Published in final edited form as:

J Am Soc Mass Spectrom. 2013 September ; 24(9): 1376–1384. doi:10.1007/s13361-013-0681-2.

Individuality Normalization when Labeling with Isotopic Glycan Hydrazide Tags (INLIGHT): A Novel Glycan Relative Quantification Strategy

S. Hunter Walker¹, Amber D. Taylor¹, and David C. Muddiman^{1,*}

¹W.M. Keck Fourier Transform Mass Spectrometry Laboratory, Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695

Abstract

The INLIGHT strategy for the sample preparation, data analysis, and relative quantification of *N*-linked glycans is presented. Glycans are derivatized with either natural (L) or stable-isotope labeled (H) hydrazide reagents and analyzed using reversed phase liquid chromatography coupled online to a Q Exactive mass spectrometer. A simple glycan ladder, maltodextrin, is first used to demonstrate the relative quantification strategy in samples with negligible analytical and biological variability. It is shown that after a molecular weight correction due to isotopic overlap and a post-acquisition normalization of the data to account for both the systematic variability, a plot of the experimental H:L ratio vs. the calculated H:L ratio exhibits a correlation of unity for maltodextrin samples mixed in different ratios. We also demonstrate that the INLIGHT approach can quantify species over four orders of magnitude in ion abundance. The INLIGHT strategy is further demonstrated in pooled human plasma, where it is shown that the post-acquisition normalization is more effective than using a single spiked-in internal standard. Finally, changes in glycosylation are able to be detected in complex biological matrices, when spiked with a glycoprotein. The ability to spike in a glycoprotein and detect change at the glycan level validates both the sample preparation and data analysis strategy, making INLIGHT an invaluable relative quantification strategy for the field of glycomics.

INTRODUCTION

The ubiquity of glycosylation in biological systems is evident in that the majority of translated gene products are glycosylated,[1, 2] and glycosylation is involved in cellular recognition, cell-cell interactions, cell division and adhesion, protein folding and stability, and protein function.[3] Due to the importance of glycosylation in biological systems, it is necessary to develop analytical strategies to profile and quantify glycans in a high-throughput and reproducible manner. This applies to not only sample preparation and the analysis platform (often liquid chromatography coupled online to mass spectrometry (LC-MS)), but also to the bioinformatic processing strategy. Recently, numerous strategies have been described in the literature for the relative quantification of *N*-linked glycans using LC-MS-based stable-isotope labeled derivatization strategies.[4–11] However, because these strategies have only recently been developed and there are limited bioinformatic platforms for the analysis of LC-MS glycomic data, few glycomics experiments can be repeated, compared, or even reproducibly analyzed across laboratories.

*Author for Correspondence, David C. Muddiman, Ph.D., W.M. Keck Fourier Transform Mass Spectrometry Laboratory, Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695, Phone: 919-513-0084, david_muddiman@ncsu.edu.

Stable-isotope labeling (SIL) for the relative quantification of released glycans has been described for metabolic SIL[4] and several different types of SIL glycan derivatization strategies including permethylation,[5, 7, 11] reductive amination,[6, 10] PMP derivatization,[8] and hydrazone formation.[9] These strategies serve to globally measure differences in liberated glycan abundances between samples. While these strategies do not provide glycosylation site information or glycan occupancy on a given protein, global relative quantification of glycans can be effectively used to study the up and down regulation of glycans in disease. Thus, glycan-based biomarker discovery efforts will greatly benefit from these strategies. Additionally, it must also be noted that the regulation of glycotransferases plays a role in disease.[12] Because glycotransferases have the capability of non-specifically adding monosaccharides to proteins, a global glycan profiling strategy may be a more effective method for studying aberrant glycosylation. Finally, a global relative quantification strategy is useful to initially determine glycans of interest in a biological system. For example, the glycans changing in a biological system can be measured using a relative quantification strategy, and then one can focus on specific glycans for more detailed and informative analyses such as site of glycosylation, protein information, and occupancy. Thus, these techniques are capable of being biologically useful and powerful analytical technologies for a large array of glycomics applications.

Stable-isotope labeled derivatization strategies involve differentially labeling two glycan samples *via* either a 'light' or 'heavy' reagent. This allows the two samples to be mixed together, analyzed in the same LC-MS run, and separated by mass. The relative ion abundances of a specific 'light' and 'heavy' glycan pair can then be measured, and is proportional to the relative amounts of the glycan in each sample. In these types of studies, there are numerous decisions that must be made on how to reproducibly measure the relative amounts of the light- and heavy-labeled glycans. These decisions include whether to quantify based on ion abundance ratios or extracted ion chromatogram (EIC) peak area ratios, whether to use the monoisotopic peak or the entire isotopic distribution for quantification, how to control for differential isotopic shift and/or isotopic overlap, and how to normalize for any systematic variability introduced due to the parallel sample preparation and systematic biological variability. Thus, a strategy must be developed in order to process SIL relative quantification LC-MS data in a reproducible manner. Each of the relative quantification strategies that has been developed is unique such that one must process the data slightly differently. For example, the isotopic permethylation strategies and IDAWG (Isotopic Detection of Aminosugars With Glutamine)[4] will have a different m/z shift for each glycan depending on the size, branching, and constituent monosaccharides present. In contrast, techniques such as reductive amination, PMP derivatization, and hydrazone formation introduce a fixed number of stable-isotopes resulting in a fixed m/z shift. The stable-isotope label also must be noted. Because ^{13}C and ^{15}N are not chromatographically resolved from ^{12}C and ^{14}N , respectively, it is practical to use LC-MS and quantify by EIC peak area. However, SIL techniques which incorporate deuterium should use caution when paired with LC-MS due to the possibility of chromatographic separation,[13–15] which can lead to skewed quantification based on different mobile phase composition and different electrospray ionization efficiency. Thus, these studies will be the most effective using direct infusion ESI or MALDI and quantifying based on ion abundance.

