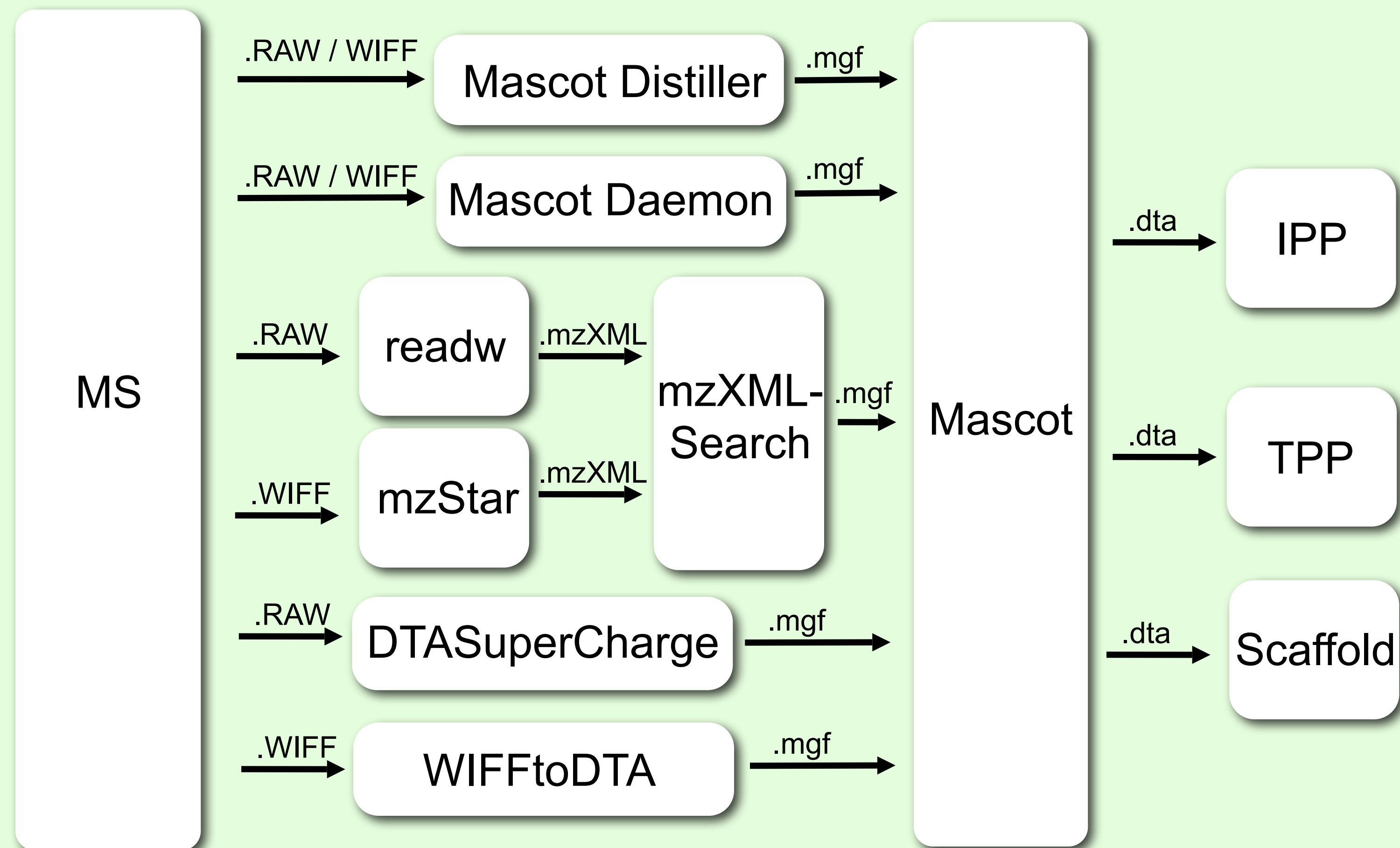


Comparison of Methods of Applying the Trans-Proteomic Pipeline to Mascot Peptide Matches

Bryan J. Prazen¹, Tom Rouwette², Leonard J. Foster³, Oliver Schilling³, Brian S. Pratt¹, Alain Doucet³, Oded Kleifeld³, Christopher M. Overall³, Erik J. Nilsson¹
¹Insilicos, Seattle, Wa; ²University of Nijmegen, Nijmegen, Netherlands; ³University of British Columbia, Vancouver BC, Canada

