Tutorial for using the TPP

1. **Download TPP**

Download and install the TPP on Windows system. Follow the Windows Installation Guide making sure that you select to download the latest version of TPP from TPP’s Sourceforge download site <http://sourceforge.net/projects/sashimi/files/>

1. **Log into Petunia, the TPP GUI**

Log in to *Petunia* by double-clicking on the **Trans-Proteomic Pipeline** (**TPP**) flower icon on your computer’s Desktop or through the Start menu. Use the credentials guest and guest as user name and password to log in.

1. **Download and install the test data, database, and parameters file**
2. **Test data** – To convert the Thermo Fisher LTQ.XL raw data format (\*.RAW) to the open mzML format

Install the free Thermo MS File Reader: <http://sjsupport.thermofinnigan.com/public/detail.asp?id=703>

Then download the raw data as a zip file

Make the folder: C:\Inetpub\wwwroot\ISB\data\practice\

Make the folder: C:\Inetpub\wwwroot\ISB\data\practice\tandem

Then copy or move the data files(\*.raw) into the folder

C:\Inetpub\wwwroot\ISB\data\practice\tandem

1. **Database**

You will need to download a FASTA database. FASTA is a specific format for each protein sequence found in the database. Databases are compiled by the NCBI ([National Center for Biotechnology Information](http://www.ncbi.nih.gov/)). Search Google with the words ncbi entrez and click on [www.ncbi.nlm.nih.gov/Entrez](http://www.ncbi.nlm.nih.gov/Entrez) (Entrez cross-database search). Scroll down and find and click onto the Protein: sequence database. Type in human as the search term and click Search. On the right side of the page, scroll down and find the lise that says Top Organisms [Tree]. Click on the number to the right of the word homo sapiens. This will bring up a list of protein sequences. It may take a few seconds for your computer to bring up the list. Click on the Send to arrow. Click on File. For Format, click FASTA. Click Create file. You will download a .gz file which your computer will unzip with your computer’s unzip software. The usable database file will need to have a \*.fasta file name. Follow the same directions for other species or groups of species. Download the database over night since it may take a few hours. Save the database to C:\Inetpub\wwwroot\ISB\data\dbase. You can change the file name of the database to make it user friendly.

1. **Parameters file**

If you have not already downloaded the parameters file, download the parameters file at:

<ftp://ftp:a@ftp.peptideatlas.org/pub/PeptideAtlas/Repository/TPP_Demo2009/TPP_Demo2009_tandemParams.zip> and unzip. Copy or move the tandem parameters file tandem.xml into the folder C:\Inetpub\wwwroot\ISB\data\parameters. Note if the tandem parameters file tandem.xml is already in C:\Inetpub\wwwroot\ISB\data\demo2009\tandem that should work as well; just remember where it is.

1. **Convert raw data file to the mzML format.**
2. Select the analysis pipeline Tandem in the pop-up menu below the Welcome message.
3. Move over the Analysis Pipeline (tandem) portion of the navigation links near the top of the Home page (just below the – home title). Select the mzML/mzXML item in the pop-up menu.
4. In Section 1, make sure that Thermo.RAW is chosen from the pop-up menu.
5. In Section 2, Click add the .RAW file (s) containing your file name(s) using the path of c:inetpub/wwwroot/ISB/data/practice/tandem/your\_filename.RAW. Click the Select button.
6. In Section 3, do not change anything.
7. In Section 4 (display page will call this Section 3), Click Convert to mzML.
8. A wait page will appear with a Session ID number, Job, Location, Start date/time, Actions, Status, and Output. Under Status, a red box will appear saying “running”. Once the job is done the Status box will turn orange and say “\*finished”. It may take 5 min to 30 min to convert one .RAW file. Once the orange box appears, under Output click Refresh. Under the Command section, you can see whether the conversion was successful. You can also see the output filename. It will be:

c:inetpub/wwwroot/ISB/data/practice/tandem/your\_filename.mzML

Under the output file, you can click on the Pep3D link to get an image of the raw file chromatogram.

Note: If you get a filed Return code 13568 message under the Command section, you are most likely missing a Windows component. For example the Windows MSVCP100.dll file was not found when I first attempted this task. It worked after I downloaded the file from Microsoft Windows.

