ELUCIDATING N-GLYCAN MICROHETEROGENEITY THROUGH SITE-SPECIFIC MASS SPECTROMETRY APPROACHES

by

TREVOR MATTHEW ADAMS

(Under the Direction of Lance Wells)

ABSTRACT

N-glycans are essential to eukaryotic biology and are important modulators of glycoprotein folding, stability, binding, and other functions. Glycoproteins often host multiple sites of N-glycosylation, each of which is modified by an extensive non-template process as the glycoprotein folds and travels through the secretory pathway. Each site of N-glycosylation on an individual glycoprotein hosts its own distinct distribution of glycoforms which often denote details of their incomplete processing, a phenomenon termed 'microheterogeneity'. Due to the importance of N-glycans in the development of biologics and biosimilars, there is great interest in both understanding and controlling N-glycan microheterogeneity to improve the quality control and efficacy of therapeutics. This dissertation looks to understand the underlying causes of microheterogeneity from the perspective of the glycoprotein that is modified and with a focus on the stages of processing that delineate the three major classes of N-glycan: high-mannose, hybrid, and complex. By utilizing modern high-resolution high-accuracy mass spectrometry methods and glycopeptide-focused data analysis, we characterized a set of reporter glycoproteins that host varying numbers and diversities of N-glycans, totaling 38 sites of N-glycosylation. Following the development of this enzyme-specific activity database we looked to identify

surrounding structural features that influence N-glycan processing. We identify the importance of protein tertiary structure in defining the relative efficiencies of glycan-processing enzymes towards differing sites. Using protein surface modeling, we identify convexity as a predictor of N-glycan processing efficiency. We also find that proximal phenylalanines may influence the balance between high-mannose and hybrid/complex structures. Additionally, we share a new tool (ppmFixer) for the mass spectrometry search engine pGlyco that dramatically improves the accuracy of its output. Future studies will continue to characterize downstream glycan-processing enzymes as well as the structural determinants that define glycoprotein-glycosyltransferase reactions.

INDEX WORDS: Glycoproteins, N-glycosylation, Microheterogeneity, Mass Spectrometry, Glycobiology

ELUCIDATING N-GLYCAN MICROHETEROGENEITY THROUGH SITE-SPECIFIC MASS SPECTROMETRY APPROACHES

by

TREVOR MATTHEW ADAMS

B.S., University of Georgia, 2016

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

ELUCIDATING N-GLYCAN MICROHETEROGENEITY THROUGH SITE-SPECIFIC MASS SPECTROMETRY APPROACHES

by

TREVOR ADAMS

Major Professor:	Lance Wells
Committee:	Kelley Moremen
	Natarajan Kannan
	Robert Haltiwanger

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2023

ACKNOWLEDGEMENTS

I would like to thank Lance Wells, my mentor who has helped guide me through the murky world of glycobiology and graduate school. I appreciate his patience, understanding, and guidance he has given me, but above all else he has helped me gain the confidence that shapes me into the scientist I am today. I would like to thank my current committee members: Kelley Moremen, Bob Haltiwanger, and Natarajan Kannan. I would also like to thank Michael Tiemeyer and Gerry Hart for their guidance on the O-GlcNAc side of things and for challenging me to think bigger.

To Hannah Stephen, Linda Zhao, and Digantkumar Chapla: thank you for letting me learn and collaborate with you. To my labmates who have come and gone over the years: you all have a special place in my heart. In no particular order: Linda Zhao, Anna Grady, Hannah Stephen, Chelsea Desbiens, Stephanie Halmo, Osman Sheikh, Alyssa Wright, Sally Boyd, Jeremy Praissman, Erin Suh, Johnathan Mayfield, Terrell Carter, David Steen, Ashley Carter, Naomi Hitefield, Xiaolin Dong, Jeff Fairley, Sydney Bedillion and a big thank you to Rob Bridger for always being there when the Orbitraps bewildered me.

I'd like to thank my parents for their unrelenting support and patience during the past quarter century of my education and my brother for letting me join his D&D campaign despite being distracted with manuscripts. And of course, I have to thank my cats: Bustelo, Mr. Wilson, and Steve. Finally, I'd like to thank my loving partner Cassidy Reeser for always being there and understanding, no matter how bright or how gloomy the day is. None of this would be here without you.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW: THE BIOLOGICAL BASIS OF N-GLYCAN MICROHETEROGENEITY AND ITS ANALYSIS VIA MASS SPECTROMETRY

Adams, T.M., Zhao, P., Wells, L. Submitted to *Molecular Cellular Proteomics*.

Abstract

N-glycosylation is an abundant and essential post-translational modification that is preserved across all eukaryotes. N-glycans have important functions in protein stability and protein-protein interactions. N-glycans exhibit a high degree of heterogeneity, even within an individual site on the same protein, a phenomenon that is termed 'microheterogeneity' that is the focus of this review. Much of what we know about N-glycan microheterogeneity is informed by in vitro experiments with free N-glycans that monitor the rate of transfer to specific branches. Traditional analytical approaches with released glycans are limited in their usefulness in studying microheterogeneity due to most glycoproteins having more than one site of N-glycosylation. Since specific N-glycans at specific sites can confer important functions to glycoproteins, this presents a significant gap between the information content of glycomics and glycoproteomics experiments. More recently, intact glycopeptide analyses using tandem mass spectrometry is used to analyze glycans while retaining site-specific information. The microheterogeneity of glycoproteins presents a significant analytical challenge not only during mass spectrometry analyses but also in downstream data processing. Use of specialized search engines followed by extensive manual validation are required for accurate and in-depth glycoproteomics. Overall, recent advances in analytical technology and data processing present exciting new opportunities to analyze N-glycans in a site-specific manner. Understanding the underlying basis of N-glycan microheterogeneity has implications across many fields, including the development of glycoprotein biologics.

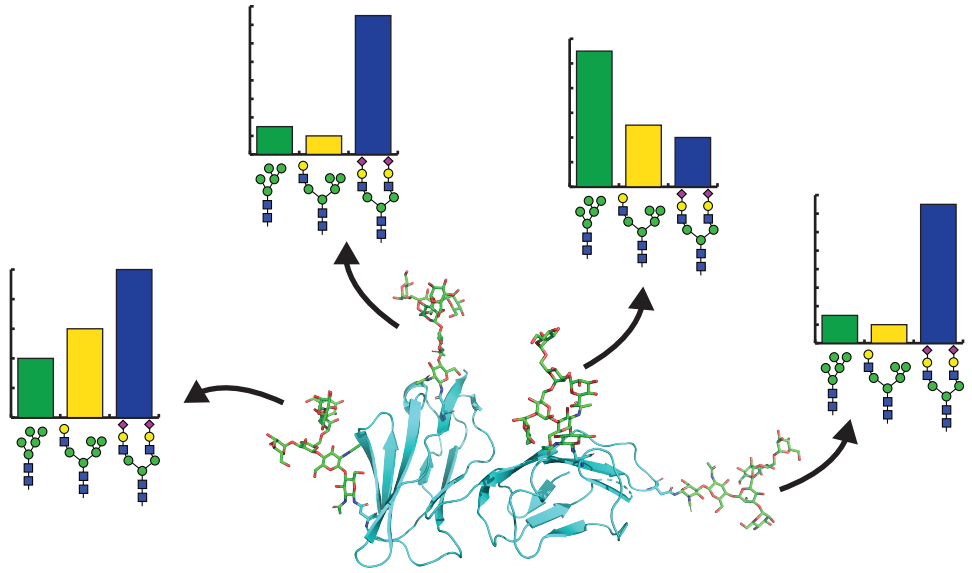


Figure 1.1: Graphical abstract.

Introduction

N-linked glycosylation is an essential eukaryotic post-translational modification that is required for proper embryogenesis and glycoprotein stability (1). A significant amount of cellular resources are dedicated to the production of N-glycans and their nucleotide sugar precursors (2). One reason why glycan processing is important to higher eukaryotes is that it serves to distinguish host tissue from bacterial and yeast cells, which present their own distinct classes of glycans (3). For this reason, it is both a clinical and regulatory priority that glycosylation of manufactured biologics have consistent, human-compatible glycosylation (4). Thus, it is important that we understand the biological basis of N-glycan microheterogeneity (Fig. 1.1). It should be noted that O-glycans also have microheterogeneity, but this topic is beyond the scope of this review.

Due to their size, diversity, and numerous biological functions, N-glycans have been a point of focus for glycobiologists for several decades. N-glycans are initially transferred *en bloc*

to asparagines within the conserved sequence Asn-X-Ser/Thr (termed 'sequons') (5), where X is any amino acid besides proline. The presence of a sequon is usually necessary but not sufficient for N-glycan transfer (6); however there are exceptions such as occasional addition of N-glycans to Asn-X-Cys groups (6), Asn-X-Val groups (6), and potentially even N-G groups (6). This suggests that the oligosaccharyltransferase is able to accommodate a variety of small amino acids to a limited extent near the active site (7). The sugars that predominate in human N-glycans are mannose (Man), N-acetylglucosamine (GlcNAc), galactose (Gal), N-acetylneuraminic acid (Neu5Ac), and fucose (Fuc) though other monosaccharides are incorporated at lower frequencies (Fig. 1.2). The donor species for sugar transfer are nucleotide sugars such as UDP-GlcNAc, GDP-Fuc, or CMP-Neu5Ac. N-glycans have a common $Man_3GlcNAc_2$ core structure that forms a minimum of two branches. The reducing-end GlcNAc, which is attached to the Asn side-chain, can also be fucosylated, a phenomenon termed 'core fucosylation' (Fig. 1.2).

Overview of N-glycan processing

N-glycans are initially transferred to sequons on the lumenal face of the endoplasmic reticulum (ER) during translation. The initial structure is a $Glc_3Man_9GlcNAc_2$ glycan that is quickly trimmed down to $GlcMan_9GlcNAc_2$, a substrate that is recognized by the lectins calnexin and calreticulin, whose interactions with the nascent glycoprotein are important in protein folding (8). The mannoses extending from the α-6 linked branch act as a quality control timer in that one or two terminal mannoses are gradually removed by ER mannosidase I as the glycoprotein remains in the ER; eventually these sub-Man9 glycans are recognized by ER degradation enhancing α-mannosidase-like protein (EDEM) and are subjected to ER-associated degradation. Properly folded glycoproteins are recognized by mannose-binding lectins such as

ERGIC-53 and are transported out of the ER and through the ER-Golgi intermediate compartment (ERGIC).

As the glycoprotein linked to high-mannose N-glycans passes into the Golgi, we begin to see divergence into multiple N-glycan classes. N-glycan structures can be grouped into three categories: high-mannose, hybrid, and complex. High-mannose glycans represent less-processed structures that resemble the initial N-glycan structure transferred to sequons in the ER, albeit minus the terminal glucoses . High-mannose structures are trimmed down by Golgi mannosidase I to form a common $GlcNAc_2Man_5$ structure. This structure marks the branching point between high-mannose and hybrid N-glycans; MGAT1 adds a β,2-linked GlcNAc to the α,3-linked mannose (Fig. 1.2), which forms a hybrid structure that can be further elaborated on that specific branch (9). The mannosidase MAN2A1 can then cleave the two non-reducing end mannoses on the α,6-linked mannose branch, which is termed a paucimannose structure. MGAT2 can then add a β,2-linked

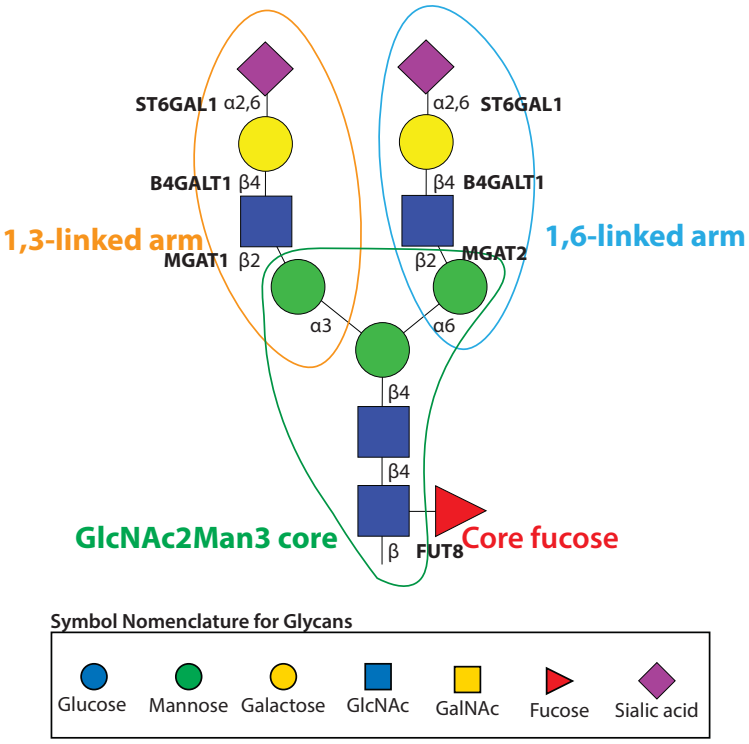


Figure 1.2 – Anatomy of a biantennary complex N-glycan.
The α1,3-linked arm and α1,6-linked arm extend out from the chitobiose core, which may or may not be core fucosylated. Glycans represented using the Symbol Nomenclature for Glycans (SNFG). The conserved core consists of two GlcNAcs extended by a tri-mannose fork. The two mannoses that can be extended into long branches can be thought of as two 'arms', one being 1,3-linked and the other being 1,6-linked. Additionally, the central mannose can be GlcNAc-modified, forming a 'bisected' structure.

GlcNac to the α,6-linked mannose which forms the simplest of complex structures. Biantennary complex structures are one of the more common glycans observed *in vivo,* especially on the Fc domain of antibodies, and a cartoon of a typical complex N-glycan is shown in Figure 1.2.

The addition of GlcNAc to non-reducing end mannoses is the key step in the branching of N-glycans. Additional GlcNAcs can be added to the two branching mannoses through the activity of enzymes such as MGAT4A, MGAT4B, and MGAT5. These GlcNAcs can be extended into sugar chains by enzymes such as B4GALT1, ST6GAL1/2, and ST3GAL1-6, which sequentially add β1-4 linked galactose, α2-6 linked Neu5Ac, and α2-3 linked Neu5Ac, respectively. Alternatively, lactosamine (LacNAc) structures can extend out from branching GlcNAcs as repeating units of galactose-β1,4-GlcNAc. Sialic acids are seen as 'capping' structures and are rarely further elaborated. The GlcNAc-transferase MGAT3 adds a bisecting GlcNAc to the central mannose in the trimannose core. MGAT3 activity truncates N-glycan processing and is inhibited by the extension of branches with galactose (10). α1,6-linked core fucosylation of the Asn-bonded GlcNAc by the fucosyltransferase FUT8 often requires a terminal GlcNAc on the α3-linked mannose, and hence requires the activity of MGAT1 (11), although interestingly there are exceptions with certain fully-folded proteins (12). Like sialic acid, fucosylation of the branching arms is usually a capping structure in human N-linked glycans that does not allow for further extension.

Historical Perspective

The earliest references to microheterogeneity in literature registered in PubMed refer to protein microheterogeneity due to small distributions of differential activity, size, or separation within a purified sample of protein (13). In some of these cases, it was discovered that the

underlying source of these differences were due to sugar modifications creating subpopulations of glycoproteins (14, 15). The term ‘microheterogeneity’ began being applied to glycans as advances in glycan characterization developed during the 1960s eventually leading to an appreciation of the number of possible structural variants that can exist at any given site on a glycoprotein (16).

Studies of readily obtained glycoproteins such as ovalbumin (17, 18) and fetuin (19) demonstrated common linkages between observed sugars, but slight variations in the composition of the glycan led to some confusion of what the representative structure could be. This is largely because studies at the time were limited to identifying relative contributions to the total carbohydrate content of a population of glycoproteins via digestion with broad-specificity proteases, hydrolysis/hydrazinolysis, and paper chromatography. Later experiments with thyroglobulin by Spiro & Spiro investigated microheterogeneity and found that the glycoprotein contained two distinct glycan units which were separable with extensive dialysis (20). Further, it was demonstrated that much of this heterogeneity could be homogenized with neuraminidase treatment, demonstrating the contribution of sialic acids to observed protein heterogeneity in electrophoeresis (20, 21). Along with studies characterizing ovalbumin (22, 23), immunoglobulin G (IgG) (24), and other proteins (15, 25–32) it was eventually recognized that this heterogeneity was likely due to incomplete processing by glycosyltransferases rather than amino acid-like substitutions due to the specificity of glycosyltransferases (16, 33, 34), and that this incomplete processing is likely influenced by protein configuration (30).

In the early 1970s, new analytical techniques such as isoelectric focusing were applied and additional proteins were shown to demonstrate neuraminidase-sensitive microheterogeneity (35–37). It was also shown that the carbohydrate content of glycoproteins in mouse embryos

may be altered over the course of development (38) due to an increase in sialyltransferase activity (39). The identification of various specific endo- and exo-glycosidases also became an invaluable tool for carbohydrate identification during this time (40). Work by the Kobata group eventually led to the elucidation of high-mannose structures (41), characterization of hybrid structures (42), and elucidation of complex biantennary N-glycan structures (43), a notable step forward from the compositional analysis of the 1960s. At this time, enough site-specific N-glycan data became available that it became possible to look at general trends of N-glycans within the context of the glycoprotein's primary structure. Pollack and Atkinson found that high-mannose glycans tended to be more common toward the C-terminal end of proteins, suggesting that protein folding may impact the accessibility of N-glycans (44), a finding that was supported by other statistical studies (45).

Mass spectrometry combined with methylation (46) and chromatographic separation also took several steps forward at this time that facilitated analyses of glycans and glycoproteins. Fast atom bombardment-mass spectrometry (FAB-MS) and glycosidase treatment were used to determine the structure of N-linked lactosaminoglycans (47, 48). FAB-MS was used in conjunction with neutral gas collision in order to generate fragments that were informative of oligosaccharide structure, even between isomeric groups (49). FAB-MS, combined with EndoH treatment and classical peptide separation techniques, was shown to be capable of deducing which sequons were occupied on a multiply glycosylated yeast invertase (50). Electron impact-field desorption mass spectrometry was used to determine the structure of larger high-mannose N-glycans such as $GlcNAc_2Man_8$ (51). NMR analysis was also used to identify less populous species that contribute to N-glycan microheterogeneity in ovalbumin (52). While all of these approaches have contributed greatly to our understanding of N-glycans and glycoproteins in

general, it is important to note that many of them do not consider site-specific N-glycosylation, in which detailed analysis has only been recently made possible with the development of new analytical techniques (see section *Site-specific analysis of N-glycans*).

Biological determinants of N-glycan microheterogeneity

Impact of cell type on N-glycan microheterogeneity

Cell type has been shown to impact N-glycan microheterogeneity. Early studies with vesicular stomatitis virus (VSV) showed differences in glycan composition when the VSV glycoprotein was expressed in different cell types (53). Thy-1 (54) and OX2 (55) glycoproteins purified from rats have different carbohydrate compositions when purified from brain versus thymocytes. Purified pig endopeptidase-24.11 has been shown to have differential glycosylation when purified from intestine rather than kidney, particularly with respect to fucose content (56). Galactosylation of human IgG fluctuates based on the age of the donor, among multiple other factors, indicating that site-specific glycosylation can change over the life of an individual (57–59). However, many of these early studies lacked the instrumentation necessary to deduce site-specific glycan data for these multiply N-glycosylated proteins and were thus limited to glycoprotein-specific information rather than site-specific information.

Studies of γ-glutamyl transpeptidase by the Kobata group revealed inter-organ differences in the glycosylation of the protein expressed in liver vs. kidney, with the liver glycans all containing sialic acid groups and the kidney all neutral (60).Additional studies with Thy-1, this time empowered by NMR spectroscopy, demonstrated that thymus-expressed Thy-1 is sialylated while brain-expressed Thy-1 is not (61). Interestingly, neural-expressed Thy-1 appears to have strong conservation of glycan-type across species, even though there are

significant inter-species differences at the polypeptide level (62). Using Sindbis virus glycoproteins, Hakimi et al. showed that related glycoproteins translated in the same compartment can result in differential glycosylation of high-mannose glycopeptides, suggesting the required mannosidases have varying activities towards glycan substrates (63).

A comparative analysis of native human and recombinantly expressed fetuin clearly demonstrates the impact that cell type can have on post-translational modifications. Lin et al. expressed recombinant human fetuin (rhFetuin) in HEK293 cells and used a combined native and bottom-up mass spectrometry approach to compare it to human fetuin from serum (hFetuin), which is mostly synthesized in the liver (64). They found that rhFetuin was more likely to express terminally galactosylated and core fucosylated structures while hFetuin expressed mostly sialylated structures. Some of these differences may arise from slightly different processing of the protein moiety in the two cell types, leading to structural differences. Comparison of closely related major histocompatibility complexes (MHCs) using radiolabeled sugars and site-specific release of glycans demonstrated that small structural variations can have dramatic consequences on the extent of branching found in complex glycans, and that glycosylation patterns are reproducible when expressed in the same tissue over time (in the case of this study, several months) (65).

Impact of species on N-glycan microheterogeneity

Studies of γ-glutamyl transpeptidase by the Kobata group revealed inter-species differences in transpeptidase glycosylation between rats, cattle, and mice (66). An interesting finding was the presence of bisected N-glycans in mouse and human kidney that were absent in bovine and rat kidney γ-glutamyl transpeptidase (67, 68). A study of viral glycoproteins

demonstrated significant differences in carbohydrate composition depending on the host cell that the viral glycoprotein was expressed in, and site-specific analysis of trypsin-digested peptides revealed that this composition was due to diversity within individual sites (69). A comparison of bovine and human fetuin showed similarities at shared sites, but slight differences in branching preferences (64).

Efficiency of Core Glycosylation

The initial transfer of the Glc3Man8GlcNAc2 to the N-X-S/T sequon on acceptor glycoproteins by OST is a key step in producing heterogeneity in that incomplete transfer, which is a common occurrence, will result in a subpopulation of unglycosylated sequons. The proportion of occupied vs. unoccupied sites can vary drastically even on a single glycoprotein, as illustrated with the ranging occupancies of HIV Env glycoprotein (70). For some glycoproteins, the specific nature of the glycans themselves is often of less importance than their general presence. For viral glycoproteins like influenza A virus hemagglutinin, glycans shield lysines, arginines, and aromatic residues that are the target of defensive host proteases like trypsin and chymotrypsin for delivery of glycopeptides to the MHC (71, 72). Indeed, inhibition of glycosylation with tunicamycin increases the rate of influenza nucleoprotein degradation (73). However, for complex organisms processing of the glycan can be important for proper development. Mice lacking MGAT1 die before coming to term, indicating that the formation of hybrid and complex N-glycans rather than the presence of N-glycans in general is important for higher eukaryotic development (74). Complex-type N-glycans are expressed as early as the 4-to-8 cell stage of embryonic development, and mouse embryos cultured with the OST inhibitor

tunicamycin fail to develop past 9-11 days (75). The interplay between site occupancy and N-glycan microheterogeneity has been termed “macro-heterogeneity” (76).

As this initial step of N-glycosylation often happens co-translationally, it may be helpful to consider this transfer within the context of a nascent polypeptide rather than a mature, folded glycoprotein. Transfer of the oligosaccharide by OST takes place within 30-40Å of the ER lumen, which is roughly the length of 12-14 peptide residues (77). For this reason, it is likely that in some cases core glycosylation may be in competition with protein folding in the context of secondary structure formation. This is supported by the lack of glycosylation found on cysteines known to participate in disulfide bonds, and that limiting disulfide bond formation can promote glycosylation at sites that are proximal to disulfide bonds (78). Consequently, one can imagine a sequon that is synthesized after the formation of several disulfide bonds will be presented to OST in a different structural context than one synthesized before the formation of disulfide bonds. The primary sequence of glycoproteins seemingly has the most impact on glycans at the point of core glycosylation. The presence of a threonine rather than a serine in the +2 position of the sequon increases the efficiency of core glycosylation of rabies virus glycoprotein, although it is unclear how this impacts further processing by Golgi enzymes (79). Also, roughly 75% of unoccupied sequons have a serine in the +2 position (80). Sequons with aromatic residues in the -2 and -1 positions are more likely to be glycosylated (80). The presence of proline in the either ‘X’ position or immediately C-terminal to the hydroxy-amino acid (S/T) abrogates core glycosylation (81). Close proximity to the signal sequence may also be responsible for poor core glycosylation efficiency (82). On the other end of the nascent peptide, C-terminal N-glycans are thought to interact more weakly with STT3A (OST-A) due to the formation of secondary and tertiary structure as the protein nears full synthesis, and these C-terminal sequons are thought to

be modified post-translationally primarily by STT3B (OST-B) (7). In further studies with rabies virus glycoprotein, it was shown that the 'X' moiety can have some impact on site occupancy as large, hydrophobic residues such as tryptophan appear to reduce the efficiency of core glycosylation (83). These negative effects on core glycosylation tend to be more pronounced at sequons that contain a serine rather than a threonine in the +2 position (80). Moving forward, knowledge of primary and tertiary structural features that help define core glycosylation could be utilized to engineer novel sites of N-glycosylation into glycoprotein targets, such as biologics.

Forks in N-glycan processing

As glycoproteins travel through the secretory pathway, their glycans are presented to glycosyltransferases and glycosyl hydrolases with varying activities and substrate specificities. These glycan-processing enzymes compete with one another to process glycan substrates along differing pathways that may or may not be mutually exclusive. The easiest way to grasp the flux of N-glycans during their processing is by recognizing "forks" in processing. These forks are best understood with the system first described by Schachter in his excellent review (84), in which a fork can either mark a point at which a glycan is no longer a substrate for certain enzymes (GO-NOGO) or a point at which a glycan becomes a valid substrate for certain enzymes (NOGO-GO).

Two major GO-NOGO forks take place during early cis-Golgi N-glycan processing. The first is the addition of a bisecting GlcNAc onto $GlcNAc_3Man_5$ structures blocks the action of MAN2A1, hence locking the glycan into a hybrid structure (85). This bisecting GlcNAc also blocks core fucosylation by FUT8 (11) and additional branching of the α-3 linked mannose via

MGAT4 (86). B4GALT1 activity is also reduced, but not totally ablated, by the presence of a bisecting GlcNAc (87).

The addition of blood group antigens have their own GO-NOGO forks. The presence of an α-3 or α-4 linked fucose to the β-2 linked GlcNAc blocks the formation of A, B, and H(O) antigens on the terminal galactose of the N-glycan (84). Thus, the competing activities and specificities of glycan-processing enzymes can directly impact host recognition and host-pathogen interactions.

The addition of a β-4 linked galactose to the α-3 linked mannose blocks the activity of MAN2A1 (85), MGAT2 (9), MGAT3 (86), MGAT4 (86), MGAT5, and FUT8. Hence, the activity of B4GALT1 can take a nascent hybrid structure (such as the MGAT1 GlcNAc2Man5 product) and lock it into a hybrid conformation by blocking pathways to complex glycosylation. These restrictions can be rationalized by considering many these enzymes require a common recognition branch of GlcNAcβ1-2Manα1-3Manβ1-4 for efficient transfer (84). During the conversion of high-mannose to hybrid and complex N-glycans, it is important to differentiate between the α-3 linked mannose and the α-6 linked mannose in the $GlcNAc_2Man_3$ core. For instance, B4GALT1 preferentially transfers to the α-3 branch of the glycan compared to the α-6 branch (88, 89). In biantennary glycans, α2-6 linked sialic acid is preferentially added to the α-3 linked branch in comparison to the α-6 linked branch (90). Another example of competing specificities can be found in the interplay between polyLacNAcylation and sialic acid linkage. By separating N-glycans by LacNAc chain length through tomato lectin binding, Merkle and Cummings found that in shorter LacNAc chains α2-6 linked sialic acid is favored while the α2-3 linkage is more abundant in longer LacNAc chains (91).

NOGO-GO forks occur when the addition of a sugar converts a glycan nonsubstrate into a glycan substrate for a specific set of enzymes. The most important NOGO-GO fork occurs with the activity of MGAT1, which adds a terminal β-2 linked GlcNAc to the α-3 linked mannose in the GlcNAc2Man3 core. The activity of MGAT1 is necessary for the formation of hybrid and complex structures, and the presence of the terminal GlcNAc is required for activity by MGAT2, MGAT3, MGAT4, and in some cases FUT8 (84). It should be noted that these are general rules that apply to N-glycan processing at large and exceptions exist in which these forks are contradicted. However, these oddball products often represent small fractions of the overall glycoprotein population and are likely not significant contributors to the glycoprotein's overall function.

Impact of secondary structure on N-glycan microheterogeneity

The initial transfer of N-glycans to N-X-S/T sequons seems to be promoted by the acceptor being an unfolded peptide, similar to its state during co-translational transfer (92). It is perhaps for this reason that N-glycans are frequently found on β-turns (6, 93) where the underlying peptide backbone remains readily accessible after folding is complete. Additionally, sites of N-glycosylation are enriched at points in which the secondary structure changes (80), which raises the question as to whether glycosylation promotes the formation of these secondary structures. Interestingly, N-glycans are enriched on β-sheets when compared to their distribution with proteins generally (94).

Time-resolved fluorescence energy transfer has been used on glycopeptides to demonstrate that glycosylation can affect the conformations that the peptide is able to adopt, indicating that N-glycosylation alters folding and potentially acts as a nucleation event for

secondary/tertiary structure formation (95). The requirement of N-glycosylation for proper glycoprotein folding is domain-dependent, as has been shown with the HIV surface glycoprotein gp120, whose V1 and V2, but not V3, domains require glycosylation for proper folding, likely due to recognition by calnexin/calreticulin chaperones (96).

Impact of tertiary structure on N-glycan microheterogeneity

It has long been thought that N-glycan processing is impacted by the tertiary structure of the acceptor glycoprotein. Early studies used differential sensitivity to EndoH to illustrate that some sites enriched in high-mannose N-glycans only become accessible to the enzyme when denatured (97). It is reasonable to assume that initial N-glycan transfer in these cases is only possible in the early stages of protein folding.

A common example of protein structure impacting N-glycan processing is found in the glycosylation of IgG heavy and light chains, where the heavy chain N-glycan is mostly unbisected biantennary complex while the light chain N-glycan is predominantly bisected (98). The reason behind this is revealed by the crystal structure of serum IgG. The 3D structure shows that the Manα1-6Man linkage of the heavy chain N-glycan is in a gauche-gauche conformation (ω=-60°) (99) which is not a valid conformation for the bisected MGAT3 product that solely adopts the trans-gauche conformation (ω=+180°) (100). Essentially, the light chain N-glycan is free to adopt an ω of -60° or +180° while the heavy chain N-glycan is restricted by the tertiary structure of the glycoprotein to -60°, restricting its processing. The interactions between IgG glycans and the underlying protein are of great interest as these protein-glycan interactions are important in receptor binding and inducing antibody-dependent cellular cytotoxicity. Modification of the region surrounding the heavy chain N-glycan can improve the antibody's

overall stability (101). Modification of the protein-glycan interface of IgG can lead to more extensive processing of the N-glycan, inducing hypergalactosylation and hypersialylation (102). Replacement of certain hydrophobic residues on IgG Fc with alanine causes an increase in heavy chain galactosylation and sialylation, demonstrating the role of the underlying protein in defining N-glycan heterogeneity (103). For more information on immunoglobulin site occupancy and glycan heterogeneity, please see the recent review by Čaval et al (76).

Much work has been completed recently by the Aebi group in deducing the underlying basis of N-glycan microheterogeneity using the model protein yeast protein disulfide isomerase 1 (Pdi1p). Pdi1p has five sites of N-glycosylation, one of which (site 4) is much less processed than the others, and molecular dynamics simulations indicate that the glycan at this site interacts with the protein backbone of Pdi1b (104). Upon removing the region of the protein that the glycan was found to interact with, ER processing of the glycan improved. The Aebi group has also developed an *in vivo* kinetics model of N-glycan processing using parallel reaction monitoring (PRM) mass spectrometry (105).

