Technology

Peptide mapping analysis of etanercept – a highly glycosylated dimeric fusion protein

Peptide mapping including N- and O-linked glycopeptide analysis using the SCIEX TripleTOF® 6600 LC-MS/MS System and Protein Metrics Inc. software

Ryo Yokoyama[1], Sibylle Heidelberger[2], Kerstin Pohl[3]
[1]SCIEX, Japan; [2]SCIEX, UK, [3]SCIEX, USA

In this work, a comprehensive examination of a complex Fc fusion protein using peptide mapping will be presented. A range of high to very low abundant N- and O-linked glycosylations were identified, and the sequence of the protein was confirmed using high-quality data derived from the SCIEX TripleTOF 6600 System. Automated data analysis for rapid and efficient characterization was done using Byos software (Protein Metrics Inc.).

Biotherapeutics are becoming more complex, resulting in increasing challenges for in-depth characterization. Their characterization involves a myriad of analytical methods, which include, but are not limited to, amino acid sequence confirmation as well as identification and localization of post-translational modifications (PTMs).[1] Glycosylations, in particular, are considered critical quality attributes as their chemistry and levels can affect the safety and efficacy of the product. However, glycosylations tend to be present in a highly heterogenous manner in terms of structure and abundancies, increasing the complexity of their analysis.

In the current technical note, different proteolytic digests of etanercept were analyzed, resulting in excellent sequence coverage for a highly complex protein and the identification of various N- and O-linked glycopeptides. High sensitivity, high resolution, accurate MS and MS/MS data acquired on the SCIEX TripleTOF 6600 System was used for confidence in primary sequence coverage, as well as identification and localization of PTMs.

Key features of the SCIEX solution for fusion protein peptide mapping

- Confirmation of sequences alongside with PTM identification and localization (like N- and O-linked glycans) for in-depth characterization from standard mAbs to new modalities leveraging high-quality MS and MS/MS data

- Confidence in correct assignments with automated data-dependent acquisition maintaining data quality (high sensitivity and resolution) independently of high acquisition speeds with the TripleTOF 6600 System

- Direct compatibility of raw data files with state-of-the-art processing software from Protein Metrics Inc., especially suited for complex biotherapeutic characterization
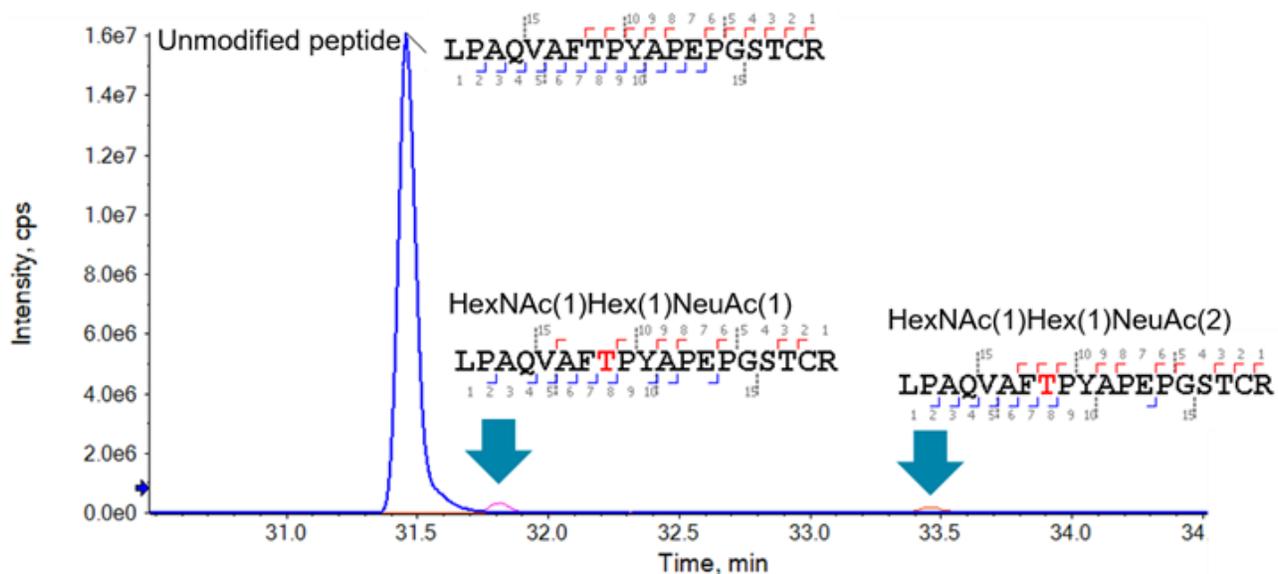

Figure 1. O-linked glycosylations identified on peptide LPAQVAFTPYAPEPGSTCR. The most intense peak of the peptide did not contain any glycosylation. Two low abundant O-linked glycopeptides were identified at later retention times as indicated.

