Peptide to Spectrum Matching

Shotgun proteomics relies on the assignment of vast number of spectra to theoretical peptides derived from a sequence database. Various search engines have been developed for this task, each with its own advantages and shortcomings. We are going to search the mgf file obtained in Chapter 2 against the database obtained in chapter 1 using OMSSA[**1**](#_ENREF_1) and X!Tandem[2](#_ENREF_2" \o "Craig, 2004 #46), two freely available search engines. The necessary spectrum and database files can be found in the resources folder.

Peptide 1

Peptide 2

Peptide 3

Peptide 4

Peptide 1

Peptide 2

Peptide 3

Peptide 4

Peptide 1

Peptide 2

Peptide 3

Peptide 4

Spectrum Collection



Protein Database

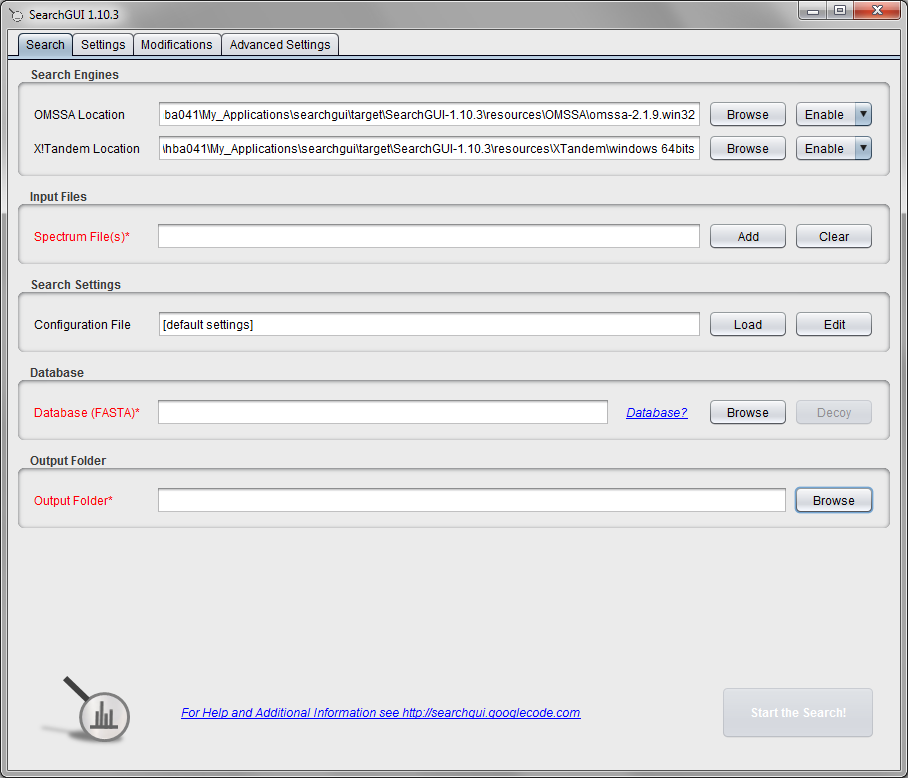


Results



Search Engine

Every search engine has its own specificity and it is recommended to study them on their respective web pages: [http://pubchem.ncbi.nlm.nih.gov/omssa](http://pubchem.ncbi.nlm.nih.gov/omssa/) and [http://www.thegpm.org/tandem](http://www.thegpm.org/tandem/). However, it is possible to use them together *via* a simple interface called SearchGUI[3](#_ENREF_3). SearchGUI for windows is provided in the software folder together with OMSSA and X!Tandem. For Mac and Linux versions, please see the SearchGUI web page: [http://searchgui.googlecode.com](http://searchgui.googlecode.com/). Start SearchGUI by double clicking on the file SearchGUI-X.Y.Z.jar (replace X.Y.Z with the current SearchGUI version number).

