

A Software Shell for MS Data Processing and Management

Yang Su^{1,5}, Sequin Huang^{1,3,6}, Hua Huang^{1,3}, David H. Perlman^{1,3}, Mark E. McComb^{1,3}, Catherine E. Costello^{1,2,4}

1. Cardiovascular Proteomics Center, 2. Mass Spectrometry Resource, 3. Department of Medicine, 4. Department of Biochemistry, Boston University School of Medicine, 5. Bioinformatics Program, Boston University, Boston, MA, 6. Present address: Waters Corp, Milford, MA

Introduction: Here we present a stand-alone software application named BUDSS (Boston University Database Search Shell) which is designed to aid the processing and management of various MS data files. For data processing it performs two separate but related functions: automatic conversion of processed MS data files (includes both mzXML and mzData formats) into several common formats accepted by different software applications and the submission of these converted data files to several web based database searching engines (Mascot, Ms-Fit, BUPID). For data management, the software supports multiple users/projects and organizes data and settings individually for each user/project. Finally, all these functions are accessible through a user-friendly graphical interface and the executable program will be available for free download.

Methods: BUDSS was developed using Microsoft Visual Studio 6.0. MSXML 4.0 was used as an XML parser to decode Base64 encoded peak list data in the mzXML/mzData file. A C++ library was also built to improve computational efficiency of decoding. Additional supported data formats are intermediate files converted from raw data files using manufacturers' software: LC MS/MS data is processed with Analyst QS (ABI/Sciex), MassLynx/PLGS2.2 (Waters); MALDI MS data with MoverZ (Proteometrics LLC); and FTMS data with BUDA (BUSM). These files can then be converted into a variety of database searching engine formats (web-based formats include Mascot, Ms-Fit, BUPID and stand-alone formats include PLGS and Sequest).

Results: We tested BUDSS using large volumes of MALDI-TOF MS, MALDI-FT MS and LC MS/MS data sets obtained in house. After initial processing using vendor-specific programs BUDSS converted processed files to the formats of several commercially and publicly available search engines. Files were then submitted for protein identification to the search engines with the search settings specified by the user. Search settings are retained and applied to multiple search sessions once they had been entered for the first search. Results files are automatically saved in HTML format and can then be viewed directly inside the program. Both data conversion and database searching were run in the batch mode to handle large amount of data with little manual intervention (BUPID). BUDSS supports multiple users/projects and for each user/project stores data/result files in his own directory for easy data access. Each user/project can also save his usual search settings into XML files and such files can be automatically loaded and applied for future searches. By providing support for individual user/project, BUDSS provides an excellent way of managing and archiving proteomics datasets.

Conclusions: BUDSS provides an easy-to-use graphical interface for performing MS data processing and management in unattended batch mode. It can also be easily be expanded for more MS data types and linked to more database search engines.

Acknowledgement:

This project was funded by NIH NHLBI contract N01-HV-28178 and NIH NCRR grants P41-RR10888 and S10-RR15942.

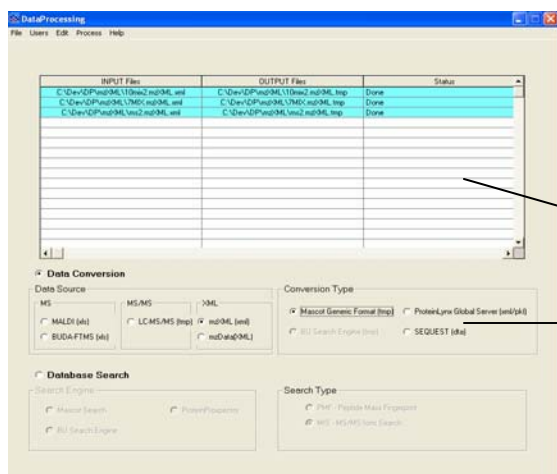


Figure 1: Front page of BUDSS showing an example of the batch processing of the conversion of LC MS/MS mzXML files to Mascot generic format.

File grid shows the files being processed and processing status.

Selection panel lets user choose between different functions.

