

# Decoding Protein Glycosylation by an Integrative Mass Spectrometry-Based *De Novo* Sequencing Strategy

Jing Gao,<sup>†</sup> Hongxu Chen,<sup>†</sup> Hongrui Yin,<sup>†</sup> Xin Chen, Zhicheng Yang, Yuqiu Wang, Jianhong Wu, Yingping Tian, Hong Shao, Liuqing Wen,<sup>\*</sup> and Hu Zhou<sup>\*</sup>



Cite This: *JACS Au* 2025, 5, 702–713



Read Online

ACCESS |

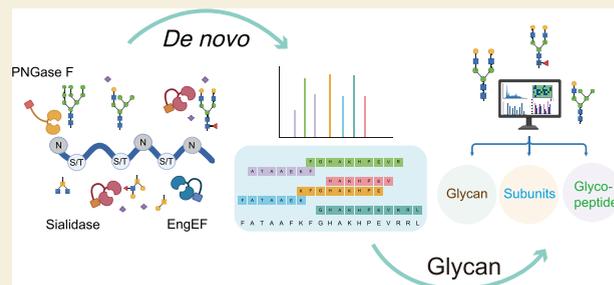
Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Glycoproteins, representing more than 50% of human proteins and most biopharmaceuticals, are crucial for regulating various biological processes. The complexity of multiple glycosylation sites often leads to incomplete sequence coverage and ambiguous glycan modification profiles. Here, we developed an integrative mass spectrometry-based approach for decoding unknown glycoproteins, which is featured with the combination of deglycosylation-mediated *de novo* sequencing with glycosylation site characterization. We utilized the enzymatic deglycosylation of *N*-/*O*-glycans to achieve comprehensive sequence coverage. Additionally, EThcD fragmentation enables the identification of high-quality long peptides, facilitating precise protein assembly. We subsequently applied this method to *de novo* sequencing of the highly glycosylated therapeutic fusion protein Etanercept (Enbrel). We also sequenced three new tumor necrosis factor receptor:Fc-fusion biologics with largely unknown sequences, unveiling subtle distinctions in the primary sequences. Furthermore, we characterized *N*- and *O*-glycosylation modifications of these proteins at subunit, glycopeptide, and glycan levels. This strategy bridges the gap between the *de novo* sequencing and glycosylation modification, providing comprehensive information on the primary structure and glycosylation modifications for glycoproteins. Notably, our method could be a robust solution for accurate sequencing of the glycoproteins and has practical value not only in basic research but also in the biopharmaceutical industry.

**KEYWORDS:** *de novo* protein sequencing, mass spectrometry, glycoproteins, *N*-/*O*-glycosylation, Etanercept, TNFR:Fc-fusion biologics



## INTRODUCTION

Protein glycosylation, as one of the most widespread and complex post-translational modifications (PTMs), plays fundamental roles in many biological processes.<sup>1,2</sup> A comprehensive analysis of the primary sequences of glycoproteins, including the identification of glycosylation sites and associated glycan structures, forms the cornerstone of glycoprotein functional research.<sup>3,4</sup> It also serves as a critical quality attribute (cQA) for biopharmaceuticals, offering insights into protein structure, function, disease associations, and biotechnological applications.

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) has emerged as a powerful tool for protein identification and post-translational modification discovery.<sup>5</sup> Diverging from conventional database searching-based protein detection strategies that are widely used in glycoproteomics, *de novo* protein sequencing is a novel strategy that focuses on the “black-box” protein without prior knowledge of DNA/amino acid sequence.<sup>6–8</sup> In this approach, overlapping peptides are generated by multiple protease digestion, then analyzed by LC-MS/MS, identified by *de novo* algorithms, and assembled into full-length proteins using a combination of software and manual interpretation. In-depth sequencing of peptides is

essential to generate arrays of adjacent overlapping peptides that reveal the full coverage of protein sequence. However, owing to the suppression effects of glycan moieties on glycopeptide sequencing efficiency,<sup>9</sup> it is hard to get informative glycopeptide fragmentation spectra to support *de novo* sequencing. Moreover, the complexity of spectra comprising both the peptide and glycan fragment, makes it hard to *de novo* interpret glycopeptide fragmentation spectra for highly accurate glycopeptide sequencing.<sup>10</sup> As a result, extensive glycosylation often results in gaps in certain regions, limiting the application of *de novo* sequencing to monoclonal antibodies (mAbs) with consistent structures and known *N*-linked glycosylation sites.<sup>6–8,11</sup>

Glycoproteomics focuses on the system-wide identification, cataloging, and characterization of glycoproteins, offering insights into how glycosylation influences biological pro-

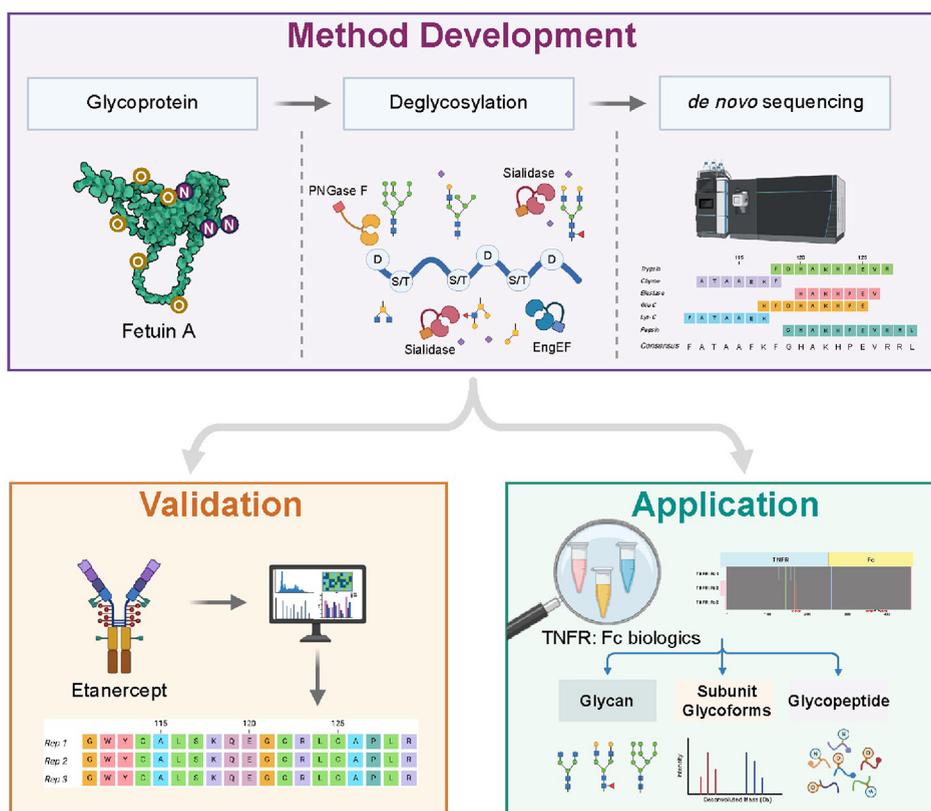
Received: October 14, 2024

Revised: December 31, 2024

Accepted: January 3, 2025

Published: January 22, 2025





**Figure 1.** Overview of the *de novo* protein glycosylation analysis study design.

cesses.<sup>12,13</sup> Recent advancements in glycoproteomics, including improved sample preparation strategies such as sequential deglycosylation,<sup>14–16</sup> and the development of specialized glycopeptide identification tools like Byonic, pGlyco, and GPQuest, have greatly enhanced the accuracy and confidence of glycopeptide analysis.<sup>17–19</sup> By integrating these approaches and methods with *de novo* sequencing, it helps to overcome the technical challenges in glycoprotein *de novo* sequencing, the sequencing, it becomes possible to comprehensively decode glycoproteins, thus providing insights into both protein sequences and their associated glycosylation patterns. In this study, we employed a hybridization approach to decode protein glycosylation, combining deglycosylation-mediated *de novo* sequencing with glycosylation site characterization. The *de novo* sequencing involves enzymatic deglycosylation of N-/O-glycans to significantly reduce the difficulty of MS/MS fragmentation and the complexity of glycopeptide-derived fragmentation spectra, thereby achieving full sequence coverage. EThcD fragmentation enables the identification of high-quality long peptides to facilitate precise protein assembly. To test the above integrative workflow, we used a group of highly glycosylated therapeutic recombinant TNFR:Fc-fusion proteins, Etanercept, and three sequence unknown TNFR:Fc-fusion biologics. These proteins contain diverse N-/O-glycosylation sites as well as an artificial hinge domain, demonstrating the applicability of our integrated approach in some of the most challenging scenarios. Furthermore, we performed characterization of glycosylation by multilevel analysis, including N-glycosylation site identification, released glycan analysis, glycoprotein subunit analysis, and glycopeptide analysis, providing comprehensive information on both N- and O-glycosylation patterns. This method is a powerful strategy

for sequencing glycoproteins and is likely to find applications in various biopharmaceuticals as well as biotechnological fields.

