Retention time prediction and intact mass validation improves automated antibody sequencing in PEAKS AB 3.5

Shakiba Zeinali and Kyle Hoffman

Bioinformatics Solutions Inc., Waterloo, Canada

Abstract:

Monoclonal antibodies (mAb) are vital in modern medicine due to their ability to specifically target and bind to antigens, making them powerful tools in the treatment of various diseases. Monoclonal antibody sequencing is crucial in biotechnology and medicine as it allows for the development of targeted therapies for various diseases, including cancer, autoimmune disorders, and infectious diseases. Here, we advance our PEAKS AB software capabilities by incorporating retention time prediction for selection of peptides used in antibody sequence assembly, and an intact mass deconvolution tool. As a validation step, each assembled sequence is automatically matched to intact mass peaks for both heavy and light chains. These new features reduce the need for manual sequence editing and improve the efficiency and accuracy of antibody sequencing.

Introduction:

Monoclonal antibodies (mAbs) are increasingly important in the pharmaceutical market, with half of the top-selling drugs in 2020 being mAbs. Their development into targeted therapies for various diseases including cancer, autoimmune disorders, and infectious diseases [1-2] requires thorough physicochemical and biophysical characterization due to their complexity and susceptibility to modifications [3]. Monoclonal antibody sequencing and post-translational modification (PTM) analysis is a crucial part of this process. By understanding the exact amino acid sequences of monoclonal antibodies and modification states researchers can replicate, modify, and enhance their efficacy, stability, and specificity. Overall, due to the continuous increase in the pharmaceutical application of monoclonal antibody, their characterization is foundational for advancing personalized medicine and drug development.

Achieving 100% accuracy in antibody sequencing is essential for reverse engineering and recombinant expression of an antibody that will effectively neutralize the target antigen. Differentiating between isobaric residues is one of the main challenges in this process. There are several cases where the mass of two amino acids is equivalent to the mass of one (for example, GA = Q, GG = N, and GV = R), and the lack of fragmentation ions can result in sequencing errors for this reason. Moreover, the mass of deamidated Q and N is equivalent to the mass of E and D, respectively. Retention time (RT) prediction can be leveraged to resolve this issue, despite there being incomplete fragmentation of the peptide, since amino acids such as Q, N, and R have very different hydrophobic properties compared to GA, GG, and GV, respectively. Calculating the difference between measured and predicted RTs (deltaRT) can be applied to decipher whether a peptide contains, for example, a deamidated N or Q versus D or E. Since each of these isobaric residues possess different properties, their differentiation is critical, especially in complementary determining regions (CDRs). In addition to enhanced peptide selection for sequence assembly using RT prediction, new features in PEAKS AB 3.5 include in-depth N- and O-linked glycan profiling, support for timsTOF data, enhanced intact mass deconvolution, and I/L differentiation is now possible for the sequence validation workflow. Taken together, these features will allow for a better understanding of antibody structure and function and help guide the development and engineering of new antibody therapeutics and diagnostic tools.

Methods:

Digestion

In-solution endoproteinase digestion of the monoclonal antibody (mAb) was carried out for sequencing analysis. The antibody was reduced with DTT and alkylated with iodoacetamide. The sample was then divided into five aliquots, each subjected to digestion with one of five enzymes—Asp N, Chymotrypsin, Elastase, Trypsin, and Pepsin. After reverse-phase cleanup of each digest, the samples were frozen until mass spectrometry analysis.

Instrumental and Data Analysis

Intact mass analysis included DTT reduction and PNGase F deglycosylation, followed by LC-MS, using a Thermo Scientific Orbitrap Fusion Lumos Tribrid mass spectrometer equipped with a heated electrospray ionization source in positive ion mode, paired with a Thermo Fisher Ultimate 3000 RSLCnano HPLC system. For sequencing, the digests were analyzed on an Orbitrap analyzer (Orbitrap Fusion Lumos, Thermo Fisher Scientific). MS2 data were acquired using HCD and ETD followed by HCD (EThcD) fragmentation methods. All raw data files were used for data analysis using the PEAKS AB 3.5 software. Carbamidomethylation was chosen as fixed PTM, while deamidation, oxidation and Pyro-glu from Q were set as variable PTM. Parent ion mass error tolerance was 10 ppm and 0.02 Da was set for fragment ion mass error tolerance.

