METABOLIC PATHWAYS FOR NATURAL AND UNNATURAL SIALIC ACIDS

APPROVED BY SUPERVISORY COMMITTEE

Jennifer Kohler Ph.D.

Joachim Seemann Ph.D.

David Mangelsdorf Ph.D.

Zhijian Chen Ph.D.

DEDICATION

First and foremost I need to thank my boss (the boss) Jennifer Kohler for taking a chance on me so many years ago. I have learned how to think about science, how to ask good questions, and also how to form complete sentences when giving a talk or writing a proposal. Especially, I am thankful for the freedom you have given me over the years to pursue my crazy ideas (and at my crazy hours).

While writing my thesis, I couldn't help but think back to the first thesis I ever read, Jim Fleming's, the first graduate student I worked for when I was an undergraduate at Stanford. We had a great time listening to Ben Folds and I still use lessons I learned from you when I perform science today. I also am grateful for Justin Du Bois for giving me my first opportunity in science and teaching me the mantra, 'work hard, play harder.' I must also thank William Nelson for letting me into his lab and introducing me to biology research.

There are so many people to thank at UT southwestern. Thank you to my thesis committee for taking time out of their busy days. I hope you learned at least one thing from me. Thank you Xiaodong Wang and Gang Yu for letting me rotate through your labs. I

learned so much in such a short amount of time. My colleagues from medical school, who always make fun of me for still being a student, have been an important source of joy and you guys have kept me sane over the years. The students of the MSTP, the only people who really know what this experience is like and how difficult it can be at times. Sometimes there is just no replacement for sharing my misery with another MSTP. I would like to thank Subu, not only for helping me with my project but for the many scientific discussion we have had over the years. I hope we continue to talk science in the future.

In terms of Kohler lab members: Pete and Randy, it seems like yesterday that we started. Michelle and Chad, you guys helped me out so much when I first joined the lab. Chad you taught me how hustle things from vendors and Michelle, I really did enjoy our scientific "discussions". I hope you remember them fondly too. Seokho Yu, Bin Li, Akiko, Fan, and Janet, you are all so technically gifted and I am lucky to have gotten a chance to see you operate. I am extremely thankful for the opportunity to mentor Charles Fermaintt, a talented rotation student. Yibing, we have spent many late nights together here and I enjoyed our many many scientific discussions. I also appreciate your hard work and dedication to science. Andrea, I always appreciated your enthusiasm and energy. You really brought the lab closer together and the lab has been a bit more fun since you joined. Lastly, I must thank Amberlyn. I never would have made it to the end without your help. We are more than lab friends – we are real friends. I have confided in you about so many things over the years both science related and personal and you have become like a sister to me. Thank you for everything, I promise I will always help you with computer/electronic stuff and I hope that we will still be making fun of each other years from now.

METABOLIC PATHWAYS FOR NATURAL AND UNNATURAL SIALIC ACIDS

by

NAM DINH PHAM

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

July, 2014

METABOLIC PATHWAYS FOR NATURAL AND UNNATURAL SIALIC ACIDS

Nam Dinh Pham

The University of Texas Southwestern Medical Center at Dallas, 2014

Supervising Professor: Jennifer Kohler Ph.D.

Carbohydrates, along with lipids, nucleic acids, and amino acids constitute the four main building blocks of life. This thesis will focus on sialic acid, a unique, nine-carbon sugar that is critically important in humans. In **Chapter 1**, I will introduce sialic acid biology and the many diseases that result when sialic acid metabolism is disrupted. Furthermore, I will discuss metabolic carbohydrate engineering as a valuable tool for studying sialic acid biology. Lastly, I will introduce an ancient and conserved family of proteins, the glycosyltransferases, which are responsible for transferring sugars, such as sialic acid, onto glycoproteins and glycolipids.

Disruptions in the sialic acid biosynthetic pathway are implicated in hereditary inclusion body myopathy (hIBM), a disease of aging. In **Chapter 2,** I determined that sialic acid biosynthesis alters the levels of UDP-GlcNAc, a product of the hexosamine biosynthetic pathway. This results in changes in the branch structure of N-glycans and impairs galectin-1 binding to cells. Furthermore, I found that sialylation of N-glycans can further influence N-glycan branching. Taken together, my work suggests unexplored mechanisms for hIBM pathogenesis.

The use of unnatural sialic acids, such as SiaDAz, and metabolic carbohydrate engineering is enabling deeper understanding of the biological roles of sialic acid. In **Chapter 3**, I determined that different cell lines have impairments at different metabolic steps during SiaDAz synthesis. The results of my work reveals the importance of cell-specific enzymes during the metabolism of Ac_4ManNDAz to ManDAz to SiaDAz. Furthermore, I found that cells poorly deprotect Ac_5 1-OMe sialic acids, both natural and unnatural. The results of this chapter inspire strategies to improve the efficiency and generality of this technology.

Mature glycans on the cell surface or on therapeutically active small molecules are formed by the action of glycosyltransferases. In **Chapter 4**, I apply statistical coupling analysis to reveal new insights into glycosyltransferases. Specifically, I was able to identify a network of key residues, termed a sector, which contained non-obvious key residues that appear to be important for catalytic function. Furthermore, sector residues correlated with surface sites that exert allosteric control over enzyme activity.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

Pham, N. D., Parker, R. B., and Kohler, J. J. (2013) Photocrossslinking approaches to interactome mapping. *Cuurent Opinion in Chemical Biology* **17**, 90-101.

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF DEFINITIONS

hIBM – Hereditary Inclusion Body Myopathy

sIBM – Sporadic Inclusion Body Myositis

NeuAc – N-acetyl Neuraminic acid

Gal – Galactose

ManNAc – N-acetyl Mannosmine

GlcNAc – N-acetyl Glucosamine

GalNAc – N-acetyl Galactosamine

LacNAc – N-acetyl Lactosamine

HBP – Hexosamine Biosynthetic Pathway

UDP-GlcNAc – Uridine Diphosphate N-acetyl Glucosamine

ManNAc-6-P – N-acetyl Mannosamine 6-Phosphate

CMP-NeuAc – Cytodine Monophosphate Neuraminic Acid

NeuAc-9-P – Neuraminic acid 9-phosphate

UPR – Unfolded Protein Response

GT – Glycosyltransferase

SCA – Statistical Coupling Analysis

CDG – Congenital Disorder of Glycosylation

MSA – Multiple Sequence Alignment

CAZy – Carbohydrate Active enZYme database

CHAPTER ONE
Introduction to Sialic Acid Biology

INTRODUCTION

Carbohydrates: essential building blocks of life

Carbohydrates, along with lipids, nucleic acids, and amino acids constitute the four main building blocks of life. In fact, genomic studies suggest that 3-5% of the genome is dedicated to carbohydrate metabolism (2). Yet despite their equal and fundamental importance, research into the cellular biology of carbohydrates and their roles beyond energy production have historically been lacking in comparison to the other critical biomolecules. However, genetic studies have revealed that a severe and diverse class of human diseases is caused by mutations to genes devoted to processing carbohydrates (3). More importantly, the pathology of these diseases is likely related to the functions these carbohydrates serve as receptors, ligands, or structural molecules within critical biological processes (4). To this end, the recent development of unique tools to study glycobiology has, in part, surged an explosion in research dedicated to understanding the biological roles of carbohydrates within these cellular processes.

This introductory chapter will focus on sialic acid, an acidic, nine-carbon sugar. First, the complex biosynthetic pathway of sialic acid will be described, as well as the human diseases that are caused by disruptions in sialic acid biosynthesis. Next, the development of novel sialic acid analogs as tools for studying sialic acid biology will be examined. Finally, sialic acids are transferred onto substrates by a family of enzymes called sialyltransferases- a

sub-family of the glycosyltransferase (GT) superfamily of proteins. The last part of the chapter will focus on the diversity of glycosyltransferase (GT) structure and function, as well as highlight efforts to engineer GT's to control their biological behavior.

Sialic acids: a family of nine-carbon sugars

Sialic acids are a family of nine carbon polyhydroxylated α-keto acids utilized by deuterostomes, and are extremely prominent in all vertebrates (5). The predominant sialic acid found in humans is N-acetyl neuraminic acid (NeuAc) (figure 1.1). It is typically found as the terminal sugar on glycans, and its outermost location on the cell surface places it in an ideal position to mediate cellular interactions (6). Indeed, containing glycoconjugates on leukocytes serve as receptors for E-selectin and P-selectin on endothelial cells during adhesion events (7,8). NeuAc can be linked to galactose (Gal) or N-acetylgalactosamine (GalNAc) in either an α 2,3 or an α 2,6 linkage (figure 1.1). NeuAc can also be connected to other NeuAc molecules in an α 2,8 linkage to form a dimer or a homopolymer of polysialic acid. The type of linkage that sialic acid is presented can drastically change the ability of proteins to interact with the glycoconjugate. For example, the hemagglutinin protein from different strains of influenza virus has various specificities for different linkages of sialic acid (9). In addition to acting as a ligand for proteins, placement of NeuAc on proteins can prevent interactions. Indeed, sialylation of NCAM is thought to prevent spurious binding events (10). Thus transfer of sialic acid onto the cell surface can have dramatic consequences, and its biosynthesis should be regulated.

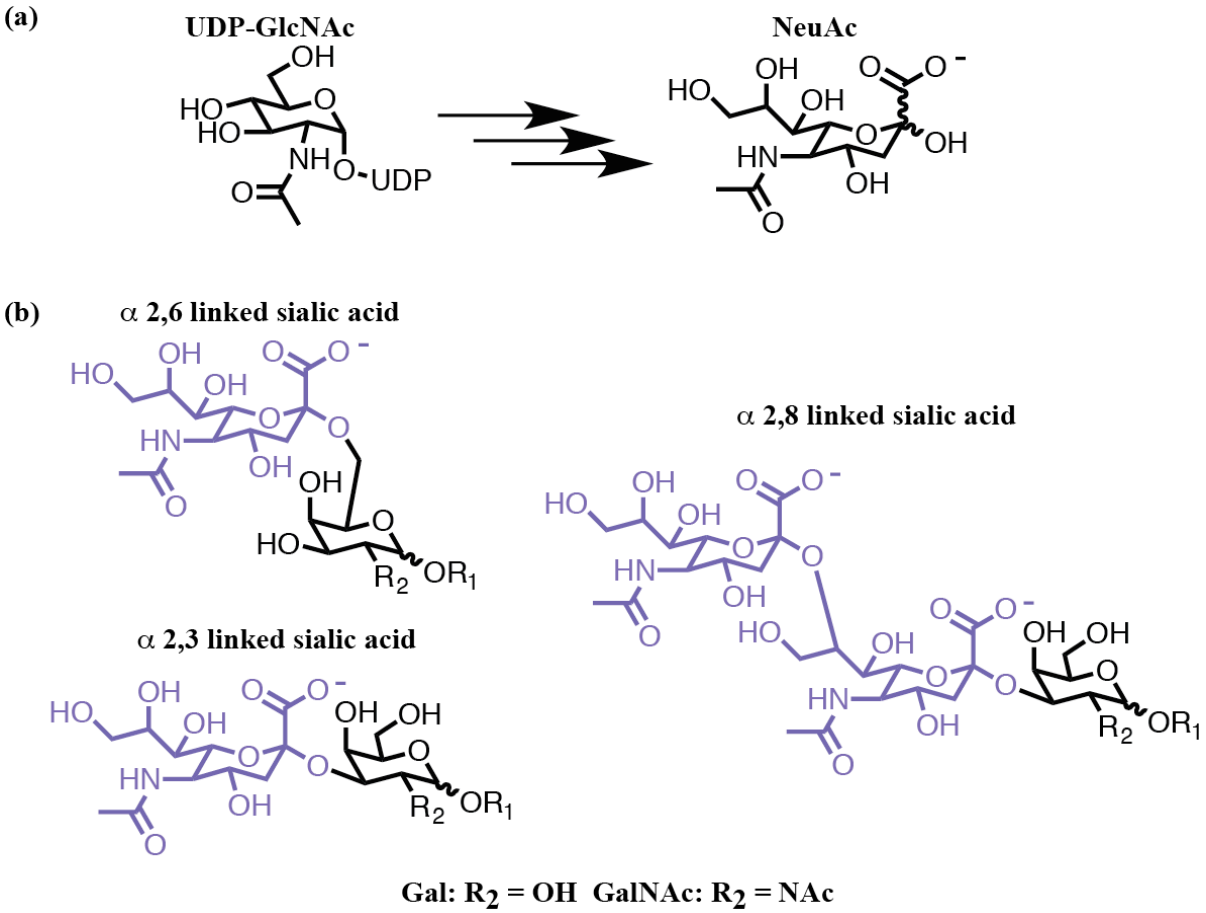

Figure 1.1. The fundamentals of sialic acid biology a) Sialic acid is made in several steps from UDP-GlcNAc. b) Sialic acid is found in glycoconjugates in three different linkages. It is typically attached to either Gal, GalNAc, or another sialic acid molecule.

Sialic acid biosynthesis and regulation in humans

NeuAc biosynthesis is unusually complex and encompasses multiple cellular organelles (Figure 1.2) (11). NeuAc is synthesized in several steps from Uridine diphosphate N-acetyl glucosamine (UDP-GlcNAc), a key product of the hexosamine biosynthetic pathway. The GNE protein catalyzes the first committed step in sialic acid biosynthesis. UDP-GlcNAc is first epimerized to N-Acetyl mannosamine (ManNAc) by GNE a bifunctional enzyme that also phosphorylates ManNAc to ManNAc-6-phosphate (ManNAc-

6-P). In mice, a homozygous knockout of *Gne* is early embryonic lethal (12). ManNAc-6-P is then condensed with phosphoenolpyruvate to from NeuAc-9-Phosphate by NANS (NeuAc-9-P). NeuAc-9-P is then dephosphorylated by NANP to form free NeuAc. Free NeuAc can also be scavenged in the lysosome, whereby the SIALIN protein transfers sialic acid from the lysosome into the cytosol (13).

The biosynthesis of NeuAc is thought to occur in the cytosol but the activation of NeuAc to cytodine monophosphate NeuAc (CMP-NeuAc) by CMAS occurs within the nucleus. NeuAc enters the nucleus and CMP-NeuAc exits the nucleus presumably by simple diffusion. NeuAc is the only sugar that is activated to its nucleotide sugar form within the nucleus. While the localization of CMAS in the nucleus is somewhat mysterious, its localization is conserved in vertebrates and multiple nuclear localization signals for the protein have evolved (14,15). Following activation, CMP-NeuAc exits the nucleus and into the cytosol, and it is then exchanged into the secretory pathway by SLC35A1, an anti-porter. Once in the Golgi, CMP-NeuAc serves as a substrate for 1 of 20 human sialyltransferases (16).

There are several points of regulation of sialic acid biosynthesis. For instance, the activity of GNE is inhibited by CMP-NeuAc (17). Also, the chemical-physical nature of sialic acid can be modified by acetylation by an unknown sialic acid acetyl transferase. These acetyl groups can then be removed by the acetyl esterase SIAE (18) and disruptions in the enzyme are implicated in chronic inflammatory disorders (19). Next, the expression of sialic acid on the cell surface can be altered by differential expression of sialyltransferases.

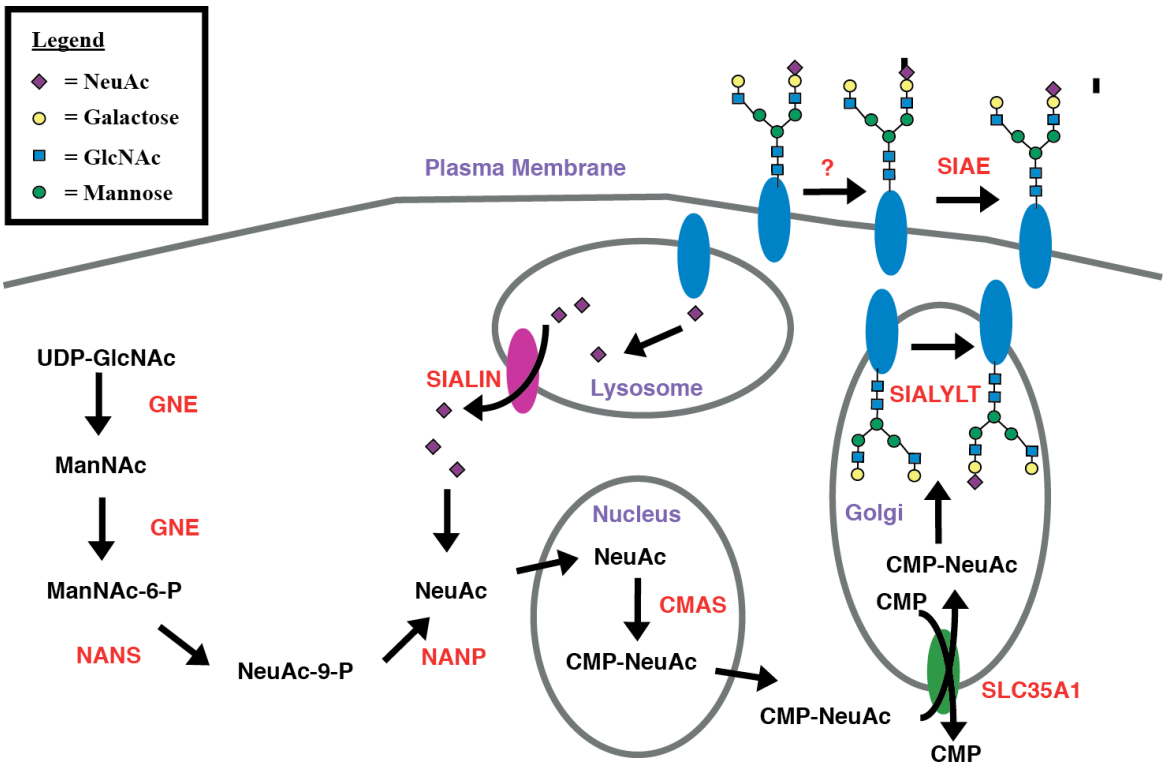

Figure 1.2. Sialic acid biosynthesis in humans. The enzyme GNE catalyzes the first committed step of sialic acid biosynthesis. GNE is a bifunctional enzyme that converts UDP-GlcNAc to ManNAc and phosphorylates it to ManNAc-6-P. NANS condenses ManNAc-6-P with phosphoenolpyruvate to form NeuAc-9-P which is dephosphorylated to NeuAc by NANP. NeuAc is then activated to CMP-NeuAc by the CMAS protein located within the nucleus. CMP-NeuAc is then exchanged into the Golgi for CMP. Within the Golgi, CMP-NeuAc can then act as a donor for 1 of 20 known human sialyltransferase. Additionally, NeuAc can be scavenged from lysomal transporter SIALIN. Once on glycoproteins, sialic acid can be acetylated by an unknown acetyltransferase and these ester groups are removed by the esterase SIAE.

Indeed, cancers are thought to upregulate sialyltransferases like ST6GALI (20). This is thought to help in the avoidance of the immune system and cell death (21). Furthermore, recent evidence suggests that upregulation of a sialyltransferase such as ST6GALNACV can influence cancer metastasis (22). Finally, once NeuAc is on the cell surface, a class of enzymes known as sialidases can remove it. This class of enzymes are also implicated in

cancer progression (23). Thus, sialic acid is clearly important to the normal function of a cell and it is perhaps unsurprising that alterations in sialic acid biosynthesis influences human health.

GNE and human health and disease

Multiple human diseases result from mutations in the sialic acid biosynthetic pathway (Table 1.1). Mutations to GNE lead to two different human diseases. In the disease sialuria, autosomal dominant mutations to GNE remove the feedback inhibition by CMP-NeuAc and cells overproduce sialic acid (17). This disease is characterized by mild development delay, course facies, and massive amounts of sialic acid in urine (24). More severe symptoms of this disease are pulmonary obstruction, hepatomegaly, and seizures. Other mutations in GNE cause hereditary inclusion body myopathy (hIBM), a disease of aging that is inherited in an autosomal recessive manner (25). Patients with the disease are normal at birth, but at around 20 years of age they begin to develop relentlessly progressive asymmetric muscle wasting. Over 60 mutations in both the protein domains of GNE have been linked to this disease, but how these mutations in GNE lead to disease is poorly understood (26).

Table 1.1. Disruptions in sialic acid biology lead to a variety of human diseases. Additionally, sialyltransferases are not listed on the table, but have been implicated in a number of human diseases.

Disease name	Gene implicated	Gene Function	Symptoms
Hereditary inclusion body myopathy (hIBM)	*GNE*	Catalyzes first committed step in sialic acid biosynthesis	Normal at birth, muscle weakness and wasting beginning in young adulthood
Sialuria	*GNE*		High amounts of secreted sialic acid. Possible liver, pulmonary, neurological disfunction
Sialic acid storage disease	*SIALIN*	Lysosomal sialic acid transporter	Hypotonia, cerebellar ataxia, and mental retardation
Congenital disorder of glycosylation (Type-IIF)	*SLC35A1*	Exchanges CMP-Sialic acid for CMP at the Golgi	thrombocytopenia, abnormal platelets
Chronic autoinflammatory disorders (Type I diabetes, rheumatoid arthritis, etc…)	SIAE	Sialic acid esterase	Depends on organ affected but increased autoimmune activity

The simplest prediction is that hIBM *GNE* mutations lead to decreases in sialic acid production and inadequate sialylation of critical cell surface molecules. Indeed, *GNE* with hIBM-causing mutations encodes a protein with reduced *in vitro* enzymatic activity, but levels of residual enzymatic activity do not correlate with disease severity (27,28). Furthermore, reduced *GNE* activity does not necessarily result in reduced total cell surface sialic acid content (29,30). While total cell surface sialic acid levels are similar between hIBM patients and normal humans, certain glycoproteins appear to be hyposialylated in hIBM patients but the clinical relevance of this finding has not been clearly demonstrated

(31-33). Most of the research on hIBM has focused on extracellular sialic acid, but it is possible that hIBM mutations in GNE alter intracellular metabolic pathways such as the hexosamine biosynthetic pathway. Given the intersection of sialic acid biosynthesis and human disease, the field of sialic acid biology is worthy of deeper exploration.

Chemical tools for studying sialic acid biology

Traditional biochemical techniques have advanced the field of glycobiology, but they are significantly augmented by the use of metabolic carbohydrate engineering. Metabolic carbohydrate engineering refers to the usage of carbohydrate analogs that contain bioorthogonal functional groups – that is functional groups that are tolerated by endogenous machinery but are not normally found in nature (34). To date, a large variety of sialic acid analogs that harbor a variety of functional groups have been developed (Figure 1.3) (35). Sialic acid analogs can be introduced into cells by culturing them with either the sugar itself or its ManNAc analog precursor (Figure 1.3). This variety is possible due to the intrinsic promiscuity of endogenous sialic acid metabolic machinery (36). Sialic acid and ManNAc analogs can be used as unprotected sugars, but these polar sugar analogs typically exhibit poor cell permeability and require high concentrations to enter the cell. A pro-drug strategy, in which the hydroxyl groups are acylated and the carboxylic acid group is esterified, makes the analogs more cell-membrane permeable (37). Once inside the cell, the acyl protecting groups are presumably removed by endogenous esterases (Figure 1.4) (38).

These unnatural sugar analogs help overcome weaknesses associated with using traditional biochemical techniques. Carbohydrate-protein interactions are typically transient,

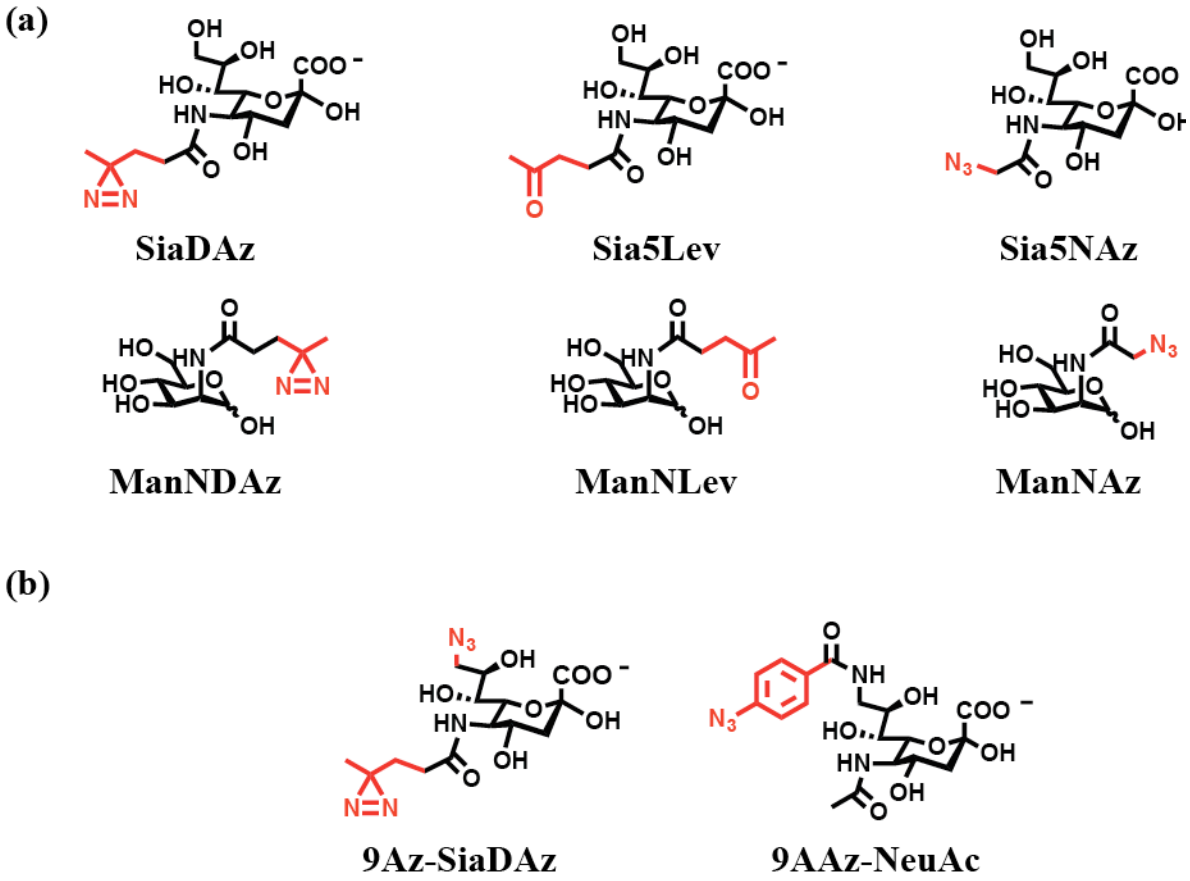

Figure 1.3. Examples of sialic acid analogs and ManNAc analogs. (a) Mono-functional sialic acid analogs and the ManNAc analog precursors that can be used to generate them. (b) Sialic acid analogs that cannot be generated from ManNAc analogs *in cellulo*. Large substitutions at the C6 position of ManNAc are not tolerated by metabolic machinery.

low-affinity, and sub-stoichiometric. Thus, methods such as immunoprecipitation are ill suited to characterize these types of interactions. For instance, to aid in identifying sialic acid containing conjugates, a purification handle (azide) can be added to sialic acid,

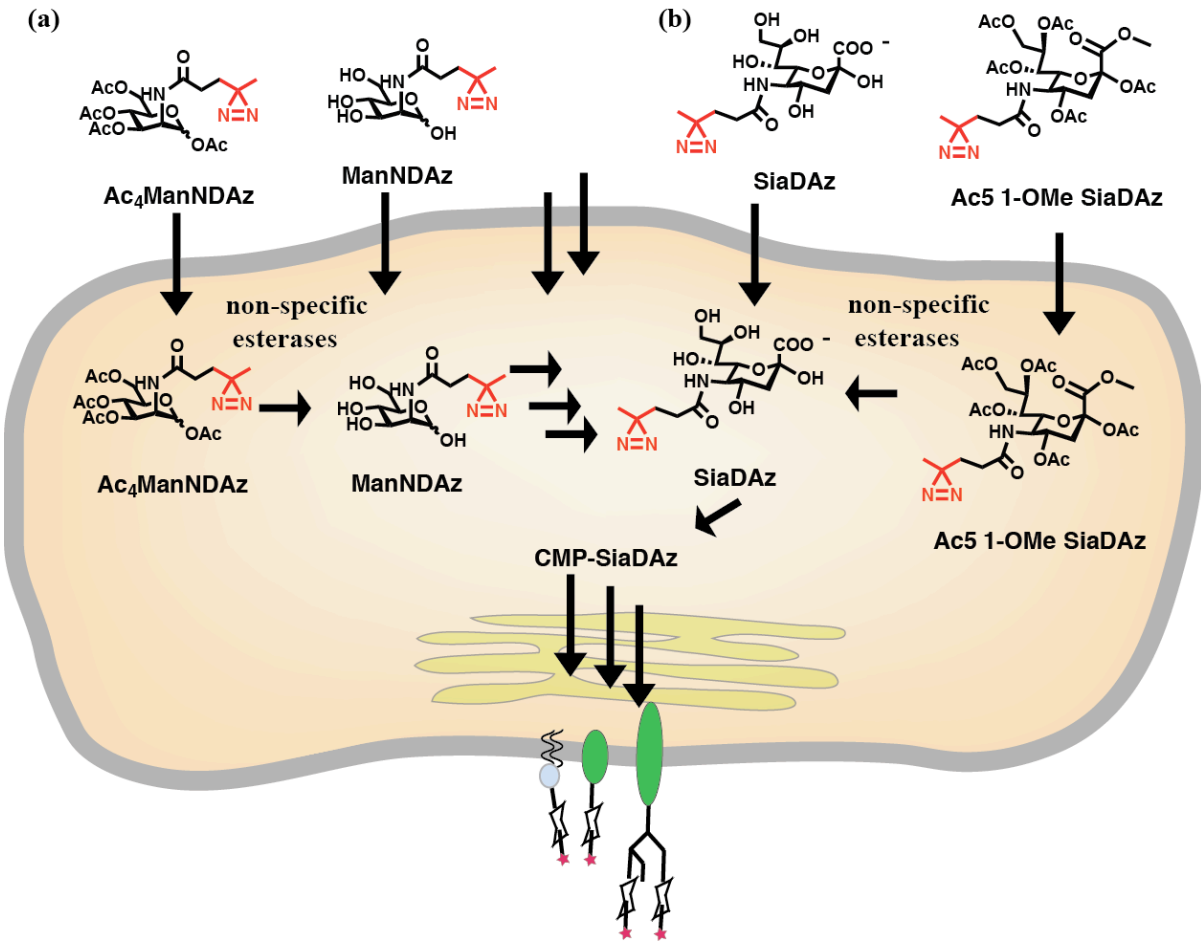

Figure 1.4. Metabolic carbohydrate engineering utilizing unnatural sialic acids and their precursors. (a) A sialic acid analog precursor such as ManNDAz can be provided to the cell in its free form or after full acylation. Acetylation makes the compound more hydrophobic so that it can cross the membrane at lower concentrations. (b) Alternatively, free sialic acids analogs such as SiaDAz can be supplemented in cell culture media. Esterification of all the free alcohols and the carboxylic acid, Ac_5 1-OMe SiaDAz in this example, makes the compound more hydrophobic allows it to traverse the cell membrane at lower concentrations. Cellular machinery converts the analogs to the activated nucleotide sugar form where it is transferred to glycoproteins and glycolipids.

Sia5NAz, and sialic acid containing glycoconjugates can be purified following chemoselective chemistry (39) (Figure 1.5). Alternatively, sialic acid binding proteins can be isolated using an analog containing a photocrosslinking group such as a diazirine, SiaDAz

(40). Once on the cell surface, UV-irradiation activates the functional group and covalently captures nearby proteins. These complexes can then be isolated using immunoprecipitation of the protein of interest (Figure 1.5). These functionalities are not mutually exclusive, indeed, bifunctional analogs containing both a crosslinking group and purification handle have been recently characterized (41) (Figure 1.5). The development of sialic acid analogs has enabled scientists to glean great insights into sialic acid biology that would be otherwise impossible. For example, the in-vivo binding partner of the sialic acid binding protein CD22, an important regulator of b-cell signaling, was identified using a sialic acid analog and metabolic engineering (42,43). However, metabolic roadblocks that currently limit the use of sialic acid analogs still need to be addressed to improve the generality of this technology. Overall, the use of sialic acid analogs is still in its infancy, and in time, its use will provide insight into the biological mechanisms of how sialic acid mediates physiological and pathological processes.

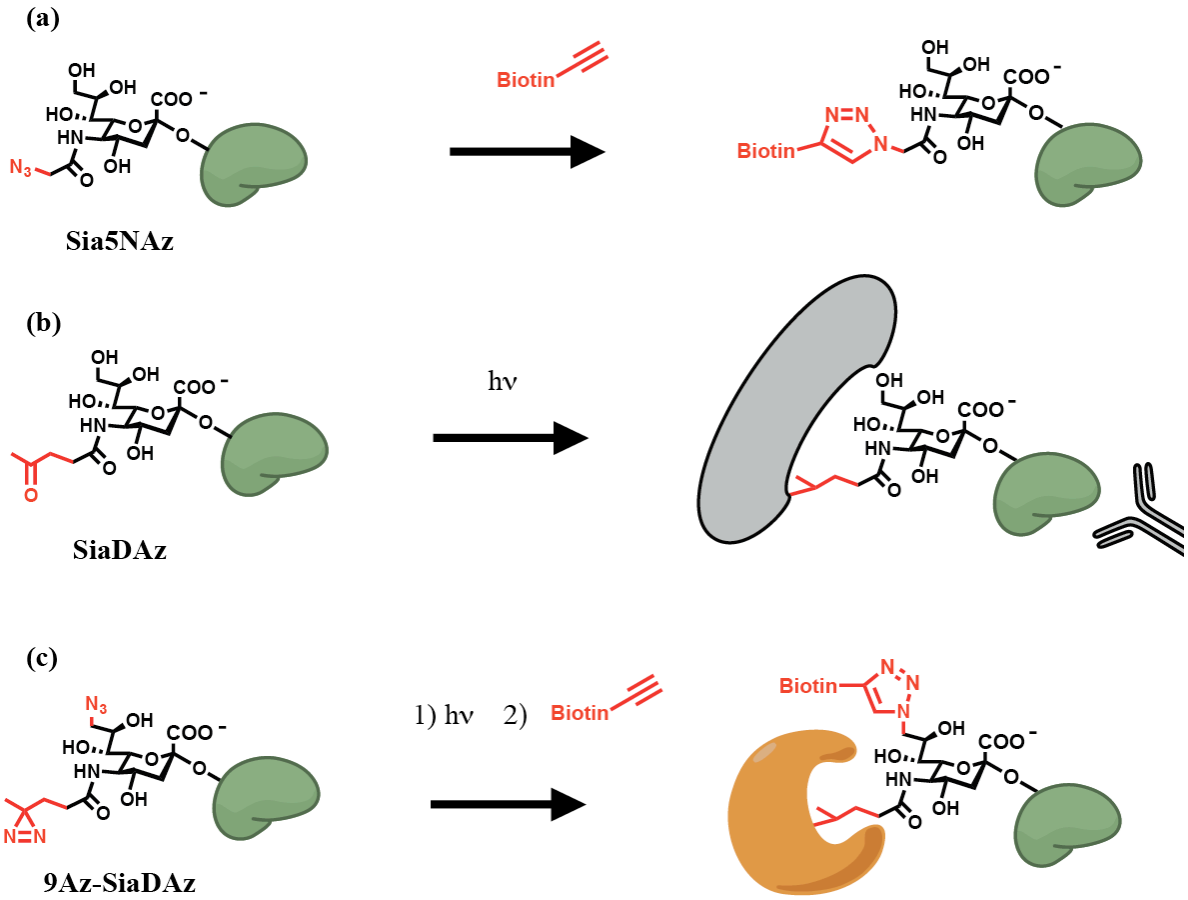

Figure 1.5. Sialic acid analogs are powerful tools to study glycobiology. (a) A sialic acid analog such as Sia5Az can be utilized to purify/identify sialylated proteins. The azide can easily be reacted with a purification handle (such as biotin) or a fluorophore. (b) Sialic acid -protein interactions can be ascertained using a photo-crosslinking analog like SiaDAz. Crosslinked complexes are formed following UV irradiation and can be purified by traditional immunoprecipation techniques. The covalent bond stabilizes the traditionally weak sugar-protein interaction. (c) Sialic acid analogs can contain multiple functionalities. In this example, 9Az-SiaDAz contains a diazirine to crosslink a binding partner and an azide for purification. This would be particularly useful when good antibodies do not exist for the proteins of interest.

