A Chemoenzymatic Route to Unnatural Sugar Nucleotides and Their Applications
&
Enzymatic Synthesis of Rare Sugars with Aldolases *In vitro* and *In vivo*

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy
in the Graduate School of The Ohio State University

By

Li Cai, M.S.

Graduate Program in Chemistry

The Ohio State University

2011

Dissertation Committee:

Dr. Peng George Wang, Advisor

Dr. David J. Hart

Dr. Craig J. Forsyth

Dr. Zhongfa Liu
Copyright by

Li Cai

2011
CARBOHYDRATE represents a class of biomacromolecules which is quite different from polynucleotide and polypeptide, given its branching character and non-template based biosynthesis. Carbohydrates exist in cells or on cell surfaces mainly as glycoconjugates: conjugation of a sugar chain (glycan) with a protein (glycoprotein), a lipid (glycolipid) or both lipid and protein. Cell surface carbohydrates are of paramount significance in biochemical recognition processes and carbohydrate complexes have been widely used in pharmaceutical areas such as prevention of infection, neutralization of toxins and cancer immunotherapy.

It is well known that nucleic acids are biosynthesized using dNTP (or NTP) and proteins using activated amino acids. Our focus on this research (Chapters 1-4) is to find a facile way to produce the key building blocks for enzymatic carbohydrate synthesis—sugar nucleotide donors. We discovered a pathway in which N-acetylglucosamine/N-acetylgalactosamine (GlcNAc/GalNAc) and their analogues were phosphorylated by a 1-kinase called NahK to generate sugar-1-phosphates in good yields. These sugar-1-phosphates were subsequently pyrophosphorylated by GlmU from bacteria or AGX1 from mammals to give the corresponding sugar nucleotide donor molecules. We thus obtained a library of UDP-GlcNAc/GalNAc analogues on a preparative scale since these involved enzymes proved to be rather promiscuous. With the modified sugar nucleotide
donors in hand, we are primed to incorporate the structurally modified GlcNAc/GalNAc analogues into different carbohydrate-containing biomolecules to either investigate carbohydrate metabolic pathways or construct a glycorandomized library of new glycoforms.

In addition, the high stereoselectivity of aldolases in C-C bond construction confers upon them tremendous applications as synthetic biocatalysts. Among the aldolases, dihydroxyacetone phosphate (DHAP)-dependent aldolases are particularly attractive as a set of four possible diastereomers of vicinal diols can be synthesized conferring upon them the potential to be used in the synthesis of rare sugars and other hydroxylated natural products. Unfortunately, the strict requirement for the donor substrate DHAP, a rather expensive and unstable compound, limits aldolase use in large-scale preparation. Therefore, the capability to generate DHAP from inexpensive sources could ultimately broaden the scope of aldolase reactions making it an attractive challenge. In Chapter 5, we employed DHAP-dependent aldolases in the synthesis of several rare sugars via a one-pot four enzyme reaction system in which DHAP is generated from the oxidation of cheap DL-glycerol 3-phosphate. Subsequently, the DHAP generated *in situ* is coupled with glyceraldehydes by aldolases to give different rare sugars. More significantly, we used engineered bacteria as ‘chemists’ to carry out the aldolase reactions *in vivo*. The bacteria were fed with green starting materials (*e.g.* glycerol or glucose) and DHAP is generated *in vivo* via bacterial glycolysis. Desired rare sugars could thus be produced by simple fermentation and isolated from the culture media.
DEDICATION

Dedicated to my father, Shengli Cai, for his loyal support and generous love.
ACKNOWLEDGMENTS

First and foremost I would like to acknowledge my research advisor Dr. Peng George Wang, who is always enthusiastic, energetic and full of sparkling ideas. He keeps training me as a ‘real professor’ and helps me develop a strong scientific independence.

I am also grateful to all my teachers during these four years, especially Dr. David J. Hart, who is the best organic teacher I have ever met and kindly served as a committee member of my first year oral exam, my candidacy exam and my thesis defense.

I also owe my deepest gratitude to the people I work with—the Wang group, a family of talented and kindhearted people from all over the world. Wanyi Guan, Zijie Li, Weiqing Han, Nick Pettit are my collaborators and this thesis would not have been possible without their help. Dr. Wenlan Chen (Alex), my best friend in the lab, has provided good argument about organic theory and reactions. Fishing, nice food, occasional secret smoking, lots of happy memories with him are truly treasure of my life.

I also want to give my thanks to my lovely wife, Liang Zha, M.D., who is always by my side, encourages me and supports me. Finally, I thank my parents and the rest of my family, for their loyal support and generous love throughout my life.
VITA

2005.................................................................................B.S., Pharmaceutical Sciences
.................................................................Peking University, China

2007.......................................................................................M.S., Medicinal Chemistry
Peking University, China

2008 to 2009 .................................................................Teaching Assistant
The Ohio State University

2009 to present ...............................................................Research Assistant
The Ohio State University

PUBLICATIONS

Research Publications


1: Equal contribution to this paper

FIELDS OF STUDY

Major Field: Chemistry
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Abstract</strong></td>
</tr>
<tr>
<td><strong>Dedication</strong></td>
</tr>
<tr>
<td><strong>Acknowledgments</strong></td>
</tr>
<tr>
<td><strong>Vita</strong></td>
</tr>
<tr>
<td><strong>List of Schemes</strong></td>
</tr>
<tr>
<td><strong>List of Figures</strong></td>
</tr>
<tr>
<td><strong>List of Tables</strong></td>
</tr>
<tr>
<td><strong>List of Abbreviations</strong></td>
</tr>
<tr>
<td><strong>Chapter 1: Introduction to enzymatic synthesis of carbohydrates</strong></td>
</tr>
<tr>
<td><strong>Chapter 2: A chemoenzymatic route to UDP-GlcNAc/GalNAc library</strong></td>
</tr>
<tr>
<td><strong>Chapter 3: Chemoenzymatic synthesis of UDP-2-ketoGlc as C2-carbon isostere of UDP-GlcNAc and the applications</strong></td>
</tr>
<tr>
<td><strong>Chapter 4: Enzymatic synthesis and properties of UDP-(β-thio)-GlcNAc/GalNAc</strong></td>
</tr>
<tr>
<td><strong>Chapter 5: Synthesis of rare sugars by in vitro and in vivo aldolase reactions</strong></td>
</tr>
<tr>
<td><strong>Bibliography</strong></td>
</tr>
<tr>
<td><strong>Appendix: 1H NMR, 13C NMR and 31P NMR Spectra</strong></td>
</tr>
</tbody>
</table>
# LIST OF SCHEMES

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scheme 1.1</td>
<td>Enzymatic glycosylation with glycosyltransferase</td>
<td>5</td>
</tr>
<tr>
<td>Scheme 1.2</td>
<td>Possible stereochemical outcomes of GT reactions</td>
<td>6</td>
</tr>
<tr>
<td>Scheme 1.3</td>
<td>Common/primary sugar nucleotides</td>
<td>7</td>
</tr>
<tr>
<td>Scheme 1.4</td>
<td>General sugar nucleotides biosynthesis</td>
<td>8</td>
</tr>
<tr>
<td>Scheme 1.5</td>
<td>Biosynthetic pathways of NDP-sugars</td>
<td>9</td>
</tr>
<tr>
<td>Scheme 1.6</td>
<td>Biosynthesis of CMP-sialic acid</td>
<td>11</td>
</tr>
<tr>
<td>Scheme 1.7</td>
<td>General synthetic strategy for NDP-sugars</td>
<td>15</td>
</tr>
<tr>
<td>Scheme 1.8</td>
<td>General synthetic strategy for α-sugar-1-phosphate</td>
<td>17</td>
</tr>
<tr>
<td>Scheme 2.1</td>
<td>Biosynthesis pathways of UDP-GlcNAc/GalNAc</td>
<td>25</td>
</tr>
<tr>
<td>Scheme 2.2</td>
<td>Chemical synthesis of UDP-GlcNAc</td>
<td>26</td>
</tr>
<tr>
<td>Scheme 2.3</td>
<td>Straightforward synthesis of UDP-GlcNAc</td>
<td>27</td>
</tr>
<tr>
<td>Scheme 2.4</td>
<td>Synthesis of acyl chain modified GlcNAc/GalNAc</td>
<td>31</td>
</tr>
<tr>
<td>Scheme 2.5</td>
<td>Synthesis of 2-azido-glucose/galactose</td>
<td>32</td>
</tr>
<tr>
<td>Scheme 2.6</td>
<td>Synthesis of N-acetyllallosamine</td>
<td>33</td>
</tr>
<tr>
<td>Scheme 2.7</td>
<td>Synthesis of C-4 modified GlcNAc/GalNAc</td>
<td>34</td>
</tr>
<tr>
<td>Scheme 2.8</td>
<td>Isothiocyanate formation using NaH, CS₂ and MeI reagents</td>
<td>35</td>
</tr>
<tr>
<td>Scheme 2.9</td>
<td>Synthesis of 6-modified GlcNAc analogues</td>
<td>36</td>
</tr>
</tbody>
</table>
Scheme 2.10  Synthesis of 6-modified GalNAc derivatives route I............ 37
Scheme 2.11  Synthesis of 6-modified GalNAc derivatives route II ......... 38
Scheme 2.12  Enzymatic synthesis of GlcNAc/GalNAc-1-P derivatives .... 40
Scheme 2.13  Biosynthesis of UDP-GlcNAc with bifunctional GlmU....... 47
Scheme 2.14  Enzymatic synthesis of UDP-sugar derivatives with GlmU... 49
Scheme 3.1   Retrosynthetic analysis of 2-ketoGlc..................................132
Scheme 3.2   Iodoacetyoxylation of glucal triacetate ...........................133
Scheme 3.3   Improved synthesis of β-gluco type 2-iodo sugar..............134
Scheme 3.4   Continued synthesis of 2-ketoGlc.....................................135
Scheme 3.5   Enzymatic synthesis of UDP-2-ketoGlc ..............................136
Scheme 4.1   Enzymatic synthesis of UDP(βS)-GlcNAc/GalNAc .............160
Scheme 5.1   Two types of aldolase mechanisms ..................................177
Scheme 5.2   Iminocyclitols and polyhydroxylated pyrrolidines .............179
Scheme 5.3   Preparation of DHAP and DHAP degradation ....................180
Scheme 5.4   Synthesis of L-fructose with RhaD ....................................183
Scheme 5.5   Synthesis of D-sorbose and D-psicose with RhaD ..............183
Scheme 5.6   One-pot four enzyme synthesis with RhaD .......................186
Scheme 5.7   One-pot four enzyme synthesis with FucA (I).................189
Scheme 5.8   One-pot four enzyme synthesis with FucA (II) .................191
Scheme 5.9   One-pot four enzyme synthesis of carbohydrates with FruA 195
Scheme 5.10  Production of DHAP from cheap green carbon source.......198
Scheme 5.11  In vivo production of rare sugars using engineered E.coli..... 199
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Cell surface glycoconjugates ........................................ 2</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Prevalence of GlcNAc and GalNAc ................................. 23</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>NahK SDS-PAGE and catalyzed reaction ......................... 28</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Modification of GlcNAc and bioorthogonal chemistry .......... 29</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>$^1$H NMR spectrum of AllNAc........................................ 33</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Structures of GlcNAc/GalNAc-1-P derivatives ................. 46</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>GlmU crystal structures and SDS-PAGE......................... 48</td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>Prokaryotic and eukaryotic UDP-GlcNAc pathway ............ 55</td>
</tr>
<tr>
<td>Figure 2.8</td>
<td>AGX1 SDS-PAGE and catalyzed reaction ....................... 56</td>
</tr>
<tr>
<td>Figure 2.9</td>
<td>Donor specificity investigation of GT by SAMDI-TOF......... 63</td>
</tr>
<tr>
<td>Figure 2.10</td>
<td><em>In vitro</em> and <em>in vivo</em> glycodiversification .............. 65</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Metabolic carbohydrate engineering with 2-ketoGal .......... 130</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>O-GlcNAc glycoprotein detections using UDP-2-ketoGal ...... 131</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>$^1$H NMR of peracetylated 2-keto glucose .................... 135</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Metabolic incorporation of ketone isotope on <em>E.coli</em> surface 137</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Flow cytometry analysis of bacterial surface labeling ........ 140</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>Fluorescent microscopy analysis of bacterial labeling ....... 141</td>
</tr>
</tbody>
</table>
Figure 3.7  LPS core oligosaccharide and SDS-PAGE............................... 142
Figure 4.1  Structures of some typical phosphorothionate analogues ...... 158
Figure 4.2  Glycosylation using sugar nucleotide thio analogue.......... 159
Figure 4.3  $^{31}$P NMR comparison of the thio containing analogues...... 161
Figure 4.4  LgtA mediated reaction using UDP(βS)-GlcNAc ............... 163
Figure 4.5  Effect of divalent ions and EDTA .................................. 163
Figure 4.6  In solution glycosylation using an azido modified lactose .... 164
Figure 4.7  Double reciprocal plots .................................................... 172
Figure 4.8  Secondary plots from the primary plot ............................. 173
Figure 5.1  Stereocomplementary sets of DHAP-dependent aldolases... 178
Figure 5.2  HPLC profile of RhaD reaction ......................................... 184
Figure 5.3  Separation profile of D-sorbose and D-psicose ................. 185
Figure 5.4  HPLC profile of FucA reaction using D-glyceraldehyde...... 190
Figure 5.5  HPLC profile of FucA reaction using L-glyceraldehyde ...... 193
Figure 5.6  Co-expression of RhaD and YqaB..................................... 200
Figure 5.7  HPLC profile of the fermentation process ......................... 201
Figure 5.8  HPLC Standard curves for quantification.......................... 203
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>Enzymatic synthesis of NDP-sugars</td>
<td>12</td>
</tr>
<tr>
<td>Table 2.1</td>
<td>Yields for C-2 modified GlcNAc-1-P analogues</td>
<td>41</td>
</tr>
<tr>
<td>Table 2.2</td>
<td>Yields for C-2 modified GalNAc-1-P analogues</td>
<td>43</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>Yields for C-3/C-4 modified sugar-1-P analogues</td>
<td>44</td>
</tr>
<tr>
<td>Table 2.4</td>
<td>Yields for C-6 modified GlcNAc/GalNAc-1-P analogues</td>
<td>45</td>
</tr>
<tr>
<td>Table 2.5</td>
<td>Yields for UDP-GlcNAc analogues by GlmU</td>
<td>51</td>
</tr>
<tr>
<td>Table 2.6</td>
<td>Yields for UDP-GalNAc analogues by GlmU</td>
<td>52</td>
</tr>
<tr>
<td>Table 2.7</td>
<td>Yields for UDP-sugar synthesis by AGX1 and GlmU</td>
<td>58</td>
</tr>
<tr>
<td>Table 2.8</td>
<td>Substrate specificity of AGX1 and GlmU</td>
<td>60</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Summary of kinetic data for LgtA catalyzed reaction</td>
<td>166</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Aldolases classified by their donor specificity</td>
<td>176</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Summary of RhaD catalyzed reactions</td>
<td>187</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>Summary of FucA catalyzed reactions</td>
<td>193</td>
</tr>
<tr>
<td>Table 5.4</td>
<td>One-pot four enzyme synthesis of carbohydrates with FruA</td>
<td>196</td>
</tr>
<tr>
<td>Table 5.5</td>
<td>Summary of results for in vivo production of rare sugars</td>
<td>202</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>alpha</td>
</tr>
<tr>
<td>Ac</td>
<td>acetyl</td>
</tr>
<tr>
<td>br</td>
<td>broad (NMR)</td>
</tr>
<tr>
<td>β</td>
<td>beta</td>
</tr>
<tr>
<td>n-Bu</td>
<td>normal-butyl</td>
</tr>
<tr>
<td>t-Bu</td>
<td>tert-butyl</td>
</tr>
<tr>
<td>Bz</td>
<td>benzoyl</td>
</tr>
<tr>
<td>°C</td>
<td>degrees Celsius</td>
</tr>
<tr>
<td>calcd</td>
<td>calculated</td>
</tr>
<tr>
<td>δ</td>
<td>chemical shift in parts per million downfield from tetramethylsilane</td>
</tr>
<tr>
<td>d</td>
<td>doublet (spectra); day(s)</td>
</tr>
<tr>
<td>DMAP</td>
<td>4-(N,N-dimethylamino)pyridine</td>
</tr>
<tr>
<td>DMF</td>
<td>N,N-dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethylsulfoxide</td>
</tr>
<tr>
<td>eq.</td>
<td>equivalent</td>
</tr>
<tr>
<td>Et</td>
<td>ethyl</td>
</tr>
<tr>
<td>γ</td>
<td>gamma</td>
</tr>
</tbody>
</table>
g  gram(s)
GT  glycosyltransferase
h  hour(s)
J  coupling constant in Hz (NMR)
k  kilo
L  liter(s)
m  milli; multiplet (NMR)
μ  micro
M  moles per liter
Me  methyl
MHz  megahertz
min  minute(s)
mol  mole(s)
Ms  methanesulfonyl
MS  mass spectrometry; molecular sieves
m/z  mass to charge ratio (MS)
NMR  nuclear magnetic resonance
p  para
Ph  phenyl
ppm  parts per million
py  pyridine
q  quartet (NMR)
rt  room temperature
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>singlet (NMR)</td>
</tr>
<tr>
<td>SAMs</td>
<td>self-assembled monolayers</td>
</tr>
<tr>
<td>t</td>
<td>tertiary (tert)</td>
</tr>
<tr>
<td>t</td>
<td>triplet (NMR)</td>
</tr>
<tr>
<td>TBAF</td>
<td>tetrabutylammonium fluoride</td>
</tr>
<tr>
<td>TBDPS</td>
<td>t-butyldiphenylsilyl</td>
</tr>
<tr>
<td>Tf</td>
<td>trifluoromethanesulfonyl</td>
</tr>
<tr>
<td>THF</td>
<td>tetrahydrofuran</td>
</tr>
<tr>
<td>TLC</td>
<td>thin layer chromatography</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION TO ENZYMATIC SYNTHESIS OF CARBOHYDRATES

1.1 Introduction

Three major repeating biomacromolecules, nucleic acid, protein, and carbohydrate carry out most of the information transfer in living systems. Nucleic acid carries genetic information in the form of DNA and RNA; protein acts as molecular machines or bio-catalysts of reactions in living cells. Carbohydrate represents a class of biopolymers which is different from polynucleotide and polypeptide because of its branching character and most significantly, its non-template based biosynthetic pathway.

It is well known that all cells are coated with a network of polysaccharides that project from cellular surfaces, called ‘glycocalyx’. These polysaccharides or oligosaccharides are usually parts of glycoconjugates: conjugation of a sugar chain (glycan) with a protein (glycoprotein), a lipid (glycolipid) or both lipid and protein (e.g. glycosylphosphatidylinositol (GPI)-anchored proteins).\(^1\)\(^-\)\(^2\) (Figure 1.1)

Cell surface carbohydrates (as shown in Figure 1.1) are of paramount significance in biochemical recognition processes such as cellular adhesion, cell differentiation, viral infections, immune responses and numerous signal transduction events.\(^1\)\(^,\)\(^3\)\(^-\)\(^5\) Moreover,
carbohydrate complexes have been widely used in pharmaceutical areas such as prevention of infection, neutralization of toxins and cancer immunotherapy.

Figure 1.1. Cell surface glycoconjugates: glycoproteins, glycolipids and glycosylphosphatidylinositol (GPI)-anchored proteins.

PCR (Polymerase chain reaction), a revolution technique developed in 1983 by Kary Mullis,\(^6\) has become an indispensable tool in biological and biochemical researches for nucleic acid (DNA) synthesis. The starting materials are deoxynucleoside
triposphates (dNTPs) and two primers. On the other hand, solid-phase peptide synthesis, pioneered by R. B. Merrifield,\(^7\) has made it possible for the preparation of desired peptides and proteins \textit{in vitro} in a synthetic manner. Protected amino acids or even unnatural ones can be used as building blocks. As for the carbohydrates which are highly branched molecules, how can they be assembled in a facile manner and what are the key biosynthetic building blocks (donors) are questions we are trying to answer in this Chapter.

\subsection*{1.2 Enzymatic approach to carbohydrates}

The difficulties in isolating and purifying structurally defined carbohydrates from natural sources have made it attractive to synthesize naturally existing carbohydrates as well as their derivatives. Methods for both chemical and enzymatic synthesis of carbohydrates have gained great advances in the past few years. Although chemical methods have made some oligosaccharides and glycoconjugates synthetically accessible\(^8\), it is still very difficult and costly to generate long-chain polysaccharides, glycolipids and glycoproteins chemically. Unlike proteins and DNA, even oligomers of carbohydrates come in a far greater complexity: the stereo-centers that constitute the ring in addition to the ring size, linkage position, branching, as well as posttranslational modifications such as sulfation, methylation, and phosphorylation.\(^9\)

Given that all chemical carbohydrate synthetic approaches are characterized by tedious time consuming protection and deprotection steps\(^10\) and poor stereoselectivities in glycosylation,\(^10\) enzymatic approach which has the reputation of high regio- and stereoselectivities seems to be an attractive alternative. Three classes of enzymes,
glycosidases\textsuperscript{11-12}, glycosynthases\textsuperscript{13-14} and glycosyltransferases are key enzymes associated with carbohydrate processing in cells and enzymatic carbohydrate synthesis.

Glycosidases catalyze oligosaccharides and polysaccharides hydrolysis. They can be used to form glycosidic bond under \textit{in vitro} conditions in which a hydroxyl group of another sugar acts as a nucleophile instead of water itself. However, use of glycosidase catalyzed oligosaccharide synthesis is restricted by low yields and poor regioselectivity.

Glycosynthases are actually a class of glycosidase mutants which can provide active sites with correct steric environment for the formation of a reactive glycosyl donor but no hydrolytic activity.\textsuperscript{15} Glycosyl fluorides are usually used as activated donor for glycosynthases. Yields are greatly improved for glycosynthases but difficulties involved in obtaining the enzymes are the major disadvantages for this strategy.

Glycosyltransferases are the most popular carbohydrates synthesis enzymes which can be divided into two categories, non-Leloir-type and Leloir-type based on their donor specificity. Both types can transfer the sugar moiety from a specific donor molecule to an acceptor. Non-Leloir-type glycosyltransferases use sugar-1-phosphate or oligosaccharides as donors.\textsuperscript{16-17} Since the reaction is normally reversible, synthesis of carbohydrates using non-Leloir-type glycosyltransferases is plagued with hydrolysis and polymerization reactions, which limit their synthetic potential.

Leloir-type glycosyltransferases\textsuperscript{18-20} catalyze the donation of a monosaccharide from its activated donor form (sugar nucleotides) to various acceptors (sugar chain, peptide, lipid) (Scheme 1.1); they are responsible for the synthesis of most glycoconjugates in cells. Leloir-type glycosyltransferases are able to delivery highly regiospecific and stereoselective glycosidic bonds. Although they are generally specific
to substrates, minor modifications on donor or acceptor structures can be tolerated\(^{21}\) which makes it possible to the synthesis of carbohydrates and glycoconjugates derivatives.

![Scheme 1.1. Enzymatic glycosylation with Leloir-type glycosyltransferases.](image)

**1.3 Mechanisms of glycosyltransferases**

Based on the stereochemistry of the newly formed glycosidic bond and that of the original glycosidic bond in the sugar nucleotide donor, two mechanisms (retaining and inverting) were identified for Leloir-type glycosyltransferases (Scheme 1.2).\(^{22}\)
Structural, mechanistic, and computational studies suggest a general base catalyzed $S_{N}2$ mechanism for inverting glycosyltransferases. The lone-pair electrons of the ring oxygen atom facilitate the formation of an oxocarbenium-like transition state by...
donating electron density to the $\sigma^*$ orbital of the anomeric C-O bond. For retaining glycosyltransferases, it was originally proposed that a double displacement mechanism was involved in which a covalent enzymatic intermediate was formed. However, a general lack of conserved amino acid residue on the $\beta$-face of the anomeric carbon atom acting as nucleophilic catalyst$^{24}$ suggests an alternative mechanism in which the nucleophilic acceptor attacks the anomeric carbon from the same face as the leaving group NDP departure in a concerted manner through a highly dissociative oxocarbenium-like transition state ($S_{Ni}$ transition state).$^{25-26}$

Scheme 1.3. Common sugar nucleotides (primary sugar nucleotides): UDP-D-Glc (1-1), UDP-D-Gal (1-2), UDP-D-GlcNAc (1-3), UDP-D-GalNAc (1-4), UDP-D-GlcA (1-5), UDP-D-Xyl (1-6), GDP-D-Man (1-7), GDP-L-Fuc (1-8), CMP-L-Neu5Ac (1-9).
1.4 Sugar nucleotide donors and their biosynthetic pathways.

Despite the structural complexity of oligosaccharides and glycoconjugates, there are surprisingly only nine common building blocks for eukaryotic glycoproteins and glycolipids (Scheme 1.3): glucose (Glc), galactose (Gal), N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc), glucuronic acid (GlcA), xylose (Xyl), mannose (Man), fucose (Fuc) and sialic acid (Neu5Ac). Although there are some posttranslational modifications on these sugars after glycosyltransfer, most eukaryotic glycan structural diversity derives from change in the number and type of these sugar moieties and in the linkages between sugar components. To participate in the construction of different sugar chains, these building blocks must be activated by transforming into sugar nucleotide donors (Scheme 1.3). They are called primary sugar nucleotides since they are generated from sugar-1-phosphate in vivo. Sugar nucleotides that are derived from modifications of primary sugar nucleotides are known as secondary sugar nucleotides. The attachment of a phosphonucleotidyl moiety serves as a recognition element for related enzymes (e.g. glycosyltransferases) and acts as a good leaving group in glycosyltransfer reactions.

Scheme 1.4. General sugar nucleotides biosynthesis reactions.
The role of sugar nucleotides in carbohydrate metabolic pathways was first discovered by the Nobel laureate Luis F. Leloir.\textsuperscript{29} He showed that the first sugar
nucleotide UDP-Glc acted as a cofactor in the following transformation: Gal-1-phosphate + UDP-Glc ⇌ Glc-1-phosphate + UDP-Gal. Half a century later, the biosynthetic pathways of these nine sugar nucleotides are now well established (Scheme 1.5). Regardless of the sugar and its origin, sugar nucleotides are biosynthesized by a kinase and a nucleotidylyltransferase (reaction 1 in Scheme 1.4), or generated from a previously synthesized sugar nucleotides (reaction 2 and 3 in Scheme 1.4). Most NDP sugars are derived from glycolytic intermediates glucose-6-phosphate and fructose-6-phosphate or from galactose. The general biosynthetic pathway for common sugar nucleotides is summarized in Scheme 1.5.

For Glc, GlcNAc and Man, the routes to their sugar nucleotides are (1) sugar to sugar-6-phosphate by a kinase, (2) sugar-6-phosphate to sugar-1-phosphate by a mutase, and (3) sugar-1-phosphate to sugar nucleotides by a nucleotidylyltransferase (pyrophosphorylase), and this is commonly seen in eukaryotes salvage pathways. There is no C6 sugar kinase reported for Gal and GalNAc and Fuc. This is probably because the 6-OH phosphorylation was sterically unfavorable for GalNAc and Gal due to the existence of the axial 4-OH group and apparently, there is no 6-OH for Fuc. Thus, these three monosaccharides are phosphorylated on the anomeric position to give the sugar-1-P directly. There are several key enzymes responsible for the inter-conversion of UDP-Glc ⇌ UDP-Gal, UDP-GlcNAc ⇌ UDP-GalNAc, such as UDP-Gal 4’-epimerase, UDP-GalNAc 4’-epimerase, UDP-Glc 4’-epimerase, UDP-GlcNAc 4’-epimerase, and Gal-1-P uridylyltransferase. UDP-GlcA is universally biosynthesized from oxidation of UDP-Glc and the biosynthesis of UDP-xyl consists of decarboxylation of UDP-GlcA catalyzed by UDP-GlcA decarboxylase. These enzymes are very useful in sugar
nucleotides synthesis as they can convert one sugar nucleotide to another, or provide both sugar nucleotides at the same time.

A notable exception is the biosynthesis of CMP-sugars (such as CMP-sialic acid).\textsuperscript{30-32} Nucleotidylyltransfer reaction can happen directly on Neu5Ac without a sugar-1-P intermediate from CTP to generate CMP-Neu5Ac. Neu5Ac may either come from condensation of ManNAc and pyruvate catalyzed by an aldolase, or from ManNAc via three steps of biotransformation. (Scheme 1.6)

\begin{itemize}
\item \textbf{ATP}
\item \textbf{CMP}
\item \textbf{pyruvate}
\end{itemize}

Scheme 1.6. Biosynthesis of CMP-sialic acid.

1.5 \textbf{Enzymatic production of sugar nucleotides.}

Chemical synthesis of most sugar nucleotides and their derivatives are available but scale up of chemical synthesis to obtain large amounts proves to be not economical and practical. With respect to the identified enzymes in the sugar nucleotides biosynthetic pathways\textsuperscript{28}, the large-scale synthesis of sugar nucleotides and/or their \textit{in situ} regeneration
is partly summarized in Table 1.1. The availability of key enzymes and major glycobiotechnologies are emphasized.

Table 1.1. Enzymatic synthesis of NDP-sugars.

<table>
<thead>
<tr>
<th>Sugar nucleotides</th>
<th>Year</th>
<th>Enzyme source</th>
<th>Scale</th>
<th>Yield</th>
<th>Ref.</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UDP-α-D-Glc</strong></td>
<td>1996</td>
<td>Fermentation and isolation</td>
<td>Ko&lt;sup&gt;33&lt;/sup&gt;</td>
<td>Improved method</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1983</td>
<td>Kinase, mutase, UDP-Glc pyrophosphorylase</td>
<td>Wong&lt;sup&gt;34&lt;/sup&gt;</td>
<td>NTP mixture</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1992</td>
<td>UDP-Glc pyrophosphorylase</td>
<td>Heidlas&lt;sup&gt;35&lt;/sup&gt;</td>
<td>Regeneration of UDP-Glc</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>UDP-α-D-Gal</strong></td>
<td>1992</td>
<td>UDP-Gal pyrophosphorylase</td>
<td>4.2 mmol, 2.5 g</td>
<td>43%</td>
<td>Heidlas&lt;sup&gt;35&lt;/sup&gt;</td>
<td>Regeneration of UDP-Glc in situ</td>
</tr>
<tr>
<td></td>
<td>1999</td>
<td>Gal-1-P uridylyltransferase</td>
<td>1.9 mmol</td>
<td>40%</td>
<td>Elling</td>
<td>Shift the equilibrium by converting Glc-1-P to Glc-6-P</td>
</tr>
<tr>
<td></td>
<td>1998</td>
<td>E. coli overexpressing PPA, UDP-Glc pyrophosphorylase combined with cells that produce UTP</td>
<td>94 g</td>
<td>Space time yield 46 g.1&lt;sup&gt;1&lt;/sup&gt;,d&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Koizumi&lt;sup&gt;36&lt;/sup&gt;</td>
<td>Whole cells as bioreactors</td>
</tr>
<tr>
<td><strong>UDP-α-D-GlcNAc</strong></td>
<td>1992</td>
<td>Extract from yeast</td>
<td>1.7 mmol, 1.1 g</td>
<td>17%</td>
<td>Heidlas&lt;sup&gt;37&lt;/sup&gt;</td>
<td>From GlcNAc-6-P</td>
</tr>
<tr>
<td></td>
<td>1991</td>
<td>Extract from calf liver</td>
<td>0.6 mmol</td>
<td>60%</td>
<td>Korf&lt;sup&gt;38&lt;/sup&gt;</td>
<td>From GlcNAc-1-P</td>
</tr>
<tr>
<td></td>
<td>2009</td>
<td>NahK, GlmU, PPA</td>
<td>~ 40 mg</td>
<td>40%</td>
<td>Guan&lt;sup&gt;39&lt;/sup&gt;</td>
<td>Modified analogues were also synthesized</td>
</tr>
<tr>
<td><strong>UDP-α-D-GalNAc</strong></td>
<td>1970</td>
<td>Gal-1-P uridylyltransferase to produce UDP-GalN followed by N-acetylation</td>
<td>60 μmmol</td>
<td>65%</td>
<td>Maley&lt;sup&gt;40&lt;/sup&gt;</td>
<td>Chemoenzymatic approach. Shift the equilibrium by converting Glc-1-P to Glc-6-P</td>
</tr>
<tr>
<td></td>
<td>1992</td>
<td>Gal-1-P uridylyltransferase to produce UDP-GalN followed by N-acetylation</td>
<td>2.7 mmol, 1.75 g</td>
<td>34%</td>
<td>Heidlas&lt;sup&gt;35&lt;/sup&gt;</td>
<td>Chemoenzymatic approach. Shift the equilibrium by converting Glc-1-P to regenerate UDP-Glc</td>
</tr>
<tr>
<td></td>
<td>2009</td>
<td>NahK, GlmU, PPA</td>
<td>~ 40 mg</td>
<td>65%</td>
<td>Guan&lt;sup&gt;39&lt;/sup&gt;</td>
<td>Modified analogues were also synthesized</td>
</tr>
</tbody>
</table>

Continued
As shown in Table 1.1, some enzymes for the synthesis of sugar nucleotides are highly expressed recombinant proteins while some of them can only be characterized as biocatalysts for production purposes (e.g. extract from yeasts or organs). However,
nowadays enzymes from microbial sources or recombinant mammalian glycosyltransferases produced in suitable hosts are well investigated and thus mainly employed for sugar nucleotides synthesis (see later Chapters).

1.6 Chemical synthesis of Sugar nucleotides.

Although there are only nine different sugar nucleotides that have been identified in mammalian organisms, the number of sugar nucleotides is tremendous in other organisms, especially in prokaryotes. Naturally occurring sugar nucleotides as well as their structural derivatives are of paramount interest as substrates for enzymatic synthesis of oligosaccharides and glycoconjugates. Therefore, methods for the efficient preparation of natural and non-natural sugar nucleotides can greatly benefit synthetic, biological and medicinal chemistry.

The chemical synthesis of sugar nucleotides is non-trivial and complicated by a number of factors: the poor solubility of sugar nucleotides in organic solvents, the presence of several polar or charged functional groups, and the instability of the glycosidic and pyrophosphate bonds to hydrolytic cleavage. Despite the fact that enzymatic approaches are available for most natural and some unnatural sugar nucleotides,

chemical methods are still indispensable, especially where non-natural compounds need to be prepared and are not accepted as substrates for enzymatic routes.

There are two major strategies for the preparation of NDP-sugars from smaller building blocks (Scheme 1.7). One which has gained less attention is to build sugar nucleotides from a nucleoside diphosphate and a monosaccharide donor (strategy a, scheme 1.7). The major disadvantages for this approach are the difficulties involved in
the identification of a suitable sugar donor for the diphosphate acceptor and the stereoselectivity issue as well.

Scheme 1.7. General synthetic strategy for NDP-sugars.

The most common approach, which was developed by Khorana and co-workers, requires two monophosphate precursors: a nucleoside monophosphate and a glycosyl phosphate (strategy b, Scheme 1.7). One of the two reaction precursors needs to be activated prior to the pyrophosphate linkage formation. This allows the condensation...
reaction to proceed under lower temperature and also prevents the formation of homodimers. Coupling conditions were generally anhydrous and in very polar solvents like pyridine or DMF to prevent the hydrolysis of the starting material and product and dissolve the very polar reaction partners.\textsuperscript{51-53} Considerable advantages of this approach over the glycosylation with nucleoside disphosphate can be easily noticed: predetermined anomeric configuration of the sugar in the product sugar nucleotide and both starting materials are readily available. Even though there are several variants of this method reported, such as addition of tetrazole\textsuperscript{51} or Lewis acid\textsuperscript{54-55} as catalysts, introduction of carbonyldiimidazole (CDI) as activating reagent\textsuperscript{55-56}, it is still the most popular chemical method for the preparation of sugar nucleotides as well as other pyrophosphate containing molecules.

Usually, the nucleoside monophosphate is the activated species since activation of the sugar-1-phosphate would give sugar-1,2-cyclic phosphate instead of the desired coupling product\textsuperscript{57} (see inserted box in Scheme 1.7). Phosphoric acid amides, especially phosphoromorpholidates have become the most widely used activated forms of nucleoside monophosphate for the preparation of NDP sugars.\textsuperscript{52,58-59} Many natural nucleoside phosphoromorpholidates are now commercially available since the synthesis is straightforward: coupling reaction between nucleoside monophosphates and morpholine catalyzed by \textit{N},\textit{N}’-dicyclohexylcarbodiimide (DCC).\textsuperscript{49}

Sugar-1-phosphates with either anomeric configuration could be prepared from the corresponding protected sugar precursor via several synthetic manipulations. For making UDP-sugars which represent a major class of primary sugar nucleotides (Scheme 1.3) and also the focus of the following Chapters, the necessary \(\alpha\)-sugar-1-phosphate
could be furnished by the following typical steps (Scheme 1.8)\textsuperscript{52,60-62}: protection, selective anomic deprotection, phosphoramidite chemistry, oxidation, and deprotection. The resulting sugar-1-phosphate could either be protected or unprotected depending on the coupling reaction conditions. Wong and co-workers reported the stereoselectivity of phosphitylation reactions\textsuperscript{61}. For Glc and Gal, the less stable β-phosphates were the major products regardless of the anomic ratio of the starting sugar-1-OH while Man and Rha produced exclusively α-phosphates. Besides, only α-phosphates were obtained for amino sugars GlcNAc and GalNAc, which will be mentioned again in Chapter 2.

![Scheme 1.8. General synthetic strategy for α-sugar-1-phosphate.](image)

In summary, chemical approach apparently could be applied to the synthesis of any sugar nucleotides of interest. However, it is mainly plagued with such serious drawbacks as long coupling reaction time and generally low to moderated overall yields. On the other hand, enzymatic approaches which could overcome the above disadvantages suffer from limited availability of key enzymes or limited substrate tolerance of enzymes. Thus a combination of chemical synthesis and enzymes which can furnish the key linkages that are usually difficulty for chemical methods provides a new platform now
called chemoenzymatic synthesis. This method takes advantage of both disciplines and may be a promising solution for sugar nucleotides preparation.

