



Published in final edited form as:

Nat Methods. 2015 June ; 12(6): 561–567. doi:10.1038/nmeth.3366.

Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis

Christina M Woo¹, Anthony T Iavarone², David R Spiciarich¹, Krishnan K Palaniappan^{1,5}, and Carolyn R Bertozzi^{1,3,4}

¹Department of Chemistry, University of California, Berkeley, California, USA

²QB3/Chemistry Mass Spectrometry Facility, University of California, Berkeley, California, USA

³Department of Molecular and Cell Biology, University of California, Berkeley, California, USA

⁴Howard Hughes Medical Institute, Berkeley, California, USA

Abstract

Protein glycosylation is a heterogeneous post-translational modification (PTM) that plays an essential role in biological regulation. However, the diversity found in glycoproteins has undermined efforts to describe the intact glycoproteome via mass spectrometry (MS). We present IsoTaG, a mass-independent chemical glycoproteomics platform for characterization of intact, metabolically labeled glycopeptides at the whole-proteome scale. In IsoTaG, metabolic labeling of the glycoproteome is combined with (i) chemical enrichment and isotopic recoding of glycopeptides to select peptides for targeted glycoproteomics using directed MS and (ii) mass-independent assignment of intact glycopeptides. We structurally assigned 32 N-glycopeptides and over 500 intact and fully elaborated O-glycopeptides from 250 proteins across three human cancer cell lines and also discovered unexpected peptide sequence polymorphisms (pSPs). The IsoTaG platform is broadly applicable to the discovery of PTM sites that are amenable to chemical labeling, as well as previously unknown protein isoforms including pSPs.

Glycosylation of proteins generates greater proteomic diversity than any other PTM. In eukaryotes, most glycans are attached to proteins at asparagine (N-glycans) or serine or threonine (O-glycans) side chains. These forms of glycosylation regulate a range of biological processes, including protein folding and trafficking¹, signaling² and cell-cell communication³. Dynamic changes in protein glycosylation have been found to accompany disease states such as Alzheimer's⁴ and cancer progression⁵, suggesting that specific protein

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

Correspondence should be addressed to C.R.B. (crb@berkeley.edu).

⁵Present address: Google Inc., Mountain View, California, USA.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS

C.R.B. conceived of and directed the project. C.M.W. designed and synthesized the probe. C.M.W. designed and performed cell culture and enrichment studies. A.T.I. and C.M.W. collected MS data. D.R.S. and C.M.W. optimized the IsoStamp algorithm. C.M.W. analyzed the data and composed the manuscript. K.K.P. performed preliminary studies. All authors revised the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

glycoforms⁶, and the enzymes that generate them⁷, may serve as disease biomarkers and therapeutic targets.

Methods to profile the intact glycoproteome would enable correlation of glycosylation variants with disease state, yet structural characterization of glycoproteins poses a great challenge in biology^{8,9}. MS is often the method of choice for profiling the proteome, but expansion of MS methods for profiling the glycoproteome is uniquely problematic. In contrast to proteomic profiling of uniform PTMs (such as phosphorylation^{10,11} and acetylation¹²), intact-glycopeptide analysis faces challenges on two fronts: detection by MS and computational identification via database searching. Glycan heterogeneity produces a number of substoichiometric modifications to the glycopeptide that effectively reduce the quantity of any particular glycoform. Because of this, as well as the lower ionization efficiency of glycoconjugates¹³, the detection of intact glycopeptides often requires increased sample quantities and enrichment. Moreover, the heterogeneity and nontemplated nature of glycopeptides stymie computational analysis. Crucially, most high-throughput proteomic platforms rely on database searches for peptide identification¹⁴. Nontemplated PTMs (for example, glycosylation) present significant computational challenges on the proteome level. Whereas a consensus sequence exists for N-glycosylation (N-X-S/T, where X is not proline), there are no defined consensus motifs for the various types of O-glycosylation. Indeed, the sheer number of possible glycoforms associated with the entire proteome yields an unmanageable sum of combinations for database search algorithms to match.

Without a universal method to identify glycoforms by database searches, enrichment of specific glycoprotein subtypes and reduction of glycan heterogeneity are presently required for glycoproteomic profiling. Enrichment of specific glycoprotein subtypes has been achieved via metabolic labeling^{15–18}, chemical tagging^{19,20}, enzymatic labeling²¹ and lectin chromatography^{22–24}. Nonetheless, identification of glycopeptides from whole proteomes requires reduction of the glycans to a uniform and predictable mass. Reduction of the glycan has been achieved for N-glycoproteins by enzymatic digestion (by PNGase F and EndoH)²⁵ and for O-glycans by genetic engineering (SimpleCell)^{26,27}. However, truncation of the glycan destroys information correlating glycan structural heterogeneity to the protein attachment site and undermines further correlation with biological activity. Methods for broadly profiling intact glycopeptides on the whole-proteome scale have until now remained elusive.

Here we present a mass-independent chemical glycoproteomics platform that enabled us to identify hundreds of intact glycopeptides from whole cell glycoproteomes. The platform, termed isotope-targeted glycoproteomics (IsoTaG), comprises four central components: (i) metabolic labeling with a chemically functionalized glycan, (ii) chemical tagging and enrichment using an isotopic recoding affinity probe, (iii) directed tandem MS and (iv) targeted glycopeptide assignment. In earlier work, we developed a computational algorithm, termed isotopic signature transfer and mass pattern prediction (IsoStamp)²⁸, for the detection of recoded species in full-scan mass spectra. Here, we show that an IsoTaG approach improved glycopeptide targeting and led to the identification of 32 intact N-glycopeptides and over 500 intact O-glycopeptides from 250 glycoproteins across three

cancer cell lines. Approximately 50% of the identified proteins were not previously known to be glycosylated. Furthermore, the specific peptide sequences for 60% of the glycopeptides were assigned as glycosylated for the first time.

RESULTS

Design and implementation of the IsoTaG platform

Critical to the performance of IsoTaG is the design and synthesis of dibrominated biotin tag **1** that allows for click chemistry–based affinity enrichment, recovery of the glycopeptide, and isotopic recoding of the intact glycopeptide (Fig. 1, Supplementary Fig. 1 and Supplementary Methods). Several cleavable linkers have been developed for proteomics applications;²⁹ we selected the silane scaffold³⁰ because of our observation that glycosidic linkages tolerate its mild acid cleavage conditions (2% formic acid in H₂O). The natural abundance of the stable isotopes, ⁷⁹Br and ⁸¹Br (1:1), provided a ready source for isotope recoding. In previous work we found that the dibromide motif, which convolutes the mass envelope of a peptide with a triplet signature (1:2:1 (M, M + 2, M + 4)), could be detected computationally using a pattern-recognition algorithm that we developed in-house²⁸.