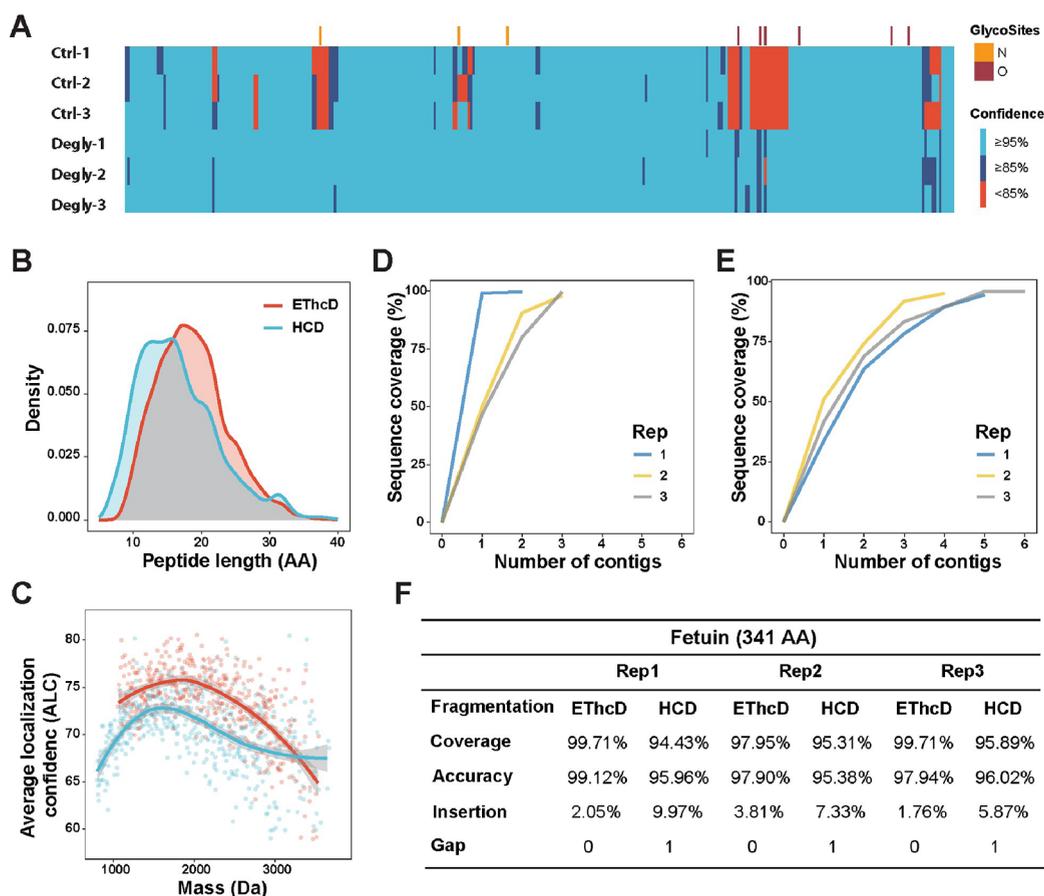
## RESULTS AND DISCUSSION

### Overview of the *De Novo* Protein Glycosylation Analysis Study Design

In this study, we present an integrative mass spectrometry-based approach to *de novo* decode protein glycosylation (Figure 1). We initially developed the *de novo* sequencing strategy for glycoproteins using Fetuin-A (UniProt Accession no.P12763), a commonly used model glycoprotein containing sialylated N-glycans and O-glycans (Figure S1). We employed glycosidases for the hydrolysis of glycosidic bonds to simplify the mass spectrometry analysis and the peptide sequence determination and obtained the primary sequence from the deglycosylated protein by an optimized *de novo* sequencing process. Next, we employed the highly complicated glycosylated biopharmaceutical, Etanercept, to demonstrate and validate the method performance. Subsequently, we applied the strategy to reveal the amino acid sequences of three unknown TNFR:Fc-fusion biologics to explore their similarity with the originator drug Etanercept. In this regard, the glycosylation characterization was carried out on multilevels, including N-glycosylation site identification, released glycan analysis, glycoprotein subunit analysis, and glycopeptide analysis, to pinpoint the N-/O-glycosylation.

### Development of *De Novo* Sequencing Method for Glycoproteins

We employed both PNGase F and O-glycosidase EngEF to remove N-glycans and O-glycans on Fetuin A, respectively. As the EngEF is active only on unsubstituted O-glycan cores, the



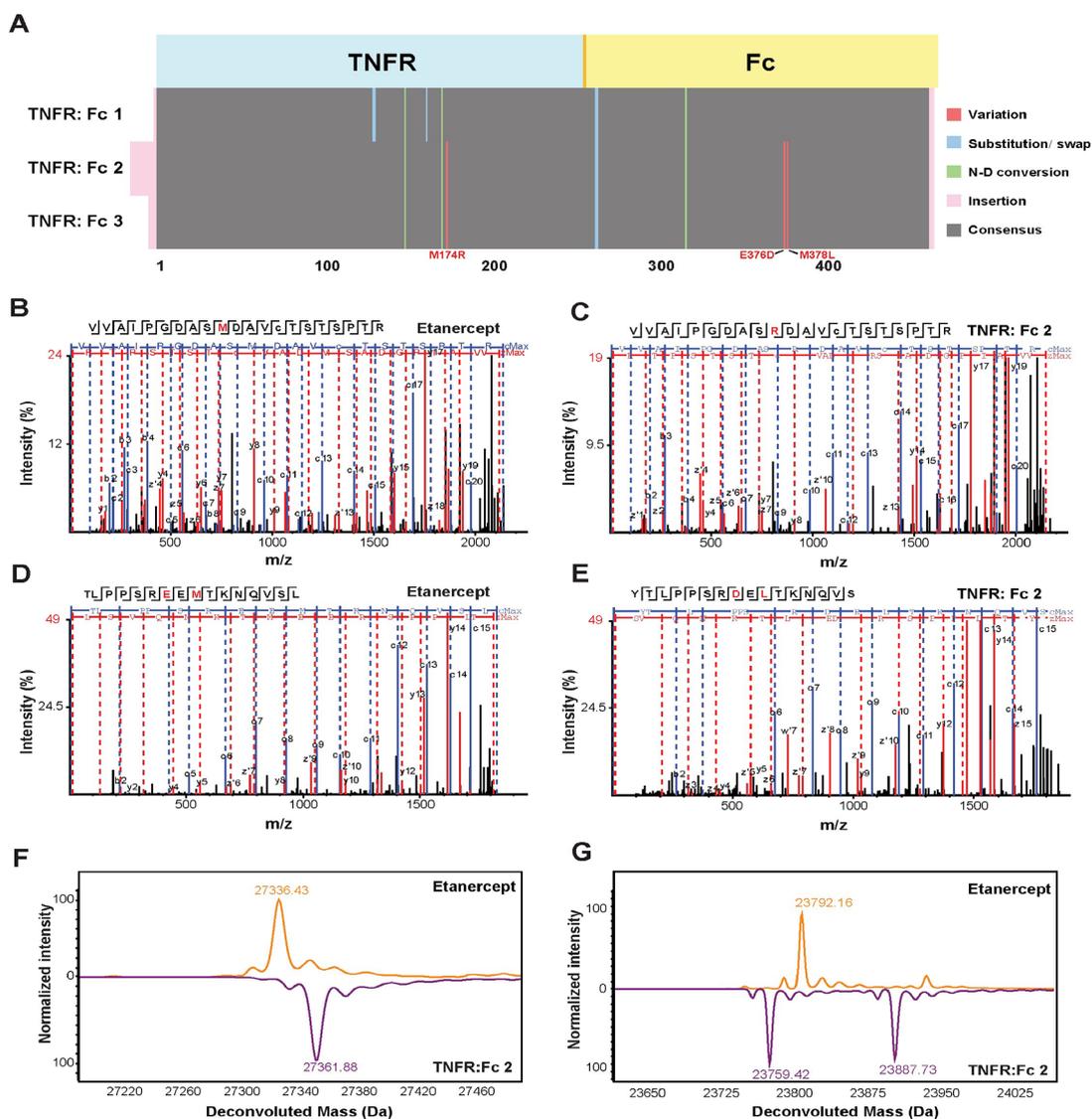
**Figure 2.** Development of the *de novo* sequencing method for glycoproteins. (A) Comparing the impact of deglycosylation on the confidence of sequence identification through the integration of results from multiple proteases (LysC, elastase, pepsin, chymotrypsin, GluC, and trypsin). (B) Density plot of the peptide length distribution. (C) Average localization confidence (ALC) of *de novo* peptide candidates. Assembly performance of ETHcD data sets (D) and stepped HCD data sets (E). It shows the full-length protein sequence coverage (*y*-axis) when certain numbers of contigs were used (*x*-axis). (F) Assembly performance comparison between ETHcD and stepped HCD using *de novo* sequenced peptides.

digest was performed in combination with Sialidase. The deglycosylation effect was confirmed by intact mass analysis (Figure S2). As expected, the sequence coverage improved significantly in most of the enzymatic digestion groups after the removal of glycans (Figure S3). For chymotrypsin, Glu-C, pepsin, and trypsin digestion groups, over 90% sequence coverage was achieved, laying the foundation for starting *de novo* glycoprotein sequencing. Considering the complementarity of different enzymatic methods, we tried to integrate them with the assumption that the robustness would be improved. Largely like the single enzyme digest results, the integrated results also confirmed the higher confidence of sequence identification after deglycosylation (Figure 2A). Indeed, the glycopeptide backbone identification was more confident after glycan removal (Figure S4), allowing the in-depth identification of the glycosylation sites containing sequence regions (Figure S5).

Extensive fragmentation of peptide ions is essential to generate arrays of adjacent fragments that reveal the amino acid sequence. Given the ETHcD fragmentation significantly increases the number of fragment ions containing more sequence information in MS/MS scans, which is beneficial for inferring the peptide sequence and post-translational modification,<sup>20</sup> we assessed the *de novo* peptide sequencing performance of the ETHcD fragmentation method compared with the commonly used stepped higher-energy collision

dissociation HCD fragmentation. We collected data using stepped HCD and ETHcD fragmentation separately. Our results revealed that despite the total number of *de novo* peptide candidates being fewer than that of the stepped HCD method (Figure S6), ETHcD methods outperformed the stepped HCD methods in terms of peptide sequence length (Figure 2B and Figure S7A) and spectrum quality (Figure 2C and Figure S7B). Moreover, the ETHcD method with longer *de novo* peptides of the ETHcD spectrum facilitates the assembly of protein. The top 3 contigs (set of overlapping peptides assembled by our approach is referred to as a contig) with overlapped sequences of ETHcD covered nearly complete of the full length (Figure 2D), while the HCD method generated ~5 contigs covering the ~95% sequence (Figure 2E). Importantly, considering the sequencing accuracy, ETHcD method achieved an overall average sequence accuracy of 98.3% (Figure 2F and Figure S8), while the stepped HCD method resulted in 95.8% overall average sequence accuracy as well as 1 gap (Figure 2F, Figure S9).