Recent work by our group has demonstrated that tertiary structure plays a critical role in dictating N-glycan microheterogeneity (106). We expressed a series of reporter glycoproteins in MGAT1- cells, resulting in an enrichment of the GlcNAc2Man5 glycoform at a total of 38 sites of glycosylation on proteins ranging from having 3 to 22 sites. We then used these purified glycoproteins as substrates for purified glycosyltransferases and MAN2A1 and quantified the relative ratios of substrates and products via tandem mass spectrometry to determine the rate of the reactions over time. Sites that were enriched with less processed structures (e.g. high-mannose and hybrid) were found to be worse substrates for the enzymes than their more-processed counterparts on the same glycoprotein. These transfer rate differences were abrogated

when the reporter glycoprotein was reduced and digested into glycopeptides prior to enzyme addition, indicating that the tertiary structure of the glycoprotein is a key player in microheterogeneity (106).

The quaternary structure of glycoproteins has been shown to impact N-glycan microheterogeneity, which is reasonable considering the additional restrictions on 3-dimensional space that occur as multimers form. The proteins Mac-1 and LFA-1 both have identical β-chains and differing α-chains at the polypeptide level, and both form multimers before reaching the Golgi. The N-glycosylation of the β-chain differs from each multimeric species, indicating that quaternary structure can impact N-glycan processing (107).

Effects of Precursor Availability

One potential source for N-glycan heterogeneity is the availability of nucleotide sugar donors. These precursors have relatively low concentrations in the cytosol (or nucleus in the case of CMP-Sia) where they are metabolized but are concentrated in the Golgi by nucleotide sugar transporters (108). Once inside the Golgi, the nucleotide sugars can be used as substrates for glycosylation. Degradation of these substrates is counterintuitively important for the proper cycling of metabolites due to nucleotide sugar transporters being antiporters that require monophosphate nucleotides (such as UMP or GMP) in order to transport the activated sugar from the cytosol (108). In support of this, inhibition of GDPase in yeast causes defects in N-glycosylation (109). Additionally, local cytosolic concentration of nucleotide sugar donors is controlled pyrophosphatases that degrade nucleotide sugars (for UDP/GDP-sugar groups) (110) or a specific hydrolase (for CMP-Sia) (111), which is thought to act as a way to regulate import of these substrates by antiporters.

However, nucleotide sugar availability cannot be a sufficient explanation for observed N-glycan microheterogeneity. When a glycoprotein with multiple N-glycan sites passes through the secretory pathway, not all thes sites are processed in the same way despite passing through the same subcellular space, and presumably through the same concentrations of nucleotide sugars. For similar reasons, availability of divalent cofactors is also unlikely to be responsible for observed microheterogeneity.

Supplementation of tissue slices with dolichylphosphate promotes core glycosylation generally, indicating that this important N-glycan precursor may not be saturated within the cell (112). CDG type I, which results in global decrease in N-glycosylation levels, can be caused by a defect in the phosphomannomutase that converts glucose to mannose-1-phosphate, a critical substrate for the formation of the sugar nucleotide donor GDP-mannose (113, 114).

Effects of Enzyme Availability

The potential diversity of N-glycans observed is a product of the repertoire of glycan-processing enzymes that are expressed by the cell. In humans, ribonucleases (RNases) isolated from different tissues have slightly different glycan occupancies (115). Specific inhibition of glycosyltransferases contributes to changes in glycosylation seen during development. Inhibition of MGAT1 by GnT1IP promotes the production of high-mannose species during spermatogenesis (116). Studies with Lec1 cell lines, which are deficient in MGAT1, have been critical in understanding the impact of complex N-glycan formation and branching (117). By using cooler culture conditions to slow the trafficking of proteins through the secretory pathway, it is possible to increase the amount of poly-LacNAc groups on lysosomal membrane glycoproteins (LAMPs), presumably via an increase in incubation time with *trans*-Golgi

enzymes (118). However, while the repertoire of glycan processing enzymes dictates the variety of structures that are possible in a cell, it is unlikely that their availability has an impact on differences in microheterogeneity between different sites on the same protein.

Effect of development and disease state on N-glycan microheterogeneity

There is much interest in using glycans as biomarkers for disease, and thus it is important to consider how N-glycan diversity is affected by age and disease state. There is a particular interest in biomarker availability in serum due to its ease and non-invasiveness (119). An important point to remember is that glycoproteins from serum will generally be biased towards sialylated structures, as non-silalylated structures are continuously removed from circulation by the asialoglycoprotein receptor. Thus, the steady-state level of glycosylation in serum, while biologically relevant, is not necessarily an accurate representation of the glycoprotein population that is actually synthesized. Another challenge in site-specific glycopeptide analysis is a well-developed understanding of the normal range of site-specific glycosylation across healthy patients of varying ages, sexes, and ethnic groups. Developing a deeper understanding of these control groups will require close collaboration with clinicians (120).

Site-specific N-glycosylation can have functional consequences in physiology and disease. The abundant urine glycoprotein uromodulin has eight sites of N-glycosylation, six of which are processed to complex structures and two of which remain high-mannose. Uromodulin forms nanometer-scale filaments, and it has been found that only the high-mannose N-glycan on the filamentous arm is recognized by *E. coli* adhesins, which aggregate on the filaments and are dispelled to prevent urinary tract infections (121). Patients with rheumatoid arthritis have been found to have reduced levels of IgG galactosylation and also correlated with disease severity

(122). NMR studies of agalactosylated IgG glycans demonstrate that in the absence of galactose, the underlying glycan is liberated from glycan-protein interactions adopts a more flexible conformation that allows the terminal GlcNAcs to be recognized by the complement-activating lectin mannose-binding protein (MBP) (123). Individuals with sepsis have been found to have an increase in fetuin fucosylation (124).

The effects of glycan flexibility

Glycans are flexible molecules and exist as a population of various conformations. For this reason, crystal structures often do not provide much information on the position of the glycan. Due to this inherent flexibility, molecular dynamics studies can be an enlightening method of exploring the structure-function properties of glycans that are not observable via crystallographic methods (125). N-glycans on NMDA receptors influence the protein structure by stabilizing an "open-clamshell" conformation that is important in facilitating transport (126). Further studies with NMR and mass spectrometry confirmed this finding and also found that processing of two of the N-glycans is restricted by these glycan-protein contacts (127). Recent evidence from studies of the bacterial glycosidase EndoS suggests that the N-glycan in IgG heavy chains is able to adopt a 'flipped-out' conformation that may explain its ability to mature into complex structures despite spatial restrictions (128).

Site-specific analysis of N-glycans

Bottom-up glycoproteomics – LC-MS/MS

Glycoproteomics by liquid chromatography coupled with mass spectrometry presents challenges in both instrumentation and data analysis due to the complexity of glycopeptides

introduced by glycan heterogeneity (Figs. 1.1, 1.3). Traditional approaches in bottom-up glycoproteomics often involve partial or total deglycosylation of glycoproteins and glycopeptides followed by separate analyses of the protein (proteomics) (70, 94, 129–132) and glycan entities (glycomics) (133–138) , which benefits from retaining linkage information but information on site-specific glycosylation becomes lost. Advances of mass spectrometry have allowed confident assignment of glycosylation sites in intact glycopeptides. While collision induced dissociation (CID) results in loss of the glycan moiety at the expense of peptide backbone fragments, electron-based dissociation, such as electron transfer dissociation (ETD) and electron capture dissociation (ECD), produces peptide backbone fragments with any labile modification attached, therefore enabling unambiguous site localization (139–141). Several instrument methods have been developed combining CID and ETD for a comprehensive characterization of intact glycopeptides (142–144) .

The implementation of beam-type CID, known as high energy C-trap dissociation or higher energy collisional dissociation (HCD) (145), in hybrid ion tap-orbitrap mass spectrometers improved the characterization of intact glycopeptides significantly: (1) the generation of distinct Y1 ions (peptide backbone fragments with a single *N*-acetylhexosamine (HexNAc) attached) by HCD provides useful information in localizing the site of glycosylation, (2) the glycan-specific oxonium ions produced by HCD can be used as diagnostic fragments to sequence the glycan moiety, and (3) the secondary fragmentation that is inherent to beam-type dissociation offers backbone fragments to sequence the peptide entity. Scott et al. combined HCD and CID to analyze the N-linked glycoproteome of *Campylobacter jejuni* (142). Zhao et al. combined HCD and ETD for the analysis of O-linked glycosylation (146), which led to the

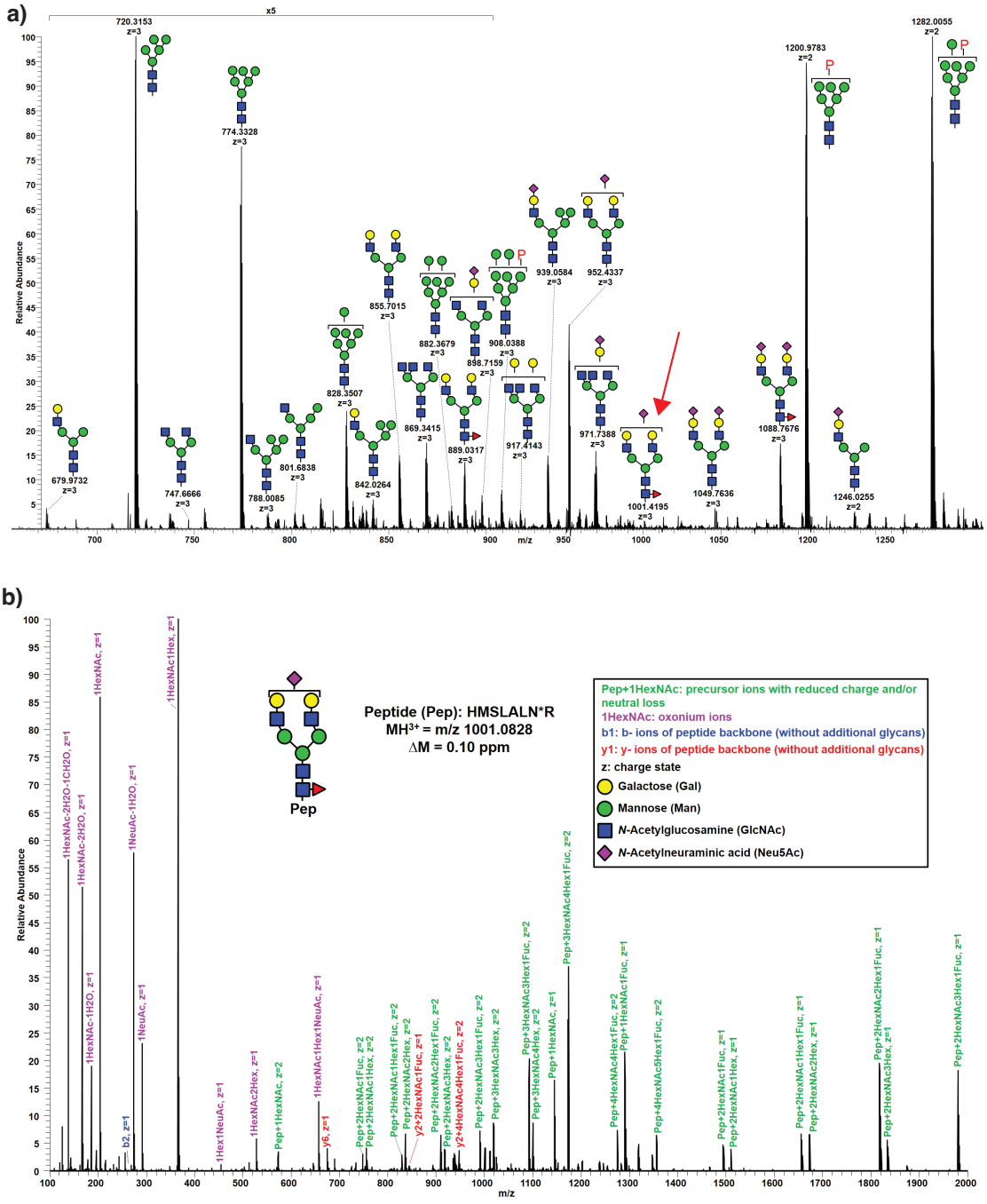


Figure 1.3 – N-glycan microheterogeneity at a single sequon.
A, The presented spectrum was merged from two sets of MS1 spectra averaged from different periods of retention times of a Fabrazyme peptide: (1) 48 MS1 spectra averaged from retention time 94 through 97 min; (2) 15 MS1 spectra averaged from retention time 103 through 104 min. Brackets indicate the precise branch placement of a terminal residue cannot be confirmed, "P" indicates a phosphate group. B, The annotated stepped HCD spectrum of the glycopeptide species at m/z 1001 in Fig. 2A (indicated with red arrow) representing the Fabrazyme peptide modified by a single sialylated biantennary glycan with core fucosylation ($Neu5Ac_1Gal_2GlcNAc_2Man_3GlcNAc_2Fuc_1$).

implementation of HCD product ion-triggered ETD (HCD-pd-ETD) in the hybrid ion trap-orbitrap mass spectrometers from Thermo Scientific™.

However, optimizing a normalized collision energy (NCE) of HCD that is universally suited for fragmenting a wide range of glycopeptides differing in glycan moieties, peptide sequences, and charge states, to yield the highly informative sequential neutral losses of glycans as well as peptide backbone fragments remained as a challenge. Wu et al. evaluated individual HCD NCEs for fragmenting intact glycopeptides, and found different optimal NCE values for glycan moieties and peptide backbones (147). Later on, the option of stepped collision energy (SNCE) became available, where a central value and a variation value are defined by users to obtain three NCE values. Using HCD with SNCE, or stepped HCD, the same precursor ions are injected three times and fragmented with the three NCEs correspondingly, the product ions from three fragmentation are combined and sent to a user selected analyzer for detection resulting a multiplexed spectrum (Fig. 1.3). The strategy of stepped HCD has been widely adopted and applied in intact glycopeptide analysis (106, 148–150). Recently, a more versatile feature of NCE selection mode has been developed to facilitate the fragmentation of small molecules, the assisted collision energy, where a single optimal energy would be selected in real time from a list of energies input by users (151).

Bottom-up glycoproteomics – Data analysis

Advances in dedicated software solutions and bioinformatics tools for high-throughput glycopeptide identification have aided the automation of glycoproteomic analysis in recent years (152–160). However, challenges still remains, particularly for the quality control of database search output, where a false discovery rate (FDR) evaluation is needed for both glycan parts and

peptide parts of the glycopeptide matches. For peptide identification, the target-decoy strategy is usually employed to estimate the peptide and protein FDR, and a sequence-based decoy database generated from the target database is routinely used. Unlike homogenous peptides, glycans are often heterogeneous branched trees due to their non-template-driven synthesis, which creates difficulties in constructing a corresponding tree-based target-decoy database that is necessary for FDR estimation. The use of a FDR for the glycan assignment increases the specificity and sensitivity in identifying complex intact glycopeptides and allows for FDR control of both the peptides and glycans. However, most available software currently only calculates FDR on peptides but not on glycans (152–155, 157). There are only a few exceptions where the glycan FDR is considered.

pGlyco addressed the glycan FDR issue by introducing a spectrum-based target-decoy method to estimate glycan FDR of glycopeptides (156). In this method, the theoretical target glycopeptide spectrum was first generated after the masses of the Y ions (peptide backbone fragments with glycan attached) were deduced based on the putative peptide backbone mass, then a random mass ranging from 1 to 30 Da was added to the mass of each deduced Y ion to generate a theoretical decoy spectrum, and the glycan FDR for the glycopeptides was calculated based matches against both databases. A finite mixture model (FMM) was also employed to adjust the bias from assuming "the number of incorrect identifications from target or decoy sequences are equally likely" (161) in a target-decoy approach (156, 162).

In their latest version of the software, MSFragger also addressed the glycan FDR issue for N-linked glycopeptides by including Y-ions and oxonium ions for glycan identification and by introducing an associated glycan FDR estimation method (158). Their new glycan FDR estimation approach first matches the peptide sequence, then matches the mass difference

between the peptide sequence mass and the observed precursor mass to candidate glycans to determine the composition. After that, pairwise comparison of all candidates determines the best match to the spectrum based on unique fragment ions for each candidate and mass and isotope errors. Finally, a composite glycan score is generated from a variety of spectral evidence, including Y-ions, oxonium ions, and the observed mass and precursor isotope errors, and FDR is computed using the score distribution of target and decoy glycans (158, 163).

A recent evaluation of existing software and informatics solutions for large-scale glycopeptide analysis revealed comparable performance of freeware and commercial products, with similar limitations, especially for matching glycans with similar or identical masses (*N*-acetylneuraminic acid, *N*-glycolylneuraminic acid, multi-fucose, methionine oxidation, cysteine carbamidomethylation) (164). They are frequently mis-assigned by the evaluated search engines, demonstrating the need for improvement in matching isobaric and near isobaric glycopeptides. The search outputs in terms of specificity (accuracy) and sensitivity (coverage) were also variable with different search engines, indicating that orthogonal searches and pool of results could be useful for a comprehensive glycoproteomics analysis. The search parameter settings also contributed to the discrepancy associated with search results, especially the post-processing tools used to filter search results (164). For these reasons, some level of manual validation is often used for confirmation of search engine results.

Utilization of mixed workflows and glycosidases

While glycoproteomics can provide site-specific glycan information and general glycan topology, it can be difficult to differentiate between the compositions of specific glycan antennae since much of the information is derived from fragmentation products via oxonium ions. Mass

spectrometry-driven glycoproteomics can also be facilitated by glycomics, often known as glycomics-assisted or glycomics-informed glycoproteomics (148). In a glycomics-informed glycoproteomics study, the profile of glycans is released by enzymatic treatment or chemical reaction, and analyzed for their linkage and structural details that is often unattainable or ambiguous from the analysis of intact and un-derivatized glycopeptides. This is especially useful for the elucidation of terminal and internal or core structures, such as differentiating core and terminal fucosylation, which is often ambiguous in intact glycopeptide analysis since both features produce the same oxonium ions and needs specific Y ions to clarify. Not only does glycomics provide additional structural information and an orthogonal validation to the glycopeptide analysis, it also helps expand the application of glycoproteomics to complex biological samples such as protein extracts from cells, tissues, and bodily fluid (165).

Linkage-specific glycosidases can also be a useful method in elucidating N-glycan structures with greater specificity. An example of this can be seen with the determination of sialic acid linkages in complex N-glycans. Sialic acids on N-glycans are usually found in either an α-2,6 or an α-2,3 linkage. Since bottom-up glycoproteomics is unable to distinguish between these two different linkages, linkage-specific sialidases are useful tools for determining the identity of specific sialylated structures (166). While most sialidases are either α-2,3-specific or of a broad α-2,3/α-2,6/α-2,8 specificity, the recent development of an α-2,6-specific sialidase may be able to further power these kinds of analyses (167).

A method for determining the relative ratios of occupancy and N-glycan classes while retaining site-specificity is described by Cao et al (168). This approach resembles a standard bottom-up proteomics workflow, but after protease digestion utilizes the differing specificities of endoglycosidases EndoH and PNGaseF in the presence of ^{18}O-water. The end result is that

sequon-containing peptides will exist in some or all of three forms that denotes their occupancy: 1) Asn-GlcNAc (+204 Da difference, high-mannose/hybrid) 2) ^{18}O-Asp (+3 Da difference, complex) and 3) unmodified (+0 Da difference, unoccupied). A major benefit of this approach is that it reduces the heterogeneity of the glycopeptide population into three distinct subpopulations which have a greater signal than any specific intact glycopeptide on its own, thus reducing the required sensitivity of the mass spectrometry analysis. Additionally, it reduces the differences in ionization between differing glycopeptide species which streamlines quantitation.

Other Approaches

Other mass spectrometry approaches are applicable to N-glycopeptides but have their own disadvantages. Native mass spectrometry (native MS) has recently arisen as another way to observe N-glycan microheterogeneity and benefits from the ability to maintain protein complexes during analysis (169). However, it suffers from the disadvantage that site-specific information cannot be obtained for proteins with more than one N-glycosylation site. Advances in native MS can provide valuable information regarding the meta-heterogeneity of N-glycans, and is particularly applicable to glycoproteins which have ample glycomics and glycoproteomics data available (169). This approach can be used in tandem with lectin affinity purification in order to enrich for and characterize specific classes of N-glycans and their binding to lectins(170). For less complex samples such as purified IgG glycopeptides, MALDI-TOF-MS is a valid approach with the benefits of a robust, high-throughput workflow, but at the cost of reduced sensitivity compared to other in-line mass detection methods (171).

Conclusion

Microheterogeneity is an inherent characteristic of N-glycans, and in order to understand these biomolecules in their entirety their diversity needs to be taken into account. N-glycans have been shown to impact glycoprotein function through both the actions of specific glycoforms and through the general properties of N-glycans. Site-specific glycosylation can be altered in disease states and much work still needs to be done to develop useful site-specific glycan biomarkers for early detection of diseases such as cancer.

Bottom-up glycoproteomics is the most direct way to characterize site-specific N-glycan diversity. This is a method that is still rapidly developing as both analytical and computational technologies improve and meet the demanding specificity and sensitivity of glycan analysis. However, this method must be complemented by glycomics and traditional glycosidase treatment to gain a full understanding of linkage information and relative glycan quantities. Other methods that interrogate the structure of glycoproteins such as crystallography, cryo-EM, and molecular dynamics are all necessary to fully understanding the role of the glycan in the context of the glycoprotein.

Advances in understanding what causes a glycoform to be enriched at a specific site are critical for further advances in 'glycoengineering' biologics that have defined and consistent N-glycans. As our technology and understanding of N-glycans grows, look to this field for exciting new opportunities in our understanding of disease and the development of therapeutics.

Acknowledgements

We thank Robert Bridger for assistance with mass spectrometry and glycoproteomics and Dr. Kelley Moremen for his insight into N-glycan microheterogeneity. This work was supported in

part by a grant from the National Institutes of Health (R01GM130915) (to L.W.) and the Glycoscience Training Program (5T32GM107004) from NIGMS (to T.M.A.).

CHAPTER 2

SEQUENTIAL IN VITRO ENZYMATIC N-GLYCOPROTEIN MODIFICATION REVEALS SITE-SPECIFIC RATES OF GLYCOENZYME PROCESSING (106)

---

(106)Adams, T.M., Zhao, P., Chapla, D., Moremen, K.W., and Wells, L. 2022. *Journal of Biological Chemistry.* 10.1016/j.jbc.2022.102474

Abstract

N-glycosylation is an essential eukaryotic post-translational modification that affects various glycoprotein properties, including folding, solubility, protein-protein interactions, and half-life. N-glycans are processed in the secretory pathway to form varied ensembles of structures, and diversity at a single site on a glycoprotein is termed 'microheterogeneity'. To understand the factors that influence glycan microheterogeneity, we hypothesized that local steric and electrostatic factors surrounding each site influence glycan availability for enzymatic modification. We tested this hypothesis via expression of reporter N-linked glycoproteins in N-acetylglucosaminyltransferase *MGAT1*-null HEK293 cells to produce immature $Man_5GlcNAc_2$ glycoforms (38 glycan sites total). These glycoproteins were then sequentially modified *in vitro* from high-mannose to hybrid and on to biantennary, core-fucosylated, complex structures by a panel of N-glycosylation enzymes, and each reaction time-course was quantified by LC-MS/MS. Substantial differences in rates of *in vitro* enzymatic modification were observed between glycan sites on the same protein, and differences in modification rates varied depending on the glycoenzyme being evaluated. In comparison, proteolytic digestion of the reporters prior to N-glycan processing eliminated differences in *in vitro* enzymatic modification. Furthermore, comparison of *in vitro* rates of enzymatic modification with the glycan structures found on the mature reporters expressed in wild type cells correlated well with the enzymatic bottlenecks observed *in vivo*. These data suggest higher-order local structures surrounding each glycosylation site contribute to the efficiency of modification both *in vitro* and *in vivo* to establish the spectrum of microheterogeneity in N-linked glycoproteins.

Introduction

Glycans are important modulators of protein properties and functions across all clades of life (172, 173). N-glycosylation is a conserved and essential co- and post-translational modification in higher eukaryotes (174) and plays an important role in protein homeostasis (175). N-glycans are co- and/or post-translationally attached *en bloc* by oligosaccharyltransferase (176) to the conserved motif N-X-S/T(C), also known as a "sequon", where 'X' can be any amino acid except proline (5, 79, 177). The initial oligosaccharide is mannose-rich and it is trimmed by a series of glycoside hydrolases in the endoplasmic reticulum (ER), eventually exposing the core structure of N-glycans, which is made up of a chitobiose core (GlcNAc-GlcNAc) with branching mannoses, $Man_5GlcNAc_2$. N-glycans are generally categorized as belonging to one of three classes based on the extent of their processing: high-mannose, hybrid, or complex. High-mannose glycans are the least processed and most closely resemble the initial oligosaccharide that is transferred onto proteins, while complex glycans are the most processed and can take a variety of forms. This can include branching, extensions, and core fucosylation (1). However, the efficiency of glycan maturation at a given acceptor site on glycoproteins can often be incomplete, most notably (but not solely) because of steric or electrostatic factors that impact enzyme-substrate recognition (1, 105, 178). This often results in heterogeneous ensembles of glycan structures on glycoprotein acceptor (85), and even on individual glycosites on the same glycoprotein (179). This phenomenon is termed 'microheterogeneity' and is a hallmark of protein glycosylation that has been a focus of biochemical analysis for several decades (180, 181).

An important branching point in N-glycan processing is the addition of a branching β-2-linked GlcNAc to the α3 mannose of the $Man_5GlcNAc_2$ structure by the GT-A family

glycosyltransferase MGAT1 (9, 182). This step marks a class switch from high-mannose (e.g. $Man_5GlcNAc_2$) to hybrid (e.g. $GlcNAcMan_5GlcNAc_2$) glycans, as this branching GlcNAc can be further elaborated by other glycosyltransferases into a variety of structures. The activity of MGAT1 is also necessary for the subsequent trimming of the two terminal mannoses from the α6 mannose by the glycoside hydrolase MAN2A1 (85), which results in a $GlcNAcMan_3GlcNAc_2$ structure. This $GlcNAcMan_3GlcNAc_2$structure acts as the substrate for MGAT2 (179), which marks the point at which hybrid N-glycans transition to complex N-glycans, with two branching GlcNAc moieties that serve as a base for highly elaborated bi-, tri-, or tetra-antennary structures.

There is much interest in understanding the underlying criteria that define N-glycan microheterogeneity. N-glycans have been shown to be important modulators in antibody-receptor interactions, both with respect to glycosylation of the antibody (183–185) and their receptors (186, 187). In particular, the contribution of glycosylation to the properties of therapeutics is of particular interest in the development and manufacturing of biologics (175, 188) and biosimilars (189). Glycosylation of these therapeutics is known to impact their stability and pharmacokinetics (175). Additionally, N-glycans are a vital component of viral glycoprotein properties and are known to impact host immune surveillance (190, 191) and host receptor interactions involved in viral entry (148).

There have been several approaches to studying N-glycan microheterogeneity. Early studies demonstrated that N-glycan microheterogeneity is reproducible on a site-by-site basis (65), and that access by glycosyl hydrolases is predictive of N-glycan processing (192). NMR studies of N-glycan structures suggest that glycan interactions with the protein backbone can alter glycan conformations in ways that can impact N-glycan processing (98, 100, 193). There is evidence that changing nearby amino acids can alter N-glycan heterogeneity (178, 194). Some recent studies

have taken a systems approach, often by monitoring *in vivo* processing of N-glycans in cell culture systems (105, 178). The involvement of the peptide-glycan interactions in affecting glycan conformation, and thus potentially N-glycan processing, has also been supported by molecular dynamics simulations using yeast protein disulfide isomerase (PDI) as a model reporter glycoprotein (104, 195). Additionally, a meta-analysis of site-specific glycoproteomics papers found that solvent accessibility is related to the extent of branching and core fucosylation (196). A recent study by Mathew et al. studied early N-glycan processing steps through molecular dynamics simulations and *in vitro* processing of N-glycans on the yeast protein disulfide isomerase, a similar approach as this study (197). They found that the shape of the surrounding protein environment and subsequent glycan conformation can influence the rate at which individual sites are processed. They studied early mannose trimming and class switching from high-mannose to hybrid glycans using three mannosidases and the GlcNAc transferase MGAT1 on the processing of PDI. Our work here expands on this by using enzymes involved in later N-glycan processing using not only the model yeast protein PDI but also multiple N-linked glycoproteins of interest to human health.

In this study, we report extensive site-specific *in vitro* N-glycan processing data for five multiply N-linked glycosylated proteins, with 38 different sites of N-glycosylation in total. By enriching all sites of all glycoproteins with a common $Man_5GlcNAc_2$ substrate and then monitoring N-glycan processing through time-course reactions, we were able to identify key bottlenecks that prevent specific sites on glycoproteins from being converted from high-mannose to complex N-glycans. These bottlenecks appear to persist *in vivo* upon microheterogeneity analysis of each site of the reporter proteins when expressed in wild type cells. Additionally, we found that removing the tertiary structure of the protein abolished all site-specificity of N-glycan processing, highlighting the importance of protein tertiary structure in defining N-glycan microheterogeneity.

Results

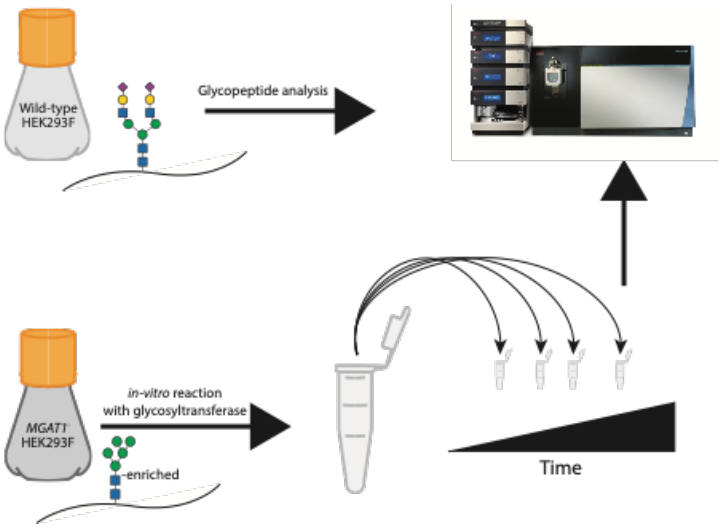
Expression of reporter proteins in WT-HEK293F cells

In order to probe individual steps of N-glycan processing, we first established a set of reporter proteins to be used as case studies (**Table** 2.1, **Fig.** S1). These proteins were selected based on their various applications in biology, virology, and use as therapeutics as well as their diversity in displayed glycans and the availability of quality crystal structures. CD16a (Fc $_{\gamma}$ receptor IIIa) is an IgG receptor that is known to have differential affinities to antibodies depending on its glycan presentation which impacts downstream signaling (187, 194, 198). Protein disulfide isomerase (PDI) is a resident ER glycoprotein that has been used as a model protein for studying N-glycan processing due to its ease of expression, analysis, and well-defined site-specific glycan heterogeneity (104, 105, 178, 195, 197). Etanercept is a bioengineered therapeutic fusion protein of a TNFα receptor and an IgG1 Fc domain commonly used to help treat auto-immune disorders (199). Erythropoietin is a therapeutic glycoprotein that stimulates red-blood cell growth, and its glycosylation is known to impact its pharmacokinetics (200–202). SARS-CoV-2 spike glycoprotein is a highly glycosylated trimer that is responsible for the viral entry of the associated coronavirus SARS-CoV-2 via binding to the human receptor ACE2 (148, 203).

Table 2.1: Glycoprotein reporters

Reporter	Uniprot ID	Amino acids	Glycan sites	Glycan structures	PDB
Etanercept TNFR-IgG Fc fusion	P20333 + P01857	1-235 + 236-467	3	Varied per site	3ALQ 3AVE
Erythropoietin	P01588	28-193	3	Tri- and tetra-antennary	1EER
SARS-CoV2 Spike	P0DTC2	1-1208	22	Varied high man to complex	6VSB
CD16a Fc ɣ receptor IIIa	P08637	19-192	5	Mostly complex bi-, tri-, and tetra-antennary	5BW7
Pdi1p (yeast)	P17967	29-522	5	Complex, varied	2B5E

The reporter proteins were first transiently expressed in high yields in wild type HEK293F cells then harvested from supernatant and purified with Ni-NTA chromatography (**Figs. S1b-f)**. Glycopeptide analysis using LC-MS/MS was performed on these purified proteins in order to determine their glycan occupancy and diversity when expressed in a "wild type" background (**Fig.** 2.1). With liquid chromatography and tandem mass spectrometry techniques, we were able to obtain a detailed characterization of the N-glycan profile at each site on the reporter proteins, some of which contained dozens of different glycan moieties with a variety of terminal structures including sialylation, as well as core fucosylation (**Fig.** 2.2a-c).



