Comprehensive characterization of etanercept glycosylation by subunit LC-MS analysis

Featuring the ZenoTOF 7600 system and Biologics Explorer software from SCIEX

Haichuan Liu[1], Xuezhi Bi[2],Zoe Zhang[1]
[1]SCIEX, USA; [2]BTI, Singapore

This technical note describes an in-depth characterization of N- and O-linked glycosylations of the complex fusion protein etanercept by subunit LC-MS analysis. The methodology described can be applied broadly to the characterization of new therapeutic proteins with increased complexity.

The characterization of modern therapeutic proteins, such as fusion proteins and multi-specific antibodies, can pose analytical challenges due to increasingly complex glycoform distributions. Intact protein analysis, routinely performed at different stages of monoclonal antibody (mAb) development, is essential for confirming identity and monitoring product quality. However, it is challenging to obtain detailed information about glycosylation on complex therapeutic proteins. The presence of sialic acids and O-glycosylation in these molecules significantly increases the complexity of the mass spectrum, making data interpretation and accurate peak assignment a daunting task.

An LC-MS workflow to characterize N- and O-glycosylation for etanercept, on the subunit level, using state-of-the-art hardware and software was developed. The characterization of desialyated N- and O-glycosylation, separately or simultaneously, in shorter sequences, enabled more confident peak assignment and better understanding of the location of the glycosylation.

Key features of the SCIEX solution for analysis of complex therapeutics

- Excellent data quality, mass accuracy, and sensitivity for intact and subunit analysis; facilitating in-depth characterization of therapeutic proteins
- Easy-to-use hardware and software accessible to experienced as well as walk-up LC-MS users
- Highly flexible instrumentation accommodates multiple flow-regimes and biopharma workflow coverage
- Intuitive data analysis software with optimized workflow templates and powerful visualization tools

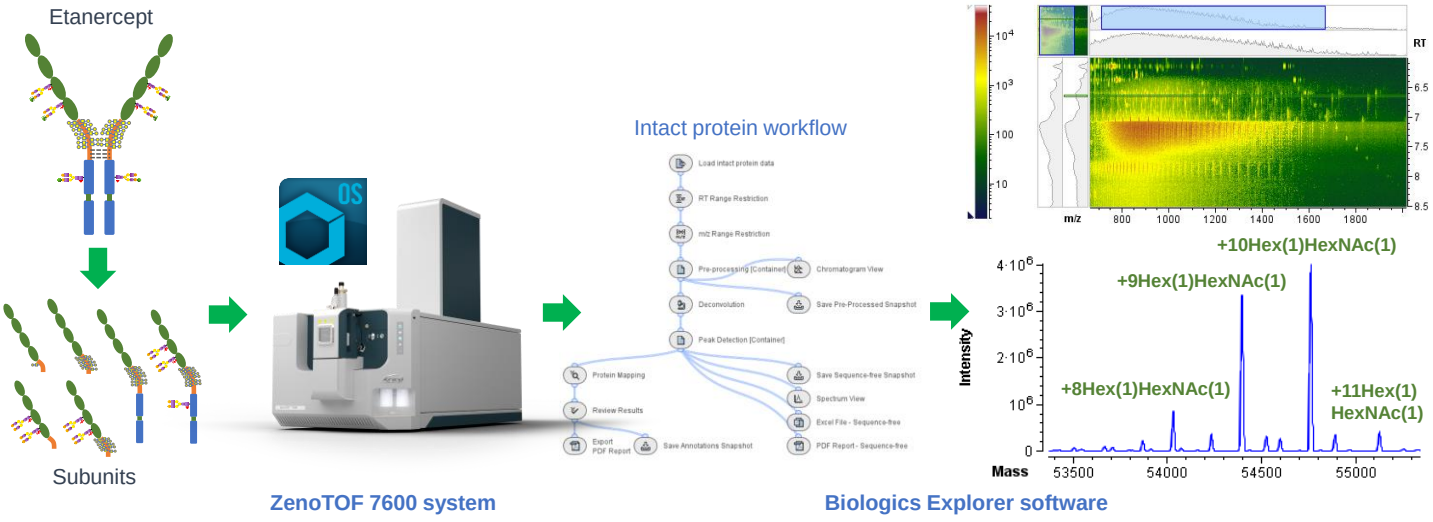

Figure 1. Subunit LC-MS analysis workflow of etanercept using the ZenoTOF 7600 system and Biologics Explorer software. The ZenoTOF 7600 system offers high-quality MS data and excellent mass accuracy for intact mass analysis. The Biologics Explorer software provides an intuitive intact protein workflow template optimized for SCIEX QTOF data and powerful visualization tools for reviewing the data and results.

