

Tutorial 9: SWATH data analysis in Skyline

In this tutorial we will learn how to perform targeted post-acquisition analysis for protein identification and quantitation using a data-independent dataset acquired in a QqTOF instrument (5600 TripleTOF, AB Sciex).

Using Skyline we will query acquired datasets for a given subset of proteins. Initially, we will set all the parameters in the Skyline session required to work with data-independent datasets and then we will proceed to extract the quantitation information from the raw data files.

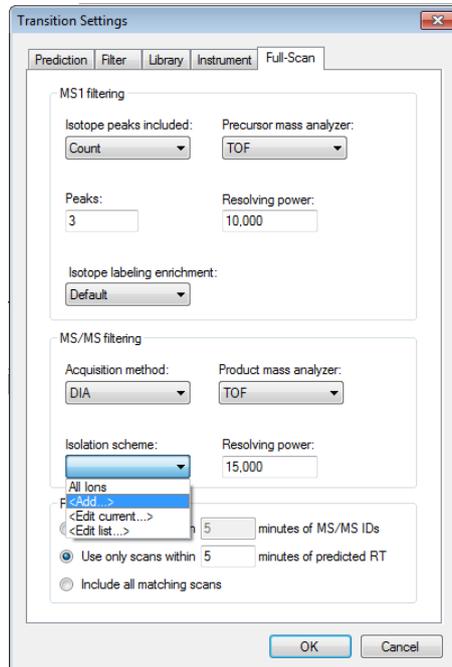
1. Preparing Skyline session for data-independent acquisition (SWATH)

In this section we are going to prepare a Skyline session with the appropriate settings for SWATH analysis. We will tune the settings for data-independent analysis, and then we will extract our proteins of interest from the SWATH file.

- Open the Skyline document *TarProCourse_Tut9_swath.sky* from the folder *Tutorial-9_SWATH*
Note! In the “Targets window” there are the iRT peptides already listed and later we will populate the window with additional targets.

1.1. Defining the data-independent acquisition isolation scheme

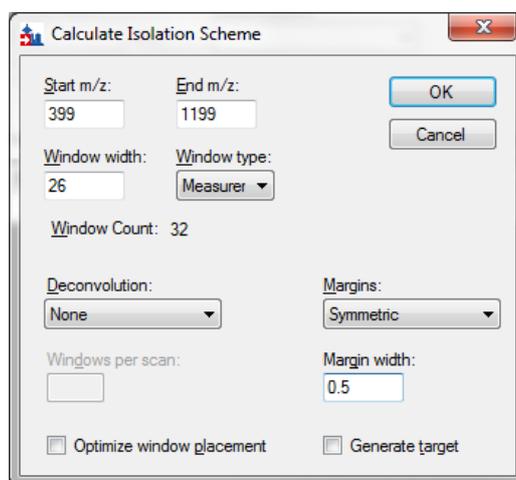
- Go to “Settings” → “Transition Settings” → “Full-Scan”
Fill the window as indicated in the screen shot.
Note! The resolving power depends on the state of the instrument during SWATH data acquisition. The optimum might be slightly different for each dataset.



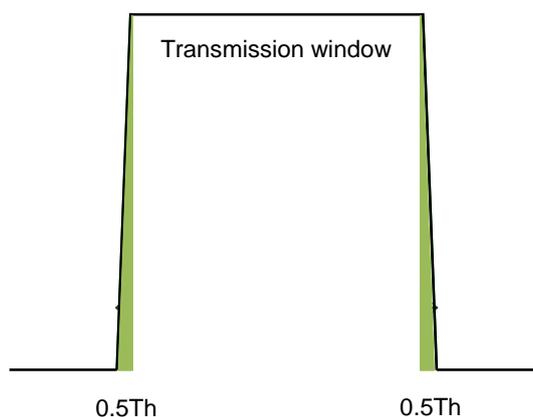
Now we need to define a new “Isolation scheme” according to the parameters defined on the instrument for data-independent acquisition.

Note! In our case we used 32 windows of 26 m/z each that covered the range from 400 to 1200 m/z with 1 m/z overlap.