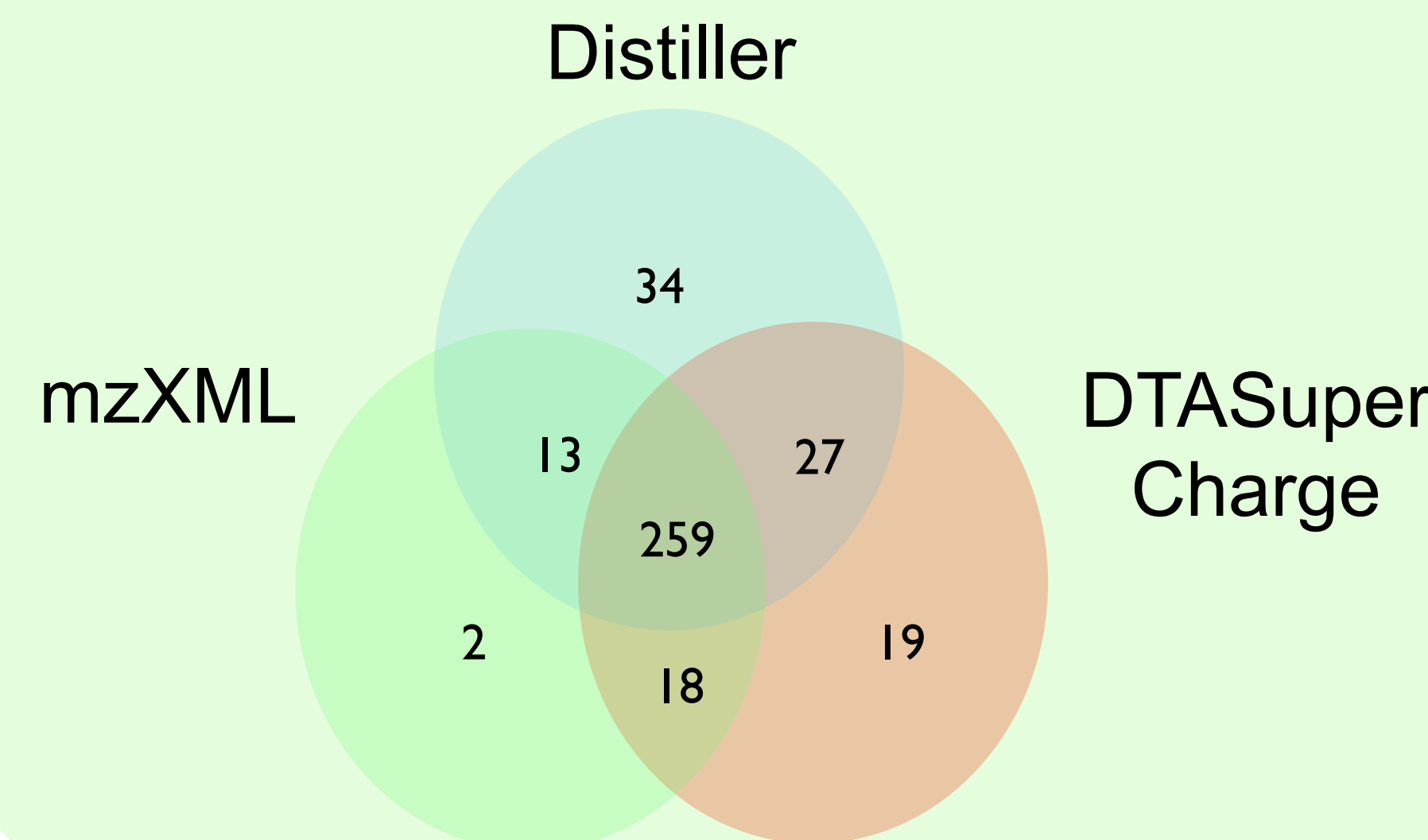
Orbitrap & QSTAR Data Processing Paths



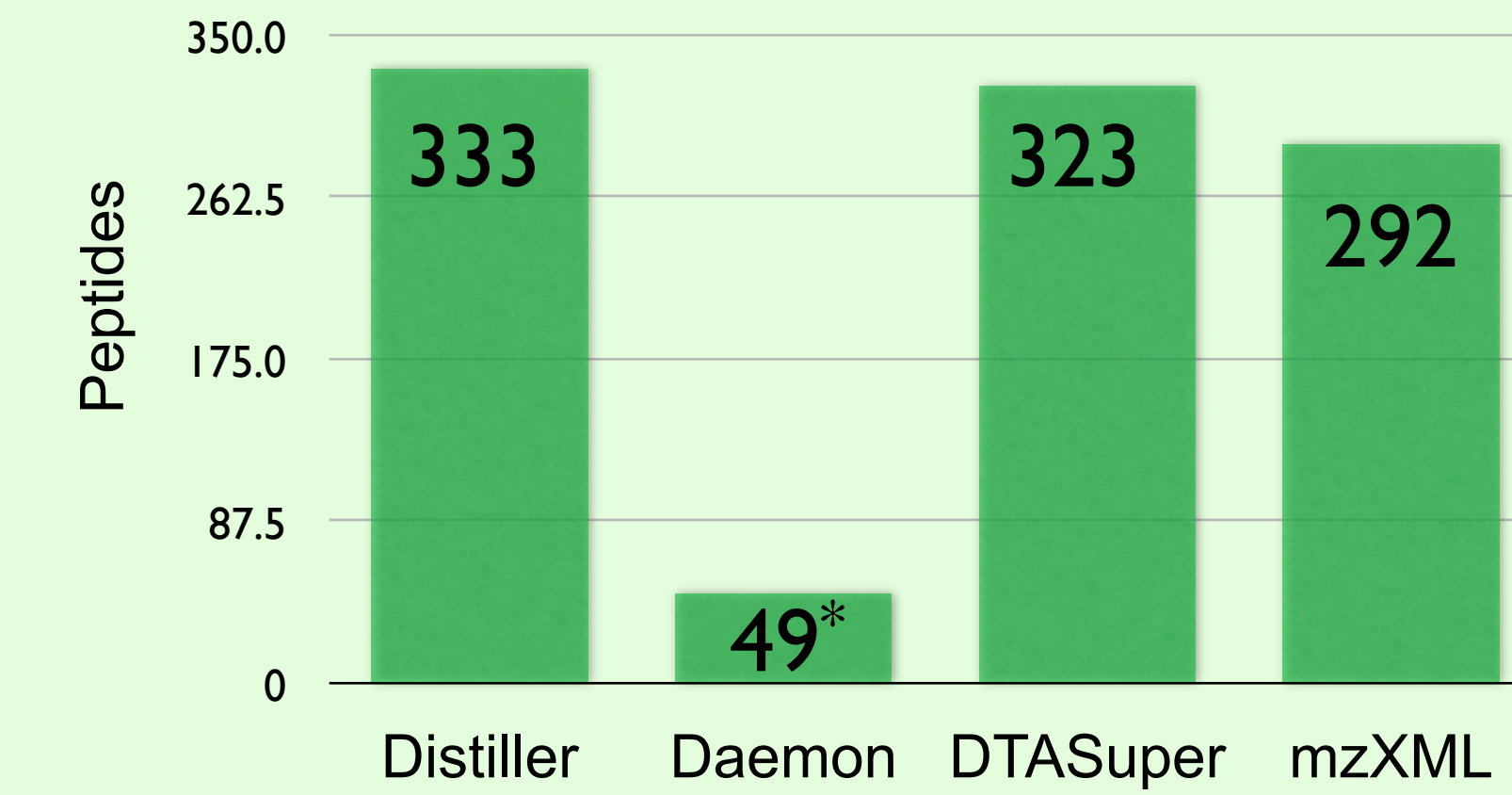
There are multiple paths available for analyzing your tandem MS data with Mascot and the trans-proteomic pipeline. This poster explores four popular methods.

Mascot Distiller processes raw data into high quality, de-isotoped peak lists by fitting each peak to a calculated isotope distribution. Mascot, Mascot Distiller and Mascot Daemon are distributed by Matrix Science. DTASuperCharge is an application for converting one or more Finnigan .RAW files to Mascot search input files that is distributed by the Center for Experimental Bioinformatics (CEBI). WIFFtoDTA is distributed by Sickmann group, Rudolf-Virchow University of Wuerzburg, Germany. The TPP is available through the Seattle Proteome Center. IPP is available through Insilicos. And Scaffold is available through Proteome Software.

Peptides Identified with Orbitrap Data by Four Data Processing Paths



A total of 372 peptides were identified with all four data processing paths. Peptides shown are Mascot peptide identifications with a ion score > 20.



