# PLGS Threshold Inspector : Instructions for use with PLGS3.0.3

## Contents

## About PLGS Threshold Inspector

PLGS Threshold Inspector (PLGS-TI) is a tool for automatically finding the optimal low energy (LE) and Elevated energy (HE) thresholds for MSe data processing in PLGS3.0.3. PLGS-TI was designed to make use the MSe processing and search executables in PLGS3.0 and later, and the instructions below can also be used with earlier versions of PLGS3 simply by changing the file paths to reflect the PLGS version number that you have access to.

PLGS Threshold Inspector can be downloaded from here :
*http://sourceforge.net/projects/plgsthresholdin/?source=directory*

PLGS Threshold Inspector is distributed as freeware. It is not covered by Waters or Nonlinear Dynamics support packages.

Similarly these guidelines for using PLGS Threshold Inspector are informal advice and do not imply official support for PLGS-TI. These guidelines are not an official Waters Corp document, and neither Waters or Nonlinear Dynamics are liable for any errors contained in this document.

## Why use PLGS Threshold Inspector?

- It is very important to optimise Apex3D parameters before processing your MSe or HD-MSe data
    - There is no universally applicable low energy (LE) and elevated (HE) settings for PLGS3.0.3 MSe data processing. The default settings are correct for the Waters system suitability sample, but not necessarily optimal for other samples.
    - Slightly counter-intuitively, setting parameters low 'to be on the safe side' does not result in more identifications. Setting the low and/or elevated energy parameters too low will detect 'noise' peaks. At the search step these peaks are penalised, resulting in fewer proteins being identified. In an extreme example of this, Waters support have seen data that gave 70 identifications with default parameters, and 3500 identifications (in PLGS) with optimised parameters.
    - Detecting and processing 'noise' peaks makes processing slower than it need be.
    - In some circumstances the extra peak processing required by the low and elevated energy threshold being too low may overload the processing resources of the PC, resulting in a PLGS3.0.3 crash.
- Optimising MSe processing parameters manually is incredibly tedious.
- PLGS Threshold Inspector does it automatically .

## Requirements

To use PLGS-TI you will need the following :

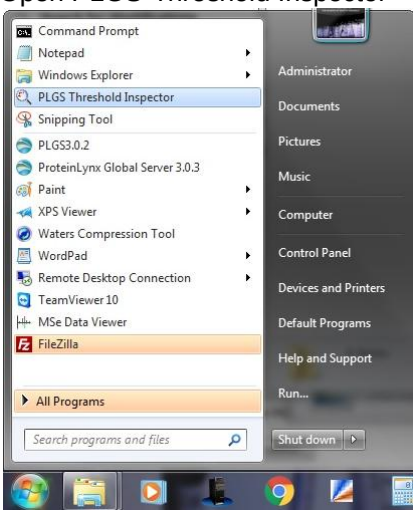
I. An appropriate MSe .raw data file
II. PLGS3.0.3 (you need to access the Apex3D64.exe, Peptide 3D.exe and iaDBS.exe files)
III. A workflow search file (an example is provided)
IV. A sequence database (fasta file) that is appropriate for the sample used to generate the raw data and referenced by the workflow file.
V. A PC running Windows 7 64 bit Professional Edition and that meets the PLGS3.0.3 minimum recommended specifications.

