

Overview

The Accurate Mass and Time (AMT) tag pipeline combines Liquid Chromatography (LC)-Mass Spectrometry (MS) and tandem mass spectrometry (MS/MS) experiments to provide high confidence identifications and quantitative information¹

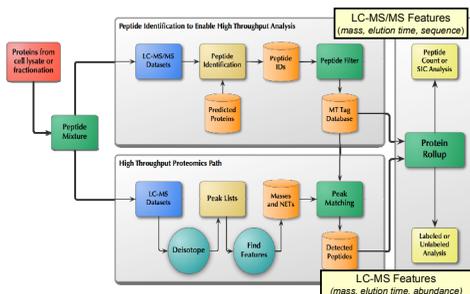


Figure 1. The AMT tag pipeline for quantitative, high throughput proteomics

Peptides and proteins are characterized and identified using sophisticated software for:

- Management of samples and datasets
- Feature discovery in LC-MS and LC-MS/MS data
- Alignment and quantitation of datasets
- Automated processing with workflows

Tools required for implementation of the AMT tag pipeline are available at: <http://ncrr.pnl.gov>
Raw and processed proteomics is available at <http://omics.pnl.gov>

Data Management

Store, track, and automatically analyze proteomics data

The Proteomics Research Information Storage and Management (PRISM) system includes a tracking database, data storage, processing nodes, and a web-based interface for user interaction

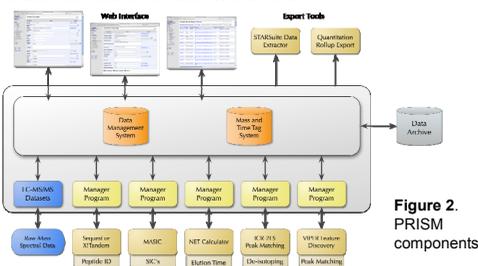


Figure 2. PRISM components

LC-MS/MS Feature Discovery

Characterize peptides to assemble AMT tag databases

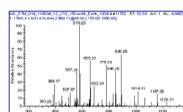


Figure 3. MS/MS fragmentation spectrum

- Samples are characterized using gradient LC separations coupled to ion trap MS/MS analyses to obtain large numbers of accurate mass and time (AMT) tags
- High-resolution hybrid LC-MS/MS instruments (e.g., the LTQ Orbitrap) typically produce low resolution MS/MS data along with high resolution precursor data.
- Fragmentation spectra are processed with SEQUEST and X!Tandem; we are evaluating OMSSA and InsPect

DeconMSn

- Software tool that determines the monoisotopic mass and charge state of each parent ion chosen for fragmentation on a hybrid LC-MS/MS instrument
- Creates spectra files (.Dta for SEQUEST or .Mgf for MASCOT or X!Tandem) that have accurate mass and charge information for each parent ion
- Allows for MS/MS searching with tighter mass tolerances, leading to lower FDR values

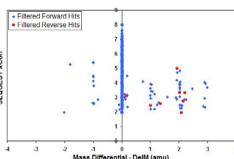


Figure 4. SEQUEST XCorr vs. mass difference (identified peptide mass vs. accurate parent ion mass)

MTDB Creator

- Application that allows external researchers to align LC-MS/MS results and create a standalone AMT tag database for use with MultiAlign

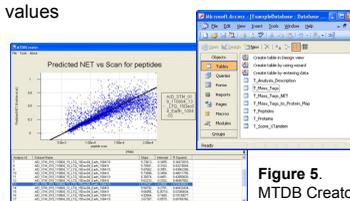


Figure 5. MTDB Creator

LC-MS Feature Discovery

Characterize features observed on high resolution MS instruments

Decon2LS

- High throughput LC-FTICR-MS analyses are used to measure intact peptide masses and elution times for various conditions
- Decon2LS software deisotopes high resolution mass spectra to generate a data file that describes the features
- A variety of algorithms are used in the deconvolution process, including noise reduction, peak detection, prediction of theoretical isotopic envelopes, and scoring functions that quantify the quality of observed signatures

- Can process data from several different vendors in addition to mzXML files
- Decon2LS has been used for automated analysis of 12,000 datasets since Feb. 2007