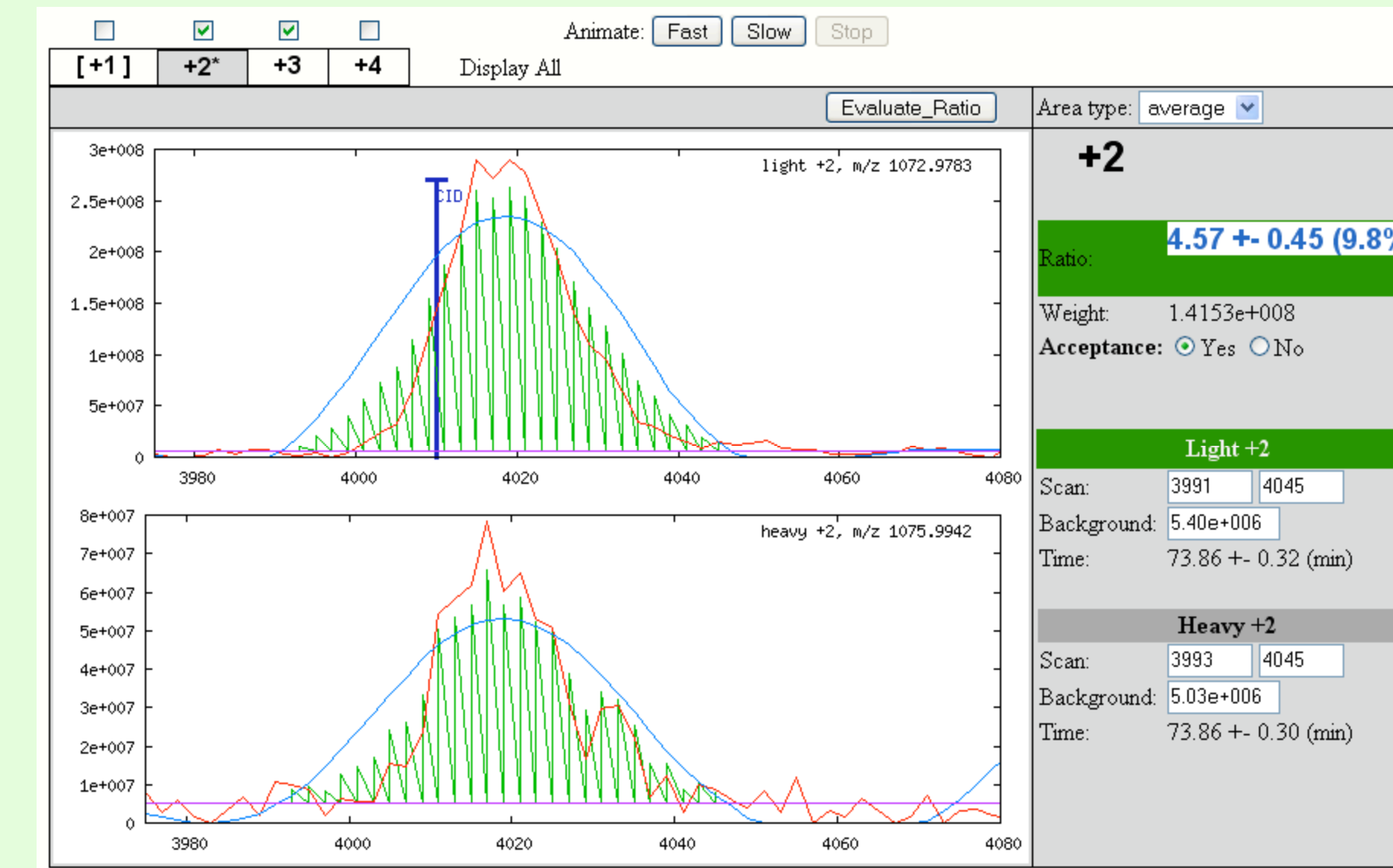
* Distribution of the peptides identified with the Mascot Daemon converter indicate that the converter may not have been able to process the entire .RAW file.

Quantitation

Recent improvements in the Insilicos distribution of the trans-Proteomics Pipeline (IPP) allow for the ICAT type quantitative analysis of all Mascot results originating from Mascot Daemon and DTASuperCharge.

.wiff file -> Mascot Daemon -> Mascot -> IPP quantitation

.RAW file -> DTASuperCharge -> Mascot -> IPP quantitation



Previously it was nearly impossible to associate the spectra from DTASuperCharge or .wiff files processed through Mascot Daemon with spectra in the mzXML file. Without an association between the Mascot results and the mzXML data, quantitation using ASAPRatio and XPRESS was impossible. Quantitation with IPP works for all known data processing paths.

