Engineering an Orthogonal O-Glycosyltransferase and Donor Sugar Pair

by

Lauren Jan Sarbo Wagner

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Carolyn R. Bertozzi, Chair
Professor Matthew B. Francis
Professor Gerard Marriott

Spring 2015

Engineering an Orthogonal O-Glycosyltransferase and Donor Sugar Pair

© 2015

By Lauren Jan Sarbo Wagner

Abstract

Engineering an Orthogonal O-Glycosyltransferase and Donor Sugar Pair

by

Lauren Jan Sarbo Wagner

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Carolyn R. Bertozzi, Chair

Despite its role in fundamental biological processes, GalNAc-type O-linked glycosylation has proven extremely difficult to study and remains one of the most poorly understood post-translational modifications. The ppGalNAcT family catalyzes the first committed step in O-glycosylation, the installation of GalNAc from the donor UDP-GalNAc onto serine and threonine (Ser/Thr) residues of acceptor polypeptides. Thus, the substrate specificity of the 20 human ppGalNAcT family members determines which proteins are O-glycosylated. Progress in the field is hampered by a limited understanding of ppGalNAcT protein substrate specificity and of the regulation underlying O-glycosylation dynamics.

The individual ppGalNAcTs display distinct spatial and temporal expression patterns during mammalian and fly development, which may reflect their unique roles governing cell-specific glycosylation. *In vitro* studies have shown that ppGalNAcT family members have both unique and redundant peptide substrate specificities, although validation of these results *in vivo* has met with limited success. Efforts to isolate the function of individual isoforms by genetic knockdown demonstrate unpredictable functional redundancy among family members— ablation of several ppGalNAcTs in mice yields only partially penetrant phenotypes, while five ppGalNAcTs are essential for *Drosophila* viability. Chemically targeting one isoform with inhibitors or substrate analogs is challenging, as members of the ppGalNAcT family demonstrate a high level of structural homology and all utilize the nucleotide sugar UDP-GalNAc as a donor substrate. All current efforts to monitor the activity of individual ppGalNAcTs *in vivo* lack an unambiguous method that can confirm whether a target protein is naturally glycosylated by a specific ppGalNAcT in a biological context.

Towards this end, we have rationally designed an individual family member, ppGalNAcT2, with an enlarged binding pocket and a UDP-GalNAc analog with a "bump" and a chemical handle. This bump-hole pair is designed to uniquely label the protein substrates of ppGalNAcT, ultimately enabling proteomic analysis. The pair that we have developed behaves orthogonally, meaning that the ppGalNAcT2 mutant and the UDP-GalNAc analog, UDP-GalNAzMe(S), demonstrate little reactivity

1

with the native system. UDP-GalNAzMe(S) is poorly accepted by native ppGalNAcTs *in vitro* and has minimal background *in vivo*. This pair can glycosylate and tag a known acceptor substrate of ppGalNAcT2, enabling downstream labeling with an affinity handle for enrichment and analysis.

As described in this dissertation, our goal was to develop a ppGalNAcT and UDP-GalNAc analog pair for identification of ppGalNAcT protein substrates. Chapter 1 describes the background for this project including the biological functions of O-glycosylation, the ppGalNAcT family, and recent progress in the field. It then discusses the utility of a bump-hole strategy and significant work using bump-hole engineering in other protein families. Chapters 2 and 3 describe the enzyme and substrate panels generated for this research. Chapter 2 explores the development of a panel of ppGalNAcT mutants with enlarged binding pockets and a mammalian expression system to secrete transferase soluble domains. Chapter 3 is focused on the chemoenzymatic synthesis of a panel of UDP-GalNAc analogs with an azide chemical handle.

Ongoing work is described in Chapter 4. We report the identification of a promising bump hole pair from *in vitro* screens with a model peptide substrate. Preliminary work demonstrating that UDP-GalNAzMe(S) is orthogonal to native ppGalNAcTs in cell lysates is also discussed. We report progress delivering UDP-GalNAzMe(S) to cells, as well as further confirmation that the analog is orthogonal to native ppGalNAcTs in cells. Ongoing work is delineated at the end of Chapter 4. We establish an ongoing collaboration with the Gerken lab at Case Western Reserve University in which a peptide library is being used to confirm that the bump-hole pair conserves ppGalNAcT substrate specificity. We also describe a strategy to measure the K_M and V_{MAX} for the candidate bump-hole pair with a model peptide substrate. Chapter 4 concludes with future directions for this research in living systems.

This dissertation is dedicated to my family, who have always supported me, and to my husband, who has stood by my side throughout this entire process.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank many people for making this possible.

I would like to thank Carolyn Bertozzi for being a constant supporter of this project. Despite the many challenges that I have faced, she has always been enthusiastic about my progress and she has encouraged me to set my sights high. I am very grateful that she has encouraged other people to continue this work, including the two current post-docs who will carry the torch, Marjoke Debets and Junwon Choi. I would also like to thank Carolyn for sharing the art of brilliant presentation with everyone in the lab. I have learned a great deal about the power of conciseness and simplicity from Carolyn's example.

I would like to thank everyone in the Bertozzi lab throughout the years for their support, humor, and generosity.

I would like to thank members of my room past and present for making lab a good place to work. I would like to thank Isaac Miller for being my first friend in lab, letting me know that graduate school would be difficult, and for introducing me to awesome music. I would also like to thank him for letting me help name is daughter before I was technically even a graduate student. I would like to thank Mason Appel for teaching me the cool kid lingo and for showing me that good scientists can wear many hats. I would like to thank Kimberly Beatty for being a good friend, running partner, and mentor. I learned a lot about redheads and cats from her, even though I'm married to a redhead and I've had cats my whole life.

I would like to thank Karen Dehnert for mentoring me during my first summer in the lab. Now that I know how Carolyn assigns summer students, I recognize that Karen was probably rather surprised to be mentoring biologist trying to learn chemistry for the first time. I am grateful to Karen for teaching me how to run a column, for pointing me towards an interesting project, and for not despairing when it turned out I didn't really have any synthetic experience. Most importantly, I'd like to thank Karen for her hilarious fear of Minnesota weather, which still amuses me whenever I think about it.

I am grateful to Megumi Ito for working with me for 3 years as an undergraduate. It was a privilege to train her, and she contributed throughout the entire scope of this project. Her efforts are showcased in Chapters 2-4, including her involvement in cloning ppGalNAcT mutants, enzymatically synthesizing UDP-GalN, chemically synthesizing UDP-GalNAc analogs, and validating full-length ppGalNAcTs in cells. Megumi's eagerness, hard work, and kindness helped me persevere in trying times.

I would like to thank other past and present members of this project. Jase Gehring took the risk of working with me during his rotation in our lab. He introduced me to Golden Gate cloning, which was a fantastic tool for synthesizing ppGalNAcT1, 7, and

10, and Jase synthesized the panel of ppGalNAcT1 mutants in Chapter 2. I would also like to thank Junwon Choi for recently joining the project in order to help characterize ppGalNAcT kinetics. He has forged ahead learning about UDP-sugars, and I look forward to seeing his progress on this project over the next months and years.

I would like to thank Ben Swarts, a former post-doc in the Bertozzi lab, for the idea to use diazotransfer on amino acids to make a stereochemically-defined panel of bumps. I would also like to thank him for his suggestion to use a mild diazotransfer reagent. These suggestions were instrumental in making this project streamlined and elegant. I would also like to thank Ben for being a great role model.

I would like to thank Anthony Iavarone for his thoroughness, his willingness to drop everything to answer my questions, and for his infinite knowledge of mass spectrometry. Tony performed the mass spectrometry experiment in Figure 4-5, and although we have validated some of those results on the LTQ mass spectrometer, Tony made it possible to analyze a large number of samples efficiently. When a Ph.D. project includes combinatorial sample analysis, processing many samples well becomes serious business. I am very grateful for Tony.

Two friends and lab mates have contributed significantly to aspects of this project. I would like to thank David Spiciarich for immense generosity teaching me to use the LTQ mass spectrometer (hours, days, weeks). David has been a great friend in graduate school, not just by being generous with his own time, but also by helping me feel that I had some wisdom to share. I am very grateful for all of the hours that I have spent taking science strategy with David, and also eating treats. I would like to thank Doug Fox for being a great listener and for sharing a passion for discussing big scientific ideas. Doug is a scientific dreamer, and he reminds me of where my passion for science comes from. Besides these things, Doug is profoundly generous with this time and effort in all aspects of life and work. Doug has put a huge amount of himself into making the Dionex work, and he made the HPAEC experiments in Chapter 4 possible. Without his devoted help, we probably would not have been able to do those experiments in house.

I would like to thank the members of Room 817, past and present, for always helping to cheer me up. Without the long talks, stuffed animal love, and funny videos, I might not have made it. I would like to thank Room 810 for letting me sit in their desk chairs while contemplating the future of humanity. I need to thank them for not kicking me out when I tried to live in their room with them.

I would like to thank Brian Belardi, Ellie Smith, and Jessica Kramer for being good friends throughout this process. Brian's keen scientific sense and thoroughness made him a lab role model. But really it was talking about bonobos and teaching Brian that women can be strong (athletically) that made it all worth it. Ellie has been a kind but tough ear, a great exercise partner, and a hilarious margarita fan. I look forward to keeping up with Ellie as our journeys continue. Jessica is a fellow animal

lover, house-sitter extraordinaire, and a newfound running buddy. I hope we have a blast this summer finishing experiments and exercising around the clock until the lab moves to Stanford.

I would like to thank Marjoke Debets for joining this project and for being a constant source of support over the past 2 years. Despite the risky nature of the project, Marjoke decided to help solve "the delivery problem," so that this project could move from *in vitro* studies into cells. It has been such a pleasure to have someone with whom to discuss the scope of the project as well as small technical details. I am glad to have shared my biology knowledge with Marjoke and to see how fast she carried that knowledge into new and interesting experiments. Marjoke contributed immensely to the work related to delivery presented in Chapter 4 including synthesis of protected sugars, HPAEC analysis of UDP-sugar formation, developing AGX mutants, and analyzing analogs in cells.

 I would like to thank three friends and lab mates for help editing this dissertation. Chelsea Gordon, Paresh Agarwal, and Peter Robinson all helped me elevate the style and language of this document, in addition to making it coherent! Paresh and Peter stepped up to help me at the last minute without complaint, despite having no warning. Chelsea helped wrangle an extremely stubborn chapter into shape. Without the three of them, I would feel much less satisfied with the product I have now. I would also like to thank Peter for being a great teaching buddy when we taught Chem 271 together. I always look forward to exchanging scientific strategy with Peter during his routine drop-ins by my desk on his was across Room 832.

I would like to thank the members of my entering class Gabriela de Almeida, Brendan Beahm, Chelsea, and Paresh for being great colleagues and great friends. The five of us became a fantastic team during qualifying exams (practice quals, GRS practices, and celebrating together). Our 5th year meetings helped me make it through many difficult experiments and reminded me of the light at the end of the tunnel. Chelsea and Paresh have been a particular support during this final year of graduate school. I am so grateful for their friendship. Chelsea and I have been close friends throughout this whole process. I always know that she is there to lend a willing ear no matter what the issue. Her wisdom, kindness, and scientific brilliance have been so important to my journey. I have her beautiful 6-legged bird drawing to remind me that scientists can be wildly creative beyond the scope of academics.

I would like to thank my friend Anna Goldstein for sharing many important moments in life and in graduate school with me. Her friendship, support, and love have helped me every step of the way. I would like to thank my family, Jeanette Sarbo, John Wagner, and Jamie Wagner for their love and encouragement. Knowing that they were rooting for me kept me going in the darkest of times. Finally, I would like to thank my husband Jimmy Nelson for sharing his life with me—especially because sharing our lives has included sharing our dissertation processes. As I finish this milestone, I cannot wait for what comes next in our life together.

Chapter 1. The ppGalNAcT family and bump-hole engineering

1.1. Introduction

Glycosylation is a dynamic and structurally diverse post-translational modification that alters protein structure and function. It is estimated that 20-50% of proteins within the human proteome are glycosylated[1-3]. O-GalNAc glycosylation (hereafter O- or O-linked glycosylation unless otherwise noted) is a major type of glycosylation that is critical to processes such as cellular adhesion and cell signaling, and thus is involved in embryogenesis, immune function, and cancer metastasis[4,5]. Despite its role in fundamental biological processes, O-linked glycosylation has proven extremely difficult to study and remains one of the most poorly understood post-translational modifications[4,6]. Elucidating the structure-function relationship of O-glycoproteins is a long-standing challenge.

There are 3 aspects of O-linked glycoprotein structure that are predicted to be relevant to biological function: 1) the presence or absence of O-glycans on a given protein, 2) within a given glycoprotein, the location of glycan attachment to the protein primary sequence, and 3) the overall structure of the fully elaborated glycan chain[7-9]. One family of enzymes, the polypeptide N-acetylgalactosaminyltransferases (ppGalNAcTs), regulates the first two aspects of O-glycosylation[6,10]. The ppGalNAcTs catalyze the first committed step in O-linked glycosylation, the installation of *N*-acetylgalactosamine (GalNAc) from the donor UDP-GalNAc onto serine and threonine (Ser/Thr) residues of acceptor proteins[11,12]. Thus, the substrate specificity of the 20 human ppGalNAcT family members determines which proteins are O-glycosylated and at what position[4,12]. Glycan structure is then determined by many other glycosyltransferases and glycosidases that modify the glycoprotein as it proceeds through the secretory pathway[4,10].

As illustrated in Figure 1-1, identifying the glycosylated substrates modified by a single family member is nontrivial. All current efforts to monitor the activity of individual ppGalNAcTs *in vivo* lack an unambiguous method that can confirm whether a target protein is naturally glycosylated by a specific ppGalNAcT in a biological context[6,13,14]. By identifying the substrates of individual glycosyltransferases, we can begin to identify the signaling pathways that alter glycosylation during processes such as development and disease. The objective of the work in this dissertation is to use bump-hole engineering to uniquely label and identify the substrates of a single polypeptide GalNAc transferase.
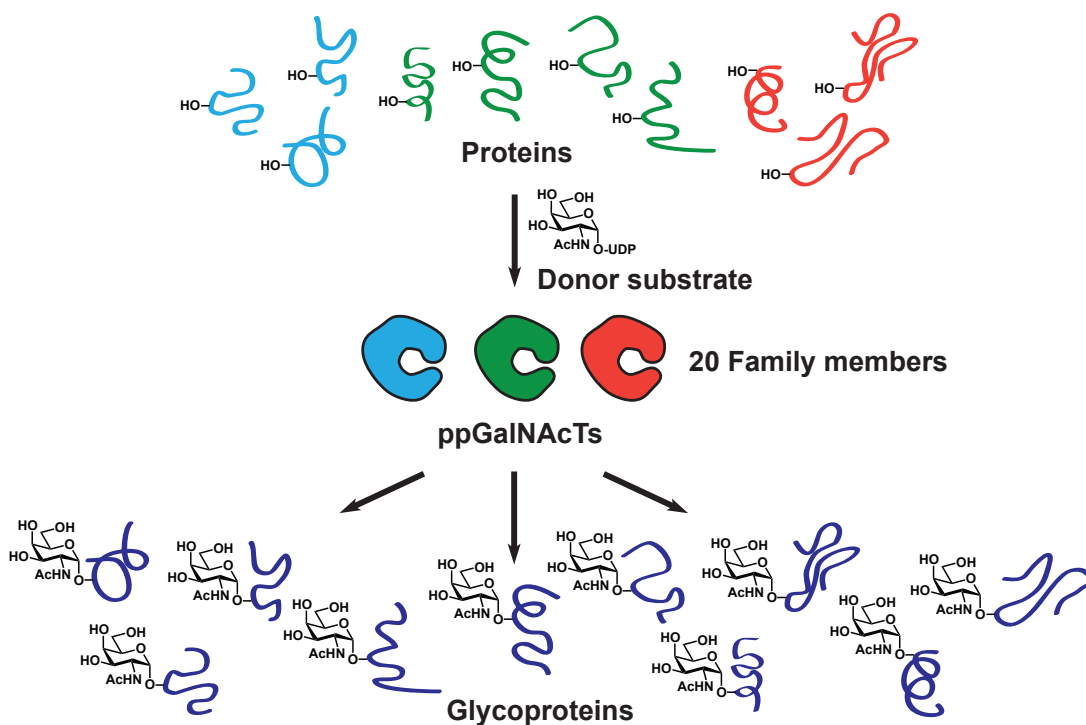
Figure 1-1. Substrate specificity of individual ppGalNAcT family members determines which proteins are O-glycosylated. Each ppGalNAcT modifies multiple proteins, but current tools cannot distinguish which glycoproteins are glycosylated by which of the 20 ppGalNAcT family members.

Our strategy involves engineering an orthogonal ppGalNAcT and UDP-GalNAc analog pair that installs a chemical handle onto O-glycoproteins (Figure 1-2). The donor substrate UDP-GalNAc is modified with a "bump" that contains a chemical handle. This orthogonal UDP-GalNAc analog has minimal activity with the native ppGalNAcTs, but is readily accepted by a modified version of a specific ppGalNAcT. This ppGalNAcT is mutated with a corresponding "hole" in the active site to facilitate binding to the bumped glycan. Thus, the ppGalNAcT mutant is able to append the tagged GalNAc moiety to polypeptide acceptors, and the glycoprotein targets of that family member are subsequently labeled with an affinity handle or fluorescent probe for further analysis[11,15]. This technique enables identification of the unique glycoproteins that are modified by an individual ppGalNAcT, a cornerstone in understanding the biological function of O-GalNAc glycosylation.
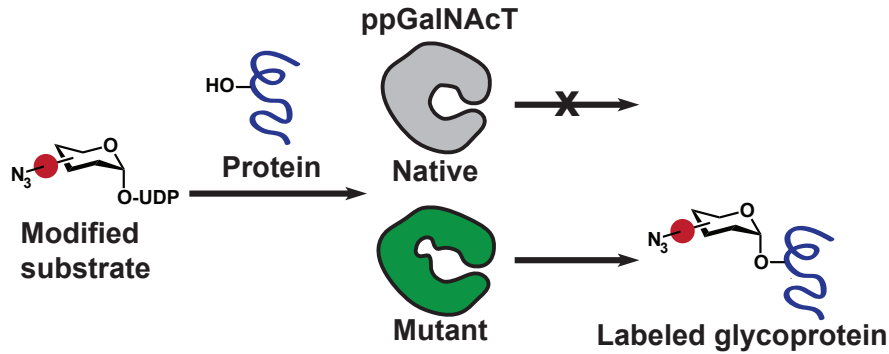
Figure 1-2. Bump-hole approach to a sugar donor-glycosyltransferase pair. A mutant ppGalNAcT is generated by enlarging the active site of one family member, making it possible to bind a donor substrate with a bump and chemical handle (red). A labeled GalNAc is appended to Ser/ Thr on acceptor proteins, generating a pool of labeled glycoprotein substrates. The native enzyme does not react with the modified donor substrate.

1.2. O-GalNAc glycosylation

In the post-genomic era, it is now appreciated that higher organisms have a vast array of post-translational modifications that modulate the structure and function of proteins[6,11,16-18]. Among these, glycosylation is conserved across all domains of life and is recognized as the most structurally complex post-translational modification[19]. Unlike DNA transcription and translation, glycan biosynthesis is not template driven. This poses analytical challenges that have significantly hampered progress in understanding glycosylation. Many glycosylation patterns are responsive to the cellular environment and are rapidly tuned by the glycosyltransferases and glycosidases that append or remove particular modifications[4,17,20]. Thus, glycoproteins elaborated with diverse glycan structures can encode subtle information with high temporal resolution. O-GalNAc glycosylation is one of the most abundant forms of glycosylation in higher eukaryotes[5,10,21]. It has been shown to play a role in embryonic development, the immune response, host-pathogen interactions, and cancer progression[5,10,21].

O-GalNAc glycosylation has long been known as mucin-type glycosylation because O-GalNAc glycosylation was first reported on secreted glycoproteins in epithelial mucosa[22,23]. These glycoproteins, called mucins, are major components of the mucus layer that serves as a protective barrier between the external environment and epithelial cells in the gastrointestinal, respiratory, and urogenital tracts[24]. Mucins are recognized by tandem repeats of proline, serine, and threonine that are densely glycosylated and result in rigid, extended protein structures[24-27]. The definition of mucins was later expanded to include mucin-type glycoproteins, after the discovery of membrane-bound proteins on epithelial cells characterized by the same repetitive primary sequences rich in O-glycosylation found in mucins[28]. As of 1990, individual glycoproteins with these characteristic mucin-domains became members of the MUC gene family as they were discovered[28]. O-GalNAc glycosylated proteins that lack some of the distinctive structural features of mucins have subsequently

3

been discovered. These proteins were labeled as mucin-like glycoproteins or as mucin-type because they contained O-GalNAc glycosylation[29].

New paradigms in the field of O-GalNAc glycosylation have arisen in the past few years in large part due to the influx of new data. One of the most dramatic advances has been the use of cells with biosynthetically truncated O-glycan structures to enable O-glycoproteomics. Mass-spectrometry-based proteomics relies on databases of proteins and their derivative peptides to categorize individual ions[30]. To identify post-translational modifications, predicted masses are supplied to these protein databases[30]. Complex O-glycans are large and heterogeneous, making it technically infeasible to input all possible masses. Without a means of chemically or enzymatically de-glycosylating these structures, peptides containing O-glycosites cannot be identified by traditional proteomic methods[31]. By truncating O-glycans during their biosynthesis, new O-GalNAc glycosylated proteins and glycosites have been identified[31-33]. Many of these glycoproteins and glycosites deviate significantly from the classical definition of a mucin[8]. Although O-GalNAc glycoproteins with these characteristics were known previously, the diversity of proteins with O-GalNAc glycosylation and the location of individual O-GalNAc sites demonstrates that mucins are not the predominant glycoproteins modified with O-GalNAc. Although all O-GalNAc glycosylated proteins used to be characterized as mucins, mucin-type, or mucin-like glycoproteins, it now appears that mucins are only a subset of proteins with this posttranslational modification. Therefore, O-GalNAc glycosylation is a more accurate description of this branch of O-glycosylation, rather than mucin-type glycosylation.

1.2.1. Structure and biosynthesis of O-GalNAc glycans

In animals, serine (Ser) and threonine (Thr) hydroxyl groups of proteins in the secretory pathway may be modified with complex O-glycans (Figure 1-3)[29,34]. These glycan structures are composed of an α-linked O-GalNAc (GalNAcα-O-Ser/Thr) moiety, known as the Tn antigen, which can be elaborated with other monosaccharides to form several core glycan structures (Figure 1-3)[10,34]. Different core structures can occur on a single protein scaffold, and their presence depends on the competing activity of many glycosyltransferases[34]. The initial installation of α-GalNAc to generate Tn antigen is carried out by the ppGalNAcT family[7,34].

Subsequently, T synthase (β1-3 galactosyltransferase), which is expressed in most cell types, can transfer galactose (Gal) to generate the evolutionarily conserved, mammalian core 1 structure (Galβ1-3GalNacα-O-Ser/Thr)(Figure 1-3)[10,34]. *C. elegans* and Drosophila also express the core 1 β1-3 galactosyltransferase gene, and one and four transferases have been functionally described in each species, respectively[10]. If the core 1 structure is not elaborated by another glycosyltransferase, it is antigenic and is known as the T antigen[34-36]. Core 1 is the most common elaboration of O-GalNAc, and upon the addition of N-acetylglucosamine (GlcNAc) to GalNAc, it can become the core 2 structure (GlcNAcβ1-6(Galβ1-3)GalNAcα-O-Ser/Thr)(Figure 1-4)[10,34]. Conversion of core 1

4

structures by β1-6 N-acetylglucosaminyltransferase to core 2 structures is regulated during embryonic development, lymphocyte activation, and cytokine stimulation[34].
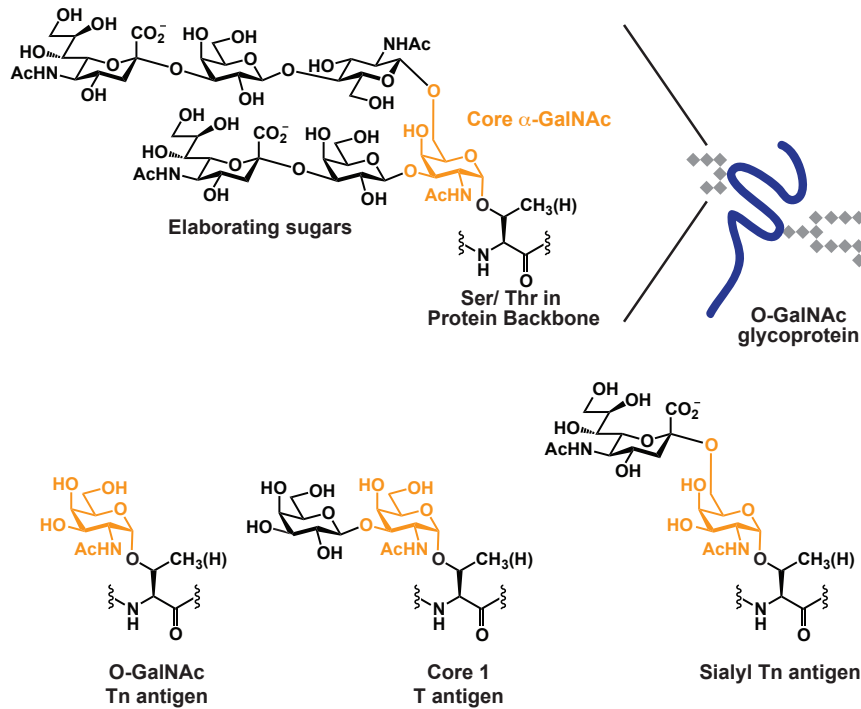
Figure 1-3. Complex O-glycans and common antigenic structures.

Alternately, the Tn antigen can be elaborated by β1-3 N-acetylglucosaminyltransferase to generate the core 3 structure, which is predominantly found in mucous epithelia of the gastrointestinal and respiratory tracts and the salivary glands[34]. Core 3 (GlcNAcβ1-3GalNAcα-O-Ser/Thr) and the branched elaboration of core 3 that forms core 4 (GlcNAcβ1-6(GlcNAcβ1-3)GalNAcα-O-Ser/Thr) have restricted expression in healthy tissue (Figure 1-4)[4,34]. Other less common core structures are often found in cancerous tissue[34]. Sialic acid (Sia) can cap the core structures to yield a range of antigenic structures often associated with malignancy including the Sialyl-Tn antigen (Siaα2-6GalNAcα-O-Ser/Thr)(Figure 1-3)[34,35,37]. Some core structures can be extended by further branching and elongated linear chains that may carry antigens such as the glycans that determine the ABO and Lewis blood groups[34,38].
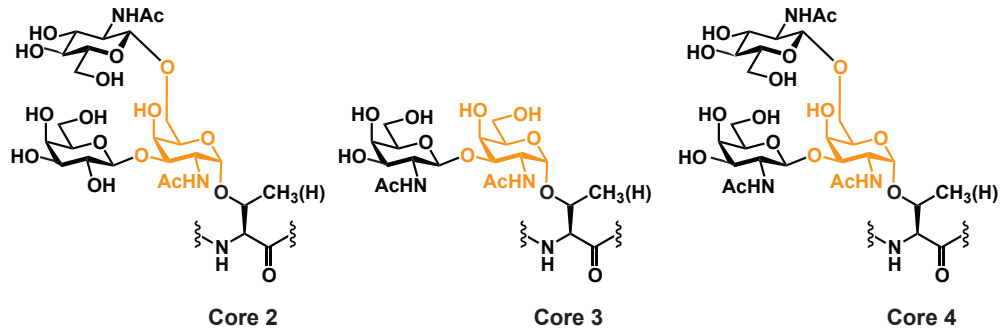
Figure 1-4. Other core structures of O-GalNAc glycans.

The glycosyltransferases described above, and many others, closely regulate the cell-surface expression of O-GalNAc glycans in a spatially and temporally specific fashion. A cell's fate is frequently tied to its glycosylation state, as exemplified by cellular reprogramming during cancer[39]. Many tumor types, particularly those of epithelial origin, are characterized by altered cell-surface O-glycosylation. Reduced activity of elaborating glycosyltransferases or increased activity of ppGalNAcTs can cause overexpression of truncated glycan epitopes including Tn, Sialyl-Tn, and T antigens, which impair cell-cell contact and increase metastatic potential[10].

1.2.2. The biology of mucin-type glycoproteins

Mucin-type glycoproteins are found on the cell surface and in the extracellular matrix and are characterized by densely O-glycosylated domains that confer an extended, rigid protein structure that can be hundreds of nanometers in length[5]. These mucin domains may occur 10-100 times within a particular glycoprotein, resulting in a biomolecule that is greater than 50% glycan by molecular weight[36,40]. Mucin-type glycan structures are often branched and composed of a variety of monosaccharides appended in different combinations, such that heterogeneity of glycan structures occurs on a single protein scaffold[34]. These structural features make it possible for mucins to extend above the cell surface, where they can mediate interactions with neighboring cells and the extracellular environment[5,34].

The aberrant glycosylation of the mucin MUC1 has been implicated in most carcinomas including breast cancer, but the source of this altered glycosylation has not been identified [37,41]. The GALNT6 gene encoding ppGalNAcT6 is upregulated in the majority of breast cancers[41]. Park *et al.* demonstrated that siRNA-mediated knockdown of either GALNT6 or MUC1 improved cellular morphology and adhesion in diverse breast carcinoma lines[41]. Furthermore, ppGalNAcT6 glycosylated MUC1 *in vitro* and *in vivo* and appeared to stabilize MUC1 protein, as measured by higher levels of MUC1 protein with increased GALNT6 expression, independent of MUC1 transcript levels. Finally, GALNT6 or MUC1 depletion reduced cell proliferation in a cell culture model. These results imply a role for ppGalNacT6 in the glycosylation of the MUC1 antigen, resulting in altered cell adhesion and carcinoma proliferation[41].

Mucin O-glycoproteins have been shown to play a direct role in modulating the immune response. In the case of the tyrosine phosphatase CD45, glycans mediate protein-protein interactions on the surface of T-cells[21]. As a monomer with three mucin domains, the cytoplasmic domain of CD45 associates with the protein kinase Lck, stimulating downstream T-cell activation. However, alternative splicing to remove the mucin domains induces cell-surface CD45 homodimer formation and prevents the Lck signaling cascades, returning T-cells to their resting state[21]. It has also been shown that these mucin domains play a role in galectin-1 mediated T-cell death. In particular, core-2 glycans on CD45 are required for high affinity interactions with galectin-1, reducing tyrosine phosphatase activity and stimulating apoptosis[42]. In this example, both the presence of glycans and the particular glycan structure plays a critical role in T-cell maturation.

O-glycans also mediate ligand receptor binding and pathogen entry. Glycoprotein B (gB) of herpes simplex virus I (HSV-1) must be glycosylated with sialic acid to bind a cellular receptor critical to viral entry[43,44]. This cellular receptor, paired immunoglobulin-like type 2 receptor α (PILRα), is found on the surface of hematopoietic and nonhematopoietic cells and normally binds its cellular ligand CD99 via sialylated residues on the ligand[43,44]. Sialylated residues on gB mimic this interaction, and point mutations to remove critical glycosylated threonine residues T53A or T53/480A on gB eliminate recognition by PILRα[43,44]. Loss of this interaction abrogates viral entry into cells expressing PILRα[43,44]. Thus, by expressing sialylated O-GalNAc epitopes, HSV-1 mimics natural ligand-receptor interactions and facilitates infection.

Mucins are critical during development, such that loss of particular mucin structures yields embryonic death in Drosophila and mice[10]. Due to their role in cell-cell adhesion, mucins play a critical role in embryonic patterning and cell migration[5,10]. Loss of the T synthase gene during mouse development impairs formation of the vascular and tubular systems, such that gene knockout induces fatal brain hemorrhage by embryonic day 14, while reduced enzyme activity causes kidney failure and death within 200 days[5,10]. Ablation of T synthase is also lethal in Drosophila due to fatal morphogenesis defects of the ventral nervous cord and brain hemispheres[45].

1.3. The ppGalNAcT family

Due to the range of critical functions attributed to glycans, it is very important to understand the extent and the timing of these post-translational modifications[46,47]. However, unlike the biosynthesis of nucleic acids and proteins, glycosylation is not genetically encoded and cannot be controlled directly[48]. Therefore, one of the most promising techniques to discern the function of particular glycan modifications is to study the glycosylation enzymes that install them, particularly by genetic or chemical perturbation[12,47,49,50]. Among the many glycosyltransferases involved in mucin biosynthesis, the ppGalNAcT family plays a unique role in selecting the proteins destined to become O-GalNAc glycoproteins (Figure 1-5). This makes them

an attractive target to study the regulation of O-glycosylation and to identify therapeutic glycoprotein targets.
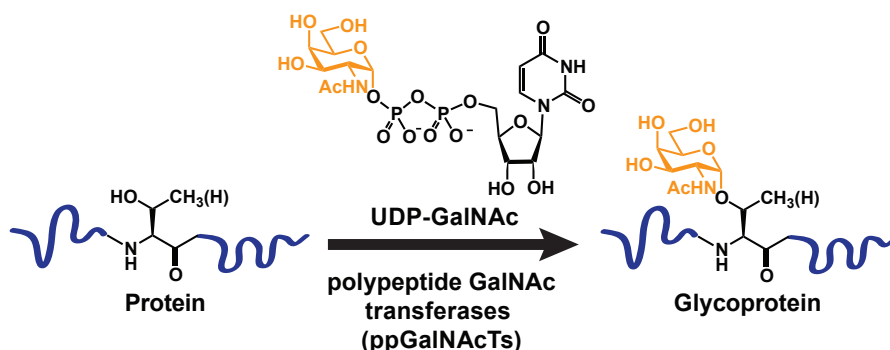
Figure 1-5. The ppGalNAcTs initiate O-GalNAc glycosylation. The ppGalNAcTs accept the nucleotide sugar UDP-GalNAc as a donor substrate. Acceptor peptides or proteins are glycosylated at Ser/ Thr residues with an α-GalNAc moiety.

The ppGalNAcTs are Golgi resident enzymes whose sub-localization within the Golgi network may reflect their individual roles regulating O-GalNAc protein glycosylation[13,51,52]. Like other Golgi-resident glycosyltransferases, the ppGalNAcTs are type II transmembrane proteins with an N-terminal cytoplasmic region, a transmembrane domain, a variable stem region, and a soluble catalytic domain[7,10]. However, unlike other glycosyltransferases, the ppGalNAcTs also have a soluble lectin domain that can bind GalNAc already installed on a peptide backbone[10,53].

As shown in Figure 1-6, several ppGalNAcT family members may act sequentially on the same protein substrate within the Golgi, altering the affinity of other ppGalNAcT family members for that substrate[14,21,53]. The ppGalNAcTs are categorized based on their substrate preferences for peptide substrates *in vitro*, specifically whether the peptide has been previously glycosylated. These substrate preferences likely reflect the order in which these different family members act on a protein as it progresses through the secretory pathway. Those ppGalNAcTs that preferentially modify unglycosylated peptides are believed to initiate glycosylation on protein substrates, while ppGalNAcTs that prefer to modify glycopeptides likely act later on the same protein substrates. Recent work by Gill *et al.* suggests that the ppGalNAcTs are sometimes trafficked to the endoplasmic reticulum[7,51,54]. This results in denser O-GalNAc glycosylation patterns in the absence of competition by the elaborating glycosyltransferases that act on these substrates in the Golgi[51,54].
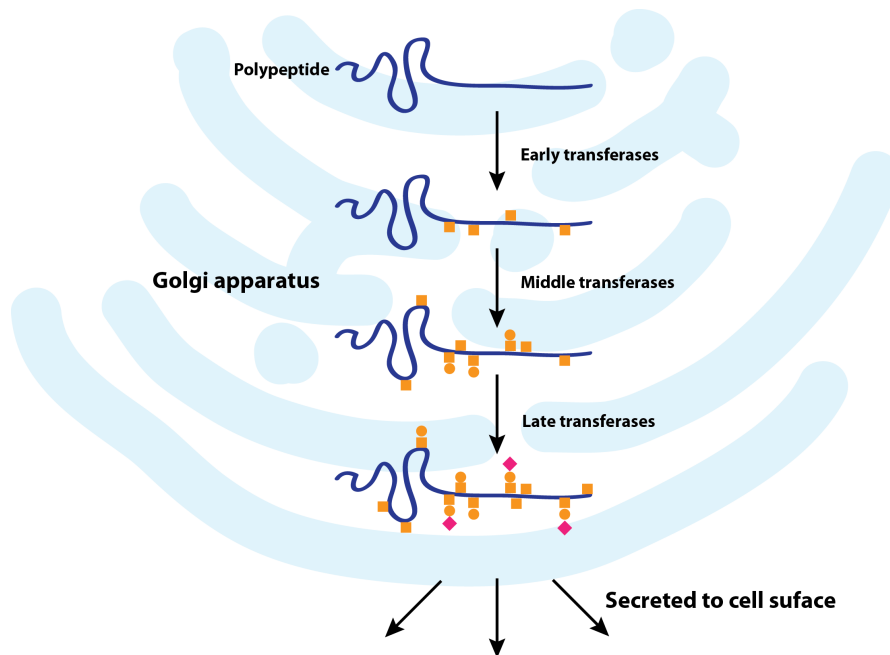
Figure 1-6. **Distribution of Golgi-resident ppGalNAcTs may reflect their acceptor substrate preferences.** Different ppGalNAcT family members add GalNAc (orange square) from the cis to the trans Golgi. Monosaccharides such as galactose (orange circle) and sialic acid (pink diamond) are added by elaborating glycosyltransferases during glycoprotein maturation. Early ppGalNAcTs prefer unmodified substrates while late ppGalNAcTs act on more heavily glycosylated substrates.

The individual ppGalNAcTs display distinct spatial and temporal expression patterns during mammalian and fly development, which may reflect their unique roles governing cell-specific glycosylation[10,12,55]. Efforts to establish the roles of individual family members by genetic knockdown have demonstrated unpredictable functional redundancy, such that ablation of ppGalNAcTs 1, 3, 4, 5, 10, 13, and 4/5 in mice has yielded only partially penetrant phenotypes[5,10,12,13,56,57]. Until 2012, only one ppGalNAcT family member was known to be essential for embryonic development, the Drosophila homolog *pgant35A*. There is not a human homolog that is known to have similar function, and although human GALNT11 shares the greatest sequence similarity to *pgant35A*, GALNT11 does not rescue the loss of *pgant35A*[56].

However, recent work by Ten Hagen and colleagues demonstrates that five of the 14 known ppGalNAcT homologs (*pgants*) are necessary for viability. The authors generated Drosophila lines where the inducible Gal4 promoter controls expression of RNAi targeting each individual *pgant*. When these lines are crossed with flies expressing the Gal4 driver, RNAi expression targeting these family members is lethal. All 5 *pgants* are necessary in a tissue-specific fashion, as confirmed by crosses with Gal4 driver lines that induce RNAi expression in individual tissues of the developing embryo. Lethality is observed for some tissues unless multiple *pgants* are expressed, including the digestive tract and mesoderm. Four of the 5 *pgants* are critical to respiratory system function, and one of these is essential for viability. The

human homologs to which these 5 *pgants* correspond most closely include GALNT1, 7, 10, and 11.

The ppGalNAcT family has increasingly been implicated in many human diseases, particularly cancers[58-62]. Although the ppGalNAcTs are correlated to disease occurrence and outcome, there are significantly fewer data about the specific glycosylated substrates that are dysregulated in these diseases. Also, there are few examples of proven causative relationships between the ppGalNAcTs and diseases, particularly in the absence of individual glycoproteins to monitor. In recent years, a few targets of individual ppGalNAcTs have been identified and verified biochemically. Fibroblast growth factor-23 (FGF23) was one of the first glycoproteins shown to be glycosylated by a specific ppGalNAcT[63,64]. Mutations in both GALNT3 and FGF23 were both identified in patients with familial tumoral calcinosis (FTC), which ultimately led researchers to link T3 with FGF23 glycosylation[64,65].

O-glycosylation of proteins plays a role in protein secretion by regulating proprotein convertase (PC) processing by proteases. This proprotein proteolysis is a major post-translational modification and a critical step in protein maturation of many cell-surface and secreted proteins including hormones, cytokines, growth factors, and receptors[66,67]. O-glycosylation by T3 and T2 regulates the PC processing of FGF23 and angiopoietin-like protein 3 (ANGPTL3), respectively. Improper O-glycosylation results in defective proteolysis, which causes a failure to activate or inactivate the glycosylated protein.

ppGalNAcT	Correlated diseases	Known substrates	Known relationship between ppGalNAcT + substrate glycosylation + disease
1	Bladder cancer	mouse osteopontin, mouse bone sialoprotein[68], PSGL-1	
2	Gastric[69], brain, hepatocellular carcinomas[62]; oral squamous cell carcinoma[59]; dislipidemia[70]	ANGPTL3[67,70], EGFR, APOCIII, IgA hinge, and lecithin-cholesterol acyltransferase	EGFR (hepatocellular[62], oral[59]); APOCIII and ANGPTL3 (dyslipidemia)
3	Familial tumoral calcinosis[64,71]; Ovarian cancer		FGF23(calcinosis)[64]
6	Breast[41] and gastric[72] carcinomas; Pancreatic cancer	fibronectin	Muc 1 (breast)
7	Cervical cancer; melanoma		
9	neuroblastoma		
10	Gastric carcinoma[73]		
11	Chronic lymphocytic leukemia[60]; heterotaxy disorder[74]		
12	Gastric, colorectal, breast cancers		
13	Lung cancer; neuroblastoma		
14	Pancreatic carcinoma; non-small-cell lung carcinoma; melanoma; ovarian cancer[58,61]		Muc 13 (ovarian)
Pgant4 (Drosophila)		TANGO1[75]	

Table 1-1. Some ppGalNAcTs, correlated human diseases, and known substrates.

In the case of FGF23, O-glycosylation by ppGalNAcT3 adjacent to the PC processing site prevents proteolysis, and results in the secretion of active FGF23[64,76]. Patients with FTC were found to have any of the following mutations that result in the loss of active, circulating FGF23: mutations in GALNT3 resulting in loss of glycosylation activity, mutations to the O-glycosylation site adjacent to the PC processing site in FGF23, or mutations in FGF23 causing the improper folding and secretion of full-length FGF23[64]. Loss of FGF23 or GALNT3 in mice mimics FTC with

hyperphosphatemia[63,65,77], suggesting a causative relationship between FGF23 glycosylation by ppGalNAcT3 and familial tumoral calcinosis[78]. It is with the knowledge of the ppGalNAcT and its specific target that the causative role of glycosylation in this disease has become clear.

Even more intriguing, recent work has implicated phosphorylation of FGF23 at a residue adjacent to the glycosylation site in the differential glycosylation of FGF23[76]. When phosphorylated, FGF23 is not glycosylated at the residue critical for PC processing. Thus, phosphorylation increases PC processing, in contrast to glycosylation. Both phosphorylation and O-GlcNAcylation, dynamic and interchangeable modifications of nuclear and cytoplasmic proteins, are involved in crosstalk[18,79]. However, because most known kinases are cytoplasmic and therefore should not phosphorylate proteins in the secretory pathway, there is little evidence of crosstalk between O-GalNAc glycosylation and phosphorylation. Recently identified kinases that are resident in the secretory pathway have been shown to phosphorylate FGF23[76].

1.3.1. Technological advances to studying O-GalNAc glycosylation

During the last decade, there has been a focus on understanding the substrate specificity of the ppGalNAcTs *in vitro*. Studies have shown that ppGalNAcT family members have both unique and redundant peptide substrate specificities[13,14,53,68,80,81]. Although a consensus sequence for the ppGalNAcTs is not known, extensive work with peptide libraries by Gerken and colleagues suggest amino acid motifs near the O-glycosylation site that are favored by individual ppGalNAcTs[8,82]. However, the only common site preference that has been identified is for proline residues adjacent to the Ser/Thr that will be glycosylated[8,63].

In recent efforts to study O-glycoproteins and their sites of glycosylation *in cellulo*, Clausen and coworkers have generated cell lines with truncated glycans (SimpleCells). These cell lines exploit a known mutation found in the Jurkat and LSC cell lines, where the glycosyltransferase chaperone protein cosmc is inactivated[31]. To engineer SimpleCells, zinc-finger nucleases were used to stably knockout *COSMC* (*COSMC*-/-). In the absence of cosmc, T synthase cannot generate the core 1 mucin structure. This results in the loss of both core 1 and 2 structures, which usually account for the majority of mucin structures in healthy tissue and in most cell lines[31].

In general, proteomic analysis of O-glycoproteins is prohibitively difficult because the complex glycan structures cannot be reduced to a single uniform mass that can be identified by database search. By simplifying the structural complexity of the O-glycan structures in SimpleCells, Clausen and colleagues demonstrated that it was possible to enrich and identify glycopeptides with their corresponding O-glycosites[31]. Application of SimpleCell engineering to many human-derived cell lines, followed by lectin enrichment of the simplified glycans and peptide and O-

glycopeptide identification by mass spectrometry, has resulted in the discovery of many new glycoproteins and specific O-glycosylation sites.

Due in large part to the O-glycoproteomics enabled with SimpleCells, many existing paradigms regarding O-GalNAc glycosylation have come into question. As described earlier in this chapter, although mucin domains were once considered a hallmark of O-GalNAc glycoproteins, they are only a subset of the structural motifs displaying O-GalNAc glycosylation. O-GalNAc glycosylation has been found in type-I and type-2 transmembrane stem regions of proteins, in linker regions, and in conserved folded domains[8]. Although mucins play many important biological roles, these findings expand the known space of O-glycoproteins and suggest other possible roles for O-GalNAc glycosylation.
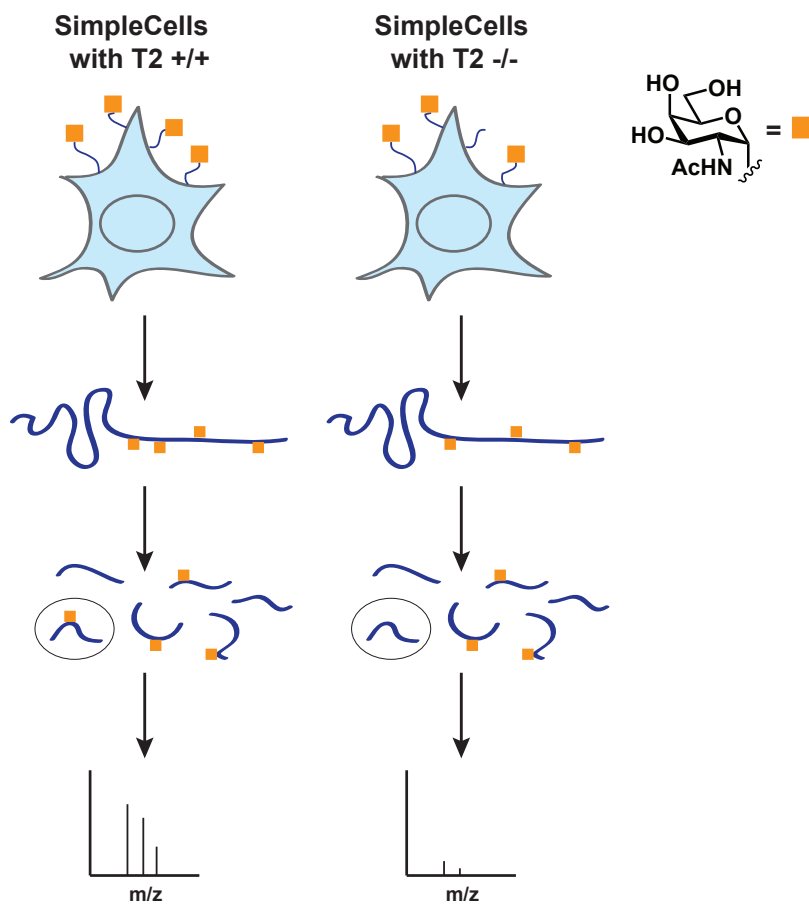
Figure 1-7. SimpleCell (SC) engineering to identify ppGalNAcT2 substrates. 1) ppGalNAcT2 knockout eliminates specific glycosites, 2) cells are lysed, 3) proteins are digested to form individual peptides and glycopeptides, 3) samples are subjected to MS, and the T2-specific glycopeptide is only detected in the SC sample.

Clausen and colleagues have carried out the first proteomics study to isolate the substrates of individual ppGalNAcTs using SimpleCells and ppGalNAcT2 +/+ and -/- cell lines[83]. In this study, hepatocyte cell lines with truncated glycans (hepatocyte SimpleCells) were evaluated for the loss of individual O-GalNAc epitopes when

ppGalNAcT2 was knocked out (Figure 1-7). This experimental design is challenging because the results depend on the detection of a loss of signal. Compared to other ppGalNAcTs, T2 is hypothesized to be an early transferase that glycosylates many substrates. Therefore, it was one of the best candidates to have detectable loss of signal in this study. The authors found 73 glycosites that were only identified in T2 +/+ cells and 30 sites only in T2 -/- cells. In theory, the 73 glycosites unique to T2 +/+ cells were substrates of T2 specifically.

There were several candidate T2 protein substrates identified based on the analysis described above, although validation of these results was non trivial. The authors generated synthetic peptides corresponding to some of the peptides identified in the proteomic analysis and subjected them to purified T2, T1, T4, or T11 *in vitro*. Of the 16 peptides selected from hits in T2 +/+ cells, which theoretically contain glycosites specific to T2, 8 were glycosylated by T2, 5 by T2 and other family members, and 3 by none of the ppGalNAcTs tested. Of the 5 peptides based on hits in T2 -/- cells, 2 were glycosylated by T2 and other family members and 3 were not glycosylated by any of the 4 ppGalNAcTs tested. These results demonstrate the challenges associated with making conclusive ppGalNAcT substrate identifications in a cellular context, particularly in cells with biochemically perturbed glycosylation.

Despite the advances made possible by truncated glycans in the SimpleCell technology, the loss of major glycan structures limits the utility of this technology beyond cell lines. Additionally, SimpleCell technology requires the detection of an absence of signal, which is a significant hindrance to studying individual ppGalNAcTs. The ppGalNAcTs are known to compensate for the loss of individual family members, meaning that sites unique to a specific ppGalNAcT in a normal cellular context might be glycosylated by other ppGalNAcT family members when that ppGalNAcT is knocked out. Hence, this long-standing complication is not resolved by SimpleCell technology combined with an individual ppGalNAcT knockout. For a glycoprotein-preferring ppGalNAcT, the knockout system must compete with all other sites of O-GalNAc glycosylation from the early-acting ppGalNAcTs.

1.4. Bump-hole engineering

In humans there are 20 ppGalNAcT family members working in concert during the biosynthesis of O-GalNAc glycans[5]. Chemically targeting one isoform is difficult because members of the ppGalNAcT family demonstrate a high level of structural homology and all utilize the nucleotide sugar UDP-GalNAc as a donor substrate[6,12,47,84,85]. Efforts to isolate the function of an individual isoform by genetic knockdown have demonstrated unpredictable functional redundancy among the ppGalNAcTs[12,85]. The challenges associated with studying the ppGalNAcT family using classical genetics or inhibitors have demonstrated the limitations of these approaches. Genetic mutation and knock down can be executed with limited temporal control, and small molecules can rapidly alter protein function, but often lack the desired specificity[49,50].

A bump-hole technique combines the specificity of genetic modification with the temporal control of a chemical strategy by mutating the target gene to be uniquely susceptible to a small molecule. Of particular value is that fact that as a non-native substrate, the small molecule can be equipped with a chemical handle, enabling downstream isolation and identification of biomolecules containing that handle[15]. As described in section 1.1, the bump-hole strategy involves modification of an individual ppGalNAcT family member with an enlarged binding pocket and a UDP-GalNAc analog with a chemical handle. This pair will glycosylate and label the specific polypeptide substrates of that ppGalNAcT, enabling downstream identification of labeled glycoproteins[15,86]. Unlike all existing technologies to study ppGalNAcTs, a bump-hole pair enables selective enrichment of ppGalNAcT-modified glycoproteins independent of modification site, degree of glycosylation, or simultaneous glycosylation by other ppGalNAcTs.

1.4.1. Steps toward bump-hole engineering

Bump-hole engineering has its roots in early efforts to achieve pharmacological perturbations in systems of interest[49,87]. Treatment of biological systems, and the proteins and other biomolecules within them, with drug-like molecules provides insight into the system being studied. Proteins have been the focus of pharmacological research because they have myriad critical cellular functions and they naturally interact with small molecules such as cofactors, steroid hormones, neurotransmitters, and natural products[87,88]. Ideally, a small molecule targets a specific protein, but frequently it targets a family of proteins with common structural features. The small molecule might be an antagonist like an inhibitor or an agonist such as a substrate analog.

Chemically perturbing a protein of interest makes it possible to observe its function within a system. Generally, agonists and antagonists are opposing technologies that provide different but often complementary information about the target. An antagonist such as an inhibitor enables the analysis of phenotypes due to the loss of function of the target. A substrate analog agonist mimics the natural substrate, and can activate the system by exogenous application of the analog. Substrate analogs equipped with a chemical handle can undergo a bioorthogonal reaction, specifically tagging the products of a target enzyme[15,80,89]. Bioorthogonal reactions proceed readily at physiological pH and temperature, and selectively in the presence of other biological functionalities to form a unique and stable ligation product[90].

Metabolic engineering takes advantage of promiscuous biosynthetic enzymes to introduce substrate analogs. Cells and animals treated with analogs to natural metabolites such as nucleotides, amino acids, and sugars can incorporate these analogs into biomacromolecules such as DNA, protein primary sequences, and post-translational modifications of proteins. Metabolic engineering using tagged analogs makes it possible to monitor the fate of the analog within the cell. A variety of classes of proteins have been studied using small molecule analogs equipped with

the well-established azide or alkyne bioorthogonal handles[90]. However, these techniques are generally not very specific for a protein of interest. To monitor an individual enzyme that uses a metabolite, a substrate that is accepted by a range of enzymes is undesirable. A small molecule analog that is not accepted by any native enzymes in the biological system is said to be 'orthogonal' to that system. By engineering a protein to be uniquely reactive with an orthogonal substrate analog, the function of a specific protein can be monitored.

Bump-hole steric complementation uses rational design to match expansions (or contractions) in the protein structure to corresponding structural changes in the substrate analog. As shown in Figure 1-8, bump-hole steric complementation can lead to enzyme or substrate orthogonality with the native system. Bump-hole steric complementation techniques have been used to engineer enzymes such as ADP-ribosyltransferases, ATPases (myosin and kinesin, GroEL), GTPases (G-proteins), acetyltransferases, and glycosyltransferases[50,88]. Other classes of proteins include phosphatases (inhibitors only), proteases (unique peptide sequences), receptors, and DNA/RNA binding proteins[88].
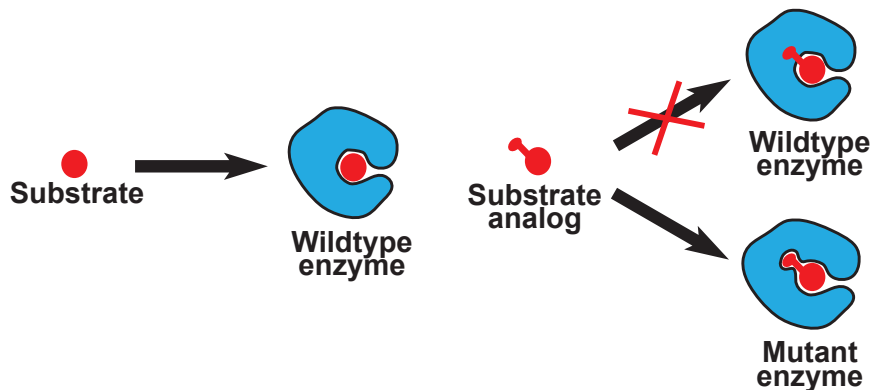
Figure 1-8. Bump-hole steric complementation.

1.4.2. Precedent for bump-hole engineering

A central goal of bump-hole engineering is to ensure that the enzyme-substrate pair maintains native activity. In the context of ppGalNAcT engineering, the goal is to maintain native glycosyltransferase activity and polypeptide substrate specificity. Early precedent engineering proteins such as GTPases, 7-transmembrane receptors, kinases, and nuclear hormone receptors to accept non-native ligands has demonstrated that native function can be recapitulated with an orthogonal system[49]. In particular, pioneering work by Shokat and colleagues demonstrated the efficacy of a bump-hole approach in the study of phosphorylation and the roles of individual protein kinases[49,50,91-93]. Much like O-GalNAc glycosylation, the hundreds of kinases that act together to phosphorylate proteins all utilize the common phosphate donor adenosine triphosphate (ATP), and it is similarly challenging to isolate the activity of a single kinase[50,94]. Studies have shown that an engineered

kinase with an expanded ATP binding pocket can accept enlarged ATP analogs or kinase inhibitor analogs[15,92]. Engineered kinases are known as analog-sensitized (AS) kinases. Complementation experiments have demonstrated that an AS-kinase can rescue the knockdown of the native kinase allele, which is one demonstration that the bump-hole pair preserves native function[95,96]. Because the bump on ATP analogs are installed on the adenosine base, these ATP analogs are termed adenosine* triphosphate (A*TP).

Shokat and colleagues have used AS-kinases extensively with inhibitors to study cellular phenotypes arising for temporally-controlled loss of the kinase[97-99]. For these experiments, engineering to generate the AS kinase must preserve its activity with ATP, so that in the absence of the inhibitor, the kinase functions normally[99]. When AS-kinases are studied with ATP analogs, the AS kinase needs to use the analog preferentially to ATP[100]. However, engineered AS kinases studied with bumped inhibitors or A*TP are based on the same mutations in the active site. Thus, the AS-kinases studied do not have very dramatically reduced or absent activity with ATP. Thus, AS kinases utilize ATP much like the native kinase, which has made it straightforward to validate that the AS kinase functions like the native kinase in a biological context[101]. There are many well understood phenotypes specific to the loss of individual kinases, enabling Shokat and coworkers to demonstrate that knockout or depletion of the wildtype kinase can be rescued by the AS-kinase.

1.4.3. Substrate labeling with bump-hole pairs

The majority of the initial work with kinase bump hole engineering was based on adding bumps to biologically-validated kinase inhibitors. Such an approach was particularly useful because ATP and its analogs are not cell permeable, whereas the inhibitors allowed a wide variety of *in vivo* experiments. After early work with AS kinase inhibitors *in cellulo*, Shokat and coworkers generated mice with AS-kinase alleles[102]. Because A*TP could not be used in mice, all studies in this animal model used bumped inhibitors.

Due to the cell permeability limitation of A*TP, the majority of experiments aimed at specifically labeling and identifying kinase substrates have been carried out in cell lysates. In early work, lysates doped with a purified AS-kinase, or containing an AS kinase expressed in cells prior to lysis, were treated with A*TP radiolabeled at the gamma phosphate position. When the AS-kinase phosphorylates its protein substrates with the A*TP analog, the ^{32}P label is transferred to the phosphoprotein. This makes it possible to identify the radiolabeled substrates of a particular AS kinase. Although lysate studies have made it possible to identify kinase substrates, in the absence of the native cellular context, the biological relevance of a particular phosphorylation event is unclear. In 2007, Allen *et al.* successfully used digitonin-permeabilized cells to deliver A*TP with a thiophosphate at the gamma phosphate (A*TP-γ-S)[15]. Although permeabilized cells only survive for 20-30 minutes, that is long enough to achieve labeling of AS-kinase substrates[15,103].
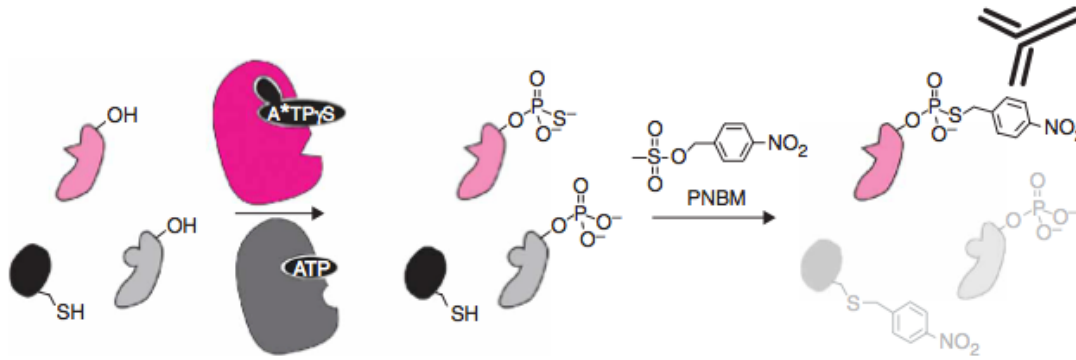
Figure 1-9. A*TP-γ-S kinase substrate labeling. Figure reprinted with permission[15]. Copyright 2007, Nature Publishing Group.