All classes of glycans were observed at most sites (**Fig.** 2.3, S5-S42), with SARS-CoV-2 spike glycoprotein pictured separately due to its large number of sites (**Fig.** S2a). Of particular interest are the sites on reporter glycoproteins that greatly differ from other sites on the same protein: sequons 2 and 4 on CD16a (**Fig.** 2.3a) and sequon 4 on PDI (**Fig.** 2.3b) are predominantly less-processed high-mannose and hybrid structures, while the other sites of N-glycosylation on the same proteins are mostly highly processed complex structures. This is in contrast to etanercept (**Fig.** 2.3c) and erythropoietin (**Fig.** 2.3d), which have more homogenous N-glycan presentations. The

Figure 2.1: Graphical representation of approach. Reporter proteins were expressed in HEK293F WT and MGAT1- cells, analyzed via LC-MS/MS, and then processed by purified glycosyltransferases and hydrolases in vitro.

SARS-CoV-2 spike glycoprotein had a diversity of N-glycan presentations on its 22 sites, with most sites enriched with complex N-glycans and certain sites mostly presenting high-mannose N-glycans (**Fig.** S2). Interestingly, we noted that Man5GlcNAc2 was always the most abundant high mannose structure on 36 of our 38 sites with two exceptions being N0234 and N0717 of SARS-CoV-2 that both contain less than 15% complex structures (**Figs**. S2, S27, S35).

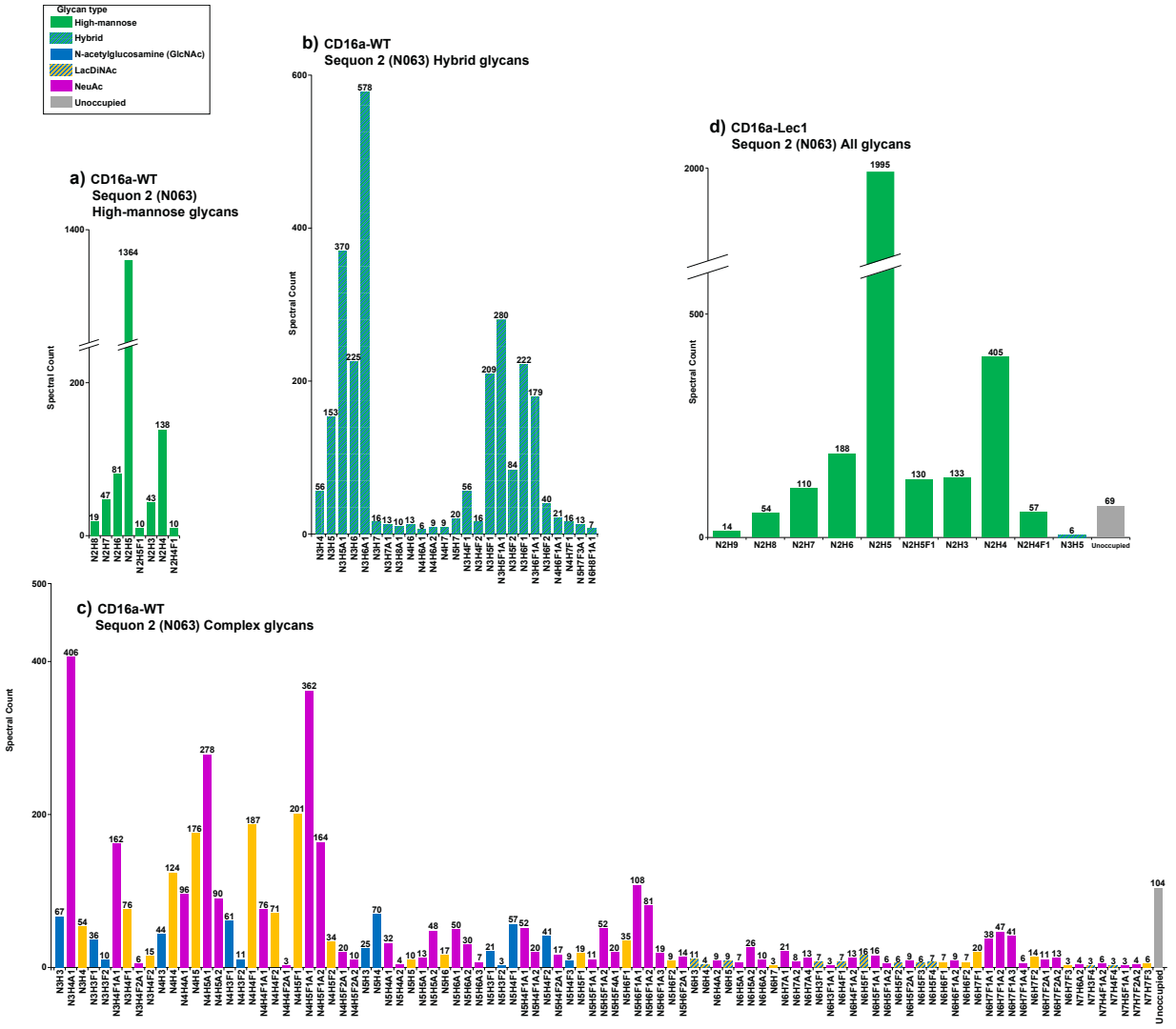


Figure 2.2: Microheterogeneity at a single site on CD16a (Site 2, N063). Glycopeptide analysis with individual glycan types quantified via spectral count when expressed in a wild-type HEK293F cells or Lec1-HEK293F (MGAT1-) cells, divided by glycan class. Column color denotes terminal features associated with composition *a*, WT-expressed CD16a high-mannose N-glycans. *b*, WT-expressed CD16a hybrid N-glycans. *c*, WT-expressed CD16a complex N-glycans. d, Lec1-expressed CD16a N-glycans (all). Colored bars denote glycan terminal features.

Expression of reporter proteins in Lec1-HEK293F cells

Next, these same proteins were transiently overexpressed in HEK293S GnTI- (MGAT1 null) cells. The activity of MGAT1 is necessary for the formation of both hybrid and complex N-glycans, as the addition of GlcNAc to the non-reducing end α3-linked mannose is needed for further elaboration and capping by downstream enzymes. Knockout of MGAT1 substantially reduces the diversity of glycans at all sites of N-glycosylation and causes a significant enrichment of $Man_5GlcNAc_2$ structures N-glycans on expressed glycoproteins, as shown on sequon 2 of CD16a (**Fig.** 2.2d) as well as the other reporter sites (**Figs.** S5-S42). This is useful because it allows for the *in vitro* processing of all N-glycans on a glycoprotein to begin from a common substrate. This enrichment was successful for most sites on all reporter proteins (**Figs.** 2.3, S2b). However, there is an exception at Sequon 2 of etanercept, which contained a significant population of apparent hybrid N-glycans by an unknown processing event that we are currently exploring. These appear to be true hybrid structures with an attached GlcNAc on the α-3-linked mannose based on MS2 fragmentation data (**Fig.** S16b).

Conversion of high-mannose glycans to hybrid glycans

In order to probe the effects of tertiary structure on N-glycan processing, we first monitored the conversion of $Man_5GlcNAc_2$ N-glycans to $GlcNAcMan_5GlcNAc_2$ N-glycans via the addition of GlcNAc by the glycosyltransferase MGAT1 on intact reporter proteins expressed and purified from MGAT1 deficient cells. We did this through a series of time-course reactions using purified protein and MGAT1 in the presence of the nucleotide sugar donor UDP-GlcNAc followed by analysis and quantitation via LC-MS/MS (**Fig.** 2.4, example TIC can be found in **Fig.** S43). The ratio of enzymes to molarity of reporter N-glycan sites was kept constant so that we could compare

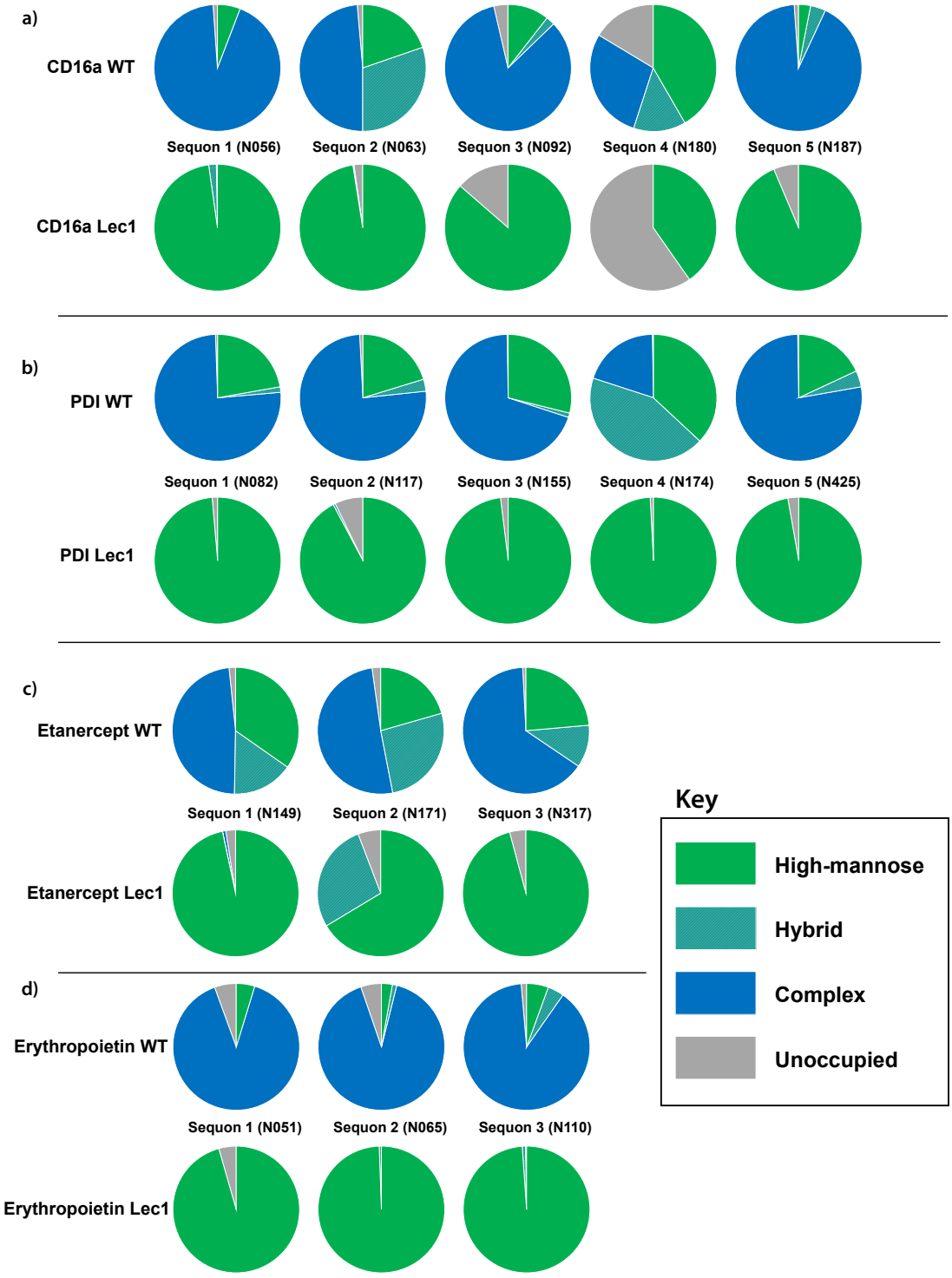


Figure 2.3: Site occupancy of reporter proteins expressed in WT-HEK293F and Lec1-HEK293F cells. Relative proportion of glycan classes at each site on reporter proteins when expressed in a wild-type or Lec1 (MGAT1-) background. *a*, CD16a. *b*, PDI. *c*, Etanercept. *d*, Erythropoietin. Relative populations were ascertained with glycopeptide analysis and quantified with spectral counts.

the inter-protein as well as intra-protein rates of N-glycan processing. We observed site-specific rates of GlcNAc addition across the range of our 38 sites of N-glycosylation. Generally, these rates corresponded well to the distributions of N-glycans that were found on the respective sites when the reporters were expressed in wild type HEK293F cells (**Figs.** 2.3, S5-42). For example, sequons 2 and 4 on CD16a are both modified relatively slowly by MGAT1 (**Fig.** 2.4a) and are also enriched with high-mannose structures when expressed in wild type HEK293 cells (**Fig.** 2.3a). This also holds for sequon 4 of PDI (**Fig.** 2.3b), which is the only site on PDI that is modified slowly by MGAT1 (**Fig.** 2.4b). This pattern is also demonstrated on SARS-CoV-2 spike glycoprotein: we observed a broad distribution of MGAT1 activity rates across its 22 sites of N-glycosylation (**Figs.** 2.4e, S3), and the fastest and slowest sites are enriched in complex and high-mannose N-glycans, respectively (**Fig.** 2.4f). This is best represented by sites N0234 and N0717 on Spike glycoprotein, which had slow transfer rates and primarily were occupied by high-mannose N-glycans when expressed in a wild type background (**Figs.** 2.4f, S2).

In contrast to the demonstration of site-specificity for PDI and CD16a, all sites on etanercept have relatively high levels of high-mannose glycans (**Fig.** 2.3c), and all are processed slowly by MGAT1 (**Fig.** 2.4c). Additionally, erythropoietin's three sites of N-glycosylation are all processed efficiently by MGAT1 (**Fig.** 2.4d), and when expressed in wild type HEK293 cells mostly produce complex N-glycans (**Fig.** 2.3d).

Conversion of hybrid glycans to complex glycans

The conversion of hybrid N-glycans such as $GlcNAcMan_5GlcNAc_2$ to complex N-glycans requires the activity of two enzymes: the glycoside hydrolase MAN2A1 and the GlcNAc transferase MGAT2. In order to probe the site-specific rates of MAN2A1 activity, we first reacted

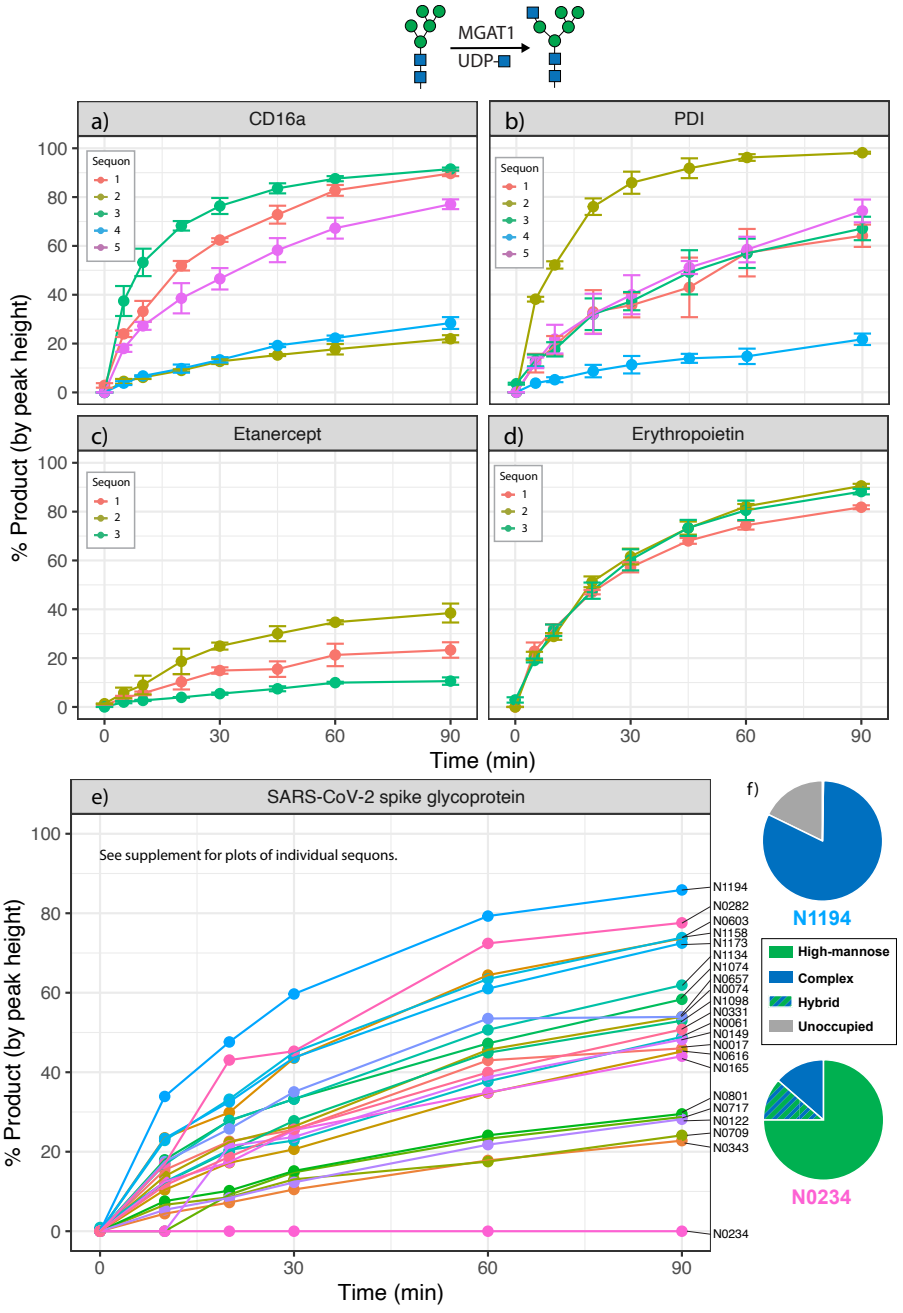


Figure 2.4: Site-specific monitoring of MGAT1 activity. Time-course reaction of GlcNAc addition to reporter proteins with recombinant MGAT1. *a*, CD16a. *b*, Protein disulfide isomerase. *c*, Etanercept. *d*, Erythropoietin. *e*, SARS-CoV-2 spike glycoprotein. *f*, wild-type glycopeptide profiles for the sites on SARS-CoV-2 spike glycoprotein with the fastest (N1194) and slowest (N0234) rates of MGAT1 activity. Error bars and legend omitted for SARS-CoV-2 spike glycoprotein due to large number of sites. Reaction progress calculated as proportion of the sum of monoisotopic peak heights of product ($Man_5GlcNAc_3$) vs. the sum of product and reactant ($Man_5GlcNAc_2$) peak heights. Experiments performed in triplicate, error bars represent standard deviation.

the reporters expressed in the MGAT1 null cell line with an excess of MGAT1 to enrich $GlcNAcMan_5GlcNAc_2$ structures. After ensuring >80% conversion to product at each site, we then examined conversion of the $GlcNAcMan_5GlcNAc_2$ product to $GlcNAcMan_3GlcNAc_2$ following digestion with MAN2A1 (**Fig.** 2.5). Similar patterns of site-specific rates were seen as with MGAT1, with sequons 2 and 4 of CD16a (**Fig.** 2.5a) and sequon 4 of PDI (**Fig.** 2.5b) all having much lower levels of activity compared to the other sites on the same protein. Again, all sites on etanercept (**Fig.** 2.5c) were processed much more slowly than those on erythropoietin (**Fig.** 2.5d). Notably, no cleavage products were observed at sequon 5 (N0149) on SARS-CoV-2 spike glycoprotein (**Fig.** S3). Inspection of the $GlcNAcMan_4GlcNAc_2$ intermediate in MAN2A1 processing revealed that at this site, only one mannose was able to be removed (**Fig.** S4). Despite this *in vitro* observation, when expressed in a wild type background this site produces an abundance of complex-type N-glycans (**Figs.** S2, S25). This is in contrast to sites N0234 and N0717 on the spike glycoprotein, which exhibit both slow transfer rates and an enrichment of high-mannose N-glycans (**Figs.** 2.5f, S2-S3), but still form the $GlcNAcMan_3GlcNAc_2$ product.

The next step in the formation of complex glycans is the addition of a β-2-linked GlcNAc to the α6-mannose of the $GlcNAcMan_3GlcNAc_2$ moiety. We reacted the reporter proteins expressed in the MGAT1 null cells with an excess of MGAT1, MAN2A1, and the UDP-GlcNAc donor in order to enrich the $GlcNAcMan_3GlcNAc_2$ substrate, then performed another set of time-course reactions with MGAT2. Sites which could not be efficiently converted to $GlcNAcMan_3GlcNAc_2$ structures by MGAT1 and MAN2A1 treatment (e.g. sequon 4 on PDI (**Fig.** 2.5b)) were excluded from further analyses. Similar patterns of modification were observed in the MGAT2 reactions as were seen in the MGAT1 and MAN2A1 experiments, with lower levels of activity observed at sequons 2 and 4 on CD16a compared to other sites on the same protein (**Fig.**

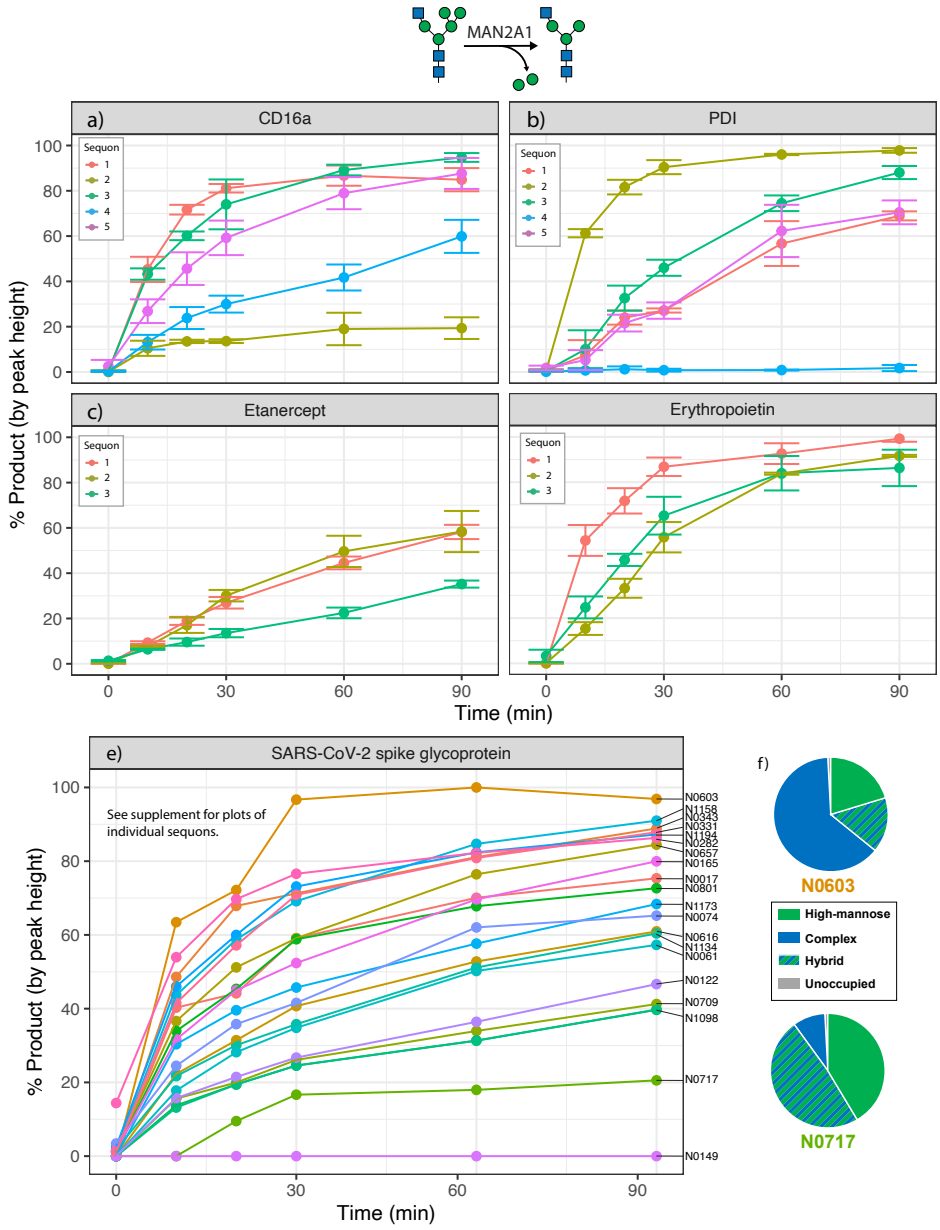


Figure 2.5: Site-specific monitoring of MAN2A1 activity. Time-course reaction of GlcNAc addition to reporter proteins with recombinant MGAT1. *a*, CD16a. *b*, Protein disulfide isomerase. *c*, Etanercept. *d*, Erythropoietin. *e*, SARS-CoV-2 spike glycoprotein. *f*, wild-type glycopeptide profiles for the sites on SARS-CoV-2 spike glycoprotein with the fastest (N0603) and slowest (N0747) rates of MAN2A1 activity for a site that was able to form product. Legend omitted for SARS-CoV-2 spike glycoprotein due to large number of sites. Reaction progress calculated as proportion of the sum of monoisotopic peak heights of product ($Man_3GlcNAc_3$) vs. the sum of product and reactant ($Man_5GlcNAc_3$) peak heights. Experiments performed in triplicate, error bars represent standard deviation.

2.6a). Additionally, all sites on etanercept (**Fig.** 2.6c) were processed more slowly than those on erythropoietin (**Fig.** 2.6d). A broad range of processing rates was observed on SARS-CoV-2 spike glycoprotein (**Fig.** 2.6e). Sites that were modified fastest in our *in vitro* modification studies were also enriched in complex N-glycans when expressed in wild type HEK293 cells and the sites that were the slowest for *in vitro* modification corresponded to sites that were relatively enriched with high-mannose N-glycans when generated in wild type cells (**Figs.** 2.6f, S3).

Core fucosylation of N-glycans by FUT8

Following the above experiments, we wanted to see if similar patterns of site-specific N-glycan processing rates would apply to core fucosylation. Core fucosylation is the attachment of an α1,6-linked fucose to the GlcNAc that is directly attached to the asparagine at the core of N-linked glycans, a reaction which is catalyzed by the fucosyltransferase FUT8. This reaction is generally specific to complex N-glycans (12), and thus we generated $GlcNAc_2Man_3GlcNAc_2$ glycans on our collection of reporter proteins by reacting with an excess of MGAT1, MAN2A1, and MGAT2 in the presence of the UDP-GlcNAc sugar donor. We then examined the rates of modification of the respective glycans with FUT8 (**Fig.** 2.7). Interestingly, at many sites we found substantial core fucosylation prior to *in vitro* processing despite the reporter proteins being expressed in an MGAT1-null cell line and thus lacking complex N-glycans (**Figs.** 2.7a-b, d-e, S5-42). Otherwise, we observed a diversity of fucosylation rates among the different sites. Sequon 2 on CD16a (**Fig.** 2.7a) and sequon 1 and 5 on PDI were markedly slow (**Fig.** 2.7b), as well as sequon 3 on etanercept (**Fig.** 2.7c). All sites on erythropoietin were fucosylated rapidly, which possibly reflects their high initial levels of fucosylation even before the FUT8 reaction (**Fig.** 2.7d). Most sites on SARS-CoV-2 spike glycoprotein were efficiently fucosylated (**Fig.** 2.7e), with a few

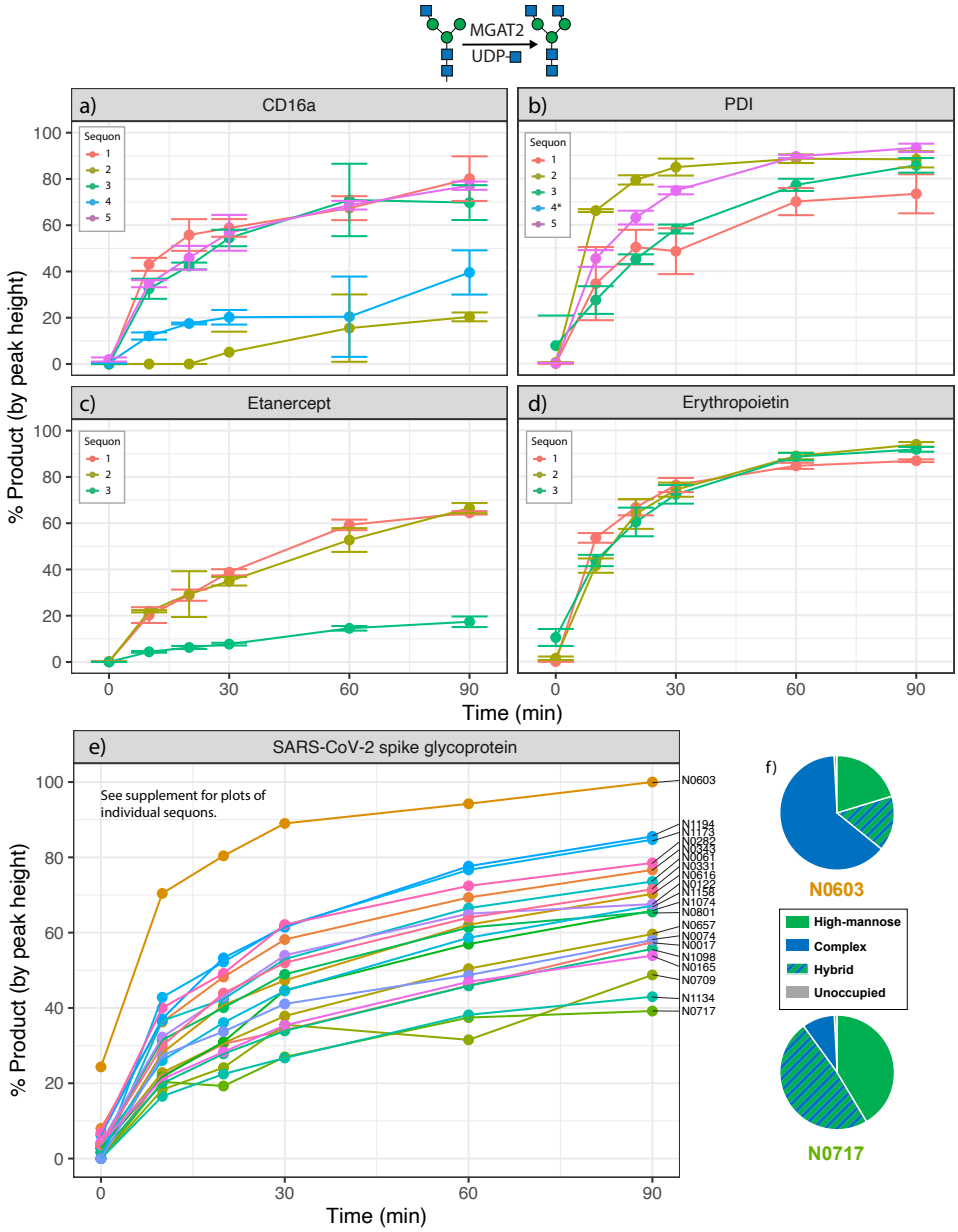


Figure 2.6: Site-specific monitoring of MGAT2 activity. Time-course reaction of GlcNAc addition to reporter proteins with recombinant MGAT2. *a*, CD16a. *b*, Protein disulfide isomerase. *c*, Etanercept. *d*, Erythropoietin. *e*, SARS-CoV-2 spike glycoprotein. f, wild-type glycopeptide profiles for the sites on SARS-CoV-2 spike glycoprotein with the fastest (N0603) and slowest (N0717) rates of MGAT2 activity. Error bars and legend omitted for SARS-CoV-2 spike glycoprotein due to large number of sites. Asterisks on site legend indicate that not enough substrate was generated from previous N-glycan processing steps to monitor reaction progress. Reaction progress calculated as proportion of the sum of monoisotopic peak heights of product ($Man_3GlcNAc_4$) vs. the sum of product and reactant ($Man_3GlcNAc_3$) peak heights. Experiments performed in triplicate, error bars represent standard deviation.