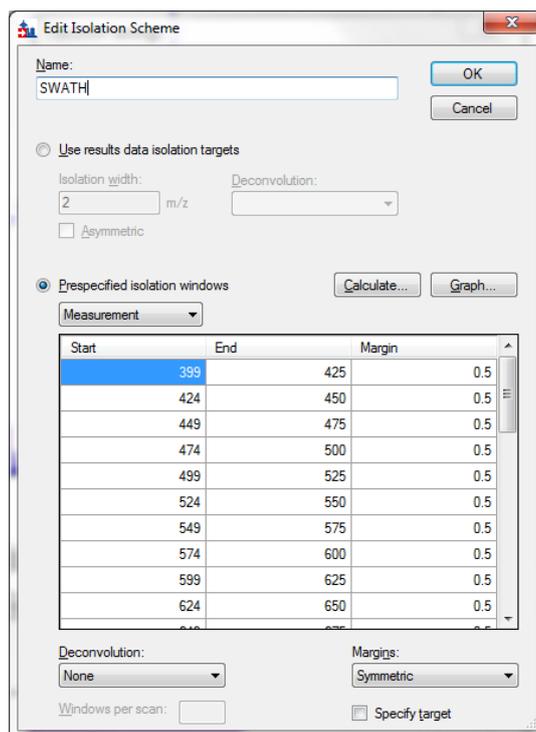
- From “Isolation scheme” drop-down → select “Add...” → fill the name as SWATH
 - Select “Pre-specified isolation windows” → click on “Calculate...”
 - Fill the “Calculate Isolation Scheme” window as indicated:
 - Start m/z = 399
 - End m/z = 1199
 - Window width = 26
 - Window type = “Measurement”
 - Margins = Symmetric
 - Margin width = 0.5



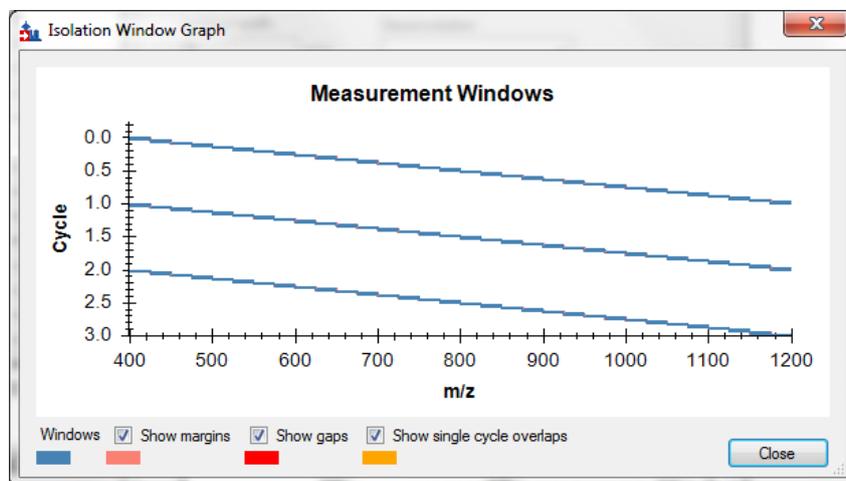
Note! As the transmission windows are not perfect, the margin option allows to specify how much of the window edges should not be used for extraction. Skyline will then extract from *Start+margin* until *End-margin*.



- Click “OK”
- Check that the “Start” of the first window is 400 m/z and that the “End” of your last window is 1200 m/z. Adjust them if necessary.



- Click “Graph” to see how the isolation windows cover the specified range.

