate the activity of mycobacterial AfTs. They subsequently applied the same methodology to prepare a panel of DPA derivatives varying in the length of the polyprenyl lipid chain.[121] This approach used a phosphoramidite coupling strategy 2-cyanoethyl *N*,*N*-diisopropyl-chlorophosphoramidite (**1.14**) coupled first to decaprenol followed by 2,3,5-tri-*O*-t-butyldimethylsilyl-D-Ara*f* (**1.15**, Scheme 1-2). Oxidation and deprotection gave DPA in an 11% yield; the major product (obtained in 48%) was the undesired α-epimer of DPA.[16, 121]



Scheme 1-2. Synthesis of DPA via phosphoramidite coupling.

In 2005, Liav and Brennan reported a second, stereoselective, procedure for the synthesis of polyprenylphosphate derivatives of DPA when they completed the synthesis of farnesylphosphoryl-β-D-Ara*f* (FPA) and the shorter nerylphosphoryl-β-D-Ara*f*,[122] followed a year later with the synthesis of DPA.[123] Their approach involved first the stereoselective synthesis of a protected β-D-Ara*f* 1-phosphate derivative (**1.18**) followed by coupling with a trichloroacetimidate-activated polyprenol (**1.19**, Scheme 1-3). Using this method, they were able to stereoselectively access DPA without the formation of any α-isomer.



Scheme 1-3. Stereoselective synthesis of DPA.

Liav and coworkers later applied their method to the synthesis of farnesylphosphoryl β-D-Rib*f* (FPR),[124] which has subsequently been used to study the activity of the DprE1/DprE2 enzymes,[29] as discussed in section 1.2.2. Joe and Lowary,[125] and Smellie *et al.*[126] have also applied this method to the synthesis of 2′-deoxy-2′-fluoro-DPA and 5′-modified derivatives of FPA (**1.20** and **1.21**, Figure 1-28), respectively, to serve as mechanistic probes of mycobacterial AfTs. However, there have been no subsequent reports discussing the use of these compounds.

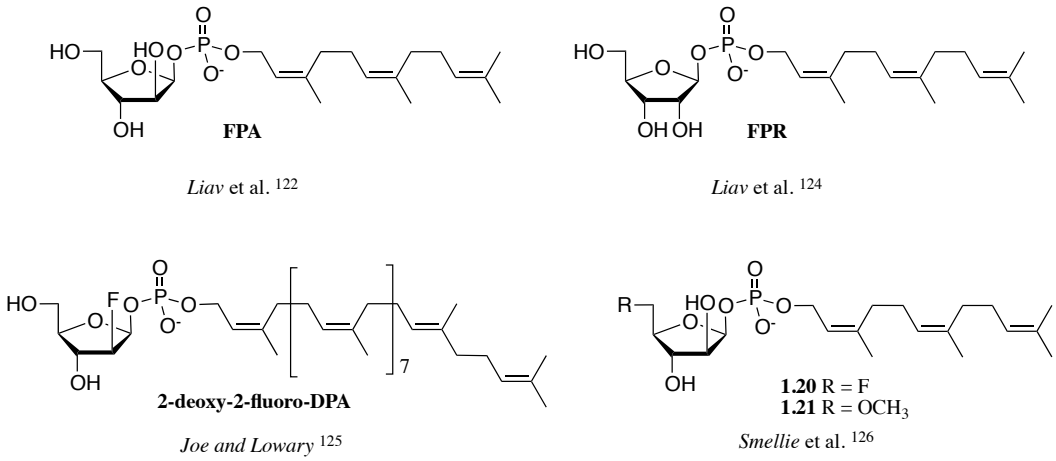


Figure 1-28. Representative examples of synthetic DPA and DPR analogs.

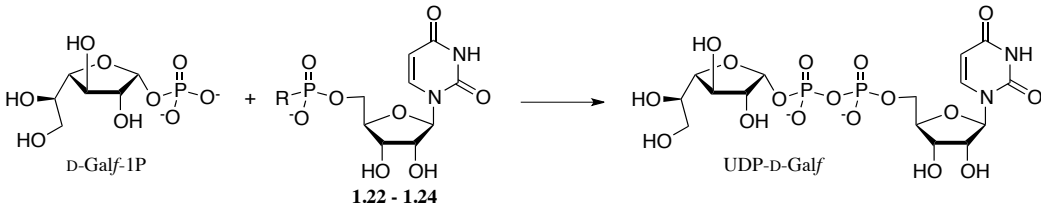
1.5.2 Synthesis of UDP-D-Gal*f*, UDP-L-Ara*f*, and their analogs

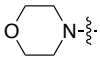
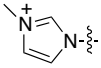
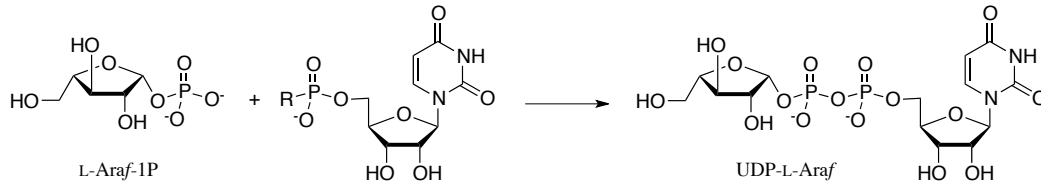
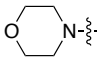
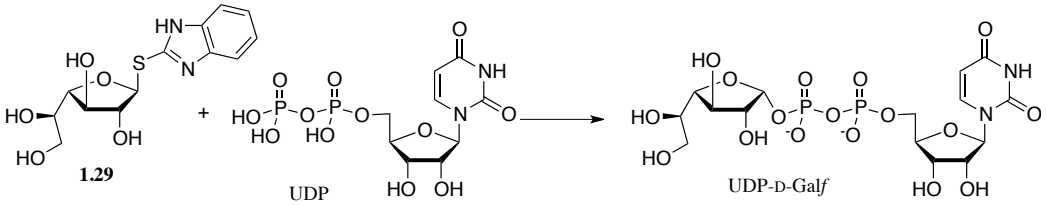
UDP-D-Gal*f* can be prepared enzymatically from UDP-D-Gal*p* using UDP-galactopyranose mutase enzymes. In fact, the first report of a mutase assay, in 1996, isolated this intermediate from reactions with recombinant *E. coli* UGM.[127] The first assays for GlfT2 also used UGM to prepare UDP-D-[^{14}C]-Gal*f* *in situ* from UDP-D-[^{14}C]-Gal*p*.[128] However, the enzyme equilibrium favors the pyranose ring form in a ratio of ~9:1, as mentioned in section 1.3, which limits the utility of this method for preparing larger quantities of UDP-D-Gal*f*. This has lead to an interest in synthetic methods to access this substrate.

Tsvetkov and Nikolaev reported the first chemical synthesis of UDP-D-Gal*f* in 2000, which involved coupling of α-D-Gal*f*-1-phosphate with UMP-imidazolide **1.22** (Table 1-1).[129] After a 19 hour reaction, this method gave the desired UDP-D-Gal*f* in a 23% yield. In the same year, Zhang and Liu also reported a synthesis of UDP-D-Gal*f* in a comparable yield using a different activated UMP derivative, UMP-morpholidate **1.23**, and used the product to quantify the activity of the *E. coli* K12 UGM.[63] Zhang and Liu subsequently applied this same methodology to prepare 2′′- and 3′′-deoxyfluoro analogs of UDP-D-Gal*f* (**1.25**, **1.26**, Figure 1-29),[61] as well as UDP-L-Ara*f* (Table 1-1),[130] which served as mechanistic probes of UGM. A year later, Marlow and Kiessling used yet another activated UMP derivative, UMP *N*-methylimidazolide, to improve the reaction yield to 35% and to shorten the reaction time to 2 hours.[131] This method was subsequently used to prepare deoxyfluoro analogs of UDP-D-Gal*f* **1.27** and **1.28** (Figure 1-29), which were

used as mechanistic probes of the galactofuranosyltransferase GlfT2, but in this case only low yields (8–11%) of the products were obtained.[132]

Table 1-1. Chemical methods for UDP-D-Gal*f* synthesis.



R =	Conditions	Yield	Reference
 **1.22**	DMF, 19 h, RT	23%	Tsvetkov & Nikolaev[129]
 **1.23**	Pyridine, tetrazole, 40 h, RT	20%	Zhang & Liu[63]
 **1.24**	CH_3CN, 0 °C, 2 h	35%	Marlow & Kiessling[131]
			
 **1.23**	Pyridine, tetrazole, 40 h, RT	42%	Zhang & Liu[130]
			
	DMF, 10 min,	32%	Peltier *et al.*[133]

A more recent report involving the direct coupling between the free acid of UDP and a 1-β-thiomidoyl-Gal*f* derivative (**1.29**, Table 1-1) to access UDP-D-Gal*f* was reported by Ferrières and coworkers.[133] The disadvantage of this method, however, is that the product is obtained in only a moderate 32% yield and as a 1:2 mixture of α to β anomers, where only the α-UDP-D-Gal*f* is biologically relevant. This approach was also employed to prepare UDP-D-Gal*f* analogs including 6-deoxy-6-fluoro-UDP-D-Gal*f* (**1.28**, Figure 1-29), which was later used as mechanistic probes of UGM[134] and GlfT1[135] enzymes.

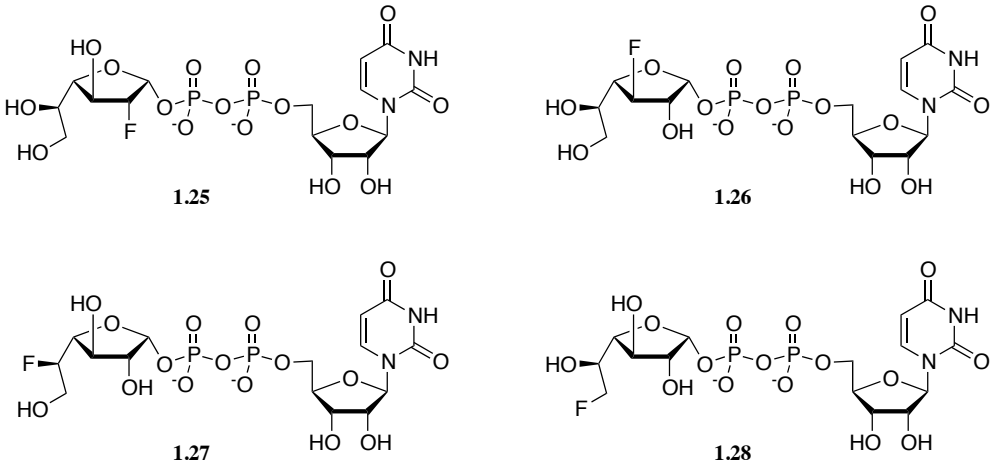


Figure 1-29. Deoxyfluoro analogs of UDP-D-Gal*f*.

In addition to the purely chemical methods to access UDP-D-Gal*f*, a chemo-enzymatic method was developed in 2004 by Errey *et al*.[136] and then later optimized by Rose *et al*.[137] This chemo-enzymatic method allows for the preparation of UDP-D-Gal*f* and UDP-D-Gal*f* derivatives from the corresponding D-Gal*f*-1-P or D-Gal*f*-1-P derivatives, UTP, and uridine-5′-diphosphate-D-glucose (UDP-D-Glc) using three enzymes from the galactose salvage pathway

(Figure 1-30). An equivalent of UMP is transferred from UDP-D-Glc to D-Gal*f*-1-P, catalyzed by galactose-1-phosphate uridylyltransferase (GalPUT), to form UDP-D-Gal*f* and D-glucose-1-phosphate (D-Glc-1-P). This procedure takes advantage of the reduced substrate specificity of the GalPUT enzyme of *E. coli*, which has been shown to tolerate a wide range of sugar-1-phosphate derivatives as substrates, including the furanose sugars D-Gal*f*-1-P and L-Ara*f*-1-P.[136, 138] UTP and the enzyme UDP-glucose pyrophosphorylase (GalU) react with D-Glc-1-P to regenerate the UDP-D-Glc, generating inorganic pyrophosphate (PPi) as a byproduct. To avoid product inhibition of GalU, the PPi is hydrolysed to inorganic phosphate (Pi) by inorganic pyrophosphatase (IPP). This procedure has been used successfully to produce UDP-D-Gal*f* and UDP-L-Ara*f* in multi-milligram scale with reported yields up to 80% and 22%, respectively, from the corresponding furanose-1-phosphates.[138]

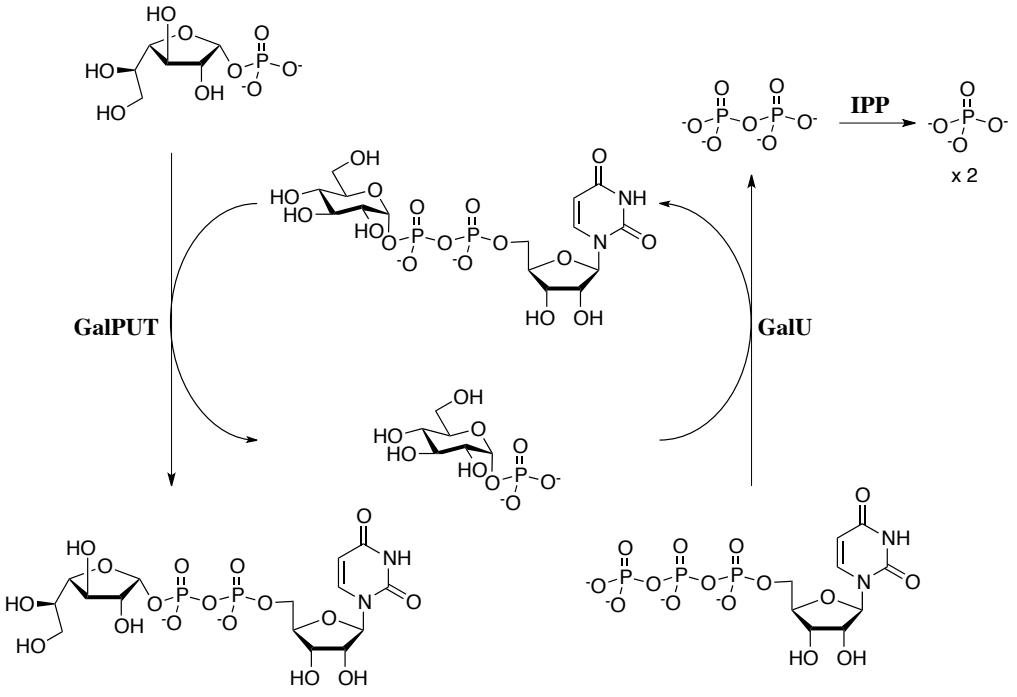


Figure 1-30. Chemo-enzymactic synthesis of UDP-D-Gal*f*.

In summary, a number of different chemical and chemo-enzymatic procedures have been reported for the synthesis of UDP-D-Gal*f* and related UDP-furanose derivatives. Nearly all of the chemical methods involve pyrophosphate coupling between D-Gal*f*-1-P (or derivatives) and an activated UMP derivative.[63, 129, 131] Yields of these reactions vary substantially (8–43%) between different reports, and depending on the product being produced. Of the currently available methods, the chemo-enzymatic approach appears to offer the highest yields of UDP-D-Gal*f*.

1.6 Overview of thesis research

Although there has been much progress in the last two decades to elucidate the biosynthesis of the furanose sugars as discussed above, there are still many questions that remain unanswered. In particular, we still do not have a clear picture of protein structural motifs involved in the recognition of these activated furanose donors. There are also numerous additional furanose sugars found in the glycoconjugates of bacteria and other microorganisms not discussed above. For example, the O-antigen in the LPS of *Pectinatus frisingensis* consists of a homopolysaccharide of 6-deoxy-L-altrofuranose (6d-L-Alt*f*),[139] and 6-deoxy-D-*altro*-heptofuranose (6d-D-*altro*-Hep*f*) and 6-deoxy-D-Gal*f* (D-Fuc*f*) are prevalent in the antigenic polysaccharide of *Eubacterium saburreum* strains.[140-143] D-Fuc*f* is also found in the O-antigen of *E. coli* O52,[87] which also contain a *glf* gene homolog (*fcf2*) implicated in the biosynthesis of TDP-D-Fuc*f*.[86]

As discussed above, a number of *glf* gene homologs have been identified in bacteria from which no D-Gal*f*-containing glycoconjugates have been isolated.

Of particular interest to this thesis are serotype HS:2 and HS:41 strains of *Campylobacter jejuni*, which contain *glf* homologs in the capsular polysaccharide (CPS) gene locus.[144] Neither serotype has been shown to contain D-Gal*f* in its glycoconjugates. Instead, the HS:2 serotype of *C. jejuni* produces a CPS composed of a tetrasaccharide repeating unit containing β-D-Gal*f*NAc (Figure 1-31 A),[145] and the HS:41 serotype produce a CPS composed of two trisaccharide repeating units where the carbohydrates are exclusively in the furanose ring form (Figure 1-31 B), consisting of 6d-D-*altro*-Hep*f,* 6d-L-Alt*f,* D-Fuc*f* and L-Ara*f*.[146] The research described herein describes the characterization of enzymes involved in the biosynthesis of activated precursors to these furanose sugars. Examining the structure and activity of these *glf* homologs provides us an opportunity to explore structural motifs responsible for the specificity of these different enzymes.

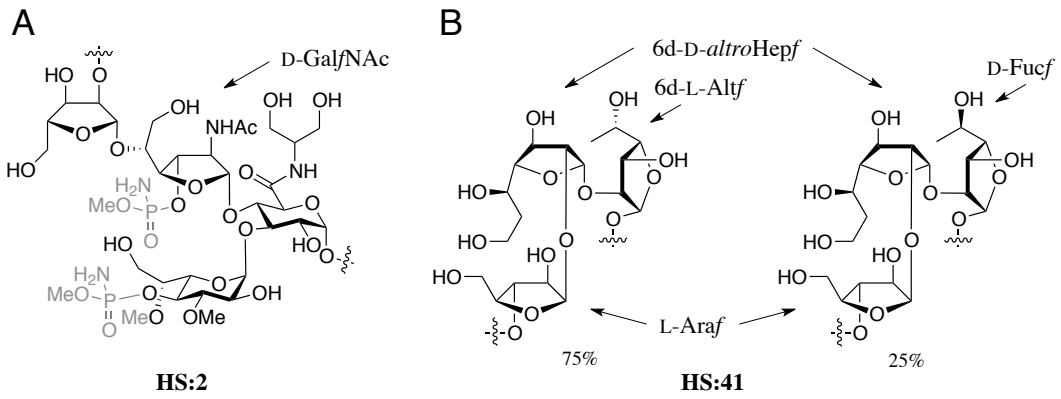


Figure 1-31. Structure of the CPS repeating units from *C. jejuni* serotype HS:2 (A) and HS:41 (B) shown to contain sugars in the furanose ring form.

In Chapter 2, I describe the characterization of a UGM homolog, encoded by the *cj1439* gene of *Campylobacter jejuni* 11168. Using an *in vitro* HPLC assay and complementation studies we demonstrate that this enzyme functions as a UDP-N-acetyl-galactopyranose mutase (UNGM). This UNGM enzyme has a relaxed specificity and can use either UDP-D-Gal or UDP-D-GalNAc as a substrate, compared to other bacterial UGMs, which are specific for UDP-D-Gal. Using site-directed mutagenesis I identify two active site amino acid residues involved in the recognition of the UDP-D-GalNAc substrate.

On the basis of the biochemical characterization of UNGM described in Chapter 2, I sought to further explore the difference in substrate specificity between this enzyme and homologous UGM from *E. coli* and *K. pneumoniae*. To accomplish this, in Chapter 3 I describe the synthesis of a panel of singly modified UDP-D-Gal*f* derivatives and evaluate their activity as substrates or inhibitors of these mutase enzymes. Further techniques, including saturation transfer difference nuclear magnetic resonance (STD-NMR) spectroscopy and protein X-ray crystallography were explored to further map the differences in the binding epitopes for these enzymes.

I expand our studies of pyranose–furanose mutase enzymes specificity to explore the activity of three putative UGM homologs in *C. jejuni* serotype HS:41 involved in the biosynthesis of the L-Ara*f*, D-Fuc*f*, 6d-L-Alt*f* and 6d-D-*altro*-Hep*f* found in the CPS (Figure 1-31). These studies required me to synthesize putative sugar nucleotide substrates of these three enzymes and then test their activity with the recombinant *C. jejuni* UGM homologs (Chapter 4).

In addition to the pyranose–furanose mutase enzymes described above, my research has also focused on exploring the specificity of a galactofuranosyl-transferase (GlfT2) involved in mycobacterial cell wall galactan biosynthesis. In Chapter 5 I examine the importance of donor substrate binding interactions on GlfT2 activity using the panel of UDP-D-Gal*f* derivatives synthesized in Chapter 3. This study has allowed me to examine the role these interactions play in substrate turnover and their influence on the reaction specificity of this enzyme. GlfT2 catalyzes the formation of alternating β-(1→5) and β-(1→6) glycosides, although the exact mechanism of this specificity remains to be established. Recent crystallographic studies have identified putative protein–carbohydrate interactions predicted to influence this alternating β-(1→5) and β-(1→6) activity of the enzyme. In this Chapter, I also explored the influence of these interactions on the alternating activity of GlfT2 by characterizing the structure of products produced by GlfT2 mutant enzymes. The data obtained supports a model of substrate recognition developed from the X-ray crystallographic studies.[100]

1.7 Bibliography

1. Pan, F.; Jackson, M.; Ma, Y. F.; McNeil, M., *J. Bacteriol.* **2001,** *183*, 3991-3998.

2. Schmalhorst, P. S.; Krappmann, S.; Vervecken, W.; Rohde, M.; Muller, M.; Braus, G. H.; Contreras, R.; Braun, A.; Bakker, H.; Routier, F. H., *Eukaryotic Cell* **2008,** *7*, 1268-1277.

3. Kleczka, B.; Lamerz, A. C.; van Zandbergen, G.; Wenzel, A.; Gerardy-Schahn, R.; Wiese, M.; Routier, F. H., *J. Biol. Chem.* **2007,** *282*, 10498-10505.

4. Pedersen, L. L.; Turco, S. J., *Cell. Mol. Life Sci.* **2003,** *60*, 259-266.

5. Köplin, R.; Brisson, J. R.; Whitfield, C., *J. Biol. Chem.* **1997,** *272*, 4121-4128.

6. Whitfield, C.; Richards, J. C.; Perry, M. B.; Clarke, B. R.; Maclean, L. L., *J. Bacteriol.* **1991,** *173*, 1420-1431.

7. Brennan, P. J.; Nikaido, H., *Annu. Rev. Biochem.* **1995,** *64*, 29-63.

8. Bhamidi, S.; Scherman, M. S.; Rithner, C. D.; Prenni, J. E.; Chatterjee, D.; Khoo, K. H.; McNeil, M. R., *J. Biol. Chem.* **2008,** *283*, 12992-13000.

9. Brennan, P. J., *Tuberculosis* **2003,** *83*, 91-97.

10. Bishop, C. T.; Blank, F., *Can. J. Microbiol.* **1958,** *4*, 35-42.

11. Previato, J. O.; Mendoncapreviato, L.; Lewanczuk, R. Z.; Travassos, L. R.; Gorin, P. A. J., *Exp. Parasitol.* **1982,** *53*, 170-178.

12. McConville, M. J.; Thomasoates, J. E.; Ferguson, M. A. J.; Homans, S. W., *J. Biol. Chem.* **1990,** *265*, 19611-19623.

13. McConville, M. J.; Ferguson, M. A. J., *Biochem. J.* **1993,** *294*, 305-324.

14. Wolucka, B. A., *FEBS J.* **2008,** *275*, 2691-2711.

15. Wolucka, B. A.; McNeil, M. R.; Dehoffmann, E.; Chojnacki, T.; Brennan, P. J., *J. Biol. Chem.* **1994,** *269*, 23328-23335.

16. Lee, R. E.; Mikusova, K.; Brennan, P. J.; Besra, G. S., *J. Am. Chem. Soc.* **1995,** *117*, 11829-11832.

17. Kus, J. V.; Kelly, J.; Tessier, L.; Harvey, H.; Cvitkovitch, D. G.; Burrows, L. L., *J. Bacteriol.* **2008,** *190*, 7464-7478.

18. Mergaert, P.; Dhaeze, W.; FernandezLopez, M.; Geelen, D.; Goethals, K.; ClaudeProme, J.; VanMontagu, M.; Holsters, M., *Mol. Microbiol.* **1996,** *21*, 409-419.

19. Harvey, H.; Kus, J. V.; Tessier, L.; Kelly, J.; Burrows, L. L., *J. Biol. Chem.* **2011,** *286*, 28128-28137.

20. Pieper, S.; Unterieser, I.; Mann, F.; Mischnick, P., *Carbohydr. Res.* **2012,** *352*, 166-176.

21. Schneider, P.; Nikolaev, A.; Ferguson, M. A. J., *Biochem. J.* **1995,** *311*, 307-315.

22. Schneider, P.; McConville, M. J.; Ferguson, M. A. J., *J. Biol. Chem.* **1994,** *269*, 18332-18337.

23. Singh, S.; Hogan, S. E., *Microbios* **1994,** *77*, 217-222.

24. Klutts, J. S.; Hatanaka, K.; Pan, Y. T.; Elbein, A. D., *Arch. Biochem. Biophys.* **2002,** *398*, 229-239.

25. Scherman, M. S.; KalbeBournonville, L.; Bush, D.; Deng, L. Y.; McNeil, M., *J. Biol. Chem.* **1996,** *271*, 29652-29658.

26. Huang, H. R.; Scherman, M. S.; D'Haeze, W.; Vereecke, D.; Holsters, M.; Crick, D. C.; McNeil, M. R., *J. Biol. Chem.* **2005,** *280*, 24539-24543.

27. Mikušová, K.; Huang, H. R.; Yagi, T.; Holsters, M.; Vereecke, D.; D'Haeze, W.; Scherman, M. S.; Brennan, P. J.; McNeil, M. R.; Crick, D. C., *J. Bacteriol.* **2005,** *187*, 8020-8025.

28. Sassetti, C. M.; Boyd, D. H.; Rubin, E. J., *Mol. Microbiol.* **2003,** *48*, 77-84.

29. Trefzer, C.; S'Kovierova, H.; Buroni, S.; Bobovska, A.; Nenci, S.; Molteni, E.; Pojer, F.; Pasca, M. R.; Makarov, V.; Cole, S. T.; Riccardi, G.; Mikusova, K.; Johnsson, K., *J. Am. Chem. Soc.* **2012,** *134*, 912-915.

30. Makarov, V.; Manina, G.; Mikusova, K.; Mollmann, U.; Ryabova, O.; Saint-Joanis, B.; Dhar, N.; Pasca, M. R.; Buroni, S.; Lucarelli, A. P.; Milano, A.; De Rossi, E.; Belanova, M.; Bobovska, A.; Dianiskova, P.; Kordulakova, J.; Sala, C.; Fullam, E.; Schneider, P.; McKinney, J. D.; Brodin, P.; Christophe, T.; Waddell, S.; Butcher, P.; Albrethsen, J.; Rosenkrands, I.; Brosch, R.; Nandi, V.; Bharath, S.; Gaonkar, S.; Shandil, R. K.; Balasubramanian, V.; Balganesh, T.; Tyagi, S.; Grosset, J.; Riccardi, G.; Cole, S. T., *Science* **2009,** *324*, 801-804.

31. Buroni, S.; Pasca, M. R.; Ribeiro, A.; Degiacomi, G.; Molteni, E.; Riccardi, G., *Appl. Microbiol. Biotechnol.* **2012,** *94*, 907-916.

32. Trefzer, C.; Rengifo-Gonzalez, M.; Hinner, M. J.; Schneider, P.; Makarov, V.; Cole, S. T.; Johnsson, K., *J. Am. Chem. Soc.* **2010,** *132*, 13663-13665.

33. Batt, S. M.; Jabeen, T.; Bhowruth, V.; Quill, L.; Lund, P. A.; Eggeling, L.; Alderwick, L. J.; Fütterer, K.; Besra, G. S., *Proc. Natl. Acad. Sci. USA* **2012,** *109*, 11354-11359.

34. Tam, P. H.; Lowary, T. L., *Curr. Opin. Chem. Biol.* **2009,** *13*, 618-625.

35. Berg, S.; Kaur, D.; Jackson, M.; Brennan, P. J., *Glycobiology* **2007,** *17*, 35R-56R.

36. Alderwick, L. J.; Birch, H. L.; Mishra, A.; Eggeling, L.; Besra, G. S., *Biochem. Soc. Trans.* **2007,** *35*, 1325-1328.

37. Alderwick, L. J.; Seidel, M.; Sahm, H.; Besra, G. S.; Eggeling, L., *J. Biol. Chem.* **2006,** *281*, 15653-15661.

38. Seidel, M.; Alderwick, L. J.; Birch, H. L.; Sahm, H.; Eggeling, L.; Besra, G. S., *J. Biol. Chem.* **2007,** *282*, 14729-14740.

39. Zhang, J.; Angala, S. K.; Pramanik, P. K.; Li, K.; Crick, D. C.; Liav, A.; Jozwiak, A.; Swiezewska, E.; Jackson, M.; Chatterjee, D., *ACS Chem. Biol.* **2011,** *6*, 819-828.

40. Escuyer, V. E.; Lety, M. A.; Torrelles, J. B.; Khoo, K. H.; Tang, J. B.; Rithner, C. D.; Frehel, C.; McNeil, M. R.; Brennan, P. J.; Chatterjee, D., *J. Biol. Chem.* **2001,** *276*, 48854-48862.

41. Zhang, N.; Torrelles, J. B.; McNeil, M. R.; Escuyer, V. E.; Khoo, K. H.; Brennan, P. J.; Chatterjee, D., *Mol. Microbiol.* **2003,** *50*, 69-76.

42. Shi, L. B.; Berg, S.; Lee, A.; Spencer, J. S.; Zhang, J.; Vissa, V.; McNeil, M. R.; Khoo, K. H.; Chatterjee, D., *J. Biol. Chem.* **2006,** *281*, 19512-19526.

43. Alderwick, L. J.; Lloyd, G. S.; Ghadbane, H.; May, J. W.; Bhatt, A.; Eggeling, L.; Fütterer, K.; Besra, G. S., *PLoS Pathog.* **2011,** *7*.

44. Pan, F.; Jackson, M.; Ma, Y.; McNeil, M., *J. Bacteriol.* **2001,** *183*, 3991-3998.

45. Stevenson, G.; Neal, B.; Liu, D.; Hobbs, M.; Packer, N. H.; Batley, M.; Redmond, J. W.; Lindquist, L.; Reeves, P., *J. Bacteriol.* **1994,** *176*, 4144-4156.

46. Liu, D.; Reeves, P. R., *Microbiology-Uk* **1994,** *140*, 49-57.

47. Linnerborg, M.; Weintraub, A.; Widmalm, G., *Eur. J. Biochem.* **1999,** *266*, 460-466.

48. Linnerborg, M.; Wollin, R.; Widmalm, G., *Eur. J. Biochem.* **1997,** *246*, 565-573.

49. Perry, M. B.; MacLean, L. L., *Carbohydr. Res.* **2004,** *339*, 1399-1402.

50. Tefsen, B.; Ram, A. F. J.; van Die, I.; Routier, F. H., *Glycobiology* **2012,** *22*, 456-469.

51. Richards, M. R.; Lowary, T. L., *ChemBioChem* **2009,** *10*, 1920-1938.

52. Frey, P. A., *FASEB J.* **1996,** *10*, 461-470.

53. Nassau, P. M.; Martin, S. L.; Brown, R. E.; Weston, A.; Monsey, D.; McNeil, M. R.; Duncan, K., *J. Bacteriol.* **1996,** *178*, 1047-1052.

54. Weston, A.; Stern, R. J.; Lee, R. E.; Nassau, P. M.; Monsey, D.; Martin, S. L.; Scherman, M. S.; Besra, G. S.; Duncan, K.; McNeil, M. R., *Tuber. Lung Dis.* **1998,** *78*, 123-131.

55. Beverley, S. M.; Owens, K. L.; Showalter, M.; Griffith, C. L.; Doering, T. L.; Jones, V. C.; McNeil, M. R., *Eukaryotic Cell* **2005,** *4*, 1147-1154.

56. Bakker, H.; Kleczka, B.; Gerardy-Schahn, R.; Routier, F. H., *Biol. Chem.* **2005,** *386*, 657-661.

57. Barlow, J. N.; Girvin, M. E.; Blanchard, J. S., *J. Am. Chem. Soc.* **1999,** *121*, 6968-6969.

58. Caravano, A.; Sinay, P.; Vincent, S. P., *Bioorg. Med. Chem. Lett.* **2006,** *16*, 1123-1125.

59. Barlow, J. N.; Blanchard, J. S., *Carbohydr. Res.* **2000,** *328*, 473-480.

60. Burton, A.; Wyatt, P.; Boons, G. J., *Journal of the Chemical Society-Perkin Transactions 1* **1997**, 2375-2382.

61. Zhang, Q. B.; Liu, H. W., *J. Am. Chem. Soc.* **2001,** *123*, 6756-6766.

62. Errey, J. C.; Mann, M. C.; Fairhurst, S. A.; Hill, L.; McNeil, M. R.; Naismith, J. H.; Percy, J. M.; Whitfield, C.; Field, R. A., *Org. Biomol. Chem.* **2009,** *7*, 1009-1016.

63. Zhang, Q. B.; Liu, H. W., *J. Am. Chem. Soc.* **2000,** *122*, 9065-9070.

64. Sanders, D. A. R.; Staines, A. G.; McMahon, S. A.; McNeil, M. R.; Whitfield, C.; Naismith, J. H., *Nat. Struct. Biol.* **2001,** *8*, 858-863.

65. Fullerton, S. W. B.; Daff, S.; Sanders, D. A. R.; Ingledew, W. J.; Whitfield, C.; Chapman, S. K.; Naismith, J. H., *Biochemistry* **2003,** *42*, 2104-2109.

66. Huang, Z. H.; Zhang, Q. B.; Liu, H. W., *Bioorg. Chem.* **2003,** *31*, 494-502.

67. Soltero-Higgin, M.; Carlson, E. E.; Gruber, T. D.; Kiessling, L. L., *Nat. Struct. Mol. Biol.* **2004,** *11*, 539-543.

68. Gruber, T. D.; Westler, W. M.; Kiessling, L. L.; Forest, K. T., *Biochemistry* **2009,** *48*, 9171-9173.

69. Sun, H. G.; Ruszczycky, M. W.; Chang, W. C.; Thibodeaux, C. J.; Liu, H. W., *J. Biol. Chem.* **2012,** *287*, 4602-4608.

70. Oppenheimer, M.; Valenciano, A. L.; Kizjakina, K.; Qi, J.; Sobrado, P., *PLoS One* **2012,** *7*.

71. Beis, K.; Srikannathasan, V.; Liu, H.; Fullerton, S. W. B.; Bamford, V. A.; Sanders, D. A. R.; Whitfield, C.; McNeil, M. R.; Naismith, J. H., *J. Mol. Biol.* **2005,** *348*, 971-982.

72. Chad, J. M.; Sarathy, K. P.; Gruber, T. D.; Addala, E.; Kiessling, L. L.; Sanders, D. A. R., *Biochemistry* **2007,** *46*, 6723-6732.

73. Yuan, Y.; Bleile, D. W.; Wen, X.; Sanders, D. A. R.; Itoh, K.; Liu, H. W.; Pinto, B. M., *J. Am. Chem. Soc.* **2008,** *130*, 3157-3168.

74. Yuan, Y.; Wen, X.; Sanders, D. A. R.; Pinto, B. M., *Biochemistry* **2005,** *44*, 14080-14089.

75. Yao, X. H.; Bleile, D. W.; Yuan, Y.; Chao, J.; Sarathy, K. P.; Sanders, D. A. R.; Pinto, B. M.; O'Neill, M. A., *Proteins* **2009,** *74*, 972-979.

76. Gruber, T. D.; Borrok, M. J.; Westler, W. M.; Forest, K. T.; Kiessling, L. L., *J. Mol. Biol.* **2009,** *391*, 327-340.

77. Partha, S. K.; Bonderoff, S. A.; van Straaten, K. E.; Sanders, D. A. R., *Acta Cryst. F* **2009,** *65*, 843-845.

78. Partha, S. K.; van Straaten, K. E.; Sanders, D. A. R., *J. Mol. Biol.* **2009,** *394*, 864-877.

79. Partha, S. K.; Sadeghi-Khomami, A.; Slowski, K.; Kotake, T.; Thomas, N. R.; Jakeman, D. L.; Sanders, D. A. R., *J. Mol. Biol.* **2010,** *403*, 578-590.

80. Oppenheimer, M.; Poulin, M. B.; Lowary, T. L.; Helm, R. F.; Sobrado, P., *Arch. Biochem. Biophys.* **2010,** *502*, 31-38.

81. Oppenheimer, M.; Valenciano, A. L.; Sobrado, P., *Biochem. Biophys. Res. Commun.* **2011,** *407*, 552-556.

82. Dhatwalia, R.; Singh, H.; Oppenheimer, M.; Karr, D. B.; Nix, J. C.; Sobrado, P.; Tanner, J. J., *J. Biol. Chem.* **2012,** *287*, 9041-9051.

83. van Straaten, K. E.; Routier, F. H.; Sanders, D. A. R., *Acta Cryst. F* **2012,** *68*, 455-459.

84. van Straaten, K. E.; Routier, F. H.; Sanders, D. A. R., *J. Biol. Chem.* **2012,** *287*, 10780-10790.

85. Dhatwalia, R.; Singh, H.; Oppenheimer, M.; Sobrado, P.; Tanner, J. J., *Biochemistry* **2012,** *51*, 4968-4979.

86. Wang, Q.; Ding, P.; Perepelov, A. V.; Xu, Y. L.; Wang, Y.; Knirel, Y. A.; Wang, L.; Feng, L., *Mol. Microbiol.* **2008,** *70*, 1358-1367.

87. Feng, L.; Senchenkova, S. N.; Yang, J. H.; Shashkov, A. S.; Tao, J.; Guo, H. J.; Cheng, J. S.; Ren, Y.; Knirel, Y. A.; Reeves, P. R.; Wang, L., *J. Bacteriol.* **2004,** *186*, 4510-4519.

88. Cunneen, M. M.; De Castro, C.; Kenyon, J.; Parrilli, M.; Reeves, P. R.; Molinaro, A.; Holst, O.; Skurnik, M., *Carbohydr. Res.* **2009,** *344*, 1533-1540.

89. Wing, C.; Errey, J. C.; Mukhopadhyay, B.; Blanchard, J. S.; Field, R. A., *Org. Biomol. Chem.* **2006,** *4*, 3945-3950.

90. Guan, S.; Clarke, A. J.; Whitfield, C., *J. Bacteriol.* **2001,** *183*, 3318-3327.

91. Mikušová, K.; Beláňová, M.; Kordulakova, J.; Honda, K.; McNeil, M. R.; Mahapatra, S.; Crick, D. C.; Brennan, P. J., *J. Bacteriol.* **2006,** *188*, 6592-6598.

92. Beláňová, M.; Dianišková, P.; Brennan, P. J.; Completo, G. C.; Rose, N. L.; Lowary, T. L.; Mikusova, K., *J. Bacteriol.* **2008,** *190*, 1141-1145.

93. Mikušová, K.; Yagi, T.; Stern, R.; McNeil, M. R.; Besra, G. S.; Crick, D. C.; Brennan, P. J., *J. Biol. Chem.* **2000,** *275*, 33890-33897.

94. Rose, N. L.; Completo, G. C.; Lin, S. J.; McNeil, M.; Palcic, M. M.; Lowary, T. L., *J. Am. Chem. Soc.* **2006,** *128*, 6721-6729.

95. Alderwick, L. J.; Dover, L. G.; Veerapen, N.; Gurcha, S. S.; Kremer, L.; Roper, D. L.; Pathak, A. K.; Reynolds, R. C.; Besra, G. S., *Protein Expression Purif.* **2008,** *58*, 332-341.

96. Szczepina, M. G.; Zheng, R. B.; Completo, G. C.; Lowary, T. L.; Pinto, B. M., *ChemBioChem* **2009,** *10*, 2052-2059.

97. May, J. F.; Levengood, M. R.; Splain, R. A.; Brown, C. D.; Kiessling, L. L., *Biochemistry* **2012,** *51*, 1148-1159.

98. Levengood, M. R.; Splain, R. A.; Kiessling, L. L., *J. Am. Chem. Soc.* **2011,** *133*, 12758-12766.

99. May, J. F.; Splain, R. A.; Brotschi, C.; Kiessling, L. L., *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106*, 11851-11856.

100. Wheatley, R. W.; Zheng, R. B.; Richards, M. R.; Lowary, T. L.; Ng, K. K. S., *J. Biol. Chem.* **2012**.

101. Konishi, T.; Ishii, T., *Trends in Glycoscience and Glycotechnology* **2012,** *24*, 13-23.

102. McNeil, M.; Darvill, A. G.; Fry, S. C.; Albersheim, P., *Annu. Rev. Biochem.* **1984,** *53*, 625-663.

103. Sun, R. C.; Sun, X. F.; Tomkinson, I., Hemicelluloses and their derivatives. In *Hemicelluloses: Science and Technology*, Gatenholm, P. T. M., Ed. 2004; Vol. 864, pp 2-22.

104. Wilkie, K. C. B., *Adv. Carbohydr. Chem. Biochem.* **1979,** *36*, 215-264.

105. Konishi, T.; Takeda, T.; Miyazaki, Y.; Ohnishi-Kameyama, M.; Hayashi, T.; O'Neill, M. A.; Ishii, T., *Glycobiology* **2007,** *17*, 345-354.

106. Konishi, T.; Ohnishi-Kameyama, M.; Funane, K.; Miyazaki, Y.; Konishi, T.; Ishii, T., *Carbohydr. Res.* **2010,** *345*, 787-791.

107. Bollig, K.; Lamshoeft, M.; Schweirner, K.; Marner, F. J.; Budzikiewicz, H.; Waffenschmidt, S., *Carbohydr. Res.* **2007,** *342*, 2557-2566.

108. Aspinall, G. O.; Matheson, N. K.; Cottrell, I. W., *Can. J. Biochem.* **1972,** *50*, 574-580.

109. Bar-Peled, M.; O'Neill, M. A., Plant Nucleotide Sugar Formation, Interconversion, and Salvage by Sugar Recycling. In *Annual Review of Plant Biology, Vol 62*, Merchant, S. S.; Briggs, W. R.; Ort, D., Eds. Vol. 62, pp 127-155.

110. Bolwell, G. P.; Northcote, D. H., *Planta* **1981,** *152*, 225-233.

111. Fry, S. C.; Northcote, D. H., *Plant Physiol.* **1983,** *73*, 1055-1061.

112. Nunan, K. J.; Scheller, H. V., *Plant Physiol.* **2003,** *132*, 331-342.

113. Ishii, T.; Konishi, T.; Ito, Y.; Ono, H.; Ohnishi-Kameyama, M.; Maeda, I., *Phytochemistry* **2005,** *66*, 2418-2425.

114. Konishi, T.; Ono, H.; Ohnishi-Kameyama, M.; Kaneko, S.; Ishii, T., *Plant Physiol.* **2006,** *141*, 1098-1105.

115. Dhugga, K. S.; Tiwari, S. C.; Ray, P. M., *Proc. Natl. Acad. Sci. U. S. A.* **1997,** *94*, 7679-7684.

116. Delgado, I. J.; Wang, Z. H.; de Rocher, A.; Keegstra, K.; Raikhel, N. V., *Plant Physiol.* **1998,** *116*, 1339-1349.

117. Saxena, I. M.; Brown, R. M., *Trends Plant Sci.* **1999,** *4*, 6-7.

118. Guerry, P.; Monteiro, M. A. U.S. Patent 13/117,215, 2011.

119. Karlyshev, A. V.; Champion, O. L.; Churcher, C.; Brisson, J. R.; Jarrell, H. C.; Gilbert, M.; Brochu, D.; St Michael, F.; Li, J. J.; Wakarchuk, W. W.; Goodhead, I.; Sanders, M.; Stevens, K.; White, B.; Parkhill, J.; Wren, B. W.; Szymanski, C. M., *Mol. Microbiol.* **2005,** *55*, 90-103.

120. Poly, F.; Serichatalergs, O.; Schulman, M.; Ju, J.; Cates, C. N.; Kanipes, M.; Mason, C.; Guerry, P., *J. Clin. Microbiol.* **2011,** *49*, 1750-1757.

121. Lee, R. E.; Brennan, P. J.; Besra, G. S., *Bioorg. Med. Chem. Lett.* **1998,** *8*, 951-954.

122. Liav, A.; Brennan, P. J., *Tetrahedron Lett.* **2005,** *46*, 2937-2939.

123. Liav, A.; Huang, H. R.; Ciepichal, E.; Brennan, P. J.; McNeil, M. R., *Tetrahedron Lett.* **2006,** *47*, 545-547.

124. Liav, A.; Swiezewska, E.; Ciepichal, E.; Brennan, P. J., *Tetrahedron Lett.* **2006,** *47*, 8781-8783.

125. Joe, M.; Lowary, T. L., *Carbohydr. Res.* **2006,** *341*, 2723-2730.

126. Smellie, I. A.; Bhakta, S.; Sim, E.; Fairbanks, A. J., *Org. Biomol. Chem.* **2007,** *5*, 2257-2266.

127. Lee, R.; Monsey, D.; Weston, A.; Duncan, J. K.; Rithner, C.; McNeil, M., *Anal. Biochem.* **1996,** *242*, 1-7.

128. Pathak, A. K.; Besra, G. S.; Crick, D.; Maddry, J. A.; Morehouse, C. B.; Suling, W. J.; Reynolds, R. C., *Bioorg. Med. Chem.* **1999,** *7*, 2407-2413.

129. Tsvetkov, Y. E.; Nikolaev, A. V., *J. Chem. Soc., Perkin Trans. 1* **2000**, 889-891.

130. Zhang, Q. B.; Liu, H. W., *Bioorg. Med. Chem. Lett.* **2001,** *11*, 145-149.

131. Marlow, A. L.; Kiessling, L. L., *Org. Lett.* **2001,** *3*, 2517-2519.

132. Brown, C. D.; Rusek, M. S.; Kiessling, L. L., *J. Am. Chem. Soc.* **2012,** *134*, 6552-6555.

133. Peltier, P.; Daniellou, R.; Nugier-Chauvin, C.; Ferrières, V., *Org. Lett.* **2007,** *9*, 5227-5230.

134. Eppe, G.; Peltier, P.; Daniellou, R.; Nugier-Chauvin, C.; Ferrières, V.; Vincent, S. P., *Bioorg. Med. Chem. Lett.* **2009,** *19*, 814-816.

135. Peltier, P.; Beláňová, M.; Dianišková, Petronela; Zhou, R.; Zheng, R. B.; Pearcey, J. A.; Joe, M.; Brennan, P. J.; Nugier-Chauvin, C.; Ferrières, V.; Lowary, T. L.; Daniellou, R.; Mikušová, K., *Chem. Biol.* **2010,** *17*, 1356-1366.

136. Errey, J. C.; Mukhopadhyay, B.; Kartha, K. P. R.; Field, R. A., *Chem. Commun.* **2004**, 2706-2707.

137. Rose, N. L.; Zheng, R. B.; Pearcey, J.; Zhou, R.; Completo, G. C.; Lowary, T. L., *Carbohydr. Res.* **2008,** *343*, 2130-2139.

138. Peltier, P.; Guegan, J. P.; Daniellou, R.; Nugier-Chauvin, C.; Ferrières, V., *Eur. J. Org. Chem.* **2008**, 5988-5994.

139. Senchenkova, S. N.; Shashkov, A. S.; Moran, A. P.; Helander, I. M.; Knirel, Y. A., *Eur. J. Biochem.* **1995,** *232*, 552-557.

140. Kondo, W.; Nakazawa, F.; Sato, M.; Ito, T., *Carbohydr. Res.* **1981,** *97*, 279-283.

141. Sato, N.; Nakazawa, F.; Sato, M.; Hoshino, E.; Ito, T., *Carbohydr. Res.* **1993,** *245*, 105-111.

142. Hoffman, J.; Lindberg, B.; Skaug, N.; Hofstad, T., *Carbohydr. Res.* **1980,** *84*, 181-183.

143. Kondo, W.; Sato, N.; To, F., *J. Dent. Res.* **1980,** *59*, 896-896.

144. Karlyshev, A. V.; Champion, O. L.; Churcher, C.; Brisson, J.-R.; Jarrell, H. C.; Gilbert, M.; Brochu, D.; St Michael, F.; Li, J.; Wakarchuk, W. W.; Goodhead, I.; Sanders, M.; Stevens, K.; White, B.; Parkhill, J.; Wren, B. W.; Szymanski, C. M., *Mol. Microbiol.* **2005,** *55*, 90-103.

145. St Michael, F.; Szymanski, C. M.; Li, J.; Chan, K. H.; Khieu, N. H.; Larocque, S.; Wakarchuk, W. W.; Brisson, J.-R.; Monteiro, M. A., *Eur. J. Biochem.* **2002,** *269*, 5119-5136.

146. Hanniffy, O. M.; Shashkov, A. S.; Moran, A. P.; Prendergast, M. M.; Senchenkova, S. N.; Knirel, Y. A.; Savage, A. V., *Carbohydr. Res.* **1999,** *319*, 124-132.

Chapter 2

Characterization of a Bifunctional Pyranose–furanose Mutase from *Campylobacter jejuni* 11168

A version of this chapter has been published:
Poulin, M. B.; Nothaft, H.; Hug, I.; Feldman, M. F.; Szymanski, C. M.; Lowary, T. L. *J. Biol. Chem.* **2010**, *285*, 493-501.

Dr. Harald Nothaft, a research associate in the group of our collaborator Professor Christine Szymanski, performed the cloning of the *Campylobacter jejuni* and *Escherichia coli glf* genes, and the complementation experiments of the *C. jejuni* 11168 Δ*glf* mutant strain. Dr. Isabelle Hug, a former postdoctoral fellow in the group of our collaborator Professor Mario Feldman, performed the complementation experiments of the *E. coli* Δ*glf* mutant strain.

2.1 Introduction

As discussed in the previous chapter, hexose sugars exist predominantly in the thermodynamically favored pyranose ring form; however, hexose sugars in the furanose ring form are found in many bacteria, fungi, and parasites.[1, 2] For example, D-galactofuranose (Gal*f*) is a component in many microbial cell surface oligosaccharides,[1, 2] and is a major structural component of the mycobacterial cell wall.[3] In many pathogenic microorganisms these Gal*f* residues are essential for cell viability or play a crucial role in cell physiology.[4, 5] For this reason, and because hexofuranose sugars are absent in mammalian cell glycocongugates,[6] there has been a surge of interest in studying and identifying inhibitors of Gal*f* biosynthesis.[7]

The sugar nucleotide UDP-Gal*f* (**2**) is the precursor of Gal*f*, and is incorporated into growing oligosaccharides via galactofuranosyltransferase-mediated reactions.[8] First identified in *Escherichia coli*,[9] the enzyme UDP-D-galactopyranose mutase (UGM) is responsible for the biosynthesis of UDP-Gal*f* via the ring contraction of UDP-galactopyranose (UDP-Gal*p* **1**, Figure 2-1). The *E. coli* UGM (ecUGM) is encoded by the *glf* gene, for which homologues have since been identified and characterized in *Klebsiella pneumoniae* (kpUGM),[10] mycobacterial species (mtUGM),[11] *Deinococcus radiodurans* (drUGM),[12] and in various eukaryotic pathogens.[13, 14] Since the advent of rapid genome sequencing, a number of putative UGMs have been identified throughout the microbial species; however, few of the gene products have been confirmed by biochemical analysis.

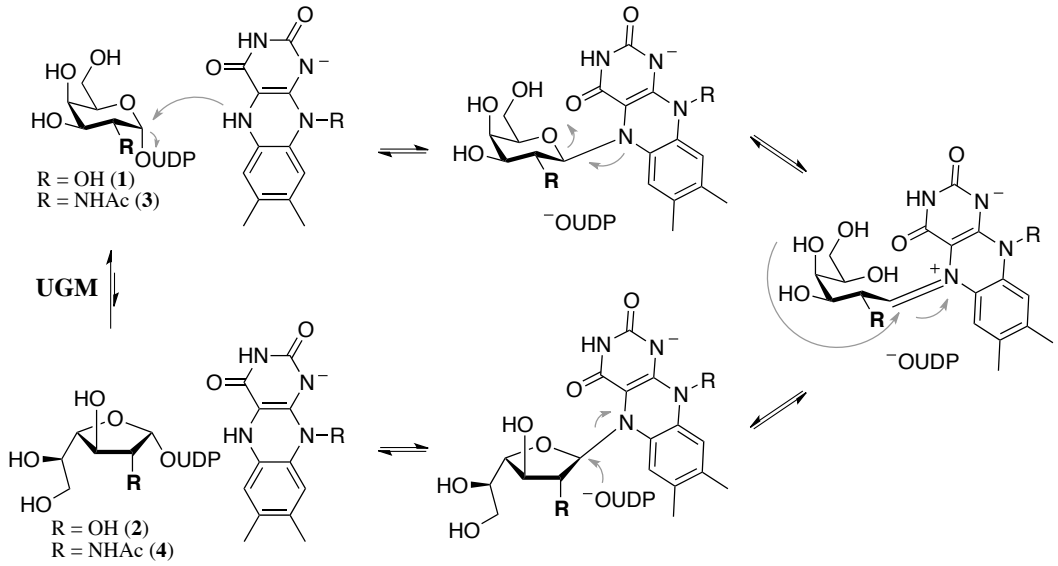


Figure 2-1. Reaction mechanism for the UGM catalyzed ring contraction of UDP-Gal*p*.

UGMs are flavoproteins, and catalyze the reversible ring contraction of UDP-Gal*p* to UDP-Gal*f* via a unique mechanism (Figure 2-1).[15] The non-covalently bound flavin adenine dinucleotide (FAD) co-factor is directly involved in catalysis, and must be in the reduced form for the enzyme to be active.[16] Because of the interest in UGM as a drug target[7] significant work has been done to study its mechanism. It was shown that the reduced $FADH^-$ acts as a nucleophile and directly displaces the anomeric UDP to form a covalent intermediate.[17] Formation of an iminium ion breaks the O5–C1 bond of the galactose moiety leading to a covalently bound acyclic intermediate, which has been indirectly detected by reduction with sodium cyanoborohydride and isolation of the reduced aduct.[18] This iminium ion species can then cyclize to the furanose ring form.

Although the enzyme mechanism is generally understood, there are still many unanswered questions about enzyme–substrate interactions. For example, UGM proteins contain a mobile loop region, which adopts either an open or closed form in the crystal structures that have been determined to date[15, 19] with the closed structure being the catalytically active form. This loop has been shown to close upon substrate binding[20] and a conserved arginine (R174 in *K. pneumoniae*, R170 in *E. coli*, R180 in *M. tuberculosis*, and R196 in *D. radiodurans*) has been found to be essential for UGM activity.[21] This arginine appears to stabilize the negatively charged diphosphate backbone of the sugar nucleotide substrate. Many synthetic analogues[22-27] have been used to probe the mechanism of UGM and investigate substrate binding, but until recently no ligand bound crystal structures have been available. Tryptophan fluorescence[15] and molecular modeling have predicted that the uridine of the UDP-Gal*p* stacks with W160 (numbering for *K. pneumoniae*);[28] in contrast, recent crystal structures of the *K. pneumoniae* UGM with bound UDP-Glc*p*[29] and UDP-Gal*p*,[30] and the *D. radiodurans* UGM bound to UDP-Gal*p*[12] all show that the uridine stacks against tyrosine 155 (kpUGM) or tyrosine 179 (drUGM) in the active site. This discrepancy demonstrates that many of the key binding interactions responsible for the substrate specificity of UGM still remain to be elucidated.

Although Gal*f* is the most common naturally occurring hexofuranose, it is not unique. 6-Deoxy-D-galactofuranose (D-Fuc*f*),[31] 6-deoxy-L-altrofuranose (L-6d-Alt*f*),[32] and 2-acetamido-2-deoxy-D-galactofuranose (D-Gal*f*NAc),[33, 34] among others, have also been identified in bacterial saccharide structures.

However, little is known about the biosynthesis of these other hexofuranose sugars. The work with the UGM from *K. pneumoniae*, which has been the most studied, has established that is unable to catalyze the synthesis of either UDP-D-Fuc*f* or UDP- D-Gal*f*NAc.[27, 35] It was recently demonstrated that a homologue of the *glf* gene, *fcf2* in *E. coli* O52, acts as a TDP-D-fucopyranose mutase enzyme for the biosynthesis of TDP-Fuc*f*.[35] This protein has 60% identity to the *K. pneumoniae* UGM, but the origin of the difference in substrate tolerance is unknown.

The bacterium *Campylobacter jejuni* is a foodborne pathogen that is a leading cause of diarrheal disease worldwide.[36] Infections by this organism have also been linked to the development of the neurological disorder Guillain–Barré syndrome.[37] Previous work showed that the capsular polysaccharide (CPS) from the 11168 strain contains a Gal*f*NAc residue (Figure 2-2).[34] *C. jejuni* 11168 also contains a homologue of the *glf* gene, *cj1439c*, among the CPS biosynthetic genes. Because no Gal*f* residues have been found in *C. jejuni* 11168 glycoconjugates, it has been proposed that the *cj1439c* gene product is responsible for the biosynthesis of UDP-Gal*f*NAc from UDP-Gal*p*NAc.[34] Herein we report the first studies on the protein produced by expression of *cj1439c* and demonstrate its activity as an UDP-*N*-acetyl-galactopyranose mutase (UNGM). We also demonstrate the enzyme can use both UDP-Gal and UDP-GalNAc as substrates and have investigated the origins of this substrate selectivity using site-directed mutagenesis to identify key residues involved in UDP-GalNAc turnover.

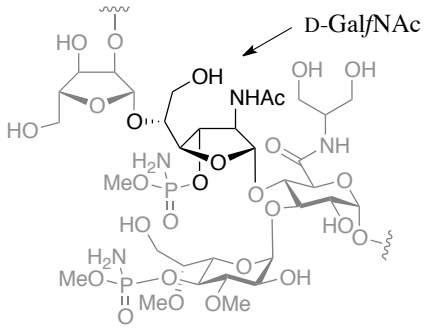


Figure 2-2. *C. jejuni* 11168H CPS tetrasaccharide repeat. The Gal*f*NAc residue is highlighted.

2.2. Results

2.2.1. *C. jejuni cj1439c* gene product functions as a UNGM *in vitro*

When the genome was first sequenced, the *cj1439c* gene from *C. jejuni* 11168 was annotated as a *glf* gene based on sequence homology,[38] despite the lack of Gal*f* residues in its glycoconjugates. Although it was later suggested the *glf* gene product may function in biosynthesis of Gal*f*NAc found in the CPS tetrasaccharide,[34] the function of the protein had never been examined. In this chapter, the activity of recombinant *C. jejuni* UNGM was examined by incubation with UDP-Gal*p*NAc (**3**) under the reducing conditions modified from Liu and Zhang,[16] and conversion to products was monitored by reversed-phase HPLC. Formation of a longer retention time product (12.5 min) was observed and, at equilibrium, a ratio of 7:93 of the product peak, with respect to the UDP-Gal*p*NAc peak, resulted as seen in Figure 2-3A (*i.* and *ii.*). This reflects the equilibrium ratios observed for the *K. pneumoniae*[27] and *E. coli*[16] UGM with UDP-Gal*p* (**1**). Scaling up the reaction gave access to sufficient quantities of the reaction product to characterize by ^{1}H NMR spectroscopy (Figure 2-4) and mass spectrometry. The observed mass was consistent with the product being UDP-

Gal*f*NAc (**4**), and the resonances in the ^{1}H NMR spectrum closely matched those previously reported for Gal*f*NAc residues in the CPS of *C. jejuni* 11168[34] and the O-specific polysaccharide of *Proteus penneri* strain 22.[33] In particular, the characteristic coupling pattern for the Gal*f*NAc H-4 proton and the upfield shift of the H-5 proton confirmed a galactofuranose ring configuration.[10] Incubation of the isolated product **4** with UNGM under the same reducing conditions gave, as expected, an equilibrium ratio of 7:93 UDP-Gal*f*NAc to UDP-Gal*p*NAc (Figure 2-3A *iv*).

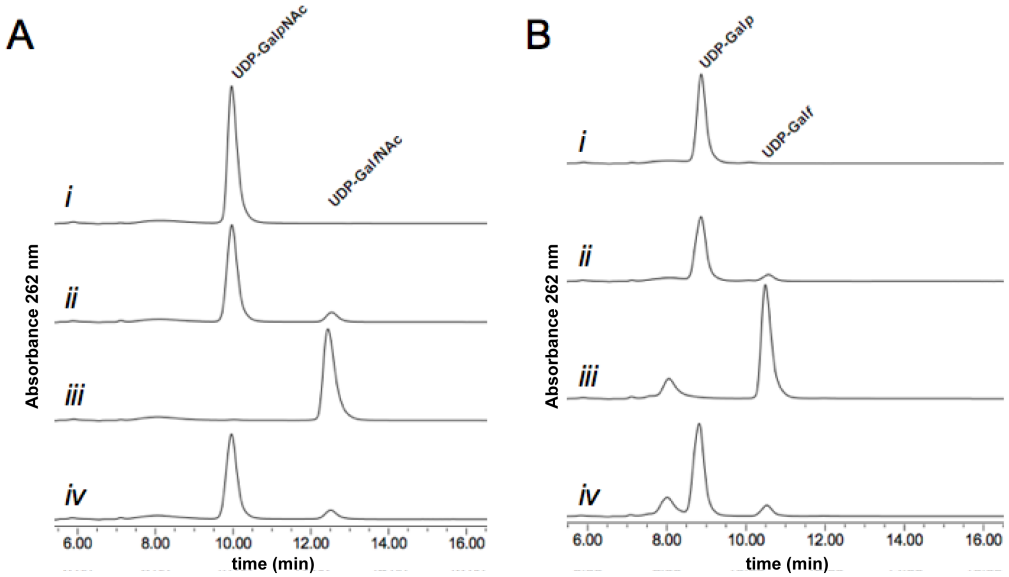


Figure 2-3. Functional characterization of cjUNGM (A) with UDP-GalNAc as the substrate. *i*. UDP-Gal*p*NAc standard. *ii.* Incubation of UDP-Gal*p*NAc with *C. jejuni* UNGM at equilibrium. *iii.* Purified reaction product (UDP-Gal*f*NAc). *iv.* Incubation of UDP-Gal*f*NAc with *C. jejuni* UNGM at equilibrium. The same ratio of 93:7 UDP-Gal*p*NAc to UDP-Gal*f*NAc is seen as in *ii*. (B) With UDP-Gal as the substrate. *i*. UDP-Gal*p* standard. *ii.* Incubation of UDP-Gal*p* with UNGM at equilibrium. *iii.* UDP-Gal*f* standard. *iv.* Incubation of UDP-Gal*f* with *C. jejuni* UNGM at equilibrium. The same ratio of 93:7 UDP-Gal*p* to UDP-Gal*f* seen in *ii,* is also observed. The peak at ~8 min corresponds to UMP.

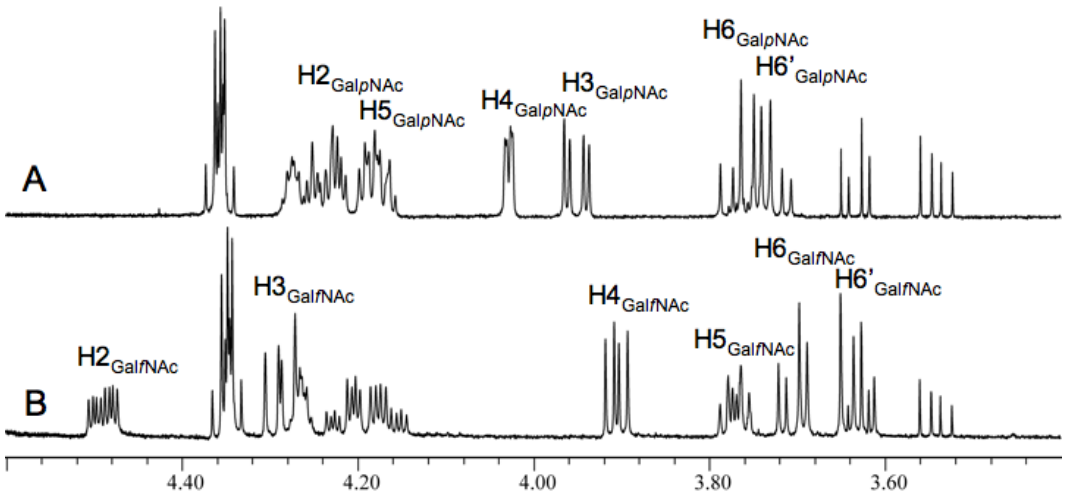


Figure 2-4. ^{1}H NMR analysis of the product of cjUNGM reaction with UDP-Gal*p*NAc. (A) A region of the ^{1}H NMR spectrum of UDP-Gal*p*NAc. (B) The same region of the ^{1}H NMR spectrum for the UNGM reaction product. The observed chemical shifts and coupling constants are consistent for UDP-GalNAc in the furanose ring form.

2.2.2 The *C. jejuni* UNGM interconverts UDP-Gal*f* and UDP-Gal*p in vitro*

The native CPS of *C. jejuni* 11168 contains only Gal*f*NAc residues and no Gal*f* residues.[34] To probe whether this observation is due to the specificity of the *C. jejuni* UNGM, the protein was incubated with both UDP-Gal*p* (**1**) and UDP-Gal*f* (**2**).[39] Interconversion between **1** and **2** was observed in both cases (Figure 2-3B). As was observed for incubations with **3**, the equilibrium ratio was consistent with the 7:93 UDP-Gal*f* to UDP-Gal*p* ratio previously reported for other UGMs.[16]

In a second experiment, co-incubation of UNGM with a 50:50 mixture of UDP-Gal*f* (**2**) and UDP-Gal*f*NAc (**4**) was carried out. This experiment was intended to provide an estimate of the relative specificity of the enzyme for these

two substrates. After approximately 60% conversion, the ratio of the corresponding pyranose product was 2.48:1 in favor of UDP-Gal*p* as shown in Figure 2-5, demonstrating the cjUNGM is bifunctional and that the enzyme preferentially recognized UDP-Gal*p*

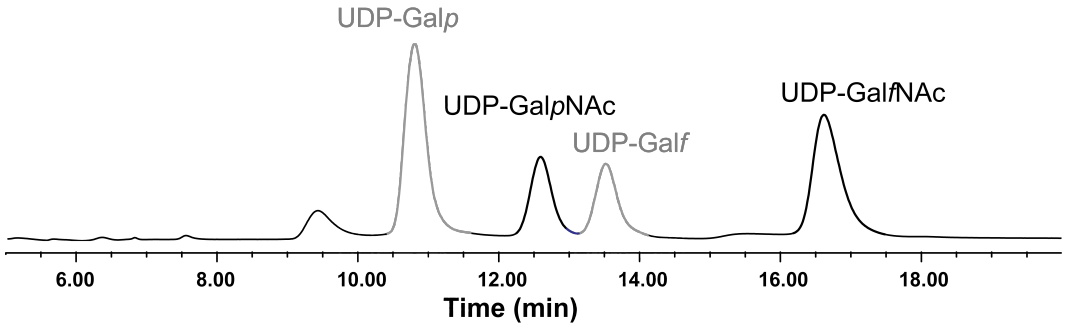


Figure 2-5. Co-incubation of cjUNGM with a 50:50 mixture of UDP-Gal*f* (2) and UDP-Gal*f*NAc (**4**). The ratio of the UDP-Gal*p* (1) and UDP-Gal*p*NAc product peaks was found to be 2.48:1 after ~60% total conversion.

2.2.3 *E. coli* UGM does not interconvert UDP-Gal*p*NAc and Gal*f*NAc

Earlier studies established that the UGM of *K. pneumoniae* can not interconvert UDP-Gal*p*NAc and UDP-Gal*f*NAc,[27] but no published data was available for the *E. coli* UGM. The *E. coli* UGM has a higher sequence identity, 60%, with the *C. jejuni* 11168 UNGM than with the *K. pneumoniae* UGM (38% identity) with which it shares a function. Because we determined the cjUNGM can turnover both UDP-Gal*p* and UDP-Gal*p*NAc, we wanted to explore if the closely homologous *E. coli* UGM recognized UDP-Gal*p*NAc as a substrate. We found that no conversion of UDP-Gal*p*NAc was detected using the *E. coli* UGM, even upon prolonged incubations (60 min). A similar experiment, using UDP-

Gal*f*NAc as the substrate, also demonstrated no conversion. To ensure that the protein was active, both UDP-Gal*p* and UDP-Gal*f* were used in the assay and activity levels consistent with previously published results[16] were observed.

2.2.4 CPS production is restored by expressing the *cj1439c* allele *in trans*

To test the role of the *glf*-His allele *in vivo*, the corresponding proteins were expressed *in trans* in *C. jejuni* wild-type (wt) and the *glf* mutant strain (*Δglf*). CPS formation was followed by silver staining of crude CPS preparations (Figure 2-6). Expression of the Cj-*glf*-His allele resulted in restoration of CPS in the *cj1439c*(*glf*) mutant, whereas no CPS formation was observed upon expression of Ec-*glf*-His. The expression of an additional copy of Cj-*glf*-His, Ec-*glf*-His or the presence of the empty plasmid (pWM1007/cat, negative control) in wt *C. jejuni* 11168 cells did not affect CPS formation (Figure 2-6A).

2.2.5 *C. jejuni* UNGM has UDP-galactopyranose mutase activity in *E. coli*

Because the specificity of cjUNGM allows it to act as a UGM *in vitro*, we desired a method to test this activity in a cellular environment similar to the Δ*glf* complementation experiment in *C. jejuni.* Unfortunately, no *C. jejuni* serotypes containing Gal*f* residues have been identified. Conveniently, common *E. coli* laboratory strains such as W3110 are derived from serogroup O16, which contain Gal*f* as a component of the lipopolysaccharide (LPS) O-antigen. In this species, the O-antigen is comprised of a pentasaccharide repeating unit (Figure 2-7) with a β-D-Gal*f* residue at the non-reducing end of the repeat unit.[40]

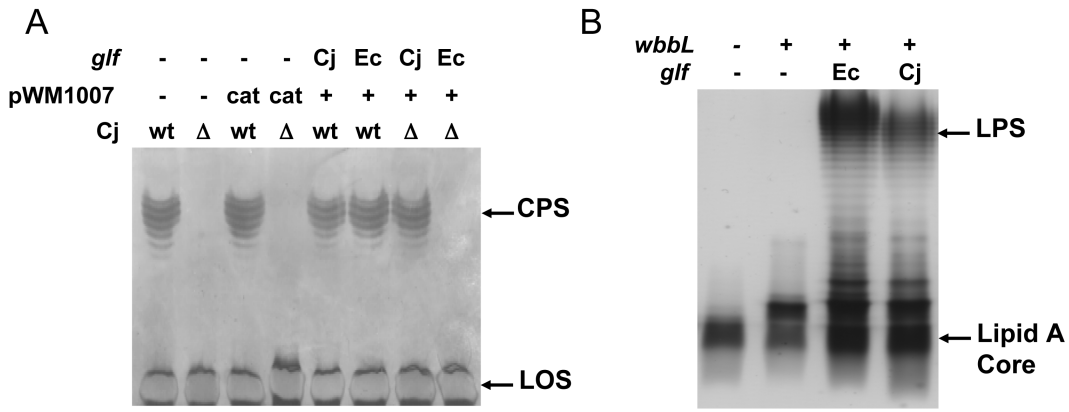


Figure 2-6. *C. jejuni glf* gene complements CPS in *C. jejuni* Δ*glf* strain and LPS in *E. coli* Δ*glf* strain. (A) Separation of *C. jejuni* 11168 crude CPS preparations by16.5% deoxycholate PAGE. Equivalent amounts of sample were loaded in each lane that originated from bacterial cell cultures adjusted to an absorbance (*A*600) of 3.0. Silver staining showed formation of CPS in the Cj-wt strain as well as Cj-wt with pMW1007/*cat*, pMW1007/Cj-*glf,* or pMW1007/Ec-*glf* plasmids (*wt* is wild type). The *C. jejuni 1439c* (*glf*) knock-out strain (Cj-Δ*glf*) and the Cj-Δ*glf* strain with pMW1007/*cat* plasmid showed no CPS formation. Cj-Δ*glf* strain complemented with pMW1007/Cj-*glf* showed restoration of CPS production, and no CPS was observed in Cj-Δ*glf* complemented with pMW1007/Ec-*glf.* In all strains lipooligosaccharide (*LOS*) formation was not affected. (B) LPS of *E. coli* strains derived from MFF1 were extracted after overnight growth and separated by SDS-PAGE. Equivalent amounts of sample were loaded in each lane that originated from bacterial cell cultures adjusted to an *A*600 of 0.45. Silver staining shows the production of a fast migrating band composed of lipid A core plus a GlcNAc residue (*wbbL* $^{-}$, *glf* $^{-}$); a band of higher molecular weight due to addition of an incomplete O antigen subunit to the lipid A core, which is only deficient of the Gal*f* residue (*wbbL* $^{+}$, *glf* $^{-}$); and smooth LPS when both *wbbL* and either the *E. coli* or the *C. jejuni glf* are present (*wbbL* $^{+}$, Ec-*glf* and *wbbL* $^{+}$, Cj-*glf*), respectively.

