

**MECHANISMS AND ENGINEERING  
OF GLYCOSYLTRANSFERASES**

by

**LUKE LEE LAIRSON**

B.Sc., University of Guelph, 2002

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Chemistry)

**THE UNIVERSITY OF BRITISH COLUMBIA**

October 2007

© Luke Lee Lairson, 2007

## ABSTRACT

In order to gain insight into the natural evolution of enzyme mechanism and to increase the utility of a class of sugar modifying enzymes known as glycosyltransferases, several representative enzymes were subjected to mechanistic and engineering studies.

The chemical and kinetic mechanisms of the GT-A fold inverting sialyltransferase Cst II were investigated by detailed kinetic analysis, protein X-ray crystallography and mutagenesis. This enzyme catalyzes the general  $S_N2$ -like direct displacement mechanism used by all inverting glycosyltransferases. However, the chemical strategies utilized to facilitate reaction are more akin to those of the GT-B fold enzymes, indicating a convergence in mechanism between these two clans of enzymes.

By analogy to retaining glycosidases, retaining glycosyltransferases had been thought to use a double displacement mechanism involving an enzymatic nucleophile. However, a comparison of the X-ray crystal structures of multiple retaining glycosyltransferases indicates a complete lack of conserved structural architecture in the region that would be occupied by this critical catalytic residue. This lack of conserved architecture, precedence for cationic enzymatic mechanisms, and the inherent differences in reactivities of glycosyltransferase and glycosidase substrates all support a notion that the majority of retaining glycosyltransferases utilize a mechanism involving the formation of a short-lived ion pair intermediate species. However, a protein engineering approach was used to explore the possibility of nucleophilic catalysis in the retaining galactosyltransferase LgtC. The results of this work led to the first direct observation of a catalytically relevant covalent glycosyl-enzyme intermediate for a retaining glycosyltransferase.

It was demonstrated that catalytically active LgtC could be displayed on the surface of M13 bacteriophage as a pIII fusion protein. Further, LgtC phage display was successfully performed in the context of a water-in-oil emulsification procedure that may have significant utility in the development of directed evolution screening approaches. A substrate engineering strategy was developed which allowed the substrate specificity of wild type LgtC to be broadened to allow exclusive formation of  $\alpha$ -1,2,  $\alpha$ -1,3 or  $\alpha$ -1,4 linkages at synthetically useful rates with various alternative acceptor substrates. Finally, it was demonstrated that Cst II and LgtC will utilize alternative donor substrates in the presence of their respective natural nucleotide products.

# TABLE OF CONTENTS

<b>Abstract</b> .....	<b>ii</b>
<b>Table of Contents</b> .....	<b>iv</b>
<b>List of Tables</b> .....	<b>xii</b>
<b>List of Figures</b> .....	<b>xiii</b>
<b>List of Schemes</b> .....	<b>xxi</b>
<b>List of Abbreviations</b> .....	<b>xxii</b>
<b>Acknowledgements</b> .....	<b>xxv</b>
<b>Co-Authorship Statement</b> .....	<b>xxvii</b>
<b>Chapter 1. Introduction</b> .....	<b>1</b>
1.1. Carbohydrates in Biology .....	2
1.2. Biological Glycosyl Group Transfer Reactions .....	4
1.3. Glycosyltransferases .....	6
1.3.1. Glycosyltransferase Activities .....	6
1.3.2. Glycosyltransferase Folds .....	7
1.3.3. Glycosyltransferase Classification .....	14
1.3.4. Inverting Glycosyltransferases .....	17
1.3.4.1. Mechanism of Inverting Glycosidases .....	17
1.3.4.2. Mechanism of Inverting Glycosyltransferases .....	18
1.3.4.3. Inverting GT-A Glycosyltransferases (Clan I) .....	18
1.3.4.4. Inverting GT-B Glycosyltransferases (Clan II) .....	22
1.3.5. Retaining Glycosyltransferases .....	27
1.3.5.1. Mechanism of Retaining Glycosidases .....	27

1.3.5.2. Mechanism of Retaining Glycosyltransferases .....	30
1.3.5.3. Challenges in Studying the Mechanisms of Retaining Glycosyltransferases .....	32
1.3.5.4. Retaining GT-A Glycosyltransferases (Clan III) .....	33
1.3.5.5. Retaining GT-B Glycosyltransferases (Clan IV) .....	43
1.3.5.6. An Alternative “S <sub>N</sub> i-Like” Mechanism .....	50
1.4. Aims of the Thesis .....	53
<b>Chapter 2. Structural and Mechanistic Investigations of an Inverting Glycosyltransferase .....</b>	<b>56</b>
2.1. Summary .....	57
2.2. Background .....	57
2.3. Determining the Kinetic Mechanism of the Bifunctional Inverting $\alpha$ -(2,3/8) Sialyltransferase Cst IIOH <sub>4384</sub> .....	61
2.3.1. Preliminary Kinetic Analysis .....	61
2.3.2. Potential Kinetic Mechanisms .....	63
2.3.3. Distinguishing Between a Sequential or Ping Pong Mechanism .....	65
2.3.4. Product Inhibition Studies with Cst II .....	67
2.3.5. Synthesis and Inhibition Studies of a Dead-End Donor Substrate Analogue .....	71
2.3.6. Conclusions on the Kinetic Mechanism of Cst II .....	73
2.3.7. Calculating Kinetic Parameters for Cst II .....	76
2.4. Three-Dimensional X-Ray Crystallographic Analysis of Cst II in Complex with a Stable Donor Substrate Analogue .....	78
2.5. Investigating the Role of Putative Catalytic Residues .....	81
2.5.1. Tyr156 and Tyr162 Facilitate CMP Departure .....	82

2.5.2. His188 Acts as the Catalytic Base .....	84
2.6. Concluding Remarks .....	90
<b>Chapter 3. Structural and Mechanistic Investigations with Retaining Glycosyltransferases .....</b>	<b>93</b>
3.1. Summary .....	94
3.2. Investigating the Catalytic Mechanism of the Retaining $\alpha$ -(1,4) Galactosyltransferase LgtC .....	94
3.2.1. Background .....	94
3.2.2. Engineering the Mechanism of LgtC .....	98
3.2.3. Kinetic Analysis of LgtC Q189E .....	99
3.2.4. Intermediate Trapping by ESI-MS .....	100
3.2.5. Identification of the Site of Labelling .....	104
3.2.6. Analysis of the D190N Variant of LgtC .....	107
3.2.7. Crystallographic Analysis of LgtC Q189E .....	108
3.2.8. Attempts to Accumulate a Covalent Intermediate using Alternative Active Site Modifications .....	109
3.2.9. Interpretation of the LgtC Q189E Labelling Results .....	111
3.3. Investigating the Retaining $\alpha$ -(1,2) Mannosyltransferase Kre2 .....	115
3.3.1. Background .....	115
3.3.2. Synthesis of a Non-Reactive Donor Substrate Analogue .....	116
3.3.3. Kinetic Analysis of Kre2 .....	117
3.3.4. Attempts to Observe a Covalent Glycosyl-Enzyme Intermediate with Kre2 .....	117
3.3.5. X-Ray Crystal Structure of Kre2 with Bound GDP 2F Mannose .....	118

3.4. ESI-MS Analysis of the Retaining $\alpha$ -(1,3) Galactosyltransferase $\alpha$ 3GalT .....	122
3.4.1. Background .....	122
3.4.2. Attempts to Observe a Covalent Glycosyl-Enzyme Intermediate with $\alpha$ 3GalT .....	124
3.5. Reasons Why Retaining Glycosyltransferases May Have Evolved a Mechanism Not Involving Nucleophilic Catalysis .....	126
3.6. Concluding Remarks .....	135
<b>Chapter 4. Glycosyltransferase Protein Engineering .....</b>	<b>142</b>
4.1. Summary .....	143
4.2. Engineering the Nucleotide Specificity of LgtC .....	144
4.2.1. Background .....	144
4.2.2. Mutation of a Conserved Active Site Side Chain Alters the Base Specificity of GTP Binding Proteins .....	147
4.2.3. Identification and Modification of a Base Specificity Conferring Site Within the LgtC Active Site .....	151
4.2.4. Synthesis of CDP Gal .....	152
4.2.5. Testing the Nucleotide Sugar Specificity of Wild Type and N10D LgtC .....	153
4.2.6. Concluding Remarks .....	154
4.3. Compartmentalized Phage Display of LgtC .....	156
4.3.1. Background .....	156
4.3.2. Phage Display of LgtC as an M13 pIII Fusion Protein .....	163
4.3.3. Compartmentalized LgtC Phage Display .....	168
4.3.4. A High Throughput Screen for LgtC Directed Evolution Based on Compartmentalized Phage Display .....	170

4.3.5. Synthesis of a Biotinylated Lactose Conjugate Acceptor Substrate .....	175
4.3.6. Generating a Library of LgtC Variants .....	176
4.3.7. Immobilization of the Carbohydrate Binding Domain of Shiga-Like Toxin SLT-1B .....	177
4.3.8. Multivalent Phage Display of Streptavidin as a pVIII fusion protein ...	179
4.3.9. Simultaneous M13 Phage Display of LgtC and Streptavidin .....	181
4.3.10. Conclusions and Future Directions .....	183
<b>Chapter 5. Glycosyltransferase Substrate Engineering .....</b>	<b>187</b>
5.1. Summary .....	188
5.2. Engineering Alternative Acceptor Substrates for Glycosyltransferases .....	188
5.2.1. Background .....	188
5.2.2. Exploring the Acceptor Substrate Specificity of LgtC .....	190
5.2.3. Determining the Ability of LgtC to Transfer to Equatorial Hydroxyl Substituents .....	192
5.2.4. Exploring the Synthetic Utility of LgtC Catalyzed Galactosyl Transfer to Alternative Monosaccharide Acceptor Substrates .....	193
5.2.5. A Substrate Engineering Strategy the Uses Aromatic Substituents as Active Site Anchors .....	194
5.2.6. Using Substrate Engineering to Control Regio-Selective $\alpha$ 3GalT Catalyzed Galactosyl Transfer to a Monosaccharide Acceptor Substrate .....	196
5.2.7. Using Substrate Engineering to Facilitate Cst II Catalyzed Sialyl Transfer to a Monosaccharide Acceptor .....	198
5.2.8. The Nature of the Substituent on an Alternative Acceptor Substrate can Affect the Regio-Selectivity of LgtC .....	198
5.2.9. Controlling the Regio-Selectivity of LgtC-Catalyzed Galactosyl Transfer to Mannose .....	202

5.2.10. Controlling the Regio-Selectivity of LgtC-Catalyzed Galactosyl Transfer to Xylose .....	205
5.2.11. Concluding Remarks .....	206
5.3. Engineering Alternative Donor Substrates for Glycosyltransferases .....	209
5.3.1. Background .....	209
5.3.2. An Activated Alternative Donor Substrate for Cst II that Mimics the Product Binding Mode .....	209
5.3.3. An Activated Alternative Donor Substrate for LgtC .....	213
5.3.3.1. Testing $\alpha$ -Linked Aryl Galactosides .....	213
5.3.3.2. Testing $\beta$ -Linked Aryl Galactosides .....	214
5.3.3.3. Mechanistic Implications for LgtC .....	216
5.3.4. Concluding Remarks .....	217
<b>Chapter 6. Summary and Future Directions .....</b>	<b>219</b>
<b>Chapter 7. Methods .....</b>	<b>226</b>
7.1. General .....	227
7.2. Chapter 2 Experimental .....	227
7.2.1. Acknowledgements and Generous Gifts .....	227
7.2.2. Expression and Purification of Cst II .....	228
7.2.3. A UV Spectroscopy Based Enzyme Coupled Continuous Assay for Cst II Kinetic Characterization .....	228
7.2.4. A Capillary Electrophoresis Based Stopped Assay for Cst II Kinetic Characterization .....	229
7.2.5. Synthesis of CMP 3F NeuAc .....	230
7.2.6. Generation of Site Directed Mutants .....	231

7.2.7. Chemical Rescue of the H188A Mutant of Cst II .....	232
7.3. Chapter 3 Experimental .....	232
7.3.1. Acknowledgements and Generous Gifts .....	232
7.3.2. Expression and Purification of LgtC-25 .....	233
7.3.3. Expression of LgtC-25 as a C-terminal 6-Histidine Fusion .....	234
7.3.4. A UV Spectroscopy Based Enzyme Coupled Continuous Assay for LgtC and Kre2 Kinetic Characterization .....	235
7.3.5. ESI-MS Analysis of LgtC-25 Q189E .....	236
6.3.5.1. ESI-MS Conditions .....	236
6.3.5.2. Labelling of LgtC-25 Q189E .....	237
6.3.5.3. Turn-over of Covalently Labelled LgtC-25 Q189E .....	238
7.3.6. Generation of the I143D, N153D, Q185D, and D190N Mutants of LgtC-25 .....	238
7.3.7. Synthesis of GDP 2F Gal .....	239
7.3.8. ESI-MS Analysis of Kre2 .....	240
7.3.9. Expression and Purification of $\alpha$ 3GalT as a Male Fusion Protein .....	241
7.3.10. ESI-MS Analysis of $\alpha$ 3GalT .....	242
7.4. Chapter 4 Experimental .....	242
7.4.1. Acknowledgements and Generous Gifts .....	242
7.4.2. Synthesis of CDP Gal .....	243
7.4.3. Generation of the N10D Mutant of LgtC-25 .....	244
7.4.4. Kinetic Analysis of CDP Gal as a Substrate for LgtC-25 and LgtC-25 N10D .....	244
7.4.5. Subcloning LgtC-25 for Expression as an M13 pIII Fusion Protein ....	245
7.4.6. VCSM13 Helper Phage Production .....	246

7.4.7. Small Scale PEG Precipitated LgtC-25 Displaying Phage Preparation .....	247
7.4.8. Quantification of Phage by UV Spectroscopy .....	248
7.4.9. Western Blot Analysis of pLL01 and pSJF6 Derived Phage .....	248
7.4.10. Compartmentalized Phage Display Procedure .....	249
7.4.11. Testing Infectivity of Emulsified LgtC-25 displaying Phage .....	249
7.4.12. Synthesis of a Biotin Lactose Conjugate .....	250
7.4.13. Generation of a Library of LgtC-25 Variants by Error Prone PCR .....	251
7.4.14. Subcloning Verotoxin for Expression as a C-Terminal 6-Histidine Fusion Protein .....	252
7.4.15. Subcloning of Streptavidin-pVIII Fusion Protein .....	254
7.4.16. ELISA Analysis of pSAV and pLL34 Derived Phage .....	255
<b>7.5. Chapter 5. Experimental .....</b>	<b>256</b>
7.5.1. Acknowledgements and Generous Gifts .....	256
7.5.2. LgtC-Catalyzed Galactosylation of Modified Alternative Acceptor Substrates .....	256
7.5.3. NMR Analysis of LgtC Products .....	257
7.5.4. $\alpha$ 3GalT-Catalyzed Galactosylation of Alternative Acceptor Substrates .....	261
7.5.5. Cst II-Catalyzed Sialylation of Alternative Acceptor Substrates .....	262
7.5.6. Testing Alternative Cst II Donor Substrates .....	263
7.5.7. Testing Alternative LgtC Donor Substrates .....	264
<b>References .....</b>	<b>266</b>
<b>Appendix A – IUPAC Recommended Nomenclature for Reaction Mechanisms ..</b>	<b>296</b>
<b>Appendix B – Graphical Representation of Kinetic Data .....</b>	<b>297</b>

## LIST OF TABLES

<b>Table 2.1.</b> Apparent Michaelis-Menten kinetic parameters for the transferase and hydrolase activities of Cst II .....	61
<b>Table 2.2.</b> Expected product inhibition patterns for select Bi Bi enzyme mechanisms .	68
<b>Table 2.3.</b> Michaelis-Menten kinetic parameters for Cst II catalyzed sialyl transfer to 3'-sialyl lactose .....	78
<b>Table 2.4.</b> Apparent Michaelis-Menten parameters for CMP NeuAc with various Cst II mutants .....	84
<b>Table 3.1.</b> Michaelis-Menten parameters for the transferase and hydrolase activities of WT and various mutants of LgtC-25 with respect to UDP Gal and UDP Glc donor substrates measured at pH 7.5, 37°C .....	100
<b>Table 4.1.</b> Common forms of nucleotide sugar donor substrates found in Nature and utilized by glycosyltransferases .....	144
<b>Table 4.2.</b> Michaelis-Menten kinetic parameters for wild type and N10D versions of LgtC using either CDP Gal or UDP Gal as the varied substrate at a fixed saturating concentration of lactose acceptor (160mM) .....	153
<b>Table 5.1.</b> Kinetic parameters for various LgtC acceptor substrates .....	191

## LIST OF FIGURES

- Figure 1.1.** Overall folds observed for glycosyltransferase enzymes. The GT-A fold (A) is represented by the inverting enzyme SpsA from *Bacillus subtilis* (pdb 1qgq) and that of the GT-B fold (B) by bacteriophage T4  $\beta$ -glucosyltransferase ..... 10
- Figure 1.2.** Like glycosidases, glycosyltransferases catalyze glycosyl group transfer with either inversion or retention of the anomeric stereochemistry with respect to the donor sugar ..... 14
- Figure 1.3.** Glycosyltransferase (GT) classification system ..... 16
- Figure 1.4.** A direct displacement  $S_N2$ -like reaction results in inverted anomeric configuration via a single oxocarbenium ion-like transition state ..... 17
- Figure 1.5.** Comparison of the active sites of several metal-dependent inverting glycosyltransferases from Clan I ..... 20
- Figure 1.6.** Active site of wild type T4 bacteriophage BGT from family GT63 ..... 24
- Figure 1.7.** The double displacement mechanism established for retaining glycosidases proceeds via two oxocarbenium ion-like transition states with the intermediate formation of a discrete covalently bound glycosyl-enzyme species, resulting in overall retention of anomeric configuration ..... 29
- Figure 1.8.** Proposed double displacement mechanism for retaining Glycosyltransferases ..... 31
- Figure 1.9.** Comparison of the active sites of several retaining glycosyltransferases from Clan III ..... 35

<b>Figure 1.10.</b> Comparison of the active sites of several retaining glycosyltransferases from Clan IV .....	47
<b>Figure 1.11.</b> The $S_{Ni}$ mechanism for the decomposition of alkyl chlorosulphites is a special case of an $S_{N1}$ process that leads to complete retention of stereochemistry in the product .....	51
<b>Figure 1.12.</b> Proposed “ $S_{Ni}$ -like” mechanism for retaining glycosyltransferases and glycogen phosphorylase involving a direct “front-side $S_{N2}$ -like” displacement proceeding through a highly dissociative oxocarbenium ion-like transition state resulting in retention of anomeric configuration .....	52
<b>Figure 2.1.</b> Cleland notation schematic representations of ping pong (A), Random sequential (B), ordered sequential (C), and Theorell-Chance (D) Bi Bi kinetic mechanisms .....	64
<b>Figure 2.2.</b> Double reciprocal and secondary plots of initial rate data with respect to CMP NeuAc at various constant concentrations of 3'-sialyl lactose (3'-SL) .....	67
<b>Figure 2.3.</b> Representative CE data used to monitor Cst II sialyltransferase activity ...	69
<b>Figure 2.4.</b> Inhibition of Cst II catalyzed sialyl transfer to FCHASE 3'-SL by the product CMP with respect to donor (A) and acceptor (B) substrates .....	70
<b>Figure 2.5.</b> Inhibition of Cst II catalyzed sialyl transfer to 3'-SL by the dead end substrate analogue CMP 3F NeuAc with respect to donor (A) and acceptor (B) Substrates .....	72
<b>Figure 2.6.</b> Cleland notation schematic representation of the steady state iso ordered Bi Bi mechanism .....	75

<b>Figure 2.7.</b> Overall structure of Cst II. Top down (A) and side (B) views of the overall tetrameric arrangement of Cst II .....	80
<b>Figure 2.8.</b> Cst II active site highlighting the hydrogen bond interactions between the side chain hydroxyls of Y156 and Y162 and the departing CMP leaving group .....	83
<b>Figure 2.9.</b> Cst II active site highlighting the side chain imidazole of His188 suitably positioned on the $\alpha$ -face of the donor substrate to play the role of base catalyst .....	86
<b>Figure 2.10.</b> pH dependence of $k_{cat}/K_m$ for Cst II. The pKa value was determined by best fit of the data to a single titration model in Grafit 4.0 .....	87
<b>Figure 2.11.</b> Chemical rescue of nucleophile (A) and acid/base (B) mutant glycosidases with small molecules .....	88
<b>Figure 2.12.</b> Chemical rescue of Cst II H188A .....	89
<b>Figure 3.1.</b> LgtC active site with bound substrate analogues UDP 2F Gal and 4'-deoxy lactose highlighting the proximity of the side chain amide of Gln189 and the 6'-hydroxyl of lactose to the anomeric reaction centre .....	96
<b>Figure 3.2.</b> Mechanism of retaining $\beta$ -hexosaminidases involving anchimeric assistance leading to the formation of a stabilized oxazolinium ion intermediate .....	96
<b>Figure 3.3.</b> ESI-MS analysis of labelled and unlabelled species of LgtC Q189E .....	102
<b>Figure 3.4.</b> Decay data for the turn over of the glycosyl-enzyme intermediate species resulting from the incubation of UDP Glc with LgtC Q189E .....	103
<b>Figure 3.5.</b> ESI-MS analysis of the fragmented, purified Gal-labelled peptide .....	106

<b>Figure 3.6.</b> Superimposition of the wild type and Q189E mutant active sites of LgtC .....	108
<b>Figure 3.7.</b> LgtC active site highlighting the side chains of residues within close proximity to that of Asp190 .....	111
<b>Figure 3.8.</b> Acyl migration of the galactosyl moiety from the side chain of Q189 to that of D190 is difficult to rationalize .....	112
<b>Figure 3.9.</b> Kre2 active site with bound intact donor substrate analogue GDP 2F Man .....	119
<b>Figure 3.10.</b> Active site of $\alpha$ 3GalT with bound non-reactive donor substrate analogue UDP 2F Glc highlighting the position of Glu317 suitably positioned to play the role of a catalytic nucleophile in a double displacement mechanism .....	124
<b>Figure 3.11.</b> Comparison of the potential differences in the free energy barriers for $S_N1$ -like and $S_N2$ -like pathways that could be utilized by (A) retaining glycosidases (GH) and (B) retaining glycosyltransferases (GT) .....	127
<b>Figure 3.12.</b> Chemical similarities between the substrate used in the first step of the glycosyltransferase reaction and the covalent glycosyl-enzyme intermediate turned over in the second step of a $\beta$ -retaining glycosidase catalyzed reaction .....	129
<b>Figure 3.13.</b> The reactions catalysed by FPP synthetase .....	133
<b>Figure 3.14.</b> Proposed $D_N^*A_{Nssip}$ nucleophilic substitution reaction for retaining glycosyltransferases .....	141

<b>Figure 4.1.</b> Hydrogen bonding interactions between guanosine and a side chain carboxylate (A) or xanthosine and a side chain amide (B) determines the base specificity of EF-Tu and FtsY .....	148
<b>Figure 4.2.</b> Hydrogen bonding interactions between guanosine and a side chain carboxylate that are found to control base specificity between guanosine and xanthosine in the case of EF-Tu (A) and FtsY (B) .....	149
<b>Figure 4.3.</b> Hydrogen bonding interactions between uridine and a side chain amide (A) or cytidine and a side chain carboxylate (B) proposed to determine base specificity ...	150
<b>Figure 4.4.</b> Active sites of the GT family 8 members glycogenin (A) and LgtC (B) revealing the conserved interaction of a side chain amide with the 4-carbonyl group of the bound uridine ring .....	151
<b>Figure 4.5.</b> Graphical representation of the macromolecular anatomy of an Ff filamentous phage particle .....	157
<b>Figure 4.6.</b> Schematic representations of particles resulting from various phage display systems available for the surface display of foreign peptides or entire proteins .....	159
<b>Figure 4.7.</b> Recombinant construct within the MCS of pLL01 encoding an LgtC-pIII fusion protein .....	165
<b>Figure 4.8.</b> TLC analysis (7:2:1 EtOAc:MeOH:H <sub>2</sub> O) of the LgtC activity of twice PEG precipitated M13 bacteriophage derived from pLL01 phagemid .....	166
<b>Figure 4.9.</b> Western blot analysis of various phage particles using anti-c-myc-HRP for detection .....	167

<b>Figure 4.10.</b> TLC analysis (7:2:1 EtOAc:MeOH:H <sub>2</sub> O) of the LgtC activity of emulsified M13 bacteriophage derived from pLL01 phagemid .....	169
<b>Figure 4.11.</b> Proposed screening procedure for the <i>in vitro</i> evolution of LgtC involving compartmentalized phage display .....	174
<b>Figure 4. 12.</b> Three-dimensional x-ray crystal structure of Shiga-like toxin I B subunit with bound starfish molecule .....	178
<b>Figure 4.13.</b> Phage ELISA for SAV display .....	181
<b>Figure 4.14.</b> Phage ELISA for SAV display .....	182
<b>Figure 5.1.</b> <sup>1</sup> H-COSY NMR spectrum of <i>para</i> -Nitrophenyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 4)-2,3,6-tri-O-acetyl- $\beta$ -D-glucoside .....	193
<b>Figure 5.2.</b> Acceptor binding site of LgtC containing a hydrophobic platform (Phe132) for binding the glucose sugar ring of lactose or an aromatic substituent .....	195
<b>Figure 5.3.</b> <sup>1</sup> H-COSY NMR spectrum of Benzyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 3)-2,4-di-O-acetyl- $\beta$ -D-xyloside .....	196
<b>Figure 5.4.</b> <sup>1</sup> H-COSY NMR spectrum of <i>para</i> -Nitrophenyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 3)-2,4,6-tri-O-acetyl- $\beta$ -D-galactoside .....	197
<b>Figure 5.5.</b> <sup>1</sup> H-COSY NMR spectrum of <i>n</i> -Octyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 3)-2,4,6-tri-O-acetyl- $\beta$ -D-glucoside .....	199
<b>Figure 5.6.</b> <sup>1</sup> H-COSY NMR spectrum of <i>n</i> -Octyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 4)-2,3,6-tri-O-acetyl- $\beta$ -D-galactoside .....	200

<b>Figure 5.7.</b> $^1\text{H}$ -COSY NMR spectrum of <i>n</i> -Octyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 3)-2,4-di-O-acetyl- $\beta$ -D-xyloside .....	201
<b>Figure 5.8.</b> A model for the substrate engineering strategy using 6-OBz Man .....	202
<b>Figure 5.9.</b> Overlay of the MM2 minimized structure of 6-OBz Man (green) with the structure of 4-deoxylactose (orange) bound within the active site of LgtC .....	203
<b>Figure 5.10.</b> $^1\text{H}$ -COSY NMR spectrum of (2,3,4,6-Tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 2)-1,3,4-tri-O-acetyl-6-O-benzoyl- $\alpha$ -D-mannose .....	204
<b>Figure 5.11.</b> A model for the substrate engineering strategy using 4-O-Benzoyl-D-xylose (4-OBz Xyl) .....	205
<b>Figure 5.12.</b> $^1\text{H}$ -COSY NMR spectrum of (2,3,4,6-Tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 2)-1,3,-di-O-acetyl-4-O-benzoyl- $\alpha/\beta$ -D-xylose .....	206
<b>Figure 5.13.</b> Substrate engineering strategy applied to various alternative LgtC acceptor substrates .....	208
<b>Figure 5.14.</b> Representation of the product binding mode of Cst II (A) and the proposed mode of binding of $\alpha$ -linked pNP NeuAc that would facilitate <i>in situ</i> generation of the natural donor substrate .....	211
<b>Figure 5.15.</b> (A) Monitoring Cst II catalyzed glycosylation of fluorescent fluorescein lactose conjugate acceptor substrate using <i>p</i> -nitrophenyl $\alpha$ -sialoside as an alternative donor substrate in the presence or absence of CMP by TLC (4:2:1:0.1 EtOAc: MeOH: H <sub>2</sub> O: AcOH). (B) Reaction catalysed by Cst II using the alternative donor substrate <i>p</i> -nitrophenyl $\alpha$ -sialoside .....	212

**Figure 5.16.** (A) Monitoring LgtC catalyzed glycosylation of fluorescent fluorescein lactose conjugate acceptor substrate using 2,4-dinitrophenyl  $\beta$ -galactoside as an alternative donor substrate in the presence or absence of UDP by TLC (7:2:1:0.1 EtOAc: MeOH: H<sub>2</sub>O: AcOH). (B) Reaction catalysed by LgtC using the alternative donor substrate 2,4-dinitrophenyl  $\beta$ -galactoside ..... 216

**Figure 7.1.** TLC analysis of  $\alpha$ -2,3 specific neuraminidase treatment of glycosylation product derived from Cst II using pNP  $\beta$ -galactoside as an alternative acceptor substrate ..... 263

**Figure 7.2.** TLC analysis of  $\alpha$ -2,3 specific neuraminidase treatment of glycosylation product derived from Cst II using *p*-nitrophenyl  $\alpha$ -sialoside as an alternative donor substrate in the presence CMP monitored by TLC ..... 264

**Figure A.1.** Michaelis-Menten plots of various Cst II activities ..... 297

**Figure A.2.** Michaelis-Menten plots of wild type LgtC activity with various acceptor substrates ..... 298

## LIST OF SCHEMES

<b>Scheme 1.1.</b> Overall reactions catalyzed by (a) glycosidases, (b) glycosyltransferases, (c) phosphorylases, and (d) polysaccharide lyases .....	5
<b>Scheme 2.1.</b> Monofunctional and bifunctional sialyltransferase activities of Cst II enzymes .....	60
<b>Scheme 2.2.</b> Enzyme coupled continuous assay used to measure the kinetic parameters of sialyltransferases .....	62
<b>Scheme 2.3.</b> Enzymatic synthesis of the dead end donor substrate analogue CMP 3F NeuAc .....	71
<b>Scheme 3.1.</b> Galactosylation reaction catalyzed by LgtC .....	95
<b>Scheme 3.2.</b> Equilibrium for LgtC catalyzed hydrolysis of UDP Gal in the presence of pyruvate kinase (PK) and phosphoenolpyruvate (PEP) .....	104
<b>Scheme 3.3.</b> Mannosylation reactions catalyzed by Kre2 .....	116
<b>Scheme 3.4.</b> Synthesis of the non-reactive donor substrate analogue GDP 2F Mannose via a morpholidate coupling reaction .....	116
<b>Scheme 3.5.</b> Galactosylation reaction catalyzed by $\alpha$ 3GalT .....	122
<b>Scheme 4.1.</b> Chemical synthesis of the desired unnatural nucleotide sugar CDP Gal .	152
<b>Scheme 4.2.</b> Synthesis of a biotin lactose conjugate acceptor substrate for LgtC .....	175

## LIST OF ABBREVIATIONS

Ab	antibody
ABTS	2,2'-azino-bis[3-ethylbenziazoline-6-sulphonic acid]
ADP	adenosine 5'-diphosphate
ATP	adenosine 5'-triphosphate
BSA	bovine serum albumin
CDP	cytidine 5'-diphosphate
CDP Gal	cytidine 5'-diphosphogalactose
CE	capillary electrophoresis
CMP	cytidine 5'-monophosphate
CMP NeuAc	cytidine 5'-monophospho-N-acetylneuraminic acid
CTP	cytidine 5'-triphosphate
DCC	dicyclohexylcarbodiimide
DMF	dimethylformamide
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphates
DTT	dithiothreitol
dNP	2,4-dinitrophenyl
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
Et <sub>3</sub> N	triethylamine
ESI-MS	electrospray ionization mass spectrometry
EtOAc	ethyl acetate
FCHASE	6-(5-fluorosceincarboxamido)-hexanoic acid succinidyl ester
GAG	glycosaminoglycan
Gal	galactose
GalNAc	N-acetyl-D-galactosamine
GDP	guanosine 5'-diphosphate
GDP Man	guanosine 5'-diphosphomannose

GDP 2F Man	guanosine 5-diphospho-(2''-deoxy-2''fluoro)- $\alpha$ -D-mannose
Glc	glucose
GlcA	glucuronic acid
GlcNAc	N-acetyl-D-glucosamine
Glc1P	$\alpha$ -D-glucose 1-phosphate
GMP	guanosine 5' monophosphate
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulphonic acid
hGH	human growth hormone
HPLC	high performance liquid chromatography
HR	high resolution
HRP	horse radish peroxidase
ICAT	isotope-coded affinity tags
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl $\beta$ -D-thiogalactopyranoside
IUBMB	International Union of Biochemistry and Molecular Biology
IUPAC	International Union of Pure and Applied Chemists
KIE	kinetic isotope effect
LDH	lactate dehydrogenase
ManNAc	N-acetyl-D-mannosamine
MeOH	methanol
MCS	multiple cloning site
NAD <sup>+</sup>	$\beta$ -nicotinamide adenine dinucleotide
NADH	$\beta$ -nicotinamide adenine dinucleotide, reduced
NDP	nucleoside diphosphate
NMPK	nucleoside monophosphate kinase
NMR	nuclear magnetic resonance
6OBz Man	6-O-benzoyl mannose
4OBz Xyl	4-O-benzoyl xylose
PBS	phosphate buffered saline
PEG	polyethylene glycol
PEP	phospho(enol)pyruvic acid

PK	pyruvate kinase
PLP	pyridoxal phosphate
PNP	<i>para</i> -aminophenyl
RBS	ribosome binding site
RNA	ribonucleic acid
SAV	streptavidin
SDS	sodium dodecyl sulphate
SDS PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SL	3'-sialyllactose
SSL	3'8''-disialyllactose
TDP	thymidine 5'-diphosphate
TRIS	2-amino-2-(hydroxymethyl)-1,3-propanediol
Tween	polyethylene glycol sorbitan monolaurate
UDP	uridine 5'-diphosphate
UDP Gal	uridine 5'-diphosphogalactose
UDP 2F Gal	uridine 5'-diphospho-(2''-deoxy-2''fluoro)- $\alpha$ -D-galactose
UDP Glc	uridine 5'-diphosphoglucose
UDP 2F Glc	uridine 5'-diphospho-(2''-deoxy-2''fluoro)- $\alpha$ -D-glucose
UDP GlcA	uridine 5'-diphosphoglucuronic acid
UMP	uridine 5'-monophosphate

## ACKNOWLEDGEMENTS

First and foremost I would like to thank Professor Stephen G. Withers for his support during the course of my doctoral studies at UBC. He has provided me not only with tremendous scientific and personal insights, but also with a research environment that is limited only by one's creativity, scientific rigour and self-motivation. I would especially like to thank him for allowing me to pursue multiple research projects. The diversity of these projects has facilitated the development of skills in a broad range of research techniques and a foundation firmly rooted in the logic of chemistry yet not limited by a classification of academic discipline.

I would also like to thank Dr. Warren W. Wakarchuk at the National Research Centre in Ottawa for providing me with clones for multiple enzymes and for allowing me to work under his direct supervision in order to develop molecular biology research skills. In addition, Professor Natalie C.J. Strynadka has provided me with extraordinary insight into the world of protein X-ray crystallography and has shown constant interest in my research with glycosyltransferases. Prof. Tony Warren has provided me with laboratory space and with tremendous insights into molecular biology and microbiology techniques. The incredible knowledge base of members of the Withers laboratory has played a critical role in my research. I would especially like to thank Dr. Chris Tarling, Dr. Andrew Watts, Dr. Amir Aharoni, Dr. James MacDonald, and Dr. Omid Hekmat. I must also thank my former research mentors at the University of Guelph for fostering my interest in research science, especially Prof. Bernard Grodzinski and Prof. Adrian Schwan. My family (Kathy, Lloyd and Jaden Lairson) has always supported me, for which I am truly

grateful. I would also like to thank my friends, especially Huxley, for their support and well needed intermittent reality checks during the course of my doctoral studies.

Finally, I am truly grateful for the financial support of the Natural Science and Engineering Research Council of Canada (NSERC) and the Michael Smith Foundation for Health Research (MSFHR).

## **Co-Authorship Statement**

The research described in this thesis was identified and designed by Luke L. Lairson under the guidance of Prof. Stephen G. Withers. Unless explicitly stated, Luke L. Lairson performed all of the research and data analyses described in this thesis. Luke L. Lairson wrote the entirety of this thesis with critical readings and suggestions provided by Prof. Stephen G. Withers. Modified versions of portions of previously published text that appear in this thesis were both originally written and subsequently modified by Luke L. Lairson.

# CHAPTER 1

## INTRODUCTION

\* A version of portions of this chapter has been published:

Lairson, L.L. and Withers, S.G. (2004) Mechanistic Analogies Amongst Carbohydrate Modifying Enzymes. *Chemical Communications*. (20): 2243. - Reproduced by permission of The Royal Society of Chemistry (RSC)

## 1.1. Carbohydrates in Biology

The central dogma of molecular biology consists of a flow of biological information in a template-controlled fashion from DNA to RNA to protein. The development of this paradigm has served to revolutionize both the modern view of life and the ways by which a deep understanding of biology are explored. An appreciation of these precise template-controlled mechanisms has allowed the development of tools that can be used to manipulate one class of molecules based upon an understanding of the other. Examples include recombinant DNA technologies using restriction enzymes and the expression of designer proteins using engineered DNA. However, a more complete view of life at the molecular level incorporates the roles of other major classes of biological molecules including glycans, lipids and small molecules. These classes are not derived in a template-controlled fashion, making their *in vitro* synthesis a more formidable task. This has resulted in a relative lag in the progression of our understanding of their roles in biology compared to that of DNA and proteins.

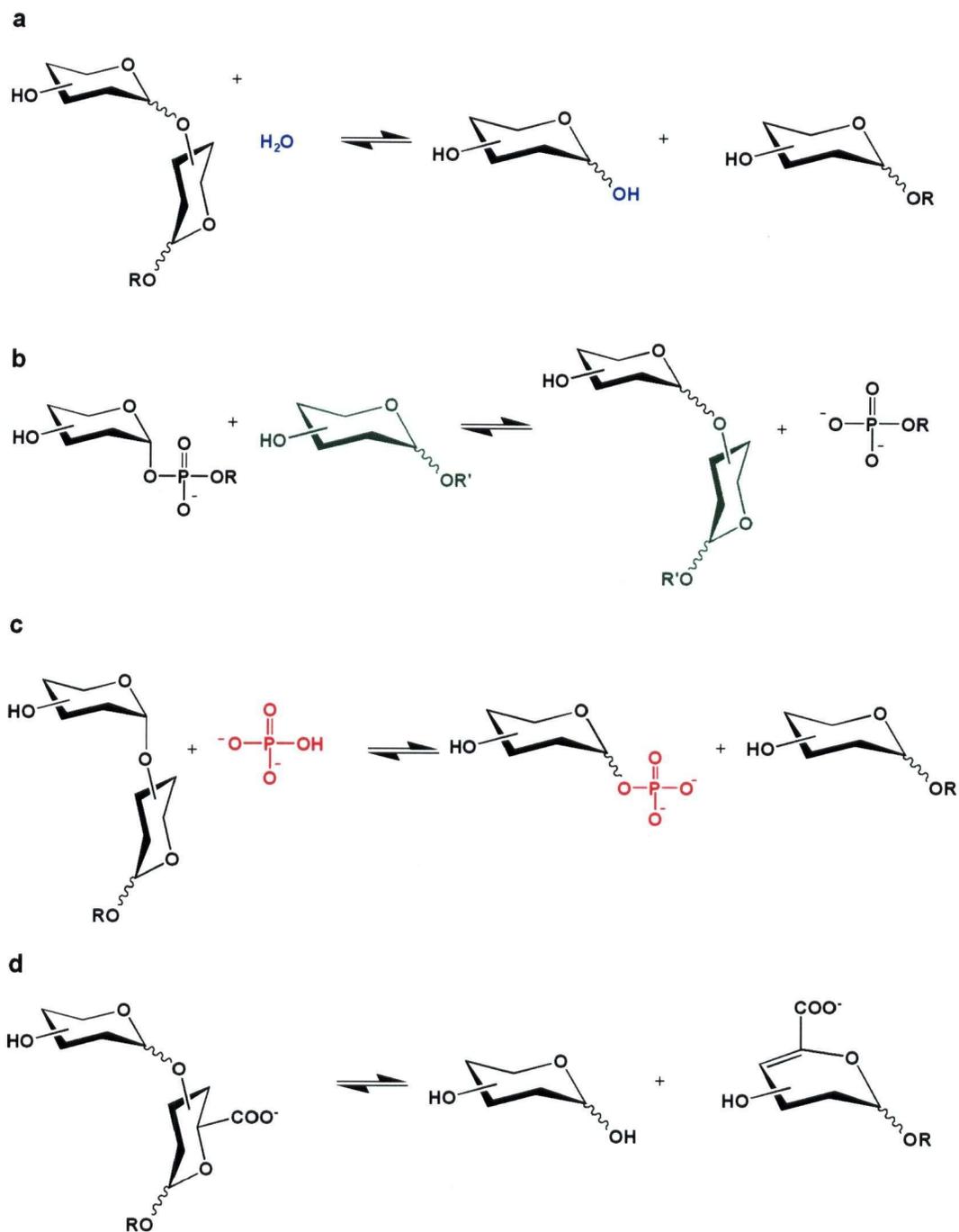
In terms of sheer biomass, carbohydrates are in fact the largest class of biological macromolecules. Carbohydrates are traditionally known for their roles in providing structural platforms and in serving as energy storage devices. Cellulose in the cell walls of plants, chitin in the exoskeletons of insects, and glycosaminoglycans (GAGs) in the extracellular matrices of various mammalian tissues are important examples of critical structural roles that carbohydrates play in biology. The base metabolic currencies used to generate and store energy are sugars, with starch and glycogen serving as the reservoirs used by plants and animals respectively to store the initial input of glycolysis (glucose). A coming together of biochemistry, synthetic carbohydrate chemistry, and enzymology

led to development of an interdisciplinary field of study that was formally termed Glycobiology in 1988 (Rademacher et al., 1988). From this, a more recent understanding of the important roles that carbohydrates play in molecular recognition events dictating a variety of critical biological phenomena has arisen.

Because of the range of available monosaccharide building blocks and the various manners by which individual units can be connected via alternative glycosidic linkages, glycan structures can achieve a level of structural diversity that exceeds that of DNA and even proteins. Whereas both DNA and proteins are linear polymers with a single type of linkage, the individual units of oligosaccharides can in theory be connected by either an  $\alpha$ - or  $\beta$ -linkage to any of several positions of another sugar. This point is illustrated by considering the number of variations possible from the combination of three different amino acids, nucleotide bases, or monosaccharides in a peptide, DNA, or oligosaccharide polymer respectively. In the case of both the peptide and DNA only six variations are possible. In contrast, three hexoses could produce up to 27,648 unique trisaccharide structures (Laine, 1994). Nature has taken advantage of the potential of this diverse structural space by using glycans as information display systems that are recognized by specific proteins (lectins, microbial carbohydrate binding proteins, GAG-binding proteins, etc.). These protein-glycan interactions, which occur in the nucleus and cytoplasm and on both inter- and extracellular membrane surfaces, are involved in processes ranging from cell-cell recognition and signal transduction to the entry of pathogens into host cells (Rademacher et al., 1988).

## 1.2. Biological Glycosyl Group Transfer Reactions

The enormous complexity of the various oligosaccharide structures found in nature is derived from a rational orchestration of the enzymatic formation and breakdown of glycosidic linkages achieved by glycosyltransferases, glycosidases, phosphorylases, and polysaccharide lyases (Scheme 1.1). The importance of these classes of enzymes is illustrated by the fact that ~1-3% of the proteins encoded by most organisms are those of enzymes that catalyse the synthesis and breakdown of glycosidic bonds (Davies et al., 2005). With the exception of the polysaccharide lyases that utilize mechanisms involving  $\beta$ -elimination (Ernst et al., 1995; Lee et al., 2002a; Rye and Withers, 2002) and a recently described mechanism involving transient remote redox and  $\beta$ -elimination steps utilized by a unique family of glycosidases (Yip et al., 2004), these classes of enzymes are generally believed to utilize mechanisms involving straightforward direct nucleophilic substitution at the anomeric reaction centre to catalyse glycosyl group transfer from a donor substrate to an acceptor. Enzyme catalysed glycosyl group transfer reactions proceed through transition states similar to those for non-enzymatic nucleophilic substitution reactions at acetal centres. Such transition states have considerable oxocarbenium ion-like character in which both the incoming nucleophile and departing leaving group interact weakly with a reaction centre possessing a significant degree of positive charge that is stabilized by the lone pair of electrons donated by an endocyclic oxygen atom.



**Scheme 1.1.** Overall reactions catalysed by (a) glycosidases, (b) glycosyltransferases, (c) phosphorylases, and (d) polysaccharide lyases.

While the mechanistic strategies used by glycosidases to catalyse glycosidic bond hydrolysis are fairly well understood on both a structural and chemical level (Davies, 1998); (Zechel and Withers, 2000), the characterization and mechanistic understanding of the glycosyltransferases responsible for glycoside bond formation has lagged far behind. In the case of glycosidases, a glycosyl moiety is transferred to water, which acts as the acceptor substrate, while in the case of glycosyltransferases the acceptor is typically the hydroxyl group of an acceptor other than water. Despite a lack of evolutionary relatedness, by simple chemical analogy, glycosyltransferases had been thought to use mechanistic strategies that directly parallel those used by glycosidases and transglycosidases (Davies, 1998). However, some distinct differences are becoming apparent, as will be discussed below. Sandwiched between these two enzyme groups are the phosphorylases; some of which clearly follow a glycosidase type of mechanism while others show much greater mechanistic and structural parallels to the glycosyltransferases.

### **1.3. Glycosyltransferases**

#### **1.3.1. Glycosyltransferase Activities**

The IUBMB classification system denotes glycosyltransferase enzymes with an EC number of 2.4.x.y. However, several of these enzymes utilize disaccharides, oligosaccharides or polysaccharides as donor sugar substrates. Examples include cyclodextrin glucanotransferases (2.4.1.19), dextransucrases (2.4.1.5), and xyloglucan endotransferases (2.4.1.207). These enzymes are structurally, mechanistically and evolutionarily related to glycosidases and are more accurately termed transglycosidases and are therefore more appropriately classified along with glycosidases. As such, this

thesis will reserve the term glycosyltransferase for those enzymes that utilize an activated donor sugar substrate that contains a (substituted) phosphate leaving group. Donor sugar substrates are most commonly activated in the form of nucleoside diphosphate sugars (e.g. UDP Gal, GDP Man, etc.), however nucleoside monophosphate sugars (e.g. CMP NeuAc) and lipid phosphates (e.g. dolichol phosphate oligosaccharides) are also used. Nucleotide sugar dependent glycosyltransferases are often referred to as Leloir enzymes, in honor of Luis F. Leloir who discovered the first sugar nucleotide and was awarded the Nobel prize in chemistry in 1970 for his enormous contributions to our understanding of glycoside biosynthesis and sugar metabolism. The acceptor substrates utilized by glycosyltransferases are most commonly other sugars but can also be a lipid, protein, nucleic acid or antibiotic small molecule. In addition, while glycosyl transfer most frequently occurs to the nucleophilic oxygen of a hydroxyl substituent of the acceptor, it can also occur to nitrogen (e.g. the formation of *N*-linked glycoproteins), sulphur (e.g. the formation of thioglycosides in plants) and carbon (e.g. C-glycoside antibiotics) nucleophiles.

### **1.3.2. Glycosyltransferase Folds**

As has been done for other classes of “Carbohydrate-Active enZYmes” (CAZY), glycosyltransferases are classified into families based on gene sequence similarities (Campbell et al., 1997; Coutinho et al., 2003). This information is maintained and continually updated on the CAZY website (<http://cazy.org>). Indicative of the pace and progress of current genome sequencing endeavors, in the ten years since Campbell and colleagues first started the classification of glycosyltransferases, the number of distinct

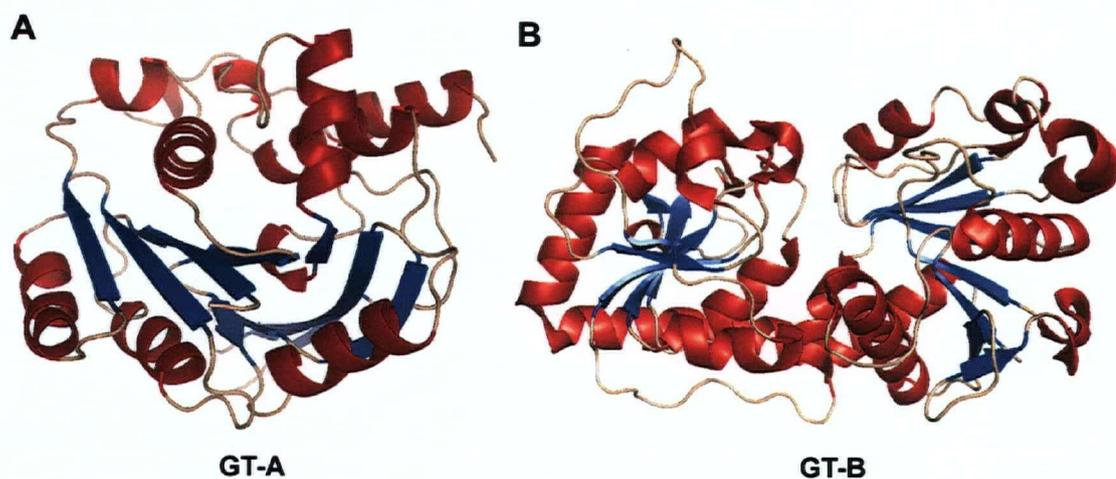
families has grown from 27 to 90. However, many of these sequences are simply open reading frames that encoded putative glycosyltransferases and detailed biochemical characterization of this class of enzymes remains relatively scarce compared to that of the glycosidases. This relative lag results from the fact that the vast majority of glycosyltransferases are membrane-associated proteins, thereby causing problems associated with the production of insoluble and therefore uncharacterizable material when expressed as full-length proteins. Using recombinant DNA technology, this problem is being overcome by the creation of truncated, soluble forms of the enzymes consisting solely of the catalytic domains, thus greater progress should be forthcoming in the near future.

Structural and mechanistic characterization of representatives from a large number of the 110 families of glycosidases has been achieved. These studies have revealed an extraordinary degree of diversity in overall fold, despite considerable commonalities in active site features and catalytic mechanisms. This would indicate a convergence in mechanism during the course of evolution. In contrast, a recent burst of reported glycosyltransferase structures has revealed a quite different situation as only two general folds, called GT-A and GT-B, have been observed for all structures of NDP sugar-utilizing glycosyltransferases solved to date (Coutinho et al., 2003; Hu and Walker, 2002; Unligil and Rini, 2000). Further, threading analysis has revealed that the majority of uncharacterized families are also predicted to adopt one of these two folds. This finding is probably a consequence of the structural constraints of a nucleotide-binding motif and may indicate that the majority of glycosyltransferases have evolved from a small number of progenitor sequences. Tantalizing support for this latter notion is derived from the fact

that only two glycosyltransferase families (GT2 having the GT-A fold and GT4 having the GT-B fold) are possessed by primitive Archaea, suggesting that these may have served as the origins from which the vast majority of glycosyltransferases have evolved (Coutinho et al., 2003). In addition, within the GT4 family, there exist members that utilize donor substrates activated with a nucleoside diphosphate and others that utilize donor substrates activated with a phosphate group, suggesting an evolutionary link between enzymes that utilize these two substrate forms.

The GT-A fold was first described for the inverting enzyme SpsA from *Bacillus subtilis*, for which both apo- and UDP-bound three-dimensional X-ray crystal structures were obtained (Charnock and Davies, 1999). Consisting of an open twisted  $\beta$  sheet surrounded by  $\alpha$  helices on both sides, the overall architecture of the GT-A fold is reminiscent of the "Rossmann-like" fold of nucleotide-binding proteins. Two tightly associated  $\beta/\alpha/\beta$  domains, the sizes of which vary, abut closely leading to the formation of a continuous central  $\beta$  sheet (Figure 1.1A). For this reason some describe the GT-A fold as a single domain fold. However, distinct nucleotide and acceptor binding domains are present (Unligil and Rini, 2000). Eukaryotic members possessing the GT-A fold typically have a short N-terminal cytoplasmic domain that is followed by transmembrane and stem regions leading to the globular catalytic region (Breton and Imberty, 1999). It should be noted that not all enzymes that possess a GT-A fold are glycosyltransferases. Indicating the possibility of a divergence in mechanism during the course of evolution, the sugar-1-phosphate pyrophosphorylase/nucleotidyl transferase superfamily of enzymes, responsible for the synthesis of nucleoside diphosphate sugars, have been shown to possess the GT-A architecture (Brown et al., 1999). Finally, most GT-A

enzymes possess what is termed a DXD sequence motif, in which the carboxylates coordinate a divalent cation and/or a ribose (Breton et al., 1998; Wiggins and Munro, 1998). These are not sequence-invariant motifs and although frequently described as a determining characteristic of GT-A glycosyltransferases, examples exist of enzymes from this superfamily that do not possess the DXD “signature” (Pak et al., 2006). In addition, as ~50% of all protein sequences (119,805) in SwissProt are found to possess a DxD, it certainly cannot be used as a diagnostic for a potential glycosyltransferase (Coutinho et al., 2003).



**Figure 1.1.** Overall folds observed for glycosyltransferase enzymes. The GT-A fold (A) is represented by the inverting enzyme SpsA from *Bacillus subtilis* (pdb 1qgq) and that of the GT-B fold (B) by bacteriophage T4  $\beta$ -glucosyltransferase (pdb 1jg7).

In 1994, the first determined three-dimensional structure for a nucleoside diphosphate utilizing glycosyltransferase was reported and was that of a DNA modifying  $\beta$ -glucosyltransferase from bacteriophage T4 (Vrielink et al., 1994). The overall fold of

this protein was found to be homologous to that of glycogen phosphorylase and, now that it has been observed for multiple other glycosyltransferases it is termed the GT-B fold. Like the GT-A fold, the architecture of GT-B enzymes consists of two  $\beta/\alpha/\beta$  Rossmann-like domains, however, in this case the two domains are less tightly associated and face each other with the active site lying within the resulting cleft (Figure 1.1B). Also analogous to the GT-A fold, there are discrete domains associated with the donor and acceptor substrate binding sites. In addition, non-glycosyltransferase enzymes are also known to adopt the GT-B fold. One such example is the case of UDP GlcNAc 2-Epimerase (Campbell et al., 2000).

A third glycosyltransferase fold termed GT-C was recently predicted based on iterative BLAST searches (Liu and Mushegian, 2003), however a structure has yet to be solved for any protein from this proposed superfamily. This report predicted eight families (GT22, GT39, GT48, GT50, GT53, GT57, GT58, GT59) to possess the GT-C fold. Following the publication of this report, four (GT66, GT83, GT85, GT86) of the 15 glycosyltransferase families subsequently created were also predicted to adopt this fold by the CAZY database. The predicted architecture of the GT-C fold is that of a large hydrophobic integral membrane protein located in the ER or on the plasma membrane having between eight and 13 transmembrane helices and an active site located on a long loop region (Maeda et al., 2001; Strahlbolsinger et al., 1993; Takahashi et al., 1996). Consistent with this predicted architecture is the fact that ten of the twelve families (all but GT48 and GT53) predicted to adopt the GT-C fold utilize lipid-phosphate activated donor sugar substrates. Perhaps indicative of the limitations of prediction based solely on BLAST analysis, a subsequent comparison of glycosyltransferase families using a

“profile hidden Markov method” of sequence analysis resulted in classification of family GT48, which uses UDP Glc as donor substrate, within the GT-A superfamily (Kikuchi et al., 2003). Members of family GT53, predicted to adopt the GT-C fold, utilize UDP L-arabinose as donor substrates and, as such, it will be interesting to see if this family does in fact adopt the proposed GT-C fold.

In addition, not all enzymes that utilize lipid phosphate-activated donor sugars are predicted, based on sequence analysis, to adopt the GT-C fold. For example, members of family GT51 utilize lipid II diphosphate activated donor sugars and family GT76 use dolichol phosphate activated donor sugars and are not predicted to adopt this fold. These are both members of a handful of orphan families that are not predicted to adopt either the GT-A, GT-B or proposed GT-C fold. Finally, certain families, for example GT2 that adopts the GT-A fold, contains both nucleoside diphosphate donor sugar-utilizing members and lipid phosphate donor sugar-utilizing members.

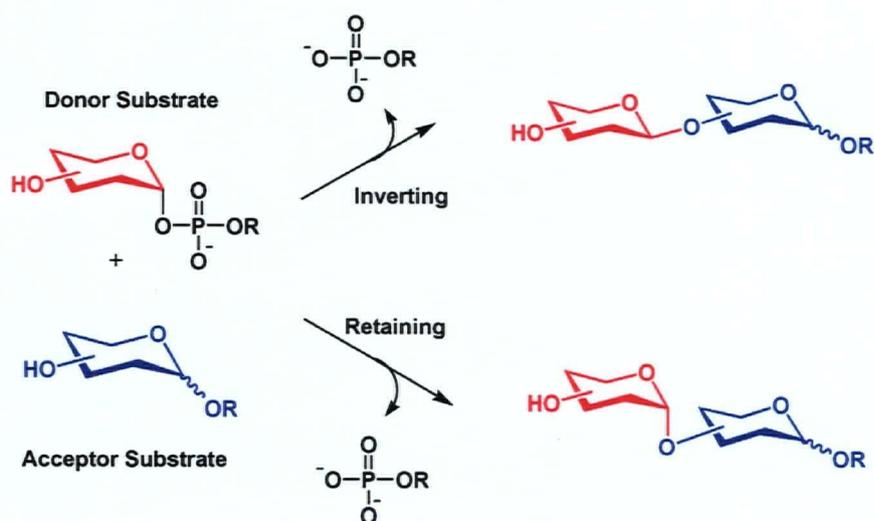
A structural and mechanistic study of chitobiose phosphorylase from *Vibrio proteolyticus* revealed that enzymes from family GT36 share more of a structural and evolutionary relationship with glycosidases of clan GH-L having an  $(\alpha/\alpha)_6$  fold (Hidaka et al., 2004). As a result, this family was subsequently reclassified as family GH-94. This may either be interpreted as evidence supporting the notion of a mechanistic continuum between glycosidases and glycosyltransferases or may simply be indicative of inherent limitations of the methods used for classification. Similarly, very recently, much-anticipated three-dimensional X-ray crystal structures of a peptidoglycan glycosyltransferase from family GT51 were solved with or without a bound substrate analogue/inhibitor (Lovering et al., 2007; Yuan et al., 2007). As mentioned above,

members of this inverting family utilize lipid II phosphate activated donor substrates and the structure of the glycosyltransferase domain was not found to show similarity to either the GT-A, GT-B or proposed GT-C folds. In fact, the  $\alpha$ -helical fold and several active site features are more akin to the lysozyme family of enzymes and this glycosyltransferase family will likely be reclassified amongst the glycosidase-like phosphorylases in the near future. This would be the first case of a lipid phosphate-dependent enzyme being classified amongst the glycosidases and again lends support to the notion of a mechanistic continuum.

Future results with the other orphan families currently classified as glycosyltransferases, but not predicted by sequence analysis to adopt either of the GT-A, GT-B or proposed GT-C folds, will undoubtedly shed more light on the evolutionary origin of glycosyltransferase activities and their relationships to the other major classes of carbohydrate active enzymes. Indeed, the release of the coordinates of a member of one such family (GT70), that have been deposited but not released in the RCSB Protein Data Bank, is much anticipated amongst aficionados. Whether these families will be found experimentally to actually adopt the GT-A, GT-B or putative GT-C folds; to form new glycosyltransferase superfamilies; or simply succumb to reclassifications as described above remains to be determined. In light of the described results with family GT 51, it is possible that those families proposed to adopt a GT-C fold will in fact be found to share more evolutionary and mechanistic commonalities with the glycosidases thereby leading to an appropriate mass reclassification of what will be defined as true glycosyltransferases.

### 1.3.3. Glycosyltransferase Classification

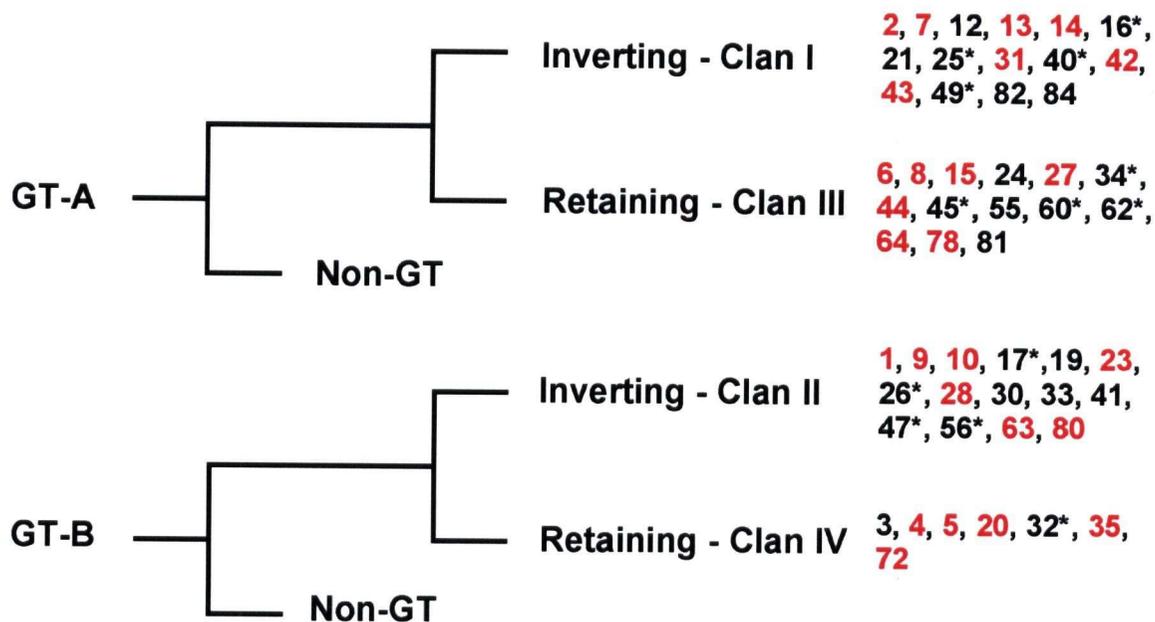
Two stereochemical outcomes are possible for reactions that result in the formation of a new glycosidic bond: the anomeric configuration of the product can either be retained or inverted with respect to the starting material (Figure 1.2). As such, enzymes catalysing glycosyl group transfer are classified as either inverting or retaining, depending on the outcome of the reaction. Logically, it follows that this stereochemical outcome must result from the utilization of different mechanisms by the two classes of enzyme.



**Figure 1.2.** Like glycosidases, glycosyltransferases catalyse glycosyl group transfer with either inversion or retention of the anomeric stereochemistry with respect to the donor sugar.

With the interesting exception of members of glycoside hydrolase (GH) family 4 (Thompson et al., 1998), the enzymes within a given GH family catalyse hydrolysis with the same stereo-selectivity. This is also believed to be the case amongst glycosyltransferase families. As such, glycosyltransferase families are classified into clans

depending on their fold and the stereochemical outcome of the reactions that they catalyse (Figure 1.3). The overall fold of the enzyme does not dictate the stereochemical outcome of the reaction that it catalyses as examples of both inverting and retaining glycosyltransferases have been identified within both the GT-A and GT-B fold superfamilies (Coutinho et al., 2003). Indicative of this phenomenon are recent findings from studies of a mannosylglycerate synthetase, which transfers mannose to the 2-OH of D-lactate, D-glycerate or glycollate with net retention of configuration (Flint et al., 2005). Based on amino acid sequence, this enzyme was initially classified amongst the GT-A inverting family GT2, however, structural and mechanistic studies led to its reclassification amongst the GT-A retaining family GT78. To date, all enzymes predicted to adopt the GT-C fold belong to inverting glycosyltransferase families.

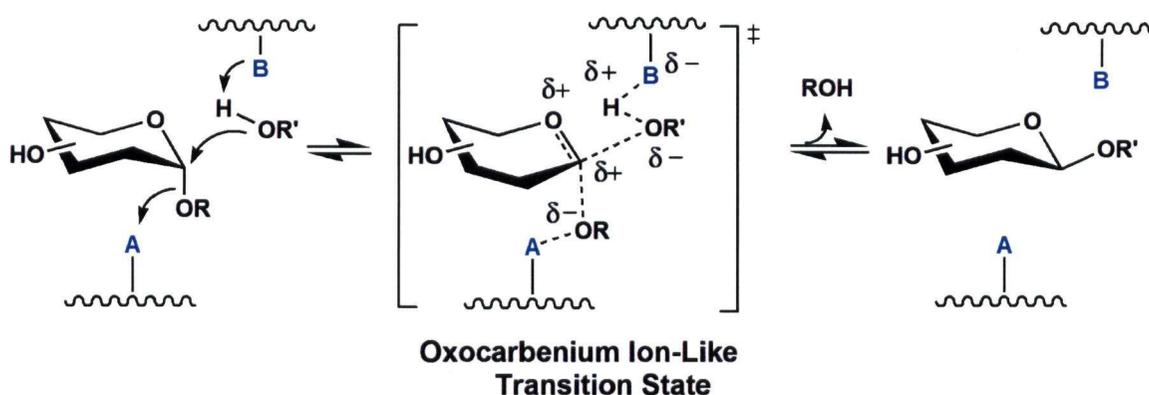


**Figure 1.3.** Glycosyltransferase (GT) classification system proposed by Coutinho et al. (Coutinho et al., 2003). Families are classified into clans based on their fold and activity. GT family numbers belonging to each clan are indicated on the far right. *Bona fide* families are indicated in red, having members with solved three-dimensional structures. The remaining families are those predicted to adopt either the GT-A or GT-B fold. Families identified with an asterisk are those with structures predicted to adopt either the GT-A or GT-B fold solely by Liu and Mushegian (Liu and Mushegian, 2003) and those not denoted with an asterisk have GT-A or GT-B structures as predicted by both Liu and Mushegian and the CAZY website. This classification system does not include 39 of the 90 glycosyltransferases. Members from twelve (GT22, GT39, GT48, GT50, GT53, GT57, GT58, GT59, GT66, GT83, GT85, GT86) of those families not included are predicted to adopt a proposed GT-C fold. Based on a determined three-dimensional structure, family GT36 was reclassified amongst the glycosidases as family GH94. Structural characterization of the remaining 26 orphan families (GT11, GT18, GT29, GT37, GT38, GT46, GT51, GT52, GT54, GT61, GT65, GT67, GT68, GT69, GT70, GT71, GT73, GT74, GT75, GT76, GT77, GT79, GT87, GT88, GT89, GT90) will provide insights into the strengths and limitations of predictive bioinformatic tools.

## 1.3.4. Inverting Glycosyltransferases

### 1.3.4.1. Mechanism of Inverting Glycosidases

The mechanistic strategy employed by inverting glycosidases is that of a direct displacement  $S_N2$ -like reaction. A pair of carboxyl groups exist within the active site, typically separated by 7-11 Å, one acting as an acid (A) protonating the glycosidic oxygen, the second acting as a base (B) that activates the incoming water nucleophile facilitating a reaction that proceeds via an oxocarbenium ion-like transition state (Zechel and Withers, 2000)(Figure 1.4).



**Figure 1.4.** A direct displacement  $S_N2$ -like reaction results in inverted anomeric configuration via a single oxocarbenium ion-like transition state. For glycosidases, R = a carbohydrate derivative and  $R'OH = H_2O$  or phosphate (phosphorylases classified as glycosidases). For glycosyltransferases, R = a nucleoside diphosphate, a lipid phosphate, or phosphate (phosphorylases classified as glycosyltransferases) and  $R'OH =$  an acceptor group (e.g. another sugar, a protein, or an antibiotic).

### 1.3.4.2. Mechanism of Inverting Glycosyltransferases

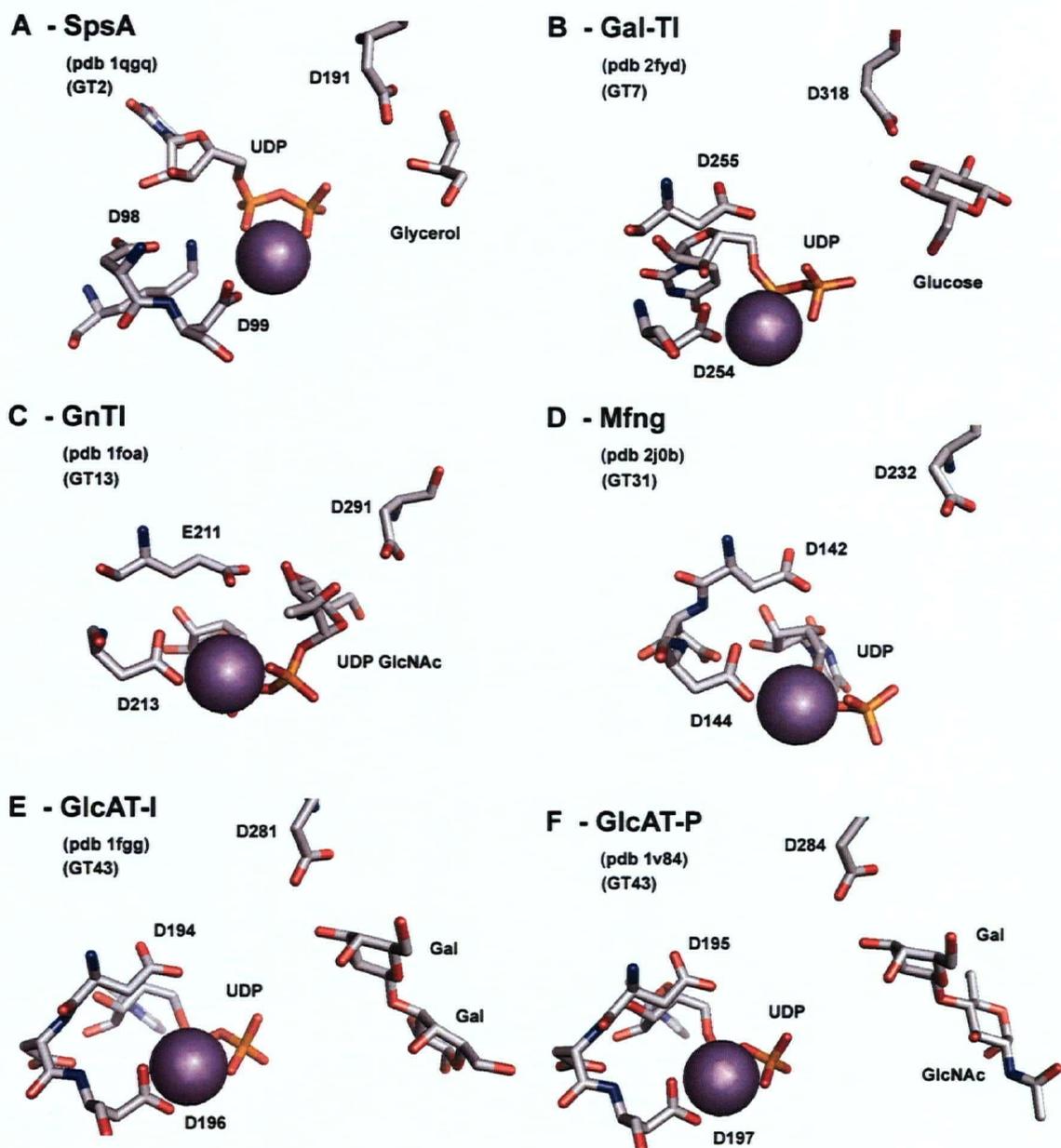
Catalysis by inverting glycosyltransferases is believed to directly parallel that of inverting glycosidases wherein a base (B) deprotonates the incoming nucleophile of the acceptor facilitating direct  $S_N2$  displacement of the activated (substituted) phosphate leaving group (Figure 1.4). Indeed, structural and mechanistic studies clearly point to such a mechanism with the major difference between these two classes simply being the manner by which leaving group activation/departure is facilitated. The key questions in examining the catalytic mechanism of inverting glycosyltransferases are, therefore, the identity of the base catalyst and the method used to facilitate departure of the (substituted) phosphate leaving group.

### 1.3.4.3. Inverting GT-A Glycosyltransferases (Clan I)

Excluding members of the sialyltransferase family GT42 that are the subject of the detailed analysis described in chapter 2, structural and mechanistic studies of nine representatives from six families of inverting glycosyltransferases with the GT-A fold have been undertaken. These studies have revealed several commonalities and differences amongst members of Clan I glycosyltransferases.

As mentioned, the family GT2 enzyme SpsA from *Bacillus subtilis* was the first glycosyltransferase experimentally shown to adopt the GT-A fold (Charnock and Davies, 1999). This family represents the largest and evolutionarily most ancient of inverting glycosyltransferases. Unfortunately, the natural donor and acceptor sugars of this enzyme are not known. However, fortuitously, a molecule of the cryoprotectant glycerol was found bound within the SpsA active site in a suitable position to mimic the natural

acceptor sugar and hydrogen bonded to the side chain carboxylate of Asp191, leading to the proposition that this residue played the role of base catalyst in the proposed direct displacement mechanism (Figure 1.5A)(Charnock and Davies, 1999). This notion was later supported by superpositioning of the SpsA structure on the solved GT-A structures of the enzymes lactose synthases (Gal-T1) from family GT7 (Gastinel et al., 1999; Ramakrishnan et al., 2002; Ramakrishnan and Qasba, 2001, 2002; Ramakrishnan et al., 2006; Ramakrishnan et al., 2001; Ramasamy et al., 2005), GnT-I from family GT13 (Gordon et al., 2006; Unligil et al., 2000), and GlcAT-I from family GT43 (Pedersen et al., 2002; Pedersen et al., 2000; Tarbouriech et al., 2001). These enzymes all have a conserved Asp residue within their active sites occupying a position equivalent to that of Asp191 in SpsA (Figure 1.5). Similarly, an analogously positioned conserved Asp or Glu residue has been observed in the subsequently obtained structures of the enzymes C2GnT-L from family GT14 (Pak et al., 2006), Mfng from family GT31 (Jinek et al., 2006), and GlcAT-P from family GT43 (Kakuda et al., 2004; Ohtsubo et al., 2000) (Figure 1.5). Structures with bound acceptor substrates indicate that these conserved carboxylates are within hydrogen bonding distance of the nucleophilic hydroxyl undergoing reaction, consistent with their roles as base catalysts (Figure 1.5). Further support comes from site-directed mutagenesis studies of the family GT2 enzyme ExoM, in which it was shown that mutation of Asp187 (structurally homologous to Asp191 in SpsA) abolished all *in vitro* glycosyltransferase activity (Garinot-Schneider et al., 2000).



**Figure 1.5.** Comparison of the active sites of several metal-dependent inverting glycosyltransferases from Clan I. A conserved side chain carboxylate is located in a near identical relative position within all active sites on the  $\beta$  face of the donor substrate and (when present) within hydrogen bonding distance of the nucleophilic hydroxyl of the acceptor sugar and plays the role of base catalyst in a direct displacement mechanism.  $Mn^{2+}$  cations are shown as magenta spheres.

The vast majority of enzymes from glycosyltransferase Clan I that have been subjected to biochemical analysis use an essential divalent cation (usually  $Mn^{2+}$  or  $Mg^{2+}$ ), coordinated by the so-called DXD motif, to facilitate departure of the nucleoside diphosphate leaving group by electrostatically stabilizing the developing negative charge. A notable exception to this strategy has recently been reported for the metal ion-independent  $\beta$ -1,6-GlcNAc transferase C2GnT-L from family GT14 (Pak et al., 2006). This enzyme uses the basic amino acids Arg378 and Lys401 to electrostatically stabilize the nucleoside diphosphate leaving group. This strategy is reminiscent of that used by metal ion-independent glycosyltransferases possessing a GT-B fold (as discussed below), indicative of a convergence of mechanism amongst these two superfamilies.

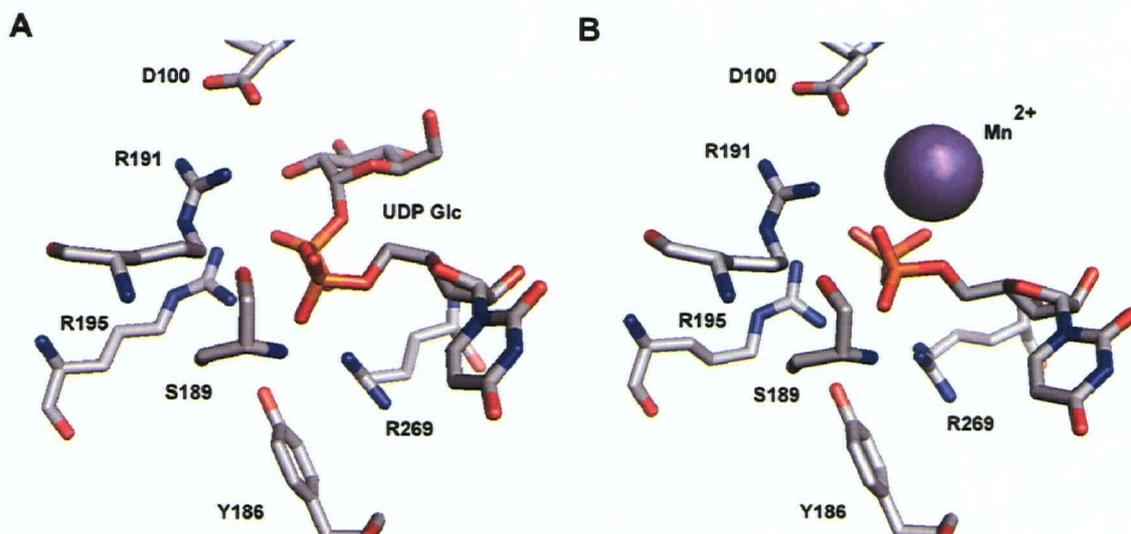
Theoretical support for a concerted  $S_N2$ -like displacement mechanism for inverting glycosyltransferases from Clan I is derived from a hybrid quantum mechanical/molecular mechanical (QM/MM) study of the  $\beta$ -1,2-GlcNAc transferase GnT-I (Kozmon and Tvaroska, 2006). Using the 1.8 Å resolution three-dimensional X-ray crystal structure of GnT-I with bound UDP GlcNAc (Unligil et al., 2000) and a knowledge of the ordered sequential Bi Bi kinetic mechanism of this enzyme (Nishikawa et al., 1988), a theoretical model of the Michaelis complex was generated and subjected to QM/MM analysis. The results supported a catalytic mechanism involving a concerted  $S_N2$ -type transition state involving near simultaneous nucleophilic attack, facilitated by proton transfer to the catalytic base (Asp291), and leaving group dissociation steps. An activation energy of  $\sim 19$  kcal/mol was estimated for the proposed transition state.

#### 1.3.4.4. Inverting GT-B Glycosyltransferases (Clan II)

To date, structural and mechanistic studies have been reported for 14 representatives from seven families of inverting glycosyltransferases that have been shown to possess a GT-B fold. Despite sharing homologous three-dimensional architectures, members of this superfamily seem to display a greater degree of diversity in the selected modes of catalysing glycosyl group transfer between, and in some cases even amongst, families compared to what has been observed for the GT-A superfamily. This greater diversity in mechanism is perhaps a reflection of the greater diversity of chemistries catalysed by this superfamily of enzymes, compared to the GT-A superfamily.

The prototype of this superfamily is the  $\beta$ -glucosyltransferase (BGT) from T4 bacteriophage, the first nucleoside diphosphate utilizing glycosyltransferase for which a three-dimensional X-ray crystal structure was obtained (Vrieling et al., 1994). This unique enzyme, the sole member of family GT63, transfers glucose from UDP Glc to the hydroxymethyl substituents of modified cytosine bases. It is proposed that this mechanism has evolved as a defence mechanism to prevent genome degradation from phage and host nucleases (Kornberg et al., 1961). In addition to the initially reported structures with bound metal ions and UDP present (Morera et al., 2001; Vrieling et al., 1994), a subsequent ternary complex X-ray crystal structure with bound UDP and DNA was obtained (Lariviere and Morera, 2002). The ternary complex structure revealed the ability of this glycosyltransferase to facilitate selective glycosylation by inducing the DNA to “flip out” in a fashion reminiscent of DNA glycosylases, methyltransferases and endonucleases. Analogous to what has been found with the GT-A superfamily,

mutagenesis studies revealed Asp100 to be a likely candidate for base catalyst. This is supported by the observation that co-crystallization of the wild-type enzyme with UDP Glc results in complete hydrolysis of the donor sugar resulting in the obtainment of UDP product complexes, while analogous co-crystallization procedures with the D100A mutant results in the observation of intact donor sugar within the active site (Lariviere et al., 2003). It should be noted that the D100A mutation prevents the formation of an inter-domain salt bridge with Arg191, preventing complete formation of the active site, which could also be the cause of the observed decrease in substrate turnover. A crystal structure with bound intact donor substrate was also obtained by a brief soaking of wild-type BGT crystals with UDP Glc (Figure 1.6A). Interestingly, this donor substrate complex had no metal cation bound within the active site despite being soaked with high concentrations of  $Mg^{2+}$  (Lariviere et al., 2003). Instead, positively charged side chains were found to neutralize the negative charges of the pyrophosphate group of the bound donor substrate (Figure 1.6A). In contrast, in the UDP product complex, obtained by co-crystallization with  $Mn^{2+}$  and UDP, an  $Mn^{2+}$  cation was found coordinating the pyrophosphate group and was located in the region occupied by the glucose moiety in the UDP Glc complexed structure (Figure 1.6B). This led the authors to propose that the divalent cation plays a role in facilitating product release and not necessarily in the cleavage of the glycosidic linkage of the donor substrate. This alternative mode of leaving group activation, compared to the GT-A superfamily, has also been proposed for several other GT-B enzymes (described below).