Because the purpose of developing glycan relative quantification strategies is most often to measure differences in the glycan profiles of biological systems, it is necessary to minimize the random analytical variability and account/correct for any systematic analytical and biological variability such that the biological change of interest can be Identified. The most effective strategy for minimizing sample preparation variability in glycan SIL relative quantification studies are those that involve metabolic labeling (e.g. IDAWG) due to the mixing of the two samples at the onset of sample preparation. However, the majority of

glycan relative quantification studies are not capable of being performed in cell culture. Thus, many of the glycan SIL strategies involve parallel sample preparation in which variability is introduced at each step and must be controlled and corrected for in order to avoid detecting differences in relative abundance due to sample preparation variability. This can be accomplished by incorporating internal standards at the onset of sample preparation and has been demonstrated in previous glycan literature.[9, 16–21] The most common internal standards used in glycan relative quantification experiments include free sugars (such as maltoheptaose). While these internal standards are capable of correcting for sample preparation variability in the purification of *N*-linked glycans, derivatization efficiency, and other technical variability, studies which require the release of *N*-linked glycans from proteins often continue to use free glycans as internal standards. However, variability can be introduced when using enzymatic cleavage (PNGases) due to lot-to-lot variability, loss of enzyme activity over time, and when using an enzyme in an undefined biological matrix that could contain endogenous species which may inhibit the enzyme. Thus, a recent report has demonstrated the use of a plant glycoprotein (horseradish peroxidase) as an internal standard that can be used to account for all sample preparation systematic variability, including enzymatic release efficiency, when processing samples in parallel.[20]

Though using a single internal standard has been shown to be effective for samples with similar biological matrices (aliquots of pooled human plasma),[20] often very little is known about the specific matrix of each sample. For example, when comparing a biological specimen from two individuals, the total protein concentration may be different, the total glycosylation may be different, and/or inhibitors to PNGase F may be present. This is defined here as the “individuality” of each biological sample. Thus, each of these situations complicates relative quantification, and an internal standard will not be able to account for systematic biological variation due to the individuality of each sample. Additionally, a single glycan internal standard may act significantly different than other types of glycans (e.g. high mannose, complex, hybrid, fucosylated, acidic), creating a biased correction based on the chosen internal standard. Attempts can be made to normalize to the total protein concentration; however, this is not always proportional to the total glycosylation between different biological samples and biological states.

When measuring changes in a biological system (e.g. onset of disease) every effort must be made to define the inter-individuality (*i.e.* the variability of biological constituents in a population[22, 23]) of a specific biological sample set. Though these changes are often random between samples and can only be defined but not minimized, the previously mentioned systematic biological variations due to individuality differences between samples must be accounted for in order to identify biological change. Because these situations cannot be accounted for using a spiked-in internal standard, other methods must be developed such that both analytical and biological systematic variability can be accounted for. Strategies for normalization in proteomic label free relative quantification experiments have been reported.[24–29] These studies discuss how to identify biological change while accounting for variations in the sample preparation. Two proteomic normalization strategies can be applied to glycans: 1) normalization to selected proteins (analogous to spiking in a glycan internal standard – *vide supra*),[29] and 2) normalization based on total spectral counts (TSpC).[24, 25, 29, 30] These two normalization strategies have been recently compared[29] and show that TSpC normalization performs more effectively than the normalization to selected proteins. This gives insight into possible glycan normalization methods and leads one to believe that normalization to the total glycosylation of each sample may be a more effective strategy than using an individual spiked-in internal standard.

Herein, the INLIGHT method for generating and analyzing glycan SIL relative quantification data is described in detail, where *N*-linked glycans are derivatized with fixed

mass shift SIL reagents[9] *via* hydrazone formation[31, 32] and analyzed by reversed phase liquid chromatography (RPLC) coupled to a Q Exactive mass spectrometer. The reagents used for derivatization have been previously reported,[9] and offer significant advantages (increased ionization efficiency,[31, 32] near stoichiometric derivatization,[31, 32] analysis by RPLC,[33] no post-derivatization clean up step, and <4 hr derivatization time) in addition to the ability to perform relative quantification in LC-MS. The INLIGHT strategy for relative quantification involves generating an extracted ion chromatogram for each light- and heavy-derivatized glycan based on the monoisotopic peak at a MMA of ± 5 ppm, correcting for the isotopic overlap of large molecular weight (MW) glycans, correcting for systematic analytical and biological variability, and determining the relative amounts of *N*-linked glycans in each sample. A method for isotopic overlap correction of glycan SIL relative quantification data is presented, and a post-acquisition data normalization strategy is described that is capable of accounting for both analytical and biological systematic variability. The INLIGHT strategy is demonstrated to be effective for the relative quantification of glycans using both a standard glycan ladder and pooled human plasma while minimizing the systematic variability in the measurement. The success of the INLIGHT strategy is ultimately demonstrated by the detection of significant change in *N*-linked glycan abundances when a glycoprotein is spiked into pooled human plasma.