Methods

Sample preparation: Figure 2 shows a schematic of the sample preparation for this study. Briefly, 250 µg of etanercept were incubated with 40 units/µL of FabALACTICA (IgdE from Genovis) at 37°C overnight. One half of the resulting solution was treated with SialEXO (Genovis) at 37°C for 4 hours for removal of sialic acid (SA), while the other half was incubated with SialEXO and OglyZOR (Genovis) under the same condition to remove SA and O-glycans, simultaneously. Subsequently, one part of these two samples was incubated with N-glycanase (PNGase F from Agilent Technologies) at 37°C overnight to remove N-glycans. All the samples (four in total) were further treated with 7M guanidine-HCl in 50 mM Tris-HCl and dithiothreitol (DTT) to reduce disulfide bonds. The final solutions contained ~60 µg of etanercept subunits (~0.5 µg/µL). 2-4 µL (1-2 µg) of etanercept subunits were injected for LC-MS analysis.

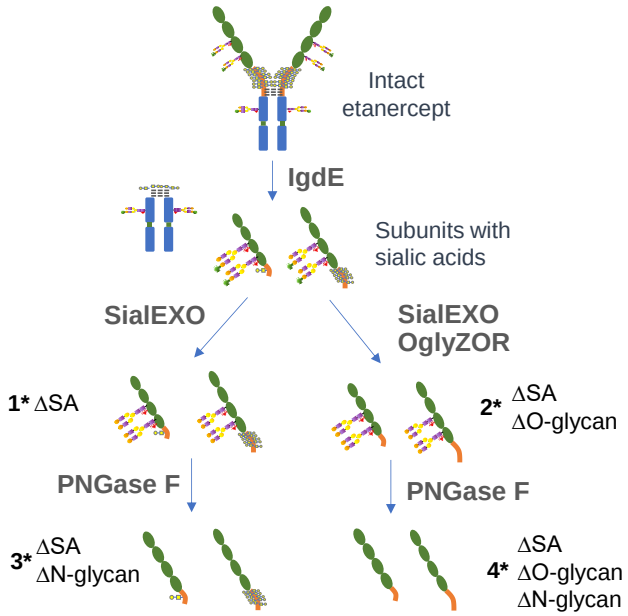

Figure 2. Sample preparation scheme for subunit LC-MS analysis of etanercept. Schematic is limited to species of interest, while other species were observed**.** Δ indicates the losses of specific glycan moieties by treatment with enzymes as shown. * indicates samples were treated with reducing agent prior to measurement. Sample 1-4 were analyzed by LC-MS.

Chromatography: Etanercept subunits were separated using an ExionLC system installed with an ACQUITY UPLC Protein BEH C4 column (300 Å, 1.7 µm, 2.1 x 50 mm, Waters), which was kept at 60°C. Chromatographic separation was performed at a flow rate of 0.3 mL/min using an 8-minute linear gradient (Table 1). Mobile phases A and B consisted of 0.1% formic acid in water and 0.1% FA in acetonitrile, respectively.

Mass spectrometry: TOF-MS data were acquired using a SCIEX ZenoTOF 7600 system with intact protein mode enabled. Key MS settings used in this study are listed in Table 2.

Table 1. LC gradient for separation of etanercept subunits.

