Software and Algorithm for Differential Characterization of Post-Translational Modifications

Vivek N. Bhatia, David H. Perlman, Catherine E. Costello, Mark E. McComb
Cardiovascular Proteomics Center, Boston University School of Medicine, Boston, MA

PROJECT AIMS

Protein post-translational modifications (PTMs) play a critical role in normal cellular function. Changes in PTMs are known to affect the mechanisms of disease, thus studying PTM identification and localization is a significant priority in contemporary biomedical research. To facilitate differential PTM characterization, we have developed STRAP PTM (Software Tool for Rapid Annotation of Proteins, PTMs), for automated analysis of global changes in PTMs on proteins.

STRAP PTM

- collates data from large-scale proteomics experiments
- performs multi-way differential analyses by scoring and ranking proteins, peptides and PTM identifications
- provides a visual means to identify important PTMs that are differentially observed between multiple sample groups.
- uses a unique algorithm to find differential PTMs

BACKGROUND

MS-based proteomics typically yields thousands of protein and post-translational modification (PTM) identifications in an experiment. Post-identification, information about PTMs must be interpreted to gain insight into their biological roles in disease. While there is a great interest in PTMs due to their significant roles in disease, there are few tools for differentially comparing PTMs across multiple different sample groups. An example workflow is shown in Figure 1.

THE PROTEOMICS WORKFLOW

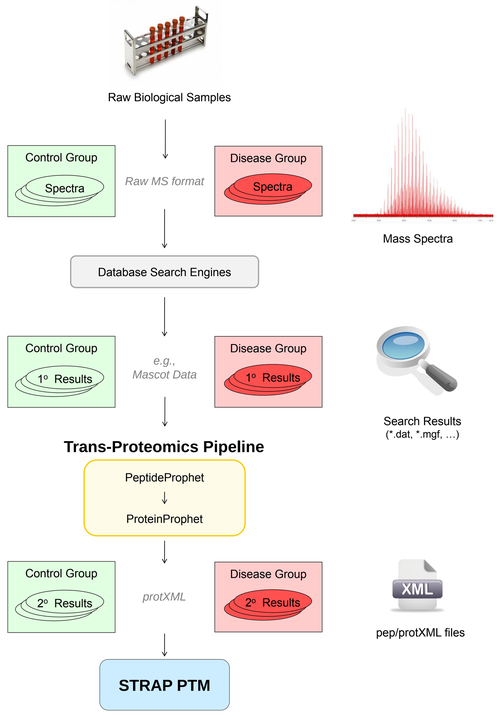

Figure 1. Overview of a proteomics data processing scheme. Each biological and technical replicate within each sample group is analyzed by an LC-MS/MS experiment, producing multiple MS/MS spectra per replicate. The resulting raw data are then submitted to search engines for peptide and protein identification. These search results are analyzed and grouped to produce a consolidated summary of the search results. In this case, we used the TPP, an open- source framework, to obtain these results as protXML files. The resulting protXML files were then differentially analyzed by STRAP PTM.

METHODS

STRAP PTM is written in C# and runs on Microsoft Windows XP or higher with version 4.0 of the Microsoft.NET Framework. The application's code is written to import all parsed data into an object model. This data can be queried for alternate applications, a feature which makes STRAP PTM easily extensible.

STRAP PTM uses protXML as the input format. ProtXML is an open format established by the Institute for Systems Biology (ISB). These files may be produced by the Trans-Proteomics Pipeline (TPP) from the ISB, or by using commercial tools such as Proteome Discoverer from Thermo Scientific. After importing the protXML files, STRAP PTM compiles the peptide hits and lists them by protein, rendering information in a tabular format. Each protein entry is linked to the associated PTMs, peptides, sample or group origin.

