Comprehensive structural and positional profiling of intact complex *O*-glycans in biologic drugs using O-Glycoprotease (IMPa)

Andy Hanneman, Ph.D., Edwin Escobar, Ph.D., Saulius Vainauskas, Ph.D., Chris Taron, Ph.D., New England Biolabs

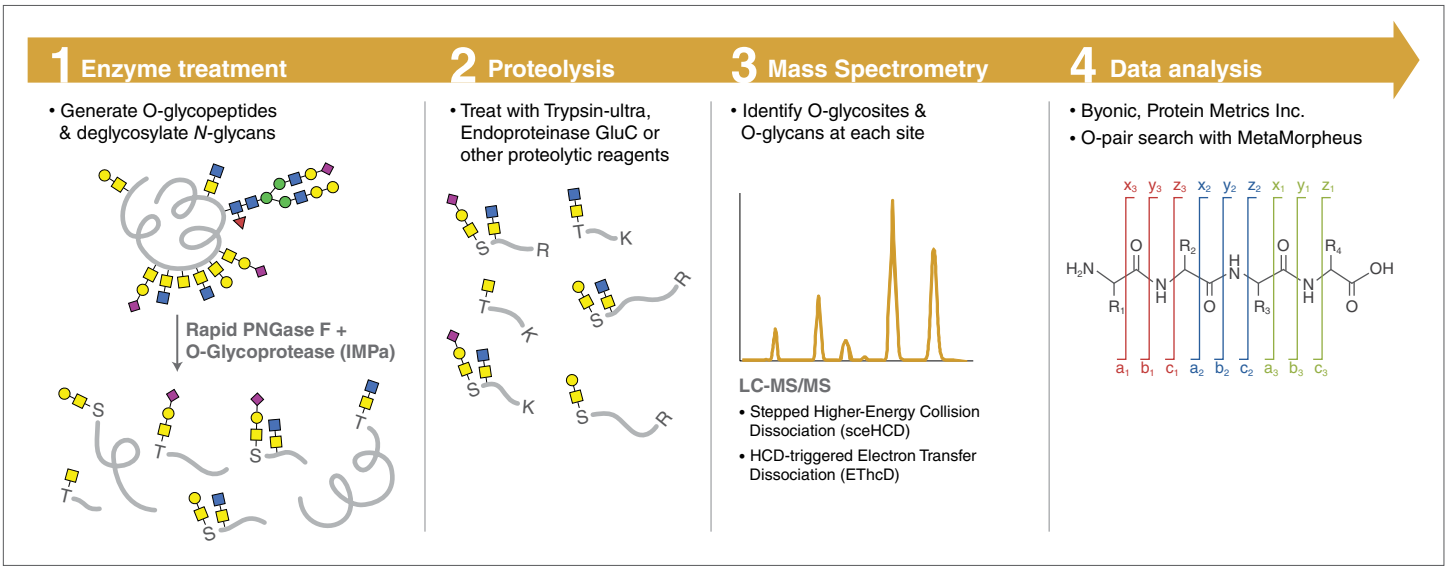
INTRODUCTION

Many important classes of biologic drugs are secreted glycoproteins (e.g., monoclonal antibodies, cytokines, growth factors) that possess post-translational carbohydrate modifications of asparagine (*N*-glycans) or serine/threonine residues (*O*-glycans). Glycan structure and composition can significantly influence key drug properties including stability, bioactivity, and pharmacokinetics. Thus, glycosylation is considered a critical quality attribute (CQA) that must be monitored during manufacturing processes, as well as to demonstrate biosimilarity of follow-on biologics (1). For these reasons, glycan composition profiling is often required for assessing batch-to-batch consistency. Nevertheless, *O*-glycosylated proteins are among the most challenging molecules presented to a biopharmaceutical analytical development laboratory because composition profiling of *O*-glycosylation has historically been daunting. One analytical approach involves releasing *O*-glycans from the glycoprotein prior to structural profiling. Unfortunately, the lack of a broad-specificity enzyme able to release complex *O*-glycans intact complicates their enzymatic characterization, whereas chemical release procedures present various technical challenges. Additionally, released glycan profiling provides no information regarding occupancy sites on the protein. *O*-glycosylation site analysis is further impeded by the absence of a clear amino acid consensus sequence defining possible *O*-glycosites. *O*-glycans can be studied at the glycopeptide or glycoprotein level. Nevertheless, characterizing an *O*-linked protein by reporting both occupancy (macro-heterogeneity) and the diverse *O*-glycan structures at each site (micro-heterogeneity) often evades even well-established peptide mapping strategies. As *O*-glycosylated proteins often contain stretches of serine or threonine residues, one of the main drivers of discovery has been a lack of high-performance enzymes reliably capable of cleaving proteins within clustered *O*-glycosylation domains. Therefore, glycopeptides prepared using standard enzymatic