## Setting up PLGS-TI using PLGS3.0.3 executables

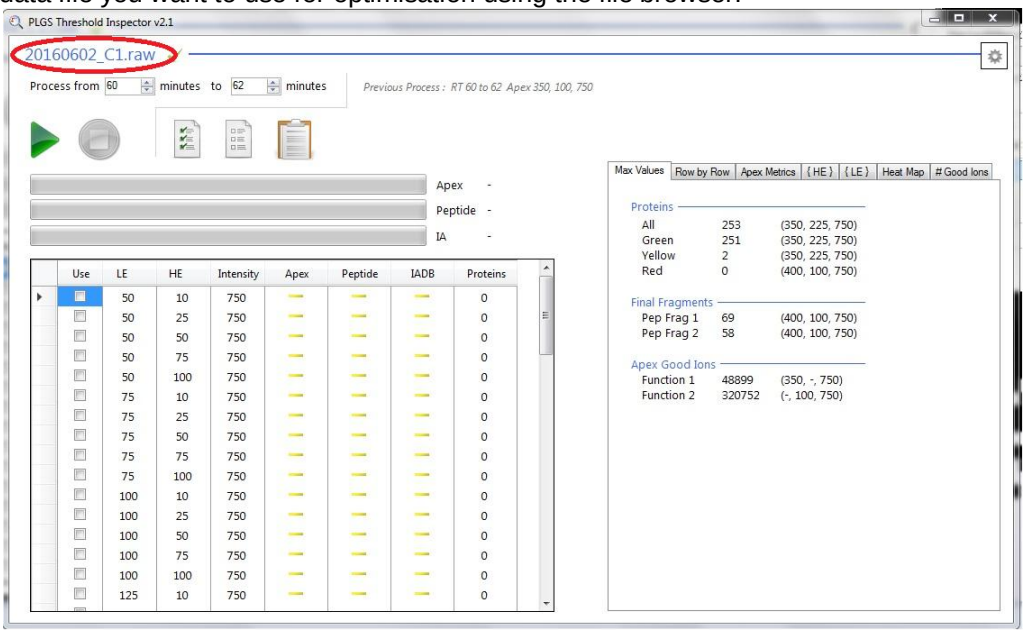
1. Download and install PLGS Threshold Inspector
2. Open the embedded xml file below, and save it to your documents folder

Default_MSe_Search_Workflow.xml

3. Open PLGS-Threshold Inspector

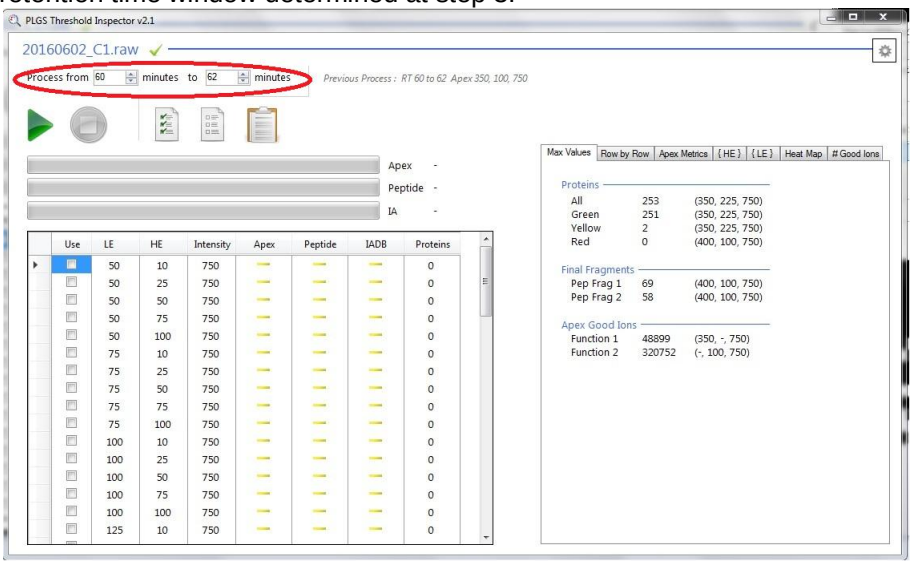



4. Click on the icon at the top left of the PLGS Threshold Inspector window, and select the raw data file you want to use for optimisation using the file browser.

