Skyline Data Independent Acquisition

Data-independent acquisition (DIA)[1,2] is an advanced technique for performing large scale targeted proteomics experiments. Targeted acquisition approaches such as selected reaction monitoring (SRM) and parallel reaction monitoring (PRM) are limited to only a handful of peptides without acquisition scheduling or tens to hundreds of peptides per mass spec run with scheduling. DIA allows the measurement of much larger numbers of peptides (thousands or even whole proteomes) with only modest sacrifices in sensitivity, selectivity and reproducibility relative to SRM. DIA also has the advantage that the peptides to be measured do not need to be specified in advance or scheduled, but rather, product ion chromatograms for any desired peptide within a broad range of precursor m/z can be extracted after acquisition from a DIA run.

Since support for chromatogram extraction from DIA data was first implemented in Skyline, in October 2010, this support has been greatly improved, so that as of version 2.6, Skyline supports several popular strategies and workflows for analyzing DIA data. Skyline also supports all DIA-capable instruments including Q-TOF instruments from AB SCIEX, Agilent, Bruker and Waters and Q-Orbitrap instruments from Thermo.

One effective DIA workflow starts with any number of initial data-dependent acquisition (DDA) runs on a certain instrument type and chromatography set-up. It is beneficial to carry out these initial shotgun measurements on the same instrument where DIA will be run. However, it is also possible to transfer targeted assays between instrument platforms if similar fragmentation techniques and chromatography are used. For these initial DDA measurements, samples may be fractionated or otherwise simplified to reach higher proteome coverage. The DDA runs are processed with a peptide spectrum matching pipeline, and the resulting peptide ID's, spectra and retention times are then used to create either a spectral library and retention time (iRT) library (in Skyline) or an extended transition list (with similar information for a subset of fragment ions) called an "assay library" (in other tools). These libraries of fragment ion relative abundance and normalized retention time (iRT) can be used in any number of subsequent DIA runs employing the same instrumentation and standard peptides for retention time alignment.

Although many methods exist for converting DDA search results to a library suitable for this library DIA analysis approach (described as the "Prior Knowledge Workflow" in Skyline Tutorial Webinar #2), the most straightforward and accessible way to start DIA analysis is simply to intersperse DDA runs with DIA runs on the same instrument and use the spectra and retention times of the DDA results as predictions for the DIA runs. Either approach allows for the creation of more precise libraries of fragment ion relative abundance and normalized retention times, known as chromatogram libraries (detailed in the Panorama Chromatogram Libraries tutorial), actually using DIA runs to predict future DIA runs.

In this tutorial, you will learn how to set up, import, and process DIA runs in Skyline using the simple interspersed DDA/DIA approach (though the tutorial also contains pointers to alternative, more advanced methods, which are described elsewhere).

Getting Started

To start this tutorial, download the following ZIP file:

https://skyline.gs.washington.edu/tutorials/DIA.zip

Note that the file is very large (4.5 GB download, and 6.0 GB uncompressed), because it includes DIA runs which are generally very large files, often in the order of 100 to 200 times larger than files for SRM runs. If this download takes you too long or you do not have the available disk space, you can alternatively download a smaller (73 MB download, and 113 MB uncompressed) version of the tutorial here:

https://skyline.gs.washington.edu/tutorials/DIASmall.zip

If you choose the small version, lacking the mass spec raw data files, you will need to skip some steps in the tutorial, which will be noted in the text below. Regardless of which ZIP file you choose, the next step is to extract the files in it to a folder on your computer, like:

C:\Users\damodei\Documents

This will create a new folder:

C:\Users\damodei\Documents\DIA

It will contain the files necessary for this tutorial. Open the file DIABlank.sky in this folder, either by double-clicking on it in Windows Explorer, or by choosing **Open** from the **File** menu (Ctrl-O) in a running instance of Skyline.

Setting Up an Isolation Scheme for DIA

The Skyline document you have opened should be completely blank, with no transitions, no mass spec data, and no special settings. To illustrate the entire process for analyzing DIA data in Skyline, you will build up a Skyline document appropriate for DIA analysis from scratch, filling in the necessary settings, transitions, spectral libraries, and retention time information.

If you were actually performing an experiment using the DIA with interspersed DDA workflow, you would first configure your instrument for both the DIA and DDA runs, with only very general consideration of any peptide targets you might be interested in, like making sure the precursor m/z range covered contains the targets. The DDA method is left to you, but Skyline can help you set up the DIA method by allowing you define your "isolation scheme" (a pattern of precursor isolation windows to fragment for MS/MS). Even if you have already collected DIA data, you will

2

need to define the isolation scheme you used to allow Skyline to process your DIA runs. To define an isolation scheme for this tutorial experiment, take the following steps:

- On the **Settings** menu, click **Transition Settings**.
- Click the **Full-Scan** tab.
- In the **Acquisition method** dropdown list, select 'DIA' to tell Skyline that you will be importing DIA data.
- In the **Isotope peaks included** dropdown list, select 'Count'.
- In the **Peaks** field, enter '3'.

For most tryptic peptides, the first 3 isotope peaks are the most intense. You can also use an intensity threshold based on percentage of the base (most intense) isotope peak, but these settings are a reasonable default for tryptic peptides.

- Set the **Precursor mass analyzer** and **Product mass analyzer** to 'Orbitrap'. (This data was collected on a Q-Exactive, in which both MS1 and MS2 scans were performed using the Orbitrap.)
- Set the **Resolving power** under **MS1 filtering** to '35000' **At** '200' and under **MS/MS filtering** to '17500' **At** '200'.

The **Transition Settings** form should look something like this:

You have set the basic full-scan instrument parameters. Next, you need to specify the DIA isolation scheme, or the pattern of precursor *m/z* ranges that the instrument cycles through when performing DIA. For example, in the data set for this tutorial, the Q Exactive instrument used a cycle of precursor *m/z* ranges starting with 500 to 520 *m/z*, then 520 to 540 *m/z*, all the way up to 880 to 900 *m/z* (or 20 consecutive 20 m/z windows from 500 to 900 *m/z*), and then repeated the cycle. To specify this isolation scheme in Skyline, follow these steps:

- On the **Isolation scheme** dropdown list menu, select **<Add...>**.

The **Edit Isolation Scheme** form should appear as follows:

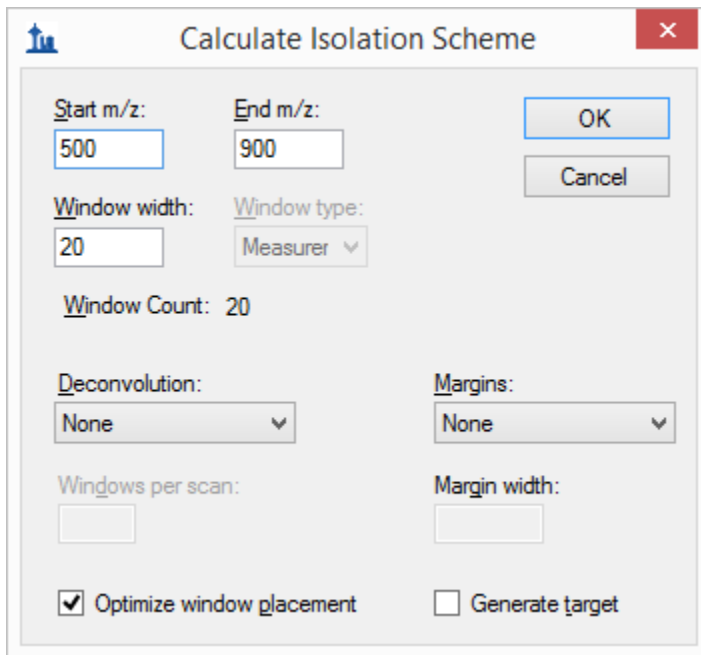- Choose **Prespecified isolation windows**.

This will enable a grid allowing you to specify the isolation windows. You can enter window boundaries manually in the grid, but since the cycle of window boundaries is highly regular in this case (from 500 to 900 *m/z* in increments of 20 *m/z*), there is a faster way to specify the boundaries:

- Click the **Calculate** button to show the **Calculate Isolation Scheme** form.
- In the **Start m/z** field, enter '500'.
- In the **End m/z** field, enter '900'.
- In the **Window width** field, enter '20'.

- Check the **Optimize window placement** check box.

This will place window boundaries in regions where peptides do not generally occur, alleviating the need for the overlapping windows (with 0.5 m/z margins on either side) proposed in the original paper on SWATH[2] if the Q1 isolation is reasonably efficient at the borders[3]. If the DIA data you are working with has already been acquired, it is important that the isolation scheme defined here reflects the instrument settings that have been used for acquisition.

The **Calculate Isolation Scheme** form should look like this:

- Click the **OK** button.

Skyline will automatically fill in the boundaries of the 20 windows required to cover 500 to 900 m/z in intervals of 20 m/z. The **Edit Isolation Scheme** form should look like this:

Skyline also allows you to visualize the precursor *m/z* range isolations over time, to check that what you have entered is correct:

- Click the **Graph** button beside the **Calculate** button you just clicked.

You should see a graph of the cycle of isolation windows as it steps forward in time, with cycles/time on the y axis and *m/z* on the x-axis:
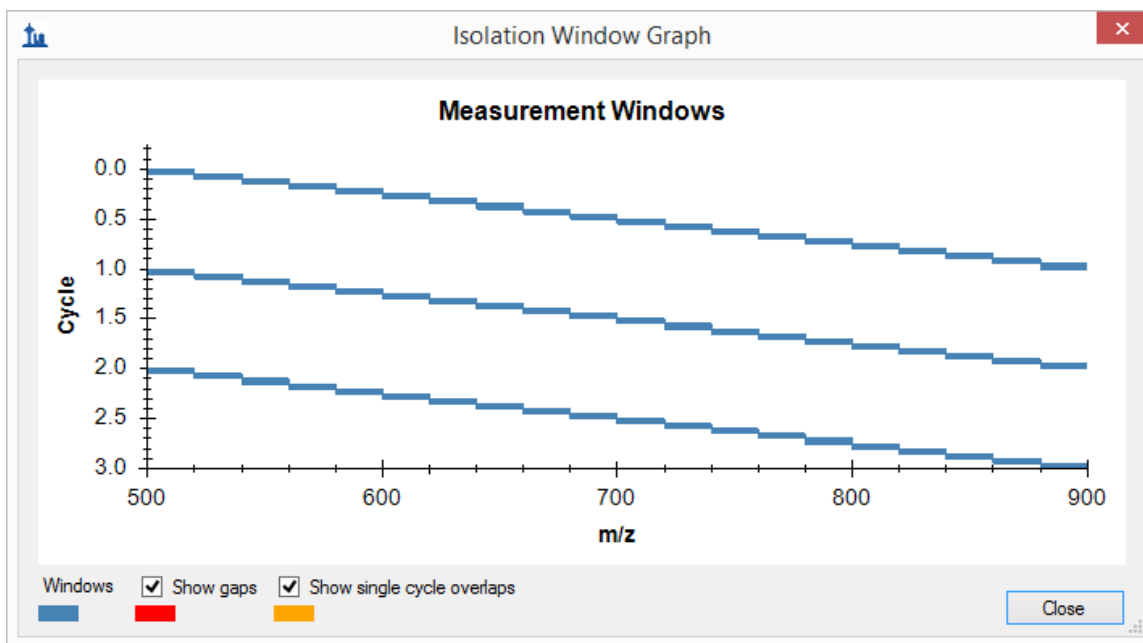
- Click the **Close** button.
- In the **Name** field, enter the name '500 to 900 by 20' for this isolation scheme.
- Click the **OK** button in the **Edit Isolation Scheme** form.
- Click the **OK** button in the **Transition Settings** form.