As shown in Figure 1-9, the specific enrichment of AS-kinase substrates was made possible by bumped thiophosphorylated A*TP (A*TP-γ-S). This thiophosphate takes advantage of the concepts behind bioorthogonal chemistry. Unlike a fully bioorthogonal reaction, the reaction partner for thiophosphate does not react uniquely with thiophosphate. To be able to visualize thiophosphorylated proteins, the alkylating agent *p*-nitrobenzyl mesylate (PNBM) is used[15,104]. PNBM also alkylates other nucleophiles such as cysteine. Therefore an antibody specific to the alkylated thiophosphate ester was developed that allowed the preferential enrichment of this epitope. This methodology and the thiophosphate modification make it possible to label the protein substrates of an individual kinase. By carrying out a ligation on the thiophosphorylated protein substrates of the bump hole pair, these substrates can be enriched and analyzed by mass spectrometry.

With a strategy to label and enrich kinase substrates, Shokat and colleagues carried out the first AS-kinase substrate labeling experiment in a living system[105]. They injected zebrafish embryos at the one-cell stage with AS-kinase mRNA and with the appropriate A*TP-γ-S[105]. Zebrafish injected at the 1-cell stage can share the AS-kinase mRNA and A*TP-γ-S with all of their daughter cells, which makes it possible to deliver ATP analogs intracellularly.

There are some limitations to the A*TP-γ-S ligation strategy. The ability to utilize the thiophosphate at the gamma position is a property of some wildtype kinases, and is not a product of protein engineering on the kinases studied. Thus, some native kinases can accept ATP-γ-S, while others cannot[106]. The AS kinases that have been studied with A*TP-γ-S were only engineered to have an enlarged binding pocket near then N6 position of the adenosine base of ATP so that they could utilize A*TP[15,103]. The outcome of this is that only a subset of all possible AS kinases can accept A*TP-γ-S, unless further kinase bump hole engineering is carried out near the binding pocket for the gamma phosphate.

The bump-hole technology is now well established, and recent work by Luo and coworkers with engineered protein methyltransferases has demonstrated the

versatility of the bump-hole technique with other families of enzymes. Methyltransferases use S-adenosyl methionine (SAM) as a methyl donor to modify proteins or DNA. By expanding the methyl group on SAM to bulkier moieties containing an alkyne, they were able to generate orthogonal SAM analogs that are only recognized by the engineered methyltransferase[107,108]. These alkyne-SAM analog allowed them to label the protein targets of individual methyltransferases[89,109].

Enzyme-substrate engineering has been performed on a very limited number of glycosyltransferases and is an emergent field. A small portion of the work engineering glycosyltransferases has aimed to introduce non-natural sugar analogs into the enzymes' repertoire[110-114]. No glycosyltransferase enzyme-substrate engineering has utilized substrates that are orthogonal in a cellular context.

1.5. Bump-hole engineering to label specific O-glycoproteins

We have generated a bump-hole system by engineering an orthogonal ppGalNAcT and UDP-GalNAc analog pair that can install a chemical handle onto O-glycoproteins. The donor substrate UDP-GalNAc is modified with a "bump" that contains a chemical handle. This orthogonal UDP-GalNAc analog is poorly accepted by native ppGalNAcTs, but is readily accepted by a modified version of a specific ppGalNAcT isoform. This ppGalNAcT isoform is mutated with a corresponding "hole" in the active site to facilitate binding to the bumped glycan. Thus, the ppGalNAcT mutant can append the tagged GalNAc moiety to protein acceptors, and the glycoprotein targets of that isoform can be subsequently labeled with an affinity handle or fluorescent probe for further analysis[11,15].

Generating a bump-hole pair requires the development of a panel of mutant enzymes and a panel of UDP-GalNAc analogs (Figure 1-10). We rationally designed the panel of ppGalNAcT mutants based on available crystal structures. The design and cloning of these mutants, as well as their expression in mammalian cells and affinity purification is described in Chapter 2. The panel of UDP-GalNAc analogs was based on hexosamine analogs that have been successfully processed in cells. As described in Chapter 3, we designed the panel to be synthetically tractable and highly diversifiable. With these panels in hand, we set about screening for enzyme-substrate pairs that demonstrate 1) reactivity with each other, 2) orthogonality of the analog with the native enzymes, and 3) unaltered preferences for acceptor peptides/proteins compared to the wildtype enzyme. Our work screening the panels and validating them in a biological context is described in Chapters 4.
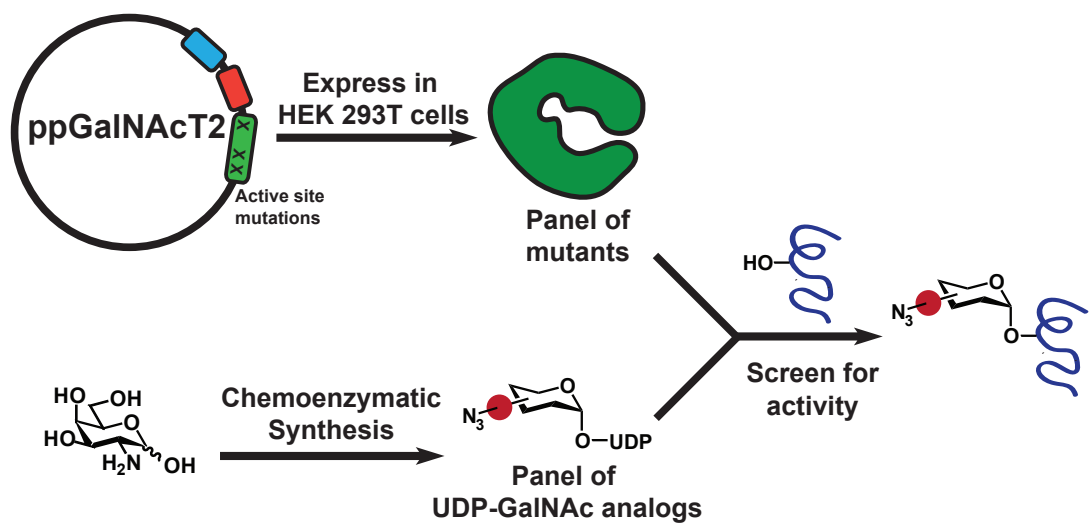
Figure 1-10. Strategy to develop a ppGalNAcT bump-hole pair.

Chapter 2. A panel of ppGalNAcT mutants with enlarged binding pockets

2.1. Bump hole design

Previous work has shown that single amino acid mutations in the active site of β-1,4-galactosyltransferase (β4Gal-T1) and other glycosyltransferases expand their donor substrate specificities[110-113]. β4Gal-T1 normally accepts UDP-galactose, from which it transfers galactose to terminal GlcNAc residues. The point mutation Y289L expands its donor specificity to include modifications at the C2 position of galactose, which allows β4Gal-T1-Y289L to use UDP-GalNAc as efficiently as UDP-Gal[111]. β4Gal-T1-Y289L is also known to utilize sugars modified with a chemical handle at the C2 position such as C2-keto-galactose and GalNAz[112,113]. Activity with C2-keto-Gal and GalNAz is robust, which allows β4Gal-T1-Y289L to modify O-GlcNAcylated proteins with Gal analogs and selectively install a chemical handle for downstream analysis. These precedents set the stage for rationally designed ppGalNAcT mutants that accept novel donor substrates.

2.1.1. Selection of the GalNAc binding pocket
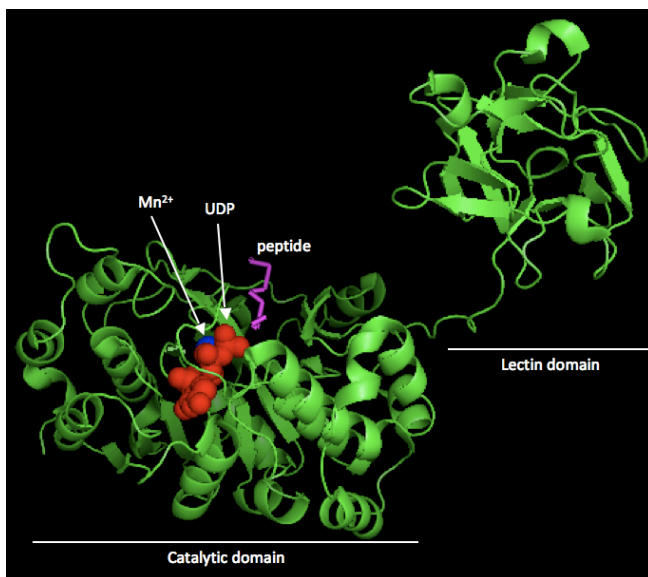
Figure 2-1. Crystal structure of ppGalNAcT2 with Mn^{2+}, UDP, and EA2 acceptor peptide bound. PDB ID: 2FFU.

The catalytic domain of the ppGalNAcTs contains binding sites for both the donor and acceptor substrates[115,116]. Unlike other glycosyltransferases, the ppGalNAcT family is characterized by a unique lectin domain that contributes to acceptor substrate recognition (Figure 2-1)[115-117]. As the bump-hole strategy aims to preserve native acceptor substrate preferences, we sought to introduce mutations only in the donor-binding site of the catalytic domain.
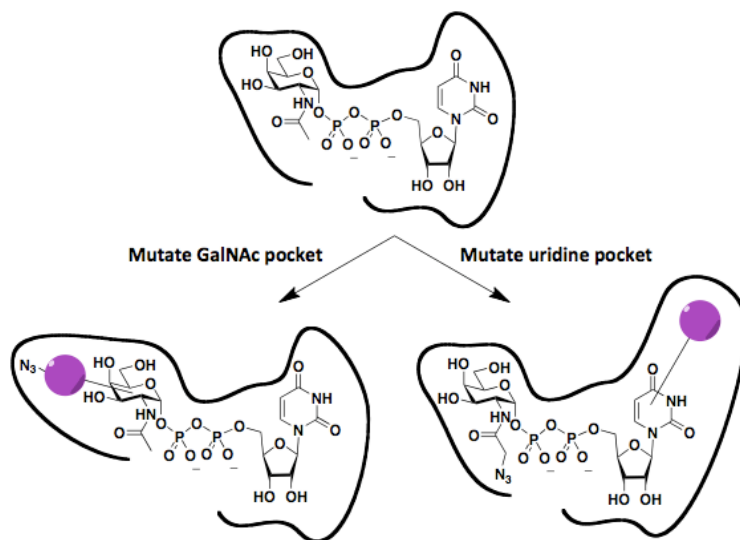
Figure 2-2. Strategies for mutating the ppGalNAcT active site. The structure of UDP-GalNAc constrains which portions of the ppGalNAcT active site should be mutated. The position of the chemical handle, here an azide, would depend on the location of the bumped moiety.

To enlarge a suitable pocket for a bumped substrate, two regions within the enzyme active site were considered: the amino acids surrounding either the uridine base or the GalNAc moiety (Figure 2-2). Correspondingly, the selection of amino acid residues in the enzyme active site is dependent on the ultimate position of the bump on UDP-GalNAc. The uridine or GalNAc moieties are both potential positions for donor substrate modification. In determining what functional groups to install on our UDP sugar analogs, we took inspiration from Shokat and coworkers. They observed that the addition of bulky substituents such as benzyl[91,93], cyclopentyl[118], and phenethyl[119] groups on the adenosine base of ATP are ideal for a kinase bump-hole pair. We reasoned that we might find similar success and used these functional groups as a starting point.

One concern we anticipated is the dearth of studies that explore the bioavailability of nucleotide sugars, especially those with modified nucleotide bases. Unlike the cytoplasmically resident kinases originally studied with bump-hole engineering, the ppGalNAcTs are Golgi-resident and thus UDP-sugars suffer from dual membrane barriers to efficient delivery. For our UDP-GalNAc analogs to be correctly localized in cells, they must traverse not only the cell membrane, but also the Golgi stacks via nucleotide sugar transporters[120,121]. It is unknown if these Golgi transporters tolerate modifications to the uridine base, which was not a concern for the ATP analogs. As an additional confounding factor, recent work demonstrates that UDP-GalNAc modified at the uridine base inhibits glycosyltransferase activity by impeding conformational changes necessary for enzyme activity[122].

Ultimately, our goal is to use the orthogonal ppGalNAcT-donor substrate pair to identify glycosylated substrates that contain GalNAc analogs with a chemical handle. In designing UDP-GalNAc analogs, modification of the GalNAc moiety can simultaneously install both the bump and chemical handle. Shokat and colleagues found that, although the adenosine base of ATP affords a convenient site to introduce a bump, identification of specific phosphorylated kinase targets requires the installation of a chemical handle on a second site on ATP to enable phosphoprotein detection or enrichment. Initially, the gamma-phosphate was radiolabeled[91]. Later iterations utilized a thiophosphate group at this same position for chemical enrichment[15,104,106]. Similarly, installation of the bump on the uridine base of UDP-GalNAc demands an additional chemical handle on the GalNAc moiety to label glycosylated substrates (Figure 2-2). However, if the bump is located on the GalNAc moiety, both the bump and the chemical handle can be installed at one position simultaneously. This simplifies the design of the UDP-GalNAc analogs and is more synthetically tractable. Because of these key advantages and lack of major disadvantages, we elected to target the GalNAc pocket in the enzyme active site in our initial mutation design.

2.1.2. Modification of the C8 position of GalNAc

The simultaneous installation of both the bump and chemical handle on the GalNAc moiety is an attractive design approach. In theory, if the ppGalNAcTs can be engineered to accept the corresponding UDP-GalNAc analog, then any position on GalNAc represents a potential modification site. Fortunately, the literature is rich with descriptions of many such UDP-GalNAc analogs with GalNAc-oriented modifications. Examples include 2-azido[46], 3-, 4-, and 6-methoxy[123], 6-biotinyl[124], and 6-azido GalNAc, all of which promisingly abrogate native ppGalNAcT activity[21,46]. Early ppGalNAcT engineering efforts to take advantage of 6-azido substrate orthogonality were unsuccessful[6]. Despite the orthogonality conferred by the modifications of these analogs, other considerations limited us from pursuing them further.

While the orthogonality of these early UDP-GalNAc analogs appeared promising, another major design goal of bump-hole engineering is to minimally perturb the biological system. This additional consideration provided motivation to exclude modifications at several positions around the GalNAc ring. The orientation of the hydroxyl group at the C4 position distinguishes GalNAc from GlcNAc, therefore modification at this position could result in off-target labeling. C4 analogs might disrupt not only ppGalNAcT recognition, but also recognition by enzymes that utilize GlcNAc or glucose[34]. The C3 and C6 positions are necessary for the elaboration of the Tn antigen to complex glycan structure during biosynthesis. Modifications to the C5 position might interfere with glycan elaboration at the C6 position and are also synthetically challenging. These potential perturbing effects dampened our enthusiasm for modification at C3, C4, C5, and C6.

Two major reasons guided us to focus on engineering the C8 position of GalNAc. First, modifications at the C8 position have been shown to be tolerated well by cells[85,125,126], thereby fostering the goal of biocompatibility and increasing the likelihood of achieving good signal to noise. Second, modifications at the C8 position are much more synthetically tractable, allowing the generation of a much larger panel than would be possible if modifying the other positions.

One possible liability of modifying the C8 position of GalNAc is the vulnerability of the N-acyl side chain of hexosamine sugars to enzymatic hydrolysis. Specifically, after GlcNAc is converted to GlcNAc-6-phosphate, the N-acyl side chain can be metabolized and removed to yield glucosamine and acetate[127-129]. The acetate can then be activated on CoA and participate in protein acetylation[127-129]. N-acyl modified GlcNAc analogs can be metabolized in the same fashion, including N-glycolylglucosamine[128] and N-(4-pentynoyl)glucosamine[130](Figure 2-3). Indeed, this alternative metabolic pathway has been used to selectively deliver an alkyne tag to acetylated proteins by feeding cells 1-deoxy-N-pentynyl glucosamine[129].

N-glycolyl glucoasmine

N-(4-pentynoyl)glucoasmine
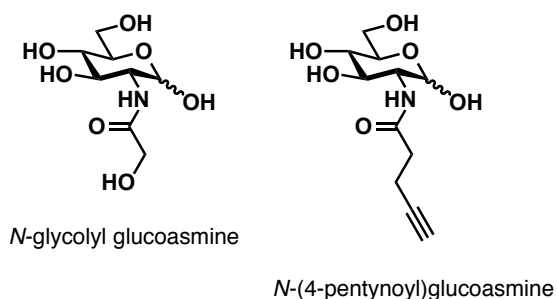
Figure 2-3. GlcNAc analogs known to be de-N-acetylated.

Such a fate is unlikely or irrelevant to our bumped GalNAc analogs. Metabolism of a GalNAc analog to the GlcNAc analog and subsequent de-N-acetylation requires the concerted activity of several enzymes. First, the C4-epimerase, GALE, would need to convert the UDP-GalNAc analog to the corresponding UDP-GlcNAc analog[127,131]. UDP-GlcNAc can be converted into ManNAc and then to both sialic acid and GlcNAc[127,128]. Only after GlcNAc is phosphorylated to form GlcNAc-6-phosphate can the N-acyl side chain be removed[127-129]. Not only would this metabolic fate require the action of many biosynthetic enzymes, but formation of the UDP-GlcNAc analog would also enable labeling of other sugar pathways. Loss of the N-acyl side chain would be one of several possible causes of off-target labeling if the GalNAc analog could be converted to the UDP-GlcNAc analog. Thus, lability of the N-acyl side chain is a negligible concern when selecting a position on GalNAc to install the bump and chemical handle.

2.1.3. Identifying gatekeeper residues

Having resolved to "bump" the C8 position of the GalNAc moiety within UDP-GalNAc, we next needed to identify the amino acid residues to mutate in order to create a corresponding "hole" within the enzyme active site. We turned to

ppGalNAcT crystal structures to inform our selection of a target isoform and residues for mutagenesis. Three ppGalNAcT isoforms, T1, T2, and T10, have been crystallized[115,116,132]. Among these, T1 and T2 are the most ubiquitously expressed isoforms in the embryonic mouse[12,13]. *In vitro* data indicate that T1 and T2 specifically append α-GalNAc to mono- and unglycosylated peptides, whereas T10 binds and modifies only highly glycosylated peptides[12-14]. Together, these data suggest that T1 and T2 may have a more accessible substrate pool, enabling easy identification of transferase acceptor substrates from naked peptide pools.

While murine T1 has been crystallized as an apoenzyme, human T2 (hT2) has been crystallized with UDP and co-crystallized with UDP and the acceptor peptide[116,132]. This crystal structure is an excellent guide for design and makes hT2 a promising first target. Although the hT2 crystal structure is not ideal for selecting residues that interact with the donor substrate due to the absence of the GalNAc moiety, hT10 has been co-crystallized with Mn^{2+}, UDP, and GalNAc[115,116,133]. The high degree of conservation in ppGalNAcT sequences and the structural conservation among glycosyltransferases enable hT0 to complement the h2 structure as a good model for the donor substrate binding site[6,12].
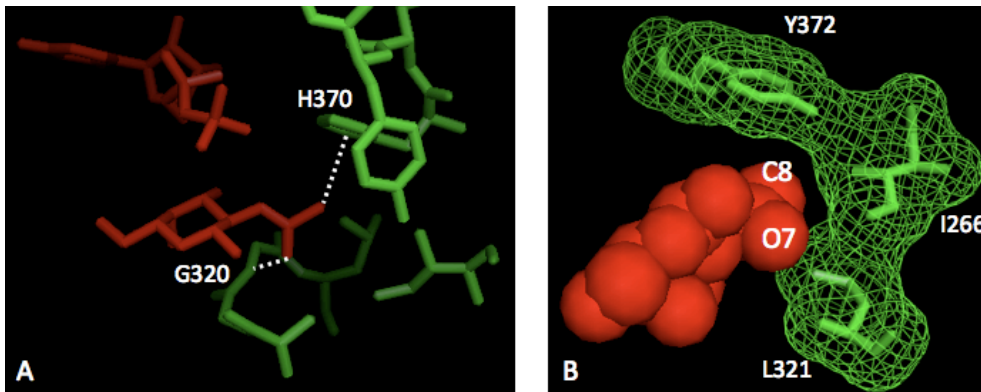
Figure 2-4. Residues within 5 Å of GalNAc. A. Hydrophobic interactions (dashed line) between imidazole ring of H370 and GalNAc C8; hydrogen bonding (dashed line) between backbone NH of G320 and GalNAc O7 B. Gatekeeper residues (green) are promising targets for mutagenesis (PDB ID: 2D7I).

Analysis of crystal structures yields several candidate gatekeeper residues. The crystal structures reveal several residues that interact directly with GalNAc. Two interact with components of the C2 N-acetyl group: His370 with C8 and Gly320 with O7 (Figure 2-4a)[115]. Unfortunately, neither of these residues is a suitable target for mutation. His370 also coordinates Mn^{2+}, which makes it critical for enzyme activity, and Gly320 does not possess a side chain that can be targeted with rational design[115]. Promisingly, our modeling of space-filling interactions in the hT10 active site indicates 3 potential "gatekeeper" residues within 5Å of GalNAc: Ile266, Leu321, and Tyr372 (Figure 2-4b).

Homologous residues in human ppGalNAcT2 were identified by comparing an amino acid sequence alignment of hT10 and hT2. Both I266 and L321 are conserved in T2 as I253 and L310, respectively. Y372 demonstrates moderate conservation as the aromatic residue F361 (Figure 2-5). Among these gatekeeper residues, F361 is the bulkiest and therefore is the most promising single amino acid change to expand the GalNAc binding pocket (Figure 1-6). However, mutations to F361 might interfere with substrate specificity, as it has been shown to interact with the acceptor peptide in the co-crystal of hT2, Mn^{2+}, UDP, and EA2 peptide[116,133]. Specifically, F361 is one of 4 residues that form a hydrophobic pocket where the +3 proline in the peptide sequence binds. ppGalNAcT2 strongly prefers proline at the +3 position of the acceptor sequence, so F361 may be critical for acceptor substrate recognition or in glycosyltransferase activity. We address this concern by developing several mutants: F361A and F361S, as well as I253A, L310A, and the double mutant I253A/L310A.

```
 T9  --LPAIDNI--   --IGCSFVV--   --AHIERTR--
 T8  --SPVFDNI--   --MG-ILAA--   --AHLERHH--
T10  --CPMIDVI--   --AGGLFAV--   --GHIYRKY--
 T7  --VPLIDVI--   --AGGLFAI--   --GHIYRLE--
T11  --CPVIDII--   --AGGLFAM--   --GHIFRKR--
 T5  --CPVIEVI--   --AGGLFSI--   --GHIFRND--
T15  --SPVIDVI--   --PGEVVAM--   --GHIYQNQ--
 T3  --SPDIASI--   --AGGLFSI--   --GHVFRSK--
 T6  --SPDIVTI--   --AGGLFSI--   --GHVFRTK--
 T4  --CPVIDTI--   --AGGLFAV--   --GHVFPKR--
T12  --CPVIDVI--   --AGGLFAV--   --GHVFPKQ--
 T1  --CPIIDVI--   --AGGLFSI--   --GHVFRKA--
T13  --CPIIDVI--   --AGGLFSI--   --GHVFRKA--
 T2  --SPIIDVI--   --AGGLFVM--   --GHVFRKQ--
T14  --CPVIDII--   --AGGLFVI--   --GHVFRKK--
```

Figure 2-5. Alignment of human ppGalNAcT family members 1-14. Isoforms are grouped according to overall sequence homology using a CLUSTALW alignment. Sequences near proposed mutation sites corresponding to hT2 positions I253, L310, and F361 (bold) are shown.

Among human ppGalNAcT isoforms 1-14, the amino acid corresponding to position 361 in T2 is conserved almost exclusively as either a Phe or a Tyr (Figure 2-5)[116]. Therefore, Ala and Ser were selected to replace F361 because these smaller residues emulate the nonpolar or hydrogen bond donating character, respectively, and should maintain a similar electronic environment while expanding the active site by occupying less space. Similarly, Ala mutants I253A, L310A, and I253A/L310A maintain the nonpolar character of the original amino acids, while occupying less space in the active site.
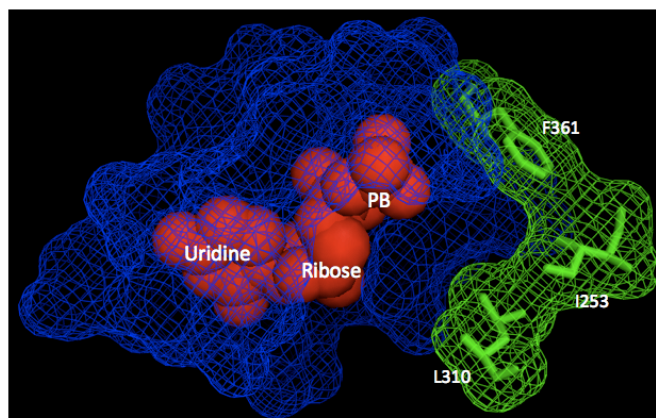
Figure 2-6. Gatekeeper residues in ppGalNAcT2. Space-filling representation of residues within 4 Å of UDP (blue mesh) with gatekeeper residues (green mesh) and UDP (red). PDB ID: 2FFU.

2.2. Development of ppGalNAcT expression constructs

We set out to engineer ppGalNAcT2 using structure-guided rational design. To this end, we developed a panel of hT2 mutants for *in vitro* studies. We also cloned ppGalNAcT1 and the analogous mutants. The T1 panel will aid in evaluating whether trends in bump-hole reactivity are easily translatable between ppGalNAcT family members. Promising results with an hT2 double mutant and UDP-GalNAc analog pair (Chapter 4) inspired us to clone the transferases hT7 and hT10 and generate the wildtype and double mutant in both family members. The glycopeptide-preferring ppGalNAcT isoforms 7 and 10 are considered late transferases, and are uniquely suited to studies with bump-hole engineering.

2.2.1. Secretion constructs of hT2

Tabak and colleagues previously developed secretion constructs of several ppGalNAcT isoforms for expression in cultured mammalian cells[13,134]. The Tabak laboratory generously provided full-length wildtype human ppGalNAcT2 in the pCMV NTAP vector. pCMV NTAP contains an N-terminal FLAG and 6x His tags as well as a kanamycin/neomycin resistance cassette for selection. We began developing mammalian expression protocols for COS-7 and HEK 293T cells using this construct. We evaluated protein expression in conditioned media and cell lysates by SDS-PAGE and α-FLAG Western blotting. Concurrently, we generated the active site mutants F361S and F361A by site-directed mutagenesis.

The codons selected for the alanine (GCC) and serine (TCC) mutations are based on the frequency of codon use in the human genome[135]. Initial efforts to generate the mutants F361A and F361S according to the QuikChange™ site-directed mutagenesis protocol using fully overlapping primers were unsuccessful. However, a modified protocol described in Zheng *et al*. (2004) using staggered forward and reverse

primers yielded positive clones of F361S and F361A[136]. This modified QuikChange™ protocol was used for all point mutations generated by site-directed mutagenesis.

To facilitate the expression of secreted proteins, full-length wildtype and mutant hT2 were cloned into the commercial mammalian secretion vector p3xFLAG-CMV™-8, which contains an N-terminal preprotrypsin leader sequence and 3xFLAG tag and an ampicillin resistance cassette. We then generated soluble hT2 by removing the N-terminal transmembrane domain from hT2 WT, F361A, and F361S. The soluble domains of ppGalNAcT2 include both the catalytic and lectin domains, and our truncation constructs were designed based on work by Tabak and coworkers[137,138].

Early experiments expressing these full-length and truncated hT2 constructs in COS-7 and HEK 293T cells indicated that, as anticipated, truncated hT2 is expressed, secreted, and purified more efficiently than the full-length construct. Full-length hT2 should not be secreted because the transmembrane domain inhibits secretion and solubility. Therefore, we focused on optimizing the purification of truncated hT2 from conditioned media by α-FLAG immunoprecipitation. We found that proteins were not readily purified with the 3xFLAG tag due to poor elution from the α-FLAG resin.

To improve the elution of constructs during FLAG purification, we chose to move them into a vector with a single FLAG tag. We cloned the truncated hT2 WT, F361A, and F361S constructs from p3xFLAG-CMV™-8 into pFLAG-Myc-CMV™-19, which contains an N-terminal preprotrypsin leader sequence and FLAG tag, as well as an ampicillin resistance cassette. I253A, L310A, and I253A/L310A were generated via site-directed mutagenesis from WT truncated hT2 in pFLAG-Myc-CMV™-19. We confirmed that active transferases were generated as described in Chapter 4.

Figure 2-7. Design of ppGalNAcT secretion constructs.

After the hT2 I253A/L310A mutant demonstrated promising reactivity trends with UDP-GalNAc and analogs, we were interested in evaluating its activity *in cellulo*. Thus, we designed full-length hT2 wildtype and double mutant constructs for transient, intracellular expression. We did not possess full-length constructs of I253A/L310A, so we amplified the full-length wildtype hT2 from the original construct provided by the Tabak lab and cloned it into the pFLAG-CMV™-2 vector. The pFLAG-CMV™-2 vector is designed for intracellular expression of proteins and contains an N-terminal FLAG tag and an ampicillin resistance cassette. We then generated the double mutant by two rounds of site-directed mutagenesis as described above. However, we recently discovered a mutation at the N-terminus of the WT and double mutant constructs that arose by mis-priming when the full-

length gene was first cloned into pFLAG-CMV™-2. We are currently correcting this mutation in both constructs.

2.2.2. Other ppGalNAcT family members

To determine whether the bump-hole technology could be extended to other ppGalNAcT family members, we generated a panel of ppGalNAcT1 mutants homologous to the hT2 panel. Human T1 was selected because it is also considered an early transferase, but it is evolutionarily divergent from hT2[63]. The N-terminal truncation of hT1 was designed based on online bioinformatics tools to predict the transmembrane region, as well as previously published T1 truncation mutants[138]. We used hT1 nucleotide sequence with accession number X85018 in the NCBI database. We synthesized the soluble portion of wildtype ppGalNAcT1 by modified Golden Gate assembly using gene segments (gBlocks) that were made commercially[139]. The truncated wildtype gene was cloned into pFLAG-Myc-CMV™-19. The mutants (F346A, F346S, I238A, L295A, and I238A/L310A) are homologous to the hT2 mutants and were generated using site-directed mutagenesis.

Golden Gate assembly relies on the generation of unique sticky ends among all components to be assembled in the final plasmid. The target vector is subjected to restriction digest with NotI and EcoRI to generate sticky ends for directional cloning of the synthesized gene. The gBlocks are designed to create unique sticky ends upon BsaI cleavage so that the gBlocks can only assemble in one orientation. The 5' end of the first gBlock (5'-gBlock1) generates a NotI overhang for annealing with the digested vector. The 3' end of gBlock1 (3'-gBlock1) generates an overhang complementary to 5'-gBlock2, and 5'-gBlock2 generates an overhang complementary to 3'-gBlock1. 3'-gBlock2 generates an overhang complementary to 5'-gBlock3, and 5'-gBlock3 generates an overhang complementary to 3'-gBlock2. The assembly of internal gBlocks can proceed indefinitely in this manner. The 3' end of the final gBlock generates an EcoRI overhang for annealing with the digested vector.
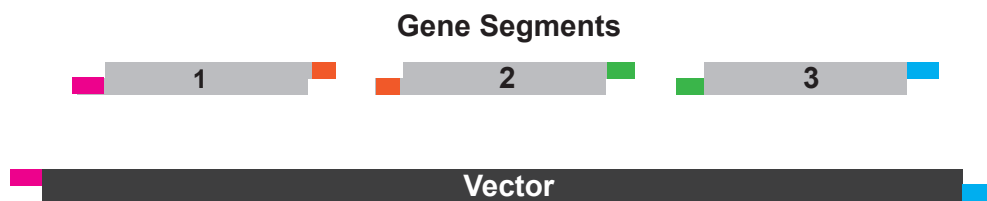
Figure 2-8. Gene assembly by Golden Gate cloning.

Exploring the substrate specificities of late transferases represents a uniquely powerful application of our bump-hole technology. Because the glycoprotein substrates on which they act are heavily glycosylated, it is extremely difficult to confirm late ppGalNAcT activity *in vivo*. Thus, we synthesized truncated and full length variants of ppGalNAcT7 and T10 for both the wildtype and double mutant

genes. The double mutants of hT7 (I330A/L391A) and hT10 (I266A/L321A) are homologous to the hT2 double mutant.

Based on the synthesis of human T1, we synthesized the genes for human T7 and T10 by Golden Gate assembly using gBlocks. The nucleotide sequences in the NCBI database were referenced to synthesize hT7 (AJ002744) and hT10 (AJ_505950). The N-terminal truncations were made by cleaving at the end or middle of the disordered domain that follows the transmembrane domain. These domains were identified with online protein secondary structure predictors, and published truncations were also referenced for T7[140] and T10[115,141,142].

GBlock gene segments were used to synthesize both genes in a modular fashion, such that combinatorial assembly of different gBlocks generates truncated and full-length and wildtype and double mutant alleles in parallel. With different combinations of gBlocks, we cloned 2 full-length and 2 truncated constructs corresponding to the wildtype and double mutant of each gene, producing 8 constructs in total of hT7 and hT10. The truncated constructs were cloned into the secretion plasmid pFLAG-Myc-CMV™-19 vector, while the full-length constructs were cloned into the pFLAG-CMV™-2 vector. The soluble constructs are intended for *in vitro* characterization, and full-length constructs will be utilized for experiments *in cellulo*.

Experiments with the full-length constructs of T7 and T10 including immunocytochemistry and cellular fractionation indicated that the N-terminal FLAG might have been disrupting proper localization of the full-length enzymes. This interference should have been anticipated, as most work with membrane-resident Golgi enzymes uses a C-terminal tag to minimize interference with the protein's natural signal sequence. In the case of the ppGalNAcTs, the transmembrane domain serves as the signal sequence. Although the naturally-occurring protein has basic residues upstream of the transmembrane domain (as the FLAG tag is also), these residues reside on the cytoplasmic side of the Golgi and may be involved in binding to chaperones or tethering proteins that help maintain the proper position of the enzyme within the Golgi stack[63]. Because the FLAG tag is upstream of these residues, it could interfere with this or other N-terminally directed methods of properly localizing the enzyme to the Golgi.

We relocated the FLAG tag in all of the full-length constructs from the N- to the C-terminus. We employed a PCR-based strategy for insertions and deletions that uses the same principles as site-directed mutagenesis[143]. We first removed the FLAG tag by amplifying the entire plasmid and gene excluding the FLAG tag. We then introduced the FLAG tag at the C-terminus of the hT2 gene in a similar fashion using overhanging primers. The wildtype and double mutant of full-length T7 and T10 were all re-cloned in parallel. We will use the same protocol to relocate the FLAG tag in the full-length hT2 constructs.

2.3. Preparation of secreted transferase soluble domains

We have optimized a protocol for expression and purification of active transferases from Human Embryonic Kidney (HEK) 293T cells. As described below, the protocol is initiated by transient transfection, followed by 48 hours of protein expression, clarification of the conditioned medium, and α-FLAG immunoprecipitation. All 12 hT2 and hT1 constructs have been purified by this methodology. As described in Chapter 4, the activity of the purified proteins has been validated with the model acceptor peptide EA2 and UDP-GalNAc and/or UDP-GalNAz (**2.1**) for 8 of the 12 proteins.
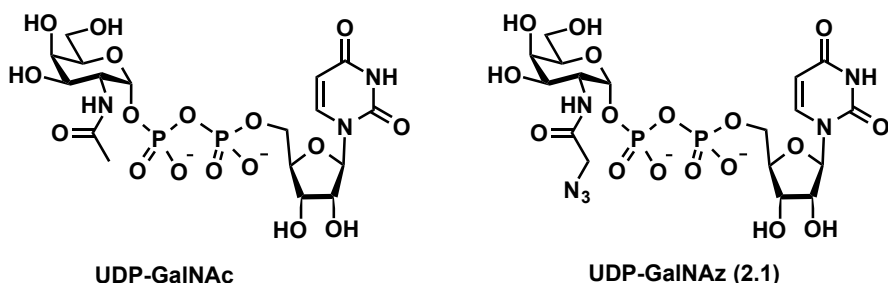
Figure 2-9. UDP-GalNAc and azide-modified analog UDP-GalNAz are ppGalNAcT substrates.

2.3.1. Mammalian transfection and purification

Twenty-four hours prior to transfection, HEK293T cells are split at a density that achieves 60 – 70% confluence when the cells are transfected. To achieve comparable cell densities on all experimental plates, several confluent plates are treated with trypsin, the single-cell suspensions are pooled, and this suspension is divided evenly among seven 10-cm plates (1 plate per ppGalNAcT construct and 1 mock transfection). Transfection is carried out the following day according to the manufacturer's protocol. We found that it is challenging to reproducibly express the wildtype and mutant enzymes at the same concentrations. It is important to achieve equal expression of the secreted protein because protein concentrations are low and are therefore difficult to quantitate accurately. To improve confidence that the same quantities of enzymes are compared in our activity screens, we focused on reproducing our early experiments in which all constructs were expressed at equal levels. These successful prior experiments demonstrated that uneven expression levels are not intrinsic to the mutants themselves. We found that transfecting plates of HEK293T cells at equal densities with high quality DNA dissolved at the same concentration in the stock solutions (~1 μg/μL) gives equal expression of all mutants consistently.

We found that when preparing large amounts of DNA from *E. coli*, as is necessary for mammalian expression experiments, the precipitated DNA pellet isolated at the final step is often difficult to re-solubilize completely. The pellet is glassy and difficult to see, and can be easily over-dried after precipitation, resulting in non-homogenous

31

DNA solutions. We hypothesized that larger DNA pellets generated by Qiagen Maxiprep (300 – 500 µg DNA total) were more difficult to solubilize than the DNA pellets generated by Midiprep (75 – 100 µg DNA total). We found that by dividing the eluted DNA from a Maxiprep prior to isopropanol precipitation (or by using Midipreps), and by drying the pellets for no longer than 10 minutes in air, the smaller pellets were dissolved readily. We also diluted the DNA samples to the same concentration (based on the lowest concentration sample), to ensure that different mutants had identical conditions. These modifications resulted in excellent reproducibility among mutants and across experiments. Good protein production by the HEK 293T cells is correlated with cells detaching from the monolayer on the surface of the plate. At 48 h post-transfection, a large number of cells are floating, resembling cells that have been treated with trypsin. This may reflect cytotoxicity due to protein expression, or changes in the cell surface and/ or extracellular matrix due to ppGalNAcT overexpression. We saw this phenomenon with all mutants, including those that did not appear to be active *in vitro*.

Our protocol recovers soluble proteins from the conditioned media, so we changed the media 24 h post-transfection to remove any transfection agent. This was to minimize possible contamination with the transfection reagent, which might increase cell toxicity, affect protein stability or solubility, or lead to poor reproducibility. Early experiments evaluating protein expression levels in which the media was replaced every 24 h for 96 h demonstrated that the highest protein levels were achieved by expressing protein from 24 to 48 h post-transfection. Although slightly higher protein levels could be achieved by expressing protein from 24 to 72 h without a media change at 48 h, cells detached from the plate in large numbers by 48 h post-transfection. We were concerned that cell death in the 48 to 72 h window might result in lower reproducibility and possible protein degradation. Consequently, for the experiments presented in this dissertation, the media was changed 24 h after transient transfection, and the conditioned media containing secreted protein was harvested after 48 h. Proteins were purified in column format using a 1.25 mL bed of α-FLAG agarose resin according to the manufacturer's protocol. Protein samples for all constructs were purified in parallel to minimize variation across ppGalNAcT mutants.
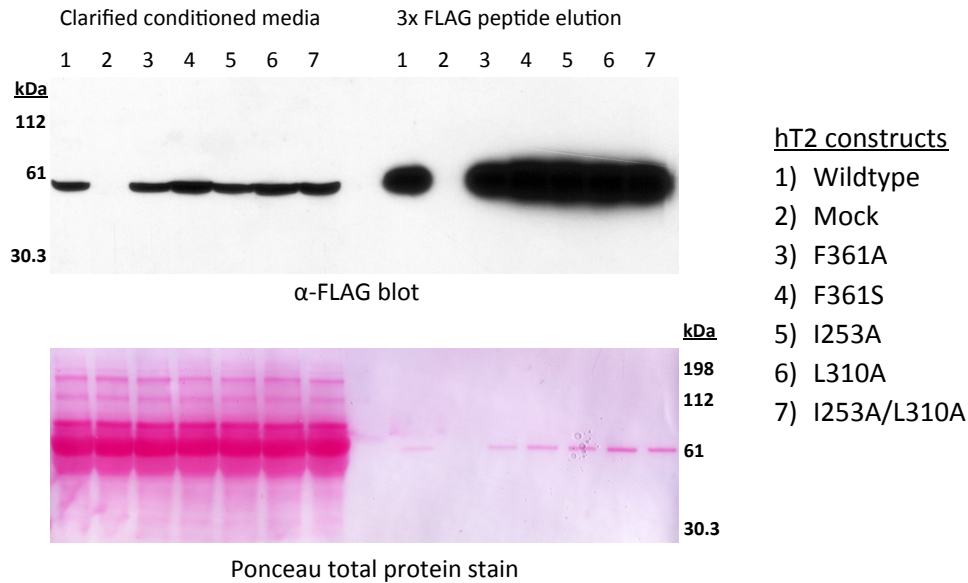
Figure 2-10. ppGalNAcT2 secretion construct expression and purification. Predicted MW after signal peptide cleavage 61.6 kDa.
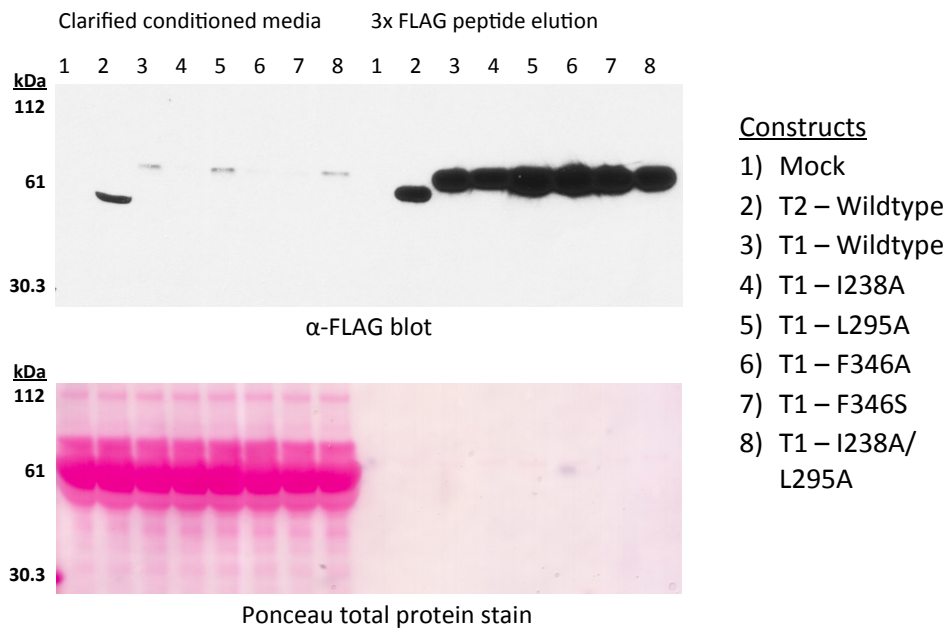
Figure 2-11. ppGalNAcT1 secretion construct expression and purification.

2.3.2. Protein quantitation

After ppGalNAcT α-FLAG immunoprecipitation, we attempted a variety of strategies to quantitate the protein. Because protein was eluted using 3x FLAG peptide, protein

quantitation required correcting for protein concentration due to the presence of the peptide. We first attempted to use a traditional BCA assay comparing the quantity of protein purified from mock-transfected cells to the protein purified from cells expressing ppGalNAcTs. We found that the protein concentration was too low to measure accurately in this fashion. We attempted to use 30 kDa molecular weight spin cutoff filters to concentrate the protein and perform a buffer exchange that would remove the peptide. Despite our attempts to optimize the manufacturer's protocol, as well as attempting purification and concentration on different scales, our overall yields were significantly depleted, presumably caused by protein loss due to adhesion to the filter. We then attempted to utilize a fluorescence-based assay in a very similar fashion to the BCA quantitation experiments. This assay is supposed to be sensitive for lower protein concentrations, but we found that we could not distinguish signal above background when comparing protein from mock transfected versus transfected cells.

We had initially rejected gel-based quantitation assays due to their limited sensitivity. However, we realized that, if adequately sensitive, a gel would allow us to separate the peptide from the ppGalNAcT protein during the quantification analysis itself. Colloidal blue staining has been shown to behave linearly when comparing a protein standard to a protein of interest with a similar molecular weight[144]. A standard curve of albumin was generated to determine the lower limit of detection, which is approximately 15 – 30 ng of protein, and the linear range, which is approximately 60 – 1000 ng of protein. Recombinant, monomeric human albumin is a similar molecular weight to the ppGalNAcTs, and is a good protein standard for colloidal blue quantification of ppGalNAcT expression.
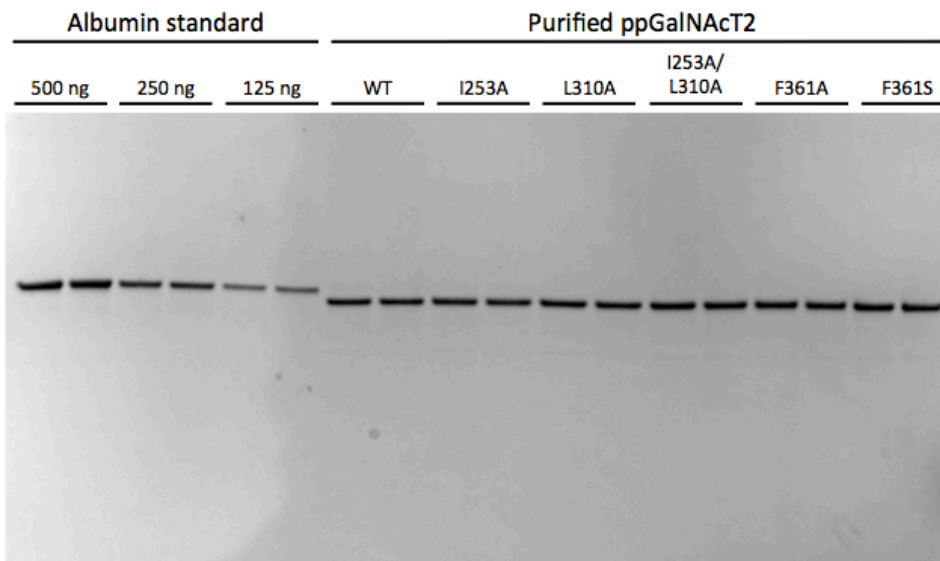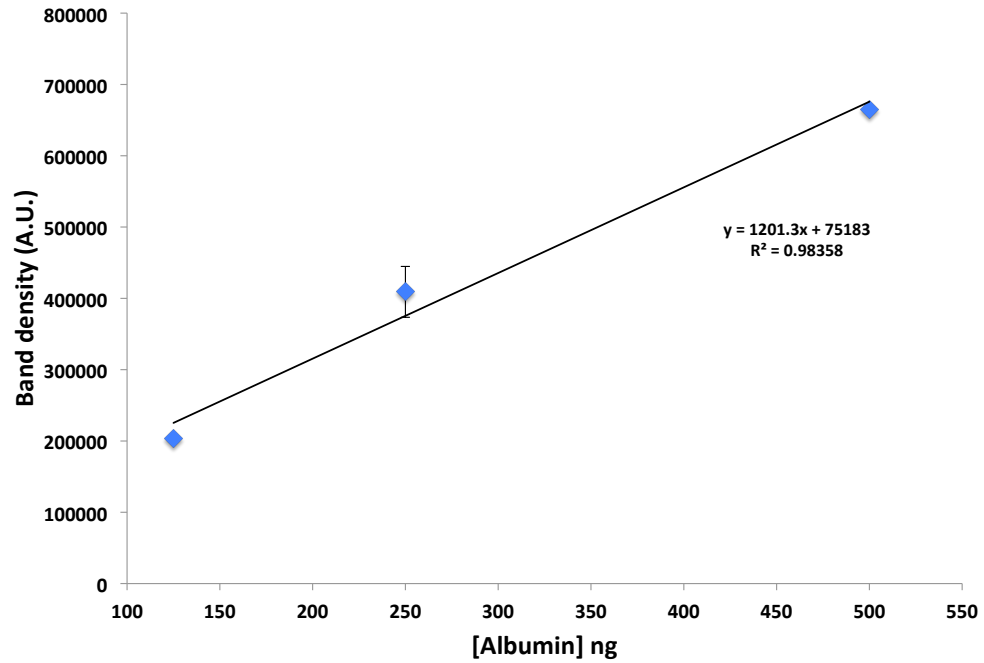
Figure 2-12. Colloidal blue assay for ppGalNAcT protein quantification. A) Proteins visualized with SDS-PAGE and colloidal blue staining. Lanes 1-6 albumin standard at 3 concentrations in duplicate. Lanes 7-18 equal volumes of purified ppGalNAcTs in duplicate. B) Standard curve of albumin concentrations quantified by densitometry.

2.4. Experimental

2.4.1. General

All biological reagents were molecular biology or cell culture grade and from commercial sources, unless otherwise specified. Chemically competent *E. coli* strain One Shot® Top10 and S.O.C. medium were purchased from Invitrogen. Bacteria were grown in LB broth (Lennox L Broth Base, Invitrogen) supplemented with 100 μg/mL ampicillin (LB/amp) or 50 μg/mL kanamycin (LB/kan) and grown on LB/amp or LB/kan agar plates. The Plasmid Midi and Maxi kits were from QIAGEN.

NotI, NotI-HF™, EcoRI, EcoRI-HF™, BsaI, DpnI, Antarctic phosphatase, and T4 DNA ligase were purchased from New England Biolabs (NEB). PfuUltra II Fusion HS DNA polymerase was from Agilent. 1 Kb and 100 bp DNA ladders were from Invitrogen. DNA Clean and Concentrator™-5, Zyppy™ Plasmid Miniprep Kit, and Zymoclean™ Gel DNA Recovery kit were from Zymo Research. Elim Biopharmaceuticals, Inc (Hayward, CA) synthesized all primers and performed all plasmid DNA sequencing. Integrated DNA Technologies (IDT) synthesized all gBlocks.

Dulbecco's Modified Eagle Medium with high glucose (DMEM), Dulbecco's Phosphate-Buffered Saline without Calcium or Magnesium (DPBS), and Penicillin/Streptomycin were purchased from Hyclone. Fetal bovine serum (FBS) was obtained from Omega Scientific and 0.25% trypsin/EDTA and Opti-MEM® Reduced Serum Medium were purchased from Invitrogen. TransIT®-293 Transfection Reagent was obtained from Mirus Bio. COmplete EDTA-free Protease inhibitor cocktail tablets were purchased from Roche. Anti-FLAG M2 Affinity Gel was from Sigma (Cat. A2220).

Criterion™ XT Bis-Tris Precast gels, XT-MES Running Buffer, nitrocellulose membrane (0.45 micron), Precision Plus Protein™ Kaleidoscope™ Prestained Standards (10-250 kDa), and Kaleidoscope™ Prestained Standards (7.6-216 kDa) were obtained from BioRad. SuperSignal West Pico Chemilluminescent Substrate was purchased from Fisher Scientific. Monoclonal α-FLAG M2 antibody was from Sigma (Cat. F1804) and Goat Anti-Mouse Kappa-HRP was from Southern Biotech (Cat. 1050). Recombinant human albumin expressed in S. cerevisiae was from Sigma (Cat. A6608).

MacPyMOL was used to model the crystal structures of ppGalNAcT2 (with and without EA2 peptide) and of ppGalNAcT10 (PDB accession: 2FFU, 2FFV, 2D7I).

2.4.2. General Methods

Manufacturer recommended protocols for Zymo Research kits were subject to the following modifications: Prior to elution, spin columns were dried by centrifugation at maximum speed for 2 min. DNA was eluted with water added directly to the filter, followed by 1 min incubation and 1 min centrifugation at maximum speed).

DNA was quantitated using a Nanodrop 2000 UV-vis spectrophotometer.

DNA samples were pre-mixed for sequencing: 500 ng plasmid DNA, 8 pM primer, 15 μL total in water.

Human Embryonic Kidney (HEK) 293T cells (ATCC) were maintained in a 5 % CO_2, water-saturated atmosphere at 37 °C in DMEM supplemented with 10 % FBS, penicillin (100 U/mL), and streptomycin (0.1 mg/mL). Adherent cell densities were maintained between 1 x 10^5 and 2 x 10^6 cells.

2.4.3. General Protocol 1: Standard PCR conditions

PCR amplifications were carried out in duplicate for all cloning applications as recommended by the manufacturer: 25 ng template DNA, 0.2 μM forward primer, 0.2 μM reverse primer, 250 μM of each dNTP, 3% DMSO, 1x PfuUltra II buffer, and 1 μL PfuUltra II Fusion HS DNA polymerase in a final reaction volume of 50 μL in water.

2.4.3.1. General Protocol 1a: Standard PCR conditions for insert amplification

Thermocycle conditions were 95 °C for 2 min; 30 cycles of 95 °C for 20 sec, annealing temp for 20 sec, 72 °C extension for 15sec/ kb; 72 °C for 3 min; 4 °C hold. Annealing temperatures were generally T_M – 5 °C and are listed with the primers. For primers with 5' overhangs, annealing temperatures are based on the T_M of the portion of the primer designed to anneal to the template. Amplicon size and extension times are listed with the primers.

2.4.3.2. General Protocol 1b: Standard PCR conditions for site-directed mutagenesis

PCR for site-directed mutagenesis was carried out according to modified QuikChange™ site-directed mutagenesis protocols as described by Zheng *et al*. (2004) for point mutations and Liu *et al*. (2008) for insertions and deletions[136,143].

Thermocycle conditions were 95 °C for 2 min; 18 cycles of 95 °C for 20 sec, annealing temperature for 20 sec, 72 °C extension for 15sec/ kb; 72 °C for 3 min; 4 °C hold. For primers with mismatches for point mutations, annealing temperatures were generally 55 °C and are listed with the primers. For primers with 5' overhangs for insertions/ deletions, annealing temperatures were generally T_M – 5 °C corresponding to the portion of the primer designed to anneal to the template. Amplicon size and extension times are listed with the primers.

2.4.4. General Protocol 2: Standard protocol for bacterial cloning

Bacterial transformations for cloning were carried out in One Shot® Top10 *E. coli* according to the manufacturer's instructions as described below. Modifications for

specific applications are in the second paragraph. Bacteria (50 µL) were thawed on wet ice for 10 min, 2.5 – 125 ng DNA was added based on the application, tubes were flicked gently to mix, and *E. coli* were incubated on wet ice for 25 min. *E. coli* were subjected to heat shock at 42 °C for 30 sec, incubated on wet ice 2 min, and 250 µL rt S.O.C. medium was added. *E. coli* were grown shaking at 225 rpm at 37 °C for 35 min (Amp^R) to 45 min (Kan^R). Based on the application, 25-100 µL *E. coli* were spread onto LB agar plates with 100 µg/mL ampicillin or 50 µg/mL kanamycin. Plates were incubated at 37 °C for approximately 16 h and then stored at 4 °C. Individual colonies were picked and used to inoculate 5 mL LB with antibiotics (100 µg/mL ampicillin or 50 µg/mL kanamycin). 5 mL cultures were grown shaking at 225 rpm at 37 °C for approximately 16 h, and cells were pelleted at 3,650 x g at 4 °C for 15 minutes. Supernatant was removed, pellets were flicked and then suspended in 600 µL water. DNA was extracted using a Zyppy™ Plasmid Miniprep Kit according to the manufacturer's protocol with modifications as described in the General Methods (30 µL elution).

Quantities of DNA added for transformations and the volume of transformed *E. coli* plated varied based on the application. When plasmid DNA was previously purified from *E. coli*, 0.25 µL (25-125 ng DNA) was used for transformation, and 25-35 µL *E. coli* were plated. When plasmid DNA was prepared via ligation, 2 µL ligation (2.5 ng DNA) was used for transformation, and 100-200 µL *E. coli* were plated. When plasmid DNA was prepared via PCR for site-directed mutagenesis, 2.5 ng DNA (0.25-5 µL) was used for transformation, and 100-200 µL *E. coli* were plated.

2.4.5. General Protocol 3: Standard protocol for restriction digest cloning

Insert DNA was amplified as described in General Protocol 1a. Target vectors were subjected to restriction digest (2.5 µg plasmid DNA, 1x NEB buffer, 1x BSA as necessary, 1 µL of each restriction enzyme in a 50 µL digest, 2 h at 37 °C,). Both the vector and insert were isolated with a 0.8% agarose gel and purified with a Zymoclean™ Gel DNA Recovery kit according to the manufacturer's protocol with modifications as described in the General Methods (12 µL elution). The insert was subjected to restriction digest (volume of purified PCR product, 1x NEB buffer, 1x BSA as necessary, 1 µL of each restriction enzyme in a 50 µL digest, 2 h at 37 °C,). The vector was treated with Antarctic phosphatase (volume of digested DNA, 1x Antarctic phosphatase buffer, 2 U (.4 µL) Antarctic phosphatase, and a 20 µL final volume in water, 1 h at 37 °C). Vector and insert were purified before ligation with DNA Clean and Concentrator™-5 according to the manufacturer's protocol with modifications as described in the General Methods (10 µL elution). A 3:1 molar ratio of insert to vector (25 ng vector) were ligated (1x NEB T4 DNA ligase buffer, 0.5 µL (1000 U) NEB T4 DNA, and a 10 µL final volume in water, 30 min at rt). Bacteria were transformed and clones were isolated as described in General Protocol 2.

2.4.6. General Protocol 4: Vector preparation for Golden Gate assembly

Target vectors were digested with NotI and EcoRI (2.5 µg plasmid DNA, 1x NEB4 buffer, 1x BSA, 1 µL (20U or 100 U) NotI-HF™, 1 µL (20U or 100 U) EcoRI-HF™ in a 50 µL digest, 2 h at 37 °C). The vector was isolated with a 0.8% agarose gel and purified with a Zymoclean™ Gel DNA Recovery kit according to the manufacturer's protocol with modifications as described in the General Methods (12 µL elution). The vector was treated with Antarctic phosphatase (volume of digested DNA, 1x Antarctic phosphatase buffer, 0.4 µL (2 U) Antarctic phosphatase in a 20 µL final volume in water, 1 h at 37 °C) and was purified with DNA Clean and Concentrator™-5 according to the manufacturer's protocol with modifications as described in the General Methods (10 µL elution).

2.4.7. General Protocol 5: SDS-PAGE

Samples were prepared by adding 4x SDS protein loading buffer (final concentrations 1x sodium dodecyl sulfate, 2.5 % 2-mercaptoethanol) and denatured by boiling at 95 °C for 5 min. Samples were centrifuged at maximum speed for 15 sec and loaded onto a 12 % or 4-12 % Bis-Tris precast gel. Electrophoresis was carried out in XT-MES buffer at 180 V for approximately 45 minutes. Gels were rinsed briefly with water following electrophoresis.

2.4.8. General Protocol 6: Western blotting

For western blotting, proteins were transferred from an SDS-PAGE gel to a nitrocellulose membrane in Tris/Glycine buffer (BioRad, 20% MeOH) at 50V for 2 h. The membrane was incubated in Ponceau total protein stain (0.2% Ponceau, 3% acetic acid) for 5 min, rinsed with water, and imaged. The membrane was blocked with 5% non-fat powdered milk in TBST (25 mM Tris-Cl pH 7.4, 150 mM NaCl, 0.1% TWEEN-20) mixing at rt for 30 min. The blocking solution was discarded, α-FLAG M2 primary antibody (diluted 1:1000 in 3% milk/TBST) was added, and it was incubated mixing overnight at 4 °C. The blot was rinsed 3 x with TBST for 10 min and goat α -mouse light chain-HRP secondary antibody conjugate (diluted 1:5000 in 3% milk/TBST) was added. After incubating mixing at rt for 1 h, the blot was washed 3 x with TBST for 10 min. The blot was imaged with an HRP detection kit and film was exposed to the blot for 1 sec, 10 sec, 1 min, and 10 min.

2.4.9. ppGalNAcT2 cloning

2.4.9.1. Site-directed mutagenesis to generate F361A and F361S

Full-length human ppGalNAcT2 in pCMV NTAP (Kan[R]) was generously provided by the L. Tabak Laboratory (NIH, Bethesda, MD). Sequencing in that plasmid was carried out with T3F, T3R, T7F, and T7R (Table 2-1). F361A and F361S were generated via site-directed mutagenesis. PCR amplifications were carried out as in General Protocol 1b except PCR reactions contained 50 µM of each dNTP (200 µM total), and primers are in Table 2-2. PCR reactions were digested with 1 µL (20 U)

DpnI at 37 °C for 1 h, and duplicate reactions were combined (100 μL total) and purified. F361S reactions were purified with DNA Clean and Concentrator™-25 (50 μL elution) and F361A reactions were purified with DNA Clean and Concentrator™-5 (10 μL elution) according to the manufacturer's protocol with modifications as described in the General Methods. Bacteria were transformed and clones were isolated as in General Protocol 2. Clones were sequenced with primers in Table 2-1 as in the General Methods.

Primer ID	Sequence	Annealing Temp (°C)
T3F	5'- aattaaccctcactaaaggg	50
T3R	5'- ccctttagtgagggttaatt	50
T7F	5'- gccctatagtgagtcgtattac	50
T7R	5'- gtaatacgactcactatagggc	50

Table 2-1. Sequencing primers for pCMV NTAP.

