MOD^i : A Scalable Approach to Identify Multiple Peptide Modifications from Tandem Mass Spectra

Sangtae Kim[1], Seungjin Na[2], Heejin Park[3], Kong-Joo Lee[4], and Eunok Paek[2], [1]Dept. of Computer Science and Engineering, University of California, San Diego, USA, [2] Dept. of Mechanical and Information Engineering, University of Seoul, Seoul, South Korea, [3]College of Information and Communications, Hanyang University, Seoul, South Korea, [4]Center for Cell Signaling Research, Division of Molecular Life Sciences & College of Pharmacy, Ewha Womans University, Seoul, South Korea.

We present an efficient method called MOD^i to interpret a tandem mass spectrum of a peptide having multiple post-translational modifications (PTMs). The proposed method is scalable in that it performs well when more than a hundred modification types are considered and the number of potential PTMs in a peptide gets bigger. This method consists of four steps: peak selection, *de novo* sequencing, database search for candidate peptides, and identification of PTMs. In the peak selection step, we select peaks with relatively high intensities. In the *de novo* sequencing step, we perform partial *de novo* sequencing on the selected peaks to identify all the tags (partial amino acid sequences that do not contain PTMs), and then make a list of identified tags together with locations of peaks used to generate the tag. In the database search step, we search the peptide database for peptide sequences that have the identified tags, and then we align the tags to each candidate peptide sequence to generate a chain of tags and in-between gaps. When aligning the tags, we use the location information of the selected peaks. It must be noted that the peptide database we use does not contain any modification information, thus meeting the scalability requirement. Finally, in the PTM identification step, for each gap, we identify PTMs that best explain the gap. We first enumerate candidate PTMs that explain the size of each gap, and then select the best one by comparing the theoretical spectra generated by the candidate PTMs with a partial tandem mass spectrum. By taking a hybrid approach, combining de novo sequencing and DB search, peptides with three or more modifications are identified effectively. MOD^i is equipped with a graphical tool called MassPective developed to display MOD^i's output in a user-friendly manner and helps users understand MOD^i's output quickly.

Supported by MOST 21C Functional Proteomics Center Project (FPR03B3-04-110, FPR02A7-32-110) and by KOSEF through the Center for Cell Signaling Research (CCSR) at Ewha Womans University.