Time [min]	A [%]	B [%]
Initial	*80**	*20 **
2	*80*	*20**
10	*65*	*35*
10.5	*10*	*90*
11.5	*10*	*90*
12	*80*	*20**
17	*80**	*20**

**75% A and 25% B were used as the start and end LC conditions for sample 4 without glycans (Figure 2).*

Table 2. MS parameters

Parameter	Values
Scan mode	*TOF MS*
Intact protein mode	*ON*
Polarity	*Positive*
Ion source gas 1	*50 psi*
Ion source gas 2	*50 psi*
Curtain gas	*35 psi*
CAD gas	*5*
Source temperature	*400 ºC*
Spray voltage	*5500 V*
TOF start mass	*700 m/z*
TOF stop mass	*2500 m/z*
Accumulation time	*0.5 s*
Declustering potential	*150 V*
Collision energy	*10 V*
Time bins to sum	*40*

Data processing: All the data were analyzed using the Intact Protein workflow template in the new Biologics Explorer software from SCIEX.[1] The default processing parameters optimized for SCIEX TOF MS data were used except for the following settings that required manual input (Table 3).

Table 3. Biologics Explorer parameters

Parameters	Values
RT range restriction	*Sample dependent (4 min – 9 min)*
m/z range restriction	*600-2500 m/z*
Deconvolution	*Min. mass = 10 kDa* *Max. mass = 70 kDa* *Mass step: 0.5 Da* *RT ranges = manual*
Protein mapping	*Sequence* *Sequences: etanercepts subunits were entered from text* *Mass tolerance: 50 ppm* *Modifications* *samples 2 and 4: None for de-O-glycan Hex(1)HexNAc(1) ST as variable modification for O-glycan samples 1 and 3 (Maximum: 13 per sequence)* *Glycosylation* *Deglycosylation for samples 3 and 4 (Allowed sites: Only N-linked)* *glycosylated for samples 1 and 2 (Library: CHO N-glycans small; allowed sites: only N-linked)* *Disulfide: fully reduced*

Overview of subunit LC-MS workflow

Etanercept is a fusion protein consisting of two tumor necrosis factor receptor (TNFR)-Fc chains with six N-glycosylation and 26 O-glycosylation sites that are partially sialylated (Figure 3).

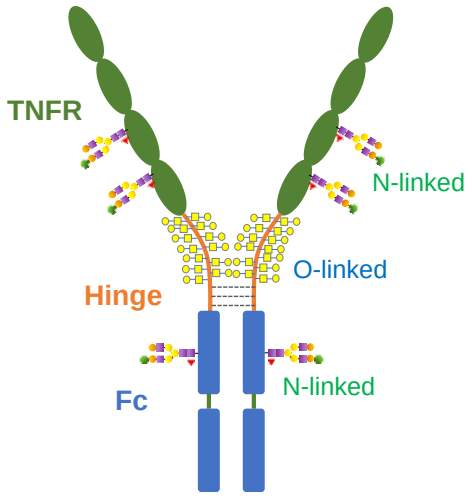

Figure 3. Illustration of etanercept structure. Etanercept is a dimeric fusion protein composed of two TNFR, each of which is fused with an Fc domain of an IgG1 antibody. Each TNFR-Fc chain contains three N- and 13 O-glycosylation sites. Two chains are connected through 3 inter-disulfide bonds.

Structural elucidation of fully intact etanercept is challenging due to complexity and heterogeneity induced by O-glycosylation and sialylation. To gain insight into the structure of etanercept, the mAb was cleaved into subunits by IgdE followed by removal of sialyation using SialEXO and reduction with DTT (Figure 2), which reduced sample heterogeneity and data complexity. The monomeric TNFR, Fc, and TNFR-Fc subunits with or without N- and O-glycans were analyzed using the high-performance ZenoTOF 7600 system. While the data on subunits without glycosylation provided information on the integrity of IgdE fragments, the results from subunits with N- and/or O-glycans enhanced the understanding of the glycosylations present in the etanercept sample used.

The high-quality LC-MS data were processed using the intuitive Biologics Explorer software, which provides powerful visualization tools for data review (Figure 4), optimized workflow templates, and proven algorithms for reliable spectrum processing and deconvolution.

Aglycosylated etanercept subunits

The main species detected from the fully aglycosylated and reduced etanercept (sample 4 in Figure 2) were monomeric TNFR-Fc without C-terminal lysine (Lys, 51.1 kDa) and subunits TNFR1-187 (20.3 kDa) and TNFR1-223 (24.1 kDa) corresponding to cleavages at the C-terminus of M^{187} and M^{223}, respectively (Figure 4), above the hinge region by IgdE.