**Figure 1.6.** Active site of wild-type T4 bacteriophage BGT from family GT63. (A) Donor substrate complex with bound UDP Glc (pdb 1j39). Note the triad of positively charged Arg residues as well as side chain hydroxyls suitably positioned to stabilize developing charge on the UDP leaving group in the absence of a divalent cation. (B) Product complex with bound Mn<sup>2+</sup> cation partially occupying the glucose binding site (1nvk).

Glycosyltransferases from family GT1 adopt the GT-B fold and are responsible for the glycosylation of various organic structures such as terpenes, anthocyanins, cofactors, steroids, peptide antibiotics, and macrolides. The various products produced by this family of enzymes serve critical biological functions, making this one of the most intensely studied families of glycosyltransferases. Three family GT1 enzymes (GtfA (Mulichak et al., 2003); GtfB (Mulichak et al., 2001); and GtfD (Mulichak et al., 2004) involved in the biosynthesis of the peptide antibiotic vancomycin have been subjected to structural and biochemical analysis. The results of these studies suggest that these enzymes use Asp13 as the catalytic base and that leaving group departure is facilitated in a metal ion-independent fashion using a helix dipole and interactions with side chain

hydroxyl and imidazole groups to stabilize the developing negative charge. A second group of related enzymes from family GT1 have also been characterized, revealing interesting differences from the Gtf enzymes. These include the multifunctional terpene/flavonoid glycosyltransferase UGT71G1 (Shao et al., 2005); the flavonoid glucosyltransferase VvGTI (Offen et al., 2006); the macrolide glycosyltransferases OleD and OleI (Bolam et al., 2007); and the human drug metabolizing glucuronyltransferase UGT2B7 (Miley et al., 2007). Although these enzymes were found to use a metal ion-independent mechanism, analogous to that described above, to facilitate leaving group departure, the catalytic base was determined to be a side chain imidazole that interacts with an adjacent conserved side chain carboxylate group. These conclusions were based on comparisons to a three-dimensional X-ray crystal structure with bound acceptor substrate (Offen et al., 2006), in which case the putative side chain imidazole was found within hydrogen bonding distance of the reactive acceptor substrate hydroxyl, and confirmed by kinetic analysis of enzyme mutants.

Other inverting GT-B enzymes that have been subjected to structural and mechanistic analysis include the heptosyltransferase WaaC from family GT9 (Grizot et al., 2006); the fucosyltransferases FucT from family GT10 (Sun et al., 2007) and FUT8 from family GT23 (Ihara et al., 2007); and the GlcNAc transferase essential for bacterial cell wall synthesis known as MurG (Ha et al., 2000; Hu et al., 2003). These enzymes all appear to use metal ion-independent methods for stabilizing the departure of nucleoside diphosphate leaving groups analogous to that described above. In the cases of WaaC and FucT, structural and kinetic analysis of enzyme mutants supports the roles of the side

chain carboxylates of Asp13 and Glu95 (respectively) as base catalysts. In the cases of FUT8 and MurG, the identity of the base remains less clear.

A rather intriguing story is unfolding involving the multifunctional sialyltransferase PmST1 from family GT80. This metal ion-independent enzyme has been reported to possess four distinct enzymatic activities with differing pH optima consisting of: an  $\alpha$ -2,3-sialyltransferase activity with an optimal pH of 7.5-9.0; an  $\alpha$ -2,6-sialyltransferase activity with an optimal pH of 4.5-6.0; an  $\alpha$ -2,3-sialidase activity with an optimal pH of 5.0-5.5; and an  $\alpha$ -2,3-trans-sialidase activity with an optimal pH of 5.5-6.5 (Yu et al., 2005). It should be noted that the presence of CMP could be the cause of the observed sialidase and trans-sialidase activities, as this enzyme, like most other glycosyltransferases, presumably catalyses the hydrolysis of its donor sugar substrate. At the high concentrations of enzyme used in determining these activities, back reaction in the presence of CMP would lead to the formation of CMP NeuAc *in situ*, which could then be hydrolysed (leading to the observed sialidase activity) or used as a substrate in the catalysed transfer of sialic acid to an alternative acceptor (leading to the observed trans-sialidase activity). The three-dimensional X-ray crystal structure of this enzyme has been solved under conditions (pH 7.5) that favour the  $\alpha$ -2,3-sialyltransferase activity in apo and product complex form with bound CMP (Ni et al., 2006). Subsequent co-crystallization with CMP 3F NeuAc, followed by soaking with lactose acceptor, led to the production of a ternary complex structure (Ni et al., 2007). This revealed the presence of an Asp/His pair, reminiscent of several of the previously described family GT1 enzymes, whose side chain functional groups were suitably positioned for either residue to play the role of base catalyst. The D141A mutant was found to have a transferase activity 20,000-

fold lower than wild-type enzyme, compared to a 350- to 1000-fold lower activity measures for the H112A mutant. As such, in contrast to what has been suggested with the analogous family GT1 enzymes, Asp141 was proposed to play the role of base catalyst (Ni et al., 2007).

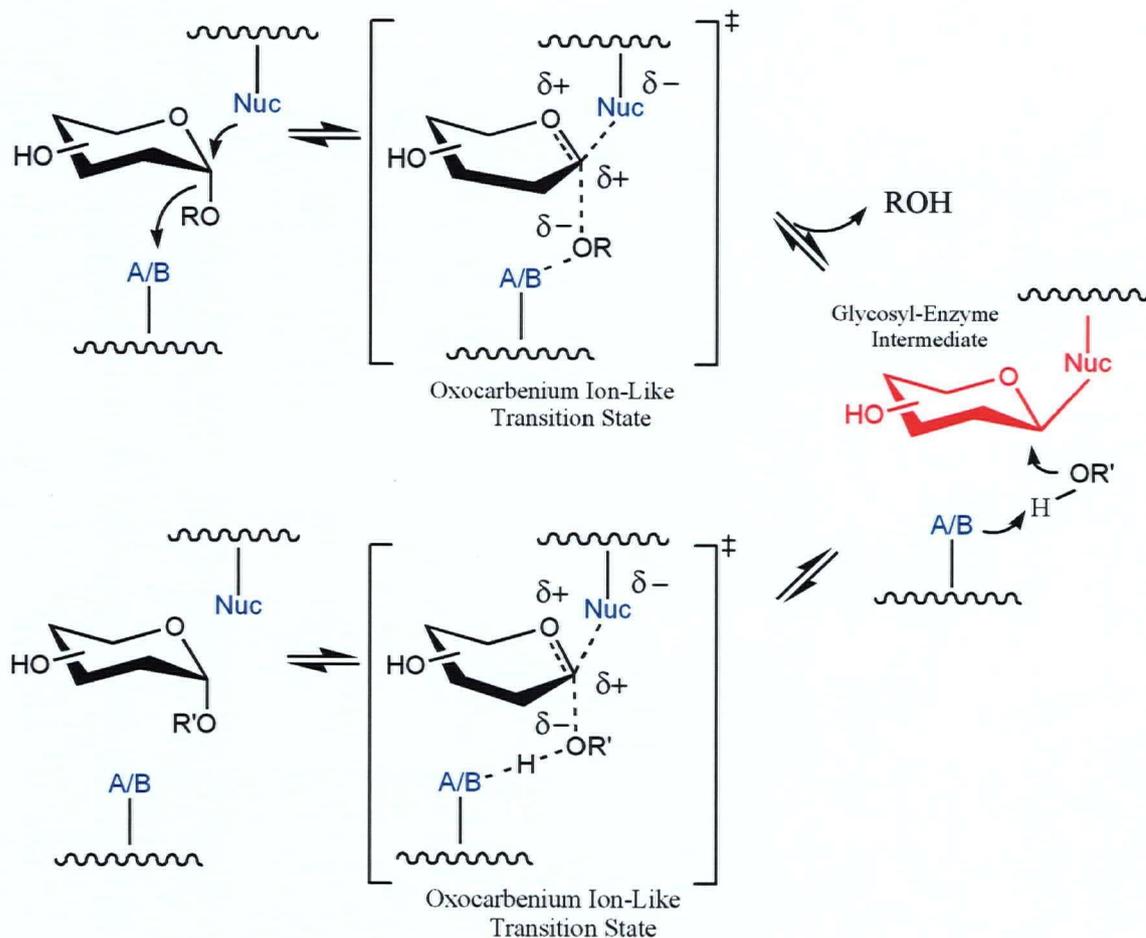
## **1.3.5. Retaining Glycosyltransferases**

### **1.3.5.1. Mechanism of Retaining Glycosidases**

With the notable exception of members from family GH 4 that use a recently described elimination mechanism involving transient remote oxidation (Yip et al., 2004) and retaining  $\beta$ -hexosaminidases that use an intramolecular nucleophile (reviewed in (Mark and James, 2002), the typical mechanism of retaining glycosidases is that of a double-displacement reaction involving a covalently bound glycosyl-enzyme intermediate species (Zechel and Withers, 2000). This mechanism was first put forward by Koshland (Koshland, 1953), who realized that, because inversion of configuration is a fundamental and universal property of bimolecular displacement at saturated carbon centres, glycosidases that gave sugar products with the same anomeric configuration as the substrate must catalyse the reaction using two distinct nucleophilic displacement steps involving an enzymatic catalytic nucleophile. An alternative  $S_N1$  mechanism, involving the formation of a discrete enzyme-stabilized oxocarbenium ion intermediate species that is shielded on one face by the enzyme, thereby preventing nucleophilic attack from the opposite face of the reaction centre and leading to complete retention of anomeric configuration in the product, was subsequently proposed for the retaining glycosidase hen egg white (HEW) lysozyme by Phillips (Phillips, 1966, 1967). It should be noted that

these two mechanisms are the extremes of a range of possible nucleophilic substitution reactions. The free energy of the intermediate, and therefore the associated transition states, for the Koshland mechanism would be lower than for those associated with the Phillips mechanism. The glycosidic linkage between two sugars is very stable. For example, the half-lives for the spontaneous hydrolysis of starch and cellulose at room temperature are in the range of 5 million years (Wolfenden et al., 1998). As such, the fact that glycosidases are observed to catalyse the hydrolysis of these materials with rate constants of up to  $1000 \text{ s}^{-1}$  would lead one to believe that this class of enzyme has evolved a mechanism that would involve an intermediate species with the lowest possible free energy, thereby facilitating as much of a decrease in the activation barrier as possible, in order to achieve such a formidable task with such proficiency. With some notable exceptions (Fersht, 1999), this logic, a large array of mechanistic data and relentless touting from its supporters, has led most in the field to accept the Koshland mechanism as the mechanism utilized by virtually all retaining glycosidases.

Analogous to the inverting glycosidases, a pair of carboxylates exist within the active site of retaining glycosidases, in this case 5 Å apart, with one acting as a general acid/base catalyst (A/B) while the other acts as a nucleophile (Nuc), the reaction again proceeding via oxocarbenium ion-like transition states (Figure 1.7). Amongst the glycosidases, notable exceptions as to the nature of the catalytic nucleophile exist amongst certain hexosaminidases in cases where an *N*-acetamido group acts as an intramolecular nucleophile (Mark and James, 2002) and a recently characterized trans-sialidase in which case a tyrosine plays the role (Watts et al., 2003).



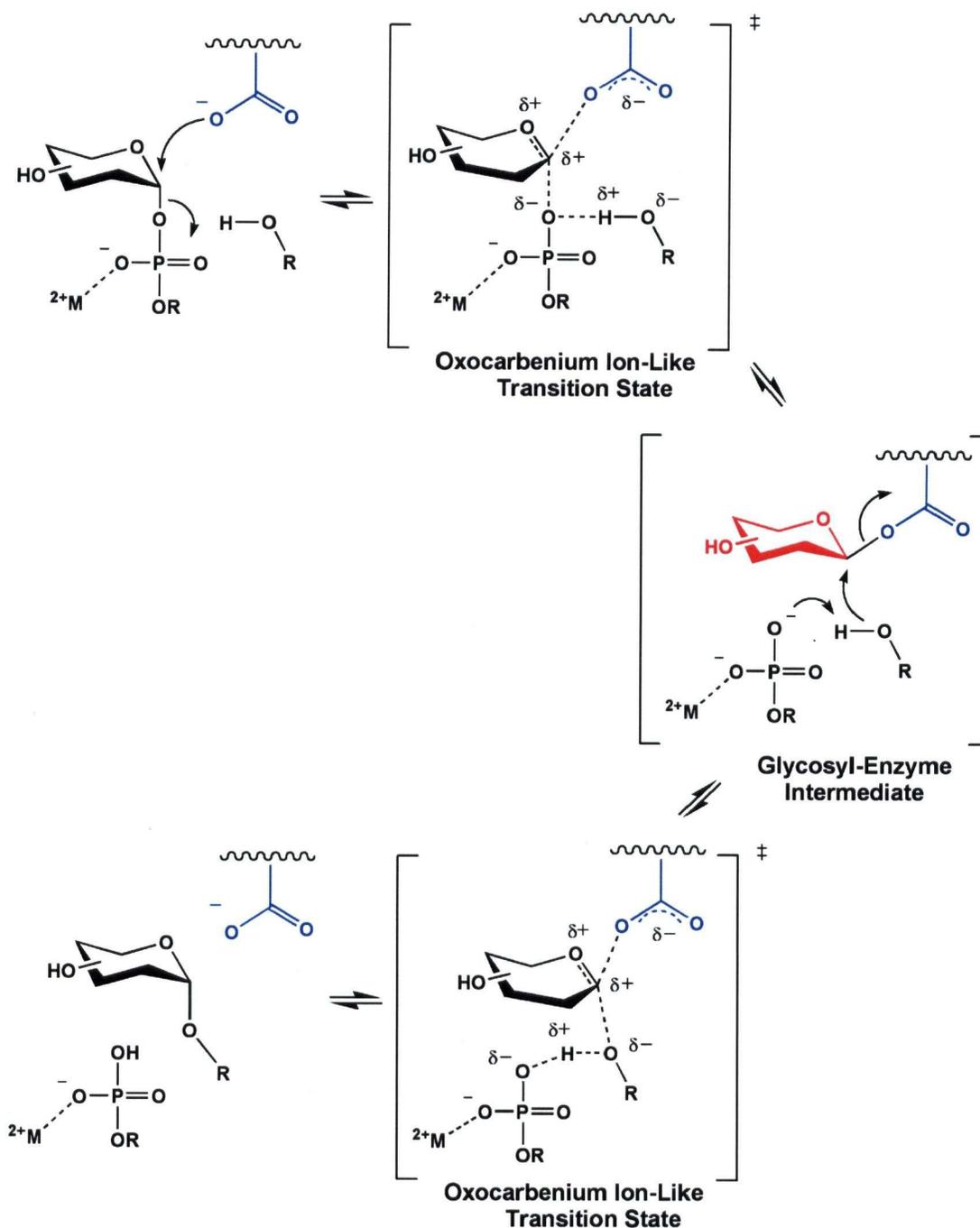
**Figure 1.7.** The double displacement mechanism established for retaining glycosidases proceeds via two oxocarbenium ion-like transition states with the intermediate formation of a discrete covalently bound glycosyl-enzyme species, resulting in overall retention of anomeric configuration. For glycosidases, R = a carbohydrate derivative and R'OH = H<sub>2</sub>O or phosphate (phosphorylases classified as glycosidases).

Evidence supporting an S<sub>N</sub>2-like mechanism involving enzymatic nucleophilic catalysis comes from the observation of normal secondary  $\alpha$  deuterium KIEs for *both* transition states (indicating rehybridization of C1 from sp<sup>3</sup> to sp<sup>2</sup>) (Kempton and Withers, 1992; Sinnott and Souchard, 1973; Vocadlo et al., 2002), as well as the observation of primary <sup>13</sup>C KIEs >1.01 consistent with S<sub>N</sub>2 pathways (Berti and Tanaka, 2002; Huang et al., 1997). Additionally, the covalent glycosyl-enzyme intermediates from

representatives of multiple families have been isolated and characterized using mass spectrometry and X-ray crystallography (Zechel and Withers, 2000). This includes the characterization of the intermediate of lysozyme, a result that contradicts the classical textbook mechanism involving the formation of a discrete oxocarbenium ion intermediate species (Vocadlo et al., 2001).

### **1.3.5.2. Mechanism of Retaining Glycosyltransferases**

Again by direct comparison to retaining glycosidases, the mechanism of retaining glycosyltransferases has been proposed to be that of a double displacement mechanism involving a covalently bound glycosyl-enzyme intermediate (Figure 1.8), demanding the existence of an appropriately positioned nucleophile within the active site (Davies, 1998). A divalent cation or suitably positioned positively charged side chains or helix dipoles would presumably play the role of the Lewis acids as was described above for the inverting glycosyltransferases. The leaving diphosphate group itself probably plays the role of a base catalyst activating the incoming acceptor hydroxyl group for nucleophilic attack. The role of substrate/product phosphates acting as base catalysts has been previously postulated in the mechanism of farnesyl diphosphate synthase (FPPS) from *E. coli* (Hosfield et al., 2004) and *T. Cruzi* (Gabelli et al., 2006).



**Figure 1.8.** Proposed double displacement mechanism for retaining glycosyltransferases. For glycosyltransferases, R = a nucleoside diphosphate (e.g. UDP, GDP), a lipid phosphate, or phosphate (phosphorylases classified as glycosyltransferases) and R'OH = an acceptor group (e.g. another sugar, a protein, or an antibiotic). In addition, for glycosyltransferases the role of base catalyst is likely played by the departing (substituted) phosphate leaving group.

### **1.3.5.3. Challenges in Studying the Mechanisms of Retaining Glycosyltransferases**

Mechanistic characterization of this class of enzymes has proven to be a challenging task. The conclusive identification of a catalytic nucleophile and observation of a true, kinetically and catalytically covalent intermediate has yet to be reported for any retaining transferase despite years of exhaustive studies using techniques that have been successfully applied to the characterization of retaining glycosidases. Although this may be interpreted as evidence against the double displacement mechanism, it could also be the result of the inapplicability of these techniques to the study of transferases due to inherent differences in the nature of the substrates being studied. The most successful approach used for the characterization of retaining glycosidases has involved the use of fluorinated substrate analogues (Withers and Aebersold, 1995). The introduction of an electronegative fluorine at either the 2 or 5 position of a pyranose ring inductively destabilizes the oxocarbenium ion-like transition states through which both steps of the double displacement reaction proceed and in some cases also removes key hydrogen-bonding interactions, resulting in a significant decrease in the rate of the overall reaction. By introducing a good leaving group (e.g. dinitrophenol or fluoride), the first step is “rescued”, resulting in the accumulation of the intermediate species with a significant lifetime that allows mass spectrometric and X-ray crystallographic characterization. Alternatively, or used in combination with the fluoro-sugar approach, removal of the acid/base catalyst of a retaining glycosidase by mutagenesis also leads to a decrease in the rates of both steps of the reaction and again, by using a substrate with a highly activated

leaving group, the glycosylation step can be “rescued” leading to the accumulation of the glycosyl-enzyme intermediate.

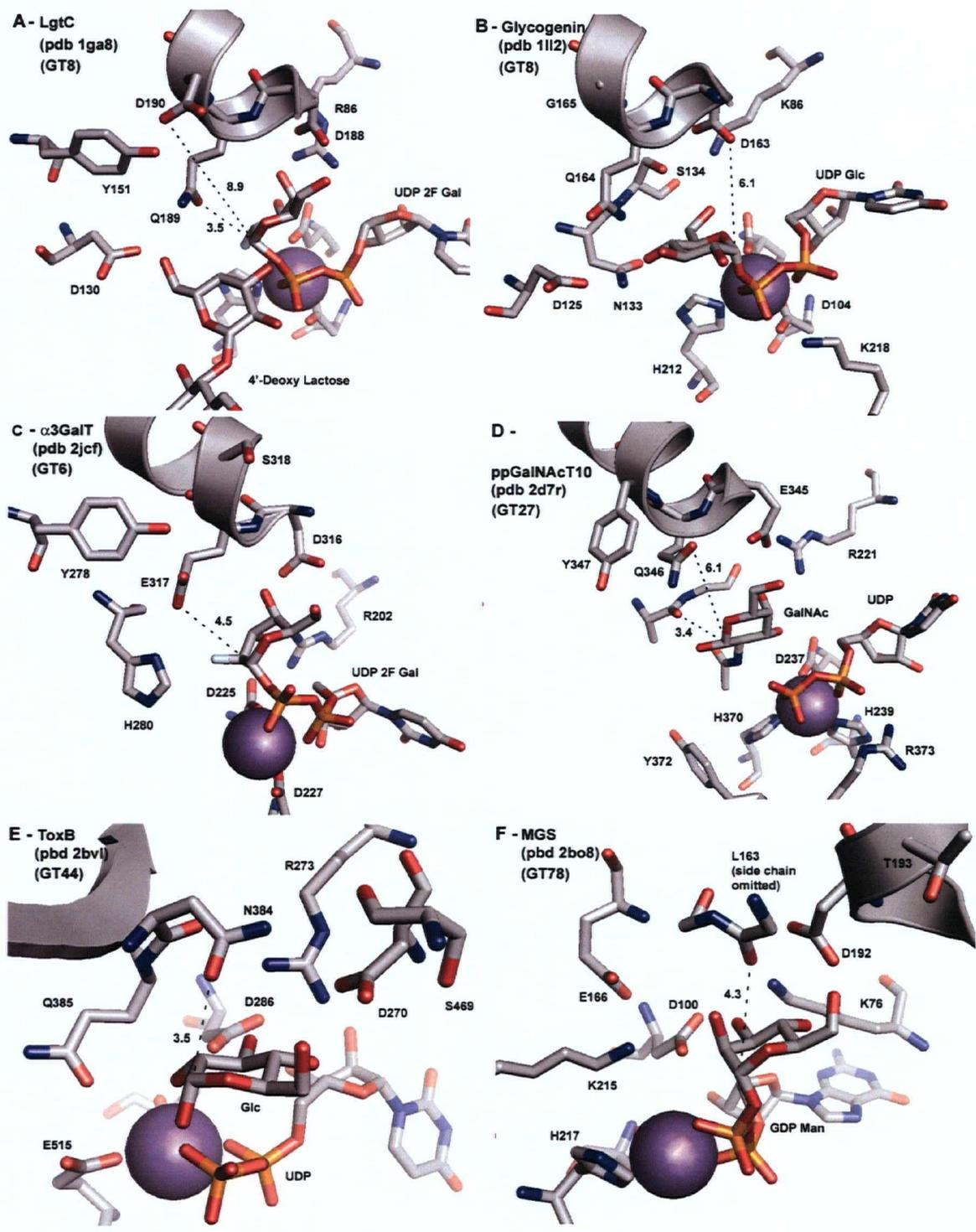
However, because of the strict requirement of the glycosyltransferases for their NDP leaving group, the relative leaving group ability cannot be manipulated, thus the relative rates of the glycosylation and deglycosylation steps cannot be altered. In order to prevent hydrolysis of metabolically expensive (substituted) phospho donor sugar substrates, it would be expected that glycosyltransferases would have evolved a mechanism in which the step involving cleavage of the bond between the sugar and its activated leaving group would be rate limiting thereby preventing the accumulation of an observable intermediate species which could be easily hydrolysed. In addition, because the leaving groups are themselves believed to play the role of base catalyst, reactions cannot be slowed down for this class of enzyme by mutagenesis of the protein catalyst. This inability to alter the relative rates of glycosylation versus deglycosylation has rendered the fluoro sugar approach ineffective in the trapping of intermediates on retaining transferases (Ly et al., 2002).

#### **1.3.5.4. Retaining GT-A Glycosyltransferases (Clan III)**

To date, structural and mechanistic studies of twelve representatives from seven families of retaining glycosyltransferases that have been shown to possess the GT-A fold have been undertaken. The results of these studies support generally utilized strategies amongst members of Clan III glycosyltransferases for facilitating leaving group departure and activating the nucleophilic group of the incoming acceptor. However, the observed structures have not revealed a conserved structural architecture on the  $\beta$ -face of the donor

sugar substrate in the region that would be expected to be occupied by the obligate enzymatic nucleophile of a proposed double displacement mechanism.

The family GT8 galactosyltransferase LgtC from *Neisseria meningitidis* was the first retaining nucleoside diphosphate-utilizing retaining glycosyltransferase for which a three-dimensional X-ray crystal structure was obtained (Persson et al., 2001). Using non-reactive UDP 2F Gal and 4'-deoxy lactose as donor and acceptor substrate analogues respectively, a well-resolved ternary complex structure was obtained (Figure 1.9A). As was seen with the inverting GT-A enzymes, this structure revealed the presence of a  $Mn^{2+}$  cation, coordinated within the active site by the carboxylate side chains of a DXD motif and the donor substrate diphosphate leaving group. It is presumed that the  $Mn^{2+}$  acts as Lewis acid to facilitate leaving group departure. In the region that would be occupied by the reactive axial 4'-hydroxyl of the acceptor substrate analogue, the only functional group that is suitably positioned to activate the incoming nucleophile is the leaving group oxygen of the diphosphate leaving group. This suggests that the departing diphosphate moiety acts as the base catalyst. Surprisingly, the only functional group that is suitably positioned within the active site on the  $\beta$ -face of the donor substrate to play the role of the catalytic nucleophile is that of the side chain amide of Gln189. The amide carbonyl oxygen of this residue is perfectly positioned 3.5 Å away and with an ideal trajectory for nucleophilic attack on the anomeric reaction centre. However, the results of mutagenesis studies were not consistent with an essential role for this residue as the catalytic nucleophile, since the corresponding alanine mutant retained 3% activity. This enzyme was the subject of the mechanistic studies described in section 3.2 and a detailed description of previous mechanistic studies are provided therein.



**Figure 1.9.** Comparison of the active sites of several retaining glycosyltransferases from Clan III. Descriptions of conserved structural features, and a lack thereof, are provided in the text.  $Mn^{2+}$  cations, which play the role of a Lewis acid catalyst that facilitates leaving group departure, are shown as magenta spheres. Distances are indicated in Angstroms.

Subsequently, the three-dimensional X-ray crystal structure of rabbit muscle glycogenin, another family GT8 enzyme, with intact UDP Glc donor substrate bound within the active site was reported (Gibbons et al., 2002). This enzyme catalyses self-glucosylation of a tyrosine residue in another monomer of the enzyme, in a process that is the initial step of glycogen biosynthesis, plus successive glycosylations of the glucosyl tyrosyl formed. Following the formation of an  $\alpha$ -1,4-linked oligosaccharide covalently bound to glycogenin, glycogen synthase and the branching enzyme complete the process of polysaccharide synthesis. The positioning of the tyrosyl hydroxyl that undergoes glucosylation, 21 Å away from the anomeric reaction centre on the surface of the same monomer, suggests that the initial glucosylation reaction occurs in an inter-subunit process with further glucose elaboration possibly occurring in subsequent intra-molecular reactions (Gibbons et al., 2002). This would be consistent with previous mechanistic studies. As was the case with LgtC, a  $Mn^{2+}$  cation was observed within the active site of glycogenin coordinated by the side chain carboxylates of a DXD motif positioned to act as a Lewis acid that activates the departing diphosphate leaving group (Figure 1.9B). However, unlike LgtC, the side chain most suitably positioned on the  $\beta$ -face of the donor substrate to act as the catalytic nucleophile is that of the carboxylate of Asp163, albeit at a distance of 6.1 Å from the reaction centre (Figure 1.9B). This conserved residue corresponds to Asp188 of LgtC, which is in close proximity to the positively charged side chain of Lys86. As will be discussed, this residue pairing constitutes the only conserved structural motif that can be identified on the  $\beta$ -face of the donor substrate binding sites of retaining glycosyltransferases. In LgtC the side chain carboxylate of Asp188 is within hydrogen bonding distance of the C6 and C4 hydroxyls of the galactose moiety of the

donor substrate and the positively charged side chain of Arg86 (Figure 1.9A). In the glycogenin structure, the side chain amide of Gln164, which corresponds to Gln189 in LgtC, is within hydrogen bonding distance of the C4 hydroxyl of the glucose donor substrate (Figure 1.9B). These observations indicate that either different donor sugar binding modes exist amongst enzymes of the same family and/or that significantly different ground state structures exist, indicating a high degree in structural plasticity for this class of enzyme, for which snapshots can be observed by crystallography.

Bovine  $\alpha$ 3GalT from family GT6 is another retaining galactosyltransferase possessing a GT-A fold that has been the subject of significant structural and mechanistic investigations. The initial crystal structure reported suggested the observation of a covalently bound  $\beta$ -galactosyl-enzyme intermediate but this was dogged by considerable disorder and limited resolution (Gastinel et al., 2001). Confidence in the interpretation of this putative covalently modified species is significantly compromised by the fact that a second donor galactose residue from a non-covalently bound UDP Gal species was also present. In addition, the electron density for the “covalently bound” galactose was weak. Higher resolution structures were subsequently reported of complexes of  $\alpha$ 3GalT with UDP (Boix et al., 2002), UDP and acceptors (Zhang et al., 2003) as well as the non-reactive donor substrate analogue UDP 2F Gal (Jamaluddin et al., 2007). The structure with bound UDP 2F Gal clearly reveals the side chain carboxylate of Glu317, structurally equivalent to Gln189 in LgtC, to be suitably positioned to play the role of the proposed catalytic nucleophile in a double displacement mechanism (Figure 1.9C). Interestingly the UDP 2F Gal is bound in an essentially identical “folded” conformation to that seen in LgtC. The conserved structural motif of the side chain carboxylate from residue Asp316

within hydrogen bonding distance of donor sugar hydroxyls and the positively charged side chain of residue Arg202 is also present (Figure 1.9C). Once again these structures all reveal the presence of a bound divalent cation within the active site suitably positioned to play the role of a Lewis acid that facilitates leaving group departure. In addition, a product complex with bound UDP contained a second  $Mn^{2+}$  bound within the active site in a way that suggests a role in facilitating product release (Jamaluddin et al., 2007). Bovine  $\alpha 3GalT$  is the subject of the mechanistic investigations described in section 3.4 and a discussion of previous mechanistic studies is provided therein.

The glycosyltransferases responsible for addition of the specificity-determining Gal/GalNAc moieties to the cell surface glycolipids and glycoproteins that constitute the ABO(H) blood groups are another group of family GT6 enzymes that have been subjected to intensive structural scrutiny (Watkins, 1957; Yamamoto et al., 1990). Glycosyltransferase A (GTA) specifically uses UDP GalNAc as a donor substrate to modify the 3-OH of the terminal galactose residue of the H-antigen (terminal Fuc- $\alpha$ (1,2)-Gal- $\beta$ -R acceptors) thereby creating the A blood group antigen. Conversely, glycosyltransferase B (GTB) specifically uses UDP Gal to modify the same hydroxyl of H-antigen to generate the B blood group antigen. Amazingly, these two enzymes differ from each other by only 4 out of 354 amino acids. The three-dimensional X-ray crystal structures of these two enzymes with both UDP and H-antigen bound revealed that distinction between the two donor substrates is achieved by the substitutions of residues Leu266 and Gly268 in GTA with the more bulky side chains of Met266 and Ala268 in GTB (Patenaude et al., 2002). Similarly, a crystal structure of the inactive mutant of this enzyme, expressed by individuals of the O blood group wherein terminal H-antigens are

not modified, revealed the ability of a single mutation, that changes Gly/Ala268 with the bulky side chain of Arg, to block donor substrate binding resulting in the ablation of catalytic activity (Lee et al., 2005). The role of intramolecular hydrogen bonding within acceptors (Nguyen et al., 2003) and the role of the two residues (Gly/Ser235 and Leu/Met266) (Letts et al., 2006) in the assembly of Type I and II H antigens by GTA and GTB has also been revealed by X-ray crystallographic analysis. Finally, the role of various missense mutations leading to the production of enzymes of weak dual specificity has been investigated (Hosseini-Maaf et al., 2007; Persson et al., 2007).

Family GT27 contains a series of UDP GalNAc:polypeptide  $\alpha$ -*N*-acetylgalactosaminyltransferases (ppGalNAcTs) that initiate the formation of mucin-type *O*-linked glycans by catalysing the transfer of GalNAc to serine or threonine residues on protein surfaces, thereby generating the Tn antigen (GalNAc- $\alpha$ -*O*-Ser/Thr). These enzymes are unique amongst the glycosyltransferases in that they possess C-terminal lectin domains. The majority of ppGalNAcTs, termed peptide transferases, are able to catalyse transfer to both unmodified peptides as well as to glycopeptides. Others, termed glycopeptide transferases, are only able to transfer to peptides that have been modified by an initial GalNAc residue (Ten Hagen et al., 2001; Ten Hagen et al., 1999). The three-dimensional X-ray crystal structures of murine ppGalNAcT-1 isozyme with only Mn<sup>2+</sup> bound (Fritz et al., 2004), human ppGalNAcT-10 isozyme in complex with either hydrolysed donor or a GalNAc-serine acceptor (Kubota et al., 2006) and human ppGalNAcT-2 isozyme with UDP and an acceptor peptide bound (Fritz et al., 2006) have been reported. These structures indicate that activation of the UDP leaving group is achieved in the usual fashion by the presence of a coordinated Mn<sup>2+</sup> cation within the

active site poised to act as a Lewis acid (Figure 1.9D). In addition, the conserved motif of a side chain carboxylate (Glu345 in ppGalNAcT-10 and Glu334 in ppGalNAcT-2) positioned within hydrogen bonding distance of donor substrate hydroxyls and a positively charged side chain (Arg221 in ppGalNAcT-10 and Arg208 in ppGalNAcT-2) is again observed (Figure 1.9D). The side chain amides of Gln346 of ppGalNAcT-10 and Asn335 of ppGalNAcT-2 are located in a structurally analogous position to that of Gln189 in LgtC and are therefore the most likely candidate to act as the catalytic nucleophile based simply on the available structural information (Figure 1.9D). These structural studies have also revealed how dynamic association between the lectin and catalytic domains, facilitated by a flexible tether, permits the formation of a range of binding modes for various glycopeptide acceptor substrates, thereby facilitating the production of the characteristic high density glycosylation patterns associated with mucins.

The catalytic domain of *Clostridium difficile* Toxin B (ToxB) is a retaining family GT44 glucosyltransferase possessing a GT-A fold that modifies host Rho proteins by glycosylating key threonine residues (Just et al., 1995). A three-dimensional X-ray crystal structure of the catalytic domain with bound  $Mn^{2+}$ , UDP and glucose has been reported (Reinert et al., 2005). This structure revealed the conserved retaining GT-A architectural features of a bound  $Mn^{2+}$  that acts as a Lewis acid activator, as well as a side chain carboxylate (Asp270) within hydrogen bonding distance of the donor sugar hydroxyls and a positively charged side chain (Arg273) (Figure 1.9E). The side chain amide of Asn384 is the most suitably positioned candidate for a putative catalytic nucleophile (Figure 1.9E).

Mannosylglycerate synthase (MGS) from *Rhodothermus marinus* is a family GT78 mannosyltransferase that is responsible for the synthesis of the stress protectant 2-*O*- $\alpha$ -D-mannosylglycerate. An X-ray crystal structure of MGS with bound  $Mn^{2+}$  and intact GDP Man bound within the active site also revealed the presence of a coordinated divalent cation Lewis acid activator and a side chain carboxylate (Asp192) within hydrogen bonding distance of donor sugar hydroxyls and a positively charged side chain (Lys72) (Figure 1.9F)(Flint et al., 2005). The main chain amide of Leu163 is the only functional group suitably positioned on the  $\beta$ -face of the anomeric reaction centre of the mannosyl donor substrate to act as a catalytic nucleophile (Figure 1.9F).

Three-dimensional X-ray crystal structures have been reported for the retaining enzymes Kre2 (Lobsanov et al., 2004), a yeast mannosyltransferase from family GT15, and EXTL2 (Pedersen et al., 2003), a mouse N-acetylhexosaminyltransferase involved in heparan biosynthesis from family GT64. Both of these enzymes were shown to possess a GT-A fold, however the absence of bound donor sugar substrates limits the utility of these structures in identifying candidate catalytic nucleophiles. Both reported structures had bound divalent cations coordinating the nucleoside diphosphate product. A ternary complex structure of EXTL2 with bound UDP and an acceptor substrate supports a role for the phosphate leaving group in acting as the base catalyst that activates the incoming nucleophile. Kre2 is the subject of the structural and mechanistic studies described in section 3.3.

A comparison of representative retaining glycosyltransferases, possessing the GT-A fold, from various families illuminates some conserved structural features within some regions of the active site and a lack thereof in other regions for this class of enzyme. The

mode of leaving group activation is similar to that used by the majority of inverting GT-A enzymes. A divalent cation, coordinated by the side chains of a conserved DXD motif and the side chain imidazole of a structurally conserved histidine residue, acts as a Lewis acid that facilitates leaving group departure. In contrast to the conserved features in the region surrounding the leaving group and incoming acceptor (the  $\alpha$ -face of the donor substrate), there appear to be very few conserved architectural features within the active site region that accommodates the  $\beta$ -face of the donor. The only strictly conserved feature observed on this face of the reaction centre is a side chain carboxylate (from an Asp/Glu residue typically situated on helix 6) that is within hydrogen bonding distance of the donor sugar hydroxyls and a positively charged side chain. The side chain of this Arg/Lys residue is in turn hydrogen bonded to a side chain carboxylate, derived from a residue of the DXD motif, which is in turn coordinated to the bound divalent cation. Interestingly, in the case of glycogenin, the side chain from this  $\beta$ -face residue (Asp163) is the most suitably situated group to act as the catalytic nucleophile in a double displacement mechanism, indicating how subtle differences in the mode of donor sugar binding can influence our interpretation of mechanism when based solely on structural information. In most cases, a side chain or main chain amide is most suitably positioned in an appropriate trajectory and at a reasonable distance from the anomeric reaction centre to play the role of the catalytic nucleophile in a putative double displacement mechanism. An exception to this observation is found with family GT6 enzymes that have a side chain carboxylate at this position. This apparent lack of conserved active site architecture amongst retaining GT-A glycosyltransferases, all of which appear to have diverged from a common ancestor, is in stark contrast to what has been observed for the analogous

retaining glycosidases. In those cases, for which strong evidence has been provided supporting a double displacement mechanism, highly conserved pairs of carboxylates separated by a defined distance are situated within the active site with one being clearly positioned to play the role of catalytic nucleophile. A vast multitude of observed three-dimensional folds all having the same disposition of catalytic active site carboxyl groups indicates a convergence in mechanism amongst unrelated retaining glycosidases. By comparison, if all retaining glycosyltransferases utilize the analogous double displacement mechanism, it would seem likely that during the course of divergent evolution the most stringently conserved active site feature would have been the relative positioning of the most important component of the catalytic machinery. Based on the available crystal structures, the conservation of this active site feature appears to be absent. This observation may very well suggest that retaining glycosyltransferases utilize a mechanism that differs from that of retaining glycosidases.

#### **1.3.5.5. Retaining GT-B Glycosyltransferases (Clan IV)**

To date, structural and mechanistic studies of eleven representatives from five families of retaining glycosyltransferases possessing the GT-B fold have been undertaken. As was the case with the inverting glycosyltransferases possessing a GT-B fold, these enzymes also use a metal ion-independent method for facilitating leaving group departure. Crystal structures with substrates or inhibitors bound within the donor sugar site reveal the conserved presence of two active site functional groups on the  $\beta$ -face of the donor substrate.

Members of family GT35 catalyse the phosphorolysis of  $\alpha(1,4)$ -linked glucans leading to the production of  $\alpha$ -D-glucose-1-phosphate (Glc1P), which is in turn isomerised to glucose-6-phosphate and fed into the glycolytic pathway. Although the equilibrium constant favours the synthetic direction of the reaction, under physiological conditions the reaction is driven in the disfavoured phosphorolysis direction by the high cellular levels of phosphate and the metabolic flux of glucose-1-phosphate into the glycolytic pathway. The enzymes of this family play an essential role in energy storage/mobilization in organisms ranging from bacteria to mammals. This family of enzymes has a noteworthy history of investigation, with a Nobel prize being awarded to Carl and Gerty Cori in 1947 for the discovery of rabbit muscle glycogen phosphorylase (Green et al., 1942). Given the key role that these enzymes play in metabolism, it is not surprising that the activities of the eukaryotic enzymes are found to be under strict allosteric regulation by covalent modification (serine phosphorylation) and by the binding of small molecule activators (e.g. AMP) and inhibitors (e.g. glucose and caffeine)(Madsen, 1986). In contrast, the activity of the bacterial enzymes is controlled at the level of expression (Raibaud and Schwartz, 1984). Because of their potential application as targets of therapeutics in the treatment of type II diabetes, the development of inhibitors of human liver glycogen phosphorylase is an active area of research (for example (Rath et al., 2000a).

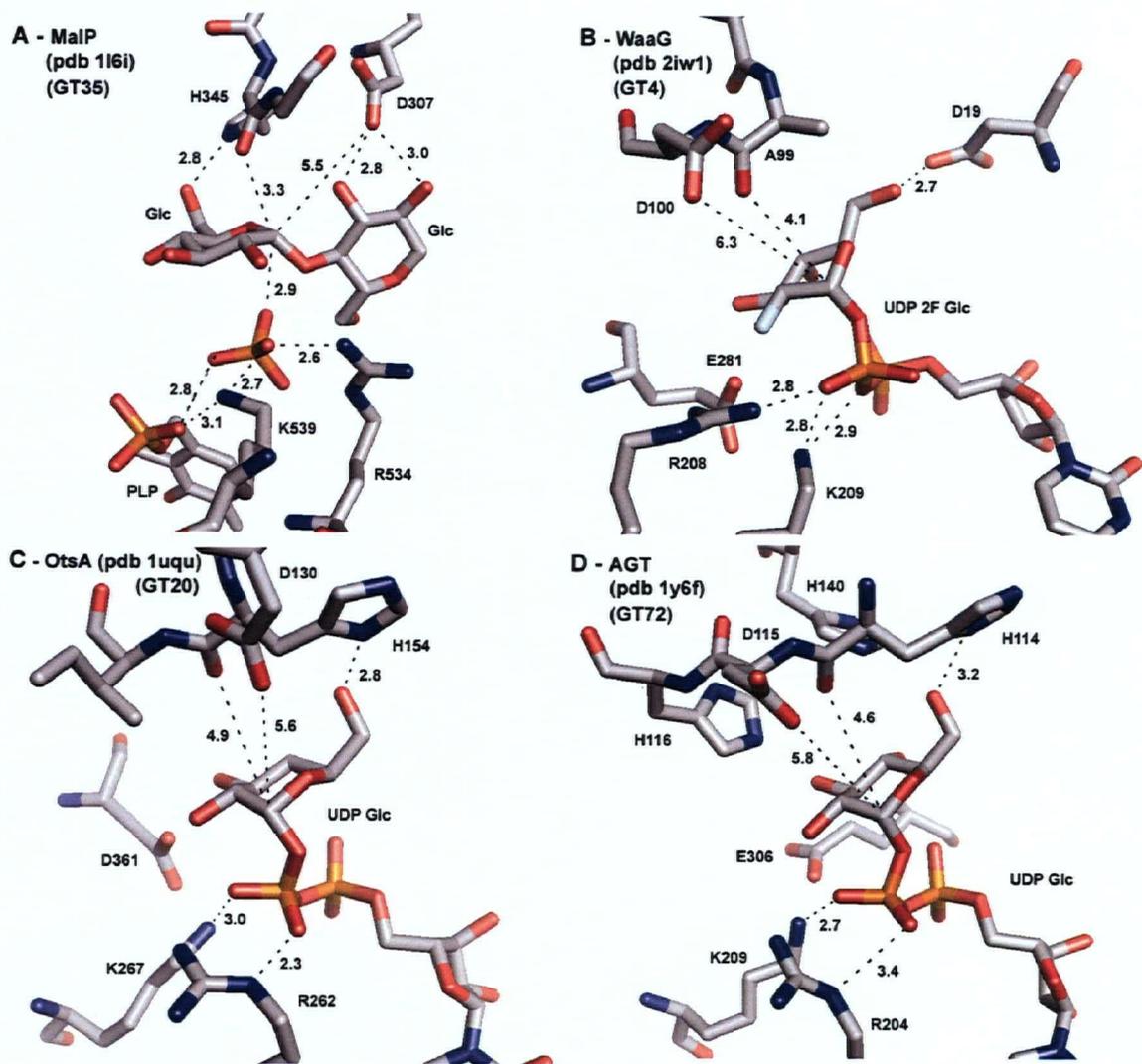
The most intensively studied member of family GT35 is that of rabbit muscle glycogen phosphorylase (rmGP), with over 30 structural investigations having been reported to date (for example see (Mitchell et al., 1996; Watson et al., 1994). However, despite a long history of exhaustive and laborious research with rmGP, a detailed

understanding of the catalytic mechanism remains elusive. Crystal structures are also available for bacterial maltodextrin phosphorylase (MalP) (Geremia et al., 2002; O'Reilly et al., 1999; O'Reilly et al., 1997; Watson et al., 1999); yeast glycogen phosphorylase (Lin et al., 1996); human muscle glycogen phosphorylase (Lukacs et al., 2006); and human liver glycogen phosphorylase (Ekstrom et al., 2002; Klabunde et al., 2005; Rath et al., 2000b; Wright et al., 2005).

The enzymes of this family are unique amongst the glycosyltransferases in that they contain a pyridoxal phosphate (PLP) group covalently bound within their active site via a Schiff base to a lysine residue. The exact role of PLP in the mechanism of catalysis has been a topic of heated debate. A proposed role as an acid catalyst was disfavoured by pH-dependence studies using PLP analogues with altered electronic demands at the phosphate position (Stirtan and Withers, 1996; Withers et al., 1982). The current view is that PLP acts as a surrogate for the nucleoside monophosphate portion of a nucleoside diphosphate during the evolution from an NDP sugar-dependent GT-B glycosyltransferase to a GT-B glycosyltransferase-like phosphorylase. This notion is supported by the fact that glycogen phosphorylase reconstituted with synthetic "pyridoxaldiphosphoglucose" is able to regio- and stereo-specifically transfer a glucose residue to the non-reducing end of a glycogen acceptor (Withers et al., 1981).

A ternary complex three-dimensional X-ray crystal structure of rmGP, with the transition state analogue nojirimycin and phosphate bound within the active site, revealed the main chain amide of His377 to be the functional group most suitably positioned to play the role of catalytic nucleophile (Mitchell et al., 1996). More recently, soaking of (Glc)<sub>4</sub> or (Glc)<sub>5</sub> malto-oligosaccharides into crystals of *E. coli* MalP with bound GlcP

resulted in the production of ternary complex structures with both transfer product and phosphate bound within the active site (Geremia et al., 2002)(Figure 1.10A). Like rmGP, the main chain amide of His345 (corresponding to His377 in rmGP) was found to be the most suitably positioned candidate to act as the catalytic nucleophile (Figure 1.10B). The formation of a hydrogen bond between the side chain imidazole of His 345 and the C6-hydroxyl of what would have been the glucose donor sugar results in the closure of a flexible loop region that makes up a significant portion of the maltose acceptor recognition site (Figure 1.10A). The only other suitable functional group positioned on the  $\beta$ -face of the donor substrate to act as a catalytic nucleophile is the side chain carboxylate of Asp307 (Figure 1.10A). However, this carboxylate is situated within hydrogen bonding distance of the C2- and C3-hydroxyls of the acceptor glucose residue and does therefore most likely serve to recognize and orient the incoming acceptor. A conformational change away from the position found in the observed structure would have to occur for this residue to play a nucleophilic role. The ternary complex structure with bound GlcP substrate is virtually identical to the product complex shown in Figure 1.9A with a TRIS molecule acting as an acceptor surrogate making similar interactions with the side chain of Asp307. This indicates that if such a conformational change were to occur it would have to happen during the course of catalysis. The bound phosphate product is located within hydrogen bonding distances of the positively charged side chains of Arg534 and Lys539, indicating a role for these residues in stabilizing the departing phosphate leaving group (Figure 1.10A).



**Figure 1.10.** Comparison of the active sites of several retaining glycosyltransferases from Clan IV. Distances are indicated in Angstroms.

Three-dimensional X-ray crystal structures with bound intact donor substrates/analogues are also available for retaining GT-B fold glycosyltransferases from families GT4, GT20 and GT 72. The structure of *E. coli*  $\alpha$ -(1,3) glycosyltransferase WaaG from family GT4, involved in lipopolysaccharide biosynthesis, was solved with bound non-reactive donor sugar substrate analogue UDP 2F Glc (Figure 1.10B) (Martinez-Fleites et al., 2006). Similarly, structures of *E. coli* OtsA from family GT20,

responsible for the biosynthesis of the very interesting stress response molecule  $\alpha,\alpha$ -1,1 trehalose-6-phosphate, were obtained with either UDP 2F Glc or UDP Glc (Figure 1.10C) bound (Gibson et al., 2004). Finally, ternary complex structures of AGT from family GT72, the retaining DNA modifying glucosyltransferase counterpart from T4 bacteriophage of inverting BGT (described in section 1.3.5.3.), were obtained with UDP Glc and various DNA acceptor substrates bound (Lariviere et al., 2005)(Figure 1.10D). Like BGT, AGT was shown to use a “base-flipping” mechanism to active DNA acceptors. Crystal structures of bacterial (Buschiazzo et al., 2004) and archaeal (Horcajada et al., 2006) glycogen phosphorylases from family GT5 are also available. However, the absence of the sugar moiety of the donor substrate limits their utility in identifying candidate nucleophilic residues.

A comparison of these structures, along with those of the rmGP and MalP ternary complexes, facilitates an initial identification of several conserved active site structural features amongst GT-B fold glycosyltransferases. On the  $\beta$ -face of the bound donor sugar substrate, a main chain amide is most suitably positioned, at an appropriate distance and trajectory, to act as the catalytic nucleophile in a double displacement mechanism. The conserved presence of hydrogen bonding partners, donated from main chain carbonyl or side chain amide groups, in proximity to the main chain NH group could help facilitate this catalytic role. The only other candidate for the role of nucleophile is that of the side chain carboxylate of a structurally conserved aspartate (Asp307, Asp100, Asp130, Asp115 in MalP, WaaG, OtsA and AGT respectively, Figure 1.10) situated  $\sim 6$  Å from the anomeric reaction centre. The product complex structure of MalP suggests that this residue is more likely involved in acceptor substrate recognition and orientation (Figure

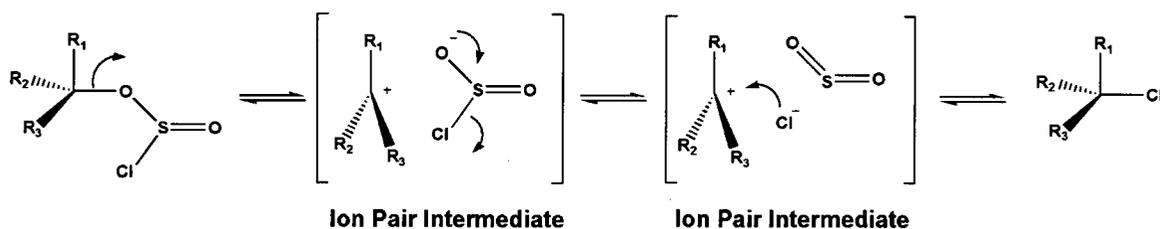
1.10A). The plausibility of a nucleophilic role for this residue is further limited by mutagenesis investigations with AGT, in which it was found that the D115A displayed ~10% residual transferase activity (Lariviere et al., 2005). Another conserved feature on the  $\beta$ -face of the donor is the formation of a hydrogen bond between the donor sugar C6-hydroxyl and, with the exception of WaaG in which case Asp19 is the bonding partner, the side chain imidazole of the His residue whose main chain amide is positioned to act as a nucleophile (His354/377, His154 and His114 in MalP/rmGP, OtsA and AGT respectively, Figure 1.10). As mentioned, the formation of this interaction is believed to cause the closure of a flexible loop leading to the complete formation of the acceptor recognition site. On the  $\alpha$ -face of the donor substrate, a structurally conserved Arg/Lys pair of positively charged side chains is present within hydrogen bonding distance of the phosphate leaving group, poised to replace the role of a divalent cation in GT-A fold enzymes, electrostatically stabilizing leaving group departure (Figure 1.10). The ternary complex structure of MalP with bound products suggests that the substrate/product phosphate group is most suitably positioned to play the role of the base catalyst that activates the functional group of the incoming acceptor (Figure 1.10A).

### 1.3.5.6. An Alternative “S<sub>N</sub>i-Like” Mechanism

Albeit rather rare, reactions involving a single nucleophilic displacement step can lead to the formation of a product in which the stereochemistry of the reaction centre is completely retained. This is a special case of the S<sub>N</sub>1 mechanism that involves the formation of discrete ion pair intermediates, with lifetimes longer than a molecular vibration, which can collapse to give back the starting material or a product. This unique form of S<sub>N</sub>1 reaction, termed S<sub>N</sub>i wherein the “i” indicates internal return, was first described to account for the observed products and the nature of the reaction involved with the decomposition of alkyl chlorosulphites (Cowdrey, 1937; Hughes, 1941). Support for this proposed mechanism was later provided by studies showing that the decompositions of alkyl chlorosulphites were in fact, under certain conditions, first order reactions that lead to the formation of alkyl halides with completely retained stereochemistry (Lewis and Boozer, 1952).

The observed stereochemical outcome is proposed to result from a multistep mechanism involving the formation of two discrete intermediate species of intimately interacting ion pairs (Figure 1.11). Multiple studies have since been reported providing support for the formation of ionic species in this S<sub>N</sub>1 process (Boozer and Lewis, 1953; Cram, 1953; Schreiner et al., 1993). The unique feature of the S<sub>N</sub>i mechanism that accounts for the exclusive retention of stereochemistry is the fact that the leaving group undergoes decomposition leading to the production of a nucleophile that is held as an ion pair on the same face as the leaving group. Decomposition of the initial intermediate species and attack by the formed nucleophile occurs at a rate that exceeds that of solvent and ion pair reorganisation required for nucleophilic attack from the back face. The

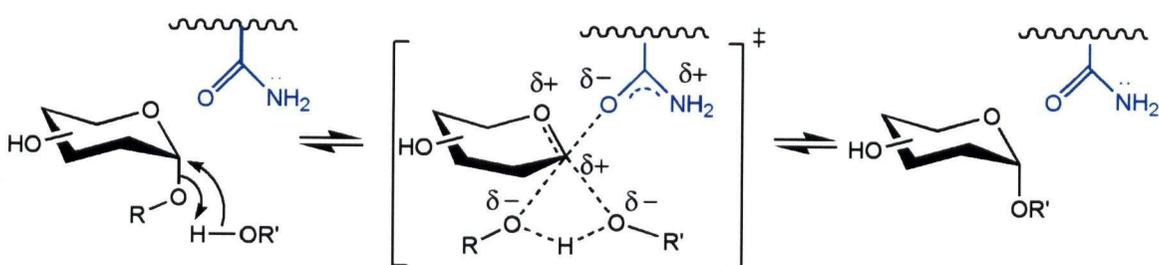
formation of intimate ion pairs makes the rate of attack by a nucleophile on the back face slow, as the effective concentration of the formed nucleophile would be low.



**Figure 1.11.** The  $S_{Ni}$  mechanism for the decomposition of alkyl chlorosulphites is a special case of an  $S_{N1}$  process that leads to complete retention of stereochemistry in the product.

In light of structural, mutagenesis and mechanistic results with LgtC, an alternative mechanism to the double displacement, termed  $S_{Ni}$ -like, was proposed (Persson et al., 2001). This was described as involving a single transition state in which attack by the incoming nucleophile of the acceptor and departure of the leaving group of the donor occur on the same face, contradicting one of the most classic tenets of physical organic chemistry (Figure 1.12). The single exploded four-centre transition state for such a process was proposed to be cyclic and late with highly developed oxocarbenium ion character. It must be noted that this four-centre transition state is formally forbidden by the rules of Woodward and Hoffman (Moss et al., 2007; Schreiner et al., 1994; Woodward and Hoffman, 1969). In such a mechanism, the enzyme acts as a scaffold that precisely orients substrates in close proximity, decreasing the energy of the transition state by stabilizing the oxocarbenium ion-like species and activating the leaving group. To prevent antibonding interactions and to allow for the development of density in the  $\sigma^*$  orbital between the anomeric carbon and the leaving group of the donor, departure of the

leaving group and front side attack must occur in an asynchronous fashion. Chemical precedent for this type of a mechanism comes from detailed conformational studies of the solvolysis of glucose derivatives in mixtures of ethanol and trifluoroethanol by Sinnott and Jencks (Sinnott and Jencks, 1980). They suggest that a sufficiently “open” transition state can allow for nucleophilic push on the reaction centre by the entering solvent molecule on either the opposite or the same face as the leaving group, thereby directly challenging the classical Ingold view of  $S_N2$  displacement reactions involving Walden inversion.



**Figure 1.12.** Proposed “ $S_{Ni}$ -like” mechanism for retaining glycosyltransferases and glycogen phosphorylase involving a direct “front-side  $S_N2$ -like” displacement proceeding through a highly dissociative oxocarbenium ion-like transition state resulting in retention of anomeric configuration.

The  $S_{Ni}$ -like mechanism had previously been proposed for glycogen phosphorylase (Klein et al., 1986) and has since been proposed for the structurally defined retaining transferases Kre2 (Lobsanov et al., 2004), ToxB (Reinert et al., 2005), Ext12 (Pedersen et al., 2003), Mgs (Flint et al., 2005), WaaG (Martinez-Fleites et al., 2006), OtsA (Gibson et al., 2002) and AGT (Sommer et al., 2004) based on the lack of an appropriately positioned nucleophile within their active sites.

## 1.4. Aims of the Thesis

Despite their undeniable biological significance and potential applicability in the synthesis of defined oligosaccharide structures, glycosyltransferases are a relatively poorly characterized class of enzymes that have yet to see deserved widespread practical applications: a situation that this thesis seeks to address. The work described in this thesis is divided into two parts. In the first part, the structural features and mechanisms of representative inverting and retaining glycosyltransferases are explored. Following these structure/function investigations, glycosyltransferases are subjected to engineering strategies with the goal of increasing their synthetic utility.

A detailed understanding of the chemical mechanisms by which glycosyltransferases function not only provides a rational basis for their engineering and application in both the development and synthesis of new classes of therapeutic agents, but also provides insight into the role of convergence in the natural evolution of enzyme function. Has nature converged on analogous strategies for catalysing reactions in opposite directions using different unrelated enzymes (i.e. glycoside hydrolysis by glycosidases and glycoside formation by glycosyltransferases) and/or what is the nature of this mechanistic continuum?

The kinetic and chemical mechanism of the inverting bifunctional family GT42 sialyltransferase Cst II will be explored guided by protein X-ray crystallography using the techniques of site directed mutagenesis, detailed Michaelis-Menten kinetic analysis, and chemical rescue. The kinetic mechanisms of glycosyltransferases are generally found to be ternary complex ordered Bi Bi, consistent with the generally observed feature that complete formation of an acceptor binding site requires the ordering of a flexible loop

region caused by the binding of a donor substrate within the nucleotide binding domain. The bifunctional nature of Cst II, being able to catalyse sialyl transfer reactions using two significantly different acceptor substrates, makes the study of its kinetic mechanism a particularly interesting pursuit that could provide insights into how this useful bifunctional activity is achieved. The general mechanism of both inverting glycosidases and glycosyltransferases is that of a direct displacement  $S_N2$ -like pathway. Such a mechanism requires an appropriately positioned base catalyst to activate the nucleophile of the incoming acceptor and functional groups that facilitate leaving group departure. The identities of these catalytic groups will be determined.

Following studies of this inverting enzyme, the chemical mechanisms of the retaining enzymes LgtC,  $\alpha3GalT$  and Kre2 (from families GT8, GT6 and GT15 respectively) will be investigated. The mechanism of retaining glycosyltransferases remains unclear. By analogy to the retaining glycosidase enzymes, a double displacement mechanism has been proposed. However, to date no definitive evidence in support of this proposition has been provided and an alternative  $S_{Ni}$ -like mechanism, with little chemical precedence, has become the generally quoted mechanism of choice. The potential for nucleophilic catalysis amongst this class of enzyme will be explored by altering the reactivity of an appropriately positioned enzymatic functional group within the active site of LgtC. The resulting engineered enzyme in combination with substrate and substrate analogues will be subjected to ESI-MS and protein X-ray crystallographic analysis. In addition,  $\alpha3GalT$ , an enzyme which seems to be one of the most likely to use a double displacement mechanism amongst characterized glycosyltransferases, will be subjected to ESI-MS analysis with the goal of directly observing the formation of a covalently bound

glycosyl-enzyme intermediate. Finally, Kre2 will be subjected to both ESI-MS analysis and protein X-ray crystallographic analysis with the goal of identifying potential enzymatic catalytic nucleophiles.

Stringent substrate specificity, the high cost of nucleotide sugar donor substrates, and the general lack of availability of soluble/active enzymes limit the synthetic utility of glycosyltransferases. To overcome these issues, glycosyltransferases will be subjected to engineering strategies involving either a modification of the protein catalyst or in a novel approach, modification of substrates. In addition to their potential practical applications, engineering results may provide information about mechanism. With the goal of developing a generally applicable screening strategy for directed evolution, the ability to display LgtC on the surface of M13 bacteriophage in the context of “water-in-oil” emulsions will be explored. In addition, a generally applicable rational engineering approach will be used to change the nucleotide specificity of LgtC.

## CHAPTER 2

# STRUCTURAL AND MECHANISTIC INVESTIGATIONS OF AN INVERTING GLYCOSYLTRANSFERASE

\* A version of portions of this chapter has been published:

Chiu, C.P.C., Watts, A.G., Lairson, L.L., Gilbert, M., Lim, D., Wakarchuk, W.W., Withers, S.G., and Strynadka, N.C.J. (2004) Structural Analysis of the Sialyltransferase Cst II from *Campylobacter jejuni* in Complex with a Substrate Analog. *Nature Structural and Molecular Biology*. **11**: 163.

## 2.1. Summary

Inverting glycosyltransferases are generally believed to catalyse group transfer via a direct displacement  $S_N2$ -like mechanism. The kinetic and catalytic mechanism of the bifunctional sialyltransferase Cst II was investigated. A three-dimensional X-ray crystal structure of the enzyme was obtained in complex with a non-reactive donor substrate analogue. The resulting crystal structure was used to identify putative catalytic residues within the enzyme active site. Site directed mutagenesis was used to investigate the catalytic importance of the identified residues, thereby facilitating a detailed account of the reaction mechanism.

## 2.2. Background

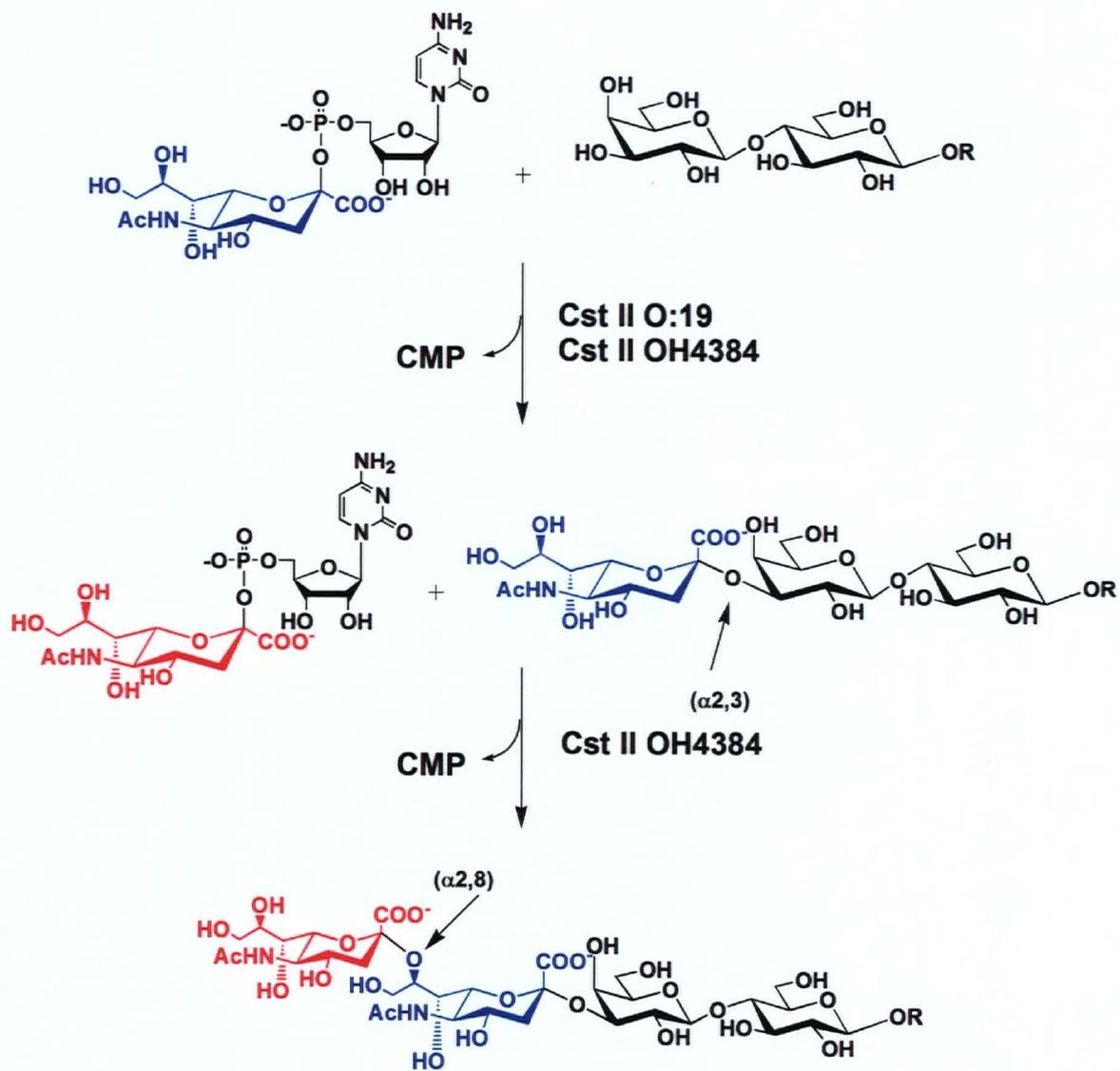
Glycoproteins, glycolipids and polysaccharides present on cell surfaces are known to play central roles in multiple biological processes including cell-cell recognition events controlled by receptor-ligand interactions. Many biologically active glycans are terminated by the 9-carbon sugar sialic acid, or *N*-acetyl-neuraminic acid (NeuAc), that bears a negative charge at physiological pH values and is added to glycoconjugates by a class of glycosyltransferases known as sialyltransferases. A common feature of malignant transformation and tumor progression is a change in surface glycosylation profile. These carbohydrate profile changes represent a significant portion of known tumor markers. A classic example is the observed increase in cell surface sialic acid content of many animal tumor types and a positive correlation of this change with metastatic potential (Dennis et al., 1982; Ogata et al., 1995; Yogeewaran and Salk,

1981). Sialic acid is related to cancer pathology for several reasons: sialylated cell surfaces can either prevent cell-cell interactions through non-specific charge repulsion or may facilitate specific cell adhesion events by binding to adhesion molecules of the selectin or siglec families; terminal sialic acids can conceal underlying oligosaccharide structures and prevent recognition by other lectin-like molecules; sialyltransferase regulatory elements can be the target of specific cell signaling pathways as is seen with the upregulation of ST6Gal I by ras oncogene (Dall'Olio and Chiricolo, 2001). Ganglioside glycosphingolipids constitute the major sialic acid-bearing glycans of vertebrate nerve cells, as well as being the major sialic acid-bearing glycoconjugates in the brain (Vyas and Schnaar, 2001). In the bacterial world, *Campylobacter jejuni*, a human mucosal pathogen, has been shown to express variable cell surface carbohydrate mimics of gangliosides associated with virulence (Guerry et al., 2002; Penner and Aspinnall, 1997). Molecular mimicry between *C. jejuni* LOS outer core structures and gangliosides is believed to act as a trigger for autoimmune mechanisms in the development of Guillain-Barré syndrome (Endtz et al., 2000). Compared to their eukaryotic counterparts, such bacterial enzymes are more convenient systems to study mechanistically since both expression and crystallization are more feasible. A detailed understanding of sialyltransferase structure and mechanism is clearly a topic of intense interest within the glycobiology community.

The genes responsible for the biosynthesis of the ganglioside mimics in *C. jejuni* have been identified (Gilbert et al., 2000; Gilbert et al., 2002). Cst II is a membrane-associated enzyme from family GT42 that possesses an inverting  $\alpha$ -2,3-sialyltransferase activity. A sialic acid moiety is transferred to terminal  $\beta$ -galactose-containing acceptor

substrates using CMP NeuAc as the donor with net inversion of anomeric configuration with respect to the donor substrate (Scheme 2.1). Interestingly, multiple versions of Cst II from different strains of *C. jejuni* were identified. Two of these enzymes, Cst II<sub>O:19</sub> and Cst II<sub>OH4384</sub> from serotypes O:19 and OH4384 respectively, possess very high sequence identity (97.3%) and yet the Cst II from serotype OH4384 was found to have bifunctional activity. Both enzymes possess the  $\alpha$ -2,3-sialyltransferase activity, however, the bifunctional enzyme is able to use the initially formed  $\alpha$ -2,3-linked sialic acid as an acceptor substrate in a second transfer step, as shown in Scheme 2.1 (Gilbert et al., 2000).

Despite the clear importance of sialylated glycans in normal and pathological physiologies, to date very few structural or mechanistic investigations of sialyltransferases from any organism have been undertaken. This lack of understanding results from the fact that members of this class of enzyme are generally membrane associated or membrane bound which limits the expression of soluble protein for characterization. A designed soluble and catalytically active construct of the bifunctional Cst II<sub>OH4384</sub>, in which 32 basic and hydrophobic residues are deleted from the predicted C-terminal membrane association domain have been removed (here after referred to simply as Cst II), was the subject of the detailed structural and mechanistic investigations described below.



**Scheme 2.1.** Monofunctional and bifunctional sialyltransferase activities of Cst II enzymes.

## 2.3. Determining the Kinetic Mechanism of the Bifunctional Inverting $\alpha$ -(2,3/8) Sialyltransferase Cst II<sub>OH4384</sub>

### 2.3.1. Preliminary Kinetic Analysis

The donor substrate used by sialyltransferases is CMP NeuAc. In order to determine kinetic parameters by monitoring the release of the CMP leaving group, the enzyme-coupled continuous assay typically used to characterize nucleoside diphosphate-utilizing glycosyltransferases had to be adapted. This was accomplished by addition of nucleoside monophosphate kinase (NMPK) and ATP to the typical coupling scheme. Using this coupled assay, glycosyltransferase activity is monitored as a function of CMP release via the consumption of NADH and the resulting decrease in absorbance at  $\lambda = 340$  nm (Scheme 2.2). Michaelis-Menten parameters were determined for the transferase and hydrolase activities of Cst II (Table 2.1).

**Table 2.1.** Apparent Michaelis-Menten kinetic parameters for the transferase and hydrolase activities of Cst II.

Substrate Varied	$k_{cat}$ ( $\text{min}^{-1}$ )	$K_m$ (mM)	$k_{cat}/K_m$ ( $\text{mM}^{-1} \text{min}^{-1}$ )
CMP NeuAc (hydrolase) <sup>a</sup>	2.0	0.50	4.0
CMP NeuAc (transferase) <sup>b</sup>	45	0.50	90
Lactose <sup>c</sup>	40	35	1.1
3'-Sialyl Lactose <sup>c</sup>	55	3.5	16

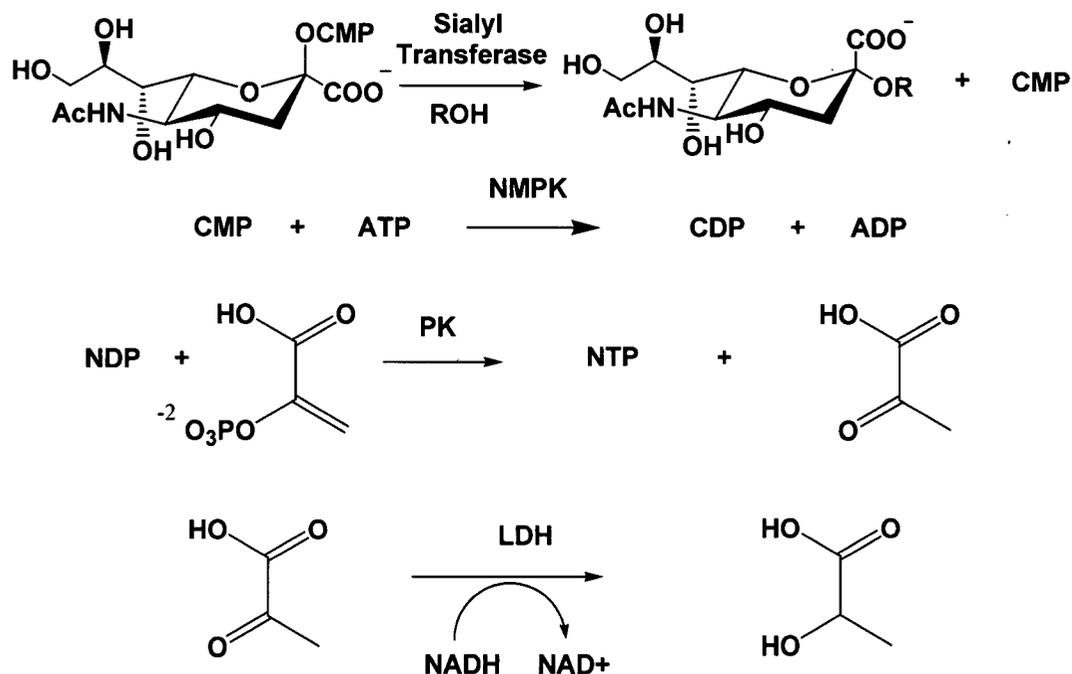
\* - error range in data is between 5-15%.

a - determined in the absence of acceptor substrate

b - determined at a constant lactose concentration of 160 mM.

c - determined at a constant CMP NeuAc concentration of 1 mM.

Careful attention was required during kinetic analysis to ensure that rates being determined were in fact those for the transferase activity of Cst II. A significant rate of spontaneous background hydrolysis of CMP NeuAc was observed and subtracted from all enzyme catalysed activities and determined the lower limit of detection for all assays. This significant rate of spontaneous donor substrate hydrolysis also limits the maximum concentration of CMP NeuAc that could be assayed. Above ~1 mM CMP NeuAc, the rate of spontaneous hydrolysis is too great to allow for measurement of sialyltransferase activity. As such, when  $K_m$  values are greater than 1 mM, errors are significantly larger because enzyme activity cannot be measured under truly saturating concentrations of donor substrate.



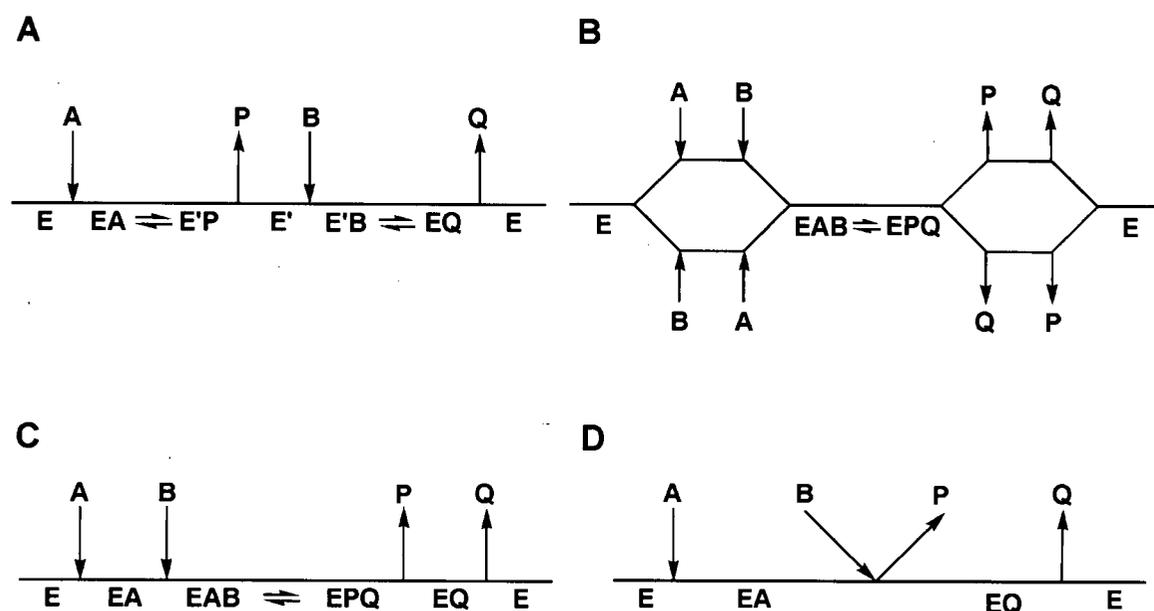
**Scheme 2.2.** Enzyme-coupled continuous assay used to measure the kinetic parameters of sialyltransferases.

In addition to catalysing the transfer of sialic acid to an acceptor sugar, Cst II was also found to catalyse the hydrolysis of CMP NeuAc at an appreciable rate, as determined by incubating Cst II with CMP NeuAc in the absence of acceptor (Table 2.1). To ensure that the observed hydrolytic activity was not due to a transfer of sialic acid from CMP NeuAc to an additional molecule of donor, a result that would not be completely unexpected considering the observed bifunctional acceptor specificity, TLC and mass spectroscopic analysis was performed and confirmed that no such transfer product was formed. The quantification of each of these three possible transformations of CMP NeuAc proved particularly important during the analysis of several Cst II mutants.

### **2.3.2. Potential Kinetic Mechanisms**

In both of the non-hydrolytic reactions catalysed by Cst II, two substrates are converted to two products. As such, Cst II is designated a Bi Bi enzyme according to the notation developed by Cleland for multisubstrate enzymes (Cleland, 1963). The mechanism of a multisubstrate enzyme is designated as either ping pong if the reaction proceeds with the release of one or more products prior to the association of all substrates or as ternary complex (or sequential) if all substrates bind to the enzyme prior to the formation of the first product. Sequential reactions are further designated as being either random if there is no obligatory order of substrate association and product release or as ordered if the substrates bind to and products are released from the enzyme in a defined order. A special case of the ordered mechanism is called the Theorell-Chance mechanism in which there is an obligatory order of substrate association and product release but without the accumulation of the ternary complex under the reaction conditions.

Schematic representations of these four basic types of kinetic mechanism are shown in Figure 2.1 using the notation developed by Cleland.



**Figure 2.1.** Cleland notation schematic representations of various Bi Bi kinetic mechanisms. (A) ping pong, (B) random sequential, (C) ordered sequential, and (D) Theorell-Chance.

A further distinction of a kinetic mechanism is made depending on the effect that substrate association and product release have on the overall observed rate of reaction. Reactions are termed rapid equilibrium, if the chemical steps are slower than all substrate association and product release steps, or steady state, if the rates of substrate association and product release contribute to the overall rate of reaction.