PTM SCORING

We have explored several scoring algorithms to rank the most significant differentially observed PTMs in a data set. In one such scheme, PTMs on a particular peptide with a delta mass, δ, are scored and ranked according to the following scoring function, S:

$$S(\delta) = 100\ (\bar{u} + 0.1)\left[(1 - F_d) + \frac{p(\text{PTM in protein})}{p(\text{PTM in all proteins})}\right]$$

ū = average of PeptideProphet peptide scores for unmodified forms of the peptide in any sample group, offset by 0.1 to minimize bias against peptides without unmodified forms

Enrichment of a particular PTM within a protein against the background of all proteins within an experiment

F_d = the differential PTM fraction

RESULTS

PTM COUNTING VS. LABEL-FREE QUANTIFICATION

STRAP PTM was used to analyze data from a label-free, LC-MS-based oxidant sensitivity study, in which a target protein containing several suspected redox-sensitive residues was subjected *in vitro* to treatment with increasing concentrations of peroxinitrite (PN). In this example, we report on the ligand fragment, CD40L_HUMAN, which is known to be sensitive to oxidation.

The identification of the resulting oxidized species was determined by trypsin digestion, LC-MS/MS and PTM-tolerant database searching. Relative abundances were quantified by peptide ion abundances or by PTM counting. Among the different modifications detected, we observed various forms of cysteine oxidation (Figure 2) and nitration. Nitration showed distinct and differing degrees of formation with respect to PN concentration, as depicted in the label-free analysis shown in Figure 3. We observed increases in both nitrotyrosine and di-nitrotyrosine formation as PN concentration was increased.

Nitration and other PTMs were rapidly elucidated using STRAP PTM, as shown in Figure 4. The findings were comparable to the label-free LCMS results.

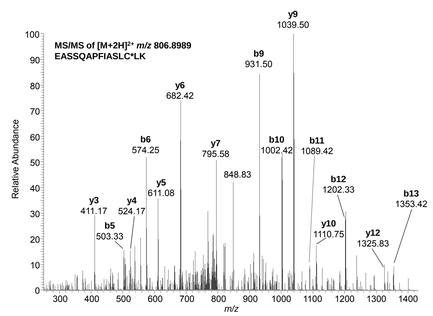

Figure 2. Tandem mass spectrum and assignments of a redox PTM on CD40L_HUMAN treated with PN. We observed cysteine sulfonic acid formation on PN-treated CD40L at C194, indicated by the b13 and y3 ions. The b14 ion (not shown) was observed at *m/z* 1466.58.

RESULTS II

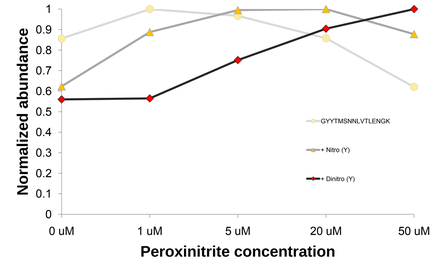

Figure 3. PTM expression trend for nitration and di-nitration of the CD40L peptide, GYYTMSNNLVTLENGK. The amount of the unmodified peptide decreased with increasing PN concentration. The singly nitrated version of this peptide increased as CD40L was exposed to lower concentrations of PN, but decreased in abundance at higher concentrations of PN as it was converted to the di-nitrotyrosine form. The di-nitrotyrosine-bearing peptide increased with increasing PN. The abundance values, obtained from Progenesis LCMS, were normalized to the maximum value in each series.

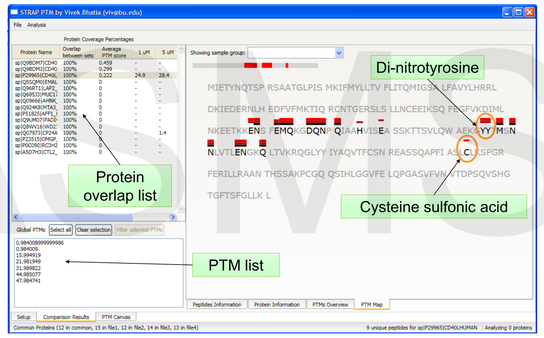

Figure 4. Differential PTM map of CD40L_HUMAN. Red bars are placed above modified amino acids; the brighter reds correspond to increased observation of PTMs associated with higher PN concentration. PTMs of interest are indicated.

ENHANCED GLOBAL PTM DATA ANALYSIS

To illustrate STRAP PTM's global PTM analysis capabilities, we utilized data from an experiment in which ocular cells were exposed to different treatment conditions to study the biochemistry of wound-healing. The resulting analysis allowed for the rapid selection of biologically significant PTMs within the complicated proteome involved in the context of ocular wound healing. Results from this study were presented in detail by A. Kehasse *et al.*, MP484.

