DYNAMIC ISOTOPIC DETECTION OF AMINOSUGARS WITH GLUTAMINE (IDAWG) FOR RELEASED GLYCANS AND O-GLCNAC MODIFIED PROTEINS

by

CHELSEA DESBIENS

(Under the Direction of

LANCE WELLS)

ABSTRACT

The Isotopic Detection of Aminosugars With Glutamine (IDAWG) method was originally developed for the glycomics field as a quantitative tool that takes advantage of the hexosamine biosynthetic pathway, isotopically labeling nitrogen-containing glycans in cell culture systems via the use of ^{15}N-Gln. Here, we present an adaptation of this method, Dynamic IDAWG, that allows for the calculation of half-lives of released glycans in a given sample following analyses by mass spectrometry. Simultaneously, the amount of sialic acid recycling for an identified glycan can also be calculated. An additional benefit to using this method is that the cycling rates of the post-translational modification O-linked β-N-acetylglucosamine (O-GlcNAc) can also be determined. O-GlcNAc is found on thousands of nuclear and cytosolic proteins in mammals and is thought to be a regulatory modification playing a role in a variety of cellular processes. O-GlcNAc is thought to be a dynamic modification; that is, the modification exhibits a shorter turnover rate than

the modified protein. However, dynamics have only been evaluated on a small number of O-GlcNAc modified proteins due to the laborious and insensitive methods that are available. Therefore, there is an urgent need in the field to develop high-throughput, sensitive methods to evaluate the dynamics of O-GlcNAc on a global scale. Here, we illustrate the utility of Dynamic IDAWG in cell culture systems to evaluate the turnover of complex N- and O-linked glycans as well as delineating the turnover rate of O-GlcNAc on nuclear and cytosolic proteins.

INDEX WORDS: Glycomics, O-GlcNAc, Released Glycans, Dynamics, IDAWG, Glycoproteomics

DYNAMIC ISOTOPIC DETECTION OF AMINOSUGARS WITH GLUTAMINE (IDAWG) FOR RELEASED GLYCANS AND O-GLCNAC MODIFIED PROTEINS

by

CHELSEA DESBIENS

BS, Valdosta State University, 2015

BS, Valdosta State University, 2016

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

DCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2021

DYNAMIC ISOTOPIC DETECTION OF AMINOSUGARS WITH GLUTAMINE (IDAWG) FOR RELEASED GLYCANS AND O-GLCNAC MODIFIED PROTEINS

by

CHELSEA DESBIENS

Major Professor:	Lance Wells
Committee:	Jeffrey Urbauer
	Jonathan Amster

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
December 2021

DEDICATION

To never giving up.

ACKNOWLEDGEMENTS

First and foremost, I want to thank my advisor, Dr. Lance Wells, and the rest of my committee, Dr. Jeffrey Urbauer and Dr. Jonathan Amster, for encouraging me and pushing me through this entire process to become the independent scientist that I am now. I am very thankful to have worked with such a wonderful group of lab mates both past and present over the years, but I am especially grateful for Hannah Stephen, Stephanie Halmo, and Rob Bridger. I couldn't have asked for a more supportive and fun work environment, it is one that I will truly miss!

To my amazing parents and brother: you all have supported me and all of my goals and for that I am grateful. Thank you for nodding along while I try to explain the science and all of the different projects I've worked on over the years.

To whom I can only describe as the best group of friends and constant support system anyone could ask for: Kellie and Adam Downey, Samantha and Jonathan Ramey, Katie and Logan Brown, Colleen Bradley, Carla and Logan Franks, Lettie and Jeff Wysong, Bianca Babcock, Melissa and Ian Delahunty, Marilyn Naughton, Samantha Brown, Brad Gall, Rob Williams, Hope Patrick, and Danielle Revenue. Whether you were here in Athens with me these last 5 years or cheering me on from far away, you all are like family to me and I don't know how I got so lucky to have all of you in my life.

Lastly, to my soon to be husband and best friend. Malcolm, you and I have been through a lot together throughout our studies since undergrad and throughout our relationship. I am so blessed and so lucky to have had you in my corner cheering me on, being my rock, and my shoulder to cry on before, during, and soon after graduate school. You have done more for me than you will ever know!

Thank you all!

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Introduction to Dissertation

This chapter serves as an introduction and overview to the dissertation and a literature review covering topics that will be discussed throughout. The glycobiology field currently lacks effective methods to study the turnover of glycans and effectively compare those dynamics across cell types or disease states. The original Isotopic Detection of Aminosugars With Glutamine (IDAWG) takes advantage of the hexosamine biosynthetic pathway in cell culture, where ^{15}N is incorporated into GlcNAc, GalNAc, and Sialic Acid residues through ^{15}N-Gln supplementation. This allowed for quantification of present glycans to be performed without the need for radiolabeling. Herein, we propose the method termed Dynamic IDAWG which will allow for the quantification of glycans, defining the half-life of structures, as well as delve into the extent recycling of sialic acids is occurring. There are two main focal points of this dissertation that center around the development of this Dynamic IDAWG method. First, in chapter 2 we will discuss Dynamic IDAWG for released glycans studies. Through the combination of glutamine supplementation in cell culture, pulse-chase, and mass spectrometry, we have developed a high-throughput method capable of determining the half-life for O-linked glycans. In addition to this, the amount of recycling occurring in a

subset of structures containing sialic acid can be approximated simultaneously. Second, in chapter 3 we turn our attention to the post-translational modification O-linked β-N-acetylglucosamine (O-GlcNAc) and its ability to cycle on and off proteins due to the presence of the enzymes O-GlcNAc Transferase (OGT) and O-GlcNAcase (OGA) respectively. O-GlcNAc is known to play a number of roles ranging from transcription, neural development, and stress response, just to name a few. However, with this modification's ability to cycle on and off proteins it is necessary to define an effective high-throughput and accessible method, Dynamic IDAWG, that allows for the turnover of O-GlcNAc to be probed for specific proteins and specific modification sites. With some modifications from the original method presented here for released glycans, this secondary version will allow the necessary insight to understand the impact cycling rates play as well as how they can change under various conditions. Further application of these techniques can allow for a deeper understanding of O-GlcNAc's role across different diseases and cellular processes.

Analytical Techniques for Glycomics and Glycoproteomics

Glycomic analyses enable researchers to characterize the oligosaccharides that are released from the proteins they modify. The field also centers around defining functional roles glycosylation plays throughout biological systems.[1,2] Profiling glycan expression is necessary to understand glycoconjugate functions and changes that may occur.[3] What makes the study of glycomes so challenging is that there is no model for what is structure is expected during the formation of glycans. This is due to the complex set of glycosyltransferases and glycosidases

competing in the biosynthesis and processing of glycan structures throughout production.[2]

There are a number of mass spectrometric techniques and strategies used for conducting glycomics studies. Matrix assisted laser desorption ionization (MALDI-MS), electrospray ionization (ESI-MS), and liquid chromatography mass spectrometry (LC-MS) are among those techniques.[4] MALDI and ESI are widely used where the glycan is detected as a metal adduct, commonly sodium or lithium is used, and they can be detected as protonated or deprotonated species; glycopeptides can be detected in a similar fashion.[5] While both can provide a glycan profile, ESI does generate multiply charged species that can complicate analyses. MALDI has its own limitations, where it is unable to separate isomers and labile groups (i.e. sialic acid and fucose) are sometimes lost. Tandem LC-MS however, allows for more definitive structural assignments and can resolve the permethylated species to quantify individual glycans, where this can also be completed by direction infusion MS. For tandem MS, collision induced dissociation (CID) is utilized and produces both glycosidic and cross-ring cleavages aiding in composition, but also deducing the type of linkage that is employed for that glycan.[5]

In order to quantify the glycan profile observed for a given biological sample, isotopic labeling is broadly used in order to quantify at the MS^1 level where heavy and light labeled glycans can be analyzed based on their isotopic peaks and the shift observed by the heavy labeled species. $^{12}C/^{13}C$ and ^{18}O are two of the more commonly used isotopes used for glycan labeling.[3] These produce a +1 and +2 Da shift respectively. Labeling efficiency is however a concern with isotopic

labeling and some researchers have turned to stable isotopic labeling of amino sugars in cell culture (SILAC).[6] Although primarily used in proteomics studies, the original IDAWG method makes this metabolic labeling strategy useful for glycomics quantification as well.[3] This methodology will be discussed at length in a later section.

Our ability to quantify at the glycopeptide level is slightly more complicated. While glycomics analysis allows us to see the bigger picture and overall heterogeneity within a glycoproteome, comprehensive characterization of a glycoprotein is desired.[3] In recent years though there have been advancements in sample prep and overall functionality of instrumentation. Previously discussed SILAC experiments using heavy lysine and arginine have been widely used as a metabolic incorporation approach for glycopeptide quantification as well. Another approach, termed isotope-targeted glycoproteomics (IsoTaG), incorporates azide containing sugars. In brief, the azide sugar is incorporated into glycans which is then joined with a biotin tag that allows for glycopeptide enrichment via streptavidin beads. When the tag is later cleaved, targeted mass spectrometry can characterize the observed mass shift.[3] While the field has a number of quantitative approaches available for glycomic and glycoproteomic studies, there is still a need for the development and expansion of bioinformatic tools that are accessible and can accurately analyze such data.

Glycan Turnover

The processing of glycans is accomplished through a variety of glycosyltransferases and glycosidases. The synthesis and degradation of glycans

is extremely structured, where each enzyme involved has a defining role. Glycosyltransferases generally have high specificity for their donor and acceptor substrates as they elongate glycan moieties and also initiate the biosynthesis of glycoproteins. There are differences in where N- and O-glycans are initiated. For eukaryotes, we find that N-glycans are cotranslationally transferred in the ER to Asparagine residues, before passing through the Golgi for processing, giving rise to high mannose, hybrid, and complex N-glycans. In contrast, mucin-type O-glycan production is initiated in the ER where a single monosaccharide is added to Serine or Threonine residues. The moieties are then extended in the Golgi following the folding of the protein. In the lysosome, we find most glycoconjugates being degraded, while a subset of those structures can be salvaged for further synthesis.

The O-GlcNAc Transferase (OGT)

Highly conserved across eukaryotes, OGT, and subsequently the O-GlcNAc modification, is essential for life, where loss of the enzyme is fatal in the embryonic stage[7]. The transferase has been mapped to a region of the X-chromosome, specifically on a locus near the centromere, that has notably been linked to Parkinson's disease[8]. OGT is divided into two different domains: a catalytic domain located at the C-terminus and a tetratricopeptide repeat (TPR) domain located at the N-terminus[7–10]. There are several tandem repeats of the TPR motif in OGT, leading researchers to believe OGT interacts with proteins by means of the TPR domain, making it a regulatory complex[11]. There are various forms of OGT, where the full length OGT is termed nucleocytoplasmic OGT (ncOGT) and it consists of 13.5 encoded TPRs[9]. Thousands of proteins have been

identified as targets for OGT, however the cycling rates of the O-GlcNAc modification has only been established for a portion of these targets. While it has been established that dynamically the addition of O-GlcNAc and phosphate share similarities, phosphorylation differs in the sense that it is determined by hundreds of kinases rather than a singular enzyme as with OGT. However, they are sometimes found in the same enzyme complex[8].

The O-GlcNAc Modification

The essential and dynamic post-translational modification (PTM), O-GlcNAc, is found on thousands of nuclear and cytosolic proteins. It has been shown to be conserved from *Caenorhabditis elegans* to humans and is thought to be a regulatory modification playing a role in a variety of cellular processes[12]. The discovery of O-GlcNAc in 1984 contradicted previous hypotheses stating that glycosylation was restricted only to the luminal domain and the cell surface[12–15]. The hexosamine biosynthetic pathway (HBP) is responsible for production of uridine diphosphate N-acetylglucosamine (UDP-GlcNAc), the donor molecule for O-GlcNAc prior to glycosylation of serine and threonine residues. Here, 2-5% of available fructose-6-phosphate is diverted from glycolysis for utilization in the HBP to synthesize glucosamine-6-phosphate, marking the start of this pathway[16,17]. The resulting UDP-GlcNAc then goes onto the Golgi and ER for O-linked and N-linked glycosylation in addition to modifying nuclear and cytosolic proteins by a single O-GlcNAc moiety. This takes place due to a single gene that encodes for the enzyme O-GlcNAc-transferase (OGT), allowing for the addition of O-GlcNAc onto Serine (Ser) and Threonine (Thr) residues. another gene encodes for O-GlcNAc-

hydrolase (OGA), which facilitates the removal of the aforementioned modification[8].

In contrast to other known forms of glycosylation, the O-GlcNAc modification has several unique and distinguishing qualities. This modification is found exclusively on nuclear and cytosolic proteins, whereas membrane and secretory proteins travel through the endoplasmic reticulum (ER)-Golgi pathway and undergo glycosylation. In contrast to other glycan modifications that are extended to become complex moieties, O-GlcNAc has the ability to exist as a single monosaccharide. As mentioned previously, this PTM also has the ability to cycle on and off modified proteins and is hypothesized to be dynamic where it exhibits a shorter half-life than the protein in which it modifies[8,18]. Furthermore, it is believed that dynamics vary among different modification sites as well as the proteins that are modified, where levels, localization, and kinetics of cycling influence this property[18,19]. This is in contrast to extracellular glycosylation which remains static regardless of changes in the cellular environment[20]. The cycling of O-GlcNAc is seen as analogous to phosphorylation, where studies have shown that there is extensive influence amongst the two for the addition and removal of one another, or cross talk, impacting the regulation of various cellular processes[8,21,22]. Evidence has also suggested that O-GlcNAc serves a regulatory role, impacting protein-protein interactions, as well as protein activity and stability.

Additionally, evidence shows O-GlcNAc modifications act as nutrient sensors within the cellular environment in processes including stress, insulin signaling, and gene transcription[8,20,22,23]. As a cellular response, the quantity of O-

GlcNAc modified proteins increases as cells are exposed to various forms of stress[24]. This reaction acts as a defense mechanism against disruptions in the system. Examples of this include increased O-GlcNAc modifications on specific proteins following a heart attack in order to protect cardiac muscle[8]. Alternatively, when considering gene transcription, RNA polymerase II transcription factors are typically found to have multiple sites modified by O-GlcNAc. This is necessary for the pre-initiation complex to form. The removal of these modifications is then necessary for transcriptional elongation to occur preceding the addition of phosphate to the C-terminal repeat domain of the polymerase[8,14].

It is apparent O-GlcNAc plays a key regulatory role in many aspects of cellular systems. Because of this, this PTM's involvement in various chronic diseases should not come as a shock[8,14,25,26]. Hyperglycemia, for example, occurs as a result of an imbalance between O-GlcNAc and phosphorylation, where an upregulation of O-GlcNAc moieties hinders signaling controlled by phosphorylation and changes in the regulation of transcription can be observed[8]. While this increase of O-GlcNAc modifications is not harmful in the short-term, extended periods of increased modifications due to hyperglycemia can result in unfavorable changes in cell functionality. Transcription factors, such as Sp1, which is linked to diabetes, are modified by O-GlcNAc. In fact, under these conditions transcription factors are hyper-O-GlcNAc modified in diabetes patients causing deviations in functionality[8,20]. For Sp1 specifically, it is thought that the over-abundance of O-GlcNAc sites affects promotor specificity[20]. In the brain, there is an ample amount of O-GlcNAc modified proteins, where changes in the metabolism of glucose has

been linked to Alzheimer's disease, as well as Parkinson's disease[8,25]. Glucose metabolism is hindered causing an overall decrease of O-GlcNAc in the brain leading to tau phosphorylation, however all proteins associated with the disease have been found to still be modified by O-GlcNAc[8]. There are obviously distinct differences in the role O-GlcNAc plays in each disease, however the potential changes in cycling, and in turn the abundance, of the modification is a key factor in each disease, as well as the impacts on cross-talk with phosphorylation and other protein translational modifications[23,27,28]. Unfortunately, dynamics have only been evaluated on a small number of O-GlcNAc modified proteins due to the laborious and insensitive methods utilized. Therefore, there is a need in the field to develop high-throughput, sensitive methods to evaluate the dynamics of O-GlcNAc and O-GlcNAc modified proteins to address these changes and how they contribute to disease.

While O-GlcNAc is an abundant modification it is more difficult to detect with standard analytical methods. The modification can go undetected with typical methods, being lost at the source when introduced to conditions for electrospray mass spectrometry. Similarly, the modification is extremely labile under collision-induced fragmentation, resulting in poor peptide fragmentation[29,30]. This culminates in low signals for the modified peptides compared to the signals generated from unmodified peptides making it difficult to accurately quantify glycan structures present. There are *in vitro* methods that exist for quantification of glycans, using label free and isotope tags alike, however samples are not enriched for the O-GlcNAc modification specifically making it necessary to work with pure

O-GlcNAc peptides for the method to work well[31,32]. In order to further study the dynamics of this specific modification and identify modified proteins, it is necessary for sophisticated methods to be developed that select specifically for the O-GlcNAc modification.

Enriching of O-GlcNAc Modified Proteins

There are a variety of enrichment tools that have been developed over the years that have allowed researchers to enrich for O-GlcNAc modified proteins such as antibodies, chemoenzymatic derivatization, and hydrophilic interaction liquid chromatography (HILIC) to name a few[33,34]. Lectin based enrichment is also among these strategies known to be useful tools in the enrichment of glycoproteins. More specifically, the lectin wheat-germ agglutinin (WGA) is a well-known for its ability to bind both GlcNAc and Sialic Acid residues[35–37]. WGA is capable of binding tightly to the terminal GlcNAc residues due to the formation of dimers with two binding sites. Studies have found that WGA typically has a lower binding affinity than desired for an O-GlcNAc peptide enrichment strategy, but passing those same peptides over a WGA column, also known as lectin weak affinity chromatography (LWAC), has shown to be effective[38]. A secondary type of WGA, succinylated WGA (sWGA), is also available for enrichment and in contrast is incapable of binding to sialic acid residues. It is important to note that enriching at the protein level is more desirable versus enriching at the peptide level as there is a possibility for peptides to be modified by prototypical glycans that will interfere with enrichment[39]. Taking advantage of this enrichment method is more desirable

because of its accessibility to researchers in contrast to other methods that tend to be more exclusive and laborious.

Isotopic Detection of Amino Sugars with Glutamine (IDAWG)

Isotopic Detection of Amino Sugars With Glutamine, or IDAWG, was originally developed as a *in vivo* quantification tool for the glycomics field, allowing for the incorporation of mass tags into glycan structures of cultured cells[40,41]. IDAWG was originally designed as a tool to look at released glycans rather than intact glycopeptides. This method takes advantage of the necessity for Glutamine supplementation into cell culture media, amide-^{15}N-Gln ("heavy Glutamine") is supplemented into culture media lacking standard amide-^{14}N-Gln ("light Glutamine"). This allows for the ^{15}N tag to be incorporated exclusively into the Gln residues in protein backbones and select glycan structures through the HBP. The glycolysis intermediate fructose-6-phosphate is first converted to glucosamine-6-phosphate in the rate limiting step of the pathway. This conversion is due to the interaction with the present source of Glutamine, where the side chain amide is the sole source of nitrogen for the final product of the pathway, UDP-GlcNAc (**Fig. 1.1**).[40,41] Using Gln for isotopic labeling is ideal as it limits where the tag will be incorporated. Gln involved in the start of the HBP loses its amide group and is converted into Glutamic Acid (Glu) which is the start for other amino acids to be synthesized, therefore limiting the ^{15}N tag in amino acids solely to Gln. The Gln present that is not involved in the HBP can be found incorporated in the peptide backbone of proteins. Incorporation of amide-^{15}N-Gln acts as the sole source of

nitrogen for the synthesis of GlcNAc, GalNAc, and sialic acid in the HBP increases these molecular weights by one Dalton[40,41].

Following harvest of the labeled cells, N- and O-linked are released, via PNGase F digestion and β-elimination respectively, and permethylated as previously described and subjected to MS analysis[32,42]. Identification of present glycans can be completed and also confirmed by manual interpretation of glycan structures present based on fragmentation patterns[40]. This method has been shown to result in virtually complete incorporation of ^{15}N into both N-linked and O-linked glycans and subsequently allows for quantification and provides additional information for glycan fragmentation when samples are subjected to mass spectrometry analysis[40].

While IDAWG is powerful alone for quantifying released glycans, it can be adapted further. The foundation of this method, coupled with pulse-chase analysis, gives us what we have termed Dynamic IDAWG. We benefit from this incorporation of pulse-chase analysis in a couple of different ways. First, in relation to released glycans, the half-life of identified glycans can be calculated by following the isotopic peak distributions for the heavy and light forms of a structure over time. Simultaneously, we can determine the amount of remodeling occurring for sialylated structures. Second, Dynamic IDAWG can be used for glycoproteomics experiments in order to study intact O-GlcNAc modifications. Because O-GlcNAc exists as a labile moiety, the methodology is combined with specific mass spectrometry techniques, incorporating triggered ETD and stepped HCD fragmentation alike. In a similar fashion to the workup of released glycans, the

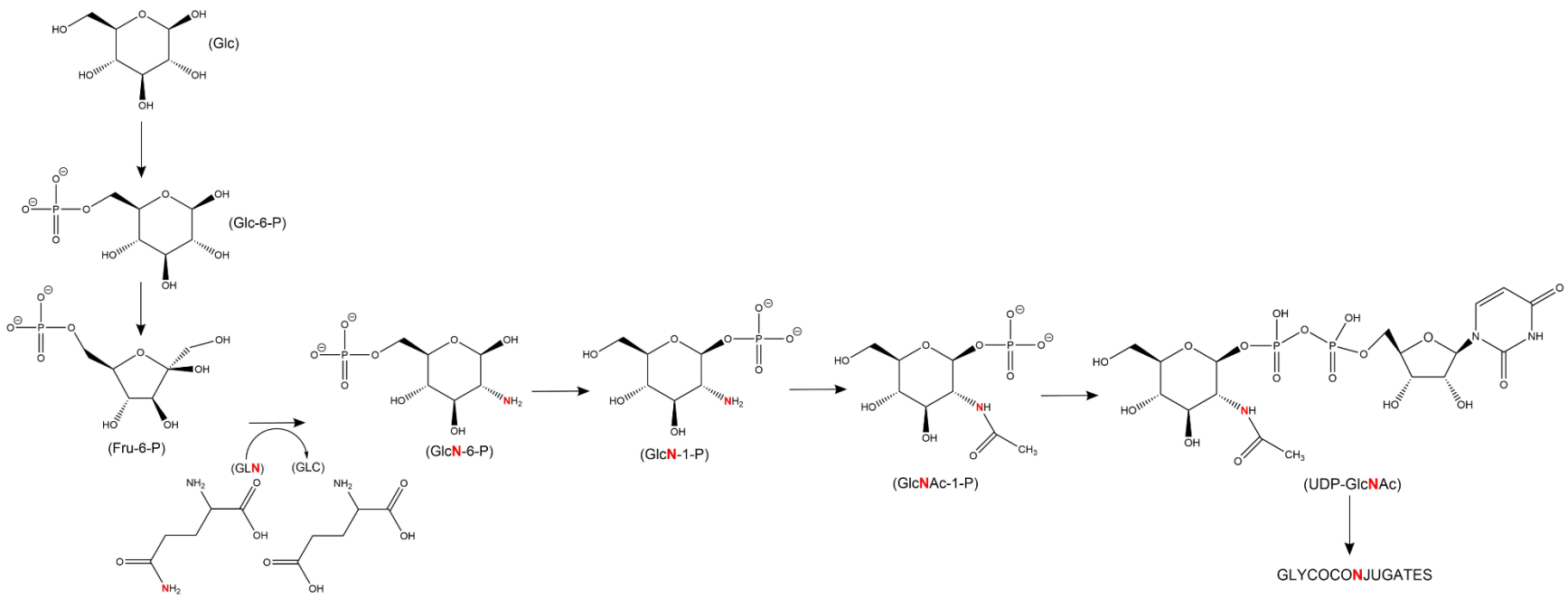


Figure 1.1: The Hexosamine Biosynthetic Pathway (HBP). Schematic describing the HBP where the Fru-6-P is first converted to GlcN-6-P in the reaction at the rate limiting step of the pathway. The ^{15}N is gained form the amide side chain of Gln supplemented in cell culture. The ^{15}N tag is taken throughout the pathway up until the production of UDP-GlcNAc, in addition to UDP-GalNAc and CMP-Neu5Ac. The mass of these aminosugars will increase by 1 Da.

overall dynamics of the modification can be defined at numerous sites across identified proteins in your cell line of choice based on the spectral data[43]. HCD produces b and y ions predominately, which can be further fragmented. This fragmentation incorporates a wider mass range and produces highly accurate MS2 spectra, allowing diagnostic peaks of PTMs to be recognized that would otherwise go undetected. In contrast, ETD is known for its ability to aid in the structural characterization of a peptide allowing for cleavage of the peptide backbone and generating c and z ions. This type of fragmentation is useful when studying labile PTMs such as O-GlcNAc and allows for the identification of peptide sequences and location of modifications[5,30,43]. Applying this method to released glycans can help identify changes to a system due to various stressors or how the glycome is altered in various disease states. More importantly, Dynamic IDAWG can be adapted for other studies such as dynamics of O-GlcNAc specifically, assisting in closing the gap in knowledge attributed to O-GlcNAc cycling rates and the role those cycling rates play in various diseases and disorders including neurological disorders, cancer, and Alzheimer's[8].

References

1. Ruhaak, L. R., Xu, G., Li, Q., Goonatilleke, E. & Lebrilla, C. B. Mass Spectrometry Approaches to Glycomic and Glycoproteomic Analyses. *Chem. Rev.* **118,** 7886–7930 (2018).
2. Kailemia, M. J. *et al.* Recent Advances in the Mass Spectrometry Methods for Glycomics and Cancer. *Anal. Chem.* acs.analchem.7b04202 (2017). doi:10.1021/acs.analchem.7b04202
3. Delafield, D. G. & Li, L. Recent advances in analytical approaches for glycan and glycopeptide quantitation. *Mol. Cell. Proteomics* **20,** 0–21 (2021).
4. Mechref, Y., Hu, Y., Desantos-Garcia, J. L., Hussein, A. & Tang, H. Quantitative glycomics strategies. *Mol. Cell. Proteomics* **12,** 874–884 (2013).
5. Han, L. & Costello, C. E. Mass spectrometry of glycans. *Biochem.* **78,** 710–720 (2013).
6. Ong, S. E. The expanding field of SILAC. *Anal. Bioanal. Chem.* **404,** 967–976 (2012).
7. Iyer, S. P. N. & Hart, G. W. Roles of the Tetratricopeptide Repeat Domain in O -GlcNAc Transferase Targeting and Protein Substrate Specificity. *J. Biol. Chem.* **278,** 24608–24616 (2003).
8. Hart, G. W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling,

Transcription, and Chronic Disease. *Annu. Rev. Biochem.* **80,** 825–858 (2011).

9. Yang, X. & Qian, K. Protein O-GlcNAcylation: emerging mechanisms and functions. *Nat. Rev. Mol. Cell Biol.* **18,** 452–465 (2017).
10. Worth, M., Li, H. & Jiang, J. Deciphering the Functions of Protein O-GlcNAcylation with Chemistry. *ACS Chem. Biol.* **12,** 326–335 (2017).
11. Kreppel, L. K., Blomberg, M. A. & Hart, G. W. Dynamic Glycosylation of Nuclear and Cytosolic Proteins. *J. Biol. Chem.* **272,** 9308–9315 (1997).
12. Torres, C.-R. & Hart, G. W. Topography and Polypeptide Distribution of Terminal N-Acetylglucosamine Residues on the Surfaces of Intact Lymphocytes. *J. Biol. Chem.* **259,** 3308–3317 (1984).
13. Butkinaree, C., Park, K. & Hart, G. W. O-linked β-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. *Biochim. Biophys. Acta - Gen. Subj.* **1800,** 96–106 (2010).
14. Hardivillé, S. & Hart, G. W. Nutrient regulation of signaling, transcription, and cell physiology by O- GlcNAcylation. *Cell Metab.* **20,** 208–213 (2014).
15. Wells, L., Vosseller, K. & Hart, G. W. Glycosylation of Nucleocytoplasmic Proteins : Signal Transduction and O-GlcNAc. *Science (80-.).* **291,** 2376–2379 (2001).
16. Ruba, A. & Yang, W. O-GlcNAc-ylation in the Nuclear Pore Complex. *Cell. Mol. Bioeng.* **9,** 227–233 (2016).
17. Vaidyanathan, K., Durning, S., Wells, L. & Carbohydrate, C. Functional O-

GlcNAc modifications: Implications in molecular regulation and pathophysiology. *Crit Rev Biochem Mol Biol* **49,** 140–163 (2016).

