# Adapting Data-Independent Acquisition for Mass Spectrometry-Based Protein Site-Specific N-Glycosylation Analysis

Kuan-Ting Pan,[†] Chen-Chun Chen,[‡] Henning Urlaub,[*,†,§] and Kay-Hooi Khoo[*,∥]

[†]Bioanalytical Mass Spectrometry Group, Max Plank Institute for Biophysical Chemistry, Goettingen 37077, Germany
[‡]Genomics Research Center, Academia Sinica, Taipei 11529, Taiwan
[§]Department of Clinical Chemistry, University Medical Center Goettingen, Goettingen 37075, Germany
[∥]Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan

S *Supporting Information*

ABSTRACT: A hallmark of protein N-glycosylation is extensive heterogeneity associated with each glycosylation site. In human cells, the constituent glycoforms differ mostly in numerous ways of extensions from an invariable trimannosyl core and terminal modifications. The efficient identification of these glycoforms at the glycopeptide level by mass spectrometry (MS) requires a precursor sampling technique that is not dictated by signal intensity or by preset targets during MS2 data acquisition. We show here that the recently developed data-independent acquisition (DIA) approach is best suited to this demanding task. It allows postacquisition extraction of glycopeptide-specific fragment-ion chromatograms to be aligned with that of precursor MS1 ion by nanoLC elution time. For any target glycoprotein, judicious selection of the most favorable MS1/MS2 transitions can first be determined from prior analysis of a purified surrogate standard that carries similar site-specific glycosylation but may differ in its exact range of glycoforms. Since the MS2 transitions to be used for extracting DIA data is common to that glycosylation site and not dictated by a specific MS1 value, our workflow applies equally well to the identification of both targeted and unexpected glycoforms. Using a case example, we show that, in targeted mode, it identified more site-specific glycoforms than the more commonly used data-dependent acquisition method when the amount of the target glycoprotein was limited in a sample of high complexity. In discovery mode, it allows detection, with supporting MS2 evidence, of under-sampled glycoforms and of those that failed to be identified by searching against a predefined glycan library owing to unanticipated modifications.
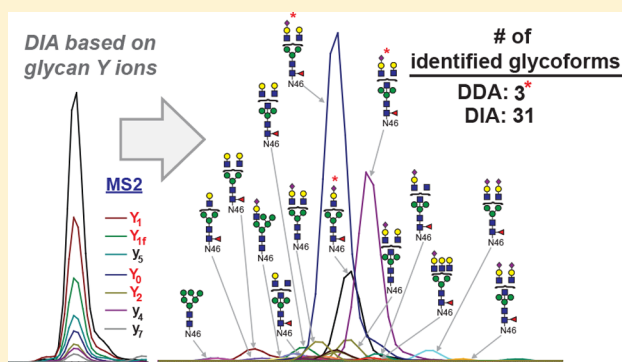
Protein glycosylation involves a series of nontemplate-encoded steps of elongation and branching setting out from several distinctive core structures. It thus generates a seemingly undefined, yet nonrandom, heterogeneity that is specific for each cell state, cell type, and protein site. The "holy grail" of mass spectrometry (MS)-based glycoproteomics[1] is to determine efficiently this vast array of glycosylation variants, or glycoforms, so that one can better understand their functional effects. Confident identification of a glycopeptide invariably relies on the ability to induce sequence-informative fragmentation with and without retaining the glycan moiety through MS2 techniques such as the higher-energy collisional dissociation (HCD) implemented on the Orbitrap series and the electron transfer dissociation (ETD).[1,2] Recent advances in MS instrumentation, complemented by software and search-engine development, now allow the efficient semiautomated identification of several hundreds of N-glycopeptides from large LC–MS/MS data sets.[3,4] However, a common problem (other than false identifications) is that usually only a fraction of all glycoforms for a specific site is detected at sufficient MS1 intensity to be selected randomly for MS2 by data-dependent

acquisition (DDA). Indeed, the stochastic nature of DDA in picking precursors, being based primarily on signal intensity, often results in a less reproducible peptide identification list that is biased toward the more abundant ions.[5] The performance of DDA is severely affected by the relative abundance of a target peptide and hence by the sample's complexity. This further implies that depth of coverage often comes at the price of enrichment and fractionation, which in turn necessitates greater amounts of starting sample, which must be available in a form conducive to purification procedures. All of these factors are unfavorable for the analysis of endogenous membrane glycoproteins with complex substoichiometric glycosylation patterns.

Understanding the technical shortcomings of DDA has spurred on recent developments in alternative data-acquisition methods, which capitalize on previously acquired MS2 data or

spectral libraries to derive targeted assays for specific peptides.[6] This includes the single, multiple, or parallel reaction monitoring (SRM/MRM/PRM) methods[7,8] that aim to identify selectively predefined target peptides in a hypothesis-driven manner. These targeted methods provide higher sensitivity, dynamic range, and reproducibility at the cost of discarding all other "irrelevant" signals. In contrast, the increasingly popular data-independent acquisition (DIA) method[9,10] accumulates fragments indiscriminately from all precursors at once, with or without dividing them into consecutive scan events based on predefined m/z ranges. DIA measurement is then followed by data extraction from the digital archive that theoretically contains all product ions from all precursors occurring within that mass range.[11] A main advantage of DIA compared with SRM/MRM/PRM is that the DIA data set not only allows postacquisition interrogation of targeted peptides but also preserves the possibility of future exploration of any other peptides not initially known or targeted for analysis.

For modified peptides, however, such advantageous DIA approaches have been less widely adopted, partly owing to the need first to evaluate and build up the optimum MS1/MS2 transitions by using synthetic peptides that carry the modification at the specific sites. In the case of glycoproteomics, deriving a reliable spectral library by using a comprehensive collection of synthetic glycopeptides is even less feasible. Additional issues include the way glycopeptides fragment, which more readily produces highly abundant glyco-oxonium ions and a series of glycosyl residue losses from precursors of little peptide-identification values. Nonetheless, given the new generations of MS instruments that can achieve high-resolution/accurate mass detection with ever increasing speed and sensitivity, the technical aspects of glycopeptide analysis by either the targeted[12−14] or the DIA approach[15,16] merit more critical evaluation than has yet been undertaken. The goal is to devise the most appropriate, efficient and productive workflow, which can complement current DDA analysis.

