APPLICATIONS OF THE IDAWG TECHNIQUE TO QUANTITATIVE GLYCOMICS OF HUMAN EMBRYONIC STEM CELLS

by

MENG FANG

(Under the Direction of LANCE WELLS)

ABSTRACT

As the first reported *in-vivo* stable isotopic labeling strategies for quantitative glycomics, the IDAWG (Isotopic Detection of Aminosugars With Glutamine) takes advantage of the hexosamine biosynthetic pathway that provides the nitrogen to all aminosugars from the side chain of glutamine (Gln). As a result, if the cells are fed with Gln-free media and glutamine with a ^{15}N-labeled side chain is added, all the aminosugars produced in cells will be labeled with ^{15}N and thus the mass of all glycan structures will be shifted by approximately +1 dalton per aminosugar. In this work, we demonstrate the utility of the IDAWG technique to study both static and dynamic glycomics of human embryonic stem cells.

To compare N-linked and O-linked glycan expression levels in hESCs and hDE and to quantify the changes in glycan expression that accompany differentiation of hESCs into hDEs, four cell populations including light and heavy hESCs, light and heavy hDEs have been processed and N- and O-linked glycans have been released. After the permethylated glycan samples were analyzed by ESI-LTQ-Orbitrap, more than 20 O-linked glycan structures and 90 N-linked glycan structures were assigned in hESCs and

hDEs. Further, these glycans were quantified to show the prevalence of each structure in the total glycan profiles. Based on the mass spectrometry data of mixed samples of hES and hDE, glycan expression level changes were quantified for a subset of the glycans.

To apply the IDAWG technique to assess the dynamics of turnover and synthesis for individual glycans and to provide a dynamic view of the glycome, a pulse-chase labeling of hESCs was performed. Heavy labeled cells were shifted to light media for 0, 6, 12, 24 or 36 hr before harvesting, isolation and permethylation. Analysis of these samples allowed us to approximate the turnover and synthesis rate of the glycans. Interestingly, by comparing the glycan spectra of pulse-chase samples with the spectra of real mixtures of light and heavy sample, we noticed an isotope mixture that could only be explained by remodeling of the glycans and were able to approximate the amount of remodeling occurring for a subset of structures.

INDEX WORDS: IDAWG, stable isotopic labeling, quantitative glycomics, cell culture, glycan analysis, human embryonice stem cells, human definitive endoderms, Pulse-Chase, dynamics, glycan turnover, remodeling

APPLICATIONS OF THE IDAWG TECHNIQUE TO QUANTITATIVE GLYCOMICS OF HUMAN EMBRYONIC STEM CELLS

by

MENG FANG

B.S. University of Science and Technology of China, China, 2005

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2011

APPLICATIONS OF THE IDAWG TECHNIQUE TO QUANTITATIVE GLYCOMICS OF HUMAN EMBRYONIC STEM CELLS

by

MENG FANG

Major Professor: Lance Wells

Committee: I. Jonathan Amster
Ron Orlando

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2011

DEDICATION

To my mother and father, thank you for your unconditional love and support throughout my life. Words cannot express my gratitude for your sacrifice. Without your encouragement, I would not be able to achieve anything. I love you all.

To my grandfather, who passed away in China while I was in the U.S. May you rest in peace. I miss you and love you forever and always.

ACKNOWLEDGEMENTS

First and most importantly, I would like to express my gratitude and appreciation to my advisor, Dr. Lance Wells, for giving me the opportunity to study and work in his laboratory with an extraordinary environment. I am really grateful to his constant guidance, support and encouragement throughout all the years of my graduate school.

I would like to thank my committee, Dr. I. Jonathan Amster and Dr. Ron Orlando, for their invaluable time and expertise. I would also like to thank Dr. Will York, Dr. John Miller, Dr. Michael Tiemeyer, Dr. Kazuhiro Aoki and Jun Han for their countless help.

Many thanks go to the past and present members of the Wells' laboratory: Jae-min Lim, Enas Gad Elkarim, Chin Fen Teo, Stephanie Hammond, Peng Zhao, Anu Rajesh Koppikar, Sandii Brimble, Edith Hayden, Krithika Vaidyanathan, Sean Durning, Seongha Park, Ryan Stuart, Jeremy Praissman, Robert Bridger and Crissy Dobson for your kindness and assistance. Best wishes in their future endeavors.

Last but not least, I would like to thank my family, my girlfriend and all my friends for their love and support.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 GLYCOSYLATION

Post-translational modifications (PTMs) are covalent processing events that change the properties of a protein by proteolytic cleavage or by addition of a modifying group to one or more amino acids. Almost all proteins are modified by during or after synthese by cleavage or chemical modifications. Post-translational modification regulates protein activity and function and significantly increases the diversity and complexity of the proteome. The majority of protein-based biopharmaceuticals approved or in clinical trials bears some form of post-translational modification, which can profoundly affect protein properties relevant to their therapeutic application[1,2]. Several hundred different forms of PTMs are known and glycosylation, which is a process of covalent attachment of glycan structures on protein backbones, is one of the most common post-translational modifications of proteins[3-5]. The glycans are complex carbohydrates that are oligosaccharides chains usually linked to proteins and lipids[6,7]. Generally, glycans are built up by monosaccharide residues including N-acetylgluctosamine (GlcNAc), N-acetylgalactosamine (GalNAc); glucose (Glc), galactose (Gal), mannose (Man), fucose (Fuc) and N-acetylneuraminic acid (Neu5Ac, sialic acids). There are five main ways that carbohydrate chains are linked to protein backbones: 1. N-linked glycan is linked through the side chain of an asparagine residue present in the tripeptide consensus sequence – Asn-X-Thr/Ser (where X can be any amino acid except proline). In some rare cases, N-glycosylation can occur even when the serine or threonine residue is replaced with a cysteine as Asn-X-Cys[8-12]. N-linked

glycans can be classified into three types based on compositions of different branches and they are high mannose structures, hybrid structures and complex structures. 2. O-linked glycan is attached to the oxygen on the side chain of a serine or threonine and there is no consensus amino acid sequence for O-glycosylation sites[13-15]. O-linked glycans can be categorized as several core structures such as core 1, core 2, core 3 and core 4 structures. 3. Glycophosphatidylinositol (GPI) anchors link lipids to the carboxyl terminus of proteins and serve to anchor these proteins to cell membranes[16-18]. 4. Glycosaminoglycans (GAGs) are long unbranched polysaccharides containing a repeating disaccharide unit, which are linked to the hydroxy oxygen of serine[19,20]. 5. A mannose sugar attaches to the carbon of tryptophan residues in Thrombospondin repeats, which is called C-mannosylation[21-23].

It has been estimated that glycosylation is occurring in up to half of all gene products[24] and glycosylated proteins account for approximately 60-80% of all mammalian proteins at some point during their existence and nearly 100% of all membrane and secreted proteins[25]. Glycans often play critical roles in various physiological processes such as cell development and differentiation,[26-31] cell-cell recognition and interaction,[32,33] tumor growth and metastasis,[34-39] and immune recognition and response[40-44]. A correlation between glycosyltransferase expression levels and subsequent changes in glycan structures has been reported in several diseases, including inherited diseases[45], the progression of cancer[46] and autoimmune diseases[47-49]. The recycling of glycoproteions has also been reported and thus the remodeling of glycans becomes possible. In 1986, the recycling of cellular glycoproteins to the site of Golgi mannosidase I, an enzyme of asparagine-linked oligosaccharide synthesis, was

studied in K562 human erythroleukemia cells by metabolically labeling the cells in the presence of deoxymannojirimycin, a reversible inhibitor of Golgi mannosidase I[50]. This study of membrane traffic revealed the recycling of the total N-linked glycoprotein pool through the mannosidase I compartment and implied that the recycling of glycoproteins might allow the posttranslational modifications to be continuously repaired. The recycling of glycoproteins into the *trans*-Golgi compartment that contains sialyltransferase, Galactosyltransferase and the early Golgi region that contains α-mannosidase I but yet *cis*-Golgi has been demonstrated[51,52]. In 1998, the study of N-glycans demannosylation of cell surface glycoproteins revealed that cell surface glycoproteins can be modified by mannosidases without further processing their glycans to the complex form and this glycan remodeling can occur at the cell surface or during endocytosis and recycling back to the cell surface[53].

1.2 GLYCOMICS

Just following the naming convention established by genomics, which deals with genes and proteomics, which deals with proteins, the term glycomics covers a range of scientific disciplines that are applied to study the structure and function of carbohydrates in biological systems. As a comprehensive study of all glycan structures in a given biological system, the glycomics has drawn more and more attentions[54,55]. In contrast to nucleic acids and proteins, the study of glycans is much more challenging due to the complexity of glycan structures and diversity of in vivo glycosylation. First of all, the biosynthesis of glycans is not template-driven. Instead it is regulated by certain glycotransferases, glycosidase and carbohydrate enzymes, which are encoded from a small portion of genes. As the post-genomic products, the structures of glycans are not

predictable. Second, several qualities of glycans have given rise to the enormous structural heterogeneity and these facts include: constituent carbohydrate isomers, like glucose (Glc), galactose (Gal), mannose (Man); distinct arrangements of the same components; multiply linkage points; α or β anomeric linkage configurations; and a wide range of branches. For example, the chemical structures of an oligosaccharide composed by six monosaccharides could theoretically be more than 1×10^{12}. Last but not least, additions of different glycan chains onto the proteins that have several various susceptible sites to glycosylation would lead to a considerable number of glycoforms[32,55,56]. These fundamental and interrelated aspects of glycans make the field of glycomics challenging and render the lagging phase of glycomics behind the genomics and proteomics.

The study of glycomics is challenging yet important. Protein glycosylation occurs in the endoplasmic reticulum (ER) and Golgi compartments of the cell and involves membrane-bound glycosyltransferases and glycosidases, most of which are sensitive to the surrounding environment. As affected by different factors, the expression of glycans in a biological system can vary with its species, tissue, developmental stage and even the genetic and physiological state and it has also been noticed that altered carbohydrates expression is a common feature of many types of cancers[37]. For example, glycosylation alteration has been identified in some systemic genetic disease like congenital disorder of glycosylation (CDG) syndrome, caused by genetic defects affecting the activity of specific glycotransferases in biosynthesis pathways of glycans[57]. Besides, particular carbohydrate modifications may be altered more specifically or frequently than their underlying core protein in certain disease state, which is a potential advantage of using glycans for diagnostics[58,59]. With the awareness of the different expression levels of

glycosylation in diseases, extensive biomedical and biological applications have been opened up for glycomics[60]. Biomarker discovery for diagnosis and prognosis, new glycan therapeutics have become the frontier in glycomics[54,61,62].

1.3 ANALYTICAL INSTRUMENTS AND METHODS IN GLYCOMICS

1.3.1 Mass Spectrometry

Mass spectrometry is the primary technique for characterizing the structures of individual glycans when only small quantities are available. An advantage of mass spectrometric glycan profiling is that multiple glycans of any given subtype can be profiled at once, increasing the throughput of the glycomic analysis. Typically, a mass spectrometric analysis consists of three procedures. First, the sample is converted into gas-phase and charged species in the ion source. Second, all ions are introduced into a mass analyzer, which separates ion species based on their mass-to-charge ratio (m/z) in vacuum. Last, the ions are detected and results are exported in the form of a mass spectrum.

Ion Source

The first step of MS analysis is to convert the analyte into gas-phase ions, which is a requirement imposed by the mass analyzer. In early time mass spectrometers were equipped with conventional ionization methods such as chemical ionization (CI) or electron ionization (EI), which required samples to be volatile. Since most biological samples are nonvolatile, polar and thermally unstable, they were not amenable to being analyzed by MS using these ionization techniques at first. The invention of other ionization methods in 1980's, matrix-assisted laser desorption/ionization (MALDI)[63] and electrospray ionization (ESI)[64] in particular, rendered MS a major breakthrough, which

results in the widespread application of MS in the analysis of nonvolatile macromolecules, especially biomolecules like protein, peptide and glycan. These ionization techniques are both soft-ionization techniques, resulting with little or no fragmentation[65].

The MALDI method involves the ionization of a protein with the proper combination of laser wavelength and matrix[66]. To prepare for MALDI ionization the sample must be crystallized in a low molecular weight ultraviolet-absorbing matrix and then irradiated with a laser beam to generate gas-phase ions. The matrix is used to isolate the sample molecules from each other for lower desorption energy, and to absorb laser light with a wavelength at which the analyte has a very weak absorbance. Absorption of energy from the laser beam causes evaporation and ionization of analytes[67]. Despite different lasers can be used in MALDI, ultraviolet lasers including Nitrogen lasers are most commonly used. The MALDI ionization has an advantage that the ionization efficiency isn't affected by the increase of the mass and size of the molecules,[66] however, it also suffers from the disadvantage of the presence of metastable ions formed by ions decomposing during flight. MALDI primarily produces intact singly charged molecular ions. MALDI method is reasonably tolerant toward the presence of salts, buffers and other addictives[68] and is very sensitive in the range from 100 femtomole to 2 picomole. As a result, MALDI is very useful as one of the two widely used ionization methods for biopolymers such as proteins and glycans for mass spectrometry.

In 1980s, John Bennett Fenn and coworkers introduced ESI as an important interface for biological macromolecules solution samples to mass spectrometry[69,70.] In ESI, a stream of liquid containing the sample is passed through a capillary and an

electrospray is produced by applying a high voltage to the capillary. A flow of dry nitrogen is often used to expedite evaporation of the solution and droplet shrinkage. Finally, the analyte molecules are stripped of solvent molecules to form gas phase ions[71]. There are two major theories that explain the final production of gas-phase ions. The ion evaporation model (IEM) suggests that as the droplet reaches a certain radius the field strength at the surface of the droplet becomes large enough to assist the field desorption of solvated ions[72]. In the other hand, the charged residue model (CRM) suggests that electrospray droplets undergo evaporation and fission cycles, eventually leading progeny droplets that contain on average one analyte ion or less. The gas-phase ions form after the remaining solvent molecules evaporate, leaving the analyte with the charges that the droplet carried[73]. Unlike MALDI technique, ESI usually produces multiply charged ions, which allows the analysis of very large molecules in case that the detection range of the mass analyzer is relatively small. This ionization technique can be coupled with almost all types of mass analyzers available. Another important advantage of ESI is that it can be coupled to liquid separation techniques such as HPLC with mass spectrometer in a real-time mode[74]. Nanoelctrospray (nanoESI) operates at much lower flow rates (1000~10 nl/min compared to 1000~10 μl/min for normal ESI) generate much smaller initial droplets by significantly reducing the needle tip diameter from 20~150 μm to 2~50 μm, which ensure improved ionization efficiency[75]. Nanospray brings the benefits of high sensitivity at femto-attomole level, better tolerance toward buffer and lower sample requirements compared to low-flow rate ESI[76].

Linear Ion Trap

In a linear ion trap, ions are confined radially by a two-dimensional (2D) radio frequency (RF) field, and axially by stopping potentials applied to end electrodes. In comparison to three-dimensional (3D) Paul traps[77], linear traps have higher injection efficiencies and higher ion storage capacities[78]. In a Thermo LTQ XL mass spectrometer (LTQ stands for ‘‘linear trap quadrupole’’), the linear ion trap is a square array of precision-machined and precision-aligned hyperbolic rods. In each quadrupole rod section, rods opposite each other in the array are connected electrically. Thus, the four rods of each section can be considered to be two pairs of two rods each. DC potentials are applied to trap ions in the Z direction while An AC voltage of constant frequency and amplitude is applied to the X and Y rods to trap ions in the X and Y direction. Helium is utilized as dampening gas when trapping to decrease the kinetic energy of the ions being trapped. The motion of ions in quadrupole fields is described mathematically by the solutions to a second-order linear differential equation described by Mathieu in 1868[79], which can only be solved numerically or equivalently by computer simulations. The mass spectrum is obtained by scanning the fields at which ions are ejected from the analyzer. LTQ allows a larger volume chamber, in possession of improved trapping efficiency and increased ion storage capacity and is capable of MS^n analysis. As a result, LTQ provides higher sensitivity, resolution and mass accuracy at similar mass range than conventional 3-D ion trap and is a robust tool for biochemical analysis[80,81].

Orbitrap

Orbitrap technology was first described in 2000[82] and has now achieved gold-standard status in the world of mass spectrometry. Compared to conventional ion traps

such as the Paul and Penning traps, the orbitrap uses only electrostatic fields instead to confine and analyze injected ion populations[83]. The Orbitrap mass analyzer is axially symmetrical and consists of a spindle-shape central electrode surrounded by a pair of bell-shaped outer electrodes. In the Orbitrap, stable ion trajectories combine rotation around an axial central electrode with harmonic oscillations along it. The frequency ω of these harmonic oscillations along the z-axis depends only on the ion mass-to-charge ratio m/q and the field curvature (k). The outer electrode of the orbitrap, which has two split halves, detect the image current produced by the oscillating ions. By Fast Fourier Transformation (FFT) of the image current, the instrument obtains the frequencies of these axial oscillations and therefore the mass-to-charge ratios of the ions. As the first high-performance mass analyzer to employ ion trapping in an electrostatic field, the Orbitrap mass spectrometer easily achieves ultra-high resolution (>100,000) with high mass accuracy (<1 ppm), a wide dynamic range (up to 5,000), fast scanning, and uncompromised sensitivity. In addition, its relatively low cost, simple design and high spacecharge capacity make it suitable for tackling complex scientific problems in which high performance is required[84]. Coupled with linear ion trap which possesses the MS^n capability, Orbitrap has proven to be a robust mass analyzer for proteomics, glycomics and studies of post translational modifications.

Electron Transfer Dissociation (ETD)

In 1998, McLafferty and coworkers[85] introduced a unique method for peptide/protein ion fragmentation: electron capture dissociation (ECD). In ECD, low energy electrons are reacted with peptide cations in the magnetic field of a Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR-MS). With ECD, peptide

backbone cleavage is relatively indifferent to either peptide sequence or length. More importantly, ECD does not cleave chemical modifications from peptide and thus labile post-translational modifications remain intact with ECD. However, ECD requires an FT-ICR mass spectrometer, which is the most expensive MS instrument[86]. Also, ECD is relatively inefficient since it takes a long time for a scan so it is hard to combine with a chromatography. As a new method using ion/ion chemistry to fragment peptides, ETD was introduced to the proteomics community[86,87]. ETD fragments peptide by transferring an electron from a radical anion to a peptide cation. The ion/ion reaction can be described like this:

$$(M + nH)^{n+} + A^{-}\bullet \rightarrow (M + nH)^{(n-1)+\bullet} + A \rightarrow \text{fragments}$$

Here $A^{-}\bullet$ is the reacting anion. This reaction results in the fragmentation of peptide backbone, causing the same cleavage just like what ECD does and producing c and z ions. ETD preserves post-translational modifications that are labile with CAD and makes it possible to obtain the sequence information of the peptide[87]. A large-scale study of ETD performance has been done, as compared to ion trap CAD, which proves that ETD outperforms CAD for all charge state greater than two[88]. We can also use ETD to address the problems we may encounter when using CAD to deal with the labile post-translational modifications, such as phosphorylation and glycosylation[89-95].

1.3.2 Assignment of Glycan Structures

The advances in the mass spectrometry technologies have extremely facilitate the study of glycomics, especially the research focusing on structural analysis of all the glycans in a given biological system, which is prerequisite to fully understanding the functional glycomics[96,97]. Due to several facts of glycan structures, namely, the diversity

of the anomeric configurations, interglycosidic linkages and variety of branching attached to the glycosylation sites, the assignment of glycan structures can be more complicated than peptide sequencing. A typical mass spectrometric analysis of glycan structures involves release of the intact glycans using enzymatic or chemical methods, appropriate derivatization to improve detection sensitivity and resolution and characterization by full mass spectrum scan or MS^n analysis. Separation and characterization by high performance liquid chromatography is sometimes a necessity for better recovery of glycan identifications.

Glycans can be released from proteins either by chemical reaction or enzymatic release. Since there is no specific enzyme to cleave O-linked glycans from peptide backbone, O-linked oligosaccharides are typically released chemically by a β-elimination reaction[98-100]. N-glycans cab be released from glycoproteins using a wide variety of endoglycosidases, exoglycosidases and glycoamidases, however, the most widely used enzyme for releasing intact N-glycans is Peptide N-glycosidase F (PNGase F)[101-104].

Prior to mass spectrometric analysis, glycans are usually processed by derivatization to enhance the hydrophobicity of glycans and increase the signal strength and provide more structural information in tandem mass spectrometry analysis. Permethylation is the most widely used derivatization method used for mass spectrometric analysis of oligosaccharides[105]. The addition of methyl groups stabilizes the sialic acid residues by converting highly polar –OH and –COOH groups into non-polar. Methylated glycans are hydrophobic so they can be easily separated from salts other impurities. And also fragmentation of methylated glycans is more predictable than their native counterparts leading to accurate structural assignment[106]. Besides

permethylation, fluorescent tags to the reducing ends such as 2-aminobenzoic acid (2-AA), 2-aminobenzamide (2-AB) and 2-aminopyridine (2-AP) are used for chromatographic detectability[107,108].

Extensive research has been done to study the fragmentation mechanisms for glycans since tandem mass spectrometry is extremely useful to achieve directly or indirectly structural information of the composition, sequence, branching and interglycosidic linkages of glycans[109-111]. Based on the nomenclature developed by Domon and Costello[112] (**Figure 1.1**), glycans fragmentation ions containing the non-reducing terminus are labeled with capital letters from the beginning, as A, B, C; those containing reducing end are labeled with letters from the end, as X, Y, Z; subscripts indicate the position of the cleavage corresponding to each end; superscripts mean the bond positions of cleavages in cross-ring fragmentation. The total ion mapping (TIM) was utilized to detect and quantify the prevalence of individual glycans in the total glycan profile[113,114]. In a TIM mode, automated MS and MS/MS spectra are obtained in collection window of a given width. The m/z range from 500 to 2000 is scanned in these successive collection windows, which also have overlap with each other. Due to the immaturity of high-throughput glycan identification software, manual structural assignment of glycans according to individual MS^n spectra is still required. In recent years, more and more automatic or semi-automatic computer programs for glycan structure assignment based on mass spectra such as SimGlycan[115], GlycoWorkbench[116] and etc. have been developed to save enormous time and effort.

1.4 QUANTITATIVE GLYCOMICS

It is one of the major challenges in the fields of –omics to develop relative-quantitative analysis technologies that are able to generate meaningful data to compare the expression levels of targeted molecules in different biological samples or developmental states. For proteomics people already have some powerful tools in hand. In two-dimensional (2D) gel electrophoresis, which has been traditionally performed for quantitative proteomics, the relative quantitation is achieved by recording differences in the staining intensity of protein spots derived from two states of differently labeled cell populations or tissues[117-119]. Since mass spectrometry has fulfilled its role as a rapid and reliable method for proteomics study, MS signal intensities, ion chromatograms, spectral counts, accurate mass retention time pairs and more other protein/peptide behaviors under MS have been used as label-free methods to quantify the changes in protein abundances [120-123]. Notably mass spectral peak intensities, peptide counts, spectral counts that add all the MS/MS spectra observed for peptides derived from a single protein in an LC-MS/MS analysis become very important tools for label-free quantitative proteomics because proteins that are more abundant in a sample have a higher probability of being identified during data dependent MS/MS scanning[124-127]. Other than label-free methods, stable isotopic labeling methods are more popular in recent years because unlike label-free methods, they are not suffering some shortcomings such as the varying response of instrument from time to time and different ionization efficiencies depending on species of analytes or other factors. Stable isotope labels can be incorporated by chemical approaches (*in vitro*) or through metabolic incorporation (*in vivo*). For example, ICAT (Isotope-Coded Affinity Tags) developed by Gygi chemically target specific amino acids,

typically cysteine, in the peptide sequence for differential labeling[128-130]. In the ^{18}O labeling strategy, the heavy ^{18}O isotope is introduced into the C-terminus of peptides by performing the digestion in $H_2{}^{18}O$. The proteolytic reaction involved in ^{18}O labeling can be performed in a variety of enzymes, such as trypsin, LysC, GluC, etc.[131,132]. Another in-vitro labeling strategy for quantitative proteomics that gets really popular in recent years is iTRAQ (Isobaric tag for relative and absolute quantification), which allows the relative and absolute quantification at the MS2 level of up to eight samples[133,134]. Stable isotopes can also be introduced into biological systems through metabolic labeling. For instance, SILAC (Stable Isotope Labeling with Amino Acids in Cell Culture) provides a simple, robust and powerful method for the incorporation of an isotopic label into proteins prior to MS based proteomics[135-137]. In SILAC, Cells are differentially labeled by growing them in light medium with normal amino acids (lysine or arginine) or medium with heavy amino acids. Metabolic incorporation of the amino acids into the proteins results in a mass shift of the corresponding peptides. This mass shift can be detected by a mass spectrometer. When both samples are combined, the ratio of peak intensities in the mass spectrum reflects the relative protein abundance. An advantage of this method is the early combination of samples after labeling in cell culture, which means that they will be subjected to identical downstream workflow throughout sample processing and analysis, thus minimizing technical variability.

Compared to proteomics, analytical technology for comparable glycomics is not that mature. However, people in the field of glycomics are always trying to follow the steps of proteomics and adapt some of the quantitative proteomics tools for glycomic analysis successfully (**Table 1.1**). For example, the method of total ion mapping (TIM),

which is a label-free strategy, is capable of identifying glycan structures based on fragmentation information from tandem MS and quantifying the prevalence of each glycan structure by reconstructed ion chromatogram or by normalizing each ion intensity seen on the mass spectrum to the total of them[113]. As for in-vitro isotopic labeling methods, several groups have used heavy methyl iodide ($^{13}CH_3I$, $^{12}CDH_2I$, $^{12}CHD_2I$, and/or $^{12}CD_3I$) to label glycans during the permethylation procedure which is normally performed before MS analysis for both N- and O-linked glycans. Then the glycan samples incorporated with heavy isotopes will be mixed with samples processed with standard permethylation using light methyl iodide ($^{12}CH_3I$)[113,138,139].

