DEVELOPMENT OF CHEMICAL TOOLS FOR STUDYING TWO UNUSUAL FORMS OF PROTEIN *O*-GLYCOSYLATION

by

Garrett E. Whitworth
B.Sc., University College of the Cariboo, 2003

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

DOCTOR OF PHILOSOPHY

In the
Department of Chemistry
Faculty of Science

SIMON FRASER UNIVERSITY

Fall 2011

APPROVAL

Name: **Garrett E. Whitworth**

Degree: **Doctor of Philosophy**

Title of Thesis: **Development of chemical tools for studying two unusual forms of protein *O*-glycosylation.**

Examining Committee:

Chair: **Dr. Steven Holdcroft**
Professor

Dr. David Vocadlo
Senior Supervisor
Professor, Department of Chemistry

Dr. Erika Plettner
Supervisor
Professor, Department of Chemistry

Dr. Andrew Bennet
Supervisor
Professor, Department of Chemistry

Dr. Tim Storr
Internal Examiner
Assistant Professor
Department of Chemistry

Dr. Mark Nitz
External Examiner
Professor, Department of Chemistry
University of Toronto

Date Defended/Approved: 5 December 2011

Partial Copyright Licence

ABSTRACT

The development of tools, techniques and methods has been important to the evolution of glycobiology and the elucidation of the function of many forms of protein glycosylation. *O*-GlcNAcylation of nucleocytoplasmic proteins and *O*-glucosylation of Notch are two unusual forms of protein *O*-glycosylation that require further investigation in order to comprehend their biological roles. The *O*-GlcNAc modification of proteins is a dynamic process, catalyzed by *O*-GlcNAc transferase, which installs the *O*-GlcNAc modification, and *O*-GlcNAcase, a glycosidase that cleaves *O*-GlcNAc from proteins. *O*-GlcNAc modification of proteins has been implicated in various disease states such as diabetes and Alzheimer's disease. Unfortunately, many of the studies that link *O*-GlcNAcylation of proteins to these disease states use inhibitors of *O*-GlcNAcase that also affect functionally-related enzymes. In this thesis, an *O*-GlcNAcase selective inhibitor termed NButGT was developed as a tool to study the biological role of *O*-GlcNAcase without affecting the functionally related enzymes. A transition state analysis of two potent inhibitors of *O*-GlcNAcase was also undertaken to elucidate the likeness of these inhibitors to the transition state of the substrate during *O*-GlcNAcase catalyzed cleavage of *O*-GlcNAc. This study will influence the design of future generations of *O*-GlcNAcase inhibitors. *O*-Glucosylation, the second unusual form of protein *O*-glycosylation that was investigated in this thesis, modifies Notch a key transmembrane protein of the developmentally essential Notch signalling pathway. *O*-Glucosylation of Notch and elongation to a trisaccharide plays a role in Notch structure and function. Complete characterization of this trisaccharide is necessary to further our understanding of the role that it plays in the Notch signaling pathway. Di- and trisaccharide standards were synthesized, an α-xylosidase that could cleave the trisaccharide was identified, and a new capillary electrophoresis method for

identifying this *O*-glycan on proteins was developed. These new tools enabled us to unequivocally assign the structure of this trisaccharide as D-Xyl-*α*1-3-D-Xyl-*α*1-3-D-Glc. After establishing the identity of the trisaccharide modifying Notch, we incorporated the trisaccharide into a peptide for use as an antigen to raise polyclonal antibodies. This antigen along with the other new chemical tools and methodologies that were developed will prove valuable in future work directed towards clarifying the roles of these non-canonical forms of *O*-glycosylation.

DEDICATION

I dedicate this thesis to my wife, Daina, my two boys, Benjamin and Jacob and my parents, Hugh and Gail.

ACKNOWLEDGEMENTS

I would like to thank all of the Vocadlo lab members (past and present) for their help. More specifically, I would like to thank Dr. Vocadlo for taking a chance on me in the beginning and always supporting me. Dr. Stubbs for spending so much time with me on the finer aspects of carbohydrate chemistry. Dr. Macauley, Dr. Gloster, Scott Yuzwa and Dave Shen for schooling me in the art of biochemistry. Dr. Zandberg for helping with the capillary electrophoresis and tissue culture work, as well as Dr. Yadav and Simon Walker for helping me with the trisaccharide project and for synthetic advice. I would also like to thank all of the people that I mentioned for their friendship, encouragement, and help during my PhD.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SCHEMES

GLOSSARY

Ac	Acetyl protecting group
AcBr	Acetyl bromide
AcO_2	Acetic anhydride
AcOH	Acetic Acid
ADAM10	A disintegrin and metalloprotease 10
Ala	Alanine
Ank	Ankrin repeat
Ar	Aromatic ring
Arg	Arginine
Asn	Asparagine
ATP	Adenosine triphosphate
BF_3OEt	Borontrifluoro etherate
Bn	Benzyl
C1	Anomeric carbon
Calcd	Calculated
CAZy	Carbohydrate Active enZyme database
$CDCl_3$	Deuterated chloroform
CDG	Congenital disorder of glycosylation
CD_3OH	Deuterated methanol
CH_2Cl_2	Dichloromethane
Cl_3CCN	Tirchloroacetonitrile
CMP	Cytidine monophosphate
CPMV	Cowpea mosaic virus
CR	Cysteine rich domain
CSA	Camphor sulphonic acid
CuAAC	Copper-catalyzed azide-alkyne cycloaddition
Cys	Cysteine
DBU	1,8-Diazabicyclo[5.4.0]-undec-7-ene
DMAP	4-Dimethylaminopyridine
DMF	Dimethyl formamide
D_2O	Deuterium oxide
ECD	Extracellular domain
EGF	Epidermal growth factor
E●I	Enzyme, inhibitor complex
Et_3N	Triethylamine
Et_2O	Diethyl ether
EtOAc	Ethylacetate
E●S	Enzyme, substrate complex
E●S●I	Enzyme, substrate, inhibitor complex

Fmoc	9-fluorenylmethyl carbamate protecting group
Fuc	Fucose
ΔG	Change in Gibbs free energy
GAG	Glycosaminoglycans
Gal	Galactose
GalNAc	*N*-acetylgalactosamine
GDP	Guanidine diphosphate
Glc	Glucose
GlcA	Glucuronic acid
GlcNAc	*N*-acetylglucosamine
GlcNAz	*N*-azidoacetylglucosamine
GS	Ground State
HBSP	Hexosamine biosynthetic pathway
HCl	Hydrochloric acid
Hex A	β-Hexosaminidase A
Hex B	β-Hexosaminidase B
HCl	Hydrochloric acid
HNK-1	Human natural killer-1
HPLC	High performance/pressure liquid chromatography
HRESIMS	High resolution electrospray ionization mass spectrometry
H_2SO_4	Sulfuric acid
ICD	Intracellular domain
IdoA	Iduronic acid
Int	Intermediate
LFER	Linear Free Energy Relationship
LOGNAc	*N*-acetylglucosaminono-1,-5-lactone oxime
Man	Mannose
ManNAc	*N*-acetylmannosamine
ManNAz	*N*-azidoacetylmannosamine
MS	Mass spectrometry
N_2	Nitrogen
NaH	Sodium hydride
k_{cat}	Catalytic rate constant (turnover number)
K_2CO_3	Potassium carbonate
K_M	Michaelis constant (see appendix 1 for further detail)
K_{MAPP}	Apparent Michaelis constant of substrate (in the presence of an inhibitor)
K_I	Dissociation constant
KIE	Kinetic isotope effect
MeOH	Methanol
$MgSO_4$	Magnesium sulphate
MU-GlcNAc	4-Methylumbelliferyl 2-acetamido-2-deoxy-β-D-glucopyranoside
NAG-Thiazoline	1,2-dideoxy-2′-methyl-α-D-glucopyranoso-[2,1-d]-Δ2′-thiazoline
$NaHCO_3$	Sodium bicarbonate

NaOAc	Sodium acetate
NaOH	Sodium hydroxide
NaOMe	Sodium methoxide
NeuAc	*N*-acetylneuraminic acid
NMR	Nuclear magnetic resonance
Notch	A transmembrane protein embedded in the plasma membrane of the cell and involved in the notch signaling pathway
NRR	Notch regulatory region
OGA	*O*-GlcNAcase
O-GlcNAc	2-acetamido-2-deoxy-*β*-D-glucopyranoside linked to serine or threonine residues
O-GlcNAcase	*O*-glycoprotein 2-acetamido-2-deoxy-*β*-D-glucopyranoside cleaving enzyme
OGT	*O*-GlcNAc transferase
P1	Product 1 (the leaving group)
P2	Product 2 (GlcNAc)
PEST	Region rich in proline (P), glutamic acid (E), serine (S), threonine (T)
Pfp	Pentaflourophenyl ester activating group
PMM	Phosphomannomutase
POMT1	Protein *O*-mannosyl-transferase 1
Pro	Proline
PUGNAc	*O*-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-*N*-phenylcarbamate
RAM	Recombining binding protein associated module
Ser	Serine
TCA	Trichloroacetamide group
Thr	Threonine
TLC	Thin layer chromatography
TMSOTf	Trimethylsilyl trifluoromethanesulfonate
Trp	Tryptophan
TS	Transition state
TSR	Thrombospondin type 1 repeat
UV	Ultra violet
UDP	Uridine diphosphate
Xxx	Any amino acid
Xyl	Xylose

1: General introduction

1.1 An introduction to carbohydrates

Of the four major biological building blocks of life, carbohydrates, proteins, nucleic acids, and lipids, carbohydrates are the most abundant and yet they are arguably the least well understood. The major role of carbohydrates was traditionally viewed as being a source of energy accessed through glycolysis, which serves to convert glucose into pyruvate and in the process generate ATP, the primary source of energy within cells. Carbohydrates are also often primarly considered to be components of various biomaterials that confer structural properties to many organisms through their assembly into long polymers such as chitin, the principle component of arthropod exoskeletons, and cellulose, which provides rigidity to plants (Figure 1.1A and D). However, carbohydrates have also been implicated in the regulation of critical cellular processes. In 1965 Watkins and Morgan showed that the ABO blood group epitopes, which are important factors defining blood type compatibility for transfusions, were defined by carbohydrates[1]. Sialyl Lewis^x, another complex glycan structure (Figure 1.1B), plays a critical role in directing white blood cells, known as leukocytes, during the inflammatory process[2-4]. In addition to playing various other roles, *N*-glycans are pivotal in the protein quality control pathway within the endoplasmic reticulum and so aid proper protein folding (Figure 1.1C)[5-8]. While the impact of carbohydrates in diverse biological processes is undeniable and our understanding of their role is continually expanding, few of the advances that have been made would be possible without the development of carbohydrate specific tools and techniques.

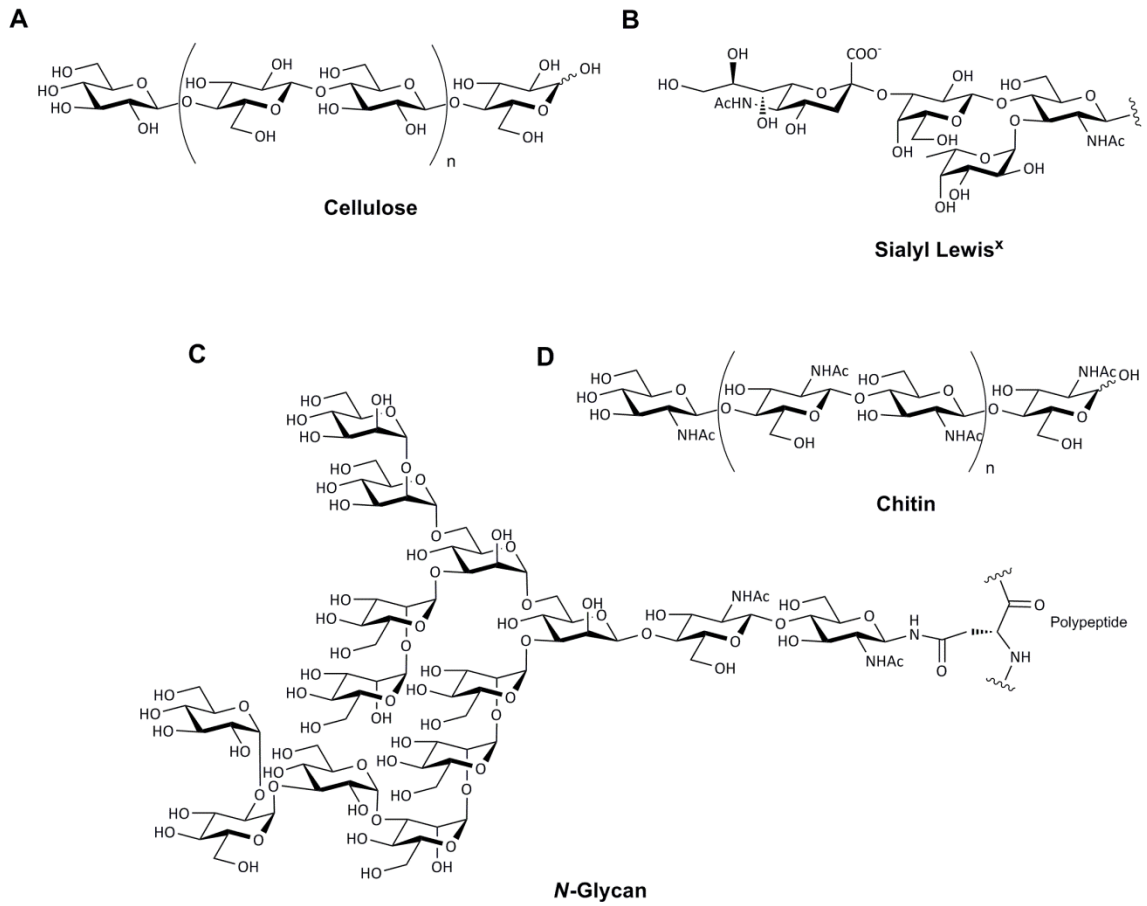


Figure 1.1: Several biologically relevant glycans

A) Cellulose is a polymer of repeating *β*1-4-linked glucose units that provides rigidity to plants. B) Sialyl Lewis^x is displayed on the terminus of glycolipids that are present on the surface of white blood cells. C) An *N*-glycan found in the endoplasmic reticulum. Upon cleavage of two of the terminal *α*1-2- and *α*1-3-linked glucose moieties the *N*-glycan modified protein can enter the calnexin/calreticulin protein folding pathway. D) Chitin is a polymer of repeating *β*1-4-linked GlcNAc units that is the principle component of arthropod exoskeletons.

1.1.1 Carbohydrate structure

To better understand the roles played by carbohydrates in biology and to provide sufficient background to appreciate the chemistry of carbohydrates, some background knowledge is essential. At the core of this background is the question, “what is a carbohydrate?” This question was not clearly answered until 1891 when Emil Fischer deduced the structure of glucose (and other carbohydrates) through a series of transformations including oxidation of glucose with nitric acid, and the use of phenylhydrazine to form various osazones[9,10]. To

accurately describe the complex stereochemical information present within glucose and all other carbohydrates, Fischer developed a method for depicting carbohydrates known as the Fischer projection (Figure 1.2). In a Fischer projection the carbon chain is drawn vertically with horizontal bonds projecting towards the viewer and vertical bonds projecting away from the viewer. Fischer projections are used to distinguish between different carbohydrates and the two enantiomers, D and L, that exist for each carbohydrate. D and L designations are assigned based on the stereochemical configuration of the stereogenic centre in the molecule that is furthest from the carbonyl centre (Figure 1.2B, R = D, S = L). R (for rectus, which is latin for right) and S (for sinister, which is latin for left) configuration is determined by the relative priority of the substituents attached to the stereogenic centre assigned according to the Cahn-Ingold-Prelog system, (reviewed by March[11]). Carbohydrates can also be broken down into two groups, aldoses and ketoses (Figure 1.2A). In aldoses the carbonyl group is always at the top of the Fischer projection and the aldose carbon is referred to as C1, with numbering of carbon atoms continuing down the chain (Figure 1.2B). With ketoses the carbonyl carbon is usually the second carbon from the top of the chain and is referred to as C2 (Figure 1.2C).

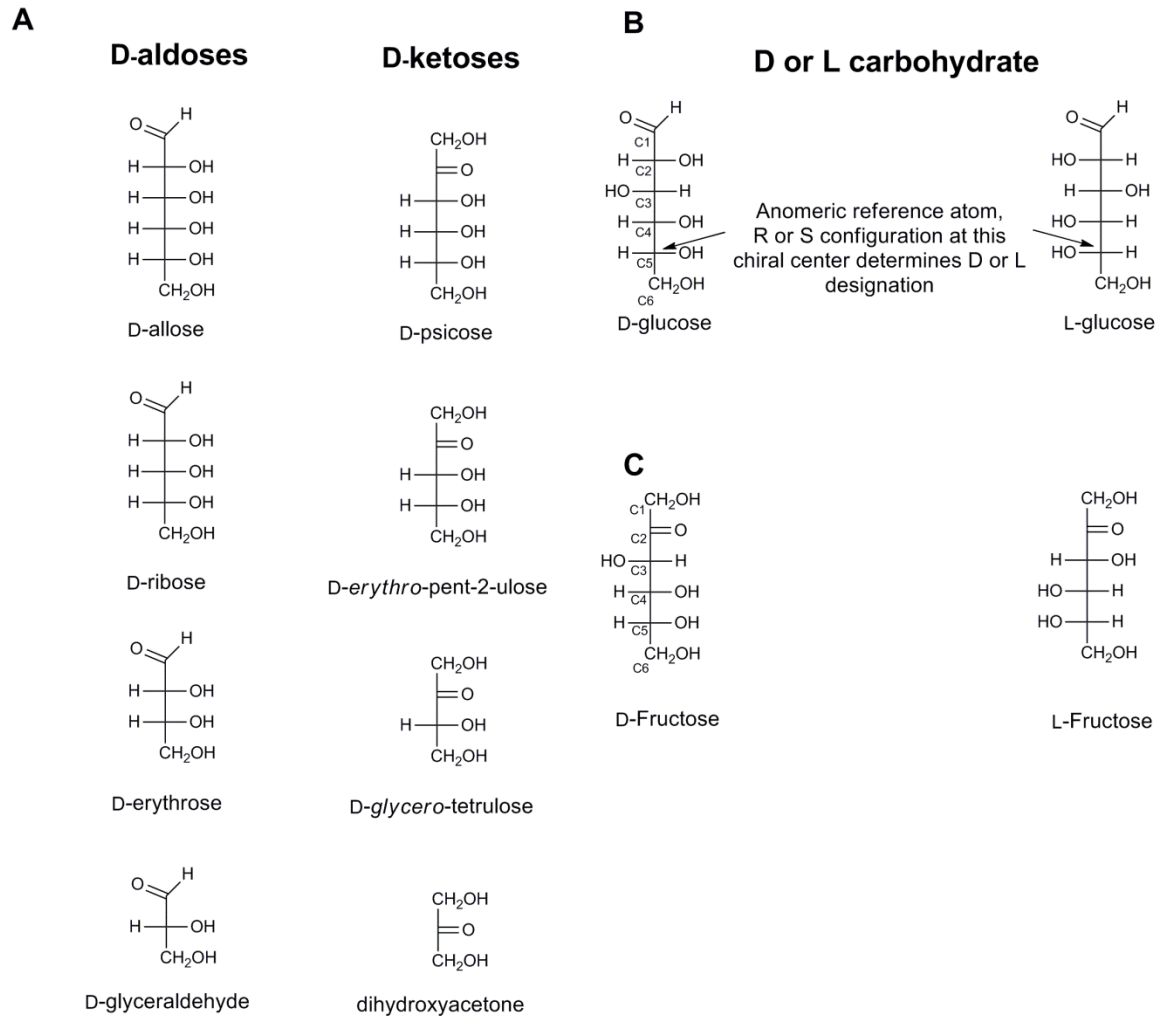


Figure 1.2: Fischer projection of some aldoses and ketoses

A) Examples of 3-6 carbon unit aldoses and ketoses in the D configuration shown as Fischer projections. B) The determination of the D or L configuration of glucose depends on the stereochemical configuration (R or S) at the chiral centre (C5 for glucose) furthest from the aldehyde at C1, C) or ketone at C2, when dealing with most ketoses.

While Fischer projections convey many of the structural aspects of monosaccharides, most monosaccharides actually exist in a cyclic form. In reference to D-glucose, the C5 hydroxyl group reacts intramolecularly with the C1 aldehyde to form a hemiacetal (Figure 1.3). The resulting pyranose ring structure preferentially adopts a non-planar conformation so as to reduce ring strain. The chair conformation is the most stable and energetically favourable conformation adopted (Figure 1.3) by most aldoses including glucose. Depending on which

face of the anomeric carbonyl group is attacked (re or si face) the resulting cyclization gives rise to either an α- or β-anomer. The α-anomer is formed when the oxygen attached to the anomeric carbon (O1 in glucose) and the oxygen (O5 in glucose) attached to the anomeric reference atom (C5 in glucose) are on the same side in the Fischer projection (Figure 1.3). When these two oxygen atoms are on different sides in the Fischer projection the β-anomer is formed. Because the α- and β-anomers interconvert and are defined by changes in stereochemistry at C1, C1 is also referred to as the anomeric carbon.

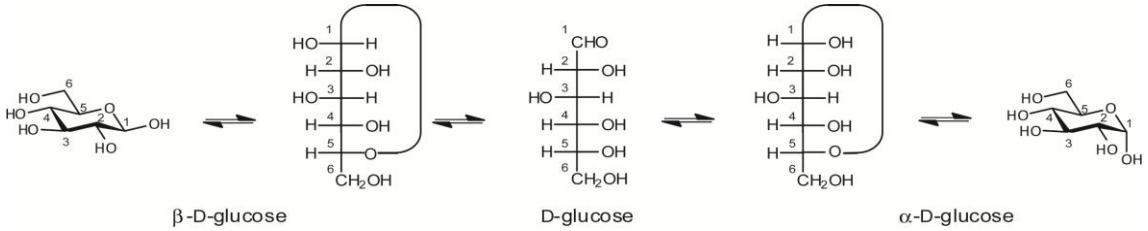


Figure 1.3: Cyclization of D-glucose to *β*-D-glucose or *α*-D-glucose

D-Glucose exists primarily in the pyranose chair formation. Cyclization from the Fischer projection of D-glucose to *β*-D-glucose or *α*-D-glucose is shown. Carbon centres are numbered to indicate where they appear upon cyclization.

The α- and β-anomers at C1, as illustrated for D-glucose, that are formed upon cyclization can also interconvert through a process known as mutarotation (Figure 1.3). This interconversion occurs because of the chemical properties of the hemiacetal functionality (Figure 1.4A). The hemiacetal (Figure 1.4B) formed when a nucleophilic alcohol (the C5 hydroxyl in the case shown here) attacks the carbonyl centre (C1) is unstable and can revert back to the original aldehyde and alcohol (Figure 1.4A). A more stable acetal functionality is formed when the alcohol group of the hemiacetal is replaced with an alkoxy group. The resulting species do not mutarotate and are known as glycosides. The C1-O1 bond of these species is known as the glycosidic bond or glycosidic linkage (Figure 1.4B). Proteins, lipids, and other carbohydrates are the most common types of biomolecules conjugated to carbohydrates via glycosidic linkages.

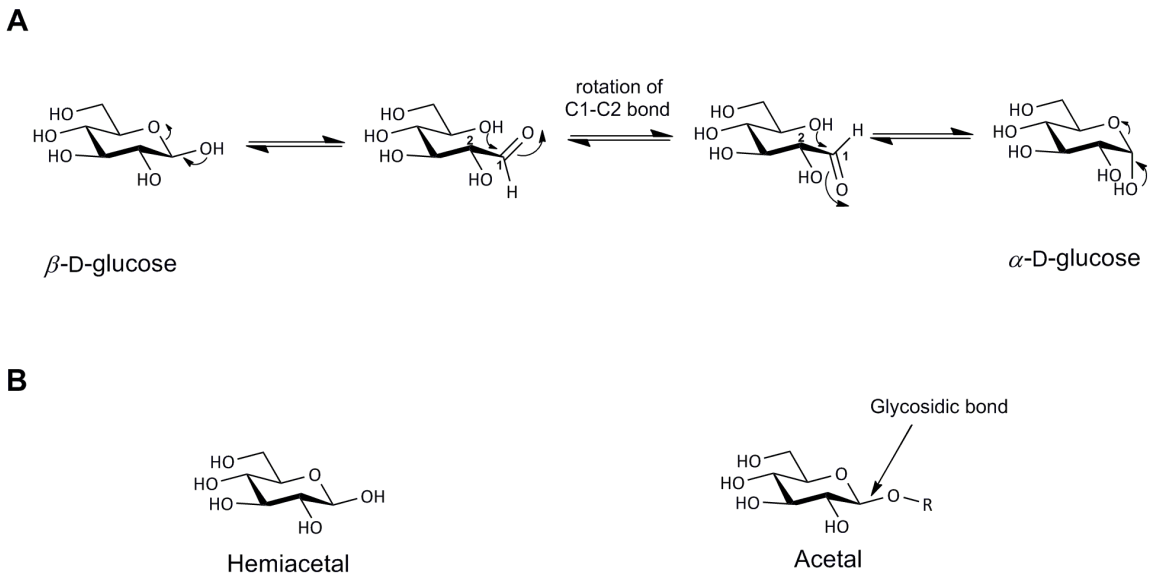


Figure 1.4: Mutarotation of *α*- and *β*-anomers of D-glucose

A) Interconverion between *α*- and *β*-anomers of D-glucose occurs rapidly due to the instability of the hemiacetal functionality. Looking at mutarotation from the *β*- to the *α*-anomer of D-glucose, the *β*-D-glucose ring opens to give the aldehyde functionality, rotation of the C1-C2 bond allows O5 to now attack the other face of the carbonyl to form the *α*-anomer. B) The hemiacetal of *β*-D-glucose (left) and the acetal (right), with the glycosidic bond between C1 of the donor and the oxygen atom of the acceptor indicated by an arrow.

1.2 Glycosyltransferases and glycosidases

Carbohydrates can be *O*-glycosidically linked together to form di-, oligo-, and polysaccharides. In eukaryotes the assembly of these carbohydrate chains, known as glycans, and their attachment to other biomolecules to form glycoconjugates, occurs within subcellular compartments known as the endoplasmic reticulum and golgi apparatus. After their proper assembly, glycoconjugates usually end up on the outer surface of the plasma membrane, which surrounds the cell, or are secreted from the cell into the extracellular mileu. Glycoproteins and glycolipids form a dense layer around the cell called the glycocalyx. The glycocalyx is involved in many cellular processes, including for example inflammation regulation[2-4], cell to cell signalling[12,13], and cell adhesion[14]. Most glycoconjugates including proteins and lipids bearing glycans are eventually transported to the lysosome for degradation and recycling. The enzymes

involved in the assembly and degradation of glycans is of central relevance to this thesis.

1.2.1 Classification of glycosidases and glycosyltransferases

Glycosidic linkages within biomolecules are formed and broken down by carbohydrate processing enzymes. These enzymes account for between 1 and 3 % of the genome of most organisms[15]. The two types of carbohydrate active enzymes that are relevant to this thesis are glycosyltransferases (enzymes that catalyze the transfer of carbohydrates to proteins, lipids, and other carbohydrates) and glycosidases (enzymes that cleave carbohydrates from proteins, lipids, and other carbohydrates). The Carbohydrate Active enZyme (CAZy) database was developed to classify the carbohydrate processing enzymes into "families" based on primary sequence similarity[16]. Members of each family of glycosidases have conserved protein folds, active site architectures, catalytic residues, and catalytic mechanisms. For glycosyltransferases the CAZy database is less predictive but still very useful. Thus, the CAZy database is a powerful tool enabling the characterization of new carbohydrate processing enzymes[17].

1.2.2 Glycosyltransferases

Glycosyltransferases use high energy donor sugar substrates that contain a phosphate or diphosphate leaving group. There are ten different sugars commonly found in eukaryotic glycoconjugates (Figure 1.5), most of them are activated in the form of a nucleoside diphosphate sugar such as UDP-Glu, UDP-Gal, and GDP-fucose. However, CMP-NeuAc is a monophosphate donor sugar, while iduronic acid does not have a nucleoside sugar donor because it is enzymatically produced by epimerization of glucuronic acid already incorporated into glycosaminoglycan (GAG) chains[18]. While the most common type of bond formed by glycosyltransferases is an *O*-glycosidic linkage to another carbohydrate, lipid, or protein, these enzymes also catalyze the formation of

carbon-nitrogen (N-glycans)[19], carbon-sulfur (glycopeptide antibiotics produced by bacteria)[20], and carbon-carbon (C-mannosylation)[21] bonds (Figure 1.6).

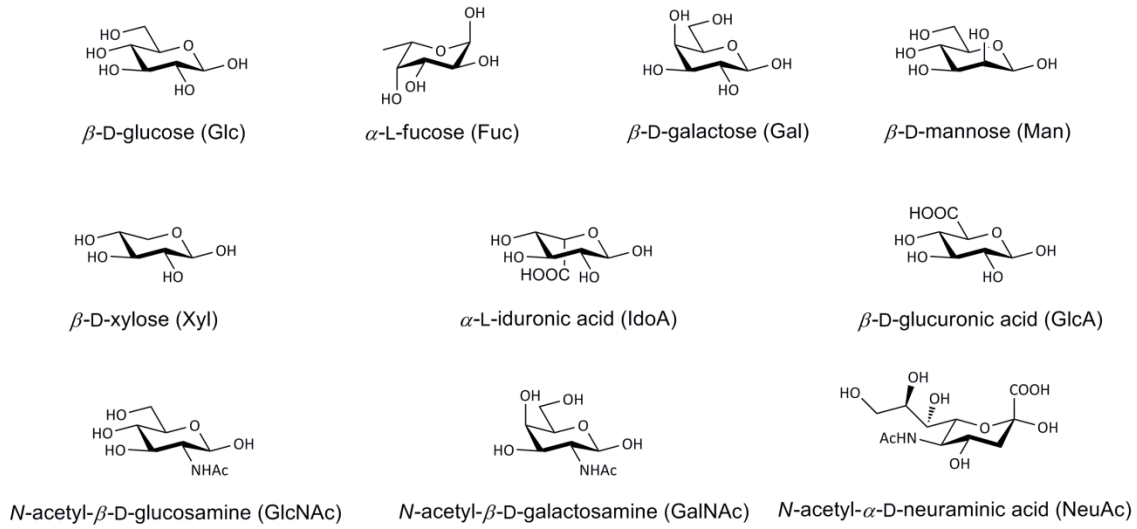


Figure 1.5: Carbohydrates used by mammals

The ten sugars found in eukaryotic glycoconjugates. Each sugar is activated to a nucleoside (di)phosphate donor with the exception of IdoA which is formed from GlcA already incorporated into a glycan.

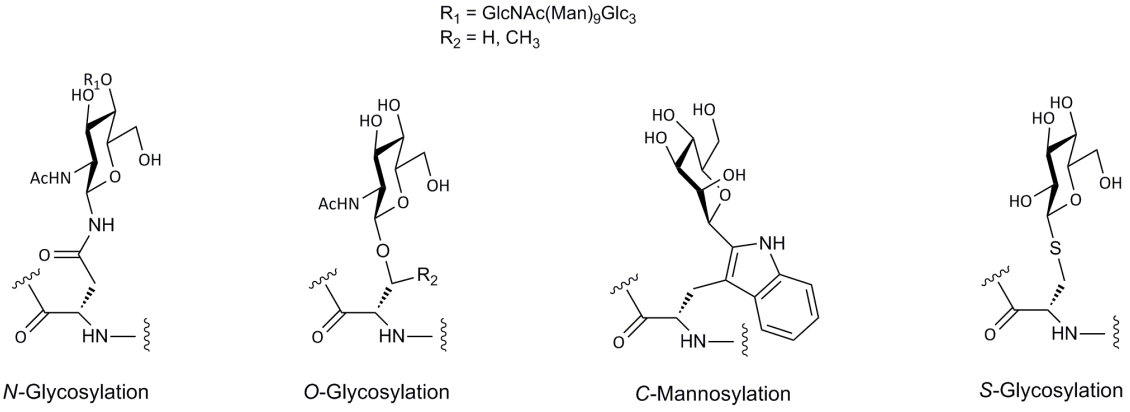


Figure 1.6: Glycosidic Linkages

Examples of *N*-, *O*-, and *C*-glycosidic linkages to proteins and a recently identified *S*-linked glycopeptide; *N*-glycosylation to the amide group of asparagine with an *N*-acetylglucosamine residue (left), *O*-glycosylation to the hydroxyl group of serine or threonine with an *N*-acetylglucosamine residue (centre-left), *C*-mannosylation of the C2 atom of the side chain of tryptophan with a mannose residue (centre-right), and *S*-glycosylation of the glycopeptide sublancin with a glucose residue.

The glycosidic linkage formed by glycosyltransferases can occur with retention or inversion of stereochemistry at the anomeric centre of the transferred sugar unit. The stereochemical outcome of the reaction is used to categorize glycosyltransferases as either inverting, where the stereochemistry changes, or retaining, where the stereochemistry remains the same. The general mechanism for inverting glycosyltransferases is generally thought to be a direct displacement S_N2-like reaction. An active site side chain acts as a general base catalyst, aiding nucleophilic attack of the glycosyl acceptor and concomitant departure of the (di)phosphate leaving group via an oxocarbenium ion-like transition state to give the stereochemically inverted product (Figure 1.7A)[22]. The catalytic general base for this reaction was found to be the side chain of either an aspartic acid[22], glutamic acid[23], or histidine[24] residue. To facilitate departure of the (di)phosphate leaving group a Mn^{2+} or Mg^{2+} cation coordinates to a DXD motif in the enzyme and thereby stabilizes the developing negative charge on the leaving group. One proposed mechanism for retaining glycosyltransferases is a two-step, double-displacement mechanism involving transient formation of a glycosyl-enzyme intermediate on an enzymic nucleophile with inversion of stereochemistry at the anomeric centre[25,26]. Soya *et al*. recently provided evidence to support the formation of the proposed glycosyl-enzyme intermediate by mutating the catalytic nucleophile of two glycosyltransferases from Glu to Cys[25]. Incubation of these mutated transferases with the requisite donor sugar produced a covalent glycosyl-enzyme intermediate that was detected by mass spectrometry[25]. In a second step the nucleophilic acceptor group displaces the enzyme nucleophile causing a second inversion of stereochemistry and leading to overall stereochemical retention (Figure 1.7B). Following the recent proposal that the phosphate leaving group acts as a general base catalyst in the mechanism of the farnesyl diphosphate synthases[27,28], the nucleoside diphosphate leaving group has been predicted to act as the general base catalyst for the incoming acceptor hydroxyl group for the glycosyltransferase double-displacement mechanism[29]. A second mechanistic possibility for overall stereochemical retention that has received more mechanistic support is an S_Ni-like mechanism wherein a highly

dissociative transition state, or formation of a discrete oxocarbenium ion intermediate, allows for nucleophilic attack at the same face as the leaving group, thus retaining overall stereochemistry at the anomeric centre[30-32]. It has been proposed that the leaving group forms a hydrogen bond with the incoming nucleophile to direct it to the correct face[29,33,34]. Although there has been recent progress, the study of glycosyltransferases has lagged behind glycosidases due to the paucity of chemical tools for determining catalytic mechanisms, as well as the great difficulty associated with purification and structural determination of membrane bound glycosyltransferases by X-ray crystallography. Two other major reasons for the limited knowledge regarding glycosyltransferases as compared to glycosidases, is the more complex nature of both the acceptor substrate (proteins, carbohydrates, and lipids) and the donor substrate (nucleoside (di)phosphate sugars).

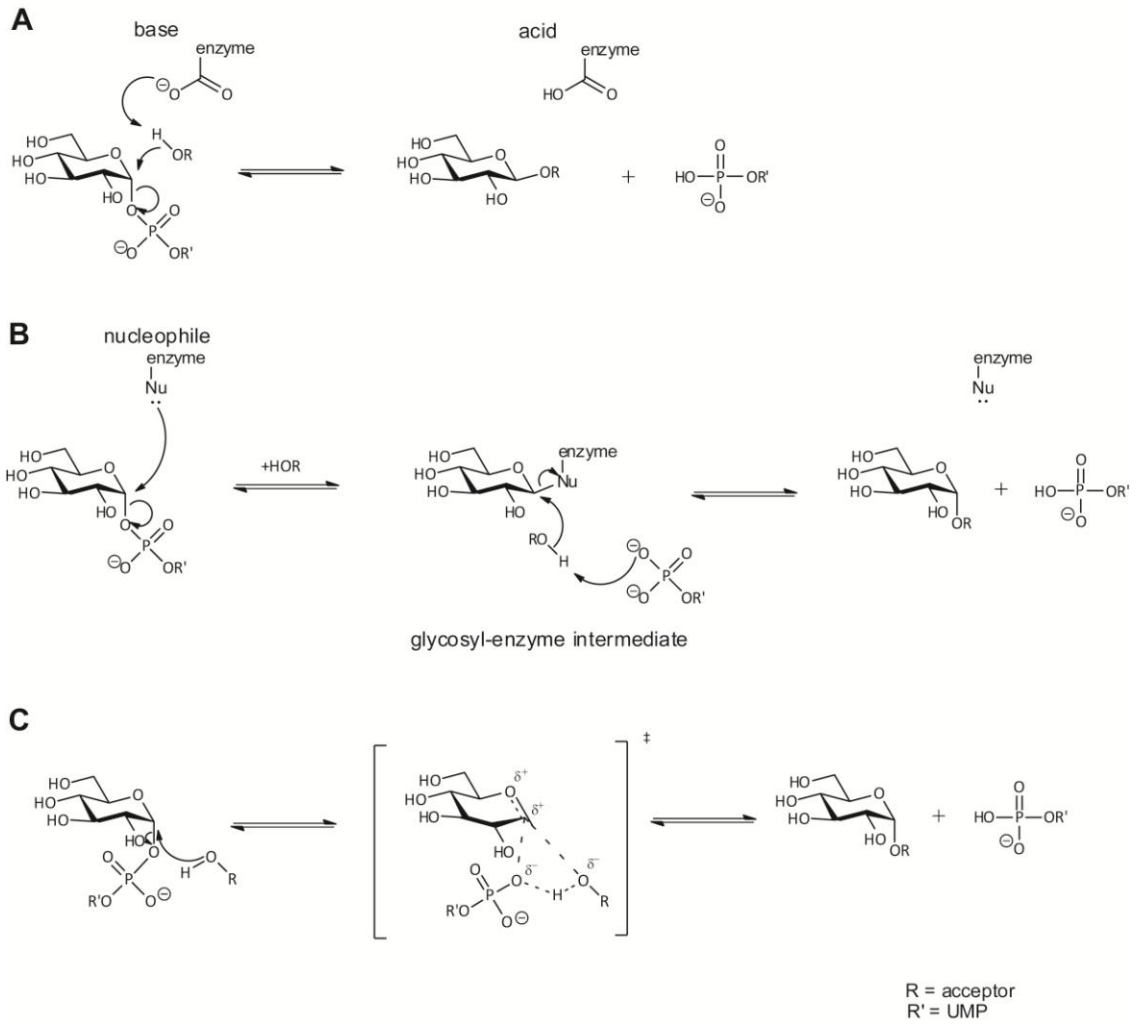


Figure 1.7: Putative general inverting and retaining glycosyltransferase mechanisms.

A) A one step inverting mechanism involving a direct-displacement S_N2-like reaction. B) A two step retaining glycosyltransferase mechanism where a glycosyl-enzyme intermediate is formed on an enzymic residue before being displaced by the acceptor. C) A one step retaining S_Ni-type mechanism which may involve transient formation of an oxocarbenium ion intermediate.

1.2.3 Glycosidases

As previously mentioned, glycosidases catalyse the hydrolysis of glycosidic bonds. This class of enzymes are arguably the best studied of the carbohydrate-processing enzymes. In depth details of glycosidases mechanism, structure, function, and active site composition are readily available for different glycosidases from many different families. This is due to several detailed kinetic isotope studies[35-37], development of various potent inhibitors[38-44], and a large

number of X-ray crystallography studies[45-47]. Similar to glycosyltransferases, glycosidases can be categorized as inverting or retaining, however, they can also be further classified as *endo*- (cleaving a non-terminal glycosidic linkage)[48] or *exo*- (cleaving a terminal glycosidic linkage at the reducing or non-reducing end)[49] processing (Figure 1.8).

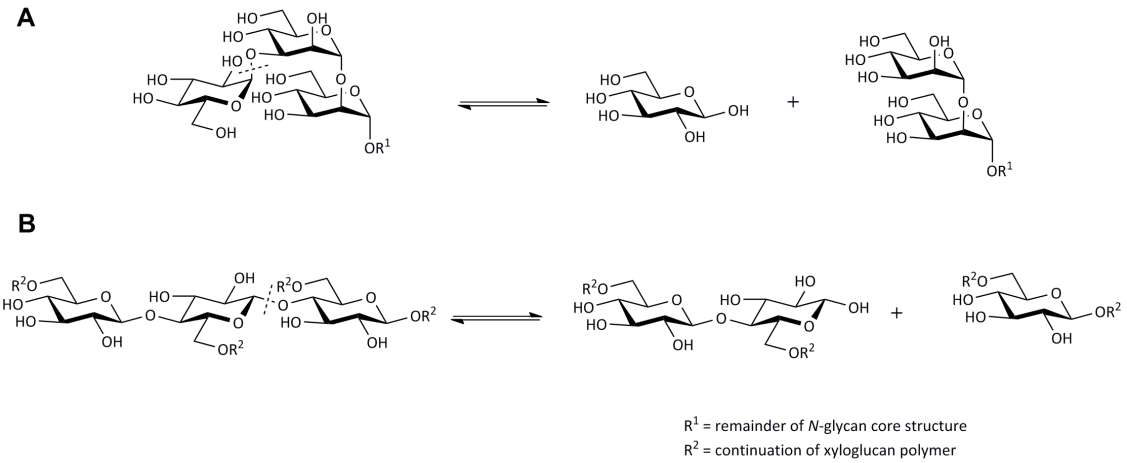


Figure 1.8: Classification of *exo*- and *endo*-glycosidases.

A) Sites of cleavage of the Glc-*α*1-3-Man glycosidic linkage (depicted by dashed line through the glycosidic bond) of *N*-glycans by *exo*-glucosidase II[49]. B) Cleavage of the Glc-*β*1-4-Glc glycosidic linkage (depicted by dashed line through the glycosidic bond) found in xyloglucan polymers of plant cell walls by an *endo*-xyloglucan *β*-glycosidase[48].

1.2.4 Basic mechanism

In 1953, Koshland proposed the conceptual basis for glycosidase catalyzed cleavage of the glycosidic linkage occurring with either retention or inversion of stereochemistry at the anomeric centre of the product relative to the stereochemistry of the anomeric centre of the substrate[50]. This proposal has stood the test of time, with mechanisms for both retaining and inverting glycosidases having been determined (Figure 1.9). Retention of anomeric stereochemistry occurs in two steps. In the first step a nucleophilic residue (usually aspartate or glutamate) in the active site of the enzyme attacks the anomeric carbon, displacing the leaving group and inverting stereochemistry at the anomeric centre to form a covalent enzyme-substrate intermediate. During the second step, the intermediate is hydrolyzed, resulting in a second inversion

of stereochemistry and thus retaining the initial anomeric stereochemistry in the product of the overall reaction (Figure 1.9A). During this process an active site general acid/base residue facilitates the departure of the leaving group and in the second step acts as a general base to aid attack of a molecule of water to hydrolyze the enzyme intermediate. The inverting mechanism occurs through the use of two active site residues (typically aspartate or glutamate). In this one step mechanism, an enzymic general base catalytic residue facilitates the attack of water at C1 while an enzymic general acid catalytic residue aids aglycon departure (Figure 1.9B).

Physical organic evidence supporting these mechanisms comes from various studies. Kinetic isotope effects (KIEs) are a particularly useful technique for discerning the geometry, charge distribution, and the extent of bond dissociation/association at the transition state of an enzyme catalysed reaction[36,51,52]. KIE studies generally require two comparable substrates differing only in that one is enriched with a stable heavy isotope and the other with the lighter isotope. The KIE is determined by measuring the difference in the reaction rate caused by the presence of the two different isotopes. This needs to be done when the atoms are involved in the chemistry of the rate determining step of the reaction. Three different KIEs can be measured: primary (isotopic substitution of the atom that is undergoing a change in bond order), *α*-secondary (isotopic substitution of an atom one bond away from the atom undergoing a change in bond order), and *β*-secondary (isotopic substitution of an atom two bonds away from the atom undergoing a change in bond order). In a classic study, Sinnott and Souchard used *α*-secondary KIEs to propose the existence of a covalent glycosyl enzyme intermediate[36]. More recently Chan *et al*. used a combination of primary and *α*-secondary KIEs to propose a transition state structure and reaction coordinate for the sialadase-catalyzed hydrolysis of sialyllactoside substrates[37]. A second useful probe of transition state charge distribution and geometry, are linear free energy relationships (LFERs). The most common LFERs used in enzyme kinetic studies are the Brønsted relationship (probes the charge development on the glycosidic oxygen), Taft-like relationship (used to

measure both steric and electronic effects of the substituent), and the Bartlett relationship (probes the resemblance of an inhibitor to the transition state). All three relationships involve structural alteration of the substrate, inhibitor, enzyme, or a combination of the three. To illustrate with an example, building on the work of Sinnott and Souchard[36], Kempton and Withers used a series of aryl glucosides with phenol p*K*a values ranging from 4 to 10 to construct a Brønsted relationship[53]. They observed a concave-downward plot consistent with a two-step mechanism involving a glucosyl-enzyme intermediate for a *β*-glucosidase from *Agrobacterium faecalis*. Furthermore, the slope of the leaving group portion of the Brønsted plot indicated a large degree of bond cleavage at the transition state[53]. Use of the Brønsted relationship has been very useful in clarifying the mechanism of enzyme action[54-57]. Recently, Lee *et al*. took advantage of all three LFERs to provide evidence for the existence of a S_Ni-type reaction in a retaining glycosyltransferase[34], while a classical example of a Taft-like LFER pertaining to glycosidases and the role of an acetamido group in substrate assisted catalysis was undertaken by Yammamoto in 1974[58].

Figure 1.9: General inverting and retaining glycosidase mechanisms

A) General mechanism for the retention of anomeric stereochemistry. An enzyme based nucleophile forms a covalent enzyme-substrate intermediate before a molecule of water attacks at the anomeric carbon, forming the product with retention of anomeric stereochemistry. B) General mechanism for the inversion of anomeric stereochemistry. An enzymic general base catalytic residue facilitates the attack of water on the anomeric centre with aglycon departure assisted by the general acid catalytic residue.

1.2.5 Transition state stabilization

Remarkably, glycosidases are able to enhance the rate of catalysis 10^{17} fold over the uncatalyzed reaction[59]. In 1948 Pauling proposed that this rate enhancement is achieved through binding interactions realized preferentially with the transition state rather than with the ground state (Pauling's transition state theory and inhibitor design is discussed further in Chapter 3). Tight binding of the fleeting transition state structure by the enzyme stabilizes it, resulting in lower activation energies and enormous rate accelerations[59]. In 1973 Sinnott shed more light on the mechanism by which carbohydrate active enzymes achieve this impressive catalytic proficiency. An oxocarbenium ion-like transition state with a trigonal anomeric centre was proposed for both inverting and retaining mechanisms (Figure 1.9)[36]. This idea implied that at the transition state the bonds being broken and/or formed were highly dissociative, causing significant charge development at the anomeric centre. The positive charge that develops at the anomeric centre was thought to be stabilized by overlap between p-like orbitals of the endocyclic oxygen and C1. To accommodate this p-orbital stabilization it was postulated that the transition state adopts a coplanar arrangement of C5, O5, C1, and C2 (Figure 1.10B)[36]. This coplanar arrangement could only be satisfied if the carbohydrate in question strayed from its energetically favoured $^{4}C_{1}$ (chair) configuration. From Stoddards pseudorotational itinerary[60], the only conformations that fulfill this coplanar arrangement are the: $^{2,5}B$ and $B_{2,5}$ conformations (boat conformations, numbers to the left and superscript represent atoms above the plane of the ring and numbers to the right and subscript represent atoms below the plane of the ring) and the $^{4}H_{3}$ and $^{3}H_{4}$ (half chair) conformations (Figure 1.10B). Using secondary α-deuterium and β-deuterium kinetic isotope effects Sinnott was able to provide evidence for this hypothesis[36,61,62]. Lending credence to this theory, in 1996 two different groups observed substrates distorted away from the energetically favourable $^{4}C_{1}$ conformation toward envelope and skew-boat conformations[63-65]. While Koshland and Sinnott's insightful proposals provided the conceptual basis for the aforementioned glycosidase mechanisms, tools such as inhibitors gave

credence to the theory. Many of the potent inhibitors discussed below, regardless of whether they were discovered in nature or are synthetic products of rational design, mimic the transition state conformation and/or charge development proposed by Sinnott.

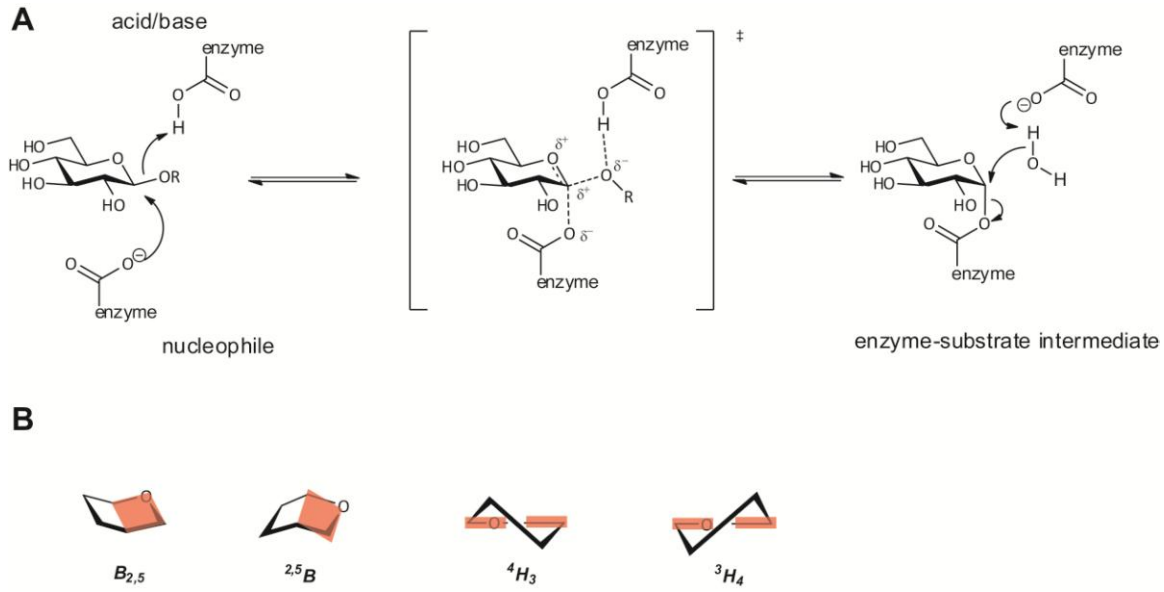


Figure 1.10: Glycoside catalyzed formation of an oxocarbenium ion-like transition state and its geometric requirements

A) Shown here for the first step catalyzed by a retaining glycosidase, the catalytic mechanism of glycosidases proceeds through a highly dissociative oxocarbenium ion-like transition state with low bond orders between the anomeric centre and both the nucleophile and the leaving group. Positive charge development at C1 is stabilized by overlap between the occupied p-like orbitals of the endocyclic oxygen and the vacant p-like orbital at C1. B) Conformations from Stoddard's pseudorotational itineraries in which C5, O5, C1, and C2 have a planar geometry required for efficient orbital overlap. The plane is highlighted in red.

1.3 Chemical and biochemical tools for probing glycans and glycan processing enzymes

The range of tools and techniques that are important for the study of glycosidases, glycosyltransferases, and the glycans that they process; include for example, X-ray crystallography, site directed mutagenesis, affinity labelling, inhibitors, lectins, carbohydrate specific antibodies, and kinetic isotope studies. With the development of new tools and methodologies for improving our

understanding of specific biologically relevant glycans and glycosidases being the main goals of this thesis, I will provide a brief overview of some of these tools. Considering a major focus of this thesis is on hydrolase inhibitors, more specifically non-covalent glycosidase inhibitors, I will start by reviewing some of the inhibitors relevant to the field.

1.3.1 Inhibitors in nature

The impetus for studying or developing glycosidase inhibitors is generally either to generate a potential therapeutic agent for the treatment of a disease or as an aid for understanding catalytic mechanism or cellular function of an enzyme. Nojirimycin (Figure. 1.11), an inhibitor of both α- and β-glycosidases, was initially isolated from a strain of *Streptomyces* in 1966[38], and stimulated researchers to generate a slew of analogues. Deoxynojirimycin (Figure. 1.11), a close analogue and also a potent inhibitor of both α- and β-glycosidases, was synthesized in 1967 by Paulsen and coworkers[39] prior to being found in nature[66]. Current drugs Miglitol[67], used to treat type II diabetes, and Zavesca/Miglustat[68,69], a glucosylceramide synthase inhibitor used in substrate reduction therapy to treat Gaucher disease, build upon the nojirimycin scaffold (Figure. 1.11). Other examples of potent inhibitors of glycosidases have been found in nature, for instance, castanospermine inhibits a wide range of glycosidases[70]. It was used prominently to aid in determining the role and mode of action of calnexin by inhibiting endoplasmic reticulum (ER) *α*-glucosidases, thus preventing the binding of calnexin to the mannose capped antennae of *N*-glycan modified proteins[5]. Swainsonine, a potent inhibitor of the *N*-linked glycan-processing enzyme *α*-mannosidase II, has been shown to reduce tumor cell metastasis, enhance cellular immune responses, and reduce solid tumor growth in mice[71,72]. Acarbose, another good example of a glycosidase inhibitor found in nature[73], is a potent, oligosaccharide-based inhibitor of intestinal *α*-amylase used in the treatment of diabetes[74]. These and other natural glycosidase inhibitors also offer a wealth of design based knowledge to those who wish to chemically synthesize potent inhibitors for specific enzymes.

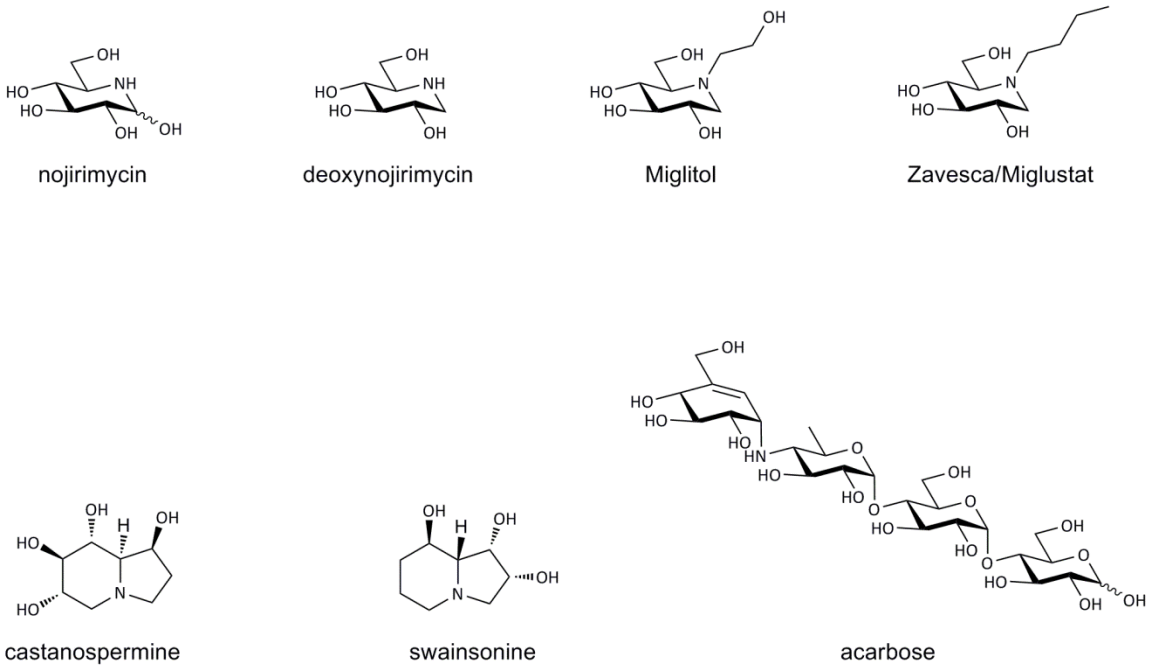


Figure 1.11: Naturally occurring glycoside hydrolase inhibitors and the glucosylceramide synthase inhibitor, Miglustat

Five glycosidase inhibitors found in nature and two pharmaceuticals elaborating on the nojirimycin scaffold; nojirimycin, deoxynojirimycin, Miglitol, Zavesca/Miglustat, castanospermine, swainsonine, and acarbose.

1.3.2 Inhibitor design theory

The vast majority of rationally designed unnatural glycosidase inhibitors incorporate elements of the canonical oxocarbenium ion-like transition state structure predicted by enzymological observations[36,62,75]. Accordingly, two principle design features have been extensively exploited[40,41,76]. The first principle design feature is the installation of a nitrogen atom in place of C1 or O5 to mimic the relative positive charge development at these centres within the transition state. Isofagomine and azafagomine (Figure 1.12), as well as nojirimycin (Figure 1.11), castanospermine (Figure 1.11), and related inhibitors are good examples of transition state charge distribution mimicry[77-79]. Secondly, an sp^2-hybridized centre is often installed at C1 of the pyranose ring to mimic the geometric requirements of the putative planar (planar geometry at C5, O5, C1,

and C2) oxocarbenium ion-like transition state. This “geometric transition state mimicry” was first proposed by Sinnott[36] and rationalized by low bond orders for the transition state structure, between the nucleophile and the anomeric centre, as well as the leaving group and the anomeric centre (Figure 1.10A). Two potent inhibitors that take advantage of an sp^2 hybridized anomeric centre are glucotetrazole and *O*-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-*N*-phenylcarbamate (PUGNAc) (Figure 1.12), which were both synthesized by Vasella and coworkers[42,43]. Other potent synthetic and natural inhibitors such as *gluco*-nagstatin[80] (Figure 1.12) and acarbose[73] (Figure 1.10) have been developed/discovered that force the pyranose ring to adopt a transition state-like conformation ($^4H_3$ and $^{2,5}B$ respectively, Figure 1.11B)[76,81,82]. To increase potency many inhibitors have been designed to exploit both of these traditional design features. For example glucotetrazole and *gluco*-nagstatin incorporate a nitrogen in place of O5 and take advantage of a sp^2 hybridized anomeric centre. Synthetic rationally designed glycosidase inhibitors and natural glycosidase inhibitors have been used effectively in the determination of glycosidase mechanisms, biological processes, and as pharmaceuticals to treat diseases and viral infections. Inhibitors such as Relenza[83] and Tamiflu[84], neuraminidase inhibitors that slow the cellular spread of the influenza virus by inhibiting a viral neuraminidase, have been developed into pharmaceuticals.

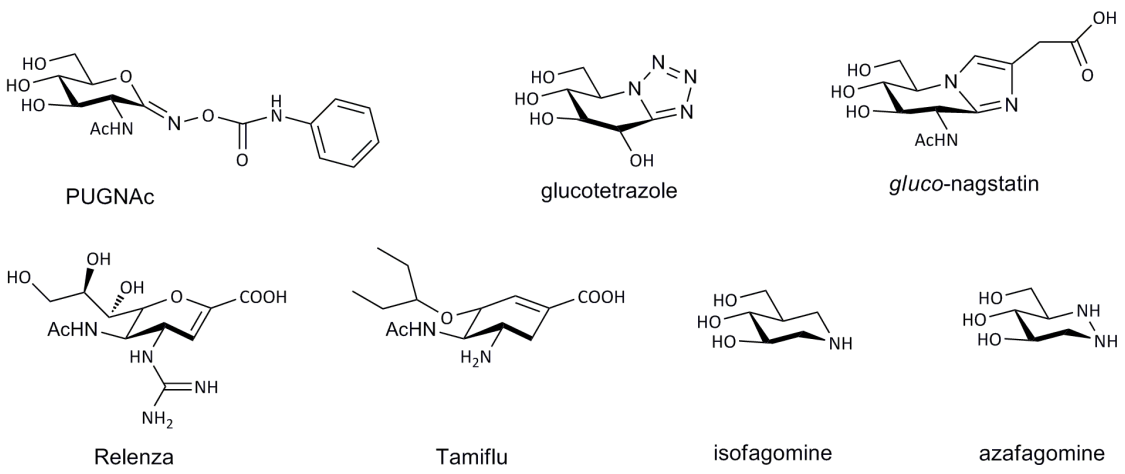


Figure 1.12: **Rationally designed glycosidase inhibitors**

Rationally designed inhibitors that contain traditionally important design features; PUGNAc, glucotetrazole, *gluco*-nagstatin, Relenza, and Tamiflu are all conformationally constrained inhibitors, while azafagomine and isofagomine incorporate nitrogen at C-1 and are transition state charge mimics.

1.4 Methods of detecting glycoproteins

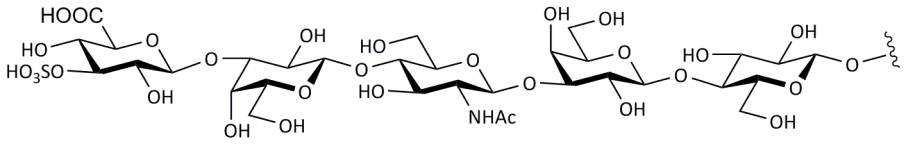
While there are many types of glycoconjugates, the focus of this thesis is on glycoproteins, so only leading methods of detection that can be applied to the characterization of glycoproteins will be discussed. The glycosylation of proteins is a topic of great current interest since glycosylation is among the most common and complex forms of protein post-translational modification[85]. It has been estimated that over 50% of all proteins are modified by glycans[86]. Identification and analysis of this wide variety of mono-, poly-, or oligosaccharides decorating many different types of proteins can be difficult at the best of times. With the growing realization that carbohydrates play essential roles in biological systems[87-90], the development of chemical tools to isolate, probe, and analyze the large variety of glycoproteins and the mono-, poly-, or oligosaccharides that decorate them is increasingly important. I will focus on a general description of the tools most relevant to this thesis, for more comprehensive discussions of analytical methods and tools to monitor glycans there are several excellent reviews that can be accessed[85,91-93].

1.4.1 Lectins

Lectins are carbohydrate binding proteins that recognize particular glycans or segments of glycans. Currently, more than 300 lectins have been discovered[94], with the majority of these being derived from plants and act as toxins. Mammalian lectins are also known and these play important roles in diverse biological processes[95-97]. For example galectins, a family of lectins, bind the *N*-acetyllactosamine (Gal-*β*1-4-GlcNAc) unit of *N*-glycans[98]. Galectin-3 has been shown to regulate the surface retention time of cytokine receptors through binding of *N*-glycans on the cytokine receptor and the formation of a lattice[99]. As research tools it is common for lectins to be immobilized on the stationary phase of a column and used to bind and enrich proteins modified with a specific glycan structure[100,101]. Concanavalin A (ConA) is a lectin originally extracted from *Canavlia ensiformis* that binds to high mannose and complex type *N*-glycans. In 1976 ConA was coupled to sepharose to generate an affinity chromatography column specific for high mannose and complex type *N*-glycan modified proteins[102]. This method has been widely applied towards the study of *N*-linked glycoproteins and glycopeptides[103]. Ricinus communis agglutinin (RCA) binds *N*-glycans with terminal galactose units, RCA affinity chromatography has been used to show that patients with overt primary hypothyroidism had increased levels of terminal NeuAc and galactose units[104]. Lectins can also be incorporated onto surfaces in an array and used to evaluate cell surface sugars[105], and lectin-carbohydrate interactions[106-108]. One disadvantage of using lectins for analysis or purification of glycoproteins is that some non-specific binding can occur, leading to the capture/analysis of proteins/peptides that bear incorrect glycans or have an absence of glycans[85]. For this reason lectins are primarily used as a method of purifying/enriching glycans or glycoconjugates, a downstream analysis technique is typically required to characterize the isolated glycans or glycoconjugates.

1.4.2 Antibodies

Carbohydrate specific antibodies are powerful tools for the identification of proteins modified by a particular glycan[109-111], localization of glycoproteins[112], and enrichment of glycoconjugates for analysis[92]. Unfortunately, the development of carbohydrate specific antibodies is perceived to be difficult due to the low immunoreactivity of most glycans. This problem can be overcome by generating glycopeptide site-specific antibodies, where the antibody is targeted towards a specific peptide sequence as well as the glycan modifying the peptide sequence[113]. However, antibodies specifically targeting just specific glycan structures are available and, indeed, autoantibodies against glycans are found in humans[114]. These glycans often contain charged sialic acid or glucuronic acid carbohydrate moieties and some of the antibodies targeting such species have been used to monitor and predict tumor progression and malignancy[115,116]. For example, monoclonal antibodies directed towards Sialyl Lewis^x(Figure 1.1) are valuable research tools and have been used to predict the invasive nature and metastatic outcome of primary bladder carcinomas by monitoring the expression of Sialyl Lewis^x on glycosphingolipids from tumors[117]. The human natural killer-1 (HNK-1) carbohydrate epitope (Figure 1.13) is an adhesion molecule and is recognized as an important mediator of molecular recognition in the normal development of the nervous system[110,118]. It is expressed on recognition molecules such as immunoglobulins and proteoglycans[110]. This important carbohydrate epitope was first identified by immunoaffinity purification using the HNK-1 antibody[111]. The successful application of carbohydrate specific antibodies as tools for detection and analysis has prompted multiple groups to develop techniques for increasing the immunogenicity of carbohydrate based antigens[119,120]. These new techniques should enable development of a broader array of carbohydrate specific antibodies that should spur research in glycobiology.



HNK-1 carbohydrate epitope

Figure 1.13: HNK-1 carbohydrate epitope

The HNK-1 carbohydrate epitope is the tetrasaccharide sulphate-3-D-GlcA-β1-3-D-Gal-β1-4-D-GlcNAc-β1-3-D-Gal-β1-4-D-Glc

1.4.3 Metabolic labeling

Metabolic labeling of a wild type cell line or a mutant cell line with radiolabeled saccharides can be used to identify the subcellular localization of glycans through biochemical fractionation[121-123] or the tissue specific localization of a specific glycan or a group of glycans containing the radiolabeled saccharide[124]. Mutant cell lines with genetic deficiencies can be used to prevent metabolic conversion and incorporation of the labeled saccharides into certain glycan structures[13,125]; thus ensuring the label is favourably incorporated into the targeted glycan[126]. Mutant cell lines are particularly useful in cases where the labeled carbohydrate can be readily metabolically converted into other carbohydrates and/or incorporated into multiple glycans in wild type cells, which would make it difficult to track the specific glycans of interest. A more recent metabolic labelling methodology involves the incorporation of unnatural carbohydrates into the carbohydrate biosynthetic pathway. Saxon and Bertozzi showed that modification of the *N*-acyl side chain of *N*-acetyl-D-mannosamine (ManNAc), the physiologoical precursor to neuraminic acid (NeuAc), with an azido functionality (*N*-azidoacetyl mannosamine, (ManNAz))[127] leads to its conversion to *N*-azidoacetyl neuraminic acid (NeuAz) and incorporation into cell surface sialoglycoconjugates. Labelling was demonstrated using a bio-orthogonal reaction known as the Staudinger ligation to attach a biotinylated probe to the cell surface azide groups. Vocadlo *et al.* were able to expand the metabolically incorporated azido sugar library with *N*-azidoacetyl-D-glucosamine (GlcNAz) and

show that this saccharide was incorporated into *O*-GlcNAc[128] modified glycoproteins (dicussed in more detail below)[129]. Various chemoselective probes that react with azide groups have since been developed and these have been used to identify proteins bearing specific glycans, to aid glycoconjugate purification, or enable glycan localization studies, and even study developmental glycan dynamics[130].

This thesis will discuss efforts to generate selective inhibitors of glycosidases for use as research tools, coupled with other techniques such as LFERs and X-ray crystal structure analysis, to probe the mechanism of certain glycosidases. This thesis will also describe the synthesis of glycans of interest and look at their use in the development of carbohydrate targeted antibodies to detect particular glycans.

1.5 Two unusual forms of *O*-glycosylation

Two forms of *O*-glycosylation are the main focus of this thesis, the first of which is a unique form of *O*-glycosylation that is simply a monosaccharide, and the second is a structurally ambiguous trisaccharide modifying essential signalling proteins. These two unusual *O*-glycans and the enzymes and proteins that they interact with will be discussed in greater detail below.

1.5.1 *O*-GlcNAcylation

In 1984 Torres and Hart made the remarkable discovery that proteins in the nucleus and cytoplasm were modified by *O*-linked *N*-acetylglucosamine (*O*-GlcNAc)[128]. *O*-GlcNAc is estimated to modify over 1000 different proteins in metazoans and involves addition of only a single unit of GlcNAc. No elongation to generate a disaccharide or oligosaccharides has ever been found although one report has shown that *O*-GlcNAc can be phosphorylated[131]. Furthermore, the *O*-GlcNAc modification of proteins, which is installed by *O*-GlcNAc transferase (OGT) using UDP-GlcNAc as the donor substrate[132], is a dynamic process and can be cycled on and off of proteins during their lifetime. *O*-GlcNAcase (OGA), a glycosidase, cleaves *O*-GlcNAc off from proteins. (Figure 1.14)[133,134]. This cycling

differs from most forms of glycosylation which remain covalently attached to the protein until the protein and glycan are broken down in the lysosome.

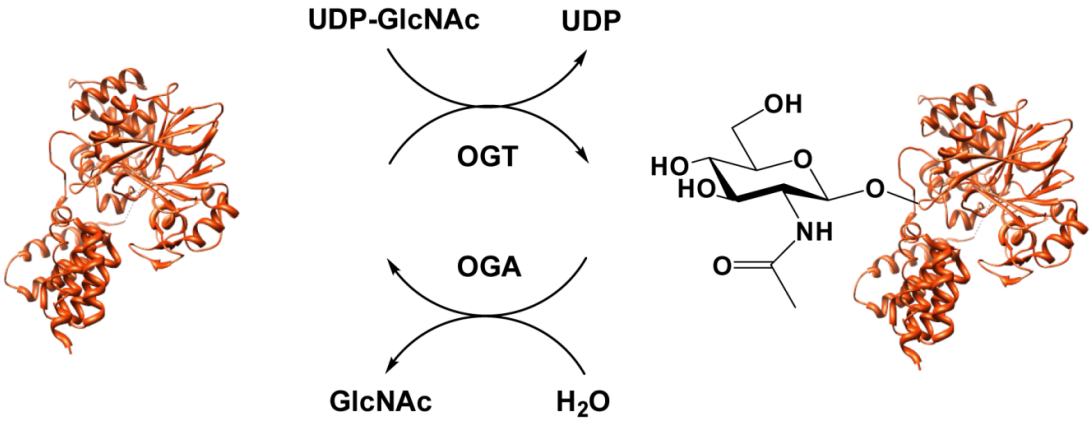


Figure 1.14: The *O*-GlcNAc post-translational modification

The dynamic *O*-GlcNAc modification is installed by *O*-GlcNAc transferase (OGT), which uses UDP-GlcNAc to modify proteins with *O*-GlcNAc. *O*-GlcNAcase (OGA) cleaves *O*-GlcNAc off of proteins to return them to their unmodified state.

1.5.2 Significance of *O*-GlcNAcylation of proteins

O-GlcNAc is synthesized within eukaryotic cells from glucose via the hexosamine biosynthetic pathway (HBSP). In 1991 Marshall *et al.* demonstrated that increased flux through the HBSP leads to the onset of insulin resistance[135]. This led to a proposal that the *O*-GlcNAc modification is involved in nutrient sensing, and that the insulin desensitizing effects of hyperglycemia[136] stem from sustained high blood glucose levels that cause increased *O*-GlcNAc levels and so lead to type II diabetes[137]. To test this theory, multiple research groups have used PUGNAc to inhibit OGA in cultured cells and in tissues ex vivo, thus pharmacologically raising *O*-GlcNAc levels[138-140]. Increased *O*-GlcNAc levels due to treatment with PUGNAc induced insulin resistance in 3T3-L1 adipocytes[139], rat skeletal muscle[138], and rat primary adipocytes[141]. However two functionally related *N*-acetylglucosaminidases, lysosomal *β*-hexosaminidase A (HexA), and B (HexB), are possibly also inhibited by PUGNAc since these enzymes also process terminal *β*-linked GlcNAc from glycoconjugates, including gangliosides,

which have been implicated in insulin resistance. Genetic inactivation of HexA and HexB results in an accumulation of ganglioside GM2 within lysosomes and has proven deleterious effects in mammals[142,143]. Therefore, drawing conclusions on the basis of inhibition of OGA by PUGNAc without assessing whether functionally related HexA and HexB enzymes are inhibited by this compound is somewhat risky. Therefore selective inhibitors of OGA are needed to clarify the biological impact of inhibiting OGA and increasing *O*-GlcNAc levels before functional conclusions can be made.

Another possibility for the role of *O*-GlcNAc modification stems from a proposal that cells respond to a variety of cellular stresses such as heat, oxidation and UV light based stresses by elevating their *O*-GlcNAc levels[144]. Blocking *O*-GlcNAcylation of proteins, or reducing it, renders cells more sensitive to stress and results in decreased cell survival. In contrast, increased *O*-GlcNAc levels on proteins protects cells against these types of stresses. In any event *O*-GlcNAc is of high interest, yet the role played by this post-translational modification is still unclear in many cases despite the growing body of research focusing on this unusual monosaccharide modification. I was therefore attracted to this promising area of research and aimed to contribute through the development of new chemical biology tools.

1.5.3 *O*-Glucosylation

O-linked glucose is a non-canonical form of protein glycosylation that was first identified in 1988 as a modification decorating the epidermal growth factor (EGF)-like domains of bovine blood coagulation factors VII and IX[145]. The *O*-linked glucose residue was found attached to the hydroxyl group of serine residues; it was also revealed that glucose was elongated by two xylose residues to form a trisaccharide that appeared to be D-Xyl-*α*1-3-D-Xyl-*α*1-3-D-Glc (XXG)[146]. Nishimura and coworkers proposed a consensus sequence governing which serine residues become modified: Cys^1-Xxx-Ser-Xxx-Pro-Cys^2, where Cys^1 and Cys^2 represent the first and second conserved cysteine residues of the EGF-like repeats and Xxx represents any amino acid[147]. More recent work by the

Haltiwanger group has shown that mammalian Notch is also modified with *O*-glucose[148]. Similar to the trisaccharide identified by Hase and coworkers, this non-canonical *O*-glucosylation of mammalian Notch was also elongated to a trisaccharide, however the identity of the two terminal carbohydrates and the linkages that bound them to glucose was not determined. The physiological or molecular role played by this glycan present on Notch was entirely unknown, however, Notch is an important cell surface receptor that plays key roles in development and cancer, which makes this modification of significant interest.

1.5.4 The Notch signalling pathway and *O*-glucosylation

The Notch signalling pathway is an essential pathway for the proper development of all metazoans[149]. Abnormal Notch signaling has been linked to developmental disorders of the kidney, liver, heart, eye, and skeleton,[150,151] and has also been implicated in human diseases such as cancer[152]. Notch, a large transmembrane protein and essential component of the notch signalling pathway, was named after a phenotype described in *Drosophila* over 90 years ago which showed irregular “notches” in the tips of the wings[153]. The protein comprises an intracellular domain, a transmembrane domain, and a extracellular domain. The intracellular domain (ICD) consists of a recombining binding protein associated module (RAM) domain, six ankyrin repeats and a C-terminal region rich in proline (P), glutamine (E), serine (S), and threonine (T) (PEST domain)(Figure 1.15)[154,155]. The transmembrane domain spans the cell membrane and links together the intracellular and extracellular domains. The large extracellular domain is composed of between 29 and 36 EGF-like repeats, and a Notch regulatory region (NRR) composed of 3 cysteine rich Lin12/Notch repeats and a region that links to the transmembrane domain (Figure 1.15)[149,156]. There are four Notch homologues in mammals (Notch1-4) that can be activated by different ligands binding to the extracellular domain of Notch[157].