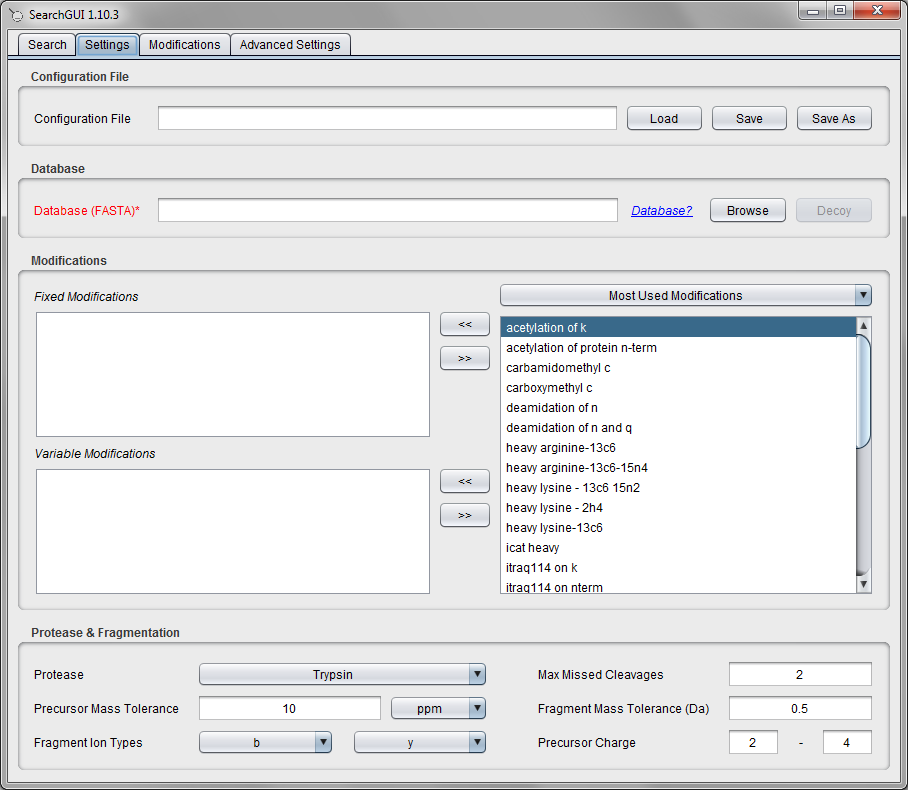


You will notice that the locations for OMSSA and X!Tandem are predefined. In fact, keen observers may have already noticed that these folders exist in the SearchGUI home folder. This means that when you have downloaded the SearchGUI zip file and unzipped it (which comprises the entire installation procedure), you have also already downloaded OMSSA and X!Tandem along with it!

*Is this legal? Can the SearchGUI authors do this? They did not make OMSSA or X!Tandem?*

In order to perform the search, we need to provide the spectra, the database and search settings which are experiment dependant. Load the previously created mgf file Velos005137.mgf (available in the resources folder). *Note that you can load multiple mgf files and even entire folders.*

We are now going to set the search settings in the Settings tab, click on ‘Edit’ in after the ‘Configuration File’ field (or simply select the tab at the top):



Note that at the top of this tab, you can load or save the configuration you will enter - this makes it easy to keep track of your parameters, and to reuse them *verbatim* later on. These can also be loaded from the Settings tab as a 'Configuration File'. For now, we will define the parameters by hand.

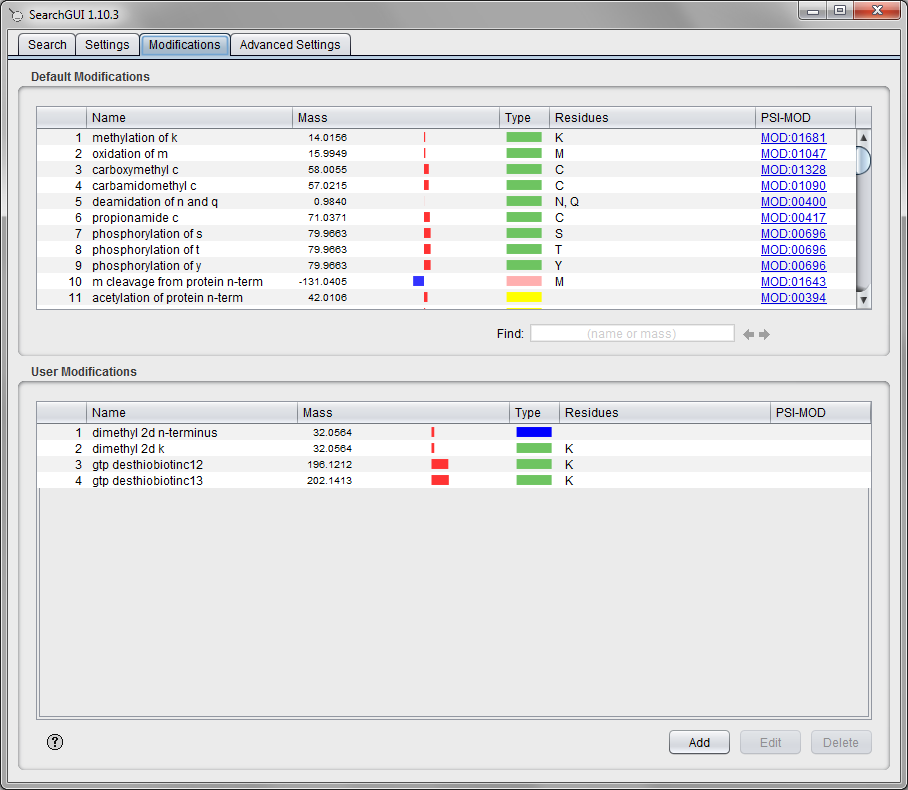
First, we need to specify the modifications: as fixed modifications choose carbamidomethyl c, and as variable modifications choose phosphorylation of s, phosphorylation of t, phosphorylation of y and oxidation of m. *Tip: CTRL + Click allows you to select multiple entries.* *Are these all the modifications you would expect for a standard shotgun experiment? How do you define which modifications are variable and which are fixed?*