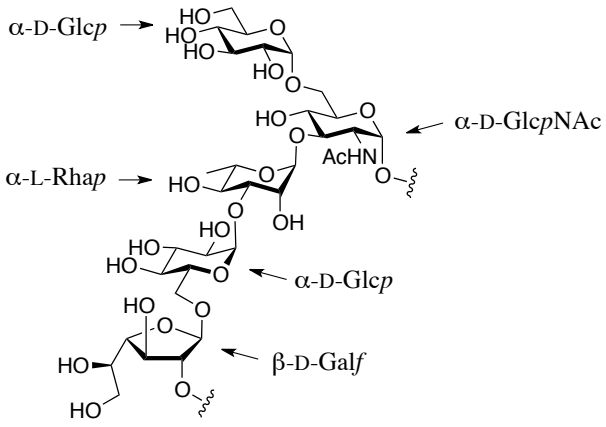


Figure 2-7. The pentasaccharide repeat unit of E. coli O16 LPS contains β-D-Gal*f* in the repeating unit.

The W3110 strain does not display a smooth LPS phenotype, due to an insertion sequence disrupting a gene encoding the rhamnosyltransferase (*wbbL*), which is required for the attachment of the second sugar of the O-antigen subunits. Re-introduction of the *wbbL* gene on a plasmid restores smooth LPS biosynthesis.[41, 42] By mutating the *E. coli* O16 *glf* gene in a *wbbL* complemented strain, only one incomplete O chain subunit, devoid of Gal*f*, is attached to the lipid A core (Figure 2-6B, lane 2[40]). Wild type LPS production is restored when *E. coli glf* is reintroduced (Figure 2-6B, Lane 3[40]). To prove *in vivo* UGM activity of the *C. jejuni* UNGM, we repeated the complementation of the *E. coli Δglf* mutation using a plasmid carrying *C. jejuni glf*. As shown in lane 4 of Figure 2-6B, the resulting strain produced full-length LPS, indicating that intact O chain subunits containing Gal*f* were synthesized and polymerized. Incorporation of Gal*f*NAc into the O-chain can be ruled out because subunits of the O16 LPS are connected via an α-(1→2) linkage to Gal*f* (Figure 2-7), which

would be prevented by the presence of an acetamido group at position 2, as in Gal*f*NAc. Furthermore, to the best of our knowledge, serogroup O16 *E. coli* strains do not produce UDP-Gal*p*NAc. Thus, the *E. coli* complementation studies indicate that *C. jejuni* UNGM can interconvert UDP-Gal*f* and UDP-Gal*p* *in vivo*.

2.2.5 Modeling the active site of cjUNGM suggests the origin of UDP-Gal*p*NAc recognition

To investigate the origin of the increased substrate scope of *C. jejuni* UNGM compared to the highly homologous *E. coli* UGM, we examined differences in the active site residues. No crystal structure for the *C. jejuni* UNGM had been determined; therefore, a homology model was generated based on the *E. coli* UGM for which a structure is available.[15] After building the protein model, UDP-Gal*p*NAc (Figure 2-8A) and UDP-Gal*f*NAc (Figure 2-8B) were docked into the active site. Inspection of the resulting structure showed that the nucleotide portion was bound in a similar conformation to that observed in the crystal structure of kpUGM or drUGM bound to UDP-Gal*p*,[12, 29] which were recently reported. As was seen in these structures, a key base stacking interaction is predicted to occur between the uridine base and a tyrosine residue (Y150) in the cjUNGM.

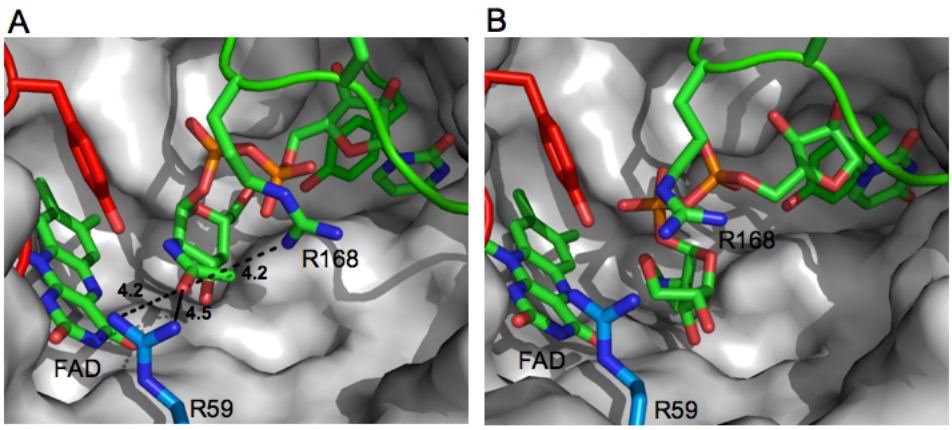


Figure 2-8. Homology model of the active site of *C. jejuni* UNGM, (A) with UDP-Gal*p*NAc or (B) UDP-Gal*f*NAc docked. UDP-Gal*p*NAc appears to be bound in an active conformation with the Gal*p*NAc C-1 located in proximity to the $FADH^-$ cofactor. Non-conserved active site residues R59 and R168 and their distance from the acetamido carbonyl oxygen are highlighted. The UDP-Gal*f*NAc appears bound in an inactive conformation (no possible interaction between Gal*f*NAc C-1 and $FADH^-$ cofactor).

The active site residues are highly conserved between the four bacterial UGM for which crystal structures have been determined (*E. coli*, *K. pneumonia*, *M. tuberculosis*, and *D. radiodurans*) and the *C. jejuni* UNGM. Only one active site residue, arginine 59 in the cjUNGM, differed from the conserved histidine found in all of the four bacterial UGMs (Table 2-1). A second residue, arginine 168 in cjUNGM, differs from the conserved lysine found in the ecUGM and kpUGM. As seen in Figure 2-8A, R59 in our docked homology model appears to be in close proximity to the carbonyl oxygen of the acetamido moiety of UDP-Gal*p*NAc. The other residue, R168, is located adjacent to the conserved arginine, in this case R169, in the mobile loop region which has previously been shown to be essential for UGM activity.[21]

Table 2-1. Numbering of conserved active site residues in bacterial UGM compared to cjUNGM

E. coli	*K. pneumoniae*	*D. radiodurans*	*M. tuberculosis*	*C. jejuni*
H56	H60	H88	H65	H56
H59	H63	H91	H68	***R59***
Y151	Y155	Y179	Y161	Y150
T152	T156	T180	T162	T151
W156	W160	W184	W166	W155
R170	R174	R198	R180	R169
Y181	Y185	Y209	Y191	Y180
N268	N270	N296	N282	N267
R278	R280	R305	R292	R277
E298	E301	E325	E315	E297
Y311	Y314	Y335	Y328	Y310
R340	R343	R364	R360	R339
Y346	Y349	Y370	Y366	Y345

2.2.6 Mutagenesis of arginine 59 and arginine 168 reduces the ability of UNGM to catalyze interconversion of UDP-Gal*f*NAc to UDP-Gal*p*NAc

The two non-conserved active site arginine residues, Arg 59 and 168, were examined for their role in the interconversion of UDP-Gal*f*NAc and UDP-Gal*p*NAc by cjUNGM. Site directed mutagenesis was used to convert these residues to the corresponding amino acid from the *E. coli* / *K. pneumoniae* UGMs; arginine 59 to histidine, and arginine 168 to lysine. The double mutant was also constructed with both R59H and R168K mutations. The specificity of the mutant UNGM was determined using a co-incubation assay with the same 50:50 mixture of UDP-Gal*f* and UDP-Gal*f*NAc described above (section 2.2.2). Again, the ratio of the corresponding pyranose products was used to determine the relative selectivity of each mutant for the two substrates (Figure 2-9). The

R59H mutation resulted in a decrease in the conversion of UDP-Gal*f*NAc and the selectivity for UDP-Gal*f* conversion increased by greater than ten-fold over the wild-type cjUNGM. The R168K mutation had a less pronounced effect causing only an approximate two-fold increase in the selectivity for UDP-Gal*f* compared to the wild-type enzyme. Mutation of both residues R59H and R168K resulted in the largest decrease in the conversion of UDP-Gal*f*NAc to UDP-Gal*p*NAc, and a selectivity increase for UDP-Gal*f* of greater than twelve-fold in comparison to the wild-type UNGM.

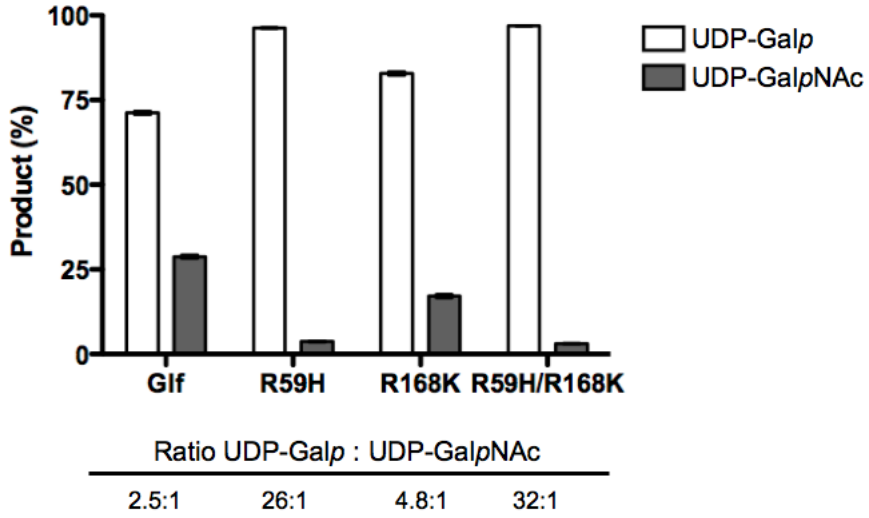


Figure 2-9. Co-incubation assay of *C. jejuni* UNGM with 50:50 UDP-Gal*f*NAc and UDP-Gal*f*. After co-incubation the relative amount of UDP-Gal*p* and UDP-Gal*p*NAc were determined based on the total amount of products observed. Shown are the average of three co-incubation assays for each of the *C. jejuni* UNGM, and the R59H, R168K, and R59H/R168K UNGM mutants. Also shown is the specificity of the protein represented by the ratio of UDP-Gal*p* to UDP-Gal*p*NAc.

2.2.7 Analysis of cjUNGM and mutant kinetics with UDP-Gal*f* and UDP-Gal*f*NAc supports results of the co-incubation assay

At equilibrium, the pyranose ring form is favored in the reaction of pyranose–furanose mutase enzymes. Therefore, substrate kinetics are often measured starting with the furanose form and monitoring the formation of the thermodynamically favored pyranose isomer.[43] Thus, the kinetic parameters were determined for cjUNGM and each of the mutants using UDP-Gal*f* (**2**) and UDP-Gal*f*NAc (**4**) as substrates. As seen in Table 2-1, the K_M of wild type UNGM is approximately the same for both UDP-Gal*f* and UDP-Gal*f*NAc; however, the k_{cat} value is larger for UDP-Gal*f*. The R59H mutant showed only small changes in the K_M for both substrates as compared to the wild-type cjUNGM. At the same time, there is a measurable decrease in k_{cat} observed for UDP-Gal*f*NAc in addition to an increase in the k_{cat} for UDP-Gal*f*. Conversely, the R168K mutant displayed approximately the same k_{cat} for both substrates, but the K_M was larger for UDP-Gal*f*NAc as the substrate. For each protein, the first order rate constants were approximated by calculating k_{cat}/K_M for each substrate (Table 2-2). The ratio of the first order rate constants for UDP-Gal*f* over UDP-Gal*f*NAc could be used to approximate the specificity of each protein for UDP-Gal*f* as the incubation times for each substrate remained constant.[44] The calculated specificities in each case mirror those determined using the co-incubation assay (Figure 2-9), but the former approach underestimates these differences.

Table 2-2. Kinetic parameters for the *C. jejuni* UNGM and mutants.

Enzyme	Substrate	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m ($\mu M^{-1}min^{-1}$)	UDP-Gal / UDP-GalNAc
Wild-type UNGM	UDP-Gal*f*	45 ± 3	178 ± 4	4.1	1.5
	UDP-Gal*f*NAc	40 ± 6	114 ± 5	2.8	
R59H	UDP-Gal*f*	92 ± 20	590 ± 50	6.4	15
	UDP-Gal*f*NAc	61 ± 7	26 ± 1	0.43	
R168K	UDP-Gal*f*	65 ±9	96 ± 4	1.5	3.8
	UDP-Gal*f*NAc	231 ± 40	89 ± 7	0.39	
R59H/ R168K	UDP-Gal*f*	77 ± 9	380 ± 20	4.9	21
	UDP-Gal*f*NAc	59 ± 6	13.8 ± 0.5	0.23	
E. coli UGM	UDP-Gal*f*NAc	--[a]	--[a]	--[a]	--[a]

[a] No turnover detected

2.3 Discussion

In this chapter, we have tested a previously proposed hypothesis [45] that the *C. jejuni* 11168 gene *cj1439c* encodes a protein responsible for the biosynthesis of UDP-Gal*f*NAc. Our investigations have shown that, unlike the highly homologous UGM from *E. coli* and *K. pneumoniae*, the *C. jejuni* enzyme is able to convert the sugar nucleotide UDP-Gal*p*NAc to UDP-Gal*f*NAc, the expected precursor of Gal*f*NAc in the CPS. Thus, the protein is a UDP-*N*-acetyl-galactopyranose mutase (UNGM). In addition, we have demonstrated that the *C. jejuni* UNGM also converts UDP-Gal*p* to UDP-Gal*f* both *in vitro* and *in vivo* in *E. coli*. We also identified two amino acids in the active site of UNGM that play a role in the interconversion of UDP-Gal*f*NAc and UDP-Gal*p*NAc. Our co-incubation assay allowed us to examine, in a single reaction, the substrate specificity of wild-type UNGM protein and each of the UNGM mutants.

2.3.1 *cj1439c* encodes for a protein with both UGM and UNGM activity.

This report represents the first demonstration of an enzyme involved in the interconversion of UDP-Gal*p*NAc and UDP-Gal*f*NAc. This enzyme, which bears a high sequence similarity to known UGMs, produces UDP-Gal*f*NAc from UDP-Gal*p*NAc presumably via a similar ring contraction mechanism. Both *in vitro* investigations and an *in vivo* complementation experiment support the role of UNGM in the biosynthesis of the Gal*f*NAc in *C. jejuni* 11168. In addition, we have demonstrated that the *C. jejuni* UNGM has dual substrate specificity and can interconvert the furanose and pyranose isomers of both UDP-Gal and UDP-GalNAc. Furthermore, the UGM activity of the *C. jejuni* enzyme has been demonstrated *in vivo* where it is able to complement the activity of the *E. coli* UGM in a Δ*glf* gene knock-out.

It is not unusual that bacteria with compact genomes express enzymes that exhibit more than one activity. These bifunctional enzymes are widespread among bacteria and allow for the synthesis of many complex structures advantageous to the survival of the organism while still maintaining a small genome size. An example is the *wbbO* gene product from *K. pneumoniae*, a galactosyltransferase that catalyzes the transfer of both Gal*p* and Gal*f* residues in the biosynthesis of the lipopolysaccharide O1 antigen.[46] Bifunctional enzymes have also been characterized in *C. jejuni*. For example, a single UDP-GlcNAc/Glc 4-epimerase was shown to be involved in the biosynthesis of three cell surface glycoconjugates in strain 11168.[47] Because the CPS structures in *C.*

jejuni are highly variable between serotypes, it is reasonable to hypothesize that a bifunctional UNGM would be advantageous. However, no Gal*f* residues have been identified in any glycoconjugate from this organism. Thus, the inclusion of Gal*f*NAc, instead of Gal*f,* into the CPS appears to be due to the specificity of the cognate glycosyltransferase.

2.3.2 Comparing calculated specificity of *C. jejuni* UNGM to the experimentally determined specificity

In this study, two methods were used to determine the substrate specificity of the *C. jejuni* UNGM. The first was the direct determination from the ratio of product conversion observed in the co-incubation assay (Figures 2-5 and 2-9). The second was calculated using the observed kinetic parameters for each protein with UDP-Gal*f* or UDP-Gal*f*NAc (Table 2-2). The observed trend of substrate specificity is the same by either method; however, the calculated specificity in each case is lower than that determined using the co-incubation assay. The calculated substrate specificity may not accurately represent the observed specificity as the combined rate of the two competing reactions may be greater than, equal to, or lower than the rate of each individual reaction.[48] This may aid in explaining the difference in specificity measured for the competitive assay compared to that calculated from the enzymatic rates of UNGM for its two substrates UDP-Gal*f* and UDP-Gal*f*NAc.

2.3.3 The effect of two active site arginines on UDP-GalNAc recognition by cjUNGM

We hypothesized that residues in, or in proximity to, the active site would play a role in the different substrate specificity of the cjUNGM compared to other known bacterial UGM enzymes. We also rationalized that these key residues would be, at least partially, conserved in the other UGM but not in the *C. jejuni* enzyme. Two residues that fit these preliminary criteria were R59 and R168. In both cases, these residues were found to be other basic amino acids, histidine and lysine respectively, in the *E. coli* and *K. pneumoniae* UGM. In the drUGM and mtUGM, R59 was also replaced by histidine. However, R168 was found to be an alanine and threonine, in drUGM and mtUGM, respectively, rather than the lysine observed in ecUGM and kpUGM.

Mutagenesis of R168K only resulted in a two-fold increase in selectivity for UDP-Gal*f* over UDP-Gal*f*NAc in the co-incubation assay; however, there were more significant changes observed for the kinetic parameters. This amino acid change resulted in a decrease in the catalytic activity of the protein for both substrates compared to the wild-type UNGM but most interesting was that this substitution resulted in a increase in the K_M with UDP-Gal*f*NAc but not with UDP-Gal*f*. As the K_M approximates the dissociation constant for all enzyme–substrate complexes, this supports the notion that R168 has a role in binding and stabilizing UDP-GalNAc in the active site. It has previously been proposed that amino acid residues in the mobile loop of UGMs are involved in substrate recognition.[20, 21] Because R168 is located in the mobile loop, its ability to

stabilize the UDP-GalNAc enzyme–substrate complex is consistent with the role of the mobile loop in substrate recognition. This arginine residue is located 4.2 Å from the acetamido carbonyl oxygen in our homology model with docked UDP-Gal*p*NAc (Figure 8A). Although this distance is long for a hydrogen bonding interaction, the cjUNGM homology model is not based on the crystal structure of the catalytically-active closed form of the enzyme; therefore, it is likely that the residue resides closer in the active closed conformation.

Mutation of the other active site arginine (R59) indicates that it also plays an important role in the catalytic activity of UNGM. Mutagenesis of R59H results in a greater than four-fold decrease in k_{cat} for UDP-Gal*f*NAc, while simultaneously resulting in a three-fold increase in k_{cat} for UDP-Gal*f*. The K_M is also changed for both substrates but this appears minor in comparison to the observed changes in k_{cat}. Considering our homology model, R59 is located within 4.5 Å of the carbonyl oxygen of the acetamido group of UDP-Gal*p*NAc (Figure 2-8A). In the *E. coli* UGM crystal structure (1I8T) the equivalent residue H59 sits below the active site (not shown) and does not protrude into the active site as does R59 in UNGM. Therefore, it appears that R59 is able to interact with the UDP-GalNAc substrate to stabilize the intermediate. This may occur by preventing the formation of non-productive oxazoline-like intermediates, which could be formed by an intramolecular reaction of the acetamido group (Figure 2-10). When the substrate is UDP-Gal, then there is no possibility of forming such intermediates. Therefore, R59 does not aid in catalysis. Instead, because

arginine is more bulky than histidine, it could lower the catalytic rate due to steric interactions in the active site during catalysis.

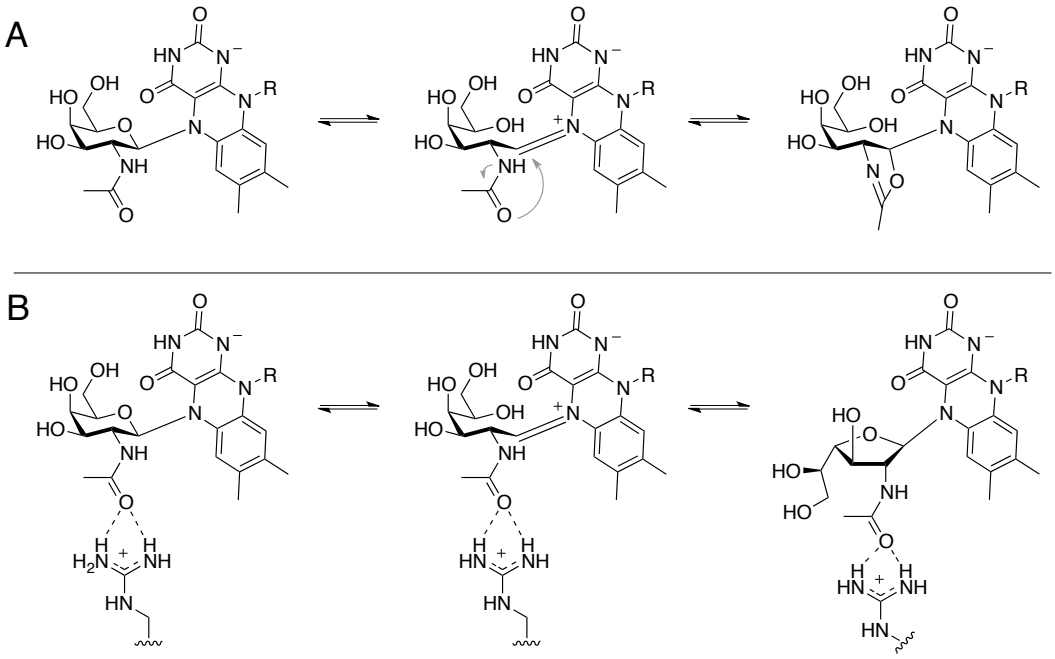


Figure 2-10. Formation of oxazoline intermediate. (A) The acetamido group of Gal*p*NAc could potentially undergo an unproductive intramolecular reaction to form an FAD bound oxazoline intermediate that prevents conversion to Gal*f*NAc. (B) Possible interaction between R59 and the Gal*p*NAc acetamido group that could prevent formation of an oxazoline intermediate and facilitate conversion to Gal*f*NAc.

The mutagenesis of both R59H and R168K results in a larger decrease in turnover of UDP-Gal*f*NAc while simultaneously increasing turnover of UDP-Gal*f*. This results in an increased selectivity for UDP-Gal*f* than observed for either of the single mutants (Figure 2-9; Table 2-2). Somewhat surprisingly, the increase in K_M for UDP-Gal*f*NAc seen for the R168K mutant was not observed in the case of the double mutant. This suggests that the amino acids play a

synergistic role in allowing the enzyme to interconvert UDP-Gal*f*NAc and UDP-Gal*p*NAc, rather than an additive role in which R59 is the major determinant.

2.3.4 Subtle amino acid substitutions result in changes in cjUNGM substrate tolerance

Despite the high sequence identity of *C. jejuni* UNGM with the *E. coli*, *M. tuberculosis*, *D. radiodurans* and *K. pneumoniae* UGM, the UNGM possesses the ability to recognize UDP-GalNAc as a substrate. Most of the sequence variability between these proteins occurs in solvent-exposed residues and the amino acids making up the active site are nearly identical in all five mutase enzymes (Figure 2-8, Table 2-1). The two residues identified here as playing an important role in allowing recognition of UDP-GalNAc by cjUNGM are relatively conservative replacements of the residues found in the UGM enzymes; however, they nevertheless have a notable effect on the substrate selectivity of the enzyme. It should be appreciated that small variations in amino acids leading to changes in substrate specificity are well known in carbohydrate active-enzymes. For example, the blood group GTA and GTB glycosyltransferases, which use UDP-Gal*p*NAc *vs* UDP-Gal*p* respectively, as the donor substrate*,* differ in only four amino acids.[49] Similarly, in *Neisseria meningitidis,* a single amino acid change in the capsule polymerase determines the substrate specificity for either Glc*p* or Gal*p* transferase activity.[50]

Although the two arginine residues identified in this study influence the substrate specificity, the mutagenesis of either residue or both failed to result in a

complete loss in UNGM activity. It is therefore clear that other amino acids further removed from the active site also contribute to the specificity of the enzyme, and these remain to be elucidated. As well, we have demonstrated that *C. jejuni* UNGM can function as a UGM *in vivo*; however, there have been no Gal*f* residues reported in *C. jejuni* glycoconjugates. In this context, another unresolved issue is if other *C. jejuni* strains containing the *glf* gene possess Gal*f*-containing glycoconjugates and investigations into the specificity of the transferases that utilize the products of this bifunctional enzyme are warranted. This study further demonstrates the intricacies in bacterial glycoconjugate biosynthesis. Detailed understanding of these systems could allow for the development of novel antimicrobials targeting these pathogen-specific pathways.

2.4. Experimental Details

General Procedures

All reagents were purchased from commercial sources without further purification. Primers were purchase from Integrated DNA Technologies (IDT) and are listed in Appendix A (Table A-1). NMR spectra were obtained at 27 °C using a Varian DirectDrive two-channel spectrometer operating at 500 MHz for ^{1}H. Samples were prepared at concentrations of ~1.5 mM in D_2O. Chemical shifts were externally-referenced to 0.1% acetone signal (2.225 ppm). All coupling constants are reported in Hz. Electrospray mass spectra were recorded on samples suspended in CH_3Cl or CH_3OH and added NaCl.

Preparation of UDP-sugars

UDP-Gal*p* and UDP-Gal*p*NAc were obtained as the disodium salt from Aldrich and used without further purification. UDP-Gal*f* was prepared from synthetic Gal*f*-1-phosphate[51] using galactose-1-phosphate uridylyltransferase and UDP-glucose pyrophosphorylase.[39, 52] All stock solutions were prepared by dissolving the appropriate quantity of UDP-sugar in 100 mM potassium phosphate (pH 7.4). Before use, stock solutions were calibrated by HPLC co-injection with a known concentration of UDP-Glc.

Cloning, expression and purification of Glf proteins

For MFF1 (Ec-Δ*glf* mutant strain), and Cj1439c (Cj-Δ*glf* mutant strain), complementation and *in vivo* analyses, *glf* alleles were put under the control of the constitutive Cj-Ec shuttle promoter replacing the *gfp* gene on plasmid pWM1007.[53] Cj-*glf* alleles were amplified by PCR using oligonucleotides CS261 and CS262 (Appendix Table A-1) that introduce restriction sites for *EcoR*I and *BsrG*I while Ec-*glf* was amplified with oligonucleotides CS362 and CS363 introducing an *EcoR*I site in the 5' end of Ec-*glf*. For both species a C-terminal His-tag was introduced via PCR, where the plasmid pJHCV32[54] and chromosomal DNA of *C. jejuni* 11168-V26[55] served as template DNA, respectively. The *EcoR*I-*BsrG*I digested Cj-*glf* PCR product was ligated with the purified 8643 bp pWM1007 vector DNA fragment obtained after digest with the same enzymes. The Ec-*glf* PCR product, subsequently treated with T4 DNA polymerase and *EcoR*I, was inserted into the purified 8297 bp pWM1007 vector

subsequently treated with *Sfu*I, T4 DNA polymerase and *EcoR*I. For expression in *C. jejuni* the *kan* (kanamycin) cassette within the pWM1007-Cj-*glf* construct was replaced by the *cat* (chloramphenicol) cassette after *EcoR*V digestion of the vector and ligation with a 842 bp DNA fragment containing the *cat* cassette isolated from plasmid pRY109[56] after *Sma*I digestion. A similar strategy was carried out for the pWM1007 Ec-*glf* constructs, except that the 8428 bp vector fragment was purified after partial digest with *EcoR*V. Orientation of the *cat* gene on the resulting plasmids (same orientation as the non-polar *kan* cassette) was verified by restriction analyses. For high yield expression of Ec-*glf*, the corresponding gene was amplified by PCR using oligonucleotides CS372 and CS373 to introduce *Nde*I and *Xho*I restriction sites, respectively. Plasmid pJHCV32[54] served as template DNA for the PCR. After restriction digestion, the purified DNA fragment was ligated into plasmid pET22b cut with the same enzymes. Dr. Harald Nothaft, in the group of our collaborator Professor Christine Szymanski, performed the protein cloning experiments.

Expression of soluble C-terminal hexa-histidine-tagged *E. coli* UGM protein from plasmid pET22b:E*c-glf* in BL21 was observed after induction with 0.01 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) for 2 h at room temperature (22 °C). The protein was subsequently purified by Ni-NTA affinity chromatography as per the manufactures direction to give approximately 20 mg of >95% pure soluble protein per liter of culture. C-terminal hexa-histidine-tagged *C. jejuni* UNGM was expressed in *E. coli* DH5α from plasmid pWM1007:(*kan*)-Cj-*glf* after growth for 18 h at 28 °C. Soluble UNGM-His_6

proteins were purified by Ni-NTA affinity chromatography as above, giving approximately 3 mg of >95% pure protein per liter of culture. The quantity of protein was determined using the BCA method with Bovine Serum Albumin standards.

cjUNGM activity assay

The activity of purified wild-type and mutant proteins was assayed by incubating a mixture of sugar nucleotide (UDP-Gal*p*, UDP-Gal*f*, UDP-Gal*p*NAc, or UDP-Gal*f*NAc, 1 mM) and mutase protein (3.9 μM) in 30 μL of 100 mM potassium phosphate buffer (pH 7.4) containing freshly prepared sodium dithionite (20 mM) for 2, 5, 10, and 20 min periods at 37 °C. Reactions were monitored by HPLC (Varian Prostar 210) following conditions similar to those previously reported by Zhang and Liu.[16] A C_{18} column (Microsorb-MV, Varian, 4.6 × 250 mm) was used with an isocratic elution of 50 mM triethylammonium acetate buffer (pH 6.5) containing 1.5% acetonitrile. A flow rate lower than that described earlier[16] (0.6 mL/min) was used to increase separation and the UV detector was set to a wavelength of 262 nm. Base line resolution for all substrates was achieved and the retention times for UDP-Gal*p*, UDP-Gal*f*, UDP-Gal*p*NAc, and UDP-Gal*f*NAc were found to be 8.8 min, 10.6 min, 10.0 min, and 12.5 min, respectively. The amount of conversion was determined by integration of the product and starting material peaks.

Enzymatic synthesis of UDP-Gal*f*NAc (4)

To confirm the identity of the product of the UNGM incubation with UDP-Gal*p*NAc (**3**), a milligram scale reaction was carried out. UDP-Gal*p*NAc (1 mM, 25 mg) was incubated with 4.7 μM UNGM in 2 mL of 100 mM potassium phosphate buffer (pH 7.4) containing freshly prepared sodium dithionite (20 mM) for 1 h at 37 °C. The reaction product was purified by HPLC using a C_{18} column (Microsorb, Varian, 21.4 × 250 mm) with 50 mM triethylamonium acetate (pH, 6.5) containing 1.5% acetonitrile at a flow rate of 7.0 mL/min. The product retention time was found to be 22.3 min under these conditions. This product was further purified by chromatography on a Sephadex G-15 column (2.1 × 250 cm) eluting with milliQ H_2O. The product was lyophilized to obtain a white powder (1 mg, 4% yield). ^{1}H NMR (500 MHz, D_2O) δ 7.92 (d, *J* = 8.1 Hz, 1H, H-6), 5.98 (d, *J* = 4.6 Hz, 1H, H-1′), 5.95 (d, *J* = 7.9 Hz, 1H, H-5), 5.64 (dd, *J* = 5.4, 4.6 Hz, 1H, H-1′′), 4.49 (ddd, *J* = 9.3, 4.4, 2.5 Hz, 1H, H-2′′), 4.38 – 4.32 (m, 2H, H-2′/H-3′), 4.29 (dd, *J* = 9.3, 7.6 Hz, 1H, H-3′′), 4.27 – 4.24 (m, 1H, H-4′), 4.22 (ddd, *J* = 11.8, 4.6, 2.6 Hz, 1H, H-5a′), 4.17 (ddd, *J* = 11.8, 5.7, 3.1 Hz, 1H, H-5b′), 3.91 (dd, *J* = 7.6, 5.0 Hz, 1H, H-4′′), 3.77 (app. dt, *J* = 7.3, 4.6 Hz, 1H, H-5′′), 3.71 (dd, *J* = 11.8, 4.4 Hz, 1H, H-6a′′), 3.63 (dd, *J* = 11.8, 7.4 Hz, 1H, H-6b′′), 1.90 (s, 3H, NHC(O)C*H*$_3$); HRMS (ESI) *m*/*z* Calc. for $(M–2H)^{2-}$ $C_{17}H_{25}N_3O_{17}P_2$: 302.5335. Found: 302.5339.

Determination of the pyranose–furanose distribution at equilibrium

The distribution of pyranose to furanose ring forms were measured at equilibrium and determined by integration of the appropriate peaks from the HPLC chromatogram. The ratio was determined for both UDP-Gal and UDP-GalNAc substrates in both the forward and reverse directions. For the forward reaction UDP-Gal*p* or UDP-Gal*p*NAc (1 mM) was incubated with the appropriate mutase protein (3.9 μM) in 30 μL of 100 mM potassium phosphate buffer (pH 7.4) with freshly prepared sodium dithionite (20 mM) until equilibrium was reached as indicated by a constant product/substrate ratio by HPLC. The same procedure was used for the reverse reaction but instead using UDP-Gal*f* or UDP-Gal*f*NAc (1 mM) as the starting substrate.

Complementation of *C. jejuni* 11168 Δ*glf* knock out

The *C. jejuni* complementation experiments were performed by Dr. Harald Nothaft in the group of our collaborator Professor Christine Szymanski. Shuttle plasmids (pWM1007/*cat*-derivatives) expressing the *E. coli glf* and *C. jejuni glf* genes were mobilized into *C. jejuni* 11168 wild-type and the Δ*glf* mutant as described.[57] Capsular polysaccharides of the resulting strains were prepared as previously described[58] and visualized after 16.5% deoxycholate polyacrylamide gel-electrophoresis by silver staining using the protocol of Tsai and Frasch[59] with the modification that fixing was performed for only 2 h.

Testing cjUNGM activity in *E. coli*

The *E. coli* Δ*glf* mutant strain MFF1[40] was transformed with plasmids containing *E. coli wbbL* (pMF19)[40] and either empty vector pWM1007, or pWM1007 containing *glf* from *E. coli* or from *C. jejuni*. This was done by Dr. Isabelle Hug in the group of our collaborator Professor Mario Feldman. Strains were grown overnight after induction with 0.5 mM IPTG. Lipopolysaccharide (LPS) were extracted following a modified protocol of Marolda *et al*.[60] Briefly, cells were adjusted to an optical density of 3.0 (600 nm wave length), resuspended in 150 μL lysis buffer (2% SDS, 4% β-mercaptoethanol, 10% glycerol, 0.5 M Tris, pH 6.8) and incubated at 100 °C for 10 min. After addition of 2 μL proteinase K (20 mg/mL), samples were incubated for 2 h at 60 °C. To this solution was added 150 μL hot phenol and samples were incubated at 70 °C for 15 min, followed by 10 min on ice. After centrifugation, the aqueous phase was mixed with 250 μL EtOH, centrifuged, and the precipitated LPS were dried at room temperature. LPS corresponding to a cell optical density of 0.45 were separated by SDS-PAGE, and LPS silver staining was performed according to Tsai and Frasch.[59]

Homology model of cjUNGM with docked UDP-Gal*p*NAc

A homology model of *C. jejuni* 11168 UNGM was generated based on the crystal structure of the *E. coli* UGM in the closed ring form (1I8T chain A),[15] using ESyPred3D.[61] Chain A was used as it has the mobile loop region in the 'closed' conformation, which better represents the enzyme's active

conformation. Using Autodock 4.0,[62] the UDP-Gal*p*NAc and UDP-Gal*f*NAc substrates were modeled into the active site. For each substrate, the low energy conformation, in which the nucleotide conformation agreed with the published crystal structure data of UGM bound to UDP-Glc*p*[29] or UDP-Gal*p*[12, 30] was chosen to represent a plausible binding mode.

Site–directed mutagenesis of cjUNGM

The R59H, R168K, and R59H/R168K mutants were prepared using the Quikchange XL II protocol by Stratagene.[63] In the case of the R59H and R168K, mutants were prepared from the pWM1007 plasmid containing the *glf* gene isolated from *E. coli* DH5α cells using *R59H_F* and *R59H_R*, or *R168K_F* and *R168K_R* (Appendix Table A-1) as primer pairs, respectively. In the case of R59H/R168K, the double mutant was prepared using the R59H mutant pWM1007 plasmid and the *R168K_F* and *R168K_R* primer pair. After mutagenesis, the plasmid DNA was isolated from the XL10-gold cells and sequenced before being transformed into DH5α cells for protein expression.

Competitive substrate specificity assay

A mixture of UDP-Gal*f* (0.5 mM) and UDP-Gal*f*NAc (0.5 mM) was incubated with the appropriate mutase protein (3.9 μM) in 100 mM potassium phosphate buffer (pH 7.4) containing freshly prepared sodium dithionite (20 mM) for 5 min. The reactions were again monitored using the same modified HPLC conditions of Zhang and Liu.[16] In this case, the use of a lower

concentration of acetonitrile (1.25%) and a flow rate of 0.6 mL/min allowed for better resolution of all four product and substrate peaks (UDP-Gal*p*, UDP-Gal*f*, UDP-Gal*p*NAc, and UDP-Gal*f*NAc). Using these conditions, these compounds were found to have retention times of 10.0 min, 12.4 min, 11.6 min, and 15.4 min, respectively. The relative specificity was determined from the ratio of the integrated UDP-Gal*p* and UDP-Gal*p*NAc peaks.

Determination of the cjUNGM and cjUNGM mutant kinetic parameters

Kinetic parameters for wild-type *C. jejuni* UNGM and each of the mutants were determined following a kinetic assay modified from the procedure reported by Zhang and Liu.[16] Reactions were prepared containing an appropriate concentration of the desired protein with UDP-Gal*f* or UDP-Gal*f*NAc (10, 12.5, 25, 50, 100, 250, and 500 μM) in a final volume of 60 μL of 100 mM potassium phosphate (pH 7.4) containing 20 mM of freshly prepared sodium dithionite. Incubations were carried out for 5 min at 37 °C and then promptly quenched by heating to 90 °C for 5 min. The sugar nucleotides do not decompose by loss of UDP or UMP under these quenching conditions. The incubation mixtures were monitored by HPLC as described above. In all cases, a concentration of protein was used so that less than 40% conversion to the pyranose product was observed. The concentrations of UDP-Gal*p* or UDP-Gal*p*NAc were determined by integration of the appropriate peaks on the HPLC trace, and these were used to determine the initial velocities. Each assay was performed in duplicate for all proteins for each of the furanose substrates. Kinetic parameters K_M and k_{cat} were

obtained by nonlinear regression analysis of the Michaelis–Menten equation using GraphPad PRISM 4 (GraphPad Software, San Diego, CA) (see Appendix A). The calculated enzyme specificity was determined using a modification of the Michaelis–Menten equation for two competing substrates:[44]

$$\frac{v_A}{v_B} = \frac{\left(k_{cat}/K_M\right)_A[A]}{\left(k_{cat}/K_M\right)_B[B]}$$

2.5. Bibliography

1. de Lederkremer, R. M.; Colli, W., *Glycobiology* **1995,** *5*, 547-552.
2. Whitfield, C., *Trends Microbiol.* **1995,** *3*, 178-185.
3. Brennan, P. J.; Nikaido, H., *Annu. Rev. Biochem.* **1995,** *64*, 29-63.
4. Lee, R. E.; Smith, M. D.; Nash, R. J.; Griffiths, R. C.; McNeil, M.; Grewal, R. K.; Yan, W. X.; Besra, G. S.; Brennan, P. J.; Fleet, G. W. J., *Tetrahedron Lett.* **1997,** *38*, 6733-6736.
5. Pan, F.; Jackson, M.; Ma, Y. F.; McNeil, M., *J. Bacteriol.* **2001,** *183*, 3991-3998.
6. Peltier, P.; Euzen, R.; Daniellou, R.; Nugier-Chauvin, C.; Ferrières, V., *Carbohydr. Res.* **2008,** *343*, 1897-1923.
7. Pedersen, L. L.; Turco, S. J., *Cell. Mol. Life Sci.* **2003,** *60*, 259-266.
8. Beláňová, M.; Dianišková, P.; Brennan, P. J.; Completo, G. C.; Rose, N. L.; Lowary, T. L.; Mikušová, K., *J. Bacteriol.* **2008,** *190*, 1141-1145.
9. Nassau, P. M.; Martin, S. L.; Brown, R. E.; Weston, A.; Monsey, D.; McNeil, M. R.; Duncan, K., *J. Bacteriol.* **1996,** *178*, 1047-1052.

10. Köplin, R.; Brisson, J. R.; Whitfield, C., *J. Biol. Chem.* **1997,** *272*, 4121-4128.

11. Weston, A.; Stern, R. J.; Lee, R. E.; Nassau, P. M.; Monsey, D.; Martin, S. L.; Scherman, M. S.; Besra, G. S.; Duncan, K.; McNeil, M. R., *Tuber. Lung Dis.* **1998,** *78*, 123-131.

12. Partha, S. K.; van Straaten, K. E.; Sanders, D. A. R., *J. Mol. Biol.* **2009,** *394*, 864-877.

13. Beverley, S. M.; Owens, K. L.; Showalter, M.; Griffith, C. L.; Doering, T. L.; Jones, V. C.; McNeil, M. R., *Eukaryotic Cell* **2005,** *4*, 1147-1154.

14. Bakker, H.; Kleczka, B.; Gerardy-Schahn, R.; Routier, F. H., *Biol. Chem.* **2005,** *386*, 657-661.

15. Sanders, D. A. R.; Staines, A. G.; McMahon, S. A.; McNeil, M. R.; Whitfield, C.; Naismith, J. H., *Nat. Struct. Biol.* **2001,** *8*, 858-863.

16. Zhang, Q. B.; Liu, H. W., *J. Am. Chem. Soc.* **2000,** *122*, 9065-9070.

17. Sun, H. G.; Ruszczycky, M. W.; Chang, W. C.; Thibodeaux, C. J.; Liu, H. W., *J. Biol. Chem.* **2012,** *287*, 4602-4608.

18. Soltero-Higgin, M.; Carlson, E. E.; Gruber, T. D.; Kiessling, L. L., *Nat. Struct. Mol. Biol.* **2004,** *11*, 539-543.

19. Beis, K.; Srikannathasan, V.; Liu, H.; Fullerton, S. W. B.; Bamford, V. A.; Sanders, D. A. R.; Whitfield, C.; McNeil, M. R.; Naismith, J. H., *J. Mol. Biol.* **2005,** *348*, 971-982.

20. Yao, X. H.; Bleile, D. W.; Yuan, Y.; Chao, J.; Sarathy, K. P.; Sanders, D. A. R.; Pinto, B. M.; O'Neill, M. A., *Proteins-Structure Function and Bioinformatics* **2009,** *74*, 972-979.

21. Chad, J. M.; Sarathy, K. P.; Gruber, T. D.; Addala, E.; Kiessling, L. L.; Sanders, D. A. R., *Biochemistry* **2007,** *46*, 6723-6732.

22. Caravano, A.; Sinay, P.; Vincent, S. P., *Bioorg. Med. Chem. Lett.* **2006,** *16*, 1123-1125.

23. Itoh, K.; Huang, Z. S.; Liu, H. W., *Org. Lett.* **2007,** *9*, 879-882.

24. Carlson, E. E.; May, J. F.; Kiessling, L. L., *Chem. Biol.* **2006,** *13*, 825-837.

25. Zhang, Q. B.; Liu, H. W., *J. Am. Chem. Soc.* **2001,** *123*, 6756-6766.

26. Yuan, Y.; Bleile, D. W.; Wen, X.; Sanders, D. A. R.; Itoh, K.; Liu, H. W.; Pinto, B. M., *J. Am. Chem. Soc.* **2008,** *130*, 3157-3168.

27. Errey, J. C.; Mann, M. C.; Fairhurst, S. A.; Hill, L.; McNeil, M. R.; Naismith, J. H.; Percy, J. M.; Whitfield, C.; Field, R. A., *Org. Biomol. Chem.* **2009,** *7*, 1009-1016.

28. Yuan, Y.; Wen, X.; Sanders, D. A. R.; Pinto, B. M., *Biochemistry* **2005,** *44*, 14080-14089.

29. Gruber, T. D.; Borrok, M. J.; Westler, W. M.; Forest, K. T.; Kiessling, L. L., *J. Mol. Biol.* **2009,** *391*, 327-340.

30. Gruber, T. D.; Westler, W. M.; Kiessling, L. L.; Forest, K. T., *Biochemistry* **2009,** *48*, 9171-9173.

31. Feng, L.; Senchenkova, S. N.; Yang, J. H.; Shashkov, A. S.; Tao, J.; Guo, H. J.; Cheng, J. S.; Ren, Y.; Knirel, Y. A.; Reeves, P. R.; Wang, L., *J. Bacteriol.* **2004,** *186*, 4510-4519.

32. Hanniffy, O. M.; Shashkov, A. S.; Moran, A. P.; Prendergast, M. M.; Senchenkova, S. N.; Knirel, Y. A.; Savage, A. V., *Carbohydr. Res.* **1999,** *319*, 124-132.

33. Arbatsky, N. P.; Shashkov, A. S.; Mamyan, S. S.; Knirel, Y. A.; Zych, K.; Sidorczyk, Z., *Carbohydr. Res.* **1998,** *310*, 85-90.

34. St Michael, F.; Szymanski, C. M.; Li, J. J.; Chan, K. H.; Khieu, N. H.; Larocque, S.; Wakarchuk, W. W.; Brisson, J. R.; Monteiro, M. A., *Eur. J. Biochem.* **2002,** *269*, 5119-5136.

35. Wang, Q.; Ding, P.; Perepelov, A. V.; Xu, Y. L.; Wang, Y.; Knirel, Y. A.; Wang, L.; Feng, L., *Mol. Microbiol.* **2008,** *70*, 1358-1367.

36. Allos, B. M., *Clin. Infect. Dis.* **2001,** *32*, 1201-1206.

37. Kaldor, J.; Speed, B. R., *Br. Med. J.* **1984,** *288*, 1867-1870.

38. Parkhill, J.; Wren, B. W.; Mungall, K.; Ketley, J. M.; Churcher, C.; Basham, D.; Chillingworth, T.; Davies, R. M.; Feltwell, T.; Holroyd, S.; Jagels, K.; Karlyshev, A. V.; Moule, S.; Pallen, M. J.; Penn, C. W.; Quail, M. A.; Rajandream, M. A.; Rutherford, K. M.; van Vliet, A. H. M.; Whitehead, S.; Barrell, B. G., *Nature* **2000,** *403*, 665-668.

39. Errey, J. C.; Mukhopadhyay, B.; Kartha, K. P. R.; Field, R. A., *Chem. Commun.* **2004**, 2706-2707.

40. Feldman, M. F.; Marolda, C. L.; Monteiro, M. A.; Perry, M. B.; Parodi, A. J.; Valvano, M. A., *J. Biol. Chem.* **1999,** *274*, 35129-35138.

41. Liu, D.; Reeves, P. R., *Microbiology-Uk* **1994,** *140*, 49-57.

42. Stevenson, G.; Neal, B.; Liu, D.; Hobbs, M.; Packer, N. H.; Batley, M.; Redmond, J. W.; Lindquist, L.; Reeves, P., *J. Bacteriol.* **1994,** *176*, 4144-4156.

43. Lee, R.; Monsey, D.; Weston, A.; Duncan, J. K.; Rithner, C.; McNeil, M., *Anal. Biochem.* **1996,** *242*, 1-7.

44. Fersht, A., *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*. W. H. Freeman and Company: New York, NY, 1999.

45. Karlyshev, A. V.; Champion, O. L.; Churcher, C.; Brisson, J. R.; Jarrell, H. C.; Gilbert, M.; Brochu, D.; St Michael, F.; Li, J. J.; Wakarchuk, W. W.; Goodhead, I.; Sanders, M.; Stevens, K.; White, B.; Parkhill, J.; Wren, B. W.; Szymanski, C. M., *Mol. Microbiol.* **2005,** *55*, 90-103.

46. Guan, S.; Clarke, A. J.; Whitfield, C., *J. Bacteriol.* **2001,** *183*, 3318-3327.

47. Bernatchez, S.; Szymanski, C. M.; Ishiyama, N.; Li, J. J.; Jarrell, H. C.; Lau, P. C.; Berghuis, A. M.; Young, N. M.; Wakarchuk, W. W., *J. Biol. Chem.* **2005,** *280*, 4792-4802.

48. Cha, S., *Mol. Pharmacol.* **1968,** *4*, 621-629.

49. Rose, N. L.; Palcic, M. M.; Evans, S. V., *J. Chem. Educ.* **2005,** *82*, 1846-1852.

50. Claus, H.; Stummeyer, K.; Batzilla, J.; Muhlenhoff, M.; Vogel, U., *Mol. Microbiol.* **2009,** *71*, 960-971.

51. de Lederkremer, R. M.; Nahmad, V. B.; Varela, O., *J. Org. Chem.* **1994,** *59*, 690-692.

52. Poulin, M. B.; Lowary, T. L., *Methods Enzymol.* **2010,** *478*, 389-411.

53. Miller, W. G.; Bates, A. H.; Horn, S. T.; Brandl, M. T.; Wachtel, M. R.; Mandrell, R. E., *Appl. Environ. Microbiol.* **2000,** *66*, 5426-5436.

54. Marolda, C. L.; Welsh, J.; Dafoe, L.; Valvano, M. A., *J. Bacteriol.* **1990,** *172*, 3590-3599.

55. Carrillo, C. D.; Taboada, E.; Nash, J. H. E.; Lanthier, P.; Kelly, J.; Lau, P. C.; Verhulp, R.; Mykytczuk, O.; Sy, J.; Findlay, W. A.; Amoako, K.; Gomis, S.; Willson, P.; Austin, J. W.; Potter, A.; Babiuk, L.; Allan, B.; Szymanski, C. M., *J. Biol. Chem.* **2004,** *279*, 20327-20338.

56. Yao, R. J.; Alm, R. A.; Trust, T. J.; Guerry, P., *Gene* **1993,** *130*, 127-130.

57. Nothaft, H.; Liu, X.; McNally, D. J.; Li, J. J.; Szymanski, C. M., *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106*, 15019-15024.

58. Bacon, D. J.; Szymanski, C. M.; Burr, D. H.; Silver, R. P.; Alm, R. A.; Guerry, P., *Mol. Microbiol.* **2001,** *40*, 769-777.

59. Tsai, C.-M.; Frasch, C. E., *Anal. Biochem.* **1982,** *119*, 115-119.

60. Marolda, C. L.; Lahiry, P.; VInes, E.; Saldias, S.; Valvano, M. A., *Methods Mol. Biol.* **2006,** *347*, 237-252.

61. Lambert, C.; Leonard, N.; De Bolle, X.; Depiereux, E., *Bioinformatics* **2002,** *18*, 1250-1256.

62. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., *J. Comput. Chem.* **1998,** *19*, 1639-1662.

63. Bian, X.-L.; Rosas-Acosta, G.; Wu, Y.-C.; Wilson, V. G., *J. Virol.* **2007,** *81*, 2899-2908.

Chapter 3

Structure and Specificity of a *Campylobacter jejuni* UNGM

Parts of this Chapter were previously published as part of:

Poulin, M. B.; Zhou, R.; Lowary, T. L., *Org. Biomol. Chem.* **2012**, *10*, 4074-4087.

Ruokun Zhou, as part of his M.Sc. thesis, synthesized the Gal*f* methyl glycosides used in this chapter. Carla Protsko, an MSc student from the lab of our collaborator Professor David A. R. Sanders (University of Saskatchewan) performed the protein crystallography of cjUNGM described in this Chapter.

3.1 Introduction

Campylobacter jejuni are commensal bacteria in many animals but they are the most common cause of diarrheal disease worldwide in humans.[1] Infection by *C. jejuni* is also regarded as the primary predetermining factor for the development of the neurological disorder Guillain–Barré syndrome (GBS).[2] Although, the mechanism of infection by *C. jejuni* is not well understood, it is believed that mimicry of human gangliosides by lipooligosaccharides (LOSs) is responsible for the development of GBS.[3, 4]

In addition to the LOSs, *C. jejuni* also produces a high molecular weight capsular polysaccharide (CPS), which is one of the only clear virulence factors of the organism. It is required for chicken colonization,[5] the adhesion and invasion of human epithelial cells, serum resistance and the maintenance of cell surface charge by the organism.[6] The CPS is the main serodeterminant for the Penner serotyping of *C. jejuni*.[7] These structures are composed of structurally diverse carbohydrates, including unusual heptoses and sugars in the furanose ring form.[8-10] Given the importance of surface carbohydrates for these bacteria, there has been an interest in understanding biosynthesis of these structurally diverse carbohydrate structures, and, specific to this chapter, the furanose sugar found in the CPS of *C. jejuni* 11168 (Figure 3-1A).

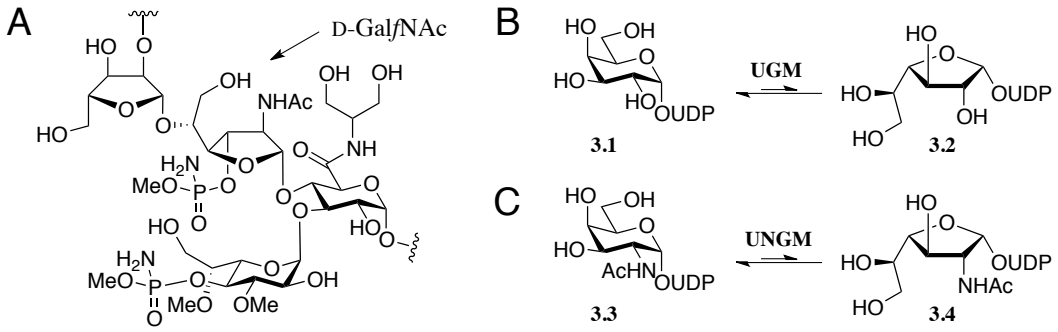


Figure 3-1. Structure of the *C. jejuni* serotype HS:2 CPS highlighting the Gal*f*NAc residue (A). The reaction of UGM for the biosynthesis of UDP-Gal*f* (B), and the reaction of UNGM for the biosynthesis of UDP-Gal*f*NAc (C) are also shown.

The UDP-galactopyranose mutase (UGM) enzymes, encoded by *glf* genes,[11] are responsible for the biosynthesis of galactofuranose (Gal*f*) sugars found in many microorganisms.[12-16] These enzymes catalyze the isomerization of UDP-galactopyranose (UDP-Gal*p*) to form UDP-Gal*f*, the precursor to Gal*f* residues found in the glycoconjugates of these organisms. My previous work (Chapter 2) showed that a homolog of the *glf* gene (*cj1439c*), encoding a UDP-*N*-acetyl-galactopyranose mutase (UNGM), is involved in the biosynthesis of the *N*-acetyl-galactofuranose (Gal*f*NAc) found in the CPS tetrasaccharide repeating unit of *C. jejuni* 11168 (Figure 3-1A).[12] UNGM is a bifunctional enzyme and is able to catalyze the isomerization of both UDP-Gal*p* and UDP-Gal*p*NAc to the corresponding furanose sugar nucleotides (Figure 3-1 B and C, respectively). Although no Gal*f* residues have been found in glycoconjugates of *C. jejuni* 11168, cjUNGM is able to complement UGM activity in an *E. coli* ΔUGM knockout strain restoring lipopolysaccharide (LPS) biosynthesis.[12] Thus cjUNGM can function *in vivo* as a UGM.

Both UGM and UNGM are flavoenzymes, in which the FAD cofactor is directly involved in catalysis. The N-5 of reduced $FADH^-$ reacts directly with the UDP-Gal(NAc) substrate to displace UDP and form a covalent iminium ion intermediate (Figure 3-2 *b*).[13, 14] The cjUNGM sequence is highly homologous to known bacterial UGMs, in particular, the UGM from *E. coli* (ecUGM, 60% identity). Despite the high sequence homology, only UNGM accepts UDP-GalNAc as a substrate. Two arginine residues in cjUNGM (R59 and R168) play a role in the recognition and turnover of UDP-GalNAc, and we hypothesized that an interaction, between R59 and the acetamido group of UDP-GalNAc, is required to prevent the formation of unproductive side products (Figure 3-2 *d*). We wanted to further test this hypothesis by examining any differences in the binding interactions between this cjUNGM and other UGMs, which do not turn over UDP-GalNAc.

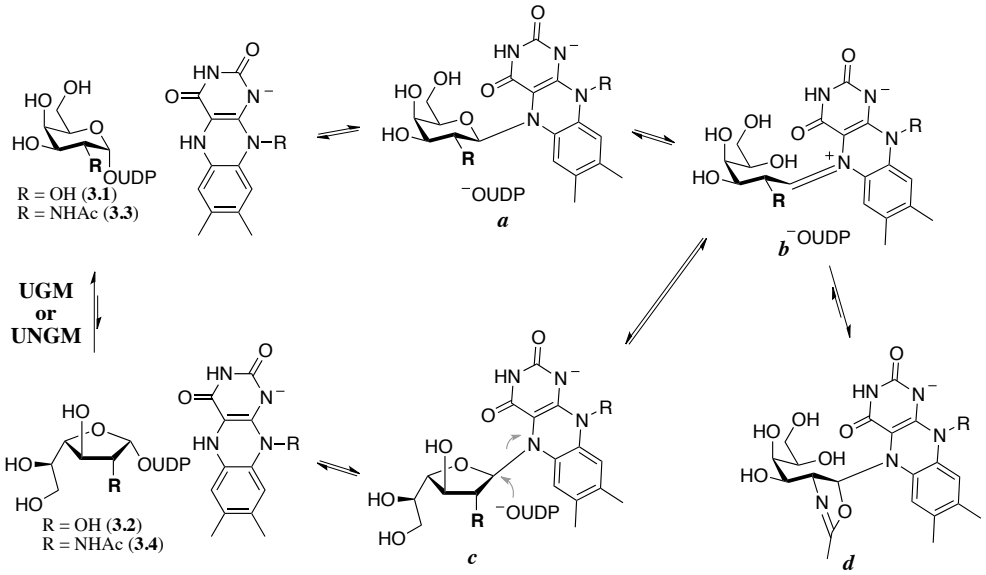


Figure 3-2. Putative mechanism for the cUNGM enzyme based on the proposed UGM mechanism. With UDP-Gal*p*NAc (**3.3**) it was hypothesized that UGMs lacking R59 (i.e., ecUGM) form an unproductive oxazoline intermediate (*d*) preventing turnover of this substrate.[12]

The use of saturation transfer difference NMR (STD-NMR) spectroscopy has proven useful to characterize protein–ligand binding interactions.[15, 16] In these experiments, protons in closer proximity to the protein amino acid residues experience a greater saturation transfer effect, which allows the binding epitope, the ligand surface interacting with the protein, to be mapped. This technique was used to examine the binding of UDP-Gal*p*[17] and UDP-Gal*f*[18] to the *Klebsiella pneumoniae* UGM (kpUGM) and could be used to predict the bound conformation of these substrates. Although protein crystallography[19-21] later showed that the predicted binding mode was incorrect, these STD-NMR spectroscopic investigations were successful in predicting many of the galactose–protein binding interactions.

The use of singly modified carbohydrate analogs has also been an effective tool to explore carbohydrate–protein binding interactions.[22-24] Measuring the activity of an enzyme with these analogs allows the evaluation of both hydrogen bonding interactions (with deoxy analogs) and steric constraints (with methoxy analogs) leading to substrate recognition.

In this chapter, we describe the synthesis of a panel of singly modified UDP-Gal*f* analogs and their use in assessing substrate-binding interactions of cjUNGM. Using these analogs, as well as STD-NMR spectroscopy, we explored whether the substrate binding interactions, between the UDP-Gal(NAc) substrates and cjUNGM and other bacterial UGM, could help explain the different substrate tolerance of these pyranose–furanose mutase enzymes.

3.2 Results and Discussion

3.2.1 No covalent intermediate forms between ecUGM and UDP-GalNAc

In the previous Chapter, I described that ecUGM is unable to turn over UDP-GalNAc whereas the highly homologous cjUNGM does. I attributed the difference in activity partially to the presence of an arginine in the cjUNGM active site (R59), which interacts with the GalNAc acetamido moeity. It is possible that the ecUGM, which contains instead a histidine (H59) at this position, could become trapped after formation of the iminium ion (Figure 3-2 *b*) and form an unproductive oxazoline (Figure 3-2 *d*) or, alternatively, the ecUGM may not bind to UDP-GalNAc at all. To distinguish between such possibilities we sought to directly observe the formation of any of these species.

Flavin derivatives often exhibit characteristic UV–visible absorbances, which can be used to identify species that are produced during an enzymatic reaction. For example, it was previously shown that the covalent iminium intermediate of the UGM reaction (Figure 3-2 *b*) gives rise to an increased absorbance at 500 nm with an isosbestic point around 475 nm.[14] We hypothesized that a similar covalent intermediate would exist en route to the proposed unproductive oxazoline intermediate (Figure 3-2 *d*) with ecUGM and UDP-Gal*p*NAc (**3.3**). Therefore, we sought to detect such an iminium ion intermediate from the absorbance spectrum of the ecUGM in the presence of UDP-Gal*p*NAc. First, we examined the absorbance of ecUGM with increasing amounts of UDP-Gal*p* (**3.1**). Similar to published results with the kpUGM,[14, 20] we see an increase in the absorbance at 500 nm corresponding to the

concentration of UDP-Gal*p* added to the ecUGM (Figure 3-3 A). When UDP-Gal*p*NAc instead was added, there was no change in the absorbance of the sample. As seen in Figure 3-3 B, the constant FAD absorbance with increasing concentrations of UDP-Gal*p*NAc suggests there is no iminium ion species formed between the ecUGM FAD co-factor and this substrate, at least under the conditions tested. These absorbance measurements provide further evidence that ecUGM does not recognize UDP-Gal*p*NAc as a substrate.[12]

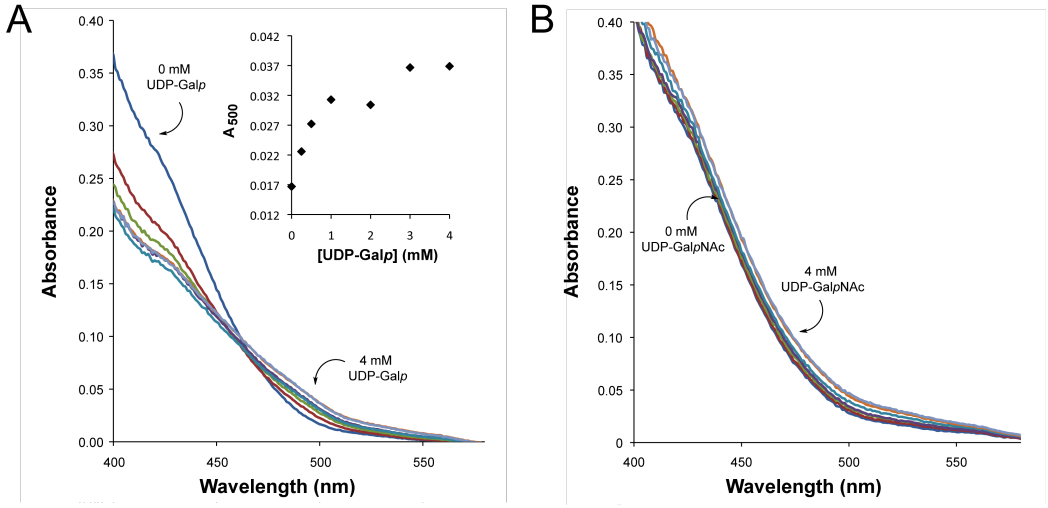


Figure 3-3. UV–visible spectrum for titrations of ecUGM. Concentrated UDP-Gal*p* (A) or UDP-Gal*p*NAc (B) was titrated into reduced ecUGM. The formation of a covalent adduct between the FAD co-factor and the substrate can be seen by the increase in A_{500} with increasing UDP-Gal*p* (A, inset) reaching a maximum at ~3 mM. Titrating with UDP-Gal*p*NAc showed no ΔA_{500}.

3.2.2 Differentiating the binding epitopes for ecUGM and cjUNGM with UDP-Gal*p* and UDP-Gal*p*NAc by STD-NMR spectroscopy

The *C. jejuni* UNGM catalyzes the isomerization of both UDP-Gal*p* (**3.1**) and UDP-Gal*p*NAc (**3.2**) into the corresponding furanose sugar nucleotides,

whereas the highly homologous *E. coli* UGM can isomerize only UDP-Gal*p*. The UV–visible spectroscopic data provided in the previous section suggest no reaction occurs between the ecUGM and UDP-Gal*p*NAc. However, we wished to probe this further by determining any differences in binding interactions between the ecUGM and cjUNGM with both of these substrates. Therefore, saturation transfer difference NMR (STD-NMR) experiments were performed to map the binding epitopes of both UDP-Gal*p* and UDP-Gal*p*NAc with the ecUGM and cjUNGM enzymes. Because the active site amino acids are largely conserved between these enzymes, any difference in the saturation transfer for a given substrate between the two enzymes is likely due to differences in the bound substrate conformation.

*3.2.2.1 Mapping the UDP-Gal***p** *binding epitope for ecUGM and cjUNGM*

Significant saturation transfer was observed in the STD-NMR spectrum of the UDP-Gal*p* protons with both the ecUGM and cjUNGM enzymes (Figure 3-4). As with previous studies with kpUGM,[17] the overlapping H-1′/H-5 protons showed the largest STD effect, and so the intensity of this peak was used as a reference to measure the relative saturation transfer of the other protons. The ribose protons, in general, showed higher STD effects with both enzymes compared to the protons of the Gal*p* moiety, consistent with previous reports on kpUGM.[17]

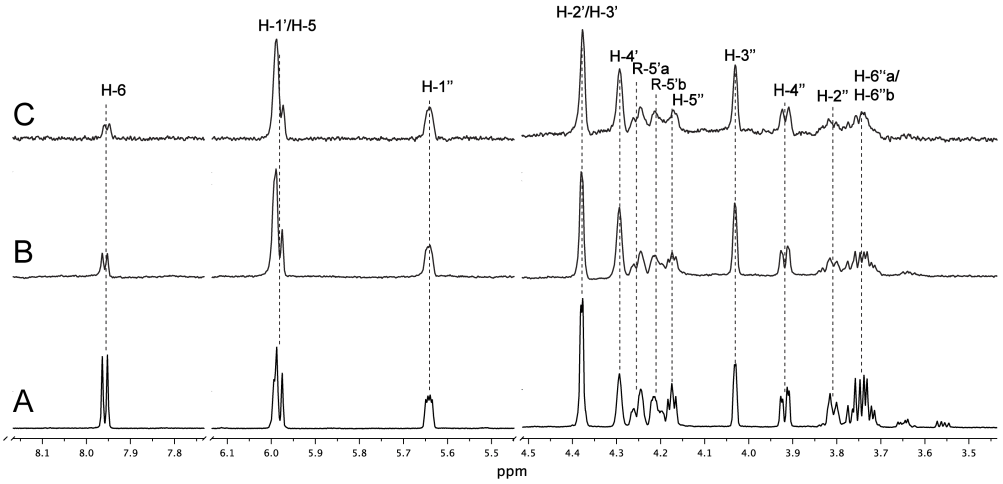


Figure 3-4. Expansions of the reference ^{1}H NMR spectrum for UDP-Gal*p* (A) and the STD-NMR spectra recorded in the presence of ecUGM (B) or cjUNGM (C). Resonance assignments are shown above the peaks. The protons of the uracil are indicated with no prime, those on the ribose with a single prime (′) and those on the Galp with a double prime (″).

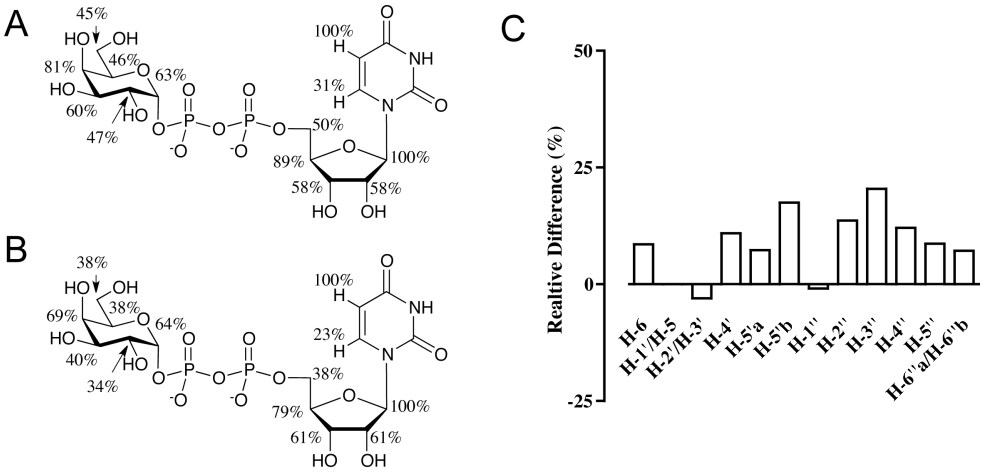


Figure 3-5. Epitope mapping of UDP-Gal*p* in the binding site of ecUGM (A) or cjUNGM (B). The differences in the relative saturation transfer were calculated by subtracting the relative intensity from the cjUNGM STD spectra from that of the ecUGM (C). The relative STD intensities are shown with the H-1′/H-5 set to 100%. For (A) and (B) the average intensity is shown for H-5′a/H-5′b. A positive relative difference indicates that the substrate binds more strongly to ecUGM than to cjUGNM; a negative relative difference indicates the reverse.

Nearly all the galactose and ribose protons showed lower relative intensities in the spectrum for cjUNGM when compared to the ecUGM spectrum (Figure 3-5). Most notably, the relative intensity of the H-2′′, H-3′′ and H-5′b protons was 14–20% lower with cjUNGM then with ecUGM. Despite these differences, both enzymes showed the same general trend with respect to the relative intensity of the protons. Because nearly all protons show lower intensity in the cjUNGM spectrum, this data suggests that UDP-Gal*f* forms weaker interactions with the enzyme active site than those observed with the ecUGM. These observations are not surprising as the biologically relevant substrate for cjUNGM in *C. jejuni* 11168 is UDP-Gal*p*NAc and not UDP-Gal*p*, which is the natural substrate for ecUGM.[11]

*3.2.2.2 Differences in the UDP-Gal*p*NAc binding epitope for ecUGM and cjUNGM*

Both the ecUGM and cjUNGM showed signals in the STD spectra recorded in the presence of UDP-Gal*p*NAc (Figure 3-6). This observation demonstrates that UDP-Gal*p*NAc binds to both ecUGM and cjUNGM, despite it not being a substrate for the former enzyme. As with UDP-Gal*p*, the largest STD effects were observed for the H-1′/H-5 protons of UDP-Gal*p*NAc, which were again used as a reference to calculate the relative saturation transfer to each proton. In addition to the UDP-Gal*p*NAc protons, significant STD effects were seen for the product UDP-Gal*f*NAc protons (indicated with an asterisk in Figure 3-6) in the cjUNGM STD spectrum. The absence of these signals in the ecUGM

STD spectrum provides further evidence that UDP-Gal*p*NAc is not a substrate for this enzyme. As these protons are only observed for cjUNGM, they were not used in the comparison of relative saturation transfer intensities of each proton. This may, however, lead to an underestimate of the relative STD effects for cjUNGM with UDP-Gal*p*NAc.