A necessary feature of all carbohydrate metabolic engineering strategies requires the transfer of the unnatural sugar to a protein/carbohydrate/lipid via the action of a glycosyltransferase (GT). Sialic acids, whether natural or unnatural, are transferred to the cell surface by a class of enzymes known as sialyltransferases (ST's) – a subfamily of the GT superfamily. There are 20 known human ST's and it is currently unclear which of these enzymes can accept unnatural sugars. Understanding whether these enzymes can accept unnatural sugars and discovering the key residues that control substrate specificity would accelerate the field of carbohydrate metabolic engineering.

Glycosyltransferase function and engineering

GT's are an ancient family of proteins that catalyze the transfer of activated nucleotide sugar donor to a protein, lipid, or another carbohydrate acceptor (44) (Figure 1.6). The transfer of sugar can either occur with an inversion or retention of stereochemistry, and GT's are classified as inverting and retaining, respectively. GT's are encoded by one of the largest gene families and every non-viral genome contains at least one GT gene. In humans, there are over 200 GT's according to the CAZy database (45). GT's that utilize nucleotide sugars have probably evolved twice into two structural superfamilies, GT-A and GT-B (Figure 1. 7), with both structural families containing inverting and retaining enzymes.

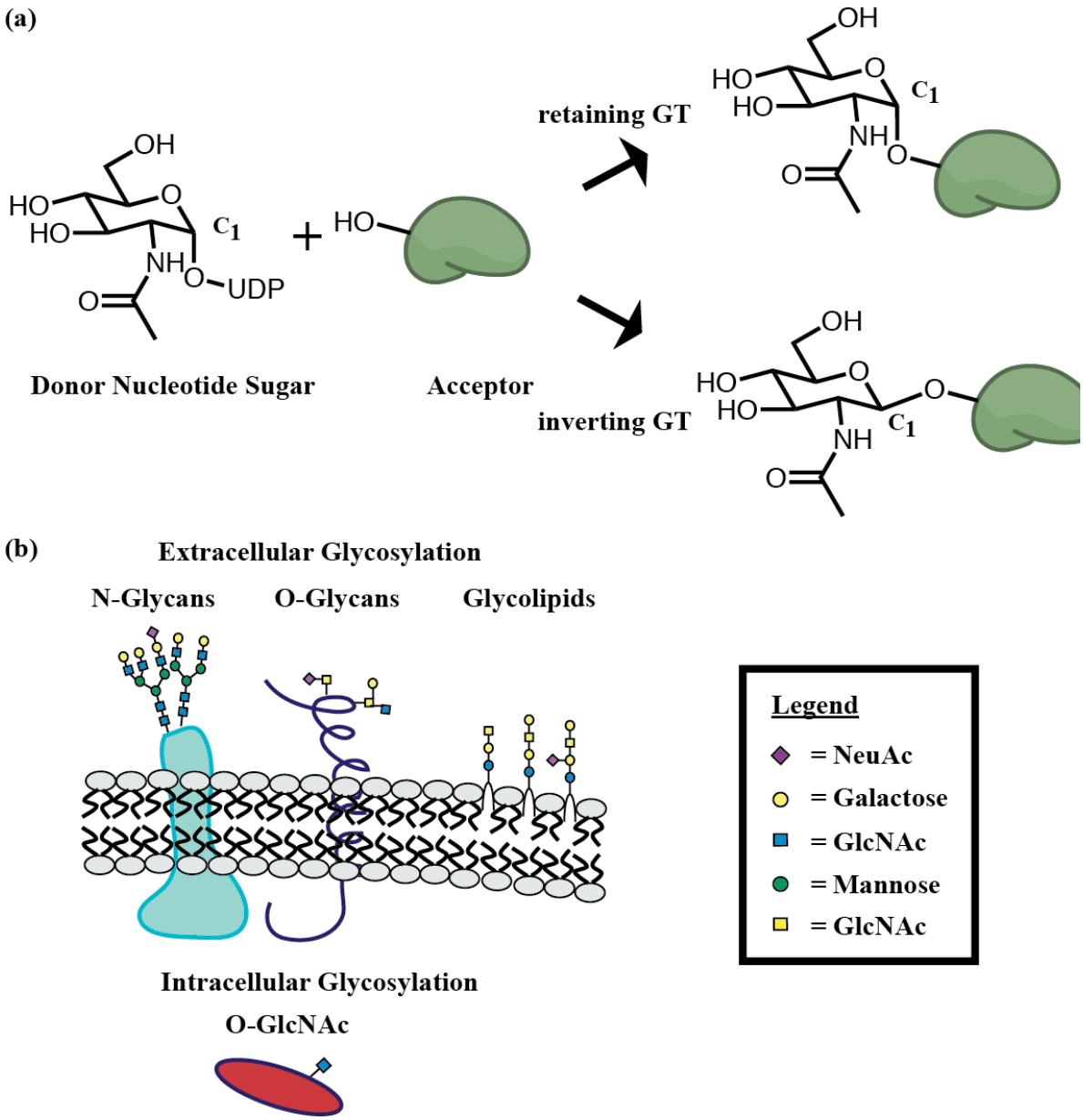

Figure 1.6. The glycosyltransferase reaction and common human glycans. (a) Glycosyltransferases discussed in this thesis utilize a donor nucleotide sugar (UDP-GlcNAc in this example) and an acceptor that could be a protein, glycan, or lipid. Glycoysltransferases are classified as being retaining or inverting depending on whether they retain the stereochemistry at C_1 of the sugar being transferred. (b) A small sample of the common classes of glycans in humans. Each class has much more diversity than depicted and there are many more types of glycosylation.

The structural families are characterized as having two rossmann folds, which in the GT-A structural class are joined together, but are connected by a flexible linker in that of GT-B. Remarkably, sequences within each structural class are only ~12% similar (Figure 1.7). The GT-A family of GT's has a "conserved" DxD motif that is found in many, but not all, GT-A enzymes, while the GT-B enzyme has no conserved sequence motif. This large diversity in sequence space reflects the variety in the sugar being transferred, the linkages by which it can be connected to its substrate, and the number of possible substrates (46). Indeed, there are often multiple gene products that carry out very similar reactions – in humans, there are 20 sialyltransferases, 20 ppGalNAcT's that transfer a single N-acetyl galactosamine (GalNAc) to serines/threonines, and there are at least 7 N-glycan branching enzymes (Fig 1.F).

Understanding the residues that control glycosyltransferase function is still in its early stages. In this regard, the various GT crystal structures have provided insight into the nature of the nucleotide donor site and acceptor binding site (47). While steric bulkiness can affect binding, it is not easy to predict without co-crystal structures, and in terms of catalytic efficiency, mutations far from the active site of GT's have been demonstrated to have profound effects. For instance, the GT's that control the ABO blood group antigens, GTA and GTB, utilize UDP-GalNAc and UDP-Gal as the donor sugar, respectively. (Not to be confused with the GT-A and GT-B structural class). Remarkably, these proteins are 354 amino acids in length and are identical except at 4 residues (GTA/GTB: Arg/Gly^{176}, Gly/Ser^{236}, Leu/Met^{266}, Gly/Ala^{268}) (48). Indeed, steric bulk at position M266 and A268 prevents the GTB enzyme from binding the bulkier GalNAc. However, position 176 is not

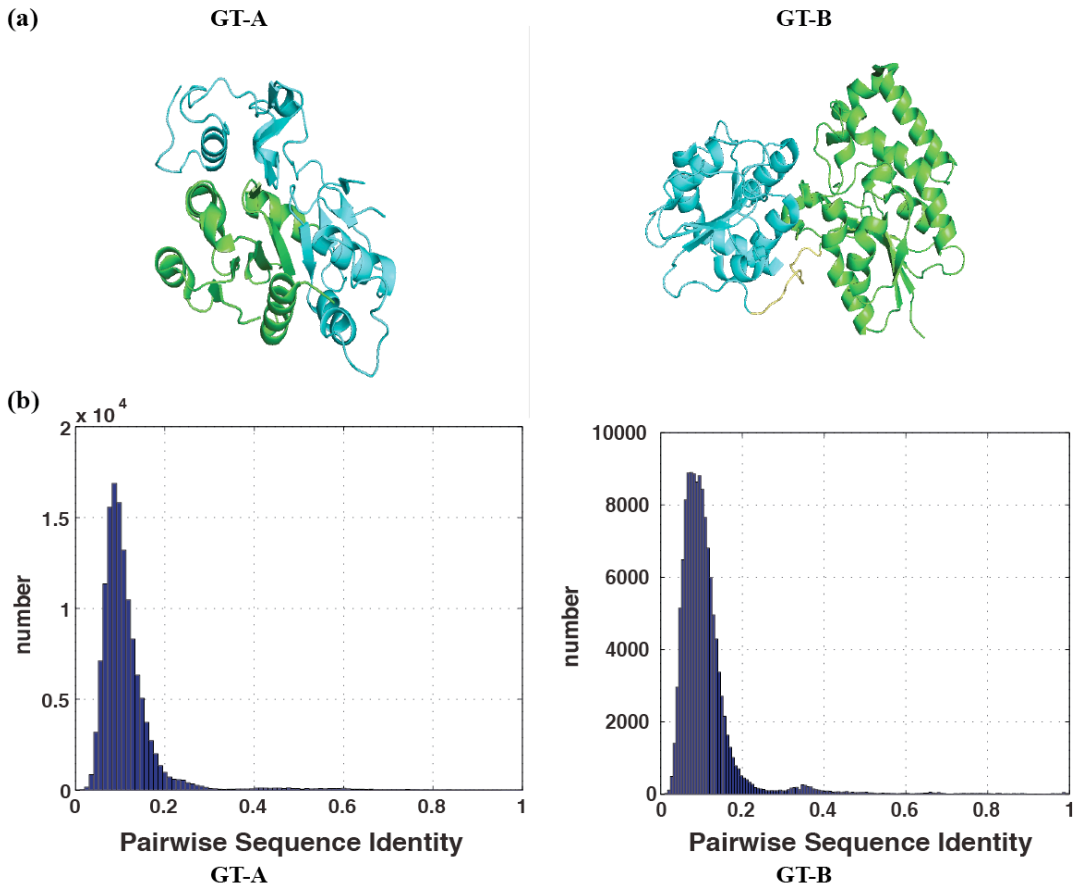

Figure 1.7. Glycosyltransferase structural classes and sequence diversity. (a) GT's are found in two main structural classes, GT-A and GT-B. Both have two rossmann folds (one colored in green and one colored in blue) In the GT-A structural class they are closely adjoined whereas in the GT-B class they are connected by a flexible linker (yellow). (b) Characterized GT's within each class have ~12% pairwise sequence identity. Despite the common fold, sequences within each class are extremely diverse.

directly within the active site, but has large effects on catalytic efficiency by effecting the conformational changes that occur during catalysis (49). Furthermore, in the case of OleD1 GT, an enzyme involved in the biosynthesis of macrolide antibiotics, the protein has natural promiscuity, but a directed evolution experiment revealed three mutations outside the active site that further increased promiscuity and improved catalytic efficiency (50,51). However, it is unclear if these key residues exist across the entire superfamily. Obtaining insight into the

critical residues that govern GT function would not only have an impact on metabolic carbohydrate engineering as described earlier, but also because many natural products with therapeutic value are glycosylated, the engineering of these enzymes to diversify their glycan structure would have profound pharmaceutical implications (52).

CONCLUSIONS

The importance of glycobiology is beginning to be recognized by the scientific community at large due to the increased understanding that disruptions in glycobiology lead to human disease and also due to the development of new tools to explore this relationship. Indeed, hereditary inclusion body myopathy (hIBM), a disease of aging, as well as a variety of other human diseases appear to be sensitive to the sialic acid biosynthetic pathway. **Chapter 2** of this thesis will explore the intersection of hIBM, sialic acid metabolism and the hexosamine biosynthetic pathway. Furthermore, the use of unnatural sialic acids has enabled deep insight into the biological roles of sialic acid, particularly in determining sialic acid – protein interactions. **Chapter 3** examines the metabolic limitations to utilizing these unnatural sialic acids and provides insight into improving the generality of this technology. Finally, sialic acids are transferred to the cell surface by GT's, and a deeper understanding of this family of enzymes will improve efforts to engineer GT's to perform specific tasks. I apply a mathematical analysis of this large, diverse set of enzymes and identify non-obvious key residues that appear to be important for catalytic function in **Chapter 4**.

CHAPTER TWO
Effects of Sialic Acid Biosynthesis on the Hexosamine Pathway and N-linked Glycan Structure

INTRODUCTION

Sialic acid is an essential sugar in mammals – homozygous inactivation of the enzyme that catalyzes the first committed step in sialic acid biosynthesis, *Gne*, is embryonic lethal in mice (12). In humans, over 60 mutations in GNE are linked to hereditary inclusion body myopathy (hIBM), a rare disease of aging that is inherited in an autosomal recessive manner(26). Patients with the disease are normal at birth, but at around 20 years of age they begin to develop relentlessly progressive asymmetric muscle wasting. Despite being clearly linked to sialic acid biosynthesis, how mutations in GNE lead to disease pathogenesis remains a mystery. GNE is a bifunctional protein with an N-terminal epimerase domain that converts UDP-GlcNAc to ManNAc and a C-terminal kinase domain that phosphorylates ManNAc to ManNAc-6-P. Intriguingly, hIBM mutations have been found in both domains but only subtly alter enzyme activity *in vitro,* and this residual enzyme activity does not correlate with disease severity (28-30).

Sialic acid is typically the terminal sugar on glycoproteins and glycolipids, and its position allows it to mediate cellular interactions. For this reason, early work on hIBM focused on potential cell surface hyposialylation (32,33). However, reduced *GNE* activity does not necessarily lead to reduced total cell surface sialic acid content, and sialic acid levels are similar between humans with and without hIBM. A similar phenomenon can be

observed in a mouse model containing the hIBM GNE D176V mutation, in which the mouse is normal at birth but develops muscle weakness after it ages (53). While muscle weakness can be prevented by providing sialic acid or its precursor ManNAc (54), interestingly sialic acid levels are still lower than that of control mice. This implicates a role outside of sialylation during the pathogenesis of hIBM.

Sialic acid is made from UDP-GlcNAc, a product of the hexosamine biosynthetic pathway (HBP), and disruption of hexosamine metabolism has been implicated in the aging process (55). To this end, recent work has linked HBP to the unfolded protein response (UPR) and the prevention of proteotoxicity (55,56). This finding is intriguing in the context of hIBM, as defects in proteostasis may be responsible for the inclusion bodies that are characteristic of this disease. In addition to its influence on the UPR, UDP-GlcNAc levels control N-glycan branching. In mammals, homozygous inactivation of *Mgat5* in mice, the enzyme most responsible for synthesizing branched tetraantennary N-glycans, results in fewer muscle satellite cells and accelerated muscle weakness as the mice age (57). Furthermore, increased hexosamine flux extends the life span in *C. elegans*, possibly through increased N-glycosylation (56). Taken together, alterations in HBP metabolism secondary to subtle alterations in sialic acid biosynthesis may underlie hIBM pathogenesis. Indeed, the role of intracellular changes that might occur as a result of GNE hIBM mutants has been underexplored.

In this chapter, I utilize BJAB K20 cells that do not express GNE to study the effects of hIBM mutants on intracellular metabolism – particularly effects on the HBP. Indeed, I

found that GNE activity does alter UDP-GlcNAc levels, and this results in changes in the branch structure of N-glycans on the cell surface. Furthermore, I found that sialylation of N-glycans can further influence N-glycan structure. Taken together, the results of this chapter shed new light on the relationship between sialic acid synthesis, HBP, and N-glycosylation, and opens new avenues of exploration of hIBM pathogenesis.

RESULTS

Generation of cell lines that express different GNE constructs

To assess the function of hIBM GNE mutants, I utilized BJAB K20 cells that do not express GNE (58). This was a necessity because hIBM is an autosomal recessive disease, and the expression of wild-type enzyme would dominate the mutant enzyme. I made four mutants of GNE – 1) GNE D176V, an hIBM mutant with the mutation in the epimerase domain, 2) GNE M712T, an hIBM mutant with the mutation in the kinase domain, 3) GNE R266Q, a mutant that causes the disease sialuria and sialic acid is overproduced, and 4) GNE Kinase, a construct where the epimerase domain is missing and the enzyme cannot make sialic acid. I also compared K20 cells to K88 cells, a related BJAB cell line. The D176V and M712T mutations were chosen because they occur in separate domains, and because a mouse model containing the GNE mutant has been established. Importantly, both mouse models have a disease phenotype – the GNE M712T mouse has a renal defect likely due to hyposialylation, and intriguingly the GNE D176V mouse model has an hIBM like disease.

The various GNE constructs were introduced into K20 cells via lentiviral infection that allowed for long-term stable expression. Cell lines were subjected to subcellular fractionation, and the resulting intracellular and membrane sialic acid was quantified by DMB derivatization and HPLC analysis (59)(Figure 2.1). As expected, Kinase cells have similar levels of sialic acid as K20 parental cells, and hIBM D176V and M712T mutants produce similar levels of membrane sialic acid as wild type GNE. This reinforces previous evidence that the mutant enzymes have similar sialic acid levels as cells expressing the wild-type enzyme. Surprisingly, Sialuria cells had normal levels of sialic acid (opposite of previous reports) (17). (Overall, early passage Sialuria cells had a more pronounced

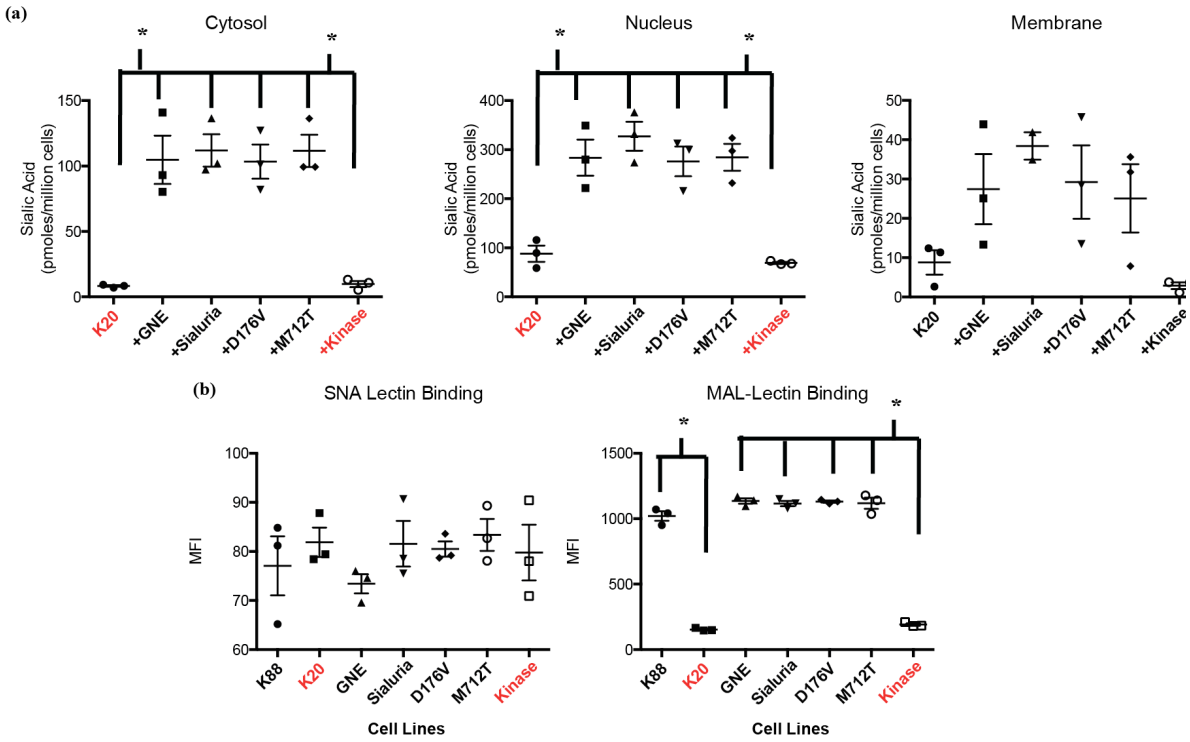

Figure 2.1. Characterization of cells expressing various GNE constructs. K20 cells stably expressing GNE constructs were produced by lentiviral infection. (a) Cells were fractionated and their cytosolic, nuclear and membrane sialic acid was quantified by DMB derivatization. (b) Characterization of cell surface sialic acid by lectin staining and flow cytometry. Cell names in red do not make sialic acid.

difference, but this may have been too difficult for the cells to sustain over time). To my knowledge this is the first time hIBM mutant cells have been fractionated and their intracellular and extracellular sialic acid content quantified. The fractionation protocol suggests that the hIBM GNE mutants produce intracellular amounts of sialic acid similar to GNE WT. The membrane levels of sialic acid were difficult to measure with high precision utilizing the DMB-derivatization method, so I utilized lectin analysis and flow cytometry to more precisely compare cell surface sialic acid levels. All cell lines have similar levels of α 2,6 linked sialic acid as measured by SNA staining, but cell lines that cannot make sialic acid (and have lower levels of intracellular sialic acid) have lower levels of α 2,3 linked sialic

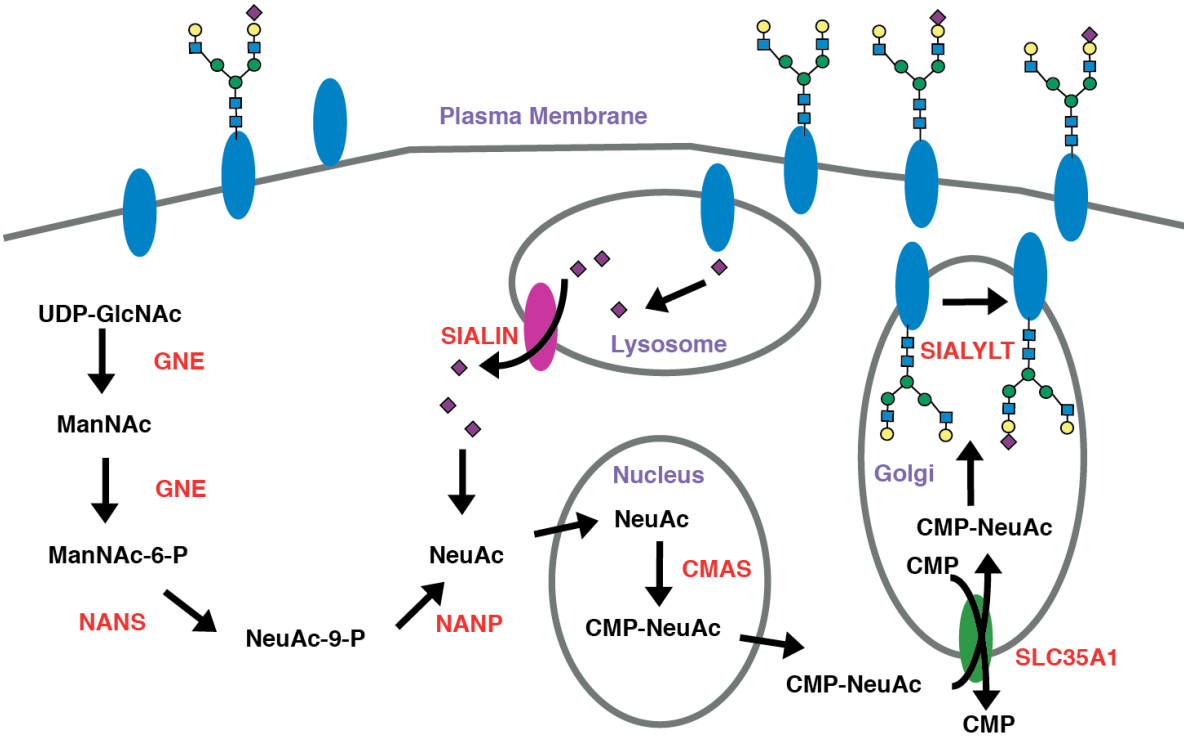

Figure 2.2. Biosynthesis of sialic acids. Sialic acid has a complicated biosynthetic pathway that encompasses multiple cellular organelles. Curiously, sialic acid must enter the nuclear where it is activated to CMP-sialic acid. The purpose of this localization is unknown.

acid as measured by MALII staining. (It should be noted that K20 cells and Kinase cells still have sialic acid because they are able to acquire it from media).

GNE mutants do not cause gene expression changes

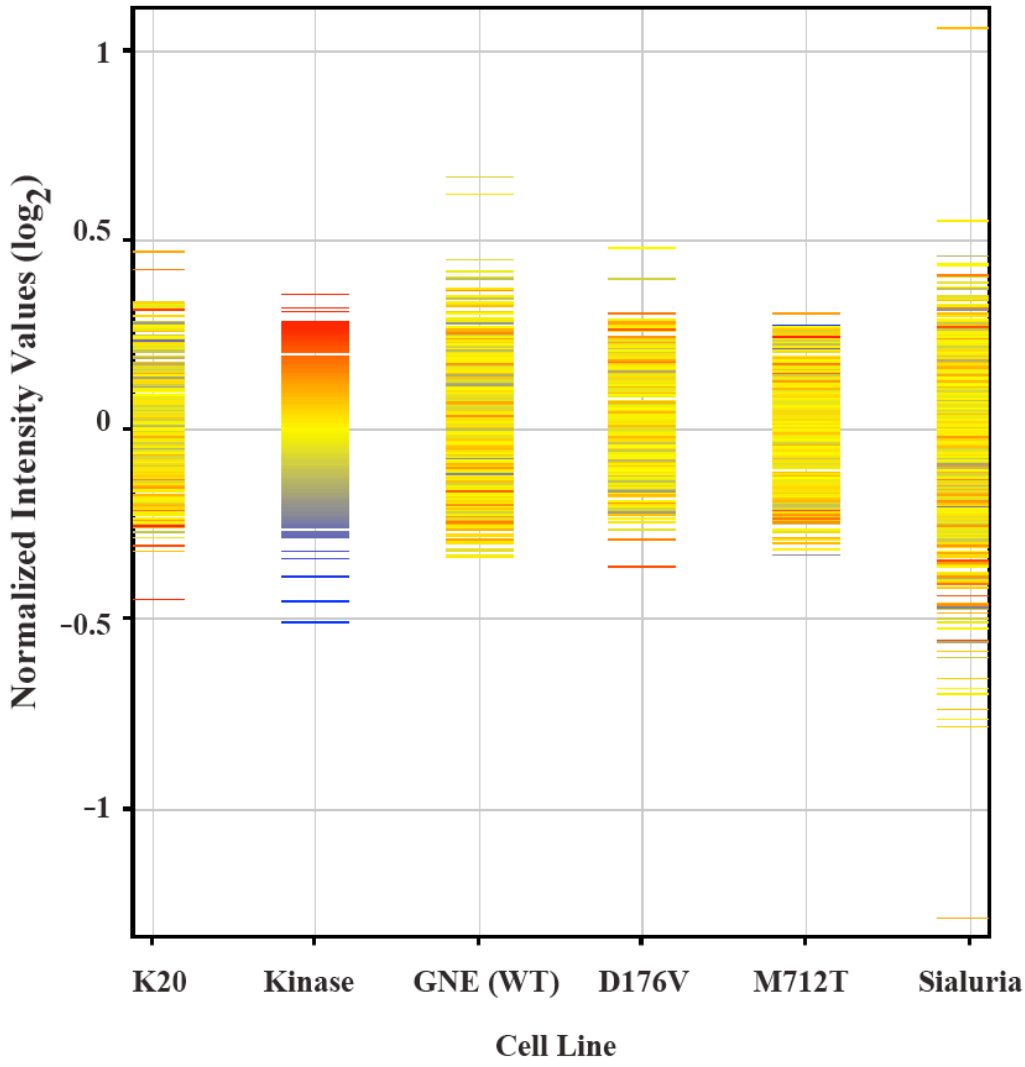

Figure 2.3. Microarray analysis of cell lines that do and do not make their own sialic acid. Cells were analyzed in duplicate on an Illumina HumanHT-12 v4 BeadChip. Cells were nearly identical to each other.

Sialic acid has a complex biosynthetic pathway that encompasses multiple cellular compartments (Figure 2.2). Sialic acid is the only sugar that is activated to its nucleotide sugar form (CMP-sialic acid) in the nucleus. Furthermore, the localization of the sialic acid activating enzyme, CMAS, appears to be conserved (14). I found this intriguing, and I wanted to explore the possibility that sialic acid might act as a transcriptional regulator, or at least initiate signaling events that would lead to changes in gene expression. Indeed, it has been suggested that sialic acid can act as a signaling molecule in certain cell lines (60,61). I took an unbiased approach and performed a microarray experiment looking for gene expression changes between cell lines that expressed active GNE and cells that did not produce sialic acid. Surprisingly, the microarray revealed that all cell lines had very similar levels of gene expression (Figure 2.3). No gene in any cell line was more than 2 fold expressed than from baseline. Furthermore, there were no genes that were expressed consistently at the same level in cells that could make sialic acid and cells that could not. Even in the absence of gene expression changes, there could still be metabolic differences in cells that make sialic acid versus cells that cannot.

GNE activity alters the hexosamine biosynthetic pathway

I next decided to examine whether sialic acid synthesis altered levels of intracellular metabolites. First, I examined the levels of its precursor, UDP-GlcNAc. Cells were cultured in media for 1 or 2 days, and their intracellular metabolites were harvested and examined by high performance anion exchange chromatography with UV detection (HPAEC-UV) (62)(Figure 2.4). After one day, there were no detectable differences between cell lines, but after two days, presumably at lower glucose levels, there appears to be a trend where cells that make more sialic acid have lower levels of intracellular UDP-GlcNAc. To my knowledge, this is the first report that sialic acid production can alter the hexosamine biosynthetic pathway.

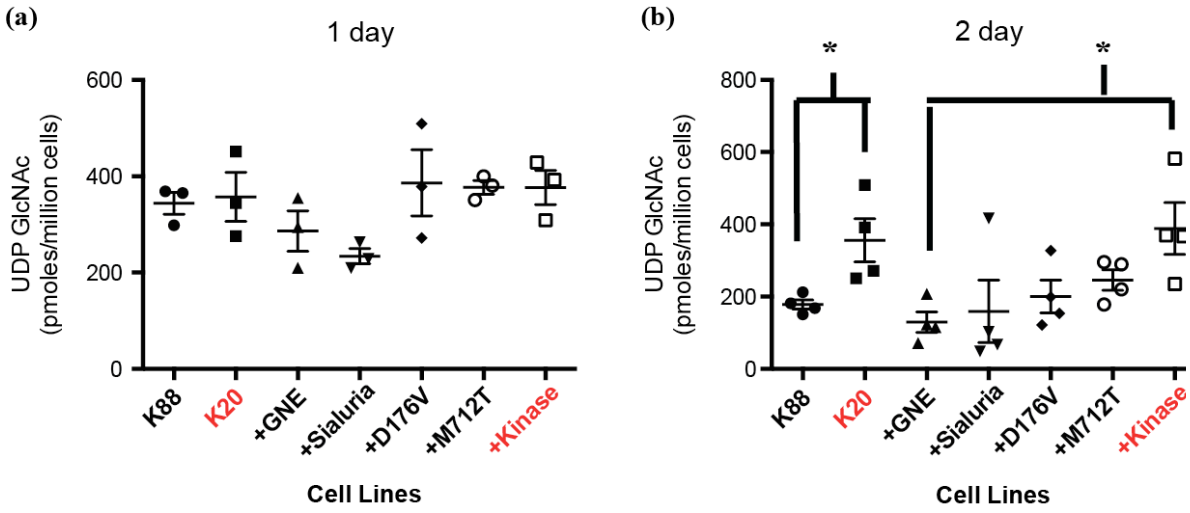

Figure 2.4. Sialic acid production influences UDP-GlcNAc levels. Cells were grown in normal media and harvested at the (a) 1 day and (b) two days. Intracellular metabolites were harvested and examined by HPAEC-UV. Cell names in red do not make sialic acid.

Alterations in intracellular UDP-GlcNAc lead to changes in intracellular O-glycosylation (O-GlcNAc). Furthermore, O-glycosylation levels increase under cell stress (63). For these reasons, I examined the levels of O-glycosylation in different cellular compartments (Figure 2.5). Interestingly, I did not find any changes in O-glycosylation.