Methods

Sample preparation: To 100 µg protein, 25 µL 8 M guanidine hydrochloride was added before the sample was boiled for 1 minute. Tris-HCl (1 M, pH 8.0, 2 µL) was added to the sample before 1 µL of 0.4 M DTT was added. Samples were incubated for 30 min at 37 °C. Subsequently, 2 µL of 0.4 M iodoacetamide was added to each sample and the samples were incubated for 20 min in darkness at room temperature. Samples were diluted by adding 350 µL of 50 mM Tris-HCl (pH 7.9). For the double digests used in this experiment, 4 µg Asp-N. or 10 µL of 1 µg/µL chymotrypsin was added to the sample. Samples were incubated for 4 hrs at 37°C for Asp-N and 25°C for 16 hours for chymotrypsin. For the secondary digest, 10 µL of 1 µg/µL trypsin/Lys C was added for an additional 16 hrs at 25°C to both samples. Digestion was stopped by the addition of formic acid and the pH was confirmed to be below 3. An injection volume of 20 µL (approximately 5 µg) was applied for analysis.

Chromatography: A total of 20 µL per sample of enzymatically digested etanercept (approximately 5 µg) was injected onto a Waters Acquity CSH C18 column (2.1×150 mm, 1.7 µm particle size, 130 Å) using an ExionLC™ AD System. The aqueous mobile phase (A) consisted of water with 0.1% formic acid while the organic phase (B) consisted of acetonitrile with 0.1 % formic acid. Temperature was set to 40°C on the column and a gradient profile was used (Table 1) with 0.3 mL/min.

Table 1. Chromatographic gradient.

Time	%A	%B
3.0	99	1.0
55	65	35
60	40	60
61	10	90
65	10	90
65.1	99	1.0
70	99	1.0

Mass spectrometry: Mass spectrometry data were acquired in positive ionization mode using the SCIEX TripleTOF 6600 System fitted with an IonDrive™ Turbo V Ion Source. An information dependent acquisition (IDA) approach was used. MS conditions are listed in Table 2 while optimized rolling collision energy values for glycosylated peptides are listed in Table 3.[2]

Table 2. Mass spectrometry parameters.

Parameter	MS	MS/MS
Scan mode	TOF-MS	IDA dependent
Polarity	Positive	
Gas 1	60 psi	
Gas 2	60 psi	
Curtain gas	35 psi	
Source temperature	250°C	
Ion spray voltage	5200 V	
Declustering potential	20 V	
Collision energy	6 V	rolling
CAD gas	4	
Maximum candidate ion	5	
Intensity threshold	125 cps	
Charge states	1 to 6	
Exclusion time	8 s after 1 occurrence	
Start mass	300 m/z	100 m/z
Stop mass	1,800 m/z	1,800 m/z
Accumulation time	0.25 s	0.05 s
Time bins to sum	4	8

Table 3. Rolling collision energy values.[2]

Charge	Slope	Intercept
Unknown	0.037	-1
1	0.038	5
2	0.037	-1
3	0.036	-2
4	0.038	-2
5	0.038	-2

Data processing: Data processing was performed using the PTM workflow in Byos software (Protein Metrics Inc.) which allowed for a combined search across all digests. An extensive list of 52 N-glycans and 70 O-glycans was used during the data processing to identify all glycosylated peptides.

Sequence coverage

Etanercept is a dimeric fusion protein (Figure 2), which consists of the Fc portion of human IgG1 fused together with two extracellular domains of the tumor necrosis factor receptor 2 (TNFR2). This Fc fusion protein has molecular weight of approximately 100 kDa without any modifications being taken into account. However, the protein furthermore contains the complexity of the Fc IgG1 region with two N-linked glycosylation sites, as well as 4 additional N-glycosylated sites located in the TNFR subunit. Up to 14 O-glycosylated sites have been identified resulting in a final modified form of approximately 130 kDa in molecular weight.[1,3,4] The primary function of etanercept is as an inhibitor of TNFα, which plays an important role in inflammatory cell responses.