Primer ID	Sequence	Annealing Temp (°C)
S361stag_F	5'- GACACGTGTcCCGGAAGCAGCAC	55
S361stag_R	5'- CTTCCGGgACACGTGTCCCAC	55
A361long_F	5'- GGACACGTGgcCCGGAAGCAGCACCCCTACACGTTC	55
A361long_R	5'- GCTTCCGGgcCACGTGTCCCACACGGCTGCAC	55

Table 2-2. Point mutagenesis primers for hT2 F361A and F361S. Lowercase letters indicate the site of base pair mismatch. S361stag_F and S361stag_R in PCR reactions were each 0.8 μM. PCR amplicon is 6.2 kb (1 min 45 sec extension, mistakenly 6 min).

2.4.9.2. Secretion constructs 1.0: p3xFLAG-CMV-8 cloning and generation of truncated constructs

Full-length hT2 WT, F361A, F361S were transferred from pCMV NTAP to p3xFLAG-CMV™-8 (Amp^R). These clones were generated as in General Protocol 3, hT2 in pCMV NTAP was the insert template, and p3xFLAG-CMV™-8 (4.8 kb) was the target vector. Primers for insert amplification are in Table 2-3. p3xFLAG-CMV™-8 was subjected to restriction digest with NotI and EcoRI (1x NEB3 buffer, 1x BSA, 1 μL (20U or 100 U) NotI in a 50 μL digest, 2 h at 37 °C, addition of 1 μL (20U or 100 U) EcoRI and 2 h at 37 °C). Insert was subjected to restriction digest with NotI and EcoRI (1x NEB3 buffer, 1x BSA, 1 μL (20U or 100 U) NotI, 1 μL (20U or 100 U) EcoRI). Clones were sequenced with primers in Table 2-4 as in the General Methods. The same protocol was used to clone soluble truncations of hT2 WT, F361A, and F361S from pCMV NTAP into p3xFLAG-CMV™-8. Truncations were based on published hT2 truncations[137,138].

Primer ID	Sequence	Annealing Temp (°C)
CMV9_hT2_F2	5'- gacaagcttgcggccgcgATGCGGCGGCGCTCGCG	68
CMV9_hT2_R2	5'- cgatgaattcCTACTGCTGCAGGTTGAGC	68
hT2trunc_F	5'- gacaagcttgcggccgcgGAGGACTGGAATGAAATTG	54

Table 2-3. Cloning primers for hT2 in p3xFLAG-CMV™-8. hT2 sequence is uppercase and plasmid sequence is lowercase. CMV9_hT2_F2 and CMV9_hT2_R2 were used to generate the full length constructs and hT2trunc_F and CMV9_hT2_R2 were used to generate the trunc constructs. PCR amplicon is 1.8 kb for the full-length gene and 1.6 kb for the truncated gene (45 sec extension). Primers accidentally named CMV9 instead of CMV8.

Primer ID	Sequence	Annealing Temp (°C)
CMV30short_F	5'- aatgtcgtaataaccccgccccg	53
CMV24_R	5'- tattaggacaaggctggtgggcac	53
CMV30_F	5'- aatgtcgtaataaccccgccccgttgacgc	60
CMV24new_R	5'- ctccagtgcccaccagccttgtcctaata	59

Table 2-4. Sequencing primers for p3xFLAG-CMV™-8, pFLAG-Myc-CMV™-19, and pFLAG-CMV™-2. The sequencing primers CMV30short_F and CMV30_F were used interchangeably, as were CMV24_R and CMV24new_R.

2.4.9.3. Secretion constructs 2.0: pFLAG-Myc-CMV-19 cloning

Truncated hT2 WT, F361A, and F361S constructs were transferred from p3xFLAG-CMV™-8 into pFLAG-Myc-CMV™-19 (Amp[R]). Clones were generated using General Protocol 3, although instead of insert amplification by PCR, truncated hT2 (1.6 kb) was removed from p3xFLAG-CMV™-8 by restriction enzyme digest with NotI and EcoRI (2.5 µg plasmid DNA, 1x NEB4 buffer, 1x BSA, 0.4 µL (8 U or 40 U) NotI-HF, 0.4 µL (8 U or 40 U) EcoRI-HF in a 20 µL digest). pFLAG-Myc-CMV™-19 (4.8 kb) was the target vector, and it was restriction enzyme digested under the same conditions as the insert. The transformants were screened by colony PCR using primers in Table 2-4, and promising clones were sequenced with the same primers as in the General Methods.

2.4.9.4. Site-directed mutagenesis to generate I253A, L310A, and I253A/L310A

I253A, L310A, and I253A/L310A were generated via site-directed mutagenesis from WT truncated hT2 in pFLAG-Myc-CMV™-19 (Amp[R]). PCR amplifications were carried out as in General Protocol 1b, and primers are in Table 2-5. PCR reactions were digested with 1 µL (20 U) DpnI at 37 °C for 1 h, and duplicate reactions were combined (100 µL total) and purified with DNA Clean and Concentrator™-5 according to the manufacturer's protocol with modifications as described in the General Methods (10 µL elution). Bacteria were transformed and clones were isolated as in General Protocol 2. Clones were sequenced with primers in Table 2-4 as in the General Methods. I253A/L310A was generated from the I253A clone with the L310A primers.

Primer ID	Sequence	Annealing Temp (°C)
I253A_F	5'- CACCCATCgcCGATGTCATTAATATGGACAAC	55
I253A_R	5'- GACATCGgcGATGGGTGACACAACCCGAGTC	55
L310A_F	5'- GCTGGTGGGgccTTTGTGATGGATAAGTTC	55
L310A_R	5'- CATCACAAAggcCCCACCAGCAATCATGGGG	55

Table 2-5. Mutagenesis primers for hT2 I253A, L310A, and I253A/L310A. Lowercase letters indicate the site of base pair mismatch. PCR amplicon is 6.4 kb (1 min 45 sec extension, mistakenly 6 min).

2.4.10. Synthesis of human T1, T7, and T10 and point mutagenesis

For all secretion constructs, N-terminal truncations were made by cleaving at the middle or end of the stem region that follows the transmembrane domain, as identified by online protein secondary structure predictors GlobPlot (2.3) and HMMTOP (2.0), and published truncations were also referenced. Point mutations homologous to those in hT2 were selected based on a multiple sequence alignment (CLUSTALW) of ppGalNAcT isoforms 1-14.

2.4.10.1. Golden Gate assembly of wildtype ppGalNAcT1

We synthesized the gene for truncated human T1 by a modified Golden Gate assembly protocol using gBlocks (IDT) and cloned it into pFLAG-Myc-CMV™-19. The published nucleotide sequence in the NCBI database for hT1 X85018 was used. Published constructs were referenced for N-terminal truncations of hT1[138].

The original design for gene assembly uses gBlocks with NotI and EcoRI restriction sites at the 5'- and 3'- ends of the gene, respectively (Table 2-6). These gBlocks (GB1 and GB4) were subjected to restriction digest prior to Golden Gate assembly, such that after the gene was assembled, it could be ligated to the digested vector to form the desired plasmid. This approach was unsuccessful. Instead, following Golden Gate assembly, PCR was used add BsaI sites to the 5'- and 3'- ends of assembled gene. The amplified gene was digested with BsaI to produce NotI and EcoRI overhangs and was successfully ligated to the digested vector.

gBlock ID	Sequence
GB1.ppGalNAc-T1	GACAAGCTTGCGGCCGCGGATGAAAAAAAGGAGAGAGGACTTCCTGC TGGAGATGTTCTAGAGCCAGTACAAAAGCCTCATGAAGGTCCTGGAG AAATGGGGAAACCAGTCGTCATTCCTAAAGAGGATCAAGAAAAGATG AAAGAGATGTTTAAAATCAATCAGTTCAATTTAATGGCAAGTGAGAT GATTGCACTCAACAGATCTTTACCAGATGTTAGGTTAGAAGGGTGTA AAACAAAGGTGTATCCAGATAATCTTCCTACAACAAGTGTGGTGATT GTTTTCCACAATGAGGCTTGGAGCACACTTCTGCGAACTGTCCATAGT GTCATTAATCGCTCACCAAGACACATGATAGAAGAAATTGTTGAGACC TGGTGTG
GB2.ppGalNAc-T1	GACAGGAGGTCTCATTGTTCTAGTAGATGATGCCAGTGAAAGAGACT TTTTGAAAAGGCCTTTAGAGAGTTATGTGAAAAAACTAAAAGTACCA GTTCATGTAATTCGAATGGAACAACGTTCTGGATTGATCAGAGCTAG

	ATTAAAAGGAGCTGCTGTGTCTAAAGGCCAAGTGATCACCTTCCTGGA TGCCCATTGTGAGTGTACAGTGGGATGGCTGGAGCCTCTCTTGGCCAG GATCAAACATGACAGGAGAACAGTGGTGTGTCCCATCATCGATGTGAT CAGTGATGATACTTTTGAGTACATGGCAGGCTCTGATATGACCTATGG TGGGTTCAACTGGAAGCTCAATTTTCGCTGGTATCCTGTTCCCCAAAG AGAAATGGACAGAAGGAAAGGTGATCGGACTCTTCCTGTCAGGACAC CTACCATGGCAGGAGGCCTTTTTTCAATAGACAGAGATTACTTTTGAG ACCTGGTGTG
GB3.ppGalNAc-T1	GACAGGAGGTCTCACTTTCAGGAAATTGGAACATATGATGCTGGAAT GGATATTTGGGGAGGAGAAAACCTAGAAATTTCCTTTAGGATTTGGC AGTGTGGAGGAACTTTGGAAATTGTTACATGCTCACATGTTGGACAT GTGTTTCGGAAAGCTACACCTTACACGTTTCCAGGAGGCACAGGGCAG ATTATCAATAAAAATAACAGACGACTTGCAGAAGTGTGGATGGATGA ATTCAAGAATTTCTTCTATATAATTTCTCCAGGTGTTACAAAGGTAGA TTATGGAGATATATCGTCAAGAGTTGGTCTAAGACACAAACTACAAT GCAAACCTTTTTCCTGGTACCTAGAGAATATATATCCTGATTCTCAAA TTCCACGTCACTATTTCTCATTGGGAGAGATACGAAATGTGGAAACGA ATCAGTGTCTAGATAACATGGCTAGAAAAGAGAATGAAAAAGTTGGA TGAGACCTGGTGTG
GB4.ppGalNAc-T1	GACAGGAGGTCTCATGGAATTTTTAATTGCCATGGTATGGG**T**GGTAA TCAGGTTTTCTCTTATACTGCCAACAAAGAAATTAGAACAGATGACCT TTGCTTGGATGTTTCCAAACTTAATGGCCCAGTTACAATGCTCAAATG CCACCACCTAAAAGGCAACCAACTCTGGGAGTATGACCCAGTGAAATT AACCCTGCAGCATGTGAACAGTAATCAGTGCCTGGATAAAGCCACAGA AGAGGATAGCCAGGTGCCCAGCATTAGAGACTGCAATGGAAGTCGGTC CCAGCAGTGGCTTCTTCGAAACGTCACCCTTCCAGAAATATTCTGAGA ATTCATCGATAG

Table 2-6. gBlocks to synthesize wildtype ppGalNAcT1. Sequences are listed 5' – 3'. Silent mutations are **bold**.

Corresponding gBlock	Primer ID	Sequence	Annealing Temp (°C)
GB1	GBT1-1.F1	cacaccaggtctcaGGCCGCGGATGAAAAAAAGGAGA GAGGACTT	53
GB4	GBT1-4.R1	cacaccaggtctctAATTCTCAGAATATTTCTGGAAG GGTGACGT	53

Table 2-7. Primers for extension of gBlocks 1 and 4. Amplicon <1 kb (15 sec extension). Annealing portion of primers uppercase.

The Golden Gate assembly reaction was carried out as follows: 2.375 μL (20 fmol) GB1, 0.6 μL (20 fmol) GB2, 0.6 μL (20 fmol) GB3, 2.145 μL (20 fmol) GB4, 1x NEB T4 DNA ligase buffer, 1 μL (2000 U) NEB T4 DNA ligase, and 1 μL (10 U) BsaI in a final reaction volume of 15 μL in water. Thermocycle conditions used to enhance reaction yield were as follows: 25 cycles of 37 °C for 2 min, 16 °C for 5 min; 60 °C for 10 min; 80 °C for 20 min; 4 °C hold. Following Golden Gate assembly, the gene was amplified by PCR according to General Protocol 1a, and primers are in Table 2-7. The PCR product was digested with BsaI and purified with DNA Clean and Concentrator™-5

according to the manufacturer's protocol with modifications as described in the General Methods (10 μL elution).

Target vector pFLAG-Myc-CMV™-19 (Amp^R) was digested with NotI and EcoRI as described in General Protocol 4, with the modification of a 50 μL phosphatase reaction (1 μL (5 U) of Antarctic phosphatase). The vector and insert were ligated (25 ng digested vector, 25 ng insert, 1x NEB T4 DNA ligase buffer, 0.5 μL (1000 U) NEB T4 DNA ligase, 10 μL final volume in water, 30 min at rt). Bacteria were transformed and clones were isolated as described in the General Protocol 2. Clones were sequenced with primers in Table 2-4 and Table 2-8 as in the General Methods.

Primer ID	Sequence	Annealing Temp (°C)
ppGalNAcT-1seqR1	5'- TCACATAACTCTCTAAAGGCCTTTTCAAA	53
ppGalNAcT-1seqF2	5'- TAGTAGATGATGCCAGTGAAAGAGACT	53
ppGalNAcT-1seqR2	5'- ATGGTAGGTGTCCTGACAGGA	53
ppGalNAcT-1seqF3	5'- AGAAATGGACAGAAGGAAAGGTGATC	53
ppGalNAcT-1seqR3	5'- CCAATGAGAAATAGTGACGTGGAATTTGA	53
ppGalNAcT-1seqF4	5'- TTTTCCTGGTACCTAGAGAATATATATCCTGATT	53

Table 2-8. Sequencing primers for ppGalNAcT1.

2.4.10.2. Site-directed mutagenesis of hT1

Five mutants were obtained from WT truncated hT1 in pFLAG-Myc-CMV™-19 (Amp^R) by site-directed mutagenesis (F346A, F346S, I238A, L295A, and I238A/L310A). PCR amplifications were carried out as in General Protocol 1b, and primers are in Table 2-9. PCR reactions were digested with 1 μL (20 U) DpnI at 37 °C for 1 h, and duplicate reactions were combined (100 μL total) and purified with DNA Clean and Concentrator™-5 according to the manufacturer's protocol with modifications as described in the General Methods (10 μL elution). Bacteria were transformed and clones were isolated as in General Protocol 2. Clones were sequenced with primers in Table 2-4 and Table 2-8 as in the General Methods.

Mutation	Primer ID	Sequence
F346A	ppGalNAcT-1.F346A.F1	5'- CATGTGgccCGGAAAGCTACACC
	ppGalNAcT-1.F346A.R1	5'- TTCCGggcCACATGTCCAACATGTG
F346S	ppGalNAcT-1.F346S.F1	5'- CATGTGTccCGGAAAGCTACACCTTAC
	ppGalNAcT-1.F346S.R1	5'- TTCCGggACACATGTCCAACATGTG
I238A	ppGalNAcT-1.I238A.F1	5'- CCATCgcCGATGTGATCAGTGATGATAC
	ppGalNAcT-1.I238A.R1	5'- ACATCGgcGATGGGACACACC
L295A	ppGalNAcT-1.L295A.F1	5'- GAGGCgccTTTTCAATAGACAGAGATTACTTTC
	ppGalNAcT-1.L295A.R1	5'- GAAAAggcGCCTCCTGCCATGG

Table 2-9. Primers for site-directed mutagenesis of T1 mutants. Lowercase letters indicate the site of base pair mismatch. PCR amplicon is 6.4 kb (1 min 45 sec extension). Annealing temperatures from 52 – 55 **°C** were used.

2.4.10.3. Golden Gate assembly of ppGalNAcT7 and T10

Genes for human T7 and T10 were assembled by Golden Gate cloning. Published nucleotide sequences in the NCBI database AJ002744 (hT7) and AJ_505950 (hT10) were used. Published constructs were referenced for N-terminal truncations of hT7[140] and hT10[115,141,142]. The double mutants generated for hT7 (I330A/L391A) and hT10 (I266A/L321A) are homologous to the hT2 double mutant.

GBlocks of up to 1000 base pairs were available from IDT, so genes were divided into 3 segments. 5 gBlocks were required to build the 4 constructs of T7 (Table 2-10), and the same was true for T10 (Table 2-11). For each family member, 2 gBlocks (GB1) correspond to truncated and full-length versions of the first third of the gene. 2 gBlocks (GB2) correspond to WT or double mutant variants of the middle of the gene. The final gBlock (GB3) corresponds to the C-terminal portion of the gene that is common amongst all four constructs of that family member. Truncated constructs were cloned into the pFLAG-Myc-CMV™-19 expression vector while the full-length constructs were cloned into the pFLAG-CMV™-2 expression vector.

BsaI cleavage sites were found within hT7 and hT10, so silent mutations were introduced to prevent BsaI cleavage of each gene during Golden Gate assembly. Codons for silent mutation were based on the codon frequency in the human genome[135]. In GB2 of hT7, the codon for G403 was mutated from GGT to the slightly more prevalent codon GGA to remove the BsaI restriction site. In hT10, the codon for G514 in GB3 was mutated from GGA to the equally prevalent codon GGG.

gBlock ID	Sequence
T7_GB1_FULL	GACAGGAGGTCTCAGGCCGCGATGAGGCTGAAGATTGGGTTCATCTTA CGCAGTTTGCTGGTGGTGGGAAGCTTCCTGGGGCTAGTGGTCCTCTGG TCTTCCCTGACCCCGCGGCCGGACGACCCAAGCCCGCTGAGCAGGATG AGGGAAGACAGAGATGTCAATGACCCCATGCCCAACCGAGGCGGCAAT GGACTAGCTCCTGGGGAGGACAGATTCAAACCTGTGGTACCATGGCCT CATGTTGAAGGAGTAGAAGTGGACTTAGAGTCTATTAGAAGAATAAA CAAGGCCAAAAATGAACAAGAGCACCATGCTGGAGGAGATTCCCAGA AAGATATCATGCAGAGGCAGTATCTCACATTTAAGCCTCAGACATTCA CCTACCATGATCCTGTGCTTCGCCCAGGGATCCTCGGTAACTTTGAAC CCAAAGAACCTGAGCCTCCTGGAGTGGTTGGTGGCCCTGGAGAGAAAG CCAAGCCATTGGTTTTGGGACCAGAATTCAAACAAGCAATTCAAGCCA GCATTAAAGAGTTTGGATTTAACATGGTGGCAAGTGACATGATCTCA CTGGACCGCAACGTCAATGACTTACGCCAAGAAGAATGCAAGTATTGG CATTATGATGAAAACTTGCTCACTTCGAGCGTTGTCATTGTCTTCCAT ATGAGACCTGGTGTG
T7_GB1_TRUNC	GACAGGAGGTCTCAGGCCGCGCCGCTGAGCAGGATGAGGGAAGACAG AGATGTCAATGACCCCATGCCCAACCGAGGCGGCAATGGACTAGCTCC TGGGGAGGACAGATTCAAACCTGTGGTACCATGGCCTCATGTTGAAGG AGTAGAAGTGGACTTAGAGTCTATTAGAAGAATAAACAAGGCCAAAA ATGAACAAGAGCACCATGCTGGAGGAGATTCCCAGAAAGATATCATG

	CAGAGGCAGTATCTCACATTTAAGCCTCAGACATTCACCTACCATGAT CCTGTGCTTCGCCCAGGGATCCTCGGTAACTTTGAACCCAAAGAACCT GAGCCTCCTGGAGTGGTTGGTGGCCCTGGAGAGAAAGCCAAGCCATTG GTTTTGGGACCAGAATTCAAACAAGCAATTCAAGCCAGCATTAAAGA GTTTGGATTTAACATGGTGGCAAGTGACATGATCTCACTGGACCGCAA CGTCAATGACTTACGCCAAGAAGAATGCAAGTATTGGCATTATGATG AAAACTTGCTCACTTCGAGCGTTGTCATTGTCTTCCATATGAGACCTG GTGTG
T7_GB2_WT	GACAGGAGGTCTCACATAATGAAGGATGGTCAACCCTCATGAGAACA GTCCACAGTGTAATTAAAAGGACTCCAAGGAAATATTTAGCAGAAAT TGTGTTAATTGACGATTTCAGTAATAAAGAACACTTAAAAGAAAAAC TGGATGAATATATTAAGCTGTGGAATGGCCTAGTGAAGGTATTTCGA AATGAAAGAAGGGAAGGTTTAATTCAAGCACGAAGTATTGGTGCTCA GAAGGCTAAACTTGGACAGGTTTTGATATACCTTGATGCCCACTGTGA AGTGGCAGTTAACTGGTATGCACCACTTGTAGCTCCCATATCTAAGGA CAGAACCATTTGCACTGTGCCGCTTataGATGTCATAAATGGCAACACA TATGAAATTATACCCCAAGGGGGTGGTGATGAAGATGGGTATGCCCG AGGAGCATGGGATTGGAGTATGCTCTGGAAACGGGTGCCTCTGACCCC TCAAGAGAAGAGACTGAGAAAAACAAAAACTGAACCGTATCGGTCCC CAGCCATGGCTGGGGGAttaTGTGCCATTGAACGAGAGTTCTTCTTTGA ATTGGG**A**CTCTATGATCCAAGTCTCCAGATTTGGGGTGGTGAAAACTT TGAGATCTCATACAAGATATGGCAGTGTGGTGGCAAATTATTATTTG TTCCTTGTTCTCGTGTGAGACCTGGTGTG
T7_GB2_DOUBLE	GACAGGAGGTCTCACATAATGAAGGATGGTCAACCCTCATGAGAACA GTCCACAGTGTAATTAAAAGGACTCCAAGGAAATATTTAGCAGAAAT TGTGTTAATTGACGATTTCAGTAATAAAGAACACTTAAAAGAAAAAC TGGATGAATATATTAAGCTGTGGAATGGCCTAGTGAAGGTATTTCGA AATGAAAGAAGGGAAGGTTTAATTCAAGCACGAAGTATTGGTGCTCA GAAGGCTAAACTTGGACAGGTTTTGATATACCTTGATGCCCACTGTGA AGTGGCAGTTAACTGGTATGCACCACTTGTAGCTCCCATATCTAAGGA CAGAACCATTTGCACTGTGCCGCTTgccGATGTCATAAATGGCAACAC ATATGAAATTATACCCCAAGGGGGTGGTGATGAAGATGGGTATGCCC GAGGAGCATGGGATTGGAGTATGCTCTGGAAACGGGTGCCTCTGACCC CTCAAGAGAAGAGACTGAGAAAAACAAAAACTGAACCGTATCGGTCC CCAGCCATGGCTGGGGGAgccTGTGCCATTGAACGAGAGTTCTTCTTT GAATTGGG**A**CTCTATGATCCAAGTCTCCAGATTTGGGGTGGTGAAAA CTTTGAGATCTCATACAAGATATGGCAGTGTGGTGGCAAATTATTAT TTGTTCCTTGTTCTCGTGTGAGACCTGGTGTG

gBlock ID	Sequence
T7_GB3	GACAGGAGGTCTCACGTGTTGGACATATCTACCGTCTTGAGGGCTGGC AAGGAAATCCTCCGCCCATTTATGTTGGGTCTTCTCCAACTCTGAAGA ATTATGTTAGAGTTGTGGAGGTTTGGTGGGATGAATATAAAGACTAC TTCTATGCTAGTCGTCCTGAATCGCAGGCATTACCATATGGGGATATA TCGGAGCTGAAAAAATTTCGAGAAGATCACAACTGCCAAAGTTTTAA GTGGTTCATGGAAGAAATAGCTTATGATATCACCTCACACTACCCTTT GCCACCCAAAAATGTTGACTGGGGAGAAATCAGAGGCTTCGAAACTGC TTACTGCATTGATAGCATGGGAAAAACAAATGGAGGCTTTGTTGAAC TAGGACCCTGCCACAGGATGGGAGGGAATCAGCTTTTCAGAATCAATG AAGCAAATCAACTCATGCAGTATGACCAGTGTTTGACAAAGGGAGCT GATGGATCAAAAGTTATGATTACACACTGTAATCTAAATGAATTTAA GGAATGGCAGTACTTCAAGAACCTGCACAGATTTACTCATATTCCTTC AGGAAAGTGTTTAGATCGCTCAGAGGTCCTGCATCAAGTATTCATCTC CAATTGTGACTCCAGTAAAACGACTCAAAAATGGGAAATGAATAACA TCCATAGTGTTTAGGAATTTGAGACCTGGTGTG

Table 2-10. gBlock sequences for hT7. Sequences are listed 5' – 3'. Silent mutations are **bold**.

gBlock ID	Sequence
T10_GB1_FULL	GACAGGAGGTCTCAGGCCGCGATGAGGCGGAAGGAGAAGCGGCTCCTG CAGGCGGTGGCGCTGGTGCTGGCGGCCCTGGTCCTCCTGCCCAACGTG GGGCTTTGGGCGCTGTACCGCGAGCGGCAGCCCGACGGCACCCCTGGG GGATCGGGGGCGGCGGTGGCGCCGGCGGCGGGACAGGGCTCACACAGT CGACAAAAGAAAACGTTTTTCTTGGGAGATGGGCAGAAGCTGAAGGA CTGGCATGACAAGGAGGCCATCCGGAGGGACGCTCAGCGCGTAGGAAA TGGAGAACAAGGAAGACCTTACCCCATGACCGATGCTGAGAGAGTGG ATCAGGCATACCGAGAAAATGGATTTAACATCTACGTCAGTGATAAA ATCTCCTTGAATCGCTCTCTCCCAGATATCCGGCACCCAAACTGCAAC AGCAAGCGCTACCTGGAGACACTTCCCAACACAAGCATCATCATCCCC TTCCACAACGAGGGCTGGTCCTCCCTCCTCCGCACCGTCCACAGTGTGC TCAATCGCTCGCCTCCAGAGCTGGTCGCCGAGATTGTACTGGTCGACG ACTTCAGTGATCGAGAGCACCTGAAGAAGCCTCTTGAAGACTACATGG CCCTTTTGAGACCTGGTGTG
T10_GB1_TRUNC	GACAGGAGGTCTCAGGCCGCGCCTGGGGGATCGGGGGCGGCGGTGGCG CCGGCGGCGGGACAGGGCTCACACAGTCGACAAAAGAAAACGTTTTTC TTGGGAGATGGGCAGAAGCTGAAGGACTGGCATGACAAGGAGGCCAT CCGGAGGGACGCTCAGCGCGTAGGAAATGGAGAACAAGGAAGACCTT ACCCCATGACCGATGCTGAGAGAGTGGATCAGGCATACCGAGAAAAT GGATTTAACATCTACGTCAGTGATAAAATCTCCTTGAATCGCTCTCTC CCAGATATCCGGCACCCAAACTGCAACAGCAAGCGCTACCTGGAGACA CTTCCCAACACAAGCATCATCATCCCCTTCCACAACGAGGGCTGGTCC TCCCTCCTCCGCACCGTCCACAGTGTGCTCAATCGCTCGCCTCCAGAGC TGGTCGCCGAGATTGTACTGGTCGACGACTTCAGTGATCGAGAGCACC TGAAGAAGCCTCTTGAAGACTACATGGCCCTTTTGAGACCTGGTGTG
T10_GB2_WT	GACAGGAGGTCTCACTTTTCCCCAGTGTGAGGATTCTTCGAACCAAGA AACGGGAAGGGCTGATAAGGACCCGAATGCTGGGGGCCTCAGTGGCA ACTGGGGATGTCATCACATTCTTGGATTCACACTGTGAAGCCAATGTC AACTGGCTTCCCCCCTTGCTTGACCGCATTGCTCGGAACCGCAAGACC ATTGTGTGCCCGATGAATGATGTAATTGACCATGACGACTTTCGGTAC

	GAGACACAGGCAGGGGATGCCATGCGGGGAGCCTTTGACTGGGAGAT GTACTACAAGCGGATCCCGATCCCTCCAGAACTGCAGAAAGCTGACCC CAGCGACCCATTTGAGTCTCCCGTGATGGCCGGTGGACTGTTCGCCGT GGATCGGAAGTGGTTCTGGGAACTCGGCGGGTATGACCCAGGCTTGGA GATCTGGGGAGGGGAGCAGTATGAAATCTCCTTCAAGGTGTGGATGT GTGGGGGCCGCATGGAGGACATCCCCTGCTCCAGGGTGGGCCATATCT ACAGGAAGTATGTGCCCTACAAGGTCCCGGCCGGAGTCAGCCTGGCCC GGAACCTTAAGCGGGTGGCCGAAGTGTGGATGGATGAGTACGCAGAG TTGAGACCTGGTGTG
T10_GB2_DOUBLE	GACAGGAGGTCTCACTTTTCCCCAGTGTGAGGATTCTTCGAACCAAGA AACGGGAAGGGCTGATAAGGACCCGAATGCTGGGGGCCTCAGTGGCA ACTGGGGATGTCATCACATTCTTGGATTCACACTGTGAAGCCAATGTC AACTGGCTTCCCCCCTTGCTTGACCGCATTGCTCGGAACCGCAAGACC ATTGTGTGCCCGATGGCCGATGTAATTGACCATGACGACTTTCGGTAC GAGACACAGGCAGGGGATGCCATGCGGGGAGCCTTTGACTGGGAGAT GTACTACAAGCGGATCCCGATCCCTCCAGAACTGCAGAAAGCTGACCC CAGCGACCCATTTGAGTCTCCCGTGATGGCCGGTGGAGCCTTCGCCGT GGATCGGAAGTGGTTCTGGGAACTCGGCGGGTATGACCCAGGCTTGGA GATCTGGGGAGGGGAGCAGTATGAAATCTCCTTCAAGGTGTGGATGT GTGGGGGCCGCATGGAGGACATCCCCTGCTCCAGGGTGGGCCATATCT ACAGGAAGTATGTGCCCTACAAGGTCCCGGCCGGAGTCAGCCTGGCCC GGAACCTTAAGCGGGTGGCCGAAGTGTGGATGGATGAGTACGCAGAG TTGAGACCTGGTGTG
T10_GB3	GACAGGAGGTCTCAGAGTACATTTACCAGCGCCGGCCTGAATACCGCC ACCTCTCCGCTGGGGATGTCGCAGTCCAGAAAAAGCTCCGCAGCTCCC TTAACTGCAAGAGTTTCAAGTGGTTTATGACGAAGATAGCCTGGGAC CTGCCCAAATTCTACCCACCCGTGGAGCCCCCGGCTGCAGCTTGGGGG GAGATCCGAAATGTGGGCACAGGGCTGTGTGCAGACACAAAGCACGG GGCCTTGGGCTCCCCACTAAGGCTAGAGGGCTGCGTCCGAGGCCGTGG GGAGGCTGCCTGGAACAACATGCAGGTATTCACCTTCACCTGGAGAGA GGACATCCGGCCTGG**G**GACCCCCAGCACACCAAGAAGTTCTGCTTTGA TGCCATTTCCCACACCAGCCCTGTCACGCTGTACGACTGCCACAGCAT GAAGGGCAACCAGCTGTGGAAATACCGCAAAGACAAGACCCTGTACCA CCCTGTCAGTGGCAGCTGCATGGACTGCAGTGAAAGTGACCATAGGAT CTTCATGAACACCTGCAACCCATCCTCTCTCACCCAGCAGTGGCTGTT TGAACACACCAACTCAACAGTCTTGGAAAAATTCAATAGGAACTGAG AATTTGAGACCTGGTGTG

Table 2-11. gBlock sequences for hT10. Sequences are listed 5' – 3'. All gBlocks except T10_GB1_FULL were synthesized by IDT. T10_GB1_FULL was synthesized by primer extension as described below. Silent mutations are **bold**.

Target vectors pFLAG-Myc-CMV™-19 (Amp[R]) and pFLAG-CMV™-2 (Amp[R]) were digested with NotI and EcoRI as described in General Protocol 4. Golden gate assembly reactions were carried out as follows: 1 µL (20 fmol) GB1, 1 µL (20 fmol) GB2, 1 µL (20 fmol) GB3, 20 fmol vector DNA, 1x Promega T4 DNA ligase buffer, 0.25 µL (5 U) Promega T4 DNA ligase, and 0.5 µL (5 U) BsaI in a final reaction volume of 10 µL in water. Thermocycle conditions used to enhance reaction yield were as follows: 25 cycles of 37 °C for 2 min, 16 °C for 5 min; 50 °C for 5 min; 80 °C for 10 min; 4 °C hold.

Following the Golden Gate assembly, the reaction was treated with T4 DNA ligase (1x NEB T4 DNA ligase buffer, 0.6 μL (1200 U) NEB T4 DNA ligase, and a 12 μL final volume in water, 25 min at rt). Bacteria were transformed (2.4 μL ligation mixture, 200 μL bacteria spread on plates) and clones were isolated as described in General Protocol 2. Clones were sequenced with primers in Table 2-12 (T7) and Table 2-13 (T10) as in the General Methods. There are BsaI cleavage sites within the vector that should re-anneal and ligate in the Golden Gate reaction, which we sequenced to confirm proper reassembly of the vector (Table 2-14).

Primer ID	Sequence	Annealing Temp (°C)
T7seq_F1	5'- GGCACCAAAATCAACGGGACTTTCCAAAATGTC	58
T7seq_F2	5'- CCTGGAGAGAAAGCCAAGCCATTGGTTTT	58
T7seq_F3	5'- TGCCCGAGGAGCATGGGATTGGAGTAT	58
T7seq_F4	5'- GCCACAGGATGGGAGGGAATCAGCTTTT	58
T7seq_R1	5'- GCTGGTGGGCACTGGAGTGGCAACTT	58
T7seq_R2	5'- CCTGCGATTCAGGACGACTAGCATAGAA	58
T7seq_R3	5'- GTCCAAGTTTAGCCTTCTGAGCACCAATACTT	58
T7seq_R5	5'- GTCCACTTCTACTCCTTCAACATGAGGCCAT	58

Table 2-12. hT7 sequencing primers.

Primer ID	Sequence	Annealing Temp (°C)
T10seq_F1	5'- CACGGGGATTTCCAAGTCTCCACCCCATT	58
T10seq_F2	5'- GAGAGAGTGGATCAGGCATACCGAGAAAAT	58
T10seq_F3	5'- GATCCCGATCCCTCCAGAACTGCAGAA	58
T10seq_F4	5'- GCCTGGAGACCCCCAGCACACCAA	58
T10seq_R1	5'- GCTCGCGGTACAGCGCCCAAA	58
T10seq_R2	5'- CGGGTCCTTATCAGCCCTTCCCGTTT	58
T10seq_R3	5'- CAGTTAAGGGAGCTGCGGAGCTTTTT	58
T10seq_R5	5'- CTGGTGGGCACTGGAGTGGCAACTT	58

Table 2-13. hT10 sequencing primers

Primer ID	Sequence	Annealing Temp (°C)
cmv_BsaI1_F	5'- cacggggatttccaagtctccaccccatt	58
cmv_BsaI1_R	5'- gagagagtggatcaggcataccgagaaaat	58
cmv_BsaI2_F	5'- gatcccgatccctccagaactgcagaa	58
cmv_BsaI2_R	5'- gcctggagacccccagcacaccaa	58

Table 2-14. Vector BsaI site sequencing primers.

The amount of T7_GB3 was insufficient for Golden Gate assembly, so we amplified more of this gBlock by PCR. T7_GB3 was amplified as in General Protocol 1a, and the primers are listed in Table 2-15. This gBlock was used at the same concentration and volume as other gBlocks for Golden Gate assembly.

Primer ID	Sequence	Annealing Temp (°C)
T7_GB3_amplify_F	5'- GACAGGAGGTCTCACGTGTTGGACATATCTA	58

| T7_GB3_amplify_R | 5'- CACACCAGGTCTCAAATTCCTAAACACTATGGAT | 58 |

Table 2-15. Primers used to amplify T7_GB3. Amplicon <1 kb (15 sec extension).

This paragraph describes how T10_GB1_FULL was generated. Because the N-terminal region of full-length T10 is GC-rich, IDT was unable to synthesize T10_GB1_FULL (Table 2-11). We generated the full-length GB1 by multiple PCR extensions of T10_GB1_TRUNC (Table 2-11). PCRs were carried out as in General Protocol 1a, and the primers are listed in Table 2-16. Duplicate reactions were combined (100 µL total) and purified with DNA Clean and Concentrator™-5 according to the manufacturer's protocol with modifications as described in the General Methods (10 µL elution). The purified PCR product was used as the template for the next PCR extension. Once T10_GB1_FULL was synthesized, the sequence was confirmed with the primers in Table 2-17. The gBlock was used at the same concentration and volume as the other gBlocks for Golden Gate assembly.

Primer ID	Sequence	Annealing Temp (°C)
hT10_FULL_GB1_frag1	5'- GACAGGAGGTCTCAGGCCGCgatgaggcggaaggagaagc	51
hT10_FULL_GB1_frag2	5'- GATGAGGCGGAAGGAGAAGCggctcctgcaggcggt	51
hT10_FULL_GB1_frag3	5'- GGCTCCTGCAGGCGGTGGcgctggtgctggcgg	51
hT10_FULL_GB1_frag4	5'- CGCTGGTGCTGGCGGCCCTGgtcctcctgcccaacgtgg	51
hT10_FULL_GB1_frag5	5'- GTCCTCCTGCCCAACGTGGGGCTTTgggcgctgtaccgcga	51
hT10_FULL_GB1_frag6	5'- GGGCGCTGTACCGCGAgcggcagcccgacgg	51
hT10_FULL_GB1_frag7	5'- GCGGCAGCCCGACGGCACCcctgggggatcggggCG	51
hT10_FULL_GB1_REV	5'- CACACCAGGTCTCAAAAGGGCCATGTAGTCTTCAA	51

Table 2-16. T10 GB1 extension primers. Annealing portion of the primer is lowercase and the overhang is uppercase. hT10_FULL_GB1_frag7 was the forward primer for the first extension reaction and hT10_FULL_GB1_REV was the reverse primer for every extension. PCR amplicon is <1 kb (15 sec extension). Although a mistake occurs in hT10_FULL_GB1_frag7 (underlined CG), the correct nucleotides (GC) are in the template hT10 gene, as well as in the full-length hT10 constructs we generated.

Primer ID	Sequence	Annealing Temp (°C)
T10_GB1seq_F	5'- CTCAGCGCGTAGGAAATGGAGAACAA	57
T10_GB1seq_R	5'- CTGGGAGAGAGCGATTCAAGGAGATTTT	57

Table 2-17. T10_GB1_FULL sequencing primers.

Site-directed mutagenesis of hT10 WT constructs in pFLAG-Myc-CMV™-19 (Amp[R]) and pFLAG-CMV™-2 (Amp[R]) was used to correct a single nucleotide error

originating from gBlock T10_GB2_WT. PCR amplifications were carried out as in General Protocol 1b, and primers are in Table 2-18. PCR reactions were digested with 1 µL (20 U) DpnI at 37 °C for 1 h, and duplicate reactions were combined (100 µL total) and purified with DNA Clean and Concentrator™-5 according to the manufacturer's protocol with modifications as described in the General Methods (10 µL elution). Bacteria were transformed and clones were isolated as in General Protocol 2. Clones were sequenced with primers in Table 2-13 as in the General Methods.

Primer ID	Sequence	Annealing Temp (°C)
T10_WT_N>I_F	5'- CCGATGAtTGATGTAATTGACCATGACGACTTTC	55
T10_WT_N>I_R	5'- CATCAaTCATCGGGCACACAATGGTCTTGCGG	55

Table 2-18. T10_GB2_W2 site-directed mutagenesis primers. Lowercase letters indicate the site of base pair mismatch. PCR amplicon is 6.7 kb (1 min 45 sec extension).

2.4.10.4. Relocating the FLAG tag in pFLAG-CMV-2

For full-length hT7 and hT10 in pFLAG-CMV™-2 (Amp[R]), we removed the FLAG tag from the N-terminus by PCR and reinstalled it at the C-terminus by PCR. PCR amplifications to remove the N-terminal FLAG tag were carried out as described in General Protocol 1b. The primers (Table 2-19) were designed to amplify the entire vector in opposite directions, as with site-directed mutagenesis, and included 3' overhangs that removed the FLAG tag , allowing the vector to self anneal with nicks like those introduced during site-directed mutagenesis[143]. PCR reactions were digested with 1 µL (20 U) DpnI at 37 °C for 1 h, and duplicate reactions were combined (100 µL total) and purified with DNA Clean and Concentrator™-5 according to the manufacturer's protocol with modifications as described in the General Methods (10 µL elution). Bacteria were transformed and clones were isolated as in General Protocol 2. Clones were sequenced with primers in Table 2-4 as in the General Methods. Once the deletion of the FLAG tag was confirmed, the C-terminal FLAG tag was installed after the hT2 coding sequence directly prior to the stop codon. The C-terminal Flag tag was installed as described above for the removal of the FLAG tag.

Primer ID	Sequence	Annealing Temp (°C)
hT7_10N_FLAGdel_F	5'- gaattgatctaccaagcttgcggccgcg^ATGAG	55
hT2N_FLAGdel_R	5'- ggccgcaagcttggtagatcaattctgacggttcactaaacg	55
hT7C_FLAGIns_F	5'- gactacaaagacgatgacgacaagtaggaattcatcgatagatctgatatcggtacc	55
hT7C_FLAGIns_R	5'- cttgtcgtcatcgtctttgtagtcAACACTATGGATGTTATTCATTTCCCATTTTTGAG	55

51

hT10C_FLA GIns_F	5'- gactacaaagacgatgacgacaagtgagaattcatcgatagatctgatatcggtacc	55
hT10C_FLA GIns_R	5'- cttgtcgtcatcgtctttgtagtcGTTCCTATTGAATTTTTCCAAGACTGT TGAGTTGG	55

Table 2-19. Primers to move the FLAG tag in pFLAG-CMV™-2. hT7_10N_FLAGdel_F and hT2N_FLAGdel_R were used to remove the FLAG tag from hT7 and hT10 constructs. To install the C-terminal FLAG tag, hT7C_FLAGIns_F and R were used for hT7 and hT10C_FLAGIns_F and R were used for hT10. Plasmid sequences are lowercase, FLAG tag is underlined, and hT7 and hT10 sequences are uppercase. The site of the FLAG deletion is marked with ^. PCR amplicon for hT10 is 6.5 kb and for hT7 is 6.7 kb (1 min 45 sec extension).

2.4.11. Mammalian cell culture and protein expression

Human Embryonic Kidney (HEK) were cultured as described in the General Methods. Cells were plated at density of approximately 1-3 x 10^5 cells/mL in a 10-cm plate 24 h prior to transfection. To obtain sufficient amounts of plasmid DNA for mammalian cell transfection, a Midiprep kit (Qiagen) was used according to the manufacturer's protocol (DNA pellets were solubilized in water). Cells were transiently transfected with human ppGalNAcT constructs described above using *Trans*IT®-293 reagent (10-cm plate: 15 μg DNA, 45 μL transfection reagent, 1.5 mL serum free media) according to the manufacturer's protocol. The transfection mixture was added dropwise to the cells and plates were mixed gently. The media was changed 24 h post-transfection.

At 48 h, the conditioned media (10 mL) was collected and the cells were harvested by scraping in 3 mL cold PBS. The conditioned media and cells were pelleted by centrifugation for 15 min at 3650 x g at 4 °C. The clarified conditioned media was removed from the pelleted cell debris, transferred to a clean tube, and EDTA-free protease inhibitors were added. The supernatant from the harvested cells was discarded.

2.4.12. Anti-FLAG purification

The secreted protein in the clarified conditioned media (10 mL) was purified at rt using columns packed with 1.25 mL α-FLAG M2 agarose resin from Sigma according to manufacturer's protocol. Prior to purification, *10 μL of clarified conditioned medium was removed* for SDS-PAGE/WB analysis and stored at -20°C. Conditioned medium was applied to column under gravity flow, and the eluent was applied 2 x (3 passes total) to ensure maximal binding. *10 μL of the final flow-through was removed* for SDS-PAGE/WB analysis and stored at -20°C. the column was washed TBS (25 mM Tris pH 7.4, 150 mM NaCl, 2 x 10 mL) to remove any unbound protein. Protein was in 5 mL (fractions E1-4) using 100 μg/mL 3x FLAG peptide in TBS buffer(with EDTA-free protease inhibitors). A fifth elution fraction (E5, 1.25 mL) was used to asses elution efficiency. *5 μL of E1-4 and 1.25 μL of E5 was removed* for SDS-PAGE/WB analysis and stored at -20°C. 50 μL of E1-4 was removed for colloidal blue

analysis and stored at -20°C. Glycerol was added to elution fractions 1-4 to a final concentration of 25% glycerol (80% glycerol stock) and proteins were stored at –80 °C.

For column reuse, the column was washed with three bed volumes of 0.1 M glycine HCl, pH 3.5, and then immediately washed with 10 mL of TBS to neutral pH. The column was treated with 5 mL of 50% glycerol (TBS buffer, 0.02 % sodium azide), and after 2-3 mL had passed through the column, it was stored at 2-8 °C. The column was washed with 5-10 mL TBS directly prior to column reuse. After column reuse, protein yields decreased slightly and some anti-FLAG antibody was shed from the resin. Therefore, column reuse was limited to optimization experiments when possible.

2.4.13. SDS-PAGE and western blotting of ppGalNAcTs

The expression and purification of ppGalNAcT2 and the mutants were monitored by SDS-PAGE and western blot analysis. Whole cell lysates were prepared by resuspending the cell pellet in 2 mL water, removing a 75 μL aliquot, adding 25 μL protein loading buffer (100 μL total), and boiling at 90 °C for 10 min. 15 μL of this whole cell lysate sample was loaded onto the gel.

Aliquots removed during anti-FLAG purification correspond to 0.1% of the total sample volume at respective step. TBS was added to sample purification aliquots to a final volume of 10 μL. SDS protein loading buffer (3.75 μL) and water (1.25 μL) were added. SDS-PAGE was carried out as in General Protocol 5 followed by western blotting as in General Protocol 6.

2.4.14. Colloidal blue analysis

50 μL of E1-4 from the anti-FLAG purification of each ppGalNAcT construct was quantitated by SDS-PAGE followed by colloidal blue staining. SDS protein loading buffer (16.67 μL) was added to E1-4. Recombinant human albumin expressed in *S. cerevisiae* was used as a protein standard. Albumin stocks (250 ng/μL and 62.5 ng/μL in TBS) were diluted in TBS and SDS protein loading buffer to generate solutions at 16.67, 8.33, and 4.17 ng/ μL. 30 μL of all samples and standards (500, 250, 125 ng albumin) were loaded in duplicate onto an 18-well 12 % Bis-Tris gel, and SDS-PAGE was carried out as in General Protocol 5. Colloidal blue staining (Colloidal Blue Staining Kit, Invitrogen) was carried out according to the manufacturer's instructions.

Chapter 3. Chemoenzymatic synthesis of UDP-GalNAc analogs

The previous chapter discusses ppGalNAcT active site engineering, and this chapter will address the development of UDP-GalNAc analogs. There is substantial evidence that GalNAc modified with a chemical handle allows labeling of the glycosylation products of the ppGalNAcTs, which is the ultimate goal of this research. As described in Chapter 2, we chose to generate the bumped UDP-GalNAc analog via modification of the GalNAc moiety, rather than modification of the uridine base. This allows the bump and chemical handle to be a single modification that is installed in one synthetic step rather than two distinct syntheses. Due to the extensive precedent with C8-modified sugars being processed by cells, as well as the synthetic ease of modifying the N-acyl position, we chose to engineer the C8 position of GalNAc.

The chemical handle appended to UDP-GalNAc analogs is designed to undergo a selective ligation reaction in a biological context. This enables specific labeling of biomolecules modified with the analog and downstream identification and analysis of those biomolecules. Selective labeling of the chemical handle in a biological setting is made possible by bioorthogonal chemistry.

3.1. Bioorthogonal chemistry

The field of bioorthogonal chemistry has grown substantially since its inception in the late 1990s. Bioorthogonality is a characteristic that describes chemical reactions that can proceed selectively within a biological context. Importantly, bioorthogonal reactions are orthogonal to biological functional groups, small molecules, and metabolites, meaning that these reactions occur with minimal interference from or influence on native biology[90]. Bioorthogonal ligation reactions involve two reactive partners that are largely absent from nature that can react rapidly at physiological temperature and pH in water to form a stable product[90].

When bioorthogonal reactions are used to monitor physiological processes, one reaction partner, known as the chemical handle, must be small and easily incorporated into the biomolecule of interest. This bioorthogonal handle should be small enough that it does not perturb the normal biological function of the tagged biomolecule. The status of the biomolecule is monitored as it is metabolized or trafficked within a living system, using a probe equipped with the reactive partner complementary to the bioorthogonal handle. Most commonly, the probe is a fluorophore for visualization or an affinity handle for enrichment of the biomolecule of interest.

The azide is currently the most widely used and versatile bioorthogonal handle. The first bioorthogonal reaction for the azide was developed by Bertozzi and colleagues, and is based on a modification of the Staudinger reduction[125]. An azide can react with a triarylphosphine to form an amide bond in the Staudinger ligation[90,125]. However, the Staudinger ligation is not ideal for many applications. The reaction

proceeds slowly and the triarylphosphine is easily oxidized, which can lead to lower signal than is required for many applications[90].

The next major advance, independently reported by Sharpless[145] and Meldal[146], was that the [3+2] cycloaddition between a linear alkyne and an azide can be accelerated with Cu(I). This rapid and selective reaction is now termed copper click. Despite the advantages of copper click, the utility of this reaction is limited because Cu(I) is toxic to cells and animals[147,148]. The development of copper coordinating ligands has reduced the toxicity of Cu(I), accelerated the reaction, and improved the sensitivity in complex samples[148-150].Optimization of the reaction has made it possible to carry out copper click labeling on the surfaces of live cells and zebrafish, but toxicity still precludes its use within living animals[150,151]. Copper click has been particularly well-suited as a tool for lysate labeling experiments with downstream proteomic applications[130,152].

The next bioorthogonal reaction for the azide, copper-free click chemistry, was developed by Bertozzi and coworkers. The reaction partner for copper-free click chemistry is a cyclooctyne, which was chosen because its strain energy makes it significantly more reactive than a linear alkyne. Copper-free click reactions are nontoxic and are faster than the Staudinger ligation by up to two orders of magnitude, enabling robust labeling in a variety of contexts including cells and living animals.

The 4th bioorthogonal ligation that is widely used is the inverse-electron-demand Diels-Alder reaction between tetrazines and strained alkenes such as *trans*-cyclooctene[153] and cyclopropene[154]. Although developed more recently than other bioorthogonal reactions, the tetrazine ligation is very fast, and this ligation has been rapidly adopted for biological labeling experiments[155,156]. The tetrazine ligation has been used successfully in many contexts including cell[157,158] and animal studies, such as labeling of mouse tumor models[159]. However, many of the fastest tetrazines are unstable in aqueous media, and other suffer from poor solubility in water, which limits the available tetrazine reactive partners [160].

As described in the next section, the azide was selected as the chemical handle modifying all UDP-GalNAc analogs in this dissertation. Therefore, the bioorthogonal reactions utilized in this dissertation include the Staudinger ligation and copper and copper-free click chemistry.

3.2. Design of UDP-GalNAc analogs

3.2.1. Choice of chemical handle

Choosing the proper chemical handle is an important aspect of UDP-GalNAc analog design. The earliest chemical handle used to label a glycan was a ketone, which was incorporated into an N-acetylmannosamine (ManNAc) analog[161]. Ketones and aldehydes can participate in selective reactions with hydrazides and aminooxy

reagents to label extracellular biomolecules. Unfortunately, the hydrazone and oxime linkages that form are somewhat labile in aqueous conditions, which has led to the development of new carbonyl reactive partners that form hydrolytically-stable adducts[162]. However, many intracellular biomolecules, including pyruvate and glucose, contain aldehydes or ketones. As a result, aldehydes and ketones are not fully bioorthogonal handles, nor are their reaction partners.

Due to their small size and absence from biological systems, the alkyne and azide are excellent bioorthogonal handles. Placement of an alkyne on the biomolecule and an azide on the probe is preferred, as live cell feeding followed by Cu-click lysate labeling suggests that this configuration results in better signal to noise ratios than the reverse[130,148,152,163]. However, a biomolecule modified with an alkyne is limited by having only one type of chemistry available: copper-click. Due to the liabilities of copper- click described earlier, alkyne handles are not ideal for GalNAc analogs that are to be utilized in cells and animals. More bioorthogonal reactions are available to azide handles, thereby expanding the scope of possible biological experiments that can be done on different timescales and in different contexts.

3.2.2. Design of bump

When designing possible bumps for the C8 position, we considered three types of modifications that might result in orthogonality: long chains, bulky moieties such as aromatics, and branched alkyl structures.

Long chains explore a relatively limited chemical space and generally lack significant stereochemical features. To develop an orthogonal modification, significant steric clashes with the native enzyme active site need to be introduced, which are more likely with bulky aromatics and branched alkyl structures. The more complex stereochemistry of these structures allows for the exploration of more chemical space. Compared to a flexible, extended chain, the structural rigidity of these other modifications might allow for tighter "lock and key" binding and better activity.

There is precedent for GalNAc and GlcNAc analogs with extended linkers that have promiscuous and/or orthogonal behavior. Bateman *et al.* found that *N*-propargyloxycarbamate monosaccharides GlcPoc and GalPoc, shown in Figure 3-1, are accepted by the GlcNAc and GalNAc salvage pathways[164]. Both analogs label O-GlcNAcylated proteins, which suggests that UDP-GalPoc can be epimerized to UDP-GlcPoc[164]. These data indicate that UDP-GlcPoc is utilized by OGT and these modifications are not orthogonal the enzymes in salvage pathways for these enzymes. In another example, GalNAz is accepted by the GalNAc salvage pathways, is interconverted by the C4-epimerase GALE, and is accepted by glycan biosynthetic pathways(Figure 3-1)[17,46]. However, GlcNAz experiences a bottleneck in the GlcNAc salvage pathway that results in minimal formation of UDP-GlcNAz[131]. Similarly, a GalNAc analog with a carbon isostere of the N-acetyl group, a C2-keto-GalNAc analog, is metabolized in cells and presented on cell surfaces(Figure 3-1)[165]. This

analog, C2-keto-GalNAc, likely occupies the core position of O-GalNAc glycans[165]. In contrast, labeling with the C2-keto-GlcNAc analogs does not result in signal above background, indicating that C2-keto-GlcNAc is orthogonal to glycan biosynthetic pathways.

Many analogs with extended moieties at the C8 position are good substrates for some glycan biosynthetic enzymes. However, many appear to be poor substrates for the C4-epimerase the converts UDP-GlcNAc and UDP-GalNAc, suggesting partial orthogonality introduced by these C8 modifications. Kohler and coworkers designed a GlcNAc analog with a diazirine connected via an extended alkyl linker at the C8 position(Figure 3-1)[114]. As intended, GlcNDAz is not orthogonal to O-GlcNAc transferase (OGT). However, it is orthogonal to enzymes in the GlcNAc salvage pathway, and GlcNDAz is not epimerized to form GalNDAz in cells. Zaro *et al.* reported GalNAlk, an alkyne modified analog that appears to label N-linked glycans and potentially mucin-type glycans in cells[130]. However, this analog does not appear to be converted to the GlcNAlk analog, suggesting that this analog is orthogonal to the C4 epimerase GALE(Figure 3-1)[130].
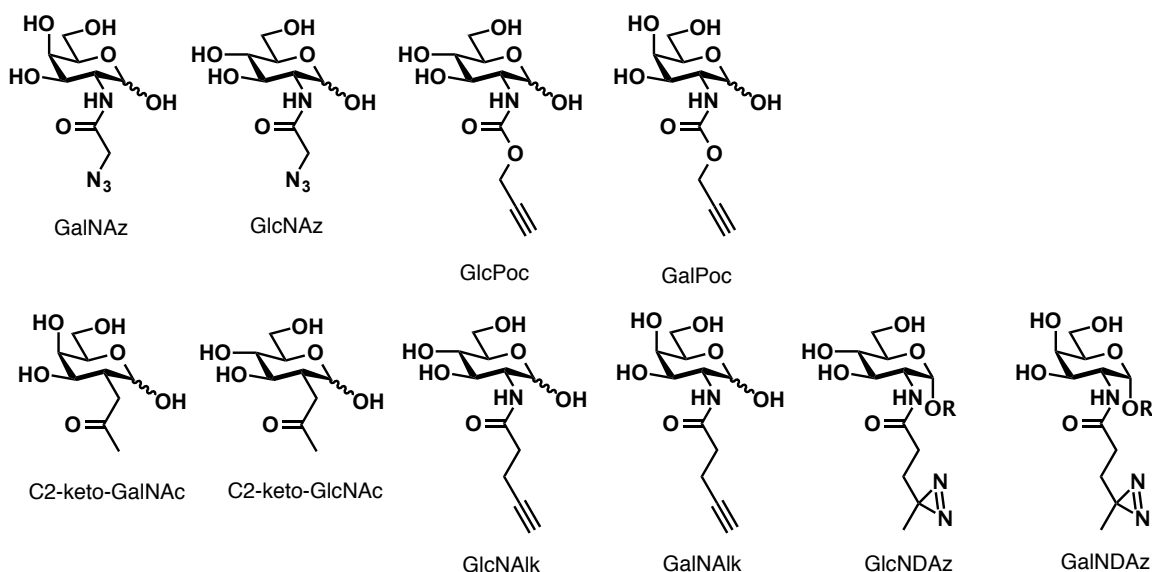
Figure 3-1. GlcNAc and GalNAc analogs with bioorthogonal handles. (Top row) Chemical handles that have been shown to allow interconversion of GlcNAc and GalNAc analogs. (Bottom row) Chemical handles that have not shown ready conversion.

Other enzyme classes have also been studied with substrate analogs modified with long chains. Luo and coworkers developed several SAM analogs with extended, partially unsaturated chains capped with alkynes or azides[108]. Many of these analogs are non-orthogonal and are utilized by the native methyltransferases, and only a few structures have been appropriate for bump-hole engineering[108,109]. Other analogs with saturated chains demonstrated no activity with wildtype or mutant enzymes, presumably due to the flexibility of the chemical handle, which results in low binding affinity[166]. Due to the instances of non-orthogonality achieved with these

modifications, we did not consider long chains to be as desirable as other modifications when engineering a bump hole.

Bulkier substituents such as aromatics are successful with ATP analogs designed for kinase bump-hole engineering. Shokat and colleagues found that benzyl- and phenylethyl-modified adenosine fits well into pockets generated via mutation of isoleucine and other bulky residues to alanine or glycine[167]. Additionally, because we mutated the Phe/Tyr gatekeeper residues in some of our ppGalNAcT mutants, we were interested in evaluating UDP-GalNAc analogs with aromatic bumps. As described in Chapter 4, the Phe mutants did not demonstrate enzymatic activity thus far, which may be due to the loss of acceptor substrate binding. As described in Chapter 2, Phe has been shown to coordinate with the acceptor peptide in the crystal structure of T2.

Concurrently, our initial experiments with the branched methyl analog UDP-GalNAzMe(S) demonstrated promising activity with the Ile/Leu double mutant. As described in Chapter 4, the double mutant was promising because UDP-GalNAz partially rescued loss of activity compared to UDP-GalNAc. Once our efforts focused on the double mutant, we followed up with other alkyl substituents, rather than evaluating aromatic bumps.

3.3. Chemical synthesis of UDP-GalNAz

Work by Bertozzi and colleagues has shown that the azide is an ideal chemical handle and that sugars functionalized with an azide on the N-acyl side chain can be processed by the cell[85,125,126]. Studies *in vitro* using the UDP-GalNAc analog uridine 5'-diphospho-2-azidoacetoamido-2-deoxy-α-D-galactopyranose (UDP-GalNAz) have shown that it can act as a donor substrate for the ppGalNAcTs at about one third the efficiency of UDP-GalNAc[6,80]. This indicates that an azide installed alpha to the carbonyl is a "bump" that is too small to make UDP-GalNAz fully orthogonal to the cell's glycan processing machinery. Therefore, UDP-GalNAz was synthesized as control substrate **2.1**, and analogs **3.13a-g** with substituents on the N-acyl side chain are designed for loss of reactivity with native ppGalNAcTs.