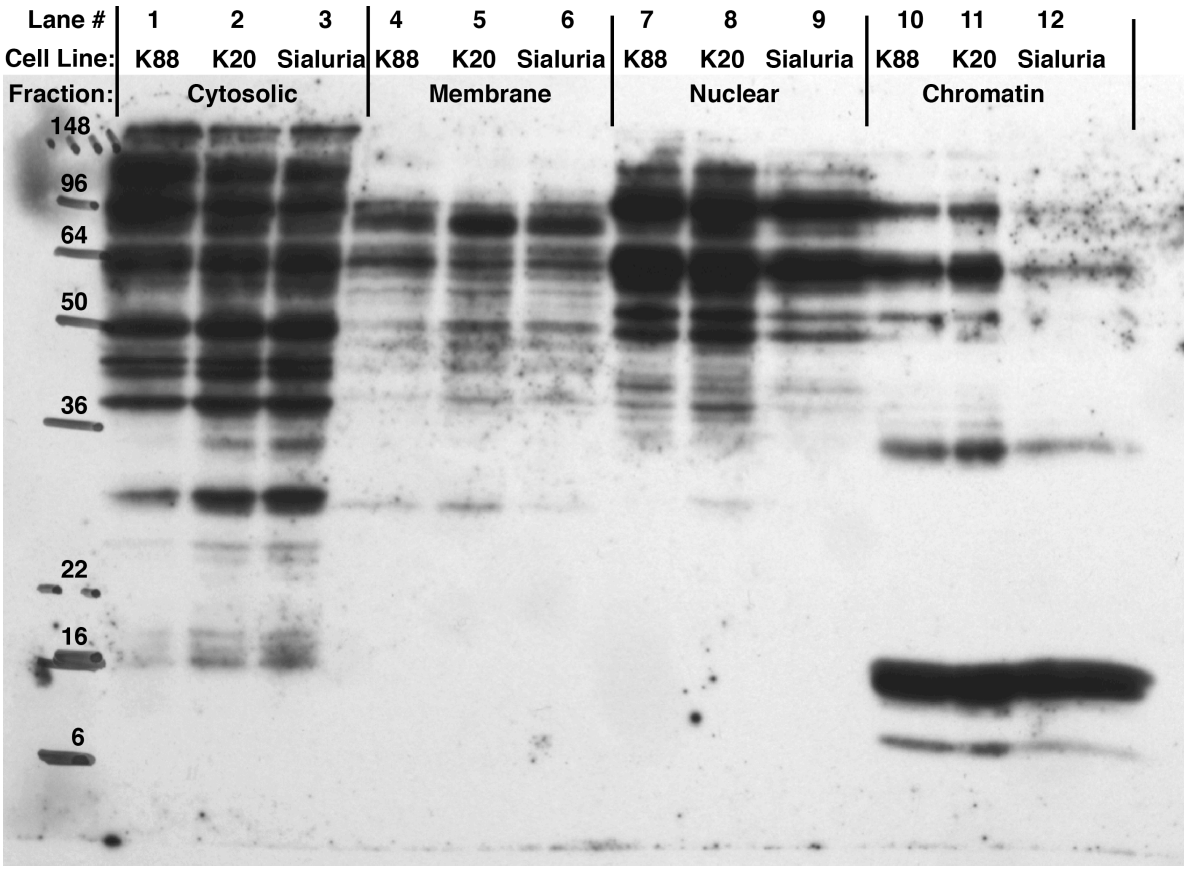

Figure 2.5. O-GlcNAc blot reveals no global changes between cell lines. Cells were fractionated and proteins were probed for O-GlcNAc. K88 and Sialuria cells make sialic acid, K20 cells do not.

While this was surprising at first, changes in O-glycosylation secondary to changes in UDP-GlcNAc or cell stress are not permanent or stable, and are generally only observed for short time points (< 24hrs) (Mike Boyce and John Hanover, personal communication).

The hexosamine biosynthetic pathway is a critically important metabolic pathway, and it is thought that 5% of glucose is routed to this pathway (64). Given the central importance of the hexosamine pathway, I decided to take a metabolomics (65) approach and evaluate if there were any differences in intracellular metabolites in cells that can make sialic

acid versus cells that cannot using liquid chromatography and tandem mass spectrometry. I collaborated with Sunil Laxman in this endeavor, a postdoctoral fellow in Ben Tu's lab. After harvesting intracellular metabolites, I analyzed metabolites using three solvent systems (formic acid: positive mode, TBA: negative mode, ammonium acetate: positive mode) and analyzed the levels of >100 metabolites. I did not see any significant changes between the cell lines. (It should be noted however that the hexosamine metabolites GlcN-6-P, GlcNAc-6-P, GlcNAc-1P, and UDP-GlcNAc couldn't be detected by this system). I was initially disappointed and surprised by these results, but upon further reflection I realized that metabolic pathways are complex, and it is possible that other metabolic pathways buffered any stress imposed by sialic acid synthesis.

Sialic acid production reduces N-glycan branching

It has been reported that the degree of N-glycan branching is ultrasensitive to the intracellular concentration of UDP-GlcNAc (66). In other words, small changes in the UDP-GlcNAc levels result in large differences in N-glycoforms. Furthermore, this has significant signaling implications, as the N-glycan branching can affect the cell surface retention of proteins (66-69). I thus hypothesized that N-glycan branching levels would be altered in cells that produce sialic acid. Indeed, L-PHA lectin flow cytometry demonstrates that there is less branching in cells that produce sialic acid (Figure 2.6). To ensure that sialylation does not affect L-PHA binding, cell surface sialic acid was first removed by treating cells with sialidase, and cells that do not make sialic acid still had higher levels of N-glycan branching than cells that do make sialic acid (Figure 2.7). Taken together, the results of the L-PHA flow

cytometry experiments appear to reflect a difference in N-glycosylation and not in sialylation. Surprisingly, this difference in N-glycan branching appears to be true even after 1 day of passage, a time when UDP-GlcNAc levels are similar between different cell lines. Thus, sialic acid synthesis places a strain on available UDP-GlcNAc even if total levels appear unaltered. Alternatively, it is possible that levels of UDP-GlcNAc may be different after 1 day, but this difference cannot be detected with my method. Regardless, this strain on the hexosamine pathway is further evidenced by alterations in poly-LacNAc formation. UDP-GlcNAc is important for polyLacNAc formation and extension that occurs on branched glycans. Cells that do not make sialic acid had less cell surface LacNAc as measured by LEL lectin flow cytometry analysis. I have submitted samples to Anne Dell and Stuart Haslam at Imperial College of London to confirm these findings by glycan mass spectrometry.

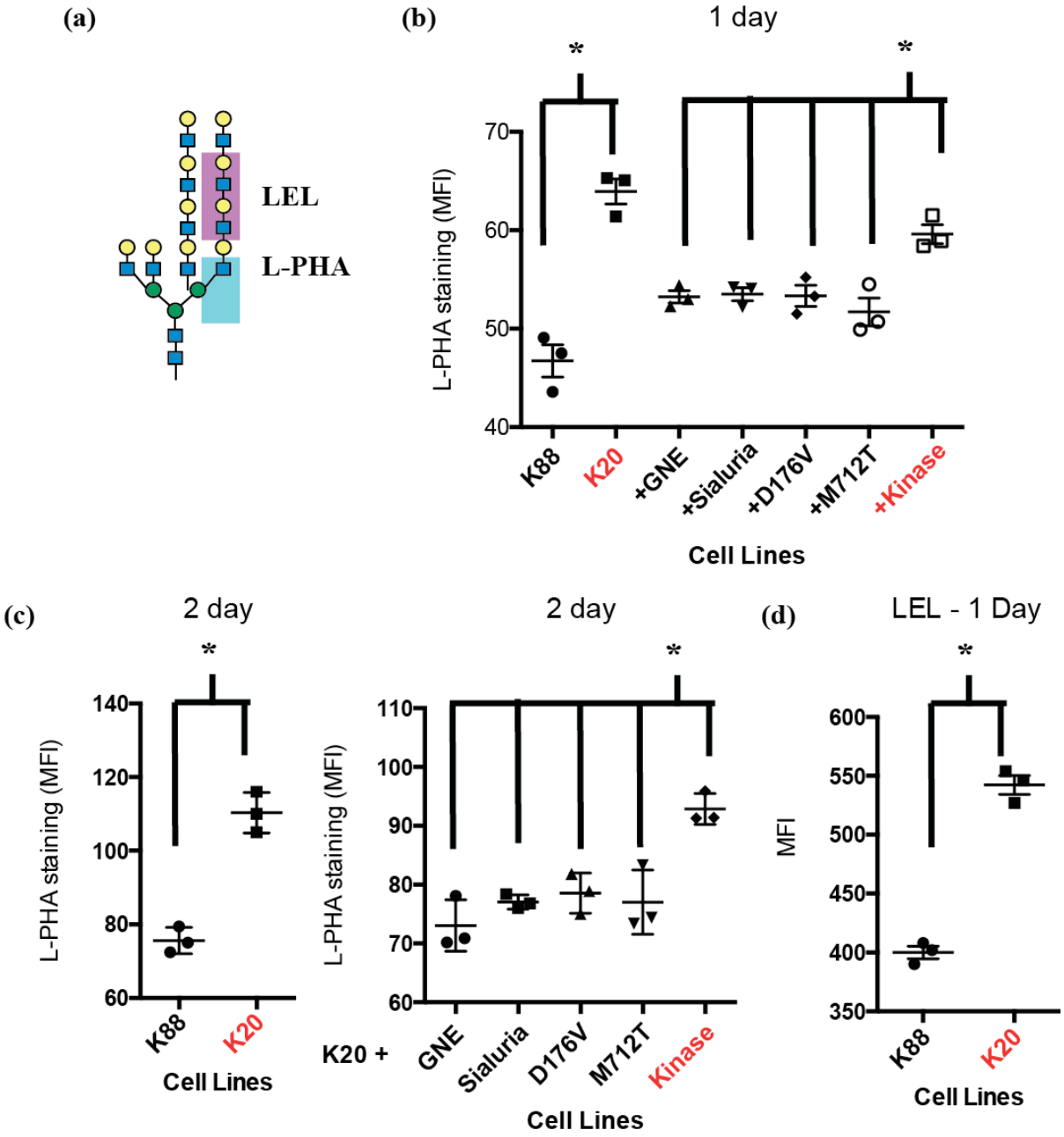

Figure 2.6. Cells that do not make sialic acid have increased branching levels and polyLacNAc extension. (a) L-PHA lectin and LEL lectin bind to the tetrantennary branch and polyLacNAc respectively. (b) 1 day and (c) 2 days after passage cells that do not make sialic acid have higher levels of L-PHA staining by flow cytometry. (d) K20 cells have higher levels of LEL lectin staining when compared to K88 cells. Cell names in red do not make sialic acid.

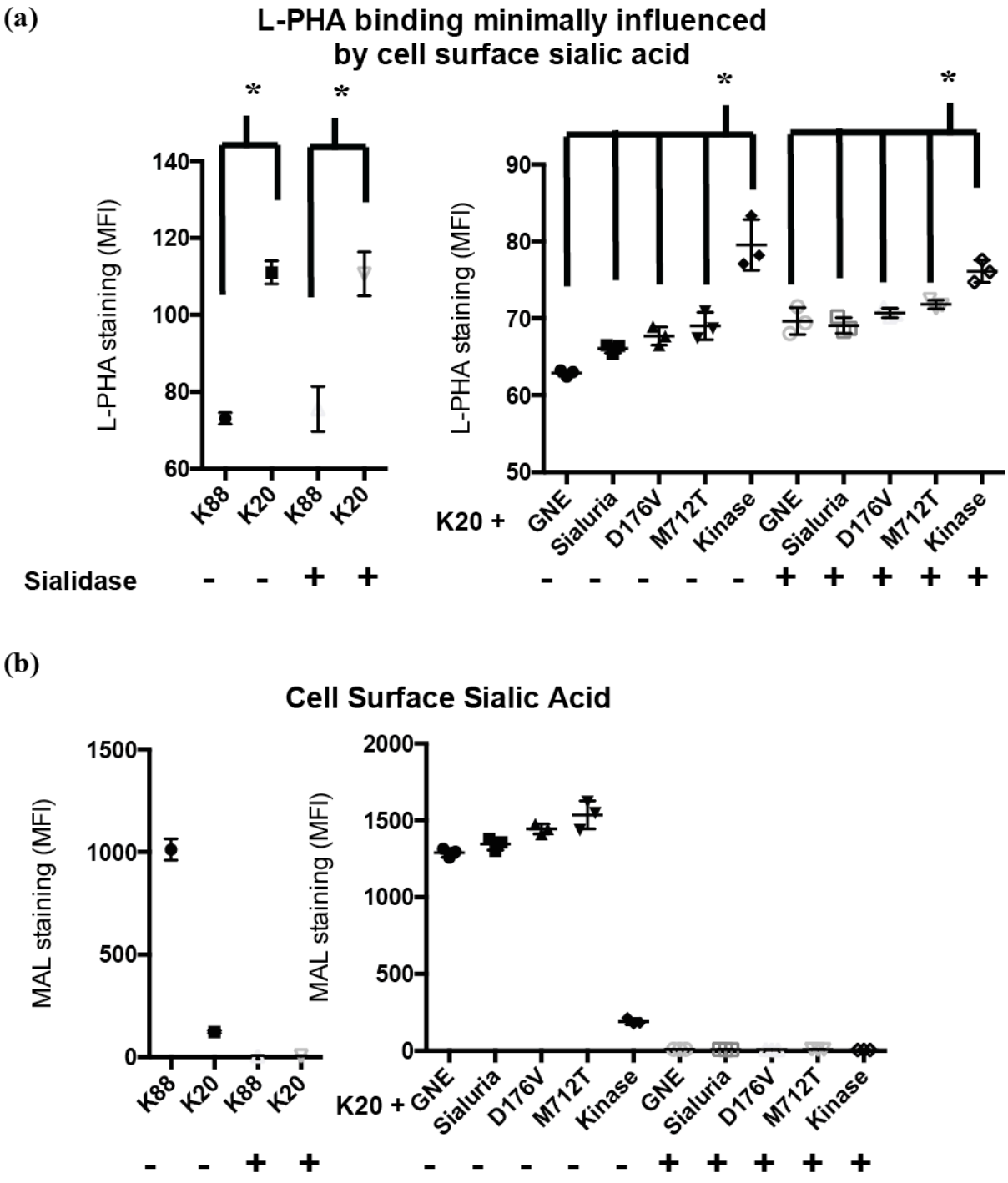

Figure 2.7. Sialic acid does not influence L-PHA binding. After 1 day of passage, cells were either treated with or without sialidase and (a) L-PHA staining was measured by flow-cytometry and activity of sialidase was confirmed by (b) MALII lectin.

Altered N-glycosylation impairs galectin-1 binding

Theses differences in the N-glycome can have widespread impact because N-glycosylation is one of the most common post-translational modifications, and branched and extended N-glycans are ligands for a family of extracellular proteins called galectins (70). Galectins are a family of proteins that form an extracellular lattice (71). Through their ability to bind to LacNAc found on N-glycans, they can affect the retention of proteins on the cell surface and affect the clustering of proteins (72). Indeed, galectin-1 mediated clustering of T-cells results in apoptosis (73). I performed an in-cell ELISA assay adapted from Sandra Schmid's lab (UT Southwestern) and determined that there was more galectin-1 bound to cells that could not make sialic acid (and thus had higher levels of UDP-GlcNAc and higher

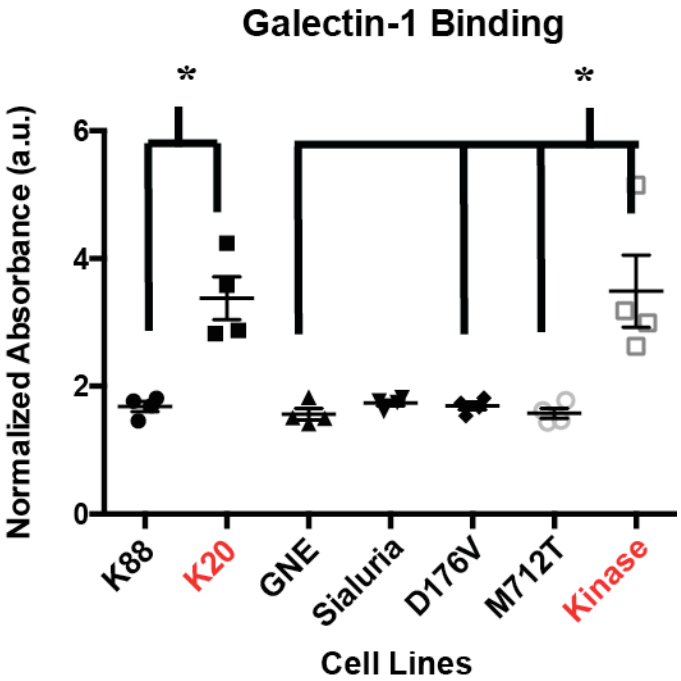

Figure 2.8. Cells that do not make their own sialic acid bind better to galectin-1. Cells wer cultured in media for 1 day and binding to biotinylated galectin-1 was assayed by an in-cell ELISA protocol adapted from Sandra Schmids lab. Cell names in red do not make sialic acid.

levels of branching) (Figure 2.8). While galectin-1 binding can be influenced by α 2,6 sialic acid, α 2,6 sialic acid levels among these cell lines are similar, and thus does not influence these results (74). Although α 2,3 sialic acid is different among these cell lines, galectin-1 is not sensitive to this linkage (75). Thus, alterations in sialic acid synthesis affect the HBP and influences cell surface interactions through changes in the N-glycan branching.

Increased sialylation reduces N-glycan branching

N-glycan branching is a process governed by ultrasensitivity – that is small changes in UDP-GlcNAc levels leads to large changes in N-glycan structure. Indeed, it has been established that N-glycan branching is increased following an increase in intracellular glucose (which feeds into the hexosamine pathway to increase UDP-GlcNAc levels), which can result in altered retention of cell surface proteins such as the glucose transporters (66,67,76). Thus mechanisms linking glucose to GlcNAc to branching has been delineated, but the role of sialylation in mediating this process has not been studied. However, supplementing cells with 40 mM GlcNAc (69) did not cause branching to be increased (Figure 2.9b), suggesting that in addition to its effect on the hexosamine pathway, increased sialic acid itself may block branching pathways. To test this hypothesis, I supplemented cells with Ac_4ManNAc, a compound that rapidly crosses the membrane and metabolizes to sialic acid. As expected, cell surface sialic acid increased as measured by MALII lectin and flow cytometry (Figure 2.9a). Surprisingly, there was less L-PHA binding, indicating that sialylation itself can inhibit N-glycan branching. It should be noted that K20 cells supplemented with Ac_4ManNAc had higher L-PHA staining than K88 cells supplemented

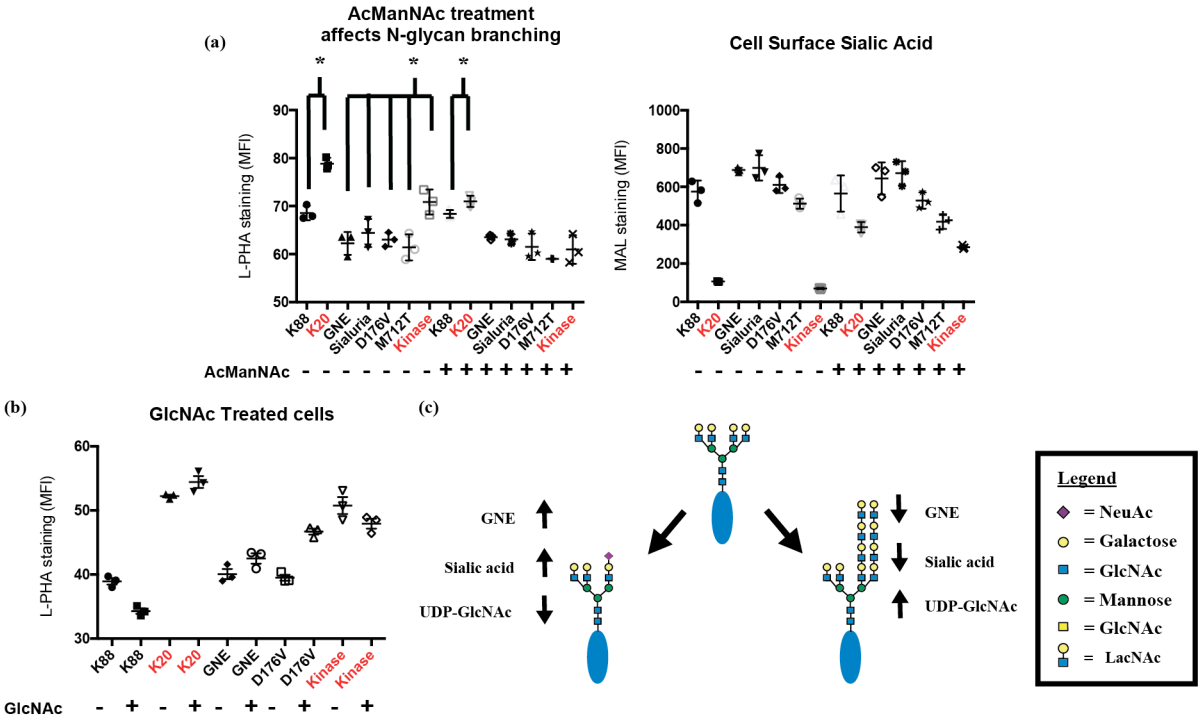

Figure 2.9. Sialylation influences N-glycan structure. (a) Cells were supplemented with Ac_4ManNAc and cell surface sialic acid was increased (right) and L-PHA binding and thus N-glycan branching was abrogated. (b) GlcNAc treatment cannot increase branching levels. (c) Model for how sialic acid metabolism influences N-glycan structure through altering sialyliation and influencing the HBP. Cell names in red do not make sialic acid.

with Ac_4ManNAc. Taken together altered sialic acid metabolism can have a two-pronged effect on N-glycan structure: 1) increased sialic acid production causes fluctuations in UDP-GlcNAc that are sensed by branching enzymes, causing decreased N-glycan branching, and 2) sialylation of N-glycan structures prevents branching from occurring (Figure 2.9c).

I began these experiments with the purpose of discovering pathways that could be disrupted in cells containing hIBM GNE mutants. While I did not discover any differences in cell lines containing hIBM mutants versus cell lines that contained wild-type GNE, I did discover an unappreciated role of sialic acid in regulating the hexosamine pathway and N-

glycan structure. Taken together, these experiments reveal a new layer of regulation, as well as the impacts of sialic acid synthesis, on both a core metabolic pathway and a ubiquitous post-translational modification.

DISCUSSION

Disruptions in sialic acid biology are implicated in hIBM, a disease of aging. Despite the number of mutations in GNE that are linked to the disease, mechanistic understanding of hIBM pathogenesis is lacking. Early studies focused primarily on alterations in cell surface sialic acid, while the effect that GNE mutants might have on intracellular metabolism or glycan structures have been underexplored. Here I use a cell line that does not express wild-type GNE to demonstrate that GNE activity alters the hexosamine pathway and subsequently alters N-glycan branching. Furthermore, I demonstrate that sialylation of N-glycans alters N-glycan processing.

The hexosamine pathway, glycan structure, and aging

There is a growing relationship between metabolism and aging. In previous studies, increased hexosamine flux increased longevity (9,55). In contrast, the results of my studies suggest that UDP-GlcNAc levels are increased in hIBM mutants and this should improve survival. One possible explanation for this discrepancy is that the early increase in UDP-GlcNAc is not detrimental and thus hIBM patients are normal at birth but this energetic excess is detrimental later in life. Indeed, dietary restriction has long been associated with

increased longevity (77,78). Perhaps hIBM mutants activate aging mechanisms that occur during dietary excess.

Furthermore, my finding that sialic acid production can influence the structure of cell surface glycans and alter galectin-1 binding is consistent with the idea that changes in N-glycan interactions are associated with aging. Indeed, N-glycan structures can alter growth factor signaling, and aged cells have been demonstrated to have lower levels of EFGR (79). Consistently, mice with out the branching enzyme, Mgat5, had fewer muscle stem cells (satellite cells) (57). Furthermore, an enzyme involved in N-glycan remodeling, Klotho, is also implicated in aging (80). Klotho was identified in mice that had multiple phenotypes resembling aging – shortened life span, infertility, skin atrophy and, interestingly, muscle atrophy (81). Moreover, Cha *et al* demonstrated that Klotho putatively acts by removing cell surface sialic acid, thereby exposing LacNAc branches on proteins such as the Ca^{2+} channel TRPV (82). These LacNAc branches then interact with galectin-1, increasing the retention of TRPV on the cell surface. In this context, I propose a mechanism where hIBM mutants accelerate the aging process in muscle through alterations in the hexosamine biosynthetic pathway and they may function like Klotho by simultaneously reducing cell surface sialic acid and increasing N-glycan branching and polyLacNAc extension. Furthermore, this proposed mechanism is consistent with findings that hIBM mutants subtly alter GNE activity. I propose that subtle decreases in sialic acid flux in hIBM mutant cells could result in less sialylation and increased N-glycan branching. Synergistically, there is increased UDP-GlcNAc which further drives N-glycan branching.

Generality of the sialic acid – hexosamine – N-glycan pathway

The results outlined in this chapter was observed in one cell line, a B-cell lymphoma, and only observed in the extreme case where cells could not produce their own sialic acid. However, due to the lack of expression of GNE in K20 cells, they were the optimal cell line to perform my initial experiments. Nevertheless, obtaining more cell lines to demonstrate this phenomenon would be informative. To this end, new genome editing tools like Zinc Finger nucleases and the CRISPR/CAS system can allow for gene disruption in potentially any desired cell line (83). I attempted to perform CRISPR/CAS mediated knockout of GNE in HeLa cells, but was initially unsuccessful. For cell lines to be useful, both alleles of GNE would have to be disabled. The probability of this occurring is somewhat low and many clones would have to be screened. I attempted to perform a selection experiment using lectins to select for a population of CRISPR/CAS GNE knockout cells by sorting for cells with low cell surface sialic acid staining. Unfortunately, sorted cells when re-analyzed, had no evidence of GNE knockout. More time would be necessary than available to screen for colonies with GNE knockout following CRISPR/CAS targeting.

Furthermore, while K20 cells have less cell surface sialic acid than K88 cells, both cell lines still have significant amount of sialic acids because they acquire it from the media. Typical RPMI media with 10% FBS contains a ~200 uM concentration of sialic acid, presumably mostly bound to glycoproteins. I propose that sialic acid from media was able to mask differences in gene expression, signaling, and metabolic differences between cells that make sialic acid and cells that don't. I attempted performing experiments under serum free

media where there is no sialic acid in the media. While this is effective in lowering cell surface and intracellular sialic acid, experiments were difficult to perform consistently due to extremely low cell viability. I also attempted some preliminary experiments under low serum (2.5% FBS), and the cells grew well and were viable. While the cells did have lower levels of α 2,3 linked sialic acid (by MALII staining), they had similar levels of α 2,6 linked sialic acid (by SNA staining) and I decided not to pursue it further. The development of normal serum but completely lacking sialic acid would be very useful in studying the effects of hIBM mutants on intracellular metabolism and signaling.

Another possible explanation for observing few differences outside of the HBP pathway could be due to the use of stable cells. When I started the project, I was adamant about using stable cell lines because I wanted the observations not to be dependent on sudden and dramatic cellular changes (i.e. through transfection). However, the downside of using stable cell lines is that cells can utilize homeostatic mechanisms to buffer out any consequences of increased GNE expression and sialic acid production. It would be interesting to utilize transient transduction (with lentivirus) and see if the sudden onset of sialic acid production causes changes in gene expression, metabolism, and O-glycosylation.

Implications for sporadic Inclusion Body Myositis (sIBM)

Sporadic inclusion body myositis (sIBM) is the most common acquired muscle disease associated with aging (84-87). Like hIBM, sIBM is a muscle wasting disease of aging, but it typically afflicts individuals older than 50 years of age. The disease is relentlessly progressive and there is currently no effective treatment. Intriguingly, the

pathology of sIBM is thought to be similar to hIBM in that muscle weakness is associated with the presence of protein inclusion bodies (25). Thus principles underlying hIBM progression may underlie sIBM pathogenesis. It would be interesting to evaluate the N-glycan structure of sIBM patients, and possibly consider providing GlcNAc therapeutically.

ACKNOWLEDGEMENTS

I would like to thank Linda Baum for generously providing biotinylated galectin-1. I would also like to thank Stuart Haslam and Anne Dell for helping with the glycan mass spectrometry. I would like to thank Yibing and Andrea for helping with early experiments utilizing K20 cells growing in serum free media. I especially would like to think Sunil Laxman and Ben Tu for being generous with their LC-MS/MS machine. Sunil particularly was extremely helpful in designing the mass spectrometry experiment. I would also like to thank the National Institutes of Aging for awarding me a fellowship (F30AG040909) to pursue this work.

METHODS

Cell lines and culturing conditions

BJAB K20 cells (58)(obtained from Michael Pawlita (German Cancer Research Center) and James Paulson (The Scripps Research Institute)) and Jurkat cells were cultured in RPMI 1640 media containing 2 mM glutamine and supplemented with 10% fetal calf serum, 100 U/mL penicillin, and 100 μg/mL streptomycin at 37 °C, 5% CO_2 in a water-

saturated environment. Ac_4ManNAc was synthesized as previously described (40). GlcNAc (Sigma).

Cloning of GNE and mutagenesis

GNE was cloned from human brain cDNA (Origene CH-1001) using the primers:

GNE-Forward: 5'- AAAGCTAGC ATG GAG AAG GGA AAT AAC C-3'

GNE-Reverse: 5'- TTT CTCGAG CTAGTAGATCCTGCGTGTTGTG-3'

Pcr product was Blunt TOPO cloned (Invitrogen) into pCR4 Blunt-TOPO vector. Sequencing revealed that the sequence matched accession NM_005476.4. To prepare the gene for insertion into the lentiviral plasmid, I switched the restriction sites by performing PCR using the primers:

GNE-F-AgeI: 5'- AAA ACCGGT ATG GAG AAG AAT GGA AAT AAC C-3'

GNE-R-SbfI: 5'- TTT CCTGCAGG CTA GTA GAT CCT GCG TGT TGT G-3'

The PCR product was TOPO cloned into another pCR4 Blunt-TOPO vector. Quickchange mutagenesis was performed in this pCR4 Blunt-TOPO GNE vector with the following primers:

GNE-R266Q-F: 5'- gatggttcgagtgatgcagaagaagggcattgagc -3'

GNE-R266Q-R: 5'- gctcaatgcccttcttctgcatcactcgaaccatc -3'

GNE-D176V-F: 5'- catgtgtgaggaccatgttcgcatccttttggcag -3'

GNE-D176V-R: 5'- ctgccaaaaggatgcgaacatggtcctcacacatg -3'

GNE-M712T-F: 5'- ctgggtgctgccagcacggttctggactac -3'

GNE-M712T-R: 5'- gtagtccagaaccgtgctggcagcacccag -3'

Furthermore, the kinase only construct was obtained using the following primer:

AgeI-Kinase-start: 5'- AAAACCGGTATGACTCTAAGTGCCTTGGCCGTTG-3'

The reverse primer from above was used. All constructs were validated by sequencing and there were no secondary mutations present in any construct. pCR4 Blunt-TOPO Plasmids then were then cut using AgeI and SbfI and ligated into pRRL CAGpNFLAG BAF155 IRES GFP (Addgene, 24561) cut with AgeI and PstI. All restriction enzymes were purchased from NEB. Lentivirus was produced as described below.

Production of lentivirus and infection of K20 cells to express GNE constructs

Virus was produced utilizing the third generation packaging system (88). Briefly, HEK-293T cells were transfected with a pRRL GNE IRES GFP plasmid (WT or mutants) accompanied with pRRE (12251, Addgene), pRSV-REV (12253, Addgene) and pMD2.G (C12259, Addgene) in the presence of Fugene 6 (E2691, Promega) to generate lentivirus. Media was replaced after 20 h. After 2 days, supernatant containing lentivirus was harvested and filtered through a .45 uM PVDF membrane. K20 cells (~200k) aliquots were incubated with 1:1, 1:5, 1:10 lentivirus stock solution diluted with RPMI media and supplemented with 4 μg/mL polybrene (AL-118, Sigma-Aldrich) to enhance infection efficiency. Cells were spun 500 g for 2hrs. Cells were resuspended in fresh media and placed in a 6-well dish. After 48 hrs, successful infection was determined by performing flow cytometry and looking for GFP fluorescences. To achieve homogeneity, cells underwent two rounds of cell sorting

with either an Aria or MoFlo cell sorter. Viral dilution that caused less than 20% of cells to GFP+ was utilized for the first round of sorting to increase the probability that cells were infected once. After two rounds of sorting in this manner, all cell lines were greater than 95% GFP+. Even without selection, GFP expression was extremely stable.

Gene expression array

Cells were cultured for 2 days and. ~1.6 million cells were harvested and washed twice with dPBS. RNA was extracted using the Aurum RNA extraction kit (Bio-Rad). 1 ug of each line was submitted in duplicate the UTSW microarray core. They validated the quality of RNA using an Agilent 2100 Bioanalyzer. Samples were then loaded onto a Illumina HumanHT-12 v4 BeadChip. Data was analyzed using BeadStudio and GeneSpring.

Flow cytometry analysis

Reagents used and concentrations consisted of SNA-biotin (Vector labs, 1:100), MAL-II (Vector labs, 1:100), allophycocyanin-streptavidin (APC-Strep) (Life Sciences, S-868), 1:200), LEL-biotin (Vector labs, 1:1000), L-PHA rhodamine (Vector labs, 1:100).

K20 cells were cultured for the indicated time points with or without supplementation collected and washed twice with dPBS, then resuspended in dPBS at 2.0×10^6 cells/mL. Then 200 µl of cell suspension was transferred to V-bottom 96-well plate and centrifuged once more. The cell pellets were incubated with 100 µl for 60 min at 4 °C (LEL lectin incubation was 30 min), then washed with D-PBS three times. The cells were then incubated with 1:200 allophycocyanin-streptavidin(Life Sciences, S-868) for 45 minutes. Fluorescence

was analyzed by flow cytometry on a FACSCalibur instrument (BD Biosciences) equipped with dual lasers at 488-nm and 635-nm. For sialidase experiments, before lectin staining, cells were incubated with Sialidase A (Glykotech) for 90 minutes in dPBS + .1% BSA at 37° C.

Cellular fractionation and DMB-derivitzation of sialic acids

Freshly harvested cells were fractionated as described previously (89). Briefly, cells were counted and harvested by centrifugation and then suspended in hypotonic lysis buffer containing protease inhibitors for 15 min on ice. Cells were lysed by extrusion through a 25 gauge needle. Nuclei and unbroken cells were removed from the post-nuclear supernatant by two rounds of centrifugation at 1000*g* for 15 min at 4 °C. Next, the post-nuclear supernatant was transferred to heavy-walled polycarbonate tubes and centrifuged at 100,000*g* for 1 h in a Beckman TLA 120.2 rotor. The supernatant was designated the cytosolic fraction (C). The pellet was washed twice with 400 μL cold hypotonic lysis buffer followed by centrifugation at 100,000*g* for 1 h after each wash. The remaining pellet was designated the membrane fraction (M). Samples were flash-frozen and the solvent removed by vacuum overnight.