EXPERIMENTAL

Materials

Maltodextrin was purchased from V-Laboratories, Inc. (Covington, LA), and horseradish peroxidase, pooled human plasma, and other consumable chemicals were purchased from Sigma Aldrich (St. Louis, MO). Peptide: *N*-glycosidase F (PNGase F) was purchased from New England BioLabs (Ipswich, Ma). HPLC grade solvents were purchased from Burdick & Jackson (Muskegon, MI). The hydrazide reagent, 4-phenethylbenzohydrazide, was synthesized previously in both the natural and stable-isotope labeled form in the Department of Chemistry at North Carolina State University.[9]

N-linked Glycan Derivatization Procedure

Maltodextrin and *N*-linked glycans cleaved from pooled human plasma were derivatized with either the light or heavy 4-phenethylbenzohydrazide (P2GPN) using the method previously described.[32] Briefly, a 1 mg/mL solution of each reagent was made immediately prior to reaction in 25:75 acetic acid:methanol (*v/v*). To the dried glycan sample, 200 μ L of the reagent solution was added. The sample was vortexed, centrifuged, and incubated at 56°C for 3 hr. The samples were dried *in vacuo* at 55°C, and the dried powder was stored until analysis at -20°C.

Maltodextrin Sample Preparation

The maltodextrin sample has been prepared such that each vial contains a 50 μ g aliquot of dried glycan powder. The samples were either derivatized with the light or heavy P2GPN reagent, according to the procedure detailed above. Before analysis, the samples were reconstituted in 200 μ L of the initial LC conditions. The samples were then vortexed, centrifuged, and the supernatant was extracted and analyzed.

N-linked Glycan Extraction from Pooled Plasma

N-linked glycans were cleaved from pooled human plasma aliquots following methods detailed previously.[21, 33] Briefly, at the onset of sample preparation 200 μ g of a plant glycoprotein internal standard, horseradish peroxidase (HRP), was added. PNGaseF was used to liberate the *N*-linked glycans from 50 μ L aliquots of pooled human plasma, the reaction was incubated for 18 hr, and an ethanol precipitation step was performed

immediately after glycan cleavage to remove a majority of the plasma proteins. The remaining glycans were further purified using solid phase extraction (SPE). Following SPE, the samples were dried and stored at -20°C until analysis. The samples were either derivatized with the light or heavy P2GPN reagent, according to the procedure detailed above. Before analysis, the samples were reconstituted in 200 μL of the initial LC conditions. The samples were then vortexed, centrifuged, and the supernatant was extracted and analyzed.

cHiPLC nano-Flow Reversed Phase Chromatography

The derivatized glycan samples were analyzed using reversed phase liquid chromatography. Ten microliters of the glycan samples were injected onto the column using an EASY-nLC II autosampler and liquid chromatography (LC) system (Thermo Fisher Scientific, San Jose, CA) in mobile phase A at 650 nL/min. The sample was then washed with an additional 2 μL of mobile phase A. Mobile phase A and B consist of 98/2/0.2% and 2/98/0.2% water/ acetonitrile/formic acid, respectively. The separation of derivatized *N*-linked glycans was performed on a cHiPLC system (Eksigent, Dublin, CA) in the vented column configuration, [34] using a ChromXP C18-CL 75 $\mu\text{m} \times 15\text{ cm}$ analytical column for both the trap and analytical column. This dual-analytical column setup allowed for a total column length of 30 cm, and glycans were separated at a flow rate of 300 nL/min. The initial conditions of 2% mobile phase B were held for 1 min, and then, mobile phase B was ramped to 22% over 1 minute. The mobile phase B composition was further ramped to 35% over 35 min, and the column was washed at 95% mobile phase B and re-equilibrated for 15 min at the initial solvent conditions.

Q Exactive Mass Spectrometry

The RP chromatography system is coupled online to a Q Exactive mass spectrometer equipped with a nanospray source (Thermo Fisher Scientific, San Jose, CA).[35] Glycans were ionized by applying 2.25 kV to the union between the outlet of the LC system and the emitter tip. The MS inlet capillary was heated to 225°C . Glycans were detected using data dependent acquisition, where up to 5 precursor ions per full scan were chosen for MS/MS fragmentation. The 5 precursor ions were chosen based on ion abundance, and once an ion is chosen, it is put on an exclusion list for 25 s so that other precursors may be fragmented in the next cycle. In the full scan, precursor ions are detected in the Orbitrap at a resolving power of 70,000 (FWHM) at $m/z = 200$. The automatic gain control (AGC) was set to 1×10^6 ions, and the maximum injection time was set to 250 ms. Upon selection for fragmentation, the precursor ions are isolated and fragmented in the higher energy collision dissociation (HCD) cell at a normalized collision energy (NCE) of 20. In the product ion scan, ions are detected in the Orbitrap at a resolving power (FWHM) of 17,500 at $m/z = 200$. The AGC was set to 2×10^4 ions, and the maximum injection time was set to 120 ms. The instrument was calibrated just before the data set per the manufacturer's specifications. Unique glycan compositions were manually identified *via* accurate mass ($\pm 5\text{ ppm}$) and analyzed using the manufacturer's software (Xcalibur v.2.2). The *N*-linked glycan database, detailed integration, and processing strategy are all included in the supplementary material.