We tested probe **1** in a glycoproteomics analysis of Jurkat cell lysates using the IsoTaG workflow (Fig. 2a). To demonstrate the flexibility of the platform for various glycoprotein subtypes, we labeled Jurkat cells separately with tetraacetylated *N*-azidoacetylgalactosamine (Ac₄GalNAz), tetraacetylated *N*-azidoacetylmannosamine (Ac₄ManNAz) or dimethylsulfoxide (DMSO) vehicle for 48 h. Ac₄ManNAz is metabolized to the corresponding *N*-azidoacetyl sialic acid (SiaNAz) to yield labeled sialoglycoproteins³¹. Ac₄GalNAz is processed by the cellular GalNAc salvage pathway to form UDP-GalNAz, which is inter-converted by the UDP-galactose 4-epimerase (GALE) with its C4 epimer, UDP-GlcNAz¹⁵. Thus, metabolic labeling with Ac₄GalNAz produces GalNAz-labeled mucin-type O-glycoproteins, GlcNAz-labeled N-glycoproteins and O-GlcNAzylated proteins. We harvested labeled Jurkat cells by centrifugation and collected the media by centrifugal filtration for analysis of secreted glycoproteins. In addition, the cell pellet was homogenized and separated into soluble and insoluble fractions. Three milligrams of protein per fraction were used as a starting point for glycoproteomics enrichment.

Fractionated Jurkat cell lysates were conjugated to probe **1** via copper-catalyzed azide-alkyne [3+2] cycloaddition (CuAAC; Fig. 2a). Excess probe was removed by methanol precipitation, and proteins were resuspended in 1% RapiGest surfactant in PBS. Glycoproteins were enriched by streptavidin-agarose affinity chromatography and digested on-bead with trypsin to release all nonconjugated peptides for glycoprotein identification. Treatment with 2% formic acid recovered the glycopeptides from the beads. The efficiency of glycoprotein capture and release was assessed by western blot (Supplementary Figs. 2–4). The released glycopeptides were analyzed by reversed-phase nanoflow liquid chromatography coupled to a Thermo LTQ–Orbitrap XL mass spectrometer.

In traditional proteomics approaches, tandem MS is performed on the most abundant species in the full-scan mass spectra to the exclusion of lower-abundance species (Fig. 2b). In contrast, IsoTaG enables mass-independent, targeted glycoproteomics (Fig. 2c).

Glycopeptides displaying the isotopic signature are computationally detected by the pattern-searching algorithm²⁸, which produces an inclusion list of m/z values and retention times (± 2 -min window) for ions bearing isotopically recoded envelopes. We then performed an inclusion list-directed tandem MS run, where we typically observed a <0.5 -min shift in retention time. Femtomole quantities of tagged proteins at a signal-to-noise ratio of 2.5:1 were detectable with the pattern-recognition algorithm²⁸. Owing to the versatility of the algorithm²⁸, species up to 12 kDa that have incorporated multiple tags are also directed to the inclusion list and have an equal chance of selection for tandem MS.

IsoTaG improves selection of low-abundance glycopeptides

To assess the advantage of IsoTaG over traditional data-dependent acquisition, we performed a series of back-to-back runs with and without the inclusion list and evaluated the total number of isotopically recoded species selected for tandem MS. A fourfold improvement in tagged glycopeptides selected for tandem MS was observed across all fractions and glycan types (Supplementary Fig. 5). In total, over 1,500 isotopically recoded species were selected for tandem MS from Ac₄GalNAz-labeled and Ac₄ManNAz-labeled Jurkat cell lysates using the inclusion list, as compared to 385 without. This selection factor indicates that IsoTaG effectively targets low-abundance species and mitigates the need for extensive fractionation, mass spectrometer analysis time and computation time. To assess the rate of off-target labeling, we quantified the number of isotopically recoded species selected for tandem MS from the DMSO control samples. Using the same procedure, we found 52 isotopically recoded off-target species in the DMSO control (off-target rate = 3%).

We then repeated the enrichment procedure with PC-3 and MCF-7 cell lines. Because of our interest in identifying low-abundance glycopeptides, tandem MS was performed using collision-induced dissociation (CID) to achieve the highest sensitivity. Figure 2c illustrates the identification of an N-glycoprotein from Ac₄GalNAz-labeled Jurkat lysates. Inclusion list-triggered tandem MS of an isotopically recoded precursor ion generated a tandem mass spectrum (MS2) for assignment of glycan structure from charged and neutral losses. Subsequent data-dependent CID of the three most intense ions in the MS2 spectrum generated MS3 spectra for glycopeptide amino acid sequence identification by database searching. In a typical MS duty cycle, we performed inclusion list-driven MS2 on the top three precursor ions, and the three most intense fragment ions in each MS2 spectrum were selected for MS3 analysis, for a total of nine events following MS1.

By virtue of their isotopically recoded mass envelopes, glycopeptides could be identified in a targeted, mass-independent manner (Fig. 3). MS2 and MS3 spectra of isotopically recoded precursor ions were searched against the Swiss-Prot human proteome using the Sequest HT algorithm within Proteome Discoverer software. Unassigned spectra were then searched in an iterative fashion using the UniProtKB human proteome and Byonic, a program specialized for variable-modification searches. Glycans were assigned manually from the MS2 spectra, and glycan structures were inferred from the observed neutral or charged losses (Supplementary Fig. 6). Ambiguous glycan types (for example, GlcNAz vs. GalNAz) were inferred from the subcellular localization of the underlying peptide or the most likely glycan structure (for example, **O10** and **O11**; Fig. 3b). All assignments were validated for

isotopic distribution, exact mass (mass error, $\text{mass} < 5$ p.p.m.), charged or neutral loss of the glycan, and peptide fragment ions. None of the assigned glycopeptides was detected in the DMSO control samples. To assess reproducibility, we performed biological replicates of the enrichment procedure with the Jurkat cell line and found that approximately half of the glycopeptides were identified in all three replicates, and 78% were found in two of three replicates (Fig. 4a). In aggregate, 32 N- and over 500 O-glycopeptides from 250 glycoproteins were identified, of which 220 peptides and 120 proteins were not previously known to be glycosylated.

Mass spectrometry data analysis

The identified glycopeptides represent a range of proteins and glycan types. Sixteen N-glycan types were identified (**N1–N16**; Fig. 3a), in which immature and high-mannose structures were more predominant than we expected (Fig. 3c). In addition to the 32 fully characterized N-glycopeptides, we partially characterized 156 N-glycopeptide spectra (for example, the glycan was sequenced but not the underlying peptide). The lower number of fully identifiable N-glycopeptides mainly reflects difficult elucidation by MS (up to four SiaNAz glycans were observed in a single species, but the entire glycoform was unresolved) but may also be caused by hindered access of probe **1** or reduced ionization efficiency. Alternatively, metabolically labeled immature N-glycans may be more abundant than elaborated N-glycans in the samples evaluated, a hypothesis supported by previous observations using lectin chromatography²⁴. To fully characterize the peptides carrying N-glycans, we performed the enrichment using conditioned medium from PC-3 cells metabolically labeled with Ac₄GalNAz, Ac₄ManNAz or DMSO. After trypsin digestion, we used PNGase F in ¹⁸O-water to release only peptides with N-glycans and found 54 formerly N-glycosylated peptides at 1% false discovery rate. These samples represent one-ninth of the total fractions evaluated herein, indicating that N-glycoproteins are labeled and captured by our enrichment procedure. Unfortunately, we could no longer characterize the structures that occupied the glycosite using this established method²³.

Identified O-glycan types corresponded to O-GlcNAc (**O1**), Tn (**O2**), STn (**O5**), Cores 1–4 (**O3**, **O11**, **O4** and **O10**, respectively) and sialylated glycans (**O6**, **O7**, **O8** and **O9**) (Fig. 3b). The preponderance of O-GlcNAcylated peptides (Fig. 3d) reflects, at least in part, the efficiency of the GALE epimerase to biosynthesize UDP-GlcNAz from Ac₄GalNAz¹⁵. Additionally, terminal glycans (for example, GlcNAz) may be easier to tag with probe **1**. Nonetheless, O-GlcNAcylated proteins are highly abundant on the basis of other analyses³².