Inspection of the assemblies, the highest error frequency occurred on the Ile-Leu (I-L) and Asn-Asp (N-D) conversion. The three N-D conversions are in good agreement with the N-glycosylation sites, on which the attached N-glycan was released by PNGase F (Figure S8), converting glycosylated Asn residues into Asp.<sup>21</sup> This result indicated that both the stepped HCD and ETHcD methods can faithfully reveal the



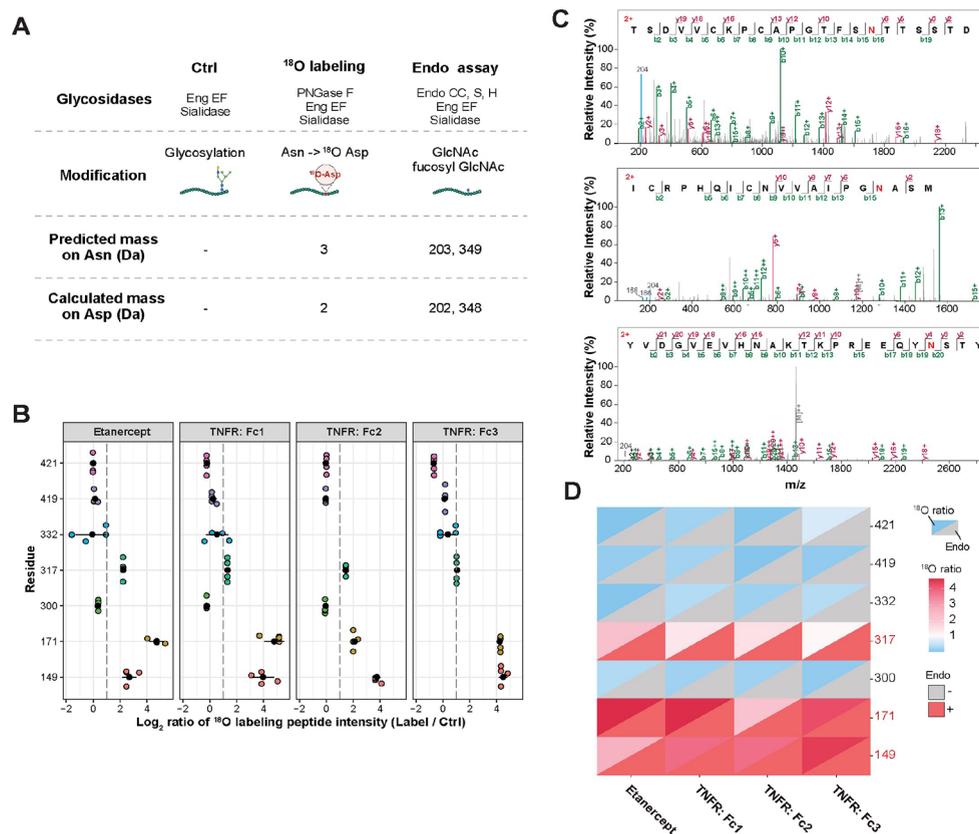
**Figure 3.** *De novo* sequencing of TNFR:Fc-fusion biopharmaceuticals. (A) Sequence similarity of three new TNFR:Fc-fusion biologics and Etanercept. *De novo* sequencing results of biologics were aligned to the known sequence of Etanercept. (B) MS/MS spectra of the original peptide containing M174 in Etanercept. (C) MS/MS spectra of the variant peptide containing M174R in TNFR:Fc 2. (D) MS/MS spectra of original peptide containing E376 and M378 in Etanercept. (E) MS/MS spectra of variant peptide containing E376D, M378L in TNFR:Fc 2. (F) Mirror plot of deconvoluted mass spectra of TNFR fragment of Etanercept and TNFR:Fc 2. (G) Mirror plot of deconvoluted mass spectra of Fc fragment of Etanercept and TNFR:Fc 2.

actual amino acids of the peptide mixture. The Asn residues that are *N*-glycosylated can be determined by the  $^{18}\text{O}$  labeling-based peptide mapping method as well as the glycosidase-based method (see later). Notably, the EThcD method showed its advantage in reliable Leu/Ile determination. With the diagnostic  $w'$ -ion in the EThcD spectrum (characteristic loss of  $\text{C}_3\text{H}_7$  ( $-43.05$  Da) for Leu or  $\text{C}_2\text{H}_5$  ( $-29.04$  Da) for Ile from the side chain of particular  $z$ -ions),<sup>22</sup> we reliably identified 12 out of 13 I-L conversions occurring in HCD data sets (peptide DIEIDTLETTCH as an example, Figures S10 and S11).

Consequently, EThcD mass acquisition simplifies the protein assembly and enables higher accuracy for protein *de novo* sequencing. In combination with glycan removal and EThcD mass spectrometry acquisition, we developed a robust method for glycoprotein *de novo* sequencing, enabling precise sequence coverage in the glycosylation region.

### De Novo Sequencing of Etanercept, a Highly Glycosylated TNFR:Fc-Fusion Biopharmaceutical

To test the robustness of our *de novo* strategies and demonstrate their application for highly glycosylated biopharmaceuticals, we applied the developed method to sequence a marketed therapeutic Fc-fusion protein original drug Etanercept. As one of the most complex highly glycosylated Fc-fusion biopharmaceuticals, Etanercept possesses at least 3 *N*- and 13 *O*-glycosylation sites, a soluble fusion protein of the tumor necrosis factor receptor extracellular domain, linked to a Fc part of IgG1.<sup>23</sup> We confirmed the deglycosylation efficacy for Etanercept by SDS-PAGE (Figure S12) and intact mass spectrometry measurement (Figure S13). The MS/MS spectra were obtained by EThcD fragmentation for the following *de novo* protein sequencing. Owing to the deep sequence coverage achieved by deglycosylation treatment, we can assemble the entire



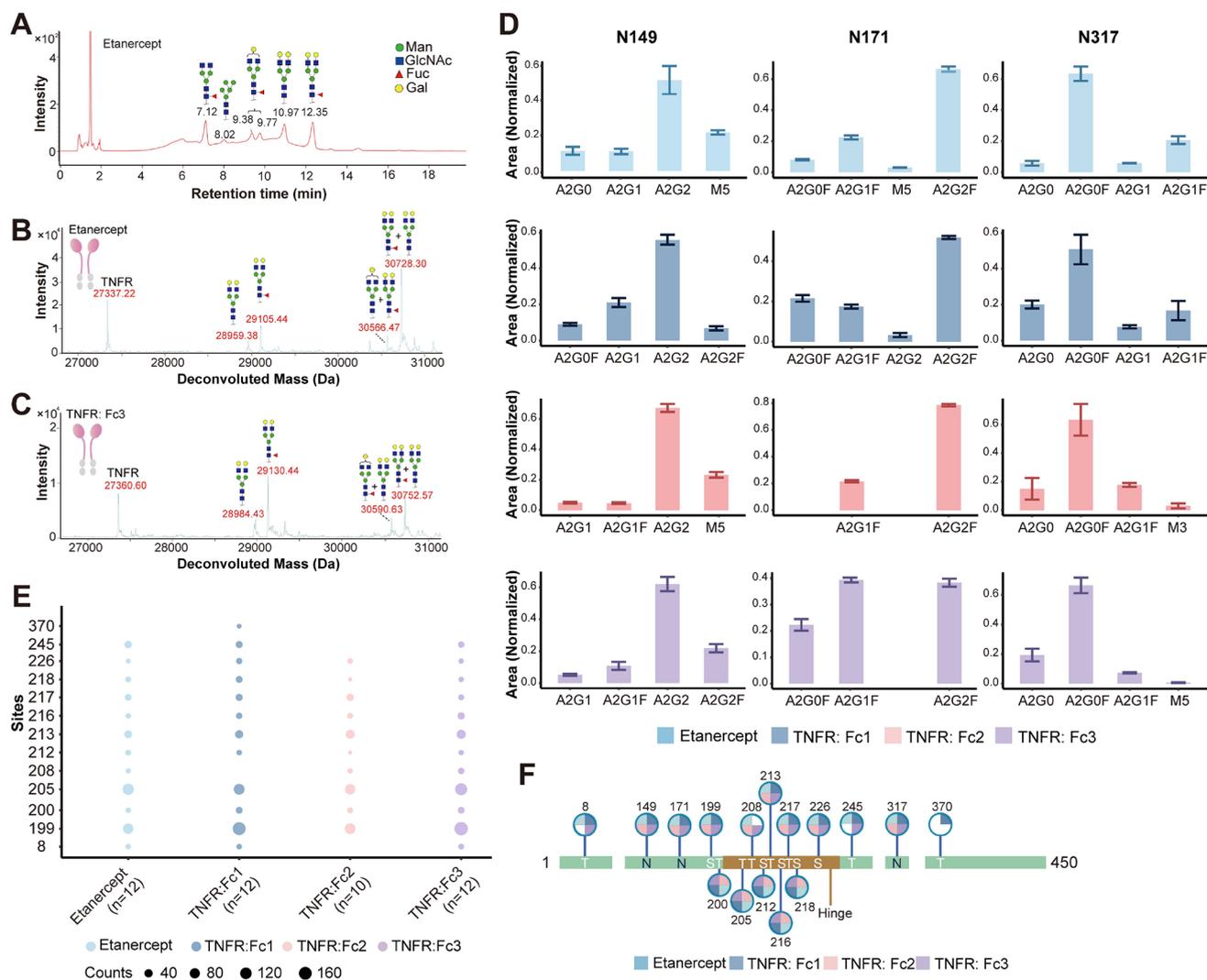
**Figure 4.** Identification of *N*-glycosylation sites. (A) Schematic overview of the protocol with mass signatures for peptide glycosylation sites with <sup>18</sup>O labeling and treatment with a cocktail of endoglycosidases. (B) <sup>18</sup>O labeling based quantitative screening of *N*-glycosylation sites. We compared the intensity peak area of <sup>18</sup>O labeled peptides from labeled to control samples to quantify the relative distribution of the *N*-glycosylation state at each glycosylation site detected. (C) Exemplary MS/MS spectra in support of the assigned *N*-glycosylation sites on Etanercept. The peptide sequence and fragment coverage are indicated in the top left of each spectrum spectra, with *b* ions indicated in green and *y* ions in magenta. The same color annotation is used for peaks in the spectra, with additional peaks such as intact/charge reduced precursors, neutral losses, and immonium ions indicated in gray. The *de novo* sequencing output sequences, with potential Asn to Asp conversions, were used for database searching and corrected to Asn with the *N*-glycosylation sites identification results. GlcNAc (N + 203.10 Da) modification sites are indicated in red. (D) Diagonal heatmap illustrating the consistency of *N*-glycosylation site identification through quantitative screening using <sup>18</sup>O labeling and validation by endoglycosidase treatment. Each row corresponds to a distinct glycosylation site, while each column represents a specific biopharmaceutical sample. The upper triangle displays the log<sub>2</sub> ratio of <sup>18</sup>O-labeled peptide intensities for each site. The lower triangle indicates the presence of GlcNAc or GlcNAc-Fuc modifications identified for the same site. Color reflects the level of consistency between the two methods, with red indicating higher agreement.

