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 project. 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 new Skyline MS1 Full-Scan Filtering 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:

- Generating Skyline spectral libraries from discovery datasets
- 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, 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.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.

Configuring a Skyline Document for a Data Dependent Dataset

To begin working with a DDA experiment in Skyline, you will create a spectral library from the output of one of the many peptide search engines supported by Skyline. The steps necessary to build spectral libraries are covered in detail in the "Targeted Method Editing" tutorial. Because these spectral libraries are critical to working with MS1 filtering in Skyline, however, the steps are included again below. 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.

Configuring Digestion and Modification Settings

Follow the steps below to configure the peptide settings for your new document:

1. On the **Settings** menu, click **Peptide Settings**.
2. Click the **Digestion** tab.
3. In the **Enzyme** dropdown list select 'Trypsin [KR/P]'.
4. In the **Max missed cleavages** dropdown list select '2'.
5. In the **Background Proteome** dropdown list select 'None'.
6. Click the **Modifications** tab.
7. In the **Structural modifications** list, check 'Carbamidomethyl Cysteine'.

If the list does not contain any of the modifications:

- Phospho (ST)
- Phospho (Y)
- Oxidation (M)

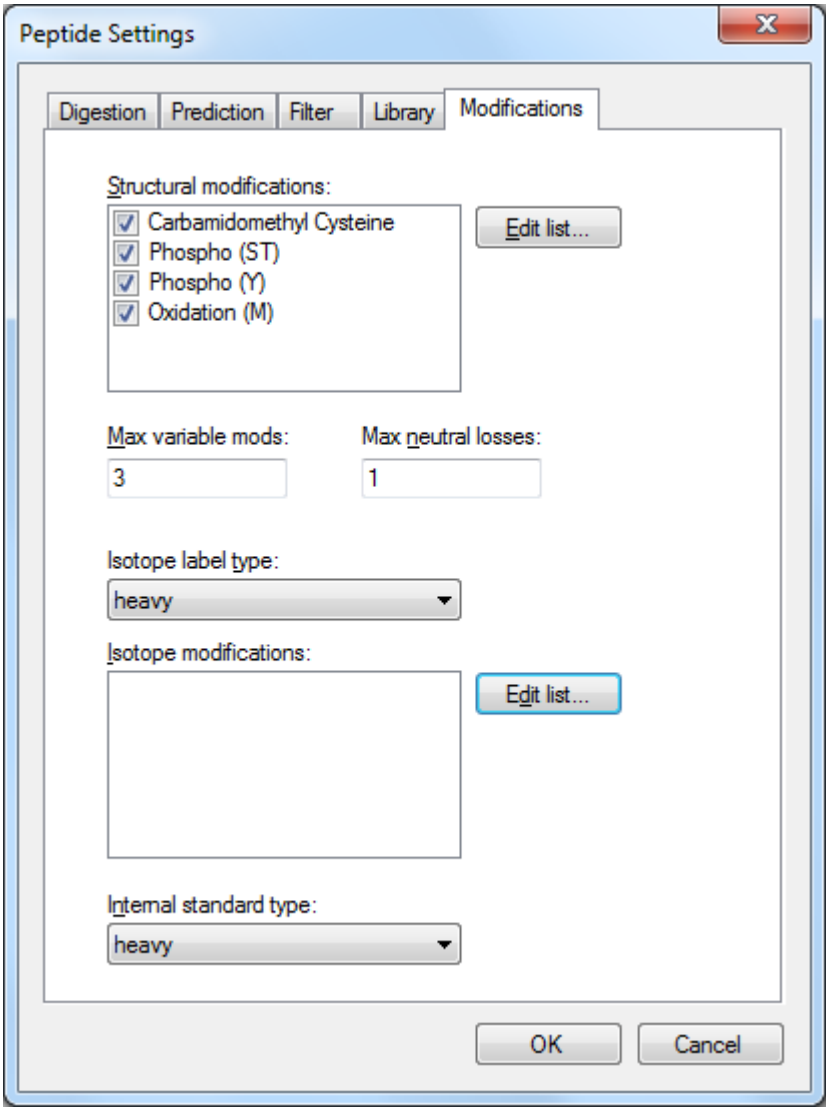
Perform the following steps to add them:

1. Click the **Edit List** button beside the **Structural modifications** list.
2. Click the **Add** button.
3. In the **Name** dropdown list, select one of the modifications named above.
4. Check the **Variable** check box.
5. Click the **OK** button.
6. Repeat 2-5 until all of the above modifications are in the list.
7. Click the **OK** button on the **Edit Structural Modifications** form.

To complete the modification configuration for the peptide search you will work with in this tutorial:

1. Check the 3 modifications listed above.

The form should look something like:



2. Click the **OK** button on the **Peptide Settings** form.

Building a Spectral Library

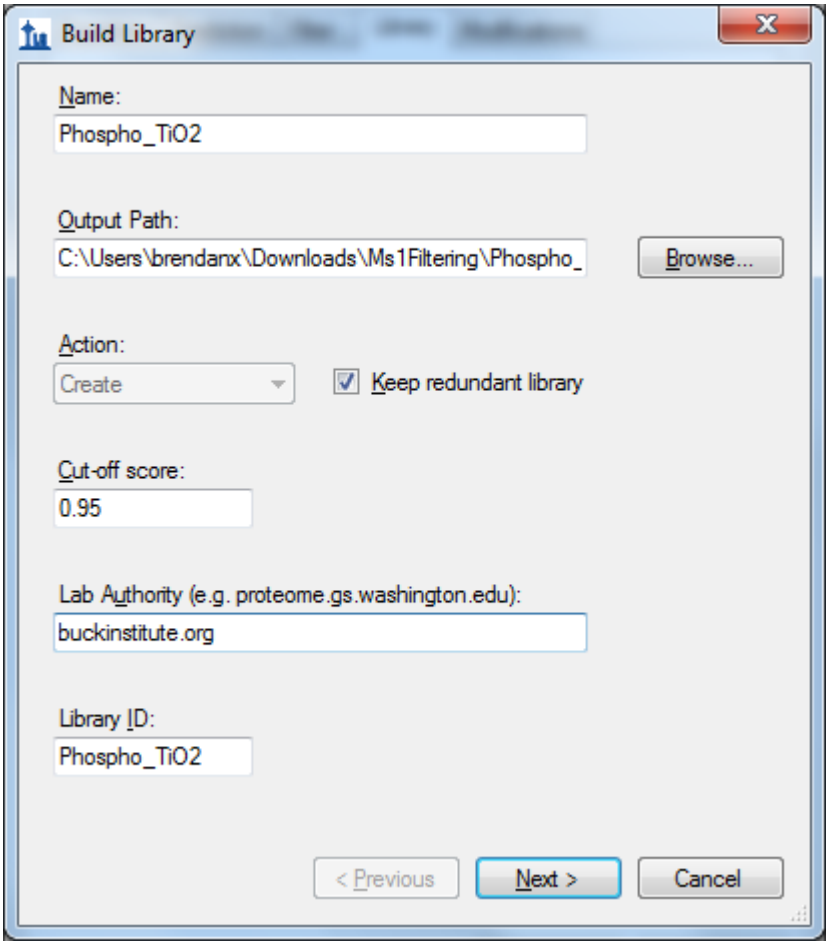
Now you are ready to generate the spectral library for this tutorial, which will be critical in supporting new Skyline MS1 Filtering features. Unlike SRM, with MS1 Filtering it is also useful to keep all of the redundant spectra for each peptide identified, since Skyline will make it possible for you to inspect these spectra during your review of the MS1 chromatogram results.

To create the spectral library, perform the following steps:

1. On the **Settings** menu, click **Peptide Settings**.
2. Click the **Library** tab.
3. Click the **Build** button.
4. In the **Name** field of the **Build Library** form, enter 'Phospho_TiO2'.
5. Click the **Browse** button, beside the **Output Path** field.