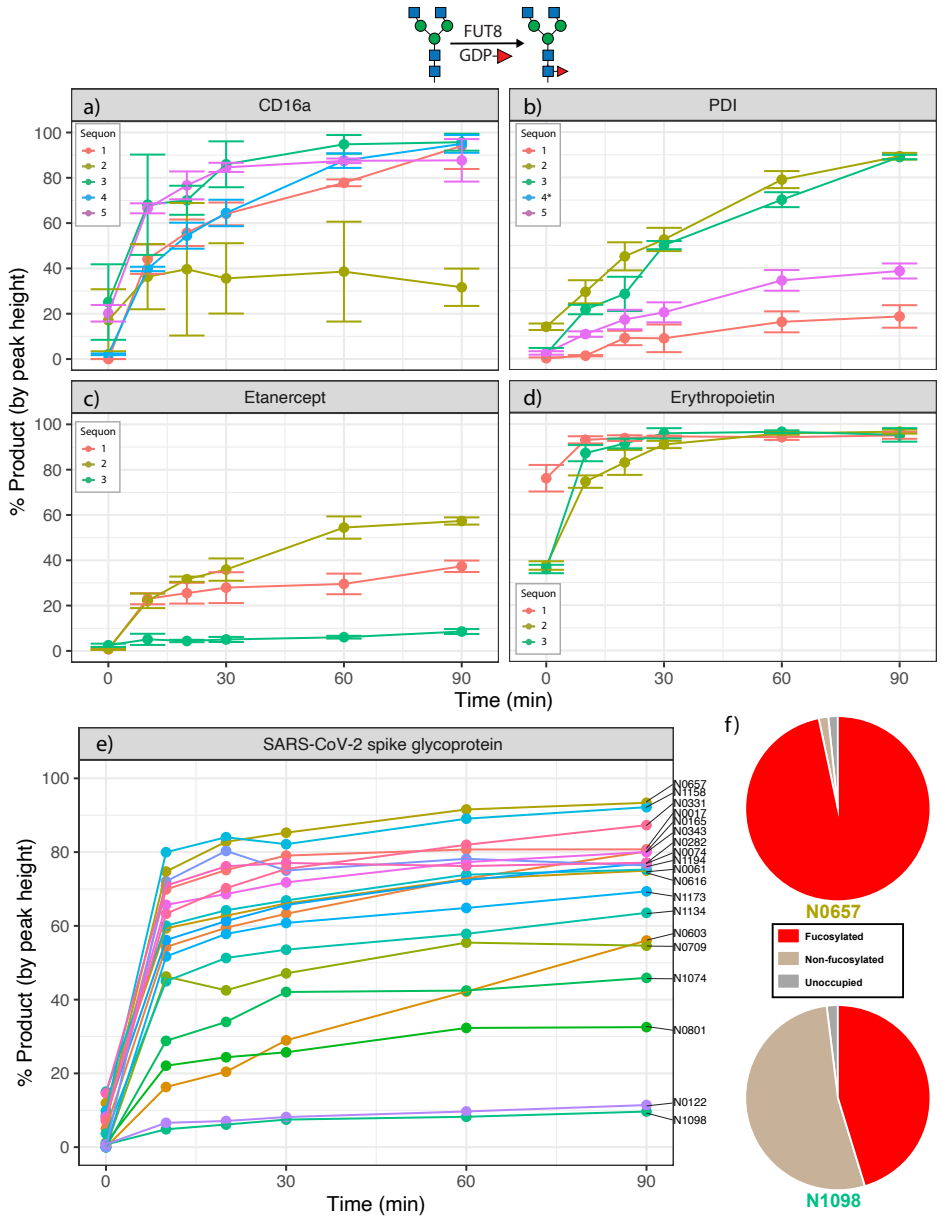


Figure 2.7: Site-specific monitoring of FUT8 activity. Time-course reaction of Fucose addition to reporter proteins with recombinant FUT8. *a*, CD16a. *b*, Protein disulfide isomerase. *c*, Etanercept. *d*, Erythropoietin. *e*, SARS-CoV-2 spike glycoprotein. *f*, wild-type glycopeptide profiles for the sites on SARS-CoV-2 spike glycoprotein with the fastest (N0657) and slowest (N1098) rates of FUT8 activity. Error bars and legend omitted for SARS-CoV-2 spike glycoprotein due to large number of sites. Asterisks on site legend indicate that not enough substrate was generated from previous N-glycan processing steps to monitor reaction progress. Reaction progress calculated as proportion of the sum of monoisotopic peak heights of product ($Man_3GlcNAc_4$) vs. the sum of product and reactant ($Man_3GlcNAc_4Fuc_1$) peak heights. Experiments performed in triplicate, error bars represent standard deviation.

exceptions (N0122, N0801, N1074, N1098) (**Fig.** S3). Glycopeptide analysis data reflects that sites modified more rapidly by FUT8 *in vitro* had higher levels of fucosylation *in vivo* (**Fig.** 2.7f).

Impact of tertiary structure on site-specificity

In order to determine whether these site-specific differences in glycosyltransferase rates were due to tertiary structure, we repeated the transfer of GlcNAc onto PDI-$Man_5GlcNAc_2$ glycans using MGAT1, but first digested the protein with trypsin to cleave the fully-folded protein substrate into glycopeptides. Without the reporter protein tertiary structure, all site-specificity of GlcNAc transfer rate was lost and the overall rate of transfer was reduced (**Fig.** 2.8a compared to **Fig.** 2.4B). Since FUT8 activity requires access to the core GlcNAc linked to the Asn residue of the peptide backbone, we were curious to see if the glycoprotein being cleaved to glycopeptides would eliminate the site-specificity observed on intact protein. Similar to MGAT1, all site-specific modification by FUT8 was lost following cleavage of CD16a to glycopeptides, with all sites (including the slow site 2 (Fig. 2.2a)) exhibiting similar rates of modification (**Fig.** 2.8b). CD16a was used instead of PDI for the FUT8 experiment to more fully illustrate that local tertiary structure was important for

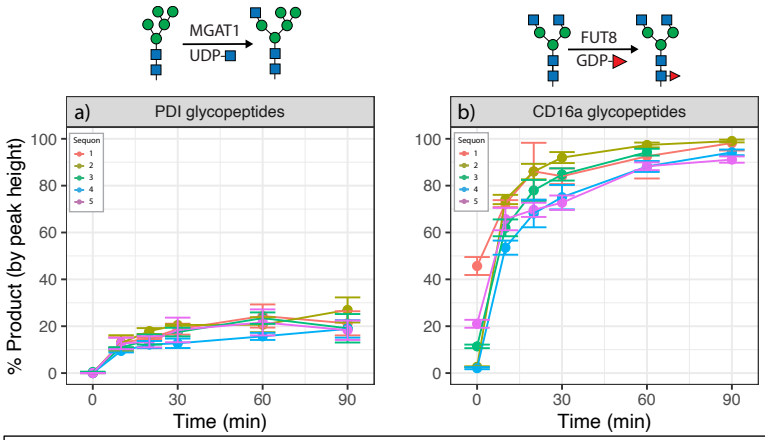


Figure 2.8: Tertiary structure is responsible for site-specificity of N-glycan processing rates. Time-course reaction of N-glycan processing on proteins digested into glycopeptides with proteases. *a*, Reaction of MGAT1 with PDI glycopeptides. *b*, Reaction of FUT8 with CD16a glycopeptides. Reaction progress calculated as proportion of the sum of monoisotopic peak heights of product vs. the sum of product and reactant peak heights. Experiments performed in triplicate, error bars represent standard deviation.

more than just one of our reporter proteins and because site 4 on PDI cannot adequately form complex N-glycans due to low MAN2A1 activity (**Fig.** 2.5b).

Discussion

While N-glycans are a crucial component in the production of membrane bound and secreted glycoproteins, the determinants that define the diversity of N-glycan structures at any given site are not well understood. Factors that have been suggested to influence N-glycan microheterogeneity include the expression of glycosyltransferases and glycoside hydrolases (204, 205), the availability of nucleotide sugar donors (206), secretory pathway trafficking (104, 105, 205), and the accessibility of the acceptor site (102, 178, 197). The impact of enzyme availability in cells has mostly been probed through genetic engineering approaches (204). However, availability of enzymes and sugar nucleotides cannot sufficiently explain site specific differences on the same polypeptide nor can protein trafficking in the secretory pathway. Thus, we hypothesize that it is the impact of acceptor site accessibility that allows for site specific differences on the same protein. This could occur by multiple mechanisms including the substrate glycan interacting with the substrate protein backbone at a specific site and thus hampering engagement of the glycoenzyme with the substrate glycan. An example of this is site 4 on PDI that was elegantly demonstrated by Aebi and colleagues (178). Additionally, the glycosylation on the Fc fragment of etanercept (site 3, N317) differs from what is normally seen on IgG as there is an abundance of high-mannose glycans accompanying the expected biantennary-complex structures (207), which may indicate altered accessibility to the N-glycan at that site in this non-naturally occurring fusion protein. Another possibility is local secondary and tertiary structure of the substrate protein at individual sites of modification and the glycoenzyme active site resulting

in steric or electrostatic clashes that prohibit optimal binding for catalysis. Interestingly, our analysis revealed that proteolytically digesting substrate proteins before transfer reactions abolished site specific rate differences (**Fig.** 2.8), and in the case of MGAT1 reduced the overall rate of reaction. This strongly support higher-order structure of the substrate playing an important role in transfer rates and in agreement with proposed mechanisms for microheterogeneity.

Molecular dynamics studies and Markov state modeling by Mathew et al. demonstrated that the relative amount of time a glycan spends extended away from the protein and exposed to solvent correlates with site-specificity of glycan-processing rates on the yeast model protein disulfide isomerase (197). Additionally, they monitored *in vitro* N-glycan processing rates with ER mannosidases as well as MGAT1 and MAN2A1, and the results with overlapping enzymes in our present study agree well. Site 4 of PDI was identified as a "slow" site (**Figs.** 2.4b, 2.5b), and their Michaelis-Menten analysis of PDI processing kinetics are also consistent with our observations (**Figs.** 2.4b, 2.5b). Additionally, their studies on the kinetics of earlier glycan-processing steps involving ER mannosidase I and Golgi mannosidase IB show that this site-specificity is conserved in earlier steps of N-glycan processing. However, their approach to reduce tertiary structure through reduction and alkylation prior to *in vitro* modification led to results contrasting with our glycopeptide experiments (**Fig.** 2.8**)**, and they observed differences in modification rate at different sites. This may be due to some secondary structures of the protein not being completely disrupted without the use of protease digestion to cleave the model protein utilized, or perhaps the denatured protein can still influence site kinetics.

Generally, these results indicate that the tertiary structure specific to an acceptor site can be an important factor in defining the types of N-glycans seen at a given site. In particular, the efficiency (or lack thereof) of MGAT1 and MAN2A1 appears to be highly predictive of high-

mannose-type glycans at a sequon. *In vivo* it is likely that MGAT1 is rate-limiting as the most common high-mannose structure at most sites when expressed in a wild type background is the MGAT1 substrate $Man_5GlcNAc_2$ (**Figs.** S5-42), and its activity is required for downstream processing by enzymes like MAN2A1. The rate limiting role of MAN2A1 was also observed in our *in vitro* studies, particularly at sites that were also poorly modified with MGAT1. This may be partly due to lower activity of the recombinant enzyme employed in our *in vitro* studies; four times as much MAN2A1 had to be used in assays compared to the glycosyltransferases (1:250 enzyme:substrate molar ratio for MAN2A1 vs. 1:1,000 for glycosyltransferases). MAN2A1 processing has previously been identified as a potential bottleneck in N-glycan processing (197). Potential steric barriers to MAN2A1 action are suggested by the structure of MAN2A1:substrate complex that demonstrates a significant portion of the total N-linked glycan must fit into the active site of the enzyme for efficient binding and catalysis (208). If a site is not processed totally by MAN2A1, it may form hybrid structures, but cannot form complex structures due to the necessity of mannose trimming on the α6 branch of the tri-mannosyl N-glycan core. By contrast, the active site structure of MGAT1 involved in acceptor recognition has not yet been determined, but likely also presents significant steric barriers for access to some poorly modified sites.

While these studies provide a sound starting point for determining what structural features may be important in determining N-glycan destiny, much work remains. We purposely chose reporter proteins and processing enzymes with experimentally determined structures (**Table** 2.1), (203, 209–212). We are currently utilizing MD simulations of glycosylated reporter proteins and site-specific docking of specific glycan modified reporters with glycoenzymes to determine site-specific glycans interacting with the reporter protein as well as clashes between the reporter sites and the glycoenzymes. This will guide future work involving mutagenesis studies to influence the

rate at which glycosyltransferases and glycosyl hydrolases are able to modify acceptor sites. There is evidence that this approach can indeed alter the distribution of N-glycans at a specific site, as evidenced through modification of a tyrosine residue proximal to sequon 4 on protein disulfide isomerase (178). Taking a systematic approach that involves site-specific rate monitoring coupled with modeling and mutagenesis should result in common rules that not only will allow prediction of microheterogeneity but will allow us to tune it.

Experimental procedures

Expression and purification of glycoprotein reporters and glycosylation enzymes for in vitro modification.

Expression constructs encoding the reporter proteins were generated with either NH_2-terminal fusion tags (CD16a (low affinity immunoglobulin gamma Fc region receptor III-A, FCGR3A), UniProt P08637, residues 19-193; Erythropoietin (EPO), UniProt P01588, residues 28-193; Etanercept (TNF receptor-IgG1 fusion), GenBank AKX26891, residues 1-467) or C-terminal fusion tags (yeast PDI1 (protein disulfide-isomerase), UniProt P17967, residues 1-494; SARS-CoV-2 Spike glycoprotein, UniProt P0DTC2, residues 1-1208). The constructs employing N-terminal fusion sequences employed the pGEn2 expression vector while the PDI1 construct was generated in the PGEc2 vector as previously described (213).For the pGEn2 constructs, the fusion protein coding region was comprised of a 25-amino acid signal sequence, an His_8 tag, AviTag, the "superfolder" GFP coding region, the 7-amino acid recognition sequence of the tobacco etch virus (TEV) protease followed by the catalytic domain region for reporter proteins (213). Constructs encoding MGAT1, MAN2A1, MGAT2 and FUT8 employed the pGEn2 vector and were expressed and purified as previously described (213). For the PDI1 construct the pGEc2 vector

was employed and encoded the segment of *Saccromyces cerevisiae* PDI1 indicated followed by an SGSG tetrapeptide, the 7 amino acid TEV recognition sequence, the "superfolder" GFP coding region, and an His_8 tag (213). For SARS-CoV-2 Spike, the construct contained an additional COOH-terminal trimerization sequence and His6 tag as previously described (214). The recombinant reporter proteins were expressed as a soluble secreted proteins by transient transfection of suspension culture HEK293F cells (FreeStyle™ 293-F cells, Thermo Fisher Scientific, Waltham MA) for wild type glycosylated structures and in HEK293S (GnTI-) cells (ATCC) to generate $Man_5GlcNAc_2$-Asn glycan structures (213, 215). Cultures were maintained at 0.5–3.0x10^6 cells/ml in a humidified CO_2 platform shaker incubator at 37°C with 50% humidity. Transient transfection was performed using expression medium comprised of a 9:1 ratio of Freestyle™293 expression medium (Thermo Fisher Scientific, Waltham MA) and EX-Cell expression medium including Glutmax (Sigma-Aldrich). Transfection was initiated by the addition of plasmid DNA and polyethyleneimine as transfection reagent (linear 25-kDa polyethyleneimine, Polysciences, Inc., Warrington, PA). Twenty-four hours post-transfection the cell cultures were diluted with an equal volume of fresh media supplemented with valproic acid (2.2 mM final concentration) and protein production was continued for an additional 5 days at 37°C[3]. The cell cultures were harvested, clarified by sequential centrifugation at 1200 rpm for 10 minutes and 3500 rpm for 15 minutes at 4°C, and passed through a 0.8 μM filter (Millipore, Billerica, MA). The protein preparation was adjusted to contain 20 mM HEPES, 20 mM imidazole, 300 mM NaCl, pH 7.5, and subjected to Ni-NTA Superflow (Qiagen, Valencia, CA) chromatography using a column preequilibrated with 20 mM HEPES, 300 mM NaCl, 20 mM imidazole, pH 7.5 (Buffer I). Following loading of the sample the column was washed with 3 column volumes of Buffer I followed by 3 column volumes of Buffer I containing 50 mM imidazole, and eluted with Buffer I

containing 300 mM imidazole at pH 7.0. The protein was concentrated to approximately 3 mg/ml using an ultrafiltration pressure cell (Millipore, Billerica, MA) with a 10-kDa molecular mass cutoff membrane and buffer exchanged with 20 mM HEPES, 100 mM NaCl, pH 7.0, 0.05% Sodium azide and 10% glycerol.

In vitro N-Glycan processing

For the time-course reactions, purified reporter proteins generated in *HEK293S (GnTI-)* cells were used. Reactions were performed at 37°C in 1.5-mL Eppendorf tubes in a reaction volume of 150 μL, with 20 mM HEPES (VWR) pH 7.5 and 300 mM NaCl (Fisher). For glycosyltransferases, the corresponding nucleotide sugar, UDP-GlcNAc (Sigma) for MGAT1 and MGAT2 and GDP-Fucose (CarboSynth) for FUT8, was kept in excess at 1 mM. MGAT1 and MGAT2 reactions were supplemented with 1 mM $MnCl_2$ (Sigma). The concentration of total N-glycans for each reaction was kept at 5 μM: for example, for a reporter protein with 5 sites of N-glycosylation, the concentration of the protein would be 1.25 μM. For MGAT1, MGAT2, and FUT8 reactions, a 1:1,000 enzyme-to-glycan ratio was used with the concentration of respective enzyme at 5 nM; for MAN2A1 reactions, a 1:250 enzyme-to-glycan ratio was used with the concentration of MAN2A1 at 20nM. Prior to adding enzymes, time-course reaction vessels were equilibrated at 37°C for 15 minutes. At each time point, 20 μL of samples were taken and reactions were deactivated by heating at 95°C for 5 minutes. The samples were then digested by proteases and processed for LC-MS/MS analysis. In order to prepare substrate reporter proteins for N-glycan processing steps downstream of MGAT1 (e.g. MAN2A1, MGAT2, FUT8), reporters were reacted with the appropriate combination of enzymes and sugar nucleotides at a 1:100 enzyme:glycan ratio

for 3 hours. Reporters were confirmed to have been >80% converted to desired product by LC-MS/MS, detailed below.

Enzymatic digestion of PDI1, etanercept, EPO, CD16a, and SARS-CoV-2 spike from wild type and HEK293S (GnTI-) cells

All proteins were reduced by incubating with 10 mM of dithiothreitol (Sigma) at 56 °C and alkylated by 27.5 mM of iodoacetamide (Sigma) at room temperature in dark. For the intact glycopeptide analysis, aliquots of PDI1 proteins were digested respectively using trypsin (Promega), a combination of trypsin and Glu-C (Promega), or a combination of trypsin and AspN (Promega); aliquots of etanercept proteins were digested respectively using trypsin (Promega), or AspN (Promega); aliquots of EPO proteins were digested respectively using a combination of trypsin and Glu-C (Promega), or Glu-C (Promega); aliquots of CD16a proteins were digested respectively using chymotrypsin (Athens Research and Technology), AspN (Promega), or a combination of chymotrypsin (Athens Research and Technology) and Glu-C (Promega); aliquots of S proteins were digested respectively using alpha lytic protease (New England BioLabs), chymotrypsin (Athens Research and Technology), a combination of trypsin and Glu-C (Promega), or a combination of Glu-C and AspN (Promega). For the analysis of deglycosylated glycopeptides, aliquots of PDI1 proteins were digested respectively using trypsin (Promega), or a combination of trypsin and Glu-C (Promega); aliquots of etanercept proteins were digested respectively using trypsin (Promega), or AspN (Promega); aliquots of EPO proteins were digested respectively using a combination of trypsin and Glu-C (Promega), or Glu-C (Promega); aliquots of CD16a proteins were digested respectively using chymotrypsin (Athens Research and Technology), or AspN (Promega); aliquots of S proteins were digested respectively using chymotrypsin (Athens Research

and Technology), a combination of trypsin and Glu-C (Promega), or AspN (Promega). Following digestion, the proteins were deglycosylated by Endo-H (Promega) followed by PNGaseF (Promega) treatment in the presence of ^{18}O water (Cambridge Isotope Laboratories).

LC-MS/MS analysis of glycopeptides of PDI1, etanercept, EPO, CD16a, and SARS-CoV-2 spike from wild type and HEK293S (GnTI-) cells

The resulting peptides from respective enzymatic digestion of each protein were separated on an Acclaim PepMap RSLC C18 column (75 μm x 15 cm) and eluted into the nano-electrospray ion source of an Orbitrap Fusion™ Lumos™ Tribrid™ or an Orbitrap Eclipse™ Tribrid™ mass spectrometer (Thermo Fisher Scientific) at a flow rate of 200 nL/min. The elution gradient for PDI1, etanercept, EPO, and CD16a proteins consists of 1-40% acetonitrile in 0.1% formic acid over 220 minutes followed by 10 minutes of 80% acetonitrile in 0.1% formic acid. The elution gradient for S protein consists of 1-40% acetonitrile in 0.1% formic acid over 370 minutes followed by 10 minutes of 80% acetonitrile in 0.1% formic acid. The spray voltage was set to 2.2 kV and the temperature of the heated capillary was set to 275 °C. For the intact glycopeptide analysis, full MS scans were acquired from m/z 200 to 2000 at 60k resolution, and MS/MS scans following higher-energy collisional dissociation (HCD) with stepped collision energy (15%, 25%, 35%) were collected in the orbitrap at 15k resolution. For the deglycosylated glycopeptide analysis, full MS scans were acquired from m/z 200 to 2000 at 60k resolution, and MS/MS scans following collision-induced dissociation (CID) at 38% collision energy were collected in the ion trap.

For time-course reactions, a shorter LC gradient was used, and digests of the same reporters were combined prior to analysis for higher throughput. The elution gradient used for PDI, etanercept, EPO, and CD16a proteins was 1-80% acetonitrile in 0.1% formic acid over 60 minutes

followed by 5 minutes of 80% acetonitrile in 0.1% formic acid. The peptides were eluted into the source of an Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo Fisher Scientific). The spray voltage was set to 2.25 kV and the temperature of the heated capillary was set to 280°C. Full MS scans were acquired from m/z 300 to 2000 at 60k resolution, and MS/MS scans following collision-induced dissociation (CID) at 38% collision energy were collected in the ion trap. The elution gradient used for SARS-CoV-2 spike glycoprotein was 1-80% acetonitrile in 0.1% formic acid over 300 minutes followed by 10 minutes of 80% acetonitrile in 0.1% formic acid. The peptides were eluted into the source of an Orbitrap Eclipse™ Tribrid™ mass spectrometer (Thermo Fisher Scientific). The spray voltage was set to 2.25 kV and the temperature of the heated capillary was set to 275°C. Full MS scans were acquired from m/z 300 to 1900 at 60k resolution, and MS/MS scans following collision-induced dissociation (CID) at 38% collision energy were collected in the ion trap.

MS data analysis

For the intact glycopeptide analysis, the raw spectra were analyzed using pGlyco3 (156) for database searches with mass tolerance set as 20 ppm for both precursors and fragments. The database search output was filtered to reach a 1% false discovery rate for glycans and 10% for peptides. The filtered result was further validated by manual examination of the raw spectra. For isobaric glycan compositions, fragments in the MS/MS spectra were evaluated to provide the most probable topologies. Quantitation was performed by calculating spectral counts for each glycan composition at each site. Any N-linked glycan compositions identified by only one spectra were removed from the quantitation. For the deglycosylated glycopeptide analysis, the spectra were analyzed using SEQUEST (Proteome Discoverer 1.4 and 2.5, Thermo Fisher Scientific) with mass

tolerance set as 20 ppm for precursors and 0.5 Da for fragments. The search output from Proteome Discoverer 1.4 was filtered using ProteoIQ (v2.7, Premier Biosoft) to reach a 1% false discovery rate at protein level and 10% at peptide level. The search output from Proteome Discoverer 2.5 was filtered within the program to reach a 1% false discovery rate at protein level and 10% at peptide level. Occupancy of each N-linked glycosylation site was calculated using spectral counts assigned to the ^{18}O-Asp-containing (PNGaseF-cleaved) and/or HexNAc-modified (EndoH-cleaved) peptides and their unmodified counterparts.

For time-course reactions, quantitation was performed through manual inspection of MS1 spectra using Thermo Freestyle 1.7 (Thermo Fischer Scientific). The intensities of monoisotopic peak heights for all observable charge states for reactants and products were determined and then summed and averaged in triplicate to determine percent conversion to product over time. Plots generated using RStudio (1.4.1717).

Data availability

All data generated or analyzed during this study are included in this article and supporting information files. The glycopeptide analysis MS data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD032149. MS data for time-course reactions available upon request.

Supporting Information

This article contains supporting figures (**Figs.** S1-S43) and tables (Tables S1-S3), which can be found online with the original published article at the DOI below.

10.1016/j.jbc.2022.102474

Acknowledgements

We thank Dr. Henrik Clausen for providing cell lines for this project.

Funding and additional information

This work was supported in part by National Institutes of Health Grant 5R01GM130915 from NIGMS (to K.W.M. and L.W.) and the Glycoscience Training Program (5T32GM107004) from NIGMS. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations and nomenclature

ER – endoplasmic reticulum, GlcNAc – N-acetylglucosamine, MS – mass spectrometry, MS/MS – tandem mass spectrometry, LC – liquid chromatography

CHAPTER 3

MONITORING THE SITE-SPECIFICITY OF ELONGATING N-GLYCOSYLTRANSFERASES AND THEIR STRUCTURAL DETERMINANTS

Introduction

N-glycans are important contributors to protein stability, binding, and function (1). In addition to these properties, N-glycans are of much interest in the development of biologics and biosimilars where N-glycosylation is recognized as a critical quality attribute (cQA) (216, 217). For these reasons, there is great interest in understanding what defines a particular N-glycosylation site's glycoform profile. There are efforts towards controlling the glycan profile of glycoprotein therapeutics through adjustments to their production and processing (200, 218). However, unlike proteins and nucleic acids, N-glycans are synthesized and processed through a non-template driven process that results in a distribution of glycoforms with differing extents of branching and elongation, many of which result in divergent terminal sugars (219). Glycoprotein characterization is further complicated by the phenomenon of N-glycan microheterogeneity, in which glycoproteins with multiple sites of N-glycosylation show differing distributions of glycans at each site (see Chapter 2).

There is evidence that protein structure impacts N-glycan microheterogeneity. One common model for studying microheterogeneity is protein disulfide isomerase, a resident endoplasmic reticulum (ER) glycoprotein that hosts five sites of N-glycosylation. Losfeld et al. found that at one less-processed site of N-glycosylation, mutagenesis of a tyrosine in proximity

of the glycosite to an alanine resulted in the production of more-processed fucosylated complex structures when expressed in Chinese hamster ovary (CHO) cells (178). The inhibitory effect of the tyrosine was made more pronounced by a phenylalanine. In addition, by mutating an aspartate proximal to a more-processed site of glycosylation into a tyrosine, they were able to slow N-glycan processing at that specific site. Thus, the Aebi group was able to both slow and accelerate N-glycan processing depending on the substitution made (178). Mathew et al. demonstrated that protein disulfide isomerase (PDI) N-glycan processing is altered when the tertiary structure of the protein has an influence on N-glycan microheterogeneity (197), a finding that has since been supported by our previous work (106).

In this chapter, we further explore the relationships between site-specific N-glycan processing rates and wild-type glycan abundance with a focus on the elongating glycosyltransferases B4GALT1 and ST6GAL1. We probed these transfer rates through the use of a roster of reporter glycoproteins expressed in HEK293F cells, all of which have well-characterized crystal structures (Table 3.1). Additionally, we identify acceptor structural features that may have an impact on N-glycan microheterogeneity. We then use statistical analyses to determine which features may be relevant and which are irrelevant and compare our findings with previous work.

Table 3.1: Glycoprotein reporters					
Reporter	**Uniprot ID**	**Amino acids**	**Glycan sites**	**Glycan structures**	**PDB**
Etanercept TNFR-IgG Fc fusion	P20333 + P01857	1-235 + 236-467	3	Varied per site	3ALQ 3AVE
Erythropoietin	P01588	28-193	3	Tri- and tetra-antennary	1EER
SARS-CoV2 Spike	P0DTC2	1-1208	22	Varied high man to complex	6VSB
CD16a Fc ɣ receptor IIIa	P08637	19-192	5	Mostly complex bi-, tri-, and tetra-antennary	5BW7
Pdi1p (yeast)	P17967	29-522	5	Complex, varied	2B5E

Methods

Glycosyltransferase reactions and mass spectrometry analysis were largely carried out as described in Chapter 2 (106). Briefly, glycoproteins and glycosyltransferases were expressed in a secreted form from MGAT1$^{-/-}$ HEK293-F cells and subsequentially purified via Ni-NTA chromotography. These purified glycoproteins were then modified *in vitro* with recombinant glycan-processing enzymes (MGAT1, MAN2A1, and MGAT2) in excess to form substrates for B4GALT1 and ST6GAL1. For the time course reactions for B4GALT1, a 1:100 enzyme:substrate ratio was used while for ST6GAL1 a 1:500 enzyme:substrate ratio was used. Time course reactions took place over 90min, with individual timepoints being boiled after sampling to halt enzymatic activity. Samples were then processed through a standard C18 cleanup as described previously (106). Processed timepoints were then analyzed using tandem LC-MS/MS with a Thermo Orbitrap Fusion mass spectrometer as described previously (106). Data analysis was completed using Thermo Freestyle 1.7 (Thermo Fischer Scientific), where substrate and product ratios were determined by comparing the intensities of their respective monoisotopic peaks. Data analysis used R (4.1.1).

All modeling was completed using PyMol and Chimera to analyze the crystal structures described in Table 3.1. Convexity at each glycosite was determined using Chimera's attribute calculator, where convexity = atom.areaSAS/atom.areaSES with 'atom' being the terminal nitrogen of the asparagine at the glycosite. Statistical analysis used R (4.1.1).

Results

We continued progress from previous experiments with monitoring site-specific N-glycan processing, this time using downstream enzymes that are responsible for extending GlcNAc-terminal hybrid and complex structures into galactosylated and sialylated structures as described in our study from last year (see Chapter 2: *Sequential in vitro enzymatic N-glycoprotein modification reveals site-specific rates of glycoenzyme processing*, or see reference from original publication)(106). Due to a problematic co-precipitation of CD16a and B4GALT1, this previously characterized reporter glycoprotein was excluded from these experiments. Briefly, a suite of four reporter glycoproteins were used as acceptors for glycosyltransferase time-course reactions. Glycopeptide analysis of the individual timepoints allowed us to observe the progress of glycosyltransferase reactions over time in a site-specific manner.

We first monitored the rate of galactosylation of $GlcNAc_3Man5$ N-glycans via B4GALT1 (Fig. 3.1). We chose to transfer to a hybrid N-glycan, as differing rates of elongation along each arm of N-glycans could complicate analysis (88, 89). Rates of galactosylation for erythropoietin (Fig. 1a) and etanercept (Fig. 3.1b) were fairly similar across all sites of their respective glycoprotein, with transfer to erythropoietin greatly outpacing transfer to etanercept. For protein disulfide isomerase, transfer to four of the five sites of N-glycosylation were modified rapidly while a lone site (sequon 4) was not modified at all (Fig. 3.1c). For the reporter glycoproteins

listed so far, this mostly recapitulates results from upstream glycosyltransferase experiments (see Chapter 2). For the SARS-CoV-2 spike glycoprotein, our sample size was unfortunately limited to 14 of the 22 N-glycan sites due to problems in protein digestion and sample processing. However, it is interesting to note that the range of transfer rates for B4GALT1 were distributed more narrowly than what was observed for upstream N-glycan processing enzymes.

