

Skyline MS1 Full-Scan Filtering

The Skyline Targeted Proteomics Environment provides informative visual displays of the raw mass spectrometer data you import into your Skyline documents. These displays allow you to manipulate the data by performing tasks such as refining the peptides and transitions you are measuring, and tuning integration boundaries. Originally developed to analyze quantitative assays from selected reaction monitoring (SRM – also referred to as multiple reaction monitoring or MRM) mass spectrometry, Skyline has been expanded to extract time-intensity chromatograms from MS1 scans for use in peptide quantification experiments involving mass spec runs with data dependent MS/MS.

The Skyline MS1 Full-Scan Filtering¹ supports importing data sets from discovery-type proteomics experiments where mass spectrometers were operated in data dependent acquisition (DDA) mode. After importing the raw data, new and previously existing Skyline features facilitate quantifying peptide precursor MS1 signal across many replicate acquisitions.

This tutorial will cover the following areas critical to making effective use of Skyline MS1 Filtering:

- Setting up a Skyline document for MS1 filtering
- Importing raw data and using retention time information from spectral libraries to direct peak picking during MS1 filtering
- Further processing of MS1 filtered peptides to obtain quantitative information across acquisition replicates

Skyline aims to provide a vendor-neutral platform for targeted proteomics investigation. It can import raw data for MS1 filtering from the instrument vendors AB Sciex, Agilent, Bruker, Thermo-Scientific and Waters. The ability to import data across various instrument platforms greatly facilitates cross-instrument comparisons and large multi-site studies.

Getting Started

To start this tutorial, download the following ZIP file:

https://skyline.gs.washington.edu/tutorials/MS1Filtering_2.zip

Extract the files in it to a folder on your computer, like:

C:\Users\brendanx\Documents

This will create a new folder:

C:\Users\brendanx\Documents\MS1Filtering

It will contain all the files necessary for this tutorial. Now start Skyline, and you will be presented with a new empty document.

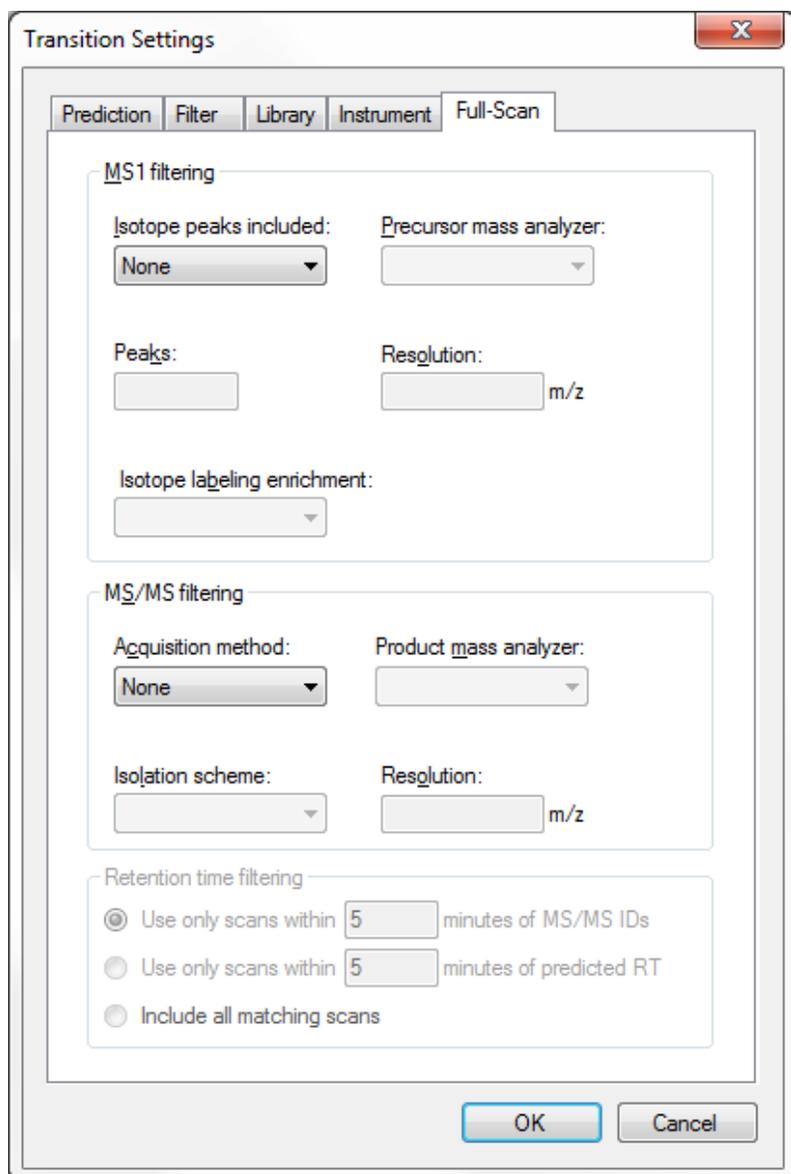
Importing a Data Dependent Peptide Search into a Skyline Document

The easiest way to get a peptide search based on data dependent acquisition MS/MS into a Skyline document is to use the **Import Peptide Search** wizard.

If you have been working with full-scan data in Skyline before, do the following to reset your full-scan settings for this tutorial:

- On the **Settings** menu, click **Transition Settings**.
- Click the **Full-Scan** tab.
- In the **M1 filtering** section, set the **Isotope peaks included** field to 'None'.
- In the **MS/MS filtering** section, set the **Acquisition method** field to 'None'.

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



The screenshot shows the 'Transition Settings' dialog box with the 'Full-Scan' tab selected. The dialog is divided into three main sections: MS1 filtering, MS/MS filtering, and Retention time filtering. In the MS1 filtering section, 'Isotope peaks included' is set to 'None'. In the MS/MS filtering section, 'Acquisition method' is set to 'None'. In the Retention time filtering section, the first radio button is selected, indicating 'Use only scans within 5 minutes of MS/MS IDs'.

Section	Field	Value
MS1 filtering	Isotope peaks included:	None
	Precursor mass analyzer:	
	Peaks:	
	Resolution:	
MS/MS filtering	Acquisition method:	None
	Product mass analyzer:	
	Isolation scheme:	
	Resolution:	
Retention time filtering	Use only scans within	5 minutes of MS/MS IDs
	Use only scans within	5 minutes of predicted RT
	Include all matching scans	

If you have been working with isotope labeled peptides, you may also want to go to the **Peptide Settings** – **Modifications** tab and turn off all isotope modifications.

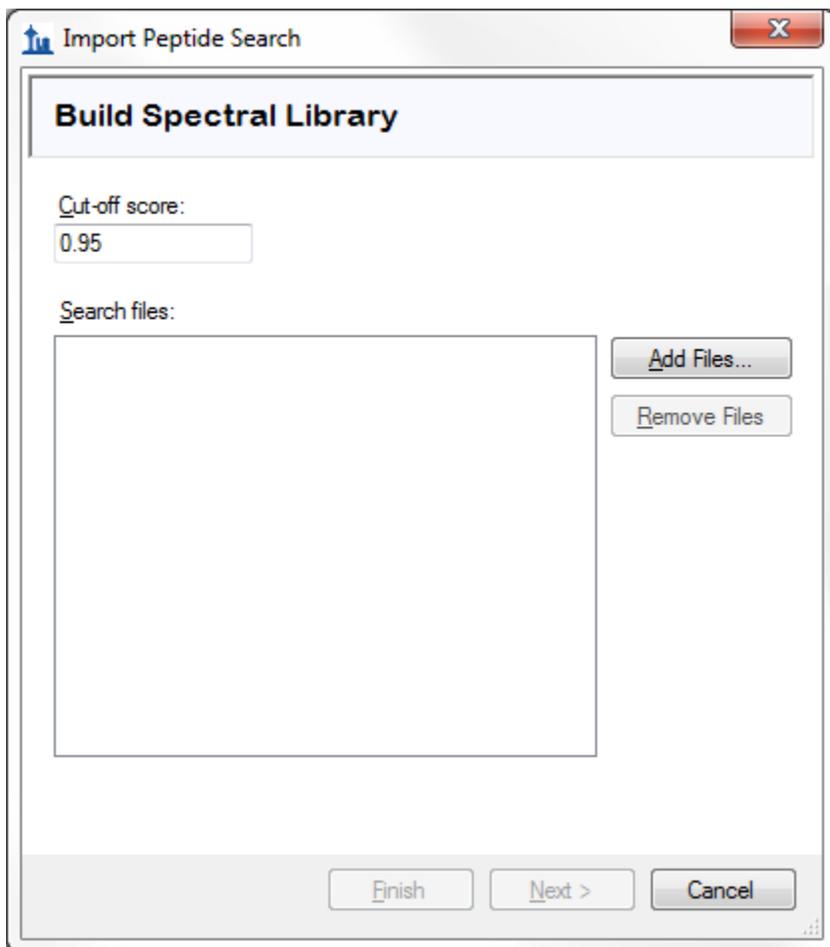
Next, do the following to save your new document:

- Click the **Save** button on the toolbar (Ctrl-S).
- Navigate to the MS1Filtering folder you created for this tutorial.
- In the **File name** field, enter 'Ms1FilterTutorial.sky'.
- Click the **Save** button.

Now, initiate the **Import Peptide Search** wizard as follows:

- On the **File** menu, choose **Import** and click **Peptide Search**.

Skyline should present a form that looks like:



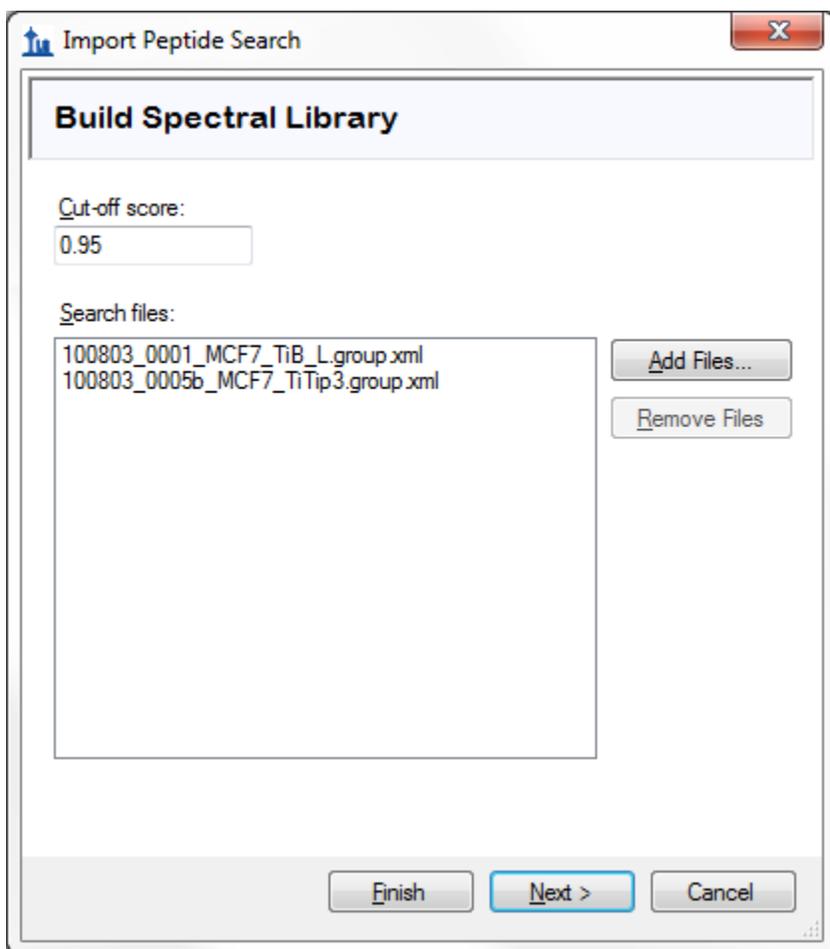
Next, you will use this wizard to create a spectral library from the output of one of the many peptide search engines supported by Skyline. For complete details on the supported search pipelines, consult the "[Targeted Method Editing](#)" tutorial. Note that the files you will use in this tutorial have been

reduced to the minimum information necessary to complete the tutorial, in the interest of keeping the ZIP file download a reasonable size.

Add the included search results to your library by doing the following:

- Click the **Add Files** button.
- Select both of the .group.xml files in the MS1Filtering folder you created for this tutorial.
- Click the **Open** button.

The wizard form should now look like:



- Click the **Next** button.

Skyline will build a new spectral library associated with your Ms1FilterTutorial.sky document, showing progress as it does so. In the MS1Filtering folder where Skyline saved your new spectral library, you should now find two new files:

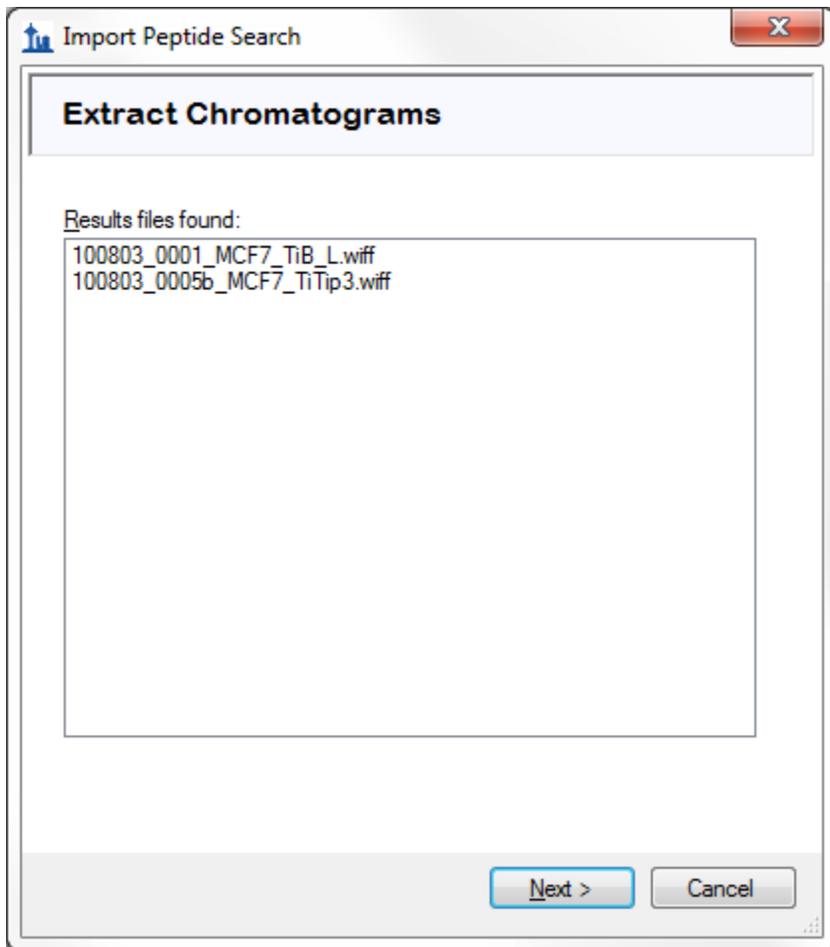
- The non-redundant library “MS1FilteringTutorial.blib” containing best matching spectra.
- The redundant library called “MS1FilteringTutorial.redundant.blib” containing all matching spectra.

