Improved Data Mining by Using TPP-based Analysis Workflows for Searching MS/MS Data

Author(s):
Quandt, A.; Malström, L.; Lam, Henry; Shteynberg, D.; Aebersold, Ruedi

Publication date:
2009-08

Permanent link:
https://doi.org/10.3929/ethz-b-000085708

Originally published in:
Molecular & Cellular Proteomics 8(8, S1), https://doi.org/10.1074/mcp.S800608-MCP200

A.15

Improved Data Mining by Using TPP-based Analysis Workflows for Searching MS/MS Data

A. Quandt[1], L. Malstroem[1], H. Lam[2], D. Shteynberg[3], and R. Aebersold[1]

[1]Institute for Molecular Systems Biology, ETH Zurich, Switzerland; [2]Department of Chemical and Biomolcular Engineering, The Hong Kong University of Science and Technology, Kowloon, Hong Kong; [3]Institute for Systems Biology, Seattle, WA

The identification and characterisation of peptides from tandem mass spectrometry (MS/MS) data represents a critical aspect of proteomics. In the past, many software packages have been developed to tackle this problem. Beside the development of new analysis tools, recent publications describe also the pipelining of different search programs to increase the identification rate. Unfortunately, it remains still unclear for the practical user when to apply which software or parameter set to retrieve optimal results. Most people rely on the usage of a identification software what consequently often leads to a significant proportion of the experimental spectra which are not going to be identified. Hence, the usage of different tools and parameter combinations is crucial but seldom approached in reality. Among other reasons such as manual result validation and the handling of various data formats, the main problem still remains in the automated combination of different identification tools. Here, we present a workflow approach which is based on the TransProteomicPipeline (Keller *et al.*, 2005; Pedrioli *et al.*, 2004) and combines multiple search tools and strategies with the result to retrieve a more complete list of peptide identifications and a higher protein coverage. For us, a workflow is the combination of various identification tools and search strategies in parallel and/or in sequential order. In the example workflow, we are going to present here, two classical search engines (X!Tandem and OMSSA) have been combined with a spectral library search search (SpectraST). In fact, we use the output of the first two search tools to dynamically create a spectral library which is searched afterwards. To compare the performance of our workflow with the results of the individual search engines, we used the 18-protein-mix dataset, which has been especially created to benchmark different search tools and pipelines. On these initial dataset, we are able to show that the combination of various identification tools in so-called workflows leads to an increased trust into the results by lowering the level of accepted false positives identifications and a significant increased protein coverage leading to a more reliable search results.

A.16

Microwave-assisted Phosphoproteomics

P. Liu and W. Sandoval

Genentech, Inc., S. San Francisco, CA

Reversible protein phosphorylation controls a multitude of important biological functions. Elucidation of the exact site of phosphorylation is often necessary to further understand the intricate mechanisms involved in intracellular sites. We have investigated the use of microwave enhanced tryptic digestions on phosphopeptide recovery. We have incorporated the use of stable isotopically labeled peptides to quantitate differences observed. In addition we have quantitatively evaluated the use of stabilizing chemical modification of phosphopeptides and phosphoproteins for subsequent analysis by mass spectrometry or Edman degradation. Finally we have applied our findings to a global phosphoproteomic analysis of peroxide treated cell lysate.

A.17

Proteome Analysis of Apoplastic Proteins in Rice Shoot Respond to Salt Stress

Y. Song[1], A. L. Burlingame[2], and Y. Guo[1,2]

[1]Institute of Molecular and Cell Biology, Hebei Normal University, Shijiazhuang, China; [2]Mass Spectrometry Facility, Department of Pharmaceutical Chemistry, University of California, San Francisco, CA

Plants have evolved sophisticated systems to cope with the adverse environmental conditions such as cold, drought and salinity due to their sessile nature. Although a lot of stress response networks have been proposed, the roles of plant apoplast have been obviously ignored in plant stress response. To investigate the role of apoplastic proteins in the salt-stress response, 10-day-old rice plants were treated with 200 mM NaCl for 1, 6 or 12 hours, and the soluble apoplast proteins were extracted and for differential analysis compared to untreated controls by 2-D DIGE saturation labeling techniques. 122 significant changed (1-ANOVA p-value <0.05) spots are identification by LC-MS/MS, and 117 spots representing 69 proteins have been identified. Of these, 37 proteins are apoplastic proteins according to bioinformatic analysis. These proteins are mostly involved in carbohydrate metabolism, oxido-reduction, protein processing and degradation. According to the results of functional categories and cluster analysis, a stress response model of apoplastic proteins has been proposed. These dates indicate that apoplast is an important portion in plant stress signal reception and response.

LC-MS/MS data was provided by the Bio-Organic Biomedical Mass Spectrometry Resource at UCSF (A. L. Burlingame, Director) supported by the Biomedical Research Technology Program of the NIH National Center for Research Resources, NIH NCRR P41RR001614 and NIH NCRR RR019934.