From the outset, we envisaged that DIA would be most useful for mapping, as comprehensively as possible, all the glycoforms of one or more glycosylation site(s) of a target glycoprotein. This is currently an unmet goal in DDA-based approaches for glycopeptide analysis. Arguably, it is best addressed by target analysis. Focusing on this aspect and starting with N-glycosylation for this work, we showed that the extracted ion chromatograms (XIC) of shared peptide-specific fragment ions can be used to filter out all related glycoforms of a specific glycosylation site while alignment with the matching MS1 precursor XICs allows discrimination of the individual glycoforms. As would be expected, the precursor isolation width for MS2 is a critical parameter affecting successful identification, especially when the glycoforms coelute. Overall, using the human serum IgM as a case study and surrogate glycoprotein standard for a future targeted assay of surface IgM/B cell receptor, our DIA analysis on a high-end quadrupole-orbitrap hybrid-MS instrument allowed the identification of more unique glycoforms from 6 different sites including 1 from the conjunctive IgJ, compared with the DDA method, when sample amount is limiting in the presence of spiked-in complex cell lysates. We further demonstrated how the DIA method can be strategically adapted from targeted to discovery mode, not only to "rescue" glycopeptides missed by DDA, but also to map additional glycoforms with unexpected modifications.

■ RESULTS AND DISCUSSIONS

Identification of IgM N-Glycopeptides by DDA. Purified human serum IgM was reduced, alkylated, and digested sequentially with trypsin and endopeptidase GluC before being subjected to triplicate LC−MS/MS analyses on a Q Exactive HF mass spectrometer operated under a general DDA Top20 setting. The raw files were processed and searched against a protein database containing only sequences of human IgM and IgJ together with a pre-established mammalian N-glycan database using Byonic.[17] At a false discovery rate (FDR) of 1%, more than 1500 unique peptide-to-spectrum matches (PSMs) were identified as glycopeptides from triplicate analyses of 1 μg of IgM digests injected on column (Table s1), around 30% of which did not meet our manual verification criteria (see Supplemental Methods) and therefore were not included in the final reference list. In total, the reference list contains 327 glycopeptides or 219 unique glycopeptides if those occurring in more than one charge states were only counted once each, corresponding to a total of 124 non-redundant site-specific glycoforms from five N-linked glycosylation sites (N46, N209, N272, N279, and N439) on IgM and one (N71) on IgJ (Table s2A).

Since the ability to detect a particular glycopeptide by DDA depends largely on its abundance and sample complexity, we next prepared human IgM at various concentrations, with or without adding in yeast lysates to mimic real sample. With the same DDA settings on the MS instrument and the same data-processing approach and PSM acceptance criteria, the total number of nonredundant site-specific glycoforms identified dropped by more than 30% from 124 to 84 (or 219 to 155 unique glycopeptides; Figure 1A) when the amount of IgM injected was reduced from 1 μg to 500 ng on column. At 50 ng, only 29 site-specific glycoforms, of which none corresponded to the N439 site, were identified. A further 5-fold drop in the number of glycoforms identified was observed when the same quantity (50 ng) of IgM was used to spike 250 μg of yeast lysate, yielding only 6 unique glycoforms from three glycosylation sites (N46 and N272 on IgM and N71 on IgJ). Moreover, with lower amounts of spiked-in IgM, the overall quality of MS2 spectra was greatly reduced (Figure 1B), which resulted in lower identification scores for the PSMs.

Glycosylation Site-Specific MS2 Ions for DIA. Beam-type CID of N-linked glycopeptides often induces a series of sequential losses of glycosyl residues, leading to a characteristic daughter ion corresponding to its naked peptide backbone, with only a single GlcNAc retained at the Asn. This is commonly referred to as the Y1 ion.[2,18] For N-glycopeptides carrying core fucosylation, Y1 + Fuc (Y1f) is also frequently observed, whereas ions corresponding to the peptide core itself (Y0) and the larger ones retaining more than a GlcNAc, namely, Y1 + GlcNAc (Y2), Y2 + Man (Y3), Y2 + 2Man (Y4), and Y2 + 3Man (Y5), etc., hereafter referred to as "glycan Y ions", can also be detected to different extent. Under HCD on an Orbitrap, the Y1 ion is often the most prominent peak at above m/z 600, although this depends somewhat on the peptide sequence and the glycan. As shown in Figure 1B, both Y1 and Y1f ions for the selected N46 glycopeptides remained clearly detectable when only small amounts of target glycopeptides were present. This was also the case for most other IgM glycopeptides identified by DDA here. Since these
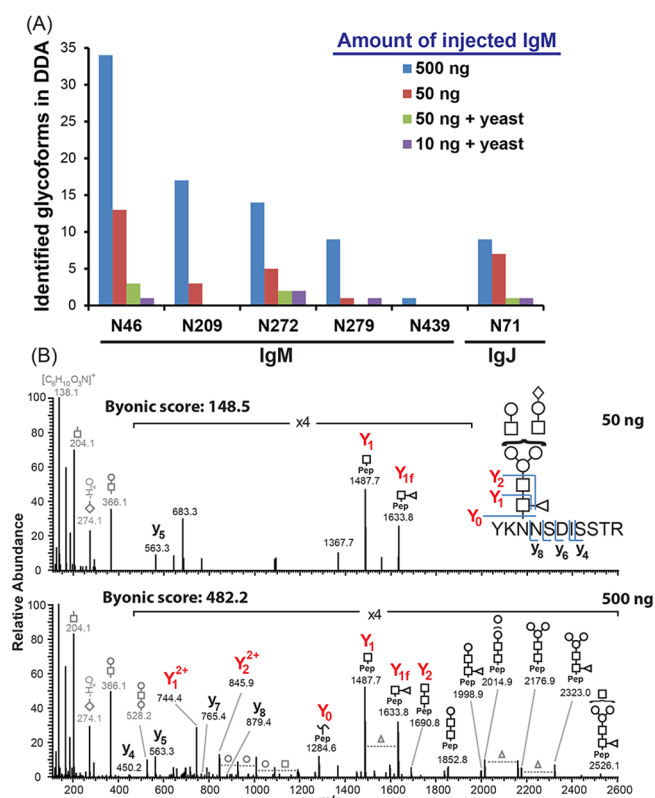
Figure 1. Identification of human IgM glycopeptides using the DDA approach. Various amounts of enzymatically digested human serum IgM, with or without mixing with yeast lysate, were analyzed with a Q-Exactive mass spectrometer by the DDA method; this was followed by data-processing and a database search using Byonic. (A) Numbers of identified N-linked glycoforms on various glycosylation sites were affected by both the concentration of IgM/IgJ and the sample complexity. (B) MS2 spectra of N46 glycopeptide from 50 ng (above) or 500 ng (below) of IgM digests, showing a typical pattern of dominant Y0/Y1/Y1f ions obtained in HCD fragmentation mode. Glycan Y ions are labeled in red and the respective Byonic identification scores for the two PSMs are indicated.

fragment ions carry peptide mass-specific information, they would represent a good balance of sensitivity and selectivity and hence a better choice for DIA transition selection, compared with the highly abundant, ubiquitous low-molecular-weight glyco-oxonium ions shared by all glycopeptides and used in other targeted approaches.[12,13,15] Extracting these oxonium ions across the entire mass range from DIA data would nevertheless give a good indication of which MS2 swath carries most of the glycopeptide signals and when they are eluted.

