

# Electron-based dissociation is needed for O-glycopeptides derived from OpeRATOR proteolysis

Nicholas M. Riley<sup>1</sup>, Stacy A. Malaker<sup>1</sup>, Carolyn R. Bertozzi<sup>1,2\*</sup>

<sup>1</sup>Department of Chemistry and Stanford ChEM-H, Stanford University, Stanford, California, USA

<sup>2</sup>Howard Hughes Medical Institute, Stanford, California, USA

\*Correspondence should be to C.R.B., email: bertozzi@stanford.edu

## ABSTRACT

The recently described O-glycoprotease OpeRATOR presents exciting opportunities for O-glycoproteomics. This bacterial enzyme purified from *Akkermansia (Sp). muciniphila* cleaves N-terminally to serine and threonine residues that are modified with (preferably asialylated) O-glycans. This provides orthogonal cleavage relative to canonical proteases (e.g., trypsin) for improved O-glycopeptide characterization with tandem mass spectrometry (MS/MS). O-glycopeptides with a modified N-terminal residue, such as those generated by OpeRATOR, present several potential benefits, perhaps the most notable being *de facto* O-glycosite localization without the need of glycan-retaining fragments in MS/MS spectra. Indeed, O-glycopeptides modified exclusively at the N-terminus would enable O-glycoproteomic methods to rely solely on collision-based fragmentation rather than electron-driven dissociation because glycan-retaining peptide fragments would not be required for localization. The caveat is that modified peptides would need to reliably contain only a single O-glycosite. Here we use methods that combine collision- and electron-based fragmentation to characterize the number of O-glycosites that are present in O-glycopeptides derived from OpeRATOR digestion of four known O-glycoproteins. Our data show that over 50% of O-glycopeptides generated from combined digestion using OpeRATOR and trypsin contain multiple O-glycosites, indicating that collision-based fragmentation alone is not sufficient. Electron-based dissociation methods are necessary to capture the O-glycopeptide diversity present in OpeRATOR digestions.

## INTRODUCTION

Mucin-type O-glycosylation is a prevalent post-translational modification on extracellular and secreted proteins that drives both biochemical and biophysical interactions at the cell surface.<sup>1-5</sup> This important modification, characterized by an initiating  $\alpha$ -N-acetylgalactosamine ( $\alpha$ -GalNAc) monosaccharide on serine and threonine residues, is challenging to study due to several inherent features. These include non-template driven elaboration of the  $\alpha$ -GalNAc residue into four major core structures, the lack of a well-defined sequence motif, and occurrence in densely glycosylated regions rich in serine and threonine residues.<sup>6</sup> The combination of these attributes leads to O-glycosylated sequences that are heterogeneous both in the glycosites that are occupied and the glycans that modify them, requiring site-specific characterization. Tandem mass spectrometry (MS) is the premier method for O-glycosite mapping, but O-glycopeptides are largely intractable using standard glycoproteomic approaches, which have focused almost entirely on N-glycans. To this end, recent efforts have focused on improving O-glycopeptide analyses,<sup>7-10</sup> ranging from developments in sample preparation,<sup>11-15</sup> data acquisition,<sup>16-20</sup> and post-acquisition data analysis.<sup>21-25</sup>

One promising avenue to emerge from these investigations is the description of O-glycoproteases that can generate O-glycopeptides more amenable to MS characterization. These can be grouped into two main classes: 1) mucinases, such as StcE from *E. coli*, that specifically recognize mucin-type O-glycosylated serine and threonine residues in densely glycosylated mucin domains;<sup>26,27</sup> and 2) the broadly acting O-glycoprotease OpeRATOR from *A. muciniphila*, which appears to recognize any mucin-type O-glycan modified serine or threonine with minor preferences for flanking residues. OpeRATOR cleaves N-terminally to O-glycosylated serine and threonines, generating O-glycopeptides that retain the modified residue at their N-terminus.<sup>28-30</sup> In the handful of studies to employ OpeRATOR thus far, collision-based fragmentation, namely higher-energy collisional dissociation (HCD), has been used to identify O-glycopeptides. The operating presumption in these studies is that site-specific O-glycosite localization can be reported by attributing the entire glycan mass identified for the peptide to an O-glycan modifying the N-terminal serine or threonine.

We recently compared dissociation methods for glycosite localization in both N- and O-glycopeptides.<sup>31</sup> Our unequivocal conclusion, one that is shared by others,<sup>9,16</sup> was that electron-based dissociation methods, such as electron transfer dissociation (ETD) and ETD with supplemental HCD activation (EThcD), are necessary for O-glycopeptide characterization. Our conclusions support several studies that said, OpeRATOR-derived O-glycopeptides differ in that an O-glycosite is localized at the N-terminus by the nature of the enzyme. In order to preclude the need for electron-based fragmentation, however, O-glycopeptides from OpeRATOR proteolysis must not have missed cleavages. Missed cleavages by this definition would mean that internal O-glycosites also exist within the peptide sequence rather than exclusively at the N-terminus, ultimately leading to misattribution of a portion of the identified glycan aggregate mass to the N-terminal O-glycosite.

Here we investigate the number of internal O-glycosites that exist in addition to N-terminal O-glycosites in O-glycopeptides derived from OpeRATOR proteolysis. Through a combination of product-dependent methods to collect HCD and ETD/EThcD spectra and our newly developed O-Pair Search approach specifically designed to localize O-glycosites,<sup>32</sup> we analyze O-glycopeptides generated through sequential OpeRATOR and trypsin digestion of O-glycoprotein standards (fetuin, monocyte differentiation antigen [CD14], fibronectin, and P-selectin glycoprotein ligand 1 [PSGL-1]). Our data show that over half of OpeRATOR-derived O-

glycopeptide spectra harbor multiple glycosites. As such, O-glycosite localization exclusively at N-terminal residues cannot be assumed, rendering collisional activation inadequate for site-specific analysis. Our data underscore the need for electron-driven dissociation in methods aiming to characterize O-glycopeptides, even those generated using OpeRATOR.

## EXPERIMENTAL

A mixture of O-glycopeptides generated by sequential OpeRATOR, PNGaseF, and trypsin digestion was generated using four glycoproteins: bovine fetuin (alpha-2-HS-glycoprotein, P12763), recombinant human CD14 (P08571), human fibronectin (P02751), and recombinant human P-selectin glycoprotein ligand 1 (PSGL1) (Q14242). OpeRATOR proteolysis was coupled with sialidase co-treatment. Following digestion and desalting, peptides were combined in equal parts by mass for the four proteins and analyzed by 90-min LC-MS/MS product-dependent methods constructed using ETD and EThcD triggered scans, i.e., HCD-pd-ETD and HCD-pd-EThcD methods.<sup>33–36</sup> All raw data were searched using O-Pair Search implemented in MetaMorpheus (0.0.308), which is available at <https://github.com/smith-chem-wisc/MetaMorpheus>.<sup>32</sup> A glycan database representing 12 common O-glycans was used, and the “Maximum OGlycan Allowed” setting was set to 5 unless otherwise noted. Non-specific digestion was enabled for peptides ranging from 5 to 25 residues. Only Level 1 and 1b identifications were retained for further analysis, meaning all identifications had spectral evidence in ETD or EThcD spectra for localized glycosites (or had only one plausible glycoform). For glycan modifications shown as H#N#, H represents the number of hexose residues (galactose in O-glycans), and N represents the number of *N*-acetylhexosamines, which can be *N*-acetylgalactosamine or *N*-acetylglucosamine in O-glycans. Spectral annotation was aided by the Interactive Peptide Spectral Annotator (IPSA, <http://www.interactivepeptide-spectralannotator.com>).<sup>37</sup> Graphs were generated using OriginPro 2018 with the exception of the alluvial diagram, which was generated using the RAWGraphs web app, licensed under CC BY-NC-SA 4.0.<sup>38</sup> WebLogo was used for Logo plot generation (<https://weblogo.berkeley.edu/logo.cgi>).<sup>39</sup> More details are available in the **Supporting Information**.