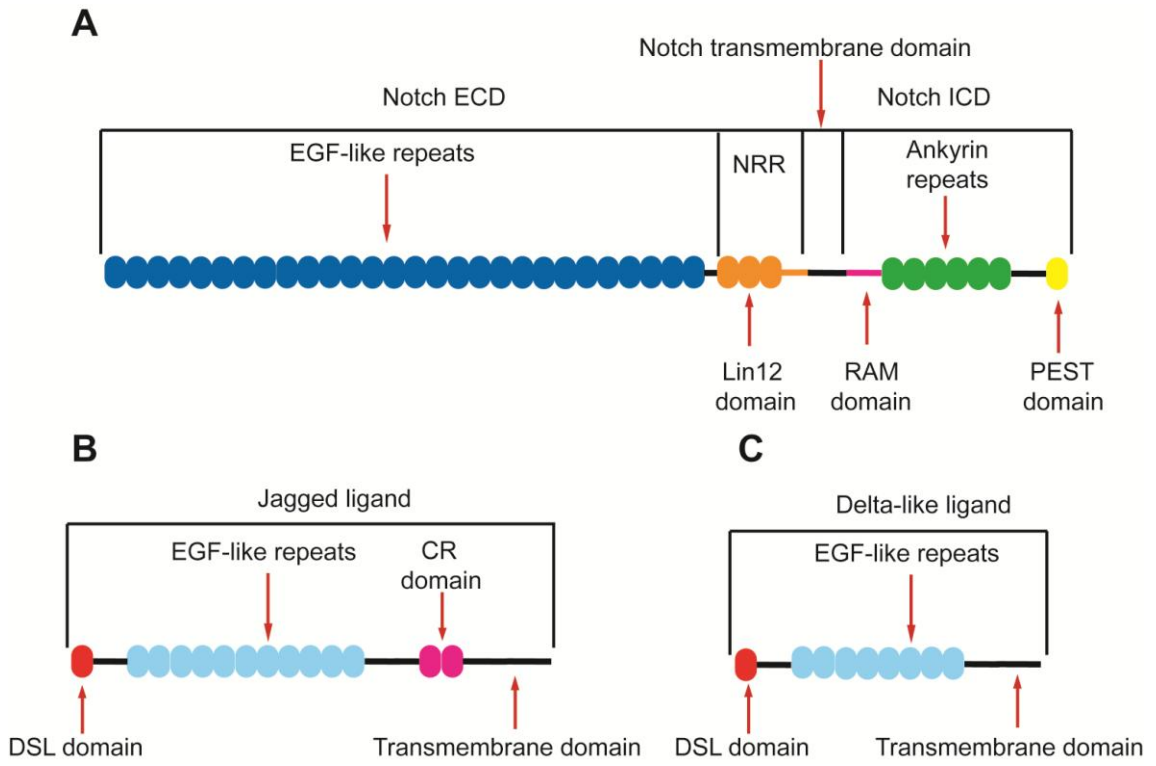


Figure 1.15: Notch, Delta, and Jagged

Diagram of Notch and its ligands A) The Notch transmembrane protein is composed of: EGF-like repeats (dark blue), a Notch regulatory region (NRR, orange) composed of three Lin12/Notch repeats and a small linker bound to the transmembrane domain, a transmembrane domain, a recombining binding protein associated module (RAM) domain (pink), ankyrin repeats (green), and a PEST domain (yellow).The EGF-like repeats are involved in ligand binding, the NRR prevents ADAM10 cleavage at the S2 site until ligand binding occurs, the RAM domain interact with transcription factors, the ankyrin repeats bind to transcription factors, and the PEST domain is involved in regulating proteolytic degradation of Notch.[155,156] B) Jagged is composed of a: Delta Serrate Lag (DSL) domain (red) that is essential for interactions with the Notch receptor, EGF-like repeats (light blue), cysteine rich domains (CR, purple), and a transmembrane domain. C) Delta-like ligand is composed of: a DSL domain (red), EGF-like repeats (light blue), and a transmembrane domain.

There are five mammalian protein ligands that can initiate the Notch signalling pathway, Delta-like 1, 3, and 4 and Jagged 1 and 2 (Figure 1.15). Similar to Notch, these ligands are also transmembrane proteins and are composed primarily of EGF repeats. Upon ligand binding the conformation of the Notch regulatory region is altered[158,159], thus allowing cleavage to occur (Figure 1.16)[160]. Proteolytic cleavage at the S2 site in the extracellular domain is performed by ADAM10, a membrane bound protein with A Disintegrin And Metalloprotease domain, this separates the extracellular domain from the

transmembrane domain [161-163]. The γ-secretase complex catalyzes a second cleavage at the S3 site of the transmembrane domain of Notch[160]. This releases the intracellular domain which is subsequently translocated to the nucleus where it interacts with transcription factors and so activates the expression of various target genes[164-166].

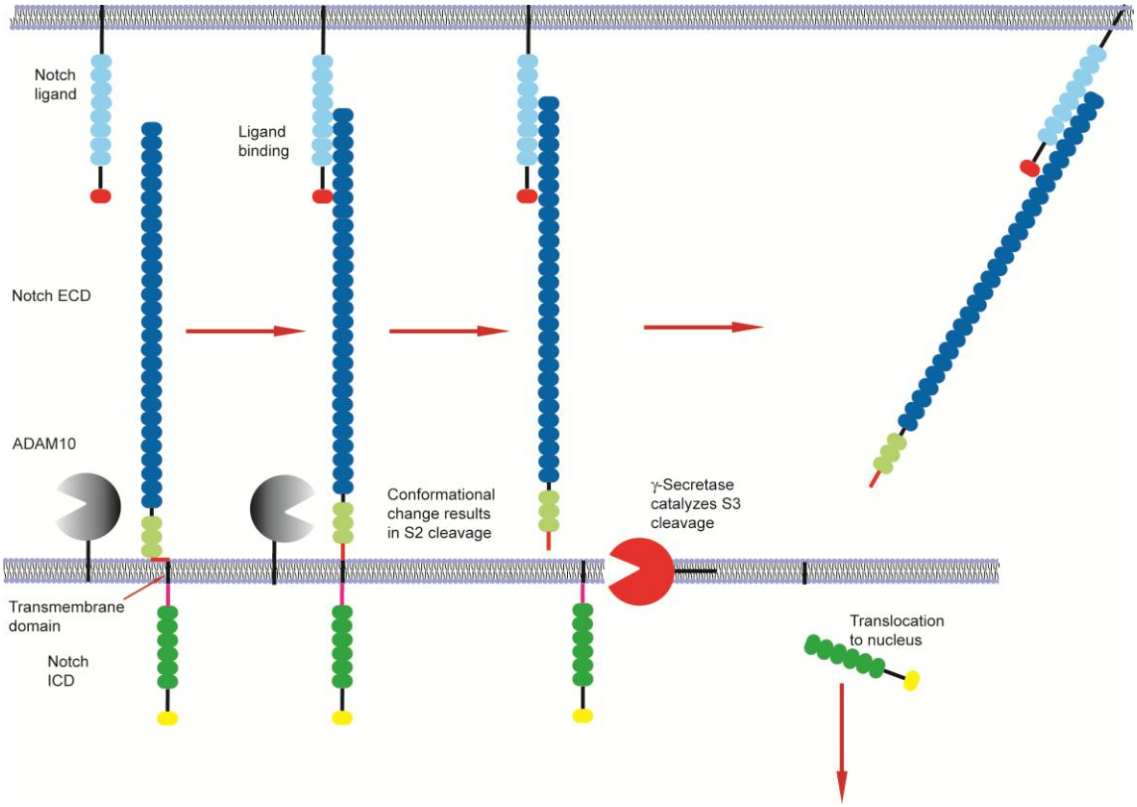


Figure 1.16: The Notch signalling pathway

Prior to ligand binding Notch is not susceptible to ADAM10 catalyzed S2 cleavage, upon binding with a notch ligand (Delta 1,3, or 4 or Jagged 1 or 2) the conformation of the NRR is altered making Notch susceptible to ADAM10 catalyzed S2 cleavage. The ECD is released from the cell membrane and subsequent S3 cleavage catalyzed by the γ-secretase complex releases the intracellular domain. The intracellular domain is then translocated to the nucleus.

It is already known that the Notch signaling pathway can be modulated by a different form of *O*-glycosylation of the EGF-like repeats of the Notch extracellular domain (ECD). *O*-Fucose is added by protein *O*-fucosyltransferase-1 (Pofut1)[167] to Ser/Thr residues within the consensus sequence Cys^2-Xxx-Xxx-Xxx-Xxx-(Ser/Thr)-Cys^3 found within the EGF-like repeats of Notch[168]. While the

role of Notch *O*-fucosylation is essential to proper Notch function, the precise mechanism has recently become quite convoluted[166,169]. However, glycosylation of *O*-fucosylated Notch by Fringe, an *N*-acetylglucosaminyl (GlcNAc) transferase, is known to regulate ligand binding. Addition of GlcNAc to the *O*-fucose residues catalyzed by Fringe enhances Delta binding to Notch and decreases Serrate (a *Drosophila* homologue of Jagged) binding, thus glycosylation modulates the interactions of Notch with its ligands[170,171].

O-Glucosylation is a more recently discovered form of Notch *O*-glycosylation, however, its role in the Notch signaling pathway or elsewhere is not apparent and little research effort has been undertaken since its initial discovery in 1988[145]. In 1991 Bjoern and coworkers designed a Ser to Ala mutant of Factor VII, a vitamin K dependent plasma glycoprotein involved in the blood coagulation pathway. This mutation prevented the XXG modification of Factor VII and led to a decrease of approximately 60% of the coagulation ability of mutant Factor VII as compared to wild type Factor VII[172]. While nothing additional has been discovered involving the XXG modification of the blood coagulation proteins Factor VII, Factor IX, or protein Z, glucose modification of Notch continues to be investigated. This effort has been stimulated by the realization that Fringe-mediated GlcNAc modification of *O*-fucose on Notch regulates its function in the Notch signaling pathway[12,13,173,174]. The recent discovery by Acar *et al*. that the transferase responsible for adding *O*-glucose to Notch in *Drosophila* is encoded by the gene *Rumi*, has greatly stimulated interest in the *O*-glucose modification of Notch[175]. Acar *et al*. showed that loss of Rumi *O*-glucosyltransferase activity leads to a temperature sensitive Notch phenotype, which points to glucosylation of Notch playing a critical role in proper Notch structure or function[166,175,176]. To enable research in the area new tools are needed to both fully elucidate the structure of the Notch *O*-glucose trisaccharide and to aid in defining its role in the proper functioning of the Notch signaling pathway.

1.6 Aims of this thesis

Despite being the most abundant biological building block of life carbohydrates are arguably the least well understood. This is due to early perceptions of carbohydrates as being simply a source of energy, as well as structural component of various organisms. In the last couple of decades the realization that carbohydrates play more critical roles in diverse biological processes than had been anticipated has dawned on the scientific community. Carbohydrates are now implicated in blood group determination[2-4], inflammation[2-4], protein folding[5-8], adhesion[110], cell-cell signaling[170,171], nutrient sensing[136], regulation of cellular stress[144], tumour progression[115-117], as well as bacterial virulence factors[177-179], and being involved in viral replication[83,84]. This enlightenment is due in part to the development of carbohydrate specific tools and methods of analysis, which has enabled much research in the carbohydrate and glycobiology fields. To continue furthering our understanding of carbohydrates and their biological roles, new tools and methods are critically important. The aims of this thesis are to elucidate the catalytic mechanism of mammalian *O*-GlcNAcase and to use this information to inform and develop selective and potent inhibitors of this enzyme. A goal is that these inhibitors should be selective for *O*-GlcNAcase over the functionally related lysosomal hexosaminidases. I will also look at the *O*-glucosylation of Notch with the aim of fully characterizing the trissacharide glycan structure and developing polyclonal antibodies to aid studies on the cellular function of this modification.

2: Human *O*-GlcNAcase uses a catalytic mechanism involving substrate-assisted catalysis: kinetic analysis and development of highly selective, cell permeable, mechanism-based inhibitors.

The manuscript below is reprinted from:

Macauley, M. S.; Whitworth, G. E.; Debowski, A. W.; Chin, D.; Vocadlo, D. J., *O*-GlcNAcase uses substrate-assisted catalysis: Kinetic analysis and development of highly selective mechanism-inspired inhibitors. *Journal of Biological Chemistry* **2005,** 280, (27), 25313-25322.

Updates in the field relating to this Chapter can be found in appendix 1. The designated numbers given to synthetic or target compounds in this Chapter relate only to this Chapter.

2.1 Contributions

I synthesized the thiazoline series of inhibitors and worked with Matthew Macauley to analyze the data. Matthew Macauley tested the inhibitors and fluorinated substrates and worked with Alex Debowski to express the enzymes. Danielle Chin synthesized the fluorinated substrates.

2.2 Abstract

The post-translational modification of serine and threonine residues of nuclear and cytosolic proteins with the saccharide 2-acetamido-2-deoxy-D-glucopyranose (GlcNAc) is a reversible process implicated in protein stability, cellular stress, and signaling. *O*-GlcNAcase, the enzyme catalyzing the cleavage of β-*O*-linked GlcNAc (*O*-GlcNAc) from modified proteins, is a member of the family 84 glycosidases. The family 20 β-hexosaminidases bear no apparent sequence similarity yet are functionally related in so far as both cleave terminal GlcNAc residues from glycoconjugates. The family 20 human lysosomal β-hexosaminidase, is known to use substrate-assisted catalysis involving the 2-acetamido group of the substrate, however, the catalytic mechanism of human *O*-GlcNAcase is unknown. In order to probe the catalytic mechanism of *O*-GlcNAcase a series of 4-methylumbelliferyl 2-deoxy-2-acetamido-β-D-glucopyranoside substrates with varying degrees of fluorine substituted *N*-acyl derivatives were prepared. A Taft-like linear free energy analysis of the Taft parameter (σ*) of the *N*-acyl group of each substrate plotted against the logarithm of the second order rate constants measured for both enzymes indicates that *O*-GlcNAcase, like β-hexosaminidase, uses a catalytic mechanism involving anchimeric assistance. A bicyclic thiazoline (NAG-thiazoline) that mimics the putative oxazoline intermediate found in the catalytic mechanism of family 20 glycosidases is also shown to be a potent competitive inhibitor of both *O*-GlcNAcase (K_I = 70 nM at pH 6.5) and β-hexosaminidase (K_I = 70 nM at pH 4.25). A series of NAG-thiazoline analogues were prepared and tested in order to develop a potent and selective inhibitor of *O*-GlcNAcase. The most selective inhibitor demonstrated a remarkable 1500-fold selectivity for *O*-GlcNAcase (K_I = 230 nM) over β-hexosaminidase (K_I = 340 000 nM). These inhibitors are cell permeable and act to modulate the activity of *O*-GlcNAcase in tissue culture. Because both O-GlcNAcase and β-hexosaminidase have vital roles in organismal health, these potent and selective inhibitors of *O*-GlcNAcase should

prove useful in studying the role of this enzyme at the organismal level without generating a complex chemical phenotype stemming from concomitant inhibition of β-hexosaminidase.

2.3 Introduction

Humans have three genes encoding enzymes that cleave terminal *N*-acetylglucosamine residues from glycoconjugates. The first of these, *O*-GlcNAcase, is a member of family 84 of glycosidases that includes enzymes from organisms as diverse as prokaryotic pathogens and humans[180,181]. The substrates of *O*-GlcNAcase are post-translationally modified glycoproteins bearing the monosaccharide 2-acetamido-2-deoxy-β-D-glucopyranoside linked to serine or threonine residues (*O*-GlcNAc)[128,182,183]. This post-translational modification is abundant in mammalian cells[128] and is found on many cellular proteins having a wide range of vital cellular functions including, for example, transcription[184-187], proteasomal degradation[188], cellular signaling [139], and is also found on many structural proteins[189-191]. Consistent with the abundance of *O*-GlcNAc, it appears to have roles in the etiology of several disease states including type II diabetes,[139,192] Alzheimer's[190,193,194], as well as cancer[195]. The mammalian enzymes appear to be bifunctional, acting to cleave the *O*-GlcNAc from the serine and threonine residues of proteins[196] as well as having histone acetyltransferase activity[197]. *O*-GlcNAcase has been cloned[198] and partially characterized[199] but little is known about its catalytic mechanism.

HEXA and *HEXB* are the two other genes that encode enzymes catalyzing the hydrolytic cleavage of terminal *N*-acetylglucosamine residues from glycocolipids. The gene products of *HEXA* and *HEXB* predominantly yield two dimeric isozymes. The homodimeric isozyme, hexosaminidase B (ββ) is composed of two β-subunits, and the heterodimeric isozyme, hexosaminidase A (αβ) is composed of an α- and a β-subunit. Both of these enzymes are normally localized within the lysosome. The two subunits bear a high level of sequence identity and both are members of family 20 of glycosidases. The dysfunction of either of these isozymes results in the accumulation of gangliosides and glycoconjugates within the lysosome causing the inheritable neurodegenerative disorders known as Tay-Sach's and Sandhoff's disease[143]. Dysfunction of either of these subunits modulates ganglioside levels at the organismal level, the

deleterious effects of which are still being uncovered[200]. In contrast, the consequences of genetic deletion of *O*-GlcNAcase remain unknown.

As such, small molecule inhibitors of glycosidases, including β-*N*-acetyl-glucosaminidases,[43,44,201,202] have received a great deal of attention[41] both as tools for elucidating the role of these enzymes in biological processes as well as in developing therapeutic interventions with minimal side effects. Indeed, the control of glycosidase function using small molecules offers several advantages over genetic knockout studies including the ability to rapidly vary dose or entirely withdraw treatment. A major hurdle, however, in the generation of selective inhibitors of human glycosidases is that multiple functionally related enzymes are often present within a given organism. The promiscuity of these inhibitors and the resultant inhibition of multiple enzymes render them of limited utility in studying the cellular and organismal physiological role of one particular enzyme since they necessarily generate complex phenotypes. In the case of β-*N*-acetylglucosaminidases no potent inhibitor is currently known that is selective for nucleocytoplasmic *O*-GlcNAcase over the lysosomal β-hexosaminidases.

Two existing inhibitors of *O*-GlcNAcase, however, have enjoyed use in studies of the *O*-GlcNAc post-translational modification within both cells and tissues[138,139,203-205]. Both of these compounds, however, suffer from having secondary effects within the cellular and organismal environment. Streptozotocin (STZ) has long been used as a diabetogenic compound because it has a particularly detrimental effect on β-islet cells[206]. STZ exerts its cytotoxic effects through both the alkylation of cellular DNA[206,207] as well as the generation of radical species including nitric oxide[208]. The resulting DNA strand breakage promotes the activation of poly(ADP-ribose) polymerase (PARP)[209] with the net effect of depleting cellular NAD^+ levels and, ultimately, cell death[210,211] Other investigators have proposed instead that STZ toxicity is a consequence of the irreversible inhibition of *O*-GlcNAcase which is highly expressed within β-islet cells[203,212]. This hypothesis has, however, been brought into question by two independent research groups[213,214]. Because cellular *O*-GlcNAc levels increase in response to many forms of cellular stress[144] it seems possible that STZ results

in increased *O*-GlcNAc levels by inducing cellular stress rather than through any specific and direct action on *O*-GlcNAcase. Indeed, STZ has been shown to act as a poor, and somewhat selective, inhibitor of *O*-GlcNAcase[215] over β-hexosaminidase and although it has been proposed that this compound acts to irreversibly inhibit *O*-GlcNAcase there has been no clear demonstration of this mode of action[216].

The second compound that has enjoyed use as an inhibitor of *O*-GlcNAcase is *O*-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-*N*-phenylcarbamate (PUGNAc, Chapter 1, Figure 1.12). The groups of Vasella and Rast described the potent inhibitory properties of PUGNAc on β-*N*-acetylglucosaminidases from *Canavalia ensiformis*, *Mucor rouxii,* and the family 20 β-hexosamindase from bovine kidney[43]. This compound has also been shown to be a highly potent inhibitor of mouse β-hexosaminidase[217]. Studies by Haltiwanger and coworkers have more recently demonstrated that PUGNAc is also a potent inhibitor of the family 84 *O*-GlcNAcase that lacks the cytotoxic effects of STZ[134]. Despite its apparent lack of short-term toxicity, this compound is a potent inhibitor of both the family 20 and family 84 glycosidases and therefore *in vivo* studies using this compound will yield a complex phenotype stemming from the inhibition of both of these functionally related enzymes.

Given these considerations, we felt that the development of a potent, cell permeable, inhibitor selective for *O*-GlcNAcase over the lysosomal hexosaminidase would be a valuable tool in studying the physiological role of *O*-GlcNAc at the cellular and organismal level. A logical starting point for the design of such an inhibitor for *O*-GlcNAcase would take into consideration the catalytic mechanism of these two functionally related enzymes. Although the catalytic mechanism of action of the family 20 human hexosaminidase A and B have been established[218] that of the family 84 *O*-GlcNAcase remains unknown. We therefore aimed to firstly elucidate the catalytic mechanism of human *O*-GlcNAcase and secondly, to use this information in designing simple inhibitors that would be potent, cell permeable, and highly selective for *O*-GlcNAcase over the lysosomal β-hexosaminidase.

2.4 Results and discussion

2.4.1 Comparative analysis of the catalytic mechanisms of *O*-GlcNAcase and β-Hexosaminidase.

Three realistic mechanistic alternatives exist for *O*-GlcNAcase and family 84 glycosidases. The first alternative is an inverting mechanism (Figure 2.1A) such as that found for α-*N*-acetyl-galactosaminidase[219] from family 27 of glycosidases. This catalytic mechanism involves the base-catalyzed nucleophilic attack of water at the anomeric centre concomitant with the acid-catalyzed departure of the leaving group[219]. The second mechanistic possibility is the canonical two step double displacement mechanism that results in retention of configuration at the anomeric centre (Figure 2.1B). This mechanism is used by most retaining β-glycosidases and involves, in the first step, attack of an enzymic nucleophile at the anomeric centre with the resulting formation of a transient covalent glycosyl enzyme intermediate[220]. Departure of the aglycon leaving group is facilitated by an enzymic general acid/base catalytic residue. In the second step the enzymic acid/base catalytic residue facilitates the attack of a water molecule at the anomeric centre, cleaving the intermediate to liberate the regenerated enzyme and the hemiacetal product with retained stereochemistry. β-*N*-Acetylglucosaminidases from family 3 of glycosidases have been shown to use this mechanism as have the C-type lysozymes from family 22[221]. The third mechanistic alternative involves the nucleophilic participation of the 2-acetamido group of the substrate in place of an enzymic catalytic nucleophile (Figure 2.1C). This last mechanistic option is exploited by β-hexosaminidases from family 20 of glycosidases[44,64,222]. These three mechanisms differ in several aspects. The inverting mechanism is a one step reaction that results in the formation of a product with inverted stereochemistry. The other two alternatives are retaining in stereochemistry and differ from each other primarily in the nature of the intermediate; in the second mechanism this species is a covalent enzyme adduct

whereas in the third case it is believed to be a bicyclic oxazoline or oxazolinium ion.

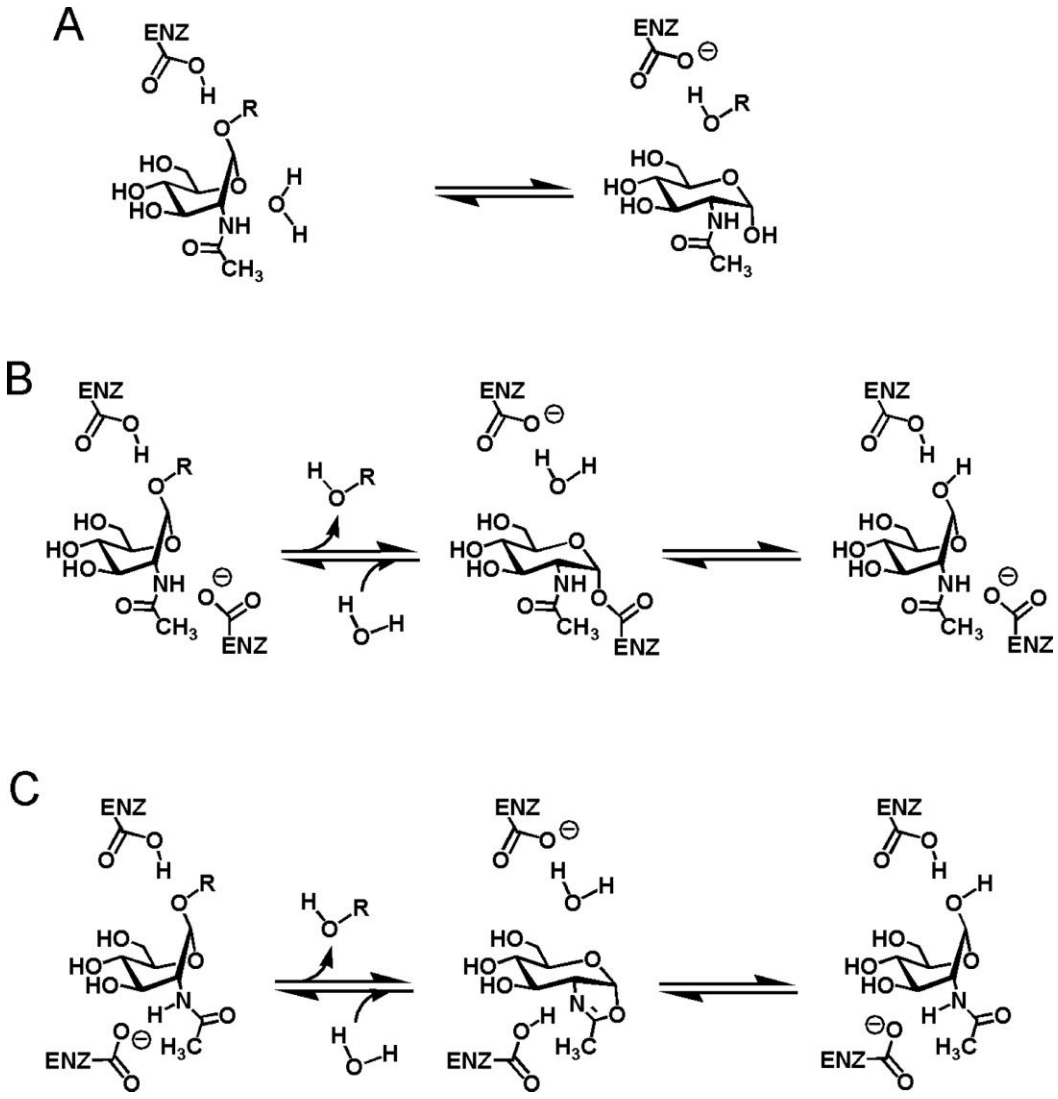


Figure 2.1: Three possible catalytic mechanisms for *O*-GlcNAcase

Pathway A; the single step inverting mechanism; pathway B; the double displacement retaining mechanism involving formation and breakdown of a covalent glycosyl enzyme intermediate; pathway C; the double displacement retaining mechanism involving formation and breakdown of a bicyclic oxazoline intermediate.

Table 2.1: Michaelis-Menten parameters for the β-hexosaminidase and O-GlcNAcase-catalyzed hydrolysis of a series of 4-methylumbelliferone 2-N-acetyl-2-deoxy-β-D-glucopyranosides

Substrate	σ*[a]	Enzyme	K_M (mM)	$V_{max}/[E]_T$ (μmol min^{-1} mg^{-1})	$V_{max}/K_M[E]_T$ (μmol $mM^{-1}min^{-1}$ mg^{-1})
MuGlcNAc (**5**)	0.0	Hexosaminidase	0.88 ± 0.05	8.3 ± 0.2	9.5 ± 0.8
		O-GlcNAcase	0.43 ± 0.04	0.74 ± 0.02	1.7 ± 0.2
MuGlcNAc-F_1 (**5a**)	0.8	Hexosaminidase	3.8 ± 0.4[b]	3.3 ± 0.2[b]	0.40 ± 0.07[d]
		O-GlcNAcase	0.49 ± 0.06	0.60 ± 0.03	1.2 ± 0.2
MuGlcNAc-F_2 (**5b**)	2.0	Hexosaminidase	ND[c]	ND[c]	0.069 ± 0.001[d]
		O-GlcNAcase	0.45 ± 0.05	0.16 ± 0.01	0.36 ± 0.06
MuGlcNAc-F_3 (**5c**)	2.8	Hexosaminidase	ND[c]	ND[c]	0.0077 ± 0.001[d]
		O-GlcNAcase	0.38 ± 0.03	0.033 ± 0.001	0.0077 ± 0.0008

[a]Values used were obtained from Hansch and Leo[223].
[b]Values were estimated by non-linear regression of the Michelis-Menten data. Note that substrate concentrations assayed did not exceed K_M due to limited substrate solubility.
[c]These values could not be determined as saturation kinetics were not observed owing to limitations in substrate solubility.
[d]Values were determined by linear regression of the second order region of the Michelis-Menten plot.

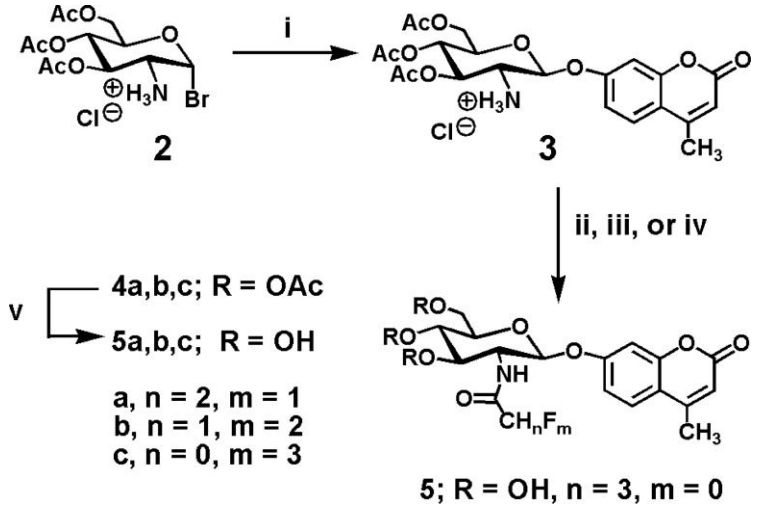
A key difference between these mechanisms is the involvement of the 2-acetamido group of the substrate. This moiety may actively participate in catalysis as a nucleophile, as for the lysosomal β-hexosaminidases, or it may act as a bystander (Figure 2.1). To address the role of the 2-acetamido group of the substrate we synthesized several substrate analogues bearing increasing levels of fluorine substitution on the *N*-acetyl group (Scheme 2.1). The electron withdrawing fluorine substituents decrease the basicity of the carbonyl group and the expected effect on catalysis would be to slow down the enzymatic reaction if it uses anchimeric assistance. These compounds were first tested with the lysosomal human β-hexosamindase since it is known to proceed *via* a mechanism involving anchimeric assistance (Figure 2.2A). Although Michaelian saturation kinetics were not observed for all substrates (Table 2.1) $V_{max}[E]_o/K_M$, which is proportional to the second order rate constant governing the enzyme catalyzed reaction, could be determined from the initial slope of the Michaelis-Menten plot (Figure 2.2A Inset). A plot of log $V_{max}[E]_o/K_M$ against the Taft

parameter (σ*) of the *N*-acyl substituent shows a negative linear correlation on increasing fluorine substitution (Figure 2.2C). As expected, decreasing the basicity of the carbonyl oxygen has a deleterious effect on catalysis. The steep negative slope of the Taft-like linear free energy analysis (ρ* = -1.0 ± 0.1) reveals that the carbonyl group is interacting with a cationic centre in the transition state. This data is consistent with a mechanism involving electrophilic migration of the anomeric centre and consequent oxocarbenium-ion like transition states that have been generally proposed for enzyme-catalyzed glycoside hydrolysis[220,221] For *O*-GlcNAcase, Michaelian saturation kinetics were observed for all four substrates and therefore both kinetic parameters were determined (Table 2.1, Figure 2.2B). For *O*-GlcNAcase the value of $V_{max}[E]_o/K_M$ is also dependant on the extent of fluorine substitution but the slope is more shallow (ρ* = -0.42 ± 0.08, Figure 2.2C). On the basis of these results it appears that *O*-GlcNAcase, in common with lysosomal β-hexosaminidase, uses a catalytic mechanism involving anchimeric assistance. The slope of the correlations in the Taft analysis can be considered a function of both an electronic component (σ*) and a steric component (E_s) according to the following equation:

$$\rho^* = x\sigma^* + yE_s \qquad \text{(Eqn 1)}$$

The difference between the slopes measured for lysosomal β-hexosaminidase and *O*-GlcNAcase may thus reflect the position of the transition state along the reaction coordinate or may indicate that the lysosomal β-hexosaminidase has a more sterically constrained active site architecture than does *O*-GlcNAcase. Indeed, a common popular misconception is that fluorine (1.47 Å Van der Waals radius and 138 pm C-F bond length) is often considered to have a negligible difference in size as compared to hydrogen (1.2 Å Van der Waals radius and 109 pm C-H bond length). Therefore, it is possible that unfavorable steric interactions between the substrate and the active site of human β-hexosaminidase may play an additional role, beyond electronics, in discriminating between the varying levels of fluorine substitution. Indeed, the

recent crystal structures of human hexosaminidase B revealed a carefully structured pocket that tightly nestles the acetamido group between three tryptophan residues[218,224]. For *O*-GlcNAcase, however, no structure is available but the relatively constant K_M values measured for all of the substrate analogues suggests that steric effects are not a major contributor and that the electronic effect of the *N*-acyl-fluorine substituents predominate. An earlier study on an isolated enzyme of unknown family using two *para*-nitrophenyl 2-acetamido-2-deoxy-β-D-glucopyanoside (pNP-GlcNAc) with either two or three fluorine substituents in the acetamido group yielded a ρ* of -1.41 ± 0.1. This value is greater than that found for the lysosomal β-hexosaminidase in this study (ρ* = -1.0 ± 0.1) suggesting that the enzyme Kosman and Jones studied from *Aspergillus niger* is more likely a member of family 20 of glycosidases than family 84.



Scheme 2.1: Synthesis of substrate derivatives
(***i***) ***a***, acetone, 4-MU, Na4-MU; ***b***, K_2CO_3, Et_2O; ***c***, HCl; (***ii***) ***a***, Dowex 50-H^+, $NaOOCCH_2F$; ***b***, DMF, Et_3N, Py, DCC, 4 °C; (***iii***) DMF, Et_3N, Py, DCC, $F_2HCCOOH$, 4 °C; (***iv***) DMF, Et_3N, Py, DCC, $(F_3CCO)_2O$, 4 °C; (***v***) ***a***, NaOMe, MeOH; ***b***, Dowex 50-H^+.

2.4.2 Inhibition of *O*-GlcNAcase and β-hexosaminidase with NAG-Thiazoline

As a further test of whether *O*-GlcNAcase uses a catalytic mechanism involving anchimeric assistance we tested the inhibitor NAG-thiazoline (**9a,** Scheme 2.2) with this enzyme. NAG-thiazoline, designed as a mimic of the bicyclic oxazoline intermediate, has been demonstrated to function as an inhibitor of the family 20 hexosaminidases[44,222]. Using *para*-nitrophenyl 2-acetamido-2-deoxy-β-D-glucopyranoside (pNP-GlcNAc (**5**)) as a substrate we found NAG-thiazoline to be a potent inhibitor of the family 84 human *O*-GlcNAcase and a clear pattern of competitive inhibition was observed (Figure 2.3). Non-linear regression revealed a K_I value of 180 nM at pH 7.4 and analysis using 4-methylumbelliferyl 2-acetamido-2-deoxy-β-D-glucopyranoside (MU-GlcNAc) revealed a K_I value of 70 nM at pH 6.5 (Table 2.2).

Table 2.2: Inhibition constants and selectivity of inhibitors for both O-GlcNAcase and β-hexosaminidase
All K_I values were determined using linear regression.

Compound	*O*-GlcNAcase K_I (μM)	β-Hexosaminidase K_I (μM)	Selectivity Ratio (β-Hex K_I / *O*-GlcNAcase K_I)
GlcNAc	1500	1200	0.8
STZ	1500	47000	31
9a	0.070	0.070	1
9b	0.12	32	270
9c	0.23	340	1500
9d	4.9	4600	450
9e	57	11000	100
9f	1.6	720	700
9g	5.7	4000	190

NAG-Thiazoline is therefore a potent inhibitor of *O*-GlcNAcase, binding approximately 21000-fold more tightly than the parent saccharide, GlcNAc (K_I = 1.5 mM) at pH 6.5. This potent inhibition may be attributed to the resemblance of NAG-thiazoline to a putative oxazoline intermediate or a structurally related transition state. Indeed, the observed inhibition data is similar to that measured

by us for the family 20 human lysosomal β-hexosaminidase (K_I = 70 nM, 17000-fold more tightly than GlcNAc for which the K_I value is 1.2 mM) and strongly supports the Taft-like analysis indicating that *O*-GlcNAcase, in common with the family 20 β-hexosaminidases, uses a catalytic mechanism involving anchimeric assistance. Indeed, glycosidases from families 18[225] and 56[226], which are *endo*-glycosidases acting to cleave oligosaccharide chains, have been shown to use a mechanism involving anchimeric assistance[227]. Thus families 18, 20, 56 and 84 are all retaining glycosidases using a catalytic mechanism involving anchimeric assistance from the acetamido group of the substrate. This is in marked contrast to retaining β-glycosidases from families 3 and 22 that also act on substrates bearing an acetamido group at the 2-position of their respective substrates[221].

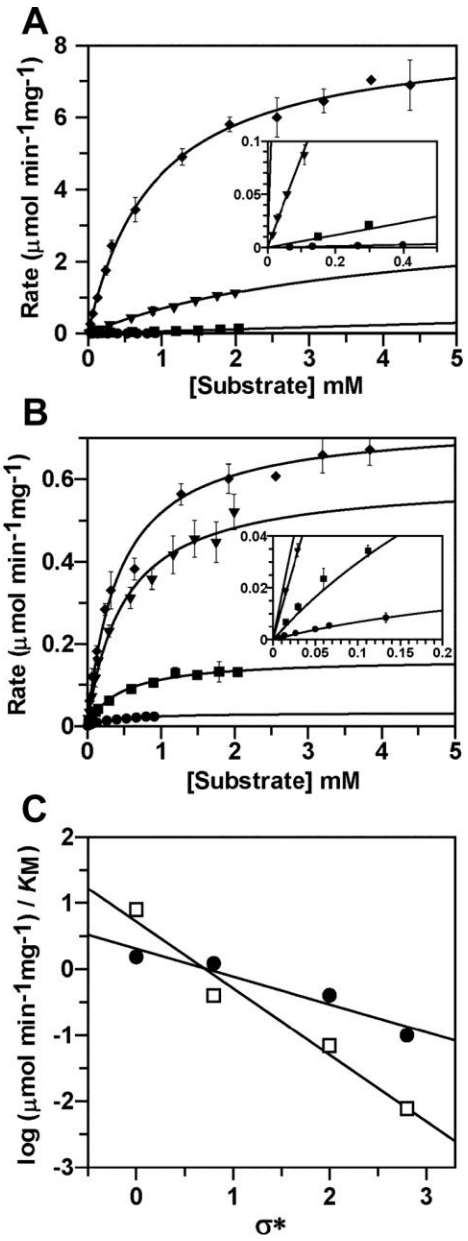


Figure 2.2: Activity of *O*-GlcNAcase and β-hexosaminidase with *N*-fluoroacetyl derivatives of MU-GlcNAc

A, initial velocities of the human β-hexosaminidase-catalyzed hydrolysis of *N*-fluoroacetyl derivatives of MU-GlcNAc; ♦, MU-GlcNAc (**5**); ▾, MU-GlcNAc-F (**5a**); ▪, MU-GlcNAc-F_2 (**5b**); •, MU-GlcNAcF_3 (**5c**). *Inset*, detail of the region of the plot at the intersection of the axes. *B*, initial velocities of the human *O*-GlcNAcase-catalyzed hydrolysis of *N*-fluoroacetyl derivatives of MU-GlcNAc; ♦, MU-GlcNAc (**5**); ▾, MU-GlcNAc-F (**5a**); ▪, MU-GlcNAc-F_2 (**5b**); •, MU-GlcNAcF_3 (***5c***). *Inset*, detail of the region of the plot at the intersection of the axes. *C*, linear free energy analysis plotting the Taft parameter (σ*) of the *N*-fluoroacetyl substituent of MU-GlcNAc substrate analogues against the log $V_{max}/[E]_0$ K_M values measured for each substrate as shown in *A* and *B* with *O*-GlcNAcase (•) and β-hexosaminidase (□).

2.4.3 Design, synthesis, and testing of selective *O*-GlcNAcase inhibitors

With the mechanism of human *O*-GlcNAcase, and by extension other members of family 84 of glycosidases, established we turned our attention to using this information in the design of inhibitors that would be selective for this enzyme over the human lysosomal hexosaminidase. Because both β-hexosaminidase and *O*-GlcNAcase use a mechanism involving anchimeric assistance we chose NAG-thiazoline as a scaffold that could be elaborated to generate the required selectivity. Three observations provided a starting point in the design of the inhibitor. The first is that the slope of the Taft-like analysis for the lysosomal enzyme was much steeper than that measured for *O*-GlcNAcase suggesting that the bulk of the *N*-acyl group may be a determinant in substrate recognition (*vide supra*). The second, and related, consideration is that the structure of the human lysosomal β-hexosaminidase B reveals a snug pocket into which the methyl group of the acetamido substituent is poised[218]. The third is that STZ, which bears a bulky *N*-acyl substituent shows some selectivity for *O*-GlcNAcase over β-hexosaminidase[215].

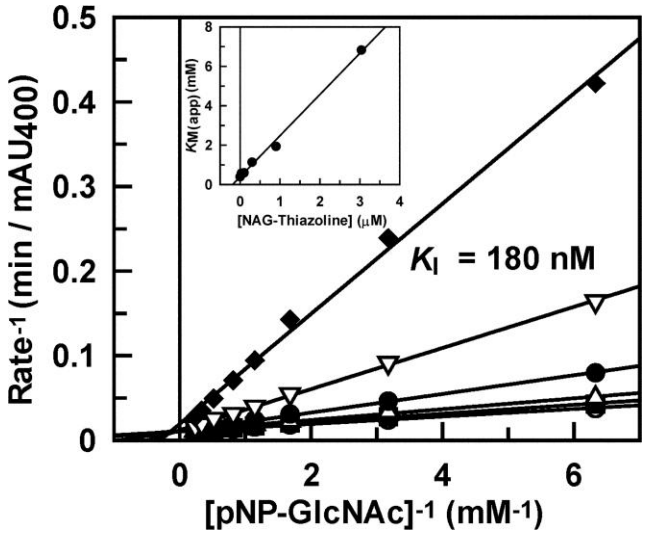
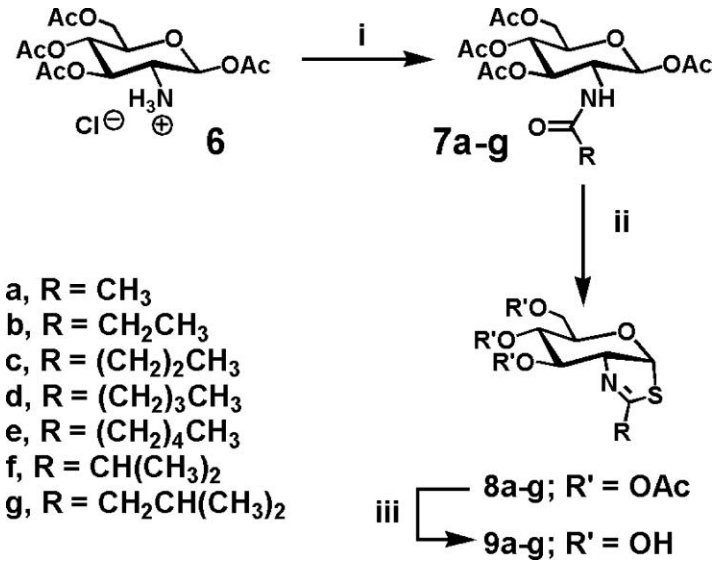


Figure 2.3: Inhibition of human *O*-GlcNAcase-catalyzed hydrolysis of *p*NP-GlcNAc (5) by NAG-thiazoline (9a) shows a pattern of competitive inhibition
The concentrations of **9a** (μm) used were 3.04 (♦), 0.900 (▽), 0.300 (•), 0.100 (△), 0.033 (▪), and 0.00 (○). *Inset*, graphical analysis of K_I from plotting K_M apparent against NAG-thiazoline (**9a**) concentration.



Scheme 2.2: Synthesis of a series of thiazoline based inhibitors
(*i*) RCOCl, Et_3N, CH_2Cl_2, 0 °C; (*ii*) Lawesson's reagent, *Tol* (toluene), Δ; (*iii*) *a*, NaOMe, MeOH; *b*, AcOH, MeOH.

We therefore prepared a series of seven inhibitors in which the thiazoline ring was elaborated with aliphatic chains of increasing length in the expectation that these compounds would allow the discriminative inhibition of *O*-GlcNAcase over lysosomal hexosaminidase. The synthesis of this panel of inhibitors is outlined in Scheme 2.2. This facile synthetic route enables the production of quantities of inhibitor from commercially available starting materials in three steps or from the inexpensive starting material 2-amino-2-deoxy-glucopyranose in six steps. Analysis of the inhibition of human β-hexosaminidase reveals that increasing the chain length resulted in a marked decrease in the potency of these inhibitors (Figure 2.4 and Table 2.2). The inclusion of even one methylene unit (compound **9b**) resulted in a 460-fold increase in the value of K_I for β -hexosaminidase as compared to the parent compound NAG-thiazoline (**9a**). Further increases in the chain length lead to still greater increases in K_I values. For *O*-GlcNAcase, however, the situation is markedly different (Figure 2.4 and Table 2.2). Increases in chain length are much better tolerated and the inclusion of two methylene units yields a compound (**9c**) for which the K_I value (K_I = 230 nM) is only 3-fold greater than that measured for the parent compound (**9a**) NAG-

thiazoline (K_I = 70 nM). Branching of the aliphatic chain also does not abrogate binding since both compounds **9f** and **9g** are good inhibitors of *O*-GlcNAcase. From analysis of the data it can be seen that compounds **9b, 9c** and **9f** are potent inhibitors of *O*-GlcNAcase and show a remarkable selectivity for *O*-GlcNAcase over lysosomal hexosaminidase. Indeed, the selectivity ratio for *O*-GlcNAcase over hexosaminidase is 1500-fold for compound **9c,** 700-fold for compound **9f**, and 270-fold for compound **9b** (Figure 2.4 and Table 2.2). The K_I values for STZ with these enzymes are unknown and so we determined these values (Table 2.2) for both *O*-GlcNAcase (K_I = 1.5 mM) and β-hexosaminidase (K_I = 47 mM). The selectivity of STZ for *O*-GlcNAcase over β-hexosaminidase is surprisingly modest given the bulk of the *N*-acyl group of this compound. Perhaps the thiazoline compounds demonstrate greater selectivity than STZ by virtue of the fact that they may emulate a transition state or tightly bound intermediate. Not surprisingly, GlcNAc showed no selectivity for *O*-GlcNAcase (K_I = 1.2 mM) as compared to β-hexosaminidase (K_I = 1.5 mM). The remarkably high selectivities of these thiazoline containing compounds should greatly facilitate finding the appropriate dose to obtain entirely selective inhibition of *O*-GlcNAcase at the organismal level.

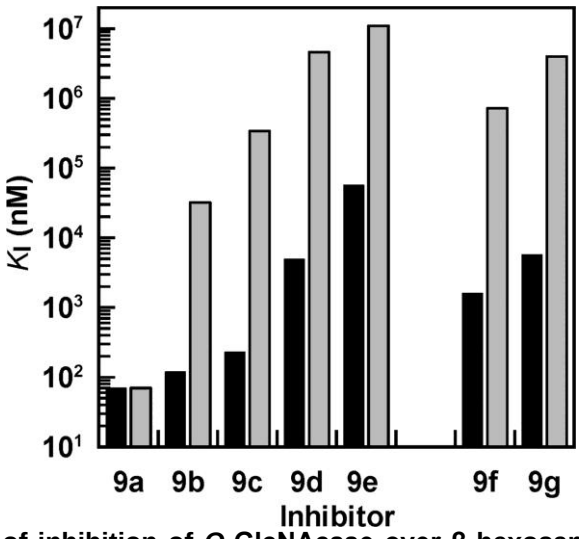


Figure 2.4: Selectivity of inhibition of *O*-GlcNAcase over β-hexosaminidase by a panel of thiazoline inhibitors

Bar graph of the K_I values of the panel of thiazoline inhibitors (**9a–f**) measured for the inhibition of the *O*-GlcNAcase- (■) and β-hexosaminidase (▨)-catalyzed hydrolysis of MU-GlcNAc.

These data also support the view, *vide supra*, that the steric bulk of the *N*-acyl group of the fluorine-substituted substrates does not contribute greatly to the slope of the Taft-like analysis ($\rho^* = -0.42$) measured with *O*-GlcNAcase. The slope ($\rho^* = -1.0$) measured with lysosomal β-hexosaminidase, however, appears likely to be a composite of both electronic and steric effects according to equation 1. The electronic effect of the fluorine substituents ($x\sigma^*$) may therefore be the same for both enzymes but the significant steric contribution (yE_s) results in an apparently steeper slope for the lysosomal β-hexosaminidase than that measured for *O*-GlcNAcase. Together, the Taft-like linear free energy analysis and the selective inhibition data suggest that the active site of *O*-GlcNAcase has considerable more space in the region surrounding the 2-acetamido group of the substrate than does the lysosomal β-hexosaminidase. Clarification of the precise molecular differences in the active site architectures of these two enzymes awaits structural elucidation.

2.4.4 Evaluation of selective inhibitors in cell culture

Having demonstrated the selectivity of these compounds *in vitro* we now turned to evaluating their use in living cells. COS-7 incubated in flasks with 50 μM of inhibitor **9a**, **9c,** or **9g** revealed no abnormalities in proliferation rate or morphology as compared to control cells (data not shown). Cellular *O*-GlcNAc levels within cells cultured for 40 hours in the presence of inhibitors **9a**, **9c** or **9g**, or in their absence was carried out using the *O*-GlcNAc directed monoclonal antibody[228] mAbCTD101.1. Marked increases in cellular *O*-GlcNAc levels within the cells were observed as compared to the control (Figure 2.5A) indicating that these compounds readily gain access to the interior of the cell where they act to block *O*-GlcNAcase function. Western blot analysis of blots probed with α-β-actin (Figure 2.5B) followed by the appropriate secondary HRP-conjugate revealed that in all cases the sample loading was equivalent. It is worth noting

that the parent compound NAG-Thiazoline has been demonstrated to enter into the cells where it exerts an effect on lysosomal β-hexosaminidase[229].

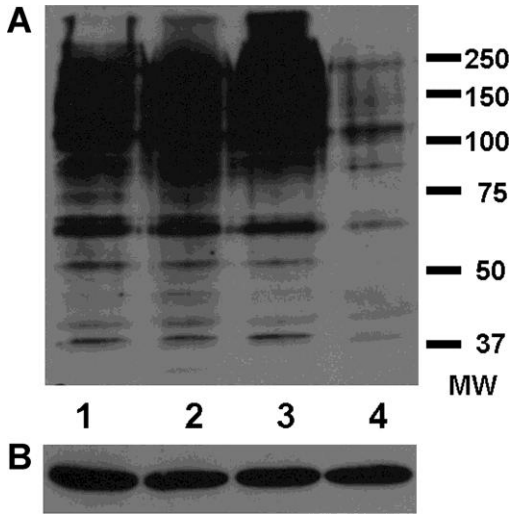


Figure 2.5: Western blot analysis of proteins from COS-7 cells cultured for 40 hours in the presence or absence of 50 μM of different thiazoline inhibitors
Lane 1, thiazoline **9a**; lane 2, thiazoline **9c**; lane 3, thiazoline **9g**; lane 4, no inhibitor. (A) Western blot analysis of cellular *O*-GlcNAc levels using anti-*O*-GlcNAc MAb CTD 110.6 followed by an anti-mouse IgG-HRP conjugate. (B) Western blot of samples loaded in (A) treated with anti-β-actin mAb Clone AC-40 followed by an anti-mouse IgG-HRP conjugate reveals equivalent β-actin levels in each sample.

2.5 Conclusion

In summary we show that catalytic mechanism of human *O*-GlcNAcase and by extension members of family 84 of glycosidases are retaining glycosidases and use a mechanism involving substrate-assisted catalysis from the 2-acetamido group of the substrate. In this regard the functionally related human enzymes *O*-GlcNAcase and lysosomal β-hexosaminidase, which both act to cleave terminal *N*-acetylglucosamine from glycoconjugates, use similar catalytic mechanisms despite the lack of sequence similarity. *O*-GlcNAcase, however, has an active site pocket that more readily tolerates moderately bulky substituents on the acetamido group of the substrate than does that of lysosomal β-hexosaminidase. Despite the similarity of the catalytic mechanism of these enzymes we have exploited the differences in active site architectures between

these two enzymes to generate a small panel of potent inhibitors some of which are remarkably selective for nucleocytoplasmic *O*-GlcNAcase over lysosomal β-hexosaminidase. Furthermore, these compounds are shown to work in cell culture where they result in dramatic increases in cellular *O*-GlcNAc levels. We anticipate that these compounds will be powerful tools in accurately dissecting the role of *O*-GlcNAc in model organisms and tissues without the generation of complex phenotypes stemming from the concomitant inhibition of functionally related proteins.

2.6 Materials and methods

2.6.1 General procedures for synthesis of compounds:

All buffer salts used in this study were obtained from Sigma-Aldrich. Dry methanol and toluene were purchased from Acros Organics, dicloromethane and triethylamine were dried by distillation over CaH_2 prior to use. β-Hexosaminidase was purchased from Sigma (Lot 043K3783). All other reagents were purchased from Sigma-Aldrich and used without further purification. Milli-Q (18.2 mΩ cm^{-1}) water was used to prepare all buffers. Synthetic reactions were monitored by TLC using Merck Kieselgel 60 F_{254} aluminum-backed sheets. Compounds were detected by charring with 10% ammonium molybdate in 2 M H_2SO_4 and heating. Flash chromatography under a positive pressure was performed with Merck Kieselgel 60 (230-400 mesh) using the specified eluants. ^{1}H NMR spectra were recorded on a Varian AS500 Unity Innova spectrometer at 500 MHz (chemical shifts quoted relative to $CDCl_3$, CD_3OD or $(CD_3)_2SO$ where appropriate). ^{19}F NMR spectra were recorded on a Varian AS500 Unity Innova spectrometer at 470 MHz and are proton-coupled with CF_3CO_2H as a reference. ^{13}C NMR spectra were recorded on a Varian AS500 Unity Innova spectrometer at 125 MHz (chemical shifts quoted relative to $CDCl_3$, CD_3OD, or $(CD_3)_2SO$). Elemental analyses of all compounds used in cell culture and enzyme assay were performed at the Simon Fraser University Analytical Facility.

2.6.2 Synthesis of 4-methylumbelliferone 2-deoxy-2-acetamido-β-D-glucopyranosides

4-methylumbelliferyl 2-amino-2-deoxy-β-D-glucopyranoside hydrochloride (**3**) was prepared essentially as described by Roeser and Legler[230] and was used without further purification.

4-Methylumbelliferyl 3,4,6-tri-O-acetyl-2-deoxy-2-trifluoroacetamido-β-D-glucopyranoside **-** To a solution of the hydrochloride salt **3** (0.10 g, 0.2 mmol) in a solution of dimethylformamide (DMF, 6 mL) was added triethylamine (0.06 mL, 0.42 g, 0.41 mmol). The reaction mixture was then cooled to 0 ˚C and trifluoroacetic anhydride (0.08 mL, 0.12 g, 5.7 mmol) was added *via* syringe. The resulting solution was allowed to stand for 16 h at 0 ˚C after which time the reaction was judged complete by TLC analysis. The reaction mixture was then diluted with ethyl acetate (20 mL) and a solution of saturated sodium chloride (40 mL) was added. The organic layer was collected and the aqueous layer was extracted twice with ethyl acetate. The combined organic extracts were washed successively with water, twice with saturated sodium bicarbonate, and finally with a solution of saturated sodium chloride. The organic extracts were dried over $MgSO_4$, filtered, and the solvent removed *in vacuo* to yield a light yellow syrup. The desired product was purified using flash column silica chromatography using a gradient solvent system (1:1; hexanes - ethyl acetate) to yield the desired crude compound as an amorphous white solid (0.93 g, 0.17 mmol, 82 %) that was used in the next step without further purification.

4-Methylumbelliferyl 3,4,6-tri-O-acetyl-2-deoxy-2-difluoroacetamido-β-D-glucopyranoside - To a solution of the hydrochloride salt **3** (0.15 g, 0.3 mmol) in a solution of dimethylformamide (DMF, 6 mL) was added triethylamine (0.09 mL, 0.063 g, 0.62 mmol) and dry pyridine (3 mL). Dicyclohexylcarbodiimide (DCC, 0.48 g, 2.3 mmol) and difluoroacetic acid (0.12 mL, 0.18 g, 1.3 mmol) was added to the reaction mixture *via* syringe. The resulting solution was allowed to stand

for 16 h at 0 ˚C after which time two drops of difluoroacetic acid were added. After a further 3.5 h later the reaction was judged complete by TLC analysis. The solvent was partially removed *in vacuo* after which ethyl acetate (50 mL) and a solution of saturated sodium chloride (20 mL) were added. The organic layer was collected and the aqueous layer was extracted twice with ethyl acetate. The combined organic extracts were washed successively with water, twice with saturated sodium bicarbonate, and finally with a solution of saturated sodium chloride. The organic extracts were dried over $MgSO_4$, filtered, and the solvent removed *in vacuo* to yield a light yellow syrup. The desired product was purified using flash column silica chromatography using a gradient solvent system (1:1; hexanes - ethyl acetate) to yield the crude desired compound as a white amorphous solid (0.10 mg, 0.19 mmol, 64 %) that was used in the next step without further purification.

4-Methylumbelliferyl 3,4,6-tri-O-acetyl-2-deoxy-2-fluoroacetamido-β-D-glucopyranoside - To a solution of the hydrochloride salt **3** (0.50 g, 1.0 mmol) in a solution of dimethylformamide (DMF, 10 mL) was added triethylamine (0.3 mL, 0.21 g, 2.1 mmol) and dry pyridine (10 mL). To a stirred mixture of dry DMF (45 mL) containing dried DOWEX-50 H^+ resin (6 g) was added sodium fluoroacetate (0.9 g, CAUTION: Fluoroacetic acid is highly toxic and proper precautions must be taken during its use and disposal!). After one hour dicyclohexylcarbodiimide (DCC, 1.6 g, 7.8 mmol) and 30 mL of the fluoroacetic acid solution (6.0 mmol) was added *via* canula to the reaction vessel containing the hydrochloride salt 3. The resulting solution was allowed to stand for 16 h at 0 ˚C after which time the reaction was judged complete by TLC analysis. After a further 3.5 h the reaction was judged complete by TLC analysis. The solvent was partially removed *in vacuo* after which ethyl acetate (50 mL) and a solution of saturated sodium chloride (20 mL) were added. The organic layer was collected and the aqueous layer was extracted twice with ethyl acetate. The combined organic extracts were washed successively with water, twice with saturated sodium bicarbonate, and finally with a solution of saturated sodium chloride. The organic extracts

were dried over $MgSO_4$, filtered, and the solvent removed *in vacuo* to yield a light yellow syrup. The desired product was purified using flash column silica chromatography (2:1; ethyl acetate - hexanes) to yield the crude desired compound as an amorphous white solid (356 mg, 0.68 mmol, 68 %) that was used in the next step without further purification.

2.6.3 General procedure for the synthesis of 4-methylumbelliferyl 2-deoxy-2-fluoroacetamido-β-D-glucopyranosides

To a solution of each glycoside in dry methanol was added a spatula tip of anhydrous sodium methoxide. The resulting basic solution was stirred under nitrogen until the reaction was judged complete by TLC analysis. Dowex H^+ resin was added to the stirred reaction mixture until the pH of the solution became neutral. The suspension was filtered and the filter cake rinsed extensively with methanol after which the solvent from the combined filtrates was removed *in vacuo.* The desired deprotected glycosides were isolated by flash column silica chromatography using the following solvent systems: ethyl acetate – methanol –water (12:1:1) for the *N*-tri- and *N*-difluoroacetyl derivatives (**5b** and **5c**) and ethyl acetate - methanol (1:1) for the *N*-monofluoroacetyl derivative (**5a**). Products were recrystallized from ethanol and diethyl ether to yield the desired products with the overall yields over two steps of 66% for the *N*-trifluoroacetyl derivative (**5c**), 37% for the *N*-difluoroacetyl derivative (**5b**), and 45% for the *N*-fluoroacetyl derivative (**5a**).

*4-Methylumbelliferyl 2-deoxy-2-fluoroacetamido-β-D-glucopyranoside (**5a**)*-
^{1}H-NMR (500 MHz, d6-DMSO) δ 7.71 (1 H, d, $J_{H5AR-H6AR}$ = 8.8, H-5_{AR}), 7.00 (1 H, $J_{H8AR-H6AR}$ = 2.4, H-8_{AR}), 6.96 (1 H, dd, H-6_{AR}), 6.26 (1 H, d, $J_{H3AR-CH3}$ = 1.1, H-3_{AR}), 5.21 (1 H, d, J_{H1-H2} = 8.5, H-1), 4.81 (2 H, d, J_{H-F} = 47.0, CH_2F), 3.86 (1 H, dd, J_{H2-H3} = 9.9, H-2), 3.74 (1-H, dd, $J_{H6-H6´}$ = 11.7, J_{H6-H5} = 1.7, H-6), 3.53-3.46 (2 H, m, H-3, H-6´), 3.38 (1 H, ddd, J_{H5-H4} = 9.6, $J_{H5-H6´}$ = 6.0, H-5), 3.20 (1 H, dd, J_{H4-H3} = 8.9, H-4), 2.39 (1 H, d, CH_3) ppm; ^{19}F-NMR (500 MHz, d6-DMSO) -

225.24 ppp (dd, J_{F-H} = 53). Analytical calculated for $C_{18}H_{20}FNO_8$; C, 54.41; H, 5.07; N, 3.53; Experimental C, 54.20; H, 4.97; N 3.59.

*4-Methylumbelliferyl 2-deoxy-2-difluoroacetamido-β-D-glucopyranoside (**5b**)*- ^{1}H-NMR (500 MHz, d6-DMSO) δ 7.69 (1 H, d, $J_{H5AR-H6AR}$ = 8.8, H-5_{AR}), 6.97 (1 H, $J_{H8AR-H6AR}$ = 2.4, H-8_{AR}), 6.92 (1 H, dd, H-6_{AR}), 6.24 (1 H, d, $J_{H3AR-CH3}$ = 1.2, H-3_{AR}), 6.21 (1 H, d, J_{H-F} = 53.6 Hz, CHF_2), 5.16 (1 H, d, J_{H1-H2} = 8.5, H-1), 3.79 (1 H, dd, J_{H2-H3} = 10.3, H-2), 3.72 (1-H, dd, $J_{H6-H6´}$ = 11.8, J_{H6-H5} = 1.9, H-6), 3.53-3.42 (2 H, m, H-3, H-5, H-6´), 3.20 (1 H, dd, J_{H4-H5} = 9.6 Hz, J_{H4-H3} = 8.9 Hz, H-4), 2.38 (1 H, d, CH_3) ppm; ^{19}F-NMR (500 MHz, d6-DMSO) -127.32 (d, J = 54 Hz); Analytical calculated for $C_{18}H_{19}F_2NO_8$; C, 52.05; H, 4.61; N, 3.37; Experimental C, 51.92; H, 4.62; N, 3.31.

*4-Methylumbelliferyl 2-deoxy-2-trifluoroacetamido-β-D-glucopyranoside (**5c**)*- ^{1}H-NMR (500 MHz, d6-DMSO) δ 7.70 (1 H, d, $J_{H5AR-H6AR}$ = 8.8, H-5_{AR}), 6.86 (1 H, $J_{H8AR-H6AR}$ = 2.3, H-8_{AR}), 6.91 (1 H, dd, H-6_{AR}), 6.25 (1 H, d, $J_{H3AR-CH3}$ = 1.2, H-3_{AR}), 5.16 (1 H, d, J_{H1-H2} = 8.5, H-1), 3.80 (1 H, dd, J_{H2-H3} = 10.2, H-2), 3.72 (1-H, dd, $J_{H6-H6´}$ = 11.7, J_{H6-H5} = 1.8, H-6), 3.50-3.42 (2 H, m, H-6´, H-3, H-5), 3.22 (1 H, dd, J_{H4-H5} = 9.5 Hz, J_{H4-H3} = 8.9 Hz, H-4), 2.38 (1 H, d, CH_3) ppm; ^{19}F-NMR (500 MHz, d6-DMSO) -77.29; Analytical calculated for $C_{18}H_{18}F_3NO_8$; C, 49.89; H, 4.19; N, 3.23; Experimental C, 49.80; H, 4.29; N, 3.11.

2.6.4 Kinetic analysis of *O*-GlcNAcase and β-hexosaminidase

All assays were carried out in triplicate at 37 ^{o}C for 30 minutes using a stopped assay procedure in which the enzymatic reactions (25 μL) were quenched by the addition of a 6 fold excess (150 μL) of quenching buffer (200 mM glycine, pH 10.75). Assays were initiated by the addition, *via* syringe, of enzyme (3 μL), and in all cases the final pH of the resulting quenched solution was greater than 10. Time dependent assay of β-hexosaminidase and *O*-GlcNAcase revealed that both enzymes were stable over this period in their respective buffers; 50 mM

citrate, 100 mM NaCl, 0.1% BSA, pH 4.25 and 50 mM NaH_2PO_4, 100 mM NaCl, 0.1% BSA, pH 6.5. The progress of the reaction at the end of thirty minutes was determined by the measuring the extent of 4-methylumbelliferone liberated as determined by fluorescence measurements using a Varian CARY Eclipse Fluoresence-Spectrophotometer 96-well plate system and comparison to a standard curve of 4-methylumbelliferone under identical buffer conditions. Excitation and emission wavelengths of 368 and 450 nM were used, respectively, with 5 mm slit openings. Human placental β-hexosaminidase was purchased from Sigma-Aldrich (Lot 043K3783). The cloning and expression of *O*-GlcNAcase will be described elsewhere. Both enzymes were dialyzed against PBS buffer and their concentrations determined using the Bradford assay. The concentration (μg/μl) of β-hexosaminidase and *O*-GlcNAcase used in assays with fluorinated substrates were as follows: for 4-methylumbelliferyl 2-acetamido-2-deoxy-β-D-glucopyranoside (**5**); 0.00077, 0.0126, **5a;** 0.0031, 0.0189, **5b;** 0.0154, 0.0756, and for **5c;** 0.0154, 0.01523. In addition, β-hexosaminidase and *O*-GlcNAcase were used at a concentration (μg/μL) of 0.0154 and 0.0378, respectively to test the inhibitors using substrate **5** at a concentration of 0.64 mM. All inhibitors were tested at eight concentrations ranging from 5 times to $1/5^{th}$ K_I with the exception of the assay of inhibitor **8e** with β-hexosaminidase, where a such high concentrations of inhibitor could not reached owing to the high K_I value of **8e**. Where necessary, assays were carried out in triplicate and error bars are included in plots of the data.

2.6.5 Synthesis of thiazoline inhibitor panel

General procedure for the synthesis of 1,3,4,6-tetra-O-acetyl-2-deoxy 2-N-acyl-β-D-glucopyranoses: To a solution of the hydrochloride salt of 2-amino-2-deoxy-1,3,4,6-tetra-*O*-acetyl-β-D-glucopyranose (**6**)[231] in 1 volume of dry dichloromethane was added 2 equivalents of dry triethylamine at which time the starting material dissolved. The reaction mixture was cooled to 4 ºC and 1.2 equivalents of the appropriate acyl chloride was added *via* syringe. The resultant

mixture was stirred for approximately two hours at room temperature. When the reaction mixture was judged complete by TLC analysis, 5 volumes of ethyl acetate were added. The resulting organic phase was washed successively with water, 1 M NaOH, and saturated sodium chloride. The organic phase was dried over $MgSO_4$, filtered, and concentrated to yield a white crystalline solid. The material was recrystallized using a mixture of ethyl acetate and hexanes to yield the desired *N*-acylated materials in yields ranging from 46 to 74 %.

1,3,4,6-tetra-O-acetyl-2-deoxy-2-N-propionyl-β-D-glucopyranose (**7b**)- ^{1}H NMR (500M Hz, $CDCl_3$) δ: 5.68 (1 H, d, $J_{H1,H2}$ = 8.8 Hz, H-1), 5.42 (1 H, d, $J_{NH,H2}$ = 6.7 Hz, NH), 5.14 (1 H, dd, $J_{H3,H4}$ = 8.7 Hz, H-3), 5.12 (1 H, dd, $J_{H4,H5}$ = 8.7 Hz, H-4), 4.34 (1 H, ddd, $J_{H2,H3}$ = 8.7 Hz, H-2), 4.27 (1 H, dd, $J_{H6,H6'}$ = 12.5 Hz, $J_{H5,H6}$ = 4.6 Hz, H-6), 4.12 (1 H, dd, $J_{H5,H6'}$ = 2.1 Hz, H-6'), 3.79-3.75 (1 H, m, H-5), 2.12 (3 H, s, OAc), 2.10 (3 H, s, OAc), 2.04 (3 H, s, OAc), 2.02 (3 H, s, OAc), 2.15-2.10 (2 H, m, H-7), 0.90 (3 H, t, $J_{H8,H7}$ = 7.4 Hz, H-8) ppm.

1,3,4,6-tetra-O-acetyl-2-deoxy-2-N-butanonyl-β-D-glucopyranose (**7c**)- ^{1}H NMR (500 MHz, $CDCl_3$) δ: 5.73 (1 H, d, $J_{H1,H2}$ = 8.8 Hz, H-1), 5.58 (1 H, d, $J_{NH,H2}$ = 9.5 Hz, NH), 5.21-5.14 (2 H, m, H-3/H-4), 4.38 (1 H, ddd, H-2), 4.31 (1 H, dd, $J_{H6,H6'}$ = 12.5 Hz, H-6), 4.16 (1 H, dd, $J_{H5,H6'}$ = 2.2 Hz, H-6'), 3.84 (1 H, ddd, $J_{H5,H6}$ = 4.7 Hz, H-5), 2.16 (3 H, s, OAc), 2.13 (3 H, s, OAc), 2.18-2.10 (2 H, m, H-7), 2.08 (3 H, s, OAc), 2.05 (3 H, s, OAc), 1.64 (2 H, ddd, H-8), 0.94 (3 H, t, $J_{H8,H9}$ = 7.4 Hz, H-9) ppm.

1,3,4,6-tetra-O-acetyl-2-deoxy-2-N-pentanoyl-β-D-glucopyranose (**7d**)- ^{1}H NMR (500 MHz, $CDCl_3$) δ: 5.68 (1 H, d, $J_{H1,H2}$ = 8.8 Hz, H-1), 5.50 (1 H, d, $J_{NH,H2}$ = 9.5 Hz, NH), 5.18-5.12 (2 H, m, H-3/H-4), 4.36-4.30 (1 H, m, H-2), 4.26 (1 H, dd, $J_{H6,H6'}$ = 12.5 Hz, H-6), 4.13 (1 H, dd, $J_{H5,H6'}$ = 2.2 Hz, H-6'), 3.82-3.77 (1 H, m, $J_{H5,H6}$ = 4.5 Hz, H-5), 2.11 (3 H, s, OAc), 2.11 (2 H, m, H-7), 2.09 (3 H, s, OAc), 2.04 (3 H, s, OAc), 2.02 (3 H, s, OAc), 1.56 (2 H, dd, $J_{H7,H8}$ = 7.5 Hz, H-8), 1.52

(2 H, dd, $J_{H7',H8}$ = 7.5 Hz, H-8), 1.26 (2 H, ddq, $J_{H8,H9}$ = 7.5 Hz, H-9), 0.94 (3 H, dd, $J_{H9,H10}$ = 7.5 Hz, H-10) ppm.

1,3,4,6-tetra-O-acetyl-2-deoxy-2-N-hexanoyl-β-D-glucopyranose (**7e**)- ^{1}H NMR (500 MHz, $CDCl_3$) δ: 5.69 (1 H, d, $J_{H1,H2}$ = 8.8 Hz, H-1), 5.52 (1 H, d, $J_{NH,H2}$ = 9.5 Hz, NH), 5.14 (2 H, m, H-3/H-4), 4.37-4.30 (1 H, m, H-2), 4.26 (1 H, dd, $J_{H6,H6'}$ = 12.5 Hz, H-6), 4.15-4.08 (1 H, m, H-6'), 3.82-3.77 (1 H, m, $J_{H5,H6}$ = 4.7 Hz, H-5), 2.11 (3 H, s, OAc), 2.09 (3 H, s, OAc), 2.04 (3 H, s, OAc), 2.02 (3 H, s, OAc), 1.58-1.52 (2 H, m, H-7), 1.33-1.08 (6 H, m, H-8, H-9, H-10), 0.87 (3 H, t, $J_{H11,H10}$ = 7.1 Hz, H-11) ppm.

1,3,4,6-tetra-O-acetyl-2-deoxy-2-N-isobutanoyl-β-D-glucopyranose (**7f**)- ^{1}H NMR (500 MHz, $CDCl_3$) δ: 5.70 (1 H, d, $J_{H1,H2}$ = 8.8 Hz, H-1), 5.58 (1 H, d, $J_{NH,H2}$ = 9.6 Hz, NH), 5.18-5.12 (2 H, m, H-3/H-4), 4.36-4.30 (1 H, m, H-2), 4.27 (1 H, dd, $J_{H6,H6'}$ = 12.5 Hz, H-6), 4.12 (1 H, dd, $J_{H5,H6'}$ = 2.2 Hz, H-6'), 3.84-3.87 (1 H, m, $J_{H5,H6}$ = 4.8 Hz, H-5), 2.10 (3 H, s, OAc), 2.08 (3 H, s, OAc), 2.28 (1 H, m, H-7), 2.04 (3 H, s, OAc), 2.03 (3 H, s, OAc), 1.08 (6 H, t, $J_{H8,H7}$ = 2.8 Hz, H-8) ppm.

1,3,4,6-tetra-O-acetyl-2-deoxy-2-N-isopentanoyl-β-D-glucopyranose (**7g**)- ^{1}H NMR (500 MHz, $CDCl_3$) δ: 5.67 (1 H, d, $J_{H1,H2}$ = 8.8 Hz, H-1), 5.58 (1 H, d, $J_{NH,H2}$ = 9.5 Hz, NH), 5.18-5.10 (2 H, m, H-3/H-4), 4.38-4.32 (1 H, m, H-2), 4.27 (1 H, dd, $J_{H6,H6'}$ = 12.5 Hz, H-6), 4.12 (1 H, $_{1}$, $J_{H5,H6'}$ = 2.2 Hz, H-6'),3.82-3.78 (1 H, m, $J_{H5,H6}$ = 4.6 Hz, H-5), 2.10 (3 H, s, OAc), 2.08 (3 H, s, OAc), 2.06-2.01 (2 H, m, H-7), 2.04 (3 H, s, OAc), 2.03 (3 H, s, OAc), 1.98 (1 H, ddd, H-8), 0.94 (6 H, d, $J_{H8,H9}$ = 6.5 Hz, H-9) ppm.

2.6.6 General procedure for the synthesis of 3,4,6-tri-*O*-acetyl-1,2-dideoxy-2′-alkyl-α-D-glucopyranoso-[2,1-d]-Δ2′-thiazoline (3,4,6-tri-*O*-acetyl-NAG-thiazoline analogues):

To a solution of the appropriate 1,3,4,6-tetra-*O*-acetyl-2-*N*-acyl-2-deoxy-β-D-glucopyranose (**7a-g)** in anhydrous toluene was added Lawesson's Reagent (0.6 equivalents) and the reaction mixture was refluxed for 2 h after which time the reaction was judged to be complete by TLC analysis. The solution was cooled to room temperature and the solvent was removed *in vacuo.* The residue was dissolved in toluene and the desired material was isolated by flash column silica chromatography using a solvent system of hexanes and ethyl acetate in ratios ranging from 4:1 to 1:2 as appropriate. Products were isolated in yields ranging from 62 to 83 %. 3,4,6-tri-*O*-acetyl-1,2-dideoxy-2′-ethyl-α-D-glucopyranoso-[2,1-d]-Δ2′-thiazoline (**8a**) has been previously prepared and all sprectral characterization agreed with the literature values.(28)

3,4,6-tri-O-acetyl-1,2-dideoxy-2'-ethyl-α-D-glucopyranoso-[2,1-d]-Δ2'-thiazoline ***(8b)*** **-** ^{1}H NMR (500 MHz, $CDCl_3$) δ: 6.28 (1 H, d, $J_{H1,H2}$ = 7.4 Hz, H-1), 5.60 (1 H, dd, $J_{H3,H4}$ = 3.3 Hz, H-3), 4.97 (1 H, d, $J_{H4,H5}$ = 9.3 Hz, H-4), 4.56-4.52 (1 H, m, H-2), 4.18-4.10 (2 H, m, H-6/ H-6'), 3.58 (1 H, ddd, $J_{H5,H6}$ = 3.2 Hz, H-5), 2.74-2.67 (1 H, m, H-7), 2.14 (3 H, s, OAc), 2.11 (3 H, s, OAc), 2.09 (3 H, s, OAc), 1.29 (3 H, t, $J_{H8,H7}$ = 7.60 Hz, H-8) ppm.

