

# PLGS Threshold Inspector : Configuring for use with Progenesis QI-Proteomics v3.0

## Contents

PLGS Threshold Inspector : Configuring for use with Progenesis QI-Proteomics v3.0.....	1
About PLGS Threshold Inspector .....	2
Why use PLGS Threshold Inspector? .....	2
Requirements.....	2
Setting up PLGS-TI using Progenesis QI-Proteomics v3.0 executables.....	3
Running PLGS-TI :.....	8
Using the optimised thresholds in Progenesis QI-Proteomics.....	9
About search workflows and sequence databases .....	10
An example xml workflow template for manual manipulation. ....	10

## About PLGS Threshold Inspector

PLGS Threshold Inspector (PLGS-TI) is a tool for semi-automatically finding the optimal low energy (LE) and elevated energy (HE) thresholds for MSe data processing in PLGS and Progenesis QI-Proteomics v3.0. In Progenesis QI-Proteomics these are referred to as Apex3D parameters. Although PLGS-TI was designed to use the MSe processing and search executables in PLGS3.0 and later, it can also make use of the same executables that are installed with Progenesis QI-Proteomics v3.0 for MSe processing. This document aims to explain how to do this.

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 a Waters or Nonlinear Dynamics product, and is not covered by Waters or Nonlinear Dynamics support packages.

These guidelines for using PLGS Threshold Inspector were not written by or approved by Waters Corp, 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 Progenesis QI-Proteomics MSe data processing. The default settings are just there to get you started with the demonstration data.
  - 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 Progenesis QI-Proteomics 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. Progenesis QI-Proteomics V3.0 (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 Progenesis QI-Proteomics minimum recommended specifications.

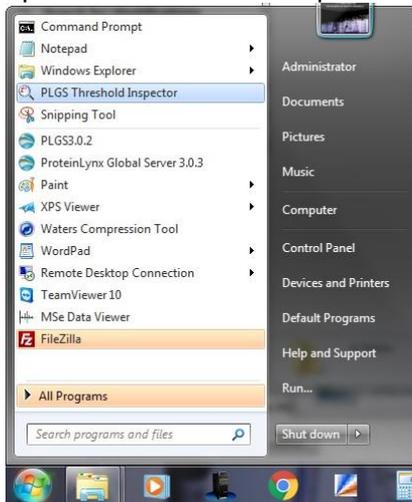
## Setting up PLGS-TI using Progenesis QI-Proteomics v3.0 executables

1. Download and install PLGS Threshold Inspector
2. Copy the embedded xml file below (this is an xml file, not just an icon) to your documents folder . This contains the sequence databank search parameters.



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.