ProteinProphet Performance

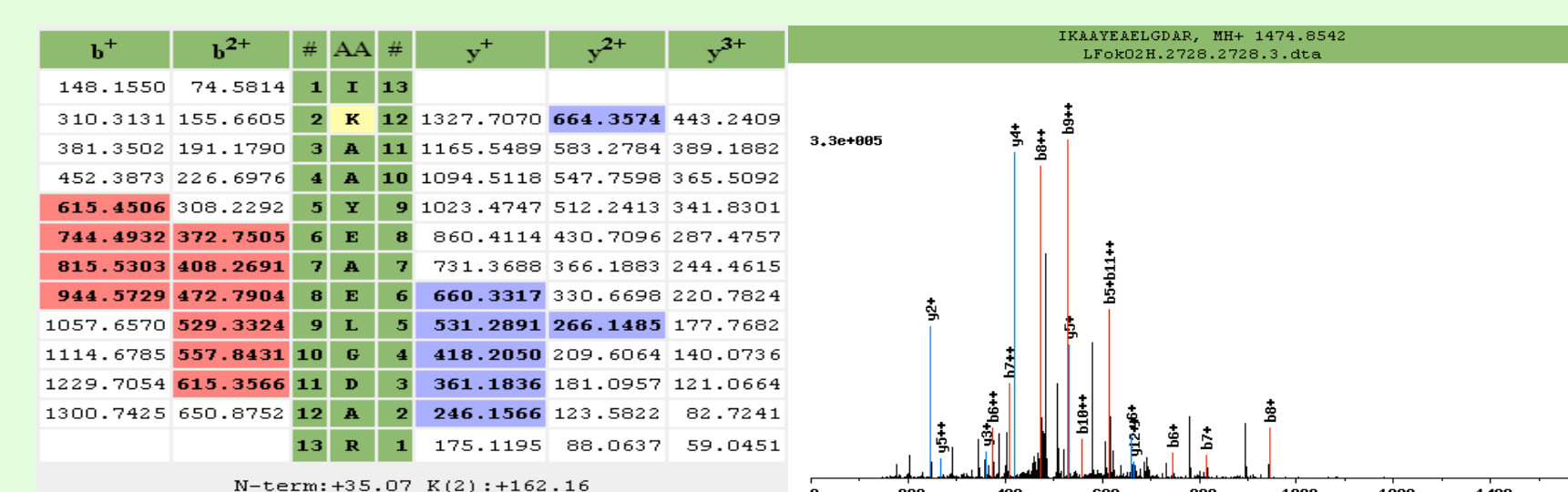
ProteinProphet Task	TPP	IPP	Improvement
461 KB pepXML	37s	7s	5x
151 MB pepXML	38m 20s	1m 49s	21x
755 MB pepXML	39hr	24m	97.5x

IPP now includes a much faster version of ProteinProphet. Insilicos and LabKey translated the Perl code for ProteinProphet to C++ and optimized the code. For very large datasets this update led to nearly a 100-fold improvement in performance.

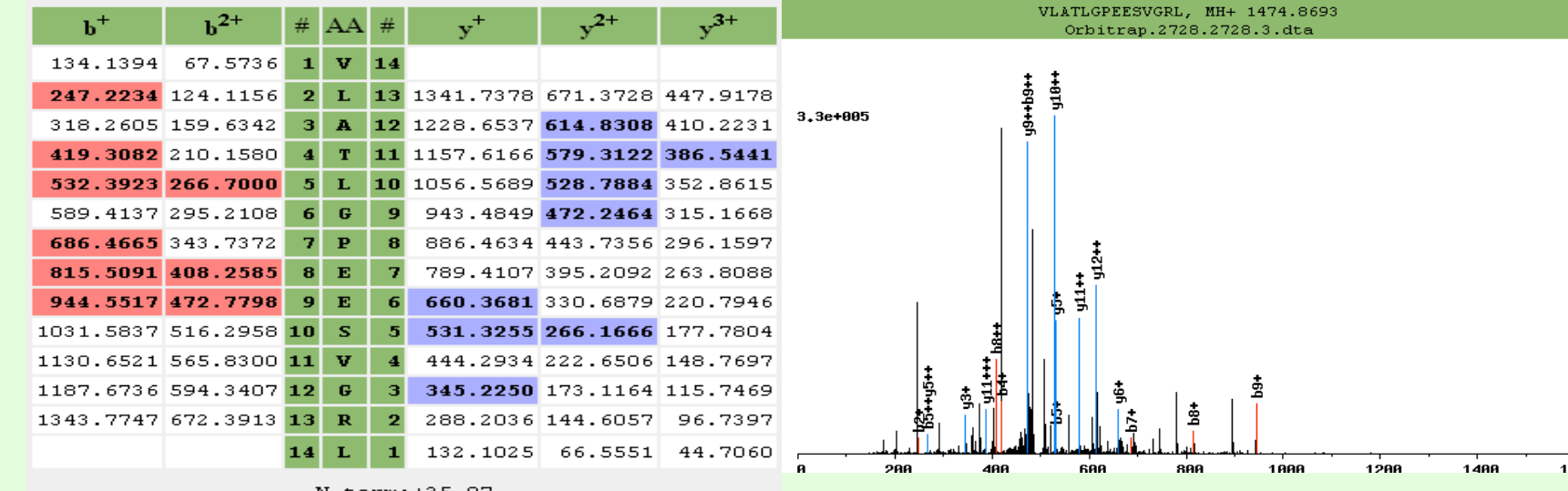
Peptide Inconsistencies Between Data Processing Paths