MATERIALS

- O-Glycoprotease (IMPa) (NEB #P0761S)
- Rapid PNGase F (NEB #P0710S)
- Trypsin-ultra™, Mass Spectrometry Grade (NEB #P8101S)
- Endoproteinase GluC (NEB #P8100S)
- Dithiothreitol (DTT) (Thermo Fisher® No-Weigh format, A39255)
- Iodoacetamide (IAA) (Thermo Fisher No-Weigh format, A39271)
- Guanidine hydrochloride (8.0M, Thermo Fisher 24115)
- Methionine (Millipore Sigma, 64319)

FIGURE 1: **Overview of workflow for peptide and *O*-glycopeptide mapping**



proteolysis procedures (e.g., trypsin digestion) may often exhibit a high degree of structural heterogeneity and can be difficult to recover and particularly challenging to observe and characterize accurately by mass spectrometry.

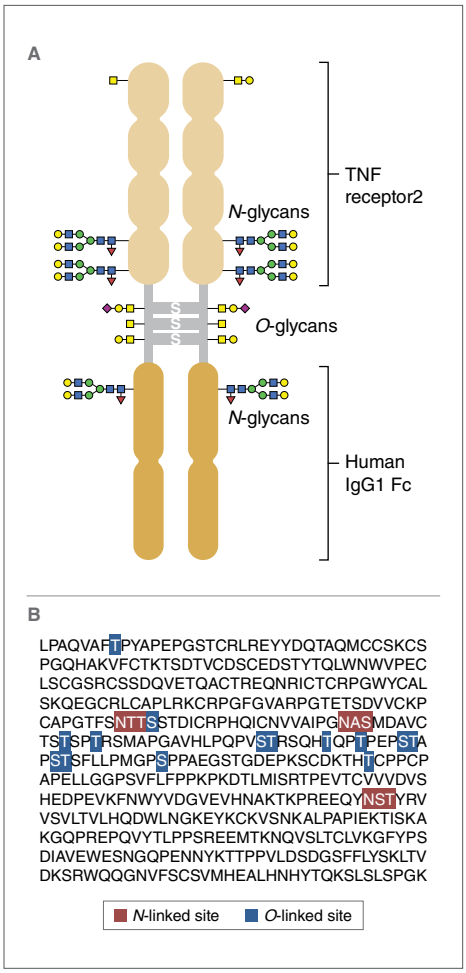
Applications of proteomic methods to address *O*-glycosylation have recently been undergoing rapid developments (2-4). New workflows that employ *O*-glycan-specific proteases (*O*-glycoproteases), enzymes that cleave a glycopeptide/glycoprotein adjacent to an *O*-glycosylated serine or threonine, are increasingly being used to generate and analyze *O*-glycopeptides (5). The first commercial enzyme of this kind was OgpA (OpeRATOR™ Genovis), a protease that cleaves proteins adjacent to *O*-glycans consisting of Gal-GalNAc (Core1). Peptides produced with OgpA have been used to map the macroheterogeneity of *O*-glycosites in therapeutic proteins (6). However, OpeRATOR does not cleave adjacent to truncated *O*-glycans consisting of a single GalNAc residue, or more complex branched or sialylated *O*-glycans that must first be digested with sialidase. As such, OgpA is not suited for analysis of *O*-glycan structural microheterogeneity. Recently, a new broad-specificity O-Glycoprotease (IMPa) that is capable of digesting *O*-glycosylated proteins containing sialylated and asialylated *O*-glycans of varying complexity has been released by NEB (7,8). In this Application Note, we present a protocol for the use of O-Glycoprotease (IMPa) in a typical peptide mapping workflow to characterize the complex biologic drug etanercept (Enbrel®). The workflow presented delivers on the long-sought aim of mapping *O*-glycosites to determine the *O*-glycan structures at each site in a single experiment (Figure 1, page 1).

