1) Introduction

OMSSAGUI is an open source user interface for searching peak list files with the Open Mass Spectrometry Search Algorithm. It uses the Java Swing API, the XmlPullParser library and, obviously, the NIH's C++ implementation of OMSSA.

2) Installing

To install, just download the two JAR files from the website and save them to a place you will remember.

3) How to Use

The first thing you will need to do is download the formatdb.exe executable from the NCBI website. The executable is included in the NCBI's release of the BLAST package, which can be found here:

ftp://ftp.ncbi.nih.gov/blast/executables/release/2.2.18/

The specifically Windows version of this executable has been uploaded to this site for convenience.

Next, the OMSSA program must be installed. It can be downloaded from the OMSSA website at:

http://pubchem.ncbi.nlm.nih.gov/omssa/download.htm

Finally, you must move the formatdb.exe executable, to the \NCBI\omssa-2.0.2.win32 folder, which will be the same folder that the omssacl.exe executable is in. If you got the formatdb.exe executable from the NCBI website, it will be in the bin folder of your BLAST folder. If you got the formatdb.exe executable from this website, it will be wherever you saved it.

Once the OMSSA program is installed, this GUI can be run by clicking the icon for the jar folder.

When the software starts, a prompt will open for the path to the location of the OMSSA program. You will only have to input this path once, as long as you do not move the OMSSA folder. If you do move the folder, the GUI will prompt again for the path.

To run a search, you will need (1) a FASTA-formatted database file, and (2) an MGF-formatted, PKL-formatted, or DTA-formatted file which are the spectra from that same search.

There are a few options for accomplishing (1). First, you can export a subset database of your hits from a previous search using an outside parsing program, such as ProteomeSoftware's Scaffold program. In Scaffold, this can be done by a simple dropdown menu. Second, it can be done manually from the results of a Mascot search. This would entail recording the accession numbers of the hits from the Mascot search and downloading a corresponding list of entries from the online database repositories. For example, the Swiss-Prot website allows you to download a FASTA-formatted list of entries by uploading a text file which contains a list of accession numbers:

http://ca.expasy.org/sprot/sprot-retrieve-list.html

Finally, you can simply download a universal database in FASTA format from a website such as NCBI, ExPASy, or the EBI.

There are also a few options for accomplishing (2). First, the simplest way, again, is to use a

program such as Scaffold to export all your original spectra hits in mgf format. This is once again a dropdown menu in Scaffold. Second, it can be done manually by going back and retrieving the MGF files which were originally searched by Mascot. This would be the output of whatever program is being used as a converter between RAW files and MGF files, such as Mascot Daemon, Mascot Distiller, mzXML-Search, DTASuperCharge, or WIFFtoDTA. PKL and DTA files are generated by Micromass ProteinLynx and Sequest, respectively. Finally, there is a new tool from Richard Smith's group which may be used to do peak extract and charge state re-calculation from the command line, DeconMSn. It is available here:

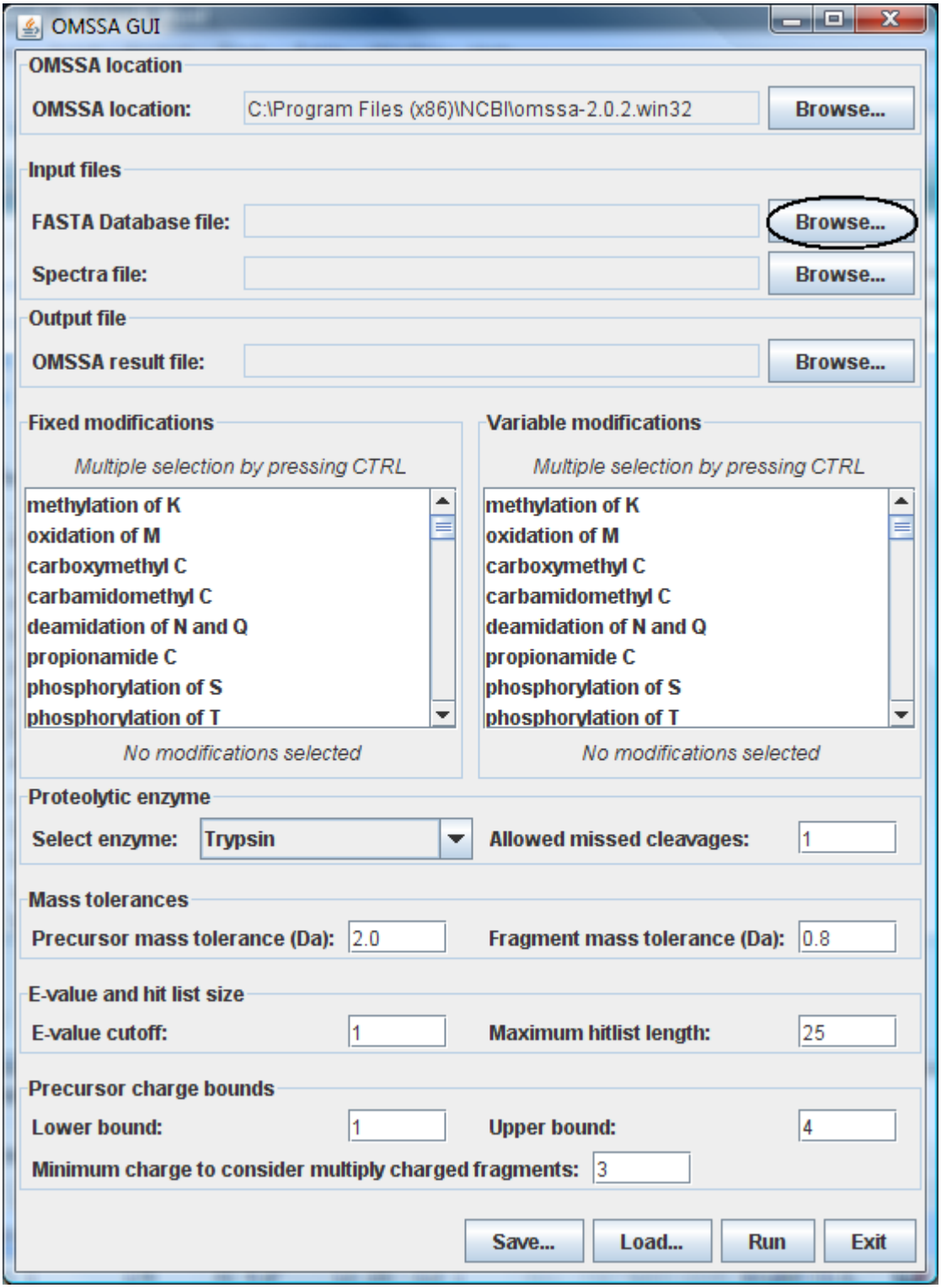
http://omics.pnl.gov/software/DeconMSn.php

This can create MGF, DTA, or PKL files, all of which can be read by OMSSA_GUI.