RESULTS AND DISCUSSION

In this study, two samples, a hexose sugar ladder and pooled human plasma, were used to develop and demonstrate the effectiveness of the INLIGHT relative quantification strategy. Maltodextrin, a linear glucose polymer, was used as a standard for developing a reproducible strategy for the analysis of SIL glycan relative quantification data. Using maltodextrin as a standard affords several advantages including minimized sample preparation variability in comparison to a biological sample, a large molecular weight range

(~150-10,000 Da), and a wide abundance range (>5 orders of magnitude). This allows one to determine the effectiveness of the glycan quantification strategy and processing method at a wide variety of glycan sizes and abundances, mimicking a biological sample, while minimizing the variability.

The native and SIL P2GPN reagents (described previously[9]) were used in this experiment to derivatize glycans and have a nominal mass difference of 6 Da. At low molecular weights (~1000 Da), a 6 Da shift in the SIL tag mass is sufficient to fully separate the isotopic distributions (Figure 1a). However, as the molecular weight of the glycans becomes larger, the isotopic distributions begin to overlap (Figure 1b – d). This poses a problem because the current method for analyzing SIL relative quantification data is by generating extracted ion chromatograms of both the light and heavy monoisotopic peaks and creating a peak area ratio. Because the monoisotopic peaks (A peak) are used to extract the chromatograms for quantification, any overlapping isotopic distribution from the light-labeled sample will contribute to the area of the heavy-labeled sample. In order to account for this isotopic overlap of large molecular weight glycans, the isotopic distribution was modeled for each maltodextrin analyte detected. Using the theoretical isotopic distribution, the ratio of the A +6_{Light} peak to the A_{Light} peak can be determined for each molecular weight. This ratio can then be used determine the amount of area contributed by A+6_{Light} peak and subtracted from the $[M+nH^+]^{n+}_{Heavy}$ area. Thus the corrected heavy peak area ($[M+nH^+]^{n+}_{Heavy_Corrected}$) can be calculated using Equation 1, and the H:L ratio can be calculated using Equation 2.

$$\sum [M+nH^+]^{n+}_{Heavy_Corrected} = \sum [M+nH^+]^{n+}_{Heavy} - \left(\frac{A+6_{Light}}{A_{Light}}\right) \times \sum [M+nH^+]^{n+}_{Light} \quad (\text{Equation 1})$$

$$\text{Relative Glycan Abundance Ratio} = H:L = \frac{\sum [M+nH^+]^{n+}_{Heavy_Corrected}}{\sum [M+nH^+]^{n+}_{Light}} \quad (\text{Equation 2})$$

In order to demonstrate contribution of the isotopic overlap to the relative quantification data and the effectiveness of correcting the data using Equations 1 and 2, two aliquots of maltodextrin were differentially labeled (one with the light and one with the heavy reagent), the two samples were mixed in a 1:1 ratio, and analyzed. It is shown in Figure 2a that the expected 1:1 glycan ratios are not correctly measured in the raw data. Additionally, Figures 2b and 2c show that the ratios of these data are skewed by a combination of both the increasing molecular weight and the decreasing ion abundance in the spectra. Upon closer observation, it is clear that at low glycan ion abundances, the incorrect measured ratio is due to the glycan concentrations in the sample approaching the detection limits of the instrument. Here, one peak of the ion pair is at or just above the detection limit, causing the detected glycan ratios to be non-representative of the data. These glycans are shaded in the blue box in both Figures 2b and 2c. Thus, based on this observation and the specification of our instrument, the data will be truncated below a normalized abundance value of 1×10^{-4} .

After the normalized abundance cutoff was applied, the data were further corrected using Equations 1 and 2. By correcting for the isotopic overlap of the glycans, the molecular weight dependence of the ratio is eliminated (Figure 2e). Although it is shown that the isotopic overlap can be corrected for, the glycans are still not detected with the ideal 1:1 ratio in Figure 2d–f. Thus, this data set was used to demonstrate the effectiveness of a Total Glycan Normalization Factor (TGNF) for the correction of systematic variability in the sample and sample preparation (Figure 2g–i). This TGNF is analogous to TSPC normalization[24, 25, 29] in label free proteomic methods and is described by Equation 3. This method sums both the heavy glycan areas and light glycan areas (post molecular weight

correction) and takes the ratio in order to calculate the TGNF (larger value always on top). This factor is then multiplied by each of the glycan areas in the sample with the smaller of the two sums in order to normalize the two data sets. Equation 3 sums the 30 most abundant glycan areas for each sample. All glycans are not used due to the mentioned adverse effects of low abundant glycans (*vide supra*). For studies involving multiple samples and comparisons, the same 30 glycans were used for all samples. The 30 most abundant glycans from a single sample (arbitrarily chosen) were used to calculate the TGNF normalization for all samples. Even though these may not be the most abundant glycans for all samples, it is assumed that the relative abundances are similar for glycans in all samples.