METHODOLOGY AND RESULTS

In this Application Note, etanercept was employed to demonstrate the use of O-Glycoprotease (IMPa) to characterize *O*-glycosylation macro- and micro-heterogeneity. Etanercept, a homodimeric receptor-Fc (IgG1) fusion protein with 3 *N*-linked glycosylation sites and 14 or more occupied *O*-linked sites per chain (Figure 2), is especially challenging to characterize due to a considerable number of clustered *O*-glycosylation sites (9-13).

Here we demonstrate that O-Glycoprotease (IMPa) can be readily incorporated into an existing peptide map workflow, enabling identification of occupied *O*-glycosites and the range of complex *O*-glycan structures at each site. Methods employed for comprehensive peptide and *O*-glycopeptide site-mapping are described in a four-stage workflow (Figure 1, page 1). The overall scheme is a typical peptide mapping procedure suitable for a biotherapeutic protein characterization package. The added steps are introduction of O-Glycoprotease (IMPa) proteolysis as a first step, and use of an electron-based tandem mass spectrometry instrument method for *O*-glycopeptide analysis.

FIGURE 2: **Structure of Etanercept (Enbrel)**



A. Structure of etanercept, a homodimeric receptor-Fc fusion protein with three *N*-linked sites and multiple reported *O*-linked glycosylation sites per chain.

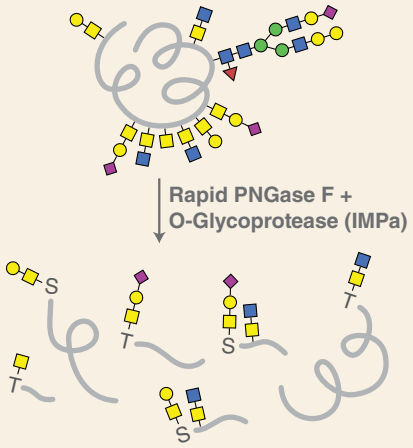
B. The amino acid sequence of etanercept highlighting *N*-linked sites, and 14 previously reported *O*-linked sites.

1 RAPID PNGASE F AND O-GLYCOPROTEASE (IMPa) TREATMENT

To start the peptide map workflow, etanercept was first processed using Rapid PNGase F for *N*-deglycosylation. After Rapid PNGase F treatment, etanercept was then processed with O-Glycoprotease (IMPa) to generate *O*-glycopeptides.

1. Optional step: etanercept was used for this study. To deformulate the commercial product, 12 mg of etanercept (NDC 58406-425-34 - AMGEN® Enbrel® - 25 mg/vial) was dialyzed using a Protein Slide-A-Lyzer cassette (Thermo Scientific, cat# A52971, 3mL) to 5 mM ammonium acetate (AA) solution using 3 buffer changes (500 mM AA for 2 hours, 50 mM AA for 2 hours, 5 mM AA overnight) at 4°C. The required sample volume, containing 50 µg in a 1.5 mL centrifuge tube, was taken to dryness using a refrigerated speed vac overnight at 10°C.
2. Dissolve 50 µg etanercept in 50 µL of 50 mM Tris-HCl, pH 7.5, containing 4 M guanidine-HCl, 20 mM methionine.
3. Reduction: add DTT (10 mM final concentration) and incubate for 45 minutes at 37°C.*
4. Alkylation: add IAA (37.5 mM final concentration) and incubate for 45 min at room temperature (RT).*
5. Quenching: add DTT (5 mM final concentration) and incubate for 10 min at RT.*
6. Buffer exchange in an ultrafiltration (UF) device Amicon 10 kDa (Millipore: UFC501096) using 3 x 400 µL additions of 50 mM Tris-HCl, pH 7.5, 20 mM methionine (Millipore PHR-124) to a final volume of 20 µL.
7. Add 5 uL of 5x Rapid PNGase F buffer and 1 µL of Rapid PNGase F. Incubate at 50°C for 10 minutes in Eppendorf ThermoMixer at 500 rpm.
8. UF (2 x 400 µL 50 mM Tris-HCl pH 7.5, 20 mM methionine) to remove *N*-glycans (collect the flow-through and dry-down the *N*-linked glycans for a separate *N*-glycan analysis, if desired).
9. Recover the UF retentate (~40-50 µL) and adjust to 50 µL using 50 mM Tris-HCl, pH 7.5, 20 mM methionine.
10. Add 5 µL O-Glycoprotease (IMPa).
11. Incubate for 5 hours at 37°C at 500 rpm using an Eppendorf ThermoMixer.