 Ms1FilterTutorial.blib	10/15/2013 10:26 ...	BLIB File	803 KB
 Ms1FilterTutorial.redundant.blib	10/15/2013 10:26 ...	BLIB File	1,160 KB

You may also find a file 'MS1FilterTutorial.slc', which is a 'Skyline Library Cache' file to improve library load time. It can be deleted, and Skyline will simply recreate it as needed.

If you have used Skyline in the past to build spectral libraries, you may be used to naming them as you please, and putting them wherever you like. In this case, Skyline creates the library as a document-specific spectral cache, very similar to the way it stores chromatograms specific to your document. You can add more search results later and you can remove search results, just as you may be used to doing with chromatogram data.

When library building has completed, Skyline will present the following page in the wizard:

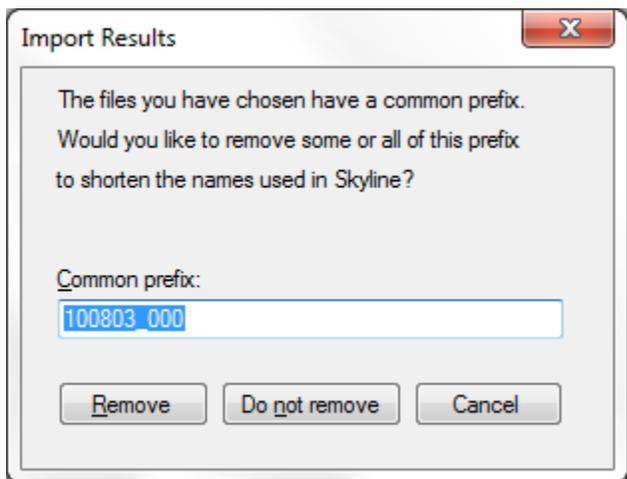


In this case, Skyline has found original WIFF data files matching the spectrum source files used to build your library, and the library appears to have the retention time information Skyline will need to locate identified MS/MS spectra on the chromatograms it will extract. If Skyline could not find suitable data files for chromatogram extraction, it would ask you to locate them. If the library build had failed to find retention time information in the imported peptide search files, Skyline would notify you of that.

Consult the “Verifying Library Retention Time Information” section below for more information on troubleshooting issues Skyline may have with determining spectrum source files or retention times from peptide search results files. To continue this tutorial:

- Click the **Next** button.

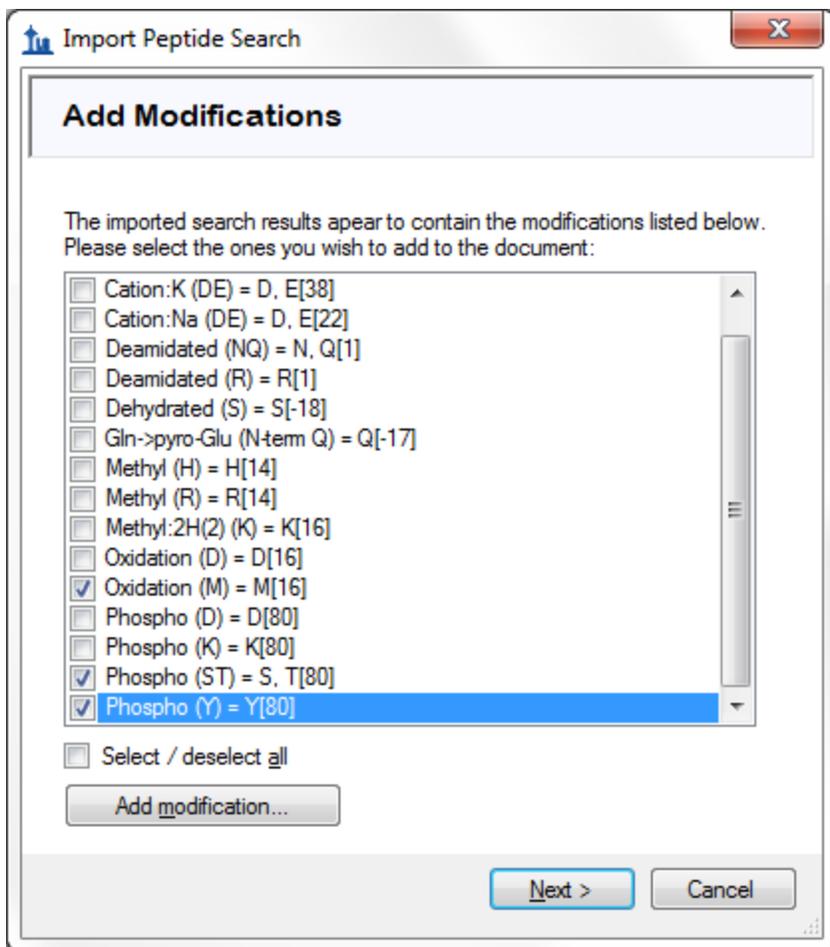
A form should appear asking you how to handle the prefix shared by the two WIFF files:



- Click the **Remove** button.

The wizard will advance to a page to the **Add Modifications** page, where it lists all of the amino acid modifications it has found in the search results that do not already exist in the document. Where possible it will suggest specific modifications from the Unimod site that match the amino acid, mass combinations found in the search.

For this tutorial, you only need the ‘Phospho (ST)’, ‘Phospho (Y)’ and ‘Oxidation (M)’ modifications. Check them in the list, and the wizard should look like this:



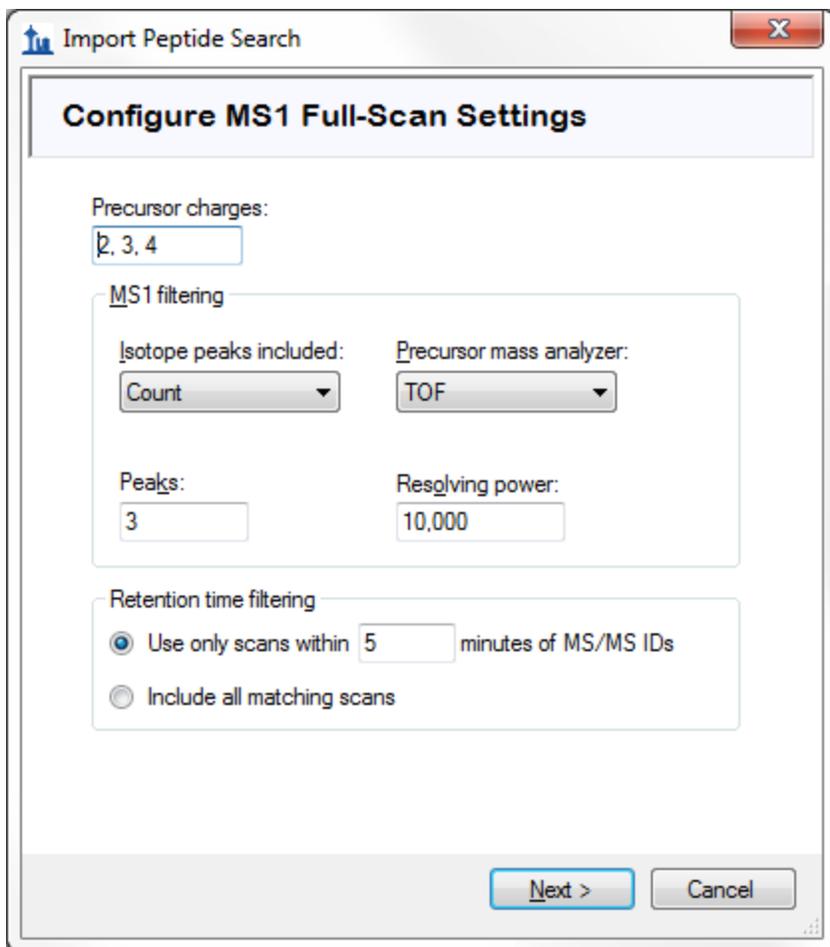
It is possible that your document will already have one or more of these modifications defined (e.g. Oxidation (M)), in which case the list presented may look different.

- Click the **Next** button.

The wizard will advance to the **Configure MS1 Full-Scan Settings** page, where you should do the following:

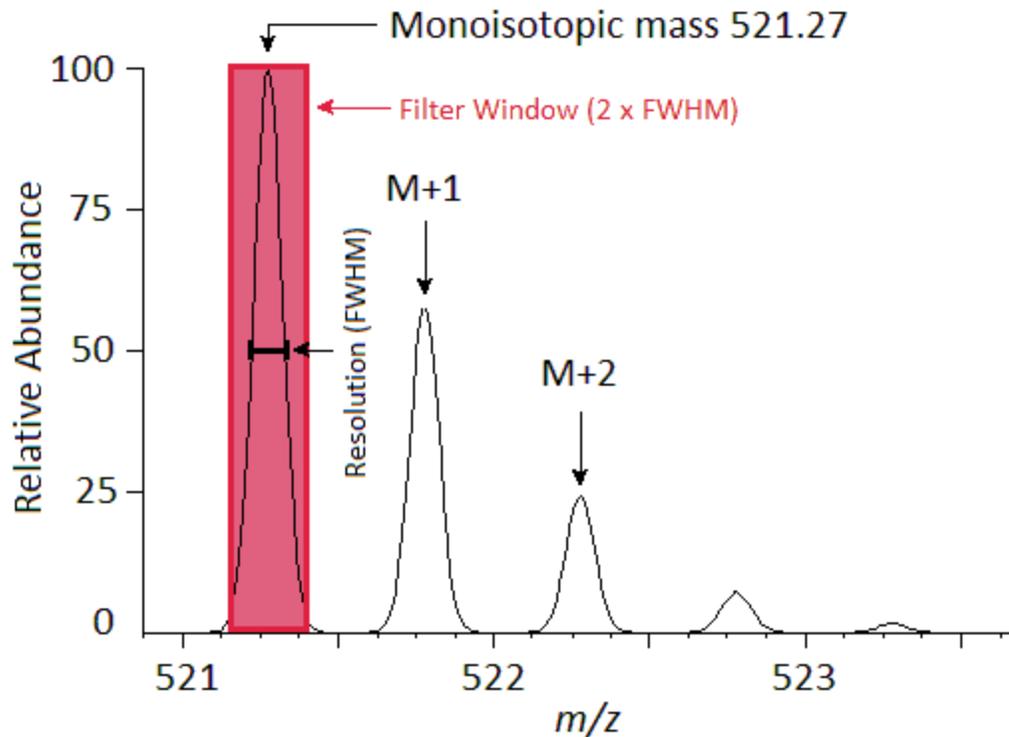
- In the field **Precursor charges**, enter '2, 3, 4'.

All of the other fields in this page should default to values you can use for this tutorial, leaving the wizard looking like:

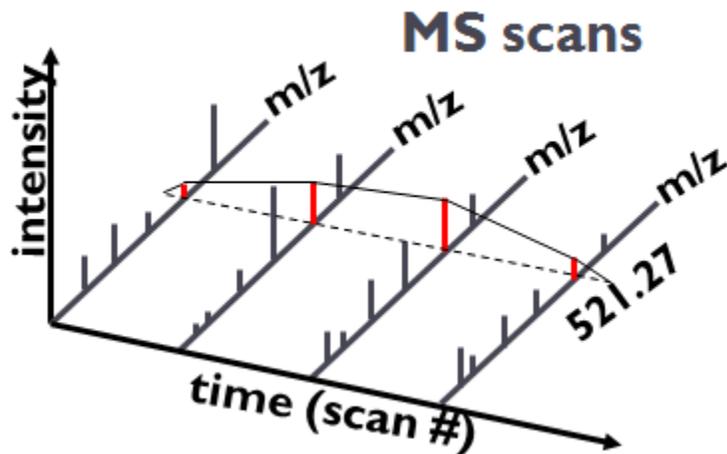


In the **MS1 filtering** section, you will see the following default settings.

1. In the **Isotope peaks included** dropdown list, the value 'Count' is chosen.
2. The **Peaks** field contains '3' to have Skyline filter the first 3 isotope peaks (M, M+1 and M+2) from this high resolution data.
3. In the **Precursor mass analyzer** dropdown list, the value 'TOF' is chosen (as the data were acquired on a QSTAR Elite).
4. In the **Resolving power** field, the default value is '10,000'. This field defines the width of the MS1 Filtering window around each precursor m/z to be filtered. Skyline uses this value to predict full width at half maximum (FWHM) for a peak in the m/z dimension, and uses 2 times that FWHM value as the filtering window as shown below. (Note: For other datasets and experiments, the resolution setting can be adjusted depending on instrument capabilities).



A series of these intensities extracted over time make up the chromatograms you will see in Skyline:



In the **Retention time filtering** section, note the choice to **Use only scans within [5] minutes of MS/MS IDs**. This means for a peptide with only 1 ID, Skyline will extract a 10 minute chromatogram around that ID. For a set of IDs over a 3 minute range, Skyline will extract a 13 minute chromatogram with 5 minutes added on either side of the IDs. When a run lacks any IDs for a particular peptide, Skyline will use retention time alignment of IDs in other runs to determine the time range over which to extract a chromatogram.

- Click the **Next** button.

This should bring you to the **Import FASTA** page in the wizard. You could import a FASTA file of all human SwissProt protein entries to get a comprehensive identified peptide list (this MS experiment involves a human MCF7 breast cancer cell line sample and subsequent phosphopeptide enrichment), however, for file size reasons, you will import a much smaller FASTA file containing only 12 human proteins by performing the following steps:

- In the **Max missed cleavages** dropdown list, pick '2'.
- Click the **Browse** button.
- Select the '12_proteins.062011.fasta' file in the MS1Filtering folder you created for this tutorial.
- Click the **Open** button in the **Open FASTA** form.