Double digestions of trypsin-Asp-N and trypsin-chymotrypsin were performed to provide complete sequence coverage and detailed information on the glycan profile (Figure 3 and 4). The sequence coverage of etanercept based on MS/MS data was 87.15% for trypsin-Asp-N and 82.44% for trypsin-chymotrypsin respectively, resulting in 92.72% coverage in total. Double digest
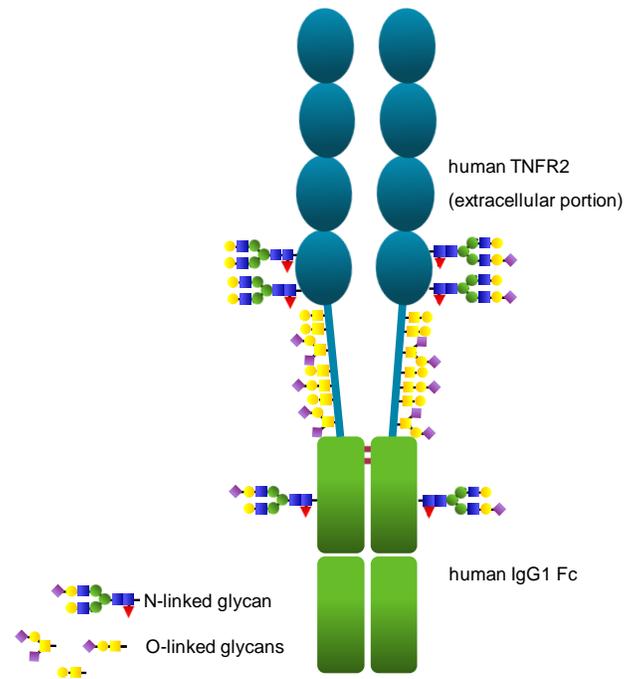

Figure 2. Schematic of the fusion protein etanercept.
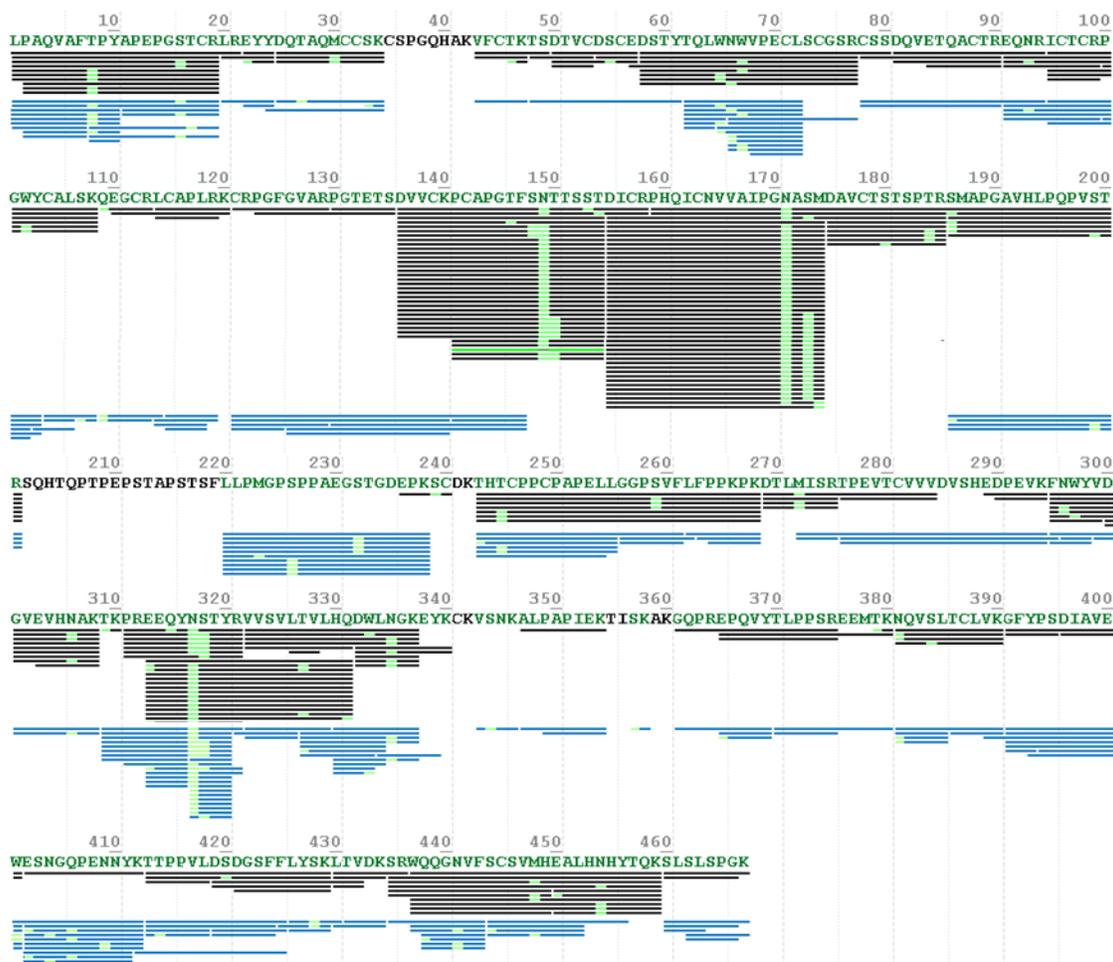

