PTMFinder Based on PEAKS™ De Novo Sequencing Result

Lei Xin[1]; Baozhen Shan[1]; Mingjie Xie[1]; Gilles Lajoie[2]; Bin Ma[3]

[1]Bioinformatics Solutions Inc., Waterloo, ON. [2]University of Western Ontario, London, ON. [3]University of Waterloo, Waterloo, ON.

Bioinformatics Solutions Inc.

Overview

Purpose: To discover variable PTMs in MS/MS proteomic data.

Methods: PEAKS™ de novo sequencing results are used to automatically discover variable PTMs.

Introduction

Identification of post-translational modification (PTM) by tandem mass spectrometry is still a major challenge in proteomics, especially if the PTMs are unknown. In typical existing software, tandem mass spectra are searched against an enlarged-database that includes all possible combinations of modified peptides. Because the search time grows exponentially with the number of allowed modifications, only a small number of known variable modifications can be included in each search. We propose a new approach based on de novo sequencing results to identify unknown variable PTMs from an MS/MS dataset.

Methods

For a peptide with variable PTM, usually only a fraction of many copies of the same peptide are modified. Thus spectra of both modified and unmodified versions of the peptide may be seen in the dataset. By aligning the de novo sequencing results of the pair of spectra of the modified and unmodified peptides, variable PTMs can be discovered. The working diagram of the algorithm for PTMFinder is shown in Figure 1.

1. PEAKS™ software [1] is used to perform de novo sequencing on all spectra with no PTM allowed. Thus, for the modified peptides, de novo sequencing can only provide partially correct sequence tags from the spectra.

2. Sequences from de novo sequencing results are aligned to find the pairs of the modified and unmodified spectra. Each pair results in a precursor mass difference (△m) and an alignment score computed from retention time and the local confidence of de novo sequencing results. We observe that for most of the common PTMs, the retention times of the modified and unmodified peptides do not differ significantly.

3. The precursor mass differences are mapped with PTM databases to generate a list of variable PTM candidates.

4. PEAKS™ database search is used to validate the candidates. If a variable PTM is true positive, then by allowing the PTM in PEAKS database search, both the modified and unmodified peptide should be identified.

5. A list of confirmed variable PTMs with confidence scores are reported.

Results

A sample was prepared to test the validity and efficiency of the method. A complex protein mixture from C. elegans was digested with trypsin and alkylated with iodoacetamide. All spectra were acquired on a LTQ Orbitrap XL. The peptide mixture (2 ul injected) was separated via Surveyor LC equipped with MicroAS autosampler (Thermo Fisher Scientific) using a reversed phase peptide trap (100 um inner diameter, 2 cm length) and a reversed phase analytical column (75 um inner diameter, 10 cm length, 3 µm particle size, both Nanoseparations, NL), at a flow rate of 250 nl/min. A gradient of 5 - 30% acetonitrile in 90 minutes was used.

7743 MS/MS spectra were obtained along with 4805 survey scans. PEAKS™ Studio is used for de novo sequencing with carbamidomethylation on Cys. The resulted de novo sequences are aligned with respect to retention time (±4 minutes) and sequence similarity. 3200 spectrum pairs were obtained, and then a list of mass differences with the frequency was derived from the spectrum pairs. By mapping with PTM databases, five significant variable PTM candidates were discovered (Table 1). PEAKS™ database search was used for the validation of candidates by turning them on. As a result, four of the five PTMs (except for sodium) in Table 1 were confirmed as true variable PTMs.

Figure 1. Working diagram of the algorithm for PTMFinder

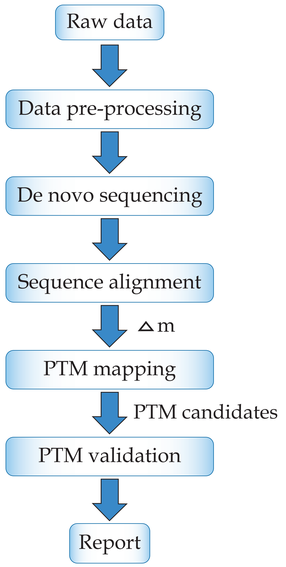
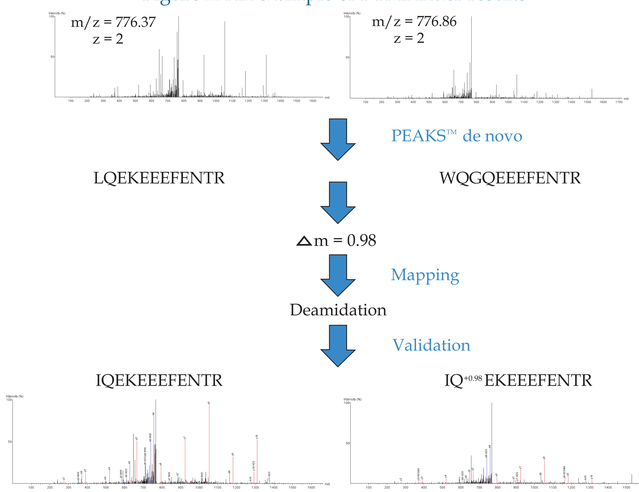


Table 1. Outcome for each round of search

△m	Score	PTM	Frequency
1	90	deamidation	120
28	67	dimethylation	16
74	60	glycerol ester	16
22	58	sodium	16
16	56	oxidation	15

An example of the PTMFinder result for two spectra is shown in Figure 2. The first spectrum m/z = 776.37 and retention time = 16.53 min has a de novo sequence of LQEKEEEFENTR. The second spectrum m/z = 776.86 and retention time = 17.44 min has the sequence of WQGQEEEFENTR. Both spectra have charge 2. A very confident mass difference △m = 0.98 was obtained. The corresponding PTM is deamidation. PEAKS™ database search confirmed that deamidation is a true variable PTM. The peptide for the first spectrum is IQEKEEEFENTR and IQ$^{+0.98}$EKEEEFENTR for the second spectrum.

Figure 2. An example of PTMFinder results



Conclusion

The experiment results showed that our PTMFinder can discover variable PTM efficiently and effectively.

Reference

B. Ma, et al., Rapid Communications in Mass Spectrometry, 17(20):2337-2342 2003.