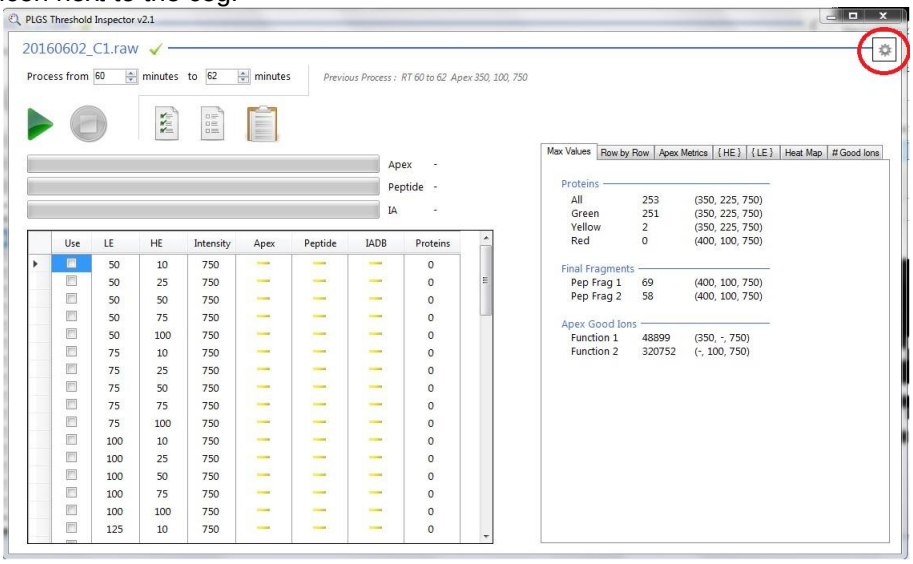



5. Have a look at the data in MassLynx>Chromatogram. Identify a suitable 5 minute retention time window in the most intense part of the data.

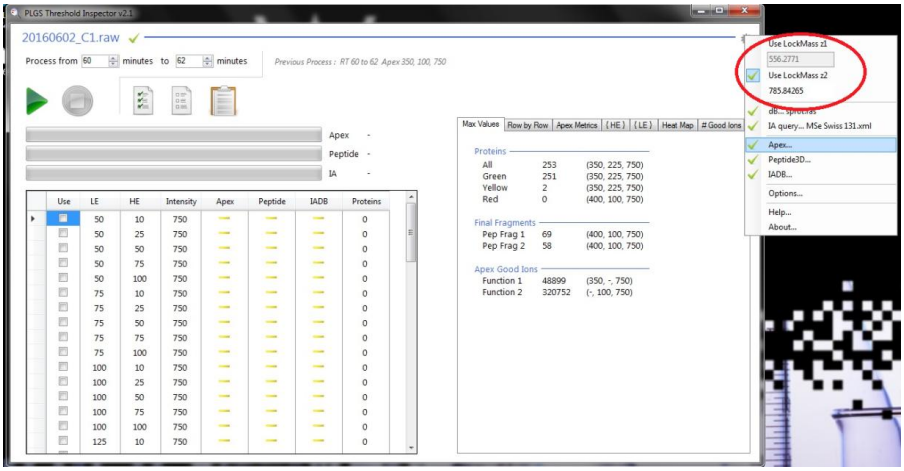
6. In PLGS-TI select "Process from *x* minutes to *y* minutes" settings based on the representative retention time window determined at step 3.



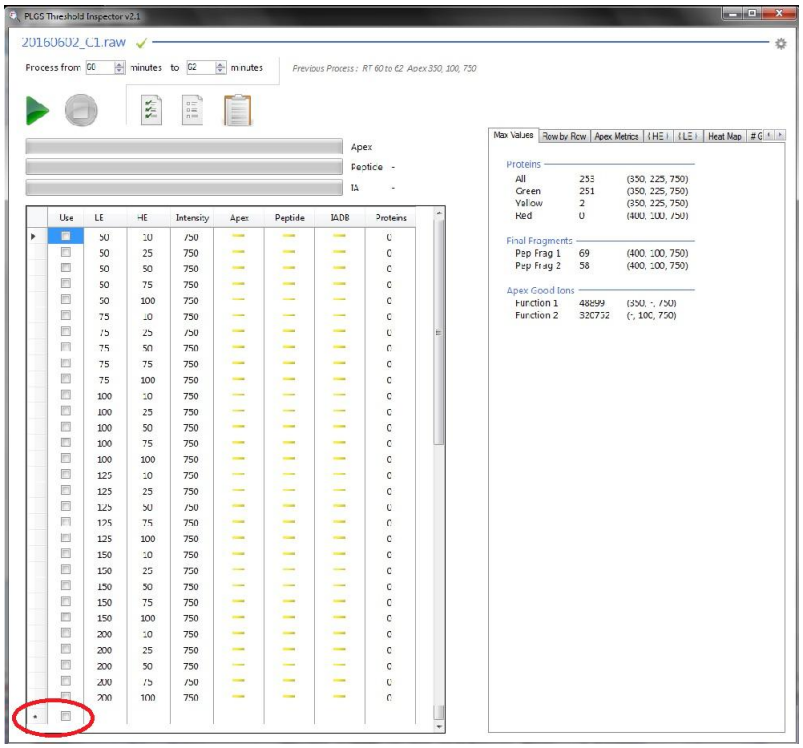

7. The cog icon at the top right of the screen opens a drop down menu of settings for PLGS-TI. If some of these haven't been set, or have been set incorrectly, there may be a flashing red icon next to the cog.



