⌘ Author's Choice

Mapping Yeast *N*-Glycosites with Isotopically Recoded Glycans*[S]

Mark A. Breidenbach‡§, Krishnan K. Palaniappan‡§, Austin A. Pitcher‡, and Carolyn R. Bertozzi‡¶‖****

Asparagine-linked glycosylation is a common post-translational modification of proteins; in addition to participating in key macromolecular interactions, *N*-glycans contribute to protein folding, trafficking, and stability. Despite their importance, few *N*-glycosites have been experimentally mapped in the *Saccharomyces cerevisiae* proteome. Factors including glycan heterogeneity, low abundance, and low occupancy can complicate site mapping. Here, we report a novel mass spectrometry-based strategy for detection of *N*-glycosites in the yeast proteome. Our method imparts *N*-glycopeptide mass envelopes with a pattern that is computationally distinguishable from background ions. Isotopic recoding is achieved via metabolic incorporation of a defined mixture of *N*-acetylglucosamine isotopologs into *N*-glycans. Peptides bearing the recoded envelopes are specifically targeted for fragmentation, facilitating high confidence site mapping. This strategy requires no chemical modification of the *N*-glycans or stringent sample enrichment. Further, enzymatically simplified *N*-glycans are preserved on peptides. Using this approach, we identify 133 *N*-glycosites spanning 58 proteins, nearly doubling the number of experimentally observed *N*-glycosites in the yeast proteome. ***Molecular & Cellular Proteomics 11: 10.1074/mcp.M111.015339, 1–10, 2012.***

Determining the functions of post-translational modifications (PTMs)[1] is a grand challenge in post-genomic biology. Locating modified sites is often an essential first step toward ascertaining the biological roles of PTMs, and liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) has emerged as a choice technology for this purpose. For example, LC-MS/MS has been successfully utilized in recent proteomic surveys of phosphorylation (1, 2), acetylation (3, 4), ubiquitination (5, 6), and glycosylation (7–10) sites. Modifications such as these can profoundly impact the molecular biology of a protein. Unambiguous assignment of PTM locations using existing LC-MS/MS methodology is sometimes difficult to achieve, particularly in cases involving low protein abundance or low site occupancy. Adding to the technical challenge of site mapping, PTMs such as glycosylation are known to reduce ionization efficiency and, therefore, detectability via MS (11–13). In highly complex proteomic samples, low intensity ions generated from glycopeptides may be overlooked because of instrument-dependent limitations on the rate at which ions can be selected for fragmentation during data-dependent acquisition (14). Here, we describe a targeted proteomic approach, in which LC-MS/MS analysis is focused specifically on peptides bearing *N*-glycosylation, a common PTM found in all major phylogenetic branches of life. In contrast to data-dependent methods that select a subset of relatively intense ions for fragmentation, our approach is specifically designed to provide intensity-independent fragmentation priority to peptides most likely to bear *N*-glycans and boost confidence in correct PTM site assignment.

In eukaryotes, *N*-glycosylation is a PTM frequently found on proteins that are translated into the endoplasmic reticulum (ER). Structurally conserved *N*-glycan precursors are synthesized via the dolichol pathway and transferred onto nascent proteins; the glycosylated Asn residues are typically found within a subset of the polypeptide N*X*(S/T) motifs, where *X* denotes a non-proline residue (15, 16). Biologically, *N*-glycans contribute to the folding, trafficking, and thermodynamic stability of proteins exposed to ER, vacuolar, and Golgi lumens, as well as those destined for exposure to the extracellular milieu (17). The *N*-glycan precursors are enzymatically edited to yield a plethora of mature glycoforms with compositions largely dependent on cell type and protein localization. Sometimes, specific *N*-glycan structures are required for protein function, whereas in other cases, *N*-glycans primarily contribute to protein stability and solubility (18). As a class, the *N*-glycosylation modification is biologically essential; chemical or genetic disruption of *N*-glycan biosynthesis is lethal, and aberrant *N*-glycosylation is associated with several human disease states (19, 20).

From the Departments of ‡Chemistry and ¶Molecular and Cell Biology and the ‖Howard Hughes Medical Institute, University of California, Berkeley, California 94720

Received October 26, 2011, and in revised form, January 15, 2012
⌘ *Author's Choice*—Final version full access.
Published, MCP Papers in Press, January 19, 2012, DOI 10.1074/mcp.M111.015339

[1] The abbreviations used are: PTM, post-translational modification; LC-MS/MS, liquid chromatography coupled to tandem mass spectrometry; ER, endoplasmic reticulum; PNGase F, peptide:*N*-glycosidase F; IW, isolation window; CID, collision-induced dissociation; GlcNAc, *N*-acetylglucosamine; UDP-GlcNAc, uridine diphosphate-GlcNAc.

Many existing LC-MS/MS approaches for mapping *N*-glycosites depend on enzymatic removal of the entire *N*-glycan following stringent sample enrichment to remove nonglycosylated peptides prior to analysis (9, 10). These methods rely on detection of a 0.98-Da mass increase resulting from the enzymatic deamination of glycosylated Asn residues by peptide:*N*-glycosidase F (PNGase F). Enzymatic deamination is often performed in the presence of ^{18}O-labeled water, imparting a 2.98-Da mass shift to the peptide to increase confidence in site assignment (7). Unfortunately, complete removal of *N*-glycans with PNGase F can lead to instances of incorrectly mapped glycosites. During the course of PNGase F treatment, spontaneous deamination of nonglycosylated Asn residues and other instances of PNGase F-independent incorporation of ^{18}O can potentially yield false positives (21–23). This drawback has led to the development of alternative strategies utilizing partial rather than total removal of *N*-glycans; for instance, treating samples with the enzyme endoglycosidase H preserves a single core GlcNAc residue, leaving direct evidence for *N*-glycosylation intact (23, 24). Unfortunately, the presence of even a single sugar residue on peptides is known to considerably suppress ionization efficiency (8, 13), potentially biasing data-dependent LC-MS/MS data acquisition against glycopeptide ions. Despite this limitation, detection of the retained glycan by LC-MS/MS provides unequivocal evidence for glycosite assignment instead of indirect evidence that the *N*-glycan modification once existed at a given site. Therefore, LC-MS/MS methods that could select low abundance ions for fragmentation could greatly benefit glycosite mapping.

Recently, we described a strategy for addressing the challenge of identifying low abundance species in complex mixtures based on a directed LC-MS/MS approach (25). Termed isotopic signature transfer and mass pattern prediction (IsoStamp), the technique exploits the perturbing effects of a dibrominated chemical tag on the isotopic envelope of a peptide. Once covalently modified with the tag, dibrominated peptides can be readily detected with high sensitivity and fidelity in high resolution LC-MS data using a novel computational pattern-searching algorithm. Pattern identification allows targeted proteomic analyses of labeled peptides in complex biological samples; inclusion lists containing *m/z* values and retention times of ions bearing recoded envelopes are used to trigger fragmentation. Thus, isotopic pattern rather than ion abundance is used to drive this directed proteomic approach. The IsoStamp method was an extension of the concept of isotopic distribution encoding tagging reported by Goodlett *et al.* (26).