* For optimal mass spectrometry data analysis, stabilize cysteine side chains by reduction and alkylation prior to deglcoyslation.

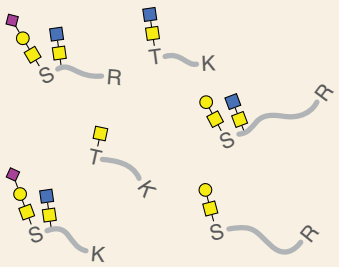


• Generate *O*-glycopeptides & deglycosylate *N*-glycans

2 PROTEOLYSIS

In the workflow presented here, we used Trypsin-ultra and GluC, however, note that other proteases can also be used by simply carrying out the digestion as recommended for the protease of choice.

1. Incubate overnight (16 hours at 37°C) with Trypsin-ultra, or Endoproteinase GluC at a 1:25 enzyme to substrate ratio using an Eppendorf ThermoMixer at 500 rpm.
2. Dilute the samples to 2% acetonitrile/0.1% formic acid final concentration for LC-MS/MS.
3. If required, samples can first be concentrated by centrifugal evaporation prior to LC-MS/MS.



• Treat with Trypsin-ultra, Endoproteinase GluC or other proteolytic reagents

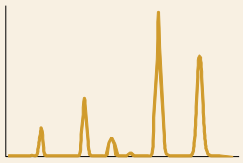
3 MASS SPECTROMETRY ANALYSIS

Two separate tandem mass spectrometry (MS/MS) instrument methods were applied for *O*-glycopeptide characterization.

Two injections of the same sample treated with O-Glycoprotease (IMPa) and either Trypsin-ultra or Endoproteinase GluC were analyzed using these two methods (3,4):

1. Stepped higher-energy collisional dissociation (sceHCD)
2. Oxonium ion-triggered electron dissociation higher-energy collisional dissociation (EThcD)

To characterize small glycopeptides with a single potential *O*-glycosylation site sceHCD is the appropriate instrument method. For larger glycopeptides with more than one serine/threonine residue, the preferred MS/MS method is oxonium ion triggered EThcD because it generates spectra conveying highly specific site occupancy information. (Note: O-Glycoprotease (IMPa) will not cleave between two adjacent occupied *O*-glycosylation sites). Refer to mass spectrometer instrument methods to see the parameters used (Appendix A: Instrument methods used).



LC-MS/MS

- Stepped Higher-Energy Collision Dissociation (sceHCD)
- HCD-triggered Electron Transfer Dissociation (EThcD)

- Identify O-glycosites & O-glycans at each site

4 DATA ANALYSIS

The data was evaluated using two data analysis packages:

1. Byonic and Byomap (Byos package) by Protein Metrics Incorporated (PMI)
2. O-Pair Search with Metamorpheus (14,15)

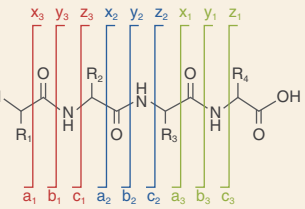
For peptide mapping, four etanercept digests were initially generated:

1. Trypsin-ultra
2. Endoproteinase GluC
3. Rapid PNGase F + Trypsin-ultra
4. Rapid PNGase F + Endoproteinase GluC

The LC-MS/MS Trypsin-ultra and Endoproteinase GluC peptide maps generated without O-Glycoprotease (IMPa) treatment each covered ~70-85% of the etanercept amino acid sequence, with the notable exception of the clustered *O*-glycosylated domain (~200–240 aa) (Figure 3). Some *O*-linked glycopeptides were observed without the use of O-Glycoprotease (IMPa), however, precise *O*-linked site localization was not typically obtained.

