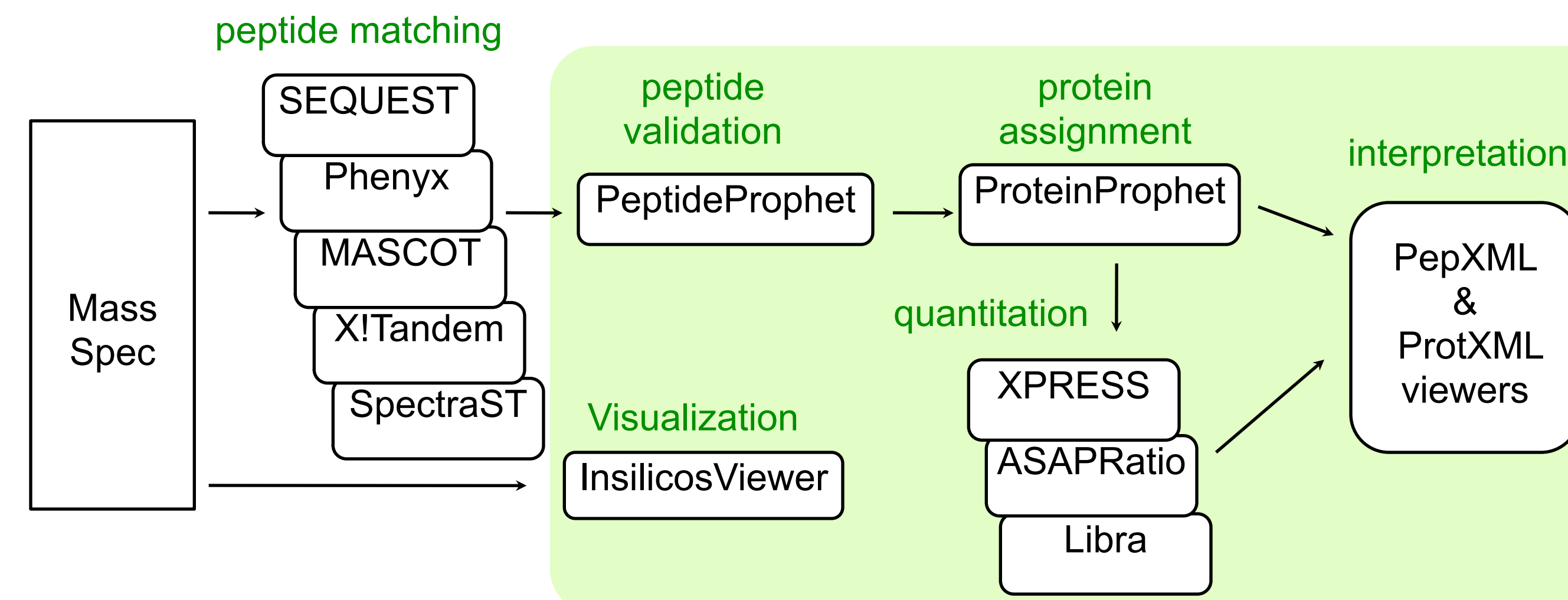


# A Fast, Robust and Reliable Data Analysis Pipeline to Produce Statistically Valid Protein Identifications and Quantitation Results from Tandem MS