The wizard should now look like:

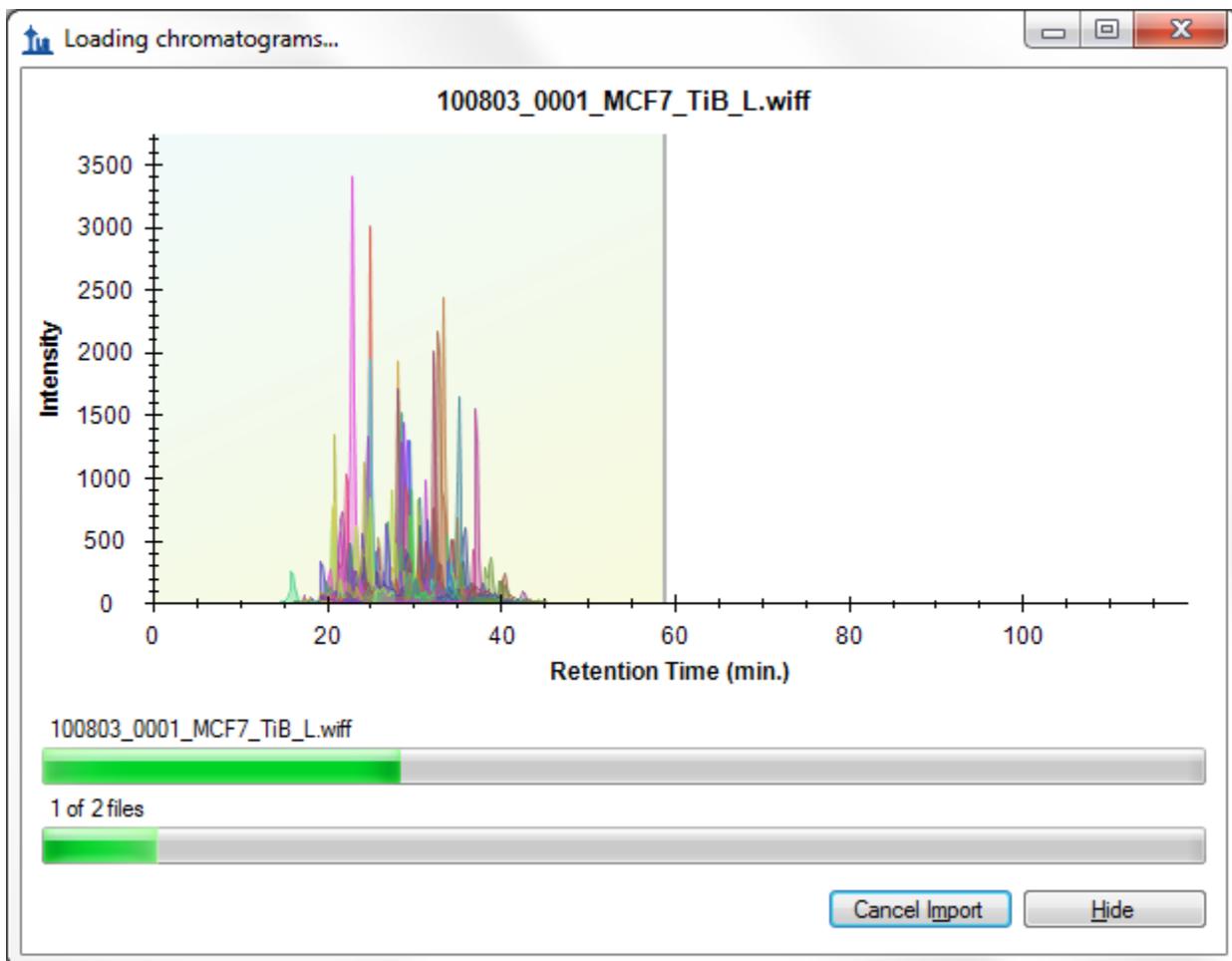
The screenshot shows a software window titled "Import Peptide Search" with a sub-header "Import FASTA". It features two dropdown menus: "Enzyme:" set to "Trypsin [KR | P]" and "Max missed cleavages:" set to "2". Below these is a text area containing a FASTA file preview with the following text:


```
>sp|Q09666|AHNK_HUMAN Neuroblast differentiation-as
MEKEETTRELLLPNWQGSGLTIAQRDDGVFVQEV
YFDNLQSGEVTQLLNTMGHHTVGLKLRKGDSPSP
YQRIYTTKIKPRLKSEDEGVEGDLGETQSRITVTRRVT/
KIKIPRHELTEISNVDVETQSGKTVIRLPSSGGAASPTG/
HSLQVTMPGIKVGSGVNVNAKGLDLGGRGGVQVP/
HGKIKFPTMKVPKFGVSTGREGQTPKAGLRVSAPEVS\
ANIEGLEGLKKGPIITGPSLEGDLGLKGAQPQGHIGVD/
DVQGGPSKLNVPKMKVPKFSVSGAKGEETGIDVTLPTC
GKMKGTKVKTPEMIIQKPKISMQDVDLSLGSPLKGDIF
VDIETPNLEGLTGPRLGSPSGKTGTCRISMSEVDLNV
PEVDVRGPKVDVSAPDVEAHGPEWNLKMPKMKMPTF
GPKVNVEAPDVNLEGLGGKLGKPDVKLPDMSVKTPKIS
```

 To the right of the text area are "Browse..." and "Clear" buttons. At the bottom of the window are "Finish" and "Cancel" buttons.

- Click the **Finish** button.

Skyline will add targets for all of the peptides from the FASTA file with matching spectra in the peptide search results which you imported, and then begin importing the two WIFF files and extracting chromatograms from them. You should see a progress graph like this:



Once the import has completed, first have an in-depth look at the spectral library you created before inspecting the chromatogram data.

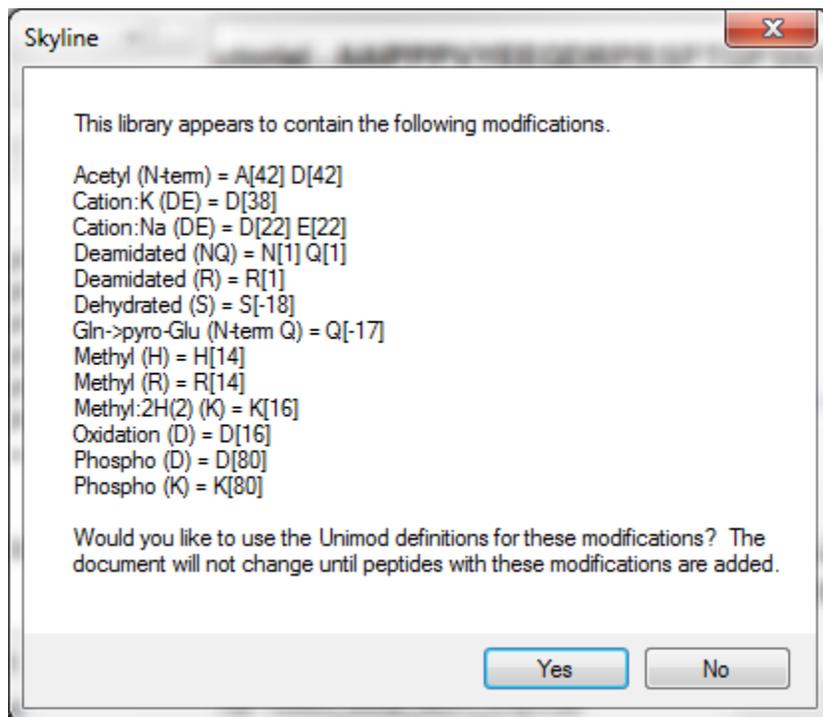
Verifying Library Retention Time Information

Anytime you are building a spectral library for MS1 Filtering from the results of a peptide search pipeline, with which you have not already done this, you should be sure that the resulting library contains the necessary retention time information to support the Skyline features explained below. One benefit of using the **Import Peptide Search** wizard is that it will inform you early when your library is lacking necessary information.

To verify that the library you just created contains the retention time information for MS1 Filtering peak picking and peak annotation, perform the following steps:

- On the **View** menu, click **Spectral Libraries**.

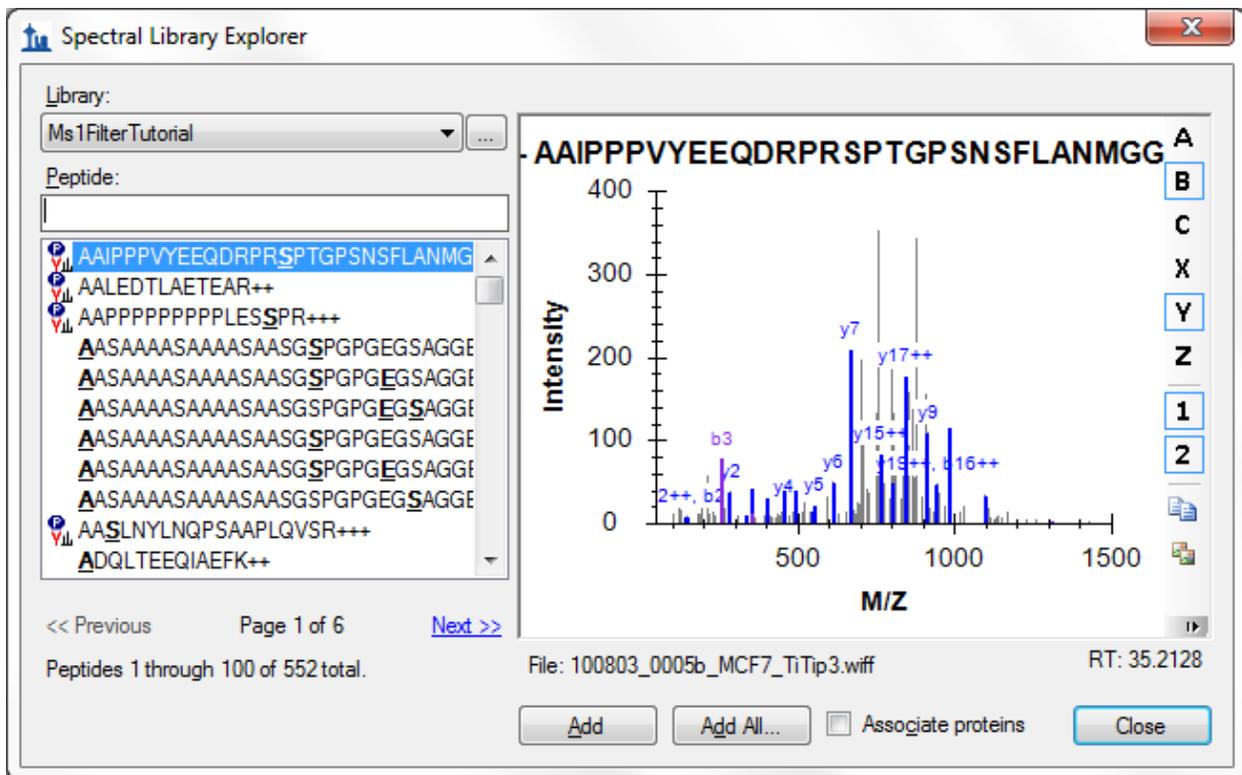
Skyline will again offer to use the modifications it detects in the library which you chose not to add to your document during the **Import Peptide Search** wizard.



Choosing to use them now in the **Spectral Library Explorer** will not add them to the current document unless you use the **Spectral Library Explorer** to add peptides that use these modifications to your document. However, these modifications are not important to this tutorial. You can continue without them by doing the following:

- Click the **No** button.

The **Spectral Library Explorer** will appear, looking something like this:



In the peptide list, the peptides without an icon to the left of the sequence text are ones that contain any of the modifications you just chose not to use.

Below the spectrum graph, you can see the text “File: 100803_005b_MCF7_TiTip3.wiff” and “RT: 35.2128”. The “RT” value tells you that retention time information is present, and the “File” value tells you that it is correctly associated with the files you have imported into Skyline. The “File” values need not be exact matches of the files you import. Skyline recognizes that many peptide search pipelines involve converting raw instrument data to formats like mzXML, mzML, MGF, MS2, etc. In general, Skyline will look for a base-name match, where ‘basename.mgf’ matches successfully with ‘basename.wiff’. Due to specific instances of pipelines that require greater flexibility, this matching is also done insensitive of case, so that ‘BASENAME.mzML’ will match ‘Basename.RAW’, and finally handling for multiple-dot extensions is included, so that ‘basename.c.mzXML’ matches ‘basename.raw’. If, however, you saw something like ‘F011852.dat’ or some other search output file that does not share a base-name with the data you intend to import into Skyline, then you will need to review your search pipeline, and possibly work with the Skyline team to fix the problem. For Mascot .dat files specifically, you are encourage to consult the [“ID Annotations Missing with Mascot Search Results”](#) page on the Skyline web site. For other issues, you are encouraged to post to the Skyline support board (on the **Help** menu, click **Support**) for help with issues of this type.

Now press the Down-Arrow key to select other peptides, and you will see that the “File” and “RT” values change. When you are done inspecting MS/MS spectra, their source files and retention times, return to the document you have created by doing the following:

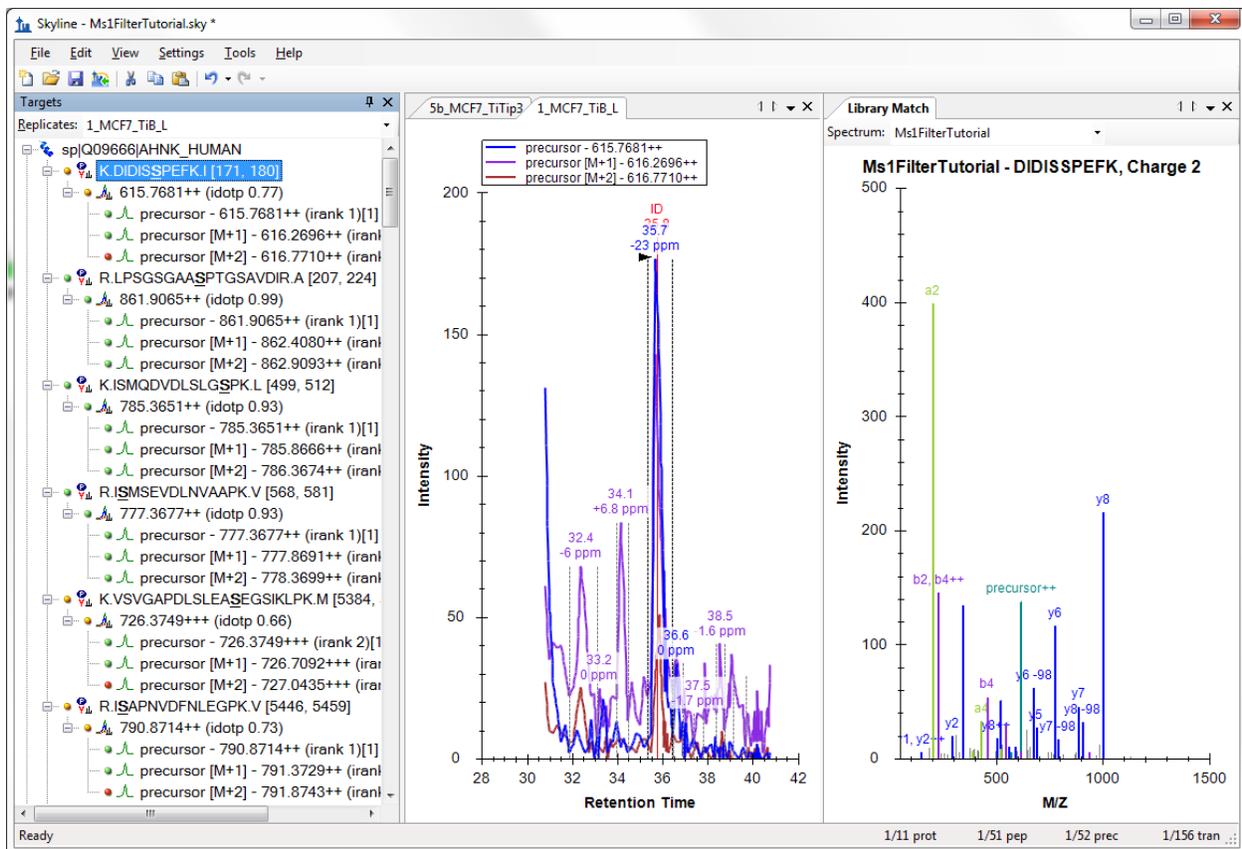
- Click the **Close** button in the **Spectral Library Explorer**.

Peptide Targets, Spectra and Chromatograms

You should see 51 peptides (counted in the status bar) in the Skyline Targets view.