## RESULTS AND DISCUSSION

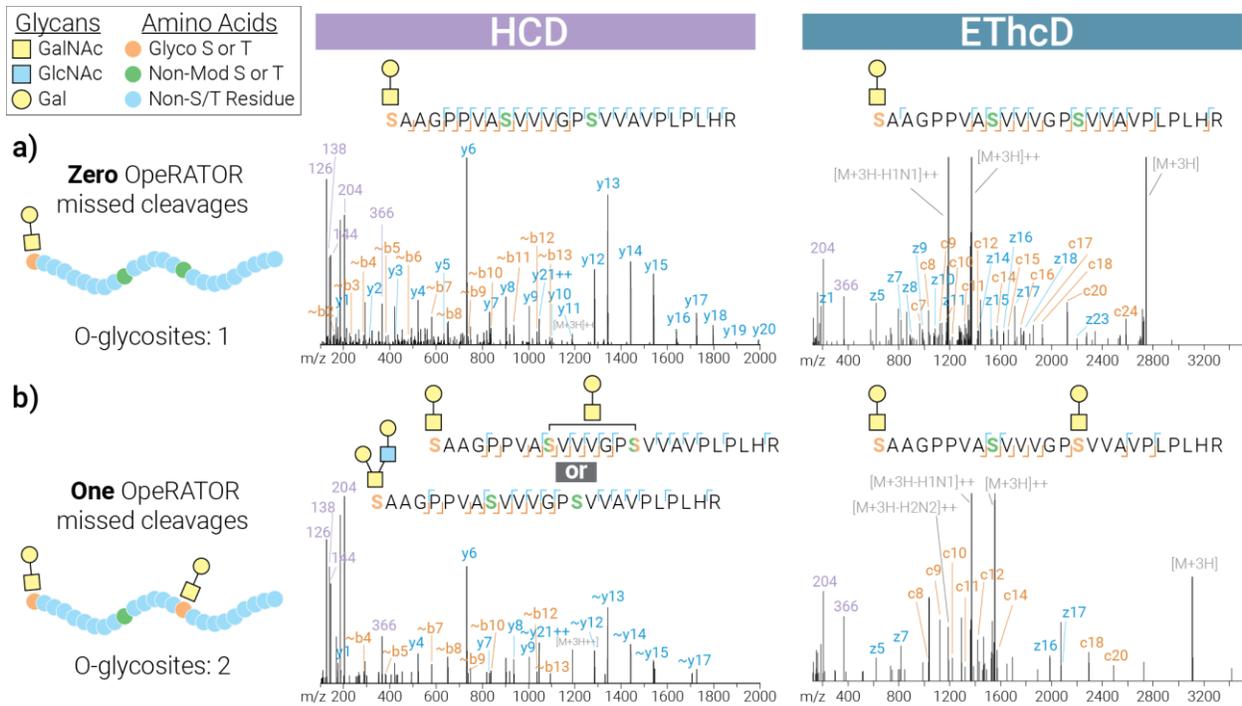
O-glycosite localization in OpeRATOR-derived O-glycopeptides has thus far relied on HCD fragmentation and the assumption that the total glycan mass observed can be attributed to a single glycosylated serine or threonine at the N-terminus. Here we investigated the reliability of this approach for O-glycosite localization by digesting four glycoproteins with known O-glycosites using OpeRATOR in combination with trypsin. OpeRATOR has been used for both in-solution digests and in chemoenzymatic methods that use solid-supports to facilitate release of O-glycopeptides upon cleavage with OpeRATOR (O-GIG from Cipollo and co-workers, and EXoO from Zhang and co-workers).<sup>28–30</sup> Here we elected to perform in-solution digests to characterize O-glycosites that occur as expected at the N-terminal site of cleavage, as well as potential missed cleavage events that occur both up-sequence and down-sequence from the cleavage site. Missed cleavage events that are up-sequence (i.e., N-terminal) to OpeRATOR cleavage sites may go undetected in solid-support-base methods because they remain attached to the solid support following O-glycopeptide elution via OpeRATOR cleavage.

The extent of missed cleavages directly relates to the utility of collision-based fragmentation for O-glycosite localization. If OpeRATOR cleavage efficiency is not close to 100% (resulting in internal O-glycosites), collision-based fragmentation is not a reliable tool for O-glycosite

localization in OperATOR-centric methods, and electron-driven dissociation must be used. This point is explicitly discussed by Zhang and co-workers when describing their EXoO method,<sup>28</sup> and Cipollo and co-workers reported a significant number of multiply O-glycosylated peptides following OperATOR digestion both in-solution and using solid-support methods.<sup>29</sup> Even so, collision-based dissociation was used to collect most of the data in both studies. In this work, OperATOR proteolysis was coupled with co-incubation of SialEXO (a sialidase mix that is recommended by the manufacturer to improve OperATOR cleavage efficiency due to preference of asialylated O-glycans), and N-glycans were removed with PNGaseF. Product-dependent ETD and EThcD methods were used to collect HCD and ETD/EThcD spectra of putative glycopeptide precursor ions, and spectra were analyzed using O-Pair Search, which provides localization levels to indicate confidence in glycosite localization. Only Level 1 and Level 1b data are reported here, meaning all O-glycosites were supported by spectral evidence (the majority of the identifications) or had only one plausible glycoform, and identifications were filtered to be sequences expected from OperATOR cleavage.

**Figure 1** provides an example of why collision-based fragmentation is insufficient for O-glycosite localization in OperATOR-derived O-glycopeptides and instead why electron-driven dissociation is necessary. An O-glycopeptide from fetuin, <sup>282</sup>SAAGPPVASVVVGPSVVAVPLPLHR<sup>306</sup>, represents two well-known O-glycosites, Ser-282 and Ser-296, in addition to Ser-290, which was identified as O-glycosylated using OperATOR in the EXoO workflow.<sup>28</sup> Two glycopeptidofoms identified using an HCD-pd-EThcD25 method include the top example, with an O-glycosylated Ser-282 (orange) and unmodified Ser-290 and Ser-296 (green) (**Figure 1a**), and the bottom example, with O-glycosylated Ser-282 and Ser-296 (orange) and unmodified Ser-290 (green) (**Figure 1b**). In the top example, b-type peptide fragment ions in the HCD spectrum did not retain the O-glycan (as denoted by “~”), but the EThcD spectrum confirmed that the assumption of an N-terminal Ser-282 modified with H1N1 is correct. Here HCD was adequate, albeit less than ideal, for glycosite localization; the H1N1 glycan mass is presumably a single core-1 GalNAc-Gal, a common mucin-type O-glycan, but O-mannose (a hexose) can modify similar regions as O-GalNAc. on the same glycan (albeit less likely to be the case).<sup>40</sup> The doubly O-glycosylated peptide in **Figure 1b**, however, demonstrates the issue with relying on HCD. The total O-glycan mass observed is H2N2. Neither b- nor y-type fragments retain glycan modifications in the HCD spectrum (marked with “~”), leaving ambiguity as to which serine residues were modified. OperATOR digestion affords the assumption of the N-terminal serine as glycosylated, but H2N2 is equally as plausible of a glycan to observe at a single O-glycosite as H1N1. Even with the assumption that the N-terminal serine has the same H1N1 modification as the top example, ambiguity remained as to whether Ser-290 or Ser-296 was modified with the other H1N1 O-glycan. On the other hand, the EThcD spectrum in the bottom provided c- and z<sup>•</sup>-type fragment ions that unambiguously localized the H1N1 modification to both Ser-282 and Ser-296.