UDP-GalNAz was synthesized as described previously (Figure 3-2)[80,168,169]. Briefly, EDC coupling of azidoacetic acid with galactosamine (**3.1**) affords sugar analog **3.2**[6]. Peracetylation of **3.2** followed by selective deprotection of the anomeric position yields **3.4**[6,169]. Protected phosphate **3.5** is synthesized via the phosphite in a phosphoramidite coupling followed by oxidation[6,170]. The phosphodiester coupling reaction and generation of the free nucleotide sugar is carried out in 3 steps. Treatment of **3.5** with sodium p-toluenesulfinate and Pd(PPh$_3$)$_4$ affords the free phosphate, which upon reaction with uridine monophosphate-N-methylimidazolide (**3.6**) yields the protected nucleotide sugar[171]. UDP-GalNAc analog **2.1** is recovered upon deprotection of the pyranose and ribose hydroxyl groups[6].
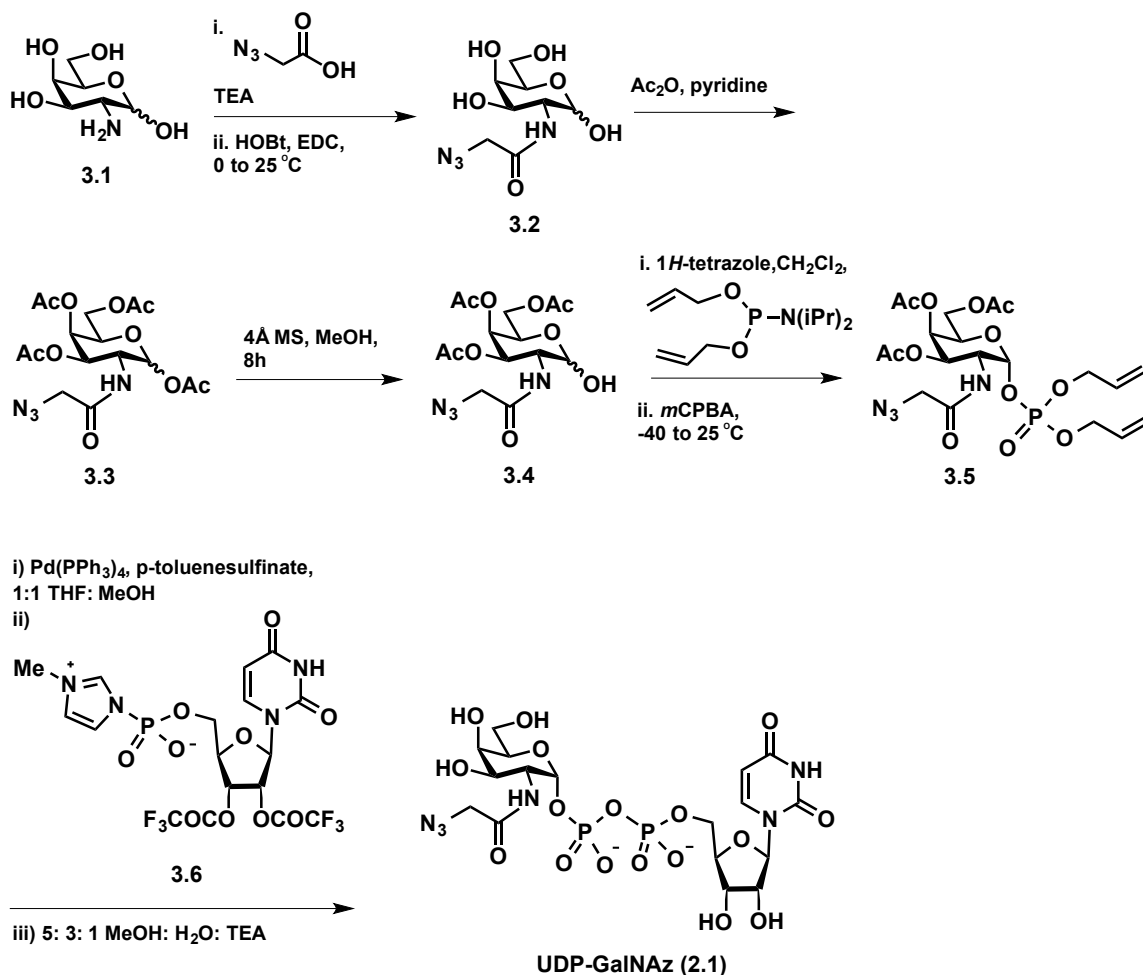
Figure 3-2. Chemical synthesis of UDP-GalNAz.

3.3.1. Challenges with chemical synthesis of nucleotide sugars

Although this synthetic route is well-known, the challenges associated with chemically synthesizing nucleotide sugars are also well-understood[172]. The generation of the phosphodiester bond is synthetically challenging and low-yielding, making these reactions difficult to perform on greater than low milligram scale. Since we began this work, synthetically straightforward routes that can be performed in water have been reported[173,174]. In 2013, good yields on the 100 mg scale were reported by Dabrowski-Tumanski *et al.*, indicating that these new routes may be valuable in the future[174]. However, we focused our efforts on enzymatic synthesis as an alternative to chemical synthesis.

The route we selected allowed us to synthesize a single nucleotide sugar, UDP-galactosamine, and functionalize it, rather than generating the phosphodiester bond for every analog. This design is synthetically streamlined relative to other methods. By generating UDP-galactosamine (UDP-GalN) enzymatically, we avoided the challenge of protecting the amine and deprotecting it under sufficiently mild conditions.

3.4. Enzymatic synthesis of the diversifiable intermediate UDP-galactosamine

There has been significant interest in the use of glycosyltransferases to synthesize carbohydrate structures, both natural and non-natural, due to their stereospecificity, high yields, lack of side products, and in the case of some enzymes, ability to tolerate alternative substrate structures. Particularly relevant to nucleotide sugars, enzymes offer the ability to produce synthetically challenging structures in high yields in one pot. Although recent work to generate nucleotide sugar analogs has focused on using enzymes in the GalNAc salvage pathway to transform non-natural sugars to their UDP-sugars, these reactions appear to have a limited substrate scope[175]. In particular, galactokinase has been shown to have limited promiscuity, although this varies by species. Due to our interest in a structurally diverse panel, we sought to minimize the limitations imposed by enzyme specificity. We selected a chemoenzymatic route where the enzymatic synthesis preceded the chemical modification.

To generate a panel of analogs efficiently, we focused our efforts on the generation of a late-stage diversifiable intermediate. Work by Whitesides and colleagues demonstrated that the biosynthetic pathway to generate UDP-Galactose can be hijacked to produce UDP-GalN[176]. UDP-GalN is the ideal intermediate for a UDP-GalNAc analog because the amine at the C-2 position can undergo an amide coupling under mild conditions to generate a "bumped" N-acyl side chain. Hence, UDP-GalN can be transformed into a UDP-GalNAc analog in a single step [176]. The Leloir enzymes are not as active with galactosamine as with their native substrate Gal. Coupling the transformations by Galactokinase (GalK) and Galactose-1-phosphate uridyltransferase (GalPUT) with the enzymatic regeneration of other starting materials drives the reaction forward (Figure 3-3)[176,177].
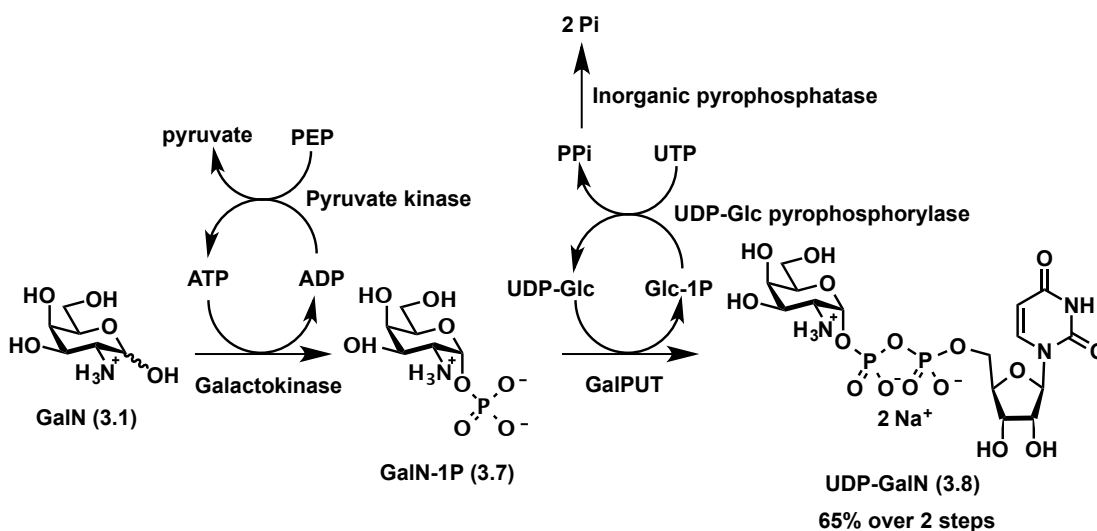
Figure 3-3. Enzymatic synthesis of UDP-GalN. A coupled enzyme system drives the conversion of GalN to UDP-GalN through the intermediate GalN-1P. The species of origin for

the biosynthetic enzymes are GalK (*E. coli*), PYK2 (*S. cerevisiae*), GalPUT (*E. coli*), UDP-GP (*Trypanosoma brucei*), PPiase (*Pasteurella multocida*). Starting materials include GalN, ATP, PEP, UDP-Glc, and UTP. Substoichiometric concentrations of ATP and UDP-Glc are used to catalyze the reaction.

3.4.1. Biosynthetic enzymes

We selected GalK and GalPUT from *E. coli* due to their ability to utilize galactosamine and galactosamine-1-phosphate, respectively, as substrates[177]. In nature, GalK uses ATP as a phosphate donor to phosphorylate galactose, generating galactose-1-phosphate (Gal-1P). GalPUT then uses UDP-glucose (UDP-Glc) as a uridine monophosphate (UMP) donor to convert Gal-1P to UDP-galactose. Based on work by Whitesides and colleagues and optimization by Errey *et al.* to generate UDP-galactosamine enzymatically, additional enzymes were used to deplete the side products of the reaction and regenerate starting materials. As shown in Figure 3-3, the first reaction was driven forward with pyruvate kinase (PYK2), which regenerated ATP from adenosine diphosphate (ADP) and phosphoenolpyruvate (PEP). UDP-glucose pyrophosphorylase (UDP-GP) was used to drive the second reaction forward by regenerating UDP-Glc from glucose-1-phosphate and uridine triphosphate. Inorganic pyrophosphatase was used to break-down pyrophosphate to inorganic phosphate, further depleting the accumulation of reaction side products and driving the reaction towards completion.

Wang and coworkers demonstrated that by immobilizing multiple enzymes onto beads, the enzymes can be used in excess while streamlining the purification of the nucleotide sugar[178]. To facilitate this strategy, we generated His-tagged versions of the enzyme constructs as necessary; produced the protein in *E. coli* expression strain BL21 (DE3); and immobilized, washed, and reacted the enzymes on Ni-NTA beads. The one pot, on-bead, coupled enzyme reaction ultimately enabled the synthesis of UDP-GalN on almost gram scale in excellent yields.

3.4.2. Optimization of enzymatic synthesis

We have developed a protocol to synthesize UDP-GalN enzymatically in tens to hundreds of milligram quantities in approximately 50 – 70 % yields. In the optimization process, the ratios of starting material and biosynthetic enzymes were investigated. The synthesis by Heidlas *et al.* emphasized the coupled enzyme system to drive the reaction forward, while using smaller quantities of the major biosynthetic enzymes GalK and GalPUT[176]. This made it possible to use catalytic amounts of ATP and UDP-Glc, while regenerating them from the less expensive reagents PEP and UTP. The yield of this enzymatic synthesis is heavily reliant on the efficient function of the supporting enzymes to remove side products and to regenerate starting materials. In contrast, Errey *et al.* used the enzymes that catalyze the main synthetic pathway in excess, compared to relatively small amounts of the coupled support system[177]. This was particularly relevant when using galactose analogs with which GalK and GalPUT had lower activity. However,

this design potentially experiences more acute buildup of side products, which can inhibit reaction progress.

The effective use of such a range of different reaction conditions indicates that an intermediate set of conditions is also feasible. For a large scale synthesis, we focused on conditions that are not heavily reliant on one aspect of the coupled reaction. Such conditions should reduce the variation between reactions and enable the maximal use of the biosynthetic enzymes purified from bacteria. Because GalK accepts Gal analogs less efficiently than GalPUT, the phosphorylation of GalN by GalK is likely rate limiting in the synthesis of UDP-GalN[177]. To overcome this bottleneck and ensure the initiation of the synthetic pathway, GalK was used in excess.

The relative amounts of the biosynthetic enzymes, as well as the ease of monitoring reaction progress, impacted our optimization of ratios of starting materials. We initially utilized a large excess of UDP-Glc to drive the uridine transfer forward. However, UDP-Glc and UDP-GalN share roughly the same R_f value in our TLC solvent system, and the presence of excess UDP-Glc obscures the UDP-GalN spot. We found that catalytic amounts of UDP-Glc made it possible to monitor the reaction progress by TLC. After optimizing ratios of enzymes and starting materials, we optimized reaction variables including concentration, amount of enzyme, and temperature. Temperature significantly impacts reaction progress and overall yield, and we hypothesize that a balance between enzyme catalysis and enzyme stability was achieved at 30 °C.

UDP-GalN was initially purified by HPLC with a tributylammonium bicarbonate buffer, and the separation of the various reaction components was excellent. The purified product was further purified by an ammonium salt exchange followed by a sodium salt exchange. The consecutive salt exchanges were necessary because the tributylammonium and ammonium salts appear to inhibit the modification of UDP-GalN at the diversification step. We hypothesize that this is due to altered reaction pH in the presence of these salts. Additionally, ammonium bicarbonate can react with the NHS-ester and quench the amide bond forming reactions[179]. Unfortunately, we found that the salt exchanges did not remove all of the undesirable salts, and further purification was required. Additionally, HPLC was limited by low throughput for large scale purifications. We found that size exclusion chromatography in water was an alternative purification where large quantities of UDP-GalN could be purified without additional processing.

3.5. Synthesis of UDP-GalNAc analogs

UDP-GalN was treated with a panel of azide-modified bumps to generate a panel of UDP-GalNAc analogs. To ensure the stability of UDP-GalN to the coupling conditions, amide bond formation was achieved via mild *N*-hydroxysuccinimide (NHS)-ester coupling in a miscible aqueous/ organic solvent mixture. The azido-bumps were based on an amino acid scaffold, enabling a huge variety of possible bumps in two stereochemical configurations corresponding to the D- and L-amino acids. The

dimensions of the UDP-GalNAc binding pocket in the enzyme active site cannot be measured precisely due to limited crystallographic data, so these analogs were chosen to span a range of sizes.

3.5.1. Diazotransfer to modify amino acids

As published in the synthesis of GalNAz (**3.2**) and UDP-GalNAz (**2.1**), installation of an azide onto a sugar analog has previously been accomplished by treating a halogenated acid such as iodoacetic acid with sodium azide, resulting in the S_N2 displacement of the primary halogen by the nucleophilic azide ion[2,26,27]. This creates an azide-modified acid that can from an amide bond with a hexosamine sugar under peptide-coupling conditions. This reaction is not ideal for our purposes, as the S_N2 reaction with azide anion is often quite slow, occurring over days, and undesired stereochemistry and elimination products can form with certain structures[180].

The installation of stereocenters is challenging, and simple, commercially available starting materials already containing stereocenters are limited. In selection of these starting materials, we were additionally constrained by the need for functionalities that could easily be converted to a carboxylic acid and an azide. Natural amino acids are stereochemically defined and contain stereocenters in both R and S configurations corresponding to the D- and L-amino acids. Additionally, a wide variety of non-natural amino acids are commercially available. Finally, amino acids contain carboxylic acid and amine functionality, and the amine can be converted to an azide in a single step.

An alternative to the installation of azides with azide anion is the synthesis of azides by diazotransfer to a primary amine. This reaction proceeds in good yields over shorter reaction times and preserves the existing stereochemistry[180]. Diazotransfer reactions have traditionally used trifluoromethanesulfonyl azide (TfN$_3$), which must be prepared in solution prior to its use, as the pure reagent is explosive. This led to the development diazotransfer reagents that are easier to prepare and handle than TfN$_3$, but react with a range of substrates comparably to TfN$_3$ [180,181]. We chose to use the tetrafluoroborate salt of imidazole-1-sulfonyl azide **3.10**, which is the most shelf-stable salt[181]. The diazotransfer reagent **3.10** was synthesized according to literature precedent[180,181]. With **3.10** in hand, we modified a panel of branched and unbranched natural and non-natural alkyl amino acids (**3.9**), and the diazotransfer was carried out according to the methods described in Goddard Borger *et al*[180].

3.5.2. Coupling reaction and current panel

To generate an amide bond between the azido acids **3.11** and UDP-GalN (**3.8**), we needed a reaction that could take place in an aqueous or mixed aqueous solvent with minimal side products or off-target reactivity. These criteria were important because UDP-GalN is charged and is poorly soluble in organic solvents. Additionally, we wanted to minimize loss of the precious nucleotide sugar to off-target reactivity or complicated purification to remove side products.

An amide bond can be generated with a shelf-stable active ester or with carboxylic acids otherwise activated for nucleophilic attack. Although isolable active esters may be less reactive than carboxylic acids activated *in situ* or by other methods, they allow a very simple reaction mixture during the amide bond formation, can be tailored for aqueous solubility, and require minimal reaction clean up[182]. Based on these criteria, we selected NHS-esters as well-characterized active esters that are stable enough to be purified but generally reactive enough to have good yields coupling in water at rates competitive with hydrolysis[182].
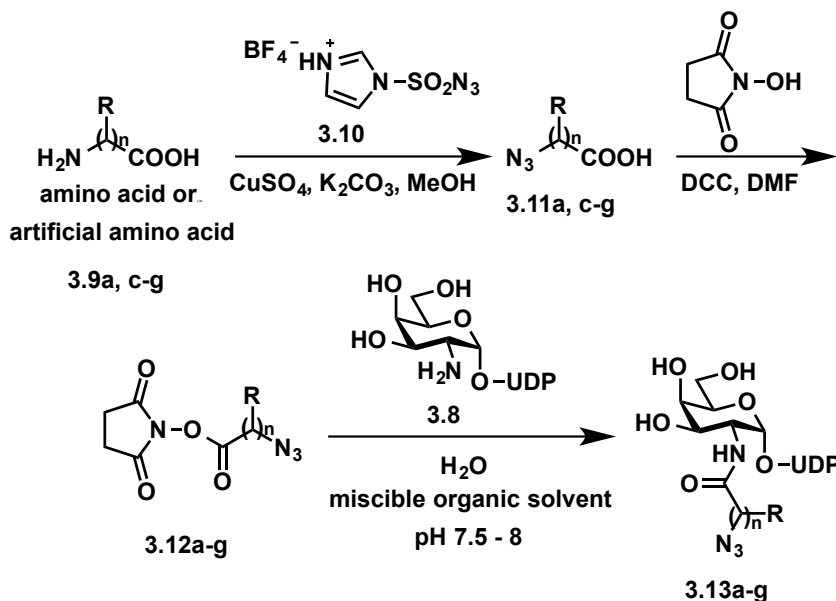
Figure 3-4. Synthesis of azido bumps and coupling to form UDP-GalNAc analogs.

With the azido acids **3.11** in hand, we generated the corresponding NHS-esters **3.12a-g** using dicyclohexylcarbodiimide (DCC). Compounds **3.12a-g** were purified by flash chromatography on C2-silica to minimize hydrolysis during purification. Couplings between the NHS-esters and UDP-galactosamine (**3.8**) were carried out in water with mixed miscible organic solvents to solubilize the NHS ester. The reactions were performed at a pH of 7.5-8, to facilitate amide bond formation by nucleophilic attack of the amine over ester hydrolysis[179]. Incremental addition of NHS-ester over a longer period of time increases the overall conversion. This is not unexpected as the ester undergoes hydrolysis over time, and extending the exposure time of **3.8** to the intact ester will improve product formation. The major byproduct from these reactions was free NHS, and unreacted UDP-GalN could be recovered during purification for reuse. Size exclusion chromatography in water was used to purify the UDP-GalNAc analogs **3.13a-g** (Figure 3-5).
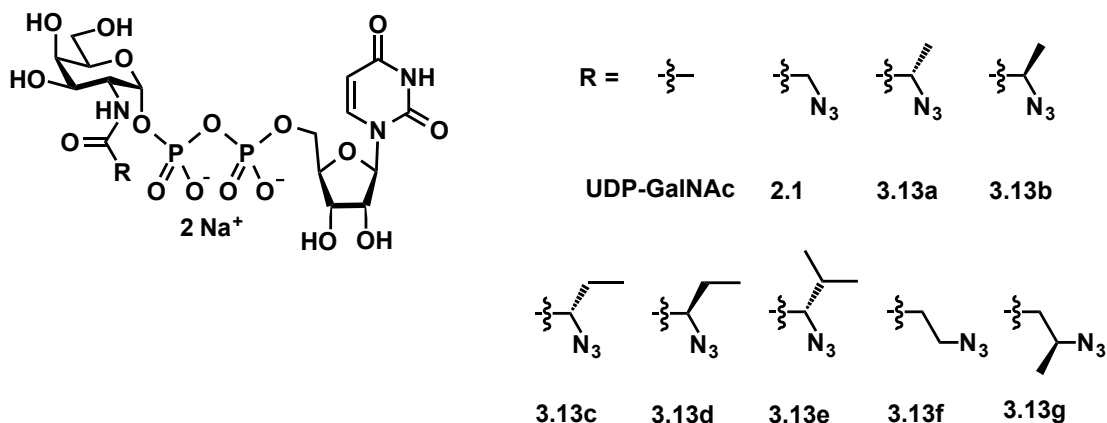
Figure 3-5. UDP-GalNAc analogs modified at C8 position.

We observed a broad range of percent conversion values for the UDP-GalNAc analogs. There is a significant increase in the conversion by NMR between analogs **3.13d** (60%) and analogs **3.13f** and **3.13g** (100% conversion). Analogs **3.13a-e** are branched at the alpha position, but **3.13f**, derived from β-alanine, and **3.13g**, derived from L-3-aminobutyric acid, contain a methylene at that position. This leads us to believe that substitutions at the alpha position introduce significant steric factors to the molecule that may influence activity with mutant ppGalNAcTs.

3.6. Experimental

3.6.1. General

All chemical reagents were analytical grade and commercially available, unless otherwise noted. Organic reactions were performed in oven-dried reaction vessels under an N_2 atmosphere, and liquid reagents were added with a dry needle unless otherwise noted. Sealed bottles of anhydrous methanol and anhydrous pyridine were purchased from Alfa Aesar and Acros Organics, respectively. All other solvents were purified as described by Pangborn *et al*.[48] Organic extracts were dried with $MgSO_4$ and solvent was removed with a rotary evaporator at reduced pressure. Flash chromatography was carried out with Silicycle SiliaFlash® P60 230-400 mesh silica gel according to the procedure described by Still *et al*.[49] Thin layer chromatography was carried out with Silicycle 250 micron 60 Å glass back silica gel plates and visualized with *p*-anisaldehyde staining or by absorbance of UV light at 245 nm.

Reverse phase-HPLC was performed using a modular Varian Prostar HPLC system with 210 and 254 nm detection on a Microsorb C_{18} Dynamax semi-preparative column (10 x 250 mm) at a flow rate of 4 mL/min or preparative column (21.4 x 250 mm) at a flow rate of 20 mL/min. Bio-Gel P-2 Gel was from BioRad. ^1H, ^{13}C, and ^{31}P NMR spectra were obtained with AVQ-400, AVB-400, DRX-500, AV-500 or AV-600 Bruker spectrometers. Chemical shifts (δ) are reported in parts per million referenced to the solvent peak and coupling constants (J) are reported in Hz. High resolution electrospray ionization (ESI) mass spectra were obtained from the UC Berkeley Mass Spectrometry Facility.

All biological reagents were molecular biology or cell culture grade and from commercial sources, unless otherwise specified. Chemically competent *E. coli* strain One Shot® Top10 and One Shot® BL21 (DE3) and S.O.C. medium were purchased from Invitrogen. Bacteria were grown in LB broth (Lennox L Broth Base, Invitrogen) supplemented with 100 µg/mL ampicillin (LB/amp) or 50 µg/mL kanamycin (LB/kan) and grown on LB/amp or LB/kan agar plates. The Plasmid Midi and Maxi kits and Ni-NTA agarose were from QIAGEN.

BamHI, BamHI-HF, NdeI, Antarctic phosphatase, and T4 DNA ligase were purchased from New England Biolabs (NEB). PfuUltra II Fusion HS DNA polymerase was from Agilent. 1 Kb and 100 bp DNA ladders were from Invitrogen. DNA Clean and Concentrator™-5, Zyppy™ Plasmid Miniprep Kit, and Zymoclean™ Gel DNA Recovery kit were from Zymo Research. Elim Biopharmaceuticals, Inc (Hayward, CA) synthesized all primers and performed all plasmid DNA sequencing. Integrated DNA Technologies (IDT) synthesized all gBlocks.

COmplete EDTA-free Protease inhibitor cocktail tablets were purchased from Roche. Criterion™ XT Bis-Tris Precast gels, XT-MES Running Buffer, nitrocellulose membrane (0.45 micron), Precision Plus Protein™ Kaleidoscope™ Prestained

Standards (10-250 kDa), and Kaleidoscope™ Prestained Standards (7.6-216 kDa) were obtained from BioRad. SuperSignal West Pico Chemilluminescent Substrate was purchased from Fisher Scientific.

Deactivated C_2 silica gel was generously provided by P. Shieh (C. Bertozzi Laboratory), and was generated by treating silica with ethyl-trichlorosilane, similar to the procedure described be Panne *et al.*[183].

3.6.2. General Methods

Manufacturer recommended protocols for Zymo Research kits were subject to the following modifications: Prior to elution, spin columns were dried by centrifugation at maximum speed for 2 min. DNA was eluted with water added directly to the filter, followed by 1 min incubation and 1 min centrifugation at maximum speed).

DNA was quantitated using a Nanodrop 2000 UV-vis spectrophotometer.

DNA samples were pre-mixed for sequencing: 500 ng plasmid DNA, 8 pM primer, 15 µL total in water.

3.6.3. General Protocol 1: Standard PCR conditions

PCR amplifications were carried out in duplicate for all cloning applications as recommended by the manufacturer: 25 ng template DNA, 0.2 µM forward primer, 0.2 µM reverse primer, 250 µM of each dNTP, 3% DMSO, 1x PfuUltra II buffer, and 1 µL PfuUltra II Fusion HS DNA polymerase in a final reaction volume of 50 µL in water.

Thermocycle conditions were 95 °C for 2 min; 30 cycles of 95 °C for 20 sec, annealing temp for 20 sec, 72 °C extension for 15sec/ kb; 72 °C for 3 min; 4 °C hold. Annealing temperatures were generally T_M – 5 °C and are listed with the primers. For primers with 5' overhangs, annealing temperatures are based on the T_M of the portion of the primer designed to anneal to the template. Extension times are listed with the primers.

3.6.4. General Protocol 2: Standard protocol for bacterial cloning

Bacterial transformations were carried out in One Shot® Top10 *E. coli* for cloning or One Shot® BL21 (DE3) *E. coli* for protein expression according to the manufacturer's instructions as described below. Modifications for specific applications are in the second paragraph. Bacteria (50 µL) were thawed on wet ice for 10 min, 2.5 – 125 ng DNA was added based on the application, tubes were flicked gently to mix, and *E. coli* were incubated on wet ice for 25 min. *E. coli* were subjected to heat shock at 42 °C for 30 sec, incubated on wet ice 2 min, and 250 µL rt S.O.C. medium was added. *E. coli* were grown shaking at 225 rpm at 37 °C for 35 (Amp[R]) – 45 (Kan[R]) min. Based on the application, 25-100 µL *E. coli* were spread onto LB plates with 100 µg/mL

ampicillin or 50 µg/mL kanamycin. Plates were incubated at 37 °C for approximately 16 h and then stored at 4 °C. Individual colonies were picked and used to inoculate 5 mL LB with antibiotics (100 µg/mL ampicillin or 50 µg/mL kanamycin). 5 mL cultures were grown shaking at 225 rpm at 37 °C for approximately 16 h, and cells were pelleted at 3,650 x g at 4 °C for 15 minutes. Supernatant was removed, pellets were flicked and then suspended in 600 µL water. DNA was extracted using a Zyppy™ Plasmid Miniprep Kit according to the manufacturer's protocol with modifications as described in the General Methods (30 µL elution).

Quantities of DNA added for transformations and the volume of transformed *E. coli* plated varied based on the application. When plasmid DNA was previously purified from *E. coli*, 0.25 µL (25-125 ng DNA) was used for transformation, and 25-35 µL *E. coli* were plated. When plasmid DNA was prepared via ligation, 2 µL ligation (2.5 ng DNA) was used for transformation, and 100-200 µL *E. coli* were plated.

3.6.5. General Protocol 3: Standard protocol for restriction digest cloning

Insert DNA was amplified as described in General Protocol 1. Target vectors were subjected to restriction digest (2.5 µg plasmid DNA, 1x NEB buffer, 1x BSA as necessary, 1 µL of each restriction enzyme in a 50 µL digest, 2 h at 37 °C). Both the vector and insert were isolated with a 0.8% agarose gel and purified with a Zymoclean™ Gel DNA Recovery kit according to the manufacturer's protocol with modifications as described in the General Methods (12 µL elution). The insert was subjected to restriction digest (volume of purified PCR product, 1x NEB buffer, 1x BSA as necessary, 1 µL of each restriction enzyme in a 50 µL digest, 2 h at 37 °C). The vector was treated with Antarctic phosphatase (volume of digested DNA, 1x Antarctic phosphatase buffer, 2 U (.4 µL) Antarctic phosphatase, and a 20 µL final volume in water, 1 h at 37 °C). Vector and insert were purified before ligation with DNA Clean and Concentrator™-5 according to the manufacturer's protocol with modifications as described in the General Methods (10 µL elution). A 3:1 molar ratio of insert to vector (25 ng vector) were ligated (1x NEB T4 DNA ligase buffer, 0.5 µL (1000 U) NEB T4 DNA, and a 10 µL final volume in water, 30 min at rt). Bacteria were transformed and clones were isolated as described in General Protocol 2.

3.6.6. General Protocol 4: SDS-PAGE

Samples were prepared by adding 4x SDS protein loading buffer (final concentrations 1x sodium dodecyl sulfate, 2.5 % 2-mercaptoethanol) and denatured by boiling at 95 °C for 5 min. Samples were centrifuged at maximum speed for 15 sec and loaded onto a 12 % or 4-12 % Bis-Tris precast gel. Electrophoresis was carried out in XT-MES buffer (BioRad) at 180 V for approximately 45 minutes. Gels were rinsed briefly with water following electrophoresis and stained with Coomassie Blue.

3.6.7. UDP-GalNAz synthesis

<u>1,3,4,6-Tetra-*O*-acetyl-2-azidoacetamido-2-deoxy-D-galactopyranoside (Ac$_4$GalNAz, **3.3**)</u>
This protocol was adapted from Hang *et al.* and Laughlin *el al.* for the chemical synthesis of UDP-GalNAz and Ac$_4$GalNAz, respectively[2,26,27]. The peracetylated sugar **3.3** is synthesized via the free sugar, **3.2**. To a 250-mL round-bottom flask was added azidoacetic acid (755 mg, 7.47 mmol), anhydrous methanol (46.7 mL), D-galactosamine hydrochloride (**3.1**, 1.00 g, 4.67 mmol), and triethylamine (1.67 mL, 12.0 mmol). The reaction mixture was stirred under an N$_2$ atmosphere for 5 min at room temperature, then cooled to 0 °C in an ice water bath. To the flask was then added 1-hydroxybenzotriazole hydrate (0.631 g, 4.67 mmol) and EDC (1.79 g, 9.33 mmol). The reaction was allowed to warm to room temperature overnight while stirring under an N$_2$ atmosphere and then concentrated *in vacuo* to give a dark orange syrup. The crude material was purified by silica gel flash chromatography eluting with 9:1 CH$_2$Cl$_2$: MeOH to afford compound **3.2**, which was carried forward without further purification.

To a 250-mL round-bottom flask was added compound **3.2** (all isolated from previous step, 5.65 mmol assuming 100% yield), anhydrous pyridine (30 mL), and acetic anhydride (30 mL, 0.317 mol). The reaction mixture was stirred under an N$_2$ atmosphere at room temperature overnight and then coevaporated with toluene in vacuo to give a pale yellow syrup. The crude product was dissolved in ethyl acetate (100 mL), washed with 1M aqueous HCl (20 mL), saturated NaHCO$_3$ (20 mL), and saturated NaCl (20 mL), dried over MgSO$_4$, and concentrated. The product was purified by silica gel flash chromatography eluting with a gradient of hexanes: ethyl acetate (3:1 to 1:2) to afford **3.3** as a white foam (0.751 g, 31% yield over two steps, 1.3: 1 α:β anomers). The α and β anomers were isolated by further purification by HPLC using C$_{18}$ column eluting with a gradient of CH$_3$CN/H$_2$O (5-50 %). The physical and spectroscopic properties match those of the literature.[2,27] α-anomer ^{1}H NMR (400 MHz, CDCl$_3$) δ 6.26 (d, NH, *J* = 9.2 Hz), 6.23 (d, 1 H, *J* = 3.6 Hz), 5.45 (dd, 1 H, *J* = 3.2, 1.2 Hz), 5.26 (dd, 1 H, *J* = 11.6, 3.2 Hz), 4.71 (ddd, 1 H, *J* = 11.6, 9.2, 3.6 Hz), 4.26 (app d, 1 H, *J* = 6.8 Hz), 4.10 (m, 2 H), 3.96 (s, 2 H), 2.19 (s, 3 H), 2.17 (s, 3 H), 2.04 (s, 3 H), 2.03 (s, 3 H); ^{13}C NMR (125 MHz, CDCl$_3$) δ 171.04, 170.41, 170.18, 168.89, 166.91, 90.96, 68.68, 67.64, 66.58, 61.24, 52.50, 47.03, 20.96, 20.75, 20.70, 20.69; β-anomer ^{1}H NMR (400 MHz, CDCl$_3$) δ 6.25 (d, NH, *J* = 8.8 Hz), 5.78 (d, 1 H, *J* = 8.4 Hz), 5.40 (dd, 1 H, *J* = 3.2, 0.8 Hz), 5.19 (dd, 1 H, *J* = 11.6, 3.6 Hz) 4.39 (dt, 1 H, *J* = 11.2, 9.2 Hz), 4.15 (m, 2 H), 4.04 (app td, 1 H, *J* = 6.4, 0.8 Hz), 3.93 (s, 2 H), 2.17 (s, 3 H), 2.13 (s, 3 H), 2.05 (s, 3 H), 2.02 (s, 3 H); ^{13}C NMR (126 MHz, CDCl$_3$) δ 170.41, 170.11, 169.35, 167.14, 166.90, 92.64, 71.91, 70.01, 66.27, 61.21, 52.62, 50.06, 20.91, 20.70, 20.69, 20.63; ESI-HRMS calcd. for C$_{16}$H$_{22}$N$_4$O$_{10}$Na [M+Na]$^+$ 453.1228, found 453.1228.

<u>3,4,6-Tri-*O*-acetyl-2-azidoacetamido-2-deoxy-D-galactopyranose (Ac$_3$GalNAz, **3.4**)</u>
<u>(3:1 mixture of α:β anomers)</u>

This protocol was adapted from Ravindranathan *et al.* for the chemical synthesis of nucleotide sugars[34]. To a 50-mL round bottom flask was added compound **3.3** (401 mg, 0.931 mmol), oven-dried, powdered 4 Å molecular sieves (0.400 g), and anhydrous methanol (10 mL). The reaction was allowed to stir at room temperature under N_2 and monitored by TLC analysis with 3% methanol in CH_2Cl_2. When significant conversion to product was detected and before major side products began to appear (8 h), the reaction was filtered and concentrated in vacuo to give a pale yellow syrup. The crude material was purified by silica gel chromatography eluting with 1% methanol in CH_2Cl_2 to give **3.4** as a white solid (142.4 mg, 39% yield). The physical and spectroscopic properties match those of the literature.[2,27] ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.78 (β, d, NH, *J* = 7.5 Hz), 6.52 (α, d, NH, *J* = 9.5 Hz), 5.42 (α, dd, 1 H, *J* = 3.0, 1.0 Hz), 5.38 (β, dd, 1 H, *J* = 3.0, 0.5 Hz), 5.34 (α, app t, 1 H, *J* = 3.0 Hz), 5.28 (α, dd, 1 H, *J* = 11.5, 3.0 Hz), 5.11 (β, dd, 1 H, *J* = 11.0, 3.0 Hz), 4.85 (β, app d, 1 H, *J* = 8.0 Hz), 4.67 (β, app t, 1 H, *J* = 8.3 Hz), 4.56 (α, m, 1 H), 4.44 (α, app t, 1 H, *J* = 6.5 Hz), 4.12 (m, 5 H), 4.03 (β, s, 2 H), 3.96 (α, d, 2 H, *J* = 0.5 Hz), 3.34 (α, d, OH, *J* = 2.0 Hz), 2.17 (s, 6 H), 2.06 (s, 9 H), 2.01 (α, s, 3 H); α-anomer ^{13}C NMR (126 MHz, $CDCl_3$) δ 170.92, 170.68, 170.34, 167.06, 92.02, 67.93, 67.42, 66.69, 62.02, 52.57, 48.05, 20.78, 20.76, 20.68. ESI-HRMS calcd. for $C_{14}H_{20}N_4O_9Na$ $[M+Na]^+$ 411.1118, found 411.1122.

Diallyl 3,4,6-Tri-*O*-acetyl-2-azidoacetamido-2-deoxy-α-D-galactopyranosyl-1-phosphate (protected GalNAz-1-P**, 3.5**)

This protocol was adapted from Hang *et al.* for the chemical synthesis of UDP-GalNAz[2,27]. To a 10-mL round bottom flask was added compound **3.4** (67.8 mg, 0.175 mmol), CH_2Cl_2 (2 mL), and 1H-tetrazole (59.3 mg, 0.846 mmol). The reaction was stirred under N_2 for 10 min at room temperature and diallyl N,N-diisopropylphosphoramidite (0.139 mL, 0.509 mmol) was added dropwise. The reaction mixture was stirred for 3 h under N_2 at room temperature, cooled to –40 °C, and mCPBA (146 mg, 0.846 mmol) was added. The reaction was stirred for 10 min at –40 °C and then warmed for 1 h in an ice water bath. The reaction mixture was diluted in CH_2Cl_2 (20 mL), washed with 10 % aqueous Na_2SO_3 (2 x 20 mL), saturated $NaHCO_3$ (20 mL), and water (20 mL). The organic layer was dried over anhydrous $MgSO_4$, filtered, and concentrated in vacuo, and the crude product was purified by silica gel chromatography eluting with ethyl acetate: hexanes (2:1) to afford **3.5** as a white solid (24.8 mg, 27 % yield). The physical and spectroscopic properties match those of the literature.[2,27] ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.55 (d, NH, *J* = 9.2), 5.95 (m, 2 H), 5.78 (dd, 1 H, *J* = 6.0, 3.2 Hz), 5.46 (dd, 1 H, *J* = 3.2, 1.0 Hz), 5.40 (ddq, 2 H, *J* = 17.2, 6.4, 1.2 Hz), 5.31 (m, 2 H), 5.24 (dd, 1 H, *J* = 11.4, 3.0 Hz), 4.60 (m, 5 H), 4.43 (app t, 1 H, *J* = 6.4 Hz), 4.11 (m, 2 H), 3.95 (s, 2 H) 2.16 (s, 3 H), 2.03 (s, 3 H), 2.01 (s, 3 H); ^{13}C NMR (151 MHz, $CDCl_3$) δ 170.02, 170.56, 170.35, 167.57, 132.36 (d, *J* = 6.6 Hz), 132.32 (d, *J* = 6.6 Hz), 119.36, 119.30, 96.56 (d, *J* = 6.15 Hz), 69.16 (d, *J* = 4.5 Hz), 69.07, 69.02 (d, *J* = 4.5 Hz), 67.56, 67.05, 61.69, 52.78, 48.25 (d, *J* = 7.5 Hz), 20.96, 20.93; ^{31}P NMR (162 MHz, $CDCl_3$) δ -2.22. ESI-HRMS calcd. for $C_{20}H_{29}N_4O_{12}PNa$ $[M+Na]^+$ 571.1410, found 571.1412.

Uridine 5'-diphospho-2-azidoacetamido-2-deoxy-α-D-galactopyranose disodium salt (UDP-GalNAz, **2.1**).

UDP-GalNAz was synthesized from **3.5** as described in the literature[2,27]. Compound **2.1** was purified by size exclusion chromatography and quantified as described for compound **3.8**. ^{1}H NMR (400 MHz, Deuterium Oxide) δ 7.92 (d, J = 8.1 Hz, 1H), 5.99 – 5.89 (m, 2H), 5.54 (dd, J = 7.2, 3.4 Hz, 1H), 4.34 (p, J = 5.2 Hz, 2H), 4.29 – 4.08 (m, 6H), 4.04 (d, J = 16.3 Hz, 2H), 3.96 (dd, J = 10.9, 3.1 Hz, 1H), 3.76 – 3.68 (m, 2H). ^{13}C NMR (101 MHz, Deuterium Oxide) δ 171.14, 166.22, 151.79, 141.67, 102.62, 94.56 (d, J = 6.3 Hz), 88.53, 83.17 (d, J = 9.1 Hz), 73.76, 72.10, 69.63, 68.37, 67.51, 65.02 (d, J = 5.5 Hz), 61.02, 51.61, 49.90 (d, J = 8.3 Hz). ^{31}P NMR (162 MHz, Deuterium Oxide) δ -11.22 (d, J = 21.0 Hz), -12.86 (d, J = 21.0 Hz). ESI-HRMS calcd. for $C_{17}H_{25}N_6O_{17}P_2$ [M – 2 Na + H]$^-$ 647.0757, found 647.0738.

3.6.8. Enzymatic synthesis of UDP-GalN

3.6.8.1. Cloning enzymes for UDP-GalN synthesis

The Galactose-1P Uridyltransferase (GalPUT, P09148, *E. coli*) in the pET15b expression plasmid was generously provided by the R. Field Laboratory (John Innes Centre, Norwich, England). Inorganic pyrophosphatase (PPiase, PmPpA, *Pasteurella multocida*) was cloned by the X. Chen Laboratory[184], was cloned into pET23a by P. Wu (Albert Einstein College of Medicine, Bronx, NY), and was provided by J. Hudak (C. Bertozzi Laboratory). The Galactokinase (GalK, pDB88, *E. coli*), Pyruvate kinase (PyK2, P413TEF-PYK2, *S. cerevisiae*), and UDP-Glc pyrophosphorylase (UDPGP, UDPGP, *T. brucei*) were purchased from Addgene. All clones from Addgene were streaked onto LB plates with the appropriate antibiotic, and DNA was isolated as described in General Protocol 2. GalK and PyK2 enzymes were cloned into the pET15b vector as described below. All enzyme constructs was transformed into BL21(DE3) *E. coli* for expression as described in General Protocol 2.

GALK and PYK2 were cloned into pET15b (5.7 kb) as described in General Protocol 3. The vector was digested at the BamHI and NdeI sites (1x NEB4 buffer, 1x BSA, 0.5 μL (10 U) BamHI-HF™, 1 μL (20 U) NdeI in a 50 μL digest, 2 h at 37 °C). GalK and PyK2 were amplified with primers in Table 3-1. The insert was digested as described for the vector. The transformants were screened by colony PCR and sequenced with primers in Table 3-1 as described in the General Methods.

Primer ID	Sequence	Annealing Temp (°C)
GalKpet15bF	5'- agccatATGAGTCTGAAAGAAAAAACACAATCTCTG	52
GalKpet15bR	5'- gccggatccTCAGCACTGTCCTGCTCCTTG	52
GalPUTseq1F	5'- caaccgcacctgtggcgccgg	59
GalPUTseq4R	5'- cctggatgctgtaggcataggc	53
PYK2petlongF	5'- ggcagccatATGCCAGAGTCCAGATTGCAGAGAC	53
PYK2petlongR	5'- gccggatccCTAGAATTCTTGACCAACAGTAGAAATGC	53

Table 3-1. Primers used for cloning enzymes needed in the enzymatic synthesis of UDP-GalN. Gene sequences are in uppercase and vector sequence is in lowercase. PCR amplicon for GalK is 1.2 kb(30 sec extension) and for PYK2 is 1.5 kb (40 sec extension). GalKpet15bF and R were used for amplification and colony PCR of GalK. PYK2petlongF and R were used for amplification and colony PCR of PYK2. GalPUTseq1F and GalPUTseq4R were used for sequencing clones.

3.6.8.2. Bacterial protein expression and purification

Single colonies of each of the five enzymes in BL21(DE3) *E. coli* were used to inoculate a 5 mL of LB/amp broth. Two cultures of GalK were inoculated. Cultures were grown at 225 rpm and 37 °C overnight to $A_{600} \geq 2$. Cultures were diluted to $A_{600} = 0.05$ into 200mL fresh LB/amp broth and grown for 2 h to $A_{600} \geq 0.8$. Protein expression was induced with 0.5 mM IPTG (1.5 mM IPTG for UDPGP and PyK2) and cultures were grown at 16°C at 225 rpm for 20 h. Aliquots (100 µL) were removed immediately (t_0), and at 3 and 20 h. After the overnight induction, the cultures of each enzyme were centrifuged at 3650 x g for 15 min at 4 °C.

Supernatant was removed and cell pellets were resuspended in Lysis Buffer (20 mM Tris-Cl pH 8.5, 500 mM NaCl, 5 mM $MgCl_2$, 1% Triton X-100, and freshly added lysozyme at 200 µg/mL and 1x EDTA-free protease inhibitors) and mixed on a shaker at room temperature for 10 min. DNAse I (2 µg/mL) was added and the samples were incubated at 225 rpm, 37 °C for 1 h. The lysate was then centrifuged at 3650 x g for 15 min and the supernatant was collected. The two GalK lysate samples were combined.

Ni-NTA agarose resin (50% slurry, Qiagen) was added to each lysate sample (volumes of slurry: GalK 8 mL, PYK2 2 mL, GalPUT 2.5 mL, UDPGP 2 mL, PPiase 3 mL). An explanation of the resin volumes used is below. The lysate slurry rotated at 4 °C for 1 h, samples were centrifuged at 1000 x g for 3 min, 4 °C and the supernatant was discarded. The pelleted agarose beads were washed twice by gently re-suspending the pelleted agarose beads in 10 mL of Wash Buffer (50 mM NaH_2PO_4, pH 8.5, 300 mM NaCl, 20 mM imidazole) with a centrifugation step in between the washes. Finally, the beads were suspended in resin volumes of Storage Buffer (20 mM Tris-Cl pH 8.5, 500 mM NaCl, 5 mM $MgCl_2$).

For expression analysis by SDS-PAGE, induction aliquots were mixed with Loading Buffer (1x sodium dodecyl sulfate (SDS), 2.5% 2-mercaptoethanol) and were lysed by boiling at 95°C for 10 min. 20µL of these lysed samples were reserved for SDS-PAGE analysis. A 40 µL aliquot of each immobilized bead slurry was removed and centrifuged. The supernatant (20 µL) was discarded and 20 µL of Elution Buffer (50 mM NaH_2PO_4 pH 8.5, 300 mM NaCl, 250 mM imidazole) was added. The sample was mixed and allowed to incubate for 1 min before centrifuging and removing the 20µL supernatant sample for SDS-PAGE analysis. The elution step was repeated twice. 20 µL elution samples, clarified lysate, flow-through, and wash samples were subjected to SDS-PAGE as described in General Protocol 4 (Figure 3-6).
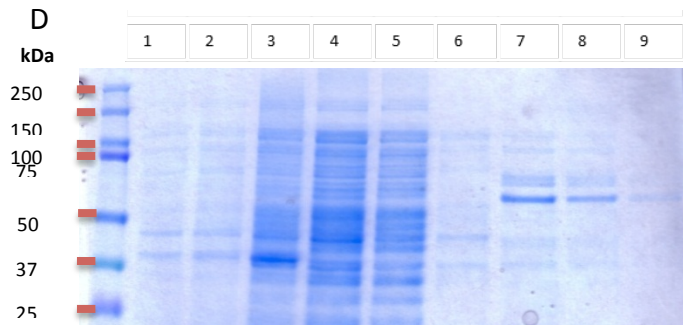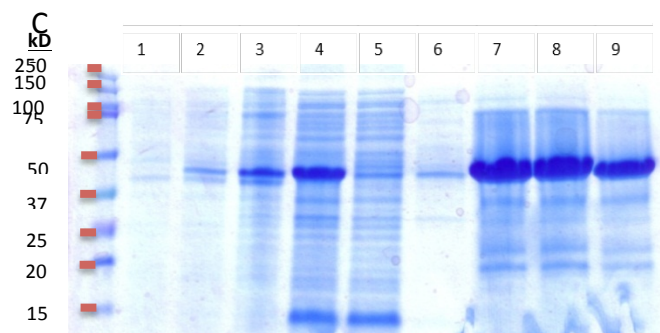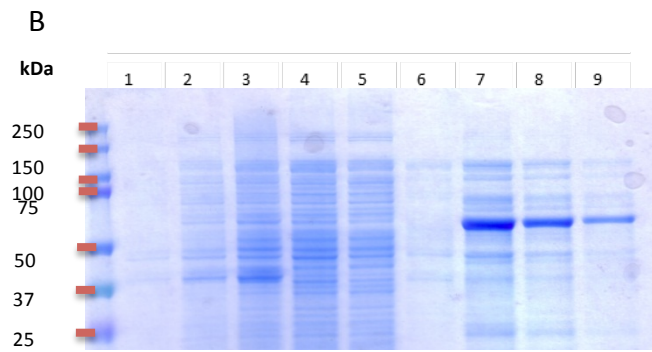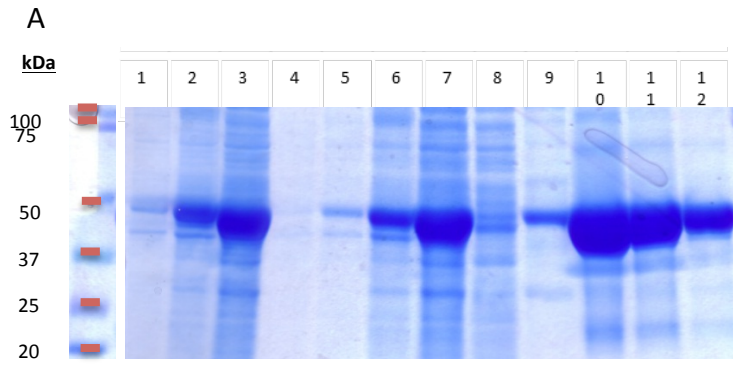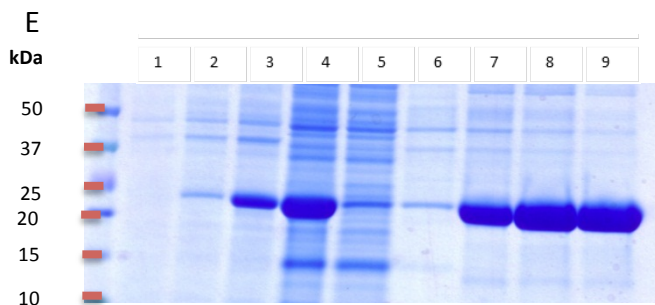
Figure 3-6. Expression and immobilization of biosynthetic enzymes. (A) GalK 41.4 kDa; Lane 1-3 induction at 0, 3, 20 h; Lane 4 lysate; Lane 5-7 induction at 0, 3, 20 h; Lane 8 lysate; Lane 9 flow-through; Lane 10 2nd wash; Lane 11-13 elutions 1-3, (B) PyK2 55.2 kDa, (C) GalPUT 42.6 kDa, (D) UDPGP 52.5 kDa, (E) PPiase 19.3 kDa. (B, C, D, E) Lane 1-3 induction at 0, 3, 20 h; Lane 4 lysate; Lane 5 flow-through; Lane 6 2nd wash; Lane 7-9 elutions 1-3.

To roughly quantitate the amount of enzyme used in each reaction, the lower binding capacity of the Ni-NTA resin (5 mg/mL resin) was used to approximate the mass of immobilized enzyme. The quantity of resin added to the lysate was adjusted according to the expected protein expression levels as indicated by previous BCA assays. The binding saturation point was found by sequential expression experiments where the volume of resin added to the lysate was incrementally adjusted so that there was minimal unbound protein in the lysate after the binding step in the protein purification.

For early protein quantitation experiments, after lysates were generated as described above, 5 mL of Ni-NTA agarose slurry was incubated with the clarified whole cell lysate at 4 °C for 1 hour. The lysate slurry was applied to a 25 mL disposable column (2.5 mL bed volume) and the flow-through was collected. The column was washed 2x with 8 volumes of Wash Buffer and the protein was eluted with one volume of Elution Buffer four times. The four elution fractions were combined in a 15 mL 30K MW cut-off spin filter and centrifuged at 3650 x *g* for 10 minutes at 4 °C. The flow-through was discarded and the filter was washed with 10 mL Dialysis Buffer (20 mM Tris-Cl pH 7.4) and centrifuged under the same conditions as the previous step. The residual volume in the filter was then collected and the filter was washed with 500 µL Dialysis Buffer and combined with the residual volume. The concentrated and purified enzymes were stored at -20 °C. The concentration was analyzed with a Pierce BCA Protein Assay Kit according to the manufacturer's instructions.

3.6.8.3. UDP-GalN synthesis

This protocol was adapted from Heidlas *et al.*, Chen *et al.*, and Errey *et al.* for the enzymatic synthesis of nucleotide sugars[176,177,185]. Galactosamine (**3.1**, 220 µmol), ATP (7.2 µmol), UDP-Glucose (6.4 µmol), phosphoenolpyruvate (214 µmol) and Uridine triphosphate (152 µmol) were combined in 6 mL reaction buffer (50 mM HEPES pH 8, 10 mM $MgCl_2$, 5 mM KCl) in a 15 mL conical, and the pH was adjusted

to 8.5 with 1M KOH. The immobilized enzymes (Table 3-2) were added, and the reaction was mixed by rotating at 30 °C and monitored by TLC (3:1 EtOH:10mM tributylammonium bicarbonate mobile phase, UV and *p*-anisaldehyde stain: 2% glacial AcOH, 4% H_2SO_4, 2% p-anisaldehyde in EtOH). The reaction was stopped after approximately 45 hours and centrifuged at 10,000 x g for 15 minutes. The supernatant was removed and the resin was washed twice with 10 mL water. The washes were added to the original reaction mixture, frozen, and lyophilized.

Enzyme	Volume of bead slurry (mL)
Galactokinase	2.6
Pyruvate Kinase	1.06
Galactose-1P Uridyltransferase	0.4
UDP-Glc Pyrophosphorylase	0.8
Inorganic Pyrophosphatase	0.4

Table 3-2. Quantity of enzymes for UDP-GalN biosynthetic reactions. Volume of bead slurry is used to approximate the amount of enzyme used in each reaction as the beads have a protein binding capacity of 5-10 mg/mL resin. We estimate that 5 mg/mL of enzyme is bound to the Ni-NTA resin.

The lyophilized samples were dissolved in water and purified by size exclusion chromatography. Bio-Gel P-2 size exclusion columns with bed volumes of 1 L and 0.5 L were packed according to manufacturer's protocol (Table 3-3). Elution volumes from column A were approximately 5mL, and UDP-GalN eluted at approximately 20 hours. The purity of the fractions was determined by TLC analysis visualized by UV and *p*-anisaldehyde staining with either 3:1 EtOH:10mM tributylammonium bicarbonate or 3:1 EtOH:0.5M NH_4OAc as the mobile phase. The fractions were combined based on purity, frozen, and lyophilized. Following lyophilization, impure fractions were dissolved in water, loaded onto column B, and collected in 2.5 mL fractions. Multiple rounds of chromatography were used to improve resolution of UDP-GalN from side products of the enzymatic synthesis, many of which are of similar molecular weight to UDP-GalN. This procedure has been carried out on larger scales of up to 2.2 mmol compound **3.1** in similar yields.

Column	Resin volume (L)	Diameter (cm)	Cross sectional area (C.S.A.) cm²	Optimal flow rate (mL/h)	Max optimal loading volume (mL)
A	1	5	19.6	98.2	20-25
B	0.5	2.5	4.9	24.5	10-15

Table 3-3. Size exclusion chromatography parameters. Optimal flow rate (calculated by manufacturer's recommendation of C.S.A. x 5).

3.6.8.4. Quantitation of UDP-GalN

After UDP-GalN was purified and analyzed by NMR, we found the mass of inorganic salts was higher than expected, so UV-vis spectrometry was used to obtain a more

accurate yield for the product. The extinction coefficient for UDP-GalN has not been published, so an approximate extinction coefficient of UDP-GalN was obtained by averaging the measured extinction coefficients of UDP-GalNAc, UDP-Glc, and UTP in water. The absorbance of a standard curve with known concentrations of each standard was measured at 260 nm. The extinction coefficients were calculated and confirmed to match those reported by the literature. A standard curve of UDP-GalN by mass was measured at 260 nm, and the molarity of these solutions was extrapolated (ε = 9374 cm^{-1}M^{-1}). The approximate mass of UDP-GalN was calculated, and approximately half of the mass of the sample was attributed to inorganic salts.
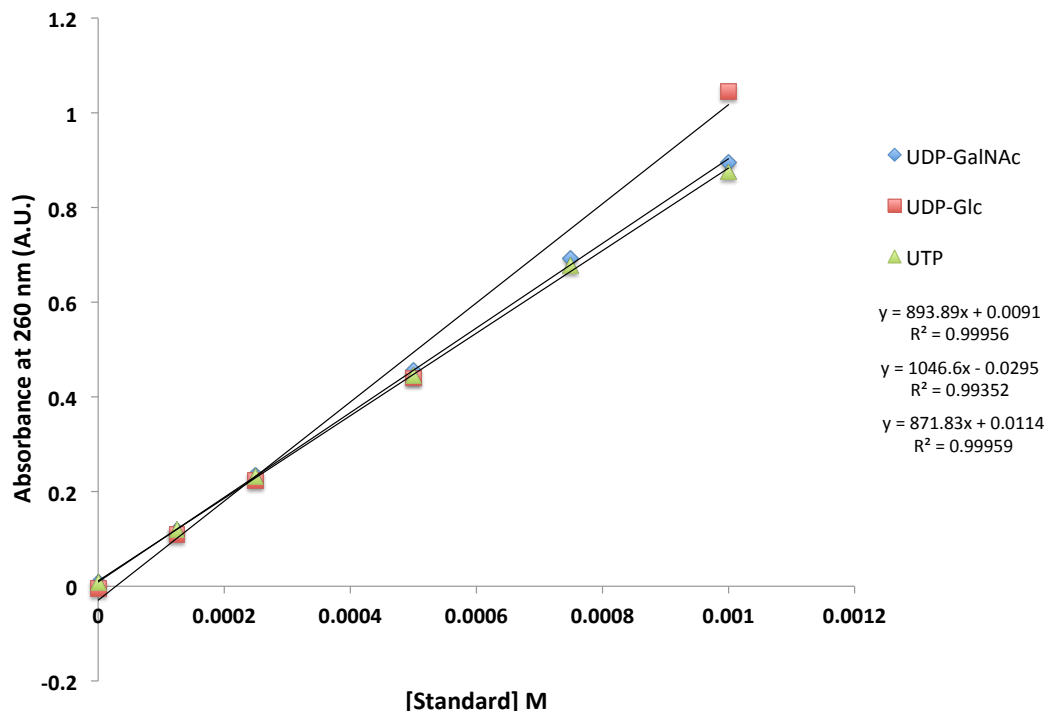
Figure 3-7. Extrapolating the extinction coefficient of UDP-GalN. Absorbance vs. concentration plots of UDP-GalNAc (ε = 8939 cm^{-1} M^{-1}), UDP-Glc (ε = 10466 cm^{-1} M^{-1}), and UTP (ε = 8718 cm^{-1} M^{-1}) afforded the average extinction coefficient of 9374 cm^{-1} M^{-1} that was used to quantitate UDP-GalN.

UDP-galactosamine disodium salt (UDP-GalN, **3.8**)
The protocol described above afforded compound **3.8** in 65% yield (143 µmol). ^{1}H NMR (400 MHz, Deuterium Oxide) δ 7.96 (d, J = 8.1 Hz, 1H), 6.01 – 5.96 (m, 2H), 5.77 (dd, J = 6.6, 3.5 Hz, 1H), 4.38 (t, J = 4.6 Hz, 2H), 4.33 – 4.18 (m, 4H), 4.04 (d, J = 2.9 Hz, 1H), 3.98 (dd, J = 11.0, 3.0 Hz, 1H), 3.83 – 3.71 (m, 2H), 3.34 (d, J = 11.1 Hz, 1H). ^{13}C NMR (101 MHz, Deuterium Oxide) δ 166.49, 151.99, 141.55, 93.95 (d, J = 5.5 Hz), 88.59, 83.02 (d, J = 9.0 Hz), 73.70, 72.21, 69.55, 68.04, 67.44, 65.06 (d, J = 5.9 Hz), 60.87, 50.80 (d, J = 8.7 Hz). ^{31}P NMR (162 MHz, Deuterium Oxide) δ -11.21 (d, J = 20.7 Hz), -13.18 (d, J = 21.1 Hz). ESI-HRMS calcd. for $C_{15}H_{24}N_3O_{16}P_2$ [M – 2 Na + H]$^-$ 564.0637, found 564.0627.