References


CHAPTER 2

A CHEMOENZYMATIC ROUTE TO UDP-GlcNAc/GalNAc LIBRARY

2.1 Introduction

It is well known that glucose, which makes up starch and cellulose, is the most abundant monosaccharide in nature. However, in our mammalian world N-acetylglucosamine (GlcNAc, 2-1) represents the most popular monosaccharide and investigations showed that GlcNAc and its 4-hydroxy epimer GalNAc (2-2) consist of over 36% of mammalian carbohydrate moieties.\(^1\) In fact, GlcNAc and GalNAc are prevalent sugar residues in organisms ranging from bacteria to vertebrates. (Figure 2.1)

Firstly, they are fundamental components of many important polysaccharides or glycoconjugates (Figure 2.1). For example, chitin, which is the simplest GlcNAc polymer, is the main component of fungi cell walls and also the exoskeletons of crustaceans and insects. Besides, alternating units of GlcNAc and N-acetylmuramic acid (MurNAc) form the backbone of peptidoglycan of bacterial cell wall.\(^2\)\(^-\)\(^3\) GlcNAc and GalNAc are also key units of glycosaminoglycans (e.g. hyaluronan, heparin), which are linear polysaccharides interacting with many biologically important proteins and dominating the chemical and biological properties of proteoglycans.\(^4\)\(^-\)\(^6\)
Figure 2.1. Prevalence of GlcNAc and GalNAc in polysaccharides and glycoconjugates
In addition, GlcNAc and GalNAc are also key components of lots of biologically active complex polysaccharides such as tumor associated carbohydrate antigens (TACAs), which are known as the glycans uniquely or excessively expressed by tumor cells. Those GlcNAc/GalNAc containing polysaccharides are thus useful targets for cancer immunotherapies and the development of therapeutic cancer vaccines.\(^7^8\) On the other hand, for glycoconjugates, all typical N-linked glycoproteins are characterized by a \(\beta\)-glycosidic bond between a GlcNAc residue and an asparagine. Mucin-Type O-GalNAc glycans are characterized by a \(\alpha\)-glycosidic linkage between a GalNAc residue and serine or threonine, which is also typical for major vertebrate O-linked glycoproteins. More significantly, O-GlcNAc, the single GlcNAcylation of Ser and Thr residues of cytoplasmic proteins of all multicellular organisms is becoming increasingly important by modulating the function/activity of proteins in cells. Studies suggest that this special type of GlcNAcylation is a key regulatory modification in the brain, contributing to transcriptional regulation and neuronal signaling (Figure 2.1, bottom right).\(^9^\text{-}^{11}\)

As mentioned in Chapter 1 that sugar nucleotides are natural donor molecules for polysaccharides and glycoconjugates biosynthesis. The corresponding donor molecules for GlcNAc/GalNAc are uridine 5'-diphospho-GlcNAc/GalNAc (UDP-GlcNAc/GalNAc). These donors are obtained from enzymatic pyrophosphorylation of GlcNAc-1-phosphate (GlcNAc-1-P), which is biosynthesized from either glucose (\textit{de novo} pathway) or GlcNAc/GalNAc (salvage pathways) via several enzymatic steps (Scheme 2.1).\(^2\) The biosynthesis of UDP-GlcNAc/GalNAc could be further simplified as sugar to sugar-6-P by a 6-kinase; the sugar-6-P is mutated to sugar-1-P by a mutase; then sugar-1-P is pyrophosphorylated with UTP to give UDP-sugar (Scheme 2.1).
Scheme 2.1. Biosynthesis pathways of UDP-GlcNAc/GalNAc.

Classical chemical synthesis of GlcNAc-1-P and UDP-GlcNAc involves multistep synthetic manipulations (Scheme 2.2) and purifications which result in a tedious process and low overall yield (36% for GlcNAc-1-P, 13% for UDP-GlcNAc in my practice). Peracetylation of glucosamine hydrochloride (2-3) afforded GlcNAc peracetate (2-4). The anomic acetate was selectively removed by hydrazine acetate and the resulting anomic hydroxyl group was phosphorylated by typical phosphoramidite chemistry followed by oxidation with mCPBA to give the protected GlcNAc-1-phosphate (2-6). Only the α-isomer was isolated. The protection groups were removed by Pd(OH)$_2$/C catalyzed hydrogenolysis and deacetylation using sodium methoxide. The chemically synthesized GlcNAc-1-P (2-7) was subsequently coupled with commercially
available UMP-morpholidate at room temperature for 5-7 days to afford UDP-GlcNAc (2-8) in poor overall yield.

Scheme 2.2. Chemical synthesis of GlcNAc-1-P (2-7) and UDP-GlcNAc (2-8).

It is thus quite clear that an enzymatic route following the biosynthetic pathways is a more promising alternative. More significantly, the enzymatic approach can be more attractive if GlcNAc (2-1) is directly phosphorylated to GlcNAc-1-P by an anomeric kinase (compared with two reactions with 6-kinase and mutase in Scheme 2.1). This straightforward strategy is illustrated in Scheme 2.3. Unfortunately, previous reported 1-
kinases that can convert a sugar directly into the corresponding sugar-1-P are limited to fucokinases,\textsuperscript{12} galactokinases,\textsuperscript{13-15} and GalNAc kinases.\textsuperscript{16-17} Site specific mutations were thus carried out extensively to alter the substrate specificity of those kinases or to generate novel mutants to accommodate structures of interest.\textsuperscript{18-21} For example, the so called ‘first’ glucokinase was created from mutation of \textit{E.coli} galactokinase.\textsuperscript{18,21}

\textbf{NahK} (EC 2.7.1.162), a recently discovered \textit{N}-acetylhexosamine kinase, is the truly first wild type gluco-type 1-kinase reported.\textsuperscript{22} This enzyme was found in \textit{Bifidobacterium longum} JCM1217 and was named NahK by its function or \textit{Lacto-}\textit{N}-biose Phosphorylase \textit{B} (LnpB) by its origin (Figure 2.2a). This enzyme could quickly and efficiently synthesize GlcNAc-1-P from GlcNAc acceptor and ATP phosphate donor (Figure 2.2b). A yield of 90\% was obtained on a preparative scale (1 gram of GlcNAc starting material). TLC analysis showed that the starting material GlcNAc was completely consumed within 19 h and a clean spot of the product GlcNAc-1-P had distinct Rf value with the byproduct ADP (Figure 2.2c).

![Scheme 2.3. Straightforward synthesis of GlcNAc-1-P and UDP-GlcNAc.](image-url)
Figure 2.2. a. SDS-PAGE of NahK expression and purification. Aliquots were removed from each step of purification and loaded onto a 12% SDS-PAGE gel followed by Coomassie Brilliant Blue staining. Lanes: M, SeeBlue Plus2 Pre-Stained Standard; 1, cell lysate; 2, cell pellet; 3, supernatant of cell lysate; 4, flow through after sample loading (the first time) onto Ni-column; 5, flow through after sample loading (the second time) onto Ni-column; 6, Ni-column purified protein; b. Preparative scale synthesis of GlcNAc-1-P with NahK; c. TLC analysis of NahK reaction mixture. Lanes: 1, GlcNAc; 2, GlcNAc-1-P authentic sample; 3, NahK reaction mixture showing GlcNAc-1-P product and complete consumption of starting material GlcNAc. (nBuOH/ACOH/H2O 2/1/1, stained with anisaldehyde sugar stain).

2.2 Substrate specificity investigation of NahK and GlcNAc/GalNAc-1-P library

Since there is increasing interest in glycan and glycoconjugate mimics with modified sugar residues for either probing the carbohydrate metabolic pathways\textsuperscript{23-24} or generating new glycoforms\textsuperscript{25-26}, our first step is to investigate the substrate specificity of NahK using structurally modified GlcNAc derivatives (Figure 2.3a) and generate a library of GlcNAc-1-P analogues if this enzyme proves to be promiscuous. Meantime, we aim to install different bioorthogonal groups (e.g. reactive ketones, azides, alkynes, etc.) since molecules with those chemical handles can be further functionalized by chemoselective reactions (Figure 2.3b).\textsuperscript{9,27-31} Moreover, if those modified substrates can tolerate in all the biosynthetic pathways and be utilized and presented by cells, we can ultimately decorate cellular or bacterial surfaces with modified glycans or
glycoconjugates comprised of a ketone or azide functionality (Figure 2.3c, also see Chapter 3).  

Figure 2.3.  

a. Structural modification of GlcNAc/GalNAc.  

b. Examples of bioorthogonal reactions of aldehydes/ketones and azides. Aldehydes and ketones can react with aminooxy (or hydrazide) reagents to form oxime (or hydrazone) linkage (top); Azido biomolecules could be labeled via Staudinger-Bertozzi reactions (middle) or via Copper catalyzed azide-alkyne 1,3-dipolar cycloaddtions (bottom).  

c. Metabolic oligosaccharide remodeling. Modified sugar bearing an azide epitope is utilized by cells or bacteria and incorporated into cell surfaces glycans. The azido group can be detected via Copper catalyzed cycloaddition or strain-promoted cycloaddition.
2.2.1 Synthesis of GlcNAc/GalNAc derivatives as NahK substrates

As designed in Figure 2.3a, GlcNAc analogues include C-2 (acyl chain), C-3, C-4, and C-6 modifications. GalNAc, the 4-hydroxy epimer of GlcNAc, could also be regarded as a ‘modified analogue’ (this sugar was also proved to be a good substrate of NahK, see later in this Chapter). We thus synthesized structurally modified compounds of both GlcNAc/GalNAc series as potential NahK substrates.

For C-2 modifications, N-propionyl, N-butyryl, N-benzoyl analogues were prepared by treating glucosamine hydrochloride (2-3) or galactosamine hydrochloride (2-9) with sodium methoxide, followed by different acid anhydrides in methanol (Scheme 2.4 top). It is well known that peracylation happens when treating glucosamine with acid anhydride in pyridine. However, early in 1956 Inouye et al. reported that treatment of glucosamine with the acid anhydrides in methanol or with acid chloride in water could both result in regioselective N-acylated glucosamines.\(^{35}\) \(N\)-chloroacetyl intermediates (2-16 and 2-17) were generated using chloroacetic anhydride and displacement of the \(\alpha\)-chloride with sodium azide at 80 °C gave the \(N\)-azidoacetyl analogues (2-18 and 2-19) bearing an azide tag (Scheme 2.4 middle). It is noteworthy that these azido compounds (2-18/2-19 or commonly called GlcNAz/GalNAz) have been very popular candidates for metabolic engineering of cell surface glycans\(^{28}\) and cytoplasmic proteins\(^{27}\) with azido groups. Unfortunately, the corresponding donor molecules UDP-GlcNAz/GalNAz were all chemically synthesized if used in these experiments.\(^{27,36}\) Meantime, the phthalide derivative (2-20) was also synthesized following the reported procedure (Scheme 2.4 bottom).\(^{37}\)
Scheme 2.4. Synthesis of acyl chain modified GlcNAc/GalNAc.

To determine the importance of the amido moiety of GlcNAc/GalNAc for NahK activity, we simply prepared 2-azido-deoxy-glucose/galactose (2-21/2-22) using newly reported diazotransfer reagent imidazole-1-sulfonyl azide which is easily prepared and shelf-stable (Scheme 2.5). [38]
Scheme 2.5. Synthesis of 2-azido-glucose/galactose (2-21/2-22).

The C-3 epimer of GlcNAc (2-23, also called N-acetyl allosamine) was synthesized as shown in Scheme 2.6. Glycosylation of peracetylated GlcNAc (2-4) with thiophenol and BF$_3$ etherate afforded β-thioglycoside (2-24) after deacetylation. $O_4$ and $O_6$ oxygen atoms were selectively protected as a benzylidene acetal (compound 2-25). 3-OH was activated as a triflate with 1.2 equiv. of triflic anhydride in pyridine/CH$_2$Cl$_2$ and it underwent an intramolecular S$_{N}2$ reaction attacking by the carbonyl oxygen to give the methyloxazoline intermediate with C-3 epimerized (2-26). Similar strategy was reported very early while using mesylate as the activation species. We noticed that if triflic anhydride was used in large excess the resulting oxazoline (2-26) would easily tautomerize to give the N-trifluoromethanesulfonyl enamine (2-27) (Scheme 2.6 bottom reaction). The tautomerization was also observed in thiazoline systems by Knapp and co-workers and had been successfully used for modification of the thiazoline ring methyl group. Ring opening of the oxazoline and benzylidene was performed under acidic conditions, after which the thio group of AlINAc thioglycoside (2-29) was further removed by NBS in presence of water to afford N-acetyl allosamine (2-23, also called AllNAc). In this case, a pyranose-furanose equilibrium of 2-23 was observed by $^1$H NMR in addition to the α:β anomic equilibrium (Figure 2.4). It is also noticed that the major
anomer of solution (2-23) is the β-D-pyranose form. The preference of β form probably results from unfavorable 1,3-diaxial interactions of α-D-AllNAc 1,3-OH groups, compared to the facts that there is no furanose form existing for GlcNAc solution and its major isomer is the α-pyranose.$^{44}$ (Figure 2.4).

Scheme 2.6. Synthesis of N-acetyl allosamine (2-23, GlcNAc-3-epimer).

Figure 2.4. $^1$H NMR spectrum of AllNAc (2-23) and its pyranose/furanose equilibrium.
For C-4 modifications, besides natural GlcNAc/GalNAc which are 4-epimers with respect to each other, we further synthesized 4-deoxy version of GlcNAc/GalNAc (2-30) and 4-azido version of GalNAc (2-31) (Scheme 2.7). Starting from GlcNAc (2-1), O4 and O6 oxygen atoms were selectively protected as a benzylidene acetal followed by acetylation of O1 and O3 to afford compound (2-32) as the only isomer. Ring opening of the benzylidene by treatment with acetyl chloride in methanol; subsequent selective acetylation of O6 gave the 4-hydroxyl compound (2-34) as a key intermediate.

Scheme 2.7. Synthesis of C-4 modified GlcNAc/GalNAc.
Triflation of compound (2-34) followed by nucleophilic substitution using sodium azide gave the triacetate (2-35) with 4-axial azido group. Deacetylation with sodium methoxide afforded the target 4-azido-GalNAc (2-31). On the other hand, conversion of alcohol (2-34) to the corresponding thionocarbonate (2-36), followed by free radical deoxygenation with tributyltinhydride and AIBN gave the 4-deoxy triacetate (2-37). Deprotection with sodium methoxide yielded the 4-deoxy analogue (2-30).

Activation of hydroxyl group as a xanthate with NaH, imidazole, CS$_2$ seemed not feasible for GlcNAc deoxygenation because an unexpected rearrangement happened to yield the isothiocyanato intermediate (2-38) which was further attacked by the adjacent oxygen. Scheme 2.8 showed the formation of a trans fused oxazoline (2-39) when we tried to synthesize 3-deoxy-GlcNAc analogue.

Scheme 2.8. Isothiocyanate formation using NaH, CS$_2$ and MeI reagents.
Scheme 2.9. Synthesis of 6-modified GlcNAc analogues.

For C-6 modifications, we synthesized 6-deoxy (2-40), 6-azido (2-41) and 6-N-acetyl (2-42) analogues of GlcNAc. The purpose for those versions was to investigate the influence of 6-bulkiness to the enzyme activity as well as the importance of the 6-hydroxyl group. The synthesis is summarized in Scheme 2.9, starting from thioglycoside (2-24). Selective 6-mesylation afforded the mesylate (2-43) in 74% yield, after which the 6-mesylate was reduced by NaBH₄ in dry DMSO at an elevated temperature to give the 6-deoxy thioglycoside (2-44). The anomeric phenylthio was removed by treatment with NBS to give the 6-deoxy GlcNAc analogue (2-40). 6-Azido GlcNAc (2-41) was prepared through seleveutive 6-tosylation of (2-24) followed by azide substitution in DMF without isolation of the 6-tosylate. A similar anomic deprotection with NBS afforded
the azido sugar (2-41) in good yields.\textsuperscript{52} Contrastingly, the 6-azido group of intermediate (2-45) was also reduced to the corresponding amine via the Staudinger reaction,\textsuperscript{53} after which it was peracetylated in a one pot fashion to afford compound (2-46). Deacetylation with sodium methoxide and subsequent anomeric deprotection furnished 6-N-acetyl-GlcNAc (2-42).\textsuperscript{54}

This reaction sequence (Scheme 2.9) went quite smoothly for GlcNAc derivatives with 4-hydroxyl group on equatorial positions. However, for GalNAc derivatives bearing an axial hydroxyl group, both NaBH\textsubscript{4} reduction and azide substitution were performed with fairly low yield (Scheme 2.10). The most probable reason lies in the fact that 6-OH of GalNAc is quite sterically hindered due to the existence of the axial 4-OH group. It is also evidenced by the observation that there is no C6 sugar kinase reported for galactose and GalNAc till date.\textsuperscript{19-20}

Scheme 2.10. Synthesis of 6-modified GalNAc derivatives route I.
To improve the yield of preparing compound (2-47) and (2-48), a second strategy was used as shown in Scheme 2.11. Starting from GalNAc (2-2), O4 and O6 oxygen atoms were selectively protected as a benzylidene acetal followed by acetylation of O1 and O3 to afford compound (2-55) as the only isomer (α). Ring opening of the benzylidene with \( p \)-toluenesulfonic acid afforded the 4,6-diol (2-56). 6-Tosylate (2-57) was obtained by selective tosylation of 6-OH followed by acetylation. Deoxygenation reaction at the C6-position of (2-57) was performed by iodination and subsequent radical reduction with tributyltinhydride. Deprotection of the 6-deoxy triacetate (2-58) by sodium methoxide afforded 6-deoxy GalNAc (2-47) in better overall yield. The 6-azido
compound (2-59) could simply be constructed from the tosylate intermediate (2-57) by azide substitution and deacetylation. Similarly, due to the sterically hindrance of 4-OH, the substitution reaction ended up with moderate yield (50%) and long reaction time.

2.2.2 Synthesis of GlcNAc/GalNAc-1-phosphate derivatives: Substrate specificity investigation of NahK.

With the GlcNAc/GalNAc derivatives ready, we performed enzymatic 1-phosphorylation reactions of those analogues with NahK using adenosine 5’- triphosphate (ATP) as a phosphate donor (Scheme 2.12). For optimized reaction details please refer to Experimental Section. NahK could also accept guanosine triphosphate (GTP) and inosine triphosphate (ITP) as phosphate donors with weaker activity as reported by Kitaoka and co-workers.\(^{22}\) It is obvious that there is high negative charge associated with the triphosphate group of ATP, which shields each phosphorus against reaction with incoming nucleophiles (anomeric OH). Actually, this property makes ATP kinetically stable in the cell. In enzyme catalysis reactions such as NahK, these charges are probably neutralized in order to facilitate nucleophilic attack by two factors: 1. Coordination with metal ions, mostly divalent magnesium (Scheme 2.12). In cells, ATP is frequently found associated with magnesium, and the true substrate is ‘Mg-ATP’.\(^{56-57}\) 2. Ion pairing with positively charged amino acids within the enzyme active site such as the guanidinium of arginine, or the lysine ammonium group.\(^{58-59}\)
Scheme 2.12. Enzymatic synthesis of GlcNAc/GalNAc-1-P derivatives with NahK.

Unless specified, all GlcNAc/GalNAc derivatives were incubated with NahK in Tris-HCl buffer (pH 9.0) and the product GlcNAc/GalNAc-1-P analogues were separated using silica gel chromatography and P2 gel filtration column for NMR purposes. Reactions were carried out on a preparative scale (20–40 mg) and the corresponding reaction yield was used as a representative of NahK substrate tolerance (Table 2.1-2.4).

Even though the NahK crystal structure is not yet available, we were able to investigate the structural requirements within the enzyme active site and summarize the structure activity relationship using these diversified GlcNAc/GalNAc analogues.
Table 2.1. Isolated yields for enzymatic 1-phosphorylation of C-2 modified GlcNAc analogues.

<table>
<thead>
<tr>
<th>Substrate Entry</th>
<th>Substrate Structure</th>
<th>1-P Yield&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Substrate Entry</th>
<th>Substrate Structure</th>
<th>1-P Yield&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td><img src="image" alt="Structure 2-1" /></td>
<td>90%</td>
<td>2-12</td>
<td><img src="image" alt="Structure 2-12" /></td>
<td>88%</td>
</tr>
<tr>
<td>2-10</td>
<td><img src="image" alt="Structure 2-10" /></td>
<td>86%</td>
<td>2-18</td>
<td><img src="image" alt="Structure 2-18" /></td>
<td>87%</td>
</tr>
<tr>
<td>2-11</td>
<td><img src="image" alt="Structure 2-11" /></td>
<td>87%</td>
<td>2-20</td>
<td><img src="image" alt="Structure 2-20" /></td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>2-21</td>
<td><img src="image" alt="Structure 2-21" /></td>
<td>58%</td>
<td></td>
<td></td>
<td>58%</td>
</tr>
</tbody>
</table>

<sup>a</sup> Isolated yield from silica gel column chromatography for 1-phosphorylation reactions.

<sup>b</sup> N/A (Not available): No product detected by MS or TLC.

Table 2.1 summarized the reaction yields for C-2 modified GlcNAc analogues.

In comparison to GlcNAc (2-1, 90% conversion), the enzyme NahK demonstrated
comparable levels of activity towards the C-2 modified compounds. Modifications include aliphatic (2-10, 2-11, 2-18) and aromatic acyl chains (2-12), indicating a relaxed specificity to the N-acyl chain derivatization. Most significantly, tolerance of compound (2-12) and (2-18), which bore a bulky phenyl group or an azido side chain, indicated the potential for installation of a small fluorescence substituent or a bio-orthogonal tag into sugar-1-P and sugar nucleotides. It is also worth noting that although it had no activity for N-Phth modified compound (2-20), 2-azido-2-deoxy-glucose (2-21) which bears no acetamido moiety turned out to be a substrate for NahK under longer reaction time (64.5 h, 58% yield). This successful example suggests we could even enlarge our sugar-1-P and sugar nucleotide library by synthesizing UDP-glucose derivatives under optimized reaction conditions.

Similarly, C-2 modified GalNAc analogues were phosphorylated in good yields, including both aliphatic and aromatic acyl chain modifications (Table 2.2). This at least means epimerization of 4-OH of GlcNAc does not significantly affect the enzyme activity. However, the bulkiness of acyl chain exhibited a slight influence on enzyme activity (compound 2-15 and 2-19) for GalNAc series; and galactose (Gal) was a very poor substrate (5% as indicated by TLC) as was the completely unreactive 2-azido-2-deoxy-galactose (2-22). Depending on the success of 2-azido-glucose (2-21), we can conclude that NahK has slightly more limited specificity towards GalNAc series than GlcNAc series.
Table 2.2. Isolated yields for enzymatic 1-phosphorylation of C-2 modified GalNAc analogues.

<table>
<thead>
<tr>
<th>Substrate Entry</th>
<th>Substrate Structure</th>
<th>1-P Yield(^a)</th>
<th>Substrate Entry</th>
<th>Substrate Structure</th>
<th>1-P Yield(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-2</td>
<td><img src="image" alt="Substrate 2-2" /></td>
<td>78%</td>
<td>2-15</td>
<td><img src="image" alt="Substrate 2-15" /></td>
<td>77%</td>
</tr>
<tr>
<td>2-13</td>
<td><img src="image" alt="Substrate 2-13" /></td>
<td>85%</td>
<td>2-19</td>
<td><img src="image" alt="Substrate 2-19" /></td>
<td>65%</td>
</tr>
<tr>
<td>2-14</td>
<td><img src="image" alt="Substrate 2-14" /></td>
<td>86%</td>
<td></td>
<td><img src="image" alt="Substrate Gal" /></td>
<td>&lt;5% (TLC)</td>
</tr>
<tr>
<td>2-22</td>
<td><img src="image" alt="Substrate 2-22" /></td>
<td>N/A(^b)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Isolated yield from silica gel column chromatography for 1-phosphorylation reactions.  
\(^b\) N/A (Not available): No product detected by MS or TLC.

Encouragingly, the enzyme NahK also displayed a wide tolerance of C-4 modifications (R\(^4\), Table 2.3, left portion). Axial hydroxyl (2-2), equatorial hydroxyl (2-1), azido (2-31), and deoxy (2-30) compounds were all accepted as good substrates,
indicating 4-hydroxyl group may not be involved in enzyme recognition, another feature that could greatly expand our perspective library of sugar nucleotide analogues.

Table 2.3. Isolated yields for enzymatic 1-phosphorylation of C-4 and C-3 modified GlcNAc and GalNAc analogues.

<table>
<thead>
<tr>
<th>Substrate Entry</th>
<th>Substrate Structure</th>
<th>1-P Yield&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Substrate Entry</th>
<th>Substrate Structure</th>
<th>1-P Yield&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td><img src="image1" alt="Structure" /></td>
<td>90%</td>
<td>2-23</td>
<td><img src="image2" alt="Structure" /></td>
<td>22%</td>
</tr>
<tr>
<td>2-2</td>
<td><img src="image3" alt="Structure" /></td>
<td>78%</td>
<td>2-70</td>
<td>Only MS detected</td>
<td></td>
</tr>
<tr>
<td>2-30</td>
<td><img src="image4" alt="Structure" /></td>
<td>70%</td>
<td>2-71</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>2-31</td>
<td><img src="image5" alt="Structure" /></td>
<td>73%</td>
<td>2-72</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Isolated yield from silica gel column chromatography for 1-phosphorylation reactions.

<sup>b</sup> N/A (Not available): No product detected by MS or TLC.

2-Acetamino-2-deoxy-D-allose (2-23, GlcNAc 3-epimer) was identified as a poor substrate (22%), indicating poor substrate tolerance for C-3 modified analogues (Table
While 1-phosphorylation product of 3-\textit{N}-acetyl compound (2-70)\textsuperscript{60} could only be detected by mass spectrometry, \textit{N}-acetylmuramic acid\textsuperscript{61-62} (2-71, bacterial cell wall component) and its methyl ester (2-72) were completely unreactive which bore a bulky and charged substituent at the C-3 position.

Contrastingly, the extent of tolerance for C-6 modifications is limited of both GlcNAc and GalNAc type analogues (Table 2.4). For example, phosphorylation of 6-deoxy GlcNAc (2-40) and 6-deoxy GalNAc (2-47) occurred in moderate yields, thus revealing that the 6-hydroxy is not necessary for enzyme activity for both series. The activity decreased, however, as the steric bulk at C-6 increased from azido (2-41) (34%) to 6-\textit{N}-acetyl (2-42) (N/A).

Table 2.4. Isolated yields for enzymatic 1-phosphorylation of C-6 modified GlcNAc and GalNAc analogues.

<table>
<thead>
<tr>
<th>Substrate Entry</th>
<th>Substrate Structure</th>
<th>I-P Yield\textsuperscript{a}</th>
<th>Substrate Entry</th>
<th>Substrate Structure</th>
<th>I-P Yield\textsuperscript{a}</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-40</td>
<td><img src="image" alt="Structure" /></td>
<td>75%</td>
<td>2-47</td>
<td><img src="image" alt="Structure" /></td>
<td>37%</td>
</tr>
<tr>
<td>2-41</td>
<td><img src="image" alt="Structure" /></td>
<td>34%</td>
<td>2-48</td>
<td><img src="image" alt="Structure" /></td>
<td>42%</td>
</tr>
<tr>
<td>2-42</td>
<td><img src="image" alt="Structure" /></td>
<td>N/A\textsuperscript{b}</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a} Isolated yield from silica gel column chromatography for 1-phosphorylation reactions.

\textsuperscript{b} N/A (Not available): No product detected by MS or TLC.
In summary, we successfully synthesized a small library of structurally modified GlcNAc/GalNAc-1-P analogues (C-2,3,4,6 modifications) with NahK (Figure 2.5). This enzyme had been proved to be very promiscuous and demonstrated very broad substrate specificity. Specifically, the relaxed tolerance to the N-acyl chain modifications could most possibly allow us to incorporate bio-orthogonal functionalities which can carry additional tags or labels into carbohydrate containing biomolecules (see this application in Chapter 3).

Figure 2.5. Structures of GlcNAc/GalNAc-1-P derivatives.
2.3 Enzymatic synthesis of UDP-GlcNAc/GalNAc analogues library

Since sugar-1-P is the key intermediate for sugar nucleotide biosynthesis (Scheme 2.1), our next aim is to convert our sugar-1-P library (Figure 2.5) to the corresponding UDP-sugar donor library with a suitable pyrophosphorylase (also called uridylyltransferase).

2.3.1 Substrate specificity investigation of GlmU and UDP-GlcNAc/GalNAc analogues library

Because the kinase (NahK) we used for sugar-1-P synthesis is a bacterial source enzyme, we first exploited a well known bacterial (Escherichia coli) uridylyltransferase GlmU as our candidate for sugar nucleotide preparation. GlmU is called GlcNAc-1-P uridylyltransferase and a bifunctional enzyme (EC 2.3.1.157 and EC 2.7.7.23).\textsuperscript{63-64} It has both N-acetyltransferase and pyrophosphorylase activities and has been proved to catalyze the last two sequential reactions in the UDP-GlcNAc biosynthesis in bacteria. (Scheme 2.13).\textsuperscript{2}

![Scheme 2.13. Biosynthesis of UDP-GlcNAc with bifunctional GlmU.](image)

47
The crystal structure of GlmU\(^{65}\) from *E.coli* is already available on Protein Data Bank (www.pdb.org 1HV9) (Figure 2.6a), which has two distinct catalytic sites: the C-terminal *N*-acetyltransferase domain and *N*-terminal pyrophosphorylase domain. Those two catalytic domains are connected by a α-helix linker. It was reported that truncated GlmU which only kept the pyrophosphorylase domain (Figure 2.6b) still retained the pyrophosphorylase activity.\(^{66}\)

![Figure 2.6.](image)

**Figure 2.6.** **a.** Full length crystal structure of *Escherichia coli* GlmU from Protein Data Bank (www.pdb.org), showing two catalytic domains: C-terminal *N*-acetyltransferase domain (purple), *N*-terminal pyrophosphorylase domain (green) and the α-helix linker (yellow); **b.** GlmU pyrophosphorylase domain showing the active site and the ligand UDP-GlcNAc; **c.** SDS-PAGE of GlmU expression and purification. Aliquots were removed from each step of purification and loaded onto a 10% SDS-PAGE gel followed by Coomassie Brilliant Blue staining. Lanes: M, SeeBlue Plus2 Pre-Stained Standard; 1, cell pellet; 2, supernatant of cell lysate; 3, flow through after sample loading onto Ni-column; 4, Ni-column purified protein; 5, filtrate during concentration of purified protein.

We cloned and expressed full length GlmU from *E.coli* K12 (Figure 2.6c) and used it as a pyrophosphorylase only (Scheme 2.14). With our GlcNAc/GalNAc-1-P
analogues as substrates, we performed enzymatic pyrophosphorylation reactions using uridine 5’-triphosphate (UTP) as the donor molecule. For optimized reaction details please refer to Experimental Section. Mg$^{2+}$ was also present to ‘neutralize’ the negative charge of the triphosphate and facilitate the departure of a pyrophosphate. In addition, yeast inorganic pyrophosphatase (PPA) was added necessarily here to hydrolyze the byproduct pyrophosphate (PPI) and to shift the equilibrium to the product, because the accumulation of PPI may inhibit the enzymatic activity of GlmU (Scheme 2.14).

Unless specified, all GlcNAc/GalNAc-1-P derivatives were incubated with GlmU in Tris-HCl buffer (pH 7.5) and the product UDP-GlcNAc/GalNAc analogues were separated sequentially by DEAE anion exchange resin chromatography and, for desalting purpose, size exclusion chromatography (P2 gel filtration). Reactions were normally carried out on a preparative scale (20~40 mg) and the corresponding reaction yield was used as a representative of GlmU substrate tolerance (Table 2.5-2.6).

Scheme 2.14. Enzymatic synthesis of UDP-GlcNAc/GalNAc derivatives with GlmU.
As shown in Table 2.5-2.6, UDP-GlcNAc (2-8) and its 4-epimer UDP-GalNAc (2-89) and the 4-deoxy version (2-88) were all obtained by GlmU at comparable levels, indicating the 4-hydroxyl group may not be necessary for enzyme recognition. In contrast, 4-azido compound (2-97) failed in the reaction, presumably resulting from a larger azido group than the hydroxyl group.

Most distinctly, the tolerance for N-acyl modifications was totally different for GlcNAc-1-P and GalNAc-1-P analogues (Table 2.5-2.6, left columns). For GlcNAc type (Table 2.5, left column), compounds with relatively small N-acyl groups were good substrates, leading to generation of compounds (UDP-GlcNPr, 2-80) and (UDP-GlcNAz, 2-82). Even the 2-azido-2-deoxy version compound (2-84) bearing no N-acyl moiety was also produced with good yield. Reaction efficiency decreased for compound (UDP-GlcNBu, 2-81), however, with its bulky acyl group, while compound (UDP-GlcNBz, 2-83) was non-isolable and only detectable by MS, demonstrating poor acceptance for bulky substitutions. Interestingly and conversely, none of the N-acyl modified GalNAc-1-P analogues were accepted by GlmU to produce the N-acyl modified UDP-GalNAc analogues (Table 2.6, left column). The changes at both the N-acyl and C-4 configuration may enlarge these sugar-1-P compounds, thus causing them to fail to enter the uridylyltransferase pocket.69
Table 2.5. Isolated yields for enzymatic synthesis of UDP-GlcNAc analogues with GlmU

<table>
<thead>
<tr>
<th>Product Entry</th>
<th>Product Structure</th>
<th>Yield&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Product Entry</th>
<th>Product Structure</th>
<th>Yield&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-8</td>
<td><img src="image" alt="Structure 2-8" /></td>
<td>40%</td>
<td>2-84</td>
<td><img src="image" alt="Structure 2-84" /></td>
<td>48%</td>
</tr>
<tr>
<td>2-80</td>
<td><img src="image" alt="Structure 2-80" /></td>
<td>57%</td>
<td>2-85</td>
<td><img src="image" alt="Structure 2-85" /></td>
<td>50%</td>
</tr>
<tr>
<td>2-81</td>
<td><img src="image" alt="Structure 2-81" /></td>
<td>27%</td>
<td>2-86</td>
<td><img src="image" alt="Structure 2-86" /></td>
<td>20%</td>
</tr>
<tr>
<td>2-82</td>
<td><img src="image" alt="Structure 2-82" /></td>
<td>44%</td>
<td>2-87</td>
<td><img src="image" alt="Structure 2-87" /></td>
<td>20%</td>
</tr>
<tr>
<td>2-83</td>
<td><img src="image" alt="Structure 2-83" /></td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2-88</td>
<td><img src="image" alt="Structure 2-88" /></td>
<td>59%</td>
</tr>
</tbody>
</table>

<sup>a</sup> Isolated yield from DEAE cellulose and P2 gel columns.

<sup>b</sup> N/A (Not available): No product detected by MS or TLC.
Table 2.6. Isolated yields for enzymatic synthesis of UDP-GalNAc analogues with GlmU

<table>
<thead>
<tr>
<th>Product Entry</th>
<th>Product Structure</th>
<th>Yield&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Product Entry</th>
<th>Product Structure</th>
<th>Yield&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-89</td>
<td><img src="image" alt="Structure" /></td>
<td>65%</td>
<td>2-94</td>
<td><img src="image" alt="Structure" /></td>
<td>N/A</td>
</tr>
<tr>
<td>2-90</td>
<td><img src="image" alt="Structure" /></td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2-95</td>
<td><img src="image" alt="Structure" /></td>
<td>55%</td>
</tr>
<tr>
<td>2-91</td>
<td><img src="image" alt="Structure" /></td>
<td>N/A</td>
<td>2-96</td>
<td><img src="image" alt="Structure" /></td>
<td>10%</td>
</tr>
<tr>
<td>2-92</td>
<td><img src="image" alt="Structure" /></td>
<td>N/A</td>
<td>2-88</td>
<td><img src="image" alt="Structure" /></td>
<td>59%</td>
</tr>
<tr>
<td>2-93</td>
<td><img src="image" alt="Structure" /></td>
<td>N/A</td>
<td>2-97</td>
<td><img src="image" alt="Structure" /></td>
<td>N/A</td>
</tr>
</tbody>
</table>

<sup>a</sup> Isolated yield from DEAE cellulose and P2 gel columns.

<sup>b</sup> N/A (Not available): No product detected by MS or TLC.
In contrast, the tolerance for 6-modified GlcNAc-1-P and GalNAc-1-P analogues was quite similar. The yields of smaller 6-deoxy UDP-sugars (compound 2-85 and 2-95) were comparable to that of UDP-GlcNAc (2-8) and decreased dramatically for 6-azido compounds bearing a relatively large substituent (compound 2-86 and 2-96). The 3-epimer of GlcNAc-1-P (2-75) was also converted to the corresponding sugar donor (compound 2-87) but with lower yield than GlcNAc-1-P. (Table 2.5-2.6)

In summary, we have successfully synthesized a set of twelve UDP-GlcNAc/GalNAc analogues with GlmU and investigated the corresponding substrate specificity of this uridylyltransferase. The enzyme showed relaxed tolerance for GlcNAc-1-P modifications (Table 2.5) at N-acyl, C-3, C-4, C-6 positions, with a preference for small substituent groups. The yields were low to moderate (10-65%) and unfortunately, some GalNAc-1-P analogues failed to generate the corresponding UDP-sugars. However, this enzymatic approach which eliminates protection and deprotection steps still exceeds cumbersome chemical synthesis\(^70\) and can quickly produce the desired UDP-sugar analogues in relatively large scale (30-50 mg). Mutants of GlmU with altered active sites or other pyrophosphorylases could be utilized to enrich the donor analogue library which would be covered in the next section (2.3.2).

2.3.2 Substrate specificity investigation of AGX1 and UDP-GlcNAc/GalNAc analogues library

As mentioned above, the recombinant \textit{E.coli} GlcNAc-1-P uridylyltransferase GlmU could not synthesize most of the UDP-GalNAc analogues and the yields for those successful examples were generally low to moderate. One possibility to overcome the
narrow substrate specificity of GlmU and enrich our sugar nucleotide analogue library is to carry out site specific mutations of GlmU active site to accommodate structural diversities. For instance, we could enhance the substrate specificity of C-6 modified compounds by mutations to alleviate the steric crowding around C-6 of the substrate. However, although engineering nucleotidyltransferases are well studied and quite a few examples are available,\textsuperscript{24,71} it is normally tedious and time consuming and it requires the crystal structure information of the target enzyme. Moreover, several mutants need be prepared because of the diversified structure of our substrates.