Fragmentation of the glycan occurs before fragmentation of the peptide backbone with CID, enabling the assignment of glycan structure—but not the underlying peptide—for several of the elaborated glycopeptides. Whereas electron-transfer dissociation (ETD) is an orthogonal fragmentation method for peptides bearing CID-labile PTMs³³, we found that halogenated glycopeptides have low ETD fragmentation efficiency. The mucin-type O-glycans identified were predominantly various sialylation states of core 1 O-glycans from PC-3 and MCF-7 cell lines. Jurkat cells displayed highly heterogeneous O-glycopeptides owing to the knockdown of core 1 O-glycan elaboration via truncation of the COSMC chaperone³⁴. Notably, core 1 O-glycans from Jurkat cells were still identified, as truncated COSMC

maintains 2–5% activity of the full-length protein. The presence of core 1 glycoproteins in the conditioned medium from Jurkat cells was confirmed by lectin staining (Supplementary Fig. 7).

Interestingly, unique subpopulations of glycopeptides were enriched with either GalNAz or ManNAz labeling (Fig. 4b). The predicted subcellular localization of the identified glycoproteins reflected the prevalence of nuclear and cytoplasmic O-GlcNAzylated peptides identified (Fig. 4c). Nonetheless, the identified glycoproteins were highly cell specific. Of the characterized glycoproteins, 60% were found exclusively in one of the three cell types, and 30% in two of the three cell types (Fig. 4d).

We identified intact glycopeptides with multiple glycoforms (Supplementary Table 1). Elaborated glycopeptides possessed a high degree of glycoform variation, including peptides derived from the same protein (O94907) or from different proteins (for example, O94907 and Q14766). Up to five glycoforms were observed for a single peptide (P14314), and several peptides possessed four detected glycoforms (P02786, Q02818, Q12841, Q14242 and Q92854).

Identification of peptide sequence polymorphisms

Approximately 60% of isotopically recoded species were assigned by database searching methods described above. The remaining 40% of tandem mass spectra of isotopically recoded species were unassigned from database searching alone. In some cases, spectral nonassignment was related to low signal-to-noise ratios of fragment ions or the presence of unidentified chemical modifications (for example, unexpected fragments and CuAAC-related off-target reactivity). In other cases, however, we suspected that failure to identify glyco-peptides from database searching was due to the absence of a matching amino acid sequence in the protein database. To investigate this possibility, we collected an additional data set on Ac₄ManNAz-labeled PC-3 cell lysates, with full-scan mass spectra and MS2 spectra acquired in the Orbitrap analyzer to obtain high-resolution measurement of both precursor and fragment ions. Targeted analysis of this data set revealed 21 glycopeptides carrying pSPs that were not present in the protein database (UniProtKB) and one pSP previously observed by genetic sequencing³⁵. For example, a total of 15 isoforms were identified, representing four glycan types and six pSPs derived from prostate-associated microseminoprotein³⁶ (Fig. 5 and Supplementary Fig. 8). Polymorphic peptides additionally carry glycoform variants (for example, P14314 and Q1L6U9; Supplementary Table 1); this is potentially related to variable levels of abundance or indicative of differential biosynthetic processing of the underlying peptide.

Identification of protein polymorphisms in cancer is an area of intense study on the genomic³⁷ and proteomic^{38,39} levels. Interestingly, all pSPs identified from PC-3 cells, a metastatic prostate cancer cell line, were mutations of alanine to serine or of proline to threonine. The selectivity of mutations observed in our data may be related to the glycan-dependent enrichment strategy. Alternatively, as aberrant glycosylation is a hallmark of cancer, particularly in metastasis⁴⁰, an increase in O-glycosylation sites may play a significant role and is a subject for future investigation.

DISCUSSION

Unlike previous methods, which require cleavage or truncation of glycans for glycosite mapping, IsoTaG identifies fully elaborated glycopeptides, revealing the inherent structural details of the glycoproteome. Use of the IsoStamp algorithm increased selection of glycopeptide species for tandem MS by fourfold and facilitated mass-independent targeted database searching for high-confidence assignment. As described, the enrichment and recovery of metabolically labeled glycopeptides with probe **1**, MS analysis and initial data analysis were all performed in 3 d. In our preliminary study we found that IsoTaG was highly reproducible using biological replicates from Jurkat lysates. With Jurkat, PC-3 and MCF-7 cell lines, 50% of identified glycoproteins and 60% of identified glycopeptides were novel, suggesting that many more glycoproteins, and particularly O-glycoproteins, remain to be discovered. Indeed, the assignment of a large number of novel O-glycopeptides further highlights the ability of IsoTaG to survey new classes of glycoproteins. A 3% off-target CuAAC rate was found with the Jurkat cell line, as quantified by the number of isotopically recoded species selected for tandem MS in labeled versus control samples. Protein glycosylation was also highly cell specific, as 60% of the identified glycoproteins were found exclusively in one of the three cell types (30% in two of three cell types). As less than 2% of proteins are expressed in a cell-specific manner⁴¹, analysis of glycoproteins may be an important source of tissue-specific diversity and disease biomarkers.

Notably, any covalent labeling strategy may be integrated into the IsoTaG platform to achieve enrichment and targeted proteomics of low-abundance species. We have illustrated the generality of the platform by using two metabolically labeled glycans: Ac₄GalNAz and Ac₄ManNAz. Use of highly glycan-specific metabolic^{17,18} or enzymatic⁴² labels may enable focused analysis of particular glycan types. Incorporation of the isotopic pattern via metabolic heavy-atom labeling is also a viable alternative⁴³. An Orbitrap mass spectrometer was used in the present study; however, other types of mass spectrometers that can achieve high resolution (for example, quadrupole time-of-flight analyzers) can also be employed²⁸. Recent advances in nested fragmentation techniques (such as higher-energy collisional dissociation with accurate mass-product-dependent ETD) show promise toward a targeted glycoproteomics approach and may be incorporated into an IsoTaG strategy^{44,45}.

IsoTaG represents the first proteome-scale platform in which the glycan structure and peptide can be identified for both intact N- and O-glycopeptides. Recognition of tagged glycopeptides by full-scan MS improved rates of selection and characterization, such that 60% of isotopically recoded species were characterized (glycan structure and peptide), including the identification of novel pSPs. These results point to a wealth of information in the proteome and glycome that are not covered in current databases. Given that on average 20% of spectra in a data set are assigned by database searching⁴⁶, we suggest that a targeted isotopic recoding approach may be broadly useful for the future discovery of novel protein isoforms, including nontemplated PTMs and pSPs.

ONLINE METHODS

Materials

Commercial solvents and reagents were used as received with the following exceptions. Dichloromethane and *N,N*-dimethylformamide were purified according to the method of Pangborn and coworkers⁴⁷. Triethylamine was distilled from calcium hydride under an atmosphere of nitrogen immediately before use. RapiGest was prepared according to the method of Lee and Compton⁴⁸. 3-[4-({Bis[(1-*tert*-butyl-1*H*-1,2,3-triazol-4-yl)methyl]amino}methyl)-1*H*-1,2,3-triazol-1-yl]propanol (BTTP) was prepared according to the method of Wu and coworkers⁴⁹. Tetraacetylated *N*-azidoacetyl galactosamine was prepared according to the method of Bertozzi and coworkers⁵⁰. Tetraacetylated *N*-azidoacetyl mannosamine was prepared according to the method of Bertozzi and coworkers⁵¹. EDTA-free protease inhibitor cocktail was obtained from Roche Diagnostics (Version 11). Streptavidin-agarose beads were obtained from Thermo Scientific and washed with PBS before use. High-sensitivity streptavidin-HRP was obtained from Pierce (catalog number 21130). PNGase F was purchased from NEB (catalog number P0704S).