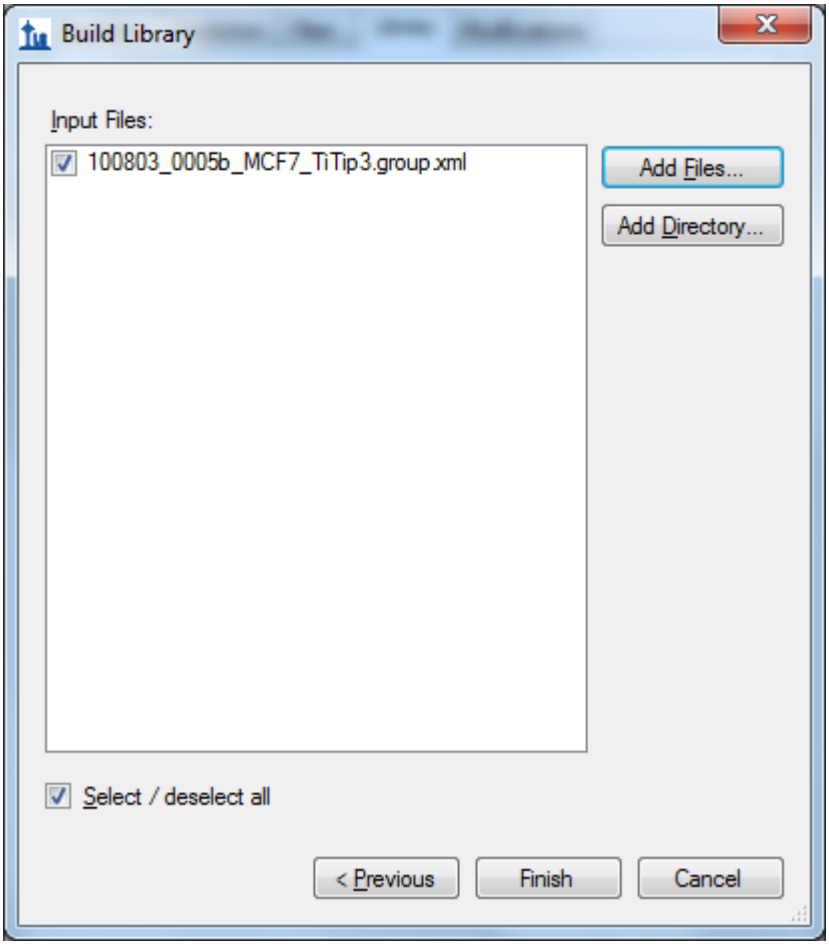
6. Navigate to the MS1Filtering folder created earlier.
7. Check the **Keep redundant library** check box.
8. In the **Cut-off score** field, enter '0.95'.
9. In the **Lab Authority** field, enter 'buckinstitute.org'

The form should look something like:



10. Click the **Next** button.
11. Click the **Add Files** button (to add a peptide search result file).
12. Navigate to the MS1Filtering folder.
13. Double-click the "100803_0005b_MCF7_TiTip3.group.xml" file in this folder.

The form should appear as below:



14. Click the **Finish** button.

The library will build inside Skyline and within a few seconds a notification appears briefly in the lower left corner of Skyline to indicate that the library has been successfully built.

To complete library configuration, perform the following steps:

15. In the **Libraries** list, check the new 'Phospho_TiO2' library.
16. In the **Pick peptides matching** dropdown list, choose 'Library'.
17. Click the OK button to close the **Peptide Settings** form.

Note that still no peptides are displayed yet in the Skyline peptide tree on the left, which you will populate later in the tutorial.

In the MS1Filtering folder where Skyline saved your new spectral library you should now find two new files:

- The non-redundant library "Phospho_TiO2.blib" containing best matching spectra.
- The redundant library called "Phospho_TiO2.redundant.blib" containing all matching spectra.

Phospho_TiO2.blib	1/11/2012 3:05 PM	BLIB File	647 KB
Phospho_TiO2.redundant.blib	1/11/2012 3:05 PM	BLIB File	672 KB

You may also find a file 'Phospho_TiO2.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.

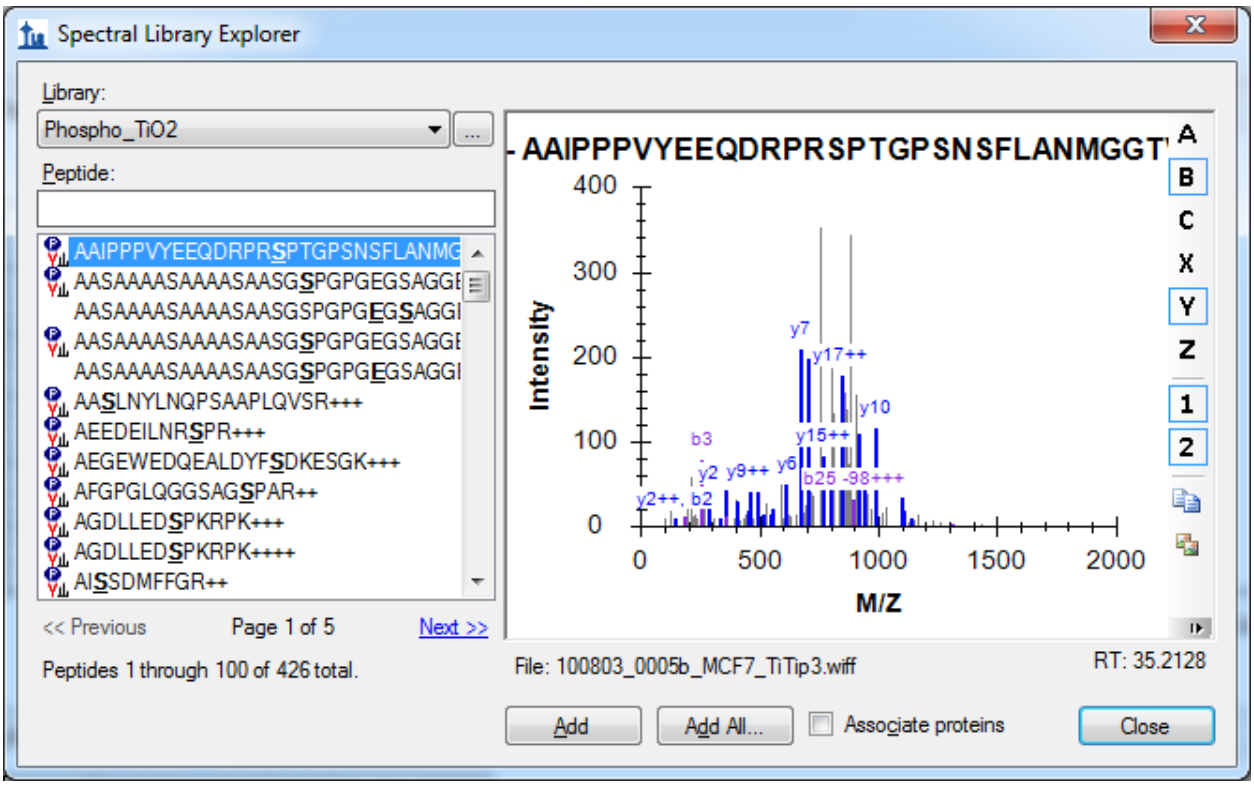
Verifying Library Retention Time Information

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. For instance, at the time of this revision, you can use Skyline to build a spectral library from mzIdentML and MGF files produced by Scaffold, but these Scaffold results do not contain retention time information, and neither will your library.

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

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

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



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 are about to import into Skyline. The "File" values

need not be exact matches of the files you will 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. You are encourage to post to the Skyline support board (on the **Help** menu, click **Support**) for help with issues of this type.

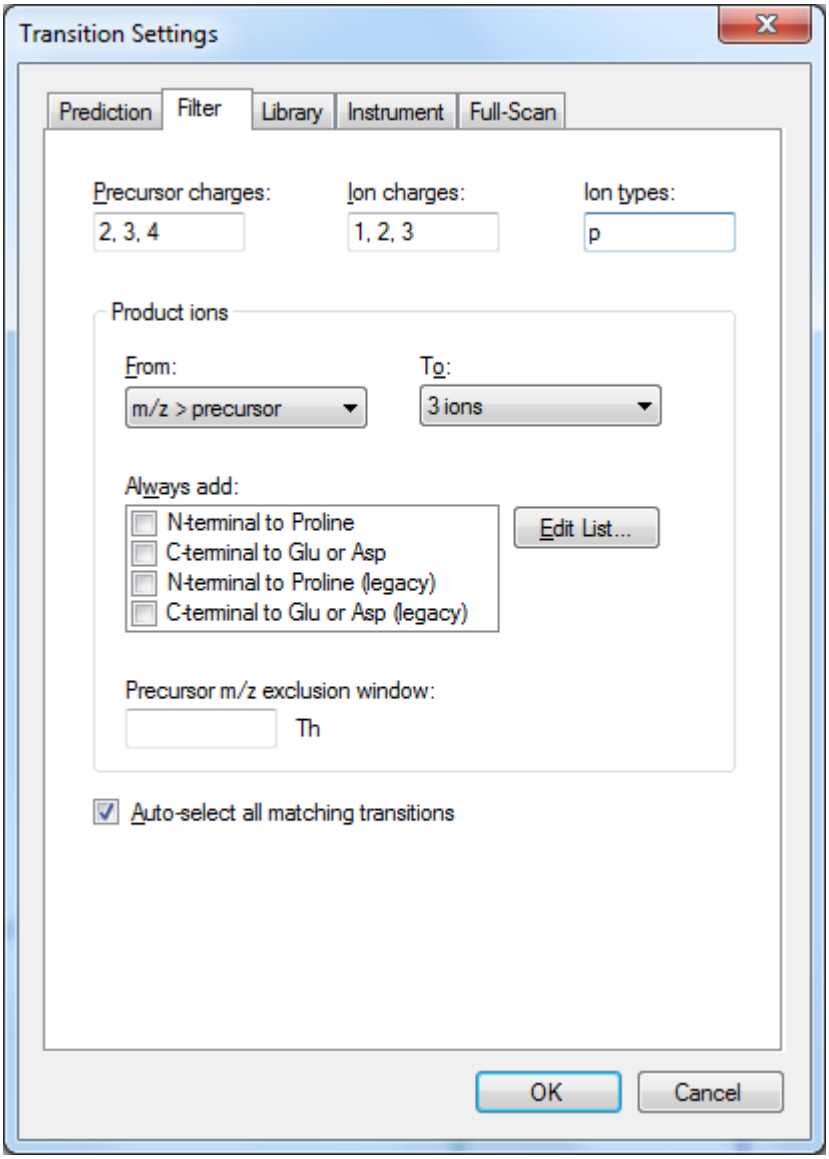
Now press the Down-Arrow key to select other peptides, and you will see that the "RT" value changes. If the library included peptide IDs from multiple mass spec files, you would also expect to see the "File" value changing. You have just verified that this library will support Skyline MS1 Filtering.

