Automated *De Novo* Sequencing of Antibodies with Isoleucine/Leucine Differentiation by using EThcD Fragmentation

Wen Zhang[1], Lin He[1], Lei Xin[1], Jonathan R. Krieger[2], Michael F. Moran[2], Baozhen Shan[1]

[1]Bioinformatics Solutions Inc., Waterloo, Ontario, Canada [2]SPARC BioCentre, Hospital For Sick Children, Toronto, Ontario, Canada

Introduction

Antibody *de novo* sequencing by LC-MS/MS has gained more interests and applications in biologics research. However, differentiation of isoleucine and leucine remains a challenging task. EThcD fragmentation scheme combines electron transfer and higher-energy collision dissociation to give extensive peptide backbone fragmentation and generate MS/MS spectra containing both b/y and c/z ions. The richer MS/MS spectra derived from EThcD have been suggested to be advantageous for peptide *de novo* sequencing and identification of post-translational modifications. Moreover, the isobaric amino acids leucine and isoleucine can be discriminated by the signature w-ions generated by EThcD.

Summary

In this work, we present our newly released software platform, PEAKS AB 2.0, for automatic antibody protein *de novo* sequencing with Ile/Leu differentiation by using combination of HCD and EThcD fragmentation methods. Furthermore, PEAKS AB 2.0 enables intact mass deconvolution from LC-MS data.

Workflow

- **Bottom-Up Sequencing Procedure**

1) Digest antibody protein into overlapping peptides using an optimized set of orthogonal enzymes
2) Analyze peptides by high resolution LC-MS/MS, supporting data generated from CID/ETD/HCD/EThcD
3) Peptide *de novo* sequencing from tandem mass spectra
4) Construct antibody sequences by assembling *de novo* peptide sequences
5) Ile/Leu differentiation by using advanced EThcD method (if any), enzyme digestion specificity and homology database analysis

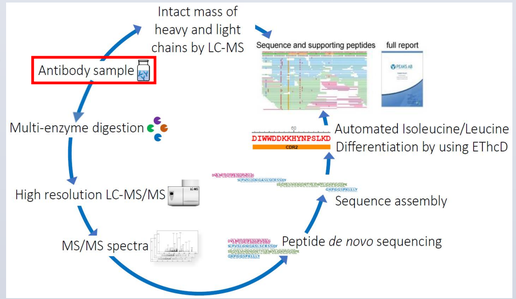


Figure 1. Workflow of antibody protein *de novo* sequencing using LC/MS coupled with PEAKS AB software platform.

- **Intact Mass Validation Procedure**

1) Reduce antibody protein to separate light and heavy chains
2) Remove N-linked glycans from the heavy chain by PNGase-F
3) Analyze the reduced mixtures by LC-MS
4) Deconvolution of acquired MS spectra to derive masses of light and heavy chains respectively

Methods

A standard antibody sample, NIST mAb, was reduced, alkylated, and digested by using multiple enzymes. The digests were analyzed on an Themo Scientific Fusion Lumos Tribrid Mass Spectometer with high resolution MS1 and MS2. All raw files were analyzed by PEAKS AB 2.0 software platform for automated antibody *de novo* sequencing with Ile/Leu differentiation by EThcD (**Figure 2**):

1) Peptide *de novo* sequencing was performed for each MS2;
2) Heavy and light chains were automatically assembled based on overlapping amino acids of the *de novo* peptides;
3) For each Ile/Leu position, signature w-ions resulted from different side chain breakage were searched for automated discrimination of Ile/Leu (**Figure 3**);
4) EThcD results, enzyme cleavage specificity, and homologous database statistics were considered for assigning Ile/Leu confidence.

