**Technical Brief**

# SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches

**Marc Vaudel[1*], Harald Barsnes[2,3*], Frode Berven[2], Albert Sickmann[1], Lennart Martens[4,5]**

[1] Leibniz – Institut für Analytische Wissenschaften - ISAS - e.V., Dortmund, Germany

[2] Proteomics Unit, Department of Biomedicine, University of Bergen, Norway

[3] Computational Biology Unit, UniBCCS, University of Bergen, Norway

[4] Department of Medical Protein Research, VIB, B-9000 Ghent, Belgium

[5] Department of Biochemistry, Ghent University, B-9000 Ghent, Belgium

[*] These authors contributed equally to this work.

***Corresponding author:***

Harald Barsnes, Department of Biomedicine, University of Bergen, Jonas Liesvei 91, N-5009 Bergen, Norway; E-mail: harald.barsnes@biomed.uib.no; Fax: (+47) 55 58 63 60.

***Abbreviations:*** OMSSA – the Open Mass Spectrometry Search Algorithm

***Keywords:*** Bioinformatics / OMSSA / Protein identification / Search Engines / X!Tandem

***Total number of words:***1862

**Abstract**

The identification of proteins by mass spectrometry is a standard technique in the field of proteomics, relying on search engines to perform the identifications of the acquired spectra. Here we present a user-friendly, lightweight and open-source graphical user interface called SearchGUI (http://searchgui.googlecode.com), for configuring and running the freely available OMSSA and X!Tandem search engines simultaneously. Freely available under the permissible Apache2 license, SearchGUI is supported on Windows, Linux, and OSX.

**Main text**

Mass spectrometry (MS) is a key technique for the identification of proteins within proteomics experiments. The proteins are first proteolytically digested, resulting in peptides which are then analyzed using tandem mass spectrometry, yielding MS/MS spectra containing peptide precursor and fragment ion mass-to-charge ratios (m/z) and their intensities. A specialized software application called a search engine then attempts to identify the acquired MS/MS spectra by matching the observed ions to theoretical ions from known protein sequences obtained from a database. The results are written to an output file where each spectrum-identification mapping gets a score value, and in most cases also an expectancy value as well as a score threshold that helps the user distinguish between true and false positives [1]. Several algorithms, both free and commercial, have been developed for this purpose, including Mascot [2], SEQUEST [3], OMSSA [4], VEMS [5] and X!Tandem [6].

Each search engine has its own strengths and weaknesses, and the best results are often achieved by using more than one algorithm [7-9]. Running multiple search engines can however be a complicated process since each algorithm requires distinct input parameters and uses separate interfaces for communication. Two common options for configuring and running a search engine are *via* web interfaces or through a command line interface. While the web interface has its advantages, including simple user interaction, it is often complicated to automate. Web interfaces are also difficult to introduce into existing proteomics pipeline software. Command line interfaces on the other hand are much easier to tailor to the individual user's needs, but most often require advanced computer skills to operate.

A solution in between these two alternatives is therefore included with the commercial Mascot search engine. This so-called Mascot Daemon client application automates the submission of data files to the Mascot server, and supports the option of batch processing (http://www.matrixscience.com/daemon.html). A similar, open source tool called OMSSAGUI [10] has also been implemented for the freely available OMSSA search engine (http://pubchem.ncbi.nlm.nih.gov/omssa). OMSSAGUI serves as a stand-alone front-end to OMSSA, and can be used as part of an informatics processing pipeline for MS driven proteomics.

We here describe a drastically extended version of the OMSSAGUI, called SearchGUI, which in addition to important improvements, now also supports X!Tandem database searches (http://www.thegpm.org/tandem) using the same input parameters. This means that the user only has to input the search parameters once in order to configure and start both OMSSA and X!Tandem searches simultaneously. SearchGUI supports the input of three common spectra file formats: MGF, PKL and DTA. It is possible to use multiple files and even combinations of input file formats, in the same search. This is a vital improvement on OMSSAGUI, which only supports one input file per search. The protein sequence database to run the search against can be provided by the user in the standard FASTA sequence format.