18. Ma, J. & Hart, G. W. O-GlcNAc profiling: From proteins to proteomes. *Clin. Proteomics* **11,** 1–16 (2014).
19. Hart, G. W., Housley, M. P. & Slawson, C. Cycling of O-linked β - N - acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446,** 1017–1022 (2007).
20. Slawson, C., Housley, M. P. & Hart, G. W. O-GlcNAc cycling: How a single sugar post-translational modification is changing the way we think about signaling networks. *J. Cell. Biochem.* **97,** 71–83 (2006).
21. Hanover, J. A., Krause, M. W. & Love, D. C. The hexosamine signaling pathway: O-GlcNAc cycling in feast or famine. *Biochim. Biophys. Acta - Gen. Subj.* **1800,** 80–95 (2010).
22. Wells, L. & Hart, G. W. O-GlcNAc turns twenty: Functional implications for post-translational modification of nuclear and cytosolic proteins with a sugar. *FEBS Lett.* **546,** 154–158 (2003).
23. Wells, L., Whalen, S. A. & Hart, G. W. O-GlcNAc: A regulatory post-translational modification. *Biochem. Biophys. Res. Commun.* **302,** 435–441 (2003).
24. Zachara, N. E. *et al.* Dynamic O-GlcNAc modification of nucleocytoplasmic proteins in response to stress: A survival response of mammalian cells. *J. Biol. Chem.* **279,** 30133–30142 (2004).
25. Banerjee, P. S., Lagerlöf, O. & Hart, G. W. Roles of O-GlcNAc in chronic

diseases of aging. *Mol. Aspects Med.* **51,** 1–15 (2016).

26. Bond, M. R. & Hanover, J. A. *O*- GlcNAc Cycling: A Link Between Metabolism and Chronic Disease. *Annu. Rev. Nutr.* **33,** 205–229 (2013).
27. Ma, X., Li, H., He, Y. & Hao, J. The emerging link between O-GlcNAcylation and neurological disorders. *Cell. Mol. Life Sci.* **74,** 3667–3686 (2017).
28. Zhu, Y., Shan, X., Yuzwa, S. A. & Vocadlo, D. J. The emerging link between O-GlcNAc and Alzheimer disease. *J. Biol. Chem.* **289,** 34472–34481 (2014).
29. Alfaro, J. F. *et al.* Tandem mass spectrometry identi fi es many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets. *PNAS* **109,** 7280–7285 (2012).
30. Myers, S. A., Daou, S., Affar, E. B. & Burlingame, A. Electron transfer dissociation (ETD): The mass spectrometric breakthrough essential for O-GlcNAc protein site assignments-a study of the O-GlcNAcylated protein Host Cell Factor C1. *Proteomics* **13,** 982–991 (2013).
31. Alvarez-Manilla, G. *et al.* Tools for glycomics: Isotopic labeling of glycans with C-13 for relative quantitation. *Glycobiology* **16,** 677–687 (2006).
32. Aoki, K. *et al.* Dynamic developmental elaboration of N-linked glycan complexity in the Drosophila melanogaster embryo. *J. Biol. Chem.* **282,** 9127–9142 (2007).
33. Ma, J., Wu, C. & Hart, G. W. Analytical and Biochemical Perspectives of Protein O-GlcNAcylation. *Chem. Rev.* **121,** 1513–1581 (2021).
34. Cecioni, S. & Vocadlo, D. J. Tools for probing and perturbing O-GlcNAc in cells and in vivo. *Curr. Opin. Chem. Biol.* **17,** 719–728 (2013).

35. Allen, A. K., Neuberger, A. & Sharon, N. The purification, composition and specificity of wheat-germ agglutinin. *Biochem. J.* **131,** 155–162 (1973).

36. Nagata, Y. & Burger, M. M. Wheat Germ Agglutinin: Molecular Characteristics and Specificity for Sugar Binding. *J. Biol. Chem.* **249,** 3116–3122 (1974).

37. Peters, B. P., Goldstein, I. J., Flashner, M. & Ebisu, S. Interaction of Wheat Germ Agglutinin with Sialic Acid†. *Biochemistry* **18,** 5505–5511 (1979).

38. Maynard, J. C. & Chalkley, R. J. Methods for enrichment and assignment of N-acetylglucosamine modification sites. *Mol. Cell. Proteomics* **20,** 1–28 (2021).

39. Zachara, N. E., Vosseller, K. & Hart, G. W. *Detection and analysis of proteins modified by O-Linked N-Acetylglucosamine*. *Current Protocols in Protein Science* **1,** (2011).

40. Orlando, R. *et al.* IDAWG: Metabolic incorporation of stable isotope labels for quantitative glycomics of cultured cells. *J. Proteome Res.* **8,** 3816–3823 (2009).

41. Fang, M., Lim, J.-M. & Wells, L. Quantitative Glycomics of Cultured Cells Using Isotopic Detection of Aminosugars with Glutamine (IDAWG). *Curr. Protoc. Chem. Biol.* **2,** 55–69 (2010).

42. Aoki, K. *et al.* The diversity of O-linked glycans expressed during Drosophila melanogaster development reflects stage- and tissue-specific requirements for cell signaling. *J. Biol. Chem.* **283,** 30385–30400 (2008).

43. Quan, L. & Liu, M. CID , ETD and HCD Fragmentation to Study Protein Post-

Translational Modifications. *Mod. Chem. Appl.* **1,** 1–2 (2013).

CHAPTER 2

QUANTITATIVE APPROACH FOR IDENTIFYING CHANGES IN GLYCOSYLATION, CALCULATING HALF-LIVES OF GLYCANS, AND EVALUATING SIALIC ACID RECYCLING[1]

[1] Chelsea Desbiens, Meng Fang, Jeremy Praissman, Michael Kulik, Sandii Constable, Jae-Min Lim, Alison Nairn, Kelley Moremen, Stephen Dalton, Will York, and Lance Wells
To be submitted to *Glycobiology*

Abstract

The isotopic detection of aminosugars with glutamine (IDAWG) method was first developed to quantify changes in glycans in cell culture upon perturbation. Using Static IDAWG comparative studies can be completed between cell types to quantify changes observed in glycosylation. Initial studies completed here quantified changes seen in human definitive endoderm cells (hDE) and human embryonic stem cells (hES). In our study here we also sought to extend this static IDAWG approach to a Dynamic IDAWG approach to study the half-life of glycans in cell culture. By performing pulse-chase-like experiments with the IDAWG labeling strategy during the cell culture of human embryonic stem cells, we are able to assess the dynamics of turnover for individual O-glycans. During the analyses of the half-life data, we uncovered an isotope pattern that could only be explained by recycling of glycans in cell culture. Our analyses revealed that this recycling was restricted, for the most part, to sialylated glycan structures. We developed bioinformatic tools to interpret the time course data in order to determine the half-life for identified glycans as well as the degree of recycling occurring to make the application and interpretation of data generated from this approach more accessible to the community. Thus, Dynamic IDAWG reveals the half-life of O-glycans in cell culture as well as uncovers the degree of recycling over time for specific structures.

Introduction

In 2009, the methodology of IDAWG, Isotopic Detection of Aminosugars With Glutamine, was reported as the first in-cellulo heavy isotopic labeling strategies for quantitative glycomics.[1] IDAWG takes advantage of the hexosamine biosynthetic pathway, and the necessity for Gln to be supplemented in cell culture. This takes advantage of the side chain of glutamine as the sole donor source of nitrogen for aminosugars in the production of sugar nucleotides. As a result, if cells are fed with media supplemented with glutamine containing a ^{15}N-labeled side chain (heavy Gln), all of the aminosugars including GlcNAc, GalNAc and sialic acid produced in cells will be labeled with ^{15}N where the mass of N- and O-linked glycans, glycolipids, and proteoglycans will all be shifted by +1 Da per labeled amino sugar. The IDAWG methodology has been used to analyze both N- and O-linked glycans released from murine embryonic stem cells successfully and can be applied to various comparative glycomic studies.[2] As an in-vivo isotopic labeling strategy for comparative glycomics, IDAWG demonstrates some significant advantages for comparative glycomics, similar to that of the SILAC technique for proteomics,[3–6] but this technique also provides new opportunities for assessing the dynamics of glycan turnover during the course of any cellular behavior that can be induced or sustained in culture. As compared to previous work with radioactive monosaccharides,[7,8] a standard glycomics workflow using mass spectrometry can be utilized with time-course samples shifted from isotopically-heavy to -light media to determine glycan turnover and new synthesis rates.

Glycans are degraded in the lysosome through families of endo- and exo-

glycosidases that remove individual monosaccharides. In addition to glycoprotein turnover in the lysosome, some glycoproteins can be sent to the trans-Golgi apparatus, where glycan additions may occur, so that they can be rapidly re-displayed on the cell surface.[9,10] It has been suggested that the recycling of glycoproteins occurs at the non-reducing ends of glycans and often involves sialyation that is consistent with the localization of many sialyltransferases to the trans-Golgi.[11–15] There is also some evidence to suggest recycling of core 2 to core 1 O-glycans occurs in addition to the recycling of sialic acid residues.[16,17] Methods for tracking and measuring recycling have been extremely limited but here we find that the Dynamic IDAWG approach that we present here identifies significant recycling occurring on sialylated O-glycans and have developed a workflow for quantifying the degree of recycling from the generated datasets.

Herein, we first wanted to apply the original IDAWG method to an actual biological question, where the original method was defined in a single cell type. Comparative studies were completed to identify quantifiable differences in O-glycans between pluripotent hES cells and differentiated definitive endoderm (hDE) cells. Second, we completed pulse-chase-like experiments with the IDAWG labeling strategy, defining our “Dynamic IDAWG” methodology, during the cell culture of human embryonic stem cells. By culturing cells with heavy-Gln media to complete labeling and then switching to light-Gln media, we are able to harvest and analyze samples over a predetermined time-course to follow turnover and new synthesis of glycans. Permethylated O-glycans released from human embryonic stem cells at the designated time points were analyzed using a linear ion trap

instrument. Using this type of instrument makes the methodology more adoptable overall in comparison to using a high-end Fourier-transform instrument (Orbitrap and ion-cyclotron resonance instruments), and we are still able to assess the dynamics of turnover for individual glycans based on the 50% degradation time of the heavy tag. Comparing the glycan spectra of each sample over time with the 50:50 true mixture spectra of the combined light and heavy samples, an isotope pattern that can only be explained by the occurrence of recycling in the system, primarily involving the terminal sialic acids on identified glycans. We also present an accessible high throughput workflow here for data analysis through the use of bioinformatics tools developed in Excel to aid in the adoption of Dynamic IDAWG. These aid determining the half-life of glycan structures in cells as well as measuring the existence and extent of recycling of glycans that may be occurring.

Results

Quantification of Glycosylation in hED and hES Cells

O-glycan comparative studies were first completed on two cell types, human definitive endoderm cells (hDE) and human embryonic stem cells (hES), in order to identify and quantify changes in glycosylation between the pluripotent and differentiated cells. Here, using Static IDAWG, we investigated and quantified the differences in the glycan profiles of pluripotent and their differentiated counterparts where light- and heavy-Gln labeled cells were grown for both cell types. Samples were analyzed via mass spec, where we were able to identify 24 O-glycan structures in both cell types. Out of the identified glycans we completed comparative quantification studies on the top 8 most abundant structures

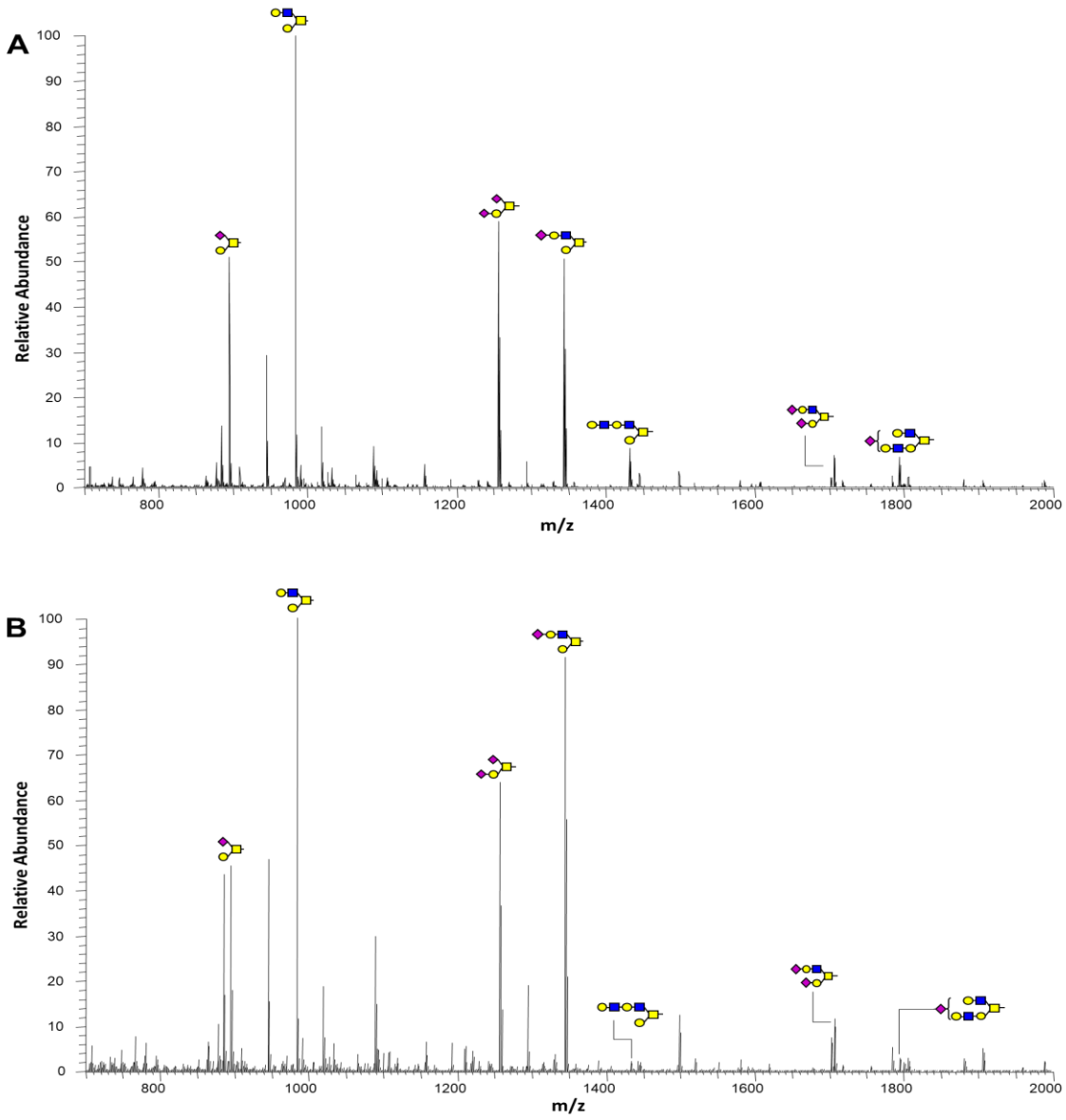


Figure 2.1: Labeled Full MS of Most Abundant O-Glycans in hDE and hES Cells. From the glycans that were identified in both cell types, labeled here are the top 8 most abundant structures (not pictured the HexNAc-Hex structure at 534 m/z) for (A) hES cells and (B) hED cells.

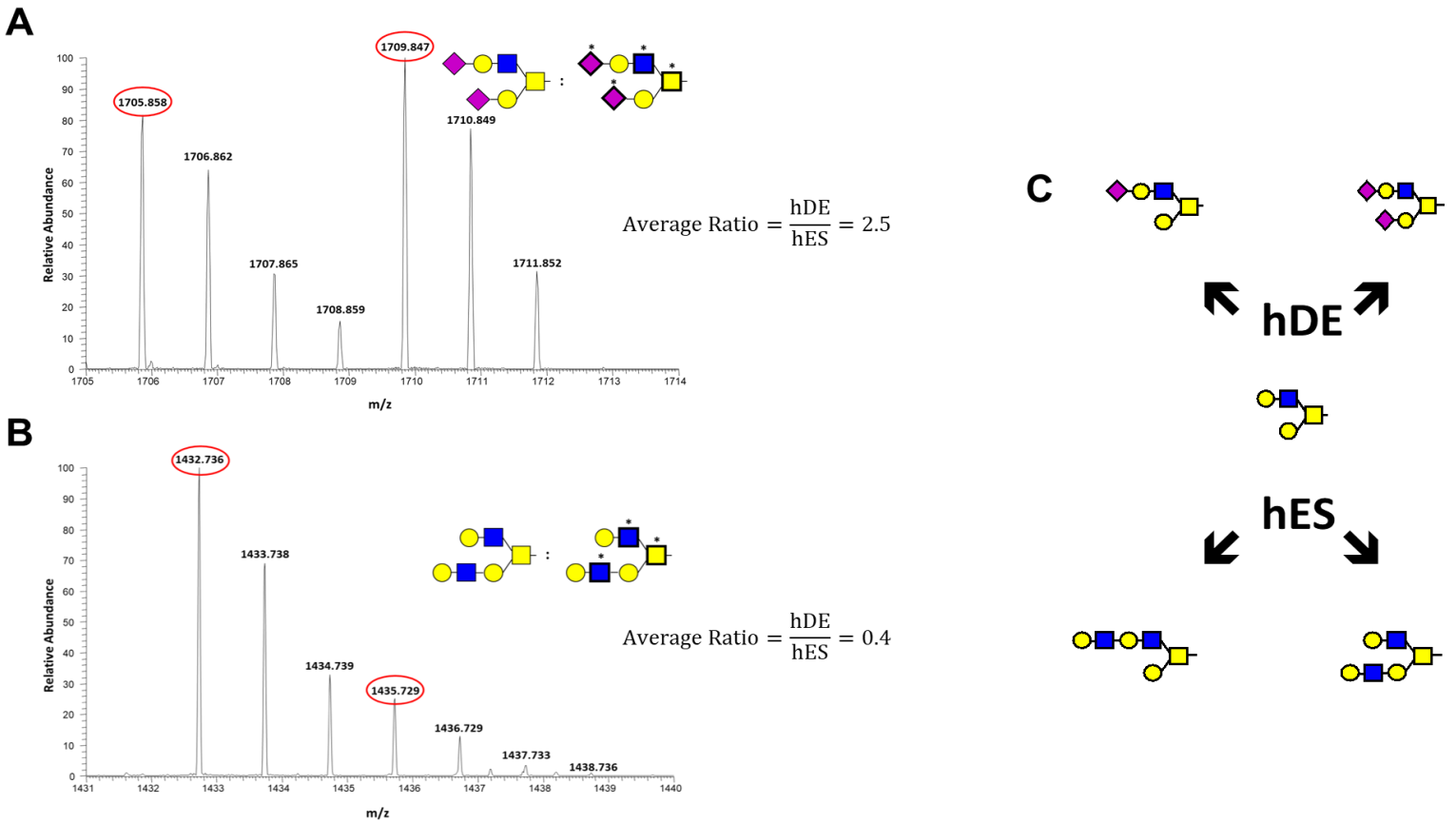
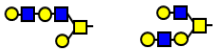
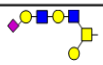


Figure 2.2: Quantification of Glycans in hDE and hES cells. For the quantification of the abundant O-glycans hES cells were labeled with standard ^{14}N-Gln and hDE cells were labeled with ^{15}N-Gln in cell culture. Samples of both could be combined in a 50:50 mixture to quantify and compare the abundance of each glycan. (A) Scan representing the light (1205) and heavy (1209) disialylated structure where the average ratio of hDE:hES was 2.5. (B) Scan representing the light (1432) and heavy (1435) poly-LacNAc structure where the average ratio of hDE:hES was 0.4. (C) Sialylated structures were found to be more abundant in hDE cells and poly-LacNAc structures were more abundant in hES cells.

Table 2.1 Quantification and comparison of O-glycan expression levels in hESCs and hDE

Structure	m/z	Ratio of $\frac{\text{hDE}}{\text{hESC}}$
	534.3	0.8 ± .2
	738.4	1.0 ± .2
	983.5	1.1 ± .2
	1432.7	0.4 ± .2
	895.5	0.9 ± .1
	1256.6	1.1 ± .2
	1344.7	2.6 ± .8
	1705.9	2.5 ± .7
	1793.9	0.6 ± .4

(**Fig. 2.1**). By combining samples from heavy-labeled hDE cells and light-labeled hES cells at a 50:50 ratio, we can observe both light and heavy species simultaneously via mass spec (**Fig. 2.2**). Based on the relative abundances of the observed heavy and light isotopic peak distributions for a given glycan we are able to quantify and compare the presence between the cell types (**Table 2.1**). What we observed was a pattern with core 2 O-glycan structures. In hDE cells core 2 structures, either mono- (1344.7 m/z) or di-sialylated (1705.9 m/z), were more abundant. In contrast, the poly-LacNAc structure (1432 m/z) was more abundant in hES cells (**Fig. 2.2C**). This correlates to the increase in the abundance of ST3 beta-galactoside alpha-2,3-sialyltransferase 1 (ST3GAL1) in hDE cells and increase of beta-1,3-N-Acetylglucosaminyltransferase 3 (B3GNT3) in hES cells.

Pulse-Chase Experiment Designed for Dynamic IDAWG

Through the hexosamine biosynthetic pathway, IDAWG takes advantage of the need for Gln supplementation into cell culture media, utilizing heavy ^{15}N-glutamine to label aminosugars with in cell culture.[1] In order for the original static IDAWG method to be able to assess the dynamics of turnover and synthesis of identified O-glycans, Pulse-chase experiments using human embryonic stem cells (hESC) were performed in order to determine the turnover rates of the hES cell O-glycome. Pulse-chase experiments were completed by growing cells in media supplement with amide-^{15}N-Gln to reach full incorporation of the ^{15}N label. For human embryonic stem cells (BG02 line) we used in our experiment, a 72-hour labeling could yield a ^{15}N incorporation rate above 95% for most glycans (data not

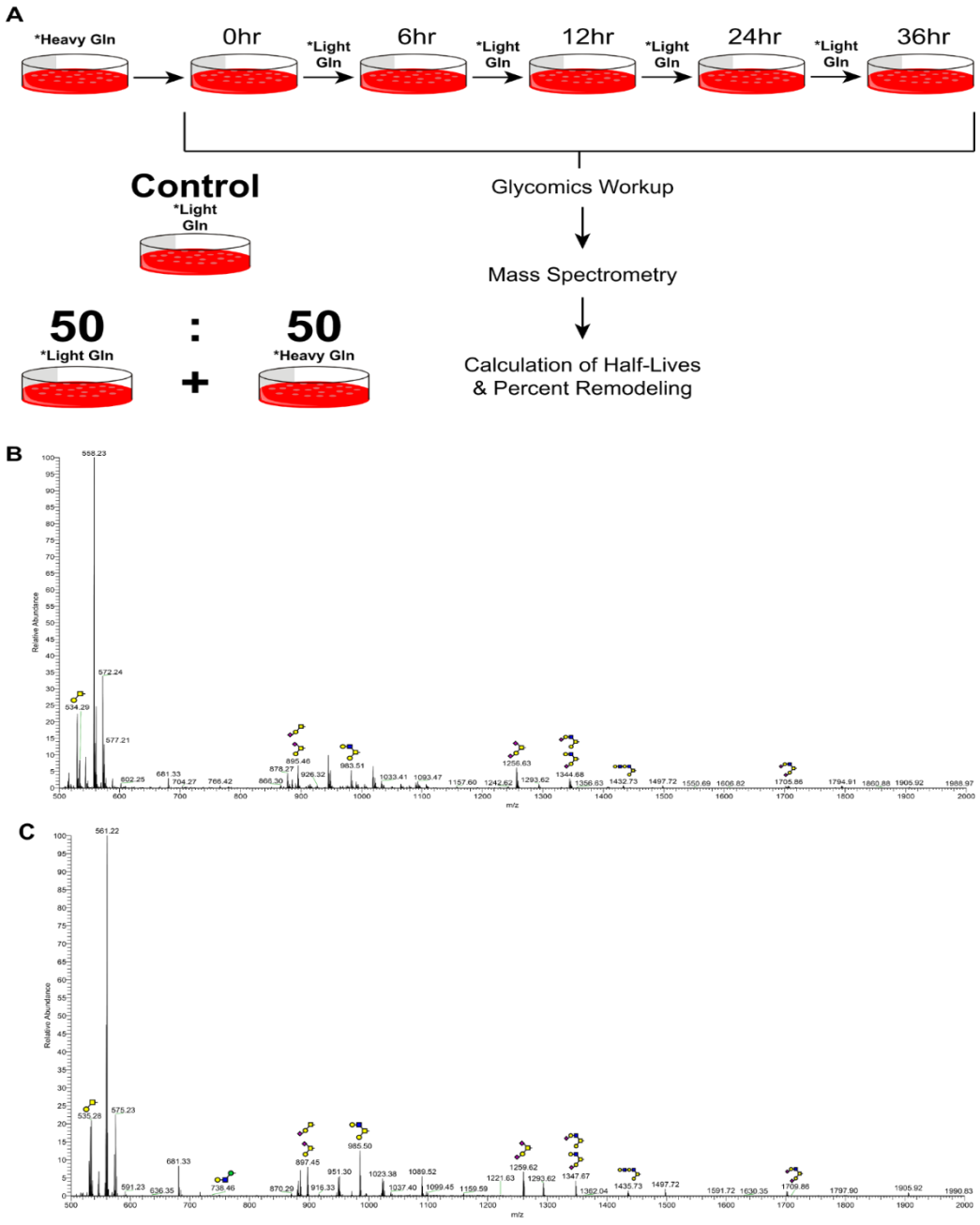


Figure 2.3: Experimental design for Dynamic IDAWG for released glycans. (A) Based upon the original IDAWG method (Orlando et al, 2009), Dynamic IDAWG incorporates the metabolic labeling technique in culmination with pulse-chase experiments and mass spectrometry. Combined these techniques give way to a high throughput method for the determination of glycan turnover rates as well as approximate remodeling. This is accomplished by evaluating the isotopic distributions of the most abundant O-glycans, (B) heavy labeled and (C) light labeled, as well as N-glycans, (D) heavy and (E) light labeled.

shown). Once cells were labeled to completion a subset of cells could be harvested the 0 hour time point as the fully heavy labeled sample. The remaining cells were transferred to new media with normal ^{14}N-glutamine ("light media"). The remaining plates of cells were then harvested at the predetermined time points. In our initial experiments cells were harvested at 0, 6, 12, 24, and 36 hours (**Fig. 2.3A**). A standard O-glycomics workup was completed for each harvested sample, permethylated, and subjected to analysis by Thermo ESI-LTQ-Orbitrap mass spectrometer. The most abundant glycans across all samples were then analyzed (**Fig 2.3B-C**).