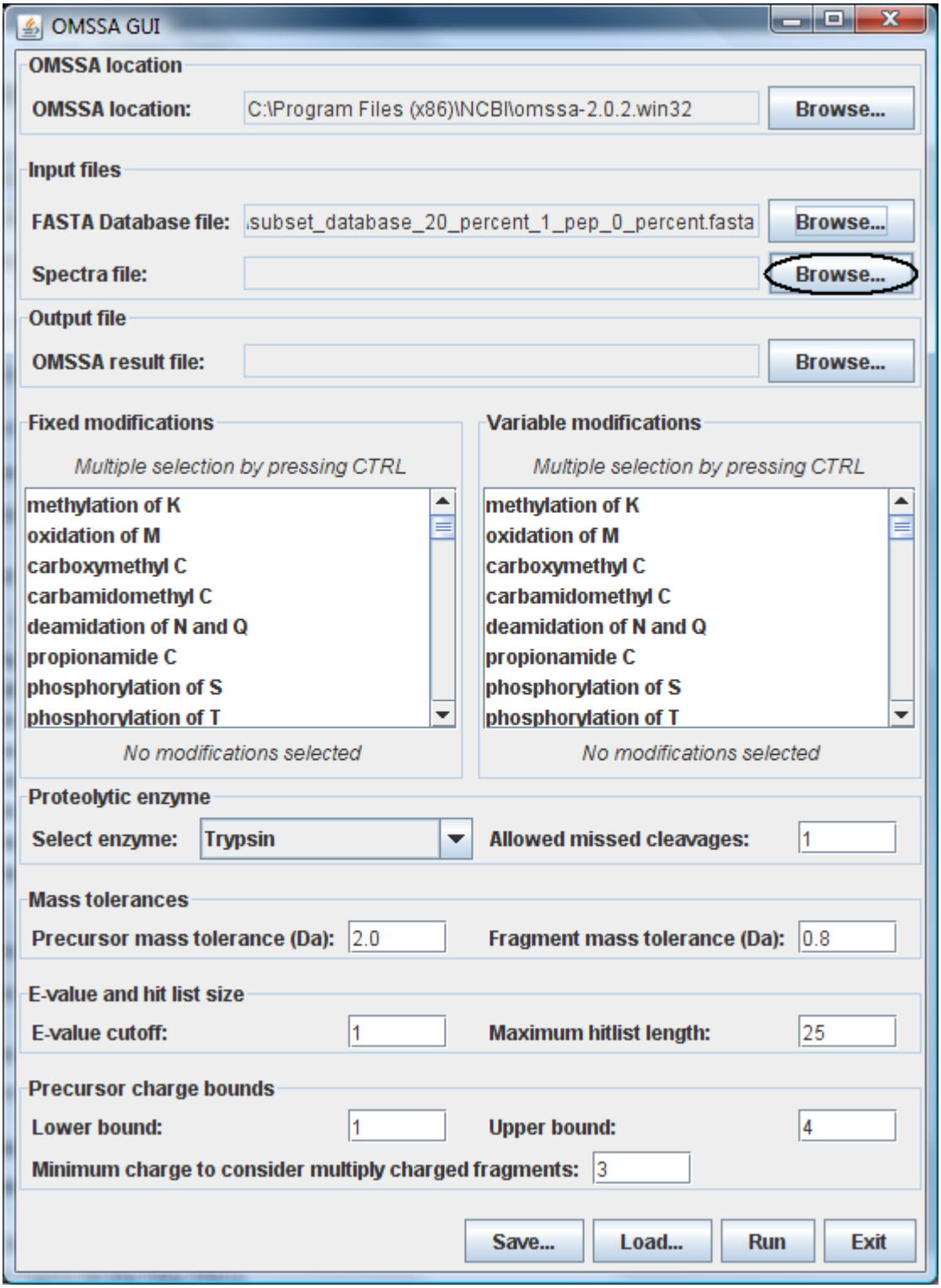
The rest of the search settings on the GUI will be filled in with defaults by the default_OMSSA_GUI.properties file. You can change these options according to your preference. If you want to change the defaults, you can edit the default_OMSSA_GUI.properties file. Of course, there are also menus for selecting the modifications you want to search.

Once you have a FASTA-formatted hits database and an MGF, PKL, or DTA-formatted spectra file, you can run a search by doing the following:

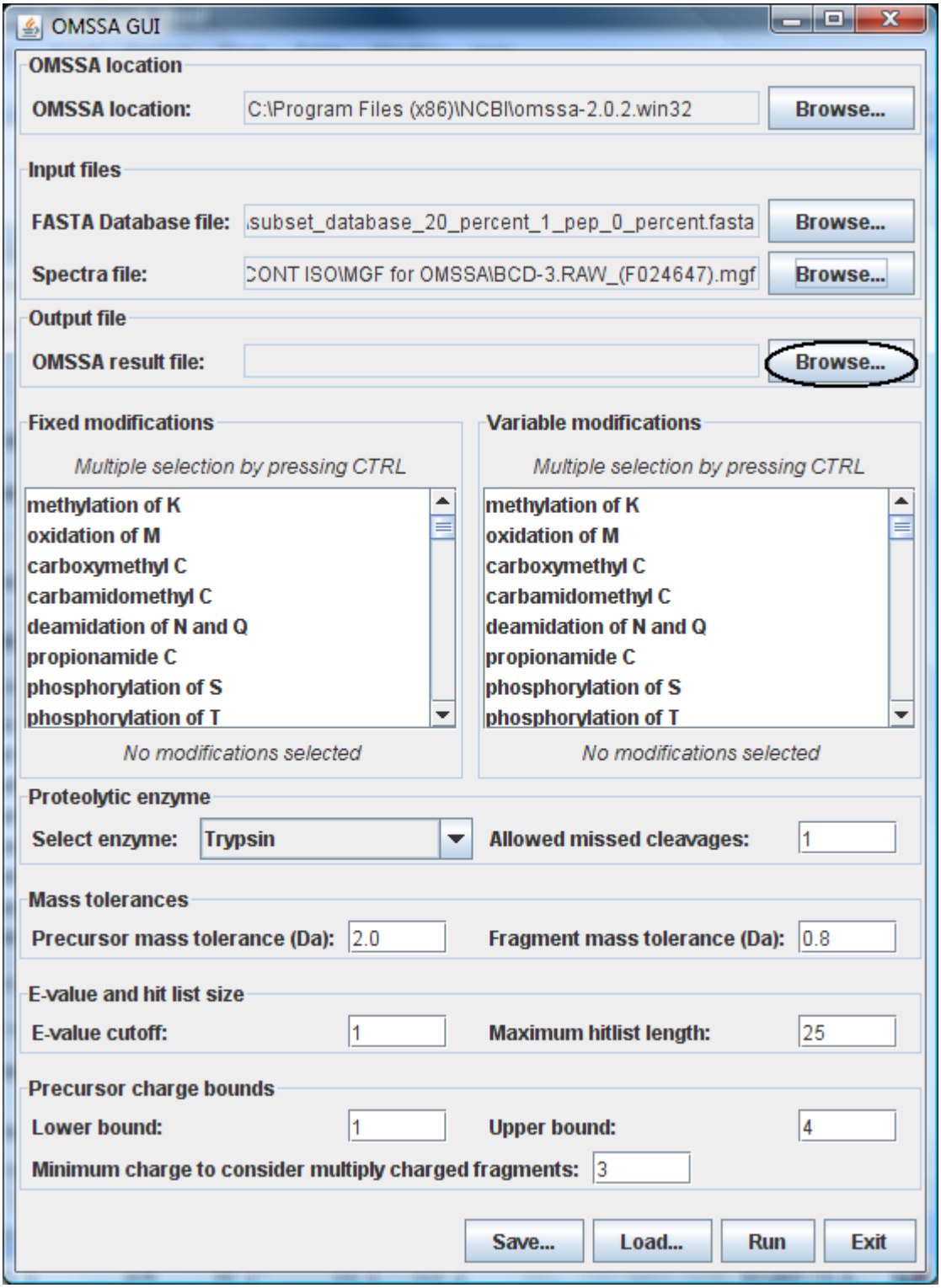
1) Input the database by clicking "Browse" for the "FASTA Database file:" line.