SearchGUI can be used as a stand-alone tool, where the user manually inserts the parameters for the search, or it can be included as a lightweight component in existing proteomics workflow software. The latter is made possible by providing the required parameters as a predefined configuration file. This file can either be created from scratch, or the parameters can be saved to a configuration file directly from the

SearchGUI graphical user interface. A code example illustrating how to incorporate SearchGUI as a component in a pipeline can be found in the online documentation of SearchGUI (http://searchgui.googlecode.com).

To start identifying peptides and proteins using SearchGUI, it is sufficient to download the latest version from the SearchGUI web page, unzip the downloaded file, and double-click the SearchGUI jar file. The tool will then start and display the first set of input parameters, referred to as the Task Editor (Figure 1). In this section the installed locations of the search engines is provided, the input files are selected, and the results folder can be specified. It is also possible to load a configuration file containing the additional parameters for the search. Assuming that no such file has been uploaded, the next step involves selecting the Parameters Editor tab, where the basic search parameters are inserted (Figure 2). Again, it is possible to load a configuration file here, and the current settings can be saved to a new or existing file as well. If a predefined file is not used, the FASTA sequence database must be provided, and the set of fixed and variable modifications need to be selected. Furthermore, the user must specify the *in silico* protease, the number of allowed missed cleavages, the precursor and fragment ion mass tolerances, the fragment ion types considered, and the lower and upper bounds for the precursor charge. Finally, a set of advanced settings can be provided in the Advanced Parameters Editor tab (Figure 3). This includes the e-value cut-off and the maximum hit list length, along with a set of OMSSA specific parameters, such as the minimum precursor charge to consider multiply charged fragments, and minimum and maximum allowed peptide lengths. These parameters provide specific fine-tuning that is particularly useful for searches using ETD spectra as well as no-enzyme or semi-tryptic digest searches.

When all parameters have been inserted, the user can decide which search engine(s) to run the search on, and then start the process by clicking the "Run" button in the Task Editor tab. During the search the user will be kept informed about the progress while any output from the search engines will be displayed on screen as well. Upon completion, the search results will be stored in separate files in the previously selected output folder.

SearchGUI comes with the default versions of OMSSA and X!Tandem for Windows, Linux and OSX platforms, but the user can specify different, independently installed versions as well in the graphical user interface. Default sets of modifications and enzymes are also included, but the user can add new ones or edit existing entries. More detailed information on these settings can be found in the SearchGUI manual provided with the tool.

Coupled to the freely available OMSSA Parser [11] and XTandem Parser [12] tools, which allow the search engine results to be visualized and further analyzed, SearchGUI provides a very user-friendly suite of tools that empower any interested user to efficiently leverage the power of multiple search engines for proteomics identifications. Importantly, the obtained identifications in OMSSA and X!Tandem format can be submitted directly to the PRIDE online proteomics identifications database [13] using PRIDE Converter [14].

SearchGUI is made freely available as open-source under the permissible Apache2 license. Additional documentation, cross platform binaries and source files can be freely downloaded from http://searchgui.googlecode.com.

**Acknowledgements**

LM would like to thank Joël Vandekerckhove for his support.

The financial support by the Ministerium für Innovation, Wissenschaft, Forschung und Technologie des Landes Nordrhein-Westfalen and by the Bundesministerium für Bildung und Forschung is gratefully acknowledged.

The authors have no competing financial or commercial interests.

**References**


[1] Aebersold, R. and Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, *422*, 198-207.

[2] Perkins, D.N., Pappin, D.J., Creasy, D.M. and Cottrell, J.S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, *20*(18), 3551-67.

[3] Eng, J., McCormack, A.L. and Yates, J.R., III, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 1994, *5*(11), 976–989.

[4] Geer, L.Y., Markey, S.P., Kowalak, J.A., Wagner, L. *et al.*, Open mass spectrometry search algorithm. *J Proteome Res* 2004, *3*(5), 958-64.

[5] Matthiesen, R., Virtual Expert Mass Spectrometrist v3.0: an integrated tool for proteome analysis. *Methods Mol Biol* 2007, *2007*(367), 121-38.

[6] Fenyo, D. and Beavis, R.C., A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal Chem* 2003, *75*(4), 768-74.

[7] Balgley, B.M., Laudeman, T., Yang, L., Song, T. and Lee, C.S., Comparative evaluation of tandem MS search algorithms using a target-decoy search strategy. *Mol Cell Proteomics* 2007, *6*(9), 1599-608.