Configuring Appropriate Transition Settings

To configure the **Transition Settings** to filter intensity values for the desired ion *m/z* peaks from the MS1 scans in a DDA raw data file, perform the following steps:

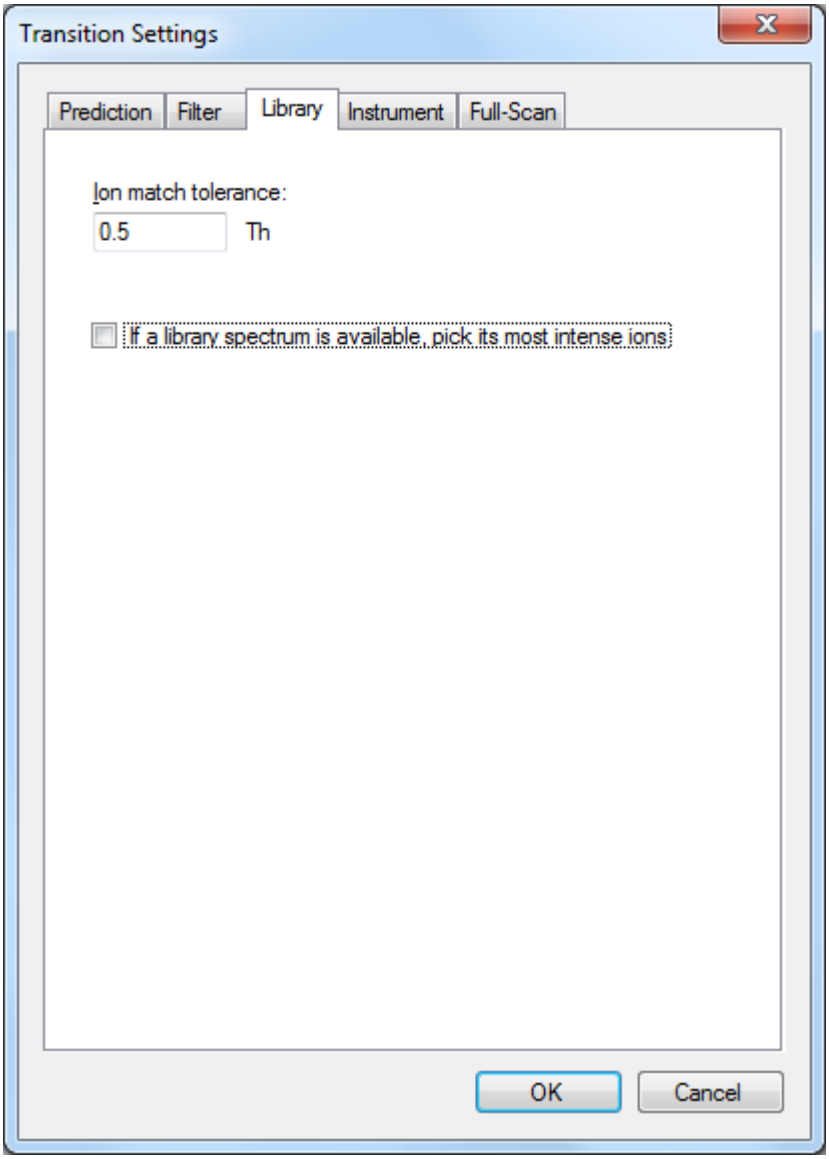
1. On the **Settings** menu, click **Transition Settings**.
2. Click the **Filter** tab.
3. In the **Precursor charges** field, enter '2, 3, 4'.
4. In the **Ion charges** field, enter '1, 2, 3'.
5. In the **Ion types** field, enter 'p' and only 'p' (p stands for precursor which is what you want Skyline to filter from MS1 scans).
6. The **Product ions** fields will be ignored with only 'p' in the **Ion types** field.
7. Check the **Auto-select all matching transitions** check box.

The **Filter** tab should now look something like:



8. Click the **Library** tab.
9. Uncheck the **If a library spectrum is available, pick its most intense ions** check box.

The **Library** tab should look like the following:



Configuring Full-Scan Settings for MS1 Chromatogram Extraction

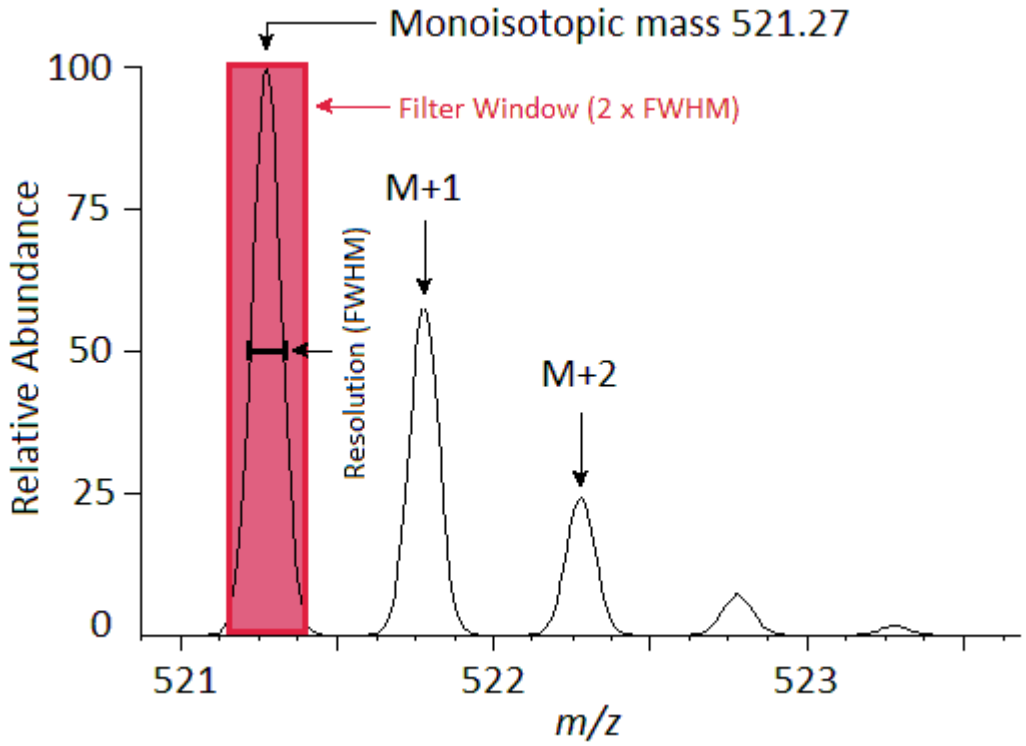
To configure Skyline to extract chromatograms from the MS1 scans in the DDA run included in this tutorial, perform the following steps:

1. Click the **Full-Scan** tab, in the **Transition Settings** dialog.

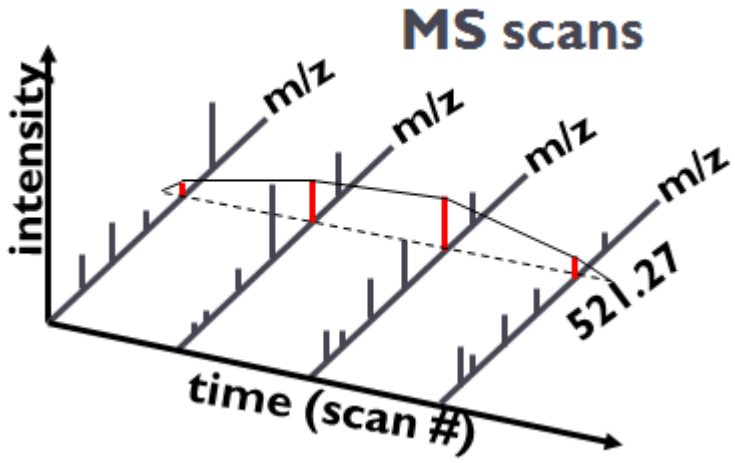
In the **MS1 filtering** section on top, you will find the settings that control chromatogram extraction from MS1 scans.