Cell culture

All cell lines were obtained from the American Type Culture Collection (ATCC) and maintained at 37 °C and 5% CO₂ in a water-saturated incubator. Jurkat and PC-3 cells were maintained in RPMI-1640 supplemented with 10% FBS and 1% penicillin/streptomycin. MCF-7 cells were maintained in DMEM supplemented with 10% FBS and 1% penicillin/streptomycin. Cell lines were metabolically labeled between passages 4–7 (MCF-7), 6–10 (Jurkat), or 17–22 (PC-3) and visually inspected daily for the presence of mycoplasma.

Cell culture flasks were seeded with 100 μM of Ac₄GalNAz, Ac₄ManNAz, or vehicle control containing DMSO (3.0 μL). A suspension of cells (2 × 10⁵ cells/mL) was added to the flask, and the flask was incubated for 48 h at 37 °C. Flasks containing adherent cells were aspirated, and the cells were washed with PBS. The contents of flasks containing suspension cells were transferred to centrifuge tubes, and the cells were pelleted (300g, 3 min). The medium was aspirated, and cell pellets were washed with PBS. Cells were resuspended in complete medium containing 100 μM glycan metabolite without FBS additive, and the cells were incubated an additional 48 h at 37 °C.

Medium (100 mL) was harvested and cleared by centrifugation (300g, 3 min). Clarified medium was spin concentrated (Amicon, 15-mL 10-kDa spin filter) to 1 mL. The concentrated residue was washed with PBS (three washes of 15 mL each: 3 × 15 mL) and transferred to an Eppendorf microcentrifuge tube as the “conditioned medium fraction.” Adherent cells were washed with PBS (1 × 10 mL) and trypsinized for 5 min at 37 °C. Cells were harvested, centrifuged (300g, 3 min), and washed with PBS. Suspension cells were pelleted by centrifugation (300g, 3 min) and washed with PBS. Cell pellets were resuspended in lysis buffer (2 mL: 10 mM HEPES, pH 7.9, 1.5 mM MgCl₂, 10 mM KCl, 0.5% Triton X-100, 1 × protease inhibitors, 1 μM thiamet G), swelled for 5 min on ice, and broken by Dounce homogenization using a tight glass hand pestle (Wheaton, 30 strokes). Insoluble material was pelleted by centrifugation (3,700g, 10 min, 4 °C). The supernatant

was collected as the “soluble fraction” and the pellet kept as the “insoluble fraction.” The insoluble fraction was resuspended in 1% RapiGest/PBS and briefly probe sonicated. The conditioned medium and soluble fractions were adjusted to 1% RapiGest/PBS. Protein concentration from the three fractions was measured by bicinchonic acid assay (Pierce) and normalized to 4.5 mg/mL.

CuAAC-assisted enrichment

GalNAz-labeled, ManNAz-labeled, or DMSO vehicle-treated cell fractions were divided into aliquots containing 3.0 mg protein. Click-chemistry reagents (40.0 μ L, 200 μ M **1**, 300 μ M CuSO₄, 600 μ M BTTP, 1.50 mM sodium ascorbate, mixed immediately before addition to lysates) were added, and the reaction was incubated for 3.5 h at 24 °C. Proteins were precipitated in methanol (1 mL) for 1 h at –80 °C. Precipitated proteins were pelleted by centrifugation, and the supernatant was discarded. Protein pellets were resuspended in 400 μ L 1% RapiGest/PBS and solubilized by probe sonication. Streptavidin-agarose resin was added, and the resulting mixture was incubated for 12 h at 24 °C with rotation. The beads were pelleted by centrifugation, and the supernatant containing uncaptured proteins was separated. The beads were washed with 1% RapiGest/PBS (1 mL), 6 M urea (2 \times 1 mL), and PBS (5 \times 1 mL).

The washed beads were reduced (5 mM DTT) and alkylated (10 mM iodoacetamide). On-bead trypsin (1.5 μ g) digestion proceeded for 12 h at 37 °C. Beads were pelleted by centrifugation, and the supernatant digest was collected. The beads were washed with PBS (1 \times 200 μ L) and H₂O (2 \times 200 μ L). Washes were combined with the supernatant digest to form the trypsin digest. The silane tag **1** was cleaved with two treatments of 2% formic acid/H₂O (200 μ L) for 30 min at 24 °C with rotation, and the eluent was collected. The beads were washed with 50% acetonitrile-water + 1% formic acid (2 \times 200 μ L), and the washes were combined with the eluent to form the cleavage fraction. The trypsin digest and cleavage fraction were concentrated using a vacuum centrifuge to 50–100 μ L. Samples were desalted by ZipTip P10 and stored at –20 °C until analysis.

Mass spectrometry

Recovered glycopeptides were analyzed using a Thermo Dionex UltiMate3000 RSLCnano liquid chromatograph that was connected in-line with an LTQ Orbitrap XL mass spectrometer equipped with a nanoelectrospray ionization (nanoESI) source. Full-scan mass spectra were acquired in the positive-ion mode over the range m/z = 400–1,800 using the Orbitrap mass analyzer, in profile format, with a mass resolution setting of 60,000. Inclusion lists were generated using the IsoStamp algorithm (<http://www.cchem.berkeley.edu/crbgrp/IsoStamp/>). In the data-dependent mode, the three most intense ions exceeding an intensity threshold of 50,000 counts were selected from each full-scan mass spectrum for tandem mass spectrometry (MS/MS, i.e., MS₂) analysis using collision-induced dissociation (CID). MS₂ spectra were acquired using the linear ion trap or the Orbitrap analyzer (in the latter case, with a resolution setting of 7,500 at m/z = 400, FWHM), in centroid format, with the following parameters: isolation width, 4 m/z units; normalized collision energy, 28%; default charge state, 3+; activation Q, 0.25; and activation time, 30 ms. The three most intense fragment ions in each MS₂ spectrum exceeding an intensity threshold of 1,000 counts were

selected for MS3 analysis using CID. MS3 spectra were acquired using the linear ion trap, in centroid format, with the same parameters as those used for MS2. Global parent mass lists (i.e., inclusion lists) were enabled to specify the m/z values and retention times of glycopeptide precursor ions detected in full-scan mass spectra by the IsoStamp isotope pattern-searching algorithm. Data acquisition was controlled using Xcalibur software (version 2.0.7, Thermo). The raw data are available upon request.