### 2.3.3. Distinguishing Between a Sequential or Ping Pong Mechanism

The rate equations for steady state ping pong and ordered Bi Bi mechanisms are shown in double reciprocal form in equations 2.1 and 2.2, respectively. In the absence of products, the equation for the Theorell-Chance “hit and run” Bi Bi mechanism is identical to that of the steady state ordered Bi Bi mechanism.

$$\frac{1}{v_o} = \frac{K_m^A}{V_{\max}[A]} + \frac{K_m^B}{V_{\max}[B]} + \frac{1}{V_{\max}} \quad (2.1)$$

$$\frac{1}{v_o} = \frac{1}{V_{\max}} + \frac{K_m^A}{V_{\max}[A]} + \frac{K_m^B}{V_{\max}[B]} + \frac{K_S^A K_m^B}{V_{\max}[A][B]} \quad (2.2)$$

$$\frac{1}{v_o} = \frac{1}{V_{\max}} + \frac{K_S^A K_m^B}{V_{\max} K_S^B [A]} + \frac{K_m^B}{V_{\max}[B]} + \frac{K_S^A K_m^B}{V_{\max}[A][B]} \quad (2.3)$$

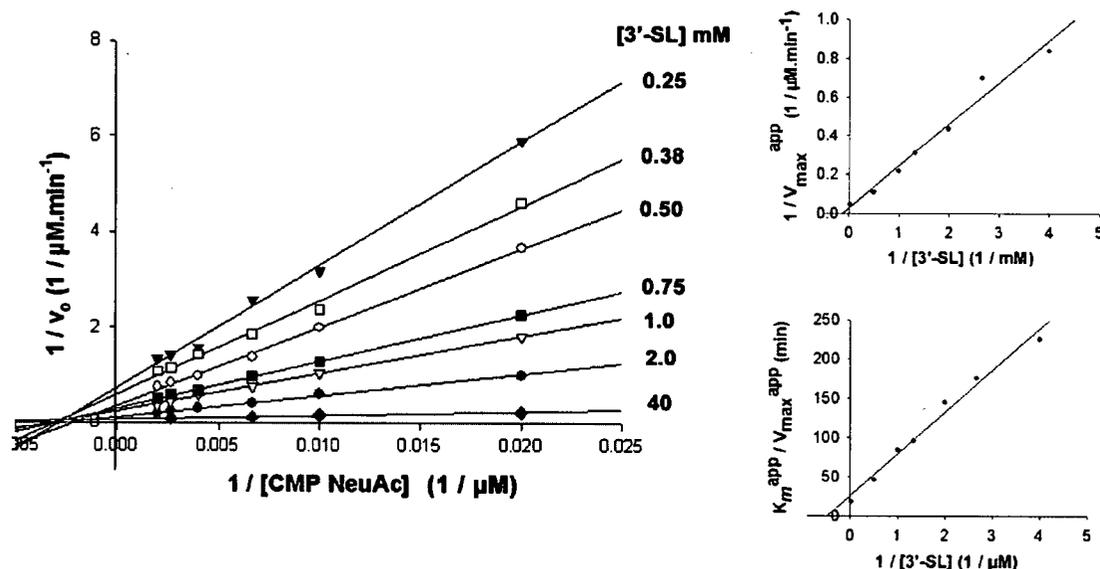
The steady state random Bi Bi mechanism is exquisitely complicated with 37 additional terms in the denominator compared to the analogous ordered mechanism (Segel, 1975). In many cases for random Bi Bi mechanisms, both substrates and both products are found to be in rapid and independent equilibrium with the enzyme, making the interconversion of EAB to EPQ rate limiting. In this special case, the mechanism is called rapid equilibrium random Bi Bi and the rate equation reduces to that shown in equation 2.3 in double reciprocal form.

Inspection of the above equations indicates that a ping pong mechanism can be readily distinguished from a sequential mechanism by plotting  $1/v_o$  versus  $1/[A]$  at various constant concentrations of substrate B. For a ping pong mechanism, such a plot will yield a series of straight lines for which the slopes are equal to  $(K_m^A / V_{max})$  and the intercepts are equal to  $(K_m^B / V_{max}[B]) + (1 / V_{max})$ . Because the slope of each of the lines is not dependent on the concentration of substrate B, the resulting lines will be a family of parallel lines. Such a series of parallel lines is diagnostic of a ping pong Bi Bi mechanism. In contrast, inspection of equations 2.2 and 2.3 indicates that a plot of  $1/v_o$  versus  $1/[A]$  at various constant concentrations of substrate B will yield families of straight lines for which the slope of each line is dependent on the concentration of substrate B. Double reciprocal plots for sequential mechanisms yield a family of straight lines that intersect to the left of the  $1/v_o$  axis.

To distinguish between a ping pong and a sequential mechanism, initial rates were measured for Cst II with increasing concentrations of CMP NeuAc at various constant concentrations of 3'-sialyl lactose using the enzyme-coupled continuous assay (Figure 2.2). The resulting family of straight lines is clearly not parallel and converges to the left of the  $1/v_o$  axis and approximately on the  $1/[CMP\ NeuAc]$  axis. A ping pong mechanism for Cst II can therefore be ruled out. For a Bi Bi mechanism, this data alone cannot be used to distinguish between the various types of sequential mechanisms.

It is not surprising that Cst II does not follow a ping pong mechanism. As it is an inverting enzyme, such a mechanism would require the intermediate formation of a stabilized oxocarbenium ion species and release of CMP prior to association and attack

by the incoming nucleophilic acceptor. Indeed, under standard aqueous conditions, such a mechanism would result in minimal transfer to a sugar acceptor.



**Figure 2.2.** Double reciprocal and secondary plots of initial rate data with respect to CMP NeuAc at various constant concentrations of  $3'$ -sialyl lactose ( $3'$ -SL).

### 2.3.4. Product Inhibition Studies with Cst II

Product inhibition studies can be used to distinguish whether substrates bind to and products are released from a sequential Bi Bi enzyme in an ordered or random fashion (Cornish-Bowden, 2004; Fromm, 1975; Segel, 1975). For ordered reactions, the patterns of product inhibition can also identify the order of substrate association and product release. The product inhibition patterns for several sequential Bi Bi mechanisms are shown in Table 2.2. For a more complete list of inhibition patterns the reader is directed to the detailed analysis of Segel (Segel, 1975).

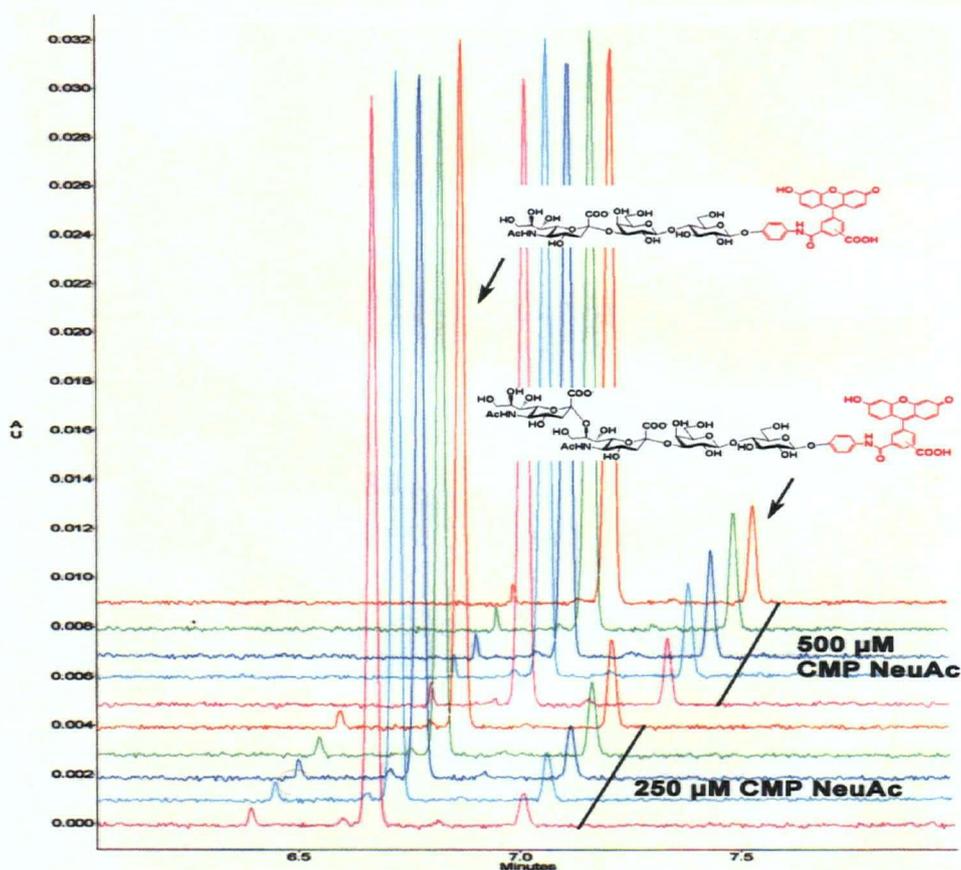
**Table 2.2.** Expected product inhibition patterns for select Bi Bi enzyme mechanisms (Segel, 1975).

Mechanism	Product Inhibitor	Varying A		Varying B	
		Unsat B	Sat B	Unsat A	Sat A
Ping Pong Bi Bi	P	Mixed	-	Comp	Comp
	Q	Comp	Comp	Mixed	-
Steady State Ordered Bi Bi	P	Mixed	Uncomp	Mixed	Mixed
	Q	Comp	Comp	Mixed	-
Steady State Iso Ordered Bi Bi	P	Mixed	Uncomp	Mixed	Mixed
	Q	<u>Mixed</u>	Mixed	<u>Mixed</u>	Uncomp
Steady State Theorell-Chance	P	Mixed	-	Comp	Comp
	Q	Comp	Comp	Mixed	-
Rapid Equilibrium Random Bi Bi	P	Comp	-	Comp	-
	Q	Comp	-	Comp	-
Steady State Random Bi Bi	P	Mixed	Mixed	Mixed	Mixed
	Q	<u>Mixed</u>	Mixed	<u>Mixed</u>	Mixed

Unsat - initial rates measured at a constant unsaturating concentration of substrate  
 Sat - initial rates measured at a constant saturating concentration of substrate  
 Comp - Competitive inhibition  
 Uncomp - Uncompetitive inhibition  
 Mixed - Mixed inhibition

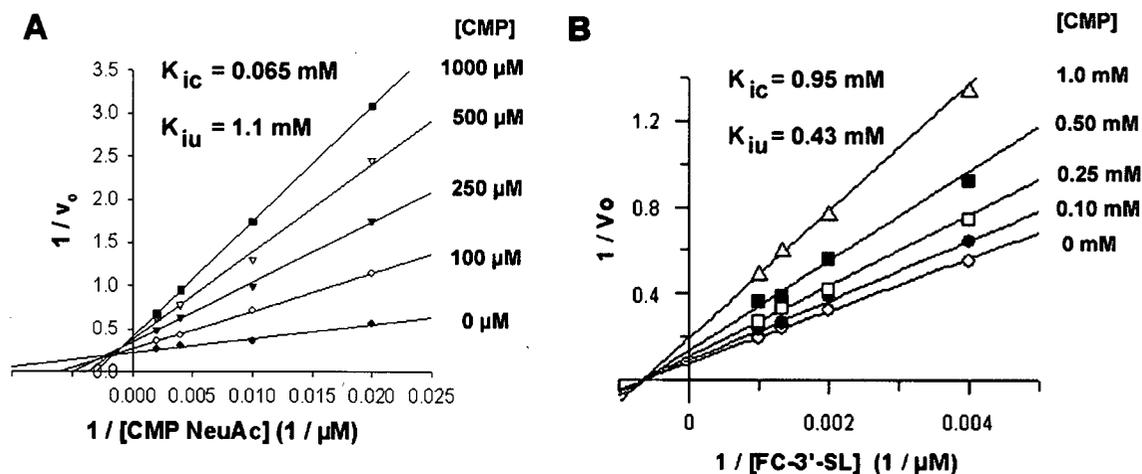
In order to determine patterns of inhibition with respect to the product CMP, because CMP release is the mode of detecting transferase activity in the enzyme-coupled continuous assay, an alternative assay was required. A stopped capillary electrophoresis (CE) based assay was established in which product formation is detected using fluorescein 3'-sialyl lactose conjugate (FCHASE 3'-SL) as the acceptor substrate. Representative CE traces for product formation at various concentrations of CMP NeuAc and CMP are shown in Figure 2.3. Because the presence of the fluorescein substituent limits the aqueous solubility of the conjugated acceptor, initial rate measurements could

only be obtained under non-saturating concentrations of the acceptor substrate. In addition, because of the instability of CMP NeuAc and resulting difficulty in obtaining donor substrate free of contaminating CMP (~10% contamination following HPLC purification), initial rates could also only be obtained under non saturating concentrations of the donor substrate.



**Figure 2.3.** Representative CE data used to monitor Cst II sialyltransferase activity. Product formation was determined based on percent conversion of FCHASE SL acceptor substrate using the peak areas of starting material and product. Initial rates were determined using 24  $\mu\text{g/mL}$  Cst II, various constant concentrations of CMP (0, 100, 250, 500, and 1000  $\mu\text{M}$ ), 500  $\mu\text{M}$  FCHASE SL, and varied concentrations of CMP NeuAc (50, 100, 250, and 500  $\mu\text{M}$ ).

Using CMP as a product inhibitor of the Cst II reaction, mixed inhibition patterns were observed with respect to both the donor substrate CMP NeuAc (Figure 2.4A) and the acceptor substrate FCHASE 3'-SL (Figure 2.4B) when the non varied substrate was held at a fixed non saturating concentration. Competitive ( $K_{ic}$ ) and uncompetitive ( $K_{iu}$ ) inhibition constants were determined by direct fit of the data to the equation for mixed inhibition using the program Grafit 4.0 and are shown on the double reciprocal plots. Somewhat surprisingly, these modes of inhibition are not consistent with either of the expected steady state ordered or Theorell-Chance Bi Bi mechanisms (Table 2.1). Mixed product inhibition patterns for both substrates in the presence of fixed non saturating concentrations of the other substrate are consistent with either a steady state random or an iso ordered Bi Bi mechanism.

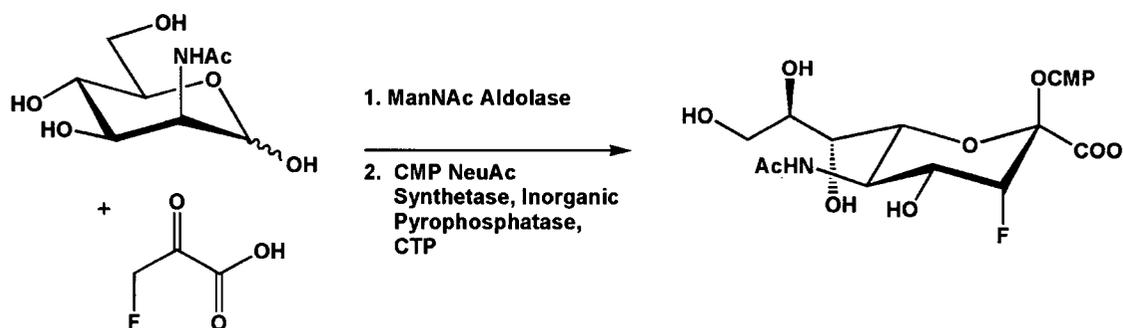


**Figure 2.4.** Inhibition of Cst II catalysed sialyl transfer to FCHASE 3'-SL by the product CMP with respect to donor (A) and acceptor (B) substrates.

### 2.3.5. Synthesis and Inhibition Studies of a Dead-End Donor Substrate Analogue

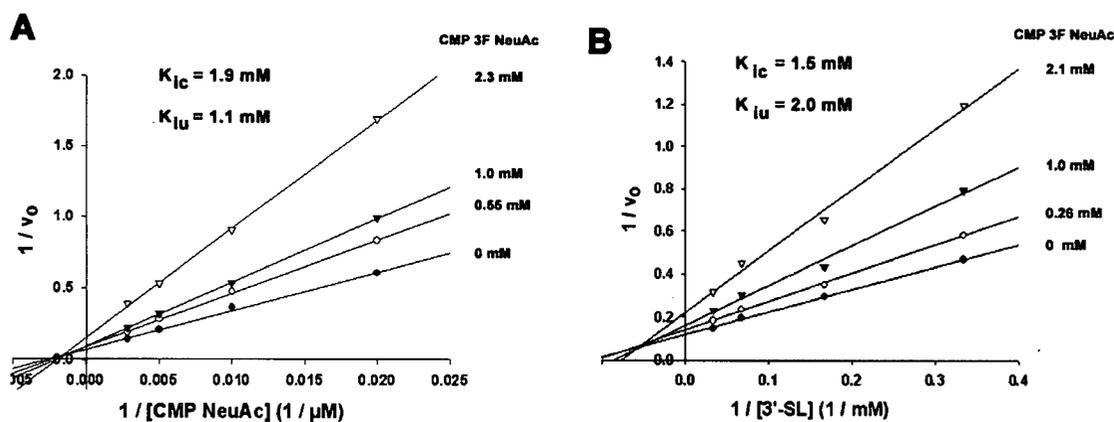
Inhibition studies using dead end substrate analogues can be used to distinguish between the mechanisms of bireactant enzyme systems (Piszkiwicz, 1977; Segel, 1975). To obtain further insight into the kinetic mechanism of Cst II, the dead end donor substrate analogue CMP 3F NeuAc was synthesized and subjected to detailed kinetic characterization.

As described in chapter 1, introduction of the highly electronegative fluorine atom adjacent to the anomeric reaction centre inductively destabilizes oxocarbenium ion-like transition states through which glycosyl group transfer reactions are thought to proceed. Indeed, in the case of glycosyltransferases, NDP 2-fluoro donor sugar analogues have been found to act as competitive inhibitors with respect to the donor substrate and do not act as slow substrates (Ly et al., 2002; Murray et al., 1997). In the case of Cst II, the desired analogue is CMP 3F NeuAc and facile access was obtained in two steps using the enzymes ManNAc aldolase and CMP NeuAc synthetase (Scheme 2.3).



**Scheme 2.3.** Enzymatic synthesis of the dead end donor substrate analogue CMP 3F NeuAc.

Due to limitations in the quantity of available inhibitor at the time, using the method of Dixon, CMP 3F NeuAc was initially reported to act as a competitive inhibitor of Cst II with respect to CMP NeuAc (Chiu et al., 2004). However, subsequent more rigorous analysis clearly indicates that using CMP 3F NeuAc as a dead end substrate analogue leads to mixed inhibition patterns with respect to both CMP NeuAc (Figure 2.5A) and 3'-sialyl lactose (Figure 2.5B) when the non-varied substrate is held at a fixed non-saturating concentration. Competitive ( $K_{ic}$ ) and uncompetitive ( $K_{iu}$ ) inhibition constants were determined by direct fit of the data to the equation for mixed inhibition using the program Grafit 4.0 and are shown on the double reciprocal plots.



**Figure 2.5.** Inhibition of Cst II catalysed sialyl transfer to 3'-SL by the dead end substrate analogue CMP 3F NeuAc with respect to donor (A) and acceptor (B) substrates.

As was observed with CMP product inhibition studies, these patterns of inhibition are not consistent with either the steady state ordered or rapid equilibrium random Bi Bi kinetic mechanisms. For both of these mechanisms, competitive inhibition behaviour would be expected for CMP 3F NeuAc with respect to CMP NeuAc (Piszkiwicz, 1977). However, the observed modes of inhibition for CMP 3F NeuAc are understandable in

light of the bifunctional nature of Cst II. An ability to associate with two distinct stable forms of the enzyme would make CMP 3F NeuAc a mixed inhibitor with respect to CMP NeuAc. This mode of inhibition is therefore consistent with a kinetic mechanism that is more complex than that of the steady state ordered Bi Bi mechanism and, although by no means conclusive in itself, supports an iso ordered Bi Bi mechanism.

### **2.3.6. Conclusions on the Kinetic Mechanism of Cst II**

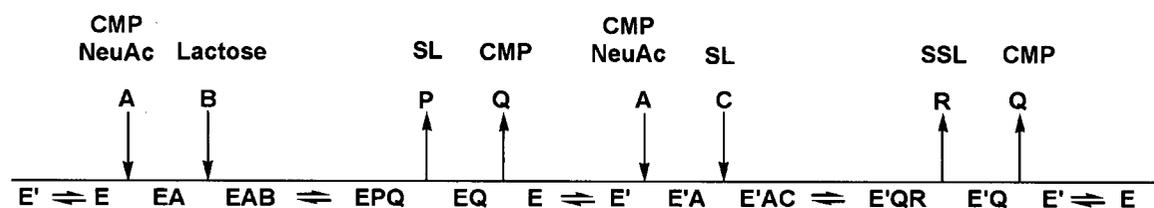
Of the few glycosyltransferases that have been subjected to detailed kinetic analysis, to date the vast majority have been shown to catalyse their respective reactions using an ordered Bi Bi ((Bendiak and Schachter, 1987) - GnTII; (Kearns et al., 1991) - XylT; (Murray et al., 1996) - FucT V; (Chen et al., 2002) - MurG; (Ly et al., 2002) - LgtC; (Zhang et al., 2006b) - TagA; (Morrison and Ebner, 1971) - Lactose Synthase; (Boix et al., 2002) -  $\alpha$ 3GalT) or Theorell-Chance ((Kamath et al., 1999) - GTB) mechanism involving the obligate binding of the nucleotide sugar donor substrate prior to the association of the acceptor. This finding has been rationalized on the basis of obtained three-dimensional X-ray crystal structures of glycosyltransferases with bound intact donor substrates. Upon binding the donor substrate, a conformational change occurs in which a flexible loop closes over the donor substrate and forms a significant portion of the acceptor substrate binding site, consistent with the observed ordered kinetic mechanisms. Interestingly, oleandomycin glucosyltransferase has also been shown to use an ordered Bi Bi mechanism, but in this case the acceptor substrate associates prior to the donor (Quiros and Salas, 1995). In addition, a polypeptide GalNAc transferase (Wragg et al., 1995) and O-GlcNAc transferase (Kreppel and Hart, 1999) were found to use rapid

equilibrium random Bi Bi mechanisms. These two enzymes both use peptide acceptor substrates.

Binding of substrates and release of products does not occur in a defined order but does contribute to the overall observed rate of reaction in the steady state random Bi Bi mechanism. In contrast to the frequently observed rapid equilibrium random Bi Bi mechanism, few examples have been reported for the steady state random mechanism. The most frequently quoted example, yeast hexokinase, required years of argument and reports of seemingly contrasting data before an authoritative study by Rudolph and Fromm closed the case (Rudolph and Fromm, 1971). This mechanism is very difficult to unambiguously identify for several reasons. The rate equation is painfully complex, making the derivation of expected modes of inhibition very difficult to predict. The initial rate data for this mechanism should in theory be non-hyperbolic, although there are several reasons why obtained data can appear hyperbolic (Segel, 1975), which can lead to erroneous interpretations of data. Unfortunately, for reasons described above, in the case of Cst II initial rates cannot be measured under truly saturating concentrations of donor substrate and thus curve fitting to non-hyperbolic equations cannot be used to support this mechanism. An additional complexity of the steady state random mechanism comes from the fact that certain portions of the mechanism can be in rapid equilibrium while others contribute to the overall rate, further complicating the derivation of the actual rate equation and the prediction of modes of inhibition.

The steady state iso ordered Bi Bi mechanism requires an obligate order of substrate binding and product release but also involves an isomerization between two stable enzyme forms that contributes to the overall observed rate of reaction (Segel,

1975). In such a mechanism, product Q will be a mixed inhibitor with respect to substrate A, as was found to be the case with Cst II, if the two can bind to different stable forms of the enzyme (Figure 2.6). Although Cleland predicted such a mechanism and the anticipated modes of inhibition in 1963 (Cleland, 1963), an example with experimental evidence was not reported until 1993 for Amaranth leaf betain aldehyde dehydrogenase (Valenzuelasoto and Munozclares, 1993). Because Cst II is a bifunctional enzyme catalysing two distinct sialyl transfer reactions, it seems quite reasonable that the enzyme could exist in two distinct stable forms. It is possible that, while monitoring transfer to 3'-sialyl lactose, CMP binds to the form of the enzyme that catalyses transfer to lactose. This would give rise to a mixed inhibition pattern with respect to CMP NeuAc if release of CMP and formation of the other stable enzyme conformation are partially rate limiting. The existence of two stable enzyme forms of Cst II is also consistent with the fact that CMP 3F NeuAc acts as a mixed inhibitor with respect to CMP NeuAc.



**Figure 2.6.** Schematic representation of the steady state iso ordered Bi Bi mechanism of Cst II. E and E' represent stable enzyme forms that catalyse sialyl transfer to lactose and 3'-sialyl lactose (SL) respectively. A represents CMP NeuAc, B represents Lactose and C represents 3'-sialyl lactose (as substrate). P represents 3'-sialyl lactose (as product), Q represents CMP and R represents the disialylated product (SSL). Product Q can associate with either stable forms of the enzyme E and E' which would make it a mixed inhibitor with respect to A. A dead end analogue of A could associate with both stable forms of the enzyme E and E' which would make it a mixed inhibitor with respect to substrate A.

Unfortunately, the described kinetic data alone cannot distinguish between the steady state random and steady state iso ordered Bi Bi mechanisms with complete certainty. However, in light of the bifunctional nature of Cst II and the precedence for the utilization of ordered mechanisms by glycosyltransferases, the iso ordered mechanism seems the most reasonable. Evidence disfavoring the steady state random mechanism could come from structural investigations indicating the occurrence of a conformational change resulting in the formation of the acceptor binding site upon donor substrate association.

### **2.3.7. Calculating Kinetic Parameters for Cst II**

Assuming an iso ordered Bi Bi mechanism, kinetic parameters for Cst II can be calculated using the data presented in Figure 2.2. Isomerization between stable enzyme forms results in the appearance of additional terms containing the concentrations of products P and Q in the denominator of the initial velocity equation that leads to changes in product inhibition patterns. However, in the absence of products the initial velocity equation of the steady state ordered and iso ordered Bi Bi mechanisms are identical (Segel, 1975). The initial rate equation of these mechanisms (equation 2.4) can be rearranged to take the general form of the Michaelis-Menten equation (equation 2.5). If the concentration of substrate A is varied while that of substrate B is held at a constant concentration, terms that do not contain A are constants. Values of the apparent Michaelis-Menten constants  $V_{\max}^{\text{app}}$  and  $K_m^{\text{app}}$  obtained at a given fixed concentration of substrate B are therefore functions of the concentration of B (equations 2.6 and 2.7, respectively). These terms can be assembled to take the form of equation 2.8. Because

equations 2.6 and 2.8 have the general form of the Michaelis-Menten equation, secondary plots of  $V_{\max}^{\text{app}}$  or  $V_{\max}^{\text{app}} / K_m^{\text{app}}$  against B will be rectangular hyperbolas through the origin that can be analysed by standard methods to obtain the four parameters  $V_{\max}$ ,  $K_m^A$ ,  $K_S^A$ , and  $K_m^B$ . Because  $K_S^B$  does not appear in equation 2.4, this parameter cannot be obtained from the analysis of secondary plots.

$$v_o = \frac{V_{\max}[A][B]}{K_S^A K_m^B + K_m^B[A] + K_m^A[B] + [A][B]} \quad (2.4)$$

$$v_o = \frac{\left( \frac{V_{\max}[B]}{K_m^B + [B]} \right) [A]}{\left( \frac{K_S^A K_m^B + K_m^A[B]}{K_m^B + [B]} \right) + [A]} \quad (2.5)$$

$$V_{\max}^{\text{app}} = \frac{V_{\max}[B]}{K_m^B + [B]} \quad (2.6)$$

$$K_m^{\text{app}} = \frac{K_S^A K_m^B + K_m^A[B]}{K_m^B + [B]} \quad (2.7)$$

$$V_{\max}^{\text{app}} / K_m^{\text{app}} = \frac{(V_{\max} / K_m^A)[B]}{(K_S^A K_m^B / K_m^A) + [B]} \quad (2.8)$$

Direct fit of the data in Figure 2.2 to rectangular hyperbola equations 2.6 and 2.8, shown in double reciprocal format in Figure 2.2, was used to calculate the kinetic parameters shown in Table 2.3. It is worth noting that the actual kinetic parameters calculated from this extensive kinetic analysis of Cst II are quite similar to the apparent values shown in Table 2.1.

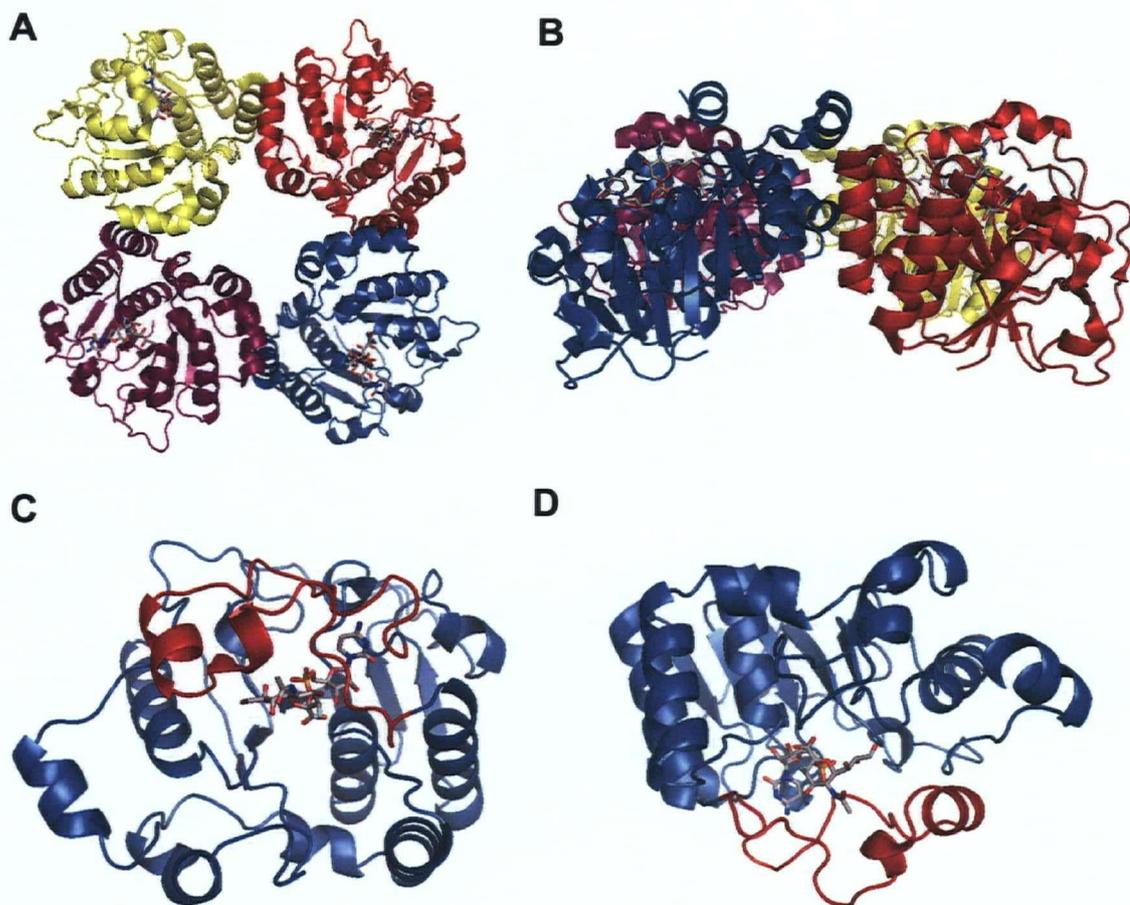
**Table 2.3.** Michaelis-Menten kinetic parameters for Cst II catalysed sialyl transfer to 3'-sialyl lactose. Values in parentheses represent the standard error.

Constant	Value
$K_m$ (CMP NeuAc)	0.40 (0.078) mM
$K_{ia}$	0.38 (0.064) mM
$K_m$ (SL)	3.7 (0.40) mM
$k_{cat}$	40 (1.4) min <sup>-1</sup>

## 2.4. Three Dimensional X-Ray Crystallographic Analysis of Cst II in Complex with a Stable Donor Substrate Analogue

In collaboration with Ms. Cecilia Chiu in Prof. Natalie Strynadka's laboratory at UBC, a co-crystal structure of the soluble truncated version of Cst II was obtained with either CMP or an intact molecule of CMP 3F NeuAc bound within the active site. Ms. Chiu found this soluble form of the enzyme to form a tetramer in solution by static light scattering as was also observed in the resulting crystal structure (Figure 2.7 A&B). The individual monomers of Cst II were found to be organized into two domains (Figure 7.6 C&D). The first domain, consisting of residues 1 to 154 and 189 to 259, is composed of a mixed  $\alpha/\beta$  fold in which a central parallel seven-stranded twisted beta sheet (topology  $\beta 8$ ,  $\beta 7$ ,  $\beta 1$ ,  $\beta 2$ ,  $\beta 4$ ,  $\beta 5$ ,  $\beta 6$ ) is surrounded by five helices on one side (D, E, F, I and J) and four helices (A, B, C and K) on the other. A Rossman fold nucleotide-binding domain is

formed by helices F, I and part of the  $\beta$ -sheet ( $\beta 8$ ,  $\beta 7$ ,  $\beta 1$ ,  $\beta 2$ , and  $\beta 4$ ). Suggesting that protein oligomerization has no direct role in catalysis, the active site of Cst II is not located at the interface between monomers. The second domain, consisting of residues 155 to 188, is composed of a coil extending from  $\beta 7$  followed by helices G and H and finally a coil that connects back to helix I. As discussed in chapter 1, this architecture is that of a GT-A fold and Cst II therefore belongs to glycosyltransferase Clan I. It is however worth noting that significant differences from other members of the GTA group and Cst II are observed in terms of secondary structure connectivity and the absence of a conserved DXD motif.



**Figure 2.7.** Overall structure of Cst II. Top down (A) and side (B) views of the overall tetrameric arrangement of Cst II. Individual monomers are colored differently and the active site location is indicated by the presence of the stick representation of CMP 3F NeuAc. Top down (C) and side (D) views of an individual monomer of Cst II with CMP 3F NeuAc bound within the active site. The two domains of the monomer are colored differently in blue and red.

A comparison of the structures of Cst II with either CMP or CMP 3F NeuAc bound reveals little change in the overall structure of the enzyme ( $\sim 0.2$  Å r.m.s deviation for 246 C $\alpha$  atoms). However, the presence of the sialic acid moiety leads to an ordering

of residues 175-187 and the formation of an ordered loop that forms a lid that effectively buries the donor substrate. Formation of this lid presumably protects the labile donor substrate from bulk solvent and forms a substantial portion of the acceptor binding site. Formation of the acceptor binding site following binding of the donor substrate is consistent with previously obtained glycosyltransferase structures and provides substantial support for the proposed iso ordered Bi Bi mechanism over the alternative random mechanism.

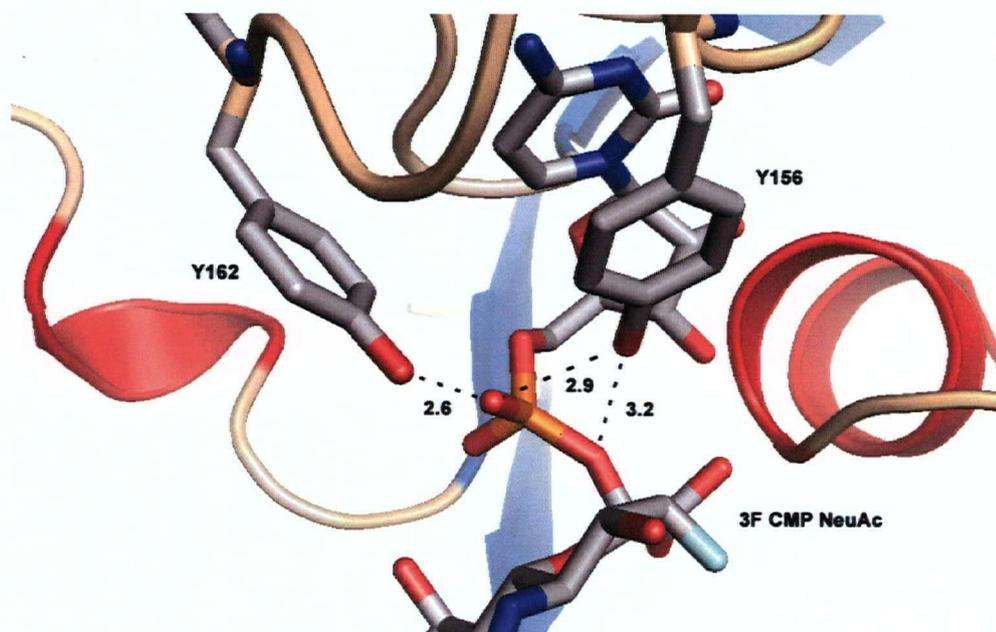
## **2.5. Investigating the Role of Putative Catalytic Residues**

To date, all structural and kinetic evidence for inverting glycosyltransferases supports a direct displacement mechanism proceeding through an oxocarbenium ion-like transition state facilitated by base assistance. Cst II catalyses two such direct displacements. In the first reaction the 3-hydroxyl of galactose acts as the nucleophile, directly attacking the anomeric centre (C2) of CMP NeuAc to form a NeuAc $\alpha$ -2,3-Gal linkage. In the second reaction the 8-hydroxyl from the sialyl moiety of the acceptor acts as the nucleophile, attacking the anomeric centre of a second CMP NeuAc bound in the active site (Scheme 2.1). Key questions about this mechanism then are the identities of the base catalyst that deprotonates the attacking nucleophile and of any acid catalyst that assists CMP departure. The presence of a carboxylate group at the anomeric centre of sialic acid alters the electronic and steric nature of the reaction centre and is likely to impose differences in the mechanism and the mode of substrate binding compared to other glycosyltransferases. The carboxylate substituent also provides a potential role in substrate-assisted catalysis.

### 2.5.1 Tyr156 and Tyr162 Facilitate CMP Departure

Unlike other glycosyltransferases with a GTA fold, bound metal was not observed in the active site of Cst II. In addition, Cst II lacks the 'DXD' sequence motif found in a wide range of glycosyltransferases that coordinates the divalent cation involved in binding of the nucleotide sugar via interaction with the diphosphate moiety (Ten Hagen et al., 1999; Wiggins and Munro, 1998). Bound metal is generally considered to act as a Lewis acid catalyst that stabilizes the departing nucleoside diphosphate. Mutagenesis studies of the conserved Asp residues in the DXD motif of glycosyltransferases from various species indicate that removal of the Asp residues, and hence the coordination of metal ions, completely eliminates the transferase activity (Busch et al., 1998; Ten Hagen et al., 1999; Wiggins and Munro, 1998). Kinetic studies with Cst II indicate that although magnesium and manganese enhance the activity somewhat (~2-fold increase for both metal ions), they are not essential for catalysis. This difference from other glycosyltransferases is not surprising given that the donor substrate in this case is a nucleoside monophosphate sugar (CMP NeuAc). Cst II therefore uses other means to activate/stabilize the leaving nucleotide.

Inspection of the structure of Cst II in complex with CMP 3F NeuAc indicates that departure of the phosphate leaving group could be facilitated by the hydrogen bonding interactions of Tyr156 and Tyr162 with the non-bridging pro S oxygen (Figure 2.8). In contrast to the alcohol leaving group with  $pK_a \approx 16$  released by glycosidases, the second ionization of the monophosphate leaving group has a  $pK_a$  of around neutrality and discrete acid catalysis is not required for Cst II. In fact, the situation is more



**Figure 2.8.** Cst II active site highlighting the hydrogen bond interactions between the side chain hydroxyls of Y156 and Y162 and the departing CMP leaving group. Indicated distances are measured in Angstroms.

equivalent to the second step of a retaining glycosidase wherein the leaving group is an enzymatic carboxylate and a tyrosine residue is often found to play a role in stabilizing the departing leaving group (Gebler et al., 1995). Mutational analysis of Cst II shows that  $k_{cat}$  values for the Y156F and Y162F mutants are 75-fold and 360-fold lower than those for the wild-type enzyme (Table 2.4). Further, mutation of both of these residues to phenylalanines results in the complete loss of any enzymatic activity above the rate of spontaneous CMP NeuAc hydrolysis. In the case of the retaining *Agrobacterium* sp.  $\beta$ -glucosidase, Tyr298 plays an equivalent role of interacting with the carboxylate leaving group and mutation to Phe likewise causes a 2500-fold decrease in its  $k_{cat}$  value (Gebler et al., 1995). This metal ion-independent method of facilitating leaving group departure

utilized by Cst II is more akin to the mode used by GT-B enzymes, as discussed in section 1.3.4.4., and may therefore indicate a convergence in mechanism.

**Table 2.4.** Apparent Michaelis-Menten parameters for CMP NeuAc with various Cst II mutants.

Cst II	$k_{cat}$ ( $\text{min}^{-1}$ )	$K_m$ (mM)	$k_{cat} / K_m$ ( $\text{mM}^{-1} \text{min}^{-1}$ )
WT	45 <sup>a</sup>	0.50	90
WT Hydrolysis <sup>b</sup>	2.0 <sup>b</sup>	0.50	4.0
R129A <sup>c</sup>	<0.20 <sup>a</sup>	0.25	0.80
H188A <sup>c</sup>	<0.20 <sup>a</sup>	0.30	0.50
Y156F	0.60 <sup>a</sup>	0.20	2.5
Y162F	0.10 <sup>a</sup>	0.45	0.30
Y156/162F <sup>d</sup>	<0.040 <sup>a</sup>	nd	-

\* - error range in data is between 5-20%

a - Kinetic parameters were determined at a constant lactose concentration of 160 mM

b - For enzyme catalysed hydrolysis kinetic parameters were determined in the absence of acceptor

c - No significant activity above that of enzyme catalysed hydrolysis (measured in the absence of acceptor) was observed

d - No activity above that of the significant rate of spontaneous background hydrolysis was detected up to an enzyme concentration of 500  $\mu\text{g/mL}$

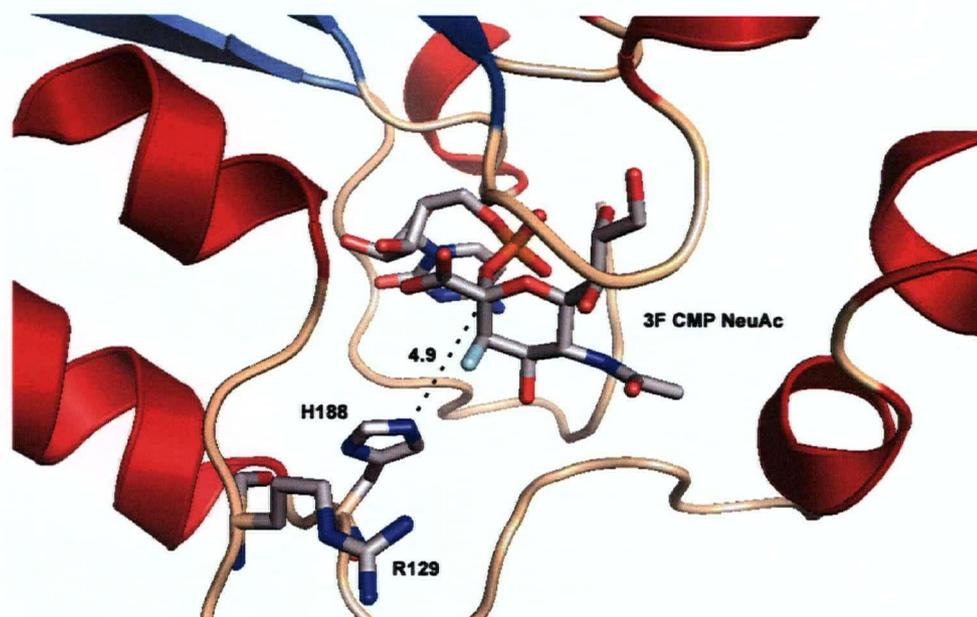
## 2.5.2 His188 Acts as the Catalytic Base

As mentioned above, a possible candidate for base catalyst is the carboxylate moiety of the substrate itself. This contention is supported by the complete absence of any basic amino acid side chains in the immediate vicinity of the carboxylate, in contrast to the three arginines seen in sialidases (Amaya et al., 2003; Burmeister et al., 1992; Crennell et al., 2000; Luo et al., 1998). The observed in-plane conformation of the

carboxylate is not consistent with a role as base catalyst, as such a role would require that the carboxylate be bound perpendicular to the ring plane (Figure 2.9). While it is still possible that the conformation would change upon formation of a ternary complex with the acceptor sugar, it is interesting to note that this near-planar conformation of the carboxylate is what has been predicted by *ab initio* calculations and supported by kinetic isotope effect experiments to exist during the transition state for the formation of a sialyloxocarbenium ion during the spontaneous hydrolysis of CMP NeuAc (Horenstein, 1997). Additionally, KIE experiments have cast serious doubt on the ability of the carboxylate of CMP-NeuAc to function as an intramolecular base catalyst during spontaneous hydrolysis (Horenstein and Bruner, 1998).

In glycosidases, including sialidases, and inverting glycosyltransferases, excluding certain GT-B fold family GT1 inverting enzymes that are also thought to use the imidazole of a His residue, the role of base is clearly played by the carboxylate group of a Glu or Asp residue. However, in the three-dimensional X-ray crystal structure of Cst II with bound CMP 3F NeuAc there are no suitably disposed acidic groups in the active site. The closest is that of Glu57 which is 14 Å away. Side chains situated close to the anomeric carbon are those of Asn31 (3.1 Å), Asn51 (4.0 Å), Ser132 (3.0 Å) and His188 (4.8 Å). His188 is the only feasible candidate and is situated appropriately on the alpha face of the sugar and adjacent to an open cleft in the enzyme that would be the most obvious site to bind the acceptor sugar (Figure 2.9). The imidazole side chain is suitably positioned to act as the base in catalysis, deprotonating the incoming hydroxyl group of the acceptor during nucleophilic attack at the anomeric carbon of the donor sugar. Precedence for the role of His188 as the base catalyst derives from the fact that side chain

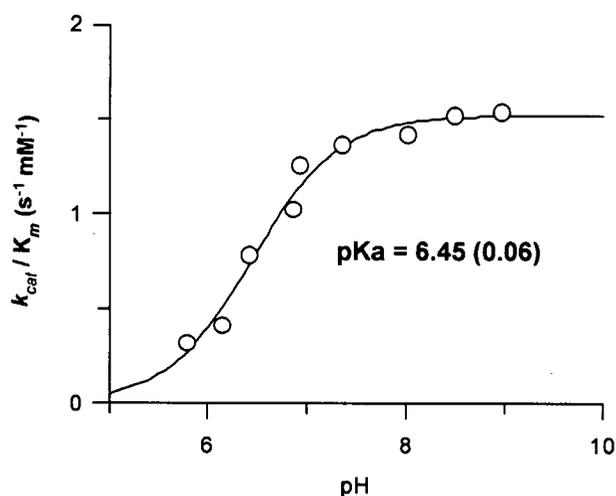
imidazoles are found to play this role in multiple other classes of enzymes (for example (Legler et al., 2002; Whiteson et al., 2007). In addition, studies on several human sialyltransferases (ST3Gal I, ST8Sia II and ST8 Sia IV) have revealed the identity of an invariant His residue that is essential for enzymatic activity (Close et al., 2003; Jeanneau et al., 2004). Finally, the recently described X-ray crystal structures of the sialyltransferases Cst I (Chiu et al., 2007) and PmST1 (Ni et al., 2007) reveal the existence of active site His residues with near identical relative positioning to that of His188 in Cst II.



**Figure 2.9.** Cst II active site highlighting the side chain imidazole of His188 suitably positioned on the  $\alpha$ -face of the donor substrate to play the role of base catalyst. Indicated distances are measured in Angstroms.

Optimal Cst II transferase activity occurs at approximately pH = 8 (Figure 2.10). At this pH, the imidazole of His 188 would be expected to be deprotonated, if it has a

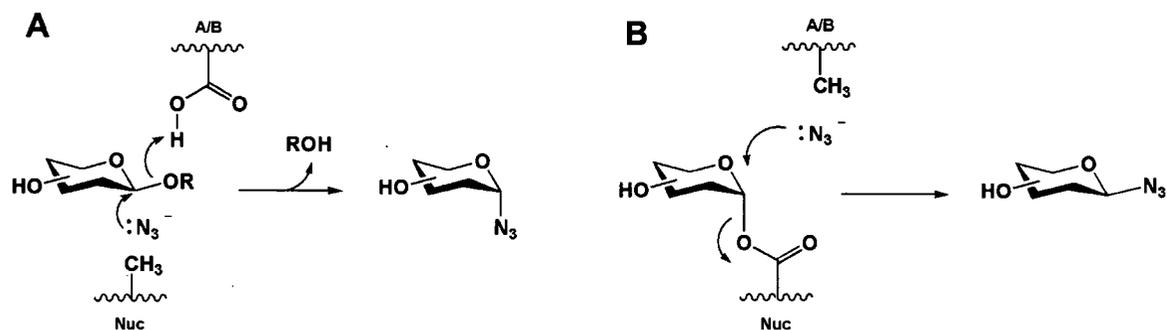
“normal” histidine pKa, as required for a base catalyst. Indeed, the experimentally determined pKa of Cst II was found to be 6.45, consistent with the typical pKa of a side chain imidazole (pKa ~7), providing support for a catalytic role for His188. Mutation of His188 to alanine results in a loss of all detectable transferase activity indicating a key catalytic role in Cst II activity (Table 2.4).



**Figure 2.10.** pH dependence of  $k_{cat}/K_m$  for Cst II. The pKa value was determined by best fit of the data to a single titration model in Grafit 4.0. The value in parenthesis represents the standard error.

Described as chemical rescue, reactivation of inactive mutants by the addition of a nucleophilic small molecule is a well established mechanistic probe for retaining glycosidases (Macleod et al., 1994; Viladot et al., 1998; Zechel and Withers, 2000). Exogenous addition of a small nucleophile (e.g. azide or formate) to a mutant enzyme, in which the catalytic nucleophile or the acid/base catalyst has been replaced by alanine, restores the activity of the mutant as a function of small molecule concentration and results in the formation of a covalent glycosidic adduct (Figure 2.11). Rescue indicates

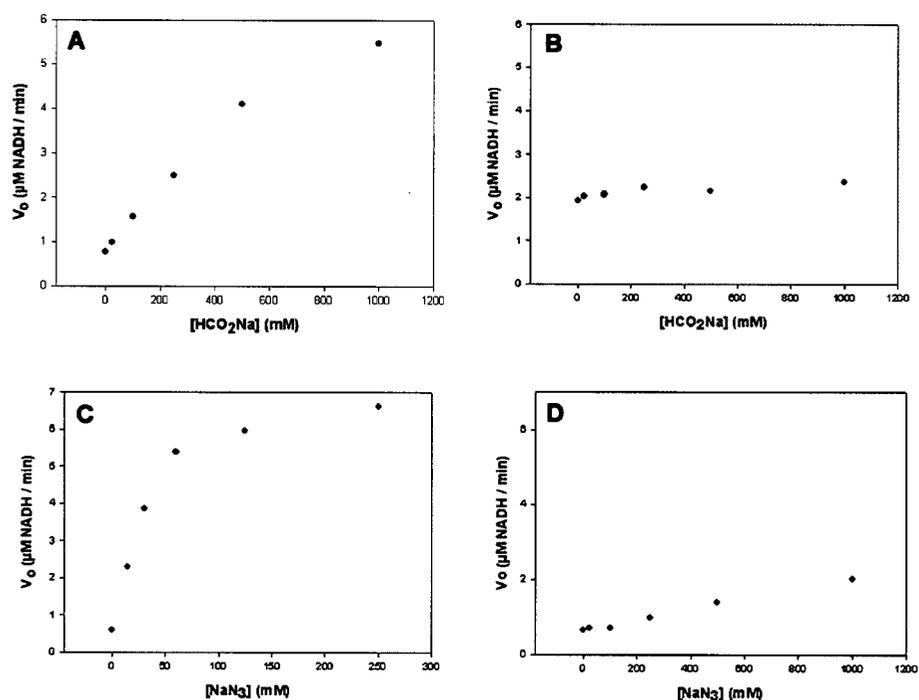
that the putative catalytic side chain is suitably positioned within the active site during the course of catalysis to play the proposed catalytic role.



**Figure 2.11.** Chemical rescue of nucleophile (A) and acid/base (B) mutant glycosidases with small molecules. When presented with a small exogenous nucleophile (e.g. azide), the nucleophile mutant of a  $\beta$ -retaining glycosidase will catalyse a group transfer reaction leading to a product of inverted anomeric configuration. In the case of the acid/base mutant of a  $\beta$ -retaining glycosidase, when deglycosylation is rate limiting, the presence of an exogenous nucleophile recovers enzymatic activity and leads to the formation of a covalent adduct with overall retention of anomeric configuration.

To further investigate the role of His188 as the base catalyst of Cst II, chemical rescue of the H188A mutant was attempted using sodium azide and sodium formate. Addition of either formate (Figure 2.12A) or azide (Figure 2.12C) to the H188A mutant resulted in significant reactivation of activity as determined using the NMPK/PK/LDH enzyme-coupled assay. Rates in the presence of acceptor were virtually identical to those in the absence of acceptor at all concentrations of exogenous nucleophile for the H188A mutant. This indicates that the observed activated reaction is not that of the transferase activity of the enzyme and presumably corresponds to the formation of a sialic acid derivative. TLC and ESI-MS was used to confirm the formation of a sialyl azide

derivative in the H188A reaction mixture. Importantly, increasing concentrations of either formate or azide did not result in a significant increase in the activity of wildtype Cst II (Figure 2.12 B and 2.12D, respectively).



**Figure 2.12.** Chemical rescue of Cst II H188A. Rates were determined using 500  $\mu\text{g/mL}$  of H188A mutant or 25  $\mu\text{g/mL}$  of WT enzyme, 500  $\mu\text{M}$  CMP NeuAc and standard Cst II assay conditions. (A) Activity of Cst II H188A in the presence of increasing concentrations of formate. (B) Activity of WT Cst II in the presence of increasing concentrations of formate. (C) Activity of Cst II H188A in the presence of increasing concentrations of azide. (D) Activity of WT Cst II in the presence of increasing concentrations of azide.

A conserved arginine residue (Arg129) lies 5.9  $\text{\AA}$  away from His188 and could act to provide an electrostatic shield favouring the neutral deprotonated form of the putative base His188 (Figure 2.9). This residue is also suitably positioned within the active site to

potentially play a role in orienting the incoming acceptor substrate. Kinetic analysis shows that mutation of Arg129 to alanine decreases the  $k_{cat}$  value by 210-fold and, similarly to mutants of His188, that the residual activity results from enzyme-catalysed hydrolysis (Table 2.4).

## 2.6. Concluding Remarks

A combination of kinetic, chemical, structural and mutagenesis techniques were used to investigate the catalytic activities of the bifunctional sialyltransferase Cst II. Kinetic analysis revealed that the overall rate ( $k_{cat}/K_m$ ) for sialyl transfer to a 3'-sialyl lactose acceptor substrate is approximately an order of magnitude greater than that for transfer to a lactose acceptor. In addition, the enzyme was found to catalyse substantial hydrolysis of the labile CMP NeuAc donor substrate. The measurement of initial rates under a matrix of donor and acceptor substrate concentrations revealed the enzyme to use a sequential mechanism and facilitated the accurate determination of kinetic parameters. Encouragingly, these kinetic parameters were found to be virtually identical to the apparent parameters determined using the less rigorous analysis used to characterize a number of active site mutants. Inhibition studies using either the product CMP or the dead end donor substrate analogue CMP 3F NeuAc were used to further refine the kinetic mechanism of Cst II and were found to be consistent with either a steady state random or steady state iso ordered Bi Bi mechanism. The bifunctional nature of Cst II makes the existence of two stable enzyme forms a reasonable supposition and the iso ordered Bi Bi mechanism the preferred choice. Evidence disfavoring the random mechanism comes from the obtained crystal structure of Cst II with an intact molecule of CMP 3F NeuAc

bound within the active site. As has been observed with other glycosyltransferases, upon binding the donor substrate, a conformational change occurs leading to the formation of a lid region over the active site that completes the formation of the acceptor binding site. This finding is consistent with an obligate order of substrate binding and product release, thereby disfavoring the alternative random mechanism.

Inspection of the X-ray crystal structure of Cst II with bound donor substrate analogue facilitated the identification of putative catalytic residues. Cst II catalyses sialyl transfer with net inversion of anomeric configuration via a  $S_N2$ -like direct displacement mechanism. In such a mechanism, the enzyme catalyses the reaction by facilitating departure of the leaving group and providing a base catalyst that activates the incoming nucleophile. In contrast to the majority of glycosyltransferases, which use a divalent cation to facilitate departure of a nucleotide leaving group, Cst II activity was not found to be metal dependent. Instead, Cst II appears to use two tyrosyl hydroxyl groups to stabilize departure of the CMP leaving group. With the exception of two other recently characterized sialyltransferases and certain members of the GT-B fold family GT1 enzymes that are also thought to use a side chain imidazole, all inverting glycosyltransferases investigated to date have a side chain carboxylate that is believed to play the role of the base that activates the incoming nucleophilic substituent of the acceptor substrate. In contrast, in the case of Cst II a side chain carboxyl group was not found to be appropriately located within the active site. Kinetic and chemical rescue analysis of the alanine mutant indicates that in the case of Cst II the side chain imidazole of His188 plays the role of base catalyst. Interestingly, a metal ion-independent method of leaving group activation and the use of an active site His residue as the catalytic base are

strategies utilized by GT-B fold enzymes. As such, the finding that Cst II utilizes a mechanism involving these chemical strategies and yet possesses a GT-A fold indicates a possible convergence in mechanism between these two classes of enzymes.

## CHAPTER 3

# STRUCTURAL AND MECHANISTIC INVESTIGATIONS WITH RETAINING GLYCOSYLTRANSFERASES\*

\* A version of portions of this chapter has been published:

Lairson, L.L., Chiu, C.P.C., Ly, H.D., He, S., Wakarchuk, W.W., Strynadka, N.C.J., Withers, S.G. (2004) Intermediate Trapping on a Mutant Retaining  $\alpha$ -Galactosyltransferase Identifies an Unexpected Aspartate Residue. *Journal of Biological Chemistry*. **279**: 28339.

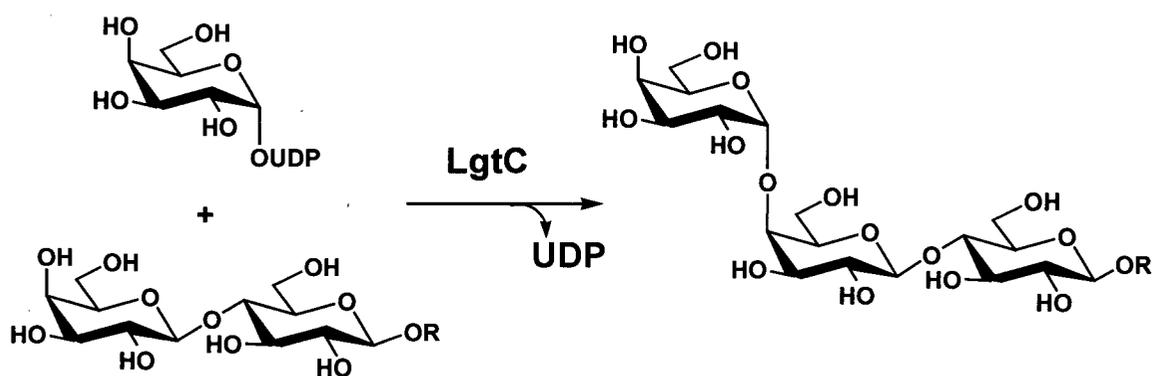
### 3.1. Summary

Despite decades of exhaustive research, the mechanism of retaining glycosyltransferases remains ambiguous. In order to investigate the potential for enzymatic nucleophilic catalysis in a double displacement mechanism, the active site of the retaining galactosyl transferase LgtC was engineered in an attempt to drive the reaction along such a coordinate. In addition, ESI-MS analysis was used to try to observe a putative glycosyl-enzyme intermediate with the galactosyltransferase  $\alpha$ 3GalT and the mannosyltransferase Kre2. Finally, the non-reactive substrate analogue GDP 2F Man was synthesized and used to obtain a three-dimensional X-ray co-crystal structure of Kre2.

### 3.2. Investigating the Catalytic Mechanism of the Retaining $\alpha$ -(1,4) Galactosyltransferase LgtC

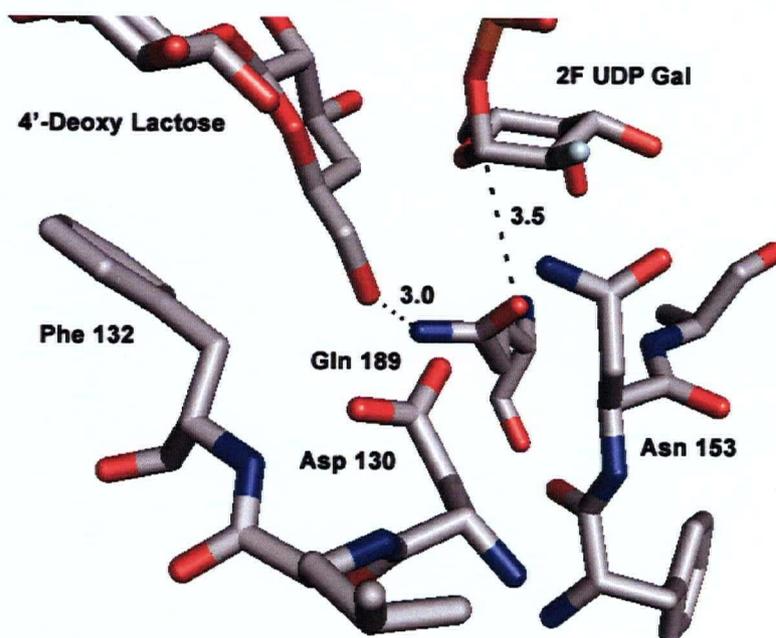
#### 3.2.1. Background

LgtC (EC 2.4.1.X ) is a family GT8  $\alpha$ -1,4 galactosyltransferase from *Neisseria meningitidis* that catalyses the transfer of galactose to the 4'-hydroxyl group of terminal lactose-containing acceptor substrates, using UDP Gal as the donor substrate, with overall net retention of anomeric configuration (Scheme 3.1). The Gal- $\alpha$ (1,4)-lactose product of the LgtC reaction is known as the P<sup>k</sup> antigen that terminates P blood group glycolipids in humans. Formation of this antigen on the surface of *Neisseria meningitidis* serves as a biomimicry strategy that allows the pathogen to evade an immune response from the host during infection.

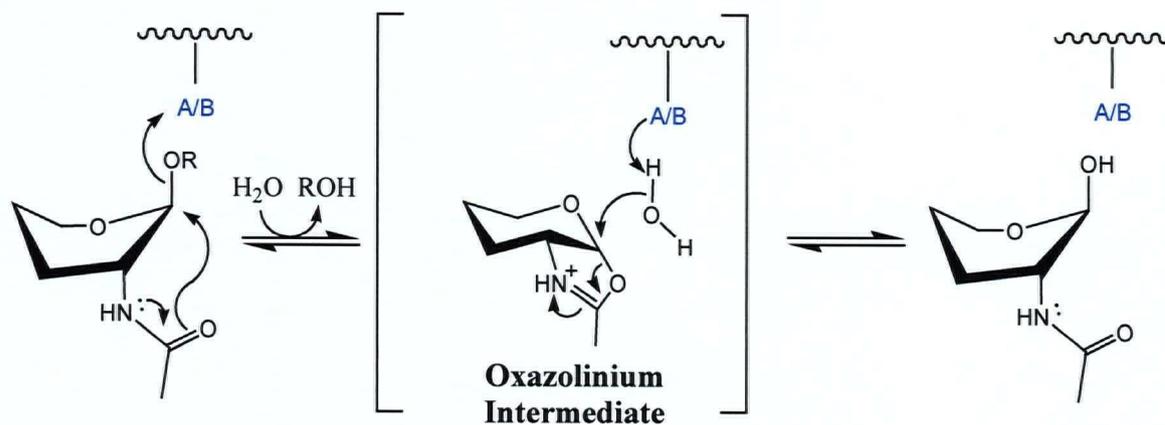


**Scheme 3.1.** Galactosylation reaction catalysed by LgtC.

A ternary complex three-dimensional X-ray crystal structure with both intact donor and acceptor substrate analogues bound is available for LgtC (Persson et al., 2001). Surprisingly, this structure revealed that the only active site side chain suitably positioned to act as the nucleophile in a double displacement mechanism is that of Gln189 (Figure 3.1). Although not generally considered a good candidate as an enzymatic nucleophile, an amide can indeed play this role as is seen with retaining hexosaminidases from families 18, 20 and 56, in which the acetamido substituent of the substrate functions as an intramolecular nucleophile leading to the formation of an oxazolinium ion intermediate (Mark and James, 2002; Mark et al., 2001) (Figure 3.2). In addition, a possible role of an amide as a nucleophile had been previously proposed for glycogen phosphorylase on the basis of a very similar active site interaction, though in that case involving a main chain amide (Mitchell et al., 1996). Subsequently, multiple other examples of amide functional groups positioned to play this role within the active sites of retaining glycosyltransferases have arisen as discussed in sections 1.3.5.4. and 1.3.5.5. .



**Figure 3.1.** LgtC active site with bound substrate analogues UDP 2F Gal and 4'-deoxy lactose highlighting the proximity of the side chain amide of Gln189 and the 6'-hydroxyl of lactose to the anomeric reaction centre (pdb 1ga8) .



**Figure 3.2.** Mechanism of retaining  $\beta$ -hexosaminidases involving anchimeric assistance leading to the formation of a stabilized oxazolinium ion intermediate.

However, despite this precedence, considerable doubt concerning a nucleophilic role for Gln189 was raised by the finding that the Q189A variant of LgtC possesses relatively high residual activity, with a  $k_{\text{cat}}$  value that is 3% of that of the wild-type enzyme (Persson et al., 2001). Doubt concerning a standard double displacement mechanism also derives from the lack of success in attempts to observe a covalently bound intermediate on LgtC, glycogen phosphorylase and other glycosyltransferases, despite exhaustive studies using techniques that have been successfully applied to observe the glycosyl-enzyme intermediates of glycosidases (Ly et al., 2002).

The possibility of a double displacement mechanism involving nonenzymatic nucleophiles has also been investigated with LgtC. In order to achieve overall net retention of anomeric configuration, a hydroxyl group from one of the substrates could act as a nucleophile that displaces the UDP leaving group in an initial step. Subsequent attack on the resulting  $\beta$ -linked galactosyl intermediate by the 4'-hydroxyl of the lactose acceptor from the  $\alpha$ -face would yield the resulting Gal-( $\alpha$ -1,4)-Gal linkage. Candidate intramolecular nucleophiles include the 4''- and 6''-hydroxyls of the UDP Gal donor. The plausibility of such roles was discounted on the basis of the inability of LgtC to turnover either of the expected  $\beta$ -1,4 or  $\beta$ -1,6 anhydrogalactose intermediates (Ly et al., 2002). This conclusion is further supported by a report that LgtC is able to use both UDP 4''-deoxygalactose and UDP 6''-deoxygalactose as substrates (Sujino et al., 2000), which would clearly not be possible if either of the above mechanisms were true. Another candidate nonenzymatic nucleophile is that of the 6'-hydroxyl of the acceptor substrate lactose, which is in fact well positioned for such a role (Figure 3.1). A mechanism involving the lactose acceptor in this way would have the advantage of minimizing

unwanted donor substrate hydrolysis, as it would obligate ternary complex formation prior to the production of a reactive intermediate species. Limiting the plausibility of such a mechanism is the fact that the resulting intermediate would be that of an inherently unreactive glycoside species. The plausibility of this mechanism was discounted based on the finding that the chemically synthesized putative intermediate (galactosyl  $\beta$ -1,6-lactose) was not turned over by the enzyme (Persson et al., 2001).

These exhaustive studies led to the proposition that LgtC uses a mechanism termed “S<sub>Ni</sub>-like” (Figure 1.12). Such a mechanism has limited chemical precedence and would prove difficult to distinguish experimentally from a classic S<sub>N1</sub>-like mechanism involving the formation of a discrete oxocarbenium ion intermediate.

### **3.2.2. Engineering the Mechanism of LgtC**

To investigate the degree of nucleophilic character contributed by Gln189 during catalysis, a Q189E variant of LgtC was created, thereby replacing the side chain amide with a more nucleophilic carboxylate. This substitution would also introduce a worse leaving group for the putative covalent glycosyl-enzyme intermediate (pK<sub>a</sub> of -0.5 to -1.0 for protonation of the oxygen of an amide versus a pK<sub>a</sub> of ~4.0 for that of a carboxylate), which may lead to a rate limiting deglycosylation step and the potential accumulation of the intermediate species. The mutant protein was subjected to kinetic, mass spectrometric and X-ray crystallographic analysis. Making the transferase more “glycosidase-like” might result in intermediate accumulation or even perhaps a more efficient enzyme with respect to the rate of transfer, albeit with the potential cost of an increased relative rate of unwanted donor substrate hydrolysis (as discussed later).

### 3.2.3. Kinetic Analysis of LgtC Q189E

Expression and purification of the Q189E variant of LgtC-25 proceeded without difficulties yielding protein that displayed the same purification behaviour, yield and CD spectrum as the wild-type enzyme. Kinetic analysis indicates that, while the  $K_M$  for UDP Gal is approximately unchanged, the measured  $k_{cat}$  value of this mutant is approximately 3% of that of the wild-type enzyme (Table 3.1). Interestingly, this is very similar to the residual activity of the Q189A mutant observed previously (Table 3.1). It might be expected that substitution of the weakly nucleophilic side chain of glutamine with that of a carboxylate-containing side chain could increase the turnover rate of an enzyme that catalysed a reaction via a double displacement mechanism, if that residue was in fact acting as the catalytic nucleophile. The observed decrease in turnover rate could be interpreted as indicating that this residue is not the catalytic nucleophile, that introduction of a worse leaving group for turnover of the glycosyl-enzyme intermediate leads to a change in the rate limiting step that leads to an overall decrease in the rate of turnover, or that an uncharged nucleophile is required for optimal activity on a charged substrate. The recent finding that an active site tyrosine acts as the nucleophile in the double displacement mechanism of a trans-sialidase, which uses charged sialic acid-containing substrates, supports this latter interpretation (Watts et al., 2003). Alternatively, this finding is consistent with the notion that this enzyme uses a mechanism that differs from that involving the discrete nucleophilic catalysis of a double displacement mechanism.

**Table 3.1.** Michaelis-Menten parameters for the transferase and hydrolase activities of WT and various mutants of LgtC-25 with respect to UDP Gal and UDP Glc donor substrates measured at pH 7.5, 37°C.

LgtC	Activity	UDP Galactose			UDP Glucose	
		$k_{cat}$ (s <sup>-1</sup> )	$K_M$ ( $\mu$ M)	$k_{cat} / K_M$ ( $\mu$ M <sup>-1</sup> -s <sup>-1</sup> )	$k_{cat}$ (s <sup>-1</sup> )	$K_M$ ( $\mu$ M)
WT	Transferase <sup>a</sup>	24	20	1.2	0.007	N/A <sup>c</sup>
	Hydrolase <sup>b</sup>	0.02	4.5	0.004	0.0003	N/A <sup>c</sup>
Q189E	Transferase <sup>a</sup>	0.68	23	0.030	0.002	N/A <sup>c</sup>
	Hydrolase <sup>b</sup>	0.003	N/A <sup>c</sup>	-	0.001	N/A <sup>c</sup>
Q189A	Transferase <sup>a</sup>	0.43	25	0.017	nd	nd
D190N <sup>d</sup>	Transferase <sup>a</sup>	0.008	N/A <sup>c</sup>	-	nd	nd
	Hydrolase <sup>b</sup>	0.010	N/A <sup>c</sup>	-		

\* - standard error range in data is from 5 – 20%

a – determined at saturating concentrations of lactose acceptor (240 mM)

b – determined in the absence of acceptor substrate

c – values could not be obtained as the high concentrations of enzyme required to obtain significant rates approached those of unsaturating donor substrate concentrations in violation of Michaelis-Menten assumptions. Apparent  $k_{cat}$  values were obtained from rates observed at saturating donor substrate concentrations (100  $\mu$ M).

d – the D190N mutant was generated in a C128/174S background

nd – not determined

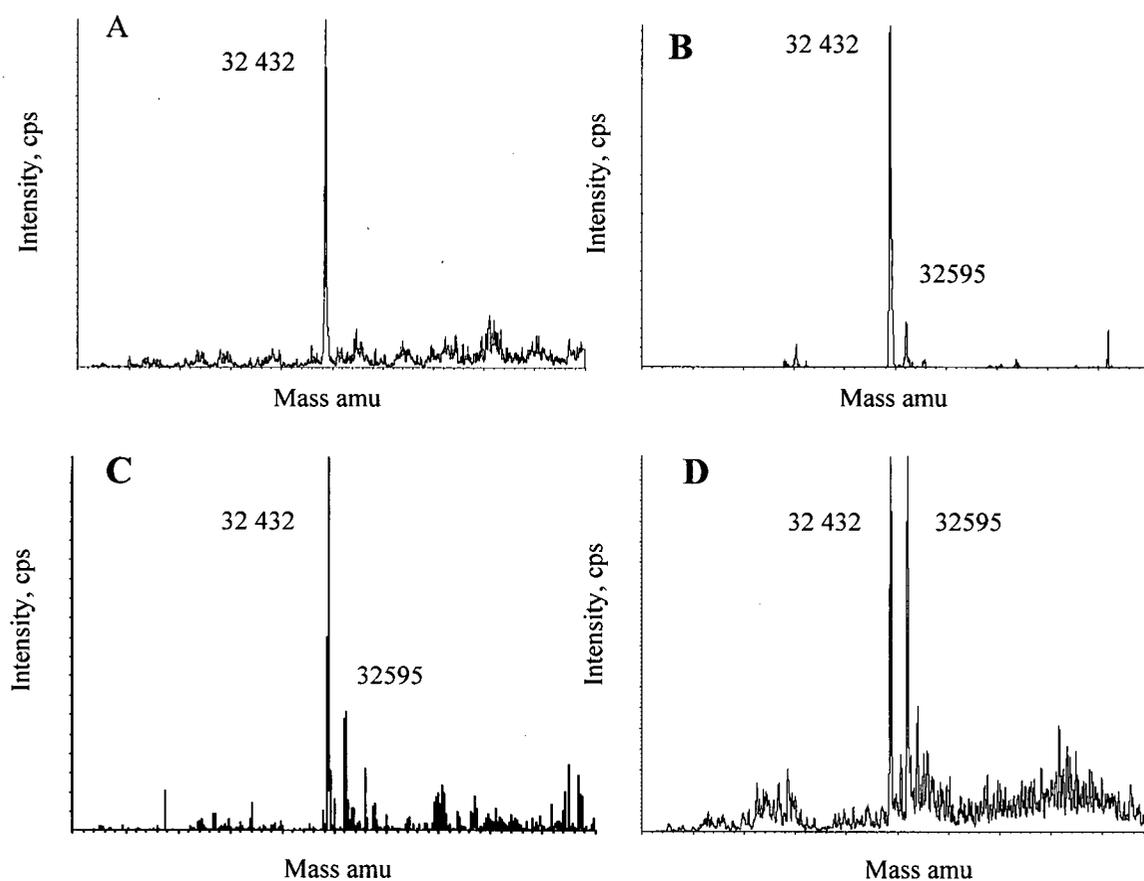
### 3.2.4. Intermediate Trapping by ESI-MS

To further investigate the contribution of nucleophilic catalysis to the mechanism of the Q189E mutant, the enzyme was incubated with various combinations and concentrations of donor and acceptor substrates to see if any covalently modified protein could be observed by ESI-MS. Interestingly, incubation with 1 mM UDP Gal for 2 minutes followed by quenching in 3 M urea resulted in what appeared to be a steady state

population (~10%) of protein for which the mass had increased by 164 mass units, corresponding to a covalently bound galactosyl moiety (Figure 3.3.B). In an attempt to increase the level of labelling, UDP Glc was employed in place of UDP Gal since it has a lower  $K_M$  value and slower rate of turnover. As hoped, UDP Glc labelled the mutant to a slightly higher level (~30%) (Figure 3.3.C). Incubation in the presence of higher concentrations of donor substrate or in the presence of various concentrations of the incompetent acceptor analogue 4'-deoxy lactose did not increase the relative amount of labelled protein. Surprisingly, quenching the reaction with acidic conditions (phosphate buffer pH 1.6) resulted in the loss of any observable labelled protein. This finding is unusual in that glycosidase intermediates covalently bound by ester linkages are typically stable under acidic conditions. A covalently modified intermediate of the Q189A variant of LgtC-25 was not observed under similar conditions.

Supporting the notion that the observed labelled portion of the mutant enzyme was in fact that of a catalytically competent steady state population, rather than a 'dead-end' inhibited population, was the observation that removal of excess UDP Gal from such mixtures by centrifugal dialysis resulted in complete loss of the labelled species as determined by ESI-MS. Presumably the galactosyl-enzyme intermediate was hydrolysed as a result of the significant rate of background hydrolytic activity (Table 3.1). However, because of the higher level of labelling and lower turnover rate, removal of excess UDP Glc from enzyme labelled with that nucleotide sugar allowed the isolation and observation of a labelled species that could be turned over in a time-dependent fashion by incubation in the presence of saturating amounts of lactose acceptor. This finding indicates that the observed intermediate is a catalytically relevant species. Indicating a

significant kinetic competence for the covalent species, fitting the rate of decay of labelled enzyme (as determined by ESI-MS) to a first order rate equation allowed the calculation of a first order rate constant of  $0.0014 \text{ s}^{-1}$  for the turnover of this intermediate via transglycosylation onto lactose, in close agreement with the observed  $k_{\text{cat}}$  value of  $0.002 \text{ s}^{-1}$  for the UDP Glc transferase activity of this mutant (Table 3.1 and Figure 3.4).



**Figure 3.3.** ESI-MS analysis of labelled and unlabelled species of LgtC Q189E. (A) unlabelled, (B) labelled with Gal, (C) labelled with Glc, and (D) labelled with Gal in the presence of pyruvate kinase and phosphoenolpyruvate.

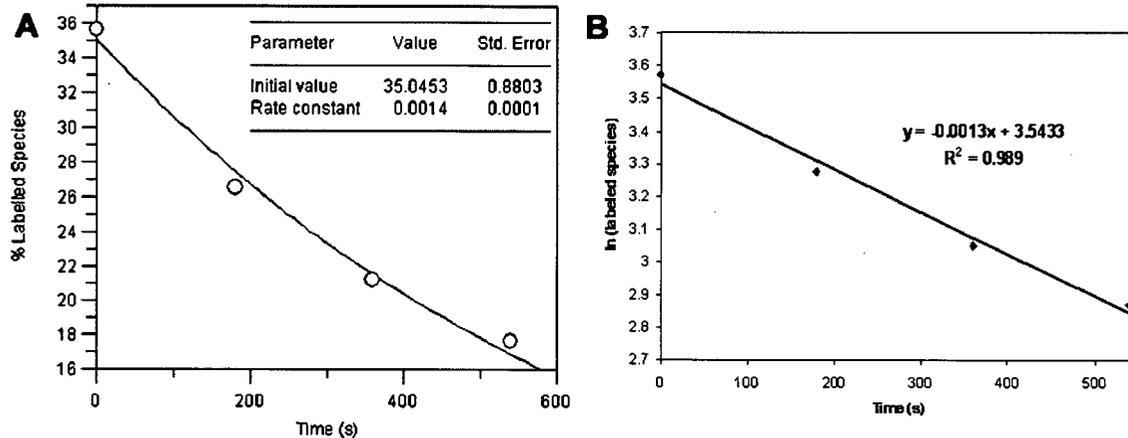
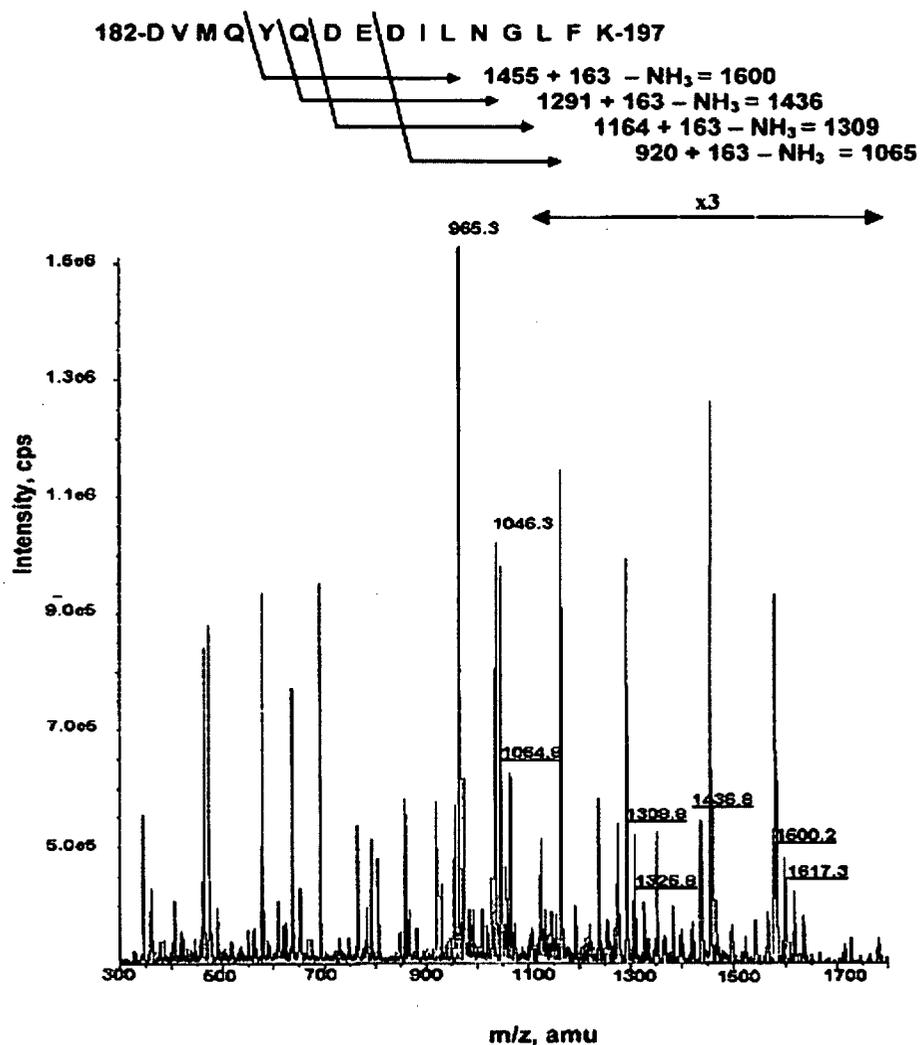


Figure 3.4. Decay data for the turnover of the glycosyl-enzyme intermediate species resulting from the incubation of UDP Glc with LgtC Q189E. Following removal of excess UDP Glc, enzyme was incubated in a solution containing saturating lactose acceptor and left to incubate for various time periods prior to quenching with 6 M urea. First order rate constants were calculated by direct fit of the data to a single exponential first order decay equation (A) or from the slope of a plot of the natural logarithm of concentration versus time (B).