Quantification of sialic acids

Quantification was performed relative to known standards, prepared as follows. Calibration curves were prepared by injecting between 50 to 750 fmol of DMB-Neu5Ac and DMB-Neu5Gc on a Dionex Acclaim® Polar Advantage C16 5μm, 4.6 x 250 mm column attached to a Dionex Ultimate 3000 HPLC with fluorescence detector. Separation was

performed using a gradient of 2-90% acetonitrile (ddH_2O) with fluorescence detection (ex. 373 nm, em. 448 nm). Linearity of the fluorescence signal of DMB-Neu5Ac confirmed in the range of 50 to 750 fmol.

To release sialic acids, 2.0 M acetic acid was added to each of the membrane-bound (50 uL) and dried cytosolic samples (100 uL). Solutions were incubated at 80 °C for 2 h. Samples were cooled to room temperature, then DMB reaction solution (1.4 M acetic acid, .75 M 2-mercaptoethanol, 18mM $Na_2S_2O_3$) was added to each sample (40 uL for membrane fraction and 80 uL for cytosolic fraction). After a 2 h incubation at 50° C, 0.2 M NaOH was added to each sample (10 uL for membrane fraction and 20 uL for cytosolic fraction). Samples were filtered through 10 kDa MWCO filters by centrifugation and the resulting flow-through was stored at −20 °C in the dark until analysis. Generally, 2 μL of the derivatized material was diluted with 98 μL ddH_2O and analyzed by fluorescence HPLC, as described for the standards. Linear regression analysis using the DMB-NeuAc standard curves was used to calculate the amount of DMB-Neu5Ac present in experimental samples.

Cellular fractionation for western blot

Cells were harvested after 2 days of culture and fractionated using the Cell Fractionation kit (PIERCE).

In-Cell Elisa assay

Cells were grown in media for the indicated amount of time and 400k cells in a volume of 200 uL were harvested and placed 96-well v-bottom plate. Cells were chilled on

ice for approximately 15 minutes and then cells were washed 3x with dPBS. Cells were then incubated with 3 uM galectin-1 (a kind gift from Linda Baum) in dPBS for 30 min on ice. Cells were spun and cells were washed with cold dPBS three times. 100 uL (half the cells) were transferred to another 96-well v-bottom plate. The protein content of these cells was analyzed by the BCA assay (PIERCE) and the absorbance was measured at 562 nm. The other half of cells were incubated with Streptavidin-POD/1%BSA/PBS for 1 hr at 4° C. Cells were washed 3x dPBS. Cells incubated with 100 uL of OPD solution (20 mL HRP buffer (.05 M phosphate-citrate buffer pH 5.0), 20 mg OPD tablet (Sigma), and 20 uL of 30% H_2O_2. The reaction was then quenched with 50 uL of 5 M H_2SO_4. A Biotek Synergy H1 Hybrid Reader acquired the absorbance at 490 nm and 650 nm. Samples were normalized in excel via the formula (A490 – A650) / A562.

Harvesting cells for glycan mass spec

40 million cells were harvested and frozen. Their membrane proteins were isolated with the following protocol. The frozen cell pellets are suspended in 2 mL of lysis buffer containing 50 mM Tris-HCl (pH 7.4), 0.1 M NaCl, 1 mM EDTA and protease inhibitor cocktail (Roche Diagnostics) and kept on ice for 20 min and then homogenized using a polytron homogenizer (15 sec, 7 times on ice bath). The homogenized cells are centrifuged at 2000g for 20 min at 4° C to precipitate nuclei. The supernatant is diluted with 2 mL of Tris-buffer (50 mM Tris-HCl (pH 7.4), 0.1 M NaCl) and then membranes pelleted by ultracentrifugation at 120000g for 80 min at 4° C. The supernatant is discarded, and the membrane pellet is suspended in 100 μL Tris-buffer. After adding 400 μL Tris-buffer

containing 1% (v/v) Triton X-114, the suspended mixture is homogenized by pipetting vigorously. The homogenate is chilled on ice for 10 min and incubated at 37° C for 20 min and then phase partitioned by centrifugation at 2000g for 2 min. The upper aqueous phase is removed. The lower detergent phase is further mixed with 1 mL of ice-cold acetone and kept at -20° C overnight to precipitate proteins and remove any detergent. After centrifugation at 1940g for 2 min, the precipitated cell membrane proteins are stored at -20° C.

Intracellular metabolite harvesting

Cells were grown various time points as indicated and cells were harvested and counted. Typically ~2 x 10^6 cells were harvested. Cells were spun, and washed with cold dPBS twice and flash frozen. Cells were then lysed with 80% "super-cold" methanol (on dry ice) (90). Lysate was spun at 2,000g for 15 min at 4° C. The supernatant was flash frozen and dried by a speed vacuum for 4-5 hours. The intracellular metabolite pellet was typically used immediately or it was stored at -80° C for future use.

Analysis of intracellular metabolites by HPAEC

Metabolite pellet was resuspended in 40 mM sodium phosphate buffer (pH 7.4 40 μL per million cells), and filtered through an Amicon® Ultra centrifugal filter unit (Millipore, 10,000 MWCO). Filtrates were analyzed by HPAEC (ICS-3000 system, Dionex) with CarboPac™PA1 (Dionex) with a pulsed amperometry detector (PAD) and UV-detector in-line (62,91). Typically, 20 μL of metabolite was injected into the sample loading loop before the sample enters a guard column (Dionex, 4 × 50 mm) and then an analytical column

(Dionex, 4 × 250 mm). The eluents used were 1.0 mM NaOH (C) and 1.0 M NaOAc and 1.0 mM NaOH (D). Low-carbonate NaOH (50% in water) was obtained from Fisher Scientific (SS254-1) and NaOAc was from (Sigma, 71183). HPAEC was run with a flow rate = 1 mL/min and the following gradient elution was performed: $T_{0\ min}$ = 95% C, T_5 = 85% C, T_{15} = 70% C, T_{20} = 60% C, T_{45} = 60% C, T_{50} = 0% C, T_{60} = 0% E2, T_{65} = 95%, T_{75} = 95%. UDP-GlcNAc standards (50 uM, 25 uM, 10 uM, and 2.5 uM) were injected at the same time as cellular samples. UDP-GlcNAc peak areas were input into excel and raw data were converted to pmoles of UDP-GlcNAc by comparing to a standard curve generated by analyzing the peak areas of the UDP-GlcNAc standards. Data was normalized to cell number.

Analysis of intracellular metabolites by LC-MS/MS

Metabolites were extracted and measured using targeted LC-MS/MS methods, expanding and modifying methods described previously (65,92). Briefly, a library of common metabolites was constructed. For each metabolite, a 1mM standard solution was infused into a Applied Biosystems 3200 QTRAP triple quadrupole-linear ion trap mass spectrometer for quantitative optimization detection of daughter ions upon collision-induced fragmentation of the parent ion using multiple reaction monitoring (MRM), using positive mode or negative mode analysis. For positive mode analysis (usually MW+1), the infusion solvent contained either 50% methanol/0.1% formic acid or 50% methanol/5 mM NH_4OAc, and for negative mode (usually MW - 1), 50% methanol/50% 5mM Tributyl amine. For each metabolite, the optimized parameters for quantitation of the two most abundant daughter ions

(i.e., two MRMs per metabolite) were selected for inclusion in further method development. For metabolites not easily detected in positive mode, a negative mode method was optimized. Metabolites were separated chromatographically on a C18-based column with polar embedded groups (Synergi Fusion, 150 ´2.0 mm 4 m, Phenomenex), using a Shimadzu Prominence LC20/SIL-20AC HPLC-autosampler coupled to the mass spectrometer. Flow rate was 0.5 ml/min. The buffers used for positive mode analysis were: Buffer A: 99.9% H_2O/0.1% formic acid, Buffer B: 99.9% methanol /0.1% formic acid. T = 0 min, 0% B; T = 4 min, 0% B; T = 11 min, 50% B; T = 13 min, 100% B; T = 15 min, 100% B, T = 16 min, 0% B; T = 20 min, stop, or Buffer A: 5mM Ammonium Acetate in H_2O, Buffer B: 5mM Ammonium Acetate in 100% methanol. For those metabolites that were more optimally detected by using tributyl amine (TBA, negative mode), a similar gradient was used with buffer A being replaced by 5mM TBA, and 100% methanol used as buffer B. These conditions enabled highly quantitative measurements of metabolites, as assessed using standards.

For running samples, extracts were dried under vacuum and then resuspended in 150 uL 0.1% formic acid or 5 mM TBA or 5 mM ammonium acetate for injection (typically 90 uL injection volume). Solution was allowed to sit on ice for 30 minutes and spun at 16,000g for 5 minutes at 4° C to pellet any precipitation. The supernatant was filtered through .2 um PVDF micro spin filter (Grace Davison Discovery Science, 8604742) at 9000 g for 5 min at 4° C twice. Mass spectrometer settings were: curtain gas = 45; collision gas = medium; IonSpray voltage = 5000; temperature = 600; ion source gas 1 = 70; ion source gas 2 = 55; interface heater = on. Many MRMs (up to 150, each targeting a parent/daughter ion

pair) were conducted simultaneously (20 ms per MRM) throughout each chromatographic run. Typically, both MRMs for a given metabolite displayed high correlation. The retention time for each MRM peak was matched with the metabolite standard. The area under each peak was then quantitated by using Analyst software, with a manual inspection for accuracy, and normalized against total ion count.

CHAPTER THREE
Cellular Metabolism of Unnatural Sialic Acid Precursors and Analogs – the Method Behind the Madness

INTRODUCTION

Sialic acids are typically the terminal sugars on glycans, and this outermost positioning on the cell surface places it in an ideal position to modulate cellular biology. Indeed, sialic acid has been implicated in a variety of physiological processes such as neural development and leukocyte homing, as well as pathological processes such as viral infections and cancer development (93). However, detailed mechanistic understanding of how sialic acid mediates these processes is lacking.

Traditional biochemical techniques have incrementally advanced the field of glycobiology, but are significantly augmented by the use of metabolic carbohydrate engineering (94). Metabolic carbohydrate engineering refers to the usage of carbohydrate analogs that contain bioorthogonal functional groups – that is functional groups that are tolerated by endogenous machinery but are not normally found in nature (34). These analogs take advantage of the intrinsic promiscuity of endogenous enzymes to process and present the analog onto glycoproteins and glycolipids (35). To date, a large variety of sialic acid analogs that harbor a variety of functional groups have been developed and utilized to study biological systems. Indeed, sialic acid analogues containing a photo-crosslinking group have illuminated sialic acid-protein interactions that would be difficult to study utilizing traditional biochemical techniques (95). For instance, the photo-crosslinking sialic acid analogue 9-

Aryl Azide NeuAc (9AAz NeuAc) has elucidated the binding partners of CD22, an important regulator of B-cell activity (42,43).

The utility of metabolic sialic acid engineering is increased if higher intracellular and cell surface levels of the sialic acid analogue can be achieved (96). For instance, high levels of a photo-crosslinking sialic acid such as 9AAz NeuAc or SiaDAz on a glycoprotein of interest would increase the likelihood of crosslinking a binding partner and provide more material for biochemical evaluation (i.e. immunoblot or mass spectrometry analysis). While 9AAz NeuAc must be supplied as the sialic acid analog, SiaDAz, a sialic acid analogue containing an N-acyl diazirine photo-crosslinking group, has the advantage that it can be generated *in cellulo* from its precursor ManNDAz (Figure 3.1) (97). ManNDAz and SiaDAz can be utilized in either their free or protected forms – Ac_4ManNDAz and Ac_5 1-OMe SiaDAz, respectively. Indeed, in general, much more SiaDAz is incorporated onto the cell surface when Ac_4ManNDAz is supplemented to cells than when free ManNDAz is given. To this end, peracetylation of ManNAc analogs has been demonstrated to be generally more effective for increasing the incorporation of the sialic acid analogue than their free sugar counterpart (36,96). However, for unclear reasons, Ac_5 1-OMe sialic acid analogues are less effective for use in metabolic carbohydrate engineering (37,41).

Another limitation of metabolic carbohydrate engineering is the variable ability of certain cell lines to process sialic acid analogues and their precursors. Indeed, Jacobs *et al* demonstrated that different cell lines have different abilities to convert ManNLev to Sia5Lev and incorporate it onto the cell surface (98). Likewise, the Kohler lab has demonstrated that

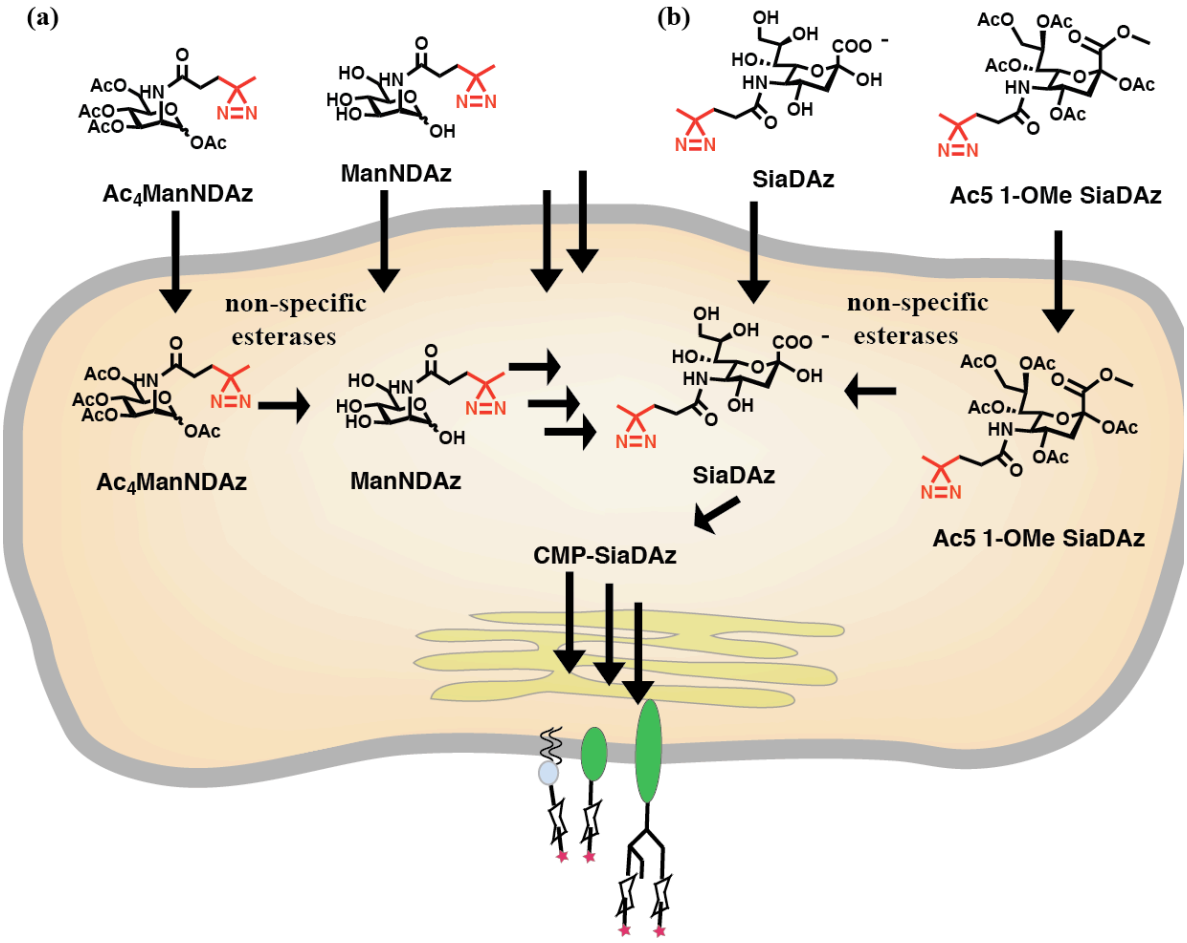

Figure 3.1. Metabolic carbohydrate engineering of the photo-crosslinking sialic acid analogue, SiaDAz and its precursor ManNDAz. (a) A sialic acid analog precursor such as ManNDAz can be provided to the cell in its free form or peracetylated form. Peracetylation makes the compound more hydrophobic so that it can cross the membrane at lower concentrations. (b) Alternatively, free SiaDAz can be supplemented in cell culture media. Esterification of all the free alcohols and the carboxylic acid, to form Ac_5 1-OMe SiaDAz makes the compound more hydrophobic allows it to traverse the cell membrane at lower concentrations. Cellular machinery converts the analogs to the activated nucleotide sugar form where it is transferred to glycoproteins and glycolipids.

various cell lines utilize ManDAz with different efficiencies, with some cell lines completely unable to incorporate SiaDAz onto the cell surface (Table 3.1). In this chapter, I will focus on the mechanistic basis of the variability of ManNDAz utilization, as well as examine the

reasons why protected sialic acid analogues yield poor incorporation efficiencies even when protected ManNAc analogues work well.

In my experimental design, the metabolism of SiaDAz and ManNDAz was investigated by employing a cellular fractionation technique to quantify the ability of cells to produce intracellular and cell surface SiaDAz. Additionally, I harvested and examined the intracellular metabolites for intermediate metabolites of the sialic acid biosynthetic pathway. Utilizing this experimental design, I was able to identify the metabolic steps that are impaired during the synthesis and incorporation of SiaDAz onto the cell surface of various cell lines. In fact, different cell lines have blockades at different steps and certain cell lines have issues at multiple steps of sialic acid analogue synthesis. The results of these experiments reveal the importance of cell-specific esterases, kinases, and phosphatases in the metabolism of Ac_4ManNDAz to ManDAz to SiaDAz. Furthermore, I found that cells poorly deprotect Ac_5 1-OMe sialic acids, both natural and unnatural. The results of these experiments detail the current limitations of metabolic carbohydrate engineering, as well as inspire strategies to improve the efficiency and generality of this technology.

RESULTS

Metabolism of ManNDAz, by different cell lines

To assess the metabolism of ManNDAz to SiaDAz and the roadblocks that prevent this conversion in certain cell lines, I supplemented cells with either Ac_4ManNDAz or its free unprotected form ManNDAz, and harvested the cells after a specified amount of time. Harvested cells were fractionated into cytosol and membrane components, and each fraction was subjected to acid hydrolysis to release sialic acids. Sialic acids were then fluorescently derivatized with 1,2-diamino-4,5-methylenedioxybenzene (DMB). Amounts of DMB-NeuAc and DMB-SiaDAz were determined by HPLC analysis with fluorescence detection (59). Sialic acids detected in the cytosol fraction are derived primarily from free sialic acids

Table 3.1. Cell surface incorporation of SiaDAz. Cells were supplemented with Ac_4ManNDAz. Percent of cell surface sialic acid that consisted of SiaDAz was measured.

Cell Line	**Media**	**% Cell surface incorporation of SiaDAz**
K88	Serum Normal	34
K20	Serum Normal	50
K20	Serum Free	100
Jurkat	Serum Normal	51
Daudi	Serum Normal	36
HeLa	Serum Normal	0
SW-48	Serum Normal	4
SW-48 Overexpressing ST6GalI	Serum Normal	4
HEK293T	Serum Normal	13
HEK293	Serum Normal	30
T84	Serum Normal	7
T84	Serum Free	15
Caco-2	Serum Normal	39
MDA-MB-231	Serum Normal	19
SK N SH	Serum Normal	15
SK SY	Serum Normal	0
PC-3	Serum Normal	0

and CMP-sialic acids, while sialic acids released from glycoproteins and glycolipids are detected in the membrane fraction. Additionally, in certain cases, intracellular metabolites were harvested from the cells and examined by high performance anion exchange chromatography (HPAEC) with either pulsed amperometric detection (PAD) or UV detection (62,91,99). MDA-MB-231, SW-48, and HeLa cell lines were studied because they represent a broad range of SiaDAz cell surface incorporation following supplementation with Ac_4ManNDAz. I performed these experiments with the aid of a rotation student, Charles Fermaintt, who worked under my direct supervision.

MDA-MB-231 cells, a breast carcinoma cell line, metabolizes Ac_4ManNDAz to SiaDAz and incorporates it into the cell surface at a moderate level (Table 3.1). After 48 hrs of supplementing cells with Ac_4ManNDAz, approximately 20% of surface sialic acids are SiaDAz (Figure 3.2a). However, supplementing cells with unprotected ManNDAz at high concentrations causes very little SiaDAz to appear on the cell surface. Evaluation of the intracellular metabolites of MDA-MB-231 cells supplemented with Ac_4ManNDAz reveals that they have high levels of intracellular SiaDAz, which are most likely in the form of CMP-SiaDAz (Figure 3.2b). Interestingly, the levels of intracellular SiaDAz are much lower in cells treated with high concentration of ManNDAz compared with cells treated with Ac_4ManNDAz. This is peculiar because Ac_4ManNDAz must be converted to ManNDAz before SiaDAz formation. This result implies that following free ManNDAz supplementation, intracellular levels of ManNDAz do not accumulate high enough to form sufficient amounts of ManNDAz-6-P to effectively compete with endogenous ManNAc-6-P (Figure 3.2c and Figure 3.7a). Taken together, these results confirm previous findings that

free sugars poorly cross the membrane and that the ManNAc kinase limits the metabolic flux of ManNAc analogs (36). Other cell lines that incorporate a higher percentage of SiaDAz into the cell surface following Ac_4ManNDAz treatment probably have intracellular kinases that facilitate ManNDAz phosphorylation.

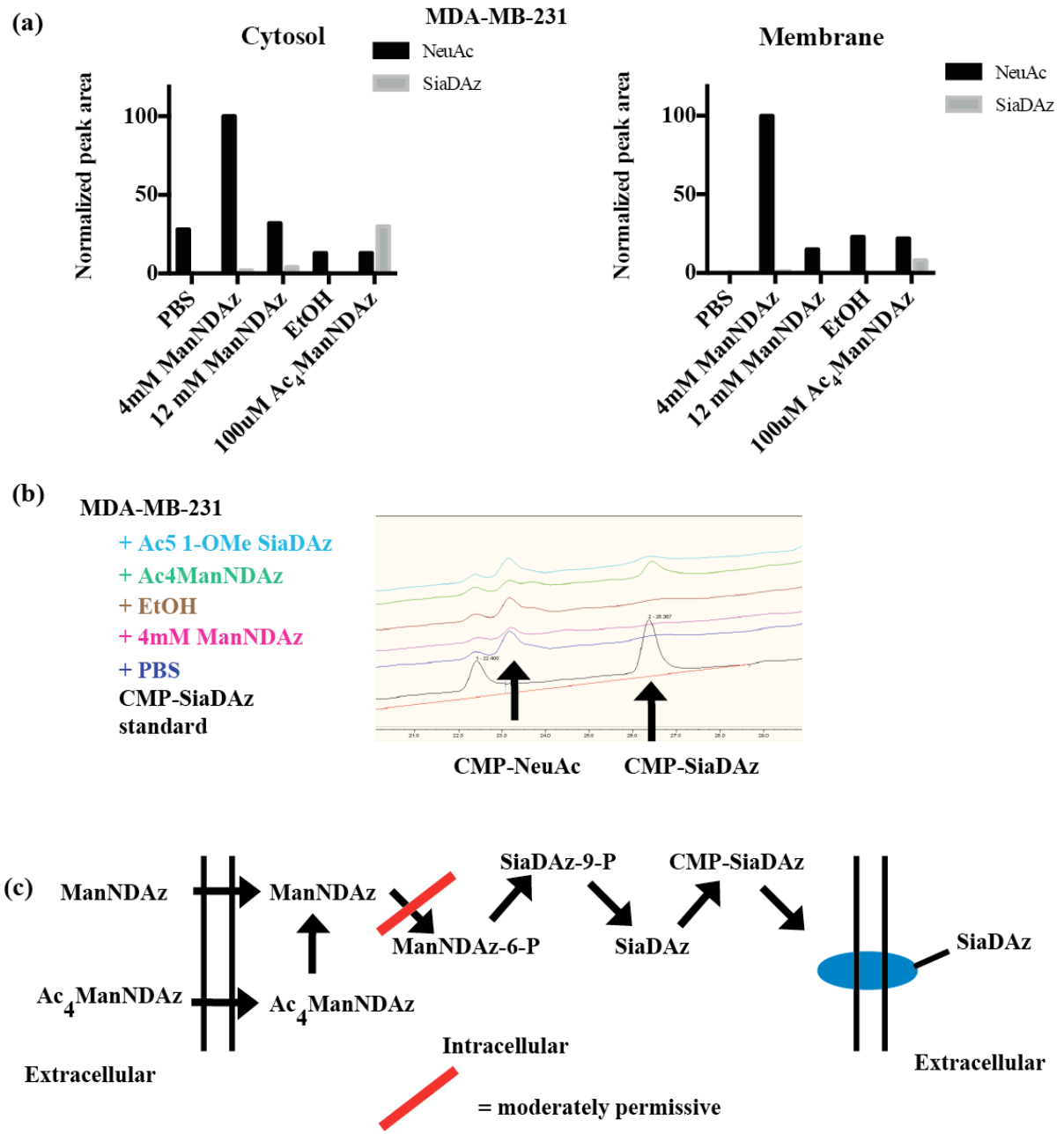

Figure 3.2. Metabolism of ManNDAz and Ac_4ManNDAz by MDA-MB-231 cells. Cells were supplemented with either ManNDAz or its protected form for 48hrs and (a) intracellular and membrane sialic acid was measured following fractionation, DMB-derivatization, and HPLC analysis. (b) Cells were supplemented for 48hrs with indicated compound and intracellular CMP-sialic acids were measured by HPAEC-UV. (c) MDA-MB-231 cells lack cellular kinases to generate high levels of ManNDAz-6-P.

SW-48 cells, a colon carcinoma cell line, metabolizes Ac_4ManNDAz to SiaDAz and incorporates it into the cell surface at a weak level (Table 3.1). SW-48 cells have very low levels of sialic acid due to low expression of sialyltransferases. To rule out lack of sialyltransferase expression as the cause of poor SiaDAz incorporation, I transduced human ST6GALI under the control of the Ef1α promoter into SW48-cells. ST6GalI is the primary sialyltransferase that acts on N-glycans; it is thought to be promiscuous in terms of accepting sialic acid analogs and has been demonstrated to transfer SiaDAz onto proteins ((100,101) and unpublished Kohler lab data). Following Ac_4ManNDAz supplementation, SW-48 cells expressing ST6GalI still have very low levels of membrane SiaDAz (although SW-48 STGalI cells have high levels of cell surface NeuAc) (Figure 3.3a). While the total amount of membrane SiaDAz is slightly higher in SW-48 ST6GalI cells when compared with wild-type cells, as a percentage they are very similar. Very surprisingly, following Ac_4ManNDAz supplementation, both cell lines have high levels of intracellular SiaDAz- approximately 40% of intracellular sialic acid is SiaDAz. There are two explanations for this: 1) CMP-SiaDAz is being formed but cannot get into the Golgi. This seemed highly unlikely because CMAS and SLC35Al, the sialic acid activating enzyme and the CMP-sialic acid transporter respectively, are both thought to be highly promiscuous and these proteins have no problem functioning in other cell lines. Furthermore, cells supplemented with Ac_4ManNAc produce high levels of CMP-NeuAc (Figure 3.3c). 2) I realized that SiaDAz-9-P might accumulate in the cell and be processed similarly to free sialic acids during the DMB derivatization step (Figure 3.3b). Indeed, I treated SW-48 cells with Ac_4ManNDAz and analyzed intracellular metabolites by

(HPAEC-PAD). A novel peak appeared in cells supplemented with Ac_4ManNDAz but not in cells supplemented with EtOH (control) or Ac_4ManNAc. UV-irradiation of the metabolites caused this peak to disappear and two new peaks to appear more quickly in the chromatograph (Figure 3.3d). This finding is characteristic for diazirine containing small molecules (91). UV-irradiation of a diazirine containing small molecule activates the photo-crosslinker to form a reactive carbene, this carbene then very quickly reacts with a nearby molecule such as water or itself. This produces two molecules that run faster on the HPAEC-PAD than the parent compound. I strongly suspect this novel peak to be SiaDAz-9-P, and that SW-48 cells lack a phosphatase to process SiaDAz-9-P and thus it accumulates within the cell (Figure 3.3e and Figure 3.7a). Confirmation by mass spectrometry is currently ongoing.

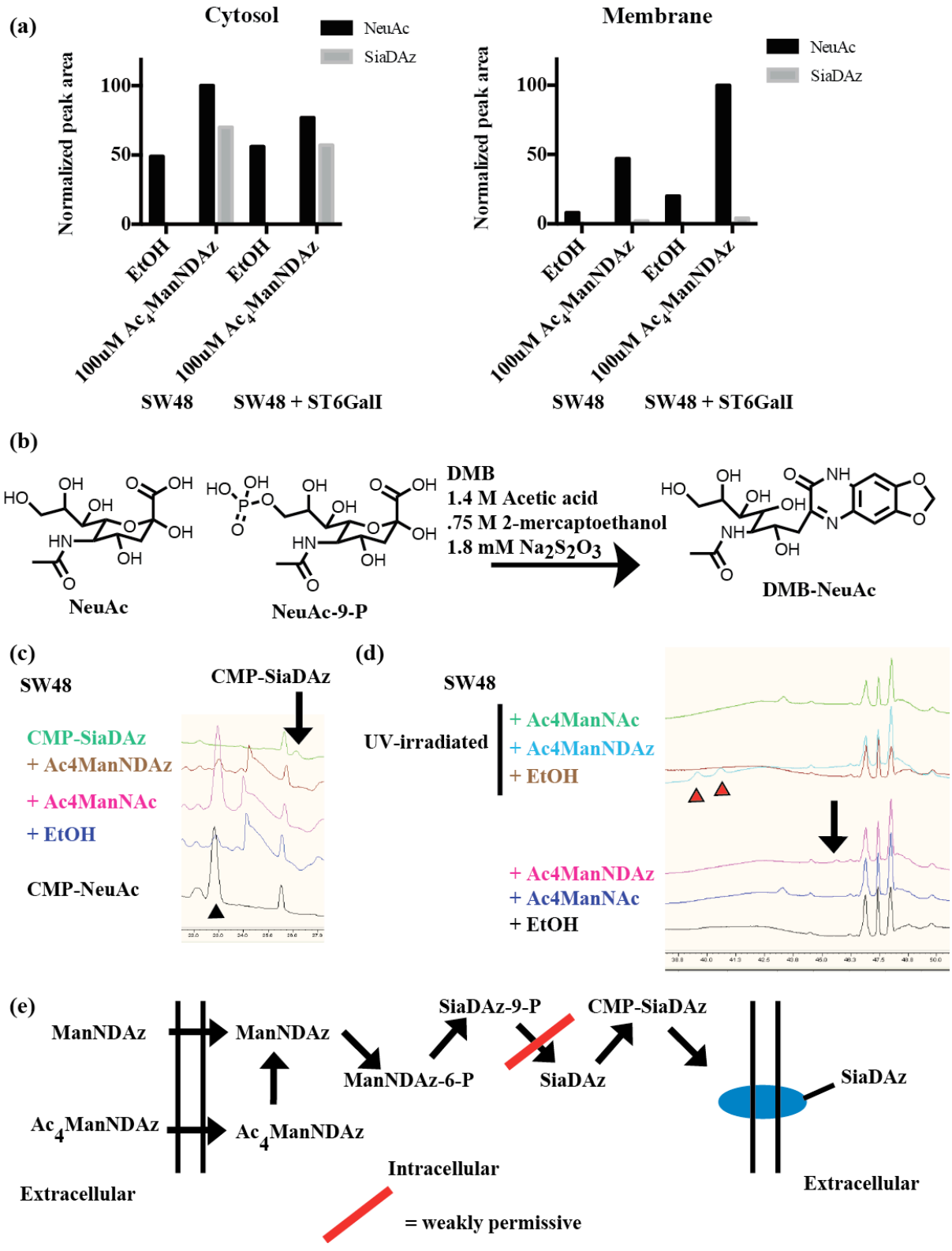
(a)
Cytosol
NeuAc
SiaDAz
Normalized peak area
100
50
0
EtOH
100uM Ac4ManNDAz
EtOH
100uM Ac4ManNDAz
SW48
SW48 + ST6GalI
Membrane
NeuAc
SiaDAz
Normalized peak area
100
50
0
EtOH
100uM Ac4ManNDAz
EtOH
100uM Ac4ManNDAz
SW48
SW48 + ST6GalI
(b)
NeuAc
NeuAc-9-P
DMB
1.4 M Acetic acid
.75 M 2-mercaptoethanol
1.8 mM Na2S2O3
DMB-NeuAc
(c)
SW48
CMP-SiaDAz
CMP-SiaDAz
+ Ac4ManNDAz
+ Ac4ManNAc
+ EtOH
CMP-NeuAc
(d)
SW48
UV-irradiated
+ Ac4ManNAc
+ Ac4ManNDAz
+ EtOH
+ Ac4ManNDAz
+ Ac4ManNAc
+ EtOH
(e)
ManNDAz
ManNDAz
ManNDAz-6-P
SiaDAz-9-P
SiaDAz
CMP-SiaDAz
SiaDAz
Ac4ManNDAz
Ac4ManNDAz
Intracellular
Extracellular
Extracellular
= weakly permissive

Figure 3.3. Metabolism of Ac_4ManNDAz by SW-48 cells. Cells were supplemented with Ac_4ManNDAz for 48 hrs. (a) intracellular and membrane sialic acid was measured. Intracellular metabolites were harvested and examined by HPAEC. (b) DMB reaction of NeuAc or NeuAc-9-P produce the same product- DMB-NeuAc. (c) Analysis of SW-48 intracellular metabolites by HPAAEC-PAD following supplementation with Ac_4ManNAc for 48hrs. Cells do not produce CMP-SiaDAz following Ac_4ManNDAz supplementation but do produce CMP-NeuAc (black arrowhead) following Ac_4ManNAc supplementation. (d) Following supplementation with Ac_4ManNDAz for 48hrs, a novel peak appears in the HPAEC-PAD chromatogram that does not appear in control cell lines (arrow). UV-irradiating metabolites caused the new peak to disappear and two new peaks to appear earlier in the chromatogram (red arrow heads). This is a characteristic finding for diazirine containing compounds. The novel peak is most likely SiaDAz-9-P.

HeLa cells, a cervical carcinoma cell line, do not incorporate SiaDAz into the cell surface following supplementation with Ac_4ManNDAz (Table 3.1). In fact, no SiaDAz is incorporated into the cell surface after supplementing HeLa cells either Ac_4ManNDAz or high concentrations of free ManNDAz (Figure 3.4a). However, in contrast to MDA-MB-231 cells, more intracellular SiaDAz is produced following ManNDAz treatment than with Ac_4ManNDAz treatment. This implies that HeLa cells cannot remove the acetyl groups from Ac_4ManNDAz to form free ManNDAz. It has always been assumed that the acetyl protecting groups are processed rapidly by “non-specific” esterases, and for some esterified ManNAc analogs this has been experimentally demonstrated (38). However these non-specific esterases must be present in the cell to carry out this function, and depending on the enzyme present in the cell, deacetylation may occur at drastically different rates (102). The inability of HeLa cells to deacetylate Ac_4ManNDAz does not reflect a general esterase defect because HeLa cells are able to metabolize Ac_4ManNAc to CMP-NeuAc (Figure 3.4c). While it is tempting to speculate that Ac_4ManNAc is more poorly metabolized than free ManNAc, keep in mind that Ac_4ManNAc is supplied at a concentration of 100 uM while ManNAc is

given at 4 mM. Furthermore, because the SiaDAz is produced intracellularly but does not get converted to CMP-SiaDAz (Figure 3.4b), this intracellular SiaDAz must be SiaDAz-9-P. Thus HeLa cells have two metabolic steps that are impaired during the metabolism of Ac_4ManNDAz: 1) deacetylation of Ac_4ManNDAz and 2) dephosphorylation of SiaDAz-9-P (Figure 3.4d).