Aiming to attain the highest DIA sensitivity, we evaluated systematically the yield of target peptide fragment and glycan Y ions from IgM glycopeptides as compared with the normalized collisional energies (NCE) applied. Our results showed that the optimum NCE, yielding the most intense fragment ions, was peptide-sequence-dependent and varied significantly from one glycosylation site to another (Figure s1). On the other hand, consistent with a previous report,[18] different glycans on the same peptide cores exerted only marginal effects on optimum NCE selection. Overall, lower NCE (22-28) resulted in higher intensities of glycan Y ions, while higher NCE (26-30) was preferred for detecting peptide fragment ions. Stepped NCE was also evaluated but was found advantageous for only a portion of the glycopeptides. Hereafter in this work, NCE 28

was applied to achieve the best balance of producing peptide fragment ions versus the glycan Y ions for DIA analyses. However, for targeted analysis on relatively pure samples, where peptide fragment ions are less critical and glycan Y ions alone are sufficient for successful identification, NCE of 22-24 would afford better analytical sensitivity.

Implementing Targeted Glycopeptide Analysis by DIA. Further to transition selection, a prime consideration in DIA is to fit as many isolation steps, or swaths, with as small width as possible to cover the entire desired mass range in a reasonably short cycle time compatible with LC peak width. Accordingly, we evaluated the performance of DIA analysis of targeted intact glycopeptides based on one survey scan (MS1) followed by 50 DIA swaths (MS2) at a fixed isolation width of 16 Th to cover the mass range of m/z 800−1600, where most tryptic glycopeptides can be found. This led to an average cycle time of ~3.8 s, which allowed 7−11 MS2 data points to be acquired for each of the eluting glycopeptide peaks. After data acquisition, intensities of predefined glycopeptide targets were extracted from the DIA raw files, either manually by using Xcalibur (Thermo Scientific) or with the Skyline software.[19,20] First, the MS1 ion chromatograms for the glycopeptides identified by preceding DDA analysis were extracted by their accurate precursor masses (Figure 2A, upper part). The IgM/IgJ glycopeptides bearing the same peptide backbone but carrying different glycans were found to be eluted nearby, within a retention time range of 3−5 min (Figure s2). Next, the ion chromatograms of all expected target peptide fragment ions and glycan Y ions were extracted from MS2 swaths corresponding to the 16 Th mass windows encompassing the target glycopeptide precursors (Figure 2A, lower part). Since glycopeptides sharing the same peptide core but bearing varying glycan structures have different precursor masses and were chromatographically resolved in this case example, their commonly associated peptide fragment and glycan Y ions of identical m/z values would be detected in different MS2 swaths at different retention times, as shown in Figure 2A. The extracted MS1 and MS2 ion chromatograms could then be aligned and the identity of each glycopeptide was confirmed by the perfect coelution of precursor and fragment ions.

In the case of IgM N272, although a total of 22 glycoforms were cumulatively identified by triplicate DDA analyses (Table s2), only 11 could be reproducibly identified in all three replicates. The remaining 8 and 3 glycoforms were identified only in two and one of the replicates, respectively (Figure 2B), largely because of their low MS signal intensities and the stochastic nature of precursor picking in the DDA method, despite a relatively large amount injected (1 μg). In contrast, we could found 14 out of the 22 glycoforms in a single DIA run with only 500 ng of IgM injected by matching the MS1 precursor signals to the corresponding MS2 signals of glycan Y ions. These include, for example, a low-abundance N272 glycopeptide carrying the glycan "dHex(1)HexNAc(5)Hex(6)-NeuAc(2)", which was only selected for MS2 and identified in one of the DDA triplicates. The sole evidence for its putative presence in the other two replicates is at MS1 level, based on the accurate glycosyl mass increment from a much more abundant neighboring glycopeptide identified as bearing a related glycan "dHex(1)HexNAc(5)Hex(5)NeuAc(2)" (Figure 2C). In the DIA approach, on the other hand, despite its low abundance and its coeluting with the highly abundant glycopeptide, we succeeded in detecting their common glycan
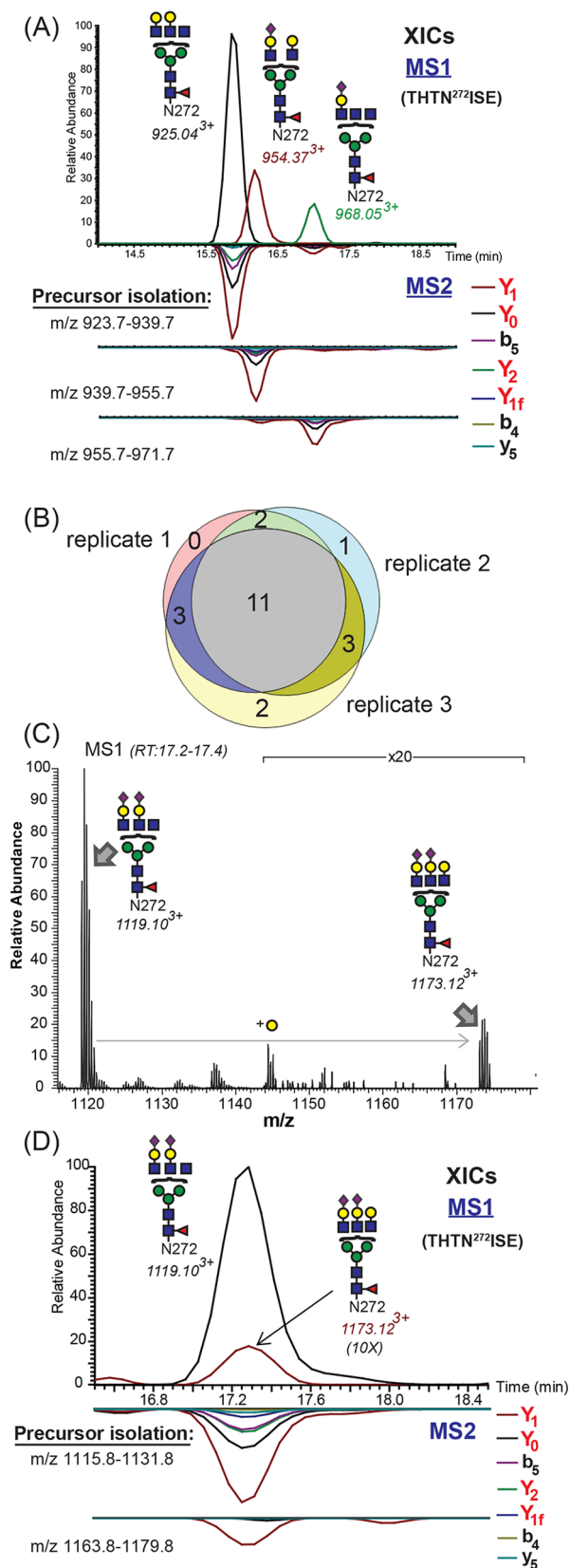
Figure 2. DIA for N-glycopeptide analysis. (A) MS1 XICs of targeted glycopeptides (upper) were aligned and matched to the MS2 XICs of peptide fragment ions and glycan Y ions extracted from various MS2 swaths corresponding to the 16 Th mass windows where the respective precursors were found. (B) Venn diagram showing the overlapping of identified unique glycoforms in three DDA replicates.

