Proteomics-2013
OMICS Group Conferences
Accelerating Scientific Discovery
3rd International Conference on
Proteomics & Bioinformatics
July 15-17, 2013 Courtyard by Marriott Philadelphia Downtown, USA

FMP: A web implementation for robust peptide mass fingerprinting using feature-matching pattern-based support vector machines

Youyuan Li
East China University of Science and Technology, China

Abstract

Peptide Mass Fingerprinting (PMF) plays an irreplaceable role in nowadays tandem proteomics due to its higher sample throughput, higher level of specificity for single peptides and lower level of sensitivity to unsuspected post-translational modifications compared to MS/MS. The PMF method would become more attractive if we could improve the accuracy of protein identification. With this motivation, we have proposed and evaluated a feature-matching based uniform approach using support vector machines (SVMs) to incorporate individual concepts and conclusions for accurate PMF. The SVMs approach focused on the inherent attributes and critical issues of theoretical spectrum (peptides), experimental spectrum (peaks) and spectrum (masses) alignment. The experimental peak intensity was introduced to the algorithm. An optimal SVMs model with 491 out of 35,640 feature-matching patterns outperformed Mascot, MS-Fit, ProFound and Aldente with a high-performance evaluation on a standard PMF set of 225 items. Now, the approach is extended with a web server, FMP, to identify protein from MS[1] data. The web implementation contains several features: (i) a local secondary database PUD (Peptide Uniqueness Database) have been constructed to provide each theoretical peptide's sequence and mass uniqueness to generate SVMs features; (ii) a model named Matched Peak Intensity Redistribution (MPIR) was used to handle mass modification type (fixed and variable) and cleavage type(proper, theoretical missed and random missed) to recalculate peak intensity for each matched theoretical peptide;(iii) 17 select SVMs features are firstly calculated in crude ranking procedure to efficiently reduce the number of candidate proteins from ten thousands to tens; (iv) robust protein prediction by a set of 491 selective and well-evaluated SVMs features; (v) dynamical interface to easily monitor the identification pipeline; and (vi) double prediction tags and probabilities plus detailed statistics to describe protein identification result. The FMP server is freely available at http://bioinformatics.datawisdom.net/fmp.

Biography

Youyuan Li completed his Ph.D in Biochemical from East China University of Science and Technology (ECUST). His research interest is application of proteomics and bioinformatics in industrial bio-processes optimization. In 2002, Li joined the faculty at School of Biotechnology, ECUST. He teaches courses ranging from Bioinformatics to Fermentation Engineering.