Results and Discussions:

To demonstrate how the new features in PEAKS AB 3.5 improve the accuracy and assembly of antibody sequences, we first assessed peptide coverage across the US Pharmacopeia USP3 monoclonal antibody standard [4]. PEAKS AB 3.5 provided high coverage of both constant and variable region (Figure 1) by assembling the peptides from all enzyme digestions and MS runs in both HCD and EtHCD mode for data acquisition. Next, for validating that the assembled sequences are correct, PEAKS AB 3.5 successfully assigned the theoretical masses generated from assembled amino acid sequences for heavy and light chains to deconvoluted LC-MS data from the intact antibody after reduction and deglycosylation (Figure 2).

Peak assignments were matched with consideration of modifications to the sequence, such as C-terminal lysine truncation of the HC. When assigned intact peaks are chosen from the table, their retention time and chromatographic peaks are indicated in the bottom-left window. The major peaks used for deconvolution are displayed in the bottom-right window. These displays allow for the user to confirm that the deconvoluted peaks are reliable and were generated from good quality data.

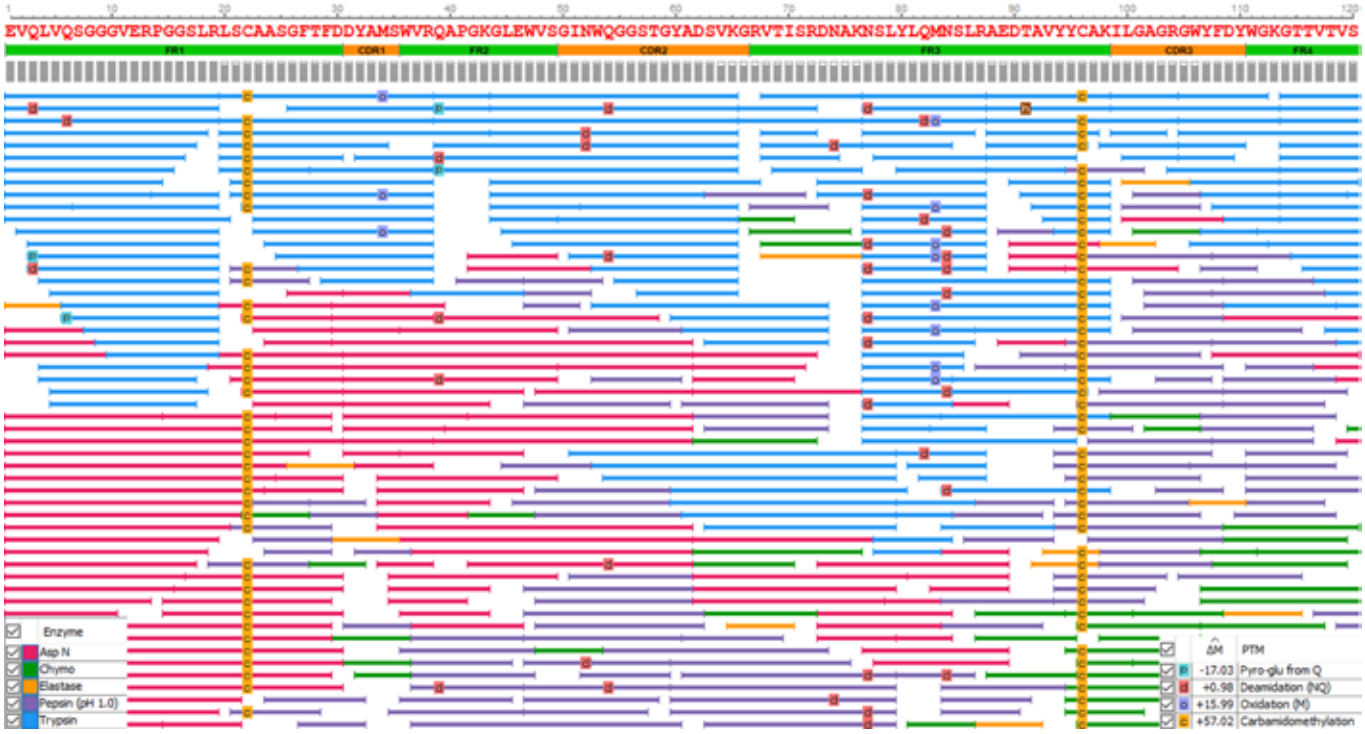


Figure 1. Peptide coverage of the variable region of USP-3 antibody.