Figure 3. Sequence coverage of etanercept using TOF-MS and MS/MS data. Digest using trypsin/Asp-N digest (black) resulted in 87.15% and trypsin/chymotrypsin (blue) in 82.4% sequence coverage. Both samples in combination resulted in 92.72% sequence coverage. N- and O-glycosylation sites are highlighted in light green.

p 3

Sample name	Protein name	Coverage summary	Coverage percent%
Currently checked	Etanercept	433 of 467	92.72
201016_Tryp_AspN_Etanercept_CES5_zrCE_5ug_1	Etanercept	407 of 467	87.15
201016_Etanercept_Tryp_Chymo_CES5_zrCE_5ug_1	Etanercept	385 of 467	82.44

Figure 4. Sequence coverage of etanercept. Digest using trypsin/Asp-N digest resulted in 87.15% and trypsin/chymotrypsin in 82.4 % sequence coverage. Both samples in combination resulted in 92.72 % sequence coverage with MS/MS.

combinations were used to reduce the complexity of the data, allowing for smaller peptides carrying glycans to be observed and improvement of the localization of the glycans.

N-linked glycosylations

Etanercept has been confirmed to have 6 N-linked glycosylation sites.[1-3] The N-linked glycosylations sites are 2 sites along the TNFR region at positions 149 and 171 and the Fc glycosylation site within EEQYNSTYR peptide at position 317 (Figure 2 and 6). All N-linked glycosylation sites were identified and the glycosylations found to be present were in alignment with those previously reported (Figure 6). Accurate and detailed identification of glycopeptides was performed using the Byos software (Protein Metrics Inc.). MS data were matched with a tolerance of 5 ppm. Subsequent data interpretation of MS/MS spectra included the identification of glycan fragments, oxonium ions and peptide as well as peptide-glycan fragments (Figure 5). For an easy review of the data, the report within Byos software was filtered to show only information on N-linked glycans and

their relative levels (Figure 6). All of the glycopeptides shown were confirmed with MS/MS data. For all three major N-linked glycosylations sites at the amino acid positions 149, 171 and 317 multiple glycans per site were identified. Their levels reached as high as 66% and as low as 0.8%. The automated color coding facilitates a quick understanding of which glycoforms are present in high, medium or low abundance.
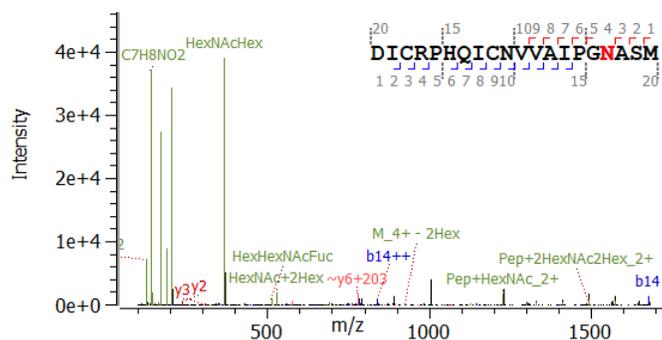

Figure 5: Identification of N-linked glycosylation site on peptide DICRPHQICNVVAIPGNASM with G2F (HexNAc(4)Hex(5)Fuc(1)) glycan.