Although the isotopic signature of a halogenated tag can effectively highlight labeled peptides within a complex LC-MS data set, its utility is restricted to those situations in which the desired subset of peptides can be chemospecifically modified. Here, we demonstrate that it is possible to impart a similar perturbation to the isotopic envelope of a peptide without the requirement for chemical tagging. Instead, our approach metabolically embeds a dibromide-like isotopic signature directly into glycans. In this study, we mimic the dibromide isotopic signature with a stoichiometrically defined mixture of GlcNAc isotopologs, referred to as a GlcNAc isomix. The isomix is metabolically installed into structurally conserved *N*-glycan core positions, marking them with a uniquely identifiable isotopic signature. We employed the technique to map occupied *N*-glycosites on proteins from whole *Saccharomyces cerevisiae* lysates. Via preferential fragmentation of isotopically recoded glycopeptides, we identified numerous *N*-glycosites within the yeast proteome, nearly doubling the number that was previously known. The isomix method offers an enhanced level of confidence for mapping glycosylation sites that was not previously available to LC-MS/MS analyses because of the unique isotopic envelope of an isomix-containing peptide. Here, we showcase the utility of isotopic recoding by surveying metabolically labeled *N*-glycans in yeast, but we believe the technology is extendable to other PTMs and organisms through a variety of labeling strategies.

EXPERIMENTAL PROCEDURES

Synthesis of GlcNAc Isotopologs—The synthesis of *N*-acetyl-D-glucosamine from D-glucosamine hydrochloride was performed in a single synthetic step as previously described (27). To a solution of the D-glucosamine hydrochloride salt (150 mg, 0.6 mmol) in water was added Dowex 200–400 mesh (OH^-) anion exchange resin, and the pH was adjusted to 7.5. The resin was removed by filtration, yielding a D-glucosamine solution in 6 ml of water. To this solution was added [$^{13}C_2$]sodium acetate (1.1 eq, Cambridge Isotope Laboratories) as a predissolved solution in water and 2-ethoxy-1-ethoxycarbonyl-1,2-dihydroquinoline (1.1 eq) as a predissolved solution in 10 ml of ethanol. The total reaction volume was brought to 40 ml by the addition of ethanol. The reaction was covered in foil and stirred for 36 h at room temperature. This reaction was repeated using 1-[^{13}C,^{15}N]D-glucosamine hydrochloride (Isotec) as the starting material. After ethanol evaporation under reduced pressure, the crude products were purified by silica gel chromatography using a 10:4:3 mixture of ethyl acetate:pyridine:water, dried, filtered, lyophilized, and stored at −20 °C.

Preparation of GlcNAc Isomix—Stock solutions of 10 mM *N*-acetyl-D-glucosamine, *N*-[1,2-$^{13}C_2$]acetyl-D-glucosamine, and *N*-[1,2-$^{13}C_2$]acetyl-D-[1-^{13}C;^{15}N]glucosamine in water were prepared. The solutions were combined to produce a mixture of the GlcNAc isotopologs at a 1:2:1 molar ratio, respectively. The mixture was analyzed by direct infusion on a Thermo-Finnigan LTQ-XL mass spectrometer set to zoom scan, with the signal averaged over 20 scans. The isotopic ratios were then adjusted by the iterative addition of the desired isotopes until a nearly perfect 1:2:1 peak intensity ratio was observed empirically (see Fig. 1*b*). The isomix sample was lyophilized and stored at −20 °C.

Cell Culture and Sample Preparation—Detailed descriptions of cell culture conditions and sample preparation are included in the supplemental information. Briefly, lysates from log phase and stationary phase GlcNAc isomix-supplemented yeast cultures were prepared for MS analysis via a modified version of the previously described filter-aided sample preparation methodology (10, 28).

Mass Spectrometric Analysis—Prior to analysis, the samples were resuspended in 50 μl of water, and 2.5–5 μl aliquots were subjected to reverse phase liquid chromatography with an Agilent 1200 LC

system connected in-line to a LTQ Orbitrap XL hybrid mass spectrometer. External mass calibration was performed prior to analysis. A binary solvent system consisting of buffer A (0.1% formic acid in water (v/v)) and buffer B (0.1% formic acid in acetonitrile (v/v)) was employed. The mass spectrometer was outfitted with a nanospray ionization source. The LC was performed using a 100-μm fritted capillary (New Objective) precolumn self-packed with 1 cm of 5 μm, 200 Å Magic C18AQ resin (Michrom Bioresources) followed by a 100-μm fused silica capillary (Polymicro Technologies) self-packed with 15 cm of 5 μm, 100 Å Magic C18AQ resin (Michrom Bioresources). After sample injection and a 20 min loading step in 2% buffer B, peptides were eluted using a gradient from 7% to 35% buffer B over 150 min, followed by a washing step in 99% buffer B for 20 min. A solvent split was used to maintain a flow rate of 400 nl min^{-1} at the column tip.

The samples were subjected to an inclusion list-driven targeted acquisition method. First, we collected only full scan mass spectra in positive ion mode over the *m/z* scan range of 400–1,700 or 400–1,800 using the Orbitrap mass analyzer in profile mode at a resolution of 60,000 (at 400 *m/z*). Noise reduction and peak detection were performed using software developed in-house making use of a continuous wavelet transform (25, 29). The centroided mzXML files were then pattern searched for recoded envelopes (25). The output was an inclusion list that contained the *m/z* value (M+2 ion in the isotopic envelope of the labeled peptide) and a retention time window (± 1.5 min, empirically determined) for each labeled peptide. The same sample, stored at 4 °C, was then reanalyzed with identical chromatographic conditions using an inclusion list-driven selection of precursor ions for fragmentation. In relatively rare instances we observed a chromatographic anomaly causing substantially shifted peptide elution profiles; in these cases the experiment was repeated. For each full scan mass spectrum, up to eight CID fragmentation events were performed in the linear ion trap. Dynamic exclusion and charge state screening were enabled to reject ions with an unknown or +1 charge state. An isolation window (IW) of 2 or 4 Da, a minimum threshold of 500 ion counts, and activation energy of 35 were used when triggering a fragmentation event.

Database Searches—Peptide identities were obtained using the SEQUEST search algorithm (30) within Proteome Discoverer 1.2 (Thermo-Fisher). CID spectra were searched against the sequence database generated from all systematically named *S. cerevisiae* ORFs (downloaded March 2011) (31) augmented with sequences from the common repository of adventitious proteins (cRAP, from the Global Proteome Machine Organization, downloaded March 2011) and the Ngt1 (*C. albicans*) and NAGK (*H. sapiens*) protein sequences, totaling 6,739 entries. Indexed databases for tryptic and chymotryptic digests were created allowing for two missed cleavages, one nonenzymatic terminus, one fixed modification (cysteine carboxyamidomethylation, +57.021 Da) and the following variable modifications: methionine oxidation (+15.995 Da) and asparagine GlcNAcylation (only the M + 2 isotopolog, +205.086 Da). Precursor ion tolerance was set to 10 ppm, and the CID fragment tolerance was set to 0.8 Da. The criteria used for filtering search results included using SEQUEST score function (*XCorr*) cutoffs assigned by a 5% maximum false discovery rate obtained from decoy database searches (using reversed sequences from each of the 6,739 entries in our database), an 8-ppm maximum allowed precursor ion mass deviation, and a 40-amino acid maximum allowed peptide length. Approximately 70% of the peptides reported here were identified within a 1% maximum false discovery rate. Importantly, all of the precursor ion and fragmentation spectra were visually inspected for the isomix signature by authors M. A. B. and K. K. P.; only peptides agreed upon by both authors were considered valid identifications. Precursor ion isomix signatures were also visually verified in the corresponding full scan data sets. In cases where the isomix signature was identifiable in the low resolution fragmentation spectra, this additional information was used to verify correct glycosite assignment, especially in peptides containing multiple Asn residues. For data sets collected with an IW of 2, we used the full scan only data set (*i.e.* the LC-MS run, which was subjected to pattern searching) to confirm the isomix signature. In cases where multiple peptide sequences covered a single glycosite, a representative peptide sequence was selected based on CID spectral quality. Factors such as protein biological function and the presence of a canonical *N*-glycosylation consensus motif (N*X*(S/T)) were not used as criteria for accepting or rejecting peptides. Instances of ambiguous PTM assignment within a peptide were resolved by manual inspection of fragment envelopes as necessary. In one case, highlighted in supplemental Fig. S4, the exact site could not be resolved among two possibilities. The primary data associated with this manuscript may be downloaded from the Tranche repository (https://proteomecommons.org/tranche/); hash codes are available in the supplemental information.