Bryan J. Prazen, Brian S. Pratt, Erik J. Nilsson  
Insilicos LLC, Seattle WA

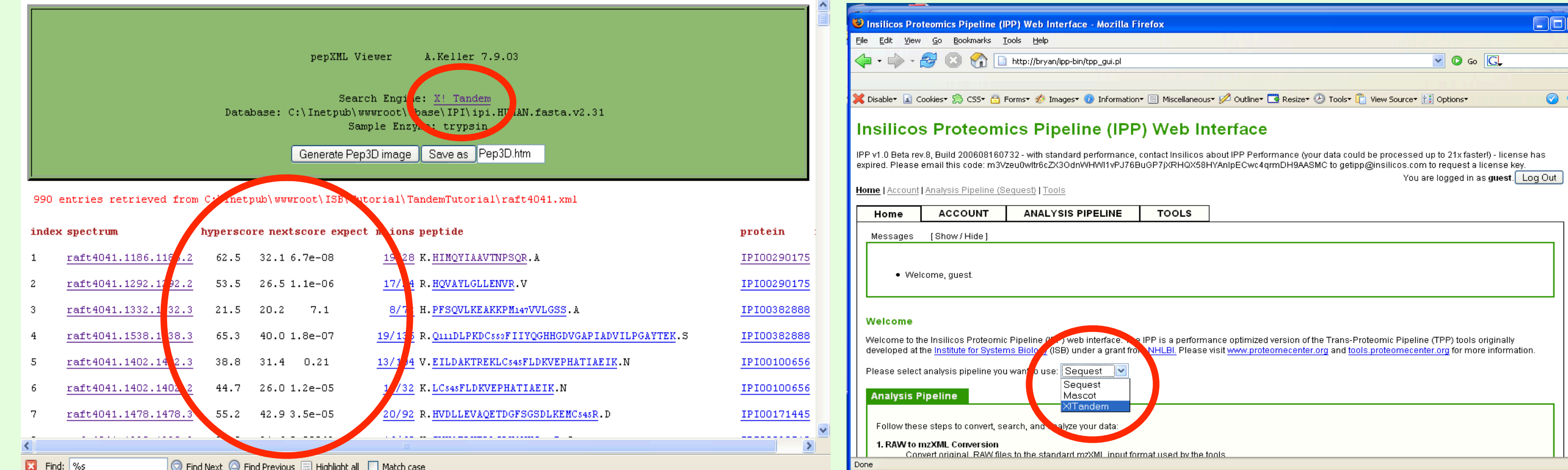
## Insilicos Proteomics Pipeline (IPP)



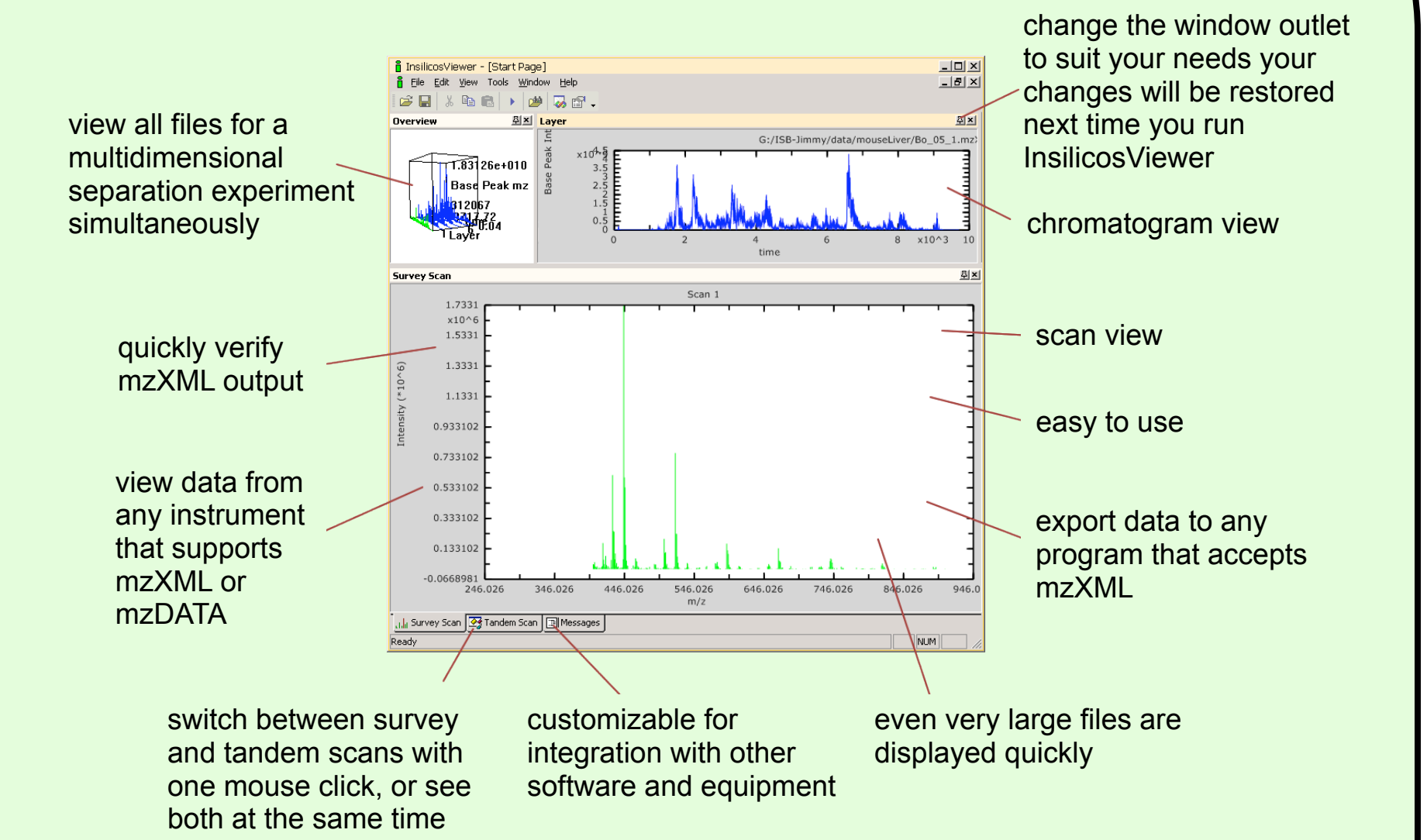
The Insilicos Proteomics Pipeline (IPP) includes modules for validation of database search results, quantitation of isotopically labeled samples and other proteomics functions. IPP is based on the Trans-Proteomic Pipeline (TPP), a popular open source toolkit for proteomics data processing, introduced by the Aebersold group at the Institute for Systems Biology (ISB).