Sequence (unformatted) ↑	GlycanShortName ↑	Mod. AAs ↑	Var. Pos. Protein ↑	MS Alias name ←	201016_Tryp_AspN_Etanercept_CES5_zrCE_5ug_1 (%)
DICRPHQICNVVAIPGNASM	G0F	N	171		1.79
	G1F	N	171		5.7
	G2F	N	171		8.09
	G2FS1	N	171		55.2
	G2FS2	N	171		26.5
	Man5	N	171		2.71
DVVCKPCAPGTFSNTTSST		N	149		66.1
	G2F	N	149		3.61
	G2FS1	N	149		17.4
	G2FS2	N	149		7.88
	Man5	N	149		5.02
EEQYNSTYR	G0	N	317		0.795
	G0F	N	317		66.3
	G1F	N	317		29.4
	G2F	N	317		3.55

Figure 6. Identification of N-linked glycosylations in etanercept on three different positions with trypsin/Asp-N digest.

O-linked glycosylations

While obtaining MS/MS data on O-linked glycosylations seems relatively straightforward, the complexity of their characterization lies in the fact that there may not be consensus sequences required for these glycosylation events to occur. In addition, they can be present on S, T or Y residues. Any peptide with a mixture of these residues therefore has the potential for multiple sites of glycosylation. Furthermore, there is a great heterogeneity of glycan structures known to occur for O-glycans, which in sum leads to an even more complex mixture compared to N-linked glycans. Here, 11 sites have been successfully identified, most with a mixture of glycans associated with the location and all identified with good quality MS/MS data, allowing for confirmation of both glycan and site (Figure 7). Of those identified, some glycans were at levels below 1% in relative abundance. Figure 8 shows the tabular results from peptide LLPMGPSPPAEGSTGDEPK, which has two serine-linked glycosylations sites and up to four different glycans upon those
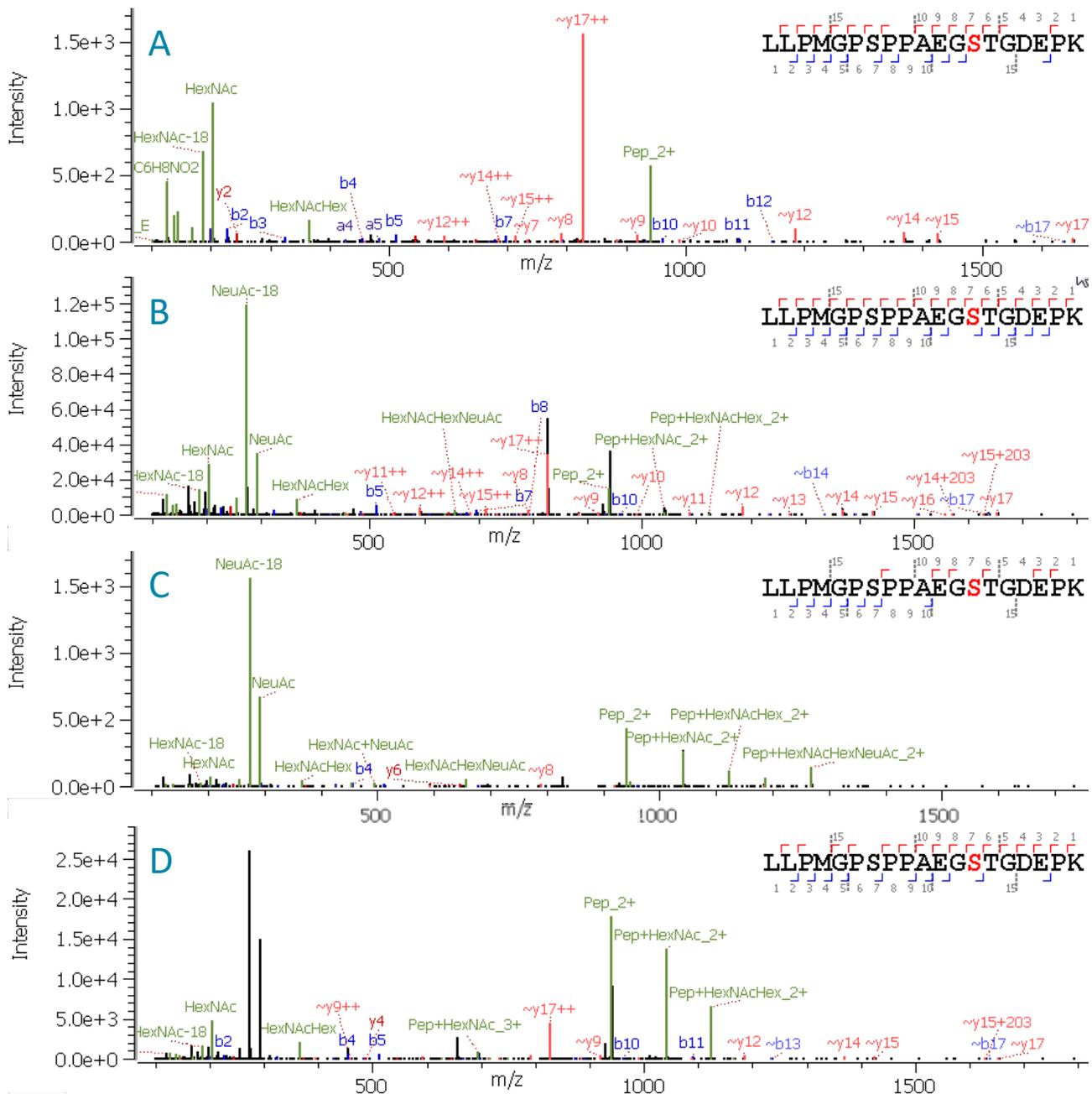