[8] Searle, B.C., Turner, M. and Nesvizhskii, A.I., Improving sensitivity by probabilistically combining results from multiple MS/MS search methodologies. *J Proteome Res* 2008, *7*(1), 245-53.

[9] Alves, G., Wu, W.W., Wang, G., Shen, R.F. and Yu, Y.K., Enhancing peptide identification confidence by combining search methods. *J Proteome Res* 2008, *7*(8), 3102-13.

[10] Tharakan, R., Martens, L., Van Eyk, J.E. and Graham, D.R., OMSSAGUI: An open-source user interface component to configure and run the OMSSA search engine. *Proteomics* 2008, *12*, 2376-8.

[11] Barsnes, H., Huber, S., Sickmann, A., Eidhammer, I. and Martens, L., OMSSA Parser: an open-source library to parse and extract data from OMSSA MS/MS search results. *Proteomics* 2009, *9*(14), 3772-4.

[12] Muth, T., Vaudel, M., Barsnes, H., Martens, L. and Sickmann, A., XTandem Parser: An open-source library to parse and analyse X!Tandem MS/MS search results. *Proteomics* 2010, *10*(7), 1522-4.

[13] Martens, L., Hermjakob, H., Jones, P., Adamski, M. *et al.*, PRIDE: the proteomics identifications database. *Proteomics* 2005, *5*(13), 3537-45.

[14] Barsnes, H., Vizcaíno, J.A., Eidhammer, I. and Martens, L., PRIDE Converter: making proteomics data-sharing easy. *Nat Biotechnol* 2009, *27*(7), 598-9.

**Figure legends**

**Figure 1:** Task Editor: the location of the search engines, the input files and the results folder is configured here. It is also possible to load a predefined configuration file containing the additional parameters for the search.

**Figure 2:** Parameters Editor: the basic search parameters are provided here, including the FASTA sequence database to search against, and the set of fixed and variable modifications. Furthermore, the protease, the number of allowed missed cleavages, the precursor and fragment ion mass tolerances, the fragment ion types, and the lower and upper bounds for the precursor charge are also selected in this frame.

**Figure 3:** Advanced Parameters: Here the e-value cut-off and the maximum hit list length, plus a set of OMSSA specific parameters, like the minimum charge to consider for multiply charged fragments, and minimum and maximum peptide lengths can be set.