$$TGNF = \frac{\sum_{i=1}^{30} Glycan_{Area_Heavy}}{\sum_{i=1}^{30} Glycan_{Area_Light}} \quad (\text{Equation 3})$$

It is shown in Figures 2g – i that the TGNF (in combination with the molecular weight correction and abundance truncation) is capable of correcting for nearly all of the systematic variation of the sample preparation and analysis method. This allows for accurate relative quantification for 4 orders of magnitude when using the INLIGHT strategy, and after filtering out the low-abundance data (< the normalized value of 1×10^{-4}), the mean $\pm$ 95% confidence interval for the TGNF and molecular weight corrected H:L ratio is calculated to be 1.06 ± 0.02 . In contrast, the mean $\pm$ 95% confidence interval for the raw data (Figures 2a – c) after a cutoff is calculated to be 1.37 ± 0.04 . Additionally, a two-tailed f-test for different variances was performed, and it was shown that the variance of the corrected data was significantly less than the variance of the raw data. This is attributed to the molecular weight correction, and the reduction in variability will allow for lower cutoffs when determining significant biological change.

An additional data set was prepared such that the light- and heavy-labeled maltodextrin samples were mixed in the following ratios to generate a total of 7 samples: 5:1, 2:1, 1.5:1, 1:1, 1:1.5, 1:2, and 1:5. Figure 3 plots the TGNF and molecular weight corrected experimental H:L ratio vs. the calculated H:L ratio for 3 representative glycans of different molecular weights: Hex₄, Hex₁₄, and Hex₂₄. The data are plotted in triplicate for each of the three glycans, and a weighted least squares regression was calculated. The slope $\pm$ 95% confidence interval of the regression in Figure 3 was found to be 1.01 ± 0.03 , and the intercept $\pm$ 95% confidence interval was found to be 0.004 ± 0.06 . These data show that the INLIGHT strategy is capable of statistically measuring the ideal slope of unity and ideal intercept of 0, and the linear range of quantification spans a 5-fold change in glycan abundance in both directions.

The maltodextrin ladder has been used as an effective model for N-linked glycans in mammalian biological systems, and it has been shown that the INLIGHT strategy is capable of quantifying glycans in a sample with ~50 oligosaccharides. However, this model glycan ladder required minimal sample preparation, thus allowing for minimal sample preparation variability. The following studies were performed in aliquots of pooled human plasma in order to determine the effectiveness of the INLIGHT strategy in samples requiring significantly more complex sample preparation strategies. Thus, variability (both systematic and random) is incorporated at each step of the sample preparation (*e.g.*, glycan cleavage, solid phase extraction, derivatization) due to the samples being quantified having to be processed in parallel. Pooled plasma aliquots were prepared according to Figure 4, where the internal standard glycoprotein (HRP) was added to each of the 10 aliquots, and 100 μ g of α -1-acid glycoprotein (AGP) were spiked into two aliquots to determine the effectiveness of

using INLIGHT to relatively quantify *N*-linked glycans in complex biological matrices (*vide infra*).

The 8 pooled plasma samples that did not contain spiked AGP were derivatized and compared according to Figure 4a in order to measure and compare the variability of using the previously reported HRP internal standard correction vs. using the TGNF post-acquisition correction. Figure 5 shows histograms of the $\log_2(\text{H:L})$ for these 7 samples in triplicate for the internal standard (**5a**) and TGNF (**5b**) corrected samples. It is shown that in these samples where the biological variability is assumed to be zero (*i.e.*, pooled plasma), that both the internal standard and the TGNF methods are capable of measuring the expected 1:1 ratio with distributions centered around zero. However, qualitatively the distribution for the TGNF method is narrower than the internal standard corrected distribution, and there are ~33% more glycans in the TGNF corrected center-most bin (0 ± 0.05). Thus, an f-test was performed, and it was shown that the TGNF method allows for significantly less variance after correction than when using the internal standard correction.

The 2 spiked AGP plasma samples (one light and one heavy) were compared to a corresponding light or heavy sample of the normal pooled plasma samples analyzed in the previous experiment. These comparisons were made in order to test whether or not the INLIGHT strategy is capable of detecting change in arguably the most complex of biological samples and matrices, plasma. Because a glycoprotein was spiked in and not a standard free glycan, there is no way to know the added glycan concentration or what the expected change should be. However, it has been shown in the literature that AGP is significantly glycosylated with bi- tri- and tetra-antennary complex-type *N*-linked glycans containing *N*-acetylneuraminic acid.[36, 37] Thus, it would be expected that one would observe significantly different glycan abundances for large, acidic glycans, whereas neutral glycans would not be determined statistically different.