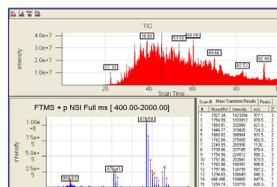


Figure 6. Decon2LS TIC and spectrum

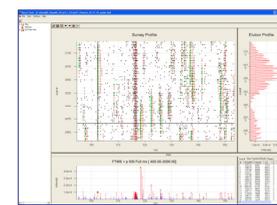


Figure 7. Two-dimensional view of features

Quantitation

Identify, quantify, and compare features across multiple datasets

MultiAlign

- Provides advanced visualization and manipulation capabilities, such as overlaid 2D plots, alignment plots, normalizations, and basic statistical comparisons

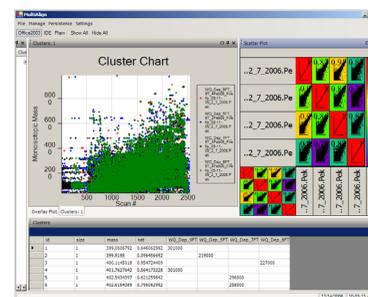


Figure 8. MultiAlign user interface

- Uses the LCMSWARP² algorithm to consolidate features from multiple LC-MS datasets into a master list

- Aligns data to an AMT tag database to identify the LC-MS features
- Alignment across datasets followed by alignment and matching to AMT tag DB results in fewer "missing values"

- Can quantify changes to both identified and unidentified features

- Able to process large numbers of datasets simultaneously to support large-scale experiments (>50 datasets)

- Can interact with MTS to automatically process datasets

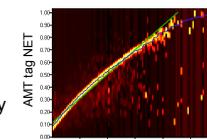


Figure 9. LCMSWarp alignment surface

DANte

- Developed to allow researchers to normalize the variations and systematic biases commonly observed in quantitative -omic results

- Allows association of each dataset with one or more factors (independent variables), which are used to appropriately weight the data during statistical analysis

- Includes a comprehensive analysis of variance (ANOVA) method that takes into account fixed and random effects, in addition to unbalanced data; this allows for statistically significant abundance changes to be discovered by analyzing abundance changes between samples

- Features an extensive array of diagnostic plots including histograms, box plots, correlation plots, and scatter plots

See also the separate poster with additional details on DANte

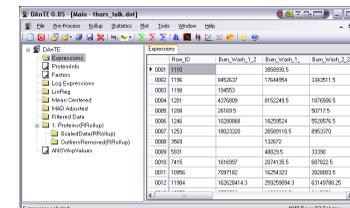


Figure 10. DANte user interface

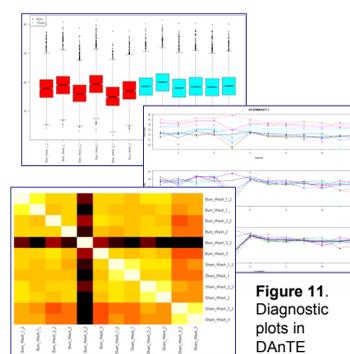


Figure 11. Diagnostic plots in DANte

Confidence Metrics

Provide measures of the quality of identified peptides and proteins



Figure 12. QC Plots

Statistical Method for Assignment of Relative Truth (SMART) score

- A probability score that combines various metrics associated with identified peptides
- Allows researchers to prioritize acceptable matches for a given false discovery rate (FDR)

See also the separate poster with details of the SMART algorithm

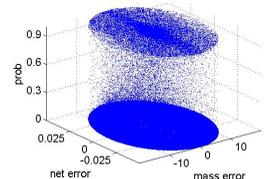


Figure 13. Distribution of components used by the SMART score

Websites

NCRR portal: <http://ncrr.pnl.gov>

- News, information, and tutorials

- Over 25 downloadable, open source software applications, including both compiled binaries and source code

Data distribution: <http://omics.pnl.gov>

- Browse by publication, organism, or sample type

- Download raw and processed data

See also the separate poster for more information

PRISMWiki

- User-editable proteomics resource for visiting scientists and staff
- Includes a knowledgebase, glossary, links, and tutorials on various aspects of the proteomics analysis process

References

1. JD Zimmer, ME Monroe, WJ Qian, and RD Smith. *Mass Spec. Rev.* 25, 450-482 (2006).
2. N. Jaitly, ME Monroe, VA Petyuk, TR Clauss, JN Adkins, and RD Smith. *Anal. Chem.*, 78 (21), 7397-7409 (2006)

