

Recent Advances in Reliability, Performance and Usability of the Trans-Proteomic Pipeline (TPP) Software Tools

Brian Pratt¹

¹ Insilicos LLC, Seattle WA

TPP Overview

The Trans-Proteomic Pipeline (TPP) is a popular open source toolkit for proteomics data processing, introduced by the Aebersold group at the Institute for Systems Biology (ISB) as part of the "Sashimi" open source software project.

This XML-based analysis pipeline includes modules for validation of database search results, quantitation of isotopically labeled samples, protein inference, as well as tools for viewing raw LC/MS data, peptide-level results, and protein-level results. The XML backbone enables a uniform analysis for LC/MS/MS data generated by a wide variety of mass spectrometer types, and assigned peptides using a wide variety of database search engines.

TPP User Issues

While the TPP is extremely useful and well-received, our monitoring of the TPP users email list (<http://groups.google.com/group/spcctools-discuss>) shows that many users experience installation, configuration, and performance issues which prohibit the use of the TPP in high throughput proteomics facilities or facilities that cannot afford to dedicate an information technology specialist to the task.

Introducing the IPP

Insilicos has created the **Insilicos Proteomics Pipeline (IPP)** which extends the applicability of the TPP by focusing on three areas:

Speed – optimized to provide results in minimal time.

Reliability – provably provides the same results from run to run, and release to release.

Convenience – easy to install or remove without expert assistance. Residue modifications, the most commonly requested customizations, are read from a text file instead of from recompiled C++ code.

Part of the TPP Community

The **TPP** is maintained and supported by users and developers at several different research institutions and commercial firms, most notably the ISB and Insilicos. Most communication around TPP topics occurs on the spcctools-discuss (<http://groups.google.com/group/spcctools-discuss>) email list. Insilicos monitors this list and offers advice and performs bug fixes on the publicly available TPP source code (<http://sashimi.cvs.sourceforge.net/sashimi/>). Consequently, the authors are intimately familiar with the TPP codebase.

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.

Introduction

The **Insilicos Proteomics Pipeline (IPP)** is a proprietary software layer around the TPP, which utilizes the core TPP logic while boosting performance dramatically (see table to the right). Insilicos' highly experienced software professionals applying fundamental software engineering tools:

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

Insilicos has thus developed the IPP, a performance enhanced version of the TPP which is significantly faster, more robust, and easier to install and use.

Insilicos continues to be an active contributor to the open source TPP code.



Results: Improved 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

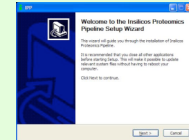
Results: Improved Ease of Use

Installation of TPP is complex, and often takes several days or even weeks. 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.

Because the installation does not depend on Cygwin, the IPP is much less likely than TPP to interfere with other tools such as Mascot or SEQUEST which can be affected by the environment changes caused by a Cygwin installation.

A longstanding user headache is solved: specifying residue modifications no longer requires understanding how to recompile the software, or waiting for someone to do it for you. Instead, if the user provides a UNIMOD-formatted file the IPP will add to or override the residue modifications in its internal table.

The IPP reads existing TPP output files, so you don't have to start over to make the switch to better performance.



Results: Improved Reliability

As veterans of many years of commercial software development experience, Insilicos understand how important it is to produce software that is both fast and reliable. By introducing convenient automated regression testing capability into the TPP codebase, Insilicos has prevented newly introduced bugs from showing up in TPP releases that would otherwise have been found by end users – or worse yet, not caught at all. This automated testing capability also allows us to verify that the IPP with its performance enhancements still yields the same results as the generic TPP. As ongoing members of the TPP open source community, Insilicos shares any and all improvements made to the core TPP code.

Conclusions

The IPP is available from Insilicos at www.insilicos.com. A free version of IPP includes several important improvements, including more modest speed improvements and the ability to read UNIMOD residue modification files.

Acknowledgments

Insilicos gratefully acknowledges the original authors and fellow ongoing maintainers of the TPP, and the user community members who continue to make the TPP better and better through their participation on the TPP user email list.

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

Insilicos, IPP, Insilicos Proteomics Pipeline, and the Insilicos logos are registered trademarks of Insilicos LLC. All other brand names, product names, or trademarks belong to their respective holders.