8. Click on the cog and set an appropriate lock mass for the selected data file

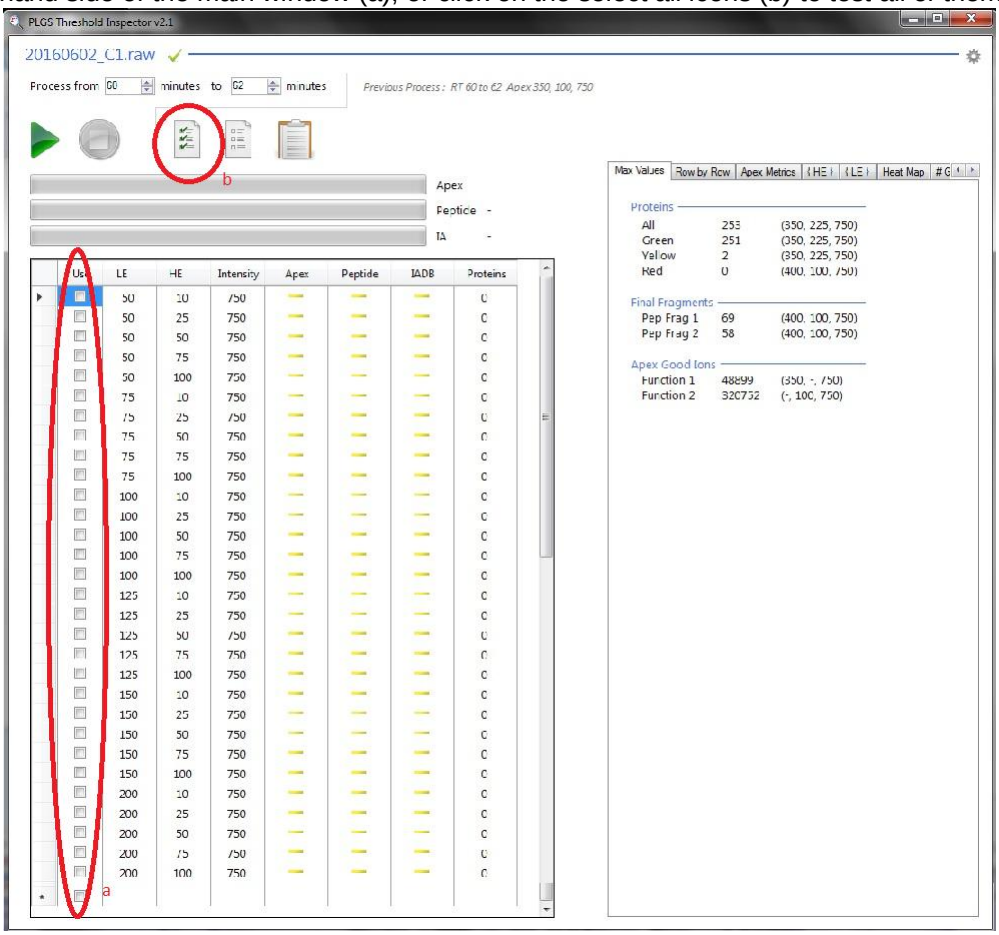



9. Click on the cog, and select Apex.... In the resulting browser window navigate *C:\PLGS3.0.3\lib\apex3d* and select Apex3D64.exe
10. Click on the cog, select Peptide3D. In the browser window navigate to *C:\PLGS3.0.3\lib\apex3d* and select Peptide3D
11. Click on the cog, select 'IADBs'. Navigate to *C:\PLGS3.0.3\bin* and select iadbs.exe
12. Click on the cog, select 'IA Query'. In the browser navigate to your documents folder and select the Default_MSe_Search_Workflow.xml file you copied there at step 2
13. Click on the cog, select dB.... This will open a browser window in which you select the fasta sequence database you want to use for parameter. The Default_MSe_Search_Workflow.xml uses the default swissprot databank installed with PLGS3.0.3. Navigate to *C:\PLGS3.0.3\demodbs* and select sprot.fas
14. LE and HE thresholds :
    a. By default PLGS-TI tests low energy (LE) thresholds ranging from 50 to 250, and high energy (HE) threshold settings from 10 to 100.
    b. Some customers find that the optimal setting for their data are higher than this.
    c. To add more combinations of parameters click on the an assigned check box at the bottom of the list of threshold combinations.



d. Click on the LE threshold setting of the new parameter group, and type a new number.
e. Click on the HE threshold setting of the new parameter group, and type a new number.