We then monitored the rate of sialylation of GlcNAc3Man5Gal1 N-glycans via ST6GAL1 (Fig. 3.2). Like with B4GALT1, we used a hybrid acceptor structure (the product from the B4GALT1 reaction) due to differing branch specificities of ST6GAL1 (90). Again, due to problems with sample processing, we were limited to only 15 of the 22 N-glycan sites on the SARS-CoV-2 spike glycoprotein. Erythropoietin's three sites of glycosylation resulted in divergent rates of sialylation by ST6GAL1 (Fig. 3.2a), which contrasts with the B4GALT1

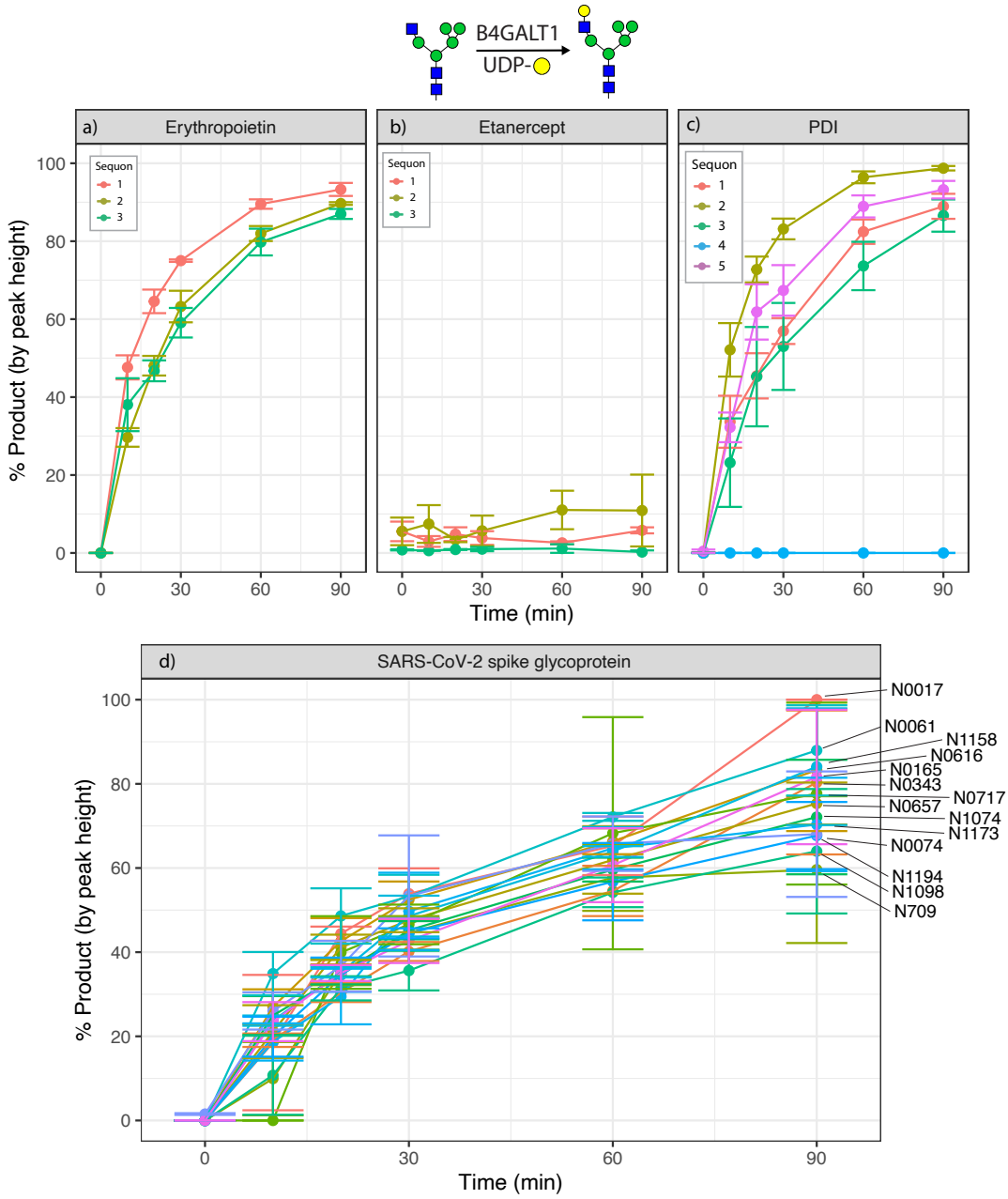


Figure 3.1: Site-specific monitoring of B4GALT1 activity. Time-course reaction of galactose addition to reporter proteins with recombinant B4GALT1. *a*, Erythropoietin. *b*, Etanercept. *c*, Protein disulfide isomerase (PDI). *d,* SARS-CoV-2 spike glycoprotein. Reaction progress calculated as proportion of the sum of monoisotopic peak heights of product ($Gal_1Man_3GlcNAc_4$) vs. the sum of product and reactant ($Man_3GlcNAc_4$) peak heights. Experiments performed in triplicate, error bars represent standard deviation.

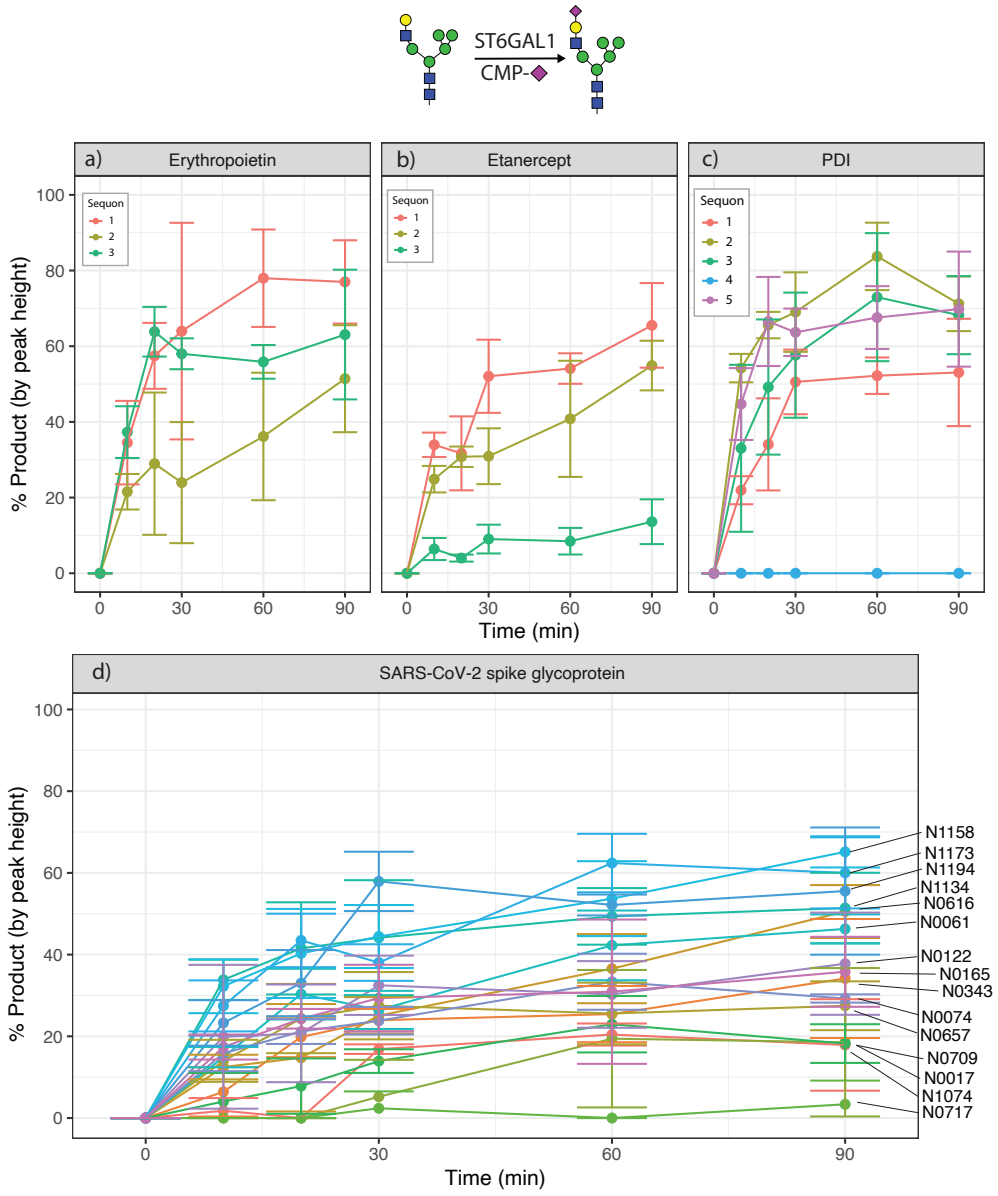


Figure 3.2: Site-specific monitoring of ST6GAL1 activity. Time-course reaction of neuraminic acid (Neu5Ac) addition to reporter proteins with recombinant ST6GAL1. *a*, Erythropoietin. *b*, Etanercept. *c*, Protein disulfide isomerase (PDI). *d,* SARS-CoV-2 spike glycoprotein. Reaction progress calculated as proportion of the sum of monoisotopic peak heights of product ($Sia_1Gal_1Man_3GlcNAc_4$) vs. the sum of product and reactant ($Gal_1Man_3GlcNAc_4$) peak heights. Experiments performed in triplicate, error bars represent standard deviation.

results (Fig. 3.1a). A similar contrast was seen with Etanercept (Fig. 3.2b), in which transfer to sequon 3 was much slower than to the other two sites. PDI behaved similarly to previous experiments, where four sites of N-glycosylation were modified rapidly while sequon 4 was slightly modified, if at all (Fig. 3.2c). For the SARS-CoV-2 spike glycoprotein, overall rates of sialylation were reduced when compared to the other reporter glycoproteins (Fig. 3.2d). A broader range of transfer rates were observed for the viral glycoprotein when compared to the B4GALT1 data (Fig. 3.1d).

With this database of N-glycosites in tow, we next looked to determine what structural features contribute to N-glycan microheterogeneity. To do this, we probed individual glycosites using previously available crystal structures on PDB (Table 3.1) and compared them to N-glycan processing rates over multiple enzymes. In addition to B4GALT1 and ST6GAL1, we included the previously characterized MGAT1, MAN2A1, MGAT2, and FUT8 (see Chapter 2). For the dependent variable in these analyses, we used the reaction progress at the 20-minute timepoint since the rate was linear for most glycosites at that point (Figs. 3.1, 3.2, and Chapter 2). Some sites were excluded from this analysis on the basis that they were not resolved in their associated protein's crystal structure (CD16a site 1, SARS-CoV-2 spike sites 1,3,5,9,20-22). While these structures likely include loops whose flexibility disrupts resolution, there was no significant difference between reaction progress distribution of unresolved and resolved glycosites for all studied glycan-processing enzymes (Fig. 3.3a).

We began by interrogating basic primary and secondary structural properties to see if we could find any relationship between them and site-specific glycan processing. First, we compared the rates of sequons when they had either a serine or a threonine in the +2 position (e.g. NxS vs. NxT) (Fig. 3.3b). We found no significant relationship between the +2 position and transfer rate

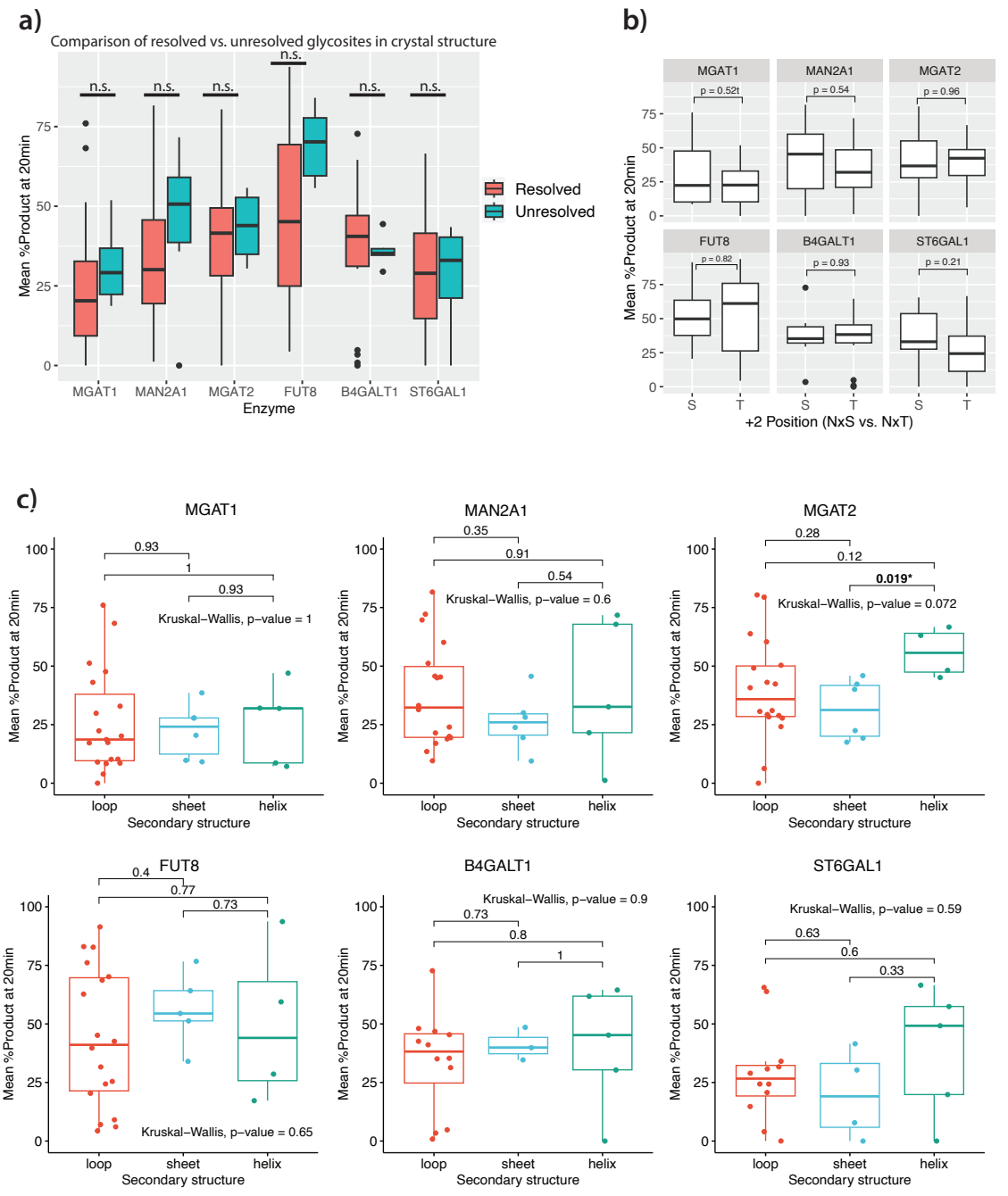


Figure 3.3: Sequence and secondary structures do not trend with most N-glycan processing. *a,* Comparison of modification rates of crystallizable vs. non-crystallizable glycosites. Wilcoxon rank sum test. *b*, Comparison of modification rates of NxS vs. NxT N-glycan sequons. Wilcoxon rank sum test. *c*, Comparison of modification rates of differing secondary structures, non-crystallizable glycosites omitted. Kruskall-Wallis test used in conjunction with Wilcoxon test for pairwise comparisons.

for any of the studied enzymes. Next, we assigned every N-glycan site to a specific secondary structure based on its published crystal structure (Fig. 3.3c). Sites that do not resolve in their crystal structure were omitted from this analysis. While secondary structure as a whole was not a significant contributor to the productivity of any glycan-processing enzyme, there were some minor differences in their distributions, most notably with slightly elevated values for MGAT2 activity at glycosites on helices when compared to sheets.

Next, we determined if proximal residues could influence N-glycan processing rates (Fig. 3.4). We investigated three ringed amino acids: tyrosine, phenylalanine, and histidine. The rationale for this was that they were all identified as residues that may contribute to rates of N-glycan processing as evidenced the by Losfeld et al. studies on protein disulfide isomerase (178). We initially searched within these residues within four residues of the glycosite in the protein sequence, as this positions a residue directly next to a neighboring residue in a helical turn, which was the rationale in the previously described study (178). For tyrosine and histidine, we found no significant relationship with any glycan-processing enzyme (Fig. 3.4a,b). However, with phenylalanine we found a significant decrease in MGAT1 activity as more phenylalanines were found in proximity (Fig. 3.4c). This relationship dissipated when instead counting the number of surface phenylalanines found within 7.5Å of the glycosylated asparagine (Fig. 3.4d). However, note that in contrast to the primary sequence analysis, this analysis omitted non-crystallizable glycosites, which likely reduces the power of the analysis.

We then probed the relationship between the shape of the protein surface and site-specific glycan processing (Fig. 3.5). The convexity at a given position on a protein surface can be calculated by finding the ratio between the solvent exposed surface area and the solvent excluded surface area over a given region of a protein. We calculated the convexity of the protein surface

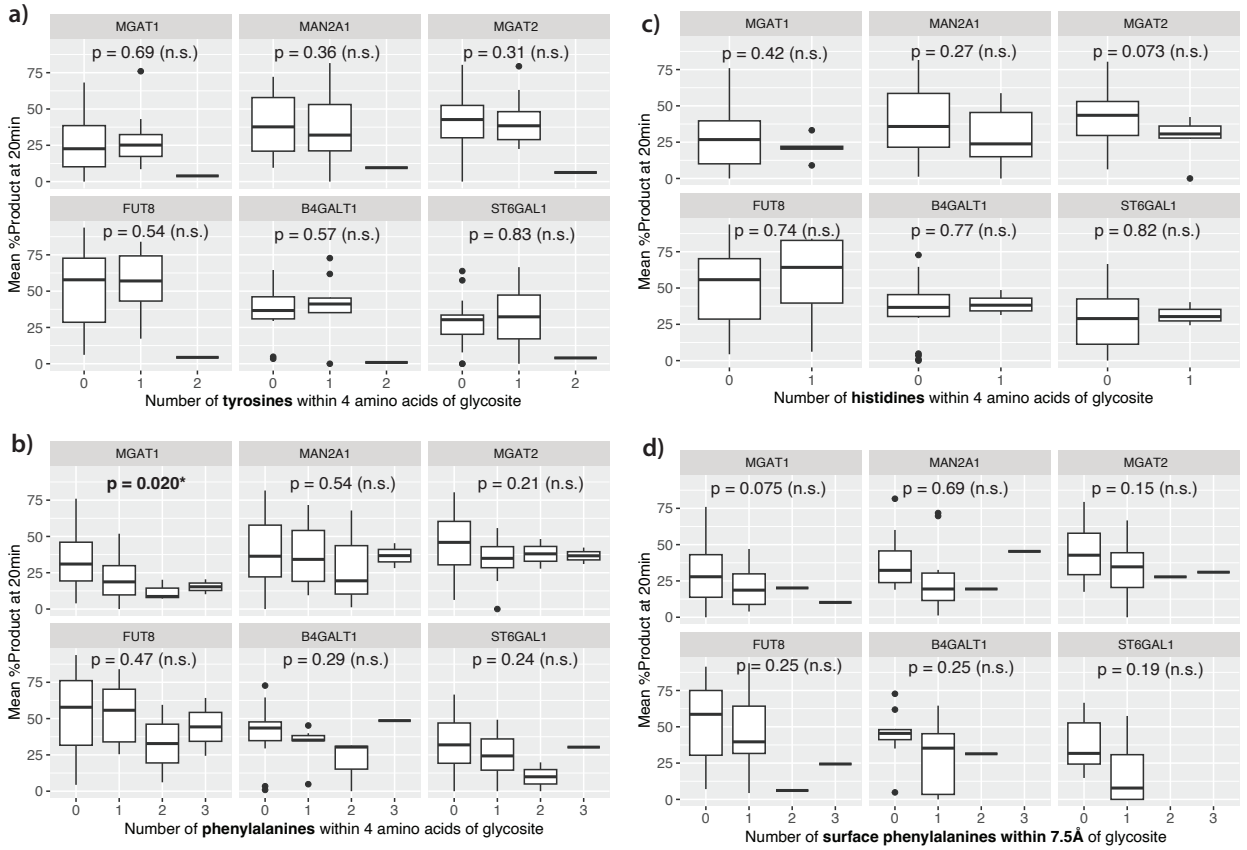


Figure 3.4: Relationship between proximal ringed amino acids and N-glycan processing. Mean N-glycan activity rates for glycosites with differing number of ringed residues. *a*, Tyrosines +/- 4 amino acids from glycosite. *b*, Histidines +/- 4 amino acids from glycosite. *c,* Phenylalanines +/- 4 amino acids from glycosite. *d,* Surface phenylalanines within 7.5A of glycosylated asparagine. For all, the statistical method used was a one-way ANOVA.

at the point of each asparagine that hosts an N-glycan sequon, where a value <1 indicates the surface is concave and a value >1 indicates the surface is convex. Most sites of N-glycosylation lie on convex regions, which is reasonable considering that the surface of a perfectly spherical globular protein would be entirely convex.

For MGAT1, we initially saw a significant but very weak relationship between convexity and GlcNAc transfer (Fig. 3.5a). However, when we omitted the densely glycosylated SARS-CoV-2 spike glycoprotein from analysis, we saw a much stronger trend. We repeated this

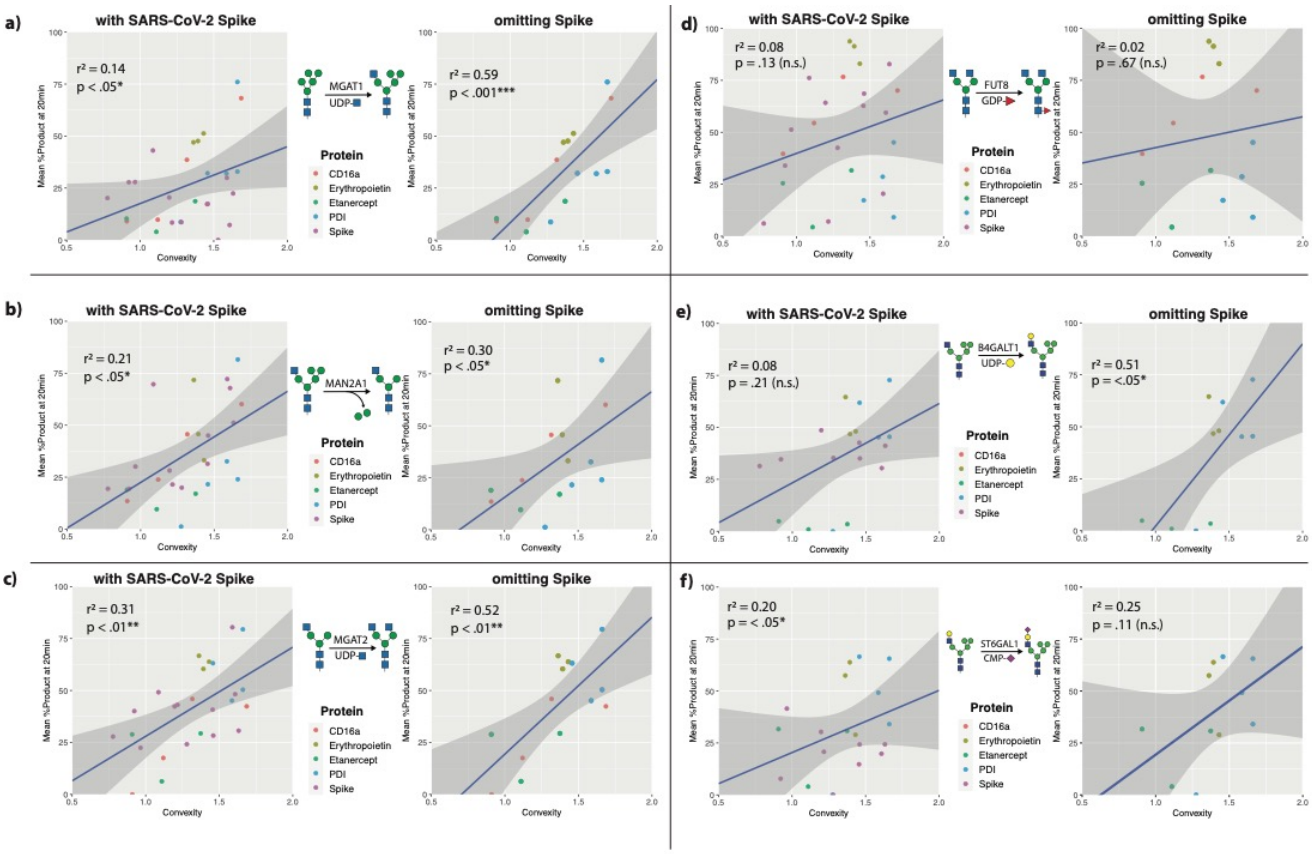


Figure 3.5: Impact of convexity on N-glycan processing. Relationship between convexity of the protein surface at glycosylated asparagines and the activity of *a,* MGAT1. *b,* MAN2A1. *c,* MGAT2. *d,* FUT8. *e,* B4GALT1. *f,* ST6GAL1. All plots used linear regressions with the shaded area denoting 95% confidence intervals.

analysis for all the other characterized glycan-processing enzymes. With MAN2A1 (Fig. 3.5b) and MGAT2 (Fig. 3.5c), we saw similar but weaker relationships between convexity and enzyme activity. Omitting the SARS-CoV-2 spike glycoprotein did result in a stronger trendline, but only slightly when compared to MGAT1. We found no relationship between convexity and the rate of fucosylation by FUT8 (Fig. 3.5d). We did see a trend with B4GALT1 that was similar to what was seen with MGAT1, including the recovery of the relationship with the omission of the spike glycoprotein (Fig. 3.5e). Finally, with ST6GAL1 we saw a significant trend when the spike

glycoprotein was included, but not when it was omitted (Fig. 3.5f). For the significant relationships, the r^2 value ranged from 0.14-0.31 (0.30-0.59 when omitting Spike), indicating that convexity of the acceptor site can explain some of, but not all, the observed variance of transfer rates between sites on the reporter glycoprotein.

Discussion

In this chapter, we continued our progress from Chapter 2 and continued characterizing N-glycan processing rates for elongating enzymes. Since we had already determined that protein structure is responsible for much of observed N-glycan microheterogeneity, we then investigated multiple structural features of each glycosite that may influence glycan-glycosyltransferase (or glycosyl hydrolase) interactions.

By far the strongest trend with N-glycan processing found in this study was protein surface convexity (Fig. 3.5). This was initially hypothesized by Mathew et al. (197) when studying protein disulfide isomerase. Interestingly, this trend was strongest for MGAT1, but only when the SARS-CoV-2 spike glycoprotein was omitted from analysis. This suggests that in the committed step towards hybrid and complex N-glycans, densely glycosylated viral glycoproteins may be defined by other features such as glycan-glycan interactions. In contrast, we saw the opposite relationship with ST6GAL1, in which removing the SARS-CoV-2 spike glycoprotein from analysis ablated the relationship between convexity and ST6GAL1 productivity. It is possible that glycan-glycan interactions dominate when the $GlcNAc_2Man_5$ structure is in place, but that its influence starts to fade as the structure is extended away from the protein surface. However, due to this extension from the surface, there is a concomitant loss of convexity's power as well, and the relationship only recovers due to a higher sample size. It would be

interesting to test this on dense high-mannose regions of glycoproteins such as those found on HIV-1 Env (220). Perhaps by knocking out surrounding sites of N-glycosylation, the relationship with protein surface convexity and early MGAT1 processing could be restored.

The relationship between MGAT1 and proximal phenylalanines is particularly interesting (Fig. 3.4a). If one wished to shift the glycoprofile of a site of N-glycosylation away from hybrid/complex structures and towards high- or paucimannose structures, perhaps one could attempt to add phenylalanines proximal to the glycosite. Since MGAT1 is the committed step in hybrid N-glycan synthesis, inhibiting this enzyme would generate a separate pool of high- or paucimannose N-glycans while the glycans that are processed by MGAT1 would be decorated with terminal structures as usual.

We identified several structural features that do not appear to impact glycosyltransferase-glycoprotein interactions (Fig. 3.3). It is reasonable that the +2 position of glycosites does not greatly impact the rate of N-glycan processing, as both serines and threonines are less bulky than asparagines, so the glycan's terminal features protrude reasonably far away from these amino acids in space. However, it should be noted that this is in contrast to core-glycosylation via oligosaccharyltransferase, which acts at the site of the asparagine rather than the glycan (79). It is somewhat surprising that secondary structure does not appear to have a strong relationship with N-glycan processing, although this may be due to enzyme evolution accounting for all basic motifs. However, N-glycans are found more often in loop regions and random coils (80), so this may just be survivorship bias.

Future work could perhaps emulate the heterogenous nature of the Golgi apparatus through a competitive assay, where multiple reporter proteins are incubated with glycosyltransferases. Interestingly, B4GALT1 and ST6GAL1 are known to be able to form

heterodimers, a phenomenon that has also been observed with the GlcNAc transferases MGAT1 and MGAT2 (221). Future experiments may look to study the effect of this these heterodimers on site-specific rates of glycan processing, as this may impact glycoprotein-glycosyltransferase interactions. Additional possibilities include the use of crowders to recapitulate the crowded nature of the secretory pathway, increasing the diversity of reporter glycoproteins studied, and further extending the characterization of glycosyltransferases to include branching enzymes such as MGAT4A, MGAT4B, MGAT5, and the truncating MGAT3.

CHAPTER 4

PPMFIXER: A MASS ERROR ADJUSTMENT FOR PGLYCO3.0 TO CORRECT NEAR-ISOBARIC MISMATCHES

Adams, T.M., Zhao, P., Wells, L. Submitted to *Glycobiology*.

Abstract

Modern glycoproteomics experiments require the use of search engines due to the generation of countless spectra. While these tools are valuable, manual validation of search engine results is often required for detailed analysis of glycopeptides as false-discovery rates are often not reliable for glycopeptide data. Near-isobaric mismatches are a common source of misidentifications for the popular glycopeptide-focused search engine pGlyco3.0, and in this technical note we share a strategy and script that improves the accuracy of the search utilizing two manually validated datasets of the glycoproteins CD16a and HIV-1 Env as proof-of-principle.

Introduction

As mass spectrometry instrumentation matures, more in-depth data is being produced than ever before. The complexity of glycoproteomic data necessitates the use of specialized search engines (e.g. Byonic (152), pGlyco (156), MSFragger (163)) to assist the automated and high-throughput data analysis of intact glycopeptides (222). However, one of the biggest challenges for glycoproteomic data analysis is the quality control of the search output, where a false discovery rate (FDR) evaluation is often underestimated leading to inaccurate glycopeptide spectral matches (GPSMs). Manual validation can increase the accuracy of GPSMs but requires large amounts of analyst time, effort, and expertise to complete as the searches often scale to thousands of GPSMs (223). Manually validated results can differ significantly from the FDR-filtered search engine output. The pGlyco search engine is of particular interest because it computes respective FDR values on peptide and glycan moieties (N-linked or O-linked) as well as an integrated FDR value on glycopeptide entities therefore providing a comprehensive FDR estimation, whereas most of the other search engines only consider FDRs on the peptides (156). To streamline the process of

validating pGlyco search outputs, in this technical note, we looked to identify common errors in glycopeptide spectral matching and build automated solutions.

Results and Discussion

When examining large sets of spectra that failed manual inspection, common themes emerge. A major issue is the monoisotopic mass being incorrectly assigned to the 1 or 2 ^{13}C-isotope peak that shifts the predicted mass of the glycopeptide by ~1-2 daltons. Glycopeptides, in general, are more likely to have large masses and a higher number of Carbon atoms than a typical non-glycosylated peptide resulting in the base peak often not being the monoisotopic peak. This incorrect calculation of the monoisotopic mass can result in common errors given that various combinations of sugars result in mass shifts of ~1 or ~2 daltons. One of the most common errors found in glycopeptide spectral matching search engines is confusion between a pair of fucoses (Fuc) and a single *N*-acetylneuraminic acid (Neu5Ac). Two fucoses are roughly 1 Dalton (1.0204) greater than a single Neu5Ac. This problem can even be doubled, where 2 sialic acids are mistaken for 4 fucoses, which leads to erroneous GPSMs with multiple fucoses when the peak with two-^{13}C isotopes is misidentified as the monoisotopic peak. Note the small difference in weight between two fucose for 1 Neu5Ac (1.0204) and a single proton (1.007276).