Synthesis of Light Structure and Degradation of Heavy Structure

By completely labeling the human embryonic stem cells with heavy Gln and then replacing it with media supplemented with light Gln for the remaining plates after harvesting time 0 plates, spectra can be acquired by running each sample on the Orbitrap. From this, the synthesis and degradation of any amino-sugar-containing glycans can be evaluated by means of the isotopic distributions of each identified structure. The 5 trapping scan spectra of a non-sialylated core 2 structure are displayed in a 7 m/z window exhibiting changes in its isotopic peak distribution over time, at 0 hrs, 6 hrs, 12 hrs, 24 hrs, and 36 hrs as the heavy label is degraded and new glycans are synthesized (**Fig. 2.4; Supplemental Figure 2.1**). With the exception of spectra for 0 hrs, all other spectra show both monoisotopic peaks for both the light and heavy structure. Comparing the relative intensities of the isotopic peaks over time, the trend of the increasing "light distribution" and decreasing "heavy distribution" is clear, suggesting that by replacing the heavy media after the

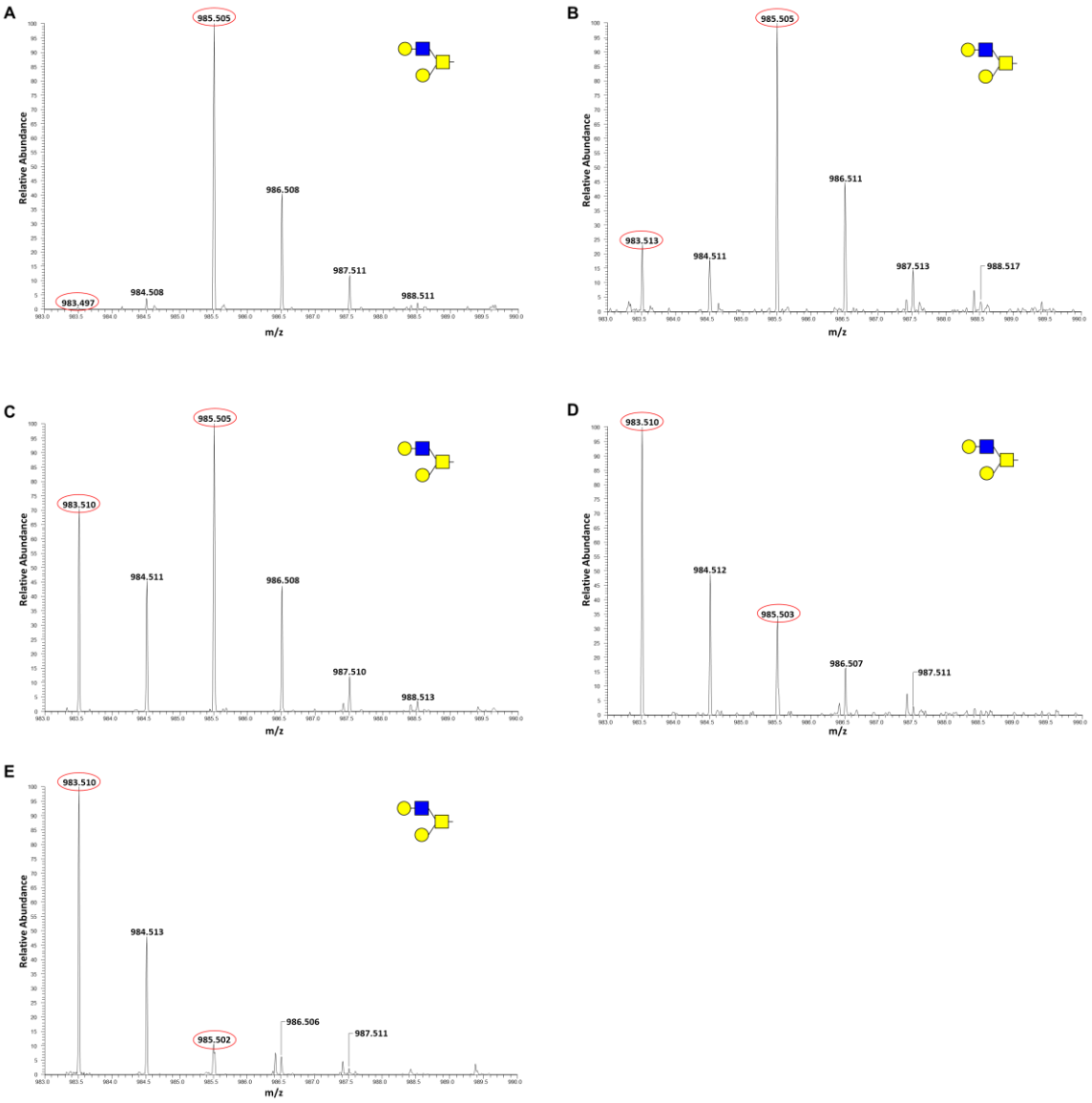


Figure 2.4: Mass spectra of a core 2 non-sialylated O-glycan during pulse chase. Released from hESCs at designated time points. By utilizing pulse chase we are able to observe the transition from the heavy labeled structure at 985 back to the light or unlabeled species at 983 at (A)0hr, (B) 6hr, (C) 12hr, (D) 24hr, and (E) 36hr. At the 12hr time point we can see that the glycan is nearing its half-life where the isotopic patterns for heavy and light are nearly equivalent.

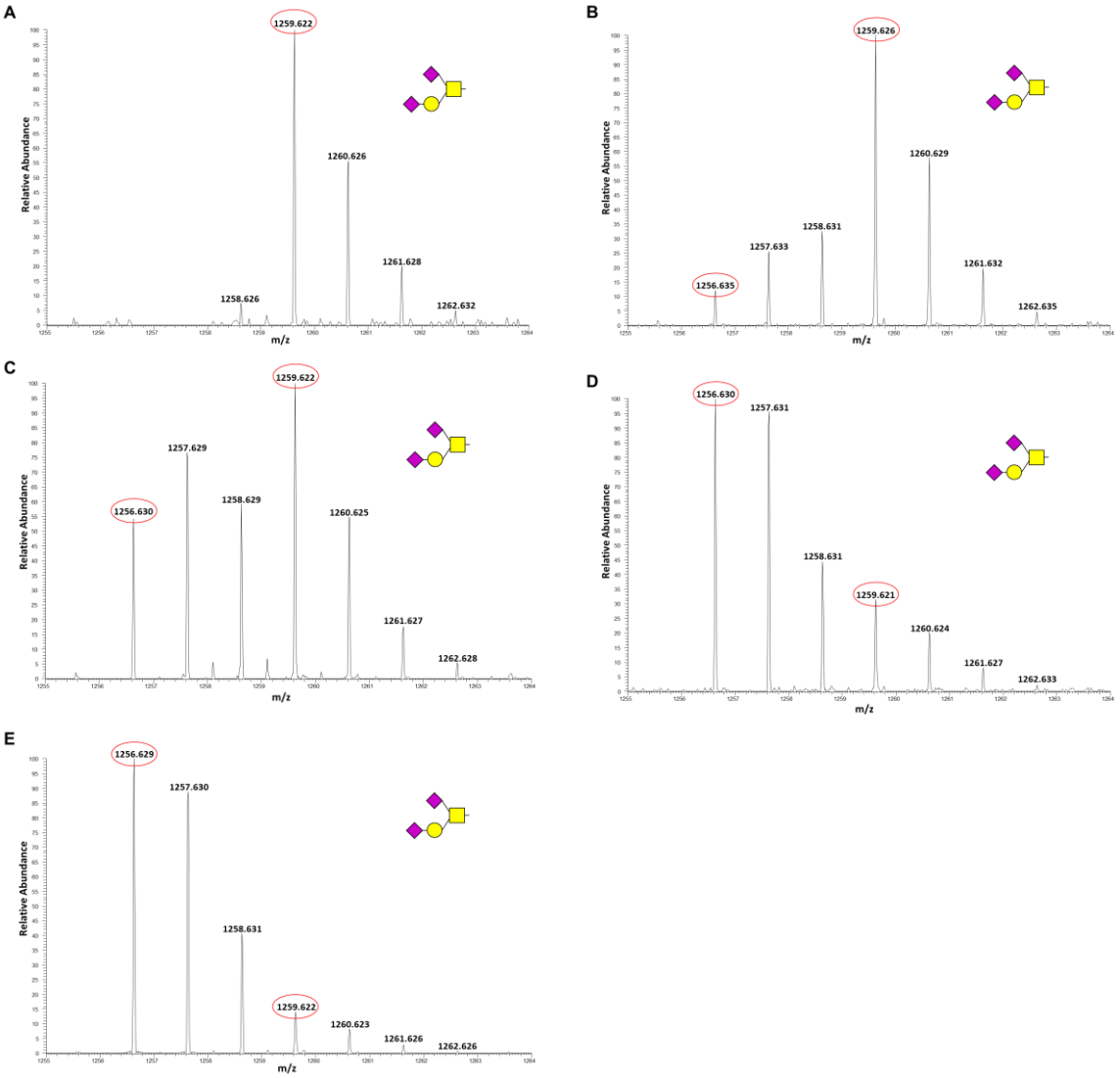


Figure 2.5: Mass spectra of disialylated t-antigen during pulse-chase. Released from hESCs at designated time points for Dynamic IDAWG. The disialylated T-antigen exhibited an unexpected shift back the fully heavy (1259) to fully light species (1256) across the (A)0hr, (B) 6hr, (C) 12hr, (D) 24hr, and (E) 36hr time points. The ratios of the isotopic distribution are skewed most notably at 12hr (C), where the peak at 1257 is more abundant than 1256.

initial time 0 harvest, the glycans containing the heavy aminosugars are degraded, while the cells are synthesizing new glycans incorporating the ^{14}N from newly supplemented light Gln from the media. The same series of spectra was acquired for the disialylated T-antigen, showing a similar trend of changes for both light and heavy structures as the previously described T-antigen (**Fig. 2.5**). However, there was a noticeable difference in the shift back to the light species, particularly at 12hrs, in contrast to the non-sialylated core 2 structure. Where the 1257 (1 ^{15}N) peak is most abundant peak in the “light” distribution and similar patterns were observed in other sialylated glycans that were identified. These results demonstrate that the metabolic labeling of glycans using IDAWG is able to provide new opportunities for assessing the dynamics of turnover and synthesis for individual glycans based on the isotopic patterns as shown in our results.

Recycling of Sialylated Structures

The mass spectrometry analysis of glycans from our pulse-chase experiments has demonstrated that both the synthesis and degradation of glycan structures are occurring simultaneously. However, the metabolic labeling of glycans using IDAWG has provided more interesting information as we compare the spectra for individual glycans identified from our pulse-chase experiments to the spectra collected for our 50:50 true mixture of equal amounts of light and heavy glycans released from differentially labeled human embryonic stem cells. We can observe during pulse-chase, cells are simultaneously degrading the heavy labeled glycan species, while light glycan species are being synthesized. When 50% of the heavy species has been degraded, the isotopic patterns of those glycans should

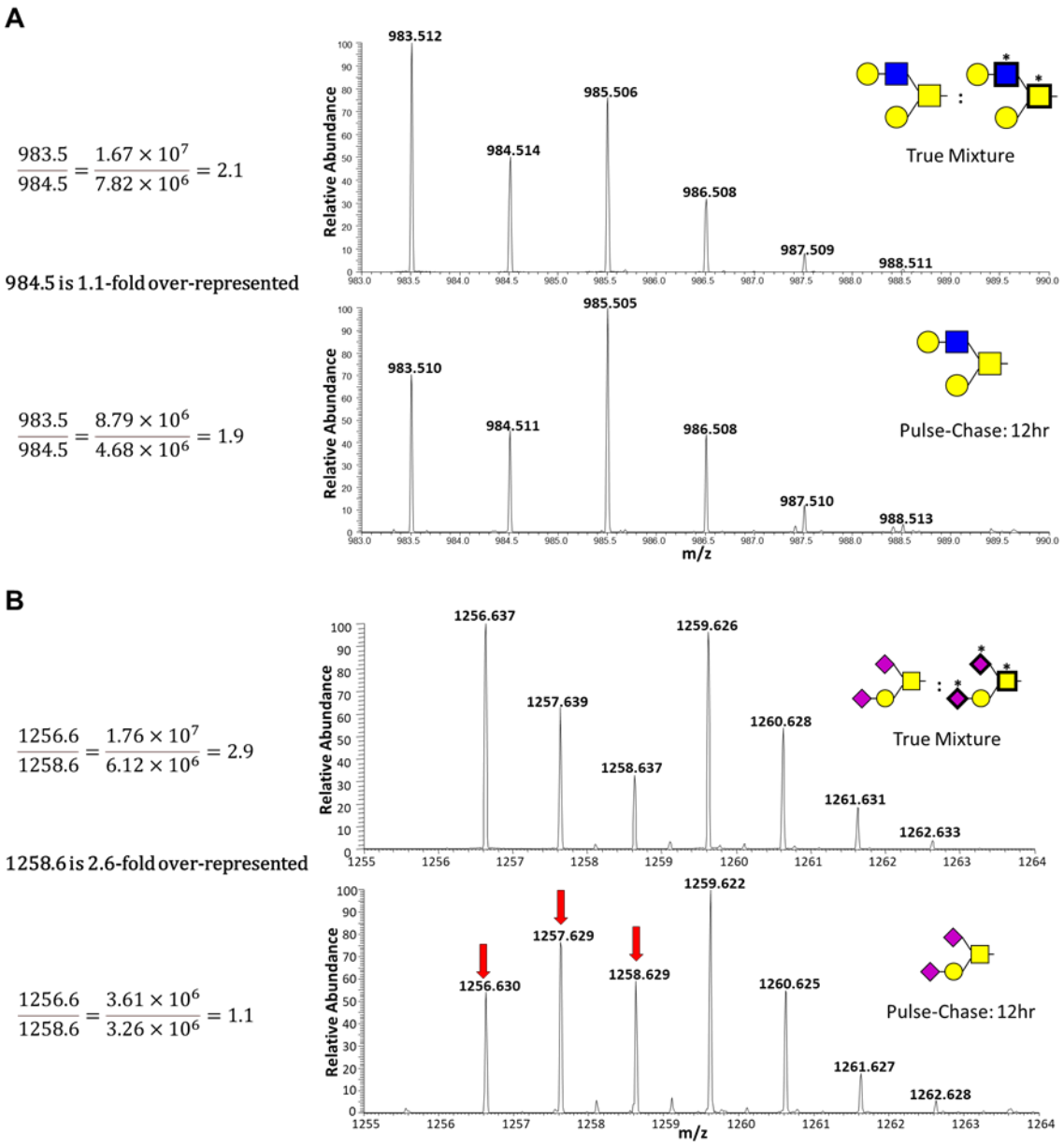


Figure 2.6: Comparison of true mixture of light/heavy structures and Pulse-Chase 12hr sample. The true 50:50 mixture represents a combined sample of equal amounts of heavy and light sample. The resulting mass spectra should represent the expected isotopic distributions at a glycan's given half-life. For the (A) disialylated T-antigen the 50:50 true mixture differs significantly from the actual distribution observed at the 12hr time point. Instead of the 3:1 ratio of intensities for the 1256 and 1258, at 12hrs there exists roughly a 1:1 ratio of the 1256 and 1258 intensities. (B) In contrast, the non-sialylated core 2 glycan exhibits approximately a 2:1 ratio of the 983 and 984 peaks for both the 12hr time point and the 50:50 true mixture.

resemble the isotopic pattern seen in the mass spectra of the true 50:50 mixture of light and heavy samples. Looking at the isotopic peak distributions observed for the core 2, non-sialylated, O-glycan (983.5 m/z) we do observe similar patterns from the 50:50 true mixture compared to the pattern observed at 12hr during the pulse-chase experiments (**Fig. 2.6A**). In contrast, the mass spectra for the disialylated T-antigen during the pulse-chase for the 12hr sample produces a very different isotopic pattern in contrast to the spectrum of the same structure in the 50:50 true mixture of light and heavy labeled glycans (**Fig. 2.6B**). In this spectra, the m/z range from 1256 to 1259 represents the isotopic distribution of the light structure and in the pulse-chase spectrum, the isotopic pattern is abnormal. This is especially noticeable looking at the ratio of the 1256 and 1258 peaks. And the 1257 peak is actually the most abundant within the light distribution. Comparing the ratios of these two peaks, during pulse-chase at 12hrs the peaks are roughly 1:1, when in the true mixture the 1256 peak is nearly 3 times more abundant than its 1258 counterpart. This difference indicates that, unlike a true mixture of light and heavy glycans, the glycan samples from the pulse-chase experiment could contain other types of intermediate structures that induce the abnormal isotopic pattern. Because sialic acid can be removed by sialidases during the life cycle of the molecule once attached to glycoconjugates, one explanation for these isotopic patterns we are observing is that glycans can be detached from their underlying core proteins and returned to the Golgi apparatus. The terminal residues, typically the sialic acids, can be replaced with new sialic acid residues.[18–20] After structures have undergone this recycling process, the glycans will be put back on their core

proteins. As a result of this, the glycans that are released from our pulse-chase samples exist as not only fully light and heavy labeled structures, but also exist as partially heavy, or intermediate, structures. Based on MS^2 spectra collected for the light, intermediate, and heavy labeled disialylated t-antigen, the core GalNAc is found to be light while the terminal sialic acids still contain a ^{15}N tag (**Supplemental Figures 2.2-2.4**). This is taking place due to the salvage pathway recycling heavy labeled sialic acid residues, which can then be reincorporated onto newly synthesized O-glycan cores. The culmination of these three labeled species of a glycan explains the differing isotopic patterns observed in the mass spectra in comparison to the spectra of the true 50:50 mixture of light and heavy structures that is observed.

Automated Modeling for the Glycan Degradation Rate and the Proportion of Recycling by the Quantitative Software

Comparing the mass spectra generated for glycans released from human embryonic stem cells harvested at designated time points in the pulse-chase experiments qualitatively depict the behaviors of glycans in living cells. However, to quantify the rates of glycan synthesis, degradation, and the amount of recycling occurring for an individual glycan, we must rely on a set of quantitative tools built in Microsoft Excel. Three workbooks have been developed as tools to process and analyze the data from the Dynamic IDAWG experiments for identified glycans. Because the synthesis and degradation along with the recycling of glycans is occurring simultaneously in living cells, each pool of released glycans from cells harvested at each time point could consist of 3 components: non-labeled,

completely heavy-labeled, or an intermediate species that is partially labeled with ^{15}N. The percentage of each species can also vary over time. Looking at the disialylated T-antigen as an example, there are three labeled aminosugars in its structure: two sialic acids and one GalNAc. This glycan has the potential to be fully labeled with 3 ^{15}N or not labeled at all; however, it could also contain only 1 or 2 ^{15}N due to recycling.

The bioinformatics tools developed in Microsoft Excel have the capacity to extract the absolute and relative intensities for each glycan's isotopic distribution, average that data across runs, and analyze the composition over time to give the percentage of each species at each time point. Taking into consideration the cell doubling time, which is 30 hours for human embryonic stem cells, as well as the theoretical isotopic distribution for the unlabeled, or light, glycan, the recycling workbook can illustrate a variation trend line of each species for each identified glycan over time (**Fig. 2.7**). The disialylated T-antigen shows a considerable amount of intermediate species being present with only one or two ^{15}N, providing direct evidence recycling of sialic acids occurring, and also explaining the abnormal isotopic patterns observed in certain mass spectra (**Fig. 2.7B**). Analyzing the MS^2 at 12hr gives us additional information on which aminosugar contains the ^{15}N, where the fragment associated with the core GalNAc is exhibits a light labeled profile compared to the shift observed in the sialic acid containing fragment due to the 15N being present (**Fig 2.7C-D**) Using our newly developed tools, we can efficiently and effectively determine the rates of glycan turnover and synthesis as well as the different extents of recycling between sialylated and non-sialylated

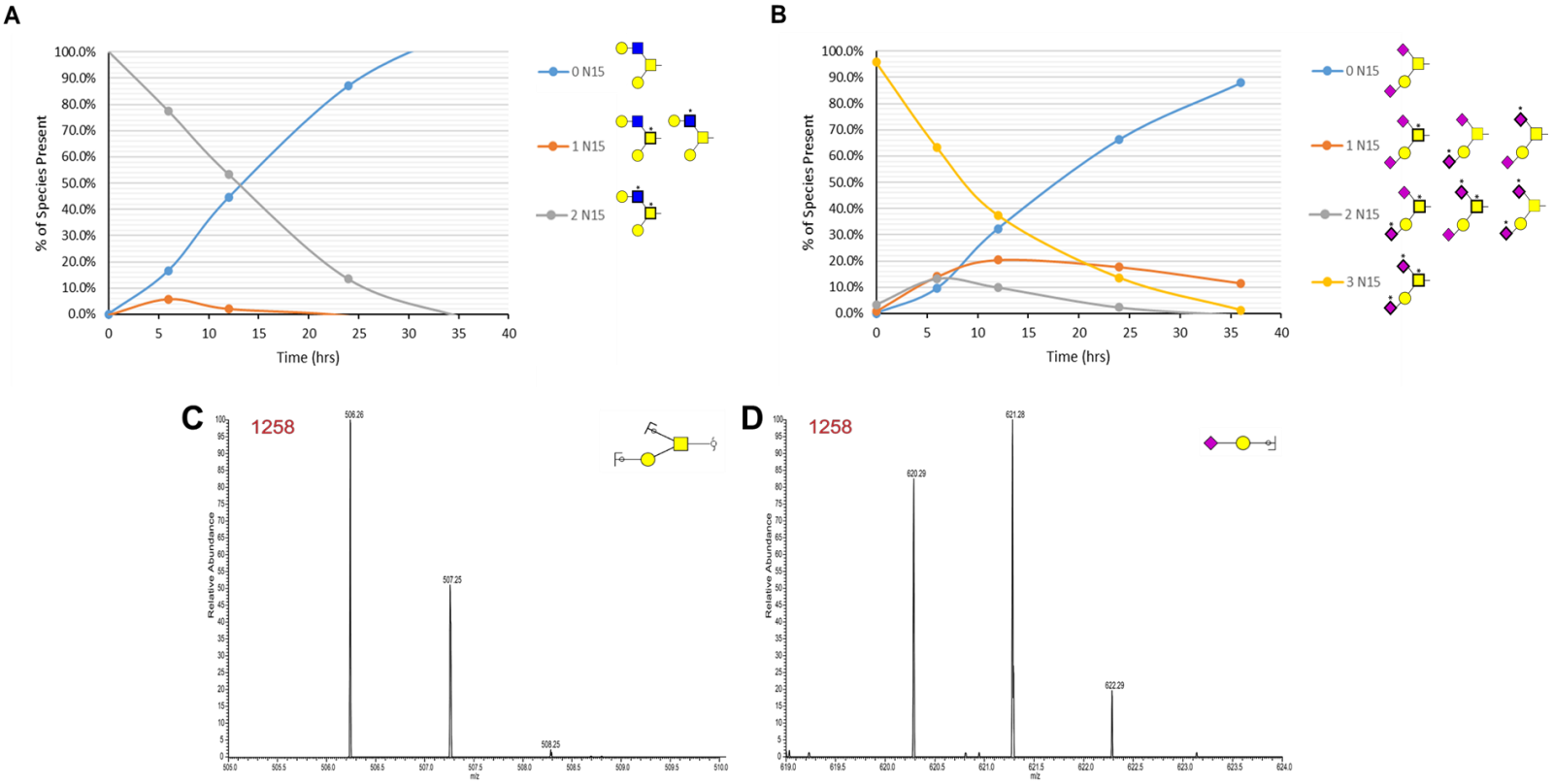
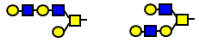
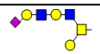


Figure 2.7: Remodeling of both sialylated and non-sialylated structures. Using the set of excel workbooks, the extracted isotopic distributions at each time point are analyzed to determine the approximate amount of remodeling that is occurring for a given structure. The remodeling workbook has the capability to calculate how much of a given species makes up the observed distribution. (A) As expected, the sialylated glycan has a greater amount of remodeling and intermediate structures with only 1 or 2 labeled residues, which explains its isotopic distribution; in contrast to (B) the non-sialylated structure that exhibits little to no remodeling.

Table 2.2: 50% "all heavy" degradation time and proportion of recycling at 50% "all heavy" degradation time for major O-glycans.

Structure	m/z	50% "All Heavy" Degradation Time (hrs)	Proportion of Recycling at 50% Degradation Time (%)	50% New Synthesis Time (hrs)	Difference Between 50% New Synthesis and 50% "All Heavy" (hrs)
	983 (+1)	13.0 ± .03	0.1 ± .007	13.5 ± .04	0.5
	534 (+1)	11.5 ± .05	0.0 ± 0.0	11.5 ± .05	0.0
	738 (+1)	7.2 ± .1	0.0 ± 0.0	7.2 ± .09	0.0
	1432 (+1)	12.8 ± .03	6.9 ± .007	15.3 ± .04	2.5
	1256 (+1)	10.3 ± .04	21.9 ± .005	18.5 ± .03	8.2
	1344 (+1)	16.2 ± .04	10.4 ± .005	20.3 ± .03	4.1
	895 (+1)	12.0 ± .04	7.6 ± .005	14.7 ± .04	2.7
	1705 (+1)	16.3 ± .03	31.8 ± .005	25.8 ± .02	9.5
	1793 (+1)	14.7 ± .03	22.4 ± .01	21.2 ± .04	6.5

structures in the most abundant O-glycans in our samples. The designated half-life for each glycan is based upon the amount of structure containing only heavy nitrogen when it has decreased to 50% of its original amount at 0 hour, or 50% “all heavy” degradation time, and is listed in **Table 2.2** for the most abundant (≥2x10^5 intensity) O-glycans. The proportion of recycling at the 50% degradation time for each glycan has also been calculated for each structure. Sialylated structures demonstrated a relatively higher percentage of recycling overall, compared to the non-sialylated glycans we analyzed. This further suggests that sialic acid has much to do with the phenomenon that was observed.

Discussion

Here, initial comparative quantification of glycosylation of hDE and hES cells was performed. These experiments successfully quantified 8 of the most abundant O-glycans and identified a pattern between cell types determining changes in O-glycosylation between pluripotent and differentiated cells. hDE cells have a higher abundance of extended core 2 structures with sialic acid while hES ells favored poly-LacNAc structures. This demonstrated that the original static IDAWG method can effectively be applied in differing cell types, where original studies were solely completed in the same cell type.

Pulse-chase experiments combined with the IDAWG labeling strategy during the cell culture of human embryonic stem cells were performed to assess the dynamics of turnover for individual glycans and define the Dynamic IDAWG method for released glycans. By fitting the mass spectra acquired for a series of samples using bioinformatics tools developed in a set of Excel workbooks we were

able to investigate the turnover rates for abundant O-linked glycans. By comparing the glycan spectra of pulse-chase samples with the spectra of real 50:50 mixtures of light and heavy sample, we noticed an isotope pattern that could only be explained by recycling of aminosugars, mostly sialic acid, where intermediate species of a structure were present in addition to fully labeled and non-labeled species. Sialylated structures were undergoing a relatively higher extent of recycling compared to non-sialylated glycans, suggesting sialic acid residues have much to do with the phenomenon observed with the help of the sialic acid salvaging pathway.

Determining the half-life, or 50% “all heavy” degradation time, and approximating the amount of recycling occurring for a subset of structures was accomplished using the set of bioinformatics tools developed in Microsoft Excel: an Extraction Workbook, an Averaging Workbook, and a Half-Life/Recycling Workbook. The combination of this Dynamic IDAWG method and these bioinformatics tools, addresses the need for an easily accessible and high throughput method to study the turnover of present glycans under a multitude of conditions without incorporating previously used radiolabeling techniques. Simultaneously, further investigations into recycling of glycan structures, specifically in glycans containing sialic acid, can be accomplished.

Materials and Methods

Cell Culture and Pulse-Chase Experiments

The BG02 line of hESCs were cultured as previously described.[21] Briefly, the hESCs were grown on gelatin coated petri dishes at 37ºC under 10% CO_2. The

culture media was composed of Dulbecco's modified Eagle's medium (DMEM, Invitrogen) supplemented with dialyzed fetal calf serum (FCS, Commonwealth Serum Laboratories), 1,000U/ml recombinant murine leukemia inhibitory factor (LIF) (ESGRO, Chemicon International) and 2 mM L-glutamine which is either normal abundance ^{14}N-isotope or amide-^{15}N-Gln (98% purity, Cambridge Isotopes, Inc). The media was changed daily and cells were allowed to grow for 72 hours at which point a subset of plates were harvested by dissociation buffer and scraping. The remaining cells were transferred to light Gln supplemented media with normal abundance amide-^{14}N-Gln and continued to grow for 6, 12, 24 and 36 hours respectively. The harvested cell pellets were washed with phosphate buffered saline (PBS) and stored at -80˚C until analysis.

Preparation of Protein Powder, Delipidation, Glycan Release and Permethylation

The isolation of permethylated O-linked glycans was completed as previously described.[1,2,22–24] Briefly, the cell pellets were disrupted on ice by Dounce homogenization and delipidated using chloroform/methanol/water with a ratio of 4:8:3 as organic extraction. After washed with cold water and cold acetone for three times, the final pellets of insoluble protein were dried under a stream of nitrogen. The resulting protein powders were treated with alkaline borohydride to release O-linked glycans by β-elimination. Released glycans were permethylated with methyl iodide, cleaned up by reverse-phase columns, and dried down.