1.5 ISOTOPIC DETECTION OF AMINOSUGAR WITH GLUTAMINE (IDAWG)

In 2009, IDAWG (Isotopic Detection of Aminosugars With Glutamin) was reported as the first in-vivo stable isotopic labeling strategies for quantitative glycomics [140]. The IDAWG methodology takes advantage of hexosamine biosynthetic pathway that shows that the side chain of glutamine is the only source to provide nitrogen when producing the aminosugars. As a result, if the cells are fed with Gln-free media and glutamine with a ^{15}N-labeled side chain, all the aminosugars including GlcNAC, GalNAc and sialic acids produced in cells will be labeled with ^{15}N and thus the mass of all glycan structures will be shifted by +1 dalton per aminosugar. As an in-vivo labeling strategy, IDAWG shares some advantages with SILAC over in-vitro labeling methods. After the labeling, the ^{15}N-labeled sample can be mixed with normal sample immediately after the cell harvest or the cell lysis. Thus the glycans from two cell types are subjected to the same experimental conditions for glycan release and permethylation until they are analyzed by MS, which will dramatically reduce the possibilities of differential losses

that can make serious problems as it does to samples treated separately in a parallel way. The IDAWG technology has been used to analysis both N- and O-linked glycans released from murine embryonic stem cells successfully[140,141]. The overlapping of isotopic peaks corresponding to light and heavy species increases the complexity of mass spectra and requires software solution due to the difficulties of data analysis. Despite this, the IDAWG strategy is still predicted to be useful to various comparative glycomic studies in future.

REFERENCE

1 Walsh, G. & Jefferis, R. Post-translational modifications in the context of therapeutic proteins. *Nat Biotechnol* **24**, 1241-1252, doi:10.1038/nbt1252 (2006).

2 Krueger, R. R. Corneal topography vs ocular wavefront sensing in the retreatment of highly aberrated post surgical eyes. *J Refract Surg* **22**, 328-330 (2006).

3 Butor, C., Diaz, S. & Varki, A. High level O-acetylation of sialic acids on N-linked oligosaccharides of rat liver membranes. Differential subcellular distribution of 7- and 9-O-acetyl groups and of enzymes involved in their regulation. *J Biol Chem* **268**, 10197-10206 (1993).

4 Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97-130 (1993).

5 Morelle, W., Faid, V., Chirat, F. & Michalski, J. C. Analysis of N- and O-linked glycans from glycoproteins using MALDI-TOF mass spectrometry. *Methods Mol Biol* **534**, 5-21, doi:10.1007/978-1-59745-022-5_1 (2009).

6 Pabst, M. & Altmann, F. Glycan analysis by modern instrumental methods. *Proteomics* **11**, 631-643, doi:10.1002/pmic.201000517 (2011).

7 Aoki-Kinoshita, K. F. An introduction to bioinformatics for glycomics research. *PLoS Comput Biol* **4**, e1000075, doi:10.1371/journal.pcbi.1000075 (2008).

8 Sato, C. *et al.* Characterization of the N-oligosaccharides attached to the atypical Asn-X-Cys sequence of recombinant human epidermal growth factor receptor. *Journal of biochemistry* **127**, 65-72 (2000).

9 Vance, B. A., Wu, W., Ribaudo, R. K., Segal, D. M. & Kearse, K. P. Multiple dimeric forms of human CD69 result from differential addition of N-glycans to

typical (Asn-X-Ser/Thr) and atypical (Asn-X-cys) glycosylation motifs. *J Biol Chem* **272**, 23117-23122 (1997).

10 Gil, G. C., Velander, W. H. & Van Cott, K. E. N-glycosylation microheterogeneity and site occupancy of an Asn-X-Cys sequon in plasma-derived and recombinant protein C. *Proteomics* **9**, 2555-2567, doi:10.1002/pmic.200800775 (2009).

11 Chi, Y. H. *et al.* N-glycosylation at non-canonical Asn-X-Cys sequence of an insect recombinant cathepsin B-like counter-defense protein. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology* **156**, 40-47, doi:10.1016/j.cbpb.2010.01.017 (2010).

12 Matsui, T. *et al.* N-glycosylation at noncanonical Asn-X-Cys sequences in plant cells. *Glycobiology* **21**, 994-999, doi:10.1093/glycob/cwq198 (2011).

13 Gerken, T. A., Owens, C. L. & Pasumarthy, M. Site-specific core 1 O-glycosylation pattern of the porcine submaxillary gland mucin tandem repeat. Evidence for the modulation of glycan length by peptide sequence. *J Biol Chem* **273**, 26580-26588 (1998).

14 Asada, M. *et al.* The AATPAP sequence is a very efficient signal for O-glycosylation in CHO cells. *Glycoconj J* **16**, 321-326 (1999).

15 Schlummer, S. *et al.* Influence of serine O-glycosylation or O-phosphorylation close to the vJun nuclear localisation sequence on nuclear import. *Chembiochem : a European journal of chemical biology* **7**, 88-97, doi:10.1002/cbic.200500212 (2006).

16 de Lederkremer, R. M., Lima, C., Ramirez, M. I. & Casal, O. L. Structural features of the lipopeptidophosphoglycan from Trypanosoma cruzi common with the glycophosphatidylinositol anchors. *European journal of biochemistry / FEBS* **192**, 337-345 (1990).

17 Puoti, A., Desponds, C., Fankhauser, C. & Conzelmann, A. Characterization of glycophospholipid intermediate in the biosynthesis of glycophosphatidylinositol anchors accumulating in the Thy-1-negative lymphoma line SIA-b. *J Biol Chem* **266**, 21051-21059 (1991).

18 Puoti, A. & Conzelmann, A. Structural characterization of free glycolipids which are potential precursors for glycophosphatidylinositol anchors in mouse thymoma cell lines. *J Biol Chem* **267**, 22673-22680 (1992).

19 Glowacki, A., Kozma, E. M., Olczyk, K. & Kucharz, E. J. [Glycosaminoglycans--structure and function]. *Postepy biochemii* **41**, 139-148 (1995).

20 Sasisekharan, R., Raman, R. & Prabhakar, V. Glycomics approach to structure-function relationships of glycosaminoglycans. *Annual review of biomedical engineering* **8**, 181-231, doi:10.1146/annurev.bioeng.8.061505.095745 (2006).

21 Krieg, J. *et al.* C-Mannosylation of human RNase 2 is an intracellular process performed by a variety of cultured cells. *J Biol Chem* **272**, 26687-26692 (1997).

22 Gonzalez de Peredo, A. *et al.* C-mannosylation and o-fucosylation of thrombospondin type 1 repeats. *Mol Cell Proteomics* **1**, 11-18 (2002).

23 Perez-Vilar, J., Randell, S. H. & Boucher, R. C. C-Mannosylation of MUC5AC and MUC5B Cys subdomains. *Glycobiology* **14**, 325-337, doi:10.1093/glycob/cwh041 (2004).

24 Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**, 4-8, doi:S0304-4165(99)00165-8 [pii] (1999).

25 Atwood, J. A., 3rd *et al.* Quantitation by isobaric labeling: applications to glycomics. *J Proteome Res* **7**, 367-374, doi:10.1021/pr070476i (2008).

26 Lin, X. Functions of heparan sulfate proteoglycans in cell signaling during development. *Development* **131**, 6009-6021, doi:131/24/6009 [pii] 10.1242/dev.01522 (2004).

27 Inatani, M., Irie, F., Plump, A. S., Tessier-Lavigne, M. & Yamaguchi, Y. Mammalian brain morphogenesis and midline axon guidance require heparan sulfate. *Science* **302**, 1044-1046, doi:10.1126/science.1090497 302/5647/1044 [pii] (2003).

28 Hwang, H. Y., Olson, S. K., Esko, J. D. & Horvitz, H. R. Caenorhabditis elegans early embryogenesis and vulval morphogenesis require chondroitin biosynthesis. *Nature* **423**, 439-443, doi:10.1038/nature01634 nature01634 [pii] (2003).

29 Lowe, J. B. & Marth, J. D. A genetic approach to Mammalian glycan function. *Annu Rev Biochem* **72**, 643-691, doi:10.1146/annurev.biochem.72.121801.161809 121801.161809 [pii] (2003).

30 Haltiwanger, R. S. & Lowe, J. B. Role of glycosylation in development. *Annu Rev Biochem* **73**, 491-537, doi:10.1146/annurev.biochem.73.011303.074043 (2004).

31 Cipollo, J. F., Awad, A. M., Costello, C. E. & Hirschberg, C. B. N-Glycans of Caenorhabditis elegans are specific to developmental stages. *J Biol Chem* **280**, 26063-26072, doi:M503828200 [pii] 10.1074/jbc.M503828200 (2005).

32 Collins, B. E. & Paulson, J. C. Cell surface biology mediated by low affinity multivalent protein-glycan interactions. *Curr Opin Chem Biol* **8**, 617-625, doi:10.1016/j.cbpa.2004.10.004 (2004).

33 Crocker, P. R. Siglecs: sialic-acid-binding immunoglobulin-like lectins in cell-cell interactions and signalling. *Curr Opin Struct Biol* **12**, 609-615, doi:S0959440X02003755 [pii] (2002).

34 Ishida, H. *et al.* A novel beta1,3-N-acetylglucosaminyltransferase (beta3Gn-T8), which synthesizes poly-N-acetyllactosamine, is dramatically upregulated in colon cancer. *FEBS Lett* **579**, 71-78, doi:S0014-5793(04)01427-9 [pii] 10.1016/j.febslet.2004.11.037 (2005).

35 Fuster, M. M., Brown, J. R., Wang, L. & Esko, J. D. A disaccharide precursor of sialyl Lewis X inhibits metastatic potential of tumor cells. *Cancer Res* **63**, 2775-2781 (2003).

36 Liu, D., Shriver, Z., Venkataraman, G., El Shabrawi, Y. & Sasisekharan, R. Tumor cell surface heparan sulfate as cryptic promoters or inhibitors of tumor growth and metastasis. *Proc Natl Acad Sci U S A* **99**, 568-573, doi:10.1073/pnas.012578299 99/2/568 [pii] (2002).

37 Dube, D. H. & Bertozzi, C. R. Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat Rev Drug Discov* **4**, 477-488, doi:nrd1751 [pii] 10.1038/nrd1751 (2005).

38 Iwai, T. *et al.* Core 3 synthase is down-regulated in colon carcinoma and profoundly suppresses the metastatic potential of carcinoma cells. *Proc Natl Acad Sci U S A* **102**, 4572-4577, doi:0407983102 [pii] 10.1073/pnas.0407983102 (2005).

39 Sasisekharan, R., Shriver, Z., Venkataraman, G. & Narayanasami, U. Roles of heparan-sulphate glycosaminoglycans in cancer. *Nat Rev Cancer* **2**, 521-528, doi:10.1038/nrc842 nrc842 [pii] (2002).

40 Crocker, P. R. Siglecs in innate immunity. *Curr Opin Pharmacol* **5**, 431-437, doi:S1471-4892(05)00072-X [pii] 10.1016/j.coph.2005.03.003 (2005).

41 Guo, Y., Mascareno, E. & Siddiqui, M. A. Distinct components of Janus kinase/signal transducer and activator of transcription signaling pathway mediate the regulation of systemic and tissue localized renin-angiotensin system. *Mol Endocrinol* **18**, 1033-1041, doi:10.1210/me.2003-0231 me.2003-0231 [pii] (2004).

42 Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A. & Dwek, R. A. Glycosylation and the immune system. *Science* **291**, 2370-2376 (2001).

43 Rudd, P. M., Wormald, M. R. & Dwek, R. A. Sugar-mediated ligand-receptor interactions in the immune system. *Trends Biotechnol* **22**, 524-530, doi:10.1016/j.tibtech.2004.07.012 S0167-7799(04)00205-7 [pii] (2004).

44 Kinjo, Y. *et al.* Recognition of bacterial glycosphingolipids by natural killer T cells. *Nature* **434**, 520-525, doi:nature03407 [pii] 10.1038/nature03407 (2005).

45 Schachter, H. Congenital disorders involving defective N-glycosylation of proteins. *Cellular and molecular life sciences : CMLS* **58**, 1085-1104 (2001).

46 Dennis, J. W., Granovsky, M. & Warren, C. E. Glycoprotein glycosylation and cancer progression. *Biochim Biophys Acta* **1473**, 21-34 (1999).

47 Delves, P. J. The role of glycosylation in autoimmune disease. *Autoimmunity* **27**, 239-253 (1998).

48 Gleeson, P. A. Glycoconjugates in autoimmunity. *Biochim Biophys Acta* **1197**, 237-255 (1994).

49 Chui, D. *et al.* Genetic remodeling of protein glycosylation in vivo induces autoimmune disease. *Proc Natl Acad Sci U S A* **98**, 1142-1147, doi:10.1073/pnas.98.3.1142 (2001).

50 Snider, M. D. & Rogers, O. C. Membrane traffic in animal cells: cellular glycoproteins return to the site of Golgi mannosidase I. *The Journal of cell biology* **103**, 265-275 (1986).

51 Huang, K. M. & Snider, M. D. Glycoprotein recycling to the galactosyltransferase compartment of the Golgi complex. *J Biol Chem* **268**, 9302-9310 (1993).

52 Neefjes, J. J. *et al.* Recycling glycoproteins do not return to the cis-Golgi. *The Journal of cell biology* **107**, 79-87 (1988).

53 Porwoll, S. *et al.* Cell surface glycoproteins undergo postbiosynthetic modification of their N-glycans by stepwise demannosylation. *J Biol Chem* **273**, 1075-1085 (1998).

54 Shriver, Z., Raguram, S. & Sasisekharan, R. Glycomics: a pathway to a class of new and improved therapeutics. *Nat Rev Drug Discov* **3**, 863-873, doi:10.1038/nrd1521
nrd1521 [pii] (2004).

55 Raman, R., Raguram, S., Venkataraman, G., Paulson, J. C. & Sasisekharan, R. Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nat Methods* **2**, 817-824, doi:10.1038/nmeth807 (2005).

56 Raman, R., Sasisekharan, V. & Sasisekharan, R. Structural insights into biological roles of protein-glycosaminoglycan interactions. *Chemistry & biology* **12**, 267-277, doi:10.1016/j.chembiol.2004.11.020 (2005).

57 Freeze, H. H. Genetic defects in the human glycome. *Nature reviews. Genetics* **7**, 537-551, doi:10.1038/nrg1894 (2006).

58 Yue, T. *et al.* The prevalence and nature of glycan alterations on specific proteins in pancreatic cancer patients revealed using antibody-lectin sandwich arrays. *Mol Cell Proteomics* **8**, 1697-1707, doi:M900135-MCP200 [pii]
10.1074/mcp.M900135-MCP200 (2009).

59 Amano, M. & Nishimura, S. Large-scale glycomics for discovering cancer-associated N-glycans by integrating glycoblotting and mass spectrometry. *Methods Enzymol* **478**, 109-125, doi:S0076-6879(10)78004-6 [pii] 10.1016/S0076-6879(10)78004-6 (2010).

60 Packer, N. H. *et al.* Frontiers in glycomics: bioinformatics and biomarkers in disease. An NIH white paper prepared from discussions by the focus groups at a workshop on the NIH campus, Bethesda MD (September 11-13, 2006). *Proteomics* **8**, 8-20, doi:10.1002/pmic.200700917 (2008).

61 Kam, R. & Poon, T. The Potentials of Glycomics in Biomarker Discovery. *Clinical Proteomics* **4**, 67-79, doi:10.1007/s12014-008-9017-9 (2008).

62 Dube, D. H. & Bertozzi, C. R. Glycans in cancer and inflammation [mdash] potential for therapeutics and diagnostics. *Nat Rev Drug Discov* **4**, 477-488 (2005).

63 Karas, M. & Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical chemistry* **60**, 2299-2301, doi:citeulike-article-id:1378325 (1988).

64 Fenn, J., Mann, M., Meng, C., Wong, S. & Whitehouse, C. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71, doi:10.1126/science.2675315 (1989).

65 Hop, C. E. C. A. & Bakhtiar, R. An introduction to electrospray ionization and matrix-assisted laser desorption/ionization mass spectrometry: Essential tools in a modern biotechnology environment. *Biospectroscopy* **3**, 259-280, doi:10.1002/(sici)1520-6343(1997)3:4<259::aid-bspy2>3.0.co;2-# (1997).

66 Tanaka, K. *et al.* Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **2**, 151-153, doi:10.1002/rcm.1290020802 (1988).

67 Beavis, R. C., Chait, B. T. & Fales, H. M. Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid Communications in Mass Spectrometry* **3**, 432-435, doi:10.1002/rcm.1290031207 (1989).

68 Beavis, R. C. & Chait, B. T. Rapid, sensitive analysis of protein mixtures by mass spectrometry. *Proc Natl Acad Sci U S A* **87**, 6873-6877 (1990).

69 Whitehouse, C. M., Dreyer, R. N., Yamashita, M. & Fenn, J. B. Electrospray interface for liquid chromatographs and mass spectrometers. *Anal Chem* **57**, 675-679 (1985).

70 Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).

71 Gaskell, S. J. Electrospray: Principles and Practice. *Journal of Mass Spectrometry* **32**, 677-688, doi:10.1002/(sici)1096-9888(199707)32:7<677::aid-jms536>3.0.co;2-g (1997).

72 Iribarne, J. On the evaporation of small ions from charged droplets. *J. Chem. Phys.* **64**, 2287 (1976).

73 Dole, M. Molecular Beams of Macroions. *J. Chem. Phys.* **49**, 2240 (1968).

74 MEHLIS *et al. Liquid chromatography/mass spectrometry of peptides of biological samples*. Vol. 352 (Elsevier, 1997).

75 Wilm, M. & Mann, M. Analytical Properties of the Nanoelectrospray Ion Source. *Analytical chemistry* **68**, 1-8, doi:10.1021/ac9509519 (1996).

76 Körner, R., Wilm, M., Morand, K., Schubert, M. & Mann, M. Nano electrospray combined with a quadrupole ion trap for the analysis of peptides and protein digests. *Journal of the American Society for Mass Spectrometry* **7**, 150-156, doi:10.1016/1044-0305(95)00626-5 (1996).

77 Paul, W. Electromagnetic traps for charged and neutral particles. *Reviews of Modern Physics* **62**, 531-540 (1990).

78 Douglas, D. J., Frank, A. J. & Mao, D. Linear ion traps in mass spectrometry. *Mass Spectrometry Reviews* **24**, 1-29, doi:10.1002/mas.20004 (2005).

79 March, R. E. An Introduction to Quadrupole Ion Trap Mass Spectrometry. *Journal of Mass Spectrometry* **32**, 351-369, doi:10.1002/(sici)1096-9888(199704)32:4<351::aid-jms512>3.0.co;2-y (1997).

80 Hager, J. W. A new linear ion trap mass spectrometer. *Rapid Communications in Mass Spectrometry* **16**, 512-526, doi:10.1002/rcm.607 (2002).

81 Schwartz, J. C., Senko, M. W. & Syka, J. E. A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom* **13**, 659-669, doi:10.1016/S1044-0305(02)00384-7 (2002).

82 Makarov, A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* **72**, 1156-1162 (2000).

83 Hu, Q. *et al.* The Orbitrap: a new mass spectrometer. *Journal of mass spectrometry : JMS* **40**, 430-443, doi:10.1002/jms.856 (2005).

84 Scigelova, M. & Makarov, A. Orbitrap mass analyzer--overview and applications in proteomics. *Proteomics* **6 Suppl 2**, 16-21, doi:10.1002/pmic.200600528 (2006).

85 Zubarev, R. A., Kelleher, N. L. & McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *Journal of the American Chemical Society* **120**, 3265-3266, doi:10.1021/ja973478k (1998).

86 Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **101**, 9528-9533, doi:10.1073/pnas.0402700101 (2004).

87 Mikesh, L. M. *et al.* The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta* **1764**, 1811-1822, doi:10.1016/j.bbapap.2006.10.003 (2006).

88 Good, D. M., Wirtala, M., McAlister, G. C. & Coon, J. J. Performance characteristics of electron transfer dissociation mass spectrometry. *Mol Cell Proteomics* **6**, 1942-1951, doi:10.1074/mcp.M700073-MCP200 (2007).

89 McAlister, G. C. *et al.* A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. *J Proteome Res* **7**, 3127-3136, doi:10.1021/pr800264t (2008).

90 Shi, S. D. *et al.* Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. *Anal Chem* **73**, 19-22 (2001).

91 Kelleher, N. L. *et al.* Localization of labile posttranslational modifications by electron capture dissociation: the case of gamma-carboxyglutamic acid. *Anal Chem* **71**, 4250-4253 (1999).

92 Chi, A. *et al.* Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc Natl Acad Sci U S A* **104**, 2193-2198, doi:10.1073/pnas.0607084104 (2007).

93 Swaney, D. L., Wenger, C. D., Thomson, J. A. & Coon, J. J. Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A* **106**, 995-1000, doi:10.1073/pnas.0811964106 (2009).

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

95 Alley, W. R., Jr., Mechref, Y. & Novotny, M. V. Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Commun Mass Spectrom* **23**, 161-170, doi:10.1002/rcm.3850 (2009).

96 Dell, A. & Morris, H. R. Glycoprotein structure determination by mass spectrometry. *Science* **291**, 2351-2356 (2001).

97 Haslam, S. M., North, S. J. & Dell, A. Mass spectrometric analysis of N- and O-glycosylation of tissues and cells. *Curr Opin Struct Biol* **16**, 584-591, doi:10.1016/j.sbi.2006.08.006 (2006).

98 Dell, A. *et al.* Mass spectrometry of carbohydrate-containing biopolymers. *Methods Enzymol* **230**, 108-132 (1994).

99 Huang, Y., Mechref, Y. & Novotny, M. V. Microscale nonreductive release of O-linked glycans for subsequent analysis through MALDI mass spectrometry and capillary electrophoresis. *Anal Chem* **73**, 6063-6069 (2001).

100 Conrad, H. E. Beta-elimination for release of O-linked glycosaminoglycans from proteoglycans. *Curr Protoc Mol Biol* **Chapter 17**, Unit17 15A, doi:10.1002/0471142727.mb1715as31 (2001).

101 Freeze, H. H. & Varki, A. Endo-glycosidase F and peptide N-glycosidase F release the great majority of total cellular N-linked oligosaccharides: use in demonstrating that sulfated N-linked oligosaccharides are frequently found in cultured cells. *Biochemical and biophysical research communications* **140**, 967-973 (1986).

102 Papac, D. I., Briggs, J. B., Chin, E. T. & Jones, A. J. A high-throughput microscale method to release N-linked oligosaccharides from glycoproteins for matrix-assisted laser desorption/ionization time-of-flight mass spectrometric analysis. *Glycobiology* **8**, 445-454 (1998).

103 Freeze, H. H. Endoglycosidase and glycoamidase release of N-linked oligosaccharides. *Curr Protoc Mol Biol* **Chapter 17**, Unit17 13A, doi:10.1002/0471142727.mb1713as48 (2001).

104 Wheeler, S. F. & Harvey, D. J. Extension of the in-gel release method for structural analysis of neutral and sialylated N-linked glycans to the analysis of sulfated glycans: application to the glycans from bovine thyroid-stimulating

hormone. *Analytical biochemistry* **296**, 92-100, doi:10.1006/abio.2001.5199 (2001).

105 Ciucanu, I. & Kerek, F. A simple and rapid method for the permethylation of carbohydrates. *Carbohydrate Research* **131**, 209-217, doi:10.1016/0008-6215(84)85242-8 (1984).

106 Kang, P., Mechref, Y., Klouckova, I. & Novotny, M. V. Solid-phase permethylation of glycans for mass spectrometric analysis. *Rapid Commun Mass Spectrom* **19**, 3421-3428, doi:10.1002/rcm.2210 (2005).

107 Anumula, K. R. High-sensitivity and high-resolution methods for glycoprotein analysis. *Analytical biochemistry* **283**, 17-26, doi:10.1006/abio.2000.4645 (2000).

108 Harvey, D. J. Electrospray mass spectrometry and fragmentation of N-linked carbohydrates derivatized at the reducing terminus. *J Am Soc Mass Spectrom* **11**, 900-915, doi:10.1016/S1044-0305(00)00156-2 (2000).

109 Ashline, D., Singh, S., Hanneman, A. & Reinhold, V. Congruent strategies for carbohydrate sequencing. 1. Mining structural details by MSn. *Anal Chem* **77**, 6250-6262, doi:10.1021/ac050724z (2005).

110 Zhang, H., Singh, S. & Reinhold, V. N. Congruent strategies for carbohydrate sequencing. 2. FragLib: an MSn spectral library. *Anal Chem* **77**, 6263-6270, doi:10.1021/ac050725r (2005).

111 Lapadula, A. J. *et al.* Congruent strategies for carbohydrate sequencing. 3. OSCAR: an algorithm for assigning oligosaccharide topology from MSn data. *Anal Chem* **77**, 6271-6279, doi:10.1021/ac050726j (2005).

112 Domon, B. & Costello, C. E. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate Journal* **5**, 397-409, doi:10.1007/bf01049915 (1988).

113 Aoki, K. *et al.* Dynamic developmental elaboration of N-linked glycan complexity in the Drosophila melanogaster embryo. *J Biol Chem* **282**, 9127-9142, doi:M606711200 [pii] 10.1074/jbc.M606711200 (2007).

114 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, doi:M804925200 [pii] 10.1074/jbc.M804925200 (2008).

115 Apte, A. & Meitei, N. S. Bioinformatics in glycomics: glycan characterization with mass spectrometric data using SimGlycan. *Methods Mol Biol* **600**, 269-281, doi:10.1007/978-1-60761-454-8_19 (2010).

116 Ceroni, A. *et al.* GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res* **7**, 1650-1659, doi:10.1021/pr7008252 (2008).

117 O'Farrell, P. H. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**, 4007-4021 (1975).

118 Zhou, G. *et al.* 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers. *Mol Cell Proteomics* **1**, 117-124 (2002).

119 Patton, W. F. & Beechem, J. M. Rainbow's end: the quest for multiplexed fluorescence quantitative analysis in proteomics. *Curr Opin Chem Biol* **6**, 63-69 (2002).

120 Liu, H., Sadygov, R. G. & Yates, J. R., 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76**, 4193-4201, doi:10.1021/ac0498563 (2004).

121 Radulovic, D. *et al.* Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry. *Mol Cell Proteomics* **3**, 984-997, doi:10.1074/mcp.M400061-MCP200 M400061-MCP200 [pii] (2004).

122 Silva, J. C. *et al.* Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem* **77**, 2187-2200, doi:10.1021/ac048455k (2005).

123 Wang, W. *et al.* Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal Chem* **75**, 4818-4826 (2003).

124 Wiener, M. C., Sachs, J. R., Deyanova, E. G. & Yates, N. A. Differential mass spectrometry: a label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. *Anal Chem* **76**, 6085-6096, doi:10.1021/ac0493875 (2004).

125 Bondarenko, P. V., Chelius, D. & Shaler, T. A. Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Anal Chem* **74**, 4741-4749 (2002).

126 Chelius, D. & Bondarenko, P. V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res* **1**, 317-323 (2002).

127 Higgs, R. E., Knierman, M. D., Gelfanova, V., Butler, J. P. & Hale, J. E. Comprehensive label-free method for the relative quantification of proteins from biological samples. *J Proteome Res* **4**, 1442-1450, doi:10.1021/pr050109b (2005).