5. **Search data with X!Tandem** (More info at http://www.thegpm.org/TANDEM/api/)

1. Make sure that Tandem is selected in the analysis pipeline pop-up menu below the Welcome in the home page.
2. Go to Analysis Pipeline (Tandem) and click on Database Search in the pop-up menu.
3. An ISB/SPC Trans Proteomic Pipline – runtandem page will appear.
4. In Section 1. Specify mzXML Input files click Add files. Click practice. Then click tandem. Select your\_filename of interest. It will have .mzML at the end of the file name. Click Select.
5. In Section 2. Specify Tandem Parameters File, click Add Files. The tandem\_params.xml file should be located at C:\Inetpub\wwwroot\ISB\data\parameters. Select the tandem\_params.xml parameters file. This file contains two modifications, the C modification due to alkylation with IAA (57.02146) and The M modification due to oxidation (15.994915). The mass tolerance is set to -2 to 4 Da. Maximum missed trypsin cleavage sites is set to 2.
6. In Section 3, Specify a sequence database. Click Add Files and download the database to be used. For this example, c:Inetpub/wwwroot/ISB/data/dbase/human\_ref.fasta will be used.
7. In Section 4, Options make sure the Convert output files to pepXML is checked.
8. In Section 5. Search! Click Run Tamdem Search. The Command Status Jobs tab will appear and the program will say running in the red box until the orange finished box appears. This may take up to 30 min. Click Refresh on the output. You will have two output files c:Inetpub/wwwroot/ISB/data/practiced/tandem/your\_filename.tandem

c:Inetpub/wwwroot/ISB/data/practiced/tandem/your\_filename.tandem.pep.xml

1. You must convert the your\_filename.tandem file to a PepXML file for use in downstream analysis. If you did not check to Convert output files to pepXML in Section 4 described above, you can still do so by following these directions. Making sure that Tandem is selected in the analysis pipeline pop-up menu below the Welcome in the home page, go to Analysis Pipeline (Tandem) and click on pepXML in the pop-up menu. Select the file c:Inetpub/wwwroot/ISB/data/practiced/tandem/your\_filename.tandem

Click on Convert to PepXML and the file c:Inetpub/wwwroot/ISB/data/practice/tandem/your\_filename.tandem.pep.xml will be generated.

5. **Search data with SpectraST (Optional)**. SpectraST is a search engine that compares acquired specta against a library of pre-identified spectra whose peptide sequences have been assigned. In spectral searching, a spectral library is compiled from a large collection of previously observed and identified peptide MS/MS spectra. The unknown spectrum can then by identified by comparing it to all the candidates in the spectral library for the best match. First the correct spectral library must be downloaded.

1. Make sure that SpectraST is selected in the analysis pipeline pop-up menu below the Welcome in the home page.
2. Go to Analysis Pipeline (SpectraST) and click on SpectrST Tools in the pop-up menu.
3. Click on the Download Spectral Libraries in the pop-up menu.
4. This will bring you to a page that shows a list of Spectral libraries. Select the spectral library database that you want and click Download Spectral library. The library file will be large (1 GB) so this download may take many hours.
5. Search the spectral library that was downloaded by moving to Analysis Pipeline (SpectraST) and click on SpectrST Search in the pop-up menu.
6. In Section 1, select the mzML data file. You can find it at:

c:/Inetpub/wwwroot/ISB/data/practice/tandem/your-filename.mzML

In Section 2, select the spectral library that you downloaded under dbase\speclibs. You can find it at c:/Inetpub/wwwroot/ISB/data/dbase/speclibs/spectral\_library\_filename.splib

1. For Section 3, select the fasta sequence database located under dbase. You can find it at:

c:/Inetpub/wwwroot/ISB/data/dbase/sequence\_database\_filename.fasta

1. Leave the other options at their default and click on Run SpectraST to initiate the search.

6. **Validation of Peptide-spectrum assignments with Peptide Prophet** (assigns a probability to each peptide-spectrum match, must be perfomed to generate ProteinProphet file later on)