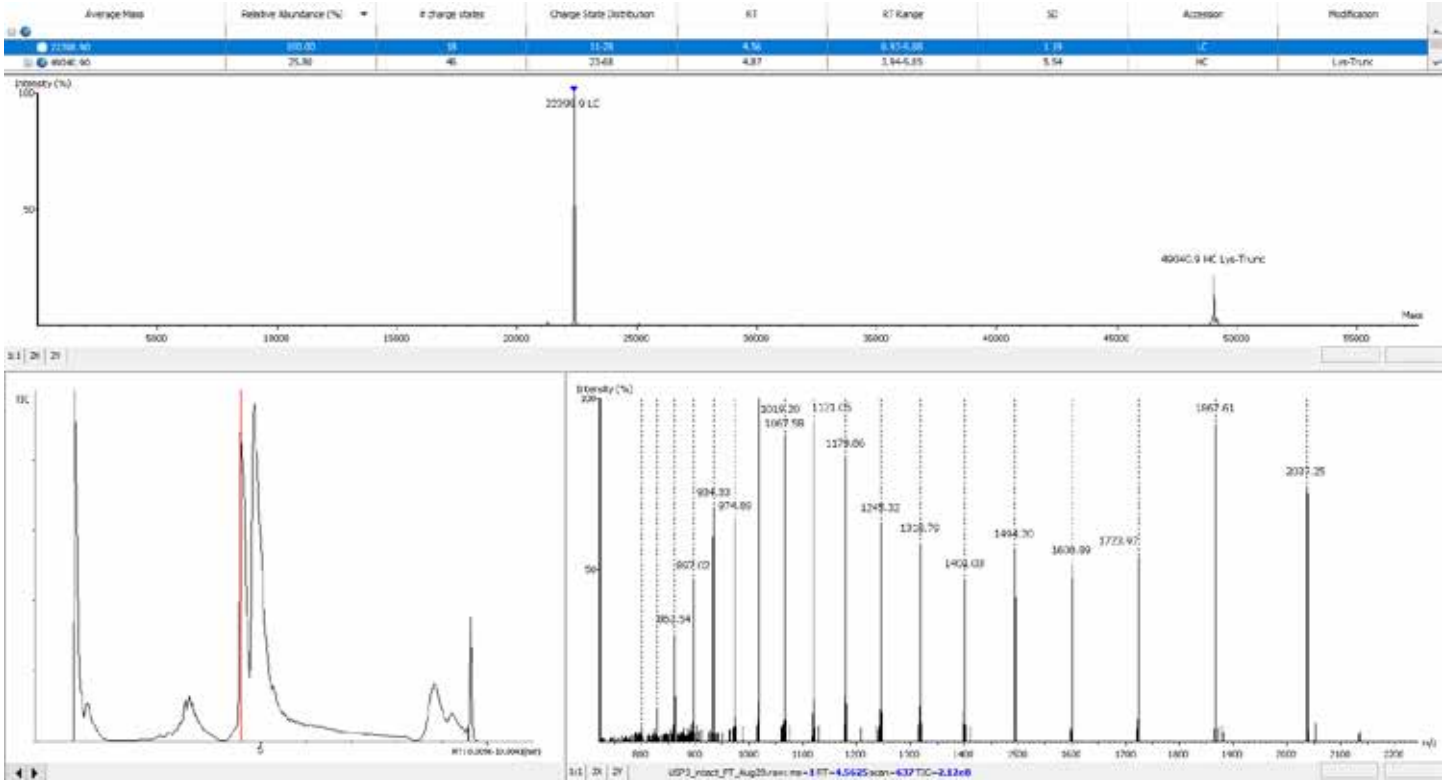


Figure 2. Peaks assigned to heavy and light chains from analysis of LC-MS data.