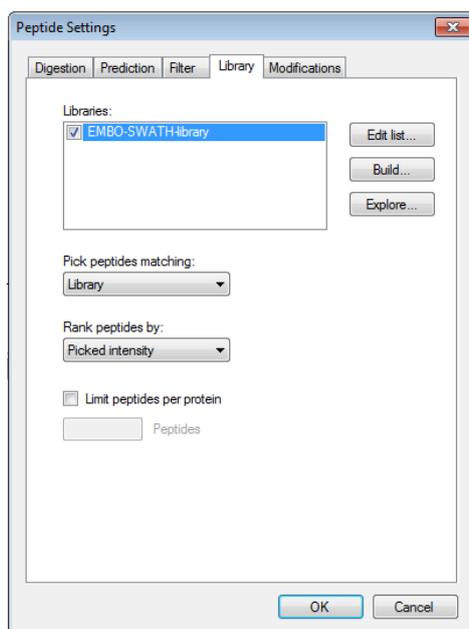


- Click “Close”
- Click “OK”
- Go to “Transition settings” → “Full Scan” and select “Use only scans within [] minutes of predicted RT”
- Set the number of minutes to 5 (for a 10 minute window around the predicted RT).
- Click OK
- Save your file as *TarProCourse_Tut9_swath_acquistionscheme.sky*

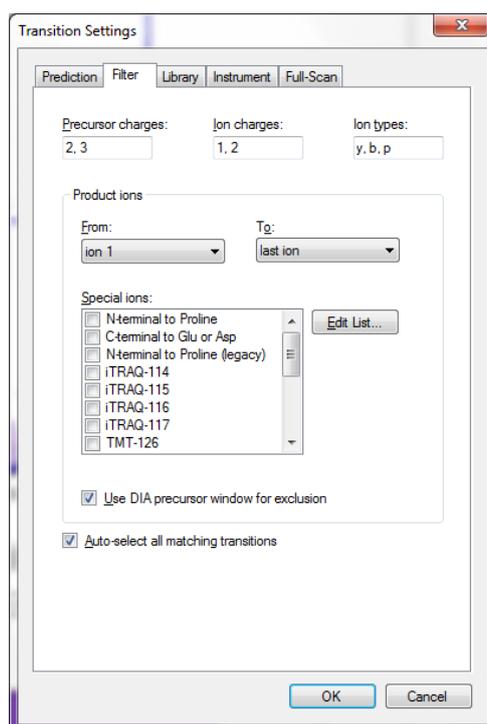
1.2. Adding a reference spectral library

A reference spectral library calibrated with standard iRT peptides was generated for *Mycobacterium tuberculosis* using synthetic peptides (Schubert et al. Cell Host Microbe, 2013) that we will use for our SWATH analyses. To upload the library in our Skyline session:

- Go to “Settings” → “Peptide Settings” → “Library”
 - Select “Edit list” → “Add...”
 - Name your new library TarProCourse-SWATH-library
 - Browse in the folder *Tutorial-9_SWATH* for the file: *TarProCourse-SWATH-library.blib* → Click twice “OK”.
 - In the “Library” tab, activate the newly added library by checking the box next to it.
 - set the “Pick peptides matching” to “Library” and “Rank peptides by:” to “Picked intensity” → Click “OK”.



- Go to “Settings” → “Transition Settings” → “Filter” tab
 - Make sure that the settings are as the ones indicated in the screen shot:

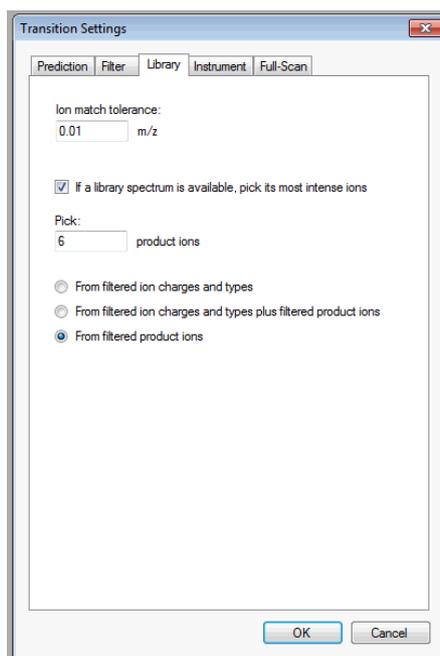


Note! “Use DIA precursor window for exclusion” means that transitions falling into the precursor window will not be selected. These transitions are typically noisy and are therefore often excluded.