For the Orbitrap data analyzed in this experiment, most of the differences in identified peptides are a result of the Mascot ion scores changing. In many of the examples from this data set the score changed by 10 points. In some cases differences in identified peptides were a result of spectra assigned to different peptides. Below is an example where processing with Mascot Distiller resulted in a different match than processing through mzXML. For this dataset 5 of the 34 peptides that were unique to Distiller were assigned to different peptides during the mzXML processing path.

Mascot Distiller

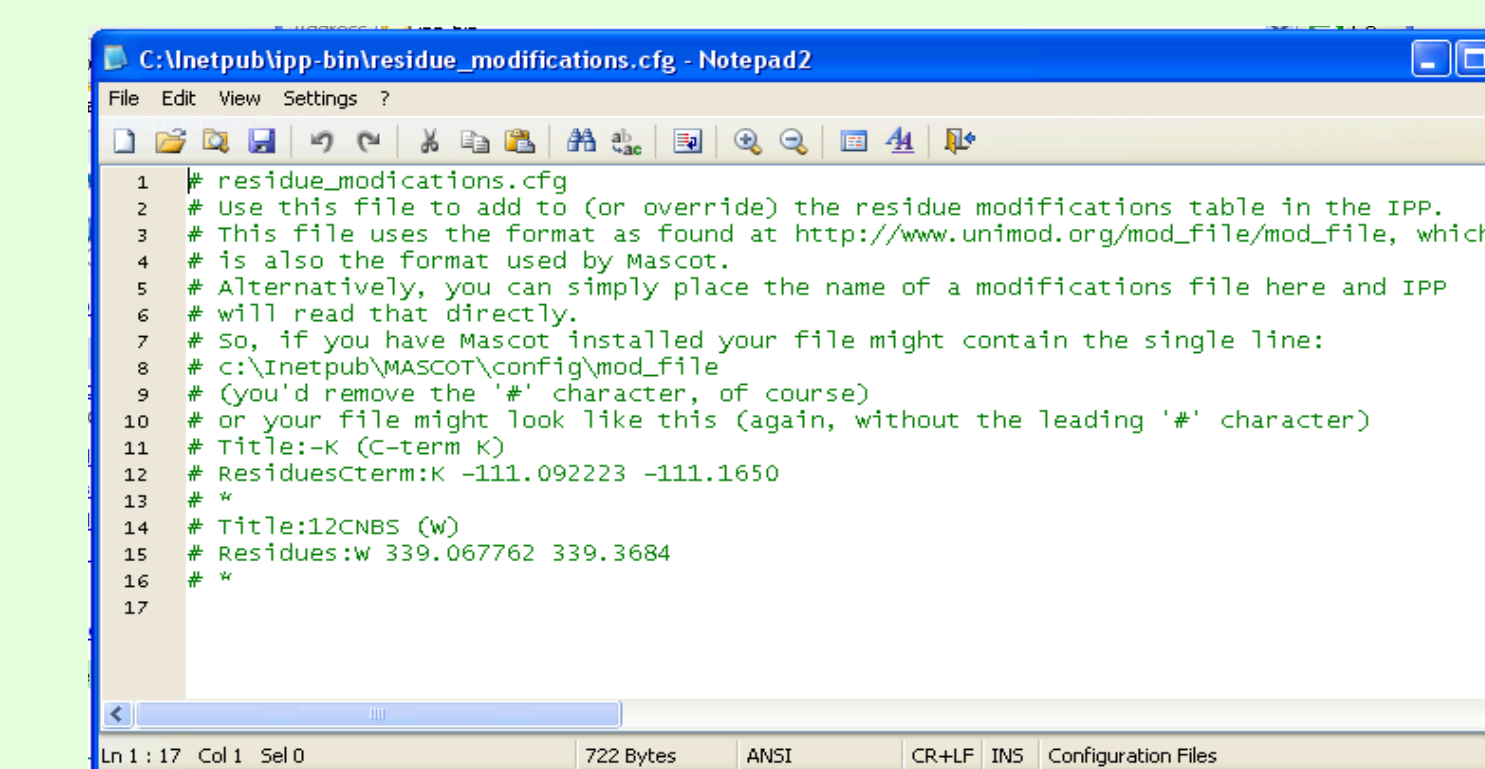


mzXML



UNIMOD-formatted Residue Modifications

IPP has adopted the UNIMOD format to specify residue modifications. With this improvement Mascot users can use the same residue modification file for IPP and Mascot and new residue modifications no longer require recompiling the proteomics pipeline, or waiting for someone to do it for you.