## IPP with X!Tandem & GeneBio's Phenyx

IPP now includes tools to run X!Tandem searches, analyze X!Tandem results and view X!Tandem results. Tools for the analysis of Phenyx results with IPP are currently under active development. A paper covering the calibration of PeptideProphet for X!Tandem data by Alexey Nesvizhskii of the University of Michigan is in preparation. The current version of IPP contains the preliminary calibration developed by Nesvizhskii.



## InsilicosViewer



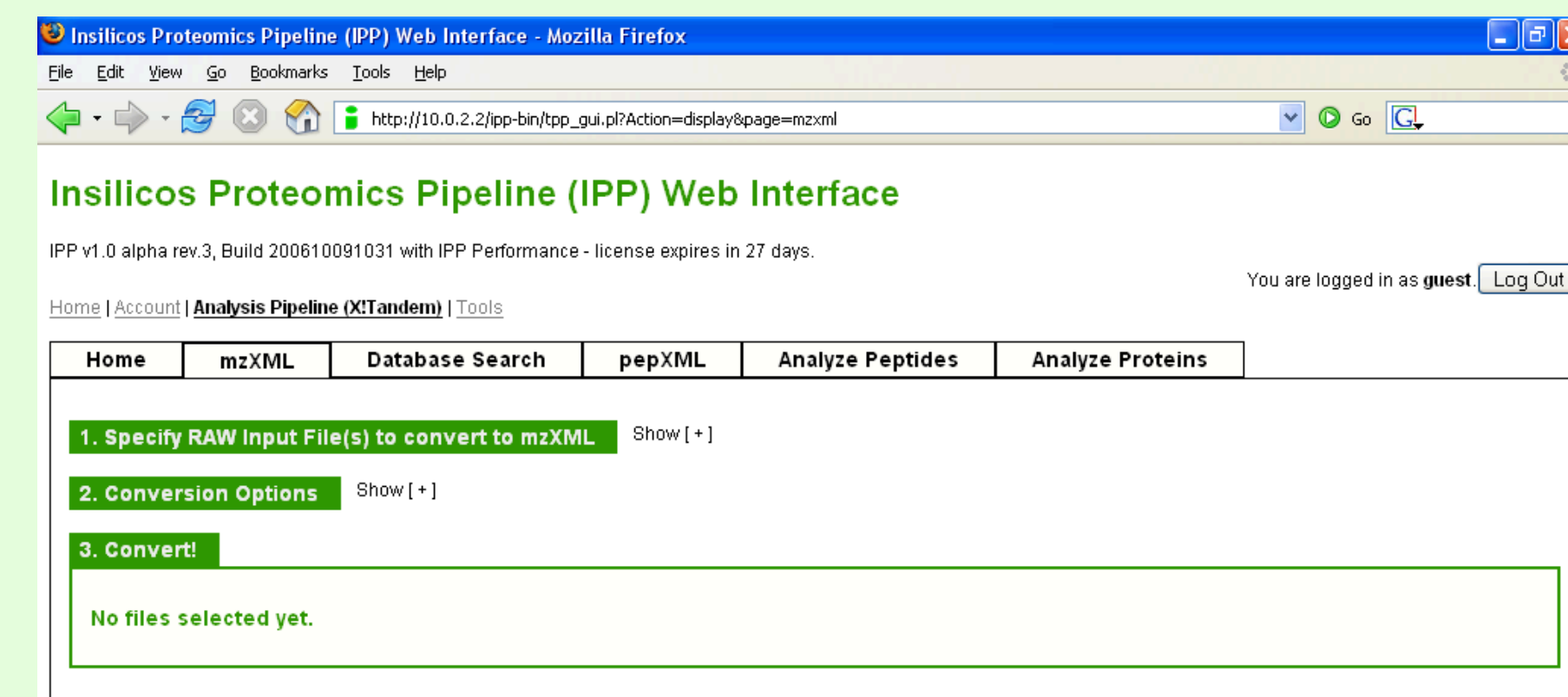
InsilicosViewer is a Windows application that reads mzDATA, mzXML, ANDI CDF, and some manufacturer specific formats. Uniquely, InsilicosViewer can combine the datasets from multiple SCX fractions, such as for a MudPIT experiment, showing the entire experiment as one 3D map. InsilicosViewer is provided free of charge as a courtesy to the scientific community and to promote the use of common MS file formats.