Then we’ll need to choose the enzyme - leave it set at Trypsin, and keep the number of allowed missed cleavages at 2*. What is a missed cleavage? Why 2 and not 0 or 1?*

Keep the precursor ion mass tolerance at 10 ppm and the fragment ion mass tolerance at 0.5 Da. *How do we choose these values? What is the difference between using a mass tolerance in ppm or Dalton?*

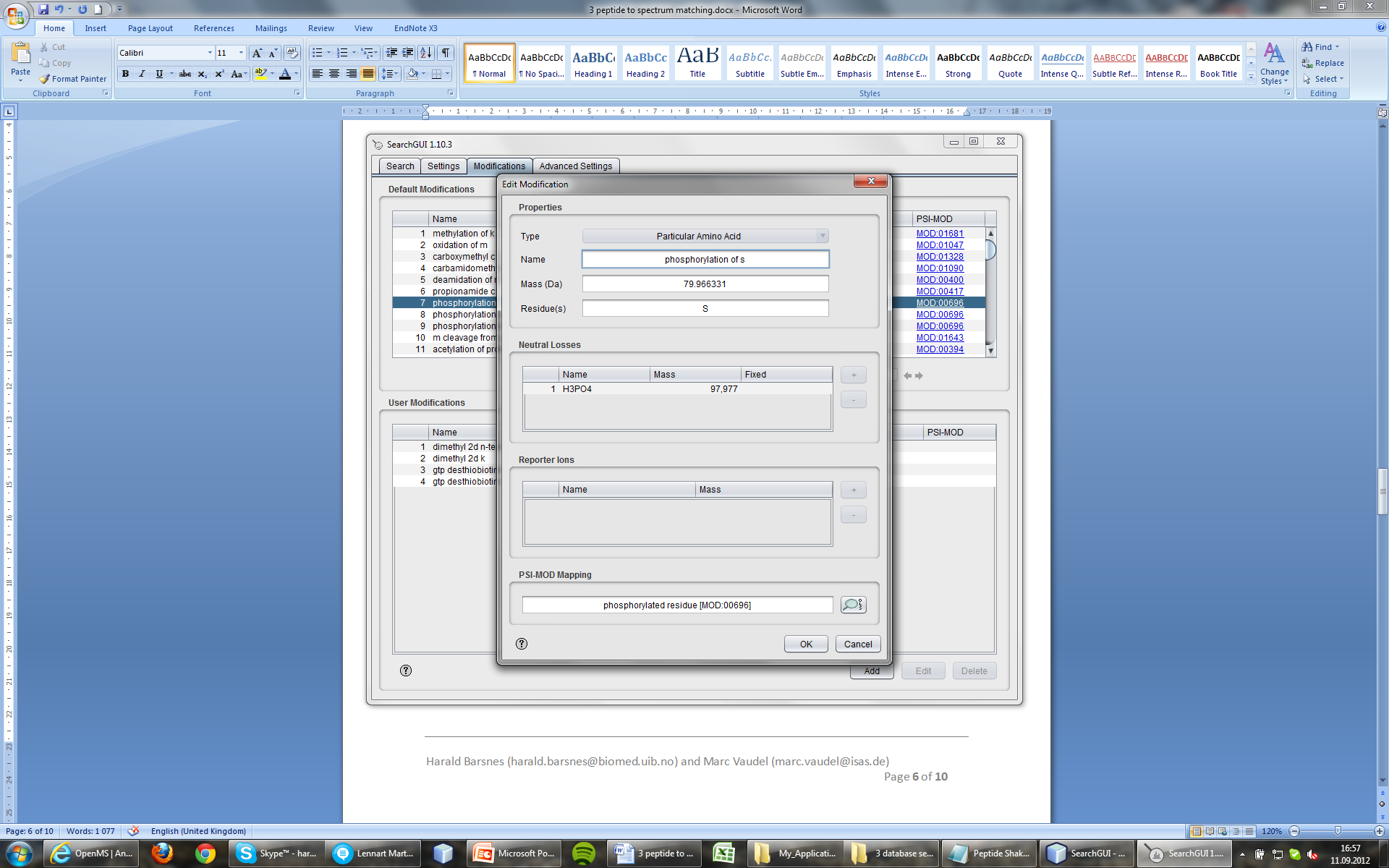
The fragment ion types and the charge bounds are fine as they are. *Why?*