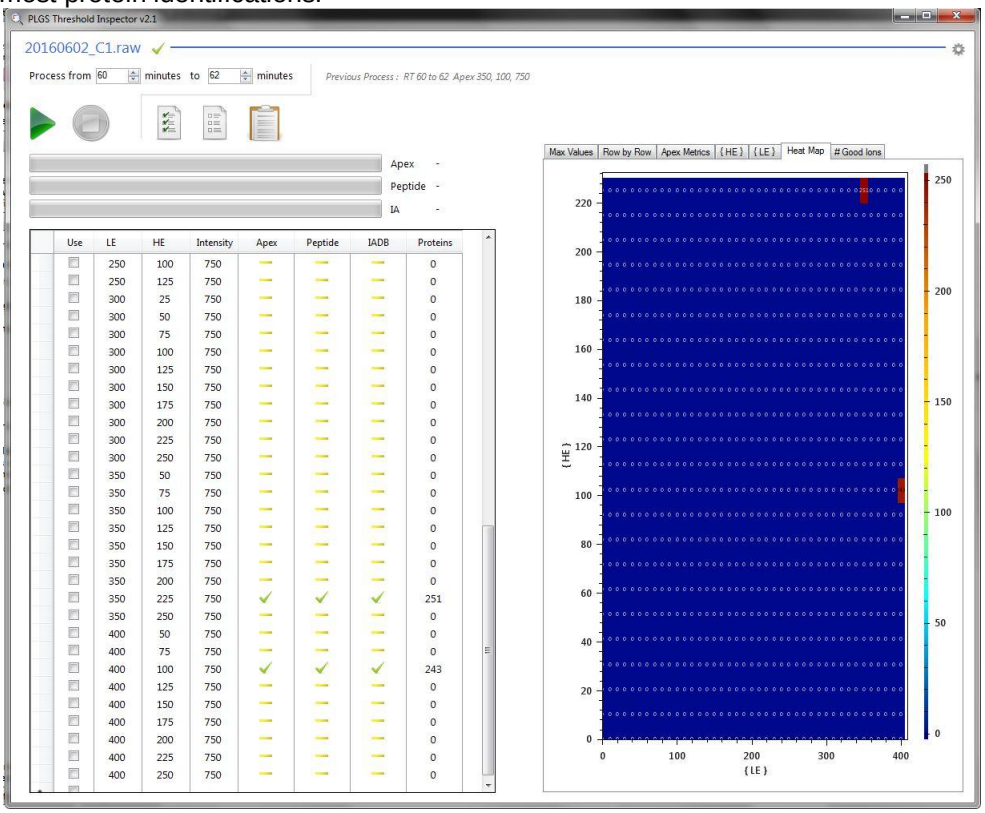
15. Intensity. PLGS-TI also uses a variable bin intensity threshold. The algorithms in PLGS3.0.3 do not use the Intensity parameter. This is only used in PLGS3.0.2 and earlier and is set to 750 by default. **Do not change this**
16. Select the parameter combinations that you want to test by clicking the tick boxes on the left hand side of the main window (a), or click on the select all icons (b) to test all of them.