Finally, in an attempt to 'pull' the labelling equilibrium over, the Q189E mutant enzyme was incubated with UDP Gal in the presence of pyruvate kinase and phosphoenolpyruvate to remove UDP formed (Scheme 3.2). Indeed, the introduction of this coupling system resulted in a dramatic increase of the relative proportion of the labelled enzyme species (Figure 3.3.D).



and fragmented by increasing the orifice voltage to 65 V. For the labelled peptide species, in addition to the fragmentation pattern found for the unlabelled peptide, peaks were observed having  $m/z$  values of 1617, 1600, 1436, 1327, 1309, and 1065 corresponding to peptide fragments containing a covalently bound galactosyl moiety (Figure 3.5). Most significantly, the peak at  $m/z = 1065$  can only correspond to the peptide fragment  ${}_{190}\text{DILNGLFK}_{197}$  lacking ammonia ( $m/z$  902) and containing the galactose label ( $m/z$  164). To our considerable surprise, therefore, the site of labelling has to be Asp190. Substitution of Gln189, suitably positioned in the ground state crystal structure to play the role of a catalytic nucleophile, with a more nucleophilic Glu residue therefore results in the accumulation of a steady-state population of covalently modified protein. However, the site of labelling is not the mutated side chain, as was anticipated, but rather a sequentially adjacent aspartate residue.



**Figure 3.5.** ESI-MS analysis of the fragmented, purified Gal-labelled peptide. The peptide species bearing the galactosyl label was identified by comparative mapping of tryptic digests of native and labelled LgtC Q189E enzyme species. The sequences of the unlabelled and labelled peptide species were deduced from the observed fragmentation patterns and the enzyme primary sequence to be that of residues 182 to 197 as shown. The location of covalent attachment can conclusively be determined to be Asp190 from the observed peak at  $m/z$  1065, as no other fragment corresponding to this mass is possible from the known peptide sequence and as this peak is only observed from the fragmented labelled peptide species. The region corresponding to  $m/z$  ratios greater than 1100 has been magnified three-fold.

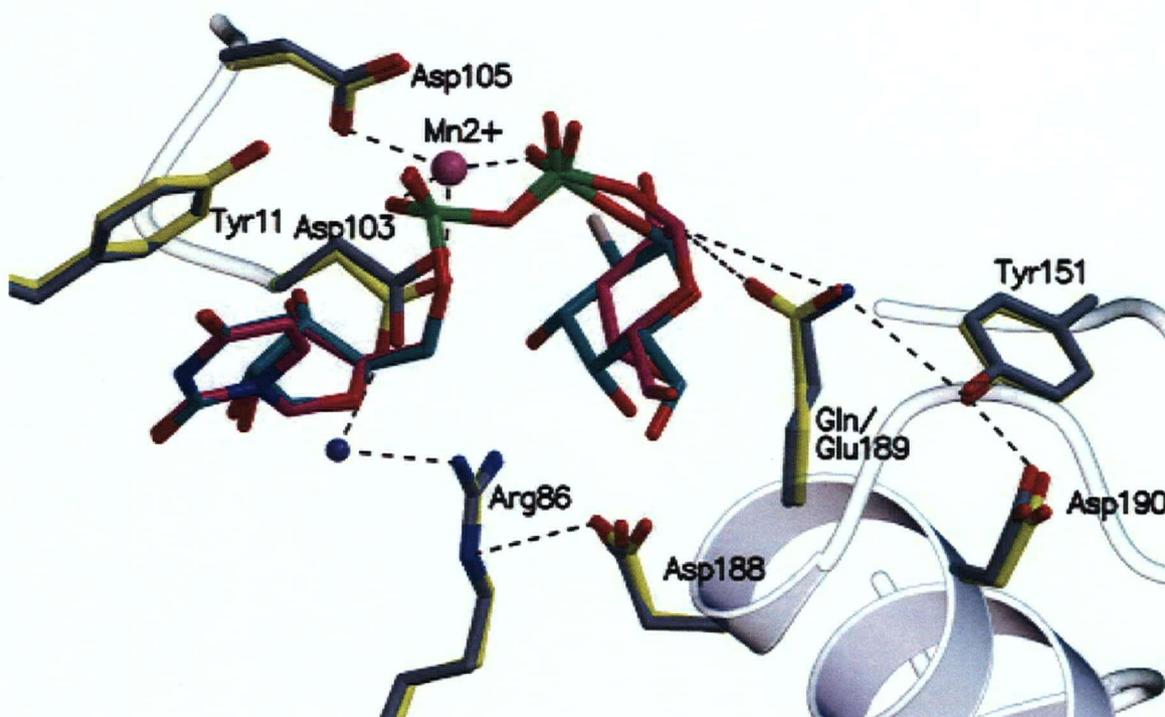
### 3.2.6. Analysis of the D190N Variant of LgtC

To probe the mechanistic importance of Asp190, the D190N mutant of LgtC was generated. Initial attempts to generate this variant based upon the wild-type enzyme as template resulted in yields of purified protein that were insufficient for characterization, presumably due to a destabilization of the protein structure. To counter this problem, the mutation was introduced into a double cysteine mutant of LgtC-25 (C128/174S) that had previously been shown to be more stable while maintaining kinetic characteristics that are indistinguishable from those of the wild-type enzyme (Persson et al., 2001). This allowed sufficient enzyme to be purified for kinetic characterisation. The  $k_{\text{cat}}$  value of the D190N variant, measured using UDP Gal as variable substrate at a fixed saturating concentration of lactose, was found to be 3000-fold less than that of the wild-type enzyme (Table 3.1). This is consistent with a potential role for this residue as the catalytic nucleophile in a double displacement mechanism. To ensure that the observed decrease in catalytic proficiency was not simply the result of unfolded or misfolded enzyme, the D190N variant was subjected to CD analysis, yielding a spectrum indistinguishable from that of the Q189E and wild-type versions of the enzyme.

An exact  $K_M$  value could not be obtained for the D190N mutant of LgtC-25, as the enzyme appeared to be saturated by substrate concentrations (5  $\mu\text{M}$ ) approaching that of the relatively high enzyme concentration needed to measure significant rates of galactosyl transfer. An upper limit of 5  $\mu\text{M}$ , however, seems reasonable and this significantly lower apparent  $K_M$  than that for the wild-type enzyme ensures that the rates measured are indeed those of the mutant and not those of contaminating wild-type enzyme (Ly and Withers, 1999).

### 3.2.7. Crystallographic Analysis of LgtC Q189E

A three-dimensional X-ray crystal structure of the Q189E mutant of LgtC was obtained at 2.6 Å resolution with a molecule of UDP 2F Gal bound in the active site by Ms. Cecilia Chiu in Prof. Natalie Strynadka's laboratory at UBC. The Q189E structure can be overlaid with the wild-type structure with a r.m.s deviation of 0.25 Å for 269 C $\alpha$  atoms with near identical relative positioning of the carboxylate oxygen of D190 and the anomeric carbon of the donor (Figure 3.6).



**Figure 3.6.** Superimposition of the wild-type and Q189E mutant active sites of LgtC. Side chains for the active site residues are labelled and depicted in CPK coloring with carbon atoms in gray (wild-type) or yellow (mutant), nitrogen atoms in blue, and oxygen atoms in red. UDP 2F Gal is shown with carbon atoms in cyan (wild-type complex) or magenta (mutant complex), nitrogen atoms in blue, oxygen atoms in red, and fluorine atoms in light gray. Mn<sup>2+</sup> ion is shown as a violet sphere and the water molecule in blue.

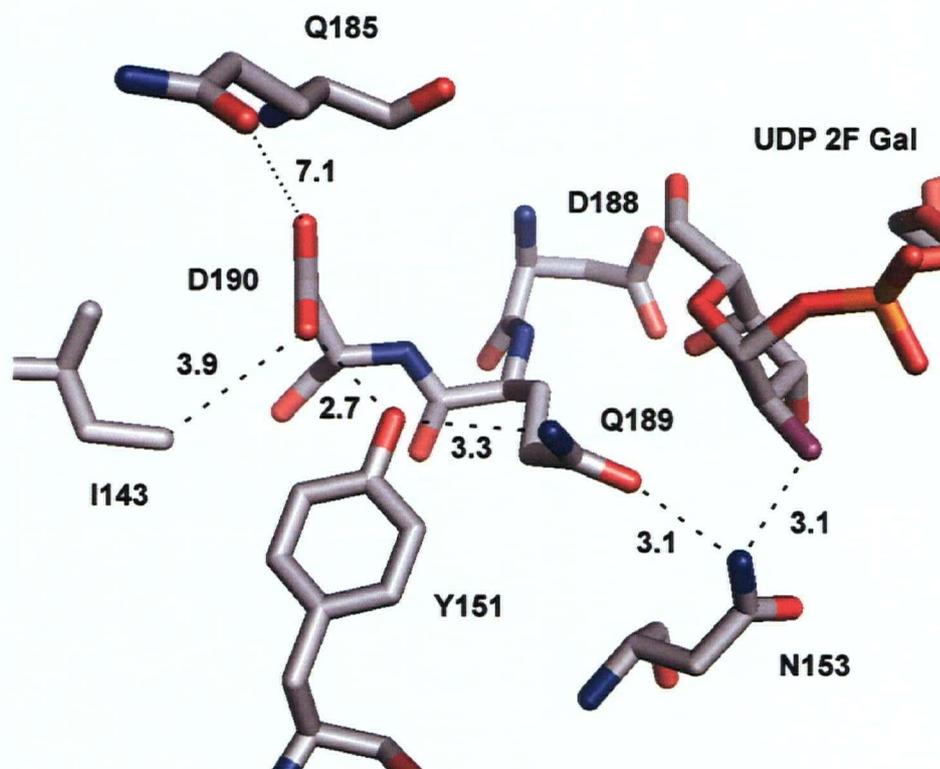
The near identity of the structures of the wild-type and Q189E mutant indicates that the observed X-ray crystal structures are not of those of the conformation that would lead to the formation of covalent glycosyl-enzyme species involving Asp190. As such, if Asp190 were to act as the nucleophile in a double displacement mechanism, a conformational change from that of the observed ground state would have to occur during the course of the reaction.

### **3.2.8. Attempts to Accumulate a Covalent Intermediate using Alternative Active Site Modifications**

If the side chain carboxylate of Asp190 does in fact act as the catalytic nucleophile in a double displacement mechanism, why does the modification of an adjacent residue result in the accumulation of an observable steady state population of the covalent glycosyl-enzyme intermediate? Because no such intermediate is observed with the wild-type enzyme, one possible explanation is that the modification causes a change in the relative rates of the glycosylation and deglycosylation steps of the mechanism. The Q189E mutation results in the introduction of a charged side chain in close proximity to that of Asp190, thus producing a potentially unfavourable charge-charge interaction. Formation of a glycosyl-enzyme intermediate would neutralize one of these charges, which could change the relative rates of the glycosylation versus deglycosylation steps leading to an increased life time for the intermediate species.

To explore this possibility, other modifications of the LgtC active site were made to determine if this would result in a similar outcome to that of the Q189E mutation. The ternary complex X-ray crystal structure was used to identify candidate uncharged residues

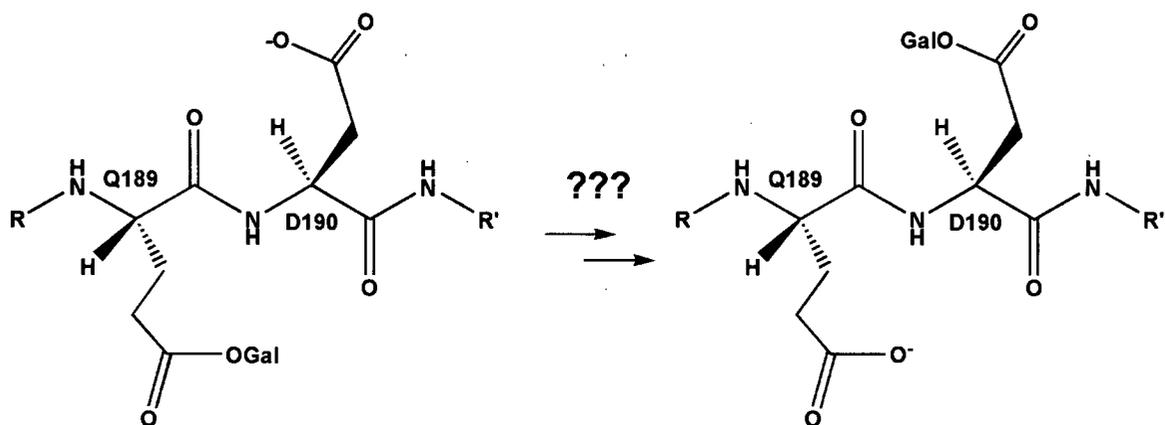
close to the side chain of Asp190. Several uncharged residues were identified that were either in direct proximity (Ile143, Gln185) or involved in a hydrogen bond network (Asn153) with Asp 190 (Figure 3.7). Site directed mutants I143D, N153D and Q185D were generated in the double cysteine background. Unfortunately, the I143D and N153D mutations resulted in a complete loss of soluble protein expression, despite exhaustive efforts to optimize expression conditions. This was not completely unexpected as both mutations result in the introduction of a charged group within the core of the protein, which is frequently found to severely compromise structural stability. Expression of the Q185D variant proceeded with expression levels comparable to that of the wild-type enzyme. However, no covalent glycosyl-enzyme species was observed despite testing a variety of labelling conditions. This could be a consequence of the side chain of this residue being close to the protein surface where the effect of the introduced charge could be neutralized by bulk solution.



**Figure 3.7.** LgtC active site highlighting the side chains of residues within close proximity to that of Asp190 (pdb 1ga8). Distances are indicated in Angstroms.

### 3.2.9. Interpretation of the LgtC Q189E Labelling Results

Several interpretations of the surprising LgtC Q189E labelling results are possible. The first possibility is that labelling does in fact occur at Glu189, but that a subsequent migration results in the label being observed at the alternate Asp190 site. However, the relative positioning of both of these residues within the active site (both on the  $\beta$  face of the galactose ring) would prevent glycosyl migration from occurring in the intact enzyme (Figure 3.6 and 3.7). Additionally, it is hard to envision a chemically sound mode of acyl migration on the denatured or digested protein (Figure 3.8).



**Figure 3.8.** Acyl migration of the galactosyl moiety from the side chain of Q189 to that of D190 is difficult to rationalize.

It must therefore be considered possible that Asp190 is in fact the catalytic nucleophile of LgtC and that retaining glycosyltransferases catalyse reactions via a double displacement mechanism involving discrete nucleophilic catalysis. The Q189E mutation may lead to a slowing down of the deglycosylation step relative to the glycosylation step, allowing the accumulation of an observable steady state population of labelled enzyme. Indeed, formation of the glycosyl-enzyme intermediate would neutralise the negative charge of Asp190, which is in close proximity to the introduced carboxylate of Glu189, eliminating charge-charge repulsion and thereby altering the relative rates of intermediate formation and breakdown. This interpretation is supported by the observed pH-dependent stability of the labelled species. At low pH the introduced carboxylate would be protonated, preventing unfavourable charge interaction with D190 and thus resulting in a decrease in the stability of the glycosyl-enzyme intermediate. This argument is in accordance with the observed lack of discernible labelling with the Q189A mutant, in which case the mutation does not introduce a negatively charged species.

It is worth noting that previous suggestions regarding the mechanism of LgtC, and retaining glycosyltransferases in general, have been based primarily on the proximity of candidate nucleophiles to the anomeric carbon of a donor sugar within a crystal structure. Indeed, on this basis Asp190 was discounted as a potential catalytic nucleophile due to the distance (8.9 Å) separating the carboxylate oxygen and the reaction centre in the crystal structure of LgtC with bound donor analogue (PDB 1ga8). If Asp190 were in fact the catalytic nucleophile, a conformational change would need to occur *during* catalysis that would bring it into an appropriate position and orientation with respect to the donor sugar substrate. The need for a significant conformational change during catalysis is also suggested by the observed crystal structure of the self-glycosylating enzyme glycogenin (EC 2.4.1.186), the other family GT8 enzyme for which a structure with a bound intact donor sugar substrate exists (Gibbons et al., 2002). This enzyme transfers a glucosyl moiety from UDP Glc to the hydroxyl of Tyr195 on the surface of another monomer. The carboxylate oxygen of the proposed nucleophile of this enzyme (Asp163) was found to be 6.1 Å from the anomeric carbon of the donor sugar while that of Asp160 (structurally equivalent to Asp 190 of LgtC) was found to be 6.9 Å away. Equally significant is the observed distance of ~12 Å separating the anomeric reaction centre of the donor substrate from the hydroxyl oxygen of the Tyr195 acceptor.

It is also possible that the observed site of labelling is simply an artefact arising from an altered active site configuration that differs significantly from that of the wild-type enzyme in which Asp190 becomes suitably positioned to attack the UDP-Gal donor as a result of the introduced mutation. Substitution of a neutral Gln residue with a charged Glu might introduce significant charge repulsion leading to an altered active site

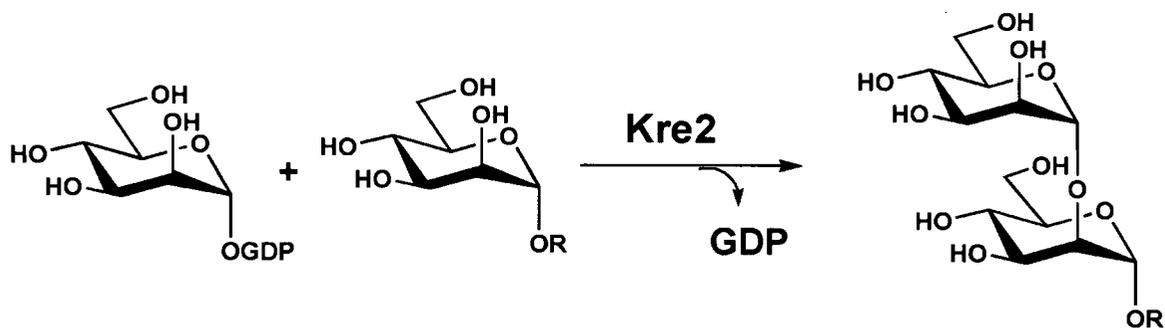
conformation. To determine whether this latter interpretation was the source of the observed unexpected site of labelling, a crystal structure of the Q189E variant was obtained at 2.6 Å with UDP 2F Gal bound in the active site and compared to the equivalent structure of the wild-type enzyme (Figure 3.6). The observed nearly identical observed active site configurations of these two structures render it unlikely that the unexpected site of labelling was the result of an altered active site conformation or mode of donor sugar binding.

Mutation of the catalytic nucleophiles of retaining  $\beta$ -glycosidases typically results in decreased observed  $k_{\text{cat}}$  values by factors of  $10^5$ - to  $10^8$ -fold (Ly and Withers, 1999). However, equivalent mutations in  $\alpha$ -retaining enzymes have been shown to result in less dramatic effects. For example, mutation of the catalytic nucleophile (Asp229) of the family GH13 cyclodextrin glycosyltransferase (CGTase, EC 2.4.1.19), the enzyme responsible for the conversion of starch and other  $\alpha$ -1,4-linked glucopyranosyl chains into cyclic oligosaccharides, to asparagine led to observed decreases in activity on the order of 2000-fold (Knegtel et al., 1995; Mosi et al., 1997). The kinetic analysis of the D190N variant may therefore be interpreted as being consistent with the role of this residue as the catalytic nucleophile in a double displacement mechanism. Recently, it was found that mutation of the proposed nucleophile of the retaining glycosyltransferase  $\alpha$ 3GT (Glu317 – structurally equivalent to Gln189 of LgtC) to glutamine reduced the observed  $k_{\text{cat}}$  for galactosyltransferase activity 2400-fold (Zhang et al., 2003).

### 3.3. Investigating the Retaining $\alpha$ -(1,2) Mannosyl Transferase Kre2

#### 3.3.1. Background

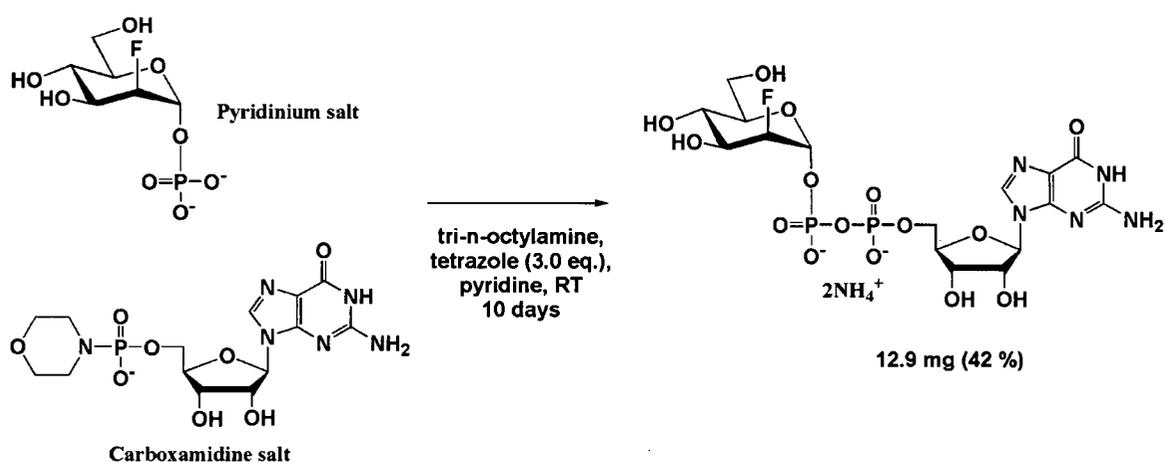
Kre2 (Mnt1p) is an  $\alpha$ -1,2 mannosyltransferase from *Saccharomyces cerevisiae* (EC 2.4.1.131) belonging to family GT15. With overall net retention of anomeric configuration, it catalyses the transfer of mannose to the 2-hydroxyl of various  $\alpha$ -D-mannopyranosides using GDP Man as the donor substrate (Scheme 3.3). A type II Golgi membrane glycoprotein, Kre2 is involved in the mannosylation of both N- and O-linked glycoproteins. Three-dimensional X-ray crystal structures of this enzyme have been reported with either GDP or GDP and the acceptor methyl  $\alpha$ -D-mannopyranoside bound in the active site (Lobsanov et al., 2004). Kre2 has the overall architecture of the GT-A fold and is therefore classified amongst Clan III glycosyltransferases. Using the ternary complex structure of LgtC (pdb 1ga8) as a reference, the donor mannose moiety was modelled into the active site of Kre2 structure with bound GDP. Based on this model, Tyr220 was proposed to act as the catalytic nucleophile in a double displacement mechanism. Kinetic analysis of the Y220F mutant revealed a  $k_{cat}$  value 3000-fold lower than that of the wild-type enzyme, which the authors interpreted as support for this catalytic role. A nucleophilic tyrosyl hydroxyl had been reported for a trans-sialidase just prior to the publication of this Kre2 structural study (Watts et al., 2003).



**Scheme 3.3.** Mannosylation reactions catalysed by Kre2. *In vitro* studies indicate acceptors with R = H, methyl, or  $\alpha$ -D-mannose to be efficiently utilized by Kre2 (Lobsanov, 2004).

### 3.3.2 Synthesis of a Non-Reactive Donor Substrate Analogue

To overcome the problem of donor substrate hydrolysis, the non-reactive analogue GDP 2F Man was synthesized using traditional morpholidate coupling methodology (Scheme 3.4). The ability of this compound to act as either a slow substrate, an inactivator or a competitive inhibitor of Kre2 was then tested. In addition, GDP 2F Man was subjected to co-crystallization trials with the enzyme.



**Scheme 3.4.** Synthesis of the non-reactive donor substrate analogue GDP 2F Mannose via a morpholidate coupling reaction.

### 3.3.3 Kinetic Analysis of Kre2

Kinetic analysis of Kre2 varying the concentration GDP Man at a fixed saturating concentration of methyl  $\alpha$ -D-mannoside acceptor substrate, using the PK/LDH continuous enzyme-coupled assay, yielded  $K_M$  and  $k_{cat}$  values of 110  $\mu$ M and 28  $s^{-1}$  respectively. These values are in approximate agreement with those obtained using a stopped radio-labelled assay (Lobsanov et al., 2004). Consistent with previous findings for the reactivity of nucleotide 2-fluoro sugar derivatives with glycosyltransferases, GDP 2F Man proved not to be a substrate for Kre2 and was found to act as a competitive inhibitor with respect to the GDP Man donor substrate with an observed  $K_i$  value of 470  $\mu$ M. Time dependent inactivation of Kre2 activity, indicative of covalent modification associated with the accumulation of a covalent glycosyl-enzyme intermediate, was not observed.

### 3.3.4 Attempts to Observe a Covalent Glycosyl-Enzyme Intermediate with Kre2

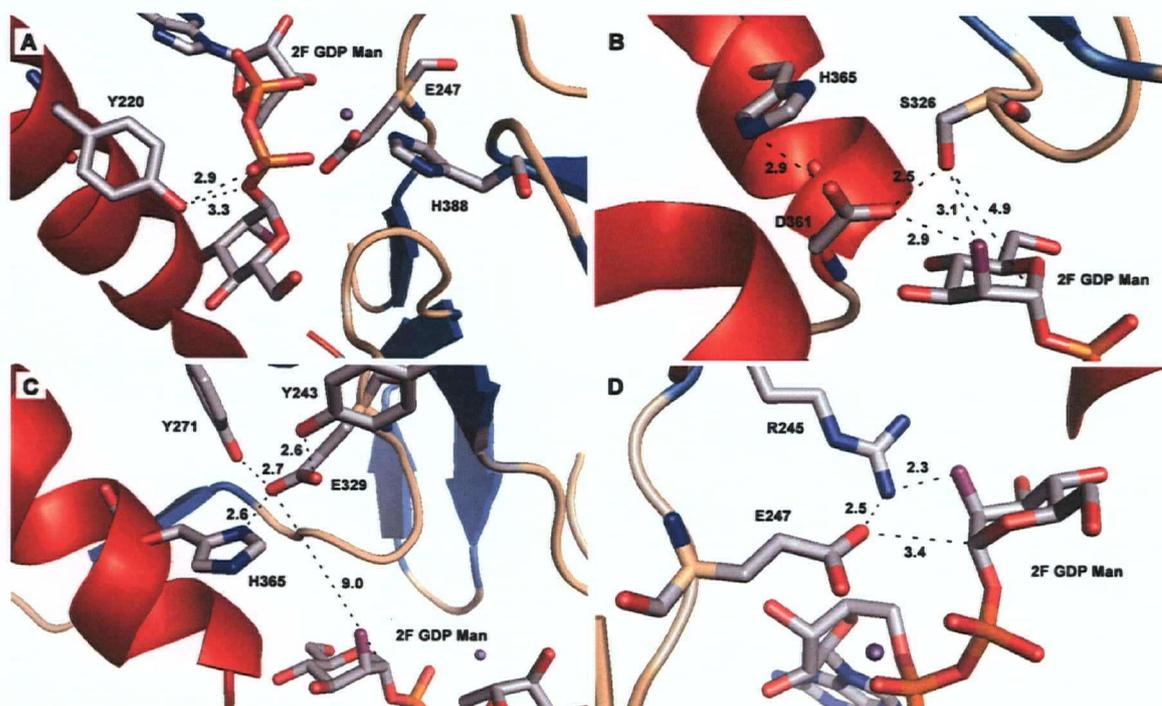
In an effort to observe a covalently bound glycosyl-enzyme intermediate, thereby providing support for the proposed double displacement mechanism, Kre2 was subjected to ESI-MS analysis in the presence or absence of various combinations of saturating concentrations of donor and acceptor substrates. Presumably due to the problem of heterogeneous glycosylation, the catalytic domain of wild-type Kre2 did not give an ESI-MS signal that could be deconvoluted to obtain an accurate mass measurement. Comparative peptide mapping of tryptic and peptic digests of Kre2 incubated under a variety of conditions, in which incubation time, quenching pH and combinations of

saturating concentrations of GDP Man, GDP 2F Man and methyl  $\alpha$ -D-mannose were varied (see methods section), did not reveal the formation of a covalent glycosyl-enzyme intermediate species.

### **3.3.5 X-Ray Crystal Structure of Kre2 with Bound GDP 2F Mannose**

In collaboration with Dr. Yuri Lobsanov in Lynne Howell's laboratory at Toronto's Hospital for Sick Children, a co-crystal structure was obtained for Kre2 with the donor substrate analogue GDP 2F Man bound within the active site. This structure, with the intact substrate analogue bound within the active site, reveals that the previously proposed nucleophilic side chain of Tyr220 is in fact situated on the alpha face of the mannose sugar ring plane (Figure 3.9A). This indicates that the described previous interpretation based on modeling was ill conceived. In order for this residue to play the role of catalytic nucleophile in a double displacement mechanism, it would need to attack the anomeric reaction centre from the  $\beta$ -face of the sugar ring in order to form the proposed beta-linked covalent glycosyl-enzyme intermediate. From this structure it is apparent that the side chain hydroxyl of Tyr220 is within hydrogen bonding distance of both the bridging and a non-bridging oxygen of the phosphate leaving group. As such, a more plausible catalytic role for this residue is that of electrostatic stabilization that facilitates leaving group departure, as was similarly observed for Tyr156 and Tyr162 in the case of CstII (Chiu et al., 2004). In the case of Cst II, these tyrosyl hydroxyl groups were proposed to stabilize departure of the CMP leaving group and mutation of either of these residues to Phe resulted in significant impairment of both transferase and hydrolase

activities, albeit to a less severe degree than that observed with the Y220F mutation of Kre2.



**Figure 3.9.** Kre2 active site with bound intact donor substrate analogue GDP 2F Man. Potential candidate catalytic nucleophiles Y220 (A), S326 (B), E329 (C) and E247 (D) are highlighted as discussed in the text.

Based on the complexed crystal structure with intact donor substrate analogue bound, amino acid side chains can be identified that exist on the beta face of the mannose sugar ring lying closest to the anomeric reaction centre in a position to act as a potential catalytic nucleophile. The side chain hydroxyl of Ser326 is situated 4.9 Å from the reaction centre within hydrogen bonding distance of the 2-hydroxyl of mannose and the side chain carboxyl group of Asp361 (Figure 3.9B). Interestingly, somewhat reminiscent of the classical catalytic triad motifs of serine proteases, albeit in a differing order, the

side chain carboxylate of Asp361 in turn forms hydrogen bonding interactions with the side chain imidazole of His365. Based simply on structure, it could be proposed that the enzyme uses an uncharged side chain hydroxyl as a nucleophile, which is activated by a neighboring carboxylate residue, to attack the charged GDP Man substrate. This proposal is reminiscent of the finding that a similarly activated tyrosyl hydroxyl serves as the nucleophile of a trans-sialidase that acts on a charged sialic acid-containing substrate (Watts et al., 2003).

The Kre2 structure in complex with GDP 2F Man also reveals the identity of two possible carboxylate nucleophiles. The side chain carboxylate of Glu247 is close (3.4 Å) to the anomeric reaction centre making it a possible candidate, however, the likelihood of a nucleophilic role for this residue is limited by several facts. Firstly, this residue is a component of the DXD motif that is responsible for coordination of the obligate  $Mn^{2+}$  ion within the active site of the enzyme. In addition, in the observed conformation, the carboxylate is not configured at an angle appropriate for nucleophilic attack on the reaction centre (Figure 3.9D). It is difficult to envisage a conformational change that would allow a loss of coordination of  $Mn^{2+}$  and reconfiguration to an angle that would facilitate nucleophilic attack. Finally, the side chain carboxylate of Glu329 of Kre2 is located 9.0 Å from the anomeric reaction centre of the donor substrate in a similar relative orientation and at a similar distance to that of Asp190 of LgtC (Figure 3.9C). Also analogous to Asp190 of LgtC (Figure 3.6), the side chain carboxylate of Glu329 hydrogen bonds to two tyrosyl hydroxyl groups.

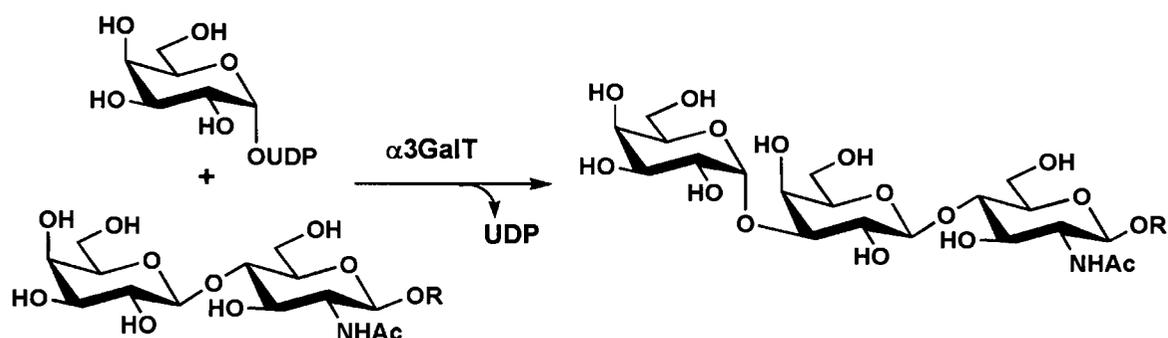
As has been the case with the majority structural investigations with retaining glycosyltransferases, crystallographic analysis of Kre2 with bound intact donor substrate

analogue GDP 2F Man does reveal the clear identity of a potential catalytic nucleophile required by the proposed double displacement mechanism. As discussed in section 1.3.5.4., the only structurally conserved motif within the active sites of retaining glycosyltransferases that adopt the GT-A fold, on the  $\beta$ -face of the donor sugar substrate, is that of a side chain carboxylate located within hydrogen bonding distance of donor sugar hydroxyls and a positively charged side chain donated by either a lysine or arginine residue, which is in turn involved in hydrogen bonding interactions with a side chain carboxylate derived from a residue of the DXD motif. This motif is observed with Kre2, consisting of the side chains of Asp361, Arg245, and the DXD side chain of Glu247 (Figure 3.9B and 3.9D). However, unlike most other glycosyltransferases wherein the carboxylate hydrogen bonds with the C3, C4 or C6 hydroxyls, in the case of Kre2 the carboxylate of Asp361 interacts with what would be the C2 hydroxyl of the mannose donor. Another unique feature is the observation that the side chain of Arg245 is located in very close proximity to the anomeric reaction centre.

## 3.4. ESI-MS Analysis of the Retaining $\alpha$ -(1,3) Galactosyl Transferase $\alpha$ 3GalT

### 3.4.1. Background

The Golgi body-residing type II transmembrane family GT6 galactosyltransferase  $\alpha$ 3GalT (EC 2.4.1.87) is expressed in New World primates and many non-primate mammals (Galili and LaTemple, 1998). Using UDP Gal as donor substrate,  $\alpha$ 3GalT catalyses the transfer of galactose to the 3'-hydroxyl of terminal LacNAc-containing acceptors with overall net retention of anomeric configuration (Scheme 3.5).

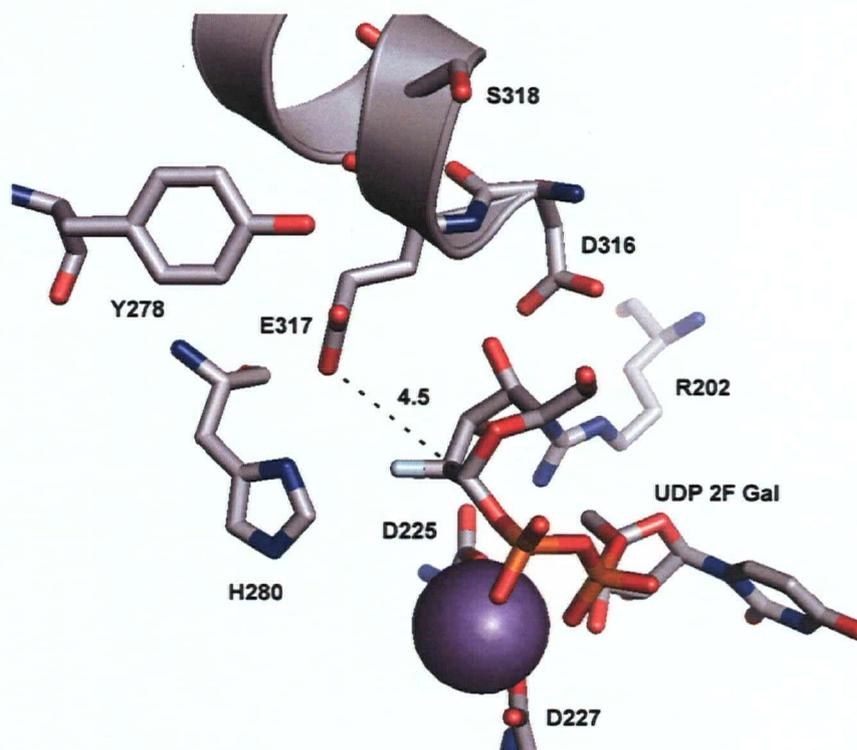


**Scheme 3.5.** Galactosylation reaction catalysed by  $\alpha$ 3GalT.

The bovine version of this enzyme has been the subject of various structural and mechanistic investigations. Based on an initial three-dimensional X-ray crystal structure, Glu317 (structurally equivalent to Gln 189 of LgtC) was proposed to act as the catalytic nucleophile in a double displacement mechanism (Gastinel et al., 2001). In fact, in this initial report, the authors claimed to observe a covalently bound intermediate species. However, a lack of order in the active site region and the presence of two crystal forms for the structure containing the proposed covalent intermediate led to a general lack of

acceptance of this interpretation. Later, a detailed study of the binding affinities of wild-type and E317Q enzyme for donor and acceptor substrates led the authors to propose that this residue is required for proper acceptor substrate orientation and does not play the role of a catalytic nucleophile (Zhang et al., 2003). However, this same study revealed that mutation of Glu317 led to a 2400-fold decrease in the observed turnover rate, indicating the critical importance of this residue (Zhang et al., 2003). Very recently, a structure with the non-reactive donor substrate analogue UDP 2F Gal bound within the active became available and, like the initial structural report, also revealed the side chain of Glu317 to be most suitably positioned to play the role of a catalytic nucleophile (Jamaluddin et al., 2007) (Figure 3.10). Also quite recently, it has been shown that the activity of the E317A mutant of  $\alpha$ 3GalT can be rescued by the addition of small exogenous nucleophiles like azide and a  $\beta$ -azido derivative was isolated and characterized (Monegal and Planas, 2006). As described in section 2.5.2, this chemical rescue strategy has been successfully employed to identify the acid/base and nucleophile catalytic residues of retaining glycosidases. As such, the chemical rescue of Glu317 is suggestive that this side chain is suitably positioned to play such a catalytic role in the mechanism of  $\alpha$ 3GalT. However, it must be kept in mind that, while chemical rescue does indicate that a side chain is suitably positioned to play a catalytic role, it does not prove that it plays such a role in the natural enzyme mechanism. It is not surprising that a mutant enzyme is found to act as a scaffold that catalyses an unnatural reaction simply by proximity effects, when the mutation leads to the introduction of a space that can accommodate a good nucleophile in proximity to an electrophilic center. Chemical rescue results alone cannot be used to deduce the natural catalytic mechanism of an enzyme. Chemical rescue experiments

prove particularly useful in the absence of structural information or, perhaps more relevant to retaining glycosyltransferases, if the catalytic significance of structural information is in question.



**Figure 3.10.** Active site of  $\alpha$ 3GalT with bound non-reactive donor substrate analogue UDP 2F Gal highlighting the position of Glu317 suitably positioned to play the role of a catalytic nucleophile in a double displacement mechanism (pdb 1jcf).

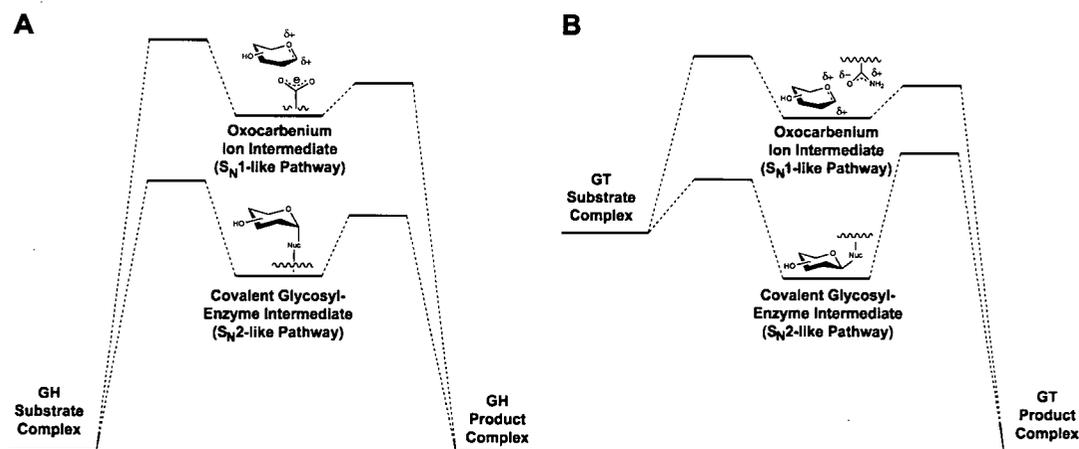
### 3.4.2. Attempts to Observe a Covalent Glycosyl-Enzyme Intermediate with $\alpha$ 3GalT

In light of the recently described chemical rescue of the E317A mutant of bovine  $\alpha$ 3GalT, providing potential support for the notion that this residue might play the role of a catalytic nucleophile in the proposed double displacement mechanism, it was deemed

prudent to determine if a covalent glycosyl-enzyme intermediate could be observed by ESI-MS. The catalytic domain was cloned as a MalE fusion protein in the laboratory of Dr. Warren Wakarchuk at the National Research Centre Institutes for Biological Sciences in Ottawa. This fusion protein was readily expressed in *E. coli*, leading to the production of active and homogenous enzyme. This non-glycosylated enzyme gave an ESI-MS signal that could be deconvoluted to obtain a mass of 75,172 Da, consistent with the theoretical mass of 75,163 Da for the MalE fusion version of the protein. However, despite exhaustive efforts using a range of labelling conditions (incubation time, quenching pH, combinations of donor and acceptor substrates), a covalent glycosyl-enzyme intermediate species of  $\alpha$ 3GalT was not observed. In addition, comparative peptide mapping of tryptic and peptic digests of  $\alpha$ 3GalT incubated with or without UDP Gal under a variety of conditions did reveal the formation of a glycosyl-enzyme adduct.

### 3.5. Reasons Why Retaining Glycosyltransferases May Have Evolved a Mechanism Not Involving Nucleophilic Catalysis

By contemplating the inherent differences in the reactivities of the substrates utilized by glycosidases and glycosyltransferases, a rationale can be developed as to why these two classes of enzymes may have evolved differing mechanistic strategies for catalysing glycosyl group transfer reactions with net retention of anomeric configuration. Glycosidases catalyse the hydrolysis of one of the most stable linkages found in nature, the glycosidic bond between two sugars, with the half lives for spontaneous hydrolysis being on the order of ~5 million years (Wolfenden et al., 1998). The stability of this linkage, under neutral conditions, is imparted by the relatively poor leaving group ability ( $pK_a \sim 14$ ) of natural glycosidase substrates. This stability leads to a large energy barrier to bond cleavage in the first step of the glycosidase reaction (Figure 3.11A). It would seem logical that retaining glycosidases would have evolved a catalytic mechanism involving an intermediate, and in accordance with the Hammond postulate therefore a transition state, of the lowest possible free energy. In other words, by selecting a mechanism with a lower energy intermediate species, catalytic efficiency is easier to achieve because the rate determining transition state being stabilized will also have a lower free energy. By this logic, the Koshland  $S_N2$ -like double displacement mechanism, involving the formation of a covalently bound glycosyl-enzyme intermediate, would have been preferred over the Phillips  $S_N1$ -like mechanism, involving the formation of a higher energy oxocarbenium ion intermediate, during the evolution of retaining glycosidase mechanism (Figure 3.11A).



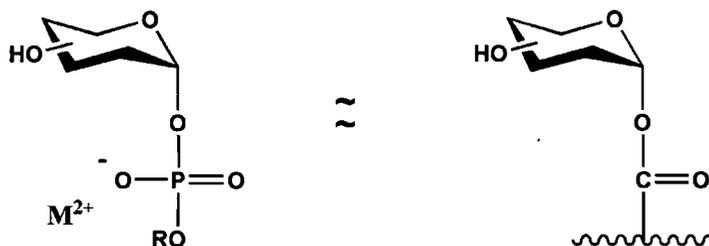
**Figure 3.11.** Comparison of the potential differences in the free energy barriers for  $S_{N1}$ -like and  $S_{N2}$ -like pathways that could be utilized by (A) retaining glycosidases (GH) and (B) retaining glycosyltransferases (GT). Because glycosyltransferases utilize high energy (substituted) phospho-sugar donor substrates containing a very good leaving group (pKa of  $\sim 2$  for the coordinated species), the free energy barrier for the first step is much less than that for the glycosidases that utilize a very stable substrate containing a poor leaving group (pKa of  $\sim 14$ ). This smaller free energy barrier for the first step of the reaction may have led to the  $S_{N1}$ -like pathway being sufficiently effective to be selected for during the course of the evolution of retaining glycosyltransferase mechanisms. In addition, for glycosyltransferases, because of the near equality in leaving group ability for both steps, utilization of a side chain carboxylate as a catalytic nucleophile (as is the case for the vast majority of retaining glycosidases) in a double displacement mechanism might lead to the deglycosylation step of the reaction being rate limiting. This would increase the probability of unwanted donor substrate hydrolysis.

In contrast, glycosyltransferases utilize high energy (substituted) phospho-sugar donor substrates. The leaving group ability for the donor substrate is much greater to start with, compared to glycosidase substrates, with a pKa of  $\sim 7$  for the second ionisation of phosphate. In addition, it is easy to imagine glycosyltransferases readily increasing this

leaving group ability electrostatically, using a coordinated divalent cation or two positively charged amino acid side chains (as discussed in sections 1.3.5.4. and 1.3.5.5.), to pKa values of  $\sim 2$  (the pKa<sub>1</sub> value for a phosphate monoester). Because of the higher free energy of the starting material and lower free energy of the first transition state, the inherent barrier for bond cleavage in the first step of the glycosyltransferase reaction is less than that of the glycosidase reaction (Figure 3.11B). Because there is not such a large barrier for the initial bond cleavage step compared to that for the analogous glycosidase reaction, an S<sub>N</sub>1-like pathway involving the formation of an intermediate oxocarbenium ion may have been accessible during the course of the evolution of the mechanisms of retaining glycosyltransferases and sufficient to have been selected for (Figure 3.11B).

In addition, in order to minimize the unwanted side reaction of donor substrate hydrolysis, the relative free energies of the starting material and intermediates might even suggest that the S<sub>N</sub>2-like pathway may have been selected against. Because of the low energy barrier for the first step of the reaction, formation of a covalent glycosyl-enzyme species could lead to an energy well that would make the second deglycosylation step rate limiting (Figure 3.11B). This can be rationalized on the basis of the relative pKa values for the leaving groups of the starting (substituted) phospho-sugar donor (pKa of  $\sim 2$  for the coordinated species) versus that of a proposed side chain carboxylate acting as a catalytic nucleophile (pKa of  $\sim 4$ ). Indeed, the chemistry involved in the first step of the glycosyltransferase-catalysed reaction is rather analogous to that of the deglycosylation step of a  $\beta$ -retaining glycosidase catalysed reaction (Figure 3.12). In the case of glycosyltransferases, the first reaction consists of carbon-oxygen glycosidic bond

cleavage with an axially-linked (substituted) phosphate leaving group. In the case of a  $\beta$ -retaining glycosidase the second step consists of carbon-oxygen glycosidic bond cleavage with an axially-linked carboxylate leaving group.



**Figure 3.12.** Chemical similarities between the substrate used in the first step of the glycosyltransferase reaction and the covalent glycosyl-enzyme intermediate turned over in the second step of a  $\beta$ -retaining glycosidase catalysed reaction.

An increased lifetime for the intermediate glycosyl donor species could lead to an increased probability for unwanted attack by water leading to donor substrate hydrolysis and increased cellular metabolic demand. At first approximation this argument may seem to violate reactivity/selectivity concepts and product partitioning principles for reactions in free solution. However, this argument is being made *exclusively* for a reaction occurring within the confines of an enzyme active site. In the case of a reaction in free solution, the selectivity of a reaction does in fact decrease with increased reactivity and the partitioning of a reactive intermediate between two competing processes will be the same whether formation of the intermediate is the rate limiting step or not. In the confines of the active site of a retaining glycosyltransferase the intermediate formed following the first bond cleavage step, whether cationic or covalently bound, has three possible fates. Attack by the (substituted) phosphate leaving group will give back starting material, attack by the nucleophilic group of the acceptor will give the desired product,

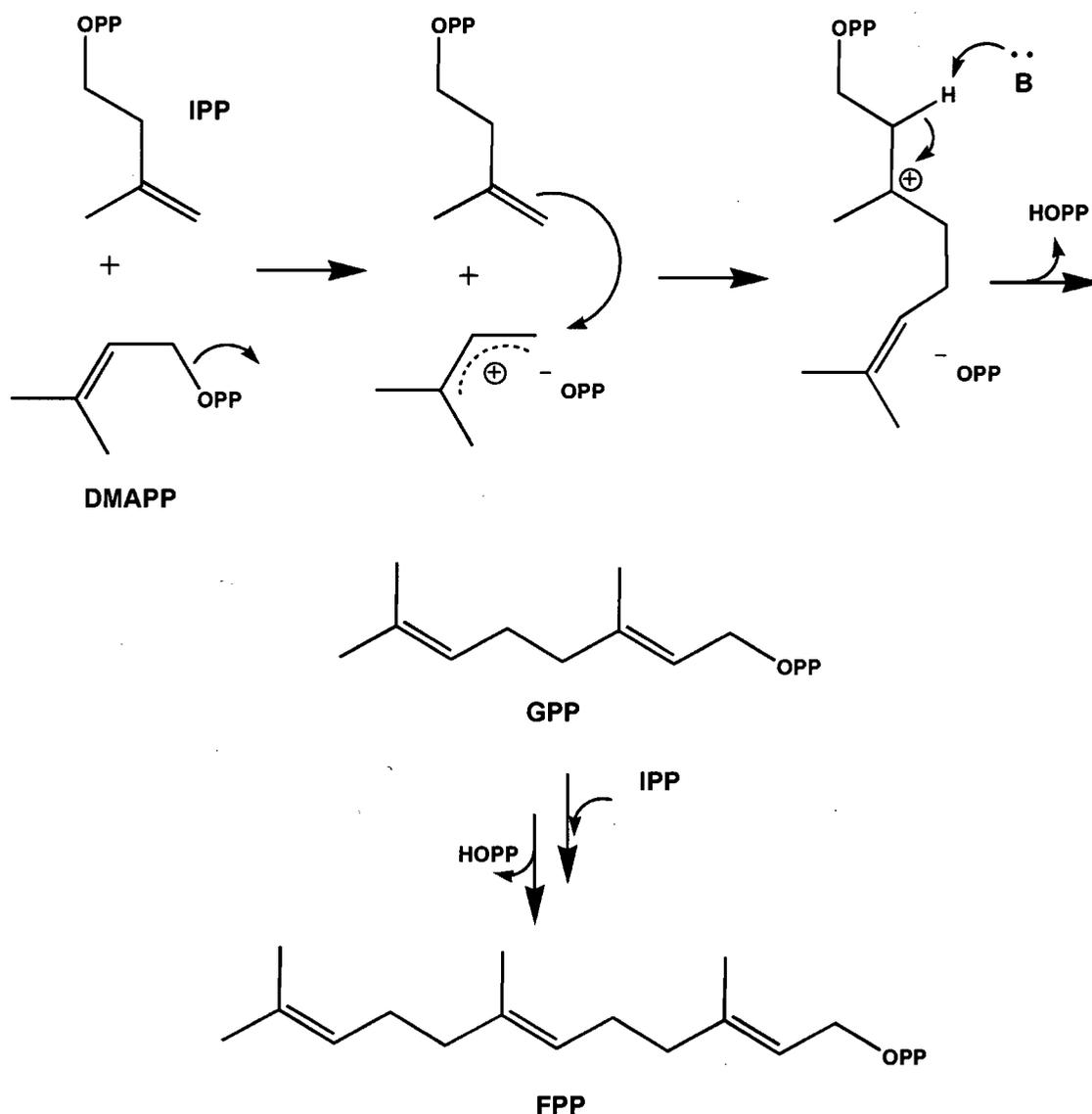
and attack by water will give the undesired hydrolysis product. Clearly, in solution, with a water concentration of 55 M, donor substrate hydrolysis would be the statistically and thermodynamically favoured outcome. However, the enzyme catalyses group transfer in part by suppressing the hydrolysis reaction by excluding water from the active site, thereby decreasing the effective concentration of water. The structural data available for this class of enzymes indicates that this anhydrous microenvironment is achieved in part by using an acceptor as a molecular "plug" that blocks bulk solvent from the region where glycosyl group transfer occurs. Association of the acceptor with the enzyme is a dynamic process with associated on and off rates. An *enzymatic* mechanism involving a longer-lived intermediate would not be able to suppress the hydrolytic reaction as efficiently, because this would increase the probability of the acceptor dissociating from enzyme with bound intermediate thereby leading to an increased effective concentration of water and an increased rate of catalysed hydrolysis. This argument could also be the reason why, if retaining glycosyltransferases were to utilize a double displacement mechanism, an amide group would be the nucleophile of choice due to its better leaving group ability (pKa of  $\sim -0.5$  to  $-1$ ), resulting in a shorter lived intermediate species.

Of course, the pKa value of a functional group, indicative of its leaving group ability for similar chemical types, varies greatly for species bound within an enzyme active site. However, a consideration of the values outside the enzymatic microenvironment indicates the starting points for the evolution of protein function. These are therefore useful values to consider when discussing the evolution of enzyme mechanism.

Precedence for enzymatic mechanisms involving S<sub>N</sub>1-like pathways, and the intermediate formation of cationic species, can be derived from enzymes involved in terpenoid biosynthesis – isopentenyl diphosphate (IPP) isomerases, farnesyl diphosphate (FPP) synthetases, and the terpene cyclases. IPP isomerase catalyses the conversion of IPP to dimethylallyl diphosphate (DMAPP) using Mg<sup>2+</sup> and Zn<sup>2+</sup> in a pathway that has been shown to occur through an S<sub>N</sub>1-like mechanism involving the intermediate formation of a tertiary carbocation intermediate. Support for the intermediate formation of a carbocationic species comes from studies quantitating decreased reactivities for fluorinated analogues of IPP (Muehlbacher and Poulter, 1988) and DMAPP (Reardon and Abeles, 1986) and potent non-covalent inhibition by ammonium-containing analogues that mimic the cationic intermediate species (Muehlbacher and Poulter, 1985, 1988; Reardon and Abeles, 1986). The cationic intermediate is well sequestered from solvent within a highly hydrophobic active site and is stabilized by the presence of an aromatic indole side chain, donated by an appropriately positioned tryptophan residue, via a pi-cation interaction (Durbecq et al., 2001). A similar mechanism, and the use of active site aromatic amino acid side chains to stabilize cationic intermediate species, has been defined for multiple terpene cyclases (Christianson, 2006; Shishova et al., 2007; Starks et al., 1997).

FPP synthetases catalyse the synthesis of the terpenoid biosynthetic C<sub>10</sub> and C<sub>15</sub> building blocks geranyl diphosphate (GPP) and farnesyl diphosphate (FPP), respectively, using the C<sub>5</sub> building blocks IPP and DMAPP (Figure 3.13). The mechanism of FPP synthetases is generally believed to be that of a three-step “ionisation-condensation-elimination” process, as shown in Figure 3.13 for the condensation of IPP and DMAPP

leading to the formation of GPP. Like Cst II, this enzyme is bifunctional in the sense that it also catalyses the synthesis of FPP in a second condensation reaction between the initially formed GPP product and an additional molecule of IPP (Figure 3.13). Evidence supporting the  $S_N1$ -like mechanism shown in Figure 3.13 comes from studies with fluorinated analogues of DMAPP (Poulter and Satterwhite, 1977) and GPP (Poulter et al., 1977, 1978) that undergo, at significantly depressed rates, FPP synthetase-catalysed condensation reactions with IPP. The rate depressions observed for the fluorinated (versus non-fluorinated) species for these enzyme-catalysed processes were compared to the rate depressions observed for the solvolysis of the identically fluorinated (versus non-fluorinated) analogues of the methanesulphonate derivatives of DMAPP and GPP. The solvolyses of these methanesulphonate derivatives are known to occur via  $S_N1$  processes involving the intermediate formation of a cationic species. The finding that a similar degree of rate depression, resulting from introduction of electronically destabilizing fluorine atoms, is observed for enzyme-catalysed and spontaneous solvolysis processes indicates that the FPP synthetase reaction proceeds via a similar  $S_N1$ -like mechanism.



**Figure 3.13.** The reactions catalysed by FPP synthetase. A three-step “ionisation-condensation-elimination” mechanism involves the intermediate formation of cationic species that are stabilized by appropriately positioned carbonyl moieties donated by conserved main chain and side chain amide groups. The majority of characterized retaining glycosyltransferases also have an analogous amide group positioned above the anomeric reaction centre to stabilize the development of positive charge. As with retaining glycosyltransferases, a substrate/product phosphate is believed to play the role of a base catalyst (B). FPP synthetase is bifunctional, being able to catalyse both the condensation of DMAPP with IPP leading to the production of GPP and the condensation of GPP with IPP leading to the production of FPP.

The chemistry involved in FPP synthetase-catalysed reactions is analogous to that involved in glycosyltransferase catalysis. In both cases, a nucleophilic substitution reaction occurs at a carbon centre with a (substituted) phosphate acting as the leaving group. Stereochemically, they are more like inverting enzymes in that they preorganize the nucleophilic alkene in a position to directly trap the carbocation and do not suffer the same stereochemical impediment of the retaining enzymes. However, analogies between the overall chemistries being catalysed provide a foundation for a comparison of the transition states chosen to be stabilized and the possibility of a convergence in mechanism during the course of evolution. As is observed for glycosyltransferases, FPP synthetases use a divalent cation and the positively charged side chains donated from a conserved Arg/Lys pair to stabilize charge development and facilitate departure of the phosphate leaving group (Gabelli et al., 2006; Hosfield et al., 2004). Intriguingly, conserved electrostatic interactions with neutral carbonyl groups, derived from the amides of a main chain (Lys202) and a side chain (Gln241), plus a side chain hydroxyl (Thr203), are positioned within the FPP synthetase active site to stabilize the positive charge distributed over the C1 and C3 atoms of the cationic intermediate (Gabelli et al., 2006; Hosfield et al., 2004). This is similar to the observed positioning of the carbonyl oxygen of a main chain or a side chain amide in close proximity (on the  $\beta$ -face) to the anomeric reaction centre within the active sites of the vast majority of structurally characterized retaining glycosyltransferases (Figures 1.9 and 1.10). In both cases, formation of a cationic intermediate is also stabilized by inherent properties of the substrates. In the case of glycosyltransferases, positive charge formation at the anomeric centre of an oxocarbenium ion is stabilized by a lone pair of non-bonding electrons donated by the

endocyclic oxygen of the sugar ring. In the case of FPP synthetases, an allylic tertiary centre stabilizes positive charge development at the carbon centre that undergoes bond fission. Also, as is proposed for retaining glycosyltransferases (sections 1.3.5.4 and 1.3.5.5.), FPP synthetases appear to situate the phosphate leaving group in such a manner that it is suitably positioned to play the role of the base catalyst (B) required for proton abstraction in the elimination step of the reaction (Figure 3.13). These described similarities in the chemistries catalysed and the observed conservation of structural architecture, suggests that retaining glycosyltransferases utilize an  $S_N1$ -like mechanism analogous to that of FPP synthetase.

### **3.6. Concluding Remarks**

Presently, the catalytic mechanism of retaining glycosyltransferases is a topic of debate with no clear and definitive answers. By analogy to the well-characterized glycosidases, a double displacement mechanism involving discrete nucleophilic catalysis and the formation of a covalently bound glycosyl-enzyme intermediate seems a likely possibility. Such a mechanism necessitates an appropriately positioned enzymatic nucleophile within the active site on the  $\beta$ -face of the donor substrate in close proximity to the anomeric reaction centre. As described in sections 1.3.5.4 and 1.3.5.5, with the exception of enzymes from family GT6 wherein a side chain carboxylate is appropriately situated, the most suitably positioned functional group found within the active sites of structurally characterized retaining glycosyltransferases is usually a main chain or side chain amide group. The results of crystallographic analysis of Kre2 with bound intact donor substrate analogue indicate that, in this case, neither of these functional groups are

found and a side chain hydroxyl from Ser326 is present in the approximate location occupied by the amide or carboxylate functional groups in other retaining enzymes.

In the case of the family GT6 retaining enzyme  $\alpha$ 3GalT, crystallographic analysis reveals a side chain carboxylate donated by Glu317 perfectly positioned within the active site to play the role of a catalytic nucleophile in a double displacement mechanism. The suitability of the positioning of this group for a catalytic role is further supported by chemical rescue experiments with the E317A mutant. The catalytic importance of this residue is supported by kinetic analysis of the E317A mutant, for which turnover is significantly decreased with respect to the wild-type enzyme. However, whether this enzyme utilizes an  $S_N1$ -like or  $S_N2$ -like pathway cannot be deduced by structural, rescue, and kinetic experiments alone. The structural and rescue results serve only to identify candidate catalytic residues and the mechanistic relevance of these studies should never be overstated. The kinetic results with the E317A mutant illustrate the catalytic importance of the Glu317, but do not indicate what that role is. Indeed, if the enzyme were to use an  $S_N1$ -like pathway, this side chain carboxylate could play a critical role in stabilizing the positively charged oxocarbenium ion intermediate and its deletion could lead to the observed decrease in rate. Conversely, the inability to observe a covalently modified version of  $\alpha$ 3GalT by ESI-MS cannot be legitimately used as evidence against the  $S_N2$ -like pathway. These experiments relied on the optimistic expectation that, if the enzyme were to use a pathway involving the formation of the glycosyl-enzyme intermediate, the deglycosylation step of the reaction would be rate-limiting and/or the lifetime of this species would be long enough for it to be observed by MS analysis. As such, the mechanism of  $\alpha$ 3GalT remains unclear.

Crystallographic analysis of the family GT8 retaining enzyme LgtC reveals the side chain amide of Gln189, structurally equivalent to Glu317 in  $\alpha$ 3GalT, to be perfectly situated to play the role of a catalytic nucleophile in a double displacement mechanism. Kinetic analysis of the Q189A mutant revealed an unexpectedly modest decrease in observed rates, inconsistent with a critical catalytic role for this residue. These kinetic data generated considerable doubt as to the likelihood of a double displacement mechanism for LgtC. However, even if the  $S_N1$ -like pathway involving formation of an oxocarbenium ion intermediate were the actual LgtC mechanism, it would be expected that Gln189 would play a critical role in stabilizing the high energy intermediate species and a larger impact on turnover would be expected when this side chain amide is removed. This may suggest that the observed ground state crystal structures are misleading in assigning the active site features essential during the course of reaction.

The observed labelling of the Q189E variant of LgtC-25 represents the first report of the direct observation of a glycosyl-enzyme intermediate covalently bound to the active site of a retaining glycosyltransferase. While it is possible that this is simply an artefact resulting from the creation of an active site mutant, it rekindles the controversy surrounding the mechanism of LgtC, which has served as a model system for the mechanistic study of retaining glycosyltransferases. Although by no means conclusive, the surprising site of labelling implicates Asp190 as an alternative candidate catalytic nucleophile. This notion is supported by kinetic analysis of the D190N mutant and by the catalytic relevance of the intermediate species. If this residue were to serve that function, a conformational change from that observed in the ground state crystal structure would be required during catalysis to allow for an appropriate positioning relative to the donor

sugar substrate. An alternative explanation for the observed labelling with the Q189E mutant could be that LgtC does in fact utilize an  $S_{N1}$ -like pathway and that the altered electrostatic environment of this mutant leads to the highly reactive electrophilic oxocarbenium ion intermediate being quenched by a nearby nucleophilic side chain. Regardless, an active site configuration other than of the observed crystal structure would have to be accessible in order to account for the observed remote site of labelling at the Asp190 position. One last possibility is that LgtC does utilize an  $S_{N1}$ -like mechanism and that the introduced mutation changes the mechanism to that of an  $S_{N2}$ -like pathway. Frustratingly, the exact catalytic mechanism of LgtC remains ill defined.

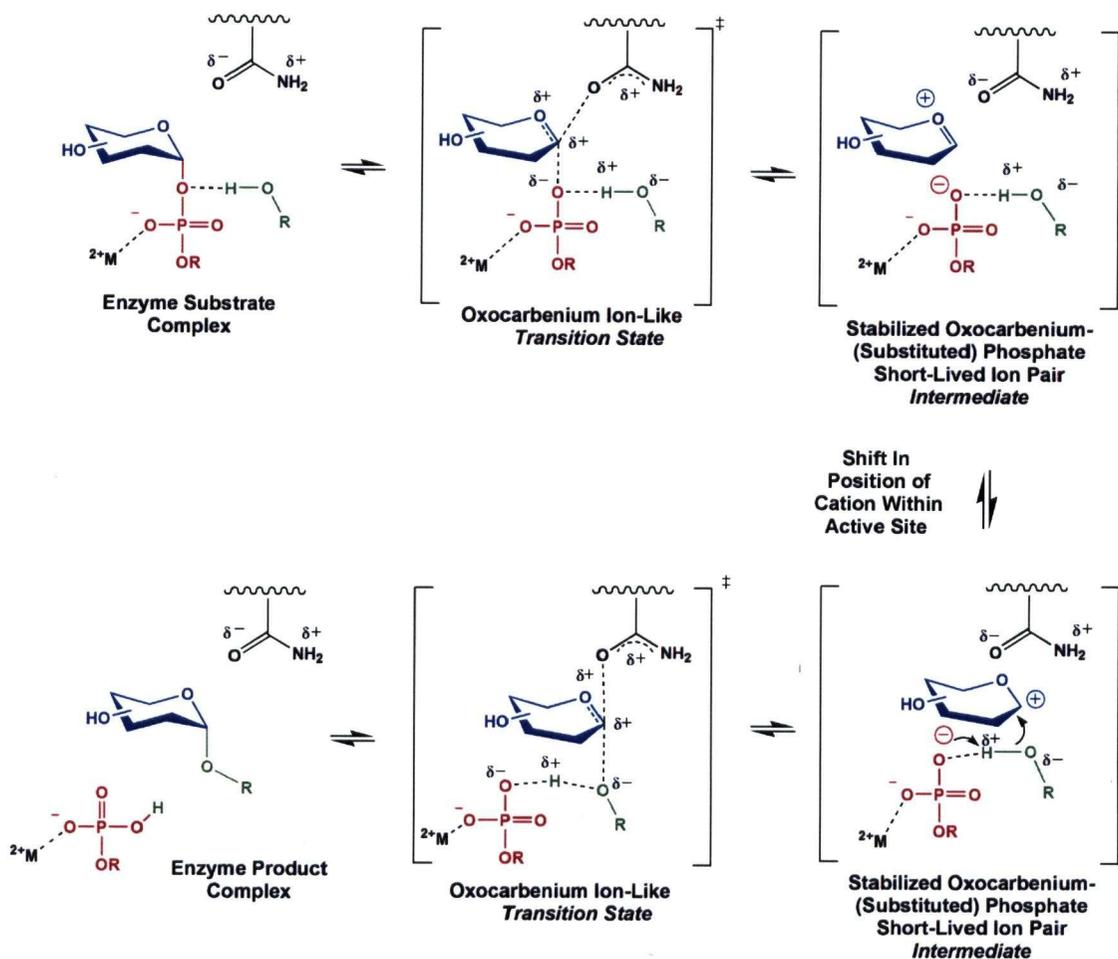
The distinct possibility that retaining glycosyltransferases utilize a mechanism that differs from the double displacement mechanism of the analogous retaining glycosidases must be considered. This notion is supported by the observed lack of conserved structural architecture in the region expected to be occupied by a catalytic nucleophile in the active sites of structurally defined glycosyltransferases (as discussed in sections 1.3.5.4. and 1.3.5.5.), in stark contrast to what has been observed with otherwise unrelated retaining glycosidases, which generally have a side chain carboxylate at this position. However, in the absence of supporting evidence, should a mechanism with little chemical precedence, that contradicts fundamental principles of organic chemistry and is formally forbidden based on orbital symmetry considerations (i.e. the previously proposed  $S_{Ni}$ -like mechanism) be suggested? Is there a more rational alternative? Consideration of the reactivity and associated free energy differences derived from the inherent differences between glycosyltransferase and glycosidase substrates, as well as analogies of chemistry and conserved active site structural features of many retaining glycosyltransferases with

FPP synthetases that use an  $S_N1$ -like mechanism, support the possibility of an ionic mechanism.

It is possible that some retaining glycosyltransferases use a distinctively  $S_N2$ -like mechanism involving formation of a covalently bound glycosyl-enzyme intermediate and others use a distinctively  $S_N1$ -like mechanism involving formation of an ionic oxocarbenium ion intermediate. More likely, however, a mechanistic continuum exists for this class of enzymes analogous to that of non-enzymatic nucleophilic substitution reactions. In non-enzymatic processes, the  $S_N2$  direct displacement and  $S_N1$  ionic mechanisms are simply the extremes of a continuum. These classic terms used to describe the extremes of nucleophilic substitution reactions are defined simply by the order of the reaction. In the extreme ionic mechanism, there is absolutely no interaction of the incoming nucleophile with the reactant in the transition state for leaving group bond cleavage, making it a first order reaction, and nucleophilic attack is possible from either face of the ionic intermediate leading to the formation of a mixture of stereochemical products. In the extreme  $S_N2$  mechanism, the covalent bond formation with the incoming nucleophile and the breaking of bonds to the leaving group occur in a synchronous fashion, with nucleophilic "push" facilitating the reaction, making it a second order reaction and leading to complete inversion of stereochemistry in the products. In between these extremes are the nucleophilic substitution mechanisms, which should more accurately be strictly defined using the IUPAC recommended designations for reaction mechanisms (Appendix A)(Guthrie and Jencks, 1989; Jencks, 1996), involving ion pair formation that can have the kinetic properties of an  $S_N2$  process yet

produce stereochemical outcomes consistent with an  $S_N1$  reaction (and vice versa)(Carey, 1990; Jencks, 1980; Winstein et al., 1956)

One such mechanism, that seems quite likely for many retaining glycosyltransferases including LgtC, is that of an ionization mechanism leading to the formation of a positively charged intermediate species, in intimate contact with an anionic leaving group counter ion and having a lifetime longer than that of a molecular vibration, that requires strong back side nucleophilic participation (without discrete covalent interaction) for formation (Figure 3.14). This ion pair could collapse to give back starting material or, following a slight shift in the active site positioning of the cation, could undergo front side attack by an incoming nucleophile leading to a product with retained stereochemistry. This mechanism would be defined as  $D_N^*A_{Nssip}$ , where the asterisk represents the formation of a short lived intermediate and “ssip” represents the formation of a “solvent separated” ion pair. Although the term “solvent separated” seems unusual in the context of an enzymatic mechanism, it is required to differentiate from an “intimate” ion pair that would not allow for attack from the front face. Indeed, considering the reactivities of the substrates undergoing reaction, the observed lack of conserved structural architecture amongst retaining glycosyltransferase  $\beta$ -face active site regions and the biological precedence for cationic mechanisms in chemically similar systems, such a mechanism, involving the formation of an oxocarbenium ion species stabilized on the back face by an enzymatic functional group and on the front side by the anionic (substituted) phosphate leaving group seems rather attractive for retaining glycosyltransferases.



**Figure 3.14.** Proposed  $D_N^*A_{Nssip}$  nucleophilic substitution reaction for retaining glycosyltransferases.

## **CHAPTER 4**

# **GLYCOSYLTRANSFERASE PROTEIN ENGINEERING**

## 4.1. Summary

The ability to alter the nucleotide specificity of a nucleotide sugar-utilizing glycosyltransferase was examined using a rational engineering strategy with potential general applicability amongst glycosyltransferases. Facile control of glycosyltransferase nucleotide specificity could provide a useful starting point for the development of chemical genetics tools for the *in vitro* exploration of glycosyltransferase activity. Upon identification and replacement of a suitably disposed conserved side chain within the LgtC active site and chemical synthesis of an alternative nucleotide sugar substrate that was thought would complement the new active site configuration, the donor substrate specificity of wild-type and mutant forms of the enzyme was investigated.

For the purpose of the development of a high throughput screen for application in the directed evolution of glycosyltransferases, the ability to display LgtC as a pIII fusion protein on M13 bacteriophage was explored. In addition, it was tested whether phage display procedures could be conducted in the context of *in vitro* compartmentalization using water-in-oil emulsification.

## 4.2. Engineering the Nucleotide Specificity of LgtC

### 4.2.1. Background

As discussed in Chapter 1, the majority of glycosyltransferases utilize nucleotide sugars as donor substrates. Interestingly, relatively few of the possible combinations of pairings of nucleotides and monosaccharides commonly exist in nature (Table 4.1).

**Table 4.1.** Common forms of nucleotide sugar donor substrates found in Nature and utilized by glycosyltransferases (Varki, 1999).

Monosaccharide	Activated Nucleotide Form
Glucose	UDP sugar
Galactose	
<i>N</i> -Acetylglucosamine	
<i>N</i> -Acetylgalactosamine	
Glucuronic acid	
Xylose	GDP sugar
Mannose	
Fucose	
Sialic acid	CMP Sialic Acid

\* Listed are the most common nucleotide sugars. Rare forms such as ADP Glucose are typically found in plants and microbes.

A consequence of this phenomenon is the fact that, despite the existence of various genes encoding glycosyltransferases that are responsible for the transfer of a given monosaccharide to a given acceptor in a spatially and temporally specific context, virtually all glycosyltransferase gene products use an identical nucleotide sugar donor

substrate. For example, sialyl-linkages terminate mammalian glycoconjugate structures and are limited to one of three possible regio-chemistries (alpha-2,3- to galactose, alpha-2,6- to galactose, N-acetylglucosamine, or N-acetylgalactosamine, or alpha-2,8- to another sialic acid). However, despite the relatively low number of possible linkages, at least 20 different sialyltransferases are encoded by the human genome all of which use the same CMP sialic acid donor substrate (Harduin-Lepers et al., 2001). Because of redundancy in acceptor substrate specificity and the fact that each sialyltransferase gene is differentially expressed in a tissue-, cell type- and stage-specific manner, very little is known about the exact role that individual gene products play in normal and pathological physiology *in vivo*. Studies that make use of non-specific inhibitors or gene deletions have yielded either conflicting results or results that are difficult to interpret.

This difficulty in delineating the roles of individual gene products in a stage-specific manner using either chemical or genetic tools individually is a common problem in cell biological studies. Small molecules typically suffer from a lack of specificity when they affect the activity of a population of enzymes that use an identical substrate. Conversely, it is often difficult to use gene-inactivation at the level of DNA to address critical temporal issues of function. Chemical genetics strategies that combine both the advantages in target specificity provided by genetics and the temporal control provided by small molecules are of tremendous value in cell biology research (Alaimo et al., 2001; Holt et al., 2002; Ubersax et al., 2003). Chemical complementation of a mutant form of a glycosyltransferase of altered nucleotide specificity with an alternate nucleotide sugar substrate or an inhibitor with an altered purine/pyrimidine scaffold would provide an

effective starting point for the development of a chemical genetics strategy to explore glycosyltransferase activities.

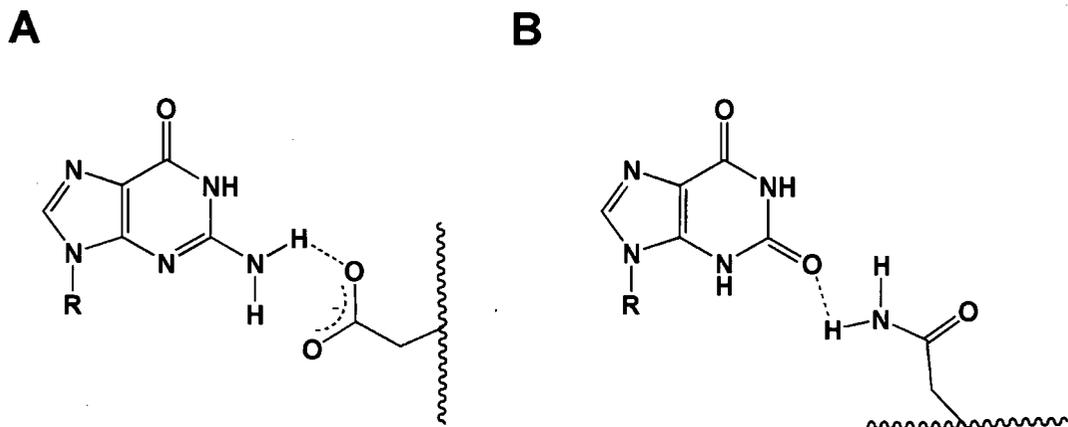
It is generally believed that glycosyltransferases are highly specific for the nucleotide sugar donor substrates that they use. The few studies that have actually explored this issue support such a notion (Baisch et al., 1996; Khaled et al., 2004; Ohrlein, 1999). For example, the GDP fucose-utilizing fucosyl transferase FucT-III was found to utilize ADP fucose ~1000 fold less efficiently in terms of  $k_{cat}/K_M$  (Khaled et al., 2004). A very recent study, describing a GT-B fold polyketide glucosyltransferase (SorF) that possesses promiscuous donor substrate specificity, has challenged this notion (Kopp et al., 2007). Despite a lack of detailed kinetic analysis, the authors report the ability of SorF to utilize dTDP glucose as efficiently as the natural donor substrate UDP glucose. Importantly for the engineering strategy described below, the nucleoside moieties of these two substrates differ only by the presence of a methyl substituent on the pyrimidine ring.

Previous examples of rational protein engineering approaches to broadening or changing the substrate specificities of glycosyltransferases have involved modification of residues involved with the binding of substituents on the sugar ring. For example, the R228K (Ramakrishnan et al., 2005) and Y289L (Ramakrishnan and Qasba, 2002) mutations to lactose synthase were found to relax donor specificities with respect to UDP Gal/UDP Glc and UDP Gal/UDP GalNAc relationships, respectively. Computational modeling was used to identify an H308R mutant of GlcAT-I that was found experimentally to utilize UDP Glc, UDP Man, and UDP GlcNAc with nearly identical efficiency of that of the natural UDP GlcA donor (Ouzzine et al., 2002). Structural and sequence data were used to significantly increase the ability of ToxB, for which the

natural donor substrate is UDP Glc, to utilize an alternative UDP GlcNAc donor substrate (Jank et al., 2005). Similar examples exist for GT-B fold glycosyltransferases (Hoffmeister et al., 2001; Hoffmeister et al., 2002; Kubo et al., 2004). Glycosyltransferase engineering was the topic of an excellent recent review to which the reader is directed (Hancock et al., 2006).