- Click on the sequence of the first phosphopeptide **K.DIDISSPEFK.I** and the MS/MS spectrum will appear. (Note that the bold, underlined residue “**S**” in the peptide sequence indicates a serine phosphorylation).
- If you do not see the MS/MS spectrum, on the **View** menu, click **Library Match**.
- If you do not see as many annotated peaks as in the image below, on the **View** menu, choose **Ion Types** and check **A, B, Y** and **Precursor**.
- If you do not see the entire chromatogram for the peptide, on the **View** menu, choose **Auto-Zoom** and click **None** (Shift-F11).
- On the **Edit** menu, choose **Expand All** and click **Precursors**.

Your Skyline document should look like the following:



The document is now fully configured for MS1 Filtering with two DDA runs imported. You will see that the chromatogram in this view is approximately 10 minutes in length (31 to 41 minutes), due to the **Use**

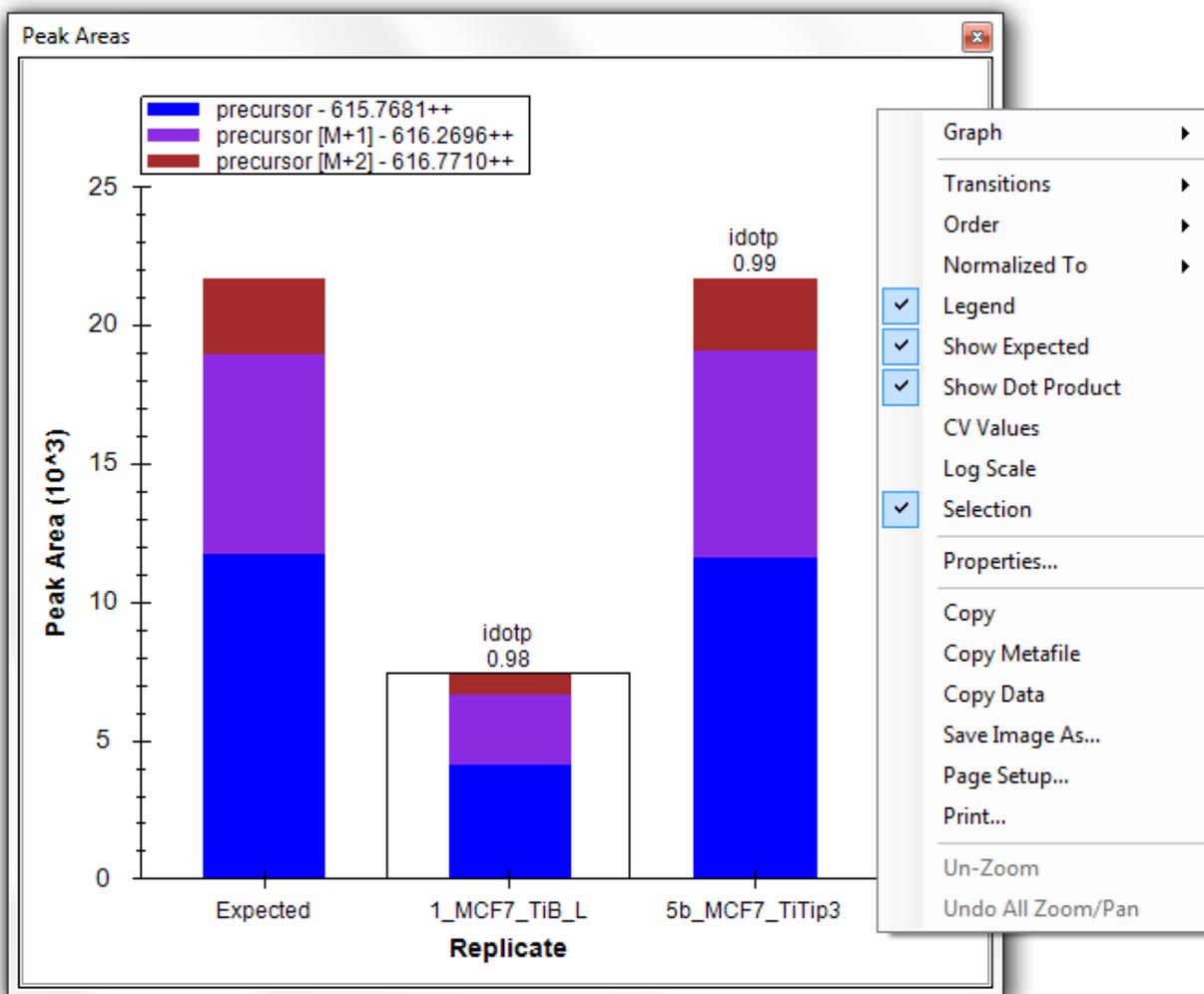
only scans within [5] minutes of MS/MS IDs setting chosen in the import wizard. Note that when the Skyline document is set up for MS1 Filtering, in the place where you would see product ion transitions (e.g. y-ions) for triple quadrupole SRM experiments, you will now see the different precursor isotope peaks, such as for peptide DIDISSPEFK: precursor - 615.7681++, precursor [M+1] - 616.2696++, and precursor [M+2] - 616.7710++.

To configure a few other features that will be helpful in general, especially to visualize certain MS1 Filtering data, perform the following steps:

- On the **Settings** menu, make sure **Integrate All** is checked.

This tells Skyline to calculate integrated areas for all chromatograms in a peak group (here precursor ions M, M+1 and M+2), regardless of whether peaks appear to be co-eluting with the largest peak.

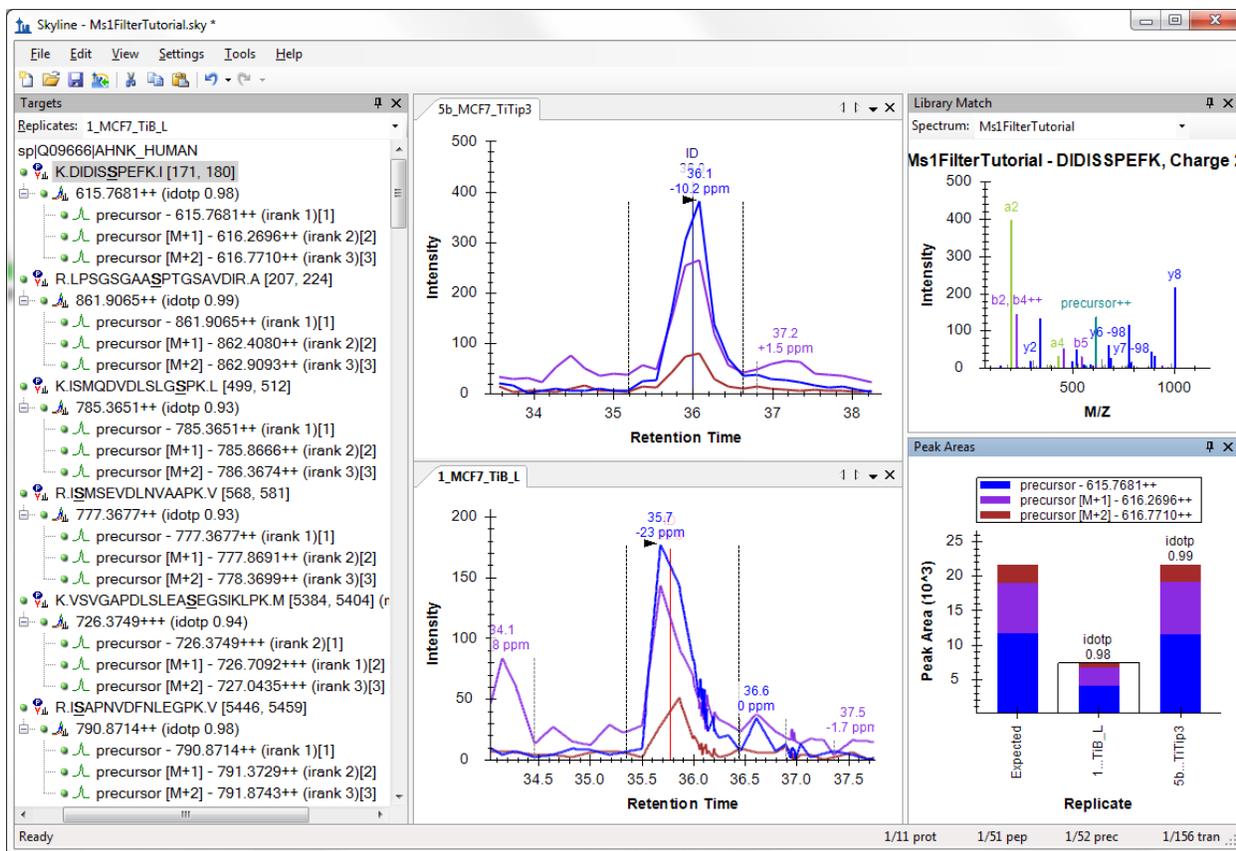
- On the **View** menu, choose **Peak Areas** and click **Replicate Comparison**.
- In the **Peak Areas** window, right-click, choose **Normalize To** and click **None**.
- In the **Peak Areas** window, right-click and check **Show Expected** and **Show Dot Product** (these two features will be explained below).



You can dock the **Peak Areas** window to your desired location by doing the following:

- Click and hold down the left mouse button, and then drag it where you want to place it, perhaps the right edge of the Skyline window.
- Then click-and-drag the **Library Match** view, and dock it above the **Peak Areas** view.
- On the **View** menu, choose **Auto-Zoom** and click **Best Peak** (F11).
- On the **View** menu, choose **Arrange Graphs** and click **Tiled** (Ctrl-T).

Your Skyline file should look something like:



If your **Library Match** view is not appropriately docked, move it above the peak area replicate view, as shown above.

The chromatogram views show the MS1 extracted ion chromatogram for all precursor isotope ions M (blue), M+1 (purple), M+2 (brown). Below the retention time annotation of the chosen peak, which may be familiar if you have used Skyline for SRM, you will see a new mass-error annotation, which is a weighted mean of the mass error in all the integrated points across the annotated chromatogram (in this case for M or the blue line). If you do not see the mass error, right-click on the chromatogram view and click **Mass Error**. The mass accuracy is not what you might expect of a modern high-resolution instrument, but as noted earlier this data is from an older QSTAR Elite.

You will also see vertical lines on the extracted ion chromatogram graphs with **ID** annotations on top, though the one for 1_MCF7_TiB_L is behind the peak annotation. The **ID** stands for “**identification**” and indicates the retention time of a sampled MS/MS spectrum with confident identification for this particular peptide. The red line indicates that this is the spectrum currently showing in the **Library Match** view. If you click on the **ID** annotation in the upper graph, the **Library Match** view will show you the identified spectrum from the 5b_MCF7_TiTip3 replicate, now stored in the library you created earlier during import. Also, the replicate name and retention time (36 min) will be selected in the **Spectrum** drop-down list at the top of the window **Library Match** view (instead of the “best” spectrum from the non-redundant library, which was selected before you clicked on the **ID** annotation). You can

switch back and forth between the two collected spectra by either clicking on the ID annotations or using the **Spectrum** drop-down list and see that they are quite similar.

Before reviewing some of the other 51 peptides in this document, first do the following:

- On the **Edit** menu, choose **Collapse All** and click **Peptides** (Ctrl-Shift-D).

Next, make sure the focus is in the **Targets** view (peptides tree) and start using the down-arrow key on your keyboard to select each peptide one at a time. For the first three phosphor peptides, you will see each identified once in each replicate, with a few prominent neutral loss ions annotated “-98” ($-H_3PO_4$) in the **Library Match** spectrum graphs.

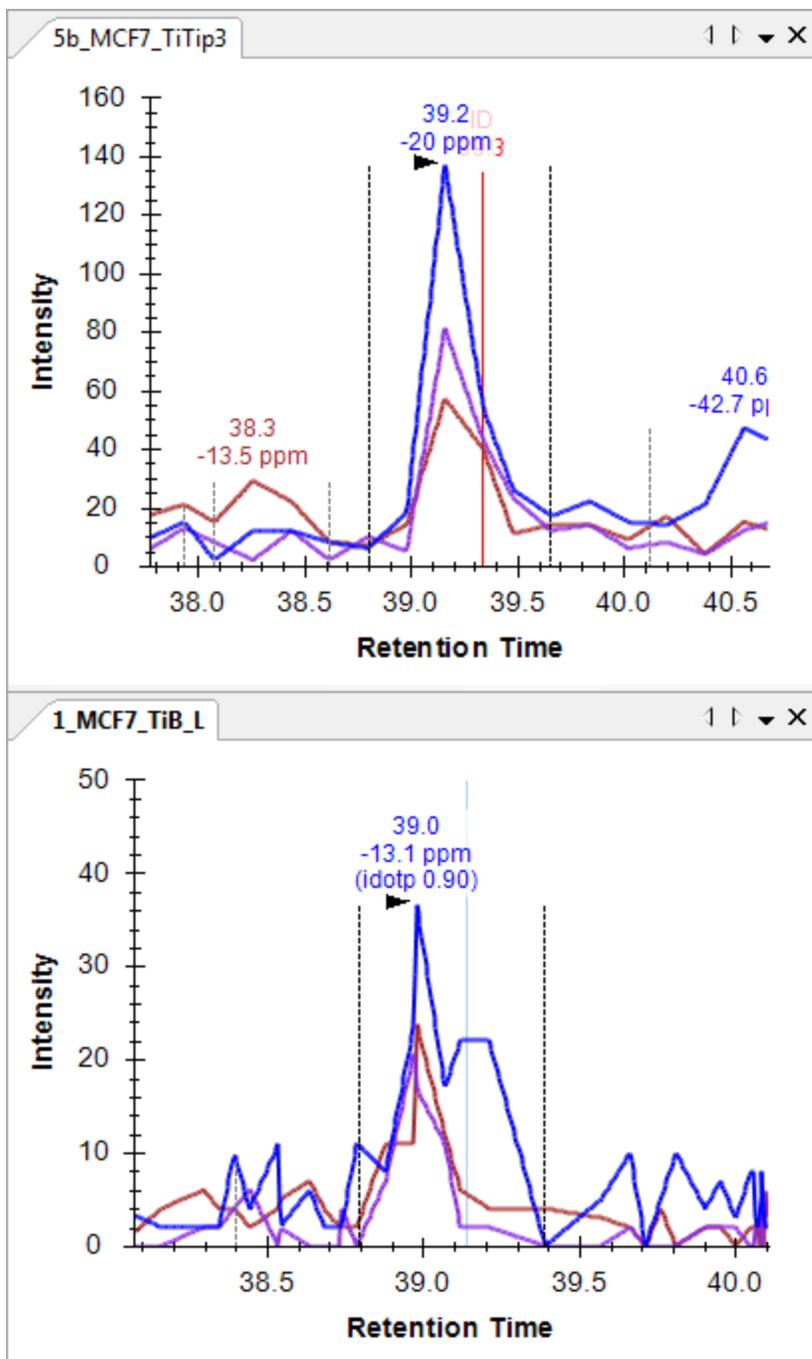
For the fourth peptide **I**SMSEVDLNVAAPK, you will notice that only the 5b_MCF7_TiTip3 replicate has an ID annotation. The peak in the 1_MCF7_TiB_L replicate was chosen using retention time alignment of the ID in 5b_MCF7_TiTip3. To see the aligned ID, do the following:

- Right-click a chromatogram graph, choose **Peptide ID Times** and click **Aligned** (if it is not checked).

You should see a light blue line appear in the 1_MCF7_TiB_L replicate inside the peak integration boundaries. Still, the peak in this case, is likely to the left of the integration boundaries. To correct this:

- Click below the x-axis at about 38.8 minutes and drag to about 39.4 minutes.