RESULTS

Metabolic Recoding of N-Glycan Cores—Because of the relative ease with which certain *S. cerev*isiae biosynthetic pathways can be manipulated, we chose to metabolically install an unnatural, dibromide-like isotopic signature into glycans. Universal isotopic recoding of *N*-glycans requires metabolic replacement of a conserved sugar residue found in all such structures with a mixture of isotopologs. All eukaryotic *N*-glycans possess a conserved GlcNAcβ1,4GlcNAc disaccharide at the peptide-proximal position; thus, a GlcNAc isomix can potentially label all *N*-glycans in the glycoproteome of the cell. However, to retain the isotopic ratio during metabolism, the GlcNAc isomix must be converted without isotopic dilution to the key metabolic intermediate uridine diphosphate-GlcNAc (UDP-GlcNAc), the donor nucleotide sugar used in construction of *N*-glycan cores. Specifically, cytosolic UDP-GlcNAc serves as the donor substrate for Alg7 in the biosynthesis of GlcNAc-diphosphodolichol; the GlcNAc residue in this precursor is ultimately covalently linked to Asn in the process of *N*-glycosylation (32). Thus, control over the uridine diphosphate-GlcNAc (UDP-GlcNAc) pool of a cell is critical for execution of the isomix method. We recently generated a *S. cerevisiae* strain that depends on an engineered salvage pathway for procuring precursors of UDP-GlcNAc (33). The *gna1*Δ yeast strain lacks the ability to perform *de novo* UDP-GlcNAc biosynthesis and instead generates UDP-GlcNAc exclusively by salvaging GlcNAc added to the culture media. In previous work, we exploited this yeast strain to achieve high efficiency replacement of GlcNAc residues with unnatural GlcNAc analogs, which are alternative substrates for the engineered salvage pathway (33).

In this work, we supplemented cultured *gna1*Δ yeast with a GlcNAc isomix designed to emulate the isotopic signature of two bromine atoms. As shown in Fig. 1*a*, the dibromide pattern is a symmetrical triplet, with major peaks at *M*, *M* + 2, and *M* + 4 at a relative intensity of 1:2:1, because of the relative abundances of the $^{79}Br_2$, $^{79}Br^{81}Br$, and $^{81}Br_2$ isotopic pairings. To replicate this pattern, a three-part isomix consisting of *N*-acetyl-

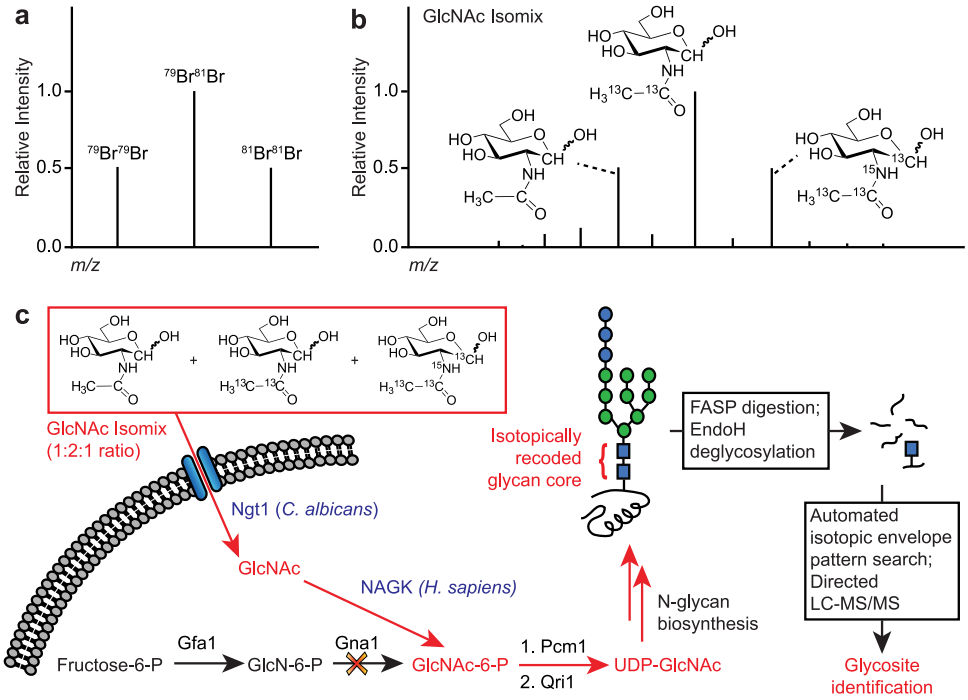

FIG. 1. **Metabolic incorporation of a GlcNAc isomix into yeast *N*-glycans.** *a*, the dibromide triplet pattern, with a 1:2:1 relative peak intensity distribution, results from the natural abundances of ^{79}Br and ^{81}Br isotope pairings. *b*, a three-component GlcNAc isomix mimics the 1:2:1 peak intensity distribution of dibromide by adjusting the concentration of each synthetically made isotopolog. *c*, the GlcNAc isomix enters the *gna1*Δ yeast hexosamine biosynthetic pathway via a heterologous salvage pathway (33). The isomix signature is subsequently embedded into UDP-GlcNAc and any glycoconjugates that utilize UDP-GlcNAc in their construction, including the structurally conserved cores of *N*-glycans. Following cell lysis, proteolysis, partial enrichment of *N*-glycopeptides, and partial deglycosylation with endoglycosidase H, the distinctive isotopic signatures of *N*-glycopeptides are detected computationally using pattern-matching software (25). Masses of putative *N*-glycopeptide ions are granted fragmentation priority in subsequent LC-MS/MS analyses for *N*-glycosite identification. The *N*-glycan precursor illustrated here is composed of two core GlcNAc residues (*blue squares*), nine mannose residues (*green circles*), and three glucose residues (*blue circles*).

D-glucosamine, *N*-[1,2-$^{13}C_2$]acetyl-D-glucosamine, and *N*-[1,2-$^{13}C_2$]acetyl-D-[1-^{13}C;^{15}N]glucosamine, mixed in a 1:2:1 molar ratio (Fig. 1*b*), was prepared and added to the *gna1*Δ yeast culture medium (Fig. 1*c*). The isomix is internalized via the GlcNAc-specific transporter Ngt1 borrowed from *Candida albicans*, and cytosolic GlcNAc is subsequently phosphorylated by the heterologously expressed human kinase NAGK. The resulting GlcNAc-6-phosphate isomix is converted to a UDP-GlcNAc isomix and subsequently installed into *N*-glycan cores by endogenous yeast machinery.