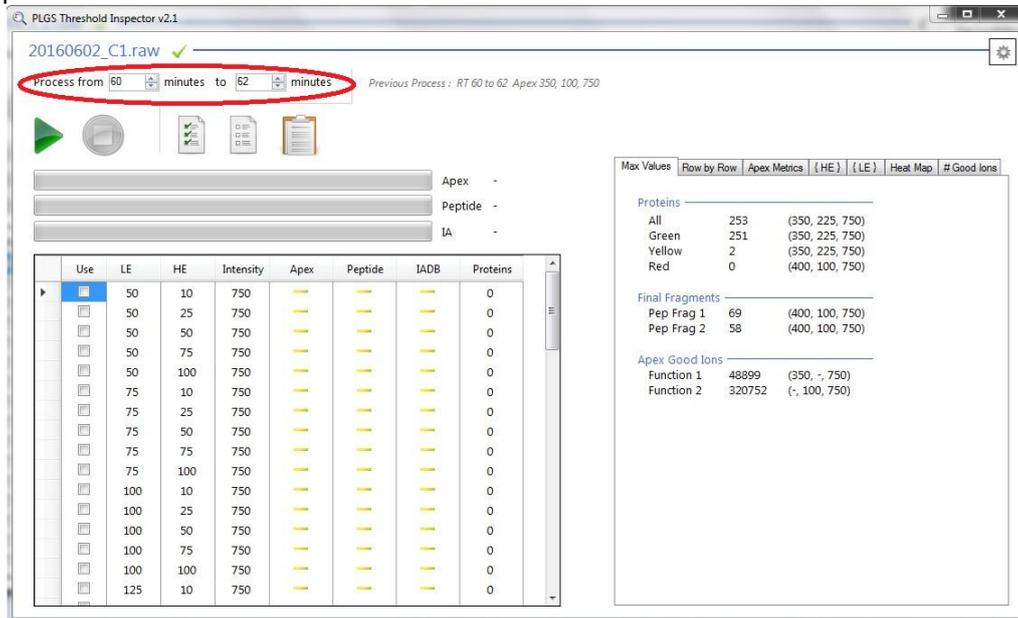
Use	LE	HE	Intensity	Apex	Peptide	IADB	Proteins
<input checked="" type="checkbox"/>	50	10	750	---	---	---	0
<input type="checkbox"/>	50	25	750	---	---	---	0
<input type="checkbox"/>	50	50	750	---	---	---	0
<input type="checkbox"/>	50	75	750	---	---	---	0
<input type="checkbox"/>	50	100	750	---	---	---	0
<input type="checkbox"/>	75	10	750	---	---	---	0
<input type="checkbox"/>	75	25	750	---	---	---	0
<input type="checkbox"/>	75	50	750	---	---	---	0
<input type="checkbox"/>	75	75	750	---	---	---	0
<input type="checkbox"/>	75	100	750	---	---	---	0
<input type="checkbox"/>	100	10	750	---	---	---	0
<input type="checkbox"/>	100	25	750	---	---	---	0
<input type="checkbox"/>	100	50	750	---	---	---	0
<input type="checkbox"/>	100	75	750	---	---	---	0
<input type="checkbox"/>	100	100	750	---	---	---	0
<input type="checkbox"/>	125	10	750	---	---	---	0

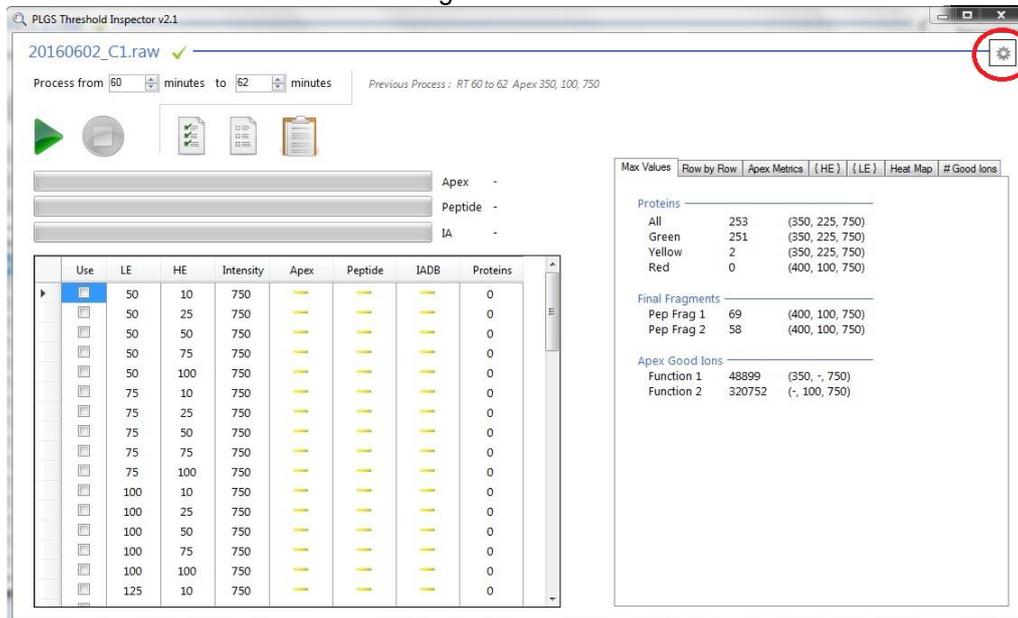
Max Values	Row by Row	Apex Metrics	{ HE }	{ LE }	Heat Map	# Good Ions
<b>Proteins</b>						
All	253	(350, 225, 750)				
Green	251	(350, 225, 750)				
Yellow	2	(350, 225, 750)				
Red	0	(400, 100, 750)				
<b>Final Fragments</b>						
Pep Frag 1	69	(400, 100, 750)				
Pep Frag 2	58	(400, 100, 750)				
<b>Apex Good Ions</b>						
Function 1	48899	(350, -, 750)				
Function 2	320752	(-, 100, 750)				