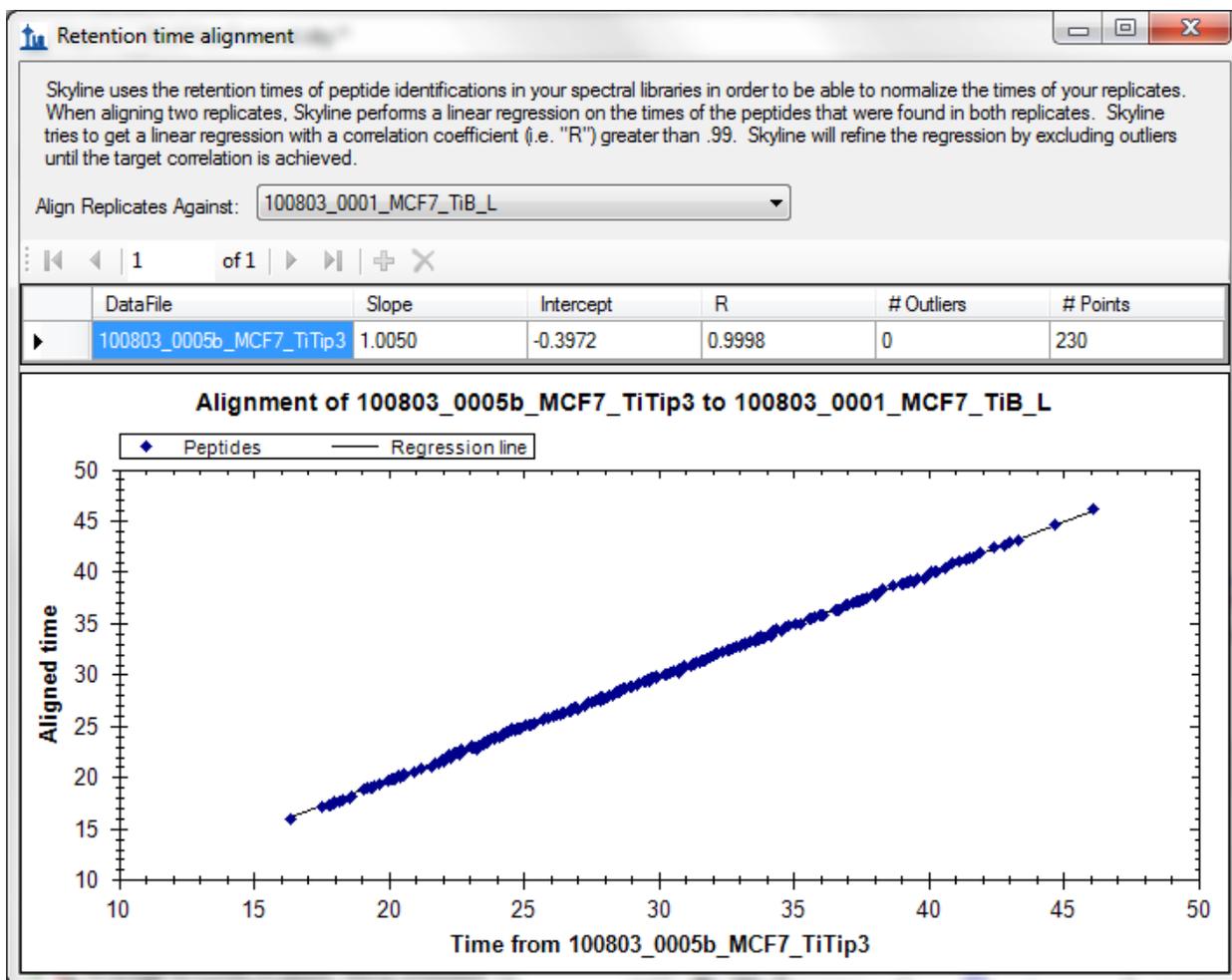
The chromatogram graphs should look something like:



To gain some insight into how the retention time alignment works, do the following:

- On the **View** menu, choose **Retention Times** and click **Alignment**.

Skyline will present a window that looks like this:



This window shows you the points for a linear regression used to align times between runs. Currently, Skyline calculates a linear regression like this between every spectrum source file in your spectral library, and every other spectrum source file. When more than 2 runs are present, you will see a row for every run other than the one selected in the **Align replicates against** dropdown list. The resulting linear equations can be used to map MS/MS ID times between runs for improved peak picking when no IDs are present in a run, as shown above. For more on using linear regression to map between retention time scales, see the [IRT Retention Time Prediction](#) tutorial.

In this case, you can see that retention time reproducibility between these two runs is quite good, with a slope of 1.005, intercept of -0.3972, a correlation coefficient (R) of 0.9998 and no outliers. As stated in the paragraph at the top of the form, when R is below 0.99, Skyline will discard outliers until it finds a set of peptides with R greater than 0.99, and use the resulting linear equation.

You may also notice that this regression is calculated with 230 points, while your document only contains 51 peptides, not all of which were identified in both runs. Remember, however, that the library you built contained 552 total peptides, many with modifications not used in this document. This seems to indicate that 230 out of the 552 were identified in both files. Skyline attempts to use all IDs present in both search result files for this regression. When multiple IDs are present in a single run, Skyline will

use the earliest ID retention time, since it is likely to be more stable than later times or even an average. For instance, we have seen cases where early eluting peptides are identified again during the gradient wash.

- Click the red X in the upper right corner of the **Retention Time Alignment** window to close it.

Reviewing the Data

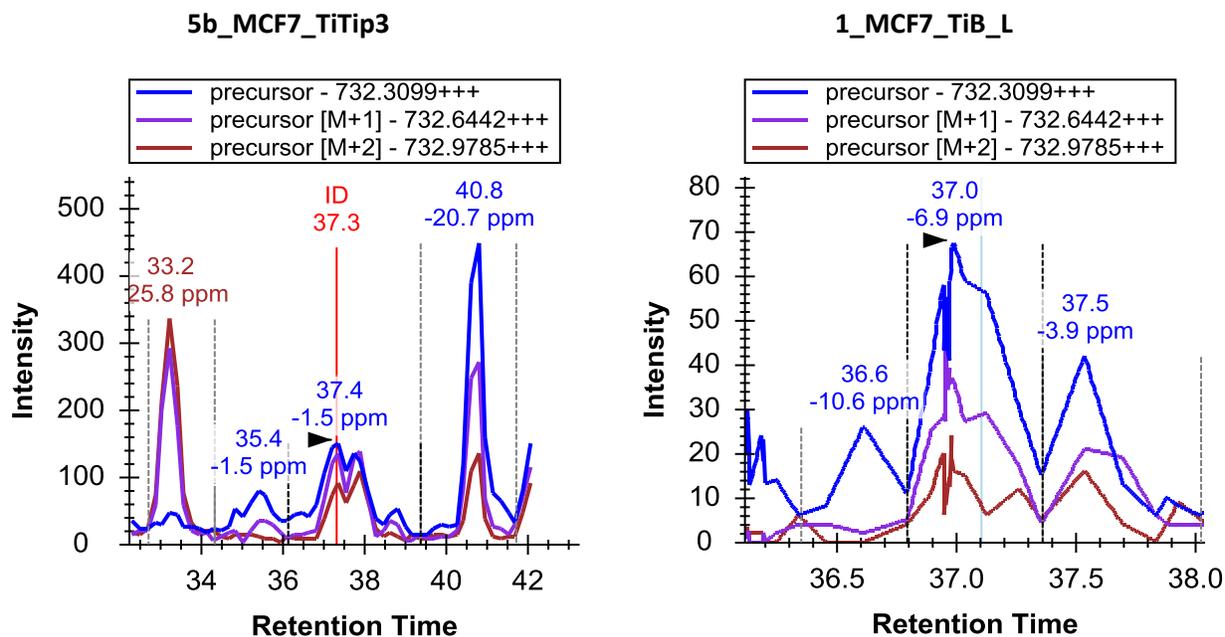
With this basic understanding and Skyline configured in this way, you can now quickly review all of the 51 peptides in this document. To do so now, just click on the **Targets** view and use the down-arrow key to select each peptide in turn. To know which peptide number out of the 51 is currently selected, you can look in the status bar at the bottom right of the Skyline window:



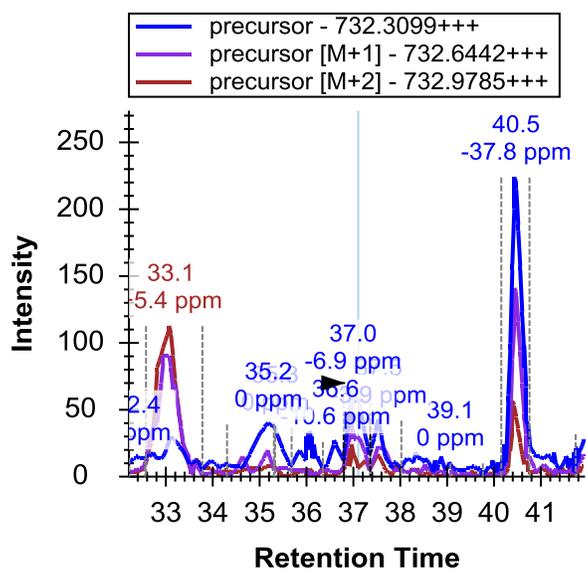
After the peptide ISMSEVDLNVAAPK, you will see 4 peptides with acceptable peak integration, though VSVGAPDLSLEASEGSIKLPK might be adjusted a little for 5b_MCF7_TiTip3. Some have an ID in both runs and some have an ID in only one run, and Skyline has used alignment to pick the correct peak.

Basic understanding of a chromatographic landscape

For the ninth peptide, SSKASLGSLEGEAEAEASSPK, the 1_MCF7_TiB_L run lacks an ID of its own, and the integration looks a little off:



Use the scroll-wheel on your mouse (scrolling back toward yourself) to zoom out of the 1_MCF7_TiB_L graph until you can see the same peaks you see in the 5b_MCF7_TiTip3 graph:



This is an extremely important aspect of working with chromatographic data: just as your target peptides can be expected to elute at highly similar times from run-to-run, so can other peptides. The two peaks on either side (33 and 40.5 minutes) of the target (37 minutes) are caused by two other peptides, which would have to be considered interference, if they coeluted with the target peptide. When they do not coelute, however, signal from other peptides can create a repetitive landscape that can help orient you to the retention time of your target, even at very low signal levels. This is especially true of methods with lower selectivity, as with MS1 filtering, since you can expect to see more peptides with signal in the chromatogram extraction ranges.

Now correct the integration range for 1_MC7_TiB_L by doing the following:

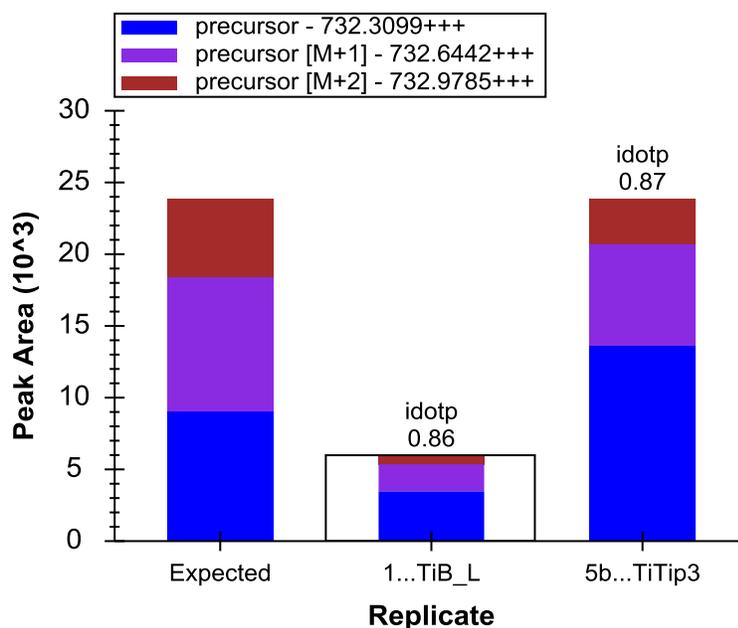
- Click and drag below the Retention Time axis between 36.5 and 38 minutes.

You will see in the **Peak Areas** graph that this improves the idotp (isotope dot-product) value for the peak from 0.87 to 0.9, and also very slightly the mass error from -6.9 to -6.5 ppm.

Before continuing to the remaining peptides, take a moment to consider the other two peaks captured by the extracted chromatograms. The peak at 40.5 minutes has very nice signal on all 3 precursor channels (M, M+1 and M+2), but you can see by the mass errors that it is also consistently lighter than expected (-20.7 and -37.8 ppm).

- Click on the labels above the peaks at 40.5 minutes.

This will cause Skyline to pick these peaks and show you in the **Peak Areas** graph that their idotp values are lower than the previously selected peaks (0.87 and 0.86 versus 0.96 and 0.90):



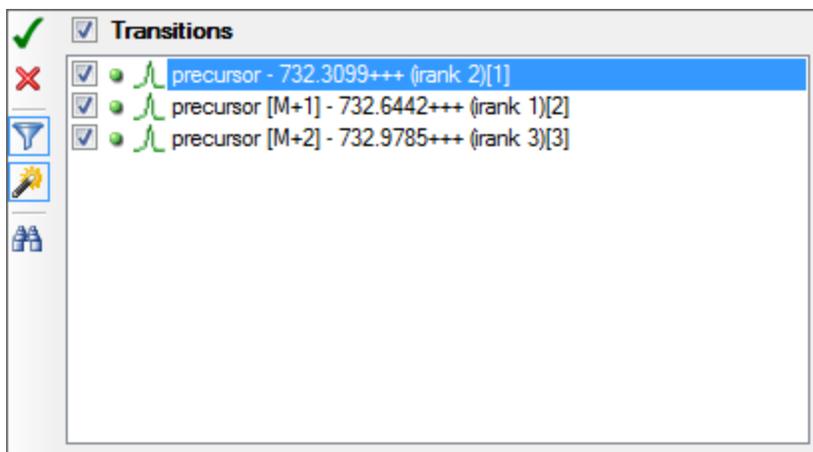
You can see from the distribution in the column marked Expected that this is because the M+2 and M+3 peaks are smaller than the predicted isotope distribution for the target peptide, which tells you that there are fewer Carbon atoms (and, therefore, less chance of getting a ^{13}C) in the peptide responsible for this peak than in the target peptide, for which there is also an ID at around 37 minutes.

Turning to the peak at 33 minutes, you will see that it lacks any real signal for the monoisotopic m/z of the target peptide, but that it has peaks of very similar intensity for M+1 and M+2, similar to the predicted isotope distribution for M and M+1 in the target peptide. The mass error for 5b_MCF7_TiTip3 is +25.8 ppm, and when fully integrated the mass error for 1_MC7_TiB_L is +5.6 ppm. Not quite as bad as the peak at 40.5 minutes, but the average error of +15.7 ppm is still much worse than the average error of -3.5 ppm for the peak at 37 minutes.

To gain a fuller understanding of the problem with the isotope distribution for the peak at 33 minutes, to the following:

- Click the + icon to the left of the peptide element in the **Targets** view to expand it.
- Hover the mouse cursor over the 732.3099+++ precursor element.
- Click the drop arrow that appears to the right of the precursor element.