Figure 2. continued

(C) Combined MS1 scan from retention time 17.2 to 17.4 min showing the peaks assigned as N272 peptides "THTNISE" bearing the glycans "dHex(1)NexNAc(5)Hex(5)NeuAc(2)" or "NexNAc(5)-Hex(6)NeuAc(2)". Signal intensity was magnified 20× for the m/z region indicated. (D) The MS1 XICs of the two glycopeptides shown in part C were matched to the MS2 XICs of their respective fragment ions found in two separate MS2 swaths. The MS1 XIC signal of m/z 1173.12 was magnified by 10-fold.

Y ions in two separate MS2 swaths matching respectively to the two glycoforms (Figure 2D).

The isolation width of MS2 swaths is a critical parameter, affecting the performance of DIA. In order to cover the m/z region of 800−1600 for glycopeptide analysis while maintaining reasonable cycle time, 16 Th was used in this work. On the one hand, 16 Th is narrow enough to separate the majority of glycopeptides with the same peptide core. For instance, 15 of the 22 glycoforms on IgM N272 were detected in separate MS2 swaths and 4 of the 7 glycoforms with overlapping swaths could be resolved chromatographically. In fact, around 55% of the total of 124 unique site-specific glycoforms from the six glycosylation sites that were identified by DDA in this study (Table s2) could be detected in a unique 16-Th-wide MS2 swath without interference from other glycopeptides sharing the same peptide core. On the other hand, we did find that the selectivity of DIA could be further improved by using an even narrower isolation width. For example, the MS1 XIC of glycopeptide "YKN[46]NSDISSTR" bearing a "HexNAc(3)-Hex(5)NeuAc(1)" glycan on IgM N46 was only partly resolved and separated from that bearing "dHex(1)HexNAc(3)Hex(4)-NeuAc(1)" (Figure 3A), but its MS2 XICs of Y1 ion from either of the two consecutive swaths were not suitably segmented to resolve it from interfering glycoforms (Figure 3B) and was thus not identified in our DIA analysis. As shown in Figure 3C,D, with narrower isolation width and thus improved selectivity, the Y1 XIC could clearly resolve the two glycoforms when 4 Th was used. Foreseeable future improvements of scanning speed on modern mass spectrometers and implementation of multiplexing would further empower the DIA approach. It should however be noted that, as is also the case for DDA, glycan linkage analysis is not possible with the DIA approach.

Current DDA-based glycopeptide analysis followed by database searching using Byonic or equivalent tools often leads to a true FDR that is much higher than the estimated FDR. Most of the glycoforms "identified" by our DDA, but not DIA, approach, including the seven missing IgM N272 glycoforms, were of low abundance. This alone would disfavor reliable extraction of MS1 precursor ion chromatogram, an issue compounded by missing glycan Y ion signals and/or interference from background signals. DDA also has a higher propensity toward false positive assignment in the first instance. For instance, the N272 glycopeptide carrying "dHex(3)-HexNAc(5)Hex(6)" differs from that bearing "dHex(1)-HexNAc(5)Hex(6)NeuAc(1)" by only 1 Da. Both glycoforms were considered to be "identified" by Byonic. However, the monoisotopic peak of the glycopeptide bearing "dHex(3)-HexNAc(5)Hex(6)" (m/z 1076.43^{3+}) could not actually be distinguished from the isotopic envelope of the other glycopeptide on the MS1 spectra to differentiate between these two glycoforms. Moreover, oxonium ions from NeuAc
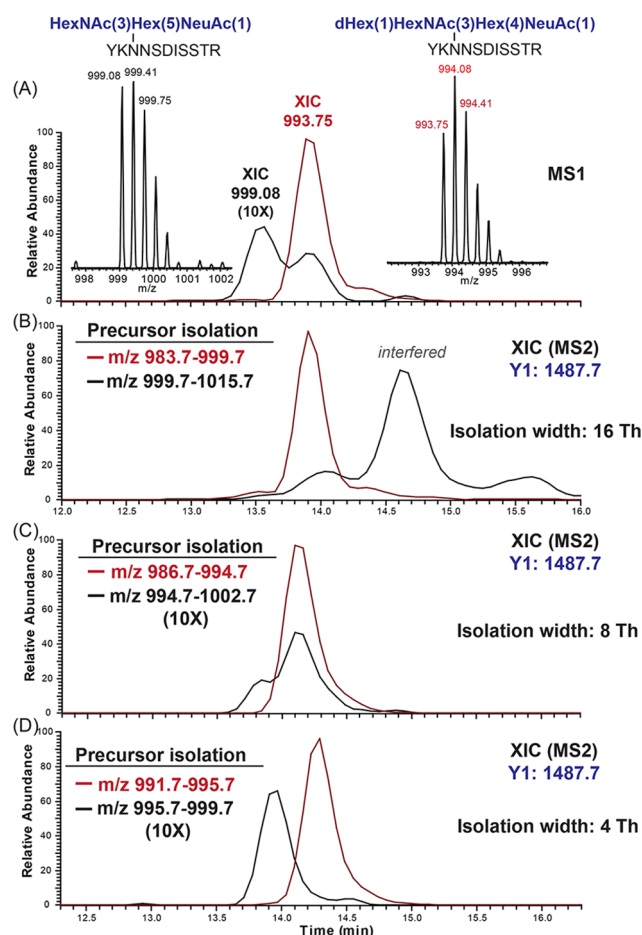
Figure 3. Improved DIA selectivity by using narrower isolation width. IgM digests were analyzed by the DIA approaches using various isolation widths or swaths, as indicated. (A) MS1 ion chromatograms of N46 peptide "YKNNSDISSTR" bearing the glycans "HexNAc(3)-Hex(5)NeuAc(1)" (m/z, 990.08; in black, 10-fold magnification) and "dHex(1)HexNAc(3)Hex(4)NeuAc(1)" (m/z, 993.75; in red) were extracted separately and overlaid. Insets show their isotopic clusters on individual MS1 scans. Their Y1 MS2 ion chromatograms were extracted, respectively, from different MS2 swaths as defined by the isolation width settings used in acquiring the DIA data, namely, 16 Th (B), 8 Th (C), or 4 Th (D).