MS Analysis of Permethylated Glycans

The glycans were analyzed as previously described on a hybrid linear ion trap Orbitrap mass spectrometer (LTQ-Orbitrap, Thermo).[1,2] Briefly, permethylated glycans were dissolved in 15 µL of 100% methanol followed by the addition of 35 µL of 1 mM NaOH in 50% methanol to make a total of 50 µL of solution and infused directly into the mass spectrometer using a nanospray ion source with a fused-silica emitter (360 × 75 × 30 µm, SilicaTip™, New Objective) at a syringe flow rate of 0.4 µL/min. The capillary temperature was set to 200°C and MS analysis was performed in positive ion mode. Full FTMS (Fourier Transform mass spectrometry) spectra in profile mode were collected at 400-2000 m/z for 30 sec with 5 microscans and 1000 maximum injection time (ms) and resolution was set to be 60,000. For fragmentation by collision-induced dissociation (CID) the centroid MS/MS spectra were acquired from 400 to 2000 m/z at 36% normalized collision energy, 0.25 activation Q, and 30.0 ms activation time by total ion mapping. Parent mass step size and isolation width was set at 2.0 m/z and 2.8 m/z respectively for automated MS/MS spectra with TIM scans. All glycan structures were manually interpreted based on in-house fragmentation rules and the GlycoWorkbench software.[25]

Determination of Half-Life and Percent Recycling

A set of bioinformatics tools were developed in Microsoft Excel, to help streamline the data analysis process. These included an extraction workbook, averaging workbooks for technical and biological replicates, and a final workbook

to calculate the half-life of individual glycans and the percent recycling. Full MS-scans were first averaged for each individual time point and the same was done across biological replicates. The averaged spectral list and corresponding peak intensities were then extracted for each and inputted into the Extraction Workbook. The workbook is set up to extract the isotopic envelope for designated glycan m/z values and can account for various charge states. This provides all glycan m/z values within the envelope, the absolute intensities, and relative intensities based on that particular envelope. By supplying the file location for the extraction data, the data can then be pulled into the averaging workbooks. Data for all technical replicates of a glycan can then be averaged across the collected time points for a single biological replicate. Additionally, each peak is normalized, and standard deviations are calculated based on the supplied data. Similarly, the biological replicate averaging workbook will pull the averaged data from all biological replicates into a single file. This final set of isotopic envelopes based on all replicates, outputs a final envelope for each time point for a given glycan of interest. These values are then pulled into the final workbook to calculate half-life and percent recycling. Calculations in the final half-life/recycling workbook are based upon the empirical or theoretical isotopic envelope of the light labeled species and the isotopic distributions for each time point as glycans transition from a fully heavy labeled to fully light. These data are pulled in from the biological replicates averaging workbook. Taking these isotopic envelopes into account, how many ^{15}N are incorporated into each glycan, and the doubling time of the cell line used (30hr was used for hES cells in these studies), the contributions of each present species

to the observed isotopic distribution can be determined over time. Half-life for these studies is defined as the time it takes to reach 50% “all heavy” degradation and the proportion of recycling is set at the time a glycan reaches said half-life.

References

1. Orlando, R. *et al.* IDAWG: Metabolic incorporation of stable isotope labels for quantitative glycomics of cultured cells. *J. Proteome Res.* **8,** 3816–3823 (2009).

2. Fang, M., Lim, J.-M. & Wells, L. Quantitative Glycomics of Cultured Cells Using Isotopic Detection of Aminosugars with Glutamine (IDAWG). *Curr. Protoc. Chem. Biol.* **2,** 55–69 (2010).

3. Ong, S. E. The expanding field of SILAC. *Anal. Bioanal. Chem.* **404,** 967–976 (2012).

4. Ong, S. E., Mittler, G. & Mann, M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat. Methods* **1,** 119–126 (2004).

5. Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1,** 376–386 (2002).

6. Ong, S. E. & Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat. Protoc.* **1,** 2650–2660 (2007).

7. Diaz, S. & Varki, A. Metabolic Radiolabeling of Animal Cell Glycoconjugates. *Curr. Protoc. Protein Sci.* **57,** 1–15 (2009).

8. Snider, M. D. Metabolic Labeling of Glycoproteins with Radioactive Sugars.

Curr. Protoc. Cell Biol. **13,** 1–11 (2002).

9. Green, S. A. & Kelly, R. B. Low density lipoprotein receptor and cation-independent mannose 6- phosphate receptor are transported from the cell surface to the Golgi apparatus at equal rates in PC12 cells. *J. Cell Biol.* **117,** 47–55 (1992).

10. Snider, M. D. Intersection of endoytic and exocytic membrane traffic in animal cells. in *Intracellular Trfficking of Proteins* (eds. Steer, C. J. & Hanover, J. A.) 361–386 (Cambridge University Press, 1991).

11. Michalski, J.-C. Normal and pathological catabolism of glycoproteins. in *Glycoproteins and Disease* (eds. Montreuil, J., Vliegenthart, J. F. G. & Schachter, H.) 55–97 (Elsevier Science B.V., 1996). doi:10.1201/9780203742815

12. Winchester, B. G. Lysosomal Metabolism of Glycoconjugates. in *Subcellular Biochemistry* (eds. LLoyd & Mason) **27,** 191–238 (Plenum Press, 1996).

13. Winchester, B. Lysosomal metabolism of glycoproteins. *Glycobiology* **15,** 1–15 (2005).

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

15. Varki, A. *et al. Essentials of Glycobiology*. (Cold Spring Harbor Laboratory Press, 2017).

16. Razawi, H. *et al.* Evidence for core 2 to core 1 O-glycan remodeling during the recycling of MUC1. *Glycobiology* **23,** 935–945 (2013).

17. Gilormini, P. A. *et al.* Chemical glycomics enrichment: imaging the recycling of sialic acid in living cells. *J. Inherit. Metab. Dis.* **41,** 515–523 (2018).

18. Schenkman, S., Jiang, M. S., Hart, G. W. & Nussenzweig, V. A novel cell surface trans-sialidase of trypanosoma cruzi generates a stage-specific epitope required for invasion of mammalian cells. *Cell* **65,** 1117–1125 (1991).

19. Roggentin, P., Schauer, R., Hoyer, L. L. & Vimr, E. R. The sialidase superfamily and its spread by horizontal gene transfer. *Mol. Microbiol.* **9,** 915–921 (1993).

20. Taylor, N. R. & Von Itzstein, M. A structural and energetics analysis of the binding of a series of N-acetylneuraminic-acid-based inhibitors to influenza virus sialidase. *J. Comput. Aided. Mol. Des.* **10,** 233–246 (1996).

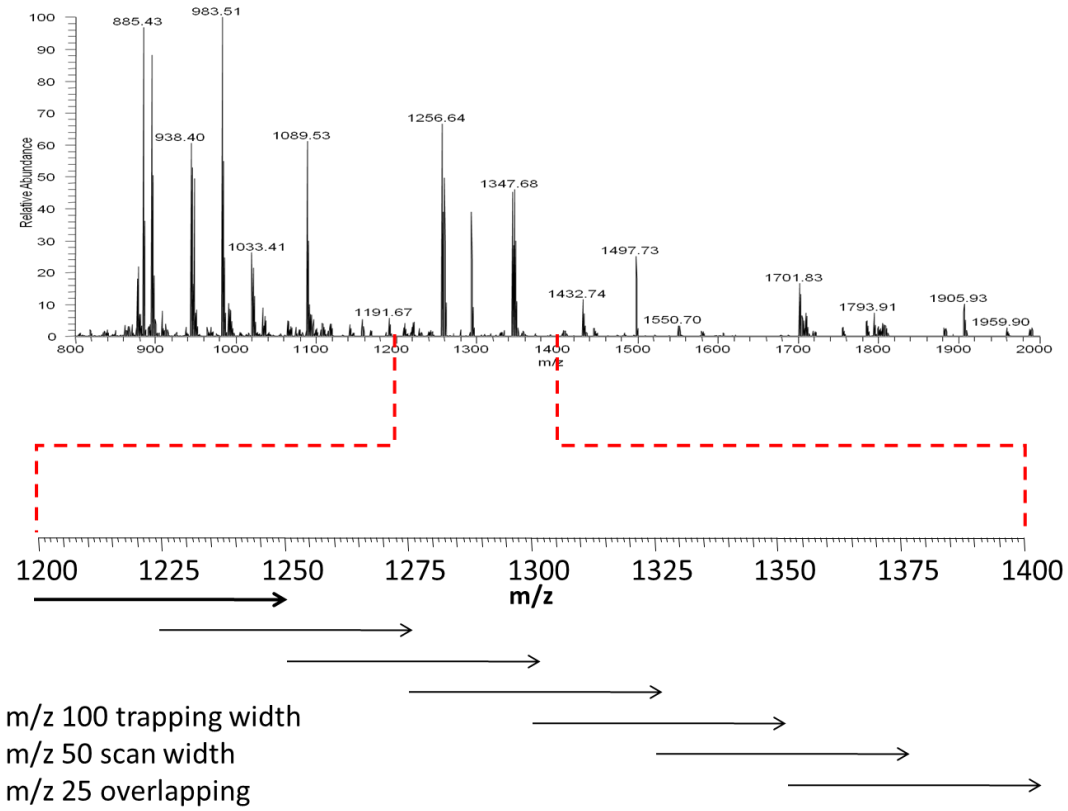
21. McLean, A. B. *et al.* Activin A Efficiently Specifies Definitive Endoderm from Human Embryonic Stem Cells Only When Phosphatidylinositol 3-Kinase Signaling Is Suppressed. *Stem Cells* **25,** 29–38 (2007).

22. Aoki, K. *et al.* Dynamic developmental elaboration of N-linked glycan complexity in the Drosophila melanogaster embryo. *J. Biol. Chem.* **282,** 9127–9142 (2007).

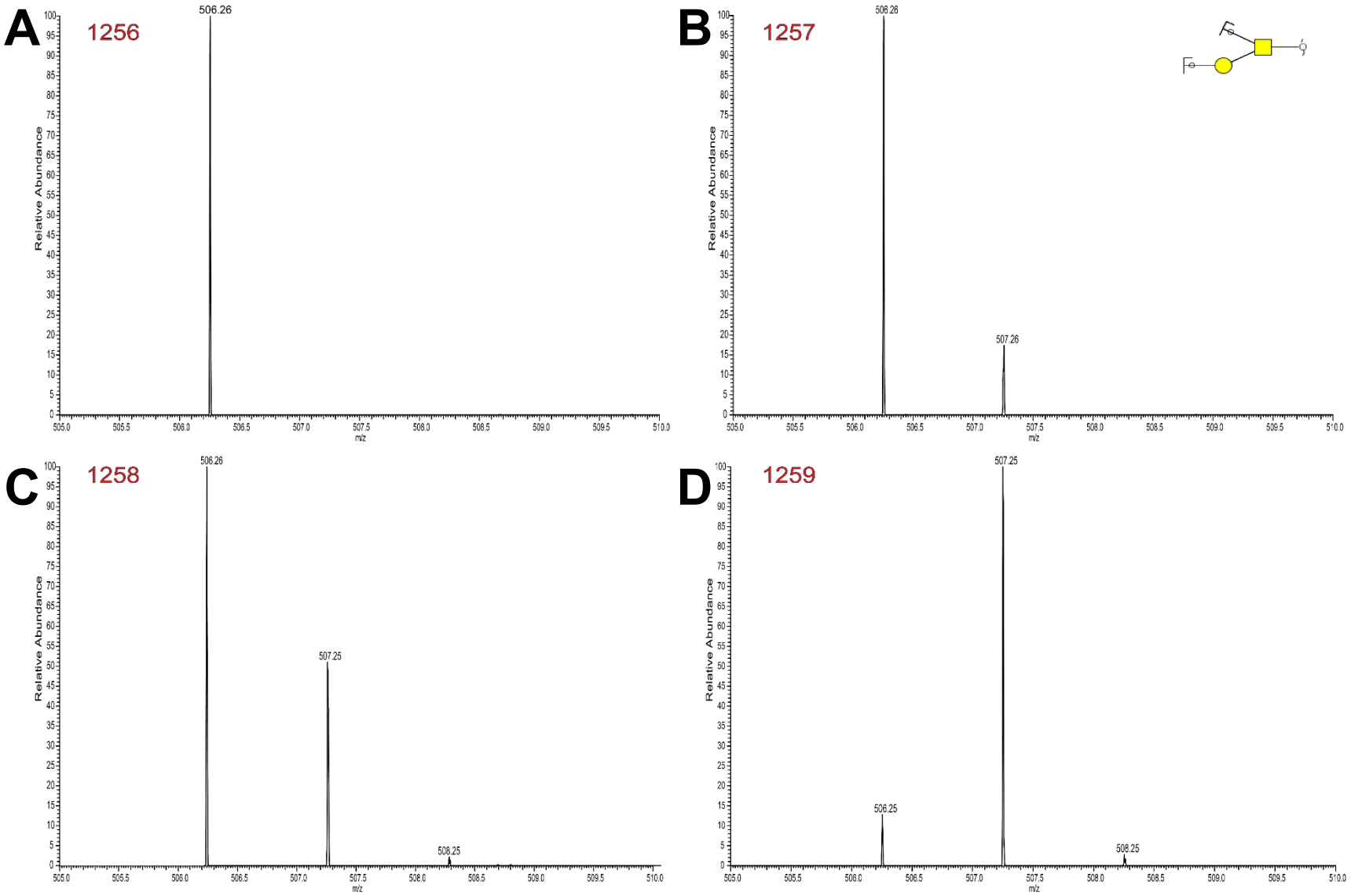
23. Aoki, K. *et al.* The diversity of O-linked glycans expressed during Drosophila melanogaster development reflects stage- and tissue-specific requirements for cell signaling. *J. Biol. Chem.* **283,** 30385–30400 (2008).

24. Kurz, S., Sheikh, M. O., Lu, S., Wells, L. & Tiemeyer, M. Separation and identification of permethylated glycan isomers by reversed phase NanoLCNSI-MSn. *Mol. Cell. Proteomics* **20,** 100045 (2021).

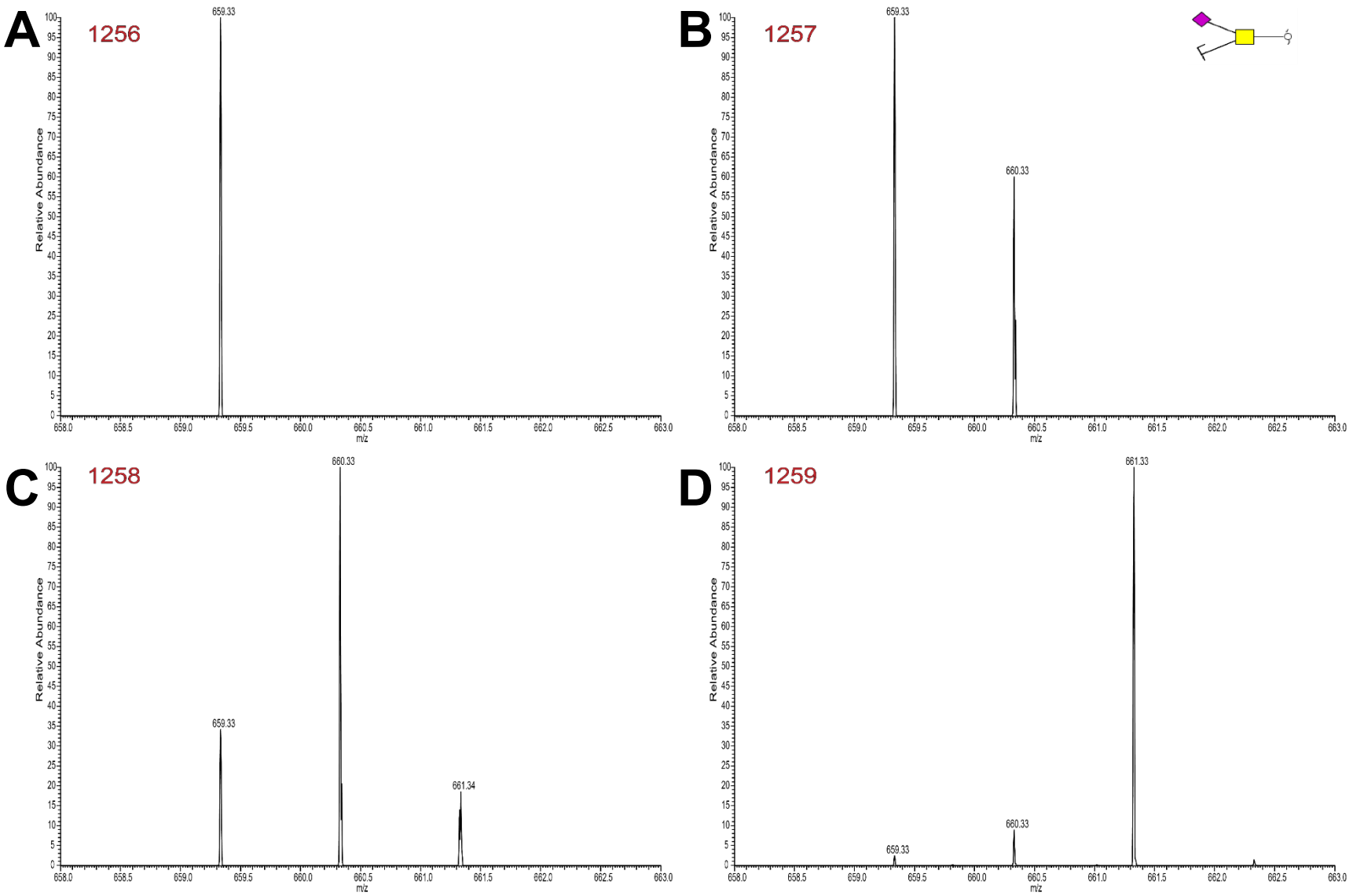
25. Ceroni, A. *et al.* GlycoWorkbench: A Tool for the Computer-Assisted Annotation of Mass Spectra of Glycans. *J. Proteome Res.* **7,** 1650–1659 (2008).



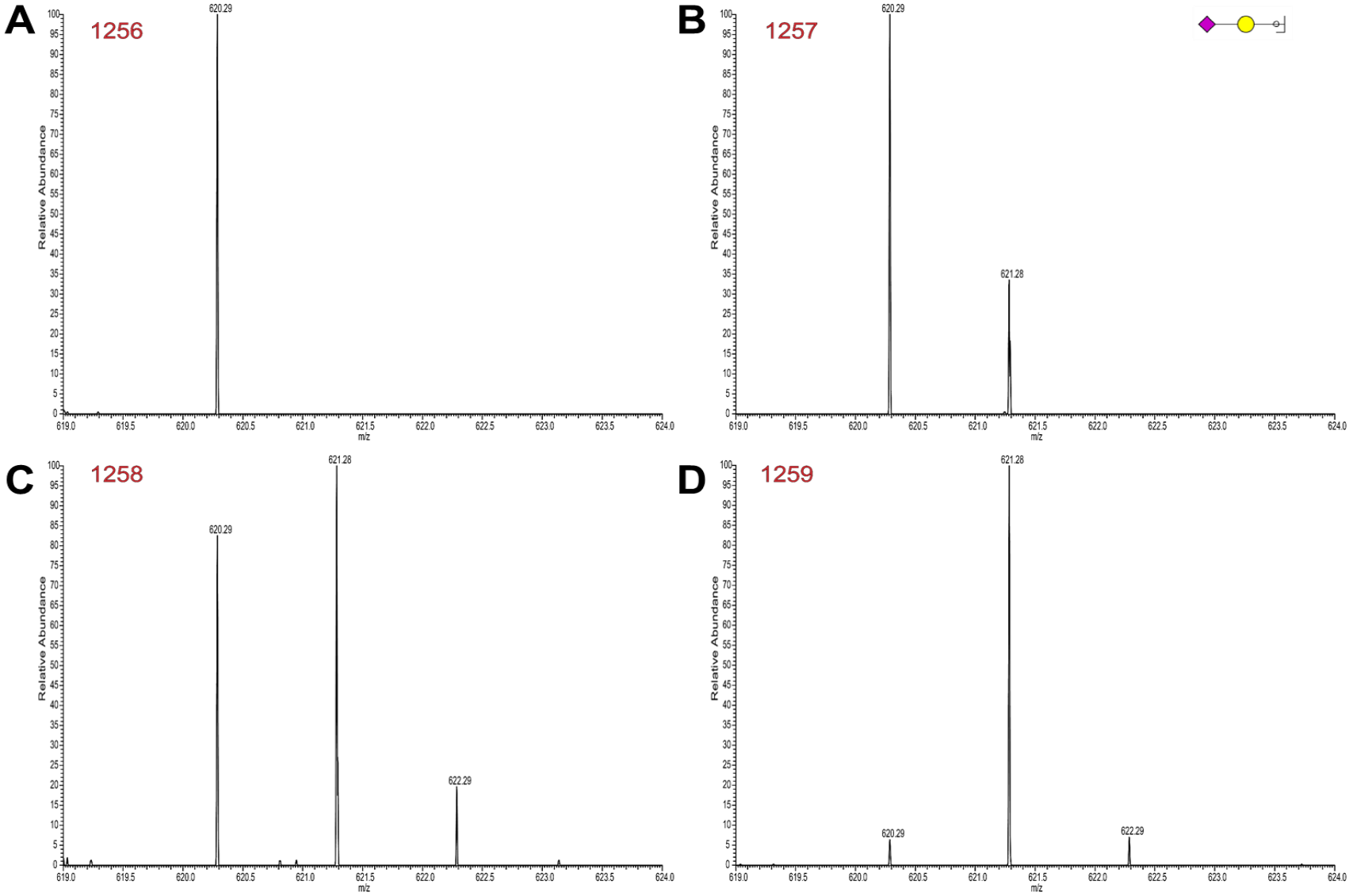
Supplemental Figure 2.1: Diagram for Mass Spectrometry Method. Rolling trapping scans were collected without collision energy, where scans were 50 m/z in width and overlapped with the following scan by 25 m/z


A
1256
506.26
Relative Abundance
m/z
B
1257
506.26
507.26
Relative Abundance
m/z
C
1258
506.26
507.25
508.25
Relative Abundance
m/z
D
1259
507.25
506.25
508.25
Relative Abundance
m/z

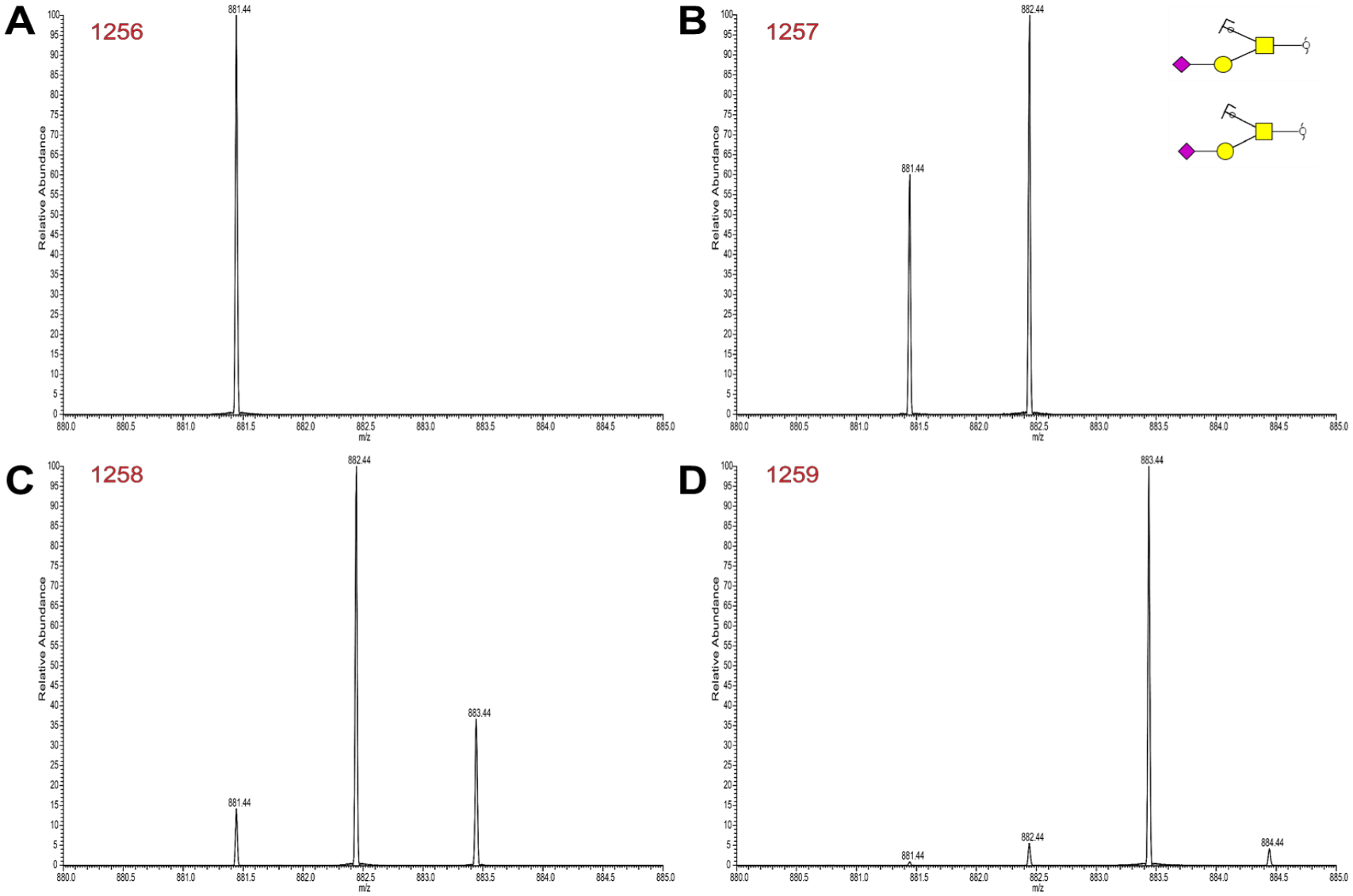
Supplemental Figure 2.2: MS2 scans of the 506 m/z fragment ion of the disialylated t-antigen. Collected at 12hrs for the disialylated t-antigen fragment at 506 m/z for (A) the fully light (1256 m/z), (B-C) the partially labeled intermediate species (1257 & 1258 m/z), and (D) the fully heavy labeled structure (1259 m/z). The most abundant peak for the intermediate species is the light or 506 peak, implying the core GalNAc is light.


A
1256
659.33
B
1257
659.33
660.33
C
1258
660.33
659.33
661.34
D
1259
661.33
660.33
659.33
Relative Abundance
m/z

Supplemental Figure 2.3: MS2 scans of the 659 m/z fragment ion of the disialylated t-antigen. Collected at 12hrs for the disialylated t-antigen fragment at 659 m/z for (A) the fully light (1256 m/z), (B-C) the partially labeled intermediate species (1257 & 1258 m/z), and (D) the fully heavy labeled structure (1259 m/z). Looking at the intermediate species 1258 m/z, the peak at 660 m/z is most abundant and we can conclude that only the sialic acid or the core GalNAc can be labeled and not both.


A
1256
620.29
B
1257
620.29
621.28
C
1258
621.28
620.29
622.29
D
1259
621.28
620.29
622.29
Relative Abundance
m/z

Supplemental Figure 2.4: MS2 scans of the 620 m/z fragment ion of the disialylated t-antigen. Collected at 12hrs for the disialylated t-antigen fragment at 620 m/z for (A) the fully light (1256 m/z), (B-C) the partially labeled intermediate species (1257 & 1258 m/z), and (D) the fully heavy labeled structure (1259 m/z). We see an 85:100:20 ratio represented by the 1258 m/z species. While the fragment containing a labeled sialic acid is the most abundant (621), there is still a significant amount of this fragment that is unlabeled (620) suggesting recycling is occurring.


A
1256
881.44
Relative Abundance
m/z
B
1257
882.44
881.44
Relative Abundance
m/z
C
1258
882.44
883.44
881.44
Relative Abundance
m/z
D
1259
883.44
882.44
884.44
881.44
Relative Abundance
m/z

Supplemental Figure 2.5: MS2 scans of the 881 m/z fragment ion of the disialylated t-antigen. Collected at 12hrs for the disialylated t-antigen fragment at 881 m/z for (A) the fully light (1256 m/z), (B-C) the partially labeled intermediate species (1257 & 1258 m/z), and (D) the fully heavy labeled structure (1259 m/z). Two of the three possible residues than can be modified are accounted for in the 881 m/z fragment. However, the fragmentation pattern for 1258 m/z suggests most of this glycan only contains 1 ^{15}N by 12 hours.

CHAPTER 3

ASSESSING THE DYNAMICS OF O-GLCNAC MODIFICATIONS FROM HUMAN ES CELLS USING DYNAMIC IDAWG[2]

[2] Chelsea Desbiens, Jeremy Praissman, and Lance Wells

To be submitted to *Glycobiology*

Abstract

It has been observed that the post translational modification (PTM), O-linked β-N-acetylglucosamine (O-GlcNAc), exists as a single non-extended monosaccharide that is both a labile and an inducible modification on thousands of nuclear and cytosolic proteins. Cycling on and off serine and threonine residues with the assistance of O-GlcNAc transferase (OGT) and O-GlcNAcase (OGA) respectively, it has been hypothesized that O-GlcNAc exists as a dynamic protein modification. As a dynamic modification, O-GlcNAc would exhibit a shorter turnover rate than that of the protein in which it modifies, where dynamics would vary by modification site, as well as what protein is modified. Currently, the field lacks high-throughput methods that are sensitive enough to evaluate the dynamics of both O-GlcNAc and the modified proteins to determine dynamics. Due to the laborious and insensitive methods that are available, only a handful of O-GlcNAc modified proteins have been evaluated. Previously, the method Dynamic IDAWG was formulated to evaluate half-lives and percent recycling occurring for released O-linked glycans. Herein, we describe an adapted take on Dynamic IDAWG to study intact O-GlcNAc modifications and the subsequent proteins to evaluate dynamics at individual modified sites.