In some cases, this problem can be easily identified with the presence of Neu5Ac oxonium ions in the MS/MS spectra (m/z 274 and m/z 292). An example of an assigned non-sialylated, triply-fucosylated glycopeptide is shown that is in fact a mono-fucosylated, mono-sialyated glycopeptide by manual examination (Fig. 4.1a, denoted using symbol nomenclature for glycans (224)). However, spotting this error becomes more difficult in glycopeptides with both and multiple Neu5Acs and fucoses, as the presence of Neu5Ac oxonium ions can be explained by the presence of a single or multiple Neu5Acs. Additionally, the Y-ions (defined as the peptide

backbone with glycan attached, Fig. 4.1a) rarely contain terminal monosaccharides, such as Neu5Acs and fucoses, due to non-reducing end fragmentation as well as limited mass range in the mass spectrometer. The neutral losses of the terminal Neu5Acs and fucoses are seldom observed for the aforementioned reasons as well. The latest version of pGlyco at the time of writing (pGlyco3.0_build20210615) attempts to find Neu5Ac oxonium ions (m/z 274 and m/z 292) in multiply fucosylated GPSMs and corrects them to assignment with Neu5Acs where the Neu5Ac

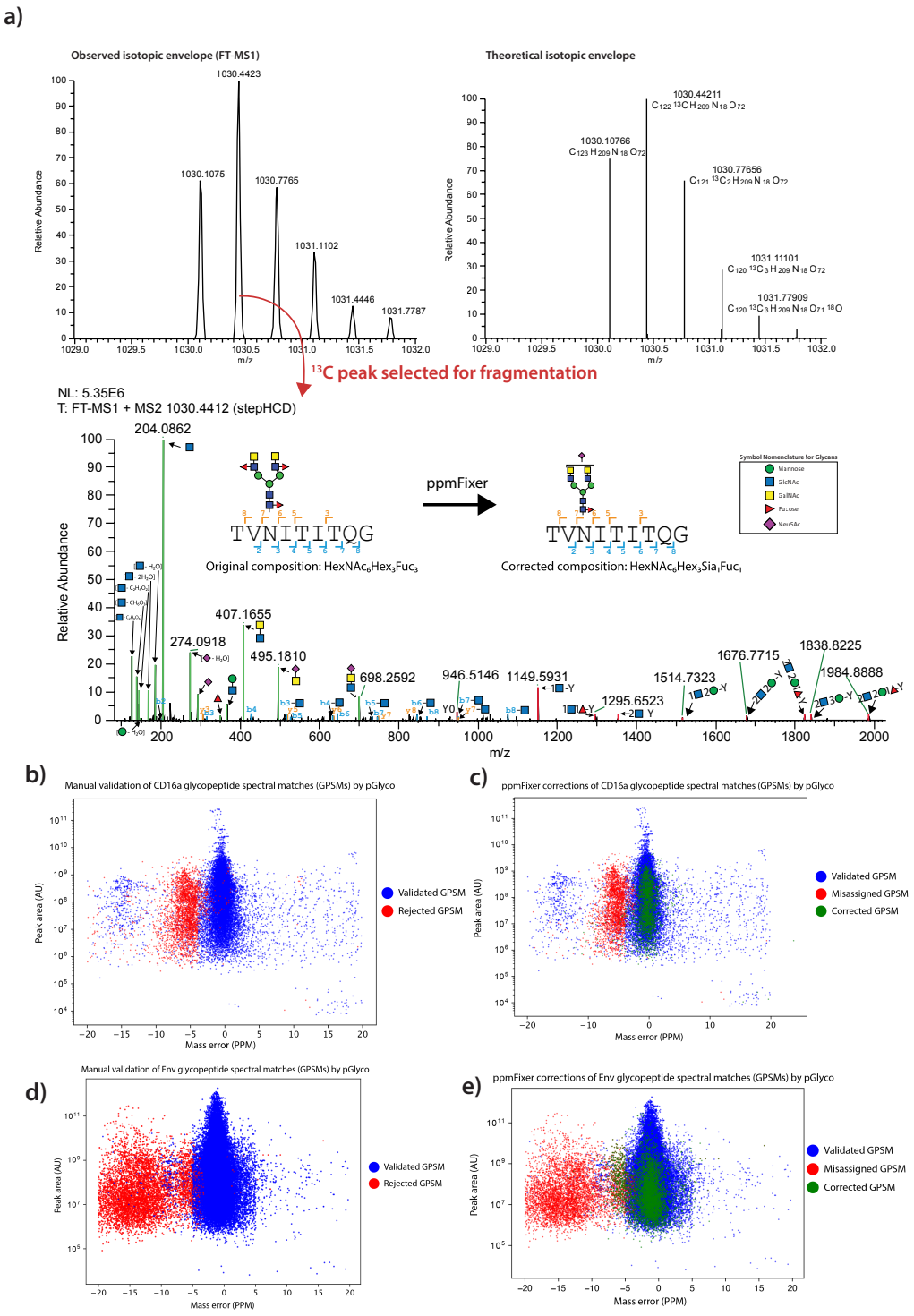


Figure 4.1: pGlyco misidentification is associated with a mass error shift. *a,* Example spectra of a misidentified CD16a glycopeptide being corrected using ppmFixer. Upper – MS1 precursor ion selected (left) and theoretical isotopic envelope (right). Lower – MS2 from precursor ion stepHCD fragmentation (Color guide - blue: peptide backbone b-ion, orange: peptide backbone y-ion, red: peptide backbone with glycan-attached Y-ion, green: oxonium ion). *b,* Mass error vs. monoisotopic peak area of identified CD16a GPSMs (glycopeptide spectral matches), colored by agreement with manual validation. *c,* Demonstration of the shift of revised CD16a GPSMs (green) from original mass error (red). GPSMs rejected in both sets excluded from plot. *d,* Demonstration of the shift of revised Env GPSMs (green) from original mass error (red). GPSMs rejected in both sets excluded from plot. *e,* Demonstration of the shift of revised Env GPSMs (green) from original mass error (red). GPSMs rejected in both sets excluded from plot.

oxonium ions are present. This results in a few additional output columns from the program (e.g. "CorrectedGlycan(x,x,x,x)") that reflect where corrections have been made. While this solution seems to fix some of the erroneous GPSMs, it does not capture all of them. This motivated us to find additional solutions to the problem.

Another common error in pGlyco search output is the confusion between four N-acetylhexosamines (HexNAc) and five hexoses (Hex) where the mass difference is approximately 2 Daltons (2.0534). For reasons discussed earlier, such as multiply-charged species reducing mass difference as well as mis-identified monoisotopic peaks, this confusion often leads to the mis-assignment of high mannose (Man) and *N*-acetylgalactosamine β1-4*N*-acetylglucosamine (GalNAcβ1-4GlcNAc, or LacDiNAc) glycans. For example, the program often mis-identifies $Man_8GlcNAc_2$ as $HexNAc_4Man_3GlcNAc_2$ with the two-^{13}C peak being misidentified as the monoisotopic peak (note that two-13C peak is a mass shift of 2.01455 daltons).

An additional abundant error is the misidentification of certain high-mannose glycans as complex or bizarre (and unlikely) hybrid N-glycans containing a single hexose. The misidentified glycans have two GlcNAcs, three fucoses, and a sialic acid assigned which is ~1 dalton (1.0581) heavier than simply being 7 hexoses.

An alternative approach to full manual validation (and the one being proposed in this manuscript) involves a comparison of mass errors to the average of the population. In order to test our approach, we utilized a highly detailed dataset of the human glycoprotein CD16a that we had previous characterized (106). CD16a has 5 sites of N-linked glycosylation and our analyses generated 18,586 GPSMs (see Supplementary Methods). When individual GPSMs from this dataset are plotted against their mass error, there is a distribution of GPSMs that are both offset from the overall distribution of mass errors and are rejected during manual validation (Fig. 4.1b). During manual validation, it was noted that a large proportion of these corrections were due to the

near-isobaric conundrums detailed previously. This is caused by these glycan species not being *exactly* isobaric; there is a small offset caused by nuclear binding energy that is reflected in the mass error of the resulting GPSMs (for example proton weighing 1.007276 daltons but 2 fucoses weighting 1.0204 daltons more than a Neu5Ac). To correct this issue, we developed a Python script titled 'ppmFixer' that does the following:

1) Identifies GPSMs with a mass error >1 standard deviation from the mean mass error.

2) Recalculates mass error for these GPSMs if they fit the following cases:

   a. Where #Fucose ≥ 2, two fucoses are replaced by a Neu5Ac, move down an isotope.

   b. Where #Fucose ≥ 4, two fucoses are replaced by a Neu5Ac, move down two isotopes.

   c. Where #HexNAc > 5, 5 hexoses are added and 4 HexNAcs removed, move down two isotopes.

   d. Where #HexNAc = 4, 7 hexoses are added and 2 HexNAcs removed and all sialics/fucoses removed, move down two isotopes.

3) If |new mass error| < |original mass error|, adjusts the composition. Outputs this adjusted composition as a new column in the pGlyco output file (CorrectedGlycan_ppmFixer). Note: If multiple cases apply, accept the revision with the lowest mass error. Additionally, the native pGlyco correction utilizing the sialic acid oxonium ion takes precedence over all other corrections due to spectral evidence.

4) Recalculates mass error average and stdDev with new corrections, cycles through script again.

5) Once a round adds no new changes, stop execution.

6) Outputs the adjusted compositions and mass errors as new columns in the pGlyco output file (CorrectedGlycan_ppmFixer, Glycan(H,N,A,F)_ppmFixer).

This script identifies a significant proportion of the misassigned and mass offset GPSMs resulting in mass error shifts more in line with the rest of the distribution (Fig. 4.1c). We also tested ppmFixer on a larger glycoproteomics dataset from the HIV-1 envelope glycoprotein (Env), that we have previously characterized with sequence perturbations many times (70, 190, 225–227), where we observed similar patterns of mass error shifts (Fig. 4.1d,e).

We probed what corrections were being made in both the CD16a and Env datasets. The two datasets showed diverging revision profiles (Fig. 4.2a), with the most common revision in the CD16a dataset being the conversion of two fucoses two a sialic acid (F2 -> A1) and the most common revision in the Env dataset being the conversion of a complex glycan to high-mannose (NxHy -> N(x-4)H(y+5)). This is consistent with the fact that the N-glycans from CD16a are dominated by highly sialylated, complex glycans while the HIV-1 Env has a large degree of high mannose N-linked glycans. This demonstrates that the revisions needed can vary from dataset to dataset and justifies the inclusion of all revision types, even if they may not greatly impact a specific experiment.

The results of the ppmFixer output were compared to fully manually validated data, which can be viewed as a positive control that automated results should seek to emulate. Utilizing the CD16a dataset, ppmFixer improved the default pGlyco output from a 76.5% match with manual validation to an 88.0% match with manual validation, a large improvement in GPSM accuracy (Fig. 4.2b). Additionally, ppmFixer erroneously adjusted only 0.37% of total GPSMs, which is a marginal error compared to the 11.7% of total GPSMs that were corrected (Fig. 4.2c). This effectively means that where ppmFixer made a change, 97% of the time it agreed with manual validation for CD16a. ppmFixer effectively cuts overall GPSM misidentifications to half of their volume when compared to the default pGlyco output for this CD16a dataset. ppmFixer performed

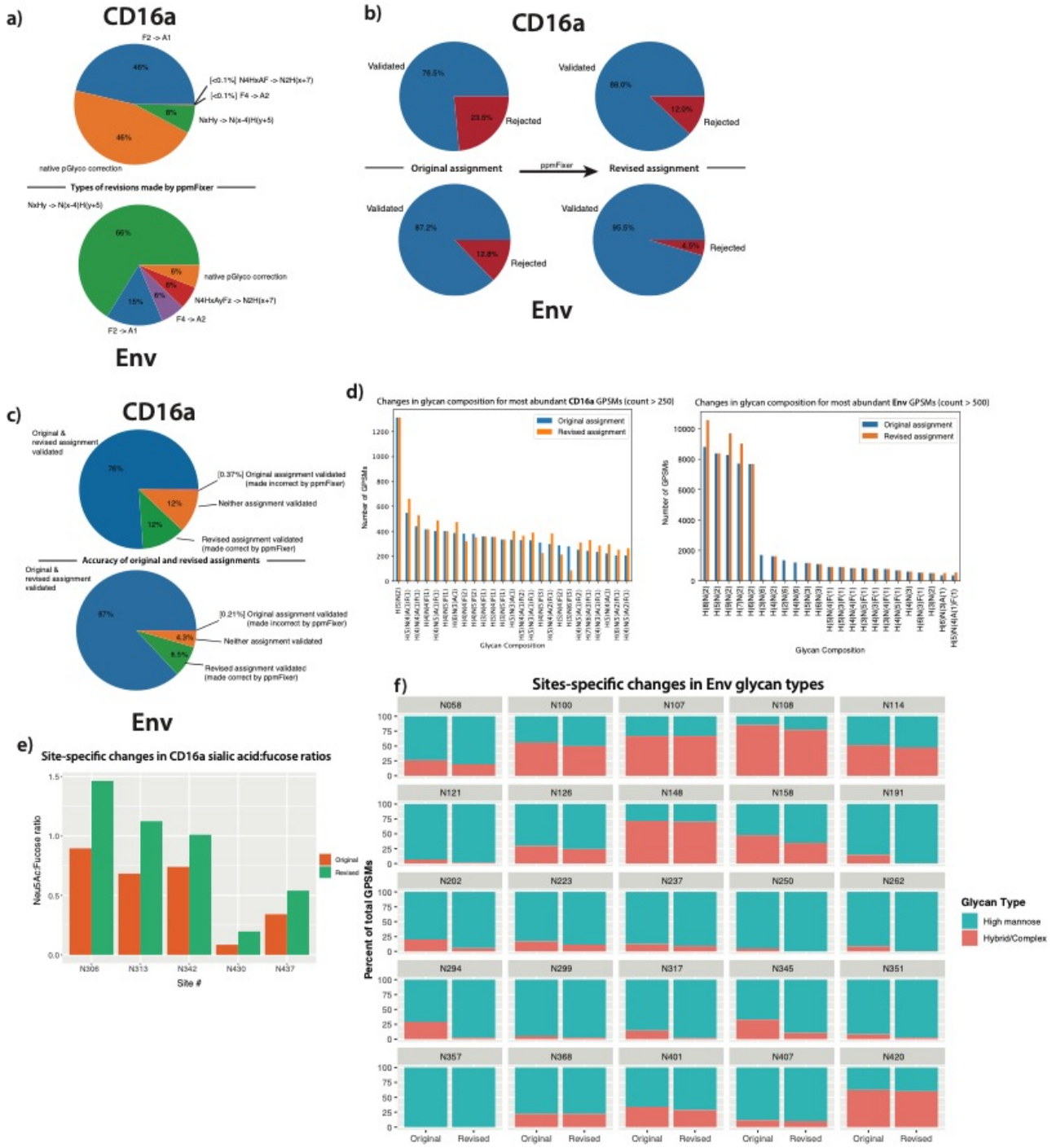


Figure 4.2: ppmFixer adjusts pGlyco results to more closely match manual validation for multiple datasets. *a,* Types of revisions made by ppmFixer in CD16a and Env datasets *b,* Comparison of agreement with manual validation between default pGlyco output and ppmFixer (revised) output. *c,* Depiction of ppmFixer accuracy.. *d,* Demonstration of the effect of the adjustment on the most abundant glycoforms (for CD16a: filtered for GPSM count > 250, for Env: filtered for GPSM count > 500). *e,* Effects of ppmFixer revisions on levels of CD16a Neu5Ac:fucose ratios. *f,* Effects of ppmFixer on site-specific Env N-glycan classes.

similarly well with the Env dataset, with agreement with manual validation increasing from 87.2% to 95.5% (Fig. 4.2b). Next, we compared the most abundant GPSMs identified in the default pGlyco output to those found in the revised ppmFixer output (Fig. 4.2d). Substantial changes were observed, with many of the top glycan compositions being affected. In particular, in the Env dataset, 3 of the top 10 most abundant glycoforms were eliminated entirely. We then compared the default and revised outputs with regards to fucosylation and sialylation for the CD16a dataset (Fig. 4.2e). Overall, ppmFixer output contained a greater proportion of sialic acids in comparison to fucoses when compared to the default pGlyco output. Identification of this error explains the previous published finding that pGlyco output contains more fucosylated glycans than other search engines (228). Finally, we examined how ppmFixer altered site-specific glycan distributions by comparing relative amounts of glycan types using the Env dataset (Fig. 4.2f). The resulting data illustrates a shift away from hybrid/complex N-glycans and towards high-mannose N-glycans. This can be explained by the most common correction in this dataset being the conversion of a complex N-glycan to high-mannose (Fig. 4.2a). This data demonstrates that ppmFixer's revisions can have a dramatic effect on site-specific glycopeptide analysis.

Overall, ppmFixer is a simple, yet effective, improvement to the current pGlyco3.0 workflow. While manual validation remains the gold standard of glycopeptide analysis, corrections to search engine results help reduce the burden of the process while also providing more accurate results especially for high-throughput analyses where manual validation isn't feasible.

Acknowledgements

Support for this work was provided by the NIH (R01AI162267 to R.K.), Emory Vaccine Center (R.K.), Georgia Research Alliance (R.K.), and Glycoscience Training Program (5T32GM107004) from NIGMS (T.M.A.).

Code availability

ppmFixer can be downloaded from https://github.com/trvadams/ppmFixer.

Methods

Protein expression

CD16a (immunoglobulin gamma Fc region receptor III-A, Unprot: P08637, residues 19-193) was expressed in a pGEn2 construct with NH_2-terminal 8xHis, AviTag, superfolder green fluorescent protein, and the 7 amino acid tobacco etch virus (TEV) protease tags in HEK293F cells. Protein was expressed and purified as previously described (213). Briefly, cell culture was maintained at 0.5-3.0 x10^6 cells/mL at 37°C, 5% CO_2, 120rpm. The plasmid was transiently transfected into HEK293F cells (Freestyle 293-F cells, Thermo Fisher Scientific) using polyethylenimine (linear 25kDa PEI, Polysciences, Inc), and protein was secreted into medium. 24 hours post-transfection, cells were diluted with an equal volume of fresh media supplemented with valproic acid (VPA) for a final concentration of 2.2mM VPA. Transfection was continued for 48 hours post-dilution, after which cells were harvested at 1200rpm for 10 minutes and the supernatant was clarified by centrifugation at 3500rpm for 15minutes at 4°C. Clarified protein extract was filtered with a 1μm syringe filter, then run over a Ni-NTA resin (EMD Millipore) gravity column. Flow through was readded, then the column was washed 3 times with 5 column

volumes (CVs) 20mM HEPES pH 7.5, 300mM NaCl, and 20mM imidazole. Protein was eluted with the previous solution but with 300mM imidazole. Protein was then dialyzed with 12,000-16,000 kDa cutoff tubing (Fisher) into 20mM HEPES pH 7.5, 300mM NaCl, 10% glycerol.

CH505.Chim.DS.SOSIP trimer protein was produced in 293FS cells and purified with VRC01 affinity and gel filtration chromatography as previously described (229).

Mass spectrometry analysis

Purified CD16a was reduced with 10mM DTT (Sigma) at 56° for 1hr and then alkylated with 27.5mM iodoacetamide (Sigma) for 45min in the dark. Aliquots of CD16a were then digested using either chymotrypsin (Athens Research and Technology), a combination of chymotrypsin and GluC (Promega), or AspN (Promega). The resulting glycopeptides were separated on an Acclaim PepMap RSLC C18 column (75 μm × 15 cm) and eluted into the nanospray source of an Orbitrap Fusion Eclipse Tribrid mass spectrometer at a flow rate of 200nl/min. The elution gradient consisted of 1% to 40% acetonitrile in 0.1% formic acid over 220 min followed by 10 min of 80% acetonitrile in 0.1% formic acid. The spray voltage was set to 2.2 kV and the temperature of the heated capillary was set to 275°C. Full MS scans were acquired from m/z 200 to 2000 at 60k resolution, and MS/MS scans following higher energy collisional dissociation with stepped collision energy (15%, 25%, 35%) were collected in the Orbitrap at 15k resolution. Raw spectra were analyzed using pGlyco3 (156). Mass tolerance was set at 20ppm for both precursors and fragments. Database output was filtered for a 1% total false discovery rate. Results were validated through manual examination of spectra. Data analysis was completed with Python.

CHAPTER 5

DISCUSSION

Conclusions

This dissertation has described the development of an experimental system that was used to investigate the role of acceptor glycoprotein tertiary structure in defining N-glycan microheterogeneity. By expressing a set of well-characterized glycoproteins in a uniform cell background we generated a large N-glycoproteomic dataset that was then used as a point of reference for a series of *in vitro* time-course reactions that recapitulated N-glycan processing step-by-step (Chapter 2). With this information, we then probed structural features at each site of N-glycosylation to find motifs that correlated with the rate of N-glycan processing in an enzyme-specific manner (Chapter 3). We also discuss some of the problems associated with modern glycopeptide analysis encountered during these studies and developed strategies for increasing the accuracy of mass spectrometry search engines (Chapter 4).

Acceptor glycoprotein structure defines N-glycan microheterogeneity

The clearest conclusion from this work is that tertiary structure of the substrate glycoprotein is directly involved with the development of N-glycan microheterogeneity. This supports earlier work, principally by the Aebi group (104, 105, 178, 197), which established the importance of local structure in defining the extent of processing of specific sites on protein disulfide isomerase (PDI), which has 5 sites of N-glycosylation. The Aebi group's studies

principally involved endoplasmic reticulum and early Golgi glycosyltransferases and glycosyl hydrolases. We continued these studies by using a broader panel of acceptor glycoproteins that extended the number of N-glycosites to 38. The finding that reducing a fully folded glycoprotein to glycopeptides abolishes all glycosyltransferase site-specificity suggests that the tertiary structure of the protein is responsible for most, if not all, of the site-specific rates of N-glycan processing enzymes. This held for both MGAT1 and FUT8, suggesting that this may apply to multiple glycosyltransferase folds.

Acceptor surface convexity is a correlated with the extent of N-glycan processing

We measured the convexity of the protein surface at the site of the asparagine for all crystallizable N-glycosites and found a significant relationship between convexity and the rates of MGAT1, MAN2A1, and MGAT2 activity, but not FUT8 activity. This relationship can be rationalized as a glycan that is protruding outwards from the protein surface (convex) would be more accessible to modification than a glycan that is recessed from the protein surface (concave). Interestingly, the r^2 of these relationships increases substantially when the SARS-CoV-2 spike glycoprotein is omitted from analysis, indicating that processing rates of densely glycosylated proteins may be more defined by other factors, such as glycan-glycan interactions. However, the power of the numerous glycosites that the viral glycoprotein appeared to improve to hint that this may dissipate as the glycosites are extended. It should be noted that many characterized N-glycan sites were not included in this analysis due to not being resolved in the glycoprotein crystal structure, thus making convexity measurements impossible. Techniques such as Cryo-EM that are able to characterize protein surfaces may serve to improve our studies in this regard.

Primary sequence and secondary structure are not predictive of N-glycan microheterogeneity

Chapters 2 and 3 described several experiments that reinforced the importance of tertiary structure in defining N-glycan microheterogeneity. The only evidence of a potential impact involves the interaction of proximal phenylalanines and MGAT1, which may be useful in restricting N-glycan processing at specific sites of multiply N-glycosylated glycoproteins. Meanwhile, secondary structure, the identity of the +2 amino acid, and proximal bulky ringed structures appear to have little to no bearing on how glycoproteins interact with glycosyltransferases. Thus, future studies will need to fully analyze the 3-dimensional space that glycans occupy and interact with.

Development of new research tools

An in vitro *experimental model for studying GT-A glycosyltransferases*

The experiments in these studies established that *in vivo* GT-A glycosyltransferases site-specificity can be recapitulated *in vitro*. As mentioned previously, this system builds on previous work and findings by the Aebi group. In contrast to the Aebi group, we used a broader range of N-glycosites and a series of cis- to medial- Golgi enzymes. The major benefit of this approach is that it allows for statistical analysis that identifies common structural features associated with either slow or fast rates of N-glycan processing, and the enzymes targeted are responsible for the divergence of N-glycans into their three major classes: high-mannose, hybrid, and complex. In addition, we applied this system to the elongating glycosyltransferases B4GALT1 and ST6GAL1 which are more relevant in predicting hybrid and complex N-glycan terminal features.

While this system worked well for GT-A fold glycosyltransferases, it did not appear to accurately recapitulate FUT8 specificity, which is a GT-B fold glycosyltransferase. Strong

evidence for this can be seen in the observed specificity of *in vitro* FUT8 versus its observed activity *in vivo*. *In vitro*, FUT8 is unable to efficiently core fucosylate $GlcNAc_2Man_5$ glycans on erythropoietin. However, when this same glycoprotein was expressed in a $MGAT1^{-/-}$ background, it resulted in substantial core fucosylation, indicating that FUT8 is indeed able to core fucosylate high-mannose N-glycans *in vivo*. In addition, we observed that the structural features that were predictive of transfer rates for other enzymes were not predictive for FUT8. However, without confirmation that the *in vitro* reactions recapitulate *in vivo* activity, we cannot conclude whether this finding is biologically relevant. Further work is needed to accurately model FUT8 activity *in vitro* and to determine if conditions in the Golgi apparatus alter enzyme specificity.

Another interesting finding was the identification of hybrid N-glycans at specific sites on glycoproteins expressed in an $MGAT1^{-/-}$ background. The presence of these hybrid N-glycans suggests that N-acetylglucosamine (GlcNAc) can be added to high-mannose N-glycans through a mechanism not involving MGAT1. Additional studies by our lab found that the GlcNAc-transferase POMGNT1 can transfer a β-2-linked GlcNAc onto $GlcNAc_2Man_5$ glycans, which was not expected since POMGNT1 normally transfers to O-mannose glycans (unpublished data). This suggests that POMGNT1 may play a role in the synthesis of these 'leaky' hybrid N-glycans, although it is unknown why the formation of hybrid N-glycans was limited to only a few of the 38 N-glycosites. Further experiments with MGAT1-POMGNT1 double-knockout cell lines are needed to confirm this finding.

While this system generated rich information, much could be done to improve experimentation. One idea is to dramatically increase the throughput of the process, as the current methodology is lengthy, high-cost, and often requires repetition due to inconsistencies in

sample processing. Since nearly all the power in our analyses rests in a single time point, and averaging over what is roughly the linear range of the curve produces similar results (data not shown), it would be reasonable to eliminate the timepoints in the experiment entirely and simply use a fixed reaction time. The current workflow can reasonably process 24 tubes at a time, and the current protocol uses every one (6 tubes, 4 replicates due to processing inconsistencies). Simplifying the workflow to a single time point would reduce risk and allow for the experiments to be carried out in triplicate instead. This, combined with the elimination of multiple time points, would effectively increasing throughput 8-fold. This may enable the screening of large selections of glycoprotein variants that could be used to study the importance of specific residues or even entire domains in defining N-glycan microheterogeneity.

Using mass error distributions to correct near-isobaric misassignments in glycoproteomics

Currently, manual validation of search engine results reveals that the rate of false positives generated by glycoproteomics search engines far exceeds their estimated false discovery rates (FDR) by as much as 10- to 20-fold. The experiments described in this dissertation necessitated extensive use of glycopeptide-specialized mass spectrometry search engine software paired with manual validation. Many of the observed false positives were due to mismatching of near-isobaric glycopeptides. As a result of this, we proposed a simple fix that substantially improves the accuracy of pGlyco3.0 (156), a popular glycopeptide-focused search engine.

This fix, called ppmFixer, is a python script that takes pGlyco output as input and adjusts glycan compositions for glycopeptide spectral matches (GPSMs) with mass errors that fall outside of a standard deviation of the average mass error for the dataset. If the adjusted mass

error is closer to the average mass error of the dataset, then the composition is adjusted in ppmFixer output. Applying these corrections dramatically improved the accuracy of pGlyco results, as much as halving the number of GPSM misidentifications. ppmFixer currently checks for four types of near-isobaric mismatches that reflect common errors made by the search engine. Due to the modular structure of ppmFixer, additional patterns of misidentification could easily be accounted for in the future.

Future Directions

The need for improvements in glycoproteomics mass spectrometry data analysis

High rates of misidentifications by search engines are currently a serious problem in glycoproteomics experiments. While manual validation of search engine results can alleviate much of this, the amount of expertise and time required for manual validation of entire datasets challenges its feasibility. As tandem LC-MS/MS and fragmentation technologies improve, more highly detailed spectra are being generated than ever before. A single digest of an isolated glycoprotein can result in thousands or even tens of thousands of spectral matches, each of which must be visually inspected by a specialized expert. As a result, analysts must make the hard decision of either largely accepting search engine outputs despite their known flaws or undergoing days (or even weeks) of manual validation for each experiment. In addition, it can hardly be expected that each analyst will find the same flaws in each dataset, which introduces additional variability and per-user bias.

To facilitate fast, accurate, and reproducible glycoproteomics experiments, further work must be done to bring the level of misidentifications of GPSMs down. Advances in tandem LC-MS/MS technology will contribute to this; as sensitivity and fragmentation techniques improve,

search engines can take advantage by lowering acceptable mass error tolerances to reduce the incidence of near-isobaric mismatches. This is especially important when glycoproteomics is applied to complex samples such as cellular lysates, where the combinatorial effect of large protein and glycan databases creates a massive set of theoretical glycopeptides that may be near enough to each other in mass to cause problems in identification. Finally, it is important that these advances are made using openly accessible and well-documented technology in order to facilitate the movement of mass spectrometry analysis away from mysterious black-box workflows and towards transparent, extendable software.

The role of structure modeling in studying N-glycan microheterogeneity

A key feature of the N-glycan microheterogeneity experiments in this dissertation is that the roster of reporter glycoproteins all had well-characterized crystal structures. This allowed us to be reasonably certain of the features of the protein surface surrounding sites of N-glycosylation. However, in most cases, they did not inform us on the conformation or flexibility of the glycan. Future work will need to use strategies such as molecular dynamics to understand how glycans can present themselves to glycosyltransferases. Some attempts towards this goal were made, but more could have been done to characterize protein-glycan interactions. One of these efforts was to model glycoproteins and MGAT1 utilizing an implicit solvent system with AMBER(230), but the overall system instability caused by superimposing the two crystal structures was insurmountable despite extensive efforts to resolve clashes and improve model stability.

Attempts at docking the glycoprotein CD16a and PDI to the glycosyltransferase MGAT1 using the popular modeling software RosettaDock (231) were unsuccessful, which was

unfortunate as a usable model could have illuminated how glycoproteins interact with glycosyltransferases. This failure was due to the glycan preferring the surface of the protein rather than the active site in all predicted conformations, even when initiated while superimposed in the binding pocket (unpublished data). RosettaDock uses a coarse-grained model, which uses an implicit solvent system and models at the resolution of individual residues (232), and perhaps this resolution struggles to effectively model glycan-glycosyltransferase interactions.

Future efforts may use a combination of the two approaches detailed above, where an all-atom model of the glycoprotein with its glycans is modeled in a generalized Born implicit solvent system(233). Perhaps by removing explicit water molecules and replacing them with a more generalized force field, greater system stability can be achieved.

Towards glycoengineering of biologics and biosimilars

While structure is a key factor in determining N-glycan microheterogeneity, the cell type defines the enzymatic repertoire that the glycoprotein is exposed to. Arigoni-Affolter et al. reconstructed N-glycan processing kinetics of the secretory pathway of CHO cells (105). Losfeld et al. recently extended this approach by incorporating different pathways through the secretory pathway that generate populations hosting distinct glycoforms(234). This information, combined with glycosyltransferase and glycosyl hydrolase expression data, are all necessary for characterizing N-glycan processing for a cell type under a given set of conditions. Additionally, information from knockout cell lines will continue to provide us with useful information on the role of individual or groups of glycan-processing enzymes (235). This could be utilized to generate cell lines that have been engineered to direct the flux of glycan-processing towards a desired glycan profile(204).

