eurofins | EAG Laboratories

WHITE PAPER

LC-MS Characterization of Therapeutic Proteins

INTRODUCTION

Proteins developed for therapeutic treatment or diagnostic testing are large, complex molecules that require specialized data analysis techniques. These analyses can provide a protein's basic sequence and can be used to indicate post translational modifications (PTM's), internal disulfide linkages, overall protein conformation, and binding activity to other molecules. The information presented in this paper will focus on four analytical techniques typically employed for the analysis of monoclonal antibodies (mAb's).

1) Intact Analysis, including glycoform differentiation
2) Peptide Mapping
3) Internal Disulfide Linkages
4) Glycopeptide Analysis

INTACT ANALYSIS

Most protein samples are not completely homologous and the protein itself may consist of multiple proteoforms. Performing an intact analysis of an ~150kD protein, with multiple populations, requires a high-resolution MS detector and a software package capable of differentiating the proteoforms without artifacts.

Antibody proteins routinely have multiple glycosylated forms, of varying population. Eurofins EAG Materials Science analyzed a denatured (non-native) confirmation of a monoclonal antibody (mAb) sample with multiple glycoforms, which are protein variants that differ with respect to number or types of attached glycans. Intact samples were analyzed using a high-resolution LC/MS system, a Thermo Q-Exactive. The data was processed utilizing Byos® software from Protein Metrics. The data as seen in Figure 1, appears as a complex spectrum. This is due to the multiple charge states of a large protein and the multiple glycoforms. Even with this complexity, a high-resolution detector is able to resolve the multiple isotopes, allowing for the deconvolution and identification of the glycoform populations. Spectral data is highly complex due to the large size, multiple charge states and multiple glycoforms present. High resolution MS and advanced software are critical for performing analyses of complex molecular samples such as this.

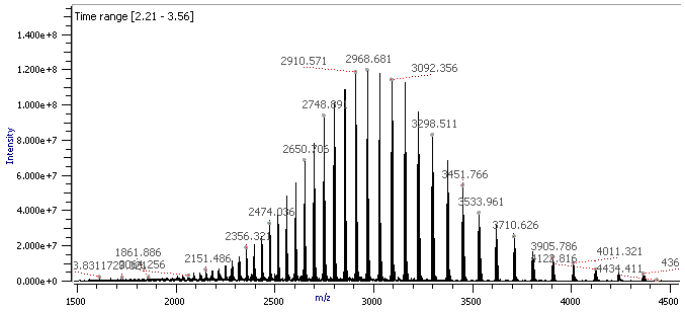

Figure 1. Full MS1 spectra of a mAb sample.

A zoomed view of the spectrum can be seen in Figure 2. Isotopic resolution of individual charge states is shown even above m/z 3000, with nearly baseline resolution. This allows for the positive identification of the glycoforms present.

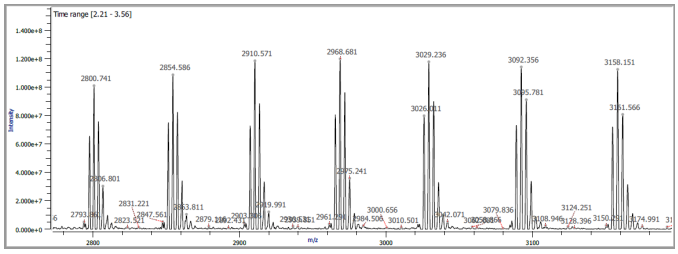

Figure 2. Zoomed view of high-resolution data of a mAb sample

High resolution data, along with advanced deconvolution software from Protein Metrics ensures accurate, high confidence identification of the populations present in this complex sample. Due to the high mass accuracy of the instrument and software package, mass errors are low for ~150kD protein. This allows for confident identification of the glycoforms present as seen in Figure 3. The expected and measured glycoform masses are compared in Table 1.

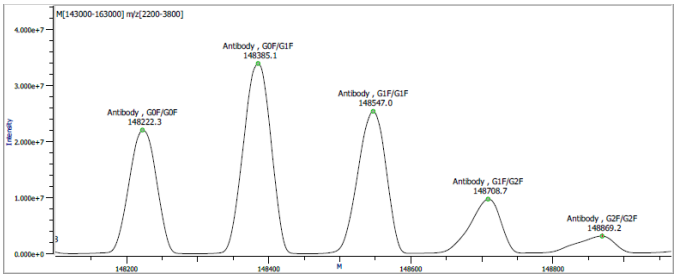

Figure 3. Identification of the five major glycoforms of the mAb.