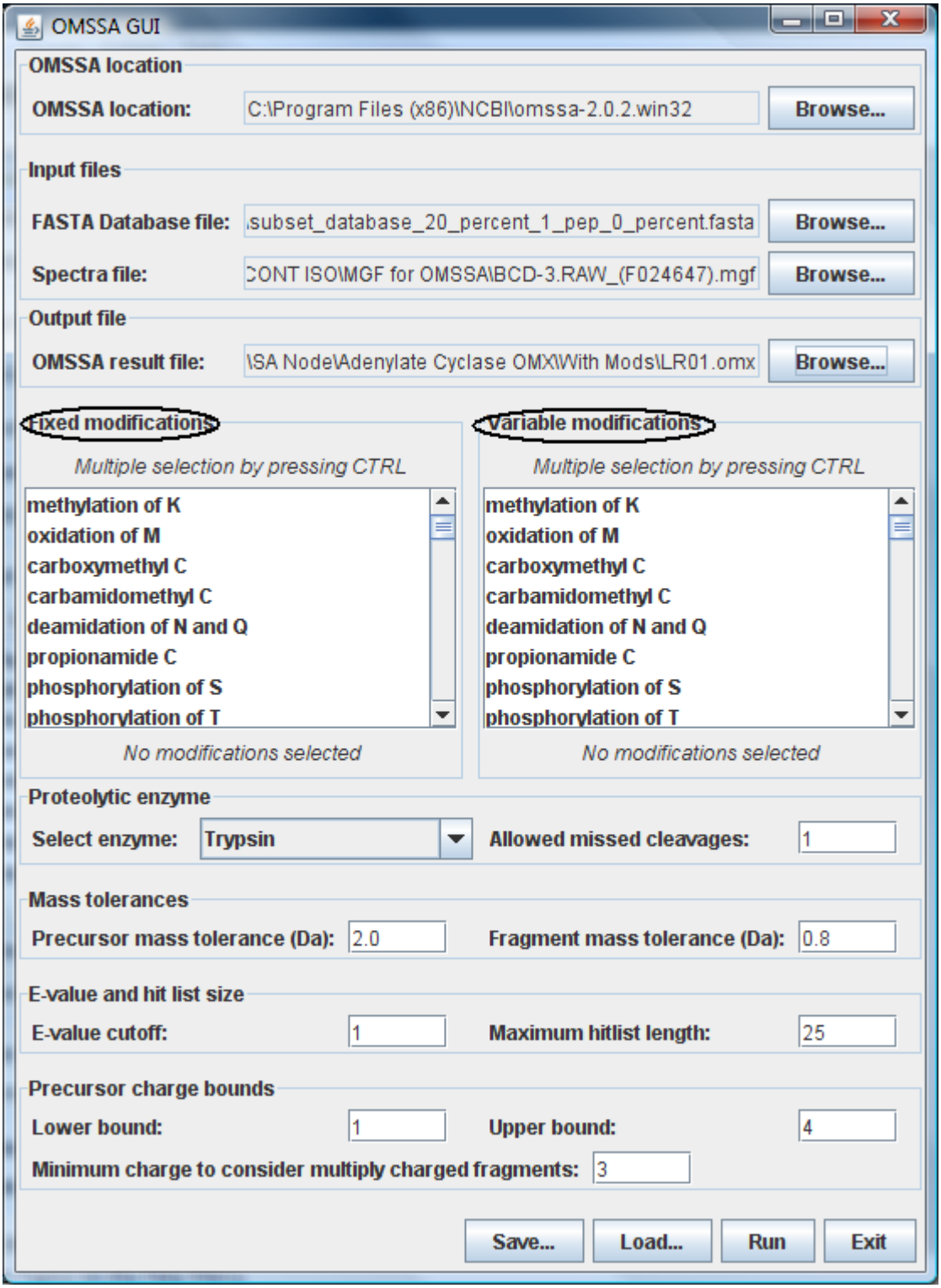
2) Input the spectra file by clicking "Browse" for the "Spectra file:" line.