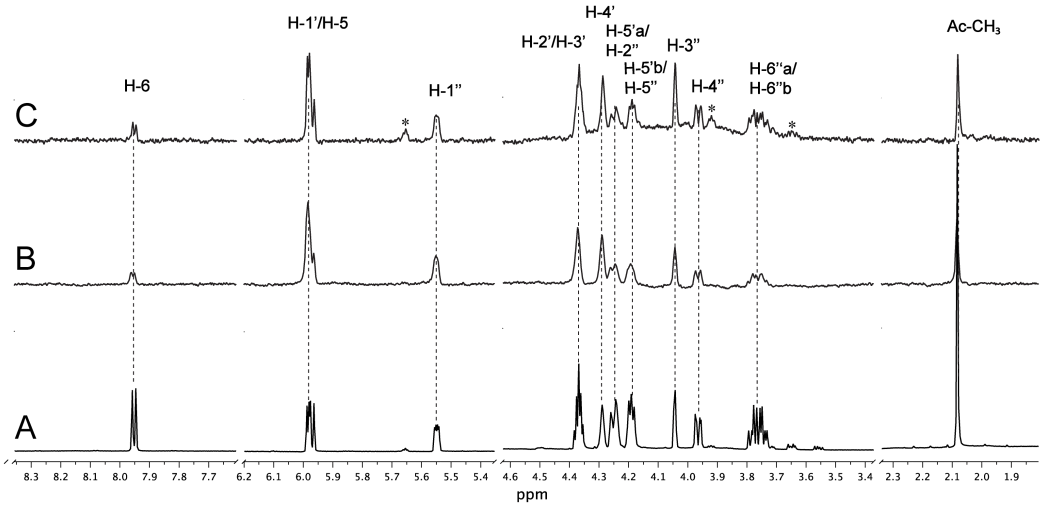


Figure 3-6. Expansions of the reference ^{1}H NMR spectra for UDP-Gal*p*NAc (A) and the STD-NMR spectra recorded in the presence of either ecUGM (B) or cjUNGM (C). Resonance assignments are above the peaks. An asterisk indicates the protons resulting from UDP-Gal*f*NAc. The protons of the uracil are indicated with no prime, those on the ribose with a single prime (′) and those on the Galp with a double prime (′′).

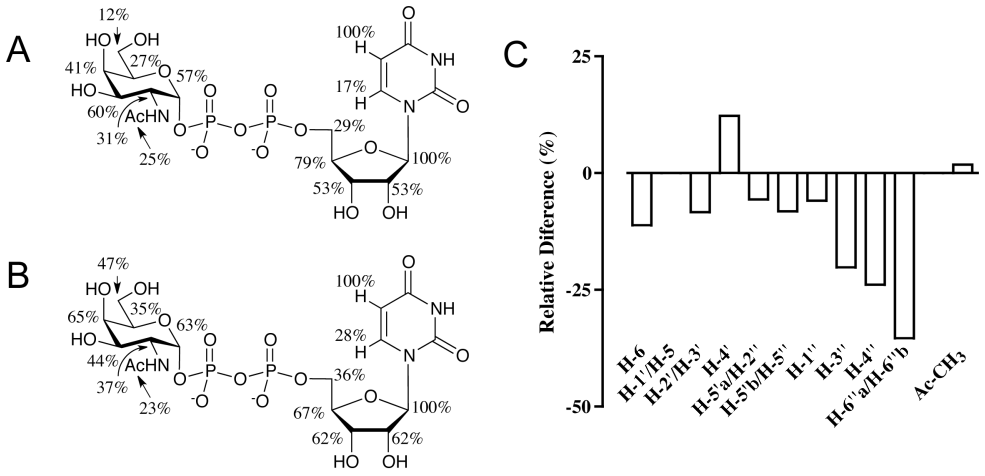


Figure 3-7. Epitope mapping of UDP-Gal*p*NAc in the binding site of ecUGM (A) or cjUNGM (B). The difference in the relative saturation transfer for each proton was calculated by subtracting the relative intensity from the cjUNGM STD spectra from that of the ecUGM (C). The relative STD intensities are shown with H-1′/H-5 set to 100%. In (A) and (B) the H-5′a/H-5′b average intensity is shown. A positive relative difference indicates that the substrate binds more strongly to ecUGM than to cjUGNM; a negative relative difference indicates the reverse.

In contrast to the observations with UDP-Gal*p*, the STD effects for UDP-Gal*p*NAc with the ecUGM were, in general, lower when compared to same effects with the cjUNGM. The largest difference was observed in the relative intensity of the Gal*p*NAc H-3″, H-4″, and H-6″a/H-6″b protons, which are 20–35% lower in intensity in the ecUGM spectrum compared to the cjUNGM spectrum. The low saturation transfer observed with ecUGM suggests that the Gal*p*NAc portion of the ligand is not in close contact to the enzyme active site amino acids. UGM enzymes contain a mobile loop region that closes upon substrate binding.[25] It is possible that binding of UDP-Gal*p*NAc to the ecUGM interferes with this loop movement minimizing the protein contacts with the

substrate, similar to the observed binding mode of UDP-Glc*p* with kpUGM[20] discussed in Chapter 1. However, further tests are required to test this hypothesis, which could explain why no activity is observed for ecUGM with UDP-Gal*p*NAc.

3.2.3 Chemo-enzymatic synthesis of UDP-Gal*f* analogs

There were observable differences in the STD effects seen for the ecUGM and cjUNGM with both substrates tested, which suggests there are different enzyme–substrate interactions in the active site of the two enzymes. Only UDP-Gal*f* is a substrate for both ecUGM and cjUNGM enzymes. Therefore, to explore the relative importance of the enzyme–substrate interactions for these enzymes, we sought to prepare a panel of singly deoxygenated and methylated UDP-Gal*f* analogs (**3.5**–**3.12**, Figure 3-8) and use them to probe the substrate specificity of the UGM from *E. coli* and *K. pneumoniae* compared to the *C. jejuni* UNGM.

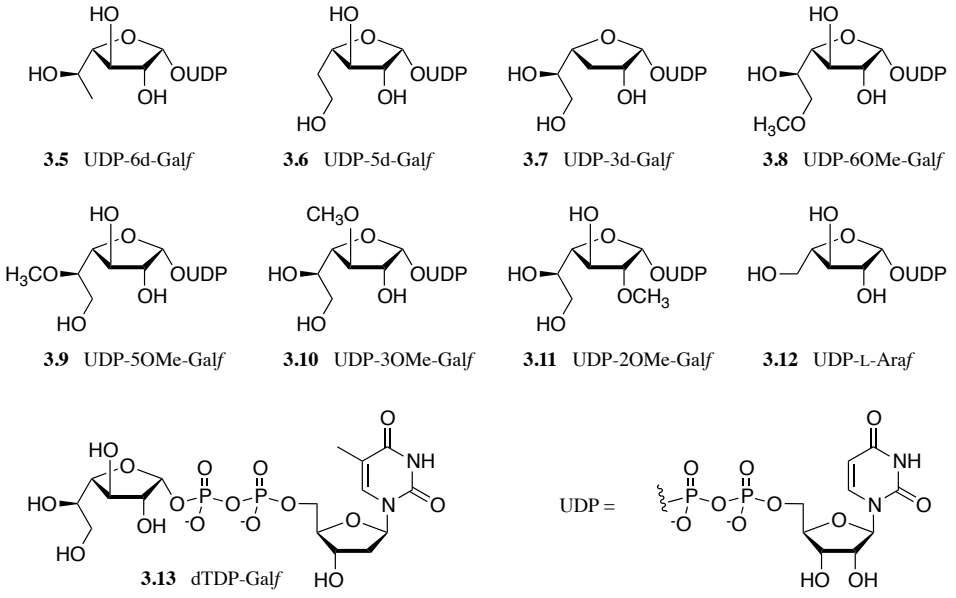
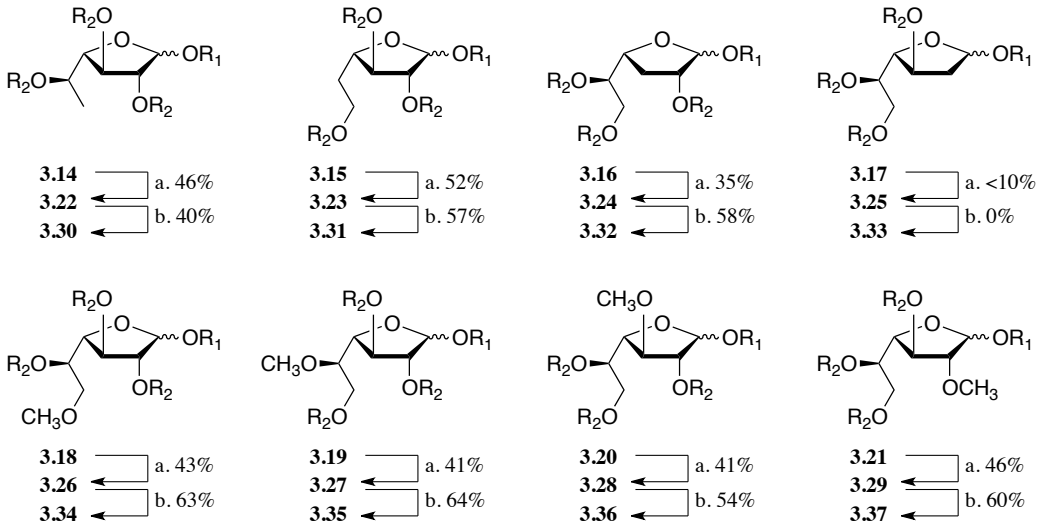


Figure 3-8. UDP-Gal*f* analogs **3.5**–**3.12** targeted for synthesis.

To prepare **3.5–3.12**, we employed a chemo-enzymatic approach, which has been previously used for the preparation of UDP-Gal*f* (**3.2**).[26-28] This method uses a three enzyme system to convert galactofuranose-1-phosphate (Gal*f*-1P) to UDP-Gal*f*. It was exploited here for the preparation of UDP-Gal*f* analogs **3.5–3.12**. To apply this approach, the Gal*f*-1P analogs **3.30–3.37** (Scheme 3-1) were first prepared from the corresponding methyl glycosides **3.14**–**3.21**, which had been previously synthesized by a former M.Sc. student in the group, Mr. Ruokun Zhou.

As illustrated in Scheme 3-1, treatment of methyl Gal*f* analogs **3.14**–**3.21** with HBr in acetic acid resulted in formation of the corresponding glycosyl bromide. Subsequent reaction of each of these bromides with dibenzyl phosphate provided the α-galactofuranosyl phosphates.[29] Unfortunately, treatment of the 2-deoxy methyl glycoside **3.17** under the above conditions resulted predominantly in the hydrolysis product and produced only small (<10%) amounts of the desired phosphate **3.25**. Furthermore, attempts to deprotect this dibenzyl phosphate intermediate resulted in further hydrolysis giving none of the desired compound **3.33**. These results are not surprising given the known lability of 2-deoxy glycosyl halides to hydrolysis. This lability is further exacerbated by the fact that these are furanose derivatives, which are more prone to hydrolysis that their six-membered ring counterparts.[30] As a result, the synthesis of the 2-deoxy analog **3.33** was abandoned. For the other analogs, the protected α-Gal*f*-1P derivatives **3.22**–**3.24** and **3.26**–**3.29** were obtained in modest (35–52%) yields. The H-1′′ to H-2′′ proton coupling constants were between 4.0–4.5 Hz for all the

products, consistent with the α-Gal*f* geometry. For all of the successful phosphorylation reactions a small amount (<5%) of the β-Gal*f*-1P was also detected in the ^{1}H NMR spectra of the isolated products. Global deprotection, first by hydrogenolysis of the benzyl groups and then removal of the benzoate esters under weakly basic conditions, afforded a 40–64% yield of the deprotected Gal*f*-1P analogs **3.30**–**3.32**, **3.34**–**3.37**.

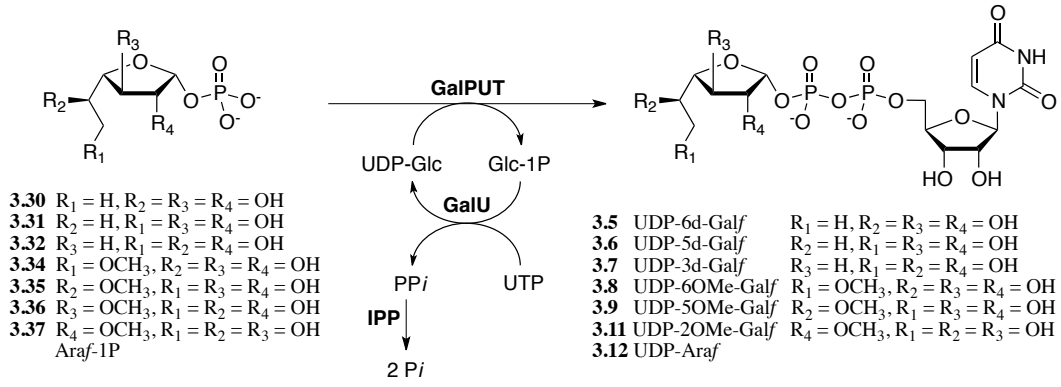


3.14 – 3.21 $R_1 = CH_3$, R_2 = Bz **3.22 – 3.29** R_1 = α-$PO(OBn)_2$, R_2 = Bz **3.30– 3.37** R_1 = α-PO_3^{2-}, R_2 = H

Scheme 3-1. Synthesis of Gal*f* 1-phosphate analogs. a. 33% HBr in AcOH, CH_2Cl_2; $HOPO(OBn)_2$, Et_3N, toluene; b. H_2, Pd–C, Et_3N, EtOAc; MeOH, H_2O, Et_3N.

We previously employed this chemo-enzymatic approach for the preparation of UDP-Gal*f* (**3.1**, Scheme 3-2).[27] The method employs a promiscuous galactose-1-phosphate uridylyltransferase (GalPUT) that has been shown to convert a variety of hexose-1-phosphate analogs, including hexose sugars in the furanose ring form, into the corresponding UDP-sugars.[30, 34, 35] This method proceeds efficiently and in high yield for a wide range of substrates,

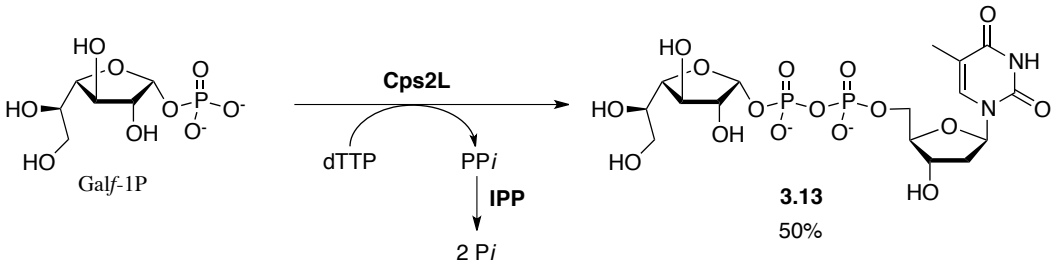
offering advantages over the entirely chemical approaches[31-34] that have been used to prepare UDP-Gal*f* and analogs thereof. In this chapter, we used this chemo-enzymatic method to produce **3.5**–**3.10** from **3.30**–**3.32**, **3.34**–**3.37**.[28] A limitation of this strategy is the ability of GalPUT to recognize and turn over the modified Gal*f*-1P analogs. The deoxy Gal*f*-1P analogs were well tolerated, yielding 35–78% of the corresponding UDP-Gal*f* derivatives. However, the methylated Gal*f*-1P analogs were very poor substrates providing <5% isolated yields of the product, and in the case of 3-OMe analog **3.36** insufficient product was produced for characterization. The method was also used to prepare UDP-Ara*f* (**3.12**) as previously described.[35]



Scheme 3-2. Chemo-enzymatic synthesis of UDP-Gal*f* analogs using GalPUT (galactose-1-phosphate uridylyltranferase), GalU (glucose-1-phosphate uridylyltransferase) and IPP (inorganic pyrophosphatase).

To prepare TDP-Gal*f* (**3.13**), we employed a different chemo-enzymatic approach, which uses a bacterial α-D-glucopyranosyl-1-phosphate thymidylyltransferase (Cps2L) to convert sugar-1-phosphates into TDP-sugars using deoxythymidine 5′-triphosphate (dTTP, Scheme 3-3).[36] Jakeman and

coworkers demonstrated Cps2L from *Streptococcus pneumoniae* tolerates a broad range of sugar-1-phosphate substrates,[36] including Gal*f*-1P and other furanosyl-1-phosphates.[37] To prepare milligram quantities of TDP-Gal*f*, we modified the reported procedure by immobilizing Cps2L (obtained from a clone provided by Professor Jakeman) on Ni-NTA agarose resin, as was done with GalPUT. This helped facilitate the product purification and, using this approach, **3.13** was obtained in 50% overall isolated yield from Gal*f*-1P.



Scheme 3-3. Chemo-enzymatic synthesis of TDP-Gal*f* using Cps2L.

3.2.4 Probing UGM (UNGM) activity with UDP-Gal*f* analogs 3.5–3.13.

3.2.4.1 Relative activity of UGM and UNGM enzymes with C-2′′, C-3′′ and C-6′′ modified analogs

With the panel of substrates in hand, each was tested with three different enzymes: ecUGM, kpUGM and cjUNGM. In cases where the analog was shown to be a substrate, the structure of the product was confirmed by ^{1}H NMR spectroscopy (see section 3.2.4.2 below). Each of the mutase enzymes examined in this study showed varying degrees of activity with the deoxy UDP-Gal*f* analogs and UDP-Ara*f* (**3.12**). For both the ecUGM and kpUGM, the 6′′-deoxy

derivative **3.5** showed the highest relative activity out of all analogs tested (24% and 19%, respectively compared to UDP-Gal*f*). UDP-Ara*f* (**3.12**) showed the next highest activity with both UGM enzymes (21% for ecUGM and 12% for kpUGM). Both of these substrates lack a 6′′-hydroxyl group, suggesting that no critical interactions occur with this hydroxyl group. The available crystal structures[19, 21] show only a single hydrogen bond present to the 6′′-OH of the UDP-D-Gal substrate, which, according to our measurements, does not appear to be critical for activity. This is also consistent with the STD-NMR measurements that showed lower relative STD affects with the H-6′′ protons.

Both UGMs also showed moderate activity with the 3′′-deoxy derivative **3.7**, but this activity was lower then with the C-6′′ modified analogs **3.5** and **3.12**. In contrast, the cjUNGM showed the same activity with both the 3′′-deoxy derivative **3.7** and 6′′-deoxy derivative **3.5** (~11%). These results suggests that interactions with the 3′′-hydroxyl group are less important for the UNGM enzyme compared to UGM. As mentioned above, STD NMR also showed a large difference in saturation transfer between the ecUGM and cjUNGM enzymes was for the UDP-Gal*p* H-3′′ consistent with the observations with these UDP-Gal*f* analogs. Water mediated hydrogen bonds to H-3′′ are observed with H63 in the kpUGM[19] crystal structure. This interaction is not possible with cjUNGM where the corresponding amino acid at this position is an arginine (R59), which could explain this difference in relative activity.

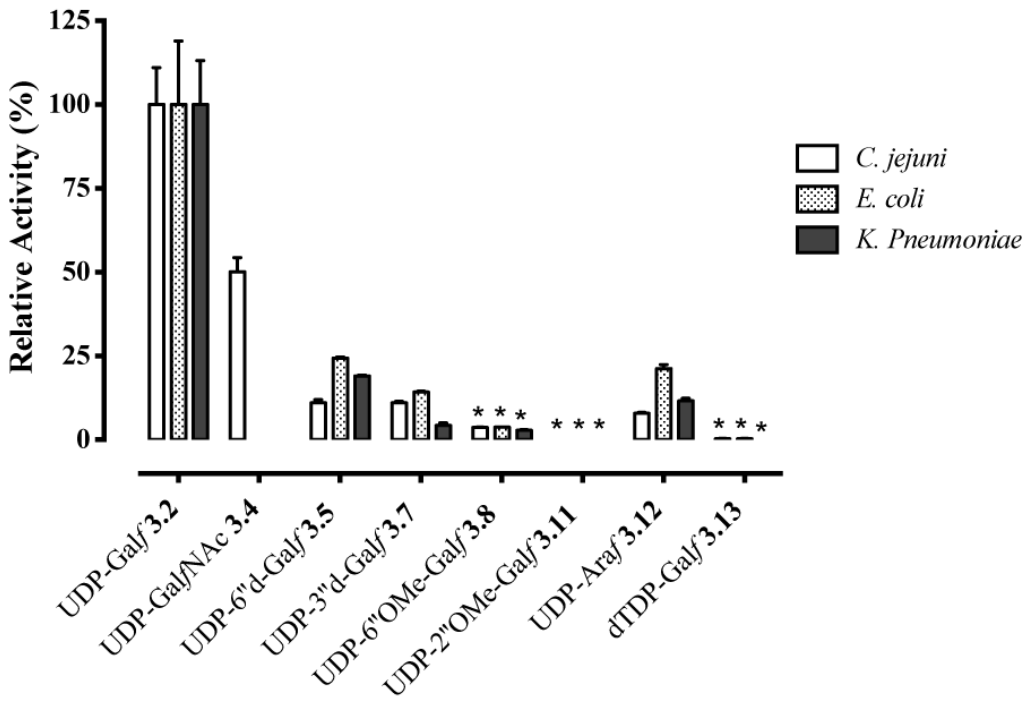


Figure 3-9. Relative activity of the ecUGM, kpUGM, and cjUNGM with UDP-Gal*f* analogs (**3.5**, **3.7**, **3.8**, and **3.11**–**3.13**). The activity for each enzyme with UDP-Gal*f* (**3.2**) was set to 100%. An asterisk indicates reactions where the structure of the product could not be confirmed. In all cases, the data for each compound is presented in the graph from left to right in the order *C. jejuni*, *E. coli* and *K. pneumoniae*.

The enzymes tested showed little to no observable activity with the C-2′′ and C-6′′-methoxy UDP-Gal*f* analogs (Figure 3-9). It appears that the enzyme will not tolerate any steric bulk at these positions, consistent with x-ray crystal structures that shown protein–substrate interactions with nearly all the Gal hydroxyl groups.[19, 21] Small levels of activity were seen with all three enzymes using 6′′-methoxy analog **3.8** as a substrate, although none of the enzymes showed sufficient turnover of this substrate to allow isolation and structural confirmation of the product. More interesting is that none of the enzymes,

including cjUNGM, showed activity with the 2′′-methoxy analog **3.11**. Use of this analog shows that the presence of a bulky substituent at C-2′′ of the Gal*f* ring abrogates turnover by these enzymes. When considering that cjUNGM recognizes UDP-GalNAc, which possesses a similarly bulky acetamido substituent (A-value ~1.6 for NHAc[38] vs. 0.79 for OCH_3[39]) at C-2′′, as its biologically relevant substrate, the result with **3.11** is somewhat paradoxical. This shows recognition of the UDP-Gal*p*NAc by cjUNGM is not controlled by steric factors, but rather, the recognition must due to a specific interaction with the 2′′-acetamido group.

3.2.4.2 Isolation and characterization of reaction products

It was reported the kpUGM showed no activity with UDP-6′′-deoxy-Gal*p* as the substrate;[40] however, we observed approximately 20% activity when using UDP-6′′-deoxy-Gal*f* (**3.5**), the isomer in which the galactose ring is in the furanose ring form. Because of this discrepancy in the reported activity, we wanted to confirm the structure of the products from these analogs. To do this each reaction was allowed to proceed to equilibrium and the product peaks were isolated by HPLC and the structure was analyzed by ^{1}H NMR spectroscopy.

As seen in Figure 3-10 A, UDP-6′′-deoxy-Gal*f* shows a characteristic doublet of doublets splitting pattern ($^3J_{H,H}$ = 6.8 and 6.9 Hz) for the H-4′′ proton, whereas the product isolated from this reaction (Figure 3-10 B) shows a broad doublet for H-4′′, characteristic for the galactopyranose configuration. The remaining proton resonances also match those expected, based on standard first-

order analysis, for UDP-6′′-deoxy-Gal*p*, confirming the structure of the reaction product. The fact that this activity was not observed previously[40] using UDP-6′′-deoxy-Gal*p* as the substrate is likely related to the enzymatic equilibrium, which largely favors the pyranose ring form. Looking at the reaction in the reverse direction, starting with the furanose sugar nucleotide, allows for easier detection of enzyme activity, in particular with substrates having lower activity. The structure of the products from reactions with **3.7** and **3.12**, which showed sufficient turnover, were also confirmed using this strategy (See Experimental Section).

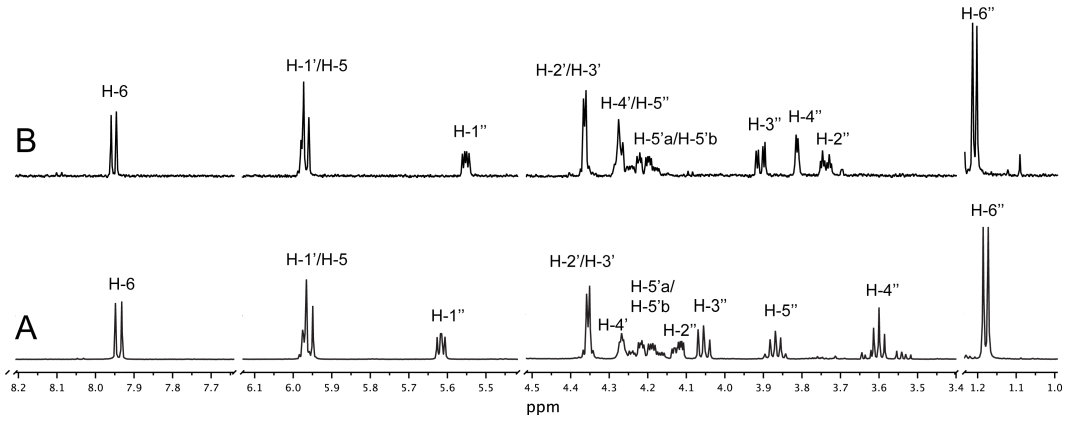


Figure 3-10. Characterization of the kpUGM reaction product with **3.5**. A ^{1}H NMR reference spectrum for UDP-6′′-deoxy-Gal*f* (**3.5**) (A) and the ^{1}H NMR spectrum for the UDP-6′′-deoxy-Gal*p* product (B).

3.2.4.3 Activity with TDP-Galf, and UNGM catalyzed nucleotide exchange

When the activity of the UGM and UNGM enzymes with synthetic TDP-Gal*f* (**3.13**) was tested only small amounts of the expected TDP-Gal*p* product

could be observed. Instead, a substantial amount of TDP was observed, likely due to hydrolysis of **3.13** during the reaction to give TDP and galactose. This hydrolysis was only observed in the presence of the UGM or UNGM enzymes, and not in reaction mixtures treated under the same conditions but lacking the enzyme.

Given the currently accepted UGM mechanism (Figure 3-2), the first step of the reaction involves cleavage of the anomeric sugar nucleotide linkage to form a covalent FAD–galactose iminium ion intermediate. Reacting with TDP-Gal*f* would also lead to this intermidiate; however, after formation of the iminium ion intermediate, the TDP could serve as a poor ligand for the protein and would then diffuse from the active site preventing further reaction leading to the TDP-Gal*p* product. It is likely that the loss of hydrogen bonding due to the deoxygenation of the ribose ring, and the additional methyl group on the thymidine moiety of TDP disrupt its binding to the UGM or UNGM active site. Thus, an activated FAD–galactose intermediate would be left in the enzyme active site.

. If a covalent FAD–galactose iminium ion intermediate is formed in this reaction of cjUNGM with **3.13**, then adding UDP to the reaction may allow for diffusion of UDP into the active site resulting in the formation of UDP-Gal*p* (**3.1**). As seen in Figure 3-11, both UDP-Gal*p* (**3.1**) and UDP-Gal*f* (**3.2**) are formed in the reaction of cjUNGM with **3.13** containing UDP.

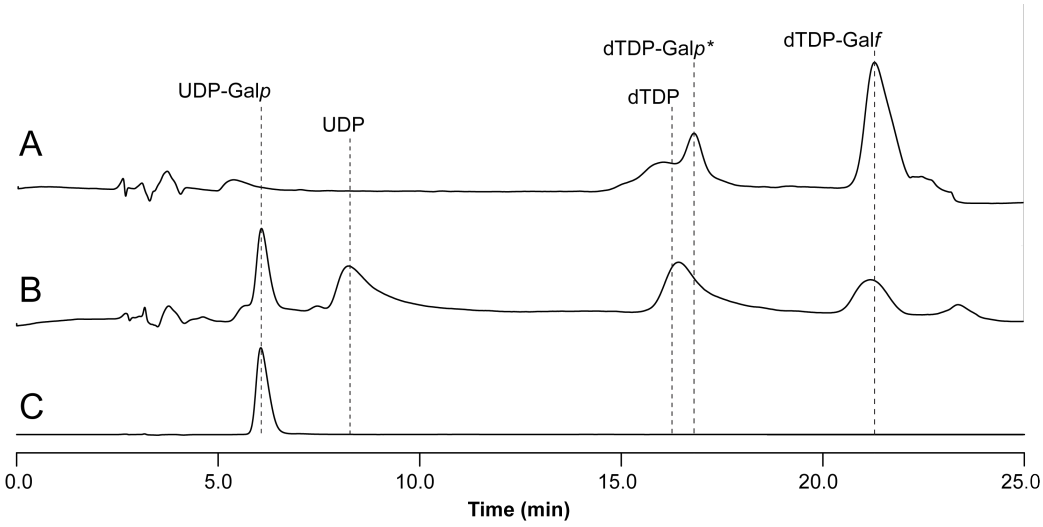


Figure 3-11. Nucleotide exchange catalyzed by cjUNGM with **3.13** and UDP. The cjUNGM reaction with **3.13** in the absence of UDP shows formation of TDP along with the putative TDP-Gal*p* product (A). In the presence of UDP none of the TDP-Gal*p* peak was observed (B) and instead a new peak was observed with the same retention time as an authentic sample of UDP-Gal*p* (**3.1**, C). An asterisk indicates the structure could not be verified. The exact retention time of TDP varies between runs as seen above.

3.2.4.4 Inhibition of UGM and UNGM with C-5 modified substrate analogs

The UDP-Gal*f* analogs **3.6** and **3.9**, which are modified at C-5″ by methylation or deoxygenation, cannot be turned over by the UGM or UNGM enzymes because they lack a free 5″-hydroxyl group required for formation of a pyranose ring. To map the protein–carbohydrate interactions at this position the inhibitory activity of **3.6** and **3.9** for the ecUGM, kpUGM, and cjUNGM were mapped. The 5″-methoxy analog **3.9** showed less then 25% inhibition of all three enzymes at all the concentrations tested (Figure 3-12). The methyl substituent on the 5″-hydroxyl group likely interferes with binding to the

enzyme in a way that is consistent with the lack of activity observed for 2′′-methoxy and 6′′-methoxy analogs, **3.8** and **3.11**. Conversely, the 5′′-deoxy-UDP-Gal*f* analog **3.6** served as a modest inhibitor of ecUGM, kpUGM, and cjUNMG with IC_{50} values of 36 ± 2.4 μM, 42 ± 3.6 μM, and 14 ± 1.2 μM, respectively. The modest activity of this compound as an inhibitor suggests any protein–substrate interactions with the 5′′-hydroxyl group are not essential for substrate recognition as the 5′′-hydroxyl group must be free to react to form the UDP-Gal*p* product.

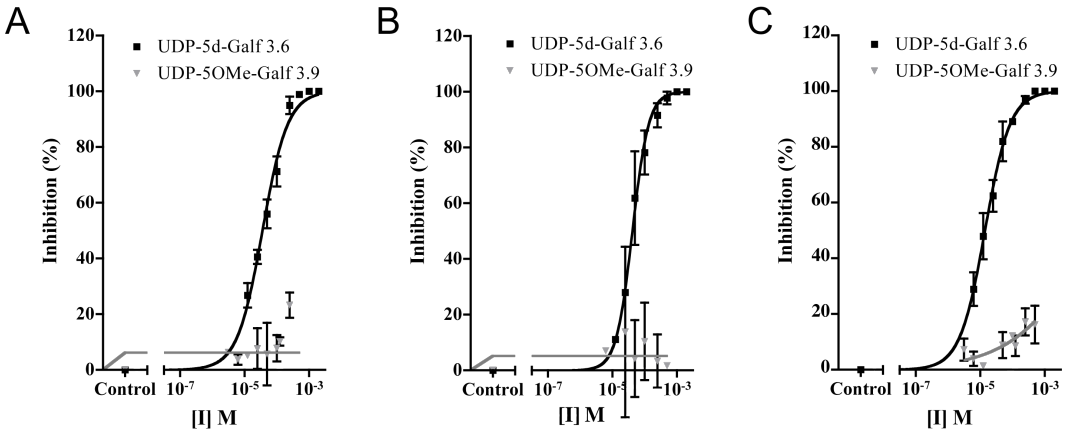


Figure 3-12. Inhibition of ecUGM (A), kpUGM (B), and cjUNGM (C) with UDP-Gal*f* analogs modified at C-5′′. Less then 20% inhibition was observed with UDP-5′′OMe-Gal*f* at all concentrations tested.

3.2.5 Structure of UNGM

In the process of studying the specificity of the cjUNGM, we were able to obtain a crystal structure of the enzyme in collaboration with Professor David A. R. Sanders at the University of Saskatchewan. Like the bacterial UGM

enzymes,[19-21, 41, 42] with which is it highly homologous, the structure of cjUNGM is a homodimer (Figure 3-13A). The obtained crystal structure, of cjUNGM, is in the oxidized form, with the mobile loop region (shown in Figure 3-13) of both monomers in the open, inactive, conformation. To date, no structure has been obstained of the cjUNGM in complex with either UDP-Gal*p* (**3.1**) or UDP-Gal*p*NAc (**3.3**), or the 5′′-deoxy-UDP-Gal*f* analog **3.6**. To obtain a better understanding of active site interactions, the structures of cjUNGM, and the oxidized ecUGM,[41] were compared to that of reduced kpUGM crystallized in complex with UDP-Gal*p* (Figure 3-14).[19]

In the kpUGM structure only a single amino acid (N84 in kpUGM) interacts with the 6′′-hydroxyl group of **3.1** (Figure 3-14B). This residue is conserved in both the ecUGM and cjUNGM (N80 in both) and is likely involved in the same interaction. Both the ecUGM and kpUGM showed similar activity with 6′′ modified substrate analogs **3.5** (6′′-deoxy), **3.8** (6′′-methoxy) and UDP-Ara*f* **3.12**, which demonstrates that this interaction is not critical for turnover.

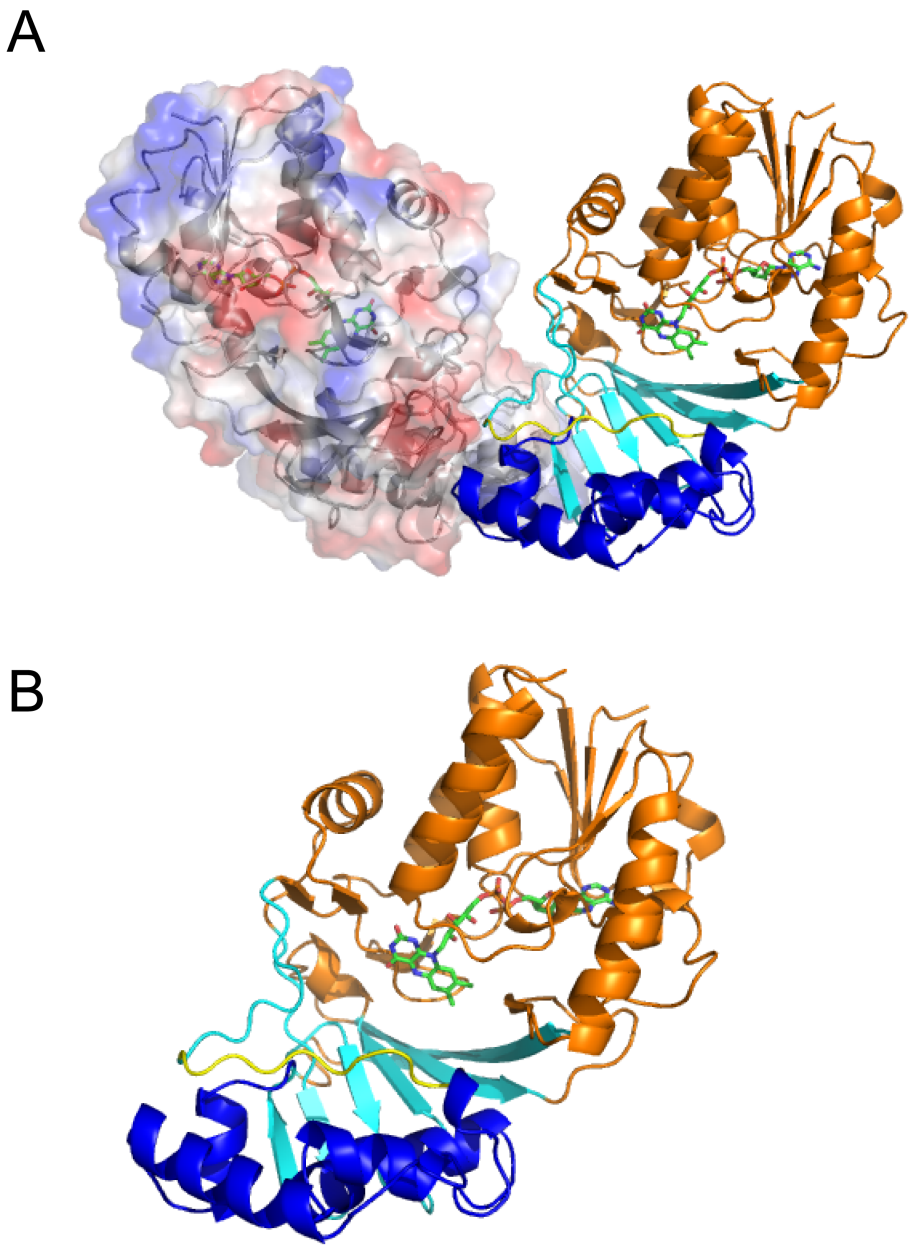


Figure 3-13. Crystal structure of cjUNGM with oxidized FAD. (A) Monomer A is shown in ribbon diagram coloured orange (domain 1), blue (domain 2) and cyan (domain 3) and the oxidized FAD co-factor is shown in green. The mobile loop region is coloured yellow. Monomer B is shown in surface representation with positively-charged regions coloured blue and negatively-charged regions coloured red. (B) Ribbon diagram of monomer A expanded.

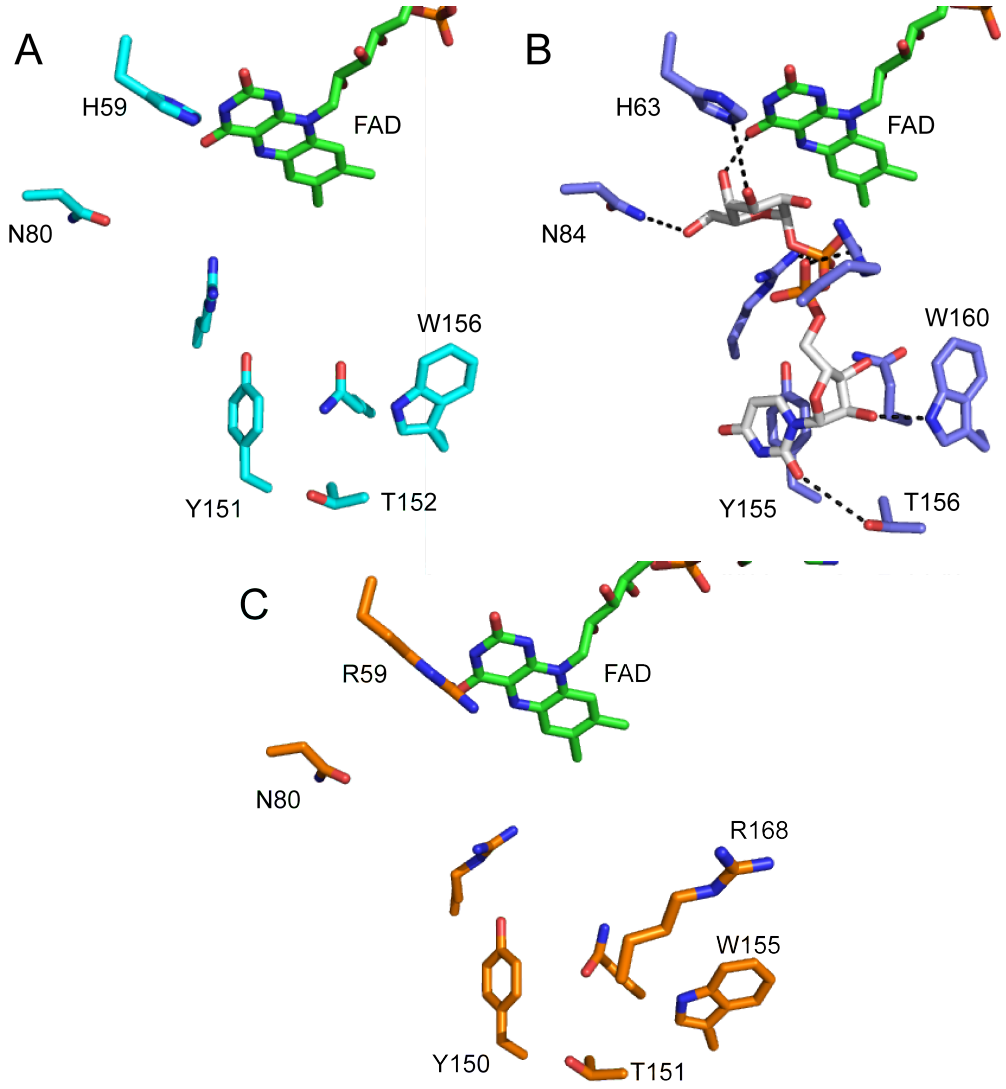


Figure 3-14. The active site residues are highly conserved between ecUGM, kpUGM, and cjUNGM. (A) A stick diagram of the ecUGM (PDB id: 1I8T_chain A) showing some of the conserved active site residues involved in contacts with the UDP-Gal*p* (**3.1**) substrate. (B) The same residues shown for the kpUGM (PDB id: 3INT_chain B) in complex with **3.1** (shown in grey). Key hydrogen bonding interactions are shown in black. (C) The active-site residues of cjUNGM. R59 is positioned to interact with the acetamido group of UDP-Gal*p*NAc (**3.3**).

Only arginine 59 (R59) of cjUNGM differs from the conserved active site amino acids of the kpUGM and ecUGM where the corresponding amino acid is a histidine (H63 and H59, respectively). This arginine is implicated in the recognition and turnover of UDP-Gal*f*NAc (**3.3**), and in the structure of cjUNGM it is positioned to interact with the substrate acetamido group. The corresponding amino acid, H63 of kpUGM (H59 in ecUGM) binds to the 3′′-hydroxyl group of UDP-Gal*p* (**3.1**). This interaction would not be seen in cjUNGM where an arginine residue (R59) replaces this histidine. These observations could explain the relative activity observed with UDP-3′′-deoxy-Gal*f* analog **3.7**, where the relative activity is of similar magnitude as with UDP-6′′-deoxy-Gal*f* analog **3.5**, compared to ecUGM and kpUGM where the activity with **3.5** is higher then that with **3.7**.

3.3 Conclusions

In this chapter, I report the synthesis of a panel of UDP-Gal*f* analogs and their use to study the substrate binding interactions of bacterial UGMs and UNGM. Using a chemo-enzymatic approach, in which the pyrophosphate moiety was formed via a promiscuous nucleotidyltransferase (GalPUT), allowed us to access deoxy UDP-Gal*f* analogs in moderate to high yield. However, this enzyme showed poor turnover for the methoxy analogs giving only small amounts of the desired sugar nucleotide products. It was nonetheless possible to isolate sufficient quantities of the 6′′-methoxy, 5′′-methoxy, and 2′′-methoxy

derivatives to allow for their evaluation. STD NMR experiments with UDP-Gal*p* and UDP-Gal*p*NAc, along with relative activity measurements using these singly modified UDP-Gal*f* derivatives, were used to elucidate the difference in substrate binding interactions between the cjUNGM and related UGM from *E. coli* and *K. pneumoniae*.

Titrating UDP-Gal*p*NAc (**3.3**) into a solution of the ecUGM demonstrated that no reaction occurred between the FAD cofactor and **3.3**. The results imply the previous hypothesis, from Chapter 2, that an unproductive oxazoline intermediate is formed in the reaction of ecUGM with **3.3**, is most likely incorrect. Instead, the differing STD-NMR intensities suggest that there are different binding modes of **3.3** to ecUGM and cjUNGM. Specifically, for ecUGM much lower STD effects were seen with the H-3′′, H-4′′ and H-6′′a/H-6′′b protons of UDP-Gal*p*NAc, compared to those observed in STD-NMR studies with cjUNGM.

More subtle differences were also observed for the binding of UDP-Gal*p*/UDP-Gal*f* with the two enzymes. For example, cjUNGM demonstrated lower STD affects for nearly all of the galactopyranose protons when compared to those observed with ecUGM. These differences imply different binding interactions in the active site of these two enzymes. Furthermore, the relative activity of these enzymes with UDP-Gal*f* analogs **3.5**–**3.13** provide further evidence of different binding interactions, particularly with the 3′′-hydroxyl group. In kpUGM, and ecUGM a histidine in the active site (H63 and H59, respectively) forms a hydrogen bond with the 3′′-hydroxyl group of UDP-Gal*p*.

This interaction is impossible in the cjUNGM in which the histidine is replaces by an arginine (R59). Instead this R59 appears to interact with the acetamido group of UDP-GalNAc according to kinetics measurements (Chapter 2).

Despite the high sequence identity of the *C. jejuni* UNGM to UGM enzymes, subtle differences in the active site residues appear to control the substrate specificity of this enzyme by influencing the conformation of the sugar nucleotide substrate. Understanding these interactions, leading to substrate specificity in these pyranose–furanose mutase enzymes, could help in annotating putative functions to the numerous other *glf* homologs being identified in other bacterial species (see Chapter 1).

3.4 Experimental Details

General Methods.

All reagents were purchased from commercial sources and were used without further purification. Reaction solvents were purified by successive passage through columns of alumina and copper under a nitrogen atmosphere using a PURESOLV-400 system (Innovative Technology Inc., Newburyport, MA). Reactions were carried out in oven-dried glassware. Unless stated otherwise, all reactions were carried out at room temperature under a positive pressure of argon and were monitored by TLC on silica gel 60-F_{254} (0.25 mm, Silicycle). Spots were detected under UV light or by charring with acidified ethanolic anisaldehyde. Unless otherwise indicated, column chromatography was

performed on silica gel 60 (40–60 μM) where the ratio of silica gel and crude product ranged from 100:1 to 20:1 (w/w). Organic solutions were concentrated under vacuum at < 40 °C (bath). Optical rotations were measured at 22 ± 2 °C on a Perkin–Elmer 241 polarimeter with a sodium D line (589 nm) and are given in units of (°·mL)/(dm·g). ^{1}H NMR spectra were recorded at 400 MHz, 500 MHz, 600 MHz or 700 MHz and chemical shifts are referenced to either TMS (0.0, $CDCl_3$) or HOD (4.67, D_2O). ^{13}C NMR spectra were recorded at 100 MHz, 125 MHz or 175 MHz, and ^{13}C chemical shifts were referenced to internal $CDCl_3$ (77.23, $CDCl_3$). Electrospray mass spectra were recorded on samples suspended in CH_3Cl or CH_3OH and added NaCl.

Enzyme preparation.

The *K. pneumoniae* UGM was prepared as previously described.[43] The *E. coli* UGM and *C. jejuni* UNGM enzymes were prepared as described in Chapter 2, with slight modifications as follows. For UV–visible spectroscopic studies, the *E. coli* UGM was prepared from 3 L cultures and eluted to a final concentration of 162 μM. For crystallography trials, the *C. jejuni* UNGM was prepared from 4 L cultures and purified by Ni-NTA agarose chromatography eluting with 250 mM imidazole in 40 mM Tris-HCl (pH 7.4) with 150 mM NaCl. After dialysis in 4 L of buffer (40 mM Tris-HCl (pH 7.4) with 150 mM NaCl, the protein was concentrated to ~7 mg mL^{-1} by centrifugal filtration in a 30,000 molecular weight cut off (Amicon® Ultra-15 Ultracel-30) centrifugal

filter device. Protein samples used for NMR analysis were lyophilized from D_2O in 1 mg aliquots directly before use, and dissolved directly in D_2O for use. Lyophilized proteins maintained >80% activity as determined using the HPLC assay described below (data not shown).

UV–visible spectroscopy.

FAD UV–visible spectroscopy was performed on a Hewlett Packard 8453 UV–VIS Spectrophotometer. The measurements were made in identical 50 µL cuvettes with, or without, 150 µM *E. coli* UGM in a 110 µL final volume of 100 mM potassium phosphate buffer (pH 7.4) containing 20 mM of freshly prepared sodium dithionite. UV–visible spectra were recorded in wave scan mode, as concentrated UDP-Gal*p* or UDP-Gal*p*NAc solutions were titrated into each cuvette as described previously.[14]

STD-NMR spectroscopy.

NMR samples were prepared containing 1.0 mg of UGM (or UNGM) in 50 mM potassium phosphate and D_2O (pH 7.4) with 20 mM freshly prepared sodium dithionite (0.7 mL) under an atmosphere of N_2. UDP, UDP-Gal*p*, or UDP-Gal*p*NAc was then added to give a final ligand concentration of 1.8 mM and ratio of ligand to protein of 50:1. Blank samples containing no ligand were also prepared to correct for protein resonances in the STD-NMR spectra.

STD-NMR spectra were recorded on a Varian/Agilent VNMRS four-channel, dual receiver 700 MHz spectrometer equipped with an inverse

detection, cryo-cooled $^{1}H\{^{15}N/^{13}C\}$ triple resonance, Z-gradient probe, at 300 K as described previously.[15, 17, 18] The spectra were recorded in 6000 scans while selectively irradiating the protein resonance at 0.0 ppm (50 ppm for reference spectra) using the pulse sequence described by Mayer and Meyer.[15] The saturation transfer double difference (STDD) NMR method[44] was used to eliminate background protein signals by subtracting a reference spectrum of the protein in the absence of ligand obtained under identical conditions. The STD effects were calculated for each resonance using $(I_o - I_{sat})/I_o$ where $(I_o - I_{sat})$ is the signal intensity of the STD spectrum and I_o is the signal intensity from a reference spectrum without saturation.

UGM (UNGM) activity assay.

To measure the relative activity of the ecUGM, kpUGM, and cjUNGM, reactions containing the UDP-Gal*f* analogs (**3.5–3.13**) (500 μM) and an appropriate amount of enzyme (adjusted to give 10–40% turnover for each substrate) in 30 μL of 100 mM potassium phosphate buffer (pH 7.4) containing sodium chloride (150 mM) and freshly prepared sodium dithionite (20 mM) were incubated at 37 °C for 10 min. The amount of enzyme was controlled to give <40% conversion to the pyranose product. The reactions were stopped by heating at 95 °C for 5 min to denature the protein. Each reaction was diluted to 100 μL with 100 mM potassium phosphate buffer (pH 7.4) prior to analysis by HPLC (Varian Prostar 210). The analysis used reversed phase ion-pairing chromatography on a C_{18} column (Varian, microsorb C-18, 4.6 × 250 mm) with

a 0.8 mL/min isocratic elution of 50 mM triethylammonium acetate buffer (pH 6.8) containing 1.5% acetonitrile. Under these conditions, baseline resolution for all substrate and product peaks was achieved. The elution times are shown in Appendix B. The relative activity was calculated based on the rate of product production, obtained from the relative integration of the product and starting material peaks.

Nucleotide exchange with TDP-Gal*f*

Reactions containing cjUNGM (600 nM), TDP-Gal*f* (250 μM), and UDP (500 μM) in 30 μL of 100 mM potassium phosphate buffer (pH 7.4) containing 20 mM freshly prepared sodium dithionite, were incubated at 37 °C for 30 min and then stopped by heating to 95 °C for 5 min. The reactions were analyzed by HPLC as described above. Reactions containing no UDP, or no cjUNGM, were run under identical conditions.

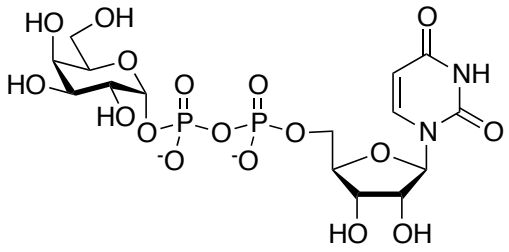
UGM (UNGM) inhibition assay

Reactions containing the UGM (or UNGM) enzyme (600 nM) in 30 μL of 100 mM potassium phosphate buffer (pH 7.4) containing 20 mM freshly prepared sodium dithionite, were pre-incubated with the potential inhibitor **3.6**, or **3.9** (2000, 1000, 500, 250, 100, 50, 25, 12.5, or 6.25 μM) for 4 min before adding UDP-Gal*f* (60 mM). The reactions were incubated for 5 min at 37 °C, then stopped by heating to 95 °C for 5 min to denature the enzyme. Reactions were analyzed by HPLC as described above, and the relative activity was

determined from the integration of the peaks for UDP-Gal*f* and UDP-Gal*p* compared directly to the activity of the reaction with no inhibitor under the same conditions. Reactions, at each inhibitor concentration, were run in duplicate. IC_{50} values were determined by non-linear regression analysis of plots of percent inhibition against inhibitor concentration.

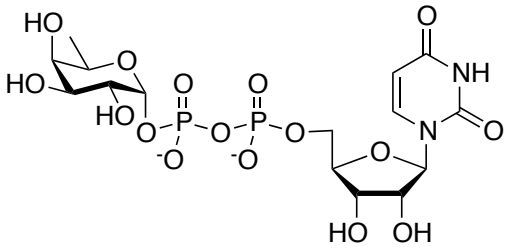
Isolation and characterization of reaction products.

Reactions of the pyranose–furanose mutase enzyme, UGM or UNGM (600 nM), containing UDP-Gal*f* (**3.2**) or analog **3.5**, **3.7**, or **3.12** (2 mM) in 100 μL of 100 mM potassium phosphate buffer (pH 7.4) containing sodium chloride (150 mM) and freshly prepared sodium dithionite (20 mM) were incubated at 37 °C for 1 h, to ensure the reaction had reached equilibrium. The reactions were stopped by heating at 95 °C for 5 min to denature the protein and were then purified reversed phase ion-pairing chromatography using a C_{18} column (Varian, microsorb C-18, 4.6 × 250 mm) with a 0.8 mL/min isocratic elution using 50 mM triethylammonium acetate buffer pH 6.8 containing 1.5% acetonitrile. The product peaks were collected and lyophilized. Before analysis the reaction products were further purified by strong anion exchange chromatography, eluting with 100 mM ammonium bicarbonate (pH 6.8). The final isolated products were lyophilized and then analyzed by ^{1}H NMR spectroscopy.



Uridine-5′-diphosphate α-D-galactopyranose (3.1):

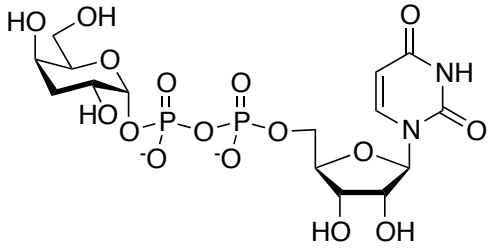
^{1}H NMR (500 MHz, D_2O) δ 7.94 (d, *J* = 8.1 Hz, 1H, H-6), 5.99–5.96 (m, 1H, H-1′), 5.96 (d, *J* = 8.1 Hz, 1H, H-5), 5.63 (dd, *J* = 7.0, 3.4 Hz, 1H, H-1′′), 4.40 – 4.33 (m, 2H, H-2′/H-3′), 4.28 (br. s, 1H, H-4′), 4.26 – 4.18 (m, 2H, H-5a′/H-5b′), 4.18 – 4.13 (m, 1H, H-5′′), 4.02 (br. d, *J* = 2.8 Hz, 1H, H-4′′), 3.90 (dd, *J* = 10.4, 2.8 Hz, 1H, H-3′′), 3.88 – 3.81 (m, 1H, H-2′′), 3.81 – 3.77 (m, 2H, H-6a′′/H-6b′′). This NMR data is consistent with previously reported NMR data for UDP-Gal*p*.[45]



Uridine-5′-diphosphate 6′′-deoxy-α-D-galactopyranose:

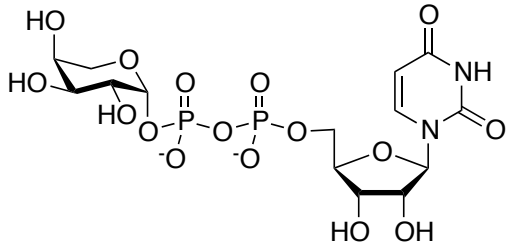
^{1}H NMR (600 MHz, D_2O) δ 7.95 (d, *J* = 8.1 Hz, 1H, H-6), 6.00–5.94 (m, 2H, H-5/H-1′), 5.55 (dd, *J* = 6.6, 3.7 Hz, 1H, H-1′′), 4.39–4.34 (m, 2H, H-2′/H-3′), 4.30–4.26 (m, 2H, H-4′/H-5′′), 4.23 (ddd, *J* = 11.9, 6.1, 3.0 Hz, 1H, H-5a′), 4.19 (ddd, *J* = 11.9, 5.5, 2.8 Hz, 1H, H-5b′), 3.91 (dd, *J* = 10.3, 3.3 Hz, 1H, H-

3′′), 3.81 (d, *J* = 3.0 Hz, 1H, H-4′′), 3.76–3.71 (m, 1H, H-2′′), 1.21 (d, *J* = 6.6 Hz, 3H, H-6′′).



Uridine-5′-diphosphate 3′′-deoxy-α-D-*xylo*-hexopyranose:

^{1}H NMR (600 MHz, D_2O) δ 7.95 (d, *J* = 8.1 Hz, 1H, H-6), 6.01–5.94 (m, 2H, H-1′/H-5), 5.58 (dd, *J* = 7.0, 3.3 Hz, 1H, H-1′′), 4.40–4.34 (m, 2H, H-2′/H-3′), 4.30–4.26 (m, 1H, H-4′), 4.26–4.17 (m, 2H, H-5′a/H-5′b), 4.11–4.05 (m, 2H, H-4′′/H-5′′), 4.04–3.98 (m, 1H, H-2′′), 3.73–3.66 (m, 2H, H-6′′a/H-6′′b), 2.02–1.97 (m, 2H, H-3′′a/H-3′′b).



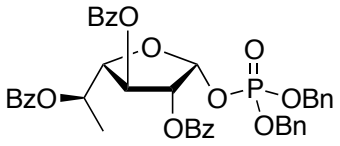
Uridine-5′-diphosphate β-L-arabinopyranose:

^{1}H NMR (700 MHz, D_2O) δ 7.96 (d, *J* = 8.1 Hz, 1H, H-6), 6.00–5.94 (m, 2H, H-1′/H-5), 5.61 (dd, *J* = 6.9, 3.5 Hz, 1H, H-1′′), 4.39–4.35 (m, 2H, H-2′/H-3′), 4.30–4.27 (m, 1H, H-4′), 4.25 (ddd, *J* = 11.7, 4.5, 2.6 Hz, 1H, H-5′a), 4.20 (ddd, *J* = 11.8, 5.6, 2.8 Hz, 1H, H-5′b), 4.12 (dd, *J* = 12.9, 0.8 Hz, 1H, H-5′′a),

4.05–4.01 (m, 1H, H-4′′), 3.93 (dd, J = 10.1, 3.4 Hz, 1H, H-3′′), 3.82 (app. dt, J = 10.2, 3.3 Hz, 1H, H-2′′), 3.73 (dd, J = 12.9, 2.1 Hz, 1H, H-5′′b).

General procedure for glycosylation with dibenzylphosphate.

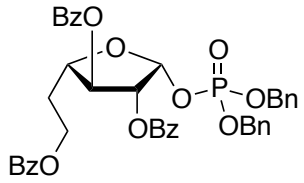
To a solution of methyl glycoside (0.15 mM, 1 equiv) in CH_2Cl_2 at 0 °C was added 33% HBr in AcOH (~1.0 mL) while keeping the temperature below 0 °C. After 3 h, the reaction mixture was diluted with toluene and concentrated. Without purification, the residue was resuspended in toluene (0.2 mM). To this solution was added dibenzylphosphate (1.08 eq) and Et_3N (1 mL). The reaction was carried out at room temperature for 1 h. The reaction mixture was then filtered to remove Et_3NHBr salts and the filtrate was concentrated to give a crude oil that was purified by column chromatography (6:1 hexanes–EtOAc).



Dibenzyl 2,3,5-Tri-*O*-benzoyl-6-deoxy-α-D-galactofuranosyl-1-phosphate (3.22):

From **3.14** (143 mg, 0.29 mmol) using the general glycosylation procedure, **3.22** (98 mg, 46%) was isolated as a colourless oil: R_f 0.30 (2:1 hexanes–EtOAc); $[\alpha]_D$ +55.2 (c 0.3, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$, δ_H) 8.18–7.12 (m, 25 H, Ar), 6.37 (dd, 1 H, J = 5.7, 4.5 Hz, H-1), 6.19 (dd, 1 H, J = 7.8, 6.9 Hz, H-3), 5.74 (ddd, 1 H, J = 7.8, 4.5, 2.0 Hz, H-2), 5.52 (dq, 1 H, J = 6.5, 4.9 Hz, H-5), 5.07 (dd, 1 H, J = 11.8, 7.3 Hz, OCH_2Ph), 4.99 (dd, 1 H, J =

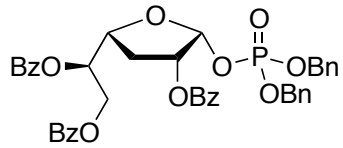
11.9, 6.6 Hz, OCH_2Ph), 4.97 (dd, 1 H, J = 11.7, 7.9 Hz, OCH_2Ph), 4.90 (dd, 1 H, J = 11.7, 7.6 Hz, OCH_2Ph), 4.48 (dd, 1 H, J = 6.9, 4.9 Hz, H-4), 1.49 (d, 3 H, J = 6.5 Hz, H-6); ^{13}C NMR (100 MHz, $CDCl_3$, δ_C) 165.7 (C=O), 165.5 (C=O), 165.3 (C=O), 135.4 (1 C, J = 3.1 Hz, Ar), 135.3 (1 C, J = 3.4 Hz, Ar), 133.6 (Ar), 133.5 (Ar), 132.9 (Ar), 130.0 (2 × Ar), 129.8 (2 × Ar), 129.8 (2 × Ar), 128.8 (Ar), 128.6 (Ar), 128.4 (5 × Ar), 128.3 (2 × Ar), 128.3 (2 × Ar), 128.2 (2 × Ar), 128.2 (2 × Ar), 127.6 (2 × Ar), 127.5 (2 × Ar), 97.5 (d, 1 C, J = 4.9 Hz, C-1), 82.4 (C-4), 76.6 (d, 1 C, J = 7.2 Hz, C-2), 73.1 (C-3), 70.2 (C-5), 69.2 (d, 1 C, J = 5.3 Hz, OCH$_2$Ph), 69.1 (d, 1 C, J = 5.4 Hz, OCH$_2$Ph), 15.6 (C-6); HRMS (ESI) *m/z* Calcd. for (M+Na) $C_{41}H_{37}O_{11}NaP$: 759.1966. Found: 759.1962.



Dibenzyl 2,3,6-Tri-*O*-benzoyl-5-deoxy-β-L-*arabino*hexofuranosyl-1-phosphate (3.23):

From **3.15** (120 mg, 0.24 mmol), following the general glycosylation procedure, **3.23** (92 mg, 52%) was obtained as a colourless oil. R_f 0.30 (2:1 hexanes–EtOAc); $[\alpha]_D$ +71.8 (*c* 0.2, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$, δ_H) 8.12–7.15 (m, 25 H, Ar), 6.29 (dd, 1 H, J = 4.7, 4.3 Hz, H-1), 5.83 (dd, 1 H, J = 6.8, 5.4 Hz, H-3), 5.66 (ddd, 1 H, J = 6.8, 4.3, 2.3 Hz, H-2), 5.04 (dd, 1 H, J = 11.8, 7.6 Hz, OCH_2Ph), 4.99 (dd, 1 H, J = 12.3, 8.4 Hz, OCH_2Ph), 4.97 (dd, 1 H, J = 11.7, 7.0 Hz, OCH_2Ph), 4.88 (dd, 1 H, J = 11.7, 8.0 Hz, OCH_2Ph), 4.47–4.42

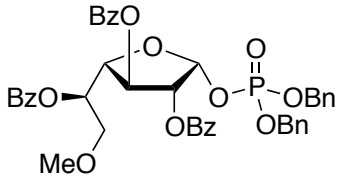
(m, 3 H, H-4/H-6a/H-6b), 2.51–2.41 (m, 1 H, H-5a), 2.37–2.27 (m, 1 H, H-5b); ^{13}C NMR (100 MHz, $CDCl_3$, δ_C) 166.2 (C=O), 165.8 (C=O), 165.5 (C=O), 135.4 (1 C, *J* = 7.3 Hz, Ar), 135.3 (1 C, *J* = 7.6 Hz, Ar), 133.5 (2 × Ar), 132.8 (Ar), 130.0 (Ar), 129.9 (2 × Ar), 129.9 (Ar), 129.8 (2 × Ar), 129.5 (2 × Ar), 128.8 (Ar), 128.6 (Ar), 128.5 (4 × Ar), 128.5 (2 × Ar), 128.4 (2 × Ar), 128.3 (Ar), 128.2 (2 × Ar), 127.7 (2 × Ar), 127.6 (2 × Ar), 98.3 (d, 1 C, *J* = 5.4 Hz, C-1), 79.8 (C-4), 78.2 (C-3), 77.0 (d, 1 C, *J* = 7.7 Hz, C-2), 69.3 (d, 1 C, *J* = 5.5 Hz, O*C*H_2Ph), 69.2 (d, 1 C, *J* = 5.5 Hz, O*C*H_2Ph), 61.2 (C-6), 34.3 (C-5); HRMS (ESI) *m/z* Calcd. for (M+Na) $C_{41}H_{37}O_{11}NaP$: 759.1966. Found: 759.1962.



Dibenzyl 2,5,6-tri-*O*-benzoyl-3-deoxy-α-D-*xylo*-hexofuranosyl-1-phosphate (3.24):

Prepared from compound **3.16** (183 mg, 0.37 mmol) using the general glycosylation procedure with dibenzyl phosphate to give **3.24** (95 mg, 35%) as a colourless oil. R_f 0.20 (2:1 hexanes–EtOAc); $[\alpha]_D$ +11.0 (*c* 0.4, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$, δ_H) 8.18–8.02 (m, 6 H, Ar), 7.61–7.16 (m, 19 H, Ar), 6.20 (dd, 1 H, *J* = 5.2, 4.0 Hz, H-1), 5.69 (ddd, 1 H, *J* = 6.5, 4.0, 4.0 Hz, H-5), 5.43 (dddd, 1 H, *J* = 11.3, 7.5, 4.0, 2.0 Hz, H-2), 5.11 (dd, 1 H, *J* = 11.8, 7.4 Hz, PhC*H*$_2$), 5.04 (dd, 1 H, *J* = 11.8, 8.0 Hz, PhC*H*$_2$), 5.03 (dd, 1 H, *J* = 11.8, 6.7 Hz, PhC*H*$_2$), 4.95 (dd, 1 H, *J* = 11.8, 7.8 Hz, PhC*H*$_2$), 4.75 (dd, 1 H, *J* = 11.9, 4.0 Hz, H-6a), 4.73–4.68 (m, 1 H, H-4), 4.61 (dd, 1 H, *J* = 11.9, 6.5 Hz, H-6b), 2.67

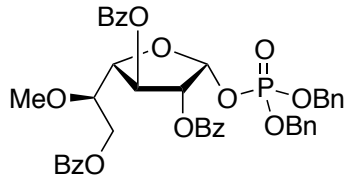
(ddd, 1 H, J = 12.0, 7.5, 6.5 Hz, H-3a), 2.37–2.27 (m, 1 H, H-3b); ^{13}C NMR (100 MHz, $CDCl_3$, δ_C) 166.0 (*C*=O), 165.9 (*C*=O), 165.7 (*C*=O), 135.6 (d, J = 7.5 Hz, Ar), 135.5 (d, J = 7.8 Hz, Ar), 133.3 (Ar), 133.2 (Ar), 133.2 (Ar), 129.9 (2 × Ar), 129.8 (2 × Ar), 129.6 (2 × Ar), 129.5 (Ar), 129.4 (Ar), 128.9 (Ar), 128.4 (6 × Ar), 128.3 (2 × Ar), 128.3 (2 × Ar), 128.2 (Ar), 128.2 (Ar), 127.7 (2 × Ar), 127.5 (2 × Ar), 97.8 (d, J = 5.1 Hz, C-1), 77.0 (C-4), 72.8 (d, J = 7.0 Hz, C-2), 72.3 (C-5), 69.2 (d, J = 5.3 Hz, Ph*C*H_2), 69.1 (d, J = 5.3 Hz, Ph*C*H_2), 63.1 (C-6), 28.3 (C-3); HRMS (ESI) *m/z* Calcd. for (M + Na) $C_{41}H_{37}O_{11}NaP$: 759.1966. Found: 759.1970.



Dibenzyl 2,3,5-tri-*O*-benzoyl-6-*O*-methyl-α-D-galactofuranosyl-1-phosphate (3.26):

Prepared from compound **3.18** (207 mg, 0.40 mmol) following the general glycosylation procedure with dibenzyl phosphate to give **3.26** (129 mg, 43%) as a colourless oil. R_f 0.31 (2:1 hexanes–EtOAc); $[\alpha]_D$ +47.8 (*c* 0.3, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$, δ_H) 8.16–7.99 (m, 6 H, Ar), 7.59–7.07 (m, 19 H, Ar), 6.37 (dd, 1 H, J = 5.7, 4.5 Hz, H-1), 6.20 (dd, 1 H, J = 7.6, 6.7 Hz, H-3), 5.75 (ddd, 1 H, J = 7.6, 4.5, 2.0 Hz, H-2), 5.59 (ddd, 1 H, J = 6.2, 5.0, 5.0 Hz, H-5), 5.06 (dd, 1 H, J = 11.8, 7.2 Hz, PhCH_2), 4.98 (dd, 1 H, J = 11.8, 6.3 Hz, PhCH_2), 4.97 (dd, 1 H, J = 11.8, 8.0 Hz, PhCH_2), 4.89 (dd, 1 H, J = 11.7, 7.6 Hz,

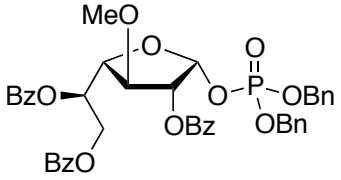
PhCH_2), 4.75 (dd, 1 H, J = 6.7, 5.0 Hz, H-4), 3.76 (dd, 1 H, J = 10.1, 5.1 Hz, H-6a), 3.70 (dd, 1 H, J = 10.1, 6.2 Hz, H-6b), 3.27 (s, 3 H, OCH_3); ^{13}C NMR (100 MHz, $CDCl_3$, δ_C) 165.7 (*C*=O), 165.5 (*C*=O), 165.4 (*C*=O), 135.4 (d, J = 7.6 Hz, Ar), 135.4 (d, J = 8.1 Hz, Ar), 133.5 (Ar), 133.5 (Ar), 133.1 (Ar), 130.0 (2 × Ar), 129.9 (2 × Ar), 129.8 (2 × Ar), 129.5 (Ar), 128.8 (Ar), 128.6 (Ar), 128.4 (4 × Ar), 128.3 (6 × Ar), 128.3 (2 × Ar), 127.7 (2 × Ar), 127.6 (2 × Ar), 97.6 (d, J = 5.1 Hz, C-1), 79.8 (C-4), 76.6 (d, J = 7.7 Hz, C-2), 73.1 (C-3), 71.8 (C-5), 69.8 (C-6), 69.2 (d, J = 5.6 Hz, Ph*C*H$_2$), 69.1 (d, J = 5.5 Hz, Ph*C*H$_2$), 59.1 (O*C*H$_3$); HRMS (ESI) *m/z* Calcd. for (M + Na) $C_{42}H_{39}O_{12}NaP$: 789.2071. Found: 789.2081.