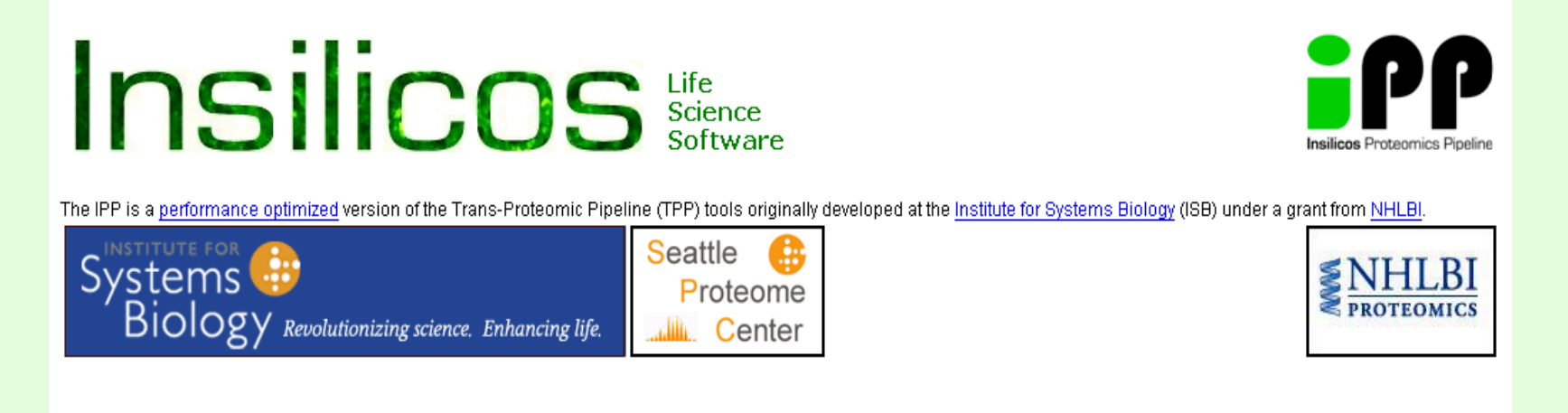
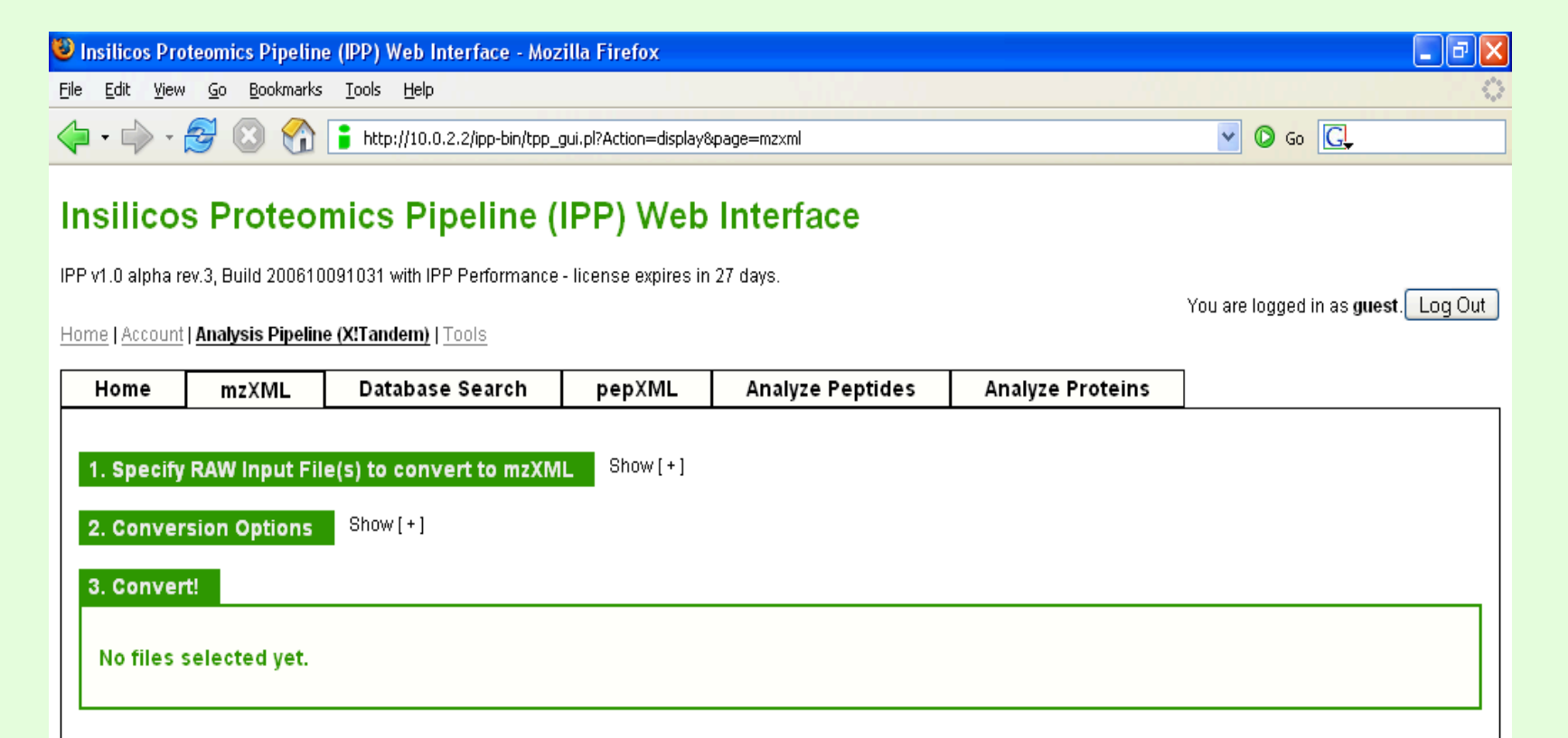
Abstract