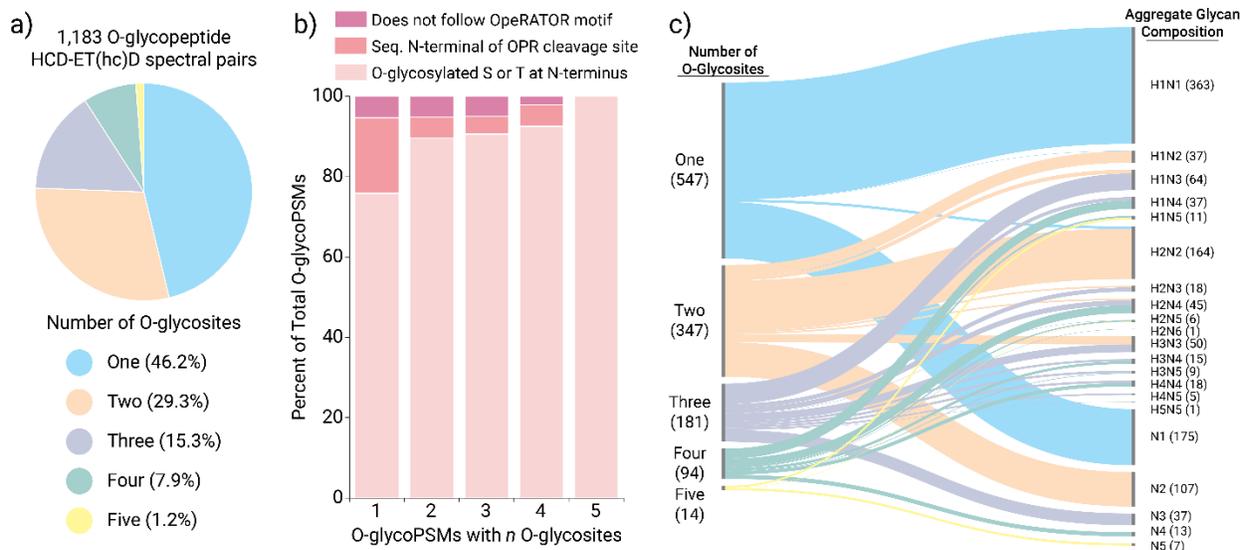
To understand the prevalence of multiply O-glycosylated peptides present in OperATOR digests, we generated a dataset comprising four different product-dependent methods with electron-driven dissociation as the triggered scan: HCD-pd-ETD, HCD-pd-EThcD15, HCD-pd-EThcD25, and HCD-pd-EThcD35 (collected in technical triplicate, 12 raw files total). **Figure 2a** provides the proportion of the total O-glycopeptide spectral pair identifications that had one or multiple localized O-glycosites. Note that the O-Pair Search returns a single identification representing two spectra, both an HCD and ETD/EThcD spectrum.<sup>32</sup> Approximately 46% of the peptides contained a single O-glycosite, indicating that over half of the identifications harbored two or more O-glycosites. We



**Figure 1. Bovine fetuin sequence  $^{282}$ SAAGPPVASVVGPSVVAVPLPLHR $^{306}$  is a tangible example of the benefits of electron-driven dissociation.** Diagrams at the left present two glycopeptidoform possibilities of this sequence, which have both O-glycosylated (orange) and non-modified (green) serine and threonine residues. These two O-glycopeptidoforms were identified in our dataset using HCD-EThcD spectral pairs, including (a) a canonical OpeRATOR-derived O-glycopeptide with a single O-glycosite at the N-terminus and (b) a doubly glycosylated O-glycopeptide from a missed OpeRATOR cleavage that has both an N-terminal O-glycosite site and a second internal O-glycosite.

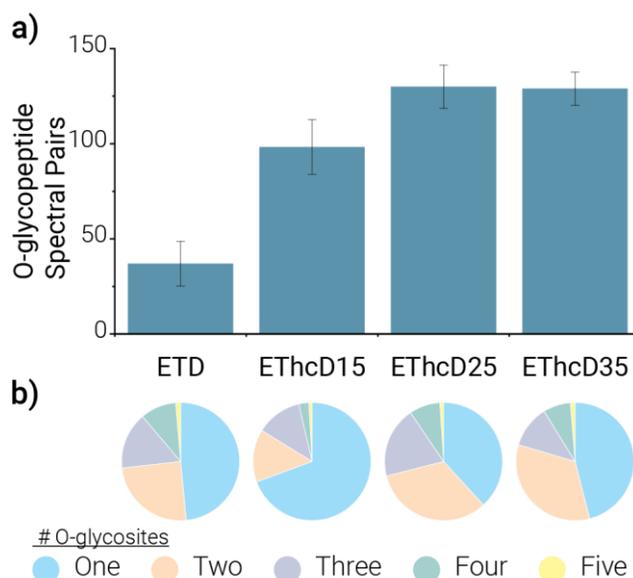
also note that our data were generated using a sialidase mixture to produce asialylated O-glycans, which is known to aid in OpeRATOR digestion. The number of missed cleavages and multiply O-glycosylated peptides would likely be even higher if sialidase treatment was omitted. Our data also corroborates a recent reanalysis of the EXoO dataset using MSFragger-Glyco (a new search algorithm that functions similarly to the O-Pair Search strategy used here), where the authors observed both a substantial number of missed OpeRATOR cleavages and larger total glycan mass modifications that indicate multiple O-glycosites.<sup>41</sup> Polasky *et al.* also identified co-occurrence of phosphorylation sites on O-glycosylated sequences, furthering the need for electron-driven dissociation to properly localize multiple sites of modification.

**Figure 2b** examines what proportion of identifications represent canonical OpeRATOR O-glycopeptides (downstream/C-terminal from the cleavage site with an O-glycosylated N-terminal residue) versus those that are sequences upstream/N-terminal from the OpeRATOR cleavage site. For O-glycopeptides with one O-glycosite, ~75% of identifications were canonical OpeRATOR peptides with a single N-terminal O-glycosylated serine or threonine residue, while nearly 20% of identified O-glycopeptides were from sequences that were upstream from an OpeRATOR cleavage site. This indicates that OpeRATOR missed cleavages do occur on either side of O-glycosites, and that some O-glycosylated peptides can evade analysis if they are retained on support matrices in solid-support approaches. Search strategies that only consider peptides starting with serine or threonine in their searches, such as those used in previous



**Figure 2. Characteristics of O-glycopeptides derived from OpeRATOR proteolysis.** **a)** A total of 1,183 O-glycopeptide spectral pairs with confidently localized glycosites were detected using HCD-pd-ETD and HCD-pd-EThcD methods. The pie graph shows the proportion of identifications that had one or multiple O-glycosites. **b)** The bar graph delineates the proportion of O-glycopeptide spectral pairs that can be grouped into three main classes: 1) identifications that have an O-glycosylated serine or threonine at the N-terminus (light pink; bottom), 2) O-glycopeptides from sequences upstream (i.e., N-terminal) of the OpeRATOR (OPR) cleavage site that indicate a missed cleavage (pink; middle), and 3) O-glycopeptides that do not appear to follow the OpeRATOR cleavage motif (dark pink; top). **c)** The alluvial diagram relates the aggregate glycan compositions that were identified on O-glycopeptides with a varying number of localized O-glycosites. Numbers in parentheses provide the number of O-glycopeptide spectral pairs identified for each condition. For glycan modifications shown as H#N#, H represents the number of hexose residues (galactose in O-glycans), and N represents the number of *N*-acetylhexosamines, which can be *N*-acetylgalactosamine or *N*-acetylglucosamine in O-glycans.