Data analysis

The raw data were processed using Proteome Discoverer 1.4 software and searched against the human-specific SwissProt-reviewed database downloaded 18 July 2014. Tandem MS data were screened for glycopeptide signifiers including isotopically recoded precursor ion in full-scan mass (i.e., MS1) spectra, and neutral or charged glycan losses in MS2 spectra. Selected MS2/MS3 spectra were documented and saved separately. Saved spectra were manually annotated for glycoform and peptide mass. Saved spectra were then searched iteratively using the Byonic search algorithm v2.0 as a node in Proteome Discoverer 1.4. Computational assignments of all spectra were validated by manual inspection for glycan and peptide fragments. High probability assignments were inspected for validity, and unassigned spectra were archived for further analysis. Unassigned spectra from the initial searches were sorted on the basis of glycan type derived from fragment ions observed in the MS2 spectra (for example, HexNAzBr₂OH, elaborated O-glycan, or elaborated N-glycan) and searched with variable modification using a focused glycan database. Finally, spectra that remained low-confidence assignments were manually inspected for similarities to assigned spectra (i.e., characteristic peptide fragment ions) or searched against the UniProtKB database (downloaded 30 September 2014) with variable modification on the specific glycan structure. Mass spectrometry spectra are provided as Supplementary Data 1–4 and annotated as losses from the MS^(*n* – 1) precursor mass as follows: glycan(number of glycan units) in reverse order of the observed losses. Glycans at separate sites are separated by a comma. The glycan annotations used are HexNAzBr₂OH = C₁₅H₂₀Br₂N₄O₇ (+527.9678, abbreviated HexNAz*), HexNAz = C₈H₁₂N₄O₅ (+244.0808), HexNAcNH₂ = C₈H₁₄N₂O₅ (+218.0903), HexNAc = C₈H₁₃NO₅ (+203.0794), NeuAzBr₂OH = C₁₈H₂₄Br₂N₄O₁₀ (+615.9839, abbreviated NeuAz*), NeuAz = C₁₁H₁₆N₄O₈ (+332.0968), NeuAcNH₂ = C₁₁H₁₈N₂O₈ (+306.1063), NeuAc = C₁₁H₁₇NO₈ (+291.0954), Hex = C₆H₁₀O₅ (+162.0528), +CO = CO (+27.9949, occurs on the O-terminus of the tag).

Western blotting procedures

PNGase F release of formerly N-glycosylated peptides—GalNAz-labeled, ManNAz-labeled, or DMSO vehicle-treated medium was enriched and trypsin digested as described above. The beads were washed with PBS (1 × 200 μL) and H₂O (1 × 200 μL). The beads were washed with 40 mM ammonium carbonate in ¹⁸O-water (1 × 40 μL). N-glycosylated peptides were cleaved by treatment with PNGase F (5 U) in ammonium carbonate in ¹⁸O-water (40 mM, 100 μL) in a 10-kDa spin filter (Corning) for 6 h at 37 °C. The eluent from the PNGase F treatment was collected by centrifugation (7,000 r.c.m., 5 min). The beads were washed with 50% acetonitrile-water (2 × 400 μL), and the washes were combined with the eluent to form the PNGase F fraction. The PNGase F fraction was

concentrated using a vacuum centrifuge (i.e., a SpeedVac, 40 °C) to 50–100 µL. Samples were desalted with a ZipTip P10 and stored at –20 °C until analysis.

Anti-biotin immunoblotting—Aliquots collected during the enrichment procedure (10.0 µL) were reduced and separated by standard SDS-PAGE (Bio-Rad, Criterion system), electroblotted onto nitrocellulose, blocked in 5% bovine serum albumin (Sigma) in PBS with 0.1% Tween-20, and analyzed by standard enhanced chemiluminescence immunoblotting methods (Pierce). The staining agent was streptavidin-HRP (Pierce, 1:100,000).

Lectin staining—GalNAz-labeled or DMSO-vehicle Jurkat conditioned medium (100 µg) in buffer (25.0 µL, 50 mM NaOAc, pH 5.5, 4 mM CaCl₂) was divided into two aliquots. One aliquot from each condition was treated with neuraminidase (4.0 µL, *Vibrio cholerae*, Roche). Aliquots were mixed and incubated at 37 °C for 12 h. Aliquots (10.0 µL) were reduced and separated by standard SDS-PAGE (Bio-Rad, Criterion system), electro-blotted onto nitrocellulose, blocked in PBS with 0.5% Tween-20, and analyzed by standard fluorescent imaging (Typhoon 9410, GE Healthcare). The staining agent was peanut agglutinin–FITC (Vector Laboratories, 1:100).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank P. Robinson and N. Rumachik for critical reading. Financial support from the US National Institutes of Health (CA200423, C.R.B.), Jane Coffin Childs Memorial Fund (C.M.W.), US National Science Foundation (Graduate Research Fellowship, D.R.S.) and Howard Hughes Medical Institute (C.R.B.) are gratefully acknowledged.

References

1. Gouyer V, et al. Inhibition of the glycosylation and alteration in the intracellular trafficking of mucins and other glycoproteins by GalNAcα-O-bn in mucosal cell lines: an effect mediated through the intracellular synthesis of complex GalNAcα-O-bn oligosaccharides. *Front Biosci.* 2001; 6:D1235–D1244. [PubMed: 11578961]
2. Zachara NE, Hart GW. Cell signaling, the essential role of O-GlcNAc! *Biochim Biophys Acta.* 2006; 1761:599–617. [PubMed: 16781888]
3. Hudak JE, Canham SM, Bertozzi CR. Glycocalyx engineering reveals a Siglec-based mechanism for NK cell immunoevasion. *Nat Chem Biol.* 2014; 10:69–75. [PubMed: 24292068]
4. Yuzwa SA, et al. Increasing O-GlcnaC slows neurodegeneration and stabilizes tau against aggregation. *Nat Chem Biol.* 2012; 8:393–399. [PubMed: 22366723]
5. Fuster MM, Esko JD. The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat Rev Cancer.* 2005; 5:526–542. [PubMed: 16069816]
6. Büll C, Stoel MA, den Brok MH, Adema GJ. Sialic acids sweeten a tumor's life. *Cancer Res.* 2014; 74:3199–3204. [PubMed: 24830719]
7. Radhakrishnan P, et al. Immature truncated O-glycophenotype of cancer directly induces oncogenic features. *Proc Natl Acad Sci USA.* 2014; 111:E4066–E4075. [PubMed: 25118277]
8. Mariño K, Bones J, Kattla JJ, Rudd PM. A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol.* 2010; 6:713–723. [PubMed: 20852609]