## IPP Software Engineering

Insilicos has developed the **Insilicos Proteomics Pipeline (IPP)**, a performance enhanced version of the TPP which is regression 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 right). Insilicos' software professionals apply fundamental software engineering tools:

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



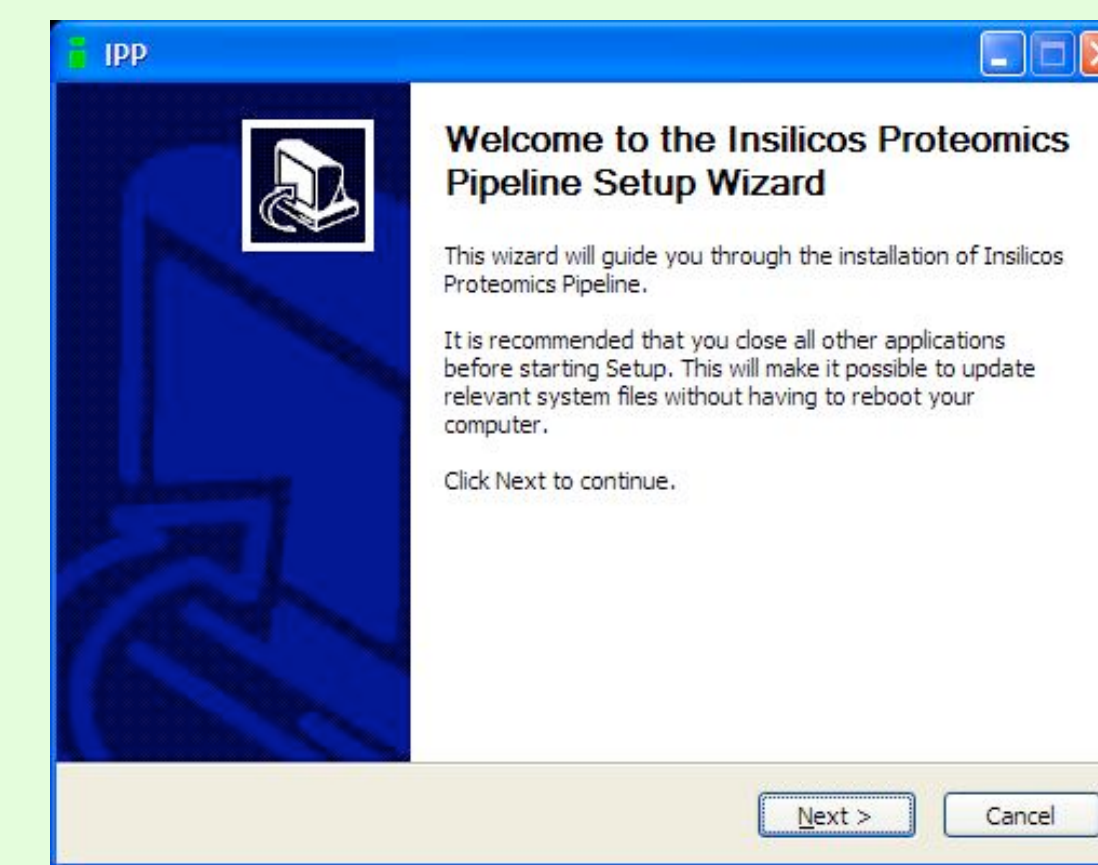
The experience of organizations such as Red Hat and Active State shows that commercial distributors can add stability and value to open-source communities. Insilicos' envisions a similar role with IPP.