Note that in DIA the instrument does *not* require a set of specific transitions (SRM) or precursors (PRM) to measure, so your document actually now has all the information needed to set up a DIA run, despite still being empty of targets. The DIA isolation scheme can be exported to an instrument as follows:

- On the **File** menu, choose **Export** and click **Isolation List**.

The **Export Isolation List** form will appear, allowing you to choose the format of the isolation list export.

- From the **Instrument type** dropdown list, choose **Thermo Q Exactive**.

The form should look like this:

- Click the **OK** button.
- In the save form that appears, navigate to the folder you created for this tutorial.
- In the **File name** field, enter the name 'DIA_tutorial_isolation_list.csv'.
- Click the **Save** button.

Open the file you saved. It should look like this:

This isolation scheme is formatted for a Thermo Q Exactive (on which the data in this tutorial was acquired), but Skyline can also export to several other instrument types.
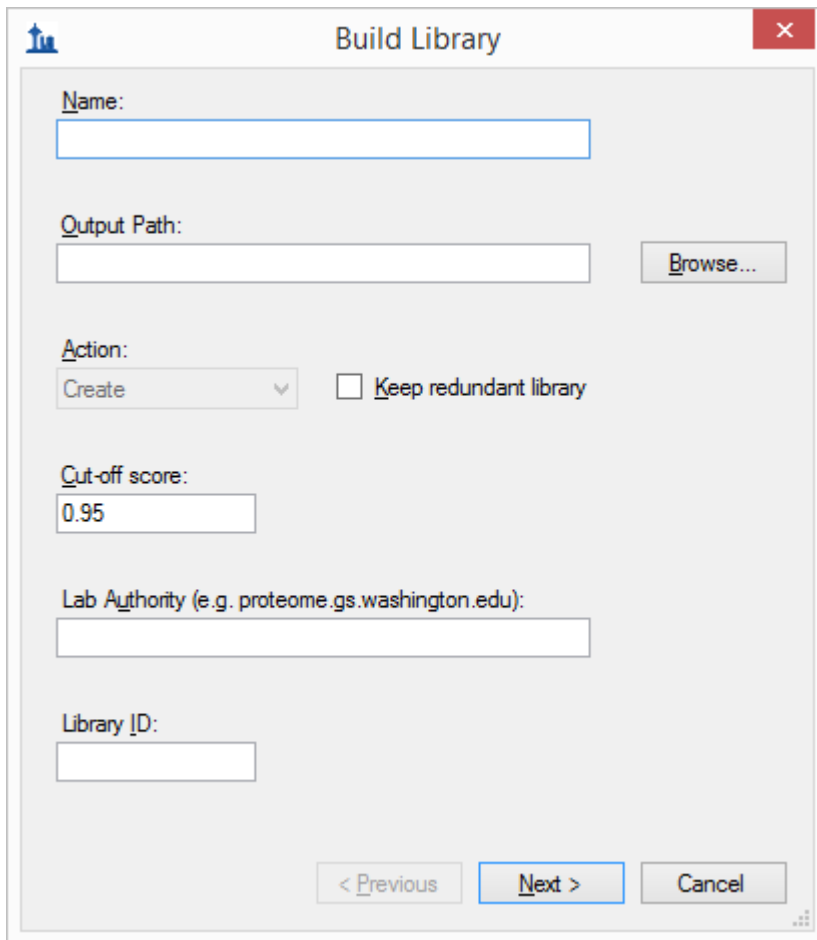
If this were a real DIA experiment, then you would use this isolation list file to run your DIA acquisition on your instrument. Alternatively, you can also simply specify the isolation scheme manually in the instrument software. Other method parameters for the data acquisition (for example MS/MS isolation width and resolution) will need to be set manually in the method file. In this tutorial, you have been provided with DIA results, although you will not actually import them until the "Importing Mass Spec Data" section below.

Setting Up Spectral Libraries from DDA Runs

Assuming now that all DDA and DIA runs for your experiment are complete, as they actually are for this tutorial, you would start the data analysis workflow by matching MS/MS spectra from the DDA runs to peptides using a search engine such as X! Tandem, resulting in a series of .xtan.xml files, or .pep.xml files, if you ran Peptide Prophet and the Trans Proteomic Pipeline (TPP). In this tutorial, if you downloaded DIA.zip (and not DIASmall.zip), you are provided a single .pep.xml file for a DDA run and a .mzXML file conversion (804 MB) of the original raw DDA data file. The first step of configuring the Skyline document for analysis of the associated DIA runs will be to import these search results into Skyline, creating a spectral library that contains MS/MS spectra and the retention times at which they were acquired. To import the DDA search results do the following:

- On the **Settings** menu, click **Peptide Settings**.
- Click the **Library** tab in the **Peptide Settings** form.
- Click the **Build** button to build a spectral library.

Skyline should present the **Build Library** form as follows:

- In the **Name** field, enter the name 'DIA_tutorial_library' for the library you will create.
- In the **Lab Authority** field, enter 'proteome.gs.washington.edu'.
- Click the **Browse** button.
- Make sure the 'DIA_tutorial_library.blib' file is located in the directory where you extracted your tutorial files.
- Click the **Save** button.
- Make sure the other fields are as shown above, i.e. **Cut-off score** at 0.95 and the box for **Keep redundant library** unchecked.

For this data set a **Cut-off score** field set to '0.95' means peptide spectrum matches with a PeptideProphet score of 0.95 or better will be included, because the DDA data was processed with the TPP (after peptide spectrum matching with SEQUEST). For other spectrum matching pipelines that use q values or expectation scores, where zero is best and 1 worst, the cut-off score uses 1 – score. So that, 0.95 would mean ≤0.05. For a reusable library, you would probably want to use a more stringent cut-off like 0.99 (i.e. ≤0.01 or 1% false discovery rate for q values). The redundant library is only necessary, if you might add more spectrum matching results to your library in the future, or if you are extracting chromatograms from the runs included in your library and you want to be able to review the original spectra for all peptide matches.

- Click the **Next** button.

The **Build Library** wizard should advance to the next page, as shown below:
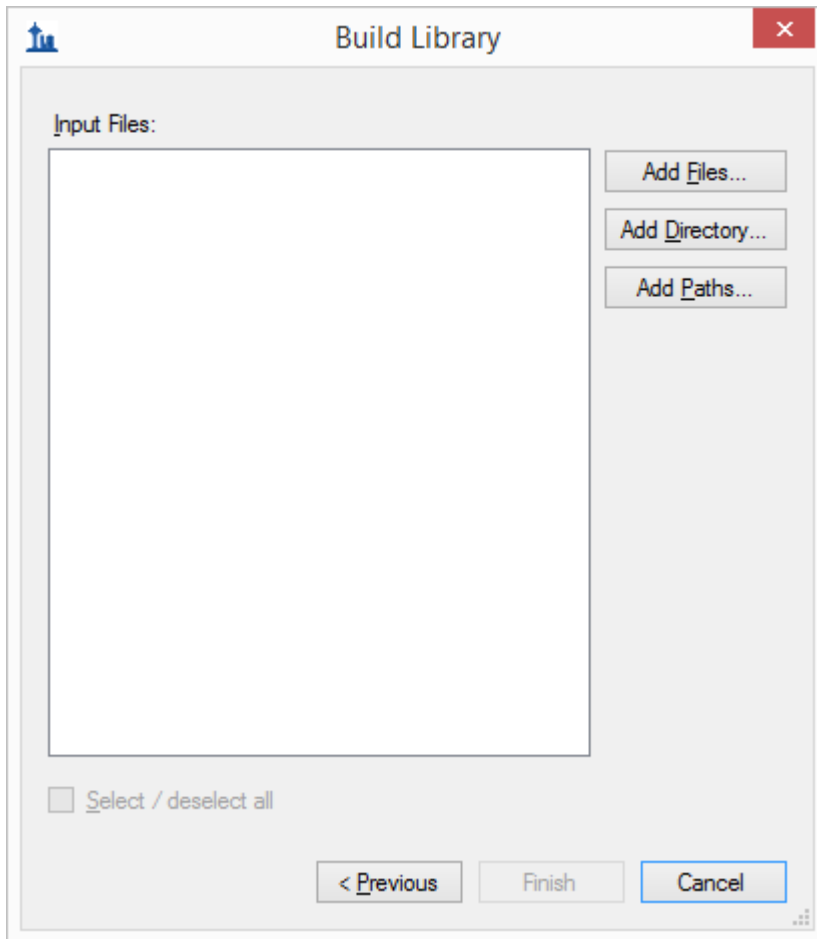
- Click the **Add Files** button.
- In the **Add Input Files** form that appears, select the following file from the directory where you put the tutorial:
  - ○ interact-20130311_DDA_Pit01.pep.xml

This file contains peptide spectrum matching results from a single DDA run. For a real experiment, you would actually perform the DDA acquisitions on a mass spectrometer, then run the output files through a search engine to produce one or more of these files (usually one for TPP-produce pepXML). Here, they have simply been provided for you. Note that the original DDA run data file (converted to mzXML), interact-20130311_DDA_Pit01.mzXML, is also present in the same folder. You do not need to import this file for chromatogram extraction, but it needs to be present in order for the library builder to find the MS/MS spectra for the library, which are not present in the .pep.xml file. Other spectrum matching pipeline outputs, such as Mascot DAT files, Proteome Discoverer MSF files and X! Tandem native XML files include all of the necessary information in a single output file.
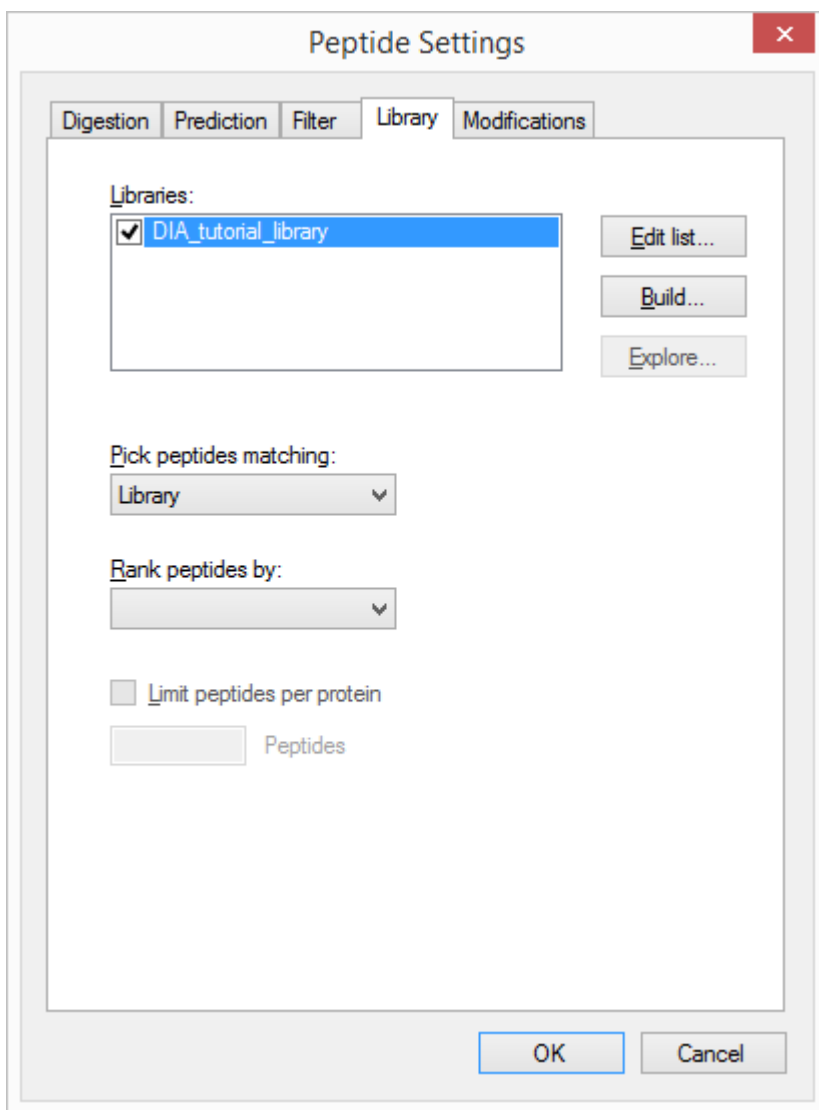
- Click **Open** to add these files to the library you are building.
- Click **Finish** in the **Build Library** form to build the library.