The relative populations of the glycoforms present in the mAb can be determined from the integrated intensities of the deconvoluted spectra as shown in Figure 4.

LC-MS Characterization of Therapeutic Proteins

Table 1. Mass spectral data from the deconvolution of the mAb sample.

Name	Expected Mass	Detected Mass	Delta Mass	Mass Error	Peak Intensity
mAb, G0F/G0F	148221.25	148222.28	1.03	6.94902E-06	2.21E+07
mAb, G0F/G1F	148383.45	148385.05	1.6	1.07828E-05	3.39E+07
mAb, G1F/G1F	148545.55	148547.01	1.45	9.76122E-06	2.54E+07
mAb, G1F/G2F	148707.75	148708.65	0.9	6.0521E-06	9.77E+06
mAb, G2F/G2F	148869.85	148869.15	-0.7	-4.7021E-06	3.18E+06

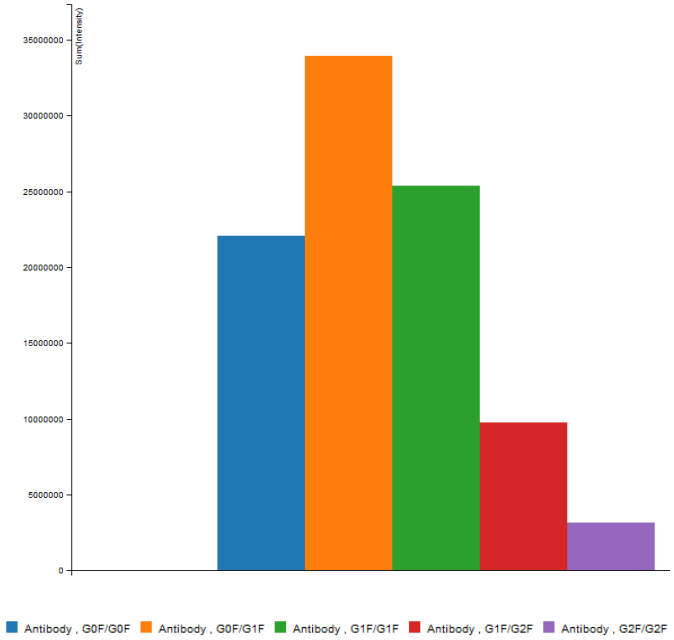

Figure 4. Relative populations of the glycoforms

PEPTIDE MAPPING

The most basic structural component of any protein is the amino acid sequence that makes up the primary structure. A peptide map can be established by analysis of the constituent amino acid sequence of a protein. Being able to map peptide sequences derived from the protein allows for confirmation of the sequence and identification of the protein. Protease digestion, mass spectral (MS) analysis and data processing allow for the identification of individual peptide sequences and matching them to the sequence of the protein. As with intact analysis, high mass accuracy, high resolution mass spectrometers allow for a confident identification of the peptides analyzed. Higher energy collisional dissociation (HCD) utilized for this analysis typically ensures fragmentation that enables nearly full sequence coverage of all amino acids, even from larger peptides.

The mAb sample was subjected to a protease digest. The digested material was analyzed using a high-resolution LC/MS system, a Thermo Q-Exactive to determine peptide sequence. The mass spectra of two of the peptides identified from the protease digestion can be found in Figure 5 and Figure 7.

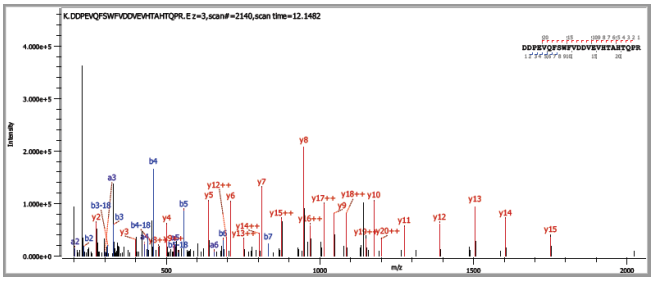

Figure 5. Spectra of peptide identified from the protease digestion of a mAb (Peptide sequence DDPEVQFSWFVDDVEVHTAHTQPR)

Of the 24 amino acids present in the first peptide sequence DDPEVQFSWFVDDVEVHTAHTQPR only the N and C terminus amino acids were not sequenced. Both the MS1 and MS2 ions detected during the analysis contained less than a 10.0pm error from the identified peptide sequence. The MS2 mass error can be found in Figure 6.