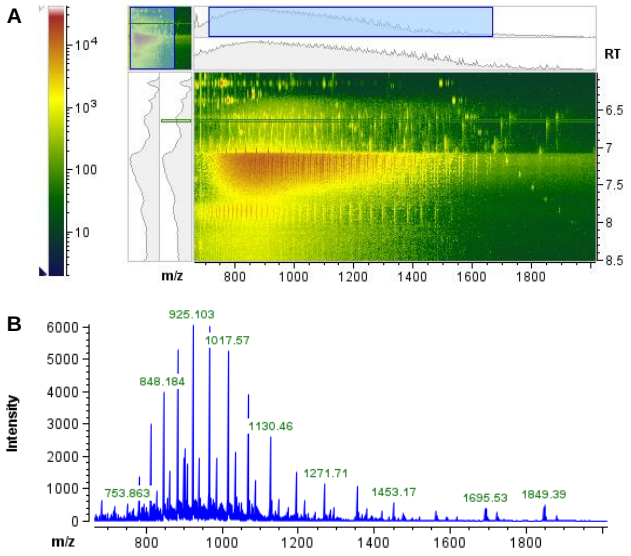

Figure 4. Ion map (A) and mass spectrum (B) viewed in the Biologics Explorer software. The ion map provides a multi-dimensional view of the data at any RT or m/z. As an example, the mass spectrum at RT = 6.63 min was displayed here.