Dibenzyl 2,3,6-tri-*O*-benzoyl-5-*O*-methyl-α-D-galactofuranosyl-1-phosphate (3.27):

From **3.19** (149 mg, 0.29 mmol) as described in the general glycosylation procedure above to give **3.27** (89 mg, 41%) as a colourless oil. R_f 0.22 (2:1 hexanes–EtOAc); $[\alpha]_D$ +53.4 (*c* 0.2, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$, δ_H) 8.08–7.93 (m, 6 H, Ar), 7.53–7.13 (m, 19 H, Ar), 6.30 (dd, 1 H, J = 5.5, 4.4 Hz, H-1), 6.26 (dd, 1 H, J = 8.0, 7.0 Hz, H-3), 5.68 (ddd, 1 H, J = 8.0, 4.4, 2.3 Hz, H-2), 5.11 (dd, 1 H, J = 10.6, 5.9 Hz, PhCH_2), 5.06 (dd, 1 H, J = 10.7, 6.4 Hz, PhCH_2), 4.97 (dd, 1 H, J = 11.7, 6.8 Hz, PhCH_2), 4.90 (dd, 1 H, J = 11.7, 7.6 Hz,

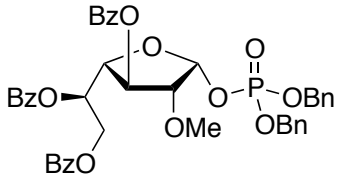
PhCH_2), 4.59 (dd, 1 H, J = 12.0, 4.6 Hz, H-6a), 4.47 (dd, 1 H, J = 12.0, 5.2 Hz, H-6b), 4.46 (dd, 1 H, J = 7.0, 5.1 Hz H-4), 3.80 (app. q, 1 H, J = 4.9 Hz, H-5), 3.58 (s, 3 H, OCH_3); ^{13}C NMR (100 MHz, $CDCl_3$, δ_C) 166.0 (*C*=O), 165.6 (2 × *C*=O), 135.8 (d, J = 8.0 Hz, Ar), 135.5 (d, J = 7.8 Hz, Ar), 133.6 (2 × Ar), 133.0 (Ar), 130.0 (2 × Ar), 129.9 (2 × Ar), 129.7 (Ar), 129.6 (2 × Ar), 128.8 (Ar), 128.7 (Ar), 128.5 (2 × Ar), 128.5 (4 × Ar), 128.4 (2 × Ar), 128.3 (4 × Ar), 127.7 (2 × Ar), 127.6 (2 × Ar), 97.4 (d, $J_{1,P}$ = 2.3 Hz, C-1), 81.0 (C-4), 78.5 (C-5), 76.6 (d, $J_{2,P}$ = 7.6 Hz, C-2), 73.1 (C-3), 69.2 (d, J = 5.6 Hz, O*C*H_2Ph), 69.2 (d, J = 5.3 Hz, O*C*H_2Ph), 63.0 (C-6), 59.3 (d, J = 2.4 Hz, O*C*H_3); HRMS (ESI) *m/z* Calcd. for (M + Na) $C_{42}H_{39}O_{12}NaP$: 789.2071. Found: 789.2082.



Dibenzyl 2,5,6-tri-*O*-benzoyl-3-*O*-methyl-α-D-galactofuranosyl-1-phosphate (3.28):

Compound **3.20** (162 mg, 0.31) was treated as described in the general glycosylation procedure to provide **3.28** (114 mg, 41%) as a colourless oil. R_f 0.30 (2:1 hexanes–EtOAc); $[\alpha]_D$ +55.7 (*c* 0.3, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$, δ_H) 8.16–7.98 (m, 6 H, Ar), 7.55–7.05 (m, 19 H, Ar), 6.23 (dd, 1 H, J = 5.7, 4.4 Hz, H-1), 5.80 (ddd, 1 H, J = 7.0, 4.6, 4.2 Hz, H-5), 5.39 (ddd, 1 H, J = 7.7, 4.4, 2.1 Hz, H-2), 4.99 (dd, 1 H, J = 11.7, 7.3 Hz, PhCH_2), 4.92 (dd, 1 H, J = 11.8, 8.3 Hz, PhCH_2), 4.91 (dd, 1 H, J = 11.7, 6.8 Hz, PhCH_2), 4.80 (dd, 1 H, J =

11.8, 7.8 Hz, PhCH_2), 4.69 (dd, 1 H, J = 11.8, 4.6 Hz, H-6a), 4.58 (dd, 1 H, J = 11.8, 7.0 Hz, H-6b), 4.43 (dd, 1 H, J = 7.4, 4.2 Hz, H-4), 4.35 (dd, 1 H, J = 7.7, 7.4 Hz, H-3), 3.50 (s, 3 H, OCH_3); ^{13}C NMR (100 MHz, $CDCl_3$, δ_C) 166.0 (*C*=O), 166.0 (*C*=O), 165.4 (*C*=O), 135.5 (d, J = 7.4 Hz, Ar), 135.4 (d, J = 7.8 Hz, Ar), 133.6 (Ar), 133.3 (Ar), 133.2 (Ar), 130.1 (2 × Ar), 129.8 (2 × Ar), 129.7 (2 × Ar), 129.6 (Ar), 129.3 (Ar), 128.8 (Ar), 128.5 (2 × Ar), 128.5 (2 × Ar), 128.4 (2 × Ar), 128.4 (2 × Ar), 128.3 (2 × Ar), 128.3 (Ar), 128.2 (Ar), 127.8 (2 × Ar), 127.6 (2 × Ar), 97.8 (d, J = 5.3 Hz, C-1), 80.2 (C-3), 79.7 (C-4), 77.9 (d, J = 7.0 Hz, C-2), 70.8 (C-5), 69.2 (d, J = 5.3 Hz, Ph*C*H_2), 69.1 (d, J = 5.1 Hz, Ph*C*H_2), 62.9 (C-6), 58.8 (O*C*H_3); HRMS (ESI) *m/z* Calcd. for (M + Na) $C_{42}H_{39}O_{12}NaP$: 789.2071. Found: 789.2079.



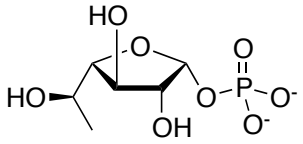
Dibenzyl 3,5,6-tri-*O*-benzoyl-2-*O*-methyl-α-D-galactofuranosyl-1-phosphate (3.29):

Compound **3.21** (497 mg, 0.95 mmol) was treated following the general procedure for glycosylation with dibenzylphosphate to give **3.29** (334 mg, 46%) as a colourless oil. R_f 0.19 (2:1 hexanes–EtOAc); $[\alpha]_D$ +18.7 (*c* 0.6, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$, δ_H) 8.10–7.90 (m, 6 H, Ar), 7.56–7.23 (m, 19 H, Ar), 6.15 (dd, 1 H, J = 5.9, 4.2 Hz, H-1), 5.89 (dd, 1 H, J = 7.8, 7.3 Hz, H-3), 5.74 (ddd, 1 H, J = 6.0, 4.5, 4.5 Hz, H-5), 5.18–5.07 (m, 4 H, 2 × OCH_2Ph), 4.71 (dd,

1 H, J = 11.9, 4.5 Hz, H-6a), 4.60–4.55 (m, 2 H, H-4, H-6b), 4.27 (ddd, 1 H, $J_{2,3}$ = 7.8, 4.2, 1.9 Hz, H-2), 3.48 (s, 3 H, OCH_3); ^{13}C NMR (100 MHz, $CDCl_3$, δ_C) 165.8 (*C*=O), 165.6 (*C*=O), 165.4 (*C*=O), 135.7 (d, J = 7.8 Hz, Ar), 135.5 (d, J = 7.7 Hz, Ar), 133.5 (Ar), 133.1 (Ar), 133.0 (Ar), 130.0 (2 × Ar), 129.8 (2 × Ar), 129.7 (2 × Ar), 129.6 (Ar), 129.5 (Ar), 128.9 (Ar), 128.5 (2 × Ar), 128.4 (Ar), 128.4 (4 × Ar), 128.3 (Ar), 128.3 (2 × Ar), 128.2 (2 × Ar), 127.9 (2 × Ar), 127.8 (2 × Ar), 97.9 (d, J = 5.0 Hz, C-1), 84.2 (d, J = 5.9 Hz, C-2), 80.1 (C-4), 74.2 (C-3), 70.8 (C-5), 69.5 (d, J = 5.6 Hz, Ph*C*H_2), 69.3 (d, J = 5.5 Hz, Ph*C*H_2), 62.7 (C-6), 58.8 (O*C*H_3); HRMS (ESI) *m/z* Calcd. for (M + Na) $C_{42}H_{39}O_{12}NaP$: 789.2071. Found: 789.2071.

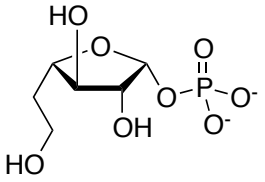
General deprotection procedure.

To a solution of protected furanose-1-phosphate analog (0.15 mM, 1 equiv) in EtOAc was added 10% Pd–C (15% by weight) and Et_3N (6 equiv). The reaction mixture was stirred under H_2 (1 atm) for 16 h, and then the catalyst was removed by filtration and the filtrate evaporated. The resulting residue was dissolved in a 10:2:1 solution of CH_3OH–H_2O–Et_3N (0.05 mM) and stirred at ambient temperature for 6 days until TLC showed complete consumption of the starting material. The solvent was removed by evaporation, and the product was purified using reversed phase C_{18} column chromatography eluting with water.



6-Deoxy-α-D-galactofuranose-1-phosphate triethylammonium salt (3.30):

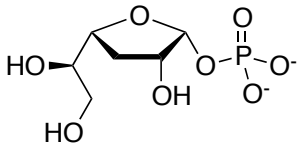
Compound **3.22** (98 mg, 0.13 mmol) was treated using the general deprotection procedure to give **3.30** as a colourless oil (23 mg, 40%). R_f 0.20 (10:2:1 MeOH–NH_4OH–H_2O); $[\alpha]_D$ +41.5 (*c* 0.1, H_2O); ^{1}H NMR (500 MHz, D_2O, δ_H) 5.50 (dd, 1 H, $J_{1,P}$ = 5.6, 4.3 Hz, H-1), 4.10 (ddd, 1 H, *J* = 8.2, 4.3, 2.2 Hz, H-2), 4.06 (dd, 1 H, *J* = 8.2, 6.8 Hz, H-3), 3.86 (dq, 1 H, *J* = 6.8, 6.6 Hz, H-5), 3.59 (app. t, 1 H, *J* = 6.8 Hz, H-4), 3.19 (q, 9 H, *J* = 7.3, 1.5 × $(CH_3CH_2)_3N$), 1.26 (t, 13.5 H, *J* = 7.3, 1.5 × $(CH_3CH_2)_3N$), 1.19 (d, 3 H, *J* = 6.6 Hz, H-6); ^{13}C NMR (125 MHz, D_2O, δ_c) 97.6 (d, 1 C, *J* = 5.7 Hz, C-1), 86.6 (C-4), 78.0 (d, 1 C, *J* = 7.7 Hz, C-2), 75.5 (C-3), 69.9 (C-5), 47.6 (1.5 × $(CH_3CH_2)_3N$), 18.4 (C-6), 9.1 (1.5 × $(CH_3CH_2)_3N$); HRMS (ESI) *m/z* Calcd. for $(M–H)^–$ $C_{27}H_{24}O_{11}P$: 555.1062. Found: 555.1059.



5-Deoxy-β-L-*arabino*-hexofuranose-1-phosphate triethylammonium salt (3.31):

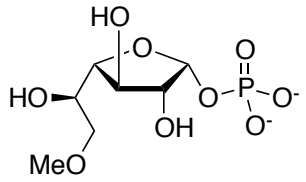
From **3.23** (90 mg, 0.12 mmol) using the general deprotection procedure gave **3.31** (28 mg, 57%) as a colourless syrup. R_f 0.23 (10:2:1 MeOH–NH_4OH–H_2O), $[\alpha]_D$ +10.9 (*c* 0.1, H_2O); ^{1}H NMR (400 MHz, D_2O, δ_H) 5.49 (dd, 1 H, *J* =

5.3, 4.8 Hz, H-1), 4.08 (ddd, 1 H, J = 7.9, 4.8, 1.3 Hz, H-2), 4.02 (dd, 1 H, J = 7.9, 7.0 Hz, H-3), 3.92–3.85 (m, 1 H, H-4), 3.76–3.69 (m, 2 H, H-6a/H-6b), 3.18 (q, 8 H, J = 7.3, 1.3 × $(CH_3CH_2)_3N$), 1.98–1.89 (m, 2 H, H-5a/H-5b), 1.26 (t, 12 H, J = 7.3, 1.3 × $(CH_3CH_2)_3N$); ^{13}C NMR (100 MHz, D_2O, δ_C) 97.9 (d, 1 C, J = 5.6 Hz, C-1), 80.3 (C-4), 78.5 (C-3), 77.4 (d, 1 C, J = 7.4 Hz, C-2), 59.3 (C-6), 47.5 (1.3 × $(CH_3CH_2)_3N$), 37.6 (C-5), 9.1 (1.3 × $(CH_3CH_2)_3N$); HRMS (ESI) *m/z* Calcd. for $(M–H)^-$ $C_6H_{12}O_8P$: 243.0264. Found: 243.0263.



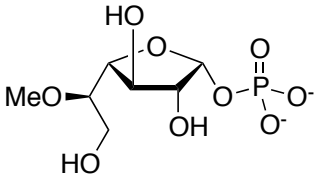
3-Deoxy-α-D-*xylo*-hexofuranose-1-phosphate triethylammonium salt (3.32):

Compound **3.24** (90 mg, 0.12 mmol) was deprotected using the general deprotection procedure and lyophilized to give **3.32** (33.7 mg, 58%) as a colourless syrup. R_f 0.21 (10:2:1 $CH_3OH–NH_4OH–H_2O$); $[\alpha]_D$ +33.5 (*c* 0.2, H_2O); ^{1}H NMR (500 MHz, D_2O, δ_H) 5.45 (dd, 1 H, J = 5.0, 4.1 Hz, H-1), 4.28 (dddd, 1 H, J = 10.7, 7.4, 4.1, 2.0 Hz, H-2), 4.10–4.04 (m, 1 H, H-4), 3.70–3.62 (m, 2 H, H-5, H-6a), 3.57–3.53 (m, 1 H, H-6b), 3.18 (q, 8 H, J = 7.3, 1.3 × $(CH_3CH_2)_3N$), 2.27 (ddd, 1 H, J = 11.8, 7.4, 6.6 Hz, H-3a), 1.86–1.78 (m, 1 H, H-3b), 1.26 (t, 12H, J = 7.3, 1.3 × $(CH_3CH_2)_3N$); ^{13}C NMR (125 MHz, D_2O, δ_C) 97.4 (d, J = 5.9 Hz, C-1), 79.1 (C-4), 75.2 (C-5), 72.6 (d, J = 7.7 Hz, C-2), 63.4 (C-6), 47.5 (1.3 × $(CH_3CH_2)_3N$), 31.7 (C-3), 9.1 (1.3 × $(CH_3CH_2)_3N$); HRMS (ESI) *m/z* Calcd. for $(M–H)^-$ $C_6H_{12}O_8P$: 243.0264. Found: 243.0263.



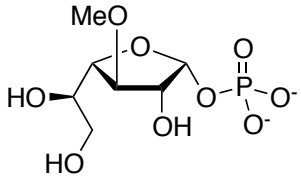
6-*O*-Methyl-α-D-galactofuranose-1-phosphate triethylammonium salt (3.34).

Compound **3.26** (123 mg, 0.16 mmol) was treated as described in the general deprotection procedure and lyophilized to give **3.34** (48 mg, 63%) as a colourless oil. R_f 0.20 (10:2:1 CH_3OH–NH_4OH–H_2O); $[\alpha]_D$ +33.4 (*c* 0.2, H_2O); ^1H NMR (400 MHz, D_2O, δ_H) 5.49 (dd, 1 H, *J* = 4.7, 4.3 Hz, H-1), 4.20 (dd, 1 H, *J* = 8.3, 7.3 Hz, H-3), 4.10 (ddd, 1 H, *J* = 8.3, 4.3, 2.0 Hz, H-2), 3.86 (ddd, 1 H, *J* = 7.2, 5.2, 4.1 Hz, H-5), 3.76 (dd, 1 H, *J* = 7.3, 5.2 Hz, H-4), 3.59 (dd, 1 H, *J* = 10.7, 4.1 Hz, H-6a), 3.51 (dd, 1 H, *J* = 10.7, 7.2 Hz, H-6b), 3.38 (s, 3 H, OC*H*$_3$), 3.18 (q, 9 H, *J* = 7.3 Hz, 1.5 × (CH_3CH_2)$_3$N), 1.26 (t, 13.5 H, *J* = 7.3 Hz, 1.5 × (CH_3CH_2)$_3$N); ^{13}C NMR (125 MHz, D_2O, δ_c) 97.5 (d, *J* = 5.7 Hz, C-1), 82.5 (C-4), 77.6 (d, *J* = 7.5 Hz, C-2), 74.7 (C-3), 73.7 (C-6), 71.0 (C-5), 59.3 (O*C*H$_3$), 47.5 (1.5 × (CH_3CH_2)$_3$N), 9.1 (1.5 × (CH_3CH_2)$_3$N); HRMS (ESI) *m/z* Calcd. for $(M–H)^-$ $C_7H_{14}O_9P$: 273.0370. Found: 273.0368.



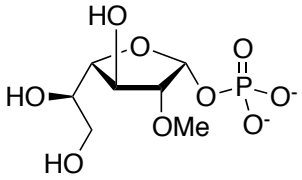
5-*O*-Methyl-α-D-galactofuranose-1-phosphate triethylammonium salt (3.35):

Deprotected compound **3.27** (88 mg, 0.12 mmol) as described in the general procedure above. The desired fractions were lyophilized to give **3.35** (33 mg, 64%) as a colourless oil. R_f 0.19 (10:2:1 CH_3OH–NH_4OH–H_2O); $[\alpha]_D$ +31.5 (*c* 0.3, H_2O); ^{1}H NMR (500 MHz, D_2O, δ_H) 5.52 (dd, 1 H, *J* = 6.3, 4.1 Hz, H-1), 4.14 (dd, 1 H, *J* = 8.2, 7.2 Hz, H-3), 4.10 (ddd, 1 H, *J* = 8.2, 4.1, 1.4 Hz, H-2), 3.87–3.82 (m, 2H, H-4, H-6a), 3.65 (dd, 1 H, *J* = 12.4, 5.7 Hz, H-6b), 3.54–3.50 (m, 1H, H-5), 3.53 (s, 3 H, OC*H*$_3$), 3.19 (q, 8 H, *J* = 7.3 Hz, 1.3 × (CH_3CH_2)$_3$N), 1.27 (t, 12 H, *J* = 7.3 Hz, 1.3 × (C*H*$_3$$CH_2$)$_3$N); ^{13}C NMR (125 MHz, D_2O, δ_c) 97.7 (d, *J* = 5.1 Hz, C-1), 83.6 (C-5), 81.5 (C-4), 77.4 (d, *J* = 6.8 Hz, C-2), 74.9 (C-3), 60.3 (C-6), 59.2 (O*C*H$_3$), 47.6 (1.3 × (CH_3*C*H_2)$_3$N), 9.1 (1.3 × (*C*H_3CH_2)$_3$N); HRMS (ESI) *m/z* Calcd. for $(M–H)^–$ $C_7H_{14}O_9P$: 273.0370. Found: 273.0372.



3-*O*-Methyl-α-D-galactofuranose-1-phosphate triethylammonium salt (3.36):

From **3.28** (111 mg, 0.145 mmol) following the general deprotection procedure gave **3.36** (37 mg, 54%) as a lyophilized syrup: R_f 0.28 (10:2:1 CH_3OH–NH_4OH–H_2O); $[\alpha]_D$ +33.2 (*c* 0.2, H_2O); ^1H NMR (500 MHz, D_2O, δ_H) 5.49 (dd, 1 H, *J* = 4.5, 4.5 Hz, H-1), 4.23 (ddd, 1 H, *J* = 7.4, 4.5, 2.0 Hz, H-2), 4.01 (dd, 1 H, *J* = 7.4, 6.5 Hz, H-3), 3.85 (dd, 1 H, *J* = 6.5, 5.0 Hz, H-4), 3.77 (ddd, 1 H, *J* = 7.0, 5.0, 4.5 Hz, H-5), 3.67 (dd, 1 H, *J* = 11.8, 4.5 Hz, H-6a), 3.59 (dd, 1 H, *J* = 11.8, 7.0 Hz, H-6b), 3.52 (s, 3 H, O*CH*$_3$), 3.18 (q, 9 H, *J* = 7.3, 1.5 × (CH_3CH_2)$_3$N), 1.26 (t, 13.5 H, *J* = 7.3 Hz, 1.5 × (CH_3CH_2)$_3$N); ^{13}C NMR (125 MHz, D_2O, δ_c) 97.9 (d, *J* = 5.7 Hz, C-1), 84.5 (C-3), 81.6 (C-4), 77.4 (d, *J* = 7.7 Hz, C-2), 73.3 (C-5), 63.2 (C-6), 58.8 (O*C*H$_3$), 47.5 (1.5 × (CH_3CH_2)$_3$N), 9.1 (1.5 × (CH_3CH_2)$_3$N); HRMS (ESI) *m/z* Calcd. for $(M–H)^-$ $C_7H_{14}O_9P$: 273.0370. Found: 273.0372.



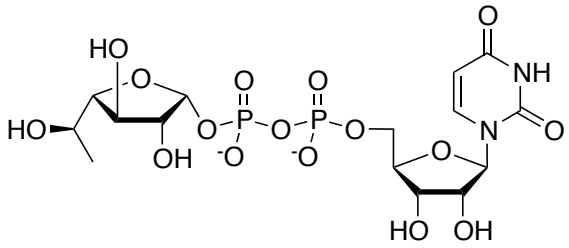
2-*O*-Methyl-α-D-galactofuranose-1-phosphate triethylammonium salt (3.37).

Compound **3.29** (302 mg, 0.39 mmol) was deprotected following the general deprotection procedure and lyophilized to give **3.37** (112 mg, 60%) as a colourless syrup. R_f 0.27 (10:2:1 CH_3OH–NH_4OH–H_2O); $[\alpha]_D$ +47.8 (*c* 0.3, H_2O); ^{1}H NMR (400 MHz, D_2O, δ_H) 5.63 (dd, 1 H, *J* = 4.2, 4.2 Hz, H-1), 4.26 (dd, 1 H, *J* = 8.2, 7.2 Hz, H-3), 3.86 (ddd, 1 H, *J* = 8.2, 4.2, 2.2 Hz, H-2), 3.78 (dd, 1 H, *J* = 7.2, 4.5 Hz, H-4), 3.70 (dt, 1 H, *J* = 7.2, 4.5 Hz, H-5), 3.65 (dd, 1 H, *J* = 11.6, 4.5 Hz, H-6a), 3.59 (dd, 1 H, *J* = 11.6, 7.2 Hz, H-6b), 3.45 (s, 3 H, OC*H*$_3$), 3.16 (q, 10 H, *J* = 7.3, 1.6 × $(CH_3CH_2)_3N$), 1.25 (t, 15 H, *J* = 7.3, 1.6 × $(CH_3CH_2)_3N$); ^{13}C NMR (100 MHz, D_2O, δ_c) 95.8 (d, *J* = 5.7 Hz, C-1), 85.9 (d, *J* = 7.9 Hz, C-2), 82.3 (C-4), 73.6 (C-3), 72.6 (C-5), 63.2 (C-6), 58.7 (O*C*H$_3$), 47.4 ($(CH_3CH_2)_3N$), 9.1 ($(CH_3CH_2)_3N$); HRMS (ESI) *m/z* Calcd. for $(M–H)^–$ $C_7H_{14}O_9P$: 273.0370. Found: 273.0377.

General procedure for chemo-enzymatic synthesis of UDP-Gal*f* analogs

UDP-glucose pyrophosphorylase (GalU) and resin immobilized-galactose-1-phosphate uridylyltransferase (GalPUT) were prepared as previously described.[26, 27, 46] To a solution of the Gal*f*-1-phosphate analog (10.5 mg, 22 μmol) in 50 mM HEPES buffer pH 8.0 containing 10 mM $MgCl_2$ and 5 mM

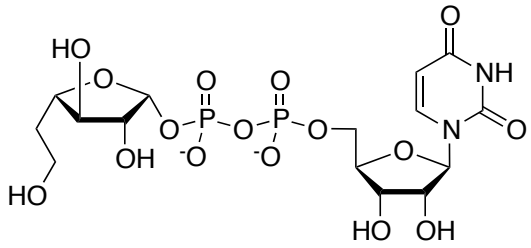
KCl, was added UTP (12.2 mg, 20 μmol), GalU (10 U), inorganic pyrophosphatase (IPP, 2 U), and immobilized GalPUT (0.6 mL, ~15 U) for a final volume of 1 mL. The reaction was initiated by the addition of UDP-Glc (91 μg, 0.15 μmol) and incubated at ambient temperature under a N_2(g) atmosphere with gentle rotation. After 1–3 days, when analysis of the reaction by HPLC[28] indicated the complete consumption of UTP, the resin bound and soluble proteins were removed by transferring the reaction mixture to a BD column cartridge, washing with Milli-Q water (3-5 mL). The flow through was filtered using a centrifugal filter device with a molecular weight cut off of 10,000 Da. The resulting filtrate was purified by semi-preparative HPLC and gel filtration chromatography as previously described[28] to give the final UDP-Gal*f* analogs as lyophilized white powders.



Uridine 5′-diphospho-6′′-deoxy-α-D-galactofuranose (3.5):

From **3.30** using the general chemo-enzymatic procedure gave **3.5** (9.7 mg, 78%) as a white powder. ^{1}H NMR (500 MHz, D_2O, δ_H) 7.95 (d, 1 H, *J* = 8.2 Hz, H-6), 5.98–5.95 (m, 1 H, H-1′), 5.96 (d, 1 H, *J* = 8.2 Hz, H-5), 5.62 (dd, 1 H, *J* = 5.9, 4.3 Hz, H-1′′), 4.38–4.32 (m, 2 H, H-2′/H-3′), 4.29–4.25 (m, 1H, H-4′), 4.23 (ddd, 1 H, *J* = 11.8, 4.4, 2.6 Hz, H-5′a), 4.18 (ddd, 1 H, *J* = 11.8, 5.7, 2.8

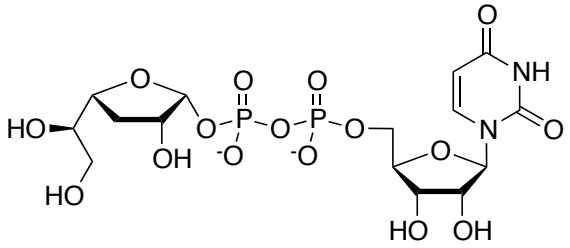
Hz, H-5′b), 4.12 (ddd, 1 H, *J* = 8.1, 4.2, 2.3 Hz, H-2′′), 4.05 (dd, 1 H, *J* = 8.1, 6.9 Hz, H-3′′), 3.85 (dq, 1 H, *J* = 6.8, 6.6 Hz, H-5′′), 3.60 (dd, 1 H, *J* = 6.9, 6.8 Hz, H-4′′), 1.18 (d, 3 H, *J* = 6.6 Hz, H-6′′); ^{13}C NMR (125 MHz, D_2O, δ_c) 167.1 (C-4), 152.7 (C-2), 142.5 (C-5), 103.6 (C-6), 98.6 (d, 1 C, *J* = 6.5 Hz, C-1′′), 89.3 (C-1′), 87.0 (C-4′′), 84.2 (d, 1 C, *J* = 9.2 Hz, C-4′), 78.1 (d, 1 C, *J* = 7.8 Hz, C-2′′), 75.4, 70.6 (C-2′/C-3′), 74.7 (C-3′′), 70.1 (C-5′′), 65.8 (d, 1 C, *J* = 6.0 Hz, C-5′), 18.2 (C-6′′); ESIMS *m/z* 549 ([M–H]$^-$, 47%), 274 ([M–2H]$^{2-}$, 100%); HRMS (ESI) *m/z* Calcd. for (M–H)$^-$ $C_{15}H_{23}N_2O_{16}P_2$: 549.0528. Found: 549.0527.



Uridine 5′-diphospho-5′′-deoxy-β-L-*arabino*-hexofuranose (3.6):

Using the general chemo-enzymatic procedure above, **3.31** was used to obtain **3.6** (7.5 mg, 61%) as a white powder. ^{1}H NMR (700 MHz, D_2O, δ_H) 7.96 (d, 1 H, *J* = 8.2 Hz, H-6), 6.00–5.98 (m, 1 H, H-1′), 5.98 (d, 1 H, *J* = 8.2 Hz, H-5), 5.63 (dd, 1 H, *J* = 5.5, 4.8 Hz, H-1′′), 4.39–4.37 (m, 2 H, H-2′/H-3′), 4.30–4.28 (m, 1 H, H-4′), 4.24 (ddd, 1 H, *J* = 11.8, 4.4, 2.6 Hz, H-5′a), 4.20 (ddd, 1 H, *J* = 11.8, 5.6, 2.8 Hz, H-5′b), 4.11 (ddd, 1 H, *J* = 8.0, 4.3, 2.2 Hz, H-2′′), 4.06 (dd, 1 H, *J* = 7.7, 7.2 Hz, H-3′′), 3.93 (ddd, 1 H, *J* = 7.4, 7.2, 5.3 Hz, H-4′′), 3.78–3.71 (m, 2 H, H-6′′a/H-6′′b), 1.99–1.95 (m, 2 H, H-5′′a/H-5′′b); ^{13}C NMR (175 MHz, D_2O, δ_C) 167.0 (C-4), 152.6 (C-2), 142.5 (C-5), 103.5 (C-6), 98.8 (d,

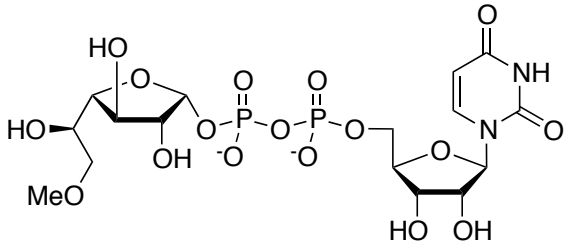
1 C, $J_{1'',P}$ = 6.0 Hz, C-1″), 89.2 (C-1′), 84.1 (d, 1 C, *J* = 8.8 Hz, C-4′), 80.5 (C-4″), 78.5 (C-3″), 77.5 (d, 1 C, *J* = 7.9 Hz, C-2″), 74.6, 70.5 (C-2′/C-3′), 65.7 (d, 1 C, $J_{5',P}$ = 3.5 Hz, C-5′), 59.2 (C-6″), 37.6 (C-5″); ESIMS *m/z* 549 ([M–H]$^-$, 83%), 274 ([M–2H]$^{2-}$, 100%); HRMS (ESI) *m/z* Calcd. for (M–H)$^-$ $C_{15}H_{23}N_2O_{16}P_2$: 549.0528. Found: 549.0527.



Uridine 5′-diphospho-3″-deoxy-α-D-*xylo*-hexofuranose (3.7):

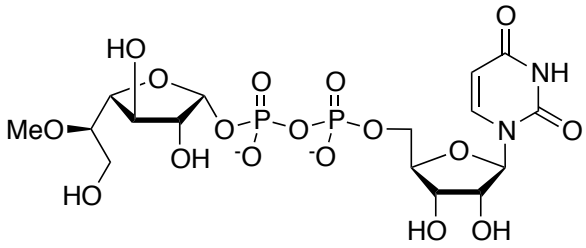
The general chemo-enzymatic procedure with **3.32** was used to give **3.7** (3.7 mg, 31%) as a white solid; ^{1}H NMR (700 MHz, D_2O, δ_H) 7.96 (d, 1 H, *J* = 8.1 Hz, H-6), 6.00–5.98 (m, 1 H, H-1′), 5.98 (d, 1 H, *J* = 8.1 Hz, H-5), 5.60 (dd, 1 H, *J* = 5.3, 4.3 Hz, H-1″), 4.39–4.36 (m, 2 H, H-2′, H-3′), 4.35–4.31 (m, 1 H, H-2″), 4.30–4.28 (m, 1H, H-4′), 4.25 (ddd, 1 H, *J* = 11.8, 4.5, 2.6 Hz, H-5′a), 4.21 (ddd, 1 H, *J* = 11.8, 5.7, 2.8 Hz, H-5′b), 4.10 (ddd, 1 H, *J* = 9.9, 6.5, 6.5 Hz, H-4″), 3.70 (ddd, 1 H, *J* = 6.8, 6.5, 3.7 Hz, H-5″), 3.66 (dd, 1 H, *J* = 12.0, 3.7 Hz, H-6″a), 3.56 (dd, 1 H, *J* = 12.0, 6.8 Hz, H-6″b), 2.31–2.29 (m, 1 H, H-3″a), 1.83 (app. q, 1 H, *J* = 10.9, H-3″b); ^{13}C NMR (175 MHz, D_2O, δ_C) 167.1 (C-4), 152.7 (C-2), 142.5 (C-5), 103.5 (C-6), 98.5 (d, 1 C, *J* = 6.0 Hz, C-1″), 89.2 (C-1′), 84.1 (d, 1 C, *J* = 9.1 Hz, C-4′), 79.7 (C-4″), 75.4 (C-5″), 74.6, 70.5 (C-2′, C-3′), 72.6 (d, 1 C, *J* = 8.2 Hz, C-2″), 65.7 (d, 1 C, *J* = 5.3 Hz, C-5′), 63.2 (C-6″),

31.5 (C-3′′); MS (ESI) *m/z* 549 ([M – H]$^{-}$, 30%), 274 ([M – 2H]$^{2-}$, 100%); HRMS (ESI) *m/z* Calcd for (M–H)$^{-}$ $C_{15}H_{23}N_2O_{16}P_2$: 549.0528. Found: 549.0529.



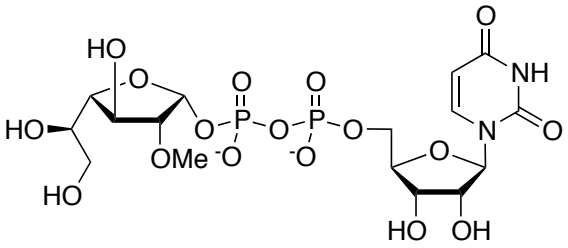
Uridine 5′-diphospho-6′′-*O*-methyl-α-D-galactofuranose (3.8):

The general chemo-enzymatic procedure with **3.34** was used to give **3.8** (~0.5 mg, <5%) as a white solid. ^{1}H NMR (700 MHz, D_2O, δ_H) 7.96 (d, 1 H, *J* = 8.1 Hz, H-6), 6.00–5.98 (m, 1 H, H-1′), 5.98 (d, 1 H, *J* = 8.1 Hz, H-5), 5.64 (dd, 1 H, *J* = 5.2, 4.2 Hz, H-1′′), 4.38–4.36 (m, 2 H, H-2′, H-3′), 4.30–4.27 (m, 1H, H-4′), 4.24 (ddd, 1 H, *J* = 11.8, 4.2, 2.6 Hz, H-5′a), 4.22–4.18 (m, 2 H, H-5′b, H-3′′), 4.15 (ddd, 1 H, *J* = 8.3, 4.2, 2.3 Hz, H-2′′), 3.90–3.87 (m, 1 H, H-5′′), 3.80 (dd, 1 H, *J* = 7.0, 6.1 Hz, H-4′′), 3.60 (dd, 1 H, *J* = 10.8, 3.9 Hz, H-6′′a), 3.53 (dd, 1 H, *J* = 10.8, 7.2 Hz, H-6′′b), 3.39 (s, 3 H, OC*H*$_3$); MS (ESI) *m/z* 579 ([M–H]$^{-}$, 5.6%), 289 ([M–2H]$^{2-}$, 100%); HRMS (ESI) *m/z* Calcd. for (M–2H)$^{2-}$ $C_{16}H_{24}N_2O_{17}P_2$: 289.0281. Found 289.0281.



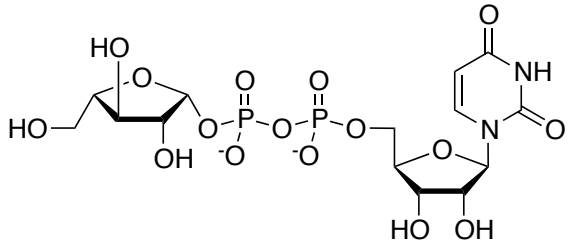
Uridine 5′-diphospho-5′′-*O*-methyl-α-D-galactofuranose (3.9).

Using the general chemo-enzymatic procedure with **3.35**, compound **3.9** (~0.5 mg, <5%) was obtained as a white solid. ^{1}H NMR (700 MHz, D_2O, δ_H) 7.96 (d, 1 H, *J* = 8.1 Hz, H-6), 6.00–5.98 (m, 1 H, H-1′), 5.98 (d, 1 H, *J* = 8.1 Hz, H-5), 5.67 (dd, 1 H, *J* = 6.7, 3.8 Hz, H-1′′), 4.39–4.37 (m, 2 H, H-2′, H-3′), 4.29–4.28 (m, 1 H, H-4′), 4.24 (ddd, 1 H, *J* = 11.8, 4.4, 2.6 Hz, H-5′a), 4.20 (ddd, 1 H, *J* = 11.8, 5.6, 2.8 Hz, H-5′a), 4.16–4.12 (m, 2 H, H-2′′, H-3′′), 3.88–3.84 (m, 2 H, H-4′′, H-6′′a), 3.62 (dd, 1 H, *J* = 12.5, 5.8 Hz, H-6′′b), 3.55 (s, 3 H, OC*H*$_3$), 3.55–3.50 (m, 1 H, H-5′′); ^{13}C NMR (175 MHz, D_2O, δ_C) 167.1 (C-4), 142.5 (C-5), 103.5 (C-6), 98.7 (d, *J* = 5.6 Hz, C-1′′), 89.2 (C-1′), 84.2 (C-5′′), 84.1 (d, *J* = 8.9 Hz, H-4′), 82.1 (C-4′′), 77.6 (d, *J* = 7.1 Hz, H-2′′), 75.1 (C-3′′), 74.6, 70.5 (C-2′, C-3′), 65.7 (d, *J* = 5.9 Hz, H-5′), 60.2 (C-6′′), 59.4 (O*C*H$_3$); MS (ESI) *m/z* 579 ([M – H]$^-$, 18%), 289 ([M – 2H]$^{2-}$, 100%); HRMS (ESI) *m/z* Calcd. for (M – 2H)$^{2-}$ $C_{16}H_{24}N_2O_{17}P_2$: 289.0281. Found: 289.0281.



Uridine 5′-diphospho-2′′-*O*-methyl-α-D-galactofuranose (3.11):

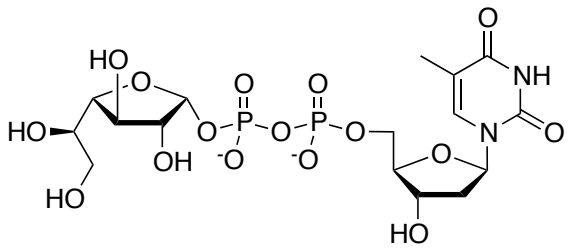
From **3.37** using the chemo-enzymatic procedure, **3.11** (~0.05 mg, <5%) was obtained as a white solid. ^{1}H NMR (700 MHz, D_2O, δ_H) 7.95 (d, 1 H, J = 8.1 Hz, H-6), 5.98–5.97 (m, 1 H, H-1′), 5.97 (d, 1 H, J = 8.1 Hz, H-5), 5.78 (dd, 1 H, J = 5.5, 4.2 Hz, H-1′′), 4.38–4.35 (m, 2 H, H-2′, H-3′), 4.28–4.26 (m, 2 H, H-4′, H-3′′), 4.23 (ddd, 1 H, J = 11.8, 4.5, 2.6 Hz, H-5′a), 4.19 (ddd, 1 H, J = 11.8, 5.7, 2.9 Hz, H-5′a), 3.94 (ddd, 1 H, J = 8.5, 4.1, 2.6 Hz, H-2′′), 3.82 (dd, 1 H, J = 7.4, 5.2 Hz, H-4′′), 3.77–3.73 (m, 1 H, H-5′′), 3.69 (dd, 1 H, J = 11.8, 4.4 Hz, H-6′′a), 3.62 (dd, 1 H, J = 11.8, 7.2 Hz, H-6′′b), 3.49 (s, 3 H, OCH_3); HRMS (ESI) *m/z* Calcd. for $(M–H)^-$ $C_{16}H_{25}N_2O_{17}P_2$: 579.0634. Found: 579.0634.



Uridine 5′-diphospho-β-L-arabinofuranose (3.12)

Compound **3.12** (6.9 mg, 58%) was prepared using the GalPUT-promoted reaction as described above and isolated as a lyophilized white powder. ^{1}H NMR (700 MHz, D_2O, δ_H) 7.96 (d, 1 H, $J_{6,5}$ = 8.1 Hz, H-6), 6.00–

5.99 (m, 1 H, H-1′), 5.98 (d, 1 H, *J* = 8.1 Hz, H-5), 5.64 (dd, 1 H, *J* = 5.7, 3.9 Hz, H-1′′), 4.39–4.36 (m, 2 H, H-2′/H-3′), 4.30–4.28 (m, 1H, H-4′), 4.24 (ddd, 1 H, *J* = 11.8, 4.4, 2.6 Hz, H-5′a), 4.20 (ddd, 1 H, *J* = 11.8, 5.6, 2.8 Hz, H-5′b), 4.17–4.13 (m, 2 H, H-2′′/H-3′′), 3.92 (ddd, 1 H, *J* = 6.6, 6.1, 3.1 Hz, H-4′′), 3.80 (dd, 1 H, *J* = 12.7, 3.1 Hz, H-5′′a), 3.70 (dd, 1 H, *J* = 12.7, 6.1 Hz, H-5′′b); ^{13}C NMR (175 MHz, D_2O, δ_c) 167.1 (C-4), 152.7 (C-2), 142.4 (C-5), 103.5 (C-6), 98.6 (d, 1 C, *J* = 7.0 Hz, C-1′′), 89.2 (C-1′), 84.1 (d, 1 C, *J* = 8.9 Hz, C-4′), 83.5 (C-4′′), 77.5 (d, 1 C, *J* = 8.6 Hz, C-2′′), 74.6, 70.6 (C-2′/C-3′), 74.1 (C-3′′), 65.8 (d, 1 C, *J* = 3.7 Hz, C-5′), 63.1 (C-5′′); MS (ESI) *m/z* 535 ([M–H]$^-$, 30%), 267 ([M–2H]$^{2-}$, 100%); HRMS (ESI) *m/z* Calcd. for (M – H)$^-$ $C_{15}H_{23}N_2O_{16}P_2$: 535.0372. Found: 535.0371.



Chemo-enzymatic synthesis of TDP-Gal*f* (3.13)

Resin immobilized Cps2L protein was prepared from *E. coli* BL21 (DE3) cells containing recombinant pSK001 plasmid.[36] Cells were grown in LB broth (1 L) supplemented with 25 μg/mL kanamycin. Production of Cps2L was induced by the addition of 375 μM IPTG (isopropyl 1-thio-β-D-galactopyranoside) at a OD_{600} of 0.6 followed by incubation at 30 °C for 4 h. Cells were collected by centrifugation at 11,300gmax for 15 min and the pellets

were then re-suspended in 40 mL of resuspention buffer (20 mM Tris-HCl, pH 8.0, containing 300 mM NaCl and 10 mM imidazole). The cells were lysed using a benchtop cell disruptor (Constant Systems Inc., NC) set to 20 Kpsi and the lysate clarified by centrifugation (105,000 × *g* for 1 h at 4 °C). The lysate was applied to a 5 mL Ni–NTA agarose column and washed with 6 column volumes of wash buffer (20 mM Tris-HCl, pH 8.0, containing 300 mM NaCl and 25 mM imidazole), followed by 6 column volumes of reaction buffer (20 mM Tris-HCl, pH 8.0, containing 300 mM NaCl).

To a solution of Gal*f*-1-phosphate (5 mg, 11 μmol) in reaction buffer containing 2.5 mM $MgCl_2$ was added dTTP (4.8 mg, 10 μmol), IPP (3.75 U), and immobilized Csp2L protein (0.5 mL) for a final volume of 0.6 mL. After incubating for 48 h at ambient temperature under a atmosphere of N_2(g) with gentle rotation, analysis of the reaction by HPLC indicated complete consumption of dTTP. The reaction was incubated for 5 h with alkaline phosphatase (AP, 10 U) to degrade unwanted TDP and dTMP in the reaction mixture. The resin bound Cps2L was again removed by transferring the reaction mixture to a BD column cartridge and washing with Milli-Q water (3-5 mL). Soluble IPP and AP proteins were removed by filtration of the resulting flow through using a centrifugal filter device with a molecular weight cut off of 10,000 Da. The filtrate was purified by reverse phase semi-preparative HPLC using conditions previously described for the purification of UDP-Gal*f*.[27, 28] Purified HPLC fractions were combined, the volume was reduced to 5 mL by evaporation under reduced pressure, and the salts were removed by gel filtration

chromatography (Sephadex G-15) eluting with Milli-Q water at a flow rate of 1 mL per min. Fractions containing the purified product were combined and lyophilized to give TDP-Gal*f* (**3.13**, 3.4 mg, 50%) as a white powder. ^{1}H NMR (600 MHz, D_2O, δ_H) 7.72 (br d, 1 H, *J* = 1.2 Hz, H-6), 6.33 (dd, 1 H, *J* = 7.5, 6.5 Hz, H-1′), 5.61 (dd, 1 H, *J* = 5.5, 4.5 Hz, H-1′′), 4.60 (app. dt, 1 H, *J* = 5.9, 3.0 Hz, H-3′), 4.21 (dd, 1 H, *J* = 8.4, 7.4 Hz, H-3′′), 4.18–4.14 (m, 3H, H-4′, H-5′a, H-5′b), 4.12 (ddd, 1 H, *J* = 8.4, 4.3, 2.4 Hz, H-2′′), 3.80 (dd, 1 H, *J* = 7.4, 5.2 Hz, H-4′′), 3.75 (ddd, 1 H, *J* = 7.2, 5.2, 4.3 Hz, H-5′′), 3.69 (dd, 1 H, *J* = 11.8, 4.3 Hz, H-6′′a), 3.61 (dd, 1 H, *J* = 11.8, 7.2 Hz, H-6′′b), 2.39-2.32 (m, 2 H, H-2′a, H-2′b); MS (ESI) *m/z* 549 ($[M - H]^-$, 47%), 274 ($[M - 2H]^{2-}$, 100%); HRMS (ESI) *m/z* Calcd. for $(M - H)^- C_{16}H_{25}N_2O_{16}P_2$: 563.0685. Found: 563.0684.

3.5 Bibliography

1. Allos, B. M., *Clin. Infect. Dis.* **2001,** *32*, 1201-1206.
2. Kaldor, J.; Speed, B. R., *Br. Med. J.* **1984,** *288*, 1867-1870.
3. Moran, A. P.; Appelmelk, B. J.; Aspinall, G. O., *J. Endotoxin Res.* **1996,** *3*, 521-531.
4. Moran, A. P.; Prendergast, M. M.; Appelmelk, B. J., *FEMS Immunol. Med. Microbiol.* **1996,** *16*, 105-115.
5. Jones, M. A.; Marston, K. L.; Woodall, C. A.; Maskell, D. J.; Linton, D.; Karlyshev, A. V.; Dorrell, N.; Wren, B. W.; Barrow, P. A., *Infect. Immun.* **2004,** *72*, 3769-3776.

6. Bacon, D. J.; Szymanski, C. M.; Burr, D. H.; Silver, R. P.; Alm, R. A.; Guerry, P., *Mol. Microbiol.* **2001,** *40*, 769-777.

7. Karlyshev, A. V.; Linton, D.; Gregson, N. A.; Lastovica, A. J.; Wren, B. W., *Mol. Microbiol.* **2000,** *35*, 529-541.

8. Karlyshev, A. V.; Champion, O. L.; Churcher, C.; Brisson, J. R.; Jarrell, H. C.; Gilbert, M.; Brochu, D.; St Michael, F.; Li, J. J.; Wakarchuk, W. W.; Goodhead, I.; Sanders, M.; Stevens, K.; White, B.; Parkhill, J.; Wren, B. W.; Szymanski, C. M., *Mol. Microbiol.* **2005,** *55*, 90-103.

9. Hanniffy, O. M.; Shashkov, A. S.; Moran, A. P.; Prendergast, M. M.; Senchenkova, S. N.; Knirel, Y. A.; Savage, A. V., *Carbohydr. Res.* **1999,** *319*, 124-132.

10. St Michael, F.; Szymanski, C. M.; Li, J. J.; Chan, K. H.; Khieu, N. H.; Larocque, S.; Wakarchuk, W. W.; Brisson, J. R.; Monteiro, M. A., *Eur. J. Biochem.* **2002,** *269*, 5119-5136.

11. Nassau, P. M.; Martin, S. L.; Brown, R. E.; Weston, A.; Monsey, D.; McNeil, M. R.; Duncan, K., *J. Bacteriol.* **1996,** *178*, 1047-1052.

12. Poulin, M. B.; Nothaft, H.; Hug, I.; Feldman, M. F.; Szymanski, C. M.; Lowary, T. L., *J. Biol. Chem.* **2010,** *285*, 493-501.

13. Sun, H. G.; Ruszczycky, M. W.; Chang, W. C.; Thibodeaux, C. J.; Liu, H. W., *J. Biol. Chem.* **2012,** *287*, 4602-4608.

14. Soltero-Higgin, M.; Carlson, E. E.; Gruber, T. D.; Kiessling, L. L., *Nat. Struct. Mol. Biol.* **2004,** *11*, 539-543.

15. Mayer, M.; Meyer, B., *J. Am. Chem. Soc.* **2001,** *123*, 6108-6117.

16. Angulo, J.; Nieto, P. M., *European Biophysics Journal with Biophysics Letters* **2011,** *40*, 1357-1369.

17. Yuan, Y.; Wen, X.; Sanders, D. A. R.; Pinto, B. M., *Biochemistry* **2005,** *44*, 14080-14089.

18. Yuan, Y.; Bleile, D. W.; Wen, X.; Sanders, D. A. R.; Itoh, K.; Liu, H. W.; Pinto, B. M., *J. Am. Chem. Soc.* **2008,** *130*, 3157-3168.

19. Gruber, T. D.; Westler, W. M.; Kiessling, L. L.; Forest, K. T., *Biochemistry* **2009,** *48*, 9171-9173.

20. Gruber, T. D.; Borrok, M. J.; Westler, W. M.; Forest, K. T.; Kiessling, L. L., *J. Mol. Biol.* **2009,** *391*, 327-340.

21. Partha, S. K.; van Straaten, K. E.; Sanders, D. A. R., *J. Mol. Biol.* **2009,** *394*, 864-877.

22. Spohr, U.; Hindsgaul, O.; Lemieux, R. U., *Can. J. Chem.* **1985,** *63*, 2644-2652.

23. Hindsgaul, O.; Kaur, K. J.; Srivastava, G.; Blaszczykthurin, M.; Crawley, S. C.; Heerze, L. D.; Palcic, M. M., *J. Biol. Chem.* **1991,** *266*, 17858-17862.

24. Laferte, S.; Chan, N. W. C.; Sujino, K.; Lowary, T. L.; Palcic, M. M., *Eur. J. Biochem.* **2000,** *267*, 4840-4849.

25. Chad, J. M.; Sarathy, K. P.; Gruber, T. D.; Addala, E.; Kiessling, L. L.; Sanders, D. A. R., *Biochemistry* **2007,** *46*, 6723-6732.

26. Errey, J. C.; Mukhopadhyay, B.; Kartha, K. P. R.; Field, R. A., *Chem. Commun.* **2004**, 2706-2707.

27. Rose, N. L.; Zheng, R. B.; Pearcey, J.; Zhou, R.; Completo, G. C.; Lowary, T. L., *Carbohydr. Res.* **2008,** *343*, 2130-2139.

28. Poulin, M. B.; Lowary, T. L., *Methods Enzymol.* **2010,** *478*, 389-411.

29. de Lederkremer, R. M.; Nahmad, V. B.; Varela, O., *J. Org. Chem.* **1994,** *59*, 690-692.

30. Wolfenden, R.; Lu, X. D.; Young, G., *J. Am. Chem. Soc.* **1998,** *120*, 6814-6815.

31. Marlow, A. L.; Kiessling, L. L., *Org. Lett.* **2001,** *3*, 2517-2519.

32. Zhang, Q. B.; Liu, H. W., *J. Am. Chem. Soc.* **2000,** *122*, 9065-9070.

33. Tsvetkov, Y. E.; Nikolaev, A. V., *J. Chem. Soc., Perkin Trans. 1* **2000**, 889-891.

34. Peltier, P.; Daniellou, R.; Nugier-Chauvin, C.; Ferrières, V., *Org. Lett.* **2007,** *9*, 5227-5230.

35. Peltier, P.; Beláňová, M.; Dianišková, Petronela; Zhou, R.; Zheng, R. B.; Pearcey, J. A.; Joe, M.; Brennan, P. J.; Nugier-Chauvin, C.; Ferrières, V.; Lowary, T. L.; Daniellou, R.; Mikušová, K., *Chem. Biol.* **2010,** *17*, 1356-1366.

36. Timmons, S. C.; Mosher, R. H.; Knowles, S. A.; Jakeman, D. L., *Org. Lett.* **2007,** *9*, 857-860.

37. Timmons, S. C.; Hui, J. P. M.; Pearson, J. L.; Peltier, P.; Daniellou, R.; Nugier-Chauvin, C.; Soo, E. C.; Syvitski, R. T.; Ferrieres, V.; Jakeman, D. L., *Org. Lett.* **2008,** *10*, 161-163.

38. Bernet, B.; Piantini, U.; Vasella, A., *Carbohydr. Res.* **1990,** *204*, 11-25.

39. Schneider, H. J.; Hoppen, V., *J. Org. Chem.* **1978,** *43*, 3866-3873.

40. Errey, J. C.; Mann, M. C.; Fairhurst, S. A.; Hill, L.; McNeil, M. R.; Naismith, J. H.; Percy, J. M.; Whitfield, C.; Field, R. A., *Org. Biomol. Chem.* **2009,** *7*, 1009-1016.

41. Sanders, D. A. R.; Staines, A. G.; McMahon, S. A.; McNeil, M. R.; Whitfield, C.; Naismith, J. H., *Nat. Struct. Biol.* **2001,** *8*, 858-863.

42. Beis, K.; Srikannathasan, V.; Liu, H.; Fullerton, S. W. B.; Bamford, V. A.; Sanders, D. A. R.; Whitfield, C.; McNeil, M. R.; Naismith, J. H., *J. Mol. Biol.* **2005,** *348*, 971-982.

43. Fullerton, S. W. B.; Daff, S.; Sanders, D. A. R.; Ingledew, W. J.; Whitfield, C.; Chapman, S. K.; Naismith, J. H., *Biochemistry* **2003,** *42*, 2104-2109.

44. Claasen, B.; Axmann, M.; Meinecke, R.; Meyer, B., *J. Am. Chem. Soc.* **2005,** *127*, 916-919.

45. Hanessian, S.; Lu, P.-P.; Ishida, H., *J. Am. Chem. Soc.* **1998,** *120*, 13296-13300.

46. Liu, Z. Y.; Zhang, J. B.; Chen, X.; Wang, P. G., *ChemBioChem* **2002,** *3*, 348-355.

Chapter 4

Exploring the Activity of Three Putative Pyranose–Furanose Mutase Enzymes in *Campylobacter jejuni* Serotype HS:41

Cloning of the *Campylobacter jejuni glf* genes was performed by Harald Nothaft and Bernadette Beadle from the group of our collaborator Christine Szymanski.

4.1 Introduction

UDP-D-galactopyranose mutase (UGM), encoded by the *glf* gene, is involved in the biosynthesis of galactofuranose (Gal*f*).[1] UGM, as has been discussed in previous chapters, is a flavoprotein that catalyzes the isomerization of UDP-D-galactopyranose (UDP-D-Gal*p*, **4.1**) into UDP-D-galactofuranose (UDP-D-Gal*f*, **4.2**), the precursor of the D-Gal*f* residues found in the glycoconjugates of numerous bacteria, fungi, and parasites.[2] More recent work has found that homologs of *glf* in *Campylobacter jejuni* 11168 and *Escherichia coli* O52, which produce no Gal*f* containing glycocongugates, instead encode pyranose–furanose mutase enzymes involved in the biosynthesis of UDP-2-acetamido-2-deoxy-D-galactofuranose (UDP-D-Gal*f*NAc),[3] and TDP-D-fucofuranose (TDP-D-Fuc*f*),[4] respectively. However, many of these *glf* homologs remain poorly characterized and the substrate specificity of the encoded enzymes is unknown.

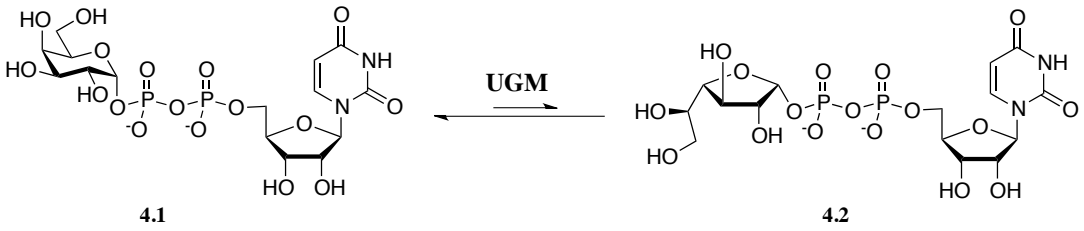


Figure 4-1. UGM catalyzed interconversion of UDP-D-Gal*p* (**4.1**) to UDP-D-Gal*f* (**4.2**).

Campylobacter jejuni, the leading cause of human gastroenteritis,[5] is known to possess numerous structurally diverse sugars in its glycoconjugates,[6] including sugars in the furanose ring conformation.[7, 8] In particular, the serotype

HS:41 strains, obtained from clinical isolates of patients with the neurological disorder Guillian–Barré syndrome (GBS),[9] produce a capsular polysaccharide (CPS) composed entirely of furanose sugars.[7] It is the lipooligosaccharide (LOS) and not the CPS, which is believed to be associated with GBS,[9] but the CPS remains a key virulence factor of the bacterium.[6] The CPS structure of this serotype contains two trisaccharide repeating units composed of 6-deoxy-D-*altro*-heptofuranose (6d-D-*altro*-Hep*f*), L-arabinofuranose (L-Ara*f*) and either 6-deoxy-L-altrofuranose (6d-L-Alt*f*), or D-Fuc*f* (Figure 4-2). In the original characterization of the CPS, the two trisaccharides made up 75% and 25% of the repeating units, respectively, which differ only the presence of either 6d-L-Alt*f* or D-Fuc*f* at the reducing end of the repeat. It is not clear whether both trisaccharide repeating unit make up a single CPS structure, or if they are found in separate polysaccharides.[7]

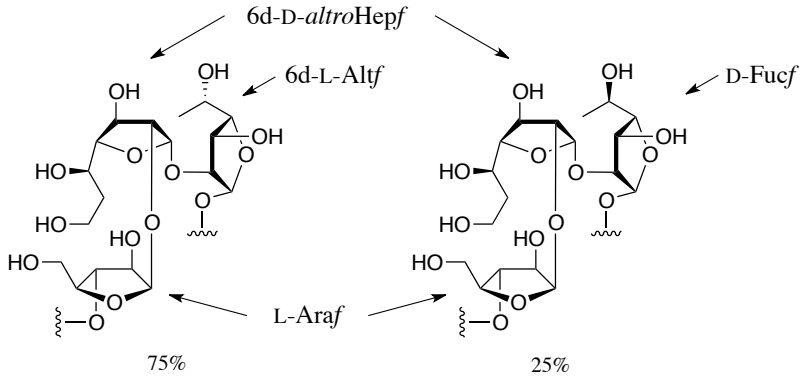


Figure 4-2. CPS repeating unit of *C. jejuni* serotype HS:41. The two trisaccharide units contain only furanose sugars and make up 75% and 25% of the CPS, respectively.

Although the structure is known, the enzymes responsible for the biosynthesis of the furanose sugars in the CPS of *C. jejuni* HS:41 have not been

elucidated. Those genes encoding for the CPS biosynthetic enzymes for this serotype have been sequenced and were found to include three genes annotated as *glf* homologs (*glf1*, *glf2*, and *glf3*), based only on their sequence homology to known bacterial UGMs.[10] However, there are no Gal*f* residues found in the glycoconjugates of these *C. jejuni* strains suggesting that the enzymes encoded by the *glf1*, *glf2*, and *glf3* genes could be involved in the biosynthesis of the furanose sugars found in the CPS repeat structure. As an extension of our work to identify the structural features of pyranose–furanose mutase enzymes responsible for substrate recognition and discrimination, we report herein the first studies to elucidate the activity of the proteins expressed by the *glf1*, *glf2*, and *glf3* genes of *C. jejuni* HS:41. Here, four of the putative furanose sugar nucleotide precursors of the 6d-D-*altro*-Hep*f*, 6d-L-Alt*f*, L-Ara*f*, and D-Fuc*f* sugars found in the CPS were synthesized and evaluated as substrates for the Glf1, Glf2 and Glf3 enzymes, which were recombinantly expressed in *E. coli*.

4.2 Results and Discussion

4.2.1 Characterization of a GDP-6d-D-*altro*-heptopyranose mutase (GaHM)

6d-*altro*-D-Hep*f* was first discovered in an antigenic polysaccharide isolate from the Gram positive *Eubacterium saburreum* strain L49,[11] and was later found in the O2,[12, 13] S29,[14] T21, and T110 strains of *E. saburreum.*[15] The presence of this monosaccharide was shown to be characteristic of the chemotype II strains of this organism.[15] Despite the widespread occurrence of

this sugar in *E. saburreum*, no work has been done to elucidate 6d-*altro*-D-Hep*f* biosynthesis in this organism, which is frequently associated with oral infections such as periodontitis.[16] In *C. jejuni* HS:41, the biosynthetic precursor of the 6d-*altro*Hep*f* is most likely a GDP sugar nucleotide, given the presence of a *hddC* (D,D-heptose-1-phosphate guanosyltransferase) gene in the CPS gene locus. The enzyme encoded by this gene was previously implicated in the biosynthesis of the GDP-heptose sugar in *C. jejuni* 11168.[10] Therefore, any pyranose–furanose mutase involed in the biosynthesis of these glycoconjugates would likely catalyze the isomerization of GDP-6-deoxy-*altro*-heptopyranose (GDP-6d-D-*altro*-Hep*p*, **4.3**) to GDP-6d-D-*altro*-Hep*f,* **4.4** (Figure 4-3). With this hypothesis in place, we synthesized GDP-6d-D-*altro*-Hep*f* as a putative substrate for the *C. jejuni* Glf enzymes.

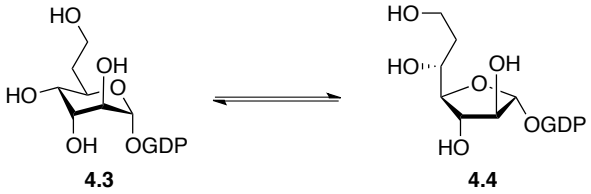
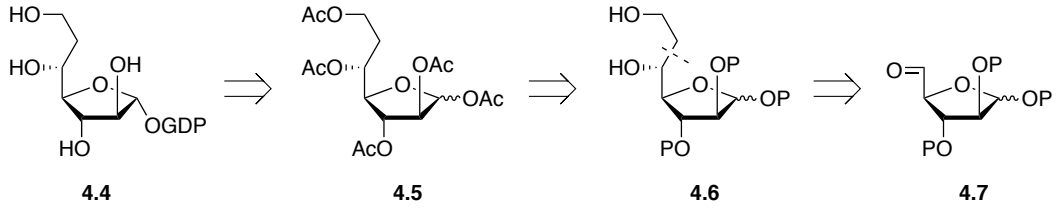


Figure 4-3. Putative reaction catalyzed by GaHM.

*4.2.1.1 Synthesis of GDP-6d-D-***altro-*Hep*f***

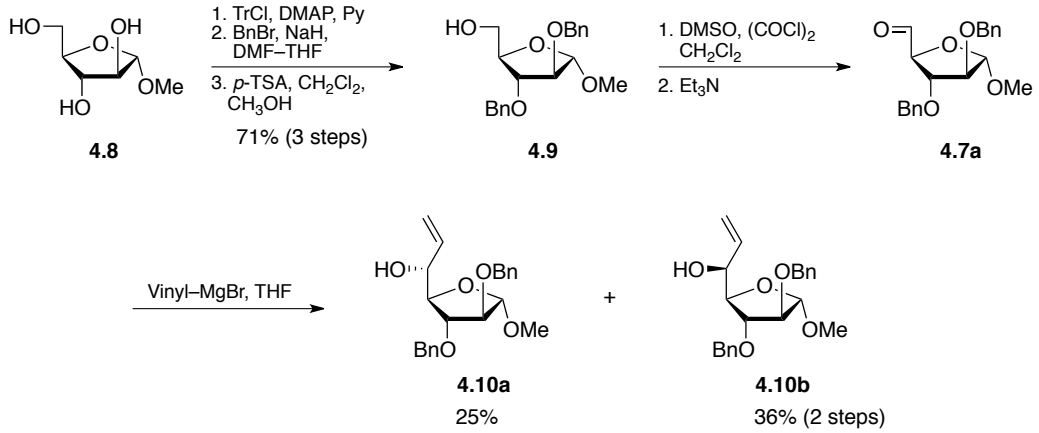
We envisioned preparing GDP-6d-D-*altro*-Hep*f* (**4.4**) from the per-*O*-acetylated 6d-D-*altro*-Hep*f* derivative **4.5** (Scheme 4-1). The stereochemistry of 6d-D-*altro*-Hep*f* is identical to that of a D-arabinofuranose (D-Ara*f*), extended by an additional two carbons at the non-reducing end. As a result, 6d-D-*altro*-Hep*f* could be derived from a suitably protected D-Ara*f* derived aldehyde **4.7** via a

route where the key step would be the diastereoselective addition of an organometallic reagent to the aldehyde to extend the carbohydrate chain.



Scheme 4-1. Retrosynthesis of GDP-6d-D-*altro*-Hep*f* (**4.4**)

Our first approach (Scheme 4-2) was to use aldehyde **4.7a** in a Grignard reaction with vinyl magnesium bromide. The aldehyde **4.7a** was synthesized from the known methyl α-D-arabinofuranoside (**4.8**),[17] by first selectively protecting the primary hydroxyl group using trityl chloride, followed by protection of the secondary hydroxyl groups as benzyl ethers. The trityl ether was then removed under acidic conditions to give the free alcohol **4.9** in 71% yield over three steps. A Swern oxidation of **4.9** yielded the desired aldehyde **4.7a**, which was used directly in the Grignard reaction with vinylmagnesium bromide to introduce the additional two carbons. We envisioned that a subsequent hydroboration–oxidation sequence on the terminal alkene could then give access to the 6d-*altro*-Hep*f*.



Scheme 4-2. Synthesis of **4.10a** and **4.10b** via a Grignard reaction with aldehyde **4.7a**.

The non-chelate addition of vinylmagnesium bromide to aldehyde **4.7a** was required to give the desired anti product (*R*-stereoisomer) **4.10a** as the major product (Figure 4-4). When the reaction was attempted, a 3:2 ratio of the diastereomers was obtained in a 61% overall yield (over two steps from **4.9**) with the *R*-stereoisomer being the minor product.

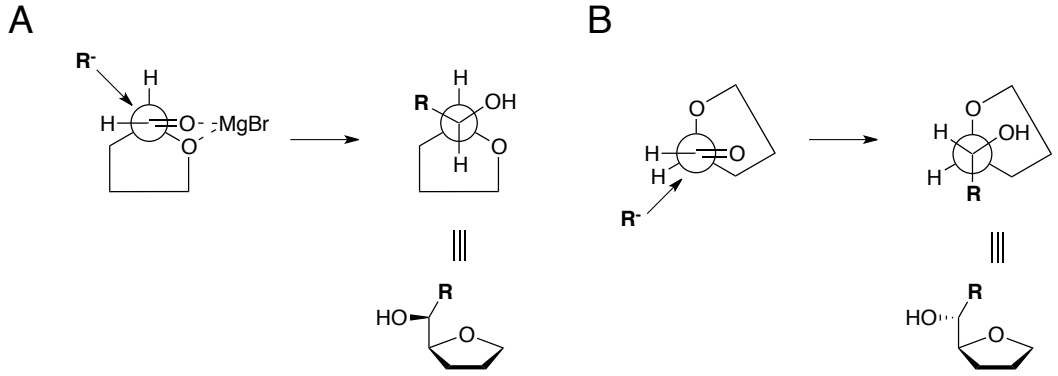
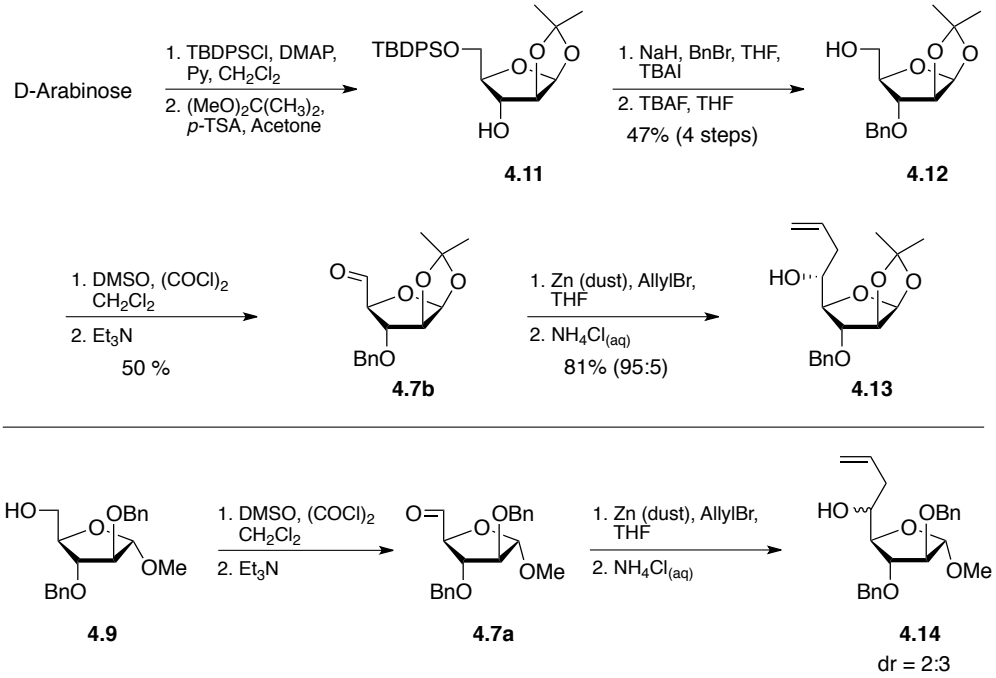


Figure 4-4. Expected diastereoselectivity of the Grignard addition to aldehyde **4.7a** via the Cram-chelate (A) or Felkin–Ahn (B) transition state.

Due to the poor stereocontrol of the Grignard reaction with aldehyde **4.7a**, we explored other reactions to extend the carbohydrate chain. In 1997, Pakulski and Zamojski reported a zinc-mediated propargylation of the D-Ara*f* derived aldehyde **4.7b** (Scheme 4-3), which proceeded with excellent diastereoselectivity (95:5) for the corresponding anti-product.[18] Zinc-mediated allylation of D-mannofuranose derived aldehydes were also reported to proceed with excellent diastereoselectivity.[19] As a result, we explored the zinc-mediated allylation of **4.7b** to access the 6d-D-*altro*-Hep*f*.