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.

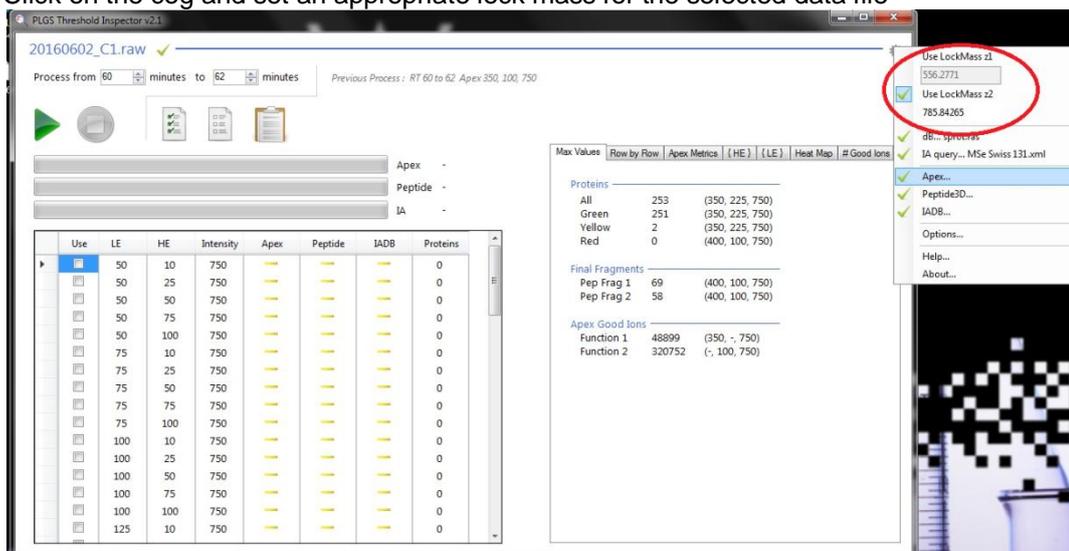
- In PLGS-TI select "Process from x minutes to y minutes" settings based on the representative retention time window determined at step 3. You can use a wider window, or the whole acquisition. But the wider the RT window, the longer the PLGS-TI will take to test each set of parameters.



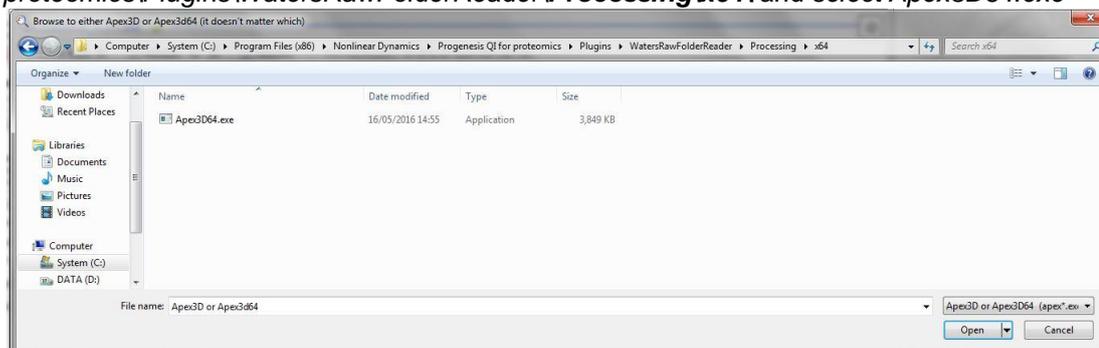
- The cog icon at the top right of the screen opens a drop down menu of settings for PLGS-TI. If some of these haven't been set, or have been set incorrectly, there may be a flashing red exclamation mark icon next to the cog.



