Monoclonal Antibody (mAb) Characterization Using PEAKS – Data Analysis Overview

Lin He, Baozhen Shan
Bioinformatics Solutions Inc. Waterloo, Ontario, Canada

Introduction

The development of reliable workflows using liquid chromatography-tandem mass spectrometry (LC-MS/MS) for antibody characterization has become an active research area. Beyond the identification of the primary sequence, researchers are also interested in post-translational modifications (PTMs), degradation analysis, and variant characterization. [1] To fulfill these requirements, we propose a reliable workflow for monoclonal antibody characterization from the purified antibody sample to MS/MS data analysis using PEAKS software.

Workflow

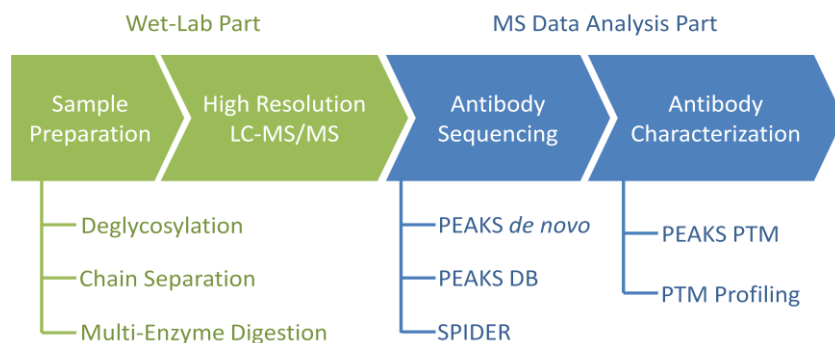The presented workflow includes wet-lab bench work, LC-MS/MS analysis, and PEAKS-based MS data analysis (Figure 1):

Figure 1. Workflow of monoclonal antibody characterization.

Sample Preparation

The heavy and light chains of the antibody are separated by SDS polyacrylamide gel electrophoresis (SDS-PAGE). Each excised band is reduced with dithiothreitol (DTT). Free cysteine residues are then alkylated using iodoacetamide. The heavy chain bands are deglycosylated with PNGase-F. The pH is adjusted for each protease and six enzyme digestions (AspN, Chymotrypsin, GluC, LysC, Protease K, Trypsin) are carried out overnight.

High-resolution LC-MS/MS

The desalted peptides are subjected to LC-MS/MS analysis on a Thermo-Fisher Scientific Q-Exactive Orbitrap mass spectrometer. The Q-Exactive runs in a data dependent mode with 10 MS/MS events per cycle. The parent ion resolution is 70,000 FWHM and the fragment ion resolution is 17,500 FWHM. Twelve data sets (six data sets for the heavy chain and the other six for the light chain) are generated for subsequent data analysis.

PEAKS-Based Antibody Sequencing

PEAKS is capable of performing *de novo*-assisted protein sequence identification and this application is well suited for antibody sequencing. In the first round search, PEAKS *de novo* and PEAKS DB help to determine the species

and antibody sequence candidates from the UniProt database, including both the constant and variable regions. Thereafter, iteratively manually (PEAKS de novo + PEAKS DB) and automatically (SPIDER) revising candidate sequences helps to unveil the exact amino acid sequences for both the heavy and light chains.

PEAKS-Based Antibody Characterization

PEAKS PTM searches the spectral data against the heavy and light chain sequences with 485 PTMs integrated in PEAKS for PTM identification with high confidence. PTM profiling, thereafter, provides the quantitative analysis of each PTM of interest to researchers. This tool provides the quantity ratio between unmodified and modified forms of a protein at a specific site with a user-selected PTM.