Introduction

Discovered in the early 1980s, the regulatory post-translational modification, O-linked β-N-acetylglucosamine (O-GlcNAc) has been found to modify thousands of proteins on serine and threonine residues.[1,2] It exists as a product of the hexosamine biosynthetic pathway (HBP), which is driven by the presence of glucose and glutamine within a system. [3–5] Interestingly, this particular modification exists as a unique non-extended monosaccharide in contrast to other glycan structures and is found on nuclear and cytosolic proteins rather than on the cell surface.[6] It has also been found to be dynamic, inducible, and modulated by extracellular stimuli. The dynamic nature of this modification is due to the collaboration of two enzymes: the nucleocytoplasmic enzyme O-GlcNAc transferase (OGT) which is responsible for the addition of O-GlcNAc onto proteins, while O-GlcNAcase (OGA) is responsible for its removal (**Fig. 3.1**).[5,7,8] The regulatory nature and cycling of the modification are analogous to phosphorylation, where phosphate cycles with the help of kinases and phosphatases. It has been shown that there is cross-talk between O-GlcNAcylation and phosphorylation modifications that aid in regulating cellular functions.[9–12]

Current methods that allow for investigation into the O-GlcNAc modification center around site mapping, proteomic studies, and quantification. For example, stable isotope labeling with amino acids in cell culture (SILAC), has proven to be particularly useful alone and in culmination with other methods to look into proteomic changes that occur due to various stimuli or stressors on a system.[13,14] Chemoenzymatic studies have involved utilizing N-azidoacetylglucosamine

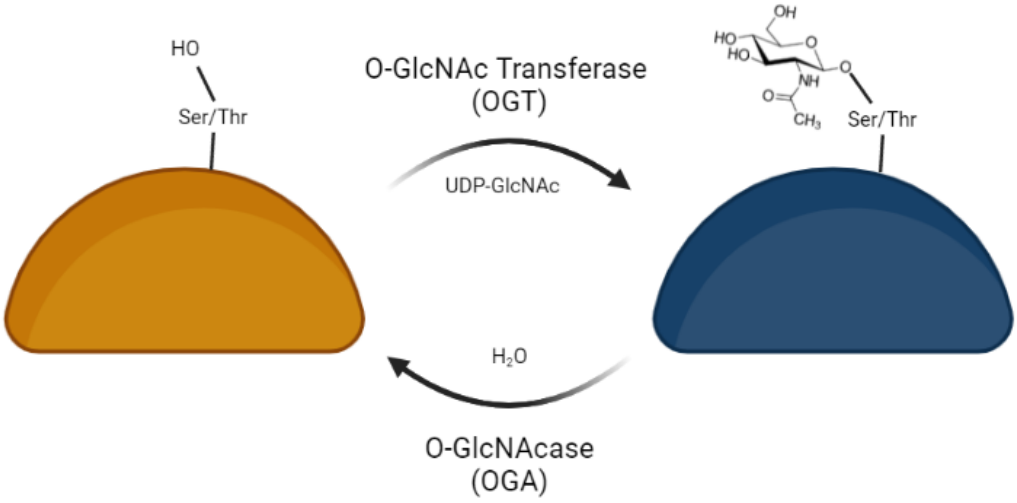


Figure 3.1: The Cycling of O-GlcNAc. Schematic showing the addition of O-GlcNAc to an acceptor substrate by OGT and the removal by OGA. (Created with BioRender.com)

(GlcNAz) for O-GlcNAc site mapping. Once supplemented, this approach allows for the salvage pathway to begin generating UDP-GlcNAz which can eventually be transferred to proteins by OGT.[15] These proteins can be digested and then identified by mass spectrometry. There are a number of other labeling and label free techniques that have been developed over the last couple of decades.[16–18] However, as important as these types of studies are the field is still lacking a more high throughput method to follow the dynamics of the modification and define the half-lives of specific O-GlcNAc sites.

As mentioned, O-GlcNAc is capable of cycling on and off proteins continuously, but little is known about the rates at which the O-GlcNAc modifications cycle. These cycling rates and half-lives could vary between proteins, and potentially between sites on the same and different proteins. Taking advantage of the HBP, supplementation of 15N-Gln in cell culture, and mass spectrometry for analysis we define herein the Dynamic IDAWG (Isotopic Detection of Aminosugars with Glutamine) method for the analysis of O-GlcNAc modification dynamics.[19] While the pulse chase experiment is similar to the experimental design described in Chapter 2, this take on the Dynamic IDAWG method allows for us to investigate the dynamics of abundant O-GlcNAc modified proteins, defining the half-lives of specific O-GlcNAc sites. Additionally, because of the incorporation of Glutamine into the protein backbone, our 15N tag allows for the half-life of specific proteins to be investigated simultaneously. Dynamic IDAWG for O-GlcNAc modified proteins has the potential to address the field's need for an

accessible and high throughput method to study the turnover rates of individual O-GlcNAc sites. It also opens the door to further answer questions regarding the relationship between O-GlcNAc and neural development, cancer, and a variety of cellular roles.[3,20–22]

Results

Experimental Design, Enrichment and Mass Spec Analysis

Similar to the original IDAWG methodology, we must utilize the hexosamine biosynthetic pathway in order to incorporate the heavy nitrogen tag into both the O-GlcNAc modification as well as the protein backbone.[19] RUES-1 WT cells were grown, labeled, and expanded out over the course of 10 days in media that was supplemented with 15N-Gln (Heavy Gln) in order to fully label each plate of cells. Once cells had been labeled to completion, a pulse chase experiment was conducted. At time zero (0hr) one set of plates was washed and harvested for the initial fully heavy labeled sample. The remaining sets of plates were rinsed and switched into 14N-Gln (Light Gln) media and then harvested at predetermined time points 1.5hr, 3hr, 6hr, and 12hr (**Fig. 3.2**). An additional set of light or unlabeled cells were also collected as a control.

Because the O-GlcNAc modification is found on nuclear and cytosolic proteins, the harvested cells underwent a nucleocytoplasmic (NC) extract. This also allows for the avoidance of cell surface protein contamination. The extracted proteins were then enriched for O-GlcNAc modified proteins overnight by means

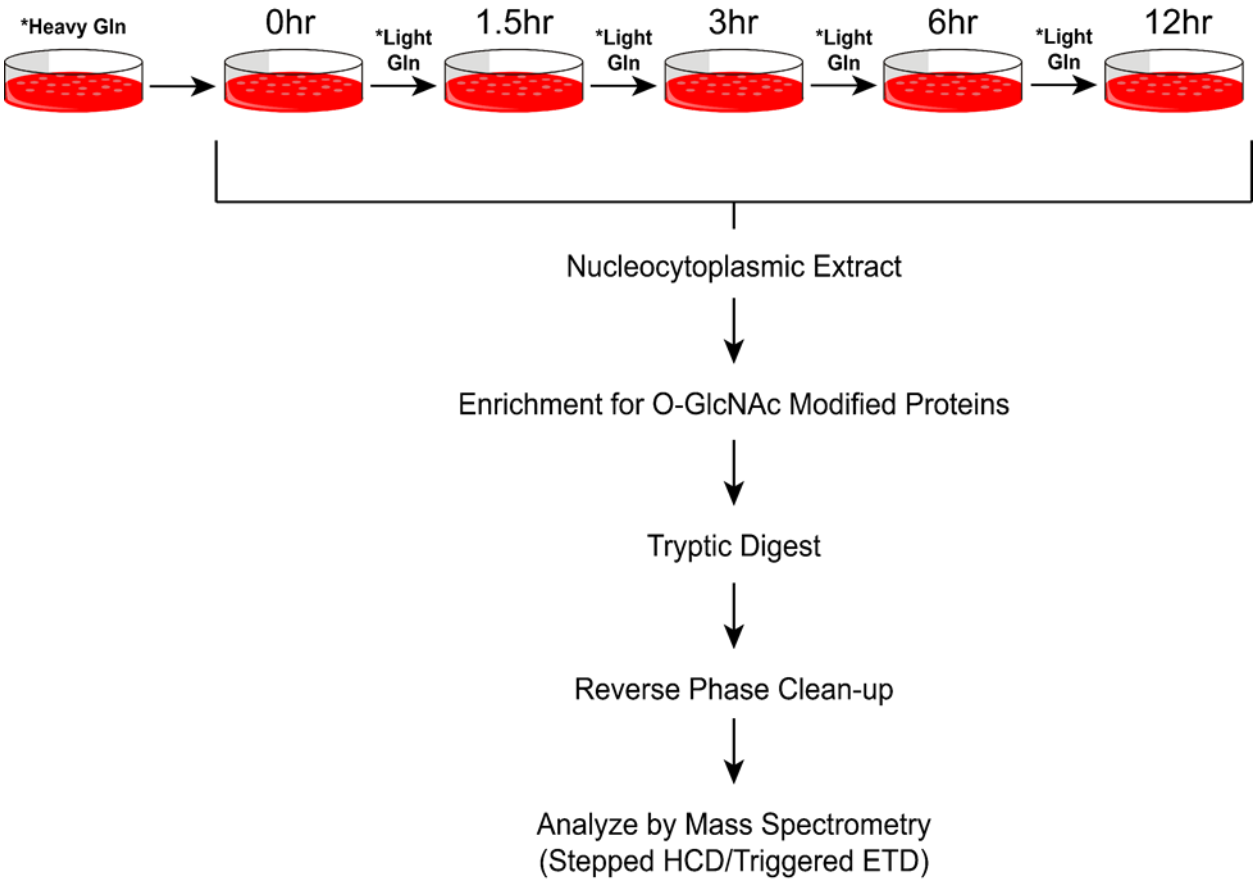


Figure 3.2: Experimental design for Dynamic IDAWG for O-GlcNAc modified proteins. Cells are grown in media supplemented with ^{15}N-Gln, undergo Pulse-Chase where they are transitioned into “light” or standard Gln supplemented media, and harvested at designated time points. Samples then undergo a nucleocytoplasmic extract, and those proteins are enriched over sWGA resin for O-GlcNAc modified proteins. Released proteins are then digested, cleaned up, and analyzed by mass spectrometry.

of succinylated wheat germ agglutinin (sWGA). By using sWGA in place of standard WGA enrichment, we avoid the potential enrichment of sialylated structures. The enriched glycoproteins were then released and separated from the sWGA resin and digested. Finally, the glycopeptides were subjected analysis via the Thermo Scientific Orbitrap Eclipse Tribrid mass spectrometer.

Dynamics of Abundant O-GlcNAc Modifications

The mass spectra that were collected for each time point, as well as an unlabeled light sample as a control, were analyzed by ProteinMetrics Inc. Byonic (PMI-Byonic) (v3.8.13). In order to define this method for the analysis of O-GlcNAc dynamics the list of peptides, possible O-GlcNAc modifications, and their corresponding protein IDs that were generated by PMI-Byonic were searched and analyzed under strict parameters. For each time point across four biological replicates, the Excel result files from the Byonic searches were filtered to only include unique peptides that contained a potential HexNAc modification. While searches varied across replicates, an average of 900 unique potential O-GlcNAc modified peptides were identified in each sample. It should also be noted that this included multiple options for the location of the modification on a single peptide. The 900 unique peptide sequences corresponded to 354 different proteins (**Supplemental Table 3.1**). Moving forward, the identified glycopeptides were then searched for across every time point for each replicate. To be considered for further analysis, the O-GlcNAc peptide had to be found in at least three of the four replicates and had to be in a majority of time points. This condensed list could also be validated through an O-GlcNAc database.[23] Lastly, the spectral and

fragmentation data had to be sufficient and contain the appropriate unlabeled/heavy-labeled, 204.08/205.08, peaks. O-GlcNAc IDs were discarded if only the peaks from the degradation of the oxonium ion were present.

With these strict parameters, we analyzed 18 unique O-GlcNAc modified peptides across 10 proteins to determine the average half-life of the modification at that particular site. Scans associated with the fragmentation of a selected peptide were averaged in Thermo Scientific™ FreeStyle™ (v1.6). The absolute and relative abundances of the averaged scans were recorded for each site for all replicates. We took advantage of the Microsoft Excel analysis tools that were developed for Dynamic IDAWG of released glycans for this analysis as well. Abundances of the 204 and 205 m/z peaks were averaged, normalized, and standard deviations were calculated. Utilizing these tools also allows for the empirical distribution of the unlabeled 204 and 205 peaks to be taken into consideration for more accurate calculation of turnover rates.

One of the more O-GlcNAc rich and abundant proteins we identified was Host Cell Factor 1 (HCFC1). While we analyzed a total of seven O-GlcNAc sites for this particular protein, it was apparent that there was an ample amount of variation between the calculated half-lives. Based on fragmentation, the peptide IPPSSAPTVLSVPAGTTIVK is modified on the second Thr residue in the sequence and predict a turnover rate of 15.3 hours (**Fig 3.3**). In contrast, a peptide such as APVTVTSLPAGVR cycles more quickly, turning over in 8.1 hours (**Fig 3.4**). While this exhibits the turnover of O-GlcNAc is not solely dependent on the protein that is being modified, the site that is modified is a contributing factor. Due to the

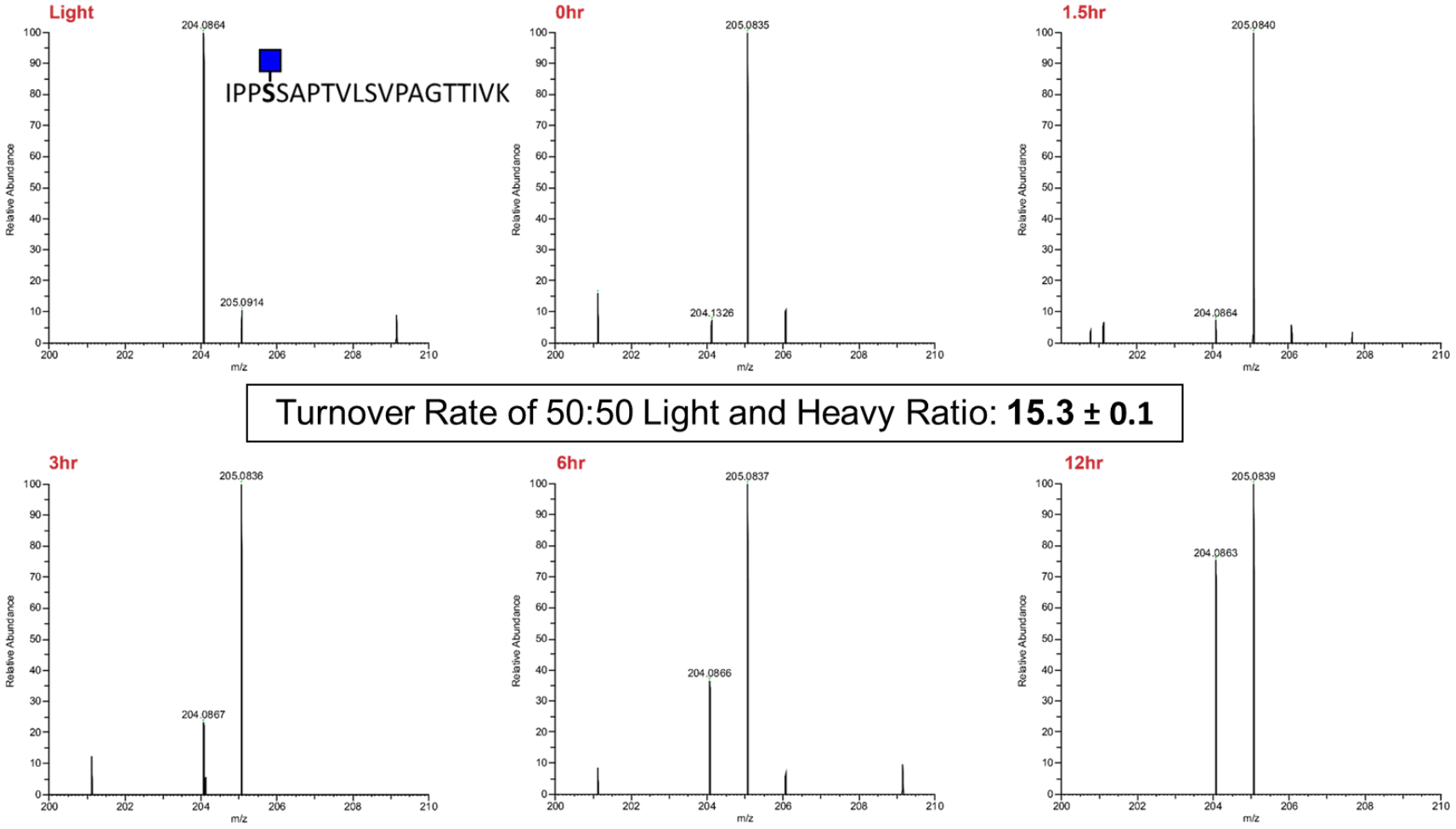

Light
204.0864
IPPSSAPTVLSVPAGTTIVK
205.0914
0hr
205.0835
204.1326
1.5hr
205.0840
204.0864
Relative Abundance
m/z
Turnover Rate of 50:50 Light and Heavy Ratio: 15.3 ± 0.1
3hr
205.0836
204.0867
6hr
205.0837
204.0866
12hr
205.0839
204.0863

Figure 3.3: Static O-GlcNAc modification of HCFC1. (A) Mass spectra collected showing the relative abundances of the unlabeled O-GlcNAc are calculated for the 204 and 205 peaks of the HCFC1 peptide, IPPSSAPTVLSVPAGTTIVK, to aid in accurately determining the turnover rate. (B-F) Relative abundances of unlabeled (204) and heavy labeled (205) O-GlcNAc on this HCFC1 peptide collected at (B) 0hr, (C)1.5hr, (D) 3hr, (E) 6hr, and (F) 12hr during pulse-chase. Following the harvest at time 0hr, cell culture media was changed to ^{14}N-Gln supplemented media for the remaining plates. This modification displays a longer turnover rate, closer to that of the protein, making the modification static.

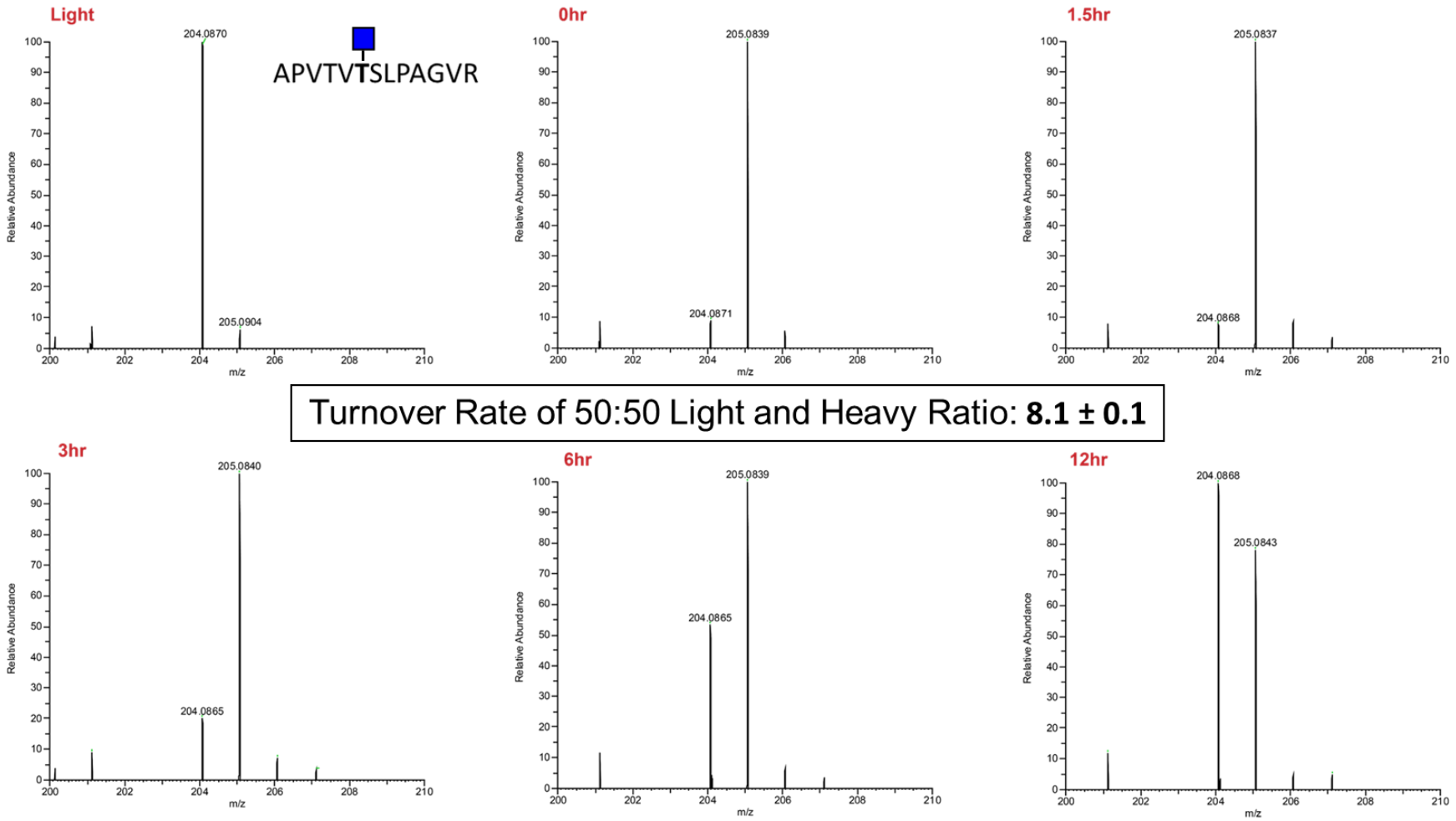

Light
204.0870
205.0904
APVTVTSLPAGVR
0hr
205.0839
204.0871
1.5hr
205.0837
204.0868
Relative Abundance
m/z
Turnover Rate of 50:50 Light and Heavy Ratio: 8.1 ± 0.1
3hr
205.0840
204.0865
6hr
205.0839
204.0865
12hr
204.0868
205.0843

Figure 3.4: Dynamic O-GlcNAc modification of HCFC1. (A) Mass spectra representing the 204/205 m/z distribution for the unlabeled O-GlcNAc modification for the HCFC1 peptide, APVTVTSLPAGVR. Pulse chase experiments allowed us to follow the changes in this distribution following the labeling of cells with 15N-Gln from (B) 0hr, (C)1.5hr, (D) 3hr, (E) 6hr, and (F) 12hr via mass spectrometry. This O-GlcNAc containing peptide exhibits a faster turnover rate than that of the protein it modifies, HCFC1, making this O-GlcNAc modification dynamic.

slow turnover rates of some modified sites, the calculated turnover rate actually exceeds that of the protein, and these are assumed to turn over with said protein. Turnover rates for our chosen top two O-GlcNAc modified proteins, each individual site identified through PMI-Byonic the turnover rates ranged from 8.1 ± 0.1 to 76.3 ± 0.1 hours (**Table 3.1**).

Dynamics in Relation to Modified Proteins

Having the ability to identify O-GlcNAc modified peptides and related proteins and following the rate in which a given site is turning over is an important tool. However, because the Dynamic IDAWG method also allows for the labeling of Gln in the protein backbone, this allowed us to calculate the turnover of the protein as well. Criteria for identifying glutamine containing peptides were equally as strict as those identifying O-GlcNAc modified peptides for analysis. In brief, the peptides had to be found in almost all time points for at least 3 of the 4 biological replicates with adequate fragmentation data. Due to these limitations, only a handful of O-GlcNAc modified proteins had a corresponding Gln containing peptide with adequate fragmentation data. The peptide GAPGQPGTILR was chosen for the comparative analysis of the protein's turnover to the identified O-GlcNAc sites on HCFC1. Based on calculations for this Gln containing peptide, HCFC1 is determined to have a turnover of roughly 15.7 hours (**Fig 3.5**). Looking at the data as a whole, O-GlcNAc on IPPSSAPTVLSVPAGTTIVK exhibits almost an identical turnover to the protein suggesting O-GlcNAc at this site is most likely static and turns over with the protein. However, we see the opposite is true for O-GlcNAc on

Table 3.1: List of modified peptides with their corresponding protein and calculated turn

(* denotes longer calculated turnover rate than protein; assumed to turnover with the protein)

Protein	Peptide	Turnover Rate
HCFC1-Host Cell Factor 1	GAPGQPGTILR	15.7 ± 0.3
HCFC1-Host Cell Factor 1	IPPSSAPTVLSVPAGTTIVK	15.3 ± 0.1
HCFC1-Host Cell Factor 1	APVTVTSLPAGVR	8.1 ± 0.1
HCFC1-Host Cell Factor 1	SPITIITTK	*36.5 ± 0.4
HCFC1-Host Cell Factor 1	TMAVTPGTTTLPATVK	*76.3 ± 0.1
HCFC1-Host Cell Factor 1	TAAAQVGTSVSSATNTSTRPIITVHK	11.8 ± 0.6
NU214-Nuclear Pore Complex Protein Nup214	VQGLLVPMK	116.4 ± 0.2
NU214-Nuclear Pore Complex Protein Nup214	KEPVLAQPAVSNSGTAASSTSLVALSAEATPATTGVPDAR	24.6 ± 0.3
NU214-Nuclear Pore Complex Protein Nup214	NNPATPSTAMGSSVPYSTAK	15.8 ± 0.4

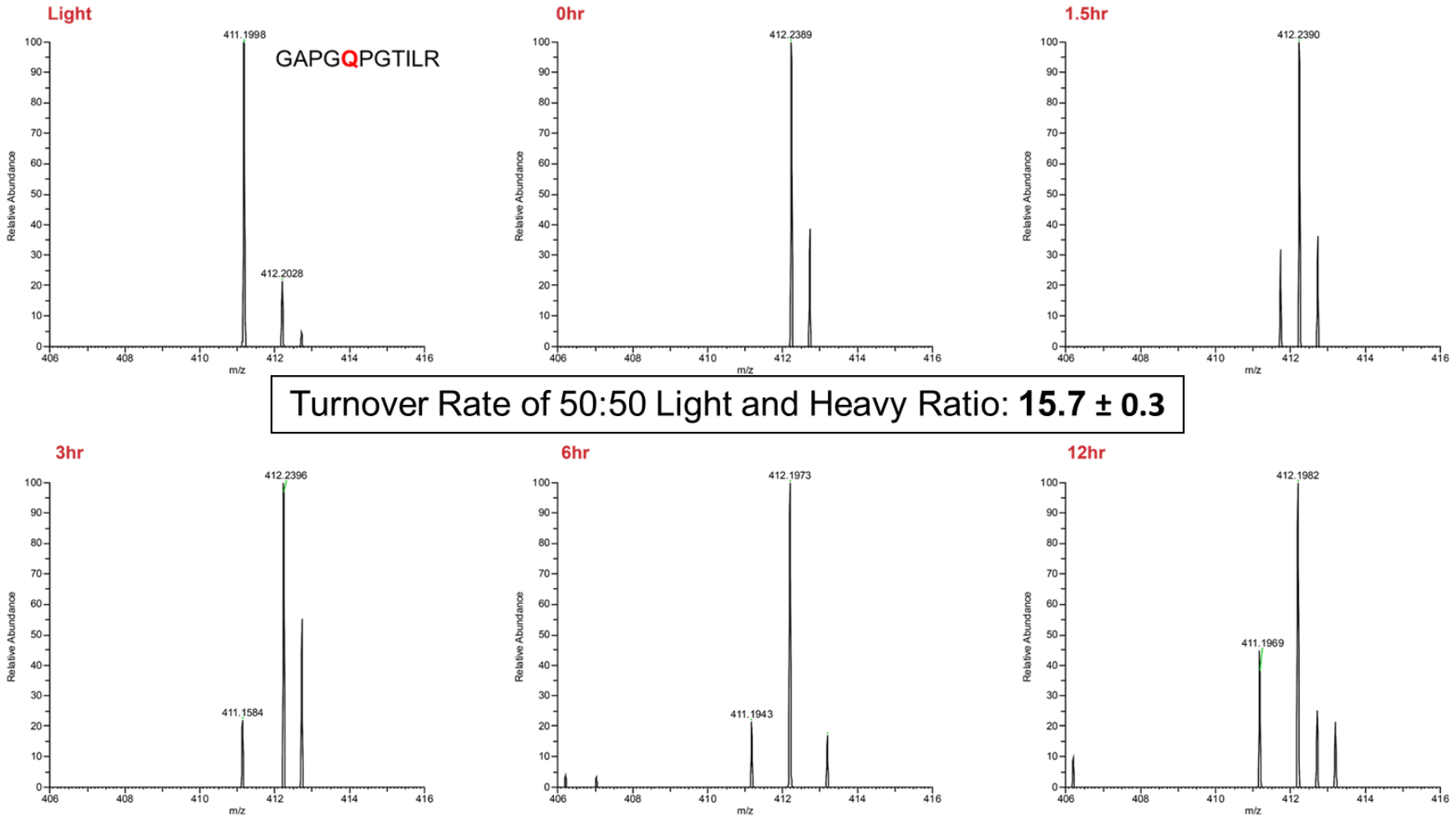

Light
411.1998
GAPGQPGTILR
412.2028
0hr
412.2389
1.5hr
412.2390
Relative Abundance
m/z
Turnover Rate of 50:50 Light and Heavy Ratio: 15.7 ± 0.3
3hr
412.2396
411.1584
6hr
412.1973
411.1943
12hr
412.1982
411.1969

Figure 3.5: Defining the turnover rate of the protein HCFC1 through a glutamine containing peptide. (A) The light unlabeled fragment ion for GAPGQPGTILR, the glutamine containing peptide of HCFC1. (B-F) Spectral data from 0-12hr of the pulse-chase experiment following the same fragment ion associated with the Gln within this HCFC1 peptide. The incorporation of the supplemented 15N-Gln into the protein backbone allows for the turnover rate to be determined for the overall protein, which is calculated to be approximately 15 hours.