protein sequence of Etanercept as one contig (Figure S14). The three replicates reached 98.93% coverage and 99.57–100% accuracy (Table S2). Meanwhile, the conventional strategy failed to assemble probably because of the ambiguous gaps rising from the complex MS/MS spectra of glycosylated peptides which could not be solved by *de novo* sequencing algorithm and multienzyme digestion strategy (Figure S15). Indeed, the gap is located in the Hinge region, which is absent in the human proteome database and reported highly *O*-glycosylated to impede the mass sequencing of peptides, indicating that it is challenging to deduce the full-length primary sequence using a standard database search approach. In summary, our method enables *de novo* sequencing of glycoproteins and achieves a deep coverage of glycosylation modification regions.

#### De Novo Sequencing of Unknown TNFR:Fc-Fusion Biologics and Comparison of Sequence Similarity with Etanercept

We then applied the method to sequence three new TNFR:Fc-fusion biologics, which are biosimilars of Etanercept with

unpublished sequences (derived from Chinese hamster ovary (CHO) cell lines, Table S1). Following the same approach, we successfully deglycosylate and evaluate the effects (Figures S16–S19). Notably, the intact mass results revealed the subtle distinctions in the primary sequences of TNFR:Fc 2 and TNFR:Fc 3 to Etanercept, while the intact mass of TNFR:Fc 1 is as same as Etanercept. Consistently, the sequences of the biologics exhibits differences at three positions, specifically: M174R (in TNFR fragment) and E376D and M378L (both in Fc fragment) (Figure 3A and Figure S20). The MS/MS spectra corresponding to the variant sites demonstrate good quality and accurate localization (Figure 3B–E and Figures S21 and S22). Moreover, the differences in protein sequences are aligned with the intact protein mass (Figure 3F,G; Figure S23 and Table S3). The sequence variant on M174R is a polymorphism in the coding region of human TNFR2 and seems to be associated with polycystic ovary syndrome,<sup>24</sup> hyperandrogenism, and systemic lupus erythematosus.<sup>25</sup> Similarly, E376D and M378L variations on the Fc fragment were also reported in commercially available TNF receptor2-Fc fusion protein products.<sup>26</sup>



**Figure 5.** *N*- and *O*-glycosylation analysis of TNFR:Fc-fusion biopharmaceuticals after desialylation. (A) HILIC-UPLC profile of the total *N*-linked glycans released from Etanercept. Deconvoluted spectrum of subunit analysis for Etanercept TNFR fragment (B) and TNFR:Fc3 TNFR fragment (C). (D) *N*-Glycopeptide profiles of Asn147, Asn179, and Asn317. Top 4 most abundant glycoforms for each *N*-glycosylation site were plotted. (E) Site-specific statistics of *O*-glycosylation. Counts represent the number of identified glycopeptide spectra. (F) Summary plot of *N*-/*O*-glycosylation modification on Etanercept and three biologics.

We validated three TNFR:Fc-fusion biologics protein sequences by searching against the sequence determined by *de novo* sequencing using the sequence validation module of the PEAKS AB software. ~99% sequence had identification confidence >85%, with multiple spectra supporting each tryptic peptide (Figure S24). Therefore, the results confirmed the high accuracy of the *de novo* sequencing results.

#### Analysis of *N*- and *O*-Glycosylation of TNFR:Fc-fusion Biopharmaceuticals

Since glycosylation can influence the biological function of glycoproteins, especially for the efficacy and safety of the biopharmaceutical,<sup>27,28</sup> a thorough assessment of glycosylation is as important as that of amino-acid sequence to decoding the detailed primary structure of glycoproteins. Here, we employed multilevel analysis, including *N*-glycosylation site identification, released glycan analysis, glycoprotein subunit analysis, and glycopeptide analysis, to pinpoint the *N*- and *O*-glycosylation.

We initially conducted quantitative PNGase F-catalyzed glycosylation site <sup>18</sup>O stable isotopic labeling analysis to determine the *N*-glycosylation sites. PNGase F specifically

deamidated the *N*-glycan-linked Asn to Asp labeled with <sup>18</sup>O-water and resulted in the mass addition of 2.98 Da on Asn which is glycosylated (Asn to Asp conversion causes an increment of +0.98 Da and <sup>18</sup>O labeling causes an increment of +2.00 Da, Figure 4A, Figure S25). By calculating the peptide intensity with <sup>18</sup>O labeled sites, three *N*-glycosylation sites (N149, N171, and N317) were screened out in all four TNFR:Fc-fusion protein samples (Ratio cutoff >2, <sup>18</sup>O-labeling sample/the control sample) (Figure 4B). The three *N*-glycosylation sites were exactly matched with the known *N*-glycosylation sites in Etanercept. Furthermore, we confirmed the *N*-glycosylation site identification results by partially cleaving *N*-glycan by a cocktail of endoglycosidases (Endo CC, Endo S, and Endo H) to generate modifications with “GlcNAc”/“GlcNAc-Fuc” on the *N*-glycosylation<sup>29</sup> (Figure 4A,C and Figure S26). The usage of a cocktail of endoglycosidases with broader cleavage specificities, covering biantennary complex-type glycans, high-mannose, and hybrid oligosaccharides, ensures comprehensive cleavage of *N*-glycans in most biopharmaceuticals. Indeed, the results are highly

consistent with the  $^{18}\text{O}$  labeling results, all three sites can be identified as *N*-glycosylation (Figure 4D). In particular, these data can be used to correct the Asn to Asp conversion in *de novo* sequencing results.

To pinpoint the content and structure of *N*-glycans on the whole protein level, enzymatically released *N*-glycans are derivatized with 2-AB labeling to allow for fluorescence and mass spectrometry detection by UPLC-HILIC-FLR-MS. The most abundant peaks of *N*-glycans released from Etanercept were F(6)G2F, biantennary, core-fucosylated *N*-glycans with two terminal galactose, followed by the F(6)G0F, biantennary, core-fucosylated with no terminal galactose residues, G2, biantennary *N*-glycans with two terminal galactose residues, and F(6)G1F, biantennary, core-fucosylated with one terminal galactose residues (Figure 5A). The results of three TNFR:Fc-fusion biologics were similar to Etanercept (Figure S27).

Furthermore, we assessed the *N*-glycan variants at the subunit level after removing the *O*-glycans as well as the sialic acid. Slight differences were observed in the samples (Figure S28). The top abundant glycoforms at the TNFR fragment of Etanercept/TNFR:Fc1 carries two different *N*-glycans, namely, G2 and G2F (Figure 5B and Figure S28). Whereas the TNFR fragment of the TNFR:Fc 3 also contained a number of the G1F glycoforms, in addition to the major glycoforms G2 and G2F (Figure 5C and Figure S28). On the other side, the most prominent glycoforms at the Fc fragment were assigned to G0F and G1F constantly in four samples, despite subtle varying degrees of Lysine-loss modification observed (Figure S28).

To reduce the complexity and variability of glycopeptides, as well as to improve the digestion efficiency of glycopeptides, we used sialidase in combination with *N*-glycosidase or *O*-glycosidase to gain insights into site-specific *N*/*O*-glycosylation.<sup>23</sup> The *N*-glycopeptides provide higher resolution of multiple glycovariants (Figure 5D, Table S4, Figure S29), namely, the N149, N171, and N317 glycosylation sites are mainly occupied by G2, G2F, and G0F, respectively. The N171 is the site that contributed to the difference of glycoforms for TNFR:Fc 3, containing a considerable amount of the G1F glycoform. The site-specific *N*-glycopeptide results are consistent with the assignment of glycoforms based on the subunit level, verifying the coherence of data integration at different digestion levels.

We identified 13 *O*-glycosylation sites on Etanercept and three TNFR:Fc-fusion biologics with the assistance of manually filtered glycopeptide spectra (Figure 5E, Table S5, Figure S30). Generally consistent with previous reports,<sup>23</sup> Etanercept displayed 12 *O*-glycosylation sites. Similar *O*-glycosylation patterns were observed in the TNFR:Fc-fusion biologics. Furthermore, our analysis of glycosylation sites revealed the highest level of glycosylation in the hinge region (Figure 5F).