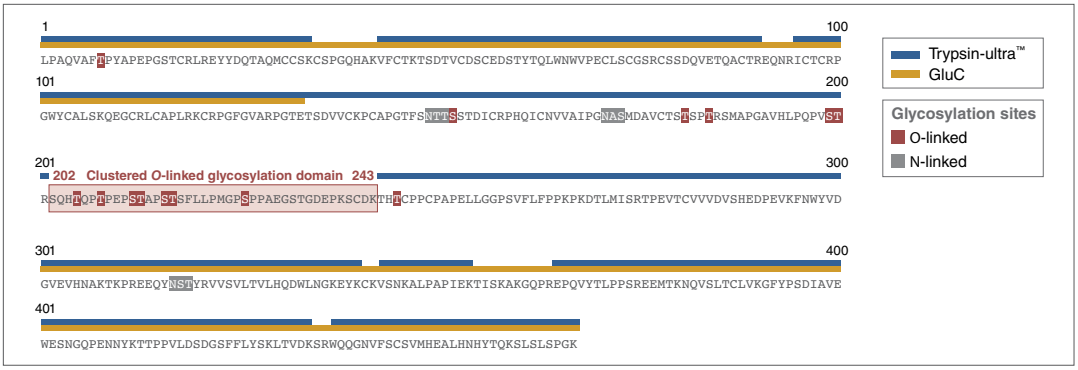
When additional digests (5 and 6 below) were prepared incorporating O-Glycoprotease (IMPa), the complete etanercept sequence including *O*-glycosylation micro-heterogeneity was obtained:

5. O-Glycoprotease (IMPa), Rapid PNGase F + Trypsin-ultra
6. O-Glycoprotease (IMPa), Rapid PNGase F + Endoproteinase GluC



- Byonic, Protein Metrics Inc.
- O-pair search with MetaMorpheus

FIGURE 3: **Peptide map sequence coverage obtained from digests 1-4**



The clustered *O*-linked glycosylation domain AA 202-243 requires O-Glycoprotease (IMPa) for characterization.

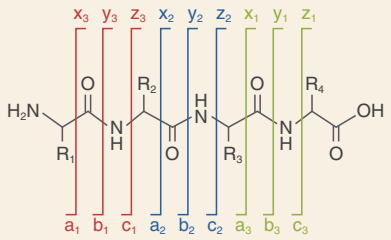
4 DATA ANALYSIS (CONT'D)

After O-Glycoprotease (IMPa) was added (steps 5 and 6), the peptide/glycopeptide digests achieved 100% amino acid sequence coverage, as well as O-glycosylation site micro-heterogeneity, providing complete characterization. By employing O-Glycoprotease (IMPa) prior to Trypsin-ultra or Endoproteinase GluC proteolysis, diverse glycopeptides are generated that are amenable to peptide-based LC-MS/MS analysis as needed to capture *O*-glycan micro-heterogeneity (Figure 4 and Table 1).

As mentioned in section 3, two complementary instrument methods were employed to analyze each digest:

1. for small glycopeptides, sceHCD.
2. for larger glycopeptides, EThcD.

Each sample type was prepared in duplicate and only high-scoring data is reported in Table 1. Data analysis used Byonic by Protein Metrics (PMI), and the MetaMorpheus *O*-glycosylation program O-Pair. Example *O*-glycopeptide IDs are shown in Figure 5, page 6.

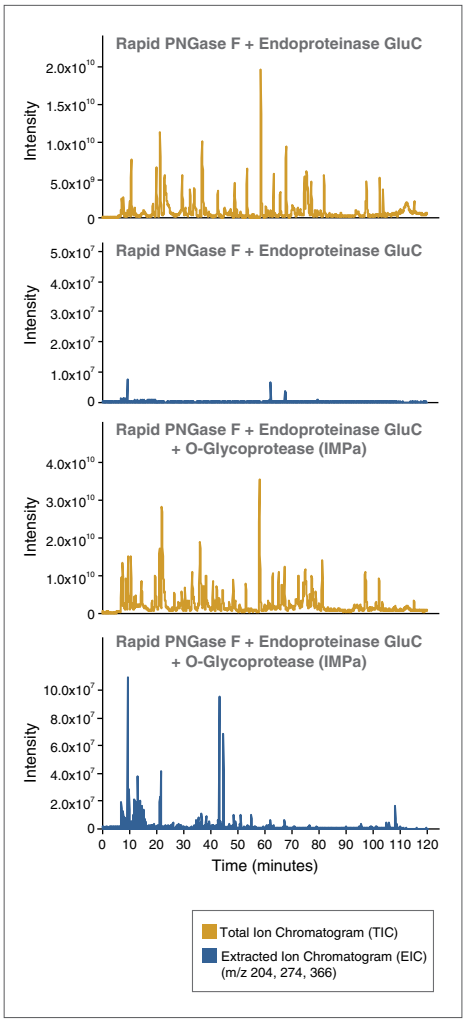


- Byonic, Protein Metrics Inc.
- O-pair search with MetaMorpheus

TABLE 1: **Reporting on etanercept micro-heterogeneity using O-Glycoprotease (IMPa)**