Note that only the most commonly used modifications are displayed here. There are more modifications available in SearchGUI, and you can also set up your own modifications. Go to the Modifications tab and type ‘phosphorylation’. You will see that many modifications are available:

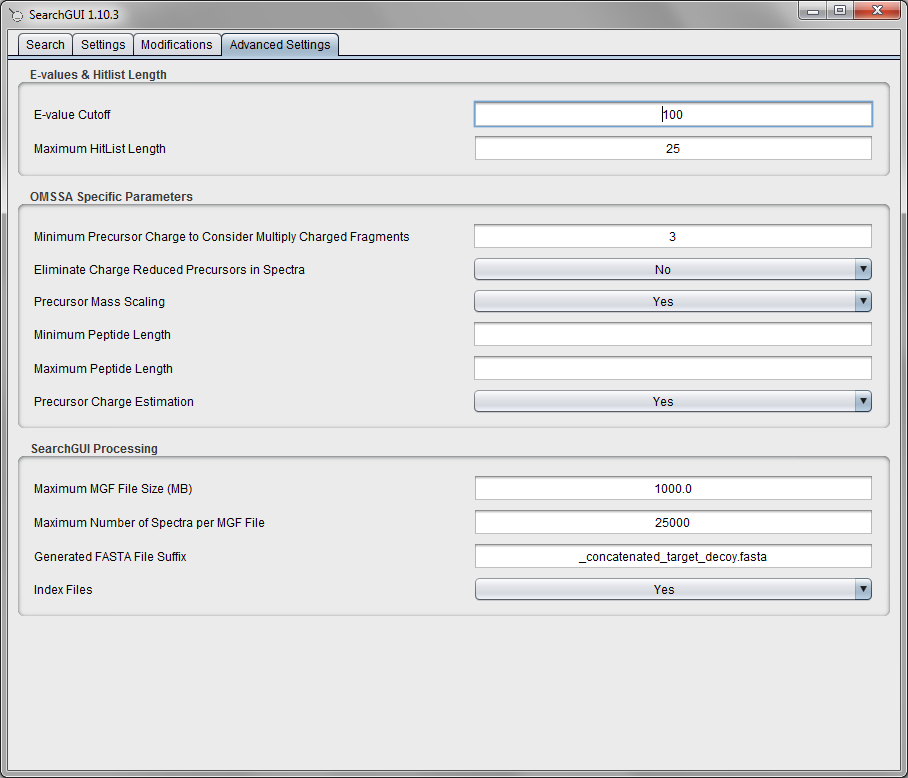


*What is the difference between the different phosphorylation possibilities?*

Double clicking on a modification will bring up the modification details:



The final tab contains advanced search settings, which we will not touch here. These settings are used for the fine tuning of the algorithms and should only be used by experienced users. The Advanced Settings tab looks like this:



Now we go back to the first tab (Search). Note that both search engines, at the top, are enabled - this means one can run the search using both OMSSA and X!Tandem, and get result files for each at the same time. However, to save some time we will just use OMSSA now. To disable the X!Tandem search, simply selected 'Disable' in the drop down menu to the right of the X!Tandem text field.