The Trans-Proteomic Pipeline (TPP) is an open source proteomics data analysis pipeline introduced by the Aebersold group at the Institute for Systems Biology. This pipeline includes tools for quantitation and statistical validation of protein identifications from tandem mass spectrometry experiments. While TPP is extremely useful and well received, many users struggle with the application of this data analysis pipeline to Mascot search results. Most of the pitfalls encountered while applying the TPP to Mascot data stem from file format converters. There are multiple data processing paths available for the analysis of Mascot search results, some of which lead to more comprehensive protein identifications than other methods. Recently, Insilicos has issued a version of the TPP that is known as the Insilicos Trans-Proteomics Pipeline (IPP). This version of the pipeline allows for analysis using MS/MS data that went through Mascot Distiller or DTASuperCharger. Previous versions of the proteomic pipeline required that MS/MS data from some instruments be converted to mzXML or mzDATA prior to searching with Mascot. Presented are results of LC-MS/MS experiments in which a number of data analysis paths are explained and compared.

IPP

Insilicos has developed the **Insilicos trans-Proteomics Pipeline (IPP)**, a performance enhanced version of the TPP which is tested, faster, more robust, and easier to install and use. Insilicos continues to be an active contributor to the open source TPP project. The IPP utilizes the core TPP logic while boosting performance dramatically (see table to the left). Insilicos software professionals applying fundamental software engineering tools:

- Profiling and targeted code optimization
- Automated builds and regression testing
- Installer/Uninstaller package



The experience of organizations such as Red Hat and Active State shows that commercial distributors can add stability and value to open-source communities. Insilicos envisions a similar role with IPP.

Experimental

- Orbitrap mass spectrometer.
- Mouse sample searched against IPI database.
- Dimethylation of N-term and lysines was used as the quantitative labeling system.
- Samples were reduced and alkylated with iodoacetamide.
- Mascot search done with Peptide Mass Tolerance: ± 10 ppm and Fragment Mass Tolerance: ± 0.5 Da.
- Default parameters were used for Mascot Distiller and DTASuperCharge.

Conclusions

- Each Mascot data process path leads to slightly different peptide identifications.
- Only IPP offers ICAT style quantitation when DTASuperCharge is used.
- Only IPP offers ICAT style quantitation when Mascot Daemon is used on .WIFF data.
- The four data processing paths led to substantially different numbers of identified peptides.
- In some cases, the peptide that was matched to a given spectra changed when the data processing method changed.

To download a copy of this poster go to www.insilicos.com