

BIOINFORMATICS METHODS FOR PROTEIN IDENTIFICATION
USING PEPTIDE MASS FINGERPRINTING DATA

Zhao Song

Dr. Dong Xu, Dissertation Supervisor

ABSTRACT

Protein identification using mass spectrometry is an important yet partially solved problem in the study of proteomics during the post-genomic era. The major techniques used in mass spectrometry are Peptide Mass Fingerprinting (PMF) and Tandem mass spectrometry (MS/MS). PMF is faster and economical compared with MS/MS and widely applicable in many fields. Our work focus on the method development for protein identification using PMF data and this work covers three subjects: (1) Protein Identification scoring function development: we developed the Probability Based Scoring Function (PBSF) which is used to quantify the degree of match between PMF data and candidate protein. The derived score is used to rank the protein and predict the identification. (2) Confidence Assessment development: scoring function may lead to false positive identification since the top hit from a database search may not be the target protein. In addition, the identification scores assigned singly by a scoring function (raw scores) are not normalized. Therefore, the ranking based on raw scores may be biased. To address the above issue, we have developed a statistical model to evaluate the confidence of the raw score and to improve the ranking of proteins for identification. (3) Software development: we implemented our computational methods in an open source package “ProteinDecision” which is freely available upon request.