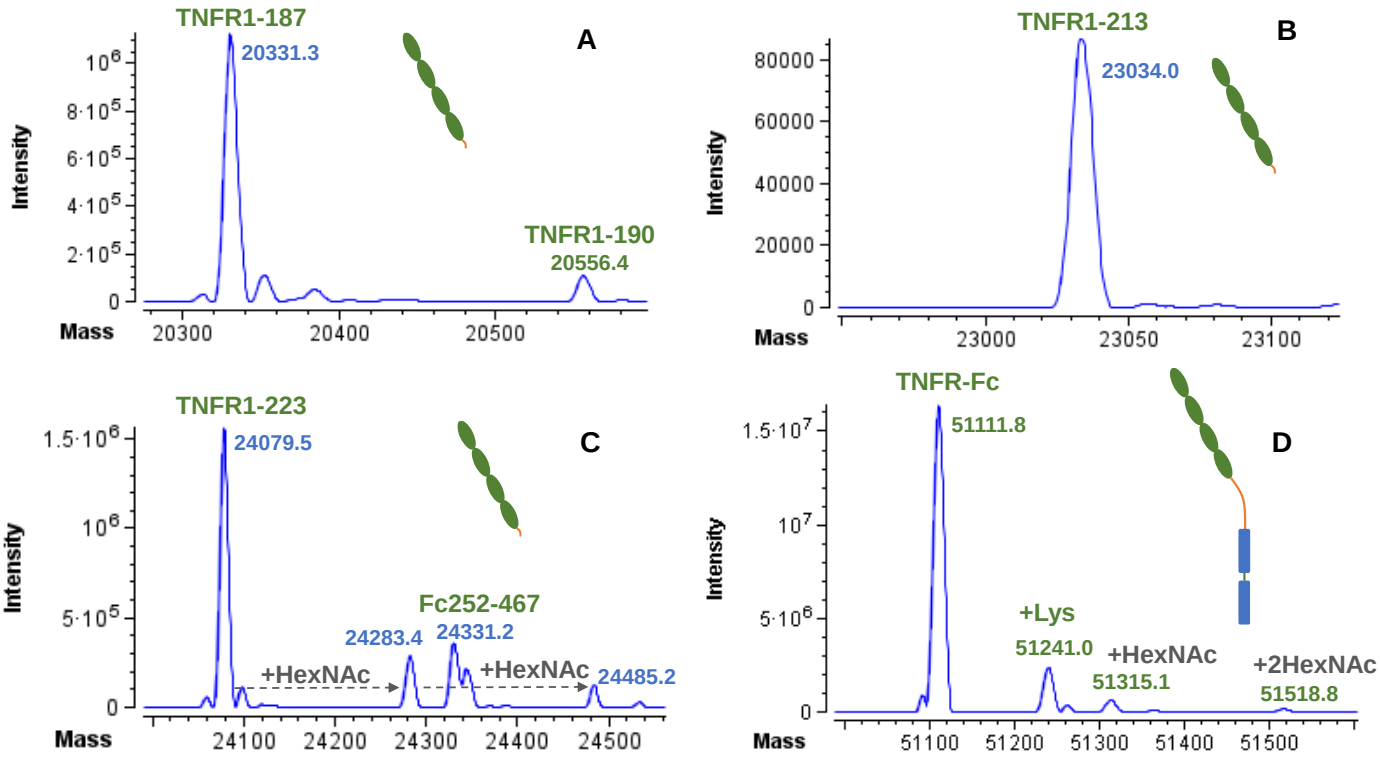

Figure 5. Deconvolution results of deglycosylated etanercept. Multiple TNFR subunits, including 1-187, 1-190, 1-213, and 1-223 cleavage products of IgdE, and one Fc subunit (253-467) were detected in this sample. Among these, the main species were found at 20.3 kDa (identified as TNFR1-187) and 24.1 kDa (identified as TNFR1-223). Another major species observed in the sample was the reduced TNFR-Fc without C-terminal Lys.

Additional subunits from TNFR (1-190 and 1-213) and Fc (253-467) were present at much lower abundance (Figure 5). Interestingly, the species with the addition of up to two HexNAc were observed for TNFR1-223 and TNFR-Fc, but not for TNFR1-187. This could be attributed to differences in structure and susceptibility to PNGase F between TNFR1-187 and its longer counterparts.

The co-existence of TNFR and TNFR-Fc subunits in the same sample allowed for characterization of N- and O-glycans in different regions of etanercept, as described below.

Glycosylated TNFR subunits

Figures 6 and 7 show aglycosylated and glycosylated forms of TNFR1-187 and TNFR1-223 observed in four samples. The major N-glycan forms observed for the two subunits are G2+G2F and 2G2F (Figure 6B and 7B), confirming the occupation of the two available N-glycosylation sites in the TNFR domain. The minor N-glycoforms with two Man5 were also detected (Figure 6B and 7B). Interestingly, drastic differences were observed between TNFR1-187 and TNFR1-223 in the samples without sialic acid and N-glycans (Figure 6C and 7C). While one O-glycan (core 1) was observed for TNFR1-187 (Figure 6C), up to nine O-glycan moieties were detected for TNFR1-223 (Figure 7C). These results indicate that most O-glycosylation sites are located in the region of TFNR187-223 containing ten Ser/Tyr. The species containing both N- and O-glycans have been observed in the samples treated only with sialidase (Figure 6D and 7D). The dominant forms are G2+G2F plus one O-glycan for TNFR1-187 (Figure 6D) and eight or nine O-glycans for TNFR1-223 (Figure 7D). Taken together, these results do not only confirm the dominant N-glycan forms, but also pinpoint the region of O-glcosylation in etanercept.

Glycosylated TNFR-Fc monomers

The aglycosylated or glycosylated TNFR-Fc monomers without C-terminal Lys were observed as one of the dominant species in all samples due to incomplete cleavage by IgdE. The presence of TNFR-Fc monomers, however, allowed for the characterization of N- and O-glycosylations simultaneously (Figure 8), providing complementary results to those described above for TNFR fragments (Figure 6 and 7).