128 Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994-999, doi:10.1038/13690 (1999).

129 Haqqani, A. S. *et al.* Protein markers of ischemic insult in brain endothelial cells identified using 2D gel electrophoresis and ICAT-based quantitative proteomics. *J Proteome Res* **6**, 226-239, doi:10.1021/pr0603811 (2007).

130 Haqqani, A. S., Hutchison, J. S., Ward, R. & Stanimirovic, D. B. Biomarkers and diagnosis; protein biomarkers in serum of pediatric patients with severe traumatic brain injury identified by ICAT-LC-MS/MS. *Journal of neurotrauma* **24**, 54-74, doi:10.1089/neu.2006.0079 (2007).

131 Schnolzer, M., Jedrzejewski, P. & Lehmann, W. D. Protease-catalyzed incorporation of 18O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. *Electrophoresis* **17**, 945-953, doi:10.1002/elps.1150170517 (1996).

132 Miyagi, M. & Rao, K. C. Proteolytic 18O-labeling strategies for quantitative proteomics. *Mass Spectrom Rev* **26**, 121-136, doi:10.1002/mas.20116 (2007).

133 Ross, P. L. *et al.* Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**, 1154-1169, doi:10.1074/mcp.M400129-MCP200 (2004).

134 Glen, A. *et al.* iTRAQ-facilitated proteomic analysis of human prostate cancer cells identifies proteins associated with progression. *J Proteome Res* **7**, 897-907, doi:10.1021/pr070378x (2008).

135 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).

136 Ong, S. E., Mittler, G. & Mann, M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat Methods* **1**, 119-126, doi:10.1038/nmeth715 (2004).

137 Ong, S. E. & Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat Protoc* **1**, 2650-2660, doi:10.1038/nprot.2006.427 (2006).

138 Alvarez-Manilla, G. *et al.* Tools for glycomics: relative quantitation of glycans by isotopic permethylation using 13CH3I. *Glycobiology* **17**, 677-687, doi:cwm033 [pii] 10.1093/glycob/cwm033 (2007).

139 Kang, P., Mechref, Y., Kyselova, Z., Goetz, J. A. & Novotny, M. V. Comparative glycomic mapping through quantitative permethylation and stable-isotope labeling. *Anal Chem* **79**, 6064-6073, doi:10.1021/ac062098r (2007).

140 Orlando, R. *et al.* IDAWG: Metabolic incorporation of stable isotope labels for quantitative glycomics of cultured cells. *J Proteome Res* **8**, 3816-3823, doi:10.1021/pr8010028 (2009).

141 Fang, M., Lim, J.-M. & Wells, L. *Quantitative Glycomics of Cultured Cells Using Isotopic Detection of Aminosugars with Glutamine (IDAWG)*. (John Wiley & Sons, Inc., 2009).

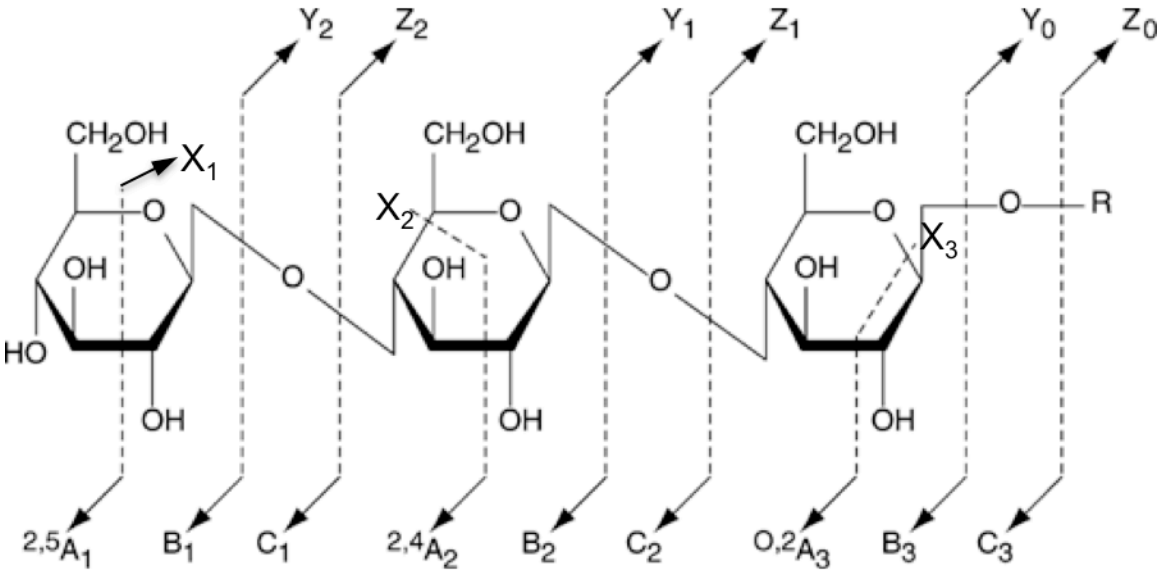


Figure 1.1 Glycans fragmentation nomenclature, after Domon and Costello[112]

Table 1.1 Quantitative Strategies for Proteomics and Glycomics

LABELING STRATEGY	PROTEOMICS	GLYCOMICS
Label-rree	Spectra count	TIM/Prevalence
In-vitro	^{18}O-H_2O labeling	^{13}C-CH_3I or CD_3I Permethylation
In-vivo	SILAC	IDAWG

CHAPTER 2

QUANTITATIVE GLYCOMICS OF CULTURED CELLS USING ISOTOPIC DETECTION OF AMINOSUGARS WITH GLUTAMINE (IDAWG)[1]

[1] Meng Fang, Jae-Min Lim and Lance Wells, *Current Protocols in Chemical Biology,* **2009**, ISBN: 9780470559277
Reprinted here with permission of the publisher

ABSTRACT

IDAWG (Isotopic Detection of Aminosugars With Glutamine) is a newly reported in vivo stable isotopic labeling strategy for quantitative glycomics of cultured cells. This unit provides detailed procedures for glycan analysis using IDAWG including labeling, release of both N- and O-linked glycans, permethylation and mass spectrometry analysis. The methods for data processing and calculations are also introduced here but have not yet been automated.

INTRODUTION

Glycomics is the comprehensive study of the entire complement of complex carbohydrates, often called glycans, that are polymers of monosaccharides often found attached to proteins and lipids[1]. As a post-translational modification, glycosylation plays critical roles in numerous physiological processes including protein folding and cellular signaling, and altered carbohydrate expression is a common feature of many types of cancers[2]. One of the important tasks for glycomics is to develop technologies capable of comparative, relative-quantitative analysis for examining global alterations of glycans between different biological samples. Following the steps of proteomics, several non-isotope-based and stable-isotope-based labeling strategies have recently been developed for quantitative glycomics by mass spectrometry[3-5].

This unit provides the detailed protocols for Isotopic Detection of Aminosugars With Glutamine, termed as IDAWG, that is the first in-vivo cell culture isotope-labeling strategy for glycan analysis[6]. The methods of labeling the aminosugars (GlcNAc, GalNAc and sialic acid) with amide-15N-Gln in cell culture and the detection of these aminosugars in both N- and O-linked glycan structures are described in the basic protocols. The mathematical calculations used to calculate Under-incorporation rate and relative ratios of light/heavy species are described in the support protocols.

EXPERIMENTAL

Basic Protocol 1: Cell Culture For IDAWG Labeling

This basic protocol is for labeling any cells in culture. Due to large differences in metabolic flux, different cells must be cultured in heavy Gln for different periods of time

to achieve a high degree (>95%) of labeling. In our hands, hESCs that divide quickly, doubling time of ~30 hours, and are very metabolically active label to completion in 72 hours (changing the media every 24 hours). However, we have also labeled differentiated 3T3-L1 adipocytes, that are post-mitotic, that required six days of labeling to achieve >90% labeling. Thus, the time of labeling must be adjusted for each individual cell line but since the label is relatively inexpensive, one week of labeling is a good starting point for an unknown cell type and under-incorporation can be calculated (see supplementary protocols). This one-week labeling is a starting point that can be adjusted based on the metabolic activity and to some extent the doubling time of the cells. In terms of scale, it is recommended to begin with a minimum of 2 X 106 cells (scaling up for making protein powder, below, is straightforward) per growth condition.

1. Grow cells as normal using Gln-free media (Invitrogen) supplemented with Gln as described in the next step.

2. Supplement the media with either amide-15N-Gln (98%) (Cambridge Isotope Laboratories, Inc) or corresponding normal abundance 14N-Gln at 2 mM final concentration.

3. Change the media daily and allow the cells to grow for at least 72 hours in heavy label.

4. Cells must not be harvested using trypsin as it will cleave off many of the cell surface glycoproteins. Instead adherent cells should be harvested by scraping in PBS.

Basic Protocol 2: N- and O-linked Glycan Analysis with IDAWG

This basic protocol describes the detailed procedure for analysis of both permethylated N- and O-linked glycans released from cultured cells via mass spectrometry. The cells will be first lysed, delipidated and turned into protein powders.

Both N- and O-linked glycans will be released from the protein and the permethylation of glycans will be performed (see Figure 2.1). Permethylated glycans are then analyzed using high-resolution mass spectrometry. To compare the quantities of each glycan structures in different samples, the protein powders of normal and ^{15}N-labeled samples are mixed together by weight (See Basic Protocol 2) or by cell number upon culture harvesting. To calculate the under-incorporation rate for each glycan, which is crucial for the calculation of the ratio of light/heavy structures, an extra analysis of pure ^{15}N-labeled sample with exactly the same procedure can be performed. (See Support Protocol 1)

1. Materials

Cell pellets from cell culture (see Basic Protocol 1)

Methanol, HPLC grade

Chloroform, HPLC grade

18.2 mega ohm (Milli-Q) water

Acetone, HPLC grade

40 mM NH_4HCO_3 buffer (pH 8)

Trypsin (2 mg/ml in 40 mM NH_4HCO_3, store at -20ºC) (pH 8)

Chymotrypsin (2 mg/ml in 40 mM NH_4HCO_3, store at -20ºC) (pH 8)

2 M urea in 40 mM NH_4HCO_3 (made fresh)

5% and 10% (v/v) acetic acid

Acetonitrile, HPLC grade

Isopropanol, HPLC grade

20% (v/v) isopropanol in 5% acetic acid

40% (v/v) isopropanol in 5% acetic acid

100 mM sodium phosphate (pH 7.5)

Peptide-N-glycosidase F (PNGase F, 7.5μg/μl, store at 4 ºC) (Prozyme)

1 M sodium borohydride (made fresh)

AG 50W – X8 resin (BIO-RAD)

To prepare the resin stock, the AG 50W – X8 resin is washed with methanol for three times and incubated overnight at room temperature. Then wash with methanol, 1M hydrochloric acid and 5% acetic acid respectively and keep the stock at 4 ºC.

1 M HCl

Acetic acid, glacial, HPLC grade

50% w/w sodium hydroxide solution

Anhydrous methanol (99.8%, Sigma)

Anhydrous dimethyl sulfoxide(99.9%, Sigma)

Iodomethane (99.5%, Aldrich)

Dichloromethane, HPLC grade

1 mM NaOH in 50% methanol

10 ml dounce homogenizer

1.5 ml ependorf tubes and 8 ml screwtop glass tubes, pre-cleaned with methanol

Heating block (e.g., Fisher Isotemp 125D)

Scientific Industries Vortex Genie 2 or equivalent

Sonicator

Incubator

Test tube rocker

Speed-Vac concentrator

BakerbondTM spe Octadecyl (C18) Disposable Extraction Columns (J.T.Baker)

VWR Clinical 50 centrifuger or equivalent

Pierce Reacti-Vap Evaporating Unit and Reacti-Therm Heating/Stirring Module

Fused-silica emitter (360 Å~ 75 Å~ 30 μm, SilicaTip™, New Objective)

LTQ-Orbitrap XL mass spectrometer with nano ESI source (ThermoFisher) or equivalent

Additional reagents and equipment for performing cell culture

2. Cell Lysis and Delipidation

1. Take cell pellet (~2 x 10_6 cells) that has been washed in PBS and add water to 100 μl and put in a 1.5 ml eppendorf tube.

2. Add 500 μl ice-cold methanol to the tube and move the cells to a 10 ml dounce homogenizer.

3. Disrupt the cells well on ice by dounce homogenization (6-8 strokes) then transfer the mixture to a 8 ml glass tube (glass tubes are weighted first).

4. Add 3.5 ml methanol, 1.5 ml water and 2 ml chloroform give a final ratio of chloroform/methanol/water equal to 4:8:3.

5. Incubate the mixture for 3 hours at room temperature with end-over-end agitation to extract lipids.

6. Centrifuge 30 minutes at 3300Å~g, 4ºC.

7. Remove the supernatant, add 4 ml methanol, 1.5 ml water and 2 ml chloroform to the insoluble materials and incubate for 2 more hours at room temperature to extract the lipids again. (Note: supernatant can be discarded or stored for analysis of glycolipids.)

8. Centrifuge 30 minutes at 3500 rpm, 4ºC and remove the supernatant.

9. Add 1 ml water to the insoluble materials in glass tube and vortex it well.

10. Add 6 ml ice-cold acetone in glass tube and vortex it well to precipitate the proteins.

11. Incubate for 10 minute on ice and then centrifuge 30 minutes at 3500 rpm, 4ºC and remove the supernatant.

12. Repeat steps 9-11 two more times.

13. Dry the insoluble protein powder under a stream of nitrogen at 40 ºC.

14. Weight the protein powders.

15. Take the same amount of protein powder (3-5 mg each) of normal and ^{15}N-labeled cell populations and mix together to release glycans. (In this protocol, we introduce the method of mixing samples by weighting protein powder. Alternatively, cells labeled light and heavy can be combined before step 1 based on accurate equal cell numbers.).

3. Release of N-linked Glycans

16. Resuspend the mixed protein powder in a 8 ml glass tube with 200 µl of 40 mM NH4HCO3 then boil at 100 ºC for 5 minutes.

17. After cooling to room temperature, add 25 µl of trypsin solution (2 mg/ml in 40 mM NH4HCO3) and 25 µl of chymotrypsin (2 mg/ml in 40 mM NH4HCO3). Then add 250 µl of 2 M urea to result in urea with final concentration of 1 M.

18. Incubate the solution for 18 hours at 37 ºC to digest the proteins. After incubation, boil the solution for 5 minutes to deactivate the enzymes.

19. After cooling to room temperature, add 500 µl 10% acetic acid to the solution then load the solution onto a Sep-Pak C18 cartridge column.

The Sep-Pak C18 column should be equilibrated before use. Normally wash the column three times with 100% acetonitrile followed by three washes with 5% acetic acid.

After loading all the sample, wash the column with 1 ml of 5% acetic acid three times.

Then put collection glass tubes under columns and elute the peptides stepwise, first with 1 ml of 20% isopropanol in 5% acetic acid, then with 1 ml of 40% isopropanol in 5% acetic acid and finally with 1 ml of 100% isopropanol. Mix the elutes and dry down in speed-vac to get rid of solvents.

20. Resuspend the peptides with 45 µl of 100 mM sodium phosphate (pH 7.5) and then add 5 µl PNGase F stock to release the N-linked glycans from the peptides for 20 hours at 37 ºC.

21. After PNGase F digestion, add 450 µl of 5% acetic acid to solution then load solution onto equilibrated Sep-Pac C18 columns (see Step 19 for equilibration). Elute the N-linked glycans with 1 ml of 5% acetic acid for three times and collect in a glass tube. Dry down in speed-vac for permethylation (proceed to step 28). (Note: peptides can be eluted as well if desired with a high organic solvent percentage in the 5% acetic acid).

4. Release of O-linked Glycans

22. In a glass tube, add 1 ml of 1 M sodium borohydride to mixed protein powder made in step 15, vortex and sonicate the sample tube in water bath quickly and then incubate for 18 hours at 45 ºC.

23. After incubation, cool the sample to room temperature. Add 10% acetic acid dropwise to the tube while vortexing it until bubbling stops.

24. Pack 1 ml bed volume of resin stock into a Pasteur pipette to make the cation exchange column and wash the column with 2 mL methanol, 1 mL 1M hydrochloric acid and 2 mL 5% acetic acid.

25. Load sample onto resin columns, elute the O-linked glycans with 7 ml of 5% acetic acid and collect in a glass tube. (Note: peptides can be recovered with high salt or high

pH wash if desired). Dry down in speed-vac.

26. Add 1.5 mL of methanol/glacial acetic acid mixture with a ratio of 9:1 and then dry the sample under a stream of dry nitrogen. Repeat this step two more times to remove the borate. Proceed to step 28 for permethylation.

5. Permethylation

27. Prepare the dry sodium hydroxide solution for permethylation:

1) In a glass tube, combine 100 µl of 50% w/w sodium hydroxide solution and 200 µl anhydrous methanol and vortex briefly.

2) Add 4 ml anhydrous dimethyl sulfoxide (DMSO) and vortex.

3) Centrifuge the tube quickly at 3000 rpm at room temperature and pipette off dry DMSO, salts and white residue with clean sodium hydroxide solution left in the tube.

4) Repeat step 2) and 3) four-five more times to remove all the white residues in the tube.

5) Once the tube is clean, add 2 ml dry DMSO and pipette up and down gently.

28. Add 200 µl dry DMSO to the released N- or O-linked glycan samples (from steps 21 or 26), purge the tube with dry N2 to remove air. Then sonicate in water bath for 2 minutes and vortex quickly to dissolve sample.

29. Add 250 µl of the prepared sodium hydroxide solution to sample tube, purge with dry N2 and sonicate in water bath quickly.

30. Add 100 µl iodomethane with 500 µl micro syringe to the sample, purge with dry N2 and vortex vigorously for 5 minutes.

31. Add 2 ml water and bubble off iodomethane with dry N2 gently. After the solution becomes clear, add 2 ml dichloromethane and vortex.

32. Centrifuge the tube quickly at 3000 rpm at room temperature and then remove the top aqueous layer.

33. Add 2 ml water, vortex and centrifuge quickly at 3000 rpm at room temperature, and then remove the aqueous layer.

34. Repeat step 33 four more times and dry the sample under gentle stream of dry N2.

6. Mass Spectrometry analysis of Glycans

35. Dissolve the permethylated N- or O-linked glycans respectively in 30 µl of 100% methanol. Take 15 µl of the solution to a 1.5 ml eppendorf tube, add 35 µl of 1 mM NaOH in 50% methanol to make a final volume of 50 µl.

36. Infuse the solution directly into the mass spectrometer using a nanospray ion source with a fused-silica emitter (360 Å~ 75 Å~ 30 µm, SilicaTipTM, New Objective) at 2.0 kV capillary voltage, 240 °C capillary temperature, and a syringe flow rate of 0.4 µl/min.

37. Acquire the full FTMS (Fourier Transform mass spectrometry) spectra at 400-2000 m/z in positive ion and profile mode with 2 microscans and 1000 maximum injection time (ms). (Note: this step must be optimized for your particular instrument of choice and can be combined with tandem mass spectrometry analysis of your analytes).

38. N- or O-linked glycan structures can be manually interpreted using glycoworkbench[7] for example but is outside the scope of this review. Calculate the ratios of the same glycan structures in each sample (normal and 15N-labeled) according to Support Protocol.

Support Protocol 1: Calculating Relative Ratios Of Glycans

The IDAWG technology is designed to compare the quantities of each glycan structure in different samples. Because the reagent with amide-15N-Gln used in cell

culture to label the glycans is not 100% pure, there will be under-incorporation of 15N for each glycan structure labeled. The under-incorporation rate could be different from one structure to another and this value is needed when calculating the relative ratios. Thus it is useful to perform an extra procedure of glycan analysis with just the 15N-labeled sample alone instead of the mixture of both light and heavy. This support protocol describes the mathematical calculations of underincorporation rate and relative ratios of light/heavy species based on an example spectra of a ^{15}Nlabeled N-linked glycan structure (Figure 2.2 A) and a mixture of a light and heavy O-linked glycan structure (Figure 2.2 B). It should be noted that a software package to automate the calculations is currently under development.

1. Calculations of Under-Incorporation Rate

To calculate the Under-Incorporation rate, take the ratio of the area of peaks that correspond to under labeling to the total area of all isotopic peaks resulting from one glycan structure. For example, in **Figure 2.2 A**, the area of isotopic peak (labeled as $^{13}C_0{}^{15}N_0$, $^{13}C_0{}^{15}N_1$, $^{13}C_0{}^{15}N_2$, $^{13}C_1{}^{15}N_2$, $^{13}C_2{}^{15}N_2$, $^{13}C_3{}^{15}N_2$, $^{13}C_4{}^{15}N_2$) is A_i (i=1, 2, 3, 4, 5, 6, 7), so the overall Under-Incorporation rate (UI) can be calculated as:

$$UI=\frac{\sum_{i=1}^{2} A_i}{\sum_{i=1}^{7} A_i}$$

For any glycan structure, if the number of nitrogens in the molecule is K and the number of isotopic peaks shown in the spectrum is M,

$$UI=\frac{\sum_{i=1}^{K} A_i}{\sum_{i=1}^{M} A_i}$$

And the incorporation rate for each nitrogen can be calculated as:

$$Incorporation = \sqrt[K]{(1-UI)}$$

2. Calculations of Relative Ratios

1. Generate the theoretical isotopic patterns for the light structure by software called "emass" written by the Somerharju Lipid Group, University of Helsinki. (http://www.helsinki.fi/science/lipids/software.html)
2. In **Figure 2.2 B**, there are 7 major isotopic peaks observed in the spectrum. Calculate the area of each peak (A_1 through A_7) using peak list and intensities extracted from the spectrum. Due to the peak overlapping and under incorporation, both of the light and heavy structures can contribute to the area of each peak. So if the actual area resulting from light structure is L_1 through L_7, from heavy structure is H_1 through H_7, then we have:

$$L_1 + H_1 = A_1 \quad (1)$$

$$L_2 + H_2 = A_2 \quad (2)$$

$$L_3 + H_3 = A_3 \quad (3)$$

$$L_4 + H_4 = A_4 \quad (4)$$

$$L_5 + H_5 = A_5 \quad (5)$$

$$L_6 + H_6 = A_6 \quad (6)$$

$$L_7 + H_7 = A_7 \quad (7)$$

3. If the ratio of theoretical isotopic pattern for the 7 peaks from the light structure is $P_1(=1):P_2:P_3:P_4:P_5:P_6:P_7$, which can be generated by using the software "emass", we can have:

$$L_1 = L_1 * P_1 \ (P_1 = 1) \quad (8)$$

$$L_2 = L_1 * P_2 \quad (9)$$

$$L_3 = L_1 * P_3 \quad (10)$$

$$L_4 = L_1 * P_4 \quad (11)$$

$$L_5 = L_1 * P_5 \quad (12)$$

$$L_6 = L_1 * P_6 \quad (13)$$

$$L_7 = L_1 * P_7 \quad (14)$$

4. By the definition of under incorporation rate shown before,

$$UI = \frac{\sum_{i=1}^{3} Hi}{\sum_{i=1}^{7} Hi} \quad (15)$$

Determine the value of UI based on the spectrum for heavy labeled glycans. Given that $H_i = A_i - L_i$, rearrange equation 15, with a substitution of H_i into A_i and L_i according to equations 1 – 14, to yield equation 16:

$$L1 = \frac{A1+A2+A3-UI*\sum_{i=1}^{7} Ai}{P1+P2+P3-UI*\sum_{i=1}^{7} Pi} \quad (16)$$

5. After calculation of L_1, calculate L_i (i=2-7) and H_i (i=1-7) based on equations 1-16. The equation for Heavy/Light ratio will be:

$$\frac{\text{Amide-}^{15}\text{N-GLN}}{\text{Amide-}^{14}\text{N-Gln}} = \frac{\sum_{i=1}^{7} H_i}{\sum_{i=1}^{7} L_i} \quad (17)$$

6. For any glycan structure, if the number of nitrogens in the molecule is K and the number of isotopic peaks shown in the spectrum of light/heavy mixture is M, the equations 16 and 17 can be rewritten as

$$L1 = \frac{\sum_{i=1}^{K} Ai-UI*\sum_{i=1}^{M} Ai}{\sum_{i=1}^{K} Pi-UI*\sum_{i=1}^{M} Pi} \quad (18)$$

$$\frac{\text{Amide-}^{15}\text{N-GLN}}{\text{Amide-}^{14}\text{N-Gln}} = \frac{\sum_{i=1}^{M} H_i}{\sum_{i=1}^{M} L_i} \quad (19)$$

Take the structures in Figure 2.2 B as an example, the ratio of theoretical isotopic pattern for the 7 peaks from the light structure generated by using the software “emass” is

$P_1(=1):P_2:P_3:P_4:P_5:P_6:P_7 = 1:0.6517:0.264:0.0798:0.0198:0.0042:0.0008$, and the total is $\sum_{i=1}^{7} Pi = 2.0203$.

The peak areas can be calculated by extracting the peak list and peak intensities from raw data and using the software "OriginPro 8" (http://www.originlab.com/) and the areas are:

$A_1 = 288182.7$; $A_2 = 180298.2$; $A_3 = 91459.51$; $A_4 = 276748.8$; $A_5 = 153348.9$; $A_6 = 54793.07$; $A_7 = 10981.6$; the total is $\sum_{i=1}^{7} Ai = 1055813$.

UI (under-incorporation rate)_of this structure is 0.17 (using the calculating method introduced before). Take all the values into equation (16) then we have

$$L1 = \frac{A1+A2+A3\text{-}UI*\sum_{i=1}^{7} Ai}{P1+P2+P3\text{-}UI*\sum_{i=1}^{7} Pi} = \frac{288182.7+180298.2+91459.51-0.17\times1055813}{1+0.6517+0.264-0.17\times2.0203} = 241979.6$$

After the calculation of L_1, based on equations (9-14), we can calculate L_2 through L_7. Based on equations (1-7), we can then calculate H_1 through H_7. And the totals are:

$\sum_{i=1}^{7} L_i = L_1 + L_2 + L_3 + L_4 + L_5 + L_6 + L_7 = 488872$

$\sum_{i=1}^{7} H_i = H_1 + H_2 + H_3 + H_4 + H_5 + H_6 + H_7 = 566941$

Based on equation (19),

$$\frac{\text{Amide-}^{15}\text{N-GLN}}{\text{Amide-}^{14}\text{N-Gln}} = \frac{\sum_{i=1}^{M} H_i}{\sum_{i=1}^{M} L_i} = \frac{566941}{488872} = 1.16$$

COMMENTARY

Background Information

Glycomics comprehensively studies all the glycan structures in a given biological system. The glycans are complex carbohydrates which are oligosaccharides chains

usually linked to proteins and lipids[1]. Glycosylation is a process of covalent attachment of glycan structures on protein backbones, which is one of the most common post-translational modifications of proteins[8]. It has been estimated that glycosylated proteins account for approximately 60-80% of all mammalian proteins at some point during their existence and nearly 100% of all membrane and secreted proteins[9]. There are two main ways that carbohydrate chains are linked to protein backbones: N-linked glycan is linked through the side chain of an asparagine residue present in the tripeptide consensus sequence – Asn-X-Thr/Ser (where X can be any amino acid except proline); O-linked glycan is attached to the oxygen on the side chain of a serine or threonine[8].
Each of these glycosylation sites can be attached with many different glycan structures.