9. Pan S, Chen R, Aebersold R, Brentnall TA. Mass spectrometry based glycoproteomics—from a proteomics perspective. *Mol Cell Proteomics*. 2011; 10:R110 003251. [PubMed: 20736408]
10. Zhou H, Watts JD, Aebersold R. A systematic approach to the analysis of protein phosphorylation. *Nat Biotechnol*. 2001; 19:375–378. [PubMed: 11283598]
11. Olsen JV, et al. Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell*. 2006; 127:635–648. [PubMed: 17081983]
12. Choudhary C, et al. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*. 2009; 325:834–840. [PubMed: 19608861]
13. Pouria S, et al. Glycoform composition profiling of O-glycopeptides derived from human serum IgA1 by matrix-assisted laser desorption ionization-time of flight-mass spectrometry. *Anal Biochem*. 2004; 330:257–263. [PubMed: 15203331]
14. Zhang Y, Fonslow BR, Shan B, Baek MC, Yates JR III. Protein analysis by shotgun/bottom-up proteomics. *Chem Rev*. 2013; 113:2343–2394. [PubMed: 23438204]
15. Boyce M, et al. Metabolic cross-talk allows labeling of O-linked β -*N*-acetylglucosamine-modified proteins via the *N*-acetylgalactosamine salvage pathway. *Proc Natl Acad Sci USA*. 2011; 108:3141–3146. [PubMed: 21300897]
16. Hubbard SC, Boyce M, McVaugh CT, Peehl DM, Bertozzi CR. Cell surface glycoproteomic analysis of prostate cancer-derived PC-3 cells. *Bioorg Med Chem Lett*. 2011; 21:4945–4950. [PubMed: 21798741]
17. Chuh KN, Zaro BW, Piller F, Piller V, Pratt MR. Changes in metabolic chemical reporter structure yield a selective probe of O-GlcNAc modification. *J Am Chem Soc*. 2014; 136:12283–12295. [PubMed: 25153642]
18. Zaro BW, Yang YY, Hang HC, Pratt MR. Chemical reporters for fluorescent detection and identification of O-GlcNAc-modified proteins reveal glycosylation of the ubiquitin ligase NEDD4-1. *Proc Natl Acad Sci USA*. 2011; 108:8146–8151. [PubMed: 21540332]
19. Zhang H, Li XJ, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol*. 2003; 21:660–666. [PubMed: 12754519]
20. Nilsson J, et al. Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat Methods*. 2009; 6:809–811. [PubMed: 19838169]
21. Khidekel N, et al. Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat Chem Biol*. 2007; 3:339–348. [PubMed: 17496889]
22. Vosseller K, et al. O-linked *N*-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol Cell Proteomics*. 2006; 5:923–934. [PubMed: 16452088]
23. Zielinska DF, Gnad F, Wi niewski JR, Mann M. Precision mapping of an *in vivo* N-glycoproteome reveals rigid topological and sequence constraints. *Cell*. 2010; 141:897–907. [PubMed: 20510933]
24. Trinidad JC, Schoepfer R, Burlingame AL, Medzihradsky KF. N- and O-Glycosylation in the murine synaptosome. *Mol Cell Proteomics*. 2013; 12:3474–3488. [PubMed: 23816992]
25. Häggglund P, Bunkenborg J, Elortza F, Jensen ON, Roepstorff P. 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*. 2004; 3:556–566. [PubMed: 15253437]
26. Steentoft C, et al. Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods*. 2011; 8:977–982. [PubMed: 21983924]
27. Steentoft C, et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J*. 2013; 32:1478–1488. [PubMed: 23584533]
28. Palaniappan KK, et al. Isotopic signature transfer and mass pattern prediction (IsoStamp): an enabling technique for chemically-directed proteomics. *ACS Chem Biol*. 2011; 6:829–836. [PubMed: 21604797]
29. Bielski R, Witczak Z. Strategies for coupling molecular units if subsequent decoupling is required. *Chem Rev*. 2013; 113:2205–2243. [PubMed: 23153040]
30. Szychowski J, et al. Cleavable biotin probes for labeling of biomolecules via azide-alkyne cycloaddition. *J Am Chem Soc*. 2010; 132:18351–18360. [PubMed: 21141861]

31. Prescher JA, Dube DH, Bertozzi CR. Chemical remodelling of cell surfaces in living animals. *Nature*. 2004; 430:873–877. [PubMed: 15318217]
32. Hart, GW.; Akimoto, Y. *Essentials of Glycobiology*. 2. Varki, A., et al., editors. Vol. Ch 18. Cold Spring Harbor Laboratory Press; 2009.
33. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA*. 2004; 101:9528–9533. [PubMed: 15210983]
34. Ju T, et al. Human tumor antigens Tn and sialyl Tn arise from mutations in *Cosmc*. *Cancer Res*. 2008; 68:1636–1646. [PubMed: 18339842]
35. Gerhard DS, et al. The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res*. 2004; 14:2121–2127. [PubMed: 15489334]
36. Valtonen-André C, et al. A highly conserved protein secreted by the prostate cancer cell line PC-3 is expressed in benign and malignant prostate tissue. *Biol Chem*. 2007; 388:289–295. [PubMed: 17338636]
37. Hindorf LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA*. 2009; 106:9362–9367. [PubMed: 19474294]
38. Tran JC, et al. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*. 2011; 480:254–258. [PubMed: 22037311]
39. Zhang B, et al. Proteogenomic characterization of human colon and rectal cancer. *Nature*. 2014; 513:382–387. [PubMed: 25043054]
40. Paszek MJ, et al. The cancer glycoalyx mechanically primes integrin-mediated growth and survival. *Nature*. 2014; 511:319–325. [PubMed: 25030168]
41. Pontén F, et al. A global view of protein expression in human cells, tissues, and organs. *Mol Syst Biol*. 2009; 5:337. [PubMed: 20029370]
42. Clark PM, et al. Direct in-gel fluorescence detection and cellular imaging of O-GlcNAc-modified proteins. *J Am Chem Soc*. 2008; 130:11576–11577. [PubMed: 18683930]
43. Breidenbach MA, Palaniappan KK, Pitcher AA, Bertozzi CR. Mapping yeast N-glycosites with isotopically recorded glycans. *Mol Cell Proteomics*. 2012; 11:M111.015339. [PubMed: 22261724]
44. Yin X, et al. Glycoproteomic analysis of the secretome of human endothelial cells. *Mol Cell Proteomics*. 2013; 12:956–978. [PubMed: 23345538]
45. Wu SW, Pu TH, Viner R, Khoo KH. Novel LC-MS² product dependent parallel data acquisition function and data analysis workflow for sequencing and identification of intact glycopeptides. *Anal Chem*. 2014; 86:5478–5486. [PubMed: 24796651]
46. Elias JE, Haas W, Faherty BK, Gygi SP. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods*. 2005; 2:667–675. [PubMed: 16118637]
47. Pangborn AB, Giardello MA, Grubbs RH, Rosen RK, Timmers FJ. Safe and convenient procedure for solvent purification. *Organometallics*. 1996; 15:1518–1520.
48. Lee, PJJ.; Compton, BJ. Destructible surfactants and uses thereof. US patent. 7,229,539. 2007.
49. Wang W, et al. Sulfated ligands for the copper(I)-catalyzed azide-alkyne cycloaddition. *Chem Asian J*. 2011; 6:2796–2802. [PubMed: 21905231]
50. Hang HC, Yu C, Kato DL, Bertozzi CR. A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation. *Proc Natl Acad Sci USA*. 2003; 100:14846–14851. [PubMed: 14657396]
51. Prescher JA, Dube DH, Bertozzi CR. Chemical remodelling of cell surfaces in living animals. *Nature*. 2004; 430:873–877. [PubMed: 15318217]