APVTVTSLPAGVR (**Fig 3.6**). This particular site turns over almost twice as fast as HCFC1 itself, with a calculated turnover rate of 8.1 hours, making O-GlcNAc on that site a dynamic modification in contrast. We see this variation again for the Nuclear Pore Complex Protein (NU214) which exhibits an extensive turnover rate of 116 hours with more dynamic O-GlcNAc modifications turning over in about a fifth of that time. **Table 3.1** exhibits the variation we see between individual O-GlcNAc sites and potentially the protein itself.

Discussion

Described here is a more robust and easily accessible method for further research into the dynamics of individual O-GlcNAc modifications. The Dynamic IDAWG method for O-GlcNAc has shown to effectively follow the turnover of the post translational modification, defining the half-lives of specific sites of the selected abundant proteins. Due to the dual labeling ability of the 15N-Gln it allows for the determination of whether or not a specific modification is static, turning over with the protein, or dynamic, turning over faster than the protein it modifies. Because the criteria in which these particular O-GlcNAc peptides were chosen were extremely stringent, opening up the criteria slightly would allow for the analysis of a greater overall number of O-GlcNAc sites.

In contrast, to the Dynamic IDAWG method for released glycans here we use MS^2 scans to quantify the overall abundance of the light and heavy species present rather than the MS^1. While it is possible to use the MS^1 in the O-GlcNAc experiments as well, you are extremely limited in the peptides/O-GlcNAc sites you can analyze. In order to analyze via the MS^1 the peptide can only have a single

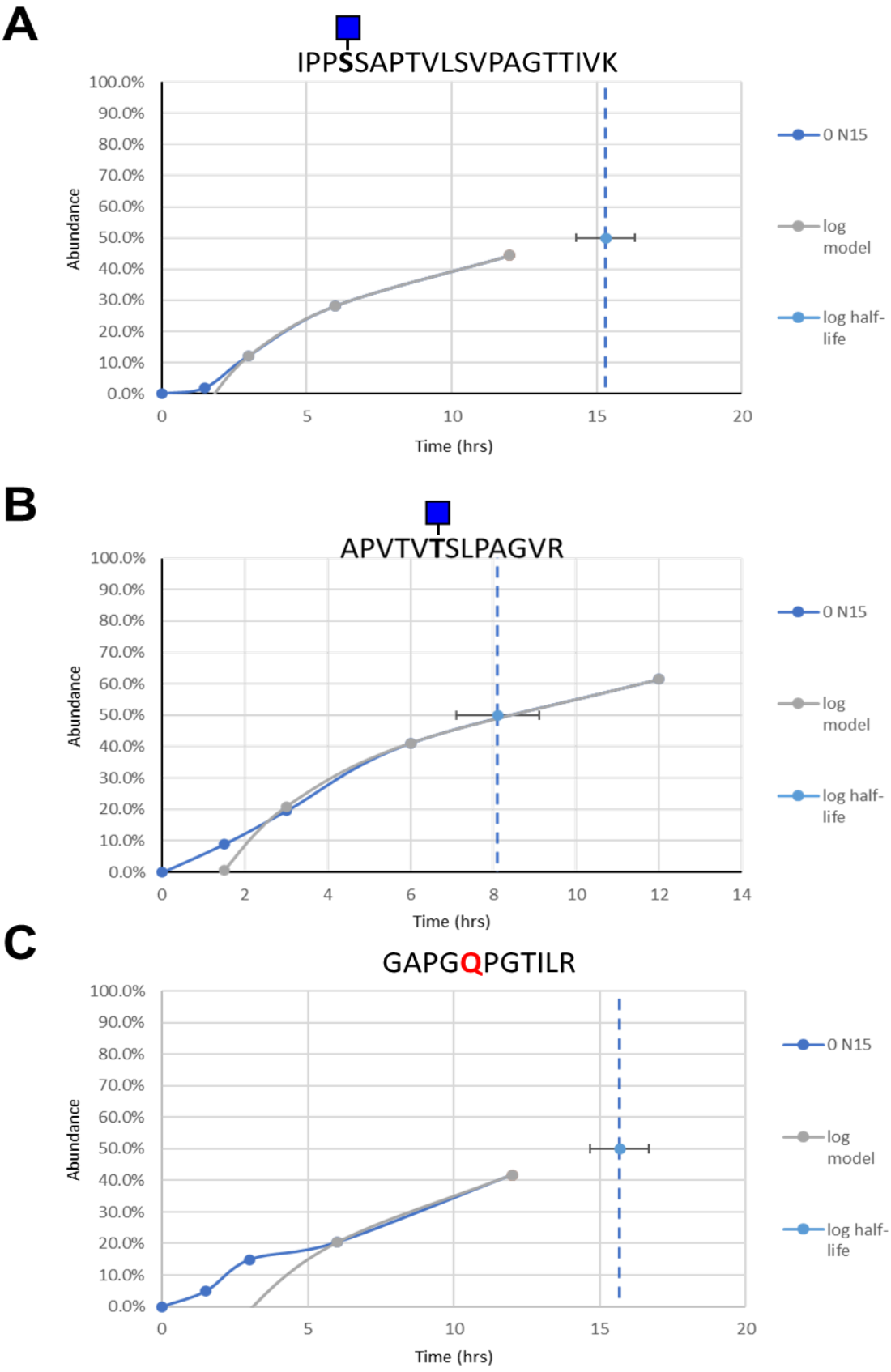

A
IPPSSAPTVLSVPAGTTIVK
100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
Abundance
0
5
10
15
20
Time (hrs)
0 N15
log model
log half-life
B
APVTVTSLPAGVR
100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
Abundance
0
2
4
6
8
10
12
14
Time (hrs)
0 N15
log model
log half-life
C
GAPGQPGTILR
100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
Abundance
0
5
10
15
20
Time (hrs)
0 N15
log model
log half-life

Figure 3.6: Calculated Turnover Rate Models of HCFC1 Peptides. Based upon the averaged and normalized spectral data across biological replicates over time and the expected relative abundances for the unlabeled species, the turnover rate can be calculated based on a logarithmic model for (A) static and (B) dynamic O-GlcNAc modifications. The same can be done for (C) the turnover of the protein can be determined based on relative abundances of the Gln containing fragment ion.

Gln or single O-GlcNAc, however by using MS^2 spectra you can quantify based on the 204/205 peaks or by the corresponding b/y fragment ions. Overall, the data shows variation in the calculated half-lives and exhibits no true pattern. The half-lives prove vary drastically even between sites on the same protein.

Having the ability to evaluate multiple sites and proteins with a single method will allow for further dynamics studies to be completed. While the actual data analysis on the back end of this method is slow-moving and less automated than the previously described Dynamic IDAWG for released glycans, it still proves to be an effective method. Incorporating this method and analysis will allow for a broader understanding of the impacts individual O-GlcNAc sites may have based on individual half-lives. Additionally, there is great potential for gaining insight into the role dynamics play and how those dynamics are impacted in regard to neural development and disease, cellular processes, diabetes, and more. Dynamic IDAWG gives us the capability to study those changes in cellulo.

Materials and Methods

Cell Culture and Pulse-Case Experiment

The RUES-1 WT line of hESCs were cultured essentially as previously described[24]. Briefly, the hESCs were grown on Matrigel® (Corning®) coated petri dishes at 37^{o}C under 5% CO_2. The base for the culture media was made in house composed of Dulbecco's modified Eagle's medium/F12/50:50 without Gln (DMEM, Invitrogen) supplemented with 10% Probumin, non-essential amino acids (Corning®), antibiotic/antimycotic, trace elements A, B, and C (Corning®), 2-mercaptoethanol, transferrin, ascorbic acid. and 2 mM L-glutamine which is either

normal abundance ^{14}N-isotope (Glutagro™-Corning®) or amide-^{15}N-Gln (98% purity, Cambridge Isotopes, Inc). Heregulin (PeproTech), Activin A (R&D Systems), IGF (Sigma®), and FGF (R&D Systems) (HAIF) were all supplemented into the base media fresh, prior to media changes. Cells were passaged every 3 to 4 days depending on density and split 1:4. The media was changed daily and the cells were allowed to grow and expand out for 10 days at which point part of them were harvested by dissociation buffer and scraping to be collected as the time 0hr sample. The remaining plates of cells were then transferred to light media supplemented with normal abundance amide-^{14}N-Gln and allowed to grow for 1.5, 3, 6, and 12 hours respectively. 8 plates per time point were collected. In addition to sets of light labeled plates for comparison. The harvested cell pellets were washed with phosphate buffered saline (PBS) and stored at -80˚C until analysis.

Nucleocytoplasmic Extract, sWGA Enrichment, and Digestion

Cell pellets that were previously harvested at designated time points, then underwent a nucleocytoplasmic (NC) extract, followed by sWGA enrichment and digestion. Briefly, cells were resuspended in cold lysis buffer A (10mM Tris-HCl pH 7.5, 500μM DTT, 500μM EDTA, protease inhibitors), incubated on ice for 10 minutes, vortexed, and placed back on ice. Vortexing and incubation were repeated twice. An equal volume of buffer B (10mM Tris pH 7.5, 800mM NaCl, protease inhibitors) was added and the cells were placed on ice for another 5 minutes. Cells were vortexed and incubated once more prior to being spun down at max speed for 10-15 minutes. Supernatant was moved to a new Eppendorf tube

and a BCA was performed on the NC extract to determine the protein concentration.

Once concentrations were determined for each sample, sWGA resin was prepped for O-GlcNAc enrichment. Approximately 3mg of protein were incubated for each time point sample with the appropriate volume of washed sWGA resin and incubated overnight at 4°C with end-over-end rotation. Following overnight incubation, samples are spun down and the flow-through/supernatant is removed. Resin is washed in wash buffer, rotating end-over-end at 4°C, a total of three times. Finally, bound O-GlcNAc proteins were released from the sWGA resin by adding 100µl of 50mM TEAB, 5% SDS buffer and boiling samples at 95°C for 10 minutes. Resin and supernatant were separated using a .2µm filter and centrifuged at 10,000xg for 2 minutes. A BCA was performed to determine the protein concentration prior to digestion. Flowthrough containing glycoproteins was then reduced, alkylated, acidified, and digested (Trypsin/Lys-C) using an S-Trap™ micro-MS sample prep kit: ≤100µg (Protifi). Samples were eluted off the column according to the kit protocol, dried down in a speed-vac, and stored at -20°C until MS analysis.

MS Analysis

Dried down glycopeptides were resuspended in 60µl, 95% Buffer A (0.1% formic acid) and 5% Buffer B (80% acetonitrile, 0.1% formic acid), where 6µl of sample was injected for each individual run on the Orbitrap Eclipse Tribrid Mass Spectrometer (Thermo Fisher Scientific). A method incorporating both stepped high collision-induced dissociation (stepped-HCD) and triggered electron transfer

dissociation (ETD) was utilized for this analysis. A mass list was curated based on the ‘heavy-labeled’ and ‘light’ O-GlcNAc modification as well as the masses associated with the degradation of the oxonium ion. When three or more of these masses were identified within an HCD scan, it triggers an ETD scan to take place, providing more information on the peptide sequence.

Evaluating the Turnover Rates of Identified O-GlcNAc Modifications

Raw files were processed through ProteinMetrics Inc. Byonic (PMI-Byonic) (v3.8.13). To first narrow down the proteomic database a light/unlabeled sample from RUES-1 WT cells was initially run against a concatenated database of SwissProt annotated human protein sequences and common contaminants. For this run and all subsequent runs, tryptic cleavage was specific at Arg and Lys, allowing for two missed cleavages. The precursor mass tolerance was 10ppm and fragment mass tolerance was 0.5Da/20ppm. Methylthio on Cys was set as a fixed modification and Oxidation of Methylthio, and HexNAc and Heavy HexNAc on Ser/Thr. The protein FDR was set at 2%. The resulting protein list from PMI-Byonic was then used to generate a focused database based upon the top proteins identified in the search. This database was utilized to run the remaining samples against.

Searches were completed to identify peptides with spectra indicating the presence of a light or heavy O-GlcNAc modification, which were identified by the 204 and 205 peak respectively. The light and heavy peaks associated with the degradation of the oxonium ion can also be identified within these same MS^2 scans. An excel file of all identified proteins and related peptides modified and

unmodified is generated for each PMI-Byonic run. Due to multiple IDs of the same peptide being included in the lists, lists were filtered to include only unique entries of peptides. Those lists were then filtered again to generate two separate lists: (Heavy) HexNAc containing peptides and Gln containing peptides. This was completed for all time points for each biological replicate. In order for a peptide of interest to be analyzed further, it was searched for across all biological replicates. If identified in at least three biological replicates and all time points of a replicate, it was chosen for turnover rate calculations. This was done for both HexNAc and Gln containing peptides. If a peptide contained both a labeled HexNAc and Gln it was discarded from this analysis.

Scans associated with identified peptides were averaged for each time point and the absolute and relative abundances were recorded for the unlabeled and heavy peaks associated with HexNAc or the fragment ion containing Gln. Using the same Excel bioinformatics tools mentioned in Chapter 2, the abundances were averaged, normalized, and standard deviations were calculated across biological replicates. Half Life calculations were based on the averaged biological replicate data over time and the empirical light isotopic data obtained from non-labeled sample data. Proteins with both a HexNAc containing peptide and separate Gln containing peptide could be further compared to determine if specific modifications were static or dynamic.

References

1. Torres, C.-R. & Hart, G. W. Topography and Polypeptide Distribution of Terminal N-Acetylglucosamine Residues on the Surfaces of Intact Lymphocytes. *J. Biol. Chem.* **259,** 3308–3317 (1984).
2. Holt, G. D. & Hart, G. W. The subcellular distribution of terminal N-acetylglucosamine moieties. Localization of a novel protein-saccharidie linkage, O-linked GlcNAc. *J. Biol. Chem.* **261,** 8049–8057 (1986).
3. Bond, M. R. & Hanover, J. A. *O*- GlcNAc Cycling: A Link Between Metabolism and Chronic Disease. *Annu. Rev. Nutr.* **33,** 205–229 (2013).
4. Nandi, A. *et al.* Global identification of O-GlcNAc-modified proteins. *Anal. Chem.* **78,** 452–458 (2006).
5. Love, D. C. & Hanover, J. A. The hexosamine signaling pathway: deciphering the 'O-GlcNAc code'. *Sci. STKE* **2005,** (2005).
6. Wells, L., Vosseller, K. & Hart, G. W. Glycosylation of nucleocytoplasmic proteins:signal transduction and O-GlcNAc. *Science (80-.).* **291,** 2376–2378 (2001).
7. Hanover, J. A., Krause, M. W. & Love, D. C. The hexosamine signaling pathway: O-GlcNAc cycling in feast or famine. *Biochim. Biophys. Acta - Gen. Subj.* **1800,** 80–95 (2010).
8. Haltiwanger, R. S., Holt, G. D. & Hart, G. W. Enzymatic additon of O-GlcNAc to nuclear and cytoplasmic proteins. Identification of a uridine diphospho-N-acetylglucosamine:peptide β-N-acetylglucosaminyltransferase. *J. Biol.*

Chem. **265,** 2563–2568 (1990).

9. Hart, G. W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. *Annu. Rev. Biochem.* **80,** 825–858 (2011).
10. Slawson, C. & Hart, G. W. Dynamic interplay between O-GlcNAc and O-phosphate: The sweet side of protein regulation. *Curr. Opin. Struct. Biol.* **13,** 631–636 (2003).
11. Slawson, C., Housley, M. P. & Hart, G. W. O-GlcNAc cycling: How a single sugar post-translational modification is changing the way we think about signaling networks. *J. Cell. Biochem.* **97,** 71–83 (2006).
12. Hart, G. W. *et al.* O-GlcAcylation of key nuclear and cytoskeletal proteins: Reciprocity with O-phosphorylation and putative roles in protein multimerization. *Glycobiology* **6,** 711–716 (1996).
13. Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1,** 376–386 (2002).
14. Ong, S. E. The expanding field of SILAC. *Anal. Bioanal. Chem.* **404,** 967–976 (2012).
15. Vocadlo, D. J., Hang, H. C., Kim, E. J., Hanover, J. A. & Bertozzi, C. R. A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 9116–9121 (2003).
16. Ma, J. & Hart, G. W. O-GlcNAc profiling: From proteins to proteomes. *Clin.*

Proteomics **11,** 1–16 (2014).

17. Cecioni, S. & Vocadlo, D. J. Tools for probing and perturbing O-GlcNAc in cells and in vivo. *Curr. Opin. Chem. Biol.* **17,** 719–728 (2013).
18. Khidekel, N. *et al.* Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat. Chem. Biol.* **3,** 339–348 (2007).
19. Orlando, R. *et al.* IDAWG: Metabolic incorporation of stable isotope labels for quantitative glycomics of cultured cells. *J. Proteome Res.* **8,** 3816–3823 (2009).
20. Zhu, Y., Shan, X., Yuzwa, S. A. & Vocadlo, D. J. The emerging link between O-GlcNAc and Alzheimer disease. *J. Biol. Chem.* **289,** 34472–34481 (2014).
21. Ferrer, C. M. *et al.* O-GlcNAcylation regulates cancer metabolism and survival stress signaling via regulation of HIF-1 pathway. *Mol. Cell* **54,** 820–831 (2015).
22. Ferrer, C. M., Sodi, V. L. & Reginato, M. J. O-GlcNAcylation in Cancer Biology: Linking Metabolism and Signaling. *J. Mol. Biol.* **428,** 3282–3294 (2016).
23. Ma, J., Li, Y., Hou, C. & Wu, C. O-GlcNAcAtlas: A database of experimentally identified O-GlcNAc sites and proteins. *Glycobiology* **31,** 719–723 (2021).
24. McLean, A. B. *et al.* Activin A Efficiently Specifies Definitive Endoderm from Human Embryonic Stem Cells Only When Phosphatidylinositol 3-Kinase Signaling Is Suppressed. *Stem Cells* **25,** 29–38 (2007).

Supplemental Table 3.1: Top O-GlcNAc Modified Proteins (≥6 unique peptides)

Protein Accession	Protein Description
HCFC1_HUMAN	Host cell factor 1 OS=Homo sapiens OX=9606 GN=HCFC1 PE=1 SV=2
P121C_HUMAN	Nuclear envelope pore membrane protein POM 121C OS=Homo sapiens OX=9606 GN=POM121C PE=1 SV=3
UBP2L_HUMAN	Ubiquitin-associated protein 2-like OS=Homo sapiens OX=9606 GN=UBAP2L PE=1 SV=2
UBAP2_HUMAN	Ubiquitin-associated protein 2 OS=Homo sapiens OX=9606 GN=UBAP2 PE=1 SV=1
RBM27_HUMAN	RNA-binding protein 27 OS=Homo sapiens OX=9606 GN=RBM27 PE=1 SV=2
S30BP_HUMAN	SAP30-binding protein OS=Homo sapiens OX=9606 GN=SAP30BP PE=1 SV=1
PRC2C_HUMAN	Protein PRRC2C OS=Homo sapiens OX=9606 GN=PRRC2C PE=1 SV=4
SC24B_HUMAN	Protein transport protein Sec24B OS=Homo sapiens OX=9606 GN=SEC24B PE=1 SV=2
NU214_HUMAN	Nuclear pore complex protein Nup214 OS=Homo sapiens OX=9606 GN=NUP214 PE=1 SV=2
WNK1_HUMAN	Serine/threonine-protein kinase WNK1 OS=Homo sapiens OX=9606 GN=WNK1 PE=1 SV=2
MGAP_HUMAN	MAX gene-associated protein OS=Homo sapiens OX=9606 GN=MGA PE=1 SV=4
QSER1_HUMAN	Glutamine and serine-rich protein 1 OS=Homo sapiens OX=9606 GN=QSER1 PE=1 SV=3
ZFR_HUMAN	Zinc finger RNA-binding protein OS=Homo sapiens OX=9606 GN=ZFR PE=1 SV=2
MYPT1_HUMAN	Protein phosphatase 1 regulatory subunit 12A OS=Homo sapiens OX=9606 GN=PPP1R12A PE=1 SV=1
KANL3_HUMAN	KAT8 regulatory NSL complex subunit 3 OS=Homo sapiens OX=9606 GN=KANSL3 PE=1 SV=2
SOX2_HUMAN	Transcription factor SOX-2 OS=Homo sapiens OX=9606 GN=SOX2 PE=1 SV=1
ANR17_HUMAN	Ankyrin repeat domain-containing protein 17 OS=Homo sapiens OX=9606 GN=ANKRD17 PE=1 SV=3
NU153_HUMAN	Nuclear pore complex protein Nup153 OS=Homo sapiens OX=9606 GN=NUP153 PE=1 SV=2
EMSY_HUMAN	BRCA2-interacting transcriptional repressor EMSY OS=Homo sapiens OX=9606 GN=EMSY PE=1 SV=2
PRSR1_HUMAN	Proline and serine-rich protein 1 OS=Homo sapiens OX=9606 GN=PROSER1 PE=1 SV=2
K2C1_HUMAN	Keratin, type II cytoskeletal 1 OS=Homo sapiens OX=9606 GN=KRT1 PE=1 SV=6
GLU2B_HUMAN	Glucosidase 2 subunit beta OS=Homo sapiens OX=9606 GN=PRKCSH PE=1 SV=2
NUP98_HUMAN	Nuclear pore complex protein Nup98-Nup96 OS=Homo sapiens OX=9606 GN=NUP98 PE=1 SV=4
CDC5L_HUMAN	Cell division cycle 5-like protein OS=Homo sapiens OX=9606 GN=CDC5L PE=1 SV=2
KINH_HUMAN	Kinesin-1 heavy chain OS=Homo sapiens OX=9606 GN=KIF5B PE=1 SV=1
CIC_HUMAN P	Protein capicua homolog OS=Homo sapiens OX=9606 GN=CIC PE=1 SV=2
XPO5_HUMAN	Exportin-5 OS=Homo sapiens OX=9606 GN=XPO5 PE=1 SV=1

Protein Accession	Protein Description
TRIPC_HUMAN	E3 ubiquitin-protein ligase TRIP12 OS=Homo sapiens OX=9606 GN=TRIP12 PE=1 SV=1
COPB_HUMAN	Coatomer subunit beta OS=Homo sapiens OX=9606 GN=COPB1 PE=1 SV=3
PLEC_HUMAN	Plectin OS=Homo sapiens OX=9606 GN=PLEC PE=1 SV=3
SIN3A_HUMAN	Paired amphipathic helix protein Sin3a OS=Homo sapiens OX=9606 GN=SIN3A PE=1 SV=2
NOP14_HUMAN	Nucleolar protein 14 OS=Homo sapiens OX=9606 GN=NOP14 PE=1 SV=3
SERA_HUMAN	D-3-phosphoglycerate dehydrogenase OS=Homo sapiens OX=9606 GN=PHGDH PE=1 SV=4
USO1_HUMAN	General vesicular transport factor p115 OS=Homo sapiens OX=9606 GN=USO1 PE=1 SV=2
ESYT1_HUMAN	Extended synaptotagmin-1 OS=Homo sapiens OX=9606 GN=ESYT1 PE=1 SV=1
IMA7_HUMAN	Importin subunit alpha-7 OS=Homo sapiens OX=9606 GN=KPNA6 PE=1 SV=1
IF4A2_HUMAN	Eukaryotic initiation factor 4A-II OS=Homo sapiens OX=9606 GN=EIF4A2 PE=1 SV=2
HNRPU_HUMAN	Heterogeneous nuclear ribonucleoprotein U OS=Homo sapiens OX=9606 GN=HNRNPU PE=1 SV=6
NUMA1_HUMAN	Nuclear mitotic apparatus protein 1 OS=Homo sapiens OX=9606 GN=NUMA1 PE=1 SV=2
UBR4_HUMAN	E3 ubiquitin-protein ligase UBR4 OS=Homo sapiens OX=9606 GN=UBR4 PE=1 SV=1
FMR1_HUMAN	Synaptic functional regulator FMR1 OS=Homo sapiens OX=9606 GN=FMR1 PE=1 SV=1
DDX21_HUMAN	Nucleolar RNA helicase 2 OS=Homo sapiens OX=9606 GN=DDX21 PE=1 SV=5
RL15_HUMAN	60S ribosomal protein L15 OS=Homo sapiens OX=9606 GN=RPL15 PE=1 SV=2
SMHD1_HUMAN	Structural maintenance of chromosomes flexible hinge domain-containing protein 1 OS=Homo sapiens OX=9606 GN=SMCHD1 PE=1 SV=2
PROM1_HUMAN	Prominin-1 OS=Homo sapiens OX=9606 GN=PROM1 PE=1 SV=1
TBA1B_HUMAN	Tubulin alpha-1B chain OS=Homo sapiens OX=9606 GN=TUBA1B PE=1 SV=1
MY18A_HUMAN	Unconventional myosin-XVIIIa OS=Homo sapiens OX=9606 GN=MYO18A PE=1 SV=3
LAMB1_HUMAN	Laminin subunit beta-1 OS=Homo sapiens OX=9606 GN=LAMB1 PE=1 SV=2
PSMD3_HUMAN	26S proteasome non-ATPase regulatory subunit 3 OS=Homo sapiens OX=9606 GN=PSMD3 PE=1 SV=2
SYIC_HUMAN	Isoleucine--tRNA ligase, cytoplasmic OS=Homo sapiens OX=9606 GN=IARS PE=1 SV=2
PHB2_HUMAN	Prohibitin-2 OS=Homo sapiens OX=9606 GN=PHB2 PE=1 SV=2
ACTN4_HUMAN	Alpha-actinin-4 OS=Homo sapiens OX=9606 GN=ACTN4 PE=1 SV=2
BCLF1_HUMAN	Bcl-2-associated transcription factor 1 OS=Homo sapiens OX=9606 GN=BCLAF1 PE=1 SV=2
LAMC1_HUMAN	Laminin subunit gamma-1 OS=Homo sapiens OX=9606 GN=LAMC1 PE=1 SV=3