Note: If you downloaded the small version of the tutorial, you should see an error message, because the original DDA .mzXML file is missing from the small version. If you downloaded the small version, you should now open the Skyline document "DIASetup.sky", which contains a successfully loaded spectral library, and skip to the next section.

If you downloaded the full version, in the **Peptide Settings** form, you should see the library you are building appear in the **Libraries** list.

- Check the check box left of the 'DIA_tutorial_library' to include it for use in analyzing the DIA data set.

The **Peptide Settings** form should look like this:

- In the **Pick Peptides Matching** drop down list, make sure 'Library' is chosen.
- Click the **OK** button to confirm the new settings and library.

All the peptides matched to MS/MS spectra and their retention times from the DDA runs are now available for inclusion in your Skyline document (which you may review under View → Spectral Libraries). Later in this tutorial, Skyline will use this information for choosing the most abundant fragment ions to extract from the DIA data, and for matching the resulting chromatogram peaks using measured fragment ion relative abundance and MS/MS spectrum retention time.

Limiting Chromatogram Extraction Retention Time Range

Before importing DIA data you still need to tell Skyline over what range of retention time (RT) to extract each chromatogram. Unlike scheduled SRM experiments, in which the instrument records every transition only during a limited time window around the expected LC elution time, in DIA MS/MS spectra covering all possible fragments for a precursor are recorded throughout the whole run. It is, therefore, possible to extract a chromatogram over the entire gradient for every targeted transition. However, in practice, extracting over the full gradient often leads to highly confusing chromatograms with many distracting interference peaks, making it difficult for both humans and automated software to find the correct peak in the chromatogram. Using full gradient chromatograms will also make your files much larger and data import times take longer. For these reasons, although Skyline allows full-gradient extraction, the Skyline team strongly recommends restricting the chromatogram range based on a RT predictor. That is, whereas in SRM chromatogram time ranges are usually limited through scheduling expressed in the instrument acquisition method, in DIA we recommend limiting chromatogram time ranges during chromatogram extraction.

Limiting the RT range requires a RT prediction for each peptide, and there are several methods for predicting RT. This tutorial focuses on the simplest and most accessible method, which is simply to use the raw peptide ID times found in DDA runs (these will be in the spectral library you have built). This fails to correct for any chromatographic changes (e.g. uniform shifts) between the DDA and DIA runs, but nevertheless many users have found it to be accurate enough for their purposes. Retention time prediction can also be achieved in Skyline using an algorithm based on the peptide sequence (SSRCalc[4]) as well as an approach based on empirical RT measurements that have been normalized and stored (termed "iRT"[5], for more information see the iRT Retention Time Prediction tutorial).

To limit the RT range based on peptide ID times in your spectral library, follow these instructions:

- On the **Settings** menu, click **Transition Settings**.
- Click the **Full-Scan** tab.
- In the **Retention time filtering** pane near the bottom of the **Full-Scan** tab, select **Use only scans within** [5] **minutes of MS/MS IDs**.

The pane should look like this:

This tells Skyline it should only extract intensities from DIA spectra acquired within 5 minutes of a peptide spectrum match found in your library. For a single peptide spectrum match, the total extraction window will be 10 minutes. If there is more than one ID for a given peptide, Skyline will extract from spectra acquired within the range between the minimum ID time minus 5 minutes and the maximum ID time plus 5 minutes. (Note that all ID retention times are stored even in the non-redundant library you have built, as you will see later.) The second option, **Use only scans within** [5] **minutes of predicted RT**, uses an RT predictor (for example SSRCalc or an iRT calculator) to determine the extraction time range. You will not use an RT predictor in this tutorial, but they are discussed in several other tutorials and the Importing Assay Libraries tip (on the Skyline web page under Tips > Working with Other Quantitative Tools)**.**

The option to **Include all matching scans** will cause Skyline to extract the chromatogram over the entire gradient (between any **Min time** and **Max time** specified in the **Instrument** tab). This is occasionally useful, but it is not recommended in general as discussed above, and you will not use it here.

- Click the **OK** button in the **Transition Settings** form.

Nothing will immediately change in your Skyline document, but when you import DIA data later in the tutorial, the chromatogram time ranges will be restricted.

Importing Targets from a FASTA File

Up to this point you have not done anything to define which proteins, peptides, or transitions will be targeted for your DIA analysis. For DIA, you do not need to do this until after the mass spec data is collected, and just before it is imported. In fact, the current Skyline document can be used as a template, over and over again, to extract any desired set of transitions from the DIA .raw files associated with the DDA run used to build the spectral library in this tutorial.

Although, you can measure hundreds to thousands of peptides (perhaps even more) in a DIA run, this tutorial looks instead at only a small set of 6 proteins, specifically targeted by the experimental design that produced the data. The tutorial folder contains a FASTA file specifying these proteins. Before importing it, first configure which peptides and transitions Skyline should include within these proteins:

- On the **Settings** menu, click **Peptide Settings**.
- Click the **Filter** tab.

- Check the **Auto-select all matching peptides** check box.

This tells Skyline to automatically populate the **Targets** list with eligible peptides, once you import the FASTA file.

For this experiment, you can leave the other settings in the **Filter** tab as they are, as they will be ignored in favor of any peptide with a spectrum match in the library you have built. If you wanted these filter settings to apply, you would have to change the **Pick peptides matching** option to any choice than 'Library'. Using 'Library and Filter' would allow only peptides both found in the library and satisfying the **Filter** tab criteria.

The peptides in this particular experiment were carbamidomethylated on cysteine, so you need to tell Skyline to add this as a fixed modification:

- Click the **Modifications** tab.
- If you see the modification "Cabomidomethyl (C)" in the **Structural modifications** list, make sure it is checked.

Otherwise:

- Under **Structural Modifications**, click the **Edit List** button.
- Click the **Add** button in the **Edit Structural Modifications** form.
- In the **Name** field, click the drop-arrow at the right, to view a list of Unimod modifications.
- Find "Carbamidomethyl (C)**"** in the list and click it.

The **Edit Structural Modification** form should now look like this:

- Click the **OK** button in the **Edit Structural Modification** form.

- Click the **OK** button in the **Edit Structural Modifications** form.
- In the **Structural modifications** list, check the 'Carbomidomethyl (C)' check box.

The **Peptide Settings** form should look like this:

- Click the **OK** button.

Finally, you need to tell Skyline which transitions to extract from each peptide. With so few proteins being targeted, you can start by specifying a very broad range of transitions, and then narrow this down later:

- On the **Settings** menu, click **Transition Settings**.
- Click the **Filter** tab.
- In the **Precursor charges** field, enter '1, 2, 3, 4'.

- In the **Ion charges** field, enter '1, 2'.
- In the **Ion types** field, enter 'y, b, p'. (The p will add precursor ions extracted from MS1 spectra.)
- In the **From** dropdown list, choose 'ion 3'.
- In the **To** dropdown list, choose 'last ion - 1'.
- Check **Auto-select all matching transitions**.
- Check **Use DIA precursor window for exclusion**.

The **Transition Settings** form should look like this:

These settings tell Skyline to include a wide range of transitions including MS1 and MS/MS scans. The **Use DIA precursor window for exclusion** tells Skyline not to include transitions that fall in the DIA isolation window (e.g. a product ion at 513 *m/z* will not be included if the precursor falls in the 500 to 520 *m/z* DIA window). This may be a desirable exclusion in some cases, because precursor ions escaping fragmentation can cause increased noise and potential interference in the MS/MS spectra over the original isolation range, making quantification based on fragment ions in this range less reliable.

- Click the **OK** button.

Now you are ready to import the FASTA file:

- On the **File** menu, choose **Import** and click **FASTA**.
- In the **Import FASTA** form, navigate to the tutorial folder you have been working in and double-click the file named 'pituitary_database.fasta'.

A list of 26 peptides belonging to the 6 proteins in the FASTA file should appear in the **Targets** pane. For every protein, only the peptides that were present in the spectral library are included in this list.

- On the **Edit** menu, choose **Expand All** and click **Precursors**.

This will show the full set of transitions Skyline has included. Note that every peptide includes all eligible precursors between 500 and 900 m/z, and that every precursor includes all possible b and y transitions as well as 3 precursor ions (M, M+1 and M+2). Note, also, that every precursor has a corresponding library spectrum. The **Targets** view should look like this:

```
Targets                                    ⊣ ×
⊟ ⃛ sp|P01222|TSHB_HUMAN                      ∧
  ⊟ ⓟ  K.YALSQDVCTYR.D [64, 74]
    ⊟ ⅃  688.3192++
       ⃛ precursor - 688.3192++ (irank 1)
       ⃛ precursor [M+1] - 688.8206++ (irank 2)
       ⃛ precursor [M+2] - 689.3213++ (irank 3)
       ⃛ L [y9] - 1141.5306+ (rank 4)
       ⃛ S [y8] - 1028.4466+ (rank 2)
       ⃛ Q [y7] - 941.4145+ (rank 5)
       ⃛ D [y6] - 813.3560+ (rank 3)
       ⃛ C [y4] - 599.2606+ (rank 7)
       ⃛ T [y3] - 439.2300+ (rank 8)
       ⃛ L [y9] - 571.2690++
       ⃛ S [y8] - 514.7269++
       ⃛ Q [y7] - 471.2109++
       ⃛ D [y6] - 407.1816++
       ⃛ V [y5] - 349.6681++
       ⃛ C [y4] - 300.1339++
       ⃛ T [y3] - 220.1186++
       ⃛ L [b3] - 348.1918+ (rank 10)
       ⃛ S [b4] - 435.2238+
       ⃛ Q [b5] - 563.2824+
       ⃛ D [b6] - 678.3093+
       ⃛ V [b7] - 777.3777+
       ⃛ C [b8] - 937.4084+
       ⃛ T [b9] - 1038.4561+
       ⃛ L [b3] - 174.5995++
       ⃛ S [b4] - 218.1155++
       ⃛ Q [b5] - 282.1448++
       ⃛ D [b6] - 339.6583++
       ⃛ V [b7] - 389.1925++
       ⃛ C [b8] - 469.2078++
       ⃛ T [b9] - 519.7317++
  ⊟ ⓟ  R.TVEIPGCPLHVAPYFSYPVALSCK.C [80, 103]
    ⊟ ⅃  677.0921++++
       ⃛ precursor - 677.0921++++ (irank 3)
       ⃛ precursor [M+1] - 677.3428++++ (irank 1)
       ⃛ precursor [M+2] - 677.5934++++ (irank 2)
       ⃛ P [y12] - 1431.6977+
       ⃛ Y [y11] - 1334.6449+        ∨
```

In DIA, it is possible to extract all of these product ions, because the instrument covers every possible precursor-product combination in the 500-900 *m/z* range (at least those within the measured range of the MS/MS spectra). However, in practice, extracting every possible product ion is often unnecessary, and in fact can hinder peptide detection by adding chromatograms which are noisy and have very low signal for the desired peptide. In most cases, we recommend choosing a few transitions that are likely to have the highest intensity. This can be done using the spectral library you built earlier from DDA search results:

- On the **Settings** menu, click **Transition Settings**.
- Click the **Library** tab.
- Check **If a library spectrum is available, pick its most intense ions**.
- In the **Pick** field, enter '5' to use 5 product ions for each precursor.
- Select the **From filtered product ions** radio button.