Despite structural similarities in their backbones, Etanercept and the TNFR:Fc-fusion biologics exhibited variability in glycoform profiles, attributed to distinct manufacturing processes and cell sources. Notably, while TNFR:Fc 1 closely resembled Etanercept in glycoform distribution across *N*-glycosylation sites, TNFR:Fc 2 and TNFR:Fc 3 exhibited profiles more similar to each other. These findings highlight potential physicochemical implications that warrant further investigation and contribute to the comprehensive characterization of Etanercept and its biosimilars.

## CONCLUSIONS

Mass spectrometry-based protein *de novo* sequencing is a powerful tool to reveal the amino acid sequence. Combined with glycosidase enzymatic dissection, EThcD-based protein sequencing, and glycosylation characterization, the novel strategy opens up new avenues in the insight of heavily glycosylated proteins. Our analysis of Etanercept demonstrates that the method provides comprehensive information not only unraveling the amino acid sequence but also pinpointing the glycosylation sites, elucidating the complex landscape of glycoform heterogeneity. Given the commonality of *N*/*O*-glycosylation in protein expressed in CHO cell lines, which comprise the majority of protein therapeutics, the study lays the framework for successful analysis of various biopharmaceuticals, which are instrumental in the treatment of a wide range of diseases.

## METHODS

### Chemicals and Materials

Sialidase from *Streptococcus pneumoniae* (NanA) was prepared as reported previously.<sup>30,31</sup> PNGase F from *Flavobacterium meningosepticum* and endoglycosidase H from *Streptomyces plicatus* (Endo H) were prepared as reported previously.<sup>32</sup> Endo- $\alpha$ -GalNAcases from *Enterococcus faecalis* (EngEF),<sup>33</sup> Endo CC from *Coprinopsis cinerea*,<sup>34</sup> and Endo S from *Streptococcus pyogenes*<sup>35</sup> were cloned into the pET-28a vector with six histidines tag (6 $\times$ His tag) for purification by Ni-affinity chromatography. Etanercept was obtained from Pfizer (Enbrel, Rijksweg, Belgium, Lot-No: GA5942). Three new therapeutic recombinant human tumor necrosis factor receptor fusion protein biologics were obtained from Sunshine Guojian Pharmaceutical (Shanghai, China, Lot-No: 301-C2031), BioRay Biopharmaceutical (Zhejiang, China, Lot-No: 0500222005), and Celgen Biopharma (Shanghai, China, Lot-No: A201222U), respectively. LC-MS grade acetonitrile and water were purchased from Thermo Fisher Scientific (Nepean, ON, Canada). Fetuin-A, sodium deoxycholate (SDC), Tris(2-carboxyethyl)phosphine (TCEP), chloracetamide (CAA), dithiothreitol (DTT), 2-AB dye, formic acid, acetic acid, ammonium bicarbonate (ABC), ammonium formate, guanidine hydrochloride, Tris-HCl, and pepsin were purchased from Sigma-Aldrich (St. Louis, MO). MS sequencing grade trypsin, elastase, AspN, chymotrypsin, and IdeZ were purchased from Promega (Madison, WI). MS sequencing grade Glu-C was purchased from Roche (Basel, Switzerland). MS sequencing grade Lys-C was purchased from FUJIFILM Wako Chemicals (Osaka, Japan).  $^{18}\text{O}$ -water (98 atom % of  $^{18}\text{O}$ ) was purchased from the Shanghai Engineering Research Center of Stable Isotope (Shanghai, China).

### Sample Preparation for *De Novo* Sequencing and Glycopeptide Analysis

30  $\mu\text{g}$  of protein was incubated in 30  $\mu\text{L}$  of 2% SDC, 50 mM Tris-HCl, 10 mM TCEP, and 40 mM CAA (pH 8.0) at 95  $^{\circ}\text{C}$  for 10 min to accomplish denature, reduction, and alkylation in one step. The sample was then diluted with 30  $\mu\text{L}$  of PBS buffer (pH 7.6). For *de novo* sequencing, *N*-glycans and *O*-glycans were released by adding a cocktail of 2U of NanA, 2U of EngEF, and 1U of PNGase F followed by incubating at 37  $^{\circ}\text{C}$  for 2h while shaking at 1200 rpm. The sample was then diluted with 60  $\mu\text{L}$  of 50 mM Tris-HCl (pH 8.0) and digested by one of the following proteases: trypsin, chymotrypsin, Lys-C, Glu-C, pepsin, and elastase at a 1:50 ratio (w/w) for 16 h. The digestion was performed at 37  $^{\circ}\text{C}$ , except for chymotrypsin reaction at 25  $^{\circ}\text{C}$ . For pepsin digestion, samples were acidified to pH 3.5 using formic acid before adding the enzyme. After digestion, SDC was removed by adding 8  $\mu\text{L}$  of pure formic acid and centrifuging at 14,000g for 5 min at 4  $^{\circ}\text{C}$ . The supernatant containing the peptides was then collected for desalting with homemade C18 tips.

The *N*-glycopeptide sample preparation was conducted for glycosylation characterization, with modifications involving NanA

and EngEF instead of a cocktail of three endoglycosidases and utilizing chymotrypsin, pepsin, and trypsin for digestion. Similarly, the *O*-glycopeptide sample preparation was performed with modifications utilizing PNGase F and NanA instead of a cocktail of three endoglycosidases and utilizing trypsin for digestion.

### Sample Preparation for *N*-Glycosylation Site Identification

30  $\mu\text{g}$  of pepsin-treated *O*-glycan removed peptides (treated with NanA and EngEF) were subjected to  $^{18}\text{O}$  labeling quantitative analysis for glycosylation site identification. Equal amounts of both *N*- and *O*-glycan (treated with NanA, EngEF, and PNGase F) released peptides were set as the control. The peptides were thoroughly dried overnight in a speed Vac and resuspended in 30  $\mu\text{L}$  of ABC buffer (50 mM ammonium bicarbonate prepared in  $\text{H}_2^{18}\text{O}$ ). Asn-linked *N*-glycans were removed by 1  $\mu\text{L}$  of PNGase F (8U). The sample was incubated for 3 h at 37  $^\circ\text{C}$  while shaking at 1200 rpm. The peptides were then collected for desalting with homemade C18 tips.

The endoglycosidases-treated peptide sample preparation was conducted as sample preparation for *de novo* sequencing, except for replacing the PNGase F with a cocktail of Endo CC, Endo S, and Endo H in a 1:10 ratio (w/w) for each enzyme.

### Sample Preparation for Intact and Subunits Mass Analysis

For the release of *N*-glycans from the protein sample, 80  $\mu\text{g}$  of protein was digested with 4 U of PNGase F in PBS buffer (pH 7.6) in a final volume of 80  $\mu\text{L}$ . The protein was incubated for 2 h at 37  $^\circ\text{C}$  while shaking at 1200 rpm.

For the release of *O*-glycans from the protein sample, 80  $\mu\text{g}$  of protein was digested with a combination of 4U of sialidase and 4U of EngEF in PBS buffer (pH 7.6) in a final volume of 80  $\mu\text{L}$ . The sample was incubated for 1 h at 25  $^\circ\text{C}$  followed by 16 h at 37  $^\circ\text{C}$  while shaking at 1200 rpm.

For reduction of the disulfide bonds, the sample was reduced with a final concentration of 20 mM DTT in 3 M guanidine hydrochloride at 57  $^\circ\text{C}$  for 45 min.

To prepare for subunits, 50  $\mu\text{g}$  of protein was digested with 1  $\mu\text{L}$  of IdeZ in PBS buffer (pH 7.6) in a final volume of 50  $\mu\text{L}$ . The sample was incubated for 1 h at 37  $^\circ\text{C}$  while shaking at 1200 rpm.

### LC-MS/MS for Peptide *De Novo* Sequencing

0.5  $\mu\text{g}$  of peptides were separated by online reversed-phase chromatography on a Vanquish Neo UHPLC system (self-packed column 75  $\mu\text{m} \times 150$  mm; 3  $\mu\text{m}$  ReproSil-Pur C18 beads, 120  $\text{\AA}$ , Dr.Maisch GmbH, Ammerbuch, Germany) coupled to an Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific, Waltham, MA). Samples were eluted over a 60 min gradient at a flow rate of 300 nL/min. Mobile phase A consisted of 0.1% (v/v) formic acid in  $\text{H}_2\text{O}$  and Mobile phase B consisted of 0.1% (v/v) formic acid in 80% acetonitrile. The gradient was set as follows: 3–12%B in 1 min; 12–30% B in 45 min; 30–80% B in 5 min; 80–99% B in 3 min; 99% B for 5 min; 99–0% B in 1 min.

For *de novo* sequencing method development with Fetuin-A, raw data was collected for the samples using stepped HCD and EThcD separately to compare the fragmentation performance. For *de novo* sequencing of Etanercept, raw data was collected using EThcD fragmentation. Peptides were analyzed with a resolution setting of 60,000 in MS1 scans. MS1 scans were obtained with a standard automatic gain control (AGC) target of 400,000, a maximum injection time of 50 ms, and a scan range of 350–2000  $m/z$ . The precursors were selected with a 2  $m/z$  window and fragmented by stepped high-energy collision dissociation (HCD) or electron-transfer high-energy collision dissociation (EThcD) separately in a duty cycle of 3 s. The stepped HCD fragmentation included steps of 27, 35, and 40% normalized collision energies (NCE). EThcD fragmentation was performed with calibrated charge-dependent electron-transfer dissociation parameters and 30% NCE supplemental activation. For both fragmentation types, MS2 scans were acquired at a 30,000 resolution, a 400,000 AGC target, and a 250 ms maximum injection time. The charge state(s) were set as 3–7 and 2–7 for EThcD and HCD fragmentation, respectively. Dynamic Exclusion was set for 30 s.