De Novo-Assisted Sequence Identification

A major benefit in mAb sequencing using PEAKS is the automated mapping from *de novo* tags to the selected candidate sequences. Exact antibody sequences, especially the complementarity determining regions (CDRs), are always absent from existing protein databases, thus a traditional database search cannot determine each amino acid in the target antibody. Fortunately, the relatively stable areas in the variable regions are recorded in public antibody sequence databases in most cases and provide clues to match *de novo* sequences. Users can correct candidate sequences by manually checking the sequence segments that are not covered by any tandem mass spectra but have partially matching *de novo* sequences. The objective of such amino acid correction is to cover the revised candidate sequences with maximum number of tandem mass spectra. Figure 2(a) shows an example of the manual correction of amino acid residues in the candidate sequence according to a *de novo* sequence.
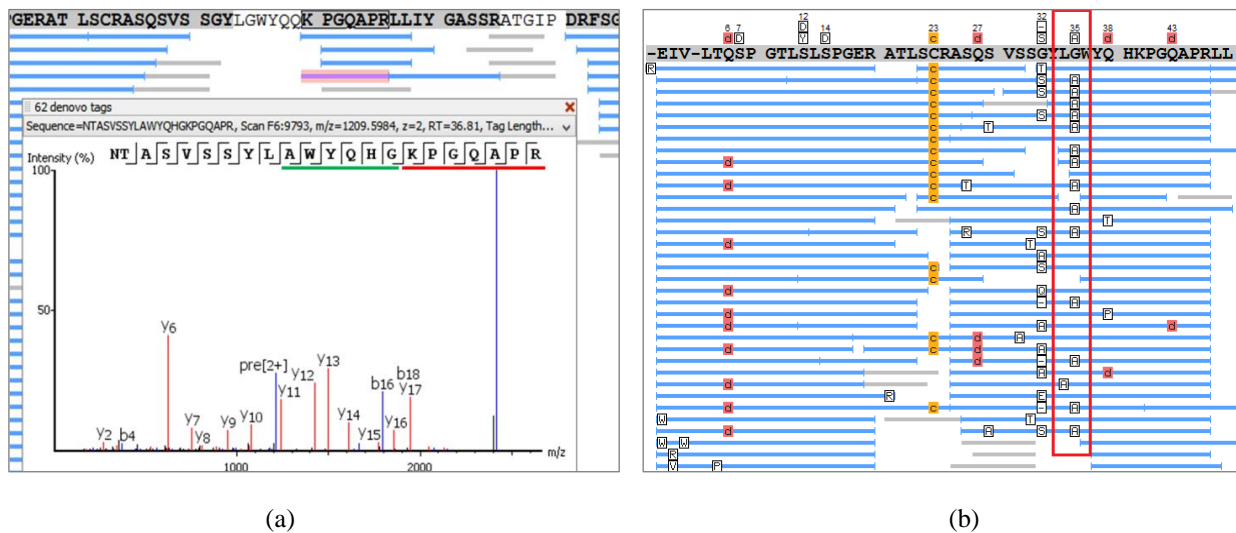

(a)   (b)

Figure 2. (a) Manual correction of amino acids in the candidate sequence. The red bar denotes the shared amino acids between the de novo sequence and the candidate sequence. The green bar denotes the amino acids to be corrected manually in the candidate sequence. (b) Automated correction of amino acids in the candidate sequence. According to SPIDER, G35, located in the red box, should be corrected to A35.

Additionally, SPIDER, provided within PEAKS, enables automated amino acid identification from PEAKS *de novo* and PEAKS DB results. SPIDER double-checks a *de novo* sequence with its mapping segment in the candidate sequence and the corresponding tandem mass spectrum to compute an optimal peptide sequence that best matches both the spectrum and the candidate sequence segment. It automatically works out three types of variants, 1) insertion, 2) deletion, and 3) substitution, and significantly helps to accelerate the antibody sequencing. Figure 2(b) shows an example of the automated correction of an amino acid residue in the candidate sequence using SPIDER.