OpeRATOR studies, would either be blind to these upstream sequences or could mis-assign their glycosites. We also created a Logo plot to examine the protein sequence regions surrounding OpeRATOR-generated O-glycopeptides (**Figure S1**). Similar to the EXoO study,<sup>28</sup> we observed a prevalence of proline residues flanking OpeRATOR cleavage sites, although the enriched proline residues at P1 and P3' were less pronounced in this dataset. Interestingly, **Figure S1** clearly shows serine and threonine residues occurring 1 to >10 positions away from the cleavage site (both up- and downstream), supporting our observations of multiple O-glycosites. For multiply O-glycosylated sequences, the majority (~90%) contained an O-glycosylated N-terminal serine or threonine in addition to an internal O-glycosite(s). Approximately 48% of all identifications came from the mucin domain of PSGL-1, which could contribute to the proportion of multiply glycosylated sequences observed. That said, mucin O-glycoproteins were discussed as major components of the previous OpeRATOR datasets, making O-glycopeptides from densely O-glycosylated mucins a mainstay rather than an exception when digesting with OpeRATOR. Lastly, a small percentage of O-glycopeptides passed filtering criteria, yet did not follow the expected OpeRATOR motif, instead showing only an internal O-glycosite (**Figure 2b**). These could be the presence of some non-glycosylation-dependent activity of OpeRATOR at serine and threonine residues or due to other non-specific cleavage (e.g., chemical degradation or from trace amounts of unknown proteases).



**Figure 3. Performance summary for ETD and EThcD methods and OpeRATOR cleavage motif analysis.** **a)** The average number of O-glycopeptide spectral pair identifications with confidently localized O-glycosites is shown for four different product-dependent methods utilizing ETD and EThcD fragmentation. Numbers after EThcD provide normalized collision energies used, and error bars show one standard deviation. **b)** Pie graphs show the proportion of identifications that had one or multiple O-glycosites for the total pool of identifications from each method.

The alluvial diagram in **Figure 2c** maps the aggregate glycan masses that were identified in O-glycopeptides harboring a varying number of O-glycosites. The only two total glycan masses that are exclusive to singly O-glycosylated peptides are N1 and H1N1, meaning identification of any other total glycan mass using HCD fragmentation can lead to ambiguity in O-glycosite assignment. As noted above, H2N2 is one aggregate glycan mass that could represent one glycan at a single glycosite or two glycans split between two sites. Here the majority of H2N2 glycans were the result of two O-glycosites rather than the total glycan mass modifying a single residue. Once multiple glycosites were considered on a peptide, we saw a substantial increase in the glycoforms (i.e., glycosites per peptide) that can explain an observed aggregate glycan mass, highlighting the need to localize O-glycans to specific residues with electron-driven methods. Presence of sialic acids, beyond the reduced OpeRATOR efficiency discussed above, would further complicate this issue because it drastically increases the number of possible glycan structures that can be observed, as noted by Polasky et al.<sup>41</sup> One potential option to eliminate the presence of core-2 structures like H2N2 would be to evaluate oxonium ion ratios, which can indicate the presence of GlcNAc residues rather than exclusively GalNAc.<sup>31,42</sup>

These data make it clear that methods like ETD and EThcD are necessary for O-glycopeptide analysis, even when using OpeRATOR. The next logical question then becomes: which electron-driven method is best? **Figure 3a** compares the average number of O-glycopeptide spectral pairs identified, showing that HCD-pd-EThcD25 and HCD-pd-EThcD35 were largely comparable. This is similar to our previous studies of O-glycopeptides derived from sequential digestion with the mucinase StcE and trypsin.<sup>31,32</sup> Interestingly, the proportion of O-glycopeptide identifications that harbored a varying number of O-glycosites differed somewhat between the methods (**Figure 3b**).

ETHcD35 showed proportions most similar the dataset on the whole, while ETHcD15 identified a significantly higher percentage of singly O-glycosylated peptides relative to sequences with two or more O-glycosites. This may be because multiply O-glycosylated peptides have lower charge density than singly O-glycosylated peptides, and supplemental energy at 15 nce was not sufficient to generate c- and z<sup>\*</sup>-type fragments for localizing multiple O-glycosites. We note that the EXoO study did use ETD to investigate the cleavage motif of OpeRATOR, but they reported identification of ETD spectra for precursors with charge state  $\geq 3$  only. By comparison, approximately a quarter to a third of our identifications were  $z = 2$  (**Figure S2**). Reliance on ETD without supplemental activation may have limited the number of multiply modified O-glycopeptides they could identify, especially without inclusion of lower charge density precursor ions.

We followed this analysis with collection of a collisional dissociation-centric dataset to understand the prevalence of presumably singly O-glycosylated peptides (i.e., N1 and H1N1 modified) compared to presumably multiply O-glycosylated peptides (i.e., those with other aggregate glycan masses). We tested eight different product-dependent HCD and stepped collision energy HCD (sceHCD) methods (**Figure S3**) and saw fairly constant performance across all HCD/sceHCD methods (**Figure S3a**). Because HCD largely does not generate glycan-retaining peptide fragments, identifications were reported without glycosite localization (Level 3) or having only one plausible glycoform (Level 1b). Nearly all of the identifications (97.2%) had the potential for multiple glycoforms (**Figure S3b**), indicating that ambiguous glycosite localization issues need to be considered. Approximately 54% of all identifications were indicated to have a single O-glycosite (**Figure S3c**), although this is based on possible glycan mass combinations that were considered in the search rather than spectral evidence. Thus, even based on aggregate glycan mass alone, a substantial portion of identifications in HCD-only methods contain two or more O-glycosites. Additionally, of those identifications purported to have one O-glycosite, only 60% had glycan modification masses that exclusively indicate a single O-glycan (i.e., N1 or H1N1). In combination with **Figure 3c** above, this means that 40% of the O-glycosylated identifications reported to be singly glycosylated (without localization evidence) are likely to have ambiguous or incorrect O-glycosite assignment if relying on the assumption of a N-terminal O-glycosite. **Figure S3d** provides the proportions of total identifications from each HCD or sceHCD method that had total glycan modifications of N1, H1N1, or some other glycan combination. Even though HCD-only methods for OpeRATOR O-glycopeptides can generate quality identifications with correctly localized O-glycosites harboring N1 and H1N1 glycans, reliance exclusively on HCD fragmentation clearly leaves room for ambiguity and fails to properly identify a significant proportion of O-glycopeptides from the sample.

Finally, **Figure S2** compares the charge state distributions for HCD/sceHCD and ETD/ETHcD methods. A concern of using ETD-based methods is a bias against lowly charged and low charge density precursor ions, which can be common for O-glycopeptides due to their glycan modifications and serine/threonine-rich sequences. Our data show that collision-based dissociation and ETHcD methods can generate relatively similar distributions of precursor ion charge states amongst identified O-glycopeptides (matching observations from StcE-trypsin O-glycopeptides<sup>31</sup>), although the HCD/sceHCD methods provide a slight advantage in identification of low charge density species. The benefit of ETHcD methods in boosting low charge state identifications (and identification numbers on the whole) over ETD is important and indicates that other hybrid ETD methods<sup>43</sup> may also be well suited for such applications.