This means that the 5 product ions selected from the library must also satisfy the filter settings defined in the **Filter** tab of the **Transition Settings**.
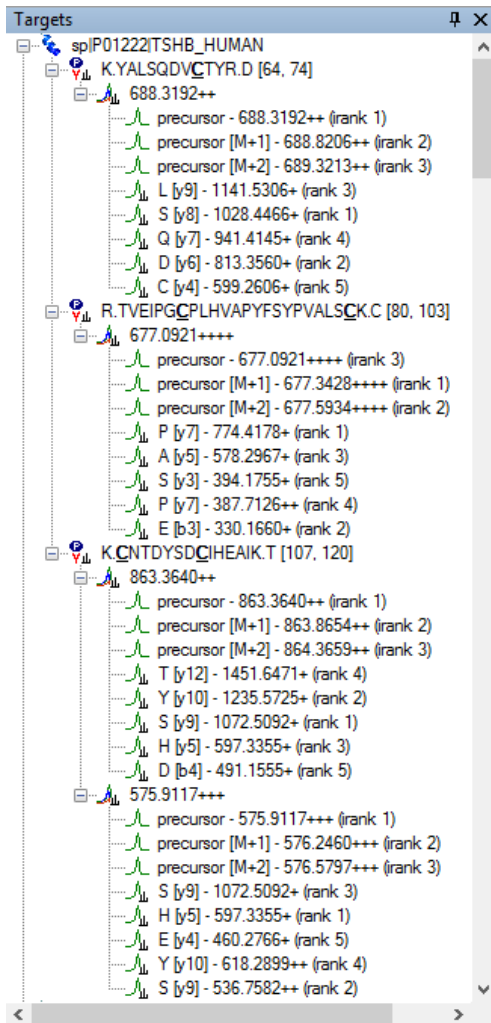
21

- Also, set the **Ion match tolerance** to 0.05, or 50 ppm at 1000 *m/z*, which is more than enough, since the DDA data was also collected with high mass accuracy.

The **Library** tab should now look like this:

- Click the **OK** button.

Look again at the **Targets** pane. It should now contain only the best 5 product ion transitions for each precursor, plus 3 precursor ion transitions:

There are many other ways to import peptides, precursors, and transitions into Skyline for use in DIA experiments, aside from FASTA files. For more details, consult the Targeted Method Editing tutorial. If you have created transition lists with other tools, you may want to consult the Existing and Quantitative Experiments tutorial. Other DIA tools may create an extended transition list format called an "assay library", which adds fragment ion relative abundance for the transitions, and retention time prediction information. Assay libraries are not used in this tutorial but are described in detail in the Importing Assay Libraries tip.

In large DIA experiments, it can be helpful to have decoy peptides[6] to train advanced peak picking (peptide detection) models, which may improve Skyline automated peak picking and allows Skyline to produce detection probability scores (q values). Decoy peptides are not strictly necessary for a DIA workflow, and are needed only if you choose to apply a custom peak scoring method. In many cases, the Skyline default peak picking performs very well and no custom peak scoring is needed, and even when custom peak scoring is used, alternatives to decoys exist (see the Advanced Peak Picking Models tutorial). So, no decoys will be added to the document for this tutorial.

Before moving to import data, it is important consider your choice of targets carefully. Any change in your targets after you import the DIA data will require the data to be reimported for updated chromatogram extraction and automated peak picking.

Importing Mass Spectrometry Data

You are now ready to actually import the DIA mass spec files included with the full version of this tutorial. These files are very large (~5 GB) and have been provided only in the full version of the tutorial. If you downloaded the small version, you should skip the rest of this section, and open the file "DIAImported.sky", which is a Skyline document with fully imported DIA runs.

If you downloaded the full version of the tutorial, then you can import the DIA data yourself. Importing DIA data is very straightforward, involving exactly the same steps as importing SRM data:

- First, on the **File** menu, click **Save As**.
- In the **File name** field, change the name to 'DIATutorial.sky'.
- Click the **Save** button.
- Next, on the **File** menu, choose **Import** and click **Results**.
- Click the **OK** button, to **Add single injection replicates in files**.
- Select the files below, using Shift+mouse-click to make a multiple selection:
  - 20130311_DIA_Pit01.raw
  - 20130311_DIA_Pit02.raw
- Click the **Open** button.

Skyline will show a form suggesting that it can remove a common prefix from the file names:
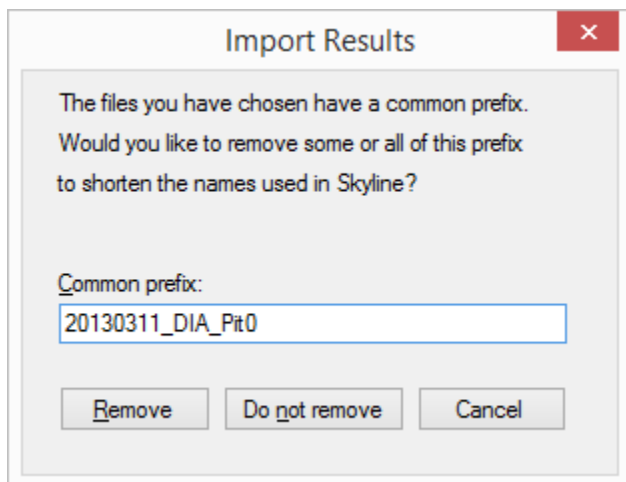
- Delete the text "DIA_Pit0" leaving just "20130311_" as the common prefix to remove.
- Click the **Remove** button.

Chromatograms for all the precursor and product ions in the **Targets** pane will now be extracted from the DIA data. This step will take a few minutes. Once loading is complete, chromatograms will be displayed for both target and decoy peptides.

Exploring DIA Results

Now that the DIA runs are imported, you can take a look at the results:

- Make sure the file DIA_Pit01 is selected in the chromatogram pane.
- On the **Edit** menu, choose **Collapse All** and click **Peptides**.
- Click on the peptide R.VLQAVLPPLPQVV**C**TYR.D in the **Targets** list. (It should be the 7th peptide down.)
- On the **View** menu, choose **Transitions** and click **Split Graph**.
- On the **View** menu, choose **Auto-Zoom** and click **Best Peak** (F11).

The chromatogram graph should look something like this:
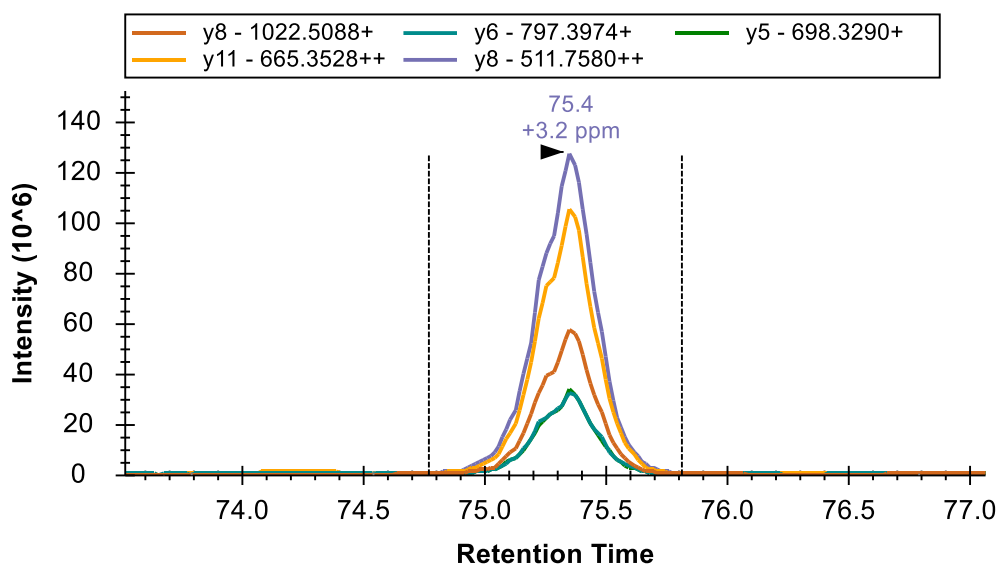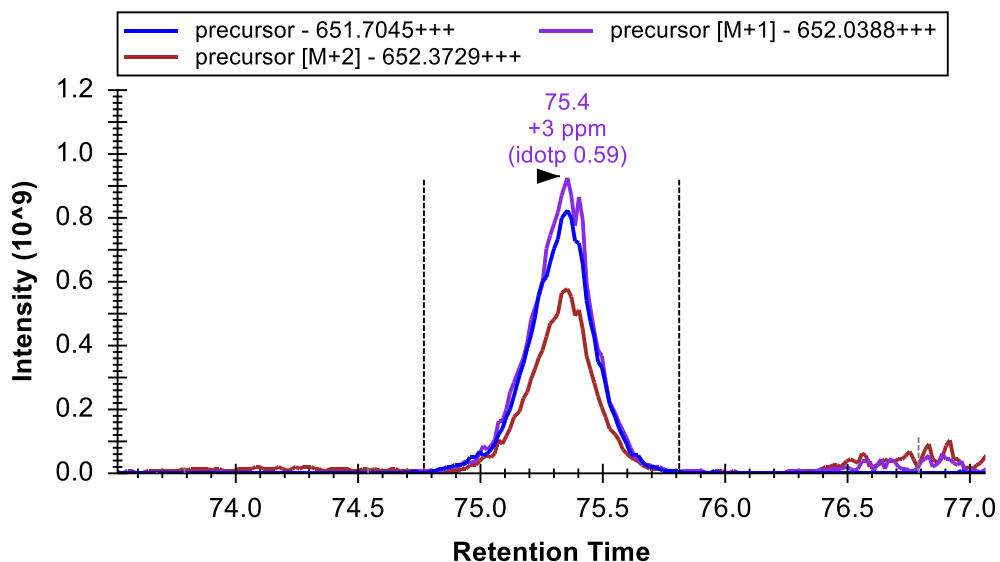
Notice that the graph containing the precursors shows "idotp 0.59". This is a measure of how well the precursor isotope peak areas correlate with the expected isotope distribution. The precursor ions come from interleaved MS1 scans that were included between cycles of the DIA MS/MS acquisition. A dot-product of 0.59 is a very low value for the expected isotope distribution, which should have lower variance than fragment ion distributions coming from MS/MS scans.

If you expand the peptide in the **Targets** view, you will see that the monoisotopic peak has a red dot, indicating that it is contributing no peak area.

```
sp|P01229|LSHB_HUMAN
    R.VLQAVLPPLPQVVCTYR.D [63, 79]
        651.7045+++ (idotp 0.59, dotp 0.83)
            precursor - 651.7045+++ (irank 2)
            precursor [M+1] - 652.0388+++ (irank 1)[1]
            precursor [M+2] - 652.3729+++ (irank 3)[2]
            P [y8] - 1022.5088+ (rank 1)[5]
            V [y6] - 797.3974+ (rank 4)[6]
            V [y5] - 698.3290+ (rank 5)[7]
            P [y11] - 665.3528++ (rank 2)[4]
            P [y8] - 511.7580++ (rank 3)[3]
```
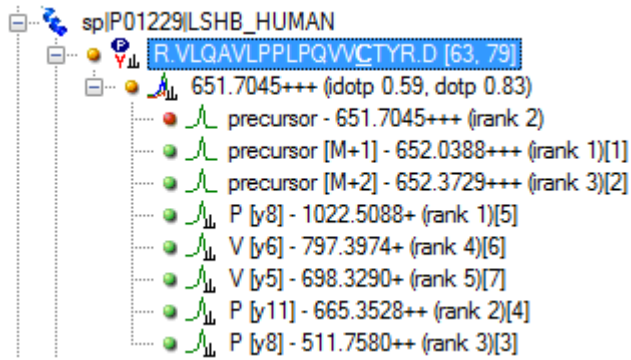
To force Skyline to include all integrated areas between the integration boundaries no matter how well peak shapes are matched, do the following:

- On the **Settings** menu, click **Integrate All**.

