analytical chemistry

pubs.acs.org/ac Article

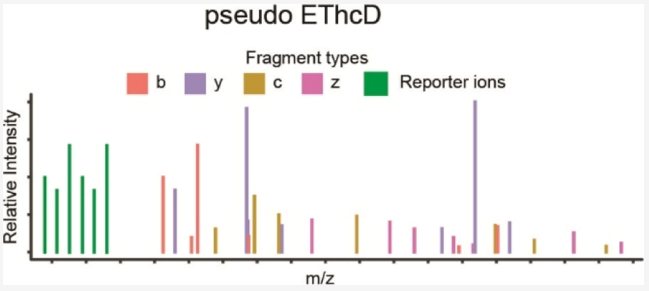
Systematic Evaluation of Fragmentation Methods for Unlabeled and Isobaric Mass Tag-Labeled O-Glycopeptides

Yang Mao,[‖] Shengjun Wang,[‖] Yuanqi Zhao, Andriana Konstantinidi, Lingyu Sun, Zilu Ye,* and Sergey Y. Vakhrushev*

Cite This: *Anal. Chem.* 2021, 93, 11167−11175 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

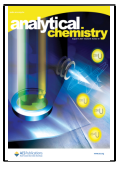
ABSTRACT: Dissecting site-specific functions of O-glycosylation requires simultaneous identification and quantification of differentially expressed O-glycopeptides by mass spectrometry. However, different dissociation methods have not been systematically compared in their performance in terms of identification, glycosite localization, and quantification with isobaric labeling. Here, we conducted this comparison on highly enriched unlabeled O-glycopeptides with higher-energy collision dissociation (HCD), electron-transfer/collision-induced dissociation (ETciD), and electron transfer/higher-energy collisional dissociation (EThcD), concluding that ETciD and EThcD with optimal supplemental activation resulted in superior identification of glycopeptides and unambiguous site localizations than HCD in a database search by Sequest HT. We later described a pseudo-EThcD strategy that *in silico* concatenates the electron transfer dissociation spectrum with the paired HCD spectrum acquired sequentially for the same precursor ions, which combines the identification advantage of ETciD/EThcD with the superior reporter ion quality of HCD. We demonstrated its improvements in identification and quantification of isobaric mass tag-labeled O-glycopeptides and showcased the discovery of the specific glycosites of GalNAc transferase 11 (GALNT11) in HepG2 cells.



Electron-activated dissociation (ExD) mass spectrometry (MS) methods, such as electron capture dissociation (ECD) and electron transfer dissociation (ETD), are a series of fragmentation techniques based on gas phase ion−electron or ion−ion reaction. Distinct from collision-activated dissociation (CAD), which tends to break the weakest bonds within precursor ions, ExD adopts radical-based fragmentation pathways that preserve labile post-translational modifications (PTMs) during peptide sequencing by tandem mass spectrometry(MS/MS).[1] This advantage makes ExD especially useful in characterizing various PTM-enriched proteomes, such as phospho-proteome,[2] glycoproteome,[3] and so on. Nevertheless, the efficiency of ExD to break peptide backbones is much lower than that of CAD, which results in reduced fragment ion abundances and a dominant presence of unreacted and charge-reduced precursor ions in a spectrum.[4] To improve the ExD efficiency, a number of hybrid fragmentation methods have been developed, usually by providing supplemental activation (SA) with orthogonal fragmentation strategies.[1,4−6] For ETD, using low-energy resonant excitation collision-induced dissociation (CID) to activate the charge-reduced precursor ions in an ion trap, electron transfer/collision-induced dissociation (ETciD) has successfully enhanced the overall fragmentation efficiency and been made available on Orbitrap Tribrid mass spectrometers.[7] Alternatively, Frese *et al.* introduced a method that transfers the whole electron−transfer reaction to a higher-energy collision dissociation (HCD) cell, which activates all ion species in the reaction by HCD, referred to as electron transfer/higher-energy collisional dissociation (EThcD).[6]

However, the methodology of improving ExD efficiency by supplemental (collision-based) activation theoretically conflicts with the purpose of preserving labile PTMs by electron-activated, radical-based fragmentations. This is particularly problematic for EThcD since PTM-containing fragment ions from the ETD can be further dissociated by the HCD. On the other hand, quantitative proteomics studies using isobaric tags (*e.g.*, TMT, iTRAQ, and DiLeu) necessitate generating abundant reporter ions, which is less efficient by ExD-based fragmentation pathways.[8] For the same reason, SA of ExD reactions in order to raise quantification quality is intrinsically incompatible with preserving labile PTMs. Consequently, a recent study by Yu *et al.* using EThcD for DiLeu and TMT-based quantitative phospho-proteomics has been focused on

Received: April 21, 2021
Accepted: July 25, 2021
Published: August 4, 2021



https://doi.org/10.1021/acs.analchem.1c01696
Anal. Chem. 2021, 93, 11167−11175

fine-tuning the MS parameters in order to find the balance between neutral loss and reporter ions.[9] However, the applicability of such fine-tuned parameters for even more labile PTMs such as glycosylation has proven tricky.[10]

A highly abundant and complex protein PTM, GalNAc-type O-glycosylation, is widely involved in a variety of physiological and pathological conditions, including proteolytic processing,[11] ligand recognition,[12] signaling transduction,[13,14] immunological response,[15,16] cell adhesion,[17−20] and so on. Dissecting the site-specific function of O-glycosylation is a difficult task and requires overcoming analytical barriers in the site-specific analysis of O-glycoproteome as well as data mining aspects to discern the functionally important from irrelevant glycosites. Due to the labile nature of glycosidic bonds, O-glycosylation is one of the most unstable PTMs during collision-based MS/MS. The O-glycopeptide easily loses its glycan modifications under collision-based dissociation conditions in the form of oxonium ions, which give information of the glycans but also make the identification of the peptide sequence and the modification site difficult. Recent progress in the data processing algorithm has successfully utilized the gas-phase-deglycosylated peptide (Y_0 ion) to identify a potential O-glycopeptide, with glycan compositions corresponding to the observed precursor and the Y_0 ion.[21] However, ExD is still a favorable solution for site-localization when characterizing O-glycoproteomes.[3,22,23] Recent advances in ETD-based MS/MS together with new enrichment methods have allowed the accumulation of 10,197 unambiguously assigned glycosites in 2039 membrane and extracellular proteins[24,25] (http://glycoproteomics.somee.com), partially solving the problem of identifying O-glycosites. In order to correlate specific glycosites to corresponding cellular functions, a differential glycoproteomics strategy combining ETD-based MS/MS with genetic knockout of GalNAc transferases (GalNAc-Ts) that determines the glycosites has been developed.[26−28] TMT-type isobaric labeling has been routinely used in regular quantitative proteomic studies and can in fact facilitate characterizing this differentially expressed glycoproteome. However, neither ExD nor CAD can simultaneously address the issue of identification, site localization, and TMT-based quantification of O-glycopeptides.