Glycans often play critical roles in various physiological processes. The important functions of glycans include but are not limited to: modulation of biological activity, cell-cell recognition and interaction, distribution in tissues and signal transduction[10]. As affected by different factors, the expression of glycans in a biological system can vary with its species, tissue, developmental stage and even the genetic and physiological state[9]. It has also been noticed that altered carbohydrates expression is a common feature of many types of cancers[2]. Moreover, particular protein glycosylations may be altered more specifically or frequently than their underlying core protein in certain disease state, which is a potential advantage of using glycans for diagnostics[11]. Given all the important roles of glycans in physiological processes, considerable effort has been taken to develop technologies to identify and quantify the glycan structures in various environments.

It is one of the major challenges in the fields of –omics to develop relative-quantitative analysis technologies that are able to generate meaningful data to compare

the expression levels of targeted molecules in different biological samples or developmental states. For proteomics people already have some powerful tools in hand. Since mass spectrometry has fulfilled its role as a rapid and reliable method for proteomics study, MS signal intensities, ion chromatograms, spectral counts, and accurate mass retention time pairs have been used as label-free methods to quantify the changes in protein abundances[12-15]. Other than label-free methods, stable isotopic labeling methods are more popular in recent years. Among all these labeling methods, ICAT (Isotope-Coded Affinity Tags) as an in-vitro labeling strategy and SILAC (Stable Isotope Labeling with Amino acids in Cell culture) as an in-vivo labeling strategy are broadly applied and already commercialized[16,17].

Compared to proteomics, analytical technology for comparable glycomics is not that mature. However, the field of glycomics has followed in the steps of proteomics and adapted some of the quantitative proteomics tools for glycomic analysis successfully. For example, the method of total ion mapping, which is a label-free strategy, is capable of identifying glycan structures based on fragmentation information from tandem MS and quantifying the prevalence of each glycan structure by normalizing each ion intensity seen on the mass spectrum to the total[4]. As for in-vitro isotopic labeling methods, several groups have used heavy methyl iodide ($^{13}CH_3I$, $^{12}CDH_2I$, $^{12}CHD_2I$, and/or $^{12}CD_3I$) to label glycans during the permethylation procedure which is normally performed before MS analysis for both N- and O-linked glycans. Then the glycan samples incorporated with heavy isotopes are mixed with samples processed with standard permethylation using light methyl iodide ($^{12}CH_3I$) prior to analysis[3-5].

In 2009, IDAWG (Isotopic Detection of Aminosugars With Glutamine) was reported as the first in-vivo stable isotopic labeling strategies for quantitative glycomics[6]. The IDAWG methodology takes advantage of the hexosamine biosynthetic pathway, which uses the side chain of glutamine as the only source of nitrogen when producing aminosugars. As a result, if the cells are fed with Gln-free media and glutamine with a ^{15}N-labeled side chain, all the aminosugars (GlcNAc, GalNAc and sialic acids) produced in the cells will be labeled with ^{15}N and thus the mass of all glycan structures will be shifted by +1 dalton per aminosugar (See Figure 2.3). As an in-vivo labeling strategy, IDAWG shares some advantages with SILAC over invitro labeling methods. After the labeling, the 15N-labeled sample can be mixed with the normal sample immediately after the cell harvest or the cell lysis. Thus the glycans from two cell types are subjected to the same experimental conditions for glycan release and permethylation until they are analyzed by MS, which will dramatically reduce technical variability. The IDAWG technology has been used to analysis both N- and O-linked glycans released from murine embryonic stem cells successfully and is predicted to be useful for various comparative glycomic studies in the future[6].

Critical Parameters

The method of quantitative glycomics of cultured cells using IDAWG consists of six steps: (1) Cell culture: cells are either fed with amide-^{14}N-Gln or amide-^{15}N-Gln. (2) Cell lysis: differently labeled cell populations are lysed, delipidated and then mixed together. (3) Release of glycans: N- and O-linked glycans are released from glycopeptides enzymatically or chemically. (4) Permethylation: released glycans are

permethylated prior to mass spectrometry analysis. (5) Mass spectrometry analysis and data collecting. (6) Calculations of relative ratios.

For the cell culture step, labeling using amide-15N-Gln for three days should be enough for embryonic stem cells to get a good incorporation rate. However, for those cells with a relatively slower metabolic rate, longer labeling time is required in order to get a high degree of labeling. A recommendation of 7 days is advised for new cell lines.

For the cell lysis step, there are three parameters to consider. First, the delipidating solvent should have the ratio of chloroform:methanol:water equal to 4:8:3 or it will introduce layers to the solvent and impair the delipidation. Second, washing with ice-cold acetone and water after delipidation is crucial because acetone helps precipitate proteins and water can dissolve oligomeric hexose ladders which are common contaminants for glycan analysis. Third, if it is possible to count the cell number, mixing the two cell populations together immediately after cell culture by the same cell number instead of by the protein weight is also an option to perform the IDAWG experiments.

For the step of releasing O-linked glycans, an extra cleanup with C18 columns may be added to the protocol to help remove the borate. The sample can be dissolved in 5% acetic acid and loaded onto the columns and eluted with 5% acetic acid. Throughout all the procedures, when drying the sample under the stream of nitrogen gas, always keep the drying time as short as possible. Staying too long under the nitrogen will likely reduce the yield of the experiments because some product could be volatilized or displaced from the tube by the nitrogen gas after the solution is dried.

Low ^{15}N incorporation rate

The 15N labeling time in cell culture is too short. Label embryonic stem cells for three days and label differentiated cell types for at least seven days to get sufficiently high (>90%) incorporation rates.

No glycans detected in mass spectrum

1. Incomplete enzymatic digestion or chemical reaction. Make sure the right amount of enzyme or chemical reagents are used and right conditions are applied (see Basic Protocol).

2. Too much contamination is introduced through all the procedures. Make sure the protein powder is washed with acetone and water properly. All the eppendorf tubes, glass tubes and pipettes must be pre-cleaned with methanol or the polymer peaks can dominant in the spectra.

Anticipated Results

Example spectra are shown in Figure 2. The relative ratio of any glycan structure with a similarly complete isotopic pattern (typically with more than 2 isotopic peaks showing up) can be calculated using IDAWG. For example, more than 125 N-linked glycans and 35 O-linked glycans have been identified by our group for mouse embryonic stem cells using these protocols. Among these glycan structures, those that could be reliably relatively quantified by IDAWG were approximately half (about 65 structures for N-linked glycans and 20 for O-linked glycans, these glycans have more than 2 isotopic peaks in the spectra which make the calculation reliable) Of course, some of these quantified glycans are in fact isobaric mixtures and this should be taken into account

when interpreting results. IDAWG is not able to separately quantify the isobaric structures and thus a ratio is calculated based on changes in a set of isobaric structures.

Time Considerations

The cell culture needs at least 72 hours to label the cells and may take longer than a week. The cell lysis and delipidation requires incubation of 5 hours to overnight. The release of N-linked glycan requires two overnight incubations for digestion and the release of O-linked glycan requires one overnight incubation for chemical reaction. The time for drying samples in the speed-vac is variable depending on the efficiency of speed-vac and solvent volume used. It takes less than 5 hours to perform permethylation and mass spectrometry analysis. Thus, the entire procedure excluding the initial labeling and data interpretation takes 3-4 days. We, working with others in the field, are currently attempting to automate the calculations for under incorporation and relative ratios. Furthermore, the introduction of spiked standard glycans is being explored so that IDAWG labeling can be used to follow glycan turnover, remodeling, and synthesis.

REFERENCE

1 Aoki-Kinoshita, K. F. An introduction to bioinformatics for glycomics research. *PLoS Comput Biol* **4**, e1000075, doi:10.1371/journal.pcbi.1000075 (2008).

2 Dube, D. H. & Bertozzi, C. R. Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat Rev Drug Discov* **4**, 477-488, doi:nrd1751 [pii] 10.1038/nrd1751 (2005).

3 Alvarez-Manilla, G. *et al.* Tools for glycomics: relative quantitation of glycans by isotopic permethylation using 13CH3I. *Glycobiology* **17**, 677-687, doi:cwm033 [pii]
10.1093/glycob/cwm033 (2007).

4 Aoki, K. *et al.* Dynamic developmental elaboration of N-linked glycan complexity in the Drosophila melanogaster embryo. *J Biol Chem* **282**, 9127-9142, doi:M606711200 [pii]
10.1074/jbc.M606711200 (2007).

5 Kang, P., Mechref, Y., Kyselova, Z., Goetz, J. A. & Novotny, M. V. Comparative glycomic mapping through quantitative permethylation and stable-isotope labeling. *Anal Chem* **79**, 6064-6073, doi:10.1021/ac062098r (2007).

6 Orlando, R. *et al.* IDAWG: Metabolic incorporation of stable isotope labels for quantitative glycomics of cultured cells. *J Proteome Res* **8**, 3816-3823, doi:10.1021/pr8010028 (2009).

7 Ceroni, A. *et al.* GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res* **7**, 1650-1659, doi:10.1021/pr7008252 (2008).

8 Morelle, W., Faid, V., Chirat, F. & Michalski, J. C. Analysis of N- and O-linked glycans from glycoproteins using MALDI-TOF mass spectrometry. *Methods Mol Biol* **534**, 5-21, doi:10.1007/978-1-59745-022-5_1 (2009).

9 Atwood, J. A., 3rd *et al.* Quantitation by isobaric labeling: applications to glycomics. *J Proteome Res* **7**, 367-374, doi:10.1021/pr070476i (2008).

10 Lowe, J. B. & Marth, J. D. A genetic approach to Mammalian glycan function. *Annu Rev Biochem* **72**, 643-691, doi:10.1146/annurev.biochem.72.121801.161809 121801.161809 [pii] (2003).

11 Yue, T. *et al.* The prevalence and nature of glycan alterations on specific proteins in pancreatic cancer patients revealed using antibody-lectin sandwich arrays. *Mol Cell Proteomics* **8**, 1697-1707, doi:M900135-MCP200 [pii] 10.1074/mcp.M900135-MCP200 (2009).

12 Wang, W. *et al.* Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal Chem* **75**, 4818-4826 (2003).

13 Liu, H., Sadygov, R. G. & Yates, J. R., 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76**, 4193-4201, doi:10.1021/ac0498563 (2004).

14 Radulovic, D. *et al.* Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry. *Mol Cell Proteomics* **3**, 984-997, doi:10.1074/mcp.M400061-MCP200 M400061-MCP200 [pii] (2004).

15 Silva, J. C. *et al.* Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem* **77**, 2187-2200, doi:10.1021/ac048455k (2005).

16 Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994-999, doi:10.1038/13690 (1999).

17 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).

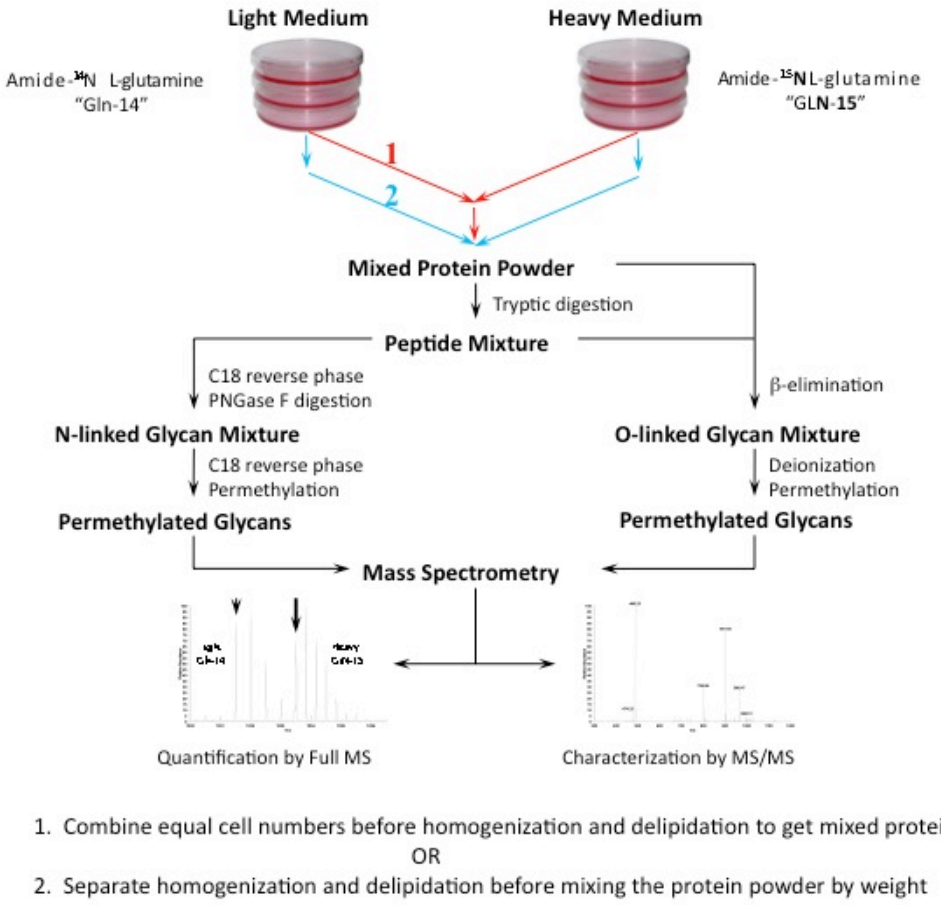


Figure 2.1 Schematic workflow of IDAWG labeling strategy for comparative glycomics. We can eigher combine equal cell numbers before homogenization and delipidation to get mixed protein powder or process the homogenization and delipidation before mixing the protein powder by weight.

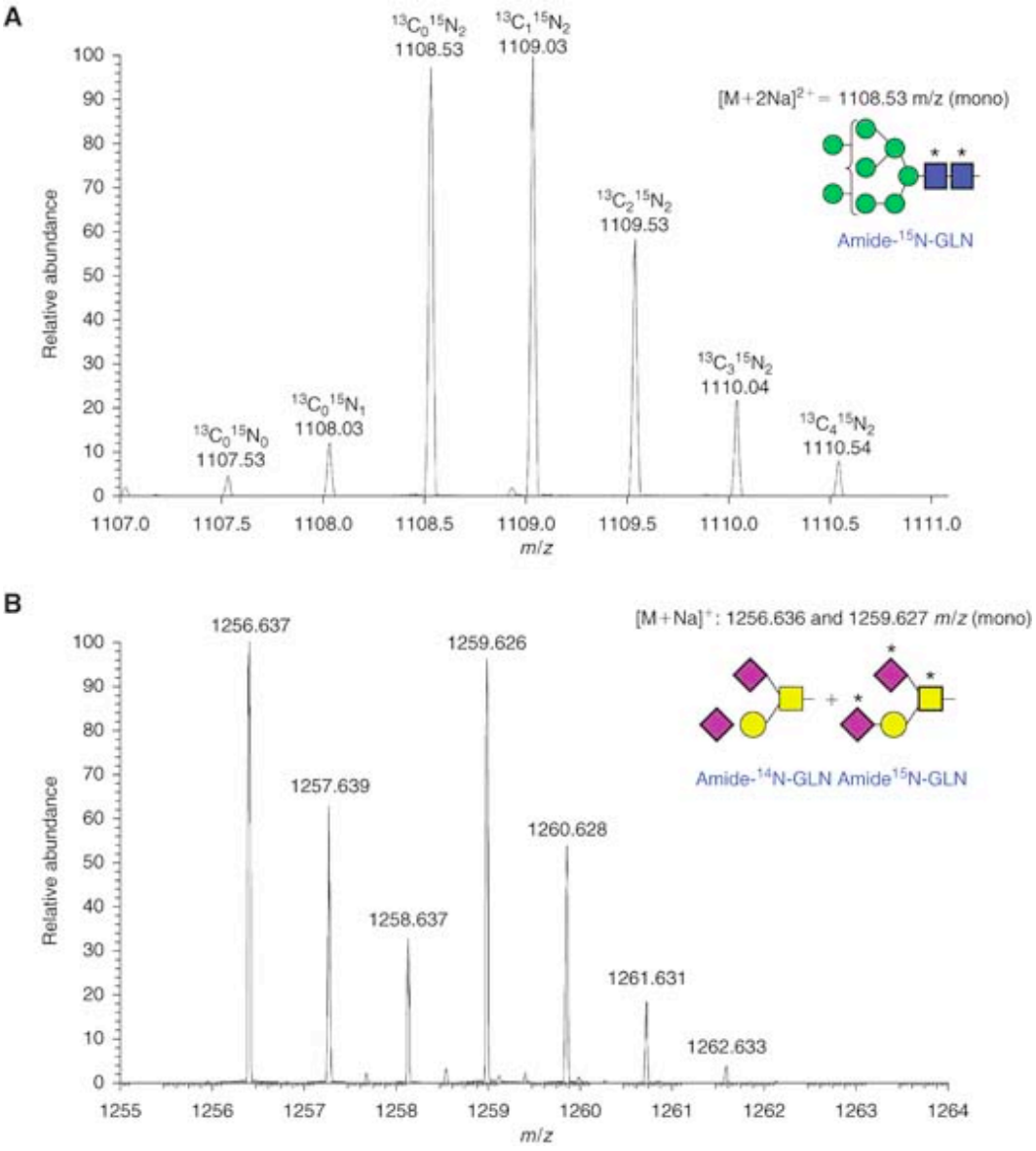


Figure 2.2 Example mass spectra of IDAWG glycans. **A.** Isotopic pattern of ^{15}N-labeled N-linked glycan structure ($Man_8GluNAc_2$). **B.** Isotopic pattern of mixture of normal and ^{15}N-labeled O-linked glycan structure**.** Green circle: mannose; Blue square: GlcNAc; Purple diamond: Neu5Ac; Yellow circle: galactose; Yellow square: GalNAc.

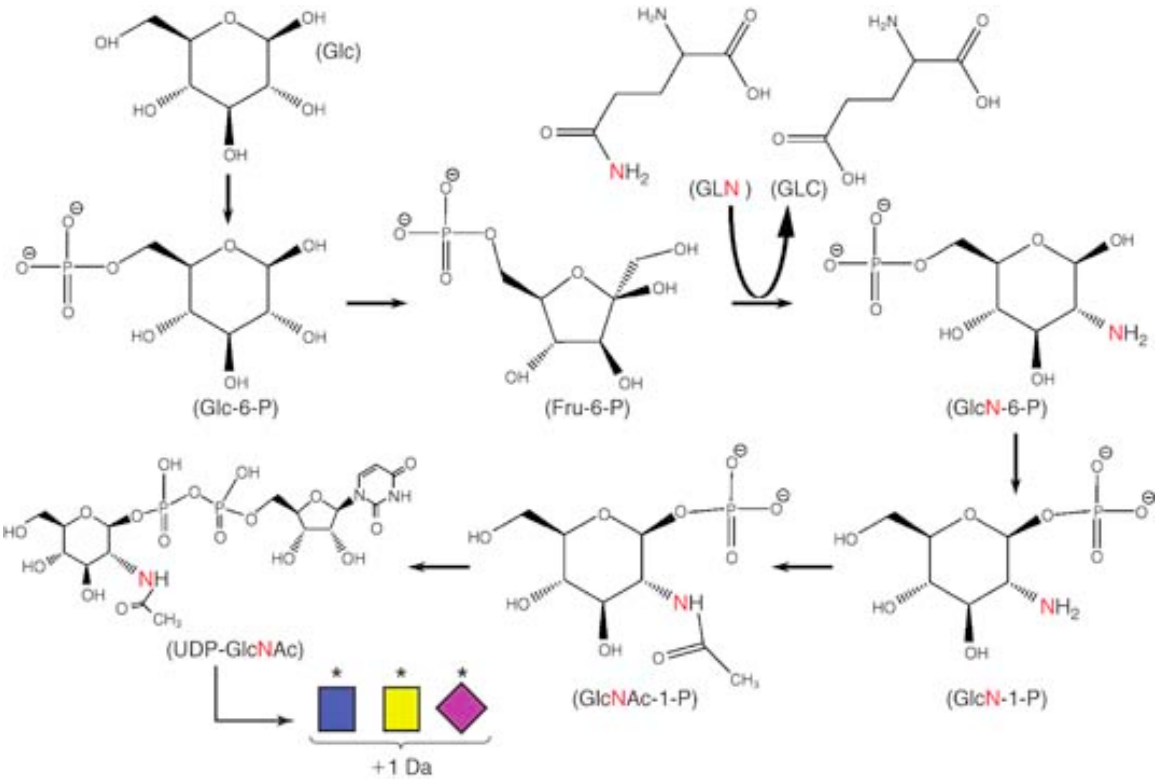


Figure 2.3 Schematic of the hexosamine biosynthetic pathway that demonstrates that the side chain of glutamine is the only source for nitrogen in the production of aminosugars. If the cells are fed with Gln-free media and glutamine with a ^{15}N-labeled side chain, all the aminosugars including GlcNAc (blue square), GalNAc (yellow square) and sialic acids (purple diamond) produced in cells will be labeled with ^{15}N and thus the mass of all glycan structures will be shifted by +1 dalton per aminosugar [6]

CHAPTER 3

GLYCOMIC PROFILING OF HUMAN EMBRYONIC STEM CELLS AND HUMAN DEFINITIVE ENDODERMS WITH RELATIVE QUANTIFICATION USING THE IDAWG TECHNIQUE[1]

[1] Meng Fang, Jae-Min Lim, Michael Kulik, Jun Han, Sandii Brimble, Alison Nairn, Kelley Moremen, Stephen Dalton, John Miller, Will York and Lance Wells
To be submitted to *Journal of Proteome Research.*

ABSTRACT

Glycosylation is known to play an important role in many physiological processes and numerous Mass Spectrometry-based technologies have been developed to compare the glycan expression levels between different biological states. IDAWG (Isotopic Detection of Aminosugars With Glutamine) was reported as the first in-vivo stable isotopic labeling strategies for quantitative glycomics, which labels all the aminosugars (GlcNAc, GalNAc and sialic acids) with ^{15}N in cell culture. In our study, differentially labeled human embryonic stem cells were used to validate the reliability and quantitative accuracy of the IDAWG technology. Meanwhile, as focusing on the O-linked glycans, we performed a global glycomic profiling of human embryonice stem cells and human definitive endoderms derived from hESCs followed by a relative quantification of glycan expression levels between hESCs and hDEs using the IDAWG technique, which revealed a reciprocal change between poly-LacNAc and sialyation of O-linked core 2 structures during the differentiation from hESCs to hDEs.

INTRODUCTION

Glycosylation is one of the most common post-translational modifications of proteins[1] and as a comprehensive study of all glycan structures in a given biological system, the glycomics has drawn more and more attentions[2,3]. It has been estimated that glycosylation is occurring in up to half of all gene products[4] and glycosylated proteins account for approximately 60-80% of all mammalian proteins at some point during their existence and nearly 100% of all membrane and secreted proteins[5]. Glycans often play critical roles in various physiological processes such as cell development and differentiation,[6-11] cell-cell recognition and interaction,[12,13] tumor growth and metastasis,[14-19] and immune recognition and response[20-24]. As affected by different factors, the expression of glycans in a biological system can vary with its species, tissue, developmental stage and even the genetic and physiological state and it has also been noticed that altered carbohydrates expression is a common feature of many types of cancers[17]. Besides, particular carbohydrate modifications may be altered more specifically or frequently than their underlying core protein in certain disease state, which is a potential advantage of using glycans for diagnostics[25,26].

Given all the important roles of glycans in physiological processes and potential application of glycans for disease diagnostics, considerable effort has been taken to develop MS-based technologies to identify and quantify the glycan structures in various environments and numerous strategies have been developed to compare the expression levels of glycans between different biological states. Among all these analytical techniques, the label-free methods[27,28] are easy to implement and rely on the direct comparison of normalized ion intensities from two or more glycan samples, however, the

instrument response can differ over time and ionization efficiency can be affected by different factors such as molecular weight, physical or chemical characters of analyte molecules and the presence of other compounds competing or interfering with the ionization of analyte. To overcome these issues, numerous stable isotope labeling methods have been developed in recent years[27,29,30]. In 2009, IDAWG (Isotopic Detection of Aminosugars With Glutamine) was reported as the first in-vivo stable isotopic labeling strategies for quantitative glycomics, which labels all the aminosugars (GlcNAc, GalNAc and sialic acids) with ^{15}N in cell culture[31]. As an in-vivo labeling strategy, IDAWG possesses a significant advantage over in-vitro labeling method, similar to SILAC for proteomics,[32] which is that after the labeling, the ^{15}N-labeled sample can be mixed with the normal sample immediately after the cell harvest or the cell lysis. Thus the glycans from two cell types are subjected to the same experimental conditions for glycan release and permethylation until they are analyzed by MS, which will dramatically reduce technical variability. IDAWG technique is expected to provide for robust quantification of glycan expression differences for various cell lines that can be differentially labeled during cell culture.

Human Embryonic Stem cells (hESCs) are derived from the inner cell mass (ICM) of blastocyst, an early stage embryo[33]. Studies of human embryonic stem cells can provide for the information about the complex process of human development and the course of some serious medical conditions such as cancer and birth defects. The generation of cells and tissues by hESCs can also provide potential therapeutic applications in tissue regeneration and transplantation[34]. Human embryonic stem cells also retain the capacity to differentiate into the embryonic germ lineages including

definitive endoderms from which a lot of organs of interest such as pancreas are derived[35,36]. Since glycans are abundant components in the stem cell surface, efforts have been put into the glycomic analysis of human embryonic stem cells for recent years. For example, the pluripotency of human embryonic stem was defined using lectin microarray analysis as a novel comprehensive approach for quality control in cell-based therapy and regenerative medicine[37]. Besides, MALDI-TOF mass spectrometry has been used to profile N-glycome of human embryonic stem cells and cells differentiated from them[38]. Here in our study, we performed a global glycomic profiling of human embryonic stem cells and human definitive endoderms derived from hESCs for both O- and N-linked glycans, followed by a relative quantification of glycan expression levels between hESCs and hDEs using the IDAWG technique. Meanwhile, by performing the IDAWG experiment using differentially labeled hES cells, we were able to validate the quantitative accuracy of the IDAWG technology and demonstrate the capability of the IDAWG technology as a powerful tool for comparative glycomic studies in cell culture.