Figure 7. O-linked modifications on serine in peptide LLPMGPSPPAEGSTGDEPK. A. HexNac(1)Hex(1), B. HexNAc(1)Hex(1)NeuAc(1) C. HexNAc(1) Hex(1)NeuAc(2) D. HexNAc(1)Hex(2)Fuc(1).

sites. From this peptide, glycans and sites of modification have been identified down to a level of 0.955% glycosylation with site of modification confirmed with MS/MS data. An example of the data from the glycosylation at position S232 has been given in Figure 7, showing the clear indication of location for each glycan.

| | | | | MS Id ← | 1 |
| | | | | MS Alias name ← | 201016_TyrCymo_Etanercept_CES5_zrCE_5ug_1 |
Sequence (unformatted) ↑	Mod. Names ↑	Mod. AAs ↑	Var. Pos. Protein ↑		(%)
LLPMGPSPPAEGSTGDEPK	OGlycan/365.1322	S	226		1.5
			232		0.955
	OGlycan/656.2276	S	226		14.6
			232		71
	OGlycan/673.2429	S	232		9.72
	OGlycan/947.3230	S	232		2.28

Figure 8. Example of peptide LLPMGPSPPAEGSTGDEPK with O-linked glycans at position 226 and 232.

Conclusions

- A straightforward characterization of the highly complex fusion protein etanercept using an automatized data-dependent acquisition strategy for peptide mapping on the SCIEX TripleTOF 6600 System and comprehensive data analysis with Protein Metrics Inc. software was demonstrated

- Using multiple digestion strategies, the sequence coverage of the glycoprotein was greater than 92%; confirmed by excellent MS/MS data

- All N-linked glycosylation sites were identified and several different glycoforms could be confirmed using high quality MS and MS/MS data

- Multiple O-linked glycosylations had both their sites and type of glycosylation identified and confirmed with MS/MS data even for peptides with multiple glycosylation sites down to below 1% relative abundance

References

1. Wohlschlager, T.; Scheffler, K.; Forstenlehner, I.C. *et al.* (2018) Native mass spectrometry combined with enzymatic dissection unravels glycoform heterogeneity of biopharmaceuticals. *Nat Commun* **9:** 1713.

2. Comprehensive analysis of low abundant mannose glycopeptides in peptide mapping of adalimumab. SCIEX technical note RUO-MKT-02-11834-A.

3. Fazel, R.; Guan, Y.; Vaziri, B. *et al.* (2019) Structural and in vitro functional comparability analysis of altebrel, a proposed etanercept biosimilar: focus on primary sequence and glycosylation. *Pharmaceuticals,* **12**(1): 14.

4. Cho, I.H.; Lee, N.; Song, D. *et al.* (2016) Evaluation of the structural, physicochemical, and biological characteristics of SB4, a biosimilar of etanercept, *mAbs,* **8(**6): 1136-1155.

Headquarters
500 Old Connecticut Path | Framingham, MA 01701 USA
Phone 508-383-7700
sciex.com

International Sales
For our office locations please call the division headquarters or refer to our website at
sciex.com/offices