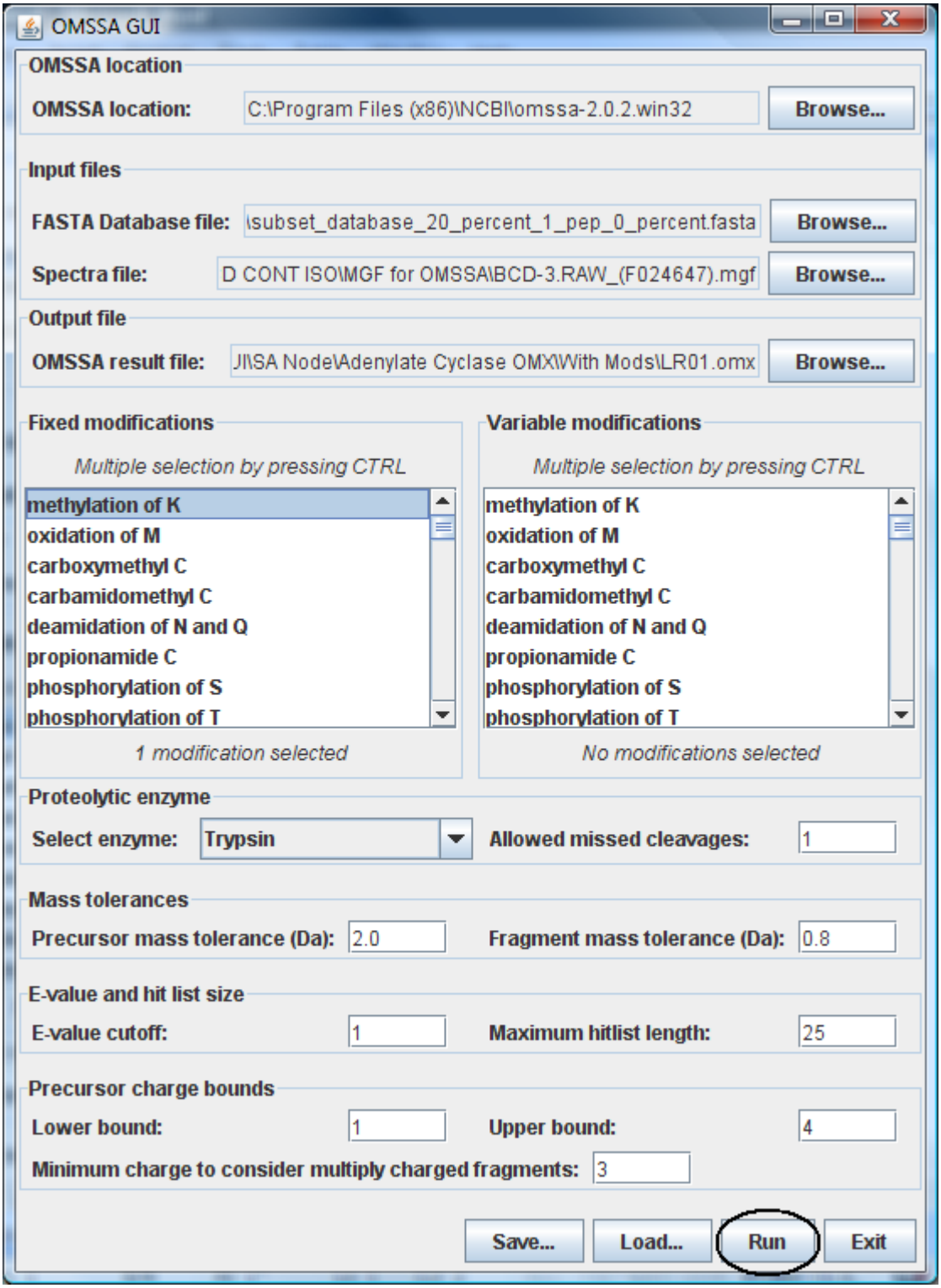
3) Select where you want the output of OMSSA to be saved by clicking "Browse" for the "OMSSA result file:" line.