Hypothetically, the dilemma of generating reporter ions and preserving O-glycan can be solved if the ExD spectrum can be merged with the CAD spectrum from the same precursor *in silico*. In that case, the ExD spectrum would provide *c*- and *z*-type fragment ions while preserving O-glycans for site identification and the CAD spectrum will contain faithful reporter ion abundances for quantification as well as additional *b*- and *y*-type fragments for identification. Indeed, earlier studies have taken advantage of the complementary fragmentation and applied them in quantitative phosphoproteomics.[29−34] In O-glycoproteomics with isobaric labeling, we have also applied a solution by concatenating exported group of reporter ions from HCD spectra with full ETciD spectra and probed the contribution of individual GalNAc-Ts on a global scale.[26,27,35]

To further optimize MS settings for O-glycoproteomics, we performed a systematic evaluation of multiple fragmentation methods. We first generated a highly enriched O-glycopeptide mixture with the T(Galβ1-3GalNAcα-O-R)- and Tn-(GalNAcα-O-R)-epitopes. Next, we sequenced the enriched O-glycopeptide mixture with HCD, ETciD, and/or EThcD with different SA energies and carefully evaluated their performance in terms of identifications and site localizations. We also created a pseudo-EThcD method for the identification and isobaric tag-based quantification of O-glycopeptides, which relies on alternating HCD and ETciD data acquisition on the same precursor and database search using the *in silico* fused spectra. We show here that the pseudo-EThcD analysis concept outperformed other methods in identifications for shotgun O-glycoproteomics and overcame the missing value quantification issue of real EThcD in a TMT-based quantitative glycoproteomics analysis.

EXPERIMENTAL SECTION

Sample Preparation. HEK293 wildtype (WT) cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and 2 mM L-glutamine. HepG2 cells (WT and GALNT11−/−) were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 2 mM L-glutamine, and 1% nonessential amino acids. Total cell lysates were prepared and Jacalin-based lectin weak affinity chromatography (LWAC) was performed as described previously.[24] In brief, packed cell pellets were lysed in 50 mM ammonium bicarbonate, 0.2% RapiGest SF Surfactant (Waters Corp.), and the lysate was homogenized by sonication. Cleared lysates were diluted in 50 mM ammonium bicarbonate to bring the final concentration of RapiGest below 0.2% before being subjected to reduction with DTT, alkylation with iodoacetamide, and digestion with trypsin (Roche Applied Science). Tryptic digest was purified using Sep-Pak C18 columns (Waters Corp.), and the peptide concentration was measured on a NanoDrop.

For glycopeptide enrichment, we employed LWAC using Jacalin lectin, which is known to recognize the T (Galβ1-3GalNAcα-O-R) antigen. Given that a majority of O-glycan structures may be further modified by sialic acids, which could interfere with LWAC enrichment and hamper glycopeptide identification, neuraminidase treatment was utilized prior to LWAC.

For TMT-based quantitative analysis, equivalent digests were labeled with TMT-6plex (Thermo Scientific) following the protocol described in the product manual.

MS Data Acquisitions. Samples were analyzed using an EASY-nLC 1200 system coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer equipped with a Nanospray Flex ion source (Thermo Fisher Scientific, San Jose, CA). Buffer A consisted of 0.1% formic acid in water and buffer B of 80% acetonitrile, 0.1% formic acid, and 19.9% water. Nano-LC was operated in a single analytical column setup using PicoFrit Emitters (New Objectives, 75 μm inner diameter) packed in-house with Reprosil-Pure-AQ C18 phase (Dr. Maisch, 1.9 μm particle size, 16 cm column length) at a flow rate of 200 nL/min. All samples dissolved in 0.1% formic acid were injected onto the column and eluted in a gradient from 3 to 32% Buffer B in 95 min, from 32 to 100% in 10 min, followed by isocratic elution at 100% for 15 min (total elution time 120 min). The nanospray ion source was operated at a 2.2 kV spray voltage and 275 °C heated capillary temperature. The mass spectrometer was set to acquire full scan MS spectra (350−1700 m/z) for a maximum injection time of 50 ms at a mass resolution of 120,000 and an automated gain control (AGC) target value of 4.0×10^5. The dynamic exclusion was set to 30 s at an exclusion window of 10 ppm with a cycle time of 3 s for all methods. In HCD scans, the AGC target value was set to 5.0×10^4 and the collision energy was 27% in fixed collision

https://doi.org/10.1021/acs.analchem.1c01696
Anal. Chem. 2021, 93, 11167−11175

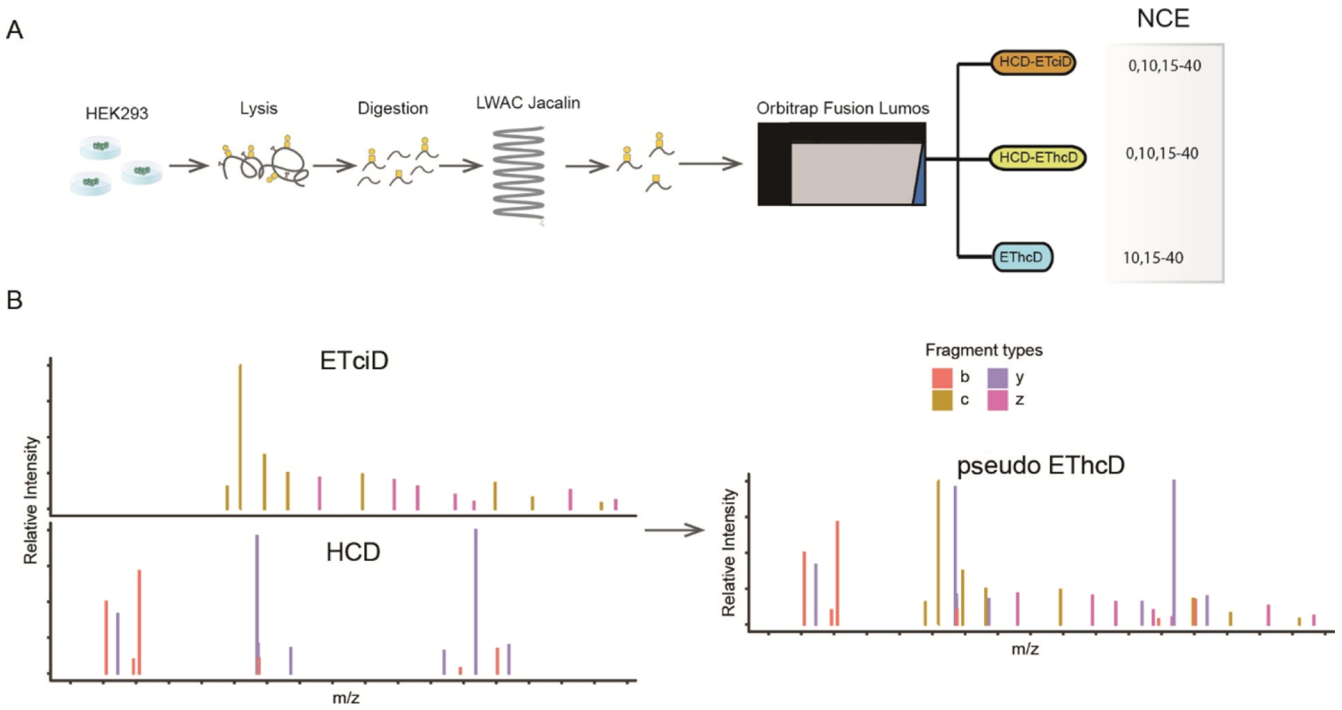


Figure 1. (A) Graphic depiction of the workflow to generate O-glycopeptide samples and MS acquisition strategies. The enrichment of glycopeptides was done with Jacalin-based LWAC. Enriched glycopeptide samples were aliquoted and then analyzed in HCD-ETciD, HCD-EThcD, and EThcD with different SA NCEs either in HCD or in CID. (B) Graphic depiction of the method to generate pseudo-EThcD spectra. ETciD and HCD spectra for the same precursors in the same duty cycle were merged into chimeric spectra designated as pseudo-EThcD spectra.

energy mode. To systematically analyze the behavior of enriched O-glycopeptides, we tested various fragmentation and acquisition strategies: (1) decision tree-based parallel HCD-ETciD fragmentation with different SA normalized collision energies (NCEs), (2) decision tree-based parallel HCD-EThcD fragmentation with different SA NCEs, and (3) EThcD fragmentation with different SA NCEs. In all HCD/ETciD/EThcD scans, the AGC target was set to 5.0×10^4 and the maximum injection time was 75 ms. All MS/MS spectra were acquired in the Orbitrap with resolution at 50,000 in profile mode.