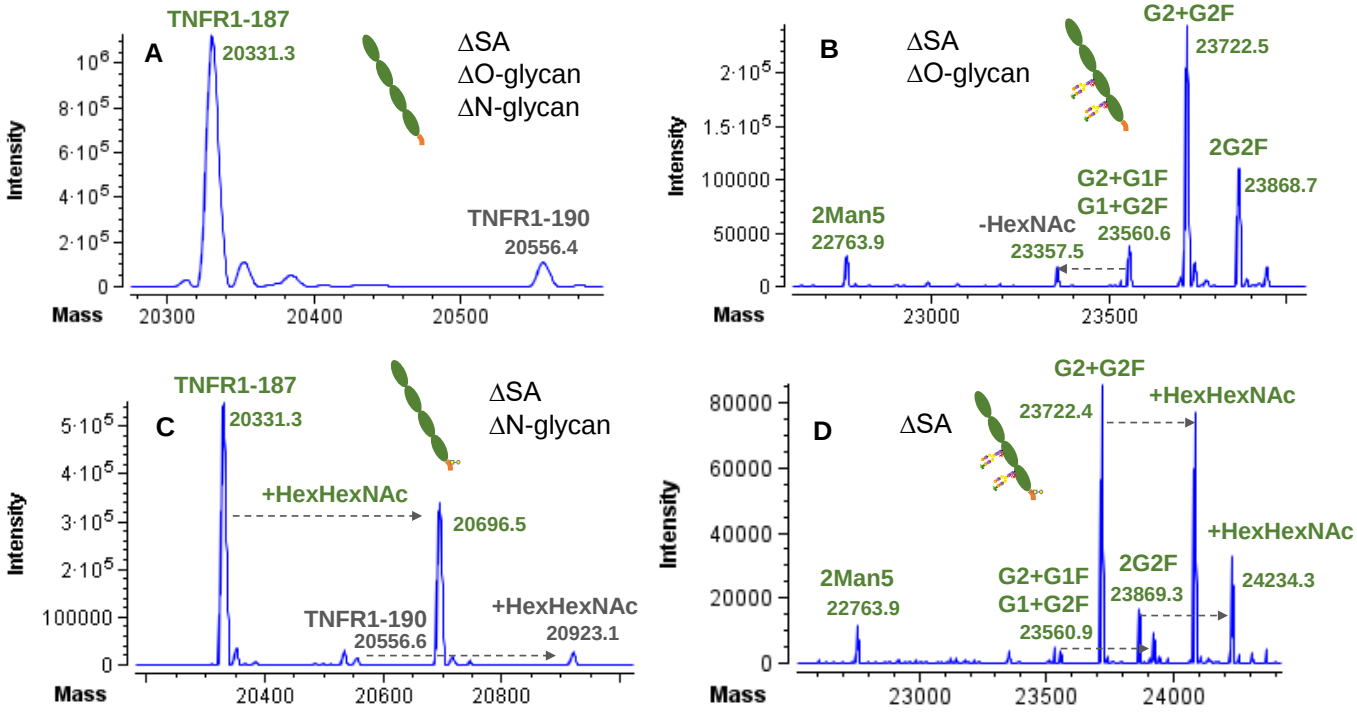

Figure 6. TNFR1-187 with (B-D) or without (A) glycosylation. SA: sialic acid. Δ represents loss of glycan moiety. The dominant N-glycan forms observed for TNFR1-187 are G2+G2F and 2G2F (B and D). Only one O-glycan (core 1) was detected in TNFR1-187 (C and D).