## CONCLUSIONS

As glycoproteomics gains popularity, a major challenge of the field is establishing accepted methods for site-specific analysis. Tandem mass spectrometry (MS/MS) is currently the main platform used for glycosite localization. Collision-based fragmentation, specifically beam-type dissociation, is ubiquitous in proteomic methods, including those used to localize sites of phosphorylation, acetylation, and ubiquitylation. Its benefits include decades of mechanistic studies on a variety of analytes to understand its gas-phase chemistry, ease of implementation on diverse instrument platforms, and rapid scan acquisition speeds. Application of collisional dissociation is less straightforward for glycosite localization in intact glycopeptides due to glycan heterogeneity, although it is largely sufficient for singly glycosylated N-glycopeptides (excluding multiply N-glycosylated sequences and those with both N- and O-glycosites).<sup>31,44</sup> A clear exception to the utility of beam-type collisional activation is site-specific analyses of O-glycopeptides. Here, it is common place to have multiple potential O-glycosites that can be modified by a heterogeneous pool of labile O-glycans. O-glycosylation instead requires alternative MS/MS fragmentation methods, namely electron-driven dissociation such as electron capture dissociation (ECD), electron transfer dissociation methods (ETD) and their derivate methods.<sup>45-</sup>

47

Nevertheless, approaches that could make site-specific O-glycopeptide analysis amenable to collisional-dissociation-centric methods would be valuable. One potential benefit of the recently described O-glycoprotease OpeRATOR is the reported ability to localize O-glycosites using collisional dissociation due to the N-terminal location of modified sites within the sequence. Here we investigated how often O-glycopeptides generated from OpeRATOR proteolysis contain only one N-terminal O-glycosite versus multiple glycosites. We show that over half of O-glycopeptides from OpeRATOR digestion contain multiple glycosites, where the total glycan mass addition cannot be solely attributed to the N-terminal residue (**Figure 2a**). This alone indicates that collision-based fragmentation (e.g., HCD) is not sufficient for site-specific characterization of OpeRATOR O-glycopeptides. Additionally, we identified cases where total glycan mass that could ostensibly be attributed to a single O-glycosite was actually distributed amongst multiple O-glycosites in OpeRATOR O-glycopeptides, as confirmed by electron-based fragmentation (**Figure 1 and 2c**).

It is worth noting here that these data do not invalidate the results of previous studies using OpeRATOR and HCD-only methods. Their results appear to contain quality identifications and confidently localized O-glycosites, especially considering the high proportion of reported H1N1 modifications. Creative alterations to workflows can improve O-glycosite localization for HCD-centric methods, e.g., the recently reported EXoO-Tn study that can indicate the number of O-glycans present in the sequence using heavy labeled monosaccharides.<sup>30</sup> Yet, ambiguities can still remain. Rather than discard previous work, this study points out 1) ambiguity is possible when using OpeRATOR for O-glycosite localization in a considerable number of identifications when solely utilizing HCD, and 2) that many multiply O-glycosylated peptides generated by OpeRATOR proteolysis go undetected when HCD-only methods are used to presume a single O-glycosite at the peptide N-terminus. As such, it is clear that use of the O-glycoprotease OpeRATOR is a valuable tool for O-glycosite mapping, but methods relying on this approach cannot forgo electron-based dissociation, especially EThcD methods, for localizing O-glycosites.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge support from Howard Hughes Medical Institute and National Institute of Health (NIH) Grant R01 CA200423 awarded to C.R.B. N.M.R. was funded through an NIH Predoctoral to Postdoctoral Transition Award (Grant K00 CA21245403). S.A.M. was supported by an NIH F32 Postdoctoral Fellowship (F32-GM126663-01).

## NOTES

C.R.B. is a co-founder and Scientific Advisory Board member of Lycia Therapeutics, Palleon Pharmaceuticals, Enable Bioscience, Redwood Biosciences (a subsidiary of Catalent), and InterVenn Biosciences, and a member of the Board of Directors of Eli Lilly & Company.

## DATA AVAILABILITY

Raw data files, the fasta file used in this study, and results for both ETD/ETHcD and HCD/sceHCD datasets have been deposited to the ProteomeXchange Consortium via the PRIDE<sup>48</sup> partner repository with the dataset identifier PXD020077.

## SUPPORTING INFORMATION

The following Supporting Information is available for this manuscript in an associated document:

Supplemental Experimental Methods

Supplemental References

Supplemental Figures:

**Figure S1.** Logo plot generated using O-glycopeptides generated from OpeRATOR proteolysis.

**Figure S2.** Charge state distributions of O-glycopeptide identifications from various methods

**Figure S3.** HCD and sceHCD methods for analyzing OpeRATOR O-glycopeptides

## REFERENCES

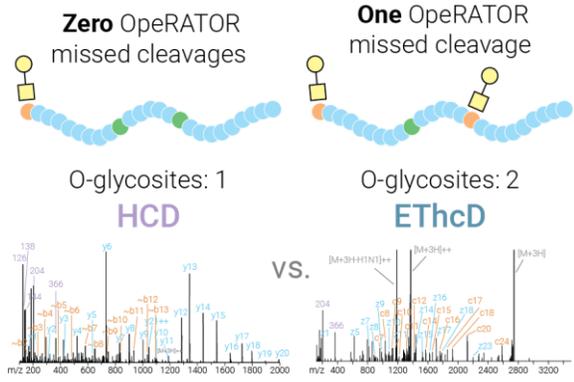
- (1) Möckl, L. The Emerging Role of the Mammalian Glycocalyx in Functional Membrane Organization and Immune System Regulation. *Front. Cell Dev. Biol.* **2020**, *8*, 253. <https://doi.org/10.3389/fcell.2020.00253>.
- (2) Woods, E. C.; Kai, F.; Barnes, J. M.; Pedram, K.; Pickup, M. W.; Hollander, M. J.; Weaver, V. M.; Bertozzi, C. R. A Bulky Glycocalyx Fosters Metastasis Formation by Promoting G1 Cell Cycle Progression. *Elife* **2017**, *6*. <https://doi.org/10.7554/eLife.25752>.
- (3) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in Health and Disease. *Nature Reviews Nephrology*. Nature Publishing Group June 1, 2019, pp 346–366. <https://doi.org/10.1038/s41581-019-0129-4>.
- (4) Kuo, J. C. H.; Gandhi, J. G.; Zia, R. N.; Paszek, M. J. Physical Biology of the Cancer Cell Glycocalyx. *Nat. Phys.* **2018**, *14* (7), 658–669. <https://doi.org/10.1038/s41567-018-0186-9>.
- (5) Paszek, M. J.; Dufort, C. C.; Rossier, O.; Bainer, R.; Mouw, J. K.; Godula, K.; Hudak, J. E.; Lakins, J. N.; Wijekoon, A. C.; Cassereau, L.; et al. The Cancer Glycocalyx Mechanically Primes Integrin-Mediated Growth and Survival. *Nature* **2014**, *511* (7509), 319–325. <https://doi.org/10.1038/nature13535>.
- (6) Brockhausen, I.; Stanley, P. Chapter 10 O-GalNAc Glycans. *Essentials Glycobiol.* **2017**, *1*, 1–9. <https://doi.org/10.1101/glycobiology.3e.010>.
- (7) Khoo, K. H. Advances toward Mapping the Full Extent of Protein Site-Specific O-GalNAc Glycosylation That Better Reflects Underlying Glycomic Complexity. *Current Opinion in Structural Biology*. Elsevier Ltd June 1, 2019, pp 146–154. <https://doi.org/10.1016/j.sbi.2019.02.007>.