3,4,6-tri-O-acetyl-1,2-dideoxy-2'-propyl-α-D-glucopyranoso-[2,1-d]-Δ2'-thiazoline ***(8c)*** **-**

^{1}H NMR (500 MHz, $CDCl_3$) δ: 6.26 (1 H, d, $J_{H1,H2}$ = 7.2 Hz, H-1), 5.58 (1 H, dd, $J_{H3,H4}$ = 3.3 Hz, H-3), 4.96 (1 H, d, $J_{H4,H5}$ = 9.2 Hz, H-4), 4.54-4.50 (1 H, m, H-2), 4.16-4.08 (2 H, m, H-6/H-6'), 3.58 (1 H, ddd, $J_{H5,H6}$ = 3.3 Hz, H-5), 2.70-2.58 (2 H, m, H-7), 2.14 (3 H, s, OAc), 2.08 (3 H, s, OAc), 2.07 (3 H, s, OAc), 1.76-1.69 (2 H, m, H-8), 1.00 (3 H, t, $J_{H9,H8}$ = 7.4 Hz, H-9) ppm.

3,4,6-tri-O-acetyl-1,2-dideoxy-2'-butyl-α-D-glucopyranoso-[2,1-d]-Δ2'-thiazoline ***(8d)*** -

^{1}H NMR (500 MHz, $CDCl_3$) δ: 6.21 (1 H, d, $J_{H1,H2}$ = 7.2 Hz, H-1), 5.57 (1 H, dd, $J_{H3,H4}$ = 3.3 Hz, H-3), 4.94 (1 H, d, $J_{H4,H5}$ = 9.4 Hz, H-4), 4.48-4.44 (1 H, m, H-2), 4.12-4.07 (2 H, m, H-6/H-6'), 3.53 (1 H, ddd, $J_{H5,H6}$ = 3.0 Hz, H-5), 2.60-2.57 (2 H, m, H-7), 2.12 (3 H, s, OAc), 2.07 (6 H, s, OAc), 1.67-1.63 (2 H, m, H-8), 1.40 (2 H, ddd, $J_{H9,H8}$ = 7.3 Hz, H-9), 0.92 (3 H, t, $J_{H10,H9}$ = 7.4 Hz, H-10) ppm.

3,4,6-tri-O-acetyl-1,2-dideoxy-2'-pentyl-α-D-glucopyranoso-[2,1-d]-Δ2'-thiazoline ***(8e)*** -

^{1}H NMR (500 MHz, $CDCl_3$) δ: 6.24 (1 H, d, $J_{H1,H2}$ = 7.2 Hz, H-1), 5.60 (1 H, s, H-3), 4.96 (1 H, d, $J_{H4,H5}$ = 9.4 Hz, H-4), 4.52-4.49 (1 H, m, H-2), 4.14-4.10 (2 H, m, $J_{H6,H6'}$ = 12.2 Hz H-6/H-6'), 3.56 (1 H, ddd, $J_{H5,H6}$ = 3.1 Hz, $J_{H5,H6'}$ = 5.6 Hz, H-5), 2.65-2.60 (2 H, m, H-7), 2.14 (3 H, s, OAc), 2.09 (3 H, s, OAc), 2.08 (3 H, s, OAc), 1.71-1.68 (2 H, m, H-8), 1.38-1.33 (4 H, m, H-9/H-10), 0.91 (3 H, t, $J_{H11,H10}$ = 6.9 Hz, H-11) ppm.

*3,4,6-tri-O-acetyl-1,2-dideoxy-2'-isopropyl-α-D-glucopyranoso-[2,1-d]-Δ2'-thiazoline (****8f****)* - ^{1}H NMR (500 MHz, $CDCl_3$) δ: 6.27 (1 H, d, $J_{H1,H2}$ = 7.2 Hz, H-1), 5.60 (1 H, m, H-3), 4.97 (1 H, d, $J_{H4,H5}$ = 9.3 Hz, H-4), 4.55-4.49 (1 H, m, H-2), 4.18-4.11 (2 H, m, H-6/H-6'), 3.60 (1 H, ddd, $J_{H5,H6}$ = 3.1 Hz, H-5), 2.55 (2 H, s, H-7), 2.15 (3 H, s, OAc), 2.09 (3 H, s, OAc), 2.08 (3 H, s, OAc), 1.07 (6 H, t, $J_{H8,H7}$ = 6.6 Hz, H-8) ppm.

3,4,6-tri-O-acetyl-1,2-dideoxy-2'-isobutyl-α-D-glucopyranoso-[2,1-d]-Δ2'-thiazoline ***(8g)*** - ^{1}H NMR (500 MHz, $CDCl_3$) δ: 6.26 (1 H, d, $J_{H1,H2}$ = 7.2 Hz, H-1), 5.60 (1 H, dd, $J_{H3,H4}$ = 3.3 Hz, H-3), 4.97 (1 H, d, $J_{H4,H5}$ = 9.3 Hz, H-4), 4.56-4.50 (1 H, m, H-2), 4.16-4.10 (2 H, m, H-6/H-6'), 3.62-3.58 (1 H, m, $J_{H5,H6}$ = 3.1 Hz, H-5), 2.57-2.52 (2 H, m, H-7), 2.15 (3 H, s, OAc), 2.09 (3 H, s, OAc), 2.08 (3 H, s, OAc), 1.28-1.23 (2 H, m, H-8), 1.02 (6 H, t, $J_{H9,H8}$ = 6.6 Hz, H-9) ppm.

2.6.7 1,2-dideoxy-2′-alkyl-α-D-glucopyranoso-[2,1-d]-Δ2′-thiazoline (NAG-thiazoline analogues):

To a solution of the appropriate protected thiazoline (**8a-g**) in dry methanol was added a spatula tip of anhydrous sodium methoxide. The basic solution was stirred until the reaction was judged complete by TLC analysis (typically 2 hours). A solution of glacial acetic acid in methanol (1:20) was added dropwise to the reaction mixture until the pH of the solution solution became neutral. The solvent was then removed *in vacuo* and the desired materials (**9a-g)** was isolated by flash column silica chromatography using a solvent system of ethyl acetate and methanol in ratios ranging from 2:1 to 6:1 as appropriate. Products were isolated in yields ranging from 86 to 99 %. 1,2-dideoxy-2′-ethyl-□-D-glucopyranoso-[2,1-d]-□2′-thiazoline (**9a**) has been previously prepared and all sprectral characterization agreed with the literature values(28) as did the elemental analysis of the sample used in these assays Analytical calculated for $C_8H_{13}O_4NS$; C, 43.82; H, 5.98; N, 6.39; Experimental C, 43.45; H, 6.23; N, 6.18.

1,2-dideoxy-2'-ethyl-α-D-glucopyranoso-[2,1-d]-Δ2'-thiazoline (**9b**) - ^{1}H NMR (500 MHz, MeOD) δ: 6.31 (1 H, d, $J_{H1,H2}$ = 7.0 Hz, H-1), 4.29-4.26 (1 H, m, $J_{H2,H3}$ = 4.32 Hz, H-2), 4.09 (1 H, dd, $J_{H3,H4}$ = 4.32 Hz, H-3), 3.70 (1 H, dd, $J_{H6,H6'}$ = 12.07 Hz, H-6), 3.57 (1 H, dd, H-6') 3.52 (1 H, dd, $J_{H4,H5}$ = 9.10 Hz, H-4), 3.32-3.29 (1 H, m, $J_{H5,H6}$ = 6.29 Hz, $J_{H5,H6'}$ = 2.21 Hz, H-5), 2.52 (2 H, dd, $J_{H7,H8}$ = 7.59 Hz, H-7), 1.18 (3 H, t, H-8) ppm. ^{13}C NMR (500 MHz, MeOD) δ: 175.1, 89.4, 80.2, 75.7, 74.5, 71.4, 62.5, 28.9, 11.1 ppm; Analytical calculated for $C_9H_{15}O_4NS$; C, 46.34; H, 6.48; N, 6.00; Experimental C, 45.95; H, 6.33; N, 5.93.

1,2-dideoxy-2'-propyl-α-D-glucopyranoso-[2,1-d]- Δ2'-thiazoline (**9c**) - ^{1}H NMR (500 MHz, MeOD) δ: 6.32 (1 H, d, $J_{H1,H2}$ = 7.0 Hz, H-1), 4.28 (1 H, m, $J_{H2,H3}$ = 4.4 Hz, H-2), 4.09 (1 H, dd, $J_{H3,H4}$ = 4.4 Hz, H-3), 3.72 (1 H, dd, $J_{H6,H6'}$ = 12.0 Hz, H-6), 3.59 (1 H, dd, H-6'), 3.53 (1 H, dd, $J_{H4,H5}$ = 9.1 Hz, H-4), 3.33 (1 H, m, $J_{H5,H6}$ = 6.3 Hz, $J_{H5,H6'}$ = 2.5 Hz, H-5), 2.51-2.48 (1 H, m, $J_{H7,H8}$ = 7.3 Hz, $J_{H7,H7'}$ = 14.9

Hz, H-7), 2.49-2.47 (1 H, m, H-7'), 1.68-1.65 (2 H, m, $J_{H8,H9}$ = 7.4 Hz, H-8) 0.98 (3 H, dd, H-9); Analytical calculated for $C_{10}H_{17}O_4NS$; C, 48.57; H, 6.93; N, 5.66; Experimental C, 48.32; H, 6.77; N, 5.45.

1,2-dideoxy-2'-butanyl-α-D-glucopyranoso-[2,1-d]-Δ2'-thiazoline (**9d**) –^{1}H NMR (500 MHz, MeOD) δ: 6.31 (1 H, d, $J_{H1,H2}$ = 7.0 Hz, H-1), 4.30-4.28 (1 H, m, $J_{H2,H3}$ = 4.4 Hz, H-2), 4.09 (1 H, dd, $J_{H3,H4}$ = 4.4 Hz, H-3), 3.71 (1 H, dd, $J_{H6,H6'}$ = 12.0 Hz, H-6), 3.59 (1 H, dd, H-6'), 3.53 (1 H, dd, $J_{H4,H5}$ = 9.1 Hz, H-4), 3.35-3.30 (1 H, m, $J_{H5,H6}$ = 6.3 Hz, $J_{H5,H6'}$ = 2.5 Hz, H-5), 2.58-2.53 (1 H, m, $J_{H7,H8}$ = 7.8 Hz, $J_{H7,H7'}$ = 14.8 Hz, H-7) 2.53-2.50 (1 H, m, $J_{H7',H8}$ = 7.6 Hz, H-7') 1.60 (2 H, ddd, $J_{H8,H9}$ = 14.9 Hz, H-8) 1.37 (2 H, ddd, $J_{H9,H10}$ = 7.4, H-9) 0.92 (3 H, t, H-10) ppm. ^{13}C NMR (500 MHz, MeOD) δ: 175.5, 89.4, 79.1, 75.2, 74.2, 71.5, 62.7, 36.9, 20.4, 15.2 ppm; Analytical calculated for $C_{11}H_{19}O_4NS$; C, 50.55; H, 7.33; N, 5.36; Experimental C, 50.68; H, 7.12; N, 5.13

1,2-dideoxy-2'-pentyl-α-D-glucopyranoso-[2,1-d]-Δ2'-thiazoline (**9e**) - ^{1}H NMR (500 MHz, MeOD) δ: 6.20 (1 H, d, $J_{H1,H2}$ = 6.9 Hz, H-1), 4.29-4.27 (1 H, m, $J_{H2,H3}$ = 6.0 Hz, H-2), 4.09 (1 H, dd, $J_{H3,H4}$ = 4.4 Hz, H-3), 3.72 (1 H, dd, $J_{H6,H6'}$ = 12.0 Hz, H-6), 3.59 (1 H, dd, H-6'), 3.53 (1 H, dd, $J_{H4,H5}$ = 9.1 Hz, H-4), 3.35-3.31 (1 H, m, $J_{H5,H6}$ = 6.3, $J_{H5,H6'}$ = 2.4 Hz, H-5), 2.52 (2 H, ddd, $J_{H7,H8}$ = 6.2 Hz, H-7), 1.37-1.32 (2 H, m, H-8), 1.35-1.30 (2 H, m, H-9), 1.66-1.63 (2 H, m, $J_{H10,H11}$ = 7.1 Hz, H-10), 0.90 (3 H, t, H-11) ppm. ^{13}C NMR (500 MHz, MeOD) δ: 174.8, 89.3, 79.2, 75.5, 73.4, 71.5, 62.2, 35.4, 30.3, 28.8, 22.1, 13.1 ppm; Analytical calculated for $C_{12}H_{21}O_4NS$; C, 52.34; H, 7.69; N, 5.09; Experimental C, 52.48; H, 7.67; N, 4.40

1,2-dideoxy-2'-isopropyl-α-D-glucopyranoso-[2,1-d]- Δ2'-thiazoline (**9f**) - ^{1}H NMR (500 MHz, MeOD) δ: 6.31 (1 H, d, $J_{H1,H2}$ = 7.0 Hz, H-1), 4.29 (1 H, m, $J_{H2,H3}$ = 4.4 Hz, H-2), 4.11 (1 H, dd, $J_{H3,H4}$ = 4.4 Hz, H-3), 3.60 (1 H, dd, $J_{H6,H6'}$ = 12.0 Hz, H-6), 3.72 (1 H, dd, H-6'), 3.55 (1 H, dd, $J_{H4,H5}$ = 9.1 Hz, H-4), 3.35-3.32 (1 H, m, $J_{H5,H6}$ = 6.4 Hz, $J_{H5,H6'}$ = 2.4 Hz, H-5), 2.87-2.83 (2 H, m, $J_{H7,H8}$ = 6.9 Hz, H-7),

1.24 (3 H, d, H-8), 1.21 (3 H, d, H-8') ppm. ^{13}C NMR (500 MHz, MeOD) δ: 175.2, 89.5, 79.7, 75.6, 72.3, 70.8, 62.3, 35.1, 30.3, 22.5, 13.8 ppm;

1,2-dideoxy-2'-isobutyl-α-D-glucopyranoso-[2,1-d]-Δ2'-thiazoline (**9g**) - ^{1}H NMR (500 MHz, MeOD) δ: 6.38 (1 H, d, $J_{H1,H2}$ = 7.1 Hz, H-1), 4.28-4.24 (1 H, m, $J_{H2,H3}$ = 6.0 Hz, H-2), 4.06 (1 H, dd, $J_{H3,H4}$ = 6.0 Hz, H-3), 3.71 (1 H, dd, $J_{H6,H6'}$ = 12.0 Hz, H-6), 3.58 (1 H, dd, H-6'), 3.58 (1 H, dd, $J_{H4,H5}$ = 9.2 Hz, H-4), 3.35-3.30 (1 H, m, $J_{H5,H6}$ = 6.3 Hz, $J_{H5,H6'}$ = 2.4 Hz, H-5), 2.46-2.40 (1 H, m, $J_{H7,H8}$ = 7.3 Hz, $J_{H7,H7'}$ = 14.1 Hz, H-7) 2.37-2.33 (1 H, m, $J_{H7',H8}$ = 7.3 Hz, H-7') 2.00 (2 H, ddd, $J_{H8,H9}$ = 6.7 Hz, H-8) 0.97 (3 H, d, H-9) ppm; Analytical calculated for $C_{11}H_{19}O_4NS$; C, 50.55; H, 7.33; N, 5.36; Experimental C, 50.68; H, 7.12; N, 5.13

2.6.8 Cell culture and inhibition.

COS-7 cells were cultured in DMEM medium (Invitrogen) supplemented with 5–10% FBS (Invitrogen). Aliquots of inhibitors (50 µl of a stock in 95% ethanol) were delivered onto tissue culture plates and the ethanol was evaporated. The cells were incubated at 37ºC for 40 hours at which time they reached approximately 80 % confluence. Control cultures without inhibitors were treated in the same manner.

2.6.9 Western blot analyses.

COS-7 cells were cultured in the presence of inhibitors **9a**, **9c**, or **9g** as described above to 90 % of confluence. A culture of control cells was treated in the same manner as follows but the cultures contained no inhibitor. Cells (2 x 10 cm plates) were harvested by scraping and pooled by centrifugation (200 × *g*, 10 min). Cells were washed once with PBS, pH 7.0 (10 mL) and pelleted (200 × *g*, 10 min). The cells could be frozen at –80ºC at this point. Cells were thawed at 4ºC, and cold lysis buffer (1 mL of 50 mM Tris, pH 8.0 containing 150 mM NaCl, 1 mM EDTA, 1 mM PMSF, 1% NP-40, 0.5% sodium deoxycholate, and 1 mM of inhibitor **9f**) was added. After 10 minutes at 4 ºC the solution was centrifuged at

14,000 rpm in an Eppendorf 5415C microcentrifuge and the supernatant was collected. SDS/PAGE loading buffer was added to an aliquot (15 µL) of each sample, and after heating at 96 ºC aliquots were loaded onto 10% or 12% Tris·HCl polyacrylamide gels. After electrophoresis, the samples were electroblotted to nitrocellulose membrane (0.45 µm, Bio-Rad). Transfer was verified by visual inspection of the transfer of prestained markers (Dual Colour Precision Plus Protein Standard - Biorad). The membrane was blocked by using 5% BSA (fraction V, Sigma) in PBS (blocking buffer A for samples probed with mouse anti-*O*-GlcNAc monoclonal IgM antibody (MAb CTD 110.6 - Covance)) or 5% low-fat dry powdered milk (blocking buffer B for samples probed with anti-β-actin), pH 7.4, containing 0.1% Tween 20 for 1 h at room temperature or overnight at 4ºC. The blocking solution was decanted, and a solution of blocking buffer A containing MAb CTD 110.6 (1:2500 of the stock) or blocking buffer B containing mouse monoclonal anti-α-actin IgG (Clone AC-40 – Sigma) was added (1:1000 dilution) as appropriate. The membrane was incubated at room temperature for 1 h or overnight at 4ºC after which the blocking buffer was decanted and the membrane was rinsed with PBS, pH 7.4, containing 0.1 % Tween 20 (wash buffer). Membranes were then rinsed for 2 × 5 min and 2 × 20 min with wash buffer. For immunological detection of *O*-GlcNAc, the membrane was incubated in blocking buffer A for 1 hour at RT and, after washing, the membrane was incubated with a secondary goat anti-mouse-IgM-HRP-conjugate (1:2500, Santa Cruz Biotech) for one hour at RT or 4 °C overnight in blocking solution. For detection of α-actin levels, the membrane was incubated with a secondary goat anti-mouse-IgG-HRP conjugate (1:100000, Sigma) for one hour at RT or 4 °C overnight in blocking solution B. Membranes were washed and detection of membrane bound goat anti-mouse-IgG-HRP conjugate was accomplished as for anti-FLAG-HRP. Detection of membrane-bound HRP-conjugates was accomplished with chemiluminescent detection using the SuperSignal West Pico Chemiluminescent Detection Kit (Pierce) and film (Kodak Biomax MR).

2.7 Acknowledgements

We thank the Canada Research Chairs Program for a Tier II Research Chair to DJV, the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support, the Protein Engineering Networks of Centres of Excellence for support through the research chair program, and Simon Fraser University for starting funds and a President's Research Grant. The authors also thank J. Hanover for the expression plasmid encoding *O*-GlcNAcase and T. Kitos and R. Cornell for assistance and access to tissue culture facilities.

3: Analysis of PUGNAc and NAG-thiazoline as transition state analogues for human *O*-GlcNAcase: Structural and mechanistic insights into inhibitor selectivity and transition state poise

The manuscript below is reprinted from:

Whitworth, G. E.; Macauley, M. S.; Stubbs, K. A.; Dennis, R. J.; Taylor, E. J.; Davies, G. J.; Greig, I. R.; Vocadlo, D. J., Analysis of PUGNAc and NAG-thiazoline as transition state analogues for human O-GlcNAcase: mechanistic and structural insights into inhibitor selectivity and transition state poise. *J Am Chem Soc* **2007,** 129, (3), 635-44.

Updates in the field relating to this Chapter can be found in appendix 1. The designated numbers given to synthetic or target compounds in this Chapter relate only to this Chapter.

3.1 Contributions

I synthesized the NAG-thiazoline series of inhibitors and I synthesized and performed kinetic studies of the ground state inhibitors. I analyzed the data and wrote the first draft of the manuscript below. Matthew Macauley synthesized and

tested the series of 4-methylumbelliferone substrates. Dr. Keith Stubbs synthesized and tested the series of PUGNAc inhibitors. Rebecca Dennis provided the crystal structures, and Dr. Ian Greig performed the molecular electrostatic potential surface (MEPs) analysis.

3.2 Abstract

O-GlcNAcase catalyzes the cleavage of beta-*O*-linked 2-acetamido-2-deoxy-beta-D-glucopyranoside (*O*-GlcNAc) from serine and threonine residues of post-translationally modified proteins. Two potent inhibitors of this enzyme are PUGNAc and NAG-thiazoline. Derivatives of these inhibitors differ in their selectivity for human *O*-GlcNAcase over the functionally related human lysosomal beta-hexosamindases, with PUGNAc derivatives showing modest selectivities and NAG-thiazoline derivatives showing high selectivities. The molecular basis for this difference in selectivities is addressed, as is how well these inhibitors mimic the *O*-GlcNAcase-stabilized transition state. Using a series of substrates, ground state inhibitors, and transition state mimics having analogous structural variations, we describe linear free energy relationships of log (K_M/k_{cat}) versus log (K_I) for PUGNAc and NAG-thiazoline. These relationships suggest that PUGNAc is a poor transition state analogue while NAG-thiazoline is revealed as a transition state mimic. Comparative X-ray crystallographic analyses of enzyme-inhibitor complexes reveal subtle molecular differences accounting for the differences in selectivities between these two inhibitors and illustrate key molecular interactions. Computational modeling of species along the reaction coordinate, as well as PUGNAc and NAG-thiazoline, provide insight into the features of NAG-thiazoline that resemble the transition state and reveal where PUGNAc fails to capture significant binding energy. These studies also point to late transition state poise for the *O*-GlcNAcase catalyzed reaction with significant nucleophilic participation and little involvement of the leaving group. The potency of NAG-thiazoline, its transition sate mimicry, and its lack of traditional transition state-like design features suggest that potent rationally designed glycosidase inhibitors can be developed that exploit variation in transition state poise.

3.3 Introduction

The development of rationally designed potent glycosidase inhibitors has been a topic of continuing research for years and has more recently enjoyed a surge of interest owing to their increased uses as both research tools[139,232] and as therapeutic agents[83,233,234]. Considerable efforts have been directed toward preparing both potent[40,41,76] and selective[235, 236] glycosidase inhibitors using transition state mimicry as a guiding design principle[40,41]. The expectation that stable molecules resembling the enzyme catalyzed transition state, in either their charge or geometry, will be potent competitive inhibitors stems from a proposal, made first by Pauling[237], that enzymes catalyze reactions by binding interactions realized preferentially with the transition state rather than with ground state[237,238]. Such tight binding stabilizes these fleeting transition state structures and thereby results in the enormous rate accelerations of up to 10^{19}-fold seen for the most proficient enzymes[59]. On the basis of this rate acceleration the theoretical limit for the dissociation constant of a perfect transition state analogue can be estimated to be approximately 10^{-22} M[59]. Of course, stable transition state analogues necessarily vary from genuine transition state structures having partial bonds and therefore it is impossible to capture all the potentially available binding energy using a stable molecule. Despite this limitation, very potent transition state analogues have been found for several different classes of enzymes[239,240].

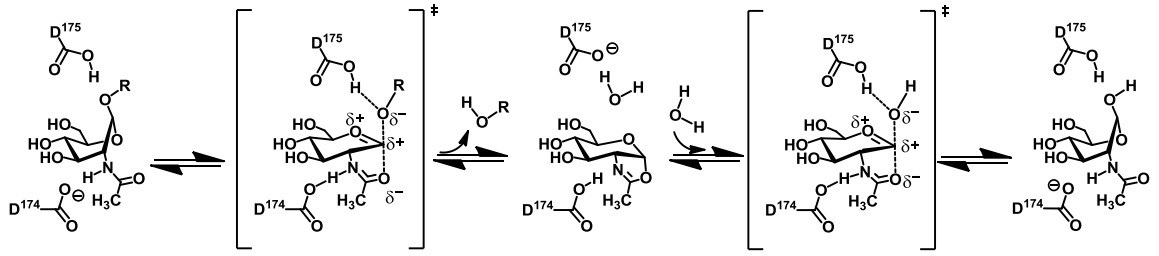


Figure 3.1: *O*-GlcNAcase uses a catalytic mechanism involving substrate-assisted catalysis

The first step of the reaction, cyclization, proceeds via attack of the 2-acetamido carbonyl oxygen on the anomeric centre to form a covalent bicyclic oxazoline intermediate. This step is facilitated by polarization of the 2-acetamido moiety by an enzymic carboxyl group most likely acting as a general base catalyst. Departure of the aglycone is facilitated by general acid catalysis provided by another carboxyl group in the enzyme active site. In the second step, ring opening, the oxazoline intermediate is broken open by the general base-catalyzed attack of a water molecule on the anomeric centre and general acid catalysis to the departing amide group. Both steps occur with inversion of stereochemistry at the anomeric centre such that the overall reaction proceeds with net retention of stereochemistry.

For certain enzyme classes, extensive kinetic isotope effect studies have been carried out that have enabled the definition of geometric and electrostatic features of the transition state[241,242]. Such studies have led to the development of very tight binding inhibitors that are now clinical drug candidates[243]. For glycosidases, however, despite considerable synthetic efforts the very best inhibitors rationally designed as transition state analogues typically have no lower than nanomolar affinities[40,41,76]. Although highly detailed kinetic isotope effect studies of the glycosidases are lacking, a significant body of data has accrued that provides limited insights into the general features of transition states found for these enzymes[62,75].

Large and normal secondary α-D kinetic isotope effect studies suggest that these transition states have a sp^2-hybridized anomeric centre with significant oxocarbenium ion-like character[36,75]. A corollary of such presumed planar oxocarbenium ion-like transition states is that the pyranose ring is expected to assume a conformation in which the C-2, C1, C5, and O-5 atoms adopt a coplanar arrangement[62]. Such limited kinetic isotope effects, in conjunction with Brønsted analyses using substrates having different leaving group abilities, have generally suggested transition states in which the leaving group is largely dissociated[75]. Accordingly, the vast majority of rationally designed glycosidase

inhibitors exploit three principle design features[40,41,76]. First, a sp^2-hybridized centre is often installed at C-1 of the pyranose ring to mimic the geometric requirements of the putative planar oxocarbenium ion-like transition state. Second, nitrogen is often incorporated in place of C-1 or O-5 to mimic the relative positive charge development at these centres within the transition state. Third, and less commonly, conformationally constrained inhibitors are used that force the pyranose ring to adopt a transition state-like conformation[76,81,82,244].

Many inhibitors are defined as transition state analogues by virtue of the principles used in their design and their superficial resemblance to predicted transition state structures. In only a very few cases have efforts been made to define whether these molecules are genuinely transition state analogues, simply serendipitous binders, or actually ground state analogues. Here we investigate the binding of two known potent inhibitors to human *O*-GlcNAcase; a family 84 glycosidase catalyzing the cleavage of *O*-linked 2-acetamido-2-deoxy-β-D-glucopyranoside (*O*-GlcNAc) from serine and threonine residues of post-translationally modified nuclear and cytoplasmic proteins. Our objective being to gain insight into the design of potent transition state analogues as well as to evaluate why *O*-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino N-phenyl carbamate (PUGNAc, Figure 3.2, **1a-g**) based derivatives bearing bulky *N*-acyl groups show weak selectivity[245,246] for human *O*-GlcNAcase over human lysosomal β-hexosaminidase while analogous derivatives of 1,2-dideoxy-2'-methyl-α-D-glucopyranoso-[2,1-*d*]-Δ2'-thiazoline (NAG-Thiazoline, Figure 3.2, **2a-g**) show high selectivities[235].

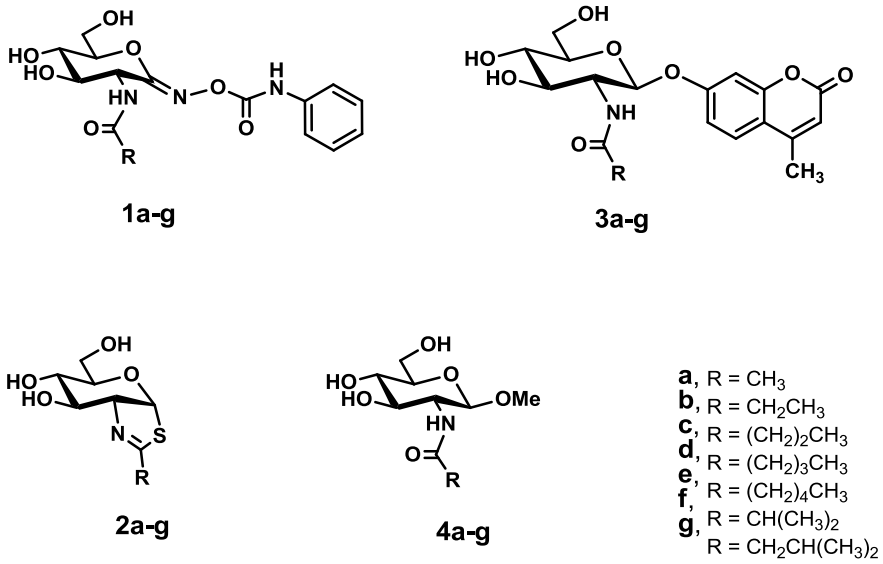


Figure 3.2: Series of inhibitors and substrates used to study transition state analogy of PUGNAc and NAG-thiazoline

O-GlcNAcase uses a two-step catalytic mechanism involving substrate-assisted catalysis[235,247] to form a transient oxazoline intermediate[248] (Fig. 3.1) that is broken down to liberate the free sugar hemiacetal and the protein. This overall catalytic mechanism is common to glycosidases from families 18, 20, 56, and 84, all of which have evolved two strategically positioned carboxyl residues located within the active site that play key catalytic roles[64,222,226,249]. Asp[174] and Asp[175] have been identified as the two key catalytic residues of human *O*-GlcNAcase[247,250]. In the first step of the reaction, the cyclization step, Asp[174] directs and polarizes the 2-acetamido group to act as a nucleophile and form the oxazoline intermediate[247]. Asp[175] meanwhile acts as a general acid, encouraging departure of the aglycone leaving group[247]. In the second step, ring-opening, Asp[174] facilitates departure of the 2-acetamido group while Asp[175] acts as a general base, promoting the attack of a molecule of water to yield the β-hemiacetal product[247]. The structure of the two transition states flanking the oxazoline intermediate have been investigated through substrate structure-function studies[235,247] and α-D kinetic isotope effects[248]. These studies have suggested an "exploded" S_N1-like transition state in which both the leaving group and nucleophilic carbonyl oxygen of the acetamido group are distant from the

anomeric centre. Similar proposals have been made for the mechanistically related family 20 β-hexosaminidases[251]. The most common interpretation of the geometric and electrostatic structure of such transition states for enzymes using substrate-assisted catalysis is that the transition states are oxocarbenium ion-like with a sp^2-hybridized anomeric centre and delocalized positive charge between the endocyclic ring oxygen and the anomeric centre[36,62,75].

The two best characterized inhibitors of *O*-GlcNAcase are PUGNAc and NAG-thiazoline. PUGNAc, first prepared by Vasella and coworkers[252], has been proposed to be a transition state analogue[253] by virtue of its sp^2-hybridized anomeric centre which is thought to have geometric resemblance to an oxocarbenium ion-like transition state (Figure 3.1). NAG-thiazoline, first prepared by Knapp and coworkers[44], is a highly potent competitive inhibitor of both the family 20 β-hexosamindases[44,222] and the family 84 *O*-GlcNAcases[235]. This molecule has an obvious geometric resemblance to the oxazoline intermediate although it has been suggested that it may bind tightly by virtue of its resemblance to a geometrically related transition state. Whether these two highly potent inhibitors of *O*-GlcNAcase are truly transition state analogues, however, has yet to be determined.

Using a combination of linear free energy analyses, computational modeling studies, and X-ray crystallographic analysis of enzyme-inhibitor complexes we evaluate whether these two inhibitors bind as transition state analogues, ground state analogues, or serendipitous binders. New insights into the structure of the transition state for the *O*-GlcNAcase catalyzed hydrolysis of glycosides are gained and, by analogy, into those of related enzymes using such a substrate-assisted catalytic mechanism.

3.4 Results

The successful application of the principles of enzymology to the design of transition state analogues can be gauged by the increasing numbers of potent, rationally designed, transition state analogues. For many inhibitors few criteria, other than their potency, are used to define them as transition state analogues.

In many cases, the designation of transition state analogue may be inadvertently misused to describe opportunistic inhibitors that simply exploit favorable binding situations within enzyme active sites. In recent years there has only been limited progress in developing still more potent transition state analogues of glycosidase inhibitors. Studies directed at clarifying which inhibitors are transition state analogues, however, have proven value in improving our understanding of transition state structures as well as in the iterative design of improved transition state analogues. We therefore have undertaken studies to define the mode of binding of two well characterized inhibitors of human *O*-GlcNAcase; NAG-thiazoline and PUGNAc.

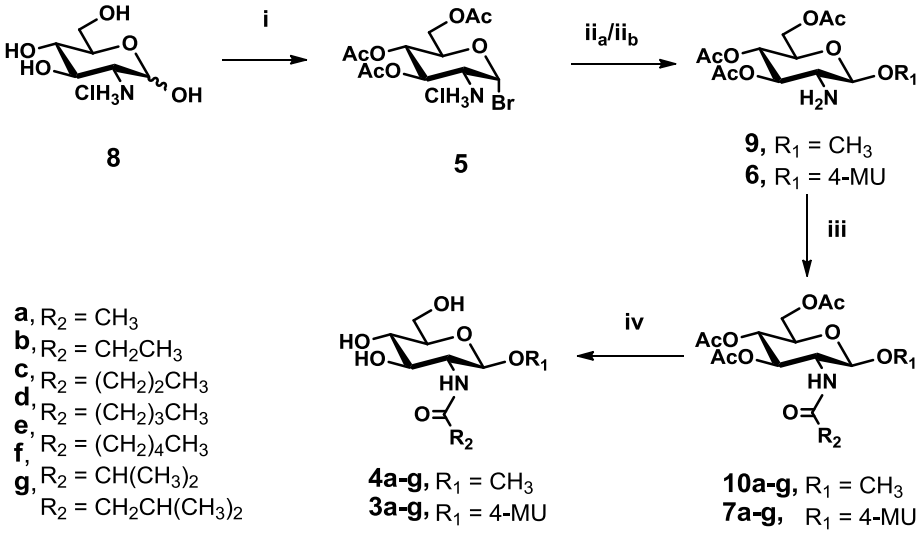
Both NAG-thiazoline and PUGNAc are potent inhibitors of *O*-GlcNAcase (Table 3.1) and bind nearly 240000-fold more tightly than the ground state inhibitor methyl 2-acetamido-2-deoxy-β-D-glucopyranoside (K_I = 11 mM, Table 3.1). Potent inhibition alone, however, is not a good measure of transition state analogy and several other more rigorous approaches have been advanced[239,254-257]. Arguably, the most rigorous approach to establishing transition state analogy has been to use free energy correlations between inhibitor binding and transition state stabilization[239,257]. Using this method to distinguish whether an inhibitor is a transition state analogue, ground state analogue, or simply a fortuitous binder, involves systematic modification of the inhibitor at one position and determining the resulting changes in K_I values. Analogous structural changes are also made within the substrate and k_{cat}/K_m values are determined, since this parameter informs on the structure of the first transition state reached along the reaction coordinate. The inverse logarithm of these values provides free energy terms that, when correlated to each other in a plot of log K_I versus log K_M/k_{cat}, may yield one of several patterns[239,257,258]. The proper interpretation of these linear free energy diagrams rests on several practical assumptions[257]; (1) the chemical step having the transition state mimicked by the inhibitors must be rate limiting for the reaction, (2) the intrinsic reactivity of the substrates must not significantly change as a result of the modifications, (3) variation of the substituent should not affect binding to the enzyme of the unmodified regions of either the substrate or

inhibitor. Once these requirements are satisfied a strong correlation with a slope of one reveals the inhibitor is a transition state analogue. This interpretation stems from the expectation that changes made to a *bona fide* transition state analogue should evoke equivalent changes in the energy of the actual transition state when introduced within the substrate[259,260]. Conversely, scattered plots having weak correlations suggest the inhibitor is a poor transition state analogue or a fortuitous binding inhibitor[40]. Ground state mimics, on the other hand, are best revealed by strong correlations having a slope of one between log K_I of the inhibitor and log K_S values for binding of the substrate (log K_I for a ground state analogue or log K_M for enzymes with a rapid equilibrium substrate binding step may also be used).

Table 3.1: Kinetic data for the three series of inhibitors and the series of substrates

R-Group	PUGNR K_I (μM)	PUGNR Selectivity Family(20/84)	GlcNR-THZ K_I (μM)	GlcNR-THZ Selectivity Family(20/84)	GlcNR *O*-Me $K_{I(GS)}$ (mM)	MU-GlcNR K_m/k_{cat} (mM min)
CH_3	0.046	1	0.07	1	11	0.13
CH_2CH_3	1.2	1	0.12	270	11	0.15
$(CH_2)_2CH_3$	2.5	11	0.25	1500	21	0.19
$(CH_2)_3CH_3$	40	6	1.5	3100	70	1.8
$(CH_2)_4CH_3$	220	≥ 5	57	100	205	1.8
$CH(CH_3)_2$	9	2	1.6	700	147	5.0
$CH_2CH(CH_3)_2$	190	≥ 6	5.7	190	224	110

It is noteworthy that in two closely related glycosidases from family 13, correlations of log K_I versus log K_M/k_{cat} have yielded slopes other than one and these have been interpreted in entirely different ways[82,244]. These two studies, however, have been carried out using enzymes in which mutations have been made at several different residues within the active sites of these enzymes. It therefore seems prudent to make systematic structural variations at only one site. The best approach, first described by the Bartlett group[239], is to make a series of *small* structural perturbations to the inhibitor at *one* site since these are most likely to report in a sensitive manner on changes in the differences between the actual transition state and the putative transition state mimic.



Scheme 3.1: Synthesis of a series of ground state analogues
(i) AcBr, heat; (ii_a) pyridine, MeOH; (ii_b)a) acetone, 4-MU, Na 4-MU; b) K_2CO_3, Et_2O; c) HCl (iii) RCOCl, Et_3N, DCM; (iv)a) NaOMe, MeOH; b) 20:1 MeOH : Acetic acid.

For *O*-GlcNAcases, we know that the N-acyl group of the substrate is critically involved in catalysis and must undergo conformational, electrostatic, and positional variation within the active site of the enzyme on proceeding from the ground state to transition state[235,248]. We speculated that modifications of this group would report clearly on the transition state structure for these enzymes and that derivatives of PUGNAc (**1**) and NAG-thiazoline (**2**), both previously synthesized within our laboratory, would be useful probes[235,245]. The structural changes within these inhibitors are small increases in the volume of the alkyl group of the N-acyl moiety and they give rise to K_I values varying by over 5-orders of magnitude when assayed against human *O*-GlcNAcase (Table 3.1)[245]. Such a wide variation in K_I values is beneficial for establishing a meaningful correlation. With these molecules already in hand, we then prepared (Figure 3.3) the requisite series of substrates with analogous modifications to the N-acyl group and measured the second-order rate constants for human *O*-GlcNAcase catalyzed hydrolysis of these compounds (Table 3.1). Satisfying the first requirement of the transition state analogy approach (*vide supra*), we have previously shown, using a combination of kinetic isotope effects[248] and a series

of substrates bearing different leaving groups[235,247], that the second-order rate constant for substrates bearing 4-methylumbelliferyl leaving group reflects the first chemical step. The second requirement of this approach is fulfilled since the spontaneous hydrolysis rates of these substrates are indistinguishable within error. Owing to limitations in the solubility of these substrates, however, we were unable to obtain reliable K_S or even K_M values making correlations with K_S or K_M values impossible. A series of methyl glycosides (Figure 3.2, **4a-g**) was also prepared (Scheme 3.1) as simple ground state substrates and were tested as competitive inhibitors to obtain inhibition constants (Table 3.1) that are taken to be equivalent to true dissociation constants. Testing parent glycoside **4** as a representative member of these ground state inhibitors revealed, as expected, a pattern of competitive inhibition (Figure 3.3) and validated their use as ground state analogues binding within the *O*-GlcNAcase active site. Together, these four sets of compounds allowed us to delineate the nature of inhibition of human *O*-GlcNAcase by NAG-thiazoline and PUGNAc.

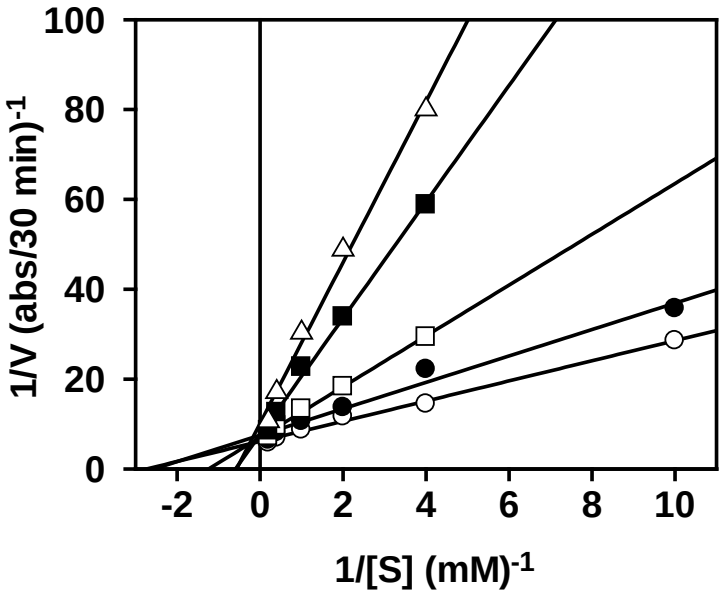


Figure 3.3: Inhibition of human *O*-GlcNAcase using ground state substrate Methyl 2-acetamido-2-deoxy-D-glucopyranoside
Inhibition of *O*-GlcNAcase-catalyzed hydrolysis of *p*NP-GlcNAc (3a) by the methyl-glycoside (4a) shows a pattern of competitive inhibition. The concentrations of 4a (mM) used were 85.0 (Δ), 42.5 (■), 11.0 (□), 5.5 (●), 0.0 (○).

Correlations of log K_I values for the series of NAG-thiazoline derivatives (Figure 3.2, **2a-g**) with the corresponding logarithm of the second-order rate constants K_M/k_{cat} for the *O*-GlcNAcase catalyzed hydrolysis of a series of 4-methylumbelliferyl glycoside substrates (Figure 3.2, **3a-g**) reveals a strong correlation ($r^2 = 0.98$) with a slope of 0.97 ± 0.06 (Figure 3.4c). This correlation indicates that the incremental differences in binding energy related to the elaboration of the thiazoline ring with aliphatic chains of increasing length are nearly identical to those perturbations observed for the transition state of the enzyme catalyzed hydrolysis of the substrates bearing analogous changes. Correlations of log K_I values with the corresponding logarithm of the inhibition constants (log $K_{I(GS)}$) for the ground state inhibitors showed no clear correlation. These two patterns are very similar to those observed by Bartlett and coworkers[239,261], which they used to show that phosphonamidates are transition state analogues for the zinc peptidases. Together these data strongly support the proposed mechanism involving substrate-assisted catalysis from the

acetamido group and reveal NAG-thiazoline as a good transition state mimic for *O*-GlcNAcase catalyzed glycoside hydrolysis.

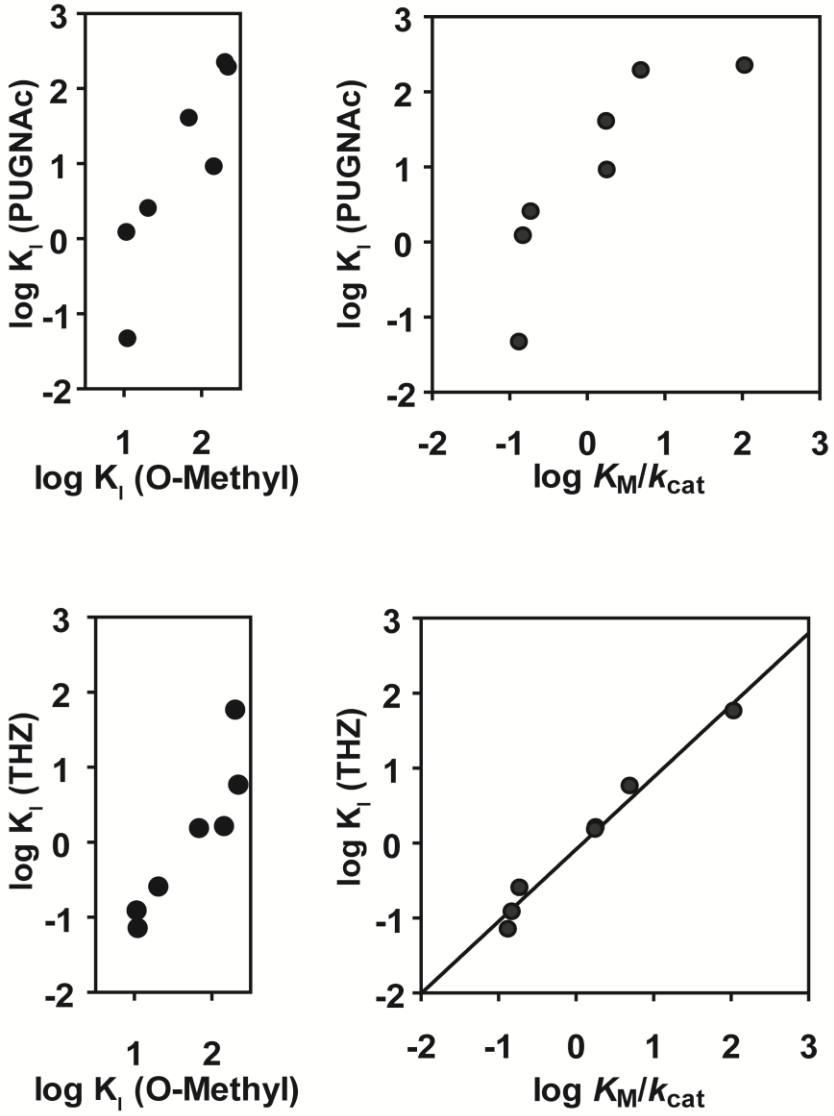


Figure 3.4: Comparison of transition state analogy free energy diagrams for NAG-thiazoline and PUGNAc derivatives
A) Comparison of thiazoline series of inhibitors (2a-g) K_I with methyl-glycoside series (4a-g) K_I. B) Comparison of thiazoline series of inhibitors (2a-g) K_I with substrate (3a-g) K_M/k_{cat}. C) Comparison of PUGNAc series of inhibitors (1a-g) K_I with methyl-glycoside series (4a-g) K_I. D) Comparison of PUGNAc series of inhibitors (1a-g) K_I with substrate (3a-g) K_M/k_{cat}.

Using this approach it is possible only to comment on the transition state resemblance in regions where the inhibitor varies in structure from the substrate. The specific structural features probed by NAG-thiazoline encompass the entire thiazoline ring including the position of all associated atoms including C-1 and C-

2 of the pyranose ring. Notably, C-1 of NAG-thiazoline is an sp^3-hybridized carbon centre rather than an sp^2-hybridized centre or basic nitrogen, features that are often assumed to be essential features of transition state mimics for glycosidases. The two best studied iminosugar inhibitors are the isofagomine class, wherein C-1 is replaced with nitrogen and the endocylic ring oxygen with a methylene unit, and the nojirimycin class of inhibitors, wherein the endocylic ring oxygen is replaced by a nitrogen atom. Both of these inhibitors have been proposed to be transition state mimics although more recent experimental data suggests that, at least for some enzymes, this is not the case. Indeed, in a crystallographic complex of GalNAc-isofagomine bound to a family 20 β-hexosamindase that uses substrate-assisted catalysis, it was observed that the protonated nitrogen engaged in an adventitious electrostatic interaction with the general acid/base catalytic residue[262]. For nojirimycin, scattered linear free energy data was observed for two related α-glucosidases, suggesting that this inhibitor is an adventitious binder to the two enzymes studied[40,244].

For the series of PUGNAc derivatives a poor correlation (r^2 = 0.73, Fig**.** 5b) with a slope of 1.07 ± 0.29 was observed in a plot of log K_I values of the inhibitor versus log K_M/k_{cat} values for the series of glycoside substrates. As well, a poor correlation is observed between log K_I values and log $K_{I(GS)}$ values for the ground state inhibitors (Figure 3.4a). Accordingly, the scatter within both of these correlations indicates that incremental changes of the N-acyl chain of the parent compound are not viewed in a similar manner within the active site of the enzyme as the incremental elongation of the N-acyl chain of the substrate, indicating that PUGNAc is either a poor mimic of the *O*-GlcNAcase-catalyzed transition state or simply a fortuitous binder. Consistent with this interpretation, Bartlett has described scatter of the data as an indication that the inhibitor may not closely resemble the transition state[257].

In the context of the traditional design principles used in developing transition state analogue inhibitors of glycosidases, the apparent weak transition state analogy of PUGNAc is surprising. The trigonal anomeric centre found within PUGNAc is often considered to be a vitally important feature and many

potent inhibitors contain such a moiety. It is worth noting that NAG-thiazoline, which lacks all of the key design features of transition state analogues, is an equally potent inhibitor as PUGNAc. NAG-thiazoline may therefore mimic a transition state that is significantly different than the canonical oxocarbenium ion-like transition state that is widely perceived to be found along the reaction coordinate of glycosidases.

In view of this possibility we considered previous results in this new context. The large secondary α-deuterium kinetic isotope effects (αD-KIE) of 1.14 ± 0.02 measured previously[248] with human *O*-GlcNAcase suggests that the enzyme stabilized transition states have oxocarbenium ion-like character with significant sp^2-character at the anomeric centre. Interpreting this KIE as revealing a strict trigonal geometry at the anomeric centre in the transition state is complicated, however, as underscored by studies of known S_N2 reactions of the methoxymethyl model system. Depending on the nature of the incoming nucleophile, αD-KIE values traditionally held to be consistent with either S_N1 (k_H/k_D = 1.18) or S_N2 (k_H/k_D = 0.99) were observed[263]. Further complicating matters is that equilibrium isotope binding effects have been observed in several cases where deuterium is present at the anomeric centre[264]. Accordingly, in the absence of other supporting heavy atom isotope effects at the reaction centre, interpreting the magnitude of αD-KIE values at acetal centres is difficult and cannot be used to discern the extent of nucleophilic participation[265]. Therefore, it seems equally possible that the αD-KIE values determined may reflect a nonplanar transition state having greater nucleophilic participation and correspondingly less p-orbital character at the anomeric centre.

Such differences in the extent of nucleophilic participation and involvement of the leaving group have been well documented for the N-ribosyl hydrolases and have developed with the support of X-ray crystallographic analyses of enzyme-inhibitor complexes into the concept of transition state poise. The transition state poise is defined by the bond orders between the anomeric centre and both the leaving group and nucleophile which can vary significantly as a function of the position of C-1 along the reaction coordinate. In this context the

glycosidases[221], like the N-ribosyl hydrolases, appear to use an electrophilic migration mechanism[266] for which the reaction coordinate diagram can be defined by the motion of the anomeric carbon and where the total bond order between the anomeric centre, nucleophile, and leaving group are significantly less than one. Within crystallographically visualized Michaelis complexes of β-glycosidases[63,64] the anomeric carbon typically adopts a position *above* the plane of the pyranose ring with the exocyclic oxygen hydrogen bonding to the general acid catalytic group. Within the newly formed product, the anomeric carbon is now positioned *below* the plane of the pyranose ring bonded to the nucleophile. For *O*-GlcNAcase, this nucleophile is the carbonyl oxygen of the acetamido group. The transition state must accordingly resemble a structure sitting somewhere along this pathway defined by these two end points and not, as widely believed, equidistant between them as would be found for an oxocarbenium ion.

Taft-like linear free energy analyses (log k_{cat}/K_M versus σ^*) of the *O*-GlcNAcase catalyzed hydrolysis of N-fluoroacetamido glycosides have strongly implicated the acetamido group in nucleophilic participation at the transition state of the rate determining chemical step[235,248]. Furthermore, the linear free energy analyses we describe here indicate that NAG-thiazoline is a transition state analogue, an observation also consistent with nucleophilic participation since the anomeric carbon and the thiazoline sulphur atom are only 1.85 Å apart from each other. However, whether NAG-thiazoline geometrically resembles the transition state along the reaction coordinate, some point along the reaction coordinate that is close to this saddle, or that the enzyme catalyzed reaction is $D_N + A_N$ with a short lived, yet discrete oxocarbenium ion, as has been suggested for β-galactosidase by Richard and coworkers[267] currently remains untested.

Assuming the transition state resembles NAG-thiazoline, as suggested by the free energy analyses, one can approximate the bond orders to both the leaving group and nucleophile. By assuming the interatomic distance between C-1 and the carbonyl oxygen in the transition state is 1.85 Å (the length of the C-S bond length within NAG-thiazoline) a Pauling bond order of 0.26 to the

nucleophile can be calculated. The Brønsted analyses and the αD-KIE values previously measured for human *O*-GlcNAcase suggest little involvement of the leaving group in the transition state. On these bases and the crystallographic analyses of the NAG-thiazoline bacterial *O*-GlcNAcase complex, we estimate the bond order to the leaving group is nearly zero (0.01). In this structure, a molecule of water sits approximately 2.8 Å above the anomeric carbon in a position that is expected to be occupied by the glycosidic oxygen of the substrate within the Michaelis complex and it is held in position by a hydrogen bond to the enzymic general base catalyst[268]. These data, in combination with other studies of glycosidase mechanism suggest that human *O*-GlcNAcase possesses an oxocarbenium ion-like transition state which is significantly stabilized by nucleophilic participation of the 2-acetamido group. This nucleophilic participation, in combination with the action of the catalytic residue hydrogen bonding with the amide nitrogen (Asp^{174} in human O-GlcNAcase and Asp^{242} in the *B. thetaiotaomicronan* β-glucosaminidase), delocalizes excess positive charge away from the C1–O5 region of the pyranose and on into the oxazoline ring. Such a dissociative transition state is consistent with the previously described Taft-analyses[235,268], kinetic isotope effects[248], and the free energy analyses we detail here. It is also reminiscent of those determined empirically for N-ribosyl transferases using a D_NA_N mechanism but most likely involving slightly greater nucleophilic participation[241,269].

Table 3.2: Data collection and refinement statistics for the structure solution of *B. thetaiotaomicron* GH84 *O*-GlcNAcase with NButGT

	*Bt*GH84 with N-butyl-thiazoline
Data collection	
Space group	C2
Cell dimensions	
a, *b*, *c* (Å)	185.1, 51.7, 172.8
α, β, γ (°)	90.0, 100.1, 90.0
Wavelength	0.9930
Resolution (Å)	40-2.25 (2.33-2.25)*
R_{sym} or R_{merge}	0.086 (0.47)
$I / \sigma I$	18 (3)
Completeness (%)	99.9 (99.9)
Redundancy	4.5 (4.4)
Refinement	
Resolution (Å)	2.25
No. reflections	77267
R_{work} / R_{free}	21 / 28
No. atoms	
Protein	9500
Ligand	32
Waters	554
B-factors	
Protein	30
Ligand/ion	19
Water	30
R.m.s deviations	
Bond lengths (Å)	0.023
Bond angles (°)	1.99

*Highest-resolution shell is shown in parentheses.

To gain greater insight into the basis by which NAG-thiazoline mimics the transition state whereas PUGNAc does not, we carried out crystallographic analyses of the selective *O*-GlcNAcase inhibitor, NButGT, complexed to a bacterial homologue of human *O*-GlcNAcase. This *Bacteroides thetaiotomicronan* β-glucosaminidase is an excellent model of human *O*-GlcNAcase since it too processes *O*-GlcNAc modified proteins and also has striking sequence conservation when compared to the human *O*-GlcNAcase. The structure of the *B. thetaiotaomicron* GH84 *O*-GlcNAcase, in complex with NButGT was solved at a resolution of 2.25Å by molecular replacement. The

structure crystallized in a C2 space group (see methods) in which two molecules of *Bt*GH84 lie in the asymmetric unit in the same orientation separated by a translation-only vector of 0, 0, 0.5. The *Bt*GH84 structure can be traced from residue 4 of the mature protein through to residue 589, with two short breaks in molecule B from 51-53 and 455-456. The C-terminal "CBM32-like" β-sheet domain, which was partially ordered in the original native structure solution is completely disordered in this crystal form and could not be modeled. We compared the structure of this complex with a previously determined structure of NAG-thiazoline bound to the same enzyme[268] as well as to a complex of PUGNAc bound within the active site of a *Clostridium perfringens* homologue[253,270] of *O*-GlcNAcase. Gratifyingly, NButGT and NAG-Thiazoline bind identically within the active site and the pyranose ring of the inhibitors within both complexes can be overlayed exactly (data not shown). We find the only difference between the two structures is where electron density corresponding to the alkyl chain of NButGT extends into an active site pocket (Fig. 5A) that has been previously predicted to accommodate such bulkier groups[235,268]. The confirmation of identical binding modes for the parent inhibitor and NButGT is important in the context of the transition state analogy studies described here since it satisfies the third implicit assumption of the approach that analogous inhibitors bind in the same orientation.

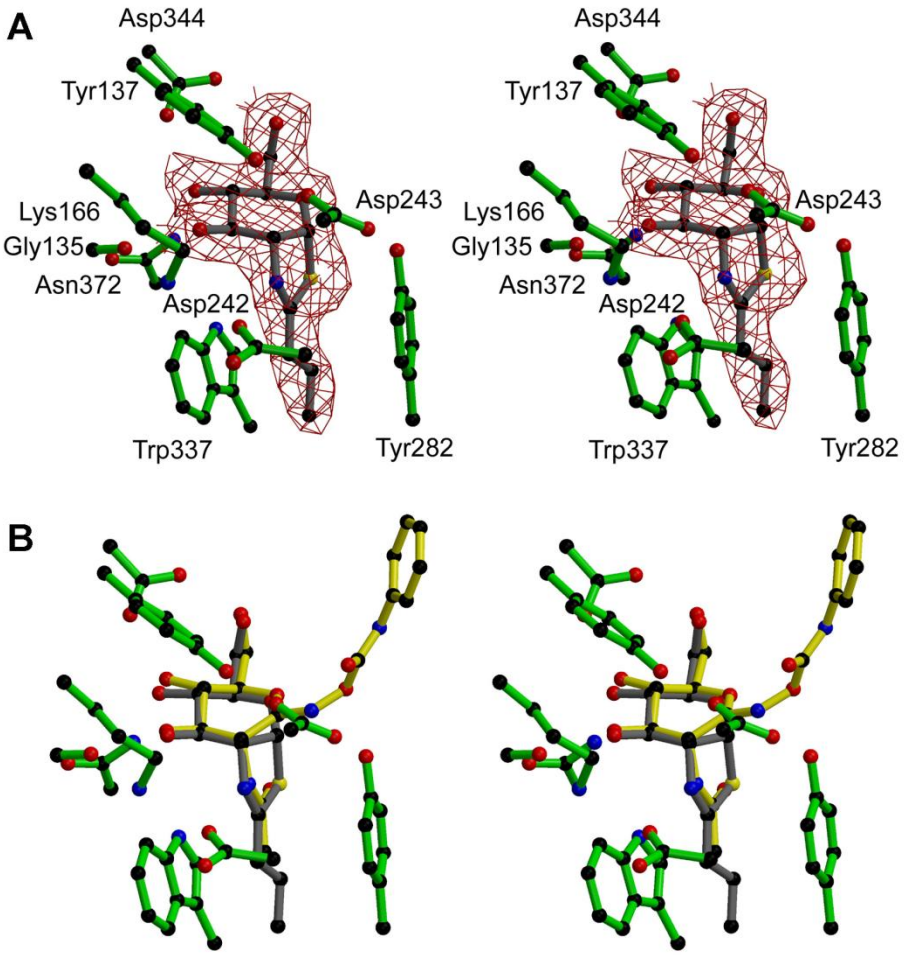


Figure 3.5: Structural analyses of the binding of NButGT to *Bt*GH84 and its comparison to binding of PUGNAc to *Cf*GH84
A) Observed electron density for the NButGT along with its flanking residues. The map shown is a $2F_{obs}$-Fcalc electron density map contoured at 1σ and is shown in divergent ("wall-eyed") stereo. B) Structural overlap of the *Bt*GH84 NButGT complex (ligand,grey; protein green) with the PUGNAc complex of the *Clostridium perfringens* GH84 enzyme[253] (yellow). The overlap was optimized using the main-chain protein atoms only, in a radius of 15Å of the ligands, using the program QUANTA (Accelrys, San Diego, USA). This figure was drawn using MOLSCRIPT[271] and BOBSCRIPT[272].

It has been noted previously by us that derivatives of PUGNAc having bulky N-acyl groups show weak selectivity toward human *O*-GlcNAcase over human β-hexosaminidase[245,246] while the analogous NAG-thiazoline derivatives show high selectivities[235]. In an effort to better understand the molecular basis for these differences in selectivity we decided to compare the binding modes of NButGT and PUGNAc by overlaying the structures of these two inhibitors bound

to the *O*-GlcNAcase homologues (Fig. 6B). The molecular basis for this discrepancy in selectivity must stem from structural differences between these molecules and how they interact with the active site of human *O*-GlcNAcase.

There are several obvious structural differences between NAG-thiazoline and PUGNAc. The first, and most obvious difference is that PUGNAc bears a phenyl carbamate moiety which is absent in NAG-thiazoline. This substituent extends out from the active site and engages in several interactions that are not found for NAG-thiazoline. The exocyclic oxime nitrogen of PUGNAc hydrogen bonds with the general acid/base catalytic group (Asp^{175} within human *O*-GlcNAcase[247] and Asp^{243} within the *B. thetaiotaomicronan* β-glucosaminidase[268]) and the phenyl group of PUGNAc also engages in several hydrophobic and stacking interactions with residues in the aglycon binding site. It is interesting to note that *N*-acetylglucosamino-1,-5-lactone oxime (LOGNAc), which lacks the phenyl carbamate group of PUGNAc, binds 30-fold more poorly[196] than PUGNAc providing an indication of the importance of these serendipitous interactions. A second structural difference between the two inhibitors is that the hybridization at the anomeric centre varies. This variation has an impact on the preferred conformation of the pyranose ring; for PUGNAc the pyranose ring adopts a ^4E conformation enforced in part, by the sp^2-hybridized anomeric centre, whereas NButGT sits in a $^4C_1$ chair that is slightly distorted toward the ^4E conformation due to the rigidity of the thiazoline ring. The structural consequence of these variations is that the substituents of the pyranose ring, particularly the 3-and 4-hydroxy groups, adopt subtly different orientations. Such apparently minor differences in the position of both ring and substituent atoms can have pronounced effects on both Van der Waals interactions and hydrogen bonds. A third obvious structural difference is that the N-acyl group of PUGNAc is free while NAG-thiazoline's orientation is locked. This difference also results in important variations in the pK_a of the protonated nitrogen at the 2-position of the pyranose ring that will be discussed in greater detail in context of the computational studies (*vide infra*). The thiazoline ring is nestled between two tryptophan residues which presumably serve to accurately position the

acetamido group of the substrate and stabilize the positive charge that develops at the transition state during formation of the oxazoline ring. Not only does the thiazoline ring engage in stacking interactions with these residues but it also orients the alkyl substituent of the inhibitors. As seen in the structural comparison, (Figure 3.5) the alkyl substituent of NButGT is pointing directly down the active site cavity, whereas the N-acetamido group of PUGNAc is pointed at a slightly different angle. Consistent with this structural view, human *O*-GlcNAcase is better able to cope with elongation of the aliphatic chain of NAG-thiazoline derivatives as compared to analogous derivatives of PUGNAc (Table 3.1). Because Van der Waals forces are powerfully distance dependent this apparently small difference may account for the significant differences in selectivity observed by us between the derivatives of these two inhibitors.

These crystallographic analyses primarily furnish details of the geometric features mediating interactions in the enzyme-inhibitor complexes but do not provide great insight into transition state structure nor the electrostatic features of the inhibitors and reaction coordinate species. We therefore reasoned that further insight into the potency, transition state analogy, and selectivity of both NAG-thiazoline- and PUGNAc-derived compounds toward human *O*-GlcNAcase could be gained from a comparison of their molecular electrostatic potential surfaces (MEPs) with that of the enzyme-bound intermediate or, better still, the enzymic transition state. Work by the Schramm group has shown that the similarity of transition states and inhibitors may be probed qualitatively and quantitatively using MEPs[273]. Accordingly, we modeled both inhibitors, PUGNAc-analogue **I** and NAG-Thiazoline **II**, as well as three species found along the reaction coordinate for the cyclization step (Figure 3.6); an intact methyl glycoside ground state **III**, the transition state **IV** we predict on the basis of the TS-analogy studies featuring a distorted oxazoline in which the C-1 carbonyl oxygen bond distance was constrained to the C-S bond distance within NAG-thiazoline, and lastly the oxazoline intermediate **V** (Figure 3.6). It is important to note that these species are expected to provide a reasonable qualitative, though not quantitatively accurate, description of the reaction coordinate.

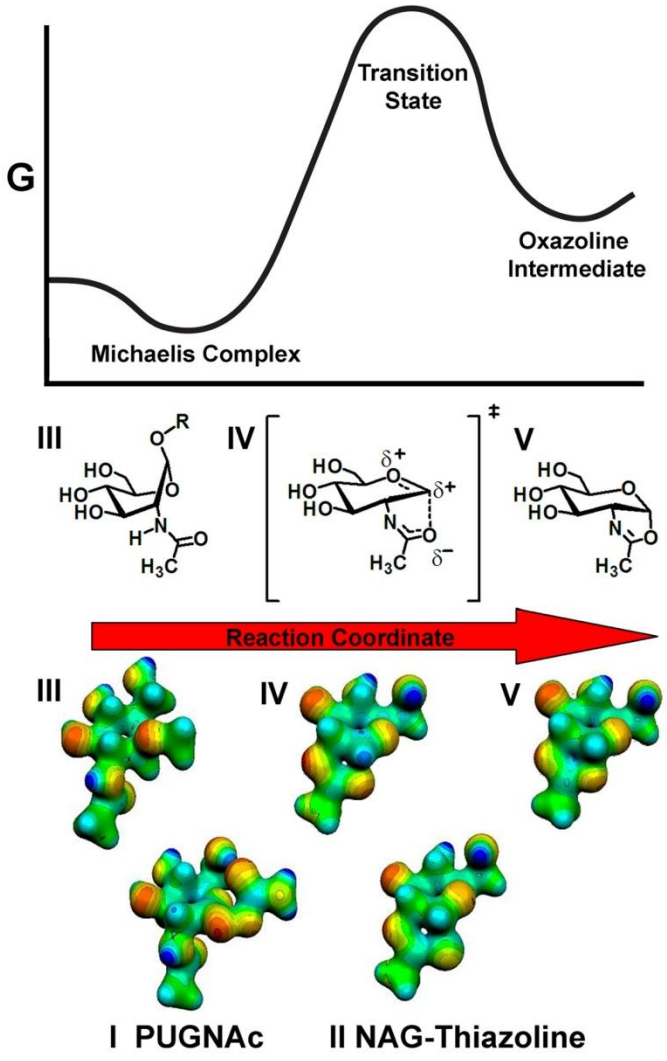


Figure 3.6: A simplified reaction coordinate for human *O*-GlcNAcase catalyzed *β*-glucosaminide hydrolysis
MEPs for PUGNAc-analogue I (modeled with methyl rather than phenyl) and NAG-Thiazoline II as well as ground state III, transition state model IV, and oxazoline V. NAG-Thiazoline II is shown to share structural and electrostatic properties of both transition state model V and oxazoline IV. PUGNAc-analogue I is shown to share significant electrostatic properties with the ground state III although it does display some transition state-like conformational properties. MEPs (which all share a common scale) are shown projected onto the 0.04 electrons / $Å^3$ isodensity surface.

A comparison of our transition state model **IV** and the oxazoline **V** demonstrates that C–O bond extension is associated with both increased planarity about the anomeric carbon centre as well as a contraction of the endocyclic C1–O5 bond. The most striking difference between oxazoline **V** and

transition state model **IV** however, is the build-up of positive charge, not only at the anomeric carbon centre, but also on the anomeric proton. This redistribution of charge is much more apparent than other changes centered about, for example, the endocyclic ring oxygen or within the oxazoline ring.

To afford a careful evaluation of this difference we verified that this change in the MEP at the anomeric centre was independent of the protonation state of oxazoline **IV** by calculating similar MEPs for the associated oxazolinium ion, thiazolinium ion and oxazolinium ion-based transition state model (in which the anomeric C–O bondlength of the oxazolinium ion was constrained to that found in the thiazolinium ion, Figure 3.7). A comparison of MEPs for the oxazolinium ion and the oxazolinium ion-based transition state model showed similarly significant changes in charge distribution on the anomeric carbon and hydrogen atoms (with an increase in positive character in the transition state model, Figure 3.7).

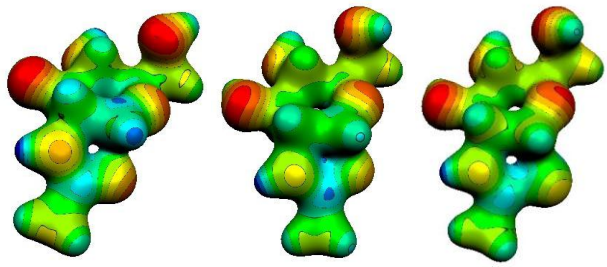


Figure 3.7: Positively charged MEPs
MEPs for an oxazolinium-ion based transition state model (left), an oxazolinium ion (center), and a thiazolinium ion (right)

An analysis of the structure and MEP of thiazoline **II** reveals that it mimics some aspects of oxazoline **V** while in other aspects it more closely resembles the transition state model **IV**. The hexopyranose ring conformations of oxazoline **V**, thiazoline **II**, and transition state model **IV** are all subtly different. The model used for the reactivity of many glycosidases is one in which electrophilic migration of the anomeric carbon centre enforces a conformational itinerary following the $^{1}S_{3}$–$^{4}H_{3}$–$^{4}C_{1}$ path from Michaelis complex to enzyme-bound intermediate (Figure 3.6).

A comparison of our optimized structures indicate that, whereas the pyranose ring of transition state model **IV** is best described as a $^4H_3$ conformation and that of NAG-thiazoline as adopting a $^4C_1$ conformation, oxazoline **V** is best described as adopting a $^4H_5$ conformation. This conformation is a result of C3–C2–C1–O5 co-planarity favored by formation of the 5-membered oxazoline ring which induces strain in the pyranose ring. Although the precise conformational itinerary traveled along the reaction coordinate is uncertain, it seems likely that the less strained $^4C_1$ conformation of thiazoline **II** lies conformationally closer to our transition state model.

Although NAG-thiazoline **II** bears a significant geometric resemblance to transition state model **IV**, a comparison of the MEPs for the three species under consideration reveal some subtle differences. The larger Van der Waals radius of sulfur compared to oxygen is readily apparent when the MEPs are visualized. The potential surrounding the sulfur atom is also appreciably less negative than that surrounding either the oxygen atom of either oxazoline **V** or transition state model **IV**. Furthermore, the potential distribution around the anomeric carbon and hydrogen atoms of thiazoline **II** most closely resemble that of oxazoline **V** when compared to that of transition state model **IV**.

The structural properties of thiazoline **II** and its MEP indicate that it is likely to be a reasonable mimic of a species that lies on the reaction coordinate between the region of the enzymic transition state and the enzyme-bound intermediate. PUGNAc-analogue **I**, in contrast, displays a number of markedly different structural and electrostatic characteristics from oxazoline **V** and the transition state model **IV**.

The hexopyranose ring of PUGNAc-analogue **I** adopts a ^4E conformation and, as such, is a marginally good mimic of the conformation of transition state model **IV**. In a number of critical respects, however, PUGNAc-analogue **I** is an unsatisfactory mimic of either the oxazoline **V** or its preceding transition state. PUGNAc-analogue **I** fails to capture any significant bonding between the *N*-acetamido moiety and the anomeric carbon centre: as indicated previously, this interaction is a critical component of the reaction pathway at the transition

state[247]. More critically, the MEP of PUGNAc-analogue **I** indicates that there is significantly more negative potential in the region of the anomeric carbon atom. This negative potential is in striking contrast with the more positive potential observed in oxazoline **V**, thiazoline **II**, and the transition state analogue **IV**. In this respect PUGNAc-analogue **I** is significantly more substrate-like (structure **III**). The marked changes in charge at this centre along the reaction coordinate are likely to be a significant target for *selective* transition state stabilization by human *O*-GlcNAcase and so any failure to mimic transition state charge at this centre is likely to hinder the ability of PUGNAc and derivatives to act as transition state analogues.

Another notable difference between NAG-thiazoline and PUGNAc is that the polarizing residue (Asp^{174} in human *O*-GlcNAcase[247] and Asp^{242} in the *B. thetaiotaomicronan* β-glucosaminidase[268]) is likely to engage in a much stronger interaction with oxazoline **IV** and thiazoline **II** as compared to PUGNAc-analogue **I**. As noted previously[247], the pK_a of the nitrogen atom will change during the cyclization step from around 15, for an amide proton, to a pK_a of about 5.5, the value found for the 2-methyl-Δ^2-oxazolinium ion[274]. The pH-rate profile ($\log(K_M/k_{cat})$ vs. pH) for the hydrolysis of *p*-nitrophenyl β-N-acetylglucosaminide allows us to assign a macroscopic pK_a value of approximately 5.0 to $Asp^{174\ 247}$. The pK_a of Asp^{174} therefore, much more closely matches those of the conjugate acids of thiazoline **II** and oxazoline **IV** than that of PUGNAc-analogue **I**. As hydrogen bonds between residues with closely matched pK_a values are stronger than those between residues whose pK_a values differ greatly we may reasonably expect thiazoline-derived inhibitors to more tightly bind to Asp^{174} than PUGNAc-derived inhibitors. From the perspective of Asp^{174}-binding thiazoline-derived analogues will therefore be better transition state analogues than PUGNAc-derived inhibitors, which possess greater ground state character in this particular interaction.

3.5 Conclusion

Using three approaches, including linear free energy correlations, X-ray crystallographic analyses, and computational modeling, we find that NAG-thiazoline is a TS-analogue for the human *O*-GlcNAcase catalyzed hydrolysis of β-glucosaminides while PUGNAc is either a poor TS-analogue or a serendipitous binding inhibitor. These observations are surprising in view of the conventional approaches used in the design of TS-analogues of glycosidases. Specifically, PUGNAc, which incorporates a traditional transition state design feature, a sp^2-hybridized anomeric centre, is a poor TS-mimic. NAG-thiazoline, which has a sp^3-hybridized anomeric centre, is a good TS-analogue even though such a geometry is not considered TS-like. It may appear counterintuitive that NAG-thiazoline, given the tetrahedral geometry of its anomeric centre, should resemble the transition state for the hydrolysis of an acetal since associated transition states are widely perceived as dissociative with significant oxocarbenium ion-like character. It seems more obvious that NAG-thiazoline may resemble the oxazoline structure found further along the reaction coordinate and be more of a mimic of a high energy oxazoline intermediate. However, the C-S bond length of the thiazoline (1.86 Å) is considerably longer than the C-O bond length (1.45 Å) expected for the corresponding oxazoline. This longer bond distance results in distortion of the pyranose ring of the thiazoline and mimics a bond order of 0.26 between the nucleophilic carbonyl oxygen and the anomeric centre; a bond order consistent with the significant nucleophilic participation established previously[235]. Crystallographic studies, computationally determined structures, and molecular electrostatic potential surfaces for NAG-thiazoline- and PUGNAc-based inhibitors of human *O*-GlcNAcase support this view. On the basis of these qualitative MEPs, thiazoline derived *O*-GlcNAcase inhibitors and related structures appear to be a better framework for capturing a substantial fraction of the transition state binding energy used by the enzyme when compared to PUGNAc-derived inhibitors. These collective results, in combination with earlier kinetics studies, also suggest late transition state poise for this electrophilic migration mechanism with significant nucleophilic

participation and essentially no involvement of the leaving group – an observation consistent with studies of some of the N-ribosyl hydrolases. This improved understanding of the reaction coordinate adopted by human *O*-GlcNAcase, however qualitative, should provide the basis for the development and refinement of transition state analogues of human *O*-GlcNAcase having optimized geometries.