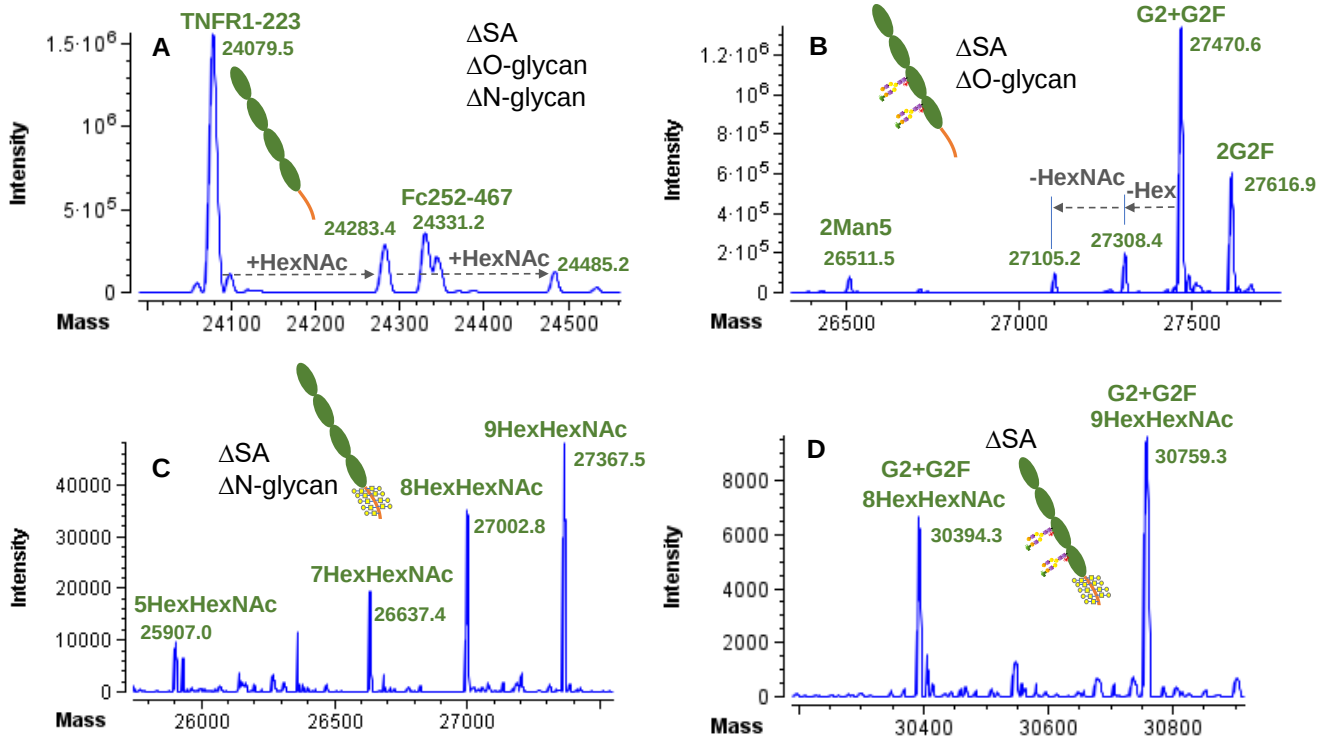

Figure 7. TNFR1-223 with (B-D) or without (A) glycosylation. As the case for TNFR1-187 (Figure 6), TNRF1-223 was found to carry G2+G2F and 2G2F as the major N-glycan forms (B and D). However, up to 9 O-glycans (core 1) were detected for TNFR1-223 (C and D), as compared to only one observed in TNFR1-187.

The monomeric TNFR-Fc carried one additional N-linked G0F or G1F in the Fc domain (Figure 8B) compared to TNFR1-187 and TNFR1-223 (Figure 6B and 7B). The dominant O-glycans observed for TNFR-Fc contain 9 and 10 HexHexNAc structures (Figure 8C and 8D). Compared to the results from TNFR1-223, 1-2 more O-glycans were observed for TNFR-Fc, indicating additional O-glycosylation site(s) close to or in the hinge region.

In summary, the subunit data of etanercept revealed that this therapeutic protein carries G2F+G2 or 2G2F as the dominant N-glycan forms in the TNFR domain and G0F or G1F in the Fc domain. One O-glycan was detected in the sequence of TNFR1-187, with the majority of O-glycans (8 out of 13) located between amino acids 187 and 223 above the hinge region (Figure 9).

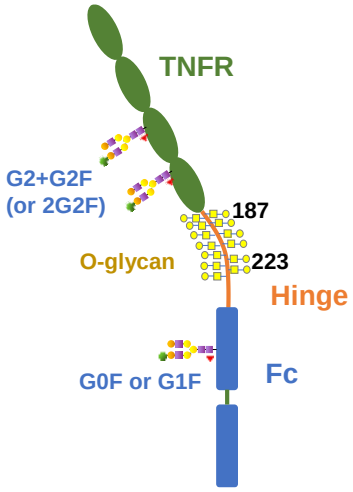

Figure 9. Illustration of the main N- and O-glycoforms in etanercept revealed from subunit LC-MS analysis. The main N-glycans observed include G2+G2F or 2G2F in the TNFR domain and G0F or G1F in the Fc domain. The majority of O-glycans (eight) were found in the region between two Met residues at 187 and 223 above the hinge region.