Protein Accession	Protein Description
TCPE_HUMAN	T-complex protein 1 subunit epsilon OS=Homo sapiens OX=9606 GN=CCT5 PE=1 SV=1
RAGP1_HUMAN	Ran GTPase-activating protein 1 OS=Homo sapiens OX=9606 GN=RANGAP1 PE=1 SV=1
PUR9_HUMAN	Bifunctional purine biosynthesis protein PURH OS=Homo sapiens OX=9606 GN=ATIC PE=1 SV=3
DUT_HUMAN	Deoxyuridine 5'-triphosphate nucleotidohydrolase, mitochondrial OS=Homo sapiens OX=9606 GN=DUT PE=1 SV=4
DYHC1_HUMAN	Cytoplasmic dynein 1 heavy chain 1 OS=Homo sapiens OX=9606 GN=DYNC1H1 PE=1 SV=5
RRP1B_HUMAN	Ribosomal RNA processing protein 1 homolog B OS=Homo sapiens OX=9606 GN=RRP1B PE=1 SV=3
SF3B3_HUMAN	Splicing factor 3B subunit 3 OS=Homo sapiens OX=9606 GN=SF3B3 PE=1 SV=4
TNPO1_HUMAN	Transportin-1 OS=Homo sapiens OX=9606 GN=TNPO1 PE=1 SV=2
ZO1_HUMAN	Tight junction protein ZO-1 OS=Homo sapiens OX=9606 GN=TJP1 PE=1 SV=3
ILF2_HUMAN	Interleukin enhancer-binding factor 2 OS=Homo sapiens OX=9606 GN=ILF2 PE=1 SV=2
KI67_HUMAN	Proliferation marker protein Ki-67 OS=Homo sapiens OX=9606 GN=MKI67 PE=1 SV=2
IMA4_HUMAN	Importin subunit alpha-4 OS=Homo sapiens OX=9606 GN=KPNA3 PE=1 SV=2
GLYM_HUMAN	Serine hydroxymethyltransferase, mitochondrial OS=Homo sapiens OX=9606 GN=SHMT2 PE=1 SV=3
ARHG2_HUMAN	Rho guanine nucleotide exchange factor 2 OS=Homo sapiens OX=9606 GN=ARHGEF2 PE=1 SV=4
NAT10_HUMAN	RNA cytidine acetyltransferase OS=Homo sapiens OX=9606 GN=NAT10 PE=1 SV=2
FBLN1_HUMAN	Fibulin-1 OS=Homo sapiens OX=9606 GN=FBLN1 PE=1 SV=4
TMOD3_HUMAN	Tropomodulin-3 OS=Homo sapiens OX=9606 GN=TMOD3 PE=1 SV=1
POGZ_HUMAN	Pogo transposable element with ZNF domain OS=Homo sapiens OX=9606 GN=POGZ PE=1 SV=2
KAT8_HUMAN	Histone acetyltransferase KAT8 OS=Homo sapiens OX=9606 GN=KAT8 PE=1 SV=2
F120B_HUMAN	Constitutive coactivator of peroxisome proliferator-activated receptor gamma OS=Homo sapiens OX=9606 GN=FAM120B PE=1 SV=1
CALU_HUMAN	Calumenin OS=Homo sapiens OX=9606 GN=CALU PE=1 SV=2
HEXA_HUMAN	Beta-hexosaminidase subunit alpha OS=Homo sapiens OX=9606 GN=HEXA PE=1 SV=2
HNRPM_HUMAN	Heterogeneous nuclear ribonucleoprotein M OS=Homo sapiens OX=9606 GN=HNRNPM PE=1 SV=3
MBB1A_HUMAN	Myb-binding protein 1A OS=Homo sapiens OX=9606 GN=MYBBP1A PE=1 SV=2
MLH1_HUMAN	DNA mismatch repair protein Mlh1 OS=Homo sapiens OX=9606 GN=MLH1 PE=1 SV=1
FRAS1_HUMAN	Extracellular matrix protein FRAS1 OS=Homo sapiens OX=9606 GN=FRAS1 PE=1 SV=2
MACF1_HUMAN	Microtubule-actin cross-linking factor 1, isoforms 1/2/3/5 OS=Homo sapiens OX=9606 GN=MACF1 PE=1 SV=4

Protein Accession	Protein Description
OSBL8_HUMAN	Oxysterol-binding protein-related protein 8 OS=Homo sapiens OX=9606 GN=OSBPL8 PE=1 SV=3
GUAA_HUMAN	GMP synthase [glutamine-hydrolyzing] OS=Homo sapiens OX=9606 GN=GMPS PE=1 SV=1
ALDOA_HUMAN	Fructose-bisphosphate aldolase A OS=Homo sapiens OX=9606 GN=ALDOA PE=1 SV=2
WNK3_HUMAN	Serine/threonine-protein kinase WNK3 OS=Homo sapiens OX=9606 GN=WNK3 PE=1 SV=3
NUP50_HUMAN	Nuclear pore complex protein Nup50 OS=Homo sapiens OX=9606 GN=NUP50 PE=1 SV=2
FLNB_HUMAN	Filamin-B OS=Homo sapiens OX=9606 GN=FLNB PE=1 SV=2
NOTC3_HUMAN	Neurogenic locus notch homolog protein 3 OS=Homo sapiens OX=9606 GN=NOTCH3 PE=1 SV=2
ITAV_HUMAN	Integrin alpha-V OS=Homo sapiens OX=9606 GN=ITGAV PE=1 SV=2
PABP1_HUMAN	Polyadenylate-binding protein 1 OS=Homo sapiens OX=9606 GN=PABPC1 PE=1 SV=2
LAMA5_HUMAN	Laminin subunit alpha-5 OS=Homo sapiens OX=9606 GN=LAMA5 PE=1 SV=8
GNL3_HUMAN	Guanine nucleotide-binding protein-like 3 OS=Homo sapiens OX=9606 GN=GNL3 PE=1 SV=2
TET1_HUMAN	Methylcytosine dioxygenase TET1 OS=Homo sapiens OX=9606 GN=TET1 PE=1 SV=2
U520_HUMAN	U5 small nuclear ribonucleoprotein 200 kDa helicase OS=Homo sapiens OX=9606 GN=SNRNP200 PE=1 SV=2
SYRC_HUMAN	Arginine--tRNA ligase, cytoplasmic OS=Homo sapiens OX=9606 GN=RARS PE=1 SV=2
EF1A1_HUMAN	Elongation factor 1-alpha 1 OS=Homo sapiens OX=9606 GN=EEF1A1 PE=1 SV=1
UIF_HUMAN U	UAP56-interacting factor OS=Homo sapiens OX=9606 GN=FYTTD1 PE=1 SV=3
LASP1_HUMAN	LIM and SH3 domain protein 1 OS=Homo sapiens OX=9606 GN=LASP1 PE=1 SV=2
RPN1_HUMAN	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 OS=Homo sapiens OX=9606 GN=RPN1 PE=1 SV=1
HEXB_HUMAN	Beta-hexosaminidase subunit beta OS=Homo sapiens OX=9606 GN=HEXB PE=1 SV=3
DNJA1_HUMAN	DnaJ homolog subfamily A member 1 OS=Homo sapiens OX=9606 GN=DNAJA1 PE=1 SV=2
C2D1A_HUMAN	Coiled-coil and C2 domain-containing protein 1A OS=Homo sapiens OX=9606 GN=CC2D1A PE=1 SV=1
HDAC2_HUMAN	Histone deacetylase 2 OS=Homo sapiens OX=9606 GN=HDAC2 PE=1 SV=2
HS105_HUMAN	Heat shock protein 105 kDa OS=Homo sapiens OX=9606 GN=HSPH1 PE=1 SV=1
CA167_HUMAN	Uncharacterized protein C1orf167 OS=Homo sapiens OX=9606 GN=C1orf167 PE=2 SV=2
DESP_HUMAN	Desmoplakin OS=Homo sapiens OX=9606 GN=DSP PE=1 SV=3
FAT1_HUMAN	Protocadherin Fat 1 OS=Homo sapiens OX=9606 GN=FAT1 PE=1 SV=2
MYO6_HUMAN	Unconventional myosin-VI OS=Homo sapiens OX=9606 GN=MYO6 PE=1 SV=4

Protein Accession	Protein Description
PSA_HUMAN	Puromycin-sensitive aminopeptidase OS=Homo sapiens OX=9606 GN=NPEPPS PE=1 SV=2
HYOU1_HUMAN	Hypoxia up-regulated protein 1 OS=Homo sapiens OX=9606 GN=HYOU1 PE=1 SV=1
CAPZB_HUMAN	F-actin-capping protein subunit beta OS=Homo sapiens OX=9606 GN=CAPZB PE=1 SV=4
TXND5_HUMAN	Thioredoxin domain-containing protein 5 OS=Homo sapiens OX=9606 GN=TXNDC5 PE=1 SV=2
EIF3B_HUMAN	Eukaryotic translation initiation factor 3 subunit B OS=Homo sapiens OX=9606 GN=EIF3B PE=1 SV=3
SPB9_HUMAN	Serpin B9 OS=Homo sapiens OX=9606 GN=SERPINB9 PE=1 SV=1
DJC10_HUMAN	DnaJ homolog subfamily C member 10 OS=Homo sapiens OX=9606 GN=DNAJC10 PE=1 SV=2
MSH6_HUMAN	DNA mismatch repair protein Msh6 OS=Homo sapiens OX=9606 GN=MSH6 PE=1 SV=2
MTREX_HUMAN	Exosome RNA helicase MTR4 OS=Homo sapiens OX=9606 GN=MTREX PE=1 SV=3
K2C6B_HUMAN	Keratin, type II cytoskeletal 6B OS=Homo sapiens OX=9606 GN=KRT6B PE=1 SV=5
TERA_HUMAN	Transitional endoplasmic reticulum ATPase OS=Homo sapiens OX=9606 GN=VCP PE=1 SV=4
AT1A1_HUMAN	Sodium/potassium-transporting ATPase subunit alpha-1 OS=Homo sapiens OX=9606 GN=ATP1A1 PE=1 SV=1
RRBP1_HUMAN	Ribosome-binding protein 1 OS=Homo sapiens OX=9606 GN=RRBP1 PE=1 SV=5
KANL1_HUMAN	KAT8 regulatory NSL complex subunit 1 OS=Homo sapiens OX=9606 GN=KANSL1 PE=1 SV=2
PFKAP_HUMAN	ATP-dependent 6-phosphofructokinase, platelet type OS=Homo sapiens OX=9606 GN=PFKP PE=1 SV=2
USP9X_HUMAN	Probable ubiquitin carboxyl-terminal hydrolase FAF-X OS=Homo sapiens OX=9606 GN=USP9X PE=1 SV=3
INO1_HUMAN	Inositol-3-phosphate synthase 1 OS=Homo sapiens OX=9606 GN=ISYNA1 PE=1 SV=1
MPRIP_HUMAN	Myosin phosphatase Rho-interacting protein OS=Homo sapiens OX=9606 GN=MPRIP PE=1 SV=3
LRP1_HUMAN	Prolow-density lipoprotein receptor-related protein 1 OS=Homo sapiens OX=9606 GN=LRP1 PE=1 SV=2
TPM1_HUMAN	Tropomyosin alpha-1 chain OS=Homo sapiens OX=9606 GN=TPM1 PE=1 SV=2
LAMB2_HUMAN	Laminin subunit beta-2 OS=Homo sapiens OX=9606 GN=LAMB2 PE=1 SV=2
RL6_HUMAN	60S ribosomal protein L6 OS=Homo sapiens OX=9606 GN=RPL6 PE=1 SV=3
FREM2_HUMAN	FRAS1-related extracellular matrix protein 2 OS=Homo sapiens OX=9606 GN=FREM2 PE=1 SV=2
VINC_HUMAN	Vinculin OS=Homo sapiens OX=9606 GN=VCL PE=1 SV=4
CARM1_HUMAN	Histone-arginine methyltransferase CARM1 OS=Homo sapiens OX=9606 GN=CARM1 PE=1 SV=3
KC1A_HUMAN	Casein kinase I isoform alpha OS=Homo sapiens OX=9606 GN=CSNK1A1 PE=1 SV=2

Protein Accession	Protein Description
PGK1_HUMAN	Phosphoglycerate kinase 1 OS=Homo sapiens OX=9606 GN=PGK1 PE=1 SV=3
1433E_HUMAN	14-3-3 protein epsilon OS=Homo sapiens OX=9606 GN=YWHAE PE=1 SV=1
SALL2_HUMAN	Sal-like protein 2 OS=Homo sapiens OX=9606 GN=SALL2 PE=1 SV=4
VGFR1_HUMAN	Vascular endothelial growth factor receptor 1 OS=Homo sapiens OX=9606 GN=FLT1 PE=1 SV=2
RBM26_HUMAN	RNA-binding protein 26 OS=Homo sapiens OX=9606 GN=RBM26 PE=1 SV=3
ITB1_HUMAN	Integrin beta-1 OS=Homo sapiens OX=9606 GN=ITGB1 PE=1 SV=2
SPS2L_HUMAN	SPATS2-like protein OS=Homo sapiens OX=9606 GN=SPATS2L PE=1 SV=2
B3GN7_HUMAN	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 7 OS=Homo sapiens OX=9606 GN=B3GNT7 PE=2 SV=1
SEM6A_HUMAN	Semaphorin-6A OS=Homo sapiens OX=9606 GN=SEMA6A PE=1 SV=2
S12A4_HUMAN	Solute carrier family 12 member 4 OS=Homo sapiens OX=9606 GN=SLC12A4 PE=1 SV=2
PUR6_HUMAN	Multifunctional protein ADE2 OS=Homo sapiens OX=9606 GN=PAICS PE=1 SV=3
PP2AB_HUMAN	Serine/threonine-protein phosphatase 2A catalytic subunit beta isoform OS=Homo sapiens OX=9606 GN=PPP2CB PE=1 SV=1
COPA_HUMAN	Coatomer subunit alpha OS=Homo sapiens OX=9606 GN=COPA PE=1 SV=2
TCPG_HUMAN	T-complex protein 1 subunit gamma OS=Homo sapiens OX=9606 GN=CCT3 PE=1 SV=4
TBCD_HUMAN	Tubulin-specific chaperone D OS=Homo sapiens OX=9606 GN=TBCD PE=1 SV=2
DHX9_HUMAN	ATP-dependent RNA helicase A OS=Homo sapiens OX=9606 GN=DHX9 PE=1 SV=4
MAP1B_HUMAN	Microtubule-associated protein 1B OS=Homo sapiens OX=9606 GN=MAP1B PE=1 SV=2
FERM2_HUMAN	Fermitin family homolog 2 OS=Homo sapiens OX=9606 GN=FERMT2 PE=1 SV=1
TSR1_HUMAN	Pre-rRNA-processing protein TSR1 homolog OS=Homo sapiens OX=9606 GN=TSR1 PE=1 SV=1
PRC2A_HUMAN	Protein PRRC2A OS=Homo sapiens OX=9606 GN=PRRC2A PE=1 SV=3
SLTM_HUMAN	SAFB-like transcription modulator OS=Homo sapiens OX=9606 GN=SLTM PE=1 SV=2
MA2B1_HUMAN	Lysosomal alpha-mannosidase OS=Homo sapiens OX=9606 GN=MAN2B1 PE=1 SV=3
DNMT1_HUMAN	DNA (cytosine-5)-methyltransferase 1 OS=Homo sapiens OX=9606 GN=DNMT1 PE=1 SV=2
XPO1_HUMAN	Exportin-1 OS=Homo sapiens OX=9606 GN=XPO1 PE=1 SV=1
PPT1_HUMAN	Palmitoyl-protein thioesterase 1 OS=Homo sapiens OX=9606 GN=PPT1 PE=1 SV=1
SDS3_HUMAN	Sin3 histone deacetylase corepressor complex component SDS3 OS=Homo sapiens OX=9606 GN=SUDS3 PE=1 SV=2

Protein Accession	Protein Description
2AAA_HUMAN	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform OS=Homo sapiens OX=9606 GN=PPP2R1A PE=1 SV=4
RS3A_HUMAN	40S ribosomal protein S3a OS=Homo sapiens OX=9606 GN=RPS3A PE=1 SV=2
FLNC_HUMAN	Filamin-C OS=Homo sapiens OX=9606 GN=FLNC PE=1 SV=3
RBM12_HUMAN	RNA-binding protein 12 OS=Homo sapiens OX=9606 GN=RBM12 PE=1 SV=1
LSG1_HUMAN	Large subunit GTPase 1 homolog OS=Homo sapiens OX=9606 GN=LSG1 PE=1 SV=2
MAP7_HUMAN	Ensconsin OS=Homo sapiens OX=9606 GN=MAP7 PE=1 SV=1
PRS7_HUMAN	26S proteasome regulatory subunit 7 OS=Homo sapiens OX=9606 GN=PSMC2 PE=1 SV=3
DHX37_HUMAN	Probable ATP-dependent RNA helicase DHX37 OS=Homo sapiens OX=9606 GN=DHX37 PE=1 SV=1
MAP4_HUMAN	Microtubule-associated protein 4 OS=Homo sapiens OX=9606 GN=MAP4 PE=1 SV=3
MYO1B_HUMAN	Unconventional myosin-Ib OS=Homo sapiens OX=9606 GN=MYO1B PE=1 SV=3
MYO1C_HUMAN	Unconventional myosin-Ic OS=Homo sapiens OX=9606 GN=MYO1C PE=1 SV=4
RL27A_HUMAN	60S ribosomal protein L27a OS=Homo sapiens OX=9606 GN=RPL27A PE=1 SV=2
DDX41_HUMAN	Probable ATP-dependent RNA helicase DDX41 OS=Homo sapiens OX=9606 GN=DDX41 PE=1 SV=2
GNS_HUMAN N	N-acetylglucosamine-6-sulfatase OS=Homo sapiens OX=9606 GN=GNS PE=1 SV=3
CAV1_HUMAN	Caveolin-1 OS=Homo sapiens OX=9606 GN=CAV1 PE=1 SV=4
GPC4_HUMAN	Glypican-4 OS=Homo sapiens OX=9606 GN=GPC4 PE=1 SV=4
AGRIN_HUMAN	Agrin OS=Homo sapiens OX=9606 GN=AGRN PE=1 SV=6
CH60_HUMAN	60 kDa heat shock protein, mitochondrial OS=Homo sapiens OX=9606 GN=HSPD1 PE=1 SV=2
UBA1_HUMAN	Ubiquitin-like modifier-activating enzyme 1 OS=Homo sapiens OX=9606 GN=UBA1 PE=1 SV=3
SND1_HUMAN	Staphylococcal nuclease domain-containing protein 1 OS=Homo sapiens OX=9606 GN=SND1 PE=1 SV=1
PTPRU_HUMAN	Receptor-type tyrosine-protein phosphatase U OS=Homo sapiens OX=9606 GN=PTPRU PE=1 SV=2
GDIB_HUMAN	Rab GDP dissociation inhibitor beta OS=Homo sapiens OX=9606 GN=GDI2 PE=1 SV=2
TPM4_HUMAN	Tropomyosin alpha-4 chain OS=Homo sapiens OX=9606 GN=TPM4 PE=1 SV=3
LRP8_HUMAN	Low-density lipoprotein receptor-related protein 8 OS=Homo sapiens OX=9606 GN=LRP8 PE=1 SV=4
TPST2_HUMAN	Protein-tyrosine sulfotransferase 2 OS=Homo sapiens OX=9606 GN=TPST2 PE=1 SV=1
SMCA5_HUMAN	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5 OS=Homo sapiens OX=9606 GN=SMARCA5 PE=1 SV=1

Protein Accession	Protein Description
CKAP5_HUMAN	Cytoskeleton-associated protein 5 OS=Homo sapiens OX=9606 GN=CKAP5 PE=1 SV=3
TR150_HUMAN	Thyroid hormone receptor-associated protein 3 OS=Homo sapiens OX=9606 GN=THRAP3 PE=1 SV=2
NAMPT_HUMAN	Nicotinamide phosphoribosyltransferase OS=Homo sapiens OX=9606 GN=NAMPT PE=1 SV=1
SPTN1_HUMAN	Spectrin alpha chain, non-erythrocytic 1 OS=Homo sapiens OX=9606 GN=SPTAN1 PE=1 SV=3
GCN1_HUMAN	eIF-2-alpha kinase activator GCN1 OS=Homo sapiens OX=9606 GN=GCN1 PE=1 SV=6
U5S1_HUMAN	116 kDa U5 small nuclear ribonucleoprotein component OS=Homo sapiens OX=9606 GN=EFTUD2 PE=1 SV=1
ACTN1_HUMAN	Alpha-actinin-1 OS=Homo sapiens OX=9606 GN=ACTN1 PE=1 SV=2
DSRAD_HUMAN	Double-stranded RNA-specific adenosine deaminase OS=Homo sapiens OX=9606 GN=ADAR PE=1 SV=4
ANXA1_HUMAN	Annexin A1 OS=Homo sapiens OX=9606 GN=ANXA1 PE=1 SV=2
TF3C4_HUMAN	General transcription factor 3C polypeptide 4 OS=Homo sapiens OX=9606 GN=GTF3C4 PE=1 SV=2
SORL_HUMAN	Sortilin-related receptor OS=Homo sapiens OX=9606 GN=SORL1 PE=1 SV=2
SET1A_HUMAN	Histone-lysine N-methyltransferase SETD1A OS=Homo sapiens OX=9606 GN=SETD1A PE=1 SV=3
OST48_HUMAN	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit OS=Homo sapiens OX=9606 GN=DDOST PE=1 SV=4
NEST_HUMAN	Nestin OS=Homo sapiens OX=9606 GN=NES PE=1 SV=2
IF4G1_HUMAN	Eukaryotic translation initiation factor 4 gamma 1 OS=Homo sapiens OX=9606 GN=EIF4G1 PE=1 SV=4
ACACA_HUMAN	Acetyl-CoA carboxylase 1 OS=Homo sapiens OX=9606 GN=ACACA PE=1 SV=2
SYMC_HUMAN	Methionine--tRNA ligase, cytoplasmic OS=Homo sapiens OX=9606 GN=MARS PE=1 SV=2
CAN1_HUMAN	Calpain-1 catalytic subunit OS=Homo sapiens OX=9606 GN=CAPN1 PE=1 SV=1
TKT_HUMAN	Transketolase OS=Homo sapiens OX=9606 GN=TKT PE=1 SV=3
DPYL2_HUMAN	Dihydropyrimidinase-related protein 2 OS=Homo sapiens OX=9606 GN=DPYSL2 PE=1 SV=1
TBB5_HUMAN	Tubulin beta chain OS=Homo sapiens OX=9606 GN=TUBB PE=1 SV=2
TOP2A_HUMAN	DNA topoisomerase 2-alpha OS=Homo sapiens OX=9606 GN=TOP2A PE=1 SV=3
SPTB2_HUMAN	Spectrin beta chain, non-erythrocytic 1 OS=Homo sapiens OX=9606 GN=SPTBN1 PE=1 SV=2
SC23B_HUMAN	Protein transport protein Sec23B OS=Homo sapiens OX=9606 GN=SEC23B PE=1 SV=2
RTCB_HUMAN	tRNA-splicing ligase RtcB homolog OS=Homo sapiens OX=9606 GN=RTCB PE=1 SV=1
SC31A_HUMAN	Protein transport protein Sec31A OS=Homo sapiens OX=9606 GN=SEC31A PE=1 SV=3

Protein Accession	Protein Description
2ABA_HUMAN	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform OS=Homo sapiens OX=9606 GN=PPP2R2A PE=1 SV=1
PKHA5_HUMAN	Pleckstrin homology domain-containing family A member 5 OS=Homo sapiens OX=9606 GN=PLEKHA5 PE=1 SV=1
CHTOP_HUMAN	Chromatin target of PRMT1 protein OS=Homo sapiens OX=9606 GN=CHTOP PE=1 SV=2
EXTL3_HUMAN	Exostosin-like 3 OS=Homo sapiens OX=9606 GN=EXTL3 PE=1 SV=1
NUDC_HUMAN	Nuclear migration protein nudC OS=Homo sapiens OX=9606 GN=NUDC PE=1 SV=1
PXDN_HUMAN	Peroxidasin homolog OS=Homo sapiens OX=9606 GN=PXDN PE=1 SV=2
NUFP2_HUMAN	Nuclear fragile X mental retardation-interacting protein 2 OS=Homo sapiens OX=9606 GN=NUFIP2 PE=1 SV=1
PCP_HUMAN	Lysosomal Pro-X carboxypeptidase OS=Homo sapiens OX=9606 GN=PRCP PE=1 SV=1
DHX36_HUMAN	ATP-dependent DNA/RNA helicase DHX36 OS=Homo sapiens OX=9606 GN=DHX36 PE=1 SV=2
PYR1_HUMAN	CAD protein OS=Homo sapiens OX=9606 GN=CAD PE=1 SV=3
LRC8A_HUMAN	Volume-regulated anion channel subunit LRRC8A OS=Homo sapiens OX=9606 GN=LRRC8A PE=1 SV=1
FEN1_HUMAN	Flap endonuclease 1 OS=Homo sapiens OX=9606 GN=FEN1 PE=1 SV=1
DDX17_HUMAN	Probable ATP-dependent RNA helicase DDX17 OS=Homo sapiens OX=9606 GN=DDX17 PE=1 SV=2
MCM7_HUMAN	DNA replication licensing factor MCM7 OS=Homo sapiens OX=9606 GN=MCM7 PE=1 SV=4
CAB45_HUMAN	45 kDa calcium-binding protein OS=Homo sapiens OX=9606 GN=SDF4 PE=1 SV=1
ITPR2_HUMAN	Inositol 1,4,5-trisphosphate receptor type 2 OS=Homo sapiens OX=9606 GN=ITPR2 PE=1 SV=2
IF4G2_HUMAN	Eukaryotic translation initiation factor 4 gamma 2 OS=Homo sapiens OX=9606 GN=EIF4G2 PE=1 SV=1
TCPA_HUMAN	T-complex protein 1 subunit alpha OS=Homo sapiens OX=9606 GN=TCP1 PE=1 SV=1
UGGG1_HUMAN	UDP-glucose:glycoprotein glucosyltransferase 1 OS=Homo sapiens OX=9606 GN=UGGT1 PE=1 SV=3
PDIA1_HUMAN	Protein disulfide-isomerase OS=Homo sapiens OX=9606 GN=P4HB PE=1 SV=3
PSME3_HUMAN	Proteasome activator complex subunit 3 OS=Homo sapiens OX=9606 GN=PSME3 PE=1 SV=1
MARK2_HUMAN	Serine/threonine-protein kinase MARK2 OS=Homo sapiens OX=9606 GN=MARK2 PE=1 SV=2
MUC18_HUMAN	Cell surface glycoprotein MUC18 OS=Homo sapiens OX=9606 GN=MCAM PE=1 SV=2
SYEP_HUMAN	Bifunctional glutamate/proline--tRNA ligase OS=Homo sapiens OX=9606 GN=EPRS PE=1 SV=5
LS14A_HUMAN	Protein LSM14 homolog A OS=Homo sapiens OX=9606 GN=LSM14A PE=1 SV=3
ACL6A_HUMAN	Actin-like protein 6A OS=Homo sapiens OX=9606 GN=ACTL6A PE=1 SV=1

Protein Accession	Protein Description
RS7_HUMAN 4	40S ribosomal protein S7 OS=Homo sapiens OX=9606 GN=RPS7 PE=1 SV=1
CLH1_HUMAN	Clathrin heavy chain 1 OS=Homo sapiens OX=9606 GN=CLTC PE=1 SV=5
HNRDL_HUMAN	Heterogeneous nuclear ribonucleoprotein D-like OS=Homo sapiens OX=9606 GN=HNRNPDL PE=1 SV=3
RL9_HUMAN 6	60S ribosomal protein L9 OS=Homo sapiens OX=9606 GN=RPL9 PE=1 SV=1
RUVB2_HUMAN	RuvB-like 2 OS=Homo sapiens OX=9606 GN=RUVBL2 PE=1 SV=3
6PGD_HUMAN	6-phosphogluconate dehydrogenase, decarboxylating OS=Homo sapiens OX=9606 GN=PGD PE=1 SV=3
VAT1_HUMAN	Synaptic vesicle membrane protein VAT-1 homolog OS=Homo sapiens OX=9606 GN=VAT1 PE=1 SV=2
BPTF_HUMAN	Nucleosome-remodeling factor subunit BPTF OS=Homo sapiens OX=9606 GN=BPTF PE=1 SV=3
COPG2_HUMAN	Coatomer subunit gamma-2 OS=Homo sapiens OX=9606 GN=COPG2 PE=1 SV=1
CHAP1_HUMAN	Chromosome alignment-maintaining phosphoprotein 1 OS=Homo sapiens OX=9606 GN=CHAMP1 PE=1 SV=2
IDHC_HUMAN	Isocitrate dehydrogenase [NADP] cytoplasmic OS=Homo sapiens OX=9606 GN=IDH1 PE=1 SV=2
CND3_HUMAN	Condensin complex subunit 3 OS=Homo sapiens OX=9606 GN=NCAPG PE=1 SV=1
SAHH_HUMAN	Adenosylhomocysteinase OS=Homo sapiens OX=9606 GN=AHCY PE=1 SV=4
HS90A_HUMAN	Heat shock protein HSP 90-alpha OS=Homo sapiens OX=9606 GN=HSP90AA1 PE=1 SV=5
MPRI_HUMAN	Cation-independent mannose-6-phosphate receptor OS=Homo sapiens OX=9606 GN=IGF2R PE=1 SV=3
NACAM_HUMAN	Nascent polypeptide-associated complex subunit alpha, muscle-specific form OS=Homo sapiens OX=9606 GN=NACA PE=1 SV=1
COIA1_HUMAN	Collagen alpha-1(XVIII) chain OS=Homo sapiens OX=9606 GN=COL18A1 PE=1 SV=5
NONO_HUMAN	Non-POU domain-containing octamer-binding protein OS=Homo sapiens OX=9606 GN=NONO PE=1 SV=4
MYH9_HUMAN	Myosin-9 OS=Homo sapiens OX=9606 GN=MYH9 PE=1 SV=4
FA98A_HUMAN	Protein FAM98A OS=Homo sapiens OX=9606 GN=FAM98A PE=1 SV=2
HSP7C_HUMAN	Heat shock cognate 71 kDa protein OS=Homo sapiens OX=9606 GN=HSPA8 PE=1 SV=1
PARP1_HUMAN	Poly [ADP-ribose] polymerase 1 OS=Homo sapiens OX=9606 GN=PARP1 PE=1 SV=4
HUWE1_HUMAN	E3 ubiquitin-protein ligase HUWE1 OS=Homo sapiens OX=9606 GN=HUWE1 PE=1 SV=3
SYDC_HUMAN	Aspartate--tRNA ligase, cytoplasmic OS=Homo sapiens OX=9606 GN=DARS PE=1 SV=2
PRKDC_HUMAN	DNA-dependent protein kinase catalytic subunit OS=Homo sapiens OX=9606 GN=PRKDC PE=1 SV=3
IPO5_HUMAN	Importin-5 OS=Homo sapiens OX=9606 GN=IPO5 PE=1 SV=4