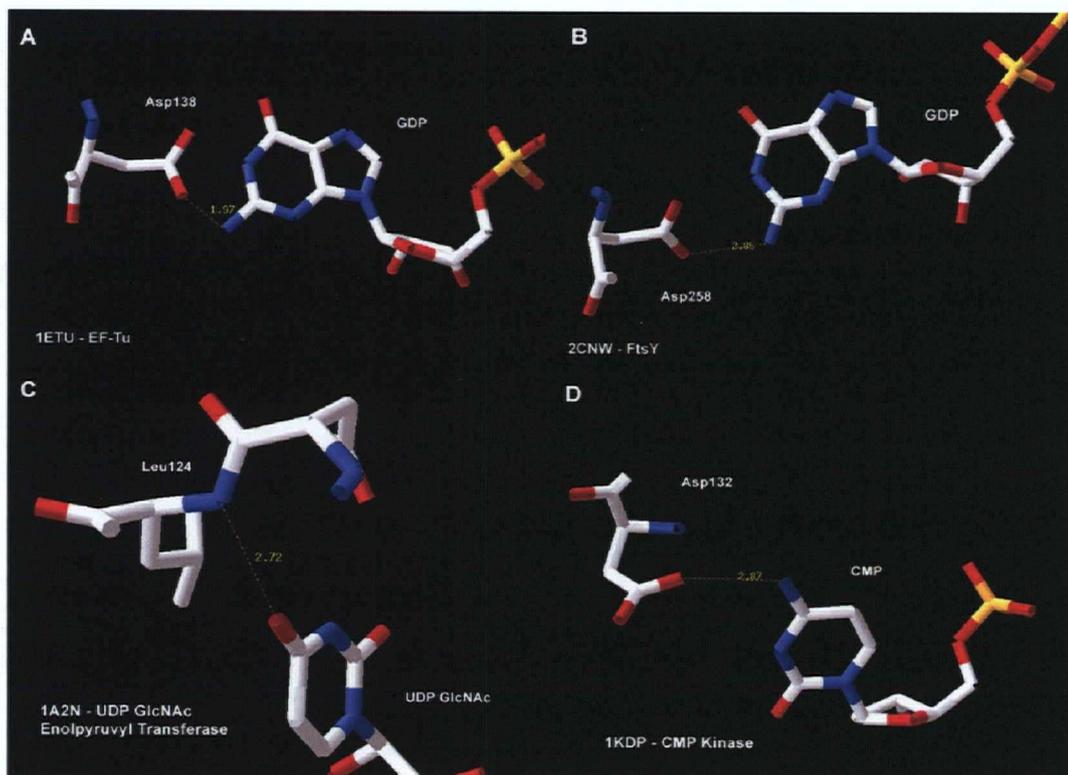
#### **4.2.2. Mutation of a Conserved Active Site Side Chain Alters the Base Specificity of GTP Binding Proteins**

To explore the base specificity of the GTP binding protein elongation factor Tu (EF-Tu), Hwang and Miller generated an array of point mutants and tested the ability of these variants to bind alternative bases (Hwang and Miller, 1987). The most interesting mutation identified was that of D138N, which was found to induce a complete change in base specificity from guanosine to xanthosine. These two bases differ only by the presence of an amino group in the case of guanosine versus a carbonyl substituent in the case of xanthosine at the 2-position of the purine ring (Figure 4.1). The ability to control the nucleotide specificity of EF-Tu was subsequently used to develop a detailed model for the interaction of EF-Tu and the ribosome (Weijland and Parmeggiani, 1993). It was rationalized that the switch in specificity for the D138N mutant results from a change in hydrogen bonding potential. In the case of the wild-type enzyme, the side chain carboxylate of Asp138 can act as a hydrogen bond acceptor and the 2-amino group of guanosine a hydrogen bond donor. Conversely, in the D138N mutant the side chain amide of N138 can act as a hydrogen bond donor and the 2-carbonyl of xanthosine as a hydrogen bond acceptor (Figure 4.1).



**Figure 4.1.** Hydrogen bonding interactions between guanosine and a side chain carboxylate (A) or xanthosine and a side chain amide (B) determine the base specificity of EF-Tu and FtsY.

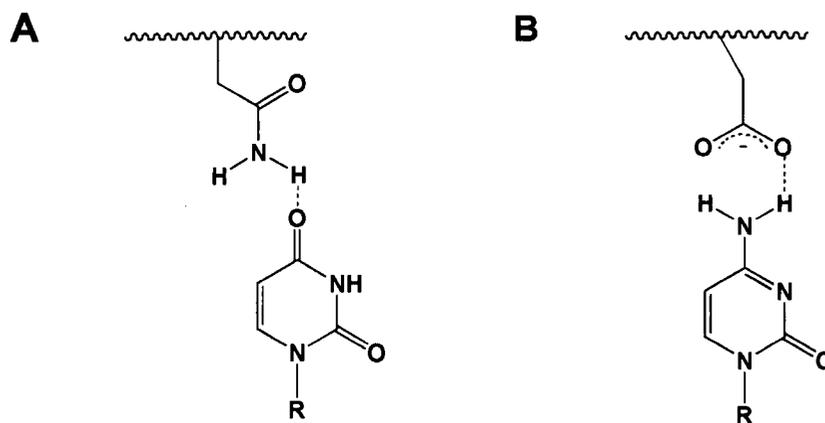
This supposition is supported by the three-dimensional X-ray crystal structure of GTP bound to EF-Tu, which clearly shows the side chain carboxylate of D138 within hydrogen bonding distance of the 2-amino group of guanosine (Figure 4.2A). In fact, this aspartate residue is completely conserved amongst GTP binding proteins. The generality of this approach for the rational control of nucleotide specificity was subsequently demonstrated by successful application to the study of the *E. coli* homologue of the eukaryotic signal recognition particle receptor FtsY (Powers and Walter, 1995). FtsY contains a GTP binding domain for which the nucleotide specificity could be altered from guanosine to xanthosine by the simple introduction of a D441N point mutation. Again, the hydrogen bonding interaction between the side chain carboxylate of D441 and the 2-amino group of guanosine is supported by the X-ray crystal structure of the *T. aquaticus* homologue of FtsY with bound GDP (Figure 4.2B).



**Figure 4.2.** Hydrogen bonding interactions between guanosine and a side chain carboxylate that are found to control base specificity between guanosine and xanthosine in the case of EF-Tu (A) and FtsY (B). Similar hydrogen bond partners are found with other nucleotide binding proteins. The 4-carbonyl oxygen of UDP GlcNAc hydrogen bonds with a main chain amide in the case of UDP Glc NAc enolpyruvyl transferase (C). The 4-amino group of CMP hydrogen bonds with a side chain carboxylate in the case of CMP kinase (D). Protein Data Bank ID numbers are indicated and distance measurements are reported in Angstroms.

It would seem reasonable that this approach of using the hydrogen bonding partnering of a carboxylate with an amino group or an amide with a carbonyl group could be extended to control the nucleotide specificity of other proteins. In the case of glycosyltransferases, a useful application would be the ability to alter the specificity of UDPbinding domains (Table 4.1). The most logical choice for base switching would be

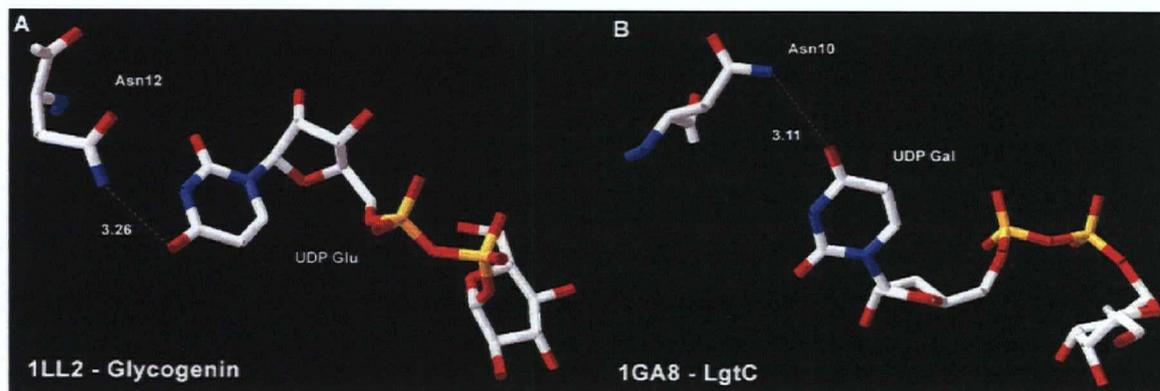
from uridine to cytosine. It would be predicted that in the case of uridine binding proteins an amide from the protein would form a hydrogen bond interaction with the 4-carbonyl oxygen of the uridine ring (Figure 4.3A). Inspection of the three-dimensional X-ray crystal structures of various uridine binding proteins reveals that such an interaction is common (Kostrewa et al., 2001; Lariviere et al., 2003; Skarzynski et al., 1998; Thoden et al., 1997)(for example Figure 4.2C). Conversely, in the case of cytosine binding proteins, it would be predicted that a side chain carboxylate would form a hydrogen bond interaction with the 4-amino group of the cytosine base (Figure 4.3B). Again, inspection of structures of cytosine binding proteins with bound ligand reveals the occurrence of such an interaction (Bertrand et al., 2002)(for example Figure 4.2D). Based on the occurrence of these interactions it would be predicted that the nucleotide specificity of a UDP sugar-utilizing glycosyltransferase could be readily altered to that of a CDP sugar-utilizing enzyme by changing a single active site amide side chain to a carboxylate. Using the UDP Gal utilizing enzyme LgtC as a model system, the feasibility of this nucleotide specificity switch approach was investigated.



**Figure 4.3.** Hydrogen bonding interactions between uridine and a side chain amide (A) or cytosine and a side chain carboxylate (B) proposed to determine base specificity.

### 4.2.3. Identification and Modification of a Base Specificity Conferring Site Within the LgtC active site

Fortunately, in the case of LgtC a three-dimensional X-ray crystal structure of the enzyme is available with UDP 2F Gal bound within the active site (Persson et al., 2001). In addition, the structure of glycogenin, another GT family 8 member, is available with bound UDP Glc (Gibbons et al., 2002). Inspection of both of these structures reveals the occurrence of a conserved interaction between the 4-carbonyl group of the uridine ring with a side chain amide within the UDP binding domain (Figure 4.4).

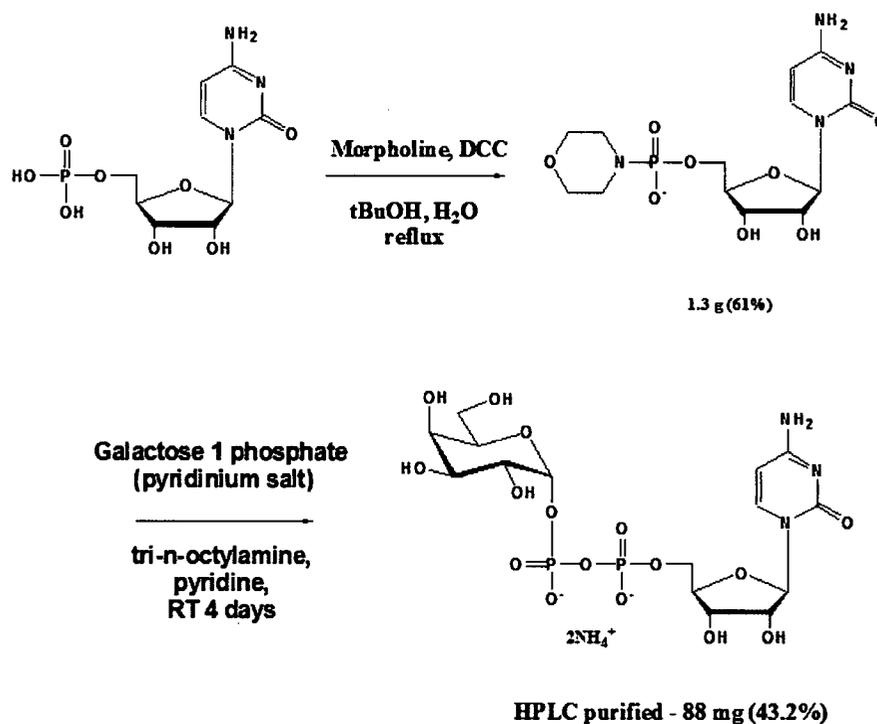


**Figure 4.4.** Active sites of the GT family 8 members glycogenin (A) and LgtC (B) revealing the conserved interaction of a side chain amide with the 4-carbonyl group of the bound uridine ring. Protein Data Bank ID numbers are indicated and distance measurements are reported in Angstroms.

It was therefore proposed that modification of the side chain amide of Asn10 of LgtC to a carboxylate would switch the nucleotide specificity of the enzyme from a UDP Gal to a CDP Gal utilizing enzyme. The N10D point mutation was introduced into LgtC and the resulting variant was expressed in *E. coli* yielding quantities of protein similar to that of the wild-type enzyme.

#### 4.2.4. Synthesis of CDP Gal

Initial studies with a UDP Gal synthesizing pyrophosphorylase enzyme from *Streptococcus pneumoniae* (spGalU)(Mollerach et al., 1998) indicated that at high enzyme and nucleotide concentrations, CTP could be substituted for UTP to generate CDP Gal. However, under preparative scale conditions, presumably due to issues with product inhibition, sufficient quantities of the desired product could not be generated using either UTP or CTP. Therefore, the unnatural nucleotide sugar CDP Gal had to be chemically synthesized using a traditional morpholidate coupling protocol. CMP morpholidate is not commercially available and had to be generated prior to coupling with  $\alpha$ -D-galactose 1-phosphate (Scheme 4.1). Following HPLC purification, the desired nucleotide sugar substrate was obtained in yields typical for this coupling technique.



**Scheme 4.1.** Chemical synthesis of the desired unnatural nucleotide sugar CDP Gal.

#### 4.2.5. Testing the Nucleotide Sugar Specificity of Wild-type and N10D LgtC

Analysis of analytical scale reactions by TLC indicated that at high enzyme concentrations (~0.5 mg/mL) both the N10D and wild-type versions of LgtC were able to use CDP Gal as an alternative donor sugar substrate. However, it was found that under similar conditions the N10D version of LgtC was able to use the natural donor substrate UDP Gal. Thus it appeared that specificity had been broadened rather than switched. To determine whether a true and useful change in nucleotide sugar preference had occurred as a result of the amide to carboxylate mutation, detailed kinetic analysis of the wild-type and N10D versions of LgtC using either UDP Gal or CDP Gal as the donor substrate was performed using the PK/LDH enzyme-coupled continuous assay (Table 4.2).

**Table 4.2.** Michaelis-Menten kinetic parameters for wild-type and N10D versions of LgtC using either CDP Gal or UDP Gal as the varied substrate at a fixed saturating concentration of lactose acceptor (160mM).

Donor	WT LgtC-25			N10D LgtC-25		
	$k_{cat}$ (min <sup>-1</sup> )	$K_m$ ( $\mu$ M)	$k_{cat}/K_m$ (min <sup>-1</sup> .mM <sup>-1</sup> )	$k_{cat}$ (min <sup>-1</sup> )	$K_m$ ( $\mu$ M)	$k_{cat}/K_m$ (min <sup>-1</sup> .mM <sup>-1</sup> )
UDP Gal	960	25	38	250	90	3.0
CDP Gal	n.d.	n.d.	1.2	n.d.	n.d.	0.25
UDP Gal/ CDP Gal			32			12

\* error in data is between 5 –15%

Inspection of the kinetic data clearly reveals that the desired nucleotide specificity switch from a UDP Gal to a CDP Gal utilizing enzyme was not achieved by mutating the side chain amide of Asn10 to a carboxylate. Both the wild-type and N10D versions of LgtC use UDP Gal in preference to CDP Gal. In fact, the N10D mutation results in a decrease in activity with respect to CDP Gal (~four-fold). A slight (~three-fold) decrease in the preference for UDP Gal over CDP Gal was achieved with the N10D mutation. However this results primarily from the effect of decreasing the natural activity of the enzyme to a greater degree (~13-fold decrease in activity with respect to UDP Gal) than the decrease in activity for the alternative activity (~four-fold decrease in activity with respect to CDP Gal).

#### **4.2.6. Concluding Remarks**

Although the results of this rational engineering approach for the facile control of glycosyltransferase nucleotide sugar substrate specificity are somewhat disappointing, significant insights and directions for future directions were obtained. Firstly, an essential assumption at the onset was that glycosyltransferases possess a stringent nucleotide sugar donor substrate specificity. The most significant finding of this study is the fact that LgtC is rather promiscuous with respect to the nature of the pyrimidine base moiety of the donor substrate it will tolerate (only a ~40-fold preference for UDP Gal versus CDP Gal). Whether this is a general theme amongst glycosyltransferases remains to be investigated, however the results provided here challenge the generally held view of strict donor substrate specificity. It is likely that this strategy for changing nucleotide specificity

would be better suited to glycosyltransferases that possess a monogamous relationship with respect to their donor substrates.

A possible explanation for the lack of success in altering the nucleotide specificity of LgtC might be the inherent importance of nucleotide binding to the structure and mechanism of the enzyme. A general theme amongst glycosyltransferases is the observation that, upon binding the nucleotide sugar donor substrate, significant conformational changes occur that lead to the formation of the active conformation of the enzyme. Perturbing this critical region of the enzyme could lead to the undesirable effect of destabilizing the active form and/or catalytic efficiency of the enzyme. Indeed, the N10D mutation leads to deleterious effects in terms of both  $K_M$  and  $k_{cat}$  with respect to the natural reaction using UDP Gal. In order for such a strategy to be successful, it is quite possible that secondary site “suppressor” mutations are necessary to rescue the effect of the specificity switch. Indeed this has been found to be the case with chemical genetics strategies involving several ATP utilizing protein kinases (Zhang et al., 2005). Future strategies to alter the nucleotide specificity of glycosyltransferases will benefit from such considerations.

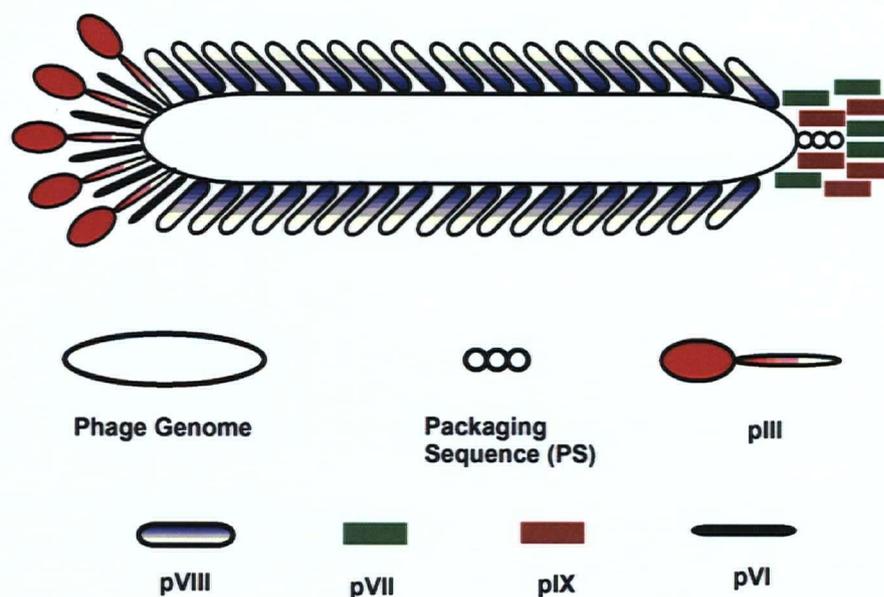
## 4.3. Compartmentalized Phage Display of LgtC

### 4.3.1. Background

The Ff class (Genus *Inovirus*) of filamentous phages (f1, fd and M13) consist of a circular single-stranded DNA (ssDNA) genome housed within a protein capsid cylinder. As their name implies, they specifically recognize the tip of the F conjugative pilus of *E. coli* containing the F' plasmid (F' strains). They are non-lytic phages and upon productive infection they do not kill their host. Using a mixture of phage- and bacterium-encoded components, single-stranded viral DNA is replicated to a double-stranded intermediate that is used as the template for the production of new phage proteins. In addition, newly synthesized single-stranded viral DNA is generated and maintained by the formation of a complex with multiple copies of single-stranded DNA binding proteins encoded by the phage genome. Viral capsid proteins exist as integral membrane proteins until they are assembled around a newly synthesized phage genome that is extruded out of the bacterial cell through a "membrane-associated assembly site" with concomitant removal of single-stranded DNA binding proteins, thereby giving rise to a newly formed organism. Infected bacteria tolerate the viral replication process with minimal effect, having replication times of just 50% longer than that of uninfected cells and with the generation of ~100 particles per generation. Indeed, the replication process can proceed for many generations leading to virus titers of up to  $10^{13}$  per mL of culture.

Having a mass of ~16.3 MD, an Ff phage particle is a protein-based cylinder (~6.5 x 930 nm) surrounding a closed circular ssDNA genome that consists of 6400 nucleotides. The entire length of the particle consists of ~2700 molecules of the 50-residue major pVIII coat protein (Figure 4.5). The 78-nucleotide hairpin packaging signal

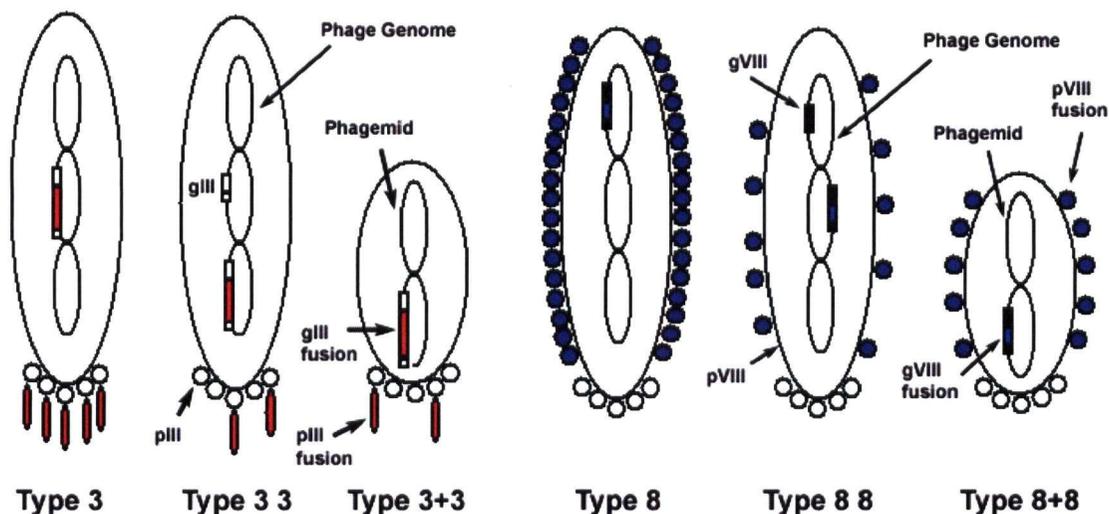
of the genome is always oriented to the capping end that consists of five molecules each of the 33-residue VII and 32-residue pIX coat proteins. The other end of the cylinder consists of five copies each of the 406-residue pIII and 112-residue pVI coat proteins. The pIII coat protein is required for the F conjugative pili recognition events required for infection. Other proteins encoded by the phage genome are required for DNA replication (pII and pX), phage assembly (pI, pIV and pXI) or binding to ssDNA (pV). All Ff phage genomes encode virtually identical amino acid sequences for the eleven gene products described above. Interestingly, two of the gene products (pX and pXI) result from translation starting at an internal methionine codon (Guycaffey et al., 1992; Rapoza and Webster, 1995). For a detailed description of the Ff phage infection process, assembly, life cycle, biology and genetics the reader is directed to the appropriate literature (Barbas III, 2001).



**Figure 4.5.** Graphical representation of the macromolecular anatomy of an Ff filamentous phage particle.

Peptides or entire proteins can be displayed on the surface of Ff phage by fusion to a coat protein. This was first described by George Smith in 1985 wherein fragments of the endonuclease *EcoRI* were fused to the N-terminal portion of pIII, resulting in the production of chimeric proteins that were packaged into phage particles (Smith, 1985). Multiple systems have since been tailored for the display of peptides and proteins on the Ff phage surface (Smith and Petrenko, 1997; Smith and Scott, 1993). Phage display systems are classified based on the resulting arrangement of coat proteins (Smith, 1993). This classification system is illustrated using pIII and pVIII fusions in Figure 4.6 and a brief description for fusions to pIII follows. In a type 3 system, the phage genome is modified to contain a single version of recombinant gIII existing as a fusion to an amino acid sequence of interest. In theory, the foreign amino acid sequence is displayed on all five copies of the pIII coat protein, thereby greatly reducing infectivity. In a type 3 3 system, the phage genome contains two copies of the gIII, encoding two different versions of the pIII protein. One version is the wild-type gene and the other is a recombinant encoding a fusion of the amino acid sequence of interest fused to gIII. Particles arising from this system contain a mixture of wild-type and fusion pIII proteins on their surface. Finally, in the 3+3 system, like in the 3 3 system, two gene versions of gIII are present. However, in this case, a wild-type version is encoded by a phage genome (described as the “helper” phage) and a recombinant version is encoded by a type of plasmid known as a “phagemid”. A phagemid carries an antibiotic resistance gene, both *E. coli* and filamentous phage origins of replication and the 78-nucleotide packaging sequence that facilitates its incorporation into newly forming progeny. In this system, an F' strain of *E. coli* is transformed with the phagemid, cultured under antibiotic selection,

and then the phage genome is introduced by superinfection with helper phage. In modern helper phage systems, the packaging sequence of the helper phage genome is disrupted to prevent packaging into newly formed progeny allowing for the production of a homogenous population of phage harboring phagemid DNA and displaying copies of both the wild-type and fusion versions of the pIII coat protein on their surface. The use of phagemid/helper phage systems is particularly well-suited to the production and screening of large libraries of protein and peptide variants. Usage of a 3+3 system is suited to the “low copy” display (~one fusion protein per particle) of whole proteins, while the 8+8 system is suited for the polyvalent display of shorter peptides.



**Figure 4.6.** Schematic representations of particles resulting from various phage display systems available for the surface display of foreign peptides or entire proteins.

Phage display has become an invaluable tool in screening procedures for the *in vitro* evolution of protein function, wherein active protein variants are selected from a

large pool of random mutants thereby mimicking the natural selection process. Although library sizes are still limited by transformation efficiencies ( $<10^9$ ), unlike traditional screening or selection strategies performed in bacterial or yeast cells, phage display allows complete control over reaction conditions (e.g. buffer, pH, metals, substrates, competing substrates, etc.). Phage display has proved to be particularly useful in the engineering of antibodies and other binding proteins with desired binding affinities. In these cases, screening is easily achieved by panning phage displaying protein variants over immobilized ligands and enriching those with desired binding properties. Limiting the application of phage display in the directed evolution of enzymes is the difficulty in obtaining a physical linkage between a phage that displays an active enzyme variant and the catalytic activity (i.e. product formation). The first applications of phage display to the evolution of catalytic activity made use of similar approaches to those used for the evolution of binding affinities. Variants of catalytic antibodies or protein enzymes displayed on phage were screened based on their ability to bind immobilized transition state analogues or suicide substrates (Fernandez-Gacio et al., 2003). Schultz and co-workers provided the first report of a phage display-based screening methodology for the directed evolution and functional cloning of enzyme activity based on turnover (Pedersen et al., 1998). Since then several other strategies have been reported, all of which involve creating a physical linkage between a substrate and a phage particle and rely on proximity effects to ensure that product formation is the result of the enzyme linked to the same phage particle (Atwell and Wells, 1999; Cesaro-Tadic et al., 2003; Demartis et al., 1999; Fa et al., 2004; Jestin et al., 1999; Strobel et al., 2003; Xia et al., 2002; Yin et al., 2004).

Clearly, a phage display-based strategy that does not rely on proximity effects and is based on turnover would be a useful development in the field.

One of the most significant recent developments in the field of directed evolution has been *in vitro* compartmentalization (IVC) procedures based on water-in-oil emulsions. Such strategies elegantly mimic those used by nature to link a genotype with a phenotype. In nature, lipid bilayer membranes of cells or organelles are used to compartmentalize a genotype with gene products, thereby facilitating a physical link between inheritable genotypes and selectable phenotypes. Similarly, the biomimetic IVC strategy involves the dispersion of a water phase in an oil phase and results in the creation of aqueous droplets (~5 fL each) that effectively restrict diffusion between compartments. Concentrations are chosen such that individual compartments contain on average less than a single gene of interest, thereby effectively recreating cell-like conditions in which transcription, translation and protein activity all take place within an isolated compartment. The system has the advantages of being high throughput ( $>10^{10}$  phenotypes per 1 mL), economical, easily performed, and being amenable to a range of reaction and temperature conditions.

Application of IVC for the directed evolution of enzyme activity was first reported for the selection of a DNA methyltransferase (Tawfik and Griffiths, 1998) and subsequent reports of application to the evolution of other DNA modifying enzymes rapidly ensued (Cohen et al., 2004; Doi et al., 2004; Ghadessy et al., 2001; Ghadessy et al., 2004; Lee et al., 2002b). Application of IVC to a non-DNA modifying enzyme has also been reported and involves two compartmentalization steps (Griffiths and Tawfik, 2003). All of these

procedures required *in vitro* transcription/translation of the protein variants, a procedure notoriously challenging to perform with a high degree of consistency and reliability.

More recently, a screening methodology has been developed that involves the use of double “water-in-oil-in-water” emulsions that can be directly sorted using fluorescence activated cell sorting (FACS) (Aharoni et al., 2005; Bernath et al., 2004). In this FACS-based procedure, product formation is detected by the production of a fluorescent signal. Individual cells expressing a unique protein variant are isolated within a compartment that prevents the diffusion of the fluorescent signal used to sort active from inactive variants.

If the fluorescent signal indicating product formation is retained within a cell, direct FACS can be performed and emulsification is not necessary. Very recently, just such an approach (non phage-based) was described for the directed evolution of the sialyltransferase Cst II using a lactose acceptor conjugated to a Bodipy fluorophore (Aharoni et al., 2006). Because the resulting 3'-sialyl lactose product is not exported by lactose permease, the enzyme product, and therefore the fluorescent signal, is retained within the *E. coli* cell following extensive washing steps thereby facilitating the direct sorting of cells harboring active protein variants.

The FACS-based screening approaches described above require product formation to be coupled to a fluorescent signal either resulting from the release of a fluorescent product or by the use of a substrate that has been modified with a fluorescent substituent. For many enzyme reactions, particularly those catalysed by transferase enzymes, fluorogenic substrates are not available. Considering the adage that “you get what you select for”, a strategy that involves the use of substrates modified with a fluorophore has the potential disadvantage of screening for an activity dependent on the alternative

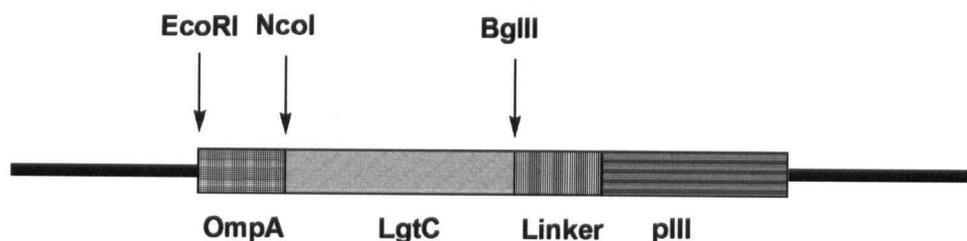
fluorophore-containing substrate. In fact, the directed evolution of Cst II resulted in the isolation of a protein variant with an evolved fluorophore binding site and increased activity was only detected using substrates that contained an expensive Bodipy substituent (Aharoni et al., 2006). In addition, *in vivo* screening procedures have the further limitation of lacking control over reaction conditions. Finally, FACS-based screening is a technologically intensive endeavor, limiting its application to only the more affluent of laboratories.

Application of IVC to phage display would provide the field of directed enzyme evolution with a generally applicable, inexpensive and facile screening technology. Unlike previously described IVC-based approaches, the robust procedure of phage display would eliminate the need for difficult *in vitro* transcription/translation and would have the advantage over *in vivo* procedures of providing complete control of reaction conditions. In addition, the use of IVC would allow the detection of product formation without reliance on proximity effects. Using the glycosyltransferase LgtC as model system, it was determined whether phage display methodologies are possible in the context of IVC and if the coupling of these two procedures could be used to devise a screening strategy for application in directed enzyme evolution.

#### **4.3.2. Phage Display of LgtC as an M13 pIII Fusion Protein**

A modified version of the commercial pBluescript SK (-)<sup>TM</sup> phagemid designated pSJF6 was obtained from the laboratory of Dr. Warren Wakarchuk at the National Research Council of Canada Institute for Biological Sciences in Ottawa. The multiple cloning site (MCS) of pSJF6 contains a DNA sequence encoding an N-terminal OmpA

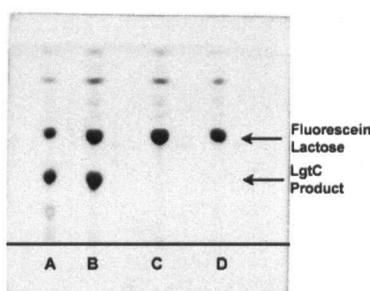
leader sequence fused to the pIII coat protein of M13 bacteriophage via a Gly/Ser rich linker that also contains a *c-myc* detection sequence. The lack of an appropriate restriction site between the OmpA sequence and the linker necessitated a subcloning strategy in which LgtC was PCR amplified using a forward primer encoding the entire OmpA leader sequence and then inserted into pSJF6 via EcoRI and BglII restriction sites. However, presumably due to the large size of the forward primer (~100 bp), the quantity of PCR product obtained was insufficient for subcloning. In order to increase the amount of DNA encoding the OmpA-LgtC fusion, a nested PCR was performed using the small amount of initial PCR product as template and the resulting nested PCR product was successfully cloned into pSJF6. Sequencing the resulting product revealed a seven base pair deletion in the OmpA leader sequence that presumably arose during the initial PCR amplification. The deleted seven base pairs were readily reintroduced using the restriction enzyme *AarI*. DNA sequencing was used to confirm the presence of the desired DNA consisting of a contiguous N'-OmpA-LgtC-linker-pIII-C' encoding sequence within the MCS, and the resulting phagemid was designated pLL01 (Figure 4.7). The introduction of an *NcoI* cut site at the start of the LgtC sequence makes pLL01 a useful vector for the creation of other pIII fusion proteins. In fact, pLL01 was successfully used by Dr. David Fox to generate a phagemid for type 3+3 display of the xylanase Bcx by subcloning Bcx into pLL01 via *NcoI* and *BglII* restriction sites (unpublished results).



**Figure 4.7.** Recombinant construct within the MCS of pLL01 encoding an LgtC-pIII fusion protein.

To determine whether LgtC remained catalytically active when expressed as a C-terminal pIII fusion protein, activity tests were performed on cell lysates derived from IPTG-induced cultures of *E. coli* harboring pLL01. As determined by TLC analysis, LgtC remained active upon fusion to the pIII M13 phage coat protein. A small scale PEG-precipitated preparation of putative LgtC-displaying phage was performed on phage derived from the XL1 Blue (f') strain of *E. coli* harboring pLL01 superinfected with VCSM13 helper phage. The resulting phage progeny were quantified using UV absorbance, and LgtC activity was tested by TLC analysis using fluorescein lactose conjugate as acceptor (Figure 4.8). Encouragingly, this analysis revealed that M13 phage derived from *E. coli* harboring pLL01 did possess LgtC activity comparable to that of wild-type (non fused) LgtC indicating the successful display of active enzyme as a pIII fusion. It must be noted that the LgtC activity observed from twice PEG precipitated phage may have resulted from fusion proteins transported to the periplasm but not incorporated into phage that had co-precipitated with phage particles. Although this interpretation cannot be completely ruled out, several facts would argue against it. It would be a significant coincidence that the level of LgtC activity resulting from non-incorporated fusion protein would directly correlate with the concentration of displayed LgtC determined based on the concentration of phage particles. In addition, bacterial

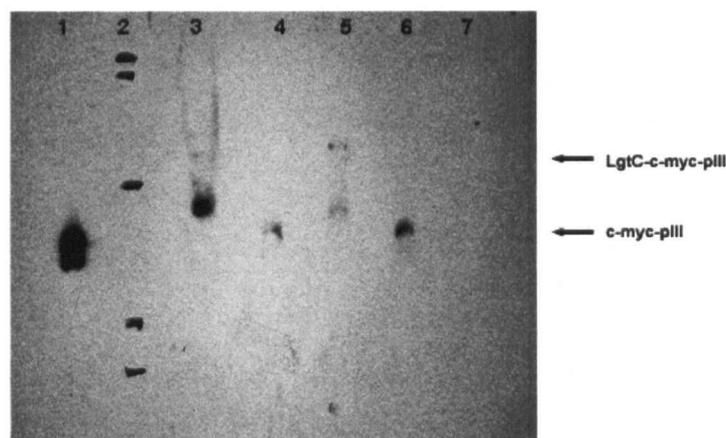
cells are removed by low speed centrifugation prior to PEG precipitation of phage from the resulting supernatant. In the absence of lysozyme, it is difficult to rationalize how an appreciable quantity of fusion protein not incorporated into phage particles would be excreted past the outer membrane. At the time, this was the first glycosyltransferase to be successfully displayed on the surface of phage. Walker and co-workers subsequently reported the successful display of the glycosyltransferase MurG (Love et al., 2006).



**Figure 4.8.** TLC analysis (7:2:1 EtOAc:MeOH:H<sub>2</sub>O) of the LgtC activity of twice PEG precipitated M13 bacteriophage derived from pLL01 phagemid. Reactions were run using 100 mM HEPES (pH 7.5), 5 mM MnCl<sub>2</sub>, 5 mM DTT, 5 mM UDP Gal, 0.5 mM fluorescein lactose and either ~30 nM purified LgtC (A), ~10 nM M13 phage derived from pLL01 (B), ~10 nM M13 phage derived from pSJF6 (C), or without enzyme source (D).

Western blot analysis of pLL01-derived phage was performed to confirm the display of LgtC-pIII fusion coat protein using rabbit anti *c-myc* Ab-horse radish peroxidase (HRP) conjugate for detection (Figure 4.9). This analysis reveals the presence of a pIII fusion protein (derived from pLL01) that differs from that of the non-fused pIII protein (derived from pSJF6). It must be noted that the theoretical mass of the LgtC-pIII fusion protein is 55.2 kDa and the major band derived from the putative fusion protein, although clearly larger than non-fused pIII, appears less than 50 kDa. This is highly

suggestive that the LgtC-pIII fusion has been subjected to proteolytic degradation, a frequent concern in phage display. Fortunately, intact LgtC-pIII with an apparent molecular weight of ~55 kDa could also be detected as a major band in other experiments (as discussed in section 4.3.3). It must also be noted that non-fused pIII protein, derived from pSJF6, migrates on an SDS PAGE gel with an apparent mass that differs from the actual mass of 23.6 kDa due to the presence of a large number of positively charged residues within the protein.



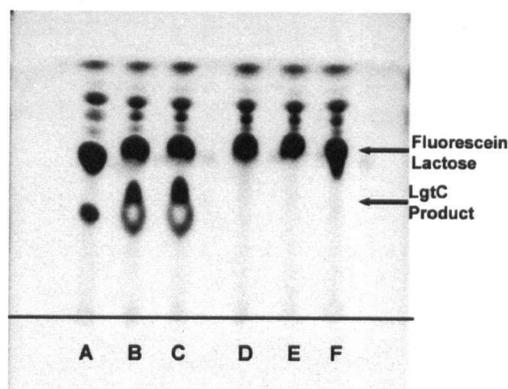
**Figure 4.9.** Western blot analysis of various phage particles using anti-c-myc-HRP for detection. Lysate from induced cultures of *E. coli* harboring pSJF6 was used as a positive control (lane 1). VCSM13 helper phage was used as a negative control (lane 7). Molecular weight standards (lane 2) were written onto the Western blot by overlaying onto the SDS PAGE gel. The molecular weights of these standards are 108, 90, 50.7, 35.5, and 28.6 kDa. PEG precipitated phage were derived directly from *E. coli* harboring either pLL01 (lane 3) or pSJF6 (lane 4) or in the context of water-in-oil emulsions from *E. coli* harboring either pLL01 (lane 5) or pSJF6 (lane 6).

### 4.3.3. Compartmentalized LgtC Phage Display

Having demonstrated that catalytically active LgtC can be displayed on M13 bacteriophage as a pIII fusion, it was determined whether this phage display procedure could be successfully conducted in the context of water-in-oil emulsions. XL1 Blue (f') *E. coli* harboring pLL01 were emulsified from an aqueous solution containing all of the components necessary for LgtC activity and VCSM13 helper phage. As a positive control, LgtC-displaying phage were generated and left to incubate under identical reaction conditions but without being subjected to the emulsification procedure. To obtain conditions in which on average less than one cell, and as such no more than one LgtC gene, was contained within each compartment,  $\sim 10^8$  cells were emulsified in  $\sim 10^9$  compartments. Washed cells derived from overnight culture were resuspended in 50 mM HEPES (pH 7.5) containing 5 mM  $\text{MnCl}_2$  and 10  $\mu\text{g/mL}$  ampicillin and immediately prior to the addition of ten volumes of an ice cold mixture of AbilEM90<sup>TM</sup> in light mineral oil (2.9% w/w), a ten-fold concentrate of LgtC reaction components (50 mM DTT, 50 mM UDP Gal, 5 mM fluorescein lactose conjugate) and  $2 \times 10^{10}$  VCSM13 helper phage were added to the resuspended cells. Emulsification on ice for five minutes at 9500 rpm using an IKA T-25<sup>TM</sup> homogenizer resulted in the production of a milky white suspension that was incubated at 37°C overnight. The following day, the mixture was centrifuged (14, 000 rpm) and the oil phase was removed. The aqueous layer was diluted five-fold with a volume of 50 mM HEPES (pH 7.5) and washed twice with water-saturated diethyl ether. Remaining ether was immediately removed from the aqueous phase using a Speed Vac. Detection of LgtC activity that occurred within compartments was confirmed by TLC analysis of the aqueous phase immediately following

centrifugation (Figure 4.10). In addition, LgtC activity of phage isolated by PEG precipitation from aqueous phases was tested by TLC analysis (Figure 4.10).

Isolated phage generated in emulsions were also subjected to Western blot analysis as described above to confirm the presence of LgtC-pIII fusion proteins (Figure 4.9). As was observed with non-emulsified phage, a significant band appears corresponding to an LgtC-pIII fusion protein that has been subjected to proteolytic degradation. However, a major band is present that corresponds to the expected mass (55 kDa) of an LgtC-pIII fusion protein.



**Figure 4.10.** TLC analysis (7:2:1 EtOAc:MeOH:H<sub>2</sub>O) of the LgtC activity of emulsified M13 bacteriophage derived from pLL01 phagemid. Reactions were run using 100 mM HEPES (pH 7.5), 5 mM MnCl<sub>2</sub>, 5 mM DTT, 5 mM UDP Gal, 0.5 mM Fluorescein Lactose and the following enzyme sources. (A) ~30 nM of purified LgtC. (B) Aqueous phase of broken emulsions containing phage derived from pLL01. (C) Twice PEG precipitated emulsified phage derived from pLL01 (~10 nM). (D) Aqueous phase of broken emulsions containing phage derived from pSJF6. (E) Twice PEG precipitated emulsified phage derived from pSJF6 (~10 nM). (F) No enzyme source.

A near identical number of phages were recovered from the aqueous phase of the emulsions compared to an equivalent volume of culture not subjected to emulsification ( $\sim 5 \times 10^{12}$  mL<sup>-1</sup>). The infectivity of emulsified LgtC-displaying phage was lower ( $\sim 4 \times 10^6$  Tu/mL from an input of  $\sim 10^8$  mL<sup>-1</sup> as) than that of non-emulsified LgtC displaying phage ( $\sim 4 \times 10^7$  Tu/mL from an input of  $\sim 10^8$  mL<sup>-1</sup>). For this reason, if compartmentalized phage display were incorporated into a directed evolution screen, to maximize the size of a library of variants, the enriched gene sequences should be recovered by lysis of enriched phages followed by PCR amplification of the gIII fusion sequences. The results described in this section indicate that phages displaying active LgtC can be generated from individual *E. coli* cells harboring an LgtC encoding phagemid when compartmentalized using water-in-oil emulsions. Catalytic activity was detected within the compartments and the phagemids of isolated phages could be readily obtained by infection of *E. coli* cells.

#### **4.3.4. A High Throughput Screen for LgtC Directed Evolution Based on Compartmentalized Phage Display**

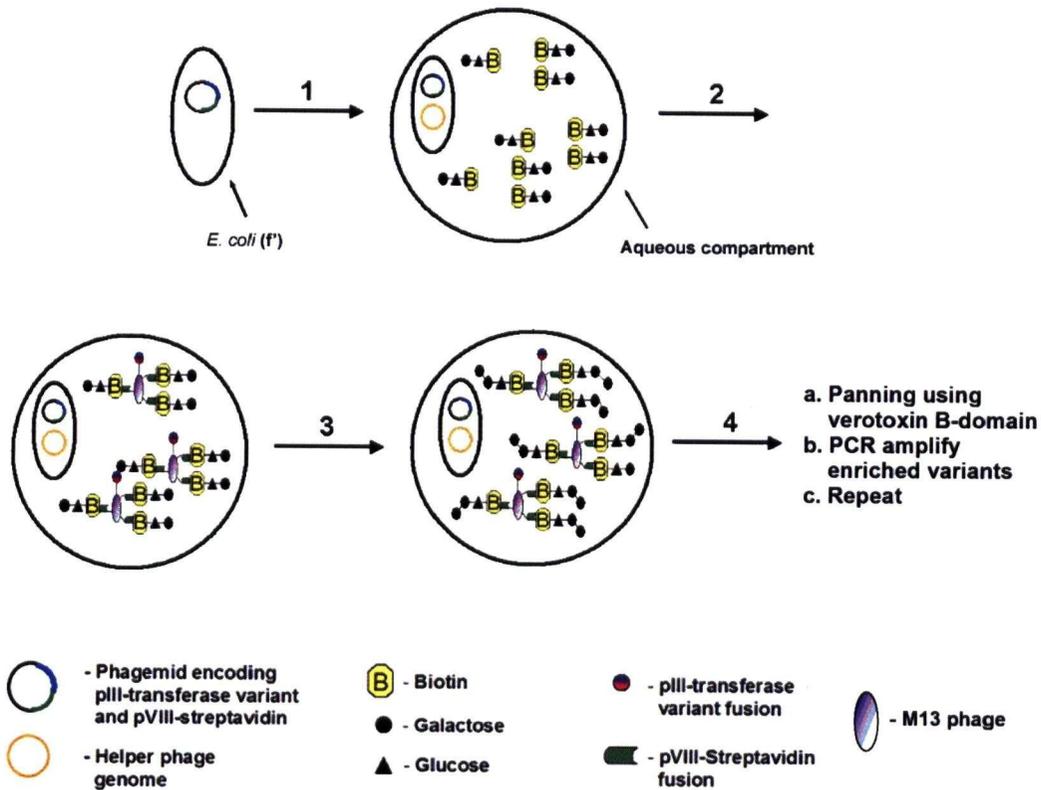
A screening methodology involving compartmentalized phage display was devised for the directed evolution of LgtC activity. The screen was developed such that it would be high throughput, inexpensive, easy to perform, reliable and readily adaptable to other enzyme systems. Methods for creating a physical linkage between phage harboring a phagemid encoding an active protein variant and the product of the reaction being catalysed, and for enriching phage linked to the product from those linked to the substrate were required.

Sidhu and colleagues have reported an evolved version of the major coat protein pVIII of M13 bacteriophage that greatly enhances the polyvalent surface display of the oligomeric protein streptavidin (SAV) (Sidhu et al., 2000). Inclusion of this construct in the system used for successful display of LgtC as a pIII fusion would result in phages that simultaneously displays LgtC (or a variant thereof) in low copy number as a pIII fusion and multiple copies of SAV as a pVIII fusion on its surface. SAV is a tetrameric protein from *Streptomyces avidinii* that is widely used in molecular biology and biochemistry applications because of its ability to bind the vitamin biotin with one of the strongest known non-covalent interactions ( $K_d \sim 10^{-15}$  M)(Chalet, 1964; Green, 1975). As such, a substrate conjugated to biotin would be essentially irreversibly linked to the surface of a phage displaying SAV. Compartmentalization of an *E. coli* cell infected with M13 helper phage and harbouring a phagemid encoding a variant of LgtC fused to pIII and SAV fused to pVIII under reaction conditions that include the presence of a biotin-lactose conjugate would lead to the production of multiple isolated phages simultaneously displaying an identical version of LgtC and having multiple copies of lactose acceptor linked to its surface. If the version of LgtC were catalytically active, the surface of each phage would be modified to have the trisaccharide product of the LgtC reaction on its surface. Because all of the phage within a particular compartment would harbour the same phagemid encoding a particular LgtC variant, this system would not rely on proximity effects to link product formation and enzyme variant on the same phage. In addition, this system would be readily adaptable to other enzyme systems by simply generating a conjugate consisting of a substrate and biotin. Following destruction of the compartments and isolation of the aqueous phase, phage harbouring phagemids encoding

active LgtC variants could be enriched from inactive variants by panning over a macromolecule that specifically recognizes and binds the trisaccharide product of the LgtC reaction.

Derived from pathogenic bacteria, verotoxins, also known as Shiga-like toxins, are able to arrest eukaryotic protein synthesis and are known to damage endothelial cells of the kidney and brain leading to renal failure and deleterious neurological effects (Coia, 1998; Kitov et al., 2000). These toxins consist of two domains. The A domain confers activity by modifying the 28S rRNA of the 60S ribosome such that it can no longer interact with EF-1 and EF-2 elongation factors, thereby preventing effective protein synthesis (Endo et al., 1988). The B domain is a homopentamer that specifically recognizes eukaryotic cell receptor globotriaosylceramide (GB<sub>3</sub>) and facilitates receptor-mediated endocytosis of the toxic A domain (Sandvig and Vandeurs, 1994). Each monomer possesses three saccharide binding sites that recognize the terminal  $\alpha$ -D-gal(1→4) $\beta$ -D-gal(1→4) $\beta$ -D-glc “P<sup>k</sup>” trisaccharide of GB<sub>3</sub>. Interaction of the B-domain pentamer with the methyl glycoside of the P<sub>k</sub> trisaccharide is known to be of rather low affinity ( $K_d \sim 1$  mM)(St. Hilaire et al., 1994). However, because of the multivalent interaction resulting from a total of 15 protein-carbohydrate interactions on the intact pentamer, the estimated affinity of the pentamer for cells expressing GB<sub>3</sub> on their surface is very high ( $K_d \sim 1$  nM)(Fuchs et al., 1986). Because the product of the LgtC reaction is the P<sup>k</sup> trisaccharide, the B domain of a verotoxin would be an ideal candidate as a molecular recognition device for isolating phage containing the LgtC product on their surface.

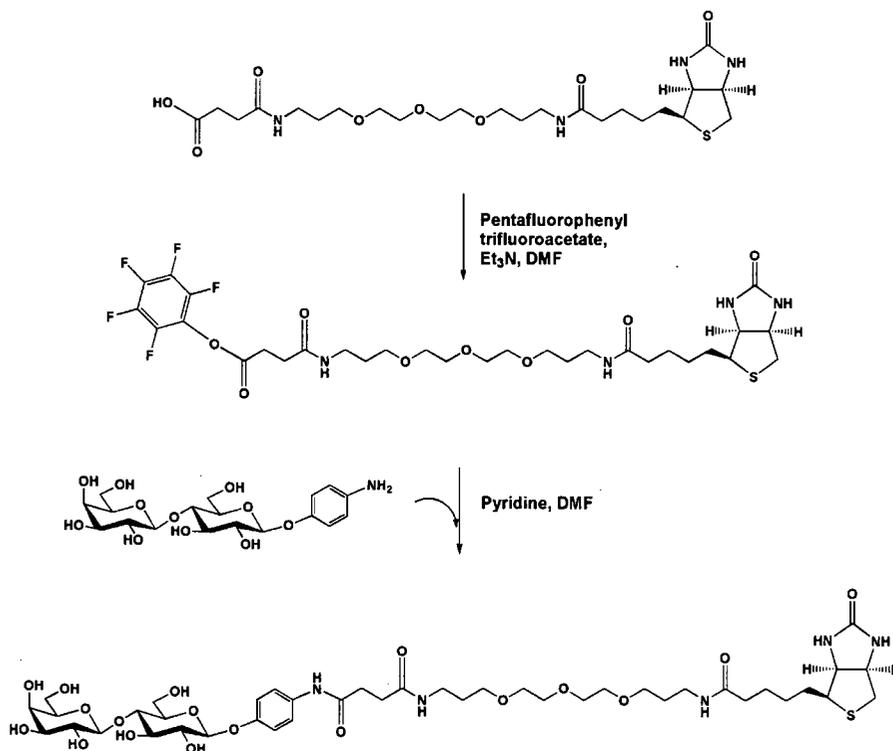
The proposed screening strategy for the directed evolution of LgtC was to simultaneously display multiple copies of SAV and a variant of LgtC in low copy number on the surface of M13 bacteriophage in water-in-oil emulsions in the presence of a desired reaction condition that includes the presence of a biotin-lactose conjugate acceptor substrate. LgtC activity would convert lactose to the P<sup>k</sup> trisaccharide, which would be effectively linked to the phage surface via the high affinity biotin-SAV interaction. Phage harbouring DNA encoding active LgtC variants could be enriched using a panning procedure involving the B-domain of a verotoxin and the LgtC encoding regions of isolated phagemids could be PCR amplified and subjected to subsequent rounds of screening. An overview of this scheme is shown in Figure 4.11.



**Figure 4.11.** Proposed screening procedure for the *in vitro* evolution of LgtC involving compartmentalized phage display. (1) Individual cells of an f<sup>+</sup> strain of *E. coli* harbouring a phagemid containing a DNA sequence for polycistronic expression of a unique variant of LgtC fused to pIII and SAV-pVIII fusion proteins are subjected to water-in-oil emulsification in the presence of the desired reaction conditions including a biotin-lactose conjugate acceptor substrate and VCSM13 helper phage. (2) Following infection with helper phage, phage harbouring phagemid DNA encoding a unique LgtC variant and having both LgtC displayed in low copy number as a pIII fusion and SAV displayed in high copy number as a pVIII fusion on their surface are excreted from the bacterial cell. (3) Depending on the level of transferase activity, the lactose acceptor attached to phage surfaces via the high affinity biotin-SAV interaction is converted to a trisaccharide product. (4) Phage with catalytically active enzyme on their surface are enriched from inactive variants by breaking emulsions, isolating phage and panning using a carbohydrate binding protein specific for the trisaccharide product of the LgtC reaction. Enriched gene variants are then isolated and subjected to subsequent rounds of screening.

### 4.3.5. Synthesis of a Biotinylated Lactose Conjugate Acceptor Substrate

A biotin lactose conjugate was readily prepared using an advanced precursor generated and provided by Dr. Spencer Williams that has been successfully applied in ICAT-based proteomic investigations of glycosidases (Hekmat et al., 2005; Williams et al., 2006). It should be noted that nearly identical precursors consisting of biotin and a PEG linker terminating in a free acid are commercially available, making the general applicability of this screening methodology feasible. The acid terminating the linker of the biotinylated precursor was activated as the pentafluorophenyl ester prior to reaction with *p*-amino-phenyl- $\beta$ -D-lactoside (Scheme 4.2). The resulting biotin lactose conjugate was subjected to preparative HPLC purification affording ~10 mg of pure acceptor substrate.



**Scheme 4.2.** Synthesis of a biotin lactose conjugate acceptor substrate for LgtC.

### 4.3.6. Generating a Library of LgtC Variants

Error-prone PCR (epPCR) is the method most often used to generate libraries of protein variants containing random mutations for directed evolution applications. PCR protocols are modified to alter and enhance the natural error rate of a DNA polymerase (Leung, 1989). Despite its bias towards AT to GC changes, because of its high natural error rate, *Taq* polymerase is commonly used. Typical conditions used to increase and alter the errors that occur during the PCR process include the use of increased concentrations of  $MgCl_2$  to stabilize non-complementary base pairing (Cadwell and Joyce, 1994); addition of  $MnCl_2$  to increase the error rate (Lin-Goerke, 1997); varying the ratio of nucleotides (Vartanian et al., 2001); the addition of nucleotide analogues (Fenton, 2002); using alternative DNA polymerase enzymes such as Stratagene's Mutazyme™ polymerase; increasing the number of doubling cycles; or changing the initial concentration of template. The length and base composition of the template DNA dictate the resulting rate of mutational frequency. As such, under identical epPCR conditions, two different genes will most likely exhibit different rates and forms of mutation.

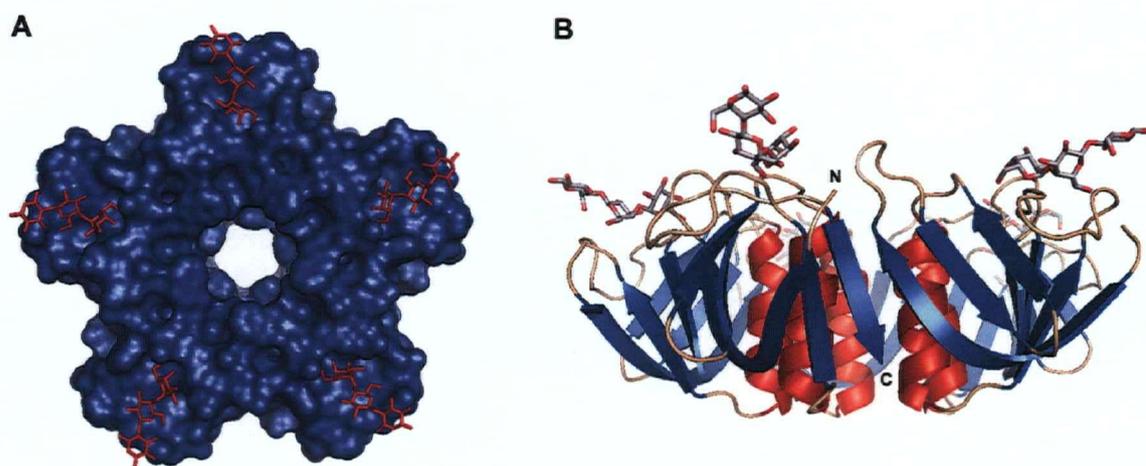
A library of LgtC variants was generated using epPCR conditions that included a 5:1 GT to AC nucleotide bias and varying concentrations of  $MnCl_2$ . Presumably due to the relatively large size of the vector, initial attempts using LgtC in pET29 (+) as template prevented the generation of libraries with a diversity of greater than  $10^5$ . Using LgtC in pUC18 a library size of  $10^8$  was achieved. Sequencing of ten random variants revealed that use of 0.25 mM  $MnCl_2$  led to the generation of a balanced mutational rate of ~two mutations per gene and use of 0.5 mM  $MnCl_2$  led to the generation of a balanced mutational rate of more than four mutations per gene. The library generated using 0.25

mM MnCl<sub>2</sub> and pUC18 containing wild-type LgtC was transformed into an expression strain of *E. coli*. Crude TLC scale activity tests on induced cell lysates revealed an activity level for the library that was ~20% that of the wild-type enzyme. Based on the mutational rate and activity, this library was deemed suitable for the directed evolution of LgtC activity.

#### **4.3.7. Immobilization of the Carbohydrate Binding Domain of Shiga-Like Toxin SLT-1B**

The three-dimensional X-ray crystal structure of the carbohydrate binding domain of Shiga-like toxin SLT-1B derived from *E. coli* 0157 with bound Pk trisaccharide was used to determine a suitable strategy for immobilizing a functional form of the protein. The C- and N-termini of the protein are located on opposite sides of the overall structure, with the ligand binding sites located on the N-terminal face (Kitov et al., 2000)(Figure 4.12). It was therefore decided to immobilize the protein on an IMAC surface as a C-terminal 6-histidine version of the macromolecule. The DNA encoding the carbohydrate binding domain of this verotoxin protein was subcloned from pVT27 vector (supplied by Dr. Warren Wakarchuk, IBS NRC, Ottawa) into pET29(+) for expression with 6-histidines fused to the C-terminus using *Nde*1 and *Xho*1 restriction sites. Sequencing of plasmid DNA isolated from a resulting clone confirmed the generation of the desired sequence. Although over-expression of this construct in *E. coli* led to the production of very high amounts of insoluble protein, ~100 mg/L of soluble protein could be readily purified using a Ni<sup>+2</sup> charged IMAC column. Indicating a rather high affinity for the IMAC column, 250 mM imidazole was required to elute bound verotoxin.

In order to test the ability of this immobilized protein to bind the P<sup>k</sup> trisaccharide, fluorescein lactose conjugate acceptor was converted to the P<sup>k</sup> product using LgtC. Ni<sup>+2</sup> charged IMAC beads were saturated with the purified protein and then incubated with either fluorescein lactose or P<sup>k</sup> conjugates. As a negative control, Ni<sup>+2</sup> charged IMAC beads without bound protein were incubated with fluorescein P<sup>k</sup>. Following washing, only IMAC beads with bound protein and incubated with the fluorescein P<sup>k</sup> conjugate remained fluorescent. This indicates the ability of the immobilized construct to selectively bind the product of the LgtC reaction and its suitability for application in the enrichment step of the described screening strategy for the directed evolution of LgtC.



**Figure 4. 12.** Three-dimensional X-ray crystal structure of Shiga-like toxin I B subunit with bound starfish molecule (pdb 1qnu). The P<sup>k</sup> trisaccharide portions of the bound ligand are shown in top down (A) or side on (B) views. The locations of the N- and C-termini of one of the subunits of the homopentamer are shown in panel B to illustrate why a C-terminal 6-histidine tag was chosen as the construct for functional immobilization on an IMAC surface.

### 4.3.8. Multivalent Phage Display of SAV as a pVIII fusion protein

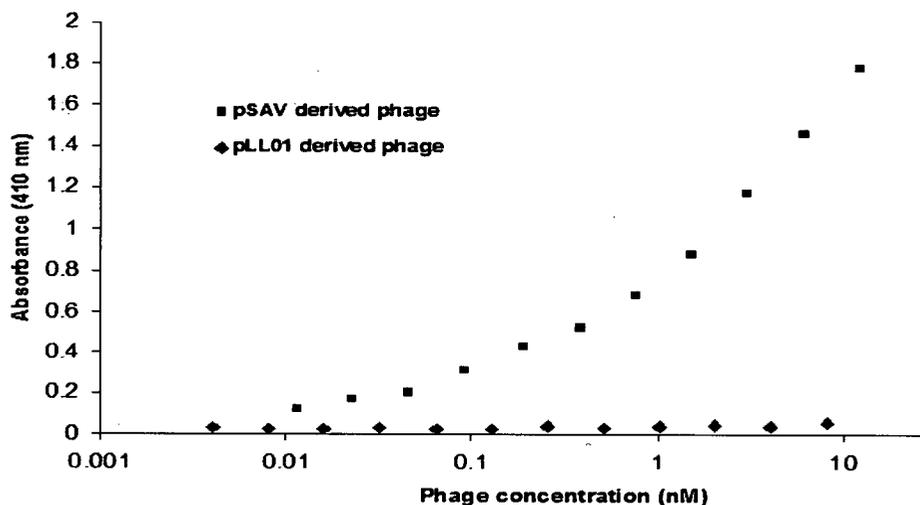
In practice, polyvalent display on pVIII is difficult. The levels of display are highly dependent on fusion length and sequence (Malik et al., 1996). Indeed, using the type 8 system (i.e. without the incorporation of wild-type pVIII coat protein) fusion of peptides greater than ten amino acids leads to the production of unstable phage (Iannolo et al., 1995). Using phagemid systems in which wild-type pVIII proteins are supplied *in trans* facilitates the display of larger peptides. However, even with the use of 8+8 systems, fusion of proteins greater than ~100 residues usually results in the display of no more than a single copy per phage particle (Kretzschmar and Geiser, 1995).

To overcome the limitations of 8+8 display, Sidhu et al. used directed evolution to isolate pVIII variants that greatly increased the surface display of monomeric and oligomeric proteins (Sidhu et al., 2000). Human growth hormone (hGH), a monomeric protein of MW = 22 kDa, was fused to the N terminus of pVIII and a library of variants in the pVIII sequence was generated. Variants that facilitated increased surface display of hGH were selected by panning over hGH receptor. After five rounds of selection, variants were isolated that increased the level of hGH display by ~100-fold.

To test the limits of type 8+8 display, the ability to display SAV (MW ~50 kDa) as a pVIII fusion was explored. Because SAV exists as a tetramer and is poorly secreted by *E. coli* (Weber et al., 1989), fusion of SAV to wild-type pVIII resulted in an undetectable level of display of functional SAV. To facilitate display of the functional tetrameric form of SAV, an amber stop codon was inserted between the sequences encoding SAV and pVIII. Phage production in an amber suppressor strain of *E. coli* leads

to translation of the majority of SAV protein as a free monomer, but also allows a small amount of read-through leading to the production of SAV-pVIII fusions. This results in the production of phage particles in which tetrameric SAV is likely displayed in a form that consists of one SAV-pVIII fusion with three associated SAV monomers. A library of pVIII variants fused to SAV was subjected to selection for increased display of SAV using anti-SAV polyclonal antibody as the capture target. Following five rounds of selection, variants of pVIII were isolated that provided an ~50-fold increase in the level of functional SAV displayed. Using this *in vitro* evolution approach, SAV display as a pVIII fusion was 'increased from an impractical level to levels that should allow for functional selection' (Sidhu et al., 2000). It was estimated that each phage particle would display 10 copies of tetrameric SAV thereby providing 40 ligand-binding sites (Sidhu, SS personal communication).

The phagemid pSAV, containing the construct encoding SAV fused to one of the selected pVIII variants, was obtained from Dr. Sachdev Sidhu (Genentech, San Francisco). An amber suppressor strain of *E. coli* (XL1 Blue (F')) was transformed with pSAV and phage particles displaying SAV as a pVIII fusion were generated by superinfection with VCSM13 helper phage. The resulting phage particles were isolated by double PEG precipitation and subjected to ELISA analysis using BSA-biotin conjugate as the capture target and anti-M13 phage monoclonal ab/horseradish peroxidase (HRP) conjugate for detection. Consistent with the published results obtained using this phagemid system (Sidhu et al., 2000), maximal response occurred at a phage concentration of ~10 nM, indicating a virtually identical level of SAV display to that obtained in the Genentech laboratories (Figure 4.13).

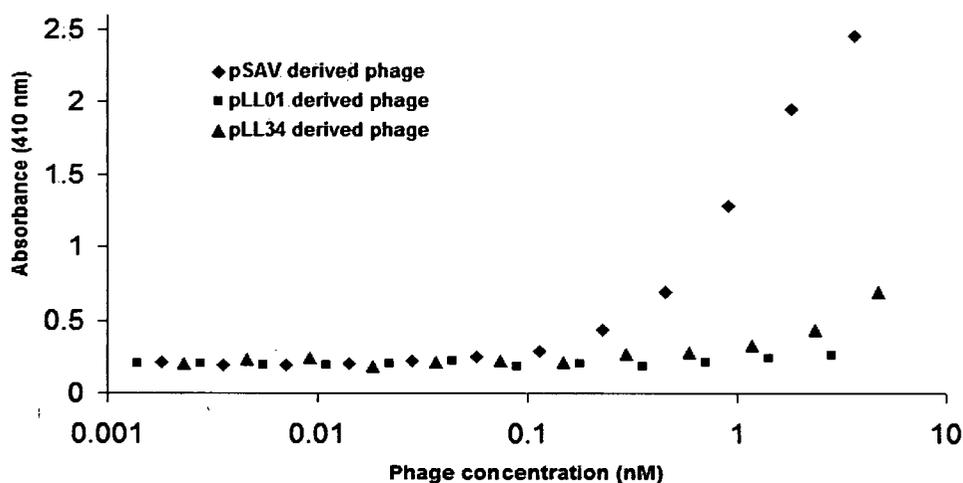


**Figure 4.13.** Phage ELISA for SAV display. Phage derived from uninduced cultures of XL1 Blue ( $f^+$ ) *E. coli* harbouring either pSAV (squares) or pLL01 (diamonds) were twice PEG precipitated and levels of SAV display investigated by determining their ability to bind immobilized biotin. Biotinylated BSA was immobilized on a 96-well plate as the capture target and following extensive washing, bound phage was detected using a monoclonal anti-M13 phage ab/HRP conjugate.

#### 4.3.9. Simultaneous M13 Phage Display of LgtC and SAV

The sequence encoding the optimized SAV-pVIII fusion protein was subcloned from pSAV into pLL01 using the *Hind*III restriction site. The resulting phagemid (pLL34) was generated for the polycistronic expression of LgtC-pIII and SAV-pVIII fusion proteins. DNA sequencing was used to confirm the presence of the desired construct consisting of the N-terminal-OmpA-LgtC-pIII fusion, a 42 bp spacer, ribosome-binding site, N-terminal-secretion sequence-SAV-amber stop-pVIII fusion within the MCS of pLL34.

It was then determined whether pLL34 could be used to generate phage particles that simultaneously displayed ~one copy of LgtC as a pIII fusion and ~ten copies of SAV as a pVIII fusion. XL1 Blue (f') *E. coli* were transformed with pLL34, pLL01 or pSAV and then superinfected with VCSM13 helper phage. The resulting phage were PEG precipitated and subjected to both LgtC activity tests and ELISA analysis to determine the level of SAV display. Phage derived from *E. coli* harboring pLL34 had an ~100-fold decrease in LgtC activity compared to those derived from pLL01, as determined by TLC analysis. In addition, display of functional SAV was significantly decreased and could barely be detected above background for phage derived from pLL34 (Figure 4.14).



**Figure 4.14.** Phage ELISA for SAV display. Phage derived from uninduced cultures of XL1 Blue (f') *E. coli* harbouring either pLL34 (triangles), pSAV (diamonds) or pLL01 (squares) were twice PEG precipitated and levels of SAV display investigated by determining their ability to bind immobilized biotin. Biotinylated BSA was immobilized on a 96-well plate as the capture target and, following extensive washing, bound phage was detected using a monoclonal anti-M13 phage ab/HRP conjugate.

In an effort to restore LgtC and SAV display on phage derived using pLL34, induction with a range of IPTG concentrations was investigated. Although induction with 0.1 mM IPTG led to a recovery of LgtC activity to within 10% of that of phages derived from pLL01, increased functional SAV display was not obtained. This is consistent with the reported observation that IPTG induction does not increase the level of SAV display and in fact resulted in a decrease in phage production (Sidhu et al., 2000). This is most likely the result of the fact that the SAV protein itself depletes bacterial cells of the essential vitamin biotin and is toxic.

It could very well be that simultaneous phage display of these two different proteins as pIII and pVIII fusion is simply too taxing for the production of viable phage. In addition, the choice of SAV as the pVIII fusion partner pushes the limits. Indeed, using wild-type pVIII as the fusion partner, SAV display could not be detected above background. Even with the optimized version of pVIII, the level of display was significantly less than the level of display achieved using an optimized version of pVIII fused to hGH. Whereas a significant ELISA response was observed at a phage concentration of  $\sim 10$  pM for hGH, in the case of SAV an equivalent response required a phage concentration of  $\sim 10$  nM (Sidhu et al., 2000).

#### **4.3.10. Conclusions and Future Directions**

The most significant finding described in this section is the observation that the process of phage display can be readily performed in the context of water-in-oil emulsions. This has significant implications for the application of phage display technology to the directed evolution of enzyme activity. Screening procedures for the

directed evolution of enzyme function involving phage display described to date have relied on proximity effects to ensure that the product of enzyme activity is derived from the protein variant attached to the same surface. This approach requires an optimization of the linker distance and, depending on the enzyme class, is limited in the number of substrate molecules accessible to the enzyme. By compartmentalizing a population of identical phage, carrying the same phenotype and displaying the same enzyme variant, multiple turnover of substrates on neighboring phage surfaces can be linked to the encoding genotype of an active variant.

In addition, it was shown that LgtC could be readily displayed on the surface of M13 phage as a pIII fusion protein. Glycosyltransferases are a class of enzyme that is particularly well suited for directed evolution involving phage display screening approaches. This applicability arises from the fact that the majority of glycosyltransferases are of membrane-associated species that contain a membrane association domain that is frequently removed for the purpose of obtaining soluble constructs. As such, the creation of a fusion in which this membrane association domain is substituted with an M13 coat protein should result in a catalytically active species. In addition, because glycosyltransferases are frequently transported to different organelles (e.g. Golgi bodies) or membrane surfaces, they should be readily secreted through the inner membrane of *E. coli*, a key requirement for successful phage display. Future work involving the phage display of LgtC should also address the issue of proteolytic cleavage by incorporating appropriate protease inhibitors. Finally, if LgtC phage display is to be performed in the absence of emulsification, kinetic analysis of phage isolated using a CsCl gradient should be performed to determine the level of activity of displayed enzyme.

Unfortunately the final component of the proposed screening procedure, simultaneous display of active LgtC and SAV on the surface of phage, was not achieved. It is possible that further labor intensive optimization, using directed evolution to optimize pIII and pVIII coat proteins or by optimizing transcription/translation levels using a different rbs or different ordering and spacing of gene sequences for polycistronic expression, might overcome the observed dilemma. However, the finding that SAV expression is toxic to *E. coli* makes it an unattractive component for a screening strategy and might lead to artifactual results arising from events leading to the control of SAV expression. As such, an alternative strategy for creating a physical linkage between the glycosyltransferase acceptor and the phage surface is appropriate. One possibility would be the use of a maleimide substrate conjugate. Maleimides react with thiols and, under alkaline conditions, with amino groups, and as such are capable of reacting with multiple sites on the surface of M13 phage. Jestin and co-workers demonstrated by SELDI-MS that a maleimide substrate conjugate could be used to generate a covalent adduct with the pVIII coat protein of M13 phage, by reaction with either the N-terminal Ala residue or the exposed side chain amino substituent of Lys8, and applied this strategy to develop a screening procedure that relied on proximity effects (Jestin et al., 1999). In order to perform directed evolution on LgtC, a maleimide-lactose conjugate could be readily obtained by reaction of the *p*-amino phenyl- $\beta$ -D-lactoside with any of a number of commercially available maleimidating agents. Substituting the biotin-lactose conjugate with this reagent during the compartmentalization procedure would negate the need for SAV display as a pVIII fusion, allowing for the direct use the pLL01 phagemid that

facilitated robust LgtC display. The work described in this section provides a foundation for the future directed evolution of LgtC activity.

## CHAPTER 5

# GLYCOSYLTRANSFERASE SUBSTRATE ENGINEERING

\* Versions of portions of this chapter have been published:

Lairson, L.L., Watts, A.G., Wakarchuk, W.W., and Withers, S.G. (2006) Using Substrate Engineering to Harness Enzymatic Promiscuity and Expand Biological Catalysis. *Nature Chemical Biology*. **2**: 724.

Lairson, L.L., Wakarchuk, W.W., and Withers, S.G. (2007) Alternative Donor Substrates for Inverting and Retaining Glycosyltransferases. *Chemical Communications*. (4): 365. - Reproduced by permission of The Royal Society of Chemistry (RSC)

## **5.1. Summary**

In an effort to increase the synthetic utility of glycosyltransferases, a process of “substrate engineering” was investigated. A strategy for expanding acceptor substrate specificity was developed in which a readily removable functional group was attached to alternative glycosyltransferase acceptor substrates to induce productive binding modes, thereby facilitating rational control of substrate specificity and regio-selectivity using wild-type enzymes. The ability of representatives of both retaining and inverting glycosyltransferases to use inexpensive alternative donor substrates was also explored. Results with these alternative donors provide more fodder for the mechanistic debate surrounding this class of enzyme and present starting points for engineering efforts to increase their potential in enzymatic glycan synthesis.

## **5.2. Engineering Alternative Acceptor Substrates for Glycosyltransferases**

### **5.2.1. Background**

Despite their unparalleled catalytic prowess and environmental compatibility, enzymes have yet to see widespread application in synthetic chemistry. This lack of application and thereby the minimisation of their enormous potential stems not only from a wariness of aqueous biological catalysis on the part of the typical synthetic chemist but also from limitations of applicability that arise from the high degree of substrate specificity possessed by most enzymes. This latter perceived limitation is being successfully challenged through rational protein engineering and directed evolution

efforts to alter substrate specificity as discussed in section 4.3.1. However, such programs require considerable effort to establish.

An alternative approach involving the chemical modification of alternative substrates using readily removable substituents, in a manner that facilitates productive binding to the enzyme by mimicking a natural binding mode, would serve as a useful method for increasing the range of starting materials that can be transformed using a wild-type enzyme and therefore their inherent synthetic utility. An excellent test case for this approach is that of oligosaccharide synthesis using glycosyltransferases, since the chemical synthesis of such molecules remains arduous. In stark contrast to the situation for peptides and oligonucleotides, a lack of synthetic capacity has delayed our understanding of the important roles that glycan structures play in various biological phenomena.

In the case of carbohydrate modifying enzymes, aromatic substituents serve as ideal synthetic mimics of the natural sugar moieties of substrates. Indeed, the use of aryl glycosides in glycosidase kinetic analysis as chromogenic surrogates and as reactive substrates in glycosidase and engineered glycosidase catalysed synthesis is ubiquitous. The ability of aryl substituents to mimic the productive binding modes of sugars is derived by their ability to mimic the hydrophobic platform of natural sugar substrates. Hydrophobic interactions with sugar ring faces are known to provide significant binding affinity and the three-dimensional structures of the sugar binding sites from many carbohydrate-utilizing proteins contain aromatic residues that provide a “hydrophobic platform” onto which a sugar ring face can bind (Nerinckx et al., 2003). In fact, the ability of aromatic substituents to mimic sugar moieties of glycosynthase substrates has

been used to alter their regio-selectivity (MacManus and Vulfson, 2000; Stick et al., 2004).

### **5.2.2. Exploring the Acceptor Substrate Specificity of LgtC**

Glycosyltransferases are generally believed to be highly selective for both the donor and acceptor substrates, although some exceptions have recently been reported (Bencur et al., 2005; Blanco et al., 2001; Borisova et al., 2004; Durr et al., 2004; Oberthur et al., 2005; Yang et al., 2005). As described in detail in chapter 3, LgtC catalyses the transfer of galactose to lactose containing acceptor substrates with net retention of anomeric configuration (scheme 3.1). Previous studies had shown that D-galactose also serves as an acceptor for LgtC, though with greatly reduced efficiency (Ly et al., 2002). To determine the actual degree of acceptor substrate specificity possessed by LgtC, a broad range of sugar and sugar-like substrates were tested as potential substrates for this enzyme. Interestingly, as shown in Table 5.1 and confirmed by ESI-MS and TLC, a range of D-aldose sugars and even *myo*-inositol were found to act as acceptor substrates for LgtC at rates comparable to that for D-galactose.

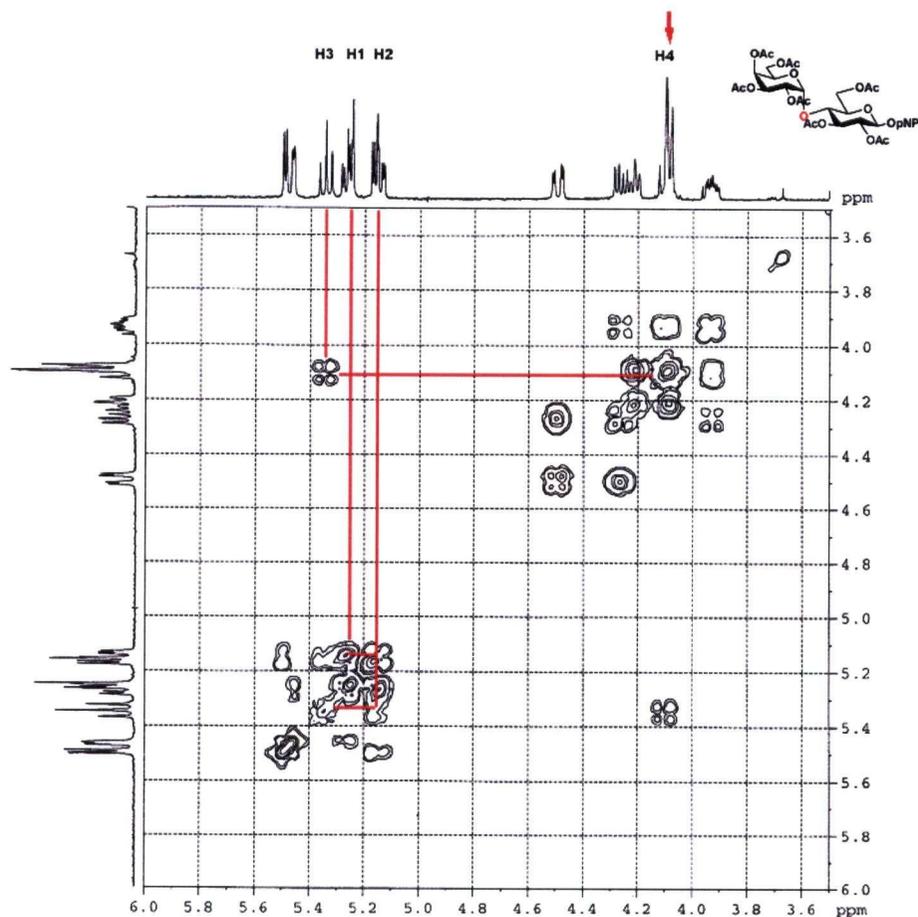
**Table 5.1.** Kinetic parameters for various LgtC acceptor substrates.