3.6 Materials and methods

3.6.1 General

All solvents were dried prior to use. Synthetic reactions were monitored by TLC using Merck Kieselgel 60 F254 aluminium-backed sheets. Compounds were detected by charring with a 10% concentrated sulfuric acid in ethanol solution and heating. Flash chromatography under a positive pressure was performed with Merck Kieselgel 60 (230-400 mesh) using specified eluants. 1H and 13C NMR spectra were recorded on a Varian AS500 Unity Innova spectrometer at 500 MHz (125 MHz for 13C) with chemical shifts quoted relative to CDCl3 or CD3OD where appropriate. Elemental analyses of all synthesized compounds used in enzyme assays were performed at the Simon Fraser University Analytical Facility.

3.6.2 Kinetic analysis of *O*-GlcNAcase:

Inhibition Assays - *O*-GlcNAcase was expressed as previously described no earlier than 14 days prior to use. All assays were carried out in triplicate at 37 °C for 30 minutes using a stopped assay procedure in which the enzymatic reactions (50 μL) were quenched by the addition of a 5-fold excess (250 μL) of quenching buffer (200 mM glycine, pH 10.75). Assays were initiated by the addition, via syringe, of enzyme (3 μL), and in all cases the final pH of the resulting quenched solution was greater than 10. Time-dependent assay of *O*-GlcNAcase revealed that the enzyme was stable in the assay buffer (50 mM NaH_2PO_4, 100 mM NaCl, 0.1% BSA, pH 6.5) and that the reaction progress was

linear over the entire assay time. The progress of the reaction at the end of 30 minutes was determined by measuring the amount of 4-nitrophenol liberated (E_{400} = 12885 M^{-1} cm^{-1}) as determined by measuring absorbance at 400 nM using a 96-well plate (Sarstedt) reader and 96-well plate reader (Molecular Devices, Model SpectraMax 340). *O*-GlcNAcase was used in the inhibition assays at a concentration of 0.0406 μg/μL using pNP-GlcNAc (Sigma) as a substrate at a concentration of 0.5 mM. All inhibitors were tested at seven concentrations ranging from 3 times to 1/3 K_I, with the exception of the assay of inhibitor **4e** where such high concentrations of inhibitor could not be reached because of the high K_I value of 205 mM and the limited solubility of the compound. In this case **4e** was tested at seven concentrations ranging from 1.5 times to 1/5 K_I. K_I values were determined by linear regression of data from Dixon plots.

Enzyme kinetics assays - Activity of *O*-GlcNAcase toward 4-methylumbelliferlyl 2-acylamido-2-deoxy-β-D-glucopyranoside derivatives measurements was performed essentially as previously described[235]. Saturation kinetics were not observed, however due to the limited solubility of the series of substrates. Substrates were dissolved in 5% DMF at a concentration of 1 mM and tested as substrates for *O*-GlcNAcase at 10, 20, 30, and 40 μM in triplicate. The slope of the line through these points furnished a value proportional to the second order rate constant. The DMF concentration was maintained at a final concentration of 2.5% in all assays and did not affect the activity of *O*-GlcNAc as determined by control assays containing no DMF.

Theoretical calculations **-** Structures of inhibitors, substrate, reaction intermediate and transition state-like species (*vide infra*) were optimised using Gaussian03 at the B3LYP/6-31+G(d,p) level of theory[275]. The orientation of hydroxyl groups around the ring (and particularly the 6-hydroxymethyl moiety) were chosen to reflect those observed in the crystal structure of thiazoline **2c**

bound to the model GH84 enzyme *Bacteroides thetaiotaomicron O*-GlcNAcase[268]. Unconstrained structures were shown to possess no imaginary frequencies and therefore correspond to local minima. Molecular electrostatic potential surfaces were generated from a single point energy calculation carried out at the RHF/STO-3G level of theory using the associated B3LYP/6-31+G(d,p) optimized structures. The RHF/STO-3G level of theory has been shown to rapidly produce useful descriptions of the MEP surface[273]. A formatted checkpoint file was generated and the Gaussian03 utility cubegen was used to generate density and potential cube files with a grid density of 6 points/Bohr. Molekel 4.3[276] was used to visualize the molecular electrostatic potential surface projected onto the 0.04 electrons / $Å^3$ isodensity surface.

3.6.3 General procedure for the synthesis of methyl tri-O-acetyl-2-N-acyl-2-deoxy-β-D-glucopyranoses

Tri-*O*-acetyl-2-amino-2-deoxy-α-D-glucopyranosyl bromide hydrobromide (**5**) and methyl tri-*O*-acetyl-2-amino-2-deoxy-α-D-glucopyranoside (**9**) were prepared essentially as described by Billing and Nilsson[277] and used without further purification. Triethylamine (2.3 eq) was added to a solution of the hydrochloride salt of Methyl tri-*O*-acetyl-2-amino-2-deoxy-β-D-glucopyranoside (**9**) in one volume of dry dichloromethane (15 mL per gram of starting material), at which time the starting material dissolved. The reaction mixture was cooled to 0°C, and the appropriate acyl chloride (1.3 eq) was added *via* syringe. The resultant mixture was then stirred at room temperature for approximately 2 h. When the reaction mixture was judged complete by TLC analysis, 5 volumes of ethyl acetate were added. The resulting organic phase was washed successively with water, 1 M NaOH, and saturated sodium chloride. The organic phase was dried ($MgSO_4$), filtered, and concentrated to yield, in each case, a white crystalline solid. These materials were recrystallized using a mixture of ethyl acetate and hexanes to yield the desired *N*-acylated materials in yields ranging from 8 to 15% when calculated from the initial starting material, D-glucosamine hydrochloride.

Methyl tri-O-*acetyl-2*-N-*acetyl-2-deoxy*-β-D-*glucopyranoside (**10a**)*: ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.44 (d, 1 H, $J_{NH,H2}$ = 8.8 Hz, NH), 5.28 (dd, 1 H, $J_{H2,H3} \approx J_{H3,H4}$ = 9.5 Hz, H3), 5.08 (1 H, dd, $J_{H4,H5}$ = 9.8 Hz, H4), 4.58 (1 H, d, $J_{H1,H2}$ = 8.3 Hz, H1), 4.28 (1 H, dd, $J_{H5,H6}$ = 4.7, $J_{H6,H6'}$ = 12.3 Hz, , H6), 4.14 (1 H, dd, $J_{H5,H6'}$ = 2.6 Hz, H6'), 3.87 (1 H, ddd, H2) 3.71 (1 H, ddd, H5), 3.50 (3 H, s, OMe), 2.09 (3 H, s, OAc), 2.03 (3 H, s, OAc), 2.02 (3 H, s, OAc), 1.96 (3 H, s, H7).

Methyl tri-O-*acetyl-2-deoxy-2*-N-*propionyl*-β-D-*glucopyranoside* (**10b**): ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.43 (1 H, d, $J_{NH,H2}$ = 8.8 Hz, NH), 5.29 (1 H, dd, $J_{H2,H3} \approx J_{H3,H4}$ = 9.4 Hz, H3), 5.08 (1 H, dd, $J_{H4,H5}$ = 9.6 Hz, H4), 4.60 (1 H, d, $J_{H1,H2}$ = 8.3 Hz, H1), 4.28 (1 H, dd, $J_{H5,H6}$ = 4.7, $J_{H6,H6'}$ = 12.2 Hz, H6), 4.14 (1 H, dd, $J_{H5,H6'}$ = 2.4 Hz, H6'), 3.86 (1 H, ddd, H2) 3.70 (1 H, ddd, H5), 3.50 (3 H, s, OMe), 2.22-2.13 (2 H, m, H7), 2.09 (3 H, s, OAc), 2.03 (3 H, s, OAc), 2.02 (3 H, s, OAc), 1.12 (3 H, t, $J_{H8,H9}$ = 7.5 Hz, H8).

Methyl tri-O-*acetyl-2*-N-*butanoyl-2-deoxy*-β-D-*glucopyranoside* (**10c**): ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.45 (1 H, d, $J_{NH,H2}$ = 8.8 Hz, NH), 5.30 (1 H, dd, $J_{H2,H3} \approx J_{H3,H4}$ = 9.5 Hz, H3), 5.07 (1 H, dd, $J_{H4,H5}$ = 9.4 Hz, H4), 4.59 (1 H, d, $J_{H1,H2}$ = 8.3 Hz, H1), 4.28 (1 H, dd, $J_{H5,H6}$ = 4.7, $J_{H6,H6'}$ = 12.2 Hz, H6), 4.14 (1 H, dd, $J_{H5,H6'}$ = 2.4 Hz, H6'), 3.86 (1 H, ddd, H2) 3.70 (1 H, ddd, H5), 3.49 (3 H, s, OMe), 2.18-2.06 (2 H, m, H7), 2.09 (3 H, s, OAc), 2.02 (3 H, s, OAc), 2.01 (3 H, s, OAc), 1.66-1.57 (2 H, m, H8), 0.91 (3 H, t, $J_{H8,H9}$ = 7.3 Hz, H9).

Methyl tri-O-*acetyl-2-deoxy-2*-N-*pentanoyl*-β-D-*glucopyranoside* (**10d**): ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.43 (1 H, d, $J_{NH,H2}$ = 8.7 Hz, NH), 5.30 (1 H, dd, $J_{H2,H3} \approx J_{H3,H4}$ = 9.3 Hz, H3), 5.07 (1 H, dd, $J_{H4,H5}$ = 9.6 Hz, H4), 4.60 (1 H, d, $J_{H1,H2}$ = 8.3 Hz, H1), 4.28 (1 H, dd, $J_{H5,H6}$ = 4.7, $J_{H6,H6'}$ = 12.2 Hz, H6), 4.15 (1 H, dd, $J_{H5,H6'}$ = 2.4 Hz, H6'), 3.85 (1 H, ddd, H2) 3.70 (1 H, ddd, H5), 3.50 (3 H, s, OMe), 2.20-2.08 (2 H, m, H7), 2.09 (3 H, s, OAc), 2.03 (3 H, s, OAc), 2.02 (3 H, s, OAc),

1.60-1.54 (2 H, m, H8), 1.35-1.28 (2 H, m, H9), 0.88 (3 H, t, $J_{H9,H10}$ = 7.3 Hz, H10).

Methyl tri-O-*acetyl-2-deoxy-2*-N-*hexanoyl*-β-D-*glucopyranoside* (**10e**): ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.42 (1 H, d, $J_{NH,H2}$ = 9.6 Hz, NH), 5.30 (1 H, dd, $J_{H2,H3} \approx J_{H3,H4}$ = 9.6 Hz, H3), 5.07 (1 H, dd, $J_{H4,H5}$ = 9.6 Hz, H4), 4.60 (1 H, d, $J_{H1,H2}$ = 8.3 Hz, H1), 4.28 (1 H, dd, $J_{H5,H6}$ = 4.7, $J_{H6,H6'}$ = 12.2 Hz, H6), 4.14 (1 H, dd, $J_{H5,H6'}$ = 2.3 Hz, H6'), 3.85 (1 H, ddd, H2) 3.70 (1 H, ddd, H5), 3.49 (3 H, s, OMe), 2.08-2.18 (2 H, m, H7), 2.09 (3 H, s, OAc), 2.02 (3 H, s, OAc), 2.01 (3 H, s, OAc), 1.62-1.54 (2 H, m, H8), 1.34-1.22 (4 H, m, H9 and H10), 0.88 (3 H, t, $J_{H10,H11}$ = 7.2 Hz, H11).

Methyl tri-O-*acetyl-2-deoxy-2*-N-*isobutanoyl*-β-D-*glucopyranoside* (**10f**): ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.43 (1 H, d, $J_{NH,H2}$ = 8.7 Hz, NH), 5.32 (1 H, dd, $J_{H2,H3} \approx J_{H3,H4}$ = 9.4 Hz, H3), 5.08 (1 H, dd, $J_{H4,H5}$ = 9.6 Hz, H4), 4.62 (1 H, d, $J_{H1,H2}$ = 8.3 Hz, H1), 4.28 (1 H, dd, $J_{H5,H6}$ = 4.7, $J_{H6,H6'}$ = 12.2 Hz, H6), 4.14 (1 H, dd, $J_{H5,H6'}$ = 2.5 Hz, H6'), 3.84 (1 H, ddd, H2) 3.70 (1 H, ddd, H5), 3.49 (3 H, s, OMe), 2.35-2.26 (1 H, m, H7), 2.08 (3 H, s, OAc), 2.02 (3 H, s, OAc), 2.01 (3 H, s, OAc), 1.12 (3 H, d, $J_{H7,H8}$ = 6.9 Hz, H8), 1.12 (3 H, d, $J_{H7,H8'}$ = 6.9 Hz, H8').

Methyl tri-O-*acetyl-2-deoxy-2*-N-*isopentanoyl*-β-D-*glucopyranoside* (**10g**): ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.42 (1 H, d, $J_{NH,H2}$ = 8.8 Hz, NH), 5.30 (1 H, dd, $J_{H2,H3} \approx J_{H3,H4}$ = 9.4 Hz, H3), 5.08 (1 H, dd, $J_{H4,H5}$ = 9.5 Hz, H4), 4.58 (1 H, d, $J_{H1,H2}$ = 8.3 Hz, H1), 4.28 (1 H, dd, $J_{H5,H6}$ = 4.7, $J_{H6,H6'}$ = 12.3 Hz, H6), 4.15 (1 H, dd, $J_{H5,H6'}$ = 2.5 Hz, H6'), 3.85 (1 H, ddd, H2) 3.70 (1 H, ddd, H5), 3.49 (3 H, s, OMe), 2.08 (3 H, s, OAc), 2.02 (3 H, s, OAc), 2.01 (3 H, s, OAc), 2.00-2.10 (2 H, m, H7), 2.04 -1.96 (1 H, m, H8), 0.93 (3 H, d, $J_{H8,H9}$ = 2.5 Hz, H9), 0.91 (3 H, d, $J_{H8,H9'}$ = 2.5 Hz, H9').

General Procedure for the Synthesis of Methyl 2-N-*acyl-2-deoxy*-β-D-*glucopyranosides* (Ground state analogues) - A spatula tip of anhydrous sodium

methoxide was added to a solution of the appropriate protected methyl glycoside (**10a-g**) in dry methanol. The resultant solution was stirred until the reaction was judged complete by TLC analysis (typically between 2-4 h). A solution of glacial acetic acid in methanol (1:20) was added dropwise to the reaction mixture until the pH of the solution was found to be neutral. The solvent was then removed *in vacuo* to yield crystalline solids. The desired materials (**4a-g**) were purified by recrystallization using a mixture of ethyl acetate and methanol in yields ranging from 55 to 97%.

Methyl 2-N-*acetyl-2-deoxy*-β-D-*glucopyranoside* (**4a**) ^{1}H NMR (500 MHz, CD_3OD) δ 4.30 (1 H, d, $J_{H1,H2}$ = 8.4 Hz, H1), 3.88 (1 H, dd, $J_{H5,H6'}$ = 2.2, $J_{H6,H6'}$ = 11.9 Hz, H6'), 3.68 (1 H, dd, $J_{H5,H6}$ = 5.7 Hz, H6), 3.64 (1 H, dd, $J_{H2,H3}$ = 8.6, $J_{H3,H4}$ = 10.1 Hz, H3), 3.46 (3 H, s, OMe), 3.42 (1 H, dd, $J_{H4,H5'}$ = 8.5 Hz, H4), 3.32-3.28 (1 H, m, H2), 3.26 (1 H, ddd, H5) 1.97 (3 H, s, H7); ^{13}C NMR (125 MHz, CD_3OD) δ 174.2, 104.0, 78.4, 76.6, 72.5, 63.2, 57.6, 57.4, 23.3; Anal. Calcd for $C_9H_{17}NO_6$: C, 45.95; H, 7.28; N, 5.95. Found: C, 45.93; H, 7.54; N, 5.76.

Methyl 2-deoxy-2-N-*propionyl*-β-D-*glucopyranoside* (**4b**): ^{1}H NMR (500 MHz, CD_3OD) δ 4.31 (1 H, d, $J_{H1,H2}$ = 8.4 Hz, H1), 3.87 (1 H, dd, $J_{H6,H6'}$ = 11.9 Hz, H6'), 3.68 (1 H, dd, $J_{H5,H6}$ = 5.7 Hz, H6), 3.65 (1 H, dd, $J_{H2,H3}$ = 8.7, $J_{H3,H4}$ = 10.1 Hz, H3), 3.46 (3 H, s, OMe), 3.43 (1 H, dd, $J_{H4,H5'}$ = 8.4 Hz, H4), 3.32-3.29 (1 H, m, H2), 3.26 (1 H, m, H5) 2.23 (2 H, t, $J_{H7,H8}$ = 7.5 Hz, H7), 1.14 (3 H, t, H8); ^{13}C NMR (125 MHz, CD_3OD) δ 177.6, 103.7, 78.0, 76.3, 72.2, 62.8, 57.1, 57.1, 30.5, 10.5; Anal. Calcd for $C_{10}H_{19}NO_6$: C, 48.19; H, 7.68; N, 5.62. Found: C, 48.26; H, 7.89; N, 5.76.

Methyl 2-N-*butanoyl-2-deoxy*-β-D-*glucopyranoside* (**4c**): ^{1}H NMR (500 MHz, CD_3OD) δ 4.30 (1 H, d, $J_{H1,H2}$ = 8.4 Hz, H1), 3.88 (1 H, d, $J_{H5,H6'}$ = 2.1, $J_{H6,H6'}$ = 11.9 Hz, H6'), 3.68 (1 H, dd, $J_{H5,H6}$ = 5.7 Hz, H6), 3.65 (1 H, dd, $J_{H2,H3}$ = 8.4, $J_{H3,H4}$ = 10.2 Hz, H3), 3.45 (3 H, s, OMe), 3.43 (1 H, dd, $J_{H4,H5}$ = 8.4 Hz, H4), 3.32-3.28 (1 H, m, H2), 3.26 (1 H, ddd, H5) 2.20 (2 H, m, $J_{H7,H8}$ = 7.4 Hz, H7),

1.65 (2 H, ddt, $J_{H8,H9}$ = 7.4 Hz, H8) 0.96 (3 H, t, H9); ^{13}C NMR (125 MHz, CD_3OD) δ 177.1, 104.1, 78.4, 76.5, 72.6, 63.2, 57.4, 57.4, 39.8, 20.8, 14.2; Anal. Calcd for $C_{11}H_{21}NO_6$: C, 50.18; H, 8.04; N, 5.32. Found: C, 50.28; H, 8.22; N, 5.54.

Methyl 2-deoxy-2-N-*pentanoyl*-β-D-*glucopyranoside* (**4d**): ^{1}H NMR (500 MHz, CD_3OD) δ 4.31 (1 H, d, $J_{H1,H2}$ = 8.4 Hz, H1), 3.88 (1 H, d, $J_{H5,H6'}$ = 2.1, $J_{H6',H6}$ = 11.9 Hz, H6'), 3.69 (1 H, dd, $J_{H5,H6}$ = 5.7 Hz, H6), 3.65 (1 H, dd, $J_{H2,H3}$ = 8.5, $J_{H3,H4}$ = 10.4 Hz, H3), 3.45 (3 H, s, OMe), 3.43 (1 H, dd, $J_{H4,H5}$ = 8.4 Hz, H4), 3.32-3.28 (1 H, m, H2), 3.26 (1 H, ddd, H5) 2.22 (2 H, t, $J_{H7,H8}$ = 7.5 Hz, H7), 1.64-1.58 (2 H, m, H8) 1.38 (2 H, ddt, $J_{H8,H9}$ = 7.4 Hz, H9), 0.94 (3 H, t, $J_{H9,H10}$ = 7.4 Hz, H10); ^{13}C NMR (125 MHz, CD_3OD) δ 177.0, 103.9, 78.1, 76.3, 72.4, 63.0, 57.2, 57.2, 38.8, 33.3, 22.8, 14.2; Anal. Calcd for $C_{12}H_{23}NO_6$: C, 51.97; H, 8.36; N, 5.05. Found: C, 51.66; H, 8.18; N, 4.72.

Methyl 2-deoxy-2-N-*hexanoyl*-β-D-*glucopyranoside* (**4e**): ^{1}H NMR (500 MHz, CD_3OD) δ 4.30 (1 H, d, $J_{H1,H2}$ = 8.4 Hz, H1), 3.88 (1 H, d, $J_{H5,H6'}$ = 2.2, $J_{H6,H6'}$ = 11.9 Hz, H6'), 3.68 (1 H, dd, $J_{H5,H6}$ = 5.7 Hz, H6), 3.64 (1 H, dd, $J_{H2,H3}$ = 8.5, $J_{H3,H4}$ = 10.4 Hz, H3), 3.45 (3 H, s, OMe), 3.43 (1 H, dd, $J_{H4,H5}$ = 8.5 Hz, H4), 3.32-3.28 (1 H, m, H2), 3.25 (1 H, ddd, H5) 2.21 (2 H, m, $J_{H7,H8}$ = 7.5 Hz, H7), 1.62 (2 H, ddt, H8) 1.38-1.30 (4 H, m, H9 and H10), 0.92 (3 H, t, $J_{H10,H11}$ = 6.9 Hz, H11); ^{13}C NMR (125 MHz, CD_3OD) δ 176.8, 103.6, 77.9, 76.0, 72.1, 62.7, 56.9, 56.9, 37.3, 32.2, 26.6, 23.4, 14.2; Anal. Calcd for $C_{13}H_{25}NO_6 \bullet 0.5H_2O$: C, 51.99; H, 8.73; N, 4.66. Found: C, 52.26; H, 8.60; N, 4.86.

Methyl 2-deoxy-2-N-*isobutanoyl*-β-D-*glucopyranoside* (**4f**): ^{1}H NMR (500 MHz, CD_3OD) δ 4.32 (1 H, d, $J_{H1,H2}$ = 8.4 Hz, H1), 3.88 (1 H, d, $J_{H5,H6'}$ = 2.0, $J_{H6,H6'}$ = 11.9 Hz, H6'), 3.68 (1 H, dd, $J_{H5,H6}$ = 5.6 Hz, H6), 3.63 (1 H, dd, $J_{H2,H3}$ = 8.5, $J_{H3,H4}$ = 10.3 Hz, H3), 3.46 (3 H, s, OMe), 3.43 (1 H, dd, $J_{H4,H5}$ = 8.5 Hz, H4), 3.32-3.28 (1 H, m, H2), 3.28-3.22 (1 H, m, H5) 2.45 (1 H, m, $J_{H7,H8} \approx J_{H7,H8'}$ = 6.9 Hz, H7), 1.12 (3 H, d, H8), 1.12 (3 H, d, H8'); ^{13}C NMR (125 MHz, CD_3OD) δ

179.6, 102.6, 76.8, 74.9, 71.1, 61.6, 55.9, 55.8, 35.4, 18.9, 18.6; Anal. Calcd for $C_{11}H_{21}NO_6$: C, 50.18; H, 8.04; N, 5.32. Found: C, 50.39; H, 8.30; N, 5.50.

*Methyl 2-deoxy-2-*N-*isopentanoyl*-β-D-*glucopyranoside* (**4g**): ^{1}H NMR (500 MHz, CD_3OD) δ 4.29 (1 H, d, $J_{H1,H2}$ = 8.5 Hz, H1), 3.88 (1 H, d, $J_{H5,H6'}$ = 2.1, $J_{H6,H6'}$ = 11.9 Hz, H6'), 3.68 (1 H, dd, $J_{H5,H6}$ = 5.7 Hz, H6), 3.65 (1 H, dd, $J_{H2,H3}$ = 8.5, $J_{H3,H4}$ = 10.4 Hz, H3), 3.45 (3 H, s, OMe), 3.42 (1 H, dd, $J_{H4,H5}$ = 8.0 Hz, H4), 3.32-3.28 (1 H, m, H2), 3.25 (1 H, ddd, H5) 2.13-2.00 (3 H, m, H7 and H8), 1.14 (6 H, t, H9); ^{13}C NMR (125 MHz, CD_3OD) δ 175.0, 102.6, 76.8, 74.9, 71.1, 61.6, 55.8, 55.8, 45.7, 26.4, 21.6, 21.4; Anal. Calcd for $C_{12}H_{23}NO_6 \bullet 0.75H_2O$: C, 49.56; H, 8.49; N, 4.82. Found: C, 49.50; H, 8.45; N, 4.78.

3.6.4 General procedure for the synthesis of 4-Methylumbelliferyl tri-O-acetyl-2-N-acyl-2-deoxy-β-D-glucopyranosides

4 - Methylumbelliferyl tri-O-acetyl-2-amino-2-deoxy-β-D-glucopyranoside hydrochloride **6** was prepared essentially as described by Roeser and Legler[230] and was used without further purification. Triethylamine (2.3 eq) was added to a solution of the hydrochloride salt **6** in 1 volume of dry dichloromethane, at which time the starting material dissolved. The reaction mixture was cooled to 0°C, and the appropriate acyl chloride (1.3 eq) was added *via* syringe. The resultant mixture was stirred at room temperature for approximately 2 h. When the reaction mixture was judged complete by TLC analysis, 5 volumes of ethyl acetate were added. The resulting organic phase was washed successively with water, 1M NaOH, and saturated sodium chloride. The organic phase was dried ($MgSO_4$), filtered, and concentrated to yield, in each case, a white crystalline solid. The desired materials were purified by recrystallization using a mixture of ethyl acetate, hexanes, and methanol to yield the desired *N*-acylated materials.

4-Methylumbelliferyl tri-O-acetyl-2-deoxy-2-N-propionoyl -β-D-glucopyranoside (**10b**): ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.49 (d,1 H, $J_{H6Ar,H8Ar}$ = 8.6 Hz, H8Ar), 6.94 (d,1 H, $J_{H5Ar,H6Ar}$ = 2.4 Hz, H5Ar), 6.92 (dd, 1 H, H6Ar), 6.17 (d, 1 H, $J_{H3Ar,CH3}$ =

1.0 Hz, H3Ar), 5.71 (d, 1 H, $J_{NH,H2}$ = 8.7 Hz, NH), 5.45 (dd, 1 H, $J_{H2,H3}$ = $J_{H3,H4}$ = 10.3 Hz, H3), 5.39 (d, 1 H, $J_{H1,H2}$ = 8.2 Hz, H1), 5.14 (dd, 1H, $J_{H4,H5}$ = 9.7 Hz, H4), 4.28 (dd, 1 H, $J_{H5,H6}$ = 5.8, $J_{H6,H6'}$ = 12.3 Hz, H6), 4.20-4.14 (m, 1 H, H2), 4.20-4.14 (m, 1 H, H6'), 3.93 (ddd, 1 H, $J_{H5,H6'}$ = 2.3 Hz, H5), 2.39 (s, 3 H, CH3), 2.17 (q, 2 H, $J_{H7,H8}$ = 7.6 Hz, H7), 2.12 (s, 3 H, OAc), 2.07 (s, 3 H, OAc), 2.07 (s, 3 H, OAc), 0.96 (t, 3 H, H8).

4-Methylumbelliferyl tri-O-acetyl-2-N-butanoyl-2-deoxy -β-D-glucopyranoside (**10c**): ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.49 (d,1 H, $J_{H6Ar,H8Ar}$ = 8.6 Hz, H8Ar), 6.94 (d,1 H, $J_{H5Ar,H6Ar}$ = 2.3 Hz, H5Ar), 6.92 (dd,1 H, H6Ar), 6.17 (d,1 H, $J_{H3Ar,CH3}$ = .9 Hz, H3Ar), 5.70 (d, 1 H, $J_{NH,H2}$ = 8.7 Hz, NH), 5.44 (dd, 1 H, $J_{H2,H3}$ = $J_{H3,H4}$ = 10.3 Hz, H3), 5.37 (d, 1 H, $J_{H1,H2}$ = 8.2 Hz, H1), 5.14 (dd, 1H, $J_{H4,H5}$ = 9.7 Hz, H4), 4.28 (dd, 1 H, $J_{H5,H6}$ = 5.8, $J_{H6,H6'}$ = 12.3 Hz, H6), 4.22-4.15 (m, 1 H, H2), 4.22-4.15 (m, 1 H, H6'), 3.93 (ddd, 1 H, $J_{H5,H6'}$ = 2.3 Hz, H5), 2.39 (s, 3 H, CH3), 2.12 (t, 2 H, $J_{H7,H8}$ = 7.6 Hz, H7), 2.11 (s, 3 H, OAc), 2.06 (s, 3 H, OAc), 2.06 (s, 3 H, OAc), 1.64-1.55 (m, 2 H, H8), 0.88 (t, 3 H, 2.12, $J_{H8,H9}$ = 7.3 Hz, H9).

4-Methylumbelliferyl tri-O-acetyl-2-deoxy-2-N-pentanoyl -β-D-glucopyranoside (**10d**): ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.49 (d, 1 H, $J_{H6Ar,H8Ar}$ = 8.7 Hz, H8Ar), 6.94 (d, 1 H, $J_{H5Ar,H6Ar}$ = 2.4 Hz, H5Ar), 6.92 (dd, 1 H, H6Ar), 6.18 (d, 1 H, $J_{H3Ar,CH3}$ = 1.0 Hz, H3Ar), 5.67 (d, 1 H, $J_{NH,H2}$ = 8.6 Hz, NH), 5.44 (dd, 1 H, $J_{H2,H3}$ = $J_{H3,H4}$ = 10.2 Hz, H3), 5.38 (d, 1 H, $J_{H1,H2}$ = 8.2 Hz, H1), 5.14 (dd, 1H, $J_{H4,H5}$ = 9.8 Hz, H4), 4.29 (dd, 1 H, $J_{H5,H6}$ = 5.8, $J_{H6,H6'}$ = 12.3 Hz, H6), 4.20-4.15 (m, 1 H, H2), 4.20-4.15 (m, 1 H, H6'), 3.93 (ddd, 1 H, $J_{H5,H6'}$ = 2.4 Hz, H5), 2.40 (s, 3 H, CH3), 2.18-2.10 (m, 2 H, H7), 2.12 (s, 3 H, OAc), 2.06 (s, 3 H, OAc), 2.06 (s, 3 H, OAc), 1.55 (dddd, 2H, $J_{H8,H7}$ = $J_{H8,H9}$ = 7.5 Hz, H8), 1.27 (m, 2H, H9), 0.96 (dd, 3 H, $J_{H10,H9}$ = 7.4 Hz, H10).

4-Methylumbelliferyl tri-O-acetyl-2-deoxy-2-N-hexanoyl -β-D-glucopyranoside (**10e**): ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.49 (d, 1 H, $J_{H6Ar,H8Ar}$ = 8.7 Hz, H8Ar), 6.95 (d, 1 H, $J_{H5Ar,H6Ar}$ = 2.4 Hz, H5Ar), 6.93 (dd, 1 H, H6Ar), 6.17 (d, 1 H, $J_{H3Ar,CH3}$ =

1.1 Hz, H3Ar), 5.71 (d, 1 H, $J_{NH,H2}$ = 8.7 Hz, NH), 5.45 (dd, 1 H, $J_{H2,H3}$ = $J_{H3,H4}$ = 10.5 Hz, H3), 5.39 (d, 1 H, $J_{H1,H2}$ = 8.3 Hz, H1), 5.14 (dd, 1H, $J_{H4,H5}$ = 9.8 Hz, H4), 4.29 (dd, 1 H, $J_{H5,H6}$ = 5.8, $J_{H6,H6'}$ = 12.3 Hz, H6), 4.22-4.15 (m, 1 H, H2), 4.22-4.15 (m, 1 H, H6'), 3.94 (ddd, 1 H, $J_{H5,H6'}$ = 2.3 Hz, H5), 2.40 (s, 3 H, CH3), 2.16-2.10 (m, 2 H, H7), 2.12 (s, 3 H, OAc), 2.07 (s, 3 H, OAc), 2.07 (s, 3 H, OAc), 1.60-1.52 (m, 2H, H8), 1.30-1.18 (m, 2H, H9), 1.30-1.18 (m, 2H, H10), 0.83 (dd, 3 H, $J_{H10,H11}$ = 7.1 Hz, H11).

4-Methylumbelliferyl tri-O-acetyl-2-deoxy-2-N-isobutanoyl -β-D-glucopyranoside (**10f**): ^{1}H NMR (500 MHz, $CDCl_3$) δ7.49 (d, 1 H, $J_{H6Ar,H8Ar}$ = 8.7 Hz, H8Ar), 6.95 (d, 1 H, $J_{H5Ar,H6Ar}$ = 2.4 Hz, H5Ar), 6.92 (dd, 1 H, H6Ar), 6.17 (d, 1 H, $J_{H3Ar,CH3}$ = 1.0 Hz, H3Ar), 5.67 (d, 1 H, $J_{NH,H2}$ = 8.7 Hz, NH), 5.44 (dd, 1 H, $J_{H2,H3}$ = $J_{H3,H4}$ = 10.3 Hz, H3), 5.37 (d, 1 H, $J_{H1,H2}$ = 8.3 Hz, H1), 5.15 (dd, 1H, $J_{H4,H5}$ = 9.8 Hz, H4), 4.29 (dd, 1 H, $J_{H5,H6}$ = 5.9, $J_{H6,H6'}$ = 12.3 Hz, H6), 4.22-4.15 (m, 1 H, H2), 4.22-4.15 (m, 1 H, H6'), 3.93 (ddd, 1 H, $J_{H5,H6'}$ = 2.3 Hz, H5), 2.39 (s, 3 H, CH3), 2.31 (dt, 1 H, H7), 2.12 (s, 3 H, OAc), 2.06 (s, 3 H, OAc), 2.06 (s, 3 H, OAc), 0.96 (dd, 6 H, $J_{H7,H8}$ = 7.1 Hz, H8).

4-Methylumbelliferyl tri-O-acetyl-2-deoxy-2-N-isopentanoyl -β-D-glucopyranoside (**10g**): ^{1}H NMR (500 MHz, $CDCl_3$) δ7.49 (d, 1 H, $J_{H6Ar,H8Ar}$ = 8.7 Hz, H8Ar), 6.94 (d, 1 H, $J_{H5Ar,H6Ar}$ = 2.4 Hz, H5Ar), 6.91 (dd, 1 H, H6Ar), 6.17 (d, 1 H, $J_{H3Ar,CH3}$ = 1.0 Hz, H3Ar), 5.72 (d, 1 H, $J_{NH,H2}$ = 8.7 Hz, NH), 5.44 (dd, 1 H, $J_{H2,H3}$ = $J_{H3,H4}$ = 10.5 Hz, H3), 5.37 (d, 1 H, $J_{H1,H2}$ = 8.3 Hz, H1), 5.15 (dd, 1H, $J_{H4,H5}$ = 9.8 Hz, H4), 4.29 (dd, 1 H, $J_{H5,H6}$ = 5.8, $J_{H6,H6'}$ = 12.3 Hz, H6), 4.24-4.16 (m, 1 H, H2), 4.24-4.16 (m, 1 H, H6'), 3.94 (ddd, 1 H, $J_{H5,H6'}$ = 2.3 Hz, H5), 2.39 (s, 3 H, CH3), 2.06-1.98 (m, 2 H, H7), 2.12 (s, 3 H, OAc), 2.06 (s, 3 H, OAc), 2.06 (s, 3 H, OAc), 2.06-1.98 (m, 1 H, H8), 0.96 (dd, 6 H, $J_{H8,H9}$ = 6.3 Hz, H9).

3.6.5 General procedure for the synthesis of 4-Methylumbelliferyl 2-N-acyl-2-deoxy-β-D-glucopyranosides

A spatula tip of anhydrous sodium methoxide was added to a solution of the appropriate protected methylumbelliferyl-glycoside (**7a-g**) in dry methanol. The resultant solution was stirred until the reaction was judged complete by TLC analysis (typically between 2-4 h). A solution of glacial acetic acid in methanol (1:20) was added dropwise to the reaction mixture until the pH of the solution was found to be neutral. The solvent was then removed *in vacuo* to yield crystalline solids, and the desired materials (**3a-g**) were purified by recrystallization using a mixture of ethyl acetate and methanol.

4-Methylumbelliferyl 2-deoxy-2-N-propanoyl-β-D-glucopyranoside (**3b**): ^{1}H NMR (500 MHz, d_6-Me_2SO) δ 7.72 (d, 1 H, $J_{NH,H2}$ = 9.1 Hz, NH), 7.68 (d, 1 H, $J_{H6Ar,H8Ar}$ = 8.8 Hz, H8Ar), 6.97 (d, 1 H, $J_{H5Ar,H6Ar}$ = 2.4 Hz, H5Ar), 6.91 (dd, 1 H, H6Ar), 6.24 (d, 1 H, $J_{H3Ar,CH3}$ = 1.1 Hz, H3Ar), 5.15 (d, 1 H, OH), 5.09 (d, 1 H, OH), 5.08 (d, 1 H, $J_{H1,H2}$ = 8.5 Hz, H1), 4.65 (dd, 1 H, $J_{OH6,H6}$ = $J_{OH6,H6'}$ = 5.4 Hz, OH6), 3.74-3.68 (m, 1 H, H6), 3.74-3.68 (m, 1 H, H2), 3.50-3.40 (m, 1 H, H5), 3.44-3.36 (m, 1 H, H6') 3.42-3.36 (m, 1 H, H3), 3.18 (1H, ddd, $J_{H3,H4}$ = $J_{H4,H5}$ = 9.3 Hz, H4), 2.38 (3 H, s, CH3), 2.00-2.08 (m, 2 H, H7), 0.96 (dd, 3 H, $J_{H7,H8}$ = 7.6 Hz, H8). ^{13}C NMR (125 MHz, d_6 Me_2SO_4) δ 173.8, 160.9, 160.7, 155.1, 154.0, 127.2, 114.9, 114.2, 112.5, 103.8, 99.7, 78.1, 74.5, 70.9, 61.3, 55.8, 29.6, 18.8, 10.8; Anal. Calcd for $C_{19}H_{23}NO_8$: C, 58.01; H, 5.89; N, 3.56. Found: C, 57.88; H, 5.89; N, 4.77.

4-Methylumbelliferyl 2-N-butanoyl-2-deoxy-β-D-glucopyranoside (**3c**): ^{1}H NMR (500 MHz, d_6-Me_2SO) δ 7.75 (d, 1 H, $J_{NH,H2}$ = 9.1 Hz, NH), 7.68 (d, 1 H, $J_{H6Ar,H8Ar}$ = 8.8 Hz, H8Ar), 6.96 (d, 1 H, $J_{H5Ar,H6Ar}$ = 2.4 Hz, H5Ar), 6.90 (dd, 1 H, H6Ar), 6.24 (d, 1 H, $J_{H3Ar,CH3}$ = 1.0 Hz, H3Ar), 5.15 (d, 1 H, OH), 5.09 (d, 1 H, $J_{H1,H2}$ = 8.3 Hz, H1), 5.08 (d, 1 H, OH), 4.65 (dd, 1 H, $J_{OH6,H6}$ = $J_{OH6,H6'}$ = 5.4 Hz, OH6), 3.68-3.76 (m, 1 H, H6), 3.68-3.74 (m, 1 H, H2), 3.50-3.40 (m, 1 H, H5), 3.44-3.36 (m, 1 H, H6') 3.42-3.36 (m, 1 H, H3), 3.19 (ddd, 1H, $J_{H3,H4}$ = $J_{H4,H5}$ = 9.2 Hz, H4),

2.38 (s, 3 H, CH3), 2.08-1.97 (m, 2 H, H7), 1.52-1.42 (m, 2 H, H8), 0.79 (dd, 3 H, $J_{H8,H9}$ = 7.4 Hz, H9). ^{13}C NMR (125 MHz, *d*$_6$ Me_2SO_4) δ 172.8, 160.8, 160.7, 155.1, 154.0, 127.2, 114.9, 114.1, 112.5, 103.8, 99.6, 78.1, 74.5, 71.0, 61.3, 55.8, 38.5, 19.4, 18.8, 14.1; Anal. Calcd for $C_{20}H_{25}NO_8$: C, 58.96; H, 6.18; N, 3.44. Found: C, 58.80; H, 6.25; N, 3.46.

4-Methylumbelliferyl 2-deoxy-2-N-pentanoyl -β-D-glucopyranoside (**3d**): ^{1}H NMR (500 MHz, *d*$_6$-Me_2SO) δ 7.75 (d, 1 H, $J_{NH,H2}$ = 9.0 Hz, NH), 7.69 (d, 1 H, $J_{H6Ar,H8Ar}$ = 8.8 Hz, H8Ar), 6.96 (d, 1 H, $J_{H5Ar,H6Ar}$ = 2.4 Hz, H5Ar), 6.90 (dd, 1 H, H6Ar), 6.24 (d, 1 H, $J_{H3Ar,CH3}$ = 1.1 Hz, H3Ar), 5.14 (d, 1 H, OH), 5.10 (d, 1 H, $J_{H1,H2}$ = 8.5 Hz, H1), 5.07 (d, 1 H, OH), 4.65 (dd, 1 H, $J_{OH6,H6}$ = $J_{OH6,H6'}$ = 5.4 Hz, OH6), 3.76-3.68 (m, 1 H, H6), 3.74-3.68 (m, 1 H, H2), 3.50-3.40 (m, 1 H, H5), 3.44-3.36 (m, 1 H, H6') 3.42-3.36 (m, 1 H, H3), 3.18 (ddd, 1H, $J_{H3,H4}$ = $J_{H4,H5}$ = 9.3 Hz, H4), 2.38 (d, 3 H, CH3), 2.08-1.98 (m, 2 H, H7), 1.48-1.36 (m, 2 H, H8), 1.26-1.12 (m, 2 H, H9), 0.77 (dd, 3 H, $J_{H9,H10}$ = 7.6 Hz, H10). ^{13}C NMR (125 MHz, *d*$_6$ Me_2SO_4) δ 173.0, 160.8, 160.7, 155.1, 154.0, 127.2, 114.9, 114.1, 112.5, 103.8, 99.6, 78.1, 74.5, 71.0, 61.3, 55.8, 36.3, 28.2, 22.3, 18.8, 14.4; Anal. Calcd for $C_{21}H_{27}NO_8$: C, 59.85; H, 6.46; 3.32. Found: C, 60.04; H, 6.57; N, 3.31

4-Methylumbelliferyl 2-deoxy-2-N-hexanoyl-β-D-glucopyranoside (**3e**): ^{1}H NMR (500 MHz, *d*$_6$-Me_2SO) δ 7.75 (d, 1 H, $J_{NH,H2}$ = 9.0 Hz, NH), 7.69 (d, 1 H, $J_{H6Ar,H8Ar}$ = 8.8 Hz, H8Ar), 6.97 (d, 1 H, $J_{H5Ar,H6Ar}$ = 2.4 Hz, H5Ar), 6.92 (dd, 1 H, H6Ar), 6.25 (d, 1 H, $J_{H3Ar,CH3}$ = 1.1 Hz, H3Ar), 5.15 (d, 1 H, OH), 5.11 (d, 1 H, $J_{H1,H2}$ = 8.4 Hz, H1), 5.08 (d, 1 H, OH), 4.65 (dd, 1 H, $J_{OH6,H6}$ = $J_{OH6,H6'}$ = 5.8 Hz, OH6), 3.74-3.68 (m, 1 H, H6), 3.74-3.68 (m, 1 H, H2), 3.50-3.40 (m, 1 H, H5), 3.46-3.36 (m, 1 H, H6') 3.46-3.36 (m, 1 H, H3), 3.19 (ddd, 1H, $J_{H3,H4}$ = $J_{H4,H5}$ = 9.3 Hz, H4), 2.39 (d, 3 H, CH3), 2.08-1.98 (m, 2 H, H7), 1.50-1.40 (m, 2 H, H8), 1.22-1.10 (m, 2 H, H9), 1.22-1.10 (m, 2 H, H10), 0.74 (dd, 3 H, $J_{H10,H11}$ = 6.9 Hz, H11). ^{13}C NMR (125 MHz, *d*$_6$ Me_2SO_4) δ 175.2, 163.2, 162.9, 157.2, 156.1, 129.4, 117.0, 116.2, 114.6, 106.0, 101.7, 80.3, 76.7, 73.1, 63.5, 58.0, 38.7, 33.5, 27.9, 24.8,

21.0, 16.6; Anal. Calcd for $C_{22}H_{29}NO_8$: C, 60.68; H, 6.71; 3.22. Found: C, 60.48; H, 6.81; N, 3.26

4-Methylumbelliferyl 2-deoxy-2-N-isobutanoyl-β-D-glucopyranoside (**3f**): ^{1}H NMR (500 MHz, d_6-Me_2SO) δ 7.68 (d, 1 H, $J_{H6Ar,H8Ar}$ = 8.8 Hz, H8Ar), 7.65 (d, 1 H, $J_{NH,H2}$ = 9.1 Hz, NH), 6.96 (d, 1 H, $J_{H5Ar,H6Ar}$ = 2.4 Hz, H5Ar), 6.90 (dd, 1 H, H6Ar), 6.24 (d, 1 H, $J_{H3Ar,CH3}$ = 1.1 Hz, H3Ar), 5.15 (d, 1 H, OH), 5.08 (d, 1 H, OH), 5.06 (d, 1 H, $J_{H1,H2}$ = 8.3 Hz, H1), 4.66 (dd, 1 H, $J_{OH6,H6}$ = $J_{OH6,H6'}$ = 5.4 Hz, OH6), 3.76-3.68 (m, 1 H, H6), 3.74-3.68 (m, 1 H, H2), 3.50-3.40 (m, 1 H, H5), 3.44-3.36 (m, 1 H, H6') 3.42-3.36 (m, 1 H, H3), 3.18 (ddd, 1H, $J_{H3,H4}$ = $J_{H4,H5}$ = 9.3 Hz, H4), 2.38 (d, 3 H, CH3), 2.32-2.24 (m, 1 H, H7), 0.97 (d, 3 H, $J_{H7,H8}$ = 6.8 Hz, H8), 0.94 (d, 3 H, $J_{H7,H8'}$ = 6.9 Hz, H8'). ^{13}C NMR (125 MHz, d_6 Me_2SO_4) δ 177.0, 161.0, 160.7, 155.1, 154.0, 127.3, 114.9, 114.1, 112.5, 103.7, 100.0, 78.1, 74.4, 71.0, 61.4, 55.6, 35.1, 20.6, 19.9, 18.8; Anal. Calcd for $C_{20}H_{25}NO_8$: C, 58.96; H, 6.18; N, 3.44. Found: C, 58.84; H, 6.35; N, 3.35

4-Methylumbelliferyl 2-deoxy-2-N-isopentanoyl-β-D-glucopyranoside (**3g**): ^{1}H NMR (500 MHz, d_6-Me_2SO) δ 7.76 (d, 1 H, $J_{NH,H2}$ = 9.0 Hz, NH),7.68 (d, 1 H, $J_{H6Ar,H8Ar}$ = 8.8 Hz, H8Ar), 6.95 (d, 1 H, $J_{H5Ar,H6Ar}$ = 2.3 Hz, H5Ar), 6.90 (dd, 1 H, H6Ar), 6.24 (s, 1 H, H3Ar), 5.13 (d, 1 H, OH), 5.10 (d, 1 H, $J_{H1,H2}$ = 8.3 Hz, H1), 5.06 (d, 1 H, OH), 4.65 (dd, 1 H, $J_{OH6,H6}$ = $J_{OH6,H6'}$ = 5.4 Hz, OH6), 3.76-3.68 (m, 1 H, H6), 3.74-3.68 (m, 1 H, H2), 3.50-3.40 (m, 1 H, H5), 3.44-3.36 (m, 1 H, H6') 3.42-3.36 (m, 1 H, H3), 3.18 (m, 1H, H4), 2.38 (s, 3 H, CH3), 1.98-1.88 (m, 2 H, H7), 1.98-1.88 (m, 1 H, H8), 0.83 (d, 3 H, $J_{H8,H9}$ = 5.1 Hz, H9), 0.80 (d, 3 H, $J_{H8,H9'}$ = 5.1 Hz, H9'). ^{13}C NMR (125 MHz, d_6 Me_2SO_4) δ 174.5, 162.9, 157.2, 156.2, 129.4, 117.0, 116.2, 114.6, 105.9, 101.6, 80.2, 76.7, 73.2, 63.5, 57.8, 48.2, 28.5, 25.2, 25.0, 21.0; Anal. Calcd for $C_{21}H_{27}NO_8$: C, 59.85; H, 6.46; 3.32. Found: C, 60.02; H, 6.48; N, 3.40

3.7 Acknowledgements

We thank the Western Canada research grid (WestGrid) for a generous allocation of resources. DJV thanks the Natural Sciences and Engineering Research Council of Canada (NSERC) for support of this research. DJV is a Michael Smith Foundation for Health Research (MSFHR) Scholar and is supported as a Tier II Canada Research Chair in Chemical Glycobiology. MSM is supported as a Canada Graduate Scholar (NSERC) and by a graduate student fellowship from the MSFHR. GJD thanks the Biotechnology and Biological Sciences Research Council and the Royal Society for funding. IRG thanks the Royal Society (UK) for a post-doctoral fellowship to Canada.

4: Mammalian Notch is modified by D-xyl-α1-3-D-xyl-α1-3-D-glc-β1-*O*-ser: Implementation of a method to study *O*-glucosylation

The manuscript below is reprinted from:

Whitworth, G. E.; Zandberg, W. F.; Clark, T.; Vocadlo, D. J., Mammalian Notch is modified by D-Xyl-alpha 1-3-D-Xyl-alpha 1-3-D-Glc-beta 1-O-Ser: Implementation of a method to study O-glucosylation. *Glycobiology* **2010,** 20, (3), 287-299.

Updates in the field relating to this Chapter can be found in appendix 2. The designated numbers given to synthetic or target compounds in this Chapter relate only to this Chapter.

4.1 Contributions

I synthesized a series of mono-, di-, and tri-saccharides for use as standards to characterize the *O*-glycans modifying Notch. The details regarding the synthesis and characterization of most of these compounds can be found in the following Chapter. I also cloned a series of epidermal growth factor (EGF) constructs encompassing the entire EGF domain of mammalian Notch. I also expressed and purified a recombinant α-xylosidase used to characterize *O*-glycans released from our recombinant Notch fragment. Finally, I wrote the

majority of the first draft of the manuscript. Thomas Clark performed the mass spectrometry analysis. Wesley Zandberg performed the cell biology and analytical capillary electrophoresis experiments described. In addition Wesley carried out the MS sample preparation and performed data analysis.

4.2 Abstract

Notch is a key cell surface protein receptor that is a vital component of intercellular signaling occurring during development. The *O*-glucosylation of the extracellular Notch epidermal growth factor-like (EGF) repeats has recently been found to play an important role in the proper functioning of Notch in Drosophila. Previous efforts to identify the fine structure of the *O*-glucose containing glycan of mammalian Notch have been hindered by limitations associated with approaches used to date. Here we report the development of an alternative strategy that can be used to study this modification from a range of different tissues. To implement this approach we have generated standards of the D-Xyl-α1-3-D-Xyl-α1-3-D-Glc trisaccharide, isomers of this structure, as well as the D-Xyl-α1-3-D-Glc disaccharide found previously on secreted EGF-containing proteins of the blood coagulation cascade. Following derivatization with 8-aminopyrene-1,3,6-trisulfonate (APTS), we use these standards in capillary electrophoretic analyses of *O*-glycans released from Notch1 EGF repeats in conjunction with *exo*-α-xylosidase digestion. These studies collectively reveal that the *O*-glucose containing glycan decorating mammalian Notch is the D-Xyl-α1-3-D-Xyl-α1-3-D-Glc trisaccharide; an assignment in accord with previous predictions. Given the demonstrated importance of this modification in the function of Notch in Drosophila we expect that the unambiguous identification of this glycan decorating mammalian Notch1 should aid studies into the functional role of *O*-glycosylation of mammalian Notch isoforms. Wider application of this approach should facilitate identification of other EGF-containing proteins bearing this *O*-glycan and aid their study.

Introduction

O-linked glucose is a non-canonical form of protein glycosylation that is installed on certain proteins as they transit through the secretory pathway. This modification was first identified as decorating the epidermal growth factor (EGF)-like domains of bovine blood coagulation factors VII and IX[145]. The *O*-linked glucose residue was found attached to the hydroxyl group of serine residues via a *β*-glycosidic linkage. Further structural analysis of the *O*-linked glucose modification found on bovine coagulation factor IX revealed it comprises primarily a trisaccharide (**1**) having the structure D-Xyl-*α*1-3-D-Xyl-*α*1-3-D-Glc (XXG, Figure 4.1A)[146]. On the basis of several known *O*-glucose modification sites, a consensus sequence governing which serine residues become modified has been proposed; NH_2-Cys^1-Xxx-Ser-Xxx-Pro-Cys^2-COOH (Figure 4.1B) where Cys^1 and Cys^2 represent the first and second conserved cysteine residues of the EGF-like repeats and Xxx represents any amino acid[147]. These EGF-like repeats are compact structures that contain six conserved cysteine residues forming three disulfide bonds. It is believed that serine residues of these EGF-containing secreted proteins within the loop region having the proposed consensus sequences are glucosylated and then subsequently elongated to generate the XXG trisaccharide (Figure 4.1A)[147].

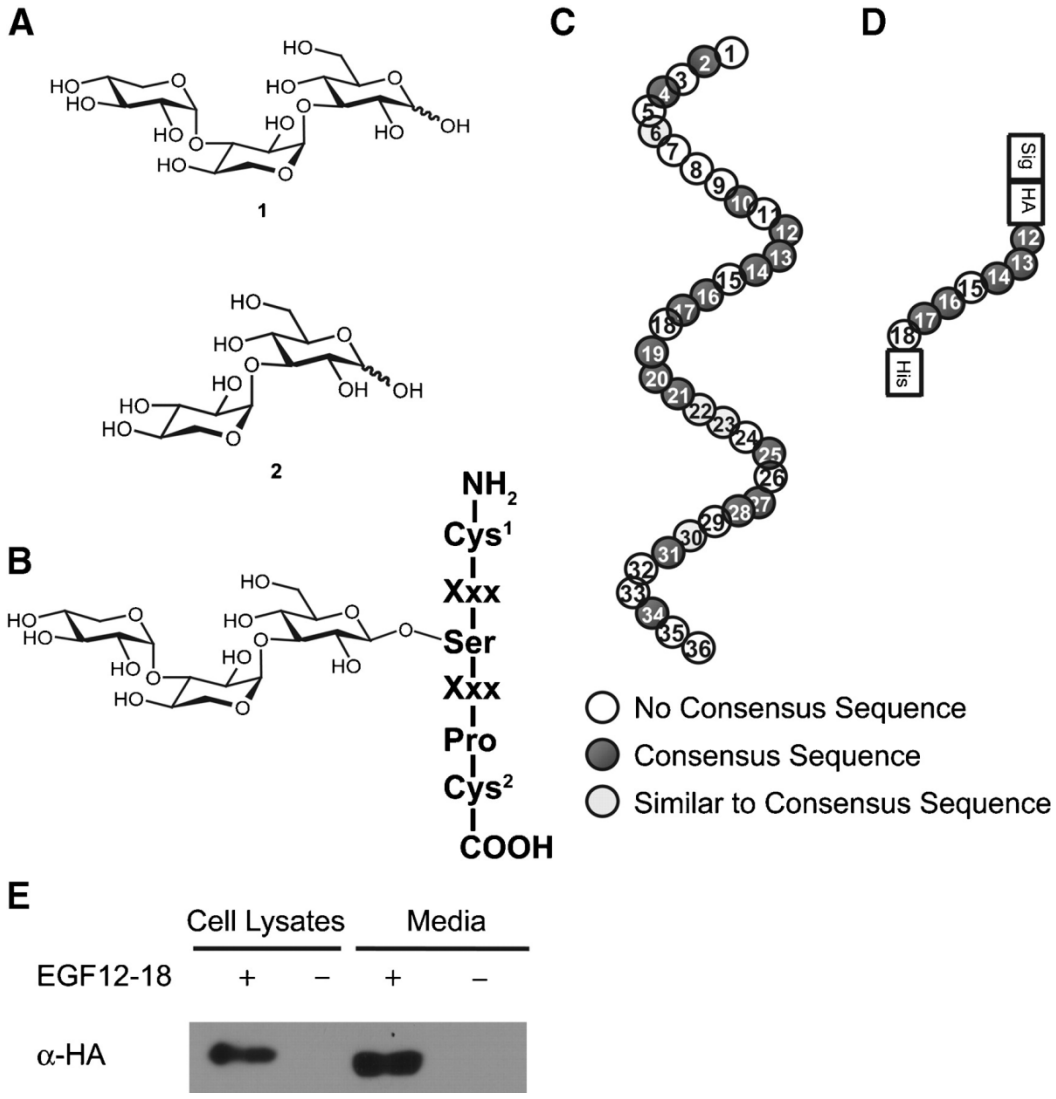


Figure 4.1: The consensus sequence directing *O*-glucosylation of Notch and protein containing the EGF repeats studied here
(**A**) Structure of the putative XXG trisaccharide (1) decorating Notch1 and the putative XG disaccharide (2). (**B**) The proposed consensus sequence directing *O*-glucosylation of EGF repeats modified by the XXG trisaccharide showing the stereochemistry of the linkage to the serine residue. (**C**) The 36 tandem EGF repeats that make up the majority of the N^{ECD} of mouse Notch1 with the *O*-glucose modification sites annotated (where "Similar to Consensus Sequence" refers to a sequence in which the serine normally modified within the consensus sequence is found shifted one residue toward either the *N*- or C-terminus). (**D**) Structure of the protein construct EGF12–18 used in these studies contains EGF repeats 12–18 bracketed by an *N*-terminal signal sequence, an HA epitope, and a C-terminal His_6 tag. (**E**) Western blot analysis of the expression level of EGF12–18 from conditioned media and cell lysates from CHO cell cultures transiently transfected with the either empty pax142 or pax142EGF12–18.

More recently, considerable effort has focused on studying the glycosylation of the extracellular EGF-like repeat domain of the cell surface

protein receptor Notch. This effort has been stimulated by the realization that glycosylation of Notch regulates its function in the Notch signaling pathway;[12,13,173,174] a cell-cell signaling pathway that uses local cellular interactions to control cell fate in metazoans[149]. Abnormal Notch signaling has been linked to developmental disorders of the kidney, liver, heart, eye, and skeleton[150,151], and has also been implicated in human diseases such as cancer[152]. Stimulated in part by these important biological roles of Notch, recent investigations have uncovered the role played by its *O*-fucosylation in tuning Notch processing and downstream signaling events in response to ligand binding[278]. Recent work by the Haltiwanger group has shown that mammalian Notch is also modified with *O*-glucose[175]. The functional role and fine structure of the complete *O*-glucose containing glycan, however, is poorly delineated when compared to our level of understanding of the *O*-fucose containing glycans. Given the functional significance of Notch *O*-fucosylation it seems possible that the *O*-glucosylation of Notch may well play similar, or related, roles and accordingly this modification merits further investigation.

Within mammals there are four different homologues of Notch (Notch1-4) while there is only one gene encoding Notch in *Drosophila melanogaster*.[174] The Notch extracellular domain (N^{ECD}), of the four mammalian homologues, are primarily composed of between 29 and 36 tandem EGF-like repeats that mediate the cell-cell interactions involving Notch[149]. Of the 36 tandem EGF repeats that comprise the N^{ECD} of mouse Notch1, sixteen of these have the putative consensus sequence that appears to direct *O*-glucosylation (Figure 4.1C). Unlike Notch *O*-fucosylation, which is known to be primarily a monosaccharide or linear tetrasaccharide[148,279], the structure of the glycan containing the non-canonical *O*-glucose linkage decorating mammalian Notch remains ill defined. The recent identification of the *O*-glucosyltransferase in *Drosophila*, known as Rumi, and the realization that deletion of Rumi compromises proper Notch functioning in *Drosophila*[175] has provided considerable stimulus for improving our understanding of the roles of *O*-glucosylation of mammalian Notch, since this protein receptor plays key roles in health and disease. Given the presence of the

proposed *O*-glucose consensus sequence within mammalian Notch and preliminary studies that have suggested that the *O*-glucose moiety is elaborated with additional sugar residues in eukaryotic cells, Haltiwanger and coworkers have proposed the structure of this glycan is the XXG trisaccharide previously identified on secreted EGF-containing blood coagulation factors. At that time, however, these researchers were unable to establish the structure of the glycan or clarify whether xylose residues were a component. Very recent studies have now suggested this glycan is a linear structure composed of two pentoses and one hexose at the reducing terminus[280]. Given, however, the tendency for structural variability of glycan structures, such as the different forms of *O*-fucosylation for example[169], the unambiguous identification of the constitution and linkages between the saccharide units in this glycan remains a topic of interest[148]. Resolution of this question will facilitate development of antibodies against these structures and functional studies of the modification in eukaryotic systems.

Existing methods of studying these non-canonical forms of *O*-glycosylation in eukaryotic cells involve metabolic labeling of mutant cell lines with radiolabeled saccharides[126,148,281]. These mutant cell lines, developed by Stanley and coworkers, have genetic deficiencies that prevent incorporation of the labeled saccharides into various glycan structures; shunting the majority of label toward incorporation into the non-canonical forms of *O*-glycosylation. Mass spectrometry has been shown to provide another complementary method for studying the composition of *O*-glycans found on Notch[280], however, this technology is limited in its ability to clarify the linkages between saccharide residues within a given glycan[282]. While these methods have provided many insights and are ideal for addressing certain questions, such as identifying proteins bearing the non-canonical forms of *O*-glycosylation, the on-going development of complementary analytical methods is required to study these modifications in a wider range of biological tissues, as well as to determine the fine structure of the *O*-glucose containing glycan on mammalian Notch.

To elucidate the fine structure of this glycan decorating mammalian Notch1, and simultaneously develop an analytical method that could be used with various tissues, we have generated a set of standards to be used in conjunction with high performance capillary electrophoresis (CE). Here we demonstrate this method and elucidate the structure of the *O*-glucose containing glycan on mammalian Notch1 as the D-Xyl-α1-3-D-Xyl-α1-3-D-Glc trisaccharide.

4.3 Results

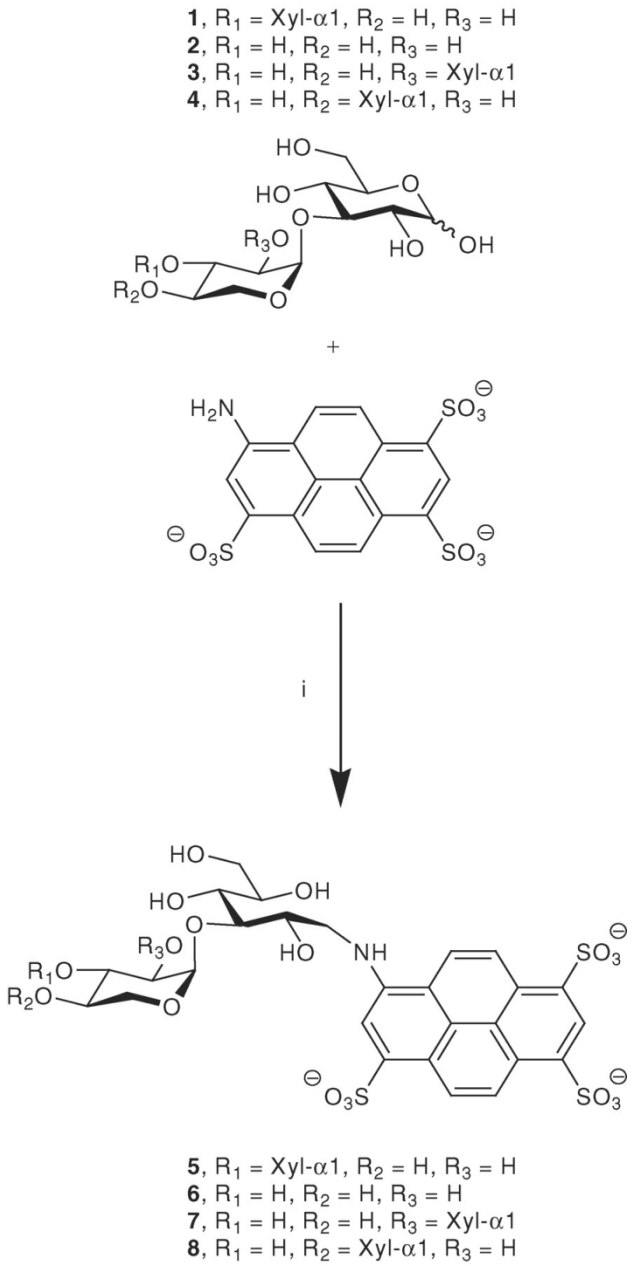
4.3.1 Expression of mammalian Notch1 construct EGF12-18

The mouse N^{ECD} contains 36 consecutive EGF repeats (Figure 4.1C), sixteen of which contain the proposed *O*-glucose consensus sequence NH_2-Cys^1-Xxx-Ser-Xxx-Pro-Cys^2-COOH. To facilitate our efforts to recombinantly express a portion of the N^{ECD} containing the proposed consensus sequence, we generated a series of constructs comprising portions of the EGF-like domains that could be used to recombinantly express the various EGF-like repeats of Notch within CHO K1 cells. These N^{ECD} fragments, encoding various sections of the EGF domain were prepared from mouse Notch1 cDNA using primers encoding a leader sequence and an *N*-terminal HA epitope tag, as well as a C-terminal His_6 tag, were ligated into the mammalian expression vector pAX142[283] to generate the desired constructs. pAXEGF constructs and the empty pAX142 vector were transiently transfected into CHO cells and the recombinant protein was purified from both cell lysates and media using metal chelate affinity column chromatography. Of these constructs, we find that after some optimization, the pAXEGF12-18 construct (encoding EGF-repeats 12-18 of Notch1) was the most effective for obtaining recombinant Notch1 EGF repeats in CHO-K1 cells. Through analysis of the purified samples by Western blot using an *α*-HA monoclonal antibody, we find this construct produces EGF12-18 (Figure 4.1D) that is primarily secreted into the media with a lesser amount being retained within cells (Figure 4.1E). Unfortunately due to the limited amount of protein expressed and purified, staining of the SDS-PAGE gels by Coomasie blue did

not reveal any detectable signal, nor did Bradford assays. The expression and secretion of the EGF12-18 Notch1 fragment suggests it properly transits through the secretory pathway and offered us reassurance that it is being exposed to the biosynthetic glycosyl transferases within the secretory pathway. Also we consistently observed that EGF12-18 with the conditioned media had a higher electrophoretic mobility, perhaps due to differences in glycosylation.

4.3.2 Preparation of the APTS-labeled glycan standards

The chemical synthesis of the XXG trisaccharide (**1**), D-Xyl-*α*1-2-D-Xyl-*α*1-3-D-Glc (X1-2XG, **3**), D-Xyl-*α*1-4-D-Xyl-*α*1-3-D-Glc (X1-4XG, **4**), and the D-Xyl-*α*1-3-D-Glc (XG, **2**) disaccharide for use as standards with CE (Scheme 4.1) will be published elsewhere. The tri-(**1, 3, 4**) and di-(**2**) saccharides were reductively aminated with APTS by adaptation of established general procedures (Evangelista *et al.* 1996, Scheme 4.1) to yield the desired fluorescent APTS-labeled standards **5**-**8**. When necessary, excess APTS and other trace impurities were removed from the labeled standards by resolving them using fluorophore-assisted carbohydrate electrophoresis (FACE), excising the bands of interest, and eluting the labeled standard from the gel into water.



Scheme 4.1: Di- and trisaccharide synthetic standards and subsequent APTS-labeling to generate the desired standards for CE

Synthetically prepared standards; D-xylose-*α*1-3-D-glucopyranose (**2**, XG), D-xylose-*α*1-3-D-xylose-*α*1-3-D-glucopyranose (**1**, XXG), D-xylose-*α*1-2-D-xylose-*α*1-3-D-glucopyranose (**3**, X1-2XG), and D-xylose-*α*1-4-D-xylose-*α*1-3-D-glucopyranose (**4**, X1-4XG). Reagents for the synthesis of **5** - **8** are (i) 0.2 M APTS, 7.5% AcOH, 1.0 M $NaBH_3CN$, THF, 37 ^{o}C.

4.3.3 UDP-D-xylose: α-D-xyloside α1-3-xylosyltransferase assay

To ensure that our EGF12-18 Notch 1 construct was transiently transfected into a cell line that possessed the appropriate xylosyl-transferases for XXG biosynthesis, we modified a procedure from Minamida *et al.*[284] to evaluate xylosyl-transferase activity in CHO microsomal fractions. CHO cell microsomes were mixed with the FACE purified XG-APTS standard. This mixture was then divided into three aliquots and supplemented with XXG-APTS, UDP-xylose, or a combination of both XXG-APTS and UDP-xylose (Figure 4.2A). As we expected, the microsomal fraction containing XG-APTS mixture spiked with XXG-APTS showed no decrease in the XG-APTS after incubation nor any increase in the peak corresponding to XXG-APTS when UDP-xylose was absent. When UDP-xylose was added to the microsomal fraction containing XG-APTS or a microsomal fraction containing XG-APTS and XXG-APTS, however, the XG-APTS standard was almost completely converted to XXG-APTS after 20 hours of incubation at 37 ºC (Figure 4.2Ai and ii). We also observed that the XXG-APTS standard does not get elongated in the presence of UDP-xylose, indicating that the CHO microsomal fraction contains no glycosyl transferases that can further elaborate the trisaccharide with additional xylose residues.

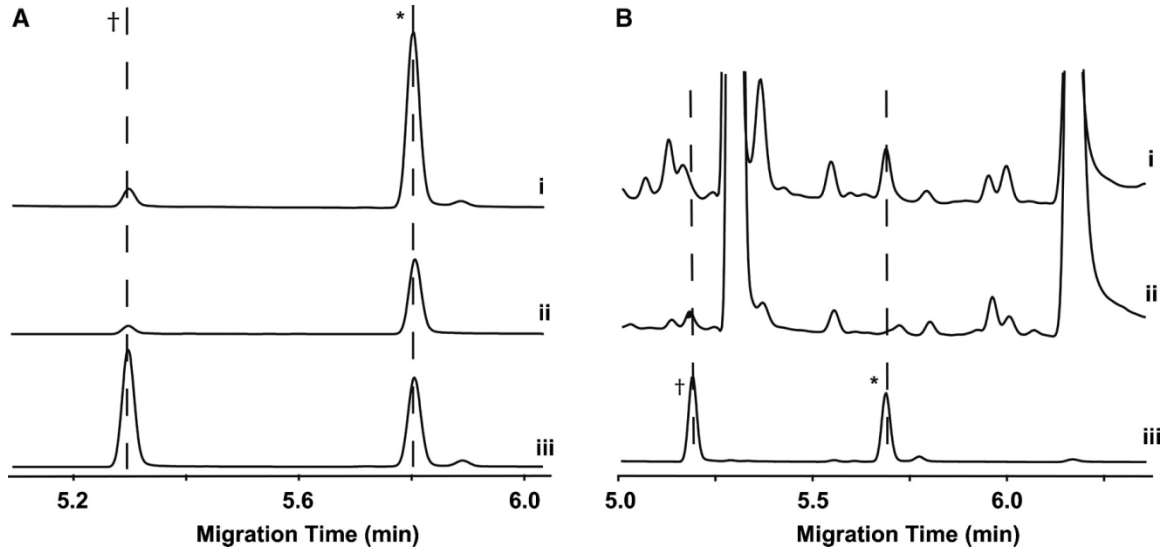


Figure 4.2: Xylosyltransferase assay and analysis of glycans found on secreted EGF12-18
(**A**) Capillary electrophoresis assay of xylosyltransferase activity in CHO microsomal fractions using 5 (XXG) and 6 (XG) as standards. (i) Assay mixture containing; CHO microsomal fraction, UDP-Xylose, APTS-labeled standards 5 (XXG, denoted by *) and 6 (XG, denoted by †), (ii) Assay mixture containing; CHO microsomal fraction, UDP-Xylose, and standard 6 (XG), (iii) Assay mixture containing; CHO microsomal fraction with standards 5 (XXG) and 6 (XG). (**B**) Analysis of APTS-labeled oligosaccharides released from purified recombinant EGF repeats isolated from the conditioned media of transfected CHO cells, (i) APTS-labeled oligosaccharides isolated from EGF constructs of CHO media, (ii) APTS-labeled oligosaccharides isolated from EGF constructs of CHO lysates, (iii) synthetic standards 5 (XXG, denoted by *) and 6 (XG, denoted by †).

4.3.4 β-Elimination, fluorescent labeling, and analysis of EGF12-18 Notch 1 *O*-glycans by capillary electrophoresis

Partially purified EGF12-18 Notch 1 and an aliquot of the vector control sample were dialyzed in parallel against a solution composed of a low concentration of volatile buffer. After lyophilization of the dialyzed EGF12-18 and control samples we obtained a white powder that we subjected to non-reducing *β*-elimination conditions to liberate all *O*-glycan moieties[285]. Following *β*-elimination and sample clean-up, we added maltose to all samples to serve as an internal standard after which we reductively aminated all free aldehydes, and ketones with APTS using standard procedures[286]. This reaction is carried out at microgram scale, which therefore limits the ability to characterize the resulting products using standard procedures. However, the reductive amination method of labeling glycans has been widely used and validated[286]. Once labeled, we analyzed the APTS-labeled *O*-glycan moieties by CE using NCHO-coated

capillaries (Figure 4.2B). We observe a large peak in the EGF12-18 sample having the same electrophoretic mobility as the XXG-APTS standard. A second peak having an apparent mobility similar to the XG-APTS standard was also observed. In contrast, on analysis of the *O*-glycan moieties released from the control sample, we only observed minor species having mobilities similar to either the XXG-APTS and XG-APTS standards (Figure 4.2Bii).

4.3.5 *Sulfolobus solfataricus* α-xylosidase catalyzed digestion of *O*-glycan moieties released from EGF12-18 and the vector control

To evaluate the structure of the APTS-glycan giving rise to the peak of interest in the trace obtained from the EGF12-18 sample we opted to digest non-reducing terminal *α*-linked xylose residues using an *exo*-acting α-xylosidase. The α-xylosidase previously used by Hase *et al.*[284] to digest this glycan is no longer available and no other enzymes having the required specificity are known. We therefore sought different sources of α-xylosidases and first assayed an α-xylosidase from *E.coli* (YicI)[287], which in our hands was incapable of digesting the XXG-APTS standard. We therefore turned to using another α-xylosidase (XylS), previously cloned from *Sulfolobus solfataricus* and recombinantly expressed in *E. coli*, that was generously provided by Prof. Moracci[288]. Using XylS, we digested, in parallel, samples containing the XXG-APTS standard, labeled *O*-glycan moieties released from EGF12-18, and the vector control (Figure 4.3A). After overnight digestion, we found that CE analysis of the XylS treated XXG-APTS standard revealed that nearly all of the XXG-APTS had been digested and a peak of similar area with a mobility matching that of the XG-APTS standard appeared (Figure 4.3Aii), revealing XylS could process the linkage of interest. We observed similar results for the labeled EGF12-18 *O*-glycan moieties, with the peak corresponding to XXG-APTS almost completely disappearing on digestion while a peak having a mobility matching that of the XG-APTS standard increased correspondingly in intensity (Figure 4.3Av). No differences could be observed in the vector control sample and we noted that the minor peak migrating close to the position of the XXG-APTS standard appeared strikingly

similar to the minor peak remaining after digestion of the EGF12-18 sample (Figure 4.3Av).

