Fragmentation Pattern-based Scoring for Peptide Database Search

Junhee Hong, Seungjin Na and Eunok Paek

Department of Computer Science, Hanyang University, Seoul 04763, Republic of Korea

For peptide identification by sequence database search in mass spectrometry-based proteomics, an experimental tandem mass spectrum is compared with a theoretical spectrum. Most algorithms make the intensities of all mass-to-charge peaks equal when constructing a theoretical spectrum. Regardless of the frequency or intensity of a particular fragment ion type observed in a spectrum, most existing algorithms unconditionally increase the match score if there is a high intensity peak at a theoretical m/z position. For instance, while b1 ions are seldom observed in tandem mass spectra, if an experimental spectrum has a high intensity peak at the b1 ion position the score gets higher, which can be deleterious. That is, using a theoretical spectrum in which all peaks have the same intensity can result in undesirable match scores between a peptide sequence and an experimental spectrum, and consequently it can disturb the identification of high-quality peptide-spectrum matches (PSMs). In this study, we analyze the fragmentation statistics of tandem mass spectra and use it to calculate the similarity score between a peptide sequence and an experimental spectrum. The fragment ion types include b-ion and y-ion, which are dominant fragment ions in an MS2 spectrum. Comet, a database search tool that uses XCorr as a similarity score, is used to evaluate the performance by applying this statistical pattern. The results show that the fragmentation pattern-based scoring improves the number of peptide-spectrum matches by about 5%.