The use of deep learning technology and RT prediction has become increasingly popular for peptide-spectral match (PSM) rescoring and improving the accuracy in peptide sequencing. For this reason, we implemented RT prediction in PEAKS AB 3.5 to find the top scoring peptide sequence for each spectrum and to differentiate between isobaric residues. Deltra RT (the difference between predicted and actual RT) is used to help assign the best de novo sequence candidate to a spectrum and can differentiate between residues such as N and GG. Figure 3 illustrates an example of how a sequencing error in the CDR3 was corrected by applying RT prediction. Without correcting such sequencing errors in antibody CDRs would likely result in the expression of an antibody lacking binding affinity to the target antigen.

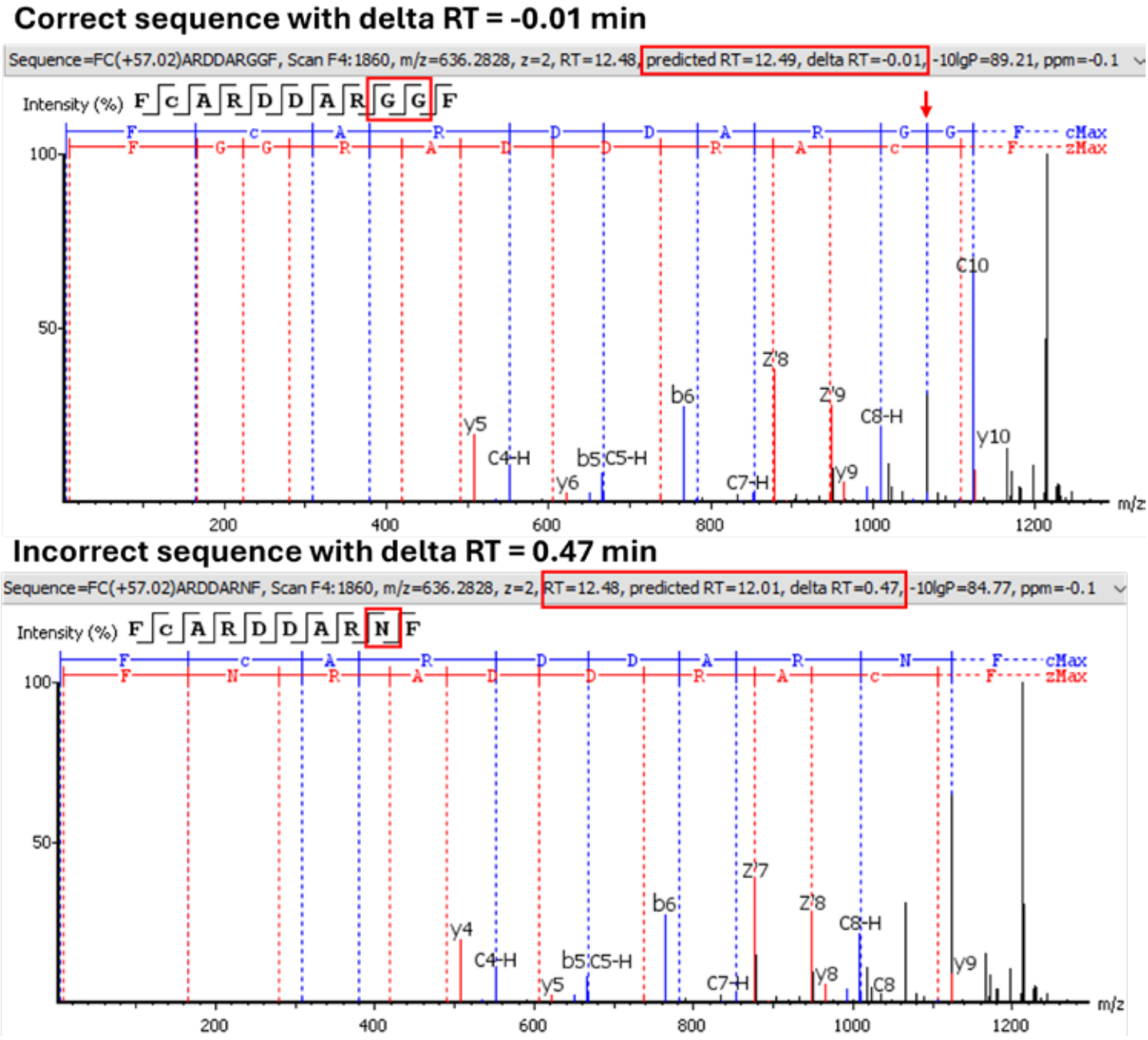
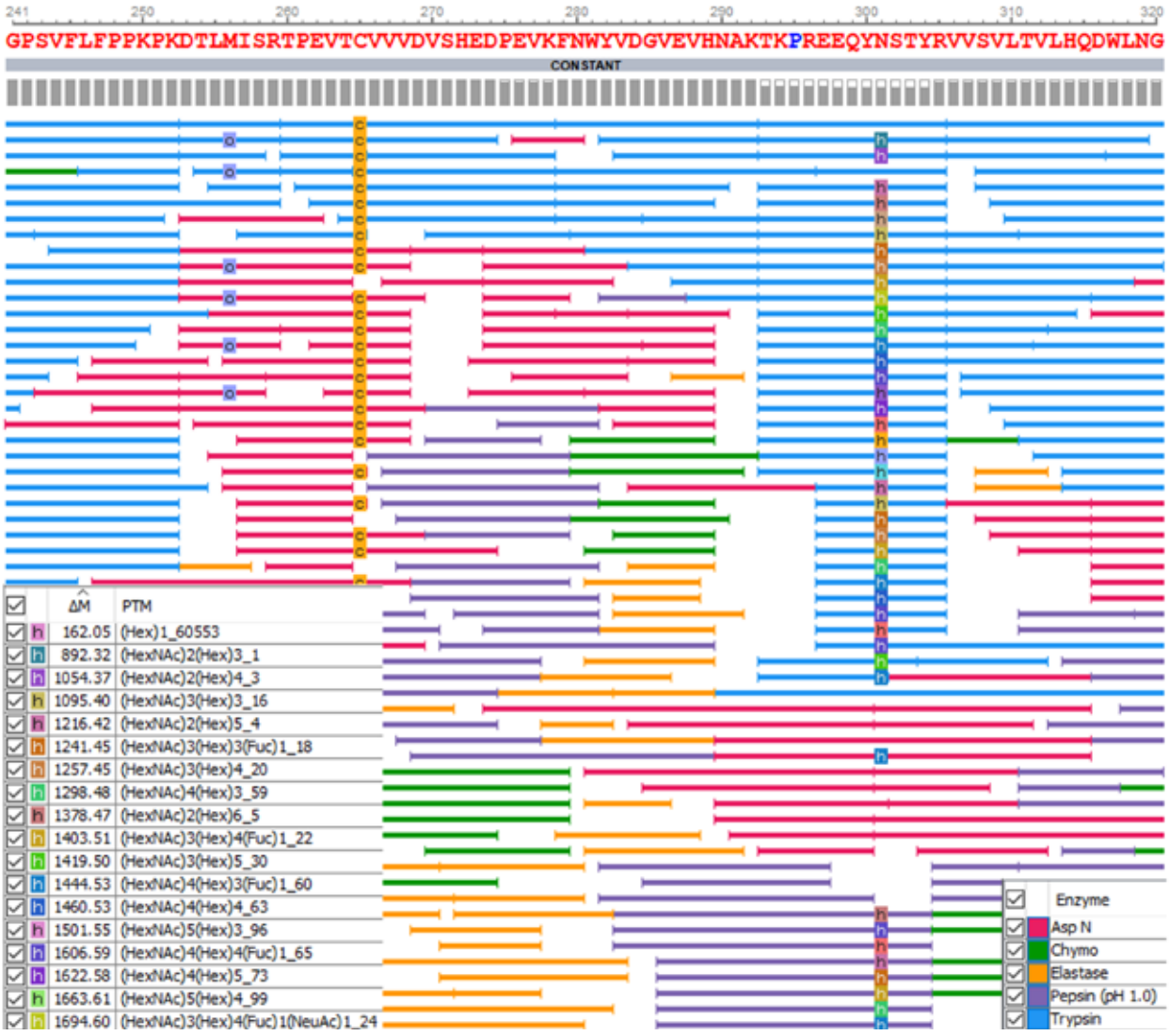


Figure 3. Retention time prediction of de novo sequence candidates. By applying RT prediction, the software correctly chooses GG using RT prediction, while without RT prediction N was assigned.