The glycan abundance ratios measured in the analyses comparing the AGP-spiked and normal pooled plasma samples are compared to the measured glycan ratios from the seven normal plasma comparisons in the previous experiment. Table 1 shows example glycan abundance ratios from the comparison of the AGP-spiked and normal plasma samples. The *N*-linked glycans shown in the top of Table 1 are examples of 4 multi-antennary glycans containing acidic monosaccharide moieties, and it is observed that when the AGP-spiked glycan ratios are compared to the normal pooled plasma glycan ratios, the $\log_2(\text{H:L})$ ratios significantly deviate from zero ($\pm$ depending on whether the AGP-spiked sample was labeled light or heavy). In contrast, examples of neutral glycan abundance ratios from the same samples are shown on the bottom half of Table 1, and no significant difference is detected. All the glycan ratios in Table 1 were compared to the glycan ratios of the 7 normal pooled plasma samples comparisons (in triplicate), and using a t-test, only the glycans with acidic monosaccharide residues were found to have significantly different abundance ratios. Furthermore, the light-labeled AGP-spiked sample ratios are detected as statistically significant negative values, whereas the heavy-labeled AGP-spiked samples ratios are measured as statistically significant positive values. This demonstrates the lack of labeling bias and further gives confidence to the fact that the molecular weight correction sufficiently minimizes the effects due to isotopic overlap. These *N*-linked glycans in which change is significantly detected are large molecular weight glycans (~3000–4000 Da), and if the change detected was due to the isotopic overlap, then the ratios would be the same for both samples instead of change in opposite directions, as seen in Table 1. Additionally, it must be noted that the amount of change detected in the AGP-spiked samples is less than a 2-fold change, which is often used as a cutoff in determining biological significance. Furthermore, as little as a 20% change in the glycan abundance ($\text{H}_6\text{N}_5\text{F}_1\text{A}_2$ glycan) was determined to be

significantly different. This is a testament to the precision and utility of the INLIGHT strategy for the relative quantification of *N*-linked glycans in complex biological samples.

CONCLUSION

The strategy presented herein, Individuality Normalization when Labeling with Isotopic Glycan Hydrazide Tags (INLIGHT), demonstrates the effective quantification of *N*-linked glycans both in a model system (maltodextrin ladder) and pooled human plasma. This method has been developed such that the SIL derivatization can be coupled to nearly any *N*-linked glycan (or any oligosaccharide with a free aldehyde at the reducing terminus) sample preparation strategy with minimal time and monetary cost (<4hr of additional sample preparation time). Because of the possible benefits to many glycomics experiments, a detailed strategy for the relative quantification of INLIGHT has been presented.

In order to measure significant biological change, up-front sample preparation strategies and back-end data analyses must be optimized in order to minimize systematic variability introduced into the measurement. These efforts have been described for the INLIGHT strategy, where variation due to isotopic overlap has been corrected. Additionally, the ability to more effectively correct for systematic analytical and biological variability using a post-acquisition normalization strategy rather than a single internal standard is presented for the first time in a glycan sample. This attention to minimizing the variability and the inherent advantages of INLIGHT make this an attractive and easily implemented strategy for those interested in measuring biological changes in glycosylation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to gratefully acknowledge the financial support received from the NIH – NCI IMAT Program (Grant #R33 CA147988-02), the W. M. Keck Foundation, and North Carolina State University.

REFERENCES

1. Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *BBA-Gen. Subjects*. 1999; 1473(1):4–8.
2. Begley, TP. *Wiley Encyclopedia of Chemical Biology*. Vol. 2. New Jersey: John Wiley & Sons, Inc.: Hoboken; 2009. p. 785
3. Varki, A.; Cummings, RD.; Esko, JD.; Freeze, HH.; Stanley, P.; Bertozzi, CR.; Hart, GW.; Etzler, ME. *Essentials of Glycobiology*. 2 ed.. New York: Cold Spring Harbor Laboratory Press: Cold Spring Harbor; 2009.
4. Orlando R, Lim JM, Atwood JA, Angel PM, Fang M, Aoki K, Alvarez-Manilla G, Moremen KW, York WS, Tiemeyer M, Pierce M, Dalton S, Wells L. IDAWG: Metabolic Incorporation of Stable Isotope Labels for Quantitative Glycomics of Cultured Cells. *J. Proteome Res*. 2009;3816–3823. [PubMed: 19449840]
5. Atwood JA, Cheng L, Alvarez-Manilla G, Warren NL, York WS, Orlando R. Quantitation by isobaric labeling: Applications to glycomics. *J. Proteome Res*. 2008;367–374. [PubMed: 18047270]
6. Xia BY, Feasley CL, Sachdev GP, Smith DF, Cummings RD. Glycan reductive isotope labeling for quantitative glycomics. *Anal. Biochem*. 2009; 387(2):162–170. [PubMed: 19454239]
7. Kang P, Mechref Y, Kyselova Z, Goetz JA, Novotny MV. Comparative glycomic mapping through quantitative permethylation and stable-isotope labeling. *Anal. Chem*. 2007;6064–6073. [PubMed: 17630715]