Acceptor	$k_{cat} / K_m$ ( $\text{min}^{-1} \cdot \text{mM}^{-1}$ )
Lactose	240
D-Galactose <sup>a</sup>	0.3
D- Lyxose <sup>a</sup>	0.3
D-Arabinose <sup>a</sup>	0.1
D-Glucose <sup>a</sup>	0.2
D-Allose <sup>a</sup>	0.05
D-Mannose <sup>a</sup>	0.3
D-Xylose <sup>a</sup>	1.8
Myo-inositol <sup>a</sup>	0.2
Benzyl- $\beta$ -D-cellobioside	1.7
Benzyl- $\beta$ -D-xyloside	40
Octyl- $\beta$ -D-glucoside <sup>a</sup>	1.4
Octyl- $\beta$ -D-galactoside <sup>a</sup>	3.8
Octyl- $\beta$ -D-xyloside <sup>a</sup>	22
6-OBz-D-Man	5.0
4-OBz-D-Xyl <sup>a</sup>	12

a - no saturation of enzyme activity was observed up to near saturated concentrations. Reliable  $k_{cat}$  and  $K_m$  values were therefore only attainable for Lactose ( $k_{cat} = 160 \text{ s}^{-1}$ ,  $K_m = 40 \text{ mM}$ ), Benzyl- $\beta$ -D-cellobioside ( $k_{cat} = 1.0 \text{ s}^{-1}$ ,  $K_m = 35 \text{ mM}$ ), Benzyl- $\beta$ -D-xyloside ( $k_{cat} = 10 \text{ s}^{-1}$ ,  $K_m = 15 \text{ mM}$ ) and 6OBz Man ( $k_{cat} = 2.5 \text{ s}^{-1}$ ,  $K_m = 30 \text{ mM}$ ).

### 5.2.3. Determining the Ability of LgtC to Transfer to Equatorial Hydroxyl Substituents

This wide acceptor range was surprising, especially given that some of these sugars possess only equatorial hydroxyl groups yet LgtC transfers to an axial hydroxyl in its natural substrate. To determine whether LgtC was indeed capable of transferring to an equatorial hydroxyl, the product formed when *p*-nitrophenyl  $\beta$ -D-glucopyranoside (pNP Glc) acts as acceptor was investigated. Analysis of the single disaccharide product, by ESI-MS and two-dimensional  $^1\text{H}$  NMR spectroscopy (Figure 5.1), clearly revealed the formation of a Gal- $\alpha$ -(1,4)-Glc linkage, indicating that pNP Glc adopts the same basic binding mode as lactose, but that transfer occurs to the equatorial 4-hydroxyl. Assignments were made by examining the couplings within the ring system and the locations of glycosidic linkages were determined on the basis of chemical shift. Because the acetate group is more electron withdrawing than a glycoside substituent, the proton attached to the carbon involved in the glycosidic bond resonates further upfield (<4.50 ppm) in a region clearly distinct from anomeric protons or protons attached to carbons bearing an acetate group.



**Figure 5.1.** <sup>1</sup>H-COSY NMR spectrum of *para*-nitrophenyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 4)-2,3,6-tri-O-acetyl- $\beta$ -D-glucoside.

#### 5.2.4. Exploring the Synthetic Utility of LgtC Catalysed Galactosyl Transfer to Alternative Monosaccharide Acceptor Substrates

This finding of relatively broad specificity led to an initial hope that wild-type LgtC could be used as a catalyst in a range of useful galactosylation reactions. Disaccharide products derived from two of the alternative acceptor monosaccharides studied, D-mannose and D-xylose, were therefore acetylated, purified and subjected to

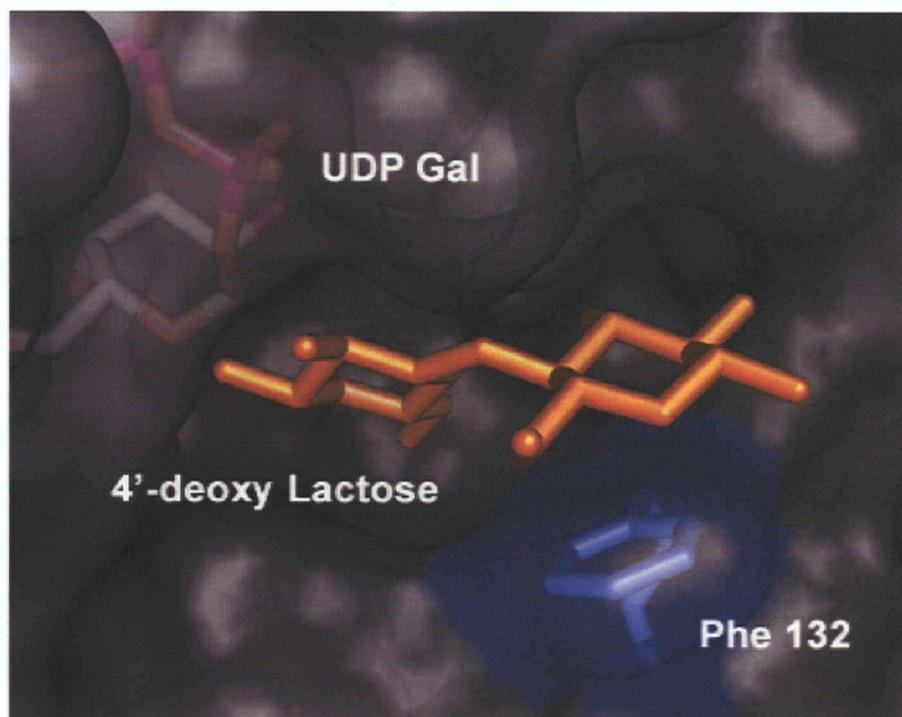
ESI-MS and NMR structural analysis. In both cases, although MS analysis revealed masses consistent with that of disaccharide product,  $^1\text{H}$  NMR analysis revealed the purified products to be mixtures of regio-isomers. This finding was quite disappointing since a catalyst that yields a mixture of regio-isomers is clearly of limited utility.

### **5.2.5. A Substrate Engineering Strategy that Uses Aromatic Substituents as Active Site Anchors**

The fact that a single regio-isomer was obtained when pNP Glc was used as acceptor suggested a strategy in which the appendage of a conveniently installed and removed substituent onto the acceptor might control its active site binding orientation in a useful manner. In fact, as mentioned, the sugar binding sites from many carbohydrate-utilizing proteins contain aromatic residues that provides a sugar binding “hydrophobic platform” (Nerincx et al., 2003). Inspection of the three-dimensional structure of LgtC (pdb 1ga8) reveals that the aromatic side chain of Phe132 is positioned to function in such a role, providing a hydrophobic platform onto which the glucose ring of the lactose acceptor or aromatic moieties may bind (Figure 5.2).

This process of substrate engineering was further explored using LgtC. A convenient appendage at the anomeric centre is a benzyl group, which can be installed by classical glycoside formation chemistry and which can be removed by hydrogenation. Correspondingly, benzyl  $\beta$ -D-cellobioside and benzyl  $\beta$ -D-xyloside were synthesized and tested as acceptors. Both TLC and ESI-MS analysis clearly indicated that they both functioned well in this role with the formation of a single product as had been seen with pNP Glc. Measurement of reaction rates revealed that, not only did the inclusion of an

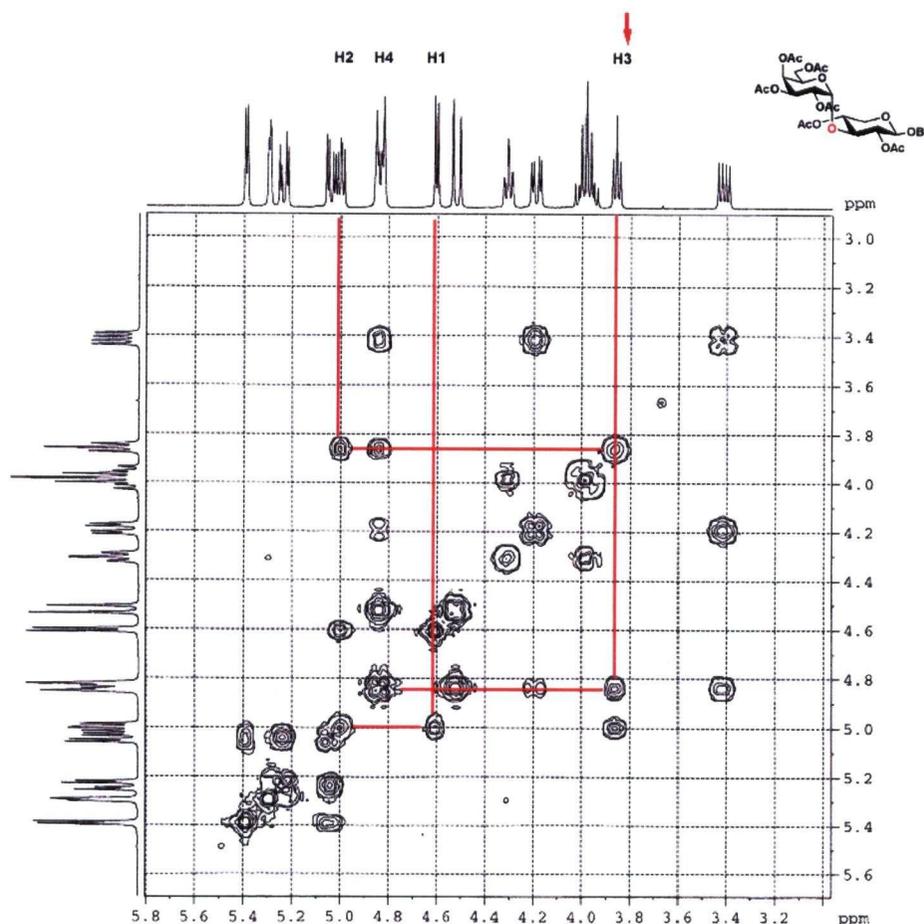
aromatic substituent control regiochemistry, but also it resulted in substantially improved reaction rates, with the  $k_{\text{cat}}/K_{\text{m}}$  value of benzyl  $\beta$ -D-xyloside being over 20-fold higher than that for its parent sugar and within a factor of four of that for the natural acceptor lactose (Table 5.1).



**Figure 5.2.** Acceptor binding site of LgtC containing a hydrophobic platform (Phe132) for binding the glucose sugar ring of lactose or an aromatic substituent (pdb 1ga8).

Analysis of the product formed from benzyl  $\beta$ -D-xyloside by two-dimensional  $^1\text{H}$  NMR spectroscopy revealed the exclusive formation of an  $\alpha$ -1,3-linkage (Figure 5.3), indicating that xylose adopts an alternative binding mode within the LgtC active site. The formation of 1,3-linkages to xylosides has been observed in transglycosylation reactions

with several glycosidases and glycosynthases with preferences, otherwise, for 1,4-linkages, providing precedence for this alternative binding mode (Mackenzie et al., 1998).

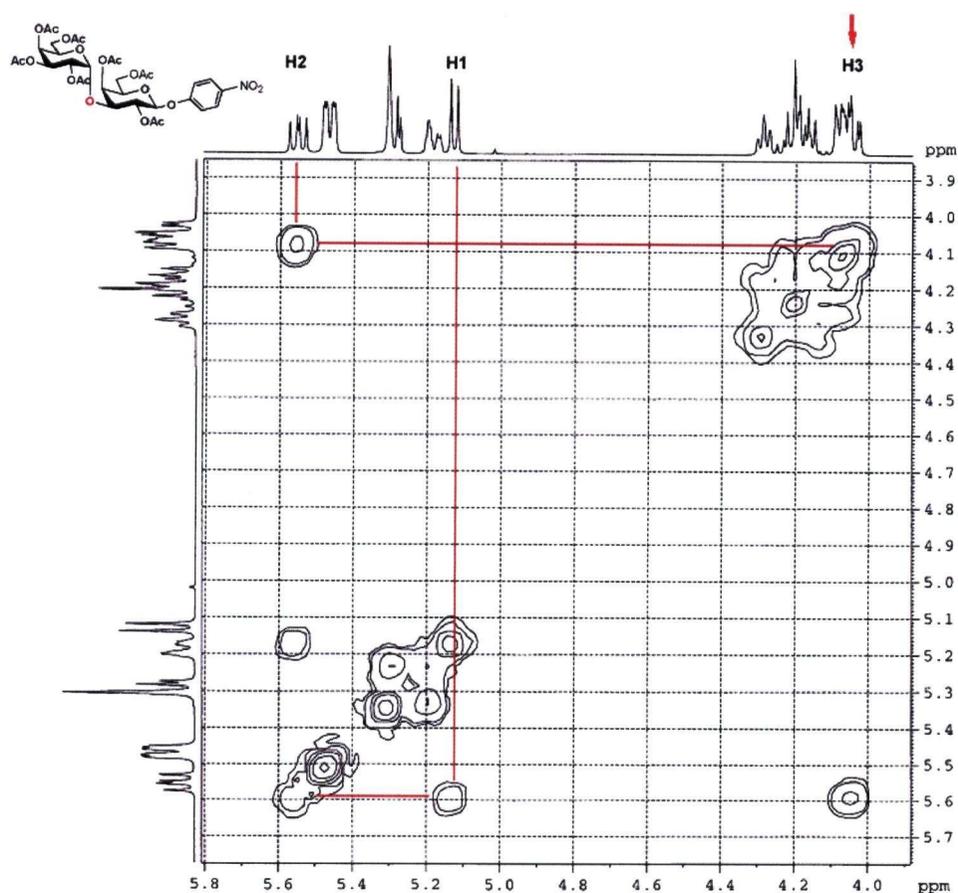


**Figure 5.3.** <sup>1</sup>H-COSY NMR spectrum of benzyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 3)-2,4-di-O-acetyl- $\beta$ -D-xyloside.

### 5.2.6. Using Substrate Engineering to Control Regio-Selective $\alpha$ 3GalT Catalysed Galactosyl Transfer to a Monosaccharide Acceptor Substrate

As described in detail in section 3.4.,  $\alpha$ 3GalT is known to transfer galactose from UDP galactose to terminal *N*-acetylglucosamine-containing glycoconjugates producing

Gal- $\alpha$ -(1,3)-Gal- $\beta$ -(1,4)-GlcNAc epitopes (Scheme 3.5). The ability of  $\alpha$ 3GalT to use D-galactose as an acceptor was investigated. As was observed with LgtC, transfer to an underivatized monosaccharide led to the formation of a mixture of regio-isomers of disaccharide products as determined by  $^1\text{H}$  NMR analysis, whereas the presence of an aromatic substituent led to the exclusive formation of a pure  $\alpha$ -(1,3)-linked regio-isomer when pNP Gal was used as acceptor (Figure 5.4.).



**Figure 5.4.**  $^1\text{H}$ -COSY NMR spectrum of *para*-nitrophenyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 3)-2,4,6-tri-O-acetyl- $\beta$ -D-galactoside.

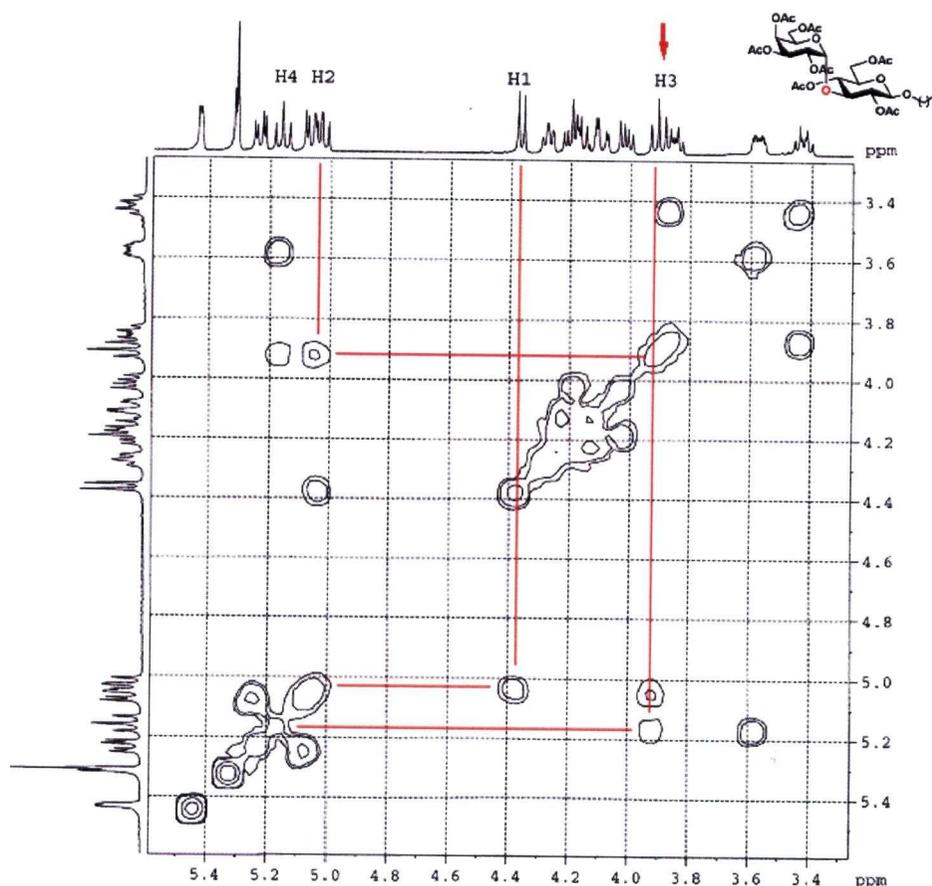
### **5.2.7. Using Substrate Engineering to Facilitate Cst II Catalysed Sialyl Transfer to a Monosaccharide Acceptor Substrate**

Cst II catalyses the transfer of sialic acid (NeuAc) from cytidine 5'-monophospho-N-acetylneuraminic acid (CMP NeuAc) to terminal lactose or 3'-sialyl lactose-containing glycoconjugates yielding  $\alpha$ -(2,3) and  $\alpha$ -(2,8) linked products respectively (scheme 2.1). Due to the high rate of enzyme catalysed hydrolysis of the donor substrate (Chiu et al., 2004), when underivatized D-galactose was used as acceptor, no product formation was observed. However, when the D-galactose was derivatized with an aromatic substituent (pNP Gal), quantitative product formation was observed in a completely regio-selective manner.

### **5.2.8. The Nature of the Substituent on an Alternative Acceptor Substrate can affect the Regio-Selectivity of LgtC**

Unexpectedly, when reactions of monosaccharides containing aromatic substituents were performed using LgtC from crude cell lysates containing the commercial protein extraction reagent BugBuster<sup>TM</sup> (Novagen), a disaccharide product was isolated that did not contain aromatic <sup>1</sup>H signals but rather possessed an eight carbon saturated alkyl chain. Inspection of the <sup>1</sup>H and <sup>13</sup>C NMR spectra revealed the product to be derived from galactosylation of the non ionic detergent octyl- $\beta$ -D-glucoside that is present in the BugBuster<sup>TM</sup> preparation. However, in contrast to the Gal- $\alpha$ -(1,4)-Glc linkage observed when pNP Glc was used as acceptor, the single product obtained from the alkyl-substituted substrate possessed a Gal- $\alpha$ -(1,3)-Glc linkage, as determined by the

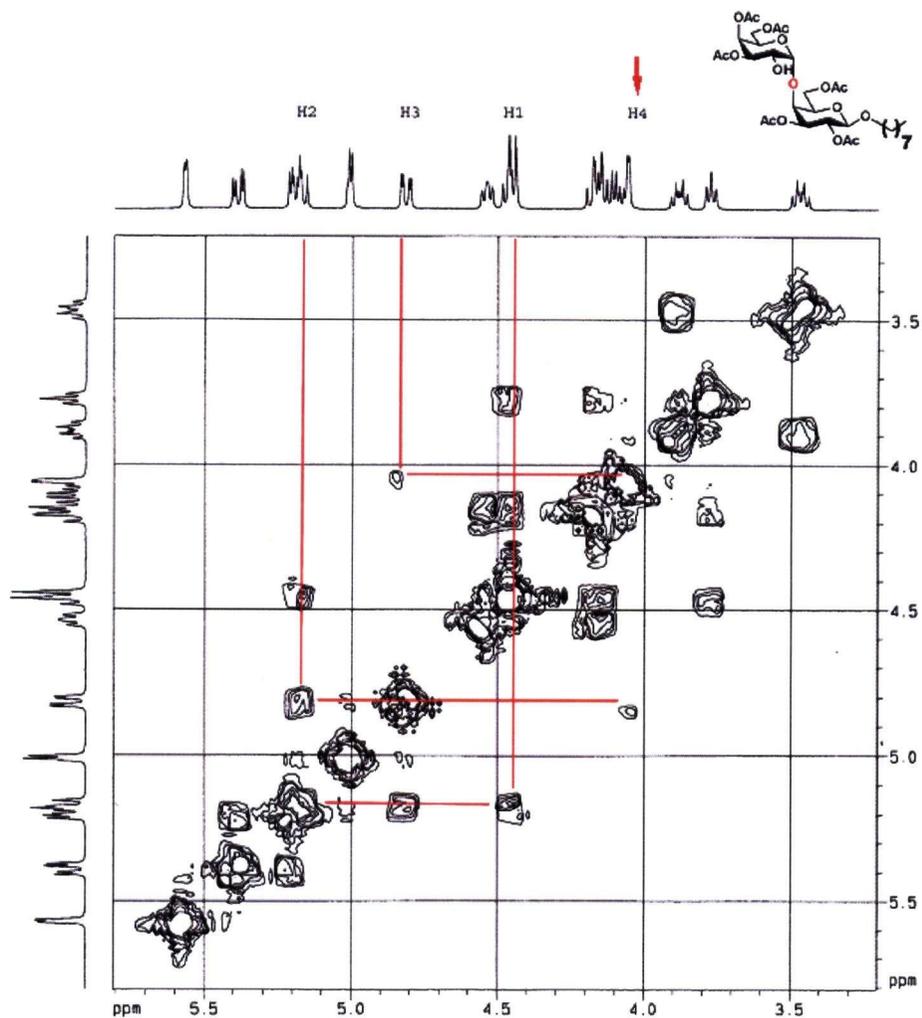
$^1\text{H}$ -COSY NMR spectra (Figure 5.5). This result was confirmed using authentic octyl- $\beta$ -D-glucoside and highly purified LgtC and all subsequent reactions were performed with pure enzyme to avoid this complication.



**Figure 5.5.**  $^1\text{H}$ -COSY NMR spectrum of *n*-octyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 3)-2,4,6-tri-O-acetyl- $\beta$ -D-glucoside.

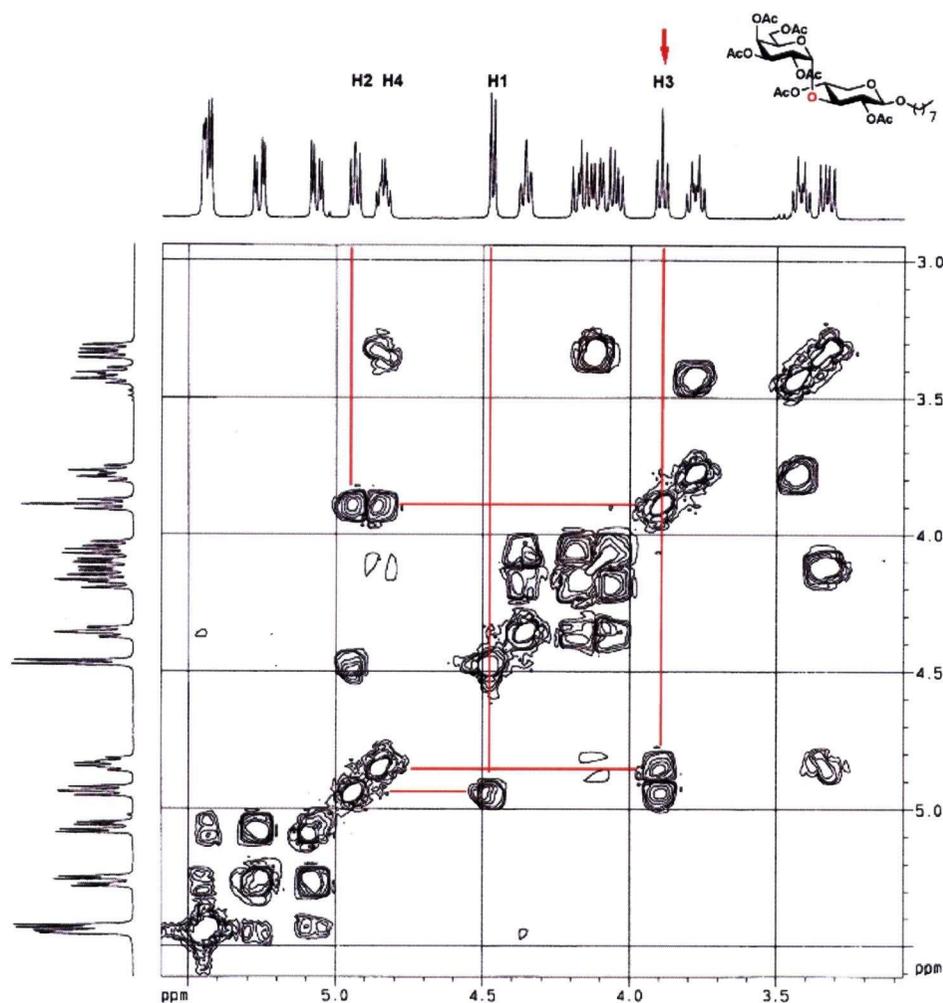
Thus, while in both cases chemical modification of the alternative substrate with either an aromatic or alkyl substituent provides control of regio-selectivity and leads to the formation of a single product, the nature of the substituent facilitated alternative productive binding modes, resulting in the production of *different* regio-isomers. Alkyl

galactoside and xyloside derivatives were also tested as alternative acceptors. In the case of octyl- $\beta$ -D-galactoside, presumably, interactions with the natural galactose moiety dominate, resulting in the exclusive formation of a Gal- $\alpha$ -(1,4)-Gal linkage (Figure 5.6).



**Figure 5.6.**  $^1\text{H}$ -COSY NMR spectrum of *n*-octyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 4)-2,3,6-tri-O-acetyl- $\beta$ -D-galactoside.

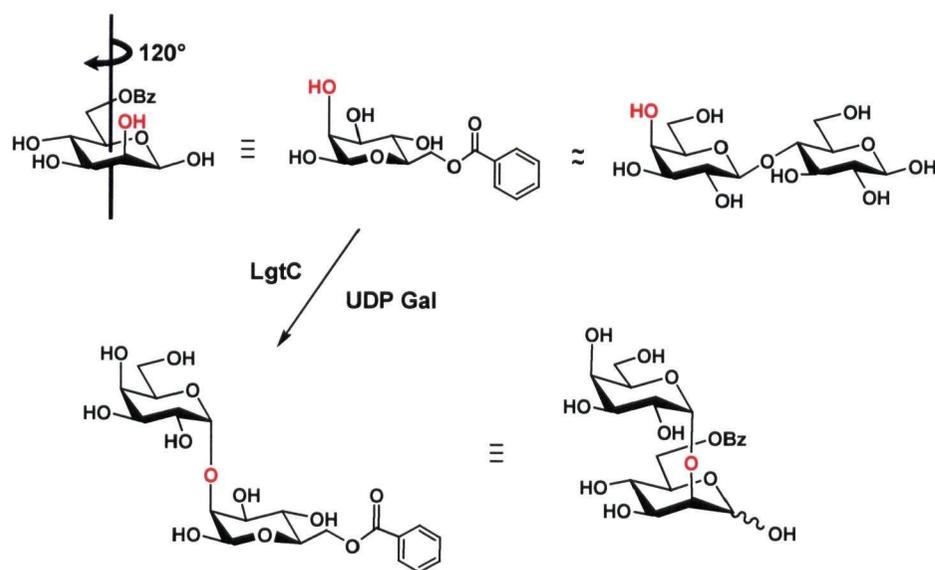
As the aromatic derivative of xylose was already found to adopt an alternative binding mode, it was not unexpected that octyl- $\beta$ -D-xyloside also yielded a Gal- $\alpha$ -(1,3)-Xyl linkage (Figure 5.7). The ability of a glycosyltransferase to use these alkylated lipid-like derivatives could prove useful in combinatorial oligosaccharide synthesis and in the production of novel glycolipids (Miura et al., 1996).



**Figure 5.7.**  $^1\text{H}$ -COSY NMR spectrum of *n*-octyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 3)-2,4-di-O-acetyl- $\beta$ -D-xyloside.

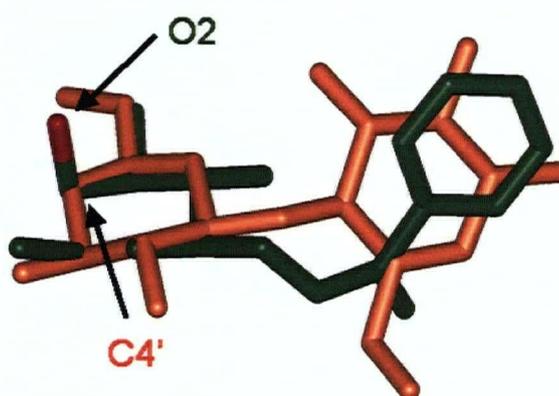
### 5.2.9. Controlling the Regio-Selectivity of LgtC-Catalysed Galactosyl Transfer to Mannose

This flexibility of sugar binding mode opened up the intriguing possibility that the regiochemical outcome could be further controlled by appending an aryl group at other positions on the sugar ring as has been reported for glycosidases and glycosynthases (MacManus and Vulfson, 2000; Stick et al., 2004). For aldohexoses, the most convenient substituent is a benzoate ester, which can be installed selectively at the 6-position at low temperatures, with limiting quantities of acylating agent, and removed under a variety of mild conditions. As shown in Figure 5.8, a binding mode for a 6-O-benzoyl sugar could be envisioned in which the sugar 2-hydroxyl group adopts the same position as the 4-hydroxyl of the galactose and the aromatic substituent takes the place of the glucose ring of lactose.



**Figure 5.8.** A model for the substrate engineering strategy using 6-OBz Man. Two representations of equivalent ( $\equiv$ ) 6-OBz Man illustrating how 6-OBz Man can approximate ( $\approx$ ) a lactose binding mode.

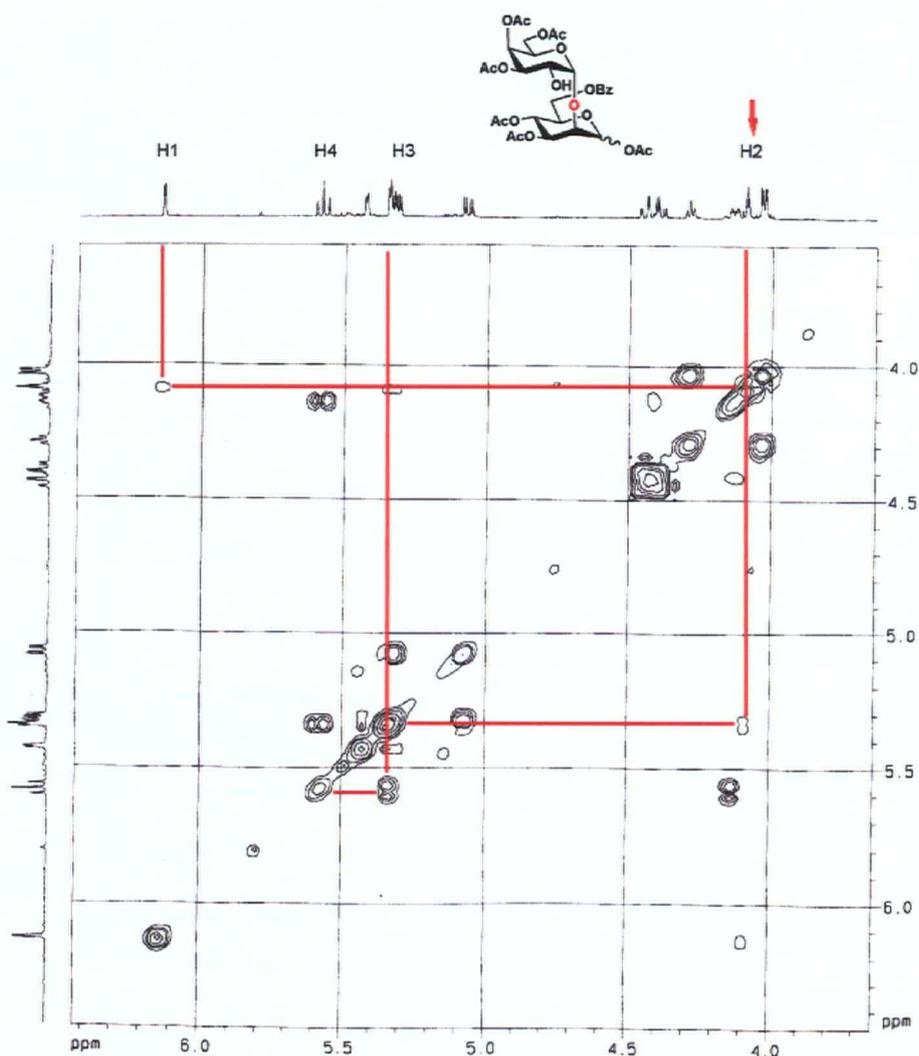
Indeed, in the case of 6-O-benzoyl-D-mannose (6-OBz Man) the hydroxyl group at this position also has the axial configuration. In fact, the molecular mechanics (MM2) energy-minimized structure of 6-OBz Man overlays well with the X-ray crystal structure of 4'-deoxy lactose bound to LgtC, with the 2-OH of mannose oriented in the position that would be occupied by the 4'-OH that undergoes glycosylation in the LgtC catalysed reaction (Figure 5.9).



**Figure 5.9.** Overlay of the MM2 minimized structure of 6-OBz Man (green) with the structure of 4-deoxylactose (orange) bound within the active site of LgtC (pdb 1GA8).

6-OBz Man was readily prepared in one step from D-mannose by low temperature benzylation and tested as an alternative acceptor substrate for LgtC, kinetic analysis revealing that the addition of the benzoyl group to mannose improved  $k_{\text{cat}}/K_m$  by approximately 20-fold (Table 5.1). To determine whether the appended aromatic functional group also provides control of regio-selectivity, the single product formed was acetylated, purified and subjected to product analysis as described. ESI-MS revealed the expected mass for a disaccharide product, with  $^1\text{H}$  NMR analysis confirming a single, pure regio-isomer.  $^1\text{H}$ -COSY NMR analysis was used to assign all proton signals,

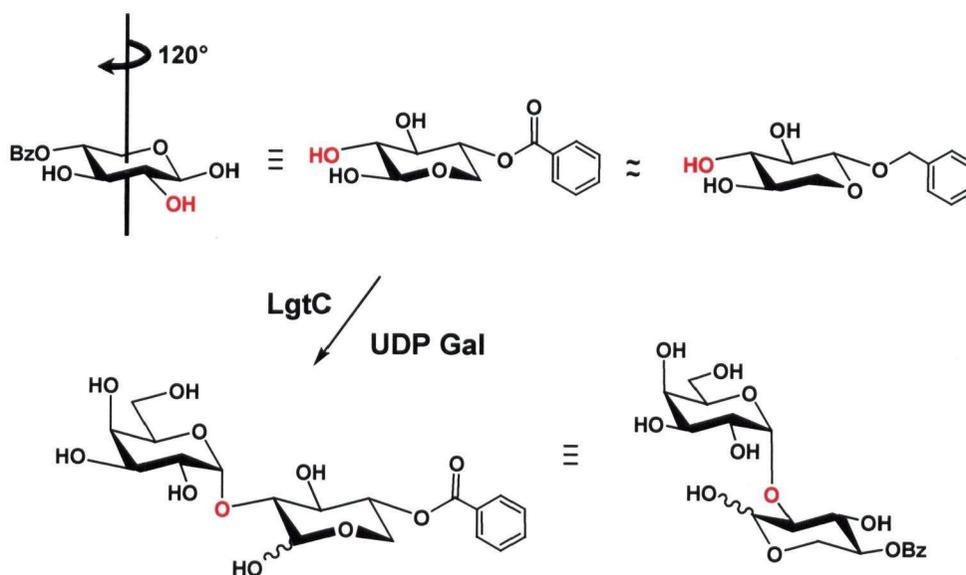
identifying the newly formed glycosidic bond as the predicted Gal- $\alpha$ -(1,2)-Man linkage (Figure 5.10).



**Figure 5.10.**  $^1\text{H}$ -COSY NMR spectrum of (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 2)-1,3,4-tri-O-acetyl-6-O-benzoyl- $\alpha$ -D-mannose.

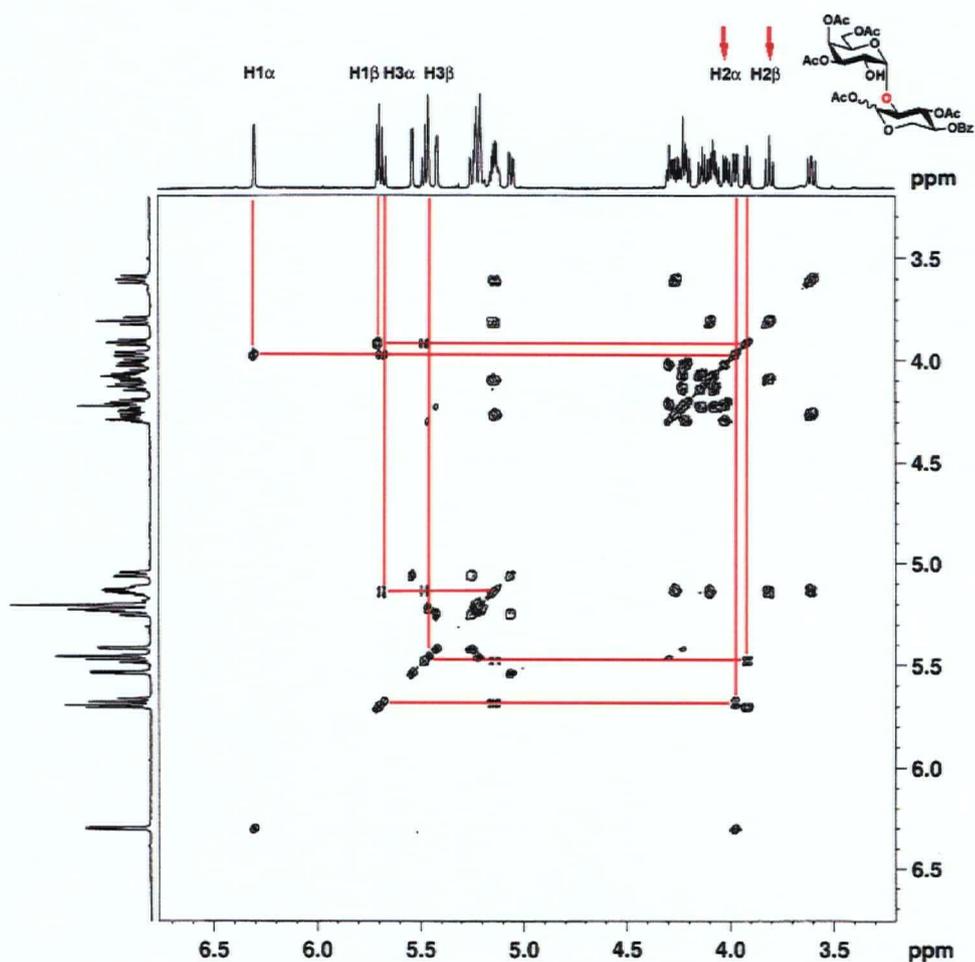
### 5.2.10. Controlling the Regio-Selectivity of LgtC-Catalysed Galactosyl Transfer to Xylose

To further test the predictability of the model, the relatively symmetric xylose moiety was subjected to further derivatization. With a knowledge of the alternative binding mode of xylosides in hand, as shown in Figure 5.11, aromatic derivatization at C4 should allow a binding mode that leads to the formation of an  $\alpha$ -(1,2)-linkage.



**Figure 5.11.** A model for the substrate engineering strategy using 4-O-benzoyl-D-xylose (4-OBz Xyl). Two representations of equivalent ( $\equiv$ ) 4-OBz Xyl illustrating how 4-OBz Xyl can approximate ( $\approx$ ) the observed binding mode of benzyl  $\beta$ -D-xyloside.

4-OBz Xyl was tested as an alternative acceptor substrate for LgtC-catalysed galactosylation. Appendage of the aromatic group again resulted in an increase in the observed rate constant ( $\sim 7$ -fold) (Table 5.1), while product analysis (ESI-MS, and  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^1\text{H}$ -COSY NMR) revealed exclusive production of the anticipated  $\alpha$ -1,2-linked disaccharide (Figure 5.12).

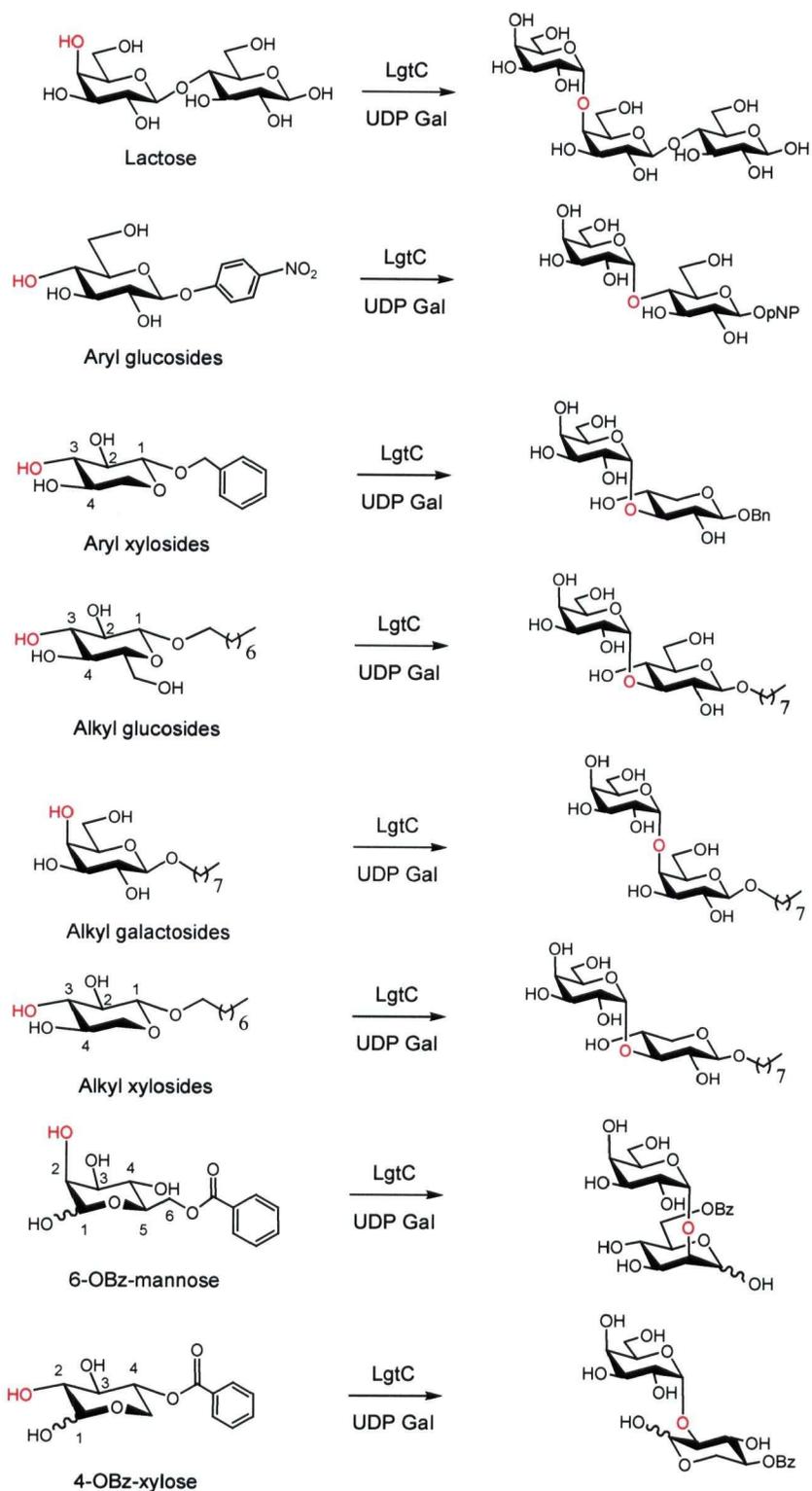


**Figure 5.12.** <sup>1</sup>H-COSY NMR spectrum of (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 2)-1,3,-di-O-acetyl-4-O-benzoyl- $\alpha$ / $\beta$ -D-xylose.

### 5.2.11. Concluding Remarks

By the simple expedient of transient attachment of an alkyl or aryl substituent to the acceptor sugar, the substrate specificity of wild-type LgtC has been broadened to allow exclusive formation of  $\alpha$ -1,2,  $\alpha$ -1,3 or  $\alpha$ -1,4 linkages at synthetically useful rates (Figure 5.13). Besides demonstrating an alternative strategy for expanding the synthetic utility of enzymatic catalysis by transient chemical modification (substrate engineering),

the findings presented here illustrate four points of significance. Firstly, they highlight the need for caution in assigning the natural role of glycosyl transferases when using unnaturally modified acceptors. Indeed, potential acceptor substrates bearing alkyl or aromatic substituents are frequently used to deduce the natural activities of newly isolated glycosyltransferases. Secondly, they demonstrate that the acceptor substrate specificity of glycosyltransferases is not quite as rigid as previously believed, with significant implications for their synthetic utility. Thirdly, this promiscuity can be harnessed by modification of the acceptor sugar with a readily removable aromatic substituent or a simple alkyl chain that facilitates a productive and predictable binding mode providing both control of regio-selectivity and an increase in observed rates. Finally, the described strategy provides facile access to glycosidic linkages for which no 'wild-type' enzymatic activities have yet been identified. This glycosyltransferase substrate engineering strategy should prove useful in the development of libraries of oligosaccharide structures or in providing facile synthetic access to a given synthetic target by allowing a single protein catalyst to be used in a broad range of reactions.



**Figure 5.13.** Substrate engineering strategy applied to various alternative LgtC acceptor substrates. Derivatization of alternative acceptors is used to generate substrates that mimic the lactose binding mode for LgtC-catalysed galactosyl transfer.

## **5.3. Engineering Alternative Donor Substrates for Glycosyltransferases**

### **5.3.1. Background**

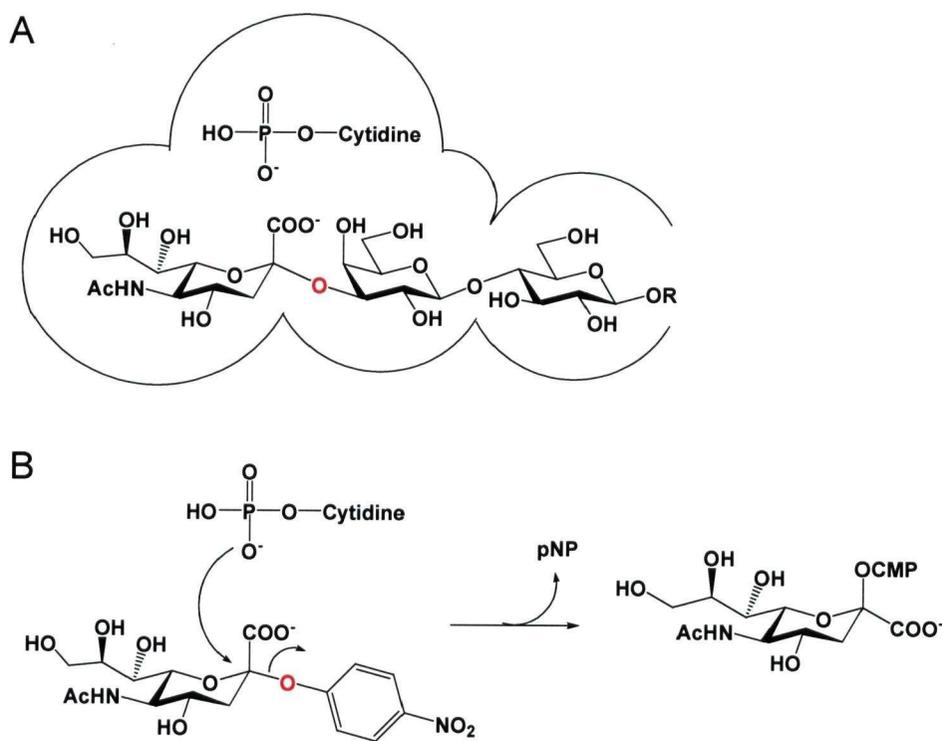
The application of glycosyltransferases in oligosaccharide synthesis has suffered from both a lack of access to the enzymes themselves, as most exist as membrane-associated species, and also from the cost of high energy glycosyl donor substrates (typically nucleotide sugars). With bacterial enzymes, the former problem is currently being overcome by successful cloning strategies involving truncations to yield recombinant soluble catalytic domains. Attempts to overcome the latter problem involve elaborate coupled enzyme recycling schemes involving *in situ* generation of the donor substrate (Gilbert et al., 1998; Ichikawa et al., 1991). A recently described alternative strategy takes advantage of the reversibility of natural product glycosyltransferases for *in situ* generation of desired nucleotide sugars (Zhang et al., 2006a). The increasing accessibility of the pyrophosphorylases and nucleotidyltransferases responsible for the formation of nucleotide sugars should further facilitate these approaches, however, the technical complexity and problems of product inhibition provide an impetus for exploring potential alternative donor substrates for glycosyltransferases.

### **5.3.2. An Activated Alternative Donor Substrate for Cst II that Mimics the Product Binding Mode**

As described in detail in chapter 2, Cst II is an inverting bifunctional  $\alpha$ -2,3/2,8 sialyltransferase that uses  $\beta$ -linked CMP sialic acid as a donor substrate and transfers the sialic acid moiety with net inversion of anomeric configuration to the 3' hydroxyl of

terminal lactose-containing acceptors (Scheme 2.1). Subsequently, Cst II will use this 3'-sialyl lactose product as an acceptor and transfer sialic acid to the 8'' hydroxyl. The three dimensional structure of this enzyme and the description of a catalytic mechanism that appears to involve a straightforward direct displacement reaction whereby active site residues facilitate departure of the CMP leaving group and activate the incoming nucleophile of the acceptor was presented in detail in chapter 2.

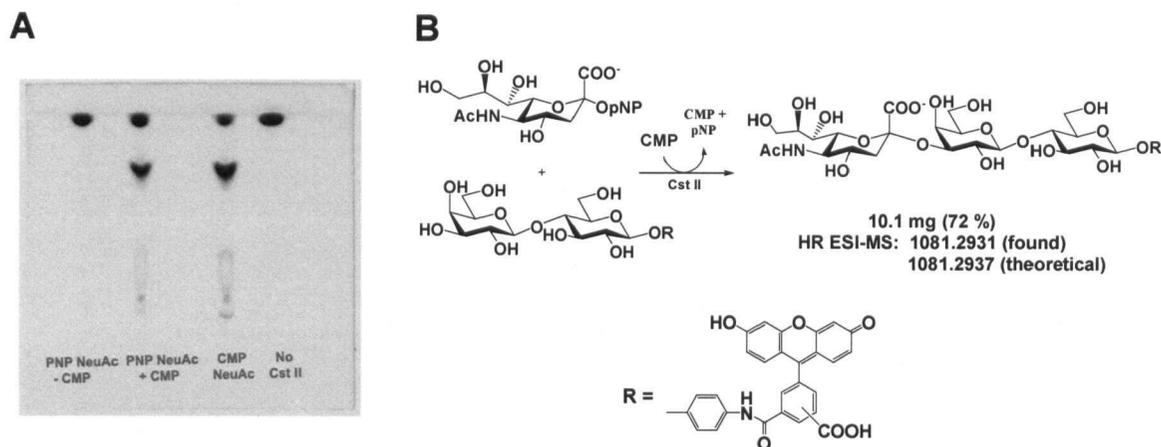
To determine whether this enzyme could use an alternative source of activated sialic acid as a donor substrate, we chose to investigate the  $\alpha$ -linked *para*-nitrophenyl derivative. Providing potential precedence for their use as alternative substrates for glycosyltransferases, nitrophenyl derivatives of ribose have been successfully employed as alternative substrates and thence used as mechanistic probes for both purine nucleoside phosphorylase and N-ribohydrolase (Mazzella et al., 1996). Because Cst II is an inverting enzyme, it was presumed that the aromatic ring of the nitrophenyl substituent could be accommodated within the active site by taking the place of the galactose ring of the acceptor which must be located on the  $\alpha$  face of the sialic acid donor (Figure 5.14). With the pNP  $\alpha$ -D-sialoside bound in such a manner, CMP could also be accommodated within the active site on the  $\beta$  face of sialic acid, allowing for the direct displacement of *p*-nitrophenolate by CMP and the *in situ* formation of  $\beta$ -linked CMP sialic acid. This could then be used as the donor substrate in a subsequent transfer reaction via the normal mechanism following release of pNP.



**Figure 5.14.** Representation of the product binding mode of Cst II (A) and the proposed mode of binding of  $\alpha$ -linked pNP NeuAc that would facilitate *in situ* generation of the natural donor substrate.

Enzyme-catalysed sialic acid transfer was monitored by TLC using a fluorescein-lactose conjugate acceptor and as is shown in Figure 5.15A, Cst II was indeed found to catalyse transfer to the fluorescent acceptor but only in the presence of CMP. Subsequent preparative scale reactions were performed to determine the efficiency of the reaction (Figure 5.15B). Product formation was confirmed by HR ESI-MS and by treatment of the purified reaction product with an  $\alpha$ -2,3 specific neuraminidase (see methods section). Although reported yield is appreciable, the synthetic utility of this particular process is limited by the high cost of pNP NeuAc, being approximately equal to the that of CMP NeuAc. However, application of this approach to other inverting glycosyltransferases, for which the cost of aryl derivatives of donor substrates are substantially less expensive than

their nucleoside diphosphate counterparts, should provide an economical alternative approach to syntheses involving this class of catalyst.



**Figure 5.15.** (A) Monitoring Cst II catalysed glycosylation of fluorescent fluorescein lactose conjugate acceptor substrate using *p*-nitrophenyl  $\alpha$ -sialoside as an alternative donor substrate in the presence or absence of CMP by TLC (4:2:1:0.1 EtOAc: MeOH: H<sub>2</sub>O: AcOH). (B) Reaction catalysed by Cst II using the alternative donor substrate *p*-nitrophenyl  $\alpha$ -sialoside.

Direct formation of the proposed CMP NeuAc intermediate could not be detected either in the presence or absence of acceptor substrate by either negative ion mode ESI-MS or TLC analysis. This is not surprising giving the labile nature of this material and the very low concentrations that would exist in solution. In addition, due to the apparent low affinity of Cst II for the donor substrate analogue and absorbance interference with the pNP substituent in the UV range used to measure rates with the NADH dependent enzyme-coupled assay, rates of product or CMP NeuAc formation could not be obtained.

### 5.3.3. An Activated Alternative Donor Substrate for LgtC

#### 5.3.3.1. Testing $\alpha$ -Linked Aryl Galactosides

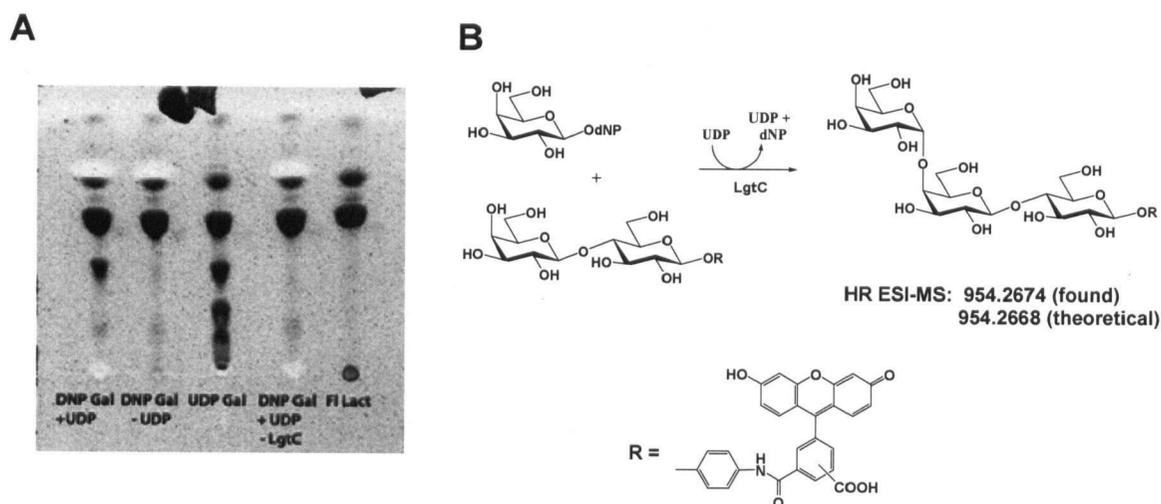
To further explore the generality of this substrate promiscuity amongst glycosyltransferases, a similar strategy was applied to the retaining  $\alpha$ -1,4 galactosyl transferase LgtC. As described in section 3.2, LgtC catalyses the transfer of galactose to the 4' hydroxyl of terminal lactose-containing acceptor substrates using  $\alpha$ -linked UDP galactose as the donor with overall net retention of anomeric configuration (Scheme 3.1). It has previously been observed that LgtC will use  $\alpha$ -galactosyl fluoride as a donor substrate analogue in the presence of UDP (Lougheed et al., 1999). In this study, the *in situ* formation of UDP Gal arising from substitution of the reactive fluoride group was demonstrated using an enzyme-coupled assay involving UDP Glc dehydrogenase (Lougheed et al., 1999). This finding is consistent with either of the proposed mechanisms of retaining glycosyltransferases. The observed activity could occur via a double displacement mechanism involving initial displacement of fluoride ion by a catalytic nucleophile, resulting in the transient formation of a covalent glycosyl-enzyme intermediate, followed by subsequent displacement by an incoming nucleophile (either the acceptor or UDP). Alternatively, an oxocarbenium ion is formed, with a lifetime that allows for a degree of departure by the fluoride ion leaving group sufficient for attack by the incoming nucleophile on the same face. To further test the limits of donor substrate promiscuity, LgtC was tested for its ability to use  $\alpha$ -linked pNP or dNP galactosides as the surrogate donor substrate. Again, glycosyl transfer was monitored by TLC using a fluorescein-lactose conjugate acceptor. Under these conditions, no trisaccharide product

was observed. This is not surprising as the dNP leaving group would have to occupy the UDP binding site of the enzyme and the binding of UDP to LgtC is known to induce a conformational change that results in the closure of a loop that makes up a significant portion of the active site (Persson et al., 2001). Therefore, an  $\alpha$ -linked alternative leaving group must be of sufficiently small size to allow for simultaneous binding of UDP, as is the case with  $\alpha$ -galactosyl fluoride.

### 5.3.3.2. Testing $\beta$ -Linked Aryl Galactosides

A strategy analogous to that described above for Cst II was then applied to LgtC in which an alternative donor with an activated leaving group of anomeric configuration *opposite* to that of the natural donor was tested. In this case, because LgtC is a retaining enzyme, the aromatic leaving group would not occupy the acceptor site as both the leaving group and the incoming nucleophile are present on the same face of the donor sugar in the active site of this class of enzyme. As such, success of this approach would rely upon fortuitous accommodation of an aromatic substituent on the  $\beta$  face of the donor within the LgtC active site. Initially, pNP  $\beta$ -D-galactoside was tested for its ability to act as a surrogate donor in the presence or absence of UDP, but no trisaccharide product was observed in either case, even at high enzyme concentrations (0.5 mg/mL) and after extended incubation times (5 days). Dependence on leaving group ability was then probed by repeating the above experiments using the more activated derivative 2,4-dinitrophenyl  $\beta$ -galactoside. Using this more activated analogue a trisaccharide was observed and again glycosyl transfer only occurred in the presence of the natural nucleotide (UDP) (Figure 5.16A). Product formation was insufficient to obtain a yield

and was confirmed solely by HR ESI-MS. Analogously to the inverting enzyme, LgtC presumably catalyses the direct displacement of the activated  $\beta$ -linked donor analogue by UDP resulting in the formation of the natural  $\alpha$ -linked UDP galactose donor substrate which is then used in a subsequent glycosylation reaction involving the acceptor substrate to yield a trisaccharide product. However, as was described in section 5.3.2., rates of formation of the product or the putative UDP Gal intermediate could not be determined due to the apparent low affinity binding of the donor substrate analogue and interference of the dNP substituent in the UV range used to measure rates with the NADH dependent glycosyltransferase and UDP Glc dehydrogenase enzyme-coupled assays. The observed rate of product formation was low and this reaction is of limited synthetic utility due to the high rate of spontaneous hydrolysis of the dNP  $\beta$ -D-galactoside donor substrate analogue. However, the results provide a potential starting point for directed evolution efforts to increase this alternative activity and provides potential insight into the structural plasticity of the active site, with associated mechanistic significance, for this enzyme.



**Figure 5.16.** (A) Monitoring LgtC catalysed glycosylation of fluorescent fluorescein lactose conjugate acceptor substrate using 2,4-dinitrophenyl  $\beta$ -galactoside as an alternative donor substrate in the presence or absence of UDP by TLC (7:2:1:0.1 EtOAc: MeOH: H<sub>2</sub>O: AcOH). (B) Reaction catalysed by LgtC using the alternative donor substrate 2,4-dinitrophenyl  $\beta$ -galactoside.

### 5.3.3.3. Mechanistic Implications for LgtC

As described in section 3.2, a catalytically relevant covalent glycosyl-enzyme intermediate was observed for an active site mutant of LgtC, however the aspartate residue that became labelled is separated from the anomeric reaction centre by a distance of 9Å in the available crystal structure. Even if this residue does not play the role of the catalytic nucleophile in a double displacement mechanism, in order for this covalent species to be formed and turned over a significant change in conformation from that observed in the crystal structure must occur during catalysis, thereby indicating a significant degree of structural plasticity. Support for this idea of significant structural plasticity within the LgtC active site is further provided by the above described

accommodation of dNP- $\beta$ -D-galactoside in a catalytically competent conformation, as it is difficult to imagine how this  $\beta$ -linked aryl substituent could be accommodated within the donor substrate binding site based on the available ternary complex crystal structure without active site reconfiguration. Although neither of these results provides definitive support for a given catalytic mechanism used by retaining glycosyltransferases, they do indicate that crystal structures for glycosyltransferase complexes should be interpreted cautiously.

### 5.3.4. Concluding Remarks

Cst II and LgtC were found to use *p*-nitrophenyl  $\alpha$ -sialoside and 2,4-dinitrophenyl  $\beta$ -galactoside, respectively, as alternative donor substrates in the presence of their respective natural nucleotide products. The results of this work illustrate how the donor substrate promiscuity of both inverting and retaining glycosyltransferases can be harnessed, providing the starting point of an alternative strategy for their application as synthetic tools. Both classes of enzymes appear to be able to use glycosyl donors with alternative activated leaving groups of *opposite* anomeric configuration, compared to the natural donor substrate, in the presence of catalytic amounts of the natural nucleotide. Presumably, the enzyme functions by catalysing the *in situ* formation of the natural nucleotide sugar donor, which is then used in a subsequent catalysed glycosylation reaction involving an acceptor substrate.

From the present understanding of structure and mechanism of the inverting Cst II, it is reasonably clear how the enzyme active site accommodates the observed alternate activity. The  $\alpha$ -linked pNP substituent could bind within the acceptor substrate binding

site ready for displacement by CMP bound in its natural site on the  $\beta$  face of sialic acid. It is therefore expected that this approach should be generally applicable amongst other inverting glycosyltransferases.

It is more difficult to rationalize the observed activity with the retaining enzyme LgtC. Based on the available ternary complex crystal structure it is not clear how the  $\beta$ -linked dNP substituent can be accommodated within the active site as both the acceptor and leaving group binding sites lie on the  $\alpha$  face of the galactosyl donor. These results indicate that alternative catalytically relevant conformation from that of the observed LgtC ternary complex crystal structure are possible, consistent with the described labelling of a remote putative catalytic nucleophile. It would be quite interesting to examine this flexibility, by obtaining a three-dimensional X-ray crystal structure of LgtC with a non-reactive aryl analogue (e.g. phenyl  $\beta$ -D-galactoside) and UDP Gal bound, to determine how this alternative substrate is accommodated within the active site in a catalytically competent manner.

## **CHAPTER 6**

### **Summary and Future Directions**

Despite their biological significance and potential applicability in oligosaccharide synthesis, glycosyltransferases remain a relatively poorly characterized and utilized class of enzyme. The work presented in this thesis sought to address deficiencies in mechanistic understanding and to increase the practical applicability of glycosyltransferase catalysis. Mechanistic studies with glycosyltransferases provided insight into the role of convergence in the evolution of enzyme mechanism not only between unrelated classes of glycosyltransferases, but also between different classes of carbohydrate modifying enzymes (i.e. glycosyltransferases and glycosidases). An understanding of the mechanisms of glycosyltransferases facilitated the success of engineering strategies. Conversely, the results of engineering studies provided insight into mechanism.

Inverting glycosyltransferases catalyze glycosyl group transfer with net inversion of stereochemistry at the anomeric reaction centre. The general mechanism of this class of enzyme is that of an  $S_N2$ -like pathway involving direct nucleophilic displacement. The chemical and kinetic mechanism of a representative inverting sialyltransferase known as Cst II was investigated by detailed kinetic analysis, protein X-ray crystallography and site directed mutagenesis. Product and dead-end donor substrate analogue inhibition kinetics, as well as X-ray crystallographic analysis, support a sequential iso-ordered Bi Bi kinetic mechanism for this bifunctional enzyme. Kinetic analysis of mutants of candidate catalytic residue, identified by X-ray crystallographic analysis of Cst II with a bound non-reactive donor substrate analogue, revealed that the tyrosyl hydroxyls of Tyr156 and Tyr162 play critical roles in facilitating departure of the nucleoside monophosphate leaving group. Chemical rescue and kinetic analysis of the H188A mutant, as well as pH

dependence studies, implicates His188 as the catalytic base that activates the incoming nucleophile of the acceptor substrate. Cst II was found to adopt a GT-A fold. Interestingly, the identities of the residues utilized to catalyze group transfer are akin to what has been observed with the GT-B fold enzymes, indicating a convergence in catalytic mechanism between these two inverting clans. Future work with Cst II should include the obtainment of ternary complex X-ray crystal structures with either lactose or 3'-sialyl lactose bound. This will elucidate the nature of the two stable enzyme forms suggested by the proposed iso ordered Bi Bi kinetic mechanism. Conformational analysis of these two structures will provide an understanding of how this enzyme evolved a bifunctional activity.

By simple analogy to the well-characterized retaining glycosidases, retaining glycosyltransferases had been thought to use a double displacement mechanism involving enzymatic nucleophilic catalysis. Such a mechanism necessitates the appropriate positioning of a catalytic nucleophile within the enzyme active site. However, a comparison of the X-ray crystal structures of representative retaining glycosyltransferases from multiple families, which are believed to have all diverged from a common evolutionary origin, indicates a complete lack of conserved structural architecture in the region that would be occupied by this critical catalytic residue. This is in stark contrast to what has been observed for the vast majority of retaining glycosidases, which have not evolved from a common evolutionary origin yet have all converged to have an identically positioned amino acid side chain carboxylate poised to play the role of a catalytic nucleophile. This revelation in itself may well support an argument against a common catalytic mechanism between these two classes of enzymes.

The possibility of a pathway involving nucleophilic catalysis for retaining glycosyltransferases was explored. A representative galactosyltransferase known as LgtC was subjected to a rational protein engineering study with the goal of driving the reaction along a coordinate involving the formation of an intermediate species covalently bound to the enzyme. This work led to the first direct observation of a catalytically relevant covalently bound glycosyl-enzyme intermediate for a retaining glycosyltransferase. However, surprisingly, the site of labelling was found to be a residue that is relatively remote from the anomeric reaction centre in the ground state crystal structure, suggesting the possible identity of a previously unidentified candidate catalytic nucleophile. If this residue were to act as a catalytic nucleophile in a double displacement mechanism, a significant conformational change would have to occur during the course of catalysis from that of the ground state ternary complex crystal structure. This would indicate that the active site conformations of retaining glycosyltransferases are highly plastic. Unfortunately, the exact mechanism of LgtC remains ambiguous. The observed labelling result could either be interpreted as evidence supporting a double displacement mechanism involving a previously unidentified nucleophile or as an artefact resulting from an altered active site electrostatic environment.

A lack of conserved architecture amongst related enzymes, precedence for cationic enzymatic mechanisms, and the inherent differences in reactivities of glycosyltransferase and glycosidase substrates all support a notion that retaining glycosyltransferases and retaining glycosidases have not converged on an identical catalytic mechanism. A likely catalytic pathway for many retaining glycosyltransferases, including LgtC, is that of a  $D_N^*A_{Nssip}$  mechanism involving the formation of a short-lived

ion pair intermediate species. If LgtC were to use such a pathway, it is possible that the highly reactive electrophilic intermediate could react with a nucleophilic side chain within the altered electrostatic environment of the engineered active site thereby accounting for the observed labelling results. Most likely, a mechanistic continuum exists between members of this class of enzyme in which some use discretely  $S_N2$ -like pathways while others use ion pair mechanisms. This would be analogous to what is known for non-enzymatic nucleophilic substitution reactions in which  $S_N1$ -like or  $S_N2$ -like pathways are simply the extremes between which a range of ion pair mechanisms exists. Future work with this class of enzyme should involve obtaining an X-ray crystal structure of the covalently bound intermediate of the LgtC Q189E mutant in order to understand how alternative active site conformations facilitate catalysis. In addition, the protein engineering strategy employed with LgtC should be applied to other retaining enzymes in order to determine if other alternative candidate nucleophiles can be identified.

It was demonstrated that catalytically active LgtC could be displayed on the surface of M13 bacteriophage as a pIII fusion protein. Further, LgtC phage display was successfully performed in the context of a water-in-oil emulsification procedure in which multiple identical phage are produced in isolated  $\sim 5$  fL nanodrops. This procedure has widespread potential in the development of phage display-based directed enzyme evolution screening methodologies that are not reliant on proximity effects. Future work with this emulsified phage display procedure should involve the development of a strategy for covalently modifying the surface of phage with multiple copies of an acceptor substrate. Upon successfully achieving this goal, the procedure should be utilized in directed evolution approaches to alter the activity of glycosyltransferases. This screening

procedure should be generally applicable to the directed evolution of various classes of enzyme catalysts.

A substrate engineering strategy was developed whereby a readily removable substituent is appended to various locations of an *alternative* glycosyltransferase acceptor substrate in a manner that facilitates a productive binding mode. The aromatic substituent mimics the hydrophobic surface of natural sugar substituents. Using this technique, the substrate specificity of wild type LgtC has been broadened to allow exclusive formation of  $\alpha$ -1,2,  $\alpha$ -1,3 or  $\alpha$ -1,4 linkages at synthetically useful rates with various alternative acceptors. Future work with this substrate engineering approach should involve its application to other carbohydrate modifying enzymes and its use in the synthesis of libraries of oligosaccharide structures.

Finally, the ability of representatives of both retaining and inverting glycosyltransferases to use inexpensive alternative donor substrates was explored. It was demonstrated that Cst II and LgtC will use *p*-nitrophenyl  $\alpha$ -sialoside and 2,4-dinitrophenyl  $\beta$ -galactoside, respectively, as alternative donor substrates in the presence of their respective natural nucleotide products. In the case of the inverting enzyme, the engineering strategy was based on an understanding of kinetic mechanism. The alternative donor substrate binds to the enzyme in a mode that mimics the product complex in which the aromatic substituent takes the place of the sugar ring of the trisaccharide product. In the case of the retaining enzyme, the observations are more difficult to rationalize. Binding of the alternative  $\beta$ -linked donor substrate would rely on a fortuitous catalytically active binding mode. Such a binding mode would result in steric clashes in the observed ground state X-ray crystal structure. This result is consistent with

the described labelling results, providing further support for the notion that the active sites of retaining glycosyltransferases are conformationally dynamic species. The ability of glycosyltransferases to use these inexpensive alternative donor substrates provides a starting point for future engineering strategies to increase the utility and economic viability of glycosyltransferase catalyzed synthetic routes. It is believed that the approach should be generally applicable to other inverting glycosyltransferases. Directed evolution should be applied to the inverting enzymes to increase the rates at which these alternative substrates are utilized.

## **CHAPTER 7**

# **EXPERIMENTAL METHODS**

## 7.1. General

All buffers and reagents were from Sigma-Aldrich Chemical Company unless otherwise stated. All  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra were recorded on Bruker Avance-400 or Avance-600 spectrometers. Chemical shifts are reported in  $\delta$  units (ppm) using residual  $^1\text{H}$  and  $^{13}\text{C}$  signals of deuterated solvents as reference:  $\delta_{\text{H}}$  ( $\text{CDCl}_3$ ) 7.27,  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ) 77.23. All  $^{19}\text{F}$  NMR spectra were recorded on a Bruker Avance-300 spectrometer using  $\text{CF}_3\text{CO}_2\text{H}$  as reference and all  $^{31}\text{P}$  NMR spectra were recorded on a Bruker Avance-400 spectrometer using 85%  $\text{H}_3\text{PO}_4$  as reference. Unless otherwise stated, all site directed mutations were generated using the method of QuikChange<sup>TM</sup> (Stratagene) according to the manufacturers' instructions. All standard molecular biology procedures were performed using the appropriate Qiagen kit according to the manufacturers' instructions.

## 7.2. Chapter 2 Experimental

### 7.2.1. Acknowledgements and Generous Gifts

Ms. Melissa Schur at the NRC IBS in Ottawa performed CE analysis of Cst II reaction mixtures. Ms. Cecilia C.P. Chiu in Prof. Natalie Strynadka's laboratory performed crystallization, X-ray diffraction and electron density refinement analysis of Cst II. Mark Chen supplied 3-fluoro sialic acid for the preparation of CMP 3F NeuAc. Dr. A.G. Watts provided *Clostridium perfringens* sialidase. Dr. Warren W. Wakarchuk supplied the H188A, Y156F and R129A mutants of Cst II, FCHASE 3'-sialyl lactose for CE analysis, and purified CMP NeuAc synthetase. CMP NeuAc was supplied as a generous gift from Neose Technologies Inc. (Horsham, PA).

### **7.2.2. Expression and Purification of Cst II**

An overnight culture of *E. coli* (BL21 (DE3)) harboring pET28a(+) containing the Cst II gene was diluted 100 fold in LB media (containing 50 µg/mL kanamycin) and grown to an OD<sub>600</sub> of 0.6 at which point IPTG was added to 0.5 mM and the induced culture was incubated overnight at 20°C. Cells were harvested by light centrifugation (4000 rpm Beckman J2-21) and the resulting cell pellets were resuspended in binding buffer (20 mM Tris (pH 8.3), 500 mM NaCl) containing Roche EDTA-free protease inhibitor cocktail and then disrupted using a French press. Cleared supernatant was loaded onto a charged Ni<sup>2+</sup> IMAC column (5 mL GE Healthcare HisTrap Fast Flow) equilibrated with binding buffer and washed with 10 mL of binding buffer and 10 mL of binding buffer containing 50 mM imidazole. Cst II was eluted using 2 column volumes each of binding buffer containing 100 mM, 250 mM, and 500 mM imidazole. Fractions containing purified Cst II, as determined by SDS PAGE analysis, were combined, concentrated and imidazole was removed using a pre-packed desalting column (PD10 Sephadex G-25M, Pharmacia Biotech) according to manufacturer's instructions. Purified protein was stored in 20 mM Tris (pH 8.3) at 4°C. Typical yields of purified Cst II were ~10 mg/L.

### **7.2.3. A UV Spectroscopy-Based Enzyme-coupled Continuous Assay for Cst II Kinetic Characterization**

A continuous coupled assay, in which the release of CMP is coupled to the oxidation of NADH ( $\lambda = 340 \text{ nm}$ ,  $\epsilon = 6.22 \text{ mM}^{-1}\text{cm}^{-1}$ ), was used to monitor the activity of all Cst II variants. Absorbance measurements were obtained using Cary 300 and Cary 4000 UV-VIS spectrophotometers equipped with a circulating water bath and a Peltier

temperature controller, respectively. GraFit version 4.0 (Erithacus) was used to calculate kinetic parameters by direct fit of initial rates to the respective equations.

The assay conditions in a 200  $\mu\text{L}$  quartz cuvette are as follows: 50 mM HEPES (pH 7.5), 100 mM KCl, 10 mM  $\text{MgCl}_2$ , 10 mM  $\text{MnCl}_2$ , 0.7 mM PEP, 0.25 mM NADH, 2 mM ATP, 1.0 mg/mL BSA, 5.9U of PK, 8.4U of LDH, 7.5U of NMPK, and varying concentrations of acceptor and donor substrates. The cuvette is left at  $37^\circ\text{C}$  until a stable rate of decreasing absorbance at 340 nm is established. This period is required to deplete the CMP present in the donor solution and any NDPs present in the ATP or NMPK solutions. The stable rate of change in absorbance is the result of the spontaneous hydrolysis of the relatively labile CMP NeuAc substrate. Once a stable background rate is established ( $\sim 20$  minutes), 4  $\mu\text{L}$  of transferase solution is added and the solution is mixed thoroughly to initiate the assay. The concentration of enzyme is chosen such that the rate of change in absorbance resulting from transferase activity is constant for a period of more than 10 minutes. The initial rate of transferase activity ( $\mu\text{M}$  /minute) is calculated using the millimolar extinction coefficient of NADH ( $6220 \text{ M}^{-1}\text{cm}^{-1}$ ).

#### **7.2.4. A Capillary Electrophoresis Based Stopped Assay for Cst II Kinetic Characterization**

A stopped assay similar to that described previously was utilized to determine inhibition kinetic parameters for Cst II in the presence of the product inhibitor CMP (Gilbert et al., 1997). Assays were run in 10  $\mu\text{L}$  volumes in triplicate in 50 mM HEPES (pH 7.5) containing 10 mM  $\text{MgCl}_2$  and 24  $\mu\text{g/mL}$  Cst II. When varying the concentration of CMP NeuAc (50 – 500  $\mu\text{M}$ ), FCHASE-3'-sialyl lactose acceptor was held at a fixed concentration of 500  $\mu\text{M}$ . When varying the concentration of FCHASE-3'-sialyl lactose

(0.25 – 1 mM), CMP NeuAc donor was held at a fixed concentration of 500  $\mu$ M. CMP concentration was varied from 0.1 to 1 mM. Assays were run for 10 minutes, during which time the rate of product formation was linear over the range of substrates used, and then quenched by the addition of an equal volume of stop solution consisting of 2% (w/v) SDS, 10 mM EDTA and 50% (v/v) MeOH and stored at  $-20^{\circ}\text{C}$  prior to analysis by CE.

Reaction mixtures were analyzed by CE using a Beckman Instruments PACE 5510 equipped with a 3 mW Argon-ion laser-induced fluorescence detector (excitation at 488 nm, emission at 520 nm). The capillary was of bare silica (75  $\mu$ m x 57 cm) with the detector at 50 cm. Following capillary conditioning (washing with 0.2 M NaOH for 2 minutes, water for 2 minutes and then equilibration with 20 mM sodium phosphate pH 7.5), samples were injected under pressure for 2-5 s and separation was performed at 18 kV, 75  $\mu$ A. Peak integration was performed using the Beckman PACE station software and data was analyzed using Grafit 4.0 (Erithacus software).

### **7.2.5. Synthesis of CMP 3F NeuAc**

CMP 3F NeuAc was synthesized using 3-fluoro sialic, CTP and CMP NeuAc synthetase according to a modified version of a previously published protocol (Simon et al., 1988); (Ohruai, 1988). 3-Fluoro sialic acid (20 mg, 61 $\mu$ mol) and CTP (50 mg, 92  $\mu$ mol) were incubated with CMP NeuAc synthetase (1 mg/mL) and inorganic pyrophosphatase (50 units) in 200 mM Tris (pH 8.5) containing 20 mM  $\text{MgCl}_2$  and 0.2 mM DTT. Prior to the addition of enzymes, the pH was adjusted to 8.5 by the addition of NaOH. The solution was incubated at room temperature for ~18 hours until reaction was complete (as determined by negative ion mode ESI-MS). Following completion of the reaction, enzyme was removed using a 5 kDa molecular weight cut off Centricon filter

(with the retained solution being washed twice with water). The combined aqueous solution containing the desired product was lyophilized, dissolved in 2 mL of water, filtered with a 0.2  $\mu\text{m}$  syringe filter and loaded onto a preparative HPLC TSKgel Amide-80 (Tosoh Bioscience) column. The desired product was purified using a flow rate of 6 mL/minute and a gradient of acetonitrile (A): water (B) (80:20 A:B hold for 20 minutes, linear gradient from 80:20 A:B to 65:35 A:B over 40 minutes, hold 65:35 A:B for 15 minutes). The 6 mL fractions containing pure CMP 3F NeuAc (as determined by negative ion mode ESI-MS) were combined, solvent removed under pressure, and the resulting aqueous solution was lyophilized to yield a white fluffy solid. 29.7 mg (77%);  $^1\text{H-NMR}$  (400 MHz)  $\delta_{\text{H}}$  ( $\text{D}_2\text{O}$ ): 7.85 (d, 1 H, J 7.6 Hz, H6), 6.01 (d, 1 H, J 7.6 Hz, H5), 5.85 (d, 1 H, J 4.4 Hz, H1'), 4.78 (dd, J 48.4 Hz, J 2.0 Hz, H3''), 4.22 – 3.95 (m, 8 H, H2', H3', H5', H5'', H4', H4'', H5'', H6''), 3.84 (m, 1 H, H8''), 3.76 (dd, 1 H, J 2.4 Hz, J 12 Hz, H9''), 3.50 (dd, 1 H, J 6.8 Hz, J 12.0 Hz, H9''), 3.33 (d, 1 H, J 9.6 Hz, H7''), 1.92 (s, 3 H, NAcCH<sub>3</sub>);  $^{19}\text{F-NMR}$  (282 MHz, proton decoupled)  $\delta_{\text{F}}$  ( $\text{D}_2\text{O}$ ); -131.52;  $^{31}\text{P-NMR}$  (162 MHz, proton decoupled)  $\delta_{\text{P}}$  ( $\text{D}_2\text{O}$ ); -4.54; HR ESI-MS:  $m/z = 631.1298$  [ $\text{M}$ ]<sup>-</sup> (expected for  $\text{C}_{20}\text{H}_{29}\text{FN}_4\text{O}_{16}\text{P}^-$ :  $m/z = 631.1300$ ).