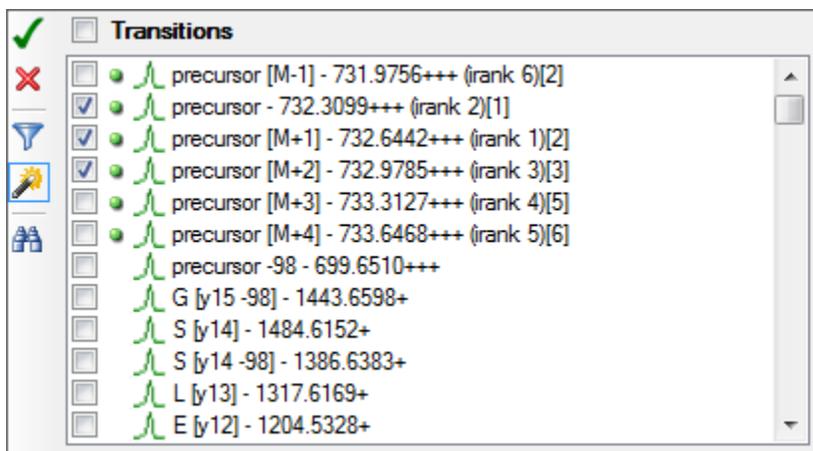
Skyline will show a pop-up that looks like:



If you see only these three precursor transitions:

- Click the funnel icon to remove transition filtering.

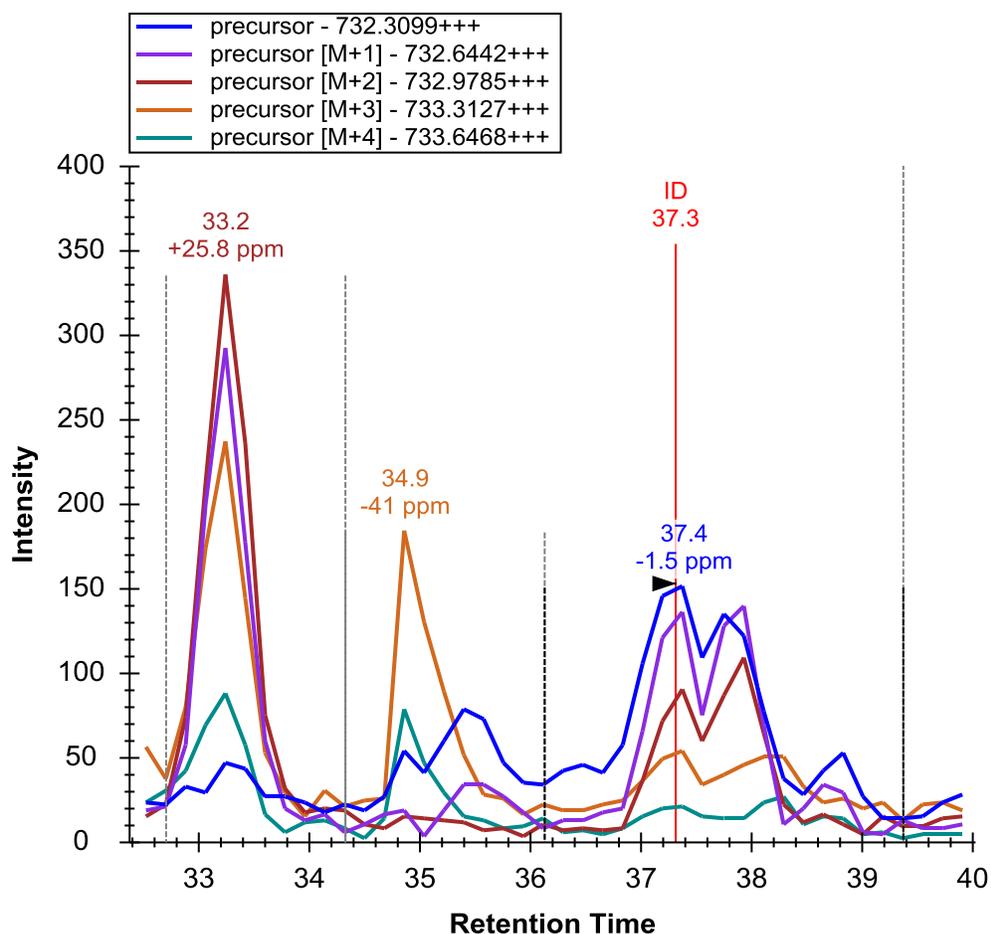
This causes Skyline to show all possible transitions for this peptide precursor:



The green dots indicate the transitions for which Skyline already has chromatogram data. Skyline automatically extracts chromatograms for all peaks in the isotope distribution which it predicts to have at least 1% of the entire distribution. Plus, it always extracts a chromatogram for M-1, since a correctly picked peak with no interference will generally have no signal at this m/z .

- Check the M+3 and M+4 transitions.
- Click the green check button in the upper left, or press Enter.

This will add chromatograms for M+3 and M+4 to the graph, and you will see more signal on them for the peak at 33 minutes than the identified peak at 37 minutes. The importance of reliable retention time reproducibility cannot be overstated in working with chromatogram data of any kind in Skyline.

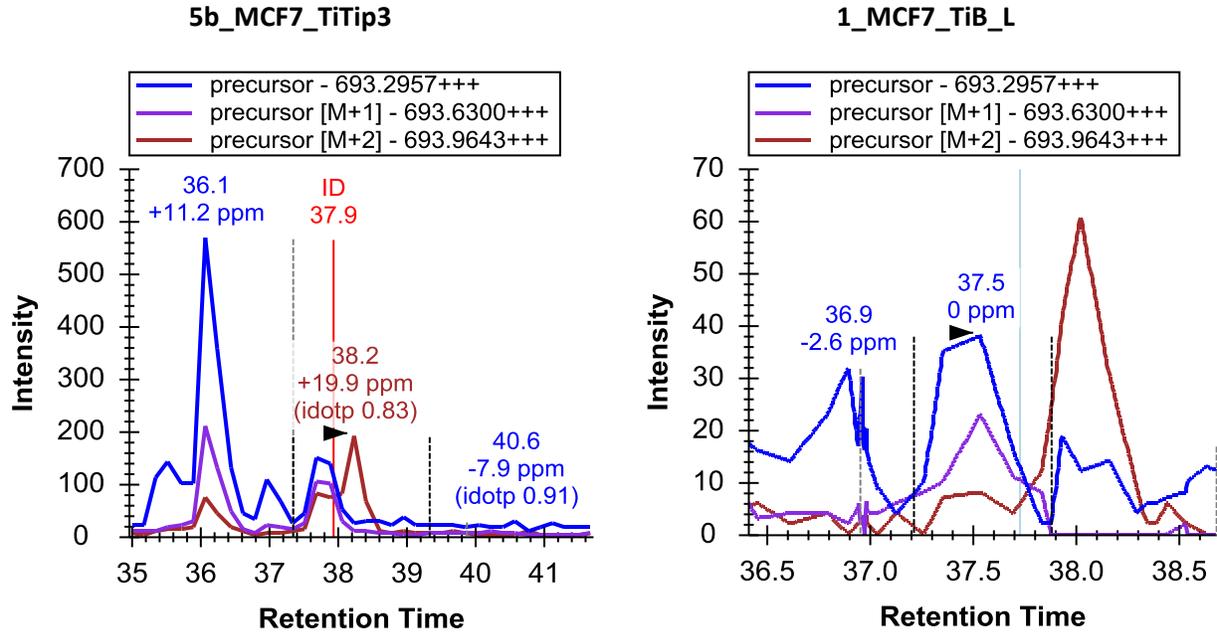


You can now feel fairly confident that the peak at 33 minutes is caused by another peptide of very similar atomic composition to the target and also triply charged, though with 1 Dalton greater monoisotopic mass.

- Click the Undo button (Ctrl-Z) 3 times, until you are back to the original integration correction.

Detecting and understanding interference

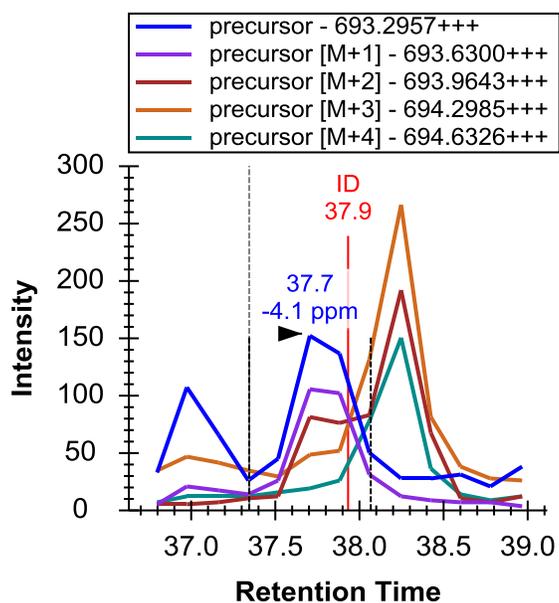
Continuing on, armed with these tools for understanding chromatographic data in Skyline, you do not have to go far to find your first peptide with true interference. You will see chromatogram graphs for the doubly-phosphorylated peptide ASLGSSLEGAEAEASSPKGK that look like:



Again, there is an ID for the peptide in 5b_MCF7_TiTip3 but not in 1_MCF7_TiB_L. The peak in 1_MCF7_TiB_L was picked based on alignment with the ID in 5b_MCF7_TiTip3. It appears to be almost free of interference from the peak to its right on its M+2 chromatogram, with a 0 ppm mass error on its most abundant peak. If you use your mouse scrollwheel to zoom out again, you will see both graphs contain a very similar peak at about 36 minutes with +11.2 and +9.6 ppm mass error and idotp values of 0.78 and 0.76 (which you can see in the **Peak Areas** view, after clicking the retention time annotations to select the peaks).

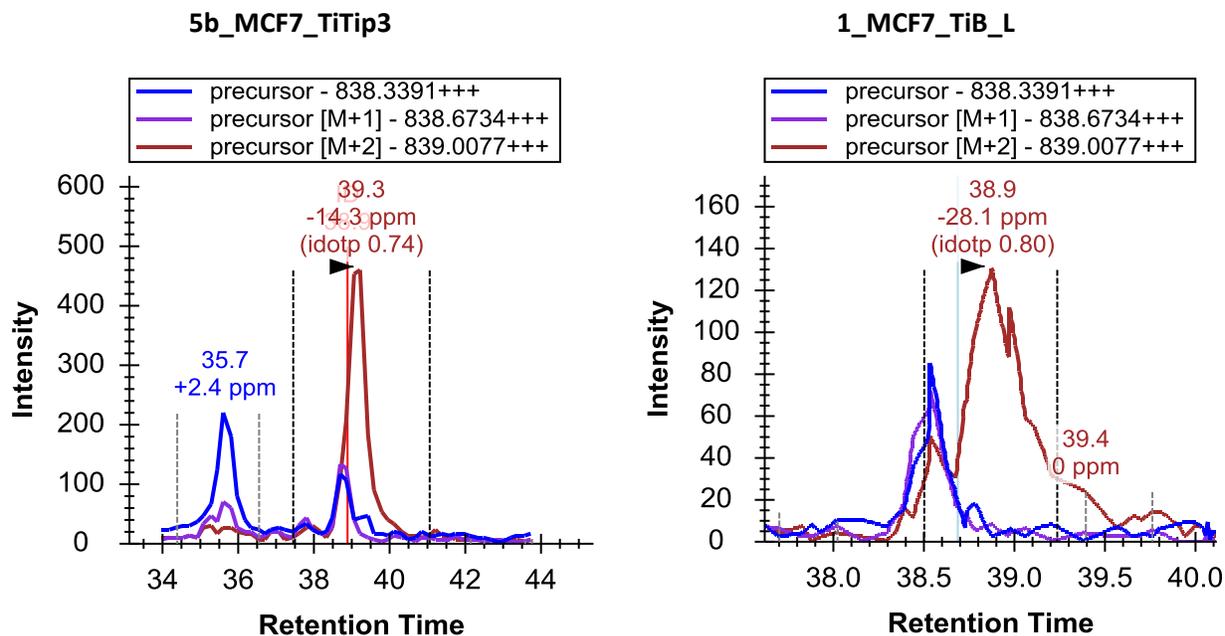
The integration boundaries for 5b_MCF7TiTip3 actually include the interference on M+2, and in fact the other peak in this chromatogram is close enough that it seems unlikely you could exclude its signal entirely even with very careful manual integration. If you try though, you can get peak integration with 0.94 idotp and -4.1 ppm mass error.

You can use the same technique of adding M+3 and M+4 to see that the interfering peak is likely caused by another triply charged peptide with 2 Da greater mass:

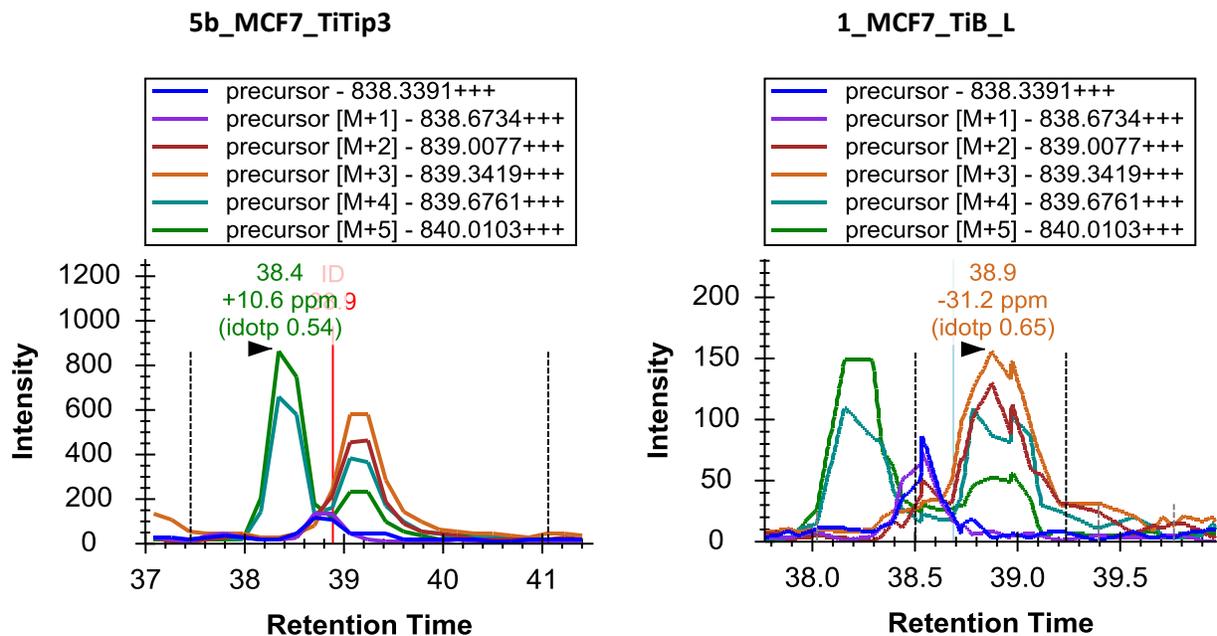


- Click the **Undo** button to revert this change and continue exploring.