The red dot will turn green, and the idotp value will change to "idotp 0.99" in the **Targets** view. The idotp value will also disappear from the chromatogram view, because it is now the best matching distribution of all peak groups detected in the extracted chromatograms.

The "dotp 0.83", following the idotp value in the **Targets** view, is a measure of how well the peak areas of the 5 selected product ions correlate with the intensities found in the matching library spectrum. Both values are now relatively high and help to confirm the association between this peak group and the targeted peptide. The combination of the two high dot-products, the matching co-elution and peak shapes, and the mass errors of around 3 ppm make a strong case that elution of the targeted peptide caused this peak.

Now do the following to show the peptide ID's that were used to predict retention time:

- Right-click in the chromatogram graph, choose **Peptide ID Times** and click **From Other Runs**.

A set of vertical light blue lines should appear in the chromatogram window, like this:
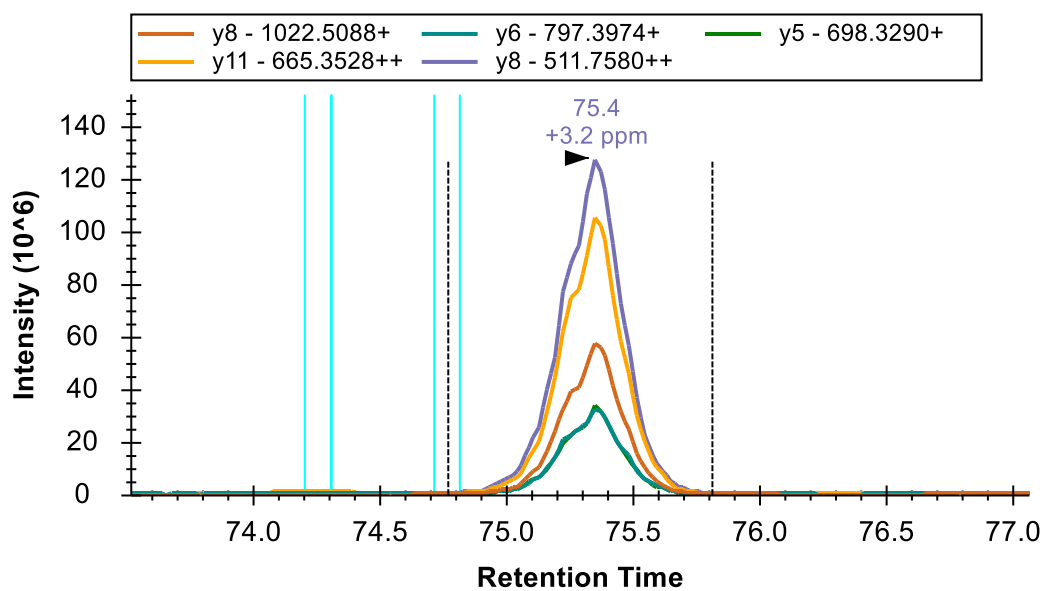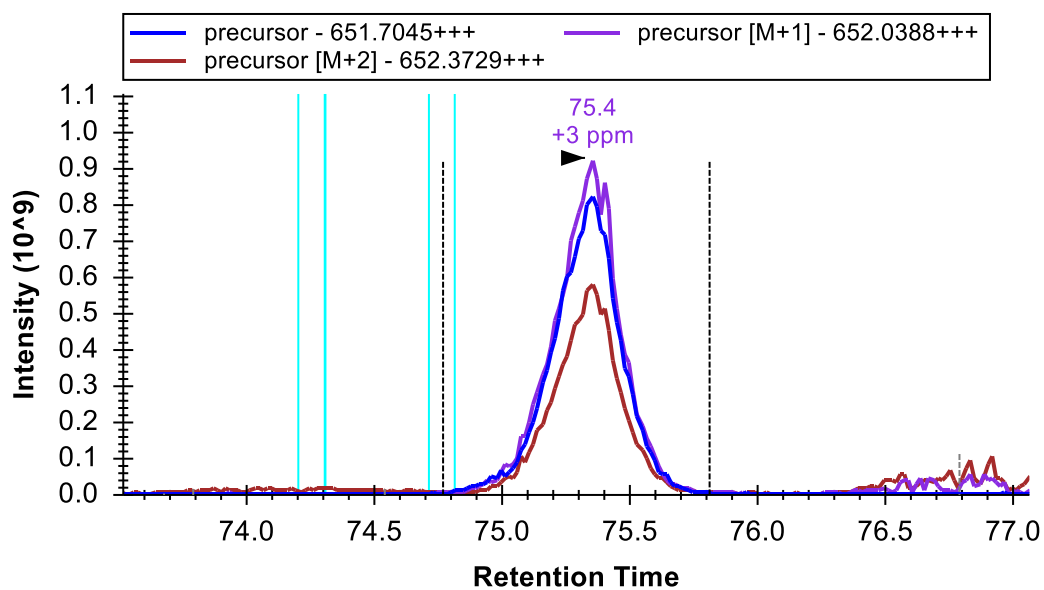
The light blue lines represent the retention times MS/MS spectra matching this peptide where acquired in the DDA run that you used to generate your spectral library. That they are within 1 minute of the integrated range adds further confirmation that the integrated peak measures the peptide of interest. (Note that it is not necessary to have a redundant library to see the retention times of all matched MS/MS spectra. All of the retention times are stored in the non-redundant library, but only one spectrum per peptide precursor.)

The Skyline automated peak picking takes into account both peak area dot-products and ID retention times when deciding which peak to pick. As you can see it has clearly chosen the correct peak in this case.

To see the full range of chromatograms extracted, do the following:

- On the **View** menu, choose **Auto-Zoom** and click **None** (Shift-F11).

This will show you that Skyline has only extracted the chromatogram in a +/-5 minute window around the DDA ID times – the ID's occur near 75 minutes and so Skyline extracts from 70 to 80 minutes.

In DIA, there can be a lot of interference, because precursor isolation windows are so wide (e.g. 10-25 m/z). However, the Skyline automated peak picking is often able to pick the correct peaks, even when a very abundant interference is present. To see this in action:

- Click the 2nd peptide R.TVEIPG**C**PLHVAPYFSYPVALS**C**K.C in the **Targets** list.
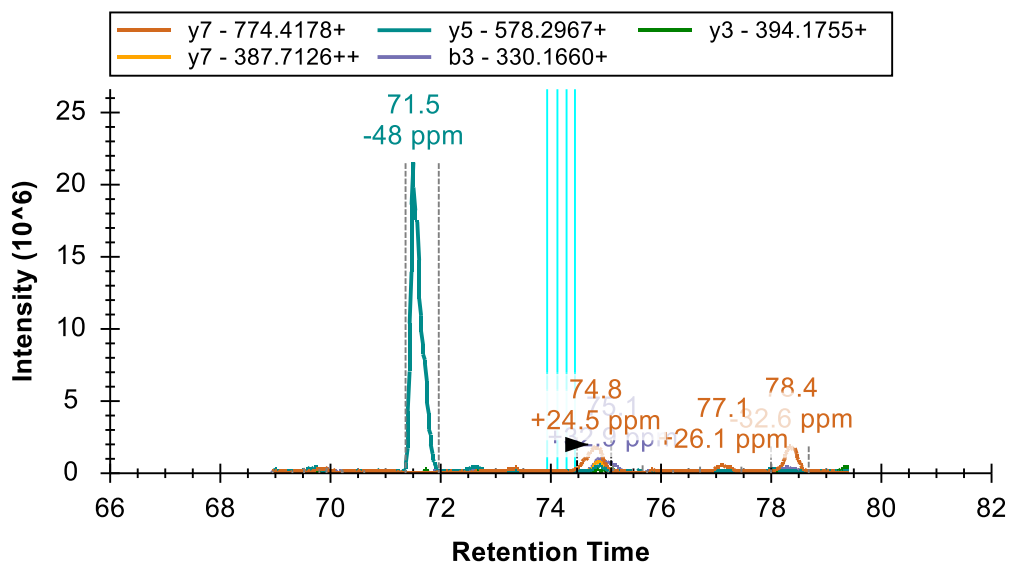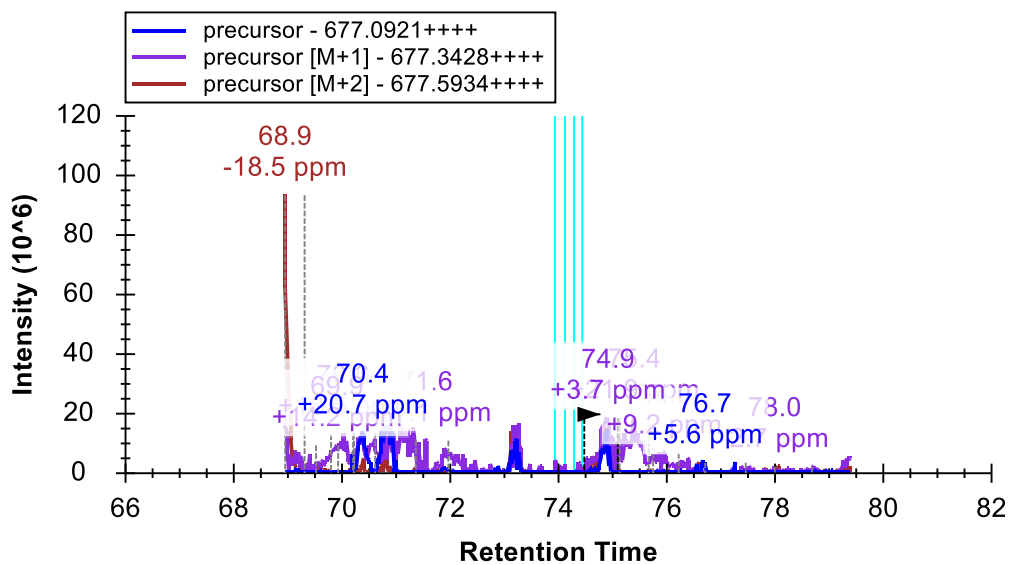
As you can see, there are a number of peak groups to choose from in the MS1 chromatograms. In the MS/MS chromatograms, there is a huge peak in one transition at 71.5 minutes and a much less intense set of co-eluting peaks at 75 minutes, closer to the predicted RT and with a high dot-product. Skyline is not distracted by the large peak (which is likely an interference) and chooses the small co-eluting peaks instead (which are certainly the most likely option).

Understanding Extracted Chromatograms

Still the +24.5 ppm mass error showing for the product ions peak is somewhat unlikely on an Orbitrap mass analyzer. To get a better understanding of what may be causing the problem do the following:

- On the **View** menu, choose **Auto-Zoom** and click **Best Peak** (F11).