Note! By checking “Auto-select all matching transition” we will automatically extract for a given peptide all transitions that match the filter and library settings.

Note! In ion types we are choosing to extract not only y and b ions but also peptide precursor (p) ions. Although targeted data-independent acquisition methods rely on MS2 quantitation, the Q-TOF 5600 from AB Sciex also records an MS1 scan in every cycle. In this tutorial, we will take advantage of this information to also visualize extracted ion chromatograms at the MS1 level.

- Go to “Settings” → “Transition Settings” → “Library” tab
Fill the window as indicated in the screen shot:

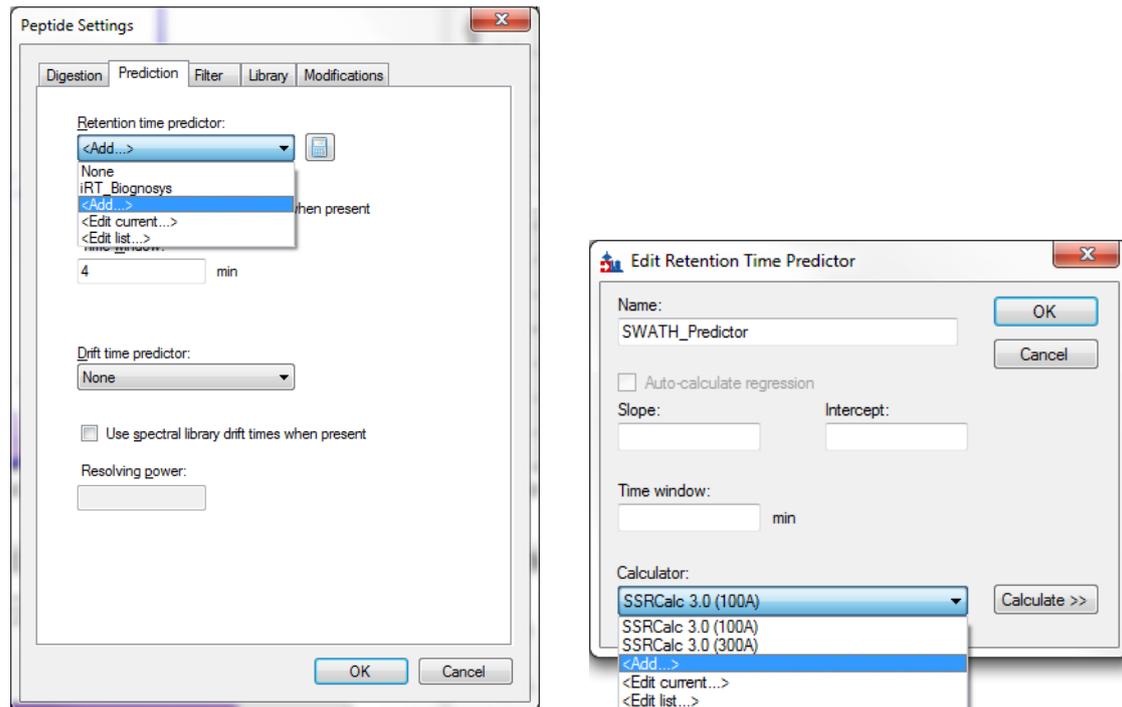


- Save your file as *TarProCourse_Tut9_swath_spectrallibrary.sky*

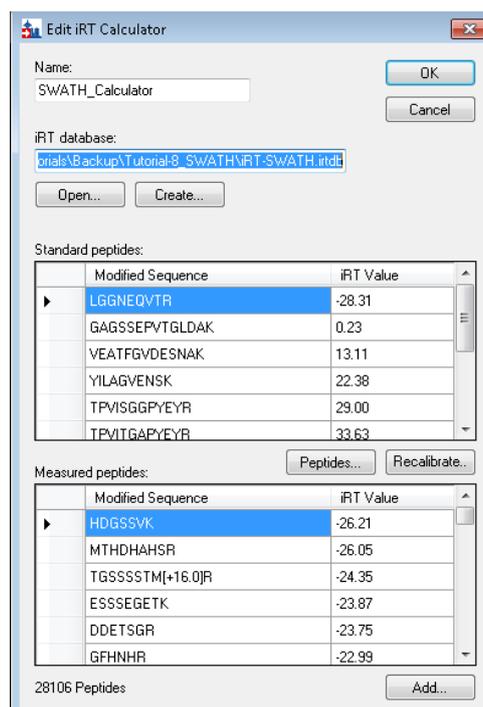
1.3. Calibration of the SWATH file

We will now generate a retention time predictor with the SWATH file data to be able to predict the retention times of our peptides of interest using the calibrated retention time information (iRT) from our spectral library.

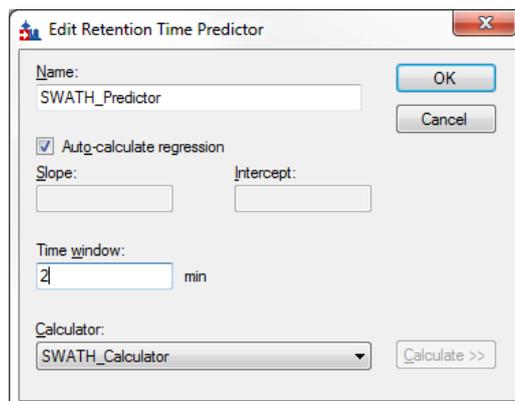
- Go to “Settings” → “Peptide Settings” → “Prediction” tab
From the “Retention time predictor” drop-down menu select the “Add” option



- In the “Edit Retention time predictor” do:
 - Fill the name as: SWATH_Predictor
 - In the calculator drop-down menu select “Add”
 - Set its name to: SWATH_Calculator → Click the “Open” button, browse to the folder *Tutorial-9_SWATH* → select the iRT database called iRT-SWATH.irtdb.
- Note!** In this case, an iRT database (database with all normalized peptide retention times) was generated when the library was first created. Therefore, there is no need to create the database again.
→ Click “OK”.



- Back in the “Edit Retention Time Predictor”
