Complete protein *de novo* sequencing from antibody mixtures by LC-MS/MS

Lin He, Lei Xin, Wen Zhang, Baozhen Shan
Bioinformatics Solutions Inc., Waterloo, Ontario, Canada

Bioinformatics Solutions Inc.

Introduction

Studies of therapeutic proteins, especially antibodies, require complete and accurate protein sequences and comprehensive characterization of chemical environment of each amino acid. With current liquid chromatography tandem mass spectrometry (LC-MS/MS) technology, we have demonstrated successful sequencing of purified monoclonal antibodies (mAb's) by using peptide *de novo* sequencing and assembly algorithms. However, challenges remain for sequencing a mixture of antibodies which cannot be further separated and purified. In this study, we present an advanced method (AbMixNovo) to completely sequence proteins from antibody mixtures.

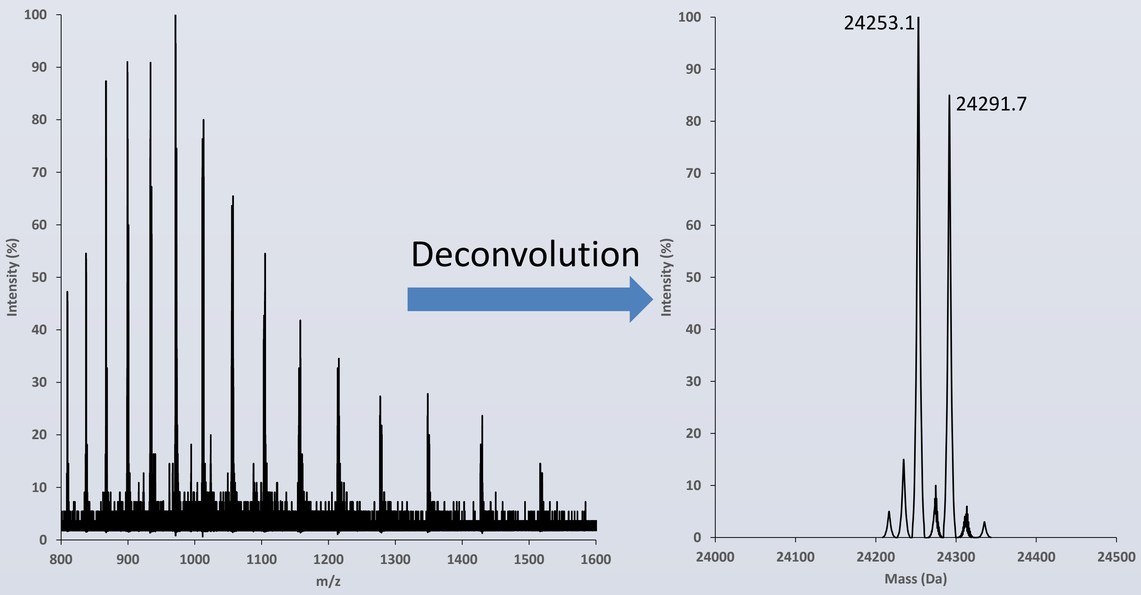
Summary

The results demonstrate that AbMixNovo can accurately sequence antibody proteins from a simple mixture of antibodies.

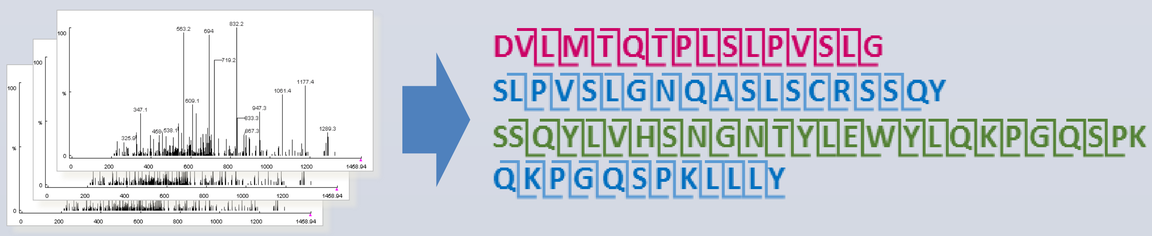
Methods

AbMixNovo takes intact mass data and LC-MS/MS data from multiple enzyme digestion as input. Four critical steps are carried out to obtain protein sequences from an antibody mixture:

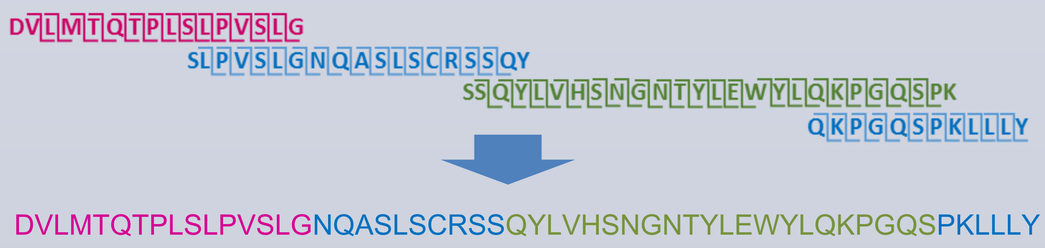
1. Intact mass data is deconvoluted to extract intact masses for all the heavy and light chains.



2. The peptide *de novo* sequencing algorithm in PEAKS AB is used to generate confident peptide tags from MS/MS spectra.



3. The *de novo* peptide tags derived from Step 2 are used to assemble protein sequences by using our published ALPS [1] system which adopts an advanced de Bruijn graph algorithm.

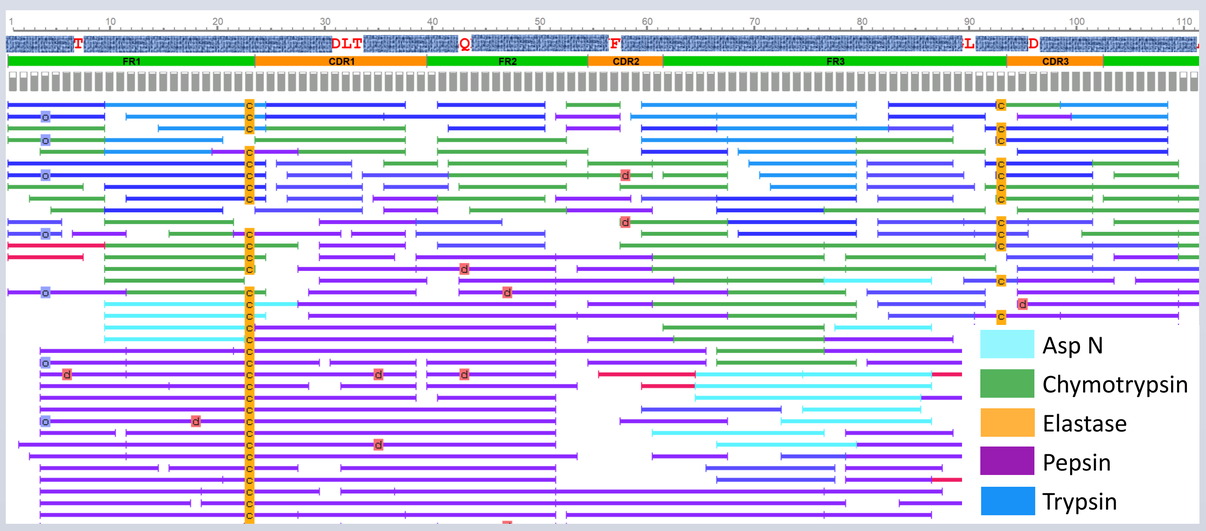


Methods

4. A refinement procedure is carried out taking into account of both the masses of the constructed sequences from Step 3 and the deconvoluted intact masses from Step 1. As shown in the figure below, all the possible combinations of regions with sequence variants confirmed by *de novo* sequencing from tandem mass spectra are generated. The deconvoluted intact masses are then used to select sequence candidates from these combinations with a certain mass error tolerance.

Results

Two mAb samples with known protein sequences were mixed deliberately. The mixture was equally divided into five aliquots for five individual enzyme digestions following manufacturer's instructions. The five digests were analyzed by a Orbitrap mass spectrometer. We used HCD fragmentation method and collected high-resolution data for complete protein sequencing.
We applied AbMixNovo for data analysis and obtained two complete protein sequences of light chains accurately.



References

1. Tran, N.H. *et al*. Complete *De Novo* Assembly of Monoclonal Antibody Sequences. Scientific Report. 6:31730. 26/08/2016.

Contact

support@bioinfor.com www.bioinfor.com Bioinformatics Solutions Inc.