ID	#	PREVIOUSLY REPORTED SITE?	MICRO-HETEROGENITY DETECTED BY SITE				
			N	H1N1	H1A1	H1N1A1	H1N1A2
T	8	Y		●		●	●
T	181	Y				●	●
T	184	Y				●	●
S	186	Y	●	●		●	
S	199	Y	●	●		●	●
T	200	Y	●	●		●	●
T	205	Y	●	●		●	●
T	208	Y	●	●	●	●	●
S	212	Y		●		●	●
T	213	Y			●	●	
S	216	Y		●	●	●	
T	217	Y		●		●	
S	226	Y	●	●		●	●
T	245	Y		●		●	●

FIGURE 4: **Using O-Glycoprotease (IMPa) to generate *O*-glycopeptides in protease digests**

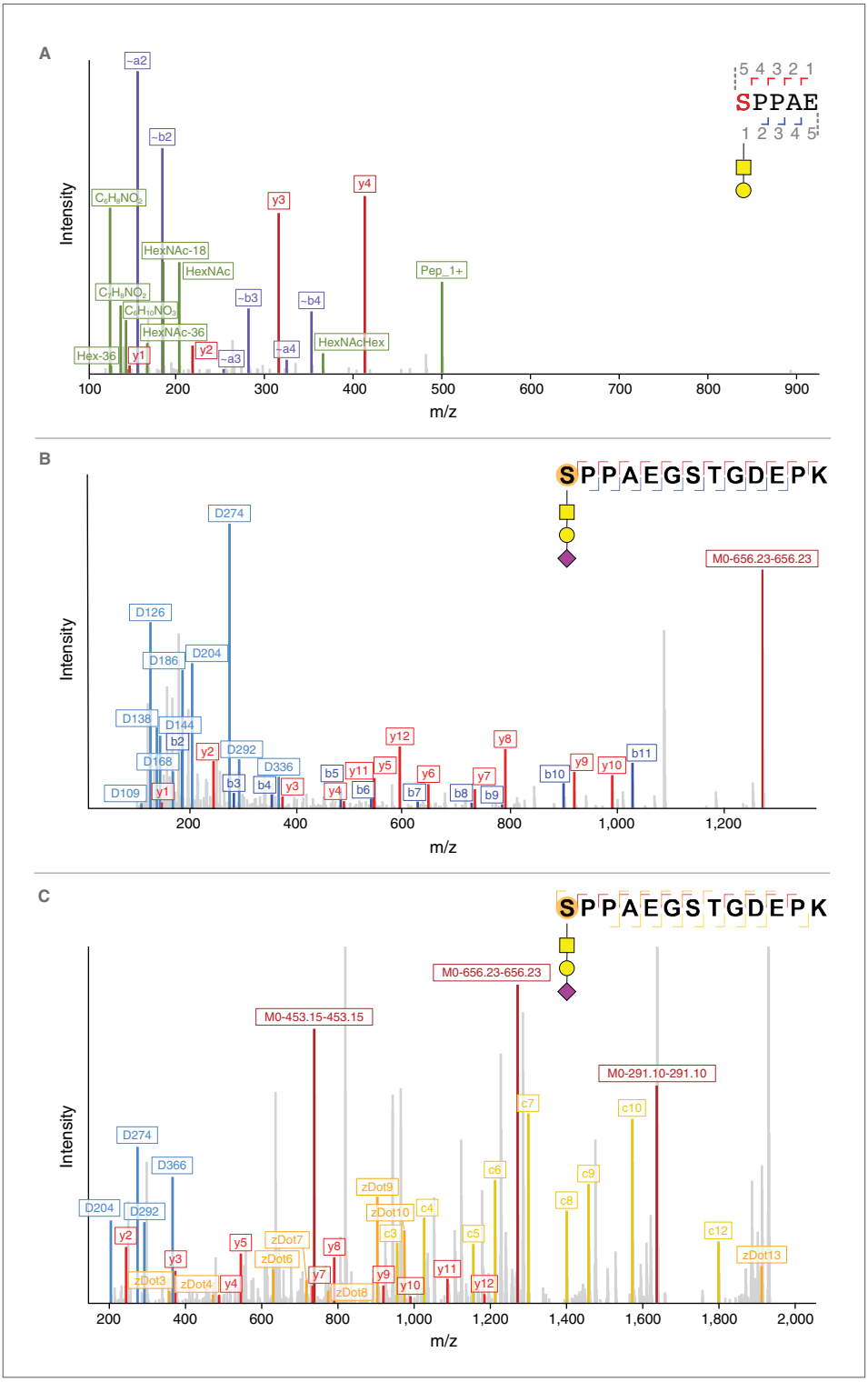


Total Ion chromatograms (TIC) and extracted ion chromatograms (EIC) show high-abundance oxonium ions (lower panel) when O-Glycoprotease (IMPa) is used to generate *O*-glycopeptides in protease digests.