3.6.9. UDP-GalNAc analog synthesis

3.6.9.1. Azido acids

Amino acid (**3.9a** and **c-g**, 0.8 mmol), K_2CO_3 (2.26 mmol, 312.35 mg), and $CuSO_4 \cdot 5H_2O$ (8.0 µmol, 2 mg) were combined in a dry flask and dissolved in 5 mL dry MeOH. Imidizole-1-sulfonyl-azide tetrafluoroborate salt (**3.10**, 0.96 mmol, 260.97 mg) was added incrementally over 1 min. The flask was evacuated and the pale blue reaction mixture was stirred under N_2 at rt overnight. The reaction mixture was evaporated to dryness and dissolved in 10-15 mL H_2O. 1N HCl was added to a pH of 1-2. The aqueous mixture was extracted with ethyl acetate (3 x 10 mL). The organic extract was dried over $MgSO_4$ and evaporated to dryness to afford a colorless to pale yellow oil. The azido acid **3.11**(**a** and **c-g**) was carried forward without further purification.

Me(S) **3.11a**
The azido acid was synthesized from L-alanine. ^{1}H NMR (600 MHz, Acetonitrile-d_3) δ 4.06 (q, J = 7.1 Hz, 1H), 1.40 (d, J = 7.2 Hz, 3H). ^{13}C NMR (151 MHz, Acetonitrile-d_3) δ 172.91, 57.89, 17.09.

Et(S) **3.11c**
The protocol described above was followed with different quantities of starting materials: L-2-aminobutryic acid (1.16 mmol, 119.62 mg), K_2CO_3 (3.05 mmol, 421.54 mg), $CuSO_4 \cdot 5H_2O$ (11.6 µmol, 2.90 mg), imidizole-1-sulfonyl-azide tetrafluoroborate salt (**3.10**, 1.392 mmol, 363.27 mg) ^{1}H NMR (400 MHz, Acetonitrile-d_3) δ 3.95 (dd, J = 7.9, 5.2 Hz, 1H), 1.91 – 1.67 (m, 2H), 0.97 (t, J = 7.4 Hz, 3H). ^{13}C NMR (101 MHz, Acetonitrile-d_3) δ 172.37, 63.71, 25.49, 10.37.

Et(R) **3.11d**
The azido acid was synthesized from D-2-aminobutyric acid. ^{1}H NMR (400 MHz, Acetonitrile-d_3) δ 3.95 (dd, J = 7.9, 5.1 Hz, 1H), 1.91 – 1.67 (m, 2H), 0.97 (t, J = 7.4 Hz, 3H). ^{13}C NMR (101 MHz, Acetonitrile-d_3) δ 172.07, 63.70, 25.48, 10.36.

Ipr(S) **3.11e**
The azido acid was synthesized from L-valine. ^{1}H NMR (600 MHz, Acetonitrile-d_3) δ 3.89 (d, J = 5.3 Hz, 1H), 2.21 – 2.12 (m, J = 6.7 Hz, 1H), 0.99 (d, J = 6.8 Hz, 3H), 0.93 (d, J = 6.8 Hz, 3H). ^{13}C NMR (151 MHz, Acetonitrile-d_3) δ 171.76, 68.40, 31.50, 19.62, 17.80.

betaAla **3.11f**
The azido acid was synthesized from beta-alanine. ^{1}H NMR (400 MHz, Acetonitrile-d_3) δ 3.51 (t, J = 6.3 Hz, 2H), 2.54 (t, J = 6.3 Hz, 2H). ^{13}C NMR (101 MHz, Acetonitrile-d_3) δ 172.82, 47.61, 33.87.

L-3-butyric **3.11g**

The azido acid was synthesized from L-3-aminobutyric acid. ^{1}H NMR (300 MHz, Acetonitrile-d_3, Reference peak = EtOAc chemical shift in CD3CN, triplet 1.20 ppm) δ 3.93 (h, J = 6.4 Hz, 1H), 2.56 – 2.37 (m, 2H), 1.26 (d, J = 6.5 Hz, 3H). ^{13}C NMR (75 MHz, Acetonitrile-d_3) δ 173.30, 55.32, 40.91, 19.59.

3.6.9.2. NHS-esters

NHS (0.8 mmol, 92.08 mg) and azido acid (**3.11a-g,** all isolated from previous step, 0.8 mmol assuming 100% yield) were combined in a dry flask and dissolved in 2-5 mL DMF. The mixture was cooled to 0 °C in an ice water bath and the flask was evacuated. DCC (0.8 mmol, 165.06 mg) was then added slowly before evacuating the flask and allowing the reaction to warm to rt overnight under N_2. A white precipitate formed approximately 30 min after the addition of DCC. The reaction was filtered, washed with MeCN, and evaporated to dryness followed by purification by C_2 silica gel column chromatography in hexanes: ethyl acetate (4:1 or 3:1). Prior to loading the sample, the modified silica was washed with four column volumes of the mobile phase. The sample was loaded in the mobile phase + DCM. The NHS-ester (**3.12a-g**) is a colorless to light yellow oil.

Me(S) **3.12a**
The yield was 55.6% over 2 steps (0.445 mmol, 94.4 mg). ^{1}H NMR (600 MHz, Acetonitrile-d_3) δ 4.53 (q, J = 7.1 Hz, 1H), 2.80 (s, 4H), 1.56 (d, J = 7.1 Hz, 3H). ^{13}C NMR (151 MHz, Acetonitrile-d_3) δ 170.60, 168.36, 56.42, 26.40, 17.29.

(MeR) **3.12b**
The commercial (D) azidoalanine (CHA salt, Chiralix) (1.61 mmol, 221.2 mg) was subjected to a sodium salt exchange followed by an acidic workup where it was dissolved in 5 mL water and 1M HCl was added to pH 1. The azido acid was then extracted with anhydrous ether (3 x 5 mL). The organic layer was dried over $MgSO_4$ and concentrated by rotary evaporation, affording a colorless liquid. The protocol for NHS-ester synthesis described above was followed with different quantities of starting materials: NHS (1.61 mmol, 185.3 mg), azido acid (1.61 mmol, 221.2 mg), DCC (1.61 mmol, 332.19 mg). The yield was 65.6% (1.07 mmol, 227.8 mg). ^{1}H NMR (400 MHz, Acetonitrile-d_3) δ 4.53 (q, J = 7.1 Hz, 1H), 2.80 (s, 4H), 1.56 (d, J = 7.1 Hz, 3H). ^{13}C NMR (101 MHz, Acetonitrile-d_3) δ 170.58, 168.29, 56.37, 26.36, 17.26.

Et(S) **3.12c**
The protocol described above was followed with different quantities of starting materials: NHS (1.16 mmol, 133.5 mg), azido acid (all isolated from previous step, 1.16 mmol assuming 100% yield), DCC (1.16 mmol, 239.3 mg). 201 mg of product was recovered with some NHS contamination, and the yield was approximated by integration values to afford a 53% yield over 2 steps (0.62 mmol). ^{1}H NMR (400 MHz, Acetonitrile-d_3) δ 4.42 (dd, J = 7.2, 5.8 Hz, 1H), 2.80 (s, 4H), 2.07 – 1.84 (m, 2H), 1.07 (t, J = 7.4 Hz, 3H). ^{13}C NMR (101 MHz, Acetonitrile-d_3) δ 170.65, 167.79, 62.03, 26.42, 25.93, 9.87.

<u>Et(R)</u> **3.12d**
The yield was 84% yield over 2 steps (0.675 mmol, 152.7 mg). ^{1}H NMR (400 MHz, Acetonitrile-d_3) δ 4.42 (dd, J = 7.2, 5.8 Hz, 1H), 2.80 (s, 4H), 2.07 – 1.84 (m, 2H), 1.07 (t, J = 7.4 Hz, 3H).^{13}C NMR (101 MHz, Acetonitrile-d_3) δ 170.65, 167.79, 62.03, 26.42, 25.93, 9.87.

<u>Ipr(S)</u> **3.12e**
The yield was 59% over 2 steps (0.47 mmol, 114.4 mg). ^{1}H NMR (600 MHz, Chloroform-d) δ 4.07 – 4.00 (m, 1H), 2.86 (s, 4H), 2.38 – 2.28 (m, J = 6.6 Hz, 1H), 1.11 (dd, J = 17.3, 6.8 Hz, 6H). ^{13}C NMR (151 MHz, Chloroform-d) δ 168.62, 165.87, 66.18, 31.67, 25.72, 19.18, 17.90.

<u>betaAla</u> **3.12f**
The yield was 62.4% over 2 steps (0.50 mmol, 105.8 mg). ^{1}H NMR (400 MHz, Acetonitrile-d_3) δ 3.66 (t, J = 6.1 Hz, 1H), 2.90 (t, J = 6.2 Hz, 1H), 2.78 (s, 2H). ^{13}C NMR (101 MHz, Acetonitrile-d_3) δ 170.91, 168.28, 47.02, 31.62, 26.35.

<u>L3butyric</u> **3.12g**
The yield was 67.5% over 2 steps (0.54 mmol, 121.8 mg). ^{1}H NMR (400 MHz, Acetonitrile-d_3) δ 4.04 (dq, J = 12.1, 6.3 Hz, 1H), 2.94 – 2.68 (m, 6H), 1.35 (d, J = 6.4 Hz, 3H). ^{13}C NMR (101 MHz, Acetonitrile-d_3) δ 170.92, 167.54, 54.93, 38.18, 26.37, 19.37.

3.6.9.3. UDP-GalNAc analogs

UDP-GalN (**3.8**, 1 equivalent) was dissolved in 4-6 mL water and NHS ester (**3.12a-g**, 2 equivalents) in 0.5 mL MeCN was added. Reaction was adjusted to pH 7.5 – 8 with 1 M KOH, if necessary. The reaction proceeded at rt for 2 h and NHS-ester addition and pH adjustment were repeated 3 x every 2 h for a total of 8 equivalents of NHS-ester. After final addition, the reaction proceeded at rt overnight. The mixture was evaporated to dryness and a crude NMR was used to approximate product conversion. The mixture was purified by size exclusion chromatography in water. The purity of the fractions was determined by TLC analysis visualized by UV and *p*-anisaldehyde staining with either 3:1 EtOH:10mM tributylammonium bicarbonate or 3:1 EtOH:0.5M NH$_4$OAc as the mobile phase. The fractions were combined based on purity, frozen, and lyophilized to yield a white powder. The purified UDP-GalNAc analog (**3.13a-g**) was quantitated by absorbance measurements at 260 nm using an extinction coefficient of 9374 cm^{-1}M^{-1} as described for UDP-GalN (**3.8**).

<u>UDP-GalNAzMe(S)</u> **3.13a**
UDP-GalN (**3.8**, 0.222mmol, 1 equivalent) was dissolved in 12.5 mL water and NHS ester (**3.12a**, 0.445 mmol, 94.4 mg, 2 equivalents) in 3.5 mL *p*-dioxane was added. Reaction was adjusted to pH 7.5 – 8 with 1 M KOH. The reaction proceeded at rt overnight. The mixture was evaporated to dryness and a crude NMR was used to approximate product conversion. The mixture was purified by size exclusion

chromatography in water. The purity of the fractions was determined by TLC analysis visualized by UV and *p*-anisaldehyde staining with either 3:1 EtOH:10mM tributylammonium bicarbonate or 3:1 EtOH:0.5M NH$_4$OAc as the mobile phase. The fractions were combined based on purity, frozen, and lyophilized. The product was purified, subjected to sodium salt exchange, lyophilized to yield a white powder. Some fractions were lost during purification, affording analog **3.13a** in 14.7% yield (23.1 mg, 28% conversion by NMR). ^{1}H NMR (400 MHz, Deuterium Oxide) δ 7.93 (d, *J* = 8.1 Hz, 1H), 6.00 – 5.90 (m, 2H), 5.54 (dd, *J* = 7.0, 3.4 Hz, 1H), 4.39 – 4.31 (m, 2H), 4.30 – 4.12 (m, 6H), 4.07 – 3.92 (m, 2H), 3.74 (h, *J* = 7.0 Hz, 2H), 1.46 (d, *J* = 7.0 Hz, 3H). ^{13}C NMR (101 MHz, Deuterium Oxide) δ 174.44, 166.23, 151.81, 141.67, 102.63, 94.70 (d, *J* = 6.3 Hz), 88.50, 83.22 (d, *J* = 9.2 Hz), 73.79, 72.12, 69.67, 68.51, 67.50, 65.02 (d, *J* = 5.6 Hz), 61.02, 58.20, 49.90 (d, *J* = 8.4 Hz), 16.89. ^{31}P NMR (243 MHz, Deuterium Oxide) δ -11.27 (d, *J* = 19.9 Hz), -13.07 (d, *J* = 19.8 Hz). ESI-HRMS calcd. for C$_{18}$H$_{27}$N$_6$O$_{17}$P$_2$ [M – 2 Na + H]$^-$ 661.0913, found 661.0897.

UDP-GalNAzMe(R) **3.13b**

The protocol described above was followed with the following modifications: a single addition of 10 equivalents of NHS-ester (**3.12b**, 1.07 mmol, 227.8 mg) in 1.5 mL MeCN, UDP-GalN (**3.8**, 0.107 mmol), and the reaction was allowed to stir at rt overnight. The product was purified, subjected to sodium salt exchange, lyophilized to yield a white powder, and quantitated as above. Analog **3.13b** was afforded in 15.7% yield (16.8 μmol, 34% conversion by NMR). ^{1}H NMR (400 MHz, Deuterium Oxide) δ 7.93 (d, *J* = 8.1 Hz, 1H), 5.98 – 5.90 (m, 2H), 5.52 (dd, *J* = 7.1, 3.5 Hz, 1H), 4.37 – 4.31 (m, 2H), 4.21 (dddd, *J* = 25.1, 11.7, 7.3, 4.3 Hz, 6H), 4.04 – 3.93 (m, 2H), 3.78 – 3.67 (m, 2H), 1.46 (d, *J* = 7.0 Hz, 3H).^{13}C NMR (101 MHz, Deuterium Oxide) δ 174.29, 166.35, 151.94, 141.82, 102.77, 94.64 (d, *J* = 6.1 Hz), 88.66, 83.34 (d, *J* = 9.1 Hz), 73.89, 72.21, 69.77, 68.54, 67.50, 65.11 (d, *J* = 4.5 Hz), 61.14, 58.38, 50.09 (d, *J* = 8.0 Hz), 16.61. ^{31}P NMR (162 MHz, Deuterium Oxide) δ -11.36 (d, *J* = 21.0 Hz), -13.09 (d, *J* = 21.2 Hz). ESI-HRMS calcd. for C$_{18}$H$_{27}$N$_6$O$_{17}$P$_2$ [M – 2 Na + H]$^-$ 661.0913, found 661.0893.

UDP-GalNAzEt(S) **3.13c**

The protocol described above was followed with different quantities of starting materials: UDP-GalN (**3.8**, 0.078 mmol), NHS ester (**3.12c**, approximately 0.62 mmol). Analog **3.13c** was afforded in 34.7% yield (27.0 μmol, 50% conversion by NMR). ^{1}H NMR (400 MHz, Deuterium Oxide) δ 7.98 (d, *J* = 8.1 Hz, 1H), 6.03 – 5.96 (m, 2H), 5.59 (dd, *J* = 6.9, 3.4 Hz, 1H), 4.42 – 4.36 (m, 2H), 4.36 – 4.18 (m, 5H), 4.15 (t, *J* = 6.7 Hz, 1H), 4.06 (d, *J* = 2.9 Hz, 1H), 4.00 (dd, *J* = 10.9, 3.1 Hz, 1H), 3.79 (h, *J* = 6.8 Hz, 2H), 1.95 – 1.78 (m, *J* = 6.9 Hz, 2H), 1.00 (t, *J* = 7.4 Hz, 3H). ^{13}C NMR (101 MHz, Deuterium Oxide) δ 173.57, 166.23, 151.81, 141.68, 102.62, 94.79 (d, *J* = 6.2 Hz), 88.50, 83.22 (d, *J* = 9.0 Hz), 73.79, 72.10, 69.66, 68.57, 67.48, 65.02 (d, *J* = 5.2 Hz), 64.00, 61.01, 49.88 (d, *J* = 8.5 Hz), 25.10. ^{31}P NMR (162 MHz, Deuterium Oxide) δ -11.16 (d, *J* = 21.0 Hz), -13.01 (d, *J* = 20.9 Hz). ESI-HRMS calcd. for C$_{19}$H$_{29}$N$_6$O$_{17}$P$_2$ [M – 2 Na + H]$^-$ 675.1070, found 675.1049.

UDP-GalNAzEt(R) **3.13d**

The protocol described above was followed with different quantities of starting materials: UDP-GalN (**3.8**, 0.084 mmol), NHS-ester (**3.12d**, 0.675 mmol). Analog **3.13d** was afforded in 23.8% yield (20.0 μmol, 61% conversion by NMR).^{1}H NMR (400 MHz, Deuterium Oxide) δ 7.94 (d, *J* = 8.1 Hz, 1H), 5.95 (d, *J* = 8.1 Hz, 2H), 5.53 (dd, *J* = 6.7, 3.4 Hz, 1H), 4.34 (t, *J* = 4.8 Hz, 2H), 4.28 – 4.24 (m, 2H), 4.23 – 4.16 (m, 3H), 4.04 (t, *J* = 6.6 Hz, 2H), 3.98 (dd, *J* = 10.9, 3.0 Hz, 1H), 3.74 (p, *J* = 6.4, 5.4 Hz, 2H), 1.83 (ddp, *J* = 28.4, 14.3, 6.9 Hz, 2H), 0.96 (t, *J* = 7.4 Hz, 3H). ^{13}C NMR (101 MHz, Deuterium Oxide) δ 173.46, 166.29, 151.85, 141.69, 102.65, 94.66 (d, *J* = 6.4 Hz), 88.50, 83.22 (d, *J* = 9.2 Hz), 73.77, 72.04, 69.64, 68.41, 67.37, 64.96 (d, *J* = 5.5 Hz), 64.14, 60.96, 49.96 (d, *J* = 8.8 Hz), 24.74, 9.33. ^{31}P NMR (162 MHz, Deuterium Oxide) δ -11.34 (d, *J* = 21.3 Hz), -13.20 (d, *J* = 21.1 Hz). ESI-HRMS calcd. for $C_{19}H_{29}N_6O_{17}P_2$ [M − 2 Na + H][-] 675.1070, found 675.1056.

UDP-GalNAzIpr(S) **3.13e**

UDP-GalN (**3.8**, 0.235 mmol, 1 equivalent) was dissolved in 15 mL water and NHS ester (**3.12e**, 0.47 mmol, 114.4 mg, 2 equivalents) in 5 mL THF was added. Reaction was adjusted to pH 7.5 – 8 with 1 M HCl. The reaction proceeded at rt for 1 h and the reaction was adjusted to pH 7.5 – 8 with 1 M KOH. The reaction proceeded at rt overnight. The mixture was evaporated to dryness and a crude NMR was used to approximate product conversion. The mixture was dissolved in water and extracted 2 x with EtOAc, the aqueous portion was evaporated to dryness, and the mixture was purified by size exclusion chromatography in water. The purity of the fractions was determined by TLC analysis visualized by UV and *p*-anisaldehyde staining with either 3:1 EtOH:10mM tributylammonium bicarbonate or 3:1 EtOH:0.5M NH4OAc as the mobile phase. The fractions were combined based on purity, frozen, and lyophilized to yield a white powder. **3.13e** was quantitated by absorbance measurements at 260 nm using an extinction coefficient of 9374 cm[-1]M[-1] as described for UDP-GalN. Some fractions were lost during purification, affording analog **3.13e** in 13.7% yield (32.3 μmol, 23% conversion by NMR). ^{1}H NMR (600 MHz, Deuterium Oxide) δ 8.02 (d, *J* = 7.8 Hz, 1H), 6.04 (d, *J* = 7.7 Hz, 2H), 5.66 – 5.61 (m, 1H), 4.44 (s, 2H), 4.41 – 4.25 (m, 5H), 4.11 (s, 1H), 4.04 (d, *J* = 10.9 Hz, 1H), 3.99 (d, *J* = 6.8 Hz, 1H), 3.86 – 3.79 (m, 2H), 2.20 (dq, *J* = 13.3, 6.7 Hz, 1H), 1.06 (dd, *J* = 12.6, 6.6 Hz, 6H). ^{31}P NMR (243 MHz, Deuterium Oxide) δ -11.21 (d, *J* = 21.0 Hz), -13.12 (d, *J* = 20.4 Hz). ESI-HRMS calcd. for $C_{20}H_{31}N_6O_{17}P_2$ [M − 2 Na + H][-] 689.1226, found 689.1221.

UDP-GalNAz-betaAla **3.13f**

The protocol described above was followed with different quantities of starting materials: UDP-GalN (**3.8**, 0.062 mmol), NHS-ester (**3.12f**, 0.50 mmol, 105.8 mg). Analog **3.13f** was afforded in 84.4% yield (52.3 μmol, 100% conversion by NMR). ^{1}H NMR (400 MHz, Deuterium Oxide) δ 7.95 (d, *J* = 8.1 Hz, 1H), 6.01 – 5.91 (m, 2H), 5.55 (dd, *J* = 6.7, 3.2 Hz, 1H), 4.35 (p, *J* = 5.3 Hz, 2H), 4.32 – 4.12 (m, 5H), 4.03 (d, *J* = 2.9 Hz, 1H), 3.96 (dd, *J* = 10.9, 3.1 Hz, 1H), 3.76 (h, *J* = 6.9 Hz, 2H), 3.60 (t, *J* = 6.4 Hz, 2H), 2.65 (td, *J* = 6.4, 4.1 Hz, 2H). ^{13}C NMR (101 MHz, Deuterium Oxide) δ 174.28, 166.19, 151.76, 141.65, 102.60, 94.70 (d, *J* = 6.3 Hz), 88.52, 83.16 (d, *J* = 9.0 Hz), 73.75, 72.07, 69.62, 68.41, 67.60, 64.99 (d, *J* = 5.0 Hz), 61.02, 49.74 (d, *J* = 8.1 Hz), 47.07, 34.84. [31]P

NMR (162 MHz, Deuterium Oxide) δ -11.21, -12.84. ESI-HRMS calcd. for $C_{18}H_{27}N_6O_{17}P_2$ [M – 2 Na + H]$^-$ 661.0913, found 661.0891.

<u>UDP-GalNAz-L3buytric</u> **3.13g**
The protocol described above was followed with different quantities of starting materials: UDP-GalN (**3.8**, 0.067 mmol), NHS-ester (**3.12g**, 0.54 mmol, 121.8 mg). Analog **3.13g** was afforded in 79.6% yield (53.3 μmol, 100% conversion by NMR). ^{1}H NMR (400 MHz, Deuterium Oxide) δ 7.97 – 7.89 (m, 1H), 5.94 (d, J = 7.2 Hz, 2H), 5.55 – 5.48 (m, 1H), 4.33 (s, 2H), 4.22 (dd, J = 26.3, 7.9 Hz, 5H), 4.01 (s, 1H), 3.95 (t, J = 10.8 Hz, 2H), 3.74 (d, J = 7.0 Hz, 2H), 2.53 (d, J = 6.5 Hz, 2H), 1.27 (d, J = 6.2 Hz, 3H). ^{13}C NMR (101 MHz, Deuterium Oxide) δ 173.92, 166.21, 151.76, 141.65, 102.60, 94.65 (d, J = 5.7 Hz), 88.47, 83.15 (d, J = 10.0 Hz), 73.74, 72.05, 69.61, 68.40, 67.49, 64.96 (d, J = 4.8 Hz), 61.00, 54.97, 49.74 (d, J = 7.6 Hz), 41.99, 18.59. ^{31}P NMR (162 MHz, Deuterium Oxide) δ -11.34 (d, J = 21.0 Hz), -12.99 (d, J = 21.0 Hz). ESI-HRMS calcd. for $C_{19}H_{29}N_6O_{17}P_2$ [M – 2 Na + H]$^-$ 675.1070, found 675.1051.

Chapter 4. Evaluating bump-hole pairs

4.1. Screening methodologies

Quantitation of glycosylation efficiency across a large number of enzyme-substrate pairs requires a robust screening methodology. The screen should be sensitive enough to detect low levels of glycosylation such that enzyme-substrate pairs with low activity can still be detected. In order to distinguish excellent enzyme-substrate pairs from good enzyme-substrate pairs, the screen should have a broad detection range over which signal saturation is not an issue. An ideal assay monitors the product of interest—the glycopeptide—directly, to most accurately reflect the formation of the desired product. Finally, the screen should have a high throughput to enable screening of all combinations of enzyme-substrate pairs with replicates in a discrete period of time.

Many biochemical screens for enzymatic activity are carried out *in vitro* to minimize sample complexity and to facilitate analysis. Early biochemical screens for glycosyltransferase activity were low throughput and used chromatography to monitor the formation of a desired product. Peptide glycosylation has been quantitated previously with the application of UDP-[^{14}C]-GalNAc to transferase reactions, which was then separated from [^{14}C]-labeled glycopeptides by high performance liquid chromatography (HPLC)[134,141,186]. If the product can be detected with antibodies or another affinity handle, a high-throughput Enzyme-Linked Immunosorbent Assay (ELISA) can be powerful. Phosphorylation is a common post-translational modification that has been studied in a variety of contexts using phosphospecific antibodies. Because there are not any antibodies appropriate for O-GalNAc glycosylation, an ELISA format that takes advantage of the azide-chemical handle for glycopeptide detection was developed[80]. Our results using the azido-ELISA are presented later in this chapter. However, the azido-ELISA is labor intensive and suffers from high variability across replicates, making is less desirable than traditional antibody-based sandwich ELISAs. Because many post-translational modifications are difficult to detect, classes of enzymes that modify proteins tend to experience challenges similar to glycosyltransferases. Recent studies of protein methyl transferase substrate specificity and enzymatic activity have utilized Matrix-Assisted Laser Desorption/Ionization (MALDI) mass spectrometry to screen reaction progress[109]. Although we evaluated MALDI for this purpose, we found it to have poor reproducibility, making it undesirable for a quantitation-based assay.

Screens for enzymatic activity can be carried out in the context of more complex biological samples if appropriate antibodies or other affinity tools for product detection are available. Whole proteins or lysates can be separated by SDS-PAGE and probed for the presence of a particular epitope by western blot. Due to the widespread availability and robustness of phosphospecific antibodies, it has been possible to identify many phosphorylated proteins in this fashion[167]. Because adequate antibodies to detect O-GalNAc glycosylation are not available, GalNAc

glycosylation can instead be monitored using chemical tags that allow the ligation of small molecules such as biotin that can be visualized with an antibody[47,80].

4.2. Screen enzyme-substrate pair with a model peptide substrate

4.2.1. UDP detection assay

Our initial efforts to confirm activity of purified ppGalNAcTs used a colorimetric assay that can be carried out in a 96-well plate format. The formation of the side product UDP is monitored in this coupled enzyme assay. This assay relies on a nucleotidease that breaks UDP to UMP and inorganic phosphate. The inorganic phosphate is detected with the small molecule malachite green. We used a commercially available kit from Pierce that has been evaluated for several glycosyltransferase targets[187]. We evaluated the activity of the purified soluble domains of wildtype and mutant ppGalNAcT2 using UDP-GalNAc as the donor sugar and biotinylated EA2 peptide (PTTDSTTPAPTTK) as the acceptor.
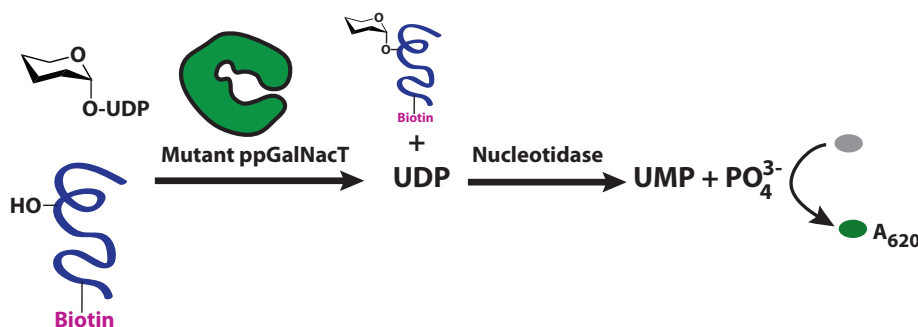
Figure 4-1. Malachite green assay to detect enzyme activity with UDP-GalNAc.

In this assay, we used a sub-saturating concentration of EA2 (50 μM) and we titrated the UDP sugar from 7.8-500 μM, spanning a range of concentrations around K_M reported for the wildtype enzyme (25 μM)[80]. The reaction was carried out for two hours such that approximately 10-20% product formation occurred. Relative to the wildtype enzyme, the I253A and L310A mutants were less active, although they still achieved good substrate turnover. In contrast the F361 mutants and the I253A/L310A double mutant performed comparably to the mock-transfected enzyme sample.
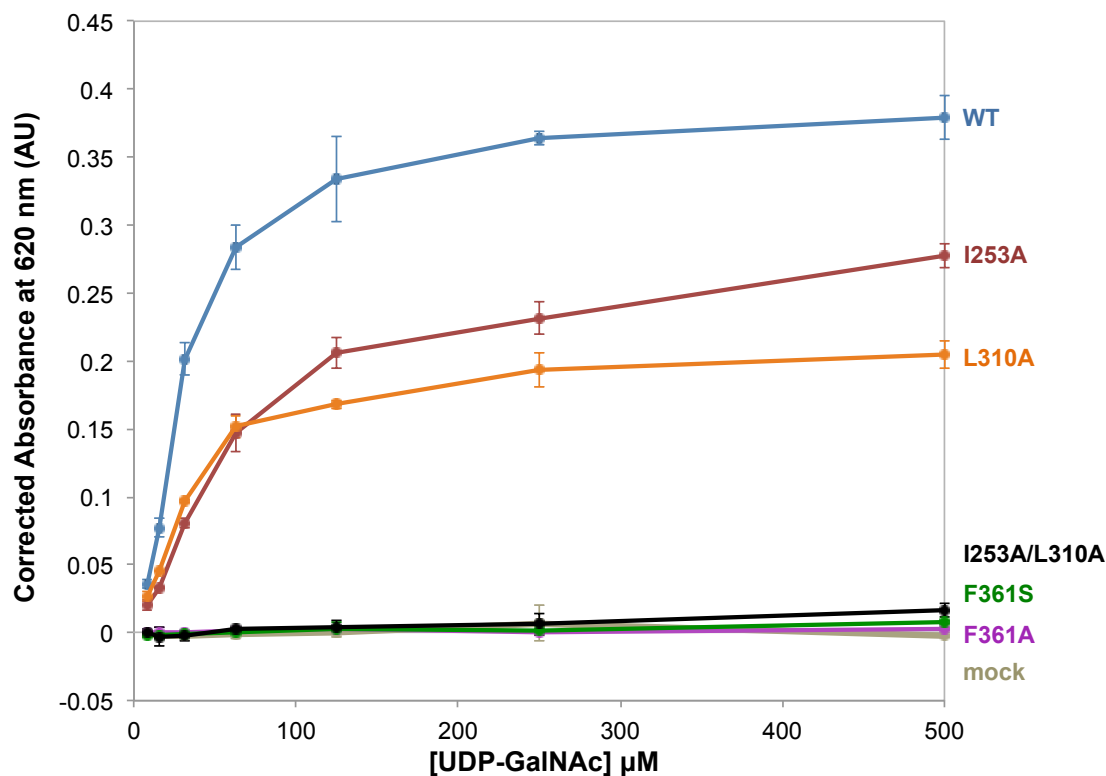
Figure 4-2. I253A and L310A show reduced activity with UDP-GalNAc compared to the wildtype enzyme. Reactions contained 50 μM EA2-biotin (5 mM stock in PBS), a two-fold dilution series of UDP-GalNAc from 500 μM – 37.81 μM (5 mM stock diluted in water), and 25 μL purified ppGalNAcT2. Reaction proceeded 2 h at 37 °C.

These results suggest that differential ppGalNAcT activity is due changes in the active site residues, and that construct design, expression, and purification result in active enzymes. However, our efforts to continue these experiments with UDP-GalNAc analogs were unsuccessful. When we repeated these experiments with UDP-GalNAc analogs, the signal increased relatively linearly independent of enzyme sample and with a dependence on UDP-GalNAc analog concentration. This suggested an unknown interfering substance might be in the samples of UDP-GalNAc analogs tested. The UDP-GalNAc analogs could be contaminated with trace amounts of UDP or inorganic phosphate that were not detected by NMR. UDP was not observed by ^{1}H or ^{31}P NMR, and inorganic phosphate was not detected by ^{31}P NMR. Experiments without nucleotidease demonstrated some signal increase that was dependent on UDP-sugar concentration, implying that UDP alone cannot be the contaminant. Finally, it is possible that the nucleotidease-dependent portion of signal increase could demonstrate increased lability of the UDP-GalNAc analogs relative to UDP-GalNAc. Based on the known routes of UDP-sugar hydrolysis, there is no reason to believe that the modifications to the N-acyl side chain would make the UDP-sugar more unstable[188-192]. Finally, we have never confirmed the presence of any contaminants that explain this background signal, nor has there been any evidence of decreased stability of the UDP-GalNAc analogs in our other assays. The malachite green was a useful assay for determining successful enzyme preparation

in a relatively high throughput fashion, but it was not used further as a screening methodology for a bump hole pair.

4.2.2. Azido ELISA

We next sought to evaluate azido-ELISA as a technique for high throughput screening of enzyme substrate pairs. In the azido-ELISA, glycosylation of the biotinylated EA2 is detected based on the presence of an azidosugar (Figure 4-3)[80]. Bertozzi and coworkers have previously used this assay to screen ppGalNAcT isoforms for activity with UDP-GalNAz, so it should be equally efficacious when screening T2 mutants and other UDP-GalNAc analogs[6,80]. Our initial work replicated published azido-ELISA assays with UDP-GalNAz and purified wildtype hT2 provided by the Tabak lab (data not shown)[6,80]. For our bump-hole screening, as described in Chapter 2, we generated soluble wildtype and mutant ppGalNAcT2, which was purified by immunoprecipitation.
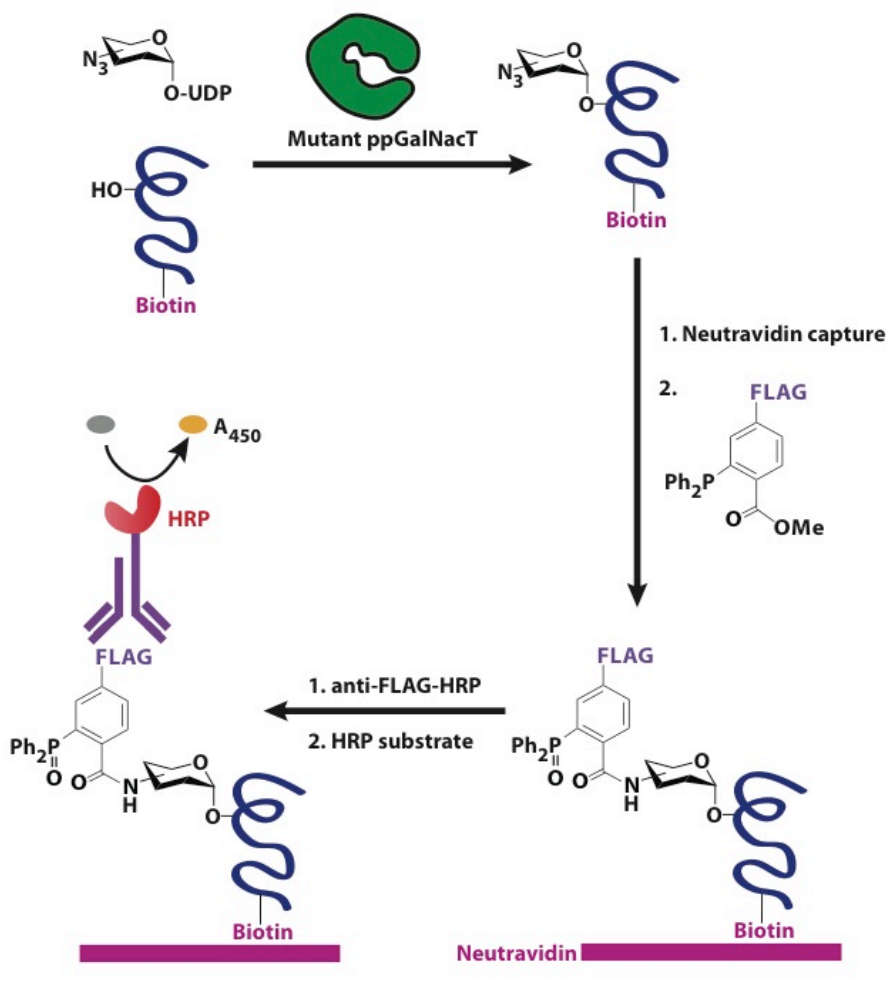
Figure 4-3. Azido-ELISA to screen ppGalNAcT2 mutants and UDP-GalNAc analogs. The azidosugar and phosphine probe undergo the Staudinger Ligation, enabling subsequent detection with α-FLAG-HRP.

We did not evaluate UDP-GalNAc in this assay because an azide handle is required to achieve signal. We used UDP-GalNAz as the first substrate because there are already parameters using this assay that have been previously published[80]. Bump-hole activity was monitored under sub-saturating concentration of EA2 (50 µM), and we titrated the UDP sugar from 3.9-250 µM, spanning a range of concentrations around K_M reported for the wildtype enzyme (25 µM)[80].

In contrast to the results with UDP-GalNAc in the malachite green assay, when UDP-GalNAz was the substrate, the activities of the wildtype, I253A mutant, and L310A mutant were comparable. This suggested a decrease in wildtype activity relative to the activity of these two mutants. Most interestingly, we saw that with UDP-GalNAz, the double mutant recovered some activity. This suggested that the double mutant was already demonstrating a property of partial orthogonality, given that it was inactive with the native substrate UDP-GalNAc but was more active with the enlarged substrate UDP-GalNAz. However, because UDP-GalNAz itself is not an orthogonal substrate to the wildtype ppGalNAcTs, we need to identify other substrates with which the double mutant is active. At this point we decided that in spite of our efforts to optimize the azido-ELISA, the large variation across replicates made it difficult to compare the relative activities of bump hole pairs.
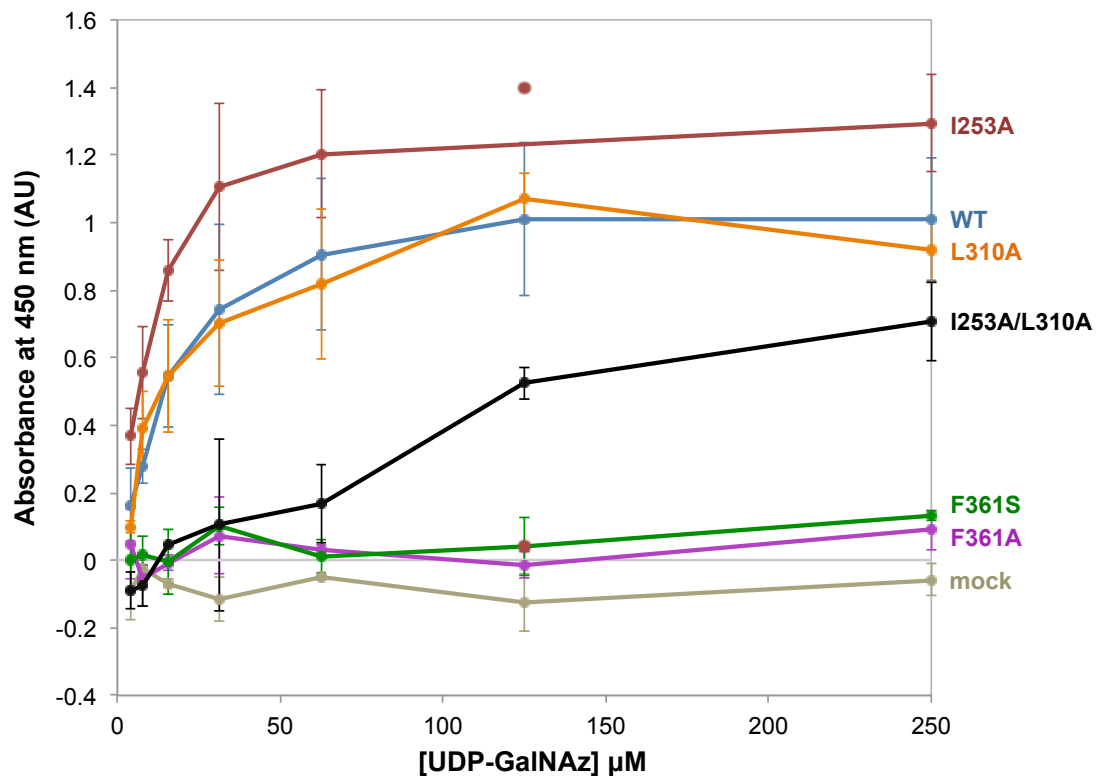
Figure 4-4. UDP-GalNAz activity with hT2 mutants. Activity of I253A and L310A is similar to that of the WT enzyme. As compared to activity with UDP-GalNAc, I253A/L310A activity is partially rescued by UDP-GalNAz. Glycosylation reactions contained 50 µM EA2-biotin, a two-fold dilution series of UDP-GalNAz from 250 µM – 3.91 µM, and 2 µL purified ppGalNAcT. Reactions proceeded for 1 h 37 °C.

The major challenge of using the azido-ELISA in a screening context is to achieve a signal to noise ratio with which differences in enzyme activity can be detected. Although the assay is currently colorimetric (abs = 450 nm), using a fluorescent or luminescent probe could improve the assay's sensitivity and dynamic range[193]. One possible modification is to use an α-FLAG antibody conjugated to alkaline phosphatase with the commercial substrate 4-methylumbelliferyl phosphate. Upon hydrolysis to 4-methylumbelliferone, detection is achieved via excitation at 360 nm and measurement of emission at 440 nm[193].

4.2.3. LS-MS assay

Ultimately, we chose to use the moderate throughput achievable from a LC-MS assay. Expected product formation on the order of picomoles to one nanomole necessitated the use of a high sensitivity instrument. Using an LTQ mass spectrometer equipped with an autosampler and nanoflow chromatography, we developed an assay with excellent sensitivity that allows us to compare levels of starting material relative to product. Initial experiments were carried out using an Orbitrap mass spectrometer by Dr. Anthony Iavarone.

Consistent with results observed with the UDP-detection assay and the azido-ELISA, we found that the F361A and F361S mutants did not demonstrate activity. This assay was carried out with purified samples from mock-transfected cells, and no glycopeptide was observed from these samples or from the F361 mutant samples. We also saw minimal activity with compound **3.13e**, UDP-GalNAzIpr(S), which only produced detectable amounts of glycopeptide when used at saturating concentrations with the double mutant (Figure 4-5).
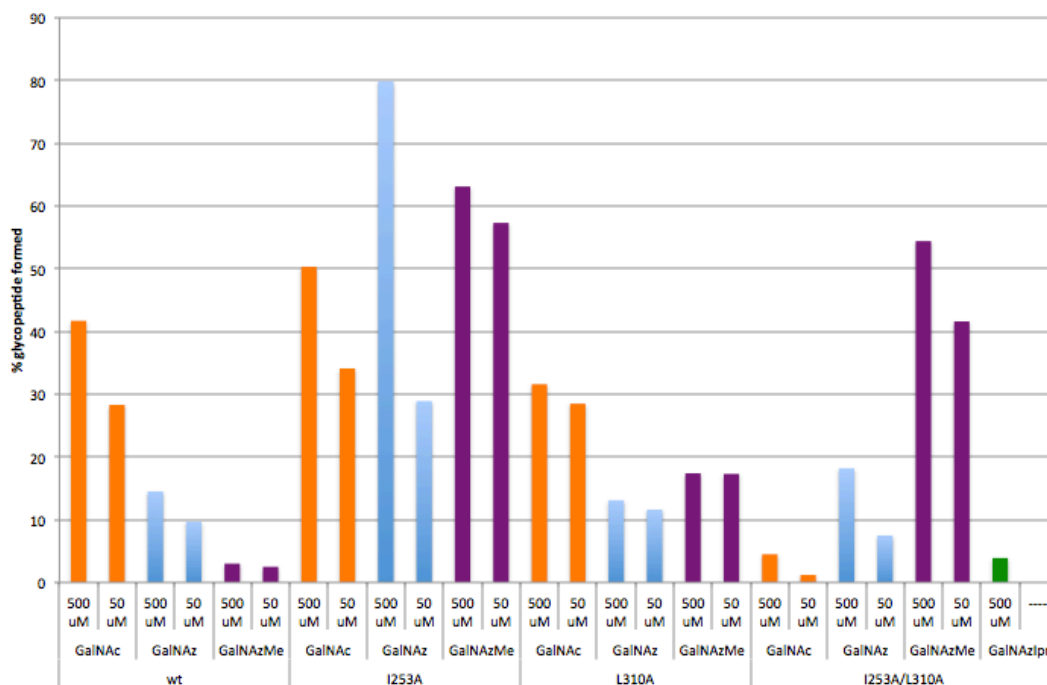
Figure 4-5. Mass spectrometry analysis of enzyme-substrate activity. WT and mutant enzymes I253A, L310A, and I253A/L310A were screened against UDP-GalNAc, GalNAz, GalNAzMe(S), and GalNAzIpr(S). Reactions (in 25 mM Tris pH 7.5, 10 mM $CaCl_2$, 10 mM $MnCl_2$, 50 μL final reaction volume) contained ppGalNAcT2 (25 μL enzyme), saturating UDP-sugar (500 μM), and EA2 acceptor peptide (50 μM). Glycosylation proceeded for 2 h, reactions were quenched with 50 μL 2% formic acid, and % glycopeptide was analyzed by LC-MS with an Orbitrap mass spectrometer. Data were not normalized for enzyme concentration, so WT and mutant enzymatic activity cannot be directly compared.

As shown in Figure 4-5, the wildtype enzyme displays a very promising trend of decreased activity as the size of the UDP-sugar increases. Glycopeptide formation decreases by approximately 3-fold when comparing UDP-GalNAc to UDP-GalNAz. In agreement with this observation, it was previously found that the ppGalNAcTs have about 1/3 the activity with UDP-GalNAz that they do with UDP-GalNAc[80]. We also observe an approximately 10-fold decrease in product formation when comparing UDP-GalNAc and compound **3.13a**, UDP-GalNAzMe(S). This trend is indicative of increasing substrate orthogonality, and may indicate that UDP-GalNAzMe(S) is somewhat orthogonal to the native ppGalNAcTs. In this experiment, the concentration of WT enzyme is less than the concentration of the other 3 enzymes, so its activity cannot be compared directly to the mutants.

The double mutant enzyme displays the opposite reactivity trend compared to the wildtype enzyme. As the size of the UDP-sugar increases, the percent glycosylation increases (Figure 4-5). This is the trend previously observed when comparing the UDP-detection assay using UDP-GalNAc, in which the double mutant was inactive (Figure 4-2), and the azido-ELISA, in which double mutant activity was partially rescued by the enlarged substrate (Figure 4-4). Approximately a 10-fold increase in product formation occurs as the substrate goes from UDP-GalNAc to UDP-

89

GalNAzMe(S). However, once the branched analog is too large, activity drops precipitously, as seen with UDP-GalNAzIpr(S).

The I253A and L310A mutants are both active with UDP-GalNAc and UDP-GalNAz as observed with previous assays. It also appears that the I253A mutant is more active than the L310A mutant, which can also be seen with the UDP-detection assay and the azido-ELISA (Figure 4-2, Figure 4-4). Interestingly, both of these mutants appear to use UDP-GalNAzMe(S) similarly to UDP-GalNAz and UDP-GalNAc. These findings are not a demonstration of enzyme orthogonality because it is likely that both the I253A and L310A mutants will utilize UDP-GalNAc *in vivo*, as well as the UDP-GalNAzMe(S), if the sugars are at similar concentrations. Such a characteristic was exploited by Shokat and colleagues with the majority of analog-sensitized kinase alleles they have studied. When the engineered enzyme is not orthogonal to the native system, it is capable of maintaining wildtype activity in the absence of the analog. This can be particularly useful when evaluating the replacement of the native allele with the mutant allele in an animal model. Both of these single mutants might have good kinetic parameters with UDP-GalNAzMe(S).
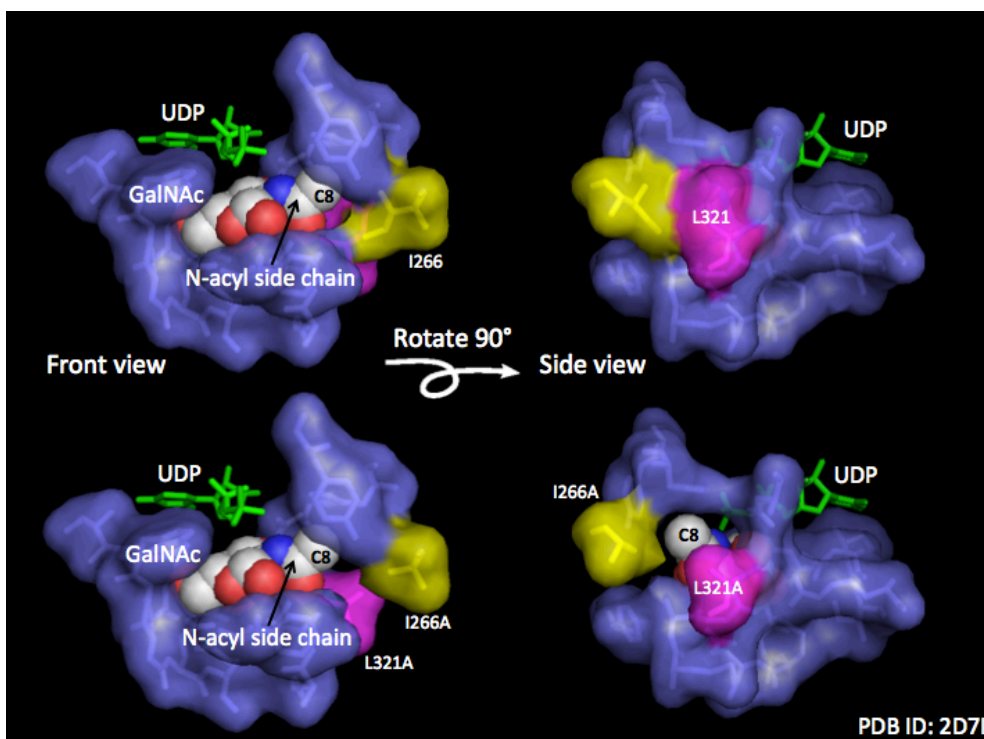
Figure 4-6. Modeling of WT and double mutant in hT10. PDB ID 2D7I.

We modeled space-filling interactions within 6 angstroms of GalNAc in the crystal structure of ppGalNAcT10. To visualize the changes that occur near the N-acyl side chain of GalNAc when bound in the active site of the double mutant ppGalNAcT, we used PyMOL mutagenesis to replace the side chains of I266 and L321 with alanine (Figure 4-6). We observed that two distinct pockets arise in the double mutant, which may be accessible to alpha-branched analogs such as UDP-GalNAzMe(S).

The remainder of our analogs have yet to be screened against all of the wildtype and mutant ppGalNAcTs. It will also be interesting to evaluate the behavior of UDP-GalNAc analog diastereomers derived from enantiomeric starting materials in order to understand the relative importance of different configurations of the stereocenter at the alpha position of the bump. Our panel of ppGalNAcT2 mutants will allow us to evaluate the most active mutant-substrate pairs. The wildtype constructs of T2, T1, T7, and T10 will allow us to evaluate the orthogonality of the UDP-GalNAc analogs with several native ppGalNAcT family members. The T1, T7, and T10 mutants will allow us to explore how well active site engineering to generate a bump-hole pair can translate into other members of the ppGalNAcT family.

ppGalNAcT secretion construct	Amino acid position		
hT2	**253**	**310**	**361**
WT	I	L	F
I253A	A	L	F
L310A	I	A	F
F361A	I	L	A
F361S	I	L	S
I253A/L310A	A	A	F
hT1	**238**	**295**	**346**
WT	I	L	F
I238A	A	L	F
L295A	I	A	F
F346A	I	L	A
F346S	I	L	S
I238A/L295A	A	A	F
hT7	**330**	**391**	**442**
WT	I	L	Y
I330A/L391A	A	A	Y
hT10	**266**	**321**	**372**
WT	I	L	Y
I266A/L321A	A	A	Y

Figure 4-7. Panel of ppGalNAcT secretion constructs (left) and UDP-GalNAc analogs (right).

Based on our initial results using LC-MS to quantify peptide glycosylation, we will screen all of the wildtype and engineered ppGalNAcTs against the panel of UDP-sugars. We will modify the conditions of the fixed time LC-MS assay described in Figure 4-5 such that full peptide glycosylation is obtained when using the wildtype ppGalNAcT in combination with the natural UDP-GalNAc substrate. Thus, all enzyme-substrate pairs will be screened under identical conditions, and based on the selectivity of the enzyme for the UDP-GalNAc analog compared to UDP-GalNAc, we will identify the most promising bump-hole pair.

4.3. Screen bump-hole pair for desired biological behaviors

As discussed in Chapter 1, one of the central challenges in engineering a bump-hole pair is recapitulating native activity. An engineered ppGalNAcT should retain both catalytic activity and specificity for its polypeptide and protein substrates. Towards this goal, bump-hole pairs were screened for activity together as described in the previous section. In this section, we describe efforts to demonstrate that the bump-hole pair has the desired behavior in a biological context. Application of the bump-hole pair in *vivo* is limited by the delivery mechanism of the sugar analog.

Even with the range of currently available delivery techniques, charged nucleotides are not cell permeable, making it is challenging to deliver a measurable quantity of the UDP-sugar to cultured cells[49,194]. Sugar analogs can be delivered to cells as UDP-sugar precursors, which are then converted to the nucleotide sugar by cellular biosynthetic enzymes. Therefore, it could be useful to visualize whether cells treated with peracetylated sugar analog result in cell-surface labeling. However, we did not believe that the results of this experiment would be meaningful until we had developed a method to detect the formation of UDP-sugar. Otherwise, any experiment with absence of signal could not be evaluated for problems with analog delivery versus ppGalNAcT orthogonality.

Because the ppGalNAcTs are Golgi resident enzymes, application of the bump hole pair to cell lysate as has been successful with kinases is unlikely to yield meaningful labeled products[15]. However, when we began the validation experiments using cellular components, we had not yet achieved a delivery method for the UDP-GalNAc analogs. Therefore, our first experiments used fractionated cell lysates to study the behavior of the first candidate bump-hole pair.

4.3.1. Confirm UDP-GalNAc analog is orthogonal to native ppGalNAcTs

With a bump-hole pair in hand, we set about demonstrating that the UDP-GalNAc analog was orthogonal to other ppGalNAcTs. This can be accomplished by screening the substrate analog against a panel of wildtype ppGalNAcT family members *in vitro*. We currently have secretion constructs of human T2, T1, T7, and T10, and we plan to use them for *in vitro* biochemical screens. However, assembling a panel comprised of all known ppGalNAcTs is prohibitive. In the future, a collaboration with the Clausen, Tabak, Ten Hagen, or Gerken labs or plasmids from the Repository of Glyco-enzyme Expression Constructs with the Complex Carbohydrate Research Center (Athens, GA) might enable screening of the analog against existing protein libraries.

Alternatively, a more accessible metric is background labeling in cells or animals treated with the analog. Most cell types express only some of the ppGalNAcT family members, so evaluating orthogonality in this fashion must be matched to the model system of study. Note that screens against cells and animals evaluate whether any cellular component utilizes the analog, not only the ppGalNAcTs. The work

presented in this dissertation is focused on developing a model system in human embryonic kidney (HEK) cells, so we first monitored background labeling in enzymatically active HEK lysates.

We first focused our efforts on fractionated cellular lysates treated exogenously with purified ppGalNAcTs and UDP-GalNAc analogs. Fractionated lysates were generated from HEK cells using a proprietary kit. Initial cell permeabilization generated a cytoplasmically enriched fraction. Subsequent membrane solubilization resulted in an ER, Golgi, and plasma membrane enriched fraction. A final step released soluble nuclear components. We evaluated the quality of the fractionation using markers of the cytoplasm (anti-HSP90), plasma membrane (anti-Na/K ATPase), and nucleus (anti-nuclear pore complex) (data not shown).
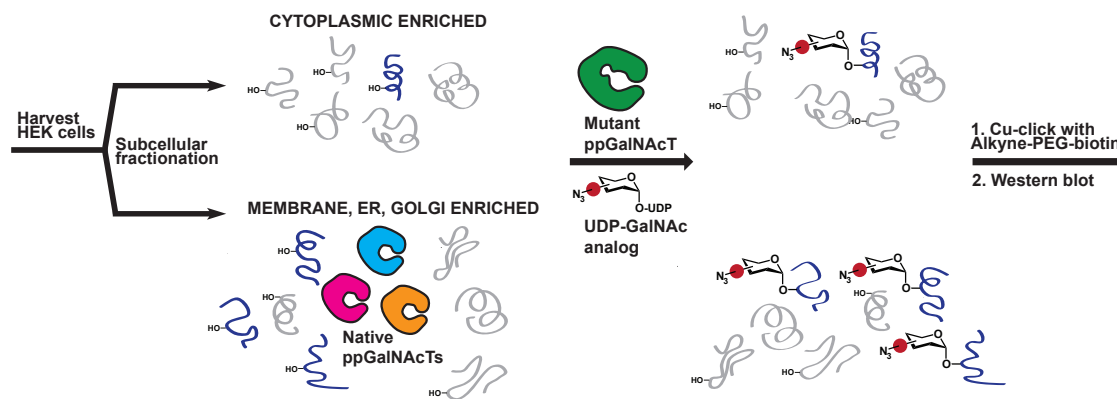
Figure 4-8. Fractionated HEK293T lysates treated with exogenous purified hT2 and UDP-GalNAc analogs.

As shown in Figure 4-8, ppGalNAcTs should be present in the membrane-enriched fraction, but not the cytoplasmic fraction. We found that both fractions contain proteins that can be glycosylated by the ppGalNAcTs, as shown Figure 4-9. However, natural ppGalNAcTs substrates should be in the secretory pathway, and therefore in the membrane fraction (Figure 4-8). These results demonstrate that, at least in the context of a lysate, there are proteins with residues that can be glycosylated when there is an excess of enzyme and donor substrate that are not natural substrates of the ppGalNAcTs. This observation has similarities to published work showing that when the ppGalNAcTs are trafficked to the ER, which is believed to occur only with certain stimuli, additional sites of glycosylation and truncated glycoforms appear[54]. This finding suggests that organelle localization is an important part of substrate selection.

As UDP-GalNAz is not orthogonal to the ppGalNAcTs, we would predict labeling with UDP-GalNAz in the membrane fraction (Figure 4-9). However, in the absence of the ppGalNAcTs, there should not be any labeling with UDP-GalNAz (Figure 4-10). In contrast, UDP-GalNAzMe(S) shows minimal labeling in both conditions, suggesting good orthogonality (Figure 4-10). Ideally, there is little detectable activity of the UDP-GalNAc analog with any native ppGalNAcTs. Although these initial studies in a

cellular context are promising, we continue to evaluate side reactivity between the analog and native ppGalNAcTs. If we encounter signal due to native ppGalNAcTs, we will evaluate the significance of this activity and may return to find another candidate bump-hole pair.
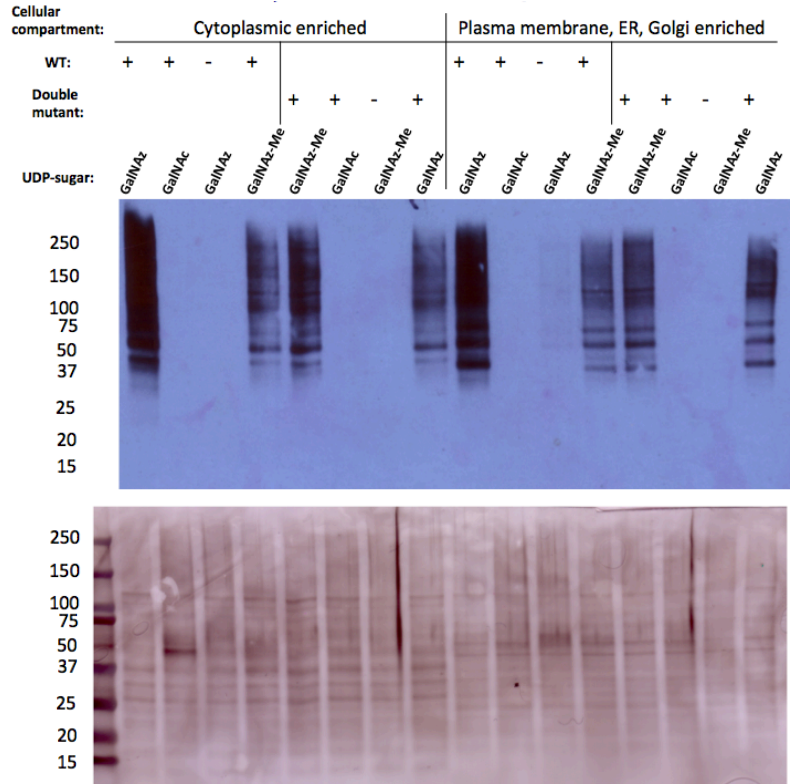
Figure 4-9. Lysates labeled with UDP-GalNAc analogs and purified hT2. Reactions (in TBS, 10 mM MnCl$_2$, 50 μL total) containing 1.33 μg/μL lysate were treated with ppGalNAcT (21 μL enzyme) and saturating UDP-sugar concentrations (250 μM) and were carried out at 37 °C overnight. Lysates were treated with alkyne-PEG-biotin and subjected to CUAAC conditions for 1 h. (A) Western blot of azide-labeled proteins. Samples were subjected to SDS-PAGE (4.9 μg protein) and Western blot and biotin was detected with anti-biotin-HRP (1:100,000). (B) India ink total protein stain.