EXPERIMENTAL PROCEDURES

Cell Culture

The BG02 line of human embryonic stem cells (hESCs) and human definitive endoderms (hDEs) differentiated from BG02 cells were cultured essentially as previously described[39]. Briefly, the hESCs or hDEs were grown on gelatin coated petri dishes at 37^{o}C 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 the cells were allowed to grow for 72 hours at which point they were harvested by dissociation buffer and scraping. The cell pellets were washed with phosphate buffered saline (PBS) and stored at -80 ^{o}C until analysis.

Preparation of Light/Heavy hESC/hDE Protein Powder and delipidation

Collected light and heavy labeled hES/hDE cell pellets were delipidated as previously described[31,40]. Briefly, the pellets were disrupted on ice by Dounce homogenization in ice-cold 100% methanol. To extract the lipids, chloroform and water were added to the solvent mixture to give a final ratio of chloroform/methanol/water equal to 4:8:3. The resulting mixture was incubated for 3 hours at room temperature with end-over-end agitation. By centrifugation at 3,000rpm for 15min at 4^{o}C, the insoluble material was extracted again and incubated for 2 more hours. After washed with cold water and cold acetone for three times, the final pellets of insoluble protein were dried under a stream of nitrogen. For the relative quantification, the differentially labeled hESC and hDE protein powders were mixed by weight at a ratio of 1:1.

Glycan Release and Permethylation

The release and permethylation of N- and O-linked glycans was essentially as previously described[27,28,31]. Briefly, the resulting protein powders were either digested by

trypsin and subjected to PNGase F treatment to release N-linked glycans or treated with alkaline borohydride to release O-linked glycans by β-elimination. Released glycans were cleaned up by reverse-phase columns and dried down before they were permethylated with methyl iodide.

MS Analysis of the Permethylated Glycans

The glycans were analyzed as previously described[27,28,31] on a hybrid linear ion trap Orbitrap mass spectrometer (LTQ-Orbitrap, Thermo). 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 (TIM). 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 glycoworkbench software[41]. The IDAWG quantification data was mainly deconvoluted and analyzed by

GlycoQuant and the manual deconvolution method described before[31,40] was used as a supplement.

RESULTS AND DISCUSSION

Comparative glycomics of the hES/hDE cells with IDAWG strategy

Before we could apply IDAWG strategy to comparative glycomics of hES and hDE cells, the natural population of each cell line was processed to generate N- and O-linked glycan profiles using ESI-LTQ-Orbitrap mass spectrometer. As an *in-vivo* labeling strategy similar to SILAC for proteomics, IDAWG strategy carries a significant advantage when applied to comparative glycomics, which is that the contribution of handling and workup to overall variability can be minimized by mixing differentially labeled cells together at the beginning of the analytic procedure: in this case, immediately after the delipidation (**Figure 3.1**). Both N- and O-linked glycans were released, permethylated and quantified by IDAWG.

For the quantification experiment, a newly designed instrumental method called rolling trapping scans was applied to increase the signal to noise ratio of full mass spectra and present more isotopic peaks for each individual glycan structures. Automated MS/MS spectra without applying collision energy were acquired in this mode. For each MS/MS scan, the instrument trapped the ions with a width of 100 m/z but only scanned 50 m/z around the center of the trapping window to ensure the trapping efficiency. The trapping center of each scan was moved automatically by 25 m/z to the higher mass region compared to its previous scan so that the range from 400 to 2000 m/z was covered by these stepwise trapping scans (**Figure 3.2**).

The glycan profiles of hES cells

N- and O-linked glycans were released respectively from hES cells and derived hDE cells and permethylated prior to the analysis on a hybrid LTQ-Orbitrap instrument. From the full spectra acquired for hES and hDE glycan sample (**Figure 3.3** and **Figure 3.S1**), several predominant oligosaccharides, mainly singly or doubly charged high mannose structures ($Man_{5-9}GlcNAc_2$) for N-linked and singly charged core 1, core 2 and core 4 structures for O-linked glycans, can be detected and recognized. All samples were subjected to total ion mapping analysis for detecting minor glycans and MS^n method was utilized to confirm structure assignments as necessary. In total, 22 O-linked glycans (**Table 3.1**) and 86 N-linked glycans (**Table 3.S1**) were identified in hES sample (isobaric structures were not counted). The N-linked glycans released from hES cells are comprised of high mannose, hybrid and complex structures which can be classified into four categories based on sialylation and fucosylation, while O-linked glycans cover core 1, core 2, core 4 and O-Man initiated structures and sialylated and fucosylated structures are also seen in O-glycan profiles. The same glycan profiles were identified in hDE sample, yet no evidences showed the existence of novel structures. To facilitate the further quantifications by IDAWG, the prevalence of each glycan identified in hES sample was calculated as the percentage of total profile which was taken as the sum of peak areas of all quantified glycans (**Table 3.1** and **Table 3.S1**).

IDAWG quantification of Light/Heavy hES cells

To demonstrate the quantitative accuracy of the IDAWG technology, hES cells were labeled in Amide-^{15}N-Glutamine media for a sufficient time and then mixed with hES cells grown in media with light glutamine by weight with a ratio of 1:1 immediately

after both cell populations were delipidated and converted to protein powder. N- and O-linked glycans were released from mixed samples and permethylated respectively and full mass spectra were acquired using rolling trapping scans (**Figure 3.4**). The experiments were performed three times to give technical triplicates. On the full spectra of N- or O-linked glycans, the monoisotopic peaks representing the light structures and their labeled counterparts are separated by a certain m/z value based on the number of aminosugars and the charge state, however, the two series of isotopic peaks for differentially labeled samples are overlapping with one another (**Figure 3.4** and **Figure 3.S2**). For the purpose of deconvoluting the spectra of mixed samples and calculating the ratios of heavy to light species, the same experiments may be processed only with heavily labeled samples to give ^{15}N incorporation rates for each quantified glycans (data not shown). GlycoQuant was used to fit the spectra and quantified the differences between light and heavy structures for all the N- and O-linked glycans with prevalence not less than 1% in hES samples (**Table 3.2** and **Table 3.S2**). The results are expressed by the ratios of heavy structure amount over light structure and normalized against the average of ratios for all glycans quantified. The ratio for each glycan shown in the tables is the average for all three replicates. The averaged ratios are close to a theoretical value of 1:1 for equally mixed samples with a range from 0.62 to 1.04 for N-linked glycans and from 0.65 to 1.32 for O-linked glycans. The percentages of error for three replicates are also calculated and the averaged error is 18.72% for N-linked glycans and 7.61% for O-linked glycans. The standard deviation for ratios of all glycan quantified is 0.21 for O-glycan and 0.13 for N-glycan. The results demonstrate that IDAWG technology is able to

quantify the difference of glycan expression levels as low as a 2-fold change with confidence.

Application of IDAWG for quantifying glycome change during differentiation from hES to hDE

Human Embryonic Stem cells which are derived from the inner cell mass (ICM) of blastocyst-stage embryos[33] serve as an excellent model to study human development and can also provide potential therapeutic applications in tissue regeneration and transplantation[34]. Human ES cells also retain the capacity to differentiate into the embryonic germ lineages including definitive endoderms from which a lot of organs of interest such as pancreas are derived[35]. In this study we applied IDAWG to compare glycan expression levels in human ES cells and human DEs. To quantify the glycome change between hES and hDE with IDAWG technology, both hES and hDE protein powders were mixed with their ^{15}N labeled counterparts respectively with the same weight to form two mixed samples --- Mixture 1 (non-labeled hES with labeled hDE) and Mixture 2 (non-labeled hDE with labeled hES), then both N- and O-linked glycans were released from these mixed samples, permethylated and subjected to the analysis by LTQ-Orbitrap instrument to acquire full spectra using rolling trapping scan method (**Figure 3.5** and **Figure 3.S3**). In total, the quantifications of O-linked glycans were performed using three sets of samples: two sets of Mixture 1 and one set of Mixture 2; the quantifications of N-linked glycans were performed using four sets of samples: two sets of Mixture 1 and two sets of Mixture 2. For O-linked glycan, each structure with prevalence over 1% was quantified and the result was expressed as the ratio of abundance of this structure in hDE cells over its abundance in hES cells and normalized against the median of ratios for all

glycans quantified for each set of sample. The ratio for each structure is averaged for all sets of samples performed (**Table 3.3**). There are 3 O-linked glycan structures that have changed more than 2-fold during the differentiation from hES to hDE and they are: a mono-sialylated core 2 structure (m/z 1344.69) elevated with a ratio of 2.64, a disialylated core 2 structure (m/z 1705.86) elevated with a ratio of 2.54 and a poly-LacNAc core 2 structure (m/z 1432.74) dropped with a ratio of 0.41. Since these 3 structures can be considered as derivatives of another core 2 structure (m/z 983.52) modified with additions of sialic acid or LacNAc, these results suggest that, during the differentiation of hES into hDE, the sialylation of core 2 structures becomes predominant and it can impede the addition of LacNAc. This result agrees with our q-PCR result, which shows the changes of glycosyltransferases in transcript level (**Figure 3.6 and Figure 3.S3**). Our q-PCR result demonstrates an increase in the abundance of the glycosyltransferases which adds the sialic acid (No.5) and a decrease in the abundance of the glycosyltransferases which adds the LacNAc (No.6) during the differentiation from hESCs to hDEs. For N-linked glycan, the ratios of 5 most abundant structures from each category in Supplemental Table 1 are weighted by their prevalence and averaged to demonstrate the trend of change for each type of N-glycan structures (**Table 3.3**). Although none of these values reaches 2-fold change, it shows a slight decrease of the abundance of high mannose structures and agrees with the results of previous studies[38].

CONCLUSION

Herein, the differentially labeled human embryonic stem cells have been used to validate the reliability and quantitative accuracy of the IDAWG technology by quantifying the expression levels of released N- and O-linked glycans. The results

demonstrate that the IDAWG technology is able to quantify the difference of glycan expression levels as low as a 2-fold change with confidence. As focusing on O-linked glycans, we perform a global glycomic analysis of human embryonic stem cells and human definitive endoderms. The comparative quantitation of glycan expression levels between hESCs and hDEs by the IDAWG technique discovers a reciprocal change between poly-LacNAc and sialyation of core 2 O-linked structures accompanying the differentiation from hESCs to hDEs, which agrees with the changes in transcript level. Along with the maturity of the IDAWG software, we anticipate that the IDAWG technique will become an important tool in the emerging field of comparative glycomics.

REFERENCES

1 Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* 3, 97-130 (1993).

2 Aoki-Kinoshita, K. F. An introduction to bioinformatics for glycomics research. *PLoS Comput Biol* 4, e1000075, doi:10.1371/journal.pcbi.1000075 (2008).

3 Shriver, Z., Raguram, S. & Sasisekharan, R. Glycomics: a pathway to a class of new and improved therapeutics. *Nat Rev Drug Discov* 3, 863-873, doi:10.1038/nrd1521
nrd1521 [pii] (2004).

4 Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* 1473, 4-8, doi:S0304-4165(99)00165-8 [pii] (1999).

5 Atwood, J. A., 3rd *et al.* Quantitation by isobaric labeling: applications to glycomics. *J Proteome Res* 7, 367-374, doi:10.1021/pr070476i (2008).

6 Lin, X. Functions of heparan sulfate proteoglycans in cell signaling during development. *Development* 131, 6009-6021, doi:131/24/6009 [pii]
10.1242/dev.01522 (2004).

7 Inatani, M., Irie, F., Plump, A. S., Tessier-Lavigne, M. & Yamaguchi, Y. Mammalian brain morphogenesis and midline axon guidance require heparan sulfate. *Science* 302, 1044-1046, doi:10.1126/science.1090497
302/5647/1044 [pii] (2003).

8 Hwang, H. Y., Olson, S. K., Esko, J. D. & Horvitz, H. R. Caenorhabditis elegans early embryogenesis and vulval morphogenesis require chondroitin biosynthesis. *Nature* 423, 439-443, doi:10.1038/nature01634 nature01634 [pii] (2003).

9 Lowe, J. B. & Marth, J. D. A genetic approach to Mammalian glycan function. *Annu Rev Biochem* 72, 643-691, doi:10.1146/annurev.biochem.72.121801.161809 121801.161809 [pii] (2003).

10 Haltiwanger, R. S. & Lowe, J. B. Role of glycosylation in development. *Annu Rev Biochem* 73, 491-537, doi:10.1146/annurev.biochem.73.011303.074043 (2004).

11 Cipollo, J. F., Awad, A. M., Costello, C. E. & Hirschberg, C. B. N-Glycans of Caenorhabditis elegans are specific to developmental stages. *J Biol Chem* 280, 26063-26072, doi:M503828200 [pii] 10.1074/jbc.M503828200 (2005).

12 Collins, B. E. & Paulson, J. C. Cell surface biology mediated by low affinity multivalent protein-glycan interactions. *Curr Opin Chem Biol* 8, 617-625, doi:S1367-5931(04)00136-X [pii] 10.1016/j.cbpa.2004.10.004 (2004).

13 Crocker, P. R. Siglecs: sialic-acid-binding immunoglobulin-like lectins in cell-cell interactions and signalling. *Curr Opin Struct Biol* 12, 609-615, doi:S0959440X02003755 [pii] (2002).

14 Ishida, H. *et al.* A novel beta1,3-N-acetylglucosaminyltransferase (beta3Gn-T8), which synthesizes poly-N-acetyllactosamine, is dramatically upregulated in colon cancer. *FEBS Lett* 579, 71-78, doi:S0014-5793(04)01427-9 [pii]

10.1016/j.febslet.2004.11.037 (2005).

15 Fuster, M. M., Brown, J. R., Wang, L. & Esko, J. D. A disaccharide precursor of sialyl Lewis X inhibits metastatic potential of tumor cells. *Cancer Res* 63, 2775-2781 (2003).

16 Liu, D., Shriver, Z., Venkataraman, G., El Shabrawi, Y. & Sasisekharan, R. Tumor cell surface heparan sulfate as cryptic promoters or inhibitors of tumor growth and metastasis. *Proc Natl Acad Sci U S A* 99, 568-573, doi:10.1073/pnas.012578299
99/2/568 [pii] (2002).

17 Dube, D. H. & Bertozzi, C. R. Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat Rev Drug Discov* 4, 477-488, doi:nrd1751 [pii] 10.1038/nrd1751 (2005).

18 Iwai, T. *et al.* Core 3 synthase is down-regulated in colon carcinoma and profoundly suppresses the metastatic potential of carcinoma cells. *Proc Natl Acad Sci U S A* 102, 4572-4577, doi:0407983102 [pii]
10.1073/pnas.0407983102 (2005).

19 Sasisekharan, R., Shriver, Z., Venkataraman, G. & Narayanasami, U. Roles of heparan-sulphate glycosaminoglycans in cancer. *Nat Rev Cancer* 2, 521-528, doi:10.1038/nrc842
nrc842 [pii] (2002).

20 Crocker, P. R. Siglecs in innate immunity. *Curr Opin Pharmacol* 5, 431-437, doi:S1471-4892(05)00072-X [pii]
10.1016/j.coph.2005.03.003 (2005).

21 Guo, Y., Mascareno, E. & Siddiqui, M. A. Distinct components of Janus kinase/signal transducer and activator of transcription signaling pathway mediate the regulation of systemic and tissue localized renin-angiotensin system. *Mol Endocrinol* 18, 1033-1041, doi:10.1210/me.2003-0231 me.2003-0231 [pii] (2004).

22 Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A. & Dwek, R. A. Glycosylation and the immune system. *Science* 291, 2370-2376 (2001).

23 Rudd, P. M., Wormald, M. R. & Dwek, R. A. Sugar-mediated ligand-receptor interactions in the immune system. *Trends Biotechnol* 22, 524-530, doi:10.1016/j.tibtech.2004.07.012 S0167-7799(04)00205-7 [pii] (2004).

24 Kinjo, Y. *et al.* Recognition of bacterial glycosphingolipids by natural killer T cells. *Nature* 434, 520-525, doi:nature03407 [pii] 10.1038/nature03407 (2005).

25 Yue, T. *et al.* The prevalence and nature of glycan alterations on specific proteins in pancreatic cancer patients revealed using antibody-lectin sandwich arrays. *Mol Cell Proteomics* 8, 1697-1707, doi:M900135-MCP200 [pii] 10.1074/mcp.M900135-MCP200 (2009).

26 Amano, M. & Nishimura, S. Large-scale glycomics for discovering cancer-associated N-glycans by integrating glycoblotting and mass spectrometry. *Methods Enzymol* 478, 109-125, doi:S0076-6879(10)78004-6 [pii] 10.1016/S0076-6879(10)78004-6 (2010).

27 Aoki, K. *et al.* Dynamic developmental elaboration of N-linked glycan complexity in the Drosophila melanogaster embryo. *J Biol Chem* 282, 9127-9142, doi:M606711200 [pii] 10.1074/jbc.M606711200 (2007).

28 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, doi:M804925200 [pii] 10.1074/jbc.M804925200 (2008).

29 Alvarez-Manilla, G. *et al.* Tools for glycomics: relative quantitation of glycans by isotopic permethylation using 13CH3I. *Glycobiology* 17, 677-687, doi:cwm033 [pii] 10.1093/glycob/cwm033 (2007).

30 Kang, P., Mechref, Y., Kyselova, Z., Goetz, J. A. & Novotny, M. V. Comparative glycomic mapping through quantitative permethylation and stable-isotope labeling. *Anal Chem* 79, 6064-6073, doi:10.1021/ac062098r (2007).

31 Orlando, R. *et al.* IDAWG: Metabolic incorporation of stable isotope labels for quantitative glycomics of cultured cells. *J Proteome Res* 8, 3816-3823, doi:10.1021/pr8010028 (2009).

32 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).

33 Thomson, J. A. *et al.* Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science* 282, 1145-1147, doi:10.1126/science.282.5391.1145 (1998).

34 Wobus, A. M. & Boheler, K. R. Embryonic Stem Cells: Prospects for Developmental Biology and Cell Therapy. *Physiological Reviews* 85, 635-678, doi:10.1152/physrev.00054.2003 (2005).

35 D'Amour, K. A. *et al.* Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat Biotech* 23, 1534-1541, doi:http://www.nature.com/nbt/journal/v23/n12/suppinfo/nbt1163_S1.html (2005).

36 Van Hoof, D., Mendelsohn, A. D., Seerke, R., Desai, T. A. & German, M. S. Differentiation of human embryonic stem cells into pancreatic endoderm in patterned size-controlled clusters. *Stem Cell Research* 6, 276-285, doi:DOI: 10.1016/j.scr.2011.02.004 (2011).

37 Toyoda, M. *et al.* Lectin microarray analysis of pluripotent and multipotent stem cells. *Genes to Cells* 16, 1-11, doi:10.1111/j.1365-2443.2010.01459.x (2011).

38 Satomaa, T. *et al.* The N-glycome of human embryonic stem cells. *BMC Cell Biol* 10, 42, doi:1471-2121-10-42 [pii] 10.1186/1471-2121-10-42 (2009).

39 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, doi:10.1634/stemcells.2006-0219 (2007).

40 Fang, M., Lim, J.-M. & Wells, L. *Quantitative Glycomics of Cultured Cells Using Isotopic Detection of Aminosugars with Glutamine (IDAWG)*. (John Wiley & Sons, Inc., 2009).

41 Ceroni, A. *et al.* GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res* 7, 1650-1659, doi:10.1021/pr7008252 (2008).

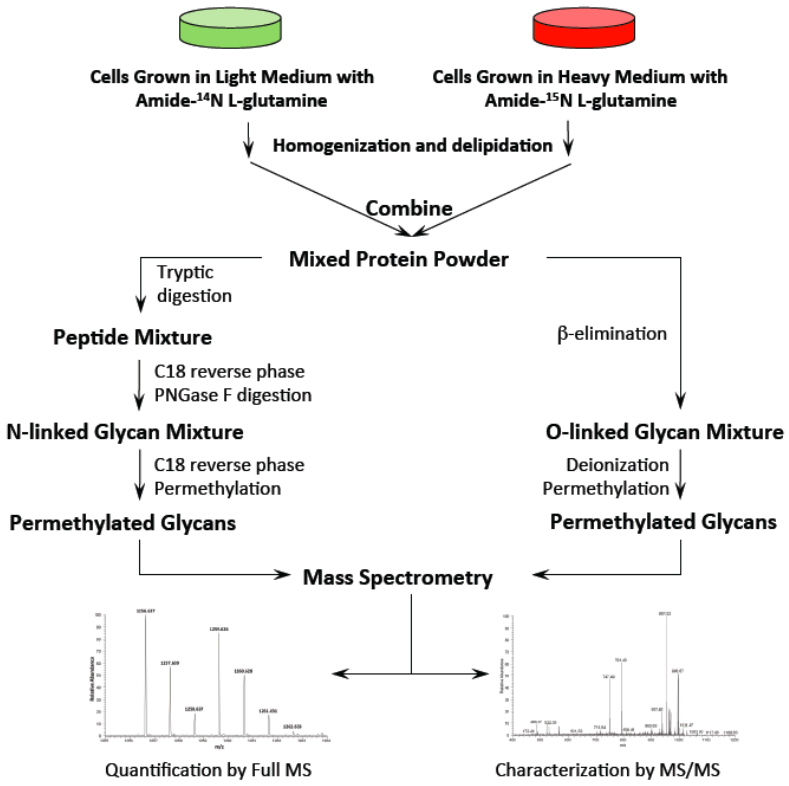


Figure 3.1: Workflow of IDAWG labeling strategy for comparative glycomics of hESC and hDE. Amide-^{15}N-Gln or Gln is added to the media to label cultured cells. Upon homogenization and delipidation, protein powders can be combined, then N-linked and O-linked glycans isolated from mixed protein powders for permethylation and tandem mass spectrometry analysis.

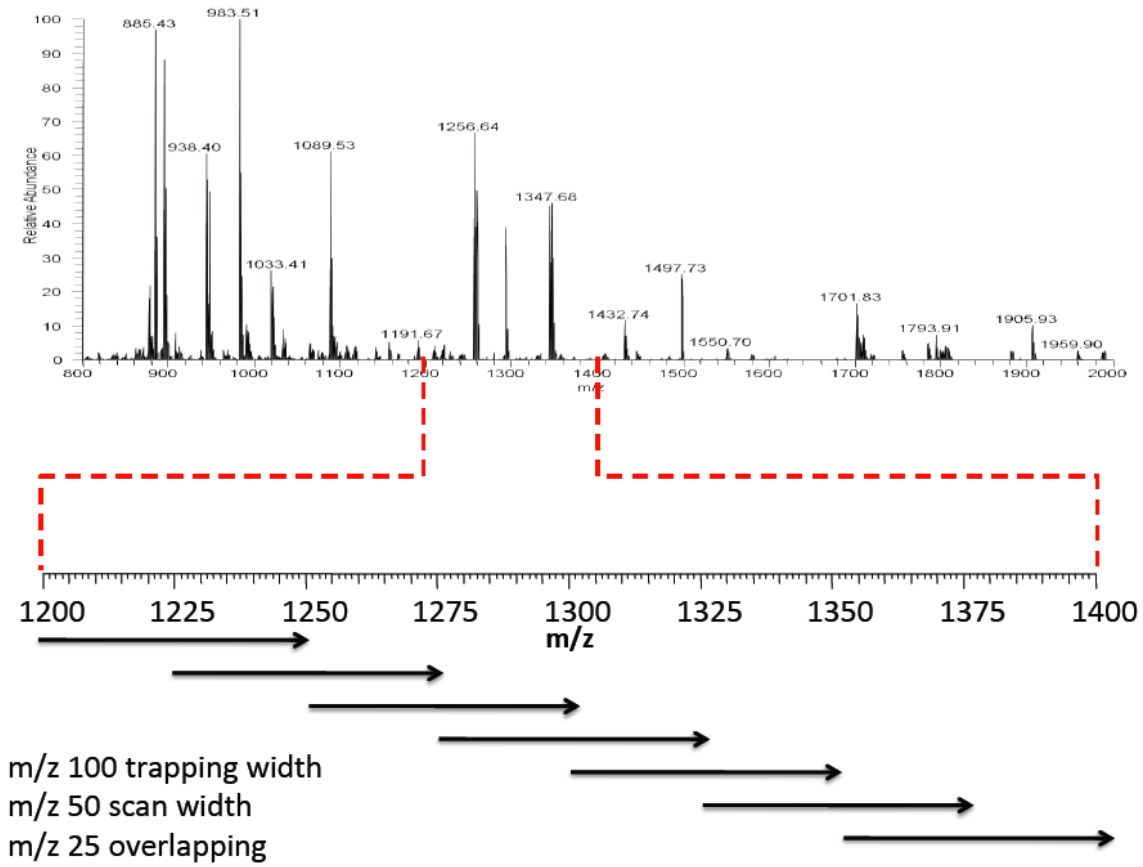


Figure 3.2: Mass Spectrometry Method: Rolling trapping scans (without collision energy). Automated MS/MS spectra without applying collision energy are acquired in this mode. For each MS/MS scan, the instrument traps the ions with a width of 100 m/z but only scans 50 m/z around the center of the trapping window to ensure the trapping efficiency. The trapping center of each scan is moved automatically by 25 m/z to the higher mass region compared to its previous scan so that the range of a typical full MS scan (shown in the upper side) can be covered by these stepwise trapping scans.