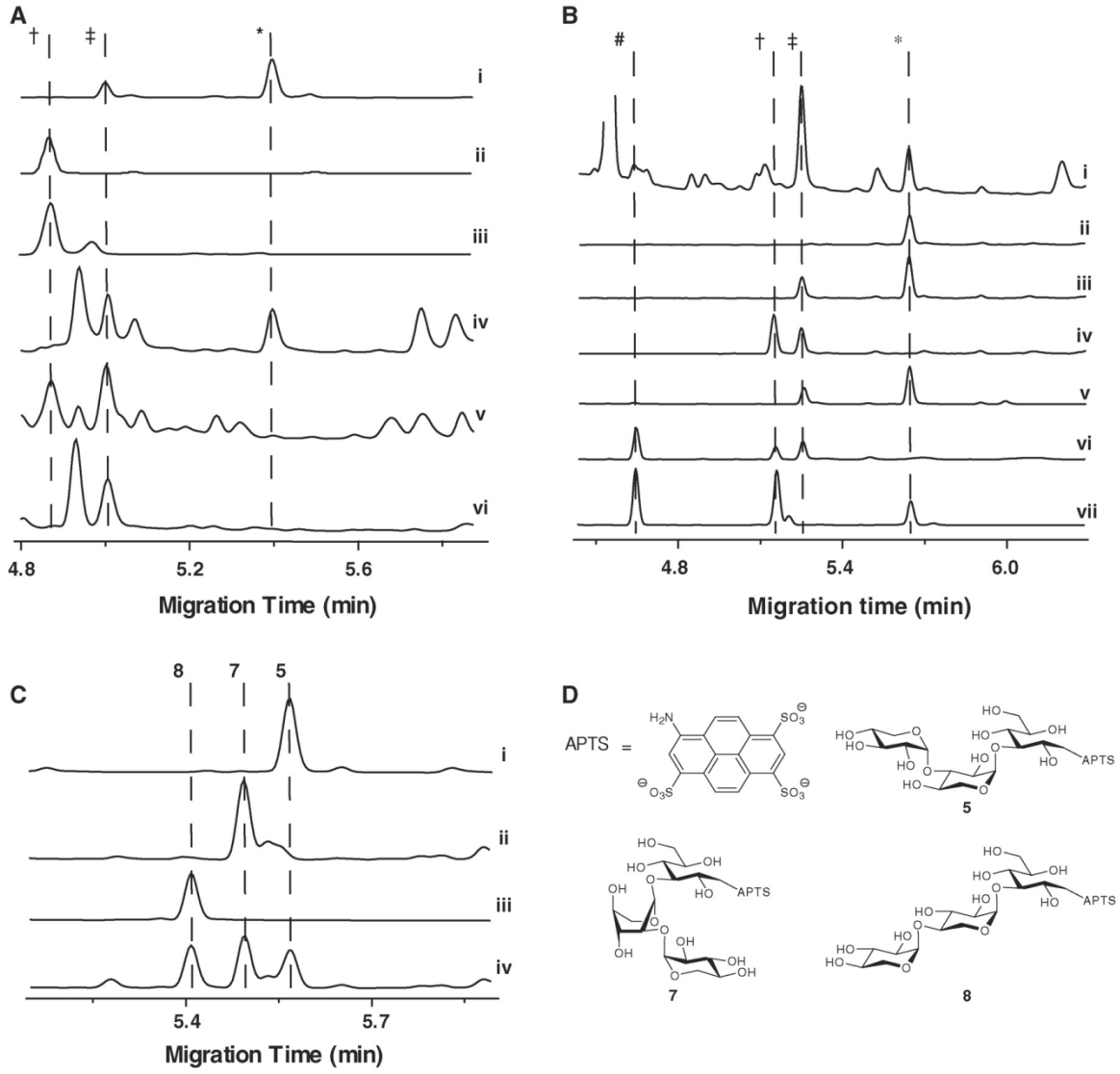


Figure 4.3: Gel purification of the putative XXG-APTS obtained from EGF12-18, *α*-xylosidase digestion and linkage assignment
(**A**) CE analysis of the *α*-xylosidase digestion of APTS-labeled oligosaccharides released from EGF12-18 isolated from the media of transfected CHO cells. Analysis of (i) synthetic standard 5 (denoted by *) and APTS-labeled maltose as the internal standard (denoted by ‡), (ii) the *α*-xylosidase digested synthetic standard 5, (iii) synthetic standard 6 (denoted by †), (iv) the APTS-labeled oligosaccharides released from recombinant EGF constructs of CHO cells and the internal APTS-labeled standard maltose, (v) the *α*-xylosidase digested APTS-labeled oligosaccharides released from recombinant EGF constructs and APTS-labeled maltose as the internal standard, (vi) the oligosaccharides released from the vector control and APTS-labeled maltose as the internal standard. (**B**) Digestion of XXG purified from mono- and oligosaccharides released from EGF12-18. Analysis of; (i) the APTS-labeled oligosaccharides released from recombinant EGF constructs of CHO cells and APTS-labeled maltose as the internal standard (denoted by ‡), (ii) the FACE-purified material corresponding to the putative XXG trisaccharide released from the EGF12-18 that has the same elution time as synthetic standard 5, (iii) the FACE-purified putative XXG trisaccharide spiked with APTS-labeled maltose as the internal standard, (iv) the *α*-xylosidase digested FACE-purified putative XXG trisaccharide released from EGF12-18 subsequently spiked with APTS-labeled maltose as the internal standard, (v) the FACE-purified putative XXG trisaccharide spiked with APTS-labeled maltose as the internal standard, (vi) the FACE-purified putative XXG trisaccharide released from EGF12-18 and further purified by HyperCarb spin-tip cartridges and then digested by the XylS *α*-xylosidase, (vii) the synthetic APTS-labeled standards **5** (denoted by *), **6** (denoted by †), and APTS labeled D-glucose (denoted by #). (**C**) CE analysis of XXG stereoisomers. (i) XXG-APTS standard (**5**), (ii) X1-2XG-APTS standard (7), (iii) X1-4XG-APTS standard (8), (iv) a mixture of the X1-4XG, X1-2XG, and XXG–APTS standards (**5**, **7**, and **8**). (**D**) APTS labeled synthetic standards; D-xylose-*α*1-3-D-xylose-*α*1-3-D-glucopyranose (**5**), D-xylose-*α*1-2-D-xylose-*α*1-3-D-glucopyranose (**7**), and D-xylose-*α*1-4-D-xylose-*α*1-3-D-glucopyranose (**8**).

4.3.6 Digestion of partially purified XXG trisaccharide from EGF12-18

We carried out further studies on the species giving rise to the peak having the same mobility as the trisaccharide standard to provide further support for its unambiguous assignment as the XXG trisaccharide. We electrophoresed the *O*-glycan moieties released from the EGF12-18 sample on a polyacrylamide gel and bands migrating at the same position as the di- and trisaccharide APTS standards were excised and eluted from the gel. Our analysis of the putative trisaccharide by CE revealed a peak migrating with an elution time matching that of the trisaccharide with no peaks in the immediate vicinity of the mono- and disaccharide standards (Figure 4.3Bii). After digesting this sample with XylS we carried out CE analysis and found a dramatic decrease in the area of the peak having the same mobility as the XXG-APTS standard and a corresponding increase in the intensity of the peak having a mobility matching that of the XG-APTS standard (Figure 4.3Biii). We speculated that Tris, a buffer present within the gel and likely co-purified during the gel purification procedure, was inhibiting

the complete digestion of the putative trisaccharide by XylS. Indeed, Tris has been shown to inhibit glycosidases[289], a finding that we also replicate here for XylS (data not shown). Therefore, we carried out a second purification step of the FACE-purified putative XXG trisaccharide, using a solid phase extraction on a porous graphite stationary phase to which glycans are known to bind, to remove any Tris or other buffer salts. This allowed us to use XylS to fully digest, the putative trisaccharide down to a monosaccharide having the same mobility as the D-glucose-*β*1-APTS (G-APTS) standard (Figure 4.3Bvi). Notably, analysis of the sample obtained from the region of the gel that would contain the APTS labeled XG disaccharide revealed no peaks having a mobility matching that of the XG-APTS standard. These collective data suggest that the *O*-glucose moieties of EGF12-18 of mouse Notch1 do not exist as the XG disaccharide in this cell line, and provides strong support for the identity of the Notch1 *O*-glucose containing glycan being the XXG trisaccharide found on the blood coagulation factors. Notably we find, consistent with previous studies of XylS[288], that this enzyme does not efficiently digest α-glucosides; digestion of a maltose-APTS standard using the same condition as used for the XXG-APTS standard revealed negligible change to the maltose-APTS standard. These results suggest that under the conditions of these studies XylS is very highly selective for terminal α-xylosidically linked residues.

4.3.7 Further linkage analysis of the XXG trisaccharide

To further investigate whether the putative XXG trisaccharide modifying EGF repeats of mouse Notch1 was indeed the same XXG trisaccharide initially identified by Hase and coworkers on EGF repeats of blood coagulation factors IIV and IX, as D-Xyl-*α*1-3-D-Xyl-*α*1-3-D-Glc[145,146], two linkage isomers were also synthesized. We prepared X1-2XG and X1-4XG and labeled these isomers with the APTS fluorophore (Figure 4.2). The three trisaccharide standards were analyzed by CE individually and as a mixture to accurately establish the electrophoretic mobility of each standard in comparison with the XXG-APTS (Figure 4.3C). Both the X1-2XG and X1-4XG standards have a higher mobility

than the XXG-APTS standard, with X1-4XG having the highest mobility. The small peak appearing alongside the X1-2XG-APTS standard is a small impurity obtained during its synthesis and purification and we identified it by NMR spectroscopy through its characteristic anomeric peaks as D-Xyl-*β*1-2-D-Xyl-*α*1-3-D-Glc-APTS standard. The resolution of the three linkage isomers provides compelling evidence that the linkage between the two xylose moieties is an α1-3 linkage. Further, given the very high sensitivity of CE for variations in fine glycan structure, it seems most likely that any linkage variations at the Xyl-Glc linkage would show similar, or more pronounced, changes in electrophoretic mobility of not only the disaccharide but the trisaccharide as well, relative to the mobility of the predicted D-Xyl-*α*1-3-D-Xyl-*α*1-3-D-Glc structure. These data therefore provide further strong support for assignment of the glycan structure decorating Notch as the XXG trisaccharide and underscore the established utility of CE for studying the fine structure of glyans.

4.3.8 Mass Spectroscopic analysis of per-methylated XXG standard and per-methylated trisaccharide released from EGF12-18

While establishing linkage information within glycans or distinguishing isomeric saccharide residues is not yet reliably possible using mass spectrometry[282], it is a powerful technique that can provide valuable data as to the mass and constitution of oligosaccharides and glycoconjugates[290]. Coupled with our previous studies using the four synthetic standards (**1**-**4**) and CE analysis, MS offered us with a complimentary technique that enabled us to further clarify the identity of the EGF-derived *O*-glycan as the XXG trisaccharide. The EGF-derived *O*-glycan, a portion of the sample not labeled with APTS, was per-methylated, purified on a Sep-Pak C_{18} cartridge, and characterized by mass spectrometry. The most intense molecular ion peak (Figure 4.4A and C), observed at *m/z* 593.30, corresponds to a sodiated trisaccharide composed of two pentoses and one reducing hexose (predicted mass of 593.29). This ion was selected for collisional activation and the MS/MS spectrum was recorded to assist in sequence assignment (Figure 4.4B and D). The Y_1 and Y_2 ions[291], at

m/z 419 and 259, are the result of the cleavages of the glycosidic linkage and confirm the loss of a reducing hexose and non-reducing terminal pentose, respectively. These losses are consistent with the results of complete XylS digestion of the EGF-derived *O*-glycan, which gave a single product that co-migrated with APTS-labeled glucose (Figure 4.3B). A less intense C_1 ion at *m/z* 215 also confirms the non-reducing terminus is a pentose while a B_2-type ion at *m/z* 357 corresponds to a fragment consisting of two linked pentose units confirming that the glycan is not branched. These results indicate that the EGF-derived trisaccharide is linear and consists of a reducing hexose and two pentose residues, in agreement with the very recent MS/MS data of the *O*-glucose containing trisaccharide present on a different construct of Notch, EGF1-5, reported by Bakker *et al.*[280] and our detailed CE-based assignment of the XXG glycan detailed here.

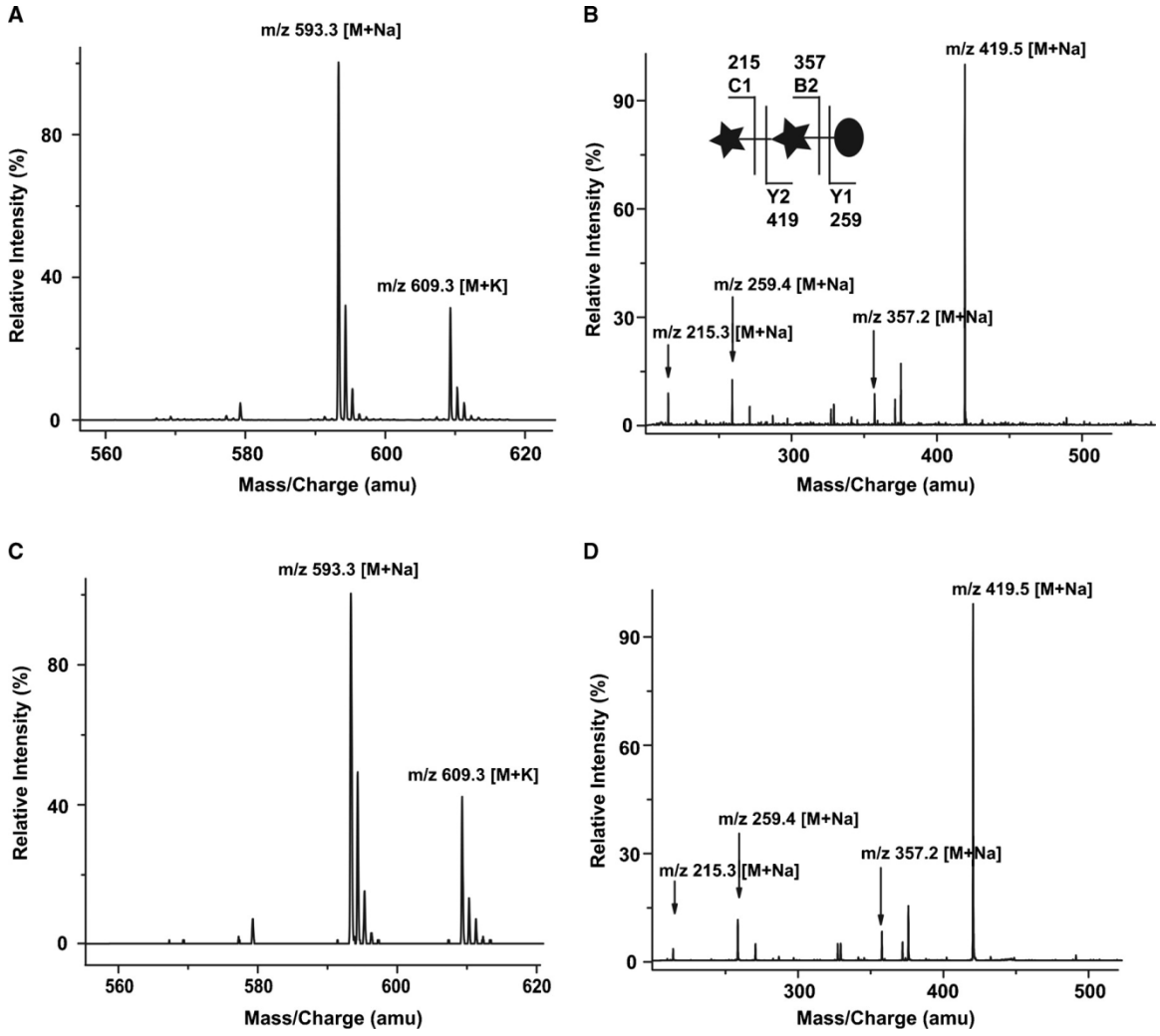


Figure 4.4: Mass spectral analysis of the putative XXG trisaccharide released from EGF12-18

(**A**) Mass spectrometric analysis of per-methylated XXG released from EGF12-18, identification of three characteristic ions; 571.3 [M+H], 593.3 [M+Na], 609.3 [M+K]. (**B**) MS/MS analysis of the sodiated parent ion from the putative XXG trisaccharide, fragmentation pattern with ions having amu's corresponding to C1 (215.3), C2 (359.2), Y2 (419.5), and Y1 (259.4). (**C**) Mass spectrometric analysis of the per-methylated XXG (1) synthetic standard, identification of two characteristic ions; 593.3 amu [M+Na], 609.3 amu [M+K]. (**D**) MS/MS analysis of the sodiated parent ion from the synthetic XXG trisaccharide (1), fragmentation pattern with ions having amu's corresponding to C1 (215.3), C2 (359.2), Y2 (419.5), and Y1 (259.4).

4.4 Discussion

Notch is a key player in the Notch signaling pathway and has been found to be modified by two unusual forms of *O*-linked glycosylation[148]. The first of these, *O*-fucosylation, has been shown to be elongated to form the tetrasaccharide, Sia-*α*2-3-Gal-*β*1-4-GlcNAc-*β*1-3-Fuc-*α*1-*O*-Ser/Thr[279,292]. Proper assembly of this tetrasaccharide has been proposed to be important in governing Delta1 and Jagged1 mediated signaling[12,173]. Elegant studies by Haltiwanger *et al*., using genetically modified Lec cell lines and radioactively labeled fucose moieties has illuminated the structure of, and revealed sites of Notch1 decorated by this non-canonical *O*-fucosyl tetrasaccharide[121,126,148]. The second form of non-canonical glycosylation found on Notch, *O*-glucosylation, can be elongated to form a more complex glycan[148,280]. Rumi, a highly conserved protein *O*-glucosyltransferase (Poglut) has been identified in Drosophila as being responsible for transferring the glucose residue to the target serine of EGF repeats of Notch. Further, Rumi has been shown in Drosophila to be essential for proper Notch signaling, an observation that underscores the functional importance of this modification and its potential importance in human health[175]. Interestingly, in Drosophila, the *O*-glucose residue installed by Rumi on Notch does not appear to be further elaborated by other sugar residues[175]. We therefore sought to elucidate the detailed structure of the *O*-glucose containing glycan present on mammalian Notch1[148] since this should facilitate studies on the role of this glycosylation in the functioning of mammalian Notch isoforms. The existing methods to study the *O*-glucose modification, involving the use of genetically modified CHO cell lines in conjugation with metabolic labeling using monosaccharides[148,293], or mass spectrometry[280], is not readily amenable for elucidating the fine structure of this glycan. Therefore, to address this question, an alternative method for studying the non-canonical *O*-glucosylation of mammalian Notch1 was required. Given that *O*-glucose linked to secreted glycoproteins, such as bovine blood clotting factors IIV and IX, and protein Z, have been shown to occur within EGF-like domains as the XXG trisaccharide, we felt it was likely that XXG is the glycan decorating mammalian Notch, as has

been previously postulated[148,175]. Accordingly, to realize our goal of developing a method that could be used, we first chemically synthesized the XXG trisaccharide (**1**), the XG disaccharide (**2**), and linkage isomers of the XXG trisaccharide, following which we subsequently fluorescently labeled these for use as standards with CE (Figure 4.2). Given the sensitivity of this CE method only very small quantities of material are necessary for such studies, limiting the need for tedious and involved chemical synthesis on large scale. Following APTS labeling, we find that purification of the resulting standards by way of gel electrophoresis removes impurities arising from the fluorophore labeling procedure and affords pure XG- and XXG-APTS samples, ideal for use as internal and external standards.

Before proceeding further we felt it prudent to ensure that the CHO K1 cell line we were aiming to use possessed the necessary xylosyltransferases required to elongate the serine linked glucose residue. We therefore tested microsomal fractions purified from CHO K1 cell lysates for xylosyltransferase activity[284] using the XG-APTS standard. In the presence of UDP-xylose, CHO microsomal fractions effectively converted XG-APTS to XXG-APTS, indicating the presence of a xylosyltransferase having the appropriate specificity (Figure 4.3A(ii)). In a similar experiment glucose was coupled to the APTS fluorophore to form G-APTS and incubated with UDP-xylose and CHO microsomal fractions. No formation of either XG-APTS or XXG-APTS was observed (data not shown), an unsurprising result given that the structure of G-APTS has the glucose residue in the open chain form, making it quite possible that the xylosyltransferase would be unable to recognize this acceptor. Regardless, the CHO microsomal assay confirms the presence of a xylosyltransferase capable of modifying a XG disaccharide with xylose to form the XXG trisaccharide.

With this knowledge in hand, we next turned to generating appropriate N^{ECD} constructs that could be analyzed for the presence of the *O*-glucose containing glycan. The 36 consecutive EGF repeats of the N^{ECD} of mouse Notch1 includes sixteen repeats that contain the proposed *O*-glucose consensus sequence NH_2-Cys^1-Xxx-Ser-Xxx-Pro-Cys^2-COOH. Of these sixteen consensus

sequences, eight occur between EGF repeats 12-21, an observation that prompted us to design multiple constructs encoding this particular region. Of the constructs that we generated, the one comprising EGF domains 12-18 were most efficiently expressed in CHO cells. This construct contains five proposed consensus sequences that may be modified with *O*-glucose. Purification of recombinant EGF12-18 from CHO cell lysates and media (Figure 4.1E), followed by *β*-elimination of the *O*-glycans and subsequent APTS labeling, allowed us to analyze the liberated *O*-glycans by CE. Peaks having the same elution time as the synthetic XG- and XXG-APTS standards were observed in the EGF12-18 sample (Figure 4.3Bi). Comparison of the standards with the vector control sample revealed a peak having a similar elution time as the XG-APTS standard as well as a minor peak in close proximity to the XXG-APTS standard. The peak having similar mobility as the XG disaccharide was found in both the EGF12-18 construct and the control sample and were both of similar intensities, suggesting this peak is some species other than the XG disaccharide. In contrast, the peak having the same elution time as the XXG-APTS standard in the sample derived from EGF12-18 was much larger than the peak having a similar mobility in the control sample, suggesting that the XXG trisaccharide does indeed decorate EGF12-18 of mouse Notch1 (Figure 4.3Bi). The complexity of the chromatogram, however, prevented conclusive assignment of this peak as the XXG trisaccharide, forcing us to more carefully investigate this possibility.

Therefore, in order to more conclusively establish whether this peak arises from the XXG trisaccharide modification of EGF12-18, we elected to use the *exo*-*α*-xylosidase XylS[288] to cleave the terminal xylose residues. This *α*-xylosidase is implicated in the degradation of plant xyloglucan and has been shown to cleave the xylose moiety off from isoprimeverose (D-xylopyranosyl-*α*1-6-D-glucopyranosyl) but does not cleave *β*-linked xylosyl residues[288]. Gratifyingly, this *α*-xylosidase was able to digest the XXG-APTS standard (Figure 4.3Av), supporting its use in these studies to aid in structural assignment of the glycans of interest. Notably, XylS did not degrade a maltose-APTS standard, indicating this enzyme is specific for terminal xylose residues when using the conditions

detailed here. Digestion of the APTS labeled *O*-glycan moieties released from the EGF12-18 sample revealed that the putative XXG trisaccharide having a mobility matching that of the XXG-APTS standard was digested using XylS and a new peak having the same elution time as that of the XG-APTS standard was observed (Figure 4.3Av). The identical electrophoretic mobility of the putative APTS labeled XXG-trisaccharide, matching that of the XXG-APTS standard, and its susceptibility to α-xylosidase digestion to yield products having a mobility matching that of the XG-APTS standard, supports assignment of the trisaccharide as having the XXG structure proposed by Hase.

Collectively, these results suggest that the *O*-glucose containing glycan of mammalian Notch is the XXG trisaccharide. As an aside, the α-xylosidase XylS should prove to be a useful tool for analysis of glycans having this D-Xyl-α1-3-D-Xyl structure, a linkage for which no commercial enzymes are currently available[288]. Due, however, to the presence of impurities in and around the region of the chromatogram in which the disaccharide is observed, analysis of these crude samples does not conclusively show the product of the trisaccharide digestion is the XG disaccharide, leaving some ambiguity in the assignment of the structure of this glycan. We therefore used the gel purified XXG trisaccharide, released from EGF12-18, to evaluate the fate of this putative trisaccharide following α-xylosidase digestion in the absence of the complicating impurities. Digestion of the gel purified trisaccharide revealed the glycan isolated from EGF12-18 of mouse Notch1 did indeed yield, as we suspected, the XG disaccharide and G monosaccharide (Figure 4.3B vi). Furthermore, when analyzed by CE, a sample of the region of the gel containing the XG disaccharide did not reveal any peaks corresponding to the XG-APTS standard (data not shown). This null observation suggests that the XG disaccharide is a building block en route to the XXG trisaccharide and that rapid stoichiometric conversion to the XXG trisaccharide occurs in CHO-K1 cells before EGF12-18 is secreted into the media. Further, we find that EGF12-18 isolated from cell lysates did not appear to bear any XXG or XG glycans, suggesting that secretion occurs rapidly after the construct is glycosylated.

Despite these findings and the established high resolution of CE we were stimulated by the possibility that the glycan isolated by us from Notch expressed in CHO cells may be an isomer of the XXG glycan (**1**), having different linkages, that simply has the same electrophoretic mobility as XXG (**1**). Therefore, to carefully clarify whether the linkages of the XXG trisaccharide modifying EGF repeats of mouse Notch1 are the same as the XXG trisaccharide identified by Hase *et al.*[145,146], we synthesized X1-2XG and X1-4XG, labeled these with APTS, and then analyzed these standards in conjunction with the XXG trisaccharide standard by CE (Figure 4.3C and D). Each linkage isomer has different electrophoretic mobilities enabling them to be completely resolved and this data further establishes that only XXG has the same electrophoretic mobility as the APTS labeled glycan released from the EGF12-18. This data provides very strong evidence to support assignment of the glycan as D-xylose-*α*1-3-D-xylose-*α*1-3-D-glucopyranose. However, to provide additional evidence as to its constitution we carried out mass spectrometric studies of the per-methylated putative XXG glycan of Notch expressed in CHO cells as well as the synthetic XXG standard. MS of both the standard and the glycan revealed a sodiated trisaccharide composed of two pentoses and one hexose (Figure 4.4A and C). Additionally, tandem mass spectrometry (MS/MS) analysis of both the per-methylated sample and standard revealed identical fragmentation patterns giving rise to signals having matching m/z values, and relative ion counts, in the daughter ion spectra. These data reveal that the hexose is coupled to the protein and that the trisaccharide is linear rather than branched (Figure 4.4B and D). While this MS and MS/MS data cannot clearly establish the linkages or anomeric configurations of the residues within the putative XXG trisaccharide, it does provide valuable information regarding order of the individual pentoses and hexoses within the linear arrangement. These MS and MS/MS results are also consistent with very recent data indicating that the glycan on Notch is a linear glycan composed of two pentoses with one hexose at the reducing terminus[280].

Collectively these experiments provide very strong evidence supporting assignment of the *O*-glucose-containing glycan modifying mouse Notch1, and

most likely the other Notch isoforms, as the XXG trisaccharide found on secreted glycoproteins of the blood coagulation cascade. While the data supporting assignment of the glycan we characterize here as the XXG glycan is very strong, there is a remote possibility that this glycan may be a different linkage isomer. Such a linkage isomer would have to have the identical electrophoretic mobility as the XXG-APTS standard as well as the same MS/MS fragmentation pattern as the synthetic XXG glycan. Such a linkage isomer would also need to be susceptible to degradation by XylS to generate a disaccharide having an identical electrophoretic mobility as the XG-APTS standard. Given the very high resolution of the CE method used here, underscored by our ability to resolve at baseline the X1-4XG, X1-2XG, and XXG-APTS standards, it seems highly unlikely that the APTS-labeled disaccharide liberated by XylS from the APTS-labeled trisaccharide would be anything other than XG-APTS, as we propose here. Therefore, we believe that these chemically synthesized XG-, XXG-, and isomeric XXG-APTS standards along with CE and using an appropriate α-xylosidase offer a new and complementary approach for studying this noncanonical *O*-glycosylation in various tissues having different origins. This strategy avoids the need for radioactivity and mutated cell lines lacking certain metabolic pathways, thus making it ideal for studying wild-type cells. Of course, this strategy necessitates availability of appropriate instrumentation and the availability of standards but it does not require access to mass spectrometry equipment. Further technical refinement of this strategy to improve sensitivity will make it a valuable tool for studying noncanonical *O*-glucosylation in tissues from model organisms at endogenous levels. With the fine structure of the XXG trisaccharide present on Notch firmly established, one obvious area of interest that we are pursuing is to generate antibodies that recognize these glycan structures in their natural context. These tools should facilitate studies directed at improving our understanding of the developmental roles of this interesting post-translation modification.

4.5 Materials and methods

4.5.1 Materials

All buffer salts in this study were purchased from BioShop Canada Inc. except for sodium dodecyl sulfate which was purchased from Biorad. Milli-Q (18.2 megaohms cm^{-1}) water was used to prepare all buffers. Mouse Notch1 cDNA was purchased from the ATCC. EDTA-free protease inhibitor tablets were purchased from ROCHE. Histrap HP 5 mL affinity column was purchased from GE Healthcare. 10 kDa molecular weight cut off centrifugal filters were purchased from Millipore Corporation. SuperSignal West Pico chemiluminescence and CL-XPosure Film were purchase from Pierce. Synthetic reactions were monitored by TLC using Merck Kieselgel 60 F_{254} aluminum-backed sheets while flash column chromatography under a positive pressure was performed using Merck Kieselgel 60 (230-400 mesh) using the specified eluants. All chemicals were purchased from Sigma-Aldrich unless otherwise noted and ^{1}H and ^{13}C NMR spectroscopy were recorded on a Varian AS500 Unity Innova spectrometer at 500 MHz (125 MHz for ^{13}C) and 600 MHz Bruker AMX spectrometer (150 MHz for ^{13}C) as indicated.

4.5.2 Production of mouse Notch1 EGF fragments

Constructs encoding EGF fragment 1-5, 6-11, 12-18, 16-23 and 24-33 of the mouse Notch1 extracellular domain were generated using PCR with pcDNA1-Notch1 as a template. Restriction sites for *EcoR*I and *Mlu*I were designed into the primers for a signal peptide and a hemagglutinin (HA) tag. The PCR products were then digested with *EcoR*I and *Mlu*I and subcloned into the corresponding sites in the mammalian expression vector pax142[283]. The EGF fragments were designed to contain between 1 and 5 predicted *O*-glucosylation sites based on the proposed consensus sequence: NH_2-Cys^1-Xxx-Ser-Xxx-Pro-Cys^2-COOH[147]. Restriction sites for *Mlu*I and *Sal*I were designed into the primers for the EGF constructs along with a His_6 tag. The primers to generate the five constructs were; EGF1-5 5′-GCGATAACGCGTGGCTTGAGATGCTCCCAGC -3′

and 5′-GCCACCCATACTGGTCCCCATCATCACCATCACCACTAAGTCGACTATCGC-3′, EGF6-11 5′-GCGTAACGCGTGGTCCCCACTGTGAACTGC-3′ and 5′-CAGTGTCTACAGGGCTACACGCATCATCACCATCACCACTAAGTCGACTATCGC-3′, EGF12-18 5′-GCGATAACGCGTCAGGGCTACACGGGACC-3′ and 5′-GGCTACCATGACCCCACGCATCATCACCATCACCACTAAGTCGACTATCGC-3′, EGF16-23 5′-GCGATAACGCGTACCACAGGGCCCAACTGTG-3′ and 5′-GCCGGCTATACAGGTCGCCATCATCACCATCACCACTAAGTCGACTATCGC-3′, EGF24-33 5′-GCGATAACGCGTTATACAGGTCGCAACTGTGAGAG-3′ and 5′-GCCACACTGGACGCCGCCATCATCACCATCACCACTAAGTCGACTATCGC-3′,. The PCR product was then digested with MluI and SalI and subcloned into the corresponding sites in the mammalian expression vector pAX142. All constructs were sequenced to confirm nucleotide sequence.

4.5.3 Cell culture and production of recombinant Notch1 EGF 12 – 18 repeats

CHO K1 cells were grown in 100 mm polystyrene dishes (Sarstedt) at 37 ºC under an atmosphere of 5% CO_2. Unless otherwise indicated, cells were grown in a mixture of Dulbeco's modified Eagle's medium (DMEM) and Ham's F12 (1:1, v/v) (Gibco) supplemented with 15% fetal bovine serum (FBS) (Gibco). Confluent monolayers of CHO K1 cells were transiently transfected with the series of His6-tagged pAXEGF constructs using Lipofectamine 2000 (Invitrogen) following the manufacturer's instructions. 12 hours after transfection the media was changed. To purify the secreted EGF-repeat-containing constructs, the cell culture media was collected every 24 h and replaced with fresh media. Collected media was supplemented with 1 x Complete^{TM} EDTA-free protease inhibitors (Roche) and *α*-D-xylose (to a final concentration of 200 mM) to inhibit endogenous xylosidases. Detached cells in the media were pelleted by centrifugation (800 rpm, 10 min, 4 ºC) and discarded. After three days cells were washed twice with 10 mL portions of cold phosphate-buffered saline (PBS), pH

7.4, scraped off of the plates, and pelleted by centrifugation (800 rpm, 10 min, 4 ºC). Both cell pellets and cell culture media were frozen in liquid nitrogen and stored at -20 °C prior to use.

4.5.4 Purification of the series of EGF Constructs and the pAX142 Vector Control from CHO Media

NaCl (500 mM) and imidazole (10 mM) were added to conditioned media, and cell lysates were centrifugated (2000 rpm and 4 ºC for 30 minutes) to pellet any precipitate before loading onto a 5 mL Histrap HP affinity column (GE Healthcare). Loading, washing, and elution of the construct was done according to the column manufacturer's protocol, and the eluate was dialyzed three times over a period of three days into a volatile buffer at pH 7.4 containing 50 mM NH_2CO_3. A small portion of the eluate was concentrated 10 fold using 10 kDa molecular weight cut off centrifugal filter (Millipore) and run on Western blot. The purified protein product could not be detected by Bradford assay or by Coomasie staining of SDS-PAGE gels, presumably due to the limited expression levels of the construct studied here using this expression system.

4.5.5 Western blot analysis

This procedure was carried out essentially as described before[294]. Samples were resolved on 12% SDS-PAGE gels and then transferred to nitrocellulose (Bio-rad) membranes. Membranes were blocked for 1 h at room temperature with 1% bovine serum albumin (BSA) in PBS containing 0.1% Tween-20 (Sigma) (PBS-T) and then subsequently probed with the anti-hemagluttinin (α-HA) primary antibodies in PBS-T, containing 1% BSA, overnight at 4 ºC. Membranes were then extensively washed with PBS-T, blocked again for 30 min with 1% BSA in PBS-T at room temperature and then probed with the appropriate horseradish peroxidase (HRP)-conjugated secondary antibody for 1 h at room temperature delivered in 1% BSA in PBS-T. Finally the membranes were washed extensively and then developed with SuperSignal West Pico

chemiluminescence substrate (Pierce) and exposed to CL-XPosure Film (Pierce).

4.5.6 Beta-elimination of EGF12-18 *O*-glycans

Purified EGF12-18 was lyophilized and the resulting powder was suspended in 1 mL of 18Ω deionized H_2O, transferred to 1.5 mL screw-capped tubes (Sarstedt), and lyophilized again. The dried protein was subjected to NH_3-based, non-reductive *β*-elimination according to the procedure of Huang *et al*. [285]. Briefly, samples were suspended in 1 mL 28% NH_4OH saturated at room temperature with $(NH_4)_2CO_3$. After the addition of more $(NH_4)_2CO_3$ (100 mg) samples were vortexed and heated at 60 ºC for 40 h with intermittent shaking to keep the samples suspended. Excess salts were removed from samples by repeated cycles of lyophilization using a SpeedVac™ (Thermo) and reconstitution in 18Ω deionized H_2O; this process typically was repeated over a period of 30 h until no salts were visible in tubes derived from mock transfected cells. The *β*-eliminated *O*-glycans, presumably trapped as glycosylammoniumcarbonate salts, were converted into reducing oligosaccharides by acidification with 10 µL 500 mM H_3BO_3 (30 min, 37 ^oC). 1 mL of CH_3OH was then added and the supernatant was collected and dried to remove borate salts. In contrast with Huang *et* al., excess salts and protein/peptides were removed by suspending the samples in 100 µL H_2O, applying them to a conditioned Hypercarb (Thermo) solid-phase extraction cartridge, washing with 3 mL H_2O and then by eluting the desired material as two separate fractions using 5% CH_3CN, 0.1% TFA and 50% CH_3CN, 0.1% TFA[295]. The latter fraction, which contains carbohydrates larger than disaccharides, was spiked with 1.0 nmol maltose and both fractions were then lyophilized and fluorescently labeled.

4.5.7 Fluorescent labeling of EGF12-18-derived *O*-glycans and sample clean-up

Following β-elimination, the dried reducing oligosaccharide material and the synthetic standards were reductively-aminated[286] with the highly fluorescent 8-aminopyrene-1,3,6-trisulfonate (APTS). For APTS labeling the manafacturer's protocol (Beckman-Coulter) was followed essentially as described. Briefly, samples were dissolved in equal volumes, (2 µL each) of 200 mM APTS (Beckman-Coulter) in 7.5% acetic acid and 1.0 M $NaBH_3CN$ (Fluka) in tetrahydrofuran (THF) and then incubated overnight, in the dark, at 37 °C. Stock APTS and $NaBH_3CN$ solutions were stored at -80 ^oC for no more than one month. After labelling, the standard oligosaccharides maltose, Xyl-α1,3-Glc and Xyl-α1,3-Xyl-α1,3-Glc were purified away from excess APTS and labeling reagents by fluorophore-assisted carbohydrate electrophoresis (FACE) according to the procedure Gao and Lehrman[296,297]. Briefly, samples were loaded onto a 1.5 mm thick 20% polyacrylamide gel, and subjected to 4.5 h of electrophoresis at 4 ^oC using a constant voltage of 250 V. When samples derived from transiently expressed EGF-repeat-containing constructs were subjected to FACE, care was taken to ensure that they were not placed in lanes adjacent to the synthetic standards XXG or XG. APTS-labelled material was visualized using a transilluminator (VWR), bands of interest were quickly excised using a fresh scalpel and the APTS-labeled material was eluted by incubating the gel slices in 2 mL of 18Ω deionized H_2O at room temperature, in the dark, with constant shaking. After 16 h the eluted APTS-labeled samples were transferred to a clean centrifuge tube and lyophilized.

4.5.8 Analysis of APTS-labeled oligosaccharides by capillary electrophoresis (CE)

96 µL 18Ω deionized H_2O was added to APTS-labeled samples to bring the total volume to 100 µL. CE separations were carried out using a ProteomeLab PA800 (Beckman-Coulter) equipped with a laser-induced fluorescence (LIF) detector and a 488 nm argon-ion laser. All separations were

carried out in reverse polarity using a 25 mM NaOAc buffer, pH 4.74, containing polymeric additives (Beckman-Coulter) at a constant voltage of 30 kV on polyvinyl alcohol coated NCHO capillaries (Beckman-Coulter), with 50 cm effective length and 50 µm i.d. as described by Guttman *et al*.[297] All injections were from the cathode side and were made by applying a pressure of 0.5 psi for 3 s.

4.5.9 UDP-D-xylose:α-D-xyloside α1-3-xylosyltransferase activity assay

Endogenous xylosyltransferase activity in CHO K1 cells was examined by using a procedure modified from Minamida *et al*.[22] 15 cm diameter plates containing confluent monolayers of CHO K1 cells (~$3x10^7$cells/plate) were washed twice with ice cold PBS (10 mL). Cells were incubated on ice for 15 min after adding cold HEPES-KOH buffer, pH 7.9, (1 mL/plate) containing 1.5 mM $MgCl_2$ and 10 mM KCl. Cells were scrapped off plates, homogenized using a mechanical homogenizer, and the lysates were centrifuged (700 g, 5 min). The resulting pellets were suspended in HEPES-NaOH buffer (pH 7.2) containing 20 mM $MgCl_2$ and 150 mM NaCl, homogenized with a Potter-Elvehjem homogenizer, and centrifuged (7000g, 4 ^oC, 5 min). The supernatant was further centrifuged (105000 g , 4 ^oC, 1 h) and the resulting pellets, each corresponding to the microsomes derived from ~$3x10^7$ cells, were suspended in HEPES-NaOH buffer (30 µL), frozen in liquid N_2, and stored at -80 ^oC until required. To test for xylosyltransferase activity, FACE purified XG-APTS (~3.75 nmol based on the amount of starting material initially labeled) and HEPES-NaOH buffer containing 5% Triton X-100 (24 µL) were gently mixed with CHO K1 microsomes (30 µL). The resulting mixture was then divided into three equal volumes which were supplemented with either; (A) FACE purified XXG-APTS (~0.12 nmol), (B) UDP-xylose (3.35 nmol), or (C) XXG-APTS (0.12 nmol) and UDP-xylose (3.35 nmol). The volumes of each reaction were adjusted to 40 µL with 18Ω deionized H_2O and incubated for 20 h at 37 ^oC. The reaction was terminated by boiling the samples for 3 min. CE analysis was performed after bringing the volume of each

sample up to 110 μL and centrifuging (13500 rpm, 1 min) to pellet any protein precipitates.

4.5.10 Optimization of α-xylosidase digestion

The pT7-SCII vector harbouring *XylS*, which encoded an archaeal *exo*-α-xylosidase (XylS) was generously provided by Moracci and coworkers. Recombinant XylS was obtained as previously described from *E. coli*[26]. To characterize XylS activity against the synthetic standards a series of assays were performed where reducing-, and APTS-labeled XG and XXG were incubated at either 37 or 60 °C for varying lengths of time. All reactions were performed in capped PCR tubes. Final conditions for the digestion of non-purified APTS-labeled mixtures (50 μL) were 60 °C, 3 h, in non-buffered solution after adding 1/10th volume of a XylS enzyme preparation. For partially digesting FACE purified material, the samples, in 50 μL H_2O, were mixed with 5 μL 5% AcOH to adjust the pH to roughly 5.5 prior to adding 1/10th volume XylS and digesting at 60 °C for 3h in the dark. Protein was removed after the digestion period by adding 3 volumes of cold EtOH and incubating samples for 30 min at -20 ^{o}C. After centrifuging the samples (13000 rpm, 4 °C, 10 min) the supernatant containing the digested oligosaccharides was dried, reconstituted in 100 μL H_2O, and analyzed by CE as described above. To remove Tris from the samples and enable further XylS digestion of the putative FACE-purified APTS labeled XXG-trisaccharide, samples were loaded onto 1 – 10 μL HyperCarb spin-tip cartridges (Thermo), preconditioned with 50 μL 25 mM TFA in 50% CH_3CN and 3 x 50 μL H_2O, washed with 50 μL H_2O and 50 μL 50 mM TFA, and then eluted with 50 μL 25mM TFA in 50% CH_3CN. The eluant was split into two equal portions and dried using a SpeedVac for 30 min before reconstitution in 15 μL of a solution containing XylS. A negative control was mixed with H_2O and both samples were incubated at 60 °C for 12 h in the dark prior to CE analysis as described above.

4.5.11 Oligosaccharide per-methylation and analysis by mass spectrometry

A portion of the EGF-derived, Hypercarb-purified *O*-glycans (not spiked with maltose or labeled with APTS) was lyophilized in 15 mL glass tubes and per-methylated according to standard procedures[298]. Briefly, dried samples were dissolved in 100 µL of glass-distilled DMSO and an equal volume of a DMSO/NaOH slurry was added and mixed. 100 µL of CH_3I was added, tubes were purged with N_2, and the mixtures were stirred at room temperature for 10 min after which the reaction was quenched by adding 2 mL of 18Ω H_2O. Excess CH_3I was removed using a stream of N_2 and per-methylated glycans were extracted into an organic phase of CH_2Cl_2. The organic layer was washed three times with H_2O and evaporated to give a solid residue. Derivatized samples were purified using a Sep-Pak C_{18} cartridge (Waters)[299] and fractions (eluting with an CH_3CN concentration between 35 and 75 %) were collected, pooled, and dried prior to analysis by mass spectrometry. Dried samples were diluted in 50 µL of 35 % CH_3CN containing 0.15 % formic acid. Mass spectra were acquired on a API 4000 Q Trap (Applied Biosystems) and mass spectrometric parameters were optimized using synthetically prepared samples per-methylated as described above. Samples were delivered by direct injection from a syringe through a micro ion spray head on a nanospray source (Applied Biosystems) with a PicoTip emitter (New Objective) at a flow rate of 300 nL/min. For conventional mass spectra the mean cumulative average of 50 scans were collected using the enhanced resolution scan type at a scan rate of 250 amu/s with a linear ion trap fill time of 10 ms. The declustering potential was set at 112 V. Collision gas was set such that instrument vacuum pressure was 4.7×10^{-5} Torr. In all cases, the sodiated parent ion had the greatest intensity and was subjected to MS/MS analysis in which a mean cumulative average of 50 MS/MS scans were collected using the enhanced product ion scan type at 250 amu/s and a collision energy of 56 V.

4.5.12 Characterization of D-xylose-α1-3-D-glucopyranose (2), D-xylose-α1-3-D-xylose-α1-3-D-glucopyranose (1), D-xylose-α1-2-D-xylose-α1-3-D-glucopyranose (3), and D-xylose-α1-4-D-xylose-α1-3-D-glucopyranose (4).

D-xylose-*α*1-3-D-glucopyranose (**2**) Rf = 0.2 (5:2, EtOAc:MeOH), ^{1}H NMR (600 MHz, D_2O) δ 3.27 (dd, 1H, *J* = 8.0, 9.3 Hz), 3.40-3.60 (m, 7H), 3.68-3.85 (m, 10H), 3.80- 3.88 (m, 2H), 4.00-4.20 (m, 2H), 4.67 (d, 1H, *J* = 8.0 Hz), 5.25 (d, 1H, *J* = 3.8 Hz. ^{13}C NMR (150 MHz, $CDCl_3$) δ 60.08, 60.25, 62.00, 62.86, 63.39, 64.70, 67.06, 68.70, 68.74, 69.10, 69.15, 69.18, 70.94, 70.98, 71.57, 72.26, 73.54, 73.64, 75.26, 75.34, 91.60, 95.41 δ HRESIMS: calcd for $[M+Na]^+$ $C_{11}H_{20}NaO_{10}$: 335.0949; found: 335.0955.

D-xylose-*α*1-3-D-xylose-*α*1-3-D-glucopyranose (**1**) Rf = 0.2 (5:2, EtOAc:MeOH), ^{1}H NMR (600 MHz, CD_3OD) δ 3.23 (dd, 1H, *J* = 7.8, 9.2 Hz), 3.41 (d, 1H, *J* = 3.7 Hz), 3.42 (d, 1H, *J* = 3.7 Hz), 3.45-3.49 (m, 4H), 3.52-3.58 (m, 7H), 3.61-3.81 (m, 14H), 3.83-3.92 (m, 5H), 4.50 (d, 1H, *J* = 7.8 Hz), 5.13 (d, 1H, *J* = 3.7 Hz), 5.16-5.19 (m, 4H). ^{13}C NMR (150 MHz, $CDCl_3$) δ 62.68, 62.68, 63.60, 63.77, 71.64, 71.69, 71.84, 71.90, 72.60, 72.69, 72.81, 72.88, 74.27, 75.12, 75.28, 77.78, 83.27, 83.31, 84.16, 86.63, 94.31, 98.51, 101.77, 101.79, 102.04, 102.19. HRESIMS: calcd for $[M+Na]^+$ $C_{16}H_{28}O_{10}$ + Na: 467.1371; found: 467.1385.

D-xylose-*α*1-2-D-xylose-*α*1-3-D-glucopyranose (**3**) Rf = 0.2 (5:2, EtOAc:MeOH), ^{1}H NMR (600 MHz, D_2O) δ 3.36 (dd, 1H, *J* = 8.7 Hz), 3.46-3.50 (m, 1H), 3.56 (dd, 3H, *J* = 3.9, 9.6 Hz), 3.60-3.78 (m, 23H), 3.80-3.90 (m, 13H), 4.67 (d, 0.8H, *J* = 8.0 Hz), 5.25 (d, 0.8H, *J* = 3.7 Hz), 5.33 (d, 2H, *J* = 3.8 Hz), 5.35 (d, 1H, *J* = 3.9 Hz), 5.36 (d, 0.8H, *J* = 3.9 Hz). ^{13}C NMR (150 MHz, $CDCl_3$) δ 22.77, 59.88, 60.05, 60.99, 61.09, 68.90, 69.25, 69.54, 69.58, 69.64, 70.75, 71.11, 71.17, 72.27, 72.52, 75.17, 78.64, 78.66, 78.94, 81.41, 91.77, 95.50, 98.51, 98.65, 98.77. HRESIMS: calcd for $[M+Na]^+$ $C_{16}H_{28}O_{10}$ + Na: 467.1371; found: 467.1383.

D-xylose-*α*1-4-D-xylose-*α*1-3-D-glucopyranose (**4**) Rf = 0.2 (5:2, EtOAc:MeOH), ^{1}H NMR (600 MHz, D_2O) δ 3.34-3.40 (m, 3H) 3.49-3.6 (m, 5H), 3.61-3.80 (m, 14H), 3.84-3.91 (m, 5H), 3.99-4.08 (m, 2H), 4.70 (d, 0.5H, *J* = 7.8 Hz), 4.71 (d, 1H, *J* = 8.1 Hz), 5.15 (d, 0.5H, *J* = 3.8 Hz), 5.15 (d, 1H, *J* = 3.8 Hz),

5.27 (d, 0.5H, *J* = 3.8 Hz), 5.39 (d, 1H, *J* = 3.7 Hz). ^{13}C NMR (150 MHz, $CDCl_3$) δ 23.25, 60.35, 60.44, 60.52, 61.49, 65.27, 69.10, 69.24, 70.09, 70.16, 70.19, 71.24, 71.31, 71.57, 71.82, 71.86, 72.40, 72.76, 72.93, 75.55, 75.67, 78.19, 78.26, 79.06, 81.37, 92.26, 95.97, 98.53, 98.72, 99.68, 100.07. HRESIMS: calcd for $[M+Na]^+$ $C_{16}H_{28}O_{10}$ + Na: 467.1371; found: 467.1382.

4.6 Acknowledgements

We thank several researchers for their generosity; Prof. R. Cornell for the pAX142 expression vector, Prof. B. Pinto for access to the CE instrumentation, Prof. S. Withers for YicI protein and pET29aYICI, Prof. M. Moracci for generous provision of the XylS protein and pXyl. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) and by provision of UDP-Xylose from CarboSource Services (provision supported in part by NSF-RCN grant #0090281). The mass spectrometry equipment was obtained through generous grants from the Canadian Foundation for Innovation and the British Columbia Knowledge Foundation to D.J.V. G.E.W. is a scholarship holder from NSERC. D.J.V. is a Scholar of the Michael Smith Foundation for Health Research and a Canada Research Chair in Chemical Glycobiology.

5: Synthesis of D-Xyl-α1-3-D-Xyl-α1-3-D-Glc, D-Xyl-α1-4-D-Xyl-α1-3-D-Glc, D-Xyl-α1-2-D-Xyl-α1-3-D-Glc, and D-Xyl-α1-3-D-Glc, and developement of tools to study a non-canonical form of Notch protein *O*-glucosylation

Updates in the field relating to this Chapter can be found in appendix 2. The designated numbers given to synthetic or target compounds in this Chapter relate only to this Chapter.

5.1 Contributions

I synthesized and characterized all compounds and wrote the first draft of the manuscript.

5.2 Abstract

A facile and efficient synthesis of D-Xyl-*α*1-3-D-Xyl-*α*1-3-D-Glc and two linkage isomers: D-Xyl-*α*1-4-D-Xyl-*α*1-3-D-Glc and D-Xyl-*α*1-2-D-Xyl-*α*1-3-D-Glc. These trisaccharides were synthesized using a block wise methodology and were used to unequivocally characterize an *O*-glycan modifying mammalian Notch. The *O*-glycan was identified as D-Xyl-*α*1-3-D-Xyl-*α*1-3-D-Glc and is known to play a critical role in proper Notch signaling. However, the precise role that this trisaccharide plays in Notch signaling is not clearly defined. To address this issue we synthesized a series of antigens incorporating the XXG trisaccharide. Each antigen is functionalized with a terminal alkyne for copper-catalyzed azide-alkyne cycloaddition to the highly immunogenic bacteriophage Qβ capsid. Future development of XXG specific antibodies from these immunogenic glycoconjugates will be useful tools to aid in the functional study of XXG modification of Notch.

5.3 Introduction

Notch is a key cell surface protein receptor that is a vital component of the Notch signaling pathway. The Notch signaling pathway plays a fundamental role in early metazoan development, it activates the expression of many genes involved in cell differentiation and tissue morphogenesis[149,300]. Abnormal Notch signaling has been linked to a slew of developmental disorders[150,151,176] and has also been implicated in human diseases such as cancer[152]. Mammals have four Notch homologues (Notch1 to Notch4) that interact with three Delta-like ligands (Delta 1, 3, and 4) and two Jagged ligands (Jagged 1, and 2)[301-305]. The mammalian delta-like and jagged ligands, located on adjacent cells, initiate the Notch signaling pathway (Figure 5-1). Upon Ligand binding the conformation of the Notch regulatory region is altered[159], thus allowing cleavage to occur[160] and subsequent translocation of the Notch intracellular domain into the nucleus activating gene transcription[165,166].

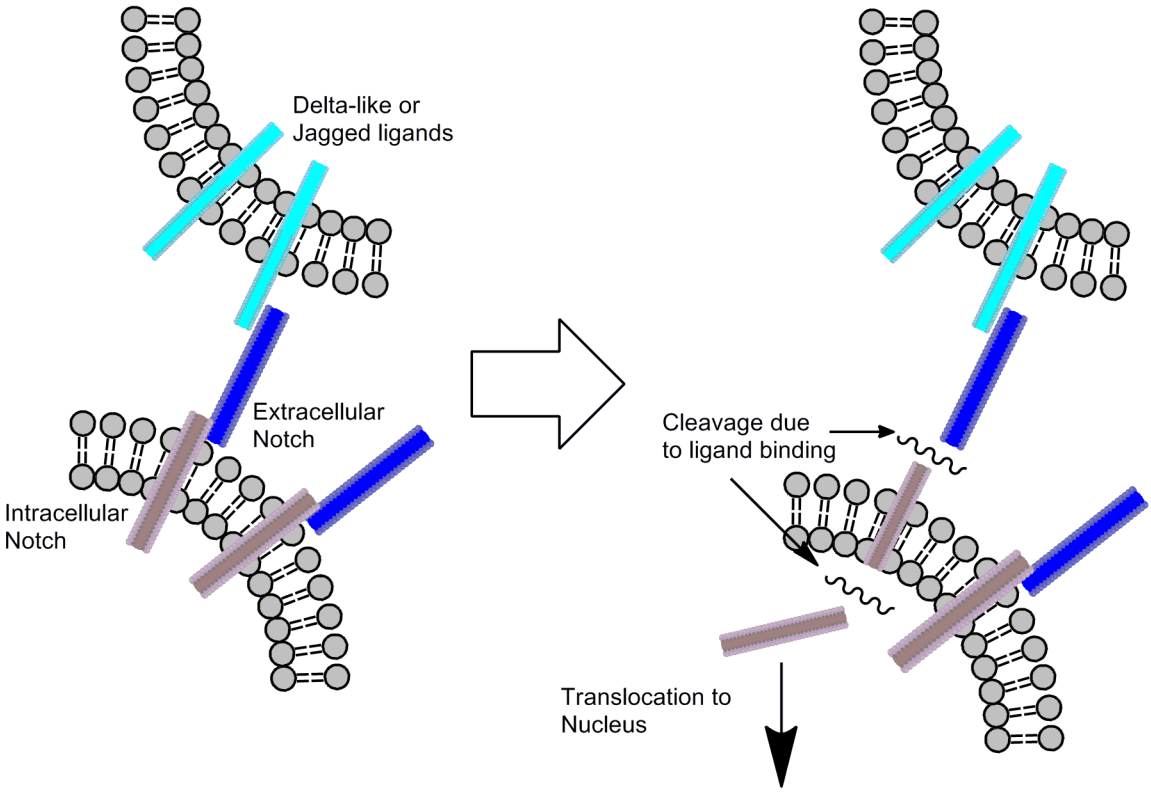


Figure 5.1: Simplified mammalian Notch signaling pathway

The Notch ECD (dark blue), is non-covalently attached to the transmembrane and intracellular domains of Notch (grey). A simplified Delta-like or Jagged ligand (light blue) on adjacent cell binds with the extracellular domain of Notch and initiates extracellular cleavage by ADAM10, a metalloprotease. After extracellular cleavage releases the Notch extracellular domain, presenilin/γ-secretase complex catalyzes the cleavage of the intracellular domain. The Notch intracellular domain is released from the membrane and transits to the nucleus to activate transcription.

Notch the key component of this signaling pathway, is a large transmembrane protein composed of three domains: an intracellular domain, a transmembrane domain, and an extracellular domain (ECD)[149]. The ECD is primarily composed of epidermal growth factor-like (EGF) repeats. These EGF repeats are modified with three different types of *O*-linked glycosylation: *O*-fucosylation, *O*-glucosylation, and *O*-GlcNAcylation[148,306,307]. In contrast to *O*-GlcNAcylation of Notch which is a recent discovery with very little known about the significance of its modification of Notch, *O*-fucosylation has pioneered the field of Notch *O*-glycosylation. Protein *O*-fucosyltransferase-1 (Pofut-1) adds fucose to Notch[167], while knockout of Pofut-1 in mice shows embryonic

lethality[308]. Furthermore, Fringe elongates the fucose moiety on Notch by the addition of GlcNAc, this disaccharide plays a role in the regulation of Notch ligand binding and inherently Notch signalling[12,122,309]. Discovery of the critical role that *O*-fucosylation of Notch, and the elongation of fucose by Fringe to the disaccharide, plays in regulation of ligand binding and proper Notch function has put an emphasis on the elucidation of the function of *O*-glycan modification of Notch and its ligands.

O-glucosylation is a non-canonical form of protein glycosylation which was first identified on EGF domains of bovine blood coagulation factors VII and IX[145]. The consensus sequence C^1xSxPC^2, where C^1 and C^2 refer to the first and second conserved cysteine residues of EGF repeats, was proposed by Nishimura *et al*. to govern the *O*-glucosylation of EGF repeats.[147] This consensus sequence was recently revised to allow alanine N-terminal to C^2: C^1xSx(A/P)C^2 [310]. In 2000, Moloney *et al*. found that Notch, which has this consensus sequence in many of its EGF repeats, was *O*-glucosylated.[148] The transferase responsible for adding this glucose moiety to Notch, was identified in *Drosophila* by Acar *et al*.[175] Loss of Rumi *O*-glucosyltransferase activity leads to a temperature sensitive Notch phenotype which points to glucosylation of Notch playing a critical role in proper Notch structure and function[166,175,176]. This is emphasized by recent studies which show embryonic lethality of mouse Rumi knockouts and implicate *O*-glucosylation of Notch in the regulation of Notch signaling at a stage between proper ligand binding and Notch cleavage[311,312]. However further exploration of this phenomenon is required to explicitly elucidate the function of Notch *O*-glucosylation.

Moloney *et al*. also discovered that the *O*-glucose modifying the ECD of mammalian Notch was elongated to a trisaccharide[148]. We recently identified the structure of that trisaccharide to be D-Xyl-*α*1-3-D-Xyl-*α*1-3-D-Glc (XXG, Figure 5-2)[306]. Two mammalian xylosyltransferases responsible for adding the first xylose unit, but not the second xylose unit, of the XXG trisaccharide to Notch, were identified by Sethi and coworkers[313]. With the defined structure of the trisaccharide determined, we decided to acquire polyclonal antibodies which

would act as useful tools to further elucidate the function of XXG modification of Notch. Unfortunately directing an immune response to carbohydrate based antigens has traditionally been difficult, often only a short-lived low affinity IgM antibody response is generated with no T-cell response. In fact it is the T-cell response which is required to produce high affinity IgG antibodies[314]. Recently Finn and co-workers successfully utilized a cowpea mosaic virus (CPMV) capsid as a carbohydrate carrier protein, and the glycoconjugate produced high total antibody titers and high levels of IgG antibody[119,315]. More recently the bacteriophage Qβ capsid conjugated to Man_4, Man_8, and Man_9, was used to elicit a strong immunogenic response to high mannose glycans[316]. This increased antibody production is most likely due to the CPMV and the bacteriophage Qβ capsids being highly immunogenic and having a patterned antigen display on their surface[119].

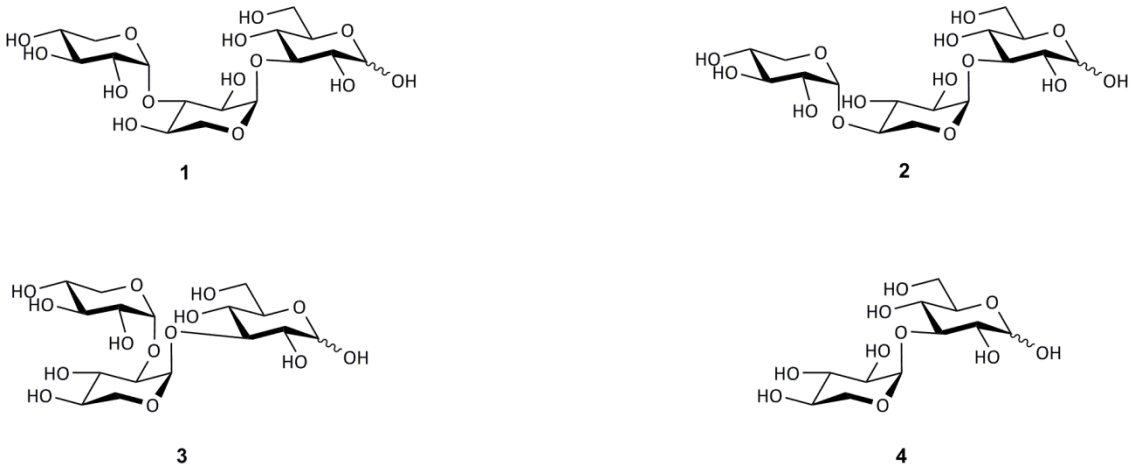
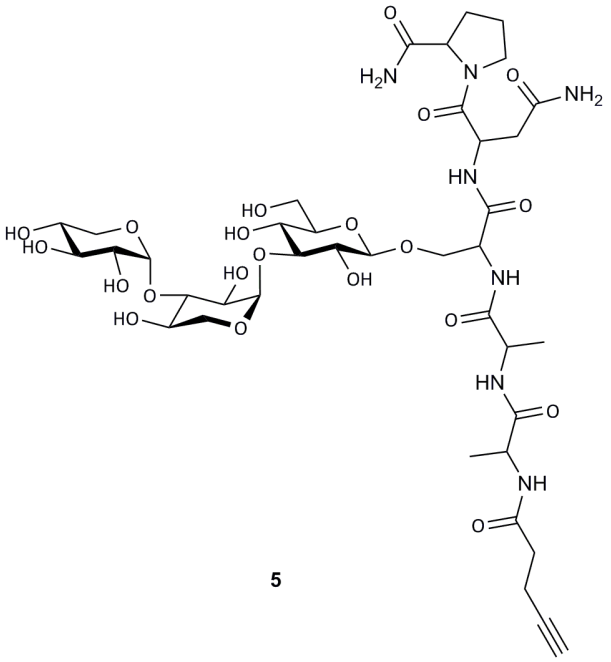


Figure 5.2: Series of synthetic di- and tri-saccharide standards

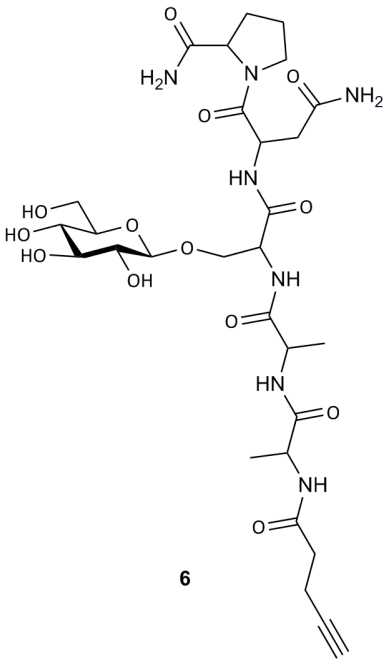
Chemically synthesized series of standards used to identify XXG modification of mammalian notch: XXG (**1**), D-Xyl-*α*1-4-D-Xyl-*α*1-3-D-Glc (X1-4XG, **2**), D-Xyl-*α*1-2-D-Xyl-*α*1-3-D-Glc (X1-2XG, **3**), and D-Xyl-*α*1-3-D-Glc (XG, **4**).

Here, we describe the synthesis of a series of trisaccharide standards that were used to elucidate the structure of the glucose containing trisaccharide modifying the ECD of mammalian Notch (Figure 5-2). We then incorporated the

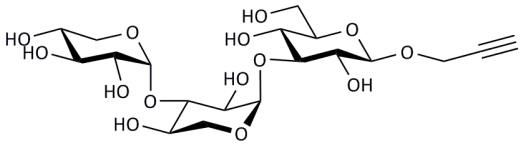
XXG trisaccharide into a peptide for use as an antigen (Figure 5-3) and coupled it to the Qβ capsid using copper-catalyzed azide-alkyne cycloaddition (CuAAC). Using this glycoconjugate we hope to develop high titers of XXG specific antibodies for use as tools to further clarify the role that XXG modification of Notch plays in the Notch signaling pathway.



5



6



7

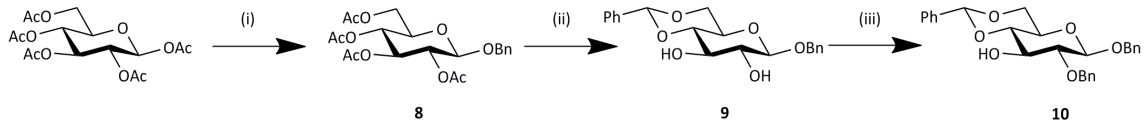
Figure 5.3: Series of antigens for development of selective antibodies

Three chemically synthesized antigens incorporating an alkyne functionality for CuAAC coupling to Qβ capsid: XXG-Pep (**5**), G-Pep (**6**), XXG-Alkyne (**7**).

5.4 Results

5.4.1 Synthesis of XXG, X(1-2)XG, X(1-4)XG and XG standards

The XXG trisaccharide (**1**), two linkage isomers: X(1-4)XG (**2**) and X(1-2)XG (**3**), and the D-Xyl-*α*1-3-D-Glc (XG, **4**) disaccharide were chemically synthesized for use as standards with CE (Figure 5-2). A block wise synthetic approach was used. The glucose building block was initially synthesized as methyl 2-*O*-benzoyl-4-6-*O*-benzylidene-*α*-D-glucopyranoside, upon completion of the fully protected XXG trisaccharide, acetolysis conditions used to convert the *α*-methyl anomeric protecting group to an acyl protecting group led to cleavage of the xyl-*α*1-3-xyl glycosidic bond (data not shown). The second attempt at synthesizing a suitable glucose building block was phenyl 2-*O*-benzoyl-4-6-*O*-benzylidene-1-thio-*β*-D-glucopyranoside. The C3-hydroxyl of this compound was not a good enough nucleophile at the low temperatures (-60 ºC) required to favour the formation of an *α*-glycosidic linkage with the donor sugar **19** (data not shown). To enhance the nucleophilicity of the C3-hydroxyl, benzyl groups were used to protect the anomeric hydroxyl and the C2-hydroxyl. This led to **10** as a suitable acceptor for **19** and allowed for a milder deprotection strategy to be used upon completion of the XXG precursor. **10** was prepared starting from 1,2,3,4,6-penta-*O*-acetyl-β-D-glucopyranose (Scheme 5-1). The anomeric position was benzylated using benzyl alcohol and boron trifluoroetherate to give the *β*-benzyl pyranoside **8**. Zemplén deprotection (NaOMe and MeOH) and subsequent *O*-benzylidination gave us **9,** using phase transfer conditions; tetrabutylammonium hydrogensulphate and benzyl bromide in a mixture of 1M NaOH and CH_2Cl_2 yielded an inseparable mixture of benzyl 2-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside and benzyl 3-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside in a 2 to 1 ratio respectively. The aforementioned mixture was acetylated, separated via flash column chromatography, and then Zemplén deprotected to give the desired product **10**.



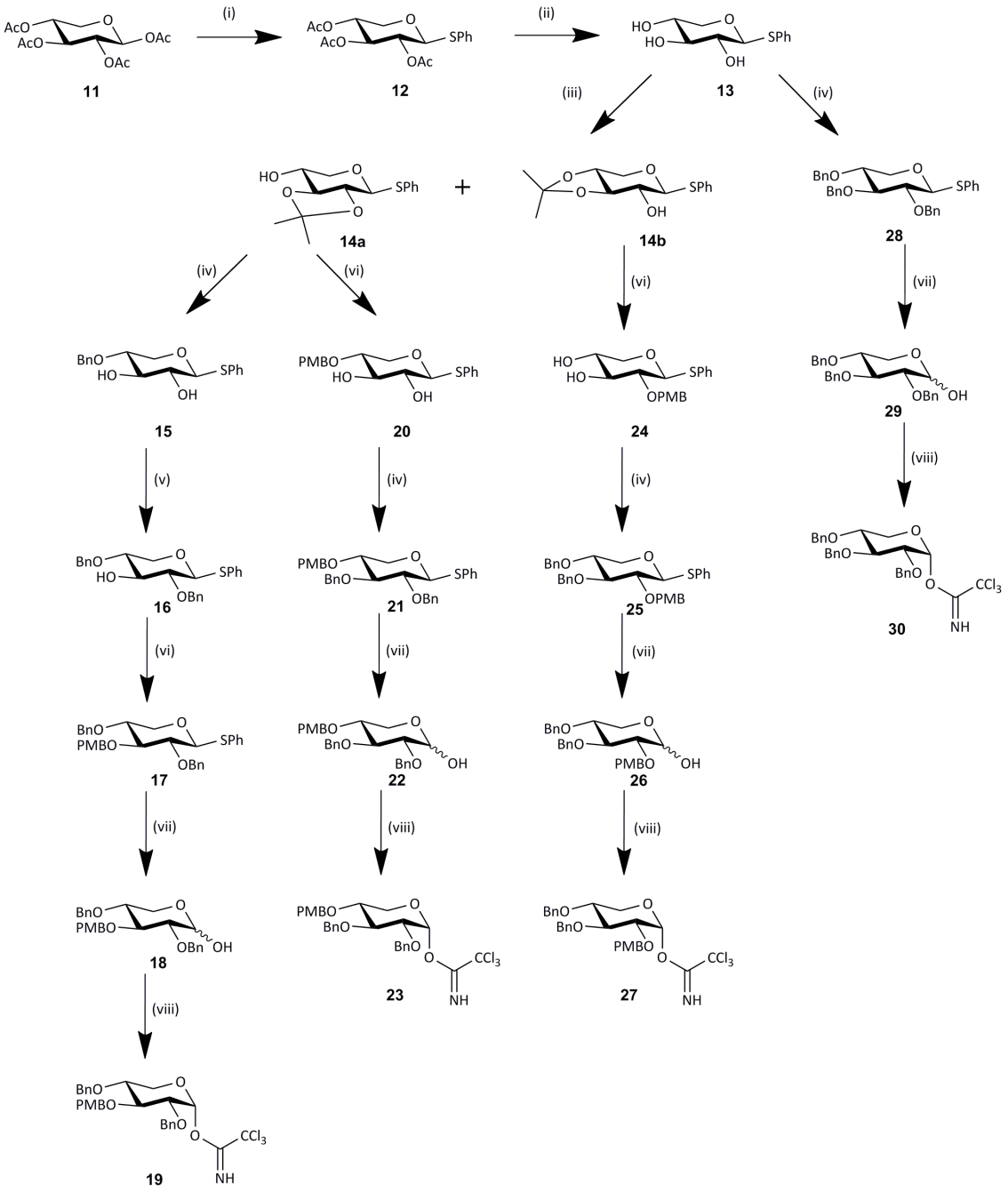
Scheme 5.1: Synthesis of selectively protected glucose building block

Reagents and conditions: (i) CH_2Cl_2, BnOH, BF_3OEt_2 (80% yield); (ii) a) NaOMe, MeOH (quantitative); b) benzaldehyde dimethyl acetal, CSA, acetonitrile (85% yield); (iii)a) Bu_4NHSO_4, BnBr, 5% NaOH in H_2O, CH_2Cl_2; b) Ac_2O, pyridine, DMAP; c) NaOMe, MeOH (55% 2-OBn, 20% 3-OBn yield over three steps).

The three different internal xylose building blocks: **19**, **23**, and **27** were prepared from xylose. Xylose was per-*O*-acetylated and then subsequently subjected to treatment with thiophenol and boron trifluoroetherate to give the thiophenyl pyranoside **12** (Scheme 5-2)[317]. **12** was then Zemplén deprotected to yield phenyl-1-thio-*β*-D-xylopyranoside[317]. Reacting this thioxyloside with 2-methoxypropene in DMF in the presence of HCl afforded the 2,3- and 3,4-acetonides. Benzylation and subsequent removal of the 2,3-acetonide using mild acid afforded the selectively 4-*O*-monobenzylated xylopyranoside **15**[318]. The key intermediate, 2,4-dibenzylated xylopyranoside **16**, was obtained using phase transfer conditions: tetrabutylammonium hydrogen sulphate and benzyl bromide in a mixture of 1 M NaOH and CH_2Cl_2. Initially phenyl (2,3,4-tri-*O*-benzyl-D-xylose)-(*α*1-3)-2,4-di-*O*-benzyl-1-thio-D-xylopyranoside was deemed a critical building block for the synthesis of XXG, unfortunately coupling **16** with **30** using Grundler and Schmidt glycosidic bond formation conditions[319] produced an inseperable mixture of α and β glycosidically linked disaccharide (data not shown). The synthetic building blocks were adjusted to target the suitably protected D-xylose-(*α*1-3)-D-glucopyranoside **31a**. **31a** incorporates *p*-methoxybenzyl protecting group which can be selectively removed in the presence of benzyl protecting groups. To achieve this **16** was *p*-methoxybenzylated to afford **17**.

The selectively removable *p*-methoxybenzyl protecting group was also incorporated into the synthesis of **23** and **27** via the following reactions (Scheme 5-2). *p*-Methoxybenzylation and subsequent removal of the 2,3-acetonide using

mild acid afforded the selectively 4-*O*-*p*-methoxybenzylated (*O*-PMB) xylopyranoside **20,** while *p*-methoxybenzylation and subsequent removal of the 3,4-acetonide using the aforementioned conditions gave the selectively 2-*O*-*p*-methoxybenzylated xylopyranoside **24**. Both 4-*O*-*p*-methoxybenzylated xylopyranoside **20** and 2-*O*-*p*-methoxybenzylated xylopyranoside **24** were dibenzylated using benzyl bromide and 60% NaH to give the other two key xylose intermediates **21** and **25** respectively. All three intermediates (**17**, **21**, and **25**) were reacted with NBS and H_2SO_4 in a water/acetonitrile solution to afford the corresponding hemiacetals[320]. Hemiacetals **18**, **22**, and **26** were converted to the desired trichloroacetimidate donors using trichloroacetonitrile, 1,8-diazabicyclo[5.4.0]-undec-7-ene in CH_2Cl_2 under N_2.