were observed on the PSMs of the "dHex(3)HexNAc(5)-Hex(6)" glycoform, which should not contain any NeuAc. Its glycosyl composition, with 3 Fuc, is also less likely given the overall glycosylation pattern, but it is not impossible. In creating the master reference list (Table s2) from DDA experiments, PSMs contributing to such dubious "identification" would not have been filtered out, since it met the criteria we used, which aimed at minimizing false negatives and thus did not include manual effort to validate glycan composition assignment against the observed glyco-oxonium ions.

Identifying Target Glycopeptides within Complex Samples by DIA. To evaluate systematically the performance of our DIA approach, we created a target glycopeptide list on Skyline and investigated it for all DDA-identified unique glycoforms. For that purpose, only a unique peptide core for each glycosylation site was selected for DIA analysis, without considering the additional variants with deamidation and/or miscleavages, except for N46, for which the MS signal of YKN[+Glycan]NSDISSTR was higher than that of N-[+Glycan]NSDISSTR and was targeted (see Table s2 for

more details). Putative glycopeptide MS1 ions not supported by the expected glycan Y ions coeluting at consistent retention time were discarded manually. This resulted in the detection of a total of 147 out of 190 target glycopeptides assayed, or 86 out of 120 unique glycoforms, from 6 glycosylation sites when 500 ng of IgM was injected (Table s3), whereas 84 target glycopeptides were identified by DDA and Byonic under the same conditions. Importantly, four out of the total of six glycosylation sites monitored showed significantly improved sensitivity, although the degree of improvement varied among glycopeptides and was more pronounced with increasing sample complexity (compare Figure 4A with Figure 1A). For example, the N46 glycopeptide "YKN[46]NSDISSTR" bearing a "dHex(1)HexNAc(4)Hex(5)NeuAc(1)" glycan was only identified by DDA approach when 100 ng or more of IgM was added to yeast lysates (red circles in Figure 4B). This is due to mistriggering of the respective precursor for MS2 analysis when the glycopeptide is coeluted with complex mixtures of yeast peptides. In contrast, the presence of this glycopeptide can still be confirmed by our DIA approach at both the MS1 level (open circles connected with a straight line in Figure 4B) and the MS2 level (color-coded cumulative bar chart in Figure 4B), even when only 0.5 ng of IgM was spiked-in. This gave a 200-fold improvement in sensitivity compared with the DDA approach. With 50 ng of IgM in yeast lysates, a total of 31 out of 46 different N46 glycoforms on YKN[46]NSDISSTR were still detected by DIA, whereas only the most abundant three (marked with red stars in Figure 4C for the overlapped XICs) were identified by DDA.

It is worth noting that a precursor with an m/z value and retention time similar to this glycopeptide was detectable in DDA MS1 survey scans and was quantifiable with sensitivity comparable to that obtained by MS1 scans by DIA, as would be expected. In both DDA and DIA, the intensities of MS1 precursor XICs could be normalized to the most abundant glycoform on each glycosylation site separately, to give a profile of site-specific glycosylation (Figure s3). However, the identity of any precursor ion that was not triggered by DDA for MS2 owing to its relatively low intensity would remain unconfirmed. In contrast, although the number and intensity of target fragment ions detectable by DIA decrease with reduced IgM amount, the glycopeptide-identifying Y1 and Y1f ions were still detectable within the right coeluting time window to validate the presence of the glycopeptide in samples with only 0.5 ng of spiked-in IgM. Figure 4B also indicated that this DIA-based quantitation could reach a dynamic range of 4 orders.