The studies described in this dissertation principally dealt with characterizing the microheterogeneity of a roster of N-glycoproteins. However, the larger goal of this project has been to develop a foundation for understanding site-specific N-glycan processing to facilitate control over the glycosylation of expressed therapeutics, such as biologics and biosimilars. Though our understanding of this process is still rudimentary, the potential in utilizing it presents a unique opportunity to enhance the benefits of existing therapies, whether that is through improving binding to receptors (236), increasing the half-life of therapeutics in serum (236), or perhaps even the development of competitive inhibitors of glycoprotein-virus interactions (148). However, to achieve these goals we will need extensive biochemical characterization of a large swathe of glycoproteins and glycan-processing enzymes to build upon the foundation of this work.

REFERENCES

1. Stanley, P., Moremen, K. W., Lewis, N. E., Taniguchi, N., and Aebi, M. (2022) in *Essentials of Glycobiology*, eds Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Mohnen D, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, Seeberger PH (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY)). 4th Ed.

2. del Val, I. J., Polizzi, K. M., and Kontoravdi, C. (2016) A theoretical estimate for nucleotide sugar demand towards Chinese Hamster Ovary cellular glycosylation. *Sci Rep* 6, 28547

3. Gagneux, P., Panin, V., Hennet, T., Aebi, M., and Varki, A. (2022) in *Essentials of Glycobiology*, eds Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Mohnen D, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, Seeberger PH (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY)). 4th Ed.

4. Zhang, P., Woen, S., Wang, T., Liau, B., Zhao, S., Chen, C., Yang, Y., Song, Z., Wormald, M. R., Yu, C., and Rudd, P. M. (2016) Challenges of glycosylation analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs. *Drug Discovery Today* 21, 740–765

5. Hart, G. W., Brew, K., Grant, G. A., Bradshaw, R. A., and Lennarz, W. J. (1979) Primary structural requirements for the enzymatic formation of the N-glycosidic bond in glycoproteins. Studies with natural and synthetic peptides. *J Biol Chem* 254, 9747–9753

6. Zielinska, D. F., Gnad, F., Wiśniewski, J. R., and Mann, M. (2010) Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* 141, 897–907

7. Ramírez, A. S., Kowal, J., and Locher, K. P. (2019) Cryo–electron microscopy structures of human oligosaccharyltransferase complexes OST-A and OST-B. *Science* 366, 1372–1375

8. Hubbard, S. C., and Robbins, P. W. (1979) Synthesis and processing of protein-linked oligosaccharides in vivo. *Journal of Biological Chemistry* 254, 4568–4576

9. Harpaz, N., and Schachter, H. (1980) Control of glycoprotein synthesis. Bovine colostrum UDP-N-acetylglucosamine:alpha-D-mannoside beta 2-N-acetylglucosaminyltransferase I. Separation from UDP-N-acetylglucosamine:alpha-D-mannoside beta 2-N-acetylglucosaminyltransferase II, partial purification, and substrate specificity. *Journal of Biological Chemistry* 255, 4885–4893

10. Ikeda, Y., Ihara, H., Tsukamoto, H., Gu, J., and Taniguchi, N. (2014) in *Handbook of Glycosyltransferases and Related Genes*, eds Taniguchi N, Honke K, Fukuda M, Narimatsu H, Yamaguchi Y, Angata T (Springer Japan, Tokyo), pp 209–222.

11. Longmore, G. D., and Schachter, H. (1982) Product-identification and substrate-specificity studies of the GDP-l-fucose: 2-acetamido-2-deoxy-β-d-glucoside (fuc→asn-linked GlcNAc) 6-α-l-fucosyltransferase in a golgi-rich fraction from porcine liver. *Carbohydrate Research* 100, 365–392

12. García-García, A., Serna, S., Yang, Z., Delso, I., Taleb, V., Hicks, T., Artschwager, R., Vakhrushev, S. Y., Clausen, H., Angulo, J., Corzana, F., Reichardt, N. C., and Hurtado-Guerrero, R. (2021) FUT8-Directed Core Fucosylation of N-glycans Is Regulated by the Glycan Structure and Protein Environment. *ACS Catal.*, 9052–9065

13. Shapira, R., and Parker, S. (1960) Artificially induced microheterogeneity in ribonuclease. *Biochemical and Biophysical Research Communications* 3, 200–204

14. Plummer, T. H., and Hirs, C. H. (1963) The isolation of ribounclease B, a glycoprotein, from bovine pancreatic juice. *J Biol Chem* 238, 1396–1401

15. Oshiro, Y., and Eylar, E. H. (1968) Physical and chemical studies on glycoproteins. *Archives of Biochemistry and Biophysics* 127, 476–489

16. Spiro, R. G. (1973) Glycoproteins. *Adv Protein Chem* 27, 349–467

17. Cunningham, L. W., Wilburn Clouse, R., and Ford, J. D. (1963) Heterogeneity of the carbohydrate moiety of crystalline ovalbumin. *Biochimica et Biophysica Acta* 78, 379–381

18. Cunningham, L., Ford, J. D., and Rainey, J. M. (1965) Heterogeneity of beta-aspartyl-oligosaccharides derived from ovalbumin. *Biochim Biophys Acta* 101, 233–235

19. Spiro, R. G. (1960) Studies on Fetuin, a Glycoprotein of Fetal Serum. *Journal of Biological Chemistry* 235, 2860–2869

20. Spiro, R. G. (1965) The Carbohydrate Units of Thyroglobulin. *Journal of Biological Chemistry* 240, 1603–1610

21. Spiro, R. G., and Spiro, M. J. (1965) The Carbohydrate Composition of the Thyroglobulins from Several Species. *Journal of Biological Chemistry* 240, 997–1001

22. Cunningham, L. W., and Ford, J. D. (1968) A Comparison of Glycopeptides Derived from Soluble and Insoluble Collagens. *Journal of Biological Chemistry* 243, 2390–2398

23. Huang, C.-C., Mayer, H. E., and Montgomery, R. (1970) Microheterogeneity and paucidispersity of glycoproteins. *Carbohydrate Research* 13, 127–137

24. Hickman, S., Kornfeld, R., Osterland, C. K., and Kornfeld, S. (1972) The Structure of the Glycopeptides of a Human γM-Immunoglobulin. *Journal of Biological Chemistry* 247, 2156–2163

25. Jackson, R. L., and Hirs, C. H. W. (1970) The Primary Structure of Porcine Pancreatic Ribonuclease. *Journal of Biological Chemistry* 245, 624–636

26. Esser, K., and Minuth, W. (1971) The Phenoloxidases of the Ascomycete Podospora anserina. Microheterogeneity of Laccase II. *Eur J Biochem* 23, 484–488

27. Yang, H. J., and Przybylska, M. (1973) The Microheterogeneity of Human Haptoglobin and Its Complex with Hemoglobin. *Can. J. Biochem.* 51, 597–605

28. Hayase, K., Reisher, S. R., and Kritchevsky, D. (1973) Microheterogeneity of N-Acetyl- -D-hexosaminidase of Bull Epididymis. *Experimental Biology and Medicine* 142, 466–470

29. Hiwada, K., and Wachsmuth, E. D. (1974) Alkaline phosphatase from pig kidney. Microheterogeneity and the role of neuraminic acid. *Biochemical Journal* 141, 293–298

30. Baenziger, J., Kornfeld, S., and Kochwa, S. (1974) Structure of the Carbohydrate Units of IgE Immunoglobulin. *Journal of Biological Chemistry* 249, 1897–1903

31. Anderson, D. R., and Grimes, W. J. (1982) Heterogeneity of asparagine-linked oligosaccharides of five glycosylation sites on immunoglobulin M heavy chain from mineral oil plasmacytoma 104E. *J Biol Chem* 257, 14858–14864

32. Williams, D. B., and Lennarz, W. J. (1984) Control of asparagine-linked oligosaccharide chain processing: studies on bovine pancreatic ribonuclease B. An in vitro system for the processing of exogenous glycoproteins. *J Biol Chem* 259, 5105–5114

33. Gottschalk, A. (1969) Biosynthesis of Glycoproteins and its Relationship to Heterogeneity. *Nature* 222, 452–454

34. Marshall, R. D., and Neuberger, A. (1970) in *Advances in Carbohydrate Chemistry and Biochemistry* (Elsevier), pp 407–478.

35. Alpert, E., Drysdale, J. W., Isselbacher, K. J., and Schur, P. H. (1972) Human α-Fetoprotein. *Journal of Biological Chemistry* 247, 3792–3798

36. Gordon, A. H., and Dykes, P. J. (1972) α1-Acute-phase globulins of rats. Microheterogeneity after isoelectric focusing. *Biochemical Journal* 130, 95–101

37. Frénoy, J.-P., and Bourrillon, R. (1974) Studies on the structure of human α2-macroglobulin IV. Analysis of the microheterogeneity by isoelectric focusing. *Biochimica et Biophysica Acta (BBA) - Protein Structure* 371, 168–176

38. Gustine, D. L., and Zimmerman, E. F. (1973) Developmental changes in microheterogeneity of foetal plasma glycoproteins of mice. *Biochemical Journal* 132, 541–551

39. Zimmerman, E. F., and Madappally, M. M. (1973) Sialyltransferase: regulation of α-foetoprotein microheterogeneity during development (*Short Communication*). *Biochemical Journal* 134, 807–810

40. Kobata, A. (1979) Use of endo- and exoglycosidases for structural studies of glycoconjugates. *Analytical Biochemistry* 100, 1–14

41. Tai, T., Yamashita, K., Ogata-Arakawa, M., Koide, N., Muramatsu, T., Iwashita, S., Inoue, Y., and Kobata, A. (1975) Structural studies of two ovalbumin glycopeptides in relation to the endo-beta-N-acetylglucosaminidase specificity. *J Biol Chem* 250, 8569–8575

42. Tai, T., Yamashita, K., Ito, S., and Kobata, A. (1977) Structures of the carbohydrate moiety of ovalbumin glycopeptide III and the difference in specificity of endo-beta-N-acetylglucosaminidases CII and H. *J Biol Chem* 252, 6687–6694

43. Endo, Y., Yamashita, K., Tachibana, Y., Tojo, S., and Kobata, A. (1979) Structures of the asparagine-linked sugar chains of human chorionic gonadotropin. *J Biochem* 85, 669–679

44. Pollack, L. (1983) Correlation of glycosylation forms with position in amino acid sequence. *The Journal of Cell Biology* 97, 293–300

45. Gavel, Y., and Heijne, G. von (1990) Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng Des Sel* 3, 433–442

46. Hakomori, S. (1964) A RAPID PERMETHYLATION OF GLYCOLIPID, AND POLYSACCHARIDE CATALYZED BY METHYLSULFINYL CARBANION IN DIMETHYL SULFOXIDE. *J Biochem* 55, 205–208

47. Krusius, T., and Finne, J. (1978) Characterization of a Novel Sugar Sequence from Rat-Brain Glycoproteins Containing Fucose and Sialic Acid. *Eur J Biochem* 84, 395–403

48. Fukuda, M., Dell, A., and Fukuda, M. N. (1984) Structure of fetal lactosaminoglycan. The carbohydrate moiety of Band 3 isolated from human umbilical cord erythrocytes. *Journal of Biological Chemistry* 259, 4782–4791

49. Carr, S. A., Reinhold, V. N., Green, B. N., and Hass, J. R. (1985) Enhancement of structural information in FAB ionized carbohydrate samples by neutral gas collision. *Biol. Mass Spectrom.* 12, 288–295

50. Reddy, V. A., Johnson, R. S., Biemann, K., Williams, R. S., Ziegler, F. D., Trimble, R. B., and Maley, F. (1988) Characterization of the glycosylation sites in yeast external

invertase. I. N-linked oligosaccharide content of the individual sequons. *Journal of Biological Chemistry* 263, 6978–6985

51. Linscheid, M., D'Angona, J., Burlingame, A. L., Dell, A., and Ballou, C. E. (1981) Field desorption mass spectrometry of oligosaccharides. *Proc. Natl. Acad. Sci. U.S.A.* 78, 1471–1475

52. Ceccarini, C., Lorenzoni, P., and Atkinson, P. H. (1984) Determination of the structure of ovalbumin glycopeptide AC-B by 1H nuclear magnetic resonance spectroscopy at 500 MHz. *J Mol Biol* 176, 161–167

53. Etchison, J. R., and Holland, J. J. (1974) Carbohydrate composition of the membrane glycoprotein of vesicular stomatitis virus grown in four mammalian cell lines. *Proc Natl Acad Sci U S A* 71, 4011–4014

54. Neil Barclay, A., Letarte-Muirhead, M., Williams, A. F., and Faulkes, R. A. (1976) Chemical characterisation of the Thy-1 glycoproteins from the membranes of rat thymocytes and brain. *Nature* 263, 563–567

55. Barclay, A. N., and Ward, H. A. (1982) Purification and chemical characterisation of membrane glycoproteins from rat thymocytes and brain, recognised by monoclonal antibody MRC OX 2. *Eur J Biochem* 129, 447–458

56. Fulcher, I. S., Chaplin, M. F., and Kenny, A. J. (1983) Endopeptidase-24.11 purified from pig intestine is differently glycosylated from that in kidney. *Biochemical Journal* 215, 317–323

57. Parekh, R., Roitt, I., Isenberg, D., Dwek, R., and Rademacher, T. (1988) Age-related galactosylation of the N-linked oligosaccharides of human serum IgG. *Journal of Experimental Medicine* 167, 1731–1736

58. Yamada, E., Tsukamoto, Y., Sasaki, R., Yagyu, K., and Takahashi, N. (1997) Structural changes of immunoglobulin G oligosaccharides with age in healthy human serum. *Glycoconj J* 14, 401–405

59. Shikata, K., Yasuda, T., Takeuchi, F., Konishi, T., Nakata, M., and Mizuochi, T. (1998) Structural changes in the oligosaccharide moiety of human IgG with aging. *Glycoconj J* 15, 683–689

60. Yamashita, K., Hitoi, A., Taniguchi, N., Yokosawa, N., Tsukada, Y., and Kobata, A. (1983) Comparative study of the sugar chains of gamma-glutamyltranspeptidases purified from rat liver and rat AH-66 hepatoma cells. *Cancer Res* 43, 5059–5063

61. Parekh, R. B., Tse, A. G., Dwek, R. A., Williams, A. F., and Rademacher, T. W. (1987) Tissue-specific N-glycosylation, site-specific oligosaccharide patterns and lentil lectin recognition of rat Thy-1. *The EMBO Journal* 6, 1233–1244

62. Williams, A. F., Parekh, R. B., Wing, D. R., Willis, A. C., Barclay, A. N., Dalchau, R., Fabre, J. W., Dwek, R. A., and Rademacher, T. W. (1993) Comparative analysis of the N-glycans of rat, mouse and human Thy-1. Site-specific oligosaccharide patterns of neural Thy-1, a member of the immunoglobulin superfamily. *Glycobiology* 3, 339–348

63. Hakimi, J., and Atkinson, P. H. (1980) Growth-dependent alterations in oligomannosyl glycopeptides expressed in Sindbis virus glycoproteins. *Biochemistry* 19, 5619–5624

64. Lin, Y.-H., Franc, V., and Heck, A. J. R. (2018) Similar Albeit Not the Same: In-Depth Analysis of Proteoforms of Human Serum, Bovine Serum, and Recombinant Human Fetuin. *J. Proteome Res.* 17, 2861–2869

65. Swiedler, S. J., Freed, J. H., Tarentino, A. L., Plummer, T. H., and Hart, G. W. (1985) Oligosaccharide microheterogeneity of the murine major histocompatibility antigens. Reproducible site-specific patterns of sialylation and branching in asparagine-linked oligosaccharides. *J Biol Chem* 260, 4046–4054

66. Kobata, A., and Yamashita, K. (1984) The sugar chains of γ-glutamyl trans-peptidase. *Pure and Applied Chemistry* 56, 821–832

67. Yamashita, K., Hitoi, A., Tateishi, N., Higashi, T., Sakamoto, Y., and Kobata, A. (1985) The structures of the carbohydrate moieties of mouse kidney γ-glutamyltranspeptidase: Occurrence of X-antigenic determinants and bisecting N-acetylglucosamine residues. *Archives of Biochemistry and Biophysics* 240, 573–582

68. Yamashita, K., Hitoi, A., Matsuda, Y., Miura, T., Katunuma, N., and Kobata, A. (1986) Structures of sugar chains of human kidney gamma-glutamyltranspeptidase. *J Biochem* 99, 55–62

69. Hsieh, P., Rosner, M. R., and Robbins, P. W. (1983) Host-dependent variation of asparagine-linked oligosaccharides at individual glycosylation sites of Sindbis virus glycoproteins. *Journal of Biological Chemistry* 258, 2548–2554

70. Yu, W.-H., Zhao, P., Draghi, M., Arevalo, C., Karsten, C. B., Suscovich, T. J., Gunn, B., Streeck, H., Brass, A. L., Tiemeyer, M., Seaman, M., Mascola, J. R., Wells, L., Lauffenburger, D. A., and Alter, G. (2018) Exploiting glycan topography for computational design of Env glycoprotein antigenicity. *PLoS Comput Biol* 14, e1006093

71. Wilson, I. A., Skehel, J. J., and Wiley, D. C. (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* 289, 366–373

72. Rudd, P. M., and Dwek, R. A. (1997) Glycosylation: Heterogeneity and the 3D Structure of Proteins. *Critical Reviews in Biochemistry and Molecular Biology* 32, 1–100

73. Elliott, T., Willis, A., Cerundolo, V., and Townsend, A. (1995) Processing of major histocompatibility class I-restricted antigens in the endoplasmic reticulum. *Journal of Experimental Medicine* 181, 1481–1491

74. Ioffe, E., and Stanley, P. (1994) Mice lacking N-acetylglucosaminyltransferase I activity die at mid-gestation, revealing an essential role for complex or hybrid N-linked carbohydrates. *Proc Natl Acad Sci U S A* 91, 728–732

75. Iwakura, Y., and Nozaki, M. (1985) Effects of tunicamycin on preimplantation mouse embryos: Prevention of molecular differentiation during blastocyst formation. *Developmental Biology* 112, 135–144

76. Čaval, T., Heck, A. J. R., and Reiding, K. R. (2021) Meta-heterogeneity: Evaluating and Describing the Diversity in Glycosylation Between Sites on the Same Glycoprotein. *Molecular & Cellular Proteomics* 20, 100010

77. Nilsson, I. M., and von Heijne, G. (1993) Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *Journal of Biological Chemistry* 268, 5798–5801

78. Bulleid, N. J., Bassel-Duby, R. S., Freedman, R. B., Sambrook, J. F., and Gething, M. J. H. (1992) Cell-free synthesis of enzymically active tissue-type plasminogen activator. Protein folding determines the extent of *N* -linked glycosylation. *Biochemical Journal* 286, 275–280

79. Kasturi, L., Eshleman, J. R., Wunner, W. H., and Shakin-Eshleman, S. H. (1995) The hydroxy amino acid in an Asn-X-Ser/Thr sequon can influence N-linked core glycosylation efficiency and the level of expression of a cell surface glycoprotein. *J. Biol. Chem.* 270, 14756–14761

80. Petrescu, A.-J. (2003) Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* 14, 103–114

81. Bause, E. (1983) Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.* 209, 331–336

82. Livi, G. P., Lillquist, J. S., Miles, L. M., Ferrara, A., Sathe, G. M., Simon, P. L., Meyers, C. A., Gorman, J. A., and Young, P. R. (1991) Secretion of N-glycosylated interleukin-1 beta in Saccharomyces cerevisiae using a leader peptide from Candida albicans. Effect of N-linked glycosylation on biological activity. *J Biol Chem* 266, 15348–15355

83. Shakin-Eshleman, S. H., Spitalnik, S. L., and Kasturi, L. (1996) The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked core-glycosylation efficiency. *J. Biol. Chem.* 271, 6363–6366

84. Schachter, H. (1986) Biosynthetic controls that determine the branching and microheterogeneity of protein-bound oligosaccharides. *Biochem. Cell Biol.* 64, 163–181

85. Harpaz, N., and Schachter, H. (1980) Control of glycoprotein synthesis. Processing of asparagine-linked oligosaccharides by one or more rat liver Golgi alpha-D-mannosidases dependent on the prior action of UDP-N-acetylglucosamine: alpha-D-mannoside beta 2-N-acetylglucosaminyltransferase I. *Journal of Biological Chemistry* 255, 4894–4902

86. Gleeson, P. A., and Schachter, H. (1983) Control of glycoprotein synthesis. *J Biol Chem* 258, 6162–6173

87. Narasimhan, S., Freed, J. C., and Schachter, H. (1985) Control of glycoprotein synthesis. Bovine milk UDPgalactose:N-acetylglucosamine 4-.beta.-galactosyltransferase catalyzes the preferential transfer of galactose to the GlcNAc.beta.l1,2Man.alpha.1,3- branch of both bisected and nonbisected complex biantennary asparagine-linked oligosaccharides. *Biochemistry* 24, 1694–1700

88. Pâquet, M. R., Narasimhan, S., Schachter, H., and Moscarello, M. A. (1984) Branch specificity of purified rat liver Golgi UDP-galactose: N-acetylglucosamine beta-1,4-galactosyltransferase. Preferential transfer of of galactose on the GlcNAc beta 1,2-Man alpha 1,3-branch of a complex biantennary Asn-linked oligosaccharide. *J Biol Chem* 259, 4716–4721

89. Blanken, W. M., van Vliet, A., and van den Eijnden, D. H. (1984) Branch specificity of bovine colostrum and calf thymus UDP-Gal: N-acetylglucosaminide beta-1,4-galactosyltransferase. *Journal of Biological Chemistry* 259, 15131–15135

90. Joziasse, D. H., Schiphorst, W. E., van den Eijnden, D. H., van Kuik, J. A., van Halbeek, H., and Vliegenthart, J. F. (1985) Branch specificity of bovine colostrum CMP-sialic acid: N-acetyllactosaminide alpha 2—-6-sialyltransferase. Interaction with biantennary oligosaccharides and glycopeptides of N-glycosylproteins. *Journal of Biological Chemistry* 260, 714–719

91. Merkle, R. K., and Cummings, R. D. (1987) Relationship of the terminal sequences to the length of poly-N-acetyllactosamine chains in asparagine-linked oligosaccharides from the mouse lymphoma cell line BW5147. Immobilized tomato lectin interacts with high affinity with glycopeptides containing long poly-N-acetyllactosamine chains. *Journal of Biological Chemistry* 262, 8179–8189

92. Lennarz, W. J. ed. (1980) *The Biochemistry of Glycoproteins and Proteoglycans* (Springer US, Boston, MA)

93. Aubert, J.-P., Biserte, G., and Loucheux-Lefebvre, M.-H. (1976) Carbohydrate-peptide linkage in glycoproteins. *Archives of Biochemistry and Biophysics* 175, 410–418

94. Zielinska, D. F., Gnad, F., Schropp, K., Wiśniewski, J. R., and Mann, M. (2012) Mapping N-Glycosylation Sites across Seven Evolutionarily Distant Species Reveals a Divergent Substrate Proteome Despite a Common Core Machinery. *Molecular Cell* 46, 542–548

95. Imperiali, B., and Rickert, K. W. (1995) Conformational implications of asparagine-linked glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* 92, 97–101

96. Fischer, P. B., Karlsson, G. B., Butters, T. D., Dwek, R. A., and Platt, F. M. (1996) N-butyldeoxynojirimycin-mediated inhibition of human immunodeficiency virus entry correlates with changes in antibody recognition of the V1/V2 region of gp120. *J Virol* 70, 7143–7152

97. Trimble, R. B., Maley, F., and Chu, F. K. (1983) GlycoProtein biosynthesis in yeast. protein conformation affects processing of high mannose oligosaccharides on carboxypeptidase Y and invertase. *J Biol Chem* 258, 2562–2567

98. Savvidou, G., Klein, M., Grey, A. A., Dorrington, K. J., and Carver, J. P. (1984) Possible role for peptide-oligosaccharide interactions in differential oligosaccharide processing at asparagine-107 of the light chain and asparagine-297 of the heavy chain in a monoclonal IgG1 kappa. *Biochemistry* 23, 3736–3740

99. Deisenhofer, J. (1981) Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from Staphylococcus aureus at 2.9- and 2.8-A resolution. *Biochemistry* 20, 2361–2370

100. Brisson, J. R., and Carver, J. P. (1983) The relation of three-dimensional structure to biosynthesis in the N-linked oligosaccharides. *Canadian Journal of Biochemistry and Cell Biology = Revue Canadienne De Biochimie Et Biologie Cellulaire* 61, 1067–1078

101. Chen, W., Kong, L., Connelly, S., Dendle, J. M., Liu, Y., Wilson, I. A., Powers, E. T., and Kelly, J. W. (2016) Stabilizing the C $_{H}$ 2 Domain of an Antibody by Engineering in an Enhanced Aromatic Sequon. *ACS Chem. Biol.* 11, 1852–1861

102. Yu, X., Baruah, K., Harvey, D. J., Vasiljevic, S., Alonzi, D. S., Song, B.-D., Higgins, M. K., Bowden, T. A., Scanlan, C. N., and Crispin, M. (2013) Engineering hydrophobic protein-carbohydrate interactions to fine-tune monoclonal antibodies. *J Am Chem Soc* 135, 9723–9732

103. Lund, J., Takahashi, N., Pound, J. D., Goodall, M., and Jefferis, R. (1996) Multiple interactions of IgG with its core oligosaccharide can modulate recognition by complement and human Fc gamma receptor I and influence the synthesis of its oligosaccharide chains. *J Immunol* 157, 4963–4969

104. Hang, I., Lin, C., Grant, O. C., Fleurkens, S., Villiger, T. K., Soos, M., Morbidelli, M., Woods, R. J., Gauss, R., and Aebi, M. (2015) Analysis of site-specific *N* -glycan remodeling in the endoplasmic reticulum and the Golgi. *Glycobiology* 25, 1335–1349

105. Arigoni-Affolter, I., Scibona, E., Lin, C.-W., Brühlmann, D., Souquet, J., Broly, H., and Aebi, M. (2019) Mechanistic reconstruction of glycoprotein secretion through monitoring of intracellular N-glycan processing. *Science Advances* 5, eaax8930

106. Adams, T. M., Zhao, P., Chapla, D., Moremen, K. W., and Wells, L. (2022) Sequential in vitro enzymatic N-glycoprotein modification reveals site-specific rates of glycoenzyme processing. *J Biol Chem* 298, 102474

107. Dahms, N. M., and Hart, G. W. (1986) Influence of quaternary structure on glycosylation. Differential subunit association affects the site-specific glycosylation of the common beta-chain from Mac-1 and LFA-1. *Journal of Biological Chemistry* 261, 13186–13196

108. Hirschberg, C. B., Robbins, P. W., and Abeijon, C. (1998) Transporters of nucleotide sugars, ATP, and nucleotide sulfate in the endoplasmic reticulum and Golgi apparatus. *Annu Rev Biochem* 67, 49–69

109. Abeijon, C., Yanagisawa, K., Mandon, E. C., Hiiusler, A., Moremen, K., and Hirschberg, C. B. (1993) Guanosine Diphosphatase Is Required for Protein and Sphingolipid Gly¢osylation in the Golgi Lumen of Saccharomyces cer isiae. *The Journal of Cell Biology* 122,

110. Martina, J. A., and Maccioni, H. J. F. (1996) UDP-sugar pyrophosphatase of rat retina: Subcellular localization and topography. *J. Neurosci. Res.* 46, 485–491

111. Lepers, A., Shaw, L., Schneckenburger, P., Cacan, R., Verbert, A., and Schauer, R. (1990) A study on the regulation of N-glycoloylneuraminic acid biosynthesis and utilization in rat and mouse liver. *Eur J Biochem* 193, 715–723

112. Carson, D. D., Earles, B. J., and Lennarz, W. J. (1981) Enhancement of protein glycosylation in tissue slices by dolichylphosphate. *Journal of Biological Chemistry* 256, 11552–11557

113. Van Schaftingen, E., and Jaeken, J. (1995) Phosphomannomutase deficiency is a cause of carbohydrate-deficient glycoprotein syndrome type I. *FEBS Lett* 377, 318–320

114. Jaeken, J., Artigas, J., Barone, R., Fiumara, A., de Koning, T. J., Poll-The, B. T., de Rijk-van Andel, J. F., Hoffmann, G. F., Assmann, B., Mayatepek, E., Pineda, M., Vilaseca, M. A., Saudubray, J. M., Schlüter, B., Wevers, R., and Van Schaftingen, E. (1997) Phosphomannomutase deficiency is the main cause of carbohydrate-deficient glycoprotein syndrome with type I isoelectrofocusing pattern of serum sialotransferrins. *J Inherit Metab Dis* 20, 447–449

115. Yamashita, K., Hitoi, A., Irie, M., and Kobata, A. (1986) Fractionation by lectin affinity chromatography indicates that the glycosylation of most ribonucleases in human viscera and body fluids is organ specific. *Archives of Biochemistry and Biophysics* 250, 263–266

116. Huang, H.-H., and Stanley, P. (2010) A testis-specific regulator of complex and hybrid N-glycan synthesis. *Journal of Cell Biology* 190, 893–910

117. Stanley, P., Narasimhan, S., Siminovitch, L., and Schachter, H. (1975) Chinese hamster ovary cells selected for resistance to the cytotoxicity of phytohemagglutinin are deficient in a UDP-N-acetylglucosamine--glycoprotein N-acetylglucosaminyltransferase activity. *Proc. Natl. Acad. Sci. U.S.A.* 72, 3323–3327

118. Wang, W. C., Lee, N., Aoki, D., Fukuda, M. N., and Fukuda, M. (1991) The poly-N-acetyllactosamines attached to lysosomal membrane glycoproteins are increased by the prolonged association with the Golgi complex. *Journal of Biological Chemistry* 266, 23185–23190

119. Hanash, S. M., Pitteri, S. J., and Faca, V. M. (2008) Mining the plasma proteome for cancer biomarkers. *Nature* 452, 571–579

120. Kailemia, M. J., Park, D., and Lebrilla, C. B. (2017) Glycans and glycoproteins as specific biomarkers for cancer. *Anal Bioanal Chem* 409, 395–410

121. Weiss, G. L., Stanisich, J. J., Sauer, M. M., Lin, C.-W., Eras, J., Zyla, D. S., Trück, J., Devuyst, O., Aebi, M., Pilhofer, M., and Glockshuber, R. (2020) Architecture and function of human uromodulin filaments in urinary tract infections. *Science* 369, 1005–1010

122. van Zeben, D., Rook, G. A., Hazes, J. M., Zwinderman, A. H., Zhang, Y., Ghelani, S., Rademacher, T. W., and Breedveld, F. C. (1994) Early agalactosylation of IgG is associated with a more progressive disease course in patients with rheumatoid arthritis: results of a follow-up study. *Br J Rheumatol* 33, 36–43

123. Malhotra, R., Wormald, M. R., Rudd, P. M., Fischer, P. B., Dwek, R. A., and Sim, R. B. (1995) Glycosylation changes of IgG associated with rheumatooid arthritis can activate complement via the mannose-binding protein. *Nat Med* 1, 237–243

124. Lin, Y.-H., Zhu, J., Meijer, S., Franc, V., and Heck, A. J. R. (2019) Glycoproteogenomics: A Frequent Gene Polymorphism Affects the Glycosylation Pattern of the Human Serum Fetuin/α-2-HS-Glycoprotein. *Molecular & Cellular Proteomics* 18, 1479–1490

125. Woods, R. J. (2018) Predicting the Structures of Glycans, Glycoproteins, and Their Complexes. *Chemical Reviews* 118, 8005–8024

126. Sinitskiy, A. V., and Pande, V. S. (2017) Simulated Dynamics of Glycans on Ligand-Binding Domain of NMDA Receptors Reveals Strong Dynamic Coupling between Glycans and Protein Core. *J. Chem. Theory Comput.* 13, 5496–5505

127. Subedi, G. P., Sinitskiy, A. V., Roberts, J. T., Patel, K. R., Pande, V. S., and Barb, A. W. (2019) Intradomain Interactions in an NMDA Receptor Fragment Mediate N-Glycan Processing and Conformational Sampling. *Structure* 27, 55-65.e3