2. In the **Isotope peaks included** dropdown list, choose 'Count'.
3. In the **Peaks** field, enter '3' to have Skyline filter the first 3 isotope peaks (M, M+1 and M+2) from this high resolution data.
4. In the **Precursor mass analyzer** dropdown list, leave the default value 'TOF' (as the data were acquired on a QSTAR Elite).

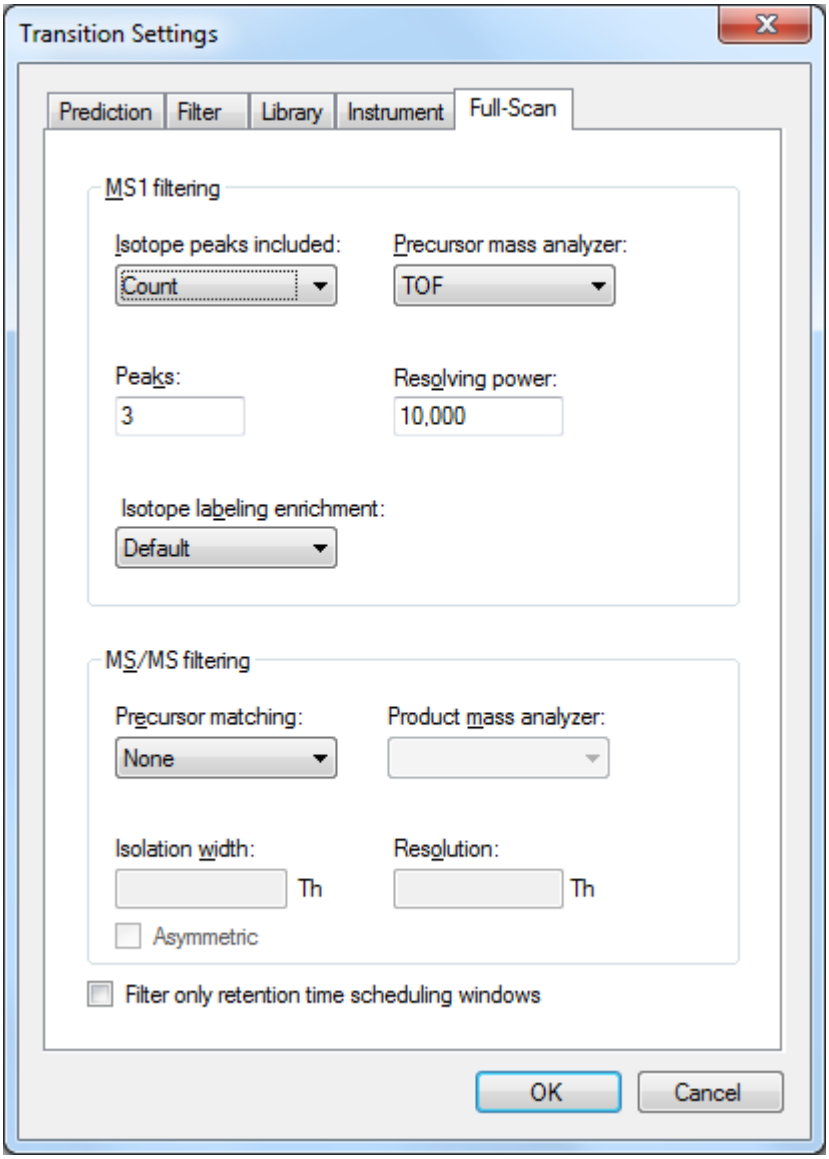
5. In the **Resolving power** field, also leave the default value '10,000'. This field defines the width of the MS1 Filtering window around each precursor *m/z* to be filtered. Skyline uses this value to predict full width at half maximum (FWHM) for a peak in the *m/z* dimension, and uses 2 times that 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 Skyline **Transition Settings** form, the **Full-Scan** tab should look like:



6. Click the **OK** button to commit your changes to the **Transition Settings**.

Populating the Skyline Peptide Tree

With the necessary configuration completed, you will next add to the document the peptides for which you would like Skyline to extract chromatograms and produce quantitative data. Skyline provides many ways to do this. You could paste a list of peptides directly or into the form presented by the **Edit** > **Insert** > **Peptides** menu item, or you could use the **Add** or **Add All** buttons in the Spectral Library Explorer to add peptides from your peptide search result for which you built the spectral library. These methods are covered in other tutorials. In this tutorial, you will import a FASTA file to quickly add all of the peptides in your spectral library for a specific set of protein sequences.

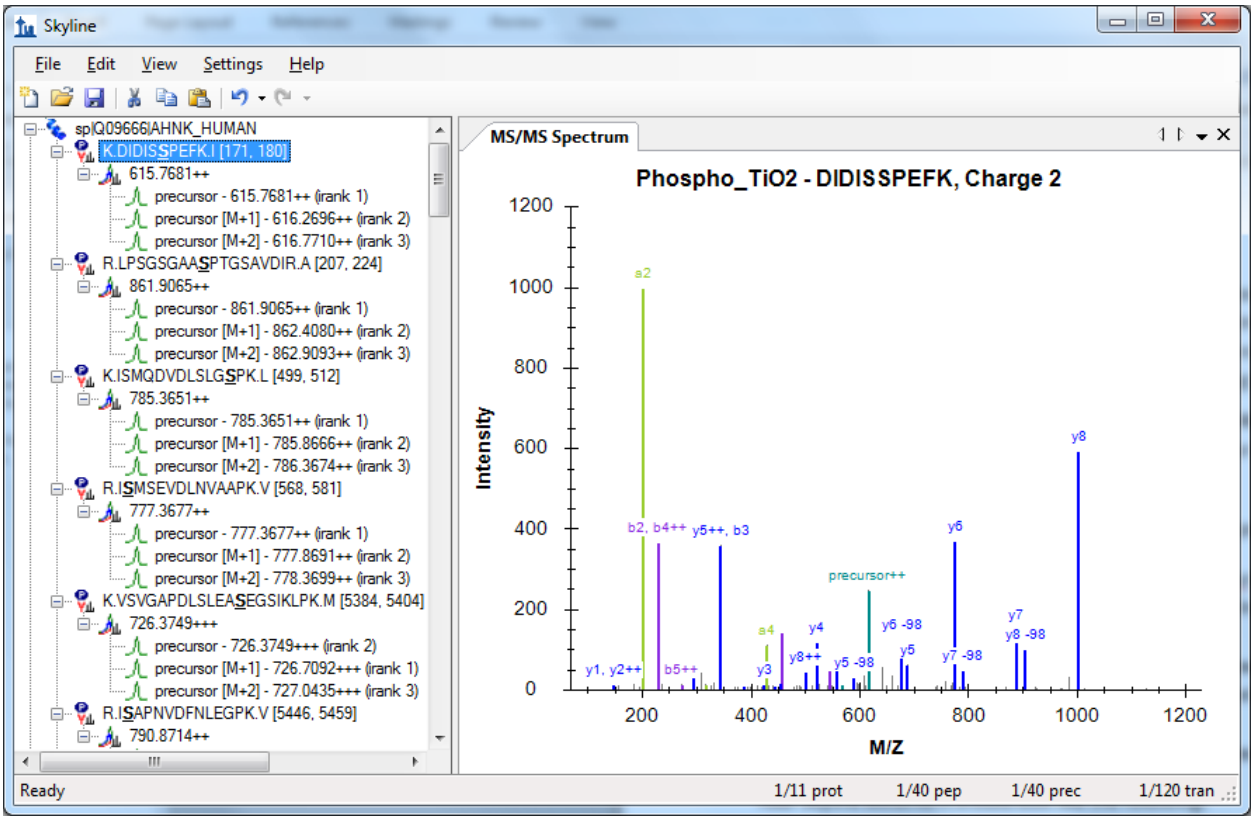
You could import a FASTA file of all human SwissProt protein entries to get a comprehensive identified peptide list (this MS experiment involves a human MCF7 breast cancer cell line sample and subsequent phosphopeptide enrichment), however, for file size reasons, you will import a much smaller FASTA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

1. On the **File** menu, choose **Import** and then click **FASTA**.
2. Navigate to the MS1Filtering folder you created for this tutorial.
3. Double-click the file '12_proteins.062011.fasta'.

You should see 40 peptides (counted in the status bar) appear in the Skyline peptide tree.