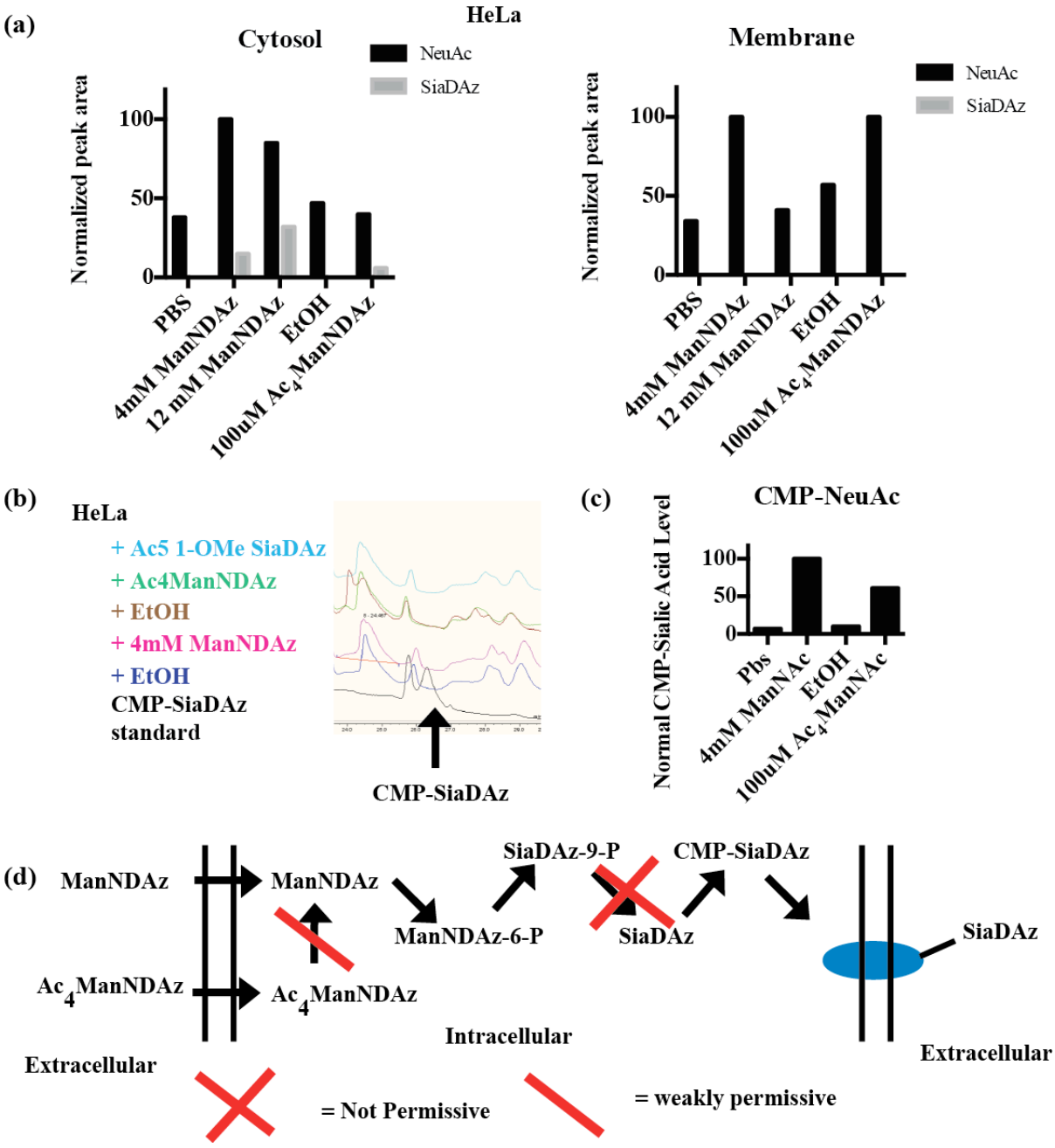

Figure 3.4. Metabolism of ManNDAz and Ac_4ManNDAz by HeLa cells. Cells were supplemented with either ManNDAz or its protected form for 48 hrs and (a) intracellular and membrane sialic acid was measured following fractionation, DMB-derivatization, and HPLC analysis. (b) Intracellular metabolites were harvested and analyzed by HPAEC-PAD. Cells do not produce CMP-SiaDAz following supplementation with sugar for 48 hrs. (c) Cells were cultures with sugar or control compound for 24 hrs and CMP-sialic acid levels were

analyzed by HPAEC-UV. (d) HeLa cells have two issues with metabolizing Ac_4ManNDAz – deacetylation of Ac_4ManNDAz to ManNDAz and dephosphorylation of SiaDAz-9-P to SiaDAz.

Protected sialic acids are poorly metabolized

All of the metabolic blockades for processing Ac_4ManNDAz and ManNDAz are upstream of the production of the free sialic acid analog. Utilizing sialic acid analogs and their protected forms can thus potentially circumvent these issues, which inspired me to study the metabolism of sialic acid analogs in greater detail. I utilized K20 and Jurkat cells because they metabolize ManNDAz very well and any defect in metabolism is unlikely due to analog intolerance.

I began my studies with K20 cells, a B-cell lymphoma, which processes Ac_4ManNDAz to SiaDAz very well and incorporates SiaDAz into the cell surface at high levels (Table 3.1). This cell line is also a good model system because it does not synthesize sialic acid due to a lack of expression of GNE, and therefore acquires all of its sialic acid from media (58). This results in lower baseline levels of cell surface sialic acids and intracellular sialic acid metabolites. I began by supplementing K20 cells with a variety of free and protected sugars (ManNAc, ManNDAz, NeuAc, Ac_4ManNAc Ac_4ManNDAz, Ac_5 1-OMe NeuAc, Ac_5 1-OMe SiaDAz) and examined the formation of CMP-Sialic acid (CMP-NeuAc and CMP-SiaDAz) by HPAEC-UV. No free SiaDAz was available for theses experiments because it is challenging to synthesis with high purity in large quantities. K20 cells supplemented with Ac_4ManNDAz produce high levels of CMP-SiaDAz, but surprisingly those supplemented with Ac_5 1-OMe SiaDAz produced very low amounts

(Figure 3.5a) (cells supplemented with free ManNDAz also have low amounts of CMP-SiaDAz but this is expected). Furthermore, cells supplemented with natural sugar variants produce high amounts of CMP-NeuAc (Ac_4ManNAc >> NeuAc > ManNAc), while very surprisingly cells given Ac_5 1-OMe NeuAc produce very low levels of CMP-NeuAc. Following supplementation with Ac_5 1-OMe NeuAc, CMP-NeuAc is not made at 6hr or at 24hrs, thereby ruling out the possibility that the fully protected NeuAc is processed extremely quickly (Figure 3.5b). K20 cells supplemented with Ac_5 1-OMe NeuAc had lower levels of cell surface sialic acid then cells given Ac_4ManNAc, free ManNAc, or free NeuAc as measured by MAL-lectin staining and flow cytometry (Figure 3.5c). Overall, these data suggest that K20 cells poorly metabolize fully protected sialic acids, both natural (Ac_5 1-OMe NeuAc) and unnatural (Ac_5 1-OMe SiaDAz).

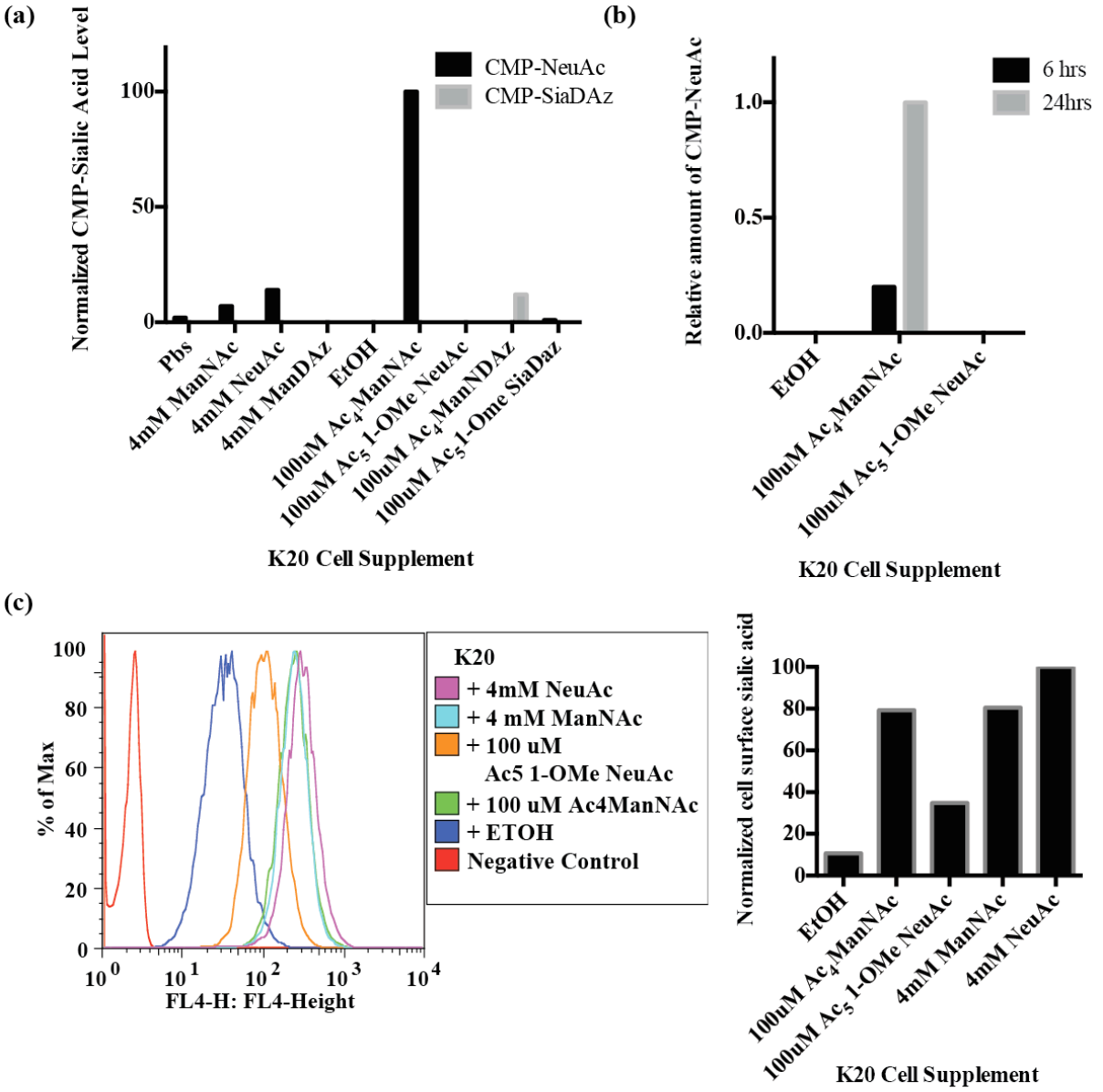

Figure 3.5. K20 cells poorly metabolize fully protected sialic acids, natural and unnatural. (a) K20 cells were supplemented with various sugars for 24hrs and intracellular levels of CMP-sialic acid were measured. (b) K20 cells supplemented Ac_5 1-OMe NeuAc do not produce CMP-NeuAc at 6hrs or 24hrs. (c) After supplementation with various sugars for 48 hrs, cell surface sialic acid was measured by flow cytometry following staining with MALII-lectin. Raw flow cytometry data (left) and data graphed (right).

To understand why protected sialic acids are metabolized poorly, I focused on Ac_5 1-OMe NeuAc as a model system because it was easy to acquire in large quantities. Utilizing Ac_5 1-OMe NeuAc also has the advantage that the free sialic acid form is natural to human cell lines and avoids confounding factors that could be associated with sialic acid analogs. Once Ac_5 1-OMe NeuAc enters the cell it must be processed by at least two different proteins to remove the two different types of protecting groups (Figure 3.6a). To understand which of these deprotection steps was most responsible for the poor metabolism of Ac_5 1-OMe NeuAc, I supplemented Jurkat cells, a T cell lymphoma cell line, with Ac_5 1-OMe NeuAc and examined the intracellular metabolites by liquid chromatography coupled to a tandem mass spectrometer (LC-MS/MS) (65). Jurkat cells contained detectable levels of intracellular Ac_5 1-OMe NeuAc and 1-OMe NeuAc, following supplementation with Ac_5 1-OMe NeuAc for 24 hrs (Figure 3.6 b and c). Thus deacetylation and demethylation of Ac_5 1-OMe NeuAc are problematic (Figure 3.7b). However, because 1-OMe NeuAc can be detected at high levels, and acetyl groups are generally easy to remove (38,103,104), the O-methyl protecting group is probably the main bottleneck for metabolizing fully protected sialic acids.

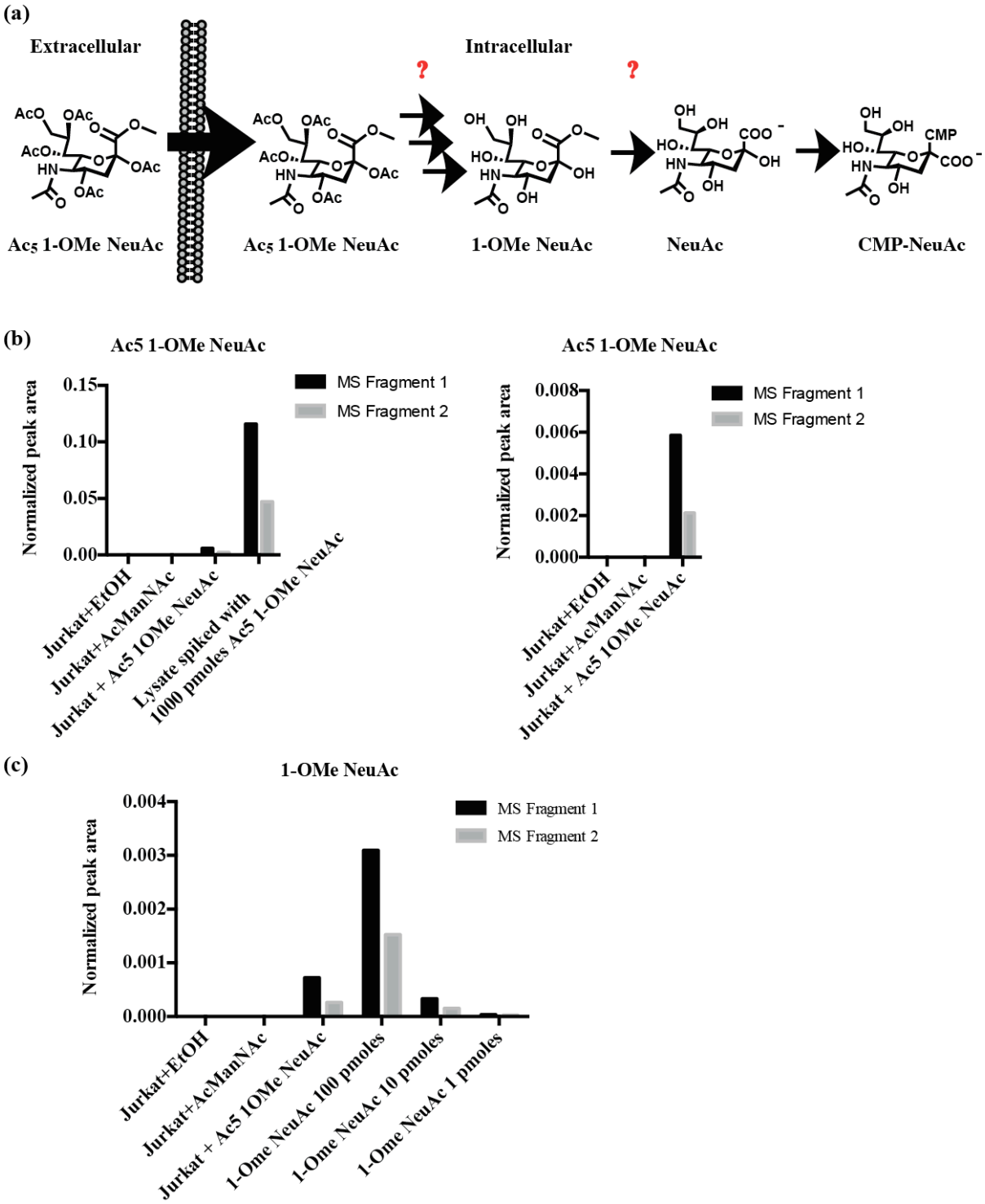

Figure 3.6. Intermediate metabolites detected intracellularly following Ac_5 1-OMe NeuAc supplementation (a) Following cellular supplementation with Ac_5 1-OMe NeuAc, at least two deprotection steps must occur, deacetylation and demethylation. The steps are drawn linearly but actual cellular breakdown can follow a complex and variable order.

Jurkat cells were supplemented with Ac_5 1-OMe NeuAc for 24hrs and intracellular metabolites were analyzed by LC-MS/MS. (b) Intracellular Ac_5 1-OMe NeuAc and (c) intracellular 1-OMe NeuAc were detected.

DISCUSSION

Sialylation is implicated in many physiologic and pathological processes, but detailed mechanistic understanding is lacking in many cases. The use of sialic acid analogs and metabolic engineering is a useful tool for studying the roles of sialic acid in biological systems. While a wide range of cell lines appear to tolerate sialic acid analogs and analogs of their ManNAc precursors, their metabolism can vary across cell lines. In the case of the photo-crosslinking sialic acid analog SiaDAz, sialic acid–protein interactions can be ascertained which would be difficult to study utilizing traditional biochemical techniques. Typically, cells are supplemented with either SiaDAz (free or as the protected Ac_5 1-OMe SiaDAz) or the SiaDAz precursor ManNDAz (free or as the protected Ac_4ManNDAz). Unfortunately, the use of this sialic acid analog is limited by variable metabolism of Ac_4ManNDAz to SiaDAz by different cell lines with some cells completely unable to present any SiaDAz onto the cell surface (Table 3.1). The metabolism of SiaDAz and ManNDAz was explored in depth by employing cellular fractionation to investigate the ability of cells to produce intracellular and extracellular SiaDAz. Furthermore, I examined the intracellular metabolites and looked for intermediate metabolites of the sialic acid biosynthetic pathway. Utilizing this experimental design, I was able to identify the blockades that impair SiaDAz synthesis and incorporation onto the cell surface of various cell lines.

Cell-specific proteins modulate the ability to metabolize ManNDAz

Metabolic engineering takes advantage of the promiscuity of cellular enzymes to process analogs of naturally occurring compounds. ManNDAz must compete with endogenous ManNAc for the utilization of the enzymes GNE, NANS, NANP, and CMAS during the production of SiaDAz and it's activated form, CMP-SiaDAz (Figure 3.7a). Underappreciated are the roles of enzymes outside of this pathway that participate in the metabolism of ManNDAz (Figure 3.7a). Indeed, K20 cells do not express GNE, a bifunctional protein that contains a ManNAc kinase domain. Despite this fact, K20 cells metabolize ManNDAz very well, suggesting that other intracellular kinases within K20 cells are able to phosphorylate ManNDAz. Furthemore, Jacobs *et al* previously reported that the ManNAc kinase works poorly on ManNAc analogs with larger N-acyl sidechains (36). Thus, even in the presence of a fully functional GNE, cells still rely on other intracellular kinases to metabolize ManNDAz. When MDA-MB-231 cells are given Ac_4ManNDAz, they can produce moderate levels of intracellular and cell surface SiaDAz, but when they are supplemented with ManNDAz very little SiaDAz is produced (Figure 3.2). This is consistent with poor phosphorylation of ManNDAz by MDA-MB-231 cells. When Ac_4ManNDAz enters MDA-MB-231 cells, it is converted to ManNDAz at high enough concentrations to be utilized be cellular kinases acting promiscuously. However, free ManNDAz has trouble traversing the plasma membrane due to its hydrophilic nature (this is generally true of free sugars) and cannot accumulate at high enough concentrations to be phosphorylated by enzymes that are not designed to specifically carry out this reaction.

Careful consideration of the metabolism of Ac_4ManNDAz revealed that cells would need esterases to act promiscuously to convert Ac_4ManNDAz to ManDAz, and could also

need other phosphatases to convert SiaDAz-9-P to SiaDAz (Figure 3.7a). SW-48 cells supplemented with Ac_4ManNDAz produce high levels of intracellular SiaDAz but very little cell surface SiaDAz. This intracellular SiaDAz produced is most likely in the form of SiaDAz-9-P. Thus SW-48 cells lack a phosphatase that can promiscuously act on SiaDAz-9-P (Figure 3.3d). To my knowledge, this is the first report that NANP poorly tolerates a sialic acid analog and the first time other cellular phosphatases have been implicated in carbohydrate metabolic engineering. Indeed, the dephosphorylation of sialic analogs may be a cellular bottleneck for processing other ManNAc analogs. It is important to note that there is no reason for these metabolic defects to be mutually exclusive. Indeed, HeLa cells lack an esterases and a phosphatase to assist in the conversion of Ac_4ManNDAz to SiaDAz. HeLa cells supplemented with either Ac_4ManNDAz or ManNDAz produce no cell surface sialic acid; but in contrast to MDA-MB-231 cells, HeLa cells given ManNDAz make more intracellular SiaDAz than cells given Ac_4ManNDAz. In this case, free ManNDAz accumulates at high enough concentrations to be utilized by kinases found in HeLa cells to form levels of ManNDAz-6-P that can effectively compete with endogenous ManNAc-6-P for NANS, the enzyme that produces NeuAc-9-P. Ac_4ManNDAz given to HeLa cells enters the cells but must not get converted to ManNDAz (otherwise it would be converted to SiaDAz-9-P). Human carboxyesterases are thought to have various expression levels in different tissues and different substrate selectivities (102) and this variation may account for the inability of HeLa cells to process Ac_4ManNDAz.

The inability of HeLa cells to deacetylate Ac_4ManNDAz was a surprising finding, especially because HeLa cells were able to metabolize Ac_4ManNAc. Furthermore, Jones *et*

al report that HeLa cells can metabolize Ac_4ManNLev (96). It is important to remember that ManNAc, ManNDAz, and ManNLev are structurally similar, but are still different compounds and it can never be assumed that an enzyme that works on one small molecule will work on the other. Furthermore, a cell having an inability to deacetylate a small molecule has precedence. HEK-293 cells cannot deacetylate the pro-drug aspirin to form active salicylic acid (105). An alternative explanation is that HeLa cells degrade Ac_4ManNDAz before it can be converted to ManNDAz. This possibility notwithstanding, this hypothetical degradation pathway could be circumvented by the presence of a fast acting esterase. ManNDAz itself appears to be well tolerated by HeLa cells as it can be utilized to produce SiaDAz-9-P.

Taken together these experiments highlight the importance of intracellular esterases, kinases, and phosphatases in the metabolism of Ac_4ManNDAz and ManNDAz. Furthermore, differential expression of these enzymes in various cells lines accounts, at least in part, for the variability in metabolizing ManNAc analogs.

Metabolism of Ac_5 1-OMe sialic acids

An alternative strategy to sialic acid metabolic engineering involves the use of the sialic acid analog itself as opposed to its ManNAc analog precursor. This strategy also allows for the use of sialic acid analogs with substitutions at the 9-position, as it was ManNAc analogs with alterations at the 6-position are not tolerated by the sialic acid biosynthetic pathway (the 6 position of ManNAc becomes the 9-position of sialic acid). While peracetylated ManNAc analogs have been demonstrated to be generally effective at

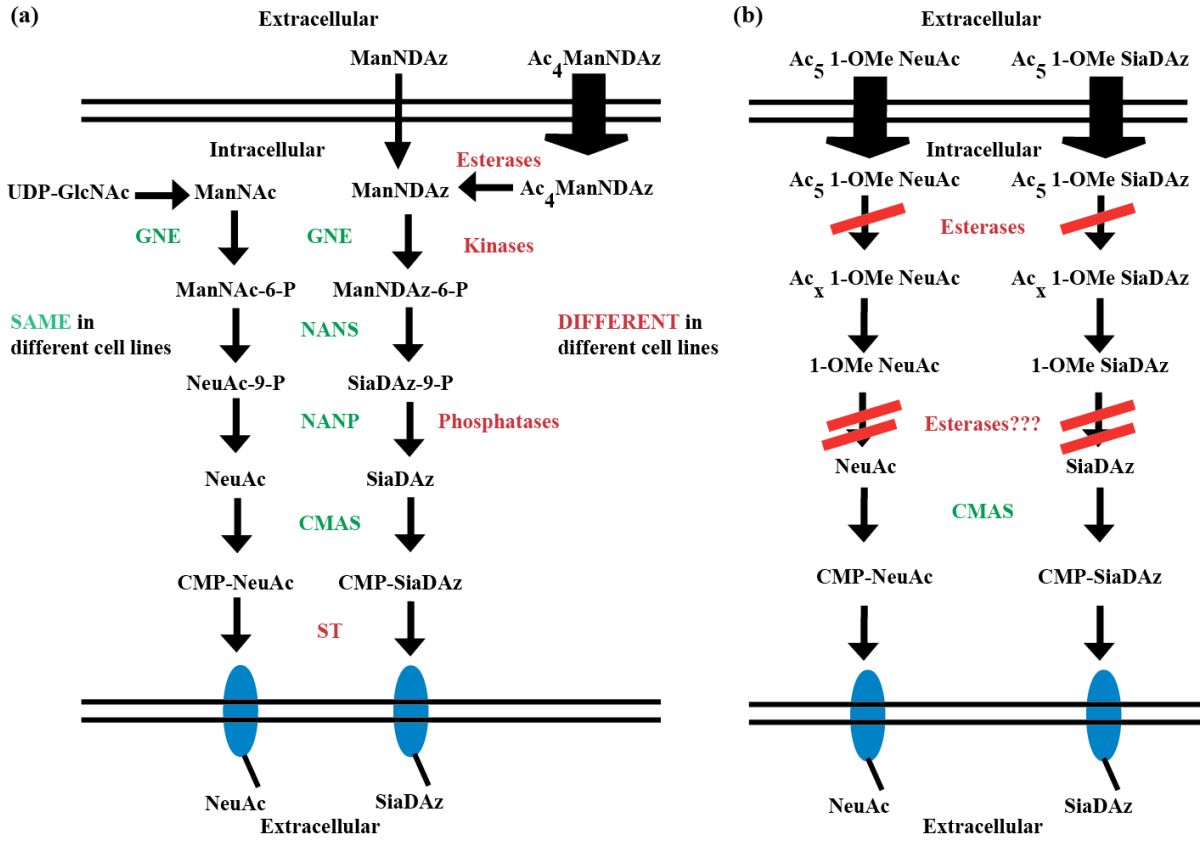

Figure 3.7. Metabolic pathways for metabolizing sialic acid analogs and precursors. (a) Analogs and their intermediate metabolites compete with endogenous sialic acid precursors for common enzymes. Cell-specific esterases, kinases, and phosphatases aid in the processing of Ac_4ManNDAz to ManNDAz to SiaDAz. (b) Ac_5 1-OMe sialic acids are able to enter into the cell and both Ac_5 1-OMe sialic acid and 1-OMe sialic acid accumulate within the cell.

increasing cell surface display of sialic acid analogs (with some exceptions), the Ac_5 1-OMe protected sialic acid analogs appear to be less effective than providing the free sialic acid itself (37). While the free form of a sialic acid analog has been utilized effectively, they are limited by the difficulty of synthesizing large quantities and their cellular toxicity. Feng *et al* reported that a bifunctional sialic acid analog inhibited cell growth when used at high concentrations (3 mM) and protected sialic acid analogs could potentially be used at

concentrations as low as 100 uM (41). I thus sought to explore the mechanisms behind the poor metabolism of protected sialic acids.

K20 cells, a cell line that metabolizes Ac_4ManNDAz very well, supplemented with Ac_5 1-OMe SiaDAz have less cell surface SiaDAz than cells supplemented with Ac_4ManNDAz (data not shown) and much less of the intracellular metabolite CMP-SiaDAz (Figure 3.5). Importantly, this finding is not exclusive to a sialic acid analog. Indeed, there is less cell surface sialic acid (natural NeuAc) and less intracellular CMP-NeuAc when K20 cells are given Ac_5 1-OMe NeuAc, when compared to cells given Ac_4ManNAc. This finding suggests that there is a fundamental problem in removing the acetyl groups protecting the alcohol groups on sialic acid and/or removing the methyl ester protecting the carboxylic acid on sialic acid. Mass spectrometry experiments confirm the presence of intracellular Ac_5 1-OMe NeuAc and 1-OMe NeuAc in Jurkat cells supplemented with Ac_5 1-OMe NeuAc, suggesting that removal of both types of protecting groups can be problematic. Consistent with this idea, Tian *et al* demonstrated that bulky substitutions on the acyl chain of ester protecting groups are poorly tolerated by cellular esterases (106). In the case of 1-OMe sialic acid, the cyclic core is a bulky substitution that could preclude the action of cellular esterases, and thus another class of enzymes may be critical in removing the O-methyl protecting group.

Degradation pathways for protected sialic acids may also be present in various cell lines, but the fact that both the fully protected and 1-OMe NeuAc can be detected intracellularly suggests that this pathway, if present, is minor and could probably be

circumvented by more effective deprotection. While certain cell lines may have trouble removing acetyl groups from sialic acid analogs (analogous to HeLa cells being unable to deacetylate Ac_4ManNDAz), this is probably less common, and I propose that developing sialic acid analogs with different protecting groups at the carboxylic acid position will yield analogs that will allow for high incorporation efficiency. Due to the general ease of removing acetyl groups, facilitating the intracellular deprotection of the carboxylic acid may drive conversion of Ac_5 1-"protected" sialic acids to free sialic acids.

Future directions for metabolic carbohydrate engineering

The multiple possible metabolic defects associated with utilizing ManNAc analogs and variable metabolism is not isolated to only ManNDAz. In fact, Jacobs *et al* previously described variable metabolism of ManNLev by various cell lines (98). I propose that many of the same principles described in this chapter apply to ManNLev and other ManNAc analogues. The metabolism of ManNAc analogues will need to be determined empirically, as closely related compounds may not behave similarly in the same cell line. In fact, HeLa cells can metabolize ManNLev to Sia5Lev and incorporate it into the cell surface but cannot metabolize ManNDAz to free SiaDAz.

While the endogenous sialic acid biosynthetic pathway is tolerant to the use of ManNAc and sialic acid analogues, not every enzyme is equally tolerant. This phenomenon is paralleled by studies detailing metabolic carbohydrate engineering technology utilizing the diazirine containing N-Acetyl glucosamine analog, GlcNDAz. Here, GlcNAc is phosphorylated to GlcNAc-6-P by NAGK then converted to GlcNAc-1-P by AGM1,

GlcNAc-1-P is then activated to UDP-GlcNAc by AGX1. Yu *et al* report that several of these enzymes are intolerant to GlcNDAz (91). They circumvent these issues by delivering a more advanced precursor GlcNDAz-1-P and mutating AGX1 to be promiscuous towards GlcNDAz-1-P. It should be possible employ a similar strategy to enzymes in the sialic acid biosynthetic pathway, that is developing enzymes that alleviate specific blockades found in certain cell lines (i.e. a ManNDAz kinase to be used in MDA-MB-231 cells). However, such enzymes may not work on other analogs and can only be applied to cells with that specific defect and no other defect. For theses reasons, I propose that future efforts should focus on developing new sialic acid protecting groups. Protecting groups that are generally easier to remove by broadly expressing cellular enzymes could be developed. Conversely, the inability of cellular enzymes to deprotect an analogue could be leveraged. A sialic acid analogue and "de-protecting" enzyme pair could be developed to utilize sialic acid analogues in a cell type specific manner. For instance, different cell lines can be co-cultured but only the cell line expressing the "de-protecting" enzyme will utilize the sialic acid analogue. This principle could also be applied to in-vivo studies. For instance, the de-protecting enzyme could be delivered to mouse neurons specifically, and then sialic acid analogues could be incorporated into neural glycoconjugates following feeding mice the sialic acid analogue (107). In this way, the roles of sialic acid in the brain could be studied by this technology.

ACKNOWLEDGEMENTS

I would like to thank Charles Fermaintt, an excellent student who I had the honor of mentoring. Under my direct supervision, he carried out several experiments described in this

chapter. Yibing, Randy, and Amberlyn provided reagents that I used to perform these experiments. I especially would like to think Sunil Laxman and Ben Tu for being generous with their LC-MS/MS machine. Sunil particularly was extremely helpful in designing the mass spectrometry experiment. I also would like to thank every member of the Kohler lab as they also helped evaluate cell lines for SiaDAz cell surface incorporation.

METHODS

Cell lines and culturing conditions

Cell lines (MD-MBA-231, HeLa, and SW-48 and HEK293T) were purchased from ATCC. All the cell lines were cultured at 37 °C, 5% CO_2 in DMEM (11995-073, Life Technologies) containing 10% fetal bovine serum (10082, Life Technologies) and supplemented with penicillin and streptomycin. BJAB K20 cells (obtained from Michael Pawlita (German Cancer Research Center) and James Paulson (The Scripps Research Institute)) and Jurkat cells were cultured in RPMI 1640 media containing 2 mM glutamine and supplemented with 10% fetal calf serum, 100 U/mL penicillin, and 100 µg/mL streptomycin at 37 °C, 5% CO_2 in a water-saturated environment.

ManNAc (Sigma), NeuAc (Carbosynth) were purchased commercially. Ac_4ManNAc, Ac_5 1-OMe SiaDAz, ManNDAz, and Ac_4ManNDAz were synthesized as previously described (40).

Generation of SW48 cells overexpressing ST6GALI

The human ST6GAL1 and eGFP genes were amplified by PCR from pCMV-Sport6-ST6Gal1 (MHS1010-97227857, Thermo Scientific). The PCR product was digested and ligated into the lentiviral destination vector, pSIN4-EF1a-mGli2-IRES-Puro (108) (a gift from Dr. Jiang Wu, University of Texas Southwestern Medical Center). Plasmid was verified by DNA sequencing (UT Southwestern DNA Sequencing Core Facility).