128. Sudol, A. S. L., Butler, J., Ivory, D. P., Tews, I., and Crispin, M. (2022) *Extensive substrate recognition by the streptococcal antibody-degrading enzymes IdeS and EndoS* (Immunology)

129. Boersema, P. J., Geiger, T., Wisniewski, J. R., and Mann, M. (2013) Quantification of the N-glycosylated secretome by super-SILAC during breast cancer progression and in human blood samples. *Mol Cell Proteomics* 12, 158–171

130. Cao, L., Diedrich, J. K., Kulp, D. W., Pauthner, M., He, L., Park, S.-K. R., Sok, D., Su, C. Y., Delahunty, C. M., Menis, S., Andrabi, R., Guenaga, J., Georgeson, E., Kubitz, M., Adachi, Y., Burton, D. R., Schief, W. R., Yates III, J. R., and Paulson, J. C. (2017) Global site-specific N-glycosylation analysis of HIV envelope glycoprotein. *Nature Communications* 8, 14954

131. Berndsen, Z. T., Chakraborty, S., Wang, X., Cottrell, C. A., Torres, J. L., Diedrich, J. K., López, C. A., Yates, J. R., van Gils, M. J., Paulson, J. C., Gnanakaran, S., and Ward, A. B. (2020) Visualization of the HIV-1 Env glycan shield across scales. *Proc Natl Acad Sci U S A* 117, 28014–28025

132. Klaver, E., Zhao, P., May, M., Flanagan-Steet, H., Freeze, H. H., Gilmore, R., Wells, L., Contessa, J., and Steet, R. (2019) Selective inhibition of N-linked glycosylation impairs receptor tyrosine kinase processing. *Dis Model Mech* 12, dmm039602

133. Wu, G., Grassi, P., MacIntyre, D. A., Molina, B. G., Sykes, L., Kundu, S., Hsiao, C.-T., Khoo, K.-H., Bennett, P. R., Dell, A., and Haslam, S. M. (2022) N-glycosylation of cervicovaginal fluid reflects microbial community, immune activity, and pregnancy status. *Sci Rep* 12, 16948

134. Zelanis, A., Serrano, S. M. T., and Reinhold, V. N. (2012) N-glycome profiling of Bothrops jararaca newborn and adult venoms. *J Proteomics* 75, 774–782

135. Cipollo, J. F., Awad, A. M., Costello, C. E., and Hirschberg, C. B. (2005) N-Glycans of Caenorhabditis elegans are specific to developmental stages. *J Biol Chem* 280, 26063–26072

136. Raghunathan, R., Polinski, N. K., Klein, J. A., Hogan, J. D., Shao, C., Khatri, K., Leon, D., McComb, M. E., Manfredsson, F. P., Sortwell, C. E., and Zaia, J. (2018) Glycomic and Proteomic Changes in Aging Brain Nigrostriatal Pathway. *Mol Cell Proteomics* 17, 1778–1787

137. North, S. J., von Gunten, S., Antonopoulos, A., Trollope, A., MacGlashan, D. W., Jang-Lee, J., Dell, A., Metcalfe, D. D., Kirshenbaum, A. S., Bochner, B. S., and Haslam, S. M. (2012) Glycomic analysis of human mast cells, eosinophils and basophils. *Glycobiology* 22, 12–22

138. Chen, Q., Müller, J. S., Pang, P.-C., Laval, S. H., Haslam, S. M., Lochmüller, H., and Dell, A. (2015) Global N-linked Glycosylation is Not Significantly Impaired in Myoblasts in Congenital Myasthenic Syndromes Caused by Defective Glutamine-Fructose-6-Phosphate Transaminase 1 (GFPT1). *Biomolecules* 5, 2758–2781

139. Håkansson, K., Cooper, H. J., Emmett, M. R., Costello, C. E., Marshall, A. G., and Nilsson, C. L. (2001) Electron capture dissociation and infrared multiphoton dissociation MS/MS of an N-glycosylated tryptic peptic to yield complementary sequence information. *Anal Chem* 73, 4530–4536

140. Mikesh, L. M., Ueberheide, B., Chi, A., Coon, J. J., Syka, J. E. P., Shabanowitz, J., and Hunt, D. F. (2006) The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta* 1764, 1811–1822

141. Mirgorodskaya, E., Roepstorff, P., and Zubarev, R. A. (1999) Localization of O-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer. *Anal Chem* 71, 4431–4436

142. Scott, N. E., Parker, B. L., Connolly, A. M., Paulech, J., Edwards, A. V. G., Crossett, B., Falconer, L., Kolarich, D., Djordjevic, S. P., Højrup, P., Packer, N. H., Larsen, M. R., and Cordwell, S. J. (2011) Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of Campylobacter jejuni. *Mol Cell Proteomics* 10, M000031-MCP201

143. Hogan, J. M., Pitteri, S. J., Chrisman, P. A., and McLuckey, S. A. (2005) Complementary structural information from a tryptic N-linked glycopeptide via electron transfer ion/ion reactions and collision-induced dissociation. *J Proteome Res* 4, 628–632

144. Wu, S.-L., Hühmer, A. F. R., Hao, Z., and Karger, B. L. (2007) On-line LC-MS approach combining collision-induced dissociation (CID), electron-transfer dissociation (ETD), and CID of an isolated charge-reduced species for the trace-level characterization of proteins with post-translational modifications. *J Proteome Res* 6, 4230–4244

145. Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007) Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* 4, 709–712

146. Zhao, P., Viner, R., Teo, C. F., Boons, G.-J., Horn, D., and Wells, L. (2011) Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. *J. Proteome Res.* 10, 4088–4104

147. Cao, L., Tolić, N., Qu, Y., Meng, D., Zhao, R., Zhang, Q., Moore, R. J., Zink, E. M., Lipton, M. S., Paša-Tolić, L., and Wu, S. (2014) Characterization of intact N- and O-linked glycopeptides using higher energy collisional dissociation. *Anal Biochem* 452, 96–102

148. Zhao, P., Praissman, J. L., Grant, O. C., Cai, Y., Xiao, T., Rosenbalm, K. E., Aoki, K., Kellman, B. P., Bridger, R., Barouch, D. H., Brindley, M. A., Lewis, N. E., Tiemeyer, M., Chen, B., Woods, R. J., and Wells, L. (2020) Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host & Microbe* 28, 586-601.e6

149. Riley, N. M., Malaker, S. A., Driessen, M. D., and Bertozzi, C. R. (2020) Optimal Dissociation Methods Differ for N- and O-Glycopeptides. *J Proteome Res* 19, 3286–3301

150. Liu, D., Wang, S., Zhang, J., Xiao, W., Miao, C. H., Konkle, B. A., Wan, X.-F., and Li, L. (2021) Site-Specific N- and O-Glycosylation Analysis of Human Plasma Fibronectin. *Front Chem* 9, 691217

151. Schwaiger-Haber, M., Stancliffe, E., Arends, V., Thyagarajan, B., Sindelar, M., and Patti, G. J. (2021) A Workflow to Perform Targeted Metabolomics at the Untargeted Scale on a Triple Quadrupole Mass Spectrometer. *ACS Meas Sci Au* 1, 35–45

152. Bern, M., Kil, Y. J., and Becker, C. (2012) Byonic: advanced peptide and protein identification software. *Curr Protoc Bioinformatics* Chapter 13, 13.20.1-13.20.14

153. Stadlmann, J., Hoi, D. M., Taubenschmid, J., Mechtler, K., and Penninger, J. M. (2018) Analysis of PNGase F-Resistant N-Glycopeptides Using SugarQb for Proteome Discoverer 2.1 Reveals Cryptic Substrate Specificities. *Proteomics* 18, e1700436

154. Yang, G., Hu, Y., Sun, S., Ouyang, C., Yang, W., Wang, Q., Betenbaugh, M., and Zhang, H. (2018) Comprehensive Glycoproteomic Analysis of Chinese Hamster Ovary Cells. *Anal Chem* 90, 14294–14302

155. Mayampurath, A., Song, E., Mathur, A., Yu, C.-Y., Hammoud, Z., Mechref, Y., and Tang, H. (2014) Label-free glycopeptide quantification for biomarker discovery in human sera. *J Proteome Res* 13, 4821–4832

156. Zeng, W.-F., Cao, W.-Q., Liu, M.-Q., He, S.-M., and Yang, P.-Y. (2021) Precise, fast and comprehensive analysis of intact glycopeptides and modified glycans with pGlyco3. *Nat Methods* 18, 1515–1523

157. Xu, F., Wang, Y., Xiao, K., Hu, Y., Tian, Z., and Chen, Y. (2020) Quantitative site- and structure-specific N-glycoproteomics characterization of differential N-glycosylation in MCF-7/ADR cancer stem cells. *Clin Proteomics* 17, 3

158. Polasky, D. A., Yu, F., Teo, G. C., and Nesvizhskii, A. I. (2020) Fast and comprehensive N- and O-glycoproteomics analysis with MSFragger-Glyco. *Nat Methods* 17, 1125–1132

159. Lu, L., Riley, N. M., Shortreed, M. R., Bertozzi, C. R., and Smith, L. M. (2020) O-Pair Search with MetaMorpheus for O-glycopeptide characterization. *Nat Methods* 17, 1133–1138

160. Praissman, J. L., and Wells, L. (2020) Getting more for less: new software solutions for glycoproteomics. *Nat Methods* 17, 1081–1082

161. Elias, J. E., and Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 4, 207–214

162. Liu, M.-Q., Zeng, W.-F., Fang, P., Cao, W.-Q., Liu, C., Yan, G.-Q., Zhang, Y., Peng, C., Wu, J.-Q., Zhang, X.-J., Tu, H.-J., Chi, H., Sun, R.-X., Cao, Y., Dong, M.-Q., Jiang, B.-Y., Huang, J.-M., Shen, H.-L., Wong, C. C. L., He, S.-M., and Yang, P.-Y. (2017) pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. *Nature Communications* 8,

163. Polasky, D. A., Geiszler, D. J., Yu, F., and Nesvizhskii, A. I. (2022) Multiattribute Glycan Identification and FDR Control for Glycoproteomics. *Mol Cell Proteomics* 21, 100205

164. Kawahara, R., Chernykh, A., Alagesan, K., Bern, M., Cao, W., Chalkley, R. J., Cheng, K., Choo, M. S., Edwards, N., Goldman, R., Hoffmann, M., Hu, Y., Huang, Y., Kim, J. Y., Kletter, D., Liquet, B., Liu, M., Mechref, Y., Meng, B., Neelamegham, S., Nguyen-Khuong, T., Nilsson, J., Pap, A., Park, G. W., Parker, B. L., Pegg, C. L., Penninger, J. M., Phung, T. K., Pioch, M., Rapp, E., Sakalli, E., Sanda, M., Schulz, B. L., Scott, N. E., Sofronov, G., Stadlmann, J., Vakhrushev, S. Y., Woo, C. M., Wu, H.-Y., Yang, P., Ying,

W., Zhang, H., Zhang, Y., Zhao, J., Zaia, J., Haslam, S. M., Palmisano, G., Yoo, J. S., Larson, G., Khoo, K.-H., Medzihradszky, K. F., Kolarich, D., Packer, N. H., and Thaysen-Andersen, M. (2021) Community evaluation of glycoproteomics informatics solutions reveals high-performance search strategies for serum glycopeptide analysis. *Nat Methods* 18, 1304–1316

165. Sethi, M. K., Downs, M., Shao, C., Hackett, W. E., Phillips, J. J., and Zaia, J. (2022) In-Depth Matrisome and Glycoproteomic Analysis of Human Brain Glioblastoma Versus Control Tissue. *Mol Cell Proteomics* 21, 100216

166. Juge, N., Tailford, L., and Owen, C. D. (2016) Sialidases from gut bacteria: a mini-review. *Biochemical Society Transactions* 44, 166–175

167. McArthur, J. B., Yu, H., Tasnima, N., Lee, C. M., Fisher, A. J., and Chen, X. (2018) α2–6-Neosialidase: A Sialyltransferase Mutant as a Sialyl Linkage-Specific Sialidase. *ACS Chem. Biol.* 13, 1228–1234

168. Cao, L., Diedrich, J. K., Ma, Y., Wang, N., Pauthner, M., Park, S.-K. R., Delahunty, C. M., McLellan, J. S., Burton, D. R., Yates, J. R., and Paulson, J. C. (2018) Global site-specific analysis of glycoprotein N-glycan processing. *Nat Protoc* 13, 1196–1212

169. Wu, D., and Robinson, C. V. (2022) Understanding glycoprotein structural heterogeneity and interactions: Insights from native mass spectrometry. *Current Opinion in Structural Biology* 74, 102351

170. Wu, D., Li, J., Struwe, W. B., and Robinson, C. V. (2019) Probing *N* -glycoprotein microheterogeneity by lectin affinity purification-mass spectrometry analysis. *Chemical Science* 10, 5146–5155

171. Baković, M. P., Selman, M. H. J., Hoffmann, M., Rudan, I., Campbell, H., Deelder, A. M., Lauc, G., and Wuhrer, M. (2013) High-Throughput IgG Fc N-Glycosylation Profiling by Mass Spectrometry of Glycopeptides. *J. Proteome Res.* 12, 821–831

172. Varki, A. (2017) Biological roles of glycans. *Glycobiology* 27, 3–49

173. Moremen, K. W., Tiemeyer, M., and Nairn, A. V. (2012) Vertebrate protein glycosylation: diversity, synthesis and function. *Nature Reviews Molecular Cell Biology* 13, 448–462

174. Stanley, P., Taniguchi, N., and Aebi, M. (2015) in *Essentials of Glycobiology, 3rd Edition* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY)), p Chapter 9.

175. Solá, R. J., and Griebenow, K. (2009) Effects of glycosylation on the stability of protein pharmaceuticals. *Journal of Pharmaceutical Sciences* 98, 1223–1245

176. Wild, R., Kowal, J., Eyring, J., Ngwa, E. M., Aebi, M., and Locher, K. P. (2018) Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science* 359, 545–550

177. Rudd, P. M., and Dwek, R. A. (1997) Rapid, sensitive sequencing of oligosaccharides from glycoproteins. *Current Opinion in Biotechnology* 8, 488–497

178. Losfeld, M.-E., Scibona, E., Lin, C.-W., Villiger, T. K., Gauss, R., Morbidelli, M., and Aebi, M. (2017) Influence of protein/glycan interaction on site-specific glycan heterogeneity. *The FASEB Journal* 31, 4623–4635

179. Narasimhan, S., Stanley, P., and Schachter, H. (1977) Control of glycoprotein synthesis. Lectin-resistant mutant containing only one of two distinct N-acetylglucosaminyltransferase activities present in wild type Chinese hamster ovary cells. *Journal of Biological Chemistry* 252, 3926–3933

180. Arima, T., Spiro, M. J., and Spiro, R. G. (1972) Studies on the Carbohydrate Units of Thyroglobulin. *Journal of Biological Chemistry* 247, 1825–1835

181. Takasaki, S., Mizuochi, T., and Kobata, A. (1982) in *Methods in Enzymology* (Elsevier), pp 263–268.

182. Gordon, R. D., Sivarajah, P., Satkunarajah, M., Ma, D., Tarling, C. A., Vizitiu, D., Withers, S. G., and Rini, J. M. (2006) X-ray Crystal Structures of Rabbit N-acetylglucosaminyltransferase I (GnT I) in Complex with Donor Substrate Analogues. *Journal of Molecular Biology* 360, 67–79

183. Cymer, F., Beck, H., Rohde, A., and Reusch, D. (2018) Therapeutic monoclonal antibody N-glycosylation – Structure, function and therapeutic potential. *Biologicals* 52, 1–11

184. Subedi, G. P., and Barb, A. W. (2015) The Structural Role of Antibody N-Glycosylation in Receptor Interactions. *Structure* 23, 1573–1583

185. Li, T., DiLillo, D. J., Bournazos, S., Giddens, J. P., Ravetch, J. V., and Wang, L.-X. (2017) Modulating IgG effector function by Fc glycan engineering. *Proc. Natl. Acad. Sci. U.S.A.* 114, 3485–3490

186. Hayes, J. M., Frostell, A., Karlsson, R., Müller, S., Martín, S. M., Pauers, M., Reuss, F., Cosgrave, E. F., Anneren, C., Davey, G. P., and Rudd, P. M. (2017) Identification of Fc Gamma Receptor Glycoforms That Produce Differential Binding Kinetics for Rituximab. *Molecular & Cellular Proteomics* 16, 1770–1788

187. Patel, K. R., Roberts, J. T., Subedi, G. P., and Barb, A. W. (2018) Restricted processing of CD16a/Fc γ receptor IIIa N-glycans from primary human NK cells impacts structure and function. *Journal of Biological Chemistry* 293, 3477–3489

188. Higel, F., Seidl, A., Sörgel, F., and Friess, W. (2016) N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. *European Journal of Pharmaceutics and Biopharmaceutics* 100, 94–100

189. Azevedo, V., Hassett, B., Fonseca, J. E., Atsumi, T., Coindreau, J., Jacobs, I., Mahgoub, E., O'Brien, J., Singh, E., Vicik, S., and Fitzpatrick, B. (2016) Differentiating biosimilarity and comparability in biotherapeutics. *Clin Rheumatol* 35, 2877–2886

190. Zhou, T., Doria-Rose, N. A., Cheng, C., Stewart-Jones, G. B. E., Chuang, G.-Y., Chambers, M., Druz, A., Geng, H., McKee, K., Kwon, Y. D., O'Dell, S., Sastry, M., Schmidt, S. D., Xu, K., Chen, L., Chen, R. E., Louder, M. K., Pancera, M., Wanninger, T. G., Zhang, B., Zheng, A., Farney, S. K., Foulds, K. E., Georgiev, I. S., Joyce, M. G., Lemmin, T., Narpala, S., Rawi, R., Soto, C., Todd, J.-P., Shen, C.-H., Tsybovsky, Y., Yang, Y., Zhao, P., Haynes, B. F., Stamatatos, L., Tiemeyer, M., Wells, L., Scorpio, D. G., Shapiro, L., McDermott, A. B., Mascola, J. R., and Kwong, P. D. (2017) Quantification of the Impact of the HIV-1-Glycan Shield on Antibody Elicitation. *Cell Reports* 19, 719–732

191. Broszeit, F., van Beek, R. J., Unione, L., Bestebroer, T. M., Chapla, D., Yang, J.-Y., Moremen, K. W., Herfst, S., Fouchier, R. A. M., de Vries, R. P., and Boons, G.-J. (2021) Glycan remodeled erythrocytes facilitate antigenic characterization of recent A/H3N2 influenza viruses. *Nat Commun* 12, 5449

192. Faye, L., Sturm, A., Bollini, R., Vitale, A., and Chrispeels, M. J. (1986) The position of the oligosaccharide side-chains of phytohemagglutinin and their accessibility to glycosidases determines their subsequent processing in the Golgi. *Eur J Biochem* 158, 655–661

193. Watanabe, Y., Bowden, T. A., Wilson, I. A., and Crispin, M. (2019) Exploitation of glycosylation in enveloped virus pathobiology. *Biochim Biophys Acta Gen Subj* 1863, 1480–1497

194. Patel, K. R., Nott, J. D., and Barb, A. W. (2019) Primary Human Natural Killer Cells Retain Proinflammatory IgG1 at the Cell Surface and Express CD16a Glycoforms with Donor-dependent Variability. *Molecular & Cellular Proteomics* 18, 2178–2190

195. Weiß, R. G., Losfeld, M.-E., Aebi, M., and Riniker, S. (2021) N-Glycosylation Enhances Conformational Flexibility of Protein Disulfide Isomerase Revealed by Microsecond Molecular Dynamics and Markov State Modeling. *J. Phys. Chem. B* 125, 9467–9479

196. Thaysen-Andersen, M., and Packer, N. H. (2012) Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. *Glycobiology* 22, 1440–1452

197. Mathew, C., Weiß, R. G., Giese, C., Lin, C., Losfeld, M.-E., Glockshuber, R., Riniker, S., and Aebi, M. (2021) Glycan–protein interactions determine kinetics of *N* -glycan remodeling. *RSC Chem. Biol.* 2, 917–931

198. Patel, K. R., Rodriguez Benavente, M. C., Lorenz, W. W., Mace, E. M., and Barb, A. W. (2020) Fc γ receptor IIIa / CD16a processing correlates with the expression of glycan-related genes in human natural killer cells. *J. Biol. Chem.*, jbc.RA120.015516

199. Liu, L., Gomathinayagam, S., Hamuro, L., Prueksaritanont, T., Wang, W., Stadheim, T. A., and Hamilton, S. R. (2013) The Impact of Glycosylation on the Pharmacokinetics of a TNFR2:Fc Fusion Protein Expressed in Glycoengineered Pichia Pastoris. *Pharm Res* 30, 803–812

200. Čaval, T., Tian, W., Yang, Z., Clausen, H., and Heck, A. J. R. (2018) Direct quality control of glycoengineered erythropoietin variants. *Nature Communications* 9,

201. Sinclair, A. M., and Elliott, S. (2005) Glycoengineering: The effect of glycosylation on the properties of therapeutic proteins. *Journal of Pharmaceutical Sciences* 94, 1626–1635

202. Banks, D. D. (2011) The Effect of Glycosylation on the Folding Kinetics of Erythropoietin. *Journal of Molecular Biology* 412, 536–550

203. Sztain, T., Ahn, S.-H., Bogetti, A. T., Casalino, L., Goldsmith, J. A., Seitz, E., McCool, R. S., Kearns, F. L., Acosta-Reyes, F., Maji, S., Mashayekhi, G., McCammon, J. A., Ourmazd, A., Frank, J., McLellan, J. S., Chong, L. T., and Amaro, R. E. (2021) A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nat. Chem.*, 1–6

204. Narimatsu, Y., Büll, C., Chen, Y.-H., Wandall, H. H., Yang, Z., and Clausen, H. (2021) Genetic glycoengineering in mammalian cells. *Journal of Biological Chemistry* 296, 100448

205. Hirschberg, K., and Lippincott-Schwartz, J. (1999) Secretory pathway kinetics and in vivo analysis of protein traffic from the Golgi complex to the cell surface. *FASEB J* 13 Suppl 2, S251-256

206. Burleigh, S. C., van de Laar, T., Stroop, C. J. M., van Grunsven, W. M. J., O'Donoghue, N., Rudd, P. M., and Davey, G. P. (2011) Synergizing metabolic flux analysis and nucleotide sugar metabolism to understand the control of glycosylation of recombinant protein in CHO cells. *BMC Biotechnol* 11, 95

207. Cobb, B. A. (2020) The history of IgG glycosylation and where we are now. *Glycobiology* 30, 202–213

208. Rose, D. R. (2012) Structure, mechanism and inhibition of Golgi α-mannosidase II. *Current Opinion in Structural Biology* 22, 558–562

209. Mukai, Y., Nakamura, T., Yoshikawa, M., Yoshioka, Y., Tsunoda, S., Nakagawa, S., Yamagata, Y., and Tsutsumi, Y. (2010) Solution of the Structure of the TNF-TNFR2 Complex. *Sci. Signal.* 3,

210. Syed, R. S., Reid, S. W., Li, C., Cheetham, J. C., Aoki, K. H., Liu, B., Zhan, H., Osslund, T. D., Chirino, A. J., Zhang, J., Finer-Moore, J., Elliott, S., Sitney, K., Katz, B. A., Matthews, D. J., Wendoloski, J. J., Egrie, J., and Stroud, R. M. (1998) Efficiency of signalling through cytokine receptors depends critically on receptor orientation. *Nature* 395, 511–516

211. Isoda, Y., Yagi, H., Satoh, T., Shibata-Koyama, M., Masuda, K., Satoh, M., Kato, K., and Iida, S. (2015) Importance of the Side Chain at Position 296 of Antibody Fc in Interactions with FcγRIIIa and Other Fcγ Receptors. *PLoS ONE* 10, e0140120

212. Tian, G., Xiang, S., Noiva, R., Lennarz, W. J., and Schindelin, H. (2006) The Crystal Structure of Yeast Protein Disulfide Isomerase Suggests Cooperativity between Its Active Sites. *Cell* 124, 61–73

213. Moremen, K. W., Ramiah, A., Stuart, M., Steel, J., Meng, L., Forouhar, F., Moniz, H. A., Gahlay, G., Gao, Z., Chapla, D., Wang, S., Yang, J.-Y., Prabhakar, P. K., Johnson, R., Rosa, M. dela, Geisler, C., Nairn, A. V., Seetharaman, J., Wu, S.-C., Tong, L., Gilbert, H. J., LaBaer, J., and Jarvis, D. L. (2018) Expression system for structural and functional studies of human glycosylation enzymes. *Nature Chemical Biology* 14, 156–162

214. Xiao, T., Lu, J., Zhang, J., Johnson, R. I., McKay, L. G. A., Storm, N., Lavine, C. L., Peng, H., Cai, Y., Rits-Volloch, S., Lu, S., Quinlan, B. D., Farzan, M., Seaman, M. S., Griffiths, A., and Chen, B. (2021) A trimeric human angiotensin-converting enzyme 2 as an anti-SARS-CoV-2 agent. *Nat Struct Mol Biol* 28, 202–209

215. Meng, L., Forouhar, F., Thieker, D., Gao, Z., Ramiah, A., Moniz, H., Xiang, Y., Seetharaman, J., Milaninia, S., Su, M., Bridger, R., Veillon, L., Azadi, P., Kornhaber, G., Wells, L., Montelione, G. T., Woods, R. J., Tong, L., and Moremen, K. W. (2013) Enzymatic basis for N-glycan sialylation: structure of rat α2,6-sialyltransferase (ST6GAL1) reveals conserved and unique features for glycan sialylation. *J Biol Chem* 288, 34680–34698

216. Walsh, G., and Jefferis, R. (2006) Post-translational modifications in the context of therapeutic proteins. *Nat Biotechnol* 24, 1241–1252

217. Kaur, H. (2021) Characterization of glycosylation in monoclonal antibodies and its importance in therapeutic antibody development. *Critical Reviews in Biotechnology* 41, 300–315

218. Hossler, P., Khattak, S. F., and Li, Z. J. (2009) Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology* 19, 936–949

219. Moremen, K. W., and Haltiwanger, R. S. (2019) Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. *Nature Chemical Biology* 15, 853–864

220. Kong, L., Lee, J. H., Doores, K. J., Murin, C. D., Julien, J.-P., McBride, R., Liu, Y., Marozsan, A., Cupo, A., Klasse, P.-J., Hoffenberg, S., Caulfield, M., King, C. R., Hua, Y., Le, K. M., Khayat, R., Deller, M. C., Clayton, T., Tien, H., Feizi, T., Sanders, R. W., Paulson, J. C., Moore, J. P., Stanfield, R. L., Burton, D. R., Ward, A. B., and Wilson, I. A. (2013) Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. *Nat Struct Mol Biol* 20, 796–803

221. Hassinen, A., Rivinoja, A., Kauppila, A., and Kellokumpu, S. (2010) Golgi N-Glycosyltransferases Form Both Homo- and Heterodimeric Enzyme Complexes in Live Cells. *Journal of Biological Chemistry* 285, 17771–17777

222. Polasky, D. A., and Nesvizhskii, A. I. (2023) Recent advances in computational algorithms and software for large-scale glycoproteomics. *Current Opinion in Chemical Biology* 72, 102238

223. Hoffmann, M., Pioch, M., Pralow, A., Hennig, R., Kottler, R., Reichl, U., and Rapp, E. (2018) The Fine Art of Destruction: A Guide to In-Depth Glycoproteomic Analyses-Exploiting the Diagnostic Potential of Fragment Ions. *PROTEOMICS* 18, 1800282

224. Neelamegham, S., Aoki-Kinoshita, K., Bolton, E., Frank, M., Lisacek, F., Lütteke, T., O'Boyle, N., Packer, N. H., Stanley, P., Toukach, P., Varki, A., Woods, R. J., and SNFG Discussion Group (2019) Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology* 29, 620–624

225. Duan, H., Chen, X., Boyington, J. C., Cheng, C., Zhang, Y., Jafari, A. J., Stephens, T., Tsybovsky, Y., Kalyuzhniy, O., Zhao, P., Menis, S., Nason, M. C., Normandin, E., Mukhamedova, M., DeKosky, B. J., Wells, L., Schief, W. R., Tian, M., Alt, F. W., Kwong, P. D., and Mascola, J. R. (2018) Glycan Masking Focuses Immune Responses to the HIV-1 CD4-Binding Site and Enhances Elicitation of VRC01-Class Precursor Antibodies. *Immunity* 49, 301-311.e5

226. Escolano, A., Gristick, H. B., Abernathy, M. E., Merkenschlager, J., Gautam, R., Oliveira, T. Y., Pai, J., West, A. P., Barnes, C. O., Cohen, A. A., Wang, H., Golijanin, J., Yost, D., Keeffe, J. R., Wang, Z., Zhao, P., Yao, K.-H., Bauer, J., Nogueira, L., Gao, H., Voll, A. V., Montefiori, D. C., Seaman, M. S., Gazumyan, A., Silva, M., McGuire, A. T., Stamatatos, L., Irvine, D. J., Wells, L., Martin, M. A., Bjorkman, P. J., and Nussenzweig, M. C. (2019) Immunization expands B cells specific to HIV-1 V3 glycan in mice and macaques. *Nature* 570, 468–473

227. Wang, S., Voronin, Y., Zhao, P., Ishihara, M., Mehta, N., Porterfield, M., Chen, Y., Bartley, C., Hu, G., Han, D., Wells, L., Tiemeyer, M., and Lu, S. (2020) Glycan Profiles of gp120 Protein Vaccines from Four Major HIV-1 Subtypes Produced from Different Host Cell Lines under Non-GMP or GMP Conditions. *J Virol* 94, e01968-19

228. Riley, N. M., Hebert, A. S., Westphall, M. S., and Coon, J. J. (2019) Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. *Nature Communications* 10,

229. Saunders, K. O., Verkoczy, L. K., Jiang, C., Zhang, J., Parks, R., Chen, H., Housman, M., Bouton-Verville, H., Shen, X., Trama, A. M., Scearce, R., Sutherland, L., Santra, S., Newman, A., Eaton, A., Xu, K., Georgiev, I. S., Joyce, M. G., Tomaras, G. D., Bonsignori, M., Reed, S. G., Salazar, A., Mascola, J. R., Moody, M. A., Cain, D. W., Centlivre, M., Zurawski, S., Zurawski, G., Erickson, H. P., Kwong, P. D., Alam, S. M., Levy, Y., Montefiori, D. C., and Haynes, B. F. (2017) Vaccine Induction of Heterologous Tier 2 HIV-1 Neutralizing Antibodies in Animal Models. *Cell Reports* 21, 3681–3690

230. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* 9, 3878–3888

231. Marze, N. A., Roy Burman, S. S., Sheffler, W., and Gray, J. J. (2018) Efficient flexible backbone protein–protein docking for challenging targets. *Bioinformatics* 34, 3461–3469

232. Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E., and Kolinski, A. (2016) Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* 116, 7898–7936

233. Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* 8, 1542–1555

234. Losfeld, M.-E., Scibona, E., Lin, C., and Aebi, M. (2022) Glycosylation network mapping and site-specific glycan maturation in vivo. *iScience* 25, 105417

235. Narimatsu, Y., Joshi, H. J., Nason, R., Van Coillie, J., Karlsson, R., Sun, L., Ye, Z., Chen, Y.-H., Schjoldager, K. T., Steentoft, C., Furukawa, S., Bensing, B. A., Sullam, P. M., Thompson, A. J., Paulson, J. C., Büll, C., Adema, G. J., Mandel, U., Hansen, L., Bennett, E. P., Varki, A., Vakhrushev, S. Y., Yang, Z., and Clausen, H. (2019) An Atlas of Human Glycosylation Pathways Enables Display of the Human Glycome by Gene Engineered Cells. *Molecular Cell*,

236. Elliott, S., Lorenzini, T., Asher, S., Aoki, K., Brankow, D., Buck, L., Busse, L., Chang, D., Fuller, J., Grant, J., Hernday, N., Hokum, M., Hu, S., Knudten, A., Levin, N., Komorowski, R., Martin, F., Navarro, R., Osslund, T., Rogers, G., Rogers, N., Trail, G., and Egrie, J. (2003) Enhancement of therapeutic protein in vivo activities through glycoengineering. *Nature Biotechnology* 21, 414–421