Getting a closer look at the integrated peak, you should see something like this:
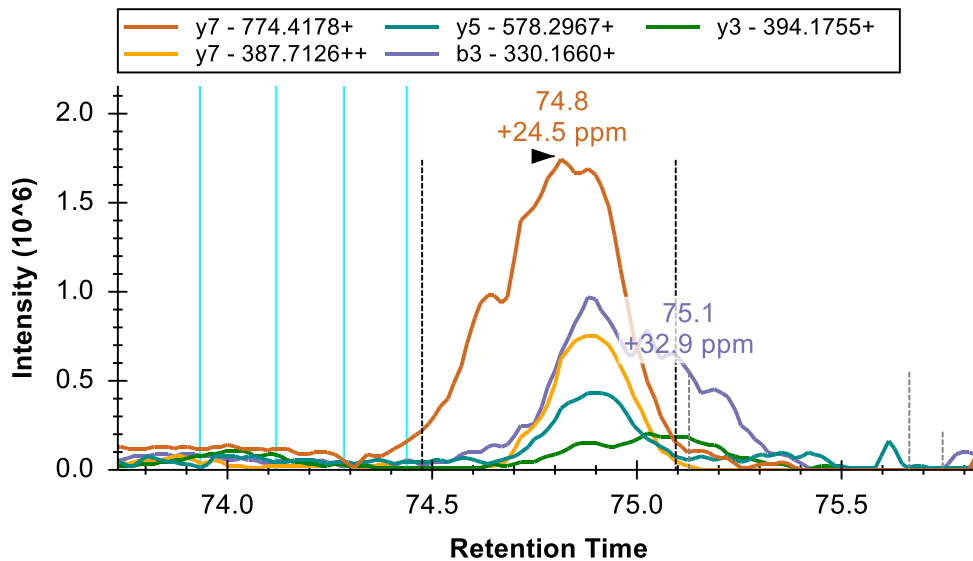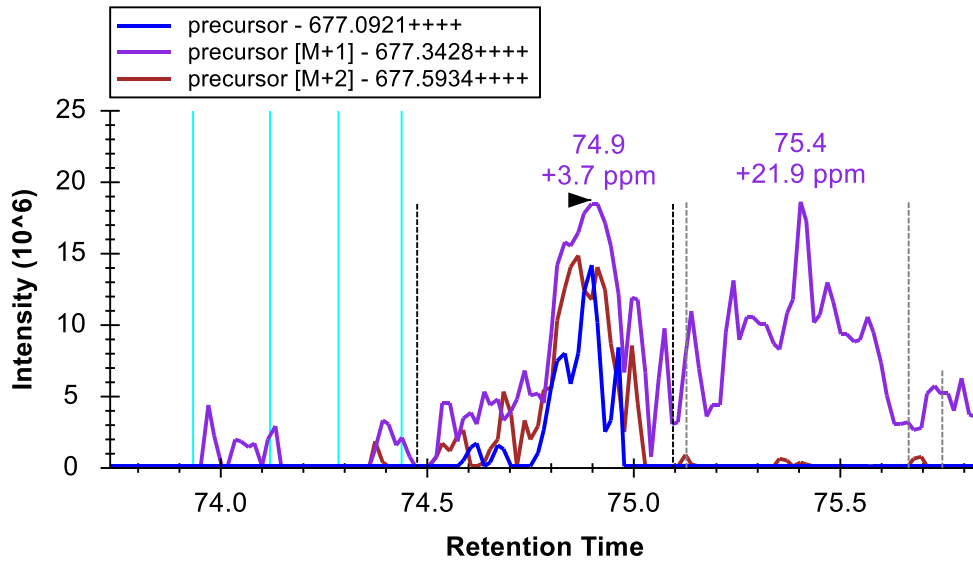
This is not quite as reassuring as the peak for R.VLQAVLPPLPQVV**C**TYR.D. Although there is clearly a lot going on in the MS1 scans around the precursor of interest near 75 minutes RT, the integrated peak has mass error of only 3.7 ppm and, if you expand the peptide in the **Targets**

view, you can see it has a 0.91 idotp. If you hover your mouse cursor over the integration boundary at 74.5 minutes in the precursors graph until you see a splitter cursor (✛) and drag right, to about 74.7 minutes, closer to the left edge of the main precursor peak, you will see the idotp value increase to 0.96, and mass error decrease to 2.3 ppm.

Note: All use of the **Full Scan** view below will only work, if you have downloaded the full tutorial data set. Otherwise, Skyline will display a message that the raw data files are missing.

To get a better sense of what might have happened in the MS1 scans from which these chromatograms were extracted, do the following:

- Hover the mouse cursor over the apex of the "precursor [M+1]" peak until you see purple circle appear, and the mouse cursor changes to a pointing hand.
- Click in the circle.

Skyline will show the **Full Scan** view displaying a range of the MS1 scan from which the precursor chromatogram points were extracted. You could see the entire MS1 scan by clicking on the button with the magnifying glass and a plus (🔍⁺) in the upper right corner of the view. Instead, do the following:

- Click and drag a rectangle around the peaks highlighted as M, M+1 and M+2 to zoom in.
- Make sure you are looking at the spectrum acquired exactly at 74.88 min (indicated in title), otherwise use the arrow buttons in the upper right corner to navigate to this spectrum.

The **Full Scan** graph should now look something like this:
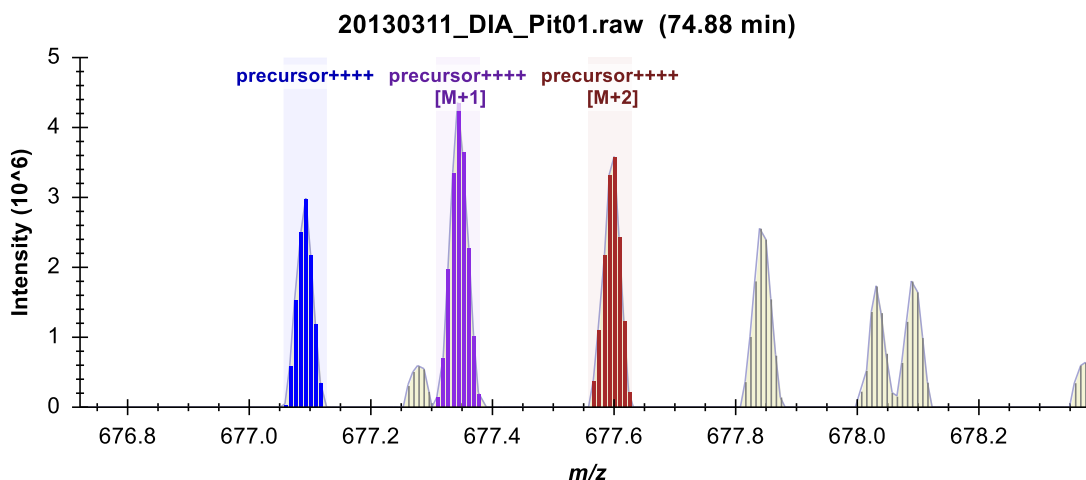
The individual intensities in the profile-mode peaks are shown as sticks, with a line and shading connecting them. The individual intensities that were summed to create the points on the chromatograms are highlighted in the chromatogram colors, and the extraction ranges are

shown as shaded regions. You can see the M+3 and the M+4 peaks, which were not extracted to the right of the shaded peaks. This looks about as one would expect for the targeted peptide. Though, you can also see two peaks (677.275 and 678.025) that come very close to interfering with the M+1 and M+4 peaks of the peptide.

- Now use the arrow buttons (⬅ ➡) in the **Full Scan** view toolbar to review scans over the elution profile of the chromatogram peak.

From about 74.77 to 74.92 minutes, you should see MS1 signal that corresponds very well with the expected isotope distribution, without anything obviously interfering in the extraction ranges. Beyond 74.92 minutes, the signal becomes less clear. If you zoom out a little (using the mouse scroll-wheel) you will see that eventually other clear peptide signal begins to take over in this *m/z* range.

To perform a similar analysis on the product ion MS/MS scans:

- Hover the mouse cursor over the apex of the "y7+" peak until you see purple circle appear, and the mouse cursor change to a pointing hand.
- Click in the circle.

You should see a graph in the **Full Scan** view that looks like this:

**20130311_DIA_Pit01.raw (74.82 min)**
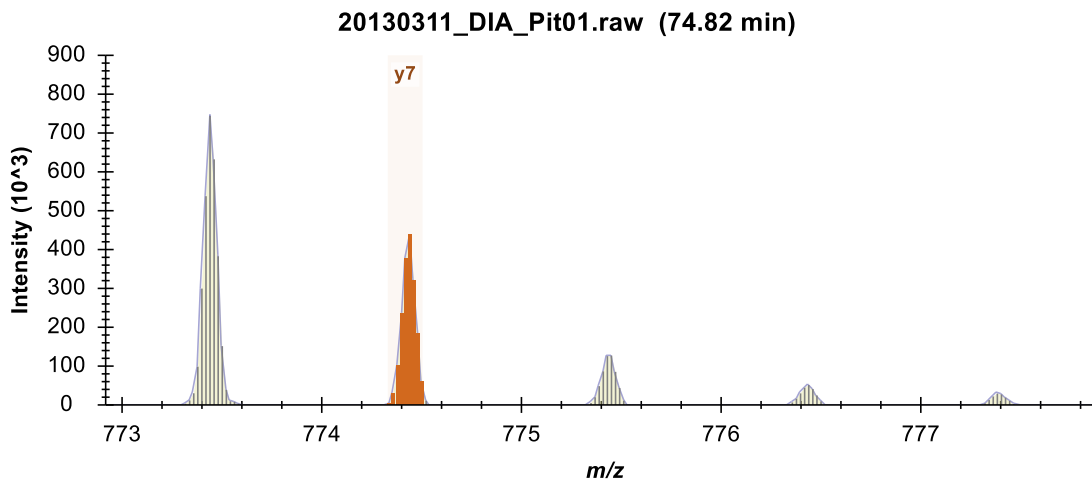
And, it is immediately obvious that the signal in the y7 chromatogram is not from the targeted fragment, because the y7 extraction range is not over the monoisotopic peak. Instead you can see that the y7 extraction range is over the second isotopic peak, while the monoisotopic peak is 1 Da lighter. Combine this with the fact that the y7 chromatogram is not co-eluting very well with the other product and precursor chromatograms, extending too far to the left, and you probably want to discard this transition (in the **Targets** pane, right-click on y7 and click **Delete**). Though, the fact that this is expected to be the most intense fragment, from the library MS/MS

spectrum and the dot-product correlation with the library spectrum is currently 0.88, sure adds some confusion.

Further, if you click on the chromatogram for the b3+ ion and zoom, you will see that its signal comes from a doubly charged ion (0.5 *m/z* between isotope peaks):

20130311_DIA_Pit01.raw (74.89 min)
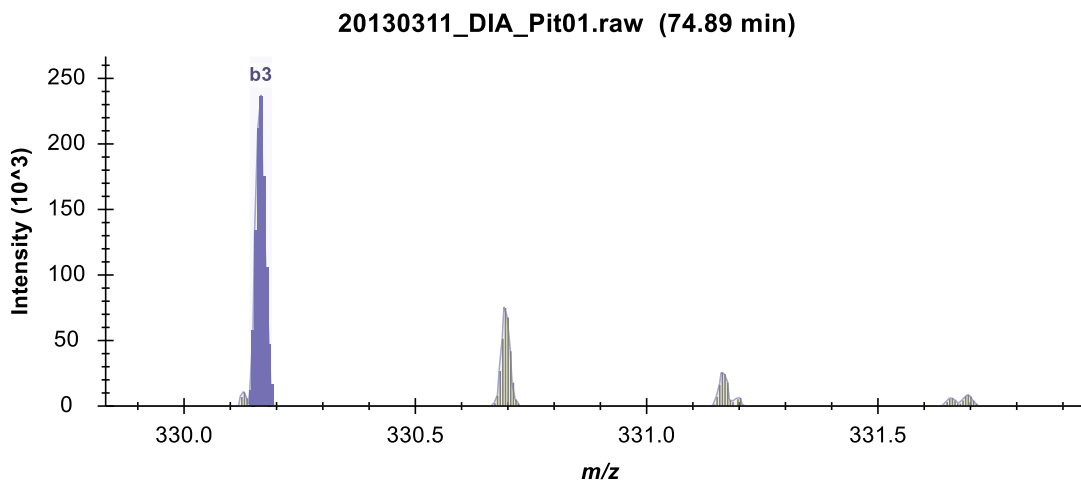
The peaks for y7++ and y5+ look reasonable, but the y3+ chromatogram clearly struggles with interference throughout its elution, and is eventually dominated by it around 74.92 minutes:

20130311_DIA_Pit01.raw (74.92 min)
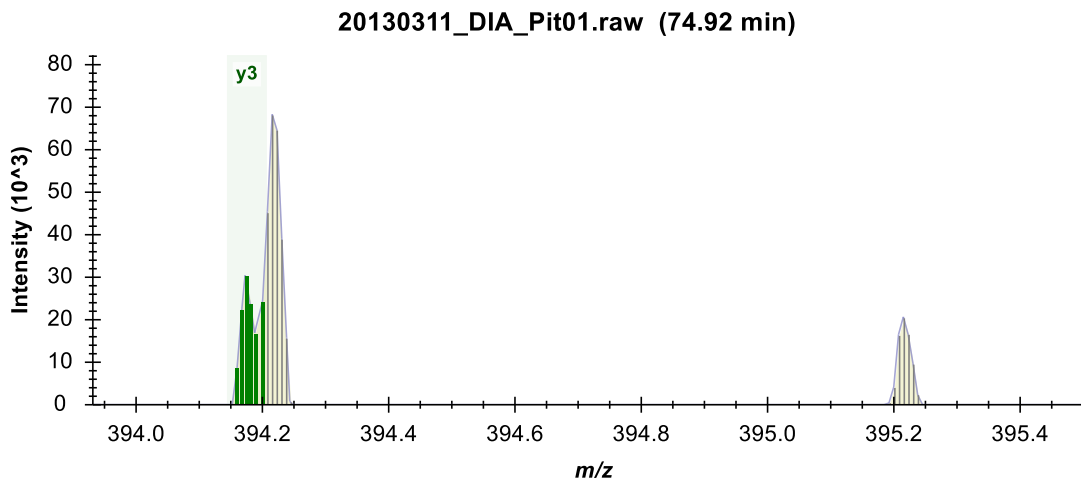
At this point, you should probably discard this peptide from consideration as quantifiable in this context. While it is possible to rerun the chromatogram extraction on new fragment ions for this peptide to find a replacement, you will not do this for this tutorial, as this would require you to first select the new transitions in the **Targets** pane and then to reimport the DIA files.

Continuing to the next peptide K.**C**NTDYSD**C**IHEAIK.T, you will see that when there are multiple charge states, the transition split view splits on charge state, but not ion type, and that the

precursor ion intensities extracted from MS1 scans typically dwarf the product ion intensities extracted from MS/MS scans:
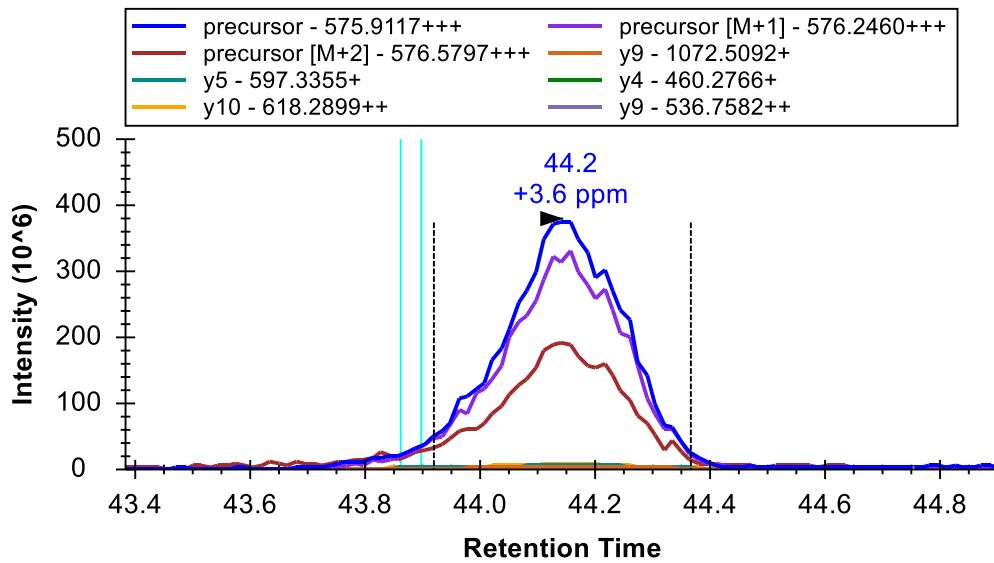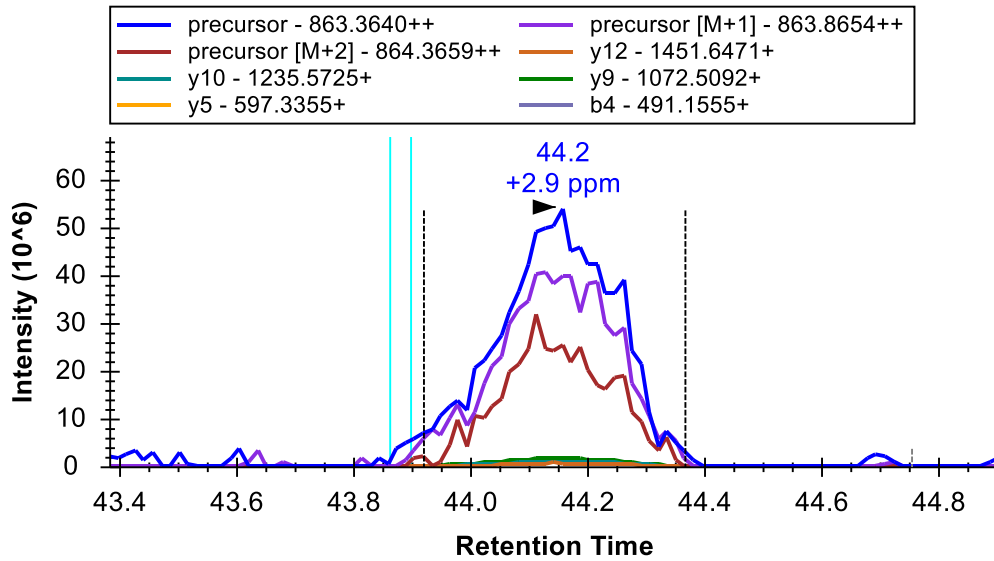

If you expand the peptide in the **Targets** view and select each precursor, you will see that both have high quality peaks with good low-noise signal on both precursor and product ion transitions, below 5 ppm mass errors, idotp values nearly 1 and dotp values greater than 0.8:
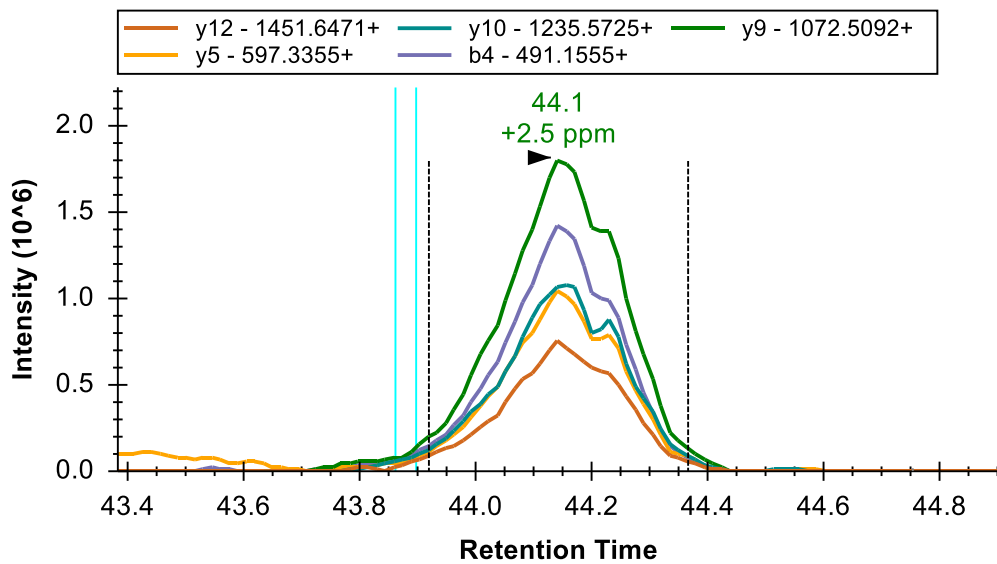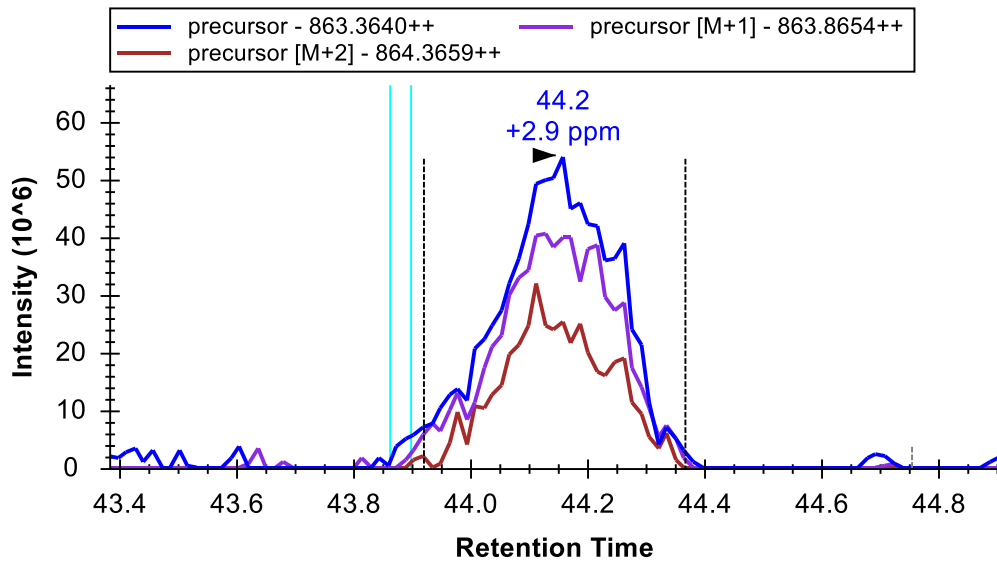
One last peptide worth considering in this tutorial is the next peptide K.ELVYETVR.V. Selecting and expanding it in the **Targets** view, you will see idotp and dotp values of 0.99, but only 3 product ion transitions. You may wonder why the y4 transition was not included, as it appears to be the second highest in the spectrum shown in the **Library Match** view:

DIA_tutorial_library - ELVYETVR, Charge 2
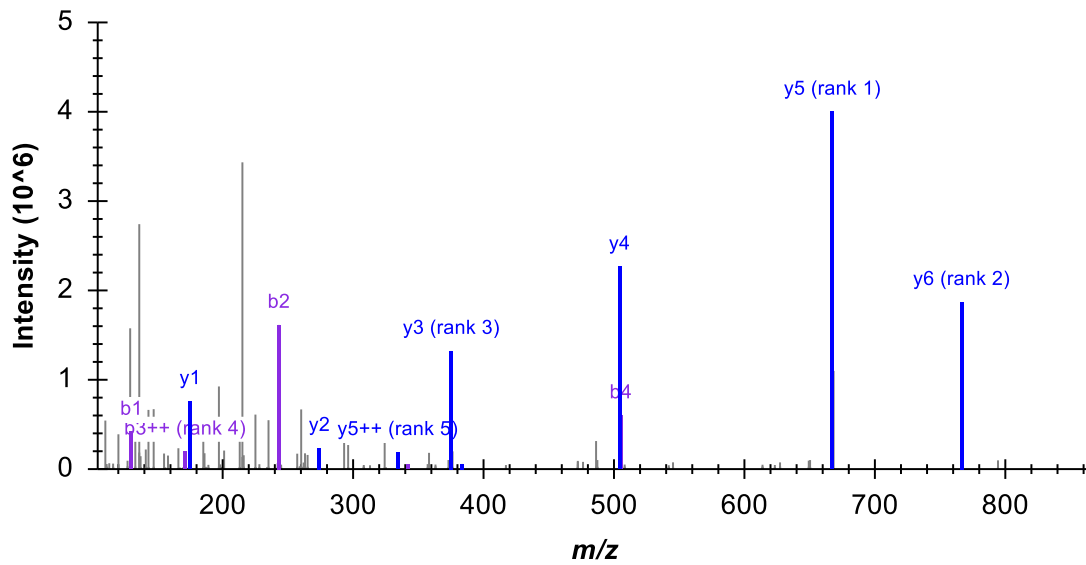
The answer to this question is that it is excluded by the **Use DIA precursor window for exclusion** checkbox you checked in the **Transition Settings – Filter** tab, since the precursor *m/z* 504.77 and the product *m/z* 504.28 both fall within the 500 to 520 isolation window. This leaves you with only y3, y5 and y6 for this peptide, which is less than ideal. You should greatly prefer to have 5 or even 6 transitions per peptide.