We recently incorporated the software algorithm from PEAKS GlycanFinder into PEAKS AB 3.5, for improved N- and O-linked glycopeptide identification and glycan profiling. Due to the nature of glycosylated peptides, their ionization efficiency is low and peptide backbone fragmentation is poor, resulting in decreased peptide mapping and sequence coverage near glycosylation sites compared to the rest of the sequence. In PEAKS AB 3.5, glycopeptides identified by the Glycan Profiling tool are now merged into peptide mapping tab. By using this workflow, in-depth glycopeptide coverage at the glycosylation site (N301) in USP3 was achieved (Figure 4a). Annotation of both peptide backbone and glycan fragment ions can be viewed by selecting any of the glycopeptides. This feature also provides the predicted structure for each glycan and is shown at the top right of each spectrum (Figure 4b).

A)



B)

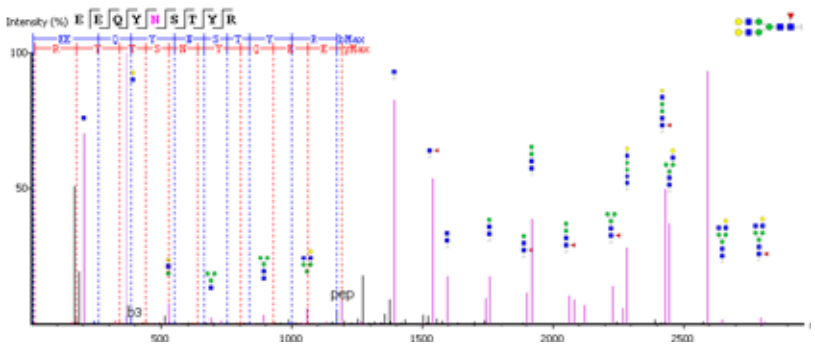


Figure 4. Glycan Profiling results for N-linked site of the USP3 antibody heavy chain.
A) Glycopeptide mapping at site N301 of USP3.
B) Representative glycopeptide spectrum. B- and y-ions are indicated in blue and red, respectively. Glycan fragment ions are shown in purple, with glycan moieties above each ion.

Since different glycans may influence the structure and function of the antibody, estimating the relative abundance of antibody glycoforms is an important step in sample characterization. Within the Glycan Profiling tool of PEAKS AB 3.5, selecting the modification site will display a pie chart that shows the relative abundance of different glycan compositions (Figure 5). Overall, glycan compositions and relative abundances are consistent with what is expected for the USP3 antibody [4], indicating that glycan profiling in PEAKS AB 3.5 is accurate in both identification and quantification.