Full Sequence Coverage with PTM Identification & Sequence Validation

Successful antibody sequencing generates both the heavy and light chain sequences, of which each amino acid is mapped typically with more than 20 peptides. The ability for each amino acid of the identified protein sequences to be validated according to a user-specified peak intensity threshold is a feature scheduled for release in PEAKS (early 2016). Figure 3 shows a protein coverage view from PEAKS PTM with amino acid validation function integrated.
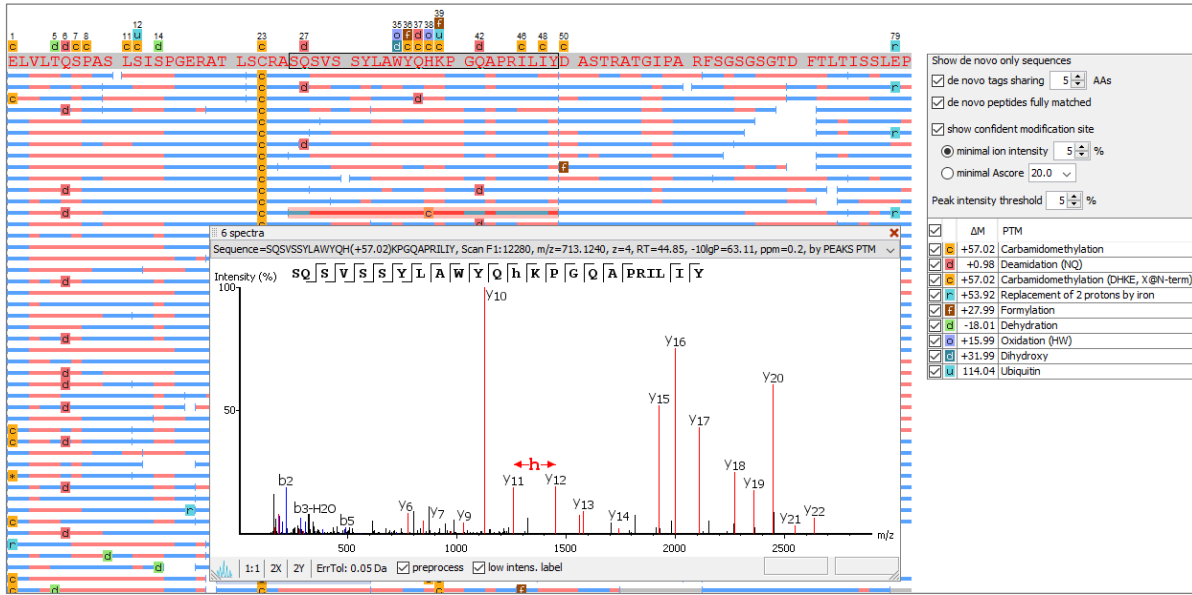

Figure 3. Full sequence coverage with PTM identification and intensity-based amino acid validation.

Figure 3 displays frequently observed PTMs from the sample listed along right. Users can adjust the "minimal ion intensity" or "minimal Ascore" to filter out PTMs with low confidence. Amino acids annotated with red color indicate the existence of supporting peaks with higher intensity than the user-specified threshold. Amino acids annotated with blue color indicate the absence of supporting peaks with higher intensity and they are identified according to the accurate precursor mass from the given protein sequences. Users can easily specify the "peak intensity threshold" in the configuration panel.
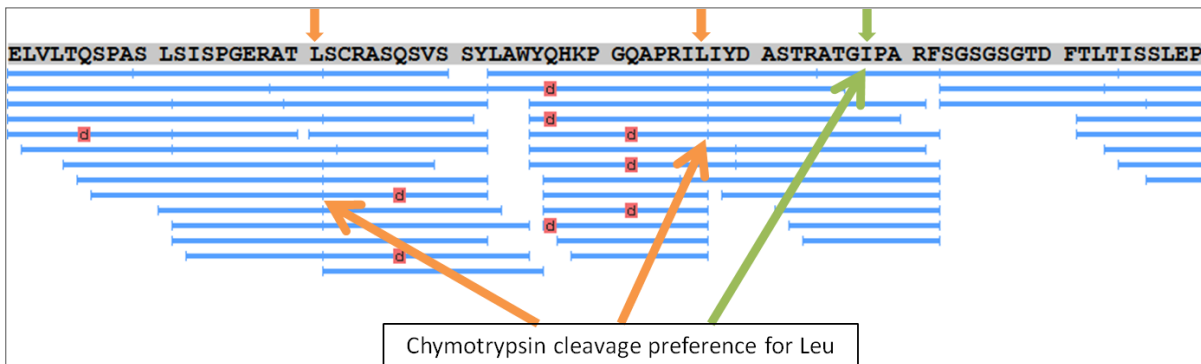

Figure 4. Leu/Ile discrimination using chymotrypsin cleavage preference.