Directed Proteomic Analysis of N-Glycosylated Peptides—Samples for LC-MS/MS analysis were generated from lysates of *gna1*Δ *S. cerevisiae* cultures grown to both mid-log and stationary phases in chemically defined, minimal medium. Tryptic and chymotryptic peptides were prepared from the lysates using a modified version of filter-aided sample preparation (10, 28). During this process, mannose-containing glycopeptides were partially enriched by binding to the lectin concanavalin A, and *N*-glycans were truncated to a single GlcNAc residue with the glycosidase endoglycosidase H. To achieve sufficient resolution for isotopic envelope pattern matching, peptides were analyzed on a Thermo-Finnigan LTQ Orbitrap XL mass spectrometer.

Recoded peptide envelopes bear distinctive peak intensity distributions, as illustrated in the simulated isotopic envelopes shown in Fig. 2*a*. The envelope of an isomix-labeled glycopeptide reflects the peptide's intrinsic envelope convolved with that of the GlcNAc isomix. The unique pattern resulting from this convolution serves as the basis for detecting putative glycopeptides in complex samples. Importantly, experimentally observed isomix-containing glycopeptide envelopes, as exemplified in Fig. 2*b*, matched our expectations based on simulations (Fig. 2*a*). This allowed us to computationally search (using the IsoStamp algorithm (25)) for the characteristic isomix signature in yeast peptides subjected to LC-MS analysis. Typically, we detected 4,000–5,000 candidate glycopeptide ions (in charge states $z = 1$ to $z = 5$) per sample. However, we did not expect all candidate precursors targeted for fragmentation to yield unique, high confidence glycosites. In some cases, high sample complexity contributed to co-eluting ions of similar masses and identical charge states, generating false positive identifications during the automated pattern search routine (25). Additionally, the precursor inclusion list inherently contains redundancy. For example, several legitimate, isomix-bearing glycopeptides were observed in multiple charge states between $z = 1$ and $z = 5$.

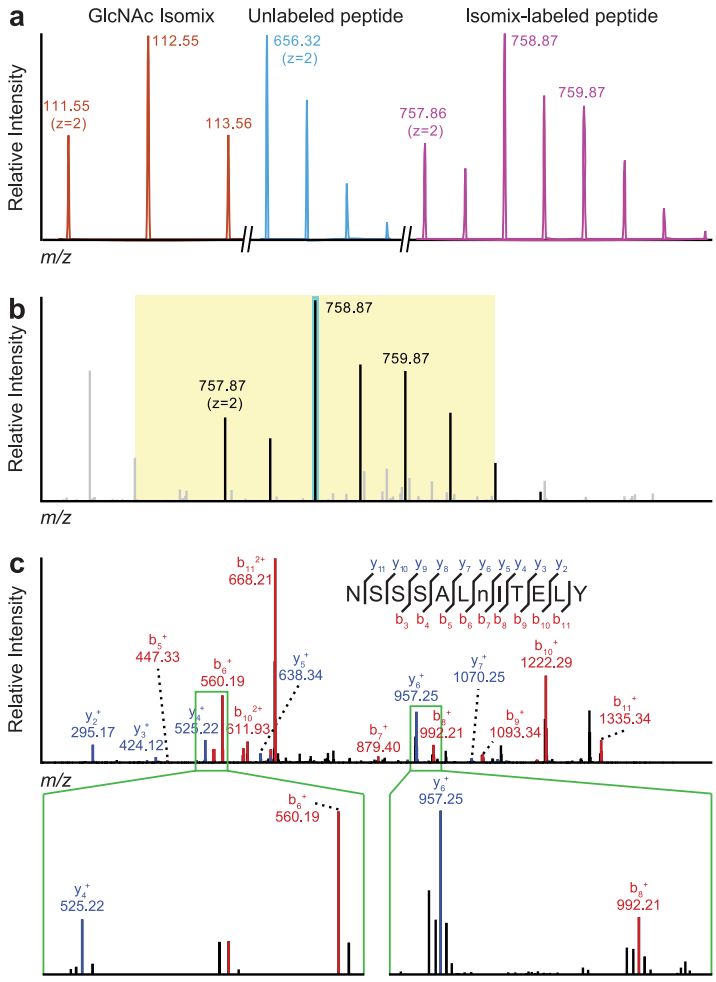

FIG. 2. **The perturbing effect of a GlcNAc isomix on a peptide isotopic envelope.** *a*, simulated isotopic envelopes ($z = 2$) for the GlcNAc isomix (*red trace*), an unlabeled peptide from the glycoprotein Ygp1 (NSSSALNITELY, *blue trace*), and the same peptide labeled with the GlcNAc isomix (*violet trace*). The isotopically recoded peptide has a visually distinctive distribution of peak intensities. *b*, in experimental LC-MS data, the isotopic envelope of the precursor ion corresponding to the NSSSALnITELY glycopeptide is shown (the modified Asn residue is shown in *lowercase*). *Highlighted* in *teal* is the precursor ion that was selected from the inclusion list, and the 4-Da isolation window used for fragmentation is shown in *yellow*. *c*, the CID fragmentation spectra and the peptide assignment for the 758.87 ion (lowercase *n* refers to the *N*-glycosite). Fragment ions that lack the GlcNAc isomix (such as the y_4^+ and b_6^+ fragment ions) have narrow isotopic envelopes, whereas fragments including the GlcNAc isomix (such as the y_6^+ and b_8^+ fragment ions) show a perturbed isotopic envelope characteristic of the isomix signature.

Also, we frequently observed more than one peptide covering the same glycosite. The combination of these factors accounts for the discrepancy between putative glycopeptide precursor ions and the high confidence glycosites we report here. Regardless, the *m/z* values and retention times of putative glycopeptide ions bearing the isomix signature were used to construct a time-resolved inclusion list for targeted fragmentation. Notably, even with the inclusion of a lectin-based glycopeptide enrichment step during sample preparation, the overwhelming majority of high intensity ions in our samples appeared to be unglycosylated, as indicated by visual and computational inspection of their isotopic envelopes.

Our directed proteomics approach (see supplemental Fig. S1) (25) required duplicate, back-to-back injections for targeted fragmentation of glycopeptides. The first injection was used exclusively to search for GlcNAc isomix signatures in LC-MS data to identify likely glycopeptide ions and inventory them into an inclusion list. In the subsequent injection, candidate glycopeptide ions were subjected to fragmentation by collision-induced dissociation (CID) (34) only if the ion abundance exceeded a defined threshold and the ions appeared in the correct retention time window. Importantly, the Asn-GlcNAc linkage is resistant to standard CID conditions used to fragment the peptide backbone (35).

Ions bearing the isomix signature were subjected to CID fragmentation to determine peptide identity and verify *N*-glycosylation. Ions selected for fragmentation were isolated in either a narrow (2-Da) or broad (4-Da) IW. Although a broad IW covers most peaks in an isomix-labeled glycopeptide envelope (Fig. 2*b*), it has the disadvantage of potentially including unrelated ions in the fragmentation spectrum. Conversely, the use of narrow IWs could yield higher quality fragmentation spectra because fewer unrelated ions will be isolated. However, in practice, the use of a 4-Da IW results in the highest number of MS/MS identifications despite the potential decrease in the precursor ion fraction isolated (36). Indeed, we observed no detrimental effects on peptide assignments in data sets collected using 4-Da *versus* 2-Da IWs. Instead, the use of a broad IW on ions bearing the isomix signature often provided additional information in the fragmentation spectra that was critical to resolving glycosite positional ambiguity. The broad IW was particularly useful for glycosite assignment in the case of peptides containing more than one Asn residue. As illustrated in Fig. 2 (*b* and *c*), when the full isotopic enve-

lope of a candidate glycopeptide was isolated for fragmentation using a broad IW, we preserved the isomix signature in all peptide fragments bearing the GlcNAc isomix modification. In contrast, unglycosylated fragments yielded unperturbed mass envelopes. This information was used to resolve cases of ambiguous glycosite assignments that resulted from automated database search engines. In some cases, doubly glycosylated peptides were observed; the presence of two *N*-glycosites on one peptide introduced a doubly convolved isomix signature, resulting in a widened isotopic envelope with a distinctive "stairstep" pattern in peak intensities (supplemental Fig. S2).