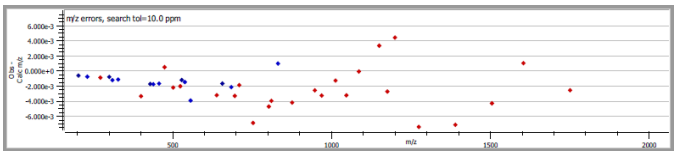

Figure 6. Mass errors of the MS2 ions detected from the peptide identified

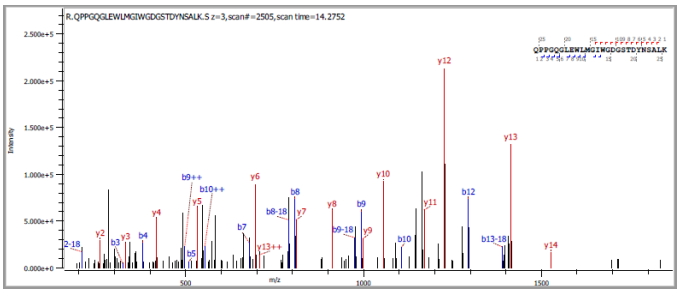

Figure 7. Spectra of peptide identified from the protease digestion of a mAb (Peptide sequence QPPGQGLEWLMGIWGDGSTDYNSALK)

Spectra of the peptide identified from the protease digestion of a different sequence can be found in Figure 7. Of the 26 amino acids present in the peptide sequence

QPPGQGLEWLMGIWGDGSTDYNSALK only four amino acids were not specifically covered in the spectra. Both the MS1 and MS2 ions detected during the analysis contained less than a 10.0pm error from the identified peptide sequence. The MS2 mass error can be found in Figure 8.

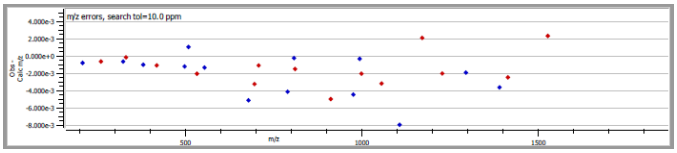

Figure 8. Mass errors of the MS2 ions detected from the peptide identified

INTERNAL DISULFIDE LINKAGES

One of the most important post translational modifications (PTM's) for any mAb are the internal disulfide linkages. Disulfide linkages enable an antibody to adopt to a conformation that will allow for the antibody to bind with the target receptor. There are multiple internal disulfide linkages necessary for the proper conformation of a mAb to be formed, but there can also be scrambling of the disulfide linkages. Hence, being able to properly identify and map these important linkages is often required to fully characterize a mAb. Due to the nature of the linkage, a protease digestion results in relatively large molecular fragments, with higher charge states and more complex MS2 spectra. Thus, the need for advanced deconvolution software from Protein Metrics that can analyze and deliver confident identifications of the disulfide linked peptides is crucial. Figure 9 is an identified spectrum from an expected disulfide linkage in a mAb. The disulfide linkage is between the cysteine amino acids (red C's in Figure 9).

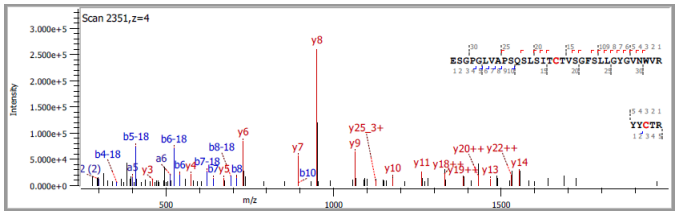

Figure 9. MS2 spectra from identified disulfide linked peptides

Excellent sequence coverage of the larger peptide and identified spectrum from the smaller peptide, allow for confident identification of the disulfide linkage. Both the MS1 and MS2 ions detected during the analysis contained less than a 20.0pm error from the identified peptide sequence. The MS2 mass error can be found in Figure 10.