4. Click on the sequence of the first phosphopeptide K.DIDIS**S**PEFK.I and the MS/MS spectrum will appear. (Note that the bold, underlined residue "**S**" in the peptide sequence indicates a serine phosphorylation).
5. If you do not see the MS/MS spectrum, on the **View** menu, click **MS/MS Spectra**.
6. 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**.
7. 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. Note that when the Skyline document is set up for MS1 Filtering, in the place where you would normally 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 K.DIDIS**S**PEFK.I: precursor - 615.7681++, precursor [M+1] - 616.2696++, and precursor [M+2] - 616.7710++. Save this Skyline document as Template_MS1 Filtering_1118_2011_3.sky in the MS1Filtering folder.

MS1 Filtering of Raw Data Imported into Skyline

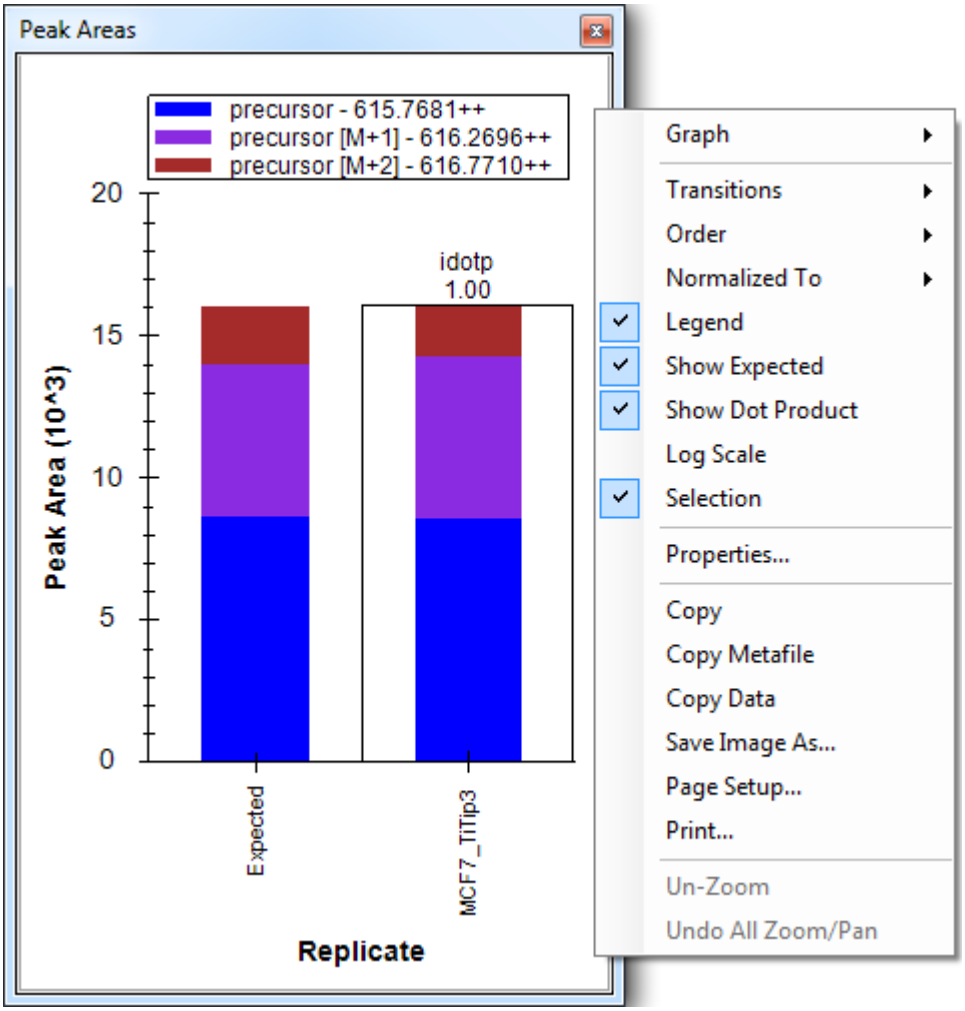
After setting up MS1 Filtering as described above, to import a DDA raw file, follow the steps below:

1. On the **File** menu, choose **Import** and click **Results**.
2. The **Import Results** form will appear. Keep the defaults, leaving the **Add single-injection replicates in files** radio button selected.
3. Click the **OK** button.
4. Navigate to the MS1Filtering folder.
5. Double-click the file 100803_0005b_MCF7_TiTip3.wiff.

Skyline will begin importing the file, with status indicated by the green progress bar at the bottom of the Skyline window, and the "**Caching file…**" note in the lower left corner. You are free, however, to continue using Skyline.

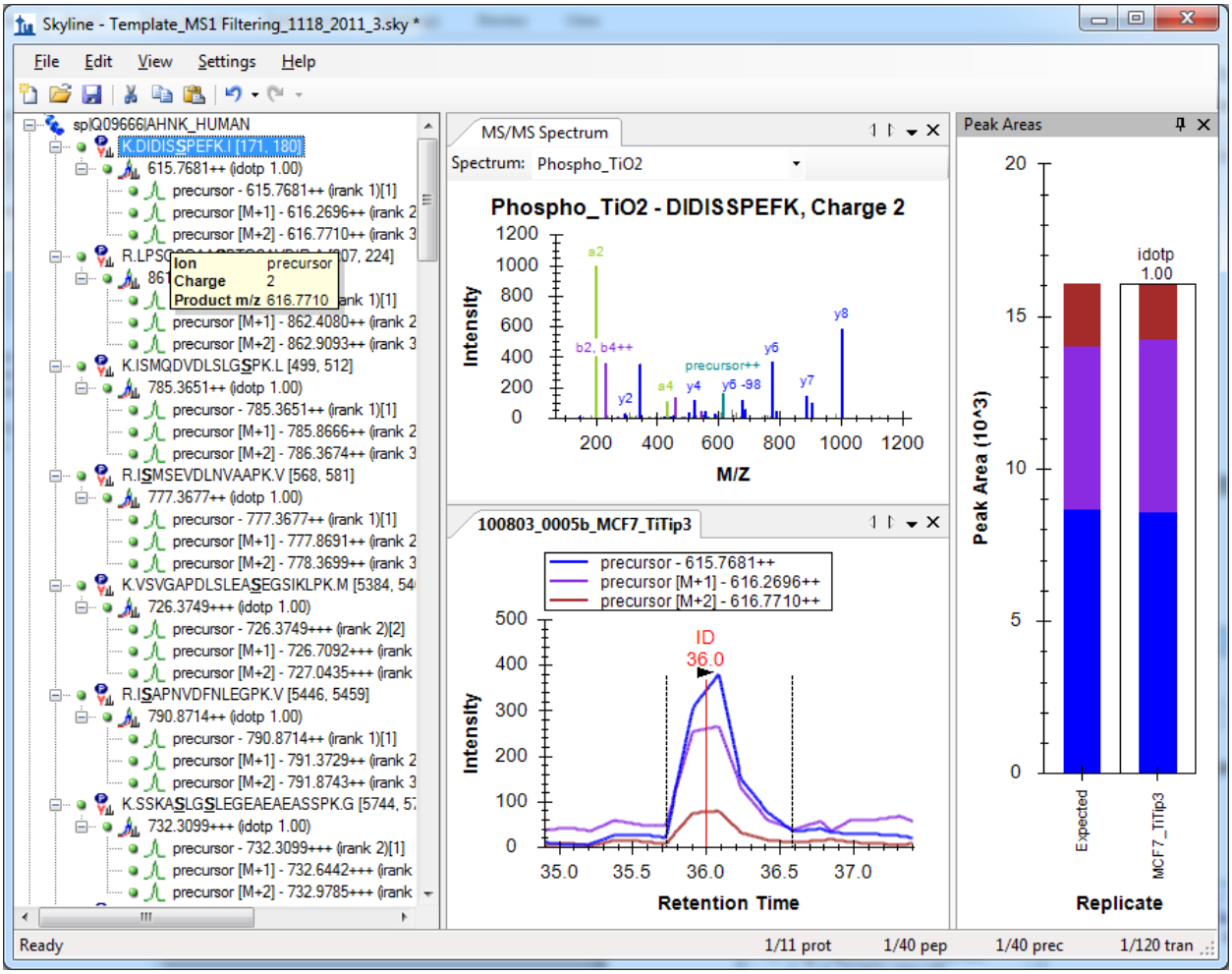
To configure a few other features that will be helpful in general, especially to visualize certain MS1 Filtering data, perform the following steps:

6. Ensure the first peptide, K.DIDIS**S**PEFK.I, is still selected in the peptide tree.
7. On the **Settings** menu, make sure **Integrate All** is checked.
8. On the **View** menu, choose **Peak Areas** and click **Replicate Comparison**.
9. In the **Peak Areas** window, right-click, choose **Normalize To** and click **None**.
10. In the **Peak Areas** window, right-click and check **Show Expected** and **Show Dot Product** (these two features will be explained below).