Alternatively, we could disregard GlmU and search in the UDP-GlcNAc biosynthetic pathways (Figure 2.7) to identify if any other enzymes are available with identical or similar functions as GlmU. Figure 2.7 shows the biosynthesis of UDP-GlcNAc of the prokaryotic and eukaryotic systems. The most significant differences are the orders for phosphate mutations/acetyltransfer reactions; and acetyltransfer and pyrophosphorylation reactions are catalyzed by two different enzymes in eukaryotes.
Specifically, in the formation of UDP-GlcNAc step, bifunctional GlmU serves in bacterial systems. In mammals, the pyrophosphorylase has also been well studied and two isoforms of this enzyme with crystal structures have been identified in humans (AGX1 and AGX2).\textsuperscript{72-73} In yeast, the pyrophosphorylase UAP1 is also a monofunctional enzyme.\textsuperscript{74} Unlike GlmU and AGX1, UAP1 requires no metal ion as a cofactor while using a conserved lysine residue to occupy the metal binding site.
Among the candidates from eukaryotic sources, we are most interested in the uridylyltransferase from humans AGX1 because it was previously reported to have an unusual specificity which accepts both GalNAc-1-P and GlcNAc-1-P in the uridylyltransfer reactions.\(^7^2\) In fact, AGX1 had already been applied to synthesize UDP-GalNAc (2-89) and UDP-N-azidoacetylgalactosamine (UDP-GalNAz, 2-90) in good yields but very small scale (~2 mg).\(^7^5\)

![Figure 2.8a](image)

**Figure 2.8 a.** SDS-PAGE of AGX1 expression and purification. Aliquots were removed from each step of purification and loaded onto a 12% SDS-PAGE gel followed by Coomassie Brilliant Blue staining. Lanes: M, SeeBlue Plus2 Pre-Stained Standard; 1, cell lysate; 2, cell pellet; 3, supernatant of cell lysate; 4, flow through after sample loading (the first time) onto Ni-column; 5, flow through after sample loading (the second time) onto Ni-column; 6, Ni-column purified protein. **b.** Enzymatic synthesis of UDP-GlcNAc/GalNAc derivatives with AGX1.

We thus cloned and expressed human AGX1 (Figure 2.8a) and tested its potency for UDP-GlcNAc/GalNAc derivatives synthesis (Figure 2.8b). With our GlcNAc/GalNAc-1-P analogues especially those that were not accepted by GlmU as substrates, we performed enzymatic pyrophosphorylation reactions also using uridine 5'-
triphosphate (UTP) as the donor molecule (Figure 2.8b). For optimized reaction details please refer to Experimental Section. AGX1 also requires a divalent cation (Mg$^{2+}$ used here) for activity and yeast inorganic pyrophosphatase (PPA) was also added to degrade the byproduct pyrophosphate (PPi) and drive the reaction forward.

To our delight, AGX1 turned out to be more efficient than GlmU since most substrates were fully converted within 0.5-2 h (typically 12 h for GlmU). Thus the AGX1 reactions were monitored by thin-layer chromatography (TLC) and were terminated when complete consumption of the starting material (Sugar-1-Ps) was observed. Reactions yields showing the AGX1 promiscuity were summarized in Table 2.7 (left yield column) and the corresponding reaction yields using GlmU were also included simply for comparison.

Nine GalNAc-1-P analogues and five GlcNAc-1-P analogues (Table 2.7), many of which had low yield or no reaction with GlmU, were tested in this study. As shown in Table 2.7, AGX1 exhibited surprising promiscuity towards both GalNAc- and GlcNAc-based structures. Eleven out of fourteen compounds tested were converted by AGX1 into the corresponding sugar nucleotides in preparative scale (10-50 mg). The natural substrates GalNAc-1-P (entry 2-89), GlcNAc-1-P (entry 2-8), and the 4-deoxy version (entry 2-88) were recognized at similar level ($\geq$72% conversion), excluding the role of 4-hydroxyl group in substrate binding. However, the axial-4-azido entry 2-97 was non-isolable and only detected by TLC and MS, probably resulting from a larger substituent than hydroxyl group.
Table 2.7. Synthesis of UDP-GalNAc/GlcNAc utilizing AGX1 and GlmU

<table>
<thead>
<tr>
<th>Product Entry</th>
<th>Product structure</th>
<th>AGX1 Yield[^a]/Time</th>
<th>GlmU Yield[^a]/12 h</th>
<th>Product Entry</th>
<th>Product structure</th>
<th>AGX1 Yield[^a]/Time</th>
<th>GlmU Yield[^a]/12 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-89</td>
<td><img src="image1" alt="Structure" /></td>
<td>79% 2 h</td>
<td>65% 12 h</td>
<td>2-88</td>
<td><img src="image2" alt="Structure" /></td>
<td>72% 2 h</td>
<td>59% 12 h</td>
</tr>
<tr>
<td>2-91</td>
<td><img src="image3" alt="Structure" /></td>
<td>57% 2 h</td>
<td>N/A[^b]</td>
<td>2-97</td>
<td><img src="image4" alt="Structure" /></td>
<td>&lt;5% &gt;24 h</td>
<td>N/A</td>
</tr>
<tr>
<td>2-92</td>
<td><img src="image5" alt="Structure" /></td>
<td>44% 24 h</td>
<td>N/A</td>
<td>2-8</td>
<td><img src="image6" alt="Structure" /></td>
<td>75% 2 h</td>
<td>40% 12 h</td>
</tr>
<tr>
<td>2-90</td>
<td><img src="image7" alt="Structure" /></td>
<td>64% 0.5 h</td>
<td>N/A</td>
<td>2-80</td>
<td><img src="image8" alt="Structure" /></td>
<td>51% 2 h</td>
<td>57% 12 h</td>
</tr>
<tr>
<td>2-93</td>
<td><img src="image9" alt="Structure" /></td>
<td>N/A &gt;24 h</td>
<td>N/A</td>
<td>2-82</td>
<td><img src="image10" alt="Structure" /></td>
<td>55% 2 h</td>
<td>44% 12 h</td>
</tr>
</tbody>
</table>

Continued
Table 2.7 continued

<table>
<thead>
<tr>
<th>Product Entry</th>
<th>Product structure</th>
<th>AGX1 Yield[\textsuperscript{a}]/Time</th>
<th>GlmU Yield[\textsuperscript{a}]/12 h</th>
<th>Product Entry</th>
<th>Product structure</th>
<th>AGX1 Yield[\textsuperscript{a}]/Time</th>
<th>GlmU Yield[\textsuperscript{a}]/12 h</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2-95</strong></td>
<td><img src="image" alt="image" /></td>
<td>77% 2 h</td>
<td>55%</td>
<td><strong>2-83</strong></td>
<td><img src="image" alt="image" /></td>
<td>N/A &gt;24 h</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>2-96</strong></td>
<td><img src="image" alt="image" /></td>
<td>61% 6 h</td>
<td>10%</td>
<td><strong>2-86</strong></td>
<td><img src="image" alt="image" /></td>
<td>74% 2 h</td>
<td>20%</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Isolated yield from DEAE cellulose and P-2 gel columns. \textsuperscript{b} N/A (Not Available): No product detected by MS or TLC. \textsuperscript{c} Determined by TLC.

Unlike GlmU, which differentiated GalNAc-1-P and GlcNAc-1-P analogues with bigger $N$-acyl modifications, AGX1 was only slightly affected by the bulkiness of $N$-acyl groups in both GalNAc-1-P and GlcNAc-1-P analogues. Briefly, GalNPr/GlcNPr-1-P (entry 2-91 and 2-80) and GalNAz/GlcNAz-1-P (entry 2-90 and 2-82) with relatively smaller $N$-propionyl and $N$-azidoacetyl groups were accepted by AGX1 with good conversion yield (>50%). UDP-GalNBu (entry 2-92) bearing a bulkier $N$-butyryl group had lower yield (44%) and longer reaction time (24 h), while a bulky $N$-benzoyl group prevents both GalNBz-1-P/GlcNBz-1-P to be accepted (entry 2-93 and 2-83).

AGX1 also shows good tolerance to 6-modified GalNAc/GlcNAc-1-P analogues. For example, not only 6-deoxy-GalNAc-1-P (entry 2-95) but also 6-azido-GalNAc-1-P...
and 6-azido-GlcNAc-1-P (entry 2-95 and 2-86) were taken by AGX1 to construct the corresponding UDP-sugars with good yields (>60%). This is another delightful result since GlmU only gave very poor yield for the 6-azido derivatives (10% and 20%).

Table 2.8. Summarized substrate specificity of AGX1 and GlmU

<table>
<thead>
<tr>
<th>Structure motif</th>
<th>AGX1</th>
<th>GlmU</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-2 modified GlcNAc-1-P</td>
<td>Broad specificity</td>
<td>Very broad specificity</td>
</tr>
<tr>
<td></td>
<td>(also see evidence in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chapter 3)</td>
<td></td>
</tr>
<tr>
<td>C-2 modified GalNAc-1-P</td>
<td>Broad specificity</td>
<td>No activity</td>
</tr>
<tr>
<td></td>
<td>Limited specificity</td>
<td></td>
</tr>
<tr>
<td>C-6 modified GlcNAc-1-P</td>
<td>Broad specificity</td>
<td>Limited specificity</td>
</tr>
<tr>
<td>C-6 modified GalNAc-1-P</td>
<td>Both limited specificity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Not affected by the configuration of OH</td>
<td></td>
</tr>
<tr>
<td>C-4 modified</td>
<td>No acceptance for groups larger than OH</td>
<td></td>
</tr>
</tbody>
</table>
The respective substrate specificity of AGX1 and GlmU is summarized in Table 2.8: Firstly, both enzymes had limited acceptance for C-4 modifications. However, the 4-OH configuration does not affect the enzyme activity. Secondly, although very broad specificity of GlmU towards C-2 modified GlcNAc-1-P analogues was observed, including 2-azido-Glc-1-P and 2-keto-Glc-1-P (see Chapter 3 for this compound), it was not a good candidate for the preparation of C-2 and C-6 modified UDP-GalNAc analogues. Last but not least, we proved AGX1 as a better choice for synthesizing most UDP-GalNAc/GlcNAc analogues with C-2, 4, 6 modifications.

In summary, owing to the promiscuity of recombinant human UDP-GalNAc pyrophosphorylase (AGX1), we have successfully prepared C-2, 4, 6 modified UDP-GalNAc/GlcNAc analogues in preparative scale with good yields. This enzyme alleviated us from designing and making GlmU mutants to accommodate the unaccepted substrate structures and finally expanded our sugar nucleotide library. We believe our library of UDP-GalNAc/GlcNAc analogues could greatly facilitate our investigation into the substrate specificity of various glycosyltransferases as well as provide a significant step towards natural product glycodiversification.

2.4 Summary and perspective

We have demonstrated a powerful enzyme, NahK, which can efficiently transform the uncommon GlcNAc/GalNAc analogues to their 1-phosphorylated products regardless of the stereo-configuration at the C-4 position of the sugar ring. More significantly, these uncommon sugar-1-phosphate analogues were enzymatically converted to the valuable corresponding sugar nucleotide donors which undoubtedly will
allow the chemical modification of macromolecules such as glycoproteins or glycolipids using suitable glycosyltransferases. This will give us the opportunity to better understand carbohydrate associated pathways and discover carbohydrate-based drugs. The following directions are currently underway or ultimate goals of this project.

2.4.1 Donor specificity investigation of glycosyltransferases (GTs)

The most common routes that are now used to characterize glycosyltransferase (GT) activity rely on radiometric, chromatographic and fluorometric assays, and have the limitations that they require labels and are tedious. In collaboration with Dr. Mrksich’s group at the University of Chicago, we used a label-free solid phase assay that is based on self-assembled monolayers (SAMs) presenting immobilized carbohydrates (lactose for example) that serve as the acceptors for GT-mediated glycosylation reactions (Figure 2.9). Our modified donor molecules were screened using on-chip reactions on the monolayer with target GT. Then the monolayer were analyzed using matrix assisted laser desorption-ionization mass spectrometry (MALDI MS) to reveal the masses of the unreacted acceptor substrate and the resulting product which can provide a measure of the yield for the glycosylation reactions. This SAMDI assay offers the benefits that is can perform true label-free assays, it requires only small quantities of enzyme and donor reagents and it is compatible with solid phase arrays, and the high throughput that they offer.

Using this technique, a library of eleven UDP-GlcNAc analogues were rapidly screened for their activities as donors for the Neisseria Meningitidis β1,3-N-acetylglucosaminyltransferase (LgtA). The results revealed that LgtA had relaxed
tolerance for C-2, C-4 and C-6 modifications, with bulky groups at these positions decreasing activity of the donors for the enzyme, and also revealed that activity is strongly affected by the stereochemistry at C-3, but not C-4, of the donor. Because GTs are important in the enzymatic synthesis of complex carbohydrates, this knowledge provides a guide to the donors that can be used in LgtA-mediated construction of oligosaccharides. This study is also significant because it demonstrates that SAMDI can be used to both profile GT activities and to provide a quantitative assessment of enzyme activity.

Figure 2.9. Donor specificity investigation of glycosyltransferases with SAMDI-TOF mass spectrometry. The acceptor-presenting SAMs were treated with GTs and donor analogues. (The SAMDI-TOF figure was adapted from Mrksich et al.76)
2.4.2 Carbohydrate metabolic engineering of bacteria

Since cells are coated with a network of polysaccharides and glycoconjugates, carbohydrate metabolic engineering is simply incorporate modified monosaccharide into cell surface polysaccharides (Figure 2.3c). When we feed the bacteria with our structurally modified GlcNAc/GalNAc analogues, the sugar analogue may be accepted by the cell, go through the metabolic pathways, and finally be presented onto cell surfaces. In addition, bioorthogonal groups (azide, ketone, or alkyne) incorporated could be utilized for cell surface labeling using chemical ligations reactions (Figure 2.3b) and allow detection of the incorporated sugar analogues. We have successfully assembled a promiscuous sugar nucleotide biosynthetic pathway in vitro using a kinase and a pyrophosphorylase of bacterial sources which could be ‘copied’ into in vivo systems. As a proof of concept experiment, bacterial polysaccharide remodeling of E.coli with a ketone modified sugar is described in Chapter 3.

2.4.3 In vitro and in vivo glycodiversification

From a chemically prepared structurally modified library of sugar substrates, we constructed a sugar-1-phosphate library and converted it to a sugar nucleotide donor library using our enzymes NahK, GlmU and AGX1. Since sugar nucleotides are activated donor molecules for carbohydrate synthesis, conjugation of the modified sugar structures with diversified acceptors including natural products by suitable glycosyltransferases (GTs) could combinatorially generate novel glycoforms such as glycoconjugate mimics or natural products of new properties. This strategy of assembling glycosylated compounds is also called glycodiversification or glycorandomization (Figure 2.10a).
This technique could quickly and efficiently alter natural product glycosylation and provide the diversity of new therapeutics.

Figure 2.10. In vitro (top) and in vivo (bottom) glycodiversification.

Alternatively, this glycorandomization reaction sequence could be completed inside the cells, called in vivo glycorandomization (Figure 2.10b). Structurally modified sugars and various potential acceptors (aglycon for instance) are fed to bacteria engineered to express NahK, GlmU or AGX1, and a wild type or engineered glycosyltransferase. This UDP-sugar donor ‘factory’ has already been proved feasible in our in vitro experiments. Aglycons conjugated with modified sugars could be isolated from the fermentation media. By using this bacteria machinery, this approach
circumvents the sugar nucleotide synthesis steps as well as enzyme purifications. However, purification of target compounds from complex culture media and large scale production are potential challenges.

2.5 Experimental Section

*Physical Properties and Spectroscopic Measurements.*

Proton nuclear magnetic resonance ($^1$H NMR) spectra and carbon-13 ($^{13}$C NMR) spectra were recorded on a Bruker DPX 400 spectrometer at 400 MHz and 100 MHz respectively, or on a Bruker DRX 500 spectrometer at 500 MHz and 125 MHz respectively. Chemical shifts and coupling constants were reported in ppm and Hz respectively. The high-resolution mass spectra were recorded on a Bruker MicrOTOF spectrometer. Analytical TLC was carried out on silica gel 60 F254 aluminum-backed plates (E. Merck). The 200–400 mesh size of the same absorbent was utilized for all chromatographic purifications. Unless noted, all compounds isolated by chromatography were sufficiently pure by $^1$H NMR analysis for use in subsequent reactions.

![2-acetamido-1,3,4,6-tetra-O-acetyl-2-deoxy-D-glucopyranose (2-4):](attachment:image.png)

To a solution of d-glucosamine hydrochloride (2-3, 8.6 g, 40 mmol) in anhydrous pyridine (70 mL) was added acetic anhydride (70 mL) and the mixture was stirred overnight at room temperature, concentrated, co-evaporated with toluene, and dried under high vacuum. The residue was dissolved in CH$_2$Cl$_2$ (250 mL) and washed with saturated
NaHCO₃ solution (200 mL) and brine (200 mL). The organic layer was dried over NaSO₄ and concentrated in vacuo to give a yellow syrup which was recrystallized in CH₃OH/EtOAc/hexane (4:100:100). White crystal was afforded (6.2 g, α-isomer) and the filtrate (mother liquid) still contains ~3.8 g product. Total yield: ~64%. ¹H NMR (400 MHz, CDCl₃, major isomer α) δ 6.17 (d, J = 3.7 Hz, 1H), 5.54 (d, J = 9.2 Hz, 1H), 5.18-5.26 (m, 2H), 4.45-4.51 (m, 1H), 4.24 (dd, J = 12.6, 4.1 Hz, 1H), 4.06 (dd, J = 12.4, 2.3 Hz, 1H), 3.97-4.01 (m, 1H), 2.19 (s, 3H), 2.08 (s, 3H), 2.05 (s, 3H), 2.04 (s, 3H), 1.93 (s, 3H).

2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-D-glucopyranose (2-5): A solution of the peracetate (2-4, 2 g, 5.136 mmol) and hydrazine acetate (519.7 mg, 5.65 mmol, 1.1 equiv.) in anhydrous DMF (30 mL) was stirred for 40 min at room temperature, then the mixture was concentrated to a dryness and dried at high vacuum to give a yellow solid. Flash chromatography (1:1(v/v) hexane-acetone) and solvent removal gave the product (2-5) as a white foam (1.6 g, 90 %, α:β > 15:1). ¹H NMR (400 MHz, CDCl₃) δ 5.97 (d, J = 9.6 Hz, 1H), 5.23-5.31 (m, 2H), 5.12 (app t, J = 9.4 Hz, 1H), 4.48 (d, J = 2.8 Hz, 1H), 4.27 (dt, J = 9.8, 3.1 Hz, 1H), 4.18-4.22 (m, 2H), 4.09-4.13 (m, 1H), 2.08 (s, 3H), 2.02 (s, 3H), 2.01 (s, 3H), 1.95 (s, 3H).
2-acetamido-3,4,6-tri-O-acety-1-O-[bis(benzyloxy)phosphoryl]-2-deoxy-\(\alpha\)-\(\alpha\)-d-gluctopyranose (2-6): (0.50 g, 1.44 mmol) compound 2-5 and tetrazole (10 equiv, 1.007 g, 14.4 mmol) were dissolved in dry CH\(_2\)Cl\(_2\) (20 mL) and the solution was cooled to -40 °C. Bis(benzyloxy)diisopropylamino)phosphine (5 equiv, 2.49 g, 7.2 mmol) was then added drop wise over a 2 min period. The mixture was stirred for 3 h while allowing the temperature to rise to 25 °C. The mixture was cooled to -70 °C, and 77% m-CPBA (10 equiv, 3.23 g, 14.4 mmol) was added slowly. The solution was stirred for 30 min at 0 °C, then overnight at room temperature. After addition of ethyl acetate (100 mL) the solution was washed with saturated Na\(_2\)SO\(_3\) (50 mL X 3), water (50 mL X 2), and saturated NaCl (50 mL X 2), dried over Na\(_2\)SO\(_4\), and concentrated to yield a yellow solid. The crude product was purified by flash column chromatography (1:1 (v/v) ethyl acetate-hexane, then 100% ethyl acetate) to give product as a white foam (2-6, 600 mg, 69 %).

\(^1\)H NMR (500 MHz, CDCl\(_3\)) \(\delta\) 7.32-7.40 (m, 10H), 5.73 (d, \(J = 9.4\) Hz, 1H), 5.66 (dd, \(J = 6.0, 3.5\) Hz, 1H), 5.02-5.17 (m, 6H), 4.34-4.39 (m, 1H), 4.12 (dd, \(J = 12.6, 4.0\) Hz, 1H), 3.98-4.01 (m, 1H), 3.92 (dd, \(J = 12.6, 2.4\) Hz, 1H), 2.02 (s, 3H), 2.00 (s, 6H), 1.71 (s, 3H). NMR data also see Heidlas et al.\(^{79}\)
2-acetamido-2-deoxy-α-D-glucopyranose-1-phosphate bis-triethylammonium salt (2-7): To a solution of dibenzyl phosphate (2-6, 240 mg, 0.395 mmol) in methanol (20 mL) was added 60 mg of 20 % Pd(OH)_2. The reaction vessel was filled with hydrogen and stirred at room temperature. A total volume of 0.137 mL of triethylamine was added after 30 min. The solution was diluted with 20 mL of MeOH and stirred for another 30 min. The catalyst was filtered off over Celite. The filtrate was concentrated and the residue was redissolved in 12 mL of dry methanol and 78 mg MeONa was added. The reaction mixture was stirred at rt. for 1 h before it was neutralized with ion-exchange resin (Dowex 50w-X8 H^+ form). After filtering to remove the resin, the filtrate was concentrated to provide the GlcNAc-1-P bis-triethylammonium salt as a colorless syrup (2-7, 160 mg, 80% over two steps). NMR data see later in this section.

Uridine 5′-diphospho-2-acetamido-2-deoxy-α-D-glucopyranose diammonium salt (2-8): GlcNAc-1-phosphate bis-triethylammonium salt (2-7, 126 mg, 0.251 mmol) was dissolved together with uridine 5′-monophosphomorpholidate 4-morpholine-N,N′-dicyclohexylcarboxamidine salt (284 mg,
0.414 mmol) in anhydrous pyridine (10 mL) and the solution was concentrated to dryness at 25 °C using oil pump. Ventilation was carried out using argon. The syrup-like residue was dissolved in 1:1 anhydrous pyridine-anhydrous DMF (5 mL) and stirred at room temperature for 7 days under argon. The mixture was concentrated in vacuo to dryness and the residue was dissolved in 2 mL water and separated with DEAE cellulose column (from Sigma) (gradient NH₄HCO₃ in H₂O as eluent, 25 mM–300 mM). The product was detected by MS. The fractions containing desired product were combined and lyophilized to give the crude coupling product as a white powder which was further purified with Biogel P2 column with degassed water as eluent, afforded the uridine 5′-(2-acetamido-2-deoxy-α-D-glucopyranosyl diphosphate as a white powder in form of its ammonium salt (2-8, 56 mg, 35%). NMR data see later in this section and also Whitesides et al.⁷⁹

**General procedure for preparation of N-acyl chain modified GlcNAc/GalNAc.** D-Glucosamine hydrochloride (2-3, 1 g, 4.637 mmol) was dissolved in 5 mL anhydrous methanol in which an equivalent amount of sodium had been added. The mixture was stirred for 10 min before acid anhydride (6.075 mmol, 1.3 equiv.) was added and the resulting reaction mixture was allowed to stir at room temperature for approximately 3 h, the precipitate was filtered off and washed with dry methanol. The combined filtrates and washings were evaporated under reduced pressure and purified by silica gel column chromatography to give the title compound as a white solid.
2-Butyramido-2-deoxy-D-glucopyranose (2-11): Compound 2-11 was prepared as described above. The crude product was purified via silica gel column chromatography (EtOAc:2-propanol:water 9:3:1) to give 2-11 as a white solid (α, β mixture, 61%). $^1$H NMR (400 MHz, D$_2$O) δ 4.98 (d, $J = 3.5$ Hz, 1H, α anomeric), 4.49 (d, $J = 8.7$ Hz, 1H, β anomeric), 3.23-3.69 (m), 2.07 (t), 1.36-1.45 (m), 0.69 (t). NMR data also see Kristova et al.$^{80}$

2-Benzamido-2-deoxy-D-glucopyranose (2-12): Compound 2-12 was prepared as described above, The crude product was purified via silica gel column chromatography (EtOAc:2-propanol:water 9:3:1) to give 2-12 as a white solid (α, β mixture, 53%). $^1$H NMR (400 MHz, D$_2$O) δ 7.42-7.71 (m, aromatic), 5.24 (d, $J = 3.5$ Hz, 1H, α anomeric), 4.77 (d, $J = 8.5$ Hz, 1H, β anomeric), 4.03 (dd, $J = 10.8, 3.7$ Hz, 1H), 3.42-3.85 (m). NMR data also see Humphrey et al.$^{81}$
2-Propionamido-2-deoxy-D-galactopyranose (2-13): Compound 2-13 was prepared as described above. The crude product was purified via silica gel column chromatography (CH$_2$Cl$_2$:MeOH 10:1 to 8:1) to give 2-13 as a white solid (α, β mixture, 75%). $^1$H NMR (500 MHz, D$_2$O) δ 5.27 (d, $J = 3.7$ Hz, 1H, α anomeric), 4.69 (d, $J = 8.4$ Hz, 1H, β anomeric), 3.72-4.20 (m), 2.36 (q, $J = 7.7$ Hz, 2H), 1.17 (t, $J = 7.7$ Hz, 3H); $^{13}$C NMR (125 MHz, D$_2$O, α, β isomer): δ = 179.0, 178.7, 95.5, 91.1, 75.2, 71.1, 70.6, 68.6, 67.9, 67.4, 61.3, 61.0, 53.6, 50.2, 29.4, 29.1, 9.5; HRMS (ESI) calcd for C$_9$H$_{17}$NO$_6$Na (M+Na)$^+$ 258.0948, found 258.0938 m/z.

2-Butyramido-2-deoxy-D-galactopyranose (2-14): Compound 2-14 was prepared as described above. The crude product was purified via silica gel column chromatography (CH$_2$Cl$_2$:MeOH 10:1 to 8:1) to give 2-14 as a white solid (α, β mixture, 83%). $^1$H NMR (500 MHz, D$_2$O): δ = 5.29 (d, $J = 3.7$ Hz, 1H, α anomeric), 4.70 (d, $J = 8.4$ Hz, 1H, β anomeric), 3.73-4.22 (m), 2.33 (q, $J = 7.2$ Hz, 2H), 1.64-1.72 (m, 2H), 0.97 (t, $J = 7.2$ Hz, 3H); $^{13}$C NMR (125 MHz, D$_2$O, α, β isomer): δ = 178.1, 177.8, 95.5, 91.1, 75.2, 71.1, 70.6, 68.7, 68.0, 67.3, 61.3, 61.0, 53.6, 50.2, 38.1, 37.7, 19.0, 12.7; HRMS (ESI) calcd for C$_{10}$H$_{19}$NO$_6$Na (M+Na)$^+$ 272.1105, found 272.1101 m/z.
**2-Benzamido-2-deoxy-D-galactopyranose (2-15):** Compound 2-15 was prepared as described above. The crude product was purified via silica gel column chromatography (CH$_2$Cl$_2$:MeOH 10:1 to 8:1) to give 2-15 as a white solid (α, β mixture, 70%). $^1$H NMR (400 MHz, D$_2$O): δ = 7.78-7.80 (m, 2H), 7.60-7.65 (m, 1H), 7.51-7.55 (m, 2H), 5.37 (d, $J = 3.6$ Hz, 1H, α anomic), 4.81 (d, $J = 8.4$ Hz, 1H, β anomic), 3.73-4.41 (m); $^{13}$C NMR (100 MHz, D$_2$O, α, β isomer): δ = 171.9, 171.6, 133.7, 133.4, 132.2, 132.2, 128.7, 128.7, 127.2, 127.1, 127.1, 95.4, 91.1, 75.2, 71.1, 70.6, 68.7, 68.0, 67.3, 61.2, 61.0, 54.3, 50.9; HRMS (ESI) calcd for C$_{13}$H$_{17}$NO$_6$Na (M+Na)$^+$ 306.0948, found 306.0957 m/z.

**2-azidoacetamido-2-deoxy-D-glucopyranose (GlcNAz, 2-18):** Glucosamine hydrochloride (2-3, 10 g, 45 mmol) was dissolved in MeOH (200 mL) and 1M NaOMe in MeOH was added (45 mL, 45 mmol). The reaction mixture was stirred at rt for 1 h at which point TEA (7 mL, 46.5 mmol) and chloroacetic anhydride (25 g, 146 mmol) were added. The reaction mixture was stirred for another 6 h. Removal of solvent provided chloro-intermediate 2-16 which was partially purified by silica gel chromatography eluting with a gradient of CH$_2$Cl$_2$:MeOH (20:1 to 7:1). The resulting
yellow oil was dissolved in DMF (100 mL), NaN₃ (30 g, 465 mmol) was added and the reaction mixture was heated to 80 °C for 2 h. Upon cooling to rt, the solvent was removed under reduced pressure and the residue was purified by silica gel column chromatography (CH₂Cl₂:MeOH 7:1 to 5:1) to give 2-18 as a white solid. ¹H NMR (400 MHz, CD₃OD) δ 5.11 (d, J = 3.5 Hz, 1H, α anomeric), 4.65 (d, J = 8.1 Hz, 1H, β anomeric), 3.33-4.04 (m).

Galactosamine hydrochloride (2-9, 800 mg, 3.695 mmol) was dissolved in MeOH (10 mL) in which an equivalent amount of sodium had been added. The reaction mixture was stirred at rt for 1 h at which point TEA (0.57 mL, 3.818 mmol) and chloroacetic anhydride (1.58 g, 9.238 mmol) were added. The reaction mixture was stirred for another 6 h. Removal of solvent provided chloro-intermediate 2-17 which was partially purified by silica gel chromatography eluting with a gradient of CH₂Cl₂:MeOH (10:1 to 7:1). The resulting oil was dissolved in DMF (20 mL), NaN₃ (2.4 g, 36.95 mmol) and 15-crown-5 (100 μL) were added and the reaction mixture was heated to 80 °C for 2 h. Upon cooling to rt, the solvent was removed under reduced pressure and the residue was purified by silica gel column chromatography (CH₂Cl₂:MeOH 10:1 to 7:1) to give compound 2-19 as a light yellow solid (600 mg, 62% over two steps). ¹H NMR (400 MHz, D₂O): δ = 5.17 (d, J = 3.7 Hz, 1H, α anomeric), 4.63 (d, J = 8.4 Hz, 1H, β anomeric), 3.66-4.14 (m); ¹³C
NMR (100 MHz, D$_2$O, α, β isomer): δ = 171.2, 170.9, 95.1, 90.9, 75.2, 70.8, 70.6, 68.6, 67.9, 67.3, 61.2, 61.0, 53.8, 52.0, 51.8, 50.4; HRMS (ESI) calcd for C$_8$H$_{14}$N$_4$O$_6$Na (M+Na)$^+$ 285.0806, found 285.0801 m/z.

A mixture of glucosamine hydrochloride (2-3, 2.156 g, 10 mmol), phthalic anhydride (1.481 g, 10 mmol) and sodium bicarbonate (1.68 g, 20 mmol) in H$_2$O (5 mL) was stirred overnight at room temperature. The resulting pale yellow solution was acidified to pH = 2 with 5% HCl, causing a precipitate to form. This was collected by suction filtration, and washed with water and dried in vacuo to give the product 2-20 as a white solid (1.5 g, 49%). $^1$H NMR (500 MHz, DMSO): δ = 7.96 (d, J = 8.0 Hz, 1H), 7.73 (d, J = 7.3 Hz, 1H), 7.48-7.58 (m, 2H), 6.34 (d, J = 3.9 Hz, 1H), 5.07 (brs, 1H), 4.91 (brs, 1H), 4.50 (brs, 1H), 4.42 (brs, 1H), 3.49-3.78 (m, 5H), 3.17-3.21 (m, 1H).

2-Azido-2-deoxy-D-glucopyranose (2-21): Imidazole sulfonyl azide hydrochloride (250 mg, 1.2 mmol) was added to the solution of glucosamine hydrochloride (2-3, 216 mg, 1mmol), K$_2$CO$_3$ (373 mg, 2.7 mmol) and CuSO$_4$·5H$_2$O (2.5 mg) in methanol (5 mL). The reaction mixture was stirred at room temperature for overnight. The mixture was concentrated under reduced pressure and the residue was
purified with silica gel column chromatography (CH$_2$Cl$_2$/CH$_3$OH 10:1) to give **2-21** as a colorless syrup (178 mg, 87%, α, β mixture). $^1$H NMR (400 MHz, D$_2$O): $\delta = 5.39$ (d, $J = 3.6$ Hz, 1H, α anomeric), 4.75 (d, $J = 8.2$ Hz, 1H, β anomeric). HRMS (ESI) calcd for C$_6$H$_{11}$N$_3$O$_5$Na (M+Na)$^+$ 228.0591, found 228.0585 m/z. Also see Chang *et al.* For this compound.$^{82}$

![Image of 2-Azido-2-deoxy-D-galactopyranose (2-22)](image_url)

**2-Azido-2-deoxy-D-galactopyranose (2-22):** Compound was prepared as described for compound **2-21**. (α, β mixture, 73%). $^1$H NMR (400 MHz, D$_2$O): $\delta = 5.43$ (d, $J = 3.7$ Hz, 1H, α anomeric), 4.69 (d, $J = 8.2$ Hz, 1H, β anomeric). Spectral data also see Lemieux *et al.*$^{83}$

![Image of Phenyl Acetamido-2-deoxy-1-thio-β-D-glucopyranoside (2-24)](image_url)

**Phenyl 2-Acetamido-2-deoxy-1-thio-β-D-glucopyranoside (2-24):** A solution of compound **2-4** (1.75 g, 4.494 mmol) in 15 mL CH$_2$Cl$_2$ was treated with thiophenol (0.55 mL, 1.2 equiv.) and BF$_3$.Et$_2$O (0.68 mL, 1.2 equiv.). The reaction mixture was stirred at rt. overnight. The reaction mixture was quenched with sat. NaHCO$_3$ and washed with sat. NaHCO$_3$ and water. The organic layer was dried over Na$_2$SO$_4$ and concentrated to dryness. The residue was purified by silica gel column chromatography (hexanes:acetone 2.5:1) to afford a white foam which was dissolved in 50 mL anhydrous methanol and CH$_3$ONa in methanol (0.5 M, 0.5 mL) was added. The
reaction mixture was stirred at rt. for overnight and the solvent was removed to give compound 2-24 as a white solid (1.09 g, 90%). $^1$H NMR (400 MHz, DMSO): $\delta$ = 7.83 (d, $J$ = 9.4 Hz, 1H), 7.39-7.41 (m, 2H), 7.27-7.31 (m, 2H), 7.18-7.22 (m, 1H), 5.06 (t, $J$ = 5.4 Hz, 2H), 4.75 (d, $J$ = 10.5 Hz, 1H), 4.59 (t, $J$ = 5.7 Hz, 1H), 3.71 (ddd, $J$ = 11.9, 5.4, 1.7 Hz, 1H), 3.60 (q, $J$ = 9.8 Hz, 1H), 3.43-3.49 (m, 1H), 3.11-3.23 (m, 2H), 1.82 (s, 3H).

Phenyl 4,6-O-Benzylidene-2-acetamido-2-deoxy-1-thio-β-D-glucopyranoside (2-25): To a solution of thio glycoside compound 2-24 (330 mg, 1.053 mmol) in dry DMF (5 mL) was added benzaldehyde dimethyl acetyl (0.32 mL, 2.11 mmol) and p-toluenesulfonic acid (20 mg, cat.), and the solution was stirred for 13 h at rt. The solvent was removed under reduced pressure and the residue was purified by silica gel column chromatography (CH$_2$Cl$_2$:MeOH 40:1) to give 2-25 as a white solid (360 mg, 85%). $^1$H NMR (500 MHz, DMSO-$d_6$) $\delta$ 7.98 (d, $J$ = 9.0 Hz, 1H), 7.31-7.45 (m, 9H), 7.23-7.26 (m, 1H), 5.61 (s, 1H), 5.42 (d, $J$ = 5.5 Hz, 1H), 4.97 (d, $J$ = 10.2 Hz, 1H), 4.19-4.22 (m, 1H), 3.62-3.74 (m, 3H), 3.46-3.53 (m, 2H), 1.84 (s, 3H); $^{13}$C NMR (125 MHz, DMSO-$d_6$) $\delta$ 169.3, 137.7, 134.5, 129.5, 129.1, 128.9, 128.1, 126.7, 126.4, 100.7, 86.4, 80.9, 71.7, 70.0, 67.7, 54.9, 23.1; HRMS (ESI) calcd for C$_{21}$H$_{23}$NO$_5$SNa (M+Na)$^+$ 424.1189, found 424.1176 m/z.
Phenyl 4,6-\(\text{O}\)-Benzyldene-1-thio-\(\beta\)-d-allopyranosido\[2,3:4',5']-2'-methyl-2'-oxazoline (2-26): Compound 2-25 (300 mg, 0.747 mmol) was dissolved in 5 mL anhydrous pyridine and a solution of trifluoromethanesulfonic anhydride (0.15 mL, 0.896 mmol) in CH\(_2\)Cl\(_2\) (2 mL) was added at 0 °C. The reaction mixture was stirred at rt. for 1 h then was diluted with CH\(_2\)Cl\(_2\) (50 mL) and washed with water (2*50 mL). The organic layer was collected and dried over Na\(_2\)SO\(_4\). The solvent was removed under reduced pressure and the residue was purified by silica gel column chromatography (Hexane:EtOAc 4:1 to 3:1) to give 2-26 as a white solid (195 mg, 68%). \(^1\text{H}\) NMR (500 MHz, CDCl\(_3\)) \(\delta\) 7.50-7.53 (m, 4H), 7.29-7.39 (m, 6H), 5.63 (s, 1H), 4.79-4.82 (m, 2H), 4.39 (dd, \(J = 9.9, 4.6\) Hz, 1H), 4.33 (dd, \(J = 9.6, 3.7\) Hz, 1H), 4.26-4.29 (m, 1H), 3.83-3.88 (m, 1H), 3.80 (t, \(J = 10.0\) Hz, 1H), 2.09 (s, 3H); \(^{13}\text{C}\) NMR (125 MHz, CDCl\(_3\)) \(\delta\) 168.3, 137.0, 133.5, 132.1, 129.4, 129.2, 128.5, 128.0, 126.4, 102.8, 86.8, 76.4, 75.2, 69.6, 68.9, 65.3, 14.4; HRMS (ESI) calcd for C\(_{21}\)H\(_{21}\)NO\(_4\)SNa (M+Na\(^+\)) 406.1083, found 406.1071 \(m/z\).