## Running PLGS-TI :

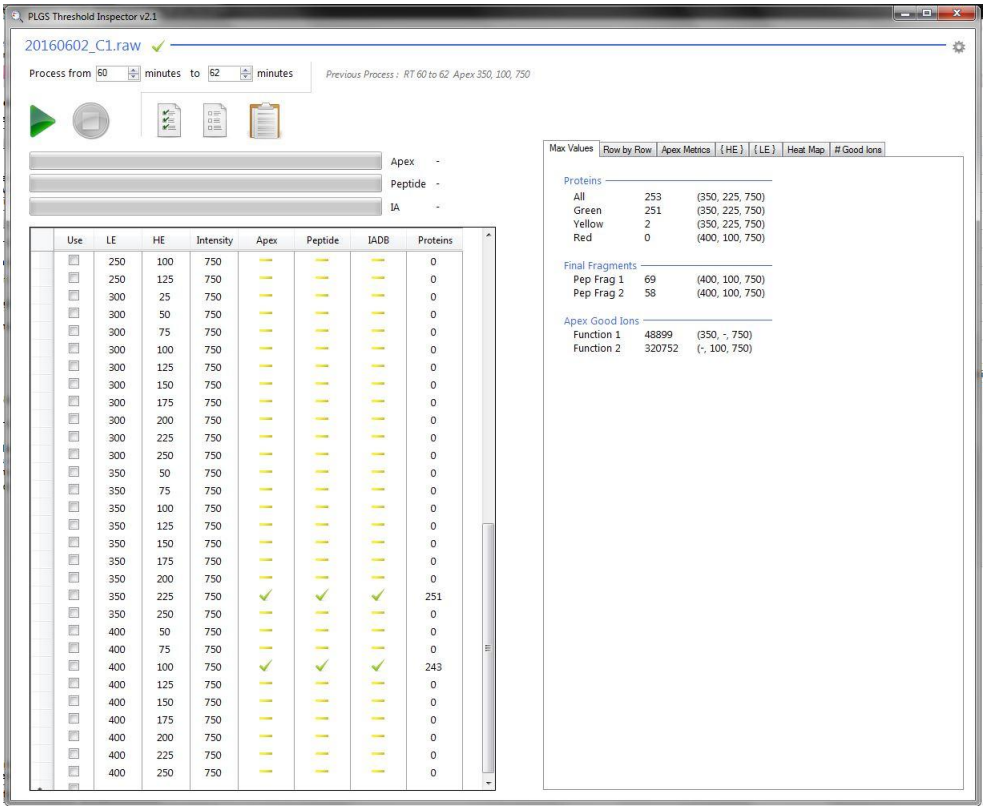
1. Press the Play button and leave it to work through the parameters and find the most appropriate combination of HE and LE settings for the data file you've selected.
   *THIS CAN TAKE A LONG TIME. BE PATIENT*

2. As PLGS-TI works through the various combinations of settings they will be ticked off the list, and a heat map will appear showing which of the parameters combinations tested gave the most protein identifications.




3. The score displayed in the heat map tab is the "Proteins, Green" setting for each parameter group. These are the protein identifications that have the highest confidence limit assigned to them by the iadbs.exe
4. The combination with the highest number of 'green' identifications will have white text.
5. If at the end of testing two sets of parameters have identical results, pick those with eh higher parameters. They will process faster in Progenesis QI
6. If at the end of testing the parameters at the top, or far right of the heat map gave the most hits, these may not be the optimal parameters. You should then add more, higher parameters to test where the real optimal parameters lie.
7. If you add more parameters at this point you can test just the additional parameters without over-writing the results of those already tested.

8. At the end of the threshold optimisation process the best thresholds will be displayed in the Max Values tab.




**Note** PLGS threshold inspector reports number of proteins, not number of hits. This means the results may be inflated due to inclusion of redundant homologues.

## Using the optimised thresholds in PLGS3.0.3

1. Open PLGS3.0.3.
2. In the Libraries manager>processing parameters wizard configure a new Electrospray MSe processing method using the optimal Apex3D parameters from the PLGS-Threshold Inspector Max Values results
3. Assign the raw data file to a well and append the processing parameters created at step 2.
4. Assign appropriate sequence search parameters in a PLGS workflow. For best results you should use a species specific databank, not the default Swissprot sequence databank provided with PLGS3.0.3. The demo databank is very out of date, and non-redundant species specific sequence databanks produce better results more quickly.
5. Process and search.

## About search workflows and sequence databases

To get you started this procedure includes a default search workflow that uses the default swissprot database installed with PLGS3.0.3. This will work. However, experienced users may prefer to use a species specific sequence databank appropriate to your sample, as this can be a little bit quicker. In PLGS3.0 and later search workflows are embedded in the projects.xml file, not stored in the documents folder. Using PLGS3.0.3 you can create or select a specific workflow for use with PLGS Threshold Inspector in the PLGS workflow library, and export it as an XML file to a known location. See PLGS help files for instructions.

Alternatively, instructions for manually editing an xml file containing a workflow that uses a different sequence database are shown below.

