

A Novel Visualization Tool for Common Mass Spectrometric File Formats

Insilicos™
Life · Science · Software

Erik J Nilsson, Brian S Pratt, and Bryan J Prazen; Insilicos, Seattle WA

Introduction

Each mass spectrometer manufacturer stores data in a unique file format, some of which are proprietary. Proprietary formats limit the ability to compare results between research groups or even between instruments maintained by the same group. Such formats also discourage the exchange of raw data files and the creation of data repositories. mzXML and mzDATA are two common file formats developed to allow for the exchange of MS data. Presented here is a fast and novel visualization tool for mzXML and mzDATA files, known as *InsilicosViewer*.

Methods

InsilicosViewer reads mzDATA, mzXML, ANDI, CDF, and some manufacturer specific formats. *InsilicosViewer* displays spectra from multiple levels of ionization and allows for customizable views of chromatographic and spectrometric data, suitable for publication. Uniquely, *InsilicosViewer* can combine the datasets from multiple SCX fractions, such as for a MudPIT experiment, showing the entire experiment as one 3D map. Furthermore, multiple experiments can be combined, displayed, and compared. *InsilicosViewer* is provided free of charge as a courtesy to the scientific community and to promote the use of common MS file formats. *InsilicosViewer* can be obtained from Insilicos (www.insilicos.com).

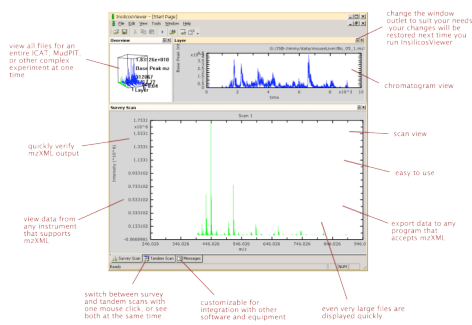
Previous Results

The *InsilicosViewer* quickly allows scientists to visually evaluate even the largest proteomics datasets and serves as a vital utility in the promotion of common file formats. Features of the *InsilicosViewer* include the ability to:

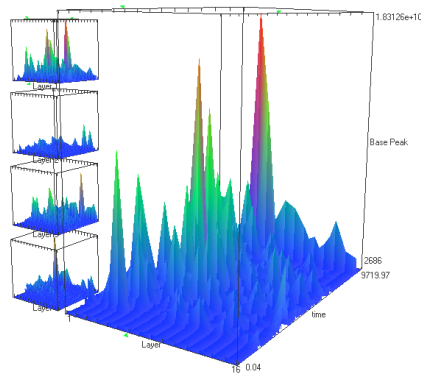
- View data from any instrument that supports mzDATA, or mzXML file formats.
- Quickly display large data files.
- Display separations, survey scans, and tandem scans in a single window.
- Customize display layout to users' needs
- Simultaneously view all files from a multidimensional separation experiment
- Evaluate sample quality and instrument performance quickly.

Compatibility

The *InsilicosViewer* runs on Windows 2000, Windows XP, and Windows XP Pro operating systems. The mzXML and mzDATA common file formats are both XML (extensible Markup Language) formats. mzXML is a data format developed by the Institute for Systems Biology (ISB), in collaboration with Insilicos and others; mzData is a format developed by the Human Proteome Organization Mass Spectrometry Standards Working Group (HUPO PSI-MS).

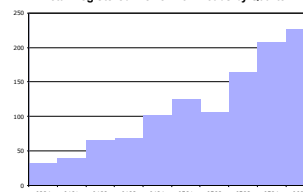


- View data from any instrument that supports mzDATA or mzXML
- Quickly display large data files
- Display separations, survey scans, and tandem scans in a single window
- Customize display layout to meet the users needs
- Simultaneous viewing of all files from a multidimensional separation
- Evaluate sample quality and instrument performance quickly
- Optimize experimental conditions
- Verify the mzDATA and mzXML file format
- Detect contaminants quickly
- Identify differences in large, complex datasets visually
- Check sample reproducibility

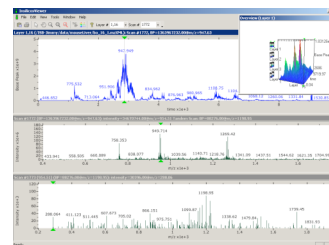


- Holistic view of very large proteomics experiments.
- Quickly display large data files.
- Multiple viewing methods: heatmap, hypecube (shown above), and chromatographic (top and right)

Total Registered Viewer Downloads by Quarter

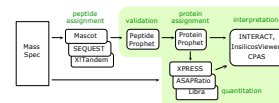


InsilicosViewer has become increasingly popular with the proteomics community. Shown above are new, registered downloads per quarter. (Repeated downloads of the same version, and unregistered downloads are not included in these totals.)



InsilicosViewer simultaneously shows the experiment, chromatogram, survey scan, and tandem scan. MS³ and higher fragmentation can also be simultaneously viewed. Multiple tandem scans per survey scan are handled via an elegant tabbed interface.

Insilicos Proteomics Pipeline



InsilicosViewer is a component of Insilicos Proteomics Pipeline (IPP), a proteomics analysis pipeline based on the Institute for Systems Biology's TPP.

mzXML

InsilicosViewer originally read only mzXML format data. mzXML is a data format developed by the Institute for Systems Biology (ISB), in collaboration with Insilicos and others. The goal of this format was a public domain standard for raw mass spectrometric data. Some mass spectrometers do not produce mzXML files directly, but there are several tools available that generate mzXML files from native acquisition files. The Sashimi website at <http://sashimi.sourceforge.net> offers a collection of converter programs for common manufacturer-specific file formats and information about software that utilizes these file formats.

mzData

mzData is a format developed by the Human Proteome Organization Mass Spectrometry Standards Working Group (HUPO PSI-MS). The goal of both formats was to provide a public domain standard maintained by a neutral organization for raw mass spectrometric data. The HUPO PSI-MS website at psidev.sourceforge.net/ms/index.html offer a collection of converter programs for common manufacturer specific file formats and information about software that utilizes the common file formats.

Other Formats

- InsilicosViewer* supports a variety of legacy and manufacturer-specific formats, including:
 - ANDI CDF
 - Legacy .RAW
 - Leco ASCII

References

- mzXML** Pedrioli PGA, Eng JK, Hubley R, Vogelzang M, Deutsch EW, Raught B, Pratt B, Nilsson E, Angeletti R, Apweiler R, Cheung K, Costello CE, Hermjakob H, Huang S, Julian RK Jr, Kapp E, McComb ME, Oliver SG, Omenn G, Paton NW, Simpson R, Smith R, Taylor CF, Zhu W, Aebersold R. (2004) "A Common Open Representation of Mass Spectrometry Data and its Application in a Proteomics Research Environment." *Nature Biotechnology* 22(11):1459-1466
- mzXML 3.0** Pedrioli, PGA, et al, *The mzXML Schema version 3.0* (Poster) ASMS 2006 ThP22 #387
- mzData** <http://psidev.sourceforge.net/ms/>

Acknowledgements

This work was funded in part by NIH grant HL079807