Protein Accession	Protein Description
NOL6_HUMAN	Nucleolar protein 6 OS=Homo sapiens OX=9606 GN=NOL6 PE=1 SV=2
DDB1_HUMAN	DNA damage-binding protein 1 OS=Homo sapiens OX=9606 GN=DDB1 PE=1 SV=1
ZN207_HUMAN	BUB3-interacting and GLEBS motif-containing protein ZNF207 OS=Homo sapiens OX=9606 GN=ZNF207 PE=1 SV=1
MDHM_HUMAN	Malate dehydrogenase, mitochondrial OS=Homo sapiens OX=9606 GN=MDH2 PE=1 SV=3
PTK7_HUMAN	Inactive tyrosine-protein kinase 7 OS=Homo sapiens OX=9606 GN=PTK7 PE=1 SV=2
CSDE1_HUMAN	Cold shock domain-containing protein E1 OS=Homo sapiens OX=9606 GN=CSDE1 PE=1 SV=2
RBP2_HUMAN	E3 SUMO-protein ligase RanBP2 OS=Homo sapiens OX=9606 GN=RANBP2 PE=1 SV=2
RL38_HUMAN	60S ribosomal protein L38 OS=Homo sapiens OX=9606 GN=RPL38 PE=1 SV=2
SF3B1_HUMAN	Splicing factor 3B subunit 1 OS=Homo sapiens OX=9606 GN=SF3B1 PE=1 SV=3
ATG9A_HUMAN	Autophagy-related protein 9A OS=Homo sapiens OX=9606 GN=ATG9A PE=1 SV=3
ROA1_HUMAN	Heterogeneous nuclear ribonucleoprotein A1 OS=Homo sapiens OX=9606 GN=HNRNPA1 PE=1 SV=5
HNRH1_HUMAN	Heterogeneous nuclear ribonucleoprotein H OS=Homo sapiens OX=9606 GN=HNRNPH1 PE=1 SV=4
AP2B1_HUMAN	AP-2 complex subunit beta OS=Homo sapiens OX=9606 GN=AP2B1 PE=1 SV=1
M4K4_HUMAN	Mitogen-activated protein kinase kinase kinase kinase 4 OS=Homo sapiens OX=9606 GN=MAP4K4 PE=1 SV=2
FLII_HUMAN	Protein flightless-1 homolog OS=Homo sapiens OX=9606 GN=FLII PE=1 SV=2
RL10A_HUMAN	60S ribosomal protein L10a OS=Homo sapiens OX=9606 GN=RPL10A PE=1 SV=2
S4A7_HUMAN	Sodium bicarbonate cotransporter 3 OS=Homo sapiens OX=9606 GN=SLC4A7 PE=1 SV=2
FAS_HUMAN	Fatty acid synthase OS=Homo sapiens OX=9606 GN=FASN PE=1 SV=3
MD2L1_HUMAN	Mitotic spindle assembly checkpoint protein MAD2A OS=Homo sapiens OX=9606 GN=MAD2L1 PE=1 SV=1
TOP2B_HUMAN	DNA topoisomerase 2-beta OS=Homo sapiens OX=9606 GN=TOP2B PE=1 SV=3
ITA6_HUMAN	Integrin alpha-6 OS=Homo sapiens OX=9606 GN=ITGA6 PE=1 SV=5
PDC6I_HUMAN	Programmed cell death 6-interacting protein OS=Homo sapiens OX=9606 GN=PDCD6IP PE=1 SV=1
MYH14_HUMAN	Myosin-14 OS=Homo sapiens OX=9606 GN=MYH14 PE=1 SV=2
KALM_HUMAN	Anosmin-1 OS=Homo sapiens OX=9606 GN=ANOS1 PE=1 SV=3
MAP2_HUMAN	Methionine aminopeptidase 2 OS=Homo sapiens OX=9606 GN=METAP2 PE=1 SV=1
ECM29_HUMAN	Proteasome adapter and scaffold protein ECM29 OS=Homo sapiens OX=9606 GN=ECPAS PE=1 SV=2

Protein Accession	Protein Description
OTUD4_HUMAN	OTU domain-containing protein 4 OS=Homo sapiens OX=9606 GN=OTUD4 PE=1 SV=4
L1CAM_HUMAN	Neural cell adhesion molecule L1 OS=Homo sapiens OX=9606 GN=L1CAM PE=1 SV=2
GLRX3_HUMAN	Glutaredoxin-3 OS=Homo sapiens OX=9606 GN=GLRX3 PE=1 SV=2
ANKH1_HUMAN	Ankyrin repeat and KH domain-containing protein 1 OS=Homo sapiens OX=9606 GN=ANKHD1 PE=1 SV=1
EPIPL_HUMAN	Epiplakin OS=Homo sapiens OX=9606 GN=EPPK1 PE=1 SV=3
SYNC_HUMAN	Asparagine--tRNA ligase, cytoplasmic OS=Homo sapiens OX=9606 GN=NARS PE=1 SV=1
MSH2_HUMAN	DNA mismatch repair protein Msh2 OS=Homo sapiens OX=9606 GN=MSH2 PE=1 SV=1
PSME2_HUMAN	Proteasome activator complex subunit 2 OS=Homo sapiens OX=9606 GN=PSME2 PE=1 SV=4
TCPB_HUMAN	T-complex protein 1 subunit beta OS=Homo sapiens OX=9606 GN=CCT2 PE=1 SV=4
MET_HUMAN	Hepatocyte growth factor receptor OS=Homo sapiens OX=9606 GN=MET PE=1 SV=4
ARP3_HUMAN	Actin-related protein 3 OS=Homo sapiens OX=9606 GN=ACTR3 PE=1 SV=3
TB182_HUMAN	182 kDa tankyrase-1-binding protein OS=Homo sapiens OX=9606 GN=TNKS1BP1 PE=1 SV=4
PURA2_HUMAN	Adenylosuccinate synthetase isozyme 2 OS=Homo sapiens OX=9606 GN=ADSS PE=1 SV=3
DDX54_HUMAN	ATP-dependent RNA helicase DDX54 OS=Homo sapiens OX=9606 GN=DDX54 PE=1 SV=2
MCAF1_HUMAN	Activating transcription factor 7-interacting protein 1 OS=Homo sapiens OX=9606 GN=ATF7IP PE=1 SV=3
MINT_HUMAN	Msx2-interacting protein OS=Homo sapiens OX=9606 GN=SPEN PE=1 SV=1
EHD4_HUMAN	EH domain-containing protein 4 OS=Homo sapiens OX=9606 GN=EHD4 PE=1 SV=1
IF4B_HUMAN	Eukaryotic translation initiation factor 4B OS=Homo sapiens OX=9606 GN=EIF4B PE=1 SV=2
NUP62_HUMAN	Nuclear pore glycoprotein p62 OS=Homo sapiens OX=9606 GN=NUP62 PE=1 SV=3
RPN2_HUMAN	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2 OS=Homo sapiens OX=9606 GN=RPN2 PE=1 SV=3
PSPC1_HUMAN	Paraspeckle component 1 OS=Homo sapiens OX=9606 GN=PSPC1 PE=1 SV=1
UBP5_HUMAN	Ubiquitin carboxyl-terminal hydrolase 5 OS=Homo sapiens OX=9606 GN=USP5 PE=1 SV=2
SEPT9_HUMAN	Septin-9 OS=Homo sapiens OX=9606 GN=SEPTIN9 PE=1 SV=2
TIF1B_HUMAN	Transcription intermediary factor 1-beta OS=Homo sapiens OX=9606 GN=TRIM28 PE=1 SV=5
TBB2B_HUMAN	Tubulin beta-2B chain OS=Homo sapiens OX=9606 GN=TUBB2B PE=1 SV=1
RBM14_HUMAN	RNA-binding protein 14 OS=Homo sapiens OX=9606 GN=RBM14 PE=1 SV=2

Protein Accession	Protein Description
EIF3D_HUMAN	Eukaryotic translation initiation factor 3 subunit D OS=Homo sapiens OX=9606 GN=EIF3D PE=1 SV=1
NUP93_HUMAN	Nuclear pore complex protein Nup93 OS=Homo sapiens OX=9606 GN=NUP93 PE=1 SV=2
KPRP_HUMAN	Keratinocyte proline-rich protein OS=Homo sapiens OX=9606 GN=KPRP PE=1 SV=1
RS10_HUMAN	40S ribosomal protein S10 OS=Homo sapiens OX=9606 GN=RPS10 PE=1 SV=1
IF2A_HUMAN	Eukaryotic translation initiation factor 2 subunit 1 OS=Homo sapiens OX=9606 GN=EIF2S1 PE=1 SV=3
LRRN1_HUMAN	Leucine-rich repeat neuronal protein 1 OS=Homo sapiens OX=9606 GN=LRRN1 PE=1 SV=1
SE1L3_HUMAN	Protein sel-1 homolog 3 OS=Homo sapiens OX=9606 GN=SEL1L3 PE=1 SV=2
SYLC_HUMAN	Leucine--tRNA ligase, cytoplasmic OS=Homo sapiens OX=9606 GN=LARS PE=1 SV=2
BGAL_HUMAN	Beta-galactosidase OS=Homo sapiens OX=9606 GN=GLB1 PE=1 SV=2
PLAK_HUMAN	Junction plakoglobin OS=Homo sapiens OX=9606 GN=JUP PE=1 SV=3
SF3B2_HUMAN	Splicing factor 3B subunit 2 OS=Homo sapiens OX=9606 GN=SF3B2 PE=1 SV=2
PLBL1_HUMAN	Phospholipase B-like 1 OS=Homo sapiens OX=9606 GN=PLBD1 PE=1 SV=2
MA2A1_HUMAN	Alpha-mannosidase 2 OS=Homo sapiens OX=9606 GN=MAN2A1 PE=1 SV=2
GTF2I_HUMAN	General transcription factor II-I OS=Homo sapiens OX=9606 GN=GTF2I PE=1 SV=2
HSPB1_HUMAN	Heat shock protein beta-1 OS=Homo sapiens OX=9606 GN=HSPB1 PE=1 SV=2
CSPG2_HUMAN	Versican core protein OS=Homo sapiens OX=9606 GN=VCAN PE=1 SV=3
LICH_HUMAN	Lysosomal acid lipase/cholesteryl ester hydrolase OS=Homo sapiens OX=9606 GN=LIPA PE=1 SV=2
NSUN5_HUMAN	Probable 28S rRNA (cytosine-C(5))-methyltransferase OS=Homo sapiens OX=9606 GN=NSUN5 PE=1 SV=2
BIP_HUMAN E	Endoplasmic reticulum chaperone BiP OS=Homo sapiens OX=9606 GN=HSPA5 PE=1 SV=2
GANAB_HUMAN	Neutral alpha-glucosidase AB OS=Homo sapiens OX=9606 GN=GANAB PE=1 SV=3
RBMX_HUMAN	RNA-binding motif protein, X chromosome OS=Homo sapiens OX=9606 GN=RBMX PE=1 SV=3
LARP1_HUMAN	La-related protein 1 OS=Homo sapiens OX=9606 GN=LARP1 PE=1 SV=2
S23IP_HUMAN	SEC23-interacting protein OS=Homo sapiens OX=9606 GN=SEC23IP PE=1 SV=1
TMTC3_HUMAN	Protein O-mannosyl-transferase TMTC3 OS=Homo sapiens OX=9606 GN=TMTC3 PE=1 SV=2
PLD3_HUMAN	Phospholipase D3 OS=Homo sapiens OX=9606 GN=PLD3 PE=1 SV=1
FDFT_HUMAN	Squalene synthase OS=Homo sapiens OX=9606 GN=FDFT1 PE=1 SV=1
K1C17_HUMAN	Keratin, type I cytoskeletal 17 OS=Homo sapiens OX=9606 GN=KRT17 PE=1 SV=2
ERF1_HUMAN	Eukaryotic peptide chain release factor subunit 1 OS=Homo sapiens OX=9606 GN=ETF1 PE=1 SV=3

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

This dissertation has described the development of two variations of the Dynamic IDAWG method: released glycans and intact O-GlcNAc modified proteins. The original IDAWG method laid the foundation for the quantification of released glycans in cell culture also allowing for comparative studies.[1] That groundwork opened the door to the development of the methodology Dynamic IDAWG that allows for turnover rates of released glycans and intact O-GlcNAc to be investigated. The conclusions and future directions herein will focus on either variation of the Dynamic IDAWG as separate entities.

Determination of Half-Life and Percent Recycling of Released Glycans

In chapter 2 we first discussed the development of Dynamic IDAWG for released glycans of whole cell proteins. Initially, our focus was O-glycans specifically where we completed comparative studies between hES and hED cells where we noticed variations in the abundance of complex sialylated and poly-LacNAc structures. In these original experiments we were able to differentially label cells with either ^{14}N or ^{15}N from supplemented Gln, quantify present O-glycans via mass spectrometry, and compare the abundances of those structures by analyzing combined 50:50 sample mixtures.

Knowing that we have the ability to do comparative studies such as this, we moved onto development of the true Dynamic IDAWG method. This method combines the same labeling technique in cell culture with ^{15}N-Gln, however we then transition to a pulse-chase experiment where the cell culture media is changed over to ^{14}N-Gln supplemented media after the initial collection heavy labeled cells at time 0hr. Remaining cells were then collected at the predetermined time points. Samples went through a released glycan protocol and subjected to mass spectrometric analysis.

Through the use of bioinformatic tools developed in house in Microsoft Excel we have the ability to easily extract the isotopic distributions of the identified O-glycans, average those distributions across replicates, normalize, and calculate standard deviation. Completing these calculations for each glycan overtime and factoring in the isotopic pattern of the unlabeled structure we are able to accurately calculate the half-life of each as well as determine the amount of recycling that is occurring for sialylated structures. While there has been debate surrounding whether or not sialylated structures experience remodeling or recycling based on the inconsistencies with the shift in isotopic distribution from heavy to light in sialylated structures. The MS2 data that was presented for the disialylated t-antigen suggests that recycling is what is occurring. The fragmentation data reveals that the core GalNAc of the glycan is light, while the sialic acids remain labeled with 15N due to the recycling of those sialic acids back onto newly generated glycans.

Dynamic IDAWG for released glycans gives us and the glycobiology field the ability to easily follow the turnover of the O-glycans identified in a cell line of interest. With this method it is also feasible to complete comparative and quantitative studies of different cell lines and identifying variations in turnover rates. Accurately analyzing N-linked glycans is a work in progress though. Through our analyses we were also able to identify released N-glycans from hES cells, these included a handful of complex N-glycans while most of the identified structures were high mannose. However, based on the half-life calculations there was recycling occurring at a higher-than-expected rate in the chitobiose core in a subset of the high mannose structures (Man-9 through Man-5). This may reflect the UDP-GlcNAc cytosolic pool instead due to the core of N-glycans being built up on dolichol in the ER which is cytosolic facing. Further investigation would need to be completed to better analyze N-glycans using the Dynamic IDAWG method.

Dynamics of O-GlcNAc Modified Proteins

The glycobiology field has been in need of methods capable of following the unique and dynamic modification O-GlcNAc that are also high throughput and easily accessible. Chapter 3 offers up a Dynamic IDAWG method, specifically designed to evaluate the dynamics of O-GlcNAc and the proteins in which they modify to address this gap in the field. Similar to the released glycan method, Dynamic IDAWG for O-GlcNAc modified proteins is capable of virtually complete labeling of O-GlcNAc moieties. Additionally, because our labeling technique is through the supplementation of ^{15}N-Gln we are simultaneously labeling the proteins within the cells as well. Here we can follow the changes in the relative

abundances in the MS^2 of the unlabeled (204) and heavy labeled (205) peaks for the O-GlcNAc modification or the corresponding peaks to the Gln containing fragment ion.

The experiments to define this method were completed across four biological replicates in RUES1 WT cells. On average we identified approximately 900 unique O-GlcNAc modified peptides; this list did include variations of a single peptide where the site of the modification varied based on fragmentation of the peptide. Taking the O-GlcNAc modified peptides we found that they modified over 300 proteins (refer back to **Supplemental Table 3.1**) We narrowed down our search results to O-GlcNAc and Gln containing peptides with sufficient fragmentation data and were identified in a majority of the time points in at least three of the four replicates. Out of those we selected a list of 17 total peptides, with Gln containing peptides correlating to the proteins that contained multiple O-GlcNAc sites, HCFC1 and NU214. While we did observe quite a bit of variation in the calculated half-lives across the selected sites and proteins, it was most interesting to see the degree in which half-lives varied on the HCFC1 protein. We selected 5 O-GlcNAc sites out of those identified across the replicates. Of those 5, 2 were found to be dynamic exhibiting a shorter half-life than the protein, based on the half-life calculations for the Gln containing peptide. The remaining 3 O-GlcNAc sites were found to be static with calculated half-lives either almost identical to the protein or excessively long half-lives which would simply turnover with the protein (refer back to **Table 3.1**).

O-GlcNAc's involvement in a number of cellular processes and functions, neural development, and an array of diseases is of high interest. Utilizing the Dynamic IDAWG O-GlcNAc method would allow researchers to investigate how fluctuations in turnover rates of specific O-GlcNAc sites on proteins of interest could aid in further understanding of its roles in disease and cellular functions. While we do have access to the same Microsoft Excel analysis tools used for the released glycan studies to aid in analysis and half-life calculations, due to the number of potential sites it would be advantageous to eventually define a more automated solution to identifying O-GlcNAc site candidates across biological replicates and each time point.

Future Directions

Recommendations for Further Released Glycan Analyses with Dynamic IDAWG

The work executed in this dissertation for released glycans was exclusively completed on whole cell protein and more specifically in hES cells. Defining the Dynamic IDAWG method for released glycans allows for broader studies to be done on any cell line of interest in addition to comparative studies between cell lines. Whole cell analysis does allow for us to see the bigger picture regarding the most abundant glycans, quantifying those structures, and defining the overall glycome of a cell line. Rather than looking at the entire glycome, we can instead narrow our focus.

Continuing to explore the dynamics of O-GlcNAc moieties in specific cell lines we can turn our sights to secreted proteins or simply individual proteins of

interest.[2] Secreted proteins maintain control over different physiological processes and must be regulated to do so; Where nearly all secreted proteins possess post translational modifications from when they were biosynthesized and passed through the secretory pathway.[3] Intracellular and intercellular communication is carried out by these secreted proteins, where this information can regard differentiation or even the need to release specific compounds to name a few.[4] Analyzing the half-life and determining the amount of recycling that may occur with this collection of the glycans associated with secreted proteins culminated with the whole cell analysis would better define the dynamics of the cell line as a whole. Similar to studies shown in chapter 2, we have the ability to compare cell lines in the same fashion. Incorporating cellular stressors into experiments would allow for observation of overall glycosylation regarding deviations in the profile, relative abundances, and potentially changes to turnover.

Recommendations for Future Studies of O-GlcNAc Dynamics

We have only begun delving into the dynamics of O-GlcNAc by defining this second Dynamic IDAWG method. Presented here was a subset of possible O-GlcNAc sites that were analyzed. While it is a daunting task, the dynamics of this unique modification can potentially be defined for the O-GlcNAcome as a whole. Defining dynamics of specific sites aids in future comparative studies as well in a cell line of choice whether its to investigate its roles in cellular function further or if its dynamics is an additional variable in diseases such as Alzheimer's.

As mentioned in chapter 3 our studies focus on utilizing the MS2 spectra rather than the MS1 in order to define the dynamics of individual sites. Typically

one would evaluate the shift in the isotopic peak distribution in the MS1 similar to the released glycan method, analyzing the MS2 spectra actually gives Dynamic IDAWG an advantage for O-GlcNAc modified proteins. While it can be dependent on fragmentation data, the MS2 spectra would actually allow for greater depth of analysis as you can determine the dynamics of multiple sites on a given peptide looking at the individual fragment ions instead of the 204/205 peaks. This is true also for peptides containing Gln residues. Overall, Dynamic IDAWG allows for more versatility.

One challenge this method does face is the amount of “sticky” proteins that may elute off of the sWGA with proteins that were actually modified with O-GlcNAc. While we still identified a large pool of modified proteins based on the PMI-Byonic searches of the MS data, there were enough unmodified proteins that could impact identification of less abundant modified proteins in our sample. No enrichment technique for O-GlcNAc is perfect though. Standard wheat germ agglutinin (WGA) also enriches for O-GlcNAc, but it does also bind to sialylated structures and has a weaker affinity.[5,6] Clostridium perfringens NagJ (CpOGA) has demonstrated itself as an effective enrichment tool.[7] Original enrichment studies for the development of the Dynamic IDAWG method were even completed using this tool. However, our goal with the development of this technique was to make it easily accessible to researchers and CpOGA is not readily available to purchase in the same way sWGA is and we become reliant on adequate protein expression and purification yield to complete your enrichments. Having also completed multiple side by side enrichments with CpOGA and sWGA, we were also able to get more

consistent results with sWGA. While all of these enrichment techniques are done at the protein level, the ability to enrich at the peptide level would be more efficient and specific, however it is something we as a field are still lacking.

As mentioned, defining this method to investigate the dynamics of O-GlcNAc moieties was accomplished in a male human embryonic stem cell line, RUES1 WT. While the methodology of Dynamic IDAWG for O-GlcNAc dynamics was completed using this cell line, it could provide baseline data for studying the impact mutations within OGT may have on the cycling of specific O-GlcNAc modifications on proteins of interest. Roughly 1-3% of the human population suffers from an intellectual disability8. Out of that population 10% of males diagnosed with an intellectual disability are considered to have an X-Linked Intellectual Disability (XLID).[8–10] XLID results from genetic anomalies in the X-chromosome and is diagnosed by an IQ below 70 and difficulties in adaptive behaviors9. Several missense mutations in OGT have been identified in patients in the tetratricopeptide repeat (TPR) domain and have shown to be causal for XLID potentially impacting protein-protein interactions and causing the phenotypes observed: L254F, A259T, R284P, A319T, and E339G (**Fig. 4.1**).[8,10–13] Mutations within OGT are not exclusively reserved to the TPR domain, additional XLID variants have been identified in the last couple of years within the catalytic domain of the enzyme as well.[14,15] As OGT is responsible for the addition of the O-GlcNAc modification investigating potential changes to the dynamics of specific sites is of particular interest. Having access to CRISPR/Cas9 engineered RUES1 cell lines of the aforementioned TPR domain variants allows for comparative studies to be

completed and could be focused on proteins specifically related to neural development. Characterizing changes or lack of changes in the cycling rates of specific O-GlcNAc modifications will hopefully shed light on if dynamics plays a role in causing the phenotypes observed in XLID patients.

In summary, we have presented here a versatile in vitro methodology that can be utilized by the field. Dynamic IDAWG can be utilized for released glycans from whole cell proteins to define the turnover rates of individual glycans, as well as compare rates and quantify across different cell lines. Additional studies can be done to investigate glycans specific to secreted proteins. Finally, this method offers the opportunity to define the dynamics of O-GlcNAc modifications on a larger scale while investigating the relationship between that and the protein's turnover rate. This opens the door to exploring the part O-GlcNAc plays in various disease states, protein function, cellular function and more.

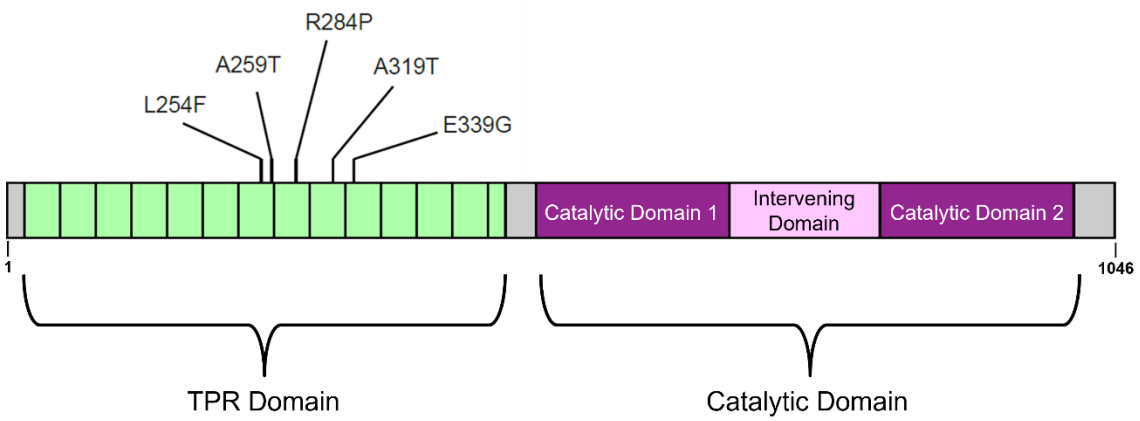


Figure 4.1: The O-GlcNAc Transferase. Representation of OGT and the domains that define it. Consisting of 13.5 tetratricopeptide repeats in the TPR domain followed by the catalytic domain. TPR variants associated with XLID are labeled.

References

1. Orlando, R. et al. IDAWG: Metabolic incorporation of stable isotope labels for quantitative glycomics of cultured cells. J. Proteome Res. 8, 3816–3823 (2009).

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

3. Yang, W. H. et al. An intrinsic mechanism of secreted protein aging and turnover. Proc. Natl. Acad. Sci. U. S. A. 112, 13657–13662 (2015).

4. Klein, R. D., Gu, Q., Goddard, A. & Rosenthal, A. Selection for genes encoding secreted proteins and receptors. Proc. Natl. Acad. Sci. U. S. A. 93, 7108–7113 (1996).

5. Thompson, J. W., Griffin, M. E. & Hsieh-wilson, L. C. Methods for the Detection , Study , and Dynamic Profiling of O-GlcNAc Glycosylation. Chemical Glycobiology Part B. Monitoring Glycans and their Interactions 598, (Elsevier Inc., 2018).

6. Maynard, J. C. & Chalkley, R. J. Methods for enrichment and assignment of N-acetylglucosamine modification sites. Mol. Cell. Proteomics 20, 1–28 (2021).

7. Selvan, N. et al. A mutant O-GlcNAcase enriches Drosophila developmental regulators. Nat. Chem. Biol. 13, 882–887 (2017).

8. Selvan, N. et al. O-GlcNAc transferase missense mutations linked to X-linked intellectual disability deregulate genes involved in cell fate determination and signaling. J. Biol. Chem. In Press, (2018).

9. Lubs, H. A., Stevenson, R. E. & Schwartz, C. E. Fragile X and X-Linked Intellectual Disability : Four Decades of Discovery. Am. J. Hum. Genet. 90, 579–590 (2012).

10. Pravata, V. M. et al. An intellectual disability syndrome with single-nucleotide variants in O-GlcNAc transferase. Eur. J. Hum. Genet. 28, 706–714 (2020).

11. Willems, A. P. et al. Mutations in N-acetylglucosamine (O-GlcNAc) transferase in patients with X-linked intellectual disability. J. Biol. Chem. 292, 12621–12631 (2017).

12. Bouazzi, H., Lesca, G., Trujillo, C., Alwasiyah, M. K. & Munnich, A. Nonsyndromic X-linked intellectual deficiency in three brothers with a novel MED12 missense mutation [c.5922G>T (p.Glu1974His)]. Clin. Case Reports 3, 604–609 (2015).

13. Vaidyanathan, K. et al. Identification and characterization of a missense mutation in the O-linked β-N-acetylglucosamine (O-GlcNAc) transferase gene that segregates with X-linked intellectual disability. J. Biol. Chem. 292, 8948–8963 (2017).

14. Pravata, V. M. et al. Catalytic deficiency of O-GlcNAc transferase leads to X-linked intellectual disability. Proc. Natl. Acad. Sci. U. S. A. 116, 14961–14970 (2019).

15. Pravata, V. M. et al. A missense mutation in the catalytic domain of O-GlcNAc transferase links perturbations in protein O-GlcNAcylation to X-linked intellectual disability. FEBS Lett. 594, 717–727 (2020).