1. Either cut and past the xml code below into Notepad, or open the default_search_workflow.xml in notepad and edit that.
2. Modify the settings highlighted in red below
    - WORKFLOW_TEMPLATE TITLE, Change this each time you create a new xml file, and save the xml file with the name you enter here
    - SEARCH_DATABASE NAME. Go to Progenesis>Indentify Peptides. Enter the name of the fasta databank specific to the sample that you're analysing in this part of the xml file. Don't put the file path of the fasta file in the XML file as that's defined in the PLGS-TI settings.
    - FASTA FORMAT VALUE. Go to Progenesis>Indentify Peptides>Edit (next to database name) The format in the xml file must match the Parsing rules setting. If you downloaded the file from Uniprot, this should be set to UNIPROT
    - <DIGESTS> If you've used something other than trypsin you'll have to manually define the properties of the protease used to digest the peptide.
3. Save As .xml using the same you used in WORKFLOW_TEMPLATE TITLE=

### An example xml workflow template for manual manipulation.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE WORKFLOW_TEMPLATE SYSTEM "file:../dtds/ProteinLynx.dtd">
<WORKFLOW_TEMPLATE TITLE="Default MSe Workflow"
WORKFLOW_TEMPLATE_ID="_13353621570860_5050763941600845">
  <PROTEINLYNX_QUERY TYPE="Databank-search">
    <DATABANK_SEARCH_QUERY_PARAMETERS>
      <SEARCH_ENGINE_TYPE VALUE="PLGS"/>
      <SEARCH_DATABASE NAME="SWISSPROT-1.0"/>
      <SEARCH_TYPE NAME="Electrospray-Shotgun"/>
      <IA_PARAMS>
        <FASTA_FORMAT VALUE="STANDARD_SPACED"/>
        <PRECURSOR_MHP_WINDOW_PPM VALUE="-1"/>
        <PRODUCT_MHP_WINDOW_PPM VALUE="-1"/>
        <NUM_BY_MATCH_FOR_PEPTIDE_MINIMUM VALUE="1"/>
        <NUM_PEPTIDE_FOR_PROTEIN_MINIMUM VALUE="1"/>
        <NUM_BY_MATCH_FOR_PROTEIN_MINIMUM VALUE="7"/>
        <PROTEIN_MASS_MAXIMUM_AMU VALUE="250000"/>
        <FALSE_POSITIVE_RATE VALUE="4"/>
        <AQ_PROTEIN_ACCESSION VALUE=""/>
        <AQ_PROTEIN_MOLES VALUE="-1"/>
        <MANUAL_RESPONSE_FACTOR VALUE="-1"/>
        <DIGESTS>
          <ANALYSIS_DIGESTOR MISSED_CLEAVAGES="1">
            <AMINO_ACID_SEQUENCE_DIGESTOR NAME="Trypsin">
              <CLEAVES_AT AMINO_ACID="K" POSITION="C-TERM">
                <EXCLUDES AMINO_ACID="P" POSITION="N-TERM"/>
              </CLEAVES_AT>
```

```
                    <CLEAVES_AT AMINO_ACID="R" POSITION="C-TERM">
                      <EXCLUDES AMINO_ACID="P" POSITION="N-TERM"/>
                    </CLEAVES_AT>
                  </AMINO_ACID_SEQUENCE_DIGESTOR>
                </ANALYSIS_DIGESTOR>
              </DIGESTS>
              <MODIFICATIONS>
                <ANALYSIS_MODIFIER STATUS="FIXED">
                  <MODIFIER MCAT_REAGENT="No" NAME="Carbamidomethyl+C">
                    <MODIFIES APPLIES_TO="C" DELTA_MASS="57.0215" TYPE="SIDECHAIN"/>
                  </MODIFIER>
                </ANALYSIS_MODIFIER>
                <ANALYSIS_MODIFIER ENRICHED="FALSE" STATUS="VARIABLE">
                  <MODIFIER MCAT_REAGENT="No" NAME="Oxidation+M">
                    <MODIFIES APPLIES_TO="M" DELTA_MASS="15.9949" TYPE="SIDECHAIN"/>
                  </MODIFIER>
                </ANALYSIS_MODIFIER>
              </MODIFICATIONS>
            </IA_PARAMS>
          </DATABANK_SEARCH_QUERY_PARAMETERS>
        </PROTEINLYNX_QUERY>
      </WORKFLOW_TEMPLATE>
```