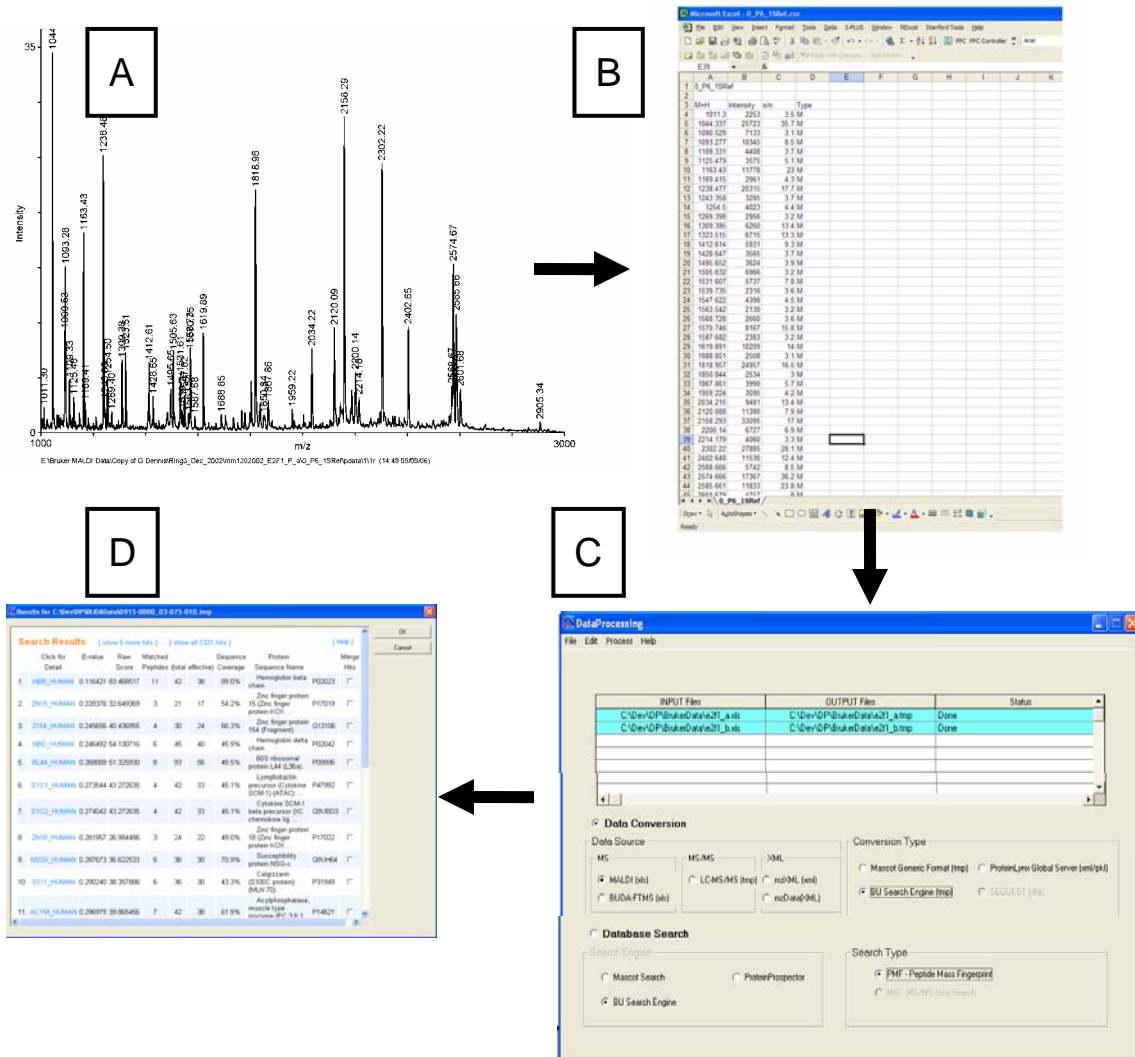


Figure 2: The pipeline of Data Analysis. 2A: A Bruker MALDI-TOFT mass spectrum analyzed using MoverZ™. 2B: Export of a processed peak list in excel format. 2C: Conversion of this file format to BUPID format using BUDSS. 2D: Automatic database search using BUDSS and results from the BUPID web server.