The chromatograms look quite good, however. The co-eluting precursors and nearby cyan ID line, now familiarly left of the peak give you added confidence in the peak assignment.
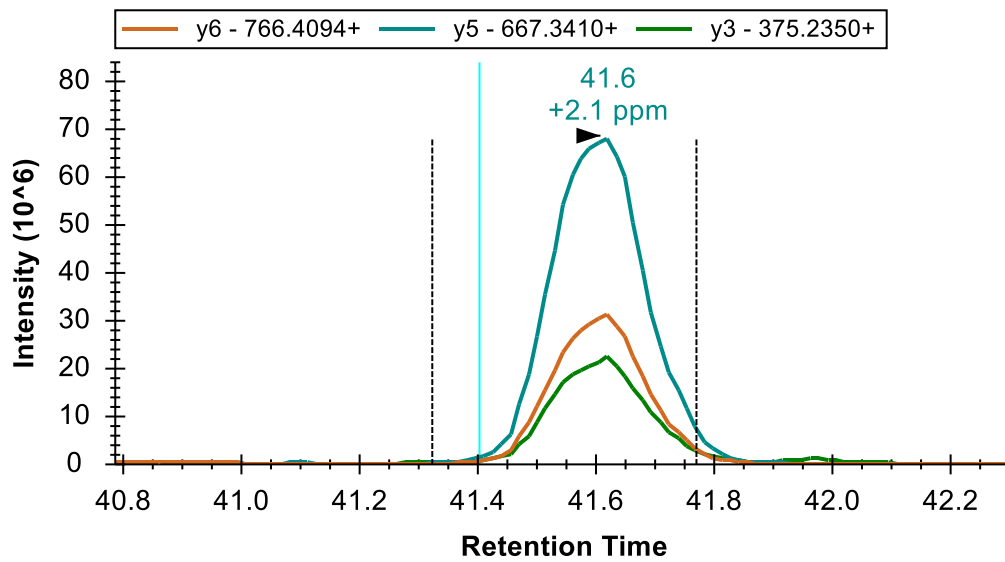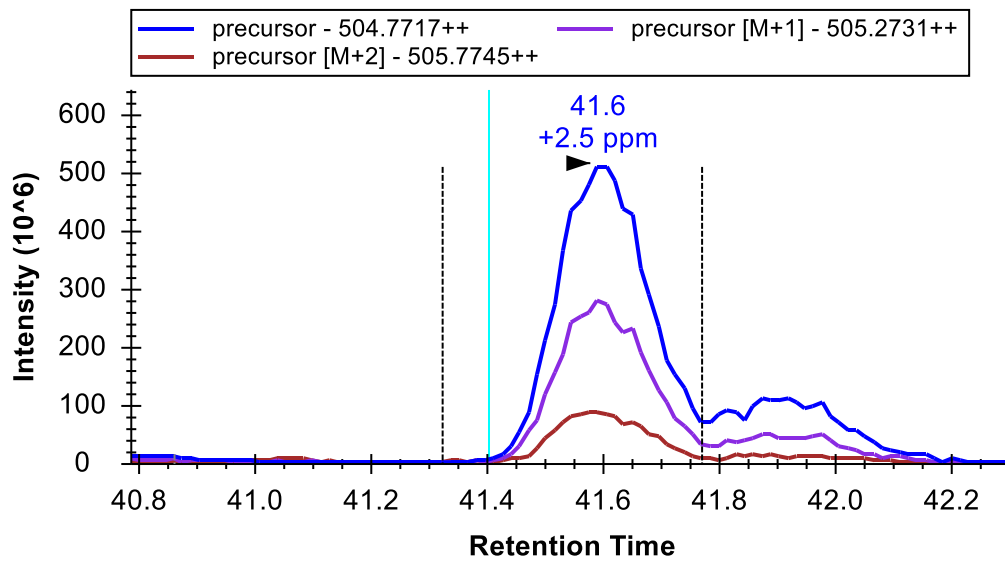
That the second peak extending from the main peak in the precursor chromatograms is an interference is made obvious by its complete absence in the product ion chromatograms. You can gain further confidence in this by clicking on it to show MS1 scans in the **Full Scan** view. You will see that the signal comes from a peptide that is 1 Da lighter than the target peptide, with clearly greater than 5 ppm mass error:

20130311_DIA_Pit01.raw (41.92 min)
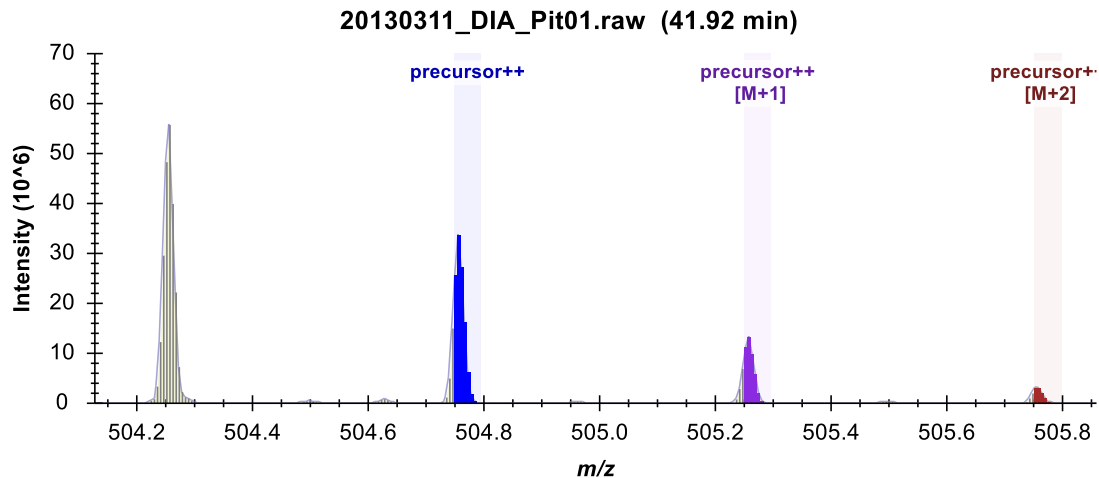
You may be thinking (quite correctly) that this peak would simply disappear, if you had used a file with centroided mass spectra (which would turn every profile peak into a single stick) and a very narrow extraction window. You can do this in Skyline by giving it centroided data and using a very high resolving power, but you should be aware that this will give you less visibility into the data as it was actually acquired by the instrument. It is also important to remember that the true profile peak resolution in the *m/z* dimension determines when you get interference, no matter whether you have used a centroiding algorithm or not. To gain a visual understanding of this:

- Click the left-arrow button (⬅) in the **Full Scan** view to traverse the MS1 scans until you reach the transition between the interference and the true peptide peak.

You will see signal from the target peptide begin to take over, and shift the center of the peaks closer to the center of the extraction range, long before the peak to the left disappears, making it clear that the signal from the two peptides is not separable where they overlap.

20130311_DIA_Pit01.raw (41.73 min)
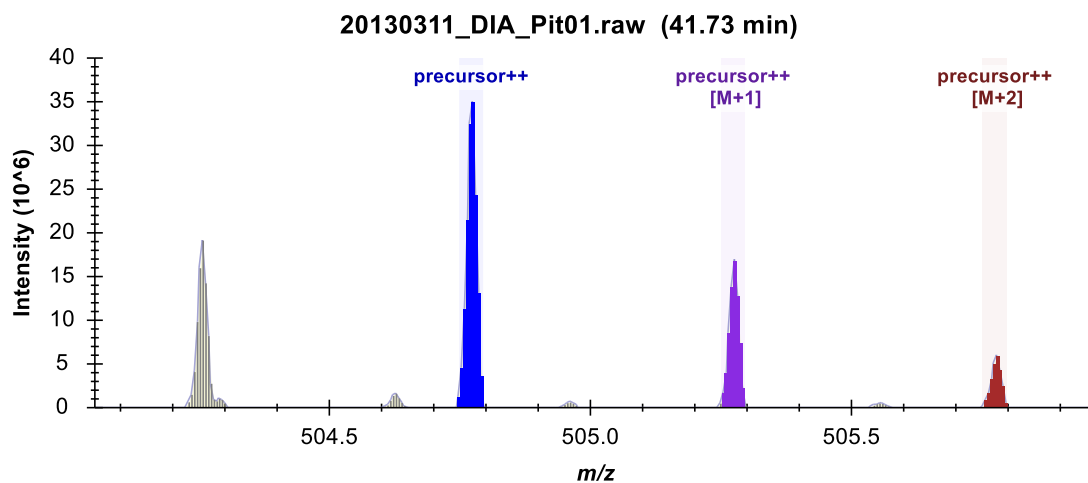
Using centroiding before chromatogram extraction can reduce the noise you see outside your targeted peptide peak, but it cannot remove signal contribution from interference within your peak of interest.

Conclusion

In this tutorial, you defined a DIA acquisition scheme for use in producing a DIA instrument method or simply data analysis, as in this tutorial. You built a spectral library from DDA search results, set up retention time restriction based on the measured RTs in the library, defined a set of transitions to be measured based on proteins of interest and matching spectra from the DDA runs, imported DIA runs, and analyzed the resulting data quality. Ultimately, as in all Skyline documents, you can produce peak areas and statistical information on the peptides of interest. Several other tutorials (Targeted Method Editing, Existing and Quantitative Experiments, iRT Retention Time Prediction, Advanced Peak Picking Models, Importing Assay Libraries and Panorama Chromatogram Libraries) have been mentioned in this tutorial for more advanced study.

This workflow should enable you to analyze any DIA data set in Skyline, provided it is accompanied by DDA runs acquired before, after or among your DIA runs. This is currently the easiest way to get started with DIA data analysis. However, if you or someone else has already built libraries containing fragment ion relative abundance information and normalized retention times for your proteins of interest, then you can use these libraries instead of acquiring and extracting information from experiment specific DDA runs. Building and using such libraries for DIA experiments is a more advanced topic, which will require a separate tutorial, or you piecing together the information found in other Skyline and Panorama tutorials, which have been mentioned here. Having completed this tutorial, however, you should now understand at least one complete workflow for quantitative proteomics analysis using DIA and Skyline.

Bibliography

1. Venable, J. D., Dong, M.-Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **1,** 39–45 (2004).
2. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics MCP* **11,** O111.016717 (2012).
3. Egertson, J. D. *et al.* Multiplexed MS/MS for improved data-independent acquisition. *Nat. Methods* **10,** 744–746 (2013).
4. Krokhin, O. V. *et al.* An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS. *Mol. Cell. Proteomics MCP* **3,** 908–919 (2004).

5. Escher, C. *et al.* Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics Accept.* (2012).

6. Reiter, L. *et al.* mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat. Methods* **8,** 430–435 (2011).