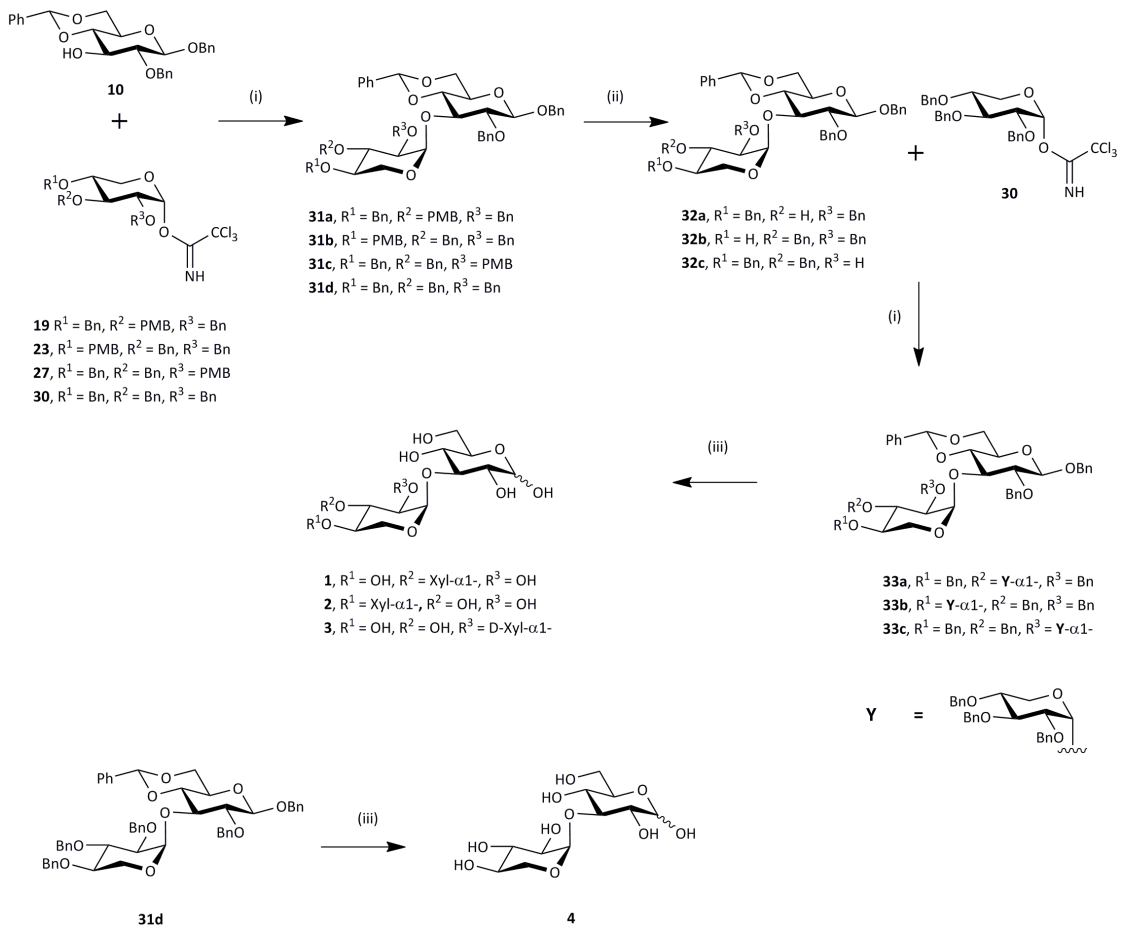
Scheme 5.2: Synthesis of xylose building blocks

Reagents and conditions: (i) CH_2Cl_2, PhSH, BF_3OEt_2 (80% yield); (ii) NaOMe, MeOH (quantitative); (iii) 2-methoxypropene, DMF, MeOH, AcCl (60% **14a**, 20% **14b** yield); (iv) a)BnBr, NaH (60%), DMF; b) MeOH, conc. HCl (quantitative); (v) Bu_4NHSO_4, BnBr, 5% NaOH in H_2O, CH_2Cl_2 (65% 2-OBn, 20% 3-OBn yield); (vi) *p*-methoxybenzyl bromide, NaH (60%), DMF (85% yield); (vii) *N*-bromosuccinimide, H_2O, ACN, conc. H_2SO_4 (90% yield); (viii) trichloroacetonitrile, DBU, CH_2Cl_2.

The third building block was synthesized from the deacetylated thiophenyl pyranoside **13**, per-*O*-benzylation gave us the benzylated thiophenyl pyranoside **28**, which was then treated with NBS and H_2SO_4 in a water/acetonitrile solution to afford the corresponding hemiacetal[320]. This hemiacetal was reacted with trichloroacetonitrile in the presence of 1,8-diazabicyclo[5.4.0]-undec-7-ene in CH_2Cl_2 under N_2 to yield the trichloroacetimidate building block **30** (Scheme 5-2)[321].

Initially disaccharide **31d** was synthesized by reacting **10** with trichloroacetimidate **30** using conditions described by Grundler and Schmidt[319]: trimethylsilyl trifluoromethanesulfonate, crushed 4 Å molecular sieves, and anhydrous diethyl ether, unfortunately **10** was not soluble in diethyl ether at temperatures lower than -15 ºC. Reactions attempted at or above -15 ºC led to a 1:1 mixture of α:β glycosidic bond formation. Greater solubility was obtained with CH_2Cl_2 as the solvent, therefore the reaction was carried out in anhydrous CH_2Cl_2 at -60 ºC in the presence of crushed 4 Å molecular sieves with trimethylsilyl trifluoromethanesulfonate acting as the Lewis acid. The α-glycosidic bond was formed in a 4:1 ratio over the β-glycosidic bond, thus giving disaccharide **31d** (Scheme 5-3). Using palladium on carbon, the benzyl and benzylidene protecting groups were hydrogenolyzed to obtain the desired XG disaccharide **4**.

Glycosylation of 10 using trichloroacetimidate donor **19**, **23**, or **27** under the modified glycosylation conditions described above, generated the α/β linkage mixture of the respective disaccharide intermediates (**31a**, **31b**, or **31c** respectively). Unfortunately these three intermediates were extremely non-polar and separation of the α/β linkage mixture was not possible. Fortunately selective deprotection of the *O*-PMB protecting group for each disaccharide, using 2,3-dichloro-5,6-dicyano-1,4-benzoquinone in a mixture of CH_2Cl_2 and H_2O, increased the polarity of each compound and allowed us to separate the α/β linkage mixtures using flash column chromatography, thus affording the three desired disaccharide intermediates **32a**, **32b**, or **32c**.



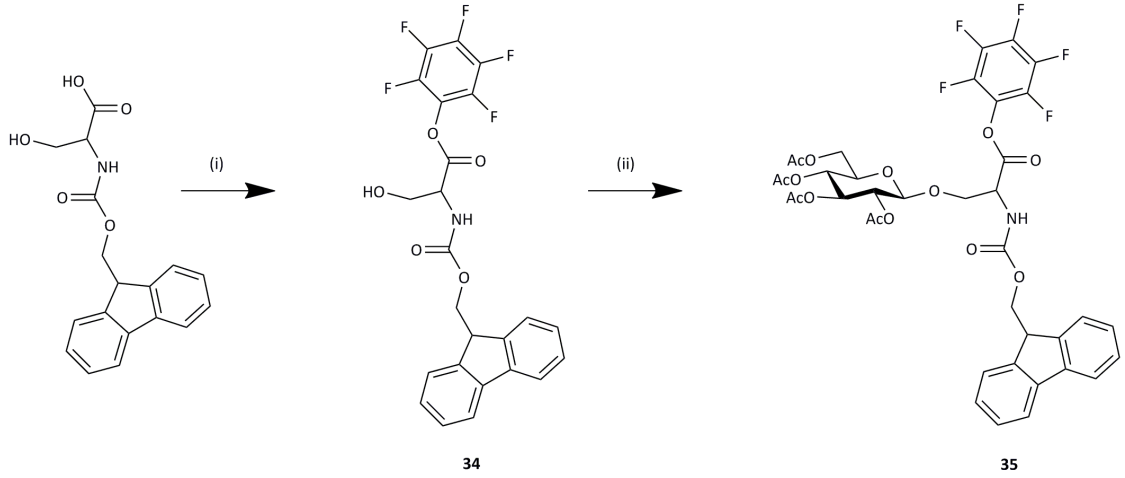
Scheme 5.3: Synthesis of di- and tri-saccharide standards

Reagents and conditions: (i) powdered 4 Å molecular sieves, TMSOTf, CH_2Cl_2; (ii) 2,3-dichloro-5,6-dicyanobenzoquinone, CH_2Cl_2:H_2O (18:1) (45-60% *α*-, 5-15% *β*- yield over two steps); (iii) H_2, 10% Pd/C, MeOH (75-90% yield).

By reacting the disaccharide intermediates **32a**, **32b**, or **32c** with trichloroacetimidate **30** using the Grundler and Schmidt[319] conditions described above, we were able to synthesize the fully protected trisaccharide linkage isomers **33a**, **33b**, and **33c** respectively. Subsequent removal of the benzyl and benzylidene protecting groups by hydrogenolysis using Pd/C, and H_2 in dry methanol gave us the three fully deprotected trisaccharides XXG (**1**), X(1-4)XG (**2**) and X(1-2)XG (**3**).

5.4.2 Synthesis of G-Pep (6), XXG-Pep (5), and XXG-Alkyne (7) for use as antigens

N-(9-Fluorenylmethoxycarbonyl)-L-Serine pentafluorophenol ester was synthesized starting from Fmoc-L-Serine and pentafluorophenol, dicyclohexylcarbodiimide (DCC) was used as the coupling reagent to give the pentafluorophenol ester of Fmoc-L-Ser (Fmoc-L-Ser-OPfp, **34**, Scheme 5-4)[322]. 1,2,3,4,6-penta-O-acetyl-β-D-glucopyranose was coupled to **34** as described by Katajisto *et al*. to give rise to **35** in good yields (Scheme 5-4)[323].



34 35

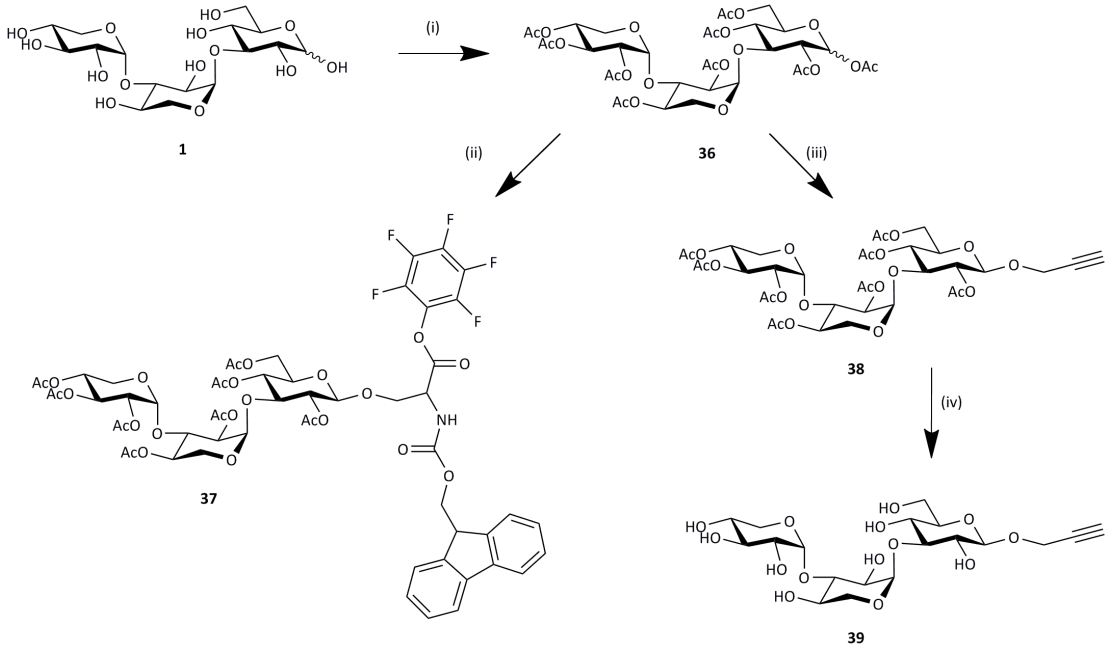
Scheme 5.4: Synthesis of G-Pep

Reagents and conditions: (i) dicyclohexylcarbodiimide, EtOAc, pentafluorophenol, 0 ºC (80% yield); (ii) BF_3OEt_2, 1,2,3,4,6-penta-O-acetyl-β-D-glucopyranose, CH_2Cl_2 (50% yield).

XXG was acetylated to yield **36** in an α:β anomeric ratio of 5:1 respectively. Boron trifluoroetherate at a reaction temperature of -10 ºC was used to form the *O*-glycosidic linkage between **34** and a mixture of *α*-*O*-acetyl (2,3,4-tri-*O*-acetyl-D-xylose)-(*α*1-3)-(2,4-di-*O*-acetyl-D-xylose)-(*α*1-3)-2,4,6-tri-*O*-acetyl)-D-glucopyranoside and *β*-*O*-acetyl (2,3,4-tri-*O*-acetyl-D-xylose)-(*α*1-3)-(2,4-di-*O*-acetyl-D-xylose)-(*α*1-3)-2,4,6-tri-*O*-acetyl)-D-glucopyranoside (**36**) thus producing the desired trisaccharide serine building block **37** (Scheme 5-5). Poor yields were achieved for this reaction due to the formation of a glycosidic linkage between only the β-anomer of **36** (data not shown). Hydrazine hydrate was used to

deprotect the anomeric acetate of **36**, subsequent acetylation regenerated the β-anomer, albeit in the poor α:β anomeric ratio of 5:1. The aforementioned Boron trifluoroetherate catalyzed coupling reaction between **36** and **34** was attempted again and all newly synthesized **37** was combined with the previous purified supply of **37**. The anomeric deprotection, acetylation, and coupling cycle described above, was repeated 3 times to increase the yield of **37**.

For the solid phase peptide synthesis a preloaded Fmoc-Pro-Rink amide AM resin was used and standard coupling procedures were followed[324]. Incorporation of (Fmoc-Ser(Ac-Glc)-*O*-Pfp) or (Fmoc-Ser(Ac-XXG)-*O*-Pfp) into the peptide was accomplished using 0.8 equivalents of (Fmoc-Ser(Ac-Glc)-*O*-Pfp) and a coupling time of 180 min instead of the standard, 6.0 equivalents of commercially available amino acids and a 20-60 min coupling time. An N-terminal alkyne was attached to the peptide, using 4-pentynoic acid, HBTU and DIEA, for CuAAC coupling of the peptide to the Qβ capsid. The Glc and XXG modified peptides were de-acetylated using NaOMe in methanol at a pH of ~10 for 36 h. The resin was cleaved from the peptide along with the remaining acid sensitive protecting groups using TFA in the presence of scavengers, EDT and triethylsilane. An Agilent HPLC equipped with a C8 column for separation was used to purify the crude peptides to give the desired antigens **5** and **6** (Figure 5-3) in > 90% purity as judged by HPLC.



Scheme 5.5: Synthesis of XXG-Alkyne and XXG-Ser building block for peptide synthesis

Reagents and conditions: (i) Ac_2O, pyridine, DMAP (70% *α*/*β* mix yield); (ii) powdered 4 Å molecular sieves, Fmoc-Ser-OPfp, BF_3OEt_2, CH_2Cl_2 (40% yield); (iii) BF_3OEt_2, propargyl alcohol, CH_2Cl_2 (60% yield); (iv) NaOMe, MeOH (quantitative).

XXG-Alkyne was synthesized from the aforementioned acetylated XXG mixture (**36**) and propargyl alcohol, the starting materials were dissolved in DCM and cooled to -20 ºC, adding BF_3OEt_2 gave compound **38**. The acetylated alkyne, **38** was Zemplén deprotected to give the desired XXG-Alkyne (**39**) antigen (Scheme 5-5).

5.5 Conclusion

We have efficiently prepared the three linkage isomers XXG, X(1-4)XG, and X(1-2)XG and the XG disaccharide using a block wise synthetic approach. XXG, X(1-4)XG, X(1-2)XG, and XG were used as standards to fully elucidate the structure of the trisaccharide modifying mammalian Notch[306]. With the identity of the Notch trisaccharide revealed, XXG was used as a synthetic platform for the development of tools which could be used to clarify the role of mammalian Notch

O-glucosylation. **36** was carefully coupled to **34** to yield **37**, which was subsequently incorporated into peptide synthesis. Due to low quantities of **37** suboptimal equivalents of **37** (0.8 equivalents) were used in peptide synthesis. Deprotection and purification of the peptide incorporating XXG afforded the desired antigen **5**. G-Pep (**6**) was synthesized in the same manor, while XXG-Alkyne (**7**) was efficiently prepared from **36**, thus rounding out the development of the three target antigens (**5**, **6**, and **7**).

5.6 Materials and methods

5.6.1 General

All solvents were dried prior to use. Synthetic reactions were monitored by TLC using Merck Kieselgel 60 F_{254} aluminium-backed sheets. Compounds were detected by charring with a 10% concentrated sulfuric acid/ethanol solution and heating. Flash chromatography under a positive pressure was performed with Merck Kieselgel 60 (230-400 mesh) using the specified eluants. ^{1}H and ^{13}C NMR spectra were recorded on a Varian AS500 Unity Innova spectrometer at 500 MHz (125 MHz for ^{13}C) with chemical shifts quoted relative to $CDCl_3$, D_2O, or CD_3OD where appropriate. For the peptide chemistry the rink amide resin, commercially available amino acid derivatives, and reagents were purchased from Chem-Impex International Inc. (IL, USA). All other chemicals were purchased from Sigma-Aldrich Co. (MO, USA) unless otherwise mentioned.

5.6.2 Synthesis of Benzyl 2-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside (10)

Benzyl 2,3,4,6-tetra-*O*-acetyl-1-*O*-*β*-D-glucopyranoside (**8**)

1,2,3,4,6-penta-O-acetyl-β-D-glucopyranose (17.0 g, 44.0 mmol) was dissolved in CH_2Cl_2 (dry, 80 mL) and benzyl alcohol (6.8 mL, 64.6 mmol) and BF_3OEt_2 (14 mL) was added. Then the resulting mixture was stirred at room temperature for 14 h. The solution was diluted with $NaHCO_3$ (60 mL) and extracted with CH_2Cl_2 (3 x 60 mL), the combined extracts were then washed with

brine (60 mL), dried ($MgSO_4$), and concentrated. Flash chromatography (EtOAc:hexanes 2:3) gave benzyl 2,3,4,6-tetra-*O*-acetyl-1-*O*-*β*-D-glucopyranoside as a white powder (15.5 g, 80% yield).

Benzyl 4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside (**9**)

To a stirred solution of benzyl 2,3,4,6-tetra-*O*-acetyl-1-*O*-*β*-D-glucopyranoside (15.5 g, 35.4 mmol) in methanol (dry, 80 mL) was added NaOMe (catalytic). The mixture was stirred for 4 h and then was neutralized with methanol:AcOH (19:1) and stirred for a further 30 min. The mixture was concentrated and crystallized using hot ethanol to afford benzyl 1-*O*-*β*-D-glucopyranoside (quantitative). ^{1}H and ^{13}C NMR spectra were in accordance with published data[325]. ^{1}H NMR (500 MHz, CD_3OD) δ 7.20-7.50 (m, 5H), 4.94 (d, 1H, *J* = 11.5 Hz), 4.61 (d, 1H, *J* = 11.5 Hz), 4.42 (d, 1H, *J* = 7.7 Hz), 3.90 (d, 1H, *J* = 2.1 Hz, *J* = 10.4 Hz), 3.62 (d, 1H, *J* = 5.0 Hz, *J* = 10.4 Hz), 3.30-3.45 (m, 4H).

Benzyl 1-*O*-*β*-D-glucopyranoside (1.8 g, 6.8 mmol) was dissolved in acetonitrile (20 mL) and benzaldehyde (1.1 mL, 10.5 mmol) was added. Camphor sulphonic acid was added and the mixture was stirred at room temperature for 5 h. Et_3N was added upon completion, the reaction was then concentrated, redissolved in toluene (5 mL) and loaded onto a flash chromatography column (EtOAc:hexanes:Et_3N, 50:49.5:0.5), which gave benzyl 4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside as a white solid (2.0 g, 84 % yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.50-7.54 (m, 2H, Ar), 7.34-7.44 (m, 8H, Ar), 5.58 (s, 1H, PhCH), 4.97 (d, 1H, *J* = 11.5 Hz, $BnCH_2$), 4.67 (d, 1H, *J* = 11.5 Hz, $BnCH_2$), 4.54 (d, 1H, *J* = 7.7 Hz, H1), 4.40 (dd, 1H, *J* = 5.0 Hz, *J* = 10.4 Hz, H6), 3.85 (m, 2H, H3, H6'), 3.61 (m, 2H, H2, H4), 3.50 (ddd, 1H, *J* = 5.0 Hz, *J* = 9.7 Hz, H5). ^{13}C NMR (150 MHz, $CDCl_3$) δ 136.93, 136.68, 129.33, 128.62, 128.38, 128.25, 128.22, 126.30, 102.12, 101.98, 80.58, 74.60, 73.14, 71.57, 68.69, 66.48. HRESIMS: calcd for $[M+H]^+$ $C_{20}H_{23}O_6$ + H: 359.1489; found: 359.1502.

Benzyl 2-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside (**10**)

Benzyl 4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside (2.0 g, 5.6 mmol) was dissolved in CH_2Cl_2 (20 mL), benzyl bromide (0.67 mL, 5.6 mmol), 5 % aqueous NaOH (5 mL) and tetrabutyl ammonium hydrogen sulphate (.47 g, 1.4 mmol) were added and the solution was stirred vigorously at room temperature for 16 h. Upon completion, the reaction was diluted with CH_2Cl_2 (100 ml), the organic layer was extracted and washed with water (10 mL), brine (10 mL), dried over $MgSO_4$ and concentrated by rotary evaporation. The residue was then dissolved in toluene and loaded onto a flash column chromatography column (EtOAc:Toluene, 1:19) which produced a inseparable mixture of benzyl 2-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside (represented as 2-O-Bn in NMR data below) and benzyl 3-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside (represented as 3-O-Bn in NMR data below) in a 2 to 1 ratio respectively (1.85 g, 74 % yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.58-7.62 (m, 3H, Ar), 7.36-7.52 (m, 19.5H, Ar), 5.65 (s, 0.5H, PhCH-3-O-Bn), 5.587 (s, 1H, PhCH (3-O-Bn)), 4.98-5.07 (m, 3H, $BnCH_2$ (3-O-Bn and 2-O-Bn)), 4.91 (d, 0.5H, *J* = 11.8 Hz, $BnCH_2$ (3-O-Bn)), 4.85 (d, 1H, *J* = 11.3 Hz, $BnCH_2$ (2-O-Bn)), 4.74 (d, 1.0H, *J* = 11.7 Hz, $BnCH_2$ (2-O-Bn)), 4.73 (d, 0.5H, *J* = 11.5 Hz, $BnCH_2$ (3-O-Bn)), 4.69 (d, 1H, *J* = 7.7 Hz, H1 (2-O-Bn)), 4.55-4.57 (m, 0.5H, H1 (3-O-Bn)), 4.45 (dd, 0.5H, *J* = 4.9 Hz, *J* = 10.4 Hz, H6 (3-O-Bn)), 4.44 (dd, 1H, *J* = 4.9 Hz, *J* = 10.4 Hz, H6 (2-O-Bn)), 3.83-3.92 (m, 2.5H, H3 (2-O-Bn), H5 (2-O-Bn), H5 (3-O-Bn)), 3.72-3.80 (m, 1.5H, H2 (3-O-Bn), H3 (3-O-Bn), H4 (3-O-Bn)), 3.62 (dd, 1H, *J* = 9.4 Hz, H4 (2-O-Bn)), 3.43-3.55 (m, 2.5H, H2 (2-O-Bn), H6' (2-O-Bn), H6' (3-O-Bn)), 2.93 (s, 1H, OH (2-O-Bn)), 2.86 (s, 0.5H, (3-O-Bn)). ^{13}C NMR (150 MHz, $CDCl_3$) δ 138.58, 138.38, 137.45, 137.24, 137.19, 137.06, 129.28, 129.09, 128.63, 128.60, 128.56, 128.49, 128.40, 128.36, 128.21, 128.16, 128.15, 128.10, 128.09, 127.95, 127.82, 126.46, 126.18, 102.85, 102.42, 101.83, 101.33, 82.02, 81.40, 80.60, 80.40, 74.95, 74.64, 74.45, 73.30, 71.58, 71.38, 68.81, 66.51, 66.22.

To separate the aforementioned mixture, benzyl 2-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside and benzyl 3-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside (1.6 g total, 3.6 mmol) were dissolved in pyridine (10 mL), Ac_2O (3 mL) and DMAP (catalytic) were added and the solution was stirred for 4 h at room temperature, methanol (5 mL) was added and the reaction was stirred for a further 30 min. The reaction was concentrated and loaded onto a column (EtOAc:toluene gradient 1:49 to 3:47) which afforded the acetylated desired compound benzyl 3-*O*-acetyl-2-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside as a white solid (1.0 g, 2.0 mmol). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.28-7.48 (m, 15H, Ar), 5.50 (s, 1H, PhCH), 5.34 (dd, 1H, *J* = 9.7 Hz, H3), 5.00 (d, 1H, *J* = 11.8 Hz, $BnCH_2$), 4.90 (d, 1H, *J* = 11.7 Hz, $BnCH_2$), 4.76 (d, 1H, *J* = 7.6 Hz, H1), 4.73 (d, 1H, *J* = 11.7 Hz, $BnCH_2$), 4.68 (d, 1H, *J* = 11.7 Hz, $BnCH_2$), 4.42 (dd, 1H, *J* = 4.9 Hz, *J* = 10.4 Hz, H6), 3.83 (d, 1H, *J* = 10.2 Hz, H2), 3.63 (d, 1H, *J* = 9.6 Hz, H4), 3.52-3.58 (m, 2H, H5, H6'). ^{13}C NMR (125 MHz, $CDCl_3$) δ 169.95, 138.01, 136.97, 136.89, 129.09, 128.57, 128.39, 128.27, 128.14, 128.10, 128.04, 127.79, 126.18, 103.13, 101.40, 79.72, 78.74, 74.58, 72.56, 71.79, 68.75, 66.20, 20.97. HRESIMS: calcd for $[M+NH_4]^+$ $C_{29}H_{34}NO_7 + NH_4$: 508.2330; found: 508.2337.

Deacetylation of benzyl 3-*O*-acetyl-2-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside (1.0 g, 2.0 mmol) using NaOMe (catalytic) all dissolved in MeOH (dry, 10 mL), stirred for 2 h, neutralized with methanol:AcOH (19:1), concentrated and loaded onto a flash chromatography column (Toluene:EtOAc, 19:1), gave the desired benzyl 2-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside (**10**) in good yield (0.9 g, 97 % yield) as a white solid. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.49-7.54 (m, 2H, Ar), 7.30-7.42 (m, 13H, Ar), 5.56 (s, 1H, PhCH), 4.98 (d, 2H, *J* = 11.2 Hz, $BnCH_2$), 4.76 (d, 2H, *J* = 11.3 Hz, $BnCH_2$), 4.72 (d, 2H, *J* = 11.6 Hz, $BnCH_2$), 4.67 (d, 2H, *J* = 7.7 Hz, H1), 4.40 (dd, 2H, *J* = 5.0 Hz, *J* = 10.5 Hz H6), 3.87 (d, 1H, *J* = 9.2 Hz, H3), 3.83 (d, 1H, *J* = 10.2 Hz, H5), 3.60 (d, 1H, *J* = 9.3 Hz, H4), 3.43-3.52 (m, 2H, H2, H6'), 2.49 (s, 1H, OH). ^{13}C NMR (125 MHz, $CDCl_3$) δ 138.20, 137.07, 137.03, 129.24, 128.54, 128.35, 128.16, 128.04,

127.96, 126.34, 102.78, 101.83, 81.09, 80.48, 74.90, 73.24, 71.58, 68.78, 66.17. HRESIMS: calcd for $[M+Na]^+$ $C_{27}H_{28}O_6$ + Na: 471.1778; found: 471.1794.

5.6.3 Synthesis of 2,4-di-*O*-benzyl-3-*p*-methoxybenzyl-D-xylopyranosyl trichloroacetimidate (19), 2,3-di-*O*-benzyl-4-*p*-methoxybenzyl-D-xylopyranosyl trichloroacetimidate (23), and 3,4-di-*O*-benzyl-2-*p*-methoxybenzyl-D-xylopyranosyl trichloroacetimidate (27)

Phenyl 2,3,4-tri-*O*-acetyl-1-thio-*β*-D-xylopyranoside (**12**)

To a stirred solution of D-xylose (15 g, 99 mmol) in Ac_2O (80 mL) was added NaOAc (.6 g, 7.3 mmol). The reaction was heated and maintained at reflux for 10 min, then allowed to cool to room temperature. Ice (200 g) was added and the reaction was stirred for 14 h. The reaction was diluted with 450 mL of EtOAc and washed with H_2O (3 x 30 mL), $NaHCO_3$ (30 mL), and brine (30 mL). The organic extracts were dried ($MgSO_4$) and concentrated, after which minimal amounts of boiling ethanol was used to crystallize, presumably, the 1,2,3,4-tetra-*O*-acetyl-*β*-D-xylopyranose (**11**) as a white crystalline compound (16.50 g, 52% yield).

The tetracetate from above (16.0 g, 50 mmol) was dissolved in CH_2Cl_2 (dry, 60 mL) and thiophenol (5.4 mL, 53 mmol) and BF_3OEt_2 (3 mL) was added. Then the resulting mixture was stirred at room temperature for 14 h. The solution was diluted with $NaHCO_3$ (60 mL) and extracted with CH_2Cl_2 (3 x 60 mL), the combined extracts were then washed with brine (60 mL), dried ($MgSO_4$), and concentrated. Flash chromatography (EtOAc:hexanes 2:3) gave phenyl 2,3,4-tri-*O*-acetyl-1-thio-*β*-D-xylopyranoside (**12**) as a white powder (16.7 g, 90% yield).

Phenyl 1-thio-*β*-D-xylopyranoside (**13**)

To a stirred solution of phenyl 2,3,4-tri-*O*-acetyl-1-thio-*β*-D-xylopyranoside (16.7 g, 45 mmol) in methanol (dry) was added NaOMe (catalytic). The mixture was allowed to stir for 4 h and was then neutralized with methanol:AcOH (19:1)

and stirred for a further 30 min. The mixture was concentrated and crystallized using hot ethanol to afford phenyl 1-thio-*β*-D-xylopyranoside (quantitative). Characterization was in agreement with Stick *et al*.[318] ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.52-7.55 (m, 2H, Ar), 7.27-7.34 (m, 3H, Ar) 4.58 (d, 1H, *J* = 9.4 Hz, H1), 3.96 (dd, 1H, *J* = 5.2 Hz, *J* = 11.2 Hz, H6), 3.49 (ddd, 1H, *J* = 5.2 Hz, *J* = 8.8 Hz, *J* = 10.1 Hz, H5), 3.37 (dd, 1H, *J* = 8.6 Hz, H3), 3.25 (dd, 1H, *J* = 10.1 Hz, *J* = 11.2 Hz, H6'), 3.23 (dd, 1H, *J* = 8.8 Hz, H4). HRESIMS: calcd for $[M+Na]^+$ $C_{11}H_{14}O_4S$ + Na: 265.0505; found: 265.0500.

Phenyl 4-*O*-benzyl-1-thio-*β*-D-xylopyranoside (**15**)

Phenyl 1-thio-*β*-D-xylopyranoside (10.0 g, 41.4 mmol) was dissolved in DMF (dry, 20 mL) and cooled to 0ºC. Acetyl chloride (1.4 mL) was added dropwise to methanol (5 mL) and cooled to 0ºC. The acetyl chloride/methanol solution was added dropwise to phenyl 1-thio-*β*-D-xylopyranoside in DMF until the solution went acidic. 2-Methoxypropene (11.0 mL, 115.8 mmol) was added and the reaction was stirred for 1 h. The reaction was quenched with triethylamine, concentrated, and flash column chromatography afforded the desired compounds **14a** and **14b** (7.0 g and 2.4 g respectively) which were used without further purification.

Compound **14a** (3.6 g, 13.0 mmol) was dissolved in DMF (dry, 30 mL), benzyl bromide (1.6 mL, 14.1 mmol) was added and the reaction was stirred at room temperature under an atmosphere of N_2. NaH (60 %, 0.56 g, 14.1 mmol) was added and the reaction was stirred for another hour. Upon completion, the reaction was quenched with methanol and allowed to stir for 30 min, after this time the reaction was concentrated under reduced pressure, taken up in EtOAc (150 mL), and washed with H_2O (30 mL) and brine (30 mL), dried over $MgSO_4$, concentrated and used without further purification.

The intermediate from above was dissolved in methanol (45 mL), concentrated HCl (450 µL) was added and the reaction was stirred until judged complete by TLC. Upon completion, the reaction was quenched with triethylamine, concentrated and crystallized using a mixture of CH_2Cl_2:hexanes. Phenyl 4-*O*-benzyl-1-thio-*β*-D-xylopyranoside was recovered as a white solid (3.6 g, 53 % yield over three steps)[318]. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.51-7.55 (m, 2H, Ar), 7.30-7.40 (m, 8H, Ar), 4.70 (d, 1H, *J* = 11.7 Hz, $BnCH_2$), 4.65 (d, 1H, *J* = 11.8 Hz, $BnCH_2$), 4.56 (d, 1H, *J* = 9.0 Hz, H1), 4.11 (dd, 1H, *J* = 4.8, 11.6 Hz, H5), 3.70 (dd, 1H, *J* = 8.4 Hz, H2), 3.50 (ddd, 1H, *J* = 4.6 Hz, *J* = 8.8 Hz, *J* = 10.0 Hz, H4), 3.41 (ddd, 1H, *J* = 2.7 Hz, *J* = 8.6 Hz, H3), 3.30 (dd, 1H, *J* = 10.0, 11.0 Hz, H5'), 2.77 (s, 2H, OH). HRESIMS: calcd for $[M+Na]^+$ $C_{18}H_{20}O_4S$ + Na: 355.0975; found: 355.0965.

Phenyl 2,4-di-*O*-benzyl-1-thio-*β*-D-xylopyranoside (**16**)

Phenyl 4-*O*-benzyl-1-thio-*β*-D-xylopyranoside (3.6 g, 10.9 mmol) was dissolved in CH_2Cl_2 (60 mL). Benzyl bromide (1.3 mL, 10.9 mmol), tetrabutylammonium hydrogen sulphate (0.74 g, 2.2 mmol) and aqueous NaOH (5 %, 30 mL) were added to the solution and the reaction was stirred vigorously at room temperature for 16 h. Once the reaction was judged complete by TLC, it was diluted with CH_2Cl_2 (180 mL), washed successively with H_2O (30 mL) and brine (30 mL), dried over $MgSO_4$, and concentrated. Flash column chromatography (EtOAc:hexanes, 1:9) afforded the desired compound, phenyl 2,4-di-*O*-benzyl-1-thio-*β*-D-xylopyranoside as a white solid (3.0 g, 65 % yield)[326]. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.57 (d, 2H, *J* = 6.8 Hz, Ar), 7.46 (d, 2H, *J* = 7.3 Hz, Ar), 7.26-7.39 (m, 11H, Ar), 4.94 (d, 1H, *J* = 10.9 Hz, $BnCH_2$), 4.77 (d, 1H, *J* = 10.9 Hz, $BnCH_2$), 4.71 (d, 1H, *J* = 11.8 Hz, $BnCH_2$), 4.65 (d, 1H, *J* = 11.4 Hz, $BnCH_2$), 4.64 (d, 1H, *J* = 9.8 Hz, H1), 4.05 (d, 1H, *J* = 5.2, 11.5 Hz, H5), 3.74 (ddd, 1H, *J* = 2.1 Hz, *J* = 8.9 Hz, H3), 3.52 (ddd, 1H, *J* = 5.1, *J* = 9.9, *J* = 14.1 Hz, H4), 3.35 (dd, 1H, *J* = 9.4 Hz, H2), 3.22 (dd, 1H, *J* = 10.7 Hz, H5'), 2.59 (d, 1H, *J* = 2.1 Hz, OH). ^{13}C NMR (125 MHz, $CDCl_3$) δ 138.12, 138.07, 133.58, 131.94,

129.03, 128.60, 128.57, 128.22, 128.05, 128.03, 127.86, 127.66, 88.08, 80.47, 77.86, 77.19, 75.24, 73.12, 67.50. HRESIMS: calcd for $[M+Na]^+$ $C_{25}H_{26}O_4S$ + Na: 445.1449; found: 445.1448.

Phenyl 2,4-di-*O*-benzyl-3-*p*-methoxybenzyl-1-thio-*β*-D-xylopyranoside (**17**)

Phenyl 2,4-di-*O*-benzyl-1-thio-*β*-D-xylopyranoside (2.5 g, 5.9 mmol) was dissolved in DMF (dry, 30 mL), *p*-methoxybenzyl bromide (1.0 mL, 7.1 mmol) was added and the solution was stirred at room temperature. To this stirred solution, 60% NaH (0.28 g, 7.1 mmol) was added and the reaction was stirred for 3 h. TLC was used to ascertain that the reaction was complete, it was then quenched with methanol and allowed to stir for 30 min. Then the reaction was concentrated under reduced pressure, taken up in EtOAc (150 mL), and washed with H_2O (30 mL), brine (30 mL), and dried over $MgSO_4$. Flash column chromatography (EtOAc:Hexanes, 1:9) was used to isolate phenyl 2,4-di-*O*-benzyl-3-*p*-methoxybenzyl-1-thio-*β*-D-xylopyranoside (3.0 g, 93 %yield) as a colourless syrup. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.56 (d, 2H, *J* = 7.6 Hz, Ar), 7.44 (d, 2H, *J* = 7.4 Hz, Ar), 7.25-7.43 (m, 13H, Ar), 6.88 (d, 2H, *J* = 8.5 Hz, Ar), 4.84-4.92 (m, 2H, $BnCH_2$), 4.74-4.82 (m, 3H, $BnCH_2$), 4.71 (d, 1H, *J* = 9.4 Hz, H1), 4.66 (d, 1H, *J* = 11.6 Hz, $BnCH_2$), 4.08 (dd, 1H, *J* = 4.5 Hz, *J* = 11.2 Hz, H5), 3.83 (s, 3H, OMe), 3.62-3.70 (m, 2H, H3, H5'), 3.46 (dd, 1H, *J* = 9.0 Hz, H5), 3.28 (dd, 1H, *J* = 11.2 Hz, *J* = 9.5 Hz, H5). ^{13}C NMR (150 MHz, $CDCl_3$) δ 159.30, 138.16, 138.14, 131.95, 130.63, 129.69, 128.98, 128.53, 128.44, 128.17, 127.93, 127.85, 127.57, 113.86, 88.41, 85.03, 80.41, 77.78, 75.45, 75.38, 73.26, 67.51, 55.31. HRESIMS: calcd for $[M+NH_4]^+$ $C_{33}H_{34}O_5S$ + NH_4: 560.2465; found: 560.2466.

2,4-di-*O*-benzyl-3-*p*-methoxybenzyl-D-xylopyranosyl trichloroacetimidate (**19**)

To a stirred mixture of phenyl 2,4-di-*O*-benzyl-3-*p*-methoxybenzyl-1-thio-*β*-D-xylopyranoside (3.0 g, 5.5 mmol) in water:acetonitrile (20 mL, 1:19) was added *N*-bromosuccinimide (1.1 g, 6.0 mmol) and concentrated H_2SO_4 (1-2 drops). The

mixture was stirred for 4 h, then concentrated, diluted in $NaHCO_3$ and extracted with CH_2Cl_2 (3 x 50 mL). The combined extracts were then washed with brine (50 mL), dried ($MgSO_4$), and concentrated. Crystallization using EtOAc and hexanes afforded 2,4-di-*O*-benzyl-3-*p*-methoxybenzyl-D-xylopyranose as a white crystalline solid (2.2 g, 90% yield). ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.25-7.40 (m, 14.4H, Ar), 6.85-6.90 (m, 2.4H, Ar), 5.15 (dd, 1H, *J* = 3.2, 3.1 Hz, H1*α*), 4.92 (d, 0.2H, *J* = 11.2 Hz, $BnCH_2$), 4.80-4.85 (m, 3.6H, $BnCH_2$), 4.77 (d, 0.2H, *J* = 11.6 Hz, $BnCH_2$), 4.76 (d, 1H, *J* = 11.6 Hz, $BnCH_2$), 4.72 (m, 1.2H, $BnCH_2$, H1*β*), 4.68 (d, 1H, *J* = 11.7 Hz, $BnCH_2$), 4.67 (d, 0.2H, *J* = 11.7 Hz, $BnCH_2$), 4.00 (dd, 0.2H, *J* = 3.4 Hz, *J* = 8.5 Hz, H3*β*), 3.88 (dd, 1H, *J* = 8.7 Hz, H3α), 3.85 (s, 3H, OMe*α*), 3.83 (s, 0.6H, OMe*β*), 3.82 (d, 1H, *J* = 10.9 Hz, H5'*α*), 3.71 (dd, 1H, *J* = 5.2 Hz, *J* = 11.2 Hz, H5*α*), 3.62-3.67 (m, 0.4H, H4*β*, H5*β*), 3.57 (ddd, 1H, *J* = 5.3 Hz, *J* = 8.5 Hz, *J* = 8.7 Hz, H4*α*), 3.52 (dd, 1H, *J* = 3.4 Hz, *J* = 8.8 Hz, H2*α*), 3.30-3.36 (m, 0.4H, H2*β*, H4*β*), 3.15 (d, 0.2H, *J* = 5.1 Hz, OH*β*), 2.96 (d, 1H, *J* = 3.0 Hz, OH *α*). ^{13}C NMR (150 MHz, $CDCl_3$) δ 159.28, 150.98, 138.40, 138.31, 138.16, 137.89, 129.71, 128.56, 128.52, 128.48, 128.09, 128.06, 128.05, 127.83, 127.79, 113.84, 97.71, 91.56, 82.23, 80.16, 79.48, 77.51, 77.47, 77.27, 77.06, 75.23, 75.19, 73.49, 73.28, 73.24, 63.73, 60.50, 55.33. HRESIMS: calcd for $[M+Na]^+$ $C_{26}H_{28}O_5$ + Na: 443.1829; found: 443.1821.

2,4-Di-*O*-benzyl-3-*p*-methoxybenzyl-D-xylopyranose (2.2 g, 4.9 mmol) was dissolved in CH_2Cl_2 (dry, 20 mL) under N_2 and Cl_3CCN (0.64 mL, 6.4 mmol) was added dropwise over 2 min. 1,8-diazabicyclo[5.4.0]-undec-7-ene (2 drops) was added and the solution was stirred for 2 h at room temperature. The solution was then concentrated and flash chromatography (EtOAc:hexanes 3:7) of the residue gave presumably the trichloroacetimidate intermediate **19** (2.0 g) as an oily product which was used without further characterization.

Phenyl 4-*O*-*p*-methoxybenzyl-1-thio-*β*-D-xylopyranoside (**20**)

Compound **14a** (3.6 g, 13.0 mmol) was dissolved in DMF (dry, 40 mL), *p*-methoxybenzyl bromide (2.1 mL, 14.1 mmol) was added and the reaction was stirred at room temperature under an atmosphere of N_2. NaH (60 %, 0.56 g, 14.1 mmol) was added and the reaction stirred for another hour. Upon completion, the reaction was quenched with methanol and allowed to stir for another 30 min, then the reaction was concentrated under reduced pressure, taken up in EtOAc (150 mL), and washed with H_2O (30 mL) and brine (30 mL), dried over $MgSO_4$, concentrated and used without further purification.

The intermediate from above was dissolved in methanol (45 mL), concentrated HCl (450 µL) was added and the reaction was stirred until judged to be complete by TLC. Upon completion the reaction was quenched with triethylamine, concentrated and crystallized using a mixture of CH_2Cl_2:hexanes. Phenyl 4-*O*-*p*-methoxybenzyl-1-thio-*β*-D-xylopyranoside (**20**) was recovered as a white solid (3.6 g, 80 % yield over two steps). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.52-7.56 (m, 2H, Ar), 7.32-7.36 (m, 3H, Ar), 7.25-7.28 (m, 2H, Ar), 6.88-6.92 (m, 2H, Ar), 4.63 (d, 1H, *J* = 11.4 Hz, $BnCH_2$), 4.59 (d, 1H, *J* = 11.2 Hz, $BnCH_2$), 4.57 (d, 1H, *J* = 8.9 Hz, H1), 4.08 (dd, 1H, *J* = 4.9 Hz, *J* = 11.6 Hz, H5), 3.82 (s, 3H, OMe), 3.68 (dd, 1H, *J* = 8.4 Hz, H3), 3.48 (ddd, 1H, *J* = 4.9 Hz, *J* = 8.6 Hz, *J* = 9.7 Hz, H4), 3.42 (dd, 1H, *J* = 8.4 Hz, *J* = 8.8 Hz, H2), 3.28 (dd, 1H, *J* = 9.7 Hz, *J* = 11.5 Hz, H5'). ^{13}C NMR (150 MHz, $CDCl_3$) δ 159.53, 132.63, 132.21, 129.93, 129.58, 129.10, 129.07, 128.10, 114.02, 88.70, 76.51, 76.39, 72.71, 71.99, 67.03, 55.31. HRESIMS: calcd for $[M+Na]^+$ $C_{19}H_{22}O_5S$ + Na: 385.1080; found: 385.1086.

Phenyl 2,3-di-*O*-benzyl-4-*p*-methoxybenzyl-1-thio-*β*-D-xylopyranoside (**21**)

Phenyl 4-*O*-*p*-methoxybenzyl-1-thio-*β*-D-xylopyranoside (3.6 g, 10.0 mmol) was dissolved in DMF (dry, 60 mL). Benzyl bromide (2.6 mL, 20.9 mmol), and NaH (60 %, 0.84 g, 20.9 mmol) were added to the solution and the reaction was stirred vigorously at room temperature for 1 h. Upon completion the reaction was

quenched with methanol and allowed to stir for another 30 min, then the reaction was concentrated under reduced pressure, taken up in EtOAc (150 mL), and washed with H_2O (30 mL) and brine (30 mL), dried over $MgSO_4$, and concentrated. Flash column chromatography (EtOAc:hexanes, 1:4) afforded the desired compound, phenyl 2,3-di-*O*-benzyl-4-*p*-methoxybenzyl-1-thio-*β*-D-xylopyranoside as a white solid (4.6 g, 85 % yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.54-7.58 (m, 2H, Ar), 7.26-7.45 (m, 15H, Ar), 6.90 (d, 2H, *J* = 8.5 Hz, Ar), 4.95 (d, 1H, *J* = 10.9 Hz, $BnCH_2$), 4.90 (d, 1H, *J* = 10.5 Hz, $BnCH_2$), 4.88 (d, 1H, *J* = 11.0 Hz, $BnCH_2$), 4.71 (d, 1H, *J* = 9.3 Hz, H1), 4.69 (d, 1H, *J* = 11.1 Hz, $BnCH_2$), 4.60 (d, 1H, *J* = 11.3 Hz, $BnCH_2$), 4.07 (dd, 1H, *J* = 4.5 Hz, *J* = 11.1 Hz, H5), 3.84 (s, 3H, OMe), 3.64-3.70 (m, 2H, H3, H4), 3.47 (dd, 1H, *J* = 8.7 Hz, H2), 3.27 (dd, 1H, *J* = 9.6 Hz, *J* = 11.1 Hz, H5'). ^{13}C NMR (150 MHz, $CDCl_3$) δ 159.43, 138.54, 138.11, 133.80, 131.94, 130.20, 129.57, 129.00, 128.46, 128.45, 128.24, 127.98, 127.89, 127.76, 127.58, 113.94, 88.44, 85.38, 80.45, 77.45, 75.69, 75.52, 72.97, 67.57, 55.32. HRESIMS: calcd for $[M+Na]^+$ $C_{33}H_{34}O_5S$ + Na: 565.2019; found: 565.2037.

2,3-Di-*O*-benzyl-4-*p*-methoxybenzyl-D-xylopyranosyl trichloroacetimidate (**23**)

To a stirred mixture of phenyl 2,3-di-*O*-benzyl-4-*p*-methoxybenzyl-1-thio-*β*-D-xylopyranoside (0.4 g, 0.72 mmol) in water:acetonitrile (10 mL, 1:19) was added *N*-bromosuccinimide (0.14 g, 0.79 mmol) and concentrated H_2SO_4 (1-2 drops). The mixture was stirred for 4 h, then concentrated, diluted in $NaHCO_3$ and extracted with CH_2Cl_2 (3 x 20 mL), the combined extracts were then washed with brine (20 mL), dried ($MgSO_4$), and concentrated. Crystallization using EtOAc and hexanes afforded 2,3-di-*O*-benzyl-4-*p*-methoxybenzyl-D-xylopyranose (**22**) as a white crystalline solid (0.3 g, 90% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.32-7.48 (m, 24H, Ar), 6.93-6.97 (m, 4H, Ar), 5.17 (dd, 1H, *J* = 3.2 Hz, H1*α*), 4.95-5.03 (m, 5H, $BnCH_2$), 4.80-4.86 (m, 2H, $BnCH_2$), 4.62-4.77 (m, 7H, $BnCH_2$, H1*β*), 3.95-4.03 (m, 3H, H3*α*, H5*β*, OH*α*), 3.89 (dd, 1H, *J* = 6.6 Hz, 10.7 Hz H5α), 3.86 (m, 6H, OMe*α*, OMe*β*) 3.62-3.74 (m, 4H, H3*β*, H4*β*, H4*α*, H5'*α*), 3.57 (dd, 1H, *J* = 3.4

Hz, *J* = 9.1 Hz, H2*α*), 3.43 (dd, 1H, *J* = 7.7 Hz, H2*β*), 3.27 (dd, 1H, *J* = 10.0 Hz, *J* = 11.4 Hz, H5'*β*). ^{13}C NMR (125 MHz, $CDCl_3$) δ 159.44, 159.39, 138.88, 138.77, 138.59, 138.06, 130.50, 130.35, 129.61, 129.56, 128.59, 128.47, 128.46, 128.17, 128.12, 128.05, 128.03, 127.77, 127.73, 114.01, 113.97, 97.93, 91.37, 83.58, 82.74, 80.76, 79.68, 77.53, 77.51, 75.66, 75.59, 74.91, 73.34, 73.05, 73.01, 63.88, 60.23, 55.35. HRESIMS: calcd for $[M+Na]^+$ $C_{26}H_{28}O_5$ + Na: 443.1829; found: 443.1821.

2,3-Di-*O*-benzyl-4-*p*-methoxybenzyl-D-xylopyranose (0.3 g, 0.49 mmol) was dissolved in CH_2Cl_2 (dry, 5 mL) under N_2 and Cl_3CCN (0.064 mL, 0.64 mmol) was added dropwise over 2 min. 1,8-diazabicyclo[5.4.0]-undec-7-ene (1 drops) was added and the solution was stirred for 2 h at room temperature. The solution was then concentrated and flash chromatography (EtOAc:hexanes 3:7) of the residue gave the presumed trichloroacetimidate intermediate **6** (0.3 g) as an oily product which was used without further characterization.

Phenyl 2-*O*-*p*-methoxybenzyl-1-thio-*β*-D-xylopyranoside (**24**)

Compound **14b** (2.4 g, 8.5 mmol) was dissolved in DMF (dry, 30 mL), *p*-methoxybenzyl bromide (1.4 mL, 9.6 mmol) was added and the reaction was stirred at room temperature under an atmosphere of N_2. NaH (60 %, 0.39 g, 9.6 mmol) was added and the reaction continued to stir for 1 h. Upon completion, the reaction was quenched with methanol and allowed to stir for 30 min, then the reaction was concentrated under reduced pressure, taken up in EtOAc (150 mL), and washed with H_2O (30 mL) and brine (30 mL), dried over $MgSO_4$, concentrated and used without further purification.

The intermediate from above was dissolved in methanol (45 mL), concentrated HCl (450 µL) was added and the reaction was stirred until judged complete by TLC. Upon completion the reaction was quenched with triethylamine, concentrated and crystallized using a mixture of CH_2Cl_2:hexanes.

Phenyl 2-*O*-*p*-methoxybenzyl-1-thio-*β*-D-xylopyranoside was recovered as a white solid (2.3 g, 75 % yield over two steps). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.54-7.57 (m, 2H, Ar), 7.30-7.37 (m, 5H, Ar), 6.91-6.94 (m, 2H, Ar), 4.90 (d, 1H, *J* = 10.9 Hz, $BnCH_2$), 4.80 (d, 1H, *J* = 9.2 Hz, H1), 4.60 (d, 1H, *J* = 10.9 Hz, $BnCH_2$), 4.15 (dd, 1H, *J* = 4.6 Hz, *J* = 11.6 Hz, H5), 3.83 (s, 3H, OMe), 3.68 (ddd, 1H, *J* = 4.7 Hz, *J* = 8.3 Hz, H4), 3.61 (dd, 1H, *J* = 7.8 Hz, H3), 3.39 (dd, 1H, *J* = 7.9 Hz, H2), 3.33 (dd, 1H, *J* = 8.8 Hz, *J* = 11.6 Hz, H5'). ^{13}C NMR (150 MHz, $CDCl_3$) δ 159.64, 134.02, 131.57, 129.95, 129.08, 127.59, 114.11, 87.64, 79.52, 76.427, 74.31, 69.38, 67.73, 55.32. HRESIMS: calcd for $[M+Na]^+$ $C_{19}H_{22}O_5S$ + Na: 385.1080; found: 385.1086.

Phenyl 3,4-di-*O*-benzyl-2-*p*-methoxybenzyl-1-thio-*β*-D-xylopyranoside (**25**)

Phenyl 2-*O*-*p*-methoxybenzyl-1-thio-*β*-D-xylopyranoside (2.3 g, 6.4 mmol) was dissolved in DMF (dry, 40 mL). Benzyl bromide (1.6 mL, 13.0 mmol), and NaH (60 %, 0.52 g, 13.0 mmol) were added to the solution and the reaction was stirred vigorously at room temperature for 1 h. Upon completion, the reaction was quenched with methanol and allowed to stir for another 30 min, at this time the reaction was concentrated under reduced pressure, taken up in EtOAc (150 mL), and washed with H_2O (30 mL) and brine (30 mL), dried over $MgSO_4$, and concentrated. Flash column chromatography (EtOAc:hexanes, 1:4) afforded the desired compound, phenyl 3,4-di-*O*-benzyl-2-*p*-methoxybenzyl-1-thio-*β*-D-xylopyranoside (**25**) as a white solid (2.9 g, 85 % yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.54-7.58 (m, 2H, Ar), 7.27-7.40 (m, 15H, Ar), 6.90 (d, 2H, *J* = 8.5 Hz, Ar), 4.94 (d, 1H, *J* = 11.0 Hz, $BnCH_2$), 4.89 (d, 1H, *J* = 11.0 Hz, $BnCH_2$), 4.83 (d, 1H, *J* = 10.0 Hz, $BnCH_2$), 4.71-4.77 (m, 2H, $BnCH_2$), 4.69 (d, 1H, *J* = 9.4 Hz, H1), 4.66 (d, 1H, *J* = 11.6 Hz, $BnCH_2$), 4.09 (dd, 1H, *J* = 4.5 Hz, *J* = 11.2 Hz, H5), 3.84 (s, 3H, OMe), 3.64-3.70 (m, 2H, H3, H5'), 3.47 (dd, 1H, *J* = 9.0 Hz, H5), 3.27 (dd, 1H, *J* = 11.2 Hz, *J* = 9.5 Hz, H5). ^{13}C NMR (150 MHz, $CDCl_3$) δ 159.40, 138.56, 133.83, 131.90, 130.26, 129.93, 129.00, 128.52, 128.46, 127.94, 127.87,

127.74, 127.55, 113.86, 88.49, 85.38, 80.17, 77.73, 75.68, 75.16, 73.27, 67.51, 55.32. HRESIMS: calcd for $[M+Na]^+$ $C_{33}H_{34}O_5S$ + Na: 565.2019; found: 565.2038.

3,4-Di-*O*-benzyl-2-*p*-methoxybenzyl-D-xylopyranosyl trichloroacetimidate (**27**)

To a stirred mixture of phenyl 3,4-di-*O*-benzyl-2-*p*-methoxybenzyl-1-thio-*β*-D-xylopyranoside (0.4 g, 0.72 mmol) in water:acetonitrile (10 mL, 1:19) was added *N*-bromosuccinimide (0.14 g, 0.79 mmol) and concentrated H_2SO_4 (1-2 drops). The mixture was stirred for 4 h, then concentrated, diluted in $NaHCO_3$ and extracted with CH_2Cl_2 (3 x 20 mL), the combined extracts were then washed with brine (20 mL), dried ($MgSO_4$), and concentrated. Crystallization using EtOAc and hexanes afforded 2,4-di-*O*-benzyl-3-*p*-methoxybenzyl-D-xylopyranose (**26**) as a white crystalline solid (0.3 g, 90% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.25-7.40 (m, 16.2H, Ar), 6.85-6.90 (m, 2.6H, Ar), 5.09 (dd, 1H, *J* = 3.2 Hz, H1*α*), 4.83-4.90 (m, 3H, $BnCH_2$), 4.70-4.75 (m, 2.6H, $BnCH_2$), 4.60-4.69 (m, 2.7H, $BnCH_2$ H1*β*), 3.97 (dd, 0.34H, *J* = 4.7 Hz, *J* = 11.5 Hz, H5*β*), 3.87 (dd, 1H, *J* = 8.6 Hz, H3α), 3.78-3.84 (m, 5H, OMe*α*, OMe*β*, H5'*α*) 3.69 (dd, 1H, *J* = 5.2 Hz, *J* = 11.2 Hz, H5*α*), 3.63 (dd, 0.34H, *J* = 4.9 Hz, *J* = 9.3 Hz, H4*β*), 3.60 (dd, 0.34H, *J* = 9.6 Hz, H3*β*), 3.55 (ddd, 1H, *J* = 4.9 Hz, *J* = 8.5 Hz, *J* = 9.6 Hz, H4*α*), 3.48 (dd, 0.4H, *J* = 3.4 Hz, *J* = 8.8 Hz, H2*α*), 3.33 (dd, 0.34H, *J* = 7.8 Hz, H2*β*), 3.29 (dd, 0.34H, *J* = 9.6 Hz, *J* = 11.6 Hz, H5'*β*), 3.22 (d, 1H, *J* = 3.1 Hz, OH *α*). ^{13}C NMR (125 MHz, $CDCl_3$) δ 159.40, 159.25, 138.70, 138.60, 138.27, 138.10, 129.82, 129.78, 128.48, 128.44, 128.38, 128.01, 127.96, 127.84, 127.81, 127.65, 113.86, 113.81, 97.93, 91.51, 82.17, 80.23, 79.00, 77.64, 77.31, 75.48, 75.37, 73.28, 73.12, 73.00, 63.73, 60.42, 55.29. HRESIMS: calcd for $[M+Na]^+$ $C_{26}H_{28}O_5$ + Na: 443.1829; found: 443.1821.

3,4-Di-*O*-benzyl-2-*p*-methoxybenzyl-D-xylopyranose (0.3 g, 0.49 mmol) was dissolved in CH_2Cl_2 (dry, 5 mL) under N_2 and Cl_3CCN (0.064 mL, 0.64 mmol) was added dropwise over 2 min. 1,8-Diazabicyclo[5.4.0]-undec-7-ene (1 drops) was added and the solution was stirred for 2 h at room temperature. The solution was

then concentrated and flash chromatography (EtOAc:hexanes 3:7) of the residue gave the presumed trichloroacetimidate intermediate **27** (0.3 g) as an oily product which was used without further characterization.

5.6.4 Synthesis of 2,3,4-tri-*O*-benzyl-D-xylopyranosyl trichloroacetimidate (30)

Phenyl 2,3,4-tri-*O*-benzyl-1-thio-*β*-D-xylopyranoside (**28**)

Phenyl 1-thio-*β*-D-xylopyranoside (6.8 g, 28mmol) was dissolved in dimethylformamide (dry) and NaH (2.2 g, 90 mmol, 95% NaH) and BnBr (10.7 mL, 90 mmol) were added. The mixture was stirred at room temperature for 1 h and then was quenched with methanol. The mixture was concentrated, then diluted with water (30 mL) and extracted with EtOAc (3 x 40 mL). The combined extracts were then washed with brine (30 mL), dried ($MgSO_4$) and concentrated. Flash chromatography (EtOAc:hexanes 1:9) gave phenyl 2,3,4-tri-*O*-benzyl-1-thio-*β*-D-xylopyranoside as a white solid (11.4 g, 79% yield). ^{1}H and ^{13}C NMR spectra were in accordance with published data.[327] ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.52-7.53 (m, 2H), 7.28-7.42 (m, 18H), 4.91 (d, 1H, *J* = 10.9 Hz), 4.87 (d, 1H, *J* = 10.3 Hz), 4.86 (d, 1H, *J* = 10.9 Hz), 4.77 (d, 1H, *J* = 10.3 Hz), 4.73 (d, 1H, *J* = 11.7 Hz), 4.69 (d, 1H. *J* = 9.4 Hz), 4.64 (d, 1H, *J* = 11.6 Hz), 4.08 (dd, 1H, *J* = 4.4, 11.4 Hz), 3.62-3.69 (m, 2H), 3.46 (dd, 1H, *J* = 9.0 Hz), 3.22-3.30 (m, 1H).

2,3,4-Tri-*O*-benzyl-D-xylopyranosyl trichloroacetimidate (**30**)

To a stirred mixture of phenyl 2,3,4-tri-*O*-benzyl-1-thio-*β*-D-xylopyranoside (11.5 g, 22.5 mmol) in water:acetonitrile (30 mL, 1:19) was added *N*-bromosuccinimide (4.4 g, 24.7 mmol) and concentrated H_2SO_4 (1-2 drops). The mixture was stirred for 4 h, then concentrated, diluted in $NaHCO_3$ and extracted with CH_2Cl_2 (3 x 50 mL), the combined extracts were then washed with brine (50 mL), dried ($MgSO_4$), and concentrated. Crystallization using EtOAc and hexanes

presumably afforded 2,3,4-tri-*O*-benzyl-D-xylopyranose as a white crystalline solid (6.4 g, 68% yield).

2,3,4-Tri-*O*-benzyl-D-xylopyranose (1.8 g, 4.3 mmol)was dissolved in CH_2Cl_2 (dry, 20 mL) under N_2 and Cl_3CCN (0.77 mL, 7.7 mmol) was added dropwise over 2 min. 1,8-diazabicyclo[5.4.0]-undec-7-ene (2 drops) was added and the solution was stirred for 2 h at room temperature. The solution was then concentrated and flash chromatography (EtOAc:hexanes 3:7) of the residue gave presumably the trichloroacetimidate intermediate **30** (2.0 g) as an oily product which was used without further characterization.

5.6.5 Synthesis of XG, XXG, X(1-4)XG, and X(1-2)XG

β-*O*-Benzyl (2,4-di-*O*-benzyl-3-*p*-methoxybenzyl-D-xylose)-(*α*1-3)-(2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (**31a**)

β-*O*-Benzyl 2-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside (0.88 g, 1.97 mmol) and 2,4-di-*O*-benzyl-3-*p*-methoxybenzyl-D-xylopyranosyl trichloroacetimidate (1.75 g, 2.95 mmol) were dissolved in toluene (dry, 20 mL, 3 times) and evaporated to remove any H_2O. *β*-*O*- Benzyl 2-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside and 2,4-di-*O*-benzyl-3-*p*-methoxybenzyl-D-xylopyranosyl trichloroacetimidate were then dissolved in DCM (dry, 20 mL), flame dried sieves were added and the mixture was stirred for twenty minutes under an atmosphere of N_2. The mixture was then cooled to -60 °C and TMSOTf (catalytic) was added and the reaction was stirred for 10 min at -60 °C and then allowed to slowly come up to room temperature. Upon completion, the reaction was diluted with DCM (80 mL), the sieves were filtered off, the solution was concentrated and loaded in toluene onto a flash chromatography column (toluene:EtOAc, 19:1) which gave an inseparable mixture (2.0 g) containing the desired compound **31a**.

β-*O*-Benzyl (2,4-di-*O*-benzyl-D-xylose)-(*α*1-3)-(2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (**32a**)

The aforementioned mixture containing *β*-*O*-benzyl (2,4-di-*O*-benzyl-3-*p*-methoxybenzyl-D-xylose)-(*α*1-3)-(2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (2.0 g) was dissolved in DCM:Water (19:1, 20 mL), DDQ (0.67 g, 0.30 mmol) was added and the reaction was stirred for 2 h. The reaction was judged complete by TLC, diluted with DCM (100 mL), the organic layer was extracted and washed with water (10 mL), saturated $NaHCO_3$ (10 mL), brine (10 mL), dried over $MgSO_4$ and concentrated by rotary evaporation. The residue was then dissolved in toluene and loaded onto a flash column chromatography column (EtOAc:Toluene, 1:24 to 1:7) which afforded *β*-*O*-benzyl (2,4-di-*O*-benzyl-D-xylose)-(*α*1-3)-(2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (**32a**) as a white solid (0.75 g, 50 % yield over two steps). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.23-7.46 (m, 23H, Ar), 7.04-7.08 (m, 2H, Ar), 5.54 (d, 1H, *J* = 3.5 Hz, $H1^B$), 5.45 (s, 1H, PhCH), 5.03 (d, 1H, *J* = 11.7 Hz, $BnCH_2$), 5.00 (d, 1H, *J* = 10.3 Hz, $BnCH_2$), 4.80 (d, 1H, *J* = 11.3 Hz, $BnCH_2$), 4.75 (d, 1H, *J* = 8.5 Hz, $H1^A$), 4.74 (d, 1H, *J* = 10.8 Hz, $BnCH_2$), 4.68 (d, 1H, *J* = 12.0 Hz, $BnCH_2$), 4.66 (d, 1H, *J* = 12.5 Hz, $BnCH_2$), 4.57 (d, 1H, *J* = 11.8 Hz, $BnCH_2$), 4.41 (dd, 1H, *J* = 5.0 Hz, *J* = 10.6 Hz, $H6^A$), 4.36 (d, 1H, *J* = 12.2 Hz, $BnCH_2$), 4.10-4.17 (m, 2H, $H3^A$, $H3^B$), 3.94 (dd, 1H, *J* = 10.5 Hz, $H5^B$), 3.83-3.88 (m, 2H, $H4^A$, $H6'^A$), 3.65 (dd, 1H, *J* = 8.4 Hz, $H2^A$), 3.53 (ddd, 1H, *J* = 5.1 Hz, *J* = 9.8 Hz, $H5^A$), 3.4-3.5 (m, 2H, $H4^B$, $H5'^B$), 3.28 (dd, 1H, *J* = 3.5 Hz, *J* = 9.6 Hz, $H2^B$), 2.65 (s, 1H, OH). ^{13}C NMR (150 MHz, $CDCl_3$) δ 138.44, 137.71, 137.86, 137.02, 136.96, 129.46, 128.72, 128.58, 128.47, 128.41, 128.40, 128.31, 128.25, 128.23, 128.10, 127.73, 127.70, 127.46, 126.42, 103.25, 102.13, 95.45, 82.42, 80.34, 77.86, 77.53, 75.58, 75.31, 72.83, 72.24, 71.75, 70.87, 68.96, 65.77, 59.76. HRESIMS: calcd for $[M+NH_4]^+$ $C_{46}H_{48}O_{10}$ + NH_4: 778.3586; found: 778.3573.

β-*O*-Benzyl (2,3-di-*O*-benzyl-4-*p*-methoxybenzyl-D-xylose)-(*α*1-3)-(2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (**31b**)

β-*O*-Benzyl 2-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside (0.25 g, 0.56 mmol) and 2,3-di-*O*-benzyl-4-*p*-methoxybenzyl-D-xylopyranosyl trichloroacetimidate (0.40 g, 0.67 mmol) were dissolved in toluene (dry, 10 mL, 3 times) and evaporated to remove any H_2O. *β*-*O*- Benzyl 2-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside and 2,3-di-*O*-benzyl-4-*p*-methoxybenzyl-trichloroacetimidate-D-xylopyranoside were then dissolved in DCM (Dry, 10 mL), flame dried sieves were added and the mixture was stirred for 20 min under an atmosphere of N_2. The mixture was then cooled to -60 °C and TMSOTf (catalytic) was added and the reaction was stirred for 10 min at -60 °C and then allowed to slowly come up to room temperature. Upon completion the reaction was diluted with DCM (40 mL), the sieves were filtered off, the solution was concentrated and loaded in toluene onto a flash chromatography column (toluene:EtOAc, 19:1) which gave an inseparable mixture (0.4 g) containing the desired compound **31b**.

Benzyl (2,3-di-*O*-benzyl-D-xylose)-(*α*1-3)-(2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (**32b**)

The aforementioned mixture containing benzyl (2,3-di-*O*-benzyl-4-*p*-methoxybenzyl-D-xylose)-(*α*1-3)-(2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (0.4 g) was dissolved in CH_2Cl_2:H_2O (19:1, 5 ml), DDQ (0.12 g, 0.54 mmol) was added and the reaction was stirred for 2h. The reaction was judged complete by TLC, diluted with CH_2Cl_2 (100 mL), the organic layer was extracted and washed with water (10 mL), saturated $NaHCO_3$ (10 mL), brine (10 mL), dried over $MgSO_4$ and concentrated by rotary evaporation. The residue was then dissolved in toluene and loaded onto a flash column chromatography column (EtOAc:toluene, 1:24 to 1:7) which afforded *β*-*O*-benzyl (2,4-di-*O*-benzyl-D-xylose)-(*α*1-3)-(2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (0.15 g, 35 % yield over two steps) as a white solid. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.47-7.52 (m, 1H, Ar), 7.25-7.39 (m, 21H, Ar), 7.18-7.22 (m, 2H, Ar), 5.48 (d, 1H, *J* = 3.3 Hz, $H1^B$), 5.47 (s, 1H, PhCH), 5.00 (d, 1H, *J* = 11.5 Hz, $BnCH_2$), 4.94 (d, 1H, *J* = 10.4

Hz, $BnCH_2$), 4.75 (d, 1H, *J* = 10.4 Hz, $BnCH_2$), 4.72 (d, 1H, *J* = 8.3 Hz, $H1^A$), 4.71 (d, 2H, *J* = 11.2 Hz, $BnCH_2$), 4.62 (d, 1H, *J* = 12.2 Hz, $BnCH_2$), 4.38 (dd, 1H, *J* = 5.1 Hz, *J* = 10.4 Hz, $H5^A$), 4.37 (d, 1H, *J* = 12.2 Hz, $BnCH_2$), 4.10 (dd, 1H, *J* = 9.3 Hz, $H3^A$), 3.87 (dd, 1H, *J* = 10.7 Hz, $H4^A$), 3.85 (dd, 1H, *J* = 9.4 Hz, $H5^B$), 3.84 (dd, 1H, *J* = 10.0 Hz, $H6^A$), 3.70 (dd, 1H, *J* = 8.7 Hz, $H3^B$), 3.63 (dd, 1H, *J* = 7.9 Hz, *J* = 9.0 Hz, $H2^A$), 3.45-3.58 (m, 3H, $H4^B$, $H5'^B$ $H6'^B$), 3.40 (dd, 1H, *J* = 3.4 Hz, *J* = 9.2 Hz, $H2^B$). ^{13}C NMR (150 MHz, $CDCl_3$) δ 138.44, 138.26, 137.77, 137.22, 136.93, 128.98, 128.52, 128.46, 128.34, 128.34, 128.18, 128.08, 128.03, 127.79, 127.70, 127.69, 127.63, 126.14, 102.97, 101.41, 100.73, 82.95, 79.52, 78.81, 77.95, 77.55, 74.83, 73.45, 73.16, 71.60, 68.85, 68.25, 66.41, 62.96. HRESIMS: calcd for $[M+NH_4]^+$ $C_{46}H_{48}O_{10} + NH_4$: 778.3586; found: 778.3573.

β-*O*-Benzyl (3,4-di-*O*-benzyl-2-*p*-methoxybenzyl-D-xylose)-(*α*1-3)-(2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (**31c**)

β-*O*-Benzyl 2-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside (0.88 g, 1.97 mmol) and 3,4-di-*O*-benzyl-2-*p*-methoxybenzyl-D-xylopyranosyl trichloroacetimidate (1.75 g, 2.95 mmol) were dissolved in toluene (dry, 20 mL, 3 times) and evaporated to remove any H_2O. *β*-*O*- Benzyl 2-*O*-benzyl-4,6-*O*-benzylidine-1-*O*-*β*-D-glucopyranoside and 3,4-di-*O*-benzyl-2-*p*-methoxybenzyl-D-xylopyranosyl trichloroacetimidate were then dissolved in CH_2Cl_2 (Dry, 20 mL), flame dried sieves were added and the mixture was stirred for 20 min under an atmosphere of N_2. The mixture was then cooled to -60 °C and TMSOTf (catalytic) was added and the reaction was stirred for 10 min at -60 °C and then allowed to slowly come up to room temperature. Upon completion the reaction was diluted with DCM (80 mL), the sieves were filtered off, the solution was concentrated and loaded in toluene onto a flash chromatography column (toluene:EtOAc, 19:1) which gave an inseparable mixture (2.0 g) containing the desired compound **31c**.

Benzyl (3,4-di-*O*-benzyl-D-xylose)-(*α*1-3)-(2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (**32c**)

The aforementioned mixture containing benzyl (3,4-di-*O*-benzyl-2-*p*-methoxybenzyl-D-xylose)-(*α*1-3)-(2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (2.0 g) was dissolved in CH_2Cl_2:H_2O (19:1, 20 mL), DDQ (0.67 g, 0.30 mmol) was added and the reaction was stirred for 2 h. The reaction was judged complete by TLC, diluted with CH_2Cl_2 (100 mL), the organic layer was extracted and washed with water (10 mL), saturated $NaHCO_3$ (10 mL), brine (10 mL), dried over $MgSO_4$ and concentrated by rotary evaporation. The residue was then dissolved in toluene and loaded onto a flash column chromatography column (EtOAc:toluene, 1:24 to 1:7) which afforded *β*-*O*-benzyl (3,4-di-*O*-benzyl-D-xylose)-(*α*1-3)-(2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (0.75 g, 50 % yield over two steps) as a white solid. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.37-7.42 (m, 2H, Ar), 7.12-7.32 (m, 23H, Ar), 5.48 (s, 1H, PhCH), 5.17 (d, 1H, *J* = 3.5 Hz, $H1^B$), 4.88 (d, 1H, *J* = 11.7 Hz, $BnCH_2$), 4.82 (d, 1H, *J* = 10.2 Hz, $BnCH_2$), 4.72 (d, 1H, *J* = 11.2 Hz, $BnCH_2$), 4.68 (d, 1H, *J* = 12.3 Hz, $BnCH_2$), 4.67 (d, 1H, *J* = 10.1 Hz, $BnCH_2$), 4.60 (d, 1H, *J* = 11.7 Hz, $BnCH_2$), 4.58 (d, 1H, *J* = 7.7 Hz, $H1^A$), 4.57 (d, 1H, *J* = 11.4 Hz, $BnCH_2$), 4.46 (d, 1H, *J* = 12.2 Hz, $BnCH_2$), 4.32 (dd, 1H, *J* = 5.0 Hz, *J* = 10.5 Hz, $H6^A$), 3.90 (dd, 1H, *J* = 9.2 Hz, $H3^A$), 3.77 (dd, 1H, *J* = 10.1 Hz, $H5^B$), 3.73 (dd, 1H, *J* = 10.0 Hz, $H6'^A$), 3.67 (dd, 1H, *J* = 9.4 Hz, $H4^A$), 3.60 (dd, 1H, *J* = 8.8 Hz, $H3^B$), 3.40-3.55 (m, 4H, $H2^A$, $H2^B$, $H4^B$, $H5'^B$), 3.37 (ddd, 1H, *J* = 5.0 Hz, *J* = 9.8 Hz, $H5^A$), 2.52 (s, 1H, OH). ^{13}C NMR (150 MHz, $CDCl_3$) δ 138.79, 138.32, 137.77, 136.92, 136.81, 129.21, 128.64, 128.55, 128.41, 128.38, 128.33, 128.32, 128.16, 128.08, 127.95, 127.78, 127.73, 127.72, 127.59, 126.04, 103.26, 101.56, 99.62, 81.74, 81.31, 80.85, 77.96, 75.36, 75.05, 73.06, 72.94, 71.79, 68.81, 65.83, 60.89. HRESIMS: calcd for $[M+NH_4]^+$ $C_{46}H_{48}O_{10}$ + NH_4: 778.3586; found: 778.3573.