(2-27): Compound 2-25 (300 mg, 0.747 mmol) was dissolved in 5 mL anhydrous pyridine and a solution of trifluoromethanesulfonic anhydride (0.38 mL, 2.24 mmol, 3 equiv.) in CH\(_2\)Cl\(_2\) (2 mL) was added at 0 °C. The reaction mixture was stirred at rt. for 1 h then was diluted with CH\(_2\)Cl\(_2\) (50 mL) and washed with water (2*50 mL).
mL). The organic layer was collected and dried over Na₂SO₄. The solvent was removed under reduced pressure and the residue was purified by silica gel column chromatography (Hexane:EtOAc 4:1 to 3:1) to give 2-27 as a white solid (357 mg, 93%). ¹H NMR (400 MHz, CDCl₃) δ 7.56-7.60 (m, 2H), 7.47-7.50 (m, 2H), 7.34-7.41 (m, 6H), 5.58 (s, 1H), 4.66-4.70 (m, 2H), 4.43-4.48 (m, 2H), 4.33-4.36 (m, 2H), 3.91-3.97 (m, 1H), 3.83 (dd, J = 9.7, 2.9 Hz, 1H), 3.78 (t, J = 10.3 Hz, 1H); ¹³C NMR (100 MHz, CDCl₃) δ 148.7, 136.4, 133.5, 131.0, 129.5, 129.1, 128.7, 128.4, 126.2, 102.8, 84.0, 75.6, 75.3, 75.2, 68.7, 66.5, 61.3; HRMS (ESI) calcd for C₂₂H₂₀F₃NO₆S₂Na (M+Na)⁺ 538.0576, found 538.0577 m/z.

Phenyl 2-Aacetamido-2-trifluoromethanesulfonamido-2-deoxy-1-thio-β-D-allopyranoside (2-28): Compound 2-27 (350 mg, 0.678 mmol) was dissolved in THF (10 mL) and treated with 3 N HCl/H₂O (3 mL). The reaction was stirred at rt. for 24 h and neutralized with NaHCO₃. The solvent was removed and the residue was purified by silica gel column chromatography (CH₂Cl₂:MeOH 10:1 to 5:1) to give 2-28 as a white solid (257 mg, 85%). ¹H NMR (400 MHz, DMSO-d₆) δ 7.56-7.59 (m, 2H), 7.26-7.35 (m, 3H), 5.60 (app t, J = 3.0 Hz, 1H), 5.00 (d, J = 10.3 Hz, 1H), 3.64-3.87 (m, 5H), 2.13 (s, 3H); ¹³C NMR (100 MHz, DMSO-d₆) δ 172.1, 135.5, 132.4, 130.2, 128.6, 85.7, 79.0, 74.5, 67.2, 62.7, 57.3, 21.0; HRMS (ESI) calcd for C₁₅H₁₈F₃NO₆S₂Na (M+H)⁺ 468.0369, found 468.0390 m/z.
Phenyl 2-Aacetamido-2-deoxy-1-thio-β-D-allopyranoside (2-29):

Compound 2-26 (187 mg, 0.488 mmol) was dissolved in THF (10 mL) and treated with 3 N HCl/H₂O. The reaction was stirred at rt. for 24 h and neutralized with NaHCO₃. The solvent was removed and the residue was purified by silica gel column chromatography (CH₂Cl₂:MeOH 10:1 to 5:1) to give 2-29 as a white solid (140 mg, 92%). ¹H NMR (400 MHz, DMSO-d₆) δ 7.43-7.46 (m, 2H), 7.23-7.35 (m, 3H), 5.10 (d, J = 4.1 Hz, 1H), 4.93 (d, J = 7.0 Hz, 1H), 4.71 (d, J = 10.1 Hz, 1H), 4.30 (dd, J = 11.9, 1.9 Hz, 1H), 4.04 (dd, J = 11.9, 7.0 Hz, 1H), 3.82-3.84 (m, 1H), 3.75-3.79 (m, 1H), 3.29-3.34 (m, 1H), 2.52 (m, 1H, in solvent), 2.02 (s, 3H); ¹³C NMR (100 MHz, DMSO-d₆) δ 170.2, 134.2, 130.2, 128.9, 126.7, 86.7, 73.6, 70.8, 67.8, 64.0, 53.8, 20.7; HRMS (ESI) calcd for C₁₄H₂₀NO₅S (M+H)+ 314.1057, found 314.1055 m/z.

2-Aacetamido-2-deoxy-D-allopyranoside (2-23): Compound 2-29 was deprotected as described for compound 2-41, white solid (pyranose α, β and furanose mixture, 90%). ¹H NMR (400 MHz, D₂O) δ 5.18 (d, J = 3.7 Hz, 1H, pyranose α anomeric), 4.99 (d, J = 8.7 Hz, 1H, pyranose β anomeric); HRMS (ESI) calcd for C₈H₁₆NO₆ (M+H)+ 222.0972, found 222.0979 m/z. NMR data also see Okumura et al.⁴³
2-Acetamino-2-deoxy-4,6-O-phenylmethylene-1,3-di-O-acetyl-α-D-glucopyranose (2-32): N-acetyl-D-glucosamine (2-1, 884 mg, 4 mmol), anhydrous DMF (19 mL), benzaldehyde dimethyl acetal (0.719 mL, 1.2 equiv) and p-toluenesulfonic acid monohydrate (56 mg) were placed in a 100 mL round bottom flask. This was attached to a rotary evaporator under diminished pressure and then rotated in a water bath at 50 °C. The suspension was rotated for 3 h until a clear solution formed. Progress of the reaction was checked by TLC (10% MeOH in CH₂Cl₂). NaHCO₃ solid (100 mg) was added to the mixture and it was stirred for 0.5 h at rt. The solvent was removed under reduced pressure and NaHCO₃ solution was added to the residue to form white precipitate which was filtered and washed with water and diethyl ether, then dried under reduced pressure before it was acetylated in pyridine and acetic anhydride at -40 °C. The mixture was evaporated under reduced pressure and the residue was recrystallized from ethanol to afford the title compound 2-32 (1.1 g) in 70% yield totally.

¹H NMR (400 MHz, CDCl₃) δ = 7.34-7.46 (m, 5H), 6.15 (d, J = 3.8 Hz, 1H), 5.67 (d, J = 9.0 Hz, 1H), 5.54 (s, 1H), 5.32 (dd, J = 10.6, 9.7 Hz, 1H), 4.44-4.50 (m, 1H), 4.31 (dd, J₁ = 4.8, J₂ = 10.4 Hz, 1H), 3.90-3.96 (m, 1H), 3.75-3.82 (m, 2H), 2.19 (s, 3H), 2.09 (s, 3H), 1.95 (s, 3H). NMR data also see Yonehara et al.⁸⁴
2-Acetamino-2-deoxy-1,3-di-O-acetyl-α-D-glucopyranose (2-33):  

MeOH (30%) in CH₂Cl₂ was added 3% AcCl and stirred for 30 min at 25 °C. Compound 2-32 (767 mg, 1.95 mmol) was added to this mixture and stirred until the starting material completely disappeared (<1 h). The reaction solution was diluted with CH₂Cl₂ (100 mL) and washed three times with sat. NaHCO₃. The organic phase was evaporated under reduced pressure and purified by silica gel column chromatography (CH₂Cl₂:MeOH 20:1) to obtain the diol 2-33 as a white powder (536 mg, 90%). ¹H NMR (500 MHz, CDCl₃): δ = 6.11 (d, J = 3.7 Hz, 1H), 5.81 (d, J = 8.8 Hz, 1H), 5.10 (dd, J = 11.0, 9.3 Hz, 1H), 4.29 (ddd, J = 11.0, 8.8, 3.7 Hz, 1H), 3.88 (t, J = 9.6 Hz, 1H), 3.79-3.85 (m, 2H), 3.70 (dt, J = 10.0, 3.2 Hz, 1H), 2.14 (s, 3H), 2.10 (s, 3H), 1.90 (s, 3H). NMR data also see Chaplin et al.⁸⁵

2-Acetamino-2-deoxy-1,3,6-tri-O-acetyl-α-D-glucopyranose (2-34): The diol 2-33 (536 mg) was dissolved in CH₂Cl₂ (15 mL) and pyridine (0.141 mL 1 equiv.) and acetic anhydride (1.1 equiv.) was added dropwise at -20 °C. The mixture was stirred over night at room temperature and concentrated under reduced pressure. The resulting residue was purified by silica gel column chromatography (CH₂Cl₂:MeOH 30:1) to afford the title compound 2-34 (440 mg, 72%). ¹H NMR (500 MHz, CDCl₃): δ = 6.15 (d, J = 3.7 Hz, 1H), 5.89 (d, J = 9.0 Hz, 1H), 5.13 (dd, J = 11.0, 9.2 Hz, 1H), 4.53 (dd, J₁ = 3.7, J₂ =
12.3 Hz, 1H), 4.33-4.38 (m, 1H), 4.22 (dd, $J = 12.3$, 2.2 Hz, 1H), 3.84-3.87 (m, 1H), 3.66 (dd, $J_1 = J_2 = 9.4$ Hz, 1H), 2.17 (s, 3H), 2.12 (s, 3H), 2.11 (s, 3H), 1.93 (s, 3H). NMR data also see Chaplin et al.\textsuperscript{85}

2-Acetamino-4-azido-2,4-dideoxy-1,3,6-tri-$O$-acetyl-$\alpha$-D-galactose (2-35): To the solution of compound 2-34 (440 mg, 1.27 mmol) in 19 mL CH$_2$Cl$_2$ was added anhydrous pyridine (0.387 mL, 4.8 mmol) and the mixture was cooled to -10 $^\circ$C. Then triflic anhydride (469 mg, 0.273 mL, 1.3 equiv.) was added through a syringe slowly. The reaction was stirred for 30 min at -10 $^\circ$C and 2 h at room temperature. The mixture was directly loaded onto a silica gel column. After flash chromatography (CH$_2$Cl$_2$), the orange syrup was redissolved in DMF (11.5 mL) and to this solution was added NaN$_3$ (832 mg, 12.8 mmol, 10 equiv.). The resulting mixture was stirred at room temperature for 18 h before the solvent was evaporated under reduced pressure. The residue was purified by flash column chromatography (2% MeOH/CH$_2$Cl$_2$) to give the 4-azido-triacetate (2-35, 357 mg, 76%). \textsuperscript{1}H NMR (400 MHz, CDCl$_3$): $\delta = 6.14$ (d, $J = 3.7$ Hz, 1H), 5.49 (d, $J = 9.2$ Hz, 1H), 5.30 (dd, $J_1 = 3.4$, $J_2 = 11.3$ Hz, 1H), 4.79 (ddd, $J_1 = 3.7$, $J_2 = 9.3$, $J_3 = 11.2$ Hz, 1H), 4.10-4.26 (m, 3H), 4.03 (m, 1H), 2.16 (s, 6H), 2.08 (s, 3H), 1.95 (s, 3H); \textsuperscript{13}C NMR (100 MHz, CDCl$_3$): $\delta = 171.3$, 170.5, 170.1, 168.8, 91.3, 70.3, 68.5, 62.7, 60.3, 47.3, 23.3, 21.0, 20.8, 20.7; HRMS (ESI) calcd for C$_{14}$H$_{20}$N$_4$O$_8$Na (M+Na)$^+$ 395.1173, found 395.1165 m/z.
2-Acetamino-4-azido-2,4-dideoxy-D-galactose (2-31): The triacetate 2-35 was treated with MeONa in dry MeOH to afford the title compound 2-31 (204 mg, 86%, α:β = 1:0.7). $^1$H NMR (400 MHz, D$_2$O): δ = 5.09 (d, J = 3.5 Hz, 1H), 4.53 (d, J = 8.4 Hz, 0.7H), 4.09 (m, 1.7H), 4.02 (m, 1.7H), 3.92 (m, 1.7H), 3.77 (dd, J$_1$=J$_2$ = 8.46 Hz, 0.7H), 3.63 (m, 4.4H), 1.94 (s, 5.1H); $^{13}$C NMR (D$_2$O, 100 MHz): δ = 174.9, 174.6, 95.3, 90.9, 73.5, 71.4, 69.0, 67.7, 63.0, 62.0, 61.1, 60.9, 53.7, 50.4, 22.1, 21.9; HRMS (ESI) calcd for C$_8$H$_{14}$N$_4$O$_5$Na (M+Na)$^+$ 269.0856, found 269.0855 m/z.

2-acetamino-2-deoxy-4-O-(phenorythiocarbonyl)-1,3,6-tri-O-acetyl-D-glucopyranose (2-36): To the mixture of compound 2-34 (700 mg, 2.02 mmol), DMAP (444 mg, 3.64 mmol) and acetonitrile (anhydrous, 8 mL) was added phenyl chlorothionoformate (416 mg, 0.334 mL, 2.42 mmol), and the solution was flushed with nitrogen and stirred at room temperature. After 18 h, the reaction mixture was filtered and the precipitate was washed with ethyl acetate. The combined filtrate and washings were concentrated under reduced pressure to a thick yellow oil which was purified on a silica gel column (2% MeOH/CH$_2$Cl$_2$). Pure product 2-36 was obtained as a white powder (690 mg 71%).$^1$H NMR (400MHz, CDCl$_3$): δ = 7.43 (m, 2H), 7.32 (m, 1H), 7.05 (m, 2H), 6.21 (d, J = 3.4 Hz, 1H), 5.86 (dd, J$_1$ = J$_2$ = 9.7 Hz, 1H), 5.64 (d, J = 9.0 Hz, 1H), 5.44 (dd, J$_1$ = J$_2$ = 10.3 Hz, 1H), 4.60 (m, 1H), 4.39 (dd, J$_1$ = 3.3, J$_2$ = 12.4 Hz, 1H).
Hz, 1H), 4.18 (m, 2H), 2.22 (s, 3H), 2.11 (s, 3H), 2.09 (s, 3H), 1.97 (s, 3H); $^{13}$C NMR (100MHz, CDCl$_3$): $\delta$ = 194.2, 171.6, 170.6, 170.0, 168.6, 153.4, 129.7, 126.9, 121.6, 90.6, 76.6, 70.8, 69.4, 61.4, 51.0, 23.1, 20.9 (2 carbons), 20.8; HRMS (ESI) calcd for C$_{21}$H$_{25}$NO$_{10}$NaS (M+Na)$^+$ 506.1091, found 506.1073 m/z.

2-acetamino-2,4-dideoxy-1,3,6-tri-O-acetyl-α-D-glucopyranose (2-37): Tributylstannane (1.6 g, 1.48 mL, 5.52 mmol) and AIBN (100 mg) were added to thiocarbonate 2-36 (666 mg, 1.38 mmol) in dry toluene (78 mL) and the mixture was stirred at 60 °C for 4 h. Concentration of the mixture and flash chromatography (2% MeOH/CH$_2$Cl$_2$) gave the 4-deoxy-triacetate 2-37 as a colorless powder (342mg, 75%).$^1$H NMR (CDCl$_3$, 400 MHz): $\delta$ = 6.19 (d, $J$ = 3.6 Hz, 1H), 5.51 (d, $J$ = 8.8 Hz, 1H), 5.20 (dt, $J$ = 11.2, 4.8 Hz, 1H), 4.30-4.35 (m, 1H), 4.09-4.17 (m, 3H), 2.17 (s, 3H), 2.09 (s, 3H), 2.08 (s, 3H), 2.07-2.09 (m, 1H), 1.95 (s, 3H), 1.78 (dd, $J$ = 24.0, 12.0 Hz, 1H).

2-Acetamino-2,4-dideoxy-D-hexopyranose (2-30): The triacetate 2-37 was treated with MeONa in dry MeOH to afford the title compound 2-30 (175 mg, 83%, α:β 1:1 as indicated by $^1$H NMR). $^1$H NMR (400 MHz, D$_2$O, α:β 1:1): $\delta$ 5.12 (d, $J$ = 3.6 Hz, 1H, α anomeric), 4.50 (d, $J$ = 8.4 Hz, 1H, β anomeric). NMR data also see Berkin et al.$^{86}$
Phenyl 4,6-O-Benzylidene-1-thio-β-D-glucopyranosido[2,3:4′,5′]-2′-methylthio-2′-oxazoline (2-39): To a solution of compound 2-25 (300 mg, 0.747 mmol) in THF (5 mL) was added NaH (60 mg, 1.49 mmol) and the mixture was stirred at rt. for 0.5 h. Imidazole (3 mg, cat.) was added and 10 min later CS$_2$ (0.38 mL) was added and the mixture was stirred at rt. for another 1 h. MeI (0.1 mL, 1.60 mmol) was added at last and the reaction was stirred at rt. for overnight. The reaction was diluted with 50 mL CH$_2$Cl$_2$ and washed with water (2*20 mL). The organic layer was collected and dried over Na$_2$SO$_4$. The solvent was removed under reduced pressure and the residue was purified by silica gel column chromatography (Hexane:EtOAc 4:1 to 3:1) to give 2-39 as a white solid (150 mg, 48%). $^1$H NMR (400 MHz, CDCl$_3$) δ 7.58-7.61 (m, 2H), 7.47-7.49 (m, 2H), 7.31-7.38 (m, 6H), 5.60 (s, 1H), 4.96 (d, $J = 10.1$ Hz, 1H), 4.40 (dd, $J = 10.5$, 4.7 Hz, 1H), 4.20 (dd, $J = 13.1$, 9.7 Hz, 1H), 4.07 (dd, $J = 9.7$, 8.5 Hz, 1H), 3.93 (app t, $J = 10.4$ Hz, 1H), 3.45-3.54 (m, 2H), 2.52 (s, 3H); $^{13}$C NMR (100 MHz, CDCl$_3$) δ 170.7, 136.6, 133.2, 131.7, 129.3, 129.0, 128.3, 128.2, 126.2, 101.5, 87.3, 85.4, 80.3, 72.3, 70.4, 68.6, 14.4; HRMS (ESI) calcd for C$_{21}$H$_{21}$NO$_4$S$_2$Na (M+Na)$^+$ 438.0804, found 438.0816 m/z.

Phenyl 2-Acetamido-2-deoxy-6-O-methylsulfonyl-1-thio-β-D-glucopyranoside (2-43): Compound 2-24 (476 mg, 1.52 mmol) was dissolved in
anhydrous pyridine (8 mL) and the solution was cooled to -15 °C. Mesyl chloride (125 μL, 1.60 mmol) was added dropwise. After 2h, the reaction was quenched by adding methanol and the solvent was removed and the resulting residue was purified by silica gel column chromatography (CH$_2$Cl$_2$:MeOH 20:1 to 15:1) to give 2-43 as a white solid (438 mg, 74%). $^1$H NMR (400 MHz, CD$_3$OD) δ 7.48-7.50 (m, 2H), 7.26-7.32 (m, 3H), 4.85 (d, 1H, in H$_2$O peak), 4.54 (dd, $J = 11.3$, 1.8 Hz, 1H), 4.37 (dd, $J = 11.3$, 6.0 Hz, 1H), 3.77 (t, $J = 10.0$, 1H), 3.54-3.58 (m, 1H), 3.50 (t, $J = 8.9$ Hz, 1H), 3.31-3.38 (m, 1H), 3.03 (s, 3H), 2.00 (s, 3H); $^{13}$C NMR (100 MHz, CD$_3$OD) δ 173.7, 135.4, 132.7, 130.2, 128.6, 88.2, 79.2, 77.3, 71.6, 71.0, 56.3, 37.7, 23.1; HRMS (ESI) calcd for C$_{15}$H$_{21}$NO$_7$S$_2$Na (M+Na)$^+$ 414.0652, found 414.0636 m/z.

Phenyl 2-Acetamido-2,6-dideoxy-1-thio-β-D-glucopyranoside (2-44): Compound 2-43 (100 mg, 0.255 mmol) was dissolved in dry DMSO (5 mL) and NaBH$_4$ (94 mg, 25.5 mmol) was added. The reaction was heated at 85-95 °C for 18 h. The reaction was quenched by adding 5% acetic acid solution at 0 °C and the solution was diluted with more water and extracted several times with ethyl acetate (200 mL total). The combined extracts were evaporated under vacuo and the residue was purified by silica gel column chromatography (CH$_2$Cl$_2$:MeOH 20:1 to 15:1) to give 2-44 as a white solid (66 mg, 87%). $^1$H NMR (400 MHz, CD$_3$OD) δ 7.43-7.47 (m, 2H), 7.25-7.32 (m, 3H), 4.74 (d, $J = 10.3$ Hz, 1H), 3.75 (t, $J = 10.0$ Hz, 1H), 3.42 (dd, $J = 9.7$, 8.9 Hz, 1H), 3.32 (m, 1H, in solvent peak), 3.06 (t, $J = 9.1$ Hz, 1H), 1.99 (s, 3H), 1.30 (d, $J = 6.2$, 3H); $^{13}$C NMR (100 MHz, CD$_3$OD) δ 173.7, 135.7, 132.6, 130.0, 128.5, 88.2, 77.6, 77.3 (two
carbons), 56.7, 23.1, 18.5; HRMS (ESI) calcd for C_{14}H_{19}NO_{4}SNa (M+Na)^{+} 320.0927, found 320.0919 m/z.

2-40 2-Acetamido-2,6-dideoxy-\(\beta\)-D-glucopyranose (2-40): The thio glycoside (2-44) was deprotected as described for compound 2-41, white solid (\(\alpha\) and \(\beta\) mixture, 90%). \(^1\)H NMR (400 MHz, D_{2}O) \(\delta\) 5.09 (d, \(J = 3.6\) Hz, 1H, \(\alpha\) anomeric), 4.63 (d, \(J = 8.4\) Hz, 1H, \(\beta\) anomeric), 3.14-3.88 (m), 1.99 (s, 3H), 1.22 (d, \(J = 6.2\) Hz, 3H). Also see Arita \textit{et al.} for this compound.\(^{51}\)

Phenyl 2-Acetamido-6-azido-2,6-dideoxy-1-thio-\(\beta\)-D-glucopyranoside (2-45): Thio glycoside 2-24 (626 mg, 2 mmol) was dissolved in anhydrous pyridine (10 mL) and a solution of \(p\)-toluenesulfonyl chloride (380 mg, 2 mmol) in CH_{2}Cl_{2} (4.0 mL) was added at 0 \(^\circ\)C and the reaction mixture was allowed to stir at room temperature. After 2 h, another amount of \(p\)-toluenesulfonyl chloride (380 mg, 2 mmol) was added. The reaction was monitored by TLC and was diluted by EtOAc (100 mL) at completion and then washed with 1N HCl (50 mL), sat. NaHCO_{3} (50 mL), and brine (50 mL). The organic layer was dried over Na_{2}SO_{4} and concentrated \textit{in vacuo}. The residue was redissolved in anhydrous DMF (10 mL) and treated with NaN_{3} (520 mg, 8 mmol) and 18-crown-6 (60 mg). The resulting mixture was stirred at 40-50 \(^\circ\)C for 48 h before the solvent was removed under reduced pressure. The residue was purified by
silica gel column chromatography (CH$_2$Cl$_2$:MeOH 25:1 to 20:1) to give 2-45 as a white solid (520 mg, 77% over two steps). $^1$H NMR (400 MHz, CD$_3$OD) δ 7.52-7.54 (m, 2H), 7.28-7.34 (m, 3H), 4.80 (d, $J = 10.4$ Hz, 1H), 3.76 (t, $J = 10.0$ Hz, 1H), 3.58 (m, 1H), 3.40-3.50 (m, 3H), 3.31 (m, 1H, in solvent peak), 2.03 (s, 3H), $^{13}$C NMR (100 MHz, CD$_3$OD) δ 173.5, 134.9, 133.2, 129.9, 128.6, 88.2, 80.5, 77.2, 72.6, 56.2, 53.0, 23.0; HRMS (ESI) calcd for C$_{14}$H$_{18}$N$_4$O$_4$SNa (M+Na)$^+$ 361.0941, found 361.0929 m/z.

2-Acetamido-6-azido-2,6-dideoxy-D-glucopyranose (2-41): Compound 2-45 (208 mg, 0.612 mmol) was dissolved in acetone/H$_2$O (v/v 9:1 10 mL) and treated with N-bromosuccinimide (NBS) (240 mg, 1.347 mmol) and the reaction was stirred at room temperature for 1 h. Upon completion, NaHCO$_3$ was added to neutralize the reaction and the solvent was removed and the residue was purified by silica gel column chromatography (CH$_2$Cl$_2$:MeOH 25:1 to 20:1) to give 2-41 as a white solid (α and β mixture, 130 mg, 86%). $^1$H NMR (400 MHz, CD$_3$OD) δ 5.10 (d, $J = 3.5$ Hz, 1H), 3.95 (ddd, $J = 9.9$, 5.8, 2.3 Hz, 1H), 3.86 (dd, $J = 10.6$, 3.5 Hz, 1H), 3.68 (dd, $J = 10.7$, 8.9 Hz, 1H), 3.51 (dd, $J = 13.1$, 2.4 Hz, 1H), 3.33-3.42 (m, 2H), 1.99 (s, 3H).

Phenyl 2,6-di-Acetamido-2,6-dideoxy-3,4-di-O-acetyl-1-thio-β-D-glucopyranoside (2-46): To a solution of compound 2-45 (100 mg, 0.296 mmol) in THF
(5 mL), was added Ph₃P (85 mg, 0.325 mmol) and water (80 μL). The mixture was stirred at rt for 24 h. The solvent was removed under reduced pressure. The residue was dissolved in 5 mL anhydrous pyridine and acetic anhydride 1.0 mL was added. The mixture was stirred at room temperature for overnight. Upon completion, the solution was diluted by 50 mL EtOAc and washed H₂O, sat. NaHCO₃ and brine. The organic layer was dried over Na₂SO₄ and concentrated in vacuo. The residue was purified by silica gel column chromatography (CH₂Cl₂:MeOH 50:1 to 40:1) to give 2-46 as a white foam (94 mg, 72 % over two steps). ¹H NMR (400 MHz, CDCl₃) δ 7.41-7.45 (m, 2H), 7.24-7.27 (m, 3H), 6.00-6.06 (m, 2H), 5.17 (app t, J = 10.0 Hz, 1H), 4.87-4.93 (m, 2H), 4.13-4.20 (m, 1H), 3.60-3.71 (m, 2H), 3.11-3.17 (m, 1H), 2.05 (s, 3H), 2.02 (s, 3H), 1.99 (s, 3H), 1.88 (s, 3H); ¹³C NMR (100 MHz, CDCl₃) δ 171.3, 170.4, 170.3, 170.0, 133.0, 132.2, 129.2, 128.1, 86.8, 76.7, 73.8, 69.7, 53.5, 40.2, 23.5, 23.3, 20.9 (two carbons); HRMS (ESI) calcd for C₂₀H₂₆N₂O₇SNa (M+Na)⁺ 461.1353, found 461.1332 m/z.

2,6-di-Acetamido-2,6-dideoxy-D-glucopyranose (2-42): Compound 2-46 (85 mg, 0.194 mmol) was dissolved in anhydrous methanol (5 mL) and the solution of MeONa in methanol (0.5 M, 0.3 mL) was added. The reaction solution was stirred at room temperature for 2h. Cation exchange resin (H form) was added to adjust the pH to ~7 and the resin was removed by filtration and washed several times with methanol. The combined filtrates were concentrated under reduced pressure and the resulting residue was dissolved in acetone/H₂O (v/v 9:1 5 mL) and treated with N-bromosuccinimide (76
mg, 0.426 mmol) and the reaction was stirred at room temperature for 2 h. Upon completion, the solvent was removed and the residue was purified by silica gel column chromatography (CH$_2$Cl$_2$:MeOH 10:1 to 5:1) to give **2-42** as a white solid (α and β mixture, 32 mg, 63% over two steps). $^1$H NMR (400 MHz, D$_2$O, α and β mixture) δ 5.22 (d, $J = 3.5$ Hz, 1H, α anomeric H), 4.73 (d, $J = 8.5$ Hz, 0.7H, β anomeric H), 3.90-3.96 (m, 2H), 3.37-3.81 (m, 8H), 2.06-2.09 (m, 10H); HRMS (ESI) calcd for C$_{10}$H$_{18}$N$_2$O$_6$Na (M+Na)$^+$ 285.1057, found 285.1052 m/z. Also see Ogawa et al. for this compound.$^{34}$

1,3,4,6-tetra-O-acetyl-2-Acetamido-2-deoxy-α-D-galactopyranoside (2-49): A solution of N-acetyl galactosamine (2-2, 2 g, 9.276 mmol) in pyridine (14 mL) was treated with acetic anhydride (9 mL) and the reaction mixture was stirred at room temperature overnight. The reaction was cooled to 10°C and poured into ice water to liberate white powder. The powder was filtered and washed with water and cold ethanol to give peracetylated GalNAc 2-49 (3.24 g, 90%). $^1$H NMR (CDCl$_3$, 400 MHz) δ 6.19 (d, $J = 3.5$ Hz, 1H), 5.48 (d, $J = 9.1$ Hz, 1H), 5.39 (d, $J = 2.8$ Hz, 1H), 5.18 (dd, $J = 11.5, 3.2$ Hz, 1H), 4.66-4.71 (m, 1H), 4.21 (app t, $J = 6.6$ Hz, 1H), 4.01-4.10 (m, 2H), 2.14 (s, 6H), 2.00 (s, 3H), 1.99 (s, 3H), 1.92 (s, 3H).
Phenyl 2-Acetamido-2-deoxy-3,4,6-tri-O-acetyl-1-thio-β-D-galactopyranoside (2-50): A solution of peracetylated GalNAc 2-49 (1.75 g, 4.494 mmol) in 15 mL CH₂Cl₂ was treated with thiophenol (0.55 mL, 1.2 equiv.) and BF₃·Et₂O (0.68 mL, 1.2 equiv.) at 0 °C. The reaction mixture was stirred at rt. for 3 days. The reaction mixture was quenched with sat. NaHCO₃ and washed with sat. NaHCO₃ and water. The organic layer was dried over Na₂SO₄ and concentrated to dryness. The residue was purified by silica gel column chromatography (hexanes:acetone 2.5:1) to afford compound 2-50 as a white foam (1.4 g, 71%). ¹H NMR (CDCl₃, 400 MHz) δ 7.51-7.54 (m, 2H), 7.30-7.35 (m, 3H), 5.39-5.41 (m, 2H), 5.22 (dd, J = 3.2, 10.8 Hz, 1H), 4.93 (d, J = 10.4 Hz, 1H), 4.12 - 4.26 (m, 3H), 3.95 (dt, J = 1.0, 6.2 Hz, 1H), 2.15 (s, 3H), 2.05 (s, 3H), 2.02 (s, 3H), 2.01 (s, 3H). NMR data also see Deng et al.⁸⁷

Phenyl 2-Acetamido-2-deoxy-1-thio-β-D-galactopyranoside (2-51): Peracetylated thio glycoside (2-50) (922 mg, 2.098 mmol) was dissolved in 50 mL anhydrous methanol and CH₃ONa in methanol (0.5 M, 0.5 mL) was added. The reaction mixture was stirred at rt. for overnight and the solvent was removed to give compound 2-51 as a white solid (650 mg, >98%). ¹H NMR (500 MHz, DMSO-d₆): δ = 7.73 (d, J = 9.4 Hz, 1H), 7.39-7.41 (m, 2H), 7.27-7.30 (m, 2H), 7.18-7.21 (m, 1H), 4.71-4.73 (m, 2H), 4.63 (t, J = 5.6 Hz, 1H), 4.58 (d, J = 4.3 Hz, 1H), 3.92 (dd, J = 19.7, 10.1 Hz, 1H), 3.73-
3.74 (m, 1H), 3.43-3.55 (m, 4H), 1.82 (s, 3H); $^{13}$C NMR (125 MHz, DMSO-$d_6$): $\delta =$ 169.3, 136.0, 128.8, 128.7, 126.0, 86.6, 79.3, 72.3, 67.6, 60.6, 50.5, 23.1; HRMS (ESI) calcd for C$_{14}$H$_{19}$NO$_5$SNa (M+Na)$^+$ 336.0876, found 336.0877 m/z.

Phenyl 2-Acetamido-6-azido-2,6-dideoxy-1-thio-$\beta$-D-galactopyranoside (2-52): Thio glycoside 2-51 (300 mg, 0.957 mmol) was dissolved in anhydrous pyridine (8 mL) and a solution of $p$-toluenesulfonyl chloride (219 mg, 1.15 mmol) in CH$_2$Cl$_2$ (2.0 mL) was added at 0 °C and the reaction mixture was allowed to stir at room temperature. After 2h, another amount of $p$-toluenesulfonyl chloride (207 mg, 1.085 mmol) was added. The reaction was monitored by TLC and was diluted by EtOAc (100 mL) at completion and then washed with 1N HCl (50 mL), sat. NaHCO$_3$ (50 mL), and brine (50 mL). The organic layer was dried over Na$_2$SO$_4$ and concentrated under vacuo. The residue was redissolved in anhydrous DMF (10 mL) and treated with NaN$_3$ (282 mg, 4.338 mmol) and 15-crown-5 (30 μL). The resulting mixture was stirred at 40-50 °C for 3 days before the solvent was removed under reduced pressure. The residue was purified by silica gel column chromatography (CH$_2$Cl$_2$:MeOH 30:1 to 25:1) to give compound 2-52 as a white solid (110 mg, 34% over two steps). $^1$H NMR (500 MHz, CD$_3$OD): $\delta =$ 7.50-7.51 (m, 2H), 7.23-7.30 (m, 3H), 4.78 (d, $J =$ 10.3 Hz, 1H), 4.06 (t, $J =$ 10.3 Hz, 1H), 3.80 (d, $J =$ 2.8 Hz, 1H), 3.60-3.66 (m, 3H), 3.34 (dd, $J =$ 11.6, 2.8 Hz, 1H), 2.00 (s, 3H); $^{13}$C NMR (125 MHz, CD$_3$OD): $\delta =$ 173.9, 135.8, 132.4, 129.9, 128.3,
88.9, 79.0, 74.0, 70.1, 52.9, 52.6, 23.0; HRMS (ESI) calcd for C\textsubscript{14}H\textsubscript{18}N\textsubscript{4}O\textsubscript{4}SNa (M+Na)\textsuperscript{+} 361.0941, found 361.0926 \textit{m/z}.

2-Acetamido-6-azido-2,6-dideoxy-D-galactopyranose (2-48): Compound 2-52 (90 mg, 0.266 mmol) was dissolved in acetone/H\textsubscript{2}O (v/v 8:2 6 mL) and treated with \textit{N}-bromosuccinimide (NBS) (104 mg, 0.585 mmol) and the reaction was stirred at room temperature for 36 h. Upon completion, the solvent was removed and the residue was purified by silica gel column chromatography (CH\textsubscript{2}Cl\textsubscript{2}:MeOH 15:1 to 13:1) to give compound 2-48 as a white solid (\(\alpha\) and \(\beta\) mixture, 46 mg, 71\%). \textit{\textsuperscript{1}}H NMR (400 MHz, D\textsubscript{2}O): \(\delta = 5.25\) (d, \(J = 3.5\) Hz, 1H, \(\alpha\) anomeric), 4.68 (d, \(J = 8.4\) Hz, 1H, \(\beta\) anomeric), 3.47-4.24 (m, 12H), 2.06 (s, 6H); \textit{\textsuperscript{13}}C NMR (100 MHz, D\textsubscript{2}O, \(\alpha\), \(\beta\) isomer): \(\delta = 175.0, 174.7, 95.4, 91.1, 73.5, 71.0, 69.0, 68.9, 68.2, 67.2, 53.5, 51.0, 50.8, 50.1, 22.2, 21.9\); HRMS (ESI) calcd for C\textsubscript{8}H\textsubscript{14}N\textsubscript{4}O\textsubscript{5}Na (M+Na)\textsuperscript{+} 269.0856, found 269.0852 \textit{m/z}.

Phenyl 2-Acetamido-2-deoxy-6-\textit{O}-methylsulfonyl-1-thio-\(\beta\)-D-galactopyranoside (2-53): Thioglycoside 2-51 (300 mg, 0.957 mmol) was dissolved in anhydrous pyridine (8 mL) and the solution was cooled to -15 \textdegree\text{C}. Mesyl chloride (81 \textmuL, 1.05 mmol) was added dropwise. After 2h, the reaction was quenched by adding methanol and the solvent was removed. The resulting residue was purified by silica gel
column chromatography (CH$_2$Cl$_2$:MeOH 25:1 to 20:1) to give compound 2-53 as a white solid (225 mg, 60%). $^1$H NMR (500 MHz, DMSO-$d_6$): $\delta$ = 7.78 (d, $J$ = 9.4 Hz, 1H), 7.40-7.41 (m, 2H), 7.28-7.31 (m, 2H), 7.20-7.23 (m, 1H), 4.97 (d, $J$ = 4.3 Hz, 1H), 4.92 (d, $J$ = 6.1 Hz, 1H), 4.81 (d, $J$ = 10.4 Hz, 1H), 4.26-4.34 (m, 2H), 3.94 (app q, $J$ = 10.0 Hz, 1H), 3.83-3.85 (m, 1H), 3.76-3.77 (m, 1H), 3.53-3.57 (m, 1H), 3.07 (s, 3H), 1.83 (s, 3H); $^{13}$C NMR (125 MHz, DMSO-$d_6$): $\delta$ = 169.4, 135.3, 129.0, 128.9, 126.3, 86.1, 75.9, 71.7, 70.3, 67.6, 50.2, 36.7, 23.1; HRMS (ESI) calcd for C$_{15}$H$_{21}$NO$_7$S$_2$Na (M+Na)$^+$ 414.0652, found 414.0664 m/z.