Isomix-directed Fragmentation Allows High Confidence N-Glycosite Identification—The unified list of high confidence *N*-glycosites in the yeast proteome (summarized in Table I) reflects a combination of results from multiple experiments, conducted with trypsinized and chymotrypsinized samples obtained from both stationary and mid-log phase cultures. A complete list of the observed tryptic and chymotryptic peptides containing these glycosites and their corresponding statistical indicators of quality are included in supplemental Table S1. We detected a total of 133 *N*-glycosites, 12 of which were previously reported in the Uniprot Knowledgebase (37). Taken collectively with glycosite information from the Uniprot database, there are now 196 experimentally mapped *N*-glycosites spanning 72 proteins in *S. cerevisiae*; 121 novel sites distributed over 52 proteins are reported here. Importantly, 50% of the proteins in Table I, highlighted in bold type, have been biochemically validated as *N*-glycosylated. Proteins were considered biochemically validated if they had previously mapped *N*-glycosites or if there was experimental observation of a gel shift following enzymatic deglycosylation (33, 37–39). The high correlation between our proteomic observations and the plethora of biochemical validation data for yeast serves as an excellent indicator that the isomix technique yielded a reliable list of *N*-glycoproteins. Moreover, our results suggest that the *gna1*Δ *S. cerevisiae* likely share similar *N*-glycosite occupancy patterns with the BY4743 strain (40) from which the *gna1*Δ haploid was derived.

For comparative purposes, we also subjected our samples to traditional intensity-driven data-dependent LC-MS/MS, in which the 10 most intense ions in each full scan mass spectrum were selected for CID fragmentation. Significantly, the intensity-dependent analyses typically identified ~60% of the glycosites found via isomix-directed fragmentation. For instance, in the case of trypsinized lysate prepared from a stationary phase culture, 52 unique glycosites spanning 25 proteins were identified using directed fragmentation. The same sample, subjected to data-dependent fragmentation, yielded only 30 *N*-glycosites in 16 proteins and contained extensive site overlap with the directed data set. This difference illustrates the advantages of directed, intensity-independent approaches for proteomic analyses. Despite being outperformed by a directed proteomic analysis here, the widely used data-dependent fragmentation approach remains a powerful method, and its performance would undoubtedly benefit from more stringent glycopeptide sample enrichment and improved chromatographic resolution of peptides.

Glycoproteins identified in this study were categorized by the manually curated ontological annotations maintained by the *Saccharomyces* Genome Database (31). An analysis of cellular localization (Fig. 3*a*) reveals that the majority of the *N*-glycoproteins identified in Table I typically reside in the yeast ER, plasma membrane, vacuole, and cell wall. Several proteins localize to substructures within these organelles, especially those involved with cytokinesis and cellular budding. Although many of the identified glycoproteins lack known molecular functions, our list was highly represented by peptidases, glycosidases, and glycosyltransferases (Fig. 3*b*). With respect to biological function, highly represented categories include vacuolar proteases and proteins involved in cell wall construction, reshaping, and maintenance (Fig. 3*c*). Several ER- and Golgi-resident proteins participating in the processes of *N*- and *O*-linked glycosylation were also detected. Overall, the ontological distribution of glycoproteins reported here is consistent with our expectations for *N*-glycoproteins in yeast.

In addition to the ontological consistency for our list of *N*-glycoproteins, we detected no plausible cases in which the *N*-glycan modification was present outside the canonical N*X*(S/T) motif. Although noncanonical *N*-glycosites have been reported in higher eukaryotes (10, 41), it is possible that the acceptor substrate promiscuity of the oligosaccharyl transferase complex, which catalyzes the *N*-glycosylation modification, differs among species. A plot of the relative residue frequencies for the 133 occupied glycosites mapped here reveals a strong, 71% *versus* 29% preference for Thr over Ser within the canonical motif (Fig. 4), consistent with *N*-glycosites mapped in other eukaryotic proteins by nonproteomic methods (42). Additionally, positions surrounding the motif are dominated by hydroxylated and small hydrophobic residues, consistent with expectations for solvent-exposed, loop regions of proteins. Charged residues, especially aspartate, are present in slightly higher frequency on the C-terminal side of the sequon, but not at the −2 position as observed in the case of bacterial *N*-glycosites (43). Collectively, these data reveal a typical acceptor substrate sequence for the yeast oligosaccharyl transferase.

DISCUSSION

We have demonstrated that a metabolically embedded isotopic signature, designed to emulate the perturbing effects of a dibrominated chemical tag, can serve as the basis for targeted proteomic analysis of a biologically important class of PTM. This approach has two major benefits. First, the ions selected for fragmentation need not be among the most intense in the sample. Second, recoded envelopes contain easily detectable cues indicating whether or not fragmentation spectra have been correctly assigned to a peptide, a

TABLE I

Occupied high confidence N-glycosites detected in the S. cerevisiae proteome

Peptides covering these glycosites are detailed in supplemental Table I. Biochemically validated *N*-glycoproteins are highlighted with bold type (33, 37–39). Previously mapped *N*-glycosites are underlined (37).