β-*O*-Benzyl (2,3,4-tri-*O*-benzyl-D-xylose)-(*α*1-3)-(2,4-di-*O*-benzyl-D-xylose)-(*α*1-3)-2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (**33a**)

β-*O*-Benzyl (2,4-di-*O*-benzyl-D-xylose)-(*α*1-3)-2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (1.3 g, 1.7 mmol) and 2,3,4-tri-*O*-benzyl-D-xylopyranosyl trichloroacetimidate (3.0 g, 5.3 mmol) were dissolved in toluene (dry, 20 mL, 3 times) and evaporated to remove any H_2O. *β*-*O*-Benzyl (2,4-di-*O*-benzyl-D-xylose)-(*α*1-3)-2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside and 2,3,4-tri-*O*-benzyl-D-xylopyranosyl trichloroacetimidate were then dissolved in CH_2Cl_2 (dry, 20 mL), flame dried sieves were added and the mixture was stirred for twenty minutes. The mixture was then cooled to -60 °C and TMSOTf (catalytic) was added and the reaction was stirred for 10 min at -60 °C and then allowed to slowly come up to room temperature. Upon completion, the reaction was diluted with CH_2Cl_2 (80 mL), the sieves were filtered off, the solution was concentrated and loaded in toluene onto a flash chromatography column (EtOAc:toluene, 1:19). *β*-*O*-Benzyl (2,3,4-tri-*O*-benzyl-D-xylose)-(*α*1-3)-(2,4-di-*O*-benzyl-D-xylose)-(*α*1-3)-2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (**33a**) was isolated as a colourless syrup (1.7 g, 85 % yield) ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.08-7.42 (m, 40H, Ar), 5.65 (d, 1H, *J* = 3.5 Hz, $H1^C$), 5.57 (d, 1H, *J* = 3.6 Hz, $H1^B$), 5.41 (s, 1H, PhCH), 5.00 (d, 1H, *J* = 11.7 Hz, $BnCH_2$), 4.97 (d, 1H, *J* = 10.4 Hz, $BnCH_2$), 4.90 (s, 2H, $BnCH_2$), 4.80 (d, 1H, *J* = 10.4 Hz, $BnCH_2$), 4.72 (d, 1H, *J* = 12.4 Hz, $BnCH_2$), 4.71 (d, 1H, *J* = 7.3 Hz, $H1^A$), 4.64 (d, 1H, *J* = 11.7 Hz, $BnCH_2$), 4.60 (d, 1H, *J* = 11.0 Hz, $BnCH_2$), 4.50-4.59 (m, 3H, $BnCH_2$), 4.48 (d, 1H, *J* = 12.0 Hz, $BnCH_2$), 4.44 (d, 1H, *J* = 12.0 Hz, $BnCH_2$), 4.37 (d, 1H, *J* = 11.7 Hz, $BnCH_2$), 4.36 (d, 1H, *J* = 5.3 Hz, *J* = 10.3 Hz, $H6^A$), 4.21 (dd, 1H, *J* = 9.2 Hz, $H3^B$), 4.15 (dd, 1H, *J* = 13.2 Hz, $H5^C$), 4.11 (dd, 1H, *J* = 9.2 Hz, $H3^A$), 3.87-3.91 (m, 3H, $H3^C$, $H4^A$, $H5^B$), 3.83 (dd, 1H, *J* = 10.3 Hz, $H5^A$), 3.63-3.68 (m, 2H, $H2^A$, $H4^B$), 3.40-3.57 (m, 6H, $H2^B$, $H2^C$, $H4^C$, $H5'^B$, $H5'^C$, $H6'^A$). ^{13}C NMR (150 MHz, $CDCl_3$) δ 163.45, 139.00, 138.70, 138.24, 138.09, 137.67, 137.42, 136.92, 128.53, 128.46, 128.43, 128.31, 128.28, 128.22, 128.15, 128.14, 128.04, 128.00, 127.64, 127.61, 127.56, 127.49, 127.47, 127.40, 126.72, 126.36, 103.25, 102.09, 96.89, 95.42, 81.08, 80.06, 79.23, 78.62, 76.35, 75.61, 75.03, 72.89, 72.59, 71.88, 71.66, 71.02, 65.78, 60.22, 59.21. HRESIMS: calcd for $[M+Na]^+$ $C_{72}H_{74}O_{14}$ + Na: 1185.4971; found: 1185.4968.

β-*O*-Benzyl (2,3,4-tri-*O*-benzyl-D-xylose)-(*α*1-4)-(2,3-di-*O*-benzyl-D-xylose)-(*α*1-3)-2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (**33b**)

β-*O*-Benzyl (2,3,-di-*O*-benzyl-D-xylose)-(*α*1-3)-2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (0.05 g, 0.066 mmol) and 2,3,4-tri-*O*-benzyl-D-xylopyranosyl trichloroacetimidate (0.055 g, 0.099 mmol) were dissolved in toluene (dry, 5 mL, 3 times) and evaporated to remove any H_2O. *β*-*O*-Benzyl (3,4-di-*O*-benzyl-D-xylose)-(*α*1-3)-2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside and 2,3,4-tri-*O*-benzyl-D-xylopyranosyl trichloroacetimidate were then dissolved in DCM (Dry, 5 mL), flame dried sieves were added and the mixture was stirred for twenty minutes. The mixture was then cooled to -60 °C and TMSOTf (catalytic) was added and the reaction was stirred for 10 min at -60 °C and then allowed to slowly come up to room temperature. Upon completion the reaction was diluted with CH_2Cl_2 (20 mL), the sieves were filtered off, the solution was concentrated and loaded in toluene onto a flash chromatography column (EtOAc:toluene, 1:19). *β*-*O*-Benzyl (2,3,4-tri-*O*-benzyl-D-xylose)-(*α*1-4)-(2,3-di-*O*-benzyl-D-xylose)-(*α*1-3)-2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (**33b**) was isolated as a colourless syrup (0.034 g, 43 % yield) ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.08-7.42 (m, 38H, Ar), 6.92 (m, 2H, Ar), 5.48 (d, 1H, *J* = 3.7 Hz, $H1^C$), 5.44 (s, 1H, PhCH), 5.12 (d, 1H, *J* = 3.6 Hz, $H1^B$), 4.98 (d, 1H, *J* = 11.7 Hz, $BnCH_2$), 4.96 (d, 1H, *J* = 11.4 Hz, $BnCH_2$), 4.90-4.95 (m, 3H, $BnCH_2$), 4.84 (d, 1H, *J* = 11.0 Hz, $BnCH_2$), 4.80 (d, 1H, *J* = 11.0 Hz, $BnCH_2$), 4.71 (d, 1H, *J* = 7.8 Hz, $H1^A$), 4.69 (d, 1H, *J* = 11.6 Hz, $BnCH_2$), 4.68 (d, 1H, *J* = 12.5 Hz, $BnCH_2$), 4.57 (m, 3H, $BnCH_2$), 4.52 (d, 1H, *J* = 12.3 Hz, $BnCH_2$), 4.36 (d, 1H, *J* = 5.0 Hz, *J* = 10.5 Hz, $H4^B$), 4.32 (d, 1H, *J* = 12.3 Hz, $BnCH_2$), 4.06-4.11 (m, 2H, $H3^A$, $H5^C$), 3.93 (dd, 1H, *J* = 9.1 Hz, $H3^C$), 3.86 (dd, 1H, *J* = 8.9 Hz, $H3^B$), 3.84 (dd, 1H, *J* = 9.1 Hz, $H4^A$), 3.82 (dd, 1H, *J* = 10.2 Hz, $H6^A$), 3.63-3.68 (m, 3H, $H2^A$, $H4^C$, $H5'^C$), 3.46-3.52 (m, 2H, $H4^B$, $H6'^A$), 3.36-3.43 (m, 4H, $H2^B$, $H2^C$, $H5^B$, $H5'^B$). ^{13}C NMR (150 MHz, $CDCl_3$) δ 138.84, 138.33, 137.90, 137.71, 137.32, 137.24, 136.55, 136.39, 128.90, 128.07, 128.01, 127.97, 127.94, 127.89, 127.78,

127.72, 127.68, 127.60, 127.58, 127.52, 127.44, 127.24, 127.22, 127.01, 126.99, 126.88, 126.86, 126.84, 126.55, 126.48, 125.89, 102.94, 101.57, 98.31, 95.69, 81.74, 80.60, 79.46, 78.86, 78.57, 78.28, 77.62, 75.37, 75.08, 74.29, 74.03, 72.84, 72.22, 71.18, 70.83, 68.46, 65.30, 60.37, 59.78. HRESIMS: calcd for $[M+Na]^+$ $C_{72}H_{74}O_{14}$ + Na: 1185.4971; found: 1185.4968.

β-*O*-Benzyl (2,3,4-tri-*O*-benzyl-D-xylose)-(*α*1-2)-(3,4-di-*O*-benzyl-D-xylose)-(*α*1-3)-2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (**33c**)

β-*O*-Benzyl (3,4-di-*O*-benzyl-D-xylose)-(*α*1-3)-2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (0.05 g, 0.066 mmol) and 2,3,4-tri-*O*-benzyl-D-xylopyranosyl trichloroacetimidate (0.055 g, 0.099 mmol) were dissolved in toluene (dry, 5 mL, 3 times) and evaporated to remove any H_2O. *β*-*O*-Benzyl (3,4-di-*O*-benzyl-D-xylose)-(*α*1-3)-2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside and 2,3,4-tri-*O*-benzyl-D-xylopyranosyl trichloroacetimidate were then dissolved in CH_2Cl_2 (Dry, 5 mL), flame dried sieves were added and the mixture was stirred for twenty minutes. The mixture was then cooled to -60 °C and TMSOTf (catalytic) was added and the reaction was stirred for 10 min at -60 °C and then allowed to slowly come up to room temperature. Upon completion the reaction was diluted with CH_2Cl_2 (20 mL), the sieves were filtered off, the solution was concentrated and loaded in toluene onto a flash chromatography column (EtOAc:toluene, 1:19). *β*-*O*-Benzyl (2,3,4-tri-*O*-benzyl-D-xylose)-(*α*1-2)-(3,4-di-*O*-benzyl-D-xylose)-(*α*1-3)-2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (**33c**) was isolated as a colourless syrup (0.032 g, 41 % yield) ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.06-7.44 (m, 38H, Ar), 6.97 (m, 2H, Ar), 5.52 (d, 1H, *J* = 3.7 Hz, $H1^C$), 5.41 (s, 1H, PhCH), 5.12 (d, 1H, *J* = 3.6 Hz, $H1^B$), 4.98 (d, 1H, *J* = 11.7 Hz, $BnCH_2$), 4.93 (d, 1H, *J* = 11.3 Hz, $BnCH_2$), 4.90-4.95 (m, 3H, $BnCH_2$), 4.84 (d, 1H, *J* = 11.0 Hz, $BnCH_2$), 4.80 (d, 1H, *J* = 11.0 Hz, $BnCH_2$), 4.73 (d, 1H, *J* = 7.8 Hz, $H1^A$), 4.69 (d, 1H, *J* = 11.5 Hz, $BnCH_2$), 4.68 (d, 1H, *J* = 12.5 Hz, $BnCH_2$), 4.53-4.59 (m, 3H, $BnCH_2$), 4.52 (d, 1H, *J* = 12.3 Hz, $BnCH_2$), 4.36 (d, 1H, *J* = 5.0 Hz, *J* = 10.5 Hz, $H4^B$), 4.33 (d, 1H, *J* = 12.5 Hz, $BnCH_2$), 4.04-

4.10 (m, 2H, $H3^A$, $H5^C$), 3.98 (dd, 1H, *J* = 9.1 Hz, $H3^C$), 3.85 (dd, 1H, *J* = 8.9 Hz, $H3^B$), 3.84 (dd, 1H, *J* = 9.0 Hz, $H4^A$), 3.81 (dd, 1H, *J* = 10.3 Hz, $H6^A$), 3.63-3.68 (m, 3H, $H2^A$, $H4^C$, $H5'^C$), 3.42-3.50 (m, 2H, $H4^B$, $H6'^A$), 3.33-3.40 (m, 4H, $H2^B$, $H2^C$, $H5^B$, $H5'^B$). ^{13}C NMR (150 MHz, $CDCl_3$) δ 138.81, 138.37, 137.92, 137.73, 137.35, 137.21, 136.52, 136.35, 128.90, 128.07, 128.01, 127.97, 127.94, 127.92, 127.79, 127.71, 127.68, 127.60, 127.58, 127.55, 127.43, 127.29, 127.24, 127.07, 126.99, 126.89, 126.83, 126.82, 126.55, 126.48, 125.93, 102.84, 101.55, 98.31, 95.69, 82.04, 80.63, 79.46, 78.76, 78.55, 78.38, 77.57, 75.37, 75.08, 74.39, 74.03, 72.78, 72.21, 71.18, 70.83, 68.46, 65.30, 60.37, 59.78. HRESIMS: calcd for $[M+Na]^+$ $C_{72}H_{74}O_{14}$ + Na: 1185.4971; found: 1185.4968.

D-Xylose-*α*1-3-D-glucopyranose (**4**)

Compound **10** (0.05 g, 0.11 mmol) and 2,3,4-tri-*O*-benzyl-D-xylopyranosyl trichloroacetimidate (0.083 g, 0.15 mmol) were dissolved in toluene (dry, 5 mL, 3 times) and evaporated to remove any H_2O. **10** and 2,3,4-tri-*O*-benzyl-D-xylopyranosyl trichloroacetimidate were then dissolved in CH_2Cl_2 (Dry, 5 mL), flame dried sieves were added and the mixture was stirred for twenty minutes. The mixture was then cooled to -60 °C and TMSOTf (catalytic) was added and the reaction was stirred for 10 min at -60 °C and then allowed to slowly come up to room temperature. Upon completion the reaction was diluted with CH_2Cl_2 (20 mL), the sieves were filtered off, the solution was concentrated and loaded in toluene onto a flash chromatography column (EtOAc:toluene, 1:19). *β*-*O*-benzyl (2,3,4-tri-*O*-benzyl-D-xylose)-(*α*1-3)-2-*O*-benzyl-4,6-*O*-benzylidine)-D-glucopyranoside (**31d**) was isolated as a colourless syrup (0.06 g, 63 % yield) and was carried forward to the next step.

MeOH (dry, 30 mL) was added to a vial containing 10% Pd/C (0.2 g), the resulting slurry was transferred to a flask containing *β*-*O*-benzyl (2,3,4-tri-*O*-benzyl-D-xylose)-(*α*1-3)-(2-*O*-benzyl-4,6-*O*-benzylidene)-D-glucopyranoside (0.03 g, 0.035 mmol). The flask was then placed under vacuum, back filled with H_2

three times, and then the reaction was stirred vigorously under H_2 for 3 h. Upon completion, the reaction was filtered through celite, concentrated, and flash column chromatography (EtOAc:MeOH, 5:2) gave the desired product, **4** as a white solid (7.0 mg, 64 % yield). ^{1}H and ^{13}C NMR spectra were in accordance with published data.[328] ^{1}H NMR (600 MHz, D_2O) δ 5.25 (d, 1H, *J* = 3.8 Hz), 4.67 (d, 1H, *J* = 8.0 Hz), 4.00-4.20 (m, 2H), 3.80-3.88 (m, 2H), 3.68-3.85 (m, 10H), 3.40-3.60 (m, 7H), 3.27 (dd, 1H, *J* = 8.0, 9.3 Hz). ^{13}C NMR (150 MHz, D_2O) δ 95.41, 91.60, 75.34, 75.26, 73.64, 73.54, 72.26, 71.57, 70.98, 70.94, 69.18, 69.15, 69.10, 68.74, 68.70, 67.06, 64.70, 63.39, 62.86, 62.00, 60.25, 60.08. HRESIMS: calcd for $[M+Na]^+$ $C_{11}H_{20}NaO_{10}$: 335.0949; found: 335.0955.

D-Xylose-*α*1-3-D-xylose-*α*1-3-D-glucopyranose (**1**)

MeOH (dry, 30 mL) was added to a vial containing 10% Pd/C (0.2 g), the resulting slurry was transferred to a flask containing *β*-*O*-benzyl (2,3,4-tri-*O*-benzyl-D-xylose)-(*α*1-3)-(2,4-di-*O*-benzyl-D-xylose)-(*α*1-3)-(2-*O*-benzyl-4,6-*O*-benzylidene)-D-glucopyranoside (0.9 g, 0.77 mmol). The flask was then placed under vacuum, back filled with H_2 three times, and then the reaction was stirred vigorously under H_2 for 3 h. Upon completion, the reaction was filtered through celite, concentrated, and flash column chromatography (EtOAc:MeOH, 5:2) gave the desired product as a colourless syrup (0.3 g, 87 % yield). ^{1}H and ^{13}C NMR spectra were in accordance with published data.[328] ^{1}H NMR (600 MHz, CD_3OD) δ 5.16-5.19 (m, 4H), 5.13 (d, 1H, *J* = 3.7 Hz), 4.50 (d, 1H, *J* = 7.8 Hz), 3.83-3.92 (m, 5H), 3.61-3.81 (m, 14H), 3.52-3.58 (m, 7H), 3.45-3.49 (m, 4H), 3.42 (d, 1H, *J* = 3.7 Hz), 3.41 (d, 1H, *J* = 3.7 Hz), 3.23 (dd, 1H, *J* = 7.8, 9.2 Hz). ^{13}C NMR (150 MHz, CD_3OD) δ 102.19, 102.04, 101.79, 101.77, 98.51, 94.31, 86.63, 84.16, 83.31, 83.27, 77.78, 75.28, 75.12, 74.27, 72.88, 72.81, 72.69, 72.60, 71.90, 71.84, 71.69, 71.64, 63.77, 63.60, 62.68. HRESIMS: calcd for $[M+Na]^+$ $C_{16}H_{28}O_{10}$ + Na: 467.1371; found: 467.1385.

D-Xylose-*α*1-4-D-xylose-*α*1-3-D-glucopyranose (**2**)

MeOH (dry, 2 mL) was added to a vial containing 10% Pd/C (0.05 g). The resulting slurry was transferred to a flask containing *β*-*O*-benzyl (2,3,4-tri-*O*-benzyl-D-xylose)-(*α*1-4)-(2,3-di-*O*-benzyl-D-xylose)-(*α*1-3)-(2-*O*-benzyl-4,6-*O*-benzylidene)-D-glucopyranoside (0.008 g, 0.007 mmol). The flask was then placed under vacuum, back filled with H_2 three times, and then the reaction was stirred vigorously under H_2 for 3 h. Upon completion, the reaction was filtered through celite, concentrated, and flash column chromatography (EtOAc:MeOH, 5:2) gave the desired product, **2** as a colourless syrup (2.0 mg, 67 % yield).^{1}H NMR (600 MHz, D_2O) δ 5.38 (d, 1H, *J* = 3.7 Hz, H1^{C}*β*), 5.24 (d, 0.45H, *J* = 3.8 Hz, H1^{A}*α*), 5.12 (d, 1.45H, *J* = 3.6 Hz, H1^{B}*β*, H1^{B}*α*), 4.68 (d, 1H, *J* = 8.3 Hz, H1^{A}*β*), 4.65 (d, 1H, *J* = 4.3 Hz, H1^{C}*α*), 3.97-4.07 (m, 2.9H), 3.84-3.91 (m, 5H), 3.61-3.80 (m, 14H), 3.49-3.6 (m, 5H), 3.34-3.40 (m, 3H) ^{13}C NMR (150 MHz, D_2O) δ 100.03, 99.63, 98.66, 98.47, 95.91, 92.21, 81.26, 78.96, 78.23, 78.16, 75.61, 75.50, 72.87, 72.70, 72.34, 71.76, 71.51, 71.25, 71.18, 70.14, 70.11, 70.03, 69.19, 69.05, 65.21, 61.44, 60.46, 60.38, 60.28. HRESIMS: calcd for $[M+Na]^+$ $C_{16}H_{28}O_{10}$ + Na: 467.1371; found: 467.1382.

D-xylose-*α*1-2-D-xylose-*α*1-3-D-glucopyranose (**3**)

MeOH (dry, 2 mL) was added to a vial containing 10% Pd/C (0.05 g), the resulting slurry was transferred to a flask containing *β*-*O*-benzyl (2,3,4-tri-*O*-benzyl-D-xylose)-(*α*1-2)-(3,4-di-*O*-benzyl-D-xylose)-(*α*1-3)-(2-*O*-benzyl-4,6-*O*-benzylidene)-D-glucopyranoside (0.005 g, 0.004 mmol). The flask was then placed under vacuum, back filled with H_2 three times, and then the reaction was stirred vigorously under H_2 for 3 h. Upon completion, the reaction was filtered through celite, concentrated, and flash column chromatography (EtOAc:MeOH, 5:2) gave the desired product, **3** as a colourless syrup (0.0014 g, 74 % yield). ^{1}H NMR (600 MHz, D_2O) δ 5.37 (d, 1H, *J* = 4.0 Hz, H1^{C}*α*), 5.36 (d, 0.5H, *J* = 3.9 Hz, H1^{C}*β*), 5.33 (d, 1.5H, *J* = 3.8 Hz, H1^{B}*β*, H1^{B}*α*), 5.26 (d, 0.5H, *J* = 3.8 Hz, H1^{A}*α*), 4.67 (d, 1H, *J* = 8.0 Hz, H1^{A}*β*), 3.84-3.94 (m, 8H), 3.60-3.8 (m, 13.5H), 3.56 (dd,

1.5H, *J* = 3.9 Hz, *J* = 9.6 Hz), 3.50 (ddd, 1H, *J* = 2.3 Hz, *J* = 5.7 Hz, *J* = 10.1 Hz), 3.36 (dd, 1H, *J* = 8.3 Hz). ^{13}C NMR (150 MHz, D_2O) δ 98.77, 98.65, 98.51, 95.50, 91.77, 81.41, 78.94, 78.66, 78.64, 75.17, 72.52, 72.27, 71.17, 71.11, 70.75, 69.64, 69.58, 69.54, 69.25, 68.90, 61.09, 60.99, 60.05, 59.88. HRESIMS: calcd for $[M+Na]^+$ $C_{16}H_{28}O_{10}$ + Na: 467.1371; found: 467.1383.

5.6.6 Synthesis of *N*-(9-fluorenylmethoxycarbonyl)-3-*O*-(2,3,4,6-tetra-*O*-acetyl-*β*-D-glucopyranosyl)-L-serine pentafluorophenyl ester (35), *N*-(9-Fluorenylmethoxycarbonyl)-3-*O*-((2,3,4-tri-*O*-acetyl-D-xylose)-(*α*1-3)-(2,4-di-*O*-acetyl-D-xylose)-(*α*1-3)-2,4,6-tri-*O*-acetyl)-D-glucopyranosyl)-L-serine pentafluorophenyl ester (37), and XXG-Alkyne (39)

N-(9-Fluorenylmethoxycarbonyl)-3-*O*-(2,3,4,6-tetra-*O*-acetyl-*β*-D-glucopyranosyl)-L-serine pentafluorophenyl ester (**35**)

The pentafluorophenyl ester of *N*-(9-fluorenylmethoxycarbonyl)-L-serine (**34**) was synthesized as described by Kisfaludy and Schon[322] from commercially available pentafluorophenol and *N*-(9-fluorenylmethoxycarbonyl)-L-serine and used without further purification.

35 was assembled from commercially available 1,2,3,4,5-penta-*O*-acetyl-*β*-D-glucopyranose (0.5 g, 1.3 mmol) and the pentafluorophenyl ester of *N*-(9-fluorenylmethoxycarbonyl)-L-serine (**34**, 0.8 g, 1.7 mmol) all dissolved in DCM (dry, 20 mL), BF_3OEt_2 (0.48 mL, 3.9 mmol) was added and the reaction was stirred at room temperature for 16 h. Upon completion, the reaction mixture was diluted with DCM (20 mL), and the solution was washed with water (15 mL). The aqueous phase was extracted with DCM (15 mL), and the combined organic phases were dried ($MgSO_4$) and concentrated in vacuo. The residue was purified by column chromatography (EtOAc:hexanes, 3:7 to 4:6) to yield *N*-(9-fluorenylmethoxycarbonyl)-3-*O*-(2,3,4,6-tetra-*O*-acetyl-*β*-D-glucopyranosyl)-L-serine pentafluorophenyl ester (0.52 g, 50 % yield)[323]. ^{1}H and ^{13}C NMR spectra were in accordance with published data. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.77 (m, 2H, Ar), 7.61 (m, 2H, Ar), 7.41 (m, 2H, Ar), 7.32 (m, 2H, Ar), 5.74 (d, 1H, *J* = 8.4

Hz, NH), 5.25 (dd, 1H, *J* = 9.5 Hz, H3), 5.11 (dd, 1H, *J* = 9.6 Hz, H4), 5.00 (dd, 1H, *J* = 8.2 Hz, *J* = 9.5 Hz, H2), 4.88 (ddd, 1H, *J* = 3.1 Hz, *J* = 7.6 Hz, CH(Ser)), 4.57 (dd, 1H, *J* = 6.8 Hz, *J* = 10.6 Hz, CH_2(Fmoc)), 4.56 (d, 1H, *J* = 8.2 Hz, H1), 4.47 (dd, 1H, *J* = 6.8 Hz, *J* = 10.6 Hz, CH_2(Fmoc)), 4.43 (dd, 1H, *J* = 3.2 Hz, *J* = 10.4 Hz, CH_2(Ser)), 4.26 (dd, 1H, *J* = 6.7 Hz, CH(Fmoc)), 4.23 (dd, 1H, *J* = 4.7 Hz, *J* = 12.3 Hz, H6), 4.15 (dd, 1H, *J* = 2.1 Hz, *J* = 12.3 Hz, H6'), 4.00 (dd, 1H, *J* = 3.4 Hz, *J* = 10.5 Hz, CH_2(Ser)), 3.72 (ddd, 1H, *J* = 2.5 Hz, *J* = 4.6 Hz, *J* = 7.2 Hz, H5) 2.06 (s, 6H, OAc, OAc), 2.04 (s, 3H, OAc), 2.00 (s, 3H, OAc).

α-*O*-Acetyl (2,3,4-tri-*O*-acetyl-D-xylose)-(*α*1-3)- (2,4-di-*O*-acetyl-D-xylose)-(*α*1-3)-2,4,6-tri-*O*-acetyl)-D-glucopyranoside and *β*-*O*-acetyl (2,3,4-tri-*O*-acetyl-D-xylose)-(*α*1-3)- (2,4-di-*O*-acetyl-D-xylose)-(*α*1-3)-2,4,6-tri-*O*-acetyl)-D-glucopyranoside (**36**)

D-xylose-*α*1-3-D-xylose-*α*1-3-D-glucopyranose (0.30 g, 0.68 mmol) was dissolved in pyridine (10 mL), Ac_2O (5 mL) and DMAP (catalytic) was added and the reaction was stirred for 4 h at room temperature. Upon completion, the reaction was diluted with MeOH and allowed to stir for a further 30 min, the reaction was then concentrated and flash column chromatography gave a mixture an anomeric mixture (5:1, α:β) of *α*-*O*-acetyl (2,3,4-tri-*O*-acetyl-D-xylose)-(*α*1-3)-(2,4-di-*O*-acetyl-D-xylose)-(*α*1-3)-2,4,6-tri-*O*-acetyl)-D-glucopyranoside and *β*-*O*-acetyl (2,3,4-tri-*O*-acetyl-D-xylose)-(*α*1-3)-(2,4-di-*O*-acetyl-D-xylose)-(*α*1-3)-2,4,6-tri-*O*-acetyl)-D-glucopyranoside (**36**) as a white solid (0.40 g, 72 % yield) ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.33 (d, 1H, *J* = 3.7 Hz, $H1^A\alpha$), 5.65 (d, 0.2H, *J* = 8.2 Hz, $H1^A\beta$), 5.38 (dd, 1H, *J* = 9.8 Hz, $H3^B\alpha$), 5.37 (dd, 0.2H, *J* = 9.8 Hz, $H3^B\beta$), 5.35 (d, 1H, *J* = 3.6 Hz, $H1^B\alpha$), 5.34 (d, 0.2H, *J* = 3.6 Hz, $H1^B\beta$), 5.22-5.26 (m, 2.4H, $H1^C\alpha$, $H1^C\beta$, $H6^A\alpha$, $H6^A\beta$), 5.17 (dd, 0.2H, *J* = 8.3 Hz, *J* = 9.4 Hz, $H2^A\beta$), 5.10 (dd, 1H, *J* = 3.7 Hz, *J* = 10.0 Hz, $H2^A\alpha$), 4.95-5.05 (m, 2.4H, $H4^B\alpha$, $H4^B\beta$, $H4^C\alpha$, $H4^C\beta$), 4.76 (dd, 1.2H, *J* = 3.8 Hz, *J* = 10.3 Hz, $H2^B\alpha$, $H2^B\beta$), 4.72 (dd, 1.2H, *J* = 3.7 Hz, *J* = 10.1 Hz, $H2^C\beta$, $H2^C\alpha$), 4.25 (dd, 0.2H, *J* = 4.3 Hz, *J* = 11.7 Hz, $H6^A\beta$), 4.22 (dd, 1H, *J* = 4.3 Hz, *J* = 11.7 Hz, $H6^A\alpha$), 4.20 (dd, 1H, *J* = 9.6 Hz,

H3^{A}*α*), 4.06-4.12 (m, 2.4H, H4^{A}*α*, H4^{A}*β*, H3^{C}*α*, H3^{C}*β*), 3.70-3.82 (m, 3.8H, H5^{B}*α*, H5^{B}*β*, H5^{C}*α*, H5^{C}*β*, H5^{C}*α*, H5^{C}*β*, H5^{A}*β*), 3.64 (dd, 1H, *J* = 10.8 Hz, H5^{B}*α*), 3.52 (dd, 0.2H, *J* = 10.8 Hz, H5^{B}*β*), 2.22 (s, 1.2H, OAc*β*, OAc*β*), 2.21 (s, 3H, OAc*α*), 2.16 (s, 3H, OAc*α*), 2.15 (s, 0.6H, OAc*β*), 2.14 (s, 0.6H, OAc*β*), 2.13 (s, 6H, Oac*α*, OAc*α*), 2.12 (s, 0.6H, OAc*β*), 2.09 (s, 0.6H, OAc*β*), 2.08 (s, 3H, OAc*α*), 2.07 (s, 0.6H, OAc*β*), 2.04-2.06 (m, 13.2H, OAc*α*, OAc*α*, OAc*α*, OAc*β*, OAc*β*). ^{13}C NMR (125 MHz, $CDCl_3$) δ 170.70, 170.68, 170.53, 170.44, 170.38, 169.92, 169.88, 169.48, 169.45, 169.41, 169.13, 169.00, 168.81, 150.99, 96.12, 95.76, 95.61, 95.54, 91.91, 89.19, 73.48, 72.78, 72.47, 72.43, 72.18, 72.15, 71.65, 71.51, 71.05, 71.02, 70.44, 70.01, 69.48, 69.18, 69.13, 61.50, 58.63, 58.57, 58.46, 30.98, 29.73, 21.03, 20.96, 20.91, 20.87, 20.84, 20.81, 20.78, 20.71, 20.66, 20.59. HRESIMS: calcd for $[M+Na]^+$ $C_{34}H_{46}O_{23}$ + Na: 845.2322; found: 845.2331.

N-(9-Fluorenylmethoxycarbonyl)-3-*O*-((2,3,4-tri-*O*-acetyl-D-xylose)-(*α*1-3)-(2,4-di-*O*-acetyl-D-xylose)-(*α*1-3)-2,4,6-tri-*O*-acetyl)-D-glucopyranosyl)-L-serine pentafluorophenyl ester (**37**)

The *α*/*β* mixture of the per-*O*-acetylated trisaccharide **36** (0.20 g, 0.24 mmol) and the pentafluorophenyl ester of *N*-(9-fluorenylmethoxycarbonyl)-L-serine (0.18 g, 0.36 mmol) all dissolved in DCM (dry, 5 mL), was mixed with BF_3OEt_2 (0.15 mL, 1.44 mmol) and the reaction was stirred at room temperature for 48 h. Upon completion, the reaction mixture was diluted with DCM (5 mL), and the solution was washed with water (3 mL). The aqueous phase was extracted with DCM (5 mL), and the combined organic phases were dried ($MgSO_4$) and concentrated in vacuo. The residue was purified by column chromatography (EtOAc:hexanes, 3:7 to 4:6), after which the crude compound (10 mg/run) was loaded onto an Agilent Zorbax 300SB-C8 (9.4 mm 9 250 mm) semi-preparative HPLC column housed in an 1100 series Agilent HPLC. Compound **37** was purified using a linear gradient of 30% acetonitrile (ACN) to 95% ACN over 45 min operating at 2 mL/min. Fractions were collected using a

Foxy Jr. fraction collector set to collect 1.0 mL fractions over the entirety of the HPLC run. The major peak eluted at approximately 34 min and was collected in fractions 65-70. These fractions were pooled and concentrated to yield *N*-(9-fluorenylmethoxycarbonyl)-3-*O*-((2,3,4-tri-*O*-acetyl-D-xylose)-(α1-3)- (2,4-di-*O*-acetyl-D-xylose)-(α1-3)-2,4,6-tri-*O*-acetyl)-D-glucopyranosyl)-L-serine pentafluorophenyl ester (0.11 g, 35 % yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.78 (m, 2H, Ar), 7.63 (m, 2H, Ar), 7.42 (m, 2H, Ar), 7.34 (m, 2H, Ar), 5.72 (d, 1H, *J* = 8.5 Hz, NH), 5.36 (dd, 1H, *J* = 9.2 Hz, $H3^C$), 5.31 (d, 1H, *J* = 3.7 Hz, $H1^C$), 5.21 (d, 1H, *J* = 3.3 Hz, $H1^B$), 5.17 (dd, 1H, *J* = 9.7 Hz, $H4^A$), 4.92-5.03 (m, 3H, $H2^A$, $H4^B$, $H4^C$), 4.88 (ddd, 1H, *J* = 3.4 Hz, *J* = 8.4 Hz, CH(Ser)), 4.74 (dd, 1H, *J* = 3.7 Hz, *J* = 10.3 Hz, $H2^C$), 4.68 (dd, 1H, *J* = 3.6 Hz, *J* = 10.2 Hz, $H2^B$), 4.56 (dd, 1H, *J* = 6.8 Hz, *J* = 10.6 Hz, CH_2(Fmoc)), 4.38-4.47 (m, 3H, $H1^A$, CH_2(Ser), CH_2(Fmoc)), 4.26 (dd, 1H, *J* = 6.9 Hz, CH(Fmoc)), 4.17 (dd, 1H, *J* = 4.7 Hz, *J* = 12.4 Hz, $H6^A$), 4.08-4.12 (m, 1H, $H6'^A$), 4.06 (dd, 1H, *J* = 9.5 Hz, $H3^B$), 3.98 (dd, 1H, *J* = 3.4 Hz, *J* = 9.8 Hz, CH_2(Ser)), 3.88 (dd, 1H, *J* = 9.4 Hz, $H3^A$), 3.74-3.80 (m, 2H, $H5^C$, $H5'^C$), 3.70 (dd, 1H, *J* = 6.2 Hz, *J* = 10.8 Hz, $H5^B$), 3.55-3.62 (m, 1H, $H5^A$), 3.50 (dd, 1H, *J* = 11.0 Hz, $H5'^B$), 2.13 (s, 3H, OAc), 2.12 (s, 3H, OAc), 2.08 (s, 3H, OAc), 2.07 (s, 3H, OAc), 2.05 (s, 6H, OAc, OAc), 2.04 (s, 3H, OAc), 1.99 (s, 3H, OAc). ^{13}C NMR (150 MHz, $(CD_3)_2CO$ and $CDCl_3$) δ 170.63, 170.41, 170.37, 169.92, 169.89, 169.83, 169.39, 169.17, 166.07, 155.80, 143.63, 143.51, 141.34, 129.42, 129.07, 127.81, 127.10, 125.00, 124.94, 120.04, 100.52, 95.98, 95.55, 72.45, 72.15, 72.04, 71.81, 71.46, 70.99, 69.63, 69.17, 69.09, 68.21, 67.28, 65.85, 61.74, 58.55, 54.11, 51.61, 50.88, 47.09, 29.68, 20.89, 20.77, 20.74, 20.72, 20.65, 20.53, 15.25. HRESIMS: calcd for $[M+NH_4]^+$ $C_{56}H_{59}O_{26}NF_5$: 1256.3240; found: 1256.3253.

2-Propyne (2,3,4-tri-*O*-acetyl-D-xylose)-(α1-3)- (2,4-di-*O*-acetyl-D-xylose)-(α1-3)-2,4,6-tri-*O*-acetyl)-D-glucopyranoside (**38**)

Trisaccharide **36** (0.02 g, 0.024 mmol), and propargyl alcohol (0.0022 mL, 0.036 mmol) were dissolved in DCM (dry, 1 mL) and cooled to -20 ºC. BF_3OEt_2

(0.025 mL, 0.19 mmol) was added and the reaction was stirred for 3 h. Upon completion, reaction was diluted with 2 mL toluene, concentrated back to ~1 mL and loaded onto a flash chromatography column (Et_3N:EtOAc:hexanes 0.1:4:5.9) which gave 2-propyne (2,3,4-tri-*O*-acetyl-D-xylose)-(*α*1-3)- (2,4-di-*O*-acetyl-D-xylose)-(*α*1-3)-2,4,6-tri-*O*-acetyl)-D-glucopyranoside as a white solid (0.012 g, 60% yield) ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.27 (dd, 1H, *J* = 9.9 Hz, $H3^C$), 5.22 (d, 1H, *J* = 3.7 Hz, $H1^C$), 5.12 (d, 1H, *J* = 3.3 Hz, $H1^B$), 5.09 (dd, 1H, *J* = 9.5 Hz, $H4^A$), 4.95 (dd, 1H, *J* = 8.0 Hz, *J* = 9.3 Hz, $H2^A$), 4.82-4.90 (m, 2H, $H4^B$, $H4^C$), 4.65 (dd, 1H, *J* = 3.7 Hz, *J* = 10.2 Hz, $H2^C$), 4.59 (dd, 1H, *J* = 3.5 Hz, *J* = 10.0 Hz, $H2^B$), 4.57 (d, 1H, *J* = 8.0 Hz, $H1^A$), 4.28 (d, 1H, *J* = 2.3 Hz, OCH_2C), 4.13 (dd, 1H, *J* = 4.6 Hz, *J* = 12.3 Hz, $H6^A$), 4.03 (dd, 1H, *J* = 2.4 Hz, *J* = 12.3 Hz, $H6'^A$), 3.98 (dd, 1H, *J* = 9.7 Hz, $H3^B$), 3.84 (dd, 1H, *J* = 9.3 Hz, $H3^A$), 3.66-3.70 (m, 2H, $H5^C$, $H5'^C$), 3.61 (dd, 1H, *J* = 6.0 Hz, *J* = 10.9 Hz, $H5^B$), 3.52 (ddd, 1H, *J* = 2.6 Hz, *J* = 4.5 Hz, *J* = 9.8 Hz, $H5^A$), 3.43 (dd, 1H, *J* = 10.9 Hz, $H5'^B$), 2.36 (dd, 1H, *J* = 2.3 Hz, CCH), 2.04 (s, 3H, OAc), 2.03 (s, 3H, OAc), 2.02 (s, 3H, OAc), 2.00 (s, 3H, OAc), 1.98 (s, 3H, OAc), 1.96 (s, 3H, OAc), 1.95 (s, 3H, OAc), 1.94 (s, 3H, OAc). ^{13}C NMR (125 MHz, $CDCl_3$) δ 170.80, 170.53, 170.45, 170.03, 169.98, 169.93, 169.53, 169.34, 98.20, 95.94, 95.52, 78.20, 75.38, 72.45, 72.15, 71.95, 71.55, 71.50, 70.97, 69.81, 69.17, 69.10, 61.74, 58.51, 58.43, 55.90, 29.74, 21.06, 20.99, 20.85, 20.81, 20.75, 20.62. HRESIMS: calcd for $[M+NH_4]^+$ $C_{35}H_{46}O_{22}$ + NH_4: 836.2819; found: 836.2818.

2-Propyne (D-xylose)-(*α*1-3)-(D-xylose)-(*α*1-3)-D-glucopyranoside (**39**)

To a stirred solution of 2-propyne (2,3,4-tri-*O*-acetyl-D-xylose)-(*α*1-3)-(2,4-di-*O*-acetyl-D-xylose)-(*α*1-3)-2,4,6-tri-*O*-acetyl)-D-glucopyranoside (0.02 g, 35.4 mmol) in methanol (dry, 80 mL) was added NaOMe (catalytic). The mixture was allowed to stir for 4 h and was then neutralized with methanol:AcOH (19:1) and stirred for a further 30 min. The reaction mixture was adsorbed onto silica and loaded onto a flash column chromatography column (EtOAc:MeOH, 5:1), which afforded 2-propyne (D-xylose)-(*α*1-3)-(D-xylose)-(*α*1-3)-D-glucopyranoside as a

white solid (0.01 g, 85% yield) ^{1}H NMR (600 MHz, D_2O) δ 5.27 (d, 1H, *J* = 3.7 Hz, $H1^{C}$), 5.26 (d, 1H, *J* = 3.8 Hz, $H1^{B}$), 4.60 (d, 1H, *J* = 8.0 Hz, $H1^{A}$), 4.44 (d, 1H, *J* = 14.9 Hz, CH_2), 4.40 (d, 1H, *J* = 14.9 Hz, CH_2), 3.72-3.82 (m, 5H), 3.52-3.67 (m, 8H), 3.48 (d, 1H, *J* = 3.9 Hz, *J* = 9.7 Hz, $H2^{B}$), 3.41-3.44 (m, 1H), 3.34-3.37 (m, 1H, $H2^{A}$). ^{13}C NMR (150 MHz, $CDCl_3$) δ 100.17, 98.70, 98.51, 81.36, 78.64, 75.21, 72.52, 71.17, 71.07, 69.57, 69.48, 69.26, 68.89, 61.08, 60.98, 59.98, 56.11. HRESIMS: calcd for $[M+H]^+$ $C_{19}H_{30}O_{14}$ + H: 482.1636; found: 482.1636.

S-D-Biotinoylated (D-xylose)-(*α*1-3)-(D-xylose)-(*α*1-3)-D-glucopyranoside (**40**)

2-Propyne (D-xylose)-(*α*1-3)-(D-xylose)-(*α*1-3)-D-glucopyranoside (**39**) (1.0 mg, 2.1 µmol) was dissolved in phosphate buffer (400 µL, 100 mM, pH 7.4) and added to azide-PEG4-biotin conjugate (1.0 mg, 2.3 µmol, Click Chemistry Tools). A 15 µL mixture of $CuSO_4$ and the ligand THPTA (7 mM, and 33 mM respectively) was added to the solution containing compound **39**. Aminoguanidine hydrochloride (25 µL, 100 mM) and sodium ascorbate (25 µL, 100 mM, prepared fresh) were added and the reaction was shaken at 37 ºC for 4 h[329]. Upon completion, as judged by TLC, the reaction mixture was diluted with water (1.0 mL) and was loaded onto an Agilent Zorbax 300SB-C8 (9.4 mm 9 250 mm) semi-preparative HPLC column housed in an 1100 series Agilent HPLC. Compound **40** was purified using a linear gradient of 2% acetonitrile (ACN) to 40% ACN over 40 min operating at 2 mL/min. Fractions were collected using a Foxy Jr. fraction collector set to collect 1.0 mL fractions over the entirety of the HPLC run. The major peak eluted at approximately 26 min and was collected in fraction 26. These fractions were pooled and concentrated to yield S-D-Biotinoylated (D-xylose)-(*α*1-3)-(D-xylose)-(*α*1-3)-D-glucopyranoside (**40**) (0.9 mg, 47% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ ^{13}C NMR (150 MHz, $CDCl_3$) δ 101.58, 99.19, 98.99, 81.93, 79.13, 75.67, 74.07, 73.00, 71.82, 71.68, 71.65, 71.44, 70.06, 70.01, 69.74, 69.66, 69.60, 69.47, 69.42, 69.37, 68.85, 68.73, 62.07, 61.99, 61.57, 61.47, 60.56, 60.24, 60.20, 58.04, 55.33, 50.05, 43.13, 39.68,

38.91, 35.66, 35.45, 30.40, 27.85, 27.67, 25.12, 25.03, 23.26, 21.64, 13.18. HRESIMS: calcd for $[M+H]^+$ $C_{37}H_{63}O_{19}N_6S$: 927.3863; found: 927.3855.

5.6.7 Synthesis of G-Pep and XXG-Pep

Peptide synthesis of G-Pep (**6**) and XXG-Pep (**5**) modified peptides

Peptide synthesis was performed manually using pre-loaded Fmoc-Pro-Rink amide resin [4-(2',4'-dimethoxyphenyl-Fmoc-aminomethyl) phenoxy resin; Chem-Impex International] with substitution levels of 0.7 mmol/g. Standard Fmoc synthesis methods were used (Atherton, E., and Sheppard, R.C. *Solid-phase peptide synthesis, a practical approach*; IRL Press: Oxford, 1989), with dimethylformamide (DMF) used as the wash, deprotection, and coupling solvent. Fmoc deprotection was done using 20% piperidine in DMF two times with shaking for 20 min. Amino acids were coupled using 3 equivalents of amino acid, 3 equivalents O-(benzotriazole-1-yl)-N,N,N',N'-tetramethyluronium (HBTU), and 6 equivalents of *N,N*-diisopropylethylamine (DIEA), two times with shaking for 20-60 min, depending on the amino acid. Amino acids used were Fmoc-Ala, Fmoc-Asn(Trt), Fmoc-Ser(tBu) (all amino acids were purchased from Chem-Impex International), a synthetically prepared pentafluorophenyl (Pfp) activated per-*O*-acetylated Glc (Ac-Glc)modified serine (Fmoc-Ser(Ac-Glc)-*O*-Pfp) and a synthetically prepared pentafluorophenyl (Pfp) activated per-*O*-acetylated XXG (Ac-XXG)modified serine (Fmoc-Ser(Ac-XXG)-*O*-Pfp). The Glc and XXG modified peptide incorporated Fmoc-Ser(Ac-Glc)-*O*-Pfp or Fmoc-Ser(Ac-XXG)-*O*-Pfp into the peptide using the following, slightly altered coupling procedure: 0.8 equivalents Fmoc-Ser(Ac-Glc)-*O*-Pfp or Fmoc-Ser(Ac-XXG)-*O*-Pfp, and 6 equivalents DIEA, two times with shaking for 180 min. The N-terminal alkyne was attached using 3 equivalents of 4-pentynoic acid, 3 equivalents HBTU, and 6 equivalents DIEA, two times with shaking for 60 min. The Glc and XXG modified peptides were de-acetylated using NaOMe in methanol at a pH of ~10 for 36 h. The resin was then washed with methanol three times and then resuspended in a solution containing trifluoroacetic acid (TFA), H_2O, triethylsilane, and 1,2-

ethanedithiol (EDT) (in a ratio of 9.4:0.25:0.1:0.25) and was allowed to shake for 2 h. The deprotected peptide was then filtered through a sintered glass frit, washed twice with fresh TFA and concentrated in vacuo using a high vacuum rotary evaporator.

Purification of the Glc and XXG modified peptides

The crude peptide (15 mg/run) was loaded onto an Agilent Zorbax 300SB-C8 (9.4 mm 9 250 mm) semi-preparative HPLC column housed in an 1100 series Agilent HPLC. The peptide was purified using a linear gradient of 5% acetonitrile (ACN) to 40% ACN over 40 min operating at 2 mL/min. Fractions were collected using a Foxy Jr. fraction collector set to collect 0.666 mL fractions over the entirety of the HPLC run. The major peak eluted at approximately 13.6 min and was collected in fractions 40–42. These fractions were pooled and high-resolution mass spectrometry was carried out to verify the correct identity of the peptide.

G-Pep (**6**) HRESIMS calcd for $[M+H]^+$ $C_{29}H_{45}N_7O_{13}$ + H: 700.3148 Da, found: 700.3156 Da.

XXG-Pep (**5**) HRESIMS calcd for $[M+H]^+$ $C_{39}H_{61}N_7O_{21}$ + H: 964.3929 Da, found: 964.4085 Da.

6: Conclusion

Glycans are now known to play critical roles in cell to cell signalling[13,149], cellular adhesion[2-4], and protein folding[5-8]. The enzymes that assemble and disassemble glycans are pharmaceutical targets for various disease states such as cancer[330,331], lysosomal storage disorders[332,333], and diabetes[67,334], as well as viral infections[335,336]. This evolution in our understanding of carbohydrates has occurred over the past couple of decades in part because of the continuing development of advancements in detection and analysis of glycans. For example, glycan specific lectin or antibody affinity columns, methods for detecting and analysing glycans, and the increasing availability of pure synthetic glycans have all aided direct glycan analysis[37,85,91-93]. In addition the mechanism of action of glycosidases and glycosyltransferases have been elucidated using X-ray crystallography[63-65], studying kinetic isotope effects[34,36,51], and LFERs[34,337,338]. The development of tools, techniques and methods have been important to the evolution of glycobiology. In this thesis efforts towards new tools and methods for studying *O*-GlcNAcylation (Chapter 2 and 3) and *O*-glucosylation (Chapter 4 and 5) are presented.

6.1 Development of potent and selective inhibitors of OGA and transition state analogy studies on NAG-thiazoline and PUGNAc

The discovery in 1984 that proteins in the nucleus and cytoplasm were modified by *O*-GlcNAc[128], and that this modification was a dynamic process[133,134], led to immediate attention on this unusual modification. This post-translational modification is abundant in mammalian cells[128] and appears to have roles in the etiology of several disease states including type II diabetes[139,192], Alzheimer's[190,193,194], and cancer[195]. However a clear understanding of the precise roles played by the enzymes involved in dynamic *O*-GlcNAc processing has been elusive. Only limited knowledge was available regarding the

mechanism of action of both OGT and *O*-GlcNAcase at the time this thesis was started. Likewise inhibitors selective for *O*-GlcNAcase over functionally related enzymes (HexA and HexB) were unknown but were needed to probe the role of increased *O*-GlcNAcylation on various disease states such as type II diabetes.

The catalytic mechanism of *O*-GlcNAcase was discussed in Chapter 2 and compared to human HexB. We compared the effect of several substrate analogues bearing increasing levels of fluorine substitution within the *N*-acetyl group. A Taft-like linear free energy analysis gave a negative slope for both human HexB and human *O*-GlcNAcase, thus revealing *O*-GlcNAcase, a GH84 enzyme, is a retaining enzyme using substrate-assisted catalysis like the GH20 human lysosomal enzymes. On the basis of the mechanistic information, which indicated there were differences in active site structures between OGA and HexA/B, efforts were made to generate selective OGA inhibitors.

We found that PUGNAc, synthesized by Vasella and coworkers[43], and NAG-thiazoline, synthesized by Knapp and coworkers[44], were not selective yet we felt these compounds could be elaborated to gain the desired selectivity. A series of NAG-thiazoline[235] and PUGNAc[245] analogues were generated by installing different *N*-acyl groups at the 2-amino position. The NAG-thiazoline series of inhibitors was found to have up to 1500-fold selectivity for human *O*-GlcNAcase while remaining a potent inhibitor. Unfortunately the PUGNAc series of inhibitors rapidly lost potency as the aliphatic chain length and steric bulk was increased and did not show significant selectivity[245,246].

With multiple good inhibitors in hand we delved into the basis for the observed potency and selectivity. Looking at the catalytic mechanism proposed in Chapter 2 for human *O*-GlcNAcase, there are clear similarities between the structure of the oxazoline intermediate and the thiazoline inhibitors. We postulated that not only did NAG-thiazoline resemble the intermediate, but perhaps also the transition state since the sulfur-carbon bond of the thiazoline ring is longer than the carbon oxygen bond in the oxazoline ring. A transition state study using linear free energy relationships supports the view that these thiazolines mimic either the transition state or a species close in energy, whereas

PUGNAc, previously proposed to be a transition state analogue, appears to be either a poor TS-analogue or a serendipitous binding inhibitor.

Crystal structures of a bacterial homologue of GH84 *O*-GlcNAcase revealed a large pocket directly below the *N*-acetyl group that is absent in the structure of the GH20 human HexB. X-ray crystal structures of *O*-GlcNAcase bound to either PUGNAc or NAG-thiazoline revealed the *N*-acetyl methyl group of NAG-thiazoline points directly into the open pocket, whereas the *N*-acetyl methyl group of PUGNAc was directed slightly away from the centre of the pocket indicating extension of the acyl chain could lead to steric clashes within the active site and accounting for the lower potency of bulkier PUGNAc analogues (Figure 3.6, Chapter 3).

These thiazoline inhibitors have spawned within the laboratory a second generation of more potent and selective inhibitors[294]. Using these various thiazoline inhibitors, a link between *O*-GlcNAcylation and Alzhiemer's disease has been made[294,339], the proposal that increased *O*-GlcNAc levels lead to insulin resistance has been challenged and other researchers in many laboratories have adopted the use of these inhibitors and used them to address various questions[340,341].

6.1.1 Future directions

As discussed in detail in the current state of the field section (Appendix 1), ThiametG and GlcNAcstatin are potent and selective inhibitors of OGA. These two inhibitors are ideal for studying the physiological role of the *O*-GlcNAc modification, in this respect the development of new inhibitors of OGA as academic tools is no longer a priority. Also the mechanism of OGA has been established (Chapter 2) along with a preliminary understanding of the transition state structure (Chapter 3 and Macauley *et al*.[248]). However a more comprehensive KIE study measuring primary, *α*-secondary, and *β*-secondary KIEs would be useful in further characterizing the geometry, charge distribution, and bond association/dissociation at the transition state. A KIE study of this sort is the most widely used method of differentiating between concerted and

dissociative mechanisms[342] and would clarify details of the mechanistic proposals for human OGA made by Greig *et al*.[338]

6.2 Elucidation of the structure of the trisaccharide modifying EGF-like repeats of mammalian Notch

The second body of work involved efforts to generate tools to probe the non-canonical *O*-glucose modification of Notch EGFs. *O*-glucosylation of proteins was discovered in 1988 by Hase and coworkers; this modification was found on blood coagulation factors VII and IX and was proposed to be elongated to the XXG trisaccharide[145,146]. Recent work by the Haltiwanger group has shown that mammalian Notch, a member of the essential Notch signaling pathway, is also modified with *O*-glucose[148]. Glucose is then glycosylated to form a trisaccharide. At the time this thesis was undertaken the identity of the terminal sugar moieties and the stereochemistry of the glycosidic linkages were not established. Rumi, the transferase responsible for adding the glucose moiety to Notch, was only recently identified in *Drosophila*[175]. Loss of Rumi activity leads to a temperature sensitive Notch phenotype, which points to *O*-glucosylation of Notch playing a critical role in proper Notch structure or function[166,175]. These findings have stimulated interest in the field but additional tools are needed to characterize these glycans and further explore this phenomenon.

A series of di- and tri-saccharide standards were synthesized and used to identify the trisaccharide modifying mammalian Notch. An α-xylosidase capable of cleaving the terminal xylose units from the trisaccharide was identified and used in conjunction with the synthetic di- and tri-saccharide standards, CE, and MS analysis to unequivocally identify the trisaccharide modifying mammalian Notch as D-Xyl-α1-3-D-Xyl-α1-3-D-Glc (XXG). The di- and tri-saccharide standards D-Xyl-α1-3-D-Glc (XG); XXG; D-Xyl-α1-2-D-Xyl-α1-3-D-Glc (X1-2XG); and D-Xyl-α1-4-D-Xyl-α1-3-D-Glc (X1-4XG), along with the α-xylosidase and the CE method developed, can be used to clarify whether any protein is modified by the XXG trisaccharide. The method should therefore prove useful for

unambiguously identifying the XXG glycan on other proteins containing EGF repeats.

After establishing the identity of the trisaccharide modifying Notch we decided to focus on the development of polyclonal antibodies raised against the XXG glycan (Chapter 5). Larger scale synthesis of XXG was carried out to generate a peptide incorporating XXG and a terminal alkyne (XXG-Pep) as well as XXG coupled to an alkyne linker (XXG-Alkyne). Incorporation of an alkyne functionality into both the XXG-Pep and the XXG-Alkyne was necessary for CuAAC chemistry. Finn and co-workers have successfully used the bacteriophage Qβ capsid conjugated, using CuAAC, to Man_4, Man_8, and Man_9 and used these conjugates to elicit a strong immunogenic response to high mannose glycans[316]. Often, carbohydrate based antigens are poor immunogens and therefore this increased carbohydrate-specific antibody production is encouraging and most likely stems from bacteriophage Qβ capsids being both highly immunogenic and having a patterned antigen display on their surface[119]. A collaboration with Professor Finn and Dr. Laufer at Scripps Research Institute in San Diego, California was initiated. We separately coupled each antigen onto the Qβ virus particle capsid using CuAAC and the resulting highly immunogenic antigens were used to immunize mice. It is anticipated that any resulting XXG selective antibodies will be powerful tools for studying the biological role of the XXG modification.

6.2.1 Future directions

The *O*-glucose modification of Notch is a relatively new field of scientific research, much work is needed to fully comprehend the significance of this glycan and the enzymes that construct it. For example the glucosyltransferase[175] and the xylosyltranseferases[313] responsible for adding the first xylose moiety have been identified, however the identitiy of the xylosyltranseferase(s) responsible for adding the terminal xylose unit is still unknown. Further, as discussed in the current state of the field sections (Appendix 2) the presence of the XXG trisaccharide is essential for proper cleavage at the S2 site in the

NRR[312]. The exact role that XXG plays, however, remains unknown. One of the major focuses of this thesis was the development of tools for studying this modification, future work in this area will certainly revolve around the characterization of XXG specific antibodies. The antibodies developed in collaboration with Professor Finn and Dr. Laufer at the Scripps research institute need to be characterized to assess their selectivity toward XXG. Development of XXG specific antibodies could be used to probe the role of XXG in disease states associated with improper Notch signaling as well as to examine the localization and abundance of XXG at different stages of development. Studies using a biotin tagged synthetic XXG trisaccharide (Figure 6.1) immobilized on streptavidin beads could be used to probe mammalian tissues to identify potential proteins/lectins that bind the XXG trissacharide. The development of XXG selective antibodies or the discovery of an XXG specific lectin would be an exciting culmination of my efforts towards the elucidation of the function of XXG modification of Notch.

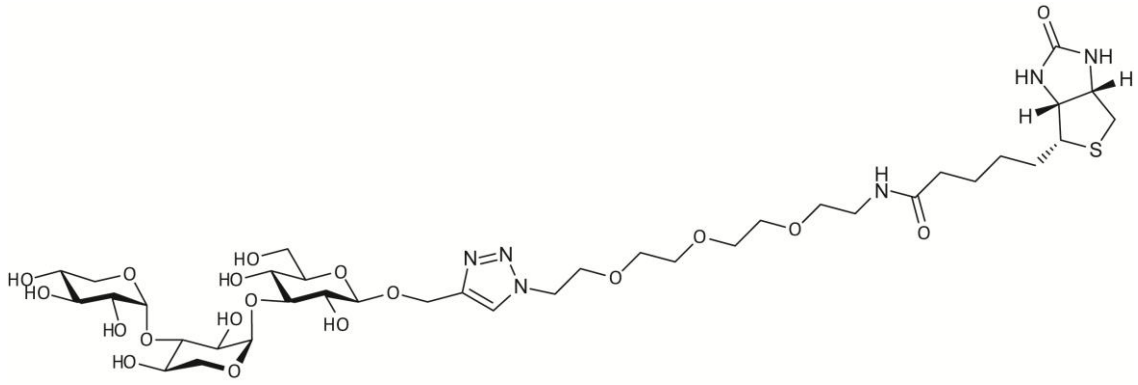


Figure 6.1: XXG-biotin

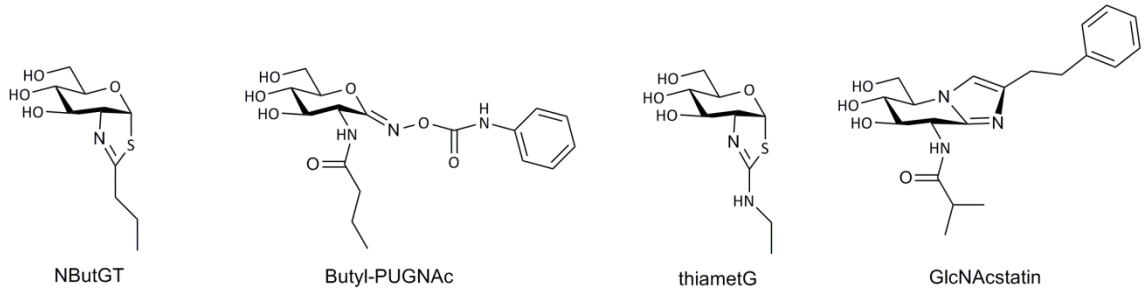
In summary, I feel that the series of selective *O*-GlcNAcase inhibitors coupled with the elucidation of the catalytic mechanism of *O*-GlcNAcase and the transition state studies, have significantly contributed towards the current understanding of *O*-GlcNAcylation and have enabled the development of still more potent inhibitors that are seeing use within the community as biological tools. The tools created and currently in development for studying *O*-

glucosylation and the XXG trisaccharide have been useful in my studies and will likely prove valuable in future work directed towards clarifying the roles of this glycan.

Appendices

Appendix 1: Current state of the field; New O-GlcNAcase inhibitors

The year following our development of selective OGA inhibitors based on the NAG-thiazoline scaffold[235], Stubbs *et al*. and Kim *et al*. reported analogous modifications to the structure of PUGNAc[245,246]. Interestingly, these derivatives were not as selective as the NAG-thiazoline derivatives for OGA over the functionally related *β*-hexosaminidase. For example, the butyl derivative of NAG-thiazoline (NButGT, Table 3.3) has a K_I of 250 nM and is 1500-fold selective for OGA, while the butyl derivative of PUGNAc (Butyl-PUGNAc, *O*-(2-deoxy-2-butamido-D-glucopyranosylidene) amino *N*-phenylcarbamate, Figure 3.7) has a K_I of 2400 nM and is only 11-fold selective for OGA (Table 3.3)[235,245]. This incongruity can be explained by the structure of OGA bound to NAG-thiazoline or PUGNAc. The alkyl group of NAG-thiazoline is oriented directly down into the active site pocket present in OGA[337] whereas the *N*-acetyl methyl group of PUGNAc is oriented at an angle and appears to be directed towards the wall of the OGA active site pocket[337].



Appendix Figure 1: Recently developd OGA inhibitors and NButGT

NButGT (left), Butyl-PUGNAc (left-centre), thiametG (right-centre), GlcNAcstatin (right)

Appendix Table 1: Inhibition constants and selectivity of recently developed inhibitors for both O-GlcNAcase and β-hexosaminidase

Compound	*O*-GlcNAcase K_I (μM)	HexB K_I (μM)	Selectivity Ratio (HexB K_I / *O*-GlcNAcase K_I)
NButGT	0.23	340	1500
Butyl-PUGNAc	2.4	26	11
ThiametG	0.021	750	37000
GlcNAcstatin	0.004	0.55	550

The NAG-thiazoline scaffold, the data acquired from the characterization of NAG-thiazoline as a transition state analogue, and detailed enzyme kinetic studies of the pH dependence of OGA and OGA mutants was used by Yuzwa *et al.* to develop a more potent and selective, second generation inhibitor called ThiametG (1,2-dideoxy-2'-ethylamino-α-D-glucopyranoso-[2,1-d]-Δ2'-thiazoline, Figure 3.7)[294]. ThiametG has a K_I of 21 nM and is 37000-fold selective for human OGA over the lysosomal *β*-hexosaminidase. The increased potency of this second generation inhibitor can be attributed to a highly favourable ionic interaction between the exocyclic nitrogen of ThiametG and the key Asp242 catalytic residue located within the active site[294]. Another potent OGA inhibitor, GlcNAcstatin ((5R,6R,7R,8S)-*N*-[6,7-dihydroxy-5-(hydroxymethyl)-2-(2-phenylethyl)-1,5,6,7,8,8a-hexahydroimazao[1,2-a]pyridine-8-yl]-2-methylpropanamide, Figure 3.7), was developed by Dorfmueller *et al*.[343] GlcNAcstatin acts in cells and has a moderate 150-fold selectivity for human OGA over human HexB, and is very potent with a K_I value of 4 nM and 550 nM for human OGA and HexB respectively[344]. More recently 6-acetamido castanospermine has been shown to be a good nanomolar inhibitor of OGA that, while showing no selectivity, works effectively in cells to increase *O*-GlcNAc levels[340]. Other GlcNAcstatin analogues have more recently been developed that show improved selectivity but these incorporate chemically reactive electrophilic groups that could circumscribe their use in cells and tissues[345]. ThiametG and GlcNAcstatin are thus good examples of next generation inhibitors that have benefitted from mechanistic and structural data gleaned from the previous generation of inhibitors.

Appendix 2: Current state of the field; Function of *O*-glucosylation

In 2008, Acar *et al*. identified Rumi as a protein *O*-glucosyltransferase in *Drosopihila*, that is capable of adding glucose to specific serine residues of Notch.[175] While the transferase activity of Rumi was shown to be essential for proper Notch function, *O*-glucosylation was not necessary for proper binding to Notch by the Notch ligand Delta. However, *O*-glucosylation was required for S2 cleavage catalyzed by ADAM10. Gordon *et al*. have shown that, after ligand binding to Notch, the notch regulatory region (NRR) undergoes a conformational change that must occur in order for S2 cleavage to occur[158,159]. These data point to *O*-glucosylation of Notch playing a role in either the conformational change of the NRR or the susceptibility to cleavage by ADAM10. The identification of Rumi as an *O*-glucosyltransferase required for proper Notch activation appears to have stimulated an increased focus on elucidating the function of Notch *O*-glucosylation.

In 2011 a study found that *Rumi*$^{-/-}$ mouse embryos die before day 9.5 with posterior axis truncation and severe defects in neural tube development, somitogenesis, cardiogenesis, and vascular remodelling[311]. Also, Jagged 1 induced Notch signaling is decreased when either Rumi activity is decreased or Notch EGF *O*-glucosylation sites are lost. The authors suggest this data supports mammalian Notch signalling being modulated by Rumi[311]. Rana *et al.* comprehensively mapped the XXG glycan sites on Notch and explored the effect that loss of mammalian Notch *O*-glucosylation has on ligand binding. They show that single serine to alanine mutations of all the predicted consensus sites on Notch 1 individually have little to no effect on ligand binding[310]. They also observed near stoichiometric glucosylation at consensus sites, efficient elongation of glucose to the XXG trisaccharide, and proposed a slightly revised consensus sequence governing *O*-glucosylation consisting of Cys^{1}-Xxx-Ser-Xxx-(Pro/Ala)-Cys^{2} [310]. Using *Drosophila*, Leonardi and coworkers corroborated the lack of impact on notch function when single mutations of the serine residue of the *O*-glucose consensus sites were made[312]. However, they propose that *O*-

glucosylation of EGF repeats of Notch is required to maintain the Notch ECD in its proper conformation and that sequential loss of additional glucosylation sites gradually prevents S2 cleavage[312]. Elucidating the role of the XXG glycan will further clarify the complex Notch signaling pathway and could offer insight into some of the disease states associated with improper functioning of this critical signaling pathway.

References

1. Painter, T. J.; Watkins, W. M.; Morgan, W. T., Serologically active fucose-containing oligosaccharides isolated from human blood-group A and B substances. *Nature* **1965,** 206, (984), 594-7.

2. Finger, E. B.; Puri, K. D.; Alon, R.; Lawrence, M. B.; von Andrian, U. H.; Springer, T. A., Adhesion through L-selectin requires a threshold hydrodynamic shear. *Nature* **1996,** 379, (6562), 266-9.

3. Lawrence, M. B.; Springer, T. A., Leukocytes roll on a selectin at physiologic flow rates: distinction from and prerequisite for adhesion through integrins. *Cell* **1991,** 65, (5), 859-73.

4. von Andrian, U. H.; Chambers, J. D.; McEvoy, L. M.; Bargatze, R. F.; Arfors, K. E.; Butcher, E. C., Two-step model of leukocyte-endothelial cell interaction in inflammation: distinct roles for LECAM-1 and the leukocyte beta 2 integrins in vivo. *Proc Natl Acad Sci U S A* **1991,** 88, (17), 7538-42.

5. Hammond, C.; Helenius, A., Folding of VSV G protein: sequential interaction with BiP and calnexin. *Science* **1994,** 266, (5184), 456-8.

6. Ou, W. J.; Cameron, P. H.; Thomas, D. Y.; Bergeron, J. J., Association of folding intermediates of glycoproteins with calnexin during protein maturation. *Nature* **1993,** 364, (6440), 771-6.

7. Sousa, M. C.; Ferrero-Garcia, M. A.; Parodi, A. J., Recognition of the oligosaccharide and protein moieties of glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *Biochemistry* **1992,** 31, (1), 97-105.

8. Suh, K.; Bergmann, J. E.; Gabel, C. A., Selective retention of monoglucosylated high mannose oligosaccharides by a class of mutant vesicular stomatitis virus G proteins. *The Journal of cell biology* **1989,** 108, (3), 811-9.

9. Fischer, E., Ueber die Configuration des Traubenzuckers und seiner Isomeren. II. *Ber. Disch. Chem. Ges.* **1891,** 24.

10. Fischer, E., Uber die Konfiguration des Traubenzuckars und seiner Isomeren. II. *Ber. Disch. Chem. Ges.* **1891,** 24.

11. Smith, M., March J., *March's advanced organic chemistry: Reactions, mechanisms, and structure*. 5th ed.; Wiley-Interscience: 2001.

12. Rampal, R.; Arboleda-Velasquez, J. F.; Nita-Lazar, A.; Kosik, K. S.; Haltiwanger, R. S., Highly conserved O-fucose sites have distinct effects on Notch1 function. *J Biol Chem* **2005,** 280, (37), 32133-40.

13. Stahl, M.; Uemura, K.; Ge, C.; Shi, S.; Tashima, Y.; Stanley, P., Roles of Pofut1 and O-fucose in mammalian Notch signaling. *J Biol Chem* **2008,** 283, (20), 13638-51.

14. Reitsma, S.; Slaaf, D. W.; Vink, H.; van Zandvoort, M. A.; oude Egbrink, M. G., The endothelial glycocalyx: composition, functions, and visualization. *Pflugers Arch* **2007,** 454, (3), 345-59.

15. Davies, G. J.; Gloster, T. M.; Henrissat, B., Recent structural insights into the expanding world of carbohydrate-active enzymes. *Curr Opin Struct Biol* **2005,** 15, (6), 637-45.

16. Henrissat, B., A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* **1991,** 280 (Pt 2), 309-16.

17. Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B., The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic acids research* **2009,** 37, (Database issue), D233-8.

18. Lindahl, U.; Backstrom, G.; Malmstrom, A.; Fransson, L. A., Biosynthesis of L-iduronic acid in heparin: epimerization of D-glucuronic acid on the polymer level. *Biochem Biophys Res Commun* **1972,** 46, (2), 985-91.

19. Kornfeld, R.; Kornfeld, S., Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* **1985,** 54, 631-64.

20. Oman, T. J.; Boettcher, J. M.; Wang, H.; Okalibe, X. N.; van der Donk, W. A., Sublancin is not a lantibiotic but an S-linked glycopeptide. *Nat Chem Biol* **2011,** 7, (2), 78-80.

21. Hofsteenge, J.; Muller, D. R.; de Beer, T.; Loffler, A.; Richter, W. J.; Vliegenthart, J. F., New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us. *Biochemistry* **1994,** 33, (46), 13524-30.

22. Charnock, S. J.; Davies, G. J., Structure of the nucleotide-diphospho-sugar transferase, SpsA from Bacillus subtilis, in native and nucleotide-complexed forms. *Biochemistry* **1999,** 38, (20), 6380-5.

23. Sun, H. Y.; Lin, S. W.; Ko, T. P.; Pan, J. F.; Liu, C. L.; Lin, C. N.; Wang, A. H.; Lin, C. H., Structure and mechanism of Helicobacter pylori fucosyltransferase. A basis for lipopolysaccharide variation and inhibitor design. *J Biol Chem* **2007,** 282, (13), 9973-82.

24. Offen, W.; Martinez-Fleites, C.; Yang, M.; Kiat-Lim, E.; Davis, B. G.; Tarling, C. A.; Ford, C. M.; Bowles, D. J.; Davies, G. J., Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *Embo J* **2006,** 25, (6), 1396-405.

25. Klassen, J. S.; Soya, N. S., N.; Fang, Y.; Palcic, M. M., Trapping and characterization of covalent intermediates of mutant retaining glycosyltransferases. *Glycobiology* **2011,** 21, (5), 547-52.

26. Lairson, L. L.; Chiu, C. P. C.; Ly, H. D.; He, S. M.; Wakarchuk, W. W.; Strynadka, N. C. J.; Withers, S. G., An active site mutant of the retaining glycosyltransferase LgtC from Neisseria meningitidis contains a covalently bound galactosyl-enzyme intermediate and reveals an alternative candidate catalytic nucleophile. *Glycobiology* **2004,** 14, (11), 28339-44.

27. Hosfield, D. J.; Zhang, Y. M.; Dougan, D. R.; Broun, A.; Tari, L. W.; Swanson, R. V.; Finn, J., Structural basis for bisphosphonate-mediated inhibition of isoprenoid biosynthesis. *J Biol Chem* **2004,** 279, (10), 8526-9.

28. Amzel, L. M.; Gabelli, S. B.; McLellan, J. S.; Montalvetti, A.; Oldfield, E.; Docampo, R., Structure and mechanism of the farnesyl diphosphate synthase from Trypanosoma cruza: Implications for drug design. *Proteins: Struct Funct Bioinf* **2006,** 62, (1), 80-8.

29. Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G., Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **2008,** 77, 521-55.

30. Cowdrey, W. A., Hughes, E.D., Ingold, C.K., Masterman, S., and Scott, A.D., Reaction kinetics and the Walden inversion. Part VI. Relation of steric orientationto mechanisms in substitutions involving halogen atoms and simple or substituted hydroxyl groups. *J. Chem. Soc.* **1937**, 1252-71.

31. Nidetzky, B.; Eis, C., Alpha-retaining glucosyl transfer catalysed by trehalose phosphorylase from Schizophyllum commune: mechanistic evidence obtained from steady-state kinetic studies with substrate analogues and inhibitors. *Biochem J* **2001,** 360, (Pt 3), 727-36.

32. Goedl, C.; Griessler, R.; Schwarz, A.; Nidetzky, B., Structure-function relationships for Schizophyllum commune trehalose phosphorylase and their implications for the catalytic mechanism of family GT-4 glycosyltransferases. *Biochem J* **2006,** 397, (3), 491-500.

33. Sinnott, M. L.; Jencks, W. P., Solvolysis of D-Glucopyranosyl Derivatives in Mixtures of Ethanol and 2,2,2-Trifluoroethanol. *J Am Chem Soc* **1980,** 102, (6), 2026-32.

34. Lee, S. S.; Hong, S. Y.; Errey, J. C.; Izumi, A.; Davies, G. J.; Davis, B. G., Mechanistic evidence for a front-side, S(N)i-type reaction in a retaining glycosyltransferase. *Nat Chem Biol* **2011,** 7, (9), 631-8.

35. Lee, J. K.; Bain, A. D.; Berti, P. J., Probing the transition states of four glucoside hydrolyses with 13C kinetic isotope effects measured at natural abundance by NMR spectroscopy. *J Am Chem Soc* **2004,** 126, (12), 3769-76.

36. Sinnott, M. L.; Souchard, I. J., The mechanism of action of beta-galactosidase. Effect of aglycone nature and -deuterium substitution on the hydrolysis of aryl galactosides. *Biochem J* **1973,** 133, (1), 89-98.

37. Chan, J.; Lewis, A. R.; Gilbert, M.; Karwaski, M. F.; Bennet, A. J., A direct NMR method for the measurement of competitive kinetic isotope effects. *Nat Chem Biol* **2010,** 6, (6), 405-7.

38. Inouye, S.; Tsuruoka, T.; Niida, T., Structure of nojirimycin a piperidinose sugar antibiotic. *J Antibiot* **1966,** 19, (6), 288-92.

39. Paulsen, H.; Sangster, I.; Heyns, K., Monosaccharide Mit Stickstoffhaltigem Ring .13. Synthese Und Reaktionen Von Keto-Piperidinosen. *CHEM BER-RECL* **1967,** 100, (3), 802-15.

40. Stutz, A. E., *Iminosugars as Glycosidase Inhibitors, Nojirimycin and Beyond*. 1999; p 397.

41. Lillelund, V. H.; Jensen, H. H.; Liang, X.; Bols, M., Recent developments of transition-state analogue glycosidase inhibitors of non-natural product origin. *Chem Rev* **2002,** 102, (2), 515-53.

42. Ermert, P.; Vasella, A., Synthesis of a Glucose-Derived Tetrazole as a New Beta-Glucosidase Inhibitor - a New Synthesis of 1-Deoxynojirimycin. *Helv Chim Acta* **1991,** 74, (8), 2043-53.

43. Horsch, M.; Hoesch, L.; Vasella, A.; Rast, D. M., N-acetylglucosaminono-1,5-lactone oxime and the corresponding (phenylcarbamoyl)oxime. Novel and potent inhibitors of beta-N-acetylglucosaminidase. *Eur J Biochem* **1991,** 197, (3), 815-8.

44. Knapp, S.; Vocadlo, D.; Gao, Z. N.; Kirk, B.; Lou, J. P.; Withers, S. G., NAG-thiazoline, an N-acetyl-beta-hexosaminidase inhibitor that implicates acetamido participation. *J Am Chem Soc* **1996,** 118, (28), 6804-5.

45. Blake, C. C. F.; Koenig, D. F.; Mair, G. A.; North, A. C. T.; Phillips, D. C.; Sarma, V. R., Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature* **1965,** 206, (4986), 757-61.

46. Davies, G.; Henrissat, B., Structures and mechanisms of glycosyl hydrolases. *Structure* **1995,** 3, (9), 853-9.

47. Sulzenbacher, G.; Bignon, C.; Nishimura, T.; Tarling, C. A.; Withers, S. G.; Henrissat, B.; Bourne, Y., Crystal structure of Thermotoga maritima alpha-L-fucosidase. Insights into the catalytic mechanism and the molecular basis for fucosidosis. *J Biol Chem* **2004,** 279, (13), 13119-28.

48. Yaoi, K.; Mitsuishi, Y., Purification, characterization, cloning, and expression of a novel xyloglucan-specific glycosidase, oligoxyloglucan reducing end-specific cellobiohydrolase. *J Biol Chem* **2002,** 277, (50), 48276-81.