### 7.2.6. Generation of Site-Directed Mutants

The Y162F and Y156/162F mutants of Cst II were generated using the following primers. For the Y162F mutant pET28a(+) plasmid DNA containing Cst II was used as template. For the Y156/162F double mutant, pET28a(+) plasmid DNA containing Y156F Cst II was used as template. All constructs were sequenced to verify the presence of the mutation of interest.

Y162Ffwd: 5'- GAA TTG ATT TTT ATC AAA ATG GGT CAT CTT TTG  
CTT TTG ATA CTA AAC AAA AAA ATC TTT T - 3'

Y162Frev: 5'- AAA AGA TTT TTT TGT TTA GTA TCA AAA GCA AAA  
GAT GAC CCA TTT TGA TAA AAA TCA ATT C - 3'

Y156/162Ffwd: 5' - GAA ATT TAT CTT TCG GGA ATT GAT TTT TTT  
CAA AAT GGG TCA TCT TTT GCT TTT G - 3'

Y156/162Frev: 5'- CAA AAG CAA AAG ATG ACC CAT TTT GAA AAA  
AAT CAA TTC CCG AAA GAT AAA TTT C - 3'

### **7.2.7. Chemical Rescue of the H188A Mutant of Cst II**

Chemical rescue of the H188A mutant of Cst II was performed using the coupled continuous assay described in section 6.2.3.. Initial rates for the H188A (500  $\mu\text{g/mL}$ ) and wild-type (25  $\mu\text{g/mL}$ ) enzyme were determined at 0.5 mM CMP NeuAc with varying concentrations of sodium formate (0 - 1000 mM) or sodium azide (0 - 250 mM). Formation of an  $\alpha$ -linked sialyl azide derivative was confirmed by negative ion mode ESI-MS and by the observed ability of *Clostridium perfringens* sialidase to convert the putative species to sialic acid as determined by TLC analysis (4: 2: 1: 0.1 EtOAc: MeOH: H<sub>2</sub>O: AcOH).

## **7.3. Chapter 3 Experimental**

### **7.3.1. Acknowledgements and Generous Gifts**

Dr. Warren W. Wakarchuk provided pCW plasmid DNA containing WT, Q189E, and Q189A versions of LgtC-25 and pCW-Bov10 plasmid containing the gene encoding  $\alpha$ 3GalT-MalE fusion protein. Ms. Aidha Shaikh performed expression optimization of the I143D, N153D and Q185D variants of LgtC-25, as well the protein expression,

purification, kinetic and ESI-MS analysis of the Q185D variant of LgtC-25. Dr. Andrew G. Watts provided 2-deoxy-2-fluoro- $\alpha$ -D-mannose-1-phosphate for the synthesis of GDP 2F Man. Ms. Cecilia C.P. Chiu in Prof. Natalie Strynadka's laboratory performed crystallization, X-ray diffraction and electron density refinement analysis of LgtC-25 Q189E. Prof. Annette Herscovics provided purified Kre2. Dr. Yuri Lobsanov in Prof. Lynne Howell's laboratory performed crystallization, X-ray diffraction and electron density refinement analysis of Kre2. Mr. Shouming He performed ESI-MS/MS analysis of labelled LgtC-25 peptides, Kre2 peptides, and whole protein and peptides of  $\alpha$ 3GalT. UDP Gal was supplied as a generous gift from Neose Technologies Inc. (Horsham, PA).

### **7.3.2. Expression and Purification of LgtC-25**

An overnight culture of *E. coli* (DH5 $\alpha$ ) harboring pCW containing the LgtC-25 gene was diluted 100 fold in 2xYT media (containing 100  $\mu$ g/mL ampicillin) and grown to an OD<sub>600</sub> of 0.6 at which point IPTG was added to 0.5 mM and the induced culture was incubated for 6 hours at 37°C. Cells were harvested by low speed centrifugation (4000 rpm Beckman J2-21) and the resulting cell pellets were resuspended in 50 mM HEPES (pH 7.5) containing 100 mM NaCl and 1 mM benzamidine and then disrupted using a French press. Following clarification by high-speed centrifugation, NaCl (to 200 mM) and PEG-8000 (to 6% (v/v)) was added. Following incubation on ice for 1 hour, the solution was centrifuged at 10, 000 rpm for 30 minutes and the supernatant was then dialyzed overnight against 20 mM Tris (pH 7.5). The following morning the dialysate was centrifuged at 25, 000 rpm, filtered with a 0.2  $\mu$ m syringe filter, diluted 1:1 with 20 mM Tris (pH 7.5) and loaded onto a 30 mL Macro Prep High Q strong anion exchange column (Bio Rad). Protein was eluted using a gradient of 0% to 30% NaCl (w/v) in 20

mM Tris (pH 7.5) over 4 column volumes using a ProSys workstation BioSeptra FPLC. Fractions containing LgtC-25 (as determined by SDS PAGE analysis), were combined, concentrated to ~3 mL and loaded onto a HiPrep 16/60 Sephacryl S-100 High Resolution size exclusion column equilibrated with 50 mM NH<sub>4</sub>Ac (pH 7.0). Fractions containing purified LgtC-25 were combined and concentrated to ~10 mg/mL and stored at 4°C. Typical yields of purified protein were ~20 mg per litre of culture.

### **7.3.3. Expression of LgtC-25 as a C-terminal 6-Histidine Fusion**

LgtC-25 was subcloned into pET29a(+) for expression as a C-terminal 6-histidine fusion protein using *NdeI* and *XhoI* restriction sites. The LgtC-25 gene was PCR amplified using pLL01 as template and the following primers:

LgtC6HisCtermpET29NdeI: 5' - GTA GCG TGC ATA TGG ACA TCG TAT  
TTG CGG CAG ACG - 3'

LgtC6HisCtermpET29XhoI: 5' - CGG TAG GCC TCG AGG TGC GGG ACG  
GCA AGT TTG CCG CGC C - 3'

The resulting PCR product was subjected to Qiagen PCR reaction clean up kit purification according to manufacturer's instructions and then subjected to *NdeI* and *XhoI* double digestion overnight. The resulting digested PCR product and similarly digested pET29a(+) vector were purified from a 1.5% agarose gel using Qiagen's gel elute kit according to the manufacturer's instructions. Ligation of insert and vector was achieved using Roche's rapid ligation kit according to manufacturer's instructions and 2 µL of the resulting ligation mixture was then used to transform 50 µL of chemically competent Topp 10 *E. coli* (Invitrogen). Plasmid DNA from 3 of the resulting clones was subjected

to DNA sequencing, revealing the desired incorporation of the LgtC-25 gene into pET29a(+).

An overnight culture of *E. coli* (BL21 (DE3)) harboring pET29a(+) containing the LgtC-25 gene was diluted 100 fold in 2xYT media (containing 50 µg/mL kanamycin) and grown to an OD<sub>600</sub> of 0.6 at which point IPTG was added to 0.5 mM and the induced culture was incubated for 6 hours at 37°C. Cells were harvested by light centrifugation (4000 rpm Beckman J2-21) and the resulting cell pellets were resuspended in binding buffer (20 mM Tris (pH 7.5), 100 mM NaCl) containing Roche EDTA-free protease inhibitor cocktail and 1 mM benzamidine and then disrupted using a French press. Cleared supernatant was loaded onto a charged Ni<sup>2+</sup> IMAC column (5 mL GE Healthcare HisTrap Fast Flow) equilibrated with binding buffer, washed with 10 mL of binding buffer and 10 mL of binding buffer containing 25 mM imidazole. LgtC-25 was eluted using 2 column volumes each of binding buffer containing 50 mM, 100 mM, 250 mM, and 500 mM imidazole. Fractions containing purified LgtC-25 (as determined by SDS PAGE analysis) were combined, concentrated and imidazole was removed using a pre-packed desalting column (PD10 Sephadex G-25M, Pharmacia Biotech) according to manufacturer's instructions. Purified protein was stored in 20 mM Tris (pH 7.5) at 4°C. Typical yields of purified LgtC-25 were ~25 mg per litre of culture.

#### **7.3.4. A UV Spectroscopy-Based Enzyme-coupled Continuous Assay for Kinetic Characterization of LgtC and Kre2**

A continuous coupled assay similar to that described previously (Gosselin et al., 1994; Ly et al., 2002), in which the release of UDP or GDP is coupled to the oxidation of NADH ( $\lambda = 340 \text{ nm}$ ,  $\epsilon = 6.22 \text{ mM}^{-1}\text{cm}^{-1}$ ), was used to monitor the activity of all LgtC

variants and Kre2. Absorbance measurements were obtained using Cary 300 and Cary 4000 UV-VIS spectrophotometers equipped with a circulating water bath and a Peltier temperature controller, respectively. GraFit version 4.0 (Erithacus) was used to calculate kinetic parameters by direct fit of initial rates to the respective equations. The assay conditions were identical to those described in section 6.1.3. with the omission of NMPK and ATP.

### **7.3.5. ESI-MS Analysis of LgtC-25 Q189E**

#### **7.3.5.1. ESI-MS Conditions**

Mass spectra were recorded on a PE-Sciex API 300 triple quadrupole mass spectrometer (Sciex, Thornhill, Ontario, Canada) equipped with an Ion spray ion source. Whole proteins and peptides were separated by reverse-phase HPLC on a LC Packing UltiMate Micro HPLC system (Dionex, Sunnyvale, CA) that was directly interfaced with the mass spectrometer.

In each of the MS experiments whole protein was loaded onto a C4 column (LC Packings, 300Å, 1 x 150 mm) equilibrated with solvent A and then eluted with a step gradient of 60:40 A:B for 10 minutes, 40:60 A:B for 8 minutes, 30:70 for 4 minutes and 20: 80 for 4 minutes followed by 100% B for 5 minutes (solvent A: 0.05% trifluoroacetic acid/2% acetonitrile in water, solvent B: 0.045% trifluoroacetic acid/80% acetonitrile in water). Solvents were pumped at a constant flow rate of 40 µL/min. Spectra were recorded in the single quadrupole scan mode (LC/MS) with the quadrupole mass analyzer scanned over a mass to charge ratio ( $m/z$ ) range of 500 – 2500 Da with a step of 0.5 Da and dwell time of 1 ms per step. The ion source voltage (ISV) was set at 5.5 kV and the

orifice energy (OR) was set at 45 V. Protein masses were deconvoluted from multiply charged species using the Analyst 1.2 software package.

For each of the MS experiments the proteolytic digest was loaded onto a C18 column (LC Packings, 100Å pepMap, 1x 150 mm) equilibrated with solvent A. Elution of the peptides was accomplished using a gradient (0-60%) of solvent B over 60 minutes followed by 85% solvent B over 20 minutes. Solvents were pumped at a constant flow rate of 50 µL/min. A post-column splitter was present in all experiments, splitting off 85% of the sample into a fraction collector and sending 15% into the mass spectrometer. Spectra were recorded in the single-quadrupole mode (LC/MS) with the quadrupole mass analyzer scanned over a  $m/z$  range of 300-2200 Da with a step of 0.5 Da and a dwell time of 1.5 ms per step. The ISV was set at 5.5 kV and the OR set at 45 V. Native or labelled peptides were sequenced in the single quadrupole mode by increasing the OR to 65 V.

### **7.3.5.2. Labelling of LgtC-25 Q189E**

LgtC-25 Q189E (25 µL, 1 mg/mL, in 20 mM HEPES buffer pH 7.5, 5 mM MnCl<sub>2</sub>, 5 mM DTT) was incubated in the presence of 1 mM UDP Gal for 2 minutes or 1 mM UDP Glc for 2 hours to obtain optimal amounts of labelled protein. All samples were quenched by the addition of an equal volume of 6 M Urea and stored frozen prior to mass spectrometric analysis. To increase the relative proportion of labelled enzyme by 'pulling' the equilibrium over, 15 units of pyruvate kinase and 5 mM phosphoenolpyruvate were added to the above incubations. Proteolysis of LgtC (25 µL in 20 mM HEPES buffer pH 7.5, native or labelled, 1 mg/mL) was achieved by incubation with trypsin (25 µL in 20 mM HEPES buffer pH 7.5, 1 mg/mL) at room temperature until digestion was complete (30 minutes).

### **7.3.5.3. Turn-over of Covalently Labelled LgtC-25 Q189E**

Excess UDP Gal or UDP Glc was removed from LgtC-25 Q189E labelled (as determined by MS analysis) with either Gal or Glc by exchanging the labelling buffer conditions (described above) with that of 20 mM HEPES buffer, pH 7.5, 5 mM MnCl<sub>2</sub> and 5 mM DTT using an Ultrafree<sup>®</sup>-0.5 Centrifugal Filter Device (Millipore) following manufacturer's instructions to obtain a final protein concentration of 2 mg/mL. When labelled with Gal, the peak corresponding to labelled enzyme had disappeared following solvent exchange (~30 minutes). When labelled with Glc, a significant peak corresponding to labelled enzyme was observed following solvent exchange. To this solution (10 µL aliquots) an equal volume of 20 mM HEPES buffer, pH 7.5, 5 mM MnCl<sub>2</sub>, 5 mM DTT (control) or 20 mM HEPES buffer, pH 7.5, 5 mM MnCl<sub>2</sub>, 5 mM DTT, 200 mM lactose was added. Aliquots were incubated for predetermined time periods prior to being quenched by the addition of 6 M Urea and stored frozen prior to MS analysis. An observed first order rate constant for the turnover of LgtC-25 Q189E labelled with Glc was obtained from a plot of the logarithm of the relative peak height of the labelled species versus time.

### **7.3.6. Generation of the I143D, N153D, Q185D, and D190N Mutants of LgtC-25**

The I143D, N153D, Q185D, and D190N mutants of LgtC-25 were generated using the following primers and pCWlgtC-25 C128/174S plasmid DNA as template. All constructs were sequenced to verify the presence of the mutation of interest.

I143Dfwd: 5'- CAG GAA GGC TAC AAA CAA AAA GAT GGT ATG GCG  
GAC GGC GA - 3'

I143Drev: 5'- TCG CCG TCC GCC ATA CCA TCT TTT TGT TTG TAG CCT  
TCC TG - 3'

N153Dfwd: 5'- GAC GGC GAA TAT TAT TTC GAT GCC GGC GTA TTG  
CT - 3'

N153Drev: 5'- AGC AAT ACG CCG GCA TCG AAA TAA TAT TCG CCG  
TC - 3'

Q185Dfwd: 5'- GAA CAA TAC AAG GAC GTG ATG GAT TAT CAG GAT  
CAG GAC ATT TTG - 3

Q185Drev: 5'- CAA AAT GTC CTG ATC CTG ATA ATC CAT CAC GTC  
CTT GTA TTG TTC - 3'

D190Nfwd: 5'- GAT GCA ATA TCA GGA TCA GAA CAT TTT GAA CGG  
GCT GTT TA-3'

D190Nrev: 5'- TAA ACA GCC CGT TCA AAA TGT TCT GAT CCT GAT  
ATT GCA TC - 3'

### 7.3.7. Synthesis of GDP 2F Man

GDP 2F Man was synthesized using a modified version of a previously published protocol (Stick and Watts, 2002; Wittmann and Wong, 1997). The triethyl ammonium salt of 2-deoxy-2-fluoro- $\alpha$ -D-mannose-1-phosphate (~100 mg) was dissolved in 1 mL of water and loaded onto a 10 mL AG 50Wx2 (Bio Rad) cation exchange column equilibrated in the pyridinium form and then thoroughly eluted using water. The eluant was lyophilized to afford the desired pyridinium salt. To this pyridinium salt (30 mg, 72  $\mu$ mol), tri-n-octylamine (63  $\mu$ L, 144  $\mu$ mol) and 5 mL of dry pyridine were added and the mixture was evaporated and then co-evaporated 3 times with 5 mL of dry pyridine. GMP morpholidate (35 mg, 48  $\mu$ mol) was then added and this mixture was again coevaporated

3 times with 5 mL of dry pyridine. Tetrazole (10 mg, 144  $\mu\text{mol}$ ) and dry pyridine (5 mL) were added and the reaction left at room temperature under dry nitrogen for 10 days. Pyridine was evaporated and the resulting residue was resuspended in 100 mM  $\text{NH}_4\text{HCO}_3$ , and washed with diethyl ether. The washed aqueous solution was lyophilized, dissolved in 5 mL of water, filtered with a 0.2  $\mu\text{m}$  syringe filter and loaded onto a preparative HPLC TSKgel Amide-80 (Tosoh Bioscience) column. The desired product was purified using a flow rate of 6 mL/minute and a gradient of acetonitrile (A): water (B) (80:20 A:B hold for 20 minutes, linear gradient from 80:20 A:B to 65:35 A:B over 40 minutes, hold 65:35 A:B for 15 minutes). The 6 mL fractions containing pure GDP 2F Man (as determined by negative ion mode ESI-MS) were combined, solvent removed under pressure, and the resulting aqueous solution was lyophilized to yield a white fluffy solid.  $^1\text{H}$ ,  $^{31}\text{P}$ , and  $^{19}\text{F}$  NMR spectra were consistent with those reported previously (Wong, 1998). 12.9 mg (42%); select data,  $^{19}\text{F}$ -NMR (282 MHz, proton decoupled)  $\delta_{\text{F}}$  ( $\text{D}_2\text{O}$ ): -128.2;  $^{31}\text{P}$ -NMR (162 MHz, proton decoupled)  $\delta_{\text{P}}$  ( $\text{D}_2\text{O}$ ): -11.1 (m, 1 P), -13.6 (m, 1 P); ESI-MS:  $m/z = 606.2$  [ $\text{M}$ ] $^-$  (expected for  $\text{C}_{16}\text{H}_{23}\text{FN}_5\text{O}_{15}\text{P}_2^-$ :  $m/z = 606.1$ ).

### 7.3.8. ESI-MS Analysis of Kre2

Kre2 (25  $\mu\text{L}$ , 1 mg/mL, in 20 mM HEPES buffer pH 7.5, 5mM  $\text{MnCl}_2$ ) was incubated in the presence of 1 mM GDP Man or 1 mM GDP 2F Man or 1 mM GDP Man and 100 mM methyl  $\alpha$ -D-mannoside or 1 mM GDP 2F Man and 100 mM methyl  $\alpha$ -D-mannoside for time periods ranging from 2 minutes to 2 hours. In an attempt to increase the relative proportion of labelled enzyme by 'pulling' the equilibrium over, 15 units of pyruvate kinase and 5 mM phosphoenolpyruvate were also added to the above incubation

conditions. Proteolysis of Kre2 (25  $\mu$ L in 20 mM HEPES buffer pH 7.5, native or putatively labelled, 1 mg/mL) was achieved by incubation with trypsin (25  $\mu$ L in 20 mM HEPES buffer pH 7.5, 1 mg/mL) or pepsin (25  $\mu$ L in 200 mM phosphate buffer pH 2.0, 1 mg/mL) at room temperature until digestion was complete (30 minutes). ESI-MS of peptide fragments was performed as described in section 6.2.5.1..

### **7.3.9. Expression and Purification of $\alpha$ 3GalT as a Male Fusion Protein**

The pCW-Bov10 plasmid was transformed into electrocompetent *E. coli* (DH5 $\alpha$ ). Protein overexpression was achieved by standard IPTG (0.5 mM) induction with overnight expression at 20°C in 2xYT media (containing 100  $\mu$ g/mL ampicillin). Cells were collected by light centrifugation (4000 rpm Beckman J2-21), resuspended in binding buffer (20 mM TRIS (pH 7.5), 200 mM NaCl) containing Roche EDTA-free protease inhibitor cocktail and disrupted with a French Press. Protein purification was achieved by loading clarified supernatant onto a 15 mL Amylose Resin Fast Flow column (New England Biolabs) equilibrated with binding buffer at a flow rate of 0.8 mL/min. Following loading, the column was washed with 12 column volumes of binding buffer and then eluted with 4 column volumes of binding buffer containing 10 mM maltose. During elution, 3 mL fractions were collected and those containing pure  $\alpha$ 3GalT (as determined by SDS PAGE analysis) were combined, concentrated and maltose was removed using a pre-packed desalting column (PD10 Sephadex G-25M, Pharmacia Biotech) according to manufacturer's instructions. Purified protein was stored as a 5 mg/mL solution in 20 mM Tris (pH 7.5). Typical yields of purified protein were ~15 mg per litre of culture.

### **7.3.10. ESI-MS Analysis of $\alpha$ 3GalT**

$\alpha$ 3GalT (25  $\mu$ L, 1 mg/mL, in 20 mM HEPES buffer pH 7.5, 5 mM  $MnCl_2$ ) was incubated in the presence of 1 mM UDP Gal or 1 mM UDP Gal and 100 mM lactose for time periods ranging from 2 minutes to 2 hours. All samples were quenched by the addition of an equal volume of 6 M Urea or 3 M phosphate (pH 2.0) and stored frozen prior to mass spectrometric analysis. In an attempt to increase the relative proportion of labelled enzyme by 'pulling' the equilibrium over, 15 units of pyruvate kinase and 5 mM phosphoenolpyruvate were also added to the above incubation conditions. Proteolysis of  $\alpha$ 3GalT (25  $\mu$ L in 20 mM HEPES buffer pH 7.5, native or putatively labelled, 1 mg/mL) was achieved by incubation with trypsin (25  $\mu$ L in 20 mM HEPES buffer pH 7.5, 1 mg/mL) or pepsin (25  $\mu$ L in 200 mM phosphate buffer pH 2.0, 1 mg/mL) at room temperature until digestion was complete (30 minutes). ESI-MS analysis of whole protein and peptide fragments was performed as described in section 6.2.5.1..

## **7.4. Chapter 4 Experimental**

### **7.4.1. Acknowledgements and Generous Gifts**

Dr. Sachdev Sidhu from the Department of Protein Engineering at Genentech in San Francisco provided the pSAV phagemid for optimized phage display of SAV. Dr. Spencer Williams provided 3-[N-(D-biotinyl)-13-amino-4,7,10-trioxatridecanylaminocarbonyl] propanoic acid for the synthesis of a lactose-biotin conjugate. Dr. Warren W. Wakarchuk provided pSJF6 phagemid as the starting vector for 3 + 3 phage display of LgtC and spGalU enzyme.

## 7.4.2. Synthesis of CDP Gal

Cytosine 5'-monophosphomorpholidate 4-morpholine-N,N'-dicyclohexylcarboxamide salt was prepared essentially as described (Moffatt, 1966, 1961). Briefly, a solution of DCC (3.8g, 18.6 mmol) in t-butyl alcohol (45 mL) was added dropwise to a refluxing solution of CMP (1.0 g, 3.1 mmol) and morpholine (1.6 mL, 18.6 mmol) in t-butyl alcohol and water (30 mL each) over a period of 3 hours. The combined mixture was then refluxed for an additional 2 hours until all of the CMP was consumed. The mixture was then cooled to room temperature, precipitate was removed by filtration, and t-butyl alcohol was removed under reduced pressure from the combined filtrates. The remaining aqueous solution was washed with diethyl ether and lyophilized to yield an off white gum. The gum was triturated twice using diethyl ether to afford a white solid. 1.3 g (61% recovered). The  $^1\text{H}$  and  $^{31}\text{P}$  NMR spectra were consistent with those previously published (Moffatt, 1966).

CDP Gal was prepared by reacting the above CMP morpholidate reagent (240 mg, 0.35 mmol) with the pyridinium salt of  $\alpha$ -D-galactose-1-phosphate (200 mg, 0.53 mmol) as described for the synthesis of GDP 2F Man in section 6.3.7.. Preparative scale HPLC purification using an Amide-80 column as described for the purification of GDP 2F Man afforded the desired CDP Gal product as a white solid.

88 mg (43%);  $^1\text{H}$ -NMR (400 MHz):  $\delta_{\text{H}}$  ( $\text{D}_2\text{O}$ ): 7.98 (d, 1 H,  $J_{5,6}$  7.24 Hz, H-6), 6.10 (d, 1 H,  $J_{5,6}$  7.24 Hz, H-5), 5.82 (d, 1 H,  $J_{1',2'}$  3.6 Hz, H-1'), 5.50 (dd, 1 H,  $J_{1'',\text{P}}$  7.2 Hz,  $J_{1'',2''}$  3.6 Hz, H-1''), 4.27-4.00 (m, 6 H, H-2', H-3', H-4', H-5<sub>a</sub>', H-5<sub>b</sub>', H-5''), 3.88 (m, 1 H, H-4''), 3.77 (dd, 1 H,  $J_{2'',3''}$  10.2 Hz,  $J_{3'',4''}$  3.0 Hz, H-3''), 3.69-3.54 (m, 3 H, H-2'', H-6<sub>a</sub>''),

H6<sub>b</sub>"); <sup>31</sup>P-NMR (162 MHz) δ<sub>P</sub> (D<sub>2</sub>O); -10.2 (d, 1 P, J<sub>PαPβ</sub> 21.0 Hz), -11.7 (d, 1 P, J<sub>PαPβ</sub> 21.0 Hz); ESI-MS: m/z = 564.3 [M]<sup>-</sup> (expected for C<sub>15</sub>H<sub>24</sub>N<sub>3</sub>O<sub>16</sub>P<sub>2</sub><sup>-</sup>: m/z = 563.1).

### 7.4.3. Generating the N10D Mutant of LgtC-25

The N10D mutant of LgtC-25 was generated by the method of QuikChange<sup>TM</sup> (Stratagene) using the following primers and pCWlgtC-25 C128/174S plasmid DNA as template. The resulting construct was sequenced to verify the presence of the mutation of interest.

LgtCN10Dfwd: 5' – CGG CAG ACG ACG ACT ATG CCG CCT ATC – 3'

LgtCN10Drev: 5' - GAT AGG CGG CAT AGT CGT CGT CTG CCG – 3'

### 7.4.4. Kinetic Analysis of CDP Gal as a Substrate for LgtC-25 and LgtC N10D

Kinetic analysis of wild-type and N10D LgtC was performed as described in section 6.2.4 using variable concentrations of CDP Gal or UDP Gal as donor substrate at a fixed saturating concentration (160 mM) of lactose acceptor substrate. For the wild-type enzyme, the concentration of UDP Gal was varied from 5 to 125 μM at an enzyme concentration of 0.20 μg/mL and the concentration of CDP Gal was varied from 25 to 500 μM at an enzyme concentration of 0.43 μg/mL. For the N10D mutant, the concentration of UDP Gal was varied from 25 to 500 μM at an enzyme concentration of 1.0 μg/mL and the concentration of CDP Gal was varied from 25 to 1000 μM at an enzyme concentration of 4.0 μg/mL.

#### 7.4.5. Subcloning LgtC-25 for Expression as an M13 phage pIII Fusion Protein

LgtC-25 was subcloned into pSJF6 phagemid for expression as a C-terminal M13 phage pIII fusion protein containing an N-terminal OmpA leader sequence using *EcoRI* and *BglIII* restriction sites. The LgtC-25 gene was PCR amplified using pCWLgtC-25 C128/174S plasmid DNA as template and the following primers:

LgtC-phage-5p: 5' – GGG GGG GAA TTC ATG AAA AAA ACC GCT ATC  
GCG ATC GCA GTT GCA CTG GCT GGT TTC GCT  
ACC GTT GCG CAA GCC ATG GAC ATC GTA TTT  
GCG GCA GAC G – 3'

LgtC-phage-3p: 5' – GGG GGG AGA TCT GTG CGG GAC GGC AAG TTT  
GCC GCG CC – 3'

The resulting PCR product was subjected to Qiagen's PCR reaction clean up kit according to manufacturer's instructions and then subjected to a nested PCR reaction to increase the overall yield of DNA product using LgtC-phage-3p and the following primer:

LgtCPh5pSh: 5'- GGG GGG GAA TTC ATG AAA AAA ACC G – 3'

The resulting PCR product was subjected to the Qiagen PCR reaction clean up kit according to manufacturer's instructions and then subjected to *EcoRI* and *BglIII* double digestion overnight. The resulting digested PCR product and similarly digested pSJF6 vector were purified from a 1.5% agarose gel using Qiagen's gel elute kit according to manufacturer's instructions. Ligation of insert and vector was achieved using Roche's rapid ligation kit according to manufacturer's instructions and 2  $\mu$ L of the resulting ligation mixture was then used to transform 50  $\mu$ L of chemically competent Topp 10 *E. coli* (Invitrogen). Colony PCR using commercial M13fwd and M13rev primers and DNA from 10 isolated transformants as template revealed one potential candidate containing an insert of the appropriate size. However, DNA sequencing of phagemid DNA isolated

from this clone revealed the presence of a 7 base pair deletion from the OmpA leader sequence region. These 7 base pairs were reintroduced by PCR amplifying the entire phagemid (containing the insert with the 7 base pair deletion) using Ultra high fidelity *pfu* DNA polymerase (Stratagene) and the following primers that contain *AarI* restriction sites and the missing 7 base pairs:

OmpA7infwd: 5'- GGA GGT CGC ACC TGC ATT GTCGCA GTT GCA CTG  
GCT GGT TTC GCT ACC GTT GCG C - 3'

OmpA7inrev: 5'- GGA GT CGC ACC TGC TAC TGC GAT CGC GAT AGC  
GGT TTT TTT CAT GAA TTC TAC C - 3'

Following PCR cycling the reaction was subjected to *DpnI* digestion to degrade all template DNA (containing the 7 base pair deletion) and then subjected to Qiagen's PCR reaction clean up kit according to manufacturer's instructions. Overnight *AarI* digestion was followed by purification from a 1.5% agarose gel using Qiagen's gene elute kit. The resulting linear *AarI* digested phagemid was ligated using Roche's rapid ligation kit and transformed into chemically competent one shot Topp10 *E. coli* (Invitrogen). Extensive DNA sequencing of phagemid DNA isolated from one of the transformants revealed the presence of the desired complete LgtC gene (fused via its C-terminus to the pIII and containing an N-terminal OmpA leader sequence) within the MCS. The resulting phagemid was designated pLL01.

#### **7.4.6. VCSM13 Helper Phage Production**

A modified version of Cold Spring Harbor protocol 10.2 was used (Barbas III, 2001). A single isolated colony of *E. coli* (XL1 Blue F') grown on LB agar plates (containing 10 µg/mL tetracycline) was used to inoculate 10 mL of SB media (containing 10 µg/mL tetracycline). This culture was incubated at 37°C (with shaking) until an OD<sub>600</sub>

of 0.3 was achieved at which point  $3 \times 10^{10}$  VCSM13 helper phage were added. Following an infection period of 30 minutes, during which the culture was incubated at 37°C without shaking to allow regrowth of sheared pili, kanamycin was added to a final concentration of 25 µg/mL and the culture was left to incubate overnight at 37°C (with shaking). The following morning, *E. coli* cells were heat killed at 65°C for 15 minutes and the culture was spun down twice at 4000 rpm in a tabletop centrifuge to remove bacterial debris. The resulting supernatant was aliquoted into 1 mL volumes and stored at 4°C. Titters of  $\sim 10^{11}$  pfu/mL were typical.

#### **7.4.7. Small Scale PEG Precipitated LgtC-25-Displaying Phage Preparation**

A modified version of Cold Spring Harbor protocol 11.2 was used (Barbas III, 2001). A single isolated colony of *E. coli* (XL1 Blue F') harboring the pLL01 phagemid was used to inoculate 4 mL of SB media (containing 100 µg/mL ampicillin and 10 µg/mL tetracycline). This culture was incubated for 11 hours at 37°C (with shaking), at which point 50 µL of helper phage solution was added and the culture was incubated at 37°C without shaking for 30 minutes and then for 90 minutes at 37°C with shaking. Following this infection period, kanamycin was added to a final concentration of 50 µg/mL and the culture was incubated overnight at 37°C with shaking. The following morning, the culture was spun down at 3500 rpm in a tabletop centrifuge for 15 minutes and to 1.2 mL of supernatant, 0.3 mL of a sterile 5x concentrated PEG/NaCl solution (200 g PEG-8000, 150 g NaCl per liter) was added and the mixture left to incubate on ice for 30 minutes. The solution was then spun down at 14,000 rpm at 4°C for 10 minutes, the supernatant was removed and the resulting white pellet was spun down again at 14,000 rpm at 4°C for

1 minute. This pellet was then resuspended in 1.5 mL of 1x PEG/NaCl solution and the described precipitation procedure repeated. The resulting pellet was then resuspended in 50  $\mu$ L of 50 mM HEPES (pH 7.5).

#### **7.4.8. Quantification of Phage by UV Spectroscopy**

A modified version of Cold Spring Harbor protocol 15.6 was used (Barbas III, 2001). Phage concentrations were determined by measuring absorbance of solutions at 269 nm according to nucleotide content of phage using an averaged extinction coefficient of  $1.006 \times 10^4 \text{ M}^{-1}\text{cm}^{-1}$  according to the following equation:

$$\text{Phage particles per mL} = \frac{(\text{adjusted } A_{269}) \times (6 \times 10^{16})}{\text{number of nucleotides in the phage genome}}$$

$$\text{where adjusted } A_{269} = (A_{269}) - (A_{269\text{-background}}) - (A_{320})$$

#### **7.4.9. Western Blot Analysis of pLL01 and pSJF6 Derived Phage**

Solutions saturated with PEG-precipitated phage were diluted 2x with SDS loading solution (containing  $\beta$ -mercaptoethanol) and boiled for 10 minutes prior to being directly loaded onto each of two identically loaded SDS PAGE Gels and proteins were then resolved by standard electrophoresis. One of the resulting gels was subjected to Coomassie blue analysis and the other to Western blot analysis using a Bio-Rad semi-dry transfer apparatus according to the manufacturer's instructions. Detection of c-myc containing proteins on the resulting blot was achieved by an immunoblotting procedure using Amersham's Anti-myc-HRP antibody according to the manufacturer's instructions.

Optimal detection was achieved by a 1-minute incubation with HRP substrate and a 40 minute exposure.

#### **7.4.10. Compartmentalized Phage Display Procedure**

*E. coli* (XL1 Blue F') cells ( $\sim 10^8$ ) harboring pLL01 were emulsified from an aqueous solution into  $\sim 10^9$  aqueous  $\sim 5$  fL compartments. Washed cells derived from overnight culture were resuspended in 50 mM HEPES (pH 7.5) containing 5 mM  $\text{MnCl}_2$  and 100  $\mu\text{g}/\text{mL}$  ampicillin and immediately prior to the addition of ten volumes of an ice cold mixture of Abilem90<sup>TM</sup> in light mineral oil (2.9% w/w), a 10 fold concentrate of LgtC reaction components (50 mM DTT, 50 mM UDP Gal, 5 mM fluorescein lactose conjugate) and  $2 \times 10^{10}$  VCSM13 helper phage were added to the resuspended cells. Emulsification on ice for 5 minutes at 9500 rpm using an IKA T-25<sup>TM</sup> homogenizer resulted in the production of a milky white suspension that was incubated at 37°C overnight. The following day, the mixture was centrifuged (14,000 rpm) and the oil phase was removed. The aqueous layer was diluted 5x with 50 mM HEPES (pH 7.5) and washed twice with water-saturated diethyl ether. Remaining ether was immediately removed from the aqueous phase using a Speed Vac and the resulting phage-containing aqueous solution was subjected to PEG precipitation as described in section 6.4.7.

#### **7.4.11. Testing Infectivity of Emulsified LgtC-25-Displaying Phage**

A modified version of Cold Spring Harbor protocol 15.2 was used (Barbas III, 2001). Starved *E. coli* (XL1 Blue F') cells were generated according to a version of the Cold Spring Harbour protocol 15.1 (Barbas III, 2001). Starved cells were infected with

PEG-precipitated phage at a ratio of 50:1 (cells : phage) by incubation for 15 minutes at room temperature (without shaking). Antibiotic resistance was induced by addition of 1 mL of LB media containing a low concentration of ampicillin (2 µg/mL). Following 1 hour incubation at 37°C (with light shaking of ~100 rpm), 200 µL of 10x, 100x, and 1000x fold dilutions were plated on LB agar plates (containing 100 µg/mL). Following overnight incubation at 37°C the number of resulting colonies was used to determine the infectivity of LgtC-displaying (or control) phages.

#### **7.4.12. Synthesis of a Biotin Lactose Conjugate**

3-[N-(D-Biotinyl)-13-amino-4,7,10-trioxatridecanylaminocarbonyl]propanoic acid was activated as the pentafluorophenyl ester as described previously (Williams et al., 2006). Briefly, to a solution of the acid (50 mg, 92 µmol) in dry DMF (2 mL), Et<sub>3</sub>N (20 µL, 140 µmol) and pentafluorophenyl trifluoroacetate (24 µL, 140 µmol) were added and the reaction left stirring at room temperature for 5 hours under dry nitrogen gas. The solvent was evaporated and the resulting residue was triturated with diethyl ether twice to afford the pentafluorophenyl ester (53.8 mg, 82% recovery). The white residue was dried under vacuum before direct use in the following coupling reaction.

The crude pentafluorophenyl ester (50 mg, 70 µmol) and *p*-aminophenyl β-D-lactopyranoside (20 mg, 46 µmol)(Toronto Research Chemicals, Toronto) were dissolved in a dry mixture of DMF and pyridine (5 and 30 mL respectively) and left stirring overnight at room temperature. Solvent was removed under reduced pressure and the resulting residue was lyophilized from water, dissolved in 2 mL of water, filtered with a 2 µm syringe filter and then subjected to preparative HPLC purification using an Amide-80 column as described for the purification of GDP 2F Man.

17.1 mg (38%);  $^1\text{H-NMR}$  (400 MHz):  $\delta_{\text{H}}$  ( $\text{D}_2\text{O}$ ): 7.45-7.38 (m, 2 H, Ar), 7.19-7.12 (m, 2 H, Ar), 5.14 (d, 1 H,  $J_{1,2}$  8.0 Hz, H-1), 4.61-4.57 (m, 1 H), 4.50 (d, 1 H,  $J_{1',2'}$  7.6 Hz, H-1'), 4.41-4.37 (m, 1 H), 4.03-3.50 (m, 30 H), 3.35-3.21 (m, 6 H), 2.98 (dd, 1 H,  $J$  5.2 Hz,  $J$  13.2 Hz), 2.80-2.59 (m, 5 H), 2.24 (m, 2 H), 1.85-1.51 (m, 8 H), 1.45-1.32 (m, 2 H);  $^{13}\text{C-NMR}$  (100 MHz)  $\delta_{\text{C}}$  ( $\text{D}_2\text{O}$ ): 176.8, 174.8, 173.5, 165.5, 154.1, 148.2, 139.8, 132.2, 123.7, 117.3, 103.1, 100.4, 78.2, 75.6, 75.1, 74.4, 72.8, 72.7, 71.2, 69.8 (x2), 69.7, 69.6, 68.8, 68.6, 68.5, 62.3, 61.2, 60.4, 60.1, 55.6, 39.9, 36.5, 35.7, 32.1, 31.5, 31.4, 28.5 (x2), 28.1, 27.9, 25.4; HR ESI-MS:  $m/z$  = 984.4104 [ $\text{M} + \text{Na}$ ] $^+$  (expected for  $\text{C}_{42}\text{H}_{67}\text{N}_5\text{O}_{18}\text{SNa}^+$ :  $m/z$  = 984.4100).

#### 7.4.13. Generating a Library of LgtC-25 Variants by Error Prone PCR

Libraries of LgtC variants were generated using established protocols (Vartanian et al., 2001). Briefly, 2-10 ng of pCW-LgtC-25C138/174S plasmid DNA was used as template to PCR amplify the LgtC-25 gene using *Taq* polymerase and reaction conditions containing dNTP ratios of 1:5 AC:TG supplemented with 0.25 or 0.5 mM  $\text{MnCl}_2$ . The following primers, used to amplify the LgtC-25 gene in these reactions, contained *EcoRI* and *SalI* restrictions site such that the resulting PCR product could be subcloned into pUC18 vector.

LgtCfwdEcoRI: 5'- GTA GCG TGG AAT TCA TGG ACA TCG TAT TTG  
CGG CAG ACG - 3'

LgtCrevSalI: 5'- CGG TAG GCG TCG ACG TGC GGG ACG GCA AGT TTG  
CCG CGC C - 3'

The PCR program consisted of 30 cycles (95°C for 0.5 minutes, 53°C for 0.5 minutes, 72°C for 2 minutes). The PCR products were subjected to PCR reaction clean up, double

digested with *EcoRI* and *SalI* overnight, and purified on a 1.5% agarose gel. The gel-purified double-digested PCR product was ligated into similarly digested pUC18 vector and then transformed into *E. coli*<sup>TM</sup> ultra competent *E. coli* cells using the UltraClone DNA ligation and transformation kit (Lucigen) according to manufacturer's instructions. Transformation efficiency, which dictates library size, was assessed by plating serially diluted transformed cells. Using the UltraClone kit and pUC18 as vector, libraries of >10<sup>7</sup> transformants were achieved. Transformed cells were grown overnight in LB (containing 100 µg/mL ampicillin) and the library plasmid DNA was isolated. DNA sequencing of plasmid DNA isolated from random colonies revealed that a desired mutation rate of ~2 amino acid mutations per gene was achieved using 2 ng of template and 0.25 mM MnCl<sub>2</sub>.

#### **7.4.14. Subcloning Verotoxin for Expression as a C-Terminal 6-Histidine Fusion Protein**

Verotoxin was subcloned from pVT21 into pET29a(+) for expression as a C-terminal 6-histidine fusion protein using *NdeI* and *XhoI* restriction sites. The verotoxin gene was PCR amplified using pVT21 as template and the following primers:

VT6HCpETFW: 5'- GTA GCG TGC ATA TGA CGC CTG ATTGTG TAA  
CTG GAA AGG TG-3'

VT6HCpETRV: 5'- CGG TAG GCC TCG AGA CCG GAG CCA GAA CGA  
AAA ATA ACT TCA CTG AAT CCC CCT CC - 3'

The resulting PCR product was subjected to the Qiagen PCR reaction clean up kit according to manufacturer's instructions and then subjected to *NdeI* and *XhoI* double digestion overnight. The digested PCR product and similarly digested pET29a(+) vector were purified from a 1.5% agarose gel. Ligation of insert and vector was achieved using

Roche's rapid ligation kit according to manufacturer's instructions and 2  $\mu$ L of the resulting ligation mixture was then used to transform 50  $\mu$ L of chemically competent Topp 10 *E. coli* (Invitrogen). Plasmid DNA from 3 of the resulting clones was subjected to DNA sequencing, revealing the desired incorporation of the verotoxin gene into pET29a(+).

An overnight culture of *E. coli* (BL21 (DE3)) harboring pET29a(+) containing the verotoxin gene was diluted 100 fold in LB media (containing 50  $\mu$ g/mL kanamycin) and grown to an OD<sub>600</sub> of 0.6 at which point IPTG was added to 0.5 mM and the induced culture was grown overnight at 25°C. Cells were harvested by light centrifugation and the resulting cell pellets were resuspended in binding buffer (20 mM Tris (pH 7.5), 100 mM NaCl) containing Roche EDTA-free protease inhibitor cocktail and then disrupted using a French press. Cleared supernatant was loaded onto a charged Ni<sup>2+</sup> IMAC column (5 mL GE Healthcare HisTrap Fast Flow) equilibrated with binding buffer and then washed with 10 mL of binding buffer and 10 mL of binding buffer containing 50 mM imidazole. Verotoxin was eluted using 2 column volumes each of binding buffer containing 100 mM, 250 mM, and 500 mM imidazole. Fractions containing purified verotoxin (as determined by SDS PAGE analysis) were combined, concentrated and imidazole was removed using a pre-packed desalting column (PD10 Sephadex G-25M, Pharmacia Biotech) according to manufacturer's instructions. Typical yields of purified verotoxin were in excess of 100 mg/L.

#### 7.4.15. Subcloning of SAV-pVIII Fusion Protein

The gene encoding the SAV-pVIII fusion optimized for type 8 + 8 phage display (Sidhu, 2000) was subcloned from pSAV into pLL01 for polycistronic expression using a single *Hind*III restriction site. The SAV-pVIII gene was PCR amplified using pSAV as template and the following primers:

SASFwdHind: 5' – TTC GAT AAA AGC TTG GAG GTT ATA TGA AAA  
TAA AAA CAG GTG CAC GCA TCC TCG C – 3'

SASRevHind: 5' – TTC GAT AAA AGC TTA ATA TCA GCT TGC TTT CGA  
GGT GAA TTT CTT AAA CAG C – 3'

The forward primer was designed to include a ribosome binding site downstream of the *Hind*III restriction site prior to the SAV start codon. The resulting PCR product was subjected to Qiagen PCR reaction clean up kit according to manufacturer's instructions and then subjected to *Hind*III digestion overnight. The resulting digested PCR product and similarly digested pLL01 vector were purified from a 1.5% agarose gel using Qiagen's gel elute kit according to manufacturer's instructions. Ligation of insert and vector was achieved using Roche's rapid ligation kit according to manufacturer's instructions and 2  $\mu$ L of the resulting ligation mixture was then used to transform 50  $\mu$ L of chemically competent Topp 10 *E. coli* (Invitrogen). Colony PCR using template derived from 10 of the resulting colonies was performed using commercial M13fwd primer and a custom internal LgtC sequencing primer to determine candidates containing a single insert with the correct orientation. Plasmid DNA from 3 such positive candidates was then subjected to DNA sequencing analysis to confirm incorporation of the SAV gene in the correct orientation into pLL01. Extensive DNA sequencing was then performed on one of the positive clones to ensure that the entire desired sequence within

the MCS was present. This phagemid containing the desired insert was designated pLL34.

#### **7.4.16. ELISA Analysis of pSAV and pLL34 Derived Phage**

PEG-precipitated phage derived from either pLL01, pSAV or pLL34 phagemid were subjected to ELISA analysis to determine the level of SAV display according to a modified version of a previously published procedure (Sidhu et al., 2000). Immobilized biotin was used as the capture target and an Amersham detection module for recombinant phage Ab system was used for detection.

Immulon 2HB (Nunc) 96-well plates were incubated with 200  $\mu$ L of BSA-biotin conjugate (10  $\mu$ g/mL in 50 mM  $\text{Na}_2\text{CO}_3$  pH 9.6) for 2 hours at room temperature in a humidified container. Wells were then emptied, blotted on paper towel and blocked by treatment with 300  $\mu$ L of blocking buffer (0.2% BSA in PBS) for 1 hour. Wells were then emptied, blotted on paper towel and washed 8 times with PBS containing 0.05 % Tween-20. In pre-blocked plates, phage were serially diluted using an equal volume of blocking buffer and incubated for 30 minutes at room temperature in a humidified container. To the wells of another pre-blocked 96-well plate, 200  $\mu$ L of the diluted phage solutions were added and left to incubate for 2 hours at room temperature in a humidified container. Following incubation, wells were emptied, blotted on paper towel and washed 8 times with PBS containing 0.05% Tween-20. Detection of immobilized phage was then achieved by treatment of wells with 200  $\mu$ L of HRP-anti-M13 Ab conjugate (diluted 1:1000 from the commercial stock in blocking buffer). Following incubation for 1 hour at room temperature in a humidified container, the wells were emptied, blotted on paper towel and washed 8 times with PBS containing 0.05% Tween-20 and then twice with

PBS. Following washing, 200  $\mu$ L of ABTS (0.2 mg/mL) / H<sub>2</sub>O<sub>2</sub> (0.02 % v/v) solution was added and left to incubate for 40 minutes, at which point immobilized phage were quantified by measuring the absorbance at 410 nm.

## **7.5. Chapter 5 Experimental**

### **7.5.1. Acknowledgements and Generous Gifts**

Dr. Andrew Watts provided 6OBz Man and 4OBz Xyl. Dr. Warren Wakarchuk provided FCHASE-Lactose. Dr. David Zechel provided dNP  $\alpha$ -D-galactoside and dNP  $\beta$ -D-galactoside. Dr. Karena Thieme provided pNP  $\alpha$ -D-sialoside. Benzyl  $\beta$ -D-cellobioside and benzyl  $\beta$ -D-xyloside were obtained from the Withers laboratory collection of custom sugars.

### **7.5.2. LgtC-Catalysed Galactosylation of Modified Alternative Acceptor Substrates**

Acceptor substrate (5 mM), UDP galactose (7.5 mM) and LgtC (~0.5 mg/mL) were incubated for ~16 hours at room temperature in 20 mM HEPES buffer (pH 7.5) containing 50 mM NaCl, 5 mM MnCl<sub>2</sub> and 5 mM dithiothreitol. Upon completion, reactions were partially purified with SepPak C<sub>18</sub>Plus Cartridges (0.39 g, Waters) by washing with HPLC grade water (10 mL) and eluting with HPLC grade MeOH (10 mL). Lyophilized products were then subjected to standard per-*O*-acetylation with pyridine/Ac<sub>2</sub>O followed by a final purification by column chromatography (10:1 to 3:1 toluene/EtOAc). All fractions containing disaccharide products were combined and analyzed by <sup>1</sup>H NMR. In contrast to the cases where underivatized monosaccharides were used as acceptors and mixtures of regio-isomers were observed, when derivatized

acceptors were used, reactions were found to be completely regio-selective beyond the limit of  $^1\text{H}$  NMR detection. Of potential utility, in the cases where pNP Glc, octyl Glc and octyl Gal were used as acceptors, a trisaccharide product was detected (confirmed by ESI MS) accounting for the lack of near quantitative yields.

### 7.5.3. NMR Analysis of LgtC Products

Specific assignments of individual  $^1\text{H}$  peaks were unambiguously determined using two-dimensional  $^1\text{H}$  COSY NMR. Assignments were made by examining the couplings within the ring system and the locations of glycosidic linkages were determined on the basis of chemical shift. Because the acetate group is more electron withdrawing than a glycoside substituent, the proton attached to the carbon involved in the glycosidic bond resonates further upfield ( $<4.50$  ppm) in a region clearly distinct from anomeric protons or protons attached to carbons bearing an acetate group.  $J_{1,2}$  coupling constants were used to assign the configuration of the anomeric linkage ( $\alpha$  or  $\beta$ ).

#### ***para*-Nitrophenyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 4)-2,3,6-tri-O-acetyl- $\beta$ -D-glucoside:**

3.6 mg (51%);  $^1\text{H}$ -NMR (400 MHz):  $\delta_{\text{H}}$  ( $\text{CDCl}_3$ ): 8.23 (m, 2 H, Ar), 7.08 (m, 2 H, Ar), 5.49 (d, 1 H,  $J_{1,2}$  4.0 Hz, H-1'), 5.46 (d, 1 H,  $J_{4',3'}$  2.5 Hz, H-4'), 5.35 (t, 1 H,  $J_{3,2} = J_{3,4}$  8.8 Hz, H-3), 5.29-5.24 (m, 2 H, H-1, H-3'), 5.18-5.12 (m, 2 H, H-2, H-2'), 4.50 (dd, 1 H,  $J_{6,5}$  2.7 Hz,  $J_{6,6}$  12.1 Hz, H-6), 4.30-4.19 (m, 2 H, H-5', H-6), 4.14-4.07 (m, 3 H, H-4, H-6', H-6'), 3.97-3.90 (m, 1 H, H-5), 2.15 (s, 3 H, Ac), 2.09 (s, 3 H, Ac), 2.08-2.06 (4xs, 12 H, 4xAc), 2.00 (s, 3 H, Ac);  $^{13}\text{C}$ -NMR (100 MHz)  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ): 171.1, 170.6, 170.4, 170.2, 170.1, 169.8, 161.3, 143.5, 128.9, 126.0 (x2), 116.8 (x2), 97.8, 96.5, 75.3, 73.0,

72.7, 71.7, 67.9, 67.7, 67.4, 67.0, 63.1, 61.7, 21.1-20.8 (x7); HR ESI-MS:  $m/z = 780.1987 [M + Na]^+$  (expected for  $C_{32}H_{39}NO_{20}Na^+$ :  $m/z = 780.1963$ ).

**Benzyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 3)-2,4-di-O-acetyl- $\beta$ -D-xyloside:**

48 mg (86%);  $^1H$ -NMR (400 MHz):  $\delta_H$  ( $CDCl_3$ ): 7.40-7.13 (m, 5 H, Ar), 5.39 (d, 1 H,  $J_{1,2}$  3.8 Hz, H-1'), 5.29 (d, 1 H,  $J_{4,3}$  2.3 Hz, H-4'), 5.23 (dd, 1 H,  $J_{3,2}$  10.9 Hz,  $J_{3,4}$  3.3 Hz, H-3'), 5.07-4.98 (m, 2 H, H-2', H-2), 4.87-4.81 (m, 2 H, H-4, BnCH), 4.60 (d, 1 H,  $J_{1,2}$  4.7 Hz, H-1), 4.52 (d, 1 H,  $J$  11.9 Hz, BnCH), 4.34-4.28 (m, 1 H, H-5'), 4.19 (dd, 1 H,  $J_{5,4}$  4.0 Hz,  $J_{5,5}$  12.2 Hz, H-5), 4.04-3.92 (m, 2 H, H-6', H-6'), 3.86 (t, 1 H,  $J_{3,2} = J_{3,4}$  6.3 Hz, H-3), 3.42 (dd, 1 H,  $J_{5,4}$  6.1 Hz,  $J_{5,5}$  12.2 Hz, H-5), 2.10 (3xs, 9 H, 3xAc), 2.00 (3xs, 9 H, 3xAc);  $^{13}C$ -NMR (100 MHz)  $\delta_C$  ( $CDCl_3$ ): 170.8, 170.5, 170.3, 170.2, 169.9, 169.4, 137.0, 129.2, 128.8, 128.4, 128.3, 128.2, 125.5, 98.4, 97.3, 75.0, 70.3, 70.2, 68.2, 68.0, 67.3, 66.9, 61.5, 60.5, 21.1, 21.0, 20.9 (x2), 20.8 (x2) ; HR ESI-MS:  $m/z = 677.2057 [M + Na]^+$  (expected for  $C_{30}H_{38}O_{16}Na^+$ :  $m/z = 677.2058$ ).

***n*-Octyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 3)-2,4,6-tri-O-acetyl- $\beta$ -D-glucoside:**

19.9 mg (52 %);  $^1H$ -NMR (400 MHz):  $\delta_H$  ( $CDCl_3$ ): 5.43 (m, 1 H, H-4'), 5.31 (d, 1 H,  $J_{1,2}$  3.2 Hz, H-1'), 5.23 (dd, 1 H,  $J_{3,2}$  11.0 Hz,  $J_{3,4}$  3.3 Hz, H-3'), 5.16 (t, 1 H,  $J_{4,3} = J_{4,5}$  9.60 Hz, H-4), 5.10-5.00 (m, 2 H, H-2, H-2'), 4.36 (d, 1 H,  $J_{1,2}$  7.8 Hz, H-1), 4.27 (m, 1 H, H-5'), 4.23-3.98 (m, 4 H, H-6, H-6, H-6', H-6'), 3.95-3.82 (m, 2 H, H-3,  $CH_2$ ), 3.61-3.54 (m, 1 H, H-5), 3.47-3.39 (m, 1 H,  $CH_2$ ), 2.13 (s, 3 H, Ac), 2.09 (2xs, 6 H, Ac), 2.08 (2xs, 6 H, Ac), 2.05 (s, 3 H, Ac), 1.98 (s, 3 H, Ac), 1.60-1.54 (m, 2 H,  $CH_2$ ), 1.34-1.24 (m, 10 H, 5x  $CH_2$ ), 0.89 (t, 3 H,  $J$  7.0 Hz,  $CH_3$ );  $^{13}C$ -NMR (100MHz)  $\delta_C$  ( $CDCl_3$ ): 171.1, 171.0,

170.5, 170.3, 169.8, 169.6, 169.1, 101.2, 96.5, 72.3, 72.1, 70.4, 70.3, 68.6, 68.4, 67.8, 67.1, 67.0, 62.3, 61.8, 32.1, 29.6, 29.5 (x2), 26.1, 22.9, 21.1-20.8 (x7), 14.8; HR ESI-MS:  $m/z = 771.3049 [M + Na]^+$  (expected for  $C_{34}H_{52}O_{18}Na^+$ :  $m/z = 771.3051$ ).

***n*-Octyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 4)-2,3,6-tri-O-acetyl- $\beta$ -D-galactoside:**

20.6 mg (54 %);  $^1H$ -NMR (400 MHz):  $\delta_H$  ( $CDCl_3$ ): 5.57 (m, 1 H, H-4'), 5.39 (dd, 1 H,  $J_{3',2'}$  11.0 Hz,  $J_{3',4'}$  3.3 Hz, H-3'), 5.23-5.14 (m, 2 H, H-2, H-2'), 5.01 (d, 1 H,  $J_{1',2'}$  3.6 Hz, H-1'), 4.82 (dd, 1 H,  $J_{3,2}$  10.8 Hz,  $J_{3,4}$  2.7 Hz, H-3), 4.54 (m, 1 H, H-5'), 4.50-4.42 (m, 2 H, H-6, H-1), 4.22-4.04 (m, 4 H, H-6', H-6', H-6, H-4), 3.93-3.84 (m, 1 H,  $CH_2$ ), 3.81-3.74 (m, 1 H, H-5), 3.52-3.342 (m, 1 H,  $CH_2$ ), 2.13 (s, 3 H, Ac), 2.10 (s, 3 H, Ac), 2.08 (2xs, 6 H, Ac), 2.05 (2xs, 6 H, Ac), 1.99 (s, 3 H, Ac), 1.62-1.54 (m, 2 H,  $CH_2$ ), 1.34-1.24 (m, 10 H, 5x  $CH_2$ ), 0.88 (t, 3 H,  $J$  7.0 Hz,  $CH_3$ );  $^{13}C$ -NMR (100MHz)  $\delta_C$  ( $CDCl_3$ ): 171.0, 170.8, 170.7 (x2), 170.4, 170.0, 169.3, 101.5, 99.6, 73.0, 72.0, 70.3 (x2), 69.0, 68.8, 68.1, 67.6, 67.3, 62.2, 60.7, 32.0, 29.6, 29.5, 29.5, 26.1, 22.9, 21.2-20.8 (x7), 14.3; HR ESI-MS:  $m/z = 771.3055 [M + Na]^+$  (expected for  $C_{34}H_{52}O_{18}Na^+$ :  $m/z = 771.3051$ ).

***n*-Octyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 3)-2,4-di-O-acetyl- $\beta$ -D-xyloside:**

27.8 mg (72%);  $^1H$ -NMR (400 MHz)  $\delta_H$  ( $CDCl_3$ ): 5.45 (m, 1 H, H-4'), 5.43 (d, 1 H,  $J_{1',2'}$  3.8 Hz, H-1'), 5.26 (dd, 1 H,  $J_{3',2'}$  10.9 Hz,  $J_{3',4'}$  3.3 Hz, H-3'), 5.07 (dd, 1 H,  $J_{2',1'}$  3.8 Hz,  $J_{2',3'}$  10.9 Hz, H-2'), 4.94 (dd, 1 H,  $J_{2,1}$  5.4 Hz,  $J_{2,3}$  7.0 Hz, H-2), 4.87-4.81 (m, 1 H, H4), 4.47 (d, 1 H,  $J_{1,2}$  5.4 Hz, H-1), 4.36 (m, 1 H, H-5'), 4.21-4.01 (m, 3 H, H-6', H-5, H-6'), 3.89 (t, 1 H,  $J_{3,2} = J_{3,4}$  7.0 Hz, H-3), 3.82-3.74 (m, 1 H,  $CH_2$ ), 3.44-3.37 (m, 1 H,  $CH_2$ ), 3.33 (dd, 1 H,  $J_{5,4}$  7.1 Hz,  $J_{5,5}$  12.0 Hz, H-5), 2.13 (s, 3 H, Ac), 2.09 (2xs, 6 H, Ac), 2.07

(s, 3 H, Ac), 2.06 (s, 3 H, Ac), 1.99 (s, 3 H, Ac), 1.62-1.54 (m, 2 H, CH<sub>2</sub>), 1.35-1.23 (m, 10 H, 5x CH<sub>2</sub>), 0.89 (t, 3 H, J 7.0 Hz, CH<sub>3</sub>); <sup>13</sup>C-NMR (100MHz) δ<sub>C</sub> (CDCl<sub>3</sub>): 170.9, 170.5, 170.4, 170.2, 169.9, 169.3, 100.2, 96.9, 75.2, 70.8, 70.7, 69.5, 68.3, 67.9, 67.4, 66.9, 61.3, 60.9, 32.0, 29.6, 29.5 (x2), 26.2, 22.9, 21.2, 21.0, 20.9 (x3), 20.8, 14.3; HR ESI-MS: m/z = 699.2848 [M + Na]<sup>+</sup> (expected for C<sub>31</sub>H<sub>48</sub>O<sub>16</sub>Na<sup>+</sup>: m/z = 699.2840).

**(2,3,4,6-Tetra-O-acetyl-α-D-galactopyranosyl)-(1→2)-1,3,4-tri-O-acetyl-6-O-benzoyl-α-D-mannose:**

12.1 mg (84%); <sup>1</sup>H-NMR (400 MHz): δ<sub>H</sub> (CDCl<sub>3</sub>): 8.08 (m, 2H, Ar), 7.52 (m, 1 H, Ar), 7.38 (m, 2 H, Ar), 6.13 (d, 1 H, J<sub>1,2</sub> 2.1 Hz, H-1), 5.58 (t, 1 H, J<sub>4,3</sub> = J<sub>4,5</sub> 10.1 Hz, H-4), 5.43 (d, 1 H, J<sub>4',3'</sub> 3.2 Hz, H-4'), 5.36-5.30 (m, 3 H, H-1', H-3, H-3'), 5.08 (dd, 1 H, J<sub>2',1'</sub> 3.9 Hz, J<sub>2',3'</sub> 11.0 Hz, H-2'), 4.48-4.37 (m, 2 H, H-6, H-6), 4.32-4.28 (m, 1 H, H-5'), 4.16-4.12 (ddd, 1 H, H-5), 4.08 (t, 1 H, J<sub>2,1</sub> = J<sub>2,3</sub> 2.4 Hz, H-2), 4.05-4.00 (m, 2 H, H-6, H-6), 2.15 (s, 3 H, Ac), 2.10 (s, 3 H, Ac), 2.06 (s, 3 H, Ac), 2.02 (s, 3 H, Ac), 2.00 (3xs, 9 H, 3xAc); <sup>13</sup>C-NMR (100 MHz) δ<sub>C</sub> (CDCl<sub>3</sub>): 171.1, 170.6, 170.2, 170.0, 169.6, 169.3, 168.1, 166.1, 133.0, 129.8 (x2), 129.7, 128.3 (x2), 97.0, 91.9, 72.9, 70.9, 70.7, 67.9, 67.8, 67.3 (x2), 65.9, 62.2, 61.9, 20.9-20.4 (x7); HR ESI-MS: m/z = 763.2063 [M + Na]<sup>+</sup> (expected for C<sub>33</sub>H<sub>40</sub>O<sub>19</sub>Na<sup>+</sup>: m/z = 763.2061).

**(2,3,4,6-Tetra-O-acetyl-α-D-galactopyranosyl)-(1→2)-1,3-di-O-acetyl-4-O-benzoyl-α/β-D-xylose:**

5.9 mg (75 %); <sup>1</sup>H-NMR (600 MHz): δ<sub>H</sub> (CDCl<sub>3</sub>): 8.01 (m, 4 H, Ar), 7.61 (m, 2 H, Ar), 7.47 (m, 4 H, Ar), 6.30 (d, 1 H, J<sub>1,2</sub> 3.4 Hz, H-1<sub>α</sub>), 5.71-5.66 (m, 2 H, H-1<sub>β</sub>, H-3<sub>α</sub>), 5.54 (d, 1 H, J<sub>1',2'</sub> 3.8 Hz, H-1<sub>α'</sub>), 5.49-5.44 (m, 2 H, H-3<sub>β</sub>, H4<sub>β'</sub>), 5.41(m, 1 H, H4<sub>α'</sub>), 5.25-5.19 (m, 4 H, H3<sub>α'</sub>, H-1<sub>β'</sub>, H-2<sub>β'</sub>, H3<sub>β'</sub>) 5.16 - 5.10 (m, 2 H, H-4<sub>α</sub>, H-4<sub>β</sub>), 5.05 (dd, 1 H,

$J_{2,1}$ , 3.8 Hz,  $J_{2,3}$ , 10.9 Hz, H-2 $_{\alpha}$ '), 4.31-4.18 (m, 4 H, H-5 $_{\beta}$ ', H-5 $_{\beta}$ , H-5 $_{\alpha}$ ', H-6 $_{\beta}$ '), 4.16-4.04 (m, 3 H, H-5 $_{\alpha}$ , H-6 $_{\alpha}$ ', H-6 $_{\alpha}$ '), 4.03-3.99 (m, 1 H, H-6 $_{\beta}$ '), 3.97 (dd, 1 H,  $J_{2,1}$  3.6 Hz,  $J_{3,2}$  10.0 Hz, H-2 $_{\alpha}$ ), 3.91(t, 1 H,  $J_{2,1} = J_{2,3}$  7.5 Hz, H-2 $_{\beta}$ ) 3.80 (dd, 1 H,  $J_{5,4} = J_{5,5}$  10.9 Hz, H-5 $_{\alpha}$ ), 3.60 (dd, 1 H,  $J_{5,4}$  9.4 Hz,  $J_{5,5}$  11.6 Hz, H-5 $_{\beta}$ ), 2.24, 2.17, 2.15, 2.14, 2.09, 2.07, 2.06, 2.06, 2.05, 2.05, 2.00, 1.98 (12xs, 36 H, 12xAc);  $^{13}\text{C}$ -NMR (100 MHz)  $\delta_{\text{C}}$  (CDCl $_3$ ): 170.7, 170.2, 169.9, 169.7, 169.2, 169.1, 165.5, 133.7, 129.9 (2x), 128.9, 128.6 (2x), 96.1, 95.5, 94.0, 88.9, 74.1, 73.8, 71.8, 70.1, 69.7, 69.5, 68.0, 67.7, 67.5, 67.4 (x2), 67.2, 67.0, 66.8, 63.1, 61.1 (x2), 60.8, 21.0-20.6 (x6); HR ESI-MS:  $m/z = 691.1840$  [M + Na] $^{+}$  (expected for C $_{30}$ H $_{36}$ O $_{17}$ Na $^{+}$ :  $m/z = 691.1850$ ).

#### 7.5.4. $\alpha$ 3GalT-Catalysed Galactosylation of Alternative Acceptor Substrates

Acceptor substrate (10 mM), UDP galactose (15 mM) and  $\alpha$ 3GalT (~0.5 mg/mL) were incubated for ~16 hours at room temperature in 50 mM HEPES buffer (pH 7.5) containing 10 mM MnCl $_2$ , 0.1% BSA, and 15 U of bovine alkaline phosphatase. Upon completion, reactions were purified as described for LgtC.

##### ***para*-Nitrophenyl (2,3,4,6-tetra-O-acetyl- $\alpha$ -D-galactopyranosyl)-(1 $\rightarrow$ 3)-2,4,6-tri-O-acetyl- $\beta$ -D-galactoside:**

41.4 mg (82 %);  $^1\text{H}$ -NMR (400 MHz):  $\delta_{\text{H}}$  (CDCl $_3$ ): 8.21 (m, 2 H, Ar), 7.09 (m, 2 H, Ar), 5.55 (dd, 1 H,  $J_{2,1}$  7.56 Hz,  $J_{2,3}$  9.98 Hz, H-2), 5.50-5.44 (m, 2 H, H-4', H-4), 5.32-5.27 (m, 2 H, H-1', H-2'), 5.22-5.16 (m, 1 H, H-3'), 5.13 (d, 1 H,  $J_{1,2}$  7.56 Hz, H-1), 4.32-4.26 (m, 1 H, H-5'), 4.26-4.02 (m, 6 H, H-6', H-6', H-6, H-6, H-5, H-3), 2.20 (s, 3 H, Ac), 2.16 (2xs, 6 H, Ac), 2.10 (s, 3 H, Ac), 2.06 (s, 3 H, Ac), 2.03 (s, 3 H, Ac), 1.98 (s, 3 H, Ac);  $^{13}\text{C}$ -NMR (100 MHz)  $\delta_{\text{C}}$  (CDCl $_3$ ): 170.5, 170.4 (x3), 169.9, 169.1, 161.4, 143.4,

129.2, 126.0 (x2), 116.8 (x2), 98.8, 93.9, 72.7, 71.7, 69.2, 67.8, 67.3, 67.1, 66.8, 65.0, 61.8, 61.4, 21.0-20.7 (x7); HR ESI-MS:  $m/z = 780.1964$   $[M + Na]^+$  (expected for  $C_{32}H_{39}NO_{20}Na^+$ :  $m/z = 780.1963$ ).

### **7.5.5. Cst II-catalysed sialylation of alternative acceptor substrates**

Acceptor substrate (10 mM), CMP NeuAc (15 mM) and Cst II (~0.5 mg/mL) were incubated at room temperature in 100 mM HEPES buffer (pH 7.5) containing 10 mM  $MgCl_2$ . Reactions were monitored by TLC (7: 2: 1: 0.1 EtOAc: MeOH:  $H_2O$ : AcOH). When Gal was used as acceptor, only donor substrate hydrolysis and no product formation was detected, even after extensive incubation. When pNP Gal was used as acceptor, near quantitative conversion to product was observed following a 6 hour incubation. This product was partially purified using a SepPak  $C_{18}$ Plus Cartridge by fractionated elution with water and characterized by HR ESI-MS (expected for  $C_{23}H_{32}N_2O_{16}Na^+$ :  $m/z = 615.1650$ ,  $m/z = 615.1648$   $[M + Na]^+$  found). Regio-selectivity was confirmed by treatment of 5 nmol of the reaction product with a selective  $\alpha$ -2,3-sialidase (New England Biolabs). Under a unit defined time condition, complete reversion of the product to pNP Gal was observed (Figure 7.1).



**Figure 7.1.** TLC analysis of  $\alpha$ -2,3 specific neuraminidase treatment of glycosylation product derived from Cst II using pNP  $\beta$ -galactoside as an alternative acceptor substrate (7:2:1:0.1 EtOAc: MeOH: H<sub>2</sub>O: AcOH). (A) pNP  $\beta$ -galactoside starting material, (B) sialylated product derived from Cst II, (C) sialylated product treated with specific neuraminidase.

### 7.5.6. Testing Alternative Cst II Donor Substrates

FCHASE lactose acceptor substrate (10 mg, 12.6  $\mu$ mol), pNP  $\alpha$ -D-sialoside (17 mg, 39  $\mu$ mol), CMP (7 mg, 19  $\mu$ mol) and Cst II ( $\sim$ 0.5 mg/mL) were incubated at room temperature in 3 mL of 100 mM HEPES buffer (pH 7.5) containing 10 mM MgCl<sub>2</sub>. Reactions were monitored by TLC (7: 2: 1: 0.1 EtOAc: MeOH: H<sub>2</sub>O: AcOH). Upon completion of the reaction ( $\sim$ 36 hours), the product was partially purified using a SepPak C<sub>18</sub>Plus Cartridge by fractionated elution with water and those fractions containing the sialylated product were combined, lyophilized, resuspended in 2 mL of water and purified by preparative HPLC using an Amide-80 column as described in section 6.3.7. The HPLC purified product (10.1 mg, 72%) was characterized by HR ESI-MS (expected for C<sub>50</sub>H<sub>53</sub>N<sub>2</sub>O<sub>25</sub><sup>-</sup>: m/z = 1081.2937, m/z = 1081.2931 [M]<sup>-</sup> found). Regio-selectivity was confirmed by treatment of 5 nmol of the reaction product with a selective  $\alpha$ -2,3-sialidase

(New England Biolabs). Under a unit defined time condition, complete reversion of the product to FCHASE-lactose was observed (Figure 7.2).



**Figure 7.2.** TLC analysis of  $\alpha$ -2,3 specific neuraminidase treatment of glycosylation product derived from Cst II using *p*-nitrophenyl  $\alpha$ -sialoside as an alternative donor substrate in the presence of CMP monitored by TLC (7:2:1:0.1 EtOAc: MeOH: H<sub>2</sub>O: AcOH). (A) Fluorescein lactose conjugate starting material, (B) sialylated product derived from Cst II using alternative donor substrate, (C) sialylated product treated with specific neuraminidase.

### 7.5.7. Testing Alternative LgtC Donor Substrates

FCHASE lactose acceptor substrate (5 mg, 6.2  $\mu$ mol), dNP  $\beta$ -D-galactoside (6.5 mg, 18.8  $\mu$ mol), UDP (4.0 mg, 8.9  $\mu$ mol) and LgtC ( $\sim$ 0.5 mg/mL) were incubated at room temperature in 3 mL of 100 mM HEPES buffer (pH 7.5) containing 10 mM MnCl<sub>2</sub>, and 10 mM DTT. Reactions were monitored by TLC (7: 2: 1: 0.1 EtOAc: MeOH: H<sub>2</sub>O: AcOH). Following a week long incubation at room temperature, although product formation could be detected, the majority of the starting acceptor substrate remained and all of the dNP  $\beta$ -D-galactoside donor substrate surrogate had been consumed (presumably

due to spontaneous hydrolysis). As such, yields were insufficient for isolation of observable masses of product and characterization was limited to HR ESI-MS (expected for  $C_{45}H_{48}N_2O_{19}^+$ :  $m/z = 954.2668$ ,  $m/z = 954.2674$   $[M]^+$  found).

## References

- Aharoni, A., Amitai, G., Bernath, K., Magdassi, S., and Tawfik, D.S. (2005). High-throughput screening of enzyme libraries: Thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. *Chemistry & Biology* 12, 1281-1289.
- Aharoni, A., Thieme, A.A.K., Chiu, C.P.C., Buchini, S., Lairson, L.L., Chen, H.M., Strynadka, N.C.J., Wakarchuk, W.W., and Withers, S.G. (2006). High-throughput screening methodology for the directed evolution of glycosyltransferases. *Nature Methods* 3, 609-614.
- Alaimo, P.J., Shogren-Knaak, M.A., and Shokat, K.M. (2001). Chemical genetic approaches for the elucidation of signaling pathways. *Current Opinion in Chemical Biology* 5, 360-367.
- Amaya, M.F., Buschiazzo, A., Nguyen, T., and Alzari, P.M. (2003). The high resolution structures of free and inhibitor-bound Trypanosoma rangeli sialidase and its comparison with T-cruzi trans-sialidase. *Journal of Molecular Biology* 325, 773-784.
- Atwell, S., and Wells, J.A. (1999). Selection for improved subtiligases by phage display. *Proceedings of the National Academy of Sciences of the United States of America* 96, 9497-9502.
- Baisch, G., Ohrlein, R., and Katopodis, A. (1996). Enzymatic fucosylations with purine-diphosphate-fucoses (PDP-fucoses). *Bioorganic & Medicinal Chemistry Letters* 6, 2953-2956.
- Barbas III, C.F., Burton, D.R., Scott, J.K., and Silverman, G.J. (2001). *Phage Display: A Laboratory Manual* (New York, Cold Spring Harbor Press).
- Bencur, P., Steinkellner, H., Svoboda, B., Mucha, J., Strasser, R., Kolarich, D., Hann, S., Kollensperger, G., Glossl, J., Altmann, F., *et al.* (2005). Arabidopsis thaliana beta 1,2-xylosyltransferase: an unusual glycosyltransferase with the potential to act at multiple stages of the plant N-glycosylation pathway. *Biochemical Journal* 388, 515-525.
- Bendiak, B., and Schachter, H. (1987). Control of Glycoprotein-Synthesis - Kinetic Mechanism, Substrate-Specificity, and Inhibition Characteristics of Udp-N-Acetylglucosamine-Alpha-D-Mannoside Beta-1-2 N-Acetylglucosaminyltransferase-Ii from Rat-Liver. *Journal of Biological Chemistry* 262, 5784-5790.

Bernath, K., Hai, M.T., Mastrobattista, E., Griffiths, A.D., Magdassi, S., and Tawfik, D.S. (2004). In vitro compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting. *Analytical Biochemistry* 325, 151-157.

Berti, P.J., and Tanaka, K.S.E. (2002). Transition state analysis using multiple kinetic isotope effects: Mechanisms of enzymatic and non-enzymatic glycoside hydrolysis and transfer. In *Advances in Physical Organic Chemistry*, Vol 37, pp. 239-314.

Bertrand, T., Briozzo, P., Assairi, L., Ofiteru, A., Bucurenci, N., Munier-Lehmann, H., Golinelli-Pimpaneau, B., Barzu, O., and Gilles, A.M. (2002). Sugar specificity of bacterial CMP kinases as revealed by crystal structures and mutagenesis of *Escherichia coli* enzyme. *Journal of Molecular Biology* 315, 1099-1110.

Blanco, G., Patallo, E.P., Brana, A.F., Trefzer, A., Bechthold, A., Rohr, J., Mendez, C., and Salas, J.A. (2001). Identification of a sugar flexible glycosyltransferase from *Streptomyces olivaceus*, the producer of the antitumor polyketide elloramycin. *Chemistry & Biology* 8, 253-263.

Boix, E., Zhang, Y.N., Swaminathan, G.J., Brew, K., and Acharya, K.R. (2002). Structural basis of ordered binding of donor and acceptor substrates to the retaining glycosyltransferase, alpha-1,3-galactosyltransferase. *Journal of Biological Chemistry* 277, 28310-28318.

Bolam, D.N., Roberts, S., Proctor, M.R., Turkenburg, J.P., Dodson, E.J., Martinez-Fleites, C., Yang, M., Davis, B.G., Davies, G.J., and Gilbert, H.J. (2007). The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. *Proceedings of the National Academy of Sciences of the United States of America* 104, 5336-5341.

Boozer, C.E., and Lewis, E.S. (1953). The Decomposition of Secondary Alkyl Chlorosulfites. II. Solvent Effects and Mechanisms. *Journal of the American Chemical Society* 75, 3182-3186.

Borisova, S.A., Zhao, L.S., Melancon, C.E., Kao, C.L., and Liu, H.W. (2004). Characterization of the glycosyltransferase activity of DesVII: Analysis of and implications for the biosynthesis of macrolide antibiotics. *Journal of the American Chemical Society* 126, 6534-6535.

Breton, C., Bettler, E., Joziassse, D.H., Geremia, R.A., and Imberty, A. (1998). Sequence-function relationships of prokaryotic and eukaryotic galactosyltransferases. *Journal of Biochemistry* 123, 1000-1009.

Breton, C., and Imberty, A. (1999). Structure/function studies of glycosyltransferases. *Current Opinion in Structural Biology* 9, 563-571.

Brown, K., Pompeo, F., Dixon, S., Mengin-Lecreulx, D., Cambillau, C., and Bourne, Y. (1999). Crystal structure of the bifunctional N-acetylglucosamine 1-phosphate uridyltransferase from *Escherichia coli*: a paradigm for the related pyrophosphorylase superfamily. *Embo Journal* 18, 4096-4107.

Burmeister, W.P., Ruigrok, R.W.H., and Cusack, S. (1992). The 2.2-Å Resolution Crystal-Structure of Influenza-B Neuraminidase and Its Complex with Sialic-Acid. *Embo Journal* 11, 49-56.

Busch, C., Hofmann, F., Selzer, J., Munro, S., Jeckel, D., and Aktories, K. (1998). A common motif of eukaryotic glycosyltransferases is essential for the enzyme activity of large clostridial cytotoxins. *Journal of Biological Chemistry* 273, 19566-19572.

Buschiazzo, A., Ugalde, J.E., Guerin, M.E., Shepard, W., Ugalde, R.A., and Alzari, P.M. (2004). Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation. *Embo Journal* 23, 3196-3205.

Cadwell, R.C., and Joyce, G.F. (1994). Mutagenic Pcr. *Pcr-Methods and Applications* 3, S136-S140.

Campbell, J.A., Davies, G.J., Bulone, V., and Henrissat, B. (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochemical Journal* 326, 929-939.

Campbell, R.E., Mosimann, S.C., Tanner, M.E., and Strynadka, N.C.J. (2000). The structure of UDP-N-acetylglucosamine 2-epimerase reveals homology to phosphoglycosyl transferases. *Biochemistry* 39, 14993-15001.