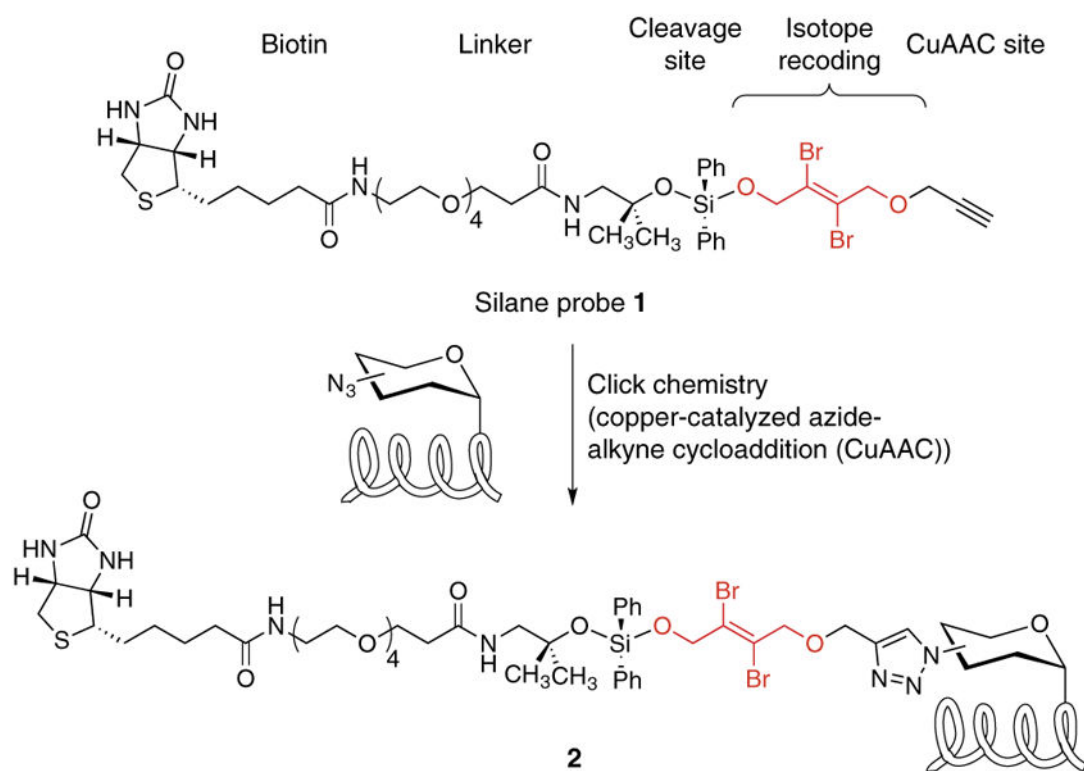


Figure 1. Structure of cleavable probe **1** for mass-independent glycoproteomics. Probe **1** tags glycoproteins via click chemistry, resulting in tagged glycoproteins **2**.

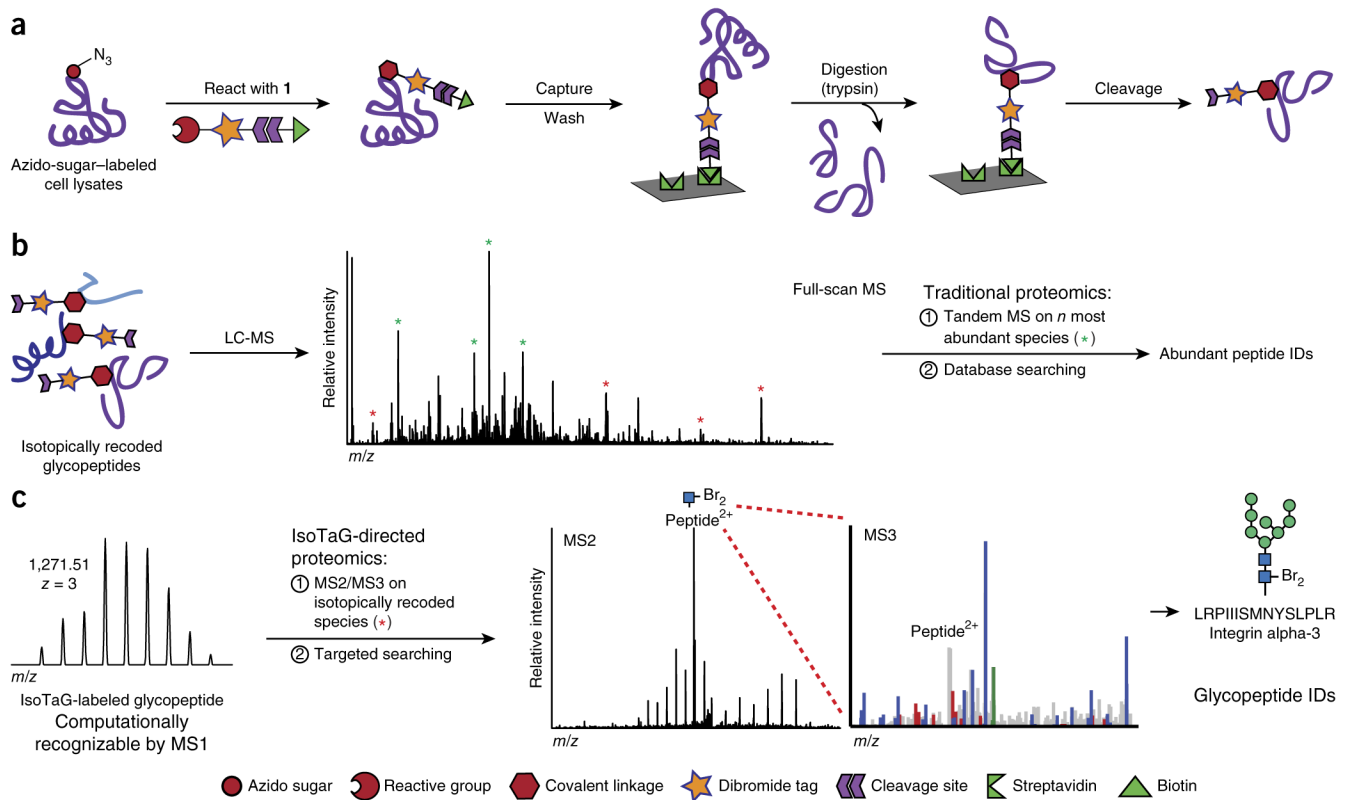


Figure 2. IsoTaG enrichment and mass spectrometry workflow. **(a)** Enrichment strategy using dibrominated silane probe **1**, shown in cartoon form. Metabolically labeled glycoproteins are enriched using streptavidin-agarose affinity chromatography. On-bead digestion removes nonglycosylated peptides, and isotopically recoded glycopeptides are cleaved and eluted from the resin. **(b)** Isotopically recoded glycopeptides are analyzed by reversed-phase liquid chromatography coupled to a mass spectrometer. **(c)** IsoTaG uses the pattern-searching algorithm to direct tandem MS (i.e., MS2 and MS3) analysis to isotopically recoded species. Targeted searching of selected species results in high-confidence glycopeptide identification. The tagged glycan is denoted with “Br₂.”