## Fast Easy Installation

The IPP uses a native Windows installer and eliminates the use of the Cygwin Linux emulation package, so installation (and removal) of the pipeline on Windows is now a simple process requiring only a few minutes.

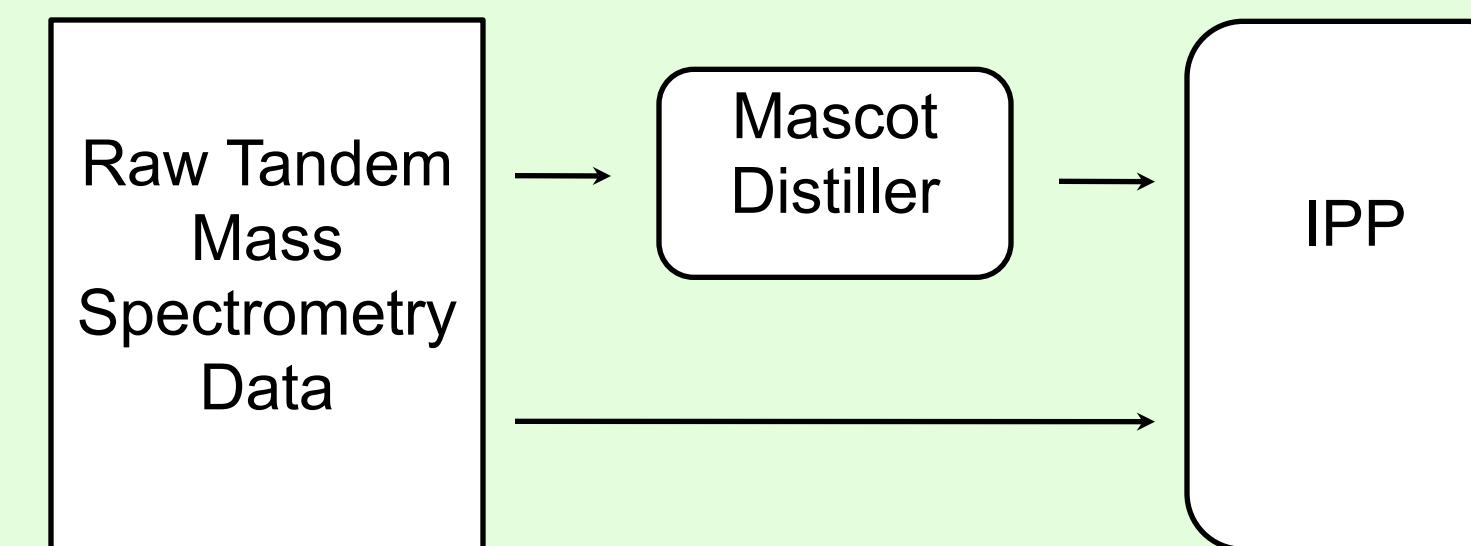
If your computer has a web server the IPP will detect it and configure IPP to the available web server. If a web server is not available the IPP installer will offer to install and configure the Apache web server.

Because the installation does not depend on Cygwin and uses an available web server, the IPP is not likely to interfere with other tools such as Phenyx, Mascot or SEQUEST which can be affected by the environment changes caused by a Cygwin installation.



## Distiller Support

The IPP now allows you to easily analyze the results from Mascot Distiller. Previously it was nearly impossible to associate the spectra matched by Mascot Distiller analysis of the raw MS data with spectra in a mzXML file. Without an association between the Mascot results and the mzXML data, quantitation using ASAPRatio, XPRESS and Libra is impossible. The IPP now works smoothly with data from Mascot Distiller.