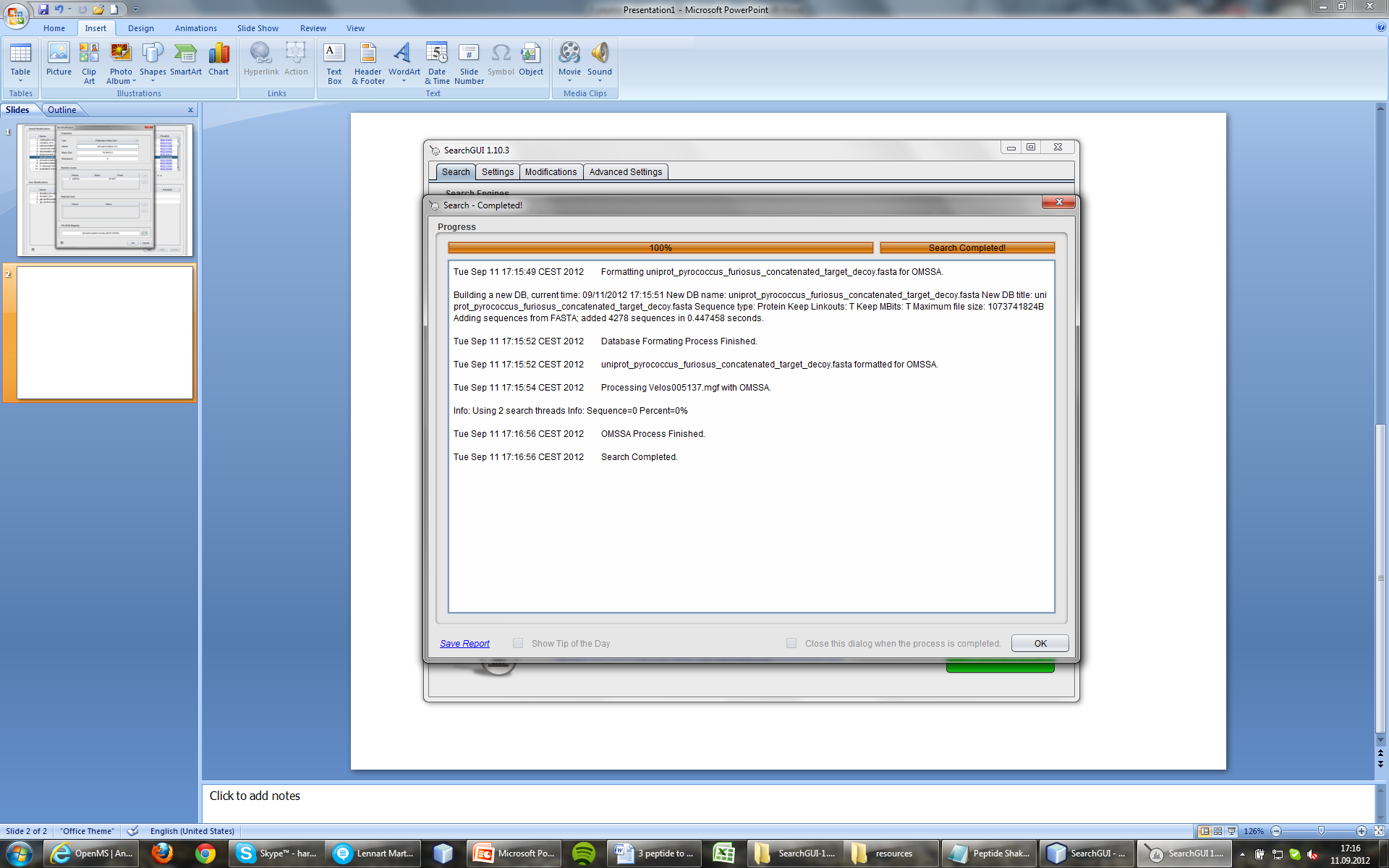
The X!Tandem search has already been performed for you though, and we will employ the results of both searches in the later analysis.

We need to specify the database to search against. Make use of the database generated in *Section 1* (also available in the resources folder). *How does the database used affect the results? Will we always find the same proteins? How does the size of the database affect the significance/score of the proteins we find?*

If you want to perform Target/Decoy searches, a button allows you to generate the corresponding database. When uniprot\_pyrococcus\_furiosus.fasta (located in the resources folder) is selected, SearchGUI offers the possibility to add decoy sequences, click on ‘Yes’. Note that the selected database has now changed. The role of this modified database will be further developed in the identification validation chapter.

Finally select an output folder and start the search by pressing the ‘Search’ button. A progress bar and scrolling text will keep you informed of the progress of the searches. *How does the size of the spectrum file affect the search time? What about the database size? The search parameters? Can all searches be performed on a standard desktop computer?*

A screenshot of the dialog after completion is shown below:



After completion, the output folder will contain several files, where the two most important are the output files for the search engines. The OMSSA output file will be called Velos005137.omx, while the X!Tandem output file will be called Velos005137\_[date] \_[time].t.xml. These files contain so-called Peptide to Spectrum Matches (PSMs) inferred by the search engines. We will see how to interpret these matches in the next chapter.

If you happen to encounter any issue with SearchGUI, consult the troubleshooting section at: <http://searchgui.googlecode.com>.

References

(1) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. Open mass spectrometry search algorithm. *J Proteome Res* **2004**, *3*, 958.

(2) Craig, R.; Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **2004**, *20*, 1466.

(3) Vaudel, M.; Barsnes, H.; Berven, F. S.; Sickmann, A.; Martens, L. SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics* **2011**, *11*, 996.