94
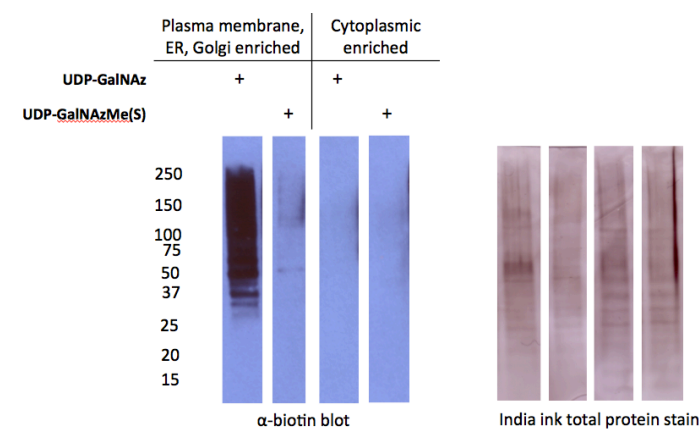
Figure 4-10. Excerpt from Figure 4-9 with longer film exposure: Labeling arising from native enzymes in enzymatically active lysates. (A) anti-biotin Western blot. (B) India ink total protein stain. Lanes shown in the order 11, 15, 3, 7.

4.3.2. Competition with UDP-GalNAc

To evaluate the ability of UDP-GalNAzMe(S) to compete with UDP-GalNAc in a cellular context, we evaluated labeling intensity of lysates doped with exogenous ppGalNAcT2 and both UDP-sugars. In one treatment, WT enzyme was added, and in the other, I253A/L310A double mutant enzyme was added. Both samples were treated with 250 μM UDP-GalNAzMe(S), which should be a saturating concentration of the sugar donor. Additionally, increasing amounts of UDP-GalNAc were titrated in 2-fold dilutions.
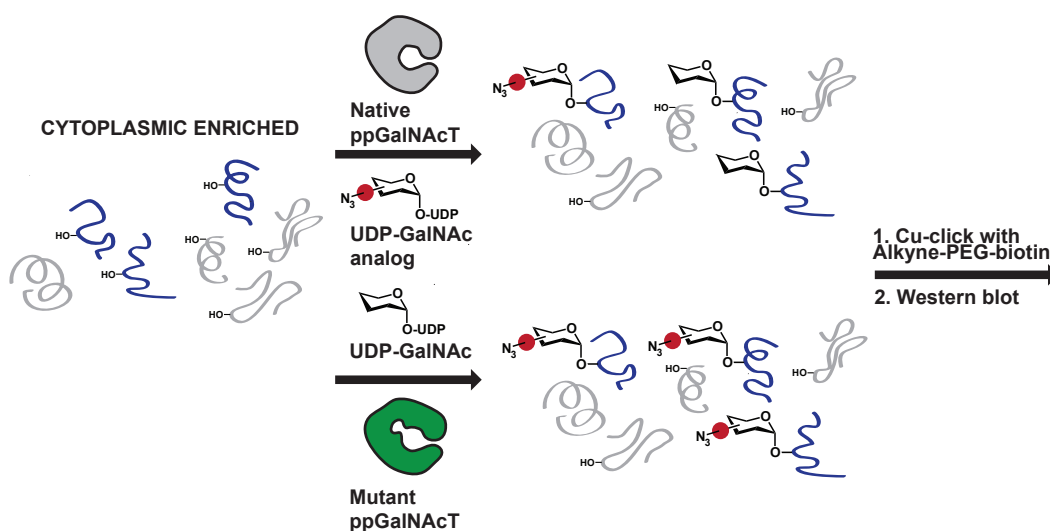
Figure 4-11. UDP-GalNAc competition in cytoplasmically-enriched HEK293T lysates. Lysates were treated with exogenous purified hT2 and UDP-GalNAzMe(S).

We saw that the addition of UDP-GalNAc did not abrogate signal generated by the double mutant enzyme. In contrast, the wildtype enzyme underwent significant reductions in signal as the amount of UDP-GalNAc increased. When the UDP-GalNAc concentration was equal to the UDP-GalNAzMe(S) concentration, the majority of the signal was depleted. When twice the concentration of UDP-GalNAc was used relative to UDP-GalNAzMe(S), the signal was virtually eliminated. These findings are very promising, indicating that the wildtype enzyme should show little labeling from the analog in the presence of UDP-GalNAc, while the double mutant can label its substrates, despite the presence of UDP-GalNAc. These are the conditions that we anticipate in cells.
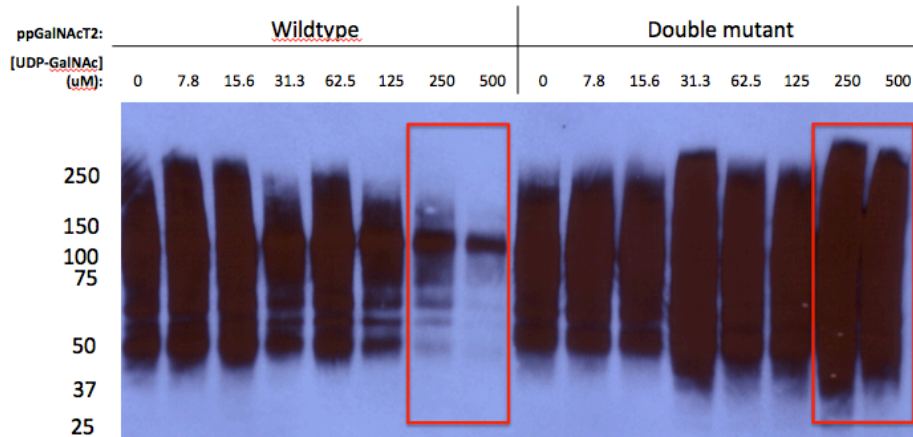
Figure 4-12. Western blot of UDP-GalNAzMe(S)-labeled proteins in UDP-GalNAc competition. Reactions (in TBS, 10 mM $MnCl_2$, 50 μL total) containing 1.33 μg/μL lysate were treated with ppGalNAcT (20 μL enzyme) and were carried out at 37 °C for 13 h. Reactions contained saturating UDP-GalNAzMe(S) concentrations (250 μM) and [UDP-GalNAc] was titrated from 0-500 μM. Lysates were treated with alkyne-PEG-biotin and subjected to CUAAC conditions for 1 h. (A) Western blot of azide-labeled proteins. Samples were subjected to SDS-PAGE (3.3 μg protein) and Western blot, and biotin was detected with anti-biotin-HRP (1:100,000). Protein concentrations are not identical due to some protein precipitation in lysates (not shown).

4.3.3. Evaluate bump-hole pair *in cellulo* / full-length ppGalNAcT constructs in cells

After treating cells with exogenous ppGalNAcT protein, the next step is to transfect cells with the ppGalNAcT of interest (either the wildtype or the double mutant), and to evaluate its activity in lysates or intact cells. Therefore, we next focused our efforts on transfecting cells with full length ppGalNAcTs and confirming their proper cellular localization. If full-length ppGalNAcT constructs are trafficked to the Golgi apparatus, they should function like the native cellular ppGalNAcTs, either in living cells or in lysates. Transfected ppGalNAcTs localized to the Golgi should utilize UDP-GalNAc analogs to modify proteins in the secretory pathway. We can then evaluate the labeling that can be achieved with overexpressed enzymes. Based on the results with the UDP-GalNAc competition described in the previous section, we hypothesize that the natural levels of UDP-GalNAc will minimize labeling with the sugar analog by wildtype enzymes. Such a result would indicate orthogonality of the analog within the biological context of a living cell. Due to the relative activities of the double mutant with these two sugars, we would expect labeling from the analog by the double mutant enzyme. We generated full-length hT7 and T10 prior to cloning full-length hT2, so we evaluated the behavior of these constructs in cells.

We transfected HEK cells with full-length constructs of ppGalNAcT7 and T10 in the pFLAG-CMV-2 vector. These constructs had an N-terminal FLAG tag followed by the entire coding sequence of the ppGalNAcT. Cells were transfected, the media was changed after 24 hours, and the cells were analyzed at 48 hours post-transfection.

To evaluate proper localization of these constructs in cells, we fractionated the cells as described previously. As shown in Figure 4-13, fractionated lysates included 1) a cytoplasmically enriched fraction, 2) an ER, Golgi, and plasma membrane enriched fraction, and 3) a fraction enriched with the soluble nuclear components. We evaluated the quality of the fractionation with markers of the cytoplasm (anti-HSP90), plasma membrane (anti-Na/K ATPase), and nucleus (anti-nuclear pore complex).
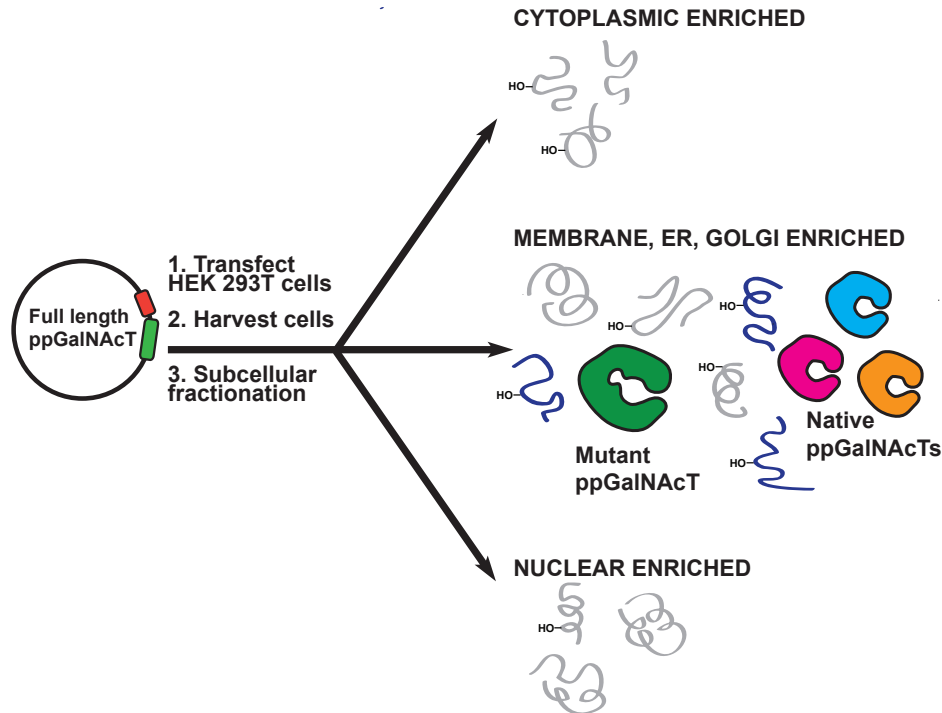

Figure 4-13. HEK293T cells transfected with full-length hT2 constructs and fractionated.

We found that fractions were enriched in the desired protein markers, but that there was some contamination across fractions. Despite our efforts to optimize the manufacturer's protocol, we were not able to improve the separation between fractions. We found that cytoplasmic marker HSP90 (heat shock protein MW 90 kDa) was most enriched in the cytoplasmic fraction and was also found in the membrane fraction. Nuclear pore complex proteins were most concentrated in the nuclear fraction, but they were also seen in the cytoplasmic fraction, and some cross-reactivity was also seen in the membrane fraction. Because our goal was to use the fractionation to evaluate ppGalNAcT localization, we moved forward at this stage.

Anti-FLAG signal corresponding to transfected ppGalNAcT7 was seen in the membrane fraction with some contamination of the nuclear fraction (Figure 4-14). As shown in Figure 4-14, a wash step following the isolation of the membrane fraction to remove any membrane contaminants did not significantly reduce this contamination. Thus, we hypothesized that contamination of the nuclear fraction

was due to incomplete membrane solubilization, rather than incomplete removal of solubilized membrane proteins. These results suggested that wildtype and double mutant hT7 was properly localized to the Golgi apparatus. In contrast, in multiple experiments, hT10 failed to localize as expected and was only observed in the boiled cell pellet (data not shown). These data are consistent with the protein being insoluble, and we hypothesized that this was due to improper trafficking of the full-length construct due to the N-terminal FLAG tag, as described in Chapter 2. To further validate these results, we performed immunocytochemistry to visualize the cellular localization of the FLAG-tagged ppGalNAcTs.
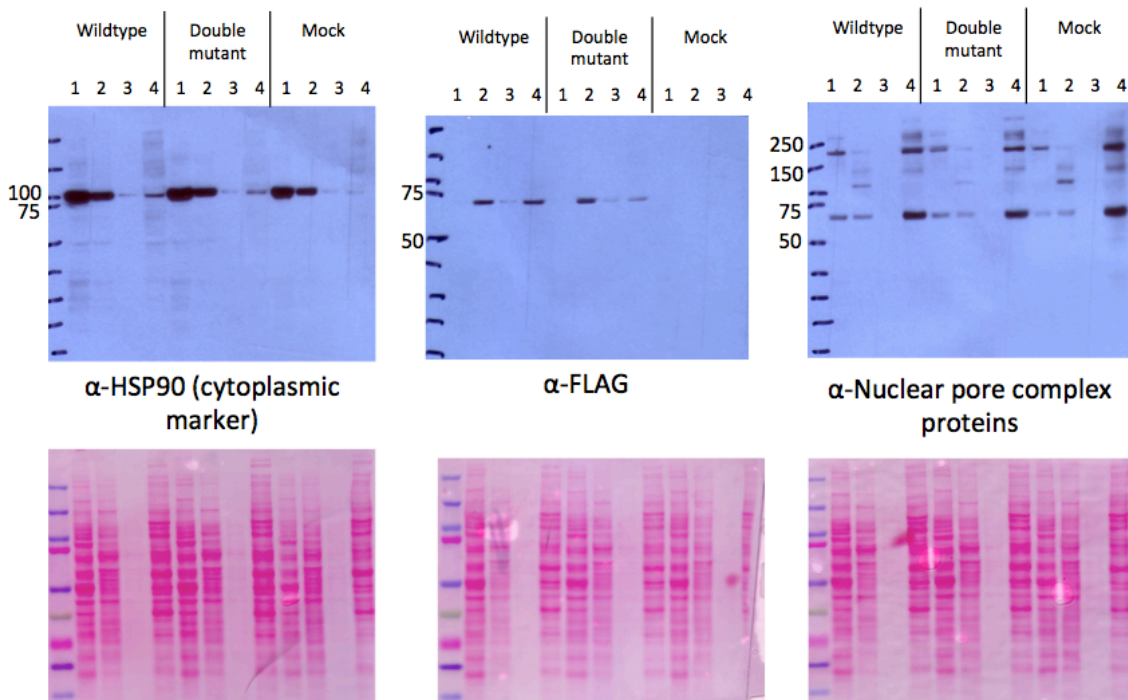
Figure 4-14. Fractionated lysates are appropriately enriched for subcellular markers. (Upper panel) Western blots of fractionated lysates (8.5 µg protein, except wash lane) probed with antibodies against known subcellular markers: lane 1) cytoplasmic extract; lane 2) membrane, ER, Golgi extract; lane 3) wash; lane 4) soluble nuclear extract. Predicted MW of hT7 = 75 kDa. (Lower panel) Ponceau total protein staining.

Cells were transfected as described for the experiments above and evaluated by immunocytochemistry 48 h post-transfection. After fixing and permeabilizing cells, cells were treated with anti-FLAG antibody and stained with DAPI to visualize nuclei. Although we planned to compare co-localization of the anti-FLAG signal to a Golgi marker, signal from the Golgi marker was too low to make this possible. Because these experiments were preliminary studies of ppGalNAcT localization, we did not optimize the immunocytochemistry or Golgi labeling. We found that, as seen with the cellular fractionation, hT7 appeared to localize to a subcellular compartment consistent with the Golgi apparatus. Although these images suggest a difference in subcellular localization between the wildtype and double mutant constructs, we found that localization was qualitatively the same, particularly when

we compared them across multiple z-planes. In contrast, hT10 was not localized to any apparent cellular compartment, and signal was observed diffusely throughout the cell (Figure 4-15). Rather than evaluate these data further, as described in Chapter 2, we re-cloned all full-length T7 and T10 constructs to transfer the FLAG tag to the C-terminus of the protein. We are repeating the experiments described in this section with the new constructs.
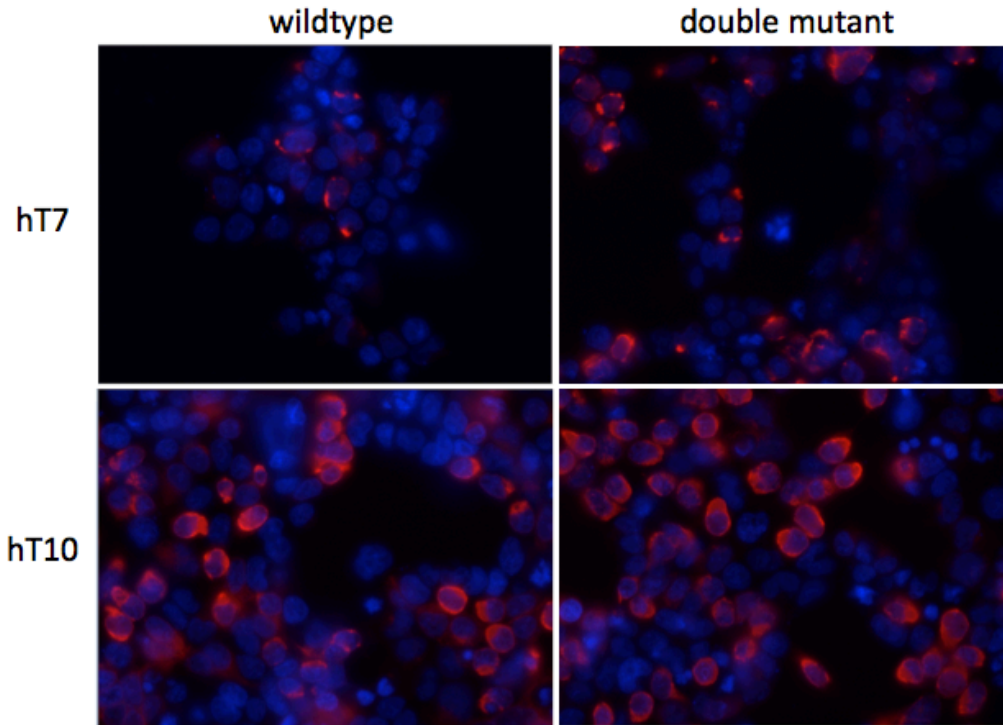
Figure 4-15. Localization of full-length hT2 constructs in HEK293T cells. AF647 (red) labels FLAG epitope tag and Hoechst dye (blue) stains nuclei.

4.4. Delivery

To achieve labeling with a sugar analog in cells, the analog has to pass from outside the cell to the interior of the cell. However, polar and charged molecules cannot passively diffuse through the cell membrane. Sugar analogs such as GalNAz are commonly delivered to the cell by protecting the polar hydroxyl groups of the sugar with relatively hydrophobic acetyls, thereby allowing diffusion across the cell membrane. Once inside the cell, non-specific esterases can deprotect the sugar, allowing it to enter the GalNAc salvage pathway. First, galactokinase glycosylates the anomeric position of the sugar. Subsequently, AGX1 converts the sugar-1-phosphate into UDP-GalNAz. UDP-GalNAz can be epimerized to UDP-GlcNAz by GALE, the C4-epimerase that interconverts these two nucleotide sugars. These steps are possible because GalNAz is not orthogonal to the enzymes in the GalNAc salvage pathway. However, when designing a bumped analog, the goal of the analog is to be modified in such a way that it is not easily accepted by native enzymes. Although our bumped analog may be orthogonal to the ppGalNAcT family, a likely side-effect is

that it may be orthogonal to some or all of the enzymes within the GalNAc salvage pathway. We anticipated that it would be necessary to develop an alternative strategy to deliver the bumped analog to the inside of the cell.
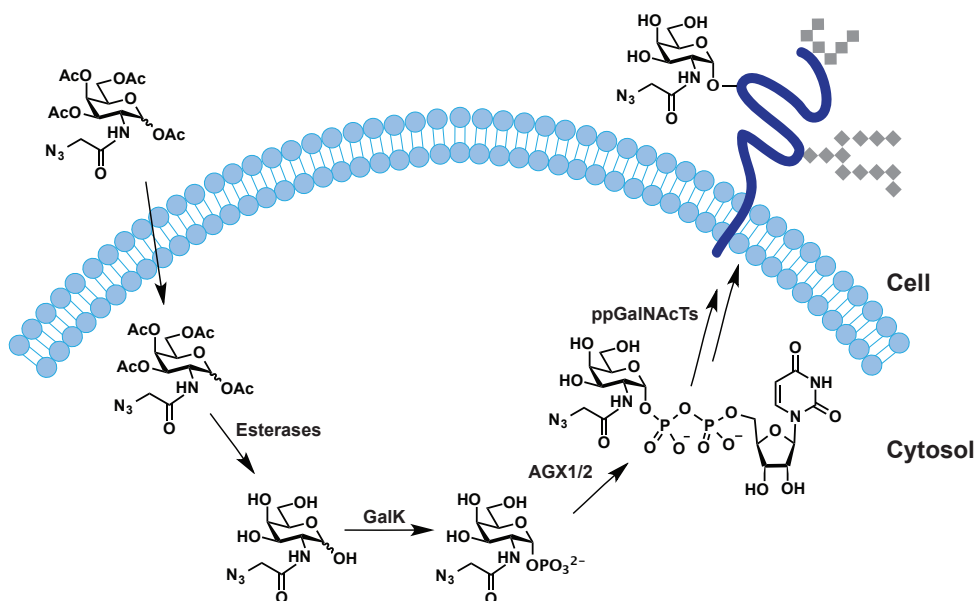
Figure 4-16. The GalNAc salvage pathway. GalK processes large analogs poorly. UDP sugars are not membrane permeable.

4.4.1. Delivery via the GalNAc salvage pathway

If the main bottleneck in the GalNAc salvage pathway is the galactokinase, it is possible to deliver a protected sugar-1-phosphate of the bumped analog. This approach has been successful with other sugar analogs including the diazirine modified GlcNAc analog developed by Kohler and colleagues. As shown in Figure 4-17, the hydroxyl groups of the sugar are acetylated while the phosphate oxygens are protected with S-acyl-2-thioethyl (SATE) protecting groups[114,195,196]. The acetyl portion of these SATE protecting groups can be removed by esterases, unmasking a self-immolative protecting group that releases ethylene sulfide and exposes the free phosphate.
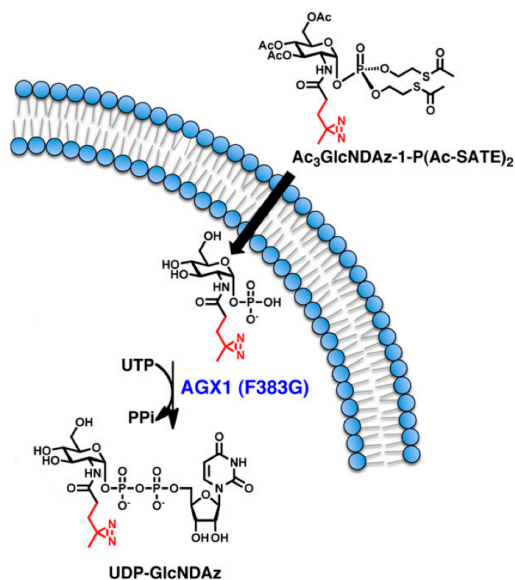
Figure 4-17. Delivery of a protected sugar-1P analog enables formation of the UDP-sugar. Figure reprinted from Kohler et al[114]. Copyright 2012, the National Academy of Sciences.

We first evaluated the whether GalNAzMe(S) could be delivered to cells as the protected monosaccharide or the protected sugar-1-phosphate. We used peracetylated GalNAz as a positive control and evaluated both the formation of the UDP sugar, as well as azidosugar labeling of glycoproteins. Nucleotide sugar formation can be evaluated by High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD or HPAEC) as described by Kohler and colleagues[114]. We first used authentic standards of UDP-GalNAc, UDP-GlcNAc, and UDP-GalNAz, and UDP-GalNAzMe(S) to confirm the elution times of these distinct nucleotide sugars. We observed that both UDP-GalNAz and UDP-GalNAzMe(S) had the same retention time by this method. HEK cells were treated for 6 hours with different protected sugar analogs, and then the cells were harvested and lysed. As shown in Figure 4-18, cells fed Ac$_4$GalNAz demonstrated significant formation of both UDP-GalNAz and UDP-GlcNAz by HPAEC. In contrast, cells treated with either Ac$_4$GalNAzMe(S) or Ac$_3$GalNAzMe(S)-1P(SATE)$_2$ did not form UDP-GalNAzMe(S).

Although no UDP-sugar formation with the GalNAzMe(S) derivatives was apparent by HPAEC, we also evaluated whether these analogs would result in detectable azidosugar labeling of cell surface glycoproteins. Cells were treated with cyclooctyne conjugated to biotin (DIBAC-PEG-biotin) and lysed. Labeled lysates were then subjected to SDS PAGE and western blot analysis. By probing with an anti-biotin HRP conjugate, the amount of azidosugar labeling was evaluated. As would be predicted based upon the HPAEC results described above, glycoproteins labeled with GalNAz were present in the positive control (Figure 4-18). In contrast, the methyl analogs showed identical labeling to that observed in the no azidosugar

and no cyclooctyne controls. These results indicate that in the absence of UDP-GalNAzMe(S) formation, cellular proteins do not incorporate the methyl analog via any other detectable pathway.
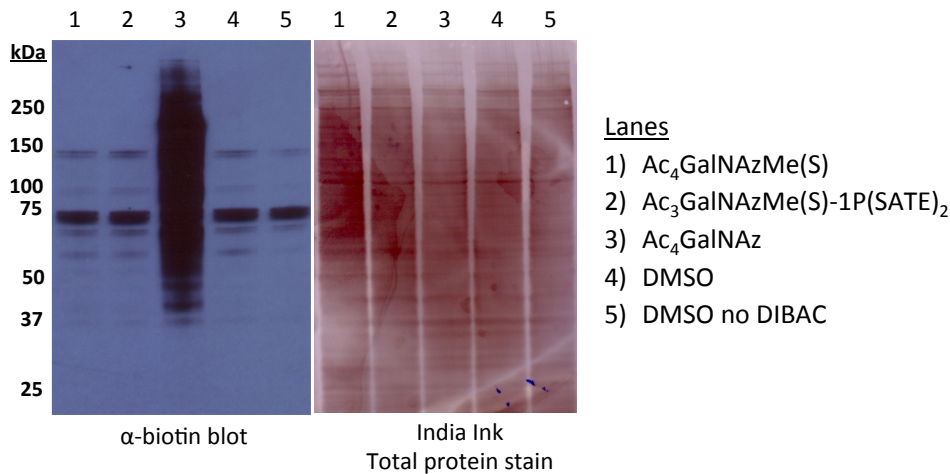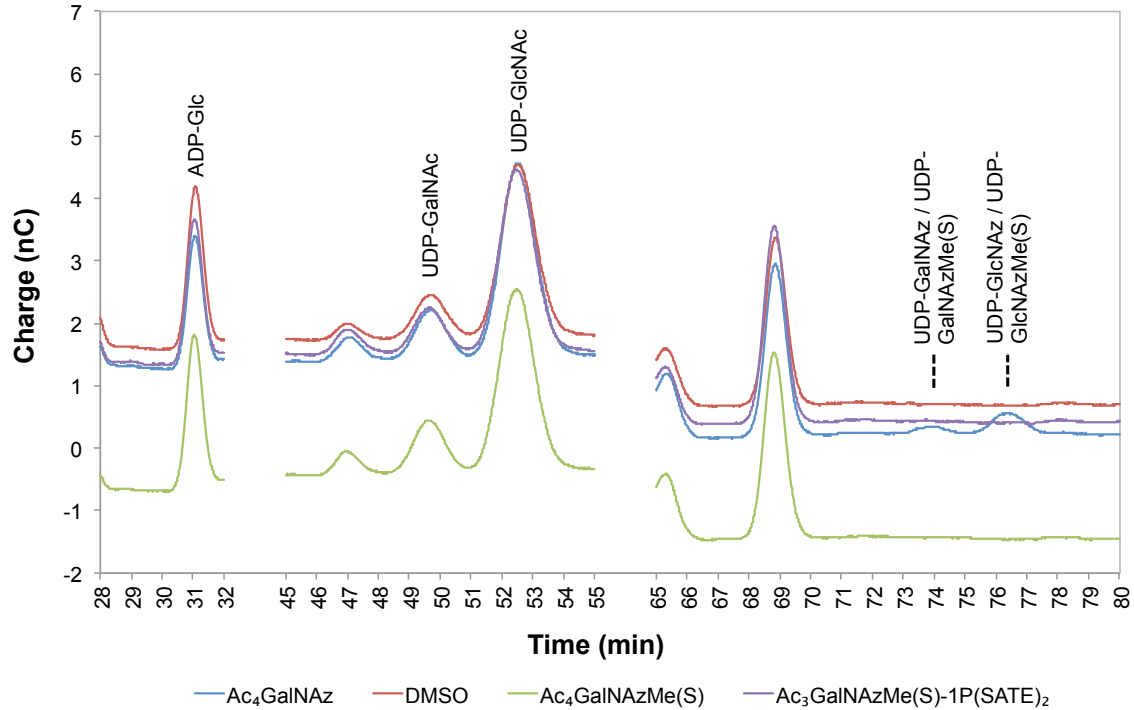
Figure 4-18. GalNAzMe(S) intermediates are not accepted by the GalNAc salvage pathway. (Top panel) HPAEC-PAD analysis of HEK 293T cells fed Ac4GalNAzMe(S), Ac3GalNAzMe(S)-1P(SATE)2, Ac4GalNAz (positive control) or DMSO (negative control). (Bottom panel) Western blot of cells fed sugar analogs, treated with DIBAC-biotin, and lysed.

4.4.2. Delivery bypassing the GalNAc salvage pathway

If the bumped sugar analog is not accommodated by the GalNAc salvage pathway, it could be bypassed by delivering the UDP-sugar directly. However, the di-anion of the UDP sugar makes this quite challenging. There are no known protecting groups that facilitate passive diffusion of the small molecule through the cell membrane. This is because protecting the phosphate oxygens makes them good leaving groups and activates the UDP sugar for hydrolysis. Therefore methods to deliver a nucleotide sugar depend on disrupting or bypassing the cell membrane. Shokat and coworkers have delivered ATP analogs to cells by permeabilizing the cell membrane with digitonin[15,103]. However, membrane permeabilization compromises cell integrity, and these cells only survive for a few hours[15,103]. In the case of kinases, phosphorylation is a rapid and dynamic modification, allowing them to achieve labeling on this short timescale. In contrast, o-GalNAc glycosylation is significantly more complex, first requiring the UDP sugar to be transported into the Golgi, followed by modification by multiple glycosyltransferases within the Golgi. Complete trafficking of the protein through the secretory pathway and finally to the cell surface is necessary to achieve labeling of mature proteins. These challenges necessitate cell survival on a significantly longer timescale than was achieved by Shokat and coworkers in order to achieve meaningful labeling.

We attempted an alternative cell permeabilization method using the streptolysin toxin, which has been shown to enable delivery of many types of cargo across cell membranes[197-199]. Treatment with Ca^{2+} has been shown to induce endocytosis, resulting in removal of the streptolysin pores from the cell membrane and making the membrane no longer permeable[198,199]. Cells treated in this fashion have been shown to survive greater for longer than 24 h[199]. Our attempts to repeat these experiments using calcein dye to visualize small molecule delivery did not indicate that this technique would be promising.

We have also considered bypassing the cell membrane to deliver the UDP-sugar. There is significant precedent encapsulating polar pharmacophores and charged biopolymers such as nucleic acids inside a vesicle that is endocytosed by the cell[200]. These molecules then escape the endosome to reach the cytoplasm by unknown mechanisms[201]. Vesicles such as liposomes are typically composed of amphiphilic molecules with a polar head group and a hydrophobic tail, and some liposome compositions are designed to decompose during acidification in the endocytic pathway[200,201].

4.4.3. Engineering the GalNAc salvage pathway

An alternative to delivering the intact UDP-sugar is to deliver an earlier intermediate by engineering an enzyme in the GalNAc salvage pathway. With confirmation that the methyl analog is not converted into the nucleotide sugar by the natural enzymes, we chose to focus our efforts on engineering either GalK or AGX1/2. AGX1 preferentially uses UDP-GalNAc over UDP-GlcNAc, while AGX2 is

known to preferentially utilize UDP-GlcNAc[131,202]. We choose to engineer AGX1 rather than GalK because it is the later bottleneck in the pathway. This is the strategy that was ultimately necessary to deliver the diazirine GlcNAc analog discussed earlier in this chapter. Kohler and coworkers found that despite the delivery of the protected sugar-1-phosphate, the analog was not processed by AGX1/2 and did not result in the formation of UDP-sugar[114]. Therefore they engineered AGX1 to introduce an enlarged binding pocket that could accommodate the diazirine. Using the crystal structure of AGX1 to design the active site mutations, Kohler and coworkers engineered AGX1 to accommodate modifications at the C8 position of UDP-GlcNAc[114]. They found that the modification of Phe383 to Gly resulted in the formation of UDP-GlcNDAz.
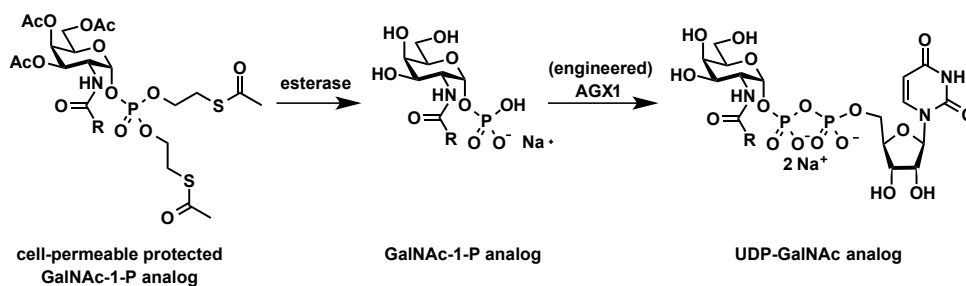
Figure 4-19. UDP-GalNAc analog synthesis using engineered AGX1.

Based on this precedent, we attempted to utilize the same AGX1 mutant that had accepted GlcNDAz. The Kohler lab generously provided us with a stable HeLa cell line transfected with a mammalian expression construct for AGX1-F383G maintained with puromycin resistance cassette. When we fed these cells GalNAzMe(S)-1P(SATE)$_2$, we did not detect the formation of the corresponding UDP sugar by HPAEC. As observed by the Kohler and coworkers, the crystal structure of AGX1 contains a hydrophobic pocket in which F381 and F383 flank the N-acyl side chain of UDP-GlcNAc[114]. We modeled residues in the active site with PyMOL and concluded that single mutants of F381 and F383 and the F381/F383 double mutant were the best candidates for active site engineering. PyMOL modeling of the glycine mutations suggested that they might create large pockets in the active site (Figure 4-20). However, we were concerned that glycine might destabilize the peptide backbone and hence destabilize ligand binding. Therefore, we developed a panel of mutants with both alanine and glycine substitutions including F381G, F381A, F383G, F383A, F381G/F383G, F381A/F383A. With the expression construct of AGX1-F383G provided by the Kohler lab, we introduced a C-terminal FLAG tag and generated wildtype AGX1 and all aforementioned AGX1 mutants.
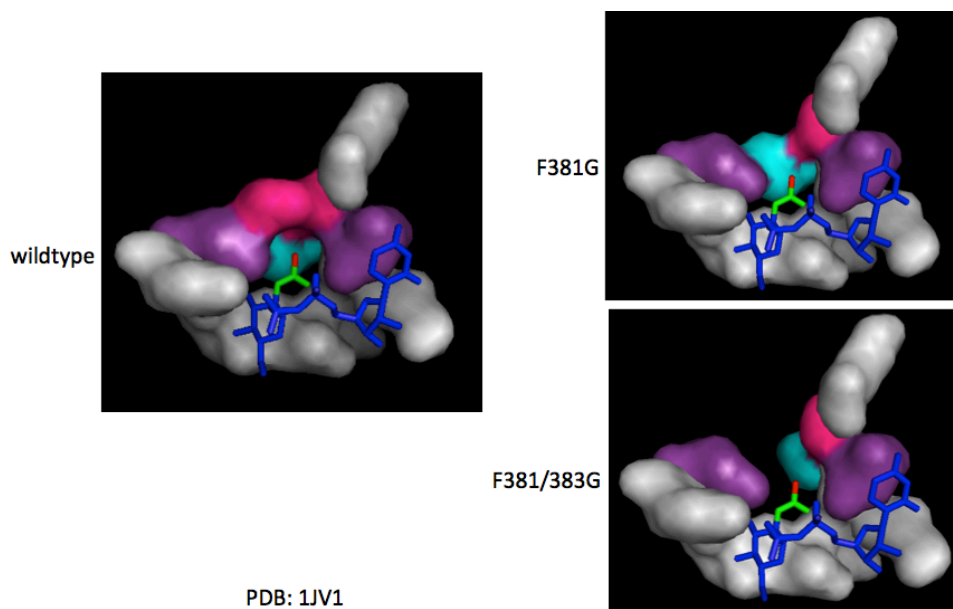
Figure 4-20. Modeling of AGX1 and mutants. Space-filling interactions of residues within 5 angstroms of UDP-GlcNAc are shown. F381 (cyan) and F383 (pink) are shown flanking the C8 methyl group (red) on the N-acyl side chain (green) of UDP-GlcNAc (blue). PDB ID 1JV1.

Transient transfection of HEK cells with these plasmids, followed by feeding $Ac_3GalNAzMe(S)-1P(SATE)_2$ resulted in the production of UDP-GalNAzMe(S) by the F383A mutant (Figure 4-21). Although these results were reproducible, we observed that the F383A mutant was expressed at much higher levels than the other mutants, and at moderately higher levels than the exogenously expressed wildtype enzyme (see Experimental, Figure 4-24). This may be due to improved stability of the F383A mRNA or protein, or this result could be caused by factors that are not intrinsic to the mutant itself. We have recently developed stable cell lines expressing the wildtype and F383A mutants in an effort to equalize protein expression during the analysis of UDP sugar formation.

105
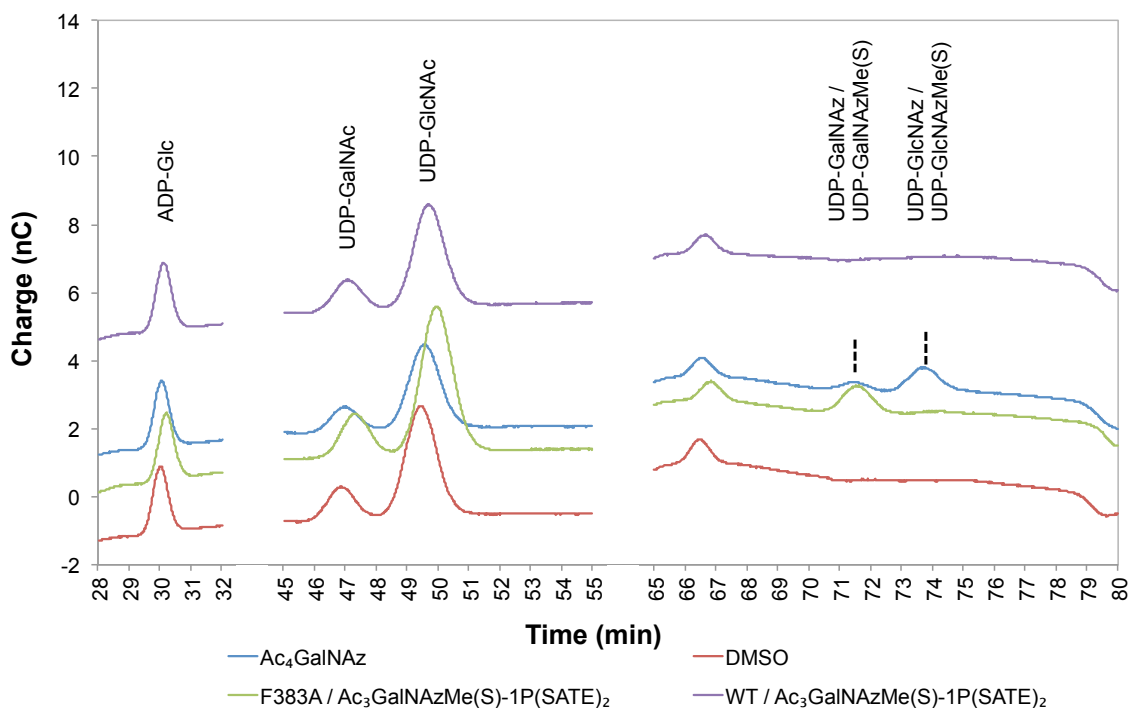
Figure 4-21. Ac$_3$GalNAzMe(S)-1P(SATE)$_2$ is converted to UDP-GalNAzMe(S) by AGX1-F383A.

Upon establishing the bump-hole pair and cellular delivery of the UDP-sugars, we are continuing our studies of the bump-hole technology *in cellulo*. For these validation experiments, we are focusing on the ppGalNAcT2 bump-hole pair, due to the wide body of literature regarding its protein substrates[2a, 23]. As described in Chapter 2, we are finalizing cloning of full-length wildtype and mutant ppGalNAcT2. We will then transiently transfect the stable HEK 293T cell lines expressing AGX1-F383A with full-length ppGalNAcT2. Ac$_3$GalNAzMe(S)-1P(SATE)$_2$ will be delivered to the cells. Cell-surface glycoproteins containing azides will be labeled by treating live cells with a cyclooctyne-biotin conjugate, followed by lysis and western blot analysis.

Ideally, the UDP-GalNAc analog continues to behave as if it is fully orthogonal, such that signal only arises in the presence of both mutant enzyme and substrate analog. The wildtype and engineered ppGalNAcT2 should both process UDP-GalNAz, so Ac$_4$GalNAz will serve as a positive controls. We will also employ the appropriate wildtype controls for both ppGalNAcT2 and AGX1 to validate the results with Ac$_3$GalNAzMe(S)-1P(SATE)$_2$. To probe the molecular details of bump-hole orthogonality, we will use the cell lysates generated in these experiments to enrich the biotinylated proteins, trypsonize, and analyze the peptides by mass spectrometry. This will result in a list of glycoproteins that are glycosylated by the engineered T2, which we can compare with substrates for wildtype ppGalNAcT2 that have been either 1) previously identified through the SimpleCell proteomics screening[2a], 2) biochemically validated, or 3) identified for the first time.
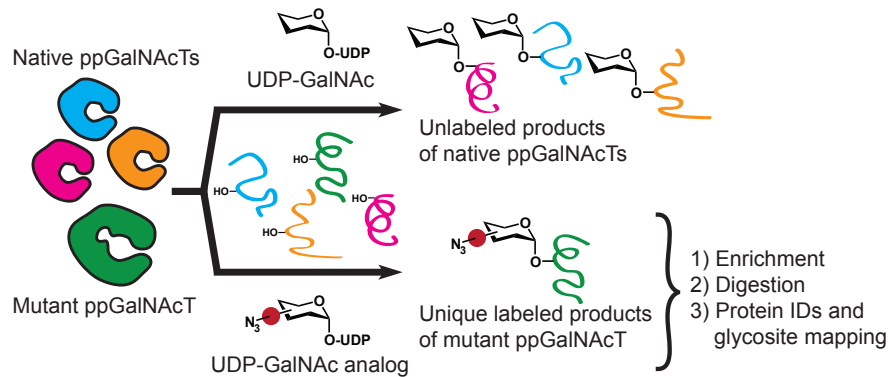
Figure 4-22. Application of the bump-hole pair to glycoproteomics

4.4.4. Confirm ppGalNAcT specificity is conserved with a peptide library

Previous work has established ppGalNAcT2 preferences for peptide primary sequences[14,53,203-205], and the Bertozzi lab has used UDP-GalNAz to evaluate preferences for various glycoforms of EA2 peptide[13]. Using a library of oriented random peptide substrates, Gerken and colleagues have shown that substrate preferences are highly conserved between Drosophila and mammalian ppGalNAcT orthologs, particularly for ppGalNAcT2 [203,204]. This suggests that it will be possible to diagnose changes in acceptor substrate profiles for mutant ppGalNAcT2. The deviation by the bump-hole pair from the preferences of wildtype ppGalNAcT2 will be evaluated, and if necessary, we may return to find another bump-hole pair.

We have considered three approaches to evaluating the substrate preferences of the candidate bump-hole pair. First, we are engaged in a collaboration with Thomas Gerken and a graduate student in his lab, Leslie Revoredo. We have provided them with purified wildtype and double mutant ppGalNAcT2, as well as UDP-GalNAz and UDP-GalNAzMe(S). They will be comparing the substrate preference of the double mutant and UDP-GalNAzMe(S) to the preferences of the native pair. They will be using the same technique as described by Perrine *et al*., where a huge number of peptides are screened in one pot, azido-glycopeptides are enriched, and preferred sequences are identified by Edman sequencing[53]. A randomized peptide library that has been previously evaluated with ppGalNAcT2 will be used to determine whether the same categories of peptides are selected by the bump hole pair.

It is also possible to use a smaller library of known substrates and to evaluate them one by one in an azido-ELISA format or using the mass spectrometry screen described in Chapter 4.1. Among the unglycosylated peptides of interest, most are available commercially, and our lab has access to the glycosylated EA2 library synthesized by Pratt and colleagues[13].

Finally, when we carry out a glycoproteomics study in cells, we plan to compare our hits to the lists generated with SimpleCells (SC) and SC/ppGalNAcT2-/- cells developed by Clausen and colleagues[83].

4.5. Future Directions

4.5.1. Measure K_M and V_{MAX} for bump-hole pair with a model peptide substrate

Our group has previously measured kinetic parameters for ppGalNAcT activity with UDP-GalNAc and UDP-GalNAz and the acceptor peptide EA2 biotin using an ELISA format[6,80]. In these experiments, a dilution series of UDP-sugar was assayed against the K_M concentration of acceptor peptide and a fixed concentration of enzyme[6]. The reactions were allowed to proceed to approximately 20% completion and were stopped with the addition of EDTA[6,80]. Because crude conditioned medium from COS-7 cells transiently expressing ppGalNAcTs was the enzyme source, the enzyme concentration was standardized but not measured quantitatively[6]. Therefore V_{MAX} was reported in units of $\mu M/min$, which corresponded to K_M concentrations of UDP-sugar in μM[6,80].

One significant disadvantage of this assay for the evaluation of enzyme kinetic is that the glycan modification must be detected differently for the native versus analog donor substrates. Detection of glycopeptides with or without a chemical handle demands a GalNAc-specific lectin and the Staudinger ligation, respectively, which precludes direct comparison of kinetic parameters[6,80,84].

The LC-MS assay described in this chapter has proven robust to different reaction parameters and UDP-GalNAc analogs. Therefore, it will be used to monitor glycosylation of acceptor peptide over a three hour time course at multiple concentrations of UDP-sugar, with saturating EA2 peptide and a fixed concentration of enzyme. These values can be used to generate initial velocity curves, from which V_{MAX}, K_M, and k_{cat} can be extrapolated. Parameters obtained from these studies will be compared to the literature values for the native pair.

4.5.2. Known substrates: ANGPTL3

We predict that ANGPTL3, a known substrate of ppGalNAcT2 and a protein that positively correlates with HDL levels, will be glycosylated by engineered hT2. Although its glycosylation serves to protect against protein degradation, glycosylation site identification has proven troublesome. A single ppGalNAcT2-targeted glycosylation site has been found in a study using small peptides that covered putative glycosylation sites[70]. This specific site has not yet been verified on full-length proteins or *in vivo*[83]. We aim to identify ANGPTL3 glycosylation sites by expressing ANGPTL3 with a FLAG tag in cells along with mutant ppGalNAcT2 and its optimized UDP-GalNAc analog. The azido-glycosylated protein will be labeled with a cleavable biotinylated probe, isolated by immunoprecipitation, and evaluated by MS/MS proteomics[30].

4.5.3. Late transferases: Discover mechanisms of cancer metastases by engineering ppGalNAcT7

After establishing the bump-hole technology as a method to selectively label glycoproteins modified by a specific ppGalNAcT, we will focus on a proteome-level discovery of the protein targets of human ppGalNAcT7. This transferase is upregulated in human cervical carcinoma and is regulated by microRNAs whose aberrant expression in cervical cancer lines and human melanoma enhances metastatic behavior in cells, animal models, and human disease[206,207]. Changes in microRNA levels directly alter hT7 expression, which leads to phenotypic changes that can be reversed by tuning hT7 levels, indicating that hT7 plays a key role in invasiveness of cervical cancer and melanoma. However, it is of note that hT7 and downstream protein glycosylation may play dissimilar roles in these cancers. Overexpression of hT7 resulted in pro-metastatic behavior in HeLa and C33A cervical cancer cells[207]. In contrast, repression of hT7 in primary melanoma cells WM35 and WM98 induced metastasis, and overexpression of ppGalNAcT7 in a mouse model of melanoma to lung cancer metastasis reduced metastatic potential[206].

To unravel the mechanism of T7 in metastasis, bump-hole methodology is critical for identifying the protein targets of T7. T7 is a family member that prefers GalNAc-glycopeptides, so it is predicted to act on previously glycosylated proteins. This poses analytical challenges to all current methods of unbiased ppGalNAcT protein substrate screening. Although the first proteomics screen for ppGalNAcT protein substrates was made possible with the truncated O-GalNAc glycans found in SimpleCell technology, this technology requires the detection of an absence of signal. For a glycoprotein-preferring ppGalNAcT, the knockout system must compete with all other sites of O-GalNAc glycosylation from the early-acting ppGalNAcTs. In contrast, the bump-hole pair enables selective enrichment of T7 modified glycoproteins independent of modification site, degree of glycosylation, or other ppGalNAcTs involved.

We will engineer a hT7 bump-hole pair using the active site mutations and UDP-GalNAc analogs developed in this work. DNA will be introduced to relevant tumor cell lines: HeLa and C33A for cervical carcinoma and WM35 and WM98 for melanoma. Creating cell lines that replace native ppGalNAcT7 with the mutant T7 should enhance the signal-to-noise ratio and improve the accuracy of protein and site IDs. To this end, we will use the CRISPR/Cas9 platform pioneered at UC Berkeley to knock out the endogenous ppGalNAcT7 gene (GALNT7) and replace it with a GALNT7 mutant and GFP transgene. With engineered cells in hand, we will confirm that the bump-hole pair can recapitulate native cell migration and invasiveness using a Transwell migration assay[206,207]. As the engineered ppGalNAcT7 should not recognize UDP-GalNAc, cells should behave as a T7 knockout that can be rescued by the UDP-GalNAc analog. Finally, we can regulate the expression of ppGalNAcT7 in these cells by altering the levels of the relevant microRNAs, and use the UDP-GalNAc analog to monitor differential glycosylation as

cells undergo changes in metastatic potential. Azido-glycoproteins will be labeled with a biotin affinity handle, enriched, and analyzed by mass spectrometry. These data will include a list of proteins with altered glycosylation in the metastatic and non-metastatic state that arises from ppGalNAcT7 glycosylation.

4.5.4. Animal studies: Identify T2 substrates involved in HDL and triglyceride metabolism

Due to the challenges associated with UDP-sugar delivery, in order to apply the bump-hole pair in a living system, the first logical step is to take advantage of the Bertozzi lab's experience injecting nucleotide sugars into zebrafish embryos[17,20]. By co-injecting the UDP-GalNAc analog and RNA encoding the modified ppGalNAcT2 into a zebrafish embryo at the 1-cell stage, it will be possible to administer the bump-hole pair to all cells in the developing animal.

Genome-wide association studies in humans have identified a correlation between ppGalNAcT2 and blood levels of both high-density lipoprotein (HDL) cholesterol and triglycerides[208]. The locus identified is an intronic region of ppGalNAcT2, which suggests that it may be involved in regulating ppGalNAcT2 expression[208,209]. As a glycosyltransferase, it has been hypothesized that ppGalNAcT2 modifies proteins involved in lipid metabolism and that this glycosylation helps regulate HDL and triglyceride levels. Normally, HDL cholesterol plays a protective role and is antiatherogenic, while triglycerides are atherogenic[209-211]. Recent studies in the mouse model indicate a role for ppGalNAcT2 in mediating plasma HDL cholesterol levels[209].The authors demonstrated that liver-specific overexpression of the GALNT2 gene significantly lowered HDL levels, while RNA-mediated knockdown of the GALNT2 gene in liver resulted in higher plasma HDL levels[209]. These results are intriguing, but in the absence of unambiguously identified substrates of a single ppGalNAcT, it is challenging to confirm the role of ppGalNAcT2 in the glycosylation of proteins that regulate plasma lipid levels.

The bump-hole technique enables the direct identification of substrates of ppGalNAcT2. We propose the application of this technique to the targeted examination of major proteins involved in HDL cholesterol and triglyceride metabolism. The most abundant protein components of the HDL particle are apolipoproteins A-I (apo A-I) and A-II (apo A-II). Less abundant protein components of the HDL particle include apo A-IV, apo C-I, apo C-II, apo C-III, apo E, apo M, serum amyloid A, and serum amyloid A-IV, among others[212]. During phases of triglyceridemia, HDL particles sequester and regulate levels of apo C-II, the activator protein for lipoprotein lipase[213]. Other proteins that interact with HDL cholesterol include lecithin-cholesterol acyltransferase, the enzyme responsible for esterification of plasma cholesterol for which HDL is a substrate[213]. Finally, the class B scavenger receptor SR-BI acts as the HDL receptor[214].

O-linked glycosylation is a common modification of apolipoproteins[215]. In particular, studies have shown that O-glycosylation of apo A-II alters its association with HDL particles[215]. Additionally, abnormal HDL particles resulting from non-enzymatic glycation of apo A-I may contribute to atherosclerosis[211,216]. Proteomic analysis of the protein components of HDL particles indicates that apo A-I and A-II both occur as multiple glycoforms, and that some of these glycoforms differ between types of HDL particles[212]. Apo C-III is also a component of very low density lipoprotein (VLDL) and is O-glycosylated only at the residue Thr-94[217]. These data suggest that it may be possible to identify a role for ppGalNAcT2 in the O-glycosylation of apolipoproteins that interact with HDL.

In future experiments, the bump-hole pair will be injected into early zebrafish embryos as described above. After developing for a few days, the fish will be sacrificed and the cells harvested. Upon cell lysis, tagged glycoproteins will be isolated by reaction with a biotin probe followed by streptavidin purification and elution. Immunoprecipitated proteins will be analyzed by SDS-PAGE and immunoblotting using anti-biotin antibodies in parallel with available antibodies against the different apolipoproteins described above. If positive hits are identified, further validation will be carried out, including treatment to remove O-glycans from purified proteins prior to western blot analysis. O-glycan removal would be expected to eliminate signal from the protein of interest on an anti-biotin blot, while leaving the anti-apolipoprotein signal intact. If positive hits are not identified, antibodies against other classes of proteins that interact with HDL cholesterol and triglycerides will next be interrogated.

The Rader lab at the University of Pennsylvania is expert in factors that regulate the function and metabolism of plasma lipoproteins. They are interested in applying our techniques to their research, and we have discussed a collaboration with their group to study ppGalNAcT2 substrates in the mouse liver. Independent of a collaboration, their expertise might be of significant value while screening apolipoprotein targets in zebrafish. Additionally, the potential application of the bump-hope technique in this context could be very powerful. They have previously used viral vectors to overexpress and knockdown GALNT2 in the mouse liver[209]. With their expertise, we may be able to explore application of the engineered hT2 and UDP-GalNAc analog in a biologically relevant model.

4.6. Experimental

4.6.1. General

All chemical reagents were analytical grade and commercially available, unless otherwise noted. Organic reactions were performed in oven-dried reaction vessels under an N_2 atmosphere, and liquid reagents were added with a dry needle unless otherwise noted. Sealed bottles of anhydrous methanol and anhydrous pyridine were purchased from Alfa Aesar and Acros Organics, respectively. All other solvents were purified as described by Pangborn *et al.*[48] Organic extracts were dried with $MgSO_4$ and solvent was removed with a rotary evaporator at reduced pressure. Flash chromatography was carried out with Silicycle SiliaFlash® P60 230-400 mesh silica gel according to the procedure described by Still *et al.*[49] Thin layer chromatography was carried out with Silicycle 250 micron 60 Å glass back silica gel plates and visualized with *p*-anisaldehyde or ninhydrin staining or by absorbance of UV light at 245 nm.

Reverse phase-HPLC was performed using a modular Varian Prostar HPLC system with 210 and 254 nm detection on a Microsorb C_{18} Dynamax semi-preparative column (10 x 250 mm) at a flow rate of 4 mL/min or preparative column (21.4 x 250 mm) at a flow rate of 20 mL/min. Bio-Gel P-2 Gel was from BioRad. ^1H, ^{13}C, and ^{31}P NMR spectra were obtained with AVQ-400, AVB-400, DRX-500, AV-500 or AV-600 Bruker spectrometers. Chemical shifts (δ) are reported in parts per million referenced to the solvent peak and coupling constants (J) are reported in Hz. High resolution electrospray ionization (ESI) mass spectra were obtained from the UC Berkeley Mass Spectrometry Facility.

All biological reagents were molecular biology or cell culture grade and from commercial sources, unless otherwise specified. Chemically competent *E. coli* strain One Shot® Top10 and S.O.C. medium were purchased from Invitrogen. Bacteria were grown in LB broth (Lennox L Broth Base, Invitrogen) supplemented with 100 μg/mL ampicillin (LB/amp) or 50 μg/mL kanamycin (LB/kan) and grown on LB/amp or LB/kan agar plates. The Plasmid Midi and Maxi kits were from QIAGEN.

DpnI and T4 DNA ligase were purchased from New England Biolabs (NEB). PfuUltra II Fusion HS DNA polymerase was from Agilent. 1 Kb and 100 bp DNA ladders were from Invitrogen. DNA Clean and Concentrator™-5, Zyppy™ Plasmid Miniprep Kit, and Zymoclean™ Gel DNA Recovery kit were from Zymo Research. Elim Biopharmaceuticals, Inc (Hayward, CA) synthesized all primers and performed all plasmid DNA sequencing.

Dulbecco's Modified Eagle Medium with high glucose (DMEM), Dulbecco's Phosphate-Buffered Saline without Calcium or Magnesium (DPBS), and Penicillin/Streptomycin were purchased from Hyclone. Fetal bovine serum (FBS) was obtained from Omega Scientific and 0.25% trypsin/EDTA and Opti-MEM® Reduced Serum Medium were purchased from Invitrogen. TransIT®-293

Transfection Reagent was obtained from Mirus Bio. COmplete EDTA-free Protease inhibitor cocktail tablets were purchased from Roche.

Criterion™ XT Bis-Tris Precast gels, XT-MES Running Buffer, nitrocellulose membrane (0.45 micron), Precision Plus Protein™ Kaleidoscope™ Prestained Standards (10-250 kDa), and Kaleidoscope™ Prestained Standards (7.6-216 kDa) were obtained from BioRad. SuperSignal West Pico Chemilluminescent Substrate was purchased from Fisher Scientific. Rabbit polyclonal anti-AGX1 (UAP1) was from GeneTex (Cat. GTX103592). Goat anti-Rabbit IgG-HRP (human adsorbed) was from Southern Biotech (Cat. 4010). AF647 conjugated Goat anti-Mouse IgG H+L was from Abcam (Cat. Ab150119)

MacPyMOL was used to model the crystal structure of AGX1 (PDB 1JVI).

4.6.2. General Methods

Manufacturer recommended protocols for Zymo Research kits were subject to the following modifications: Prior to elution, spin columns were dried by centrifugation at maximum speed for 2 min. DNA was eluted with water added directly to the filter, followed by 1 min incubation and 1 min centrifugation at maximum speed).

DNA was quantitated using a Nanodrop 2000 UV-vis spectrophotometer.

DNA samples were pre-mixed for sequencing: 500 ng plasmid DNA, 8 pM primer, 15 µL total in water.

Human Embryonic Kidney (HEK) 293T cells (ATCC) were maintained in a 5 % CO_2, water-saturated atmosphere at 37 °C in DMEM supplemented with 10 % FBS, penicillin (100 U/mL), and streptomycin (0.1 mg/mL). Adherent cell densities were maintained between 1 x 10^5 and 2 x 10^6 cells.

4.6.3. General Protocol 1: Standard PCR conditions

PCR amplifications were carried out in duplicate for all cloning applications as recommended by the manufacturer: 25 ng template DNA, 0.2 µM forward primer, 0.2 µM reverse primer, 250 µM of each dNTP, 3% DMSO, 1x PfuUltra II buffer, and 1 µL PfuUltra II Fusion HS DNA polymerase in a final reaction volume of 50 µL in water.

PCR for site-directed mutagenesis was carried out according to modified QuikChange™ site-directed mutagenesis protocols as described by Zheng *et al*. (2004) for point mutations and Liu *et al*. (2008) for insertions and deletions[136,143].