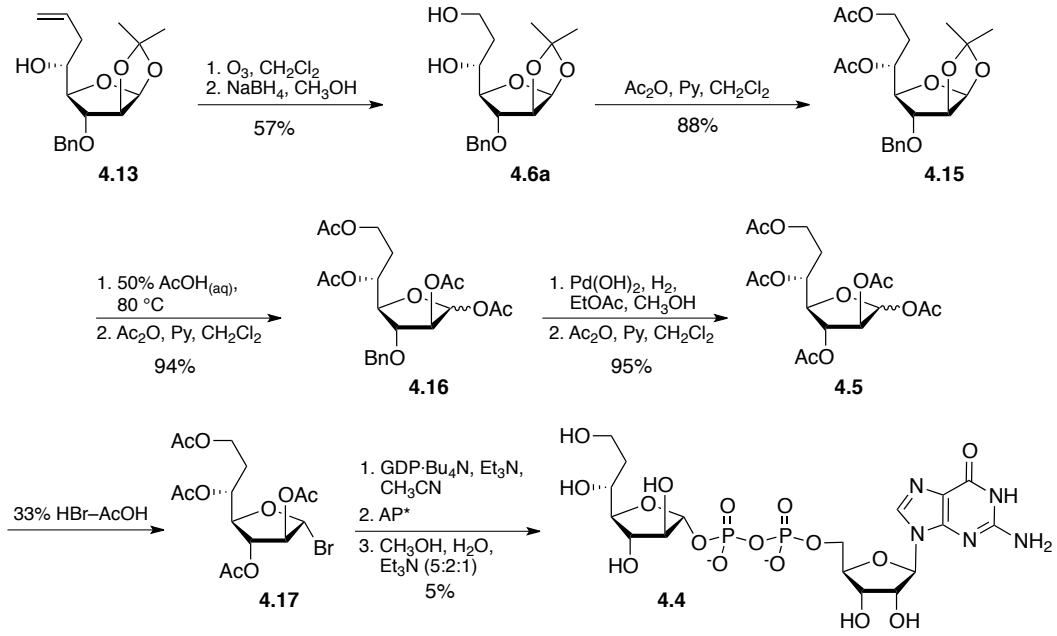


Scheme 4-3. Synthesis and zinc-mediated allylation of aldehydes **4.7a** and **4.7b.**

Thus, aldehyde **4.7b** was prepared from D-arabinose by reacting first with tertbutyldiphenylsilyl (TBDPS) chloride under DMAP catalysis to protect selecti[17]vely the primary 5-hydroxyl group, followed by trapping the furanose ring form as the 1,2 isopropylidene acetal **4.11** (Scheme 4-3) upon treatment

with dimethoxypropane with *p*-TSA in acetone.[20] Without purification, the free hydroxyl group of **4.11** was protected as a benzyl ether, and the silyl ether was removed by tetrabutylammonium fluoride treatment to give alcohol **4.12** in 47% yield over four steps. This alcohol was then converted to the aldehyde **4.7b**, again via a Swern oxidation. The key zinc mediated allylation proceeded efficiently to give an 81% yield of the desired alcohol **4.12** in a 95:5 ratio of diastereomers favoring the anti product (Scheme 4-3), which was confirmed by comparing the NMR data to that previously published.[18] To further examine the diastereoselectivity of the zinc-mediated allylation, the reaction was also tested on aldehyde **4.7a**, but in this case we observed poor selectivity, with the desired (*R*)-alcohol as the minor product.

Having developed a route to a compound with the correct stereochemistry at C-5, ozonolysis of **4.13** followed by *in-situ* reduction with sodium borohydride yielded the heptofuranose diol **4.6a**. To provide further evidence of the product stereochemistry, we hydrolyzed the isopropylidene acetal under acidic conditions and protected the resulting reducing sugar by treatment with acetic anhydride in pyridine, to give an ~1:1 mixture of the pyranose and furanose ring forms. The coupling constants for the pyranose ring isomers matched with those expected for the D-altropyranose configuration. As illustrated in Scheme 4-4, with **4.6a**, a series of protecting group manipulations led to the desired peracetate **4.5** in 78% yield over five steps.



Scheme 4-4. Synthesis of GDP-6d-D-*altro*-Hep*f*. * AP = Alkaline phosphatase

GDP-6d-D-*altro*-Hep*f* was then obtained via the direct displacement of the acylated glycosyl bromide **4.17** (Scheme 4-4), prepared by treatment of the peracetate **4.5** with 33% hydrobromic acid in acetic acid, with an organic soluble guanosine-5′-diphosphate tetrabutylammonium salt (GDP·2 Bu_4N) under the conditions developed by Timmons and Jakeman.[21] After treating the partially purified product with alkaline phosphatase to degrade any unreacted GDP, and removal of the acyl protecting groups under mild basic conditions, the desired GDP-6d-D-*altro*Hep*f*, **4.4**, was isolated by semi-preparative HPLC in 5% yield. The product was obtained in poor yield via this approach. Therefore, other routes using pyrophosphate coupling were considered.[22] In our hands, however, the 1-phosphate intermediate required for this approach could not be isolated in sufficient quantities. For example, glycosylation with dibenzylphosphate, as

described in the previous chapter for the synthesis of the modified UDP-D-Gal*f* derivatives, gave only a 40% yield of a 1:1 mixture of diastereomers, which were inseparable, and which partially hydrolyzed during column chromatography. Therefore, we were unable to obtain the pure α-1-phosphate required to attempt the pyrophosphate coupling. Using the direct displacement of the acylate glycosyl bromide method we were able to obtain sufficient amount of **4.4** to complete our studies of Glf1.

4.2.1.2 Sequence analysis suggests Glf1 is a GaHM

Of the three putative pyranose–furanose mutase enzymes encoded in the *C. jejuni* HS:41 CPS gene locus, the *glf1* gene product has the lowest homology to other bacterial UGM from *E. coli*, *Klebsiella pneumoniae*, *Mycobacterium tuberculosis* and *Deinococcus radiodurans*, or the UNGM from *C. jejuni*, all of which have had their function experimentally determined. The gene shares only 14–19% sequence identity to these mutase enzymes (~19% to EcUGM, 14% to MtUGM, 17% to CjUNGM, 15% to KpUGM, and 18% to DrUGM). Because of the low homology between the previously characterized mutases and Glf1, we hypothesized that this enzyme would be involved in the biosynthesis of GDP-6d-D-*altro*-Hep*f* **4.4**, which has the least structural similarity to UDP-D-Gal*f*, the product of UGM.

The low sequence identity did not allow us to generate an accurate homology model of Glf1; however, an analysis of the primary sequence shows that many of the UGM active site and FAD binding site amino acids are also

conserved in Glf1 (Figure 4-5). This suggests that, like the UGMs, Glf1 is a flavoprotein, and when tested, the UV-visible spectrum of recombinant Glf1 does show characteristic absorbances at 373 and 450 nm (Appendix C).

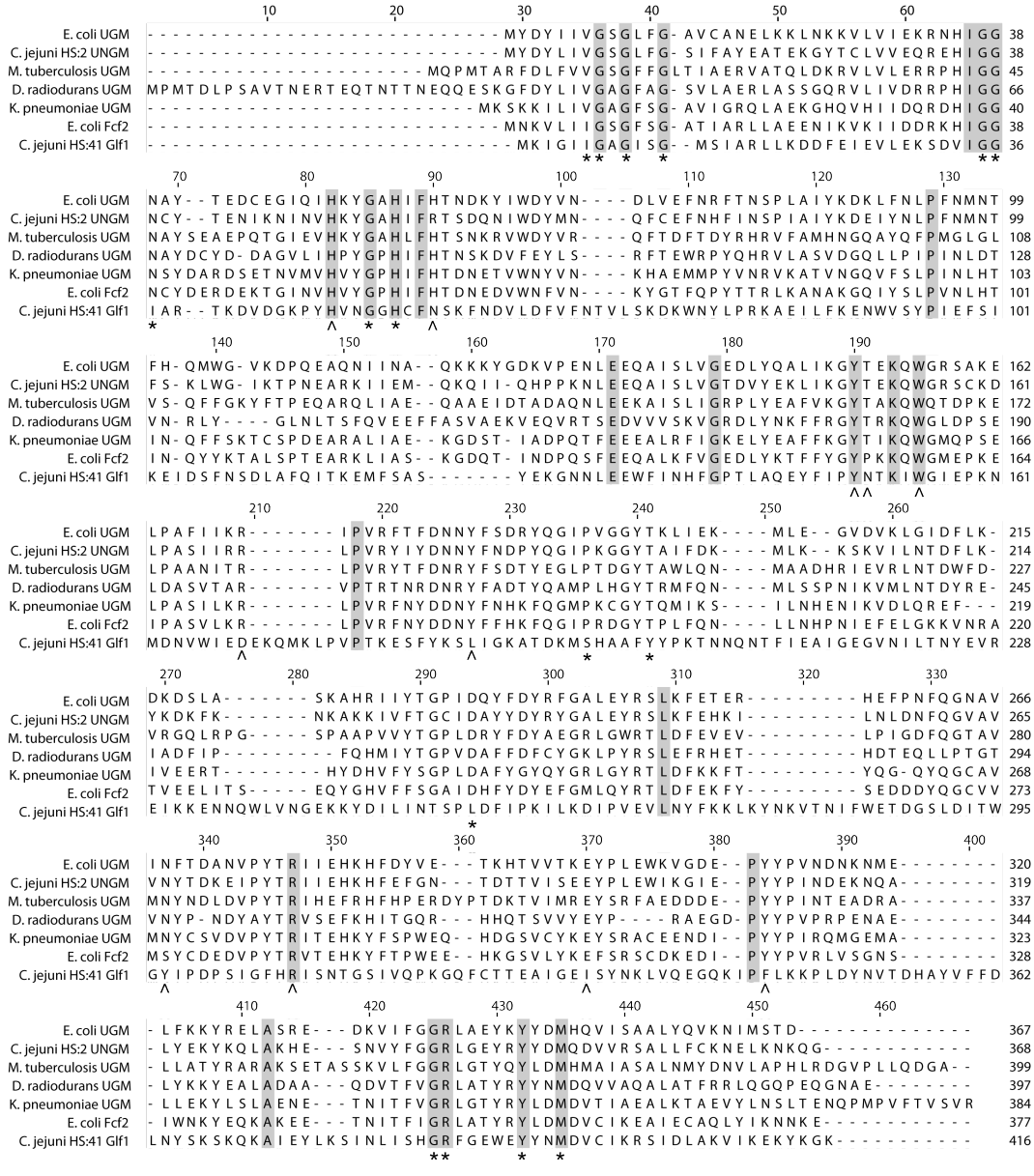


Figure 4-5. Multiple sequence alignment of Glf1 with established bacterial UGMs. Highly conserved amino acids are shaded. The conserved amino acids found in the active site are marked with an arrow. Conserved amino acids involved in FAD binding are indicated with an asterix.

*4.2.1.3 Glf1 has GDP-6d-D-***altro***-heptopyranose mutase (GaHM) activity*

The ability of Glf1 to function as a GaHM was evaluated by incubating the protein with GDP-6d-D-*altro*-Hep*f* **4.4** under reducing conditions as described for other pyranose–furanose mutases.[23, 24] When the reactions were monitored by HPLC, as described in previous chapters, hints of a new product peak were observed. However, under all conditions explored, the product peak largely overlapped with the peak for GDP-6d-D-*altro*-Hep*f* **4.4**, indicating that HPLC could not be used to analyze this reaction.

Previously, capillary electrophoresis (CE), using sodium tetraborate buffer,[25] has been used successfully to separate sugar nucleotide mixtures that are inseparable by HPLC; e.g., the separation of UDP-Glc*p*NAc and UDP-D-Gal*p*NAc from enzymatic reactions of the *C. jejuni* galactopyranose epimerase (GalE).[26] Using this CE method to analyze the Glf1 reaction with GDP-6d-*altro*-Hep*f* **4.4**, a longer retention time product (~2.2 min) was observed, which increases with the incubation time (Figure 4-6). These results suggest that, at equilibrium, a ~1:1 ratio of the product with respect to the GDP-6d-*altro*-Hep*f* **4.4** peak resulted, as seen in Figure 4-6*ii*. This differs greatly from the equilibrium ratio of ~93:7 observed for the *K. pneumoniae*[27], *E. coli*,[23] and *D. radiodurans*[28] UGM or the *C. jejuni* UNGM[3], and from the 3:1 pyranose–furanose equilibrium ratio observed for the 6d-D-*altro*-Hep reducing sugar in water.[29] The observed ratio is likely do to the lower thermodynamic stability of GDP-6d-*altro*-Hep*p*, which contains three axial hydroxyl groups.

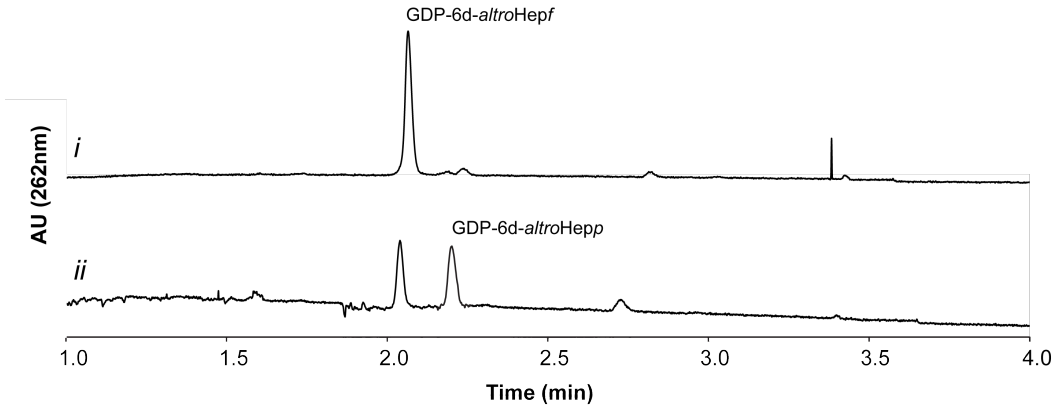


Figure 4-6. Functional characterization of the Glf1 enzyme with GDP-6d-D-*altro*-Hep*f* **4.4** as the substrate. *i.* The retention time for a standard of GDP-6d-*altro*Hep*f* **4.4**. *ii.* Reaction mixture appears to have reached equilibrium after 30 min.

We were unable to isolate and characterize the Glf1 reaction product by CE or HPLC. To confirm that the product peak observed by CE was, in fact, due to the expected GDP-6d-D-*altro*-Hep*p* (**4.3**) reaction product and not simply hydrolyzed or otherwise degraded substrate, the reaction progress was directly monitored by ^{1}H NMR spectroscopy. Like previously described bacterial UGMs, only Glf1 with a reduced FADH^{-} cofactor was found to be active, and in the absence of a reducing agent no new resonances were observed, even after 30 min. With the addition of sodium dithionite, the resonances corresponding to a single new product build in over time (Figure 4-7), consistent with the CE results. These new product resonances matched the expected chemical shifts and coupling constants for **4.3** (Appendix C). Combined, the results by CE and NMR strongly support that Glf1 functions as GaHM in vitro catalyzing the isomerization between **4.4** and **4.3** (Figure 4-3). Although we have not ruled out

the involvement of another sugar nucleotide (ADP-6d-D-*altro*-Hep, for example), the available genetic information,[10] and observed activity, supports that a GDP sugar nucleotide is the biologically relevant substrate. This is the first example of a mutase enzyme that recognizes a heptose substrate.

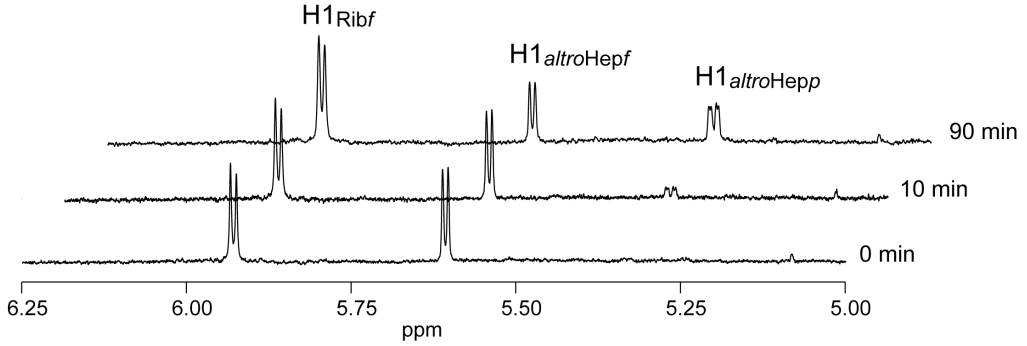


Figure 4-7. *in situ* NMR analysis of Glf1 reaction with GDP-6d-D-*altro*-Hep*f* (**4.4**) over time. After 10 min, the H1 resonance of the GDP-6d-D-*altro*-Hep*f* (δ 5.61 ppm) has decreased and the H1 resonance for GDP-6d-D-*altro*-Hep*p* appears at δ 5.33 ppm. After 90 min, the reaction has reached equilibrium with the ratio of $H1_{altro\text{-}Hepf}$ to $H1_{altro\text{-}Hepp}$ ~1:1. The ribose H1 integration remains constant over time as it is found in both the reactant and product.

4.2.2 Characterizing the activity and specificity of a bacterial UDP-L-arabinopyranose mutase (UAM)

To identify the function of the remaining putative pyranose–furanose mutase enzymes (Glf2 and Glf3) of *C. jejuni* serotype HS:41, their primary amino acid sequences were further examined. A BLAST search revealed a gene encoding for a protein with high similarlity to the Glf3 enzyme in the CPS gene locus of *C. jejuni* serotype HS:15 (GenBank accession number ADZ76280; 96%

identity; 98% similarity),[30] a strain for which the CPS structure has also been identified.[31] The CPS of the HS:15 serotype has a disaccharide repeating unit (Figure 4-8) composed of Ara*f* (absolute stereochemistry not determined, believed to be L) and 6-deoxy-D-*gulo*-heptopyranose (6d-D-*gulo*-Hep*p*). Only L-Ara*f* is common to both the HS:15 and HS:41 serotypes. A closer inspection of the CPS gene loci reveals that both serotypes also share the *galE1*, *udg*, and *galE2* genes,[10, 30] which are implicated in the biosynthesis of UDP-L-arabinose. Based on this analysis, we hypothesized that the Glf3 enzyme would function as the UDP-L-arabinopyranose mutase (UAM) involved in the biosynthesis of L-Ara*f* in the HS:41 CPS.

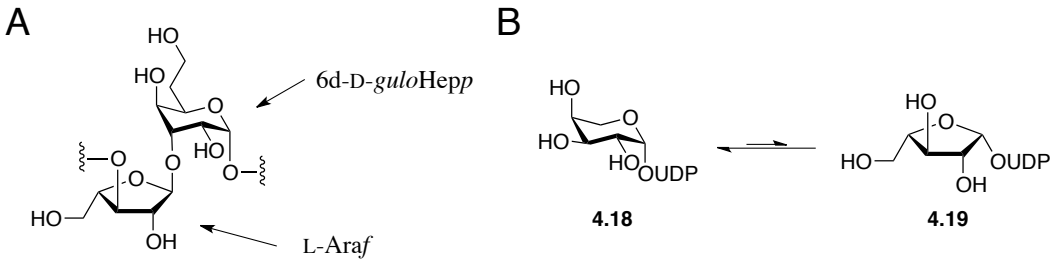


Figure 4-8. The *C. jejuni* serotype HS:15 CPS repeating unit is composed of Ara*f* and 6d-D-gulo-Hep*p* (A). The absolute configuration of the Ara*f* was not determined.[31] The proposed biosynthesis of UDP-Ara*f* catalyzed by Glf3 is also shown (B).

4.2.2.1 Glf3 has UDP-L-arabinopyranose mutase (UAM) activity in vitro

Once purified, recombinant Glf3 also showed characteristic flavin absorbances at 373 and 450 nm (Appendix C), consistent with other pyranose–furanose mutase enzymes. Moreover, the sequence similarity of Glf3 to the putative mutase of *C. jejuni* HS:15 suggested it functions as a UAM. To test this

pridiction, we incubated the Glf3 enzyme with synthetically prepared UDP-L-Ara*f* (**4.19**, described in Chapter 3) under reducing conditions,[23] and the reaction progress was monitored by reversed phase HPLC. A shorter retention time product (7.9 min) was observed (Figure 4-9A *i.*) and the ratio with respect to the peak for **4.19** was found to be ~9:1 at equilibrium, which is consistent with the product distribution observed previously for UGMs.[27, 32] NMR analysis of this reaction, as performed for Glf1, showed that the expected UDP-Ara*p* (**4.18**) is produced in this reaction (Figure 4-10). The K_M and k_{cat} values for Glf3 with **4.19** were 124 μM and 45 s^{-1}, respectively, similar to those reported for the UGM enzymes.[23, 27]

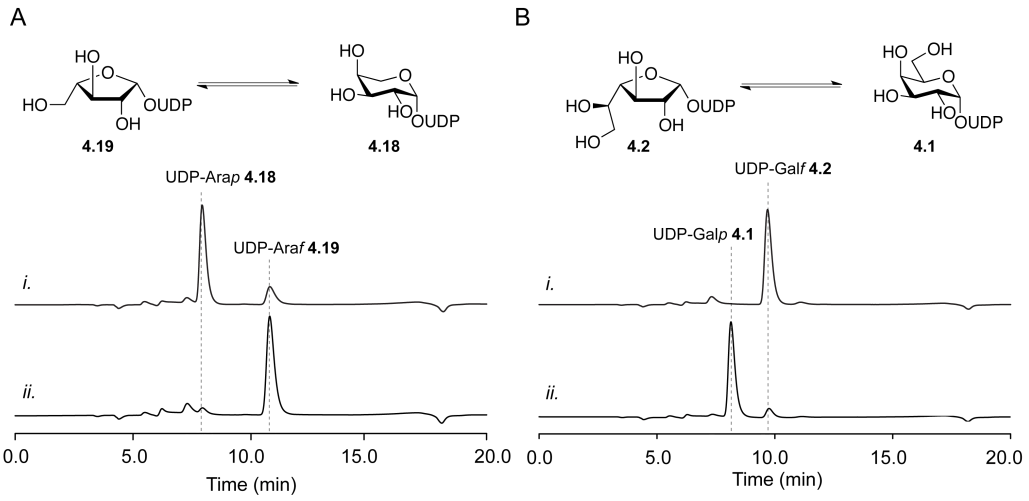


Figure 4-9. Functional characterization of *C. jejuni* Glf3 with UDP-L-Ara*f* (A) and UDP-D-Gal*f* (B). The retention time for UDP-L-Ara*f* **4.19** and UDP-D-Gal*f* **4.2** were found to be 10.8 min and 9.6 min, respectively. Reactions of Glf3 with **4.19** (A*i*.) showed the new **4.18** product peak form at 7.9 min, but reactions with **4.2** (B*i*.) showed no product formation under the same conditions. For comparison, the reactions of EcUGM with **4.19** (A*ii*.) and **4.2** (B*ii*.) are also shown.

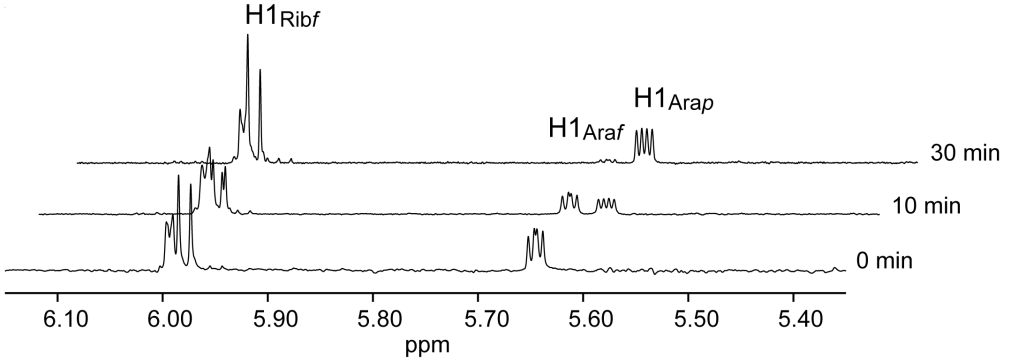


Figure 4-10. *in situ* NMR analysis of the Glf3 reaction with UDP-L-Ara*f* **4.19**. After 10 min the H1 resonance of the Ara*f* (δ 5.65 ppm) has decreased by half and the H1 resonance for UDP-Ara*p* appears at δ 5.61 ppm as previously reported.[33] After 30 min the reaction has reached equilibrium with the ratio of $H1_{Araf}$ to $H1_{Arap}$ ~1:10. The ribose H1 chemical shift (δ ~5.98 ppm) also changes slightly as the product forms.

4.2.2.2 UDP-D-Galf is not a Glf3 substrate

The structure of UDP-D-Gal*f* (**4.2**) and UDP-L-Ara*f* (**4.19**) differ only in the presence of an additional hydroxymethyl group in the Gal*f* moeity, and many UGM have been shown to use **4.19** as a substrate,[27, 32] although at reduced levels compared to **4.2** (Figure 4-9 A*ii.* and B*ii.*). This had lead to the hypothesis that a single enzyme could be responsible for the biosynthesis of both L-Ara*f* and D-Gal*f*;[32] however, to the best of our knowledge, no bacteria producing both L-Ara*f* and D-Gal*f* containing glycoconjugates has been identified. Moreover, we were interested to explore whether the *C. jejuni* Glf3 was simply a UGM enzyme with relaxed substrate specificity. When Glf3 was incubated with **4.2** using the same conditions as with above, no new products were observed (Figure 4-9 A*ii.*). Even increasing the reaction time gave only trace amounts of the potential UDP-D-

Gal*p* (**4.1**) product. From these results, it appears Glf3 does not function as a UGM and is instead specific for arabinose; thus, it should be renamed as a UAM enzyme. This represents the first example of a UAM indentified in bacteria, which differs in structure from the previously described plant UAM described in Chapter 1.[34, 35]

4.2.2.3 Modeling the Glf3 active-site suggests origin of UDP-L-Araf specificity

To further investigate the origin of the *C. jejuni* Glf3 specificity for UDP-L-Ara, a homology model of the Glf3 active site (Figure 4-11) was generated based on the crystal structure of the *K. pneumoniae* UGM bound to UDP-D-Gal*p* (**4.1**)[36] (PDB id: 3INT chain B). The active-site residues are highly conserved between Glf3 and the four bacterial UGM for which the crystal structures have been determined. However, the tyrosine in position 84 (Y84) in Glf3 differs from the corresponding residues in the other four bacterial UGM (N80 in ecUGM, N84 in kpUGM, H89 in mtUGM and H109 in drUGM). In each case, the corresponding asparagine or histidine residue is involved in hydrogen bonding to the 6-hydroxyl group of the galactopyranose moiety, an interaction that would be absent with the Y84 of Glf3. Moreover, in the model, Y84 is situated so as it would clash with the hydroxymethylene of galactopyranose (Figure 4-11, UDP-D-Gal*p* is shown in yellow). This presence of a tyrosine at this position in Glf3 appears to influence the substrate specificity of the enzyme and leads to it favoring the UDP-arabinose substrate.

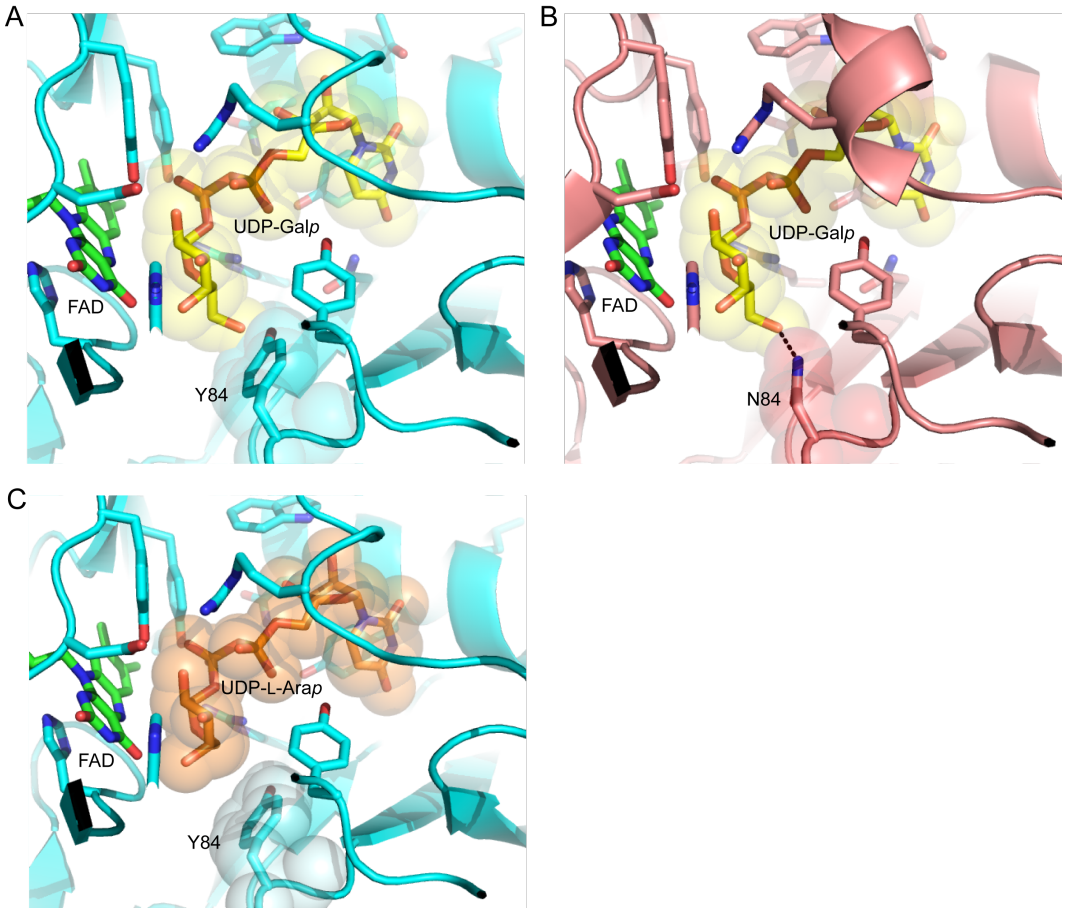


Figure 4-11. Homology model of *C. jejuni* Glf3 (A) compared to the crystal structure of *K. pneumoniae* UGM (B; PDB id: 3INT chain B). The UDP-D-Gal*p* found in the crystal structure is shown. N84 of UGM (B) makes a hydrogen bond to the Gal*p* 6-OH. The corresponding Y84 in Glf3 (A) appears to occupy the same space as the hydroxymethylene of UDP-D-Gal*p*. There is no clash with Y84, in Glf3, when UDP-L-Ara*p* is modeled in the active site (C).

4.2.2. Site directed mutagenesis results support the role of Y84 in Glf3 substrate specificity.

To test the role of Y84 in the specificity of Glf3, this amino acid was mutated to an asparagine, as found in the *K. pneumoniae*, and *E. coli* UGMs. The isolated Y84N mutant (Appendix C) was then tested for its activity with both

UDP-L-Ara*f* (**4.19**) and UDP-D-Gal*f* (**4.2**). The isolated Y84N mutant enzyme was found to be unstable in buffer, and precipitated rapidly after purification. As a result, only the relative activity between the two substrates, and not the kinetic parameters for both possible substrates, could be measured accurately as an accurate concentration of active enzyme could not be measured. The results show that the Y84N mutation resulted in a switch in the enzyme specificity to favor **4.2** as the substrate over **4.19** (Figure 4-12), in a ratio of ~3:1. This is similar to the ratio of ~5:1 seen for the ecUGM and kpUGM, which are discussed in Chapter 3. These measurements support the role of Y84 in influencing the specificity of Glf3, but kinetic measurements would be desirable to confirm these results.

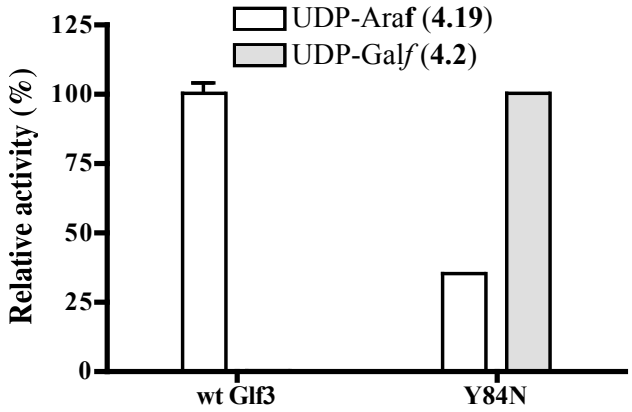


Figure 4-12. Relative activity of Glf3 and the Y84N mutant with **4.19** and **4.2**. The activity of Y84N with UDP-Gal*f* was arbitrarily set to 100% to compare the specificity to that of wt Glf3.

4.2.3 Synthesis and analysis of putative Glf2 substrates

Having established the function of Glf1 and Glf3 in the biosynthesis of 6d-D-*altro*-Hep*f* and L-Ara*f*, respectively, we hypothesized that the remaining

putative pyranose–furanose mutase, Glf2, is involved in the biosynthesis of either D-Fuc*f* or 6d-L-Alt*f*. It is also possible that Glf2 is a bifunctional enzyme involved in the biosythesis of both species, which differ only in the C-5 stereochemistry. To test these possibilities, we set out to prepare the possible furanose sugar nucleotide substrates.

Before testing the activity of Glf2, it was first necessary to select the putative donors to target for synthesis. As outlined below, we viewed the CDP or TDP derivatives as the most likely substrate for 6d-L-Alt*f*. The exact biosynthetic precursor of 6d-L-Alt*f* is not known in *C. jejuni* HS:41. A putative CDP-6d-altropyranose mutase enzyme has been implicated in the biosynthesis of CDP-6d-L-Alt*f* in *Y. pseudotuberculosis* serotype O:11,[37] based purely on sequence homology to a putative CDP-paratopyranose mutase (WbyH) of *Y. pseudotuberculosis* serotype O:1b. However, the activity of neither O:11 or O:1b enzyme has been experimentally determined. When the sequence of this putative CDP-6d-altropyranose mutase was compared to Glf2, it showed only poor sequence identity (~10%). The pyranose sugar 6d-L-Alt*p* has also been identified in the LPS of *Y. enterocolitica* O:3,[38] where it has been documented to use TDP-6d-L-Alt*p* (**4.20**, Figure 4-12) as the biosynthetic precusor. Therefore, it is possible that TDP-6d-L-Alt*f* (**4.21**) could also be a biosynthetic precursor in *C. jejuni* HS:41.

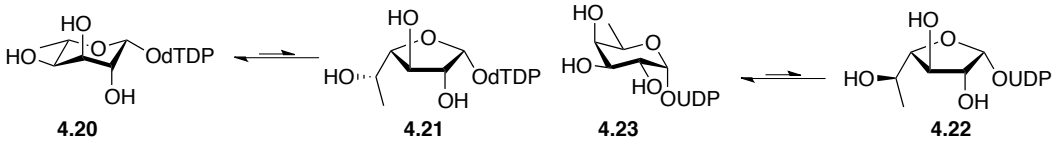
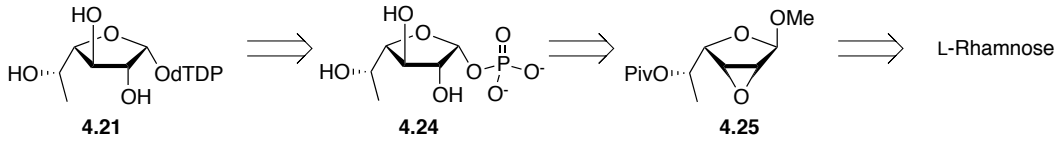


Figure 4-13. Two reactions hypothesized to be catalyzed by Glf2.

D-Fuc*f* sugar-nucleotides could also be potential substrates for Glf2. The only characterized TDP-D-fucopyranose mutase enzyme (Fcf2) has been implicated in the biosynthesis of D-Fuc*f* in *E. coli* O52.[4] This enzyme has only modest homology (35% identity) to the Glf2 enzyme. Having already synthesized an alternative potential substrate UDP-D-Fuc*f* (**4.22**, Figure 4-11) in Chapter 3, here it was evaluated as a possible Glf2 substrate. In addition, TDP-6d-L-Alt*f* (**4.21**) was synthesized and evaluated as a Glf2 substrate.

4.2.3.2 Synthesis of TDP-6d-L-Altf

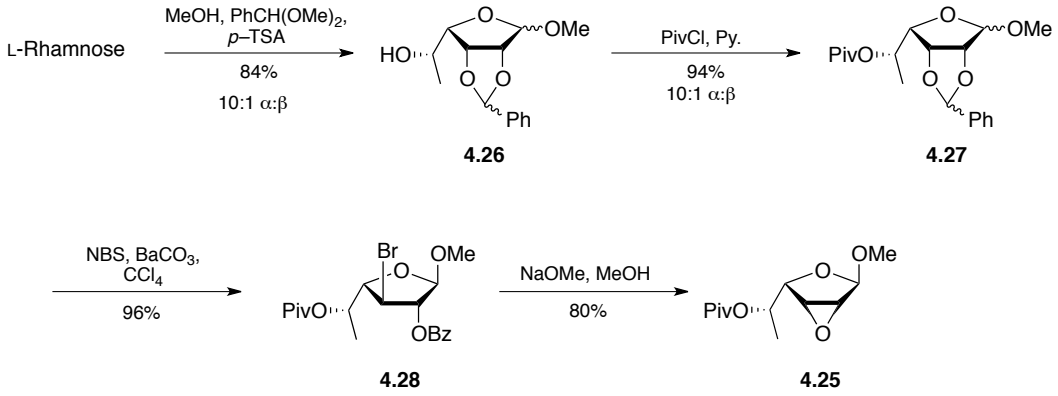
To synthesize TDP-6d-L-Alt*f* (**4.21**) we envisioned using a chemo-enzymatic approach (Scheme 4-5) employing the Cps2L enzyme that has been previously shown to tolerate furanose-1-phosphate sugars as substrates.[39] We previously used this method in Chapter 3 for the synthesis of TDP-D-Gal*f*. To use this approach, we required the 6d-L-Alt*f*-1-phosphate **4.24**.



Scheme 4-5. Retrosynthesis of TDP-6d-L-Alt*f*.

6d-L-Altrose is the 3-epimer of the more common L-rhamnose, so to synthesize **4.24** we started with commercially available L-rhamnose. Using the method developed by Florent and Monneret,[40] L-rhamnose was converted, in 84% yield, directly to the 2,3-*O*-benzylidene protected methyl glycoside **4.26**

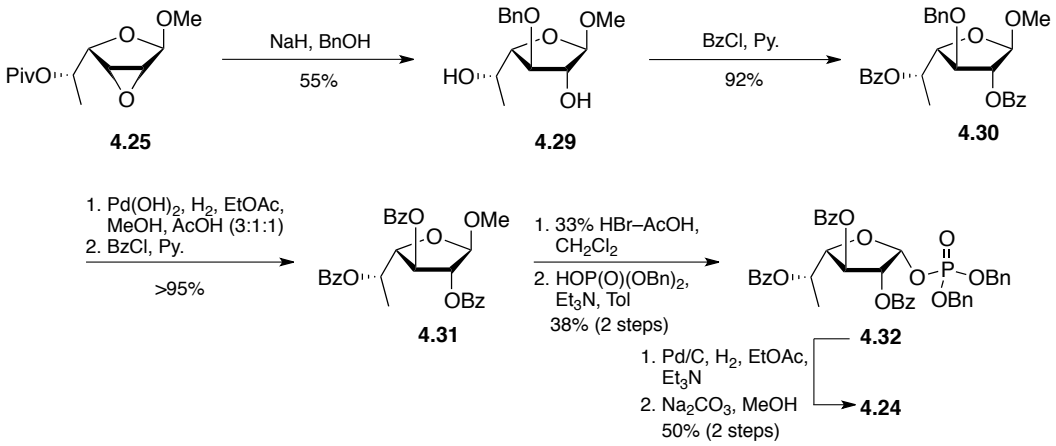
upon treatment with benzaldehyde dimethyl acetal in acidic methanol. Protection of the free hydroxyl group with pivaloyl chloride gave a 94% yield of **4.27** (Scheme 4-6). Oxidative opening of the benzylidene acetal with NBS[40] gave the 3-bromo substituted **4.28** in excellent yield. Treating **4.28** with sodium methoxide resulted in displacement of the bromide to generate the 2,3-epoxide **4.25** in 80% yield.



Scheme 4-6. Synthesis of epoxide **4.25**.

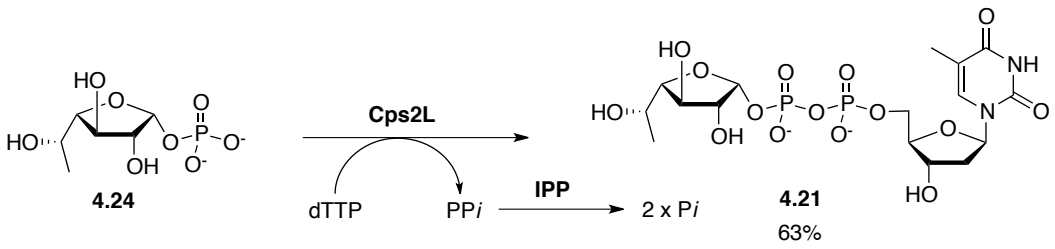
Opening the epoxide **4.25** with benzyl alcohol under basic conditions gave **4.29** in 55% yield (Scheme 4-7). To confirm the regioselectivity of the epoxide opening, the free hydroxyl groups of **4.29** were acylated with benzoyl chloride to give **4.30**. In the ^{1}H NMR spectrum of the product of **4.30**, the resonances for H-2 and H-5 were downfield compared to the spectrum of **4.29**. In particular, the H-2 resonance moved from δ 4.18 ppm (**4.29**) to 5.40 ppm (**4.30**) and H-5 from δ 4.09 ppm (**4.29**) to 5.50 ppm (**4.30**). This showed that epoxide opening occurred exclusively at C-3. Removal of the benzyl ether and

protection of the resulting hydroxyl group with benzoyl chloride gave the protected methyl glycoside **4.31** in excellent yield. The methyl glycoside was treated with hydrobromic acid followed by displacement of the resulting glycosyl bromide with dibenzyl phosphate to give the β-glycosyl-phosphate **4.32** as the major product in 38% yield over two steps. The desired 6d-L-Alt*f*-1-phosphate **4.24** was obtained in 50% yield after deprotection.



Scheme 4-7. Synthesis of 6d-L-Alt*f*-1-phosphate from epoxide **4.25**.

The 1-phosphate **4.24** was then converted to the TDP-sugar **4.21** using the Cps2L enzyme (Scheme 4-8). This enzyme normally catalyzes the transfer of a unit of dTMP from dTTP to a glucopyranose-1 phosphate substrate, but has been shown to have a broad substrate tolerance including many furanose-1 phosphate sugars.[39] Here, Cps2L effectively catalyzed the transfer of dTMP onto **4.24** in a moderate 63% yield. The isolated **4.21**, along with the previously synthesized UDP-D-Fuc*f* **4.22**, were then tested as substrates for Glf2.



Scheme 4-8. Chemo-enzymatic synthesis of **4.21** by Cps2L.

4.2.3.2 Preliminary studies to elucidate the activity Glf2

For the preliminary analysis of Glf2 we had to resort to the use of partially purified enzyme (~25% pure) due to the poor solubility and poor Ni-NTA binding of the recombinant 6 × His–tagged Glf2. Under all of the expression conditions tested, the recombinant enzyme was found predominantly in the insoluble fraction. The expression was also tested in different *E. coli* expression strains, including Rosetta(DE3)pLysS and DH5α strains, with no improvement in the soluble protein. The partially purified Glf2 enzyme was tested for activity with **4.21** and **4.22** by incubating the enzyme with the sugar nucleotides under reducing conditions. Only small amounts of product could be observed for the reactions with **4.22** (Figure 4-14A) regardless of the reaction time used, and no new product peaks were observed in the reactions with **4.21** (Figure 4-14B). These results suggest that neither UDP-D-Fuc*f* **4.22** nor TDP-6d-L-Alt*f* **4.21** is the native substrate for Glf2, but these results will need to be repeated when the pure Glf2 protein can be isolated in reasonable quantities. It is also possible that the Glf2 enzyme recognizes one, or more, of these substrates as a different nucleotide (ie. CDP-D-Fuc*f,* TDP-D-Fuc*f* or CDP-6d-L-Alt*f*), a

possibility that will be discussed in more detail below. Glf2 activity will also need to be evaluated with these alternative substrates.

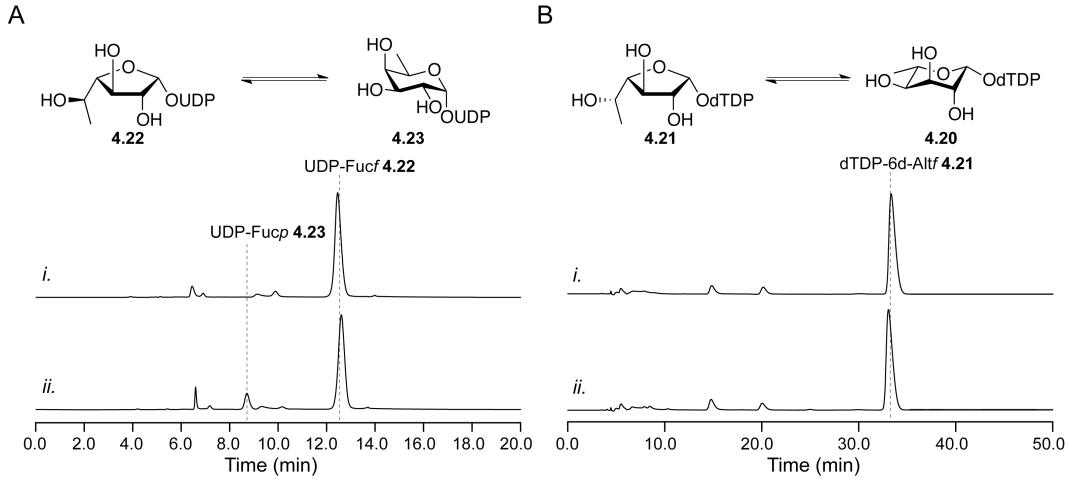


Figure 4-14. Preliminary analysis of reactions of *C. jejuni* Glf2 with **4.22** (A) and **4.21** (B). The retention time for the **4.22** (A*i*.) and **4.21** (B*i*.) standards were found to be 12.8 min and 33.4 min, respectively under the HPLC conditions used. Reactions of Glf2 with **4.22** (A*ii*.) showed small amounts of the new UDP-D-Fuc*p* (**4.23**) product peak form at 8.4 min, but reactions with **4.21**(B*ii*.) showed no product formation.

4.2.3.3 Homology model of Glf2

We prepared homology model of the *C. jejuni* Glf2 active site (Figure 4-15) based on the crystal structure of the *K. pneumoniae* UGM (PDB id: 3INT chain B) bound to UDP-D-Gal*p* (**4.1**)[36] to investigate any differences in the putative substrate binding interactions that would help in assigning biological activity of Glf2. As with Glf3, the active-site residues of Glf2 are highly conserved with the four bacterial UGM the crystal structures of which have been determined. However, the asparagine (N84) in the *K. pneumoniae* UGM involved in hydrogen bonding to the 6′′-OH of UDP-D-Gal*p* is replaced with a

hydrophobic leucine (L80) in the Glf2 model. A hydrophobic amino acid in this position is in agreement with Glf2 recognising a 6′′-deoxy substrate (ie. D-Fuc or 6d-L-Alt). There are also a number of differences in the amino acids in the nucleobase binding site of Glf2 when compared to the kpUGM crystal structure. Most notable is aspartate 150 (E150) that could be involved in binding a CDP-sugar substrate. This aspartate side chain is situated so as to be able to bind the 4-NH_2 of CDP (not shown). However, these putative interactions remain to be evaluated once we are able to produce Glf2 in sufficient quantities for further analysis.

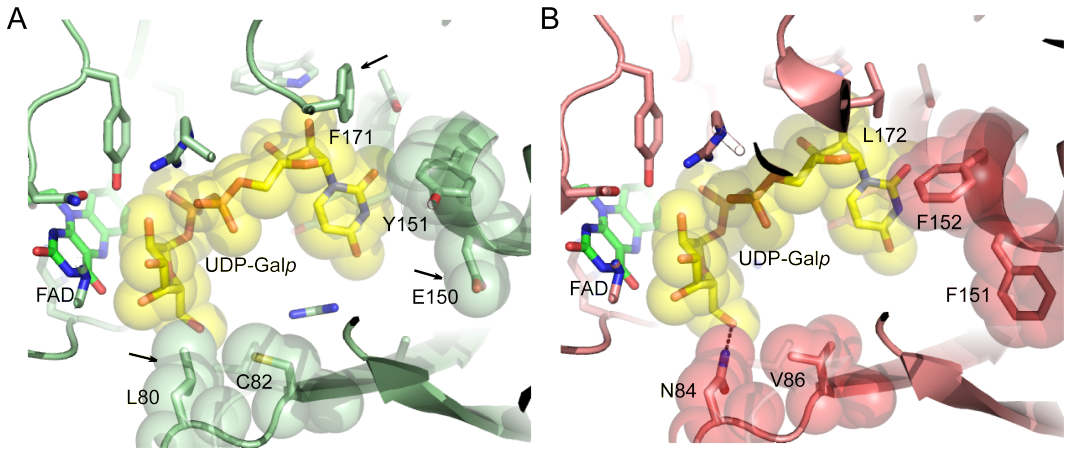


Figure 4-15. Homology model of *C. jejuni* Glf2 (A) compared to the crystal structure of *K. pneumoniae* UGM (B; PDB id: 3INT chain B). The model is shown with the UDP-D-Gal*p* found in the crystal structure shown. N84 of UGM (B) makes a hydrogen bond to the Gal*p* 6-OH. This residue is replaced by a hydrophobic L80 in Glf2 (A). A number of amino acid differences are also seen in the nucleobase binding site of Glf2 (ie. E150 vs. F151 in *K. pneumoniae* UGM). Notable amino acid differences in the Glf2 active site are indicated with arrows.

4.3 Conclusions

In this chapter, we report the synthesis of putative sugar nucleotide precursors (**4.4** and **4.21**) to the 6d-D-*altro*-Hep*f* and 6d-L-Alt*f* that comprise the CPS of *C. jejuni* serotype HS:41. Using these synthetic sugar nucleotides, as well as putative precursors to L-Ara*f* and D-Fuc*f* (**4.19** and **4.22**), we were able to demonstrate the function of two of the *C. jejuni* HS:41 Glf enzymes involved in CPS biosynthesis.

Glf1, which has the lowest sequence homology to UGM previously described, was found to exhibit GaHM activity *in vitro*, and is the first pyranose–furanose mutase described to use a heptose substrate. Like UGM, the purified Glf1 is a flavoenzyme, and likely uses a similar mechanism to UGM to catalyze the isomerization between **4.3** and **4.4**.

Using synthetic UDP-L-Ara*f* (**4.19**), we found Glf3 displays UAM activity, and was specific for this pentose sugar over the structurally related UDP-D-Gal*f* (**4.2**). A tyrosine (Y84) in the Glf3 active site appears to influence this substrate specificity. Mutation of this tyrosine to an asparagine, which is present in ecUGM and kpUGM, resulted in a switch in the specificity to favor **4.2**. The only other reported studies on the biosynthesis of L-Ara*f* address the formation of UDP-L-Ara*f* in plants.[34, 41] Interestingly, in plants this process uses a flavin-independent mutase enzyme to catalyse the isomerization of **4.18** and **4.19**. *C. jejuni* instead appears to have developed a convergent mechanism to access L-Ara*f*, using a flavin dependent pyranose–furanose mutase to catalyze

the same reaction. Thus, Glf3 represents the first flavin-dependent UAM for which the function has been unequivocally determined.

The activity of Glf2 was also examined using synthetic **4.21**, and **4.22**, but neither was found to be an effective substrate for Glf2. It is possible that this enzyme uses a CDP-derivative as the sugar nucleotide substrate. These putative CDP substrates could then be prepared from the 6d-L-Alt*f*-1P and D-Fuc*f*-1P prepared in this Chapter, and Chapter 3, through a pyrophosphate coupling reaction with CMP. Initial attempts to do this met with little success, predominantly due to difficulties purifying the CDP-sugar products. Because the function of both Glf1 and Glf3 have been established in this chapter, it is possible that Glf2 is involved in both 6d-L-Alt*f*, and D-Fuc*f* biosynthesis. However, further work is required to elucidate the function and specificity of this enzyme.

The results of this study have shown that *glf* homologs encoding flavin dependent pyranose–furanose mutase enzymes are involved in the biosynthesis of structurally diverse bacterial furanose sugars. As seen with UNGM,[3] relatively subtle amino acid differences were found to influence the substrate specificity. This is best exemplified by Glf3, in which a single amino acid residue results in a change in specificity. As a result, it would be difficult assigning a specific function to *glf* homologs based on the primary amino acid sequence, without knowledge of the furanose sugars produced by the organism.

The absence of furanose sugars in mammalian glycoconjugates,[2] has lead to an interest in targeting the furanose biosynthetic enzymes as a potentially

selective method to treat bacterial infections. As such, a better understanding of the biosynthesis of the furanose sugars in *C. jejuni* could provide potential pathogen-specific targets for the development of new therapeutics.

4.4 Experimental Details

General Methods.

All reagents were purchased from commercial sources and used without further purification. Reaction solvents were purified by successive passage through columns of alumina and copper under a nitrogen atmosphere using a PURESOLV-400 system (Innovative Technology Inc., Newburyport, MA). Reactions were performed in oven-dried glassware and, unless stated otherwise, were carried out at room temperature under a positive pressure of argon and were monitored by TLC on silica gel 60-F_{254} (0.25 mm, Silicycle). Spots were detected under UV light or by charring with acidified ethanolic anisaldehyde. Unless otherwise indicated, column chromatography was performed on silica gel 60 (40–60 μM) where the ratio of silica gel and crude product ranged from 100:1 to 20:1 (w/w). Organic solutions were concentrated under vacuum at < 40 °C (bath). Optical rotations were measured at 22 ± 2 °C on a Perkin–Elmer 241 polarimeter with a sodium D line (589 nm) and are given in units of (°·mL)/(dm·g). ^{1}H NMR spectra were recorded at 400 MHz, 500 MHz, 600 MHz or 700 MHz and chemical shifts are referenced to either TMS (0.0, $CDCl_3$) or HOD (4.78, CD_3OD; 4.67, D_2O). ^{13}C NMR spectra were recorded at 100 MHz, 125 MHz or 175 MHz, and ^{13}C chemical shifts were referenced to internal

$CDCl_3$ (77.23, $CDCl_3$), CD_3OD (48.9, CD_3OD). Electrospray mass spectra were recorded on samples suspended in CH_3Cl or CH_3OH and added NaCl. Primers were purchased from Integrated DNA Technologies (IDT), and codon optimized genes were purchase from Genscript.

Cloning, expression and purification of *C. jejuni* HS:41 Glf proteins

Harald Nothaft and Bernadette Beadle in the group of our collaborator Professor Christine Szymanski performed the cloning of the *glf1*, and *glf2* genes. The genomic DNA of *C. jejuni* serotype HS:41 served as the template for the *glf1* and *glf2* genes, which were amplified by PCR to include a 5′ *Nde*I and a 3′ *Xho*I restriction site. The PCR amplified DNA was cleaved with *Nde*I and *Xho*I and ligated into a similarly digested pET-24b plasmid. The resulting plasmids (pET-24b:*glf1* and pET-24b:*glf2*) were used to transform *E. coli* BL21. The codon optimized *glf3* gene (for expression in *E. coli*) was purchased (GenScript) as the pET-30b(+):*glf3* plasmid construct and was used to transform *E. coli* BL21 (DE3).

The optimal contditions for expression of soluble C-terminal hexa-histidine-tagged Glf1, Glf2 and Glf3 protein from *E. coli* BL21 cells carrying the pET-24b:*glf1*, pET-24b:*glf2*, or pET-30b(+):*glf3* plasmid was observed after induction with 0.25 μM isopropyl 1-thio β-D-galactopyranoside (IPTG), at an OD_{600} of ~0.6, for 6 h at 28 °C. For Glf1 and Glf2, 2× yeast extract and tryptone (2×YT) broth was used for protein expression, and Luria Bertani (LB) broth was used for the expression of Glf3. The cells were then centrifuged to pellet, re-

suspended in 100 mM potassium phosphate buffer (pH 7.4) with 150 mM NaCl, and 10 mM imidazole. The cells were lysed at 20 kpsi using a bench top cell disruptor (Constant Systems Inc.) and the lysate was clarified by centrifugation. The protein was subsequently purified by Ni-NTA affinity chromatography as per the manufacturer's directions, eluting with 250 mM imidazole in the potassium phosphate buffer described above. The proteins were dialyzed in 4 L of 100 mM potassium phosphate buffer (pH 7.4) with 150 mM NaCl for 16 h prior to use. This procedure typically yielded ~2 mg of >95% pure soluble Glf1, <2 mg of ~25% pure soluble Glf2, and 10–15 mg of >95% pure soluble Glf3 protein per liter of culture.

Sequence analysis of *C. jejuni* HS:41 Glf proteins

The percentage similarity and identity between the amino acid sequences of the Glf enzymes and other bacterial pyranose–furanose mutase enzymes were determined from the results of pairwise BLAST searches. Alignments were prepared using ClustalX2.[42]

UV–vis spectroscopy

UV–vis spectroscopy was used to measure the protein concentration of the recombinant enzymes purified from *E. coli* (BL21 DE3), and to assess the protein cofactor. Spectra were recorded on an Optizen Pop UV–visible spectrophotometer.

To determine the concentration of isolated protein the calculated molar extiction coefficients for the His_6–tagged Glf1 (ε_{280} = 82405 M^{-1} cm^{-1}), Glf2 (ε_{280} = 74150 M^{-1} cm^{-1}), Glf3 (ε_{280} = 68650 M^{-1} cm^{-1}) and Y84N_Glf3 (ε_{280} = 67285 M^{-1} cm^{-1}) were used, respectively. The molar extiction coefficient for FAD (ε_{450} = 11300 M^{-1} cm^{-1}) was used to approximate the bound FAD concentration.

Measuring GDP-6d-D-*altro*-heptopyranose mutase activity of Glf1 and capillary electrophoresis quantification

The activity of purified Glf1 was assessed by incubating a mixture of GDP-6d-D-*altro*-Hep*f*, (**4.4**, 500 μM) and Glf1 (580 nM) in 30 μL of 100 mM potassium phosphate buffer (pH 7.4) containing sodium chloride (150 mM) and freshly prepared sodium dithionite (20 mM) for 10, 20 or 30 min at 37 °C. In all cases, the reactions were stopped by heating at 95 °C for 5 min to denature the protein. Under these conditions, no appreciable degradation of the GDP-sugar nucleotides to GDP, GMP or guanosine was observed. The reactions were then diluted to 60 μL with sodium tetraborate (25 mM, pH 9.4) prior to analysis by capillary electrophoresis on a Beckman (Fullerton, CA) P/ACE 5000 series instrument equipped with a photodiode array detector (monitoring at λ = 262 nm).

Analysis of the reaction by capillary electrophoresis followed the method of Wakarchuck and co-workers[25] with minor modifications. A 75 μm × 15 cm bare silica capillary with the detector at 12 cm was used for the analysis, and the

capillary was conditioned between runs by washing with 0.2 M NaOH for 2 min, and 25 mM sodium tetraborate (pH 9.4) for 2 min. The sample was introduced by pressure injection for 5–10 s and separation was performed at 20 kV. Under these conditions, base line resolution for all substrates was achieved and the GDP-6d-D-*altro*-Hep*f* **4.4** starting material was found to elute at ~2.04 min, with a new product peak eluting at ~2.20 min. The amount of conversion was determined from the relative integration of the product and starting material peaks.

Characterization of the Glf1 reaction by ^{1}H NMR spectroscopy

Reactions containing **4.4** (500 μM), and Glf1 (580 nM) in 250 μL of D_2O–100 mM potassium phosphate buffer pH 7.4 (9:1) with freshly prepared sodium dithionite (20 mM) were monitored directly by ^{1}H NMR spectroscopy. Spectra were recorded on an Agilent/Varian VNMRS four-channel, dual receiver 700 MHz spectrometer equipped with an inverse detection, cryo-cooled $^{1}H\{^{15}N/^{13}C\}$ triple resonance, Z-gradient probe at 0, 5, 10, 30 and 90 min intervals.

Homology modeling of the Glf3 enzyme

A homology model of *C. jejuni* HS:41 Glf3 (residues 1–377) was generated using the "Automatic Modeling" mode of SWISS-MODEL[43] with default parameters. The crystal structure of the *K. pneumoniae* UGM bound to

UDP-D-Gal*f* (PDB ID: 3INT subunit B; 38% identity, 60% similarity to Glf3)[36] was used as the template.

Site-Directed Mutagenesis of Glf3

QuikChange XL II mutagenesis (Stratagene) was used to introduce mutations into the *glf3* gene sequence. The pET30b:*glf3* plasmid served as the template to introduce the Y84N mutation using 5'-gttttaccaaatggcattatttcatgaacc-gcgttaaagcatttat-3' and 5'-ataaatgctttaacgcggttcatgaaataatgccatttggtaaaac-3' as the mutagenesis primers (The Y84N codon site is underlined). Following PCR-amplification, DpnI was used to digest the template DNA, and the DpnI digested DNA was transformed into XL10-gold cells. The plasmid DNA was isolated from the XL10-gold *E. coli* cells and sequenced before being transformed into BL21(DE3) *E. coli* cells for protein expression, as described for wild-type Glf3. Typical enzyme yields were ~10 mg/1 L of initial culture at >95% purity. The protein was found to denature and lose activity after storage overnight at 4 °C, –20 ºC, or –80 °C in 100 mM potassium phosphate buffer pH 7.4 with 150 mM NaCl, or upon dialysis to remove the imidazole used in the protein purification. As a result, the relative activity measurements with the Y84N Glf3 were made directly after purification, with the reaction containing a final imidazole concentration of 25 mM.

Measuring UDP-L-arabinopyranose mutase activity of Glf3 by HPLC

Reactions of wild-type Glf3 or the Y84N mutant Glf3 (240 nM) with UDP-L-Ara*f* **4.19** (500 μM) in 30 μL of 100 mM potassium phosphate buffer (pH 7.4) containing sodium chloride (150 mM) and freshly prepared sodium dithionite (20 mM) were incubated at 37 °C for 10 min and then stopped by heating at 95 °C for 5 min to denature the protein. The reactions were then diluted to 50 μL with 100 mM potassium phosphate buffer (pH 7.4) prior to analysis by HPLC (Varian Prostar 210) following conditions similar to those previously reported by Zhang and Liu.[23] Reversed phase ion pairing chromatography was done using a C_{18} column (Phenomex, Luna 5u $C_{18}(2)$, 4.6 × 250 mm) with 0.8 mL/min isocratic elution using 200 mM triethylammonium acetate buffer pH 6.6 containing 1.5% acetonitrile. Under these conditions, base line resolution for all substrates was achieved with UDP-L-Ara*f* (**4.19**) and UDP-L-Ara*p* (**4.18**) eluting at ~10.8 and ~7.9 min, respectively. The amount of conversion was determined from the relative integration of the product and starting material peaks.

Glf3 activity with UDP-D-Gal*f*

The activity of wild-type Glf3 or Y84N mutant Glf3 with UDP-D-Gal*f* (**4.2**) as the substrate was also measured using the HPLC method as described above. Under these conditions **4.2** and UDP-D-Gal*p* **4.1** were found to elute at ~9.6 and ~8.0 min, respectively.

Characterization of the Glf3 reaction by ^{1}H NMR spectroscopy

As with the Glf1 enzyme, the activity of Glf3 was also assessed by ^{1}H NMR spectroscopy. Reactions containing **4.19** (500 μM), and Glf3 (240 nM) in 250 μL of D_2O/100 mM potassium phosphate buffer pH 7.4 (9:1) with freshly prepared sodium dithionite (20 mM) were monitored directly by ^{1}H NMR spectroscopy as described above. In this case, spectra were recorded at 0, 10 and 30 min intervals, and the reaction was found to be complete in <30 min.

Measuring Glf3 enzyme kinetics

Kinetic parameters for wild-type Glf3 were determined following a kinetic assay modified from the procedure reported by Zhang and Liu.[23] Reaction contained an appropriate amount of Glf3 was used to give <50% conversion with UDP-L-Ara*f* (15.6, 31.3, 62.5, 125, 250, 500, or 1000 μM) in a final volume of 30 μL of 100 mM potassium phosphate (pH 7.4) containing 20 mM of freshly prepared sodium dithionite. The individual reactions were incubated for 3 min at 37 °C and then promptly quenched by heating at 95 °C for 5 min. The incubation mixtures were monitored by HPLC as described above. The concentrations of UDP-L-Ara*p* produced were determined from the relative integration of the appropriate peaks on the HPLC trace, and these were used to determine the initial velocities. The assay was performed in duplicate and kinetic parameters, K_M and k_{cat}, were obtained by nonlinear regression analysis of the Michaelis–Menten equation using GraphPad PRISM 4 (GraphPad Software, San Diego, CA) (see Appendix C).

Homology modeling of the Glf2 enzyme

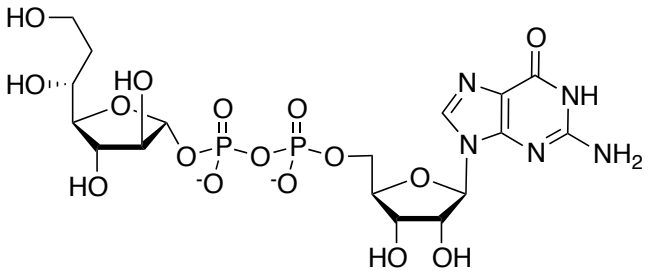
A homology model of *C. jejuni* HS:41 Glf2 (residues 1–377) was generated as described for Glf3 above. The "Automatic Modeling" mode of SWISS-MODEL[43] with default parameters and crystal structure of the *K. pneumoniae* UGM bound to UDP-D-Gal*f* (PDB ID: 3INT subunit B; 38% identity, 60% similarity to Glf3)[36] were used.

Activity of Glf2 with UDP-D-Fuc*f* and TDP-6d-L-Alt*f*

The activity of Glf2 was evaluated by incubating UDP-D-Fuc*f* or TDP-6d-L-Alt*f* (500 μM) with Glf2 (~100 nM) in 30 μL of 100 mM potassium phosphate buffer (pH 7.4) containing sodium chloride (150 mM) and freshly prepared sodium dithionite (20 mM) were incubated at 37 °C for 30 min and then stopped by heating at 95 °C for 5 min to denature the protein. The reactions were then diluted to 50 μL with 100 mM potassium phosphate buffer (pH 7.4) prior to analysis by HPLC (Varian Prostar 210) following conditions similar to those previously reported by Zhang and Liu.[23] Reversed phase ion pairing chromatography using a C_{18} column (Phenomex, Luna 5u C_{18}(2), 4.6 × 250 mm) with a 0.8 mL/min isocratic elution (of 200 mM triethylammonium acetate buffer pH 6.6 containing 1.5% acetonitrile).

Synthesis of furanose sugar-nucleotide substrates

UDP-L-Ara*f* (**4.19**), UDP-D-Gal*f*, and UDP-D-Fuc*f* were prepared using a chemo-enzymatic strategy previously described[44, 45] and reported in Chapter 3.



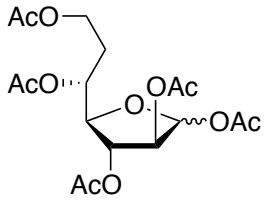
Guanosine-5′-diphosphate-6′′-deoxy-α-D-*altro*-heptofuranose (4.3):

The tetratbutylammonium salt of GDP was prepared by titrating a solution of the free acid with tetrabutylammonium hydroxide to a pH of 6, as described by Timmons and Jakeman.[21] The salt was then freeze-dried, dissolved in CH_3CN and then evaporated three times before being used. An ^1H NMR spectrum of the product revealed 2.3 equivalents of tetrabutylammonium cation per equivalent of GDP.

A solution of the GDP tetratbutylammonium salt (56 mg, 56 μmol) in CH_3CN (1.5 mL) containing Et_3N (8 μL, 60 μmol) and 4 Å molecular sieves was stirred for 30 min at room temperature, before adding a solution of glycosyl bromide **4.17** (~55 μmol) in CH_3CN (1.5 mL). The reaction mixture was then heated to 70 °C for 30 min, at which point TLC showed complete consumption of **4.17**. The molecular sieves where removed by filtration and the filtrate was concentrated. The product was dissolved in H_2O (2 mL) containing 50 EU of alkaline phosphatase and the reaction was allowed to proceed for 16 h to degrade any unreacted GDP to facilitate purification. The H_2O was then removed by evaporation under reduced pressure while ensuring the temperature stayed under 25 °C. The residue was dissolved in a solution of CH_3OH–H_2O–Et_3N 2:2:1 (2.5 mL) and stirred overnight. The solvent was then removed by evaporation and the

residue redissolved in HPLC grade H_2O (5 mL) and filtered before purification by semi-preparative HPLC.

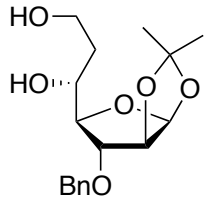
The product was purified by semi-preparative reversed phase HPLC using a C_{18} column (Microsorb, Varian, 21.4 × 250 mm) with 10 mM sodium phosphate (pH 6.8) at a flow rat of 6.0 mL/min. The product fraction was then concentrated and further purified by reversed phase ion pairing HPLC using an isocratic elution with 50 mM triethylamonium acetate (pH, 6.5) containing 1.5% acetonitrile at a flow rate of 7.0 mL/min. The final product was then obtained after evaporating the solvent and repeatedly freeze-drying to remove the excess triethylammonium acetate and give **4.4** as a white solid (2.2 mg, 5%) ^{1}H NMR (500 MHz, D_2O) δ 8.08 (s, 1H, H-9), 5.90 (d, J = 6.3 Hz, 1H, H-1′), 5.58 (d, J = 5.9 Hz, 1H, H-1′′), 4.80–4.75 (m, 1H, H-2′), 4.50 (dd, 1H, J = 4.2, 3.8, H-3′), 4.32 (s, 1H, H-4′), 4.21 (s, 1H, H-2′′), 4.20–4.16 (m, 2H, H-5a′/H-5b′), 4.10–4.04 (m, 2H, H-3′′/H-4′′), 3.90 (dt, J = 10.0, 3.2 Hz, 1H, H-5′′), 3.75–3.63 (m, 2H, H-7a′′/H-7b′′), 3.17 (q, J = 7.1 Hz, 8H, 1.3 × $(CH_3C\mathit{H_2})_3N$), 1.84–1.75 (m, 1H, H-6a′′), 1.70–1.60 (m, 1H, H-6b′′), 1.25 (t, J = 7.1 Hz, 12H, 1.3 × $(C\mathit{H_3}CH_2)_3N$); $^{31}P\{^1H\}$ NMR (202 MHz, D_2O) δ –11.34 (d, J = 20.9 Hz), –13.81 (d, J = 20.9 Hz); HRMS (ESI) *m/z* Calc. for $(M–H)^{-2}$ $C_{17}H_{25}N_5O_{16}P_2$: 308.5390. Found: 308.5391. Calc. for (M–H) $C_{17}H_{26}N_5O_{16}P_2$: 618.0855. Found: 618.0852.