Systematic name	Uniprot identification code	Standard name	Occupied *N*-glycosites
YBR023C	P29465	CHS3	Asn-323
YBR074W	P38244		Asn-121
YBR078W	P38248	ECM33	Asn-209, 267, 279, 328
YBR092C	**P24031**	**PHO3**	**Asn-162, 315, 356, 390**[a]
YBR093C	P00635	PHO5	Asn-315, 356[a], 390[a]
YBR139W	P38109		Asn-339
YBR162C	P38288	TOS1	Asn-236
YBR171W	P33754	SEC66	Asn-12
YBR229C	P38138	ROT2	Asn-173, 907
YBR286W	**P37302**	**APE3**	**Asn-85, 96, 115, 150, 162, 371, 427**
YCL043C	**P17967**	**PDI1**	**Asn-117, 155, 425**
YCL045C	P25574	EMC1	Asn-106, 420
YCR011C	**P25371**	**ADP1**	**Asn-165**
YDL095W	P33775	PMT1	Asn-390
YDR055W	**Q12355**	**PST1**	**Asn-210, 242**
YDR434W	**Q04080**	**GPI17**	**Asn-299**
YEL001C	**P40006**	**IRC22**	**Asn-93, 121**
YEL002C	P33767	WBP1	Asn-60, 332
YEL040W	P32623	UTR2	Asn-261
YEL060C	P09232	PRB1	Asn-594
YFR018C	**P43599**		**Asn-179**
YGL022W	P39007	STT3	Asn-539
YGR014W	P32334	MSB2	Asn-1088
YGR189C	P53301	CRH1	Asn-177, 201
YGR282C	**P15703**	**BGL2**	**Asn-202**
YHR028C	**P18962**	**DAP2**	**Asn-63, 139**
YHR101C	**P38813**	**BIG1**	**Asn-144**
YHR188C	**P38875**	**GPI16**	**Asn-184**
YHR202W	**P38887**		**Asn-305**
YIL015W	P12630	BAR1	Asn-268, 366, 398
YJL002C	**P41543**	**OST1**	**Asn-99, 400**
YJL171C	P46992		Asn-51, 99, 122, 174, 219, 249, 267, 300
YJL172W	P27614	CPS1	Asn-176, 228
YJL192C	P39543	SOP4	Asn-35
YKL039W	P32857	PTM1	Asn-132
YKL046C	**P36091**	**DCW1**	**Asn-203, 240**
YLR066W	Q12133	SPC3	Asn-173
YLR084C	**Q12465**	**RAX2**	**Asn-333, 524, 640**
YLR413W	**Q06689**		**Asn-194, 299**
YLR450W	P12684	HMG2	Asn-150
YML052W	P54003	SUR7	Asn-47
YML130C	**Q03103**	**ERO1**	**Asn-53, 468**
YMR006C	Q03674	PLB2	Asn-80, 193, 491
YMR008C	**P39105**	**PLB1**	**Asn-78, 123, 170, 388, 513, 541**
YMR058W	**P38993**	**FET3**	**Asn-244, 265, 292, 359, 381**
YMR200W	**Q03691**	**ROT1**	**Asn-139**
YMR238W	**Q05031**	**DFG5**	**Asn-245**
YMR297W	**P00729**	**PRC1**	**Asn-124, 479**
YMR307W	**P22146**	**GAS1**	**Asn-40, 57, 95, 149, 165, 253, 409**
YNL160W	**P38616**	**YGP1**	**Asn-58/61, 87, 94, 100, 106, 118, 172, 239, 286**
YNL327W	**P42835**	**EGT2**	**Asn-103, 161**
YNR067C	**P53753**	**DSE4**	**Asn-886**
YOL011W	**Q08108**	**PLB3**	**Asn-56, 82, 547**
YOL030W	Q08193	GAS5	Asn-24, 60, 166, 299, 344
YOL154W	Q12512	ZPS1	Asn-28, 57, 217
YOR320C	Q12096	GNT1	Asn-425
YPL123C	Q02933	RNY1	Asn-37
YPL154C	**P07267**	**PEP4**	**Asn-144, 345**

[a] Sequence ambiguity between PHO3 and PHO5.

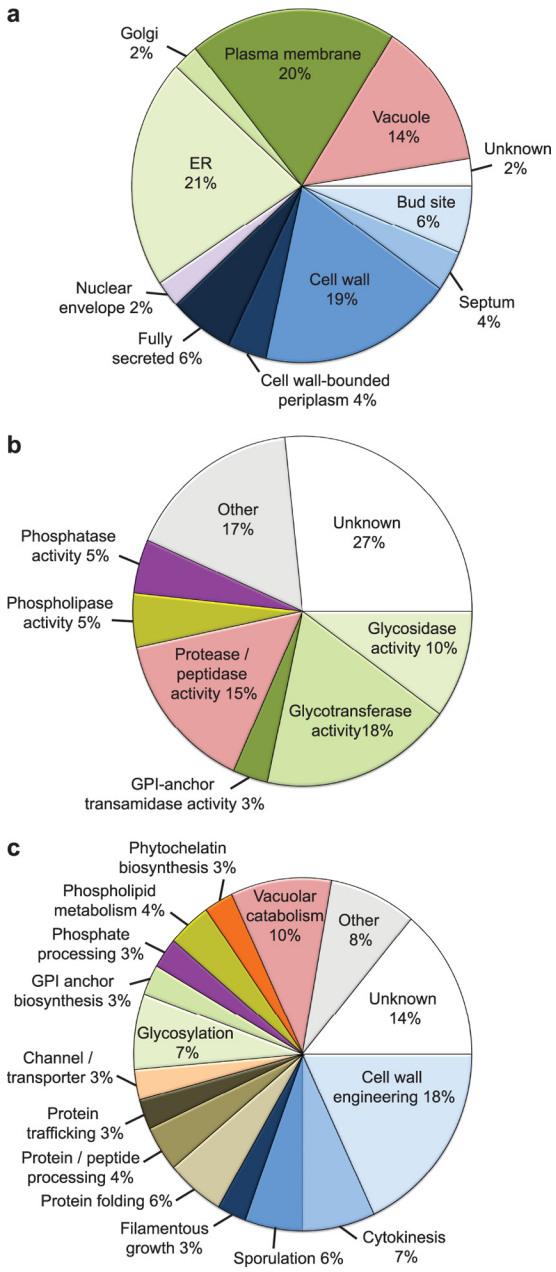

FIG. 3. **Ontological analysis of high confidence *N*-glycoproteins.** *N*-Glycoproteins were categorized according to the manually curated ontological annotations including cellular component (*a*), molecular function (*b*), and biological process (*c*) maintained by the *Saccharomyces* Genome Database (31).

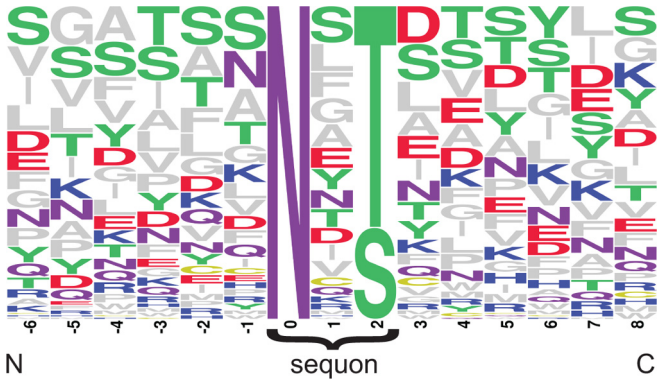

FIG. 4. **Relative frequencies of residues surrounding yeast *N*-glycosites.** Sequences of the 133 unique glycosylation sites detected by directed LC-MS/MS are aligned on the modified Asn residue. Relative heights of the surrounding amino acids are adjusted based on the frequencies of their occurrence.

feature that is particularly useful for verifying the site of a PTM. We used the approach of isomix-based targeted proteomics to build a relatively small but high confidence list of *N*-glycosites in yeast.

Systematically mapping *N*-glycosylation sites in *S. cerevisiae* provides a wealth of information pertaining to native glycoprotein structure and topology. It is well known that the mere presence of a standard N*X*(S/T) sequon is not sufficient to guarantee glycan attachment to a nascent polypeptide. A number of extenuating factors, ranging from local protein structure to nutrient availability, ultimately determine which sequons are glycosylated. In the case of soluble proteins, fully or partially occupied glycosites are an excellent indicator of sequon surface accessibility. In the case of integral membrane proteins, occupied glycosites confirm that a given Asn is ER-lumenal upon translation and can be used to confirm or reject integral membrane protein topological predictions. In our data, we detected 25 glycosylated integral membrane proteins, and in all cases the occupied glycosites confirm transmembrane hidden Markov model topology predictions (44) of lumenal *versus* extracellular domains. Additionally, glycosite mapping is a prerequisite for quantitative, comparative analysis of glycan occupancy. Although we do not attempt to quantify glycan occupancy in this study, a high confidence glycosite map serves as the foundation for investigating how glycosite occupancy changes with cell state and growth conditions; such a study can be conducted using established quantitative MS approaches focused on newly verified glycosites and is a subject of interest for future investigations.