• 100% Protein Sequence Coverage and Accuracy

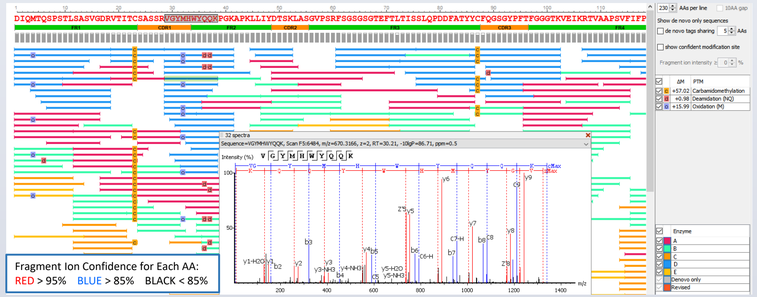
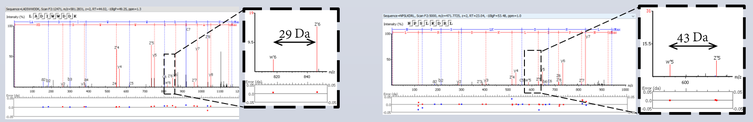


Figure 2. PEAKS AB 2.0 automatically constructs the NIST mAb protein sequences from MS/MS data with 100% protein sequence coverage and accuracy.

• Ile and Leu Differentiation

EThcD generates w-ions that distinguish Leu from Ile:



(A) (B)

Figure 3. PEAKS AB 2.0 automatically considers the different w-ion information induced by Ile/Leu in peptide sequences identified from MS/MS. A) z-ion with Ile on its N-terminus loses ethyl (-29), B) z-ion with Leu on its N-terminus loses isopropyl (-43).

Results

• Ile and Leu Confidence

Different confidence levels were given to Ile/Leu assignments based on the intensities of w-ions, the presence of the paired z-ions, mass errors, and etc. In addition, enzyme cleavage specificity and homologous database statistics were also considered. More than 90% of assigned Ile / Leu positions had high confidence.

Heavy Chain:
QVTLRESGPALVKPTQTLTLTCTFSGFSLSTAGMSVGWIRQPPGKALEWLADIWWDDKKHYNPSLK
DRLTISKDTSKNQVVLKVTNMDPADTATYYCARDMIFNFYFDVWGQGTTVTVSS…

Light Chain
DIQMTQSPSTLSASVGDRVTITCSASSRVGYMHWYQQKPGKAPKLLIYDTSKLASGVPSRFSGSGS
GTEFTLTISSLQPDDFATYYCFQGSGYPFTFGGGTKVEIK…

CDR Confidence of Ile/Leu: Red = High, Blue = Medium, Green = Low

• Intact Mass Analysis

Intact mass analysis was performed to further validate the constructed protein sequences and glycan forms attached. The intact NIST mAb was analyzed by LC/MS using SEC coupled to a Bruker Daltonics (Bremen, Germany) maXis Q-TOF mass spectrometer and PEAKS AB 2.0 was used to deconvolute the LC/MS data (**Figure 4**).

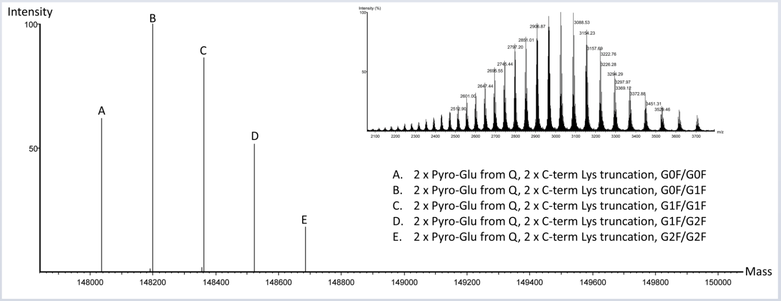


Figure 4. PEAKS AB 2.0 automatically deconvolutes the LC/MS and annotates the deconvoluted peaks with theoretical masses calculated from the input protein sequences.

Acknowledgement

We thank SPARC BioCentre for their collaboration. This work was funded by Garron Family Cancer Centre. We thank Bruker Daltonics for kindly providing intact mass data.

Contact

Bioinformatics Solutions Inc. http://www.bioinfor.com support@bioinfor.com