1,2,3,5,7-Penta-*O*-acetyl-D-*altro*-heptofuranose (4.5):

A mixture of **4.16** (255 mg, 0.56 mmol) and $Pd(OH)_2$ was suspended in 1:1 EtOAc–CH_3OH and stirred under H_2 gas (1 atm) for 3 days until no more starting material remained by TLC. The catalyst was removed by filtering the reaction mixture through Celite and the filtrate was concentrated to give a yellow syrup. Without further purification, the syrup was dissolved in 1:1 pyridine–CH_2Cl_2 (5 mL) to which was added Ac_2O (211 μL, 2.24 mmol). The reaction mixture was left overnight and the excess Ac_2O was quenched by the addition of H_2O (2 mL). The reaction mixture was concentrated and then dissolved in CH_2Cl_2 (20 mL), washed with 1 M HCl, H_2O, saturated $NaHCO_3$, brine (1 × 10 mL each), and dried over Na_2SO_4. The organic layer was then concentrated, and further purified by column chromatography (2:1 hexanes–EtOAc) to give **4.5** (213 mg, 95%) as a colourless syrup as an 9:10 α:β mixture. R_f 0.30 (2:1 hexanes–EtOAc); $[\alpha]_D$ 16.3 (*c* 0.41, $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.39 (d, *J* = 4.7 Hz, 1H, H-1β), 6.21 (s, 1H, H-1α), 5.54 (dd, *J* = 6.6, 5.3 Hz, 1H, H-3β), 5.32 (dd, *J* = 6.5, 4.7 Hz, 1H, H-2β), 5.27–5.19 (m, 3H, H-3α/H-5α/H-5β), 5.15 (d, *J* = 1.4 Hz, 1H, H-2α), 4.24 (dd, *J* = 5.8, 4.3 Hz, 1H, H-4α), 4.17–4.09 (m, 4H, H-7aα/ H-7bα/H-7aβ/H-7bβ), 4.06 (dd, *J* = 7.3, 5.3 Hz, 1H, H-4β), 2.17–2.05 (m, 2H, H-6aα/H-6aβ), 2.16 (s, 3H, C*H*$_3$CO$_2$), 2.15 (s, 3H, C*H*$_3$CO$_2$), 2.14 (s, 3H, C*H*$_3$CO$_2$), 2.14 (s, 3H, C*H*$_3$CO$_2$), 2.12 (s, 3H, C*H*$_3$CO$_2$), 2.11 (s, 3H,

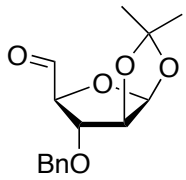
C*H*$_3$CO$_2$), 2.10 (s, 3H, C*H*$_3$CO$_2$), 2.10 (s, 3 H, C*H*$_3$CO$_2$), 2.07 (s, 6H, C*H*$_3$CO$_2$ × 2), 1.98–1.84 (m, 2H, H-6bα/H-6bβ); ^{13}C NMR (126 MHz, $CDCl_3$) δ 170.9 (*C*=O), 170.9 (*C*=O), 170.3 (*C*=O), 170.2 (*C*=O), 169.8 (*C*=O), 169.7 (*C*=O), 169.5 (*C*=O), 169.4 (*C*=O), 169.1 (*C*=O), 169.1 (*C*=O), 99.3 (C-1α), 93.8 (C-1β), 85.4 (C-4α), 82.2 (C-4β), 80.6 (C-2α), 76.0 (C-3α), 75.7 (C-2β), 74.7 (C-3β), 70.3 (C-5β), 69.1 (C-5α), 60.2 (2C, C-7α, C-7β), 29.7 (C-6α or C-6β), 29.5 (C-6α or C-6β), 21.0 (*C*H$_3$CO$_2$), 21.0 (*C*H$_3$CO$_2$), 20.9 (*C*H$_3$CO$_2$), 20.9 (*C*H$_3$CO$_2$), 20.9 (*C*H$_3$CO$_2$), 20.9 (*C*H$_3$CO$_2$), 20.8 (*C*H$_3$CO$_2$), 20.8 (*C*H$_3$CO$_2$), 20.7 (*C*H$_3$CO$_2$), 20.4 (*C*H$_3$CO$_2$); HRMS (ESI) *m/z* Calc. for (M+Na) $C_{17}H_{24}O_{11}Na$: 427.1211. Found: 427.1208.



3-*O*-Benzyl-6-deoxy-1,2-*O*-isopropylidene-β-D-*altro*-heptofuranose (4.6a):

To a solution of **4.13** (320 mg, 1.0 mmol) in CH_2Cl_2 (4 mL) at –70 °C was bubbled ozone until a dark purple colour was observed showing an excess of ozone had been introduced. The ozone atmosphere was displaced by bubbling N_2(g) through the solution and then the reaction mixture was diluted with CH_3OH (4 mL). $NaBH_4$ (113 mg, 3.0 mmol) was then added, and the solution was gradually warmed to room temperature. After 16 h, the reaction mixture was evaporated and purified by column chromatography (1:1 hexanes–EtOAc) to give **4.6a** (176 mg, 54%) as a colourless syrup. R_f 0.31 (1:1 hexanes–EtOAc);

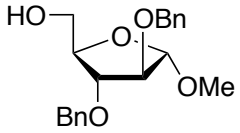
$[\alpha]_D$ 7.6 (*c* 0.3, $CHCl_3$); ^1H NMR (500 MHz, $CDCl_3$) δ 7.44–7.30 (m, 5H, Ar), 5.91 (d, *J* = 4.1 Hz, 1H, H-1), 4.70 (dd, *J* = 4.1, 1.0 Hz, 1H, H-2), 4.64 (AB q, *J* = 11.6 Hz, 2H, OC*H*$_2$Ph), 4.23 (dd, *J* = 3.4, 0.9 Hz, 1H, H-3), 4.10 (app. ddd, *J* = 12.4, 6.3, 3.2 Hz, 1H, H-5), 4.02 (dd, *J* = 6.3, 3.4 Hz, 1H, H-4), 3.96–3.85 (m, 2H, H-7a, H-7b), 2.90 (d, *J* = 3.2 Hz, 1H, 5-OH), 2.22 (dd, *J* = 6.2, 4.1 Hz, 1H, 7-OH), 1.95–1.86 (m, 1H, H-6a), 1.73 (app. dddd, *J* = 14.6, 9.1, 7.8, 4.2 Hz, 1H, H-6b), 1.55 (s, 3 H, C*H*$_3$a), 1.37 (s , 3 H, C*H*$_3$b); ^{13}C NMR (126 MHz, $CDCl_3$) δ 137.3 (Ar), 128.5 (2 × Ar), 128.0 (Ar), 127.9 (2 × Ar), 112.8 (*C*$(CH_3)_2$), 105.5 (C-1), 88.0 (C-4), 85.4 (C-2), 82.4 (C-3), 71.8 (O*C*H_2Ph), 71.0 (C-5), 61.4 (C-7), 34.5 (C-6), 27.2 (*C*H_3b), 26.4 (*C*H_3a); HRMS (ESI) *m/z* Calc. for (M+Na) $C_{17}H_{24}O_6Na$: 347.1465. Found: 347.1462.



3-*O*-Benzyl-1,2-*O*-isopropylidene-β-D-*arabino*-dialdo-1,4-furanose (4.7b):

DMSO (1.03 mL, 14.5 mmol) was added slowly to a solution of oxalyl chloride (4.35 mL of 2 M in CH_2Cl_2, 8.7 mmol) in CH_2Cl_2 (19 mL) at –70 °C, and stirred for 30 min. A solution of **4.12** (1 g, 3.57 mmol) in CH_2Cl_2 (26 mL) was then slowly added to this mixture. After 30 min, Et_3N (2.5 mL, 17.85 mmol) was added, and the reaction was allowed to proceed at –70 °C for 1 h followed by 2 h at room temperature. The reaction mixture was diluted with CH_2Cl_2 (50 mL), washed with H_2O (1 × 30 mL), brine (1 × 30 mL) and dried over Na_2SO_4.

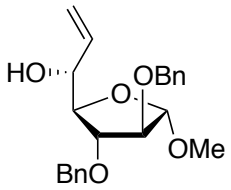
The solvent was removed by evaporation and the crude oil was purified by column chromatography (6:1 → 2:1 hexanes–EtOAc) to give **4.7b** (496 mg, 50%) as a yellow oil. R_f 0.37 (2:1 hexanes-EtOAc); $[\alpha]_D$ –20.4 (*c* 0.5, $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$) δ 9.81 (s, 1H, H-5), 7.41–7.29 (m, 5H, Ar), 6.09 (d, *J* = 3.5 Hz, 1H, H-1), 4.64 (AB q, *J* = 11.7 Hz, 2H, OC*H*$_2$Ph), 4.64 (d, *J* = 3.6 Hz, 1H, H-2), 4.55 (s, 1H, H-4), 4.35 (s, 1H, H-3), 1.47 (s, 3H, C*H*$_3$a), 1.32 (s, 3H, C*H*$_3$b); ^{13}C NMR (126 MHz, $CDCl_3$) δ 201.7 (C-5), 136.7 (Ar), 128.6 (2 × Ar), 128.2 (Ar), 127.9 (2 × Ar), 112.2 (*C*$(CH_3)_2$), 106.6 (C-1), 88.7 (C-4), 84.9 (C-3), 82.9 (C-2), 72.0 (O*C*H_2Ph), 26.2 (*C*H_3b), 25.6 (*C*H_3a); HRMS (ESI) *m/z* Calc. for (M+Na) $C_{15}H_{18}O_5Na$: 301.1046. Found: 301.1044.

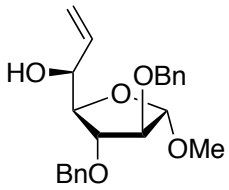


Methyl 2,3-di-*O*-benzyl-α-D-arabinofuranoside (4.9):

To a solution of **4.8** (4.0 g, 24.35 mmol) in pyridine (60 mL) was added DMAP (744 mg, 6.09 mmol) and chlorotriphenylmethane (8.14 g, 29.22 mmol). The reaction mixture was heated at 50 °C for 20 h, and the excess chlorotriphenylmethane was quenched by pouring over ice cold H_2O (200 mL). The solution was extracted with CH_2Cl_2 (2 × 100 mL), and the organic layer was dried over Na_2SO_4. The solvent was removed through evaporation under reduced pressure, and through co-evaporation with toluene (2 × 50 mL). The crude residue was further purified by column chromatography (2:1 hexanes–EtOAc) and the solvent evaporated to give a yellow oil. The oil was dissolved in 1:1

DMF–THF (100 mL) and cooled to 0 °C. To this solution was added 60% NaH in mineral oil (2.03 g, 50.75 mmol) followed by benzyl bromide (5.8 mL, 48.72 mmol). The reaction mixture was stirred at 0 °C for 30 min followed by 16 h at ambient temperature. The excess NaH was quenched through the addition of CH_3OH (5 mL) at 0 °C. The reaction mixture was washed with H_2O (1 × 100 mL) and extracted with CH_2Cl_2 (2 × 100 mL). The organic layer was further washed with H_2O (2 × 100 mL), dried over Na_2SO_4, filtered and concentrated. The crude oil was purified by column chromatography (14:1 hexanes–EtOAc) and concentrated to give a yellow oil, which was then dissolved in 5:1 (CH_2Cl_2–CH_3OH). To this solution was added *p*-toluenesulfonic acid (331 mg, 1.74 mmol) and the reaction mixture was stirred for 5 h at room temperature. The solution was neutralized with Et_3N and concentrated. The residue was then purified by column chromatography (2:1 hexanes–EtOAc) to give **4.9** (5.99 g, 71%) as a colourless oil. R_f 0.32 (2:1 hexanes–EtOAc); $[\alpha]_D$ 83.2 (*c* 0.5), $CHCl_3$; ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41–7.29 (m, 10H, Ar), 4.96 (s, 1H, H-1), 4.81–4.47 (m, 4H, OC*H*$_2$Ph), 4.17 (ddd, *J* = 6.2, 4.0, 2.9 Hz, 1H, H-4), 4.03 (dd, *J* = 2.6, 1.0 Hz, 1H, H-2), 4.00 (ddd, *J* = 6.1, 2.6, 0.6 Hz, 1H, H-3), 3.86 (ddd, *J* = 12.0, 4.5, 2.9 Hz, 1H, H-5a), 3.67 (ddd, *J* = 12.0, 8.0, 4.1 Hz, 1H, H-5b), 3.41 (s, 3H, OC*H*$_3$), 1.90 (dd, *J* = 8.0, 4.5 Hz, 1H, 5-O*H*); ^{13}C NMR (126 MHz, $CDCl_3$) δ 137.7 (Ar), 137.4 (Ar), 128.5 (Ar × 2), 128.5 (Ar × 2), 128.0 (Ar), 127.9 (Ar × 2), 127.9 (Ar), 127.9 (Ar × 2), 107.5 (C-1), 87.8 (C-2), 82.6 (C-3), 82.4 (C-4), 72.4 (O*C*H$_2$Ph), 71.9 (O*C*H$_2$Ph), 62.3 (C-5), 54.9 (O*C*H$_3$); HRMS (ESI) *m/z* Calc. for (M+Na) $C_{20}H_{24}O_5Na$: 367.1516. Found: 367.1517.





Methyl 2,3-di-*O*-benzyl-6,7-dideoxy-β-D-*altro*-hept-6-enofuranoside (4.10a) and **Methyl 2,3-di-*O*-benzyl-6,7-dideoxy-β-L-*galacto*-hept-6-enofuranoside (4.10b):**

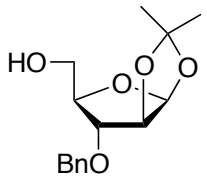
DMSO (1.03 mL, 14.5 mmol) was added slowly to a solution of oxalyl chloride (4.35 mL of 2 M in CH_2Cl_2, 8.7 mmol) in CH_2Cl_2 (26 mL) at –70 °C, and the mixture was stirred for 30 min. A solution of **4.9** (1 g, 2.9 mmol) in CH_2Cl_2 (26 mL) was then slowly added to this mixture while keeping the temperature at –70 °C. After 1 h, Et_3N (2.0 mL, 14.5 mmol) was added, and the reaction was allowed to proceed at –70 °C for 45 min followed by 1 h at room temperature. The reaction mixture was diluted with CH_2Cl_2 (25 mL), washed with H_2O (1 × 30 mL), brine (1 × 30 mL) and dried over Na_2SO_4. The solvent was removed by evaporation and thoroughly dried to give **4.7a** as a clear yellow syrup that was used without further purification.

To a solution of half of the crude **4.7a** (~500 mg, 1.45 mmol) in THF (20 mL) at –70 °C was added a solution of vinylmagnesium bromide (3.6 mL of 1.0 M in THF, 3.6 mmol). The reaction mixture was then stirred at –70 °C for 2 h before the unreacted Grignard reagent was quenched by the addition of a saturated aqueous solution of ammonium chloride (10 mL). The solution was extracted twice with CH_2Cl_2 (20 mL), and the combined organic layers were further washed with H_2O (1 × 25 mL), saturated $NaHCO_3$ (1 × 25 mL), and brine

(1 × 25 mL). The solution was then dried over Na_2SO_4 and the solvent removed by evaporation to give a crude syrup that was purified by successive column chromatography (6:1 hexanes–EtOAc × 3 columns) to give both **4.10a** (136 mg, 25%) and **4.10b** (193 mg, 36%). **4.10a** R_f 0.48 (2:1 hexanes–EtOAc); $[\alpha]_D$ 79.3 (*c* 0.7), $CHCl_3$; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.41–7.26 (m, 10H, Ar), 5.83 (ddd, *J* = 17.1, 10.7, 5.1 Hz, 1H, H-6), 5.41 (dt, *J* = 17.2, 1.7 Hz, 1H, H-7*cis*), 5.22 (dt, *J* = 10.6, 1.7 Hz, 1H, H-7*trans*), 4.97 (s, 1H, H-1), 4.55–4.45 (m, 4H, OC*H*$_2$Ph), 4.45–4.40 (m, 1H, H-5), 4.15 (dd, *J* = 4.9, 3.2 Hz, 1H, H-4), 3.99 (ddd, *J* = 4.9, 1.5, 0.8 Hz, 1H, H-3), 3.96 (dd, *J* = 1.5, 0.5 Hz, 1H, H-2), 3.40 (s, 3H, OC*H*$_3$), 2.41 (d, *J* = 2.5 Hz, 1H, 5-O*H*); ^{13}C NMR (126 MHz, $CDCl_3$) δ 137.7 (Ar), 137.2 (Ar), 135.7 (C-6), 128.5 (Ar × 2), 128.4 (Ar × 2), 128.2 (Ar ×2), 128.0 (Ar), 128.0 (Ar × 2), 127.9 (Ar), 116.6 (C-7), 107.4 (C-1), 86.6 (C-2), 85.7 (C-4), 81.1 (C-3), 72.1 (O*C*H_2Ph), 71.7 (C-5), 71.6 (O*C*H_2Ph), 54.9 (O*C*H_3); HRMS (ESI) *m/z* Calc. for (M+Na) $C_{22}H_{26}O_5Na$: 393.1672. Found: 393.1680.

4.10b R_f 0.44 (2:1 hexanes–EtOAc); $[\alpha]_D$ 65.1 (*c* 0.56), $CHCl_3$; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.48–7.24 (m, 10H, Ar), 5.90 (ddd, *J* = 17.2, 10.5, 5.7 Hz, 1H, H-6), 5.34 (dt, *J* = 17.2, 1.5 Hz, 1H, H-7*cis*), 5.20 (dt, *J* = 10.5, 1.4 Hz, 1H, H-7*trans*), 4.96 (s, 1H, H-1), 4.62–4.46 (m, 4H, OC*H*$_2$Ph), 4.20–4.12 (m, 1H, H-5), 4.09–4.04 (m, 1H, H-4), 4.03–3.93 (m, 2H, H-2/H-3), 3.38 (s, 3H, OC*H*$_3$), 2.27 (d, *J* = 6.8 Hz, 1H, 5-O*H*); ^{13}C NMR (126 MHz, $CDCl_3$) δ 137.7 (Ar), 137.3 (Ar), 137.1 (C-6), 128.5 (Ar × 2), 128.5 (Ar × 2), 128.0 (Ar × 2), 128.0 (Ar), 127.9 (Ar × 3), 116.6 (C-7), 107.4 (C-1), 87.3 (C-2), 84.5 (C-4), 83.2 (C-

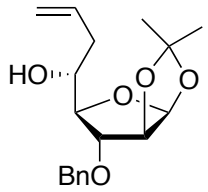
3), 72.7 (C-5), 72.3 (O*C*H_2Ph), 71.9 (O*C*H_2Ph), 54.9 (O*C*H_3); HRMS (ESI) *m/z* Calc. for (M+Na) $C_{22}H_{26}O_5Na$: 393.1672. Found: 393.1668.



3-*O*-Benzyl-1,2-*O*-isopropylidene-β-D-arabinofuranose (4.12):

To a suspension of D-arabinose (5 g, 33 mmol) and DMAP (410 mg, 3.3 mmol) in pyridine (50 mL) and CH_2Cl_2 (50 mL) at 0 °C was added TBDPSCl (9.5 mL, 37 mmol) slowly over 10 min. The reaction mixture was warmed to room temperature and stirred for 48 h. CH_3OH (80 mL) was added to quench the excess TBDPSCl and the mixture was stirred for 30 min. The solvent was removed by evaporation under reduced pressure, and then co-evaporated with 4:1 Toluene–EtOH (3 × 10 mL) to give a colourless oil. The oil was dissolved in $CHCl_3$ (40 mL) and washed with 0.5 M HCl, H_2O, and brine (1 × 20 mL each), dried over Na_2SO_4 and concentrated to give a colourless oil. Without further purification, this oil was dissolved in acetone (60 mL). To this solution was added 2,2-dimethoxypropane (60 mL, 489 mmol) and *p*-toluenesulfonic acid (560 mg, 3 mmol). The reaction mixture was stirred at room temperature for 3 h until all starting material had been consumed. The reaction mixture was then diluted with CH_3OH (60 mL) and $CHCl_3$ (50 mL) and neutralized using a saturated solution of aqueous $NaHCO_3$. The solvent was removed by evaporation under reduced pressure and the resulting residue was dissolved in CH_2Cl_2 (60

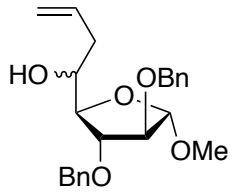
mL) and washed with brine (1 × 50 mL). The organic layer was dried over Na_2SO_4 then concentrated to give a colourless oil. A solution of the oil in THF (50 mL) was added to a solution of 60% NaH (1.18 g, 29.55 mmol) in THF (150 mL) at 0 °C and stirred at this temperature for 1 h before adding benzyl bromide (3.5 mL, 29.55 mmol). The reaction mixture was gradually allowed to warm to room temperature over 1 h, followed by the addition of tetrabutylammonium iodide (190 mg, 0.51 mmol). After 16 h the excess NaH was quenched through the addition of saturated ammonium chloride (25 mL). The mixture was extracted with CH_2Cl_2 (1 × 100 mL), the organic layer was washed with brine (1 × 100 mL) and dried over Na_2SO_4. The mixture was concentrated and purified by column chromatography (12:1 hexanes–EtOAc) to give a yellow oil. This oil was finally dissolved in THF (30 mL) and 1 M TBAF (20 mL, 20 mmol) was added to this solution. After 16 h, the reaction mixture was concentrated and purified by column chromatography (6:1 → 4:1 hexanes–EtOAc) to give **4.12** (4.36 g, 47% over 4 steps) as a colourless oil. R_f 0.31 (2:1 hexanes–EtOAc); $[\alpha]_D$ 24.3 (*c* 0.6, $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.42–7.31 (m, 5 H, Ar), 5.94 (d, *J* = 4.1 Hz, 1H, H-1), 4.71 (d, *J* = 4.1 Hz, 1H, H-2), 4.63 (AB q, *J* = 11.7 Hz, 2H, OC*H*$_2$Ph), 4.23 (app dt, *J* = 5.4, 3.5 Hz, 1H, H-4), 4.00 (br d, *J* = 3.5 Hz, 1H, H-3), 3.79–3.74 (m, 2H, H-5a/H-5b), 2.04 (app t, *J* = 6.4 Hz, 1H, 5-O*H*), 1.56 (s, 3H, C*H*$_3$a), 1.37 (s, 3H, C*H*$_3$b); ^{13}C NMR (126 MHz, $CDCl_3$) δ 137.2 (Ar), 128.6 (2 × Ar), 128.0 (Ar), 127.8 (2 × Ar), 112.9 (*C*$(CH_3)_2$), 105.6 (C-1), 85.6 (C-3), 85.3 (C-2), 82.8 (C-4), 71.9 (O*C*H_2Ph), 62.7 (C-5), 27.2 (*C*H_3b), 26.4 (*C*H_3a); HRMS (ESI) *m/z* Calc. for (M+Na) $C_{15}H_{20}O_5Na$: 303.1203. Found: 303.1201.



3-*O*-Benzyl-6,7,8-trideoxy-1,2-*O*-isopropylidene-β-D-*altro*-oct-7-enofuranose (4.13):

A solution of **4.7b** (314 mg, 1.1 mmol) in THF (5 mL) was added to a suspension of Zn dust (144 mg, 2.2 mmol) in THF (5 mL). The mixture was cooled to 0 °C and allyl bromide (190 μL, 2.2 mmol) was added. The solution was allowed to warm to room temperature. After 16 h, saturated ammonium chloride (1.25 mL) was added. The reaction mixture was then filtered through Celite to remove excess salts and the flow through was washed with H_2O (1 × 10 mL) and extracted with EtOAc (3 × 10 mL). The combined organic layers were washed with brine (1 × 15 mL), dried over Na_2SO_4 and evaporated to give a colourless syrup. This syrup was further purified by column chromatography (8:1 → 6:1 hexanes–EtOAc) to give **4.13** (283 mg, 81%) as a colourless oil. R_f 0.32 (4:1 hexanes–EtOAc); $[\alpha]_D$ 11.4 (*c* 0.5, $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.40–7.31 (m, 5H, Ar), 5.92 (d, *J* = 4.1 Hz, 1H, H-1), 5.90–5.82 (m, 1H, H-7), 5.21–5.15 (m, 2H, H8a/H8b), 4.69 (dd, *J* = 4.1, 1.0 Hz, 1H, H-2), 4.64 (AB q, *J* = 11.6 Hz, 2H, OC*H*$_2$Ph), 4.23 (dd, *J* = 3.2, 1.0 Hz, 1H, H-3), 3.99 (dd, *J* = 6.4, 3.2 Hz, 1H, H-4), 3.90 (app ddt, *J* = 8.0, 6.4, 3.9 Hz, 1H, H-5), 2.51–2.44 (m, 1H, H-6a), 2.26–2.17 (m, 1H, H-6b), 2.13 (d, *J* = 3.6 Hz, 1H, 5-O*H*), 1.55 (s, 3H, C*H*$_3$a), 1.37 (s, 3H, C*H*$_3$b); ^{13}C NMR (126 MHz, $CDCl_3$) δ 137.3 (Ar), 134.1 (C-7), 128.5 (2 × Ar), 128.0 (Ar), 127.9 (2 × Ar), 118.5 (C-8), 112.8

($C(CH_3)_2$), 105.5 (C-1), 87.7 (C-4), 85.4 (C-2), 82.4 (C-3), 71.8 (OCH_2Ph), 70.1 (C-5), 37.5 (C-6), 27.2 (CH_3b), 26.4 (CH_3a); HRMS (ESI) *m/z* Calc. for (M+Na) $C_{18}H_{24}O_5Na$: 343.1516. Found: 343.1513.

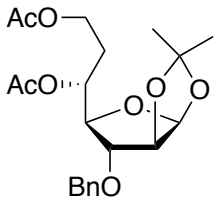


Methyl 2,3-di-*O*-benzyl-6,7,8-trideoxy-α-D-*altro*-oct-7-enofuranoside (4.14a) and methyl 2,3-di-*O*-benzyl-6,7,8-trideoxy-β-L-*galacto*-oct-7-enofuranoside (4.14b):

A solution of **4.7a** (100mg, 0.29 mmol) in THF (1 mL) was added to a suspension of Zn dust (38 mg, 0.58 mmol) in THF (1 mL). The mixture was cooled to 0 °C and allyl bromide (50 μL, 0.58 mmol) was added. The reaction was warmed to room temperature over 45 min, and then stirred at room temperature for 2 h, at which point TLC showed complete consumption of **4.7a**. Saturated ammonium chloride (0.3 mL) was added, the reaction mixture was filtered through Celite to remove excess salts and the flow through was washed with H_2O (1 × 5 mL) and extracted with EtOAc (3 × 5 mL). The combined organic layers were washed with brine (1 × 10 mL), dried over Na_2SO_4 and evaporated to give a colourless syrup. ^{1}H NMR spectroscopy of the product mixture revealed a 2:3 ratio of diastereomers. Column chromatography (6:1 hexanes–EtOAc) gave the products **4.14a** and **4.14b** (yield not determined):

4.14a; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.41–7.29 (m, 10H, Ar), 5.87 (ddt, *J* = 17.1, 10.4, 7.1 Hz, 1H, H-7), 5.19–5.12 (m, 2H, H-8a/H-8b), 4.97 (s, 1H, H-1), 4.61 (app d, *J* = 12.0 Hz, 1H, OC*H*$_2$Ph), 4.57–4.49 (m, 3H, OC*H*$_2$Ph), 4.10 (dd, *J* = 5.4, 1.4 Hz, 1H, H-3), 4.06 (dd, *J* = 5.4, 3.6 Hz, 1H, H-4), 4.00 (d, *J* = 1.4 Hz, 1H, H-2), 3.90 (app ddd, *J* = 8.0, 5.5, 3.6 Hz, 1H, H-5), 3.40 (s, 3H, OC*H*$_3$), 2.37–2.11 (m, 2H, H-6a/H-6b), 2.22 (d, *J* = 2.4 Hz, 1H, 5-O*H*).

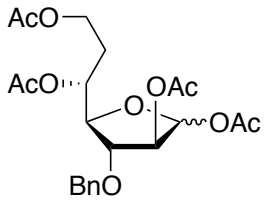
4.14b; 1H NMR (400 MHz, $CDCl_3$) δ 7.41–7.30 (m, 10H, Ar), 5.87 (ddt, *J* = 17.3, 10.2, 7.1 Hz, 1H, H-7), 5.19–5.10 (m, 2H, H-8a/H-8b), 4.96 (s, 1H, H-1), 4.61 (d, *J* = 12.0 Hz, 1H, OC*H*$_2$Ph), 4.57–4.47 (m, 3H, OC*H*$_2$Ph), 4.03–3.96 (m, 3H, H-2/H-3/H-4), 3.76–3.65 (m, 1H, H-5), 3.40 (s, 3H, OC*H*$_3$), 2.40–2.09 (m, 2H, H-6a, H-6b), 2.13 (d, *J* = 5.1 Hz, 1H, 5-O*H*).



5,7-Di-*O*-acetyl-3-*O*-benzyl-6-deoxy-1,2-*O*-isopropylidene-β-D-*altro*-heptofuranose (4.15):

To a solution of **4.6a** (250 mg, 0.77 mmol) in 1:1 pyridine–CH_2Cl_2 (7.7 mL) was added Ac_2O (290 μL, 3.08 mmol). The reaction mixture was left overnight and the excess Ac_2O was quenched by the addition of H_2O (2 mL). The reaction mixture was concentrated then dissolved in CH_2Cl_2 (20 mL), washed with 1 M HCl, H_2O, saturated $NaHCO_3$, and brine (1 × 10 mL each). The organic layer was dried over Na_2SO_4, concentrated, and further purified by

column chromatography (3:1 hexanes–EtOAc) to give **4.14** (277 mg, 88%) as a colourless syrup. R_f 0.54 (2:1 hexanes–EtOAc); $[\alpha]_D$ 30.2 (*c* 0.4, $CHCl_3$); ^1H NMR (500 MHz, $CDCl_3$) δ 7.46–7.31 (m, 5H), 5.95 (d, *J* = 4.0 Hz, 1H, H-1), 5.31 (ddd, *J* = 8.8, 8.8, 3.3 Hz, 1H, H-5), 4.67 (d, *J* = 4.1 Hz, 1H, H-2), 4.61 (AB q, *J* = 12.0 Hz, 2H, OC*H*$_2$Ph), 4.16–4.12 (m, 2H, H-7a/H-7b), 4.08 (dd, *J* = 8.9, 2.3 Hz, 1H, H-4), 3.95 (d, *J* = 2.1 Hz, 1H, H-3), 2.24 (ddd, *J* = 14.7, 7.3, 3.4 Hz, 1H, H-6a), 2.05 (s, 3H, C*H*$_3$CO$_2$), 1.97 (s, 3H, C*H*$_3$CO$_2$), 1.94–1.85 (m, 1H, H-6b), 1.61 (s, 3H, C*H*$_3$a), 1.36 (s, 3H, C*H*$_3$b); ^{13}C NMR (126 MHz, $CDCl_3$) δ 170.9 (*C*=O), 169.9 (*C*=O), 137.1 (Ar), 128.5 (2 × Ar), 128.0 (Ar), 127.8 (2 × Ar), 112.8 (*C*(CH$_3$)$_2$), 105.9 (C-1), 85.3 (C-4), 84.6 (C-2), 82.5 (C-3), 71.5 (OCH$_2$Ph), 69.7 (C-5), 60.4 (C-7), 30.3 (C-6), 26.8 (*C*H$_3$b), 26.1 (*C*H$_3$a), 20.9 (*C*H$_3$CO$_2$), 20.9 (*C*H$_3$CO$_2$); HRMS (ESI) *m/z* Calc. for (M+Na) $C_{21}H_{28}O_8Na$: 431.1672. Found: 431.1676.

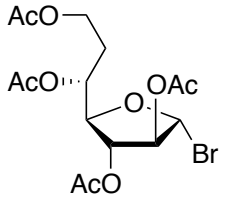


1,2,5,7-Tetra-*O*-acetyl-3-*O*-benzyl-6-deoxy-D-*altro*-heptofuranose (4.16):

A solution of **4.15** (260 mg, 0.63 mmol) in 50% aqueous AcOH (15 mL) was heated at 80 °C for 8 h. The solvent was removed through co-evaporation with toluene (3 × 10 mL) and the resulting residue was dissolved in 1:1 pyridine–CH_2Cl_2 (6.3 mL). To the solution was added Ac_2O (500 μL, 5.3 mmol) and the reaction mixture was stirred at room temperature for 16 h. Excess Ac_2O

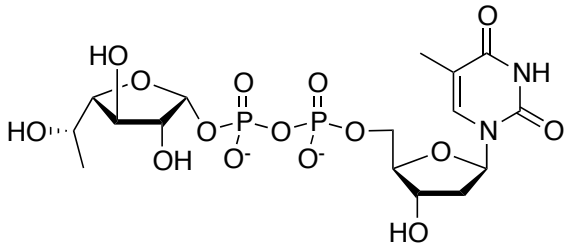
was quenched by adding H_2O (3 mL) and the solvent was removed by evaporation. The residue was dissolved in CH_2Cl_2 (20 mL), washed with 1M HCl, H_2O, saturated $NaHCO_3$, brine (1 × 10 mL each), and dried over Na_2SO_4. The organic layer was concentrated and further purified by column chromatography (4:1 hexanes–EtOAc) to give **4.16** (267 mg, 94%) as a colourless syrup as a 9:10 α:β mixture. R_f 0.30 (2:1 hexanes–EtOAc); $[\alpha]_D$ 31.2 (*c* 0.4, $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41–7.29 (m, 10H, Ar), 6.38 (d, *J* = 4.6 Hz, 1H, H-1β), 6.22 (d, *J* = 0.4 Hz, 1H, H-1α), 5.26 (dd, *J* = 6.4, 4.6 Hz, 1H, H-2β), 5.24–5.16 (m, 3H, H-5β/H-2α/H-5α), 4.79 (AB q, *J* = 12.0 Hz, 1H, OC*H*$_2$Ph), 4.65–4.57 (m, 3H, OC*H*$_2$Ph), 4.28 (dd, *J* = 6.3, 5.6 Hz, 1H, H3β), 4.25 (dd, *J* = 5.8, 4.3 Hz, 1H, H-4α), 4.17–4.03 (m, 5H, H-7aα/H-7bα/H-7aβ/H-7bβ/H-4β), 3.92 (dt, *J* = 4.3, 0.9 Hz, 1H, H-3α), 2.18–1.95 (m, 2H, H-6aα/H-6aβ), 2.15 (s, 3H, C*H*$_3$CO$_2$), 2.13 (s, 3H, C*H*$_3$CO$_2$), 2.11 (s, 3H, C*H*$_3$CO$_2$), 2.08 (s, 3H, C*H*$_3$CO$_2$), 2.05 (s, 3H, C*H*$_3$CO$_2$), 2.04 (s, 3H, C*H*$_3$CO$_2$), 2.04 (s, 3H, C*H*$_3$CO$_2$), 1.97 (s, 3H, C*H*$_3$CO$_2$) 1.93–1.77 (m, 2H, H-6bα/H-6bβ); ^{13}C NMR (126 MHz, $CDCl_3$) δ 170.9 (*C*=O), 170.8 (*C*=O), 170.2 (*C*=O), 170.1 (*C*=O), 169.7 (*C*=O), 169.6 (*C*=O), 169.5 (*C*=O), 169.1 (*C*=O), 137.3 (Ar), 137.2 (Ar), 128.5 (Ar × 2), 128.5 (Ar × 2), 128.1 (Ar), 128.0 (Ar × 3), 127.7 (Ar × 2), 99.8 (C-1α), 94.1 (C-1β), 85.9 (C-4α), 82.5 (C-3α), 82.3 (C-4β), 81.0 (C-3β), 80.2 (C-2α), 76.5 (C-2β), 72.3 (O*C*H$_2$Ph), 72.0 (O*C*H$_2$Ph), 70.7 (C-5β), 69.3 (C-5α), 60.3 (2C, C-7α, C-7β), 29.9 (C-6α), 29.8 (C-6β), 21.1 (*C*H$_3$CO$_2$), 21.0 (*C*H$_3$CO$_2$), 21.0 (*C*H$_3$CO$_2$), 20.9 (*C*H$_3$CO$_2$), 20.9 (*C*H$_3$CO$_2$), 20.8 (*C*H$_3$CO$_2$ × 2),

20.5 (CH_3CO_2); HRMS (ESI) *m/z* Calc. for (M+Na) $C_{22}H_{28}O_{10}Na$: 475.1575. Found: 475.1574.



2,3,5,7-Tetra-*O*-acetyl-α-D-*altro*-heptofuranosyl bromide (4.17):

To a solution of **4.5** (22 mg, 55 μmol) in CH_2Cl_2 (0.6 mL) at –10 °C was added a solution of 33% HBr in AcOH (37 μL, 200 μmol). After 30 min, the reaction was complete by TLC, and the solvent was removed by co-evaporation with dry toluene (3 × 5 mL) to give **4.17** as a yellow syrup, and was used directly without further purification (See above);); ^{1}H NMR (500 MHz, CD_2Cl_2) δ 6.34 (d, *J* = 1.6 Hz, 1H, H-1), 5.41 (br. s, 1H, H-2), 5.35–5.32 (m, 1H, H-5), 5.22 (ddd, *J* = 0.9, 1.6, 4.8 Hz, 1H, H-3), 4.45 (dd, *J* = 4.8, 5.2 Hz, 1H, H-4), 4.11–4.08 (m, 2H, H-7a/H-7b), 2.14 (s, 3H, C*H*$_3$CO$_2$), 2.10 (s, 3H, C*H*$_3$CO$_2$), 2.06 (s, 3H, C*H*$_3$CO$_2$), 2.03 (s, 3H, C*H*$_3$CO$_2$), 2.06–1.96 (m, 1H, H-6a), 1.93–1.87 (m, 1H, H-6b).

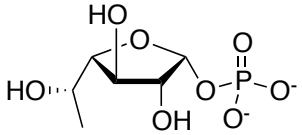


Thymidine 5′-diphospho-6-deoxy-α-L-altrofuranose disodium salt (4.21):

The soluble Cps2L protein was prepared from *E. coli* BL21 (DE3) cells containing recombinant pSK001 plasmid as previously described.[46]

To a solution of 6d-L-Alt*f*-1-phosphate **4.24** (3.6 mg, 12.5 μmol) in 40 mM Tris-HCl reaction buffer (pH 7.4) containing 2.5 mM $MgCl_2$ was added dTTP (4.6 mg, 9.1 μmol), IPP (5 U), and Csp2L protein (0.8 mg) for a final volume of 2.0 mL. After incubating for 24 h at ambient temperature under a atmosphere of N_2(g) with gentle rotation, analysis of the reaction by HPLC (using the same conditions described above for monitoring the Glf2 reaction) indicated complete consumption of dTTP. The reaction mixture was then incubated for 5 h with alkaline phosphatase (AP, 10 U) to degrade unwanted TDP and dTMP byproducts in the reaction mixture. The soluble enzymes were removed using a centrifugal filter device with a molecular weight cut off of 10,000 Da. The resulting filtrate was purified by reverse phase semi-preparative HPLC using conditions previously described for the purification of UDP-D-Gal*f*.[45, 47] The pure fractions were combined, and the volume was reduced to 5 mL by evaporation under reduced pressure. The salts were then removed by gel filtration chromatography using a Sephadex G-15 column (2.5 × 100 mm) eluting with Milli-Q H_2O at a flow rate of ≥1 mL per min. The fractions

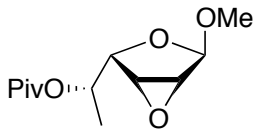
containing the purified product were combined and lyophilized to give the sodium salt of TDP-6d-L-Alt*f* **4.21** as a white powder (4.6 mg, 63%); ^{1}H NMR (500 MHz, D_2O) δ 7.82 (s, 1H, H-6), 6.42 (app t, *J* = 6.8 Hz, 1H, H-1′), 5.69 (app t, *J* = 4.9 Hz, 1H, H-1′′), 4.72–4.67 (m, 1H, H-3′), 4.33 (app t, *J* = 7.9 Hz, 1H, H-3′′), 4.29–4.23 (m, 3H, H-4′/H-5a′, H-5b′), 4.21–4.17 (m, 1H, H-2′′), 4.09 (quintet, 1H, *J* = 6.6, H-5′′), 3.84 (dd, *J* = 6.6, 4.7 Hz, 1H, H-4′′), 2.50–2.40 (m, 2H, H-2a′/H-2b′), 2.00 (d, *J* = 1.0 Hz, 3H, 5-C*H*$_3$), 1.27 (d, *J* = 6.7 Hz, 3H, H-6′′); ^{13}C NMR (126 MHz, D_2O) δ 167.6 (C-4), 152.7 (C-2), 138.3 (C-6), 112.7 (C-5), 98.3 (d, *J* = 6.1 Hz, C-1′′), 86.4 (C-1′), 86.3 (C-4′), 86.0 (C-4′′), 78.0 (d, *J* = 7.9 Hz, C-2′′), 73.9 (C-3′′), 72.0 (C-3′), 68.1 (C-5′′), 66.4 (d, *J* = 5.7 Hz, C-5′), 39.6 (C-2′), 17.7 (C-6′′), 12.6 (5-*C*H$_3$); ^{31}P{^{1}H} NMR (162 MHz, D_2O) δ –11.42 (d, *J* = 21.0 Hz, 1P), –12.73 (d, *J* = 21.0 Hz, 1P); HRMS (ESI) *m/z* Calc. for (M–H)$^-$ $C_{16}H_{25}O_{15}P_2$: 547.0736. Found: 547.0736.



6-Deoxy-α-L-altrofuranosyl phosphate disodium salt (4.24):

To a solution of **4.32** (250 mg, 0.34 mmol) in EtOAc (6.8 mL) was added 10% Pd–C (15% by weight) and Et_3N (170 μL, 1.22 mmol). The reaction mixture was stirred under H_2 (1 atm) for 16 h, and then the catalyst was removed by filtration and the filtrate was evaporated to give a colourless syrup. The resulting residue was dissolved in CH_3OH (10 mL) containing Na_2CO_3 (72 mg, 0.68 mmol) and the solution was stirred at ambient temperature for 24 h until

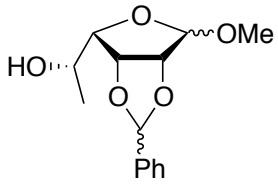
TLC showed complete consumption of the starting material. The solvent was removed by evaporation, and the product was purified using reversed phase C_{18} column chromatography eluting with H_2O. The fractions containing the product were combined and lyophilized to give the disodium salt **4.24** (46.9 mg, 50%) as a white solid. $[\alpha]_D$ 83.1 (*c* 0.3, H_2O); ^{1}H NMR (500 MHz, D_2O) δ 5.51 (app t, *J* = 4.5 Hz, 1H, H-1), 4.36 (app t, *J* = 7.7 Hz, 1H, H-3), 4.17–4.07 (m, 2H, H-2/H-5), 3.81 (dd, *J* = 7.3, 3.1 Hz, 1H, H-4), 1.23 (d, *J* = 6.7 Hz, 3H, H-6); ^{13}C NMR (126 MHz, D_2O) δ 96.7 (d, *J* = 9.8 Hz, C-1), 85.3 (C-4), 78.4 (d, *J* = 7.2 Hz, C-2), 73.3 (C-3), 67.6 (C-5), 17.1 (C-6); ^{31}P{^{1}H} NMR (162 MHz, D_2O) δ 2.05; HRMS (ESI) *m/z* Calc. for $(M–H)^-$ $C_6H_{12}O_8P$: 243.0275. Found: 243.0271.



Methyl 2,3-anhydro-5-*O*-pivaloyl-α-L-rhamnofuranoside (4.25):

To the bromide **4.28** (370 mg, 0.86 mmol) in dry CH_3OH (12 mL) was added a 1.0 M solution of sodium methoxide (1.0 mL, 1.0 mmol). The reaction was left for 16 h and then IR 120 H^+ resin was added to achieve a neutral pH (~ 7-8). The resin was removed by filtration and the solvent was evaporated under reduced pressure. The resulting residue was purified by column chromatography (9:1 Hexanes–EtOAc) to give **4.25** (168 mg, 80%) as a colorless liquid. R_f 0.27 (9:1 Hexanes–EtOAc); $[\alpha]_D$ –48.1 (*c* 0.4, $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.01 (dq, *J* = 7.2, 6.3 Hz, 1H, H-5), 4.96 (s, 1H, H-1), 3.96 (dd, *J* = 7.3, 0.6 Hz, 1H, H-4), 3.67 (dd, *J* = 2.8, 0.7 Hz, 1H, H-3), 3.65 (d, *J* = 2.8 Hz, 1H, H-2), 3.44

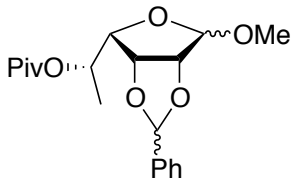
(s, 3H, OCH_3), 1.34 (d, J = 6.3 Hz, 3H, H-6), 1.24 (s, 9H, C(CH_3)$_3$); ^{13}C NMR (126 MHz, $CDCl_3$) δ 177.3 (C=O), 102.3 (C-1), 77.9 (C-4), 68.8 (C-5), 56.0 (C-2), 55.5 (OCH$_3$), 53.6 (C-3), 38.8 (C(CH$_3$)$_3$), 27.1 (C(CH$_3$)$_3$), 17.1 (C-6); HRMS (ESI) *m/z* Calc. for (M+H) $C_{12}H_{21}O_5$: 245.1384. Found: 245.1384.



Methyl 2,3-*O*-benzylidene-α-L-rhamnofuranoside (4.26):

To a suspension of L-rhamnose (2.5 g, 13.70 mmol) in CH_3OH (10.0 mL) was added benzaldehyde dimethyl acetal (13.2 mL, 89.0 mmol), and *p*-toluenesulfonic acid (0.51 g, 2.7 mmol). The reaction mixture was then heated to reflux for 5 days. The reaction mixture was poured over a saturated solution of $NaHCO_3$ (120 mL) to quench the excess acid and then extracted with CH_2Cl_2 (3 × 100 mL). The organic fractions were combined and concentrated under reduced pressure and the remaining benzaldehyde was removed by co-evaporation with toluene–H_2O (10:1, 4 × 100 mL). The crude residue was purified by column chromatography (49:1 CH_2Cl_2–CH_3OH) to give **4.26** (3.1 g, 84%) as a colourless oil in a 7:6 ratio of endo–exo benzylidene isomers, and 10:1 ratio of anomers. R_f 0.29, 0.26 (20:1 CH_2Cl_2–CH_3OH); $[\alpha]_D$ –44.4 (*c* 0.5, $CHCl_3$; ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.56–7.37 (m, 10H, Ar), 5.99 (s, 1H, PhCH-A), 5.80 (s, 0.85H, PhCH-B), 5.13 (s, 0.85H, H-1B), 5.10 (s, 1H, H-1A), 4.98 (dd, J = 5.6, 3.7 Hz, 1H, H-3A), 4.94 (dd, J = 6.2, 3.7 Hz, 0.85H, H-3B),

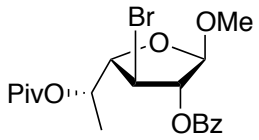
4.76 (d, *J* = 5.6 Hz, 1H, H-2A), 4.69 (d, *J* = 6.3 Hz, 0.85H, H-2B), 4.22–4.12 (m, 2H, H-5A/H-5B), 3.86 (ddd, *J* = 7.5, 3.7, 0.4 Hz, 0.85H, H-4B), 3.80 (ddd, *J* = 7.9, 3.7, 0.5 Hz, 1H, H-4A), 3.39 (s, 2.5H, OC*H*$_3$-B), 3.38 (s, 3H, OC*H*$_3$-A), 2.45 (d, *J* = 5.2 Hz, 1H, 5-O*H*-A), 2.42 (d, *J* = 5.8 Hz, 0.85H, 5-O*H*-B), 1.43 (d, *J* = 6.3 Hz, 3H, H-6A), 1.39 (d, *J* = 6.4 Hz, 2.5H, H6-B); ^{13}C NMR (126 MHz, $CDCl_3$) δ 136.3 (Ar-A), 135.7 (Ar-B), 130.0 (Ar-B), 129.8 (Ar-A), 128.5 (2 × Ar-A/2 × Ar-B), 127.0 (2 × Ar-B), 126.6 (2 × Ar-A), 107.2 (C-1A), 106.7 (C-1B), 106.3 (Ph*C*H-B), 105.8 (Ph*C*H-A), 85.5 (C-4B), 84.7 (C-4A), 83.9 (C-2A), 83.3 (C-2B), 80.8 (C-3B), 80.0 (C-3A), 66.5 (C-5B), 66.3 (C-5A), 54.6 (O*C*H$_3$-B), 54.5 (O*C*H$_3$-A), 20.5 (C-6B), 20.4 (C-6A); HRMS (ESI) *m/z* Calc. for (M+H) $C_{14}H_{19}O_5$: 267.1227. Found: 267.1225.



Methyl 2,3-*O*-benzylidene-5-*O*-pivaloyl-α-L-rhamnofuranoside (4.27):

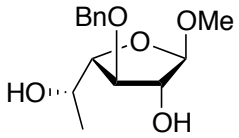
To a solution of **4.26** (1.88 g, 7.0 mmol) in pyridine (30 mL) was added pivaloyl chloride (2.22 mL, 18 mmol). The reaction was stirred for 16 h. The solvent was removed by evaporation under reduced pressure and the reaction mixture was resuspended in minimal EtOAc leaving the pyridine salts as an insoluble white solid, which were filtered. The filtrate was concentrated and residual solvent was removed by co-evaporation with toluene (3 × 30 mL) to give **4.27** (2.32 g, 94%) as a white solid in a 7:6 ratio of endo–exo benzylidene

isomers and 10:1 ratio of anomers. R_f 0.38, 0.29 (9:1 Hexanes–EtOAc); $[\alpha]_D$ – 15.2 (*c* 0.7, $CHCl_3$); ^1H NMR (500 MHz, $CDCl_3$) δ 7.50–7.47 (m, 1.85H, Ar), 7.45–7.42 (m, 1.85H, Ar), 7.41–7.36 (m, 5.5H, Ar), 5.95 (s, 1H, PhC*H*-A), 5.78 (s, 1H, PhC*H*-B), 5.22 (m, 1.85H, H-5A/H-5B), 5.09 (s, 1H, H-1A), 5.05 (s, 0.85H, H-1B), 4.82 (dd, *J* = 5.6, 3.6 Hz, 1H, H-3A), 4.79 (dd, *J* = 6.1, 3.5 Hz, 0.85H, H-3B), 4.70 (d, *J* = 5.6 Hz, 1H, H-2A), 4.65 (d, *J* = 6.2 Hz, 0.85H, H-2B), 4.02 (dd, *J* = 8.2, 3.5 Hz, 0.85H, H-4B), 3.99 (dd, *J* = 8.5, 3.6 Hz, 1H, H-4A), 3.38 (s, 3H, OC*H*$_3$-A), 3.38 (s, 2.5H, OC*H*$_3$-B), 1.43 (d, *J* = 6.2 Hz, 3H, H6-A), 1.41 (d, *J* = 6.2 Hz, 2.5H, H-6B), 1.22 (s, 9H, C(C*H*$_3$)$_3$-A), 1.20 (s, 7.6H, C(C*H*$_3$)$_3$-B); ^{13}C NMR (126 MHz, $CDCl_3$) δ 177.1 (*C*=O-A), 177.0(*C*=O-B), 136.8 (Ar-A), 136.0 (Ar-B), 129.6 (Ar-B), 129.3 (Ar-A), 128.4 (2 × Ar-A/2 × Ar-B) 126.9 (2 × Ar-B), 126.3 (2 × Ar-A), 107.1 (C-1A), 106.8 (C-1B), 106.1 (Ph*C*H-B), 105.3 (Ph*C*H-A), 85.3 (C-2B), 84.3 (C-2A), 82.0 (C-4A), 81.6 (C-4B), 80.3 (C-3B), 79.5 (C-3A), 68.1 (C-5B), 67.9 (C-5A), 54.6 (O*C*H$_3$-B), 54.5 (O*C*H$_3$-A), 38.7 (*C*(CH$_3$)$_3$-A), 38.6 (*C*(CH$_3$)$_3$-B), 27.1 (C(*C*H$_3$)$_3$-B), 27.0 (C(*C*H$_3$)$_3$-A), 17.7 (H-6B), 17.6 (H-6A); HRMS (ESI) *m/z* Calc. for (M+H) $C_{19}H_{27}O_6$: 351.1802. Found: 351.1799.



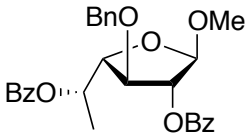
Methyl 2-*O*-benzoyl-3-bromo-3,6-dideoxy-5-*O*-pivaloyl-α-L-altrofuranoside (4.28):

To a solution of **4.27** (200 mg, 0.57 mmol) in CCl_4 (8.0 mL) was added *N*-bromosuccinimide (142 mg, 0.80 mmol), and barium carbonate (135 mg, 0.68 mmol). The reaction was heated to reflux for 3 h, during which time a solid precipitate formed. The precipitate was removed by filtration through Celite, and the filtrate was concentrated. The crude compound was then dissolved in CH_2Cl_2 (1 × 30 mL) and washed with H_2O (1 × 25 mL) and brine (1 × 25 mL), dried over Na_2SO_4, filtered and the filtrate evaporated to give a brown oil, which was purified by column chromatography (14:1 Hexanes–EtOAc) to yield **4.28** (235 mg, 96%) as a colorless oil. R_f 0.33 (9:1 Hexanes–EtOAc); $[\alpha]_D$ –6.2 (*c* 0.6, $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.05 (app dd, *J* = 8.4, 1.3 Hz, 2H, Ar), 7.62 (app t, *J* = 7.5 Hz, 1H, Ar), 7.48 (app dd, *J* = 8.1, 7.5 Hz, 2H, Ar), 5.52 (d, *J* = 2.1 Hz, 1H, H-2), 5.21 (dq, *J* = 6.5, 5.0 Hz, 1H, H-5), 5.10 (s, 1H, H-1), 4.42 (dd, *J* = 6.6, 4.9 Hz, 1H, H-4), 4.11 (ddd, *J* = 6.6, 2.1, 0.7 Hz, 1H, H-3), 3.47 (s, 3H, OC*H*$_3$), 1.40 (d, *J* = 6.5 Hz, 3H, H-6), 1.20 (s, 9H, C(C*H*$_3$)$_3$); ^{13}C NMR (126 MHz, $CDCl_3$) δ 177.6 (*C*(=O)C(CH$_3$)$_3$), 165.4 (*C*(=O)Ph, 133.6 (Ar), 129.9 (2 × Ar), 128.9 (Ar), 128.5 (2 × Ar), 106.9 (C-1), 87.2 (C-4), 86.0 (C-2), 69.2 (C-5), 55.0 (O*C*H$_3$), 45.0 (C-3), 38.9 (*C*(CH$_3$)$_3$), 27.1 (C(*C*H$_3$)$_3$), 16.7 (C-6); HRMS (ESI) *m/z* Calc. for (M+Na) $C_{19}H_{25}BrNaO_6$: 451.0727. Found: 451.0702.



Methyl 3-*O*-benzyl-6-deoxy-α-L-altrofuranoside (4.29):

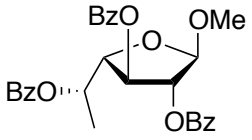
To a solution of the epoxide **4.25** (338 mg, 1.38 mmol) in BnOH (13 mL) at 0 °C was added a suspension of 60% NaH in mineral oil (220 mg, 5.5 mmol). After stirring the reaction at room temperature for 30 min a new spot on the TLC was visualized (R_f 0.18, 2:1 Hexanes–EtOAc) corresponding to the pivaloyl-deprotected intermediate. The nucleophilic epoxide opening reaction required further heating of the solution at reflux. After 24 h, the solution was cooled to room temperature and neutralized through the addition of acetic acid. The solution was diluted with EtOAc (50 mL) and washed with H_2O (1 × 50 mL) and brine (1 × 50 mL). The solution was then dried over Na_2SO_4, filtered and concentrated to give a crude oil that was purified by column chromatography (9:1 → 4:1 →1:1 hexanes–EtOAc) to yield **4.29** (232 mg, 62%) as a colourless oil. R_f 0.09 (2:1 hexanes–EtOAc); $[\alpha]_D$ –107.2 (*c* 0.4, $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41–7.30 (m, 5H, Ar), 4.91 (s, 1H, H-1), 4.66 (AB q, *J* = 12.3 Hz, 2H, OC*H*$_2$Ph), 4.18 (s, 1H, H-2), 4.11–4.04 (m, 2H, H-4/H-5), 3.87 (d, *J* = 3.3 Hz, 1H, H-3), 3.43 (s, 3H, OC*H*$_3$), 1.11 (d, *J* = 6.5 Hz, 3H, H-6); ^{13}C NMR (126 MHz, $CDCl_3$) δ 137.6 (Ar), 128.5 (2 × Ar), 128.2 (2 × Ar), 128.0 (Ar), 109.8 (C-1), 87.6 (C-4), 82.9 (C-3), 78.0 (C-2), 72.0 (O*C*H$_2$Ph), 67.0 (C-5), 55.1 (O*C*H$_3$), 18.6 (C-6); HRMS (ESI) *m/z* Calc. for (M+Na) $C_{14}H_{20}O_5Na$: 291.1203. Found: 291.1204.



Methyl 2,5-di-*O*-benzoyl-3-*O*-benzyl-6-deoxy-α-L-altrofuranoside (4.30):

To a solution of **4.29** (530 mg, 1.97 mmol) in pyridine (15 mL) was added benzoyl chloride (0.91 mL, 7.88 mmol). The reaction mixture was stirred at room temperature for 16 h and then CH_3OH (5 mL) was added. After 24 h, the solution was cooled to room temperature and was neutralized through the addition of acetic acid. The solvent was removed by evaporation under reduced pressure. The resulting residue was dissolved in CH_2Cl_2 (20 mL) and washed with 1 M HCl (1 × 20 mL), H_2O (1 × 20 mL), a saturated solution of sodium bicarbonate (1 × 20 mL), and brine (1 × 20 mL). The organic layer was dried over Na_2SO_4, filtered and the filtrate concentrated to give a crude oil, which was purified by column chromatography (9:1 hexanes–EtOAc) to give **4.30** (863 mg, 92%) as a colorless oil. R_f 0.24 (9:1 hexanes–EtOAc); $[\alpha]_D$ –31.7 (*c* 0.8, $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.00–7.91 (m, 4H, Ar), 7.62–7.56 (m, 1H, Ar), 7.53–7.47 (m, 1H, Ar), 7.42–7.33 (m, 4H, Ar), 7.31–7.24 (m, 5H, Ar), 5.50 (dq, *J* = 6.6, 4.1 Hz, 1H, H-5), 5.40 (s, 1H, H-2), 5.09 (s, 1H, H-1), 4.80 (AB q, *J* = 11.8 Hz, 1H, OC*H*$_2$Ph), 4.30–4.22 (m, 2H, H-3/H-4), 3.47 (s, 3H, OC*H*$_3$), 1.38 (d, *J* = 6.6 Hz, 3H, H-6); ^{13}C NMR (126 MHz, $CDCl_3$) δ 165.7 (C=O), 165.6 (C=O), 137.2 (Ar), 133.4 (Ar), 132.9 (Ar), 130.2 (Ar), 129.8 (2 × Ar), 129.6 (2 × Ar), 129.3 (Ar), 128.5 (2 × Ar), 128.4 (4 × Ar), 128.2 (2 × Ar), 127.9 (Ar), 107.0 (C-1), 84.7 (C-4), 83.3 (C-3), 81.6 (C-2), 72.5 (O*C*H$_2$Ph), 69.8

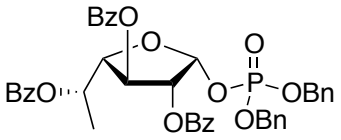
(C-5), 54.9 (O*C*H_3), 16.6 (C-6); HRMS (ESI) *m/z* Calc. for (M+Na) $C_{28}H_{28}O_7Na$: 499.1727. Found: 499.1727.



Methyl 2,3,5-tri-*O*-benzoyl-6-deoxy-α-L-altrofuranoside (4.31):

The methyl glycoside **4.30** (860 mg, 1.80 mmol) and $Pd(OH)_2$ (86 mg, 10% w/w) were dissolved in a mixture of EtOAc–CH_3OH–AcOH 3:1:1 (18 mL) and the reaction mixture was stirred at room temperature under an atmosphere of H_2 gas. After 16 h, TLC showed complete consumption of the starting material. The reaction mixture was filtered through Celite to remove the catalyst and the filtrate was concentrated to give a white solid. Without further purification, the solid was dissolved in pyridine (13 mL), and to this was added benzoyl chloride (0.33 mL, 2.8 mmol). The mixture was stirred for 16 h, at which point CH_3OH (5 mL) was added. The solvent was evaporated, and the resulting residue was resuspended in CH_2Cl_2 (20 mL) and washed with 1 M HCl (1 × 25 mL), H_2O (1 × 25 mL), a saturated solution of sodium bicarbonate (1 × 25 mL), and brine (1 × 25 mL). The organic layer was dried over Na_2SO_4 and concentrated. The crude residue was purified by column chromatography (6:1 hexanes–EtOAc) to afford **4.31** (850 mg, 96%) as a colorless oil. R_f 0.18 (9:1 hexanes–EtOAc); $[\alpha]_D$ 21.5 (*c* 0.5, $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.12–7.97 (m, 6H, Ar), 7.65–7.56 (m, 2H, Ar), 7.55–7.50 (m, 1H, Ar), 7.50–7.45 (m, 2H, Ar), 7.43–7.38 (m, 2H, Ar), 7.35–7.30 (m, 2H, Ar), 5.78 (ddd, *J* = 5.6, 1.4, 0.7 Hz, 1H, H-3), 5.63 (dq, *J*

= 6.5, 4.3 Hz, 1H, H-5), 5.43 (d, *J* = 1.4 Hz, 1H, H-2), 5.13 (s, 1H, H-1), 4.50 (dd, *J* = 5.6, 4.3 Hz, 1H, H-4), 3.52 (s, 3H, OC*H*$_3$), 1.51 (d, *J* = 6.6 Hz, 3H, H-6); ^{13}C NMR (126 MHz, $CDCl_3$) δ 165.7 (C=O), 165.5 (C=O), 165.4 (C=O), 133.5 (Ar), 133.4 (Ar), 133.0 (Ar), 130.1 (Ar), 129.9 (2 × Ar), 129.9 (2 × Ar), 129.7 (2 × Ar), 129.3 (Ar), 129.2 (Ar), 128.5 (4 × Ar), 128.3 (2 × Ar), 106.9 (C-1), 83.6 (C-4), 82.6 (C-2), 77.2 (C-3), 69.7 (C-5), 55.0 (O*C*H$_3$), 16.3 (C-6); HRMS (ESI) *m/z* Calc. for (M+Na) $C_{28}H_{26}O_8Na$: 513.1520. Found: 513.1516.



Dibenzyl 2,3,5-tri-*O*-benzoyl-6-deoxy-α-L-altrofuranosyl phosphate (4.32):

To a solution of methyl glycoside **4.31** (450 mg, 0.92 mmol) in CH_2Cl_2 (11 mL) at 0 °C was added 33% HBr in AcOH (0.6 mL) while keeping the temperature below 0 °C. After 3 h, the reaction mixture was diluted with toluene and concentrated. Without purification, the residue was resuspended in toluene (11 mL) and then dibenzylphosphate (350 mg, 1.26 mmol) and Et_3N (0.9 mL) were added. The reaction mixture was stirred overnight and then filtered to remove Et_3NHBr salts. The filtrate was concentrated to give a crude oil that was purified by column chromatography (4:1 hexanes–EtOAc) to give **4.32** (256 mg, 38%) as a white solid. R_f 0.24 (2:1 hexanes–EtOAc); $[\alpha]_D$ +113.1 (*c* 0.5, $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.08–7.93 (m, 6H, Ar), 7.63–7.52 (m, 2H, Ar), 7.48–7.42 (m, 3H, Ar), 7.40–7.30 (m, 5H, Ar), 7.30–7.21 (m, 7H, Ar), 7.14–7.08 (m, 2H, Ar), 6.32 (dd, *J* = 5.5, 4.4 Hz, 1H, H-1), 6.22 (dd, *J* = 7.3, 6.3

Hz, 1H, H-3), 5.66 (ddd, *J* = 7.4, 4.4, 2.0 Hz, 1H, H-2), 5.48 (app p, *J* = 6.3 Hz, 1H, H-5), 5.00 (ddd, *J* = 20.0, 11.7, 7.8 Hz, 2H, C*H*$_2$Ph), 4.84 (ddd, *J* = 19.3, 11.7, 7.2 Hz, 2H, C*H*$_2$Ph), 4.44 (dd, *J* = 6.7, 6.3, 1H, H-4), 1.45 (d, *J* = 6.4 Hz, 3H, H-6); ^{13}C NMR (126 MHz, $CDCl_3$) δ 165.6 (C=O), 165.4 (C=O), 165.1 (C=O), 135.5 (d, *J* = 6.0 Hz, Ar), 135.4 (d, *J* = 6.2 Hz, Ar), 133.6 (Ar), 133.5 (Ar), 132.9 (Ar), 130.1 (2 × Ar), 129.9 (Ar), 129.9 (2 × Ar), 129.7 (2 × Ar), 129.0 (Ar), 128.7 (Ar), 128.5 (2 × Ar), 128.5 (2 × Ar), 128.5 (2 × Ar), 128.4 (3 × Ar), 128.3 (Ar), 128.2 (2 × Ar), 127.8 (2 × Ar), 127.6 (2 × Ar), 97.9 (d, *J* = 4.9 Hz, C-1) 82.7 (C-4), 76.9 (d, *J* = 7.3 Hz, C-2), 74.5 (C-3), 71.3 (C-5), 69.3 (d, *J* = 5.4 Hz, *C*H_2Ph), 69.2 (d, *J* = 5.4 Hz, *C*H_2Ph), 16.5 (C-6); ^{31}P{^{1}H} NMR (162 MHz, $CDCl_3$) δ –2.73; HRMS (ESI) *m/z* Calc. for (M+Na) $C_{41}H_{37}O_{11}PNa$: 759.1966. Found: 759.1961.

4.5 Bibliography

1. Nassau, P. M.; Martin, S. L.; Brown, R. E.; Weston, A.; Monsey, D.; McNeil, M. R.; Duncan, K., *J. Bacteriol.* **1996,** *178*, 1047-1052.
2. Richards, M. R.; Lowary, T. L., *ChemBioChem* **2009,** *10*, 1920-1938.
3. Poulin, M. B.; Nothaft, H.; Hug, I.; Feldman, M. F.; Szymanski, C. M.; Lowary, T. L., *J. Biol. Chem.* **2010,** *285*, 493-501.
4. Wang, Q.; Ding, P.; Perepelov, A. V.; Xu, Y. L.; Wang, Y.; Knirel, Y. A.; Wang, L.; Feng, L., *Mol. Microbiol.* **2008,** *70*, 1358-1367.
5. Allos, B. M., *Clin. Infect. Dis.* **2001,** *32*, 1201-1206.

6. Guerry, P.; Szymanski, C. M., *Trends Microbiol.* **2008,** *16*, 428-435.

7. Hanniffy, O. M.; Shashkov, A. S.; Moran, A. P.; Prendergast, M. M.; Senchenkova, S. N.; Knirel, Y. A.; Savage, A. V., *Carbohydr. Res.* **1999,** *319*, 124-132.

8. St Michael, F.; Szymanski, C. M.; Li, J. J.; Chan, K. H.; Khieu, N. H.; Larocque, S.; Wakarchuk, W. W.; Brisson, J. R.; Monteiro, M. A., *Eur. J. Biochem.* **2002,** *269*, 5119-5136.

9. Prendergast, M. M.; Lastovica, A. J.; Moran, A. P., *Infect. Immun.* **1998,** *66*, 3649-3655.

10. Andrey V. Karlyshev; Olivia L. Champion; Carol Churcher; Jean-Robert Brisson; Harold C. Jarrell; Michel Gilbert; Denis Brochu; Frank St Michael; Jianjun Li; Warren W. Wakarchuk; Ian Goodhead; Mandy Sanders; Kim Stevens; Brian White; Julian Parkhill; Brendan W. Wren; Christine M. Szymanski, *Mol. Microbiol.* **2005,** *55*, 90-103.

11. Hoffman, J.; Lindberg, B.; Lonngren, J.; Hofstad, T., *Carbohydr. Res.* **1976,** *47*, 261-267.

12. Kondo, W.; Sato, N., *Carbohydr. Res.* **1979,** *70*, 117-123.

13. Hoffman, J.; Lindberg, B.; Skaug, N.; Hofstad, T., *Carbohydr. Res.* **1980,** *84*, 181-183.

14. Kondo, W.; Nakazawa, F.; Sato, M.; Ito, T., *Carbohydr. Res.* **1981,** *97*, 279-283.

15. Sato, N.; Nakazawa, F.; Sato, M.; Hoshino, E.; Ito, T., *Carbohydr. Res.* **1993,** *245*, 105-111.

16. Downes, J.; Munson, M. A.; Spratt, D. A.; Kononen, E.; Tarkka, E.; Jousimies-Somer, H.; Wade, W. G., *J. Med. Microbiol.* **2001,** *50*, 947-951.

17. Montgomery, E. M.; Hudson, C. S., *J. Am. Chem. Soc.* **1937,** *59*, 992-993.

18. Pakulski, Z.; Zamojski, A., *Tetrahedron* **1997,** *53*, 2653-2666.

19. Ramana, C. V.; Narute, S. B.; Gonnade, R. G.; Patil, R. S., *Synthesis-Stuttgart* **2008**, 1783-1787.

20. Varvogli, A. A. C.; Karagiannis, I. N.; Koumbis, A. E., *Tetrahedron* **2009,** *65*, 1048-1058.

21. Timmons, S. C.; Jakeman, D. L., *Org. Lett.* **2007,** *9*, 1227-1230.

22. Wagner, G. K.; Pesnot, T.; Field, R. A., *Nat. Prod. Rep.* **2009,** *26*, 1172-1194.

23. Zhang, Q. B.; Liu, H. W., *J. Am. Chem. Soc.* **2000,** *122*, 9065-9070.

24. Sanders, D. A. R.; Staines, A. G.; McMahon, S. A.; McNeil, M. R.; Whitfield, C.; Naismith, J. H., *Nat. Struct. Biol.* **2001,** *8*, 858-863.

25. Gilbert, M.; Watson, D. C.; Cunningham, A. M.; Jennings, M. P.; Young, N. M.; Wakarchuk, W. W., *J. Biol. Chem.* **1996,** *271*, 28271-28276.

26. Bernatchez, S.; Szymanski, C. M.; Ishiyama, N.; Li, J. J.; Jarrell, H. C.; Lau, P. C.; Berghuis, A. M.; Young, N. M.; Wakarchuk, W. W., *J. Biol. Chem.* **2005,** *280*, 4792-4802.

27. Errey, J. C.; Mann, M. C.; Fairhurst, S. A.; Hill, L.; McNeil, M. R.; Naismith, J. H.; Percy, J. M.; Whitfield, C.; Field, R. A., *Org. Biomol. Chem.* **2009,** *7*, 1009-1016.

28. Partha, S. K.; van Straaten, K. E.; Sanders, D. A. R., *J. Mol. Biol.* **2009,** *394*, 864-877.

29. Shashkov, A. S.; Pakulski, Z.; Grzeszczyk, B.; Zamojski, A., *Carbohydr. Res.* **2001,** *330*, 289-294.

30. Poly, F.; Serichatalergs, O.; Schulman, M.; Ju, J.; Cates, C. N.; Kanipes, M.; Mason, C.; Guerry, P., *J. Clin. Microbiol.* **2011,** *49*, 1750-1757.

31. Guerry, P.; Monteiro, M. A. U.S. Patent 13/117,215, 2011.

32. Zhang, Q. B.; Liu, H. W., *Bioorg. Med. Chem. Lett.* **2001,** *11*, 145-149.

33. Pauly, M.; Porchia, A.; Olsen, C. E.; Nunan, K. J.; Scheller, H. V., *Anal. Biochem.* **2000,** *278*, 69-73.

34. Konishi, T.; Takeda, T.; Miyazaki, Y.; Ohnishi-Kameyama, M.; Hayashi, T.; O'Neill, M. A.; Ishii, T., *Glycobiology* **2007,** *17*, 345-354.

35. Konishi, T.; Ohnishi-Kameyama, M.; Funane, K.; Miyazaki, Y.; Konishi, T.; Ishii, T., *Carbohydr. Res.* **2010,** *345*, 787-791.

36. Gruber, T. D.; Westler, W. M.; Kiessling, L. L.; Forest, K. T., *Biochemistry* **2009,** *48*, 9171-9173.

37. Cunneen, M. M.; De Castro, C.; Kenyon, J.; Parrilli, M.; Reeves, P. R.; Molinaro, A.; Holst, O.; Skurnik, M., *Carbohydr. Res.* **2009,** *344*, 1533-1540.

38. Gorshkova, R. P.; Kalmykova, E. N.; Isakov, V. V.; Ovodov, Y. S., *Eur. J. Biochem.* **1985,** *150*, 527-531.

39. Timmons, S. C.; Hui, J. P. M.; Pearson, J. L.; Peltier, P.; Daniellou, R.; Nugier-Chauvin, C.; Soo, E. C.; Syvitski, R. T.; Ferrieres, V.; Jakeman, D. L., *Org. Lett.* **2008,** *10*, 161-163.

40. Florent, J. C.; Monneret, C., *Carbohydr. Res.* **1980,** *85*, 243-257.

41. Konishi, T.; Ishii, T., *Trends Glycosci. Glycotechnol.* **2012,** *24*, 13-23.

42. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G., *Bioinformatics* **2007,** *23*, 2947-2948.

43. Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T., *Bioinformatics* **2006,** *22*, 195-201.

44. Peltier, P.; Beláňová, M.; Dianišková, Petronela; Zhou, R.; Zheng, R. B.; Pearcey, J. A.; Joe, M.; Brennan, P. J.; Nugier-Chauvin, C.; Ferrières, V.; Lowary, T. L.; Daniellou, R.; Mikušová, K., *Chem. Biol.* **2010,** *17*, 1356-1366.

45. Poulin, M. B.; Lowary, T. L., *Methods Enzymol.* **2010,** *478*, 389-411.

46. Timmons, S. C.; Mosher, R. H.; Knowles, S. A.; Jakeman, D. L., *Org. Lett.* **2007,** *9*, 857-860.

47. Rose, N. L.; Zheng, R. B.; Pearcey, J.; Zhou, R.; Completo, G. C.; Lowary, T. L., *Carbohydr. Res.* **2008,** *343*, 2130-2139.