Although LC-MS/MS is currently among the best available methods for mapping PTMs such as glycosylation, obtaining high confidence correlations between fragmentation spectra and peptide sequences is not always simple (45). The utility of tandem MS analysis is ultimately limited by the quality of fragmentation spectra. Because fragmentation spectra are rarely perfect, statistical methods for increasing confidence in peptide assignment remains an active area in research in the field of proteomics (46–48). Advances in instrumentation available to the proteomics community are also reducing the possibility of incorrect spectral assignment. Next generation hardware designed to allow rapid, sensitive, and highly accurate analyses of both precursor and fragment ion masses will undoubtedly reduce false positives in proteomic hit lists (49). Complementing high mass accuracy, isotopic recoding of PTMs provides an additional restraint for spectral assignment. As illustrated in Fig. 2, precursor and fragment ions corre-

sponding to recoded peptides bear a unique isotopic signature that could be used as a basis for accepting or rejecting a computationally assigned spectral match. Indeed, several peptides in our data were computationally matched to peptides based on favorable statistical indicators, such as high SEQUEST *XCorr* values, but were ultimately rejected because they lacked the characteristic isomix signature in their isotopic envelopes. As illustrated in supplemental Fig. S3, a peptide isolated for CID fragmentation via data-dependent selection was assigned to a GlcNAc-containing peptide from the cytosolic kinase Akl1 with high confidence. However, manual inspection of the isotopic envelope of the precursor ion revealed no evidence of the isomix signature, casting doubt on the assignment. The lack of a standard N*X*(S/T) sequon in this peptide and the cytosolic location of Akl1 are also consistent with spectral misassignment. Thus, we believe isotopic recoding of PTMs can serve as a valuable tool for rejecting incorrect spectral assignments that would otherwise pass undetected into LC-MS/MS hit lists.

By using isotopically recoded glycans, we successfully placed fragmentation priority on glycopeptide ions regardless of their relative intensities to other ions in the sample. As a result, rigorous sample enrichment for *N*-glycopeptides was not essential using the isomix strategy. Nonetheless, we speculate that imperfect enrichment with ConA allowed for the presence of many high abundance, nonglycopeptides in our sample, suppressing ionization and the successful detection and fragmentation of low abundance and poorly ionizing glycopeptides within the yeast proteome. Isotopic recoding is not intended to replace or eliminate enrichment from proteomics workflows; rather, we believe it has the potential to increase the information obtained from a given sample. We expect that enhanced glycopeptide enrichment strategies in conjunction with cellular fractionation and improved chromatographic resolution of peptides will permit detection of additional glycosites; the *S. cerevisiae* proteome has been reported to contain at least 350 *N*-glycosylated proteins (39).

The concept of isotopically recoding PTMs for targeted LC-MS/MS analysis can be extended well beyond *N*-glycosite mapping. By utilizing endogenous or engineered salvage pathways, unnatural isotopic signatures could be metabolically installed into a variety of PTMs. The relative malleability of glycan biosynthetic pathways, along with an extensive assortment of commercially available stable monosaccharide isotopologs, makes glycan-based PTMs particularly well suited for isotopic recoding. Ultimately, this approach is limited by the elemental composition and metabolic origin of the PTM being studied; although a stable, three-component GlcNAc isomix can be readily prepared and incorporated by cells, a similar phosphate-based isomix is less feasible. Importantly, there is no requirement that the isotopic signature strictly mimic that of a dibrominated probe; a variety of isotopic mixtures could impart sufficient perturbation to a peptide isotopic envelope to allow for successful targeted LC-MS/MS analyses.

The approach of isotopic recoding for targeted proteomics is not limited to the genetically modified *S. cerevisiae* strain utilized in this study. Although more challenging, we believe that pathway engineering in higher eukaryotes, with the aim of gaining compositional control over specific nucleotide-sugar pools, is feasible. Although metabolic introduction of an unnatural isotopic signature was showcased here, it is important to emphasize that similar perturbations of isotopic envelopes can also be obtained via chemospecific or enzymatic labeling of specific PTMs. For example, we are investigating the possibility of *in vitro* enzymatic transfer of isomix-bearing probes to *O*-glycosylated peptides obtained from mammalian cells; this approach would obviate the need for metabolically embedding the isomix signature. We believe the concept of isotopic recoding will be a powerful technology for those interested in profiling PTMs, including but not limited to *N*-glycans, using metabolic, enzymatic, and chemical labeling methods alike, substantially bolstering confidence in the data obtained from LC-MS/MS data sets.

Acknowledgments—We thank Rosa Viner (Thermo-Fisher) for technical assistance with data collection and analysis. We thank J. Jewett, T. Iavarone, and L. Kelly for critical reading, manuscript editing, and discussion. Images of simulated peptide envelopes were created with Isotopica (http://coco.protein.osaka-u.ac.jp/isotopica/), and the *N*-glycosylation consensus site frequency plot was generated with Weblogo (http://weblogo.berkeley.edu/logo.cgi).

* This work was supported by National Institutes of Health Grant GM066047 and Department of Defense Grant PC080659. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] This article contains supplemental material.

§ These authors contributed equally to this work.

** To whom correspondence should be addressed. Tel.: 510-643-1682; Fax: 510-643-2628; E-mail: crb@berkeley.edu.

REFERENCES

1. Zhou, H., Watts, J. D., and Aebersold, R. (2001) A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* **19,** 375–378
2. Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006) Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127,** 635–648
3. Zhang, J., Sprung, R., Pei, J., Tan, X., Kim, S., Zhu, H., Liu, C. F., Grishin, N. V., and Zhao, Y. (2009) Lysine acetylation is a highly abundant and evolutionarily conserved modification in *Escherichia coli*. *Mol. Cell. Proteomics* **8,** 215–225
4. Mischerikow, N., and Heck, A. J. (2011) Targeted large-scale analysis of protein acetylation. *Proteomics* **11,** 571–589
5. Xu, G., Paige, J. S., and Jaffrey, S. R. (2010) Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat. Biotechnol.* **28,** 868–873
6. Danielsen, J. M., Sylvestersen, K. B., Bekker-Jensen, S., Szklarczyk, D., Poulsen, J. W., Horn, H., Jensen, L. J., Mailand, N., and Nielsen, M. L. (2011) Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. *Mol. Cell. Proteomics* **3,** 10.1074/mcp.M110.003590
7. Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K., Takahashi, N., and Isobe, T. (2003) Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify *N*-linked glycoproteins. *Nat. Biotechnol.* **21,** 667–672