Carey, F.A., and Sundberg, R.J. (1990). *Advanced Organic Chemistry. Part A: Structure and Mechanisms*, Third edn (New York, Plenum Press).

Cesaro-Tadic, S., Lagos, D., Honegger, A., Rickard, J.H., Partridge, L.J., Blackburn, G.M., and Pluckthun, A. (2003). Turnover-based in vitro selection and evolution of biocatalysts from a fully synthetic antibody library. *Nature Biotechnology* 21, 679-685.

Chaiet, L.a.W., F.J. (1964). The Properties of Streptavidin, a Biotin-Binding Protein Produced by Streptomyces. *Archives of Biochemistry and Biophysics* 106, 1-5.

Charnock, S.J., and Davies, G.J. (1999). Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. *Biochemistry* 38, 6380-6385.

Chen, L., Men, H., Ha, S., Ye, X.Y., Brunner, L., Hu, Y., and Walker, S. (2002). Intrinsic lipid preferences and kinetic mechanism of *Escherichia coli* MurG. *Biochemistry* *41*, 6824-6833.

Chiu, C.P.C., Lairson, L.L., Gilbert, M., Wakarchuk, W.W., Withers, S.G., and Strynadka, N.C.J. (2007). Structural analysis of the alpha-2,3-sialyltransferase cst-i from *Campylobacter jejuni* in apo and substrate-analogue bound forms. *Biochemistry* *46*, 7196-7204.

Chiu, C.P.C., Watts, A.G., Lairson, L.L., Gilbert, M., Lim, D., Wakarchuk, W.W., Withers, S.G., and Strynadka, N.C.J. (2004). Structural analysis of the sialyltransferase CstII from *Campylobacter jejuni* in complex with a substrate analog. *Nature Structural & Molecular Biology* *11*, 163-170.

Christianson, D.W. (2006). Structural biology and chemistry of the terpenoid cyclases. *Chemical Reviews* *106*, 3412-3442.

Cleland, W.W. (1963). The Kinetics of Enzyme-Catalyzed Reactions with Two or More Substrates or Products. I. Nomenclature and Rate Equations. *Biochimica et Biophysica Acta* *67*, 104-137.

Close, B.E., Mendiratta, S.S., Geiger, K.M., Broom, L.J., Ho, L.L., and Colley, K.J. (2003). The minimal structural domains required for neural cell adhesion molecule polysialylation by PST/ST8Sia IV and STX/ST8Sia II. *Journal of Biological Chemistry* *278*, 30796-30805.

Cohen, H.M., Tawfik, D.S., and Griffiths, A.D. (2004). Altering the sequence specificity of HaeIII methyltransferase by directed evolution using in vitro compartmentalization. *Protein Engineering Design & Selection* *17*, 3-11.

Coia, J.E. (1998). Clinical, microbiological and epidemiological aspects of *Escherichia coli* O157 infection. *Fems Immunology and Medical Microbiology* *20*, 1-9.

Cornish-Bowden, A. (2004). *Fundamentals of Enzyme Kinetics*, Third edn (London, Portland Press).

Coutinho, P.M., Deleury, E., Davies, G.J., and Henrissat, B. (2003). An evolving hierarchical family classification for glycosyltransferases. *Journal of Molecular Biology* *328*, 307-317.

Cowdrey, W.A., Hughes, E.D., Ingold, C.K., Masterman, S., and Scott, A.D. (1937). *Reaction Kinetics and the Walden Inversion*. Part IV. Relation of Steric Orientation to

Mechanism in Substitutions Involving Halogen Atoms and Simple or Substituted Hydroxyl Groups. *Journal of the Chemical Society*, 1253-1271.

Cram, D.J. (1953). Studies in Stereochemistry. XVI. Ionic Intermediates in the Decomposition of Certain Alkyl Chlorosulfites. *Journal of the American Chemical Society* 75, 332-338.

Crennell, S., Takimoto, T., Portner, A., and Taylor, G. (2000). Crystal structure of the multifunctional paramyxovirus hemagglutinin-neuraminidase. *Nature Structural Biology* 7, 1068-1074.

Dall'Olio, F., and Chiricolo, M. (2001). Sialyltransferases in cancer. *Glycoconjugate Journal* 18, 841-850.

Davies, G.J., Gloster, T.M., and Henrissat, B. (2005). Recent structural insights into the expanding world of carbohydrate-active enzymes. *Current Opinion in Structural Biology* 15, 637-645.

Davies, G.S., M.L.; Withers, S.G. (1998). Glycosyl Transfer. In *Comprehensive Biological Catalysis*, M.L. Sinnott, ed. (Academic Press), pp. 119-208.

Demartis, S., Huber, A., Viti, F., Lozzi, L., Giovannoni, L., Neri, P., Winter, G., and Neri, D. (1999). A strategy for the isolation of catalytic activities from repertoires of enzymes displayed on phage. *Journal of Molecular Biology* 286, 617-633.

Dennis, J., Waller, C., Timpl, R., and Schirmacher, V. (1982). Surface Sialic-Acid Reduces Attachment of Metastatic Tumor-Cells to Collagen Type-Iv and Fibronectin. *Nature* 300, 274-276.

Doi, N., Kumadaki, S., Oishi, Y., Matsumura, N., and Yanagawa, H. (2004). In vitro selection of restriction endonucleases by in vitro compartmentalization. *Nucleic Acids Research* 32.

Durbecq, V., Sainz, G., Oudjama, Y., Clantin, B., Bompard-Gilles, C., Tricot, C., Caillet, J., Stalon, V., Droogmans, L., and Villeret, V. (2001). Crystal structure of isopentenyl diphosphate : dimethylallyl diphosphate isomerase. *Embo Journal* 20, 1530-1537.

Durr, C., Hoffmeister, D., Wohlert, S.E., Ichinose, K., Weber, M., von Mulert, U., Thorson, J.S., and Bechthold, A. (2004). The glycosyltransferase UrdGT2 catalyzes both C- and O-glycosidic sugar transfers. *Angewandte Chemie-International Edition* 43, 2962-2965.

Ekstrom, J.L., Pauly, T.A., Carty, M.D., Soeller, W.C., Culp, J., Danley, D.E., Hoover, D.J., Treadway, J.L., Gibbs, E.M., Fletterick, R.J., *et al.* (2002). Structure-activity analysis of the purine binding site of human liver glycogen phosphorylase. *Chemistry & Biology* 9, 915-924.

Endo, Y., Tsurugi, K., Yutsudo, T., Takeda, Y., Ogasawara, T., and Igarashi, K. (1988). Site of Action of a Vero Toxin (Vt2) from *Escherichia-Coli* O157-H7 and of Shiga Toxin on Eukaryotic Ribosomes - Rna N-Glycosidase Activity of the Toxins. *European Journal of Biochemistry* 171, 45-50.

Endtz, H.P., Ang, C.W., van den Braak, N., Duim, B., Rigter, A., Price, L.J., Woodward, D.L., Rodgers, F.G., Johnson, W.M., Wagenaar, J.A., *et al.* (2000). Molecular characterization of *Campylobacter jejuni* from patients with Guillain-Barre and Miller Fisher syndromes. *Journal of Clinical Microbiology* 38, 2297-2301.

Ernst, S., Langer, R., Cooney, C.L., and Sasisekharan, R. (1995). Enzymatic Degradation of Glycosaminoglycans. *Critical Reviews in Biochemistry and Molecular Biology* 30, 387-444.

Fa, M., Radeghieri, A., Henry, A.A., and Romesberg, F.E. (2004). Expanding the substrate repertoire of a DNA polymerase by directed evolution. *Journal of the American Chemical Society* 126, 1748-1754.

Fenton, C., Hao, X., Petersen, S.B., and El-Gewely, M.R. (2002). Random mutagenesis for protein breeding. In *In Vitro Mutagenesis Protocols*, J. Braman, ed. (Totowa, Humana Press), pp. 231-241.

Fernandez-Gacio, A., Uguen, M., and Fastrez, J. (2003). Phage display as a tool for the directed evolution of enzymes. *Trends in Biotechnology* 21, 408-414.

Fersht, A. (1999). *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (New York, W.H. Freeman and Company).

Flint, J., Taylor, E., Yang, M., Bolam, D.N., Tailford, L.E., Martinez-Fleites, C., Dodson, E.J., Davis, B.G., Gilbert, H.J., and Davies, G.J. (2005). Structural dissection and high-throughput screening of mannosylglycerate synthase. *Nature Structural & Molecular Biology* 12, 608-614.

Fritz, T.A., Hurley, J.H., Trinh, L.B., Shiloach, J., and Tabak, L.A. (2004). The beginnings of mucin biosynthesis: The crystal structure of UDP-GalNAc : polypeptide alpha-N-acetylgalactosaminyltransferase-T1. *Proceedings of the National Academy of Sciences of the United States of America* 101, 15307-15312.

Fritz, T.A., Raman, J., and Tabak, L.A. (2006). Dynamic association between the catalytic and lectin domains of human UDP-GalNAc : polypeptide alpha-N-acetylgalactosaminyltransferase-2 (vol 281, pg 8613, 2006). *Journal of Biological Chemistry* 281, 25000-25000.

Fromm, H.J. (1975). *Initial Rate Enzyme Kinetics* (New York, Springer-Verlag).

Fuchs, G., Mobassaleh, M., Donohue, A., Montgomery, R.K., Grand, R.J., and Keusch, G.T. (1986). Pathogenesis of Shigella Diarrhea - Rabbit Intestinal-Cell Microvillus Membrane-Binding Site for Shigella Toxin. *Infection and Immunity* 53, 372-377.

Gabelli, S.B., McLellan, J.S., Montalvetti, A., Oldfield, E., Docampo, R., and Amzel, L.M. (2006). Structure and mechanism of the farnesyl diphosphate synthase from *Trypanosoma cruzi*: Implications for drug design. *Proteins-Structure Function and Bioinformatics* 62, 80-88.

Galili, U., and LaTemple, D.C. (1998). Augmenting immunogenicity of autologous tumor vaccines by the natural anti-Gal antibody. *Faseb Journal* 12, A280-A280.

Garinot-Schneider, C., Lellouch, A.C., and Geremia, R.A. (2000). Identification of essential amino acid residues in the *Sinorhizobium meliloti* glucosyltransferase ExoM. *Journal of Biological Chemistry* 275, 31407-31413.

Gastinel, L.N., Bignon, C., Misra, A.K., Hindsgaul, O., Shaper, J.H., and Joziase, D.H. (2001). Bovine alpha 1,3-galactosyltransferase catalytic domain structure and its relationship with ABO histo-blood group and glycosphingolipid glycosyltransferases. *Embo Journal* 20, 638-649.

Gastinel, L.N., Cambillau, C., and Bourne, Y. (1999). Crystal structures of the bovine beta 4galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose. *Embo Journal* 18, 3546-3557.

Gebler, J.C., Trimbur, D.E., Warren, A.J., Aebersold, R., Namchuk, M., and Withers, S.G. (1995). Substrate-Induced Inactivation of a Crippled Beta-Glucosidase Mutant - Identification of the Labeled Amino-Acid and Mutagenic Analysis of Its Role. *Biochemistry* 34, 14547-14553.

Geremia, S., Campagnolo, M., Schinzel, R., and Johnson, L.N. (2002). Enzymatic catalysis in crystals of *Escherichia coli* maltodextrin phosphorylase. *Journal of Molecular Biology* 322, 413-423.

Ghadessy, F.J., Ong, J.L., and Holliger, P. (2001). Directed evolution of polymerase function by compartmentalized self-replication. *Proceedings of the National Academy of Sciences of the United States of America* 98, 4552-4557.

Ghadessy, F.J., Ramsay, N., Boudsocq, F., Loakes, D., Brown, A., Iwai, S., Vaisman, A., Woodgate, R., and Holliger, P. (2004). Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nature Biotechnology* 22, 755-759.

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

Gibson, R.P., Tarling, C.A., Roberts, S., Withers, S.G., and Davies, G.J. (2004). The donor subsite of trehalose-6-phosphate synthase - Binary complexes with UDP-glucose and UDP-2-deoxy-2-fluoro-glucose at 2 angstrom resolution. *Journal of Biological Chemistry* 279, 1950-1955.

Gibson, R.P., Turkenburg, J.P., Charnock, S.J., Lloyd, R., and Davies, G.J. (2002). Insights into trehalose synthesis provided by the structure of the retaining glycosyltransferase OtsA. *Chemistry & Biology* 9, 1337-1346.

Gilbert, M., Bayer, R., Cunningham, A.M., Defrees, S., Gao, Y.H., Watson, D.C., Young, N.M., and Wakarchuk, W.W. (1998). The synthesis of sialylated oligosaccharides using a CMP-Neu5Ac synthetase/sialyltransferase fusion. *Nature Biotechnology* 16, 769-772.

Gilbert, M., Brisson, J.R., Karwaski, M.F., Michniewicz, J., Cunningham, A.M., Wu, Y.Y., Young, N.M., and Wakarchuk, W.W. (2000). Biosynthesis of ganglioside mimics in *Campylobacter jejuni* OH4384 - Identification of the glycosyltransferase genes, enzymatic synthesis of model compounds, and characterization of nanomole amounts by 600-MHz H-1 and C-13 NMR analysis. *Journal of Biological Chemistry* 275, 3896-3906.

Gilbert, M., Cunningham, A.M., Watson, D.C., Martin, A., Richards, J.C., and Wakarchuk, W.W. (1997). Characterization of a recombinant *Neisseria meningitidis* alpha-2,3-sialyltransferase and its acceptor specificity. *European Journal of Biochemistry* 249, 187-194.

Gilbert, M., Karwaski, M.F., Bernatchez, S., Young, N.M., Taboada, E., Michniewicz, J., Cunningham, A.M., and Wakarchuk, W.W. (2002). The genetic bases for the variation in the lipo-oligosaccharide of the mucosal pathogen, *Campylobacter jejuni* - Biosynthesis of sialylated ganglioside mimics in the core oligosaccharide. *Journal of Biological Chemistry* 277, 327-337.

Gordon, R.D., Sivarajah, P., Satkunarajah, M., Ma, D., Tarling, C.A., Vizitiu, D., Withers, S.G., and Rini, J.M. (2006). X-ray crystal structures of rabbit N-

acetylglucosaminyltransferase I (GnT I) in complex with donor substrate analogues. *Journal of Molecular Biology* 360, 67-79.

Gosselin, S., Alhussaini, M., Streiff, M.B., Takabayashi, K., and Palcic, M.M. (1994). A Continuous Spectrophotometric Assay for Glycosyltransferases. *Analytical Biochemistry* 220, 92-97.

Green, A.A., Cori, G.T., and Cori, C.F. (1942). CRYSTALLINE MUSCLE PHOSPHORYLASE. *Journal of Biological Chemistry* 142, 447-448.

Green, N. (1975). Avidin. *Advances in Protein Chemistry* 29, 85-133.

Griffiths, A.D., and Tawfik, D.S. (2003). Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *Embo Journal* 22, 24-35.

Grizot, S., Salem, M., Vongsouthi, V., Durand, L., Moreau, F., Dohi, H., Vincent, S., Escaich, S., and Ducruix, A. (2006). Structure of the Escherichia coli heptosyltransferase WaaC: Binary complexes with ADP AND ADP-2-deoxy-2-fluoro heptose. *Journal of Molecular Biology* 363, 383-394.

Guerry, P., Szymanski, C.M., Prendergast, M.M., Hickey, T.E., Ewing, C.P., Pattarini, D.L., and Moran, A.P. (2002). Phase variation of Campylobacter jejuni 81-176 lipooligosaccharide affects ganglioside mimicry and invasiveness in vitro. *Infection and Immunity* 70, 787-793.

Guthrie, R.D., and Jencks, W.P. (1989). Iupac Recommendations for the Representation of Reaction-Mechanisms. *Accounts of Chemical Research* 22, 343-349.

Guycaffey, J.K., Rapoza, M.P., Jolley, K.A., and Webster, R.E. (1992). Membrane Localization and Topology of a Viral Assembly Protein. *Journal of Bacteriology* 174, 2460-2465.

Ha, S., Walker, D., Shi, Y.G., and Walker, S. (2000). The 1.9 angstrom crystal structure of Escherichia coli MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Science* 9, 1045-1052.

Hancock, S.M., D Vaughan, M., and Withers, S.G. (2006). Engineering of glycosidases and glycosyltransferases. *Current Opinion in Chemical Biology* 10, 509-519.

Harduin-Lepers, A., Vallejo-Ruiz, V., Krzewinski-Recchi, M.A., Samyn-Petit, B., Julien, S., and Delannoy, P. (2001). The human sialyltransferase family. *Biochimie* 83, 727-737.

Hekmat, O., Kim, Y.W., Williams, S.J., He, S.M., and Withers, S.G. (2005). Active-site peptide "fingerprinting" of glycosidases in complex mixtures by mass spectrometry - Discovery of a novel retaining beta-1,4-glycanase in *Cellulomonas fimi*. *Journal of Biological Chemistry* 280, 35126-35135.

Hidaka, M., Honda, Y., Kitaoka, M., Nirasawa, S., Hayashi, K., Wakagi, T., Shoun, H., and Fushinobu, S. (2004). Chitobiose phosphorylase from *Vibrio proteolyticus*, a member of glycosyl transferase family 36, has a clan GH-L-like ( $\alpha/\alpha$ )<sub>6</sub> barrel fold. *Structure* 12, 937-947.

Hoffmeister, D., Ichinose, K., and Bechthold, A. (2001). Two sequence elements of glycosyltransferases involved in urdamycin biosynthesis are responsible for substrate specificity and enzymatic activity. *Chemistry & Biology* 8, 557-567.

Hoffmeister, D., Wilkinson, B., Foster, G., Sidebottom, P.J., Ichinose, K., and Bechthold, A. (2002). Engineered urdamycin glycosyltransferases are broadened and altered in substrate specificity. *Chemistry & Biology* 9, 287-295.

Holt, J.R., Gillespie, S.K.H., Provance, D.W., Shah, K., Shokat, K.M., Corey, D.P., Mercer, J.A., and Gillespie, P.G. (2002). A chemical-genetic strategy implicates myosin-1c in adaptation by hair cells. *Cell* 108, 371-381.

Horcajada, C., Guinovart, J.J., Fita, I., and Ferrer, J.C. (2006). Crystal structure of an archaeal glycogen synthase - Insights into oligomerization and substrate binding of eukaryotic glycogen synthases. *Journal of Biological Chemistry* 281, 2923-2931.

Horenstein, B.A. (1997). Quantum mechanical analysis of an alpha-carboxylate-substituted oxocarbenium ion. Isotope effects for formation of the sialyl cation and the origin of an unusually large secondary C-14 isotope effect. *Journal of the American Chemical Society* 119, 1101-1107.

Horenstein, B.A., and Bruner, M. (1998). The N-acetyl neuraminyl oxocarbenium ion is an intermediate in the presence of anionic nucleophiles. *Journal of the American Chemical Society* 120, 1357-1362.

Hosfield, D.J., Zhang, Y.M., Dougan, D.R., Broun, A., Tari, L.W., Swanson, R.V., and Finn, J. (2004). Structural basis for bisphosphonate-mediated inhibition of isoprenoid biosynthesis. *Journal of Biological Chemistry* 279, 8526-8529.

Hosseini-Maaf, B., Letts, J.A., Persson, M., Smart, E., LePenec, P.Y., Hustinx, H., Zhao, Z.H., Palcic, M.M., Evans, S.V., Chester, M.A., *et al.* (2007). Structural basis for red cell phenotypic changes in newly identified, naturally occurring subgroup mutants of the human blood group B glycosyltransferase. *Transfusion* 47, 864-875.

Hu, Y.N., Chen, L., Ha, S., Gross, B., Falcone, B., Walker, D., Mokhtarzadeh, M., and Walker, S. (2003). Crystal structure of the MurG : UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proceedings of the National Academy of Sciences of the United States of America* *100*, 845-849.

Hu, Y.N., and Walker, S. (2002). Remarkable structural similarities between diverse glycosyltransferases. *Chemistry & Biology* *9*, 1287-1296.

Huang, X., Tanaka, K.S.E., and Bennet, A.J. (1997). Glucosidase-Catalyzed Hydrolysis of  $\alpha$ -D-Glucopyranosyl Pyridinium Salts: Kinetic Evidence for Nucleophilic Involvement at the Glucosidation Transition State. *Journal of the American Chemical Society* *119*, 11147-11154.

Hughes, E.D., Ingold, C.K., and Whitfield, I.C. (1941). Walden Inversion in the replacement of hydroxyl with halogen. *Nature* *147*, 206-207.

Hwang, Y.W., and Miller, D.L. (1987). A Mutation That Alters the Nucleotide Specificity of Elongation Factor-Tu, a Gtp Regulatory Protein. *Journal of Biological Chemistry* *262*, 13081-13085.

Iannolo, G., Minenkova, O., Petruzzelli, R., and Cesareni, G. (1995). Modifying Filamentous Phage Capsid - Limits in the Size of the Major Capsid Protein. *Journal of Molecular Biology* *248*, 835-844.

Ichikawa, Y., Liu, J.L.C., Shen, G.J., and Wong, C.H. (1991). A Highly Efficient Multienzyme System for the One-Step Synthesis of a Sialyl Trisaccharide - In situ Generation of Sialic-Acid and N-Acetylglucosamine Coupled with Regeneration of Udp-Glucose, Udp-Galactose, and Cmp-Sialic Acid. *Journal of the American Chemical Society* *113*, 6300-6302.

Ihara, H., Ikeda, Y., Toma, S., Wang, X.C., Suzuki, T., Gu, J.G., Miyoshi, E., Tsukihara, T., Honke, K., Matsumoto, A., *et al.* (2007). Crystal structure of mammalian  $\alpha$  1,6-fucosyltransferase, FUT8. *Glycobiology* *17*, 455-466.

Jamaluddin, H., Tumbale, P., Withers, S.G., Acharya, K.R., and Brew, K. (2007). Conformational Changes Induced by Binding UDP-2F-galactose to  $[\alpha]$ -1,3 Galactosyltransferase- Implications for Catalysis. *Journal of Molecular Biology* *369*, 1270-1281.

Jank, T., Reinert, D.J., Giesemann, T., Schulz, G.E., and Aktories, K. (2005). Change of the donor substrate specificity of *Clostridium difficile* toxin B by site-directed mutagenesis. *Journal of Biological Chemistry* *280*, 37833-37838.

Jeanneau, C., Chazalet, V., Auge, C., Soumpasis, D.M., Harduin-Lepers, A., Delannoy, P., Imberty, A., and Breton, C. (2004). Structure-function analysis of the human sialyltransferase ST3Gal I - Role of N-glycosylation and a novel conserved sialylmotif. *Journal of Biological Chemistry* 279, 13461-13468.

Jencks, W.P. (1980). When is an intermediate not an intermediate? Enforced mechanisms of general acid-base, catalyzed, carbocation, carbanion, and ligand exchange reaction, pp. 161-169.

Jencks, W.P. (1996). Quantitative approaches to reaction mechanisms and catalysis. Substitution at carbon. *Journal of Physical Organic Chemistry* 9, 337-340.

Jestin, J.L., Kristensen, P., and Winter, G. (1999). A method for the selection of catalytic activity using phage display and proximity coupling. *Angewandte Chemie-International Edition* 38, 1124-1127.

Jinek, M., Chen, Y.W., Clausen, H., Cohen, S.M., and Conti, E. (2006). Structural insights into the Notch-modifying glycosyltransferase Fringe. *Nature Structural & Molecular Biology* 13, 945-946.

Just, I., Selzer, J., Wilm, M., Voneichelstreiber, C., Mann, M., and Aktories, K. (1995). Glucosylation of Rho-Proteins by Clostridium-Difficile Toxin-B. *Nature* 375, 500-503.

Kakuda, S., Shiba, T., Ishiguro, M., Tagawa, H., Oka, S., Kajihara, Y., Kawasaki, T., Wakatsuki, S., and Kato, R. (2004). Structural basis for acceptor substrate recognition of a human glucuronyltransferase, GlcAT-P, an enzyme critical in the biosynthesis of the carbohydrate epitope HNK-1. *Journal of Biological Chemistry* 279, 22693-22703.

Kamath, V.P., Seto, N.O.L., Compston, C.A., Hindsgaul, O., and Palcic, M.M. (1999). Synthesis of the acceptor analog alpha Fuc(1 -> 2)alpha Gal-O(CH<sub>2</sub>)(7) CH<sub>3</sub>: A probe for the kinetic mechanism of recombinant human blood group B glycosyltransferase. *Glycoconjugate Journal* 16, 599-606.

Kearns, A.E., Campbell, S.C., Westley, J., and Schwartz, N.B. (1991). Initiation of Chondroitin Sulfate Biosynthesis - a Kinetic-Analysis of Udp-D-Xylose-Core Protein Beta-D-Xylosyltransferase. *Biochemistry* 30, 7477-7483.

Kempton, J.B., and Withers, S.G. (1992). Mechanism of Agrobacterium Beta-Glucosidase - Kinetic-Studies. *Biochemistry* 31, 9961-9969.

Khaled, A., Ivannikova, T., and Auge, C. (2004). Synthesis of unnatural sugar nucleotides and their evaluation as donor substrates in glycosyltransferase-catalyzed reactions. *Carbohydrate Research* 339, 2641-2649.

Kikuchi, N., Kwon, Y.D., Gotoh, M., and Narimatsu, H. (2003). Comparison of glycosyltransferase families using the profile hidden Markov model. *Biochemical and Biophysical Research Communications* 310, 574-579.

Kitov, P.I., Sadowska, J.M., Mulvey, G., Armstrong, G.D., Ling, H., Pannu, N.S., Read, R.J., and Bundle, D.R. (2000). Shiga-like toxins are neutralized by tailored multivalent carbohydrate ligands. *Nature* 403, 669-672.

Klabunde, T., Wendt, K.U., Kadereit, D., Brachvogel, V., Burger, H.J., Herling, A.W., Oikonomakos, N.G., Kosmopoulou, M.N., Schmoll, D., Sarubbi, E., *et al.* (2005). Acyl ureas as human liver glycogen phosphorylase inhibitors for the treatment of type 2 diabetes. *Journal of Medicinal Chemistry* 48, 6178-6193.

Klein, H.W., Im, M.J., and Palm, D. (1986). Mechanism of the Phosphorylase Reaction - Utilization of D-Gluco-Hept-1-Enitol in the Absence of Primer. *European Journal of Biochemistry* 157, 107-114.

Knegtel, R.M.K., Strokopytov, B., Penninga, D., Faber, O.G., Rozeboom, H.J., Kalk, K.H., Dijkhuizen, L., and Dijkstra, B.W. (1995). Crystallographic Studies of the Interaction of Cyclodextrin Glycosyltransferase from *Bacillus-Circulans* Strain-251 with Natural Substrates and Products. *Journal of Biological Chemistry* 270, 29256-29264.

Kopp, M., Rupprath, C., Irschik, H., Bechthold, A., Elling, L., and Muller, R. (2007). SorF: A glycosyltransferase with promiscuous donor substrate specificity in vitro. *Chembiochem* 8, 813-819.

Kornberg, S.R., Zimmerman, S.B., and Kornberg, A. (1961). Glucosylation of Deoxyribonucleic Acid by Enzymes from Bacteriophage-infected *Escherichia coli*. *Journal of Biological Chemistry* 236, 1487-1493.

Koshland, D.E. (1953). Stereochemistry and the mechanism of enzymatic reactions. *Biological Reviews of the Cambridge Philosophical Society* 28, 416-436.

Kostrewa, D., D'Arcy, A., Takacs, B., and Kamber, N. (2001). Crystal structures of *Streptococcus pneumoniae* N-acetylglucosamine-1-phosphate uridylyltransferase, GlmU, in apo form at 2.33 Å resolution and in complex with UDP-N-acetylglucosamine and Mg<sup>2+</sup> at 1.96 Å resolution. *Journal of Molecular Biology* 305, 279-289.

Kozmon, S., and Tvaroska, I. (2006). Catalytic mechanism of glycosyltransferases: Hybrid quantum mechanical/molecular mechanical study of the inverting N-acetylglucosaminyltransferase I. *Journal of the American Chemical Society* 128, 16921-16927.

Kreppel, L.K., and Hart, G.W. (1999). Regulation of a cytosolic and nuclear O-GlcNAc transferase - Role of the tetratricopeptide repeats. *Journal of Biological Chemistry* 274, 32015-32022.

Kretzschmar, T., and Geiser, M. (1995). Evaluation of Antibodies Fused to Minor Coat Protein-Iii and Major Coat Protein-Viii of Bacteriophage M13. *Gene* 155, 61-65.

Kubo, A., Arai, Y., Nagashima, S., and Yoshikawa, T. (2004). Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. *Archives of Biochemistry and Biophysics* 429, 198-203.

Kubota, T., Shiba, T., Sugioka, S., Furukawa, S., Sawaki, H., Kato, R., Wakatsuki, S., and Narimatsu, H. (2006). Structural basis of carbohydrate transfer activity by human UDP-GalNAc: Polypeptide alpha-N-acetylgalactosaminyltransferase (pp-GalNAc-T10). *Journal of Molecular Biology* 359, 708-727.

Laine, R.A. (1994). A Calculation of All Possible Oligosaccharide Isomers Both Branched and Linear Yields  $1.05 \times 10^{12}$  Structures for a Reducing Hexasaccharide - the Isomer-Barrier to Development of Single-Method Saccharide Sequencing or Synthesis Systems. *Glycobiology* 4, 759-767.

Lariviere, L., Gueguen-Chaignon, V., and Morera, S. (2003). Crystal structures of the T4 phage beta-glycosyltransferase and the D100A mutant in complex with UDP-glucose: Glucose binding and identification of the catalytic base for a direct displacement mechanism. *Journal of Molecular Biology* 330, 1077-1086.

Lariviere, L., and Morera, S. (2002). A base-flipping mechanism for the T4 phage beta-glycosyltransferase and identification of a transition-state analog. *Journal of Molecular Biology* 324, 483-490.

Lariviere, L., Sommer, N., and Morera, S. (2005). Structural evidence of a passive base-flipping mechanism for AGT, an unusual GT-B glycosyltransferase. *Journal of Molecular Biology* 352, 139-150.

Lee, H.J., Barry, C.H., Borisova, S.N., Seto, N.O.L., Zheng, R.X.B., Blancher, A., Evans, S.V., and Palcic, M.M. (2005). Structural basis for the inactivity of human blood group O-2 glycosyltransferase. *Journal of Biological Chemistry* 280, 525-529.

Lee, S.S., Yu, S.K., and Withers, S.G. (2002a). alpha-1,4-glucan lyase performs a trans-elimination via a nucleophilic displacement followed by a syn-elimination. *Journal of the American Chemical Society* 124, 4948-4949.

Lee, Y.F., Tawfik, D.S., and Griffiths, A.D. (2002b). Investigating the target recognition of DNA cytosine-5 methyltransferase HhaI by library selection using in vitro compartmentalisation. *Nucleic Acids Research* 30, 4937-4944.

Legler, P.M., Massiah, M.A., and Mildvan, A.S. (2002). Mutational, kinetic, and NMR studies of the mechanism of *E. coli* GDP-mannose mannosyl hydrolase, an unusual nudix enzyme. *Biochemistry* 41, 10834-10848.

Letts, J.A., Rose, N.L., Fang, Y.R., Barry, C.H., Borisova, S.N., Seto, N.O.L., Palcic, M.M., and Evans, S.V. (2006). Differential recognition of the type I and IIH antigen acceptors by the human ABO(H) blood group A and B glycosyltransferases. *Journal of Biological Chemistry* 281, 3625-3632.

Leung, D.A., Chen, E., and Goeddel, D.V. (1989). A method for random mutagenesis of defined DNA segments using a modified polymerase chain reaction. *Technique* 1, 11-15.

Lewis, E.S., and Boozer, C.E. (1952). The Kinetics and Stereochemistry of the Decomposition of Secondary Alkyl Chlorosulfites. *Journal of the American Chemical Society* 74, 308-311.

Lin-Goerke, J.L., Robbins, D.J., and Burczak, J.D. (1997). PCR-based random mutagenesis using manganese and reduced dNTP concentration. *Biotechniques* 23, 409-412.

Lin, K., Rath, V.L., Dai, S.C., Fletterick, R.J., and Hwang, P.K. (1996). A protein phosphorylation switch at the conserved allosteric site in GP. *Science* 273, 1539-1541.

Liu, J., and Mushegian, A. (2003). Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein Science* 12, 1418-1431.

Lobsanov, Y.D., Romero, P.A., Sleno, B., Yu, B.M., Yip, P., Herscovics, A., and Howell, P.L. (2004). Structure of Kre2p/Mnt1p - A yeast alpha 1,2-mannosyltransferase involved in mannoprotein biosynthesis. *Journal of Biological Chemistry* 279, 17921-17931.

Lougheed, B., Ly, H.D., Wakarchuk, W.W., and Withers, S.G. (1999). Glycosyl fluorides can function as substrates for nucleotide phosphosugar-dependent glycosyltransferases. *Journal of Biological Chemistry* 274, 37717-37722.

Love, K.R., Swoboda, J.G., Noren, C.J., and Walker, S. (2006). Enabling glycosyltransferase evolution: A facile substrate-attachment strategy for phage-display enzyme evolution. *Chembiochem* 7, 753-756.

Lovering, A.L., de Castro, L.H., Lim, D., and Strynadka, N.C.J. (2007). Structural insight into the transglycosylation step of bacterial cell-wall biosynthesis. *Science* 315, 1402-1405.

Lukacs, C.M., Oikonomakos, N.G., Crowther, R.L., Hong, L.N., Kammlott, R.U., Levin, W., Li, S., Liu, C.M., Lucas-McGady, D., Pietranico, S., *et al.* (2006). The crystal structure of human muscle glycogen phosphorylase a with bound glucose and AMP: An intermediate conformation with T-State and R-State features. *Proteins-Structure Function and Bioinformatics* 63, 1123-1126.

Luo, Y., Li, S.C., Chou, M.Y., Li, Y.T., and Luo, M. (1998). The crystal structure of an intramolecular trans-sialidase with a NeuAc alpha 2 -> 3Gal specificity. *Structure* 6, 521-530.

Ly, H.D., Loughheed, B., Wakarchuk, W.W., and Withers, S.G. (2002). Mechanistic studies of a retaining alpha-galactosyltransferase from *Neisseria meningitidis*. *Biochemistry* 41, 5075-5085.

Ly, H.D., and Withers, S.G. (1999). Mutagenesis of glycosidases. *Annual Review of Biochemistry* 68, 487-522.

Mackenzie, L.F., Wang, Q.P., Warren, R.A.J., and Withers, S.G. (1998). Glycosynthases: Mutant glycosidases for oligosaccharide synthesis. *Journal of the American Chemical Society* 120, 5583-5584.

Macleod, A.M., Lindhorst, T., Withers, S.G., and Warren, R.A.J. (1994). The Acid/Base Catalyst in the Exoglucanase/Xylanase from *Cellulomonas-Fimi* Is Glutamic-Acid-127 - Evidence from Detailed Kinetic-Studies of Mutants. *Biochemistry* 33, 6371-6376.

MacManus, D.A., and Vulfson, E.N. (2000). Regioselectivity of enzymatic glycosylation of 6-O-acyl glycosides in supersaturated solutions. *Biotechnology and Bioengineering* 69, 585-590.

Madsen, N.B. (1986). *The Enzymes* 17, 365-394.

Maeda, Y., Watanabe, R., Harris, C.L., Hong, J.J., Ohishi, K., Kinoshita, K., and Kinoshita, T. (2001). PIG-M transfers the first mannose to glycosylphosphatidylinositol on the luminal side of the ER. *Embo Journal* 20, 250-261.

Malik, P., Tarry, T.D., Gowda, L.R., Langara, A., Petukhov, S.A., Symmons, M.F., Welsh, L.C., Marvin, D.A., and Perham, R.N. (1996). Role of capsid structure and membrane protein processing in determining the size and copy number of peptides

displayed on the major coat protein of filamentous bacteriophage. *Journal of Molecular Biology* 260, 9-21.

Mark, B.L., and James, M.N.G. (2002). Anchimeric assistance in hexosaminidases. *Canadian Journal of Chemistry-Revue Canadienne De Chimie* 80, 1064-1074.

Mark, B.L., Vocadlo, D.J., Knapp, S., Triggs-Raine, B.L., Withers, S.G., and James, M.N.G. (2001). Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase. *Journal of Biological Chemistry* 276, 10330-10337.

Martinez-Fleites, C., Proctor, M., Roberts, S., Bolam, D.N., Gilbert, H.J., and Davies, G.J. (2006). Insights into the synthesis of lipopolysaccharide and antibiotics through the structures of two retaining glycosyltransferases from family GT4. *Chemistry & Biology* 13, 1143-1152.

Mazzella, L.J., Parkin, D.W., Tyler, P.C., Furneaux, R.H., and Schramm, V.L. (1996). Mechanistic diagnoses of N-ribohydrolases and purine nucleoside phosphorylase. *Journal of the American Chemical Society* 118, 2111-2112.

Miley, M.J., Zielinska, A.K., Keenan, J.E., Bratton, S.M., Radomska-Pandya, A., and Redinbo, M.R. (2007). Crystal Structure of the Cofactor-Binding Domain of the Human Phase II Drug-Metabolism Enzyme UDP-Glucuronosyltransferase 2B7. *Journal of Molecular Biology* 369, 498-511.

Mitchell, E.P., Withers, S.G., Ermert, P., Vasella, A.T., Garman, E.F., Oikonomakos, N.G., and Johnson, L.N. (1996). Ternary complex crystal structures of glycogen phosphorylase with the transition state analogue nojirimycin tetrazole and phosphate in the T and R states. *Biochemistry* 35, 7341-7355.

Miura, Y., Arai, T., and Yamagata, T. (1996). Synthesis of amphiphilic lactosides that possess a lactosylceramide-mimicking N-acyl structure: Alternative universal substrates for endo-type glycosylceramidases. *Carbohydrate Research* 289, 193-199.

Moffatt, J.G. (1966). Sugar nucleotide synthesis by the phosphomorpholidate procedure. *Methods in Enzymology* 8, 136-142.

Moffatt, J.G., and Khorana, H.G. (1961). Nucleoside Polyphosphates. X. The synthesis and some reactions of nucleoside-5' phosphomorpholidates and related compounds. Improved methods for the preparation of nucleoside-5'polyphosphates. *Journal of the American Chemical Society* 83, 649-658.

Mollerach, M., Lopez, R., and Garcia, E. (1998). Characterization of the galU gene of *Streptococcus pneumoniae* encoding a uridine diphosphoglucose pyrophosphorylase: a

gene essential for capsular polysaccharide biosynthesis. *Journal of Experimental Medicine* *188*, 2047-2056.

Monegal, A., and Planas, A. (2006). Chemical rescue of alpha 3-galactosyltransferase. Implications in the mechanism of retaining glycosyltransferases. *Journal of the American Chemical Society* *128*, 16030-16031.

Morera, S., Lariviere, L., Kurzeck, J., Aschke-Sonnenborn, U., Freemont, P.S., Janin, J., and Ruger, W. (2001). High resolution crystal structures of T4 phage beta-galucosyltransferase: Induced fit and effect of substrate and metal binding. *Journal of Molecular Biology* *311*, 569-577.

Morrison, J.F., and Ebner, K.E. (1971). Studies on Galactosyltransferase - Kinetic Investigations with N-Acetylglucosamine as Galactosyl Group Acceptor. *Journal of Biological Chemistry* *246*, 3977-&.

Mosi, R., He, S.M., Uitdehaag, J., Dijkstra, B.W., and Withers, S.G. (1997). Trapping and characterization of the reaction intermediate in cyclodextrin glycosyltransferase by use of activated substrates and a mutant enzyme. *Biochemistry* *36*, 9927-9934.

Moss, R.A., Fu, X.L., and Sauers, R.R. (2007). S(N)<sub>i</sub> fragmentations of alkoxychlorocarbenes - a perspective. *Journal of Physical Organic Chemistry* *20*, 1-10.

Muehlbacher, M., and Poulter, C.D. (1985). Isopentenyl Diphosphate Dimethylallyl Diphosphate Isomerase - Irreversible Inhibition of the Enzyme by Active-Site-Directed Covalent Attachment. *Journal of the American Chemical Society* *107*, 8307-8308.

Muehlbacher, M., and Poulter, C.D. (1988). Isopentenyl-Diphosphate Isomerase - Inactivation of the Enzyme with Active-Site-Directed Irreversible Inhibitors and Transition-State Analogs. *Biochemistry* *27*, 7315-7328.

Mulichak, A.M., Losey, H.C., Lu, W., Wawrzak, Z., Walsh, C.T., and Garavito, R.M. (2003). Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. *Proceedings of the National Academy of Sciences of the United States of America* *100*, 9238-9243.

Mulichak, A.M., Losey, H.C., Walsh, C.T., and Garavito, R.M. (2001). Structure of the UDP-Glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics. *Structure* *9*, 547-557.

Mulichak, A.M., Lu, W., Losey, H.C., Walsh, C.T., and Garavito, R.M. (2004). Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: Interactions with acceptor and nucleotide ligands. *Biochemistry* *43*, 5170-5180.

Murray, B.W., Takayama, S., Schultz, J., and Wong, C.H. (1996). Mechanism and specificity of human alpha-1,3-fucosyltransferase V. *Biochemistry* 35, 11183-11195.

Murray, B.W., Wittmann, V., Burkart, M.D., Hung, S.C., and Wong, C.H. (1997). Mechanism of human alpha-1,3-fucosyltransferase V: Glycosidic cleavage occurs prior to nucleophilic attack. *Biochemistry* 36, 823-831.

Nerinckx, W., Desmet, T., and Claeysens, M. (2003). A hydrophobic platform as a mechanistically relevant transition state stabilising factor appears to be present in the active centre of all glycoside hydrolases. *Febs Letters* 538, 1-7.

Nguyen, H.P., Seto, N.O.L., Cai, Y., Leinala, E.K., Borisova, S.N., Palcic, M.M., and Evans, S.V. (2003). The influence of an intramolecular hydrogen bond in differential recognition of inhibitory acceptor analogs by human ABO(H) blood group A and B glycosyltransferases. *Journal of Biological Chemistry* 278, 49191-49195.

Ni, L., Chokhawala, H.A., Cao, H., Henning, R., Ng, L., Huang, S., Yu, H., Chen, X., and Fisher, A.J. (2007). Crystal Structures of *Pasteurella multocida* Sialyltransferase Complexes with Acceptor and Donor Analogues Reveal Substrate Binding Sites and Catalytic Mechanism. *Biochemistry* 46, 6288-6298.

Ni, L.S., Sung, M.C., Yu, H., Chokhawala, H., Chen, X., and Fisher, A.J. (2006). Cytidine 5'-monophosphate (CMP)-induced structural changes in a multifunctional sialyltransferase from *Pasteurella multocida*. *Biochemistry* 45, 2139-2148.

Nishikawa, Y., Pegg, W., Paulsen, H., and Schachter, H. (1988). Control of Glycoprotein-Synthesis .15. Purification and Characterization of Rabbit Liver Udp-N-Acetylglucosamine-Alpha-3-D-Mannoside Beta-1,2-N-Acetylglucosaminyltransferase-I. *Journal of Biological Chemistry* 263, 8270-8281.

O'Reilly, M., Watson, K.A., and Johnson, L.N. (1999). The crystal structure of the *Escherichia coli* maltodextrin phosphorylase- $\alpha$ -carboase complex. *Biochemistry* 38, 5337-5345.

O'Reilly, M., Watson, K.A., Schinzel, R., Palm, D., and Johnson, L.N. (1997). Oligosaccharide substrate binding in *Escherichia coli* maltodextrin phosphorylase. *Nature Structural Biology* 4, 405-412.

Oberthur, M., Leimkuhler, C., Kruger, R.G., Lu, W., Walsh, C.T., and Kahne, D. (2005). A systematic investigation of the synthetic utility of glycopeptide glycosyltransferases. *Journal of the American Chemical Society* 127, 10747-10752.

Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B.G., Tarling, C.A., Ford, C.M., Bowles, D.J., and Davies, G.J. (2006). Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *Embo Journal* 25, 1396-1405.

Ogata, S., Ho, I., Chen, A.L., Dubois, D., Maklansky, J., Singhal, A., Hakomori, S., and Itzkowitz, S.H. (1995). Tumor-Associated Sialylated Antigens Are Constitutively Expressed in Normal Human Colonic Mucosa. *Cancer Research* 55, 1869-1874.

Ohrlein, R. (1999). Glycosyltransferase-catalyzed synthesis of non-natural oligosaccharides. In *Biocatalysis - from Discovery to Application*, pp. 227-254.

Ohru, H., Meguro, H., Ioa, T. (1988). N-acetyl-3-fluoro-neuraminic acid derivatives and preparation thereof, E. Patent, ed.

Ohtsubo, K., Imajo, S., Ishiguro, M., Nakatani, T., Oka, S., and Kawasaki, T. (2000). Studies on the structure-function relationship of the HNK-1 associated glucuronyltransferase, GlcAT-P, by computer modeling and site-directed mutagenesis. *Journal of Biochemistry* 128, 283-291.

Ouzzine, M., Gulberti, S., Levoine, N., Netter, P., Magdalou, J., and Fournel-Gigleux, S. (2002). The donor substrate specificity beta 1,3-glucuronosyltransferase I toward UDP-glucuronic acid is determined by two crucial histidine and arginine residues. *Journal of Biological Chemistry* 277, 25439-25445.

Pak, J.E., Arnoux, P., Zhou, S.H., Sivarajah, P., Satkunarajah, M., Xing, X.K., and Rini, J.M. (2006). X-ray crystal structure of leukocyte type core 2 beta 1,6-N-acetylglucosaminyltransferase - Evidence for a convergence of metal ion-independent glycosyltransferase mechanism. *Journal of Biological Chemistry* 281, 26693-26701.

Patenaude, S.I., Seto, N.O.L., Borisova, S.N., Szpacenko, A., Marcus, S.L., Palcic, M.M., and Evans, S.V. (2002). The structural basis for specificity in human ABO(H) blood group biosynthesis. *Nature Structural Biology* 9, 685-690.

Pedersen, H., Holder, S., Sutherlin, D.P., Schwitter, U., King, D.S., and Schultz, P.G. (1998). A method for directed evolution and functional cloning of enzymes. *Proceedings of the National Academy of Sciences of the United States of America* 95, 10523-10528.

Pedersen, L.C., Darden, T.A., and Negishi, M. (2002). Crystal structure of beta 1,3-glucuronyltransferase I in complex with active donor substrate UDP-GlcUA. *Journal of Biological Chemistry* 277, 21869-21873.

Pedersen, L.C., Dong, J., Taniguchi, F., Kitagawa, H., Krahn, J.M., Pedersen, L.G., Sugahara, K., and Negishi, M. (2003). Crystal structure of an alpha 1,4-N-acetylhexosaminyltransferase (EXTL2), a member of the exostosin gene family involved in heparan sulfate biosynthesis. *Journal of Biological Chemistry* 278, 14420-14428.

Pedersen, L.C., Tsuchida, K., Kitagawa, H., Sugahara, K., Darden, T.A., and Negishi, M. (2000). Heparan/chondroitin sulfate biosynthesis - Structure and mechanism of human glucuronyltransferase I. *Journal of Biological Chemistry* 275, 34580-34585.

Penner, J.L., and Aspinall, G.O. (1997). Diversity of lipopolysaccharide structures in *Campylobacter jejuni*. *Journal of Infectious Diseases* 176, S135-S138.

Persson, K., Ly, H.D., Dieckelmann, M., Wakarchuk, W.W., Withers, S.G., and Strynadka, N.C.J. (2001). Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nature Structural Biology* 8, 166-175.

Persson, M., Letts, J.A., Hosseini-Maaf, B., Borisova, S.N., Palcic, M.M., Evans, S.V., and Olsson, M.L. (2007). Structural effects of naturally occurring human blood group B galactosyltransferase mutations adjacent to the DXD motif. *Journal of Biological Chemistry* 282, 9564-9570.

Phillips, D.C. (1966). The three-dimensional structure of an enzyme molecule. *Scientific American* 215, 78-90.

Phillips, D.C. (1967). The Hen Egg White Lysozyme. *Proceedings of the National Academy of Sciences of the United States of America* 57, 484-495.

Piskiewicz, D. (1977). *Kinetics of Chemical and Enzyme-Catalyzed Reactions* (New York, Oxford University Press).

Poulter, C.D., Argyle, J.C., and Mash, E.A. (1977). Prenyltransferase - New Evidence for an "Ionization-Condensation-Elimination Mechanism with 2-Fluorogeranyl Pyrophosphate. *Journal of the American Chemical Society* 99, 957-959.

Poulter, C.D., Argyle, J.C., and Mash, E.A. (1978). Farnesyl Pyrophosphate Synthetase - Mechanistic Studies of 1'4 Coupling Reaction with 2-Fluorogeranyl Pyrophosphate. *Journal of Biological Chemistry* 253, 7227-7233.

Poulter, C.D., and Satterwhite, D.M. (1977). Mechanism of Prenyl-Transfer Reaction - Studies with (E)-3-Trifluoromethyl-2-Buten-1-Yl and (Z)-3-Trifluoromethyl-2-Buten-1-Yl Pyrophosphate. *Biochemistry* 16, 5470-5478.

Powers, T., and Walter, P. (1995). Reciprocal Stimulation of Gtp Hydrolysis by 2 Directly Interacting Gtpases. *Science* 269, 1422-1424.

Quiros, L.M., and Salas, J.A. (1995). Biosynthesis of the Macrolide Oleandomycin by *Streptomyces-Antibioticus* - Purification and Kinetic Characterization of an Oleandomycin Glucosyltransferase. *Journal of Biological Chemistry* 270, 18234-18239.

Rademacher, T.W., Parekh, R.B., and Dwek, R.A. (1988). Glycobiology. *Annual Review of Biochemistry* 57, 785-838.

Raibaud, O., and Schwartz, M. (1984). Positive Control of Transcription Initiation in Bacteria. *Annual Review of Genetics* 18, 173-206.

Ramakrishnan, B., Balaji, P.V., and Qasba, P.K. (2002). Crystal structure of beta 1,4-galactosyltransferase complex with UDP-Gal reveals an oligosaccharide acceptor binding site. *Journal of Molecular Biology* 318, 491-502.

Ramakrishnan, B., Boeggeman, E., and Qasba, P.K. (2005). Mutation of arginine 228 to lysine enhances the glucosyltransferase activity of bovine beta-1,4-galactosyltransferase I. *Biochemistry* 44, 3202-3210.

Ramakrishnan, B., and Qasba, P.K. (2001). Crystal structure of lactose synthase reveals a large conformational change in its catalytic component, the beta 1,4-galactosyltransferase-1. *Journal of Molecular Biology* 310, 205-218.

Ramakrishnan, B., and Qasba, P.K. (2002). Structure-based design of beta 1,4-galactosyltransferase I (beta 4Gal-T1) with equally efficient N-acetylgalactosaminyltransferase activity - Point mutation broadens beta 4Gal-T1 donor specificity. *Journal of Biological Chemistry* 277, 20833-20839.

Ramakrishnan, B., Ramasamy, V., and Qasba, P.K. (2006). Structural snapshots of beta-1,4-galactosyltransferase-1 along the kinetic pathway. *Journal of Molecular Biology* 357, 1619-1633.

Ramakrishnan, B., Shah, P.S., and Qasba, P.K. (2001). alpha-Lactalbumin (LA) stimulates milk beta-1,4-galactosyltransferase I (beta 4Gal-T1) to transfer glucose from UDP-glucose to N-acetylglucosamine - Crystal structure of beta 4Gal-T1-LA complex with UDP-Glc. *Journal of Biological Chemistry* 276, 37665-37671.

Ramasamy, V., Ramakrishnan, B., Boeggeman, E., Ratner, D.M., Seeberger, P.H., and Qasba, P.K. (2005). Oligosaccharide preferences of beta 1,4-galactosyltransferase-I: Crystal structures of Met340His mutant of human beta 1,4-galactosyltransferase-I with a

pentasaccharide and trisaccharides of the N-glycan moiety. *Journal of Molecular Biology* 353, 53-67.

Rapoza, M.P., and Webster, R.E. (1995). The Products of Gene I and the Overlapping in-Frame Gene XI are Required for Filamentous Phage Assembly. *Journal of Molecular Biology* 248, 627-638.

Rath, V.L., Ammirati, M., Danley, D.E., Ekstrom, J.L., Gibbs, E.M., Hynes, T.R., Mathiowetz, A.M., McPherson, R.K., Olson, T.V., Treadway, J.L., *et al.* (2000a). Human liver glycogen phosphorylase inhibitors bind at a new allosteric site. *Chemistry & Biology* 7, 677-682.

Rath, V.L., Ammirati, M., LeMotte, P.K., Fennell, K.F., Mansour, M.N., Danley, D.E., Hynes, T.R., Schulte, G.K., Wasilko, D.J., and Pandit, J. (2000b). Activation of human liver glycogen phosphorylase by alteration of the secondary structure and packing of the catalytic core. *Molecular Cell* 6, 139-148.

Reardon, J.E., and Abeles, R.H. (1986). Mechanism of Action of Isopentenyl Pyrophosphate Isomerase - Evidence for a Carbonium-Ion Intermediate. *Biochemistry* 25, 5609-5616.

Reinert, D.J., Jank, T., Aktories, K., and Schulz, G.E. (2005). Structural basis for the function of *Clostridium difficile* toxin B. *Journal of Molecular Biology* 351, 973-981.

Rudolph, F.B., and Fromm, H.J. (1971). Computer Simulation Studies with Yeast Hexokinase and Additional Evidence for Random Bi-Bi Mechanism. *Journal of Biological Chemistry* 246, 6611-&.

Rye, C.S., and Withers, S.G. (2002). Elucidation of the mechanism of polysaccharide cleavage by chondroitin AC lyase from *Flavobacterium heparinum*. *Journal of the American Chemical Society* 124, 9756-9767.

Sandvig, K., and Vandeurs, B. (1994). Endocytosis and Intracellular Sorting of Ricin and Shiga Toxin. *Febs Letters* 346, 99-102.

Schreiner, P.R., Schleyer, P.v.R., and Hill, R.K. (1993). Reinvestigation of the S<sub>N</sub>i reaction. The ionization of chlorosulfites. *Journal of Organic Chemistry* 58, 2822-2829.

Schreiner, P.R., Schleyer, P.v.R., and Hill, R.K. (1994). Mechanisms of front-side substitutions. The transition states for the S<sub>N</sub>i decomposition of methyl and ethyl chlorosulfite in the gas phase and in solution. *Journal of Organic Chemistry* 59, 1849-1854.

Segel, I.H. (1975). *Enzyme Kinetics: Behaviour and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems* (New York, John Wiley & Sons, Inc.).

Shao, H., He, X.Z., Achnine, L., Blount, J.W., Dixon, R.A., and Wang, X.Q. (2005). Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. *Plant Cell* *17*, 3141-3154.

Shishova, E.Y., Di Costanzo, L., Cane, D.E., and Christianson, D.W. (2007). X-ray crystal structure of aristolochene synthase from *Aspergillus terreus* and evolution of templates for the cyclization of farnesyl diphosphate. *Biochemistry* *46*, 1941-1951.

Sidhu, S.S., Weiss, G.A., and Wells, J.A. (2000). High copy display of large proteins on phage for functional selections. *Journal of Molecular Biology* *296*, 487-495.

Simon, E.S., Bednarski, M.D., and Whitesides, G.M. (1988). Synthesis of Cmp-Neuac from N-Acetylglucosamine - Generation of Ctp from Cmp Using Adenylate Kinase. *Journal of the American Chemical Society* *110*, 7159-7163.

Sinnott, M.L., and Jencks, W.P. (1980). Solvolysis of D-glucopyranosyl derivatives in mixtures of ethanol and 2,2,2-trifluoroethanol. *Journal of the American Chemical Society* *102*, 2026-2032.

Sinnott, M.L., and Souchard, I.J. (1973). Mechanism of Action of Beta-Galactosidase - Effect of Aglycone Nature and Alpha-Deuterium Substitution on Hydrolysis of Aryl Galactosides. *Biochemical Journal* *133*, 89-98.

Skarzynski, T., Kim, D.H., Lees, W.J., Walsh, C.T., and Duncan, K. (1998). Stereochemical course of enzymatic enolpyruvyl transfer and catalytic conformation of the active site revealed by the crystal structure of the fluorinated analogue of the reaction tetrahedral intermediate bound to the active site of the C115A mutant of MurA. *Biochemistry* *37*, 2572-2577.

Smith, G.P. (1985). Filamentous Fusion Phage - Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science* *228*, 1315-1317.

Smith, G.P. (1993). Surface Display and Peptide Libraries - Articles Occasioned by a Meeting at the Banbury-Center, Cold Spring Harbor Laboratory, April 4-7, 1992 - Preface. *Gene* *128*, 1-2.

Smith, G.P., and Petrenko, V.A. (1997). Phage display. *Chemical Reviews* *97*, 391-410.

Smith, G.P., and Scott, J.K. (1993). Libraries of Peptides and Proteins Displayed on Filamentous Phage. *Methods in Enzymology* *217*, 228-257.

Sommer, N., Depping, R., Piotrowski, M., and Ruger, W. (2004). Bacteriophage T4 alpha-glucosyltransferase: a novel interaction with gp45 and aspects of the catalytic mechanism. *Biochemical and Biophysical Research Communications* 323, 809-815.

St. Hilaire, P.M., Boyd, M.K., and Toone, E.J. (1994). Interaction of the Shiga-like Toxin Type 1 B-Subunit with Its Carbohydrate Receptor. *Biochemistry* 33, 14452-14463.

Starks, C.M., Back, K.W., Chappell, J., and Noel, J.P. (1997). Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science* 277, 1815-1820.

Stick, R.V., Stubbs, K.A., and Watts, A.G. (2004). Modifying the regioselectivity of glycosynthase reactions through changes in the acceptor. *Australian Journal of Chemistry* 57, 779-786.

Stick, R.V., and Watts, A.G. (2002). The chameleon of retaining glycoside hydrolases and retaining glycosyl transferases: The catalytic nucleophile. *Monatshefte Fur Chemie* 133, 541-554.

Stirtan, W.G., and Withers, S.G. (1996). Phosphonate and alpha-fluorophosphonate analogue probes of the ionization state of pyridoxal 5'-phosphate (PLP) in glycogen phosphorylase. *Biochemistry* 35, 15057-15064.

Strahlbolsinger, S., Immervoll, T., Deutzmann, R., and Tanner, W. (1993). Pmt1, the Gene for a Key Enzyme of Protein O-Glycosylation in *Saccharomyces-Cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* 90, 8164-8168.

Strobel, H., Ladant, D., and Jestin, J.L. (2003). In vitro selection for enzymatic activity: A model study using adenylate cyclase. *Journal of Molecular Biology* 332, 1-7.

Sujino, K., Uchiyama, T., Hindsgaul, O., Seto, N.O.L., Wakarchuk, W.W., and Palcic, M.M. (2000). Enzymatic synthesis of oligosaccharide analogues: UDP-Gal analogues as donors for three retaining alpha-galactosyltransferases. *Journal of the American Chemical Society* 122, 1261-1269.

Sun, H.Y., Lin, S.W., Ko, T.P., Pan, J.F., Liu, C.L., Lin, C.N., Wang, A.H.J., and Lin, C.H. (2007). Structure and mechanism of *Helicobacter pylori* fucosyltransferase - A basis for lipopolysaccharide variation and inhibitor design. *Journal of Biological Chemistry* 282, 9973-9982.

Takahashi, M., Inoue, N., Ohishi, K., Maeda, Y., Nakamura, N., Endo, Y., Fujita, T., Takeda, J., and Kinoshita, T. (1996). PIG-B, a membrane protein of the endoplasmic

reticulum with a large luminal domain, is involved in transferring the third mannose of the GPI anchor. *Embo Journal* 15, 4254-4261.

Tarbouriech, N., Charnock, S.J., and Davies, G.J. (2001). Three-dimensional structures of the Mn and Mg dTDP complexes of the family GT-2 glycosyltransferase SpsA: A comparison with related NDP-sugar glycosyltransferases. *Journal of Molecular Biology* 314, 655-661.

Tawfik, D.S., and Griffiths, A.D. (1998). Man-made cell-like compartments for molecular evolution. *Nature Biotechnology* 16, 652-656.

Ten Hagen, K.G., Bedi, G.S., Tetaert, D., Kingsley, P.D., Hagen, F.K., Balys, M.M., Beres, T.M., Degand, P., and Tabak, L.A. (2001). Cloning and characterization of a ninth member of the UDP-GalNAc : polypeptide N-acetylgalactosaminyltransferase family, ppGaNTase-T9. *Journal of Biological Chemistry* 276, 17395-17404.

Ten Hagen, K.G., Tetaert, D., Hagen, F.K., Richet, C., Beres, T.M., Gagnon, J., Balys, M.M., VanWuyckhuysse, B., Bedi, G.S., Degand, P., *et al.* (1999). Characterization of a UDP-GalNAc : polypeptide N-acetylgalactosaminyltransferase that displays glycopeptide N-acetylgalactosaminyltransferase activity. *Journal of Biological Chemistry* 274, 27867-27874.

Thoden, J.B., Gulick, A.M., and Holden, H.M. (1997). Molecular structures of the S124A, S124T, and S124V site-directed mutants of UDP-galactose 4-epimerase from *Escherichia coli*. *Biochemistry* 36, 10685-10695.

Thompson, J., Pikiš, A., Ruvinov, S.B., Henrissat, B., Yamamoto, H., and Sekiguchi, J. (1998). The gene *glvA* of *Bacillus subtilis* 168 encodes a metal-requiring, NAD (H)-dependent 6-phospho- $\alpha$ -glucosidase - Assignment to family 4 of the glycosylhydrolase superfamily. *Journal of Biological Chemistry* 273, 27347-27356.

Ubersax, J.A., Woodbury, E.L., Quang, P.N., Paraz, M., Blethrow, J.D., Shah, K., Shokat, K.M., and Morgan, D.O. (2003). Targets of the cyclin-dependent kinase Cdk1. *Nature* 425, 859-864.

Unligil, U.M., and Rini, J.M. (2000). Glycosyltransferase structure and mechanism. *Current Opinion in Structural Biology* 10, 510-517.

Unligil, U.M., Zhou, S.H., Yuwaraj, S., Sarkar, M., Schachter, H., and Rini, J.M. (2000). X-ray crystal structure of rabbit N-acetylglucosaminyltransferase I: catalytic mechanism and a new protein superfamily. *Embo Journal* 19, 5269-5280.

Valenzuelasoto, E.M., and Munozclares, R.A. (1993). Betaine-Aldehyde Dehydrogenase from Leaves of *Amaranthus-Hypochondriacus* L Exhibits an Iso Ordered Bi Bi Steady-State Mechanism. *Journal of Biological Chemistry* 268, 23818-23824.

Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., and Marth, J. (1999). *Essentials of Glycobiology* (New York, Cold Spring Harbor Press).

Vartanian, J.P., Henry, M., and Wain-Hobson, S. (2001). Simulating pseudogene evolution in vitro: Determining the true number of mutations in a lineage. *Proceedings of the National Academy of Sciences of the United States of America* 98, 13172-13176.

Viladot, J.L., de Ramon, E., Durany, O., and Planas, A. (1998). Probing the mechanism of *Bacillus* 1,3-1,4-beta-D-glucan 4-glucanohydrolases by chemical rescue of inactive mutants at catalytically essential residues. *Biochemistry* 37, 11332-11342.

Vocadlo, D.J., Davies, G.J., Laine, R., and Withers, S.G. (2001). Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* 412, 835-838.

Vocadlo, D.J., Wicki, J., Rupitz, K., and Withers, S.G. (2002). Mechanism of *Thermoanaerobacterium saccharolyticum* ss-xylosidase: Kinetic studies. *Biochemistry* 41, 9727-9735.

Vrielink, A., Ruger, W., Driessen, H.P.C., and Freemont, P.S. (1994). Crystal-Structure of the DNA Modifying Enzyme Beta-Glucosyltransferase in the Presence and Absence of the Substrate Uridine Diphosphoglucose. *Embo Journal* 13, 3413-3422.

Vyas, A.A., and Schnaar, R.L. (2001). Brain gangliosides: Functional ligands for myelin stability and the control of nerve regeneration. *Biochimie* 83, 677-682.

Watkins, W.M., and Morgan, W.T.J. (1957). Specific Inhibition Studies Relating to the Lewis Blood-Group System. *Nature* 180, 1038-1041.

Watson, K.A., McCleverty, C., Geremia, S., Cottaz, S., Driguez, H., and Johnson, L.N. (1999). Phosphorylase recognition and phosphorolysis of its oligosaccharide substrate: answers to a long outstanding question. *Embo Journal* 18, 4619-4632.

Watson, K.A., Mitchell, E.P., Johnson, L.N., Son, J.C., Bichard, C.J.F., Orchard, M.G., Fleet, G.W.J., Oikonomakos, N.G., Leonidas, D.D., Kontou, M., *et al.* (1994). Design of Inhibitors of Glycogen-Phosphorylase - a Study of Alpha-C-Glucosides and Beta-C-Glucosides and 1-Thio-Beta-D-Glucose Compounds. *Biochemistry* 33, 5745-5758.

Watts, A.G., Damager, I., Amaya, M.L., Buschiazzi, A., Alzari, P., Frasch, A.C., and Withers, S.G. (2003). *Trypanosoma cruzi* trans-sialidase operates through a covalent

sialyl-enzyme intermediate: Tyrosine is the catalytic nucleophile. *Journal of the American Chemical Society* *125*, 7532-7533.

Weber, P.C., Ohlendorf, D.H., Wendoloski, J.J., and Salemme, F.R. (1989). Structural Origins of High-Affinity Biotin Binding to Streptavidin. *Science* *243*, 85-88.

Weijland, A., and Parmeggiani, A. (1993). Toward a Model for the Interaction between Elongation Factor-Tu and the Ribosome. *Science* *259*, 1311-1314.

Whiteson, K.L., Chen, Y., Chopra, N., Raymond, A.C., and Rice, P.A. (2007). Identification of a potential general acid/base in the reversible phosphoryl transfer reactions catalyzed by tyrosine recombinases: Flp H305. *Chemistry & Biology* *14*, 121-129.

Wiggins, C.A.R., and Munro, S. (1998). Activity of the yeast MNN1 alpha-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proceedings of the National Academy of Sciences of the United States of America* *95*, 7945-7950.

Williams, S.J., Hekmat, O., and Withers, S.G. (2006). Synthesis and testing of mechanism-based protein-profiling probes for retaining endo-glycosidases. *Chembiochem* *7*, 116-124.

Winstein, S., Clippinger, E., Fainberg, A.H., Heck, R., and Robinson, G.C. (1956). Salt Effects and Ion Pairs in Solvolysis and Related Reactions. III.1 Common Ion Rate Depression and Exchange of Anions during Acetolysis<sup>2,3</sup>. *Journal of the American Chemical Society* *78*, 328-335.

Withers, S.G., and Aebersold, R. (1995). Approaches to Labeling and Identification of Active-Site Residues in Glycosidases. *Protein Science* *4*, 361-372.

Withers, S.G., Madsen, N.B., Sykes, B.D., Takagi, M., Shimomura, S., and Fukui, T. (1981). Evidence for Direct Phosphate-Phosphate Interaction between Pyridoxal-Phosphate and Substrate in the Glycogen-Phosphorylase Catalytic Mechanism. *Journal of Biological Chemistry* *256*, 10759-10762.

Withers, S.G., Shechosky, S., and Madsen, N.B. (1982). Pyridoxal-Phosphate Is Not the Acid Catalyst in the Glycogen-Phosphorylase Catalytic Mechanism. *Biochemical and Biophysical Research Communications* *108*, 322-328.

Wittmann, V., and Wong, C.H. (1997). 1H-tetrazole as catalyst in phosphomorpholidate coupling reactions: Efficient synthesis of GDP-fucose, GDP-mannose, and UDP-galactose. *Journal of Organic Chemistry* *62*, 2144-2147.

Wolfenden, R., Lu, X.D., and Young, G. (1998). Spontaneous hydrolysis of glycosides. *Journal of the American Chemical Society* *120*, 6814-6815.

Wong, C.H., and Hayashi, T. (1998). Process for preparing nucleotide inhibitors of glycosyltransferases, U.S.P. Office, ed. (United States).

Woodward, R.B., and Hoffman, R. (1969). The Conservation of Orbital Symmetry. *Angewandte Chemie-International Edition* *8*, 781-853.

Wragg, S., Hagen, F.K., and Tabak, L.A. (1995). Kinetic-Analysis of a Recombinant Udp-N-Acetyl-D-Galactosamine - Polypeptide N-Acetylgalactosaminyltransferase. *Journal of Biological Chemistry* *270*, 16947-16954.

Wright, S.W., Rath, V.L., Genereux, P.E., Hageman, D.L., Levy, C.B., McClure, L.D., McCoid, S.C., McPherson, R.K., Schelhorn, T.M., Wilder, D.E., *et al.* (2005). 5-Chloroindoloyl glycine amide inhibitors of glycogen phosphorylase: synthesis, in vitro, in vivo, and X-ray crystallographic characterization. *Bioorganic & Medicinal Chemistry Letters* *15*, 459-465.

Xia, G., Chen, L.J., Sera, T., Fa, M., Schultz, P.G., and Romesberg, F.E. (2002). Directed evolution of novel polymerase activities: Mutation of a DNA polymerase into an efficient RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* *99*, 6597-6602.

Yamamoto, F., Clausen, H., White, T., Marken, J., and Hakomori, S.I. (1990). Molecular Genetic-Basis of the Histo-Blood Group Abo System. *Nature* *345*, 229-233.

Yang, M., Brazier, M., Edwards, R., and Davis, B.G. (2005). High-throughput mass-spectroscopy monitoring for multisubstrate enzymes: Determining the kinetic parameters and catalytic activities of glycosyltransferases. *Chembiochem* *6*, 346-357.

Yin, J., Mills, J.H., and Schultz, P.G. (2004). A catalysis-based selection for peroxidase antibodies with increased activity. *Journal of the American Chemical Society* *126*, 3006-3007.

Yip, V.L.Y., Varrot, A., Davies, G.J., Rajan, S.S., Yang, X.J., Thompson, J., Anderson, W.F., and Withers, S.G. (2004). An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 beta-glycosidase from *Thermotoga maritima*. *Journal of the American Chemical Society* *126*, 8354-8355.

Yogeeswaran, G., and Salk, P.L. (1981). Metastatic Potential Is Positively Correlated with Cell-Surface Sialylation of Cultured Murine Tumor-Cell Lines. *Science* *212*, 1514-1516.

Yu, H., Chokhawala, H., Karpel, R., Yu, H., Wu, B.Y., Zhang, J.B., Zhang, Y.X., Jia, Q., and Chen, X. (2005). A multifunctional *Pasteurella multocida* sialyltransferase: A powerful tool for the synthesis of sialoside libraries. *Journal of the American Chemical Society* 127, 17618-17619.

Yuan, Y.Q., Barrett, D., Zhang, Y., Kahne, D., Sliz, P., and Walker, S. (2007). Crystal structure of a peptidoglycan glycosyltransferase suggests a model for processive glycan chain synthesis. *Proceedings of the National Academy of Sciences of the United States of America* 104, 5348-5353.

Zechel, D.L., and Withers, S.G. (2000). Glycosidase mechanisms: Anatomy of a finely tuned catalyst. *Accounts of Chemical Research* 33, 11-18.

Zhang, C., Kenski, D.M., Paulson, J.L., Bonshtien, A., Sessa, G., Cross, J.V., Templeton, D.J., and Shokat, K.M. (2005). A second-site suppressor strategy for chemical genetic analysis of diverse protein kinases. *Nature Methods* 2, 435-441.

Zhang, C.S., Griffith, B.R., Fu, Q., Albermann, C., Fu, X., Lee, I.K., Li, L.J., and Thorson, J.S. (2006a). Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions. *Science* 313, 1291-1294.

Zhang, Y.H., Ginsberg, C., Yuan, Y.Q., and Walker, S. (2006b). Acceptor substrate selectivity and kinetic mechanism of *Bacillus subtilis* TagA. *Biochemistry* 45, 10895-10904.

Zhang, Y.N., Swaminathan, G.J., Deshpande, A., Boix, E., Natesh, R., Xie, Z.H., Acharya, K.R., and Brew, K. (2003). Roles of individual enzyme-substrate interactions by alpha-1,3-galactosyltransferase in catalysis and specificity. *Biochemistry* 42, 13512-13521.

## Appendix A – IUPAC Recommended Nomenclature for Reaction Mechanisms

Symbol	Placement	Meaning
A	on the line	bond making (association)
D	on the line	bond breaking (dissociation)
+	on the line	stepwise process
*	on the line	same as +, but the intermediate is short lived
E	subscript	electrophilic at core atom
N	subscript	nucleophilic at core atom
R	subscript	homolytic at core atom
e	subscript	same as E, at a peripheral atom
n	subscript	same as N, at a peripheral atom
r	subscript	same as R, at a peripheral atom
H	subscript	same as E, with hydrogen as electrophile
h	subscript	same as H, at a peripheral atom
xh	subscript	bond making or breaking between hydrogen and hydrogen carrier reagent atom
C	on the line	diffusional combination
P	on the line	diffusional separation
int	subscript	molecules or ions weakly complexed; intimate ion pairs or equivalent pairs of uncharged molecules
ss	subscript	solvent-separated ion pairs or equivalent pairs of uncharged molecules
‡	subscript	preceding step rate limiting
{ }	on the line	repeated sequence

## Appendix B – Graphical Representation of Kinetic Data

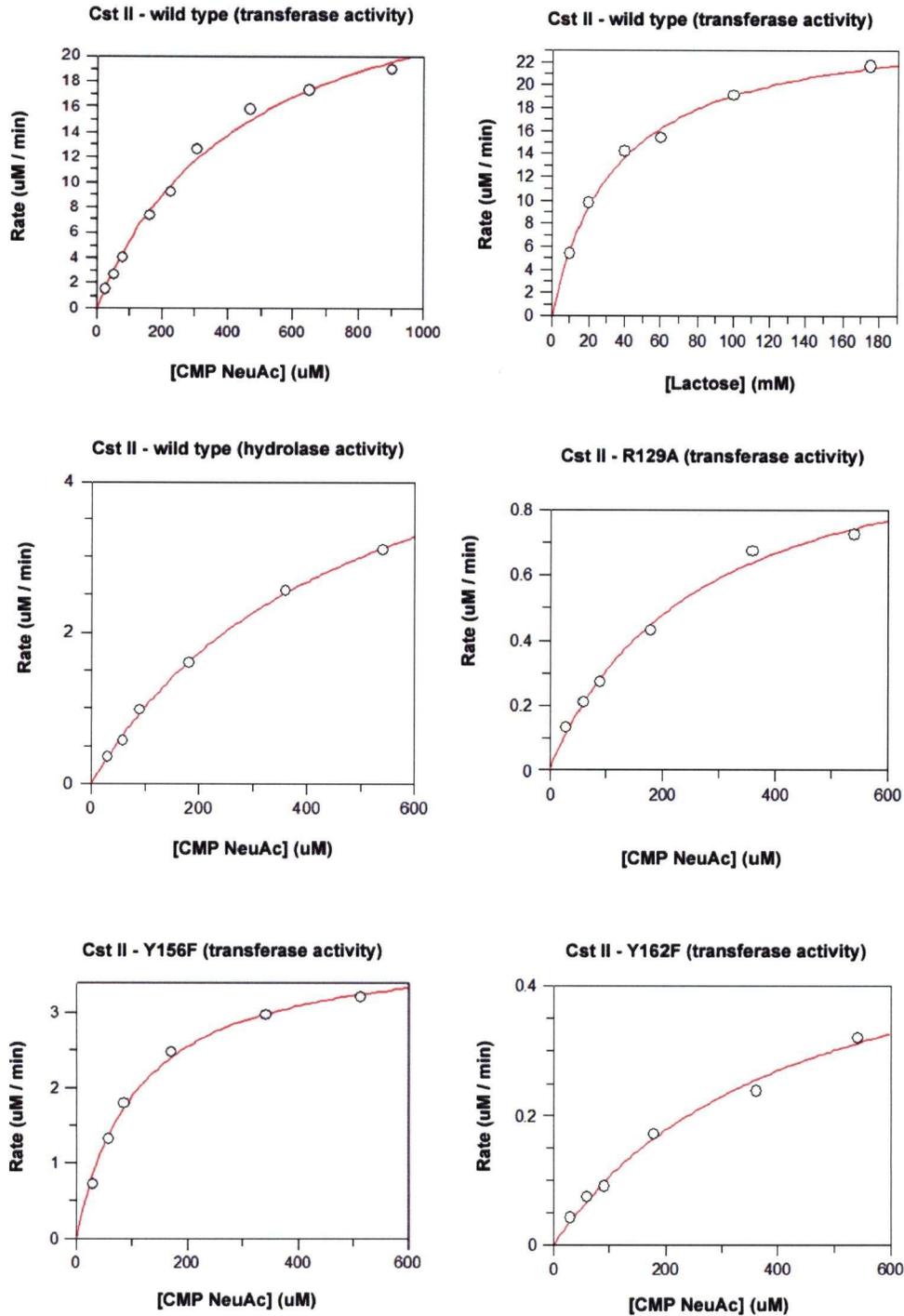
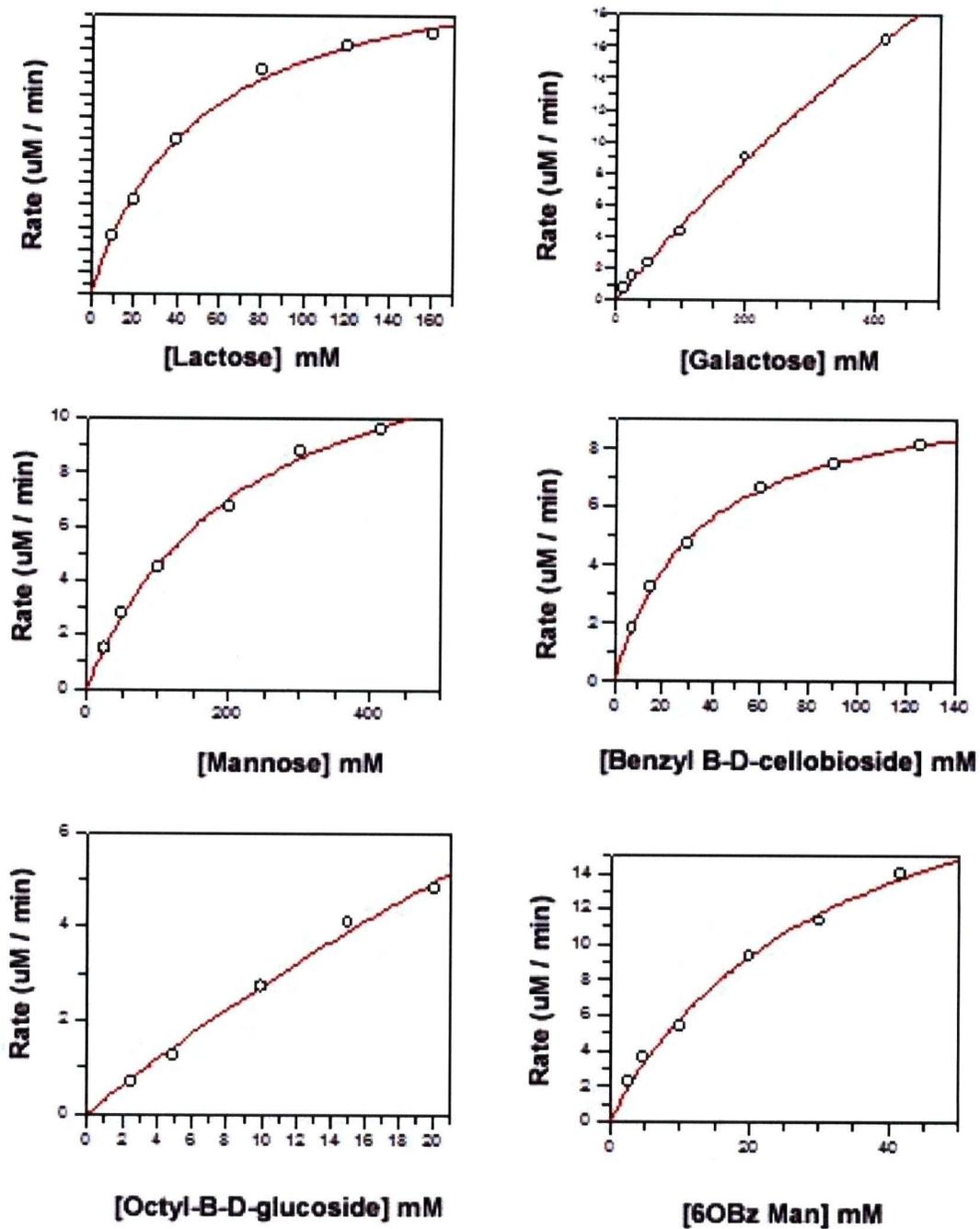


Figure A.1. Michaelis-Menten plots of various Cst II activities.



**Figure A.2.** Michaelis-Menten plots of wild-type LgtC activity with various acceptor substrates.