Two peptides further down at AEGEWEDQEALDYFSDKESGK, you will find a stronger interference, with signal that is even more difficult to exclude.



Adding the M+3, M+4 and M+5 chromatograms for this peptide clarify just how crowded this particular mass and retention time combination is in the precursor ion space:



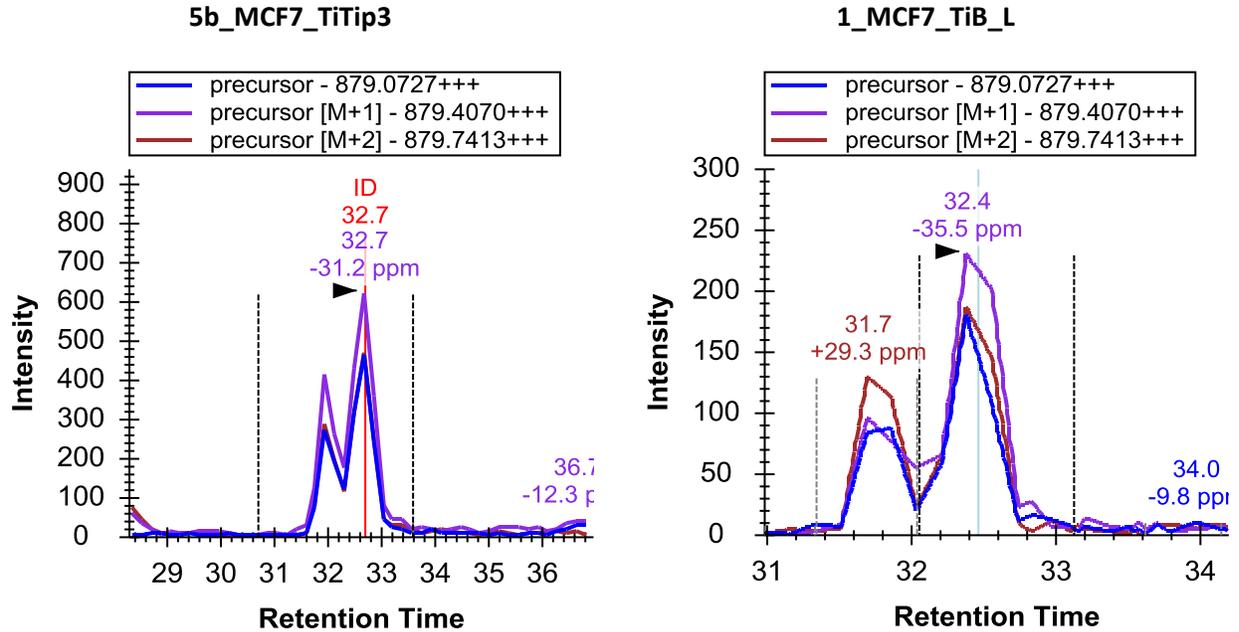
To obtain cleanly integrated peaks for this peptide you must remove all but the M and M+1 chromatograms. Do this now, and then properly adjust the integration boundaries. By now you might be wishing for a little bit more selective method, but you really can get a lot of useful quantitative data out of just MS1 scans. For quantitative statistics you may want to limit yourself to the highest ranked precursor ion without obvious interferences. With acceptable peak identification, this can limit the impact of interferences like the ones you have seen in this tutorial data.

Continue reviewing peptides, and you should only need one minor integration adjustment until you reach peptide 22 ALVEFESNPEETREPGSPPSVQR.

Differently modified forms of a peptide

Here you will find that the document contains ALVEFESNPEETREPGSPPSVQR and below it ALVEFESNPEETREPGSPSVQR, both with precursor m/z of 879.0727. The search engine, in this case Protein Pilot, has identified the former in 5b_MCF7_TiTip3 and the latter in 1_MCF7_TiB_L, but the chromatograms make it pretty clear that the same peak is being identified, both at about 32.5 min.

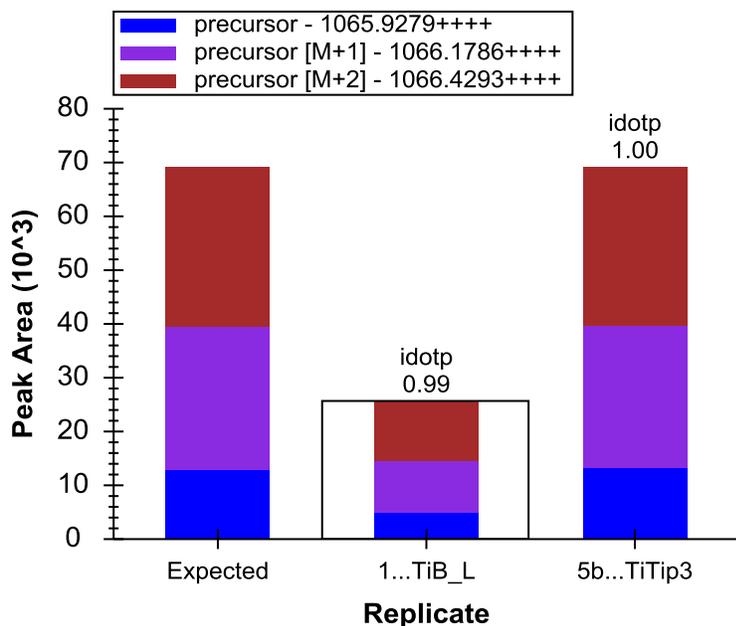
Even a little more interesting, you can see that there are actually two peaks very close together with the same m/z and at least very similar isotope distributions.



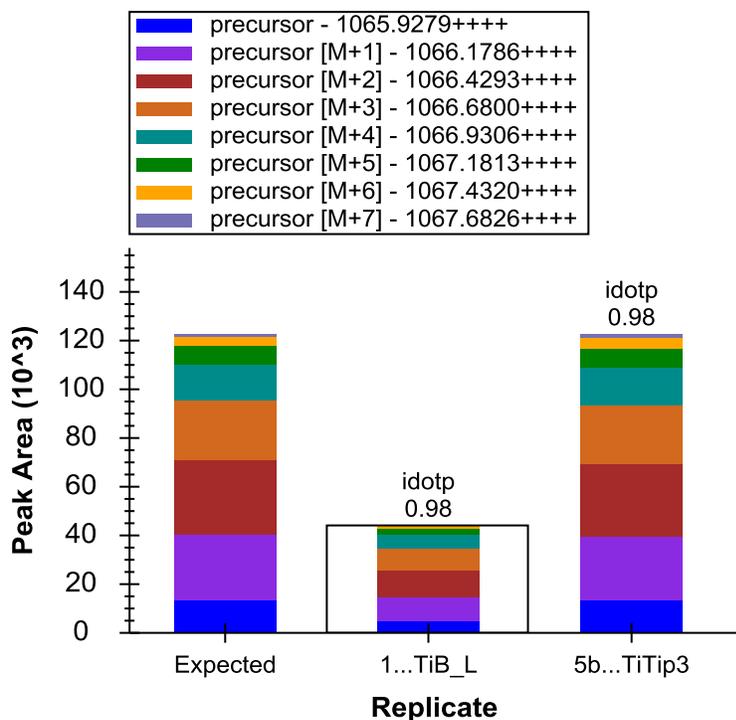
In 1_MCF7_TiB_L the isotope distributions and the mass errors make the two peaks, occurring between 31.5 and 33 minutes in both cases, look more different than in 5b_MCF7_TiTip3, but this may simply be due to variance. Adding M+3, M+4 and M+5, you can see that both peaks maintain greater than 0.9 idotp values (again integrate each and watch the **Peak Areas** view, and **Undo**). Since this peptide has 4 different possible phosphorylation sites, the two peaks may well be different single phosphorylation states of the same peptide, or phospho isoforms may have overlapping elution profiles. Careful assessment of potential isoforms is recommended during MS1 Filtering (beyond search engine outputs).

More tools for data analysis

Continuing down to peptide 25, you will find YGPADVEDTTGSGATDSKDDDDIDLFGSDDEEESEAKR, the longest and the first quadruply charged peptide precursor in this document. Because it is so large, its isotope distribution is quite different from smaller, doubly charged peptides and even the somewhat larger, triply charged peptides you have been looking at. The monoisotopic peptide, with no ¹³C atoms, is now expected to occur less frequently than both the M+1 and M+2 forms. You can see this is the case for these chromatograms, which produce idotp values of 1.0 and 0.99 with the expected distribution:



Using the transition pick-list, as you have been doing, you can add the chromatograms for M+3 through M+7, all of which should contain greater than 1% of the entire isotope distribution, and see that the idotp values remain quite high at 0.98:

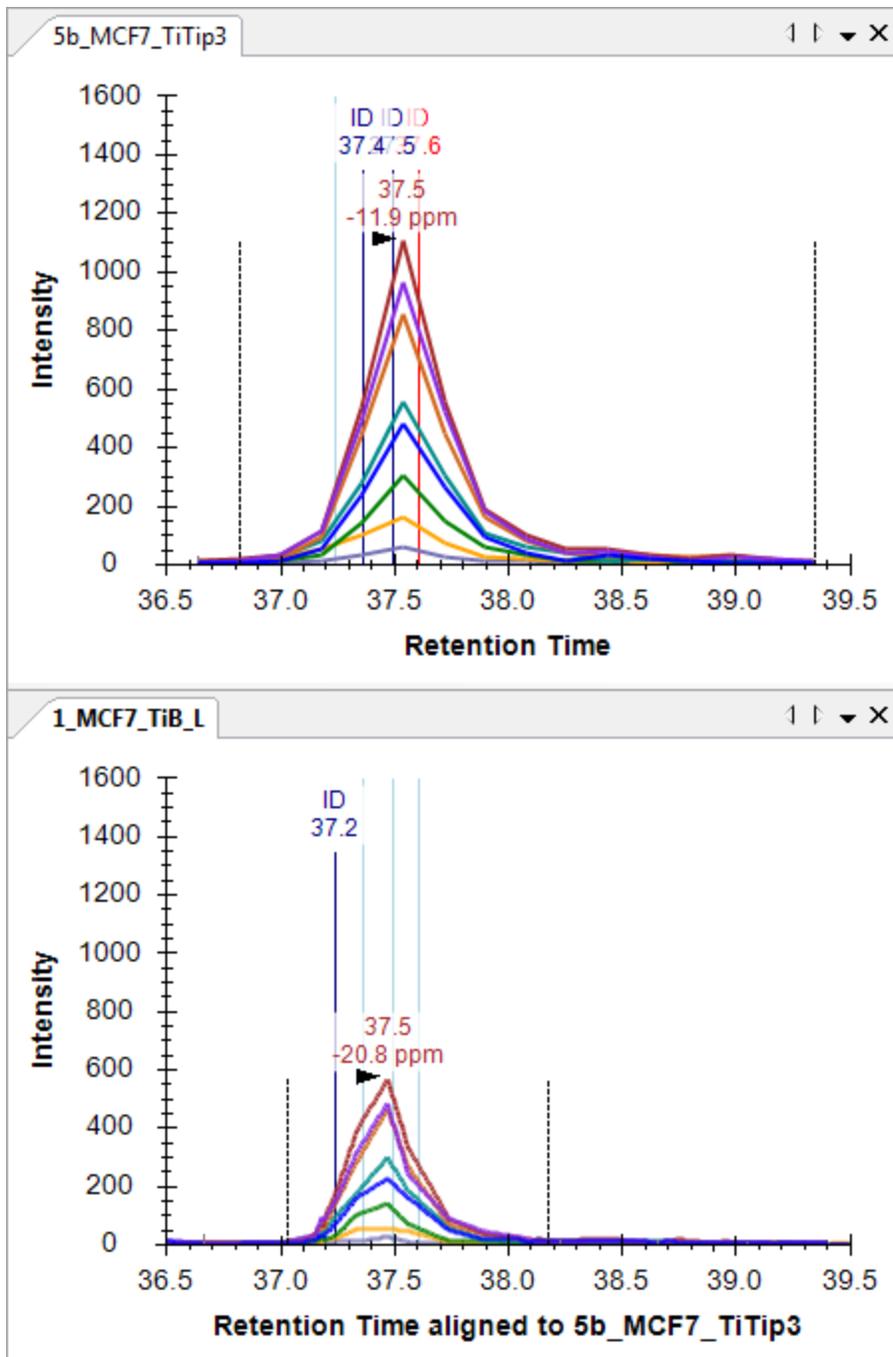


In the chromatogram graphs, you can see that this is the only peptide in the document that was identified multiple times in a single run (3 in 5b_MCF7_TiTip3).

You can do the following to put the chromatograms on the same scale and make it easier to interpret how these IDs are aligned between replicates:

- Right-click a chromatogram graph, and check **Synchronized Zooming**.
- Right-click a chromatogram graph, and uncheck **Auto-scale Y-axis**.
- Right-click the 5b_MCF7_TiTip3 chromatogram graph, and check **Align Times To 100807_0005b_MCF7_TiTip3**. (Note that this will align all peptides in this data set, until this is turned off again, and not just this current peptide.)
- Hover the mouse cursor over the plot area and roll the mouse scroll wheel back toward you to zoom out slightly.
- Click and drag a narrow rectangle around the integration range in 5b_MCF7_TiTip3.

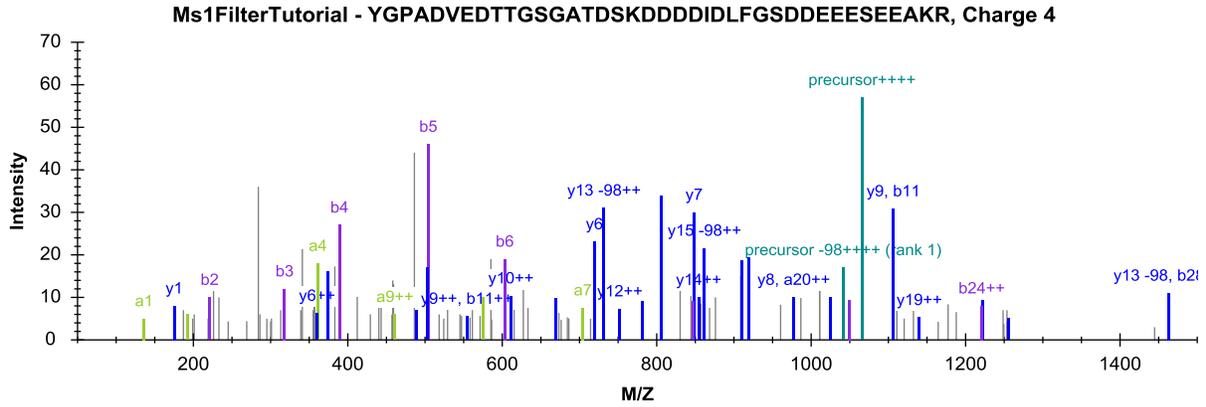
The chromatogram graphs should now look like this:



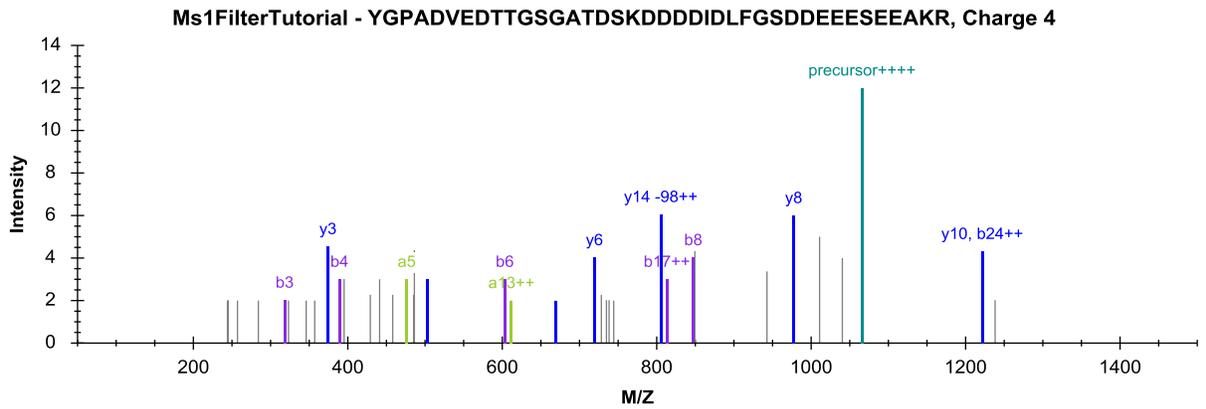
You can see the alignment makes the IDs line up quite nicely, and so do the peaks. The synchronized zooming with y-axis auto-scaling turned off gives you a sense for the relative height of the peaks.

