Protein Identification and Quantitative Analysis with N-terminal Sequencing by Mass Spectrometry

PEAKS
www.bioinfor.com

Baozhen Shan, Hao Lin
Bioinformatics Solutions Inc, Waterloo, ON

Introduction

N-terminal sequencing is in the midst of a technology transition from classical Edman degradation sequencing to mass spectrometry-based N-terminal sequencing. Using chemical derivitization of protein N-termini enhanced N-terminal sequencing of proteins by mass spectrometry. There is a need to build a platform for quantitative N-terminal analysis with LC-MS/MS, which enables identification of a wide dynamic range of original mature protein N-termini and protease cleavage products. In this work, we reported an algorithm for protein identification and quantitative analysis by N-terminal sequencing with LC-MS/MS.

Methods

1. The tandem spectra (MS2) were searched against a database with 1% of FDR at the peptide level and protein inference with a minimum of two unique peptides. The unique peptides were filtered out for quantitative analysis.
2. For each unique peptide, the peptide features were detected in the precursor scans (MS1). The intensities of the peptide features were calculated with extracted ion chromatograms.
3. The proteins with N-terminal derivative hits were filtered out for quantitative analysis. Based on the sequence similarity, a group of unique peptides one for each homologous protein are selected for quantitative analysis. The relative protein abundance ratio was estimated by the measured intensities of the peptides in the group.

Results

Datasets for ABRF PSRG 2014 study were tested [1]. Hemoglobin sample (Sigma H7379) sample was separated with SDS-PAGE gels, and bands of interest were excised, reduced with DTT, alkylated with iodoacetamide, labelled with the labeling reagent TMPP, and then digested with trypsin. The samples were analyzed by LC-ESI-MS/MS (Orbitrap Q Exactive).

The data was searched by PEAKS against SwissProt protein database at 1% of FDR with the following parameters: 1 missed cleavage trypsin; mass tolerance 10 ppm (precursor), 0.05 Da (fragment ions); caramidomethyl cysteine (fixed), TMPP (protein N-terminal, variable), oxidized methionine (variable), and deamidation (variable). Protein identifications were considered with at least two unique peptides. The peptide features were detected and quantified with PEAKS Q.

Six proteins were identified by database search: human hemoglobin alpha (P69905), human hemoglobin beta (P68871), human hemoglobin delta (P02042), human hemoglobin gamma (P69892), human myoglobin (P02144) and human hemoglobin zeta (P02008). Three proteins were mapped with N-terminal TMPP-derivative peptides: P69905, P68871, and P02042, with 100% sequence coverage (Figure 1a). The supporting peptides containing N-terminal labelling were shown in Figure 1b.

PEAKS database search was integrated with de novo sequencing. Protein N-terminal sequences could be identified in cases where the database contained only homologous sequences. The successful example was shown in Figure 2, where S and T were exchanged manually.

The proteins, P68871 and P02042, contain the same N-terminal sequence VHLTPEEK (tryptic peptide). Their sequence alignment was shown in Figure 3. Their abundance ratio was estimated by the feature intensities of the most similar unique peptides (Figure 4), which is 20:1 (P68871: P02042).

Figure 1. Identified proteins with N-terminal labelling

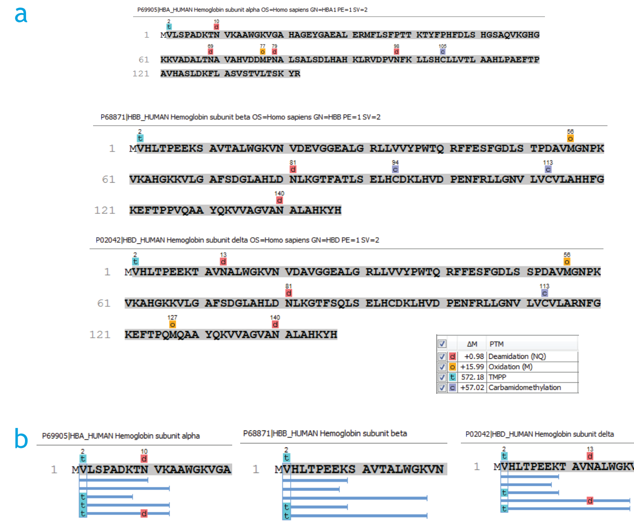


Figure 2. Homology matching for N-terminal sequencing

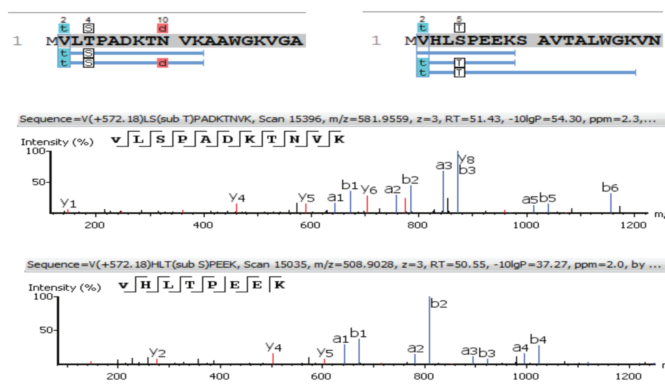


Figure 3. Sequence alignment of homologous proteins

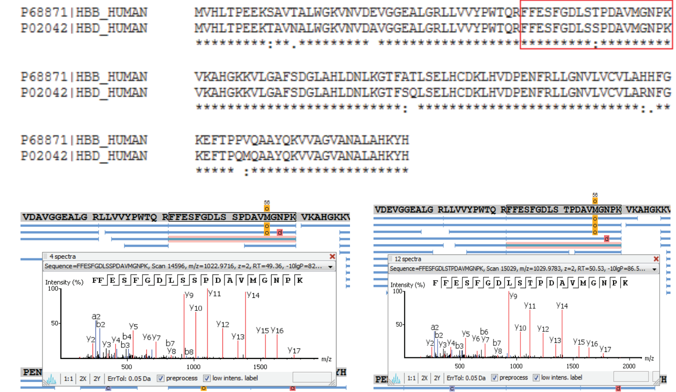
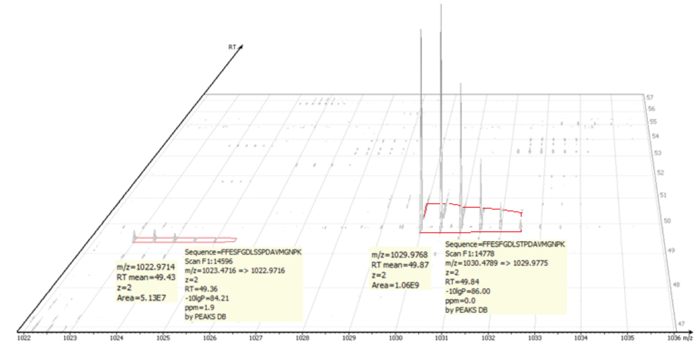


Figure 4. Quantitative analysis of unique peptides



Conclusion

An algorithm for quantitative analysis with protein N-terminal sequencing was proposed.

Reference

1. English et al. N-terminal Sequencing of Standard Proteins by N-terminal Labeling and Mass Spectrometry. Protein Sequencing Research Group. ABRF 2014.