For TMT-labeled HepG2 glycopeptide samples, a HCD-ETciD run and a HCD-EThcD run were acquired. Collision energy for HCD scans was set to 35% in both runs, and SA NCE was 30% for ETciD and 20% for EThcD. All other settings were the same as described above. All the MS data files are available via the ProteomeXchange Consortium with the data set identifier PXD025407.

Data Analysis. MS data processing for all raw files was performed using Proteome Discoverer (PD) version 2.3 software (Thermo Fisher Scientific) and further data analysis was done in R with in-house scripts. For generation of pseudo-EThcD spectra, raw spectra of the same precursors from the same duty cycles were merged in Spectrum Grouper Node and exported as .mzML files in PD to make it compatible with ptmRS and TMT quantification. In the Spectrum Grouper Node, precursor mass criterion was set to same measured mass-to-charge and max RT. The difference was 0.1 min to ensure that only HCD and ETD spectra in the same duty cycle were merged. Raw or .mzML files were searched with Sequest HT search engine against a concatenated human-specific database (UniProt, March 2019, contacting 20,355 canonical entries). Enzyme restriction was set to trypsin digestion with full specificity for a maximum of two missed cleavages. The precursor mass tolerance was set to 10 ppm and fragment ion mass tolerance to 0.02 Da. Carbamidomethylation on cysteine residues was used as a fixed modification. Methionine oxidation and HexNAc, Hex(1)HexNAc(1) attachment to serine, threonine, and tyrosine were used as variable modifications. ptmRS Node was used to determine the probability of glycosite localization.

All the high-confidence peptide-spectrum matches (PSMs) were also exported as annotated spectra. Oxonium ions, unreacted precursor ions, and charge-reduced precursor ions were annotated with in-house R scripts with the following rules:

(1) Seven ions were considered as oxonium ions with an offset of 0.01 Da including 126.055, 138.055, 144.065, 168.066, 186.076, 204.087, and 366.140;

(2) For unreacted precursor peaks

$$\text{lower\_mass} = m/z - 0.02$$

$$\text{upper\_mass} = m/z + 5/\text{charge} + 0.02$$

(3) For charge-reduced precursor peaks

$$\text{lower\_mass} = m/(z - i) - 0.02$$

$$\text{upper\_mass} = m/(z - i) + 5/\text{charge} + 0.02$$

i can be 1, 2, and 3.

For TMT-labeled samples, TMT6plex was set as a static modification on any peptide N-terminus and lysine residues. Reporter Ion Quantifier Node was applied for TMT6plex quantification and the parameter Activation Type was used to choose the corresponding types of reporter ions.

Generation of the GALNT11 Knockout HepG2 Cell Line. Guide RNAs targeting exon2 of GALNT11 were designed and cloned into dual gRNA and Cas9−2A-GFP

https://doi.org/10.1021/acs.analchem.1c01696
Anal. Chem. 2021, 93, 11167−11175

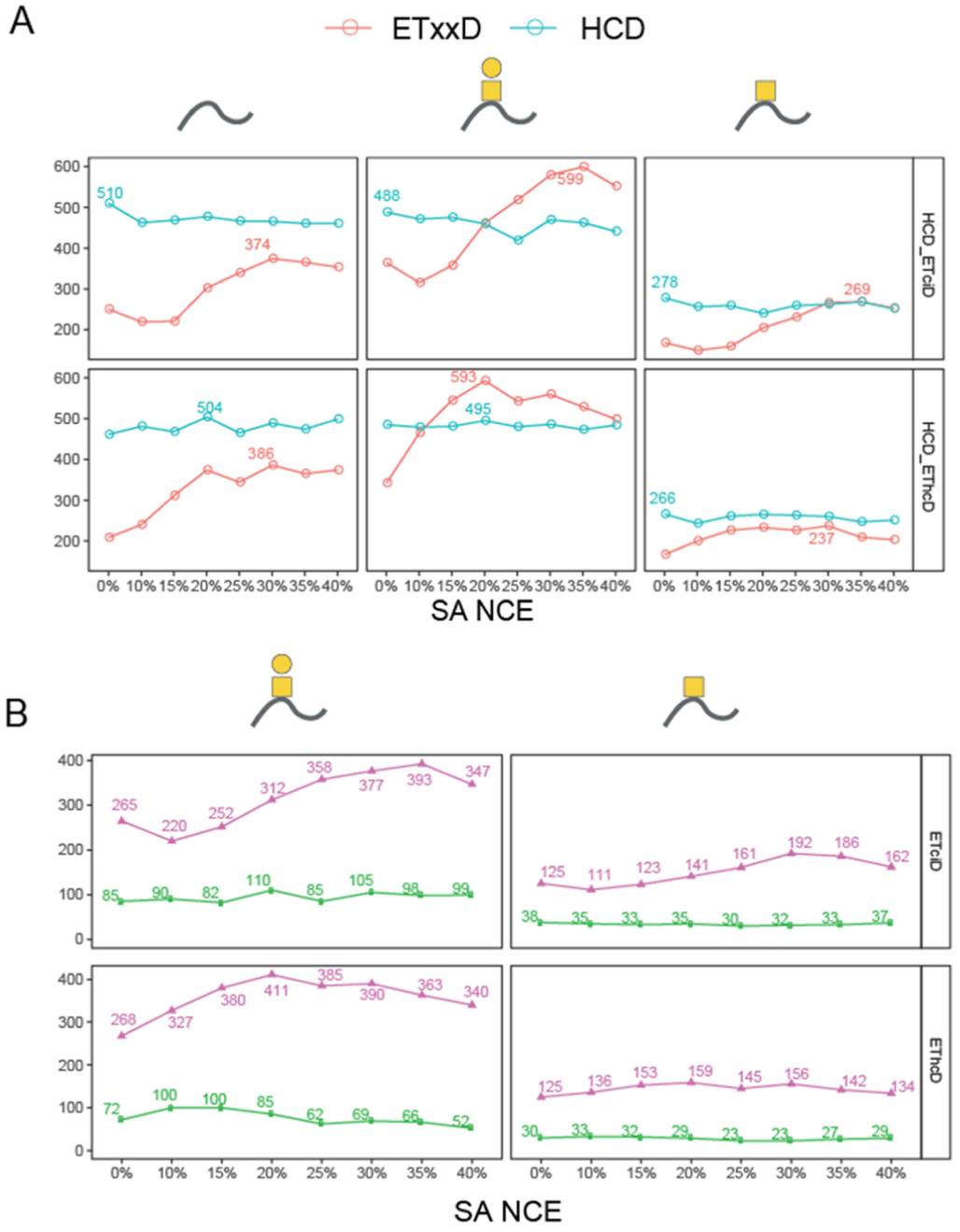


Figure 2. (A) Comparing PSMs of sequential ETD and HCD spectra at different NCEs in HCD-ETciD (top row) and HCD-EThcD (bottom row) runs. Note: the NCE in all HCD scans remained unchanged (27%) in all runs. (B) Comparing PSMs with ambiguous and unambiguous glycosite localization at different NCEs in HCD-ETciD runs (top row) and in HCD-EThcD runs (bottom row). Only ETciD or EThcD spectra of monoglycosylated peptides were subjected to the calculation of ambiguity. Glycosites with ≥ 90% of site probabilities with the ptmRS Node in PD 2.3 were considered as unambiguous sites.