Chapter 5

Synthetic UDP-galactofuranose analogs reveal critical enzyme–substrate interactions in GlfT2-catalyzed mycobacterial galactan assembly

---

A portion of this chapter has been published:

Poulin, M. B.; Zhou, R.; Lowary, T. L., *Org. Biomol. Chem.* **2012**, *10*, 4074-4087.

Site directed mutagenesis of GlfT2 and evaluation of the GlfT2 mutant activity and kinetics was performed by Ruixiang Blake Zheng.

5.1. Introduction

As noted in previous chapters, mammalian glycoconjugates contain galactose residues exclusively in the thermodynamically favored six-membered pyranose ring form (Gal*p*). However, galactose in the five-membered furanose ring form (Gal*f*) is found in many microorganisms.[1, 2] Among these organisms are mycobacteria, members of which continue to have a significant impact on world health. There is approximately one third of the world population believed to have been infected with latent *Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB), which results in nearly three million deaths annually.[1-3] Concern over TB has increased due to the emergence of multi-drug resistant and extensively-drug resistant strains of the organism.[4] The need for new anti-mycobacterial therapeutics, and the absence of Gal*f* in mammalian tissues, has lead to an interest in Gal*f* metabolism as a potential target for drug action.[5]

In mycobacteria, Gal*f* residues are found in the complex and glycan-rich cell wall, specifically, the mycolyl–arabinogalactan (mAG) complex, a lipidated polysaccharide composed almost entirely of furanose carbohydrates (Figure 5-1).[6, 7] This glycan is the largest structural component of the mycobacterial cell wall and is covalently attached to cell wall peptidoglycan through an α-L-Rha*p*-(1→3)-α-D-Glc*p*NAc-phosphate disaccharide.[7] The core of the mAG is a galactan composed of 30–35 D-Gal*f* residues connected through alternating β-(1→5) and β-(1→6) glycosidic linkages. Three arabinan domains composed of D-arabinofuranosyl (Ara*f*) residues are attached at the C-5 hydroxyl group of the

eighth, tenth and twelfth Gal*f* residue of the galactan. These arabinan domains are further esterified with mycolic acids, large C_{70}–C_{90} branched lipids that impart significant hydrophobicity to the mycobacterial cell wall.

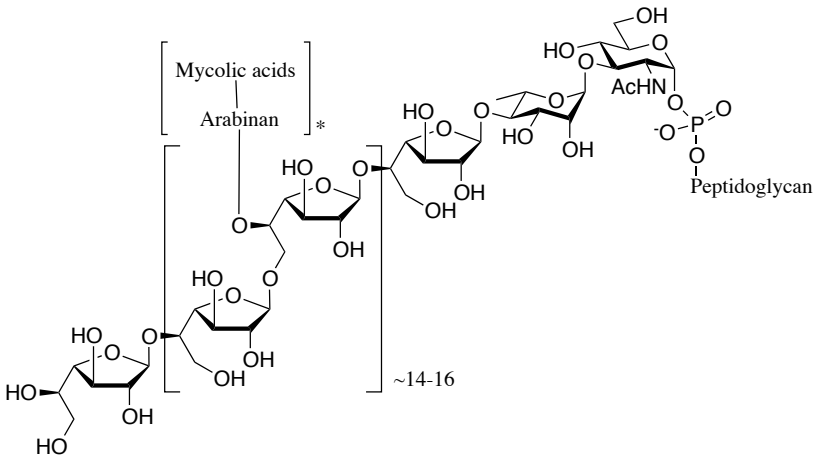


Figure 5-1. Structure of the mycobacterial mAG complex. *Three arabinan domains are attached to the galactan on the 8th, 10th and 12th Gal*f* residue.

Many of the pathophysiological features of mycobacterial infections are attributed to the hydrophobic cell wall.[8] For example, this structure contributes to difficulties in treating mycobacterial infections, by acting as a permeability barrier to antibiotics.[9] As a result, the enzymes involved in mAG biosynthesis are attractive targets for new anti-mycobacterial theraputics.[10] Two of the standard drugs currently used to treat TB, isoniazid and ethambutol, target the biosynthesis of the mycolic acid and arabinan components of the mAG, respectively.[11, 12] However, no currently used TB drugs are known to target the assembly of the mAG galactan.

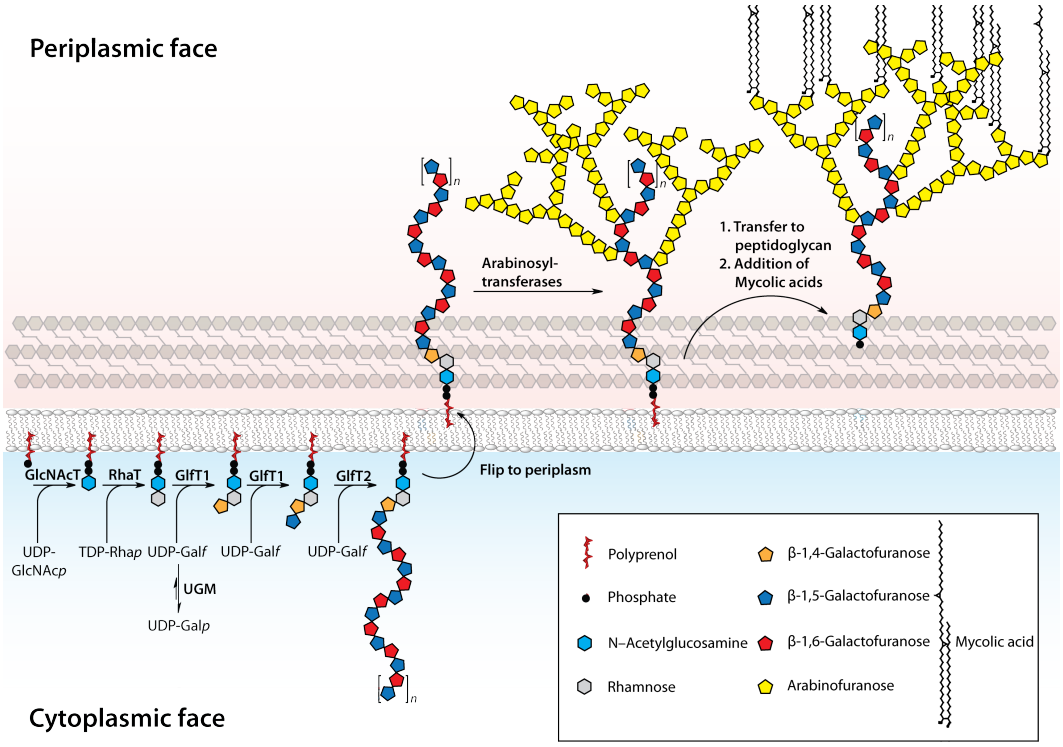
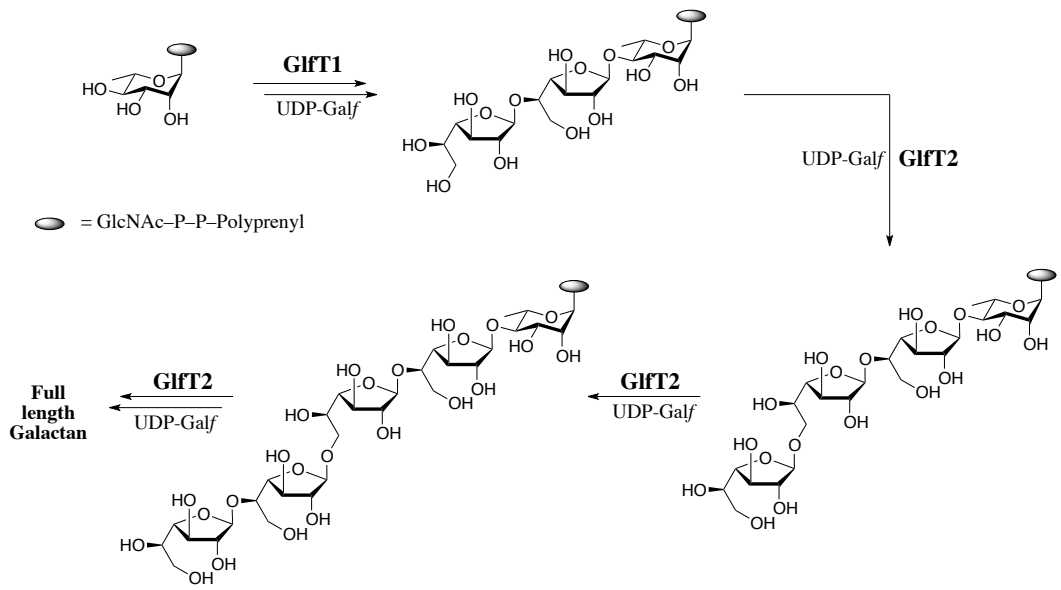


Figure 5-2. Proposed biosynthetic pathway for the biosynthesis of mAG complex showing the sub-cellular localization of each step.

The biosynthesis of mAG galactan involves two membrane associated bifunctional galactofuranosyl-transferase enzymes, GlfT1 and GlfT2, and is localized on the cytoplasmic face of the cell membrane. (Figure 5-2).[13] As the source of Gal*f*, both enzymes use the sugar nucleotide UDP-Gal*f* (**1**), which is biosynthesized from UDP-galactopyranose (Gal*p*) by the action of UDP-galactopyranose mutase (UGM).[14] GlfT1 is responsible for adding the first and second Gal*f* residues to an α-L-Rha*p*-(1→3)-α-D-Glc*p*NAc-decaprenyl phosphate acceptor.[15] The final product of the GlfT1-catalyzed reaction is the initial GlfT2 acceptor substrate.



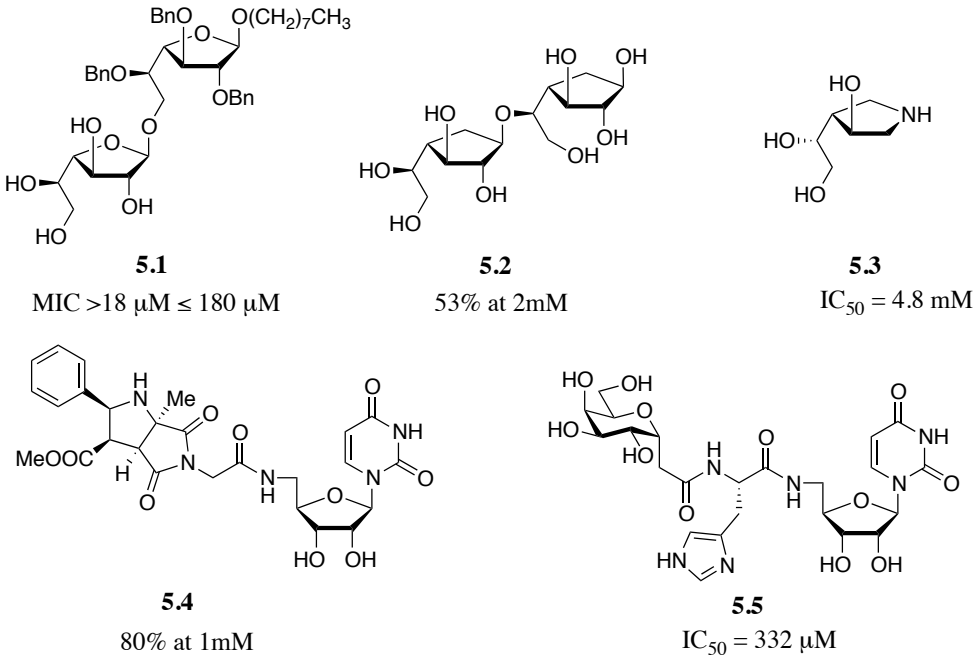
Scheme 5-1. Biosynthesis of the mAG core galactan. Alternating β-(1→6)-Gal*f* and β-(1→5)-Gal*f* transferase activity of GlfT2 are catalyzed using a single enzyme active-site.

The gene encoding GlfT2 was first identified in 2000,[16] and recombinant GlfT2 has since been expressed and purified.[15, 17] Subsequent studies on the enzyme have demonstrated that GlfT2 is bifunctional, and synthesizes both β-D-Gal*f*-(1→5)-D-Gal*f* and β-D-Gal*f*-(1→6)-D-Gal*f* linkages (Scheme 5-1). Using synthetic galactan fragments (e.g., **5.6** and **5.7**, Figure 5-4) we, and others, have demonstrated that GlfT2 requires a β-(1→5)- or β-(1→6)-linked Gal*f* disaccharide as the minimum acceptor substrate.[17, 18] Elegant saturation transfer difference NMR (STD-NMR) experiments demonstrated that trisaccharides **5.6** and **5.7**, which serve as substrate analogs for the β-(1→6) and β-(1→5) transferase activity of GlfT2 respectively, bind competitively to the same site of GlfT2.[19] This lead to a hypothesis that both the β-(1→6) and β-(1→5) transferase activity of GlfT2 arise from a single enzyme active–site. This

hypothesis was later supported through site-directed mutagenesis of the proposed active-site catalytic residues, which showed that single residue mutations affected both acctivities.[20] More recent crystallographic investigations have confirmed the presence of only one catalytic site.[21] Pulse chase experiments using isotopically labeled acceptor substrates demonstrated that GlfT2 acts as a processive polymerase[22] adding the third and subsequent Gal*f* residues to a growing galactan chain without release of the acceptor substrate after each subsequent glycosyl transfer reaction. In addition, it was shown that GlfT2 is capable of forming full length galactan polymers (~30–35 Gal*f* residues) without the need of additional enzymes, and it has been suggested that galactan length is controlled by tethering of the polyprenol aglycone to the protein.[23] However, despite these investigations, there is still relatively little known about the specific protein–carbohydrate interactions that GlfT2 uses to recognize its substrates.

The progress in establishing the mechanism of GlfT2 has not been matched by the identification of potent and selective inhibitors of this enzyme. Benzylated disaccharide acceptor analogs (e.g., **5.1**) showed toxicity at micromolar concentrations to live mycobacteria; however, this effect appeared to be due to a nonspecific surfactant effect, as these compounds failed to inhibit GlfT2 activity in a cell free assay.[20] More recent studies on carbasugar disaccharide acceptor analogs (e.g., **5.2**) showed greater then 50% inhibition of the enzyme, but only at millimolar concentrations.[24] Amino sugar analogs of Gal*f* have also been designed as mimics of the GlfT2 transition state (**5.3**), but again they afforded only weak (IC_{50} ~4.8 mM) inhibition activity.[25, 26] It is

interesting to note that the hydroxyl groups of amino sugar analog **5.3** are in the L-altrofuranose (Alt*f*) configuration and, surprisingly, the corresponding D-Gal*f* configured analog showed no inhibitory activity. A later "transition state mimic" **5.4**, incorporating UDP, gave 80% inhibition at 1 mM concentration,[27] but no further work was done to establish its specific inhibitory activity. The most potent GlfT2 inhibitor reported to date is a sugar–amino acid–nucleoside analog **5.5**, with an IC_{50} value of 332 μM.[28] However, the galactose in this analog is in the pyranose ring form and its selectivity for GlfT2, as opposed to other galactosyltrasferases, is unknown.



5.1
MIC >18 μM ≤ 180 μM

5.2
53% at 2mM

5.3
IC_{50} = 4.8 mM

5.4
80% at 1mM

5.5
IC_{50} = 332 μM

Figure 5-3. Examples of compounds tested as inhibitors of GlfT2. The minimum inhibitory concentration (MIC), IC_{50} or percent inhibition values are listed for each compound.

Singly methylated and deoxygenated carbohydrate analogs have shown great utility as biological tools to explore protein–carbohydrate binding interactions, and in some cases have led to the identification of specific glycosyltransferase inhibitors.[29-31] In this chapter, to facilitate the design of more potent and selective inhibitors of GlfT2, we utilized our panel of singly modified UDP-Gal*f* analogs **5.9**–**5.16** (Figure 5-4), whose synthesis was described in Chapter 3, to probe specific protein–carbohydrate interactions involved in substrate recognition and turnover, and to explore their effect on galactan polymerization.

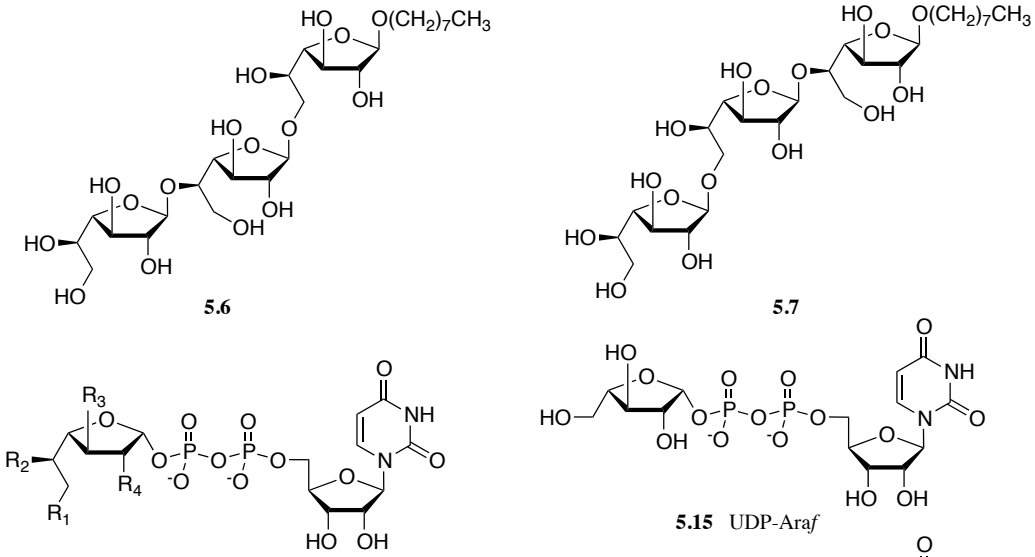


Figure 5-4. Synthetic acceptor trisaccharides **5.6** and **5.7** and synthetic UDP-Gal*f* analogs **5.8**–**5.16** used to evaluate substrate-binding interactions for GlfT2.

5.2. Results and Discussion

5.2.1. GlfT2 activity and specificity with synthetic donor analogs

Previous work from our group described a continuous spectrophotometric assay developed to monitor the activity of GlfT2.[32] This assay monitors the formation of UDP liberated from **1** upon GlfT2-mediated transfer of Gal*f* to an acceptor substrate. UDP formation is coupled to the oxidation of NADH via pyruvate kinase and lactate dehydrogenase, resulting in a decrease in NADH absorbance at 340 nm that is directly proportional to the GlfT2 activity (Figure 5-5). As the assay is not specific for UDP,[33] it can also be employed to monitor GlfT2 activity with TDP-Gal*f* through detection of liberated TDP. In this chapter, we employed this assay to probe the donor binding site specificity of GlfT2 by measuring the specific transferase activity with **5.8**–**5.16**. In these assays, the trisaccharide β-Gal*f*-(1→5)-β-Gal*f*-(1→6)-β-Gal*f*-octyl (**5.6**) served as the primary acceptor substrate.[34] Previous studies reported that **5.6** acts as the preferred GlfT2 acceptor substrate when compared to the isomeric β-Gal*f*-(1→6)-β-Gal*f*-(1→5)-β-Gal*f*-octyl trisaccharide (**5.7**).[17] Donors **5.8**–**5.10** and **5.15** have been previously shown to act as substrates for GlfT2,[35] and were further screened to compare their activity with the other substrates and to determine kinetic parameters. To determine whether GlfT2 donor specificity is influenced by the nature of the acceptor, and glycosidic linkage being formed, the donor analogs were also screened using acceptor **5.7**.

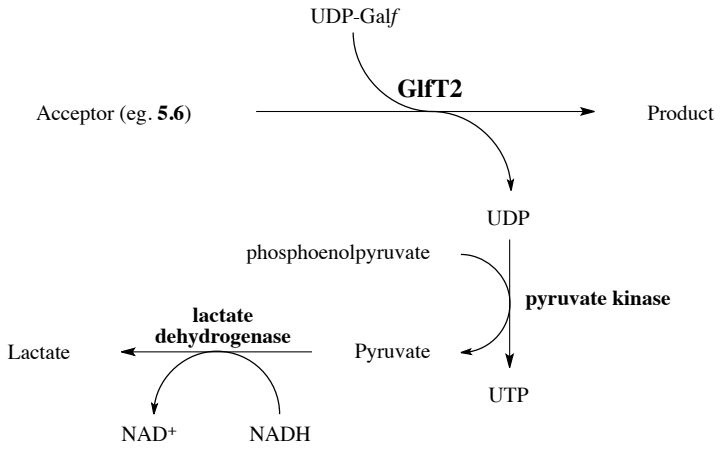


Figure 5-5. Coupled spectrophotometric assay used to measure GlfT2 activity. The decrease in NADH absorbance is directly proportional to GlfT2 activity.

5.2.1.1. Effect of deoxy UDP-Galf derivatives on GlfT2 activity

As presented in Figure 5-6 with acceptor **5.6**, deoxy UDP-Gal*f* donor analogs **5.9**–**5.11** and UDP-Ara*f* (**6**) all served as GlfT2 substrates with varying degrees of efficiency. Of these, the 6-deoxy analog **5.9** had the highest relative activity at greater than 55% compared to the natural donor substrate, **5.8**. The moderate activity observed for **5.9** suggests that any hydrogen bonding interactions between GlfT2 and the UDP-Gal*f* C-6 hydroxyl group are not critical for either substrate recognition or turnover. The much lower activity of UDP-Ara*f* analog **5.15**, which, in contrast to **5.9**, lacks both the C-6 carbon and hydroxyl group, suggests that hydrophobic interactions with C-6 help facilitate, but are not critical for, donor substrate binding and transferase activity. Analog **5.10**, lacking a C-5 hydroxyl group, also demonstrated a decrease in activity similar to **5.15**. The donor analog **5.11**, in which the C-3 hydroxyl group has been removed, displayed less then 1% relative activity, which suggests

interactions at this position are important in either substrate binding or turnover (see additional discussion below). All four of these donor analogs showed sufficient activity to allow for full kinetic characterization.

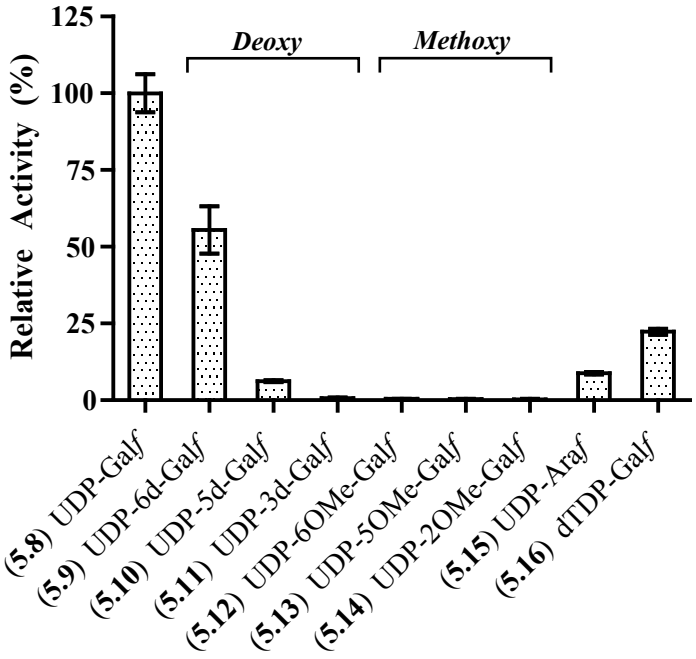


Figure 5-6. Relative activity of GlfT2 with UDP-Gal*f* analogs **5.9**–**5.16** using trisaccharide **5.6** as the acceptor substrate.

5.2.1.2. Kinetic analysis with deoxygenated UDP-Galf analogs.

To better understand the catalytic role of each of the interactions described above, the apparent kinetic constants K_M and k_{cat} were determined for analogs **5.9**–**5.11** and **5.15**, using **5.6** as the acceptor substrate. As presented in Table 5-1, the K_M value for **5.9** was ~4-fold higher compared to the native donor **5.8**, whereas only a ~1.6-fold decrease in k_{cat} was observed, indicating that removal of the hydroxyl group at C-6 has only a moderate effect on substrate

binding without substantially affecting transferase activity. Analogs **5.10** and **5.15** demonstrate a similar ~3.6-fold increase in K_M and additionally resulted in a ~12-fold and ~9-fold decrease in k_{cat}, respectively. These results show that interactions with the C-5 hydroxyl or hydroxymethyl group have only a moderate effect on substrate binding, but play a larger role in facilitating substrate turnover. Presumably, these interactions play an important, although not essential, role in stabilizing the bound substrate in the optimal conformation for turnover. The low specific activity of 3-deoxy analog **5.11** complicated kinetic analysis; however, an ~2-fold increase in K_M was observed, indicating hydrogen bonding to the C-3 hydroxyl group is not key to substrate binding. Conversely, this analog showed a greater than 100-fold decrease in k_{cat}, suggesting hydrogen-bonding interactions involving the C-3 hydroxyl group play a critical role in orienting the substrate for turnover.

Table 5-1. Summary of GlfT2 donor kinetics with analogs **5.8**–**5.11**, **5.15** and **5.16**.[a]

Donor Analog	K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M ($\mu M \cdot min)^{-1}$
5.8[b]	$2.5 (\pm 0.4) \times 10^2$	$1.31 (\pm 0.07) \times 10^2$	$5 (\pm 1) \times 10^{-1}$
5.9[b]	$1.0 (\pm 0.2) \times 10^3$	$7.9 (\pm 0.5) \times 10^1$	$7 (\pm 3) \times 10^{-2}$
5.10[b]	$9 (\pm 1.7) \times 10^2$	10 (± 1)	$1.2 (\pm 0.6) \times 10^{-2}$
5.11[b]	$5.0 (\pm 2.8) \times 10^2$	1.2 (± 0.2)	$2.2 (\pm 0.8) \times 10^{-3}$
5.15[b]	$9 (\pm 1) \times 10^2$	14.7 (± 0.7)	$1.5 (\pm 0.6) \times 10^{-2}$
5.16[b]	$1.9 (\pm 0.3) \times 10^3$	$9.7 (\pm 0.9) \times 10^1$	$2 (\pm 1) \times 10^{-2}$
5.8[c]	$1.0 (\pm 0.3) \times 10^3$	$2 (\pm 1) \times 10^1$	$2 (\pm 3) \times 10^{-2}$

[a] Full kinetics plots can be seen in Appendix D.
[b] Donor kinetics were determined with 2.0 mM **5.6**.
[c] Donor kinetics were determined with 2.0 mM **5.7**.

*5.2.1.3. Effect of O-Methylated UDP-Gal*f *derivatives on GlfT2 activity.*

The three prepared methyl donor analogs **5.12**–**5.14** were also screened for activity with GlfT2, and each demonstrated >230-fold decreases in activity (Figure 5-6). Without results for the 2-deoxy UDP-Gal*f* analog we cannot conclusively demonstrate that the decrease in activity observed for **5.14** is due to additional steric interaction rather than the result of disruption of hydrogen bonding interactions. However, the decreased activity of analogs **5.12** and **5.13**, when compared to the corresponding deoxy analogs **5.9** and **5.10**, imply that GlfT2 cannot tolerate additional steric bulk in the donor-binding site and the results for analog **5.14** are consistent with this hypothesis. The additional steric bulk of these analogs either prevents substrate binding or substantially impairs turnover after the substrate is bound. To distinguish between these two possibilities, analogs **5.12**–**5.14** were examined for their ability to inhibit GlfT2 activity. If the analogs were bound by the enzyme but poorly turned over, we expected that they would serve as inhibitors of GlfT2 activity. However, only 26% and 15% inhibition was observed for **5.12** and **5.13** respectively, and no inhibitory activity could be detected with **5.14**. It appears therefore, that the addition of the methyl groups disrupt initial substrate binding. This finding is consistent with the model developed as part of a recent X-ray crystallographic study of the enzyme in which a tight donor binding site was proposed.[21]

5.2.1.4. Effect of Acceptor on GlfT2 Donor Specificity.

GlfT2 is a carbohydrate polymerase capable of adding both the β-(1→6)- and β-(1→5)-linked Gal*f* residues to the growing mycobacterial galactan polymer.[8, 17] All the measurements discussed above have focused on the β-(1→6)-transferase activity using the synthetic acceptor **5.6**. We, and others, have previously shown that both transferase activities of GlfT2 originate from the same active site,[19, 20] but it is unknown whether the nature of the glycosidic linkage formed has an influence on the donor substrate binding and specificity of GlfT2. To address this, the same relative activity measurements were performed using the alternative acceptor substrate **5.7**, which is initially a β-(1→5)-transferase substrate.

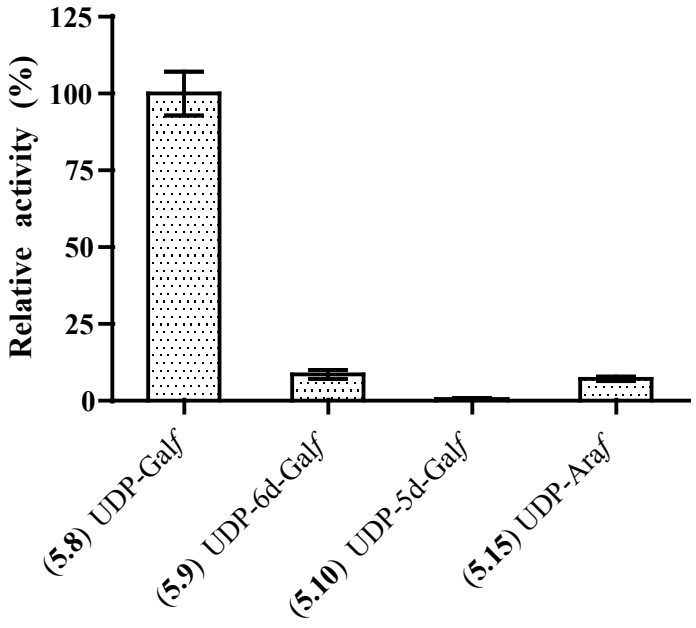


Figure 5-7. Relative activity of GlfT2 with UDP-Gal*f* analogs **5.9**–**5.16** using trisaccharide **5.7** as the acceptor substrate.

Compared to **5.6**, trisaccharide **5.7** is a much poorer GlfT2 substrate; indeed, the relative activity was only 9%. When we attempted to measure kinetic parameters we found that the data showed only a modest fit to the Michealis–Menten equation (Appendix D). GlfT2 has been shown to use a processive mechanism, where the growing galactan polymer remains bound between successive transfers of Gal*f* before dissociating, rather then a distributive mechanism, where the galactan dissociates following each addition.[22, 23] Therefore, the observed activity at higher UDP-Gal*f* concentration likely resulted from the addition of multiple Gal*f* residues. When the data was fit to the Michealis–Menten equation containing a Hill-slope factor, a better fit to the data (Appendix D) was seen. It appears, therefore, that at a low concentration of donor **5.8** we detect predominantly β-(1→5)-transferase activity of GlfT2; however, at higher donor concentrations both the β-(1→5)- and β-(1→6)-transferase activities are measured. The observed kinetics could also imply the enzyme exhibits coopertivity between multiple active sites of the GlfT2 homotetramer. As the activity with acceptor **5.7** was lower than with acceptor **5.6**, only analogs **5.9**, **5.10**, and **5.15** were screened as substrates. In this case, the relative activity values varied when compared to the values observed for the same incubations using acceptor **5.6** (Table 1), but the relative trend is the same. It appears that hydrogen bonding to the C-5 and C-6 hydroxyl groups in the donor play a more important role for the β-(1→5)-transferase activity compared to the β-(1→6)-transferase activity, as evidenced by the lower relative activities for analogs **5.9** and **5.10** with acceptor **5.7**.

*5.2.1.5. TDP-Gal**f** activity and kinetics.*

Recent work studying the substrate binding of GlfT2 using STD-NMR demonstrated that a greater relative saturation transfer is observed for the protons of the ribose sugar and nucleotide base of **5.8** relative to the Gal*f* protons, implying the nucleotide is bound more tightly than other parts of the molecule.[36] These observations are consistent with STD-NMR studies on other glycosyltransferases,[37, 38] and led to the hypothesis that modifications of the donor nucleotide would result in a large decrease in GlfT2 activity by disrupting substrate binding. When the donor nucleotide is changed from UDP (**5.8**) to TDP (**5.16**) we observed ~20% relative turnover. Kinetic analysis with **5.16** revealed an ~8-fold increase and ~1.3-fold decrease in K_M and k_{cat}, respectively, suggesting that the nucleotide portion of the donor is primarily involved in initial substrate binding and recognition, but once bound does not interfere with substrate turnover. A recent crystal structure of the enzyme with bound UDP reveals that no key hydrogen-bonding interactions with the ribose 2-hydroxyl group are present, and in addition, shows sufficient room for binding the methyl group of thymidine, which are both consistent with our observed results.[21]

*5.2.1.6. The UDP-Gal**f** analog lacking a C-3 hydroxyl group is a moderate GlfT2 inhibitor.*

As described above, UDP-Gal*f* analogs **5.10**, **5.11** and **5.15** demonstrated low relative activities with GlfT2 while still displaying moderate K_M values.

These compounds were therefore screened as enzyme inhibitors. Relative inhibition values were obtained by incubating 0.375 mM donor **1** and 2 mM acceptor **2** in the presence of 1.25 mM inhibitor. As presented in Table 5-2, only the 3-deoxy analog **5.11** demonstrated sufficient levels of inhibition under these conditions, which is not surprising as **5.11** also showed the lowest K_M of those analogs evaluated as GlfT2 inhibitor. Inhibition kinetics measurements revealed that **5.11** is, as expected, a competitive inhibitor of GlfT2 with a K_i of 120 μM; c.f. the K_M is 250 μM (Table 5-1). To our knowledge, this represents the most potent GlfT2 inhibitor reported to date.

Table 5-2. Inhibition of GlfT2 by UDP-Gal*f* donor analogs[a]

Inhibitor	Inhibition[a]	K_i (μM)
5.10	10 %	n.d.
5.11	68 %	$1.2 (\pm 0.2) \times 10^2$
5.12	26 %	n.d.
5.13	15 %	n.d.
5.14	n.d.	n.d.
5.15	<1 %	n.d.

[a] Inhibition kinetics data is shown in Appendix D.
[b] Determined at 1.25 mM inhibitor 0.375 mM donor **5.8** and 2.0 mM acceptor **5.6**.

5.2.2. Characterizing GlfT2 reaction products of synthetic UDP-Gal*f* analogs

Using synthetic UDP-Gal*f* analogs **5.9**–**5.11**, **5.15** and **5.16** we uncovered potential hydrogen bonding interactions that are important for GlfT2 activity, but we also wanted to examine the effects of these analogs on both the polymerase activity of GlfT2, and the regiochemistry of the newly formed glycosidic

linkages. To accomplish these goals, we characterized the products of the enzymatic incubations of GlfT2 with these analogs by mass spectrometry and, in cases where sufficient product could be produced, ^{1}H NMR spectroscopy.[17]

5.2.2.1. Use of TDP-Galf as the donor species has no effect on GlfT2 regioselectivity.

The mass spectrum obtained from products isolated from incubations of GlfT2 and **5.16** with a 4-fold excess of acceptor **5.6** (Figure 5-8) showed signals at *m/z* = 801 and 817, the expected mass for the sodium and potassium adducts of an octyl tetrasaccharide containing four Gal*f* residues. As excess acceptor **5.6** was used in the incubations, signals at *m/z* = 639 and 655 for the sodium and potassium adducts of this trisaccharide were also observed. We further analyzed the purified tetrasaccharide product by ^{1}H NMR spectroscopy (Figure 5-9C); a new signal at 5.03 ppm, arising from an additional anomeric proton was observed for this product when compared to trisaccharide **5.6** (Figure 5-9A). This chemical shift of this signal is identical to that observed for the product resulting from incubation of GlfT2 with the natural donor **5.8**, and is characteristic of a β-(1→6)-linked Gal*f* residue (Figure 5-9B).[17, 39] Thus, the replacement of **5.8** with **5.16** does not appear to influence the regioselectivity of the reaction.

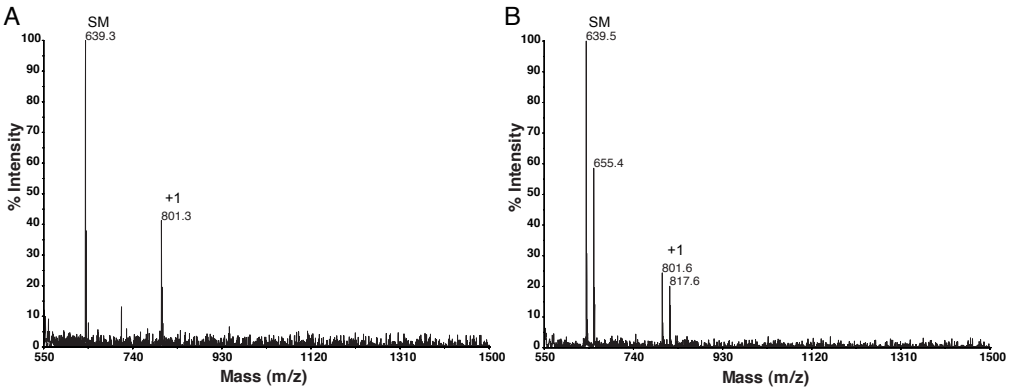


Figure 5-8. MALDI mass spectra of GlfT2 reaction products from incubations of **5.6** with **5.8** (A) or **5.16** (B). SM = starting trisaccharide **5.6**.

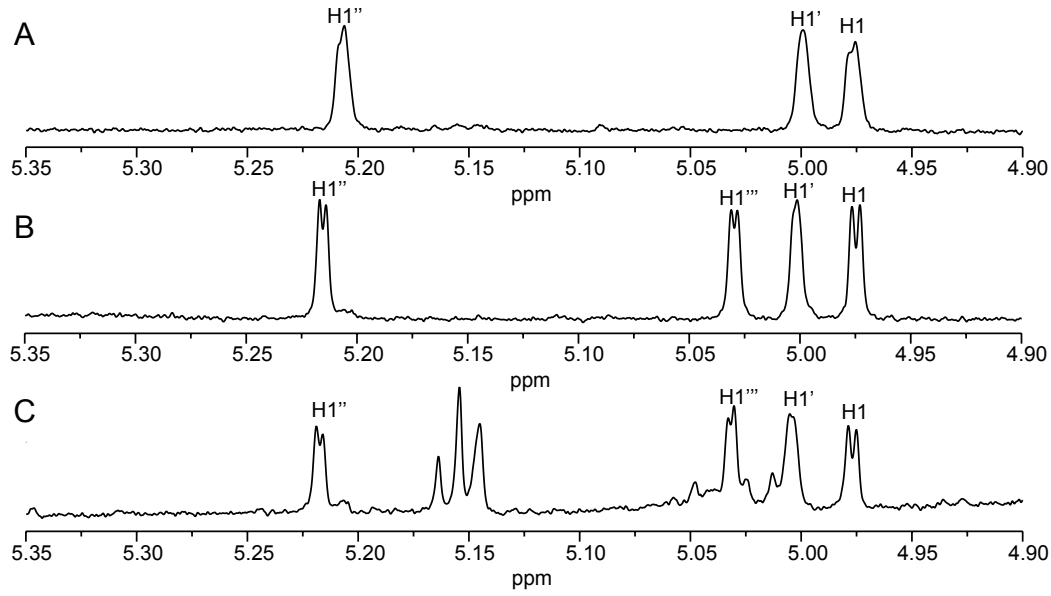


Figure 5-9. Partial ^{1}H NMR spectra of trisaccharide acceptor **5.6** (A) and GlfT2 reaction products from incubations of **5.6** with **5.8** (B) or **5.16** (C). The major signals corresponding to the anomeric hydrogen are labeled.

5.2.2.2. Products formed using UDP-Galf analogs.

The mass spectrum obtained from enzyme reactions containing 0.5 mM **5.6** and 2.0 mM of donor analogs **5.9**, **5.10** or **5.15** showed signals at *m/z* = 785 or *m/z* = 771, the expected mass for the sodium adducts of a deoxy-Gal*f* or Ara*f* containing tetrasaccharide product, respectively. In addition, no signal was observed corresponding to the starting material trisaccharide **5.6**, indicating it was completely consumed (Figure 5-11B, Figure 5-12A,B). Similar incubations with 3-deoxy donor analog **5.11** also showed a signal at *m/z* = 785 for the tetrasaccharide product containing deoxy-Gal*f*, but additionally showed a substantial signal for the starting material trisaccharide **5.6**. This latter observation further demonstrates that **5.11** is a very poor GlfT2 substrate compared to the other deoxy UDP-Gal*f* analogs examined in this chapter. ^{1}H NMR analysis of the tetrasaccharide products resulting from incubations with acceptor **5.6** revealed a new resonance at ~5.00 ppm, as expected for the anomeric proton of a new β-D-Gal*f*-(1→6)- (**5.9**, **5.10**), or α-L-Ara*f*-(1→6)-linkage (**5.15**), (Figure 5-10).

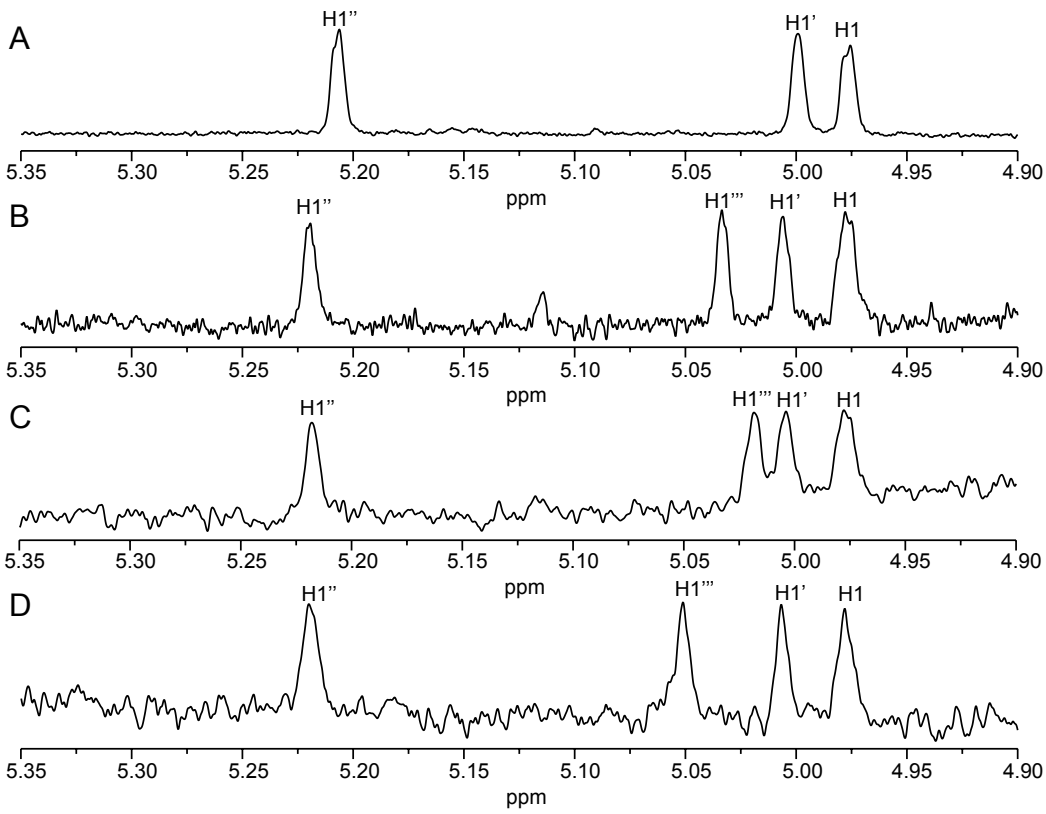


Figure 5-10. Partial ^{1}H NMR spectra of trisaccharide acceptor **5.6** (A) and GlfT2 reaction products from incubations of **5.6** with **5.9** (B), **5.10** (C) or **5.15** (D). The major signals corresponding to the anomeric hydrogen are labeled.

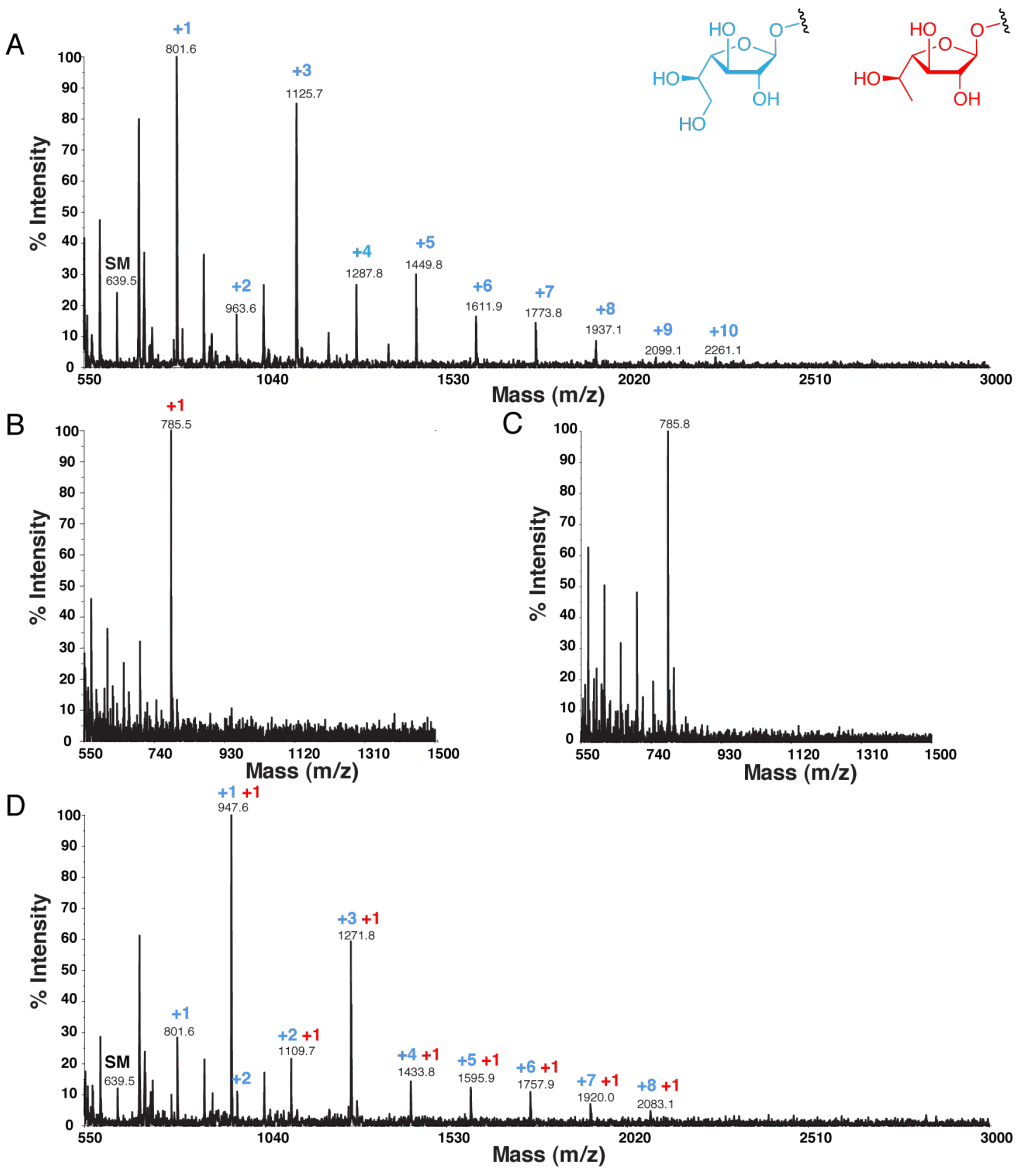


Figure 5-11. MALDI MS analysis of GlfT2 reaction products from incubations of acceptor **5.6** with donor **5.8** (A) or analog **5.9** (B). The blue numbers indicate the additional Gal*f* residues added and the red numbers indicate 6-deoxy-Gal*f* residues added. The peak at *m/z* = 639 corresponds to the sodium adduct of the starting trisaccharide **5.6**. No additional products were formed when the product isolated from incubation B were further incubated with **5.8** and GlfT2 (C). Co-incubation with donors **5.8** and **5.9** show truncated "dead end" products containing 1–8 additional Gal*f* residues and a single 6d-Gal*f* residue (D).

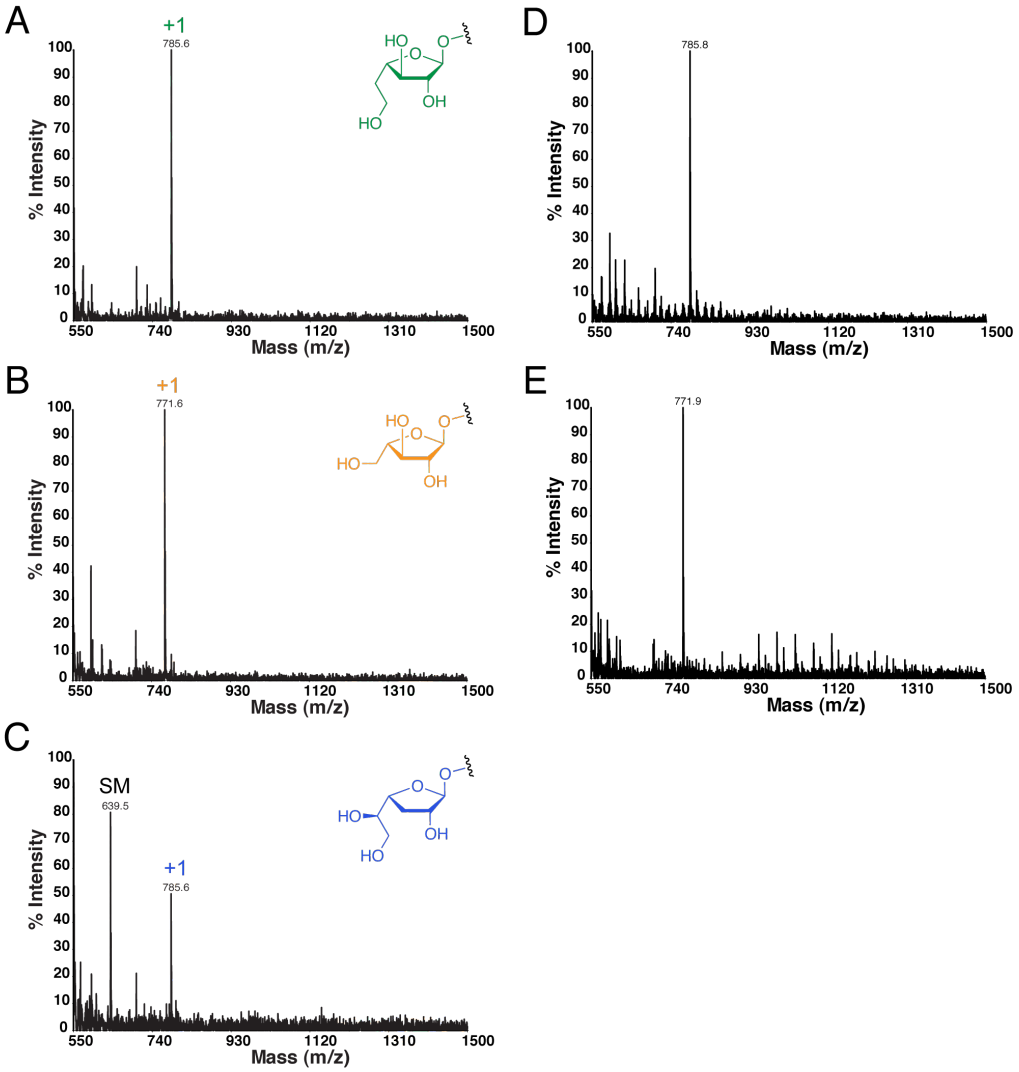


Figure 5-12. MALDI MS analysis of GlfT2 reaction products from incubations with acceptor **5.6** and donor analogs **5.10** (A), **5.15** (B), or **5.11** (C) show only tetrasaccharide products. The only product peaks observed correspond to tetrasaccharides with a single additional donor analog residue (peaks at *m/z* = 785, 771 and 785, respectively). No further products were observed from incubations of GlfT2 with donor **5.8** and the tetrasaccharide products isolated from incubations of **5.10** (D) and **5.15** (E).

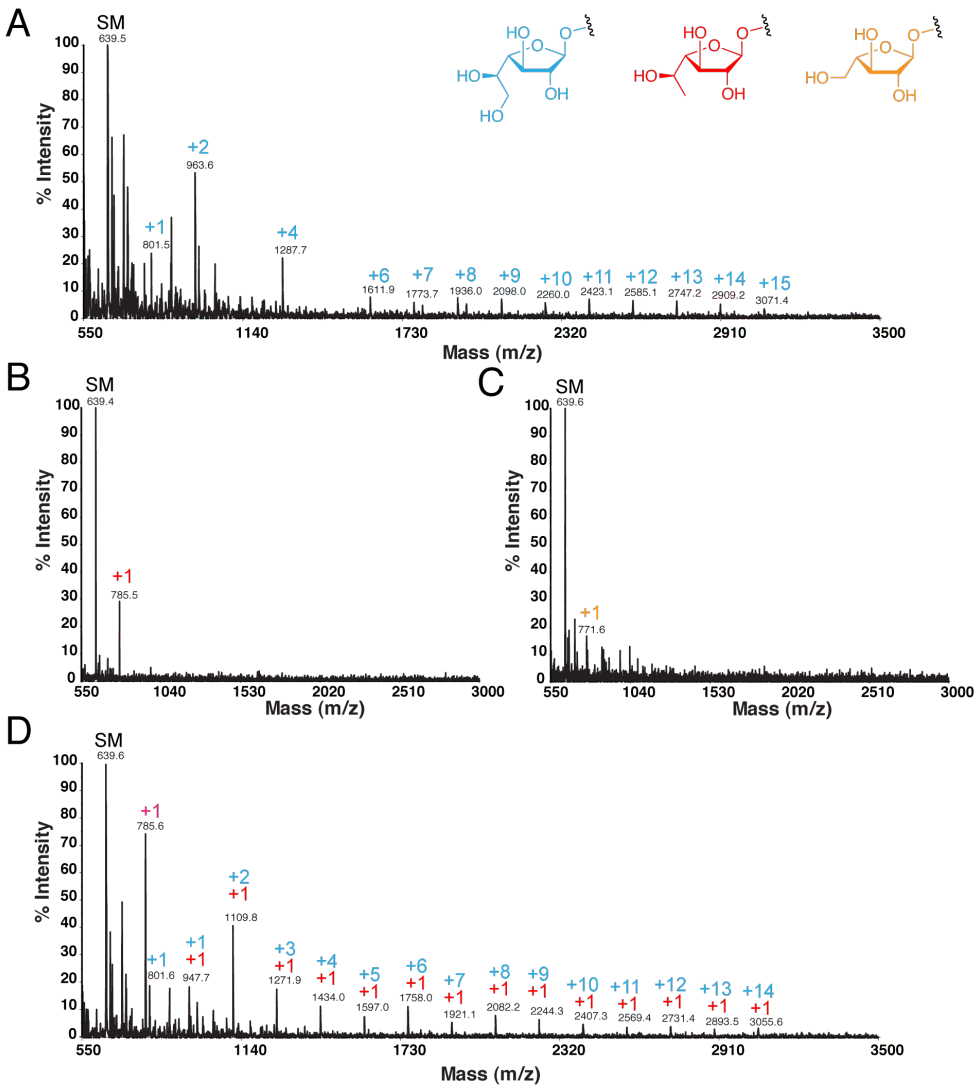


Figure 5-13. MALDI MS analysis of GlfT2 reaction products from incubations with acceptor **5.7** with donor **5.8** (A), or donor analogs **5.9** (B), or **5.15** (C). Numbers shown indicate the number of additional Gal*f* (blue), 6-deoxy-Gal*f* (red), or Ara*f* (green) residues observed. GlfT2 incubated with acceptor **5.7** and donor **5.8** resulted in polymers containing up to an additional 15 Gal*f* residues (A). For incubations of GlfT2 and **5.7** with donors **5.9** (B) or **5.15** (C), only tetrasaccharide products were observed (*m/z* = 785 and 771, respectively). Acceptor **5.7** and GlfT2 co-incubated with **5.8** and **5.9** resulted in a series of "dead-end" polymers containing between 0–14 additional Gal*f* residues and a single 6-deoxy-Gal*f* (D).

5.2.2.3. UDP-Galf analogs result in truncated galactan polymers.

Because of the polymerase activity of GlfT2, incubations with an excess of UDP-Gal*f* lead to the formation of longer galactan polymers, with the length depending on the nature of the acceptor substrate and the conditions under which the incubation is performed.[17, 22, 23] In particular, the nature of the lipid aglycone used in the acceptor has been shown to influence the product distribution.[23] In our hands, incubations containing acceptors **5.6** or **5.7**, which contain an octyl aglycone, with a four-fold excess of UDP-Gal*f* (**5.8**) resulted in polymers containing up to an additional 15 Gal*f* residues (Figure 5-11A and Figure 5-13A), while still showing a signal for the starting trisaccharide acceptor. In contrast, incubations with the deoxygenated UDP-Gal*f* analogs (eg. **5.9**) showed only tetrasaccharide product formation; no longer polymers were observed (Figure 5-11B, Figure 5-12A–C). The incubations containing acceptor **5.7** and donor **5.9** or **5.10** also showed only tetrasaccharide products; however, in these cases a substantial amount of trisaccharide **5.7** was observed (Figure 5-13B, C).

GlfT2 normally adds Gal*f* residues through alternating β-(1→6) and β-(1→5)-linkages. From a large-scale incubation of GlfT2 with acceptor **5.6** and 5-deoxy analog **5.10** the observed tetrasaccharide product **5.18** (Figure 5-14) lacks the terminal C-5 hydroxyl group required for the subsequent β-(1→5)-transferase activity, but still possesses a terminal C-6 hydroxyl group. However, we observed no longer products resulting from subsequent β-(1→6)-transferase activity. Unexpectedly, incubations of GlfT2 with acceptor **5.6** and the 6-deoxy or Ara*f* analogs **5.9** and **5.15**, whose products (**5.17** and **5.19** respectively) still

posses a terminal C-5 hydroxyl group, also produced no pentasaccharide or larger products. It is possible that the reduced activity of the modified UDP-Gal*f* donors may have prevented further extension of the observed tetrasaccharide products. Therefore, we carried out incubations of the isolated tetrasaccharide products with GlfT2 using the natural donor substrate **5.8**. Again, in all cases, the mass spectrum of the products showed only signals for the starting tetrasaccharides **5.17–5.19** and no larger polymers were observed (Figure 5-11C and Figure 5-12D and E). This experiment demonstrates that products containing a terminal Gal*f* residue lacking the C-6 or C-5 hydroxyl group, or containing a terminal Ara*f* residue are "dead end" products that no longer act as GlfT2 substrates. Similar "dead end" products have been previously observed for GlfT1, the first bifunctional galactofuranosyltransferase involved in mycobacterial galactan biosynthesis, using crude membrane preparations of the enzyme.[35]

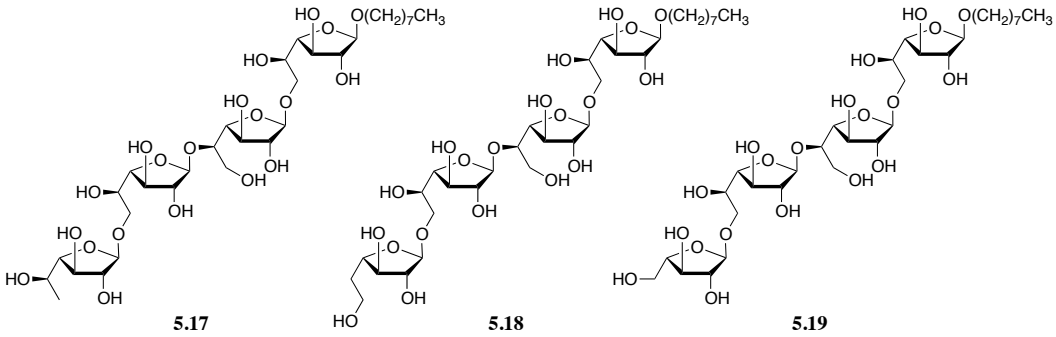


Figure 5-14. Tetrasaccharide products isolated from GlfT2 incubations with acceptor **5.6** and donor analogs **5.9** (**5.17**), **5.10** (**5.18**) and **5.15** (**5.19**).

Observing that tetrasaccharide products **5.17**, **5.18** and **5.19** were not extended led us to postulate that synthetic donor analogs **5.9**, **5.10** and **5.15**

could inhibit GlfT2 catalyzed galactan polymerization by forming prematurely terminated products. We tested this hypothesis by incubating GlfT2 and acceptor trisaccharide **5.6** with a mixture of both donor **5.8** and analog **5.9**. The reactions were monitored for the formation of truncated products by MALDI MS. As seen in Figure 5-11D, the predominant product peaks observed contain a single 6-deoxy-Gal*f* residue and between one and eight additional Gal*f* residues with the major product being a 6-deoxy-Gal*f* containing pentasaccharide (*m/z* = 947.6). Other than 6-deoxy-Gal*f* containing products, only a small amount of tetrasaccharide and pentasaccharide were produced and no trisaccharide starting material **5.6** remained. The same "dead end" products were observed when analogous reactions were performed with donor analogs **5.10** or **5.15**.

These results were not limited to acceptor trisaccharide **5.6**, incubations containing acceptor **5.7** also resulted in the formation of truncated polymers containing a single modified Gal*f* residue (Figure 5-13D). For all of these incubations, no more than a single deoxy-Gal*f* or Ara*f* residue could ever be detected in the resulting products. Combined, these results demonstrate deoxygenated UDP-Gal*f* donors modified at C-6′′ and C-5′′ can be readily incorporated into a growing galactan chain through the action of GlfT2 resulting in the production of truncated products that prevent further polymerization. This demonstrates dual recognition of both the acceptor substrates terminal C-5 and C-6 hydroxyl groups are essential for activity. This model is consistent with the interactions proposed to be present in the active site based on X-ray crystallographic investigations of the protein. Notably, the hydroxyl group

adjacent to that undergoing glycosylation is proposed to form (apparently essential) hydrogen bonds with two amino acids of the protein (Figure 5-15).[21]

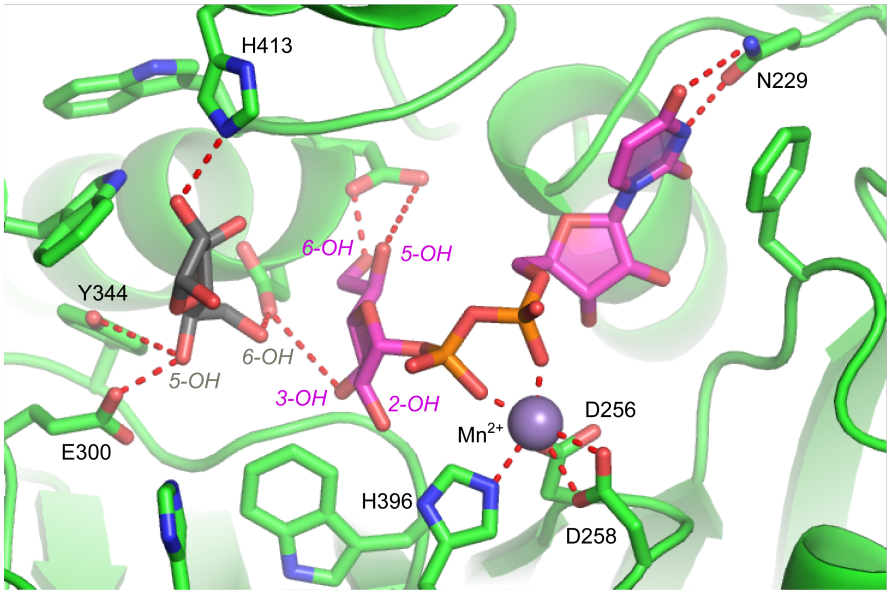


Figure 5-15. Crystal structure of GlfT2 with UDP-Gal*f* (**5.8**) modeled into the proposed active–site. The bound manganese^{2+} observed in the crystal structure is shown as a sphere, and a single Gal*f* residue of the acceptor has been modeled in. Proposed hydrogen bonding interactions are indicated by dashed lines.

5.2.4 GlfT2 Structure and Regioselectivity

Based on the crystal structure of GlfT2 with either trisaccharide acceptor **5.6** (Figure 5-16A) or **5.7** (Figure 5-16B) modeled into the proposed active site, it appears that two aromatic amino acids (W399 and H413) interact with the penultimate Gal*f* residue. The W399 residue appears to stack with the hydrophobic face and H413 hydrogen bond to the 2-OH or 3-OH of the penultimate Gal*f* residue of both acceptor **5.6** or **5.7**. Based on this model it was hypothesized that W399 and H413 could be responsible for controlling the

regioselectivity of the alternating β-(1→5) and β-(1→6) activities of GlfT2 by orienting the acceptor substrate. In addition to these two aromatic amino acids, E300 appeared to hydrogen bond with the non-reacting C5 or C6 hydroxyl group of the terminal Gal*f* residue (for **5.6** or **5.7**, respectively), which may also be critical for orienting the acceptor substrate for alternating β-(1→5), β-(1→6) activity.

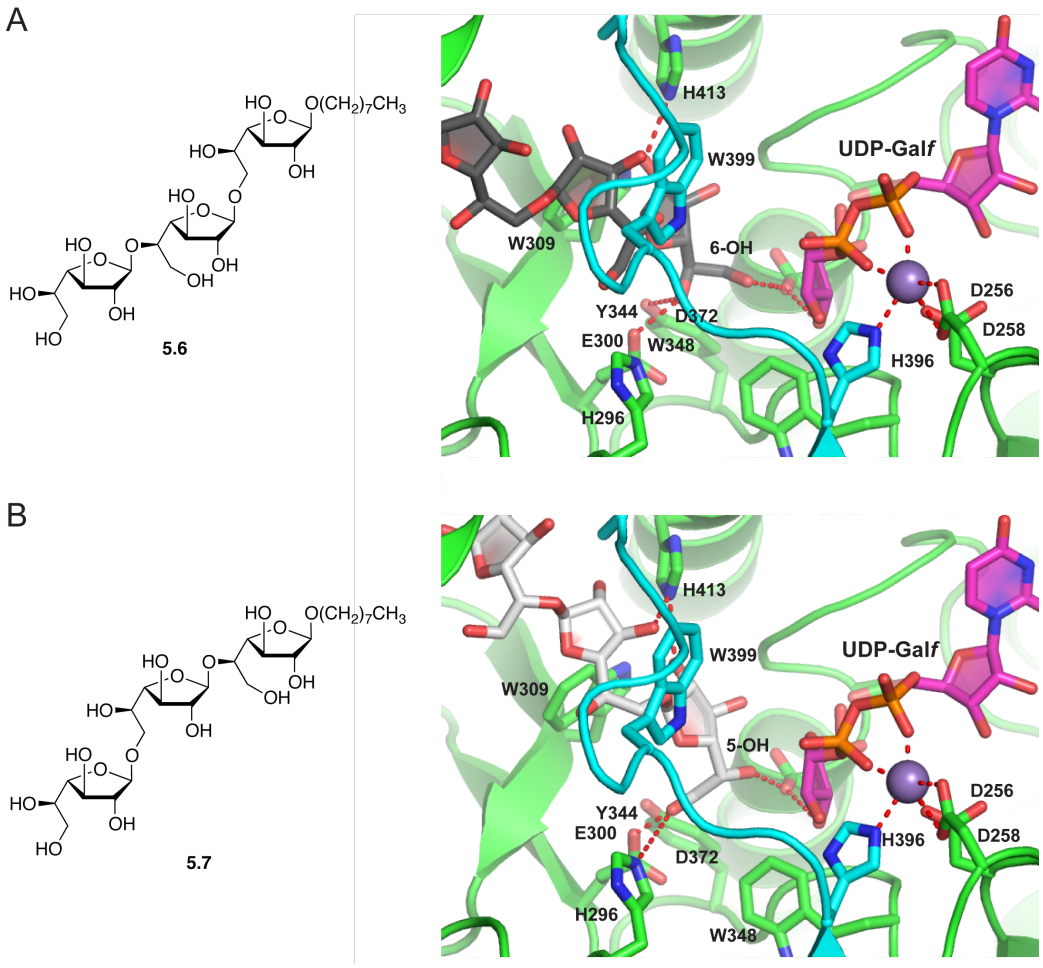


Figure 5-16. Model of acceptor **5.6** (A) and acceptor **5.7** (B) bound to GlfT2 with Mn2+ (sphere) and donor UDP-Gal*f* **5.8**. The chemical structures of acceptors **5.6** (containing a terminal (1→5)-linked β-D-Gal*f* residue) and **5.7** (containing a terminal (1→6)-linked β-D-Gal*f* residue) are shown for clarity.

5.2.4.1 GlfT2 Mutants disrupt acceptor binding and turnover

To validate the substrate binding model proposed in Figure 5-16, based on the crystal structure of GlfT2,[21] three amino acids (E300, W399, and H413) proposed to interact with the acceptor substrate (**5.6** in Figure 5-16A and **5.7** in Figure 5-16B) were mutated to serine residues and the activity of the resulting mutant proteins was evaluated by Ruixiang Blake Zheng in our group. As can be seen in Table 5-3, mutation of E300S, W399S, or H413S did not change the apparent K_M for the donor substrate (**5.8**) (Table 5-3), but caused an approximate 2–3 fold increase in the apparent acceptor **5.6** K_M values. This is to be expected if these amino acid residues interact only with the acceptor substrate as has been proposed. In addition, all three GlfT2 mutants showed a greater than 1000 fold decrease in the k_{cat}. These results support the proposed interactions between E300, W399, and H413 and the acceptor **5.6** (Figure 5-16A).

Table 5-3. Kinetic parameters for GlfT2 Mutants*

	$K_{M,\,app}$, (μM)		k_{cat} (min^{-1})
	UDP-Gal*f* **5.8**	acceptor **5.6**	
Wild-type	380 ± 60	600 ± 20	430 ± 35
E300S	420 ± 60	1780 ± 120	0.30 ± 0.01
W399S	410 ± 40	1780 ± 100	0.35 ± 0.04
H413S	430 ± 50	1540 ± 140	0.25 ± 0.01

5.2.4.2 Characterizing GlfT2 Mutant Reaction Products.

The kinetics data obtained for the GlfT2 mutant enzymes supports the role of E300, W399, and H413 interacting specifically with the acceptor substrate. We then wanted to examine if these residues play a role in the regioselectivity of the glycosylation reaction catalyzed by GlfT2 by orienting the acceptor substrate for altering β-(1→5), β-(1→6) bond formation. To test this hypothesis, the products from large-scale incubations of the GlfT2 E300S, W399S, and H413S mutants were purified and the regioselectivity of the newly formed glycosidic linkages were analyzed by ^{1}H NMR spectroscopy. The reaction conditions were controlled to give predominantly the tetra-, penta-, and/or hexasaccharide products. The tetrasaccharide product obtained with the wild-type GlfT2 showed, as expected, a newly formed β-(1→6) glycosidic linkage to the terminal Gal*f* (Figure 5-17A), and the next product isolated, the hexasaccharide in this case (Figure 5-17B), also contained an additional β-(1→5)– and β-(1→6)–Gal*f* (as the fifth and sixth residues respectively), as expected from alternating activity. A tetrasaccharide was the predominant product isolated from reactions with the E300S, W399S and H413S GlfT2 mutants (Figure 5-17C, D and F, respectively) where the newly formed glycosidic linkage was β-(1→6) in each case, identical to the wild-type tetrasaccharide. In addition, the pentasaccharide product was observed for incubations with the W399S and H413S mutants (Figure 5-17E and F, respectively). In both cases, the fifth Gal*f* residue was added in a β-(1→5) glycosidic linkage, as would be expected if the alternating activity was retained.

These results show that although the GlfT2 E300S, W399S and H413S mutants showed >1000 fold reduction in activity compared to the wild-type GlfT2, these mutations did not influence the regioselectivity of the glycosylation.