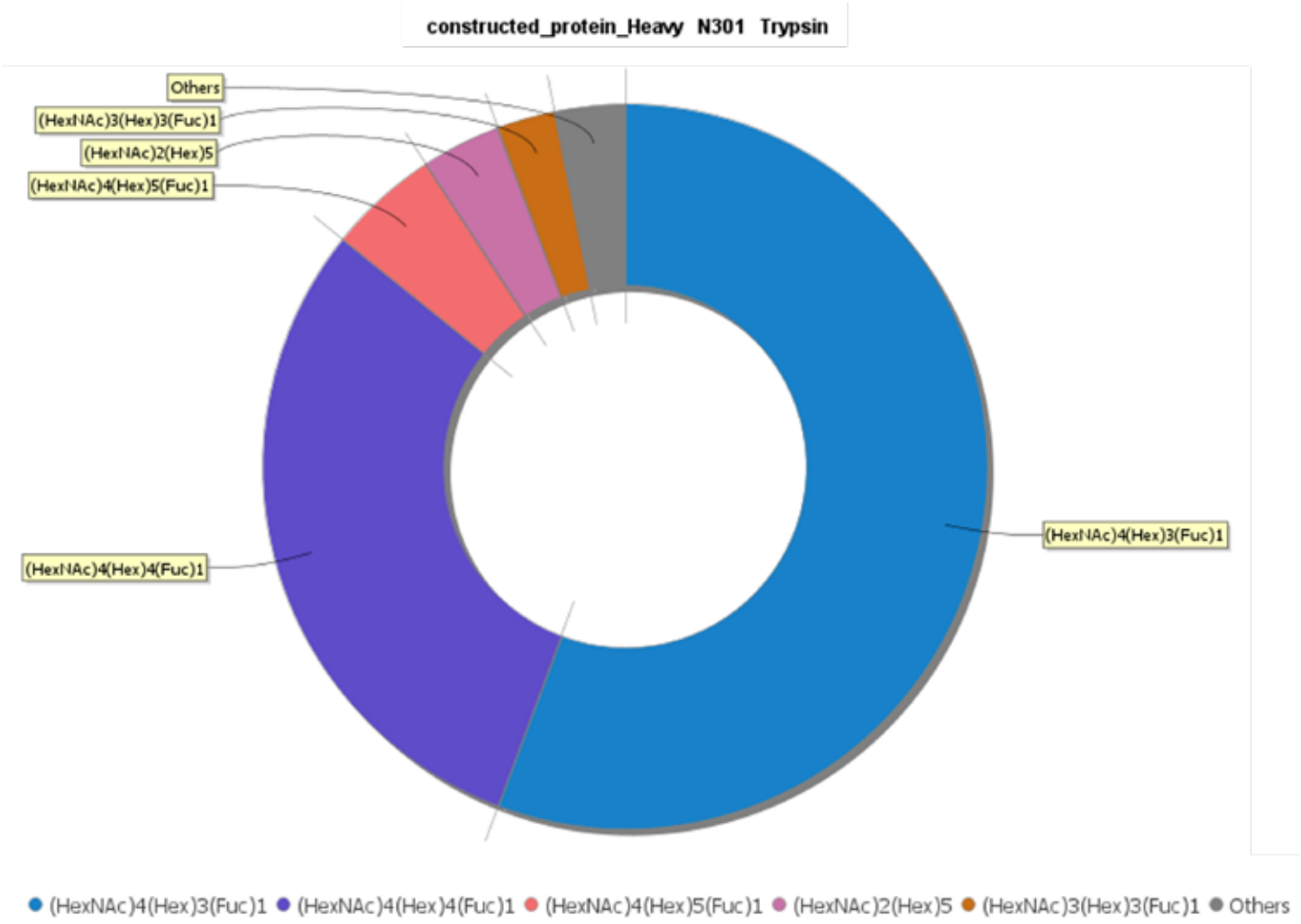


Figure 5. Relative abundance of glycans at N301 in USP3.

In conclusion, PEAKS AB 3.5 provides a comprehensive solution for accurate and automated antibody sequencing and characterization. The user-friendly interface allows easy interpretation of results and supports raw data from all major instrument vendors. Lastly, the new features introduced here, including RT prediction for peptide selection in sequence assembly, intact mass validation, and glycan profiling will directly benefit users for the development of effective therapeutics and research tools.

References:

[1] Production of monoclonal antibodies for therapeutic purposes: A review, International Immunopharmacology 2023, 120, 110376. https://doi.org/10.1016/j.intimp.2023.110376

[2] Monoclonal antibodies in breast cancer: A critical appraisal, Critical Reviews in Oncology/Hematology 2023, 183, 103915. https://doi.org/10.1016/j.critrevonc.2023.103915

[3] Analytical Techniques for the Characterization and Quantification of Monoclonal Antibodies, Pharmaceuticals 2023, 16(2), 291; https://doi.org/10.3390/ph16020291

[4] USP Reference Standard Monoclonal Antibodies: Tools to Verify Glycan Structure, Pharmaceuticals 2022, 15, 315. https://doi.org/10.3390/ph15030315

Bioinformatics Solutions, Inc.

140 Columbia St, Suite 202
Waterloo, Ontario N2L 3K8
Canada

Tel: (519) 885-8288
Fax: (519) 885-9075

sales@bioinfor.com
www.bioinfor.com