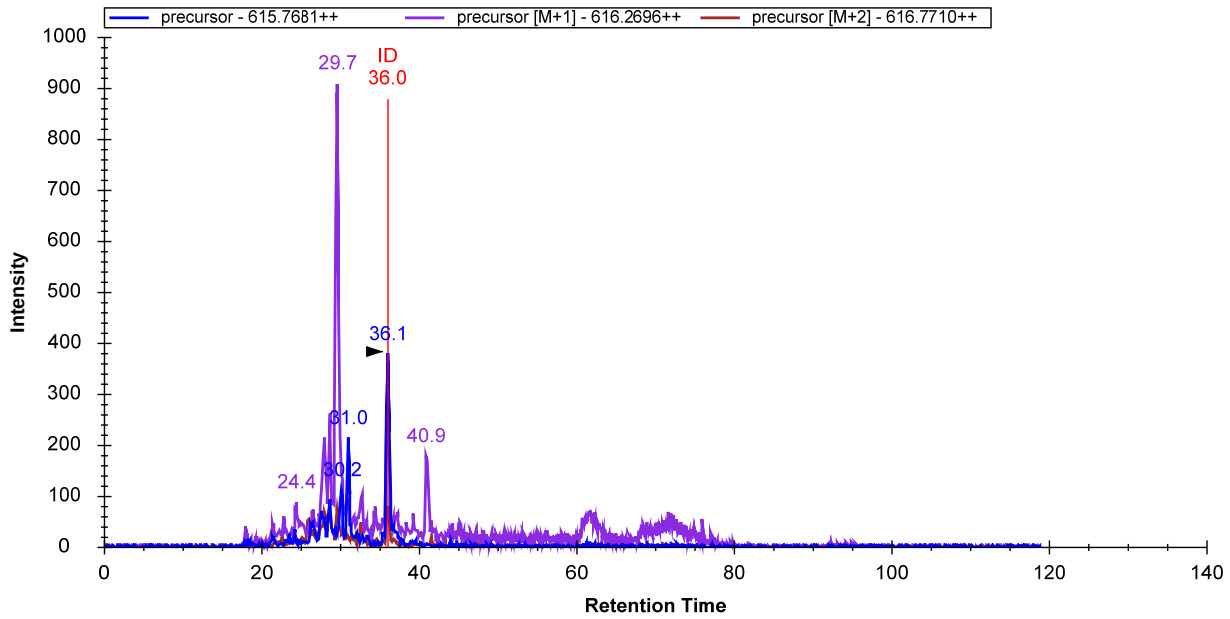
11. You can dock the **Peak Areas** window to your desired location by clicking and holding down the left mouse button, and then dragging it where you want to place it, perhaps the right edge of the Skyline window.

After importing the data, your Skyline file should look something like:



The chromatogram view shows the MS1 extracted ion chromatogram for all precursor isotope ions M (blue), M+1 (purple), M+2 (brown). Note that you will see a vertical line on the extracted ion chromatogram with the **ID** annotation on top. The **ID** stands for "**identification**" and indicates the retention time of a sampled MS/MS spectrum with confident identification for this particular peptide. If you click on the **ID** annotation, the **MS/MS Spectrum** view in the window above will show you the identified spectrum from the acquisition replicate, now stored in the redundant library you created in your library build. Also, the spectrum raw file and retention time will be selected in the **Spectrum** drop-down list at the top of the window **MS/MS Spectrum** view (instead of "best" spectrum from the non-redundant library Phospho_TiO2, which was selected before you clicked on the **ID** annotation). In this case, however, only one sampled MS/MS spectrum from a single file was identified as belonging to this peptide, and that spectrum is all there is to see.

Press Shift + F11 on the keyboard to obtain the following chromatogram graph over the entire gradient:

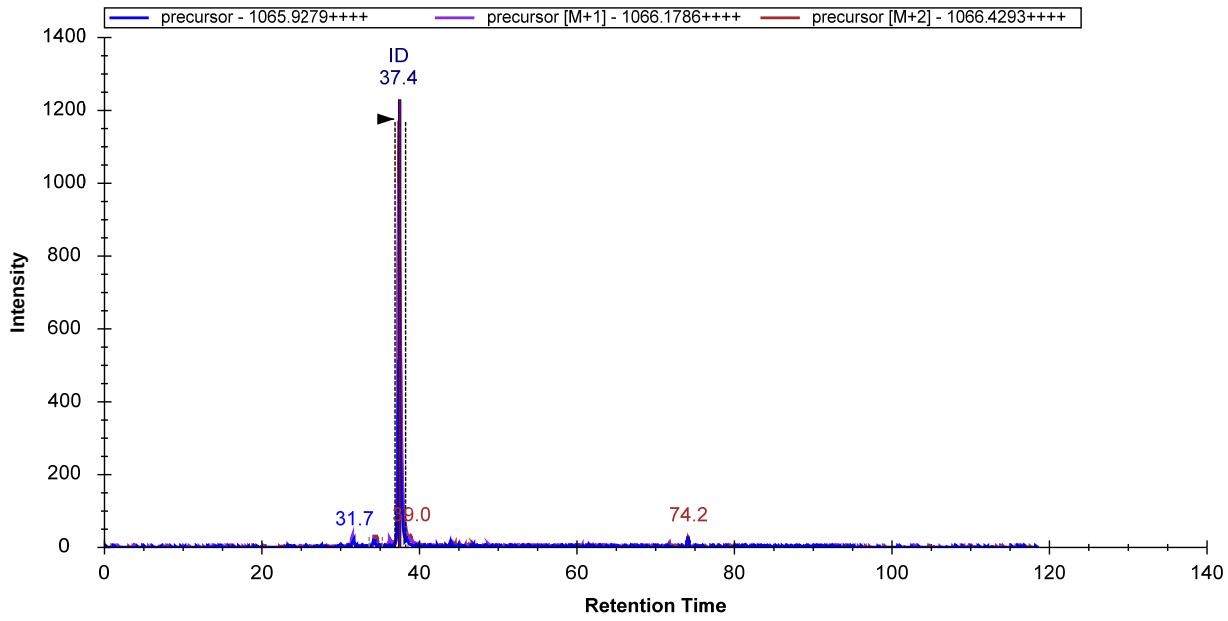


You can see how helpful the vertical ID line is for confidently assigning the correct MS1 eluting peak. During file import, identified MS/MS spectra are being used to direct MS1 peak picking and peak integration. This makes it extremely important that you build the spectral library before importing the raw data.

To discover more Skyline MS1 Filtering features and tools, you can jump to another peptide by doing the following:

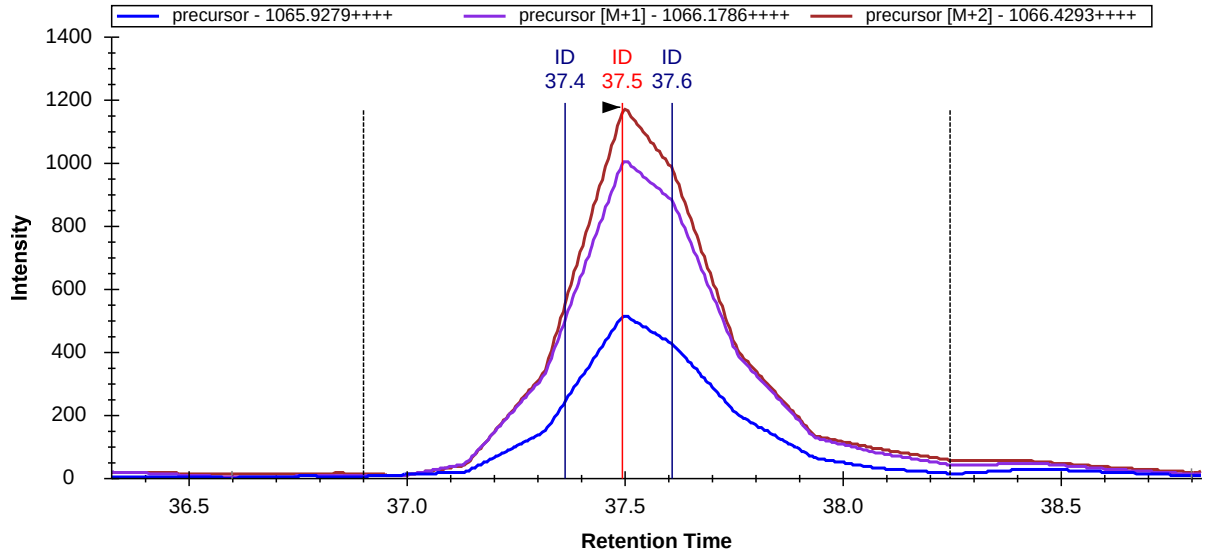
12. On the **Edit** menu, click **Find** (Ctrl-F).
13. In the **Find what** field, enter 'YGP'.
14. Click on the **Find Next** button.
15. Click the **Close** button on the **Find** dialog.
16. Click on the peptide sequence K.YGPADVEDTTGSGATDSKDDDDIDLFG**S**DDEEESEEAKR.L

In the zoomed-out chromatogram view, observe the clean ion chromatogram for this large peptide YGPADVEDTTGSGATDSKDDDDIDLFG**S**DDEEESEEAKR with a molecular weight > 4000 Da.