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



- Click on the cog, and select Apex.... In the resulting browser window navigate to *C:\Program Files (x86)\Nonlinear Dynamics\Progenesis QI for proteomics\Plugins\WatersRawFolderReader\Processing\*x64 and select *Apex3D64.exe*



- Click on the cog, select Peptide3D. In the browser window navigate to *C:\Program Files (x86)\Nonlinear Dynamics\Progenesis QI for proteomics\Plugins\WatersRawFolderReader\Processing\* and select Peptide3D
- Click on the cog, select IADBs. Navigate to *C:\Program Files (x86)\Nonlinear Dynamics\Progenesis QI for proteomics\Plugins\WatersRawFolderReader\Identification\* and select *iadbs.exe*
- 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
- 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 Progenesis QI-P. Navigate to *C:\Program Files (x86)\Nonlinear Dynamics\Progenesis QI for proteomics\Plugins\WatersRawFolderReader\Identification* and select *sprot.fas*
- LE and HE thresholds :
  - By default PLGS-TI tests low energy (LE) thresholds ranging from 50 to 250, and high energy (HE) threshold settings from 10 to 100.
  - Some users find that the optimal setting for their data are higher than this.
  - To add more combinations of parameters click on the an assigned check box at the bottom of the list of threshold combinations.

PLGS Threshold Inspector v2.1

20160602\_C1.raw ✓

Process from 60 minutes to 62 minutes Previous Process: RT 60 to 62 Apex 350, 100, 750

Apex  
Peptide -  
TA -

Use	LE	HE	Intensity	Apex	Peptide	IADB	Proteins
<input type="checkbox"/>	30	10	750	---	---	---	C
<input type="checkbox"/>	50	25	750	---	---	---	C
<input type="checkbox"/>	50	50	750	---	---	---	C
<input type="checkbox"/>	50	75	750	---	---	---	C
<input type="checkbox"/>	50	100	750	---	---	---	C
<input type="checkbox"/>	75	10	750	---	---	---	C
<input type="checkbox"/>	75	25	750	---	---	---	C
<input type="checkbox"/>	75	50	750	---	---	---	C
<input type="checkbox"/>	75	75	750	---	---	---	C
<input type="checkbox"/>	75	100	750	---	---	---	C
<input type="checkbox"/>	100	10	750	---	---	---	C
<input type="checkbox"/>	100	25	750	---	---	---	C
<input type="checkbox"/>	100	50	750	---	---	---	C
<input type="checkbox"/>	100	75	750	---	---	---	C
<input type="checkbox"/>	100	100	750	---	---	---	C
<input type="checkbox"/>	125	10	750	---	---	---	C
<input type="checkbox"/>	125	25	750	---	---	---	C
<input type="checkbox"/>	125	50	750	---	---	---	C
<input type="checkbox"/>	125	75	750	---	---	---	C
<input type="checkbox"/>	125	100	750	---	---	---	C
<input type="checkbox"/>	150	10	750	---	---	---	C
<input type="checkbox"/>	150	25	750	---	---	---	C
<input type="checkbox"/>	150	50	750	---	---	---	C
<input type="checkbox"/>	150	75	750	---	---	---	C
<input type="checkbox"/>	150	100	750	---	---	---	C
<input type="checkbox"/>	200	10	750	---	---	---	C
<input type="checkbox"/>	200	25	750	---	---	---	C
<input type="checkbox"/>	200	50	750	---	---	---	C
<input type="checkbox"/>	200	75	750	---	---	---	C
<input type="checkbox"/>	200	100	750	---	---	---	C
<input type="checkbox"/>	*						