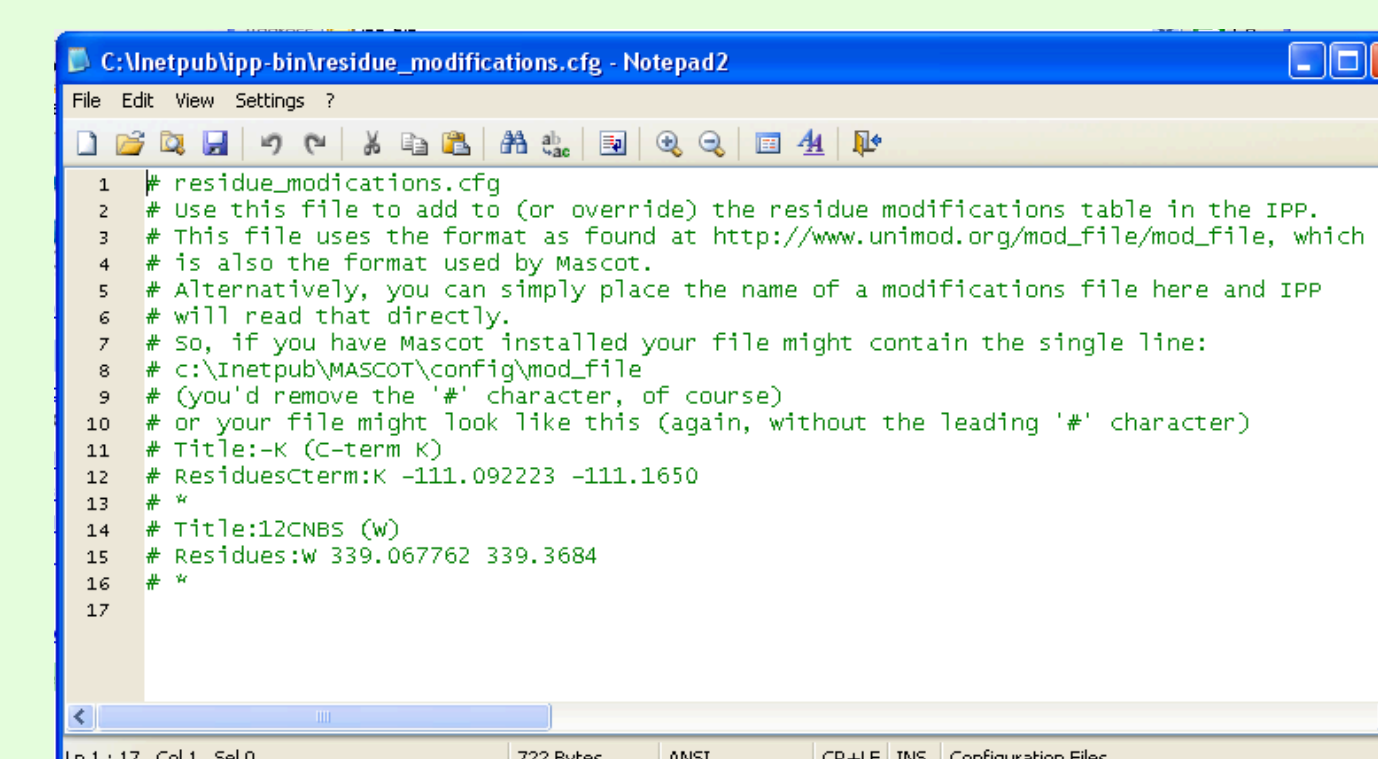
## Performance

Task	TPP	IPP	Improvement
Process Mascot search results on ICAT set "A" with ASAPRatio	3205 sec	182 sec	17.6x
Process Sequest search results on ICAT set "A" with ASAPRatio	420 sec	56 sec	7.5x
Process Sequest search results on ICAT set "A" with XPress	119 sec	37 sec	3.2x
Process Sequest search results on ITraq set "B" with Libra	539 sec	25 sec	21.6x

Performance was measured on a 3GHz PC running Windows XP Professional, comparing the TPP code from SourceForge on 5/2006 against the IPP from the same time period. The best of three times for each task is reported.

## UNIMOD-formatted Residue Modifications

IPP has adopted the UNIMOD format to specify residue modifications. The aim of UNIMOD is to create a community supported, comprehensive database of protein modifications for mass spectrometry applications. Specifying unusual residue modifications no longer requires recompiling software, or waiting for someone to do it for you. Instead, a UNIMOD-formatted file can be used to set the modifications.



## Reliability

As proteomics experiments continue to become more complex the researcher must depend on software to function reliably. 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. Insilicos does extensive testing of the open source tools which are distributed in IPP to ensure the accuracy of your results. A drawback of open source tools is that often users are the first line of defense against bugs. Using regression testing, Insilicos has prevented newly introduced bugs from showing up in the open source code. This automated testing capability also allows us to verify that the IPP with its performance enhancements yields the same results as the original code.

## Acknowledgments

Insilicos gratefully acknowledges the authors of tools contained in the IPP, fellow maintainers of the TPP, and the user community members who continue to make the IPP and TPP better through their participation.

This work was funded in part by NIH grant # HL079807.

To download a copy of this poster go to [www.insilicos.com](http://www.insilicos.com)