- (8) You, X.; Qin, H.; Ye, M. Recent Advances in Methods for the Analysis of Protein O-Glycosylation at Proteome Level. *Journal of Separation Science*. January 2018, pp 248–261. <https://doi.org/10.1002/jssc.201700834>.
- (9) Darula, Z.; Medzihradzsky, K. F. Analysis of Mammalian O-Glycopeptides - We Have Made a Good Start, but There Is a Long Way to Go. *Molecular and Cellular Proteomics*. American Society for Biochemistry and Molecular Biology Inc. January 1, 2018, pp 2–17. <https://doi.org/10.1074/mcp.MR117.000126>.
- (10) Levery, S. B.; Steentoft, C.; Halim, A.; Narimatsu, Y.; Clausen, H.; Vakhrushev, S. Y. Advances in Mass Spectrometry Driven O-Glycoproteomics. *Biochimica et Biophysica Acta - General Subjects*. Elsevier B.V. 2015, pp 33–42. <https://doi.org/10.1016/j.bbagen.2014.09.026>.
- (11) Stavenhagen, K.; Hinneburg, H.; Kolarich, D.; Wuhrer, M. Site-Specific N- and O-Glycopeptide Analysis Using an Integrated C18-PGC-LC-ESI-QTOF-MS/MS Approach. In *Methods in Molecular Biology*; Humana Press Inc., 2017; Vol. 1503, pp 109–119. [https://doi.org/10.1007/978-1-4939-6493-2\\_9](https://doi.org/10.1007/978-1-4939-6493-2_9).
- (12) Woo, C. M.; Lund, P. J.; Huang, A. C.; Davis, M. M.; Bertozzi, C. R.; Pitteri, S. J. Mapping and Quantification of over 2000 O-Linked Glycopeptides in Activated Human T Cells with Isotope-Targeted Glycoproteomics (Isotag). *Mol. Cell. Proteomics* **2018**, *17* (4), 764–775. <https://doi.org/10.1074/mcp.RA117.000261>.
- (13) Schumann, B.; Malaker, S. A.; Wisnovsky, S. P.; Debets, M. F.; Agbay, A. J.; Fernandez, D.; Wagner, L. J. S.; Lin, L.; Li, Z.; Choi, J.; et al. Bump-and-Hole Engineering Identifies Specific Substrates of Glycosyltransferases in Living Cells. *Mol. Cell* **2020**, *78* (5), 824–834.e15. <https://doi.org/10.1016/j.molcel.2020.03.030>.
- (14) Narimatsu, Y.; Joshi, H. J.; Schjoldager, K. T.; Hintze, J.; Halim, A.; Steentoft, C.; Nason, R.; Mandel, U.; Bennett, E. P.; Clausen, H.; et al. Exploring Regulation of Protein O-Glycosylation in Isogenic Human HEK293 Cells by Differential O-Glycoproteomics. *Mol. Cell. Proteomics* **2019**, *18* (7), 1396–1409. <https://doi.org/10.1074/mcp.RA118.001121>.
- (15) Steentoft, C.; Vakhrushev, S. Y.; Joshi, H. J.; Kong, Y.; Vester-Christensen, M. B.; Schjoldager, K. T.-B. G.; Lavrsen, K.; Dabelsteen, S.; Pedersen, N. B.; Marcos-Silva, L.; et al. Precision Mapping of the Human O-GalNAc Glycoproteome through SimpleCell Technology. *EMBO J.* **2013**, *32* (10), 1478–1488. <https://doi.org/10.1038/emboj.2013.79>.
- (16) Pap, A.; Klement, E.; Hunyadi-Gulyas, E.; Darula, Z.; Medzihradzsky, K. F. Status Report on the High-Throughput Characterization of Complex Intact O-Glycopeptide Mixtures. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (6), 1210–1220. <https://doi.org/10.1007/s13361-018-1945-7>.
- (17) Yu, Q.; Canales, A.; Glover, M. S.; Das, R.; Shi, X.; Liu, Y.; Keller, M. P.; Attie, A. D.; Li, L. Targeted Mass Spectrometry Approach Enabled Discovery of O- Glycosylated Insulin and Related Signaling Peptides in Mouse and Human Pancreatic Islets. *Anal. Chem.* **2017**, *89* (17), 9184–9191. <https://doi.org/10.1021/acs.analchem.7b01926>.
- (18) Kelly, M. I.; Dodds, E. D. Parallel Determination of Polypeptide and Oligosaccharide Connectivities by Energy-Resolved Collision-Induced Dissociation of Protonated O-Glycopeptides Derived from Nonspecific Proteolysis. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (3), 624–632. <https://doi.org/10.1021/jasms.9b00065>.
- (19) Darula, Z.; Pap, Á.; Medzihradzsky, K. F. Extended Sialylated O-Glycan Repertoire of Human Urinary Glycoproteins Discovered and Characterized Using Electron-Transfer/Higher-Energy Collision Dissociation. *J. Proteome Res.* **2019**, *18* (1), 280–291. <https://doi.org/10.1021/acs.jproteome.8b00587>.
- (20) Zhang, Y.; Xie, X.; Zhao, X.; Tian, F.; Lv, J.; Ying, W.; Qian, X. Systems Analysis of Singly and Multiply O-Glycosylated Peptides in the Human Serum Glycoproteome via ETHcD and HCD Mass Spectrometry. *J. Proteomics* **2018**, *170*, 14–27. <https://doi.org/10.1016/j.jprot.2017.09.014>.