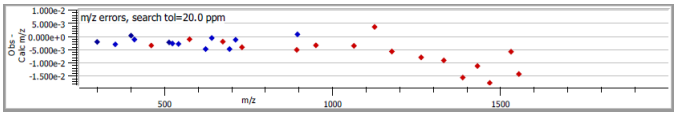

Figure 10. Mass errors of the MS2 ions detected from the disulfide linked peptide identified

GLYCOPEPTIDE ANALYSIS

Glycosylation of proteins describes the attachment of complex sugar moieties to specific amino acids. Glycosylation is an important modification to proteins such as monoclonal antibodies (mAB) and certain viral proteins. Glycosylation may play a vital role in recognition of the target binding site, or in the case of viral proteins, protection from antigen recognition and binding. In mAb's, glycosylation takes place at a specific asparagine (N) residue located in the heavy chain of the antibody. This in known as N-linked glycosylation. The glycosylated mAB was analyzed by LC/MS using a technique known as stepped collision energy (SCE). The SCE fragments the glycosylated mAB peptide to generate ions specific to both the sugar as well as the peptide backbone. This technique allows for the fragmentation of the glycopeptide at different energies and the spectra from these different fragmentations are averaged to produce the final spectrum.

The SCE fragmentation produces a large number of ions specific to the molecule. Annotation of these spectra can be complex. The software utilized (Byonic™ from Protein Metrics) is able to annotate ions produced from either the sugar, peptide, or the combination of the two. Figure 11 is the spectra for the glycosylated peptide. Annotation of the spectra is labeled green for sugar, or peptide fragments containing sugar moieties. Peptide fragments are blue (b ions), or red (y ions). Both the MS1 and MS2 ions detected during the analysis contained less than a 10.0pm error from the identified peptide sequence. The MS2 mass error can be found in Figure 12.

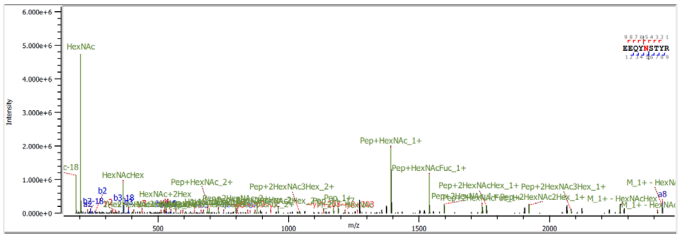

Figure 11. Identified spectrum of a known glycosylated peptide (EEQYN[+1444.534]STYR) from a mAb

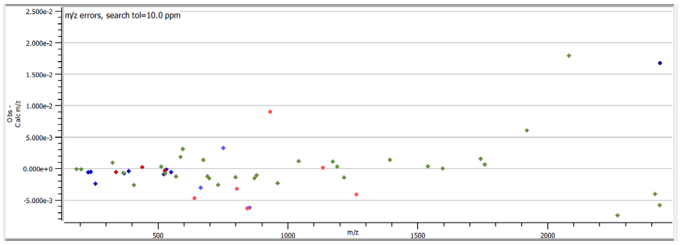

Figure 12. Mass errors for the glycosylated peptide (EEQYN[+1444.534]STYR)

LC-MS Characterization of Therapeutic Proteins

Zoomed views of both the low and high m/z range of the MS2 fragments are shown as well are shown in Figure 13 and Figure 14.

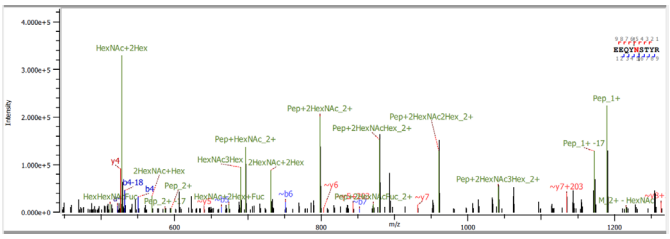

Figure 13. Zoomed view of the ions annotated from Figure 11 in the low m/z range.

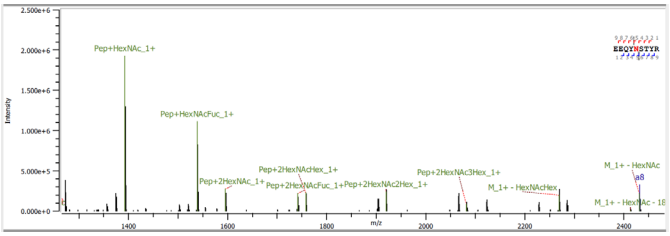

Figure 14. Zoomed view of the ions annotated from Figure 11 in the high m/z range.

SUMMARY

High Resolution LC-MS is a tool for characterizing proteins, which may be used for therapeutic or diagnostic applications. Depending on the nature of the request, it may be used to determine basic peptide sequence, post-translational modifications, internal disulfide linkages, overall protein conformation and binding activity to other molecules. Such information may be important for research and development projects, internal QC checks, or intellectual property support.