1. Go to the Analysis Pipeline (SpectraST). Click on Analyze Peptides on the pop-up menu. This will bring up the xinteract page, a general utility page that allows one to launch PeptideProphet (as well as other programs).
2. Select the c:Inetpub/wwwroot/ISB/data/practice/tandem/your\_filename.pep.xml file.
3. Under the PeptideProphet Options, find and select the option to Use accurate mass binning if this is high-accuracy data. Leave all other options set to default. Make sure that the RUN PeptideProphet box is checked.
4. Click on Run XInteract at the bottom of the page to run PeptideProphet.
5. Once the command finishes running, click on the view results link that appears in the Command Status box to view and analyze the results. The file will look like: c:/Inetpub/wwwroot/ISB/data/practice/tandem/interact.pep.xml
6. Click on PepXML and results can be viewed. This brings up an IMG:PepProphet page. (<http://tools.proteomecenter.org/wiki/index.php?title=Image:PeptideProphet.JPG>). On this page, in the sorting pop-up menu, click probability. Then click the desc button to the right and click Update Page. This will sort the list in descending order base on probabilities. The identifications at the top of the resulting list are most likely to be correct.
7. Click on any hypertext link under the PROB column for any probability. This brings up a details page IMG:PlotModel (<http://tolls.protemoecenter.org/wiki/index.php?title=Image:PlotModel.JPG>) which shows graphically how successful the modeling was. In the upper pane, it is desirable for the red curve (representing sensitivity) to hug the upper right corner, and for the green curve (representing error) to hug the lower left corner. The lower pane shows how well the data (black line) follows the PeptideProphet modeling for each charge state. The blue curve describes the modeling of the negative results, and the purple one shows the positive results. If thees two curves are well separatated and fit the black line well, the analysis for that charge state was successful.
8. If you performed SpectraST, you can go back and run this same analysis using the SpectraST results. If you want to run the iProphet program, you will have to perform SpectraST, because iProphet will compare the two analysis from Peptide Prophet and SpectraST.

7. **Further peptide-level validation with iProphet (Optional)** – iProphet will combine the analysis of the data file from PeptideProphet and SpectraST analyses.

1. Go to the Analysis Pipeline (SpectraST). Click on Combine Analysis on the pop-up menu. This will bring up the iProphet interface page.
2. Select the c:Inetpub/wwwroot/ISB/data/practice/tandem/your\_filename.pep.xml file.
3. Select the c:Inetpub/wwwroot/ISB/data/practice/practice/your\_filename.spectrast file.
4. Under Output File and Location, make sure the File path (folder) is set to Select the c:Inetpub/wwwroot/ISB/data/practice. Edit if necessary.
5. Leave all other options set to default.
6. Click on Run InterProphet at the bottom of the page to run IProphet.
7. Once the command finishes running, click on the view results link in the Command Status box to view and anlyze the results. A IMG:iprophet output file will appear. (<http://tools.proteomecneter.org/wiki/index.php?title=Image:iprophet.JPG>)

8. **Peptide Quantitation with ASAPRatio** – Toll for measuring relative expression levels of peptides and proteins from isotopically-labeled (ICAT, SILAC) samples. Use the tutorial from tools.proteonecenter.org/wiki/index.php?title=TPP\_Demo2009 to carry out this program. Other Quantitation tools available on the TPP are XPRESS and Libra which also measure isotopically-labeled samples. SuperHirn is a program that can be found in the SPC Tools under the heading LC-MS Analysis. It does not require isotopically labeled samples.

9. **Protein-level validation with ProteinProphet** – provides statistical validation of protein identifications based on PeptideProphet results generates interact.pept.xml file that can be used in APEX for quantification of spectral data from a mass spec run.

1. Go to the Analysis Pipeline (SpectraST). Click on Analyze Proteins on the pop-up menu. This will bring up the ProteinProphet interface page.
2. Select the c:Inetpub/wwwroot/ISB/data/practice/tandem/interact.pep.xml file. Make sure only one file is selected for analysis.
3. Leave all other options set to default values.
4. Click on run ProteinProphet at the bottom of the page to run ProteinProphet.
5. Once the command finishes running, click on the view results link in the Command Status box to view and analyze the results generated in an interact.prot.xml file. Protein groups are sorted in descending order by Probability so that the groups at the top of the page are the most confident identifications. The protein probabilities are the red numbers listed next to each protein group. The file is IMB:Protxml (http://tools.proteomecenter.org/wiki/index.php?title=Image:Protxml.JPG).