 - Check the “Auto-calculate regression” check box
 - set the Time window to “2” min



- Click twice “OK”
- Save your file as *TarProCourse_Tut9_swath_predictor.sky*

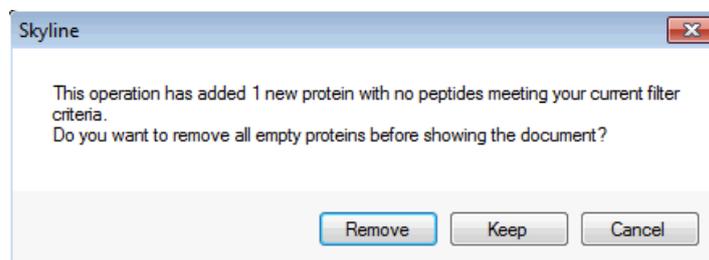
2. Adding your favourite peptides and proteins into your file to be extracted from the data-independent acquisition files.

In contrast to SRM, data-independent analysis techniques, such as SWATH MS, allow data extraction for any peptide after the mass spectrometric acquisition has been done. In this section we will illustrate this by extracting some proteins that were previously monitored by SRM, and by extracting information for a set of new proteins.

2.1.Importing a subset of proteins previously monitored by SRM

Here we will import some proteins that were previously monitored by SRM and that we now want to analyze again by SWATH MS.

- Go to “File” → “Import” → “FASTA...”
 - Open the fasta file *mtb-srm-protein-subset.fasta* located in the folder *Tutorial-9_SWATH*
 - If prompted with a message saying that one protein did not match your filtering criteria → select “Remove”.



Note! At the bottom of the “Targets window” the new proteins have been added. For each protein we can see the peptides that have a spectrum match in our spectral library. At this point you should have 5 proteins, 18 peptides, 23 precursors, and 130 transitions.

2.2. Importing a subset of proteins of interest from Uniprot

Here we will import a set of proteins related to glycolysis for which we want to have quantitative values from SWATH MS data. Note: The set of proteins selected in this exercise are related to glycolysis but note that one could select proteins involved in any other cellular process, and that this can be done iteratively to any set of proteins of interest to explore any specific hypothesis.