17. Press F11 to zoom in again.

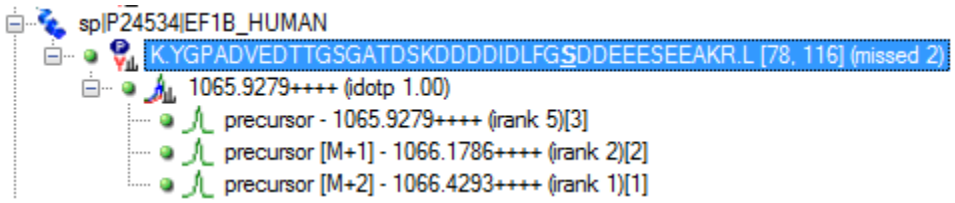
You can see that this is one peptide for which multiple collected MS/MS spectra were identified:



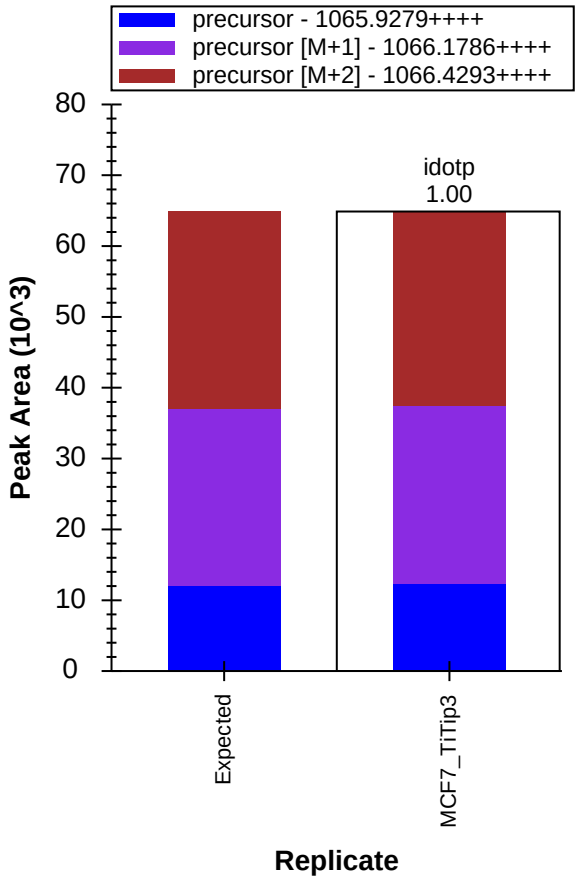
Click on the **ID** annotations, or the vertical lines, in the chromatogram plot to view the three spectra, or click on the **Spectra** dropdown list in the **MS/MS Spectrum** graph pane, and use arrow keys to page through the spectra.

Notice in the chromatogram plot that the third precursor ion isotope peak, M+2 (brown line), yields the most abundant peak area. It is typically important to filter more than just the first precursor ion isotope peak during MS1 filtering of high resolution MS1 scans. In the peptide tree for this peptide, note that the M+2 precursor with 1066.4293++++ shows an **irank** value of 1. This is the expected ranking by

isotopic abundance of the isotope peak. In square brackets to the right, the actual observed rank is shown, which also happens to be [1] in this case.



Another valuable tool to assess precursor ion isotope peak contributions is the peak area view shown on the right, which displays an **Expected** peak area precursor distribution that one can directly compare to the observed precursor ion peak area distribution. The isotope dot product (**idotp**) shown above the peak area bar in the peak area view compares observed and theoretical distributions, and should be 1.0 for a perfect match.



For MS1 Filtering, many other Skyline features can be used as described previously. Peak integration borders can be easily adjusted, and visual displays are available to help identify interferences or peak shoulders.

Note that Skyline can be used to import a large number of replicates. This tutorial has imported only one file in the interest of minimizing the size of the required download. When you are working with multi-replicate data sets, many of the features presented in the Existing and Quantitative Experiments tutorial may be used to similar advantage with MS1 filtered data.

Tuning Data Import for MS1 Filtering

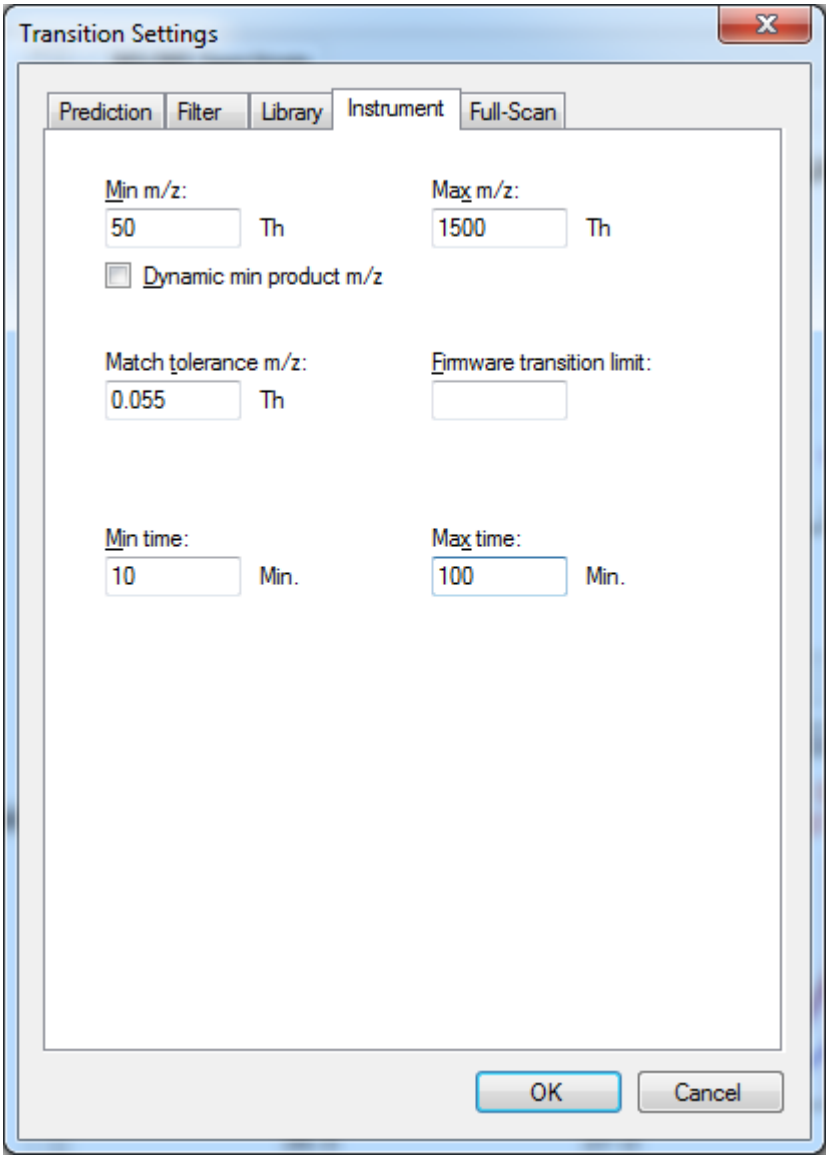
Skyline provides some additional features that can be helpful when importing raw data for MS1 filtering. For this data, unlike most SRM experiments, import speed and chromatogram data size are significant issues with multi-replicate data sets covering thousands of peptides.

Limiting the Chromatogram Extraction Time Range

Often during discovery type studies, MS1 data is still being acquired even during high organic column wash, or gradient equilibration at the end of a MS acquisition. Those MS1 chromatographic stretches typically do not contain any valuable peptides that need to be tracked by MS1 Filtering, and can be excluded from MS1 Filtering Import.

In this example, all observed peptides eluted between 10-100 min. You can specify this as follows:

1. On the **Settings** menu, click **Transition Settings.**
2. Click the **Instrument** tab.
3. In the **Min time** field enter '10' (minutes).
4. In the **Max time** field enter '100' (minutes).



5. Click the **OK** button.

This will not adjust the current file that has already been imported, but will apply to future imports of MS1 chromatographic information.

