Building and Searching Predicted Spectral Libraries for Identification of Protein Post-translational Modifications

By

Yingwei HU

Abstract

Post-translational modification (PTM) is a key step in protein biosynthesis, critical for the correct trafficking and function of the protein. However, the high-throughput identification of additional biochemical functional groups, such as phosphate and glycans, involved in PTM remains a challenge in proteomics. The traditional search strategy based on protein sequence database is time-consuming and prone to false positives because of its exponentially increased search space and incomplete theoretical fragmentation model.

Due to its advantages in efficiency and sensitivity, spectral library searching is a promising alternative to conventional sequence database searching. Our work aims to facilitate PTM identification in the spectral library search approach. In particular, we first applied the approach on two important and challenging PTMs, phosphorylation and glycosylation, and extended the method to other modifications. In phosphorylated peptide identification, the largest collision-induced dissociation (CID) tandem mass (MS2) spectral libraries of phosphorylated peptides in human and other model organisms to date have been built in an automatic platform consist of multiple state-of-art search engines and site-localization tools with strict quality control. Spectral library searching using this library significantly outperforms existing methods for detecting phosphosites in a variety of datasets. In glycopeptide identification, a spectral library searching method was developed to identify intact N-linked glycopeptides from the MS2 spectra, based upon an existing spectrum prediction tool, MassAnalyzer (Zhang, Z., Anal. Chem. 2010), to account for the special fragmentation patterns of glycopeptides. We evaluated the scoring functions, developed methods to analyze ambiguous candidates and clustered the predicted spectral library to reduce the searching cost. A novel query decoy strategy was further applied to estimate the false discovery rate (FDR) of glycopeptides. The spectral library searching strategy was successfully verified in the searching of standard N-linked glycoproteins.

Date: 23 January 2015 (Friday)
Time: 15:00
Venue: Room 5562 (Lift 27-28)

Examination Committee:
Prof. David LAM, Chairman
Prof. Henry LAM, Supervisor
Prof. Fei SUN, Prof. Tom LUO, CBME
Prof. Weichuan YU, ECE
Prof. Terence POON (External), University of Macau

- ALL ARE WELCOME -