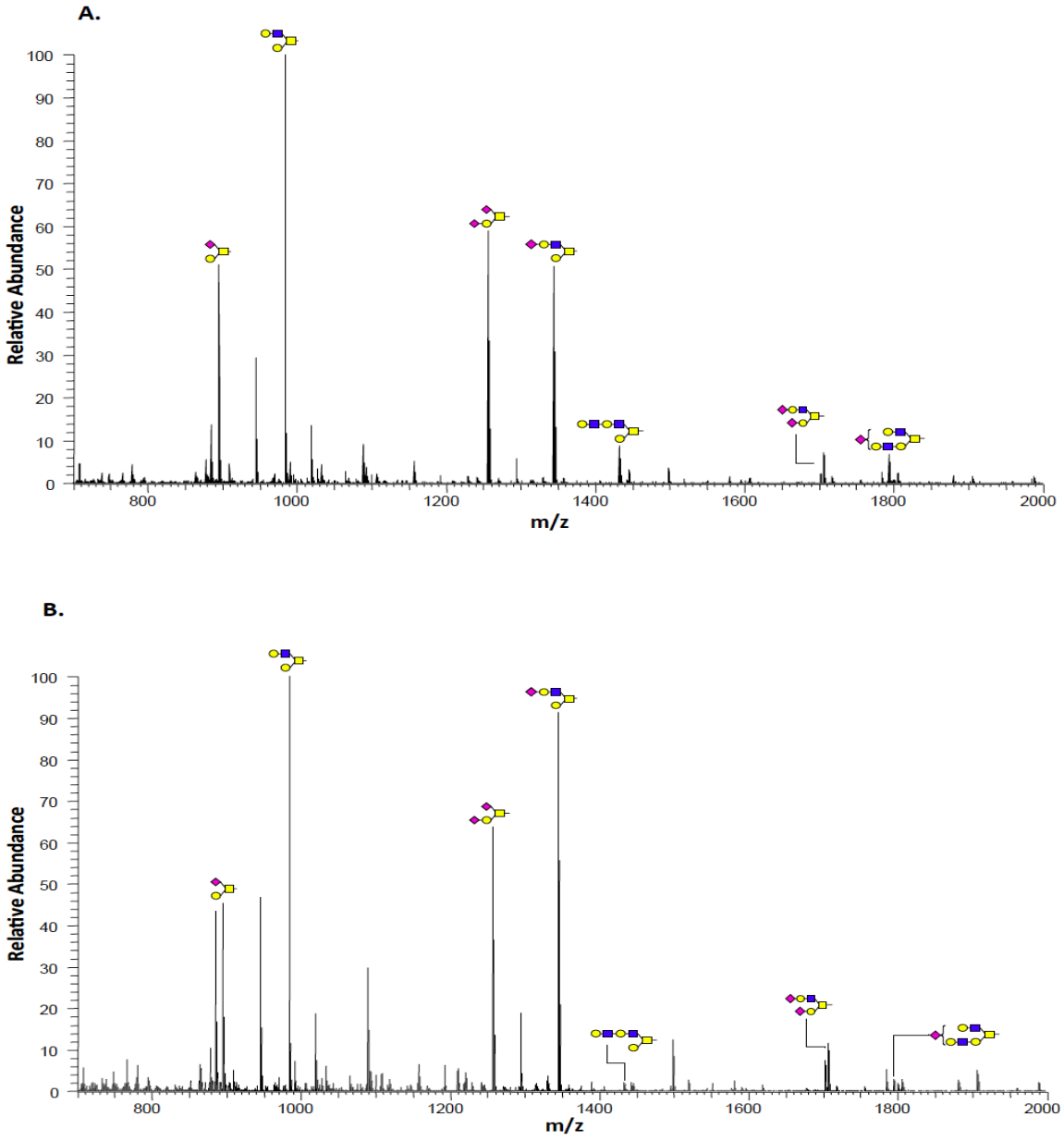


Figure 3.3: Detection and identification of O-linked glycans by LTQ-Orbitrap. Full spectra from 700-2000 m/z of the permethylated O-linked glycans released from hESCs (**A**) and hDEs (**B**). Major glycans identified by total ion mapping and MSn are shown.

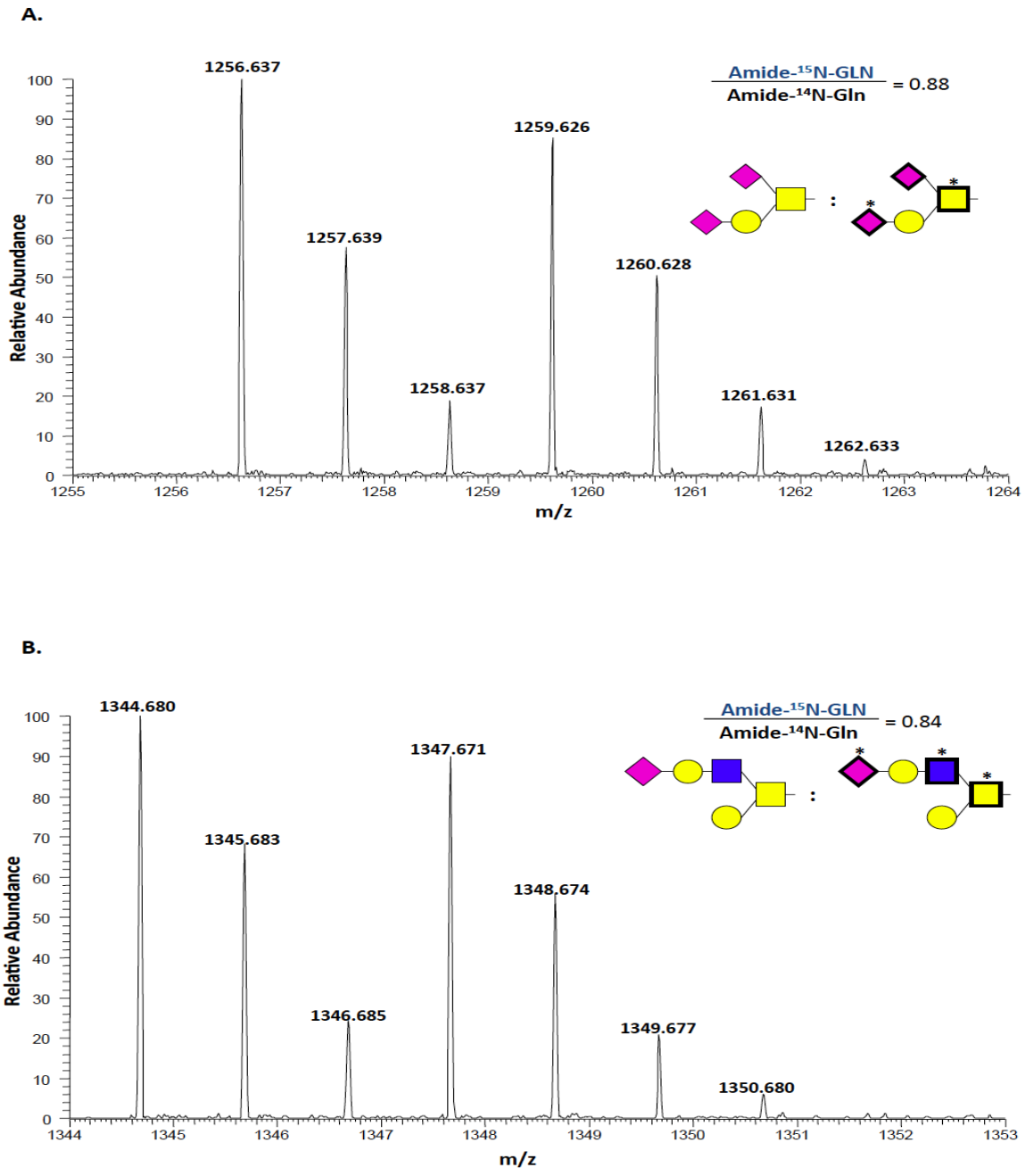


Figure 3.4: Quantitative analysis of O-linked glycans. A mixture of O-linked glycans from a 1:1 mixture of light/heavy labeled hESCs was analyzed on an LTQ-Orbitrap. **(A)** Quantification of the disialyated T-antigen. **(B)** Quantification of a monosialyated core 2 structure.

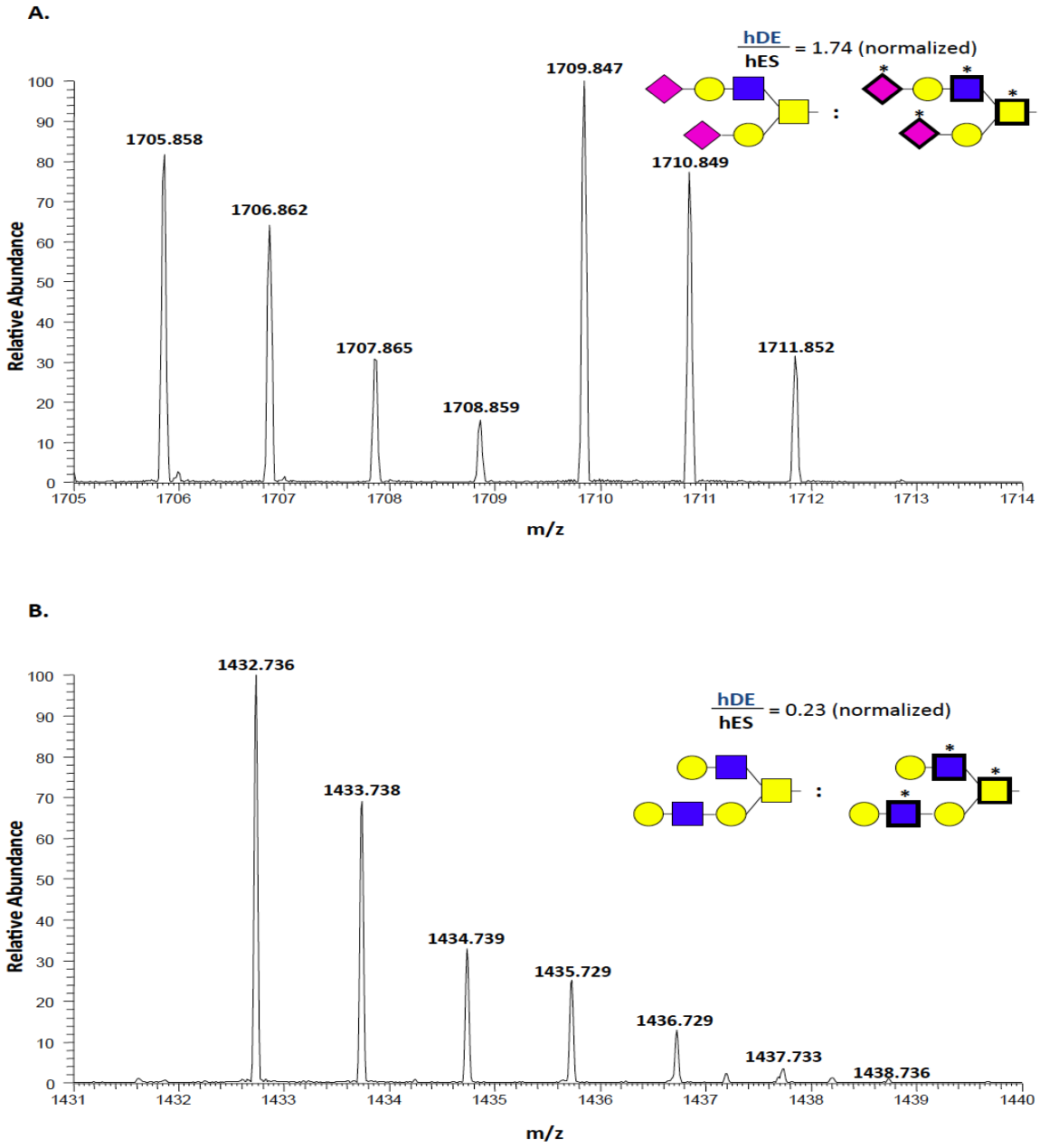


Figure 3.5: Quantitative analysis of O-linked glycans. A mixture of O-linked glycans from a 1:1 mixture of light/heavy differentially labeled hESCs/hDEs was analyzed on an LTQ-Orbitrap. (A) Quantification of a disialyated core 2 structure. **(B)** Quantification of a poly-LacNAc core 2 structure. The ratios in (A) and (B) are already normalized.

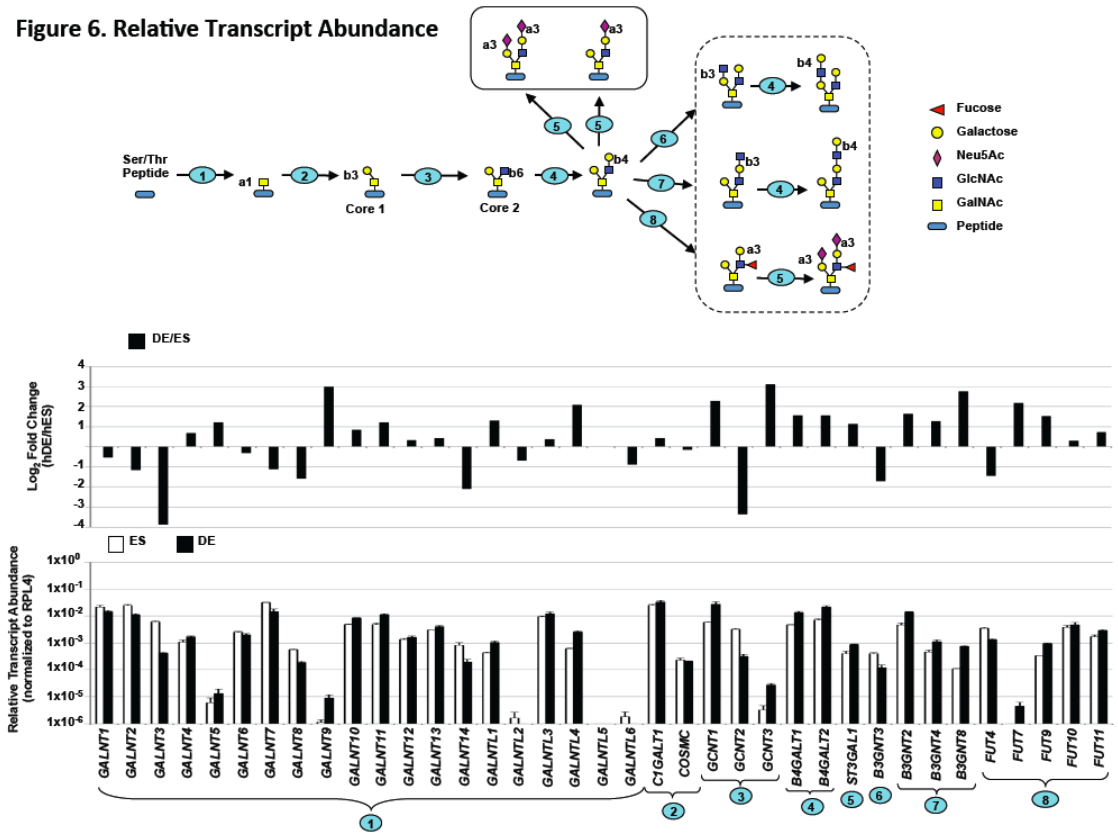


Figure 3.6: Relative Transcript Abundance of glycosyltransferases in hESCs and hDEs. The q-PCR result shows an increase in the abundance of the glycosyltransferases which adds the sialic acid (No.5) and a decrease in the abundance of the glycosyltransferases which adds the LacNAc (No.6) during the differentiation from hESCs to hDEs, which agrees with the results from the IDAWG quantification.

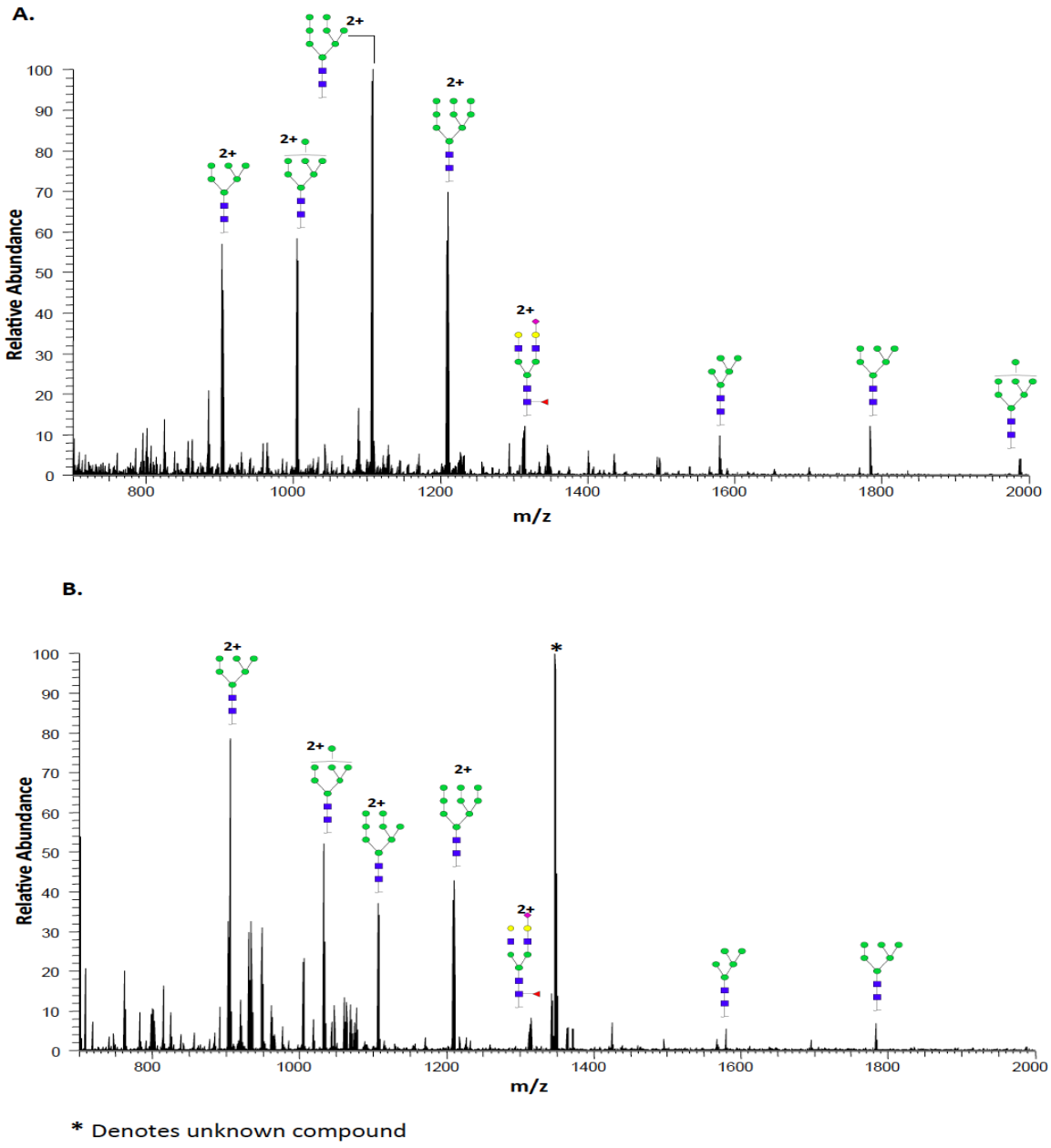


Figure 3.S1: Detection and identification of N-linked glycans by LTQ-Orbitrap. Full spectra from 700-2000 m/z of the permethylated O-linked glycans released from hESCs (**A**) and hDEs (**B**). Major glycans identified by total ion mapping and MSn are shown along with charge state (Z) if it is not singly charged.

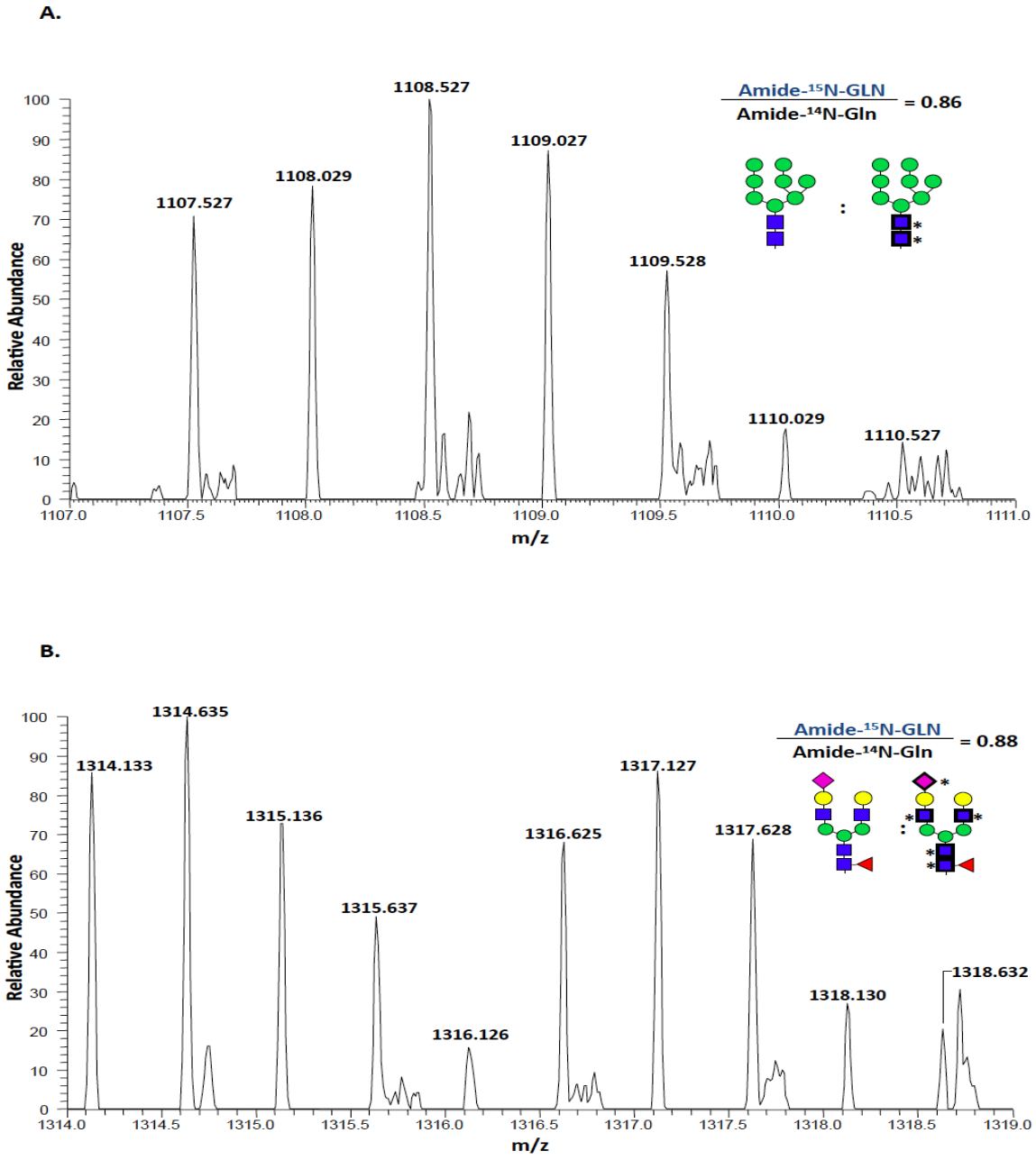


Figure 3.S2: Quantitative analysis of N-linked glycans. A mixture of N-linked glycans from a 1:1 mixture of light/heavy labeled hESCs was analyzed on an LTQ-Orbitrap. (A) Quantification of the high mannose structure. **(B)** Quantification of a monosialyated complex structure.

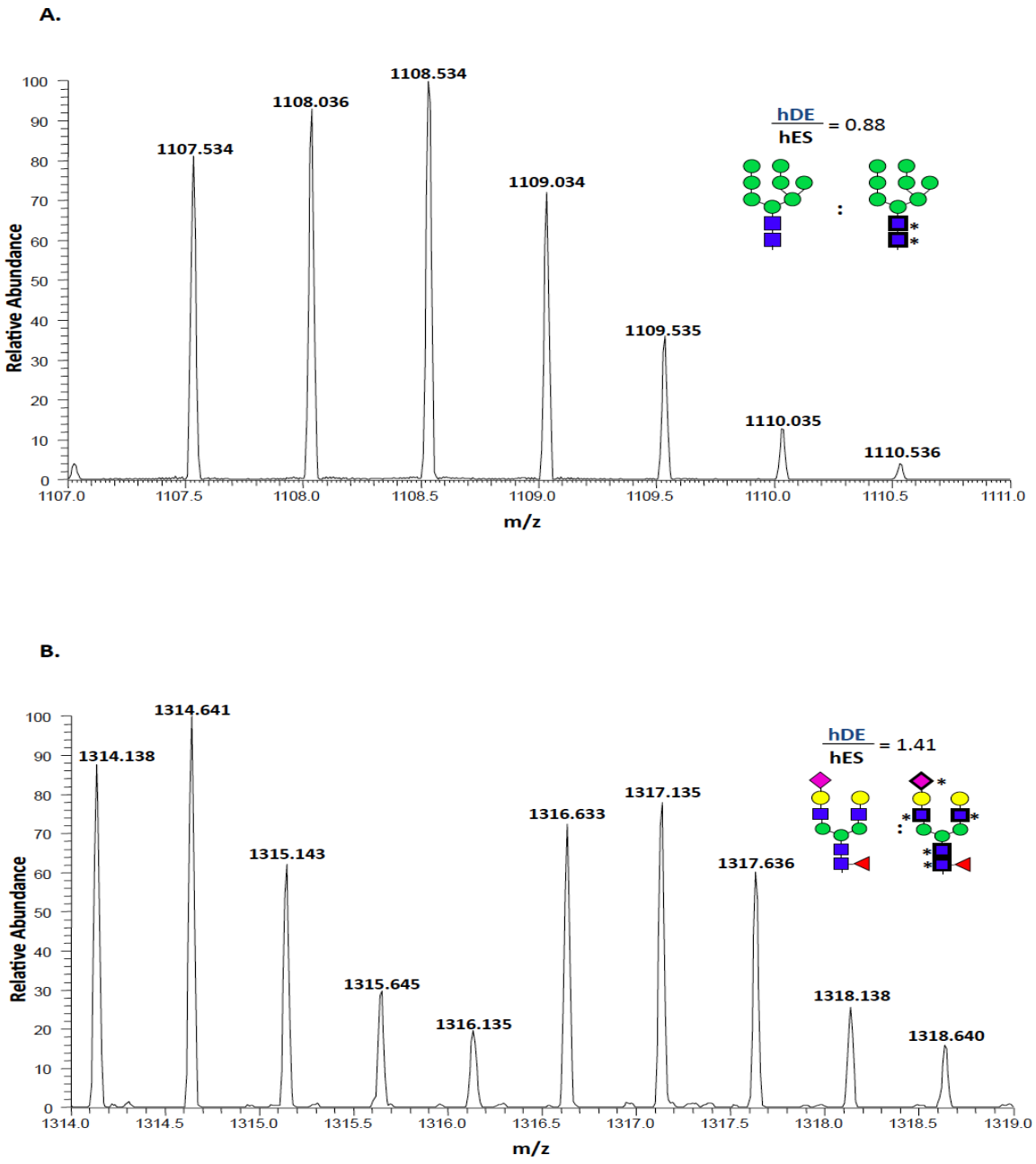


Figure 3.S3: Quantitative analysis of N-linked glycans. A mixture of N-linked glycans from a 1:1 mixture of light/heavy differentially labeled hESCs/hDEs on an LTQ-Orbitrap. (A) Quantification of the high mannose structure. **(B)** Quantification of a monosialyated complex structure.