Max Values Row by Row Apex Metrics (HE) (LE) Heat Map # G

Proteins

All	253	(350, 225, 750)
Green	251	(350, 225, 750)
Yellow	2	(350, 225, 750)
Red	0	(400, 100, 750)

Final Fragments

Pep Frag 1	69	(400, 100, 750)
Pep Frag 2	58	(400, 100, 750)

Apex Good Ions

Function 1	48899	(350, -, 750)
Function 2	320732	(-, 100, 750)

- 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 PQIP3.0 do not use the Intensity parameter. This is only used in PQIP2.0 (and PLGS3.0.2 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' icon (b) to test all of them.

Process from 00 minutes to 62 minutes Previous Process: RT 60 to 62 Apex 350, 100, 750

Use	LE	HE	Intensity	Apex	Peptide	IAD8	Proteins
<input type="checkbox"/>	50	10	750	---	---	---	U
<input type="checkbox"/>	50	25	750	---	---	---	C
<input type="checkbox"/>	50	50	750	---	---	---	C
<input type="checkbox"/>	50	75	750	---	---	---	C
<input type="checkbox"/>	50	100	750	---	---	---	C
<input type="checkbox"/>	75	10	750	---	---	---	U
<input type="checkbox"/>	75	25	750	---	---	---	C
<input type="checkbox"/>	75	50	750	---	---	---	C
<input type="checkbox"/>	75	75	750	---	---	---	C
<input type="checkbox"/>	75	100	750	---	---	---	C
<input type="checkbox"/>	100	10	750	---	---	---	C
<input type="checkbox"/>	100	25	750	---	---	---	C
<input type="checkbox"/>	100	50	750	---	---	---	C
<input type="checkbox"/>	100	75	750	---	---	---	C
<input type="checkbox"/>	100	100	750	---	---	---	C
<input type="checkbox"/>	125	10	750	---	---	---	C
<input type="checkbox"/>	125	25	750	---	---	---	C
<input type="checkbox"/>	125	50	750	---	---	---	U
<input type="checkbox"/>	125	75	750	---	---	---	C
<input type="checkbox"/>	125	100	750	---	---	---	C
<input type="checkbox"/>	150	10	750	---	---	---	C
<input type="checkbox"/>	150	25	750	---	---	---	C
<input type="checkbox"/>	150	50	750	---	---	---	C
<input type="checkbox"/>	150	75	750	---	---	---	C
<input type="checkbox"/>	150	100	750	---	---	---	C
<input type="checkbox"/>	200	10	750	---	---	---	C
<input type="checkbox"/>	200	25	750	---	---	---	C
<input type="checkbox"/>	200	50	750	---	---	---	C
<input type="checkbox"/>	200	75	750	---	---	---	U
<input type="checkbox"/>	200	100	750	---	---	---	C

Proteins

All	253	(350, 225, 750)
Green	251	(350, 225, 750)
Yellow	2	(350, 225, 750)
Red	0	(400, 100, 750)