Phenyl 2-Acetamido-2,6-dideoxy-1-thio-$\beta$-D-galactopyranoside (2-54): Compound 2-53 (144 mg, 0.368 mmol) was dissolved in dry DMSO (3 mL) and NaBH$_4$ (145 mg, 3.83 mmol) was added. The reaction was heated at 85-95 °C for 18 h. The reaction was quenched by adding 5% acetic acid solution at 0 °C and the solution was concentrated by oil pump. The residue was purified by silica gel column chromatography (CH$_2$Cl$_2$:MeOH 20:1 to 15:1) to give compound 2-54 as a white solid (25 mg, 23 %). $^1$H NMR (500 MHz, CD$_3$OD): $\delta$ = 7.45-7.47 (m, 2H), 7.23-7.30 (m, 3H), 4.74 (d, $J$ = 10.4 Hz, 1H), 4.02 (t, $J$ = 10.3 Hz, 1H), 3.59-3.67 (m, 3H), 1.99 (s, 3H), 1.28 (d, $J$ = 6.4, 3H); $^{13}$C NMR (125 MHz, CD$_3$OD): $\delta$ = 173.9, 136.2, 132.1, 129.9, 128.1, 88.5, 76.1, 74.4, 72.4, 52.6, 23.0, 17.1; HRMS (ESI) calcd for C$_{14}$H$_{19}$NO$_4$SNa (M+Na)$^+$ 320.0927, found 320.0919 m/z.
2-Acetamido-2,6-dideoxy-D-galactopyranose (2-47): Compound 2-54 was deprotected as described for compound 2-41, white solid (α and β mixture, 90%). $^1$H NMR (500 MHz, D$_2$O, α and β mixture ~1:1): δ = 5.23 (d, $J = 3.7$ Hz, 1H, α anomeric), 4.66 (d, $J = 8.5$ Hz, 1H, β anomeric), 4.27 (dd, $J = 13.4$, 6.5 Hz, 1H), 4.14 (dd, $J = 10.9$, 3.5 Hz, 1H), 3.97 (dd, $J = 10.9$, 3.1 Hz, 1H), 3.80-3.90 (m, 4H), 3.76 (dd, $J = 10.6$, 3.1 Hz, 1H), 2.10 (s, 6H), 1.32 (d, $J = 6.4$ Hz, 3H), 1.28 (d, $J = 6.7$ Hz, 3H); HRMS (ESI) calcd for C$_{8}$H$_{15}$NO$_{5}$Na (M+Na)$^+$ 228.0842, found 228.0837 m/z. Spectral data also see Horton et al.$^{88}$

2-Acetamino-2-deoxy-4,6-O-phenylmethylene-1,3-di-O-acetyl-α-D-galactopyranose (2-55): A mixture of N-acetylgalactoamine (2-2, 1 g, 4.524 mmol) in CH$_3$CN (7 mL) was treated with benzaldehyde dimethyl acetal (1.4 mL, 9.05 mmol) and p-toluenesulfonic acid monohydrate (30 mg), and the reaction mixture was stirred for overnight at rt. The reaction was filtered and the resulting white solid was washed with 15 mL CH$_2$Cl$_2$ and 15 mL water and dried overnight under reduced pressure. The white solid was redissolved in 10 mL pyridine and acetic anhydride (5 mL) was added at 0 °C. The reaction was stirred at rt. For overnight and the solvent was removed under reduced pressure upon completion. The residue was recrystallized in cold methanol to give the
fully protected compound 2-55 as a white solid. $^1$H NMR (500 MHz, CDCl$_3$): $\delta$ 7.52-7.54 (m, 2H), 7.36-7.40 (m, 3H), 6.34 (d, $J = 3.5$ Hz, 1H), 5.55 (s, 1H), 5.41 (d, $J = 9.1$ Hz, 1H), 5.26 (dd, $J = 11.7$, 3.4 Hz, 1H), 4.91 (ddd, $J = 11.5$, 9.1, 3.4 Hz, 1H), 4.35 (d, $J = 2.7$ Hz, 1H), 4.29 (dd, $J = 12.5$, 1.4 Hz, 1H), 4.04 (dd, $J = 12.7$, 1.8 Hz, 1H), 3.80 (brs, 1H), 2.17 (s, 3H), 2.12 (s, 3H), 1.94 (s, 3H); $^{13}$C NMR (125 MHz, CDCl$_3$): $\delta = 172.1$, 169.9, 169.0, 137.5, 129.3, 128.4, 126.5, 101.1, 92.4, 73.5, 69.1, 68.7, 64.6, 47.1, 23.4, 21.2 (two carbons); HRMS (ESI) calcd for C$_{19}$H$_{23}$NO$_8$Na (M+Na)$^+$ 416.1316, found 416.1317 m/z.

2-Acetamido-2-deoxy-1,3-di-O-acetyl-\(\alpha\)-D-galactopyranose (2-56): To a solution of the above crystal (2-55) in 3:1 CH$_2$Cl$_2$/CH$_3$OH (20 mL) was added \(p\)-toluenesulfonic acid monohydrate until pH = 3 was obtained. The reaction mixture was stirred at rt. for overnight. The reaction was neutralized with TEA and concentrated under reduced pressure and the residue was purified by column chromatography on silica gel (CH$_2$Cl$_2$/CH$_3$OH 20:1) to afford 2-56 as a white solid (755 mg, 55% from \(N\)-acetylgalactoamine). $^1$H NMR (400 MHz, CDCl$_3$): $\delta$ 6.21 (d, $J = 3.6$ Hz, 1H), 5.69 (d, $J = 9.2$ Hz, 1H), 5.15 (dd, $J = 11.3$, 3.0 Hz, 1H), 4.82 (ddd, $J = 11.3$, 9.1, 3.7 Hz, 1H), 4.19 (app d, $J = 1.8$ Hz, 1H), 3.83-3.91 (m, 4H), 3.10 (brs, 1H), 2.15 (s, 3H), 2.13 (s, 3H), 1.94 (s, 3H). NMR data also see Jiaang et al.$^{89}$
2-Acetamido-2-deoxy-1,3,4-tri-O-acetyl-6-O-p-toluenesulfonfyl-α-D-galactopyranose (2-57): To a solution of diol 2-56 (715 mg, 2.342 mmol) in pyridine (10 mL) was added p toluenesulfonfyl chloride (536 mg, 2.81 mmol) at 0 °C, and the mixture was allowed to warm to room temperature over 4h. The mixture was diluted with ethyl acetate and washed with aq NaHCO₃. The organic layer was dried over Na₂SO₄ and concentrated in vacuo. The residue was redissolved in acetic anhydride and pyridine (1 mL:2 mL), and the mixture was stirred at room temperature for 7h. Concentration of the mixture gave yellow syrup which was chromatographed on silica gel (hexane:EtOAc 1:1.2) to give compound 2-57 as a white foam (900 mg, 77% over two steps). ¹H NMR (500 MHz, CDCl₃): δ = 7.74 (d, J = 8.1 Hz, 2H), 7.34 (d, J = 8.3 Hz, 2H), 6.15 (d, J = 3.8 Hz, 1H), 5.50 (d, J = 9.0 Hz, 1H), 5.42 (d, J = 2.5 Hz, 1H), 5.18 (dd, J = 11.5, 3.2 Hz, 1H), 4.62-4.67 (m, 1H), 4.27 (app t, J = 6.8 Hz, 1H), 3.96-4.02 (m, 2H), 2.44 (s, 3H), 2.15 (s, 3H), 2.05 (s, 3H), 2.00 (s, 3H), 1.92 (s, 3H); ¹³C NMR (125 MHz, CDCl₃): δ = 171.2, 170.2, 170.0, 168.9, 145.4, 132.3, 130.1, 128.2, 91.1, 68.7, 67.7, 66.6, 66.2, 47.0, 23.2, 21.8, 21.0, 20.8, 20.6; HRMS (ESI) calcd for C₂₁H₂₇NO₁₁SNa (M+Na)⁺ 524.1197, found 524.1219 m/z.

2-Acetamido-2,6-dideoxy-1,3,4-tri-O-acetyl-α-D-galactopyranose (2-58): To a solution of tosylate 2-57 (340 mg, 0.678 mmol) in anhydrous dimethoxyethane
(7 mL) was added NaI (1016 mg, 6.78 mmol), and the mixture was stirred at 85 °C for 10h. After cooling to room temperature, the reaction mixture was diluted with ethyl acetate (150 mL) and washed with brine. The organic phase was dried and then concentrated in vacuo. After drying by oil pump, the residue was dissolved in dry toluene (30 mL) and deoxygenated. The solution was treated with tributyltin hydride (547 μL, 2.034 mmol) and 2,2'-azobisisobutyronitrile (18 mg). The reaction mixture was then stirred under reflux for 4 h. Upon cooling to room temperature, the mixture was concentrated in vacuo and the residue was purified by silica gel column chromatography (hexane/EtOAc 1:2) to afford compound 2-58 as a colorless syrup (189 mg, 84 % over two steps). ¹H NMR (500 MHz, CDCl₃): δ = 6.17 (d, J = 3.7 Hz, 1H), 5.42 (d, J = 9.2 Hz, 1H, NH), 5.21-5.25 (m, 2H), 4.67-4.72 (m, 1H), 4.41 (app q, J = 6.6 Hz, 1H), 2.19 (s, 3H), 2.16 (s, 3H), 2.02 (s, 3H), 1.94 (s, 3H), 1.15 (d, J = 6.6 Hz, 3H); ¹³C NMR (125 MHz, CDCl₃): δ = 171.5, 170.8, 170.2, 169.2, 91.8, 70.1, 68.4, 67.3, 47.1, 23.3, 21.2, 20.9, 20.8, 16.2; HRMS (ESI) calcd for C₁₄H₂₁NO₆Na (M+Na)⁺ 354.1159, found 354.1161 m/z.

2-Acetamido-2,6-dideoxy-D-galactopyranose (2-47): The triacetate 2-58 (173 mg) was treated with NaOCH₃/CH₃OH at room temperature to afford compound 2-47 as a colorless syrup in 71% yield.
2-Acetamido-6-azido-2,6-dideoxy-1,3,4-tri-O-acetyl-α-D-galactopyranose (2-59): To a solution of compound 2-57 (200 mg, 0.399 mmol) in DMF (5 mL) was added NaN₃ (259 mg, 3.99 mmol) and 15-crown-5 (15 μL). The solution was heated to 50 °C and stirred for 5 days. The solvent was removed under reduced pressure and the residue was purified by silica gel column chromatography (hexane/EtOAc 1:2) to afford compound 2-59 as a colorless syrup (74 mg, 50%). ¹H NMR (400 MHz, CDCl₃): δ 6.23 (d, J = 3.7 Hz, 1H), 5.44 (d, J = 9.2 Hz, 1H), 5.40 (dd, J = 3.2, 1.2 Hz, 1H), 5.21 (dd, J = 11.5, 3.2 Hz, 1H), 4.73 (ddd, J = 11.5, 9.2, 3.6 Hz, 1H), 4.08-4.12 (m, 1H), 3.43 (dd, J = 12.7, 7.3 Hz, 1H), 3.21 (dd, J = 12.7, 5.6 Hz, 1H), 2.19 (s, 3H), 2.18 (s, 3H), 2.03 (s, 3H), 1.95 (s, 3H). NMR data also see Hang et al.⁹⁰

Cloning and overexpression of NahK

Bacterial Strains and Plasmids

*Escherichia coli* DH5α [lacZΔM15 hsdR recA] was purchased from Gibco-BRL (Gaithersburg, MD). BL21 (DE3) [F⁻ompT hsdS₆(r⁻B m⁻B) gal dcm (DE3)] and plasmid pET22b were purchased from Novagen (Carlsbad, CA). Plasmid pCR2.1-BL1642 was kindly provided by Dr. Kitaoka in National Food Research Institute of Japan.

Construction of Recombinant Plasmid

The *nahK* gene from *Bifidobacterium longum* JCM1217 was cut out from plasmid pCR2.1-BL1642 with *Nde*I and *Xho*I, and ligated into pET22b at *Nde*I/XhoI sites.
recombinant plasmid was transformed into *E. coli* DH5α cells for DNA sequencing. After sequence confirmation, the plasmid was transformed into *E. coli* BL21 (DE3) for protein expression.

**Overexpression and Purification of NahK**

*E. coli* BL21 (DE3) transformant containing the pET22b-*nahK* plasmid was grown in LB broth containing 100 μg/ml ampicillin at 37 °C. When OD600 reached 0.7, IPTG was added to a final concentration of 0.1 mM to induce protein to express at 16 °C for 20 h. The cells were harvested by centrifugation at 4,000 × g for 30 min. The cell pellet was resuspended in 20 mM sodium phosphate (pH 7.4) and sonicated on ice. The lysate was clarified by centrifugation at 20,000 × g for 40 min at 4 °C and the supernatant was applied to HisTrap™ HP column. The column was then washed with washing buffer (20 mM sodium phosphate, pH 7.4, 100 mM NaCl, 10 mM imidazole) and the enzyme was eluted with elution buffer (20 mM sodium phosphate, pH 7.4, 100 mM NaCl, 500 mM imidazole). The eluant was desalted by HiTrap™ Desalting HP column with 50 mM Tris-HCl buffer, pH 7.5. The fractions containing the purified enzyme were combined and concentrated with Amicon Ultra and the enzyme was stored in 20% glycerol (50 mM Tris-HCl, pH 7.5) at -20 °C.

**Activity Measurement**

The activity of recombinant NahK was assayed in a cocktail containing 50 mM Tris-HCl, pH 7.5, 5 mM MgCl2, 5 mM GlcNAc, 5 mM ATP, and various amount of NahK incubated at 37 °C for 30 min. The reaction was terminated by boiling the mixture for 2
min followed by centrifugation. The supernatant was analyzed by normal phase silica gel thin-layer chromatography [n-butanol/acetic acid/water = 2:1:1 (v/v/v)].

**General procedure for preparation of GlcNAc/GalNAc-1-phosphate (GlcNAc/GalNAc-1-P) and their analogues using NahK.** The reaction mixture (10 mL) contained 40 mM GlcNAc or its analogues, 50 mM ATP, 10 mM MgCl₂, and 1.5 mg/mL NahK in 100 mM Tris-HCl buffer (pH 9.0). Progress of the reaction was analyzed by normal phase silica gel thin-layer chromatography [n-butanol:acetic acid:water 2:1:1 (v:v:v)]. After incubation at 37 °C for 19 h, the mixture was briefly boiled for 2 min and then centrifuged to remove protein. The supernatant was concentrated under reduced pressure and the residue was purified by normal phase silica gel column chromatography using gradient CH₂Cl₂/5 mM NH₄HCO₃ in methanol as eluent (CH₂Cl₂:5 mM NH₄HCO₃ in methanol 1:1 to 0:1). The fractions containing the products were collected and concentrated under reduced pressure.

![2-Acetamido-2-deoxy-α-D-glucosyl diammonium Phosphate](image)

**2-Acetamido-2-deoxy-α-D-glucosyl diammonium Phosphate**

(2-7): Compound 2-7 was prepared using NahK as described above. Colorless oil, 90% yield. ¹H NMR (400 MHz, D₂O) δ 5.45 (dd, J = 7.5, 3.3 Hz, 1H), 3.99-4.04 (m, 2H), 3.96 (dd, J = 12.4, 2.3 Hz, 1H), 3.83-3.90 (m, 2H), 3.57 (app t, J = 9.6 Hz, 1H), 2.14 (s, 3H).
2-Propionamido-2-deoxy-α-D-glucosyl diammomium Phosphate (2-60): Compound 2-60 was prepared using NahK as described above. Colorless oil, 86% yield. $^1$H NMR (400 MHz, D$_2$O) δ 5.55 (dd, $J = 7.4$, 3.3 Hz, 1H), 4.00-4.12 (m, 3H), 3.92-3.97 (m, 2H), 3.67 (app t, $J = 9.7$ Hz, 1H), 2.47 (q, $J = 7.8$ Hz, 2H), 1.27 (t, $J = 7.8$ Hz, 3H). NMR data also see Lazarevic et al.$^{91}$

2-Butyramido-2-deoxy-α-D-glucosyl diammomium Phosphate (2-61): Compound 2-61 was prepared using NahK as described above. Colorless oil, 87% yield. $^1$H NMR (400 MHz, D$_2$O) δ 5.52 (dd, $J = 7.2$, 3.4 Hz, 1H), 3.96-4.09 (m, 3H), 3.87-3.92 (m, 2H), 3.63 (app t, $J = 9.6$ Hz, 1H), 2.40 (t, $J = 7.5$ Hz, 2H), 1.67-1.78 (m, 2H), 1.03 (t, $J = 7.4$ Hz, 3H). NMR data also see Lazarevic et al.$^{91}$

2-benzamido-2-deoxy-α-D-glucosyl diammomium Phosphate (2-62): Compound 2-62 was prepared using NahK as described above.
Colorless oil, 88% yield. $^1$H NMR (400 MHz, D$_2$O) $\delta$ 7.88-7.90 (m, 2H), 7.56-7.69 (m, 3H), 5.58 (dd, $J = 7.4$, 3.3 Hz, 1H), 4.23-4.26 (m, 1H), 3.94-4.06 (m, 3H), 3.86 (dd, $J = 12.3$, 4.7 Hz, 1H), 3.63 (app t, $J = 9.6$ Hz, 1H); $^{13}$C NMR (100 MHz, D$_2$O) $\delta$ 171.6, 133.6, 132.2, 128.7, 127.5, 93.2 (d), 72.5, 71.1, 70.0, 60.6, 54.6 (d); $^{31}$P NMR (162 MHz, D$_2$O) $\delta$ 0.71; HRMS (ESI) calcd for C$_{13}$H$_{17}$NO$_9$P (M-H) 362.0646, found 362.0643 m/z.

2-Azidoacetamido-2-deoxy-$\alpha$-D-glucosyl diammonium Phosphate (2-63): Compound 2-63 was prepared using NahK as described above. Colorless oil, 87% yield. $^1$H NMR (400 MHz, D$_2$O) $\delta$ 5.30 (dd, $J = 7.4$, 3.3 Hz, 1H), 3.97 (two doublets, $J = 16.3$ Hz, 2H, -CH$_2$-), 3.88 (ddd, $J = 10.6$, 3.3, 2.3 Hz, 1H), 3.78-3.83 (m, 1H), 3.76 (dd, $J = 12.4$, 2.4 Hz, 1H), 3.65-3.73 (m, 2H), 3.41 (dd, $J = 10.0$, 9.2 Hz, 1H). NMR data also see Vocadlo et al. $^{27}$

2-Azido-2-deoxy-$\alpha$-D-glucosyl diammonium Phosphate (2-64): Compound 2-64 was prepared using NahK as described above only with longer reaction time (64.5 h) and additional enzyme added after the first 48 h (final enzyme concentration 2.25 mg/mL). Colorless oil, 58% yield. $^1$H NMR (400 MHz, D$_2$O) $\delta$ 5.59 (dd, $J = 7.7$, 3.3 Hz, 1H), 3.88-4.00 (m, 3H), 3.80 (dd, $J = 12.4$, 5.0 Hz, 1H), 3.53 (app t, J
= 9.6 Hz, 1H), 3.35 (ddd, J = 10.4, 3.0, 2.3 Hz, 1H); $^{13}$C NMR (100 MHz, D$_2$O) δ 93.4 (d), 72.4, 70.7, 69.8, 63.1 (d), 60.5; $^{31}$P NMR (162 MHz, D$_2$O): δ = 0.15; HRMS (ESI) calcd for C$_6$H$_{11}$N$_3$O$_8$P (M-H) $^-$ 284.0289, found 284.0300 m/z.

![2-Acetamido-2-deoxy-α-D-galactosyl diammonium](image)

**Phosphate (2-65):** Compound 2-65 was prepared using NahK as described above. Colorless oil, 78% yield. $^1$H NMR (400 MHz, D$_2$O) δ 5.37 (dd, J = 7.5, 3.5 Hz, 1H), 4.11-4.16 (m, 2H), 3.96-3.97 (m, 1H), 3.89 (dd, J = 10.9, 3.1 Hz, 1H), 3.66-3.75 (m, 2H), 2.01 (s, 3H).

![2-Propionamido-2-deoxy-α-D-galactosyl diammonium](image)

**Phosphate (2-66):** Compound 2-66 was prepared using NahK as described above. Colorless oil, 85% yield. $^1$H NMR (400 MHz, D$_2$O) δ 5.12 (dd, J = 7.4, 3.4 Hz, 1H), 3.88 (dt, J = 10.9, 2.9 Hz, 1H), 3.81-3.84 (m, 1H), 3.70-3.71 (m, 1H), 3.63 (dd, J = 10.9, 3.1 Hz, 1H), 3.38-3.48 (m, 2H), 2.00 (q, J = 7.6 Hz, 2H), 0.79 (t, J = 7.6 Hz, 3H). NMR data also see Lazarevic et al.$^{92}$
2-Butyramido-2-deoxy-\(\alpha\)-D-galactosyl diammonium Phosphate (2-67): Compound 2-67 was prepared using NahK as described above. Colorless oil, 86% yield. \(^1\)H NMR (400 MHz, D\(_2\)O) \(\delta\) 5.35 (dd, \(J = 7.3, 3.5\) Hz, 1H), 4.12 (dt, \(J = 11.0, 2.9\) Hz, 1H), 4.05-4.08 (m, 1H), 3.93-3.94 (m, 1H), 3.85 (dd, \(J = 11.0, 3.1\) Hz, 1H), 3.61-3.71 (m, 2H), 2.20 (t, \(J = 7.2\) Hz, 2H), 1.48-1.57 (m, 2H), 0.82 (t, \(J = 7.4\) Hz, 3H). NMR data also see Lazarevic et al.\(^92\)

2-benzamido-2-deoxy-\(\alpha\)-D-galactopyranosyl diammonium Phosphate (2-68): Compound 2-68 was prepared using NahK as described above. Colorless oil, 77% yield. \(^1\)H NMR (400 MHz, D\(_2\)O): \(\delta\) = 7.83-7.86 (m, 2H), 7.61-7.66 (m, 1H), 7.52-7.56 (m, 2H), 5.59 (dd, \(J = 7.5, 3.5\) Hz, 1H), 4.45-4.49 (m, 1H), 4.23-4.26 (m, 1H), 4.09-4.15 (m, 2H), 3.77-3.83 (m, 2H); \(^{13}\)C NMR (125 MHz, D\(_2\)O): \(\delta\) = 171.8, 133.6, 132.2, 128.7, 127.5, 93.6 (d), 71.7, 68.7, 67.7, 61.3, 50.7 (d); \(^{31}\)P NMR (162 MHz, D\(_2\)O): \(\delta\) = -0.10; HRMS (ESI) calcd for C\(_{13}\)H\(_{17}\)NO\(_9\)P (M-H)\(^-\) 362.0646, found 362.0660 m/z.
**2-Azidoacetamido-2-deoxy-α-D-galactopyranosyl diammonium Phosphate (2-69):** Compound 2-69 was prepared using NahK as described above. Colorless oil, 65% yield. $^1$H NMR (500 MHz, D$_2$O): $\delta$ = 5.30 (app brs, 1H), 4.10-4.15 (m, 2H), 4.04 (d, $J$ = 16.2 Hz, 1H), 3.94 (d, $J$ = 16.2 Hz, 1H), 3.91 (d, $J$ = 3.1 Hz, 1H), 3.86 (dd, $J$ = 10.8, 3.1 Hz, 1H), 3.61-3.70 (m, 2H); $^{13}$C NMR (125 MHz, D$_2$O): $\delta$ = 171.0, 92.8 (d), 71.1, 68.7, 68.1, 61.3, 51.7, 50.4 (d); $^{31}$P NMR (162 MHz, D$_2$O): $\delta$ = -0.74; HRMS (ESI) calcd for C$_9$H$_{14}$N$_4$O$_9$P (M-H) - 341.0504, found 341.0498 m/z.

\[ \text{HO}-\text{O}^\sigma\text{NH}_2^+ \]

\[ \text{O}^\sigma\text{NH}_2^+ \]

**2-Acetamido-2,4-dideoxy-α-D-galactopyranosyl diammonium Phosphate (2-73):** Compound 2-73 was prepared using NahK as described above. Colorless oil, 70% yield. $^1$H NMR (400 MHz, D$_2$O): $\delta$ = 5.47 (dd, $J$ = 7.3, 3.3 Hz, 1H), 4.16-4.22 (m, 1H), 4.04 (td, $J$ = 10.8, 4.8 Hz, 1H), 3.82 (dt, $J$ = 10.4, 2.6 Hz, 1H), 3.71 (dd, $J$ = 12.1, 3.2 Hz, 1H), 3.62 (dd, $J$ = 12.1, 6.0 Hz, 1H), 2.06 (s, 3H), 2.05-2.10 (m, 1H), 1.55 (dd, $J$ = 24.0, 12.1 Hz, 1H); $^{31}$P NMR (162 MHz, D$_2$O): $\delta$ = -0.63. NMR data also see Srivastava et al.$^{93}$
2-Acetamido-4-azido-2,4-dideoxy-α-D-galactopyranosyl diammonium Phosphate (2-74): Compound 2-74 was prepared using NahK as described above. Colorless oil, 73% yield. $^1$H NMR (400 MHz, D$_2$O): $\delta = 5.43$ (dd, $J = 7.4$, 3.0 Hz, 1H), 4.22-4.27 (m, 3H), 4.15-4.16 (m, 1H), 3.75-3.78 (m, 2H), 2.06 (s, 3H); $^{13}$C NMR (100 MHz, D$_2$O): $\delta = 174.8$, 93.4 (d), 70.1, 67.8, 62.8, 61.1, 50.2 (d), 22.0; $^{31}$P NMR (162 MHz, D$_2$O): $\delta = -0.95$; HRMS (ESI) calcd for C$_8$H$_{14}$N$_4$O$_8$P (M-H$^-$) 325.0555, found 325.0542 m/z.

2-Acetamido-2-deoxy-α-D-allosyl diammonium Phosphate (2-75): Compound 2-75 was prepared using NahK as described above. Colorless oil, 22% yield. $^1$H NMR (400 MHz, D$_2$O) $\delta$ 5.39 (dd, $J = 7.7$, 3.6 Hz, 1H), 4.02-4.13 (m, 3H), 3.94 (dd, $J = 12.3$, 2.3 Hz, 1H), 3.79 (dd, $J = 12.3$, 5.3 Hz, 1H), 3.70 (dd, $J = 10.5$, 3.4 Hz, 1H), 2.09 (s, 3H); $^{13}$C NMR (100 MHz, D$_2$O) $\delta$ 174.1, 92.3 (d), 69.9, 67.3, 66.1, 60.9, 50.3 (d), 22.0; $^{31}$P NMR (162 MHz, D$_2$O) $\delta$ 2.17; HRMS (ESI) calcd for C$_8$H$_{15}$NO$_9$P (M-H$^-$) 300.0490, found 300.0481 m/z.
**2-Acetamido-2,6-dideoxy-α-D-glucosyl diammonium Phosphate (2-76):** Compound 2-76 was prepared using NahK as described above. Colorless oil, 75% yield. $^1$H NMR (400 MHz, D$_2$O) δ 5.35 (dd, $J = 7.3, 3.4$ Hz, 1H), 4.02 (dd, $J = 9.7, 6.3$ Hz, 1H), 3.94-3.98 (m, 1H), 3.77 (dd, $J = 10.1, 9.2$ Hz, 1H), 3.25 (app t, $J = 9.4$ Hz, 1H), 2.08 (s, 3H), 1.31 (d, $J = 6.2$ Hz, 3H). NMR data also see Srivastava et al.$^{93}$

**2-Acetamido-6-azido-2,6-dideoxy-α-D-glucosyl diammonium Phosphate (2-77):** Compound 2-77 was prepared using NahK as described above. Colorless oil, 34% yield. $^1$H NMR (400 MHz, D$_2$O) δ 5.42 (dd, $J = 7.2, 3.3$ Hz, 1H), 4.05-4.09 (m, 1H), 3.97-4.01 (m, 1H), 3.76-3.85 (m, 2H), 3.67 (dd, $J = 13.6, 4.2$ Hz, 1H), 3.59 (app t, $J = 9.3$ Hz, 1H), 2.10 (s, 3H); $^{13}$C NMR (100 MHz, D$_2$O) δ 174.7, 93.1 (d), 71.0, 70.9, 70.4, 54.0 (d), 50.8, 22.0; $^{31}$P NMR (162 MHz, D$_2$O) δ 0.60; HRMS (ESI) calcd for C$_8$H$_{14}$N$_4$O$_8$P (M-H) 325.0555, found 325.0559 m/z.
Phosphate (2-78): Compound 2-78 was prepared using NahK as described above. Colorless oil, 37% yield. $^1$H NMR (400 MHz, D$_2$O) δ 5.33 (dd, $J = 6.9, 3.3$ Hz, 1H), 4.24 (app q, $J = 6.6$ Hz, 1H), 4.09-4.13 (m, 1H), 3.92 (dd, $J = 10.9, 3.1$ Hz, 1H), 3.80 (m, 1H), 2.03 (s, 3H), 1.20 (d, $J = 6.6$ Hz, 3H). NMR data also see Illarionov et al.\textsuperscript{94}

**2-Acetamido-6-azido-2,6-dideoxy-α-D-galactopyranosyl diammmonium Phosphate (2-79):** Compound 2-79 was prepared using NahK as described above. Colorless oil, 42% yield. $^1$H NMR (400 MHz, D$_2$O): δ = 5.42 (dd, $J = 7.3, 3.2$ Hz, 1H), 4.25 (app t, $J = 6.6$, 1H), 4.18-4.21 (m, 1H), 3.95-4.03 (m, 2H), 3.58-3.63 (m, 1H), 3.50-3.54 (m, 1H), 2.07 (s, 3H); $^{13}$C NMR (100 MHz, D$_2$O): δ = 174.9, 93.3 (d), 69.5, 68.6, 67.8, 50.5, 50.0 (d), 22.1; $^{31}$P NMR (162 MHz, D$_2$O): δ = 0.58; HRMS (ESI) calcd for C$_8$H$_{14}$N$_4$O$_8$P (M-H) $^-$ 325.0555, found 325.0566 m/z.

Cloning and overexpression of GlmU

*Bacterial Strains and Plasmids.*

*Escherichia coli* DH5α [*lacZΔM15 hsdR recA*] was purchased from Gibco-BRL (Gaithersburg, MD). *E. coli* BL21 (DE3) [F *ompT hsdS$_{B}(r_B m_B)$ gal dcm (DE3)] and
plasmid pET15b were purchased from Novagen (Carlsbad, CA). *E. coli* K12 (substrain MG1655) was from ATCC (#47076).

*Construction of Recombinant Plasmid.*

The *lum* gene was amplified from genomic DNA of *Escherichia coli* K12 by polymerase chain reaction (PCR) using primers: 5’-ACTG*CATATG*TTGAATAATGCTATGAGCG-3’ and 5’-CTG*ACGGATC*CTCACCTTTTTCTTTTACC GGACG-3’. The PCR fragment was digested with *Bam*HI and *Nde*I (sites in bold) and inserted into the corresponding sitesplasmid pET-15b. The recombinant plasmid was transformed into *E. coli* DH5α cells for DNA sequencing. After sequence confirmation, the plasmid was transformed into *E. coli* BL21 (DE3) for protein expression.

*Overexpression and Purification of GlmU.*

*E. coli* BL21 (DE3) transformant containing the pET15b-*glmU* plasmid was grown in LB-ampicillin medium at 37 °C until OD$_{600}$ reached 0.7. IPTG was added to a final concentration of 0.1 mM to induce protein to express at 16 °C for 20 h. The cells were harvested by centrifugation at 4,000 × g for 30 min. The cell pellet was resuspended in 20 mM sodium phosphate (pH 7.4) and sonicated on ice. The lysate was clarified by centrifugation at 20,000 × g for 40 min at 4 °C and the supernatant was applied to HisTrap™ HP column. The column was then washed with washing buffer (20 mM sodium phosphate, pH 7.4, 100 mM NaCl, 10 mM imidazole) and the enzyme was eluted with elution buffer (20 mM sodium phosphate, pH 7.4, 100 mM NaCl, 500 mM...
imidazole). The eluant was desalted by HiTrap™ Desalting HP column with 50 mM Tris-HCl buffer, pH 7.5. The fractions containing the purified enzyme were combined and concentrated with Amicon Ultra and the enzyme was stored in 20% glycerol (50 mM Tris-HCl, pH 7.5) at -20 °C.

Activity Measurement.
The activity assay of recombinant GlmU was carried out in a mixture containing 100 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 5 mM GlcNAc-1-P, 5 mM UTP, and various amount of GlmU incubated at 37 °C. The reaction was quenched by boiling the mixture for 2 min followed by centrifugation. The supernatant was analyzed by thin-layer chromatography [n-butanol/acetic acid/water = 2:1:1 (v/v/v)].

Cloning and overexpression of AGX1

Bacterial strains and plasmids.
UAP1 (NM_003115) human cDNA clone was purchased from OriGene (Rockville, MD). *Escherichia coli* DH5α [lacZΔM15 hsdR recA] was purchased from Gibco-BRL (Gaithersburg, MD). *E. coli* BL21 (DE3) [F ompT hsdSbv−rB mB− gal dcm (DE3)] and plasmid pET15b were purchased from Novagen (Carlsbad, CA).

Construction of recombinant plasmid.
The *agx1* gene was amplified from UAP1 human cDNA clone by polymerase chain reaction (PCR) with 5’-ACTGCATATGAAACATATGACCTCAAAAATCTC-3’ and 5’-CGCGGATTCCTCAAATACCATTTCACCAGC-3’ as the forward and reverse
primers, respectively. The PCR fragment was digested with NdeI and BamHI (sites in bold) and inserted into the corresponding sites in plasmid pET-15b. The recombinant plasmid was transformed into *E. coli* DH5α cells for amplification and DNA sequencing. After the sequence was confirmed, the plasmid was transformed into *E. coli* BL21 (DE3) for protein expression.

**Overexpression and purification of AGX1.**

*E. coli* BL21 (DE3) transformant containing the pET15b-*agxl* plasmid was grown in LB-ampicillin medium at 37 °C by constant shaking. When OD$_{600}$ reached 0.7, protein expression was induced with 0.1 mM IPTG (final concentration) at 16 °C for 20 h. The cells were harvested by centrifugation at 4,000 × g for 30 min and then resuspended in 20 mM sodium phosphate (pH 7.4) and sonicated on ice. The lysate was clarified by centrifugation at 20,000 × g for 40 min at 4 °C and the supernatant was loaded onto a Ni-NTA agarose affinity column equilibrated with buffer (20 mM sodium phosphate, pH 7.4, 100 mM NaCl, 10 mM imidazole). The column was washed with the same buffer and the enzyme was eluted with elution buffer (20 mM sodium phosphate, pH 7.4, 100 mM NaCl, 500 mM imidazole). The eluant was desalted by HiTrap™ Desalting HP column with 50 mM Tris-HCl buffer, pH 7.5. The protein fractions were combined and concentrated with Amicon Ultra and the enzyme was stored in 20% glycerol (50 mM Tris-HCl, pH 7.5) at -20 °C.
**Activity measurement.**

The activity of recombinant AGX1 was tested at 37 °C for 30 min in a mixture containing 100 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 5 mM GlcNAc-1-P, 5 mM UTP, 1 U/mL PPA and various amount of AGX1. The reaction was terminated by boiling the mixture for 2 min followed by centrifugation. The supernatant was analyzed by thin-layer chromatography \([n\text{-butanol/acetic acid/water} = 2:1:1 (v/v/v)]\).

**General procedure for preparation of UDP-GlcNAc/GalNAc and their analogues using GlmU.** The reaction mixture (4 mL) contained 10 mM GlcNAc/GalNAc-1-P or their analogs, 15 mM UTP, 10 mM MgCl₂, and 1 mg/mL GlmU in 100 mM Tris-HCl buffer (pH 7.5). Progress of the reaction was monitored by TLC \([n\text{-butanol:acetic acid:water} = 2:1:1 (v:v:v)]\). After incubation at 37 °C for 12 h, the mixture was briefly boiled for 2 min and then centrifuged to remove protein. The supernatant was applied to a column (10×170 mm) of DEAE-cellulose (HCO₃⁻ form) equilibrated with 50 mM ammonium bicarbonate. The column was eluted with a linear gradient of 25-300 mM NH₄HCO₃. The fractions containing the products were collected and concentrated by lyophilization. The residue was then dissolved in ddH₂O (1 mL) and applied to a Bio-Gel P-2 column. ddH₂O was used as eluent, and fractions containing the products were collected and lyophilized.

**General procedure for preparation of UDP-GlcNAc/GalNAc and their analogues using AGX1.** The reaction mixture (5 mL) contained 10 mM GalNAc/GlcNAc-1-P or their analogs, 15 mM UTP, 10 mM MgCl₂, and 0.5 mg/mL AGX1 in 100 mM Tris-HCl.
buffer (pH 7.5). Progress of the reaction was monitored by TLC [n-butanol:acetic acid:water 2:1:1 (v:v:v)]. After incubation at 37 °C for 0.5-24 h, the mixture was briefly boiled for 2 min and then centrifuged to remove protein. The supernatant was applied to a column (10×170 mm) of DEAE-cellulose (HCO₃⁻ form) equilibrated with 50 mM ammonium bicarbonate. The column was eluted with a linear gradient of 25-300 mM NH₄HCO₃. The fractions containing the products were collected and concentrated by lyophilization. The residue was then dissolved in ddH₂O (1 mL) and applied to a Bio-Gel P-2 column. ddH₂O was used as eluent, and fractions containing the products were collected and lyophilized.