Thermocycle conditions were 95 °C for 2 min; 18 cycles of 95 °C for 20 sec, annealing temperature for 20 sec, 72 °C extension for 15sec/ kb; 72 °C for 3 min; 4 °C hold. For primers with mismatches for point mutations, annealing temperatures

were generally 55 °C and are listed with the primers. For primers with 5' overhangs for insertions/ deletions, annealing temperatures were generally T_M – 5 °C corresponding to the portion of the primer designed to anneal to the template. Amplicon size and extension times are listed with the primers.

4.6.4. General Protocol 2: Standard protocol for bacterial cloning

Bacterial transformations for cloning were carried out in One Shot® Top10 *E. coli* according to the manufacturer's instructions as described below. Modifications for specific applications are in the second paragraph. Bacteria (50 μL) were thawed on wet ice for 10 min, 2.5 – 125 ng DNA was added based on the application, tubes were flicked gently to mix, and *E. coli* were incubated on wet ice for 25 min. *E. coli* were subjected to heat shock at 42 °C for 30 sec, incubated on wet ice 2 min, and 250 μL rt S.O.C. medium was added. *E. coli* were grown shaking at 225 rpm at 37 °C for 35 min (Amp^R) to 45 min (Kan^R). Based on the application, 25-100 μL *E. coli* were spread onto LB agar plates with 100 μg/mL ampicillin or 50 μg/mL kanamycin. Plates were incubated at 37 °C for approximately 16 h and then stored at 4 °C. Individual colonies were picked and used to inoculate 5 mL LB with antibiotics (100 μg/mL ampicillin or 50 μg/mL kanamycin). 5 mL cultures were grown shaking at 225 rpm at 37 °C for approximately 16 h, and cells were pelleted at 3,650 x g at 4 °C for 15 minutes. Supernatant was removed, pellets were flicked and then suspended in 600 μL water. DNA was extracted using a Zyppy™ Plasmid Miniprep Kit according to the manufacturer's protocol with modifications as described in the General Methods (30 μL elution).

Quantities of DNA added for transformations and the volume of transformed *E. coli* plated varied based on the application. When plasmid DNA was previously purified from *E. coli*, 0.25 μL (25-125 ng DNA) was used for transformation, and 25-35 μL *E. coli* were plated. When plasmid DNA was prepared via ligation, 2 μL ligation (2.5 ng DNA) was used for transformation, and 100-200 μL *E. coli* were plated. When plasmid DNA was prepared via PCR for site-directed mutagenesis, 2.5 ng DNA (0.25-5 μL) was used for transformation, and 100-200 μL *E. coli* were plated.

4.6.5. General Protocol 3: SDS-PAGE

Samples were prepared by adding 4x SDS protein loading buffer (final concentrations 1x sodium dodecyl sulfate, 2.5 % 2-mercaptoethanol) and denatured by boiling at 95 °C for 5 min. Samples were centrifuged at maximum speed for 15 sec and loaded onto a 12 % or 4-12 % Bis-Tris precast gel. Electrophoresis was carried out in XT-MES buffer (BioRad) at 180 V for approximately 45 minutes. Gels were rinsed briefly with water following electrophoresis.

4.6.6. General Protocol 4: Western blotting

For western blotting, proteins were transferred from an SDS-PAGE gel to a nitrocellulose membrane in Tris/Glycine buffer (BioRad, 20% MeOH) at 50V for 2 h.

The membrane was incubated in Ponceau total protein stain (0.2% Ponceau, 3% acetic acid) for 5 min, rinsed with water, and imaged. The membrane was blocked with 5% bovine serum albumin in TBST (25 mM Tris-Cl pH 7.4, 150 mM NaCl, 0.1% TWEEN-20) mixing at rt for 30 min. The blocking solution was discarded, antibody (diluted in 5% milk/TBST) was added, and it was incubated mixing overnight at 4 °C. The blot was rinsed 3 x with TBST for 10 min. The blot was imaged with an HRP detection kit.

4.6.7. General Protocol 5: Standard HPAEC protocol

Samples were analyzed by High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD). Four solvent compositions were prepared (A: MQ; B: 1 mM NaOH in MQ; C: 1M NaOH in MQ; D: 1 mM NaOH, 1M NaOAc in MQ) by first degassing the MQ under vacuum in a sonication bath for 25 minutes and subsequent addition of the additives. The gradient used for HPAEC was: 0 min (A: 0; B: 95; C: 0; D: 5); 20 min (A: 0; B:60; C: 0; D: 40); 60 min (A: 0; B: 60; C: 0; D: 40); 63 min (A: 0; B: 50; C: 0; D: 50); 83 min ((A: 0; B: 50; C: 0; D: 50); 87 min (A: 0; B: 0; C: 0; D: 100); 95 min (A: 0; B: 0; C: 0; D: 100); 97 min (A: 0; B: 95; C: 0; D: 5); 105 (A: 0; B: 95; C: 0; D: 5). After every five runs (one MQ and four sample injections), a column wash sequence (0 min (A: 0; B: 95; C: 0; D: 5); 5 min (A: 80; B: 0; C: 20; D: 0); 25 min (A: 80; B: 0; C: 20; D: 0); 30 min (A: 0; B: 95; C: 0; D: 5); 40 min (A: 0; B: 95; C: 0; D: 5) was run. For each sample, 25 µL was injected onto the column. The retention time of the cellular components of interest (UDP-GalNAc, UDP-GlcNAc, UDP-GalNAz, and UDP-GalNAzMe) standards were determined by running 25 µL of standard solutions (200 µM ADP-Glucose, UDP-GalNAc; UDP-GlcNAc; 100 µM UDP-GalNAz, UDP-GalNAzMe).

4.6.8. Malachite green assay
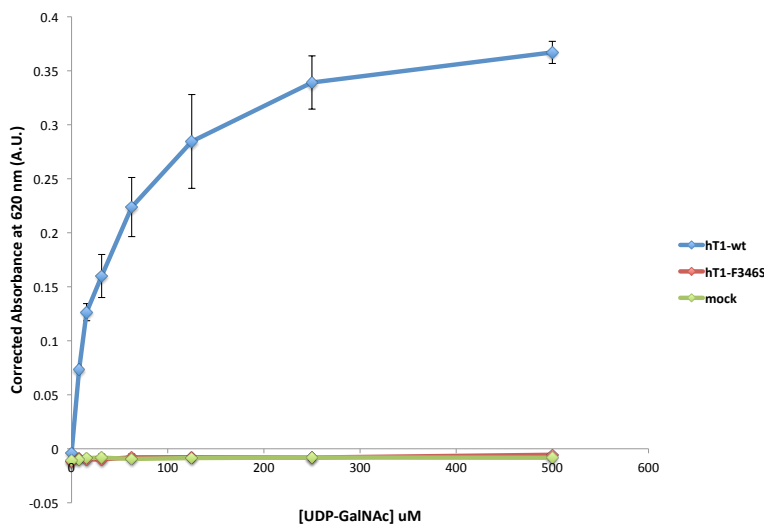
Figure 4-23. UDP-detection assay for hT1 activity.

The malachite green assay was performed according the manufacturer's protocol with the following modifications: Reactions contained 50 µM EA2-biotin (5 mM stock in PBS), a two-fold dilution series of UDP-GalNAc from 500 µM – 37.81 µM (5 mM stock diluted in water), and 25 µL purified ppGalNAcT. Note that ppGalNAcT2 proteins were purified with some departures from the protocol described in Chapter 2.4.12. The modifications include: only elutions 2-3 of protein were used, protein was stored at 4 °C, and protein quantitation had not been optimized at the time of these experiments. Note that ppGalNAcT1 proteins were purified with some departures from the protocol described in Chapter 2.4.12. The modifications include that purified enzyme was stored in 10% glycerol and t the time of these experiments, protein quantitation had not been optimized. Reactions proceeded for 2 h at 37 °C.

4.6.9. Azido ELISA assay

This protocol is adapted from Hang *et al.* for the detection of ppGalNAcT activity with UDP-GalNAz[80]. Glycosylation reactions contained reaction buffer (40 mM sodium cacodylate, 10 mM $MnCl_2$, 0.1% Triton X-100, 40 mM β-mercaptoethanol), 50 µM EA2-biotin (5 mM stock in PBS), and a two-fold dilution series of UDP-GalNAz from 250 µM – 3.91 µM (5 mM stock diluted in water). To initiate reactions, 2 µL purified ppGalNAcT (25 mM Tris-Cl pH 7.4, 150 mM NaCl, 100 µg/mL 3x FLAG peptide, 1x EDTA-free protease inhibitors) was added. Note that ppGalNAcT2 proteins were purified with some departures from the protocol described in Chapter 2.4.12. The modifications include: only elutions 2-3 of protein were used, protein was stored at 4 °C, and protein quantitation had not been optimized at the time of these experiments. The final reaction volumes were 40 µL in a half-volume 96-well plate. Reactions were mixed on a horizontal shaker at rt for 5 min and incubated at 37 °C for 1 h. Glycosylation reactions were quenched with 10 mM EDTA and percent glycosylation was evaluated as described below.

Neutravidin plates were washed 3 x with 200 µL ELISA Buffer (25 mM Tris-Cl pH 7.4, 150 mM NaCl, 0.1% BSA, 0.05% Tween-20) for 10 min. Glycosylation reactions were transferred to Neutravidin plates and incubated at rt for 1 h. Plates were washed 3 x with 200 µL ELISA Buffer for 10 min. Bound peptides were treated with 100 µL Phosphine-FLAG (0.2 M in PBS) and incubated at 37 °C for 2 h. Plates were washed 3 x with 200 µL ELISA Buffer for 10 min and 100 µL anti-FLAG-HRP (diluted 1:5000 in ELISA Buffer) was added. Reactions were incubated at rt for 1 h. Plates were washed 2 x with 200 µL ELISA Buffer for 10 min and washed with 200 µL PBS for 10 min. Bound anti-FLAG-HRP was detected with 100 µL TMB peroxide solution (premixed 1:1) and reactions were incubate at rt until blue color develops (1 – 30 min). Reactions were quenched by adding 50 µL 2 N H_2SO_4 and pipetting to mix. Signal was measured at A_{450} on a plate reader.

116

4.6.10. Mass spectrometry

Glycosylation reactions (in 25 mM Tris pH 7.5, 10 mM $CaCl_2$, 10 mM $MnCl_2$, 50 μL final reaction volume) containing ppGalNAcT2 (25 μL), saturating UDP-sugar (500 μM), and EA2 acceptor peptide (50 μM) proceeded for 2 h at 37 °C. Note that ppGalNAcT2 proteins were purified with some departures from the protocol described in Chapter 2.4.12. The modifications include increased α-FLAG resin (1.8 mL bed), only elutions 2-3 (3.6 mL total) were carried forward, and the purified enzyme was stored in 10% glycerol. At the time of these experiments, protein quantitation had not been optimized, and 25 μL purified ppGalNAcT was used in these glycosylation reactions. Reactions were quenched with 50 μL 2% formic acid, and % glycopeptide was analyzed by LC-MS with an Orbitrap mass spectrometer.

The mass spectrometry results were verified for some samples using the LC configuration and similar methods as described by Palaniappan *et al*. Samples were subjected to reverse phase chromatography with an Agilent 1200 LC system that was connected in-line with an LTQ Orbitrap XL hybrid mass spectrometer. External mass calibration was performed prior to analysis. A binary solvent system consisting of buffer A (0.1% formic acid in water (v/v)) and buffer B (0.1% formic acid in acetonitrile (v/v)) was employed. The LC was equipped with a 100 μL sample loop and the mass spectrometer was outfitted with a nanospray ionization source.

For each run, after 5 μL sample containing 100 pmol peptide was injected, sample trapping and desalting was performed for 5 min with 2% B onto a 100 μM fritted capillary (New Objective) pre-column self-packed with 1 cm of 5 μM 200 Å Magic C18AQ resin (Michrom Bioresources). The analyte was then loaded in 2% B onto a 100 μM fused silica capillary (Polymicro Technologies) self-packed with 5 cm of 5 μM 100 Å Magic C18AQ resin (Michrom Bioresources). A gradient was employed from 2% to 30% buffer B, followed by a washing step in 99% buffer B and re-equilibration in 2% buffer B. A solvent split was used to maintain a flow rate of 600 nL/ min at the column tip.

Mass spectra were recorded in a single stage of MS in positive ion mode over the *m/z* scan range of 400 to 2000 using the Orbitrap mass analyzer in full-scan, profile mode, at a resolution of 60,000 (at 400 *m/z*). Raw mass spectra were processed manually using Xcalibur (version 4.1, Thermo).

4.6.11. Cell fractionation

Human Embryonic Kidney (HEK) were cultured as described in the General Methods. HEK 293T cells were grown to confluence, the media was removed, and the cells were harvested by scraping in 3 mL cold PBS. The conditioned media and cells were pelleted by centrifugation for 15 min at 3650 x g at 4 °C. The supernatant was discarded and the cell pellet was frozen at -20 °C. Cell pellets were thawed and fractionated according to the manufacturer's protocol. Protein concentrations in

cytoplasmic, nuclear, and membrane extracts were quantitated by measuring the absorbance at 280 nm and by BCA assay. Extracts were stored at –80 °C.

4.6.12. Glycosylation reactions with fractionated lysates

Reactions were performed in TBS (10 mM $MnCl_2$, 1x EDTA-free protease inhibitors) and contained 1.33 µg/µL lysate (20 µg/µL stock). Reactions contained saturating UDP-GalNAzMe(S) concentrations (250 µM) and purified ppGalNAcT2 was added to initiate glycosylation. Note that ppGalNAcT2 (WT and I235A/L310A) was purified with some departures from the protocol described in Chapter 2.4.12. The modifications include increased α-FLAG resin (1.8 mL bed), only elutions 2-3 (3.6 mL total) were carried forward, and the purified enzyme was stored in 10% glycerol. At the time of these experiments, protein quantitation had not been optimized, and 20-21 µL purified ppGalNAcT was used in these glycosylation reactions. Glycosylation reactions were carried out at 37 °C for >12 h and reactions were quenched by boiling.

4.6.12.1. CUAAC conditions

Alkyne-PEG-biotin (1 mM in DMSO) was diluted in TBS and added to the quenched glycosylation reaction. $CuSO_4$ (50 mM in H_2O) and TBTA (30 mM in DMSO) were premixed and added to the click reaction. A fresh stock of sodium ascorbate (60 mM in H_2O) was prepared and added to the click reaction. Reactions (25 µM alkyne-PEG-biotin, 1 mM $CuSO_4$, 100 µM TBTA, and 1.5 mM sodium ascorbate, 55 µL total in TBS) were mixed and proceeded for 1 h at rt. The click reaction was quenched with EDTA (10 mM final concentration).

4.6.13. Immunocytochemistry

Transfection with full-length ppGalNAcTs was carried out as described in 2.4.11. After 48 h protein expression, the media was aspired, and 1 µg/mL Hoechst DNA stain (2 mg/mL stock diluted in media) was added to cells. Cells were incubated at 37 °C for 20-30 min, the media was aspirated, and cells were washed with PBS. PBS was aspirated and cells were fixed in 4 % formaldehyde in PBS for 10 min. The cells were washed 3 x with PBS and permeabilized in PBST (0.5 % Triton X-100) for 1 min. Cells were washed 3 x with PBS and blocked with PBST/BSA (PBS, 0.1 % Triton X-100, 3 % BSA) for 20 min. Cells were incubated with anti-FLAG primary antibody (diluted 1:1000 in PBST/BSA) for 1 h. Cells were washed 3 x with PBST/BSA and incubated with goat anti-mouse-AF647 (diluted 1:1000 in PBST/BSA) for 45 min. Cells were washed 3 x with PBS and imaged.

4.6.14. Western blotting of fractionated lysates for subcellular markers

Transfection with full-length ppGalNAcTs was carried out as described in 2.4.11. Cells were fractionated as in 4.6.11. Fractionated lysates were quantitated by BCA assay according to the manufacturer's protocol. Cellular extracts (8.5 µg protein) were analyzed by SDS-PAGE as described in General Protocol 3 and by western

blotting as described in General Protocol 4. Cytoplasmic extracts were probed with α-HSP-90 (diluted 1:1000), membrane extracts were probed with anti-FLAG M2 (diluted 1:300), and nuclear extracts were probed with anti-nuclear pore complex (diluted 1:10,000), and secondary antibodies were diluted 1:10,000.

4.6.15. Cell labeling without addition of engineered proteins

Human Embryonic Kidney (HEK) were cultured as described in the General Methods. HEK 293T cells were split in five 10-cm plates in 10 mL of DMEM at density of approximately 1-3 x 10^5 cells/mL. After 24 h, 50 μL of 10mM sugar analog stocks in DMSO or DMSO controls were added to a final concentration of 50 μM sugar analog. Samples were Ac$_4$GalNAzMe, Ac$_3$GalNAzMe-1-P(SATE)$_2$, Ac$_4$GalNAz, and DMSO (2 plates). After 48 h, the same quantity of stock solution was added again, reaching a 100 μM final concentration of sugar analog. After 60 h, the media was removed by aspiration, and cells were washed with 10 mL of cold PBS. 5 mL of FACS buffer (PBS + 1% FBS) was added to each plate and DIBAC-Biotin (1 mM, 100 μL, final concentration 20 μM) was added.

After 1 h, cells were harvested by scraping the plate, transferred to a 15 mL tube, and centrifuged at 1000 rpm for 5 min. The supernatant was removed by aspiration and the cells were re-suspended in 1 mL of PBS containing 1 mM EDTA and transferred to a 1.5 mL tube. Cells were centrifuged at 600 x g for 5 min, the supernatant was discarded, and cells were washed twice more with PBS containing 1 mM EDTA. After the final wash, cells were re-suspended in 600 μL lysis buffer (25 mM Tris-Cl, 150 mM NaCl, 1% Triton X-100, 1x EDTA-free protease inhibitors). Cells were lysed by sonication on ice (1.5 minutes total, 5 seconds on, 5 seconds off, 3W initial output). After lysis, cells were centrifuged for 30 min at 16,200 x g, the supernatant was transferred to a new 1.5 mL tube, and the cell lysate was stored at -20 °C.

The lysate was analyzed by SDS-PAGE as described in General Protocol 3 (10 μg per lane, final sample volume 15 mL). The gel was subjected to western blotting as described in General Protocol 4, and the antibody was α-biotin-HRP (1:100,000 dilution).

4.6.16. Cloning AGX mutants

4.6.16.1. Site-directed mutagenesis to make F381G and F381G/F383G

AGX1-F383G in pIRESpuro3 (Amp^R) was generously provided by J. Kohler (UT Southwestern Medical Center)[114]. F381G and F381G/F383G were generated via site-directed mutagenesis from AGX1-F383G. PCR amplifications were carried out as in General Protocol 1, and primers are in Table 4-1. PCR reactions were digested with 1 μL (20 U) DpnI at 37 °C for 1 h, and duplicate reactions were combined (100 μL total) and purified with DNA Clean and Concentrator™-5 according to the manufacturer's protocol with modifications as described in the General Methods (10

μL elution). Bacteria were transformed and clones were isolated as in General Protocol 2. Clones were sequenced as in the General Methods.

Mutation	Primer ID	Sequence
F381G	AGX1_F381G_F	5'-CAAACCCAATGGAATAAAGATGGAAAAAGGTGTCTTTGACATCTTCCAG
	AGX1_F381G_R	5'-CAAAGACACCTTTTTCCATCTTTATTCCATTGGGTTTGTCTGGCTTAATTAAC
F381G/F383G	AGX1_F381_383G_F	5'-CAAACCCAATGGAATAAAGATGGAAAAAGGTGTCGGTGACATCTTCCAG
	AGX1_F381_383G_R	5'-CACCGACACCTTTTTCCATCTTTATTCCATTGGGTTTGTCTGGCTTAATTAAC

Table 4-1. Primers for site-directed mutagenesis of AGX1 mutants. PCR extensions were 1 min 45 sec extension and annealing temperatures were 55 **°C**.

4.6.16.2. Introduction of FLAG tag

A FLAG tag was introduced to AGX1-F3831G, F383G, and F381G/F383G (Amp^R). PCR amplifications were carried out as in General Protocol 1, and primers are in Table 4-2. PCR reactions were digested with 1 μL (20 U) DpnI at 37 °C for 1 h, and duplicate reactions were combined (100 μL total) and purified with DNA Clean and Concentrator™-5 according to the manufacturer's protocol with modifications as described in the General Methods (10 μL elution). Bacteria were transformed and clones were isolated as in General Protocol 2. Clones were sequenced as in the General Methods.

Primer ID	Sequence	Annealing Temp (°C)
AGX1_FLAG_F	5'-GACTACAAAGACGATGACGACAAGTGAGCGGCCGCATAGATAACTGATCC	53
AGX1_FLAG_R	5'-GAGAATGGAGTTCATGAGCTGGTGAAAAATGGTATTGACTACAAAGACGATGACGACAAG	53

Table 4-2. Primers to introduce a FLAG tag into AGX1 constructs. PCR extensions were 1 min 45 sec extension and annealing temperatures were 55 **°C**.

4.6.16.3. Site-directed mutagenesis to make F381A, F383A, F381A/F383A, and WT

F381A, F383A, F381A/F383A, and WT were generated via site-directed mutagenesis from FLAG-tagged AGX1 clones in pIRESpuro3 (Amp^R) described above. PCR amplifications were carried out as in General Protocol 1, and primers are in . PCR reactions were digested with 1 μL (20 U) DpnI at 37 °C for 1 h, and duplicate reactions were combined (100 μL total) and purified with DNA Clean and Concentrator™-5 according to the manufacturer's protocol with modifications as described in the General Methods (10 μL elution). Bacteria were transformed and clones were isolated as in General Protocol 2. Clones were sequenced as in the General Methods.

Mutation	Primer ID	Sequence
F381A	AGX1_ F381A_ F	5'- GGAAAAAGCTGTCTTTGACATCTTCCAGTTTGC
	AGX1_ F381A_ R	5'- GATGTCAAAGACAGCTTTTTCCATCTTTATTCCATTGGG
F383A	AGX1_ F383A_ F	5'- GAATAAAGATGGAAAAATTTGTCGCTGACATCTTCCAGTTTGC
	AGX1_ F383A_ R	5'- GATGTCAGCGACAAATTTTTCCATCTTTATTCCATTGGGTTTG
F381A/ F383A	AGX1_ F381_3 83A_F	5'- GATGGAAAAAGCTGTCGCTGACATCTTCCAGTTTGC
	AGX1_ F381_3 83A_R	5'- GATGTCAGCGACAGCTTTTTCCATCTTTATTCCATTGGG
WT	AGX1_ WT_ F	5'- CCCAATGGAATAAAGATGGAAAAATTTGTCTTTGACATCTTCC
	AGX1_ WT_ R	5'- GACAAATTTTTCCATCTTTATTCCATTGGGTTTGTCTGGC

Table 4-3. Primers for site-directed mutagenesis of AGX1 mutants. PCR extensions were 1 min 45 sec extension and annealing temperatures were 55 **°C**.

4.6.17. HPAEC of AGX transfected cells

Confluent 10-cm plates (4) were used to generate a 12 mL single-cell suspension. Nine 20-cm plates with 20 mL DMEM (10% FBS, Pen/Strep) were seeded with 1.2 mL of cell suspension at density of approximately $1\text{-}3 \times 10^5$ cells/mL. To obtain sufficient amounts of plasmid DNA for mammalian cell transfection, a Midiprep kit (Qiagen) was used according to the manufacturer's protocol (DNA pellets were solubilized in water). After 36 h when cells were ~70% confluent, cells were

transfected with DNA using *Trans*IT®-293 reagent according to the manufacturer's protocol. 37.5 µg AGX plasmid DNA (134 µL of 0.28 µg/ lug stocks) was added to 3.75 mL Opti-Mem and the tubes were flicked to mix. Transit-293 reagent was vortexed vigorously, 112.5 µL was added to each tube, and the mixture was flicked to mix. The transfection mixture was incubated at rt for 25 min, added gently to the plates, and the plates were rocked to mix.

After cells were grown for 12 h, the media was replaced. After an additional 24 h, Ac$_3$GalNAzMe-1-P(SATE)$_2$ (200 µL of 10 mM solution in DMSO, 100 µM final concentration) was added to the plates transfected with AGX plasmids, Ac$_4$GalNAz (200 µL, 10 mM solution in DMSO, 100 µM final concentration) was added to 1 un-transfected control, and DMSO (200 µL) was added to 1 un-transfected control. Cells were grown for 6 h, the media was removed by aspiration, and cells were washed 2x with 20 mL cold PBS. PBS containing 1 mM EDTA (8 mL) was added and cells were harvested by scraping. The cells were pipetted thoroughly to generate a single cell suspension, transferred to a 15 mL tube, centrifuged (1000 rpm, 5 min), and the supernatant was discarded. Cells were re-suspended in PBS containing 1 mM EDTA (1 mL) and transferred to a 1.5 mL tube. Cells were centrifuged (600 x g, 5 min), supernatant was discarded, and cells were re-suspended in 1 mL PBS containing 1 mM EDTA.

200 µL of the cell suspension was set aside for cell lysis and the remaining cell suspension was transferred to O-ring tubes, centrifuged (600 x g, 5 min), and the supernatant was discarded. A layer of beads (0.1 mm zirconia/silicon beads, equal volume to the packed cell pellet) was added to the cell pellet, and 1 mL of ACN:H$_2$O (1:1) was added. The samples were lysed in a bead beater (30 sec, 6 m/s) and subsequently incubated on ice for 10 min. The resulting suspension was centrifuged at 16.2k x g for 10 min. The supernatant was transferred to a new 1.5 mL tube and solvent was removed by speed-vac (50 °C, 2 h). The pellet was re-suspended in 400 µL LC-MS grade water and ADP-Glucose (5 mM, 1.2 µL) was added as an internal standard. Samples were transferred to 10 kDa cut-off filters and centrifuged for 30 min at 14k x g. The filtrate was collected and analyzed by HPAEC as in General Protocol 5.

The 200 µL cell suspension that was set aside earlier was centrifuged (600 x g, 5 min), the supernatant was discarded, and the cell pellet was resuspended in 300 µL lysis buffer (25 mM Tris-Cl pH 7.4, 150 mM NaCl, 1% Triton X-100, 1x EDTA-free protease inhibitors). Cells were lysed by sonication (1.5 min total lysis time, 5 sec on, 5 sec off, 3W initial output). The resulting lysate was centrifuged (16.2k x g, 30 min) and the supernatant was collected.

The lysate was analyzed by SDS-PAGE as described in General Protocol 3 (20 µg per lane, final sample volume 45 µL). The gel was subjected to western blotting as described in General Protocol 4 with the following modifications: the blot was blocked with 5% non-fat powdered milk in TBST, the primary antibody was anti-

AGX1 (1:2000 dilution in 3 % milk/TBST) and secondary antibody was goat-α-rabbit-HRP (1:5000 dilution in 3 % milk/TBST).
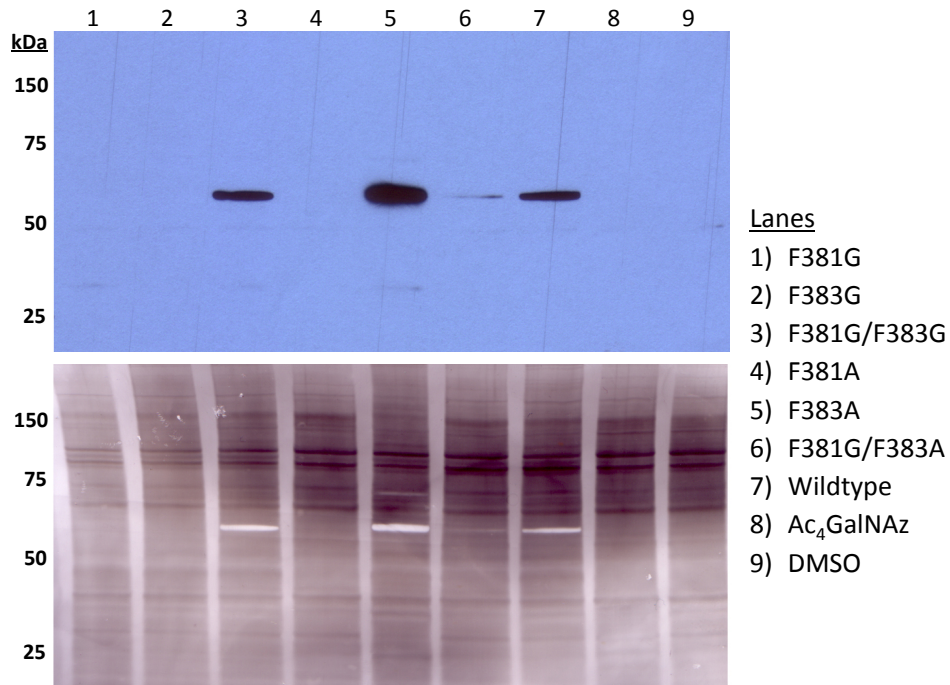
Figure 4-24. Cell lysates of cells transfected with AGX1 and fed Ac₃GalNAzMe(S)-1P(SATE)₂. α-biotin blot (top pane) and India Ink total protein stain (bottom pane). Lanes 1-7: cells transfected with WT or mutant AGX1 and fed Ac₃GalNAzMe(S)-1P(SATE)₂. Lanes 8-9: untransfected cells treated with Ac₄GalNAz (positive control) or DMSO (negative control).

4.6.18. Synthesis
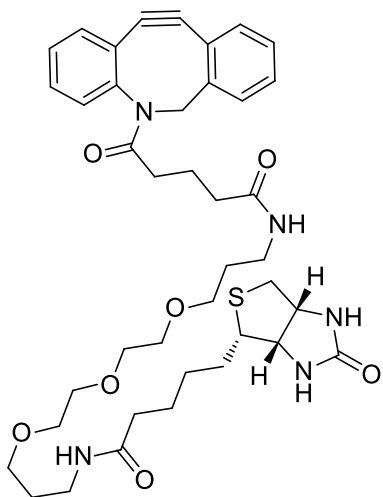
DIBAC-Biotin (5)

DIBAC-NHS [published in Debets *et al.*[218]] (20 mg, 0.05 mmol) was added to a flask under N_2-atmosphere. Dry DMF (3 mL) was added and the suspension was heated slightly to dissolve DIBAC-NHS. Next, NEt_3 (13 μL, 0.1 mmol) and Biotin-NH_2 (21 mg, 0.05 mmol) were added and the mixture was stirred overnight. DMF was evaporated and the crude product was purified by gradient column chromatography (CH_2Cl_2/MeOH, 1:0 to 9:1) to obtain DIBAC-biotin as a white solid. ^{1}H-NMR (400 MHz, $CDCl_3$): δ 7.67 (d, J = 7.3 Hz, 1H), 7.52-7.08 (m, 7H), 6.75 (br s, 1H), 6.40-6.27 (m, 1H), 6.21 (d, J = 7.8 Hz, 1H), 5.55 (d, J = 11.5 Hz, 1H), 5.14 (d, J = 13.8 Hz, 1H), 4.64-4.41 (m, 1H), 4.29-4.26 (m, 1H), 3.73-3.44 (m, 12H), 3.38-3.05 (m, 4H), 2.86 (td, J = 12.3, 4.9 Hz, 1H), 2.69 (dd, J = 23.2, 12.7 Hz, 1H), 2.33-2.09 (m, 4H), 2.03-1.84 (m, 3H), 1.81-1.58 (m, 11H), 1.53-1.36 (m, 2H). ^{13}C-NMR (151 MHz, $CDCl_3$) δ 173.1, 172.9, 172.5, 163.6, 151.6, 148.1, 132.3, 129.0, 128.6, 128.2, 128.1, 127.7, 127.1, 125.4, 123.0, 122.4, 114.8, 108.0, 70.4 (2C), 70.0, 69.9, 69.4, 61.8, 60.1, 55.6, 55.4 (2C), 40.5, 37.7, 37.2, 35.9, 35.5, 35.4, 33.8, 29.2, 28.9, 28.1, 28.0, 25.6, 21.6. HRMS (ESI+) m/z calcd for $C_{40}H_{54}N_5O_7S$ [M+H]$^+$ 748.3744, found 748.3764.
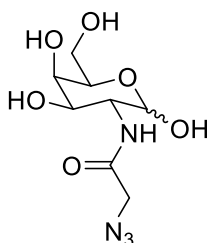
2-azido-N-((3R,4R,5R,6R)-2,4,5-trihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-3-yl)acetamide (3.2, GalNAz) Crude azidoacetic acid (2.6 g, 23 mmol) was dissolved in dry MeOH (140 mL) under N_2. Galactosamine (**3.1**, 3.0 g, 14 mmol) and NEt_3 (4.8 mL, 35 mmol) were added whereupon the reaction mixture was cooled to 0 °C. EDC (5.4 g, 35 mmol) and HOBt (1.9 g, 14 mmol) were added and the reaction was stirred overnight at room temperature. The solvents were removed under reduced pressure and the crude product was purified by gradient column chromatography (CH_2Cl_2/MeOH, 1:9 to 3:7). The impure isolated sugar was redissolve in little MeOH (20 mL) and CH_2Cl_2 (800 mL) was added. The mixture was kept at -20 °C for 3 hours whereupon the product was isolated by filtration as a white solid (2.0 g, 54%). ^{1}H-NMR (400 MHz, CD_3OD) δ: 5.50 (s, 1H), 5.14 (d, J = 3.6 Hz, 1H), 4.25 (dd, J = 10.9, 3.7 Hz, 1H), 4.04 (t, J = 6.1 Hz, 1H), 3.92 (dd, J = 10.6, 3.5 Hz, 3H), 3.82 (dd, J = 10.9, 3.2 Hz, 1H), 3.72 – 3.64 (m, 2H). Analysis was in accordance with literature precedent[219].
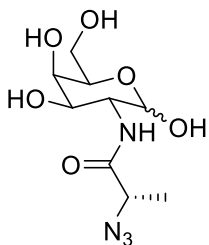
(S)-2-azido-N-((3R,4R,5R,6R)-2,4,5-trihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-3-yl)propanamide
Crude azidoalanine (6 g, 50 mmol) was dissolved in dry MeOH (200 mL) under N_2. Galactosamine (**3.1**, 5.42 g, 25 mmol) and NEt_3 (8.4 mL, 60 mmol) were added whereupon the reaction mixture was cooled to 0 °C. EDC (9.3 g, 60 mmol) and HOBt (3.38 g, 25 mmol) were added and the reaction was stirred overnight at room temperature. The solvents were removed under reduced pressure and the crude product was purified by gradient column chromatography (CH_2Cl_2/MeOH, 1:9 to 3:7). The impure isolated sugar was redissolve in little MeOH (20 mL) and CH_2Cl_2 (800 mL) was added. The mixture was kept at -20 °C for 3 hours whereupon the product was isolated by filtration as a white solid (4.0 g, 58%). ^{1}H-NMR (400 MHz, CD_3OD) δ: 5.14 (d, J = 3.6 Hz, 1H, NH), 4.60 (app. s, 1H, C_1H), 4.21 (dd, J = 10.9, 3.6 Hz, 1H, C_3H), 4.09 – 3.94 (m, 2H, C_5H, C_2H), 3.90 (d, J = 2.6 Hz, 1H, C_4H), 3.82 (dd, J = 10.8, 3.2 Hz, 1H, C_6HH), 3.77 – 3.64 (m, 2H, C_6HH, COCHCH$_3$), 1.46 (d, J = 6.9 Hz, 3H, COCHCH_3). ^{13}C-NMR (151 MHz, MeOD) δ: 173.5, 92.8, 71.7, 70.6, 69.6, 62.8, 59.5, 52.0, 17.3. HRMS (ESI-) m/z calcd for $C_9H_{15}O_6N_4$ [M-H]$^-$ 275.0997, found 275.0997.
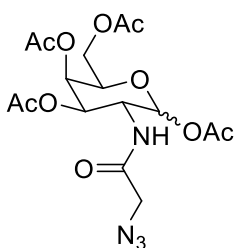
(3R,4R,5R,6R)-6-(acetoxymethyl)-3-(2-azidoacetamido)tetrahydro-2H-pyran-2,4,5-triyl triacetate
3.2 (1.57 g, 6 mmol) was dissolved in pyridine (30 mL). The solution was cooled to 0°C and Ac_2O (17 mL, 180 mmol) was added. The reaction was stirred overnight and subsequently diluted with H_2O (200 mL) and EtOAc (200 mL). The layers were separated, and the with sat. aq. $CuSO_4.5H_2O$ (4 × 200 mL), H_2O (200 mL), aqueous EDTA (50 mM, 2 × 100 mL), H_2O (200 mL) and brine (100 mL). The organic layer was dried over $MgSO_4$ and the solvents removed under reduced pressure. The crude product was dissolved in EtOAc (20 mL) and hexanes (800 mL) were added. After three hours at -20 °C, the product was isolated by filtration as a white solid (2.04 g, 80%). Prior to using the compound for cell feeding, it was further purified by HPLC purification. ^{1}H-NMR (400 MHz, $CDCl_3$) δ: 6.28 (d, J = 9.2 Hz, 1H), 6.23 (d, J = 3.7 Hz, 1H), 5.45 (dd, J = 3.3, 1.4 Hz, 1H), 5.26 (dd, J = 11.5, 3.2 Hz, 1H), 4.70 (ddd, J = 11.5, 9.2, 3.7 Hz, 1H), 4.33 – 4.22 (m, 1H), 4.16 – 4.03 (m, 2H), 3.96 (s, 2H), 2.19 (s, 3H), 2.17 (s, 3H), 2.04 (s, 3H), 2.03 (s, 3H). Analysis was in accordance with literature precedent[46].
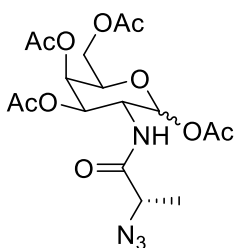
(3R,4R,5R,6R)-6-(acetoxymethyl)-3-((S)-2-azidopropanamido)tetrahydro-2H-pyran-2,4,5-triyl triacetate GalNAzMe (4 g, 14.5 mmol) was dissolved in pyridine (73 mL). The solution was cooled to 0°C and Ac_2O (41 mL, 435 mmol) was added. The reaction was stirred overnight and subsequently diluted with H_2O (200 mL) and EtOAc (200 mL). The layers were separated, and the with sat. aq. $CuSO_4.5H_2O$ (4 × 200 mL), H_2O (200 mL), aqueous EDTA (50 mM, 2 × 100 mL), H_2O (200 mL) and brine (100 mL). The organic layer was dried over $MgSO_4$ and the solvents removed under reduced pressure. The crude product was dissolved in EtOAc (20 mL) and

hexanes (800 mL) were added. After three hours at -20 °C, the product was isolated by filtration as a white solid (4.2 g, 65%). Prior to using the compound for cell feeding, it was further purified by HPLC purification. ^{1}H-NMR (400 MHz, CDCl$_3$) δ: 6.29 (d, J = 9.0 Hz, 1H, NH), 6.25 (d, J = 3.8 Hz, 1H, C$_1$$H$), 5.45 (d, J = 3.1 Hz, 1H, C$_4$$H$), 5.25 (dd, J = 11.5, 3.2 Hz, 1H, C$_3$$H$), 4.65 (ddd, J = 12.0, 9.1, 3.7 Hz, 1H, C$_5$$H$), 4.27 (t, J = 6.7 Hz, 1H, C$_6$HH), 4.19 – 4.06 (m, 2H, C$_6$$H$H + C$_2$$H$), 4.03 (q, J = 7.2 Hz, 1H, COCHCH$_3$), 2.19 (s, 3H, COCH_3), 2.17 (s, 3H, COCH_3), 2.04 (s, 3H, COCH_3), 2.03 (s, 3H, COCH_3), 1.48 (d, J = 7.1 Hz, 3H, COCHCH_3). 13C-NMR (101 MHz, CDCl$_3$) δ: 170.9 (CO), 170.3 (CO), 170.2 (CO), 170.1 (CO), 168.8 (CO), 90.8 (C_1), 68.7 (C_6), 67.6 (C_4), 66.6 (C_3), 61.2 (C_5), 59.1 (COCHCH$_3$N$_3$), 47.1 (C_2), 20.8 (COC, 20.7 (COC, 20.6 (2C, 2 × COC, 17.1 (COCHCH$_3$N$_3$). HRMS (ESI+) m/z calcd for C$_{17}$H$_{24}$O$_{10}$N$_4$Na [M+Na]$^+$ 467.1385, found 467.1379.
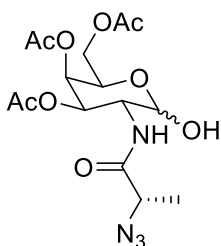
(2R,3R,4R,5R)-2-(acetoxymethyl)-5-((S)-2-azidopropanamido)-6-hydroxytetrahydro-2H-pyran-3,4-diyl diacetate Ac$_4$GalNAzMe (444 mg, 1 mmol) was dissolved in THF (4 mL). Benzylamine (218 μL, 2 mmol) was added and the reaction mixture was stirred overnight at room temperature. The solvents were removed under reduced pressure and the crude product was purified by gradient column chromatography (EtOAc/Hexanes, 1:2 to 1:1) to obtain the product as a yellow oil (280 mg, 69%). ^{1}H-NMR (400 MHz, CDCl$_3$) δ: 6.62 (d, J = 9.5 Hz, 1H, NH), 5.37 (d, J = 3.2 Hz, 1H, C$_1$$H$), 5.31 – 5.16 (m, 2H, C$_3$), 4.54 – 4.27 (m, 2H, C$_5$$H$, C$_6$$H$H), 4.06 (tt, J = 11.2, 6.8 Hz, 2H, C$_6$HH, C$_2$$H$), 3.97 (q, J = 7.0 Hz, 1H, COCHCH$_3$), 2.13 (s, 3H, COCH_3), 2.02 (s, 3H, COCH_3), 1.96 (s, 3H, COCH_3), 1.46 (d, J = 7.0 Hz, 3H, COCHCH_3). ^{13}C-NMR (101 MHz, CDCl3) δ: 170.8 (CO), 170.7 (CO), 170.5 (CO), 170.3 (CO), 91.9 (C_1), 68.0 (C_6), 67.4 (C_4), 66.4 (C_3), 61.9 (C_5), 58.9 (COCHCH$_3$), 48.1 (C$_2$), 20.7 (3C, COCH$_3$), 17.0 (COCHCH$_3$). HRMS (ESI+) m/z calcd for C$_{15}$H$_{22}$O$_9$N$_4$Na [M+Na]$^+$ 425.1279, found 425.1275.
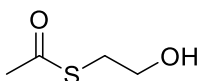
S-(2-hydroxyethyl) ethanethioate Thioacetic acid (4 mL, 57.5 mmol) was added to toluene (23 mL). The mixture was cooled to 0 °C and DBU (8.0 mL, 57.5 mmol) was dropwise over 30 minutes at 0 °C. During the addition, drops of H$_2$O were added when clumps of salts prevented stirring of the reaction. After complete DBU addition, iodoethanol (3.9 mL, 50 mmol) was added dropwise over 30 minutes at 0 °C. The resulting mixture was stirred for 3 hours at room temperature whereupon EtOAc (200 mL) and H$_2$O (200 mL) were added. The layers were separated and the organic layer was washed with H$_2$O (3 × 200 mL) and brine (200 mL) and the organic layer was dried over MgSO$_4$. The solvents were removed under reduced pressure and the crude product was purified by gradient column chromatography (CH$_2$Cl$_2$/EtOAc, 9:1 to 1:1) and the product was obtained as a red oil (2.15 g, 31%). ^{1}H NMR (400 MHz, CDCl$_3$) δ: 3.77 (t, J = 6.0 Hz, 2H), 3.09 (t, J = 6.0 Hz, 2H), 2.37 (s, 3H). Analysis was in accordance with literature precedent[195].
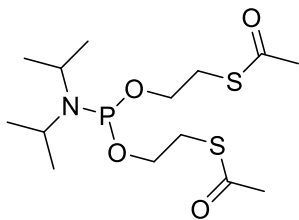
S,S'-(((((diisopropylamino)phosphanediyl) bis(oxy))bis(ethane-2,1-diyl)) diethanethioate Two separate flasks were heat-dried, evacuated and refilled with N₂-gas. In one flask, PCl₂N^{i}Pr₂ (921 µL, 5 mmol) was dissolved in dry THF (35 mL) and cooled to -78 °C. To this solution a solution of S-(2-hydroxyethyl) ethanethioate (1.2 g, 10 mmol) and NEt₃ (3 mL, 22 mmol) in dry THF (12.5 mL) was added dropwise over 1 hour. The reaction mixture was stirred for 2 hours at room temperature, whereupon the solids were filtered of. The mixture was concentrated in vacuo and the crude product was purified by column chromatography to yield a colorless oil (150 mg, 7.8%). ^{1}H-NMR (400 MHz, CDCl₃) δ: 3.90 – 3.45 (m, 6H), 3.13 (t, J = 6.5 Hz, 4H), 2.35 (s, 6H), 1.18 (d, J = 6.8 Hz, 12H). Analysis was in accordance with literature precedent[195].
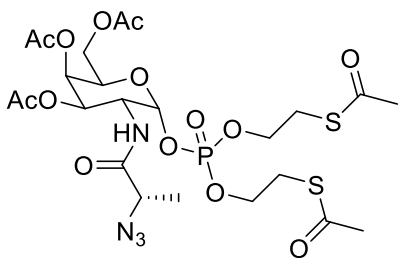
(2R,3R,4R,5R,6R)-2-(acetoxymethyl)-5-((S)-2-azidopropanamido)-6-((bis(2-(acetylthio)ethoxy)phosphoryl)oxy)tetrahydro-2H-pyran-3,4-diyl diacetate Ac₃GalNAzMe (112 mg, 0.28 mmol) and phosphoramidite **x** (150 mg, 0.39 mmol) were dissolved in dry CH₃CN (1.5 mL) in a dry flask under nitrogen-gas. The mixture was cooled to 0 °C and tetrazole (29.4 mg, 0.42 mmol) was added. The reaction was warmed to room temperature and stirred for one hour. Upon cooling to 0 °C, mCPBA (94 mg, 0.42 mmol) was added and the mixture stirred for an additional 30 minutes at room temperature. The reaction was diluted with EtOAc (10 mL) and the organic layer was washed with 10% aqueous Na₂SO₄ (2 × 10 mL), sat. aq. NaHCO₃ (2 × 30 mL) and brine (50 mL). After every wash-step the aqueous layer was back-extracted with EtOAc (10 mL) and the organic layers were combined prior to further washing. The organic layer was dried over MgSO₄ and concentrated in vacuo. The crude product was purified by gradient column chromatography (Hexanes: EtOAc, 1:1 to 0:1) to obtain the product as a clear oil (70 mg, 41%). Prior to using the compound for cell feeding, it was further purified by HPLC purification. ^{1}H-NMR (400 MHz, CDCl₃) δ: 6.74 (d, J = 9.1 Hz, 1H, NH), 5.81 (dd, J = 5.7, 3.3 Hz, 1H, C₄H), 5.50 (dd, J = 3.2, 1.4 Hz, 1H, C₁H), 5.28 (dd, J = 11.6, 3.1 Hz, 1H, C₃), 4.66 (ddt, J = 12.2, 9.1, 3.3 Hz, 1H, C₂H), 4.49 – 4.40 (m, 1H, C₆HH), 4.32 – 4.10 (m, 6H, C₆HH, C₅H, POCH_2), 4.06 (q, J = 7.0 Hz, 1H, COCHCH₃N₃), 3.34 – 3.10 (m, 4H, POCH₂CH_2), 2.40 (s, 3H, SCOCH_3), 2.39 (s, 3H, SCOCH_3), 2.19 (s, 3H, COCH_3), 2.07 (s, 3H, COCH_3), 2.03 (s, 3H, COCH_3), 1.54 (d, J = 7.1 Hz, 3H, COCHCH_3N₃). ^{13}C-NMR (100 MHz, CDCl3) δ: 195.0 (SCOCH₃), 194.7 ((SCOCH₃), 170.6 (CO), 170.5 (CO), 170.2 (CO), 170.0 (CO), 96.62 (d, J = 6.1 Hz, C_1), 68.7 (C_6), 67.1 (C_4), 66.6 (C_3), 66.5 (d, J = 5.8 Hz, POCH₂CH₂), 66.35 (d, J = 5.9 Hz, POCH₂CH₂), 61.3 (C_5), 59.0 (COCCH₃N₃), 47.8 (C_2), 30.5 (2C, SCOCH₃), 29.00 (d, J = 7.4 Hz, 2C, POCH₂CH₂), 20.6 (3C, COCH₃), 17.2 (COCHCH₃N₃). ^{31}P-NMR (162 MHz, CDCl3) δ: -3.01. Peak assignment in ^{1}H-NMR and ^{13}C-NMR based on COSY and HSQC. HRMS (ESI+) m/z calcd for C₂₃H₃₅O₁₄N₄NaPS₂ [M+Na]$^+$ 709.1221, found 709.1221.

References

(1)   Apweiler, R.; Hermjakob, H.; Sharon, N. On the Frequency of Protein Glycosylation, as Deduced From Analysis of the SWISS-PROT Database. *Biochim. Biophys. Acta.* **1999**, *1473*, 4–8.

(2)   Hart, G. W.; Copeland, R. J. Glycomics Hits the Big Time. *Cell* **2010**, *143*, 672–676.

(3)   Khoury, G. A.; Baliban, R. C.; Floudas, C. A. Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of the Swiss-Prot Database. *Sci Rep* **2011**, *1*.

(4)   Hashimoto, K.; Tokimatsu, T.; Kawano, S.; Yoshizawa, A. C.; Okuda, S.; Goto, S.; Kanehisa, M. Comprehensive Analysis of Glycosyltransferases in Eukaryotic Genomes for Structural and Functional Characterization of Glycans. *Carbohydr. Res.* **2009**, *344*, 881–887.

(5)   Tabak, L. A. The Role of Mucin-Type O-Glycans in Eukaryotic Development. *Seminars in Cell & Developmental Biology* **2010**, *21*, 616–621.

(6)   Hang, H. C. Chemical Approaches to the Study of Mucin-Type O-Linked Glycosylation. *Dept. of Chemistry, UC Berkeley* **2003**, PhDDissertation.

(7)   Stanley, P. Golgi Glycosylation. *Cold Spring Harb Perspect Biol* **2011**, *3*.

(8)   Joshi, H. J.; Steentoft, C.; Schjoldager, K. T.-B. G.; Vakhrushev, S. Y.; Wandall, H. H.; Clausen, H. Protein O-GalNAc Glycosylation: Most Complex and Differentially Regulated PTM. In *Glycoscience: Biology and Medicine*; Taniguchi, N., Endo, T., Hart, G. W., Seeberger, P. H., Wong, C.-H., Eds.; Springer Japan: Tokyo, 2014; pp 1049–1064.

(9)   Van den Steen, P.; Rudd, P. M.; Dwek, R. A.; Opdenakker, G. Concepts and Principles of O-Linked Glycosylation. *Critical Reviews in Biochemistry and Molecular Biology* **1998**, *33*, 151–208.

(10)   Tian, E.; Hagen, Ten, K. G. Recent Insights Into the Biological Roles of Mucin-Type O-Glycosylation. *Glycoconjugate J* **2008**, *26*, 325–334.

(11)   Hang, H. C.; Yu, C.; Kato, D. L.; Bertozzi, C. R. A Metabolic Labeling Approach Toward Proteomic Analysis of Mucin-Type O-Linked Glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14846–14851.

(12)   Hagen, Ten, K. G.; Fritz, T. A.; Tabak, L. A. All in the Family: the UDP-GalNAc:Polypeptide N-Acetylgalactosaminyltransferases. *Glycobiology* **2003**, *13*, 1R–16R.

(13)   Pratt, M. R.; Hang, H. C.; Hagen, Ten, K. G.; Rarick, J.; Gerken, T. A.; Tabak, L. A.; Bertozzi, C. R. Deconvoluting the Functions of Polypeptide N-Alpha-Acetylgalactosaminyltransferase Family Members by Glycopeptide Substrate Profiling. *Chem. Biol.* **2004**, *11*, 1009–1016.

(14)   Gerken, T. A.; Jamison, O.; Perrine, C. L.; Collette, J. C.; Moinova, H.; Ravi, L.; Markowitz, S. D.; Shen, W.; Patel, H.; Tabak, L. A. Emerging Paradigms for the Initiation of Mucin-Type Protein O-Glycosylation by the Polypeptide GalNAc Transferase Family of Glycosyltransferases. *Journal of Biological Chemistry* **2011**, *286*, 14493–14507.

(15)   Allen, J. J.; Li, M.; Brinkworth, C. S.; Paulson, J. L.; Wang, D.; Hübner, A.; Chou, W.-H.; Davis, R. J.; Burlingame, A. L.; Messing, R. O.; et al. A Semisynthetic Epitope for Kinase Substrates. *Nat Meth* **2007**, *4*, 511–516.

(16)   Mann, M.; Jensen, O. N. Proteomic Analysis of Post-Translational Modifications. *Nat. Biotechnol.* **2003**, *21*, 255–261.

(17)   Laughlin, S. T.; Baskin, J. M.; Amacher, S. L.; Bertozzi, C. R. In Vivo Imaging of Membrane-Associated Glycans in Developing Zebrafish. *Science* **2008**, *320*, 664–667.

(18)   Zeidan, Q.; Hart, G. W. The Intersections Between O-GlcNAcylation and Phosphorylation: Implications for Multiple Signaling Pathways. *J. Cell Sci.* **2010**, *123*, 13–22.

(19)   Szymanski, C. M.; Wren, B. W. Protein Glycosylation in Bacterial Mucosal Pathogens. *Nat Rev Micro* **2005**, *3*, 225–237.

(20)   Laughlin, S. T.; Bertozzi, C. R. Imaging the Glycome. *Proceedings of the National Academy of Sciences* **2009**, *106*, 12–17.

(21)   Hang, H. C.; Bertozzi, C. R. The Chemistry and Biology of Mucin-Type O-Linked Glycosylation. *Bioorg. Med. Chem.* **2005**, *13*, 5021–5034.

(22)   Gottschalk, A.; Graham, E. R. B. : The Neuraminidase-Susceptible Prosthetic Group of Bovine Salivary Mucoprotein. *Biochimica et Biophysica Acta* **1959**, *34*, 380–391.

(23)   Pigman, W.; Moschera, J.; Weiss, M.; Tettamanti, G. The Occurrence of Repetitive Glycopeptide Sequences in Bovine Submaxillary Glycoprotein. *Eur. J. Biochem.* **1973**, *32*, 148–154.

(24)   Strous, G. J.; Dekker, J. Mucin-Type Glycoproteins. *http://dx.doi.org/10.3109/10409239209082559* **1992**, *27*, 57–92.

(25)   Jensen, P. H.; Kolarich, D.; Packer, N. H. Mucin-Type O-Glycosylation – Putting the Pieces Together. *FEBS Journal* **2010**, *277*, 81–94.

(26)   Desseyn, J.-L.; Tetaert, D.; Gouyer, V. Architecture of the Large Membrane-Bound Mucins. *Gene* **2008**, *410*, 215–222.

(27)   Carraway, K. L.; Hull, S. R. Cell Surface Mucin-Type Glycoproteins and Mucin-Like Domains. *Glycobiology* **1991**, *1*, 131–138.

(28)   Dekker, J.; Rossen, J. W. A.; Büller, H. A.; Einerhand, A. W. C. The MUC Family: an Obituary. *Trends in Biochemical Sciences* **2002**, *27*, 126–131.

(29)   Clausen, H.; Bennett, E. P. A Family of UDP-GalNAc: Polypeptide N-Acetylgalactosaminyl-Transferases Control the Initiation of Mucin-Type O-Linked Glycosylation. *Glycobiology* **1996**, *6*, 635–646.

(30)   Woo, C. M.; Iavarone, A. T.; Spiciarich, D. R.; Palaniappan, K. K.; Bertozzi, C. R. Isotope-Targeted Glycoproteomics (IsoTaG): a Mass-Independent Platform for Intact N- and O-Glycopeptide Discovery and Analysis. *Nat Meth* **2015**.

(31)   Steentoft, C.; Vakhrushev, S. Y.; Vester-Christensen, M. B.; Schjoldager, K. T.-B. G.; Kong, Y.; Bennett, E. P.; Mandel, U.; Wandall, H.; Levery, S. B.; Clausen, H. Mining the O-Glycoproteome Using Zinc-Finger Nuclease–Glycoengineered SimpleCell Lines. *Nat Meth* **2011**, *8*, 977–982.

(32)   Vakhrushev, S. Y.; Steentoft, C.; Vester-Christensen, M. B.; Bennett, E. P.; Clausen, H.; Levery, S. B. Enhanced Mass Spectrometric Mapping of the

Human GalNAc-Type O-Glycoproteome with SimpleCells. *Mol Cell Proteomics* **2013**, *12*, 932–944.

(33)   Yang, Z.; Halim, A.; Narimatsu, Y.; Jitendra Joshi, H.; Steentoft, C.; Schjoldager, K. T.-B. G.; Alder Schulz, M.; Sealover, N. R.; Kayser, K. J.; Paul Bennett, E.; et al. The GalNAc-Type O-Glycoproteome of CHO Cells Characterized by the SimpleCell Strategy. *Mol Cell Proteomics* **2014**, *13*, 3224–3235.

(34)   Brockhausen, I.; Schachter, H.; Stanley, P. *Essentials of Glycobiology - NCBI Bookshelf*, 2nd ed.; Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., Etzler, M. E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2009.

(35)   Su, D. M.; Eguchi, H.; Yi, W.; Li, L.; Wang, P. G.; Xia, C. Enzymatic Synthesis of Tumor-Associated Carbohydrate Antigen Globo-H Hexasaccharide. *Org. Lett.* **2008**, *10*, 1009–1012.

(36)   Brockhausen, I. Mucin-Type O-Glycans in Human Colon and Breast Cancer: Glycodynamics and Functions. *EMBO Rep.* **2006**, *7*, 599–604.

(37)   Blixt, O.; Bueti, D.; Burford, B.; Allen, D.; Julien, S.; Hollingsworth, M.; Gammerman, A.; Fentiman, I.; Taylor-Papadimitriou, J.; Burchell, J. M. Autoantibodies to Aberrantly Glycosylated MUC1 in Early Stage Breast Cancer Are Associated with a Better Prognosis. *Breast Cancer Res* **2011**, *13*, R25.

(38)   Grohmann, G. P.; Schirmacher, P.; Manzke, O.; Hanisch, F.-G.; Dienes, H. P.; Baldus, S. E. Modulation of MUC1 and Blood Group Antigen Expression in Gastric Adenocarcinoma Cells by Cytokines. *Cytokine* **2003**, *23*, 86–93.

(39)   Rambaruth, N. D. S.; Dwek, M. V. Cell Surface Glycan-Lectin Interactions in Tumor Metastasis. *Acta Histochem.* **2011**, *113*, 591–600.

(40)   Hong, J. C.; Kim, Y. S. Alkali-Catalyzed Beta-Elimination of Periodate-Oxidized Glycans: a Novel Method of Chemical Deglycosylation of Mucin Gene Products in Paraffin Embedded Sections. *Glycoconjugate J* **2000**, *17*, 691–703.

(41)   Park, J. H.; Nishidate, T.; Kijima, K.; Ohashi, T.; Takegawa, K.; Fujikane, T.; Hirata, K.; Nakamura, Y.; Katagiri, T. Critical Roles of Mucin 1 Glycosylation by Transactivated Polypeptide N-Acetylgalactosaminyltransferase 6 in Mammary Carcinogenesis. *Cancer Research* **2010**, *70*, 2759–2769.

(42)   Earl, L. A.; Bi, S.; Baum, L. G. N- and O-Glycans Modulate Galectin-1 Binding, CD45 Signaling, and T Cell Death. *Journal of Biological Chemistry* **2010**, *285*, 2232–2244.

(43)   Arii, J.; Wang, J.; Morimoto, T.; Suenaga, T.; Akashi, H.; Arase, H.; Kawaguchi, Y. A Single-Amino-Acid Substitution in Herpes Simplex Virus 1 Envelope Glycoprotein B at a Site Required for Binding to the Paired Immunoglobulin-Like Type 2 Receptor   (PILR) Abrogates PILR -Dependent Viral Entry and Reduces Pathogenesis. *Journal of Virology* **2010**, *84*, 10773–10783.

(44)   Wang, J.; Fan, Q.; Satoh, T.; Arii, J.; Lanier, L. L.; Spear, P. G.; Kawaguchi, Y.; Arase, H. Binding of Herpes Simplex Virus Glycoprotein B (gB) to Paired Immunoglobulin-Like Type 2 Receptor   Depends on Specific Sialylated O-

Linked Glycans on gB. *Journal of Virology* **2009**, *83*, 13042–13045.

(45) Lin, Y.-R.; Reddy, B. V. V. G.; Irvine, K. D. Requirement for a Core 1 Galactosyltransferase in the Drosophilanervous System. *Dev. Dyn.* **2008**, *237*, 3703–3714.

(46) Hang, H. C.; Yu, C.; Kato, D. L.; Bertozzi, C. R. A Metabolic Labeling Approach Toward Proteomic Analysis of Mucin-Type O-Linked Glycosylation. *Proceedings of the National Academy of Sciences* **2003**, *100*, 14846–14851.