Final Fragments

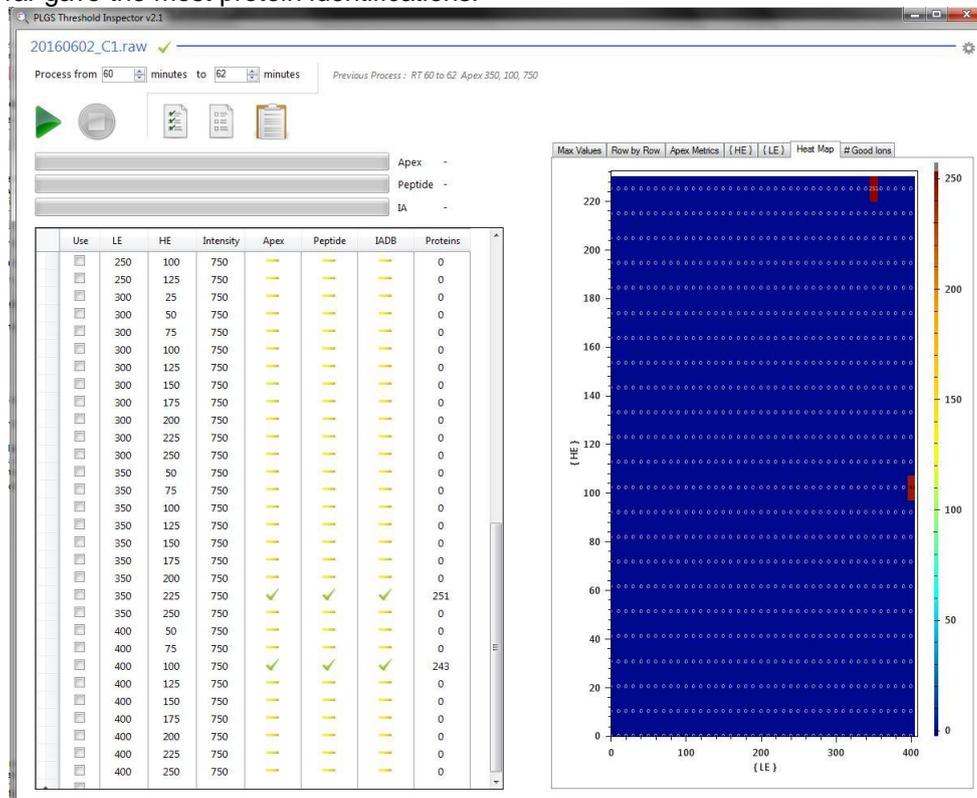
Pep Frag 1	69	(400, 100, 750)
Pep Frag 2	58	(400, 100, 750)

Apex Good Ions

Function 1	48899	(350, 1, 750)
Function 2	320732	(1, 100, 750)

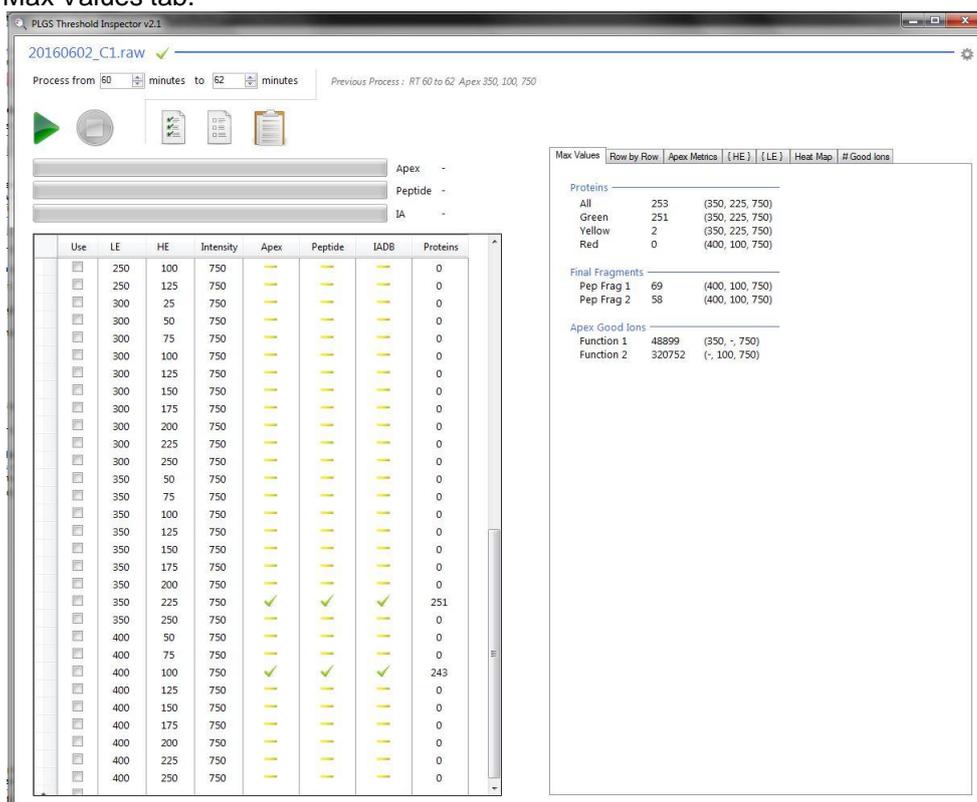
## Running PLGS-TI :