Uridine 5′-diphospho-2-acetamido-2-deoxy-α-D-glucopyranose diammonium salt (2-8): Compound 2-8 was prepared using GlmU (40% yield) or AGX1 (75% yield) as described above. White foam. ¹H NMR (400 MHz, D₂O): δ = 7.92 (d, J = 8.1 Hz, 1H), 5.92-5.95 (m, 2H), 5.48 (dd, J = 7.3, 3.2 Hz, 1H), 4.32-4.36 (m, 2H), 4.13-4.26 (m, 3H), 3.96 (dt, J = 10.5, 3.1 Hz, 1H), 3.90 (ddd, J = 10.1, 4.2, 2.2 Hz, 1H), 3.84 (dd, J = 12.5, 2.4 Hz, 1H), 3.75-3.80 (m, 2H), 3.51 (app t, J = 9.8 Hz, 1H), 2.04 (s, 3H).
Uridine 5′-diphospho-2-propionamido-2-deoxy-α-D-glucopyranose diammonium salt (2-80): Compound 2-80 was prepared using GlmU (57% yield) or AGX1 (51% yield) as described above. White foam, \( ^1H \) NMR (500 MHz, D\(_2\)O): \( \delta = 7.97 \) (d, \( J = 8.2 \) Hz, 1H), 5.98-6.00 (m, 2H), 5.53 (dd, \( J = 7.2, 3.3 \) Hz, 1H), 4.36-4.41 (m, 2H), 4.18-4.32 (m, 3H), 4.02 (dt, \( J = 10.5, 3.0 \) Hz, 1H), 3.95 (ddd, \( J = 10.1, 4.3, 2.2 \) Hz, 1H), 3.89 (dd, \( J = 12.4, 2.2 \) Hz, 1H), 3.81-3.85 (m, 2H), 3.58 (app t, \( J = 9.4 \) Hz, 1H), 2.38 (q, \( J = 7.7 \) Hz, 2H), 1.14 (t, \( J = 7.7 \) Hz, 3H). NMR data also see Lazarevic et al.\(^{91}\)

Uridine 5′-diphospho-2-butyramido-2-deoxy-α-D-glucopyranose diammonium salt (2-81): Compound 2-81 was prepared using GlmU (27% yield) as described above. White foam, \( ^1H \) NMR (500 MHz, D\(_2\)O): \( \delta = 7.98 \) (d, \( J = 8.1 \) Hz, 1H), 5.99-6.01 (m, 2H), 5.54 (dd, \( J = 7.1, 3.3 \) Hz, 1H), 4.37-4.42 (m, 2H), 4.19-4.33 (m, 3H), 4.03 (dt, \( J = 10.7, 2.9 \) Hz, 1H), 3.97 (ddd, \( J = 10.1, 4.3, 2.1 \) Hz, 1H), 3.90 (dd, \( J = 12.6, 2.1 \) Hz, 1H), 3.82-3.86 (m, 2H), 3.58 (app t, \( J = 9.6 \) Hz, 1H), 2.35 (t, \( J = 6.9 \) Hz, 1H).
= 7.5 Hz, 2H), 1.61-1.69 (m, 2H), 0.95 (t, J = 7.5 Hz, 3H). NMR data also see Lazarevic et al.91

Uridine 5'-diphospho-2-azidoacetamido-2-deoxy-α-D-glucopyranose diammonium salt (2-82): Compound 2-82 was prepared using GlmU (44% yield) or AGX1 (55% yield) as described above. White foam. ¹H NMR (500 MHz, D₂O): δ 7.82 (d, J = 8.1 Hz, 1H), 5.84-5.86 (m, 2H), 5.41 (dd, J = 7.1, 3.2 Hz, 1H), 4.22-4.26 (m, 2H), 4.04-4.17 (m, 3H), 3.98 (two doublets, J = 15.9 Hz, 2H, -CH₂-), 3.91-3.95 (m, 1H), 3.79-3.83 (m, 1H), 3.66-3.76 (m, 3H), 3.44 (app t, J = 9.7 Hz, 1H). NMR data also see Vocadlo et al.27

Uridine 5'-diphospho-2-azido-2-deoxy-α-D-glucopyranose diammonium salt (2-84): Compound 2-84 was prepared using GlmU (48% yield) as described above. White foam. ¹H NMR (500 MHz, D₂O): δ = 8.01 (d, J = 8.1 Hz, 1H), 6.00-6.02 (m, 2H), 5.71 (dd, J = 7.4, 3.3 Hz. 1H), 4.39-4.41 (m, 2H), 4.24-4.32 (m, 3H), 3.94-3.99 (m, 2H), 3.88 (dd, J = 12.5, 2.2 Hz, 1H), 3.82 (dd, J = 12.5, 4.2
Hz, 1H), 3.56 (app t, J = 10.0, 9.2 Hz, 1H), 3.40 (dt, J = 10.4, 3.0 Hz, 1H), \(^{13}\)C NMR (125 MHz, D\(_2\)O): \(\delta = 166.3, 151.8, 141.7, 102.7, 94.4 \,(d), 88.3, 83.3 \,(d), 73.8, 72.9, 70.7, 69.7, 69.4, 64.9 \,(d), 62.9 \,(d), 60.2; \)

\(^{31}\)P NMR (202 MHz, D\(_2\)O): \(\delta = -11.4 \,(J = 19.7 \,Hz), -13.4 \,(J = 19.5 \,Hz); \) HRMS (ESI) calcd for C\(_{15}\)H\(_{22}\)N\(_5\)O\(_16\)P\(_2\) (M-H)\(^-\) 590.0542, found 590.0543 m/z.

Uridine 5'-diphospho-2-acetamido-2,6-dideoxy-\(\alpha\)-D-glucopyranose diammonium salt (2-85): Compound 2-85 was prepared using GlmU (50% yield) as described above. White foam. \(^1\)H NMR (500 MHz, D\(_2\)O): \(\delta = 7.95 \,(d, J = 8.3 \,Hz, 1H), 5.96-5.98 \,(m, 2H), 5.45 \,(dd, J = 7.1, 3.3 \,Hz, 1H), 4.34-4.39 \,(m, 2H), 4.16-4.30 \,(m, 3H), 3.97-4.02 \,(m, 2H), 3.75 \,(dd, J = 11.2, 9.5 \,Hz, 1H), 3.26 \,(app t, J = 9.5 \,Hz, 1H), 2.07 \,(s, 3H), 1.30 \,(d, J = 6.2 \,Hz, 3H). \) NMR data also see Srivastava et al.\(^\text{93}\)

Uridine 5'-diphospho-2-acetamido-6-azido-2,6-dideoxy-\(\alpha\)-D-glucopyranose diammonium salt (2-86): Compound 2-86 was prepared using GlmU (20% yield) or AGX1 (74% yield) as described above. White foam. \(^1\)H NMR (500 MHz, D\(_2\)O): \(\delta = 7.97 \,(d, J = 8.1 \,Hz, 1H), 5.98-6.00 \,(m, 2H), 5.52 \,(dd, J = 7.2, 6.0 \,Hz, 1H), 4.36-4.50 \,(m, 2H), 3.93-4.00 \,(m, 2H), 3.75 \,(dd, J = 11.2, 9.5 \,Hz, 1H), 3.26 \,(dd, J = 9.5, 6.0 \,Hz, 1H), 2.07 \,(s, 3H), 1.30 \,(d, J = 6.2 \,Hz, 3H). \)
3.3 Hz, 1H), 4.36-4.40 (m, 2H), 4.18-4.31 (m, 3H), 4.00-4.09 (m, 2H), 3.81 (dd, \( J = 10.3, 9.4 \) Hz, 1H), 3.74 (dd, \( J = 13.5, 2.5 \) Hz, 1H), 3.64 (dd, \( J = 13.6, 4.1 \) Hz, 1H), 3.59 (t, \( J = 9.5 \) Hz, 1H), 2.09 (s, 3H); \( ^{13} \)C NMR (125 MHz, D\(_2\)O) \( \delta \) 174.8, 166.2, 151.8, 141.7, 102.7, 94.4 (d), 88.6, 83.2(d), 73.8, 71.7, 70.8, 70.2, 69.6, 65.0 (d), 53.7 (d), 50.6, 22.1; \( ^{31} \)P NMR (162 MHz, D\(_2\)O): \( \delta \) = -11.3 (d, \( J = 21.4 \) Hz), -13.2 (d, \( J = 21.4 \) Hz); HRMS (ESI) calcd for C\(_{17}\)H\(_{25}\)N\(_6\)O\(_{16}\)P\(_2\) (M-H)\(^{-}\) 631.0808, found 631.0806 m/z.

Uridine 5′-diphospho-2-acetamido-2-deoxy-\( \alpha \)-d-allopyranose diammonium salt (2-87): Compound 2-87 was prepared using GlmU (20% yield) as described above. White foam. \( ^{1} \)H NMR (400 MHz, D\(_2\)O): \( \delta \) = 7.96 (d, \( J = 8.1 \) Hz, 1H), 5.96-6.00 (m, 2H), 5.53 (dd, \( J = 7.1, 3.8 \) Hz, 1H), 4.36-4.40 (m, 2H), 4.21-4.30 (m, 3H), 4.08-4.15 (m, 3H), 3.92 (dd, \( J = 12.6, 2.3 \) Hz, 1H), 3.82 (dd, \( J = 12.6, 4.5 \) Hz, 1H), 3.77 (dd, \( J = 10.5, 3.2 \) Hz, 1H), 2.10 (s, 3H); \( ^{31} \)P NMR (162 MHz, D\(_2\)O): \( \delta \) = -11.0 (d, \( J = 19.9 \) Hz), -13.0 (d, \( J = 21.7 \) Hz); HRMS (ESI) calcd for C\(_{17}\)H\(_{26}\)N\(_3\)O\(_7\)P\(_2\) (M-H)\(^{-}\) 606.0743, found 606.0730 m/z.
Uridine 5′-diphospho-2-acetamido-2,4-dideoxy-α-D-glucopyranose diammonium salt (2-88): Compound 2-88 was prepared using GlmU (59% yield) or AGX1 (72% yield) as described above. White foam. $^1$H NMR (500 MHz, D$_2$O): $\delta = 7.93$ (d, $J = 8.1$ Hz, 1H), 5.94-5.97 (m, 2H), 5.54 (dd, $J = 7.0$, 3.0 Hz, 1H), 4.33-4.37 (m, 2H), 4.15-4.28 (m, 4H), 4.01 (td, $J = 11.0$, 4.8 Hz, 1H), 3.83-3.86 (m, 1H), 3.69 (dd, $J = 12.3$, 3.2 Hz, 1H), 3.60 (dd, $J = 12.1$, 5.4 Hz, 1H), 2.06 (s, 3H), 2.03-2.06 (m, 1H), 1.55 (q, $J = 12.3$ Hz, 1H). NMR data also see Srivastava et al. 93

Uridine 5′-diphospho-2-acetamido-2-deoxy-α-D-galactopyranose diammonium salt (2-89): Compound 2-89 was prepared using GlmU (65% yield) or AGX1 (79% yield) as described above. White foam. $^1$H NMR (400 MHz, D$_2$O): $\delta = 7.88$ (d, $J = 8.1$ Hz, 1H), 5.90-5.93 (m, 2H), 5.50 (dd, $J = 7.3$, 3.4 Hz, 1H), 4.28-4.34 (m, 2H), 4.11-4.24 (m, 5H), 3.99-4.00 (m, 1H), 3.91 (dd, $J = 11.0$, 3.3 Hz, 1H), 3.67-3.75 (m, 2H), 2.03 (s, 3H). NMR data also see Piller et al. 75
Uridine 5′-diphospho-2-propionamido-2-deoxy-\(\alpha\)-D-galactopyranose diammonium salt (2-91): Compound 2-91 was prepared using AGX1 (57% yield) as described above. White foam. \(^1\)H NMR (500 MHz, D\(_2\)O): \(\delta = 8.09\) (d, \(J = 8.1\) Hz, 1H), 6.09-6.12 (m, 2H), 5.67 (dd, \(J = 7.1, 3.4\) Hz, 1H), 4.49-4.52 (m, 2H), 4.30-4.42 (m, 5H), 4.17-4.18 (m, 1H), 4.10 (dd, \(J = 11.0, 3.1\) Hz, 1H), 3.86-3.94 (m, 2H), 2.49 (qd, \(J = 7.5, 2.3\) Hz, 2H), 1.26 (t, \(J = 7.6\) Hz, 3H). NMR data also see Lazarevic et al.\(^92\)

Uridine 5′-diphospho-2-butyramido-2-deoxy-\(\alpha\)-D-galactopyranose diammonium salt (2-92): Compound 2-92 was prepared using AGX1 (44% yield) as described above. White foam. \(^1\)H NMR (500 MHz, D\(_2\)O): \(\delta = 8.09\) (d, \(J = 8.1\) Hz, 1H), 6.09-6.11 (m, 2H), 5.66 (dd, \(J = 7.0, 3.4\) Hz, 1H), 4.48-4.52 (m, 2H), 4.29-4.42 (m, 5H), 4.16-4.17 (m, 1H), 4.09 (dd, \(J = 11.0, 3.1\) Hz, 1H), 3.84-3.93 (m, 2H), 2.45 (t, \(J = 7.3\) Hz, 2H), 1.70-1.79 (m, 2H), 1.04 (t, \(J = 7.4\) Hz, 3H). NMR data also see Lazarevic et al.\(^92\)
Uridine 5′-diphospho-2-azidoacetamido-2-deoxy-α-D-galactopyranose diammonium salt (2-90): Compound 2-90 was prepared using AGX1 (64% yield) as described above. White foam. \(^1\)H NMR (500 MHz, D\(_2\)O): \(\delta = 7.79\) (d, \(J = 8.0\) Hz, 1H), 5.81-5.83 (m, 2H), 5.41 (dd, \(J = 7.0, 3.4\) Hz, 1H), 4.19-4.23 (m, 2H), 4.07-4.17 (m, 3H), 3.89-4.05 (m, 5H), 3.83 (dd, \(J = 11.0, 3.2\) Hz, 1H), 3.57-3.65 (m, 2H). NMR data also see Hang et al.\(^{36}\)

Uridine 5′-diphospho-2-acetamido-2,6-dideoxy-α-D-galactopyranose diammonium salt (2-95): Compound 2-95 was prepared using GlmU (55% yield) or AGX1 (77% yield) as described above. White foam. \(^1\)H NMR (400 MHz, D\(_2\)O): \(\delta = 7.93\) (d, \(J = 8.2\) Hz, 1H), 5.94-5.97 (m, 2H), 5.46 (dd, \(J = 6.7, 3.4\) Hz, 1H), 4.32-4.37 (m, 2H), 4.14-4.30 (m, 5H), 3.95 (dd, \(J = 11.0, 3.1\) Hz, 1H), 3.82-3.83 (m, 1H), 2.06 (s, 3H), 1.22 (d, \(J = 6.6\) Hz, 3H). NMR data also see Busca et al.\(^{95}\)
Uridine 5′-diphospho-2-acetamido-6-azido-2,6-dideoxy-α-D-galactopyranose diammonium salt (2-96): Compound 2-96 was prepared using GImU (10% yield) or AGX1 (61% yield) as described above. White foam. $^1$H NMR (500 MHz, D$_2$O): δ = 7.98 (d, $J = 8.2$ Hz, 1H), 5.97-5.99 (m, 2H), 5.54 (dd, $J =$ 7.2, 3.4 Hz, 1H), 4.35-4.39 (m, 2H), 4.19-4.30 (m, 5H), 4.02 (m, 1H), 3.98 (dd, $J = 10.8$, 3.2 Hz, 1H), 3.60 (dd, $J = 12.8$, 7.2 Hz, 1H), 3.50 (dd, $J = 12.8$, 6.0 Hz, 1H), 2.09 (s, 3H); $^{13}$C NMR (100 MHz, D$_2$O) δ 175.0, 166.3, 151.8, 141.7, 102.6, 94.6 (d), 88.6, 83.2 (d), 73.8, 70.3, 69.6, 68.5, 67.5, 64.9 (d), 50.3, 49.6 (d), 22.1; $^{31}$P NMR (162 MHz, D$_2$O): δ = -11.2 (d, $J = 21.5$ Hz), -13.0 (d, $J = 21.5$ Hz); HRMS (ESI) calcd for C$_{17}$H$_{25}$N$_{6}$O$_{16}$P$_{2}$ (M-H$^-)$ 631.0808, found 631.0801 m/z.

References


(4) DeAngelis, P. L. Glycobiology 2002, 12, 9R-16R.


123


127


CHAPTER 3

CHEMOENZYMATIC SYNTHESIS of URIDINE 5′-DIPHOSPHO-2-
ACETONYL-2-DEOXY-α-D-GLUCOSE AS C2-CARBON ISOSTERE of UDP-
GlcNAc AND THE APPLICATIONS

3.1 Introduction

2-Acetamido-2-deoxy-D-glycopyranosides (GlcNAc/GalNAc) are widely
distributed in living systems as oligosaccharides and glycoconjugates, and play
significant roles in a wide range of biological processes.\textsuperscript{1-2} Hence, there is a considerable
interest in chemical modification of GlcNAc/GalNAc residues by an unnatural analogue
in such a polysaccharide chain for further understanding and modulating the targets of
these glycosides.\textsuperscript{3-8} Especially, analogues with bioorthogonal groups (e.g. reactive
ketones, azides, alkynes, \textit{etc.}) are of great interests since they can be further
functionalized by chemoselective reactions.\textsuperscript{9-11} Among the various analogues, 2-acetonyl-
2-deoxy-D-galactose (2-ketoGal, \textbf{3-1}) has attracted much attention because it bears much
resemblance to GalNAc, and most significantly, contains a ketone handle. This substrate
serves as a carbon isostere of GalNAc and was incorporated into cell surface glycoprotein
through metabolic pathways. Then the ketone epitope was chemoselectively reacted with
biotin hydrazide, stained with fluorescein isothiocyanate (FITC) labeled avidin, and detected by flow cytometry for quantification (Figure 3.1). Moreover, 2-ketoGal, which is transferred from the corresponding donor UDP-2ketoGal, was also widely served as a ketone handle to detect O-GlcNAc glycoprotein posttranslational modifications. An engineered GalT with a single mutation Y289L that enlarges the binding pocket was used to transfer the 2-ketoGal (3-1) from UDP-2-ketoGal (3-2) onto O-GlcNAc glycoprotein. The ketone group was then reacted with an aminooxybiotin and detected by chemiluminescence using streptavidin conjugated to horseradish peroxidase (Figure 3.2).

Figure 3.1. Metabolic carbohydrate engineering with 2-ketoGal (3-1).
However, the corresponding 2-keto isostere of GlcNAc (2-ketoGlc) is missing from current literature. Bertozzi and co-workers demonstrated 2-ketoGal was successfully taken as a substrate for metabolic glycoprotein engineering through the salvage pathway while the GlcNAc counterpart could not be incorporated. One explanation for this observation is because, firstly, GlcNAc is abundant and the 2-ketoGlc lies in strong competition with the endogenous GlcNAc. In addition to this, unlike GalNAc, current data suggests that there is a lack of suitable metabolic pathway (salvage pathway) which is critical for the generation of UDP-sugar donor. In this chapter, we first demonstrated that 2-keto isostere of GlcNAc (2-ketoGlc, 3-11) could be converted into the sugar nucleotide donor by our enzyme system in vitro. Then we demonstrate the incorporation of this unnatural ketone containing monosaccharide, which heterogeneously is incorporated into the cell surface oligosaccharides of E. coli. Through bioorthogonal chemistry we selectively label the cell surface oligosaccharides containing the ketone modified monosaccharide with various probes, which are used to image and identify the cell surface oligosaccharides.
3.2 Chemoenzymatic synthesis of UDP-2-ketoGlc (3-13)

To prove the sugar donor molecule of 2-ketoGlc (3-11) could be enzymatically synthesized in vitro using our two enzyme system, we first synthesized free 2-ketoGlc (3-11) chemically with an improved strategy. Scheme 3.1 shows the retrosynthetic analysis of this molecule. The keto moiety can be generated by oxidative double bond cleavage. The C-2 side chain with the double bond could be installed via intermolecular free radical coupling reactions between the radical precursor iodo-sugar (3-3) and methallytin reagent. Giese et al. analyzed the stereoselectivity of the radical coupling reaction of six-membered ring. Substituents in a β position with respect to the radical reaction center result in anti attack, which means equatorial β substituent directs equatorial installation of the side chain. Thus to get our target molecule with glucose configuration (C-2 equatorial), iodo precursor of configuration (3-3) is required. Moreover, a second equatorial β substituent (1,3-di-equitorial-OAc, structure 3-3, Scheme 3.1,) would further increase the proportion of equatorial attack. Finally, the iodo sugar could be prepared by iodoacetyoxylation reaction starting from substituted glycal.

Scheme 3.1. Retrosynthetic analysis of 2-ketoGlc.
However, starting from glucal triacetate (3-4), direct iodoacetyloxylation reaction with N-iodosuccinimide (NIS)/AcOH afforded the α-manno type product (3-5) with axial iodo substituent (Scheme 3.2). The desired β-gluco type compound (3-8 or 3-3) was only obtained as a minor isomer (<10%). Surprisingly, separation of this minor isomer from the reaction mixture turned out to be one major approach to acquire compound 3-8.\textsuperscript{10,12,19}

The reason for α-manno type product (3-5) dominance is that \( ^4\text{H}_5 \) half chair conformation predominates for glucal triacetate (3-4) and β face attack of iodonium species is stereoelectronically favored since a chair-like transition state would involve. α-Attack of acetic acid could be simply explained by anomeric effect.

![Scheme 3.2. Iodoacetyloxylation of glucal triacetate.](image)

We designed an improved route for β-gluco type configuration as shown in scheme 3.3. Change the protection groups of triacetate (3-4) to tert-butyldiphenylsilyl groups afforded compound 3-6. In this case, iodoacetyloxylation resulted isomer 3-7 with equatorial iodo substitution. For the tri-TBDPS protected glucal, the \( ^5\text{H}_4 \) half-chair conformation is thought to predominant to relieve the unfavorable steric interactions of the big substituents. Bottom face (α face) attack of iodonium species is sterically favored
at this point. Removal of the TBDPS groups with fluoride followed by peracetylation
gave the desired β-gluco type compound 3-8.

Keck radical reaction of 3-8 with methallytributyltin and AIBN afforded
compound 3-9 in 66% yield. β-Manno type product with axial C-2 side chain was also
generated as a minor isomer (Scheme 3.4). It was observed that longer refluxing time
(overnight as reported in Bertozzi et al.\textsuperscript{12}) resulted in the removal of anomeric acetyl
group. Oxidative cleavage of the terminal olefin 3-9 by OsO\textsubscript{4}-NaIO\textsubscript{4} in the presence of
weak base\textsuperscript{20} furnished peracetylated 2-ketoGlc 3-10 in good yield. The addition of a weak
base (2,6-lutidine) could greatly suppress the side reaction by neutralizing the acid
generated in the reaction.\textsuperscript{20} Further deprotection under mild deacetylation condition
provided desired 2-ketoGlc (3-11). It is worth noting that although moderate yield was
observed for the deacetylation reaction, other deacetylation methods (MeONa/MeOH or
K\textsubscript{2}CO\textsubscript{3}/MeOH) did not give our desired free sugar (<10%). The NMR data of compound
3-10 were in disagreement with those reported by Bertozzi and co-workers.\textsuperscript{12} However,
our HRMS and COSY experiments confirmed that our structural assignment was correct (see the Experimental Section and Figure 3.3).

Scheme 3.4. Continued synthesis of 2-ketoGlc.

Figure 3.3. $^1$H NMR of compound 3-10.

The free sugar 2-ketoGlc (3-11) was taken as a good substrate for both enzymes (NahK and GlmU) as shown in Scheme 3.5: 1-Phosphorylation was carried out for a
longer time (65 h), and additional NahK was added during the reaction course to fulfill the greatest conversion (79%). The product 2-ketoGlc-1-phosphate (3-12) was purified by silica gel chromatography and Bio-Gel P-2 column. Pyrophosphorylation of 3-12 was catalyzed by GlmU in the presence of UTP. Yeast inorganic pyrophosphatase (PPA) was added to drive the reaction forward by degrading the byproduct pyrophosphate (PPi). The resulting UDP-2-ketoGlc (3-13) was purified by DEAE-cellulose anion exchange resin and Bio-Gel P-2 column for desalting purpose. The good yield of both reactions indicates that the amide -NH- of GlcNAc may not be critical for the recognition of GlcNAc by NahK and GlcNAc-1-phosphate by GlmU in the UDP-GlcNAc biosynthesis. By making these changes to NahK reaction (longer reaction time and additional enzyme), we were able to broaden the scope of substrates of our system which can be utilized in the synthesis of more UDP-Glc derivatives with C-2 modifications (e.g. UDP-2-azido-Glc was also synthesized successfully as shown in Chapter 2). To the best of knowledge, this is the first chemoenzymatic synthesis of UDP-2-ketoGlc.

Scheme 3.5. Enzymatic synthesis of UDP-2-ketoGlc.
3.3 C₂-carbon isostere of GlcNAc as substrate for bacterial polysaccharide remodeling

In this section, we demonstrate that unnatural 2-ketoGlc (3-11) can be processed by *E. coli* and displayed on the cell surface (Figure 3.4). The incorporated ketone epitope allowed further detection by conjugation with commercially available aminooxybiotin reagent. Through western blot analysis, using purified lipopolysaccharide (LPS), we show that the LPS is modified with the ketone epitope.

![Metabolic incorporation of ketone isotope on E.coli surface.](image)

**Figure 3.4.** Metabolic incorporation of ketone isotope on *E.coli* surface.

### 3.3.1 Incorporation of 2-ketoGlc (3-11) in cell surface oligosaccharides

In this study we sought to remodel the bacterial cell surface by introduction of our synthesized keto sugar (2-ketoGlc, 3-11), which can be probed by using bioorthogonal
chemistry. Previous work by various groups have demonstrated the utility of modifying cell surface oligosaccharides \textit{in vivo}, typically exploiting very specific biosynthetic mechanisms, primarily in eukaryotic organisms. More specifically, we incubated 2-ketoGlc (3-11) with various bacterial strains, with the notion that sugar may act as a substrate for various enzymes within the cells’ oligosaccharide biosynthetic pathways, which could eventually be incorporated in the cell surface oligosaccharides. For example, there exists multiple salvage pathways, such as the GalNAc salvage pathway, in which a bacterium can obtain a specific sugar from the external medium, and directly convert this sugar to the required sugar nucleotide donor.\textsuperscript{22}

In the previous section, we have demonstrated that the key sugar nucleotide donor molecule (3-13) for 2-ketoGlc incorporation could be synthesized \textit{in vitro} by bacterial kinase (NahK) and pyrophosphorylase (GlmU).\textsuperscript{23} To test whether or not 2-ketoGlc (3-11) could be incorporated into the cell surface oligosaccharides, we first incubated overnight cultures of \textit{E. coli} BL21 (DE3) with 10 mM of 2-ketoGlc or peracetylated GlcNAc (negative control). Since the ketone is not a common functional group found in living organisms, by virtue of bioorthogonal chemistry, those cells that incorporate the 2-ketoGlc (3-11) into the cell surface oligosaccharides can be easily identified. Therefore, to the overnight cultures both containing and void of the 2-ketoGlc (3-11), \textit{N}-\textit{N}(aminooxyacetyl)-\textit{N}’-(D-biotinoyl) hydrazine (aminooxybiotin) in the presence of the aniline catalyst at neutral pH was added.\textsuperscript{24} Ketone groups that had been metabolically engineered into the cell surface polysaccharides can react with the aminooxybiotin reagent resulting in biotin-coated bacterial cells. The biotinylated bacterial cells can be probed by staining with FITC conjugated avidin, and the resulting fluorescence signal
can be analyzed by flow cytometry (Figure 3.5). As shown in Figure 3.5A, *E. coli* BL21 (DE3) cells fed with 2-ketoGlc (3-11) showed an increase in fluorescence intensity compared with control samples. Two negative controls are shown, with the teal line representing *E. coli* BL21 (DE3) cells that were treated with the aminooxybiotin reagent followed by subsequent coupling with FITC conjugated avidin, and the orange line representing the fluorescence intensity of just the naked *E. coli* BL21 (DE3). Cells incubated with the biotin linker and subsequently incubated with the FITC conjugated but were void of the 2-ketoGlc sugar observed no change in fluorescence (from the negative *E. coli* BL21 (DE3)) suggesting that the biotin hydrazine is reacting specifically with the ketone in 2-ketoGlc (3-11). A similar result was observed for an *E. coli* O86:B7 ΔwecA/ΔwaaL strain (Figure 3.5B), which is used a model strain to identify the location of the modified (Section 3.3.2). These results are surprising, for the simple fact that there exists no endogenous GlcNAc-1-kinase, such as NahK, which *in vitro* is utilized for the synthesis of the UDP-GlcNAc donor (3-13). Fluorescent microscopy imaging also confirmed successful incorporation of ketone group on *E. coli* cell surface (Figure 3.6). White arrows in Figure 3.6A point at FITC labeled *E. Coli* rods. Therefore, *in vivo*, these results suggest that the 2-ketoGlc is may be being processed either as a mock Glc or GlcNAc due to structural similarities, or is possibly being epimerized by GalE to either a mock Gal or GalNAc. Thus, the obtained results are suggestively very complex as it is challenging to identify which biosynthetic pathway(s) the 2-ketoGlc is traveling by.

In order to try and regulate the glycosylation we attempted at knocking out the *de novo* UDP-GlcNAc biosynthesis pathway, more specifically by removing glmM from the chromosome. To a strain of *E. coli* containing NahK (GlcNAc kinase) we attempted to
knock out *glmM* and drive UDP-GlcNAc biosynthesis through NahK, however, we were unsuccessful. This is likely due in part because of the abundance of UDP-GlcNAc required by the cell, and eliminating this biosynthetic pathway is lethal for the bacterium. Lastly, In addition to the flow cytometry results, incubation with 2-ketoGlc (3-11) appeared to have a significant inhibitory effect on the cell growth. No such effects were reported in previous studies in which unnatural sugars were incorporated with either eukaryotic or prokaryotic organisms.

---

Figure 3.5. Bacterial cell surface labeling by introduction of 2-ketoGlc (3-11), conjugated with aminooxybiotin, and detection by FITC conjugated avidin. The fluorescence signal was analyzed by flow cytometry. A) Fluorescence labeling of 2-ketoGlc fed *E. coli* BL21 (DE3). Red line: Positive reaction; Teal line: Negative control (no 2-ketoGlc but treated with aminooxybiotin and FITC-avidin; Orange fill: Negative control (naked *E. coli* BL21 (DE3)); B) Fluorescence labeling of 2-ketoGlc fed *E. coli* O86:B7 ΔwecA/ΔwaaL. Red line: Positive reaction; Teal line: Negative control (no 2-ketoGlc but treated with aminooxybiotin and FITC-avidin; Orange fill: Negative control (naked *E. coli* O86:B7 ΔwecA/ΔwaaL)).
3.3.2 Labeling LPS

Previous work by our lab has used *E. coli* O86 as a model system for the pinpoint modification of the fucosylated LPS by using metabolic pathway engineering, and in this work we sought to expand our ability to modify this model strain with 2-ketoGlc (3-11).25 Structurally 2-ketoGlc is very similar to GlcNAc, with the exception of the C-2 amide, and thus may be incorporated as GlcNAc in the LPS. In the R3-type core
oligosaccharide (Figure 3.7 left), there exist multiple GlcNAc residues, several glucose residues, and several other hexoses, thus representing several possible glycosyltransferases that may tolerate and incorporate this unnatural sugar when synthesizing the core oligosaccharide.²⁶

Figure 3.7. *E. coli* O86 (R3-type) LPS core oligosaccharide (left) and SDS-PAGE of LPS and western blot (right). Lane 1: LPS phenotype for *E. coli* O86:B7, visualized by silver staining; lane 2: LPS phenotype for *E. coli* O86:B7 ΔwecA/ΔwaaL, visualized by silver staining; lane 3: LPS phenotype for *E. coli* O86:B7 ΔwecA/ΔwaaL incubated with peracetylated GlcNAc, visualized by HRP conjugate streptavidin; lane 4: LPS phenotype for *E. coli* O86:B7 ΔwecA/ΔwaaL incubated with peracetylated 2-ketoGlc, visualized by HRP conjugate streptavidin.

To examine one possible location where the unnatural sugar is being incorporated we created a double knock out strain of *E. coli* O86:B7, which is functionally missing *wecA* and *waaL*. These enzymes are involved in the biosynthetic
pathway of the LPS, more specifically, the synthesis and ligation of the O-antigen, respectively.\textsuperscript{27-28} By removing the \textit{wecA} gene from the chromosome, we remove the glycosyltransferase which is responsible for the first sugar found in the O-antigen, thus removing the O-antigen from the LPS. Furthermore, by disruption of the \textit{waaL} gene, we ensure that there is no ligation of any oligosaccharide onto the LPS core oligosaccharide. Thus, the resulting LPS displayed on the cell surface will only contain the core oligosaccharides (Figure 3.7 left), which can be verified by the LPS phenotype observed in Figure 3.7. Figure 3.7, lane 1 shows the wild type LPS phenotype for \textit{E. coli} O86:B7, whereas lane 2 is the LPS phenotype for the \textit{E. coli} O86:B7 \textit{ΔwecA/ΔwaaL} strain. The single band observed in lane 2 is characteristic of the core oligosaccharide (Figure 3.7 left), which contains several GlcNAc residues.\textsuperscript{28} Thus, to determine one possible location where the 2-ketoGlc may be incorporated, \textit{E. coli} O86:B7 \textit{ΔwecA/ΔwaaL} was fed 2-ketoGlc (3-11) and incubated overnight, from which the LPS was isolated in analyzed with western blot. To both the \textit{E. coli} O86:B7 \textit{ΔwecA/ΔwaaL} fed 2-ketoGlc and void of 2-ketoGlc, the cells were incubated with aminooxybiotin, which biotinylates cells that incorporate the 2-ketoGlc into the cell surface oligosaccharides. After excessive washing of the cells to remove unreacted aminooxybiotin, the LPS was isolated and separated electrophorically by SDS-PAGE. Then after transfer of the LPS to a nitrocellulose membrane, the presence of biotinylated LPS was probed by using HRP conjugated to streptavidin. Figure 3.7 lane 2 shows the single band of core oligosaccharide for the \textit{E. coli} O86:B7 \textit{ΔwecA/ΔwaaL}, and in Figure 3.7 lane 4, we observe the same band in the western blot. Furthermore, absence of the band in Figure 3.7 lane 3, suggests that \textit{E. coli} O86:B7 \textit{ΔwecA/ΔwaaL} is incorporating the 2-ketoGlc into the core oligosaccharides.
3.4 Conclusion

In conclusion, we describe the chemoenzymatic synthesis of the key donor molecule (UDP-2-ketoGlc, 3-13) of 2-ketoGlc in vitro using our developed two enzyme system (NahK and GlmU). In addition, we demonstrate that the 2-ketoGlc (3-11) is incorporated in some form into the LPS core oligosaccharide, however due to the complexity of the biosynthetic pathways in which this sugar could conceivably interact with, it remains a challenge to identify the exact location that this sugar occupies on the LPS.\textsuperscript{29} The key to this success, however, likely resides in the promiscuity of sugar nucleotide biosynthetic pathways, when creating the sugar nucleotide donor UDP-2-ketoGlc (3-13). Furthermore, this result may also demonstrate the promiscuity of the glycosyltransferases which are responsible for the synthesis of the oligosaccharides that are eventually displayed on the cell surface. In conjunction with the incorporation of modified fucose derivatives into the LPS of \textit{E. coli} O86:B7, we further demonstrate proof of principle that other unique monosaccharides can be engineered into the cell surface oligosaccharides in bacteria. Unlike our previous work where we could control the exact location of the modification to the LPS, because of the possible pathways that 2-ketoGlc may take, exact identification of which residue is being modified in the core oligosaccharides is challenging. Also in conjunction with previous observations, we observed a competition effect, whereby it was critical to have an extracellular concentration of 2-ketoGlc at 10 mM. This is likely important for competing with the high intracellular concentrations of UDP-GlcNAc.\textsuperscript{12} Further studies are ongoing by incorporating different monosaccharides (GalNAc, GlcNAc) which include different modifications (\textit{e.g.} azido, alkyne) at different positions which may expand the
understanding in which biosynthetic pathways the 2-acetamido-2-deoxy-D-glycopyranosides analogues interact with.

3.5 Experimental Section

2-Deoxy-2-iodo-1,3,4,6-tetra-O-acetyl-α-D-mannopyranoside (3-5):

3,4,6-tri-O-acetyl glucal (200 mg, 0.732 mmol) was dissolved in dry CH₂Cl₂ (12 mL) under argon in a foil covered flask. To the resulting solution was added AcOH (84 μL, 1.472 mmol) followed by NIS (273 mg, 1.21 mmol). After 40 min, the starting material was almost consumed and the reaction was quenched by 10% Na₂S₂O₃ (20 mL) and the reaction mixture was stirred vigorously for 5 min. The mixture was diluted with 100 mL EtOAc, washed with 10 mL 10% Na₂S₂O₃ and brine. The organic layer was collected and dried over Na₂SO₄. The solvent was removed and the residue was purified by column chromatography (hexanes/EtOAc 4:1) to afford compound 3-5 as a white foam (238 mg, 71%). ¹H NMR (400 MHz, CDCl₃) δ 6.40 (d, J = 1.4 Hz, 1H), 5.46 (t, J = 9.7 Hz, 1H), 4.60 (dd, J = 9.5, 4.4 Hz, 1H), 4.54 (dd, J = 4.4, 1.5 Hz, 1H), 4.23 (dd, J = 12.4, 4.4 Hz, 1H), 4.11-4.19 (m, 2H), 2.17 (s, 3H), 2.13 (s, 3H), 2.12 (s, 3H), 2.08 (s, 3H). Compound 3-8 was also isolated as a minor isomer (<10%). NMR data also see Lafont et al.³⁰
**3,4,6-Tri-O-tert-butyldiphenylsilyl-glucal (3-6):** 3,4,6-tri-O-acetyl glucal (3-4, 2.01 g, 7.36 mmol) was mixed with anhydrous methanol (20 mL) and treated with 5 drops of 25 wt.% sodium methoxide solution in methanol. The mixture was stirred at rt. for 30 min and became a clear solution before it was neutralized with solid CO₂. The solvent was removed under reduced pressure to afford the triol as a light yellow syrup, which was used without further purification. The triol was redissolved in a mixture of 20 mL THF and 10 mL DMF. Then imidazole (2.5 g, 36.7 mmol) and TBDPSCl (7.2 mL, 28.1 mmol) were added and the reaction mixture was heated at 70 °C for 18h. After cooling to rt., the reaction mixture was diluted with EtOAc (50 mL) and hexanes (10 mL), and washed with water (3*50 mL). The organic layer was dried over Na₂SO₄, filtered, concentrated under reduced pressure and the residue was purified by chromatography (hexanes/EtOAc 50:1) to afford compound 3-6 as a clear syrup (4.9g, 77%). ¹H NMR (500 MHz, CDCl₃) δ = 7.09-7.56 (m, 30H), 6.20 (d, J = 6.4 Hz, 1H), 4.33-4.36 (m, 1H), 4.16-4.19 (m, 1H), 4.08 (dd, J = 11.6, 8.3 Hz, 1H), 3.89 (q, J = 1.8 Hz, 1H), 3.64-3.67 (m, 2H), 0.95 (s, 9H), 0.84 (s, 9H), 0.66 (s, 9H). NMR data also see Friesen et al.³¹

**2-Deoxy-2-iodo-3,4,6-tri-O-tert-butyldiphenylsilyl-β-D-glucopyranosyl acetate (3-7):** Compound 3-6 (2.4 g, 2.79 mmol) was dissolved in 18
mL toluene and acetic acid (1.02 mL, 17.8 mmol) and NIS (1.32 g, 5.87 mmol) were added. The reaction mixture was stirred vigorously at 100 °C for 10 min. Upon cooling, 0% aq. Na₂S₂O₃ was added until the solution was colorless. Saturated aq. NaHCO₃ (40 mL) was added slowly and the solution was then extracted with EtOAc (2*100 mL). The organic layer was dried over Na₂SO₄, filtered, concentrated under reduced pressure and the residue was purified by chromatography (hexanes/EtOAc 50:1) to afford compound 3-7 (2.42 g, 83%) as a white foam which was directly used in the next deprotection step. NMR data see Chong et al.⁴²

2-Deoxy-2-iodo-1,3,4,6-tetra-O-acetyl-β-D-glucopyranoside (3-8): To a solution of compound 3-7 (2 g, 1.91 mmol) in THF (50 mL) was added 1 M TBAF solution in THF (9 mL, 9 mmol). The reaction mixture was stirred at rt. for 4h. The solvent was removed under reduced pressure to afford the triol as a dark brown syrup which was redissolved in anhydrous pyridine (30 mL) and treated with acetic anhydride (4 mL) and DMAP (20 mg, cat.). The mixture was stirred at rt. for 3h. Then the solvent was removed under reduced pressure and the residue was purified by chromatography (hexanes/EtOAc 5:1-4:1) to afford compound 3-8 (744 mg, 85%) as a clear syrup. ¹H NMR (500 MHz, CDCl₃) δ = 5.88 (d, J = 9.6 Hz, 1H), 5.35 (dd, J = 11.2, 9.0 Hz, 1H), 5.02 (app t, J = 9.6 Hz, 1H), 4.33 (dd, J = 12.5, 4.5 Hz, 1H), 4.11 (dd, J = 12.5, 2.1 Hz, 1H), 4.00 (dd, J = 11.1, 9.6 Hz, 1H), 3.89 (ddd, J = 10.1, 4.5, 2.2 Hz, 1H), 2.18 (s, 3H), 2.11 (s, 3H), 2.08 (s, 3H), 2.03 (s, 3H). NMR data also see Lafont et al.⁴⁰
2-Deoxy-2-(methallyl)-1,3,4,6-tetra-O-acetyl-β-D-glucopyranoside (3-9): Compound 3-8 (820 mg, 1.789 mmol) was dissolved in 20 mL anhydrous benzene and methallyl-tributyltin (4.32 g, 12.53 mmol) was added under Ar. The reaction mixture was degassed by bubbling Ar through the solution for 30 min. Then AIBN (59 mg, 0.358 mmol) was added and the reaction mixture was heated at 81 °C for 4 h. The solvent was removed under reduced pressure and the residue was partitioned between CH$_3$CN (100 mL) and pentane (200 mL). The CH$_3$CN layer was washed with pentane (100 mL) and concentrated under reduced pressure. The residue was purified by chromatography (hexanes/EtOAc 5:1) to afford compound 3-9 (459 mg, 66%) as a clear syrup, β-manno (93 mg, 13%) type was also isolated as a byproduct.