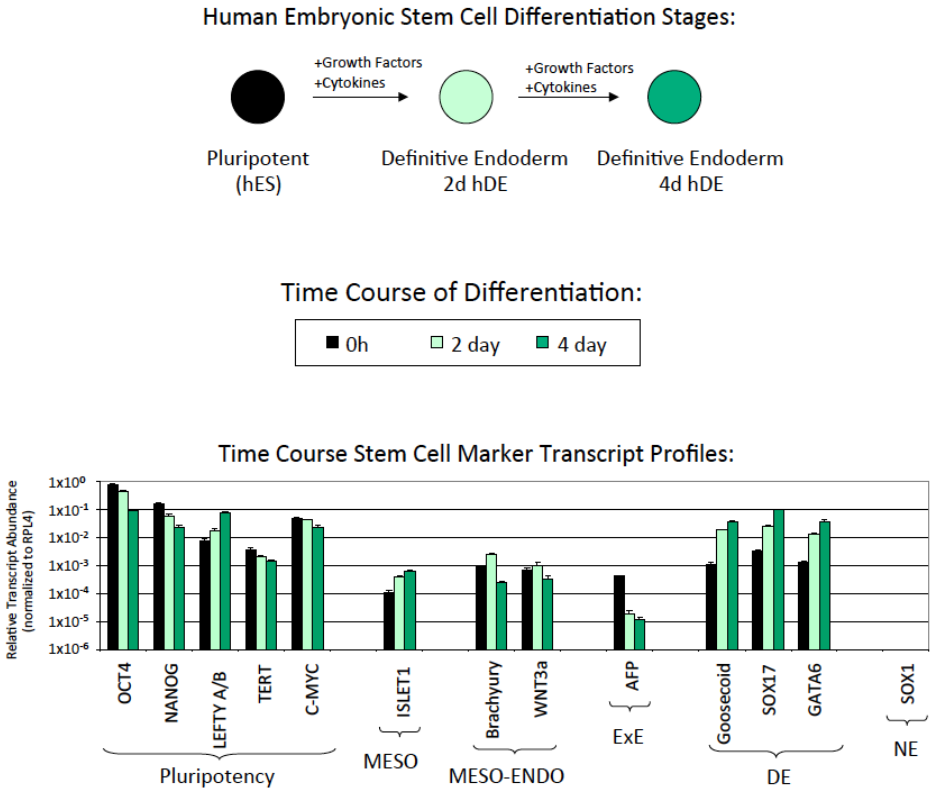


Figure 3.S4: Relative Transcript Abundance.

Table 3.1: Identification and prevalence of O-glycans released from hESCs. 24 O-glycan structures released from hESCs and hDEs are identified and they are listed in a decreasing order of their prevalence in hESCs O-glycan pool.

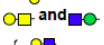
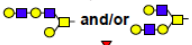
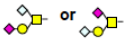
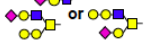
No.	Structure	M/Z	Prevalence in hESCs
1		983.5149	21.78%
2		1256.6361	20.22%
3		1344.6886	17.60%
4		895.4624	8.20%
5		534.2887	6.76%
6		1793.9148	5.81%
7		1705.8624	4.93%
8		1432.7411	4.70%
9		1879.9516	2.12%
10		1157.6041	1.99%
11		779.4151	1.03%
12		1069.5516	0.77%
13		1518.7779	0.74%
14		1099.5622	0.66%
15		1286.6467	0.56%
16		1617.8099	0.43%
17		738.3885	0.39%
18		1187.6147	0.32%
19		1140.5888	0.29%
20		912.4777	0.20%
21		925.4730	0.18%
22		1909.9622	0.15%
23		1548.7884	0.10%
24		609.3095	0.04%

Table 3.2: Quantifying O-glycans released from 1:1 mixture of light/heavy labeled hESCs.

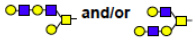
Structure	m/z	Prevalence	Average Ratio	SD	Error
	983.52	21.78%	0.89	0.11	12.58%
	1256.64	20.22%	1.03	0.07	7.23%
	1344.69	17.60%	0.97	0.06	6.35%
	895.46	8.20%	0.85	0.01	0.76%
	534.29	6.76%	1.32	0.07	4.97%
	1793.91	5.81%	1.13	0.05	4.32%
	1705.86	4.93%	0.82	0.03	4.18%
	1432.74	4.70%	1.09	0.02	1.47%
	1879.95	2.12%	1.36	0.26	18.78%
	1157.60	1.99%	0.88	0.13	15.25%
	779.42	1.03%	0.65	0.05	7.83%
			Average	0.08	7.61%
		SD for all structures		0.21	

Table 3.3: Quantifying O-glycans released from 1:1 mixture of light/heavy differentially labeled hESCs/hDEs

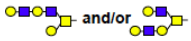
Structure	m/z	Prevalence	Ratio	Error
	983.52	21.78%	1.11	20.20%
	1256.64	20.22%	1.09	14.42%
	1344.69	17.60%	2.64	30.21%
	895.46	8.20%	0.87	13.56%
	534.29	6.76%	1.04	17.20%
	1793.91	5.81%	0.64	61.95%
	1705.86	4.93%	2.54	29.48%
	1432.74	4.70%	0.41	37.70%
	1879.95	2.12%	1.75	49.39%
	1157.60	1.99%	0.99	29.78%
	779.42	1.03%	0.79	26.42%

Table 3.S1: Identification and prevalence of N-glycans released from hESCs.

No.	Structure	Z	M/Z	Prevalence in hES
	High Man			
1	M8N2-Me	2	1107.536	22.69%
2	M9N2-Me	2	1209.586	19.36%
3	M7N2-Me	1	1987.983	13.86%
		2	1005.486	
4	M6N2-Me	1	1783.883	6.19%
5	M5N2-Me	1	1579.783	5.17%
		2	801.386	
6	Glc-M9N2-Me	2	1311.636	3.14%
7	M3N2-Me	1	1171.583	0.77%
8	M4N2-Me	1	1375.683	0.42%
9	M2N2-Me	1	967.484	0.12%
10	Glc3-M9N2-Me	2	1515.736	0.05%
11	Glc2-M9N2-Me	2	1413.686	0.02%
	Hybrid			
1	N2M5N2-Me or Gal2N2M3N2	2	1046.513	0.58%
2	GalNM3M3N2-Me	2	1128.049	0.35%
3	NM3N2-Me	1	1416.710	0.21%
4	GalNM5N2-Me	2	1026.000	0.13%
5	SAGalNM3N2-Me	1	1981.983	0.11%
		2	1002.487	
6	GalNM3N2-Me or NM4N2-Me	1	1620.810	0.06%
7				
8	NM5N2-Me	1	1824.909	0.02%
9	GalN2M3N2-Me or N2M4N2-Me	1	1865.936	<0.01%
10	GalNM4M3N2-Me	2	1230.099	<0.01%
	Complex			
	No Fuc, No SA			
1	GalN3M3N2-Me	2	1067.026	0.34%
2	N4M3N2-Me	2	1087.539	0.24%
3	N2M3N2-Me	1	1661.836	0.22%
		2	842.413	
4	Gal3N3M3N2-Me	2	1271.126	0.16%
5	GalN4M3N2-Me	2	1189.589	0.07%
6	Gal3N2M3N2-Me	2	1148.563	0.03%
7	N3M3N2-Me	1	1906.963	0.01%
8	Gal2N3M3N2-Me	2	1169.076	<0.01%
9	Gal4N2M3N2-Me	2	1250.613	<0.01%
	Fuc, No SA			
1	M3N2F-Me	1	1345.673	1.57%
2	N2M3N2F-Me	1	1835.925	1.31%
		2	929.458	
3	Gal2N3M3N2F-Me or Gal2N3FM3N2-Me	2	1256.121	0.97%
4	Gal2N3FM3N2F-Me	2	1343.165	0.52%
5	NM3N2F-Me	1	1590.799	0.51%
		2	806.894	
6	N3M3N2F-Me	2	1052.021	0.27%
7	Gal2N2M3N2F-Me	2	1133.557	0.22%
8	M2N2F-Me	1	1141.573	0.20%
9	GalN2M3N2F-Me	2	1031.508	0.18%
10	Gal2N3F2M3N2F-Me	2	1430.210	0.18%
11	Gal2N2F2M3N2F-Me	2	1307.647	0.09%
12	GalN3M3N2F-Me	2	1154.071	0.07%
13	Gal2N3FM3N2F-Me or Gal2N3F2M3N2-Me	2	1343.165	0.06%
14				
15	GalNM3N2F-Me	1	1794.899	0.04%
		2	908.944	
16	Gal3N2FM3N2-Me	2	1322.652	0.04%
17	M4N2F-Me	1	1549.772	0.03%
18	N4M3N2F-Me	2	1174.584	0.02%
19	Gal3N2M3N2F-Me	2	1235.607	0.02%
20	Gal3N3M3N2F-Me	2	1358.171	0.01%
21	Gal3N3FM3N2F-Me	2	1445.215	0.01%
22	Gal3N4M3N2F-Me	2	1480.734	0.01%
23	Gal4N3FM3N2-Me or Gal4N3M3N2F-Me	2	1460.220	<0.01%

No.	Structure	Z	M/Z	Prevalence in hESCs
	Complex			
	No Fuc, SA			
1	SAGal2N2M3N2-Me	2	1227.100	1.78%
2	SA2Gal2N2M3N2-Me	2	1407.687	0.57%
3	SA3Gal3N3M3N2-Me	2	1812.887	0.19%
		3	1216.254	
4	SAGal3N2M3N2-Me	2	1329.150	0.18%
5	SAGal2N3M3N2-Me	2	1349.663	0.18%
6	SA2Gal2N3M3N2-Me	2	1530.250	0.03%
		3	1027.830	
7	SAGalN2M3N2-Me	2	1125.050	0.03%
8	SA2Gal3N3M3N2-Me	2	1632.300	0.03%
9		3	1095.863	
10	SA4Gal3N3M3N2-Me	3	1394.675	0.03%
11	SA2Gal5N4M3N2-Me	2	1958.963	0.02%
		3	1313.638	
12	SAGalN3M3N2-Me	2	1247.613	0.02%
13	SAGal3N4M3N2-Me	2	1574.276	0.01%
14	SAGalN4M3N2-Me	2	1370.176	0.01%
15	SA2Gal4N4M3N2-Me	2	1856.913	<0.01%
		3	1245.605	
16	SAGal3N3M3N2-Me	2	1451.713	<0.01%
17	SA3Gal3N4M3N2-Me	3	1297.963	<0.01%
	Fuc, SA			
1	SAGal2N2M3N2F-Me	2	1314.144	4.88%
		3	883.759	
2	SA2Gal2N2M3N2F-Me	2	1494.731	4.21%
		3	1004.151	
3	SAGal2N2FM3N2F-Me	2	1401.189	2.22%
4	SAGal2N3M3N2F-Me	2	1436.708	1.68%
5	SAGal3N3M3N2F-Me	2	1538.757	1.41%
		3	1033.502	
6	SA2Gal3N3M3N2F-Me or SA2Gal3N3FM3N2-Me	2	1719.344	0.61%
7		3	1153.893	
8	SAGalN3M3N2F-Me	2	1334.658	0.45%
9	SA2Gal2N3M3N2F-Me	2	1617.294	0.34%
		3	1085.860	
10	SAGal2N3FM3N2F-Me	2	1524.252	0.22%
11	SA2Gal4N4M3N2F-Me	2	1943.957	0.18%
		3	1303.635	
12	SA3Gal3N3FM3N2F-Me	2	1986.976	0.05%
13	SAGal3N2FM3N2F-Me	2	1503.239	0.04%
14	SA3Gal3N3M3N2F-Me	3	1274.284	<0.01%
15	SA3Gal3N4M3N2F-Me	3	1355.993	<0.01%
16	SAGalN2M3N2F-Me	2	1212.094	NA

Table 3.S2: Quantifying N-glycans released from 1:1 mixture of light/heavy labeled hESCs.

Structure	m/z	Z	Prevalence	Average Ratio	SD	Error
	1107.54	2	22.69%	1.04	0.19	18.22%
	1209.59	2	19.36%	0.92	0.08	8.76%
	1005.49	2	11.25%	0.62	0.28	44.75%
	1783.88	1	6.19%	1.00	0.15	15.02%
	1314.14	2	4.29%	0.81	0.04	4.68%
	1579.78	1	3.81%	0.93	0.12	13.16%
	1494.73	2	3.22%	1.04	0.20	19.59%
	1987.98	1	2.61%	0.95	0.07	7.62%
	1401.19	2	2.22%	1.00	0.13	12.88%
	1227.10	2	1.78%	0.80	0.29	35.97%
	1436.71	2	1.68%	1.04	0.13	13.00%
	1345.67	1	1.57%	0.86	0.12	13.80%
	801.39	2	1.36%	0.88	0.39	44.49%
				Average	0.17	18.72%
			SD for all structures		0.13	

Table 3.S3: Quantifying N-glycans released from 1:1 mixture of light/heavy differentially labeled hESCs/hDEs. The ratios of 5 most abundant structures from each category in Supplemental Table 1 are weighted by their prevalence and averaged to demonstrate the trend of change for each type of N-glycan structures

Structure Type		Average Ratio by weight
High Mannose		0.91
Hybrid		1.54
Complex	No Fuc, No SA	0.89
	Fuc, No SA	1.04
	No Fuc, SA	1.18
	Fuc, SA	1.09

CHAPTER 4

ASSESSING THE DYNAMICS OF INDIVIDUAL GLYCANS RELEASED FROM HUMAN ES CELLS USING THE IDAWG TECHNIQUE REVEALS THE REMODELING OF SIALYATED GLYCAN STRUCTURES[1]

[1] Meng Fang, Michael Kulik, Jun Han, Sandii Brimble, Jae-Min Lim, Stephen Dalton, John Miller, Will York and Lance Wells
To be submitted to *Journal of Proteome Research.*

ABSTRACT

IDAWG, as the first reported in-vivo isotopic labeling strategy for comparative glycomics, 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. Previous glycan turnover studies that required incorporation of radioactive monosaccharide and extensive subsequent fractionation were short of the resolution to investigate individual glycan structures or glycan subsets. In our study, by performing the Pulse-Chase experiments with IDAWG labeling strategy during the cell culture of human embryonic stem cells, we are able to assess the dynamics of turnover for individual glycans. More interestingly, by comparing the glycan spectra of pulse-chase samples with the spectra of real mixtures of light and heavy sample, we have noticed an isotope mixture that could only be explained by remodeling of the glycans and are able to approximate the amount of remodeling occurring for a subset of structures, predominantly those containing sialic acid residues.

INTRODUCTION

In 2009, the methodology of IDAWG, Isotopic Detection of Aminosugars With Glutamine, was reported as the first in-vivo stable isotopic labeling strategies for quantitative glycomics[1]. The IDAWG takes advantage of the hexosamine biosynthetic pathway that uses the side chain of glutamine as its sole donor source of nitrogen for aminosugars in the production of sugar nucleotides. As a result, if the cells are fed with Gln-free media and glutamine with a ^{15}N-labeled side chain, all the aminosugars including GlcNAC, GalNAc and sialic acids produced in cells will be labeled with ^{15}N and thus the mass of N- and O-linked glycans, glycolipids, and extracellular matrix polysaccharides should all be shifted by +1 Da per aminosugar. The IDAWG technology has been used to analysis both N- and O-linked glycans released from murine embryonic stem cells successfully and it is also predicted to be useful to various comparative glycomic studies in future. As the first reported in-vivo isotopic labeling strategy for comparative glycomics, IDAWG not only shares some significant advantages of SILAC for proteomics[2-4], but the metabolic labeling of glycans using IDAWG 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.

Most glycoconjugates are degraded in lysosomes and some the liberated monosaccharides can be reused for glycoconjugate synthesis[5,6]. The degradation of glycans is quite ordered and involves highly specific glycosidases. Although studies of a series of human genetic disorders called lysosomal storage diseases have lead to the understanding of these complex pathways of degradations of different glycoconjugates such as N-[7,8] and O-linked[9] oligosaccharides chains and glycosaminoglycans, however,

the assessment of the dynamics of turnover and synthesis of individual glycans has been barely touched.

Previously, glycan turnover studies required incorporation of radioactive monosaccharide and extensive subsequent fractionation to provide specific information about the structure, sequence, and distribution of the sugar chains of glycoconjugates[10,11]. Generally, the radiolabeling techniques were capable of a very sensitive detection of glycan classes, but short of the resolution to investigate individual glycan structures or glycan subsets. With the methodology of IDAWG, the synthesis and degradation of any aminosugar-containing glycans can be evaluated by completely labeling cells with heavy glutamine and then replacing the media supplement with light glutamine. The implementation of Pulse-Chase experiments using IDAWG, with the analytic advantages of high-resolution mass spectrometry, will move forward in our understanding of the dynamics of glycan turnover with a sharp focus on individual glycans or glycan subsets. Herein, we performed the Pulse-Chase experiments with IDAWG labeling strategy during the cell culture of human embryonic stem cells. By analyzing permethylated glycans released from human embryonic stem cells at different points of time using high resolution LTQ-Orbitrap mass spectrometry, we are able to assess the dynamics of turnover for individual glycans. More interestingly, by comparing the glycan spectra of pulse-chase samples with the spectra of real mixtures of light and heavy sample, we noticed an isotope mixture that could only be explained by remodeling of the glycans and were able to approximate the amount of remodeling occurring for a subset of structures, predominantly those containing sialic acid residues.

EXPERIMENTAL PROCEDURES

Cell Culture with Pulse-Chase experiments

The BG02 line of human embryonic stem cells (hESCs) were cultured essentially as previously described[12]. Briefly, the hESCs were grown on gelatin coated petri dishes at 37^{o}C 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 the cells were allowed to grow for 72 hours at which point part of them were harvested by dissociation buffer and scraping. Then the other cells were transferred to light media supplemented with normal abundance amide-^{14}N-Gln and allowed 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 ^{o}C until analysis.

Preparation of Protein Powder, Delipidation, Glycan Release and Permethylation

The isolation of permethylated N- and O-linked glycans was essentially as previously described[1,13-15]. 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 either digested by trypsin and subjected to PNGase F treatment to release N-linked glycans or treated with alkaline borohydride to release O-linked glycans by β-elimination. Released glycans were permethylated with methyl iodide before they were cleaned up by reverse-phase columns and dried down.

MS Analysis of the Permethylated Glycans

The glycans were analyzed as previously described on a hybrid linear ion trap Orbitrap mass spectrometer (LTQ-Orbitrap, Thermo)[1,13]. 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 (TIM). 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 glycoworkbench software[16]. The IDAWG quantification data was mainly deconvoluted and analyzed by GlycoQuant.

RESULTS AND DISCUSSION

Pulse-Chase experiment designed for dynamic IDAWG

Based on the Hexosamine biosynthetic pathway, IDAWG takes advantage of heavy glutamine in media to label aminosugars with ^{15}N in cell culture.[1] To apply the IDAWG technique to assess the dynamics of turnover and synthesis for individual glycans, a Pulse-Chase experiment using human embryonic stem cells was performed to provide a dynamic view of hESC glycomes. We started the Pulse-Chase experiment with feeding the cells in media supplement with amide-^{15}N-Gln until the cells were close to completely labeled. 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 shown). The time when the cells reached complete labeling was set as 0 hour and a portion of cells were harvested at that time. The other cells were washed for several times and transferred to new media with normal glutamine. While keeping the cells in the light media, we harvested a portion of them at certain points of time counting from 0 hour. In our initial experiment, we harvested at 0, 12, 24, and 36 hour (**Figure 4.1**); to better assess the dynamics of glycan degradation and synthesis, we repeated the whole experiment with the addition of a harvesting point at 6 hour. Both N- and O-linked glycans were released from the cells harvested at different points of time, permethylated and subjected to the analysis by Thermo ESI-LTQ-Orbitrap mass spectrometer.

Synthesis of light structure and degradation of heavy structure

By completely labeling the human embryonic stem cells with heavy Gln and then replacing the media supplement with light Gln, a series of spectra can be acquired by running the glycans released from cells harvested at different time point on Orbitrap,

through which the synthesis and degradation of any amino-sugar-containing glycans can be evaluated. The 5 trapping scan spectra of the disialyated T-antigen in a 9 m/z window demonstrating its isotopic patterns at different points of time (0 hour, 6 hour, 12 hour, 24 hour and 36 hour) are shown (**Figure 4.2**). All the spectra except for 0 hour show both monoisotopic peaks for light and heavy structures. During the time course, by comparing the relative intensities of isotopic peaks, the trend of increasing "light peaks" and decreasing "heavy peaks" is very clear, which suggests that after the heavy media is replaced, while the glycans containing the heavy aminosugars are degraded, the cells are building new glycans using light Gln from the media. The same series of spectra has been acquired for a non-sialyated core2 structure, which also shows a similar trend of changes for both light and heavy structures as the T-antigen (**Figure 4.3**). This result demonstrates that the metabolic labeling of glycans using IDAWG is able to provide new opportunities for assessing the dynamics of turnover and synthesis, even for individual glycans as shown in our results.

Remodeling of sialyated structures

The mass spectrometry analysis of glycan samples of our Pulse-Chase experiments has demonstrated both the synthesis and degradation of glycan structures occur during the same time course; however, this is not the whole story. The metabolic labeling of glycans using IDAWG has provided more interesting information as we compare the mass spectra of glycan samples from Pulse-Chase experiments with another set of glycan sample, which is a true mixture of equal amount of light and heavy glycans released from differentially labeled human embryonic stem cells. If only the degradation of heavy glycans and synthesis of light glycans occurred in the cells, the glycan samples

from Pulse-Chase experiments should also be a mixture of light and heavy glycan structures and the mass spectra should present a similar isotopic pattern as the true mixture samples. However, the mass spectrum of the disialyated T-antigen in the Pulse-Chase 12 hour glycan sample presents a very different isotopic pattern from the spectrum of the same structure in another sample that is a true mixture of light and heavy labeled glycans (**Figure 4.4**). In the spectra, the m/z range from 1256 to 1259 represents the isotopic peaks of light structure and in the bottom spectrum, the isotopic pattern is abnormal, especially the first two peaks, compared to the top spectrum. This difference indicates that not like a true mixture of light and heavy glycans, the glycan samples from the Pulse-Chase experiment could contain other types of structures that induce the abnormal isotopic pattern. However, this difference in isotopic patterns is not so obvious for non-sialyated structures (**Figure 4.5**) as for the disialyated T-antigen. Since sialic acid can be removed at some point in the life cycle of the molecule once attached to glycoconjugates[17-19], one possible explanation is that other than the glycan synthesis and degradation, glycans can be detached from their underlying core proteins and returned to the Golgi apparatus, then the terminal residues especially the sialic acids can be replaced with new residues. After this remodeling process, the glycans will be put back on their core proteins. As a result of the remodeling, in the glycan pool released from our Pulse-Chase samples, besides the light and heavy structures, there also exist partially heavy structures which make the isotopic pattern in the mass spectra different from what we observe in the spectra of true mixture of light and heavy structures.

Automated modeling for the glycan degradation rate and the proportion of remodeling by the quantitative software

The comparisons of mass spectra generated for glycans released from human embryonic stem cells harvested at different points of time in the Pulse-Chase experiments depict qualitatively the behaviors of glycans in living cells including synthesis, degradation and even remodeling, however, in order to quantify the rates of glycan synthesis and degradation and how much remodeling of an individual glycan, we have to rely on a quantitative software, GlycoQuant, which is developed as a tool to process and analyze the data from the IDAWG experiment. Since the synthesis and degradation along with the remodeling of glycans occur simultaneously in the living cells, in each glycan pool released from cells harvested at different time points, there could be more than 2 components, each of which can be non-labeled, completely labeled or partially labeled with ^{15}N, and the percentage of each component can be varied at different time points. Take the disialyated T-antigen as an example, there are three aminosugars in its structure: two sialic acids and one GalNAc. Then the T-antigen can be fully labeled with 3 ^{15}N or not labeled at all; also it can contain 1 or 2 ^{15}N due to the remodeling. GlycoQuant is used to automatically fit each mass spectrum, analyze the composition and give the percentage of each component at each time point. With considering of the cell doubling time, which is set to be 30 hours for human embryonic stem cell in this project, the software can draw the variation trend line of all components for a certain glycan (**Figure 4.6**). The figure of the T-antigen shows a considerable amount of component with only one ^{15}N that is produced by replacing two terminal sialic acids, which provides the direct evidence of the existence of glycan remodeling and explains the abnormal isotopic

pattern in certain mass spectra. To demonstrate the rates of glycan turnover and synthesis as well as the different extent of remodeling between sialyated and unsialyated structures, we have examined 9 O-linked and 15 N-linked glycans for two replicates of samples. The time for each glycan when the amount of structure containing only heavy nitrogens has decreased to 50% of its original amount at 0 hour is called 50% degradation time and listed in **Table 4.1** and **Table 4.2.** The 50% degradation time demonstrates the rate of degradation for each glycan. For high-mannose N-glycans, this value is inversely proportional to the number of hexoses, especially for those with more than four mannoses (**Figure 4.7**), which means the bigger structures degrade faster than small structures in this case. In **Table 4.1** and **Table 4.**2, the proportion of remodeling structure at the 50% degradation time for each glycan is also listed. Apparently, no matter N-linked or O-linked glycans, sialyated structures are undergoing a relatively higher extent of remodeling compared to unsialyated ones and this also suggests sialic acid has much to do with the phenomenon of remodeling.

CONCLUSION

Herein, the Pulse-Chase experiments with IDAWG labeling strategy during the cell culture of human embryonic stem cells were performed to assess the dynamics of turnover for individual glycans. By fitting the mass spectra acquired for a series of samples using bioinformatics software we were able to investigate the degradation rates for major N- and O-linked glycans. For high-mannose N-glycans, there exists a negative correlation between 50% degradation time and the number of hexoses. By comparing the glycan spectra of pulse-chase samples with the spectra of real mixtures of light and heavy

sample, we noticed an isotope mixture that could only be explained by remodeling of the glycans. We were able to approximate the amount of remodeling occurring for a subset of structures using bioinformatics tools. Sialyated structures are undergoing a relatively higher extent of remodeling compared to unsialyated ones, which suggests sialic acid residues have much to do with the phenomenon of remodeling.

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, doi:10.1021/pr8010028 (2009).

2 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).

3 Ong, S. E., Mittler, G. & Mann, M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat Methods* **1**, 119-126, doi:10.1038/nmeth715 (2004).

4 Ong, S. E. & Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat Protoc* **1**, 2650-2660, doi:10.1038/nprot.2006.427 (2006).

5 Michalski, J.-C. in *New Comprehensive Biochemistry* Vol. Volume 30 (eds J. F. G. Vliegenthart J. Montreuil & H. Schachter) 55-97 (Elsevier, 1996).