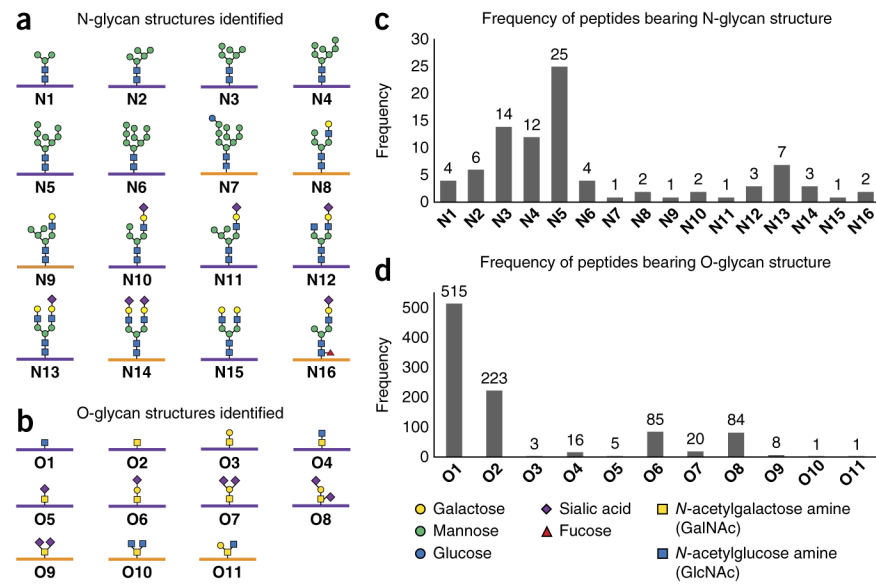


Figure 3. IsoTaG identified a diversity of intact glycopeptides. **(a)** N-glycan structures (**N1–N16**) identified. **(b)** O-glycan structures identified include O-GlcNAc (**O1**), Tn (**O2**), STn (**O5**), core 1 (**O3**), core 2 (**O11**), core 3 (**O4**), core 4 (**O10**) and sialylated variants (**O6–O9**). A purple peptide (bar) indicates a glycan and peptide sequence were identified for at least one glycopeptide. These species generated interpretable peptide fragments. An orange peptide indicates that only the glycan was assigned. These species gave unassignable peptide fragments. **(c,d)** Frequencies of observed peptides bearing N-glycans (**c**) and O-glycans (**d**).

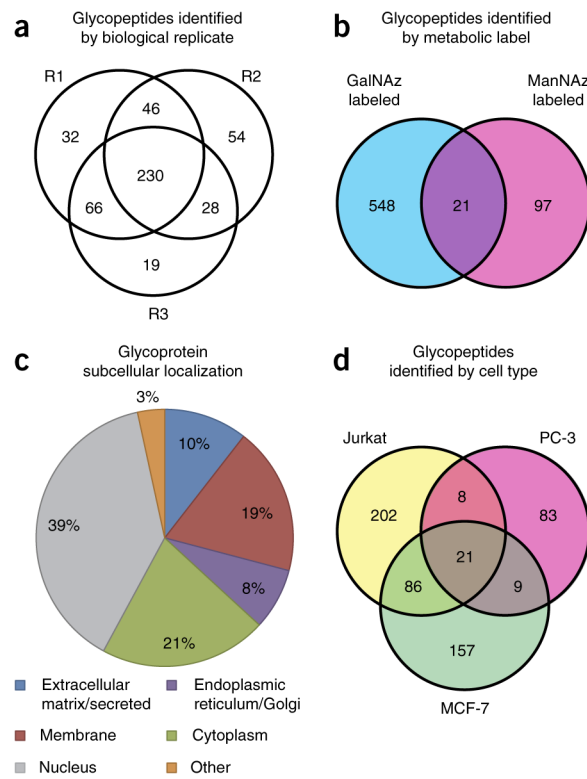


Figure 4. Distribution of identified glycopeptides. **(a,b)** Glycopeptides identified with three biological replicates (R1–R3) of Jurkat cell lysate enrichment **(a)** and from GalNAz- or ManNAz-labeled lysates **(b)**. **(c,d)** Glycoproteins identified via IsoTaG are distributed throughout the cell **(c)** and are highly cell specific **(d)**.

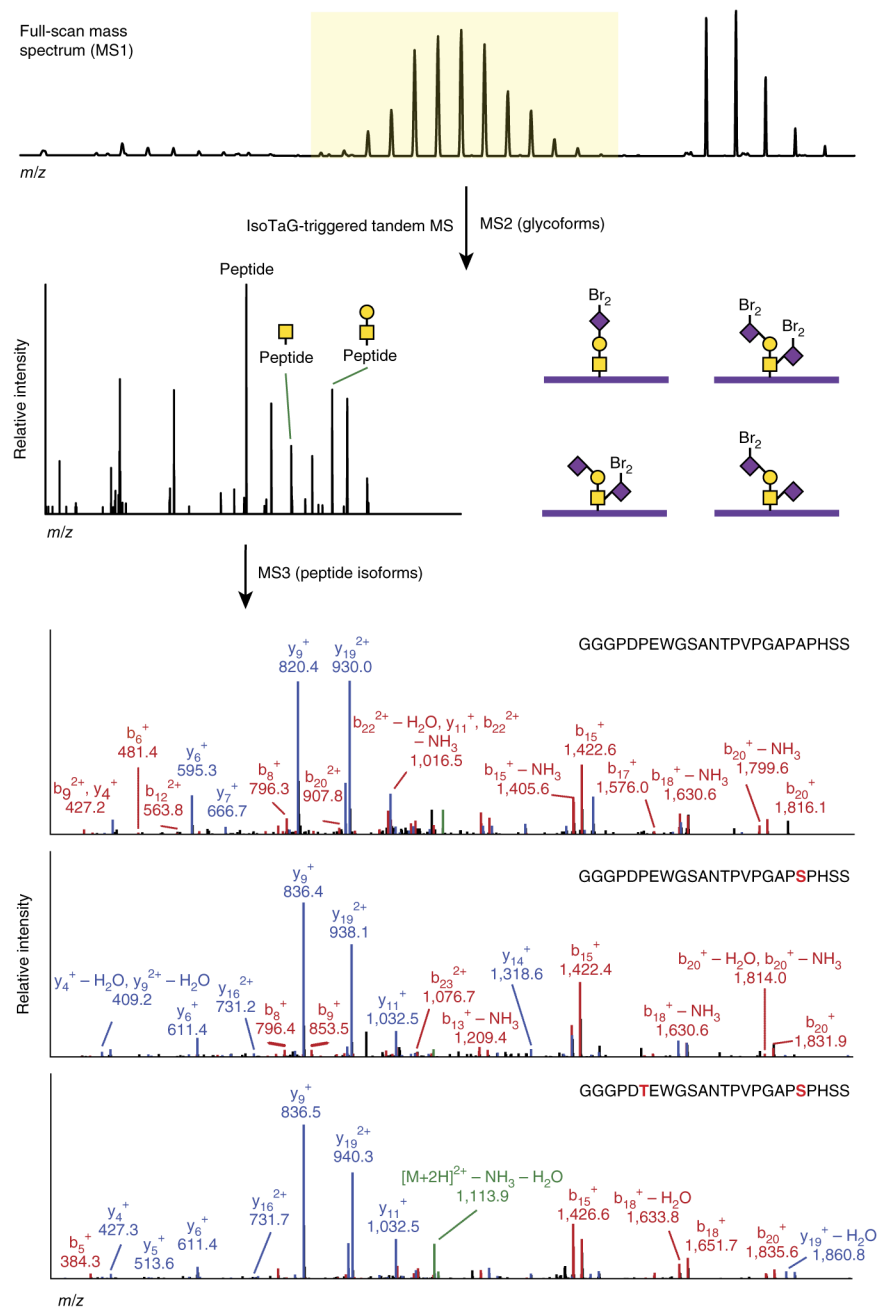


Figure 5. Targeted tandem MS analysis of isotopically recoded glycopeptides identifies both glycan and peptide isoforms. The novel O-glycopeptide from prostate-associated microseminoprotein was found as four glycoforms by MS2 and six pSPs by MS3. The tagged glycan is denoted with “Br₂.” Amino acid variants are in boldface red font. Fragmentation spectra for three peptides are shown (b ions in red, y ions in blue).