Exploring Nontargeted Glycopeptides Using DIA. The major advantage of DIA over other targeted approaches, such as SRM or PRM, is that DIA files can always be reanalyzed, even if they were not included in the initial targeted list.[21] This is especially beneficial in glycoproteomics, as there is no available way to predict all potential glycoforms on a glycosylation site. This unique advantage was attested to by our unexpected discovery of an unknown modification on IgM N272 glycopeptide "THTN[NIS]E" (Figure 5). The N272 peptide bearing glycan "dHex(1)HexNAc(5)Hex(5)" was first identified by matching the chromatographic peaks of MS2 XICs with MS1 XIC of m/z 925.04^{3+} (Figure 5A). From the MS1 ion map we found that both the sodium and the potassium adducts of the glycopeptide were coeluted at a retention time of around 16 min and were selected simultaneously for MS2 fragmentation with an isolation window of m/z 923.7−939.7 (Figure 5A, lower panel and 5B). Since we had not defined the metal
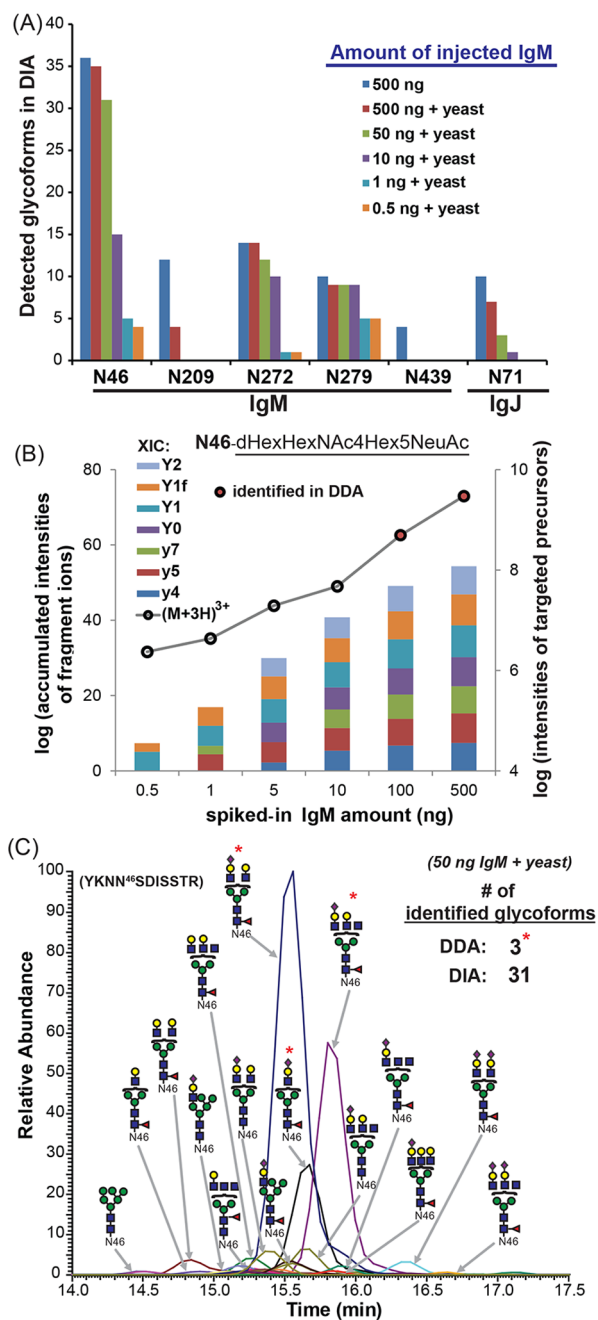
Figure 4. Target analysis by DIA approach provides better sensitivity than DDA. (A) The numbers of successfully detected site-specific IgM/IgJ glycoforms with matching MS1 precursor and MS2 fragment-ion chromatograms extracted from DIA data acquired on from 500 ng of injected IgM, compared with samples containing decreasing amounts of IgM in yeast lysates. (B) Bar chart of accumulated intensities of various fragment ions from the N46 peptide "YKNNSDISSTR", carrying the glycan "dHex(1)HexNAc(4)Hex(5)-NeuAc(1)", in yeast lysates containing increasing amount of IgM. Open circles connected by gray line indicate the XIC peak intensities of that particular N46 glycopeptide precursor in each of the samples analyzed by DIA approach. The circles are colored as red if this glycoform was also identified by the corresponding DDA analysis. (C) Overlaid precursor XICs of all IgM N46 glycoforms detected by DIA analysis of 50 ng of IgM spiked in yeast lysates. Potential glycan structures were annotated for the more abundant peaks. Glycoforms that were also identified by DDA in the same sample are indicated with an asterisk. Among the 31 glycoforms detected by DIA, only the most abundant three were identified by DDA.

adducts as variable modifications for database search, they could not be identified by DDA.

More interestingly, we further noticed that the same set of MS2 XICs in the same swath also afforded well-aligned chromatographic peaks at 17 min, which suggests the presence of another glycoform of this peptide, defined by a mass increment less than 16 Th and hence by a mass corresponding to less than a glycosyl residue from the one eluting at 16 min (Figure 5A upper panel). In fact, the only ion detected at the m/z and retention time region on MS1 ion map was 934.37^{3+}, a 27.99 Da mass increment from the originally targeted N272 glycopeptide (925.04^{3+}). MS2 spectra further confirmed that these two ions shared the same peptide core and N-linked trimannosyl glycan core structure (Figure 5B). The only difference was that fragmentation of the 934.37^{3+} precursor generated not only the usual $Y_0/Y_1/Y_2$ ions but also the product ions of which the mass was increased by 28 Da. The Y_0 + 28 ions suggested that the modification was on the peptide core rather than the glycan. In addition, the occurrence of the glycan Y ions and their +28 counterparts in the same spectrum suggested that the modification could readily fall off, at least in part, during HCD fragmentation. We have no further supporting data to define conclusively this unknown 28 Da modification. More to the point in current work is the fact that, owing to its low abundance, it is less likely for the ion to be selected for MS2 in the conventional DDA approach. Even if this precursor was selected and a good MS2 spectrum was generated, it would still not be identified in a search against a glycan library that did not allow wild-card modifications and hence would remain "undiscovered" either way. A targeted SRM/PRM approach, on the other hand, focuses only on predefined targets and would not consider this unsuspected modified glycoform in the first place. In DDA, what is not acquired is not stored and thus cannot be reinterrogated.

Concluding Summary and Perspectives. Similar to proteomic applications, the advantages of DIA are most obvious when used in a guided manner to assay for target glycopeptides of interest, particularly when the targets are of low signal intensity against an increasingly complex background. We do not anticipate its use as a substitute for DDA but rather as a complement to it, with a different strategic purpose. In fact, the workflow would always start with deriving the relevant information, including the best combination of MS1/MS2 transitions offering the optimum balance of selectivity and sensitivity, from comprehensive DDA analysis of the target glycoprotein. The concept of proteotypic peptides does not apply here. Rather, one is restricted to the specific modified peptides instead of picking from the most readily detected ones representing the proteins of interest. A prerequisite here is that the glycopeptides to be assayed should produce significant Y0/Y1/Y1f ions, ideally along with several peptide b/y fragment ions, to allow unambiguous identification of the peptide core. The IgM glycopeptides assayed in this study met this criterion and we were further able to narrow down the target list (Table s2) by focusing on the best set of nonredundant site-specific glycoforms, taking into accounts their charge states, trypsin miscleavages, and other peptide modifications. We showed that more unique glycoforms can be identified by DIA when the sample amount is limiting in the presence of cell lysates.