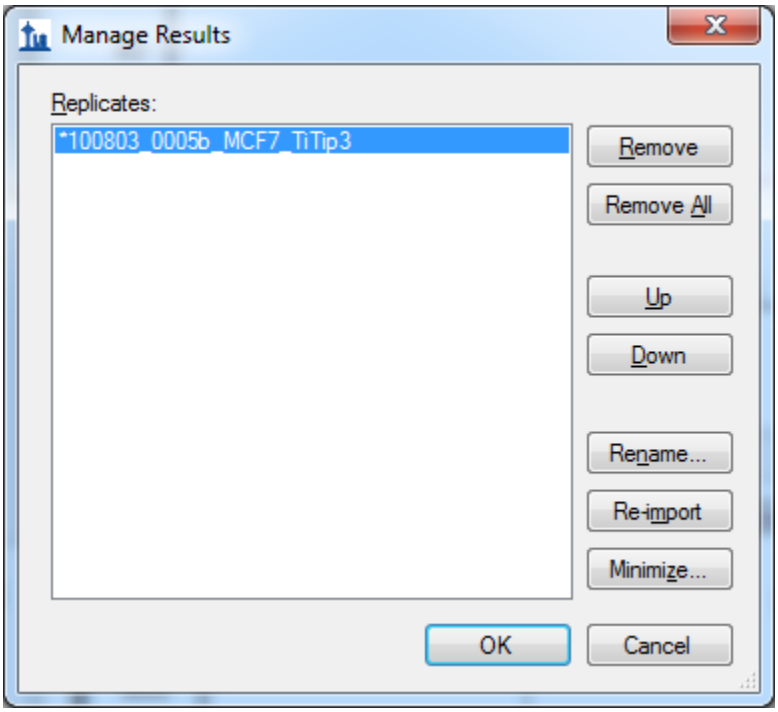
Re-importing Raw Data

If you have already imported a raw data file, but later realize that you forgot to perform MS1 filtering on a few peptides or you want to change settings like the chromatogram time range or the resolution used to filter, you can **Re-import** raw data into your existing Skyline document. The advantage of the **Re-import** feature, over simply removing the data from the document and using **File** > **Import** > **Results** to import it again, is that any manual edits you have made to integration boundaries and **Results Grid** annotations will be maintained. The **Re-import** feature can be used to add additional peptides from later mass spec runs and peptide searches to an existing document.

To **Re-import** the data file for this tutorial, with the new 10 to 100 minute time range in effect, perform the following steps:

1. On the **Edit** menu, click **Manage Results**.
2. Click the **Re-import** button.

An asterisk will be added before the selected named replicate, indicating it will be re-imported, if the changes are accepted. The **Manage Results** should look like this:



3. Click the **OK** button.

Skyline will begin to re-import the data, showing progress in green at the bottom of its window. After the re-import has completed, do the following to see that the chromatograms now range from 10 to 100 minutes:

1. Click on a new peptide.
2. Press Shift-F11 to zoom out.

You should see that the chromatogram points from 0 to 10 and 100 to 120 minutes are no longer included. In the future, Skyline will have more powerful ways of limiting time range during import, as this time range is a key factor in both import time and file size of the Skyline data cache file (.skyd). One early attempt is the **Filter only retention time scheduling windows**, which filters time ranges similar to the windows that would be used for a scheduled SRM experiment, with all the associated issues around peak truncation when the window is not accurate enough.

Because full-gradient chromatograms in MS1 Filtering experiments have been seen to result in Skyline data cache files ranging 5 Gigabytes and larger, another feature has been added to allow you to reduce

the chromatogram content in these files after you have completed a refinement and integration pass on the data.

Minimizing a Chromatogram Cache File

Before you start the next operation covered in this tutorial, you would normally want to review the integration boundaries for each of your peptides. You can do this fairly quickly for the 40 peptides targeted in this tutorial by doing the following:

1. On the **Edit** menu, choose **Collapse All** and click **Peptides**.
2. Select the first peptide in the document.
3. Use the Down-Arrow key to review each peptide.

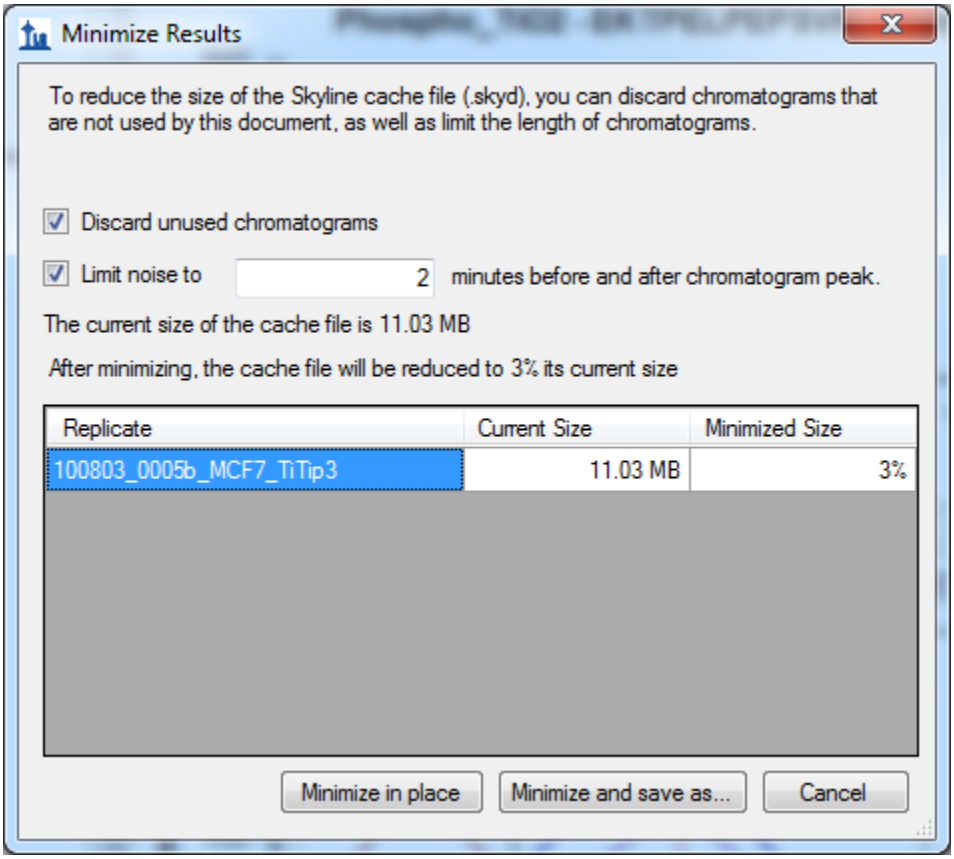
You should notice that there are a few cases of potential interference like K.ASLG**S**LEGEAEAEAS**S**PKGK.F or R.AEGEWEDQEALDYF**S**DKESGK.Q. You might delete these from your target list, and you might choose to correct integration on a few other peptides by clicking and dragging beneath the x-axis in the chromatogram chart.

For the most part, however, the data is pretty well integrated, and each peak contains a MS/MS spectrum ID for the peptide of interest. Things get a little more complicated with multi-replicate data where, because of the semi-random nature of the DDA method, not all peptides in your document will have identified MS/MS spectra for each sample injection. This is another area where Skyline will improve in the future, but where existing visual displays provide a great deal of help.

To eliminate of extraneous chromatogram data for this document, perform the following steps:

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

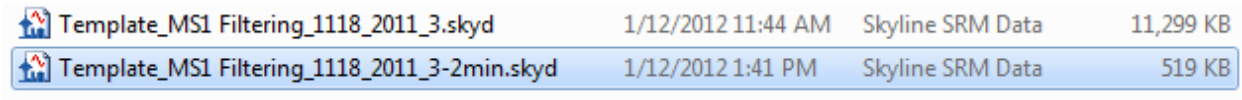
The **Minimize Results** form should now look like:



The form indicates that this operation is expected to reduce the size of the cache file from about 11 MB to 3% of its current size, or about 300 K.

6. Click the **Minimize and save as** button.
7. In the **File name** field of the **Save As** form, enter the name 'Template_MS1 Filtering_1118_2011_3-2min.sky'.
8. Click the **Save** button.

Once Skyline has completed this operation, bring up the Windows Explorer on the MS1Filtering folder, and you will see that the new .skyd file is actually 500 K, still quite a bit smaller:



If you again press 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 may 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 have the disadvantage of leaving peptides of interest unsampled by MS/MS or otherwise unidentified by peptide search tools. When this is combined with the fact that filtering even very narrow *m/z* ranges from high resolution MS1 data is less selective than sampling MS/MS product ions, picking the correct chromatogram peak for a peptide of interest becomes a challenge. 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. 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 for the tutorial document, you would perform the following steps:

1. On the **Settings** menu, click **Peptide Settings**.
2. Click the **Prediction** tab.
3. Check the **Use measured retention times when present** check box.
4. 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.
5. Click the **OK** button.
6. On the **File** menu, choose **Export** and click **Method**.
7. In the **Instrument type** dropdown list, choose 'AB SCIEX TOF'.
8. In the **Method type** dropdown list, choose 'Scheduled'.
9. In the **Template file** field, enter the path to a 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 be 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 through 2012. Even in Skyline version 1.2, however, the MS1 Filtering feature set gives proteomics investigators powerful new, cross-instrument platform tools for quantitative analysis of DDA data.