CONCLUSION

In this Application Note, we showed that O-Glycoprotease (IMPa) can be used in a traditional peptide mapping workflow to generate a complete set of *O*-glycopeptides from a therapeutic glycoprotein. This permitted both comprehensive mapping of *O*-glycosylation sites and determination of the repertoire of complex *O*-glycan structures residing at each glycosite (i.e., analysis of macro-heterogeneity and micro-heterogeneity) in a single analysis. This method represents a significant advance in *O*-glycan composition profiling of biologic drugs and is suitable for inclusion in a typical biotherapeutic characterization package.

FIGURE 5: **Mass spectra showing diverse *O*-glycopeptides with O-Glycoprotease (IMPa) treatment**



Example MS/MS spectra from GluC and Trypsin-Ultra digests:

A) sceHCD of a GluC glycopeptide,

Paired HCD & EThcD MS/MS scans of a tryptic glycopeptide:

B) HCD, C) oxonium ion-triggered EThcD (light blue highlights oxonium ions)

APPENDIX A: INSTRUMENT METHODS USED

LC-MS/MS methods

Instrument: Thermo Orbitrap Eclipse with Easy-nLC using:

Nano LC method

Analytical column:
Easy-Spray PepMap C18 25 cm, 75 µm, 2 µm

Mobile phases:
A. 0.1% formic acid in H_2O
B. 0.1% formic acid in 80% ACN/20% H_2O
Sample load: 400 ng
Column temperature: 45°C

GRADIENT TABLE	
TIME (MINUTES)	%B
0	3
6	3
86	25
102	40
106	95
116	95
118	3
120	3

Method 1: sceHCD MS/MS instrument method parameters:

This method is employed for the analysis of peptides as well as small (low charge state) glycopeptides

Spray Voltage: Positive Ion (V) = 1800
Ion Transfer Tube Temp (°C) = 250
Application Mode = Peptide

Master Scan (MS1)
Orbitrap Resolution = 120K
Scan Range (m/z) = 150-1800
Maximum Injection Time (ms) = 50
AGC Target = 100000
Microscans = 1
RF Lens (%) = 50

Dynamic Exclusion
Exclude after n times = 1
Exclusion duration (s) = 15
Mass Tolerance = ±10 ppm
Exclude isotopes = True

Data Dependent MS2
Isolation Mode = Quadrupole
Isolation Window = 2
ActivationType = HCD
Collision Energy Mode = Stepped
Collision Energies (%) = 20,30,40
Detector Type = Orbitrap
Orbitrap Resolution = 30K
Scan Range (m/z) = 150-1800
Maximum Injection Time (ms) = 150
AGC Target = 50000
Microscans = 1

Method 2. Oxonium Ion triggered HCD EThcD MS/MS instrument method parameters (MS1):

This method is applied to larger glycopeptides containing multiple serine/threonine residues

Spray Voltage: Positive Ion (V) = 1800
Ion Transfer Tube Temp (°C) = 250
Application Mode = Peptide
Orbitrap Resolution = 60K
Scan Range (m/z) = 400-1800
Maximum Injection Time (ms) = 50
AGC Target = 400000
Microscans = 1
RF Lens (%) = 50

Data Dependent HCD MS/MS settings

Dynamic Exclusion
Exclude after n times = 2
Exclusion duration (s) = 20
Mass Tolerance = ±10 ppm
Exclude isotopes = True
Charge State: Include charge state(s) = 2-8

Data Dependent Properties
Precursor Priority: Highest Charge State
Isolation Mode = Quadrupole
Isolation Window = 2 m/z
ActivationType = HCD
Collision Energy Mode = Fixed
Collision Energy (%) = 36
Detector Type = Orbitrap
Orbitrap Resolution = 30K
Scan Range (m/z) = 100-1800
Maximum Injection Time (ms) = 54
AGC Target = 50000