49. Burns, D. M.; Touster, O., Purification and characterization of glucosidase II, an endoplasmic reticulum hydrolase involved in glycoprotein biosynthesis. *J Biol Chem* **1982,** 257, (17), 9990-10000.

50. Koshland, D. E., Stereochemistry and the mechanism of enzymatic reactions. *Biol Rev Camb Phil Soc* **1953,** 28, (4), 416-36.

51. Vocadlo, D. J.; Davies, G. J., Mechanistic insights into glycosidase chemistry. *Curr Opin Chem Biol* **2008,** 12, (5), 539-55.

52. Notenboom, V.; Birsan, C.; Nitz, M.; Rose, D. R.; Warren, R. A.; Withers, S. G., Insights into transition state stabilization of the beta-1,4-glycosidase Cex by covalent intermediate accumulation in active site mutants. *Nat Struct Biol* **1998,** 5, (9), 812-8.

53. Kempton, J. B.; Withers, S. G., Mechanism of Agrobacterium beta-glucosidase: kinetic studies. *Biochemistry* **1992,** 31, (41), 9961-9.

54. Nath, R. L.; Rydon, H. N., The influence of structure on the hydrolysis of substituted phenyl beta-D-glucosides by emulsin. *Biochem J* **1954,** 57, (1), 1-10.

55. Klimacek, M.; Kavanagh, K. L.; Wilson, D. K.; Nidetzky, B., On the role of Bronsted catalysis in Pseudomonas fluorescens mannitol 2-dehydrogenase. *Biochem J* **2003,** 375, (Pt 1), 141-9.

56. Zechel, D. L.; Reid, S. P.; Stoll, D.; Nashiru, O.; Warren, R. A.; Withers, S. G., Mechanism, mutagenesis, and chemical rescue of a beta-mannosidase from cellulomonas fimi. *Biochemistry* **2003,** 42, (23), 7195-204.

57. Liu, S. W.; Chen, C. S.; Chang, S. S.; Mong, K. K.; Lin, C. H.; Chang, C. W.; Tang, C. Y.; Li, Y. K., Identification of essential residues of human alpha-L-fucosidase and tests of its mechanism. *Biochemistry* **2009,** 48, (1), 110-20.

58. Yamamoto, K., A quantitative approach to the evaluation of 2-acetamide substituent effects on the hydrolysis by Taka-N-acetyl-beta-D-glucosaminidase.

Role of the substrate 2-acetamide group in the N-acyl specificity of the enzyme. *J Biochem* **1974,** 76, (2), 385-90.

59. Wolfenden, R.; Snider, M. J., The depth of chemical time and the power of enzymes as catalysts. *Acc Chem Res* **2001,** 34, (12), 938-45.

60. Stoddard, J. F., *Stereochemistry of Carboyhydrates*. Wiley Interscience: New York, 1971.

61. Sinnott, M. L., Ions, ion-pairs and catalysis by the lacZ beta-galactosidase of Escherichia coli. *Febs Lett* **1978,** 94, (1), 1-9.

62. Sinnott, M. L., Catalytic Mechanisms of Enzymatic Glycosyl Transfer. *Chem Rev* **1990,** 90, (7), 1171-202.

63. Sulzenbacher, G.; Driguez, H.; Henrissat, B.; Schulein, M.; Davies, G. J., Structure of the Fusarium oxysporum endoglucanase I with a nonhydrolyzable substrate analogue: Substrate distortion gives rise to the preferred axial orientation for the leaving group. *Biochemistry* **1996,** 35, (48), 15280-7.

64. Tews, I.; Perrakis, A.; Oppenheim, A.; Dauter, Z.; Wilson, K. S.; Vorgias, C. E., Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. *Nat Struct Biol* **1996,** 3, (7), 638-48.

65. Davies, G. J.; Mackenzie, L.; Varrot, A.; Dauter, M.; Brzozowski, A. M.; Schulein, M.; Withers, S. G., Snapshots along an enzymatic reaction coordinate: analysis of a retaining beta-glycoside hydrolase. *Biochemistry* **1998,** 37, (34), 11707-13.

66. Yagi, M.; Kouno, T.; Aoyagi, Y.; Murai, H., Structure of moranoline, a piperidine alkaloid from Morus species. *J Ag Chem Soc Jap* **1976,** 50, (11), 571-2.

67. Katsilambros, N.; Philippides, P.; Toskas, A.; Protopapas, J.; Frangaki, D.; Marangos, M.; Siskoudis, P.; Anastasopoulou, K.; Xetteri, H.; Hillebrand, I., A double-blind study on the efficacy and tolerance of a new alpha-glucosidase inhibitor in type-2 diabetics. *Arzneimittel-Forsch* **1986,** 36-2, (7), 1136-8.

68. Zimran, A.; Elstein, D., Gaucher disease and the clinical experience with substrate reduction therapy. *Philos Trans R Soc London, B* **2003,** 358, (1433), 961-6.

69. Platt, F. M.; Neises, G. R.; Dwek, R. A.; Butters, T. D., N-Butyldeoxynojirimycin is a novel inhibitor of glycolipid biosynthesis. *J Biol Chem* **1994,** 269, (11), 8362-5.

70. Cutfield, S. M.; Davies, G. J.; Murshudov, G.; Anderson, B. F.; Moody, P. C.; Sullivan, P. A.; Cutfield, J. F., The structure of the exo-beta-(1,3)-glucanase

from Candida albicans in native and bound forms: relationship between a pocket and groove in family 5 glycosyl hydrolases. *J Mol Biol* **1999,** 294, (3), 771-83.

71. Goss, P. E.; Baker, M. A.; Carver, J. P.; Dennis, J. W., Inhibitors of carbohydrate processing: A new class of anticancer agents. *Clin Cancer Res* **1995,** 1, (9), 935-44.

72. Goss, P. E.; Reid, C. L.; Bailey, D.; Dennis, J. W., Phase IB clinical trial of the oligosaccharide processing inhibitor swainsonine in patients with advanced malignancies. *Clin Cancer Res* **1997,** 3, (7), 1077-86.

73. Truscheit, E.; Frommer, W.; Junge, B.; Muller, L.; Schmidt, D. D.; Wingender, W., Chemistry and Biochemistry of Microbial Alpha-Glucosidase Inhibitors. *Angew Chem Int Edit* **1981,** 20, (9), 744-61.

74. Lebovitz, H. E., alpha-Glucosidase inhibitors as agents in the treatment of diabetes. *Diabetes Rev* **1998,** 6, (2), 132-45.

75. Davies, G. J., Sinnott, M. L., Wither, S. G., *Comprehensive Biological Catalysis*. 1998; Vol. volume 1: Reactions of Electrophilic Carbon, Phosphorus and Sulfur.

76. Vasella, A.; Davies, G. J.; Bohm, M., Glycosidase mechanisms. *Curr Opin Chem Biol* **2002,** 6, (5), 619-29.

77. Jespersen, T. M.; Bols, M.; Sierks, M. R.; Skrydstrup, T., Synthesis of Isofagomine, a Novel Glycosidase Inhibitor. *Tetrahedron* **1994,** 50, (47), 13449-60.

78. Jespersen, T. M.; Dong, W. L.; Sierks, M. R.; Skrydstrup, T.; Lundt, I.; Bols, M., Isofagomine, a Potent, New Glycosidase Inhibitor. *Angew Chem Int Edit* **1994,** 33, (17), 1778-9.

79. Bols, M.; Hazell, R. G.; Thomsen, I. B., 1-Azafagomine: A Hydroxyhexahydropyridazine that Potently Inhibits Enzymatic Glycoside Cleavage. *Chem Eur J* **1997,** 3, (6), 940-7.

80. Shanmugasundaram, B.; Debowski, A. W.; Dennis, R. J.; Davies, G. J.; Vocadlo, D. J.; Vasella, A., Inhibition of O-GlcNAcase by a gluco-configured nagstatin and a PUGNAc-imidazole hybrid inhibitor. *Chem Commun* **2006**, (42), 4372-4.

81. Tanaka, K. S. E.; Winters, G. C.; Batchelor, R. J.; Einstein, F. W. B.; Bennet, A. J., A new structural motif for the design of potent glucosidase inhibitors. *J Am Chem Soc* **2001,** 123, (5), 998-9.

82. Mosi, R.; Sham, H.; Uitdehaag, J. C. M.; Ruiterkamp, R.; Dijkstra, B. W.; Withers, S. G., Reassessment of acarbose as a transition state analogue inhibitor of cyclodextrin glycosyltransferase. *Biochemistry* **1998,** 37, (49), 17192-8.

83. von Itzstein, M.; Wu, W. Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Van Phan, T.; Smythe, M. L.; White, H. F.; Oliver, S. W.; et al., Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **1993,** 363, (6428), 418-23.

84. Kim, C. U.; Lew, W.; Williams, M. A.; Liu, H.; Zhang, L.; Swaminathan, S.; Bischofberger, N.; Chen, M. S.; Mendel, D. B.; Tai, C. Y.; Laver, W. G.; Stevens, R. C., Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J Am Chem Soc* **1997,** 119, (4), 681-90.

85. Geyer, H.; Geyer, R., Strategies for analysis of glycoprotein glycosylation. *Biochim Biophys Acta* **2006,** 1764, (12), 1853-69.

86. Apweiler, R.; Hermjakob, H.; Sharon, N., On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1999,** 1473, (1), 4-8.

87. Parekh, R.; Roitt, I.; Isenberg, D.; Dwek, R.; Rademacher, T., Age-related galactosylation of the N-linked oligosaccharides of human serum IgG. *The Journal of experimental medicine* **1988,** 167, (5), 1731-6.

88. Parekh, R. B.; Dwek, R. A.; Sutton, B. J.; Fernandes, D. L.; Leung, A.; Stanworth, D.; Rademacher, T. W.; Mizuochi, T.; Taniguchi, T.; Matsuta, K.; et al., Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature* **1985,** 316, (6027), 452-7.

89. Peracaula, R.; Tabares, G.; Royle, L.; Harvey, D. J.; Dwek, R. A.; Rudd, P. M.; de Llorens, R., Altered glycosylation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins. *Glycobiology* **2003,** 13, (6), 457-70.

90. Saldova, R.; Royle, L.; Radcliffe, C. M.; Abd Hamid, U. M.; Evans, R.; Arnold, J. N.; Banks, R. E.; Hutson, R.; Harvey, D. J.; Antrobus, R.; Petrescu, S. M.; Dwek, R. A.; Rudd, P. M., Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG. *Glycobiology* **2007,** 17, (12), 1344-56.

91. Mechref, Y.; Novotny, M. V., Structural investigations of glycoconjugates at high sensitivity. *Chem Rev* **2002,** 102, (2), 321-69.

92. Patwa, T.; Li, C.; Simeone, D. M.; Lubman, D. M., Glycoprotein analysis using protein microarrays and mass spectrometry. *Mass spectrometry reviews* **2010,** 29, (5), 830-44.

93. Ruhaak, L. R.; Zauner, G.; Huhn, C.; Bruggink, C.; Deelder, A. M.; Wuhrer, M., Glycan labeling strategies and their use in identification and quantification. *Analytical and bioanalytical chemistry* **2010,** 397, (8), 3457-81.

94. Hu, S.; Wong, D. T., Lectin microarray. *Proteom Clin Appl* **2009,** 3, (2), 148-54.

95. Drickamer, K.; Taylor, M. E., Biology of animal lectins. *Annual review of cell biology* **1993,** 9, 237-64.

96. Sharon, N.; Lis, H., History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology* **2004,** 14, (11), 53R-62R.

97. Ghazarian, H.; Idoni, B.; Oppenheimer, S. B., A glycobiology review: carbohydrates, lectins and implications in cancer therapeutics. *Acta histochemica* **2011,** 113, (3), 236-47.

98. Hirabayashi, J.; Hashidate, T.; Arata, Y.; Nishi, N.; Nakamura, T.; Hirashima, M.; Urashima, T.; Oka, T.; Futai, M.; Muller, W. E. G.; Yagi, F.; Kasai, K., Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. *Biochim Biophys Acta* **2002,** 1572, (2-3), 232-54.

99. Lau, K. S.; Partridge, E. A.; Grigorian, A.; Silvescu, C. I.; Reinhold, V. N.; Demetriou, M.; Dennis, J. W., Complex N-glycan number and degree of branching cooperate to regulate cell proliferation and differentiation. *Cell* **2007,** 129, (1), 123-34.

100. Fang, X.; Zhang, W. W., Affinity separation and enrichment methods in proteomic analysis. *Journal of proteomics* **2008,** 71, (3), 284-303.

101. Hirabayashi, J.; Kasai, K., Separation technologies for glycomics. *J Chromatogr B* **2002,** 771, (1-2), 67-87.

102. Krusius, T.; Finne, J.; Rauvala, H., The structural basis of the different affinities of two types of acidic N-glycosidic glycopeptides for concanavalin-A-sepharose. *Febs Lett* **1976,** 71, (1), 117-20.

103. Kobata, A.; Endo, T., Immobilized lectin columns: useful tools for the fractionation and structural analysis of oligosaccharides. *J Chromatogr* **1992,** 597, (1-2), 111-22.

104. Trojan, J.; Theodoropoulou, M.; Usadel, K. H.; Stalla, G. K.; Schaaf, L., Modulation of human thyrotropin oligosaccharide structures - enhanced proportion of sialylated and terminally galactosylated serum thyrotropin isoforms

in subclinical and overt primary hypothyroidism. *The Journal of endocrinology* **1998,** 158, (3), 359-65.

105. Hsu, K. L.; Mahal, L. K., A lectin microarray approach for the rapid analysis of bacterial glycans. *Nature protocols* **2006,** 1, (2), 543-9.

106. Kuno, A.; Uchiyama, N.; Koseki-Kuno, S.; Ebe, Y.; Takashima, S.; Yamada, M.; Hirabayashi, J., Evanescent-field fluorescence-assisted lectin microarray: a new strategy for glycan profiling. *Nature methods* **2005,** 2, (11), 851-6.

107. Tao, S. C.; Li, Y.; Zhou, J.; Qian, J.; Schnaar, R. L.; Zhang, Y.; Goldstein, I. J.; Zhu, H.; Schneck, J. P., Lectin microarrays identify cell-specific and functionally significant cell surface glycan markers. *Glycobiology* **2008,** 18, (10), 761-9.

108. Katrlik, J.; Svitel, J.; Gemeiner, P.; Kozar, T.; Tkac, J., Glycan and lectin microarrays for glycomics and medicinal applications. *Medicinal research reviews* **2010,** 30, (2), 394-418.

109. Kleene, R.; Schachner, M., Glycans and neural cell interactions. *Nature reviews* **2004,** 5, (3), 195-208.

110. Schachner, M.; Martini, R., Glycans and the modulation of neural-recognition molecule function. *Trends in neurosciences* **1995,** 18, (4), 183-91.

111. Kruse, J.; Keilhauer, G.; Faissner, A.; Timpl, R.; Schachner, M., The J1 glycoprotein - a novel nervous system cell adhesion molecule of the L2/HNK-1 family. *Nature* **1985,** 316, (6024), 146-8.

112. Metcalfe, W. K.; Myers, P. Z.; Trevarrow, B.; Bass, M. B.; Kimmel, C. B., Primary neurons that express the L2/HNK-1 carbohydrate during early development in the zebrafish. *Development (Cambridge, England)* **1990,** 110, (2), 491-504.

113. Yuzwa, S. A.; Yadav, A. K.; Skorobogatko, Y.; Clark, T.; Vosseller, K.; Vocadlo, D. J., Mapping O-GlcNAc modification sites on tau and generation of a site-specific O-GlcNAc tau antibody. *Amino acids* **2010,** 40, (3), 857-68.

114. Kusunoki, S.; Kaida, K., Antibodies against ganglioside complexes in Guillain-Barre syndrome and related disorders. *J Neurochem* **2011,** 116, (5), 828-32.

115. Hakomori, S., Tumor malignancy defined by aberrant glycosylation and sphingo(glyco)lipid metabolism. *Cancer research* **1996,** 56, (23), 5309-18.

116. Dube, D. H.; Bertozzi, C. R., Glycans in cancer and inflammation - potential for therapeutics and diagnostics. *Nat Rev Drug Discov* **2005,** 4, (6), 477-88.

117. Numahata, K.; Satoh, M.; Handa, K.; Saito, S.; Ohyama, C.; Ito, A.; Takahashi, T.; Hoshi, S.; Orikasa, S.; Hakomori, S. I., Sialosyl-Le(x) expression defines invasive and metastatic properties of bladder carcinoma. *Cancer* **2002,** 94, (3), 673-85.

118. Liedtke, S.; Geyer, H.; Wuhrer, M.; Geyer, R.; Frank, G.; Gerardy-Schahn, R.; Zahringer, U.; Schachner, M., Characterization of N-glycans from mouse brain neural cell adhesion molecule. *Glycobiology* **2001,** 11, (5), 373-84.

119. Miermont, A.; Barnhill, H.; Strable, E.; Lu, X. W.; Wall, K. A.; Wang, Q.; Finn, M. G.; Huang, X. F., Cowpea mosaic virus capsid: A promising carrier for the development of carbohydrate based antitumor Vaccines. *Chem-Eur J* **2008,** 14, (16), 4939-47.

120. Ingale, S.; Wolfert, M. A.; Gaekwad, J.; Buskas, T.; Boons, G. J., Robust immune responses elicited by a fully synthetic three-component vaccine. *Nat Chem Biol* **2007,** 3, (10), 663-7.

121. Moloney, D. J.; Lin, A. I.; Haltiwanger, R. S., The O-linked fucose glycosylation pathway. Evidence for protein-specific elongation of *O*-linked fucose in Chinese hamster ovary cells. *J Biol Chem* **1997,** 272, (30), 19046-50.

122. Moloney, D. J.; Panin, V. M.; Johnston, S. H.; Chen, J.; Shao, L.; Wilson, R.; Wang, Y.; Stanley, P.; Irvine, K. D.; Haltiwanger, R. S.; Vogt, T. F., Fringe is a glycosyltransferase that modifies Notch. *Nature* **2000,** 406, (6794), 369-75.

123. Lin, A. I.; Philipsberg, G. A.; Haltiwanger, R. S., Core fucosylation of high-mannose-type oligosaccharides in GlcNAc transferase I-deficient (Lec1) CHO cells. *Glycobiology* **1994,** 4, (6), 895-901.

124. Cole, R. N.; Hart, G. W., Cytosolic O-glycosylation is abundant in nerve terminals. *J Neurochem* **2001,** 79, (5), 1080-9.

125. Stanley, P.; Narasimhan, S.; Siminovitch, L.; Schachter, H., Chinese hamster ovary cells selected for resistance to the cytotoxicity of phytohemagglutinin are deficient in a UDP-N-acetylglucosamine - glycoprotein N-acetylglucosaminyltransferase activity. *P Natl Acad Sci U S A* **1975,** 72, (9), 3323-7.

126. Shao, L.; Moloney, D. J.; Haltiwanger, R., Fringe modifies O-fucose on mouse Notch1 at epidermal growth factor-like repeats within the ligand-binding site and the Abruptex region. *J Biol Chem* **2003,** 278, (10), 7775-82.

127. Saxon, E.; Bertozzi, C. R., Cell surface engineering by a modified Staudinger reaction. *Science* **2000,** 287, (5460), 2007-10.

128. Torres, C. R.; Hart, G. W., Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J Biol Chem* **1984,** 259, (5), 3308-17.

129. Vocadlo, D. J.; Hang, H. C.; Kim, E. J.; Hanover, J. A.; Bertozzi, C. R., A chemical approach for identifying O-GlcNAc-modified proteins in cells. *P Natl Acad Sci U S A* **2003,** 100, (16), 9116-21.

130. Laughlin, S. T.; Bertozzi, C. R., In vivo imaging of Caenorhabditis elegans glycans. *ACS chemical biology* **2009,** 4, (12), 1068-72.

131. Graham, M. E.; Thaysen-Andersen, M.; Bache, N.; Craft, G. E.; Larsen, M. R.; Packer, N. H.; Robinson, P. J., A novel post-translational modification in nerve terminals: O-linked N-acetylglucosamine phosphorylation. *Journal of proteome research* **2011,** 10, (6), 2725-33.

132. Haltiwanger, R. S.; Holt, G. D.; Hart, G. W., Enzymatic addition of O-GlcNAc to nuclear and cytoplasmic proteins. Identification of a uridine diphospho-N-acetylglucosamine:peptide beta-N-acetylglucosaminyltransferase. *J Biol Chem* **1990,** 265, (5), 2563-8.

133. Chou, C. F.; Smith, A. J.; Omary, M. B., Characterization and dynamics of O-linked glycosylation of human cytokeratin 8 and 18. *J Biol Chem* **1992,** 267, (6), 3901-6.

134. Haltiwanger, R. S.; Grove, K.; Philipsberg, G. A., Modulation of O-linked N-acetylglucosamine levels on nuclear and cytoplasmic proteins in vivo using the peptide O-GlcNAc-beta-N-acetylglucosaminidase inhibitor O-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-N-phenylcarbamate. *J Biol Chem* **1998,** 273, (6), 3611-7.

135. Marshall, S.; Bacote, V.; Traxinger, R. R., Discovery of a metabolic pathway mediating glucose-induced desensitization of the glucose transport system. Role of hexosamine biosynthesis in the induction of insulin resistance. *J Biol Chem* **1991,** 266, (8), 4706-12.

136. Walgren, J. L.; Vincent, T. S.; Schey, K. L.; Buse, M. G., High glucose and insulin promote O-GlcNAc modification of proteins, including alpha-tubulin. *American journal of physiology* **2003,** 284, (2), E424-34.

137. Porte, D., Jr.; Schwartz, M. W., Diabetes complications: why is glucose potentially toxic? *Science* **1996,** 272, (5262), 699-700.

138. Arias, E. B.; Kim, J.; Cartee, G. D., Prolonged incubation in PUGNAc results in increased protein O-Linked glycosylation and insulin resistance in rat skeletal muscle. *Diabetes* **2004,** 53, (4), 921-30.

139. Vosseller, K.; Wells, L.; Lane, M. D.; Hart, G. W., Elevated nucleocytoplasmic glycosylation by O-GlcNAc results in insulin resistance associated with defects in Akt activation in 3T3-L1 adipocytes. *P Natl Acad Sci U S A* **2002,** 99, (8), 5313-8.

140. Yang, X.; Ongusaha, P. P.; Miles, P. D.; Havstad, J. C.; Zhang, F.; So, W. V.; Kudlow, J. E.; Michell, R. H.; Olefsky, J. M.; Field, S. J.; Evans, R. M., Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance. *Nature* **2008,** 451, (7181), 964-9.

141. Park, S. Y.; Ryu, J.; Lee, W., O-GlcNAc modification on IRS-1 and Akt2 by PUGNAc inhibits their phosphorylation and induces insulin resistance in rat primary adipocytes. *Experimental & molecular medicine* **2005,** 37, (3), 220-9.

142. Yamanaka, S.; Johnson, M. D.; Grinberg, A.; Westphal, H.; Crawley, J. N.; Taniike, M.; Suzuki, K.; Proia, R. L., Targeted disruption of the Hexa gene results in mice with biochemical and pathologic features of Tay-Sachs disease. *P Natl Acad Sci U S A* **1994,** 91, (21), 9975-9.

143. Triggs-Raine, B.; Mahuran, D. J.; Gravel, R. A., Naturally occurring mutations in GM2 gangliosidosis: a compendium. *Advances in genetics* **2001,** 44, 199-224.

144. Zachara, N. E.; O'Donnell, N.; Cheung, W. D.; Mercer, J. J.; Marth, J. D.; Hart, G. W., Dynamic O-GlcNAc modification of nucleocytoplasmic proteins in response to stress. A survival response of mammalian cells. *J Biol Chem* **2004,** 279, (29), 30133-42.

145. Hase, S.; Kawabata, S.; Nishimura, H.; Takeya, H.; Sueyoshi, T.; Miyata, T.; Iwanaga, S.; Takao, T.; Shimonishi, Y.; Ikenaka, T., A new trisaccharide sugar chain linked to a serine residue in bovine blood coagulation factors VII and IX. *J Biochem* **1988,** 104, (6), 867-8.

146. Hase, S.; Nishimura, H.; Kawabata, S.; Iwanaga, S.; Ikenaka, T., The structure of (xylose)2glucose-O-serine 53 found in the first epidermal growth factor-like domain of bovine blood clotting factor IX. *J Biol Chem* **1990,** 265, (4), 1858-61.

147. Nishimura, H.; Kawabata, S.; Kisiel, W.; Hase, S.; Ikenaka, T.; Takao, T.; Shimonishi, Y.; Iwanaga, S., Identification of a disaccharide (Xyl-Glc) and a trisaccharide (Xyl2-Glc) O-glycosidically linked to a serine residue in the first epidermal growth factor-like domain of human factors VII and IX and protein Z and bovine protein Z. *J Biol Chem* **1989,** 264, (34), 20320-5.

148. Moloney, D. J.; Shair, L. H.; Lu, F. M.; Xia, J.; Locke, R.; Matta, K. L.; Haltiwanger, R. S., Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on epidermal growth factor-like modules. *J Biol Chem* **2000,** 275, (13), 9604-11.

149. Artavanis-Tsakonas, S.; Rand, M. D.; Lake, R. J., Notch signaling: cell fate control and signal integration in development. *Science* **1999,** 284, (5415), 770-6.

150. Oda, T.; Elkahloun, A. G.; Pike, B. L.; Okajima, K.; Krantz, I. D.; Genin, A.; Piccoli, D. A.; Meltzer, P. S.; Spinner, N. B.; Collins, F. S.; Chandrasekharappa, S. C., Mutations in the human Jagged1 gene are responsible for Alagille syndrome. *Nat Genet* **1997,** 16, (3), 235-42.

151. Li, L.; Krantz, I. D.; Deng, Y.; Genin, A.; Banta, A. B.; Collins, C. C.; Qi, M.; Trask, B. J.; Kuo, W. L.; Cochran, J.; Costa, T.; Pierpont, M. E.; Rand, E. B.; Piccoli, D. A.; Hood, L.; Spinner, N. B., Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. *Nat Genet* **1997,** 16, (3), 243-51.

152. Shi, T. P.; Xu, H.; Wei, J. F.; Ai, X.; Ma, X.; Wang, B. J.; Ju, Z. H.; Zhang, G. X.; Wang, C.; Wu, Z. Q.; Zhang, X., Association of low expression of notch-1 and jagged-1 in human papillary bladder cancer and shorter survival. *J Urol* **2008,** 180, (1), 361-6.

153. Mohr, O. L., Character changes caused by mutation of an entire region of a chromosome in Drosophila. *Genetics* **1919,** 4, (3), 275-82.

154. Roehl, H.; Bosenberg, M.; Blelloch, R.; Kimble, J., Roles of the RAM and ANK domains in signaling by the C. elegans GLP-1 receptor. *Embo J* **1996,** 15, (24), 7002-12.

155. Bertagna, A.; Toptygin, D.; Brand, L.; Barrick, D., The effects of conformational heterogeneity on the binding of the Notch intracellular domain to effector proteins: a case of biologically tuned disorder. *Biochem Soc Trans* **2008,** 36, (Pt 2), 157-66.

156. Bray, S. J., Notch signalling: a simple pathway becomes complex. *Nat Rev Mol Cell Biol* **2006,** 7, (9), 678-89.

157. Rebay, I.; Fleming, R. J.; Fehon, R. G.; Cherbas, L.; Cherbas, P.; Artavanis-Tsakonas, S., Specific EGF repeats of Notch mediate interactions with Delta and Serrate: implications for Notch as a multifunctional receptor. *Cell* **1991,** 67, (4), 687-99.

158. Gordon, W. R.; Roy, M.; Vardar-Ulu, D.; Garfinkel, M.; Mansour, M. R.; Aster, J. C.; Blacklow, S. C., Structure of the Notch1-negative regulatory region: implications for normal activation and pathogenic signaling in T-ALL. *Blood* **2009,** 113, (18), 4381-90.

159. Gordon, W. R.; Vardar-Ulu, D.; Histen, G.; Sanchez-Irizarry, C.; Aster, J. C.; Blacklow, S. C., Structural basis for autoinhibition of Notch. *Nat Struct Mol Biol* **2007,** 14, (5), 295-300.

160. De Strooper, B.; Annaert, W.; Cupers, P.; Saftig, P.; Craessaerts, K.; Mumm, J. S.; Schroeter, E. H.; Schrijvers, V.; Wolfe, M. S.; Ray, W. J.; Goate, A.; Kopan, R., A presenilin-1-dependent gamma-secretase-like protease mediates release of Notch intracellular domain. *Nature* **1999,** 398, (6727), 518-22.

161. Wolfsberg, T. G.; Primakoff, P.; Myles, D. G.; White, J. M., ADAM, a novel family of membrane proteins containing A Disintegrin And Metalloprotease domain: multipotential functions in cell-cell and cell-matrix interactions. *The Journal of cell biology* **1995,** 131, (2), 275-8.

162. van Tetering, G.; van Diest, P.; Verlaan, I.; van der Wall, E.; Kopan, R.; Vooijs, M., Metalloprotease ADAM10 is required for Notch1 site 2 cleavage. *J Biol Chem* **2009,** 284, (45), 31018-27.

163. Mumm, J. S.; Schroeter, E. H.; Saxena, M. T.; Griesemer, A.; Tian, X.; Pan, D. J.; Ray, W. J.; Kopan, R., A ligand-induced extracellular cleavage regulates gamma-secretase-like proteolytic activation of Notch1. *Molecular cell* **2000,** 5, (2), 197-206.

164. Hayward, S. D., Viral interactions with the Notch pathway. *Seminars in cancer biology* **2004,** 14, (5), 387-96.

165. Borggrefe, T.; Oswald, F., The Notch signaling pathway: Transcriptional regulation at Notch target genes. *Cell Mol Life Sci* **2009,** 66, (10), 1631-46.

166. Takeuchi, H.; Haltiwanger, R. S., Role of glycosylation of Notch in development. *Sem Cell Dev Biol* **2010,** 21, (6), 638-45.

167. Wang, Y.; Shao, L.; Shi, S.; Harris, R. J.; Spellman, M. W.; Stanley, P.; Haltiwanger, R. S., Modification of epidermal growth factor-like repeats with O-fucose. Molecular cloning and expression of a novel GDP-fucose protein O-fucosyltransferase. *J Biol Chem* **2001,** 276, (43), 40338-45.

168. Shao, L.; Haltiwanger, R. S., O-fucose modifications of epidermal growth factor-like repeats and thrombospondin type 1 repeats: unusual modifications in unusual places. *Cell Mol Life Sci* **2003,** 60, (2), 241-50.

169. Luther, K. B.; Haltiwanger, R. S., Role of unusual O-glycans in intercellular signaling. *Int J Biochem Cell Biol* **2008**.

170. Xu, A.; Haines, N.; Dlugosz, M.; Rana, N. A.; Takeuchi, H.; Haltiwanger, R. S.; Irvine, K. D., In vitro reconstitution of the modulation of Drosophila Notch-ligand binding by Fringe. *J Biol Chem* **2007,** 282, (48), 35153-62.

171. Panin, V. M.; Papayannopoulos, V.; Wilson, R.; Irvine, K. D., Fringe modulates Notch-ligand interactions. *Nature* **1997,** 387, (6636), 908-12.

172. Bjoern, S.; Foster, D. C.; Thim, L.; Wiberg, F. C.; Christensen, M.; Komiyama, Y.; Pedersen, A. H.; Kisiel, W., Human plasma and recombinant factor VII. Characterization of O-glycosylations at serine residues 52 and 60 and effects of site-directed mutagenesis of serine 52 to alanine. *J Biol Chem* **1991,** 266, (17), 11051-7.

173. Shi, S.; Ge, C.; Luo, Y.; Hou, X.; Haltiwanger, R. S.; Stanley, P., The threonine that carries fucose, but not fucose, is required for Cripto to facilitate Nodal signaling. *J Biol Chem* **2007,** 282, (28), 20133-41.

174. Stanley, P., Regulation of Notch signaling by glycosylation. *Curr Opin Struct Biol* **2007,** 17, (5), 530-5.

175. Acar, M.; Jafar-Nejad, H.; Takeuchi, H.; Rajan, A.; Ibrani, D.; Rana, N. A.; Pan, H.; Haltiwanger, R. S.; Bellen, H. J., Rumi is a CAP10 domain glycosyltransferase that modifies Notch and is required for Notch signaling. *Cell* **2008,** 132, (2), 247-58.

176. Jafar-Nejad, H.; Leonardi, J.; Fernandez-Valdivia, R., Role of glycans and glycosyltransferases in the regulation of Notch signaling. *Glycobiology* **2010,** 20, (8), 931-49.

177. Szymanski, C. M.; Burr, D. H.; Guerry, P., Campylobacter protein glycosylation affects host cell interactions. *Infection and immunity* **2002,** 70, (4), 2242-4.

178. Hendrixson, D. R.; DiRita, V. J., Identification of Campylobacter jejuni genes involved in commensal colonization of the chick gastrointestinal tract. *Molecular microbiology* **2004,** 52, (2), 471-84.

179. Karlyshev, A. V.; Everest, P.; Linton, D.; Cawthraw, S.; Newell, D. G.; Wren, B. W., The Campylobacter jejuni general glycosylation system is important for attachment to human epithelial cells and in the colonization of chicks. *Microbiology (Reading, England)* **2004,** 150, (Pt 6), 1957-64.

180. Henrissat, B.; Bairoch, A., Updating the sequence-based classification of glycosyl hydrolases. *Biochem J* **1996,** 316, 695-6.

181. Henrissat, B.; Bairoch, A., New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* **1993,** 293 (Pt 3), 781-8.

182. Wells, L.; Vosseller, K.; Hart, G. W., Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science* **2001,** 291, (5512), 2376-8.

183. Hanover, J. A., Glycan-dependent signaling: O-linked N-acetylglucosamine. *Faseb J* **2001,** 15, (11), 1865-76.

184. Lamarre-Vincent, N.; Hsieh-Wilson, L. C., Dynamic glycosylation of the transcription factor CREB: a potential role in gene regulation. *J Am Chem Soc* **2003,** 125, (22), 6612-3.

185. Kelly, W. G.; Dahmus, M. E.; Hart, G. W., RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *J Biol Chem* **1993,** 268, (14), 10416-24.

186. Roos, M. D.; Su, K.; Baker, J. R.; Kudlow, J. E., O glycosylation of an Sp1-derived peptide blocks known Sp1 protein interactions. *Mol Cell Biol* **1997,** 17, (11), 6472-80.

187. Jackson, S. P.; Tjian, R., O-glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. *Cell* **1988,** 55, (1), 125-33.

188. Zhang, F.; Su, K.; Yang, X.; Bowe, D. B.; Paterson, A. J.; Kudlow, J. E., O-GlcNAc modification is an endogenous inhibitor of the proteasome. *Cell* **2003,** 115, (6), 715-25.

189. Lubas, W. A.; Smith, M.; Starr, C. M.; Hanover, J. A., Analysis of nuclear pore protein p62 glycosylation. *Biochemistry* **1995,** 34, (5), 1686-94.

190. Griffith, L. S.; Schmitz, B., O-linked N-acetylglucosamine is upregulated in Alzheimer brains. *Biochem Biophys Res Commun* **1995,** 213, (2), 424-31.

191. Cole, R. N.; Hart, G. W., Glycosylation sites flank phosphorylation sites on synapsin I: *O*-linked *N*-acetylglucosamine residues are localized within domains mediating synapsin I interactions. *J Neurochem* **1999,** 73, (1), 418-28.

192. McClain, D. A.; Lubas, W. A.; Cooksey, R. C.; Hazel, M.; Parker, G. J.; Love, D. C.; Hanover, J. A., Altered glycan-dependent signaling induces insulin resistance and hyperleptinemia. *P Natl Acad Sci U S A* **2002,** 99, (16), 10695-9.

193. Liu, F.; Iqbal, K.; Grundke-Iqbal, I.; Hart, G. W.; Gong, C. X., O-GlcNAcylation regulates phosphorylation of tau: a mechanism involved in Alzheimer's disease. *Proc Natl Acad Sci U S A* **2004,** 101, (29), 10804-9.

194. Yao, P. J.; Coleman, P. D., Reduction of O-linked N-acetylglucosamine-modified assembly protein-3 in Alzheimer's disease. *J Neurosci* **1998,** 18, (7), 2399-411.

195. Chou, T. Y.; Hart, G. W., O-linked N-acetylglucosamine and cancer: messages from the glycosylation of c-Myc. *Adv Exp Med Biol* **2001,** 491, 413-8.

196. Dong, D. L. Y.; Hart, G. W., Purification and characterization of an O-GlcNAc selective N-acetyl-beta-D-glucosaminidase from Rat spleen cytosol. *J Biol Chem* **1994,** 269, (30), 19321-30.

197. Toleman, C.; Paterson, A. J.; Whisenhunt, T. R.; Kudlow, J. E., Characterization of the histone acetyltransferase (HAT) domain of a bifunctional protein with activable O-GlcNAcase and HAT activities. *J Biol Chem* **2004,** 279, (51), 53665-73.

198. Gao, Y.; Wells, L.; Comer, F. I.; Parker, G. J.; Hart, G. W., Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-N-acetylglucosaminidase from human brain. *J Biol Chem* **2001,** 276, (13), 9838-45.

199. Wells, L.; Gao, Y.; Mahoney, J. A.; Vosseller, K.; Chen, C.; Rosen, A.; Hart, G. W., Dynamic O-glycosylation of nuclear and cytosolic proteins: further characterization of the nucleocytoplasmic beta-N-acetylglucosaminidase, O-GlcNAcase. *J Biol Chem* **2002,** 277, (3), 1755-61.

200. Zhou, G. C.; Parikh, S. L.; Tyler, P. C.; Evans, G. B.; Furneaux, R. H.; Zubkova, O. V.; Benjes, P. A.; Schramm, V. L., Inhibitors of ADP-ribosylating bacterial toxins based on oxacarbenium ion character at their transition states. *J Am Chem Soc* **2004,** 126, (18), 5690-8.

201. Legler, G.; Lullau, E.; Kappes, E.; Kastenholz, F., Bovine N-acetyl-beta-D-glucosaminidase: affinity purification and characterization of its active site with nitrogen containing analogs of N-acetylglucosamine. *Biochim Biophys Acta* **1991,** 1080, (2), 89-95.

202. Liu, J.; Shikhman, A. R.; Lotz, M. K.; Wong, C. H., Hexosaminidase inhibitors as new drug candidates for the therapy of osteoarthritis. *Chem Biol* **2001,** 8, (7), 701-11.

203. Konrad, R. J.; Mikolaenko, I.; Tolar, J. F.; Liu, K.; Kudlow, J. E., The potential mechanism of the diabetogenic action of streptozotocin: inhibition of pancreatic beta-cell O-GlcNAc-selective N-acetyl-beta-D-glucosaminidase. *Biochem J* **2001,** 356, (Pt 1), 31-41.

204. Liu, K.; Paterson, A. J.; Zhang, F.; McAndrew, J.; Fukuchi, K.; Wyss, J. M.; Peng, L.; Hu, Y.; Kudlow, J. E., Accumulation of protein O-GlcNAc modification inhibits proteasomes in the brain and coincides with neuronal apoptosis in brain areas with high O-GlcNAc metabolism. *J Neurochem* **2004,** 89, (4), 1044-55.

205. Parker, G.; Taylor, R.; Jones, D.; McClain, D., Hyperglycemia and inhibition of glycogen synthase in streptozotocin-treated mice: role of O-linked N-acetylglucosamine. *J Biol Chem* **2004,** 279, (20), 20636-42.

206. Junod, A.; Lambert, A. E.; Orci, L.; Pictet, R.; Gonet, A. E.; Renold, A. E., Studies of the diabetogenic action of streptozotocin. *P Soc Exp Biol Med* **1967,** 126, (1), 201-5.

207. Bennett, R. A.; Pegg, A. E., Alkylation of DNA in rat tissues following administration of streptozotocin. *Cancer research* **1981,** 41, (7), 2786-90.

208. Kroncke, K. D.; Fehsel, K.; Sommer, A.; Rodriguez, M. L.; Kolb-Bachofen, V., Nitric oxide generation during cellular metabolization of the diabetogenic N-methyl-N-nitroso-urea streptozotozin contributes to islet cell DNA damage. *Biol Chem H-S* **1995,** 376, (3), 179-85.

209. Yamamoto, H.; Uchigata, Y.; Okamoto, H., Streptozotocin and alloxan induce DNA strand breaks and poly(ADP-ribose) synthetase in pancreatic islets. *Nature* **1981,** 294, (5838), 284-6.

210. Yamada, K.; Nonaka, K.; Hanafusa, T.; Miyazaki, A.; Toyoshima, H.; Tarui, S., Preventive and therapeutic effects of large-dose nicotinamide injections on diabetes associated with insulitis. An observation in nonobese diabetic (NOD) mice. *Diabetes* **1982,** 31, (9), 749-53.

211. Burkart, V.; Wang, Z. Q.; Radons, J.; Heller, B.; Herceg, Z.; Stingl, L.; Wagner, E. F.; Kolb, H., Mice lacking the poly(ADP-ribose) polymerase gene are resistant to pancreatic beta-cell destruction and diabetes development induced by streptozocin. *Nat Med* **1999,** 5, (3), 314-9.

212. Roos, M. D.; Xie, W.; Su, K.; Clark, J. A.; Yang, X.; Chin, E.; Paterson, A. J.; Kudlow, J. E., Streptozotocin, an analog of N-acetylglucosamine, blocks the removal of O-GlcNAc from intracellular proteins. *P Assoc Am Physicians* **1998,** 110, (5), 422-32.

213. Gao, Y.; Parker, G. J.; Hart, G. W., Streptozotocin-induced beta-cell death is independent of its inhibition of O-GlcNAcase in pancreatic Min6 cells. *Arch Biochem Biophys* **2000,** 383, (2), 296-302.

214. Okuyama, R.; Yachi, M., Cytosolic O-GlcNAc accumulation is not involved in beta-cell death in HIT-T15 or Min6. *Biochem Biophys Res Commun* **2001,** 287, (2), 366-71.

215. Hanover, J. A.; Lai, Z.; Lee, G.; Lubas, W. A.; Sato, S. M., Elevated O-linked N-acetylglucosamine metabolism in pancreatic beta-cells. *Arch Biochem Biophys* **1999,** 362, (1), 38-45.

216. Liu, K.; Paterson, A. J.; Konrad, R. J.; Parlow, A. F.; Jimi, S.; Roh, M.; Chin, E., Jr.; Kudlow, J. E., Streptozotocin, an O-GlcNAcase inhibitor, blunts insulin and growth hormone secretion. *Molecular and cellular endocrinology* **2002,** 194, (1-2), 135-46.

217. Miller, D. J.; Gong, X.; Shur, B. D., Sperm require beta-N-acetylglucosaminidase to penetrate through the egg zona pellucida. *Development (Cambridge, England)* **1993,** 118, (4), 1279-89.

218. Mark, B. L.; Mahuran, D. J.; Cherney, M. M.; Zhao, D.; Knapp, S.; James, M. N., Crystal structure of human beta-hexosaminidase B: understanding the molecular basis of Sandhoff and Tay-Sachs disease. *J Mol Biol* **2003,** 327, (5), 1093-109.

219. Garman, S. C.; Hannick, L.; Zhu, A.; Garboczi, D. N., The 1.9 A structure of alpha-N-acetylgalactosaminidase: molecular basis of glycosidase deficiency diseases. *Structure* **2002,** 10, (3), 425-34.

220. Zechel, D. L.; Withers, S. G., Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc Chem Res* **2000,** 33, (1), 11-8.

221. Vocadlo, D. J.; Davies, G. J.; Laine, R.; Withers, S. G., Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* **2001,** 412, (6849), 835-8.

222. Mark, B. L.; Vocadlo, D. J.; Knapp, S.; Triggs-Raine, B. L.; Withers, S. G.; James, M. N., Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase. *J Biol Chem* **2001,** 276, (13), 10330-37.

223. Hansch, C., Leo, A., *Substituent Constants for Correlation Analysis in Chemistry and Biology*. John Wiley and Sons Inc.: New York, 1979; p 1-339.

224. Maier, T.; Strater, N.; Schuette, C. G.; Klingenstein, R.; Sandhoff, K.; Saenger, W., The X-ray crystal structure of human beta-hexosaminidase B provides new insights into Sandhoff disease. *J Mol Biol* **2003,** 328, (3), 669-81.

225. Terwisscha van Scheltinga, A. C.; Armand, S.; Kalk, K. H.; Isogai, A.; Henrissat, B.; Dijkstra, B. W., Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and X-ray structure of a complex with allosamidin: evidence for substrate assisted catalysis. *Biochemistry* **1995,** 34, (48), 15619-23.

226. Markovic-Housley, Z.; Miglierini, G.; Soldatova, L.; Rizkallah, P. J.; Muller, U.; Schirmer, T., Crystal structure of hyaluronidase, a major allergen of bee venom. *Structure* **2000,** 8, (10), 1025-35.

227. James, M. N. G.; Mark, B. L., Anchimeric assistance in hexosaminidases. *Canadian Journal of Chemistry-Revue Canadienne De Chimie* **2002,** 80, (8), 1064-74.

228. Comer, F. I.; Vosseller, K.; Wells, L.; Accavitti, M. A.; Hart, G. W., Characterization of a mouse monoclonal antibody specific for O-linked N-acetylglucosamine. *Anal Biochem* **2001,** 293, (2), 169-77.

229. Tropak, M. B.; Reid, S. P.; Guiral, M.; Withers, S. G.; Mahuran, D., Pharmacological enhancement of beta-hexosaminidase activity in fibroblasts from adult Tay-Sachs and Sandhoff Patients. *J Biol Chem* **2004,** 279, (14), 13478-87.

230. Roeser, K. R.; Legler, G., Role of sugar hydroxyl groups in glycoside hydrolysis. Cleavage mechanism of deoxyglucosides and related substrates by beta-glucosidase A3 from Aspergillus wentii. *Biochim Biophys Acta* **1981,** 657, (2), 321-33.

231. Bergmann, M.; Zervas, L., Syntheses with glucosamine. *Berichte Der Deutschen Chemischen Gesellschaft* **1931,** 64, 975-80.

232. Tulp, A.; Barnhoorn, M.; Bause, E.; Ploegh, H., Inhibition of N-linked oligosaccharide trimming mannosidases blocks human B cell development. *Embo J* **1986,** 5, (8), 1783-90.

233. Butters, T. D.; Dwek, R. A.; Platt, F. M., New therapeutics for the treatment of glycosphingolipid lysosomal storage diseases. *Glycobiology and Medicine* **2003,** 535, 219-26.

234. Mendel, D. B.; Tai, C. Y.; Escarpe, P. A.; Li, W.; Sidwell, R. W.; Huffman, J. H.; Sweet, C.; Jakeman, K. J.; Merson, J.; Lacy, S. A.; Lew, W.; Williams, M. A.; Zhang, L.; Chen, M. S.; Bischofberger, N.; Kim, C. U., Oral administration of a prodrug of the influenza virus neuraminidase inhibitor GS 4071 protects mice and ferrets against influenza infection. *Antimicrob Agents Chemother* **1998,** 42, (3), 640-6.

235. Macauley, M. S.; Whitworth, G. E.; Debowski, A. W.; Chin, D.; Vocadlo, D. J., O-GlcNAcase uses substrate-assisted catalysis - Kinetic analysis and development of highly selective mechanism-inspired inhibitors. *J Biol Chem* **2005,** 280, (27), 25313-22.

236. Siriwardena, A.; Strachan, H.; El-Daher, S.; Way, G.; Winchester, B.; Glushka, J.; Moremen, K.; Boons, G. J., Potent and selective inhibition of class II alpha-D-mannosidase activity by a bicyclic sulfonium salt. *Chembiochem* **2005,** 6, (5), 845-8.

237. Pauling, L., Nature of Forces between Large Molecules of Biological Interest. *Nature* **1948,** 161, (4097), 707-9.

238. Jencks, W. P. K., N. O. Kennedy, E. P., In *Current Aspects of Biochemical Energetics*. *Academic* **1966**, 273-98.

239. Bartlett, P. A.; Marlowe, C. K., Phosphonamidates as Transition-State Analog Inhibitors of Thermolysin. *Biochemistry* **1983,** 22, (20), 4618-24.

240. Lewandowicz, A.; Tyler, P. C.; Evans, G. B.; Furneaux, R. H.; Schramm, V. L., Achieving the ultimate physiological goal in transition state analogue inhibitors for purine nucleoside phosphorylase. *J Biol Chem* **2003,** 278, (34), 31465-8.

241. Schramm, V. L., Enzymatic transition state poise and transition state analogues. *Accounts Chem Res* **2003,** 36, (8), 588-96.

242. Berti, P. J.; Blanke, S. R.; Schramm, V. L., Transition state structure for the hydrolysis of NAD(+) catalyzed by diphtheria toxin. *Journal of the American Chemical Society* **1997,** 119, (50), 12079-88.

243. Faderl, S.; Gandhi, V.; O'Brien, S.; Bonate, P.; Cortes, J.; Estey, E.; Beran, M.; Wierda, W.; Garcia-Manero, G.; Ferrajoli, A.; Estrov, Z.; Giles, F. J.; Du, M.; Kwari, M.; Keating, M.; Plunkett, W.; Kantarjian, H., Results of a phase 1-2 study of clofarabine in combination with cytarabine (ara-C) in relapsed and refractory acute leukemias. *Blood* **2005,** 105, (3), 940-7.

244. Berland, C. R.; Sigurskjold, B. W.; Stoffer, B.; Frandsen, T. P.; Svensson, B., Thermodynamics of inhibitor binding to mutant forms of glucoamylase from Aspergillus niger determined by isothermal titration calorimetry. *Biochemistry* **1995,** 34, (32), 10153-61.

245. Stubbs, K. A.; Zhang, N.; Vocadlo, D. J., A divergent synthesis of 2-acyl derivatives of PUGNAc yields selective inhibitors of O-GlcNAcase. *Org Biomol Chem* **2006,** 4, (5), 839-45.

246. Kim, E. J.; Perreira, M.; Thomas, C. J.; Hanover, J. A., An O-GlcNAcase-specific inhibitor and substrate engineered by the extension of the N-acetyl moiety. *J Am Chem Soc* **2006,** 128, (13), 4234-5.

247. Cetinbas, N.; Macauley, M. S.; Stubbs, K. A.; Drapala, R.; Vocadlo, D. J., Identification of Asp174 and Asp175 as the key catalytic residues of human O-GlcNAcase by functional analysis of site-directed mutants. *Biochemistry* **2006,** 45, (11), 3835-44.

248. Macauley, M. S.; Stubbs, K. A.; Vocadlo, D. J., O-GlcNAcase catalyzes cleavage of thioglycosides without general acid catalysis. *J Am Chem Soc* **2005,** 127, (49), 17202-3.

249. van Aalten, D. M.; Komander, D.; Synstad, B.; Gaseidnes, S.; Peter, M. G.; Eijsink, V. G., Structural insights into the catalytic mechanism of a family 18 exo-chitinase. *Proc Natl Acad Sci U S A* **2001,** 98, (16), 8979-84.

250. Toleman, C.; Paterson, A. J.; Kudlow, J. E., Location and characterization of the O-GlcNAcase active site. *Biochim Biophys Acta* **2006,** 1760, (5), 829-39.

251. Vocadlo, D. J.; Withers, S. G., Detailed comparative analysis of the catalytic mechanisms of beta-N-acetylglucosaminidases from families 3 and 20 of glycoside hydrolases. *Biochemistry* **2005,** 44, (38), 12809-18.

252. Beer, D.; Maloisel, J. L.; Rast, D. M.; Vasella, A., Synthesis of 2-Acetamido-2-Deoxy-D-Gluconhydroximolactone-Derived and Chitobionhydroximolactone-Derived N-Phenylcarbamates, Potential Inhibitors of Beta-N-Acetylglucosaminidase. *Helv Chim Acta* **1990,** 73, (7), 1918-22.

253. Rao, F. V.; Dorfmueller, H. C.; Villa, F.; Allwood, M.; Eggleston, I. M.; van Aalten, D. M., Structural insights into the mechanism and inhibition of eukaryotic O-GlcNAc hydrolysis. *Embo J* **2006,** 25, (7), 1569-78.

254. Rahil, J.; Pratt, R. F., Characterization of covalently bound enzyme inhibitors as transition-state analogs by protein stability measurements: phosphonate monoester inhibitors of a beta-lactamase. *Biochemistry* **1994,** 33, (1), 116-25.

255. Dale, M. P.; Ensley, H. E.; Kern, K.; Sastry, K. A.; Byers, L. D., Reversible inhibitors of beta-glucosidase. *Biochemistry* **1985,** 24, (14), 3530-39.

256. Snider, M. J.; Wolfenden, R., Site-bound water and the shortcomings of a less than perfect transition state analogue. *Biochemistry* **2001,** 40, (38), 11364-71.

257. Mader, M. M.; Bartlett, P. A., Binding energy and catalysis: The implications for transition-state analogs and catalytic antibodies. *Chem Rev* **1997,** 97, (5), 1281-1301.

258. Bartlett, P. A.; Giangiordano, M. A., Transition state analogy of phosphonic acid peptide inhibitors of pepsin. *J Org Chem* **1996,** 61, (10), 3433-8.

259. Wolfenden, R., Transition state analogues for enzyme catalysis. *Nature* **1969,** 223, (207), 704-5.

260. Lienhard, G. E., Enzymatic catalysis and transition-state theory. *Science* **1973,** 180, (82), 149-54.

261. Phillips, M. A.; Kaplan, A. P.; Rutter, W. J.; Bartlett, P. A., Transition-State Characterization - a New Approach Combining Inhibitor Analogs and Variation in Enzyme Structure. *Biochemistry* **1992,** 31, (4), 959-63.

262. Mark, B. L.; Vocadlo, D. J.; Zhao, D.; Knapp, S.; Withers, S. G.; James, M. N., Biochemical and structural assessment of the 1-N-azasugar GalNAc-isofagomine as a potent family 20 beta-N-acetylhexosaminidase inhibitor. *J Biol Chem* **2001,** 276, (45), 42131-7.

263. Knier, B. L.; Jencks, W. P., Mechanism of Reactions of N-(Methoxymethyl)-N,N-Dimethylanilinium Ions with Nucleophilic-Reagents. *Journal of the American Chemical Society* **1980,** 102, (22), 6789-98.

264. Lewis, B. E.; Schramm, V. L., Binding equilibrium isotope effects for glucose at the catalytic domain of human brain hexokinase. *Journal of the American Chemical Society* **2003,** 125, (16), 4785-98.

265. Huang, X. C.; Tanaka, K. S. E.; Bennet, A. J., 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* **1997,** 119, (46), 11147-54.

266. Fedorov, A.; Shi, W.; Kicska, G.; Fedorov, E.; Tyler, P. C.; Furneaux, R. H.; Hanson, J. C.; Gainsford, G. J.; Larese, J. Z.; Schramm, V. L.; Almo, S. C., Transition state structure of purine nucleoside phosphorylase and principles of atomic motion in enzymatic catalysis. *Biochemistry* **2001,** 40, (4), 853-60.

267. Richard, J. P.; Huber, R. E.; Heo, C.; Amyes, T. L.; Lin, S., Structure-reactivity relationships for beta-glactosidase (Escherichia coli, lac Z) .4. Mechanism for reaction of nucleophiles with the galactosyl-enzyme intermediates of E461G and E461Q beta-galactosidases. *Biochemistry* **1996,** 35, (38), 12387-401.

268. Dennis, R. J.; Taylor, E. J.; Macauley, M. S.; Stubbs, K. A.; Turkenburg, J. P.; Hart, S. J.; Black, G. N.; Vocadlo, D. J.; Davies, G. J., Structure and mechanism of a bacterial beta-glucosaminidase having O-GlcNAcase activity. *Nat Struct Mol Biol* **2006,** 13, (4), 365-71.

269. Scheuring, J.; Berti, P. J.; Schramm, V. L., Transition-state structure for the ADP-ribosylation of recombinant Gialpha1 subunits by pertussis toxin. *Biochemistry* **1998,** 37, (9), 2748-58.

270. Ficko-Blean, E.; Boraston, A. B., Cloning, recombinant production, crystallization and preliminary X-ray diffraction studies of a family 84 glycoside hydrolase from Clostridium perfringens. *Acta Crystallograph Sect F Struct Biol Cryst Commun* **2005,** 61, (Pt 9), 834-6.

271. Kraulis, P. J., Molscript - a Program to Produce Both Detailed and Schematic Plots of Protein Structures. *J Appl Crystallogr* **1991,** 24, 946-50.

272. Esnouf, R. M., An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J Mol Graph Model* **1997,** 15, (2), 132-4.

273. Bagdassarian, C. K.; Schramm, V. L.; Schwartz, S. D., Molecular electrostatic potential analysis for enzymatic substrates, competitive inhibitors, and transition-state inhibitors. *Journal of the American Chemical Society* **1996,** 118, (37), 8825-36.

274. Porter, G. R.; Rydon, H. N.; Schofield, J. A., Nature of the reactive serine residue in enzymes inhibited by organo-phosphorus compounds. *Nature* **1958,** 182, (4640), 927.

275. *al,* F. M. J. *e. Gaussian03, Revision C.02*, Gaussian Inc.: 2004.

276. Portmann, S.; Luthi, H. P., MOLEKEL: An interactive molecular graphics tool. *Chimia* **2000,** 54, (12), 766-70.

277. Billing, J. F.; Nilsson, U. J., Cyclic peptides containing a delta-sugar amino acid-synthesis and evaluation as artificial receptors. *Tetrahedron* **2005,** 61, (4), 863-74.

278. Struhl, G.; Adachi, A., Nuclear access and action of notch in vivo. *Cell* **1998,** 93, (4), 649-60.

279. Harris, R. J.; van Halbeek, H.; Glushka, J.; Basa, L. J.; Ling, V. T.; Smith, K. J.; Spellman, M. W., Identification and structural analysis of the tetrasaccharide NeuAc alpha(2-->6)Gal beta(1-->4)GlcNAc beta(1-->3)Fuc alpha 1-->O-linked to serine 61 of human factor IX. *Biochemistry* **1993,** 32, (26), 6539-47.

280. Bakker, H.; Oka, T.; Ashikov, A.; Yadav, A.; Berger, M.; Rana, N. A.; Bai, X.; Jigami, Y.; Haltiwanger, R. S.; Esko, J. D.; Gerardy-Schahn, R., Functional UDP-xylose transport across the endoplasmic reticulum/Golgi membrane in a Chinese hamster ovary cell mutant defective in UDP-xylose Synthase. *J Biol Chem* **2009,** 284, (4), 2576-83.

281. Stanley, P., Glycosylation engineering. *Glycobiology* **1992,** 2, (2), 99-107.

282. Dell, A.; Morris, H. R., Glycoprotein structure determination by mass spectrometry. *Science* **2001,** 291, (5512), 2351-6.

283. Kalmar, G. B.; Kay, R. J.; LaChance, A. C.; Cornell, R. B., Primary structure and expression of a human CTP:phosphocholine cytidylyltransferase. *Biochim Biophys Acta* **1994,** 1219, (2), 328-34.

284. Minamida, S.; Aoki, K.; Natsuka, S.; Omichi, K.; Fukase, K.; Kusumoto, S.; Hase, S., Detection of UDP-D-Xylose: alpha-D-xyloside alpha 1->3Xylosyltransferase activity in human hepatoma cell line HepG2. *Journal of Biochemistry* **1996,** 120, (5), 1002-6.

285. Huang, Y. P.; Mechref, Y.; Novotny, M. V., Microscale nonreductive release of O-linked glycans for subsequent analysis through MALDI mass spectrometry and capillary electrophoresis. *Analytical Chemistry* **2001,** 73, (24), 6063-9.

286. Evangelista, R. A.; Guttman, A.; Chen, F. T. A., Acid-catalyzed reductive amination of aldoses with 8-aminopyrene-1,3,6-trisulfonate. *Electrophoresis* **1996,** 17, (2), 347-51.

287. Lovering, A. L.; Lee, S. S.; Kim, Y. W.; Withers, S. G.; Strynadka, N. C., Mechanistic and structural analysis of a family 31 alpha-glycosidase and its glycosyl-enzyme intermediate. *J Biol Chem* **2005,** 280, (3), 2105-15.

288. Moracci, M.; Ponzano, B. C.; Trincone, A.; Fusco, S.; De Rosa, M.; van der Oost, J.; Sensen, C. W.; Charlebois, R. L.; Rossi, M., Identification and molecular characterization of the first alpha-xylosidase from an Archaeon. *J Biol Chem* **2000,** 275, (29), 22082-9.

289. Ghalanbor, Z.; Ghaemi, N.; Marashi, S. A.; Amanlou, M.; Habibi-Rezaei, M.; Khajeh, K.; Ranjbar, B., Binding of Tris to Bacillus licheniformis alpha-amylase can affect its starch hydrolysis activity. *Protein Pept Lett* **2008,** 15, (2), 212-4.

290. Haslam, S. M.; North, S. J.; Dell, A., Mass spectrometric analysis of N- and O-glycosylation of tissues and cells. *Curr Opin Struct Biol* **2006,** 16, (5), 584-91.

291. Domon, B.; Costello, C. E., Structure elucidation of glycosphingolipids and gangliosides using high-performance tandem mass spectrometry. *Biochemistry* **1988,** 27, (5), 1534-43.

292. Nishimura, H.; Takao, T.; Hase, S.; Shimonishi, Y.; Iwanaga, S., Human factor IX has a tetrasaccharide O-glycosidically linked to serine 61 through the fucose residue. *J Biol Chem* **1992,** 267, (25), 17520-5.

293. Shao, L.; Luo, Y.; Moloney, D. J.; Haltiwanger, R. S., O-Glycosylation of EGF repeats: identification and initial characterization of a UDP-glucose: protein O-glucosyltransferase. *Glycobiology* **2002,** 12, (11), 763-70.

294. Yuzwa, S. A.; Macauley, M. S.; Heinonen, J. E.; Shan, X. Y.; Dennis, R. J.; He, Y. A.; Whitworth, G. E.; Stubbs, K. A.; McEachern, E. J.; Davies, G. J.; Vocadlo, D. J., A potent mechanism-inspired O-GlcNAcase inhibitor that blocks phosphorylation of tau in vivo. *Nature Chemical Biology* **2008,** 4, (8), 483-90.

295. Charlwood, J.; Birrell, H.; Camilleri, P., Efficient carbohydrate release, purification, and derivatization. *Analytical Biochemistry* **1998,** 262, (2), 197-200.

296. Gao, N.; Lehrman, M. A., Non-radioactive analysis of lipid-linked oligosaccharide compositions by fluorophore-assisted carbohydrate electrophoresis. *Methods Enzymol* **2006,** 415, 3-20.

297. Guttman, A.; Chen, F. T. A.; Evangelista, R. A.; Cooke, N., High-resolution capillary gel electrophoresis of reducing oligosaccharides labeled with 1-aminopyrene-3,6,8-trisulfonate. *Analytical Biochemistry* **1996,** 233, (2), 234-42.

298. Alvarez-Manilla, G.; Warren, N. L.; Abney, T.; Atwood, J., 3rd; Azadi, P.; York, W. S.; Pierce, M.; Orlando, R., Tools for glycomics: relative quantitation of glycans by isotopic permethylation using 13CH3I. *Glycobiology* **2007,** 17, (7), 677-87.

299. Jang-Lee, J.; North, S. J.; Sutton-Smith, M.; Goldberg, D.; Panico, M.; Morris, H.; Haslam, S.; Dell, A., Glycomic profiling of cells and tissues by mass spectrometry: fingerprinting and sequencing methodologies. *Methods Enzymol* **2006,** 415, 59-86.

300. Kopan, R.; Ilagan, M. X., The canonical Notch signaling pathway: unfolding the activation mechanism. *Cell* **2009,** 137, (2), 216-33.

301. Bettenhausen, B.; Hrabe de Angelis, M.; Simon, D.; Guenet, J. L.; Gossler, A., Transient and restricted expression during mouse embryogenesis of Dll1, a murine gene closely related to Drosophila Delta. *Development (Cambridge, England)* **1995,** 121, (8), 2407-18.

302. Henrique, D.; Adam, J.; Myat, A.; Chitnis, A.; Lewis, J.; Ish-Horowicz, D., Expression of a Delta homologue in prospective neurons in the chick. *Nature* **1995,** 375, (6534), 787-90.

303. Chitnis, A.; Henrique, D.; Lewis, J.; Ish-Horowicz, D.; Kintner, C., Primary neurogenesis in Xenopus embryos regulated by a homologue of the Drosophila neurogenic gene Delta. *Nature* **1995,** 375, (6534), 761-6.

304. Myat, A.; Henrique, D.; Ish-Horowicz, D.; Lewis, J., A chick homologue of Serrate and its relationship with Notch and Delta homologues during central neurogenesis. *Developmental biology* **1996,** 174, (2), 233-47.

305. Shawber, C.; Boulter, J.; Lindsell, C. E.; Weinmaster, G., Jagged2: a serrate-like gene expressed during rat embryogenesis. *Developmental biology* **1996,** 180, (1), 370-6.

306. Whitworth, G. E.; Zandberg, W. F.; Clark, T.; Vocadlo, D. J., Mammalian Notch is modified by D-Xyl-alpha 1-3-D-Xyl-alpha 1-3-D-Glc-beta 1-O-Ser: Implementation of a method to study O-glucosylation. *Glycobiology* **2010,** 20, (3), 287-99.

307. Matsuura, A.; Ito, M.; Sakaidani, Y.; Kondo, T.; Murakami, K.; Furukawa, K.; Nadano, D.; Matsuda, T.; Okajima, T., O-linked N-acetylglucosamine is present on the extracellular domain of notch receptors. *J Biol Chem* **2008,** 283, (51), 35486-95.

308. Shi, S.; Stanley, P., Protein O-fucosyltransferase 1 is an essential component of Notch signaling pathways. *Proc Natl Acad Sci U S A* **2003,** 100, (9), 5234-9.

309. Bruckner, K.; Perez, L.; Clausen, H.; Cohen, S., Glycosyltransferase activity of Fringe modulates Notch-Delta interactions. *Nature* **2000,** 406, (6794), 411-5.

310. Rana, N. A.; Nita-Lazar, A.; Takeuchi, H.; Kakuda, S.; Luther, K. B.; Haltiwanger, R. S., O-glucose trisaccharide is present at high but variable stoichiometry at multiple sites on mouse Notch1. *J Biol Chem* **2011**.

311. Fernandez-Valdivia, R.; Takeuchi, H.; Samarghandi, A.; Lopez, M.; Leonardi, J.; Haltiwanger, R. S.; Jafar-Nejad, H., Regulation of mammalian Notch signaling and embryonic development by the protein O-glucosyltransferase Rumi. *Development* **2011,** 138, (10), 1925-34.

312. Leonardi, J.; Fernandez-Valdivia, R.; Li, Y. D.; Simcox, A. A.; Jafar-Nejad, H., Multiple O-glucosylation sites on Notch function as a buffer against temperature-dependent loss of signaling. *Development* **2011,** 138, (16), 3569-78.

313. Sethi, M. K.; Buettner, F. F. R.; Krylov, V. B.; Takeuchi, H.; Nifantiev, N. E.; Haltiwanger, R. S.; Gerardy-Schahn, R.; Bakker, H., Identification of Glycosyltransferase 8 Family Members as Xylosyltransferases Acting on O-Glucosylated Notch Epidermal Growth Factor Repeats. *J Biol Chem* **2010,** 285, (3), 1582-6.

314. Buskas, T.; Thompson, P.; Boons, G. J., Immunotherapy for cancer: synthetic carbohydrate-based vaccines. *Chemical Communications* **2009**, (36), 5335-49.

315. Kaltgrad, E.; Sen Gupta, S.; Punna, S.; Huang, C. Y.; Chang, A.; Wong, C. H.; Finn, M. G.; Blixt, O., Anti-carbohydrate antibodies elicited by polyvalent display on a viral scaffold. *Chembiochem* **2007,** 8, (12), 1455-62.

316. Astronomo, R. D.; Kaltgrad, E.; Udit, A. K.; Wang, S. K.; Doores, K. J.; Huang, C. Y.; Pantophlet, R.; Paulson, J. C.; Wong, C. H.; Finn, M. G.; Burton, D. R., Defining Criteria for Oligomannose Immunogens for HIV Using Icosahedral Virus Capsid Scaffolds. *Chemistry & Biology* **2010,** 17, (4), 357-70.

317. Janczuk, A. J.; Zhang, W.; Andreana, P. R.; Warrick, J.; Wang, P. G., The synthesis of deoxy-alpha-Gal epitope derivatives for the evaluation of an anti-alpha-Gal antibody binding. *Carbohydr Res* **2002,** 337, (14), 1247-59.

318. Stick, R. V.; Stubbs, K. A.; Watts, A. G., Modifying the regioselectivity of glycosynthase reactions through changes in the acceptor. *Australian Journal of Chemistry* **2004,** 57, (8), 779-86.

319. Grundler, G.; Schmidt, R. R., Glycosyl Imidates .13. Application of the Trichloroacetimidate Procedure to 2-Azidoglucose and 2-Azidogalactose Derivatives. *Liebigs Annalen Der Chemie* **1984**, (11), 1826-47.

320. Motawia, M. S.; Olsen, C. E.; Denyer, K.; Smith, A. M.; Moller, B. L., Synthesis of 4'-O-acetyl-maltose and alpha-D-galactopyranosyl-(1-->4)-D-glucopyranose for biochemical studies of amylose biosynthesis. *Carbohydr Res* **2001,** 330, (3), 309-18.

321. Bien, F.; Ziegler, T., Chemoenzymatic synthesis of glycosylated enantiomerically pure 4-pentene 1,2- and 1,3-diol derivatives. *Tetrahedron-Asymmetry* **1998,** 9, (5), 781-90.

322. Kisfaludy, L.; Schon, I., Preparation and Applications of Pentafluorophenyl Esters of 9-Fluorenylmethyloxycarbonyl Amino-Acids for Peptide-Synthesis. *Synthesis-Stuttgart* **1983**, (4), 325-7.

323. Katajisto, J.; Karskela, T.; Heinonen, P.; Lonnberg, H., An orthogonally protected alpha,alpha-bis(aminomethyl)-beta-alanine building block for the construction of glycoconjugates on a solid support. *J Org Chem* **2002,** 67, (23), 7995-8001.

324. Atherton, E. S., R.C., *Solid phase peptide synthesis : a practical approach*. IRL press at Oxford University Press: Oxford, England, 1989.

325. Withers, S. G.; Rye, C. S., Elucidation of the mechanism of polysaccharide cleavage by chondroitin AC lyase from Flavobacterium heparinum. *Journal of the American Chemical Society* **2002,** 124, (33), 9756-67.

326. Suzuki, K.; Ohtsuka, I.; Kanemitsu, T.; Ako, T.; Kanie, O., Single-step multisyntheses of glycosyl acceptors: Benzylation of n-1 hydroxyl groups of phenylthio glycosides of xylose, mannose, glucose, galactose, 2-azido-2-deoxy-glucose, and 2-azido-2-deoxy-galactose. *Journal of Carbohydrate Chemistry* **2005,** 24, (3), 219-36.

327. Driguez, H.; Faure, R.; Saura-Valls, M.; Brumer, H.; Planas, A.; Cottaz, S., Synthesis of a library of xylogluco-oligosaccharides for active-site mapping of xyloglucan endo-transglycosylase. *J Org Chem* **2006,** 71, (14), 5151-61.

328. Fukase, K.; Hase, S.; Ikenaka, T.; Kusumoto, S., Synthesis of New Serine-Linked Oligosaccharides in Blood-Clotting Factor-Vii and Factor-Ix and Protein-Z - the Syntheses of O-Alpha-D-Xylopyranosyl-(1-]3)-D-Glucopyranose, O-Alpha-D-Xylopyranosyl-(1-]3)-O-Alpha-D-Xylopyranosyl-(1-]3)-D-Glucopyranose, and Their Conjugates with Serine. *Bulletin of the Chemical Society of Japan* **1992,** 65, (2), 436-45.

329. Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G., Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angewandte Chemie (International ed* **2009,** 48, (52), 9879-83.

330. Yamaoka, K.; Mishima, K.; Nagashima, Y.; Asai, A.; Sanai, Y.; Kirino, T., Expression of galectin-1 mRNA correlates with the malignant potential of human gliomas and expression of antisense galectin-1 inhibits the growth of 9 glioma cells. *Journal of neuroscience research* **2000,** 59, (6), 722-30.

331. Yoshii, T.; Inohara, H.; Takenaka, Y.; Honjo, Y.; Akahani, S.; Nomura, T.; Raz, A.; Kubo, T., Galectin-3 maintains the transformed phenotype of thyroid papillary carcinoma cells. *International journal of oncology* **2001,** 18, (4), 787-92.

332. Fan, J. Q.; Ishii, S.; Asano, N.; Suzuki, Y., Accelerated transport and maturation of lysosomal alpha-galactosidase A in Fabry lymphoblasts by an enzyme inhibitor. *Nat Med* **1999,** 5, (1), 112-5.

333. Sawkar, A. R.; Cheng, W. C.; Beutler, E.; Wong, C. H.; Balch, W. E.; Kelly, J. W., Chemical chaperones increase the cellular activity of N370S beta -glucosidase: a therapeutic strategy for Gaucher disease. *Proc Natl Acad Sci U S A* **2002,** 99, (24), 15428-33.

334. Krentz, A. J.; Bailey, C. J., Oral antidiabetic agents: current role in type 2 diabetes mellitus. *Drugs* **2005,** 65, (3), 385-411.

335. Vonitzstein, M.; Wu, W. Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Phan, T. V.; Smythe, M. L.; White, H. F.; Oliver, S. W.; Colman, P. M.; Varghese, J. N.; Ryan, D. M.; Woods, J. M.; Bethell, R. C.; Hotham, V. J.; Cameron, J. M.; Penn, C. R., Rational Design of Potent Sialidase-Based Inhibitors of Influenza-Virus Replication. *Nature* **1993,** 363, (6428), 418-23.

336. Hayden, F. G.; Treanor, J. J.; Betts, R. F.; Lobo, M.; Esinhart, J. D.; Hussey, E. K., Safety and efficacy of the neuraminidase inhibitor GG167 in experimental human influenza. *Jama* **1996,** 275, (4), 295-9.

337. Whitworth, G. E.; Macauley, M. S.; Stubbs, K. A.; Dennis, R. J.; Taylor, E. J.; Davies, G. J.; Greig, I. R.; Vocadlo, D. J., Analysis of PUGNAc and NAG-thiazoline as transition state analogues for human O-GlcNAcase: mechanistic and structural insights into inhibitor selectivity and transition state poise. *J Am Chem Soc* **2007,** 129, (3), 635-44.

338. Greig, I. R.; Macauley, M. S.; Williams, I. H.; Vocadlo, D. J., Probing synergy between two catalytic strategies in the glycoside hydrolase O-GlcNAcase using multiple linear free energy relationships. *J Am Chem Soc* **2009,** 131, (37), 13415-22.

339. Yuzwa, S. A.; Vocadlo, D. J., O-GlcNAc modification and the tauopathies: insights from chemical biology. *Current Alzheimer research* **2009,** 6, (5), 451-4.

340. Macauley, M. S.; He, Y.; Gloster, T. M.; Stubbs, K. A.; Davies, G. J.; Vocadlo, D. J., Inhibition of O-GlcNAcase using a potent and cell-permeable inhibitor does not induce insulin resistance in 3T3-L1 adipocytes. *Chem Biol* **2011,** 17, (9), 937-48.

341. Macauley, M. S.; Shan, X.; Yuzwa, S. A.; Gloster, T. M.; Vocadlo, D. J., Elevation of Global O-GlcNAc in rodents using a selective O-GlcNAcase inhibitor does not cause insulin resistance or perturb glucohomeostasis. *Chem Biol* **2011,** 17, (9), 949-58.

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

343. Dorfmueller, H. C.; Borodkin, V. S.; Schimpl, M.; Shepherd, S. M.; Shpiro, N. A.; van Aalten, D. M., GlcNAcstatin: a picomolar, selective O-GlcNAcase inhibitor that modulates intracellular O-glcNAcylation levels. *J Am Chem Soc* **2006,** 128, (51), 16484-5.

344. Dorfmueller, H. C.; Borodkin, V. S.; Schimpl, M.; van Aalten, D. M., GlcNAcstatins are nanomolar inhibitors of human O-GlcNAcase inducing cellular hyper-O-GlcNAcylation. *Biochem J* **2009,** 420, (2), 221-7.

345. van Aalten, D. M. F.; Dorfmueller, H. C.; Borodkin, V. S.; Schimpl, M.; Zheng, X. W.; Kime, R.; Read, K. D., Cell-Penetrant, Nanomolar O-GlcNAcase Inhibitors Selective against Lysosomal Hexosaminidases. *Chemistry & Biology* **2010,** 17, (11), 1250-5.