4. Add any fixed or variable post-translational modifications.

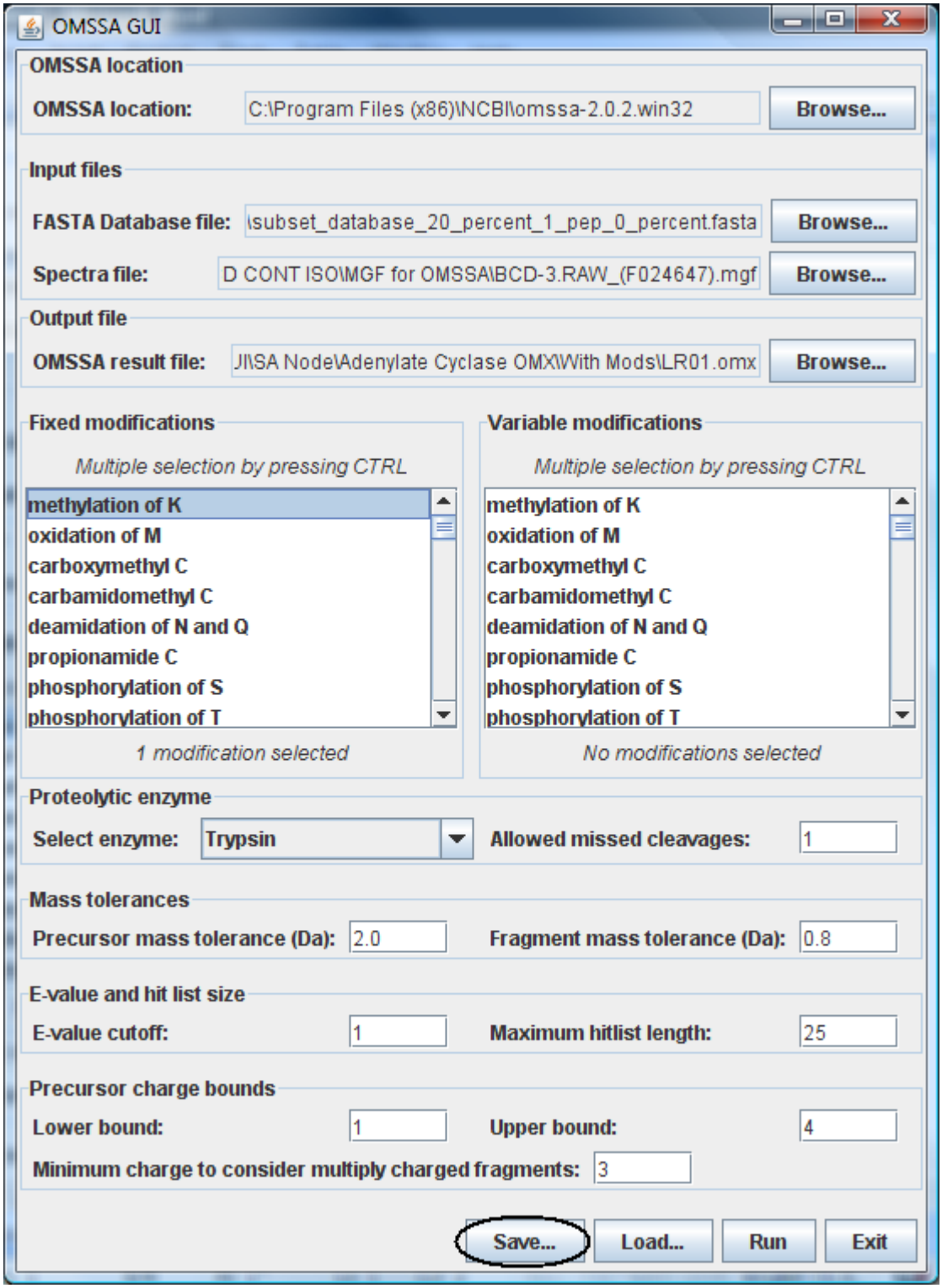


5. Check the rest of the settings, then click the "run" button to start the search.

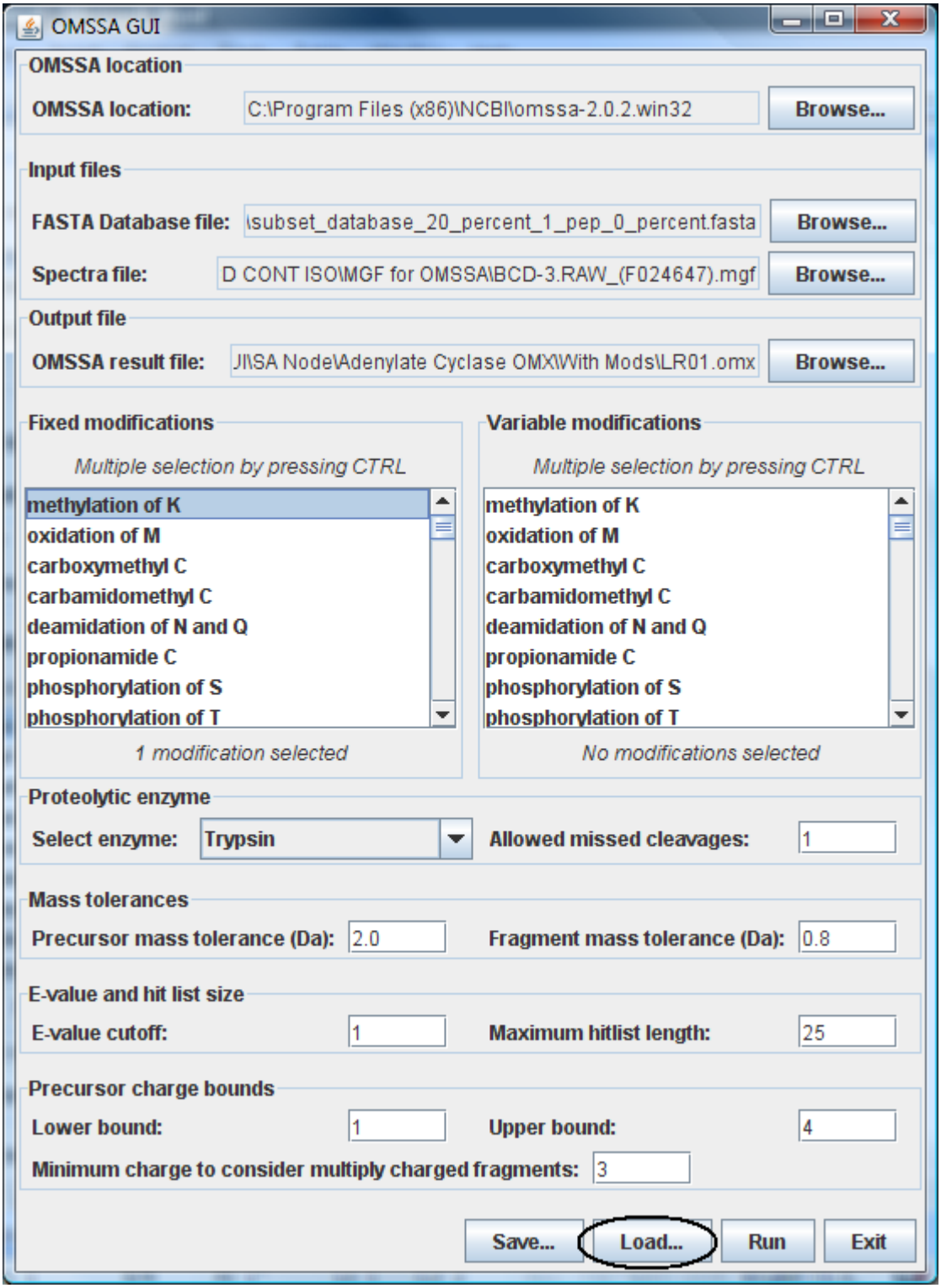


6. Once the search is complete, an OMSSA results file will have been written to the directory you selected. Go to that directory and open the file to view your results.

If you want to save the settings that you ran the search with, you can click "save" at the bottom of the GUI.



This will open a dialogue which will allow you to save a .PROPERTIES file, which will save your settings, the post-translational modifications you selected, and the database file you selected. To re-load the settings, just click "load" and select your .PROPERTIES file.



You can also write your .PROPERTIES file, as long as the file is in the following format:

DATABASE_FILE=

ENZYME=
FIXED_MODIFICATIONS=2. carboxymethyl C//5. propionamide C
VARIABLE_MODIFICATIONS=7. phosphorylation of T
MISSED_CLEAVAGES=
PRECURSOR_MASS_TOLERANCE=
FRAGMENT_MASS_TOLERANCE=
EVALUE_CUTOFF=
MAXIMUM_HITLIST_LENGTH=
PRECURSOR_CHARGE_LOWER_BOUND=
PRECURSOR_CHARGE_UPPER_BOUND=
PRECURSOR_CHARGE_TO_CONSIDER_MULTIPLY_CHARGED_FRAGMENTS=

The fields should be filled out with the values you want, after the equal sign. Note that the modifications should be separated with double slashes, as in the example here.

4) What the settings mean

Select Enzyme is the digestion enzyme you used.

Allowed missed cleavages is how many times the program will guess that the enzyme missed a cleavage site, before giving up on the peptide.

The mass tolerances should depend on the sensitivity of the mass spectrometry instrument.

E-value cutoff is the maximum e-value allowed in the hit list.

Maximum hitlist length is the maximum number of hits retained for each spectrum.

The lower bound and upper bound of precursor charge is the minimum and maximum precursor charge, respectively, that the software will look for. The minimum charge to consider multiply charged fragments is the minimum charge at which the software will guess that the precursor was multiply charged.

5) About the Software

This software is released as open source code by the JHMI NHLBI Bayview Proteomics Center and the European Bioinformatics Institute, and is under the Apache 2.0 Software License.