### LC-MS/MS for *N*-Glycosylation Site Identification

0.5  $\mu\text{g}$  of peptides were separated by online reversed-phase chromatography on an EASY-nLC1000 system (self-packed column 75  $\mu\text{m} \times 150$  mm; 3  $\mu\text{m}$  ReproSil-Pur C18 beads, 120  $\text{\AA}$ , Dr.Maisch GmbH, Ammerbuch, Germany) coupled to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific, Waltham, MA). Peptides were eluted over a 60 min gradient at a flow rate of 300 nL/min. Mobile phase A consisted of 0.1% (v/v) formic acid in  $\text{H}_2\text{O}$ , and mobile phase B consisted of 0.1% (v/v) formic acid in acetonitrile. The gradient was set as follows: 1–4%B in 1 min; 4–26% B in 43 min; 26–32% B in 5 min; 32–90% B in 2 min; 90% B for 9 min.

Peptides were analyzed with a resolution setting of 60,000 in MS1 scans. MS1 scans were obtained with an AGC target of 3,000,000, a maximum injection time of 20 ms, and a scan range of 150–1700  $m/z$ . The top 20 precursors were selected with a 1.2  $m/z$  window and fragmented by HCD with 27% NCE. MS2 scans were acquired at a 15,000 resolution, with an AGC target of 100,000 and a maximum injection time of 100 ms. Dynamic exclusion was set for 30 s.

### LC-MS/MS for *N*-Glycopeptide Analysis

*N*-Glycopeptide analysis was performed as described previously with slight modifications.<sup>36</sup> 0.5  $\mu\text{g}$  of peptides were separated by online reversed-phase chromatography on an EASY-nLC1200 system (self-packed column 75  $\mu\text{m} \times 150$  mm; 3  $\mu\text{m}$  ReproSil-Pur C18 beads, 120  $\text{\AA}$ , Dr.Maisch GmbH, Ammerbuch, Germany) coupled to an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific, Waltham, MA). Samples were eluted over a 60 min gradient at a flow rate of 300 nL/min. Mobile phase A consisted of 0.1% (v/v) formic acid in  $\text{H}_2\text{O}$ , and mobile phase B consisted of 0.1% (v/v) formic acid in 80% acetonitrile. The gradient was set as follows: 2–8%B in 1 min; 12–30% B in 45 min; 30–50% B in 5 min; 50–100% B in 3 min; 100% B for 5 min.

Orbitrap spectra (AGC target 400,000, maximum injection time of 50 ms) were collected from 375 to 2000  $m/z$  at a resolution of 120,000 followed by oxonium ion-triggered data-dependent HCD MS/MS at a resolution of 30,000 using an isolation width of 1  $m/z$  for 20% collision energy and 2  $m/z$  for 33% collision energy. Charge state screening was enabled to reject unassigned and singly charged ions. A dynamic exclusion time of 30 s was used to discriminate against previously selected ions.

### LC-MS/MS for *O*-Glycopeptide Analysis

0.5  $\mu\text{g}$  of glycopeptides were separated by online reversed-phase chromatography on a Vanquish Neo UHPLC system coupled to an Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific, Waltham, MA). The LC-MS/MS parameters were the same as those for peptide *de novo* sequencing with EThcD fragmentation.

### Intact Mass Analysis

Intact proteins were analyzed using an Agilent 1290 Infinity II LC coupled to a 6545 Q-TOF mass spectrometer (Agilent, Santa Clara, CA). 1  $\mu\text{g}$  of protein was injected for each LC/MS run. Protein sample was separated with an Agilent PLRP-S column (1.0  $\times$  50 mm, 5  $\mu\text{m}$ , 1000  $\text{\AA}$ ) and eluted over a 12 min gradient (hold at 5% B for 5 min, 5–95% B for 5 min, 95% hold for 2 min) at a flow rate of 0.300 mL/min. Mobile phase A was made up of water with 0.1% formic acid, while mobile phase B was made up of acetonitrile with 0.1% formic acid. The mass spectrometry instrument parameters were set as follows: the dry gas flow rate was set at 10.0 L/min at 325  $^\circ\text{C}$ , the nebulizer was set at 50 psig, the capillary voltage was set at 4.5 kV and the scan range was from 500 to 3000  $m/z$  at 1 Hz. The capillary voltage was set at 4.5 kV and the scan range was from 500 to 3200  $m/z$  at 1 Hz for Fetuin, reduced, or middle-down TNFR:Fc-fusion biologics samples digested by IdeZ and from 500 to 6000  $m/z$  at 1 Hz for native TNFR:Fc-fusion biologics samples. The LC/MS raw data was processed using MassHunter BioConfirm (Version 10.0, Agilent, Santa Clara, CA, USA) to deconvolute the protein average masses.

### Released *N*-Glycans Analysis

The labeled released *N*-glycans were analyzed on a Waters UPLC-Xevo G2-S QTOF system (Waters, Milford, MA). The separation of

*N*-glycans was conducted at 60 °C using a Glycan BEH Amide column (2.1 × 150 mm, 1.7 μm, 130 Å, Waters, Milford, MA). Mobile phase A was 50 mM ammonium formate in water (pH 4.4), while mobile phase B was 100% acetonitrile. The gradient was set as 30–47% A over 33 min at a flow rate of 0.4 mL/min, 47–80% A in 2 min, and maintained at 80% A for 3 min at a flow rate of 0.25 mL/min, and 80–30% A in 1 min and maintained at 30% A for 5.00 min at a flow rate of 0.4 mL/min for re-equilibration. The FLR detector was normalized and then set at 330 nm excitation wavelength and 420 nm emission wavelength. The MS settings were set as follows: scan range, 200–2000 *m/z*; capillary voltage, 3.0 kV; cone voltage, 40 V; desolvation temperature, 350 °C; source temperature, 120 °C. Acquired MS data of the glycans were automatically processed using the UNIFI 1.9.3 Scientific Information System, and a Waters Glycan glucose unit (GU) library was used for glycan identification. The UPLC retention times of the labeled glycans were first converted to GU values based on a cubic spline calibration curve against a dextran ladder labeled with 2-AB. These GU values were then searched against the Glycan GU library for glycan structural identification and were further confirmed using accurate mass data. The library searches used a GU tolerance of 0.15 GU and a mass error of 10 ppm.

### Data Processing for Peptide *De Novo* Sequencing and Protein Assembly

Automated peptide *de novo* sequencing was performed with PEAKS AB (version 2.0, Bioinformatics Solutions Inc.).<sup>11</sup> Scans were not filtered or merged, and precursors were corrected by mass only. Precursor and product mass tolerance were set to 10 ppm and 0.02 Da, respectively. Enzymes were specified on a sample-by-sample basis. Carbamidomethylation (Cys +57.02 Da) was set as a fixed modification, with oxidation (Met +15.99 Da) as a variable modification. The peptide candidates with average local confidence (ALC) scores above 50 were used for protein assembly. The peptide candidates were subjected to the protein assembly process, performed with the Multiple Contigs & Scaffolding (MuCS) algorithm using the default parameters.<sup>37</sup> All the Ile/Leu were manually verified with the diagnostic *w*<sup>-</sup>ion in the EThcD spectrum, with the characteristic loss of C<sub>3</sub>H<sub>7</sub> (−43.05 Da) for Leu or C<sub>2</sub>H<sub>5</sub> (−29.04 Da) for Ile from the side chain of particular *z*-ions.

### Sequence Validation for Protein *De Novo* Sequencing

Sequence validation was performed with PEAKS AB (version 2.0, Bioinformatics Solutions Inc.) using the *de novo* sequencing output sequences as reference sequences with default parameters.

### Data Process for *N*-Glycosylation Site Identification

Quantitative <sup>18</sup>O labeling *N*-glycosylation sites analysis was performed with Biopharma Finder (version 4.0, Thermo Fisher Scientific, Waltham, MA) to search against the *de novo* sequencing output sequences, with potential Asn to Asp conversions caused by the deglycosylation of the Asn-linked glycan by PNGase F. The parameters for peptide mapping included: enzyme, partial pepsin; missed cleavages allowed, two; fixed modification, carbamidomethylation (Cys +57.02 Da); variable modifications, Asp-Asn substitute (Asp −0.98 Da), <sup>18</sup>O labeling (Asp +2.00 Da), and oxidation (Met +15.99 Da); peptide tolerance 10 ppm; and MS/MS tolerance 0.02 Da.

Endoglycosidases cocktail-treated mass spectrometric data were analyzed using pFind (version 3.2.0)<sup>38</sup> to search against the *de novo* sequencing output sequences, with potential Asn to Asp conversions. The parameters included: enzyme, partial pepsin; missed cleavages allowed, two; fixed modification, Carbamidomethylation (Cys +57.02 Da); variable modifications, Asp-Asn substitute (Asp −0.98 Da), GlcNAc (Asp +202.10 Da), GlcNAc-Fuc (Asp +348.15 Da), and oxidation (Met +15.99 Da); precursor tolerance 20 ppm; and fragment tolerance 20 ppm.

### Data Process for *N*-Glycopeptide Identification

*N*-Glycopeptide identification was performed with Biopharma Finder (version 5.2, Thermo Fisher Scientific, Waltham, MA) to search

against the *de novo* sequencing output sequences with corrected Asn-Asp conversion. The following parameters were used: carbamidomethyl (Cys +57.02 Da) as the fixed modification and oxidation (Met +15.99 Da); Lys loss on C terminal (−128.095 Da) and the built-in *N*-glycan repertoire for Chinese hamster ovary cell lines (Table S6) as the variable modification; precursor tolerance 20 ppm; and fragment tolerance 20 ppm. A minimum confidence score of 95% was set for the peptide identification. All *N*-glycopeptide spectra were manually reviewed to confirm the accuracy of assignments and remove potential false positives.