expression vector pX458 (Addgene plasmid no. 48138). gRNA plasmids were transfected into HepG2 cells using Lipofectamine 3000 (Thermo Fisher Scientific) and GFP-tagged Cas9 nuclease was used to enrich for high Cas9 expression by fluorescence-activated cell sorting 48 h after transfection, followed by 1−2 weeks in culture, and cloning by limiting dilution in 96-well plates as recently reported.[36] Clones with frame shift mutations were identified by Indel Detection by Amplicon Analysis (IDAA)[37] using the following primers: GALNT11-F: 5-AGCTGACCGGCAGCAAAATT-GATTCGCTGACTAACTTCACTCT-3; GALNT11-R: 5-GCATTGGGAAATATGAACCCCG-3 (Figure S1). Selected clones were further verified by Sanger sequencing.

RESULTS AND DISCUSSION

Optimizing ETciD and EThcD Parameters for Unlabeled O-Glycopeptides. To find an optimum quantitative proteomics solution for O-glycoproteomics, we first tried different MS settings to maximize O-glycopeptide identifications. As the SA during any ETD process risks breaking labile glycosidic bonds and negatively affecting the identification and site-localization, we first aimed to optimize the NCE to increase the PSMs of O-glycopeptides. Using a Jacalin-

https://doi.org/10.1021/acs.analchem.1c01696
Anal. Chem. 2021, 93, 11167−11175

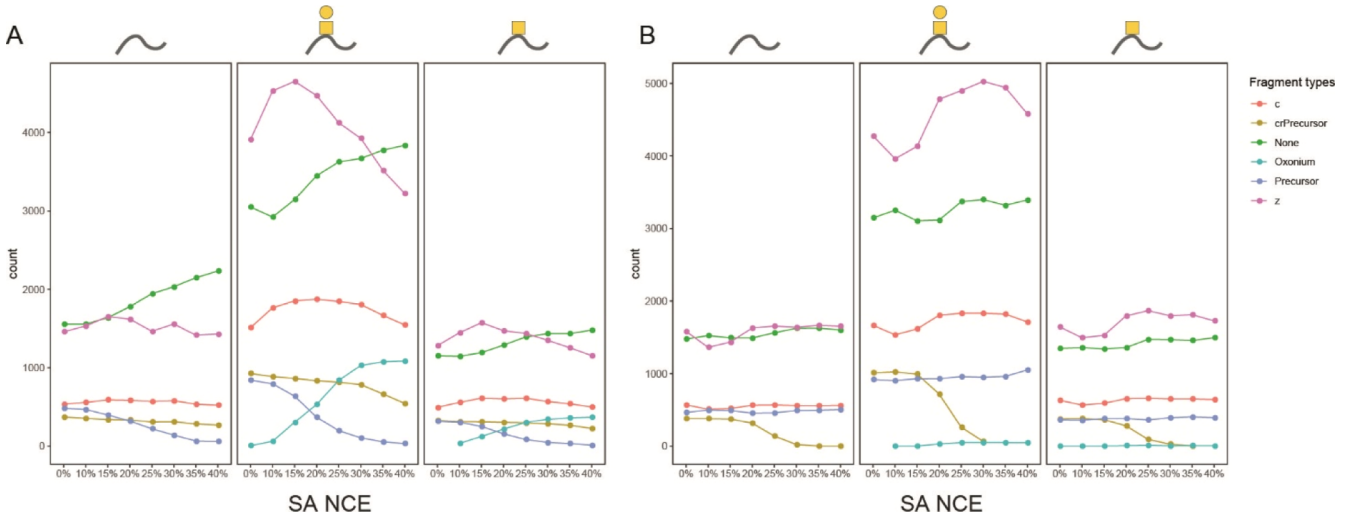


Figure 3. Changes in numbers of different types of fragment ions from shared PSMs at different NCEs in HCD-EThcD runs (A) and HCD-ETciD runs (B). All PSMs were exported as annotated spectra from the PD with default settings. Fragment ions with annotations were then extracted and further annotated. Annotation of *c*- and *z*-ions was inherited from identifications with the Sequest HT search engine. Annotation of charge-reduced precursors (crPrecursor), unreacted precursors, and oxonium ions was done with in-house-written R-scripts. Fragment ions without annotations were calculated as "None". Note: if secondary fragmentations happen on *c*- or *z*-ions on either glycosidic bonds or peptides backbones, they will no longer be defined as *c*- or *z*-ions. Instead, they will be counted as "None".

enriched O-glycoproteome prepared from HEK293 cells, we systematically evaluated three different MS/MS activation methods (HCD, ETciD, and EThcD) over a wide range of collision energies (0, 10, 15, 20, 25, 30, 35, and 40% NCE) (Figure 1A). We set the data collection to have sequential HCD and ETciD or EThcD fragmentation to each precursor ion, so that the performance of both ETciD and EThcD can be gauged against that of HCD regardless of their very different duty cycle times. Additionally, we also tested an EThcD-only method over the same range of collision energies to evaluate the impact of spending extra time on HCD on the overall identification depth. Because Jacalin binds to both of the most common O-glycosylation epitopes, T and Tn antigen, and partially co-eluted naked peptides, we were able to monitor the performance of the different activation methods on three different types of analytes (Supporting Information Note 1, Figure S2). As expected, HCD outperformed both ETciD and EThcD when no or low SA was provided presumably due to inefficient fragmentation of ETD in the absence of SA (ETnoD[1]). The efficiency of both ETciD and EThcD quickly improved along with the increasing NCE. For glycopeptides bearing the T antigen (T-peptides hereafter), it eventually caught up and surpassed that of HCD, confirming the advantage of both ETciD and EThcD in identifying O-glycopeptides. For glycopeptides carrying the Tn antigen (Tn-peptides hereafter), however, HCD kept its identification advantage throughout the energy ramping. The results seemed to suggest that HCD's identification efficiency be offset by collision-induced glycan loss when glycan size increased and to imply that ETD, specifically ETciD, would be a superior method in O-glycopeptides with larger glycan modifications, such as sialyl-T and disialyl-T. Due to the use of neuraminidase in our workflow, we were unable to prove or disprove this hypothesis. However, for both T and Tn peptides, the optimum SA energy was found to be 30 to 35% in ETciD mode and ∼20% in EThcD mode in both HCD-EThcD and EThcD-only runs (Figures 2A, S3). Notably, at their optimum SA condition, ETciD and EThcD demonstrated comparable identification capabilities for T-peptides while ETciD held an advantage for Tn-peptides over EThcD. For our highly enriched glycopeptide sample, we found that sacrificing extra time on HCD data acquisition did not obviously affect the identification depth of both glycopeptides although the PSMs of the co-eluted naked peptides were ∼15% higher when HCD was removed from the data acquisition sequence (Figures 2A, S3).