8. Wang, Z., Udeshi, N. D., O'Malley, M., Shabanowitz, J., Hunt, D. F., and Hart, G. W. (2010) Enrichment and site mapping of *O*-linked *N*-acetylglucosamine by a combination of chemical/enzymatic tagging, photochemical cleavage, and electron transfer dissociation mass spectrometry. *Mol. Cell. Proteomics* **9,** 153–160
9. Zhang, H., Li, X. J., Martin, D. B., and Aebersold, R. (2003) Identification and quantification of *N*-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **21,** 660–666
10. Zielinska, D. F., Gnad, F., Wiśniewski, J. R., and Mann, M. (2010) Precision mapping of an *in vivo N*-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141,** 897–907
11. Peterman, S. M., and Mulholland, J. J. (2006) A novel approach for identification and characterization of glycoproteins using a hybrid linear ion trap/FT-ICR mass spectrometer. *J. Am. Soc. Mass Spectrom.* **17,** 168–179
12. Pouria, S., Corran, P. H., Smith, A. C., Smith, H. W., Hendry, B. M., Challacombe, S. J., and Tarelli, E. (2004) Glycoform composition profiling of *O*-glycopeptides derived from human serum IgA1 by matrix-assisted laser desorption ionization-time of flight-mass spectrometry. *Anal. Biochem.* **330,** 257–263
13. Wang, Z., and Hart, G. W. (2008) Glycomic approaches to study GlcNAcylation: Protein identification, site-mapping, and site-specific *O*-GlcNAc quantitation. *Clin. Proteomics* **4,** 5–13
14. Domon, B., and Aebersold, R. (2006) Mass spectrometry and protein analysis. *Science* **312,** 212–217
15. Dean, N. (1999) Asparagine-linked glycosylation in the yeast Golgi. *Biochim. Biophys. Acta* **1426,** 309–322
16. Weerapana, E., and Imperiali, B. (2006) Asparagine-linked protein glycosylation: From eukaryotic to prokaryotic systems. *Glycobiology* **16,** 91R–101R
17. Imperiali, B., and Rickert, K. W. (1995) Conformational implications of asparagine-linked glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* **92,** 97–101
18. Kukuruzinska, M. A., and Lennon, K. (1998) Protein *N*-glycosylation: Molecular genetics and functional significance. *Crit. Rev. Oral Biol. Med.* **9,** 415–448
19. Elbein, A. D. (1984) Inhibitors of the biosynthesis and processing of *N*-linked oligosaccharides. *CRC Crit. Rev. Biochem.* **16,** 21–49
20. Marquardt, T., and Freeze, H. (2001) Congenital disorders of glycosylation: Glycosylation defects in man and biological models for their study. *Biol. Chem.* **382,** 161–177
21. Robinson, N. E., and Robinson, A. B. (2001) Deamidation of human proteins. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 12409–12413
22. Angel, P. M., Lim, J. M., Wells, L., Bergmann, C., and Orlando, R. (2007) A potential pitfall in ^{18}O-based *N*-linked glycosylation site mapping. *Rapid Commun. Mass Spectrom.* **21,** 674–682
23. Hägglund, P., Bunkenborg, J., Elortza, F., Jensen, O. N., and Roepstorff, P. (2004) A new strategy for identification of *N*-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J. Proteome Res.* **3,** 556–566
24. Zhang, W., Wang, H., Zhang, L., Yao, J., and Yang, P. (2011) Large-scale assignment of *N*-glycosylation sites using complementary enzymatic deglycosylation. *Talanta* **85,** 499–505
25. Palaniappan, K. K., Pitcher, A. A., Smart, B. P., Spiciarich, D. R., Iavarone, A. T., and Bertozzi, C. R. (2011) Isotopic signature transfer and mass pattern prediction (IsoStamp): An enabling technique for chemically-directed proteomics. *ACS Chem Biol.* **6,** 829–336
26. Goodlett, D. R., Bruce, J. E., Anderson, G. A., Rist, B., Pasa-Tolic, L., Fiehn, O., Smith, R. D., and Aebersold, R. (2000) Protein identification with a single accurate mass of a cysteine-containing peptide and constrained database searching. *Anal. Chem.* **72,** 1112–1118
27. Zhu, Y., Pan, Q., Thibaudeau, C., Zhao, S., Carmichael, I., and Serianni, A. S. (2006) [^{13}C,^{15}N]2-Acetamido-2-deoxy-D-aldohexoses and their methyl glycosides: Synthesis and NMR investigations of J-couplings involving ^{1}H, ^{13}C, and ^{15}N. *J. Org. Chem.* **71,** 466–479
28. Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* **6,** 359–362
29. Du, P., Kibbe, W. A., and Lin, S. M. (2006) Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics* **22,** 2059–2065
30. Eng, J. K., Mccormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5,** 976–989
31. Cherry, J. M., Ball, C., Weng, S., Juvik, G., Schmidt, R., Adler, C., Dunn, B., Dwight, S., Riles, L., Mortimer, R. K., and Botstein, D. (1997) Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature* **387,** 67–73
32. Barnes, G., Hansen, W. J., Holcomb, C. L., and Rine, J. (1984) Asparagine-linked glycosylation in *Saccharomyces cerevisiae*: Genetic analysis of an early step. *Mol. Cell. Biol.* **4,** 2381–2388
33. Breidenbach, M. A., Gallagher, J. E., King, D. S., Smart, B. P., Wu, P., and Bertozzi, C. R. (2010) Targeted metabolic labeling of yeast *N*-glycans with unnatural sugars. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 3988–3993
34. Biemann, K. (1990) Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. *Methods Enzymol.* **193,** 455–479
35. Conboy, J. J., and Henion, J. D. (1992) The determination of glycopeptides by liquid-chromatography mass-spectrometry with collision-induced dissociation. *J. Am. Soc. Mass Spectrom.* **3,** 804–814
36. Michalski, A., Cox, J., and Mann, M. (2011) More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J. Proteome Res.* **10,** 1785–1793
37. (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.* **38,** D142–D148
38. Gelperin, D. M., White, M. A., Wilkinson, M. L., Kon, Y., Kung, L. A., Wise, K. J., Lopez-Hoyo, N., Jiang, L., Piccirillo, S., Yu, H., Gerstein, M., Dumont, M. E., Phizicky, E. M., Snyder, M., and Grayhack, E. J. (2005) Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes Dev.* **19,** 2816–2826
39. Kung, L. A., Tao, S. C., Qian, J., Smith, M. G., Snyder, M., and Zhu, H. (2009) Global analysis of the glycoproteome in *Saccharomyces cerevisiae* reveals new roles for protein glycosylation in eukaryotes. *Mol. Syst. Biol.* **5,** 308
40. Brachmann, C. B., Davies, A., Cost, G. J., Caputo, E., Li, J., Hieter, P., and Boeke, J. D. (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14,** 115–132
41. Titani, K., Kumar, S., Takio, K., Ericsson, L. H., Wade, R. D., Ashida, K., Walsh, K. A., Chopek, M. W., Sadler, J. E., and Fujikawa, K. (1986) Amino acid sequence of human von Willebrand factor. *Biochemistry* **25,** 3171–3184
42. Gavel, Y., and von Heijne, G. (1990) Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: Implications for protein engineering. *Protein Eng.* **3,** 433–442
43. Kowarik, M., Young, N. M., Numao, S., Schulz, B. L., Hug, I., Callewaert, N., Mills, D. C., Watson, D. C., Hernandez, M., Kelly, J. F., Wacker, M., and Aebi, M. (2006) Definition of the bacterial *N*-glycosylation site consensus sequence. *EMBO J.* **25,** 1957–1966
44. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305,** 567–580
45. Nesvizhskii, A. I., and Aebersold, R. (2005) Interpretation of shotgun proteomic data: The protein inference problem. *Mol. Cell. Proteomics* **4,** 1419–1440
46. Aebersold, R., and Goodlett, D. R. (2001) Mass spectrometry in proteomics. *Chem. Rev.* **101,** 269–295
47. Käll, L., Storey, J. D., MacCoss, M. J., and Noble, W. S. (2008) Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J. Proteome Res.* **7,** 29–34
48. Klammer, A. A., Park, C. Y., and Noble, W. S. (2009) Statistical calibration of the SEQUEST XCorr function. *J. Proteome Res.* **8,** 2106–2113
49. Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009) A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* **8,** 2759–2769