8. Zhang P, Zhang Y, Xue XD, Wang CJ, Wang ZF, Huang LJ. Relative quantitation of glycans using stable isotopic labels 1-(d0/d5) phenyl-3-methyl-5-pyrazolone by mass spectrometry. *Anal. Biochem.* 2011; 418(1):1–9. [PubMed: 21803021]
9. Walker SH, Budhathoki-Uprety J, Novak BM, Muddiman DC. Stable-isotope labeled hydrophobic hydrazide reagents for the relative quantification of N-linked glycans by electrospray ionization mass spectrometry. *Anal. Chem.* 2011; 83(17):6738–6745. [PubMed: 21774516]
10. Bowman MJ, Zaia J. Comparative Glycomics Using a Tetraplex Stable-Isotope Coded Tag. *Anal. Chem.* 2010; 82(7):3023–3031. [PubMed: 20230064]
11. Alvarez-Manilla G, Warren NL, Abney T, Atwood J, Azadi P, York WS, Pierce M, Orlando R. Tools for glycomics: relative quantitation of glycans by isotopic permethylation using (CH₃I)-C-13. *Glycobiology.* 2007:677–687. [PubMed: 17384119]
12. Kobata A, Amano J. Altered glycosylation of proteins produced by malignant cells, and application for the diagnosis and immunotherapy of tumours. *Immunol. Cell Biol.* 2005; 83(4):429–439. [PubMed: 16033539]
13. Filer CN. Isotopic fractionation of organic compounds in chromatography. *J. Labelled Compd. Radiopharm.* 1999:169–197.
14. Filer CN, Fazio R, Ahern DG. (+/-)-[methyl-H-3 and methyl-H-2]mianserin - participants in a dramatic instance of HPLC isotopic fractionation. *J. Org. Chem.* 1981:3344–3346.
15. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 1999:994–999. [PubMed: 10504701]
16. Dixon RB, Bereman MS, Petite JN, Hawkrigide AM, Muddiman DC. One-year plasma N-linked glycome intra-individual and inter-individual variability in the chicken model of spontaneous ovarian adenocarcinoma. *Int. J. Mass Spectrom.* 2011; 305(2 3):79–86. [PubMed: 21845070]
17. de Leoz MLA, Young LJT, An HJ, Kronewitter SR, Kim JH, Miyamoto S, Borowsky AD, Chew HK, Lebrilla CB. High-Mannose Glycans are Elevated during Breast Cancer Progression. *Mol. Cell Proteomics.* 2011; 10(1):1–9.
18. Snovidia SI, Perreault H. A 2,5-dihydroxybenzoic acid/N,N-dimethylaniline matrix for the analysis of oligosaccharides by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 2007; 21(22):3711–3715. [PubMed: 17952888]
19. Kaneshiro K, Watanabe M, Terasawa K, Uchimura H, Fukuyama Y, Iwamoto S, Sato TA, Shimizu K, Tsujimoto G, Tanaka K. Rapid Quantitative Profiling of N-Glycan by the Glycan-Labeling Method Using 3-Aminoquinoline/alpha-Cyano-4-hydroxycinnamic Acid. *Anal. Chem.* 2012; 84(16):7146–7151. [PubMed: 22830976]
20. Walker SH, Taylor AD, Muddiman DC. The Use of a Xylosylated Plant Glycoprotein as an Internal Standard Accounting for N-linked glycan Cleavage and Sample Preparation Variability. *Rapid Commun. Mass Spectrom.* 2013 In Press.
21. Bereman MS, Young DD, Deiters A, Muddiman DC. Development of a Robust and High Throughput Method for Profiling N-Linked Glycans Derived from Plasma Glycoproteins by NanoLC-FTICR Mass Spectrometry. *J. Proteome Res.* 2009; 8(7):3764–3770. [PubMed: 19435342]
22. Harris EK. Effects of Intraindividual and Interindividual Variation on Appropriate Use of Normal Ranges. *Clin. Chem.* 1974; 20(12):1535–1542. [PubMed: 4430131]
23. Hawkrigide AM, Muddiman DC. Mass Spectrometry-Based Biomarker Discovery: Toward a Global Proteome Index of Individuality. *Annual Review of Analytical Chemistry.* 2009:265–277.
24. Carvalho PC, Fischer JS, Chen EI, Yates JR, Barbosa VC. PatternLab for proteomics: a tool for differential shotgun proteomics. *BMC Bioinformatics.* 2008;9. [PubMed: 18182098]
25. Dong MQ, Venable JD, Au N, Xu T, Park SK, Cociorva D, Johnson JR, Dillin A, Yates JR. Quantitative mass spectrometry identifies insulin signaling targets in C-elegans. *Science.* 2007; 317(5838):660–663. [PubMed: 17673661]
26. Zybailov B, Mosley AL, Sardi ME, Coleman MK, Florens L, Washburn MP. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* 2006; 5(9):2339–2347. [PubMed: 16944946]