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 the heat map tab is populated, showing which of the parameters combinations tested so far 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 the 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.

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



Use	LE	HE	Intensity	Apex	Peptide	IADB	Proteins
<input type="checkbox"/>	250	100	750	---	---	---	0
<input type="checkbox"/>	250	125	750	---	---	---	0
<input type="checkbox"/>	300	25	750	---	---	---	0
<input type="checkbox"/>	300	50	750	---	---	---	0
<input type="checkbox"/>	300	75	750	---	---	---	0
<input type="checkbox"/>	300	100	750	---	---	---	0
<input type="checkbox"/>	300	125	750	---	---	---	0
<input type="checkbox"/>	300	150	750	---	---	---	0
<input type="checkbox"/>	300	175	750	---	---	---	0
<input type="checkbox"/>	300	200	750	---	---	---	0
<input type="checkbox"/>	300	225	750	---	---	---	0
<input type="checkbox"/>	300	250	750	---	---	---	0
<input type="checkbox"/>	350	50	750	---	---	---	0
<input type="checkbox"/>	350	75	750	---	---	---	0
<input type="checkbox"/>	350	100	750	---	---	---	0
<input type="checkbox"/>	350	125	750	---	---	---	0
<input type="checkbox"/>	350	150	750	---	---	---	0
<input type="checkbox"/>	350	175	750	---	---	---	0
<input type="checkbox"/>	350	200	750	---	---	---	0
<input type="checkbox"/>	350	225	750	✓	✓	✓	251
<input type="checkbox"/>	350	250	750	---	---	---	0
<input type="checkbox"/>	400	50	750	---	---	---	0
<input type="checkbox"/>	400	75	750	---	---	---	0
<input type="checkbox"/>	400	100	750	✓	✓	✓	243
<input type="checkbox"/>	400	125	750	---	---	---	0
<input type="checkbox"/>	400	150	750	---	---	---	0
<input type="checkbox"/>	400	175	750	---	---	---	0
<input type="checkbox"/>	400	200	750	---	---	---	0
<input type="checkbox"/>	400	225	750	---	---	---	0
<input type="checkbox"/>	400	250	750	---	---	---	0

**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 Progenesis QI-Proteomics

- Open Progenesis.
- Set Apex3D processing parameters based on the PLGS-Threshold Inspector Max Values results
  - If importing the MassLynx raw data into a new Progenesis Experiment, enter the LE and HE parameters when you are asked to specify MSe processing settings
  - If you are reprocessing data in an existing project, navigate to the Identify Peptides window, select Change Apex 3D parameters, and enter the optimised settings.
- Assign appropriate sequence search parameters in Progenesis. For best results you should use a species specific databank, not the default Swissprot sequence databank provided with Progenesis QI-Proteomics

## 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 Progenesis Q1-Proteomics v3.0. 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. Instructions for creating 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>Identify 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>Identify 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">
    <DATABASE_SEARCH_QUERY_PARAMETERS>
      <SEARCH_ENGINE_TYPE VALUE="PLGS"/>
      <SEARCH_DATABASE NAME="SWISSPROT-1.0"/>
      <SEARCH_TYPE NAME="Electrospray-Shotgun"/>
    </DATABASE_SEARCH_QUERY_PARAMETERS>
    <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"/>
    </IA_PARAMS>
    <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>
  </PROTEINLYNX_QUERY>
</WORKFLOW_TEMPLATE>
```

```
</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>
```