**Figure 1:**

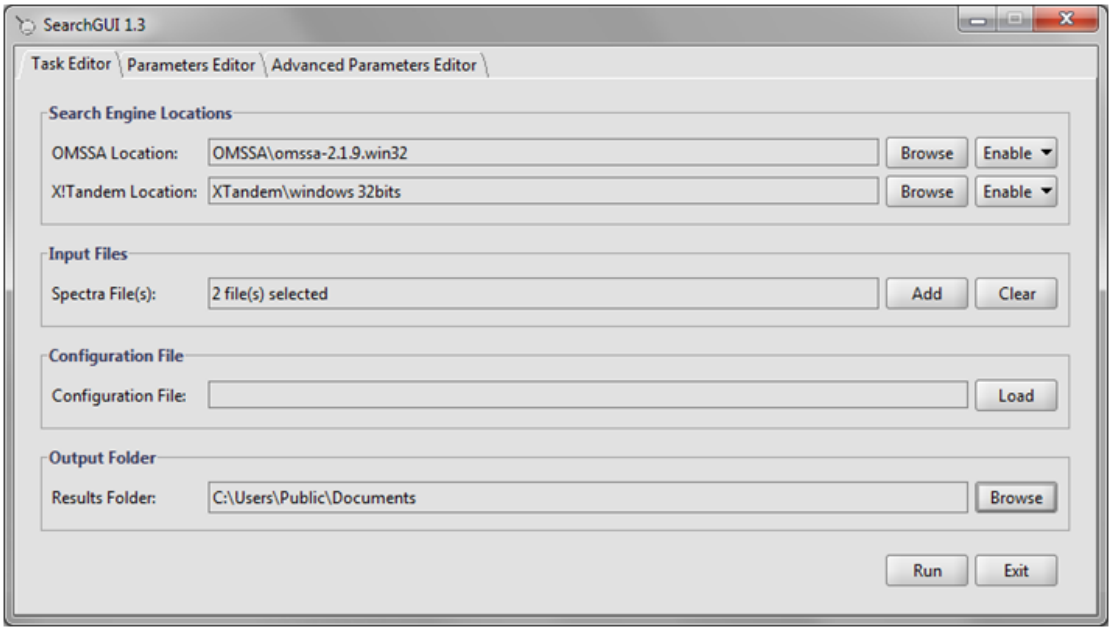

SearchGUI 1.3
Task Editor
Parameters Editor
Advanced Parameters Editor
Search Engine Locations
OMSSA Location:
OMSSA\omssa-2.1.9.win32
Browse
Enable
X!Tandem Location:
XTandem\windows 32bits
Browse
Enable
Input Files
Spectra File(s):
2 file(s) selected
Add
Clear
Configuration File
Configuration File:
Load
Output Folder
Results Folder:
C:\Users\Public\Documents
Browse
Run
Exit

**Figure 2:**

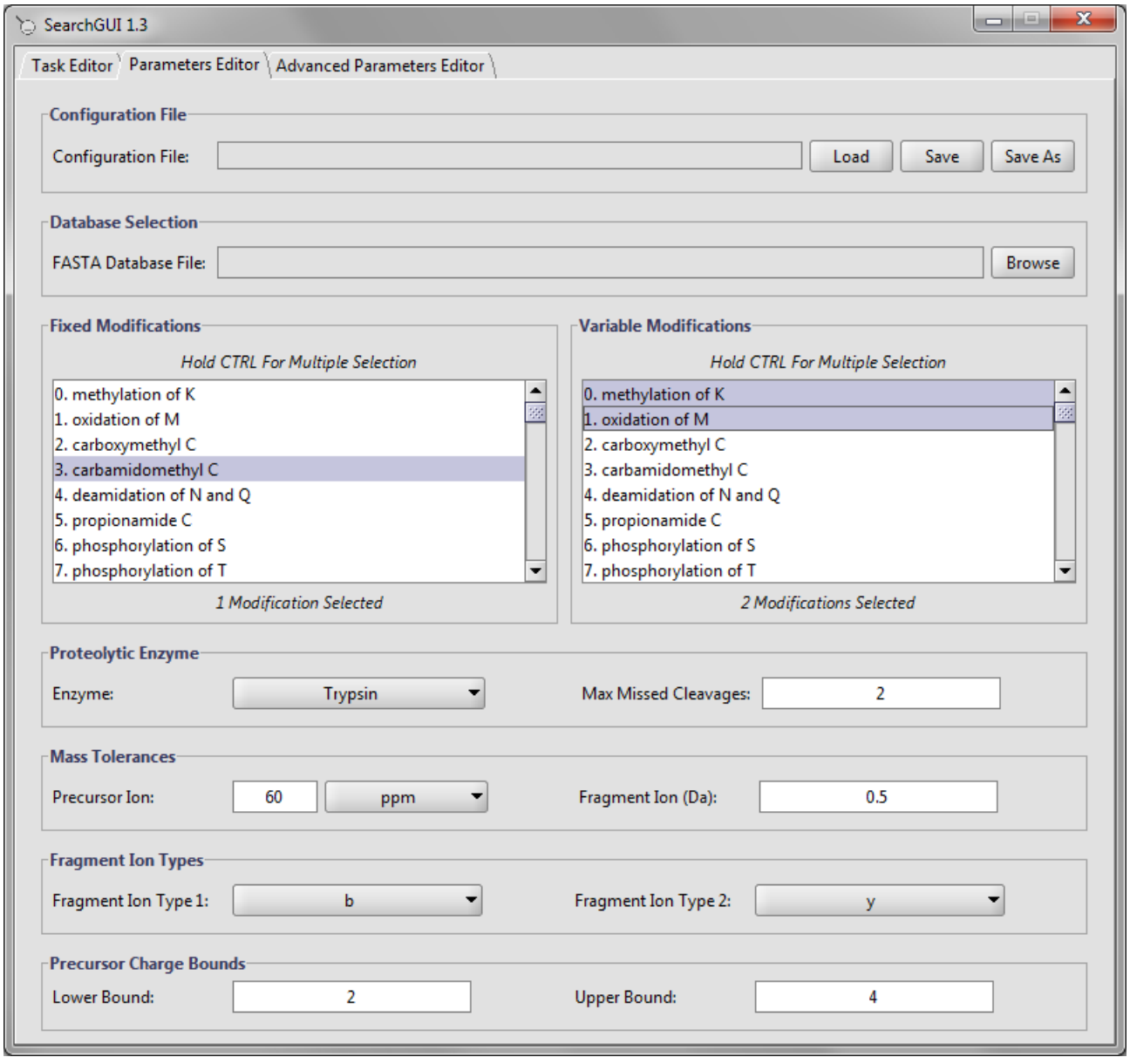
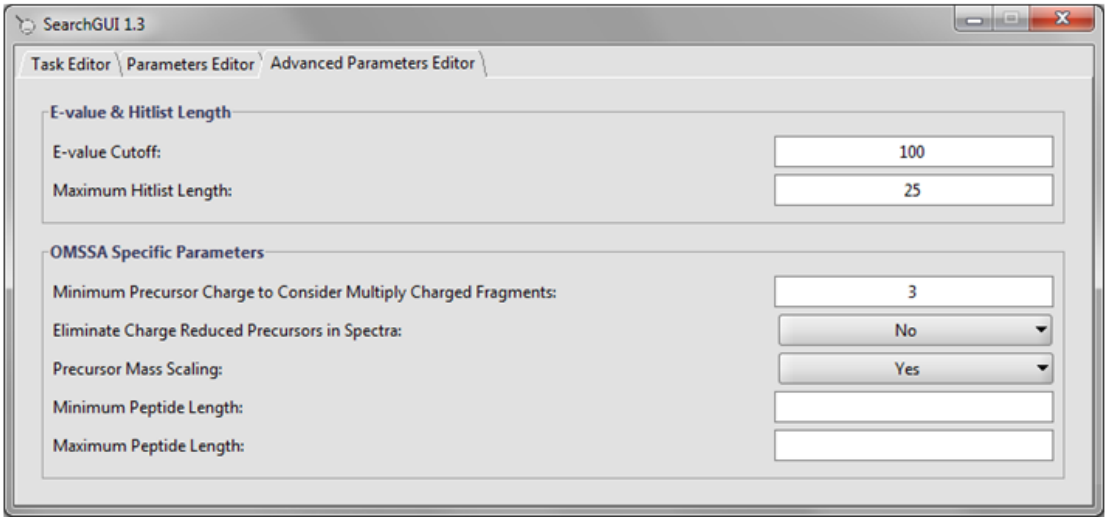