### Data Process for *O*-Glycopeptide Identification

For *O*-glycosylation site identification, *O*-glycan data were analyzed by pGlyco (version 3.1).<sup>39</sup> The parameters of pGlyco included a specific enzyme, allowing for up to five missed cleavages. The fixed modification was carbamidomethyl (Cys +57.02 Da) and the variable modification was oxidation (Met +15.99 Da). Peptide lengths range from 6 to 40 amino acids with a precursor tolerance of 10 ppm and a fragment tolerance of 20 ppm. The glycopeptide FDR was set at 0.01. The *O*-glycan database consisted predominantly of *O*-glycans expressed in CHO cells, with the inclusion of reported *O*-glycans specific to Etanercept (Table S7).<sup>23,40,41</sup> The total score threshold was set to >15.<sup>42</sup>

## ■ ASSOCIATED CONTENT

### Data Availability Statement

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE<sup>43</sup> partner repository (<https://www.ebi.ac.uk/pride/archive/>) with the data set identifier PXD055221 and PXD059329.

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacsau.4c00960>.

(Figures S1–S10) Analysis of Fetuin-A, including amino acid sequence, deglycosylation effects, peptide mapping, fragmentation comparisons (EThcD vs HCD), and I-L spectra; (Figures S12–S28) analysis of Etanercept and TNFR:Fc-fusion biologics, covering deglycosylation, mass spectrometry (ESI-Q TOF), peptide mapping, glycosylation profiling, and variant identification; (Tables S1–S3) summaries of TNFR:Fc-fusion biopharmaceuticals, *de novo* sequencing performance, and molecular mass comparisons (PDF)

(Figure S11) I-L spectra of Fetuin-A manually checked and corrected (PDF)

(Figure S29) *N*-Glycopeptide MS/MS spectrum of TNFR:Fc-fusion biopharmaceuticals (PDF)

(Figure S30) *O*-Glycopeptide MS/MS spectrum of TNFR:Fc-fusion biopharmaceuticals (PDF)

(Table S4) *N*-Glycan composition identified by glycopeptide analysis (XLSX)

(Table S5) *O*-Glycopeptide identification by pGlyco (XLSX)

(Table S6) *N*-Glycan database (PDF)

(Table S7) *O*-Glycan database (XLSX)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Liuqing Wen** – Analytical Research Center for Organic and Biological Molecules, State Key Laboratory of Drug Research, Carbohydrate-Based Drug Research Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China; School of Chinese Materia Medica, Nanjing University of Chinese Medicine, Nanjing, Jiangsu

210023, China; [orcid.org/0000-0001-9187-7999](https://orcid.org/0000-0001-9187-7999);

Email: [lwen@simmm.ac.cn](mailto:lwen@simmm.ac.cn)

**Hu Zhou** – Analytical Research Center for Organic and Biological Molecules, State Key Laboratory of Drug Research, Carbohydrate-Based Drug Research Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China; School of Chinese Materia Medica, Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210023, China; University of the Chinese Academy of Sciences, Beijing 100049, China; School of Pharmaceutical Science and Technology, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China; [orcid.org/0000-0001-7006-4737](https://orcid.org/0000-0001-7006-4737);  
Email: [zhouhu@simmm.ac.cn](mailto:zhouhu@simmm.ac.cn)

## Authors

**Jing Gao** – Analytical Research Center for Organic and Biological Molecules, State Key Laboratory of Drug Research, Carbohydrate-Based Drug Research Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China

**Hongxu Chen** – School of Chinese Materia Medica, Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210023, China

**Hongrui Yin** – NMPA Key Laboratory for Quality Control of Therapeutic Monoclonal Antibodies, Shanghai Institute for Food and Drug Control, Shanghai 201203, China

**Xin Chen** – School of Chinese Materia Medica, Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210023, China

**Zhicheng Yang** – Analytical Research Center for Organic and Biological Molecules, State Key Laboratory of Drug Research, Carbohydrate-Based Drug Research Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China; University of the Chinese Academy of Sciences, Beijing 100049, China; [orcid.org/0000-0001-7477-2984](https://orcid.org/0000-0001-7477-2984)

**Yuqiu Wang** – Analytical Research Center for Organic and Biological Molecules, State Key Laboratory of Drug Research, Carbohydrate-Based Drug Research Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China; Department of Otolaryngology, Eye & ENT Hospital, Fudan University, Shanghai 200031, China; [orcid.org/0009-0001-8931-8853](https://orcid.org/0009-0001-8931-8853)

**Jianhong Wu** – Thermo Fisher Scientific, Shanghai 201206, China

**Yinping Tian** – Analytical Research Center for Organic and Biological Molecules, State Key Laboratory of Drug Research, Carbohydrate-Based Drug Research Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China; [orcid.org/0000-0002-6998-0612](https://orcid.org/0000-0002-6998-0612)

**Hong Shao** – NMPA Key Laboratory for Quality Control of Therapeutic Monoclonal Antibodies, Shanghai Institute for Food and Drug Control, Shanghai 201203, China

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/jacsau.4c00960>

## Author Contributions

<sup>†</sup>J.G., H.C. and H.Y. contributed equally to this work.

## Author Contributions

J.G., H.C., H.Y., L.W., and H.Z. designed the study and prepared and revised the manuscript. J.G., H.C., H.Y., X.C., Z.Y., Y.W., and Y.T. performed the experiments and analyzed the experimental data. J.W. provided technical and material support. L.W., and H.Z. supervised the studies. All authors have read and approved the article.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work is supported by the Strategic Priority Research Program of the Chinese Academy of Sciences, grant no. XDB0830000, National Key Research and Development Program (2022YFA1302902 to H.Z.), the Talent Plan of Shanghai Branch, Chinese Academy of Sciences (CASSHB-QNPD-2023-018), the Shanghai Municipal Science and Technology Major Projects etttand the and the National Science Fund for Distinguished Young Scholars, grant no. 22425703. We thank Dr. Zhibiao Mai and Prof. Gong Zhang, for the instruction in the use of MuCS software and protein assembly.

## REFERENCES

- (1) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in health and disease. *Nat. Rev. Nephrol.* **2019**, *15* (6), 346–366.
- (2) Schjoldager, K. T.; Narimatsu, Y.; Joshi, H. J.; Clausen, H. Global view of human protein glycosylation pathways and functions. *Nat. Rev. Mol. Cell Biol.* **2020**, *21* (12), 729–749.
- (3) Tian, Y.; Ma, S.; Wen, L. Towards chemoenzymatic labeling strategies for profiling protein glycosylation. *Curr. Opin. Chem. Biol.* **2024**, *80*, No. 102460.
- (4) Wang, Y.; Yuan, R.; Liang, B.; Zhang, J.; Wen, Q.; Chen, H.; Tian, Y.; Wen, L.; Zhou, H. A "One-Step" Strategy for the Global Characterization of Core-Fucosylated Glycoproteome. *JACS Au* **2024**, *4* (5), 2005–2018.
- (5) Zhang, H.; Yi, E. C.; Li, X.-j.; Mallick, P.; Kelly-Spratt, K. S.; Masselon, C. D.; Camp, D. G.; Smith, R. D.; Kemp, C. J.; Aebersold, R. High Throughput Quantitative Analysis of Serum Proteins Using Glycopeptide Capture and Liquid Chromatography Mass Spectrometry. *Mol. Cell. Proteomics* **2005**, *4* (2), 144–155.
- (6) Sen, K. L.; Tang, W. H.; Nayak, S.; Kil, Y. J.; Bern, M.; Ozoglu, B.; Ueberheide, B.; Davis, D.; Becker, C. Automated Antibody De Novo Sequencing and Its Utility in Biopharmaceutical Discovery. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (5), 803–810.
- (7) Bandeira, N.; Pham, V.; Pevzner, P.; Arnott, D.; Lill, J. R. Automated de novo protein sequencing of monoclonal antibodies. *Nat. Biotechnol.* **2008**, *26* (12), 1336–1338.
- (8) Peng, W.; Pronker, M. F.; Snijder, J. Mass Spectrometry-Based De Novo Sequencing of Monoclonal Antibodies Using Multiple Proteases and a Dual Fragmentation Scheme. *J. Proteome Res.* **2021**, *20* (7), 3559–3566.
- (9) Lu, H.; Zhang, Y.; Yang, P. Advancements in mass spectrometry-based glycoproteomics and glycomics. *National Science Review* **2016**, *3* (3), 345–364.
- (10) Cao, W.; Liu, M.; Kong, S.; Wu, M.; Zhang, Y.; Yang, P. Recent Advances in Software Tools for More Generic and Precise Intact Glycopeptide Analysis. *Mol. Cell Proteomics* **2021**, *20*, No. 100060.
- (11) Tran, N. H.; Rahman, M. Z.; He, L.; Xin, L.; Shan, B.; Li, M. Complete De Novo Assembly of Monoclonal Antibody Sequences. *Sci. Rep.* **2016**, *6* (1), 31730.
- (12) Stavenhagen, K.; Hinneburg, H.; Thaysen-Andersen, M.; Hartmann, L.; Silva, D. V.; Fuchser, J.; Kaspar, S.; Rapp, E.; Seeberger, P. H.; Kolarich, D. Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass

spectrometry signal strengths using synthetic peptides and glycopeptides. *Journal of Mass Spectrometry* **2013**, *48* (6), 627–639.

(13) Leymarie, N.; Zaia, J. Effective Use of Mass Spectrometry for Glycan and Glycopeptide Structural Analysis. *Anal. Chem.* **2012**, *84* (7), 3040–3048.

(14) Bagdonaite, I.; Malaker, S. A.; Polasky, D. A.; Riley, N. M.; Schjoldager, K.; Vakhrushev, S. Y.; Halim, A.; Aoki-Kinoshita, K. F.; Nesvizhskii, A. I.; Bertozzi, C. R.; Wandall, H. H.; Parker, B. L.; Thaysen-Andersen, M.; Scott, N. E. Glycoproteomics. *Nat. Rev. Methods Primers* **2022**, *2* (1), 48.