Although HCD only demonstrated much less T antigen-modified O-glycopeptide identification in comparison to the optimized ETciD and EThcD methods in our study, it is well known that HCD readily loses glycan modification before peptide backbone fragments, resulting in missing information for glycosite localization. Nevertheless, it could still be used for site prediction if there is no positional alternative in peptide sequences. We therefore evaluated the glycosite localization efficiency of both ETciD and EThcD methods by comparing the number of PSMs bearing unambiguous glycosites to that of ambiguous glycosites using the ptmRS node in Proteome Discoverer (Figures 2B, S4). Accordingly, the optimum SA energy found for the maximum PSMs in ETciD and EThcD also produced the best glycosite localization results. For both ETciD and EThcD, the numbers of PSMs with ambiguous glycosites, however, remained at a constant level across the whole tested range for both ETciD and EThcD likely due to the inherent efficiency of ETD and dense distributions of Ser/Thr residues in O-glycopeptides. Interestingly, EThcD produced similar site localization efficiency by this measurement despite the detrimental effect of its HCD component. In addition, both ETciD and EThcD were less efficient in site localization in poly-glycosylated peptides (Figure S4) and the numbers of Ser/Thr residues in peptide sequences played an important role in site localization in ETciD, EThcD, and HCD scans (Figure S5). However, simple counting of the PSM with ambiguous glycosites by ptmRS overlooks the distinct performance of ETciD and EThcD in preserving the labile glycan modifications as database search engines such as Sequest HT do not take into account the intensities of

https://doi.org/10.1021/acs.analchem.1c01696
Anal. Chem. 2021, 93, 11167−11175

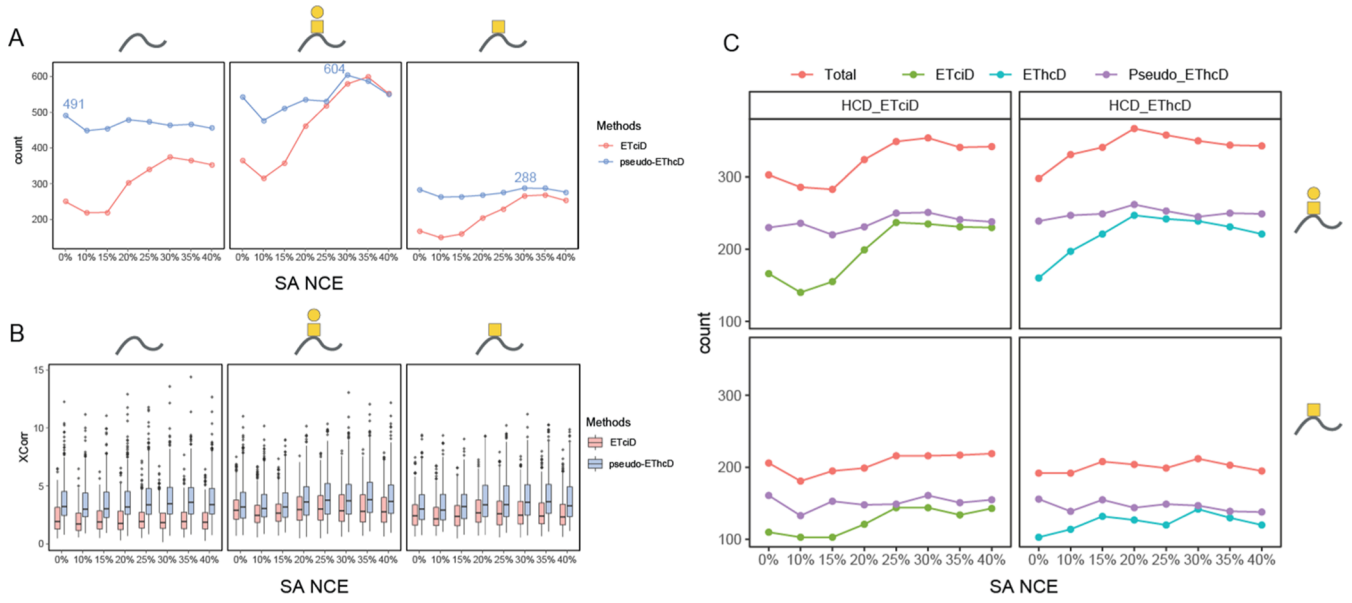


Figure 4. (A) Comparing PSMs of ETciD spectra and pseudo-EThcD spectra at different NCEs in HCD-ETciD runs. (B) Box-plots of XCorr scores of PSMs from ETciD spectra and pseudo-EThcD spectra at different NCEs in HCD-ETciD runs. (C) Comparing unambiguous glycosites of ETciD/EThcD, pseudo-EThcD, and combined HCD-ETciD and HCD-EThcD runs. The total identifications are the number of unique glycosites identified by Pseudo_EThcD methods plus those identified by ETxxD spectra alone. ETxxD stands for ETciD and/or EThcD.

fragment ions. We therefore decided to measure the change of different fragment ions during the SA ramping of ETciD and EThcD.

ETciD Preserves O-Glycosylation Better than EThcD. To make this evaluation, we extracted the fragment ions from all the glycopeptide precursors that were identified in all runs with different SA NCEs. Only one PSM with the highest XCorr score for each glycopeptide from each run was kept. A total of 84 non-glycosylated peptide, 167 T-peptide, and 70 Tn-peptide PSMs were found in the ETciD spectra and 87, 158, and 60 corresponding PSMs in EThcD spectra, respectively. The *c*- and *z*-ions were annotated by a Sequest HT search engine and residual precursor ions and oxonium ions were annotated based on the masses using in-house written R scripts. The residual precursor ions include unreacted precursor ions and charge-reduced precursor ions and they were annotated separately. We found, for EThcD, increasing SA reduced the numbers and intensities of both residual precursor ions and charge-reduced precursor ions, consistent with the mechanism of EThcD that subjects both ions to additional HCD fragmentation. However, this decrease was more pronounced in glycopeptides and was accompanied by steady increase of oxonium ions, the diagnostic fragments produced from the neutral loss of glycan moieties (Figures 3A, S6). Additionally, the ETD type of *c*- and *z*-ions in the glycopeptides initially increased with SA but sharply declined after SA reached ~15% NCE. Because the *c*- and *z*-ions in the non-glycosylated precursors are almost unchanged during the SA ramping (Figure 3A, left panel), these results support a model in which both glycosylated precursors and glycosylated *c*-/*z*-ions lose glycans during secondary HCD activation. Interestingly, the SA-dependent fragment ion depletion in glycopeptides is more pronounced with *z* ions than *c* ions for unknown reasons. Taken together, these data highlighted the detrimental effect of SA to O-glycosylation in EThcD.

On the other hand, for ETciD, although the SA ramping diminished the numbers and intensities of charge-reduced precursor ions and also reduced those of *c*-/*z*-ions after reaching the optimum NCE level, the diagnostic oxonium ions remained at a constant level across the whole SA range (Figure 3B). These results are in agreement with the resonance-based mechanism of SA in ETciD to specifically activate charge-reduced precursor ions. Therefore, our data suggest that ETciD serves better for preserving the labile O-glycosylation. All the annotated spectra that were identified in all HCD-ETciD and HCD-EThcD runs with different strategies are listed in Supporting Information Data 1 and 2.