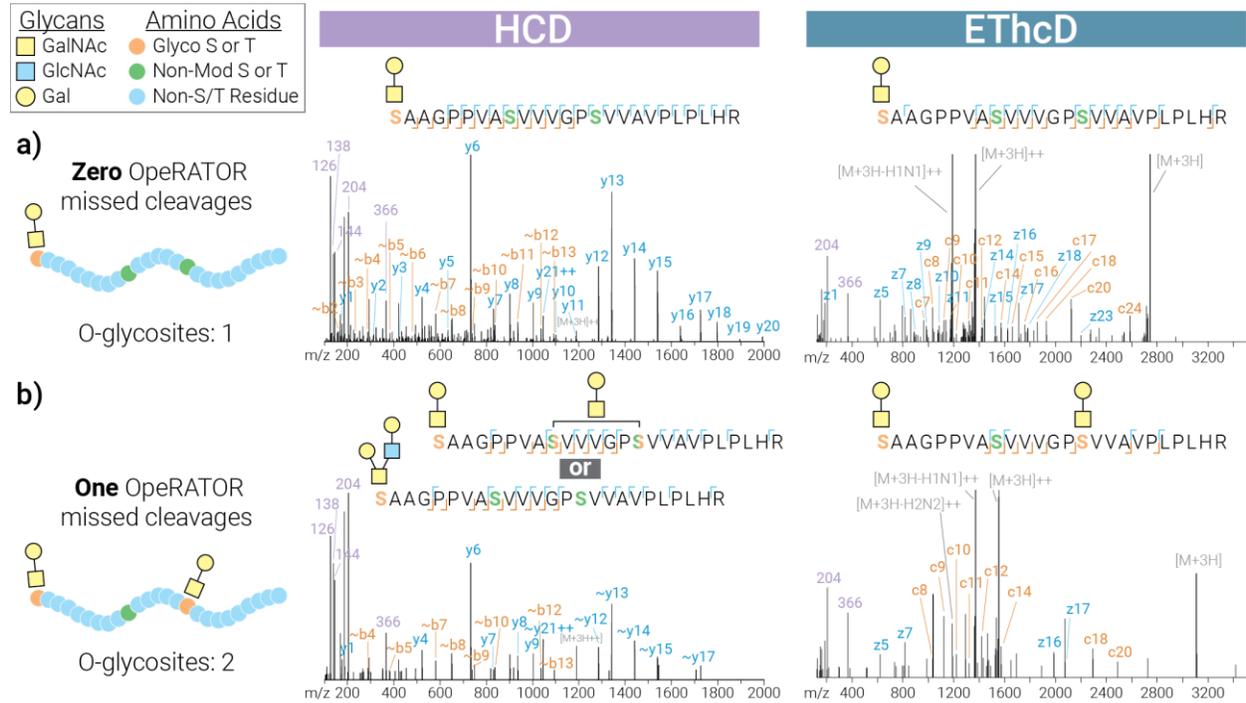
- (21) Pap, A.; Medzihradsky, K. F.; Darula, Z. Using “Spectral Families” to Assess the Reproducibility of Glycopeptide Enrichment: Human Serum O-Glycosylation Revisited. *Anal. Bioanal. Chem.* **2017**, *409* (2), 539–550. <https://doi.org/10.1007/s00216-016-9960-7>.
- (22) Huang, J.; Jiang, B.; Zhao, H.; Wu, M.; Kong, S.; Liu, M.; Yang, P.; Cao, W. Development of a Computational Tool for Automated Interpretation of Intact O -Glycopeptide Tandem Mass Spectra from Single Proteins. *Anal. Chem.* **2020**. <https://doi.org/10.1021/acs.analchem.0c01091>.
- (23) Chalkley, R. J.; Medzihradsky, K. F.; Darula, Z.; Pap, A.; Baker, P. R. The Effectiveness of Filtering Glycopeptide Peak List Files for Y Ions. *Mol. Omi.* **2020**, *16* (2), 147–155. <https://doi.org/10.1039/c9mo00178f>.
- (24) Mao, J.; You, X.; Qin, H.; Wang, C.; Wang, L.; Ye, M. A New Searching Strategy for the Identification of O-Linked Glycopeptides. *Anal. Chem.* **2019**, *91* (6), 3852–3859. <https://doi.org/10.1021/acs.analchem.8b04184>.
- (25) Zhao, X.; Zheng, S.; Li, Y.; Huang, J.; Zhang, W.; Xie, Y.; Qin, W.; Qian, X. An Integrated Mass Spectroscopy Data Processing Strategy for Fast Identification, In-Depth, and Reproducible Quantification of Protein O-Glycosylation in a Large Cohort of Human Urine Samples. *Anal. Chem.* **2019**, *92* (1), 690–698. <https://doi.org/10.1021/acs.analchem.9b02228>.
- (26) Malaker, S. A.; Pedram, K.; Ferracane, M. J.; Bensing, B. A.; Krishnan, V.; Pett, C.; Yu, J.; Woods, E. C.; Kramer, J. R.; Westerlind, U.; et al. The Mucin-Selective Protease StcE Enables Molecular and Functional Analysis of Human Cancer-Associated Mucins. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (15), 7278–7287. <https://doi.org/10.1073/pnas.1813020116>.
- (27) Shon, D. J.; Malaker, S.; Pedram, K.; Yang, E.; Krishnan, V.; Dorigo, O.; Bertozzi, C. An Enzymatic Toolkit for Selective Proteolysis, Detection, and Visualization of Mucin-Domain Glycoproteins. **2020**. <https://doi.org/10.26434/CHEMRXIV.12476369.V1>.
- (28) Yang, W.; Ao, M.; Hu, Y.; Li, Q. K.; Zhang, H. Mapping the O-glycoproteome Using Site-specific Extraction of O-linked Glycopeptides (EXoO). *Mol. Syst. Biol.* **2018**, *14* (11). <https://doi.org/10.15252/msb.20188486>.
- (29) Yang, S.; Onigman, P.; Wu, W. W.; Sjogren, J.; Nyhlen, H.; Shen, R. F.; Cipollo, J. Deciphering Protein O-Glycosylation: Solid-Phase Chemoenzymatic Cleavage and Enrichment. *Anal. Chem.* **2018**, *90* (13), 8261–8269. <https://doi.org/10.1021/acs.analchem.8b01834>.
- (30) Yang, W.; Ao, M.; Song, A.; Xu, Y.; Sokoll, L.; Zhang, H. Mass Spectrometric Mapping of Glycoproteins Modified by Tn-Antigen Using Solid-Phase Capture and Enzymatic Release. *Anal. Chem.* **2020**, *acs.analchem.0c01564*. <https://doi.org/10.1021/acs.analchem.0c01564>.
- (31) Riley, N. M.; Malaker, S. A.; Driessen, M.; Bertozzi, C. R. Optimal Dissociation Methods Differ for N- and O-Glycopeptides. *J. Proteome Res.* **2020**. <https://doi.org/10.1021/acs.jproteome.0c00218>.
- (32) Lu, L.; Riley, N. M.; Shortreed, M. R.; Bertozzi, C. R.; Smith, L. M. O-Pair Search with MetaMorpheus for O-Glycopeptide Characterization. *bioRxiv* **2020**, 2020.05.18.102327. <https://doi.org/10.1101/2020.05.18.102327>.
- (33) Singh, C.; Zampronio, C. G.; Creese, A. J.; Cooper, H. J. Higher Energy Collision Dissociation (HCD) Product Ion-Triggered Electron Transfer Dissociation (ETD) Mass Spectrometry for the Analysis of N-Linked Glycoproteins. *J. Proteome Res.* **2012**, *11* (9), 4517–4525. <https://doi.org/10.1021/pr300257c>.
- (34) Saba, J.; Dutta, S.; Hemenway, E.; Viner, R. Increasing the Productivity of Glycopeptides Analysis by Using Higher-Energy Collision Dissociation-Accurate Mass-Product-Dependent Electron Transfer Dissociation. *Int. J. Proteomics* **2012**, *2012*, 1–7.

- <https://doi.org/10.1155/2012/560391>.
- (35) Wu, S. W.; Pu, T. H.; Viner, R.; Khoo, K. H. Novel LC-MS2 Product Dependent Parallel Data Acquisition Function and Data Analysis Workflow for Sequencing and Identification of Intact Glycopeptides. *Anal. Chem.* **2014**, *86* (11), 5478–5486. <https://doi.org/10.1021/ac500945m>.
- (36) Rose, C. M.; Rush, M. J. P.; Riley, N. M.; Merrill, A. E.; Kwiecien, N. W.; Holden, D. D.; Mullen, C.; Westphall, M. S.; Coon, J. J. A Calibration Routine for Efficient ETD in Large-Scale Proteomics. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (11), 1848–1857. <https://doi.org/10.1007/s13361-015-1183-1>.
- (37) Brademan, D. R.; Riley, N. M.; Kwiecien, N. W.; Coon, J. J. Interactive Peptide Spectral Annotator: A Versatile Web-Based Tool for Proteomic Applications. *Mol. Cell. Proteomics* **2019**, *18* (8), S193–S201. <https://doi.org/10.1074/mcp.TIR118.001209>.
- (38) Mauri, M.; Elli, T.; Caviglia, G.; Uboldi, G.; Azzi, M. RAWGraphs: A Visualisation Platform to Create Open Outputs. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter - CHIItaly '17*; ACM Press: New York, New York, USA, 2017; Vol. Part F131371, pp 28:1–28:5. <https://doi.org/10.1145/3125571.3125585>.
- (39) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14* (6), 1188–1190. <https://doi.org/10.1101/gr.849004>.
- (40) Vester-Christensen, M. B.; Halim, A.; Joshi, H. J.; Steentoft, C.; Bennett, E. P.; Levery, S. B.; Vakhrushev, S. Y.; Clausen, H. Mining the O-Mannose Glycoproteome Reveals Cadherins as Major O-Mannosylated Glycoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (52), 21018–21023. <https://doi.org/10.1073/pnas.1313446110>.
- (41) Polasky, D. A.; Yu, F.; Teo, G. C.; Nesvizhskii, A. I. Fast and Comprehensive N- and O-Glycoproteomics Analysis with MSFragger-Glyco. *bioRxiv* **2020**, 2020.05.18.102665. <https://doi.org/10.1101/2020.05.18.102665>.
- (42) Halim, A.; Westerlind, U.; Pett, C.; Schorlemer, M.; Rüetschi, U.; Brinkmalm, G.; Sihlbom, C.; Lenggqvist, J.; Larson, G.; Nilsson, J. Assignment of Saccharide Identities through Analysis of Oxonium Ion Fragmentation Profiles in LC–MS/MS of Glycopeptides. *J. Proteome Res.* **2014**, *13* (12), 6024–6032. <https://doi.org/10.1021/pr500898r>.
- (43) Riley, N. M.; Hebert, A. S.; Westphall, M. S.; Coon, J. J. Capturing Site-Specific Heterogeneity with Large-Scale N-Glycoproteome Analysis. *Nat. Commun.* **2019**, *10* (1). <https://doi.org/10.1038/s41467-019-09222-w>.
- (44) Khatri, K.; Pu, Y.; Klein, J. A.; Wei, J.; Costello, C. E.; Lin, C.; Zaia, J. Comparison of Collisional and Electron-Based Dissociation Modes for Middle-Down Analysis of Multiply Glycosylated Peptides. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (6), 1075–1085. <https://doi.org/10.1007/s13361-018-1909-y>.
- (45) Riley, N. M.; Coon, J. J. The Role of Electron Transfer Dissociation in Modern Proteomics. *Analytical Chemistry*. American Chemical Society January 2, 2018, pp 40–64. <https://doi.org/10.1021/acs.analchem.7b04810>.
- (46) Zubarev, R.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, *120*, 3265–3266.
- (47) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (26), 9528–9533. <https://doi.org/10.1073/pnas.0402700101>.
- (48) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data. *Nucleic Acids Res.* **2019**, *47* (D1), D442–D450. <https://doi.org/10.1093/nar/gky1106>.