$^1$H NMR (500 MHz, CDCl$_3$) δ = 5.53 (d, $J = 9.0$ Hz, 1H), 4.97-5.03 (m, 2H), 4.66 (brs, 1H), 4.61 (brs, 1H), 4.30 (dd, $J = 12.4, 4.7$ Hz, 1H), 4.05 (dd, $J = 12.4, 2.3$ Hz, 1H), 3.75 (ddd, $J = 9.3, 4.6, 2.2$ Hz, 1H), 2.22-2.29 (m, 1H), 2.08-2.09 (m, 2H), 2.06 (s, 3H), 2.05 (s, 3H), 1.99 (s, 3H), 1.96 (s, 3H).

2-Deoxy-2-(acetonyl)-1,3,4,6-tetra-O-acetyl-β-D-glucopyranoside (3-10): To solution of compound 3-9 (127 mg, 0.329 mmol) in dioxane/H$_2$O (v/v 3:1, 6 mL) was added successively 2,6-lutidine (76 μL, 0.658 mmol), OsO$_4$ (2.5% in 2-methyl-
2 propanol, 83 μL, 0.0066 mmol) and NaIO₄ (281 mg, 1.315 mmol). The reaction mixture was stirred at rt. overnight. The mixture was partitioned between CH₂Cl₂ (30 mL) and H₂O (15 mL) and the aqueous phase was extracted with CH₂Cl₂ (10 mL). The combined organic layer was dried over Na₂SO₄, filtered, concentrated under reduced pressure and the residue was purified by chromatography (hexanes/EtOAc 2:1) to afford compound 3-10 (111 mg, 87%) as a white solid. ¹H NMR (500 MHz, CDCl₃) δ 5.70 (d, J = 9.2 Hz, 1H), 5.10 (dd, J = 10.8, 9.2 Hz, 1H), 5.03 (app t, J = 9.7 Hz, 1H), 4.31 (dd, J = 12.4, 4.5 Hz, 1H), 4.08 (dd, J = 12.4, 2.1 Hz, 1H), 3.78-3.81 (m, 1H), 2.52-2.58 (m, 1H), 2.42 (m, 2H), 2.12 (s, 3H), 2.09 (s, 3H), 2.08 (s, 3H), 2.01 (s, 3H), 2.01 (s, 3H); ¹³C NMR (125 MHz, CDCl₃) δ 205.2, 170.8, 170.6, 169.8, 169.0, 93.7, 73.5, 72.8, 68.9, 62.0, 41.4, 40.9, 30.0, 20.9, 20.86, 20.8, 20.7; HRMS (ESI) calcd for C₁₇H₂₄O₁₀Na (M+Na)⁺ 411.1262, found 411.1249 m/z.

![Chemical Structure](image)

**2-Deoxy-2-(acetyl)-β-D-glucopyranose (3-11):** Compound 3-10 (63.5 mg, 0.164 mmol) was dissolved in a mixture of MeOH/H₂O/Et₃N (v:v:v 5:2:1, 5 mL) and stirred at rt. for 18 h. The solvent was removed under reduced pressure and the residue was purified by chromatography (CH₂Cl₂/MeOH 15:1) to afford 2-ketoGlc 3-11 as a colorless syrup (α and β mixture ~1:1 as indicated by H NMR, 15 mg, 42%). ¹H NMR (500 MHz, D₂O, α and β mixture ~1:1) δ 5.24 (d, J = 3.3 Hz, 1H, H-1α), 4.69 (d, J = 8.7 Hz, 1H, H-1β), 3.91 (dd, J = 12.3, 2.1 Hz, 1H), 3.74-3.87 (m, 5H), 3.59 (dd, J =
10.9, 9.1 Hz, 1H), 3.39-3.45 (m, 3H), 2.88 (dd, $J = 17.4, 5.5$ Hz, 1H), 2.76 (dd, $J = 16.2, 5.5$ Hz, 1H), 2.65-2.71 (m, 2H), 2.28 (s, 3H), 2.276 (s, 3H), 2.24-2.30 (m, 1H), 2.06-2.12 (m, 1H); $^{13}$C NMR (125 MHz, D$_2$O, $\alpha$ and $\beta$ mixture ~1:1) $\delta$ 215.7, 215.3, 97.3 (C-1$\alpha$), 93.3 (C-1$\beta$), 76.7, 75.4, 72.6, 72.5, 71.6, 71.4, 61.7, 61.6, 46.6, 43.0, 42.5, 42.4, 30.6 (2 CH$_3$); HRMS (ESI) calcd for C$_9$H$_{16}$O$_6$Na (M+Na)$^+$ 243.0839, found 243.0829 m/z.

\[
\text{2-Deoxy-2-(acetonyl)-}\alpha\text{-D-glucosyl diammonium phosphate (3-12):} \quad \text{The reaction mixture (1.249 mL) contained 40 mM GlcNAc or its analogue (3-11), 50 mM ATP, 10 mM MgCl$_2$, and 1.5 mg/mL NahK in 100 mM Tris-HCl buffer (pH 9.0). After incubation at 37°C for 48 h, additional NahK was added to a final concentration of 2.25 mg/mL. The mixture was incubated at 37°C for another 17 h. Upon completion, the reaction mixture was briefly boiled for 2 min and then centrifuged to remove protein. The supernatant was concentrated under reduced pressure and the residue was purified by normal phase silica gel column chromatography using gradient CH$_2$Cl$_2$/5 mM NH$_4$HCO$_3$ in methanol as eluent (CH$_2$Cl$_2$/5 mM NH$_4$HCO$_3$ in methanol 1:1 to 0:1). The fractions containing the products were collected and concentrated under reduced pressure. The residue was further purified by Bio-Gel P-2 column. The fractions containing products were combined and lyophilized to provide compound 3-12 as a white foam (13.2 mg, 79%, diammonium salt). $^1$H NMR (500 MHz, D$_2$O) $\delta$ 5.47 (dd, $J = 7.7, 3.0$ Hz, 1H), 3.89-3.95 (m, 2H), 3.81 (dd, $J = 12.4, 5.2$ Hz, 1H), 3.67 (dd, $J = 11.0, 9.1$ Hz, 1H).} \]
Hz, 1H), 3.45 (app t, J = 9.6 Hz, 1H); 2.88 (dd, J = 18.0, 5.5 Hz, 1H), 2.78 (dd, J = 18.0, 7.2 Hz, 1H), 2.30 (s, 3H), 2.27-2.34 (m, 1H); \(^{13}\)C NMR (125 MHz, D\(_2\)O) \(\delta\) 214.5 (d), 94.8, 72.7, 71.7, 70.7, 60.9, 42.3 (d), 41.4, 29.9 (d); \(^{31}\)P NMR (202 MHz, D\(_2\)O): \(\delta\) 2.39; HRMS (ESI) calcd for C\(_9\)H\(_{16}\)O\(_9\)P (M-H)- 299.0537, found 299.0545 m/z.

Uridine 5'-diphospho-2-acetonyl-2-deoxy-\(\alpha\)-d-glucopyranose diammonium salt (3-13): The reaction was carried out in a cocktail (3.952 mL) containing 10 mM 2-ketoGlc-1-P (3-12), 15 mM UTP, 10 mM MgCl\(_2\), 100 mM Tris-HCl (pH 7.5), 1 mg/mL GlmU and 1U/mL inorganic pyrophosphatase. After incubating at 37 °C for 15 h, the mixture was heated in boiling water for 2 min followed by centrifugation to remove protein. The supernatant was directly loaded on DEAE-cellulose anion exchange resin using gradient ammonium bicarbonate as eluent (0 to 250 mM). The fractions containing the products were collected and freeze-dried and the residue was further purified by Bio-Gel P-2 column. The fractions containing products were combined and freeze-dried to provide compound 3-13 as a white foam (16.3 mg, 64%). \(^1\)H NMR (400 MHz, D\(_2\)O): \(\delta\) 8.01 (d, J = 8.1 Hz, 1H), 6.00-6.02 (m, 2H), 5.58 (dd, J = 7.2, 2.9 Hz, 1H), 4.39-4.43 (m, 2H), 4.19-4.32 (m, 3H), 3.87-3.95 (m, 2H), 3.82 (dd, J = 12.4, 4.4 Hz, 1H), 3.67 (dd, J = 10.6, 9.1 Hz, 1H), 3.49 (app t, J = 9.7 Hz, 1H), 2.76-2.92 (m, 2H), 2.29 (s, 3H), 2.26-2.36 (m, 1H); \(^{13}\)C NMR (100 MHz, D\(_2\)O): \(\delta\) 214.1,
166.2, 151.8, 141.7, 102.7, 96.0 (d), 88.4, 83.3 (d), 73.8, 73.1, 71.5, 70.4, 69.6, 64.9 (d), 60.5, 42.1 (d), 41.2, 29.9; $^{31}$P NMR (162 MHz, D$_2$O): $\delta$ = -11.5 (d, $J$ = 19.9 Hz), -12.9 (d, $J$ = 19.9 Hz); HRMS (ESI) calcd for C$_{18}$H$_{27}$N$_2$O$_17$P$_2$ (M-H)$^-$ 605.0790, found 605.0814 m/z.

**Experimental for in vivo remodeling**

**Bacterial strains and reagents**

*E. coli* competent cell BL21 (DE3) [F ompT hsdS$_B$ (rBmB) gal dcm (DE3)] was obtained from Stratagene. All reagents were purchased from Sigma unless otherwise noted.

**Labeling live bacterial cells via aniline-catalyzed oxime ligation**

Either *E. coli* BL21 (DE3), and *E. coli* O86:B7 $\Delta$wecA/$\Delta$waaL was inoculated in 200 µL in LB medium supplemented with the sugar analogue (peracetylated GlcNAc was used as negative control), grown overnight, at 37 °C, and shaking at 250 rpm. Cells were collected by centrifugation at 4 °C, washed 3 times with PBS buffer, and subsequently resuspended in a solution containing 20 mM MOPS, pH 6.7, 1 mM $N$-(aminooxyacetyl)-$N'$-(D-biotinoyl) hydrazine (Invitrogen), and 10 mM aniline. Cells were then incubated at 4 °C 90 minutes (on a rocker), after which the cells were harvested by centrifugation, and subsequently washed 3 times with PBS buffer. Cells were then resuspended in 0.5 mL of PBS, to which 5 µL of FITC conjugated avidin (0.5 mg/mL, Invitrogen) was added. The resuspension was incubated at 4 °C on a rocker for 90 minutes, after which the cells were extensively washed with PBS buffer, and subsequently analyzed by both
flow cytometry (BD FACS LSR II) and fluorescence microscopy (Olympus IX81 inverted microscope).

*Functional inactivation of wecA and waaL*

The *wecA* and *waaL* genes were replaced by a kanamycin resistance cassette, using the RED recombination system of phage lambda. For creation of the knock out strain we used the plasmids provided by the Quick and Easy *E. coli* Gene Deletion Kit (Gene Bridges). The *wecA* gene was first replaced from the chromosomal DNA with the kanamycin resistance cassette following the manufacturer’s protocol, which was verified by PCR. The kanamycin resistance cassette was then removed by FLP/FLPe expression and the *E. coli* O86:B7 ΔwecA was then used as the host to create the *waaL* inactivated strain, which was created following the manufacturer’s protocol. The final *E. coli* O86:B7 ΔwecA/ΔwaaL was confirmed by PCR as well as observing the change in LPS phenotype by LPS silver staining.

*LPS isolation and analysis*

To determine the relative location of the labeled sugar *E. coli* O86:B7 ΔwecA/ΔwaaL was inoculated with 10 mM of the sugar analogue and subsequently labeled with the aminooxybiotin as mentioned. Then the LPS was extracted according to the published proteinase K digestion protocol, and analyzed by both SDS-PAGE (with silver staining) and western blot. To perform the western blot analysis, the LPS was separated by SDS-PAGE, transferred to a nitrocellulose membrane, and probed with HRP conjugated
streptavidin (1:5000, Thermo Scientific). The blot was developed by using the ECL Western Blotting Detection Reagents (GE Healthcare).

References


CHAPTER 4

ENZYMATIC SYNTHESIS AND PROPERTIES OF URIDINE-5’-O-(2-THIODIPHOSPHO)-N-ACETYLGLUCOSAMINE/N-ACETYLGALACTOSAME

4.1 Introduction

In Chapter 2 and Chapter 3, we described an efficient enzymatic system for the synthesis of sugar nucleotides and especially, their derivatives. Specifically, replacement of an oxygen atom with sulfur atom in the phosphate esters represents a novel class of useful analogues which has been exploited frequently for the previous four decades in nucleotide and nucleoside chemistry.\(^1\)\(^-\)\(^3\) (Figure 4.1) The sulfur analogues are different from their oxygen counterparts and could be used as useful probes for the elucidation of enzymatic mechanisms and biological pathways. For example, nucleotide analogues nucleoside 5’-phosphorothionates which had resistance to alkaline phosphatase were synthesized by thiophosphorylation reactions.\(^1\) Synthesis of nucleotide analogues bearing a phosphorothionate moiety at the terminal of the polyphosphate chain were also reported.\(^4\) The terminal thiophosphate could be used to investigate phosphate transfer reactions in biological systems.
However, there have been only a few examples for phosphorothionate analogues of sugar nucleotides till now. Uridine-5'-O-(2-thiodiphosphoglucose) and uridine-5'-O-(2-thiodiphosphoglucuronic acid) were enzymatically prepared and their properties as substrates towards related enzymes were explored, such as glycosyltransferase, pyrophosphorylase, UDP-galactose-4-epimerase, and UDP-glucose dehydrogenase.\textsuperscript{5,6} One notable feature of thiophosphate is that it is less electrophilic than phosphate itself because sulfur atom is less electronegative than oxygen. This property improves resistance of thiophosphate to hydrolysis\textsuperscript{1,4} while does not influence the substrate binding affinities towards related enzymes. For instance, thiophosphorylated proteins showed resistance to phosphatase in dephosphorylation reactions which represented an approach to increase metabolic stability of target proteins.\textsuperscript{7}

In the Leloir-type glycosylation reactions (Figure 4.2), weak electronegativity of sulfur makes the uridine-5’-O-\(\beta\)-thiodiphosphate (\textbf{UDP(\(\beta\)S)}) a poorer leaving group and thus slow downs the reaction which could possibly result in a stable substrate/enzyme (GT) complex (Figure 4.2). This chapter describes an enzymatic approach to obtain a thio-containing UDP-GlcNAc/GalNAc analogue. We use an assay based on capture of

Figure 4.1. Structures of some typical phosphorothionate analogues.
the carbohydrate and analysis by mass spectrometry to quantitatively characterize the activity of this unnatural sugar donor in a *Neisseria meningitidis* 1,3-N-acetylglucosaminyltransferase (LgtA)-mediated glycosylation reaction.

Figure 4.2. Leloir-type glycosylation reaction using sugar nucleotide thio analogue

![Chemical structure](image)

### 4.2 Enzymatic synthesis of UDP(βS)-GlcNAc/GalNAc

Commercially available ATPγS is already commonly used as a thiophosphoryl donor to obtain thiophosphorylated nucleosides or proteins.\(^7\)-\(^9\) We replaced ATP with ATPγS in the NahK catalyzed phosphorylation reaction to construct both GlcNAc-1-P(S) (4-1) and GalNAc-1-P(S) (4-2) in good yields. (Scheme 4.1) We then used GlmU to catalyze a uridylyltransfer reaction between the anomeric thiophosphates 4-1/4-2 and UTP to produce the corresponding (UDP(βS)-GlcNAc/GalNAc, 4-3/4-4) in moderate
yields. A yeast inorganic pyrophosphatase was added to the reaction to degrade the byproduct pyrophosphate and increase the yield. We assigned the products as the β-thiophosphate product based on $^{31}$P NMR. (Figure 4.3)

Scheme 4.1. Enzymatic synthesis of UDP(βS)-GlcNAc/GalNAc

As shown in Figure 4.3a, substitution of phosphoryl oxygen by sulfur leads to the downfield shift (~44 ppm) of phosphorus (δ -1.0 vs. 43.3). Comparing the $^{31}$P NMR spectra between natural UDP-GlcNAc (2-8) (δ -11.3 (d), -13.0 (d)) and UDP(βS)-GlcNAc (4-3) (δ 42.7 (d), -12.1 (d)), only one phosphorus signal was downfield shifted due to its bonding to the sulfur atom, indicating the formation of β-thiophosphate only (Figure 4.3b). Moreover, our observation is consistent with the previous discovery that the presence of sulfur leads to the changes in chemical shifts only for phosphorus atoms directly bonded with it while the other chemical shift remains practically the same.$^{11-12}$
Figure 4.3. $^{31}$P NMR comparison of the thio containing analogues with the oxygen versions

4.3 Properties of UDP(βS)-GlcNAc (4-3)

We evaluated the ability of the Neisseria meningitidis 1,3-N-acetylglicosaminylltransferase (LgtA)$^{13}$ to use the analogue (UDP(βS)-GlcNAc, 4-3) as a donor substrate in a glycosylation reaction. The LgtA mediated glycosylation reactions were performed on self-assembled monolayers (SAMs) of alkanethiolates on gold that presented lactose acceptors at a density of 10% against a background of tri(ethylene
glycol) groups and using UDP-GlcNAc or UDP(βS)-GlcNAc as the donor. (Figure 4.4a) (Also see Chapter 2 and Figure 2.9 for detailed introduction of this technique)

The glycol groups of the monolayer prevent non-specific interactions of the enzyme with the surface and maintain the activity of immobilized carbohydrates. The assay was performed by applying a drop containing LgtA and either UDP-GlcNAc (2-8) or UDP(βS)-GlcNAc (4-3) to the monolayer, allowing the mixture to react for a short time, and then rinsing the surface and analyzing the monolayer using matrix assisted laser desorption-ionization mass spectrometry (MALDI MS) to reveal the masses of the substituted alkanethiols. Hence, this method can directly observe both the substrate and the product of the LgtA-mediated reaction, erasing the need for labels and also providing a measure of the yield for the glycosylation reaction. Representative spectra of the glycosylation are shown in Figure 4.4b. The peaks at m/z 1835 and 1851 are the sodium and potassium adducts, respectively, of the lactose-terminated disulfide. After a 2 h reaction, the mass spectrum revealed new peaks at m/z 2038 and 2054 corresponding to the trisaccharide product. Hence, the analogue described above (UDP(βS)-GlcNAc) is a substrate for LgtA. For the glycosyltransfer reaction shown in Figure 4.4a (using either natural UDP-GlcNAc or unnatural UDP(βS)-GlcNAc), we performed a kinetic experiment by repeating the experiment for reaction times ranging up to 2 hours and show plots of the time course for both natural UDP-GlcNAc (2-8) and UDP(βS)-GlcNAc (4-3) in Figure 4.4c. It is apparent that the analogue reacts much slower than the natural UDP-GlcNAc during the glycosyltransfer reaction. Additionally, when both donors were present in the same reaction, the reaction/time course was quite similar to that of natural UDP-GlcNAc alone, indicating the analogue does not inhibit the activity of LgtA.
Figure 4.4. a. LgtA mediated reaction using UDP(βS)-GlcNAc (4-3) as a donor molecule and self-assembled monolayers (SAMs) that present the lactose acceptor; b. Representative spectra of the glycosylation showing acceptor and product peaks; c. Reaction progresses of both UDP-GlcNAc and UDP(βS)-GlcNAc.

Figure 4.5. Effect of divalent ions and EDTA on the activity of UDP(βS)-GlcNAc and UDP-GlcNAc.

163
Figure 4.6. Glycosylation reactions were performed in solution using an azido modified substrate in order to determine the kinetic parameters for UDP-GlcNAc and the synthesized analogue. Reactions were terminated at different time points and then applied to the monolayers to allow immobilization and SAMDI characterization of both the lactose substrate and trisaccharide product. The lower panel shows the spectrum from a 120 min reaction of UDP(βS)-GlcNAc. Black and red label the peaks of lactose and the trisaccharide products, respectively. Letters in parenthesis showed the different ion adducts appearing in the spectrum.

To further understand the consequences of the sulfur substituted donor on enzyme activity, we first explored the influence of divalent ions on activity. Assays were performed in solution to avoid perturbations that may arise from presentation of the ligands at the surface. In this ‘pull-down’ format, the reactions were performed in
solution with an azido-modified lactose as the acceptor and either UDP-GlcNAc (2-8) or UDP(βS)-GlcNAc (4-3) as the donor (Figure 4.6). The reactions were stopped by adding cold ethanol and then spotted onto a monolayer that presented a terminal alkyne group among tri(ethylene glycol) groups (For details see Experimental Section). In this way, a copper (I)-catalyzed click reaction between the azido sugars and the alkyne monolayer served to immobilize both the trisaccharide product and disaccharide acceptor, so that SAMDI analysis can then determine the extent of glycosylation (Figure 4.6). Some glycosyltransferases are known to require the presence of a divalent metal cation for enzymatic activity, and thus it is commonplace to assay the transferase. Different enzymes can utilize different cation metals in vivo, however which metal is actually used cannot be determined. Thus, we tested a panel of divalent metal cations to determine the optimal substrate for in vitro enzymatic activity. Figure 4.5 summarizes the relative activities of different divalent ions for both of the donors. We find that the two donors share a similar metal ion preference for divalent manganese, magnesium and calcium over the other metal ions. Iron, molybdenum, palladium and cobalt salts result in diminished activity. The addition of EDTA totally abolished the activity of the enzyme. This similarity suggests that the divalent metal binds in comparable fashion to the pyrophosphate portion of both donors and the oxygen atoms, instead of the sulfur, probably make major contribution to the interaction.

The kinetic parameters for enzyme-mediated transfer of the analogue were also determined using the ‘pull-down’ format and SAMDI characterization. The enzyme assays were carried out by varying the concentration of one substrate in the presence of different fixed concentrations of the other. The initial velocities were plotted graphically
in double reciprocal form$^{16}$ to check the linearity of the relationship and to determine the patterns of the plots (Figure 4.7 in Experimental Section). The plots showed intersecting patterns and with the donor as the variable substrate, the plots intersect on the y axis. We therefore fit the data using Cleland’s$^{17}$ equation for an ordered bi-bi model: \( v_0 = \frac{V_{AB}}{(K_{iA}K_b + K_bA + K_aB + AB)} \). (\( v_0 \): initial velocity; A and B: concentration of the variable substrate and fixed substrate, respectively; \( K_{iA} \): inhibition constant of A; \( K_a \) and \( K_b \): apparent Michaelis- Menten constants) The kinetic parameters determined from a best-fit of the data are summarized in Table 4.1. Kinetic parameters for UDP-GlcNAc are consistent with the previously reported values.$^{13}$ The \( K_m \) for the analogue, by contrast, is about 150-fold greater and the \( V_{max} \) is substantially decreased. This decrease in activity likely owes to the lower electronegativity of the sulfur relative to the oxygen atom, which makes the uridine-5’-O-β-thiodiphosphate a poorer leaving group in the glycosyltransfer reaction. It is also possible that larger sulfur atom can result in a less favorable interaction between the anomeric center of the donor and the hydroxyl group of the acceptor in the transition state mimicking the oxo-carbenium ion of the S_N2 glycosyltransfer reaction (Scheme 1.2).

Table 4.1. Summary of kinetic data for LgtA catalyzed reaction using azido modified lactose and UDP(βS)-GlcNAc and UDP-GlcNAc as substrates. a: azido modified lactose; b: donor. The number is reported as the average of three parallel experiments. The number in parenthesis shows the standard error.

<table>
<thead>
<tr>
<th></th>
<th>( K_{iA, Apparate} ) (mM)</th>
<th>( K_{ai} ) (mM)</th>
<th>( K_b ) (mM)</th>
<th>( V_{max} ) (mM min(^{-1}))</th>
<th>[E(_0)] (mg mL(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDP(βS)-GlcNAc (4-3)</td>
<td>6.08 (0.93)</td>
<td>15.0 (2.75)</td>
<td>51.5 (4.3)</td>
<td>0.211 (0.194)</td>
<td>1.63</td>
</tr>
<tr>
<td>UDP-GlcNAc (2-8)</td>
<td>5.84 (0.71)</td>
<td>14.8 (1.18)</td>
<td>0.40 (0.22)</td>
<td>0.456 (0.131)</td>
<td>0.204</td>
</tr>
</tbody>
</table>
In summary, we provide here another example where the multi-milligram scale synthesis of sulfur containing UDP-GlcNAc/GalNAc analogues (4-3 and 4-4) were easily achieved using our two-enzyme pathway. Furthermore, we took advantage of the combination of self-assembled monolayers (SAMs) and MALDI MS to assay for glycosyltransferase activities. This assay is convenient, label-free and requires very little materials to achieve both quantitative and qualitative measurements.\textsuperscript{18-19} This work will benefit the investigation of more unnatural sugar donors which have chemical or biological significance in the future.

4.4 Experimental Section

General

Adenosine 5’-[\gamma-thio]triphosphate tetralithium salt (\(\gamma(S)ATP\)), \(N\)-acetyl-\(D\)-glucosamine/galactosamine were purchased from Sigma-Aldrich. Direct On-Chip MALDI-TOF mass spectrometry is performed on Applied Biosystems\textsuperscript{TM} 4800 plus MALDI TOF/TOF mass spectrometer.

NMR data

Compound 4-1, 4-2, 4-3, 4-4 were enzymatically synthesized as previously described in Chapter 2 Experimental Section.\textsuperscript{20}
Thiophosphate (4-1): $^1$H NMR (500 MHz, D$_2$O) $\delta$ 5.57 (dd, $J$ = 10.0, 3.2 Hz, 1H), 4.01-4.05 (m, 1H), 3.97-3.99 (m, 1H), 3.94 (dd, $J$ = 12.2, 2.2 Hz, 1H), 3.80-3.87 (m, 2H), 3.54 (app t, $J$ = 9.3 Hz, 1H), 2.11 (s, 3H); $^{13}$C NMR (125 MHz, D$_2$O) $\delta$ 174.8, 92.9 (d), 72.4, 71.5, 70.0, 60.7, 54.1 (d), 22.3; $^{31}$P NMR (162 MHz, D$_2$O) $\delta$ 43.3; HRMS (ESI) calcd for C$_8$H$_{15}$NO$_8$PS (M-H)$^-$ 316.0261, found 316.0267 m/z.

Thiophosphate (4-2): $^1$H NMR (500 MHz, D$_2$O) $\delta$ 5.60 (dd, $J$ = 10.0, 3.5 Hz, 1H), 4.27-4.30 (m, 1H), 4.23-4.26 (m, 1H), 4.07 (app d, $J$ = 3.0 Hz, 1H), 4.00 (dd, $J$ = 10.9, 3.2 Hz, 1H), 3.84 (dd, $J$ = 11.7, 7.7 Hz, 1H), 3.78 (dd, $J$ = 11.7, 4.8 Hz, 1H), 2.12 (s, 3H); $^{13}$C NMR (125 MHz, D$_2$O) $\delta$ 175.0, 93.1 (d), 71.5, 68.7, 68.4, 61.4, 50.3 (d), 22.4; $^{31}$P NMR (162 MHz, D$_2$O) $\delta$ 43.4; HRMS (ESI) calcd for C$_8$H$_{15}$NO$_8$PS (M-H)$^-$ 316.0261, found 316.0265 m/z.
Uridine-5′-O-(2-thiodiphospho)-N-acetylgalactosamine (4-3): $^1$H NMR (400 MHz, D$_2$O) $\delta$ 8.03 (d, $J = 8.2$ Hz, 1H), 6.01-6.04 (m, 2H), 5.72 (dd, $J = 9.8$, 3.3 Hz, 1H), 4.41-4.47 (m, 2H), 4.25-4.34 (m, 3H), 3.98-4.04 (m, 2H), 3.82-3.92 (m, 3H), 3.61 (app t, $J = 9.8$ Hz, 1H), 2.10 (s, 3H); $^{13}$C NMR (100 MHz, D$_2$O) $\delta$ 174.8, 166.3, 151.9, 141.8, 102.8, 94.7 (d), 88.3, 83.4 (d), 73.8, 73.2, 70.9, 69.9, 69.5, 65.1 (d), 60.2, 53.7 (d), 22.1; $^{31}$P NMR (162 MHz, D$_2$O) $\delta$ 42.6 (d, $J = 29.4$ Hz), 12.1 (d, $J = 29.3$ Hz); HRMS (ESI) calcd for C$_{17}$H$_{26}$N$_3$O$_{16}$P$_2$S (M-H)$^-$ 622.0514, found 622.0531 m/z.

Uridine-5′-O-(2-thiodiphospho)-N-acetylgalactosamine (4-4): $^1$H NMR (500 MHz, D$_2$O) $\delta$ 8.01 (d, $J = 8.0$ Hz, 1H), 6.02-6.04 (m, 2H), 5.75 (dd, $J = 9.7$, 3.5 Hz, 1H), 4.43-4.47 (m, 2H), 4.24-4.34 (m, 5H), 4.09-4.10 (m, 1H), 4.01 (dd, $J = 10.9$, 3.1 Hz, 1H), 3.76-3.82 (m, 2H), 2.11 (s, 3H); $^{13}$C NMR (125 MHz, D$_2$O) $\delta$ 175.0, 166.9, 152.4, 141.7, 102.8, 95.0 (d), 88.4, 83.3 (d), 73.8, 72.2, 69.9, 68.4, 67.7, 65.1, 60.9, 49.8 (d), 22.2; $^{31}$P NMR (162 MHz, D$_2$O) $\delta$ 42.6 (d, $J = 29.7$ Hz), 12.0 (d, $J = 28.1$ Hz); HRMS (ESI) calcd for C$_{17}$H$_{26}$N$_3$O$_{16}$P$_2$S (M-H)$^-$ 622.0514, found 622.0525 m/z.
Preparation of self-assembled monolayers on gold coated slides

The gold substrate was prepared as reported.\textsuperscript{19} Briefly, glass coverslips were cleaned by sonication for 30 min first in deionized ultrafiltered (DIUF) water and then in ethanol and dried under a stream of nitrogen. Titanium (5 nm) and gold (50 nm) were evaporated onto the glass coverslips using an electron beam evaporator (Boc Edwards) at a rate of 0.05-0.10 nm s\textsuperscript{-1} and at a pressure of 1.0 × 10\textsuperscript{-6} Torr. The azido-modified lactose and alkyne-terminated alkanethiol (as shown in Figure 4.6) were prepared as previously reported.\textsuperscript{21-22} Monolayers were prepared as described by Li et al.\textsuperscript{21} Briefly, gold-patterned slides were immersed in an ethanolic solution of alkyne-terminated alkanethiol (or lactose-terminated disulfide) and tri(ethylene glycol)-terminated alkanethiol (or disulfide) in a ratio of 1:9 for 12 h at room temperature (total concentration of alkanethiol or disulfide: 1 mM). The substrates were washed with ethanol and dried under nitrogen.

Enzyme assays

The enzyme buffer used in both the on-chip and pull-down assay was Tris-HCl (100 mM, pH 7.5) with MnCl\textsubscript{2} or other divalent ions (10 mM). For the on-chip assay, 2 μL reaction cocktail, which contains the enzyme buffer, LgtA (0.816 mg mL\textsuperscript{-1}) and one of the donors (2 mM), was applied to the lactose-presenting monolayer on the gold-patterned slide. Reactions were carried out for times ranging from 5 to 120 min for the reaction progress plots and stopped by adding 1 μL ethanol to the corresponding gold chip and quickly removing the mixture by pipetting. At the end of the last reaction, the slide was rinsed with water, ethanol and dried under nitrogen. For the in-solution assay, the reactions for each donor were performed under the same conditions except for higher
LgtA concentration (1.63 mg mL\(^{-1}\)) for UDP(\(^\beta\)S)-GlcNAc (4-3). The reactions for measuring relative activities of different divalent ions were stopped at 10 min for each metal. The reactions for kinetics measurements were carried out for times ranging from 2 min to 30 min for UDP-GlcNAc (2-8) and 2 min to 60 min for UDP(\(^\beta\)S)-GlcNAc (4-3) with intervals of 2 or 3 minutes for the former and 5 min for the latter and stopped by adding cold ethanol mixed with 1 mM EDTA. At the end of the last reaction, a volume of 1 \(\mu\)L of the reaction mixture from each time point was transferred onto individual gold chips of the same sizes modified with the alkyne-terminated monolayer. An aqueous solution (1 \(\mu\)L per chip) containing copper bromide (2 mM) and triethylamine (0.5 mM) was applied to each circle and the reactions were incubated at room temperature for 30 min. The slide was then rinsed with water, ethanol and dried under nitrogen. For SAMDI measurement, monolayers were treated with matrix solution (2,4,6-trihydroxyacetophenone, 5 mg mL\(^{-1}\) in acetone). The slide was allowed to dry under atmospheric pressure and loaded in the MALDI TOF mass spectrometer. A 355 nm Nd:YAG laser was used as the desorption/ionization source with an accelerating voltage of 20 kV and extraction delay time of 50 ns. All spectra were acquired using positive reflector mode, and were Gaussian-smoothed and baseline corrected. For quantification, the extent of glycosylation (R) was determined from the relative peak intensity for product (Ip) and lactose substrate (Is) on SAMDI spectra using \(R = \frac{Ip}{Ip+Is}\). Previous report shows that this peak area ratio represents the actual ratio in solution within these concentration ranges.\(^{23}\) For determination of kinetics parameters, the donor substrate concentrations range from from 10 mM to 66.7 mM and the acceptor ranges from 5 mM to 30 mM. Double reciprocal plots of initial velocities (primary plots) are plotted in
Figure 4.7. The kinetics parameters are obtained from the secondary plots (Figure 4.8) derived from the primary plots.

Figure 4.7. Double reciprocal plots
Figure 4.8. Secondary plots from the primary plot where lactose is the variable substrate and UDP(βS)-GlcNAc is the fixed substrate.

References


CHAPTER 5

SYNTHESIS OF RARE SUGARS BY IN VITRO AND IN VIVO ALDOLASE REACTIONS

5.1 Introduction

5.1.1 Aldolases

New carbon-carbon bond formation is of great importance in organic synthesis. Aldol condensation reactions constitute a powerful approach to fulfill this transformation.¹ However, developing aldol condensation with high stereoselectivity at the newly formed stereocenters by catalyst control is the most significant and difficult issue in this process.² In recent years, the search for catalytic methods that can effectively introduce desired chirality into a complex structure has become a major effort in organic methodology. Two approaches have been used for the catalytic asymmetric aldol reaction: biocatalysis and small molecule catalysis. While both approaches have limitations, enzymatic approach (especially aldolases) is gaining increasing attentions because of a number of advantages: (1) they are highly chemoselective to substrates. (2) They are regiospecific and stereoselective. (3) Reactions can be performed under mild conditions and minimum protection group chemistry is employed.
Aldolases are a special class of lyases that catalyze the reversible stereoselective addition of a ketone donor (nucleophile) onto an aldehyde acceptor (electrophile). Up to now, more than 30 aldolases (EC 4.1.2.x) are known. Usually, the stereochemistry at the newly formed stereocenters is strictly controlled by the enzyme. This feature makes aldolases especially attractive in the synthesis of carbohydrates and sugar derivatives. While most aldolases are specific for their donor molecules in the aldol reaction, they often tolerate different aldehydes as acceptors. Hence, aldolases are classified depending on their donor (nucleophile) specificity (Table 5.1): Dihydroxyacetone phosphate (DHAP) aldolases which produce 2-keto-3,4-dihydroxy adducts; pyruvate/2-oxobutyrate aldolases which produce 3-deoxy-2-keto acids; acetaldehyde aldolases which result in the
formation of 3-hydroxyaldehydes; glycine aldolases which produce β-hydroxy-α-amino acids; and a recently found dihydroxyacetone aldolase which also generates 2-keto-3,4-dihydroxy adducts.⁴

Scheme 5.1. Two types of aldolase mechanisms.
Another discrimination of aldolases concerns the different enzyme mechanisms to activate the nucleophile (Scheme 5.1). So-called class I aldolases need no cofactors, as they exhibit a conserved lysine residue in the active site which forms Schiff base\(^5\) intermediate with the donor compound to generate the nucleophilic enamine species.\(^6\) On the other hand, class II aldolases are dependent on a metal ion (mainly Zn\(^{2+}\)) in the active site. Zn\(^{2+}\) acts as a Lewis acid and polarizes the carbonyl group of the donor.\(^7\) A glutamate residue (Glu-73) usually functions as a base to remove the proton from the donor to generate a zinc enolate species.\(^8\)\(^-\)\(^10\) It is worth noting that in class II aldolases, a tyrosine residue activates the incoming aldehyde acceptor by donating a proton and stabilizes the resulting charge as well (Scheme 5.1).