27. Florens L, Carozza MJ, Swanson SK, Fournier M, Coleman MK, Workman JL, Washburn MP. Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods*. 2006; 40(4):303–311. [PubMed: 17101441]
28. Sardiù ME, Cai Y, Jin JJ, Swanson SK, Conaway RC, Conaway JW, Florens L, Washburn MP. Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. *Proc. Natl. Acad. Sci. U.S.A.* 2008; 105(5):1454–1459. [PubMed: 18218781]
29. Gokce E, Shuford CM, Franck WL, Dean RA, Muddiman DC. Evaluation of Normalization Methods on GeLC-MS/MS Label-Free Spectral Counting Data to Correct for Variation during Proteomic Workflows. *J. Am. Soc. Mass Spectrom.* 2011; 22(12):2199–2208. [PubMed: 21952779]
30. Collier TS, Randall SM, Sarkar P, Rao BM, Dean RA, Muddiman DC. Comparison of stable-isotope labeling with amino acids in cell culture and spectral counting for relative quantification of protein expression. *Rapid Commun. Mass Spectrom.* 2011; 25(17):2524–2532. [PubMed: 21818813]
31. Walker SH, Papas BN, Comins DL, Muddiman DC. Interplay of Permanent Charge and Hydrophobicity in the Electrospray Ionization of Glycans. *Anal. Chem.* 2010; 82(15):6636–6642. [PubMed: 20590124]
32. Walker SH, Lilley LM, Enamorado MF, Comins DL, Muddiman DC. Hydrophobic Derivatization of N-linked Glycans for Increased Ion Abundance in Electrospray Ionization Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 2011; 22(8):1309–1317. [PubMed: 21953184]
33. Walker SH, Carlisle BC, Muddiman DC. Systematic Comparison of Reverse Phase and Hydrophilic Interaction Liquid Chromatography Platforms for the Analysis of N-Linked Glycans. *Anal. Chem.* 2012; 84(19):8198–8206. [PubMed: 22954204]
34. Andrews GL, Shuford CM, Burnett JC, Hawkrigide AM, Muddiman DC. Coupling of a vented column with splitless nanoRPLC-ESI-MS for the improved separation and detection of brain natriuretic peptide-32 and its proteolytic peptides. *J. Chromatogr. B.* 2009; 877(10):948–954.
35. Michalski A, Damoc E, Hauschild JP, Lange O, Wiegand A, Makarov A, Nagaraj N, Cox J, Mann M, Horning S. Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Mol. Cell. Proteomics.* 2011; 10(9)
36. Imre T, Schlosser G, Pocsfalvi G, Siciliano R, Molnar-Szollosi E, Kremmer T, Malorni A, Vekey K. Glycosylation site analysis of human alpha-1-acid glycoprotein (AGP) by capillary liquid chromatography-electrospray mass spectrometry. *J. Mass Spectrom.* 2005; 40(11):1472–1483. [PubMed: 16261636]
37. Fournier T, Medjoubi-N N, Porquet D. Alpha-1-acid glycoprotein. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* 2000; 1482(1–2):157–171.

Figure 1.

a–d) Theoretical isotopic distributions for light and heavy glycan pairs in a 1:1 ratio at 1000, 3000, 4000, and 7000 Da, respectively. The dotted lines represent the abundance of the light and heavy $[M+nH]^+$ peaks and show the discrepancy in quantification if isotopic distribution overlap is not taken into account. **e)** A plot of the theoretical A+6_L to A_L ratio vs. the molecular weight of the maltodextrin ladder.

Figure 2. Maltodextrin 1:1 (H:L) mixture. **a–c)** Raw data; **d–f)** abundance truncated and isotopic overlap corrected data; **g–i)** abundance truncated, isotopic overlap, and TGNF corrected data. For each of these situations, the histogram, molecular weight, and normalized abundance have been plotted.

Figure 3.

A plot of the experimental vs. the calculated H:L ratio for the 5:1, 2:1, 1.5:1, 1:1, 1:1.5, 1:2, and 1:5 maltodextrin mixtures. The data have been plotted for 3 glycans, each in triplicate, of substantially different molecular weights (902.35, 2522.88, and 4143.41 Da). A weighted linear least squares regression is plotted for the data.

Figure 4.

The experimental design for demonstrating the effectiveness of the INLIGHT strategy in pooled plasma. Black brackets denote samples that were mixed together, and samples labeled with a red 'L' and blue 'H' are derivatized with the light and heavy P2GPN reagents, respectively. Samples in (a) were used to determine the variability of the internal standard and TGNF correction. The spiked plasma aliquots (b) were used to demonstrate the measurement of changing glycan abundances using the INLIGHT strategy.

Figure 5. Histograms of the **a)** internal standard (HRP) and **b)** TGNF corrected neat pooled plasma 1:1 *N*-linked glycan samples. An f-test confirms that there is less variation in the data corrected using the TGNF strategy.

Abundance ratios for example glycan in both of the spiked plasma samples. The log₂(H:L) are statistically compared to the neat plasma samples using a t-test, and a |t|/t_{Table} > 1 indicates significant change. Additionally, those glycans in which a significant change was detected are highlighted to show both magnitude and direction of change.

Table 1

	H ₆ N ₅ F ₁ A ₂		H ₇ N ₆ A ₃		H ₇ N ₆ F ₁ A ₃		H ₈ N ₇ A ₃		Up-Regulated Heavy Sample
	log ₂ (H:L)	t /t _{Table}	log ₂ (H:L)	t /t _{Table}	log ₂ (H:L)	t /t _{Table}	log ₂ (H:L)	t /t _{Table}	
α ₁ -Spiked Light	-0.23	2.5	-0.64	9.0	-0.57	12	-0.63	3.7	
α ₁ -Spiked Heavy	0.30	2.5	0.79	1.4	0.71	1.03	0.87	1.7	
	H ₃ N ₅		H ₇ N ₂		H ₅ N ₄		H ₆ N ₅ F ₁		Up-Regulated Light Sample
	log ₂ (H:L)	t /t _{Table}	log ₂ (H:L)	t /t _{Table}	log ₂ (H:L)	t /t _{Table}	log ₂ (H:L)	t /t _{Table}	
α ₁ -Spiked Light	-0.05	0.04	-0.26	0.82	-0.09	0.39	-0.02	0.03	
α ₁ -Spiked Heavy	0.06	0.93	0.02	0.10	0.04	0.16	0.07	0.36	

Abbreviations: H – Hexose; N – N-acetylhexosamine; F – Fucose; A – N-acetylneuraminic acid.