Leu/Ile Discrimination

The identity of Leu and Ile can be distinguished in the final antibody sequence by checking the chymotrypsin cleavage sites [2]. Figure 4 shows an example of Leu/Ile discrimination using the chymotrypsin cleavage preference.
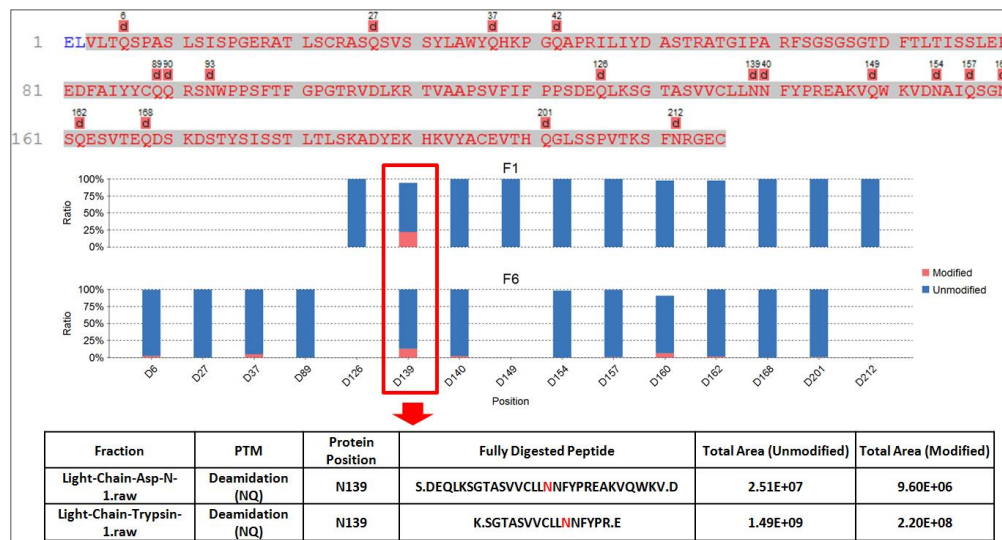

Figure 5. An example of PTM profiling for quantitatively analyzing deamidated Asn and Gln.

Quantitative Analysis of PTMs

PEAKS Studio 7.5 provides PTM profiling to quantitatively analyze PTMs identified by PEAKS DB and PEAKS PTM. Figure 6 shows an example of the deamidation PTM profiling result. In Figure 5, each bar represents a modification site on the antibody sequence, of which the modification positions are marked in the x-axis. The blue and the red sections show the percentages of accumulated quantities of unmodified and modified peptide features in the spectral data, respectively. Using the upcoming patch of PEAKS Studio the quantity information of a user-specified PTM for each modification site on a selected protein can be easily exported into spreadsheet files.

Conclusions

PEAKS provides a reliable workflow for comprehensively characterizing monoclonal antibodies. Taking advantage of both the latest LC-MS/MS technology and features presented in the PEAKS software package, the quality of the antibody characterization is guaranteed. The complete workflow, including wet-lab bench work through PEAKS-based MS data analysis, is available as the PEAKS AB service from Bioinformatics Solutions Inc (BSI).

References

[1] B. Shan & L. Xin. Integrating *de novo* sequencing and database search for monoclonal antibody sequencing. ABRF 2013.

[2] C. Poston, *et al*. A quantitative tool to distinguish isobaric leucine and isoleucine residues for mass spectrometry-based *de novo* monoclonal antibody sequencing. *J. Am. Soc. Mass Spectrom.* 25.7 (2014): 1228-1236.