- Go to www.uniprot.org
 - Select “Advanced” next to the Search field
 - Search for “Taxonomy”: 83332; “Keyword”: Glycolysis and “Reviewed proteins” as shown in the screen shot:

Searching in UniProtKB Help

Term
Taxonomy [OC] : 83332
Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv) [83332]

AND : All

Searching in UniProtKB Help

Term
Taxonomy [OC] : Mycobacterium tuberculosis (

Term
Keyword [KW] : Glycolysis
Glycolysis [KW-0324]

AND :

Searching in UniProtKB Help

Term
Taxonomy [OC] : Mycobacterium tuberculosis (

Term
Keyword [KW] : Glycolysis [KW-0324]

AND : Reviewed : Yes

- The search should retrieve 13 protein entries
- Click on “Download”
- Select “Download all (13)”; Format: FASTA (canonical); select “Uncompressed”
- Save your file into the folder *Tutorial-9_SWATH* with the name *glycolysis.fasta*

Download selected (0)

Download all (13)

Format: FASTA (canonical) ▼

Compressed Uncompressed

[Preview first 10](#)

- Back in your Skyline session:
 - Go to “File” → “Import” → “Fasta”
 - Open the fasta file you just retrieved from UniProt

Note! At the bottom of the Targets window the new proteins have been added. At this point there should be 18 proteins, 47 peptides, 64 precursors and 499 transitions.

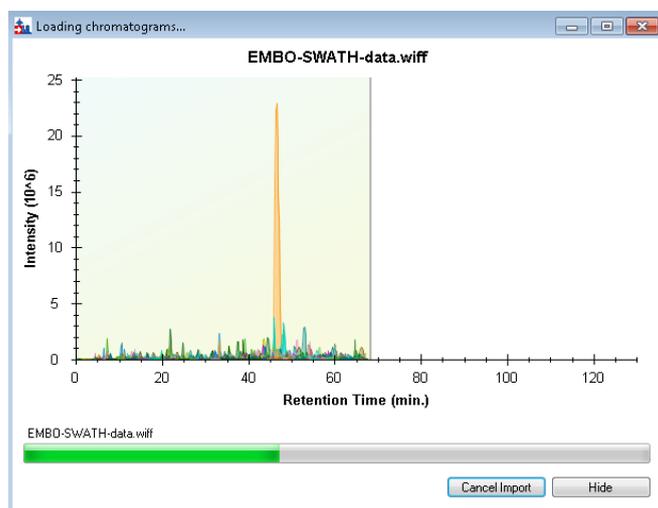
- Save your file as *TarProCourse_Tut9_newtargets.sky*

2.3. Extracting data for the new selected proteins from the SWATH data file

We will take advantage of the retention time information in the iRT calculator to extract the transitions for each peptide within a 10-minute (± 5 min) window around its predicted RT.

- Go to “Settings” → “Transition Settings” → “Full Scan”
 - Make sure that you have selected the option “Use only scans within [5] minutes of predicted RT” → Click “OK”
- Go to “File” → “Import” → “Results” → Keep the default settings and click “OK”
 - Import *TarProCourse-SWATH-data.wiff* from the *Tutorial-9_SWATH* folder.

Note! Extraction of all the transitions for all the selected peptides will take a few minutes. You can monitor this process in the Extraction window.



Once imported make sure you have selected the following visualization options:

- Go to “View” → “Ion Types” → “Y”, “B”, “Precursor”
- Go to “View” → “Transitions” → “All” and “Split Graph” (this allows you to see precursors and transitions in two separate windows).
- Go to “View” → “Transform” → “Savitzky-Golay Smoothing”
- Go to “View” → “Auto-Zoom” → “Best peak”
- Go to “View” → “Peak Areas” → “Replicate Comparison”

Tip! You might right-click on the “chromatogram” panel and uncheck “Legend” to see the chromatograms better.

Tip! You might right-click on the “Peak Areas” panel and uncheck “Legend” to see the peak area bars better.

