

AN ITERATIVE ALGORITHM TO QUANTIFY THE FACTORS INFLUENCING PEPTIDE FRAGMENTATION FOR MS/MS SPECTRUM

Chungong Yu†, Yu Lin†, Shiwei Sun, Jinjin Cai, Jingfen Zhang
Institute of Computing Technology,
Chinese Academy of Sciences, Beijing 100080, China

Zhuo Zhang, Runsheng Chen
Institute of Biophysics,
Chinese Academy of Sciences, Beijing 100035, China

Dongbo Bu
David R. Cheriton School of Computer Science University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

In protein identification through MS/MS spectrum, it is critical to accurately predict theoretical spectrum from a peptide sequence, which heavily depends on a quantitative understanding of the fragmentation process. To date, widely used database searching methods adopted a simple statistical model to predict theoretical spectrum, yielding a spectrum deviating significantly from the practical spectrum for some peptides and therefore preventing automated positive identification. Here, in order to derive an improved predicting model, we proposed a novel method to automatically learn the factors influencing fragmentation from a training set of MS/MS spectra. In this method, the determining of factors is converted into an optimization problem to minimize an objective function that measures the distance between experimental spectrum and theoretical one. Then, an iterative algorithm was proposed to minimize the non-linear objective function. We implemented the methods and tested them on experimental data. The examination of 1451 spectra is in good agreement with some known knowledge about peptide fragmentation, such as the tendency of cleavage towards the middle of peptide, and Pro's preference of N-terminal cleavage. Moreover, on a testing set containing 1425 spectra, comparison between predicted and practical spectra generates a median correlation of 0.759, showing this method's ability to predict a "realistic" spectrum. The results in this paper help to an accurate identification of protein through both database searching and de novo methods.