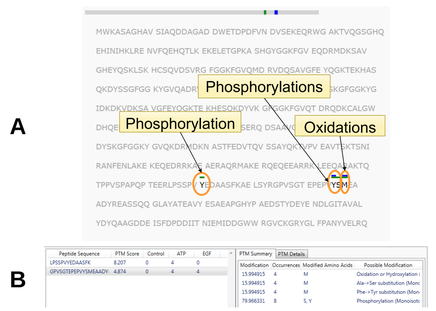

Figure 5. Example of PTM assignments using STRAP PTM: analysis of an ocular cell data set. **A**. PTM map of SRC8_HUMAN, a protein with a high PTM score relative to the proteins in the data set. The primary sequence is annotated with locations of phosphorylation and oxidation modifications. Unlike the control set, the cells exposed to treatment 1 (green) and treatment 2 (blue) had differential PTMs. **B**. STRAP PTM's table of individual peptides and associated modifications of SRC8_HUMAN.

RESULTS III

Following LC-MS/MS, RAW files were processed through Proteome Discoverer (Thermo) to obtain MGF files, which were used in the MASCOT database search. Resulting .dat files were processed through the TPP to produce pepXML and subsequent protXML files, which are imported into STRAP PTM for differential PTM analysis.

The most striking feature of this processing, shown in Figure 6, was that in under 5 minutes, STRAP PTM afforded a list of overlapping proteins observed between conditions, PTM maps of these proteins, and PTM scores to rank these proteins based on their differential expression of PTMs. The experimental data originated from 3 experimental groups, having 3 samples each. The LC-MS/MS analysis yielded a total of 6034 spectra, and 313 PTM assignments overall.

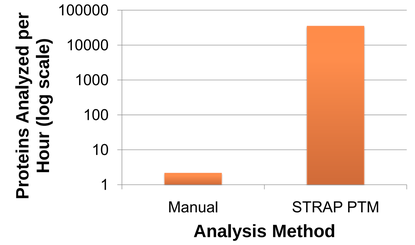

Figure 6. Comparison between the time required for manual analysis and automated analysis by STRAP PTM for the ocular injury model data. The analysis carried out involved a three-way comparison of three experimental conditions, which included roughly 3500 proteins. The manual expert analysis was not systematic as it was biased towards proteins recognized by the expert to be involved in wound-healing pathways. In addition to being faster and systematic, the STRAP PTM analysis is blinded and un-biased. A summary of all results from this study were presented by A. Kehasse *et al.*, MP484.

CONCLUSIONS

STRAP PTM performs differential comparison between files and/or sample groups, yielding differentially observed PTMs. One of the products of this comparison is a PTM map that overlays PTMs on the primary sequences of the parent proteins. This visualization aids in data interpretation, as it conveniently illustrates experimentally observed differences in modifications among several sample groups. Moreover, the scoring function built into STRAP PTM emphasizes PTMs of interest. The use of PTM maps and the scoring function together allow for global, proteome-wide PTM analysis in a manner more systematic and faster than manual analysis by an expert.

STRAP PTM accelerates differential PTM analysis that is otherwise extremely laborious when performed manually on complex data sets. By using MS search results obtained from the open standard protXML format, the user is not restricted to a specific search strategy or type of instrumentation.

Here, we demonstrated that PTMs detected by STRAP PTM via differential analysis of CD40L oxidation were comparable to those observed via label-free LC-MS. By illustrating an application to a proteome-wide analysis of a complex data set, we also demonstrated that differential PTMs may be mapped with relative ease in a substantially shortened time scale.

CURRENT AND FUTURE WORK

- Revise and validate scoring algorithm to better mimic manual analysis by an expert
- Improve visualization of data, with a more informative PTM map and alternative forms of data representation
- Improve user interface to make data manipulation and interrogation more flexible

ACKNOWLEDGMENTS

- NIH-NCRR grants P41 RR010888, S10 RR020946, S10 RR025082
- NIH-NHLBI contracts N01 HV28178 and N01 HV00239
- Special thanks to the members of the CPC and CBMS