You can click on the ID annotations in the chromatogram graphs to review the spectra that the search engine identified as this peptide, or you can click on the dropdown list at the top of the **Library Match** view and use the arrow keys to page up and down through the matched spectra. It may take a bit of imagination to convince yourself that the spectra from the different runs are from the same peptide:

5b_MCF7_TiTip3 (37.61 Min)

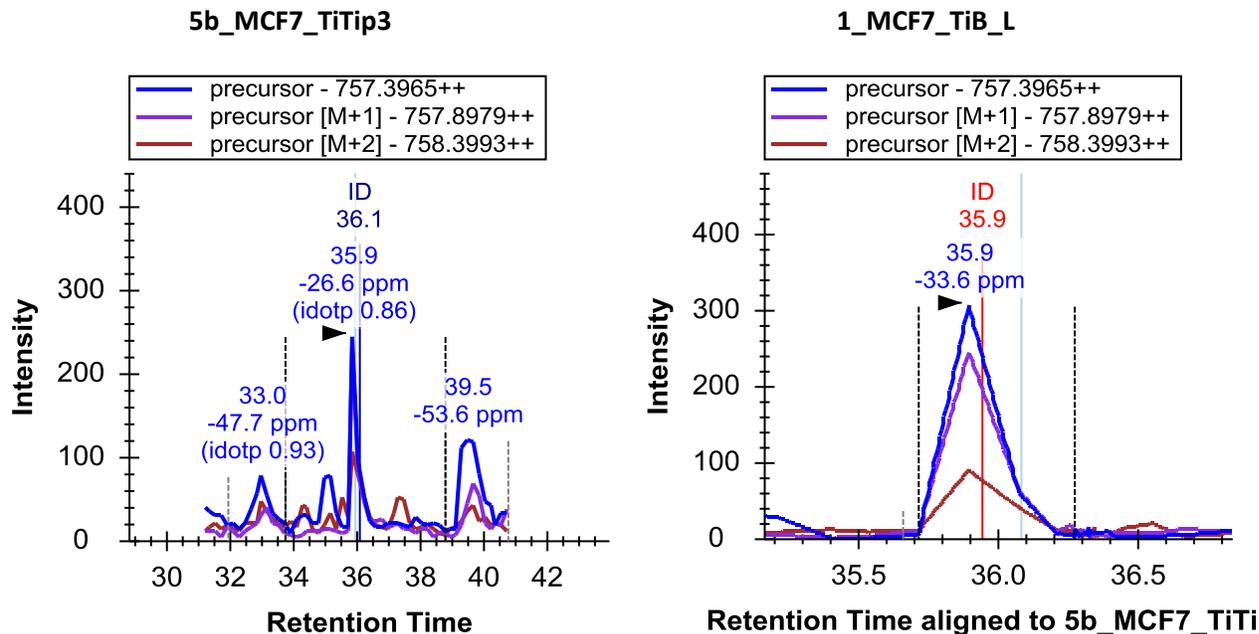


1_MCF_TiB_L (37.03 min)



But you should feel pretty confident that the chromatogram peaks in the two runs measure the same peptide molecule.

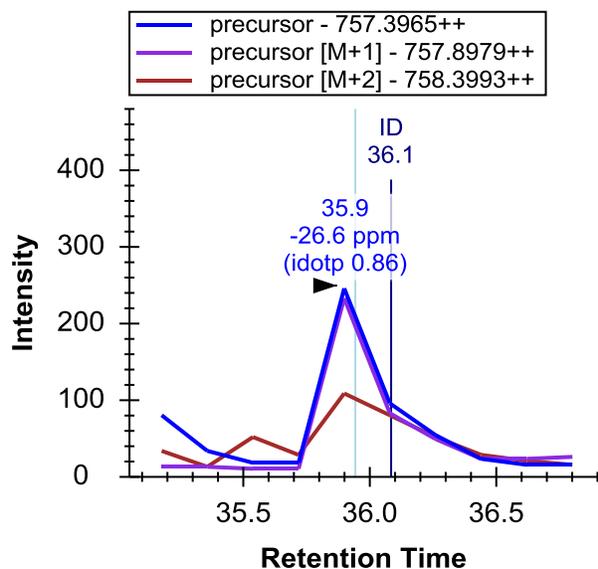
Continue down to peptide 27, GVVDSEDLPLNISR, where you will find the integration needs adjusting:



With **Synchronize Zooming** turned on, you can do the following to zoom into the peak with integration boundaries that are too far apart:

- Click in the 1_MCF7_TiB_L chromatogram plot area.
- Move the mouse scroll wheel slightly in either direction.

The graph for 5b_MCF7_TiTip3 will zoom to the same scale as the one for 1_MCF7_TiB_L:

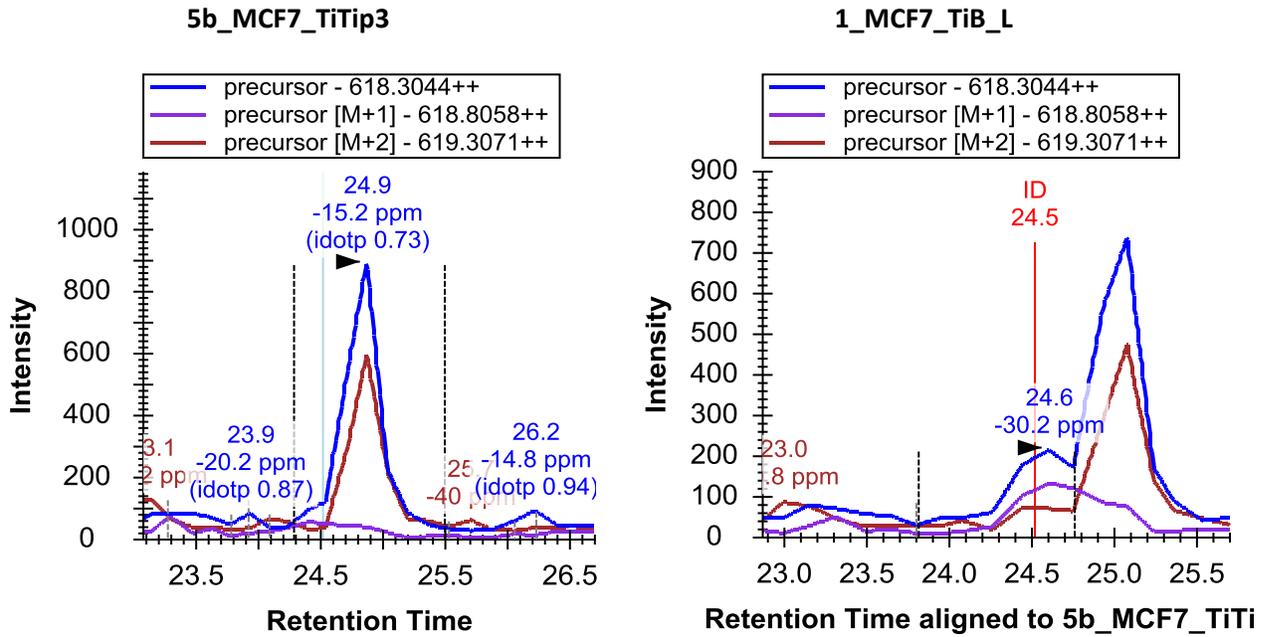


This should allow you to easily reset the integration boundaries by clicking and dragging beneath the Retention Time axis from 35.7 to 36.5 minutes. In the peak areas view, you will see that the peak areas

for both runs now total around 8-10,000, and the idotp value for 5b_MCF7_TiTip3 has improved from 0.86 to 0.97.

More fun with interference

The next peptide DQVANSAFVER has another interesting interference:



In 1_MCF7_TiB_L, the peptide was identified at 24.5 minutes. In both replicates, however, you will see a strong interfering peak around 25 minutes. Though, the interference is only on M and M+2. Because the target peptide is doubly charged, this tells you that the interfering peptide is singly charged. The target signal is weak enough in 5b_MCF_TiTip3 and the interference strong enough, that it is hard to see a peak for the target at all, even on the M+1 chromatogram.

In such a serious case of interferences in both replicate acquisitions, one should err on the careful side and disqualify this peptide from MS1 quantification. If you know you really want to measure this specific peptide, you may want to move to a more selective method, like targeted MS/MS or SRM.

The remaining seven issues in this file mostly amount to more of what you have already seen, and what you are hopefully now well equipped to understand and resolve, enumerated below for those interested, but feel free to disregard and continue on to the next section:

1. ETERASPIKMDLAPSK (31 & 32) – Repeated twice in the same protein (delete one)
2. KTGSYGALAEITASK & KTGSYGALAEITASK (34 & 35) – Peak given two phospho sites
3. TPSPKEEDEEPEPPEKK (41) – Poor chromatography mis-integrated (zoom, adjust integration)
4. KEKTPELPEPSVK (46) – Interference on M+1 and M+2
5. EKTPELPEPSVK (47) – Interference on M+2
6. VPKPEPIPEPKEPSPEKNSK & VPKPEPIPEPKEPSPEKNSK (49 & 50) – Peak given two phospho sites

7. KETESEAEDNLDDLEK (51) – Interference on M+1 from singly charged peptide

Minimizing a Chromatogram Cache File

After completing this pass through the now 50 peptides in your document, all should be pretty well integrated. Before continuing, to save the current document:

- On the **File** menu, click **Save** (ctrl-S).

Next, to eliminate extraneous chromatogram data for this document and make it as small and as easy to share as possible, perform the following steps:

- On the **Edit** menu, click **Manage Results**.
- Click the **Minimize** button.
- Check the **Limit noise** check box.
- Enter '2' in the field to specify noise **minutes before and after chromatogram peak**.
- Press the Tab key to get the form to re-estimate the size reduction.

The **Minimize Results** form should now look like:

To reduce the size of the Skyline cache file (.skyd), you can discard chromatograms that are not used by this document, as well as limit the length of chromatograms.

Discard unused chromatograms

Limit noise to minutes before and after chromatogram peak.

The current size of the cache file is 3.21 MB

After minimizing, the cache file will be reduced to 13% its current size

Replicate	Current Size	Minimized Size
1_MCF7_TiB_L	1.93 MB	14%
5b_MCF7_TiTip3	1.28 MB	13%

Minimize in place Minimize and save as... Cancel

The form indicates that this operation is expected to reduce the size of the cache file from about 3.24 MB to 13% of its current size, or about 420 K.

- Click the **Minimize and save as** button.
 - In the **File name** field of the **Save As** form, enter the name 'Ms1FilteringTutorial-2min.sky'.
1. Click the **Save** button.

If you again press Shift-F11 to zoom out, you can review the chromatograms for your peptides in this new document, and see that they now extend only 2 minutes in either direction beyond the integration boundaries.

Chromatogram minimization can be quite useful in creating a document for a large experiment which you can share as supplementary data for a manuscript. You likely still want to make the raw data available on-line, but a minimized Skyline document will provide a rich view of your data at a fraction of the download cost.

Inclusion List Method Export for MS1 Filtering

As noted above, multi-replicate studies using DDA do show undersampling of MS/MS, and not all peptides have MS/MS identification in each acquisition replicate. MS1 Filtering can overcome this problem using RT alignment as described previously. However, when you have moved from initial pure discovery to knowing even a relatively large number of peptides you wish to target, you can use Skyline to export an inclusion list method for your DDA experiment, achieving an approach termed "Accurate Inclusion Mass Screening"². The inclusion list method should increase your chances, over unguided DDA, of sampling the peptides you are interested in.

At present, Skyline can export inclusion list methods for both AB SCIEX and Thermo instruments, and we are working with Agilent and Waters. To export an inclusion list method for subsequent MS1 filtering from the tutorial Skyline document, you would perform the following steps:

- On the **Settings** menu, click **Peptide Settings**.
- Click the **Prediction** tab.
- Check the **Use measured retention times when present** check box.
- In the **Time window** field, enter an appropriate window for the expected chromatographic stability, like '10' minutes. Reducing window overlap is less important than in SRM.
- Click the **OK** button.
- On the **File** menu, choose **Export** and click **Method**.
- In the **Instrument type** dropdown list, choose 'AB SCIEX TOF'.
- In the **Method type** dropdown list, choose 'Scheduled'.
- In the **Template file** field, enter the path to an acquisition method template file for the QSTAR system...

That is about as far as you can go with instrument method export in this tutorial, unless you actually have a system with vendor instrument software installed for one of the supported vendors (AB SCIEX Analyst or Thermo Xcalibur). For all method export from Skyline, it is recommended that you perform the export function on an instance of Skyline running on the instrument control computer for the

instrument on which you intend to run your method. Since, even if your lab owns a supported instrument, it is unlikely you are doing this tutorial on it, completing the steps above is left to you when you have the need.

Conclusion

In this tutorial, you have learned some of the most basic and critical features for using Skyline to extract quantitative information from the MS1 scans in your DDA experiment data. Fortunately, much of the previously existing Skyline functionality still applies equally well to MS1 extracted chromatograms as to the SRM chromatograms for which it was originally designed. It is, therefore, recommended that you spend significant time understanding the material presented in the other Skyline tutorials and instructional videos. Though the idea of using chromatogram peak areas extracted from MS1 scans has been around for a long time, MS1 Filtering is still a relatively new feature area for Skyline. As such, you can expect it to continue improving rapidly.

Supplement

If you are interested in reviewing the real data sets processed for the original MS1 Filtering paper¹, visit the following link, where you can download documents that have been minimized as described above:

http://proteome.gs.washington.edu/supplementary_data/MS1_Filtering/minimized/

You can browse the parent directory for full Skyline documents and raw data.

Bibliography

1. Schilling, B. *et al.* Platform-independent and Label-free Quantitation of Proteomic Data Using MS1 Extracted Ion Chromatograms in Skyline APPLICATION TO PROTEIN ACETYLATION AND PHOSPHORYLATION. *Mol Cell Proteomics* **11**, 202–214 (2012).
2. Jaffe, J. D. *et al.* Accurate Inclusion Mass Screening. *Mol Cell Proteomics* **7**, 1952–1962 (2008).