(15) Cao, L.; Lih, T. M.; Hu, Y.; Schnaubelt, M.; Chen, S. Y.; Zhou, Y.; Guo, C.; Dong, M.; Yang, W.; Eguez, R. V.; Chen, L.; Clark, D. J.; Sodhi, A.; Li, Q. K.; Zhang, H. Characterization of core fucosylation via sequential enzymatic treatments of intact glycopeptides and mass spectrometry analysis. *Nat. Commun.* **2022**, *13* (1), 3910.

(16) Chuang, H.-Y.; Huang, C.-C.; Hung, T.-C.; Huang, L.-Y.; Chiu, C.-W.; Chu, K.-C.; Liao, J.-Y.; You, T.-H.; Wu, C.-Y.; Chao, P.; Shivatare, S. S.; Zeng, Y.-F.; Tsai, C.-S.; Lin, N.-H.; Wu, C.-Y. Development of biotinylated and magnetic bead-immobilized enzymes for efficient glyco-engineering and isolation of antibodies. *Bioorganic Chemistry* **2021**, *112*, No. 104863.

(17) Meng, X.; Li, L.; Wang, X. An integrated strategy for the construction of a species-specific glycan library for mass spectrometry-based intact glycopeptide analyses. *Talanta* **2021**, *234*, No. 122626.

(18) Yang, G.; Hu, Y.; Sun, S.; Ouyang, C.; Yang, W.; Wang, Q.; Betenbaugh, M.; Zhang, H. Comprehensive Glycoproteomic Analysis of Chinese Hamster Ovary Cells. *Anal. Chem.* **2018**, *90* (24), 14294–14302.

(19) Chen, Y.; Cao, J.; Yan, G.; Lu, H.; Yang, P. Two-step protease digestion and glycopeptide capture approach for accurate glycosite identification and glycoprotein sequence coverage improvement. *Talanta* **2011**, *85* (1), 70–75.

(20) Frese, C. K.; Altelaar, A. F. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2012**, *84* (22), 9668–9673.

(21) Tarentino, A. L.; Gomez, C. M.; Plummer, T. H. Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F. *Biochemistry* **1985**, *24*, 4665–4671.

(22) Zhokhov, S. S.; Kovalyov, S. V.; Samgina, T. Y.; Lebedev, A. T. An EThcD-Based Method for Discrimination of Leucine and Isoleucine Residues in Tryptic Peptides. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (8), 1600–1611.

(23) Houel, S.; Hilliard, M.; Yu, Y. Q.; McLoughlin, N.; Martin, S. M.; Rudd, P. M.; Williams, J. P.; Chen, W. N. and O-Glycosylation Analysis of Etanercept Using Liquid Chromatography and Quadrupole Time-of-Flight Mass Spectrometry Equipped with Electron-Transfer Dissociation Functionality. *Anal. Chem.* **2014**, *86* (1), 576–584.

(24) Peral, B.; San Millán, J. L.; Castello, R.; Moghetti, P.; Escobar-Morreale, H. F. Comment: the methionine 196 arginine polymorphism in exon 6 of the TNF receptor 2 gene (TNFRSF1B) is associated with the polycystic ovary syndrome and hyperandrogenism. *J. Clin. Endocrinol. Metab.* **2002**, *87*, 3977–3983.

(25) Morita, C.; Horiuchi, T.; Tsukamoto, H.; Hatta, N.; Kikuchi, Y.; Arinobu, Y.; Otsuka, T.; Sawabe, T.; Harashima, S.-I.; Nagasawa, K.; Niho, Y. Association of tumor necrosis factor receptor type II polymorphism 196R with systemic lupus erythematosus in the Japanese: Molecular and functional analysis. *Arthritis Rheum* **2001**, *44* (12), 2819–2827.

(26) Tan, Q.; Guo, Q.; Fang, C.; Wang, C.; Li, B.; Wang, H.; Li, J.; Guo, Y. Characterization and comparison of commercially available TNF receptor 2-Fc fusion protein products. *mAbs* **2012**, *4* (6), 761–774.

(27) Higel, F.; Seidl, A.; Sorgel, F.; Friess, W. N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. *Eur. J. Pharm. Biopharm* **2016**, *100*, 94–100.

(28) Martinez, V. P. M.; Tierrablanca-Sanchez, L.; Espinosa-de la Garza, C. E.; Juarez-Bayardo, L. C.; Pina-Lara, N.; Santoyo, G. G.; Perez, N. O. Functional analysis of glycosylation in Etanercept: Effects over potency and stability. *Eur. J. Pharm. Sci.* **2020**, *153*, No. 105467.

(29) Robbins, P. W.; Trimble, R. B.; Wirth, D. F.; Hering, C.; Maley, F.; Maley, G. F.; Das, R.; Gibson, B. W.; Royal, N.; Biemann, K. Primary structure of the Streptomyces enzyme endo-beta-N-acetylglucosaminidase H. *J. Biol. Chem.* **1984**, *259* (12), 7577–7583.

(30) Wei, F.; Zang, L.; Zhang, P.; Zhang, J.; Wen, L. Concise chemoenzymatic synthesis of N-glycans. *Chem.* **2024**, *10* (9), 2844–2860.

(31) Tian, Y.; Wang, Y.; Yin, H.; Luo, Y.; Wei, F.; Zhou, H.; Wen, L. A Sensitive and Reversible Labeling Strategy Enables Global Mapping of the Core-Fucosylated Glycoproteome on Cell Surfaces. *Angew. Chem., Int. Ed. Engl.* **2022**, *61* (49), No. e202206802.

(32) Luo, Y.; Wang, Y.; Tian, Y.; Zhou, H.; Wen, L. “Two Birds One Stone” Strategy for the Site-Specific Analysis of Core Fucosylation and O-GlcNAcylation. *J. Am. Chem. Soc.* **2023**, *145* (29), 15879–15887.

(33) Koutsoulis, D.; Landry, D.; Guthrie, E. P. Novel endo-alpha-N-acetylgalactosaminidases with broader substrate specificity. *Glycobiology* **2008**, *18* (10), 799–805.

(34) Higuchi, Y.; Eshima, Y.; Huang, Y.; Kinoshita, T.; Sumiyoshi, W.; Nakakita, S.-i.; Takegawa, K. Highly efficient transglycosylation of sialo-complex-type oligosaccharide using Coprinopsis cinerea endo-glycosidase and sugar oxazoline. *Biotechnol. Lett.* **2017**, *39* (1), 157–162.

(35) Huang, W.; Giddens, J.; Fan, S.-Q.; Toonstra, C.; Wang, L.-X. Chemoenzymatic Glycoengineering of Intact IgG Antibodies for Gain of Functions. *J. Am. Chem. Soc.* **2012**, *134* (29), 12308–12318.

(36) Shen, J.; Jia, L.; Dang, L.; Su, Y.; Zhang, J.; Xu, Y.; Zhu, B.; Chen, Z.; Wu, J.; Lan, R.; Hao, Z.; Ma, C.; Zhao, T.; Gao, N.; Bai, J.; Zhi, Y.; Li, J.; Zhang, J.; Sun, S. StrucGP: de novo structural sequencing of site-specific N-glycan on glycoproteins using a modularization strategy. *Nat. Methods* **2021**, *18* (8), 921–929.

(37) Mai, Z. B.; Zhou, Z. H.; He, Q. Y.; Zhang, G. Highly Robust de Novo Full-Length Protein Sequencing. *Anal. Chem.* **2022**, *94* (8), 3467–3475.

(38) Chi, H.; Liu, C.; Yang, H.; Zeng, W. F.; Wu, L.; Zhou, W. J.; Wang, R. M.; Niu, X. N.; Ding, Y. H.; Zhang, Y.; Wang, Z. W.; Chen, Z. L.; Sun, R. X.; Liu, T.; Tan, G. M.; Dong, M. Q.; Xu, P.; Zhang, P. H.; He, S. M. Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nat. Biotechnol.* **2018**, *36*, 1059–1061.

(39) Liu, M. Q.; Zeng, W. F.; Fang, P.; Cao, W. Q.; Liu, C.; Yan, G. Q.; Zhang, Y.; Peng, C.; Wu, J. Q.; Zhang, X. J.; Tu, H. J.; Chi, H.; Sun, R. X.; Cao, Y.; Dong, M. Q.; Jiang, B. Y.; Huang, J. M.; Shen, H. L.; Wong, C. C. L.; He, S. M.; Yang, P. Y. pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. *Nat. Commun.* **2017**, *8* (1), 438.

(40) Chen, X.; Liu, X.; Xiao, Z.; Liu, J.; Zhao, L.; Tan, W.-S.; Fan, L. Insights into the loss of protein sialylation in an fc-fusion protein-producing CHO cell bioprocess. *Appl. Microbiol. Biotechnol.* **2019**, *103* (12), 4753–4765.

(41) D’Atri, V.; Nováková, L.; Fekete, S.; Stoll, D.; Lauber, M.; Beck, A.; Guillard, D. Orthogonal Middle-up Approaches for Characterization of the Glycan Heterogeneity of Etanercept by Hydrophilic Interaction Chromatography Coupled to High-Resolution Mass Spectrometry. *Anal. Chem.* **2019**, *91* (1), 873–880.

(42) Hevér, H.; Xue, A.; Nagy, K.; Komka, K.; Vékely, K.; Drahos, L.; Révész, Á. Can We Boost N-Glycopeptide Identification Confidence? Smart Collision Energy Choice Taking into Account Structure and Search Engine. *J. Am. Soc. Mass Spectrom.* **2024**, *35* (2), 333–343.

(43) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Perez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, S.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaino, J. A. The PRIDE database and

related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **2019**, *47* (D1), D442–D450.