- Save your file as *TarProCourse_Tut9_swath-extracted.sky*

Note! In a normal analysis it would be important to carry out the mProphet/OpenSWATH procedure described in Tutorial 7 to improve peak picking and crucially to estimate the error rate (q-values). This is especially important for documents containing 1,000s-10,000s of peptides where manual inspection is no longer feasible.

Exercices

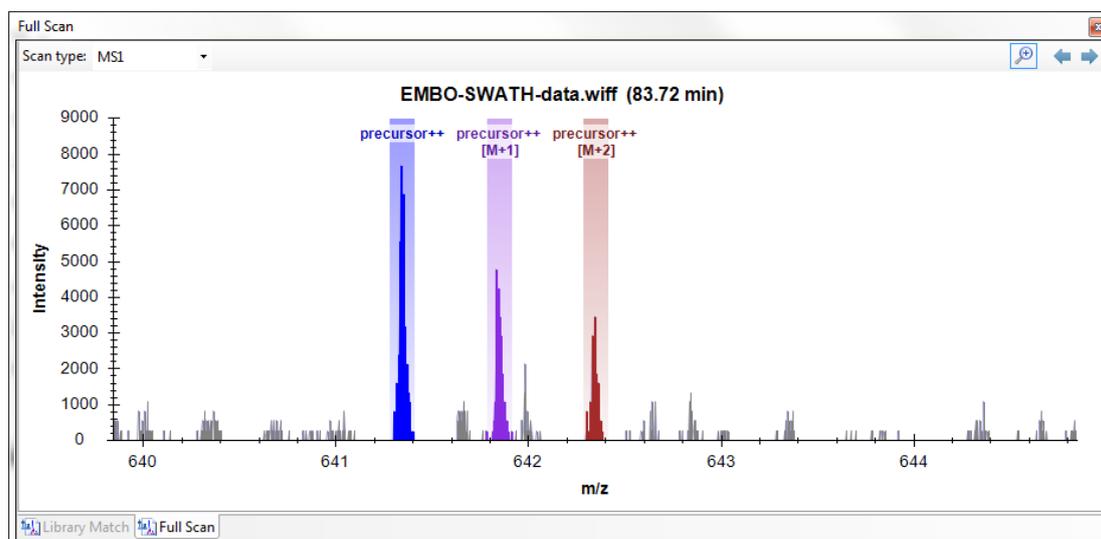
Evaluation of the SWATH data

We will now manually check the peaks that have been automatically picked by Skyline.

1. Check the peaks that have been picked and find examples of:
 - a. A peptide with clear MS1 and MS2 XICs
 - b. A peptide with bad/noisy MS1 but good MS2 XICs
 - c. A peptide with clear precursors and not MS2 XICs
 - d. A peptide with no MS1 signal but good MS2 XICs
 - e. A peptide with no MS1 signal nor good MS2 XICs

For which cases are you confident with the peptide identification? Which cases meet your quality criteria?

Note! Be sure to click on the chromatograms to see the underlying mass spectra from which the chromatograms were extracted.



Comparing SRM and SWATH data

1. Open in parallel a new Skyline session with the file *TarProCourse_Tut6_label_refined.sky* from your folder *Tutorial-6_ManualAnalysis* and compare the results with the ones obtained in the SWATH dataset. Check for some common peptides present both in the SRM and the SWATH files. Compare the signals in both types of acquisition methods and discuss what you observe (noise, interferences, intensity, etc.)

2. How many transitions can we monitor in a SRM method? How many fragments can we monitor by SWATH?
3. What is optimal number of fragments to be used for identification? And for quantitation? Are you more confident with the identifications of the SRM dataset or with the ones in the SWATH dataset?
4. Some proteins monitored by SRM and SWATH do not have the same peptides. Why?
5. For which projects would you use SRM and for which would you use SWATH MS to quantify proteins in your samples?
6. How would you quantify a thousand proteins in hundreds of samples?

We would like to thank SystemsX for supporting the Zurich Targeted Proteomics Course 2015.



SystemsX.ch
The Swiss Initiative in Systems Biology