***In Silico* Merging of the ETD and HCD Spectra.** Having determined the value of ETciD in analyzing O-glycopeptides and optimized parameters to perform it, we next aimed to improve its quantification capability by fusing the in-tandem-acquired HCD spectra with those of ETciD from the same precursors (Figure 1B). We found that the extra time spent on HCD did not significantly affect the identification depth of O-glycoproteome (Figure S3). However, it was uncertain whether such spectra fusion would improve or impair the database searching by a common proteomic search engine. The merged spectra contained *c*- and *z*-ions from the ETciD spectra and *b*- and *y*-ions from the HCD spectra, mimicking EThcD by simultaneously having all four types fragment ions; we, therefore, denominated them as pseudo-EThcD spectra. We then subjected these pseudo-EThcD spectra to the Sequest HT search engine considering *c*-, *z*-, *b*-, and *y*-ions.

Compared with unmodified ETciD, the inclusion of *b*- and *y*-ions in the pseudo-EThcD spectra greatly increased the numbers of PSMs of both O-glycosylated precursors and naked peptides in the low SA range, resembling the superior performance of HCD in this range (Figure 4A). However, in the high SA range (25−40% NCE), the additional *b*- and *y*-ions did not affect the PSMs of O-glycopeptides, which was similar to that of ETciD in this range (Figure 4A). Interestingly, the benefit of *b*-/*y*-ions to PSMs was prominent across the whole NCE range of tested SA on naked peptides. Taken together, our data suggested that the HCD-generated

https://doi.org/10.1021/acs.analchem.1c01696
Anal. Chem. 2021, 93, 11167−11175

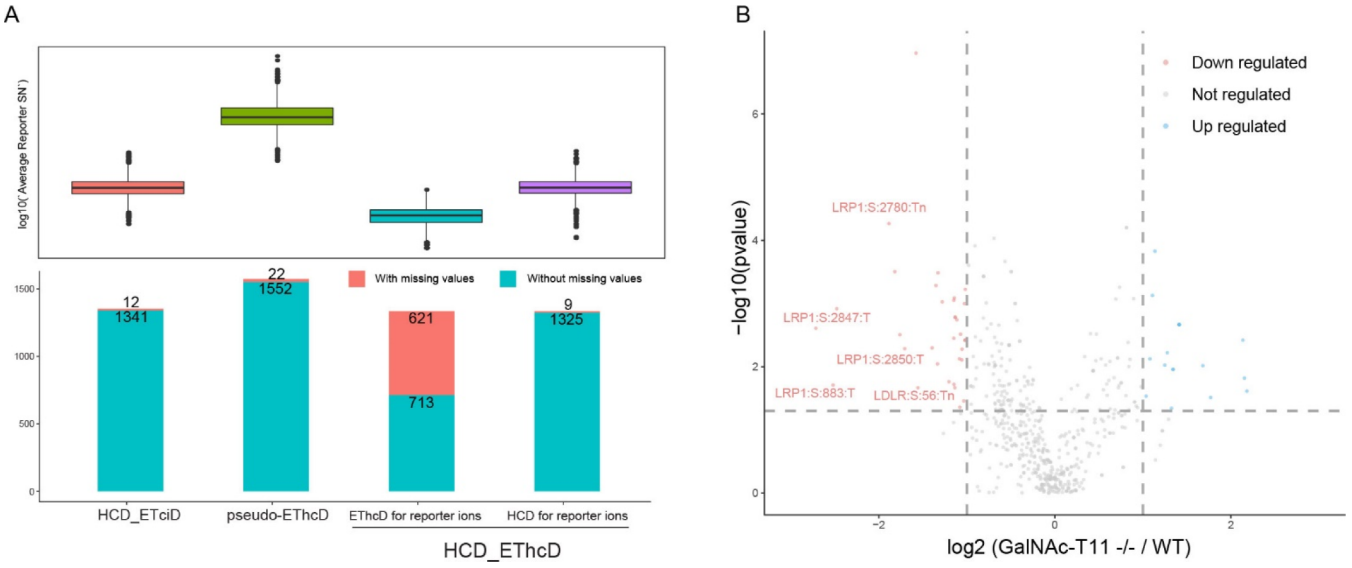


Figure 5. Testing the pseudo-EThcD method in TMT-based differential O-glycoproteomics analysis. (A) Comparing PSMs (lower panel) and reporter ion quality (upper panel) of the pseudo-EThcD strategy to those of ETciD with only reporter ions grafted from HCD spectra, unmodified EThcD, and EThcD with reporter ions grafted from HCD spectra. (B) Volcano plot of pseudo-EThcD-identified and -quantified glycosites in HepG2 cells differentially expressing GALNT11. Downregulated glycosites from the linker region of the LDLR family protein were labeled with the gene name, glycosite, and glycan type (*i.e.*, LRP1:S:2847:T denotes a glycosite with the T-epitope at S2847 in LRP1). Glycosites with fold changes ≥ 1 and *p*-values ≤ 0.05 (Student's *t*-test) were defined as up- or downregulated glycosites.

b-/*y*-ions are largely beneficial to the identification of O-glycopeptides in the low SA range but the advantage diminishes along with the increased SA presumably because the neutral loss on *b*-/*y*-ions neutralizes the benefit of including them in spectrum matches (Figure 4A). Nevertheless, the search strategy with pseudo-EThcD spectra notably improved the XCorr scores at every NCE level, in agreement with previous findings.[33] (Figure 4B, Supporting Information Data 3). We next calculated unambiguous glycosites in unmodified ETciD or EThcD spectra and pseudo-EThcD spectra in both HCD-ETciD and HCD-EThcD runs (Figure 4C). Importantly, the pseudo-EThcD spectra demonstrated higher numbers of unambiguously assigned glycosites than their unmodified counterparts (Figure 4C, pseudo-EThcD *vs* ETxxD), and they also contributed to lots of unique glycosite identifications (Figure 4C, total *vs* ETxxD). In average, the pseudo-EThcD spectra gave rise to 30−40% of unique glycosites in addition to the independent search of the corresponding pairs (Figure 4C, total *vs* ETxxD). This result therefore laid the foundation for evaluating the performance of pseudo-EThcD for isobaric tag-based quantitative O-glycoproteomics as described below.

Analysis of GALNT11-Specific Glycosites Using a Pseudo-EThcD Strategy. Since the pseudo-EThcD spectra contained fused HCD spectra, which has a known advantage in producing reporter ions for isobaric tag-based quantitative proteomics, we conducted a TMT-based differential glycoproteomics workflow to test the pseudo-EThcD strategy for mining specific glycosites of the GalNAc transferase 11-(GALNT11). Our previous studies using dimethyl labeling-based quantitative glycoproteomics revealed that GALNT11 was the only GalNAc-T isoform capable of glycosylating the short linker regions (C6XXXTC1) in LDLR class A (LA) repeats.[12,26,38] However, it was also shown that GALNT11 could glycosylate other substrates such as Notch.[39] It is interesting to test whether an improved, TMT-based quantitative O-glycoproteomic approach could provide us an expanded picture of the substrate specificity of GALNT11. We labeled tryptic peptides from three biological replicates of HepG2 WT cells and three biological replicates of HepG2 GALNT11 knockout (KO) cells with TMT-6 reagents prior to sample pooling and glycopeptide enrichment with Jacalin-based LWAC. The TMT-labeled glycopeptides were then subjected to LC−MS/MS and sequenced by either a sequential HCD-EThcD method or a sequential HCD-ETciD method using the prior optimized parameters. This allowed us to generate pseudo-EThcD spectra from the latter run and make an unbiased comparison of the quantification performance of the pseudo-EThcD with that of an unmodified EThcD.