A unique feature of site-specific glycoform analysis is that they share common fragment ions. The highly abundant glyco-oxonium ions are obviously without any discriminating value
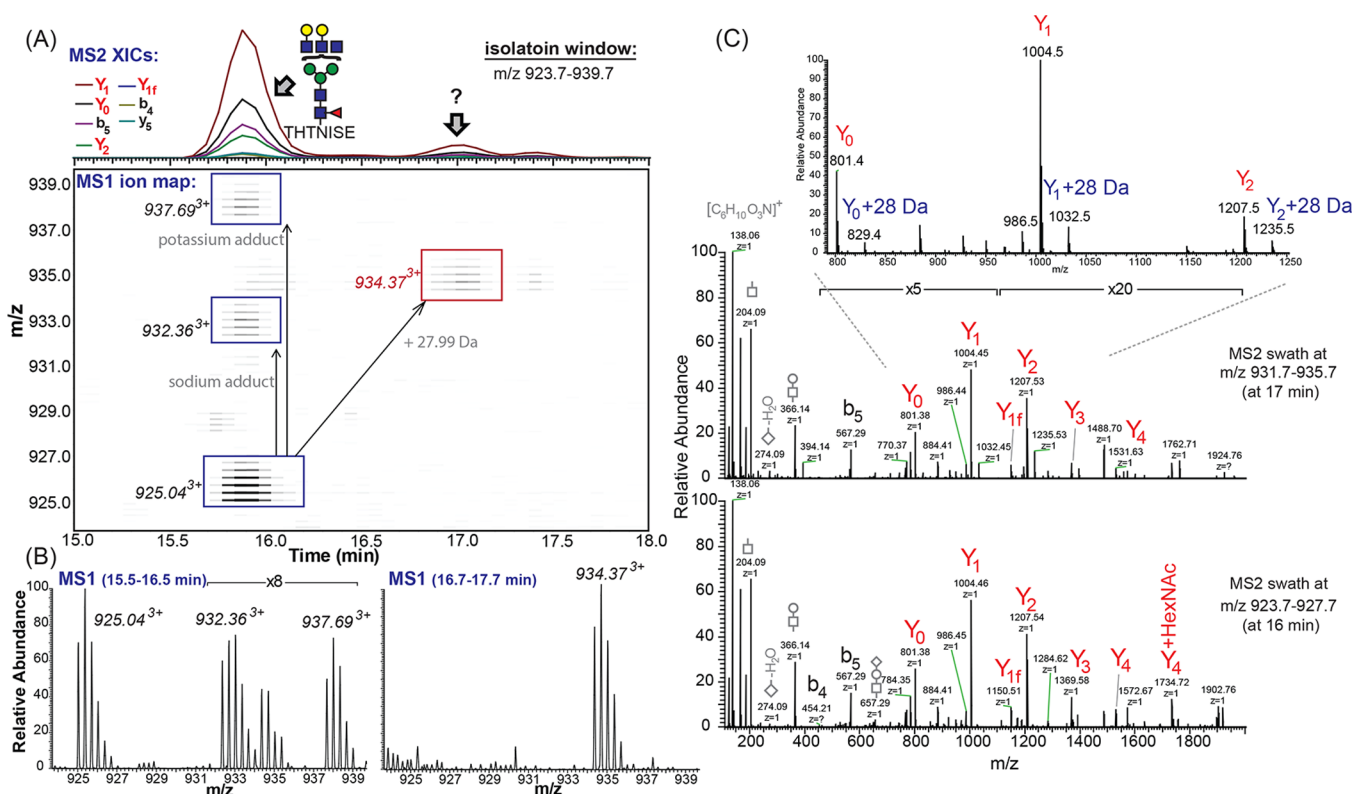
Figure 5. Discovery of untargeted glycopeptide containing unknown modification. (A) MS2 fragment-ion chromatograms of N272 peptide "THTNISE" bearing the glycan "dHex(1)HexNAc(5)Hex(5)". There were extracted from the MS2 swath m/z 923.7−939.7 and aligned with MS1 ion map based on the retention time. Isotopic clusters of different precursors detected in the same m/z region 923−939 that could be coselected were highlighted. (B) MS1 spectra averaged from 15.5 to 16.5 min (left, with region magnified as indicated) and 16.7−17.7 min (right). (C) MS2 spectra acquired from 2 different MS2 swaths at retention time 16 min (lower panel) or 17 min (middle panel) contained almost the same glycan Y ions (labeled in red), but a further zoom-in on the m/z 800−1250 region (upper panel) of the MS2 spectra acquired at 17 min revealed the additional Y0, Y1, and Y2 + 28 Da fragment ions (labeled in blue).

and are only of value for surveying the occurrence of glycopeptides across the mass range and retention time. The common glycan Y ions shared by the glycoforms of the same site impose the requirement that the precursors should have m/z values that fall into different MS2 swaths and/or are chromatographically well resolved. Neither criterion is perfectly met but advances are anticipated, particularly in respect of the attainable swath width. On the other hand, the degenerate nature of the common Y and peptide backbone ions can actually be advantageous for mapping all relevant site-specific glycoforms in a novel postacquisition discovery mode. As shown here, additional glycoforms, be they of genuinely novel structures or ones carrying unexpected modifications or adducts, could be discovered by interrogating the MS1 ion map. This is particularly relevant since no comprehensive glycan library can currently be derived for a DDA identification strategy based on database search. Moreover, such novel forms are often of too low abundance to allow high-quality MS2 with sufficient numbers of product ions to allow confident identification in the first place. There is therefore no alternative but to refer back to MS1 to pick up additional minor glycoforms, which are related to the confidently identified ones by accurate mass differences, for second-tier target analysis. Here, a digital archive is already secured by the DIA approach. Moreover, such a strategy allows one to build up the desirable ion transitions by using more readily available recombinant glycoprotein standards or that isolated from one cell type to search for its glycoform in other cell types and/or at different

pathophysiological stages. It also circumvents the technical difficulties associated with isolating full-length membrane glycoproteins to derive the transitions. We anticipate that herein will lie the most innovative and efficient uses of DIA for glycoproteomics.

■ EXPERIMENTAL SECTION

Purified human serum IgM was purchased from Sigma-Aldrich. All chemicals were also from Sigma-Aldrich unless stated otherwise. Tris(2-carboxyethyl)phosphine (TCEP) was purchased from Thermo Fisher Scientific, sequence grade trypsin from Serva, and Glu-C from Roche Diagnostics. Yeast lysate was a gift from Dr. Uzma Zaman and the laboratory of Reinhard Lührmann, Max Planck Institute for Biophysical Chemistry.

Sample Preparation and Mass Spectrometric Analysis. Filter-aided sample preparation (FASP)[22] was applied to denatured, reduced, alkylated, and digested human serum IgM to obtain the (glyco)peptides for MS analysis. To mimic complex samples, yeast lysate was mixed with varying concentration of IgM digest as described above. A Q Exactive HF (Thermo Fisher Scientific) coupled to the Dionex UltiMate 3000 LC system was used. Details of sample preparation and common settings for LC−MS/MS data-dependent acquisition are provided in the Supporting Information. For data-independent acquisition, each acquisition cycle consisted of one MS1 scan followed by 50 DIA swaths at a fixed isolation width of 16 Th to cover the mass range m/z 800−1600.