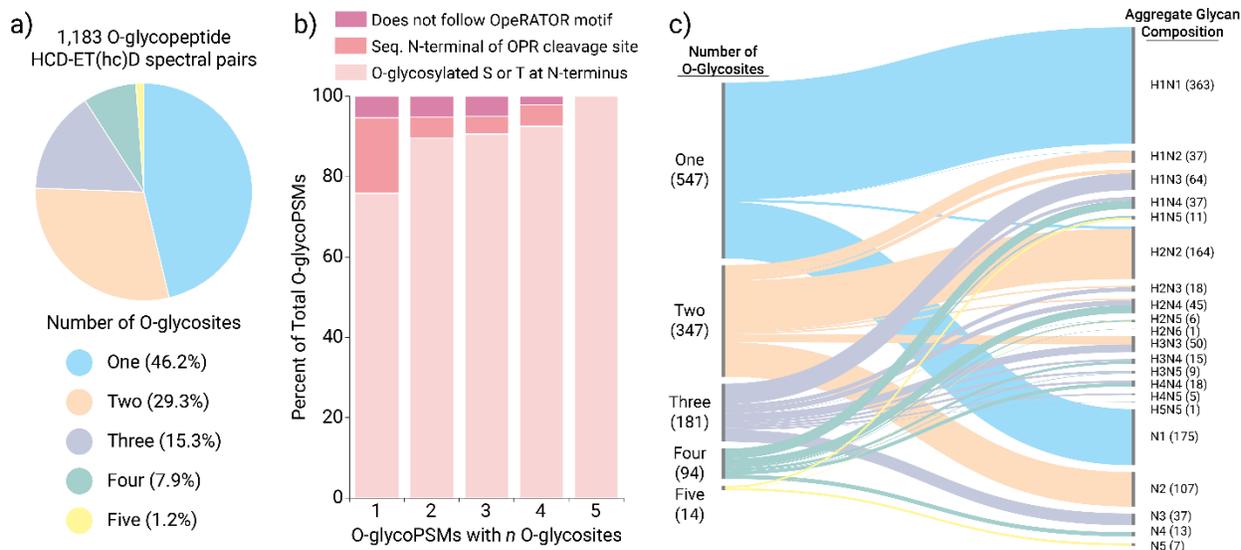
**FOR TOC ONLY**



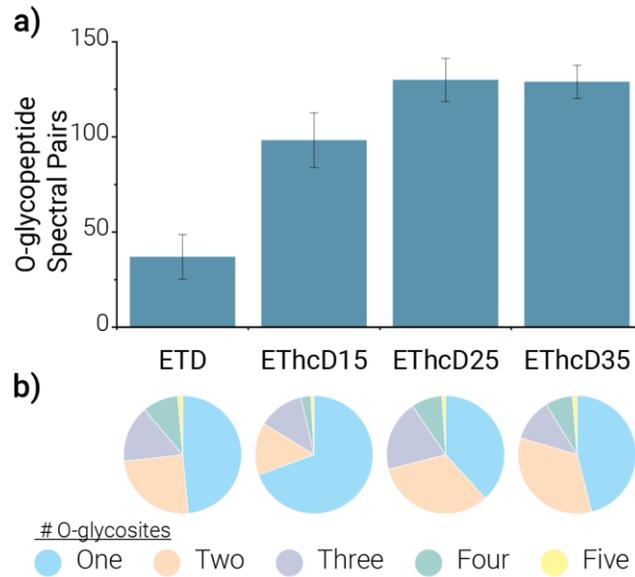
## FIGURES



**Figure 1. Bovine fetuin sequence  $^{282}$ SAAGPPVASVVVGPSVVAVPLPLHR $^{306}$  is a tangible example of the benefits of electron-driven dissociation.** Diagrams at the left present two glycopeptidoform possibilities of this sequence, which have both O-glycosylated (orange) and non-modified (green) serine and threonine residues. These two O-glycopeptidoforms were identified in our dataset using HCD-ETHcD spectral pairs, including a canonical OperATOR-derived O-glycopeptide with a single O-glycosite at the N-terminus (**a**) and a doubly glycosylated O-glycopeptide from a missed OperATOR cleavage that has a second internal O-glycosite (**b**).



**Figure 2. Characteristics of O-glycopeptides derived from OpeRATOR proteolysis.** **a)** A total of 1,183 O-glycopeptide spectral pairs with confidently localized glycosites were detected using HCD-pd-ETD and HCD-pd-ETHcD methods. The pie graph shows the proportion of identifications that had one or multiple O-glycosites. **b)** The bar graph delineates the proportion of O-glycopeptide spectral pairs that can be grouped into three main classes: 1) identifications that have an O-glycosylated serine or threonine at the N-terminus (light pink; bottom), 2) O-glycopeptides from sequences upstream (i.e., N-terminal) of the OpeRATOR (OPR) cleavage site that indicate a missed cleavage (pink; middle), and 3) O-glycopeptides that do not appear to follow the OpeRATOR cleavage motif (dark pink; top). **c)** The alluvial diagram relates the aggregate glycan compositions that were identified on O-glycopeptides with a varying number of localized O-glycosites. Numbers in parentheses provide the number of O-glycopeptide spectral pairs identified for each condition. For glycan modifications shown as H#N#, H represents the number of hexose residues (galactose in O-glycans), and N represents the number of *N*-acetylhexosamines, which can be *N*-acetylgalactosamine or *N*-acetylglucosamine in O-glycans.



**Figure 3. Performance summary for ETD and EThcD methods and OpeRATOR cleavage motif analysis.** **a)** The average number of O-glycopeptide spectral pair identifications with confidently localized O-glycosites is shown for four different product-dependent methods utilizing ETD and EThcD fragmentation. Numbers after EThcD provide normalized collision energies used, and error bars show one standard deviation. **b)** Pie graphs show the proportion of identifications that had one or multiple O-glycosites for the total pool of identifications from each method.