Figure 5.1. Stereocomplementary sets of DHAP-dependent aldolases.
DHAP-dependent aldolases constitute a family of lyases, which catalyzes the reversible aldol reaction of DHAP to a variety of aldehyde acceptors. Among aldolase families, DHAP-dependent aldolases are remarkably useful and widely investigated as the two new stereocenters generated in the aldol addition are predictable.\textsuperscript{3,11} Four reported DHAP-dependent aldolases are stereocomplementary, so that a complete set of diastereomers of vicinal diols can be achieved (Figure 5.1).\textsuperscript{3,12} The configuration of
vicinal diols produced from the aldol additional of DHAP with unnatural aldehydes usually follows that of the natural acceptors. As a result, this class of aldolases have been successfully applied to the synthesis of numerous carbohydrates and their derivatives.\textsuperscript{3,13-15} For example, the application of DHAP-dependent aldolases has been reported for the synthesis of various iminocyclitols and polyhydroxylated pyrrolidines, which are specific glycosidase inhibitors (Scheme 5.2.).\textsuperscript{14,16-18} Moreover, such complex structures as \textsuperscript{13}C-labeled sugars, long-chain sugars,\textsuperscript{19} thio sugars, and cyclitols\textsuperscript{20} which are not easily prepared by traditional methods could be obtained chemoenzymatically by DHAP-dependent aldolases using suitable acceptors.

Scheme 5.3. Preparation of DHAP and most prominent conversion of DHAP.
However, as mentioned above, DHAP-dependent aldolases show very strict requirement for their donor molecule DHAP, which is now commercial available but very expensive. Several chemical and enzymatic approaches have been developed for the synthesis of DHAP (Scheme 5.3),\textsuperscript{21} however, the yield and purity are not satisfying. Many of the reported enzymatic approaches have not advanced beyond small scale and chemical routes suffer from the use of toxic and expensive chemicals. Additionally, DHAP is usually stored as its stable precursor and generated as needed due to its unstable nature (Scheme 5.3).\textsuperscript{21} Therefore, the capability to generate DHAP from inexpensive sources could ultimately broaden the scope of aldolase reactions making it an attractive challenge.

### 5.2 Synthesis of rare sugars with DHAP-dependent aldoases \textit{in vitro}

Rare sugars are monosaccharides, and their derivatives, that are particularly uncommon in nature.\textsuperscript{22} Importantly, rare sugars possess many potential applications in the food, pharmaceutical and nutrition industries.\textsuperscript{23} In addition, rare sugars can be used as starting materials for the synthesis of intriguing natural products with important biological activities.\textsuperscript{13,23} Unfortunately, most rare sugars are quite expensive, and their synthetic routes are both limited and costly due to the expense of costly starting materials. As a specific example, D-psicose, a rare sugar and C-3 epimer of D-fructose, has the unique property of being an ideal sucrose substitute. Compared with sucrose, it has 70% the sweetness but provides no energy due to its suppressive effect toward hepatic lipogenic enzymes.\textsuperscript{24-25} Furthermore, it has been observed that foods supplemented with D-psicose exhibit higher antioxidant activity. Furthermore, D-psicose can be used as a
precursor in the synthesis of xylosylpsicoses, which are promising candidates for prebiotics, cosmetics and therapeutic uses.\textsuperscript{26} However, only two enzymes D-tagatose 3-epimerase from \textit{Pseudomonas cichorii}\textsuperscript{27-28} and D-psicose 3-epimerase from \textit{Agrobacterium tumefaciens}\textsuperscript{29} have been reported for D-psicose production. In addition, due to the fact that the interconversion between D-fructose and D-psicose is an equilibrium process, the large scale and high yield production of D-psicose remains quite challenging.

A potential solution to the above mentioned problems with rare sugar production lies in the use of aldolase reactions (\textit{i.e.} DHAP aldolase reactions), which are capable of generating vicinal diol diastereomers and are thus an ideal tool for the construction of rare sugars.\textsuperscript{12} In fact, the preparation of rare sugars, as well as their derivatives, represents the most significant applications of DHAP-dependent aldolases.\textsuperscript{13-14,30-31}

### 5.2.1 Synthesis of D-sorbose and D-psicose with rhamnulose-1-phosphate aldolase (RhaD)

Previously, L-rhamnulose-1-phosphate aldoase (RhaD) was used to couple DHAP with L-glyceraldehyde or racemic glyceraldehyde to produce L-fructose-1-phosphate. L-Fructose could then be obtained by dephosphorylation with acid phosphatase (AP).\textsuperscript{13,32} (Scheme 5.4) Furthermore, Wong and co-workers reported that RhaD preferentially made L-fructose and no diastereomers were found as by-products. Moreover, they concluded that pure L-glyceraldehyde was not necessary since D-glyceraldehyde was not consumed as a substrate by this enzyme.
Scheme 5.4. Enzymatic synthesis of L-fructose with RhaD showing perfect stereoselectivity

Scheme 5.5. Enzymatic synthesis of D-sorbose and D-psicose with RhaD and AP

Interestingly and contrasting to Wong’s results, in which D-glyceraldehyde was separated as the remaining starting material, when racemic glyceraldehyde was used, we
discovered DL-glyceraldehyde were both consumed in RhaD (from *Escherichia coli*)\textsuperscript{33} catalyzed reactions to give a complex mixture of addition products. Thus, we directly examined the RhaD reaction using pure D-glyceraldehyde and found this substrate was mostly consumed within 16 h. Subsequently, after dephosphorylation with AP, examination by TLC revealed that the corresponding product spot had the same Rf value as D-sorbose. A single peak (203.0526, [M + Na]\textsuperscript{+}) on high resolution mass spectrum was observed as well (Scheme 5.5).

![Figure 5.2](image.png)

**Figure 5.2.** HPLC (Hydrogen form, sulfonated divinyl benzene-styrene copolymer support and eluted with 5 mM H\textsubscript{2}SO\textsubscript{4}) profile of final reaction (after AP) mixture compared with authentic samples.

However, cation exchange HPLC and NMR revealed that what appeared to be one spot on TLC was actually two different products, D-sorbose and D-psicose (Scheme 5.5, Figure 5.2). While L-glyceraldehyde exclusively produced L-fructose (also proven in our one-pot reaction), the production of both D-sorbose and D-psicose interestingly
indicated aldolase RhaD had no stereo-preference when D-glyceraldehyde was the acceptor. Additionally, the stereoselectivity of this enzyme was affected by substrate configuration. The product ratio (D-sorbose/D-psicose = ~2/3) was determined by HPLC after calibration with standard curves (see Experimental Section Figure 5.8). After silica gel chromatography and gel filtration separation, the mixture containing only two rare sugars could be well separated by cation exchange resin (Ca$^{2+}$ form) chromatography$^{34-37}$ at 70 °C to provide pure D-sorbose and D-psicose (Figure 5.3).

Figure 5.3. Separation profile of D-sorbose and D-psicose with cation exchange resin (Ca$^{2+}$ form) eluted with ddH$_2$O under 70 °C.
Figure 5.3 shows the column with a thermostatic jacket and a typical separation profile of the products D-sorbose and D-psicose with the separation mainly resulting from different coordination affinities of the monosaccharides (OH groups) to the fixed Ca\(^{2+}\). Prior temperature studies indicated that the best separation was achieved under 70 °C compared with separation profiles under 50 °C, 60 °C and 80 °C.\(^{38}\)

Scheme 5.6. One-pot four enzyme synthesis of D-sorbose/D-psicose and L-fructose.

Subsequently, to improve the practicality of the reaction by avoiding the use of DHAP, a rather unstable and expensive substrate as mentioned before,\(^{21}\) we optimized our experiment to provide DHAP in situ from cheaper starting materials (Scheme 5.6). A one-pot four enzyme system was used in which DHAP was generated from the oxidation of L-glycerol 3-phosphate (L-GP) by glycerol phosphate oxidase (GPO)\(^{39-40}\). The by-product of this oxidation, H\(_2\)O\(_2\), being harmful for GPO activity was thus selectively degraded by adding catalase. The DHAP generated in situ was coupled with D-
glyceraldehyde by RhaD to give D-sorbose/D-psicose-1-phosphate. This not only prevents product inhibition of GPO by DHAP and thus promotes the conversion of L-GP, but also suppresses any decomposition of the labile DHAP since it does not accumulate.\textsuperscript{39}

Lastly, the phosphate group was removed under acidic conditions by acid phosphatase (AP) to furnish D-sorbose and D-psicose in moderate yields in a one-pot fashion. (Scheme 5.6) The product ratio (D-sorbose/D-psicose = ~1/1) was determined by HPLC. Under the same conditions, when L-glyceraldehyde was used instead of the D-configuration, L-fructose was produced exclusively with 66\% overall yield (Scheme 5.6).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Starting Material</th>
<th>Acceptor</th>
<th>Product</th>
<th>Yield</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{RhaD}_{E.\text{coli}} )</td>
<td>DHAP</td>
<td>D-glyceraldehyde</td>
<td>D-sorbose</td>
<td>D-psicose</td>
<td>73%</td>
</tr>
<tr>
<td>( \text{RhaD}_{E.\text{coli}} )</td>
<td>L-GP</td>
<td>D-glyceraldehyde</td>
<td>D-sorbose</td>
<td>D-psicose</td>
<td>48%</td>
</tr>
<tr>
<td>( \text{RhaD}_{E.\text{coli}} )</td>
<td>L-GP</td>
<td>L-glyceraldehyde</td>
<td>L-fructose</td>
<td></td>
<td>66%</td>
</tr>
</tbody>
</table>

In summary, we observed that L-rhamnulose-1-phosphate aldoalse (RhaD) loses its stereoselectivity when accepting D-glyceraldehyde for aldol addition (Table 5.2). A possible explanation could be the different enzyme concentration used. While Fessner and co-workers previously reported a diastereomeric ratio of 97:3 (syn:anti) of RhaD,\textsuperscript{41} our observed results indicated that the anti product D-psicose may be thermodynamically favored and accumulate during reaction time. Nonetheless, we utilized this property to
synthesize two rare sugars D-sorbose and D-psicose simultaneously. The enzymatically synthesized diastereomers could then be easily isolated with cation exchange resin (Ca\textsuperscript{2+} form) under elevated temperature.

5.2.2 Synthesis of rare sugars with L-fuculose-1-phosphate aldolase (FucA) from *Thermus thermophilus* HB8

FucA is a class II DHAP-dependent aldolase that catalyzes the reversible cleavage of L-fuculose 1-phosphate to DHAP and L-lactaldehyde, a central step for L-fucose metabolism in bacteria.\textsuperscript{42} Furthermore, FucA from *E.coli* (FucA\textsubscript{E.coli}) is a homotetramer of 215 amino acid residues containing one Zn\textsuperscript{2+} ion per subunit.\textsuperscript{43} Importantly, FucA has demonstrated its value in aldol reactions to selectively afford vicinal diols with the \textit{D-erythro} configuration (absolute configuration: 3\textit{R}, 4\textit{R}), which is advantageous for D-psicose construction.\textsuperscript{41,44}

Consequently, we first employed FucA\textsubscript{E.coli} in the synthesis of D-psicose via the one-pot four enzyme reaction (Scheme 5.7). Unfortunately, after silica gel and gel filtration purification, only a 12\% yield of D-psicose could be obtained (Table 5.3). Compared to the high yield of D-psicose 1-phosphate when DHAP was used directly for aldol addition,\textsuperscript{44} this low yield can most likely be attributed to the possibility that FucA\textsubscript{E.coli} may not be compatible with the one-pot system.
To improve the overall yield for this synthesis, we turned our attention towards enzymes from thermophilic bacteria, which show a great potential for biotechnological applications.\textsuperscript{45} Upon investigation, we noticed that the crystal structure of FucA from \textit{Thermus thermophilus} HB8 (FucA\textsubscript{T.HB8}) was recently reported but the enzyme had yet to be used for synthetic purposes.\textsuperscript{46} Consequently, we expressed and purified FucA\textsubscript{T.HB8} in \textit{E.coli} and then employed it in the one-pot reaction under the same conditions (Scheme 5.7). To our delight, the yield was greatly improved (Table 5.3, 67\% vs. 12\%) with the diastereomer D-sorbose being detected as a minor product when D-glyceraldehyde was used as the acceptor (Figure 5.4 maroon line). The product ratio of D-psicose/D-sorbose was \(~5:1\) as determined by ion exchange HPLC (Figure 5.4 maroon line).

Scheme 5.7. One-pot four enzyme synthesis of D-psicose/D-sorbose with FucA. (see Table 5.3 for yields and product ratio).
Figure 5.4. HPLC (Hydrogen form, sulfonated divinyl benzene-styrene copolymer support and eluted with 5 mM H₂SO₄) profile of final reaction mixture in Scheme 5.7 compared with authentic samples.

**L-Glycerol 3-phosphate (L-GP)** is a reasonably stable starting material and commercially available; however it is still quite expensive, and thus not ideal for large-scale synthesis. Theoretically, L-GP could be preparatively obtained by enzymatic phosphorylation of glycerol (Scheme 5.3). However, this reaction cannot be coupled with the oxidase reaction due to the low oxygen tolerance of glycerol kinase (GK). In contrast, racemic glycerol 3-phosphate (DL-GP) is much cheaper and it has been reported that GPO can exclusively oxidize the L-isomer. Therefore, it was envisioned that a racemic mixture could maximally deliver a 50% yield of DHAP and the remaining D-glycerol 3-phosphate could be isolated during purification. Most significantly, the synthetic cost could be greatly decreased by utilizing racemic DL-glycerol 3-phosphate (DL-GP). Thus, under the same reaction conditions, staring from racemic DL-GP, a total yield of 58% (Table 5.3) of D-psicose/D-sorbose was obtained and the ratio was
determined by HPLC (Figure 5.4 green line, D-psicose/D-sorbose 8.4/1). The only difference for reactions starting with DL-GP is that DL-GP was used in excess and the reaction yield was calculated using the acceptor as the limiting reagent (see experimental section for reaction stoichiometry). After silica gel and gel filtration chromatography, the mixture containing D-psicose/D-sorbose could be easily separated by a cation exchange resin column (Ca\(^{2+}\) form) under elevated temperature (70 °C).

Scheme 5.8. Synthesis of L-tagatose and L-fructose with FucA\(_{T,HB8}\) using L-glyceraldehyde.

Alternatively, when L-glyceraldehyde was used as the acceptor, instead of the D-isomer, two additional rare sugars, L-fructose and L-tagatose, could accordingly be synthesized as allowed by the stereoselectivity of FucA (Scheme 5.8). L-fructose is a well
known nonnutritive sweetener\textsuperscript{48} and an inhibitor of several glycosidases\textsuperscript{49} with its enzymatic synthesis by the aldolase RhaD being greatly exploited by Wong and co-workers.\textsuperscript{13,32} However, L-tagatose, which is a functional sweetener\textsuperscript{50} and a promising starting material for the synthesis of high value-added complex compounds,\textsuperscript{51} has not yet been broadly utilized due to its high cost of production. L-Tagatose can be produced via oxidation of galactitol by \textit{Klebsiella pneumoniae} \textsuperscript{40b} or generated via epimerization of L-sorbose by D-tagatose 3-epimerase from \textit{Psedomonas sp.} ST-24\textsuperscript{37}, however, both methods give low yields.

As shown in Scheme 5.8, we were able to successfully carry out the synthesis of L-fructose and L-tagatose in a one-pot fashion using DL-GP as starting material and L-glyceraldehyde as the acceptor. It is interesting to note that the acid phosphatase (AP) we previously used could not dephosphorylate L-tagatose 1-phosphate completely to give the desired sugar products. However, upon searching relevant literature we found that YqaB phosphatase from \textit{E.coli}, belonging to the haloacid dehalogenase (HAD) superfamily, shows a remarkably broad substrate range, among which the dephosphorylation activity toward D-fructo-1-phosphate is the highest.\textsuperscript{53} We thus expressed and purified YqaB phosphatase from \textit{E.coli} and the dephosphorylation reaction went quite smoothly under neutral conditions. L-Fructose and L-tagatose were afforded in moderate yield (see Table 5.3, total yield 47\%) and the ratio was determined by HPLC (Figure 5.5, L-fructose/L-tagatose 1.2/1). After silica gel and gel filtration chromatography, the mixture containing L-fructose/L-tagatose could be easily separated by cation exchange resin column (Ca\textsuperscript{2+} form) under elevated temperature (70 °C).
Figure 5.5. HPLC (Calcium form, sulfonated divinyl benzene-styrene copolymer support and eluted with H₂O) profile of final reaction mixture in Scheme 5.8 compared with authentic samples.

Table 5.3. Summary of FucA catalyzed reactions using different starting materials and acceptors.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Starting Material</th>
<th>Acceptor</th>
<th>Product(s)</th>
<th>Yield</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>FucA_E.coli</td>
<td>L-GP</td>
<td>D-glyceraldehyde</td>
<td>D-psicose</td>
<td>12%</td>
<td>1/0</td>
</tr>
<tr>
<td>FucA_T.HB8</td>
<td>L-GP</td>
<td>D-glyceraldehyde</td>
<td>D-psicose</td>
<td>D-sorbose</td>
<td>67%</td>
</tr>
<tr>
<td>FucA_T.HB8</td>
<td>DL-GP</td>
<td>D-glyceraldehyde</td>
<td>D-psicose</td>
<td>D-sorbose</td>
<td>58%</td>
</tr>
<tr>
<td>FucA_T.HB8</td>
<td>DL-GP</td>
<td>L-glyceraldehyde</td>
<td>L-fructose</td>
<td>L-tagatose</td>
<td>47%</td>
</tr>
</tbody>
</table>

As shown in Table 5.3, one thing worth noting is that FucA_T.HB8 seems to lose its stereoselectivity when accepting L-glyceraldehyde—similar to our previous discovery that L-rhamnulose-1-phosphate aldoase (RhaD) produces a single product (L-fructose) when using L-glyceraldehyde while losing its stereoselectivity when D-glyceraldehyde is the
acceptor. Nonetheless, this property of aldolases could be well utilized for the synthesis of various types of rare sugars.

In summary, we developed a one-pot four-enzyme approach for the synthesis of the rare sugars D-psicose, D-sorbose, L-tagatose, and L-fructose with aldolase FucA. All synthesized rare sugars were characterized by $^1$H NMR and compared with commercially authentic samples (see appendix for NMR spectra). To the best of our knowledge, this is the first use of FucA from a thermophilic source (*Thermus thermophilus* HB8), which proved to be more efficient than its *E.coli* counterpart. Importantly, the one-pot four-enzyme approach does not require the use of expensive DHAP and most significantly, the inexpensive starting material DL-GP was used to greatly reduce the synthetic cost. We believe this approach could ultimately contribute to the synthesis of other rare sugars and their derivatives as well.

### 5.2.3 One-pot four enzyme synthesis of carbohydrates with fructose 1,6-biphosphate aldolase (FruA)

The FruA is the most extensively studied DHAP-dependent aldolase. *In vivo*, FruA catalyzes the pivotal step of glycolysis, which is the reversible degradation of D-fructose 1,6-biphosphate to D-glyceraldehyde 3-phosphate and DHAP (Scheme 5.10a). Most importantly, the equilibrium constant for this reaction strongly favors the synthesis direction. Moreover, this enzyme has been isolated from various prokaryotic and eukaryotic sources, both as class I and class II forms.

The class I FruA isolated from rabbit muscle (RAMA) is the most widely used aldolase for preparative synthesis, due to its commercial availability and high activity.
RAMA accepts a number of unnatural aldehydes as substrates, but its operative stability in solution is low.\textsuperscript{54-55} However, FruA from bacterial sources have been cloned for overexpression which offers unusual stability for synthetic applications. For example, it was found that the class I aldolase of the \textit{Staphylococcus carnosus} is unusually temperature and pH stable.\textsuperscript{54,56} Furthermore, the \textit{S. carnosus} FruA has a similarly relaxed acceptor specificity and is much more stable than RAMA. \textit{S. carnosus} FruA is thus chosen for large-scale synthesis of various of structures of interest, such as glycosidase inhibitors and C\textsubscript{8} and C\textsubscript{9} sugars.\textsuperscript{19,57} In this section, we describe the application of RAMA and FruA from \textit{S. carnosus} (FruA\textsubscript{S.carnosus}) in the synthesis of carbohydrates using our one-pot four enzyme system.

Scheme 5.9. One-pot four enzyme synthesis of carbohydrates with FruA.

Firstly, D-fructose and L-sorbose were synthesized with FruA from \textit{S. carnosus} using our well developed one-pot system. Unlike RhaD and FucA which generated a mixture of diastereomers, FruA demonstrated excellent stereoselectivity when using D/L-glyceraldehyde as acceptor and a single product was isolated for each reaction (D-
fructose and L-sorbose respectively). (Scheme 5.9 and Table 5.4) Interestingly, FruA<sub>S. carnosus</sub> may have a discrimination for D/L-glyceraldehydes since a much higher yield was obtained for D-glyceraldehyde compared with the L-isomer.

Table 5.4. One-pot four enzyme synthesis of carbohydrates with FruA

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Acceptor</th>
<th>Product</th>
<th>Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>FruA&lt;sub&gt;S. carnosus&lt;/sub&gt;</td>
<td>D-glyceraldehyde</td>
<td><img src="image" alt="D-fructose" /></td>
<td>60%</td>
</tr>
<tr>
<td>FruA&lt;sub&gt;S. carnosus&lt;/sub&gt;</td>
<td>L-glyceraldehyde</td>
<td><img src="image" alt="L-sorbose" /></td>
<td>28%</td>
</tr>
<tr>
<td>RAMA</td>
<td>Acetaldehyde</td>
<td><img src="image" alt="Acetaldehyde" /></td>
<td>29%</td>
</tr>
<tr>
<td>RAMA</td>
<td>Propionaldehyde</td>
<td><img src="image" alt="Propionaldehyde" /></td>
<td>34%</td>
</tr>
<tr>
<td>RAMA</td>
<td>Butyraldehyde</td>
<td><img src="image" alt="Butyraldehyde" /></td>
<td>38%</td>
</tr>
</tbody>
</table>

Because FruA (especially FruA from rabbit muscle, RAMA) has been reported as an extremely flexible aldolase which accepts various aldehyde structures, we integrated this property into our one-pot system and successfully synthesized some...
uncommon carbohydrates using simple aliphatic aldehydes as acceptors (Table 5.4). The configuration of vicinal diols produced from the aldol additional of DHAP with aliphatic aldehydes follows that of the glyceraldehydes: only one diastereomer was isolated for each reaction and the products were verified by $^1$H NMR.$^{58-59}$

5.3 Let *E.coli* be the best ‘chemist’

In the previous sections, we demonstrated the synthesis of rare sugars *in vitro* with different DHAP dependent aldolases. The idea of using one-pot multi enzyme system is to generate the key intermediate DHAP, which is rather expensive and unstable, from inexpensive sources to reduce the synthetic cost as well as broaden the scope of aldolase reactions. For similar purpose, here we are introducing a novel platform with the potential of large scale production of rare sugars in which DHAP is generated and coupled with aldehydes *in vivo* and *in situ* within engineered bacterial machinery (*i.e.* *E.coli*). In fact, similar metabolic-pathway engineering approach has been reported that succeeded in producing higher titers (1 g/L in an engineered *E.coli* strain) of taxadiene—the first committed Taxol intermediate.$^{60}$
Our strategy is to produce one of the substrates DHAP through *E. coli* metabolic pathways from cheap green carbon source, glucose or glycerol for instance. As in glycolysis (Scheme 5.10a), glucose is metabolized into fructose-1,6-bisphosphate via three enzymatic steps, then the later is split by fructose-1,6-bisphosphate aldolase to release two three triose sugars, DHAP and D-glyceraldehyde-3-phosphate. An alternative route to DHAP is the phosphorylation of glycerol to glycerol phosphate and the subsequent oxidation to DHAP under aerobic conditions (Scheme 5.10b). The other substrate (glyceraldehyde) is provided by adding into the medium continuously and glyceraldehyde could easily enter the cells via glycerol facilitator. The resulting DHAP is coupled with glyceraldehyde *in situ* by aldolase to give sugar-1-phosphate which is further dephosphorylated by phosphatase *in vivo* to furnish the final sugar compound (Scheme 5.11). As shown in Scheme 5.11, D-sorbose and D-psicose were produced by fermentation as a proof-of-concept experiment.
As mentioned in Section 5.2.1, L-rhamnulose-1-phosphate aldoalse (RhaD) from *E.coli* is a DHAP-dependent class II aldolase that involves in the third step of the L-rhamnose degradation pathway.\(^\text{30}\) As no RhaD expression was observed (or at very low level) under our culture condition, RhaD overexpression was performed with pBluescript SK- plasmid. It was assumed that phosphorylated intermediates (*e.g.* aldol adduct sugar-1-phosphate) from cellular metabolism were efficiently retained by the plasma membrane and accumulation of such intracellular metabolites at high levels is toxic or, at least, a burden to the cells.\(^\text{65}\) Moreover, sugar molecule without a highly polar phosphate group will pass through the cell membrane into the medium more easily for product harvest and purification. However, acid phosphatase (AP) which operates very well *in vitro* at acidic condition is not an appropriate choice for our *E.coli* platform. YqaB, a phosphatase from *E.coli* which was successfully used to dephosphorylate L-fructose/L-tagatose-1-phosphate
(Section 5.2.2) can function under neutral or physiological conditions such as cell culture. Moreover, we already proved in our *in vitro* experiment that YqaB phosphatase also had good activity towards D-sorbose-1-phosphate and D-psicose-1-phosphate. Thus co-expression of RhaD and YqaB with pBluescript SK- plasmid was carried out in *E.coli* to complete the reaction sequence *in vivo* (Figure 5.6).

![Figure 5.6. Co-expression of RhaD and YqaB with pBluescript SK- plasmid.](image)

During the fermentation process, the initial concentration of D-glyceraldehyde (0.45 g/L) was almost consumed after 3 h as detected by HPLC. Thus, D-glyceraldehyde was fed continually once exhausted (Figure 5.7). The level of D-sorbose/D-psicose in the medium increased as the fermentation proceeded (Figure 5.7). Since D-sorbose and D-psicose are rare sugars, they were not metabolized or utilized by *E.coli* as no such enzyme exists.
Figure 5.7. HPLC profile of the fermentation process.

The concentration of the starting materials/products in the medium was monitored by HPLC as the fermentation progressed (Figure 5.7). At the end of our fermentation, cells were harvested, sonicated, and the supernatant was analyzed by TLC. The results showed that D-sorbose/D-psicose produced only existed in the medium while not in the cytoplasm, which enable an easier purification process (see experimental section). Purified product was confirmed by HPLC retention times, NMR and MS compared with authentic samples. The initial results were summarized in Table 5.5.
Table 5.5. Summary of initial results for \textit{in vivo} production of rare sugars with \textit{E.coli}_{\text{rhaD, yqaB}}.

<table>
<thead>
<tr>
<th>Biomass</th>
<th>Aldehyde</th>
<th>Fermentation scale</th>
<th>Product(s)</th>
<th>Ratio</th>
<th>Isolated yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycerol</td>
<td>D-glyceraldehyde</td>
<td>50 mL</td>
<td>D-sorbose/D-psicose</td>
<td>1.3/1</td>
<td>49 mg</td>
</tr>
<tr>
<td>Glycerol</td>
<td>D-glyceraldehyde</td>
<td>200 mL</td>
<td>D-sorbose/D-psicose</td>
<td>1.3/1</td>
<td>259 mg</td>
</tr>
<tr>
<td>Glycerol</td>
<td>DL-glyceraldehyde</td>
<td>200 mL</td>
<td>D-sorbose/D-psicose /L-fructose</td>
<td>~1/1/1</td>
<td>264 mg</td>
</tr>
<tr>
<td>Glucose</td>
<td>D-glyceraldehyde</td>
<td>1 L</td>
<td>D-sorbose/D-psicose</td>
<td>1.2/1</td>
<td>600 mg</td>
</tr>
</tbody>
</table>

In summary, we have successfully developed an \textit{in vivo} system in which DHAP was generated \textit{in vivo} from cheap green materials through bacterial metabolic pathways. Since the aldolase and phosphatase were co-expressed in the same bacteria, coupling of the resulted DHAP with aldehyde and further dephosphorylation reaction could furnish the final product which can be isolated from the fermentation medium. Construction of such systems with different aldolases (\textit{e.g.} FucA, FruA) for the production of diversified structures is currently underway.

5.4 Experimental Section

\textit{HPLC separations on Aminex resins}

Aminex carbohydrate columns separate compounds using a combination of size exclusion and ligand exchange mechanisms. For monosaccharide separations, ligand exchange is the primary mechanism which involves the binding of hydroxyl groups of the
sugars with the fixed-counterion of the resin. Ligand exchange is affected by the nature of the counterion (Ca\(^{2+}\), H\(^{+}\), etc.) and by the spatial orientation of the carbohydrate’s hydroxyl groups.\(^{66-67}\)

Aminex HPX-87H column (hydrogen form, sulfonated divinyl benzene-styrene copolymer support and eluted with 5 mM H\(_2\)SO\(_4\)) for quantification of D-sorbose and D-psicose

Aminex HPX-87C column (Calcium form, sulfonated divinyl benzene-styrene copolymer support and eluted with H\(_2\)O) for quantification of L-fructose and L-tagatose

Figure 5.8. HPLC standard curve for D-sorbose, D-psicose, L-fructose, L-tagatose quantification (peak area/concentration)
Typical procedure for the in vitro aldolase reaction starting with DHAP

Synthesis of D-sorbose/D-psicose with RhaD and AP

DHAP (0.2 M, 874 μL) solution and D-glyceraldehyde (0.5 M, 435 μL) were combined and the resulting solution was brought to a total volume of 2335 μL by adding ddH$_2$O. RhaD (18.4 mg/mL, 65 μL in pH 7.5 Tris buffer, final concentration 0.5 mg/mL) was added and the reaction mixture was shaken at rt. for 16-22 h and the reaction was monitored by TLC (developed by nBuOH/AcOH/H$_2$O 2/1/1 (v/v/v) and stained with anisaldehyde sugar stain). The pH was adjusted to pH ~5 with 6 N HCl and 2 μL acid phosphatase (3 U) was added and the mixture was shaken at 37 °C for another 24 h. The ratio of D-sorbose and D-psicose in the final mixture was determined by HPLC. (Aminex HPX-87H, 300×7.8 mm, column temperature 60 °C, 5 mM sulfuric acid as mobile phase and detected with Refractive Index Detector). After cooling to rt., the pH was adjusted to 7 with 1 N NaOH and the mixture was diluted with methanol. The solution was filtered through Celite and washed with methanol. The filtrate was concentrated under reduced pressure and the residue was purified by silica gel column chromatography (EtOAc/iPrOH/H$_2$O 9/3/1 (v/v/v)) to afford a pale yellow syrup which was further purified by Bio gel P-2 column to afford D-sorbose and D-psicose mixture (23 mg, 73% totally). The mixture containing D-sorbose and D-psicose only could be isolated as described below. NMR data were consistent with authentic samples. HRMS (ESI, D-sorbose) calcd for C$_6$H$_{12}$O$_6$Na (M+Na)$^+$ 203.0526, found 203.0530 m/z. HRMS (ESI, D-psicose) calcd for C$_6$H$_{12}$O$_6$Na (M+Na)$^+$ 203.0526, found 203.0524 m/z.
Typical procedure for the one-pot four enzyme synthesis of rare sugars starting with L-glycerol 3-phosphate

One-pot four enzyme synthesis of D-sorbose and D-psicose

To a solution of L-glycerol 3-phosphate bis(cyclohexylammonium) salt (370 mg, 1.0 mmol) in 7.56 mL ddH₂O was added D-glyceraldehyde (2.44 mL, 0.5 M, 1.22 mmol) at pH 7, glycerol phosphate oxidase (70 U, 2 mg), catalase (1000 U, 1.18 μL) and RhaD (final concentration 0.5 mg/mL). The mixture was shaken at rt. for 22 h and the reaction was monitored by TLC (developed by nBuOH/ACOH/H₂O 2/1/1 (v/v/v) and stained with anisaldehyde sugar stain). The pH was then adjusted to pH ~5 with 6 N HCl and 11 μL of acid phosphatase (18 U) added and the mixture was shaken at 37 °C for another 24 h. After cooling to rt., the pH was adjusted to 7 with 1 N NaOH and the mixture was diluted with methanol. The solution was filtered through Celite and washed with methanol. The filtrate was concentrated under reduced pressure and the residue was purified by silica gel column chromatography (EtOAc/iPrOH/H₂O 9/3/1 (v/v/v)) to afford a pale yellow syrup which was further purified by Bio gel P-2 column to afford D-sorbose and D-psicose mixture (87 mg, 48% totally). The mixture containing D-sorbose and D-psicose only could be isolated as described below. Synthesis of L-fructose was performed with the same procedure using L-glyceraldehyde (66% yield). HRMS (ESI, L-fructose) calcd for C₆H₁₂O₆Na (M+Na)⁺ 203.0526, found 203.0522 m/z.

Typical procedure for the one-pot four enzyme synthesis of rare sugars starting with DL-glycerol 3-phosphate

One-pot four enzyme synthesis of D-psicose/D-sorbose with FucA₅THB
To a solution of DL-glycerol 3-phosphate magnesium salt (548.78 mg, 2.4 mmol) in 6.86 mL ddH$_2$O was added D-glyceraldehyde (2 mL, 0.5 M, 1.0 mmol) at pH 7.0, glycerol phosphate oxidase (70 U, 2 mg), catalase (1000 U, 1.18 μL) and FucA$_{T,HB8}$ (final concentration 0.5 mg/mL). ddH$_2$O was added to bring the total volume to 10 mL if necessary. The reaction condition and isolation procedure were similar to that of using the substrate L-glycerol 3-phosphate as described above. After purification by Bio gel P-2 column, a mixture of D-psicose and D-sorbose was obtained (104.2 mg, 58% totally, D-psicose/D-sorbose 8.4:1).

Typical procedure for the one-pot four enzyme synthesis of rare sugars with aldolase and YqaB phosphatase

One-pot four enzyme synthesis of L-tagatose and L-fructose

To a solution of DL-glycerol 3-phosphate magnesium salt (548.78 mg, 2.4 mmol) in 4.7 mL ddH$_2$O was added L-glyceraldehyde (3.3 mL, 0.303 M, 1.0 mmol) at pH 7.0, glycerol phosphate oxidase (70 U, 2 mg), catalase (1000 U, 1.18 μL) and FucA$_{T,HB8}$ (final concentration 0.5 mg/mL). ddH$_2$O was added to bring the total volume to 10 mL if necessary. The mixture was shaken at rt. for 22 h and the reaction was monitored by TLC (developed by nBuOH/AcOH/H$_2$O 2/1/1 (v/v/v) and stained with anisaldehyde sugar stain). Then YqaB phosphatase (final concentration 0.25 mg/mL) and MgCl$_2$ (final concentration 2.5 mM) was added and the mixture was shaken at 37 °C for 12 h. The purification was performed as described above. After purification by Bio gel P-2 column, a mixture of L-tagatose and L-fructose was obtained (84 mg, 47% totally).
Typical procedure for isolation of diastereomeric rare sugars using cation exchange resin (Ca$^{2+}$ form)

D-Sorbose and D-psicose mixture (150 mg from one-pot reactions) was dissolved in 3 mL ddH$_2$O and applied to a cation exchange resin (Ca$^{2+}$ form, 26×100 mm) which was pre-heated to 65 °C with a thermostatic jacket. The column was eluted with ddH$_2$O (flow rate ~1.5 mL/min) and the whole isolation process was performed at 65-70 °C. Fractions were collected with an automatic fraction collector and identified by TLC and HPLC. Fractions containing pure D-sorbose or D-psicose were pooled and lyophilized to give D-sorbose (62 mg) and D-psicose (50 mg). A mixture of both sugars (~ 30 mg) was also collected which can be further separated if necessary.

Synthesis and purification of D-sorbose/D-psicose in *E.coli*

Synthesis of D-sorbose/D-psicose in *E.coli* was carried out in 500 mL flasks containing 200 mL LB medium with ~5 g/L glycerol and 100 µg/mL ampicillin. 4 mL overnight culture of strain DH5α (psk-rhaD-yqaB) was inoculated into the flask and cultivated at 37 °C, 225 rpm/min. After 3 h, IPTG was added at a final concentration of 1 mM to induce the expression of plasmid psk-rhaD-yqaB for 2 h, then D-glyceraldehyde (0.5 M, filter-sterilized stock solution) was added at a final concentration of 0.45 g/L. Samples were taken at regular intervals to monitor the consumption of D-glyceraldehyde and production of D-sorbose/D-psicose by HPLC. D-glyceraldehyde was added to the medium continuously when consumed as indicated by HPLC. The pH was kept constant at 7.0 using sodium hydroxide during the fermentation process. Upon completion, the fermentation broth was centrifuged at 4,000 × g for 20 min at 4 °C. The supernatant was
collected, decolored with active carbon (1.5% m/V) at 55 °C for 30 min, and then concentrated by rotary evaporation under reduced pressure at 40-45 °C. The residue was purified by silica gel column chromatography (EtOAc/iPrOH/H2O 9/3/1 (v/v/v)) to afford the crude D-sorbose/D-psicose mixture as a pale yellow syrup. Further purification with P-2 gel filtration was performed to afford D-sorbose/D-psicose mixture as a white solid which could be further separated by cation exchange chromatography as described above.

References


(38) Huang, W. M., W.; Jiang, B. *Food and Fermentation Industries* **2008**, *34*.


BIBLIOGRAPHY


DeAngelis, P. L. Glycobiology 2002, 12, 9R-16R.


Huang, W. M., W.; Jiang, B. Food and Fermentation Industries 2008, 34.


APPENDIX: $^1$H NMR, $^{13}$C NMR and $^{31}$P NMR spectra
chemically synthesized
2-7
Enzymatically synthesized
2-62
2-85
L-fructose
from FucA reaction
L-tagatose standard

L-tagatose from FucA reaction
\textbf{d-fructose Standard}
5-Deoxy-\(\delta\)-threo-2-pentulose

5,6-Dideoxy-\(\delta\)-threo-2-hexulose
5,6,7-Trideoxy-\(\alpha\)-threo-2-heptulose
(5-deoxy-5-ethyl-\(\alpha\)-xylulose)