HEK-293T cells were transfected with pSIN4-ST6Gal1and pSIN4-eGFP plasmids, respectively, accompanied with psPAX2 (12260, Addgene) and pMD2.G (C12259, Addgene) in the presence of Fugene 6 (E2691, Promega) to generate lentivirus. Media was replaced after 20 h. After 2 days, supernatant containing lentivirus was harvested. SW-48 cells were seeded into 6-well tissue culture wells and allowed to grow for 18 h. Then cells were incubated with 1:1 lentivirus stock solution diluted with DMEM and supplemented with 4 μg/mL polybrene (AL-118, Sigma-Aldrich) to enhance infection efficiency. After 24 h, supernatant was removed by aspiration and fresh DMEM was added to the cells. After an additional 24 h, cells were transferred to a 10-cm diameter tissue culture dish and cultured for two weeks in the presence of 1.5 μg/mL puromycin to select for infected clones.

Flow cytometry analysis

K20 cells were collected and washed twice with dPBS, then resuspended in dPBS at 2.0×10^6 cells/mL. Then 200 μl of cell suspension was transferred to V-bottom 96-well plate and centrifuged once more. The cell pellets were incubated with 100 μl of 1:100 MALII-biotin (Vector) for 60 min at 4 °C, then washed with D-PBS three times. The cells were then incubated with 1:200 allophycocyanin-streptavidin (S-868, Life Sciences) for 45 minutes.

Fluorescence was analyzed by flow cytometry on a FACSCalibur instrument (BD Biosciences) equipped with dual lasers at 488-nm and 635-nm.

Cellular fractionation and DMB-derivitzation of sialic acids

Freshly harvested cells were fractionated as described previously (89). Briefly, cells were counted and harvested by centrifugation and then suspended in hypotonic lysis buffer containing protease inhibitors for 15 min on ice. Cells were lysed by extrusion through a 25 gauge needle. Nuclei and unbroken cells were removed from the post-nuclear supernatant by two rounds of centrifugation at 1000*g* for 15 min at 4 °C. Next, the post-nuclear supernatant was transferred to heavy-walled polycarbonate tubes and centrifuged at 100,000*g* for 1 h in a Beckman TLA 120.2 rotor. The supernatant was designated the cytosolic fraction (C). The pellet was washed twice with 400 µL cold hypotonic lysis buffer followed by centrifugation at 100,000*g* for 1 h after each wash. The remaining pellet was designated the membrane fraction (M). Samples were flash-frozen and the solvent removed by vacuum overnight.

Quantification of sialic acids

Quantification was performed relative to known standards, prepared as follows. Calibration curves were prepared by injecting between 50 to 750 fmol of DMB-Neu5Ac and DMB-Neu5Gc on a Dionex Acclaim® Polar Advantage C16 5µm, 4.6 x 250 mm column attached to a Dionex Ultimate 3000 HPLC with fluorescence detector. Separation was performed using a gradient of 2-90% acetonitrile (ddH_2O) with fluorescence detection (ex.

373 nm, em. 448 nm). Linearity of the fluorescence signal of DMB-Neu5Ac confirmed in the range of 50 to 750 fmol.

To release sialic acids, 2.0 M acetic acid was added to each of the membrane-bound (M: 50 uL) and dried cytosolic samples (C: 100 uL). Solutions were incubated at 80 °C for 2 h. Samples were cooled to room temperature, then DMB reaction solution (1.4 M acetic acid, .75 M 2-mercaptoethanol, 18mM $Na_2S_2O_3$) was added to each sample (M: 40 uL and C: 80 uL). After a 2 h incubation at 50 °C, 0.2 M NaOH was added to each sample (M: 10 uL and C: 20 uL). Samples were filtered through 10 kDa MWCO filters by centrifugation and the resulting flow-through was stored at −20 °C in the dark until analysis. Generally, 10 μL of the derivatized material was diluted with 90 μL ddH_2O and analyzed by fluorescence HPLC, as described for the standards. Linear regression analysis using the DMB-NeuAc standard curves was used to calculate the amount of DMB-Neu5Ac present in experimental samples.

Intracellular metabolite harvesting

Cells were supplemented with sugar for various time points as indicated and cells were harvested and counted. Typically ~2 x 10^6 cells were harvested. Cells were spun, and washed with cold dPBS twice and flash frozen. Cells were then lysed with 80% "super-cold" methanol (on dry ice) (90). Lysate was spun at 2,000g for 15 min. The supernatant was flash frozen and dried by a speed vacuum for 4-5 hours. The intracellular metabolite pellet was typically used immediately or it was stored at -80° C for future use.

Analysis of intracellular metabolites by HPAEC

Metabolite pellet was resuspended in 40 mM sodium phosphate buffer (pH 7.4 40 μL per million cells), and filtered through an Amicon® Ultra centrifugal filter unit (Millipore, 10,000 MWCO). Filtrates and standards were analyzed by HPAEC (ICS-3000 system, Dionex) with CarboPac™PA1 (Dionex) with a pulsed amperometry detector (PAD) and UV-detector in-line. Typically, 20 μL of metabolite was injected into the sample loading loop before the sample enters a guard column (Dionex, 4 × 50 mm) and then an analytical column (Dionex, 4 × 250 mm). The eluents used were 1.0 mM NaOH (C) and 1.0 M NaOAc and 1.0 mM NaOH (D). Low-carbonate NaOH (50% in water) was obtained from Fisher Scientific (SS254-1) and NaOAc was from Sigma (71183). HPAEC was run with a flow rate = 1 mL/min and the following gradient elution was performed: $T_{0\,\mathrm{min}}$ = 95% C, T_{40} = 70% C, T_{45} = 50% C, T_{60} = 45% C, T_{65} = 0% C, T_{75} = 0% C, T_{80} = 95% C, T_{90} = 95% C. Peaks were compared to standards of CMP-NeuAc and CMP-SiaDAz. Data was normalized to cell number.

Analysis of intracellular metabolites by LC-MS/MS

LC-MS/MS with a triple quadrupole mass spectrometer (3200 QTRAP, ABSCIEX) was utilized to analyze intracellular metabolites as previously described (92). Intracellular metabolite pellet was dissolved in solvent A and allowed sit on ice for 30 minutes. Solution was spun at 16,000g for 5 minutes at 4° C to pellet any precipitation. The supernatant was filtered through .2 um PVDF micro spin filter (Grace Davison Discovery Science, 8604742) at 9000 g for 5 min at 4° C twice. Intracellular metabolites and standards were directly injected into this column and chromatographed at a flow rate of 0.5 ml/min. The column used

for HPLC separation was a Synergi 4u Fusion-RP 80A reverse phase column (Phenomenex). The solvent system consisted of 5 mM ammonium acetate in water (solvent A) and 5 mM ammonium acetate in methanol (solvent B). Injected samples were eluted with the following gradient of solvent B: 0%-0.2%/5 min, 0.2%-1%/1 min, 1%-3%/1 min, 3%-5%/1 min, 5%-25%/6 min, 25%-50%/4 min, 50%-100%/2 min. The retention time and the *m/z* (positive mode: MW+1) of the parent or product ion were as follows: Ac5 1-OMe NeuAc 12-13 min 534/414 & 534/129; 1-OMe NeuAc 2 min 324/306 & 324/288. Typically, both MRMs for a given metabolite displayed high correlation. The retention time for each MRM peak was matched with the metabolite standard. The area under each peak was then quantitated by using Analyst software, with manual inspection for accuracy. The raw peak areas were compared and analyzed in excel and graphed in Prism 6.

CHAPTER FOUR
Statistical Coupling Analysis Reveals Key Residues Governing Glycosyltransferase Activity

INTRODUCTION

Glycosyltransferases (GTs) are an ancient and conserved family of proteins essential for life. An estimated 1-2% of genes in every sequenced genome are dedicated to GTs (109). In humans, an emerging class of diseases called the congenital disorders of glycosylation (CDG) result from mutations in the myriad of pathways of glycosylation. Additionally, many glycosylated small molecules are therapeutically valuable in treating a variety of diseases (Figure 4.1). Furthermore, altering the glycan moiety can significantly alter their activities and pharmacokinetic properties (52). Thus, engineering to diversify and control GT function would have widespread biological and clinical applications (110). However, this field of research is currently limited by a lack of understanding of which amino acids control the enzymatic and structural properties of GTs.

A wealth of information about GT function has been gleaned from structural studies of the GT family (111). Despite the large number of GT crystal structures, it is still not trivial to determine residues important for catalytic function (particularly those outside the active site). For instance, the catalytic residues and mechanism utilized by retaining GT's is still being debated (112). In other cases, catalytic function of a GT can be modified by allosteric action at a distal site; however, the residues communicating action at a distal allosteric site to the active site of a protein cannot be ascertained from structure alone, particularly without the co-crystal structures.

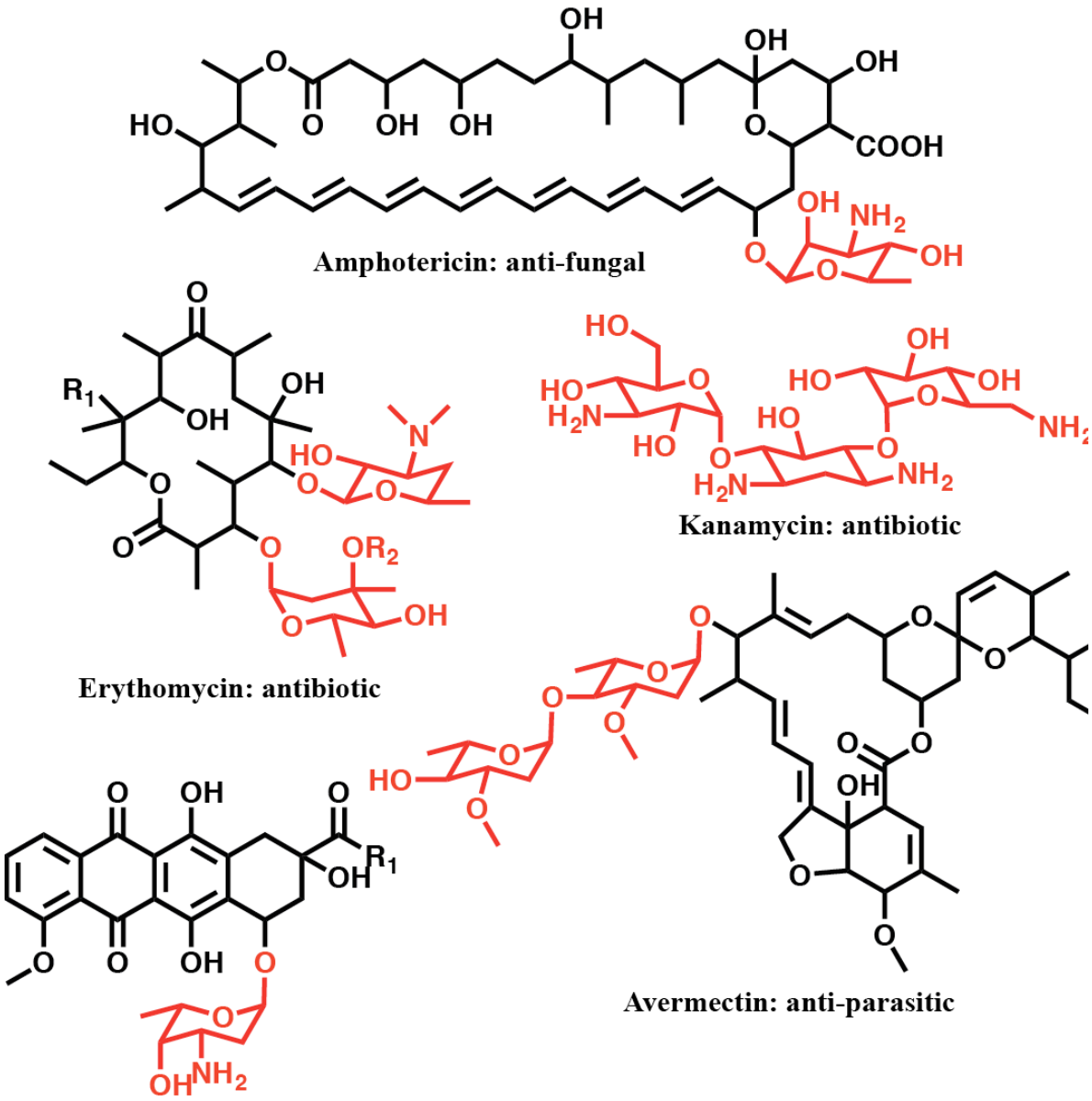

Figure 4.1. Clinically relevant glycosylated small molecules. Many different types of pharmaceuticals are glycosylated. Altering the glycan structure could potentially lead to new and useful therapeutics.

For a typical protein family, sequence conservation can complement structural studies. If a residue is especially important, it is reasonable to think that it will be conserved. However, the GT family is not a typical protein family as it extremely diverse, and there is

very little sequence homology among GT proteins. In fact, glycosyltransferase sequences are ~12% similar to each other, and since there are 20 amino acids, a completely random protein would be expected to be ~7% (~1/20) similar (factoring in the non-random distribution of amino acids – i.e., certain amino acids occur more frequently than others). As a result of low sequence homology, it is impossible to predict the enzymatic properties (such as donor or acceptor specificity) of a GT from primary sequence. Nonetheless, information about the core protein fold and the function are contained within the primary sequence (113). The challenge lies in extracting this information.

Statistical coupling analysis (SCA) is a powerful mathematical tool to extract the information contained within protein sequences (1,114). SCA is based on the premise that (i) pairs of amino acid residues in a protein, interacting functionally or structurally, are under a common selection pressure that should force the co-evolution of those residues, and (ii) this co-evolutionary constraint can be revealed by statistical analysis of a large and diverse multiple sequence alignment (MSA) of the protein family of interest. Typically, SCA performed on various protein families (e.g., GPCRs, PDZ and SH3 domains, chaperones) identifies a sparse but spatially contiguous network of residues, termed a sector, which is functionally relevant (115-117). Indeed, residues that altered substrate specificity of inteins and PDZ domains highly correlated with sector amino acids (118,119). Furthermore, sector surface position distal to an active site of a protein have been demonstrated to act as allosteric hotspots that can be utilized to modulate catalytic function (120,121).

I reasoned that SCA analysis of GTs would identify key residues involved in donor and/or acceptor binding, and could provide valuable insight into the functional properties (i.e. allostery, substrate specificity, donor specificity) of GTs. To achieve this, first I constructed two sequence alignments: one for known members of the GT-A structural class and one for known members of the GT-B structural class. Next by taking advantage of the large number of GT crystal structures, I used PROMALS3D to align a diverse and dissimilar set of sequences. Subsequent SCA analysis of the sequence alignment revealed a sector that contains key residues involved in donor and acceptor binding, along with other residues that appear to be important for catalytic function. I anticipate that this SCA analysis of GTs will lead to novel insights that could not be ascertained by traditional techniques and empower researchers in their studies of this critical class of enzymes.

RESULTS

Protein alignment of glycosyltransferases: constructing order from chaos

The GT family is difficult to align due to the low sequence homology and structural variation. For instance, the GT-B structural family contains no consensus motif, and while the GT-A class contains a "DxD" consensus motif, it is not that well conserved. Indeed a family of GT's in the GT-A class have a CG(Q/M/T) motif in place of DxD (Figure 4.2a). Structurally, the two sequences are similarly positioned and serve similar functions (122). Further complicating the matter, the protein topology within each structural class is not identical. Indeed, various GT's contain insertions and deletions that modify the core structure (Figure 4.2b). Thus an alignment program that utilizes sequence information alone

would fail to align these sequences. Therefore, I decided to utilize PROMALS3D, an alignment program that, in addition to structure prediction, utilizes real 3D structural information from the PDB database to guide MSA construction (123).

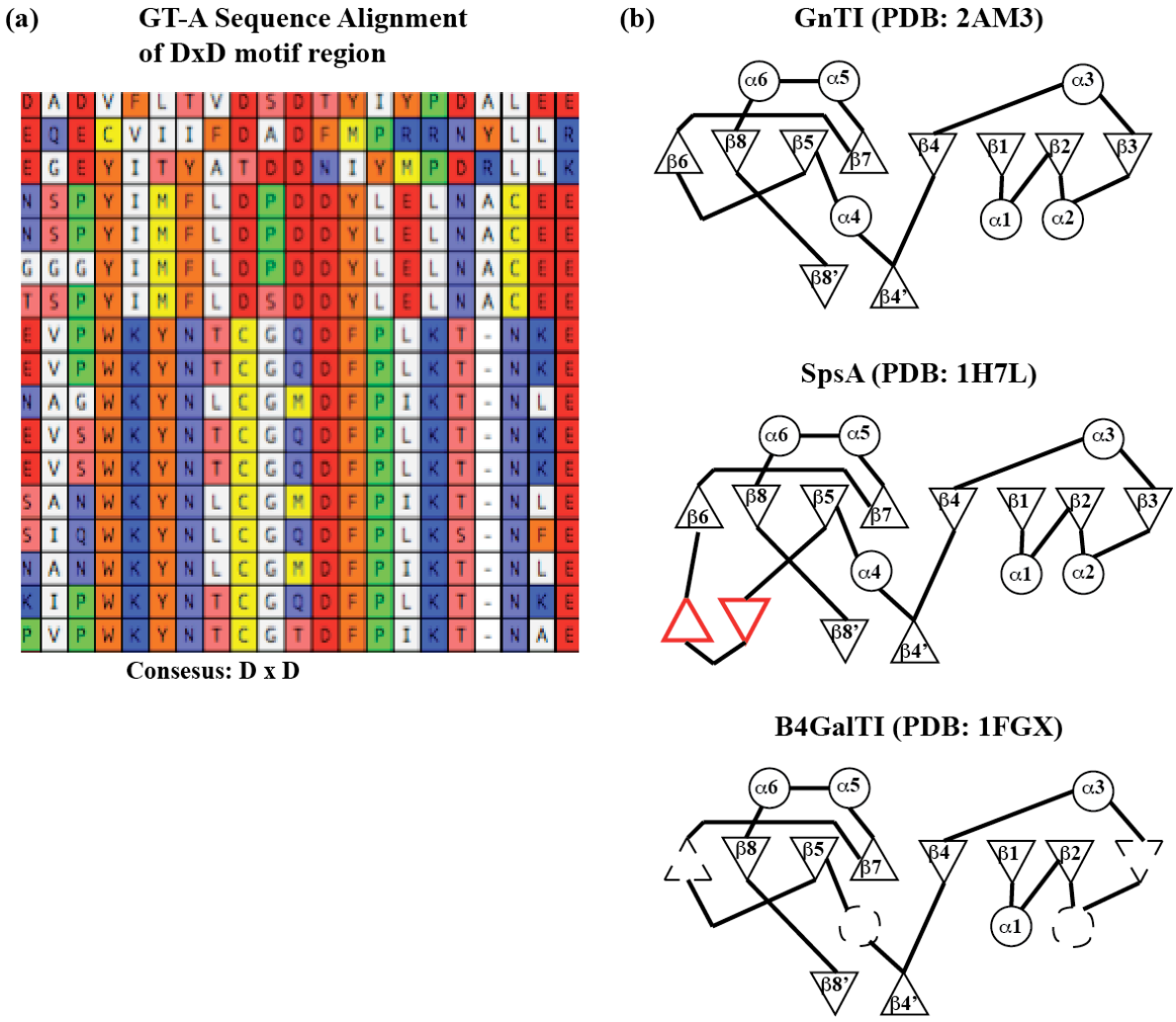

Figure 4.2. GT-A sequence and structural diversity. (a) The GT-A structural class contains the a "DxD" consensus motif that is not that well conserved. A class of GT-A enzymes have CG(Q/M/T) in place of DxD. (b) The core GT-A structure is modified by insertions and deletions in various GT structures.

A high quality alignment would ideally contain a large number of diverse, bona fide GT sequences. One method for gathering said sequences involves performing a BLAST search to find homologous proteins. Unfortunately, the low sequence homology among GT proteins makes it extremely difficult to ascertain if a random sequence with low homology to a GT is itself a GT. Fortunately, the Carbohydrate Active enZYme (CAZy) database classifies every GT sequence into a GT family and annotates each family as either inverting or retaining (45). Also, mercifully the database catalogues structural information on each family and contains a listing of GT's whose activity has been experimentally verified (Figure 4.3). I thus extracted sequences of characterized GTs from families with a structural representative from the CAZy database. Certain families with structural representatives were ignored due to large insertions or large structural variations that precluded their inclusion in the alignment (see methods section for a detailed discussion).

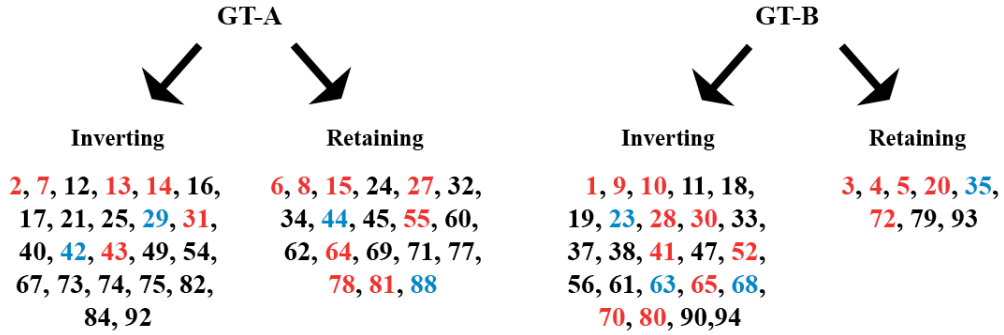

Figure 4.3. GT's are organized into families by the CAZy database. GT proteins are organized into numbered families by self-similarity. Families are classified being in the GT-A or GT-B structural class are shown. Families in red have a structural representative and characterized GT protein sequences were included in the alignment. Families in blue have a structural representative but were excluded due to presence of major structural deviations from the core fold. Families in black do not have a structural representative were excluded from structure based alignment.

When performing an alignment, the input should contain the domain of interest and contain little extraneous sequence. However, sequences acquired from the CAZy database contain the full gene of a protein possessing glycosyltransferase activity, and often, the GT activity arises from a small domain within the protein. The GT domain within large proteins was ascertained using the NCBI conserved domain search (CD-search) and any extraneous sequences were removed (124,125). Overall, great care was taken to be conservative and unbiased when removing GT families and extraneous sequences from the alignment input.

Finally, I decided to align the GT-A and GT-B structural classes separately. While the GT-A and GT-B structural class both contain two rossmann folds, the organization of the rossmann folds is dramatically different (44). In the GT-A class, the rossmann folds are directly joined together (essentially forming one fold), whereas in the GT-B class, the rossmann folds are separate and connected by a flexible linker. Furthermore, the N-terminal rossmann fold of the GT-A class binds to the nucleotide sugar, while the C-terminal rossmann fold binds to the acceptor. The opposite is true for the GT-B structural class. At then end of these pre-processing steps, >450 GT-A and > 450 GT-B characterized protein sequences were aligned using PROMALS3D.

SCA identifies a glycosyltransferase sector containing active site residues

Following the construction of a high quality protein alignment of the GT-A and GT-B classes of GT's, I analyzed the alignments utilizing the SCA Matlab toolbox. S. K. Subramanian, a graduate student in the Ranganathan lab, helped me carry out this analysis. To assess alignment quality, the SCA toolbox removes positions that have less than 80%

occupancy- meaning positions that represent idiosyncratic loops or insertions specific to a small subset of GT families are removed. The size of the remaining truncated alignment was 193 amino acids and 291 amino acids for the GT-A and GT-B classes, respectively. These sizes are reasonable given the size of most glycosyltransferases within each structural class.

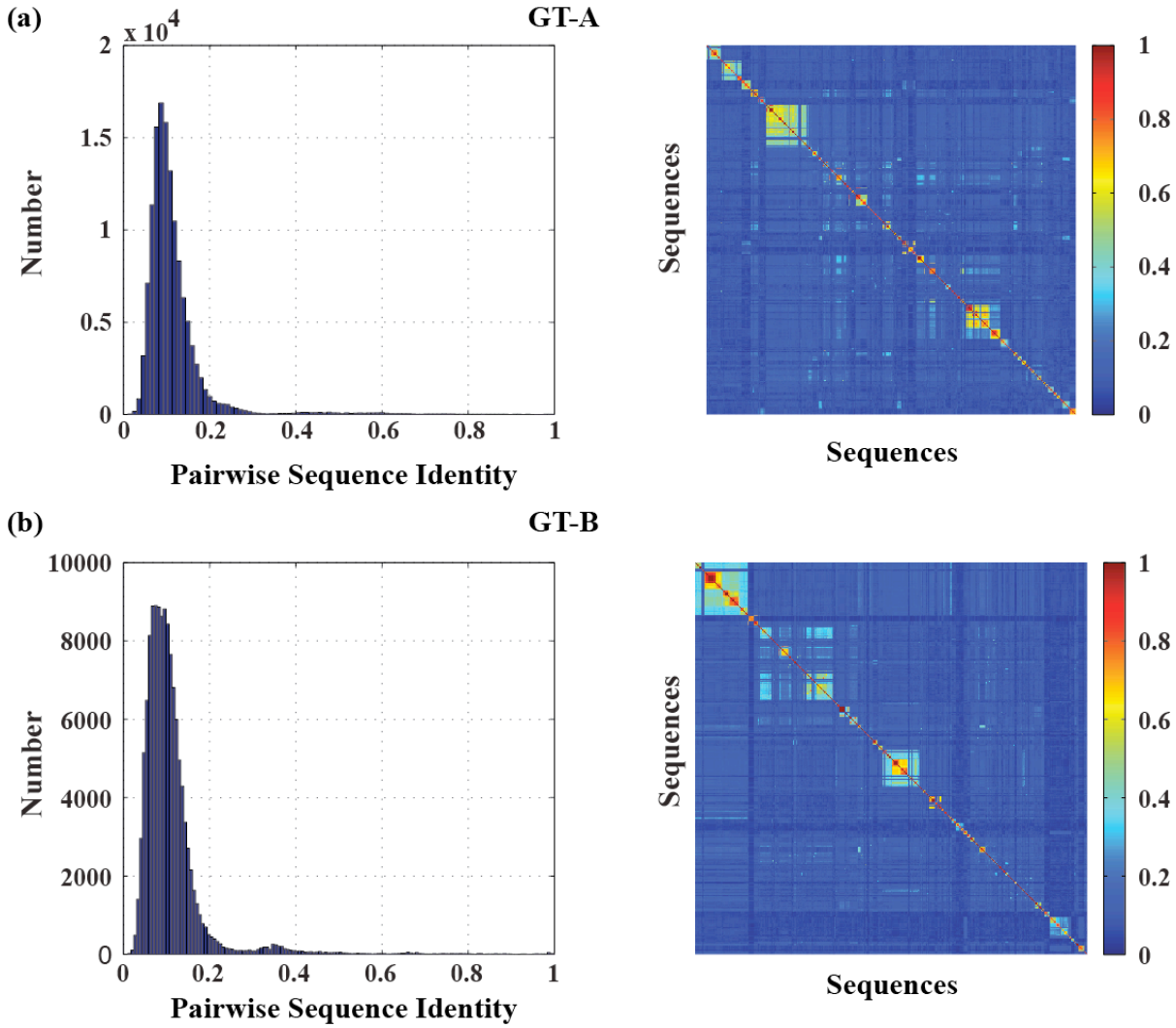

Figure 4.4. Sequence analysis of the alignment of characterized GT protein sequences. (a) Left: pairwise sequence identity between GT-A sequences. Right: heat map visualizing the self similarity among GT-A sequences included in alignment. (b) Left: pairwise sequence identity between GT-B sequences. Right: heat map visualizing the self similarity among GT-B sequences included in alignment.

Next, the diversity of the alignment was confirmed by calculating the self-similarity of the proteins contained within the alignment. Large clades of self-similar proteins can poison SCA analysis. Proteins within each class are ~12% self-similar, and while there are small clades of related proteins, the vast majority of protein sequences are different from each other (Figure 4.4).

Consistent with previous findings, sector residues do not correlate well with positional conservation (Figure 4.5) - meaning there are poorly conserved positions that are very important for protein function (1). Furthermore, examination of sector residues overlaid on top of crystal structures reveals the sectors contain many residues within the active site of

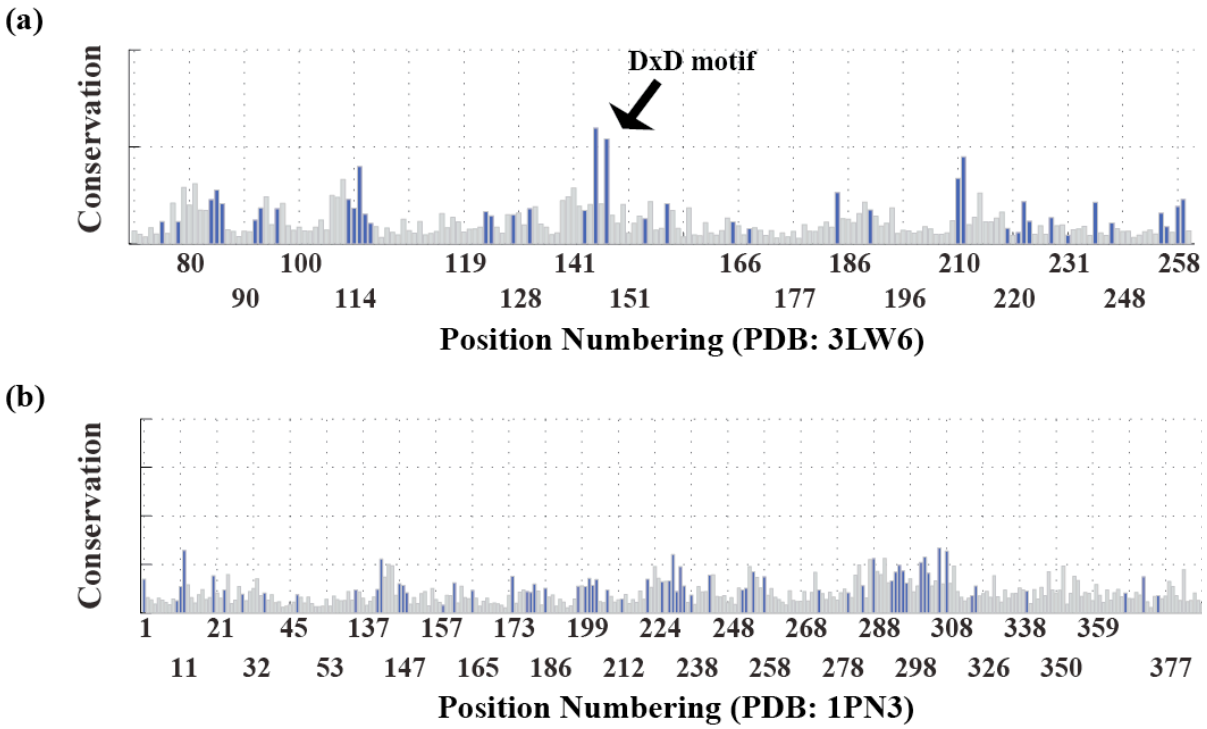

Figure 4.5. Sector positions do not strictly correlate with positional conservation. Calculated as in (1). Sector positions are in blue. (a) GT-A. (b) GT-B.

the protein (Figure 4.6). Importantly, the GT-A sector contains the aspartates of the "DxD" motif, a region known to be crucial for catalytic function. In particular, the GT-A sector contains many of the residues involved in donor nucleotide sugar binding. Conversely, while acceptor binding sites are underrepresented, sector positions were present within a nearby disordered loop thought to be important for acceptor binding (126). Remarkably, the GT-B sector contains many residues involved in donor sugar and acceptor binding (Figure 4.6b). Overall, SCA was able to identify residues that were clearly important for substrate binding. I next sought to understand the relationship between sector positions and catalytic activity.

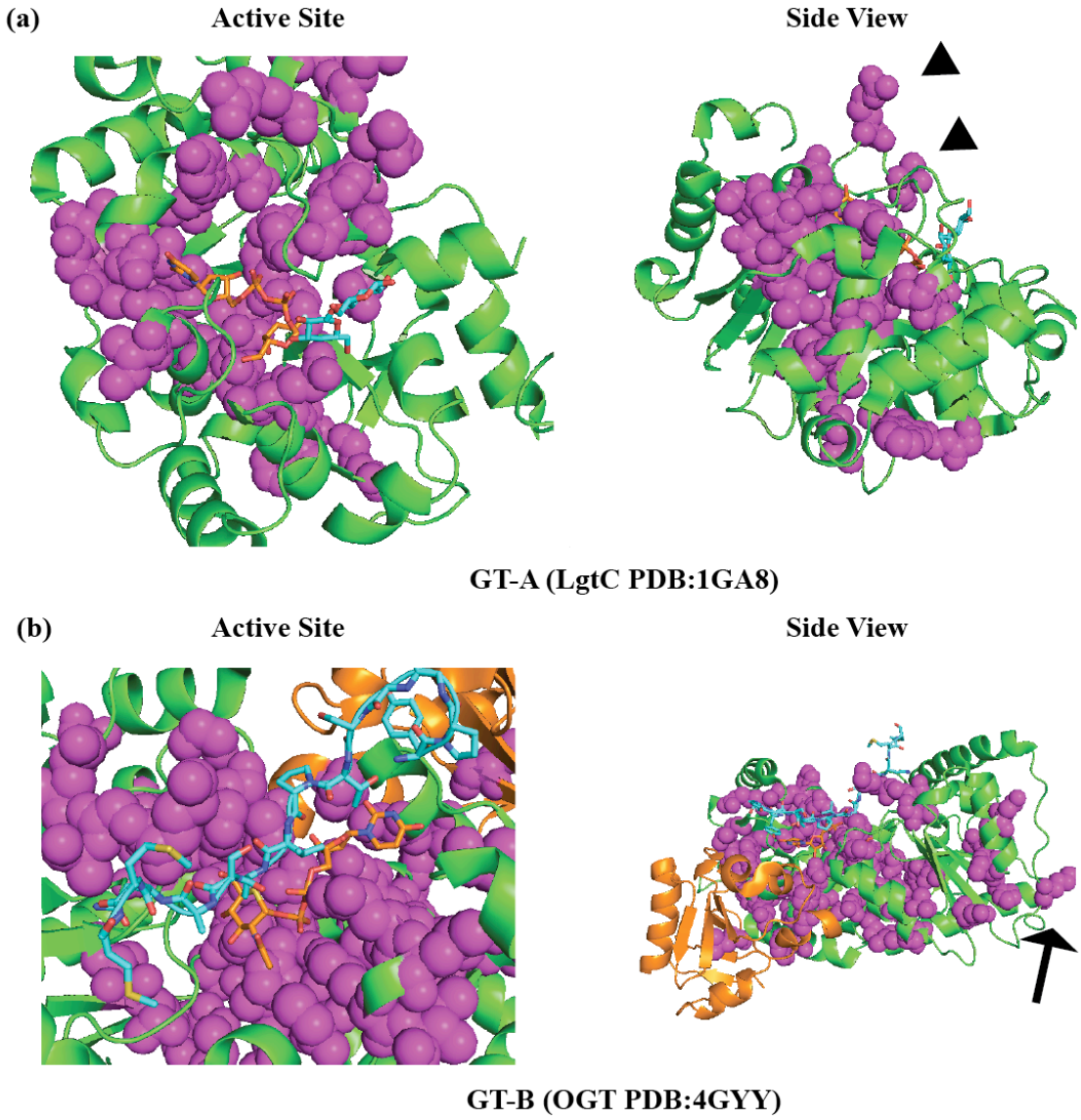

Figure 4.6. Sectors of GT's form a contiguous network of amino acids that include key residues of the active site. (a) Sector positions of a representative member of the GT-A structural class. Arrowheads point to sector residues near a disordered loop that may be important for acceptor binding and release. (b) Sector positions of a representative member of the GT-B structural class. Arrow point to sector residues distal to the active site they may represent a point of allosteric regulation. Green is the GT fold and sector positions are in magenta.