(47) Agard, N. J.; Bertozzi, C. R. Chemical Approaches to Perturb, Profile, and Perceive Glycans. *Acc. Chem. Res.* **2009**, *42*, 788–797.

(48) Prescher, J. A.; Bertozzi, C. R. Chemistry in Living Systems. *Nat Chem Biol* **2005**, *1*, 13–21.

(49) Bishop, A.; Buzko, O.; Heyeck-Dumas, S.; Jung, I.; Kraybill, B.; Liu, Y.; Shah, K.; Ulrich, S.; Witucki, L.; Yang, F.; et al. Unnatural Ligands for Engineered Proteins: New Tools for Chemical Genetics. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 577–606.

(50) Alaimo, P. J.; Shogren-Knaak, M. A.; Shokat, K. M. Chemical Genetic Approaches for the Elucidation of Signaling Pathways. *Curr. Opin. Chem. Biol.* **2001**, *5*, 360–367.

(51) Gill, D.; Clausen, H. Location, Location, Location: New Insights Into O-GalNAc Protein Glycosylation. *Trends in cell biology* **2010**.

(52) Rottger, S.; White, J.; Wandall, H. H.; Olivo, J. C.; Stark, A.; Bennett, E. P.; Whitehouse, C.; Berger, E. G.; Clausen, H.; Nilsson, T. Localization of Three Human Polypeptide GalNAc-Transferases in HeLa Cells Suggests Initiation of O-Linked Glycosylation Throughout the Golgi Apparatus. *J. Cell Sci.* **1998**, *111 (Pt 1)*, 45–60.

(53) Perrine, C. L.; Ganguli, A.; Wu, P.; Bertozzi, C. R.; Fritz, T. A.; Raman, J.; Tabak, L. A.; Gerken, T. A. Glycopeptide-Preferring Polypeptide GalNAc Transferase 10 (ppGalNAc T10), Involved in Mucin-Type O-Glycosylation, Has a Unique GalNAc-O-Ser/Thr-Binding Site in Its Catalytic Domain Not Found in ppGalNAc T1 or T2. *J. Biol. Chem.* **2009**, *284*, 20387–20397.

(54) Gill, D. J.; Chia, J.; Senewiratne, J.; Bard, F. Regulation of O-Glycosylation Through Golgi-to-ER Relocation of Initiation Enzymes. *The Journal of Cell Biology* **2010**, *189*, 843–858.

(55) Young, W. W.; Holcomb, D. R.; Hagen, Ten, K. G.; Tabak, L. A. Expression of UDP-GalNAc:Polypeptide N-Acetylgalactosaminyltransferase Isoforms in Murine Tissues Determined by Real-Time PCR: a New View of a Large Family. *Glycobiology* **2003**, *13*, 549–557.

(56) Bennett, E. P.; Chen, Y.-W.; Schwientek, T.; Mandel, U.; Schjoldager, K. T.-B. G.; Cohen, S. M.; Clausen, H. Rescue of Drosophila Melanogaster L(2)35Aa Lethality Is Only Mediated by Polypeptide GalNAc-Transferase pgant35A, but Not by the Evolutionary Conserved Human Ortholog GalNAc-Transferase-T11. *Glycoconjugate J* **2010**, *27*, 435–444.

(57) Hagen, Ten, K. G.; Tran, D. T. A UDP-GalNAc:Polypeptide N-Acetylgalactosaminyltransferase Is Essential for Viability in Drosophila Melanogaster. *J. Biol. Chem.* **2002**, *277*, 22616–22622.

(58) Wang, R.; Yu, C.; Zhao, D.; Wu, M.; Yang, Z. The Mucin-Type Glycosylating

Enzyme Polypeptide N-Acetylgalactosaminyltransferase 14 Promotes the Migration of Ovarian Cancer by Modifying Mucin 13. *Oncology Reports* **2013**, *30*, 667–676.

(59)   Lin, M.-C.; Huang, M.-J.; Liu, C.-H.; Yang, T.-L.; Huang, M.-C. GALNT2 Enhances Migration and Invasion of Oral Squamous Cell Carcinoma by Regulating EGFR Glycosylation and Activity. *Oral Oncology* **2014**, *50*, 478–484.

(60)   Libisch, M. G.; Casás, M.; Chiribao, M. L.; Moreno, P.; Cayota, A.; Osinaga, E.; Oppezzo, P.; Robello, C. GALNT11 as a New Molecular Marker in Chronic Lymphocytic Leukemia. *Gene* **2014**, *533*, 270–279.

(61)   Wang, Z. Q.; Bachvarova, M.; Morin, C.; Plante, M. Role of the Polypeptide N-Acetylgalactosaminyltransferase 3 in Ovarian Cancer Progression: Possible Implications in Abnormal Mucin O-Glycosylation. *Oncotarget* **2014**.

(62)   Wu, Y.-M.; Liu, C.-H.; Hu, R.-H.; Huang, M.-J.; Lee, J.-J.; Chen, C.-H.; Huang, J.; Lai, H.-S.; Lee, P.-H.; Hsu, W.-M.; et al. Mucin Glycosylating Enzyme GALNT2 Regulates the Malignant Character of Hepatocellular Carcinoma by Modifying the EGF Receptor. *Cancer Research* **2011**, *71*, 7270–7279.

(63)   Bennett, E. P.; Mandel, U.; Clausen, H.; Gerken, T. A.; Fritz, T. A.; Tabak, L. A. Control of Mucin-Type O-Glycosylation: a Classification of the Polypeptide GalNAc-Transferase Gene Family. *Glycobiology* **2012**, *22*, 736–756.

(64)   Kato, K.; Jeanneau, C.; Tarp, M. A.; Benet-Pagès, A.; Lorenz-Depiereux, B.; Bennett, E. P.; Mandel, U.; Strom, T. M.; Clausen, H. Polypeptide GalNAc-Transferase T3 and Familial Tumoral Calcinosis. Secretion of Fibroblast Growth Factor 23 Requires O-Glycosylation. *J. Biol. Chem.* **2006**, *281*, 18370–18377.

(65)   Benet-Pagès, A.; Orlik, P.; Strom, T. M.; Lorenz-Depiereux, B. An FGF23 Missense Mutation Causes Familial Tumoral Calcinosis with Hyperphosphatemia. *Hum. Mol. Genet.* **2005**, *14*, 385–390.

(66)   Seidah, N. G.; Prat, A. The Biology and Therapeutic Targeting of the Proprotein Convertases. *Nature Reviews Drug Discovery* **2012**, *11*, 367–383.

(67)   Schjoldager, K. T.-B. G.; Vester-Christensen, M. B.; Goth, C. K.; Petersen, T. N.; Brunak, S.; Bennett, E. P.; Levery, S. B.; Clausen, H. A Systematic Study of Site-Specific GalNAc-Type O-Glycosylation Modulating Proprotein Convertase Processing. *J. Biol. Chem.* **2011**, *286*, 40122–40132.

(68)   Miwa, H. E.; Gerken, T. A.; Jamison, O.; Tabak, L. A. Isoform-Specific O-Glycosylation of Osteopontin and Bone Sialoprotein by Polypeptide N-Acetylgalactosaminyltransferase-1. *Journal of Biological Chemistry* **2010**, *285*, 1208–1219.

(69)   Hua, D.; Shen, L.; Xu, L.; Jiang, Z.; Zhou, Y.; Yue, A.; Zou, S.; Cheng, Z.; Wu, S. Polypeptide N-Acetylgalactosaminyltransferase 2 Regulates Cellular Metastasis-Associated Behavior in Gastric Cancer. *Int. J. Mol. Med.* **2012**, *30*, 1267–1274.

(70)   Schjoldager, K. T.-B. G.; Vester-Christensen, M. B.; Bennett, E. P.; Levery, S. B.; Schwientek, T.; Yin, W.; Blixt, O.; Clausen, H. O-Glycosylation Modulates Proprotein Convertase Activation of Angiopoietin-Like Protein 3: Possible

Role of Polypeptide GalNAc-Transferase-2 in Regulation of Concentrations of Plasma Lipids. *J. Biol. Chem.* **2010**, *285*, 36293–36303.

(71)   Topaz, O.; Shurman, D. L.; Bergman, R.; Indelman, M.; Ratajczak, P.; Mizrachi, M.; Khamaysi, Z.; Behar, D.; Petronius, D.; Friedman, V.; et al. Mutations in GALNT3, Encoding a Protein Involved in O-Linked Glycosylation, Cause Familial Tumoral Calcinosis. *Nature Genetics* **2004**, *36*, 579–581.

(72)   Gomes, J.; Marcos, N. T.; Berois, N.; Osinaga, E.; Magalhães, A.; Pinto-de-Sousa, J.; Almeida, R.; Gärtner, F.; Reis, C. A. Expression of UDP-N-Acetyl-D-Galactosamine: Polypeptide N-Acetylgalactosaminyltransferase-6 in Gastric Mucosa, Intestinal Metaplasia, and Gastric Carcinoma. *J Histochem Cytochem* **2009**, *57*, 79–86.

(73)   Gao, Y.; Liu, Z.; Feng, J.; Sun, Q.; Zhang, B.; Zheng, W.; Ma, W. Expression Pattern of Polypeptide N-Acetylgalactosaminyltransferase-10 in Gastric Carcinoma. *Oncol Lett* **2013**, *5*, 113–116.

(74)   Fakhro, K. A.; Choi, M.; Ware, S. M.; Belmont, J. W.; Towbin, J. A.; Lifton, R. P.; Khokha, M. K.; Brueckner, M. Rare Copy Number Variations in Congenital Heart Disease Patients Identify Unique Genes in Left-Right Patterning. *Proceedings of the National Academy of Sciences* **2011**, *108*, 2915–2920.

(75)   Zhang, L.; Syed, Z. A.; van Dijk Härd, I.; Lim, J.-M.; Wells, L.; Hagen, Ten, K. G. O-Glycosylation Regulates Polarized Secretion by Modulating Tango1 Stability. *Proceedings of the National Academy of Sciences* **2014**, *111*, 7296–7301.

(76)   Tagliabracci, V. S.; Engel, J. L.; Wiley, S. E.; Xiao, J.; Gonzalez, D. J.; Nidumanda Appaiah, H.; Koller, A.; Nizet, V.; White, K. E.; Dixon, J. E. Dynamic Regulation of FGF23 by Fam20C Phosphorylation, GalNAc-T3 Glycosylation, and Furin Proteolysis. *Proceedings of the National Academy of Sciences* **2014**, *111*, 5520–5525.

(77)   Sitara, D.; Razzaque, M. S.; Hesse, M.; Yoganathan, S.; Taguchi, T.; Erben, R. G.; Jüppner, H.; Lanske, B. Homozygous Ablation of Fibroblast Growth Factor-23 Results in Hyperphosphatemia and Impaired Skeletogenesis, and Reverses Hypophosphatemia in Phex-Deficient Mice. *Matrix Biol.* **2004**, *23*, 421–432.

(78)   Araya, K.; Fukumoto, S.; Backenroth, R.; Takeuchi, Y.; Nakayama, K.; Ito, N.; Yoshii, N.; Yamazaki, Y.; Yamashita, T.; Silver, J.; et al. A Novel Mutation in Fibroblast Growth Factor 23 Gene as a Cause of Tumoral Calcinosis. *J. Clin. Endocrinol. Metab.* **2005**, *90*, 5523–5527.

(79)   Hart, G. W.; Slawson, C.; Ramirez-Correa, G.; Lagerlof, O. Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. *Annu. Rev. Biochem.* **2011**, *80*, 825–858.

(80)   Hang, H. C.; Yu, C.; Pratt, M. R.; Bertozzi, C. R. Probing Glycosyltransferase Activities with the Staudinger Ligation. *J. Am. Chem. Soc.* **2004**, *126*, 6–7.

(81)   Pratt, M. R.; Bertozzi, C. R. Synthetic Glycopeptides and Glycoproteins as Tools for Biology. *Chem Soc Rev* **2005**, *34*, 58–68.

(82)   Kong, Y.; Joshi, H. J.; Schjoldager, K. T.-B. G.; Madsen, T. D.; Gerken, T. A.;
         Vester-Christensen, M. B.; Wandall, H. H.; Bennett, E. P.; Levery, S. B.;
         Vakhrushev, S. Y.; et al. Probing Polypeptide GalNAc-Transferase Isoform
         Substrate Specificities by in Vitro Analysis. *Glycobiology* **2014**, *25*,
         cwu089–65.

(83)   Schjoldager, K. T.-B. G.; Vakhrushev, S. Y.; Kong, Y.; Steentoft, C.; Nudelman,
         A. S.; Pedersen, N. B.; Wandall, H. H.; Mandel, U.; Bennett, E. P.; Levery, S. B.;
         et al. Probing Isoform-Specific Functions of Polypeptide GalNAc-
         Transferases Using Zinc Finger Nuclease Glycoengineered SimpleCells.
         Proceedings of the National Academy of Sciences **2012**, *109*, 9893–9898.

(84)   Hang, H. C.; Yu, C.; Hagen, Ten, K. G.; Tian, E.; Winans, K. A.; Tabak, L. A.;
         Bertozzi, C. R. Small Molecule Inhibitors of Mucin-Type O-Linked
         Glycosylation From a Uridine-Based Library. *Chem. Biol.* **2004**, *11*, 337–
         345.

(85)   Luchansky, S. J.; Hang, H. C.; Saxon, E.; Grunwell, J. R.; Yu, C.; Dube, D. H.;
         Bertozzi, C. R. Constructing Azide-Labeled Cell Surfaces Using
         Polysaccharide Biosynthetic Pathways. *Methods Enzymol.* **2003**, *362*, 249–
         272.

(86)   Binda, O.; Boyce, M.; Rush, J. S.; Palaniappan, K. K.; Bertozzi, C. R.; Gozani, O.
         A Chemical Method for Labeling Lysine Methyltransferase Substrates.
         Chembiochem **2011**, *12*, 330–334.

(87)   Shogren-Knaak, M. A.; Alaimo, P. J.; Shokat, K. M. Recent Advances in
         Chemical Approaches to the Study of Biological Systems.
         http://dx.doi.org/10.1146/annurev.cellbio.17.1.405 **2003**.

(88)   Islam, K. Allele-Specific Chemical Genetics: Concept, Strategies, and
         Applications. *ACS Chem. Biol.* **2015**, *10*, 343–363.

(89)   Islam, K.; Bothwell, I.; Chen, Y.; Sengelaub, C.; Wang, R.; Deng, H.; Luo, M.
         Bioorthogonal Profiling of Protein Methylation Using Azido Derivative of S-
         Adenosyl-L-Methionine. *J. Am. Chem. Soc.* **2012**, *134*, 5909–5915.

(90)   Bertozzi, C. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of
         Functionality. *Angewandte Chemie International …* **2009**.

(91)   Shah, K.; Liu, Y.; Deirmengian, C.; Shokat, K. M. Engineering Unnatural
         Nucleotide Specificity for Rous Sarcoma Virus Tyrosine Kinase to Uniquely
         Label Its Direct Substrates. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3565–
         3570.

(92)   Bishop, A. C.; Shah, K.; Liu, Y.; Witucki, L.; Kung, C.-Y.; Shokat, K. M. Design
         of Allele-Specific Inhibitors to Probe Protein Kinase Signaling. *Curr. Biol.*
         1998.

(93)   Liu, Y.; Shah, K.; Yang, F.; Witucki, L.; Shokat, K. M. Engineering Src Family
         Protein Kinases with Unnatural Nucleotide Specificity. *Chem. Biol.* **1998**, *5*,
         91–101.

(94)   Bishop, A. C.; Ubersax, J. A.; Petsch, D. T.; Matheos, D. P.; Gray, N. S.;
         Blethrow, J.; Shimizu, E.; Tsien, J. Z.; Schultz, P. G.; Rose, M. D.; et al. A
         Chemical Switch for Inhibitor-Sensitive Alleles of Any Protein Kinase.
         Nature **2000**, *407*, 395–401.

(95)   Denzel, A.; Hare, K. J.; Zhang, C.; Shokat, K.; Jenkinson, E. J.; Anderson, G.;

Hayday, A. Cutting Edge: a Chemical Genetic System for the Analysis of Kinases Regulating T Cell Development. *The Journal of Immunology* **2003**, *171*, 519–523.

(96) Jaeschke, A.; Karasarides, M.; Ventura, J.-J.; Ehrhardt, A.; Zhang, C.; Flavell, R. A.; Shokat, K. M.; Davis, R. J. JNK2 Is a Positive Regulator of the cJun Transcription Factor. *Molecular Cell* **2006**, *23*, 899–911.

(97) Dar, A. C.; Shokat, K. M. The Evolution of Protein Kinase Inhibitors From Antagonists to Agonists of Cellular Signaling. *Annu. Rev. Biochem.* **2011**, *80*, 769–795.

(98) Shokat, K. Features of Selective Kinase Inhibitors. *Chem. Biol.* **2005**.

(99) Buzko, O.; Shokat, K. Magic Bullets for Protein Kinases. *Trends in cell biology* **2001**.

(100) Ulrich, S. M.; Buzko, O.; Shah, K.; Shokat, K. M. Towards the Engineering of an Orthogonal Protein Kinase/Nucleotide Triphosphate Pair. *Tetrahedron* **2000**, *56*, 9495–9502.

(101) Koch, A.; Hauf, S. Strategies for the Identification of Kinase Substrates Using Analog-Sensitive Kinases. *Eur. J. Cell Biol.* **2010**, *89*, 184–193.

(102) Wang, H.; Shimizu, E.; Tang, Y.-P.; Cho, M.; Kyin, M.; Zuo, W.; Robinson, D. A.; Alaimo, P. J.; Zhang, C.; Morimoto, H.; et al. Inducible Protein Knockout Reveals Temporal Requirement of CaMKII Reactivation for Memory Consolidation in the Brain. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4287–4292.

(103) Banko, M. R.; Allen, J. J.; Schaffer, B. E.; Wilker, E. W.; Tsou, P.; White, J. L.; Villén, J.; Wang, B.; Kim, S. R.; Sakamoto, K.; et al. Chemical Genetic Screen for AMPKα2 Substrates Uncovers a Network of Proteins Involved in Mitosis. *Molecular Cell* **2011**, *44*, 878–892.

(104) Allen, J. J.; Lazerwith, S. E.; Shokat, K. M. Bio-Orthogonal Affinity Purification of Direct Kinase Substrates. *J. Am. Chem. Soc.* **2005**, *127*, 5288–5289.

(105) Cibrián Uhalte, E.; Kirchner, M.; Hellwig, N.; Allen, J. J.; Donat, S.; Shokat, K. M.; Selbach, M.; Abdelilah-Seyfried, S. In Vivo Conditions to Identify Prkci Phosphorylation Targets Using the Analog-Sensitive Kinase Method in Zebrafish. *PLoS ONE* **2012**, *7*, e40000.

(106) Polson, A. G.; Huang, L.; Lukac, D. M.; Blethrow, J. D.; Morgan, D. O.; Burlingame, A. L.; Ganem, D. Kaposi's Sarcoma-Associated Herpesvirus K-bZIP Protein Is Phosphorylated by Cyclin-Dependent Kinases. *Journal of Virology* **2001**, *75*, 3175–3184.

(107) Wang, R.; Zheng, W.; Yu, H.; Deng, H.; Luo, M. Labeling Substrates of Protein Arginine Methyltransferase with Engineered Enzymes and Matched S-Adenosyl-L-Methionine Analogues. *J. Am. Chem. Soc.* **2011**, *133*, 7648–7651.

(108) Wang, R.; Luo, M. A Journey Toward Bioorthogonal Profiling of Protein Methylation Inside Living Cells. *Current Opinion in Chemical Biology* **2013**, *17*, 729–737.

(109) Islam, K.; Chen, Y.; Wu, H.; Bothwell, I. R.; Blum, G. J.; Zeng, H.; Dong, A.; Zheng, W.; Min, J.; Deng, H.; et al. Defining Efficient Enzyme-Cofactor Pairs

for Bioorthogonal Profiling of Protein Methylation. *Proceedings of the National Academy of Sciences* **2013**, *110*, 16778–16783.

(110)   Hancock, S. M.; Vaughan, M. D.; Withers, S. G. Engineering of Glycosidases and Glycosyltransferases. *Current Opinion in Chemical Biology* **2006**, *10*, 509–519.

(111)   Ramakrishnan, B.; Qasba, P. K. Structure-Based Design of Beta 1,4-Galactosyltransferase I (Beta 4Gal-T1) with Equally Efficient N-Acetylgalactosaminyltransferase Activity: Point Mutation Broadens Beta 4Gal-T1 Donor Specificity. *Journal of Biological Chemistry* **2002**, *277*, 20833–20839.

(112)   Khidekel, N.; Arndt, S.; Lamarre-Vincent, N.; Lippert, A.; Poulin-Kerstien, K. G.; Ramakrishnan, B.; Qasba, P. K.; Hsieh-Wilson, L. C. A Chemoenzymatic Approach Toward the Rapid and Sensitive Detection of O-GlcNAc Posttranslational Modifications. *J. Am. Chem. Soc.* **2003**, *125*, 16162–16163.

(113)   Clark, P. M.; Dweck, J. F.; Mason, D. E.; Hart, C. R.; Buck, S. B.; Peters, E. C.; Agnew, B. J.; Hsieh-Wilson, L. C. Direct in-Gel Fluorescence Detection and Cellular Imaging of O-GlcNAc-Modified Proteins. *J. Am. Chem. Soc.* **2008**, *130*, 11576–11577.

(114)   Yu, S.-H.; Boyce, M.; Wands, A. M.; Bond, M. R.; Bertozzi, C. R.; Kohler, J. J. Metabolic Labeling Enables Selective Photocrosslinking of O-GlcNAc-Modified Proteins to Their Binding Partners. *Proceedings of the National Academy of Sciences* **2012**, *109*, 4834–4839.

(115)   Kubota, T.; Shiba, T.; Sugioka, S.; Furukawa, S.; Sawaki, H.; Kato, R.; Wakatsuki, S.; Narimatsu, H. Structural Basis of Carbohydrate Transfer Activity by Human UDP-GalNAc: Polypeptide Alpha-N-Acetylgalactosaminyltransferase (Pp-GalNAc-T10). *J. Mol. Biol.* **2006**, *359*, 708–727.

(116)   Fritz, T. A.; Raman, J.; Tabak, L. A. Dynamic Association Between the Catalytic and Lectin Domains of Human UDP-GalNAc:Polypeptide -N-Acetylgalactosaminyltransferase-2. *Journal of Biological Chemistry* **2006**, *281*, 8613–8619.

(117)   Raman, J.; Fritz, T. A.; Gerken, T. A.; Jamison, O.; Live, D.; Liu, M.; Tabak, L. A. The Catalytic and Lectin Domains of UDP-GalNAc:Polypeptide Alpha-N-Acetylgalactosaminyltransferase Function in Concert to Direct Glycosylation Site Selection. *J. Biol. Chem.* **2008**, *283*, 22942–22951.

(118)   Eblen, S. T.; Kumar, N. V.; Shah, K.; Henderson, M. J.; Watts, C. K. W.; Shokat, K. M.; Weber, M. J. Identification of Novel ERK2 Substrates Through Use of an Engineered Kinase and ATP Analogs. *Journal of Biological Chemistry* **2003**, *278*, 14926–14935.

(119)   Habelhah, H.; Shah, K.; Huang, L.; Burlingame, A. L.; Shokat, K. M.; Ronai, Z. Identification of New JNK Substrate Using ATP Pocket Mutant JNK and a Corresponding ATP Analogue. *Journal of Biological Chemistry* **2001**, *276*, 18090–18095.

(120)   Hirschberg, C. B.; Robbins, P. W.; Abeijon, C. Transporters of Nucleotide Sugars, ATP, and Nucleotide Sulfate in the Endoplasmic Reticulum and Golgi Apparatus. *Annu. Rev. Biochem.* **1998**, *67*, 49–69.

(121)   Caffaro, C. E.; Luhn, K.; Bakker, H.; Vestweber, D.; Samuelson, J.; Berninsone, P.; Hirschberg, C. B. A Single Caenorhabditis Elegans Golgi Apparatus-Type Transporter of UDP-Glucose, UDP-Galactose, UDP-N-Acetylglucosamine, and UDP-N-Acetylgalactosamine. *Biochemistry* **2008**, *47*, 4337–4344.

(122)   Pesnot, T.; Jorgensen, R.; Palcic, M. M.; Wagner, G. K. Structural and Mechanistic Basis for a New Mode of Glycosyltransferase Inhibition. *Nat. Chem. Biol. 6*, 321–323.

(123)   Busca, P.; Piller, V.; Piller, F.; Martin, O. R. Synthesis and Biological Evaluation of New UDP-GalNAc Analogues for the Study of Polypeptide-A-GalNAc-Transferases. *Bioorganic & Medicinal Chemistry Letters* **2003**, *13*, 1853–1856.

(124)   Bülter, T.; Schumacher, T.; Namdjou, D. J.; Gutiérrez Gallego, R.; Clausen, H.; Elling, L. Chemoenzymatic Synthesis of Biotinylated Nucleotide Sugars as Substrates for Glycosyltransferases. *Chembiochem* **2001**, *2*, 884–894.

(125)   Saxon, E. Cell Surface Engineering by a Modified Staudinger Reaction. *Science* **2000**, *287*, 2007–2010.

(126)   Laughlin, S. T.; Bertozzi, C. R. Metabolic Labeling of Glycans with Azido Sugars and Subsequent Glycan-Profiling and Visualization via Staudinger Ligation. *Nat. Protoc.* **2007**, *2*, 2930–2944.

(127)   Bergfeld, A. K.; Pearce, O. M. T.; Diaz, S. L.; Lawrence, R.; Vocadlo, D. J.; Choudhury, B.; Esko, J. D.; Varki, A. Metabolism of Vertebrate Amino Sugars with N-Glycolyl Groups: Incorporation of N-Glycolylhexosamines Into Mammalian Glycans by Feeding N-Glycolylgalactosamine. *J. Biol. Chem.* **2012**, *287*, 28898–28916.

(128)   Bergfeld, A. K.; Pearce, O. M. T.; Diaz, S. L.; Pham, T.; Varki, A. Metabolism of Vertebrate Amino Sugars with N-Glycolyl Groups: Elucidating the Intracellular Fate of the Non-Human Sialic Acid N-Glycolylneuraminic Acid. *J. Biol. Chem.* **2012**, *287*, 28865–28881.

(129)   Zaro, B. W.; Chuh, K. N.; Pratt, M. R. Chemical Reporter for Visualizing Metabolic Cross-Talk Between Carbohydrate Metabolism and Protein Modification. *ACS Chem. Biol.* **2014**, *9*, 1991–1996.

(130)   Zaro, B. W.; Yang, Y.-Y.; Hang, H. C.; Pratt, M. R. Chemical Reporters for Fluorescent Detection and Identification of O-GlcNAc-Modified Proteins Reveal Glycosylation of the Ubiquitin Ligase NEDD4-1. *Proceedings of the National Academy of Sciences* **2011**, *108*, 8146–8151.

(131)   Boyce, M.; Carrico, I. S.; Ganguli, A. S.; Yu, S.-H.; Hangauer, M. J.; Hubbard, S. C.; Kohler, J. J.; Bertozzi, C. R. Metabolic Cross-Talk Allows Labeling of O-Linked Beta-N-Acetylglucosamine-Modified Proteins via the N-Acetylgalactosamine Salvage Pathway. *Proceedings of the National Academy of Sciences* **2011**, *108*, 3141–3146.

(132)   Fritz, T. A.; Hurley, J. H.; Trinh, L. B.; Shiloach, J.; Tabak, L. A. The Beginnings of Mucin Biosynthesis: the Crystal Structure of UDP-GalNAc:Polypeptide Alpha-N-Acetylgalactosaminyltransferase-T1. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15307–15312.

(133)   Raman, J.; Fritz, T. A.; Gerken, T. A.; Jamison, O.; Live, D.; Liu, M.; Tabak, L. A.

The Catalytic and Lectin Domains of UDP-GalNAc:Polypeptide -N-Acetylgalactosaminyltransferase Function in Concert to Direct Glycosylation Site Selection. *Journal of Biological Chemistry* **2008**, *283*, 22942–22951.

(134)   Hagen, F. K. cDNA Cloning and Expression of a Novel UDP-N-Acetyl-D-Galactosamine:Polypeptide N-Acetylgalactosaminyltransferase. *Journal of Biological Chemistry* **1997**, *272*, 13843–13848.

(135)   Wada, K.; Aota, S.; Tsuchiya, R.; Ishibashi, F.; Gojobori, T.; Ikemura, T. Codon Usage Tabulated From the GenBank Genetic Sequence Data. *Nucleic Acids Res* **1990**, *18 Suppl*, 2367–2411.

(136)   Zheng, L.; Baumann, U.; Reymond, J. L. An Efficient One-Step Site-Directed and Site-Saturation Mutagenesis Protocol. *Nucleic Acids Res* **2004**, *32*, e115.

(137)   White, T.; Bennett, E. P.; Takio, K.; Sørensen, T.; Bonding, N.; Clausen, H. Purification and cDNA Cloning of a Human UDP-N-Acetyl-Alpha-D-Galactosamine:Polypeptide N-Acetylgalactosaminyltransferase. *Journal of Biological Chemistry* **1995**, *270*, 24156–24165.

(138)   Wandall, H. H. Substrate Specificities of Three Members of the Human UDP-N-Acetyl-Alpha -D-Galactosamine:Polypeptide N-Acetylgalactosaminyltransferase Family, GalNAc-T1, -T2, and -T3. *Journal of Biological Chemistry* **1997**, *272*, 23503–23514.

(139)   Engler, C.; Gruetzner, R.; Kandzia, R.; Marillonnet, S. Golden Gate Shuffling: a One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLoS ONE* **2009**, *4*, e5553.

(140)   Bennett, E. P.; Hassan, H.; Hollingsworth, M. A.; Clausen, H. A Novel Human UDP-N-Acetyl-D-Galactosamine:Polypeptide N-Acetylgalactosaminyltransferase, GalNAc-T7, with Specificity for Partial GalNAc-Glycosylated Acceptor Substrates. *FEBS Lett.* **1999**, *460*, 226–230.

(141)   Hagen, K. G. T. Cloning and Characterization of a Ninth Member of the UDP-GalNAc:Polypeptide N-Acetylgalactosaminyltransferase Family, ppGaNTase-T9. *Journal of Biological Chemistry* **2001**, *276*, 17395–17404.

(142)   Guo, J.; Tachibana, K. K.; Kameyama, A. ScienceDirect - FEBS Letters : Characterization of a Novel Human UDP-GalNAc Transferase, Pp-GalNAc-T101. *FEBS Lett.* **2002**.

(143)   Liu, H.; Naismith, J. H. An Efficient One-Step Site-Directed Deletion, Insertion, Single and Multiple-Site Plasmid Mutagenesis Protocol. *BMC Biotechnol.* **2008**, *8*, 91.

(144)   Walker, J. M. *The Protein Protocols Handbook*; Springer Science & Business Media: New Jersey, 2002; Vol. 0.

(145)   Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angewandte Chemie* **2002**, *114*, 2708–2711.

(146)   Christian W Tornøe; Caspar Christensen, A.; Meldal, M. Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *J. Org. Chem.* **2002**,

67, 3057–3064.

(147)   Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. A Strain-Promoted [3 + 2] Azide–Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *J. Am. Chem. Soc.* **2004**, *126*, 15046–15047.

(148)   Soriano Del Amo, D.; Wang, W.; Jiang, H.; Besanceney, C.; Yan, A. C.; Levy, M.; Liu, Y.; Marlow, F. L.; Wu, P. Biocompatible Copper(I) Catalysts for in Vivo Imaging of Glycans. *J. Am. Chem. Soc.* **2010**, *132*, 16893–16899.

(149)   Besanceney Webler, C.; Jiang, H.; Zheng, T.; Feng, L.; Soriano Del Amo, D.; Wang, W.; Klivansky, L. M.; Marlow, F. L.; Liu, Y.; Wu, P. Increasing the Efficacy of Bioorthogonal Click Reactions for Bioconjugation: a Comparative Study. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 8051–8056.

(150)   Jiang, H.; Zheng, T.; Lopez-Aguilar, A.; Feng, L.; Kopp, F.; Marlow, F. L.; Wu, P. Monitoring Dynamic Glycosylation in Vivo Using Supersensitive Click Chemistry. *Bioconjugate …* **2014**, *25*, 698–706.

(151)   Chang, P. V.; Chen, X.; Smyrniotis, C.; Xenakis, A.; Hu, T.; Bertozzi, C. R.; Wu, P. Metabolic Labeling of Sialic Acids in Living Animals with Alkynyl Sugars. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 4030–4033.

(152)   Speers, A. E.; Cravatt, B. F. Profiling Enzyme Activities in Vivo Using Click Chemistry Methods. *Chem. Biol.* **2004**, *11*, 535–546.

(153)   Blackman, M. L.; Royzen, M.; Fox, J. M. Tetrazine Ligation: Fast Bioconjugation Based on Inverse-Electron-Demand Diels–Alder Reactivity. *J. Am. Chem. Soc.* **2008**, *130*, 13518–13519.

(154)   Patterson, D. M.; Nazarova, L. A.; Xie, B.; Kamber, D. N.; Prescher, J. A. Functionalized Cyclopropenes as Bioorthogonal Chemical Reporters. *Journal of the …* **2012**, *134*, 18638–18643.

(155)   Devaraj, N. K.; Weissleder, R. Biomedical Applications of Tetrazine Cycloadditions. *Acc. Chem. Res.* **2011**, *44*, 816–827.

(156)   Selvaraj, R.; Fox, J. M. Trans-Cyclooctene—a Stable, Voracious Dienophile for Bioorthogonal Labeling. *Current Opinion in Chemical Biology* **2013**, *17*, 753–760.

(157)   Devaraj, N. K.; Weissleder, R.; Hilderbrand, S. A. Tetrazine-Based Cycloadditions: Application to Pretargeted Live Cell Imaging. *Bioconjug. Chem.* **2008**, *19*, 2297–2299.

(158)   Devaraj, N. K.; Hilderbrand, S.; Upadhyay, R.; Mazitschek, R.; Weissleder, R. Bioorthogonal Turn-on Probes for Imaging Small Molecules Inside Living Cells. *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 2869–2872.

(159)   Rossin, R.; Renart Verkerk, P.; van den Bosch, S. M.; Vulders, R. C. M.; Verel, I.; Lub, J.; Robillard, M. S. In Vivo Chemistry for Pretargeted Tumor Imaging in Live Mice. *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 3375–3378.

(160)   Karver, M. R.; Weissleder, R. Synthesis and Evaluation of a Series of 1, 2, 4, 5-Tetrazines for Bioorthogonal Conjugation. *Bioconjugate …* **2011**.

(161)   Mahal, L. K.; Yarema, K. J.; Bertozzi, C. R. Engineering Chemical Reactivity on Cell Surfaces Through Oligosaccharide Biosynthesis. *Science* **1997**, *276*, 1125–1128.

(162)   Agarwal, P.; van der Weijden, J.; Sletten, E. M.; Rabuka, D.; Bertozzi, C. R. A Pictet-Spengler Ligation for Protein Chemical Modification. *Proc. Natl.*

Acad. Sci. U.S.A. **2013**, *110*, 46–51.

(163)   Charron, G.; Zhang, M. M.; Yount, J. S.; Wilson, J.; Raghavan, A. S.; Shamir, E.; Hang, H. C. Robust Fluorescent Detection of Protein Fatty-Acylation with Chemical Reporters. *J. Am. Chem. Soc.* **2009**, *131*, 4967–4975.

(164)   Bateman, L. A.; Zaro, B. W.; Chuh, K. N.; Pratt, M. R. N-Propargyloxycarbamate Monosaccharides as Metabolic Chemical Reporters of Carbohydrate Salvage Pathways and Protein Glycosylation. *Chem. Commun. (Camb.)* **2013**, *49*, 4328–4330.

(165)   Hang, H. C.; Bertozzi, C. R. Ketone Isosteres of 2-N-Acetamidosugars as Substrates for Metabolic Cell Surface Engineering. *Journal of the American Chemical ...* **2001**.

(166)   Islam, K.; Zheng, W.; Yu, H.; Deng, H.; Luo, M. Expanding Cofactor Repertoire of Protein Lysine Methyltransferase for Substrate Labeling. *ACS Chem. Biol.* **2011**, *6*, 679–684.

(167)   Elphick, L. M.; Lee, S. E.; Gouverneur, V.; Mann, D. J. Using Chemical Genetics and ATP Analogues to Dissect Protein Kinase Function. *ACS Chem. Biol.* **2007**, *2*, 299–314.

(168)   Heidlas, J. E.; Williams, K. W.; Whitesides, G. M. Nucleoside Phosphate Sugars: Syntheses on Practical Scales for Use as Reagents in the Enzymatic Preparation of Oligosaccharides and Glycoconjugates. *Acc. Chem. Res.* **1992**, *25*, 307–314.

(169)   Ravindranathan Kartha, K. P.; Mukhopadhyay, B.; Field, R. A. Practical De-O-Acylation Reactions Promoted by Molecular Sieves. *Carbohydr. Res.* **2004**, *339*, 729–732.

(170)   Marlow, A. L.; Kiessling, L. L. Improved Chemical Synthesis of UDP-Galactofuranose. *Org. Lett.* **2001**, *3*, 2517–2519.

(171)   Marchesan, S.; Macmillan, D. Chemoenzymatic Synthesis of GDP-Azidodeoxymannoses: Non-Radioactive Probes for Mannosyltransferase Activity. *Chem. Commun. (Camb.)* **2008**, 4321–4323.

(172)   Wagner, G. K.; Pesnot, T.; Field, R. A. A Survey of Chemical Methods for Sugar-Nucleotide Synthesis. *Natural Product Reports* **2009**, *26*, 1172–1194.

(173)   Tanaka, H.; Yoshimura, Y.; Jørgensen, M. R.; Cuesta-Seijo, J. A.; Hindsgaul, O. A Simple Synthesis of Sugar Nucleoside Diphosphates by Chemical Coupling in Water. *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 11531–11534.

(174)   Dabrowski Tumanski, P.; Kowalska, J.; Jemielity, J. Efficient and Rapid Synthesis of Nucleoside Diphosphate Sugars From Nucleoside Phosphorimidazolides. *European Journal of Organic Chemistry* **2013**, *2013*, 2147–2154.

(175)   Pouilly, S.; Bourgeaux, V.; Piller, F.; Piller, V. Evaluation of Analogues of GalNAc as Substrates for Enzymes of the Mammalian GalNAc Salvage Pathway. *ACS Chem. Biol.* **2012**, *7*, 753–760.

(176)   Heidlas, J. E.; Lees, W. J.; Whitesides, G. M. Practical Enzyme-Based Syntheses of Uridine 5"-Diphosphogalactose and Uridine 5-"Diphospho-N-Acetylgalactosamine on a Gram Scale. *J. Org. Chem.* **1992**, *57*, 152–157.

(177)   Errey, J. C.; Mukhopadhyay, B.; Kartha, K. P.; Field, R. A. Flexible Enzymatic and Chemo-Enzymatic Approaches to a Broad Range of Uridine-

Diphospho-Sugars. *Chem. Commun. (Camb.)* **2004**, 2706–2707.

(178)   Liu, Z.; Zhang, J.; Chen, X.; Wang, P. G. Combined Biosynthetic Pathway for De Novo Production of UDP-Galactose: Catalysis with Multiple Enzymes Immobilized on Agarose Beads. *Chembiochem* **2002**, *3*, 348–355.

(179)   Kalkhof, S.; Sinz, A. Chances and Pitfalls of Chemical Cross-Linking with Amine-Reactive N-Hydroxysuccinimide Esters. *Anal Bioanal Chem* **2008**, *392*, 305–312.

(180)   Goddard-Borger, E. D.; Stick, R. V. An Efficient, Inexpensive, and Shelf-Stable Diazotransfer Reagent: Imidazole-1-Sulfonyl Azide Hydrochloride. *Org. Lett.* **2007**, *9*, 3797–3800.

(181)   Fischer, N.; Goddard-Borger, E. D.; Greiner, R.; Klapötke, T. M.; Skelton, B. W.; Stierstorfer, J. Sensitivities of Some Imidazole-1-Sulfonyl Azide Salts. *J. Org. Chem.* **2012**, *77*, 1760–1764.

(182)   Joullié, M. M.; Lassen, K. M. Evolution of Amide Bond Formation. *Arkivoc* **2010**.

(183)   Panne, P.; Fox, J. M. Rh-Catalyzed Intermolecular Reactions of Alkynes with Alpha-Diazoesters That Possess Beta-Hydrogens: Ligand-Based Control Over Divergent Pathways. *J. Am. Chem. Soc.* **2007**, *129*, 22–23.

(184)   Lau, K.; Thon, V.; Yu, H.; Ding, L.; Chen, Y.; Muthana, M. M.; Wong, D.; Huang, R.; Chen, X. Highly Efficient Chemoenzymatic Synthesis of Beta1-4-Linked Galactosides with Promiscuous Bacterial Beta1-4-Galactosyltransferases. *Chem. Commun. (Camb.)* **2010**, *46*, 6066–6068.

(185)   Chen, X.; Fang, J.; Zhang, J.; Liu, Z.; Shao, J.; Kowal, P.; Andreana, P.; Wang, P. G. Sugar Nucleotide Regeneration Beads (Superbeads): a Versatile Tool for the Practical Synthesis of Oligosaccharides. *J. Am. Chem. Soc.* **2001**, *123*, 2081–2082.

(186)   Hagen, F.; Van Wuyckhuyse, B.; Tabak, L. A. Purification, Cloning, and Expression of a Bovine UDP-GalNAc: Polypeptide N-Acetyl-Galactosaminyltransferase. *Journal of Biological Chemistry* **1993**.

(187)   Wu, Z. L.; Ethen, C. M.; Prather, B.; Machacek, M.; Jiang, W. Universal Phosphatase-Coupled Glycosyltransferase Assay. *Glycobiology* **2011**, *21*, 727–733.

(188)   Mikkola, S. Hydrolysis and Isomerization of Sugar Phosphates and Carbohydrate Phosphodiesters. **2013**.

(189)   Nunez, H. A.; Barker, R. The Metal Ion Catalyzed Decomposition of Nucleoside Diphosphate Sugars. *Biochemistry* **1976**, *15*, 3843–3847.

(190)   Spik, G.; Six, P.; Montreuil, J. Chemical and Enzymic Degradations of Nucleoside Mono- and Diphosphate Sugars. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1979**, *584*, 203–215.

(191)   Bedford, C. T.; Hickman, A. D.; Logan, C. J. Structure-Activity Studies of Glucose Transfer: Determination of the Spontaneous Rates of Hydrolysis of Uridine 5"-Diphospho-Alpha-D-Glucose (UDPG) and Uridine 5-"Diphospho-Alpha-D-Glucuronic Acid (UDPGA). *Bioorg. Med. Chem.* **2003**, *11*, 2339–2345.

(192)   Huhta, E.; Parjanen, A.; Mikkola, S. A Kinetic Study on the Chemical Cleavage of Nucleoside Diphosphate Sugars. *Carbohydr. Res.* **2010**, *345*,

696–703.

(193) Meng, Y.; High, K.; Antonello, J.; Washabaugh, M. W.; Zhao, Q. Enhanced Sensitivity and Precision in an Enzyme-Linked Immunosorbent Assay with Fluorogenic Substrates Compared with Commonly Used Chromogenic Substrates. *Analytical Biochemistry* **2005**, *345*, 227–236.

(194) Kumar, N. V.; Eblen, S. T.; Weber, M. J. Identifying Specific Kinase Substrates Through Engineered Kinases and ATP Analogs. *Methods* **2004**, *32*, 389–397.

(195) Lefebvre, I.; Perigaud, C.; Pompon, A.; Aubertin, A.-M.; Girardet, J.-L.; Kirn, A.; Gosselin, G.; Imbach, J.-L. Mononucleoside Phosphotriester Derivatives with S-Acyl-2-Thioethyl Bioreversible Phosphate-Protecting Groups: Intracellular Delivery of 3'-Azido-2',3"-Dideoxythymidine 5-"Monophosphate. *J. Med. Chem.* **1995**, *38*, 3941–3950.

(196) Pradere, U.; Garnier-Amblard, E. C.; Coats, S. J.; Amblard, F.; Schinazi, R. F. Synthesis of Nucleoside Phosphate and Phosphonate Prodrugs. *Chem. Rev.* **2014**, *114*, 9154–9218.

(197) Provoda, C. J.; Lee, K.-D. Bacterial Pore-Forming Hemolysins and Their Use in the Cytosolic Delivery of Macromolecules. *Advanced Drug Delivery Reviews* **2000**, *41*, 209–221.

(198) Kano, F.; Nakatsu, D.; Noguchi, Y.; Yamamoto, A.; Murata, M. A Resealed-Cell System for Analyzing Pathogenic Intracellular Events: Perturbation of Endocytic Pathways Under Diabetic Conditions. *PLoS ONE* **2012**, *7*, e44127.

(199) Walev, I.; Bhakdi, S. C.; Hofmann, F.; Djonder, N.; Valeva, A.; Aktories, K.; Bhakdi, S. Delivery of Proteins Into Living Cells by Reversible Membrane Permeabilization with Streptolysin-O. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3185–3190.

(200) Allen, T. M.; Cullis, P. R. Liposomal Drug Delivery Systems: From Concept to Clinical Applications. *Advanced Drug Delivery Reviews* **2013**, *65*, 36–48.

(201) Varkouhi, A. K.; Scholte, M.; Storm, G.; Haisma, H. J. Endosomal Escape Pathways for Delivery of Biologicals. *Journal of Controlled Release* **2011**, *151*, 220–228.

(202) Wang-Gillam, A.; Pastuszak, I.; Elbein, A. D. A 17-Amino Acid Insert Changes UDP-N-Acetylhexosamine Pyrophosphorylase Specificity From UDP-GalNAc to UDP-GlcNAc. *Journal of Biological Chemistry* **1998**, *273*, 27055–27057.

(203) Gerken, T. A.; Raman, J.; Fritz, T. A.; Jamison, O. Identification of Common and Unique Peptide Substrate Preferences for the UDP-GalNAc:Polypeptide Alpha-N-Acetylgalactosaminyltransferases T1 and T2 Derived From Oriented Random Peptide Substrates. *J. Biol. Chem.* **2006**, *281*, 32403–32416.

(204) Gerken, T. A.; Hagen, Ten, K. G.; Jamison, O. Conservation of Peptide Acceptor Preferences Between Drosophila and Mammalian Polypeptide-GalNAc Transferase Ortholog Pairs. *Glycobiology* **2008**, *18*, 861–870.

(205) Hanisch, F. G.; Reis, C. A.; Clausen, H.; Paulsen, H. Evidence for Glycosylation-Dependent Activities of Polypeptide N-

Acetylgalactosaminyltransferases rGalNAc-T2 and -T4 on Mucin Glycopeptides. *Glycobiology* **2001**, *11*, 731–740.

(206)   Gaziel-Sovran, A.; Segura, M. F.; Di Micco, R.; Collins, M. K.; Hanniford, D.; Vega-Saenz de Miera, E.; Rakus, J. F.; Dankert, J. F.; Shang, S.; Kerbel, R. S.; et al. miR-30b/30d Regulation of GalNAc Transferases Enhances Invasion and Immunosuppression During Metastasis. *Cancer Cell* **2011**, *20*, 104–118.

(207)   Peng, R.-Q.; Wan, H.-Y.; Li, H.-F.; Liu, M.; Li, X.; Tang, H. MicroRNA-214 Suppresses Growth and Invasiveness of Cervical Cancer Cells by Targeting UDP-N-Acetyl-A-D-Galactosamine:Polypeptide N-Acetylgalactosaminyltransferase 7. *J. Biol. Chem.* **2012**, *287*, 14301–14309.

(208)   Kathiresan, S.; Melander, O.; Guiducci, C.; Surti, A.; Burtt, N. E. L. P.; Rieder, M. J.; Cooper, G. M.; Roos, C.; Voight, B. F.; Havulinna, A. S.; et al. Six New Loci Associated with Blood Low-Density Lipoprotein Cholesterol, High-Density Lipoprotein Cholesterol or Triglycerides in Humans. *Nature Genetics* **2008**, *40*, 189–197.

(209)   Teslovich, T. M.; Musunuru, K.; Smith, A. V.; Edmondson, A. C.; Stylianou, I. M.; Koseki, M.; Pirruccello, J. P.; Ripatti, S.; Chasman, D. I.; Willer, C. J.; et al. Biological, Clinical and Population Relevance of 95 Loci for Blood Lipids. *Nature* **2010**, *466*, 707–713.

(210)   Degoma, E. M.; Rader, D. J. Novel HDL-Directed Pharmacotherapeutic Strategies. *Nat Rev Cardiol* **2011**, *8*, 266–277.

(211)   Kontush, A.; Chapman, M. J. Functionally Defective High-Density Lipoprotein: a New Therapeutic Target at the Crossroads of Dyslipidemia, Inflammation, and Atherosclerosis. *Pharmacol. Rev.* **2006**, *58*, 342–374.

(212)   Karlsson, H.; Leanderson, P.; Tagesson, C.; Lindahl, M. Lipoproteomics II: Mapping of Proteins in High-Density Lipoprotein Using Two-Dimensional Gel Electrophoresis and Mass Spectrometry. *Proteomics* **2005**, *5*, 1431–1445.

(213)   Blum, C.; Levy, R.; Eisenberg, S. High Density Lipoprotein Metabolism in Man. *Journal of Clinical Investigation* **1977**, *60*, 795–807.

(214)   Kozarsky, K. F.; Donahee, M. H.; Rigotti, A.; Iqbal, S. N.; Edelman, E. R.; Krieger, M. Overexpression of the HDL Receptor SR-BI Alters Plasma HDL and Bile Cholesterol Levels. *Nature* **1997**, *387*, 414–417.

(215)   Remaley, A. T.; Wong, A. W.; Schumacher, U. K.; Meng, M. S.; Brewer, H. B.; Hoeg, J. M. O-Linked Glycosylation Modifies the Association of Apolipoprotein a-II to High Density Lipoproteins. *J. Biol. Chem.* **1993**, *268*, 6785–6790.

(216)   Calvo, C.; Talussot, C.; Ponsin, G.; Berthezene, F. Non Enzymatic Glycation of Apolipoprotein a-I. Effects on Its Self-Association and Lipid Binding Properties. *Biochem. Biophys. Res. Commun.* **1988**, *153*, 1060–1067.

(217)   Wopereis, S.; Grunewald, S.; Morava, E.; Penzien, J. M.; Briones, P.; Garcia-Silva, M. T.; Demacker, P. N. M.; Huijben, K. M. L. C.; Wevers, R. A. Apolipoprotein C-III Isofocusing in the Diagnosis of Genetic Defects in O-Glycan Biosynthesis. *Clinical Chemistry* **2003**, *49*, 1839.

(218)   Debets, M. F.; van Berkel, S. S.; Schoffelen, S.; Rutjes, F. P. J. T.; van Hest, J. C.

M.; van Delft, F. L. Aza -Dibenzocyclooctynes for Fast and Efficient Enzyme PEGylation via Copper-Free (3+2) Cycloaddition. *Chemical Communications* **2010**, *46*, 97–99.

(219)   Aich, U.; Loganathan, D. Synthesis of N-(B-Glycopyrano-Syl)Azidoacetamides. *Journal of Carbohydrate Chemistry* **2005**, *24*, 1–12.

Appendix of NMR Spectra

3.11a

4.082
4.070
4.058
4.046
3.791

1.948
1.944
1.940
1.936
1.932
1.403
1.391

1.00

3.16

3.11a

N₃ — CH(CH₃) — C(=O)OH

—172.906

—118.373

⌐57.887
⌐57.146

—17.088
⌐1.734
⌐1.595
⌐1.458
▪1.320
⌐1.182
⌐1.044
⌐0.907

f1 (ppm)

3.11c

3.962
3.949
3.942
3.929
3.793
1.953
1.947
1.940
1.934
1.928
1.901
1.888
1.883
1.870
1.864
1.851
1.847
1.834
1.829
1.816
1.810
1.797
1.788
1.769
1.751
1.732
1.715
1.696
1.678
0.991
0.973
0.954

1.901
1.888
1.883
1.870
1.864
1.851
1.847
1.834
1.829
1.816
1.810
1.797
1.788
1.769
1.751
1.732
1.715
1.696
1.678

3.11c

—172.375

—118.345

—63.710
—57.134

—25.493

—10.371
—1.940
—1.733
—1.527
—1.320
—1.114
—0.907
—0.701

f1 (ppm)

3.11d

3.11d

—172.066

—118.323

—63.701

—57.092

—25.483

—10.362
1.939
1.732
1.526
1.320
1.113
0.906
0.700

150

3.11e

9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
f1 (ppm)

0.94
1.00
3.09
3.13

3.895
3.886
3.794

2.203
2.192
2.181
2.170
2.159
2.149
2.138
2.127
1.948
1.944
1.940
1.936
0.999
0.988
0.931
0.920

2.22 2.20 2.18 2.16 2.14 2.12
f1 (ppm)

2.203
2.192
2.181
2.170
2.159
2.149
2.138
2.127

3.11e

—171.757

—118.315

—68.404

—57.134

—31.501

—19.617
—17.797
—1.733
—1.595
—1.457
—1.320
—1.182
—1.044
—0.907

f1 (ppm)

3.11f

3.791
3.529
3.513
3.497

2.557
2.541
2.525
1.953
1.946
1.940
1.934
1.928

2.00
2.04

f1 (ppm)

3.11f

—172.821

—171.788

—118.337

—57.094

—47.610

—33.867

1.940
1.733
1.527
1.320
1.114
0.907
0.701

f1 (ppm)

f1 (ppm)

3.11g

155

3.11g

—173.302

—118.354

57.119
55.322

—40.906

19.588
2.147
1.871
1.596
1.320
1.045
0.770
0.494

f1 (ppm)

3.12a

4.545
4.533
4.521
4.509

2.796

1.948
1.944
1.940
1.936
1.931
1.568
1.556

1.00
4.29
3.42

f1 (ppm)

3.12a

170.599
168.357

118.315

56.422

26.399

17.285
1.732
1.595
1.457
1.320
1.182
1.044
0.906

f1 (ppm)

3.12b

159

3.12b

170.581
168.292
118.340
56.371
26.360
17.257
1.939
1.733
1.526
1.320
1.113
0.907
0.700

f1 (ppm)

3.12c

2.025
2.010
2.008
2.006
1.993
1.990
1.975
1.971
1.965
1.955
1.952 CD3CN
1.947 CD3CN
1.941 CD3CN
1.936 CD3CN
1.928 CD3CN
1.918
1.901
1.883
1.865

4.440
4.425
4.422
4.407

2.801

2.025
2.010
2.008
2.006
1.993
1.990
1.975
1.971
1.965
1.955
1.952 CD3CN
1.947 CD3CN
1.941 CD3CN
1.936 CD3CN
1.928 CD3CN
1.918
1.901
1.883
1.865
1.087
1.069
1.050

1.00
4.00
1.76
3.11

3.12c

—170.603
—167.763

—118.296

—61.992

—26.396
—25.911

—9.865
—1.939
—1.732
—1.526
—1.320
—1.113
—0.907
—0.700

210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -1
f1 (ppm)

3.12d

3.12d

—170.646
—167.792
—118.307
—62.032
—26.417
—25.926
—9.869
—1.940
—1.733
—1.527
—1.320
—1.114
—0.907
—0.701

3.12e

168.618
165.872

77.372
77.160
76.948

66.181

31.675
25.720
19.180
17.899

f1 (ppm)

3.12f

3.673
3.658
3.643

2.912
2.897
2.881
2.776

1.940

3.12f

170.912
168.278

118.374

57.120

47.024

31.615
26.349

1.940
1.733
1.527
1.320
1.114
0.907
0.700

f1 (ppm)

3.12g

3.12g

— 170.921
— 167.542

— 118.340

— 57.116
— 54.928

— 38.178

— 26.368

— 19.373

— 1.938
— 1.733
— 1.527
— 1.320
— 1.114
— 0.907
— 0.701

f1 (ppm)

170

3.8

7.971
7.951

6.005
6.000
5.995
5.990
5.969
5.781
5.772
5.764
5.755

4.790
4.402
4.389
4.379
4.366
4.305
4.300
4.289
4.283
4.271
4.265
4.260
4.253
4.246
4.239
4.232
4.224
4.222
4.220
4.218
4.203
4.189
4.047
4.039
4.000
3.993
3.973
3.965
3.819
3.801
3.790
3.769
3.755
3.739
3.726
3.357
3.329

1.09

2.06
1.00

2.01
4.10
1.05
1.05
2.22
0.96

f1 (ppm)

4.389
4.379
4.305
4.300
4.271
4.265
4.260
4.232
4.224
4.222
4.220
4.218
4.203
4.047
4.039
4.000
3.973
3.790
3.769
3.755

2.01
4.10
1.05
1.05
2.22
0.96

f1 (ppm)

3.8

—166.494

—151.988

—141.553

—102.583

93.981
93.926
—88.591
83.063
82.973
73.699
72.207
—69.551
68.043
67.436
65.086
65.028
60.871
50.840
50.753

65.086
65.028

—60.871

50.840
50.753

93.981
93.926

—88.591

83.063
82.973

f1 (ppm)

3.8

— 141.542

— 102.560

94.200
94.142
— 88.522
83.066
82.976
73.698
72.162
— 69.526
68.069
67.646
65.022
64.968
60.880
50.827
50.738

94.200
94.142

— 88.522

83.066
82.976

65.022
64.968

— 60.880

50.827
50.738

f1 (ppm)

f1 (ppm)

f1 (ppm)

3.8

-11.150
-11.278
-13.116
-13.247

-11.150
-11.278
-13.116
-13.247

2.1

—171.138
—166.220

—151.787

—141.669

—102.622

94.590
94.527
—88.529
83.217
83.127
73.756
72.104
—69.632
68.368
67.514
65.043
64.988
61.025
51.607
49.937
49.855

2.1

-11.159
-11.288
-12.793
-12.923

150 130 110 90 70 50 30 10 -10 -30 -50 -70 -90 -110 -130 -150 -170 -190 -210 -230 -250
f1 (ppm)

177

3.13a

3.13a

—174.435

—166.234

—151.812

—141.671

—102.627

94.732
94.669
88.501
83.266
83.174
73.790
72.120
69.668
68.514
67.504
65.045
64.989
61.016
58.197
49.944
49.861

—16.892

65.045
64.989

61.016

58.197

49.944
49.861

94.732
94.669

88.501

83.266
83.174

f1 (ppm)

3.13a

-11.234
-11.316
-13.028
-13.109

f1 (ppm)

3.13b

3.13b

- —174.287
- —166.351
- —151.937
- —141.815
- —102.770
- 94.668
- 94.607
- —88.660
- 83.384
- 83.293
- 73.891
- 72.213
- —69.774
- 68.541
- 67.500
- 65.132
- 65.087
- 61.138
- 58.380
- 50.131
- 50.052
- —16.611

65.132
65.087

61.138

58.380

50.131
50.052

94.668
94.607

88.660

83.384
83.293

f1 (ppm)

182

3.13b

-11.295
-11.425
-13.024
-13.154

140 120 100 80 60 40 20 0 -20 -40 -60 -80 -100 -120 -140 -160 -180 -200 -220 -240
f1 (ppm)

183

3.13c

3.13c

—173.571
—166.233
—151.807
—141.679
—102.617
94.818
94.756
—88.498
83.264
83.175
73.787
72.097
—69.665
68.574
67.482
65.044
64.992
64.001
61.006
49.920
49.835
—25.100
—9.159

65.044
64.992
64.001
61.006
49.920
49.835

94.818
94.756
88.498
83.264
83.175

3.13c

-11.091
-11.220
-12.950
-13.079

3.13d

—173.457
—166.287
—151.847
—141.689

—102.645
94.688
94.625
—88.504
83.268
83.177
73.771
72.041
—69.643
—68.408
—67.368
—64.989
—64.935
—64.137
—60.957
—50.000
—49.913
—24.739
—9.334

94.688
94.625
88.504
83.268
83.177

64.989
64.935
64.137
60.957
50.000
49.913

3.13d

-11.277
-11.408
-13.140
-13.270

190

3.13e

-11.166
-11.252
-13.078
-13.163

3.13f

3.13f

—174.283

—166.186

—151.759

—141.648

—102.597

94.736
94.673
—88.519
83.201
83.112
73.749
72.073
—69.617
68.406
67.599
65.018
64.968
61.020
49.780
49.699
47.066
—34.842

3.13f

-11.211
-12.837

3.13g

7.942
7.923

5.953
5.935
5.524
5.521
5.518

4.790
4.334
4.253
4.245
4.199
4.168
4.011
3.979
3.948
3.925
3.775
3.745
3.728
2.542
2.525

1.281
1.266

195

3.13g

—173.920

—166.206

—151.758

—141.653

—102.600

94.683
94.626
—88.468
83.203
83.103
73.743
72.046
—69.607
68.401
67.493
64.989
64.939
60.998
54.972
49.777
49.702
41.989

—18.589

f1 (ppm) inset:
64.989
64.939
—60.998
—54.972
49.777
49.702

f1 (ppm) inset:
94.683
94.626
—88.468
83.203
83.103

3.13g

-11.279
-11.409
-12.926
-13.056

f1 (ppm)