6 Mononen, I., Fisher, K. J., Kaartinen, V. & Aronson, N. N., Jr. Aspartylglycosaminuria: protein chemistry and molecular biology of the most common lysosomal storage disorder of glycoprotein degradation. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **7**, 1247-1256 (1993).

7 Wang, R. J. D. a. A. M. in *The metabolic and molecular bases of inherited diseases* Vol. II (ed Scriver C.R.et al.) 2509-2528 (McGraw-Hill, 1995).

8 Aronson, N. N., Jr. & Davidson, E. A. Catabolism of mucopolysaccharides by rat liver lysosomes in vivo. *The Journal of biological chemistry* **243**, 4494-4499 (1968).

9 Aronson, N. N., Jr. & De Duve, C. Digestive activity of lysosomes. II. The digestion of macromolecular carbohydrates by extracts of rat liver lysosomes. *The Journal of biological chemistry* **243**, 4564-4573 (1968).

10 Snider, M. D. Metabolic labeling of glycoproteins with radioactive sugars. *Current protocols in cell biology / editorial board, Juan S. Bonifacino ... [et al.]* **Chapter 7**, Unit 7 8, doi:10.1002/0471143030.cb0708s13 (2002).

11 Diaz, S. & Varki, A. Metabolic radiolabeling of animal cell glycoconjugates. *Curr Protoc Protein Sci* **Chapter 12**, Unit 12 12 12 12 11-55, doi:10.1002/0471140864.ps1202s57 (2009).

12 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, doi:10.1634/stemcells.2006-0219 (2007).

13 Fang, M., Lim, J.-M. & Wells, L. *Quantitative Glycomics of Cultured Cells Using Isotopic Detection of Aminosugars with Glutamine (IDAWG)*. (John Wiley & Sons, Inc., 2009).

14 Aoki, K. *et al.* Dynamic developmental elaboration of N-linked glycan complexity in the Drosophila melanogaster embryo. *J Biol Chem* **282**, 9127-9142, doi:M606711200 [pii] 10.1074/jbc.M606711200 (2007).

15 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, doi:M804925200 [pii] 10.1074/jbc.M804925200 (2008).

16 Ceroni, A. *et al.* GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res* **7**, 1650-1659, doi:10.1021/pr7008252 (2008).

17 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).

18 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).

19 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).

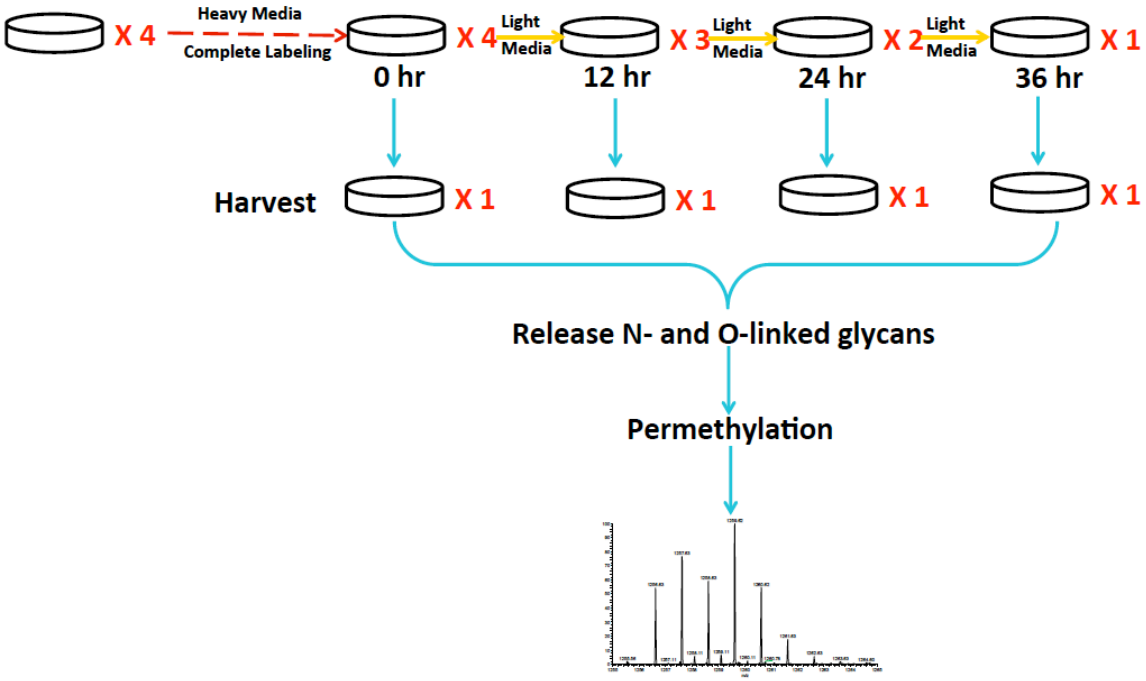


Figure 4.1: Pulse-Chase experiment designed for dynamic IDAWG. Cells are cultured in media with Amide-^{15}N-Gln until complete labeling. A portion of cells is harvested immediately and the other cells are transferred to media with normal glutamine. While kept in the light media, cells are harvested at different points of time. Both N- and O-linked glycans are released from the cells harvested at different time, permethylated and subjected to the analysis by Thermo ESI-LTQ-Orbitrap mass spectrometer.

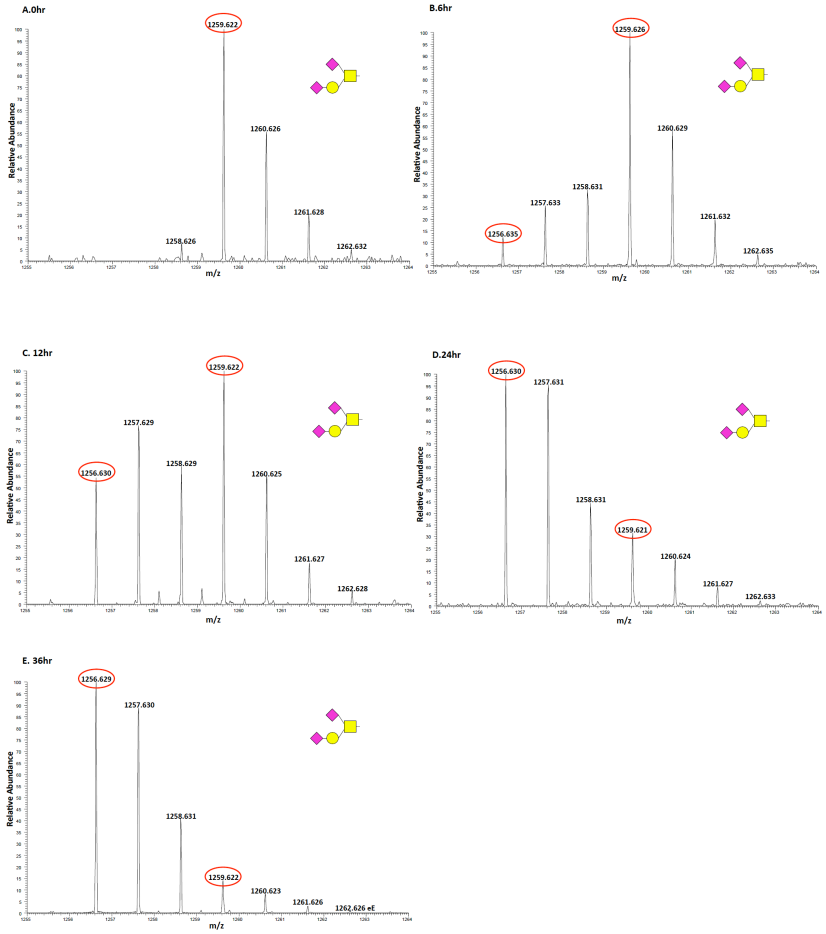


Figure 4.2: Mass spectra of one sialyted O-glycan structure released from hESCs at different points of time. Spectra of trapping scan of the permethylated disialyated T-antigen released from hESCs at 0 hour (**A**), 6 hour (**B**), 12 hour (**C**), 24 hour (**D**) and 36 hour (**E**).

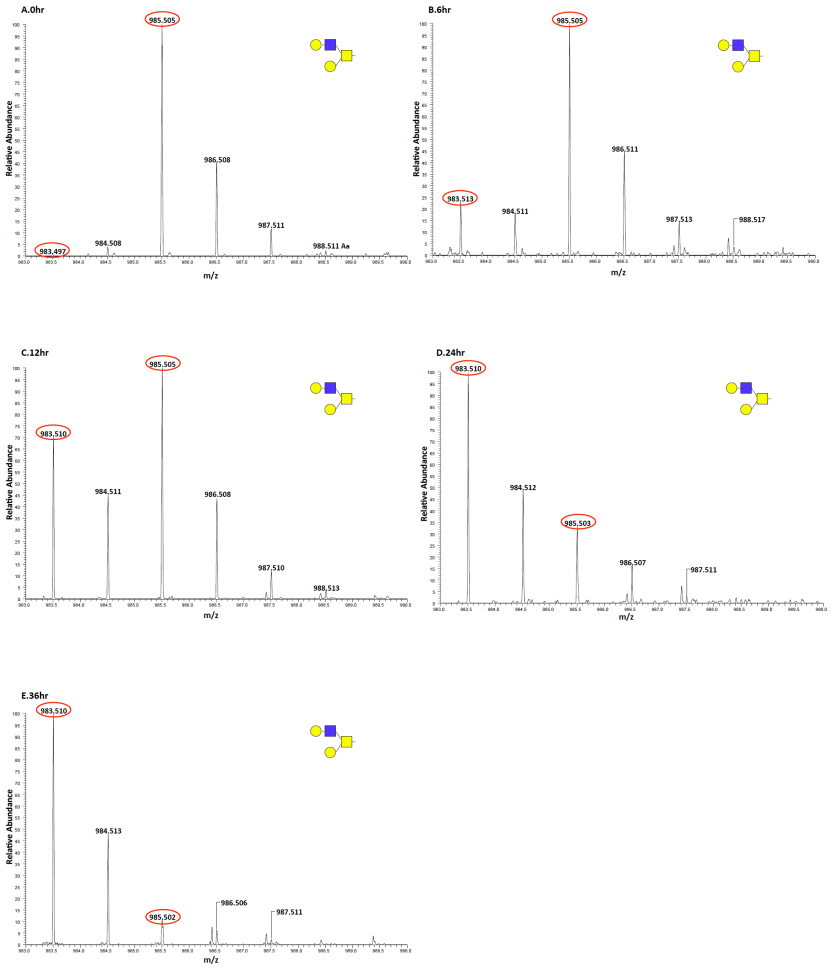


Figure 4.3: Mass spectra of one non-sialyted O-glycan structure released from hESCs at different points of time. Spectra of trapping scan of the permethylated non-sialyated core 2 structure released from hESCs at 0 hour (**A**), 6 hour (**B**), 12 hour (**C**), 24 hour (**D**) and 36 hour (**E**).

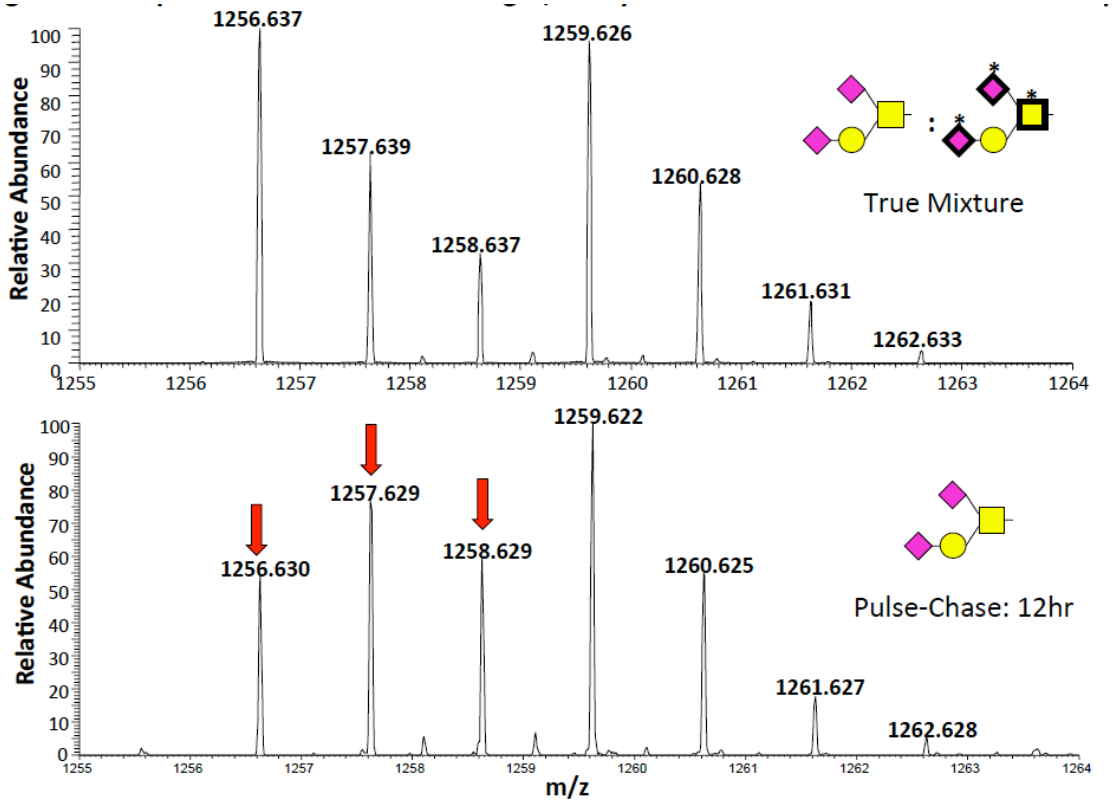


Figure 4.4: Comparison of true mixture of light/heavy structures and Pulse-Chase 12hr sample of the disialyated T-antigen. Top panel is the spectrum of true mixture of light and heavy labeled structures with a ratio of 1:1; Bottom panel is the spectrum from Pulse-Chase 12 hour sample (same as **Figure 4.2 C**). The red arrows denote the peaks with abnormal isotopic patterns.

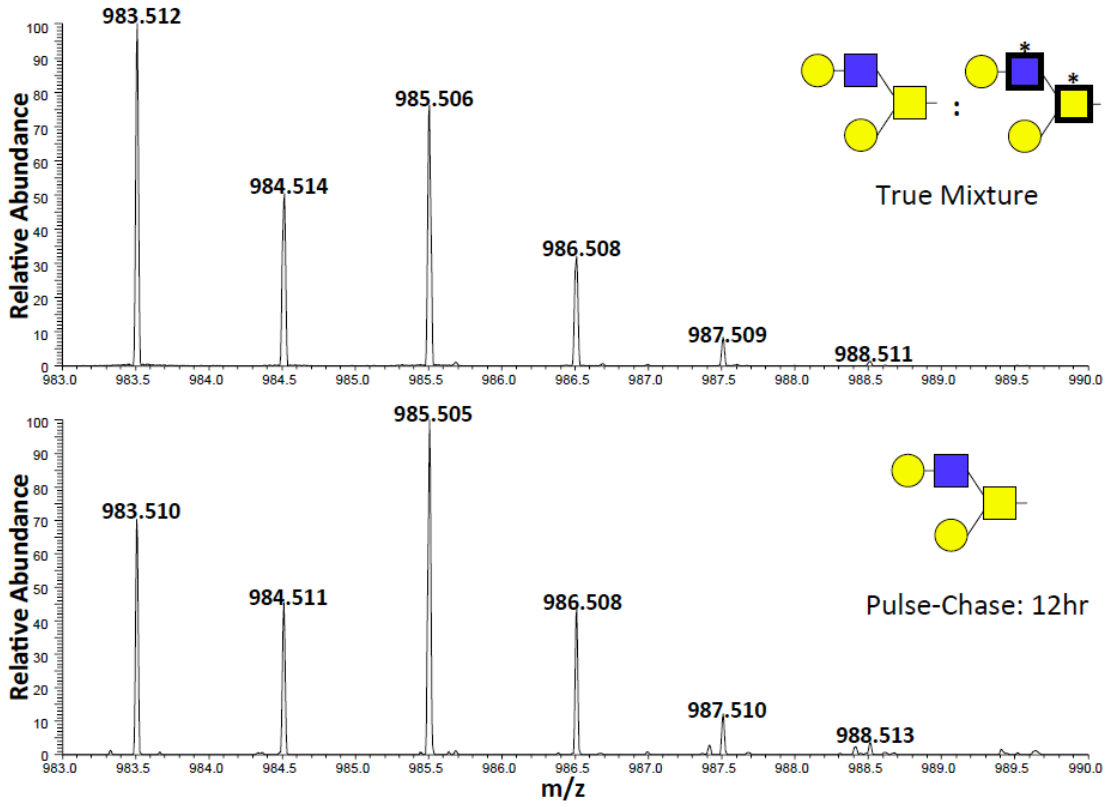


Figure 4.5: Comparison of true mixture of light/heavy structures and Pulse-Chase 12hr sample of the non-sialyated core 2 structure. Top panel is the spectrum of true mixture of light and heavy labeled structures with a ratio of 1:1; Bottom panel is the spectrum from Pulse-Chase 12 hour sample (same as **Figure 4.3 C**).

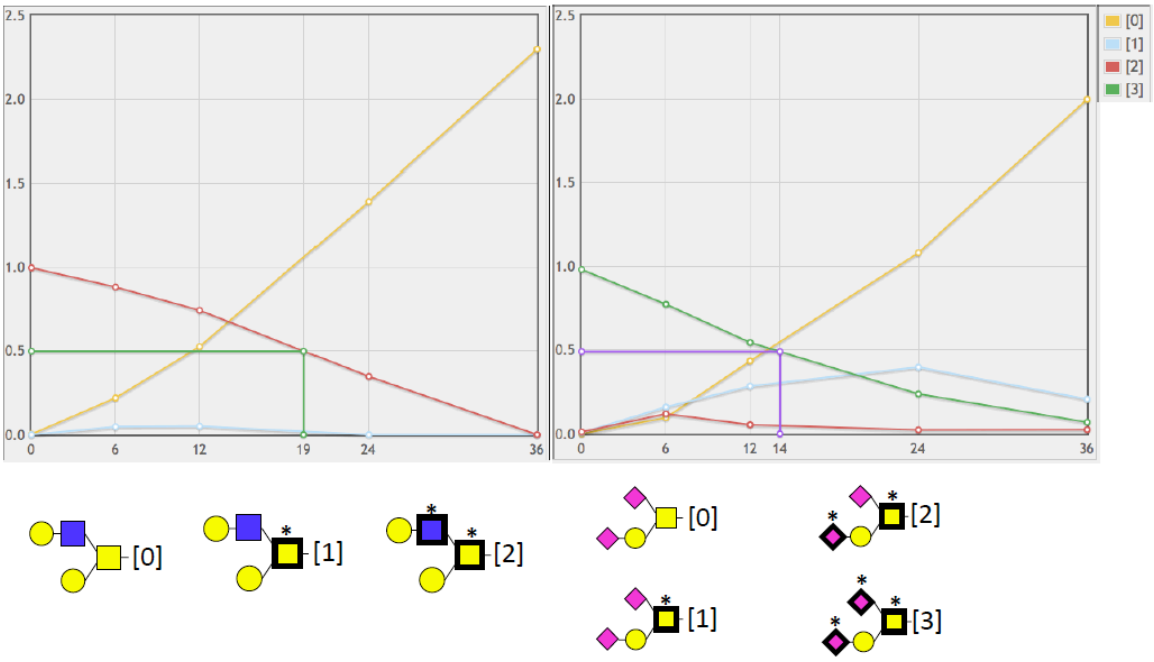


Figure 4.6. Comparison of the variation trend of each component at different points of time in the disialyated T-antigen and the non-sialyated core 2 structure reveals the remodeling of sialyated glycans. Left panel shows the trend lines for the non-sialyated core 2 structure; Right panel shows the trend lines for the disialyated T-antigen. The different line colors represent different components with a specific number of heavy nitrogens. The X-axis shows the time points of the Pulse-Chase experiments; the Y-axis shows the relative changes of each component against its original amount at 0 hour. The rectangular lines denote the 50% degradation time for components whose nitrogens are all heavy.

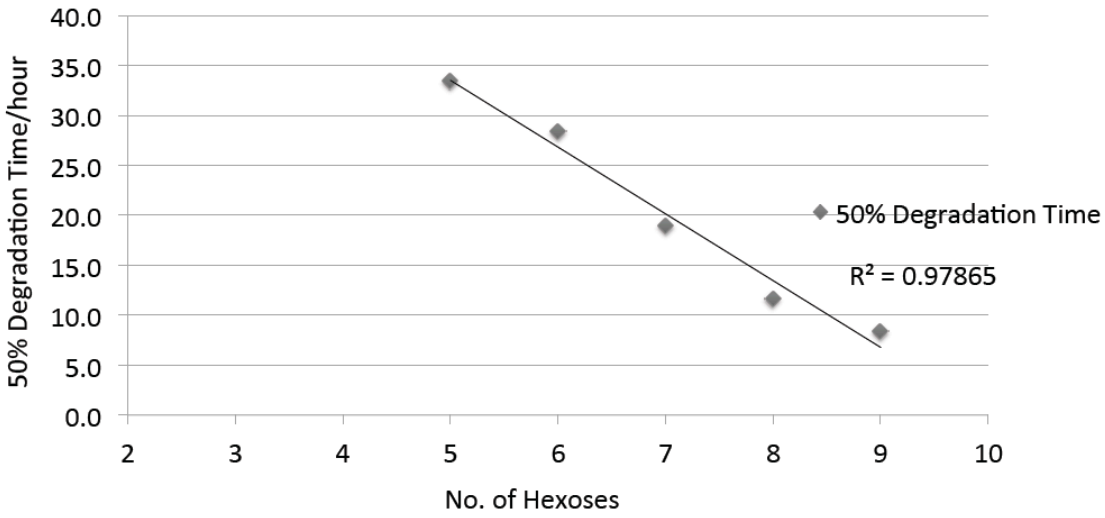


Figure 7. The relationship between the 50% degradation time and the number of hexoses in high-mannose N-linked glycans

Table 4.1: 50% degradation time and Proportion of remodeling at 50% degradation time for 9 major O-glycans

Structure	50% degradation time (hours)		Proportion of remodeling at 50% degradation time (%)	
	Rep1	Rep2	Rep1	Rep2
	18.6	19.4	1.9	1.2
	20.7	22.6	0.0	0.0
	7.7	8.4	0.0	0.0
	16.2	15.6	9.0	9.7
	14.9	14.1	26.8	25.4
	20.4	18.1	16.9	21.2
	18.4	19.8	6.3	3.4
	21.3	28.3	21.5	10.9
	22.1	19.4	27.9	27.0

Table 4.2: 50% degradation time and Proportion of remodeling at 50% degradation time for 15 major N-glycans

Structure	50% degradation time (hours)		Proportion of remodeling at 50% degradation time (%)	
	Rep1	Rep2	Rep1	Rep2
	11.0	12.2	9.2	13.7
	8.9	7.8	5.7	8.9
	29.3	27.4	9.6	8.0
	33.6	33.2	3.1	2.0
	19.2	18.7	12.6	12.0
	28.5	28.9	1.6	1.5
	36.0	31.6	0.0	0.0
	34.2	36.0	1.6	0.0

Structure	50% degradation time (hours)		Proportion of remodeling at 50% degradation time (%)	
	Rep1	Rep2	Rep1	Rep2
	21.8	16.4	26.7	30.8
	18.1	23.3	46.9	42.5
	15.1	14.1	35.1	31.9
	10.6	19.5	25.8	20.2
	9.5	13.8	42.0	39.4
	16.6	12.8	32.9	31.9
	12.0	15.9	52.6	38.3

CHAPTER 5

CONCLUSION

The overall purpose of this work was to optimize the IDAWG labeling strategy and apply the IDAWG technique to both the static and dynamic glycomic study of human embryonic stem cells.

In Chapter 2, the detailed protocols for Isotopic Detection of Aminosugars With Glutamine, termed as IDAWG, that is the first *in-vivo* cell culture isotope-labeling strategy for glycan analysis were provided. The methods of labeling the aminosugars (GlcNAc, GalNAc and sialic acid) with amide-^{15}N-Gln in cell culture and the detections of these aminosugars in both N- and O-linked glycan structures were described in the basic protocols. The mathematical calculations used to calculate Under-Incorporation rate and relative ratios of light/heavy species were described in the support protocols.

In Chapter 3, the differentially labeled human embryonic stem cells were used to validate the reliability and quantitative accuracy of the IDAWG technology by quantifying the expression levels of released N- and O-linked glycans. The results demonstrated that the IDAWG technology is able to quantify the difference of glycan expression levels as low as a 2-fold change with confidence. As focusing on O-linked glycans, we performed a global glycomic analysis of human embryonic stem cells and human definitive endoderms. The comparative quantitation of glycan expression levels between hESCs and hDEs by the IDAWG technique discovered a reciprocal change between poly-LacNAc and sialyation of core 2 O-linked structures accompanying the

differentiation from hESCs to hDEs, which agreed with the changes in transcript level. Along with the maturity of the IDAWG software, we anticipate that the IDAWG technique will become an important tool in the emerging field of comparative glycomics.

In Chapter 4, the Pulse-Chase experiments with IDAWG labeling strategy during the cell culture of human embryonic stem cells were performed to assess the dynamics of turnover for individual glycans. By fitting the mass spectra acquired for a series of samples using bioinformatics software we were able to investigate the degradation rates for major N- and O-linked glycans. For high-mannose N-glycans, there exists a negative correlation between 50% degradation time and the number of hexoses. By comparing the glycan spectra of pulse-chase samples with the spectra of real mixtures of light and heavy sample, we noticed an isotope mixture that could only be explained by remodeling of the glycans. We were able to approximate the amount of remodeling occurring for a subset of structures using bioinformatics tools. Sialyated structures are undergoing a relatively higher extent of remodeling compared to unsialyated ones, which suggests sialic acid residues have much to do with the phenomenon of remodeling.

In the future, computer software will be developed to automatically handle the datasets generated by the IDAWG experiment, which will be a great asset and mature the IDAWG strategy for quantitative glycomics. It is also possible that the glycans labeled with IDAWG can be analyzed using Multiple Reaction Monitor (MRM) strategy on a triple quadrupole mass spectrometer, which would increase the sensitivity a lot. Since the quantification using IDAWG can only be done in the full MS level, methods like iTRAQ, which can label the glycans at the reducing end for quantification in MS^2 are also in

development and will serve as a complementary strategy to IDAWG in the field of quantitative glycomics.