GT sectors and catalytic function

The GT-A enzyme's that control the ABO blood group antigens, GTA and GTB proteins (not to be confused with the GT-A and GT-B structural class) have been the subject of biochemical analysis because they are 354 amino acids in length and are identical except at 4 residues (GTA/GTB: Arg/Gly^{176}, Gly/Ser^{236}, Leu/Met^{266}, Gly/Ala^{268}) (48). This small difference is enough to impart different donor sugar specificities, with the GTA protein preferring UDP-GalNAc and the GTB protein preferring UDP-Gal as the donor sugar, respectively. Position 176 is most distal to the active site but has large effects on catalytic efficiency by effecting the conformational changes that occur during catalysis (126). These conformational changes may be communicated through nearby sector residues. Steric bulk at

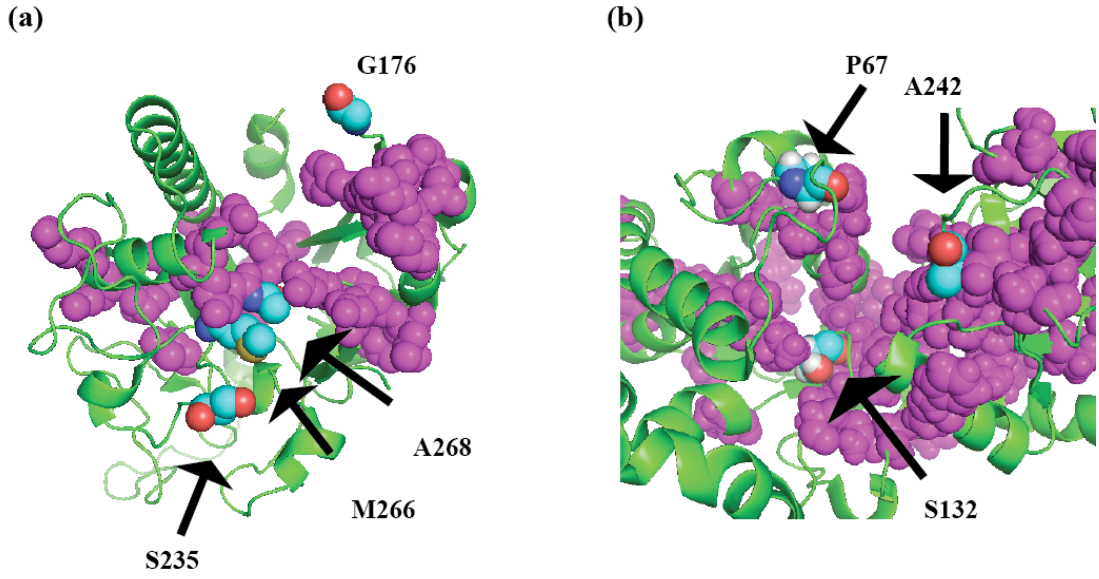

Figure 4.7. Known residues that alter GT catalytic function interact with sector residues. (a) Four residues that alter catalytic activity of the GTB protein (a member of the GT-A structural class) (b) Three residues demonstrated to significantly alter OleD activity and promiscuity interface with interface with sector residues. Green is the GT fold and sector positions are in magenta.

positions M266 and A268 prevent the GTB enzyme from binding the bulkier GalNAc. While these positions are not sector positions, they are directly adjacent to them (Figure 4.7a). Similarly, sector residues interact with critical residues governing catalytic efficiency and substrate specificity in the GT OleD, a GT-B enzyme that catalyzes the glycosylation of the macrolide antibiotic oleandomycin (50,51). Using a high throughput screen, OleD donor sugar specificity and the acceptor were diversified. Careful examination of the protein sector overlaid upon the crystral structure reveals that while the critical residues that allowed for improved catalytic function (P67/S132/A242) are not sector residues, they are all either adjacent or directly interacting with them (Figure 4.7b). This is consistent with previous findings that sectors residues correlate very highly with amino acids undergoing millisecond conformational fluctuations that occur doing enzyme catalysis (121).

To further confirm the importance of sector residues identified, I mined the literature for CDGs caused by mutations that occurred within GT proteins contained within my alignment (3,127-129). These results are summarized in Table 4.1. In one example, mutations in the GT genes responsible for heparan sulfate synthesis, *EXT1,* lead to the disease hereditary multiple exostoses. Many of the reported mutations in *EXT1* are nonsense or frameshift mutations suggesting that the loss of function of this protein is part of the etiology of this disease. This also implies that the missense mutations found in *EXT1* likely impair catalytic function. Indeed, EXT1 constructs carrying the G339D and R340C disease mutation are inactive *in vitro* and very interestingly, G339 is a sector position (130). Taken together, it appears that there is a correlation between sector positions and amino acids that are critical for catalytic function.

Table 4.1. Mutations that cause CDG's are often found in sector positions. All CDG causing mutations found in the GT domain of proteins are listed. Yellow are sector positions. Green are positions that are adjacent to a sector amino acid.

GT-A associated with human disease

GENE	OMIM		Disease	Mutations			
b4GalT7	130070		Ehlers-Danlos Syndrome	R270C	A186D	L206P	
GALNT3	211900		Familial tumoral calcinosis	T272K	T359K	C574G	
COSMC	230430		Tn Syndrome	K152E			
B3GALTL	261540		Peters Plus syndrome	G393E			
DPM1	603503		CDG-Ie	R92G			
EXT1	608177		Multiple Exostoses Type I	A486V	P496L		

GT-B associated with human disease

GENE	OMIM		Disease	Mutations			
EXT1	608177		Multiple Exostoses Type I	Q27K	D146H	G268E	R280G/S
				C298Y	N316S	G339D	R340C/L/H
				R341K	S344F	R346K	E349K
EXT2	211900		Multiple Exostoses Type II	C85R	L152R	R180T	D227N

Sector positions link active site to distal allosteric sites

Upon careful examination of the GT-A and GT-B sectors, sector positions encompassed the active site and traveled through the protein to distal surface regions (Figure 4.6). Given that surface sector positions have been identifying and engineering allosteric sites (115,120), I reasoned that these sites may define allosteric sites on GT proteins. The

EryCIII GT that functions in the synthesis of erythromycin D, an important antibiotic, is functionally inactive without its partner protein EryCII (131). EryCII binds to EryCIII to form a heterodimer, and this binding event stabilizes the dimerization EryCIII to itself. Interestingly, the EryCII:EryCIII binding interface is on the opposite side of the EryCIII:EryCIII dimerization interface. Thus binding of EryCII to EryCIII must initiate conformational changes throughout the protein, which must then be communicated through a series of internal residues. The EryCIII:EryCIII homodimerization interface consists of contacts between A424:A46, H330:T69, V257:A70, A70:S255, and R52:H330, and of these positions, H330 is a sector position and positions A424, A46 and S255 are all nearby sector positions (Figure 4.8b). EryCII binds to an insertion on EryCIII that is not general to the GT-B fold (and thus it is not a part of my alignment) but examination of the sector residues near

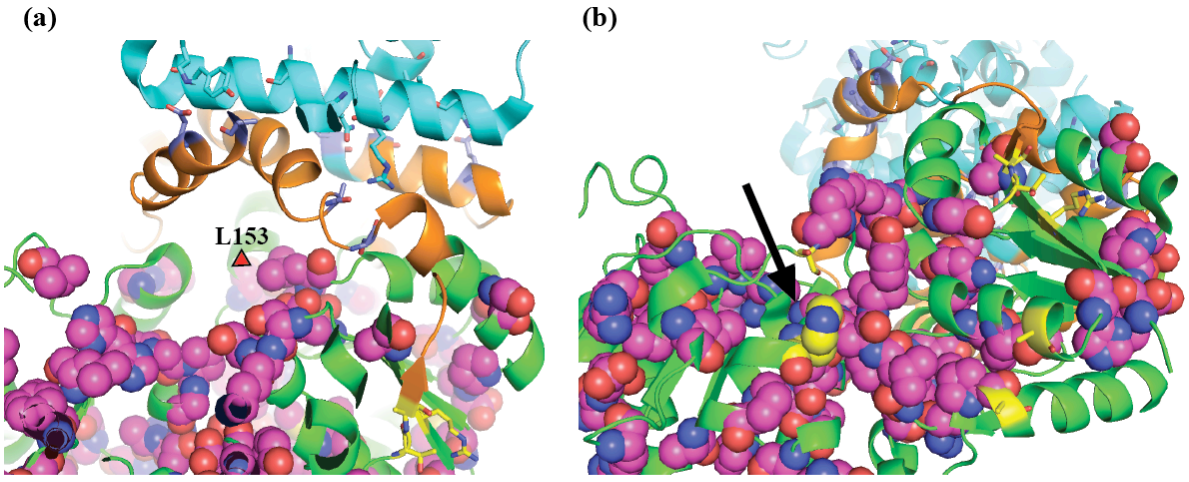

Figure 4.8. Sector positions predict the binding interface of the allosteric activator to the EryCIII GT. (a) EryCII binds to a loop that is not general to the GT-B fold and sector position L153 is nearby residues important for this interaction. (b) The homodimer interface contains sector reside H330 (arrow, yellow space filling) and several positions nearby sector amino acids. PDB: 2YJN. Sector positions are in magenta.

the EryCII:EryCIII binding interface reveals that sector position (L153 PDB:2YJN numbering) interfaces with this insertion (Figure 4.8a), Thus GT sector positions likely define a network of residues that link the active site to distal allosteric sites.

Mutagenesis of the O-GlcNAc transferase OGT sector residues alters catalytic function

The protein OGT is critically important for mammalian life as the gene is required for mammalian embryonic stem cell viability (132). OGT catalyzes the transfer of a single GlcNAc from UDP-GlcNAc onto serines/threonines of proteins to form the intracellular O-GlcNAc modification. OGT has been implicated in many important physiological roles, such as transcription and the cell stress response (133). Furthermore, aberrant OGT activity is thought to contribute to diseases such as cancer. For these reasons, research into OGT and O-glycosylation is quite active. Considering the wealth of data for this protein from multiple labs, I decided to explore the relationship between sector positions and catalytic activity of OGT (134,135). I compiled mutagenesis experiments from multiple labs and determined their effect on protein activity, as well as determined if these residues were sector positions or interacted with sector positions. The results are summarized in Table 4.2. Overall a majority of residues that affect catalytic activity also appear to be sector positions.

Table 4.2. Compilation of OGT mutations and their effect on catalytic activity. Mutations that alter catalytic activity tend to be sector positions. Table adapted from (134-136).Yellow highlights indicates sector positions. GT = glycosyltransferase. ID= intervening domain.

	<10% Activity	10-30% Activity	30-60% Activity	60-100% Activity	Domain	Notes
H558A	x				GT	
R637A	x				GT	A636 and G635 is a sector
Q839N	x				GT	
Y841A				x	GT	L840 and K842 are sectors
K842A	x				GT	
K842M	x				GT	
K898A	x				GT	interacts with sector D925 T92, possibly P559
H901A	x				GT	V902 is a sector
H920A	x				GT	
T921A		x			GT	
D925A	x				GT	poor expression
K981A/K982A				x	GT	not near sector
K742S/K745S/ K747S				x	ID	
K714S/K742S/ K745S/K747S				x	ID	
K706S/K707S/ K742S/K745S/ K747S				x	ID	
Davies Lab						
H558D	x				GT	poor expression
H558E	x				GT	poor expression
Q839E					GT	poor

						expression
H901Y					GT	poor expression see above
Daan van Aalten lab						
D554N			X		GT	interacts with sector H558
Kohler Lab						
M501G				x	GT	
M501V				x	GT	
C917G	x				GT	
C917A			x		GT	
A942G				x	GT	
RMO (Walker Lab)						
L726G	x				ID	Does Not Fold
Q906A	x				GT	
W988A			x		GT	

highlights are sector positions

Careful inspection of the OGT sector reveals that many of the residues within the active site of the protein that are responsible for donor sugar and acceptor peptide binding are sector positions. These sector residues in the active site appear to be contiguous with a network of sector amino acids that connect to the intervening domain (Figure 4.9a). OGT contains a unique insertion called the intervening domain that contacts exclusively with the C-terminal portion of OGT distal to the active site (134). In addition to its unique structure, the intervening domain is highly conserved while its function is completely mysterious. I, along with Rodrigo Meoz-Ortiz, a postdoctoral fellow in the Walker lab, decided to mutate

three sector positions at the C-terminal domain/intervening domain interface (Figure 4.9b) and test their effects on OGT catalytic function. We decided to make the Q906A, W988A, and L726G OGT mutants. Glycine instead of alanine was chosen for mutation at position 726 because a L726A mutation seemed too conservative. Interestingly, all three mutations had an effect on the catalytic function of OGT. The L726G mutant did not fold, the Q906A mutant folded properly but was functionally inactive, and the W988A mutant was less than 60% active (Figure 4.9 c and d). Taken together, these distal OGT sector positions are critical for protein function and functionally connect the intervening domain to the active site.

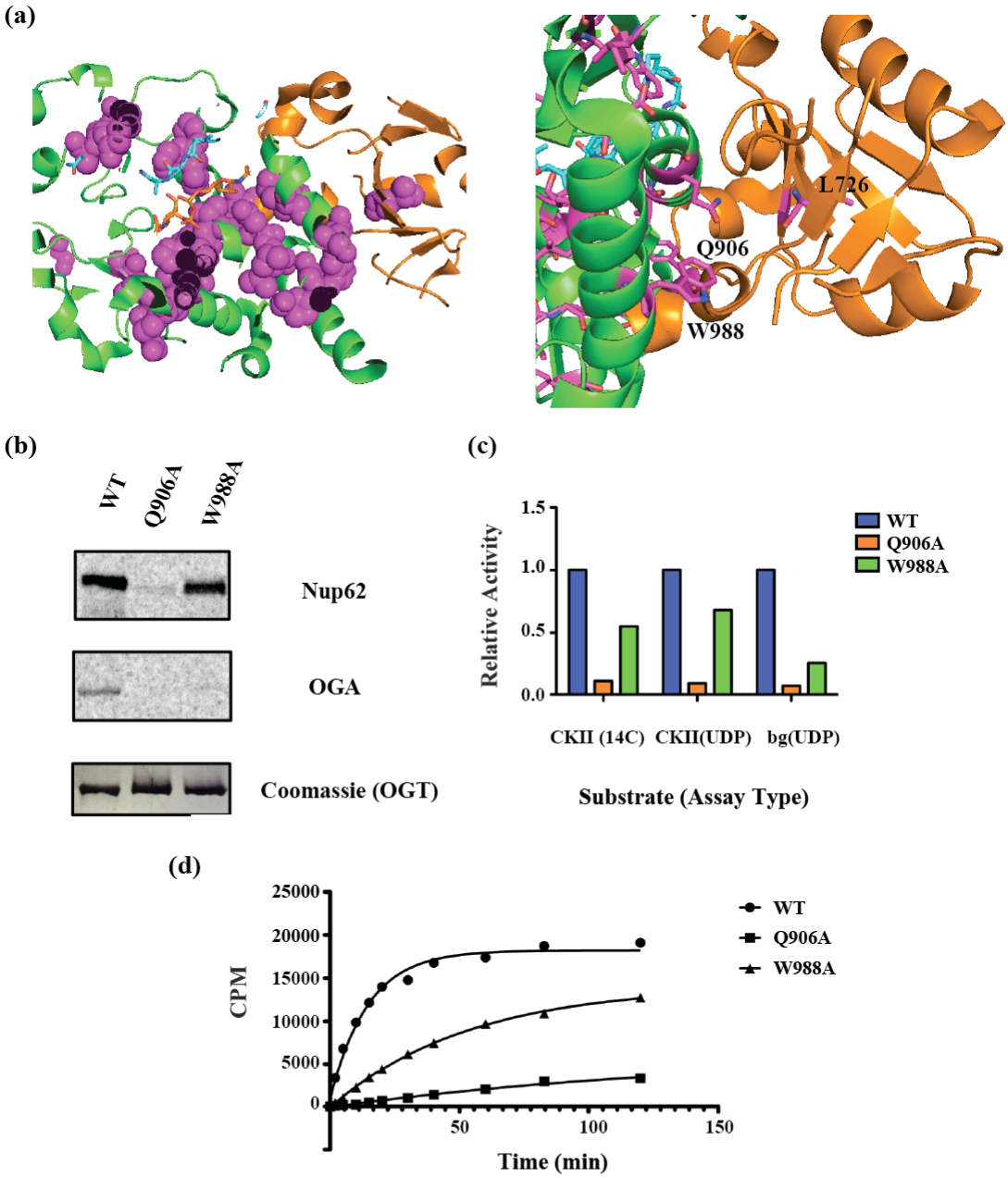

Figure 4.9. OGT sector positions connect the active site of the protein to the intervening domain. (a) Sector amino acids form a contiguous network that associates the donor sugar binding site, acceptor peptide binding site, and the intervening domain. Three sector positions appear to insert into the intervening domain. OGT sector mutants have impaired catalytic function. (b) Endogenous substrates and (c) peptide substrates are poorly O-glycosyled by sector mutants. (14C) refers to radiometric peptide capture assay. (UDP) refers to UDP-Glo assay with CKII and without CKII peptide (bg) (d) Mutants have lower maximal velocity and initial rates as measured by radiometric peptide capture assay.

DISCUSSION

GTs are a part of a large gene family that regulates many normal and abnormal cellular events. Furthermore, GTs are instrumental in the biosynthesis of many clinically relevant small molecules. Taken together, there is great interest in understanding the residues that govern the structural and functional properties of GTs. To this end, I generated a high quality MSA of the GT-A and GT-B protein families and utilized SCA to identify a protein sector that consists of a sparse but contiguous network of amino acids. Comparison of sector residues with mutations that alter GT catalytic function as well as mutations that are associated with CDGs reveal that sector residues appear to be important for substrate binding and enzyme catalysis. Furthermore, sector positions likely correspond to an allosteric network that enables interactions occurring at distal protein surface sites to affect catalysis and vice versa.

Sector residues: candidate sites for altering GT function

One method for engineering GT function involves domain swapping of highly homologous enzymes (137). In this regard, the nearly identical ABO blood group antigen GT's, GTA and GTB, have contributed significant insight into the residues important for donor sugar discrimination and the importance of disordered loops in GT catalysis (138). In another example, functionally important residues were identified in enzymes involved in synthesis of landomycin antibiotics, LanGT1 and LndGT1 (~75% similar). Here Krauth *et al* narrowed important functional regions by swapping domains between the highly homologous regions (137). Once critical regions were identified, individual positions were swapped. Due

to the general low sequence homology characteristic GT proteins, this method is also not generally applicable. Indeed, in the case of the GTA/GTB proteins as well as the lanGT1/LndGT1 protein, there are likely to many residues that are identical between the two proteins that are critical for catalytic function. However, I propose that sector analysis can extend this experimental design to less homologous proteins. Following domain swapping between less homologous proteins, efforts can then be focused on the sector positions within swapped domains.

Another method for determining functionally important residues is high-throughput screening and directed evolution. Indeed, critical residues in the OleD GT were identified by a high-throughput fluorescent screen and following initial identification of three important residues (P67/S132/A242), OleD function was further diversified by employing saturation mutagenesis targeting those positions. This strategy is also not generally applicable because very few GTs are amenable high-thoughput directed evolution screens. However, I propose that sector residues can serve as a screening method to identify catalytically important residues for mutagenesis analysis (118,119). For instance, to achieve altered substrate specificity of a GT, saturation mutagenesis of sector positions can be performed and the ability of the mutant GTs to hydrolyze the novel nucleotide donor sugar can be tested. In the case of OGT, the Kohler lab recently mutated several sites within the active site to increase the promiscuity of OGT for the novel donor sugar UDP-GlcNDAz. Of all the active site residues mutated, the residue that achieved increased promiscuity was a sector position (personal communication) (Table 4.2). Because sector positions make a small portion of the

total protein (~20%) and because they appear to be enriched for functional importance, the number of residues that need to be screened should be significantly reduced.

Surface sector residues: hot spots for engineering allosteric control of GTs

SCA analysis of proteins has revealed allosteric networks in multiple families of proteins. Protein sectors appear to correspond to allosteric network because there appears to be a high correlation between sector residues and residues that undergo microscale conformational fluctuation during protein function. In the Hsp70 family of proteins, residues that experience a large chemical shift upon peptide substrate binding correlate with Hsp70 sector positions (117,139). Furthermore, mutations in sector residues far from the binding site reduce substrate affinity. Indeed, the Ranganthan has leveraged protein sector networks to engineer allosteric control into proteins and demonstrated that surface sector sites are "hot spots" for protein insertions that can regulate catalytic function. I propose that the OGT protein is a natural manifestation of this phenomenon. Here, the intervening domain insertion is closely located to a surface sector site. Mutation of these sector positions at the intervening domain impair catalytic function suggesting that the intervening domain exerts its influence over the OGT active site through sector positions. It should be noted that I was particularly surprised that the W988A mutation was functionally relevant, because it was a surface site that did not seem obviously important from the crystal structure (whereas the other positions mutated were embedded within the protein). Taken together GT function may be controlled by engineering insertions at surface sector positions. For example, a small

molecule binding domain can be inserted at a surface sector site and perhaps the activity of the GT can switched on and off in the presence of the small molecule.

Similarly, I hypothesize that EryCIII contains an insertion that interfaces with the core GT-B sector. Analogous to the OGT intervening domain, I suggest that the EryCII:EryCIII heterodimer interaction communicates through sector position L153 and through the network of sector amino acids to stabilize the EryCIII:EryCIII homodimer interface and enable catalysis. While the interfaces do not exclusively contain sector residues, I suspect that sector amino acids are especially important for these binding events and also for transmitting events that occur at allosteric sites throughout the protein. Identifying novel allosteric surface sites would have widespread impact for the GT family as many GTs are thought to exist as homooligomers and/or heterooligomers (140). Importantly, even in the absence of co-crystal structures, SCA still predicts these interfaces to be important. Thus SCA analysis can be applied to find candidate residues even in the absence of structural data. Furthermore, it may be possible to concentrate mutagenesis efforts to alter the dependency of GT's on dimer formation or mutate sector residues at binding interfaces to create novel GT, GT-binding protein pairs (141). Indeed, this may be generally applicable as many GTs are thought to undergo homo and heterodimerization. This is unsurprising considering that sector analysis reveals general features of a protein family.

Sector positions do not perfectly correlate with critical residues, but it is important to remember that this analysis reveal global properties of a protein family. Every protein is unique and will contain idiosyncratic residues that are specific for the function of that

particular protein. This notwithstanding, these protein specific residues may function by interacting with sectors residues that are a core property of the protein. It is also important to note that SCA obtains information from evolutionary relationships contained with protein alignments. While the sequence alignment was constructed using structural information, the mathematical analysis itself does not take into account structural features. With this in mind, it is somewhat remarkable that SCA was able to identity residues important for catalytic function. Furthermore, while SCA analysis augments existing structural data, it provides useful information on its own.

Future Work: functional classification of GTs through sector positions?

Sectors have identified patterns of residues that can classify proteins. Serine proteases, nuclear hormone receptors, and GPCR's can be classified by their protein sectors. While useful, the CAZy GT family classification reveals very little about what the enzyme actually does. In fact, the vast majority of the genes in the CAZy database are functionally uncharacterized. It would be very interesting to see whether GT sectors can classify GT's by enzymatic properties (i.e. nucleotide sugar bound, type of acceptor). Such a classification system would aid in predicting the function of newly discovered GT proteins. Furthermore, having such a classification system implies that residues could be swapped to predictably alter GT function. For instance, changing a glucose transferase to a galactose transferase. In addition to engineering GT's to increase diversity of natural products and develop new pharmaceuticals, GT engineering would have an impact on the field of metabolic carbohydrate engineering. It would be incredibly useful to engineer GT's to accept novel

sugar analogs and/or engineer GT's to work on specific substrates. SCA has identified protein sectors within the GT family would be a useful tool for future studies of GT's.

In summary, SCA analysis of GT's should aid efforts to engineer GT function and increase glycodiversity. Sector positions may be good candidates to attempt saturation mutagenesis while screening for novel GT activity. Furthermore, it may be possible to take advantage of the allosteric "hot spots" identified by SCA to engineer points of control in GT's.

ACKNOWLEDGEMENTS

I would like to give special thanks to S.K. Subramanian for his advice on performing SCA analysis. I also appreciate the many scientific discussions we've had over the years. Dustin Schaeffer helped me think critically about the bioinformatics behind acquiring GT sequences and aligning them. I also would like to thank Rodrigo for his enthusiasm and his willingness to make and test the OGT mutants. I also would like to thank Jen for letting me spend time on a project that is completely unrelated to anything else going on in the lab. It was a lot of fun working on this project and I really appreciate the support and encouragement.

METHODS

Acquiring and aligning glycosyltransferase sequences from the CAZy database

The CAZy database organizes glycosyltransferases by GT family and annotates families that have a structural representative. Furthermore, the CAZy database annotates GT's as characterized if they been biochemically verified to have GT activity. I wrote a

python script (Appendix) to query the database and acquire the GenBank accession number of all characterized GT's in families with a structural representative (Figure 4.3). From the GT-A class, I excluded families 29 and 42 because they are structural variants of the GT-A class and have differences in significant differences in topology. Families 44 and 88 were also excluded because they contained large intra-domain insertions that could precluded alignment. For the GT-B class, families 23 and 68 were topological variants that could not be aligned onto the core GT-B structure. Family 63 was a unique family with a single known member. Family 35 consisted of glycogen phosphorylase family and while the active site of this protein resembles that of a classic GT-B enzyme, the protein is much larger than a typical GT-B and contains many large intradomain insertions.

Often multiple accessions are listed for a single GT. In these cases, I acquired the accession that was in bold font. I acquired GT-A sequences on 2012-06-21. I acquired GT-B sequences on 2013-01-30. I then imported theses GenBank accessions into Matlab and acquired protein information using the getgenpept command. Certain accessions were outdated and threw a warning upon protein sequence acquisition. When this happened I manually altered the accession to another accession listed on the CAZy database that did not throw an error. With all the sequences acquired, I then submitted the sequences for analysis using the CD-Search webserver to look for the amino acid sequence that contained the GT domain. I deleted sequence information that did not contain the GT-domain. Often, CD-search was only able to find a fragment of the full GT-domain. This was an unsurprising result considering the low sequence homology of the family. In these cases I deleted sequences outside a range of amino acids that was reasonable for each structural class (~350

aa for GT-A and 550 for GT-B). Additionally, certain genes contained multiple GT-domains – and sometimes not within the same fold (a GT-A and GT-B within one protein). In the case of two domains with the same fold (GT-A), I split the domains and kept them in the alignment as long as both GT domains belonged to a structurally characterized GT family. When the domains were from different structural folds, if the each domain were a member of a structurally characterized GT family then it would be pulled twice during my search.

I submitted the sequence list to the PROMALS3D webserver. I did not use the default settings. I made three changes in the advanced settings tab. I aligned homologs and input structures utilizing the DaliLite service in addition to FAST and TM-align. I also utilized PROMALS to perform the first alignment stage of closely homologous sequences. Furthermore, certain crystal structures contained extraneous domains. PROMALS3D allows for manually input structures. For the GT-A class, I included the GT-A domain of PDB accession 4FIX and 4HG6. 2Z86 and 2Z87 contained two GT-A domains and I split the PDB to each domain separately. I included the domain that contained the ligand in the binding pocket. For the GT-B class, I included only the GT domain of 3S28 and 4GYW. The latter case is the PDB for OGT, here I removed the TPR repeat and the intervening domain. At the end, I submitted 515 GT-A sequences and 507 GT-B sequences for alignment.

Statistical coupling analysis of glycotransferase protein alignments

Statistical coupling analysis (SCA) was performed using an updated version of MATLAB SCA toolbox (SCA Toolbox 5.0) using a process described in Halabi *et al* 2009.

The SCA toolbox and thecode for computing the correlation matrix and definition of sectors through spectral decomposition are available at the ranganathan laboratory web site (http://systems.swmed.edu/rr_lab/).

Briefly, the basic analytic procedure involves three steps: (1) compute a conservation-weighted correlation matrix that estimates the degree of coevolution between each pair of sequence positions in the MSA, (2) carry out spectral (or eigenvalue) decomposition of this matrix and identify its statistically significant top eigenmodes, and (3) examine the pattern of positional contributions to these top eigenmodes to deduce the number and composition of groups of coevolving amino acid positions (sectors) (121). In both families, the spectral analysis indicates three significant top eigenmodes that represent a single protein sector. Consistent with the basic principle of sectors defined I Halabi *et al*, the GT-A and GT-B sector comprises a relatively small subset of total amino acid positions (~20% based on the cutoffs indicated) and comprises a physically contiguous network of residues in the tertiary structure (Figure 4.6).

OGT mutant protein production and purificatoin

OGT protein was produced as previously described (134,142). Briefly plasmids were transformed in BL21 (DE3) cells and grown at 37° C to an O.D. = 1.1-1.3. The culture was cooled to 16° C and induced with 0.2 mM IPTG. After 24 hours, the cells were pelleted, and frozen at –80° C. Pellets were lysed with Bugbuster® (50 mL/L culture) supplemented with rLysozme®, Benzonase® (Novagen), and 40 mM imidazole, pH = 7.7. The lysate was clarified by centrifugation at 40,000 x g, then loaded onto Ni^{2+}-charged IDA column. The

column was washed with three column volumes of 50 mM imidazole, pH = 7.5, 250 mM NaCl, and the protein was eluted with 250 mM imidazole, pH = 7.5, 250 mM NaCl. The elute fraction was dialyzed against 20 mM phosphate, pH = 7.5, 150 mM NaCl, and 1 mM EDTA, then supplemented with 500 μM tris(hydroxypropyl)phosphine) and stored at -80° C.

OGT activity assays

Mutants were made from the full-length ncOGT using QuickChange mutagenesis and the primers:

L726A

5’- CAGGTCGATACCGTTACCAACGATACGGTTGTC-3’

3'-GACAACCGTATCGTTGGTAACGGTATCGACCTG-5’

Q906A

5’- GCAAACGTCAGCCAGAGCACCACGACGAACGTG-3’

3’-CACGTTCGTCGTGGTGCTCTGGCTGACGTTTGC-5’

W988A

5’-GGAGATACGCTGTTTAGCAACTTTACCACGAAC-3’

3’-GTTCGTGGTAAAGTTGCTAAACAGCGTATCTCC-5’

ncOGT activity on full length protein Nup62 and OGA were performed using ^{14}C UDP-GlcNAc at a concentration of 12 uM, substrate at a concentration of 1 uM, OGT at a concentration of 250 nM, and buffer (125 mM NaCl, 1 mM EDTA, 20 mM potassium phosphate, pH 7.4, and 500 μM tris(hydroxypropyl)phosphine). Components were incubated together at 1hr at 37° C. Proteins were then run on an SDS page gel, dried, exposed to a

phosphoimager cassette overnight, and imaged with a Typhoon instrument. Gel was stained with coomassie blue to demonstrate equal OGT loading.

ncOGT kinetic activity was determined utilizing a previously reported radiometric peptide capture assay (gross). Briefly, mixtures containing 500 μM CKII peptide (KKKYPGGSTPVSSANMM), 6 μM UDP-^{14}C-GlcNAc, 250 nM OGT (WT or mutant protein), and buffer (125 mM NaCl, 1 mM EDTA, 20 mM potassium phosphate, pH 7.4, and 500 μM tris(hydroxypropyl)phosphine) were incubated at 37° C indicated times. Reactions were then quenched by spotting onto the Whatman P81 phosphocellulose disks, washed three times for five minutes in 0.5% phosphoric acid, and counted by liquid scintillation counting.

Alternatively, ncOGT kinetics were determined by UDP-Glo assay (Promega) using the same concentrations OGT and UDP-GlcNAc (non-radioactive) as above. These experiments were done with and without CKII peptide. Experiments without CKII peptide measure amount of abortive hydrolysis.

APPENDIX A
Files for Statistical Coupling Analysis of Glycosyltransferase Sequences

Alignments.zip

Contains input sequences for PROMALS3D alignment. Also, contains non-truncated output alignments.

SCA_matlab.zip

Contains pertinent matlab files for performing SCA analysis on alignments of glycosyltransferase sequences.