Similar to the unlabeled glycoproteome, the unmodified EThcD and the pseudo-EThcD produced comparable numbers of glycopeptides PSMs, with the pseudo-EThcD outperforming by ∼20% when both Tn- and T-glycopeptides were considered (Figure 5A, lower panel). However, to our surprise, the unmodified EThcD was found inefficient to generate decent (HCD-type) reporter ions at previously optimized NCE (20%). In the 1334 glycopeptide PSMs from the raw EThcD spectra, 46.6% (621/1334) of them contained missing values in the reporter ions. A recent study by Yu *et al.* discovered that EThcD requires at least 38% NCE to produce quality TMT reporter ions.[9] Therefore, optimizing EThcD for isobaric tag-based quantitative O-glycoproteomics runs into a dilemma of either losing quantification values on glycopeptides with insufficient SA or losing identification and glycosite localization with excessive SA, as we described earlier. On the contrary, as pseudo-EThcD contains reporter ions from the real HCD scan, it completely avoids the missing quantification value issue of the EThcD without risking glycan loss from a higher NCE. Only 22 out of 1574 glycopeptide PSMs from the pseudo-EThcD spectra were found to contain missing values in the reporter ions. The improved quantification capability of the pseudo-EThcD strategy was also evidenced by the significantly

https://doi.org/10.1021/acs.analchem.1c01696
Anal. Chem. 2021, 93, 11167−11175

increased signal-to-noise ratio of the reporter ions (Figure 5A, upper panel).

As one of the major benefits of creating *in silico* pseudo-EThcD spectra is to take in the high-quality reporter ions from the HCD scans, we also tested the picking out of the TMT reporters from HCD scans, combining them with the ETciD spectra, and using only the *c*-/*z*-ions for database search. This alternative approach yielded a similar quantification ratio to the pseudo-EThcD approach, albeit with ∼14% less PSM. Similarly, taking the reporter ions from the HCD scans and using them to replace those in the unmodified EThcD spectra solved the missing quantification value issue of EThcD (Figure 5A, lower panel). In comparison, the pseudo-EThcD strategy produced the best numbers of glycopeptide PSMs among all the tested methods without significant missing quantification values. We therefore used this strategy to analyze the differential O-glycoproteome of HepG2 WT and GALNT11 KO cells.

In a single shot analysis, the pseudo-EThcD strategy allowed us to identify and quantify 581 unambiguously assigned glycosites in the differential HepG2 O-glycoproteome, of which 36 O-glycosites were downregulated in TMT channels corresponding to GALNT11 KO cells. Importantly, we re-discovered eight glycosites in the linker regions of the LDLR family protein with the consensus motif C6XXXTC1, which was previously found to be the preferred substrate of GALNT11 in a large-scale O-glycosite profiling of GALNT11 KO mouse tissues (Figure 5B, Table S1). This result confirmed our workflow as a valid and efficient approach in analyzing differential O-glycoproteomes. At the same time, the other 28 downregulated O-glycosites in our data provided new insights into other possible substrate candidates of GALNT11.

CONCLUSIONS

Currently, the fragmentation technique is still a limiting factor to quantitative O-glycoproteomics. We here systematically evaluated the fragmentation methods for unlabeled O-glycopeptides and concluded that ETciD and EThcD showed comparable performance for the identification of O-glycopeptides. When common database search engines such as Sequest HT were used, both methods displayed identification advantages over HCD with T-peptides and potentially with larger glycan modifications. Recent developments in glycoproteome-oriented database search engines could potentially increase the identification depth of HCD by considering glycan-specific neutral losses. Nevertheless, our results revealed the advantages of ETciD in preserving O-glycan modifications and unambiguous site localizations in comparison to EThcD and HCD. Hence, in our evaluation, parameter-optimized ETciD is a preferred MS/MS method for characterizing O-glycoproteomes.

Moreover, at least with our highly enriched O-glycoproteome sample, sacrificing instrument time for the pseudo-EThcD strategy generated more informative MS/MS spectra and preserved the glycan moieties on peptide fragments. For TMT-labeled O-glycopeptides, it resulted in an improved performance in terms of identification and quantification.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.1c01696.

Additional results in enrichment specificity of Jacalin-based LWAC, EThcD-only results, XCorr distribution, relative intensities of different fragment ions, and annotated ETciD spectra from shared PSMs (PDF)

Quantified and unambiguously assigned glycosites in TMT-labeled glycopeptides with the pseudo-EThcD method (XLSX)

AUTHOR INFORMATION

Corresponding Authors

Zilu Ye − *Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine, Faculty of Health Sciences, University of Copenhagen, Copenhagen 2200, Denmark;* Present Address: Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; Email: zilu.ye@cpr.ku.dk

Sergey Y. Vakhrushev − *Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine, Faculty of Health Sciences, University of Copenhagen, Copenhagen 2200, Denmark;* orcid.org/0000-0002-0418-5765; Email: seva@sund.ku.dk

Authors

Yang Mao − *School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou 510006, China; Guangdong Provincial Key Laboratory of Drug Non-Clinical Evaluation and Research, Guangzhou 510990, China;* orcid.org/0000-0002-9674-1158

Shengjun Wang − *School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou 510006, China; Guangdong Provincial Key Laboratory of Drug Non-Clinical Evaluation and Research, Guangzhou 510990, China*

Yuanqi Zhao − *School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou 510006, China*

Andriana Konstantinidi − *Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine, Faculty of Health Sciences, University of Copenhagen, Copenhagen 2200, Denmark*

Lingyu Sun − *School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou 510006, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.1c01696

Author Contributions

‖Y.M. and S.W. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (91853131, 81872786, and 31971213), Guangdong Provincial Key Laboratory of Drug Nonclinical Evaluation and Research (2018B030323024), Guangdong Basic, Applied Basic Research Foundation (2019A1515010647), the Fundamental Research Funds for the Central Universities (19ykzd27), and the Danish National Research Foundation (DNRF107).

REFERENCES

(1) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. *J. Anal. Chem.* **2007**, *79*, 477−485.