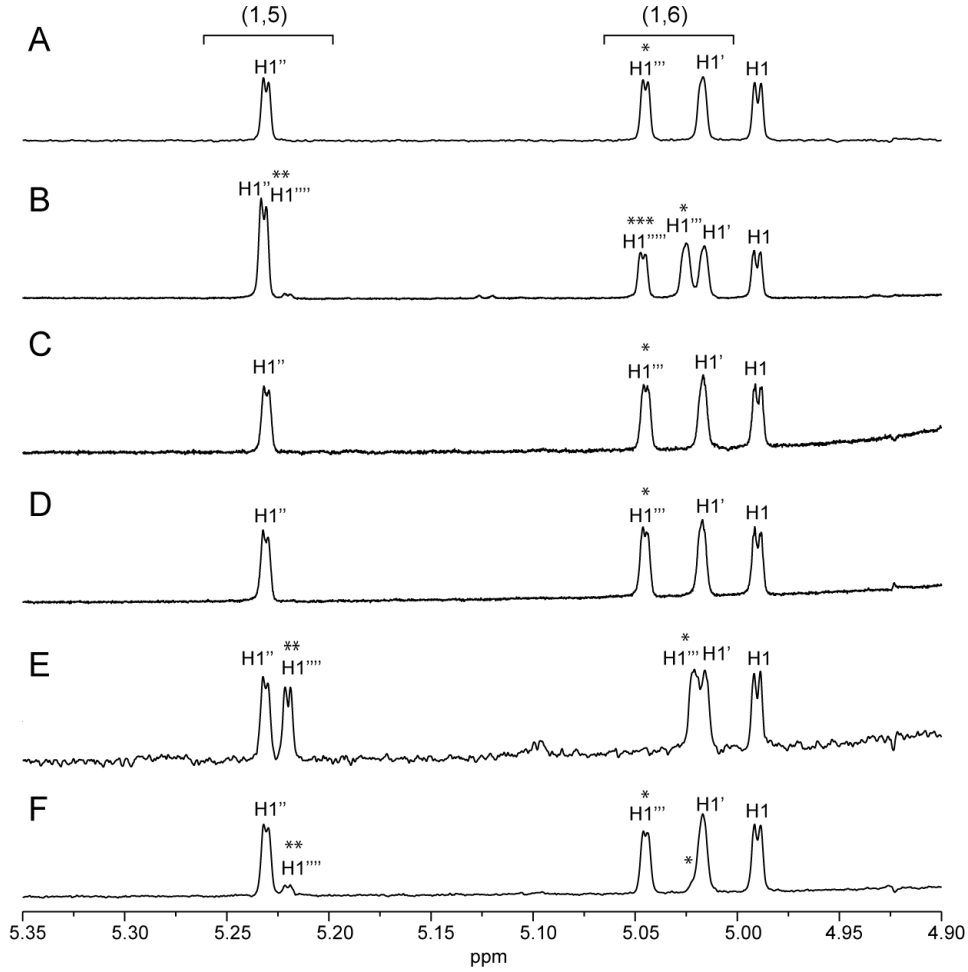


Figure 5-17. Partial ^{1}H NMR spectra of GlfT2 mutant reaction products. Both a tetrasaccharide (A) and Hexasaccharide (B) product were isolated from the incubations of acceptor **5.6** and UDP-Gal*f* **5.8** with wild-type GlfT2. Tetrasaccharide products isolated from GlfT2 E300S (C) and W399S (D) mutants are identical to wild-type tetrasaccharide. A pentasaccharide isolated from the GlfT2 W399S mutant shows an additional β-(1→5)-linkage (E). The product isolated from the GlfT2 H413S mutant (F) contains both the Tetrasaccharide product and pentasaccharide resulting from alternating β-(1→6), β-(1→5) activity. The major signals corresponding to the anomeric hydrogen are labeled, and new anomeric protons are indicated with *, **, or ***.

5.3. Conclusions

In this chapter, we examined the donor binding site specificity of GlfT2, a bifunctional galactofuranosyltrasferase involved in the biosynthesis of the mycobacterial mAG complex, using a panel of singly modified UDP-Gal*f* analogs. The results show that UDP-Gal*f* binds to GlfT2 in a sterically crowded region of the active site, consistent with a recent crystal structure model (Figure 5-15).[21] In addition, there appear to be numerous interactions between the enzyme and the carbohydrate hydroxyl groups of the UDP-Gal*f* donor, as indicated by the reduced activity observed with deoxy donor analogs **5.9**–**5.11**. These protein–carbohydrate interactions, which are consistent with a model for donor binding proposed as part of a recent crystallographic investigation of the enzyme,[21] are not essential for substrate binding, as only moderate (2 to 4-fold) increases in K_M were observed in all cases. Instead, these observations suggest hydrogen bonding to the Gal*f* hydroxyl groups assist to orient the Gal*f* ring for turnover.

Despite using a single active site,[19] our results show, similar to those reported earlier,[17, 23] that the β-(1→6)-transferase activity is more efficient that the β-(1→5)-transferase activity of GlfT2, at least with the synthetic trisaccharide acceptors used in this study. Further analysis with acceptor analogs of different lengths would reveal whether the difference between the β-(1→5)- and β-(1→6)-transferase activity is related to the size of the acceptor substrate and these studies are currently under investigation. The difference in β-(1→5)- and β-(1→6)-transferase activity is likely controlled by subtle interactions

between the enzyme and the acceptor substrate, as our results show that donor substrate binding is not substantially influenced by the nature of the glycosidic linkage formed (i.e., the same trends in donor recognition are seen both acceptors **5.6** and **5.7**). These results suggest little, if any, reconfiguration of the donor-binding pocket is required between successive glycosylation events, which is also consistent with the processive nature of the enzyme. However, the details of these interactions are not well understood and available crystallographic data[21] does not provide significant clarity on these differences.

The donor analogs used in this study had no effect on the regioselectivity of the glycosylations catalyzed by GlfT2. The alternating regioselectivity appears to be influenced exclusively by acceptor binding interactions, which, based on these observations, are not influenced by the donor. Despite our efforts to elucidate the acceptor binding interactions responsible for the glycosylation regioselectivity, the specific interaction still remain to be elucidated. Nevertheless, the C-6 and C-5 modified UDP-Gal*f* analogs **5.9**, **5.10** and **5.15** interfered with normal galactan polymerization. GlfT2 readily incorporates these analogs into a growing galactan; however, the enzyme only adds a single modified Gal*f* residue, in turn producing products that cannot be further elongated. Despite being a template independent enzyme, GlfT2 specifically forms alternating β-(1→5), β-(1→6) glycosides and is unable to add two successive Gal*f* with the same linkage. Similar "dead-end" products have been observed in the study of GlfT1,[35] but this represents the first observation of these products for GlfT2. Subsequent studies using fluoro UDP-Gal*f* analogs[40] also

observed the product termination with 5-fluoro or 6-fluoro UDP-Gal*f*, but only when the fluorine atom replaces the hydroxyl group that should react to form the next glycosidic bond. These results combined demonstrate that, in addition to requiring a hydroxyl group at the next site of glycosylation, hydrogen bonding to the C-5 or C-6 hydroxyl group, of the terminal residue, adjacent to the site of glycosylation is essential for GlfT2 activity, consistent with the mechanism for substrate recognition proposed from modeling of the acceptor substrate to the recent crystal structure of GlfT2.[21]

Although the synthetic UDP-Gal*f* analogs prepared in this study have limited potential as chemotherapeutics due to their poor cell permeability, they have proven to be useful probes for studying the process of galactan biosynthesis *in vitro*. These probes have provided valuable information regarding the importance of various carbohydrate–protein interactions occurring in the active site of GlfT2, which can now be explored for the development of novel inhibitors to target galactan biosynthesis. In addition, analogs **5.9**, **5.10** and **5.15**, which resulted in "dead end" products, could prove useful tools to elucidate the mechanism of bifunctional activity for GlfT2, a question that still remains unanswered.

5.4. Experimental Details

GlfT2 activity, kinetics, and inhibition.

The GlfT2 protein was prepared as previously described, and its activity was determined using the coupled spectrophotometric assay reported

previously.[32] Assays were performed in 384 well micotiter plates in a volume of 40 μL containing 100 mM MOPS, pH 7.6, 50 mM KCl, 20 mM $MgCl_2$, 1.1 mM NADH, 3.5 mM phosphoenolpyruvate (PEP), 7.5 U pyruvate kinase (PK, EC 2.7.1.40), 16.8 U lactate dehydrogenase (LDH, EC 1.1.1.27), 2 mM acceptor trisaccharide **5.6** or **5.7**, and 2 mM donor substrate (**5.8**–**5.16**). The amount of GlfT2 added was controlled to allow for sufficient substrate turnover. Assays were continuously monitored at 37 °C over 20 min and initial velocities were determined from the decrease in NADH absorbance at 340 nm. Specific activities were measured in duplicate for all donor analogs.

Kinetic values were determined by varying the concentration of donor analog (**5.8**–**5.16**) between 0–4000 μM while keeping the concentration of acceptor trisaccharide **5.6** fixed at 2 mM. At this concentration, the acceptor trisaccharide is saturating (reported K_M = 208 ± 50 μM for **5.6**),[19] allowing for single substrate kinetics of the donor analogs to be measured. Assays were run in duplicate and initial velocities were determined at each substrate concentration. Kinetic parameters K_M and k_{cat} were obtained by nonlinear regression analysis of the Michaelis–Menten equation using GraphPad PRISM 4 (GraphPad Software, San Diego, CA).

GlfT2 percent inhibition was determined using the coupled spectrophotometric assay with 1250 μM donor analog (**5.10**–**5.15**), and acceptor trisaccharide **5.6** and UDP-Gal*f* donor **5.8** concentration fixed at 2000 μM and 375 μM respectively. Donor analogs showing >50% inhibition were further evaluated. Inhibition kinetics were determined by varying the concentration of

donor **5.8** between 0–2500 μM while the concentration of acceptor trisaccharide **5.6** fixed at 2 mM. Assays were performed at varying concentrations of donor analog (0–1000 μM). The inhibition constant K_i was determined by nonlinear regression analysis using GraphPad PRISM 4 software.

Characterization of UDP-Gal*f* (5.8) and TDP-Gal*f* (5.16) reaction products.

Reactions containing 50 mM MOPS pH 7.6, with 20 mM $MgCl_2$, 500 μM donor **5.8** or **5.16**, 3000 μM trisaccharide acceptor **5.6**, and 100 μg GlfT2 in a total volume of 400 μL were incubated at ambient temperature under a nitrogen gas atmosphere for 4 days with gentle rotation. To ensure only singly glycosylated products were produced, a six-fold excess of trisaccharide **5.6** was used. Progress of the enzymatic reactions was monitored by thin-layer chromatography (TLC) on SiliaPlate TLC silical gel plates (Silicycle) eluting with $CHCl_3$–CH_3OH–NH_4OH–H_2O (65:25:0.5:3.6) as previously described.[17] Reaction products on TLC were visualized using 3% anisaldehyde in sulfuric acid stain. After 4 days, the reactions were diluted to 1 mL with MilliQ water, filtered through 0.22 μm Millex-GV filters, and 200 μL of the filtrate was lyophilized for MALDI MS analysis using a Voyager Elite time-of-flight spectrometer in positive ion mode. Preparative TLC was used to purify the products from the remaining 800 μL of filtrate. The silica from the area of TLC plate corresponding to the reaction product (R_f 0.31) was scraped from the plate without visualization using the 3% anisaldehyde–sulfuric acid stain. HPLC grade methanol (4 mL) was used to extract the purified reaction product, the silica was

filtered, and the methanol was evaporated. The resulting residue was re-suspended in MilliQ water and passed through a 0.22 μm Millex-GV filter, the filtrate lyophilized, resuspended in D_2O (1 mL) and again lyophilized. Products were dissolved in 700 μL D_2O and one-dimensional ^1H NMR spectra were recorded on a Varian i600 instrument with suppression of the HOD signal using a presaturation pulse sequence, irradiating at 4.67 ppm.

Isolation and characterization UDP-Gal*f* analog reaction products.

Reactions containing 50 mM MOPS pH 7.6 with 20 mM $MgCl_2$, 5 mM *β*-mercaptoethanol, 2 mM donor analog **5.9–5.11** or **5.15**, 500 μM trisaccharide acceptor **5.6** or **5.7**, 50 μg GlfT2, and 2 units of alkaline phosphatase (AP) were incubated under a nitrogen gas atmosphere at ambient temperature for 3 days with gentle rotation. To promote production of polymeric products a four-fold excess of donor was used; also, AP was added to degrade the UDP by-product produced during the reaction, which is known to inhibit GlfT2. Reaction progress was monitored by TLC, again eluting with $CHCl_3$–CH_3OH–NH_4OH–H_2O (65:25:0.5:3.6). After 3 days, the reaction products were purified using a Sep-Pak C_{18} cartridge. After washing with ~10 mL of water to remove the enzyme and unreacted donor, the reaction products were eluted using 4 mL of HPLC grade CH_3OH. The solvent was then evaporated; the products were resuspended in 1 mL water and passed through a 0.22 μm Millex-GV filter. From this solution, 100 μL was lyophilized and re-suspended in 2,5-dihydroxy benzoic acid and characterized by MALDI MS as described above. The

remaining 900 μL of the extraction solution was lyophilized and re-suspended in 600 μL D_2O. One-dimensional ^1H NMR spectra were recorded on a Varian i600 instrument with suppression of the HOD signal using a presaturation pulse sequence, irradiating at 4.67 ppm.

Incubation of tetrasaccharide reaction products with UDP-Gal*f*.

The purified and lyophilized reaction products **5.17**–**5.19** were re-suspended in 50 mM MOPS pH 7.6 with 20 mM $MgCl_2$, 5 mM *β*-mercaptoethanol, 2 mM UDP-Gal*f* **5.8**, 50 μg GlfT2, and 2 units of alkaline phosphatase (AP) in a final volume of 100 μL. The reaction mixtures were incubated under a nitrogen gas atmosphere at ambient temperature for 1 day. The products were purified and analyzed by MALDI MS as described above.

Inhibition of Galactan polymerization by UDP-Gal*f* analogs.

Reactions containing 50 mM MOPS pH 7.6 with 20 mM $MgCl_2$, 5 mM *β*-mercaptoethanol, 2 mM donor analog **5.9**, **5.10** or **5.15**, 1 mM UDP-Gal*f* **5.8**, 500 μM trisaccharide acceptor **5.6** or **5.7**, 50 μg GlfT2, and 2 units of alkaline phosphatase (AP) in a 100 μL final volume, were incubated under a nitrogen gas atmosphere at ambient temperature for 3 days with gentle rotation. The products were purified and analyzed by MALDI MS as described above.

Isolation and characterization GlfT2 mutant reaction products

Reactions containing 50 mM HEPES buffer pH 7.4 with 25 mM $MgCl_2$, 50 mM NaCl, 1 mM (1,5)(1,6)-trisaccharide acceptor, 1 mM of UDP-Gal*f*, 35–60 μg GlfT2 mutant enzyme, and 2 units of alkaline phosphatase (AP) in 500 μL total volume were incubated under a nitrogen gas atmosphere at ambient temperature. AP was used to degrade the UDP by-product produced during the reaction, which is known to inhibit GlfT2. The reaction progress was monitored by TLC eluting with EtOAc–CH_3OH—H_2O (7:2:1). After 10 min an additional equivalent of UDP-Gal*f* was added to the reaction to facilitate full conversion to the tetrasaccharide, pentasaccharide and hexasaccharide products, and an additional equivalent was then added every hour until no trisaccharide acceptor remained. The reactions were stopped by adding 1 mL of CH_3OH. The precipitated protein was removed by centrifugation (5000 ×g, 4 min), and the supernatant was evaporated to dryness. The reaction products were separated and purified by preperative TLC (EtOAc–CH_3OH—H_2O, 7:2:1), then by Sep-Pak C_{18} cartridge filtration. After washing with ~10 mL of water, the reaction products were eluted using 4 mL of HPLC grade CH_3OH. The solvent was then evaporated; the products were resuspended in 1 mL water and passed through a 0.22 μm Millex-GV filter. The solution was lyophilized and re-suspended in 650 μL D_2O. One-dimensional ^1H NMR spectra were recorded on a Varian v700 instrument with the presaturation of the HOD signal. The sample was then diluted with 5 mL water, lyophilized and then analyzed by ESI-MS.

5.5. Bibliography

1. Paolo, W. F.; Nosanchuk, J. D., *Lancet Infect. Dis.* **2004,** *4*, 287-293.
2. Davies, P. D. O., *Ann. Med.* **2003,** *35*, 235-243.
3. Coker, R. J., *Trop. Med. Int. Health* **2004,** *9*, 25-40.
4. Chan, E. D.; Iseman, M. D., *Curr. Opin. Infect. Dis.* **2008,** *21*, 587-595.
5. Pedersen, L. L.; Turco, S. J., *Cell. Mol. Life Sci.* **2003,** *60*, 259-266.
6. Brennan, P. J.; Nikaido, H., *Annu. Rev. Biochem.* **1995,** *64*, 29-63.
7. Brennan, P. J., *Tuberculosis* **2003,** *83*, 91-97.
8. Crick, D. C.; Mahapatra, S.; Brennan, P. J., *Glycobiology* **2001,** *11*, 107R-118R.
9. Jarlier, V.; Nikaido, H., *J. Bacteriol.* **1990,** *172*, 1418-1423.
10. Umesiri, F. E.; Sanki, A. K.; Boucau, J.; Ronning, D. R.; Sucheck, S. J., *Med. Res. Rev.* **2010,** *30*, 290-326.
11. Slayden, R. A.; Lee, R. E.; Barry, C. E., *Mol. Microbiol.* **2000,** *38*, 514-525.
12. Belanger, A. E.; Besra, G. S.; Ford, M. E.; Mikusova, K.; Belisle, J. T.; Brennan, P. J.; Inamine, J. M., *Proc. Natl. Acad. Sci. U. S. A.* **1996,** *93*, 11919-11924.
13. Beláňová, M.; Dianišková, P.; Brennan, P. J.; Completo, G. C.; Rose, N. L.; Lowary, T. L.; Mikušová, K., *J. Bacteriol.* **2008,** *190*, 1141-1145.
14. Weston, A.; Stern, R. J.; Lee, R. E.; Nassau, P. M.; Monsey, D.; Martin, S. L.; Scherman, M. S.; Besra, G. S.; Duncan, K.; McNeil, M. R., *Tuber. Lung Dis.* **1998,** *78*, 123-131.

15. Alderwick, L. J.; Dover, L. G.; Veerapen, N.; Gurcha, S. S.; Kremer, L.; Roper, D. L.; Pathak, A. K.; Reynolds, R. C.; Besra, G. S., *Protein Expression Purif.* **2008,** *58*, 332-341.

16. Kremer, L.; Dover, L. G.; Morehouse, C.; Hitchin, P.; Everett, M.; Morris, H. R.; Dell, A.; Brennan, P. J.; McNeil, M. R.; Flaherty, C.; Duncan, K.; Besra, G. S., *J. Biol. Chem.* **2001,** *276*, 26430-26440.

17. Rose, N. L.; Completo, G. C.; Lin, S. J.; McNeil, M.; Palcic, M. M.; Lowary, T. L., *J. Am. Chem. Soc.* **2006,** *128*, 6721-6729.

18. Pathak, A. K.; Pathak, V.; Seitz, L.; Maddry, J. A.; Gurcha, S. S.; Besra, G. S.; Suling, W. J.; Reynolds, R. C., *Bioorg. Med. Chem.* **2001,** *9*, 3129-3143.

19. Szczepina, M. G.; Zheng, R. B.; Completo, G. C.; Lowary, T. L.; Pinto, B. M., *ChemBioChem* **2009,** *10*, 2052-2059.

20. May, J. F.; Levengood, M. R.; Splain, R. A.; Brown, C. D.; Kiessling, L. L., *Biochemistry* **2012,** *51*, 1148-1159.

21. Wheatley, R. W.; Zheng, R. B.; Richards, M. R.; Lowary, T. L.; Ng, K. K. S., *J. Biol. Chem.* **2012**. <doi:10.1074/jbc.M112.347484>

22. Levengood, M. R.; Splain, R. A.; Kiessling, L. L., *J. Am. Chem. Soc.* **2011,** *133*, 12758-12766.

23. May, J. F.; Splain, R. A.; Brotschi, C.; Kiessling, L. L., *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106*, 11851-11856.

24. Frigell, J.; Pearcey, J. A.; Lowary, T. L.; Cumpstey, I., *Eur. J. Org. Chem.* **2011**, 1367-1375.

25. Cren, S.; Gurcha, S. S.; Blake, A. J.; Besra, G. S.; Thomas, N. R., *Org. Biomol. Chem.* **2004,** *2*, 2418-2420.

26. Cren, S.; Wilson, C.; Thomas, N. R., *Org. Lett.* **2005,** *7*, 3521-3523.

27. Trunkfield, A. E.; Gurcha, S. S.; Besra, G. S.; Bugg, T. D. H., *Bioorg. Med. Chem.* **2010,** *18*, 2651-2663.

28. Vembaiyan, K.; Pearcey, J. A.; Bhasin, M.; Lowary, T. L.; Zou, W., *Bioorg. Med. Chem.* **2011,** *19*, 58-66.

29. Spohr, U.; Hindsgaul, O.; Lemieux, R. U., *Can. J. Chem.* **1985,** *63*, 2644-2652.

30. Hindsgaul, O.; Kaur, K. J.; Srivastava, G.; Blaszczykthurin, M.; Crawley, S. C.; Heerze, L. D.; Palcic, M. M., *J. Biol. Chem.* **1991,** *266*, 17858-17862.

31. Laferte, S.; Chan, N. W. C.; Sujino, K.; Lowary, T. L.; Palcic, M. M., *Eur. J. Biochem.* **2000,** *267*, 4840-4849.

32. Rose, N. L.; Zheng, R. B.; Pearcey, J.; Zhou, R.; Completo, G. C.; Lowary, T. L., *Carbohydr. Res.* **2008,** *343*, 2130-2139.

33. Neverova, I.; Scaman, C. H.; Srivastava, O. P.; Szweda, R.; Vijay, I. K.; Palcic, M. M., *Anal. Biochem.* **1994,** *222*, 190-195.

34. Completo, G. C.; Lowary, T. L., *J. Org. Chem.* **2008,** *73*, 4513-4525.

35. Peltier, P.; Beláňová, M.; Dianišková, Petronela; Zhou, R.; Zheng, R. B.; Pearcey, J. A.; Joe, M.; Brennan, P. J.; Nugier-Chauvin, C.; Ferrières, V.; Lowary, T. L.; Daniellou, R.; Mikušová, K., *Chem. Biol.* **2010,** *17*, 1356-1366.

36. Szczepina, M. G.; Zheng, R. X. B.; Completo, G. C.; Lowary, T. L.; Pinto, B. M., *Bioorg. Med. Chem.* **2010,** *18*, 5123-5128.

37. Macnaughtan, M. A.; Kamar, M.; Alvarez-Manilla, G.; Venot, A.; Glushka, J.; Pierce, J. M.; Prestegard, J. H., *J. Mol. Biol.* **2007,** *366*, 1266-1281.

38. Angulo, J.; Langpap, B.; Blume, A.; Biet, T.; Meyer, B.; Krishna, N. R.; Peters, H.; Palcic, M. M.; Peters, T., *J. Am. Chem. Soc.* **2006,** *128*, 13529-13538.

39. Cyr, N.; Perlin, A. S., *Can. J. Chem.* **1979,** *57*, 2504-2511.

40. Brown, C. D.; Rusek, M. S.; Kiessling, L. L., *J. Am. Chem. Soc.* **2012**, *134*, 6552-6555.

Chapter 6

Conclusions and Future Directions

6.1 Summary and Future Directions

The enzymes involved in furanose sugar metabolism continue to gain interest as putative targets for the development of new anti-microbial therapeutics.[1] Inhibiting these enzymes could offer new methods to treat diseases ranging from tuberculosis to protozoan infections, where furanoside sugars are essential for the virulence or viability of the causative organisms.[2-4] In this thesis, I have described investigations into the activity and specificity of enzymes involved in the biosynthesis of furanose sugars found in bacterial cell wall glycans. Specifically, four putative pyranose–furanose mutase enzymes that function in the biosynthesis of activated furanoside donors used in the assembly of the capsular polysaccharide (CPS) of strains of gastrointestinal pathogen *Campylobacter jejuni* have been examined. These putative enzymes are homologs of the UDP-galactopyranose mutase (UGM) enzyme, the enzyme responsible for the synthesis of UDP-α-D-galactofuranose (UDP-D-Gal*f*) from UDP-α-D-galactopyranose (UDP-D-Gal*p*). These enzymes, encoded by the *glf* gene, have been identified in numerous bacteria, fungi, and protozoan pathogens that express D-Gal*f* in their glycoconjugates. However, there have been few reports of similar enzymes or pathways involved in the biosynthesis of other furanose sugars.

6.1.1 Pyranose–Furanose Mutases

A UGM homolog encoded by the *cj1439c* gene had been identified in *C. jejuni* 11168, an organism possessing no D-Gal*f* glycoconjugates.[5] However, this strain includes a 2-acetamido-2-deoxy-D-galactofuranose (D-Gal*f*NAc) moiety in its capsular polysaccharide repeat structure.[6] The enzyme encoded by this gene shares ~40–60% sequence identity to the bacterial UDP-galactopyranose mutase (UGM) enzymes, and had a proposed role in the isomerization of UDP-2-acetamido-2-deoxy-galactopyranose (UDP-Gal*p*NAc) to UDP-Gal*f*NAc. However, other bacterial UGM are unable to catalyze this transformation.[7] Using an *in vitro* HPLC assay and complementation studies we characterized the activity of this UGM homologue (Chapter 2). The enzyme, which we have renamed UDP-N-acetyl-galactopyranose mutase (UNGM), has relaxed specificity and will accept either UDP-D-Gal or UDP-D-GalNAc as substrates for the isomerization reaction. Complementing mutase knock-out strains of *C. jejuni* 11168 and *Escherichia coli* W3110, the latter containing Gal*f* residues in its lipopolysaccharide, with UNGM demonstrates that the enzyme also recognizes both substrates *in vivo*. Based on its primary amino acid sequence, we designed a homology model of UNGM, which, combined with site-directed mutagenesis experiments, led to the identification of two active site arginine residues that are involved in the recognition of the UDP-D-GalNAc substrate. We also examined the specificity of UNGM using a two-substrate co-incubation assay, which demonstrated, surprisingly, that UDP-D-Gal is a better substrate than UDP-D-GalNAc.

To further probe the binding specificity of this *C. jejuni* UNGM and related UGMs from *Escherichia coli* and *Klebsiella pneumoniae*, we synthesized a panel of UDP-D-galactofuranose (UDP-D-Gal*f*) derivatives and evaluated their activity as substrates for these enzymes (Chapter 3). To prepare these derivatives we used a chemo-enzymatic approach originally described by Field and coworkers,[8] and later optimized by Rose *et al*.[9] in which the pyrophosphate moieties of the UDP-D-Gal*f* derivatives were formed using a promiscuous galactopyranose nucleotidyltransferase (GalPUT). This method allowed us to prepare deoxygenated UDP-D-Gal*f* derivatives in moderate to high yield; however, the methoxy derivatives served as poor substrates for GalPUT giving low yields (<5%) of the corresponding methoxy UDP-D-Gal*f* products. Despite the low yield we were able to isolate sufficient amounts of the 6′′-methoxy, 5′′-methoxy, and 2′′-methoxy derivatives to allow for their evaluation with UNGM and the UGM of *E. coli* and *K. pneumoniae*.

The results of these studies, described in Chapter 3, show that the largest difference in the binding interactions between UNGM and the UGM enzymes occurs with the 3′′-OH group of the UDP-D-Gal*f*. These observations can be explained, at least in part, from the crystal structure of UNGM, which we obtained in collaboration with professor David A. R. Sanders (University of Saskatchewan). In the *E. coli* and *K. pneumoniae* UGM, the 3′′-OH of UDP-D-Gal*f* forms a water mediated hydrogen bond to a histidine residue (H59 and H63 respectively) in the enzyme active site.[10] In UNGM, the amino acid in this position is an arginine (R59), which would disrupt this hydrogen bonding

interaction. In addition to using the modified UDP-D-Gal*f* derivatives, substrate binding was also examined using saturation transfer difference nuclear magnetic resonance (STD-NMR) spectroscopy. The *C. jejuni* UNGM demonstrated lower STD effects for nearly all of the galactopyranose protons, when UDP-D-Gal*p* was used as the ligand, compared to the same effects for *E. coli* UGM. The opposite was seen when UDP-D-Gal*p*NAc was used as the ligand. These differences imply different binding interactions in the active site of these two enzymes; however, further studies are still required to establish the exact binding modes for these two substrates. In the future, complete relaxation and conformational exchange matrix analysis of saturation transfer (CORCEMA-ST) calculations[11, 12] based on these STD-NMR measurements could be used to model the binding of UDP-D-Gal*p* and UDP-D-Gal*p*NAc to these two enzymes. These studies will help to further explain the difference in substrate tolerance observed for UNGM and UGM enzymes.

In Chapter 4, we expanded our studies of bacterial pyranose–furanose mutase enzymes to include three putative enzymes (Glf1, Glf2, and Glf3) from *C. jejuni* serotype HS:41.[5] The CPS structure produced by this bacteria is composed entirely of sugars in the furanose ring form, including L-arabinofuranose (L-Ara*f*), 6-deoxy-L-altrofuranose (6d-L-Alt*f*), D-fucofuranose (D-Fuc*f*), and 6-deoxy-D-*altro*-heptofuranose (6d-D-*altro*-Hep*f*).[13] We expressed and purified these three enzymes as recombinant proteins in *E. coli*, all of which were found to be flavo-proteins similar to UGM. To test the activity and specificity of the recombinant enzymes, we first synthesized their putative

furanose sugar nucleotide products. Using a zinc-mediated allylation reaction[14, 15] followed by a direct glycosylation with GDP following the method described by Jakeman and coworkers,[16] we were able to prepare GDP-6d-*altro*-Hep*f* starting from D-arabinose. Using this substrate, we found that the *C. jejuni* HS41 Glf1 enzyme is able to function as a GDP-6-deoxy-*altro*-heptopyranose mutase (GaHM) enzyme *in vitro*, making it the first enzyme described for the biosynthesis of a heptofuranose sugar. GaHM, like the UGM enzymes, is a flavoprotein requiring a reduced FAD cofactor for activity. Future work will be required to establish the structure and specificity of this enzyme.

Using synthetic UDP-L-Ara*f* we found that the Glf3 enzyme of *C. jejuni* HS:41 exhibits UDP-L-arabinopyranose mutase (UAM) activity and is able to interconvert UDP-L-arabinopyranose (UDP-L-Ara*p*) and UDP-L-Ara*f*. The enzyme is specific for UDP-L-Ara over structurally related UDP-D-Gal, which has the same stereochemistry as UDP-L-Ara with an additional hydroxylmethyl group at the non-reducing end of the monosaccharide. Using homology modeling and site-directed mutagenesis, we discovered that a tyrosine residue (Y84) in the Glf3 active site controls the selectivity for UDP-L-Ara over UDP-D-Gal. Mutating this tyrosine to an asparagine (present in the *K. pneumoniae* UGM) resulted in a reversal of the specificity to now favor UDP-D-Gal as the substrate. L-Ara*f* biosynthesis has also been described in plants, which use a flavin independent UAM to catalyze the isomerization of UDP-L-Ara*p* to UDP-L-Ara*f*.[17, 18] It appears, therefore, that *C. jejuni* HS:41,which uses flavin-dependent UAM, and plants, which employ a flavin-independent UAM, have

developed convergent methods to access L-Ara*f*. Thus, the Glf3 enzyme represents the first flavin dependent UAM involved in L-Ara*f* metabolism. It should be noted, however, that some UGMs have been shown to recognize, albeit weakly, UDP-L-Ara*f* as a substrate.

In Chapter 4, we also report the synthesis of TDP-6d-L-Alt*f*, a putative substrate for the Glf2 enzyme of *C. jejuni* HS:41. This sugar nucleotide was synthesized using a chemoenzymatic approach taking advantage of the broad substrate tolerance of the Cps2L thymidylyltransferase enzyme of *Streptococcus pneumoniae*.[19, 20] We then used the synthetic TDP-6d-L-Alt*f*, along with UDP-D-Fuc*f* (prepared in Chapter 3) to evaluate the activity and substrate tolerance of Glf2. However, neither of these sugar nucleotides served as an effective substrate for the enzyme. Thus, the substrate of Glf2 remains to be determined. It is possible that this enzyme uses a CDP donor, and future work will be focused on preparing and evaluating such donors. It is also possible that Glf2 is bifunctional and recognizes both 6d-L-Alt and D-Fuc substrate, which differ only in the stereochemistry at carbon-5, and future studies will be aimed at testing this hypothesis.

6.1.2 GlfT2

In addition to the pyranose–furanose mutase enzymes discussed above, we also examined the specificity of the galactofuranosyltransferase enzyme GlfT2. Mycobacterial cell wall galactan is composed of alternating β-(1→5) and β-(1→6) galactofuranosyl residues and is assembled by the action of two

bifunctional galactofuranosyltransferases, GlfT1 and GlfT2, both of which use UDP-D-Gal*f* as the donor substrate.[21-23] In Chapter 5, we report the use of synthetic UDP-D-Gal*f* derivatives (prepared in Chapter 3) to identify critical protein–carbohydrate binding interactions involved in donor substrate recognition by the processive polymerizing galactofuranosyltransferase GlfT2. GlfT2 showed little to no activity with the methoxy UDP-D-Gal*f* derivatives, showing that the donor substrate-binding pocket is sterically crowded and cannot tolerate the additional steric bulk of a methoxy group at any of the positions tested. Evaluation of GlfT2 activity with deoxy UDP-D-Gal*f* analogs and UDP-L-Ara*f* showed that the C-6 hydroxyl group of the donor substrate is not required for substrate activity. The 3′′-deoxy UDP-D-Gal*f* derivative showed greater then 100-fold reduced activity compared to the native UDP-D-Gal*f* substrate. This derivative also functioned as a moderate GlfT2 inhibitor, which shows that interactions with the 3′′-OH of UDP-D-Gal*f* are required for substrate turnover, but do not contribute significantly to substrate binding. We propose that these interactions with this hydroxyl group could serve to orient the substrate for turnover. This hypothesis is consistent with the recently reported structure of GlfT2.[21] A model of UDP-D-Gal*f* binding based on this structure shows that the 3′′-OH group only forms a hydrogen bond to the catalytic aspartate 372 residue and may serve to anchor the donor substrate in the correct orientation for turnover.

Moreover, when a C-5 or C-6 deoxy D-Gal*f* derivative or a L-Ara*f* residue was added to the growing galactan chain it resulted in the formation of “dead

end" reaction products. These products could no longer served as an acceptor substrate for the enzyme (Chapter 5). This finding shows that GlfT2 is specific for the formation of alternating β-(1→5) and β-(1→6)-glycosides, and is incapable of adding two subsequent D-Gal*f* residues with the same linkage (i.e., two β-(1→5)-linked Gal*f* in a row). This result also reveals that recognition of both the terminal C-5 and C-6 hydroxyl groups of the acceptor substrate are required for GlfT2 activity. This interaction is consistent with a recent model developed based upon a crystal structure of the enzyme.

A series of GlfT2 mutants were also prepared and evaluated to explore the importance of these amino acids in controlling the regioselectivity of the alternating β-(1→5) and β-(1→6) activities of GlfT2. Specifically, glutamic acid 300, tryptophan 399, and histidine 413, which are predicted to form hydrogen bonds to the acceptor substrate, were mutated to serine residues and the activity of the mutant GlfT2 enzymes were evaluated. Mutation of any of these residues resulted in a greater then 1000-fold decrease in GlfT2 activity, suggesting these residues play an important role in acceptor substrate binding.[21] However, all of the products isolated from reactions with these mutant enzymes showed the same linkages as the wild-type enzyme. Thus, none of these amino acids control the mechanism of alternating regioselectivity displayed by GlfT2. That being said, the amino acids tested represent only a fraction of the putative acceptor binding interactions present in GlfT2. Future work will focus on other amino acids involved in acceptor binding such as H296 or Y344 that also appear to interact with the non-reacting terminal 5-OH or 6-OH group of the acceptor substrate.

Understanding these specific protein–carbohydrate interactions involved in substrate recognition by GlfT2 may facilitate the design of future inhibitors of mycobacterial cell wall biosynthesis that target this key enzyme.

6.2 Closing Remarks

The work described in this thesis has shown the generality of flavin dependent bacterial pyranose–furanose mutase enzymes for the biosynthesis of a variety of structurally diverse furanose sugar nucleotides, and provides a strategy to deduce the structural motifs responsible for the specificity of these enzymes. Because of the mechanistic, and structural, similarity between these enzymes, which each recognize different sugar nucleotide substrates, they provide an ideal system in which to identify specific structural motifs responsible for substrate recognition and specificity. In at least two of the enzymes studied, we identified relatively subtle amino acid substitutions in the active site that impact substrate specificity. In addition to those described in this thesis, there are a number of other furanose sugars found in bacteria where the biosynthetic pathway remains unknown. As an example, paratofuranose, found in the cell wall of Yesinia pseudotuberculosis,[22] is proposed to require a flavin dependent pyranose–furanose mutase enzyme for its biosynthesis.[23] This still needs to be experimentally validated, which will be the focus of future work. The work described here, combined with work to study the binding interactions in other furanosyltransferase enzymes, will allow for the identification specific binding interaction that can help guide the design of specific inhibitors to target bacterial

furanoside biosynthesis. These Inhibitors could then be used in the treatment of diseases such as tuberculosis.

6.3 Bibliography

1. Pedersen, L. L.; Turco, S. J., *Cell. Mol. Life Sci.* **2003,** *60*, 259-266.
2. Pan, F.; Jackson, M.; Ma, Y. F.; McNeil, M., *J. Bacteriol.* **2001,** *183*, 3991-3998.
3. Kleczka, B.; Lamerz, A. C.; van Zandbergen, G.; Wenzel, A.; Gerardy-Schahn, R.; Wiese, M.; Routier, F. H., *J. Biol. Chem.* **2007,** *282*, 10498-10505.
4. Schmalhorst, P. S.; Krappmann, S.; Vervecken, W.; Rohde, M.; Muller, M.; Braus, G. H.; Contreras, R.; Braun, A.; Bakker, H.; Routier, F. H., *Eukaryotic Cell* **2008,** *7*, 1268-1277.
5. Karlyshev, A. V.; Champion, O. L.; Churcher, C.; Brisson, J. R.; Jarrell, H. C.; Gilbert, M.; Brochu, D.; St Michael, F.; Li, J. J.; Wakarchuk, W. W.; Goodhead, I.; Sanders, M.; Stevens, K.; White, B.; Parkhill, J.; Wren, B. W.; Szymanski, C. M., *Mol. Microbiol.* **2005,** *55*, 90-103.
6. St Michael, F.; Szymanski, C. M.; Li, J. J.; Chan, K. H.; Khieu, N. H.; Larocque, S.; Wakarchuk, W. W.; Brisson, J. R.; Monteiro, M. A., *Eur. J. Biochem.* **2002,** *269*, 5119-5136.
7. Errey, J. C.; Mann, M. C.; Fairhurst, S. A.; Hill, L.; McNeil, M. R.; Naismith, J. H.; Percy, J. M.; Whitfield, C.; Field, R. A., *Org. Biomol. Chem.* **2009,** *7*, 1009-1016.

8. Errey, J. C.; Mukhopadhyay, B.; Kartha, K. P. R.; Field, R. A., *Chem. Commun.* **2004**, 2706-2707.

9. Rose, N. L.; Zheng, R. B.; Pearcey, J.; Zhou, R.; Completo, G. C.; Lowary, T. L., *Carbohydr. Res.* **2008,** *343*, 2130-2139.

10. Gruber, T. D.; Westler, W. M.; Kiessling, L. L.; Forest, K. T., *Biochemistry* **2009,** *48*, 9171-9173.

11. Yuan, Y.; Wen, X.; Sanders, D. A. R.; Pinto, B. M., *Biochemistry* **2005,** *44*, 14080-14089.

12. Yuan, Y.; Bleile, D. W.; Wen, X.; Sanders, D. A. R.; Itoh, K.; Liu, H. W.; Pinto, B. M., *J. Am. Chem. Soc.* **2008,** *130*, 3157-3168.

13. Hanniffy, O. M.; Shashkov, A. S.; Moran, A. P.; Prendergast, M. M.; Senchenkova, S. N.; Knirel, Y. A.; Savage, A. V., *Carbohydr. Res.* **1999,** *319*, 124-132.

14. Pakulski, Z.; Zamojski, A., *Tetrahedron* **1997,** *53*, 2653-2666.

15. Ramana, C. V.; Narute, S. B.; Gonnade, R. G.; Patil, R. S., *Synthesis-Stuttgart* **2008**, 1783-1787.

16. Timmons, S. C.; Jakeman, D. L., *Org. Lett.* **2007,** *9*, 1227-1230.

17. Konishi, T.; Takeda, T.; Miyazaki, Y.; Ohnishi-Kameyama, M.; Hayashi, T.; O'Neill, M. A.; Ishii, T., *Glycobiology* **2007,** *17*, 345-354.

18. Konishi, T.; Ohnishi-Kameyama, M.; Funane, K.; Miyazaki, Y.; Konishi, T.; Ishii, T., *Carbohydr. Res.* **2010,** *345*, 787-791.

19. Timmons, S. C.; Hui, J. P. M.; Pearson, J. L.; Peltier, P.; Daniellou, R.; Nugier-Chauvin, C.; Soo, E. C.; Syvitski, R. T.; Ferrieres, V.; Jakeman, D. L., *Org. Lett.* **2008,** *10*, 161-163.

20. Timmons, S. C.; Mosher, R. H.; Knowles, S. A.; Jakeman, D. L., *Org. Lett.* **2007,** *9*, 857-860.

21. Wheatley, R. W.; Zheng, R. B.; Richards, M. R.; Lowary, T. L.; Ng, K. K. S., *J. Biol. Chem.* **2012**.

22. Kondakova, A. N.; Shaikhutdinova, R. Z.; Ivanov, S. A.; Dentovskaya, S. V.; Shashkov, A. S.; Anisimov, A. P.; Knirel, Y. A., *Carbohydr. Res.* **2009,** *344*, 2421-2423.

23. Cunneen, M. M.; De Castro, C.; Kenyon, J.; Parrilli, M.; Reeves, P. R.; Molinaro, A.; Holst, O.; Skurnik, M., *Carbohydr. Res.* **2009,** *344*, 1533-1540.

Appendix A

Supporting Information for Chapter 2

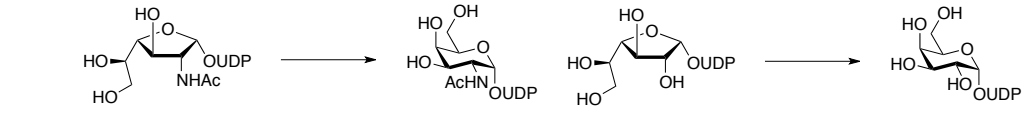
A.1 Sequences for the primers used in Chapter 2

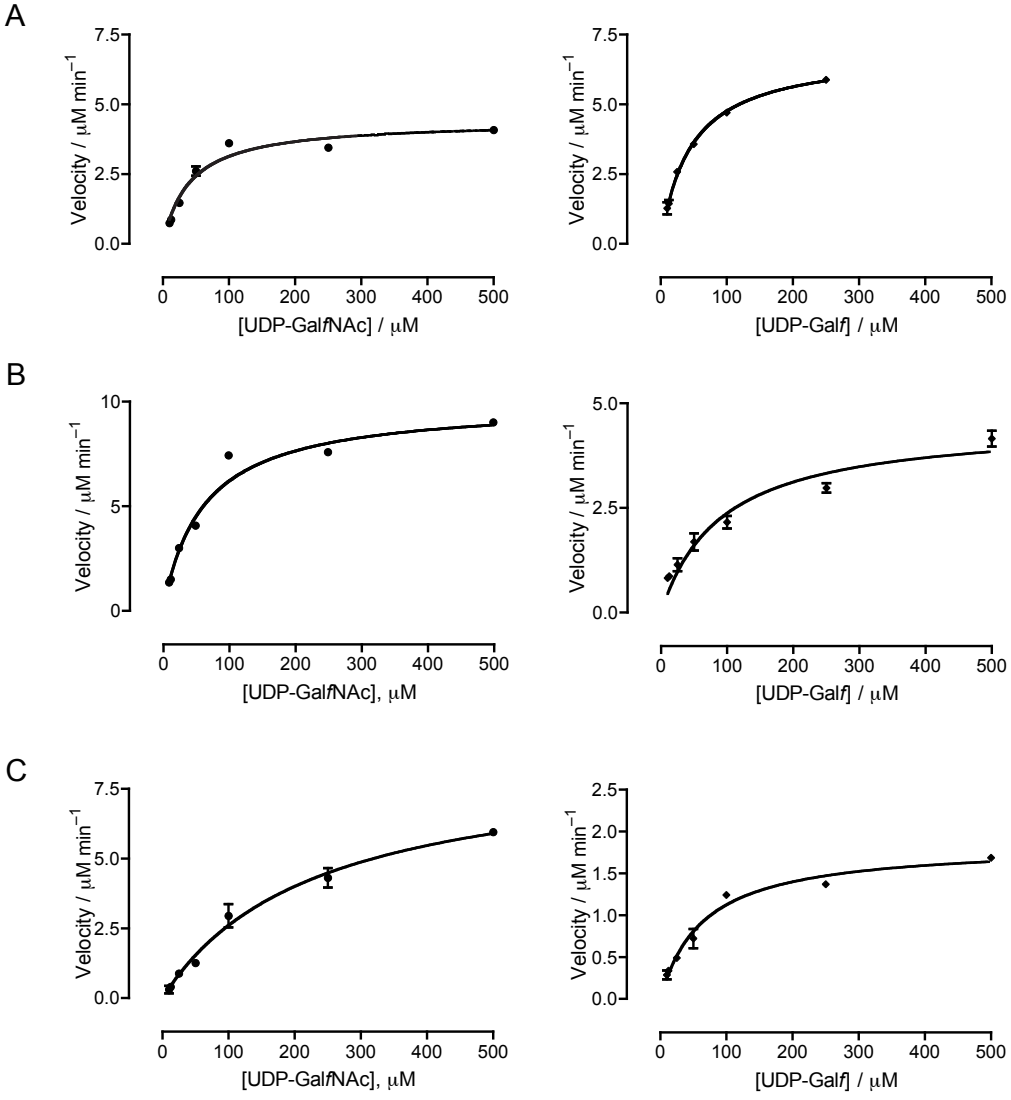
The primers used in Chapter 2 are described in Table A-1. Restriction enzyme sites are underlined and mutation sequences are shown in bold.

Table A-1. Primer sequences

Name	Sequence (5'-3')
CS-261	TTGAATTCTAATTAGGTAATGTGAGGAATGTGATTCATG (*Eco*RI)
CS-262	ATTTGTACATCAGTGGTGGTGGTGGTGGTGTCCTTGCTTATTTTTTAACTC (*Bsr*GI)
CS-362	AAGAATTCAGGTGTGAGGAATGTGATTCATGTACGATTATATCATTGTTG (*Eco*RI)
CS-363	AAATGTACATCAGTGGTGGTGGTGGTGGTGATCCGTACTCATTATATTTTTCAC
CS-372	TATAATTTTCATATGTACGATTATATCATTGTTGGTTCTGG (*Nde*I)
CS-373	TTTCTCGAGATCCGTACTCATTATATTTTTCACTTGATAAAGAGCG (*Xho*I)
R59H_F	CAATGTGCATAAATATGGTGCTCATATTTTT**CAT**ACTTCAGATCAGAAT-ATATGGG
R59H_R	CATATAATCCCATATATTCTGATCTGAAGT**ATG**AAAAATATGAGCACCA-TATTTATGCACATTGATATTTTTAATATTTTC
R168K_F	TGTAAGGATTTGCCTGCTTCTATTAT**AAA**AAGATTGCCTGTTCG
R168K_R	CGAACAGGCAATCTT**TTT**ATAATAGAAGCAGGCAAATCCTTACA

A.2 Kinetic plots for cjUGM and mutants





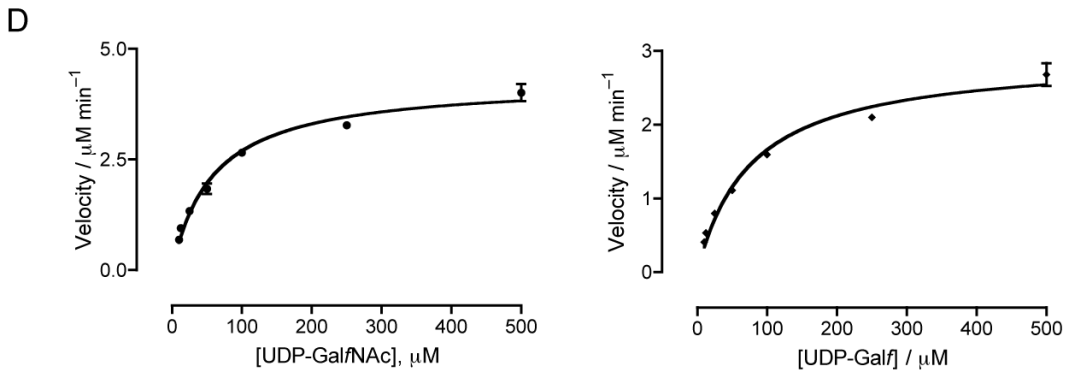


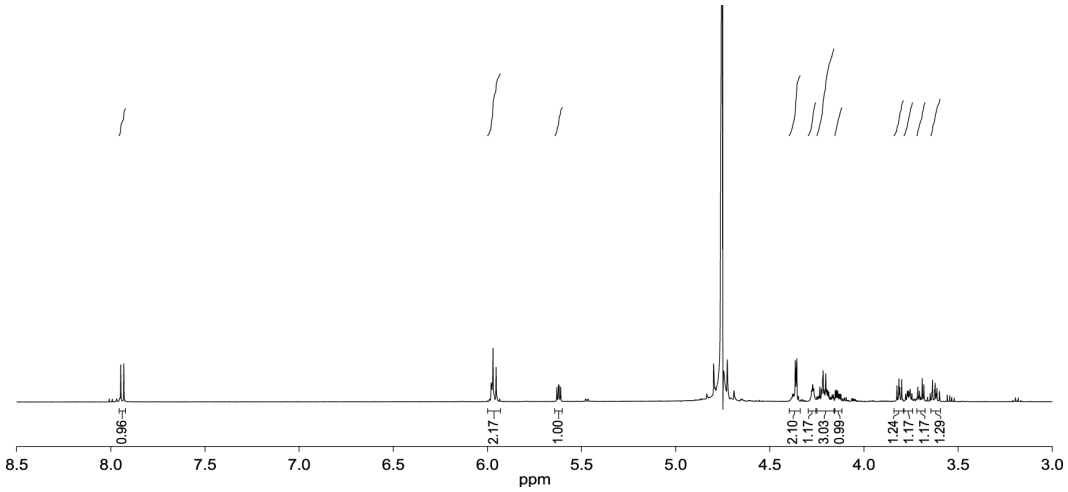
Figure A-1. Kinetic analysis of wild-type cjUNGM (A), R59H mutant cjUNGM (B), R169K mutant cjUNGM (C), and R59H/R169K double mutant cjUNGM (D) with UDP-Gal*f*NAc and UDP-Gal*f*. Curves were fit to the Michaelis-Menten equation using GraphPad PRISM 4 (GraphPad Software, San Diego, CA).

Appendix B

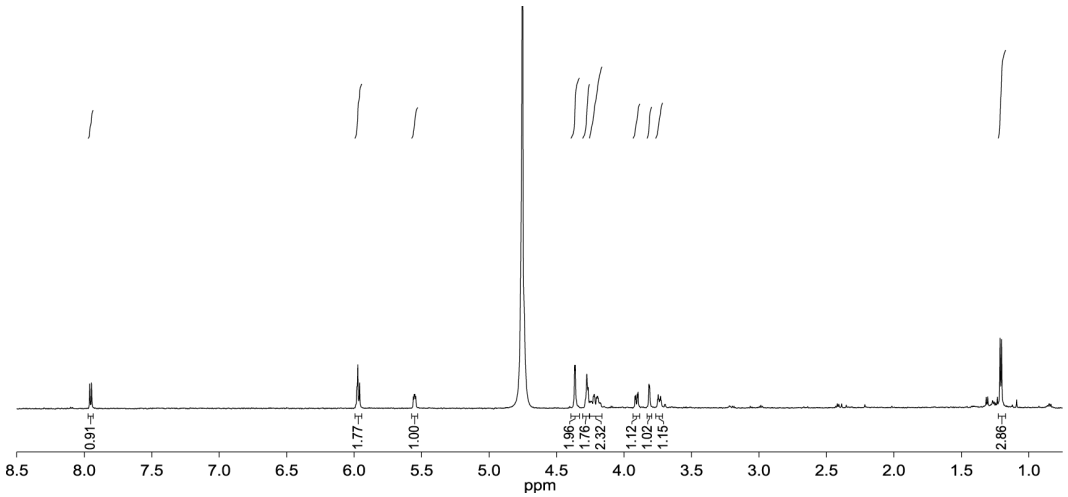
Supporting Information for Chapter 3

B.1 ^{1}H NMR of UGM/UNGM reaction products

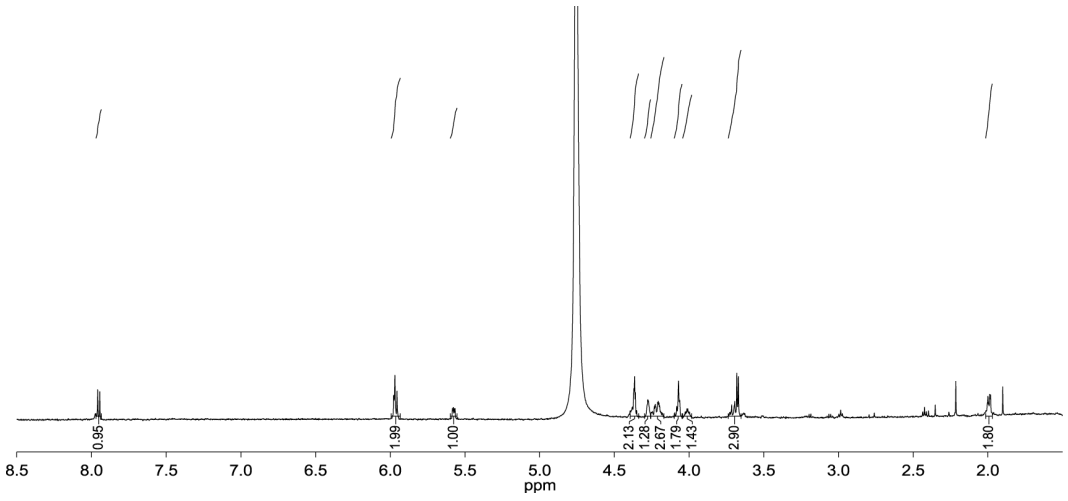
*B.1.1 UDP-Gal*p



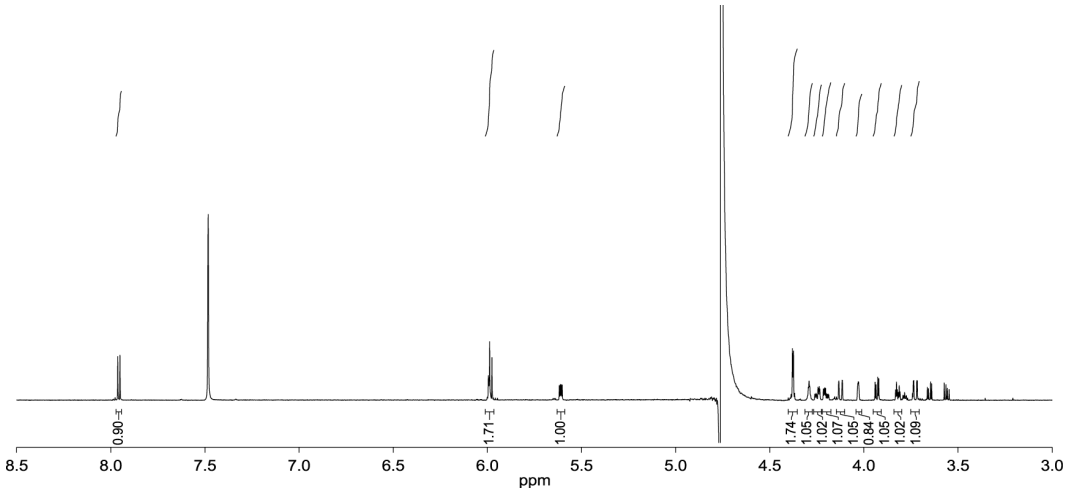
*B.1.2 UDP-6d-Gal*p



*B.1.3 UDP-3d-Gal*p



*B.1.4 UDP-L-Ara*p



Appendix C

Supporting Information for Chapter 4

C.1 Protein Characterization

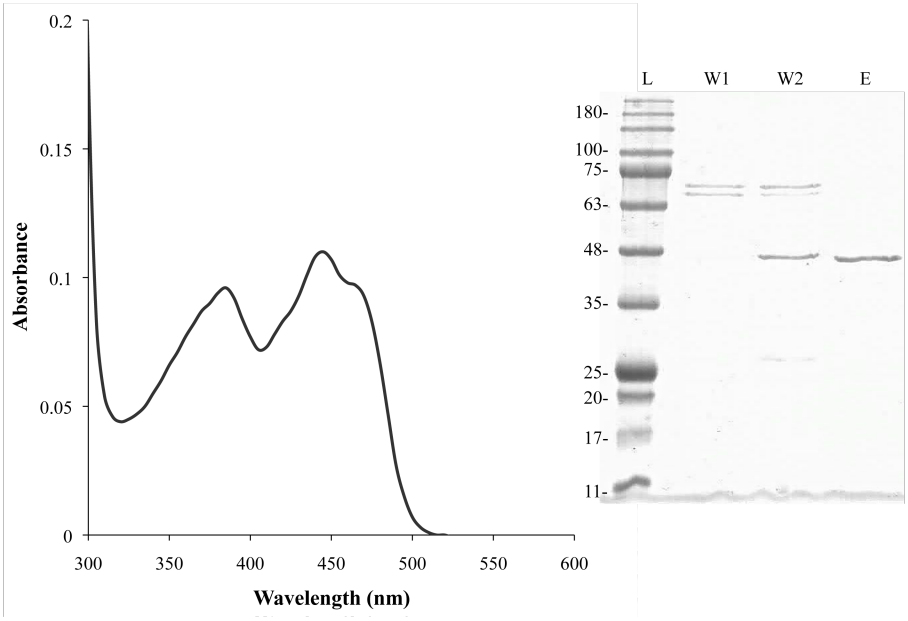


Figure C-1. UV-visible absorbance spectrum of Glf1 showing the characteristic flavin co-factor absorbances. SDS-PAGE of the purified protein is also shown. L = ladder, W1 = Wash1, W2 = Wash 2, E = Elute.

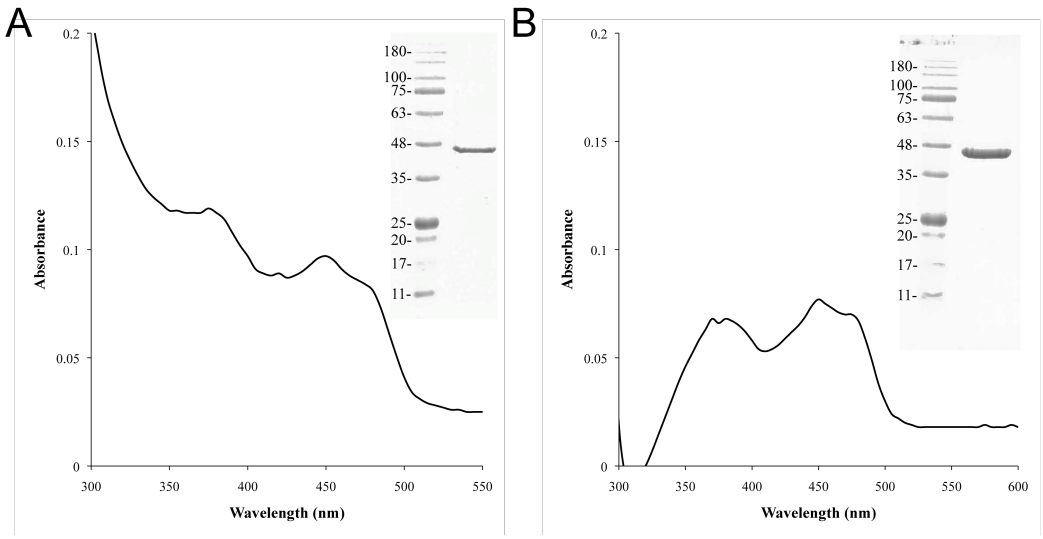
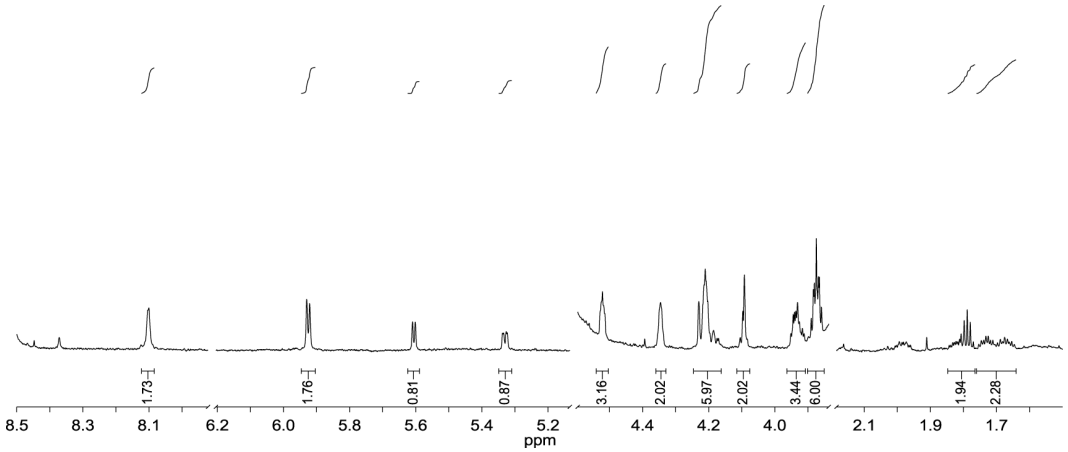


Figure C-2. UV-visible absorbance spectrum and SDS-PAGE evaluation of Glf3 (A) and Y84N Glf3 mutant (B) showing the characteristic flavin co-factor absorbances.

C.2 ^{1}H analysis of Glf1 reaction



C.3 Glf3 Kinetic analysis

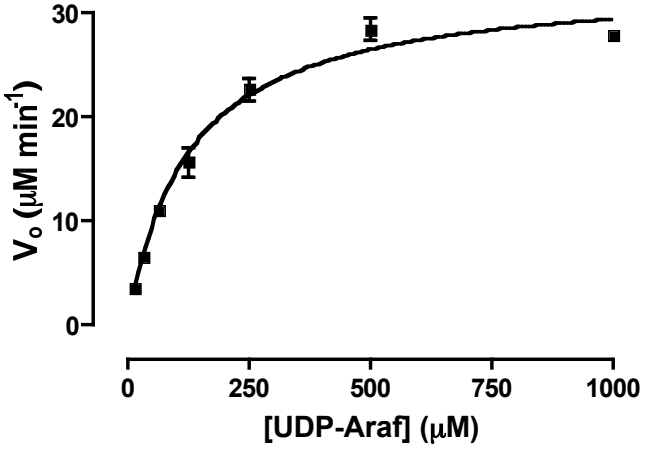
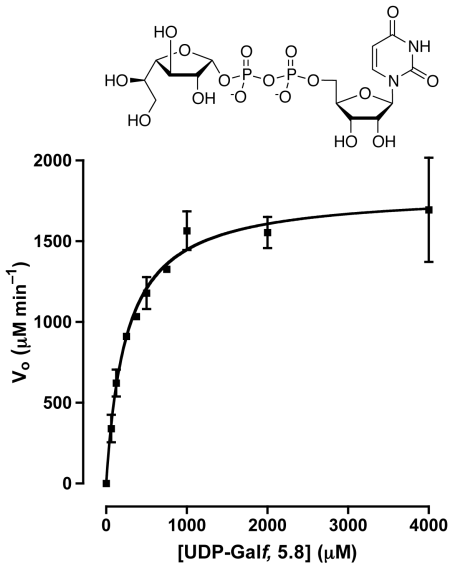


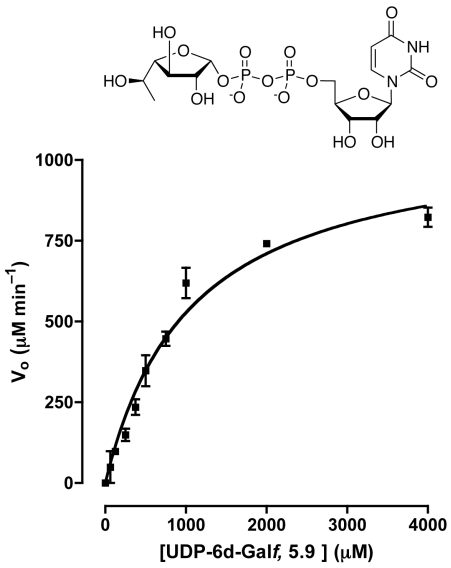
Figure C-3. Glf3 kinetic analysis with UDP-L-Ara*f* as the substrate.

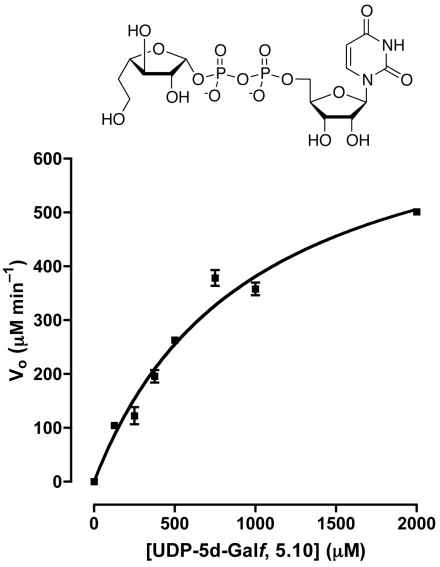
Appendix D

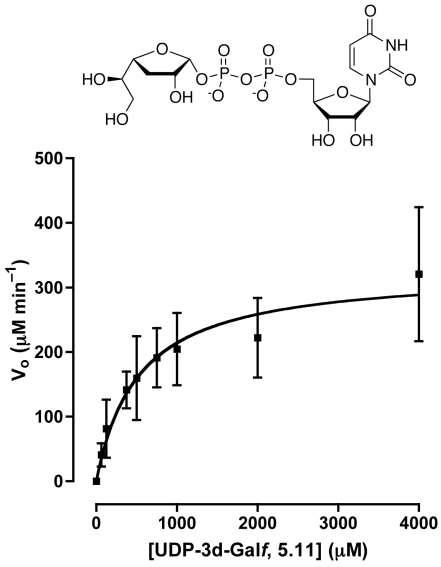
Supporting Information for Chapter 5

D.1 GlfT2 donor kinetic with UDP-Gal*f* analogs









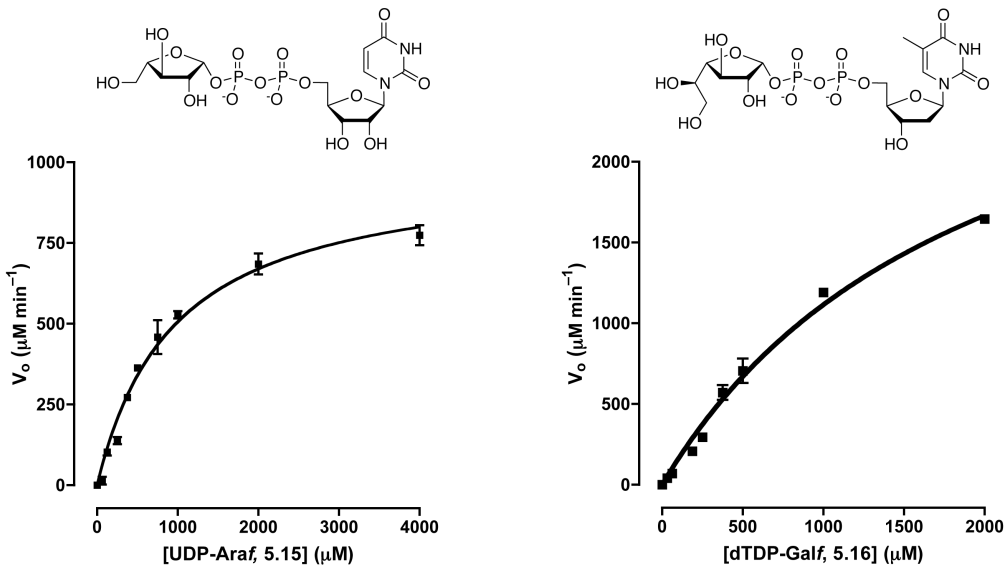


Figure D-1. Kinetic analysis of GlfT2 with UDP-Gal*f* analogs and acceptor **5.6**. Curves were fit to the Michaelis-Menten equation using GraphPad PRISM 4 (GraphPad Software, San Diego, CA).

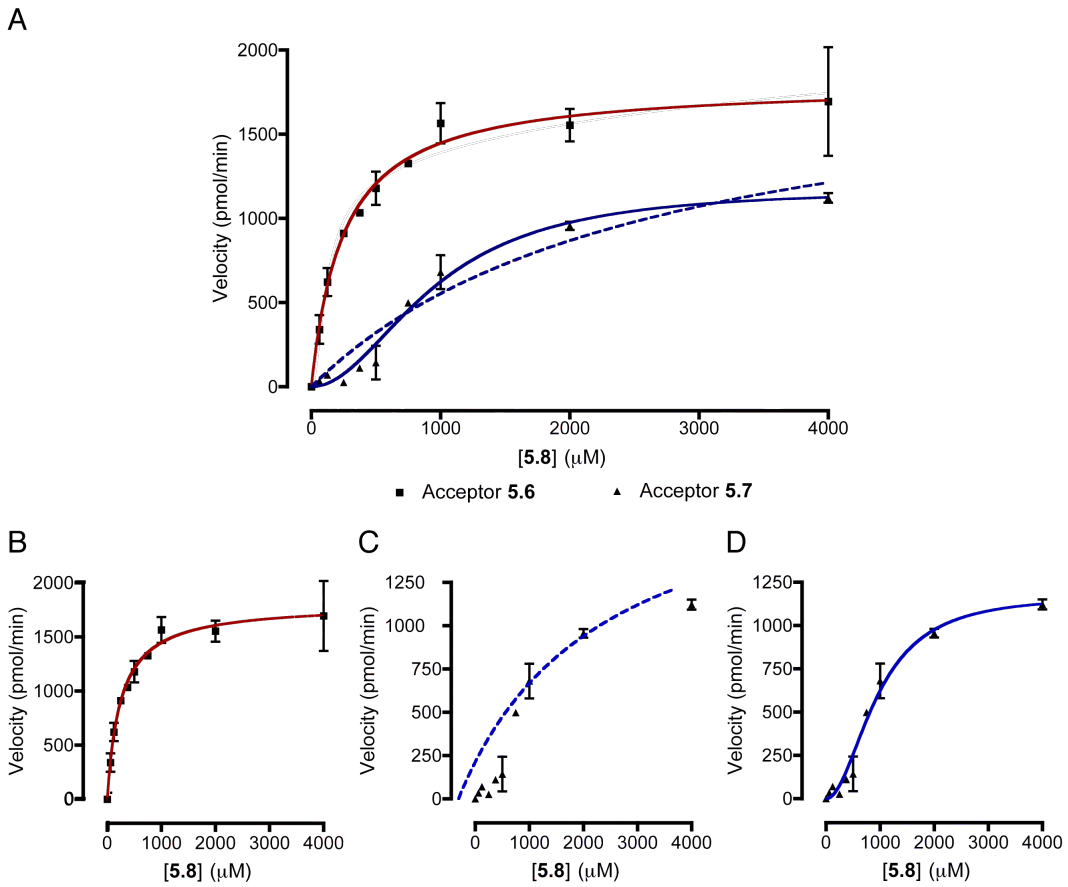


Figure D-2. GlfT2 donor kinetics data with UDP-Gal*f* and acceptor **5.6** or **5.7** (A). The donor kinetics data for acceptor **5.6** (B) and acceptor **5.7** (C) is shown fit to the Michealis–Menten equation (r^2 = 0.95 and 0.91, respectively). A better fit (r^2 = 0.96) was found with the Michealis–Menten equation containing a Hill-slope factor in the case of acceptor **5.7** (D). The Michealis–Menten equation containing a Hill-slope is provided below:

$$V_o = V_{max}*[S]^n/(K_M^2 + [S]^n)$$

Where V_o is the initial velocity, V_{max} is the maximum velocity, K_M is the Michaelis constant, [S] is the concentration of substrate and ^{n} is the hill slope value. In this case ^{n} = 1.97.