BIBLIOGRAPHY

1. Halabi, N., Rivoire, O., Leibler, S., and Ranganathan, R. (2009) Protein sectors: evolutionary units of three-dimensional structure. *Cell* **138**, 774-786
2. (2009). in *Essentials of Glycobiology* (Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., and Etzler, M. E. eds.), 2nd Ed., Cold Spring Harbor (NY). pp
3. Freeze, H. H., and Ng, B. G. (2011) Golgi glycosylation and human inherited diseases. *Cold Spring Harbor perspectives in biology* **3**, a005371
4. Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97-130
5. Angata, T., and Varki, A. (2002) Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. *Chemical reviews* **102**, 439-469
6. Schauer, R. (2009) Sialic acids as regulators of molecular and cellular interactions. *Current opinion in structural biology* **19**, 507-514
7. Phillips, M. L., Nudelman, E., Gaeta, F. C., Perez, M., Singhal, A. K., Hakomori, S., and Paulson, J. C. (1990) ELAM-1 mediates cell adhesion by recognition of a carbohydrate ligand, sialyl-Lex. *Science* **250**, 1130-1132
8. Somers, W. S., Tang, J., Shaw, G. D., and Camphausen, R. T. (2000) Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe(X) and PSGL-1. *Cell* **103**, 467-479
9. de Vries, R. P., Zhu, X., McBride, R., Rigter, A., Hanson, A., Zhong, G., Hatta, M., Xu, R., Yu, W., Kawaoka, Y., de Haan, C. A., Wilson, I. A., and Paulson, J. C. (2014) Hemagglutinin receptor specificity and structural analyses of respiratory droplet-transmissible H5N1 viruses. *Journal of virology* **88**, 768-773
10. Rutishauser, U., Acheson, A., Hall, A. K., Mann, D. M., and Sunshine, J. (1988) The neural cell adhesion molecule (NCAM) as a regulator of cell-cell interactions. *Science* **240**, 53-57
11. Tanner, M. E. (2005) The enzymes of sialic acid biosynthesis. *Bioorganic chemistry* **33**, 216-228
12. Schwarzkopf, M., Knobeloch, K. P., Rohde, E., Hinderlich, S., Wiechens, N., Lucka, L., Horak, I., Reutter, W., and Horstkorte, R. (2002) Sialylation is essential for early development in mice. *Proc Natl Acad Sci U S A* **99**, 5267-5270
13. Verheijen, F. W., Verbeek, E., Aula, N., Beerens, C. E., Havelaar, A. C., Joosse, M., Peltonen, L., Aula, P., Galjaard, H., van der Spek, P. J., and Mancini, G. M. (1999) A new gene, encoding an anion transporter, is mutated in sialic acid storage diseases. *Nature genetics* **23**, 462-465
14. Tiralongo, J., Fujita, A., Sato, C., Kitajima, K., Lehmann, F., Oschlies, M., Gerardy-Schahn, R., and Munster-Kuhnel, A. K. (2007) The rainbow trout CMP-sialic acid synthetase utilises a nuclear localization signal different from that identified in the mouse enzyme. *Glycobiology* **17**, 945-954
15. Munster, A. K., Weinhold, B., Gotza, B., Muhlenhoff, M., Frosch, M., and Gerardy-Schahn, R. (2002) Nuclear localization signal of murine CMP-Neu5Ac

synthetase includes residues required for both nuclear targeting and enzymatic activity. *J Biol Chem* **277**, 19688-19696

16. Audry, M., Jeanneau, C., Imberty, A., Harduin-Lepers, A., Delannoy, P., and Breton, C. (2011) Current trends in the structure-activity relationships of sialyltransferases. *Glycobiology* **21**, 716-726
17. Seppala, R., Tietze, F., Krasnewich, D., Weiss, P., Ashwell, G., Barsh, G., Thomas, G. H., Packman, S., and Gahl, W. A. (1991) Sialic acid metabolism in sialuria fibroblasts. *J Biol Chem* **266**, 7456-7461
18. Varki, A., Hooshmand, F., Diaz, S., Varki, N. M., and Hedrick, S. M. (1991) Developmental abnormalities in transgenic mice expressing a sialic acid-specific 9-O-acetylesterase. *Cell* **65**, 65-74
19. Surolia, I., Pirnie, S. P., Chellappa, V., Taylor, K. N., Cariappa, A., Moya, J., Liu, H., Bell, D. W., Driscoll, D. R., Diederichs, S., Haider, K., Netravali, I., Le, S., Elia, R., Dow, E., Lee, A., Freudenberg, J., De Jager, P. L., Chretien, Y., Varki, A., MacDonald, M. E., Gillis, T., Behrens, T. W., Bloch, D., Collier, D., Korzenik, J., Podolsky, D. K., Hafler, D., Murali, M., Sands, B., Stone, J. H., Gregersen, P. K., and Pillai, S. (2010) Functionally defective germline variants of sialic acid acetylesterase in autoimmunity. *Nature* **466**, 243-247
20. Swindall, A. F., Londono-Joshi, A. I., Schultz, M. J., Fineberg, N., Buchsbaum, D. J., and Bellis, S. L. (2013) ST6Gal-I protein expression is upregulated in human epithelial tumors and correlates with stem cell markers in normal tissues and colon cancer cell lines. *Cancer research* **73**, 2368-2378
21. Swindall, A. F., and Bellis, S. L. (2011) Sialylation of the Fas death receptor by ST6Gal-I provides protection against Fas-mediated apoptosis in colon carcinoma cells. *J Biol Chem* **286**, 22982-22990
22. Bos, P. D., Zhang, X. H., Nadal, C., Shu, W., Gomis, R. R., Nguyen, D. X., Minn, A. J., van de Vijver, M. J., Gerald, W. L., Foekens, J. A., and Massague, J. (2009) Genes that mediate breast cancer metastasis to the brain. *Nature* **459**, 1005-1009
23. Miyagi, T., Takahashi, K., Hata, K., Shiozaki, K., and Yamaguchi, K. (2012) Sialidase significance for cancer progression. *Glycoconjugate journal* **29**, 567-577
24. Leroy, J. G., Seppala, R., Huizing, M., Dacremont, G., De Simpel, H., Van Coster, R. N., Orvisky, E., Krasnewich, D. M., and Gahl, W. A. (2001) Dominant inheritance of sialuria, an inborn error of feedback inhibition. *American journal of human genetics* **68**, 1419-1427
25. Nishino, I., Carrillo-Carrasco, N., and Argov, Z. (2014) GNE myopathy: current update and future therapy. *Journal of neurology, neurosurgery, and psychiatry*
26. Celeste, F. V., Vilboux, T., Ciccone, C., de Dios, J. K., Malicdan, M. C., Leoyklang, P., McKew, J. C., Gahl, W. A., Carrillo-Carrasco, N., and Huizing, M. (2014) Mutation Update for GNE Gene Variants Associated with GNE Myopathy. *Human mutation*
27. Rutishauser, C., and Flammer, J. (1988) Retests in static perimetry. *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie* **226**, 75-77

28. Penner, J., Mantey, L. R., Elgavish, S., Ghaderi, D., Cirak, S., Berger, M., Krause, S., Lucka, L., Voit, T., Mitrani-Rosenbaum, S., and Hinderlich, S. (2006) Influence of UDP-GlcNAc 2-epimerase/ManNAc kinase mutant proteins on hereditary inclusion body myopathy. *Biochemistry* **45**, 2968-2977
29. Hinderlich, S., Salama, I., Eisenberg, I., Potikha, T., Mantey, L. R., Yarema, K. J., Horstkorte, R., Argov, Z., Sadeh, M., Reutter, W., and Mitrani-Rosenbaum, S. (2004) The homozygous M712T mutation of UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase results in reduced enzyme activities but not in altered overall cellular sialylation in hereditary inclusion body myopathy. *FEBS Lett* **566**, 105-109
30. Salama, I., Hinderlich, S., Shlomai, Z., Eisenberg, I., Krause, S., Yarema, K., Argov, Z., Lochmuller, H., Reutter, W., Dabby, R., Sadeh, M., Ben-Bassat, H., and Mitrani-Rosenbaum, S. (2005) No overall hyposialylation in hereditary inclusion body myopathy myoblasts carrying the homozygous M712T GNE mutation. *Biochem Biophys Res Commun* **328**, 221-226
31. Saito, F., Tomimitsu, H., Arai, K., Nakai, S., Kanda, T., Shimizu, T., Mizusawa, H., and Matsumura, K. (2004) A Japanese patient with distal myopathy with rimmed vacuoles: missense mutations in the epimerase domain of the UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase (GNE) gene accompanied by hyposialylation of skeletal muscle glycoproteins. *Neuromuscular disorders : NMD* **14**, 158-161
32. Ricci, E., Broccolini, A., Gidaro, T., Morosetti, R., Gliubizzi, C., Frusciante, R., Di Lella, G. M., Tonali, P. A., and Mirabella, M. (2006) NCAM is hyposialylated in hereditary inclusion body myopathy due to GNE mutations. *Neurology* **66**, 755-758
33. Huizing, M., Rakocevic, G., Sparks, S. E., Mamali, I., Shatunov, A., Goldfarb, L., Krasnewich, D., Gahl, W. A., and Dalakas, M. C. (2004) Hypoglycosylation of alpha-dystroglycan in patients with hereditary IBM due to GNE mutations. *Mol Genet Metab* **81**, 196-202
34. Dube, D. H., and Bertozzi, C. R. (2003) Metabolic oligosaccharide engineering as a tool for glycobiology. *Current opinion in chemical biology* **7**, 616-625
35. Du, J., Meledeo, M. A., Wang, Z., Khanna, H. S., Paruchuri, V. D., and Yarema, K. J. (2009) Metabolic glycoengineering: sialic acid and beyond. *Glycobiology* **19**, 1382-1401
36. Jacobs, C. L., Goon, S., Yarema, K. J., Hinderlich, S., Hang, H. C., Chai, D. H., and Bertozzi, C. R. (2001) Substrate specificity of the sialic acid biosynthetic pathway. *Biochemistry* **40**, 12864-12874
37. Luchansky, S. J., Goon, S., and Bertozzi, C. R. (2004) Expanding the diversity of unnatural cell-surface sialic acids. *Chembiochem : a European journal of chemical biology* **5**, 371-374
38. Mathew, M. P., Tan, E., Shah, S., Bhattacharya, R., Adam Meledeo, M., Huang, J., Espinoza, F. A., and Yarema, K. J. (2012) Extracellular and intracellular esterase processing of SCFA-hexosamine analogs: implications for metabolic

glycoengineering and drug delivery. *Bioorganic & medicinal chemistry letters* **22**, 6929-6933

39. Luchansky, S. J., and Bertozzi, C. R. (2004) Azido sialic acids can modulate cell-surface interactions. *Chembiochem : a European journal of chemical biology* **5**, 1706-1709
40. Bond, M. R., Zhang, H., Vu, P. D., and Kohler, J. J. (2009) Photocrosslinking of glycoconjugates using metabolically incorporated diazirine-containing sugars. *Nat Protoc* **4**, 1044-1063
41. Feng, L., Hong, S., Rong, J., You, Q., Dai, P., Huang, R., Tan, Y., Hong, W., Xie, C., Zhao, J., and Chen, X. (2013) Bifunctional Unnatural Sialic Acids for Dual Metabolic Labeling of Cell-Surface Sialylated Glycans. *Journal of the American Chemical Society* **135**, 9244-9247
42. Han, S., Collins, B. E., Bengtson, P., and Paulson, J. C. (2005) Homomultimeric complexes of CD22 in B cells revealed by protein-glycan cross-linking. *Nat Chem Biol* **1**, 93-97
43. Ramya, T. N., Weerapana, E., Liao, L., Zeng, Y., Tateno, H., Yates, J. R., 3rd, Cravatt, B. F., and Paulson, J. C. (2010) In situ trans ligands of CD22 identified by glycan-protein photocross-linking-enabled proteomics. *Mol Cell Proteomics* **9**, 1339-1351
44. Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008) Glycosyltransferases: structures, functions, and mechanisms. *Annual review of biochemistry* **77**, 521-555
45. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic acids research* **37**, D233-238
46. Werz, D. B., Ranzinger, R., Herget, S., Adibekian, A., von der Lieth, C. W., and Seeberger, P. H. (2007) Exploring the structural diversity of mammalian carbohydrates ("glycospace") by statistical databank analysis. *ACS chemical biology* **2**, 685-691
47. Breton, C., Fournel-Gigleux, S., and Palcic, M. M. (2012) Recent structures, evolution and mechanisms of glycosyltransferases. *Current opinion in structural biology* **22**, 540-549
48. Patenaude, S. I., Seto, N. O., Borisova, S. N., Szpacenko, A., Marcus, S. L., Palcic, M. M., and Evans, S. V. (2002) The structural basis for specificity in human ABO(H) blood group biosynthesis. *Nature structural biology* **9**, 685-690
49. Alfaro, J. A., Zheng, R. B., Persson, M., Letts, J. A., Polakowski, R., Bai, Y., Borisova, S. N., Seto, N. O., Lowary, T. L., Palcic, M. M., and Evans, S. V. (2008) ABO(H) blood group A and B glycosyltransferases recognize substrate via specific conformational changes. *J Biol Chem* **283**, 10097-10108
50. Williams, G. J., and Thorson, J. S. (2008) A high-throughput fluorescence-based glycosyltransferase screen and its application in directed evolution. *Nat Protoc* **3**, 357-362

51. Williams, G. J., Goff, R. D., Zhang, C., and Thorson, J. S. (2008) Optimizing glycosyltransferase specificity via "hot spot" saturation mutagenesis presents a catalyst for novobiocin glycorandomization. *Chemistry & biology* **15**, 393-401
52. Luzhetskyy, A., Mendez, C., Salas, J. A., and Bechthold, A. (2008) Glycosyltransferases, important tools for drug design. *Current topics in medicinal chemistry* **8**, 680-709
53. Malicdan, M. C., Noguchi, S., Nonaka, I., Hayashi, Y. K., and Nishino, I. (2007) A Gne knockout mouse expressing human GNE D176V mutation develops features similar to distal myopathy with rimmed vacuoles or hereditary inclusion body myopathy. *Hum Mol Genet* **16**, 2669-2682
54. Malicdan, M. C., Noguchi, S., Hayashi, Y. K., Nonaka, I., and Nishino, I. (2009) Prophylactic treatment with sialic acid metabolites precludes the development of the myopathic phenotype in the DMRV-hIBM mouse model. *Nat Med* **15**, 690-695
55. Wang, Z. V., Deng, Y., Gao, N., Pedrozo, Z., Li, D. L., Morales, C. R., Criollo, A., Luo, X., Tan, W., Jiang, N., Lehrman, M. A., Rothermel, B. A., Lee, A. H., Lavandero, S., Mammen, P. P., Ferdous, A., Gillette, T. G., Scherer, P. E., and Hill, J. A. (2014) Spliced X-box binding protein 1 couples the unfolded protein response to hexosamine biosynthetic pathway. *Cell* **156**, 1179-1192
56. Denzel, M. S., Storm, N. J., Gutschmidt, A., Baddi, R., Hinze, Y., Jarosch, E., Sommer, T., Hoppe, T., and Antebi, A. (2014) Hexosamine pathway metabolites enhance protein quality control and prolong life. *Cell* **156**, 1167-1178
57. Cheung, P., Pawling, J., Partridge, E. A., Sukhu, B., Grynpas, M., and Dennis, J. W. (2007) Metabolic homeostasis and tissue renewal are dependent on beta1,6GlcNAc-branched N-glycans. *Glycobiology* **17**, 828-837
58. Keppler, O. T., Hinderlich, S., Langner, J., Schwartz-Albiez, R., Reutter, W., and Pawlita, M. (1999) UDP-GlcNAc 2-epimerase: a regulator of cell surface sialylation. *Science* **284**, 1372-1376
59. Galuska, S. P., Geyer, H., Weinhold, B., Kontou, M., Rohrich, R. C., Bernard, U., Gerardy-Schahn, R., Reutter, W., Munster-Kuhnel, A., and Geyer, R. (2010) Quantification of nucleotide-activated sialic acids by a combination of reduction and fluorescent labeling. *Analytical chemistry* **82**, 4591-4598
60. Kontou, M., Bauer, C., Reutter, W., and Horstkorte, R. (2008) Sialic acid metabolism is involved in the regulation of gene expression during neuronal differentiation of PC12 cells. *Glycoconjugate journal* **25**, 237-244
61. Weidemann, W., Klukas, C., Klein, A., Simm, A., Schreiber, F., and Horstkorte, R. (2010) Lessons from GNE-deficient embryonic stem cells: sialic acid biosynthesis is involved in proliferation and gene expression. *Glycobiology* **20**, 107-117
62. Tomiya, N., Ailor, E., Lawrence, S. M., Betenbaugh, M. J., and Lee, Y. C. (2001) Determination of nucleotides and sugar nucleotides involved in protein glycosylation by high-performance anion-exchange chromatography: sugar nucleotide contents in cultured insect cells and mammalian cells. *Analytical biochemistry* **293**, 129-137

63. Zachara, N. E., O'Donnell, N., Cheung, W. D., Mercer, J. J., Marth, J. D., and Hart, G. W. (2004) Dynamic O-GlcNAc modification of nucleocytoplasmic proteins in response to stress. A survival response of mammalian cells. *J Biol Chem* **279**, 30133-30142
64. Kang, J. G., Park, S. Y., Ji, S., Jang, I., Park, S., Kim, H. S., Kim, S. M., Yook, J. I., Park, Y. I., Roth, J., and Cho, J. W. (2009) O-GlcNAc protein modification in cancer cells increases in response to glucose deprivation through glycogen degradation. *J Biol Chem* **284**, 34777-34784
65. Tu, B. P., Mohler, R. E., Liu, J. C., Dombek, K. M., Young, E. T., Synovec, R. E., and McKnight, S. L. (2007) Cyclic changes in metabolic state during the life of a yeast cell. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16886-16891
66. Lau, K. S., Partridge, E. A., Grigorian, A., Silvescu, C. I., Reinhold, V. N., Demetriou, M., and Dennis, J. W. (2007) Complex N-glycan number and degree of branching cooperate to regulate cell proliferation and differentiation. *Cell* **129**, 123-134
67. Ohtsubo, K., Takamatsu, S., Minowa, M. T., Yoshida, A., Takeuchi, M., and Marth, J. D. (2005) Dietary and genetic control of glucose transporter 2 glycosylation promotes insulin secretion in suppressing diabetes. *Cell* **123**, 1307-1321
68. Morgan, R., Gao, G., Pawling, J., Dennis, J. W., Demetriou, M., and Li, B. (2004) N-acetylglucosaminyltransferase V (Mgat5)-mediated N-glycosylation negatively regulates Th1 cytokine production by T cells. *Journal of immunology* **173**, 7200-7208
69. Wellen, K. E., Lu, C., Mancuso, A., Lemons, J. M., Ryczko, M., Dennis, J. W., Rabinowitz, J. D., Coller, H. A., and Thompson, C. B. (2010) The hexosamine biosynthetic pathway couples growth factor-induced glutamine uptake to glucose metabolism. *Genes & development* **24**, 2784-2799
70. Patnaik, S. K., Potvin, B., Carlsson, S., Sturm, D., Leffler, H., and Stanley, P. (2006) Complex N-glycans are the major ligands for galectin-1, -3, and -8 on Chinese hamster ovary cells. *Glycobiology* **16**, 305-317
71. Brewer, C. F., Miceli, M. C., and Baum, L. G. (2002) Clusters, bundles, arrays and lattices: novel mechanisms for lectin-saccharide-mediated cellular interactions. *Current opinion in structural biology* **12**, 616-623
72. Delacour, D., Greb, C., Koch, A., Salomonsson, E., Leffler, H., Le Bivic, A., and Jacob, R. (2007) Apical sorting by galectin-3-dependent glycoprotein clustering. *Traffic* **8**, 379-388
73. Perillo, N. L., Pace, K. E., Seilhamer, J. J., and Baum, L. G. (1995) Apoptosis of T cells mediated by galectin-1. *Nature* **378**, 736-739
74. Earl, L. A., and Baum, L. G. (2008) CD45 glycosylation controls T-cell life and death. *Immunology and cell biology* **86**, 608-615
75. Earl, L. A., Bi, S., and Baum, L. G. (2010) N- and O-glycans modulate galectin-1 binding, CD45 signaling, and T cell death. *J Biol Chem* **285**, 2232-2244

76. Coster, A. C., Govers, R., and James, D. E. (2004) Insulin stimulates the entry of GLUT4 into the endosomal recycling pathway by a quantal mechanism. *Traffic* **5**, 763-771
77. Weindruch, R., Walford, R. L., Fligiel, S., and Guthrie, D. (1986) The retardation of aging in mice by dietary restriction: longevity, cancer, immunity and lifetime energy intake. *The Journal of nutrition* **116**, 641-654
78. Colman, R. J., Beasley, T. M., Kemnitz, J. W., Johnson, S. C., Weindruch, R., and Anderson, R. M. (2014) Caloric restriction reduces age-related and all-cause mortality in rhesus monkeys. *Nature communications* **5**, 3557
79. Shiraha, H., Gupta, K., Drabik, K., and Wells, A. (2000) Aging fibroblasts present reduced epidermal growth factor (EGF) responsiveness due to preferential loss of EGF receptors. *J Biol Chem* **275**, 19343-19351
80. Kuro-o, M. (2009) Klotho and aging. *Biochimica et biophysica acta* **1790**, 1049-1058
81. Kuro-o, M., Matsumura, Y., Aizawa, H., Kawaguchi, H., Suga, T., Utsugi, T., Ohyama, Y., Kurabayashi, M., Kaname, T., Kume, E., Iwasaki, H., Iida, A., Shiraki-Iida, T., Nishikawa, S., Nagai, R., and Nabeshima, Y. I. (1997) Mutation of the mouse klotho gene leads to a syndrome resembling ageing. *Nature* **390**, 45-51
82. Cha, S. K., Ortega, B., Kurosu, H., Rosenblatt, K. P., Kuro, O. M., and Huang, C. L. (2008) Removal of sialic acid involving Klotho causes cell-surface retention of TRPV5 channel via binding to galectin-1. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 9805-9810
83. Mali, P., Esvelt, K. M., and Church, G. M. (2013) Cas9 as a versatile tool for engineering biology. *Nature methods* **10**, 957-963
84. Needham, M., Corbett, A., Day, T., Christiansen, F., Fabian, V., and Mastaglia, F. L. (2008) Prevalence of sporadic inclusion body myositis and factors contributing to delayed diagnosis. *J Clin Neurosci* **15**, 1350-1353
85. Needham, M., and Mastaglia, F. L. (2008) Sporadic inclusion body myositis: a continuing puzzle. *Neuromuscular disorders : NMD* **18**, 6-16
86. Dalakas, M. C. (2006) Sporadic inclusion body myositis--diagnosis, pathogenesis and therapeutic strategies. *Nat Clin Pract Neurol* **2**, 437-447
87. Wilson, F. C., Ytterberg, S. R., St Sauver, J. L., and Reed, A. M. (2008) Epidemiology of sporadic inclusion body myositis and polymyositis in Olmsted County, Minnesota. *J Rheumatol* **35**, 445-447
88. Salmon, P., and Trono, D. (2007) Production and titration of lentiviral vectors. *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]* **Chapter 12**, Unit 12 10
89. Bond, M. R., Zhang, H., Kim, J., Yu, S. H., Yang, F., Patrie, S. M., and Kohler, J. J. (2011) Metabolism of diazirine-modified N-acetylmannosamine analogues to photo-cross-linking sialosides. *Bioconjugate chemistry* **22**, 1811-1823
90. Dettmer, K., Nurnberger, N., Kaspar, H., Gruber, M. A., Almstetter, M. F., and Oefner, P. J. (2011) Metabolite extraction from adherently growing mammalian

cells for metabolomics studies: optimization of harvesting and extraction protocols. *Analytical and bioanalytical chemistry* **399**, 1127-1139

91. Yu, S. H., Boyce, M., Wands, A. M., Bond, M. R., Bertozzi, C. R., and Kohler, J. J. (2012) Metabolic labeling enables selective photocrosslinking of O-GlcNAc-modified proteins to their binding partners. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 4834-4839
92. Laxman, S., Sutter, B. M., Wu, X., Kumar, S., Guo, X., Trudgian, D. C., Mirzaei, H., and Tu, B. P. (2013) Sulfur amino acids regulate translational capacity and metabolic homeostasis through modulation of tRNA thiolation. *Cell* **154**, 416-429
93. Varki, A. (2008) Sialic acids in human health and disease. *Trends in molecular medicine* **14**, 351-360
94. Pham, N. D., Parker, R. B., and Kohler, J. J. (2013) Photocrosslinking approaches to interactome mapping. *Current opinion in chemical biology* **17**, 90-101
95. Tanaka, Y., Bond, M. R., and Kohler, J. J. (2008) Photocrosslinkers illuminate interactions in living cells. *Mol Biosyst* **4**, 473-480
96. Jones, M. B., Teng, H., Rhee, J. K., Lahar, N., Baskaran, G., and Yarema, K. J. (2004) Characterization of the cellular uptake and metabolic conversion of acetylated N-acetylmannosamine (ManNAc) analogues to sialic acids. *Biotechnology and bioengineering* **85**, 394-405
97. Bond, M. R., Zhang, H., Vu, P. D., and Kohler, J. J. (2009) Photocrosslinking of glycoconjugates using metabolically incorporated diazirine-containing sugars. *Nat Protoc* **4**, 1044-1063
98. Jacobs, C. L., Yarema, K. J., Mahal, L. K., Nauman, D. A., Charters, N. W., and Bertozzi, C. R. (2000) Metabolic labeling of glycoproteins with chemical tags through unnatural sialic acid biosynthesis. *Methods in enzymology* **327**, 260-275
99. Yu, S. H., Bond, M. R., Whitman, C. M., and Kohler, J. J. (2010) Metabolic labeling of glycoconjugates with photocrosslinking sugars. *Methods in enzymology* **478**, 541-562
100. Gross, H. J., and Brossmer, R. (1995) Enzymatic transfer of sialic acids modified at C-5 employing four different sialyltransferases. *Glycoconjugate journal* **12**, 739-746
101. Gross, H. J., Rose, U., Krause, J. M., Paulson, J. C., Schmid, K., Feeney, R. E., and Brossmer, R. (1989) Transfer of synthetic sialic acid analogues to N- and O-linked glycoprotein glycans using four different mammalian sialyltransferases. *Biochemistry* **28**, 7386-7392
102. Staudinger, J. L., Xu, C., Cui, Y. J., and Klaassen, C. D. (2010) Nuclear receptor-mediated regulation of carboxylesterase expression and activity. *Expert opinion on drug metabolism & toxicology* **6**, 261-271
103. Sarkar, A. K., Fritz, T. A., Taylor, W. H., and Esko, J. D. (1995) Disaccharide uptake and priming in animal cells: inhibition of sialyl Lewis X by acetylated Gal beta 1-->4GlcNAc beta-O-naphthalenemethanol. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 3323-3327
104. Sarkar, A. K., Brown, J. R., and Esko, J. D. (2000) Synthesis and glycan priming activity of acetylated disaccharides. *Carbohydrate research* **329**, 287-300

105. Hawley, S. A., Fullerton, M. D., Ross, F. A., Schertzer, J. D., Chevtzoff, C., Walker, K. J., Peggie, M. W., Zibrova, D., Green, K. A., Mustard, K. J., Kemp, B. E., Sakamoto, K., Steinberg, G. R., and Hardie, D. G. (2012) The ancient drug salicylate directly activates AMP-activated protein kinase. *Science* **336**, 918-922
106. Tian, L., Yang, Y., Wysocki, L. M., Arnold, A. C., Hu, A., Ravichandran, B., Sternson, S. M., Looger, L. L., and Lavis, L. D. (2012) Selective esterase-ester pair for targeting small molecules with cellular specificity. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 4756-4761
107. Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M., and Deisseroth, K. (2011) Optogenetics in neural systems. *Neuron* **71**, 9-34
108. Zhan, X., Shi, X., Zhang, Z., Chen, Y., and Wu, J. I. (2011) Dual role of Brg chromatin remodeling factor in Sonic hedgehog signaling during neural development. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 12758-12763
109. Hashimoto, K., Tokimatsu, T., Kawano, S., Yoshizawa, A. C., Okuda, S., Goto, S., and Kanehisa, M. (2009) Comprehensive analysis of glycosyltransferases in eukaryotic genomes for structural and functional characterization of glycans. *Carbohydrate research* **344**, 881-887
110. Chang, A., Singh, S., Phillips, G. N., Jr., and Thorson, J. S. (2011) Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. *Current opinion in biotechnology* **22**, 800-808
111. Thibodeaux, C. J., Melancon, C. E., and Liu, H. W. (2007) Unusual sugar biosynthesis and natural product glycodiversification. *Nature* **446**, 1008-1016
112. Gomez, H., Polyak, I., Thiel, W., Lluch, J. M., and Masgrau, L. (2012) Retaining glycosyltransferase mechanism studied by QM/MM methods: lipopolysaccharyl-alpha-1,4-galactosyltransferase C transfers alpha-galactose via an oxocarbenium ion-like transition state. *Journal of the American Chemical Society* **134**, 4743-4752
113. Socolich, M., Lockless, S. W., Russ, W. P., Lee, H., Gardner, K. H., and Ranganathan, R. (2005) Evolutionary information for specifying a protein fold. *Nature* **437**, 512-518
114. Lockless, S. W., and Ranganathan, R. (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295-299
115. Shulman, A. I., Larson, C., Mangelsdorf, D. J., and Ranganathan, R. (2004) Structural determinants of allosteric ligand activation in RXR heterodimers. *Cell* **116**, 417-429
116. Suel, G. M., Lockless, S. W., Wall, M. A., and Ranganathan, R. (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nature structural biology* **10**, 59-69
117. Smock, R. G., Rivoire, O., Russ, W. P., Swain, J. F., Leibler, S., Ranganathan, R., and Gierasch, L. M. (2010) An interdomain sector mediating allostery in Hsp70 molecular chaperones. *Molecular systems biology* **6**, 414

118. Lockless, S. W., and Muir, T. W. (2009) Traceless protein splicing utilizing evolved split inteins. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 10999-11004
119. McLaughlin, R. N., Jr., Poelwijk, F. J., Raman, A., Gosal, W. S., and Ranganathan, R. (2012) The spatial architecture of protein function and adaptation. *Nature* **491**, 138-142
120. Lee, J., Natarajan, M., Nashine, V. C., Socolich, M., Vo, T., Russ, W. P., Benkovic, S. J., and Ranganathan, R. (2008) Surface sites for engineering allosteric control in proteins. *Science* **322**, 438-442
121. Reynolds, K. A., McLaughlin, R. N., and Ranganathan, R. (2011) Hot spots for allosteric regulation on protein surfaces. *Cell* **147**, 1564-1575
122. Pak, J. E., Arnoux, P., Zhou, S., Sivarajah, P., Satkunarajah, M., Xing, X., and Rini, J. M. (2006) X-ray crystal structure of leukocyte type core 2 beta1,6-N-acetylglucosaminyltransferase. Evidence for a convergence of metal ion-independent glycosyltransferase mechanism. *J Biol Chem* **281**, 26693-26701
123. Pei, J., Tang, M., and Grishin, N. V. (2008) PROMALS3D web server for accurate multiple protein sequence and structure alignments. *Nucleic acids research* **36**, W30-34
124. Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Lu, F., Marchler, G. H., Mullokandov, M., Omelchenko, M. V., Robertson, C. L., Song, J. S., Thanki, N., Yamashita, R. A., Zhang, D., Zhang, N., Zheng, C., and Bryant, S. H. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic acids research* **39**, D225-229
125. Marchler-Bauer, A., Anderson, J. B., Derbyshire, M. K., DeWeese-Scott, C., Gonzales, N. R., Gwadz, M., Hao, L., He, S., Hurwitz, D. I., Jackson, J. D., Ke, Z., Krylov, D., Lanczycki, C. J., Liebert, C. A., Liu, C., Lu, F., Lu, S., Marchler, G. H., Mullokandov, M., Song, J. S., Thanki, N., Yamashita, R. A., Yin, J. J., Zhang, D., and Bryant, S. H. (2007) CDD: a conserved domain database for interactive domain family analysis. *Nucleic acids research* **35**, D237-240
126. Yazer, M. H., and Palcic, M. M. (2005) The importance of disordered loops in ABO glycosyltransferases. *Transfusion medicine reviews* **19**, 210-216
127. Haeuptle, M. A., and Hennet, T. (2009) Congenital disorders of glycosylation: an update on defects affecting the biosynthesis of dolichol-linked oligosaccharides. *Human mutation* **30**, 1628-1641
128. Wuyts, W., and Van Hul, W. (2000) Molecular basis of multiple exostoses: mutations in the EXT1 and EXT2 genes. *Human mutation* **15**, 220-227
129. Zak, B. M., Crawford, B. E., and Esko, J. D. (2002) Hereditary multiple exostoses and heparan sulfate polymerization. *Biochimica et biophysica acta* **1573**, 346-355
130. McCormick, C., Leduc, Y., Martindale, D., Mattison, K., Esford, L. E., Dyer, A. P., and Tufaro, F. (1998) The putative tumour suppressor EXT1 alters the expression of cell-surface heparan sulfate. *Nature genetics* **19**, 158-161

131. Moncrieffe, M. C., Fernandez, M. J., Spiteller, D., Matsumura, H., Gay, N. J., Luisi, B. F., and Leadlay, P. F. (2012) Structure of the glycosyltransferase EryCIII in complex with its activating P450 homologue EryCII. *Journal of molecular biology* **415**, 92-101
132. Shafi, R., Iyer, S. P., Ellies, L. G., O'Donnell, N., Marek, K. W., Chui, D., Hart, G. W., and Marth, J. D. (2000) The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5735-5739
133. Zachara, N. E., Molina, H., Wong, K. Y., Pandey, A., and Hart, G. W. (2011) The dynamic stress-induced "O-GlcNAc-ome" highlights functions for O-GlcNAc in regulating DNA damage/repair and other cellular pathways. *Amino acids* **40**, 793-808
134. Lazarus, M. B., Nam, Y., Jiang, J., Sliz, P., and Walker, S. (2011) Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564-567
135. Clarke, A. J., Hurtado-Guerrero, R., Pathak, S., Schuttelkopf, A. W., Borodkin, V., Shepherd, S. M., Ibrahim, A. F., and van Aalten, D. M. (2008) Structural insights into mechanism and specificity of O-GlcNAc transferase. *The EMBO journal* **27**, 2780-2788
136. Martinez-Fleites, C., Macauley, M. S., He, Y., Shen, D. L., Vocadlo, D. J., and Davies, G. J. (2008) Structure of an O-GlcNAc transferase homolog provides insight into intracellular glycosylation. *Nature structural & molecular biology* **15**, 764-765
137. Krauth, C., Fedoryshyn, M., Schleberger, C., Luzhetskyy, A., and Bechthold, A. (2009) Engineering a function into a glycosyltransferase. *Chemistry & biology* **16**, 28-35
138. Seto, N. O., Palcic, M. M., Compston, C. A., Li, H., Bundle, D. R., and Narang, S. A. (1997) Sequential interchange of four amino acids from blood group B to blood group A glycosyltransferase boosts catalytic activity and progressively modifies substrate recognition in human recombinant enzymes. *J Biol Chem* **272**, 14133-14138
139. Swain, J. F., Schulz, E. G., and Gierasch, L. M. (2006) Direct comparison of a stable isolated Hsp70 substrate-binding domain in the empty and substrate-bound states. *J Biol Chem* **281**, 1605-1611
140. Seko, A., and Yamashita, K. (2005) Characterization of a novel galactose beta1,3-N-acetylglucosaminyltransferase (beta3Gn-T8): the complex formation of beta3Gn-T2 and beta3Gn-T8 enhances enzymatic activity. *Glycobiology* **15**, 943-951
141. Skerker, J. M., Perchuk, B. S., Siryaporn, A., Lubin, E. A., Ashenberg, O., Goulian, M., and Laub, M. T. (2008) Rewiring the specificity of two-component signal transduction systems. *Cell* **133**, 1043-1054

142. Gross, B. J., Kraybill, B. C., and Walker, S. (2005) Discovery of O-GlcNAc transferase inhibitors. *Journal of the American Chemical Society* **127**, 14588-14589