Acquisition resolution was set to 30 000 for both MS1 and MS2 scans. The automated gain control (AGC) target value and maximum ion-injection times for MS and MS2 were 1×10^6 in 32 ms and 2×10^5 in 32 ms.

Data-Processing. The MS raw files acquired from all of the LC−MS/MS DDA experiments were searched by Byonic (version 2.3.5)[17] against the human Ig mu chain C region (P01871) and Ig J chain (P01591). All identified peptide-to-spectrum matches (PSMs) were examined manually and filtered by standard criteria before acceptance for comparison (see the Supporting Information for more details). No attempt was made to ensure true positives and/or accurate glycan composition assignment. For DIA analysis, Xcalibur software (Thermo Fisher Scientific) was used for manual extraction and inspection of DIA transitions from identified unique glycopeptides. A 10 ppm of mass tolerance was used for peak extraction. In cases of targeted precursors, when their monoisotopic signals did not allow extraction of well-resolved XICs, other isotopic peaks were examined manually and the one giving the most intense, interference-free signal was selected. The final list of targeted glycopeptides (Table s3) was imported to Skyline software and the XICs of corresponding DIA transitions were generated automatically. Correct peak picking was verified by manual inspection.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.anal-chem.6b04996.

> Supplemental methods; overview of the intensities of generated peptide fragment ions and glycan Y ions from various glycopeptides under different settings of normalized collisional energy (NCE); differences in chromatographic retention time of IgM/IgJ glycopeptides bearing various glycans on the same peptide backbone; site-specific glycosylation on human serum IgM/IgJ; and comparison of the observed retention times for all target glycopeptides in DDA and DIA analysis (PDF)

> List of all peptide-to-spectrum matches reported by Byonic at 1% of false discovery rate (XLSX)

> List of glycopeptides identified by DDA and Byonic search (A) and the further selected DIA target list representing unique site-specific glycoforms to be assayed (XLSX)

> List of targeted glycopeptides identified by DIA at different amount of injected sample in the presence of spiked-in yeast lysates (XLSX)

■ AUTHOR INFORMATION

Corresponding Authors
*Phone: +49-551-201-1060. E-mail: henning.urlaub@mpibpc.mpg.de.
*Phone: +886-2-2785-5696 ext. 6030. E-mail: kkhoo@gate.sinica.edu.tw.

ORCID
Kuan-Ting Pan: 0000-0001-6974-5324

Notes
The authors declare no competing financial interest.

■ REFERENCES

(1) Thaysen-Andersen, M.; Packer, N. H.; Schulz, B. L. *Mol. Cell. Proteomics* **2016**, *15*, 1773−1790.

(2) Nilsson, J. *Glycoconjugate J.* **2016**, *33* (3), 261−272.

(3) Hu, H.; Khatri, K.; Zaia, J. *Mass Spectrom. Rev.* **2016**, DOI: 10.1002/mas.21487.

(4) Lee, L. Y.; Moh, E. S. X.; Parker, B. L.; Bern, M.; Packer, N. H.; Thaysen-Andersen, M. *J. Proteome Res.* **2016**, *15* (10), 3904−3915.

(5) Domon, B.; Aebersold, R. *Nat. Biotechnol.* **2010**, *28* (7), 710−721.

(6) Ting, Y. S.; Egertson, J. D.; Payne, S. H.; Kim, S.; MacLean, B.; Käll, L.; Aebersold, R.; Smith, R. D.; Noble, W. S.; MacCoss, M. J. *Mol. Cell. Proteomics* **2015**, *14* (9), 2301−2307.

(7) Picotti, P.; Aebersold, R. *Nat. Methods* **2012**, *9* (6), 555−566.

(8) Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J. *Mol. Cell. Proteomics* **2012**, *11* (11), 1475−1488.

(9) Sajic, T.; Liu, Y.; Aebersold, R. *Proteomics: Clin. Appl.* **2015**, *9* (3−4), 307−321.

(10) Gillet, L. C.; Navarro, P.; Tate, S.; Rost, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. *Mol. Cell. Proteomics* **2012**, *11* (6), O111.016717.

(11) Chapman, J. D.; Goodlett, D. R.; Masselon, C. D. *Mass Spectrom. Rev.* **2014**, *33* (6), 452−470.

(12) Goldman, R.; Sanda, M. *Proteomics: Clin. Appl.* **2015**, *9*, 17−32.

(13) Hong, Q.; Ruhaak, L. R.; Stroble, C.; Parker, E.; Huang, J.; Maverakis, E.; Lebrilla, C. B. *J. Proteome Res.* **2015**, *14* (12), 5179−5192.

(14) Yeo, K. Y. B.; Chrysanthopoulos, P. K.; Nouwens, A. S.; Marcellin, E.; Schulz, B. L. *Anal. Biochem.* **2016**, *510*, 106−113.

(15) Zacchi, L.; Schulz, B. L. *Mol. Cell. Proteomics* **2016**, *15*, 2435−2447.

(16) Sanda, M.; Goldman, R. *Anal. Chem.* **2016**, *88* (20), 10118−10125.

(17) Bern, M.; Kil, Y. J.; Becker, C. Byonic: Advanced Peptide and Protein Identification Software. In *Current Protocols in Bioinformatics*; 2012; Chapter 13, Unit 13.20; 10.1002/0471250953.bi1320s40.

(18) Segu, Z. M.; Mechref, Y. *Rapid Commun. Mass Spectrom.* **2010**, *24* (9), 1217−1225.

(19) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. *Bioinformatics* **2010**, *26* (7), 966−968.

(20) Egertson, J. D.; Kuehn, A.; Merrihew, G. E.; Bateman, N. W.; MacLean, B. X.; Ting, Y. S.; Canterbury, J. D.; Marsh, D. M.; Kellmann, M.; Zabrouskov, V.; Wu, C. C.; MacCoss, M. J. *Nat. Methods* **2013**, *10* (8), 744−746.

(21) Guo, T.; Kouvonen, P.; Koh, C. C.; Gillet, L. C.; Wolski, W. E.; Röst, H. L.; Rosenberger, G.; Collins, B. C.; Blum, L. C.; Gillessen, S.; Joerger, M.; Jochum, W.; Aebersold, R. *Nat. Med.* **2015**, *21* (4), 407−413.

(22) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. *Nat. Methods* **2009**, *6* (5), 359−362.