(2) Swaney, D. L.; Wenger, C. D.; Thomson, J. A.; Coon, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 995−1000.
(3) Levery, S. B.; Steentoft, C.; Halim, A.; Narimatsu, Y.; Clausen, H.; Vakhrushev, S. Y. *Biochim. Biophys. Acta* **2015**, *1850*, 33−42.
(4) Liu, C.-W.; Lai, C.-C. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 57−66.
(5) Campbell, J. L.; Hager, J. W.; Blanc, J. C. Y. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1672−1683.
(6) Frese, C. K.; Altelaar, A. F. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. *Anal. Chem.* **2012**, *84*, 9668−9673.
(7) Makarov, A.; Grinfeld, D.; Ayzikov, K. *Fundamentals and Applications of Fourier Transform Mass Spectrometry*; Elsevier, 2019; pp 37−61.
(8) Phanstiel, D.; Zhang, Y.; Marto, J.; Coon, J. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1255−1262.
(9) Yu, Q.; Shi, X.; Feng, Y.; Kent, K. C.; Li, L. *Anal. Chim. Acta* **2017**, *968*, 40−49.
(10) Chen, Z.; Yu, Q.; Hao, L.; Liu, F.; Johnson, J.; Tian, Z.; Kao, W. J.; Xu, W.; Li, L. *Analyst* **2018**, *143*, 2508−2519.
(11) Goth, C. K.; Halim, A.; Khetarpal, S. A.; Rader, D. J.; Clausen, H.; Schjoldager, K. T.-B. G. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 14623−14628.
(12) Wang, S.; Mao, Y.; Narimatsu, Y.; Ye, Z.; Tian, W.; Goth, C. K.; Lira-Navarrete, E.; Pedersen, N. B.; Benito-Vicente, A.; Martin, C.; Uribe, K. B.; Hurtado-Guerrero, R.; Christoffersen, C.; Seidah, N. G.; Nielsen, R.; Christensen, E. I.; Hansen, L.; Bennett, E. P.; Vakhrushev, S. Y.; Schjoldager, K. T.; Clausen, H. *J. Biol. Chem.* **2018**, *293*, 7408−7422.
(13) Takeuchi, H.; Haltiwanger, R. S. *Biochem. Biophys. Res. Commun.* **2014**, *453*, 235−242.
(14) Hofmann, B. T.; Schlüter, L.; Lange, P.; Mercanoglu, B.; Ewald, F.; Fölster, A.; Picksak, A.-S.; Harder, S.; El Gammal, A. T.; Grupp, K.; Güngör, C.; Drenckhan, A.; Schlüter, H.; Wagener, C.; Izbicki, J. R.; Jücker, M.; Bockhorn, M.; Wolters-Eisfeld, G. *Mol. Cancer* **2015**, *14*, 109.
(15) Vichier-Guerre, S.; Lo-Man, R.; Huteau, V.; Dériaud, E.; Leclerc, C.; Bay, S. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3567−3570.
(16) von Mensdorff-Pouilly, S.; Kinarsky, L.; Engelmann, K.; Baldus, S. E.; Verheijen, R. H.; Hollingsworth, M. A.; Pisarev, V.; Sherman, S.; Hanisch, F.-G. *Glycobiology* **2005**, *15*, 735−746.
(17) Lasky, L. A.; Singer, M. S.; Dowbenko, D.; Imai, Y.; Henzel, W. J.; Grimley, C.; Fennie, C.; Gillett, N.; Watson, S. R.; Rosent, S. D. *Cell* **1992**, *69*, 927−938.
(18) Gill, D. J.; Tham, K. M.; Chia, J.; Wang, S. C.; Steentoft, C.; Clausen, H.; Bard-Chapeau, E. A.; Bard, F. A. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, E3152−E3161.
(19) Zhang, L.; Tran, D. T.; Ten Hagen, K. G. *J. Biol. Chem.* **2010**, *285*, 19491−19501.
(20) Zhang, L.; Ten Hagen, K. G. *Biochem Soc Trans* **2011**, *39*, 378.
(21) Chalkley, R. J.; Medzihradszky, K. F.; Darula, Z.; Pap, A.; Baker, P. R. *Mol. Omics* **2020**, *16*, 147−155.
(22) Riley, N. M.; Malaker, S. A.; Driessen, M. D.; Bertozzi, C. R. *J. Proteome Res.* **2020**, *19*, 3286−3301.
(23) Čaval, T.; Zhu, J.; Heck, A. J. *Anal. Chem.* **2019**, *91*, 10401−10406.
(24) Ye, Z.; Mao, Y.; Clausen, H.; Vakhrushev, S. Y. *Nat. Methods* **2019**, *16*, 902−910.
(25) Čaval, T.; de Haan, N.; Konstantinidi, A.; Vakhrushev, S. Y. *Curr. Opin. Struct. Biol.* **2021**, *68*, 135−141.
(26) Narimatsu, Y.; Joshi, H. J.; Schjoldager, K. T.; Hintze, J.; Halim, A.; Steentoft, C.; Nason, R.; Mandel, U.; Bennett, E. P.; Clausen, H.; Vakhrushev, S. Y. *Mol. Cell. Proteomics* **2019**, *18*, 1396−1409.
(27) Bagdonaite, I.; Pallesen, E. M.; Ye, Z.; Vakhrushev, S. Y.; Marinova, I. N.; Nielsen, M. I.; Kramer, S. H.; Pedersen, S. F.; Joshi, H. J.; Bennett, E. P. *EMBO Rep.* **2020**, *21*, No. e48885.
(28) Schjoldager, K. T.; Joshi, H. J.; Kong, Y.; Goth, C. K.; King, S. L.; Wandall, H. H.; Bennett, E. P.; Vakhrushev, S. Y.; Clausen, H. *EMBO Rep.* **2015**, *16*, 1713−1722.
(29) Nielsen, M. L.; Savitski, M. M.; Zubarev, R. A. *Mol. Cell. Proteom.* **2005**, *4*, 835−845.
(30) Savitski, M. M.; Nielsen, M. L.; Kjeldsen, F.; Zubarev, R. A. *J. Proteome Res.* **2005**, *4*, 2348−2354.
(31) Molina, H.; Matthiesen, R.; Kandasamy, K.; Pandey, A. *Anal. Chem.* **2008**, *80*, 4825−4835.
(32) Kim, S.; Mischerikow, N.; Bandeira, N.; Navarro, J. D.; Wich, L.; Mohammed, S.; Heck, A. J. R.; Pevzner, P. A. *Mol. Cell. Proteom.* **2010**, *9*, 2840−2852.
(33) Kim, M.-S.; Zhong, J.; Kandasamy, K.; Delanghe, B.; Pandey, A. *Proteomics* **2011**, *11*, 2568−2572.
(34) Yang, F.; Wu, S.; Stenoien, D. L.; Zhao, R.; Monroe, M. E.; Gritsenko, M. A.; Purvine, S. O.; Polpitiya, A. D.; Tolić, N.; Zhang, Q.; Norbeck, A. D.; Orton, D. J.; Moore, R. J.; Tang, K.; Anderson, G. A.; Paša-Tolić, L.; Camp, D. G., 2nd; Smith, R. D. *Anal. Chem.* **2009**, *81*, 4137−4143.
(35) Hintze, J.; Ye, Z.; Narimatsu, Y.; Madsen, T. D.; Joshi, H. J.; Goth, C. K.; Linstedt, A.; Bachert, C.; Mandel, U.; Bennett, E. P.; Vakhrushev, S. Y.; Schjoldager, K. T. *J. Biol. Chem.* **2018**, *293*, 19064−19077.
(36) Lonowski, L. A.; Narimatsu, Y.; Riaz, A.; Delay, C. E.; Yang, Z.; Niola, F.; Duda, K.; Ober, E. A.; Clausen, H.; Wandall, H. H.; Hansen, S. H.; Bennett, E. P.; Frödin, M. *Nat. Protoc.* **2017**, *12*, 581.
(37) Yang, Z.; Steentoft, C.; Hauge, C.; Hansen, L.; Thomsen, A. L.; Niola, F.; Vester-Christensen, M. B.; Frödin, M.; Clausen, H.; Wandall, H. H.; Bennett, E. P. *Nucleic Acids Res.* **2015**, *43*, No. e59.
(38) Tian, E.; Wang, S.; Zhang, L.; Zhang, Y.; Malicdan, M. C.; Mao, Y.; Christoffersen, C.; Tabak, L. A.; Schjoldager, K. T.; Ten Hagen, K. G. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 25196−25202.
(39) Boskovski, M. T.; Yuan, S.; Pedersen, N. B.; Goth, C. K.; Makova, S.; Clausen, H.; Brueckner, M.; Khokha, M. K. *Nature* **2013**, *504*, 456−459.

https://doi.org/10.1021/acs.analchem.1c01696
Anal. Chem. 2021, 93, 11167−11175