Filter: Product Ion Trigger
Trigger only when at least 2 product ions from list are detected = True
Product ion(s) must be within top n = True
Top N Product Ions = 20
Mass Tolerance = ±15 ppm
Mass list type = m/z

OXONIUM ION MASS LIST TABLE	
COMPOUND NAME	M/Z |
Hex -2H20	126.0550
HexNAc internal frag	138.0549
Hex	163.0601
HexNAc - 2H20	168.0654
HexNAc - H20	186.0760
HexNAc	204.0865
Hex - 18	144.0655
2Hex	325.1120
HexNAcHex	366.1395
Neu5Ac-H20	274.0921
Neu5Ac	292.1027

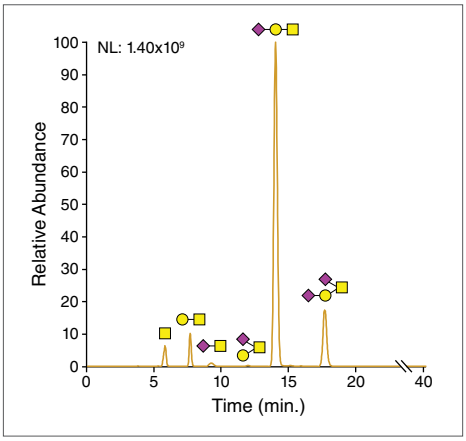
EThcD MS/MS settings

Number of Dependent Scans= 1
Scan ddMSnScan
Desired minimum points across the peak = 6
MSn Level = 2
Isolation Window = 2
Reported Mass = Original Mass
Scan Range Mode = Define m/z range
Scan Priority= 1
ActivationType = ETD
Is EThcD Active = True
Use calibrated charge dependent ETD parameters = True
Supplemental HCD activation info = 25%
Detector Type = Orbitrap
Orbitrap Resolution = 60K
Scan Range (m/z) = 200-4000
Maximum Injection Time (ms) = 400
AGC Target = 100000

APPENDIX B: RELEASED *O*-LINKED GLYCAN PROFILE

The *O*-linked glycan profile of etanercept was generated using reductive beta-elimination release, followed by permethylation and LC-MS/MS (Figure 6). The profile displays core-1 *O*-glycans typical of CHO-expression systems and is consistent with *O*-glycopeptide analysis. Prior to the introduction of O-Glycoprotease (IMPa) a significant unmet need had been the ability to assign these *O*-glycan structures among all potential *O*-glycosylation sites.

FIGURE 6: **Etanercept released *O*-linked glycan profile**



Reductive beta-elimination, permethylation, Reverse phase LC-MS/MS

References

1. Delobel, A. (2021) *Methods Mol Biol.* 2271, 1–21.
2. Bagdonaite, I., et. al. (2022) *Nat Rev Methods Primers* 2, 48.
3. Riley, N.M., et. al. (2020) *Anal Chem.* 92(22), 14878–14884.
4. Riley, N.M., et. al. (2020) *J Proteome Res.* 19(8), 3286–3301.
5. Ince, D., et. al. (2022) *Curr Opin Chem Biol.* 69, 102174.
6. Genovis Operator App Note: *Determination of O-glycan sites by OpeRATOR Digestion and LC-MS.*
7. Vainauskas, S., et. al. (2022) *Anal Chem.* 94(2), 1060–1069.
8. Taron, C., et. al. (2021) *European Pharmaceutical Review.* https://www.europeanpharmaceuticalreview.com/article/157341/o-glycan-analysis-of-therapeutic-proteins-enabled-by-o-glycoprotease.
9. Helms, A., et. al. (2023) *Anal Chem.* 2023 95(24), 9280–9287.
10. Liu, H., et. al. (2022) Sciex App Note: *Comprehensive characterization of O-linked glycosylation in etanercept by electron activated dissociation (EAD).*
11. Montacir, O., et. al. (2018) *Protein J.* 37(2), 164–179.
12. Houel, S., et. al. (2014) *Anal Chem.* 86(1), 576–84.
13. D'Atri, V., et. al. (2019) *Anal Chem.* 91(1), 873–880.
14. Riley, N.M. and Bertozzi, C. (2022) *Molecular Omics.* 18, 908.
15. Lu, L., et. al. (2020) *Nat Methods.* 202017(11), 1133–1138.