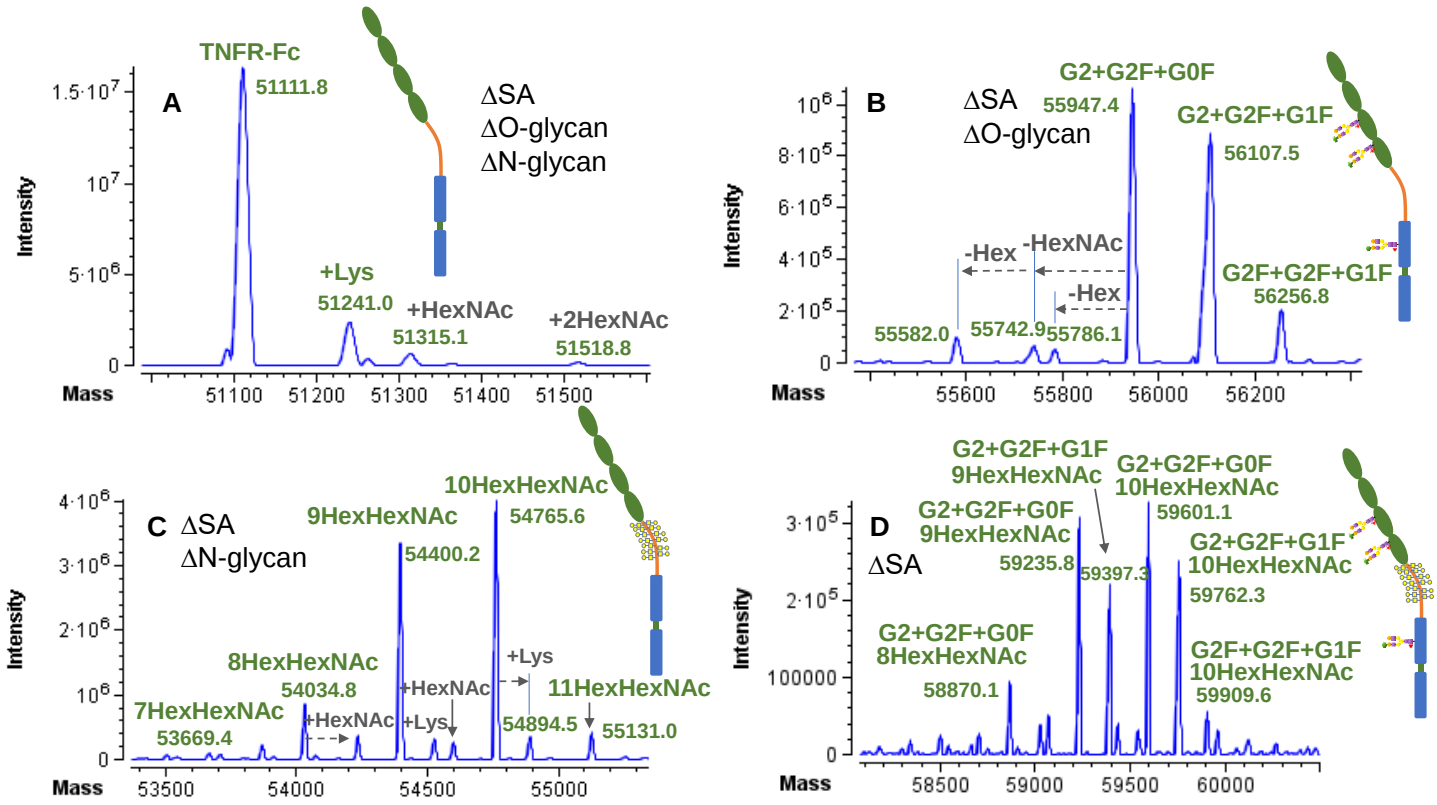

Figure 8. TNFR-Fc without (A) or with (B-D) glycosylation. The main TNFR-Fc monomer observed contains no Lys at C-terminus, as expected (A). The major N-glycan forms are G2+G2F for the TNFR domain and G0F or G1F for the Fc domain (B and D). The addition of up to 11 O-glycans (core 1) was observed for TNFR-Fc, with 9 and 10 O-glycan being the most dominant forms.

Conclusions

- A subunit LC-MS workflow leveraging of state-of-the-art hardware and software was developed to meet the challenges of characterizing complex N- and O-glycosylations in a therapeutic fusion protein.
- The high-quality MS raw data and derived deconvolution data provided insight into glycosylation types and forms in a complex therapeutic protein.
- The streamlined data acquisition and data analysis make daunting characterization tasks readily achievable in biopharmaceutical settings.

References

1. Intuitive software workflow for intact and subunit analysis of monoclonal antibodies (mAbs). SCIEX technical note, RUO-MKT-02-13505-A.

SCIEX

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