SearchGUI 1.3
Task Editor
Parameters Editor
Advanced Parameters Editor
Configuration File
Configuration File:
Load
Save
Save As
Database Selection
FASTA Database File:
Browse
Fixed Modifications
Hold CTRL For Multiple Selection
0. methylation of K
1. oxidation of M
2. carboxymethyl C
3. carbamidomethyl C
4. deamidation of N and Q
5. propionamide C
6. phosphorylation of S
7. phosphorylation of T
1 Modification Selected
Variable Modifications
Hold CTRL For Multiple Selection
0. methylation of K
1. oxidation of M
2. carboxymethyl C
3. carbamidomethyl C
4. deamidation of N and Q
5. propionamide C
6. phosphorylation of S
7. phosphorylation of T
2 Modifications Selected
Proteolytic Enzyme
Enzyme:
Trypsin
Max Missed Cleavages:
2
Mass Tolerances
Precursor Ion:
60
ppm
Fragment Ion (Da):
0.5
Fragment Ion Types
Fragment Ion Type 1:
b
Fragment Ion Type 2:
y
Precursor Charge Bounds
Lower Bound:
2
Upper Bound:
4

**Figure 3:**


SearchGUI 1.3
Task Editor
Parameters Editor
Advanced Parameters Editor
E-value & Hitlist Length
E-value Cutoff:
100
Maximum Hitlist Length:
25
OMSSA Specific Parameters
Minimum Precursor Charge to Consider Multiply Charged Fragments:
3
Eliminate Charge Reduced Precursors in Spectra:
No
Precursor Mass Scaling:
Yes
Minimum Peptide Length:
Maximum Peptide Length: